Richard Wong

Honors Multivariable Calculus

Math 32A/BH, UCLA

November 25, 2022

j

Preface

These lecture notes are written for Math 32A/BH, the honors multivariable calculus sequence at UCLA. Multivariable calculus is the mathematical language that allows us to describe the geometry of the physical world around us, such as the motion of planets in orbit, the behavior of electromagnetic forces, or the path of steepest ascent through the hills of Los Angeles.

The honors sequence differs from Math 32A/B in that it explores the topics of multivariable calculus with more mathematical rigor. In particular, in this course, you will explore not just the geometric meaning behind these definitions and theorems, but you will also learn why these definitions and theorems are true. I also plan to help you develop the reasoning and questioning skills needed to explore these mathematical ideas. Some questions you might see include the following:

> What changes, and what stays the same when we move from single variable calculus to multivariable calculus? What does it mean to take a derivative of a multivariable function?

Moreover, this course builds the foundation for more advanced topics in mathematics, such as linear algebra, real analysis, complex analysis, and differential geometry. This course is recommended for students interested in learning about advanced mathematics.

—

These lecture notes are intended to supplement the course lecture slides, and they will provide a mix of both computational and proof-based questions. They draw on material from the following texts:

- *Calculus and Analysis in Euclidean Space,* by Shurman.
- *Vector Calculus, Linear Algebra, and Differential Forms: A Unified Approach*, by Hubbard and Hubbard.
- *Multivariable Calculus*, by Rogawski and Adams.

You might find it convenient to obtain copies of these books for reference.

—

Please let me know if you notice any typos or errors, or if you have any comments or suggestions!

Contents

Chapter 1

Linear Algebra on $\mathbb{R}^n$

"Mathematics is the art of reducing any problem to linear algebra" - William Stein

Linear algebra is one of the foundational concepts in mathematics, and is essential to properly understanding multivariable calculus.

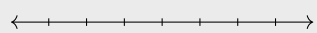
We will begin by studying certain properties of $\mathbb{R}^n$, and we will see how $\mathbb{R}^n$ and its properties can be generalized to vector spaces. We will then study how vectors spaces can be compared to each other, through linear maps.

1.1 An introduction to $\mathbb{R}^n$

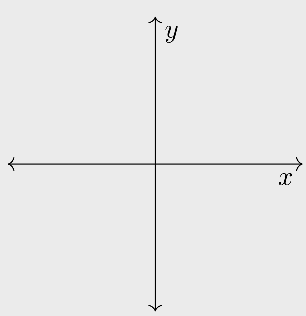
Definition 1.1.1. *Given a positive integer n, **n-dimensional Euclidean space** (denoted $\mathbb{R}^n$) is the set of all ordered n-tuples of real numbers. That is,*

$$\mathbb{R}^n := \{(x_1, \cdots, x_n) \mid x_i \in \mathbb{R}\}$$

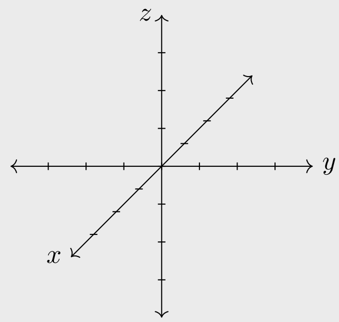
We are familiar with $\mathbb{R}^1$ (often described as the real line $\mathbb{R}$),



as well as $\mathbb{R}^2$ (often described as the Euclidean plane), which has the x and y axes.



$\mathbb{R}^3$ is often described as Euclidean space, and has 3 axes (denoted the x, y, and z axes).



We can study and understand n-dimensional Euclidean space in many different ways by imposing different *structures* on the set $\mathbb{R}^n$. For example, we might wish to consider $\mathbb{R}^n$ geometrically (more precisely, as a topological space).

Thinking geometrically, an element $\boldsymbol{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$ can be thought of as describing *coordinates* that tell us precisely where $\boldsymbol{x}$ is located inside of $\mathbb{R}^n$. In this framework, we will call elements of $\mathbb{R}^n$ points.

Definition 1.1.2. *A* ***point in*** $\mathbb{R}^n$ *(often denoted $\boldsymbol{x} \in \mathbb{R}^n$) is an element of $\mathbb{R}^n$, considered as a topological space. That is,*

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$$

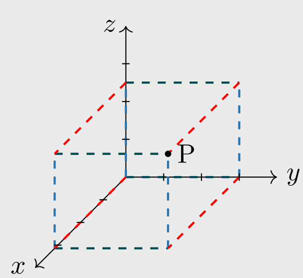
We say that x_1 is the ***first coordinate****, x_2 is the second coordinate, and so on, until x_n (which is the n-th coordinate).*

The point $O = (0, 0, \cdots, 0) \in \mathbb{R}^n$ is called the **origin** (in $\mathbb{R}^n$).

We can also consider the point $P = (\pi, 3, 2.5) \in \mathbb{R}^3$.

The first coordinate tells us that along the x-axis, we are π units in the positive direction from the origin; the second coordinate tells us we are 3 units in the positive direction along the y-axis, and the third coordinate tells us we are 2.5 units in the positive direction along the z-axis.

We can visualize P by drawing a box such that one corner is at the origin, the edges are aligned with the axes in $\mathbb{R}^3$, and the dimensions of the box correspond with the coordinates. P is then the corner of the box that is furthest from the origin.



On the other hand, we are not required to think of $\mathbb{R}^n$ in this way. Indeed, the structures that we impose on $\mathbb{R}^n$ should reflect how we want to think.

For example, we can instead think of an element $\boldsymbol{v} = (v_1, \cdots, v_n) \in \mathbb{R}^n$ not as describing coordinates, but rather, as describing ***displacement***. In other words, given a starting point P and an ending point

Q, we can measure the *change* in each coordinate. In this framework, we will call elements of $\mathbb{R}^n$ vectors:

Definition 1.1.3. *A* ***vector in*** $\mathbb{R}^n$ *(often denoted $\boldsymbol{v} \in \mathbb{R}^n$) is an element of $\mathbb{R}^n$, considered as a vector space. That is,*

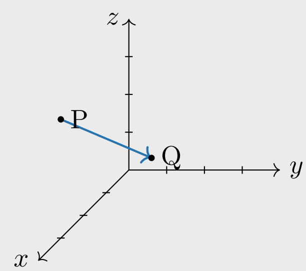
$$\boldsymbol{v} = \langle v_1, \cdots, v_n \rangle$$

We say that v_1 is the ***first component****, v_2 is the second component, and so on, until v_n (which is the n-th component).*

Suppose that you started at the point $P = (3, 0, \pi) \in \mathbb{R}^3$, and ended up at the point $Q = (4, 3, e) \in \mathbb{R}^3$. Then your total displacement along the x-axis is $4 - 3 = 1$ unit; $3 - 0 = 3$ units along the y-axis; and $e - \pi$ units along the z-axis. Then your total displacement can be described as a vector

$$\boldsymbol{PQ} = \langle 1, 3, e - \pi \rangle \in \mathbb{R}^3$$

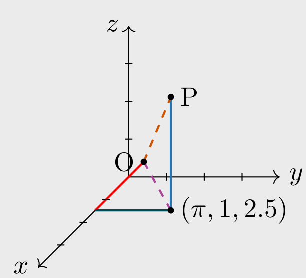
We can visualize the vector $\boldsymbol{PQ}$ as an arrow in $\mathbb{R}^3$, starting at P and ending at Q:



We say that the terminal point Q is the **head** of the vector $\boldsymbol{PQ}$, and that the initial point P is the **tail** of the vector $\boldsymbol{PQ}$

Given a vector $\boldsymbol{PQ}$, we can recover information about the relationship between the points P and Q. For example, we can calculate the distance between two points P and Q in $\mathbb{R}^3$ by iterating the Pythagorean theorem twice:

The distance between $P = (1, 1, 1)$ and $Q = (\pi, 3, 2.5)$ is $\sqrt{(\pi - 1)^2 + (3 - 1)^2 + (2.5 - 1)^2}$



Observe that the points P, $(\pi, 1, 1)$, and $(\pi, 1, 2.5)$ form a right triangle, whose hypotenuse has length $\sqrt{(\pi - 1)^2 + (2.5 - 1)^2}$. Similarly, the points P, $(\pi, 1, 2.5)$, and Q also form a right triangle, where the length of the hypotenuse is precisely the distance between P and Q.

Observe that the distance between $P = (1, 1, 1)$ and $Q = (\pi, 3, 2.5)$ is precisely the square root of the sum of the squares of the components of the vector $\boldsymbol{PQ}$. That is, if we write the components of $\boldsymbol{PQ} = \langle \pi - 1, 2, 1.5\rangle = \langle v_1, v_2, v_3\rangle$, then the distance between P and Q is

$$\sqrt{v_1^2 + v_2^2 + v_3^2}$$

This observation generalizes to $\mathbb{R}^n$, as we would need to iterate the Pythagorean theorem $n-1$ times. If we consider two points $P = (x_1, \cdots, x_n)$ and $Q = (y_1, \cdots, y_n)$, and write the components of the vector $\boldsymbol{PQ} = \langle x_1 - y_1, \cdots, x_n - y_n\rangle = \langle v_1, \cdots, v_n\rangle$, we again find that the distance between P and Q is

$$\sqrt{v_1^2 + \cdots + v_n^2}$$

Thus, we should give this quantity a name:

Definition 1.1.4. *Given a vector $\boldsymbol{v} = \langle v_1, \cdots, v_n\rangle$, its **magnitude** is given by*

$$||\boldsymbol{v}|| = \sqrt{\sum_{i=1}^{n} v_i^2}$$

Proposition 1.1.5. *The distance between two points $P = (x_1, \cdots, x_n)$ and $Q = (y_1, \cdots, y_n)$ is precisely $||\boldsymbol{PQ}||$.*

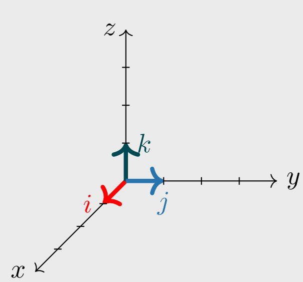
The vector $\boldsymbol{PQ} = \langle 1, 3, e - \pi\rangle \in \mathbb{R}^3$ has magnitude

$$||\boldsymbol{PQ}|| = \sqrt{1^2 + 3^2 + (e - \pi)^2}$$

Definition 1.1.6. *A **unit vector** in $\mathbb{R}^n$ is a vector $\boldsymbol{u} \in \mathbb{R}^n$ such that $||\boldsymbol{u}|| = 1$.*

In $\mathbb{R}^3$, there are special unit vectors, called the **standard basis vectors** of $\mathbb{R}^3$:

$$\boldsymbol{i} = \langle 1, 0, 0\rangle$$
$$\boldsymbol{j} = \langle 0, 1, 0\rangle$$
$$\boldsymbol{k} = \langle 0, 0, 1\rangle$$



The vector $\boldsymbol{e_{PQ}} = \left\langle \frac{1}{\sqrt{(e-\pi)^2+10}}, \frac{3}{\sqrt{(e-\pi)^2+10}}, \frac{e-\pi}{\sqrt{(e-\pi)^2+10}} \right\rangle \in \mathbb{R}^3$ is a unit vector.

What are the differences between the points perspective and the vectors perspective?

Though the two perspectives might seem similar, there are important differences:

For example, from the points perspective, we should think of elements in $\mathbb{R}^n$ as being analogous to *locations* on a map. It **doesn't make sense** to add the location of Los Angeles, California (say, coordinates with latitude 34° N, longitude 118° W) to the location of Austin, Texas (say, coordinates 30° N, 98° W).

On the other hand, from the vectors perspective, we should think of elements in $\mathbb{R}^n$ as being analogous to displacements. In this case, let us consider two displacements: The displacement from Los Angeles to Austin means travelling 4° south, and travelling 20° east. Similarly, the displacement from Austin to New York City (coordinates 41° N, 74° W) means travelling 11° north, and travelling 24° east.

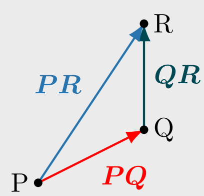
It **does make sense** to add these two displacements together, which entails travelling a total of 7° north, and 44° east. Observe that the result of this addition is the **total displacement** from Los Angeles to New York City (travelling 7° north, and 44° east).

We can formulate this in terms of vectors as follows:

Definition 1.1.7 (Vector addition)**.** *Given two vectors* $\boldsymbol{u} = \langle u_1, \cdots, u_n \rangle$, $\boldsymbol{v} = \langle v_1, \cdots, v_n \rangle$, *we can add two vectors together to form the sum* $\boldsymbol{u} + \boldsymbol{v} \in \mathbb{R}^n$. *In terms of components,*

$$\boldsymbol{u} + \boldsymbol{v} = \langle u_1 + v_1, \cdots, u_n + v_n \rangle$$

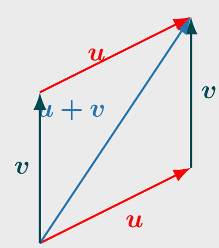
There is a nice geometric picture that accompanies vector addition: Given two vectors $\boldsymbol{PQ}$ and $\boldsymbol{QR}$, then the sum of these two vectors is the vector $\boldsymbol{PR}$ (that is, the vector with tail P, and head R)



That is, we add vectors "tip to tail".

Proposition 1.1.8. *Vector addition is commutative. That is, the vectors* $\boldsymbol{u} + \boldsymbol{v}$ *and* $\boldsymbol{v} + \boldsymbol{u}$ *are the same vector.*

This follows immediately from the definition in terms of components. However, there is a nice geometric picture that accompanies this property:



That is, if we move our vectors freely, the $\boldsymbol{u} + \boldsymbol{v}$ is the diagonal of the parallelogram spanned by $\boldsymbol{u}$ and $\boldsymbol{v}$.

Let us be more precise about moving vectors freely: we can combine the points perspective with the vector perspective. Using our previous analogies, it makes sense to add the add the *displacement* from Los Angeles to Austin to the *location* New York City.

We should interpret this as saying to start at the coordinates 41° N, 74° W, and to then travel 4° south, and 20° east. The end result is a **location** with coordinates 37° N, 54° W (somewhere in the middle of the Atlantic Ocean!).

We can formulate this in terms of vectors as follows:

Definition 1.1.9. *In terms of coordinates, given a point $P = (x_1, \cdots, x_n) \in \mathbb{R}^n$, and a vector $\boldsymbol{v} = \langle v_1, \cdots, v_n \rangle \in \mathbb{R}^n$, then we can define the **point** $P + \boldsymbol{v}$ to be the point with coordinates*

$$P + \boldsymbol{v} = (x_1 + v_1, \cdots, x_n + v_n)$$

Informally, the point $P + \boldsymbol{v}$ is the tail of the vector that starts at P, and travels along the length of $\boldsymbol{v}$. We can also make this mathematically precise through vector addition:

Definition 1.1.10. *A **position vector** in $\mathbb{R}^n$ is a vector whose tail (or initial point) is the origin.*

If the head of the position vector is the point P, then we often denote the position vector as the vector $\boldsymbol{OP}$.

Proposition 1.1.11. *Given a point $P \in \mathbb{R}^n$, and a vector $\boldsymbol{v} \in \mathbb{R}^n$, the point $P + \boldsymbol{v}$ has the same coordinates as the head (or terminal point) of the vector $\boldsymbol{OP} + \boldsymbol{v}$.*

We are now able to describe various vectors at different locations in $\mathbb{R}^n$. We have seen cases where two vectors indeed start and end at the same point, but in definition 1.1.9, we now see that it is possible for "the same vector" to have different starting points. This leads us to the following question:

When are two vectors the same?

We first must define what we mean for two vectors to be the same:

Definition 1.1.12. *A vector $\boldsymbol{w} = \boldsymbol{AB}$ is said to be **equivalent** to a vector $\boldsymbol{v} = \boldsymbol{PQ}$ if the terminal point B is the same as the point $A + \boldsymbol{v}$*

*We sometimes also say that $\boldsymbol{w}$ is a **translation** of $\boldsymbol{v}$.*

Suppose that we have a vector $\boldsymbol{v} = PQ \in \mathbb{R}^n$, and suppose that the components of $\boldsymbol{v}$ are $\boldsymbol{v} = \langle v_1, \cdots, v_n \rangle$.

Consider the point $P = (v_1, \cdots, v_n)$. Then $\boldsymbol{v}$ is equivalent to the position vector $\boldsymbol{OP}$, which starts at the origin O, and ends at the point P.

The example above is the key idea of the proof of the following proposition:

Proposition 1.1.13. *Consider the points $A, B, P, Q \in \mathbb{R}^n$. The vectors $\boldsymbol{AB}$ and $\boldsymbol{PQ}$ are equivalent if and only if they have the same components.*

We now turn to a second way to distinguish vectors:

Definition 1.1.14. *A **scalar** (often denoted $\lambda \in \mathbb{R}$) is an element of the field of real numbers, $\mathbb{R}$.*

Definition 1.1.15. *Given a vector $\boldsymbol{v} = \langle v_1, \cdots, v_n \rangle \in \mathbb{R}^n$ and a scalar $\lambda \in \mathbb{R}$, the **scalar multiple** $\lambda \boldsymbol{v}$ is the vector*

$$\lambda \boldsymbol{v} = \langle \lambda v_1, \cdots, \lambda v_n \rangle$$

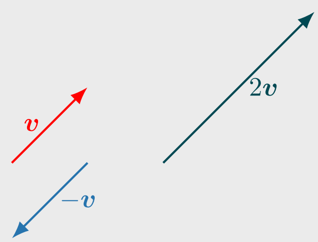
That is, we multiply λ with each component.

We can interpret scalar multiples geometrically through the following calculation:

Proposition 1.1.16. *The magnitude of a scalar multiple $\lambda \boldsymbol{v}$ is the same as the magnitude of $\boldsymbol{v}$ times the absolute value of λ. That is,*

$$||\lambda \boldsymbol{v}|| = |\lambda|(||\boldsymbol{v}||)$$

This tells us that scalar multiplication changes the magnitude of a vector (that is, it *scales* it by a factor of $|\lambda|$).



Definition 1.1.17. *Two vectors $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ are said to be **parallel** if they are scalar multiples of each other. That is, there exists a scalar $\lambda \in \mathbb{R}$ such that*

$$\boldsymbol{v} = \lambda \boldsymbol{w}$$

Consider the points $P = (3, -2, 1)$ and $Q = (2, -1, 1)$ in $\mathbb{R}^3$, and consider the vectors $\boldsymbol{u} = \langle -1, 1, 0 \rangle$, $\boldsymbol{v} = \langle -2, 2, 0 \rangle$, $\boldsymbol{w} = \langle 1, -1, 0 \rangle$, $\boldsymbol{a} = \langle 1, -1, 1 \rangle$.

- The vector $\boldsymbol{PQ}$ is equivalent to $\boldsymbol{u}$.
- The vector $\boldsymbol{PQ}$ is parallel to $\boldsymbol{v}$, with $\lambda = 2$.
- The vector $\boldsymbol{PQ}$ is parallel to $\boldsymbol{w}$, with $\lambda = -1$.
- The vector $\boldsymbol{PQ}$ is **not** parallel to $\boldsymbol{a}$. That is, there is no λ such that $\boldsymbol{PQ} = \lambda \boldsymbol{a}$.

  If such a λ, existed, that would require that the equations $-1 = \lambda 1$ and $0 = \lambda 1$ to both be true. But that would be a contradiction, since λ cannot be both -1 and 0.

Parallel vectors are so named because they give rise to parallel lines. See definition 1.1.22 in the exercises to learn how we describe lines in $\mathbb{R}^n$, and exercise 1.12.

Observe that in the example above, we saw that two vectors can be parallel, yet point in different directions. We can make this idea precise:

Definition 1.1.18. *Given a non-zero vector $\boldsymbol{v} \in \mathbb{R}^n$, its* ***direction vector*** $\boldsymbol{e_v}$ *is the unit vector such that*

$$||\boldsymbol{v}||\boldsymbol{e_v} = \boldsymbol{v}$$

Note that we used the word "the" in the definition above - in order for this definition to make sense, we should make sure to prove that the direction vectors is *unique*.

Proposition 1.1.19. *Given a non-zero vector $\boldsymbol{v} \in \mathbb{R}^n$, its* ***direction vector*** $\boldsymbol{e_v}$ *is unique. That is,*

$$\boldsymbol{e_v} = \frac{1}{||\boldsymbol{v}||}\boldsymbol{v}$$

Thus, we can now algebraically explain the difference between the vectors $\boldsymbol{PQ}$ and $\boldsymbol{QP}$:

Definition 1.1.20. *Suppose that the vector $\boldsymbol{w}$ is parallel to $\boldsymbol{v}$, with $\boldsymbol{v} = \lambda \boldsymbol{w}$. If $\lambda > 0$, then we say $\boldsymbol{w}$ has the* ***same direction*** *as $\boldsymbol{v}$. If $\lambda < 0$, then $\boldsymbol{w}$ has the* ***opposite direction*** *as $\boldsymbol{v}$.*

Proposition 1.1.21. *Two vectors are equivalent if and only if they have the same magnitude and the same direction vector.*

In other words, if two vectors have different magnitudes, then they must be different vectors. Similarly, if two vectors have different direction vectors, then they must be different vectors.

To summarize, we have learned the following:

A vector in $\mathbb{R}^n$ is determined equivalently by:

a. an initial point and a terminal point.

$$v = \boldsymbol{PQ}$$

b. its components.

$$\boldsymbol{v} = \langle x_1, \cdots, x_n \rangle$$

c. a direction and magnitude.

$$\boldsymbol{v} = ||\boldsymbol{v}||\boldsymbol{e_v}$$

Exercises

1.1 Explain the relationship between points and vectors.

More concretely, given a point in $\mathbb{R}^n$, is there a natural way to obtain a vector in $\mathbb{R}^n$? Conversely, given a vector in $\mathbb{R}^n$, is there a natural way to obtain a point in $\mathbb{R}^n$?

1.2 Sketch and find the components of the unit position vector in $\mathbb{R}^2$ that makes an angle of 30 degrees counterclockwise from the positive x-axis.

1.3 Sketch and find the components of the unit position vector in $\mathbb{R}^2$ that makes an angle of 30 degrees counterclockwise from the negative x-axis.

a. Is this vector equivalent to the vector in exercise 1.2?

b. Is this vector parallel to the vector in exercise 1.2?

1.4 Find the components of the unit position vector in $\mathbb{R}^2$ that makes an angle of θ degrees counterclockwise from the negative x-axis.

1.5 Are these vectors parallel to the vector $\boldsymbol{v} = \langle 2, -1, 1\rangle$? Are they equivalent?

a. $\boldsymbol{PQ}$, where $P = (1, 1, 0)$, and $Q = (3, 0, 1)$.

b. $\boldsymbol{QP}$, where $P = (1, 1, 0)$, and $Q = (3, 0, 1)$.

c. $\boldsymbol{AB}$, where $A = (1, 1, 0)$, and $B = (5, -1, 2)$.

d. $\boldsymbol{PQ} + 2\boldsymbol{AB}$, where $\boldsymbol{PQ}$ and $\boldsymbol{AB}$ are as above.

e. $\boldsymbol{PQ} + 2\langle 1, 0, 0\rangle$, where $\boldsymbol{PQ}$ is as above.

1.6 Let $\boldsymbol{u} = \langle 1, 1, 0\rangle$, and let $\boldsymbol{v} = \langle 0, 0, 1\rangle$.

a. Compute $\boldsymbol{u} - \boldsymbol{v}$.

b. Sketch $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{u} - \boldsymbol{v}$.

c. In general, what is the relationship between $\boldsymbol{u} - \boldsymbol{v}$ and $\boldsymbol{v} - \boldsymbol{u}$?

1.7 Given four points, A, B, C, and D in $\mathbb{R}^3$, we obtain a quadrilateral $ABCD$. If $ABCD$ is a parallelogram, and the coordinates of the corners are $A = (1, 0, 1)$, $B = (3, 3, 2)$, and $C = (2, 4, 5)$, then what are the coordinates of D?

(**Hint:** Recall that $ABCD$ is given by a line from A to B to C to D, and then back to A. Use vectors to solve this problem!).

1.8 Consider the parallelogram $ABCD$. Prove that the diagonals BC and AD intersect each other at their midpoints.

(**Hint**: Describe the diagonals as vectors).

In the following exercises, we will describe lines in $\mathbb{R}^n$.

Definition 1.1.22. *The line $\mathscr{L}$ in $\mathbb{R}^n$, passing through the point $P = (x_1, \cdots x_n)$, in the direction of the vector $\boldsymbol{v} = \langle v_1, \cdots v_n\rangle$, can be described by the vector-valued function $\boldsymbol{r}(t) : \mathbb{R} \to \mathbb{R}^n$ defined by*

$$\boldsymbol{r}(t) = \boldsymbol{r_0} + t\boldsymbol{v}$$

where $\boldsymbol{r_0}$ is the vector $\boldsymbol{r_0} = \boldsymbol{OP} = \langle x_1, \cdots x_n\rangle$.
*We call $\boldsymbol{r}(t)$ the **vector parametrization of** $\mathscr{L}$.*

Observe that by definition, the vector $\boldsymbol{r}(t)$ is a position vector for all t. The tips of these position vectors are precisely the points on the line $\mathscr{L}$.

1.9 Recall that we can describe a line in $\mathbb{R}^2$ as the set of points $L = \{(x, y) \mid y = mx + b\}$. Find a vector parametrization of the line L.

a. What is the relationship between the constant b and the vector $\boldsymbol{r_0}$?

b. What is the relationship between the slope m and the direction vector $\boldsymbol{v}$?

1.10 Find the vector parametrization of the line that passes through the points $P = (1, 0, 2)$ and $Q = (2, 5, -1)$.

1.11 Given a point P and a direction vector $\boldsymbol{v}$, is the vector parametrization $\boldsymbol{r}(t) = \boldsymbol{r_0} + t\boldsymbol{v}$ of a line $\mathscr{L}$ unique?

a. Is there a vector parametrization of $\mathscr{L}$ using the same direction vector $\boldsymbol{v}$ but a different starting point Q? If so, find the coordinates of a possible Q.

b. Is there a vector parametrization of $\mathscr{L}$ using the same starting point P, but a different direction vector, $\boldsymbol{w}$? If so, find the components of a possible $\boldsymbol{w}$.

c. Is there a vector parametrization of $\mathscr{L}$ using a different direction vector $\boldsymbol{w}$ and a different starting point Q? If so, find a possible Q and $\boldsymbol{w}$.

1.12 Suppose that $\boldsymbol{r_1}(t) = \boldsymbol{r_0} + t\boldsymbol{v}$ and $\boldsymbol{r_2}(s) = \boldsymbol{r_0} + s\boldsymbol{w}$ are two parametrizations of the same line. What is the relationship between $\boldsymbol{v}$ and $\boldsymbol{w}$?

1.13 When are $\boldsymbol{r_1}(t) = \boldsymbol{OP} + t\boldsymbol{v}$ and $\boldsymbol{r_2}(s) = \boldsymbol{OQ} + s\boldsymbol{v}$ are two parametrizations of the same line?

Definition 1.1.23. *We say that two lines $\boldsymbol{r_1}(t)$ and $\boldsymbol{r_2}(s)$ **intersect** if there is a point P lying on both lines.*

1.14 Determine whether the following two lines intersect. If they do, find the common point of intersection.

$$\boldsymbol{r_1}(t) = \langle 3, 2, 6\rangle + t\langle 1, 0, -1\rangle$$
$$\boldsymbol{r_2}(s) = \langle 4, 12, 24\rangle + s\langle -1, 10, 20\rangle$$

(**Hint:** If a point lies on both lines, then that point must satisfy both equations).

1.15 Determine whether the following two lines intersect. If they do, find the common point of intersection.

$$\boldsymbol{r_1}(t) = \langle 3, 2, 6\rangle + t\langle 1, 0, -1\rangle$$
$$\boldsymbol{r_2}(s) = \langle 4, 5, 1\rangle + s\langle -1, 3, 3\rangle$$

1.16 Suppose that you have two lines with parametrizations $\boldsymbol{r_1}(t)$ and $\boldsymbol{r_2}(s)$, respectively. What is the difference between solving the vector equation $\boldsymbol{r_1}(t) = \boldsymbol{r_2}(s)$ versus solving the vector equation $\boldsymbol{r_1}(t) = \boldsymbol{r_2}(t)$?

Definition 1.1.24. *We say that two distinct lines are **skew**, if they do not intersect each other, and they are not parallel to each other.*

1.17 Write down parametrizations of two distinct lines that are skew to each other.

1.2 Vector spaces

How can we generalize and abstract these properties of $\mathbb{R}^n$?

In these lecture notes, we will only study vector spaces whose field of scalars is $\mathbb{R}$ - that is, we are only considering *real* vector spaces[1]. So in these notes we will omit the adjective real and simply refer to them as vector spaces.

Definition 1.2.1. *A **vector space** V is a set with a distinguished element $\mathbf{0} \in V$, together with two operations, $+ : V \times V \to V$, and $\cdot : \mathbb{R} \times V \to V$ satisfying the following axioms for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$, and for all $\lambda, \alpha \in \mathbb{R}$.*

(i) $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$ *(Additive associativity)*

[1] That is, scalars must be elements of the field $\mathbb{R}$ (as in definition 1.1.14). In other classes, such as linear algebra, you will see that we can generalize by replacing the field $\mathbb{R}$ with any other field $\mathbb{F}$, and a lot of the theory of vector spaces and linear algebra still work over an arbitrary field $\mathbb{F}$.

(ii) $\boldsymbol{v} + \mathbf{0} = \mathbf{0} + \boldsymbol{v} = \boldsymbol{v}$ *(Additive identity)*

(iii) *For all* $\boldsymbol{v} \in V$, *there exists* $\boldsymbol{w} \in V$ *such that* $\boldsymbol{v} + \boldsymbol{w} = \mathbf{0}$ *(Additive inverse)*

(iv) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$ *(Additive commutativity)*

(v) $\lambda(\alpha \boldsymbol{v}) = (\lambda\alpha)\boldsymbol{v}$ *(Scalar associativity)*

(vi) $1\boldsymbol{v} = \boldsymbol{v}$ *(Scalar identity)*

(vii) $(\lambda + \alpha)\boldsymbol{u} = \lambda\boldsymbol{u} + \alpha\boldsymbol{u}$ *(Distribution over scalar addition)*

(viii) $\lambda(\boldsymbol{u} + \boldsymbol{v}) = \lambda\boldsymbol{u} + \lambda\boldsymbol{v}$ *(Distribution over vector addition)*

The elements of V *are called vectors, and elements of* $\mathbb{R}$ *are called scalars*[a].

[a]see footnote 1

When faced with an abstract definition like this, it's especially important to first find a few concrete examples to keep in mind. Luckily, from our discussion in the previous section, we have in fact checked most of these axioms:

Theorem 1.2.2. $\mathbb{R}^n$ *is a vector space.*

$\mathbb{R}^n$ is the prototypical example of a vector space[2]. What this means is that whenever you are are trying to determine whether or not something is true about vector spaces, you should try to see if it is true for $\mathbb{R}^n$. Then, you should try to translate the proof in the $\mathbb{R}^n$ case to the general case, purely using the axioms and properties of abstract vector spaces.

However, you should also be careful not to trick yourself into thinking that every vector space behaves like $\mathbb{R}^n$! Below is an example of a vector space that is quite different from $\mathbb{R}^n$. You will verify the axioms in exercise 1.29.

Theorem 1.2.3. *The set* $\mathscr{F}(\mathbb{R})$ *of functions* $f : \mathbb{R} \to \mathbb{R}$ *is a vector space with pointwise addition and pointwise scalar multiplication.*

$$(f+g)(x) := f(x) + g(x) \qquad (\lambda f)(x) := \lambda(f(x))$$

With these examples in mind, we can now look back at the axioms of a vector space (definition 1.2.1) and see *why* these axioms are important.

Axioms (i) through (iv) describe how to *add* vectors using the operation $+ : V \times V \to V$. From previous experiences with addition, we should require that $+$ should be (i) associative and (iv) commutative. Moreover, we also require that the zero vector $\mathbf{0} \in V$ should behave the way that $0 \in \mathbb{R}$ does, with regards to (ii) adding zero, as well as (iii) the existence of additive inverses (e.g. negative numbers in $\mathbb{R}$).

Axioms (v) through (vii) describe how to *scalar multiply* vectors using the operation $\cdot : \mathbb{R} \times V \to V$. The intuition for these axioms comes from our knowledge of multiplication for real numbers, which is (v) associative and (vii) distributes over addition of real numbers. Moreover, the element $1 \in \mathbb{R}$ is the *multiplicative identity* - that is, for any element $a \in \mathbb{R}$, we have that $1(a) = a$. This property of $1 \in \mathbb{R}$ is mirrored in axiom (vi).

[2]In fact, in exercise 1.51, you can show that any **finite dimensional** vector space is isomorphic to $\mathbb{R}^n$.

Warning: It is important to note, however, that scalar multiplication is **not** multiplication of vectors. Indeed, scalar multiplication is a function $\mathbb{R} \times V \to V$. That is, its input is precisely one scalar and one vector, and whose output is a vector. You will compare this to multiplication in exercise **??**.

In Section 1.4, we will discuss operations whose inputs are two vectors, and whose output is a vector.

Vector subspaces

Let us now turn to other examples of vector spaces.

Consider the set

$$W = \{(x, 0, 0) \in \mathbb{R}^3\}$$

The first observation to make is that W is a subset of $\mathbb{R}^3$ - that is, every element of W is also contained inside of $\mathbb{R}^3$.

Now, if we take any two elements in $(x_1, 0, 0), (x_2, 0, 0) \in W$, we can then add them together as elements of the vector space $\mathbb{R}^3$ to get the vector $(x_1 + x_2, 0, 0) \in \mathbb{R}^3$. Observe that the sum $(x_1 + x_2, 0, 0)$ is still an element of W - thus the vector addition in $\mathbb{R}^3$ naturally defines a vector addition on W.

Furthermore, if we take an element $(x_1, 0, 0) \in W$ and a scalar $\lambda \in \mathbb{R}$, we can again treat $(x_1, 0, 0)$ as an element in $\mathbb{R}^3$, and form the scalar multiple $\lambda((x_1, 0, 0)) = (\lambda x_1, 0, 0)$. Once again, the scalar multiple $\lambda((x_1, 0, 0))$ is still an element of W - thus the scalar multiplication in $\mathbb{R}^3$ naturally defines a scalar multiplication on W.

Thus, we have shown that if we define vector addition and scalar multiplication on W as coming from $\mathbb{R}^3$, it then follows that W satisfies the axioms a vector space.

We can now give a name to this phenomenon of a subset of a vector space turning out to be a vector space:

Definition 1.2.4. *A nonempty subset W of a vector space V, is a* ***vector subspace*** *of V if for all $v, w \in W$ and for all $\lambda \in \mathbb{R}$,*

(i) $\boldsymbol{v} + \boldsymbol{w} \in W$ *(closure under addition)*

(ii) $\lambda(\boldsymbol{v}) \in W$ *(closure under scalar multiplication)*

Again, to keep ourselves from thinking only in terms of $\mathbb{R}^n$, here is another vector subspace example:

Theorem 1.2.5. *The set of $\mathscr{C}(\mathbb{R})$ of continuous functions $f : \mathbb{R} \to \mathbb{R}$ is a vector space with pointwise addition and pointwise scalar multiplication.*

$$(f + g)(x) := f(x) + g(x) \qquad (\lambda f)(x) := \lambda(f(x))$$

Theorem 1.2.6. *$\mathscr{C}(\mathbb{R})$ is a subspace of $\mathscr{F}(\mathbb{R})$.*

Warning: It is not true that every subset of a vector space is a vector space. For a sophisticated example, see exercise **??**. For a more immediate example, consider the following:

The set of two elements $A = \{(1, 0), (0, 1)\}$ is a subset of $\mathbb{R}^2$, but is not a vector subspace of $\mathbb{R}^2$, as it fails to be closed under both addition and scalar multiplication.

In fact, we can generalize this example:

Proposition 1.2.7. *If A is a finite set of elements in a vector space contains a non-zero vector, then A cannot be a vector subspace*[a].

[a]Remember, we are considering only real vector spaces in these notes.

Proof. One proof, using vector addition, is as follows: Suppose that A be a finite set of elements in a vector space with cardinality $|A| = n$, and suppose that $\boldsymbol{a} \in A$ is a non-zero vector. Then consider the sums $\sum_{i=1}^{j} \boldsymbol{a}$ for $1 \leq j \leq n + 1$. Since $\boldsymbol{a}$ is non-zero, we have described $n + 1$ different vectors, and so at least one of the sums $\sum_{i=1}^{j} \boldsymbol{a}$ cannot be contained in A. Thus A cannot be a vector subspace.

A variation on this proof can be phrased using scalar multiplication: Let W be a vector subspace of V, and let $\boldsymbol{w} \in W$ be a non-zero vector. Then since vector subspaces are closed under scalar multiplication, W contains $\lambda \boldsymbol{w}$ for all $\lambda \in \mathbb{R}$. Moreover, since $\boldsymbol{w}$ is non-zero, we have shown that W must have at least as many elements as $\mathbb{R}$. In other words, we have described an (uncountably) infinite number of elements in W, and so any finite set of vectors in a vector space containing a non-zero vector cannot be a vector subspace.

□

To phrase this in another way, we have proved that a vector subspace with finitely many elements must be the **zero** vector space (see exercise 1.23). However, one might wonder if we can fix this example in some way:

Given a (finite) set of elements A in a vector space V, can we produce a vector subspace? Even better, can we produce the **smallest** vector space containing A (that is, without including any extraneous vectors)?

We see from our observations above to construct the smallest vector space containing A, we must first include all scalar multiples of elements in A. Once we have done so, we then need to include any possible sum of scalar multiples. This is called a **linear combination**:

Definition 1.2.8. *Let V be a vector space. A **linear combination** of vectors $\boldsymbol{v_1}, \cdots, \boldsymbol{v_k} \in V$ is a vector of the form*

$$\lambda_1 \boldsymbol{v_1} + \cdots + \lambda_k \boldsymbol{v_k}$$

where the λ_i are scalars.

Thus, we now have a candidate set for the smallest vector space containing A: the set of all linear combinations of vectors in A.

Definition 1.2.9. *Let V be a vector space, and let A be a subset of vectors in V. The set of all linear combinations of vectors in A is called the **span** of A, and is written*

$$\text{span}(A)$$

We first must prove that $\text{span}(A)$ is a vector subspace:

Theorem 1.2.10. *Let V be a vector space, and let $A = \{\boldsymbol{v_1}, \cdots, \boldsymbol{v_n}\}$ be a set of vectors in V. Then* $\operatorname{span}\{\boldsymbol{v_1}, \cdots, \boldsymbol{v_k}\}$ *is a vector subspace of V.*

Proof. We will first prove that $\operatorname{span}\{\boldsymbol{v_1}, \cdots, \boldsymbol{v_k}\}$ is closed under addition. Let $\boldsymbol{w_1}, \boldsymbol{w_2} \in \operatorname{span}\{\boldsymbol{v_1}, \cdots, \boldsymbol{v_k}\}$. Then by definition,

$$\boldsymbol{w_1} = \lambda_1\boldsymbol{v_1} + \cdots + \lambda_k\boldsymbol{v_k} \qquad \text{and} \qquad \boldsymbol{w_2} = \alpha_1\boldsymbol{v_1} + \cdots + \alpha_k\boldsymbol{v_k}$$

Then by the vector space axioms,

$$\boldsymbol{w_1} + \boldsymbol{w_2} = (\lambda_1 + \alpha_1)\boldsymbol{v_1} + \cdots + (\lambda_k + \alpha_k)\boldsymbol{v_k},$$

Hence $\boldsymbol{w_1} + \boldsymbol{w_2} \in \operatorname{span}\{\boldsymbol{v_1}, \cdots, \boldsymbol{v_k}\}$.

We now prove it is closed under scalar multiplication. Suppose λ is an arbitrary scalar. Then by the vector space axioms,

$$\lambda\boldsymbol{w_1} = (\lambda\lambda_1)\boldsymbol{v_1} + \cdots + (\lambda\lambda_k)\boldsymbol{v_k},$$

Hence $\lambda\boldsymbol{w_1} \in \operatorname{span}\{\boldsymbol{v_1}, \cdots, \boldsymbol{v_k}\}$. □

Proposition 1.2.11. *Fix a vector $\boldsymbol{v} \in V$. Then*

$$\operatorname{span}\{\boldsymbol{v}\} = \{\lambda\boldsymbol{v} \mid \lambda \in \mathbb{R}\}$$

Proof. By definition, $\{\lambda\boldsymbol{v} \mid \lambda \in \mathbb{R}\} \subseteq \operatorname{span}\{\boldsymbol{v}\}$. On the other hand, since there is only one element in the set $\{\boldsymbol{v}\}$, the sum of any elements in $\{\boldsymbol{v}\}$ is again a scalar multiple of $\boldsymbol{v}$. Thus we have that $\{\lambda\boldsymbol{v} \mid \lambda \in \mathbb{R}\} \supseteq \operatorname{span}\{\boldsymbol{v}\}$. □

For example, if $\boldsymbol{v}$ is a non-zero element in $\mathbb{R}^n$, then by definition, $\operatorname{span}\{\boldsymbol{v}\}$ is a line passing through the origin.

Let $A = \{\langle 1, 0\rangle, \langle 0, 1\rangle\}$ be a subset of $\mathbb{R}^2$. Then $\operatorname{span}(A) = \mathbb{R}^2$.

Proof. By definition, $\operatorname{span}(A) \subseteq \mathbb{R}^2$. On the other hand, pick an arbitrary element $\langle x, y\rangle$ in $\mathbb{R}^2$. We see from the axioms that $x\langle 1, 0\rangle + y\langle 0, 1\rangle = \langle x, y\rangle$. Thus $\operatorname{span}(A) \supseteq \mathbb{R}^2$. □

We can now prove that $\operatorname{span}(A)$ is the **smallest** vector subspace containing A.

Theorem 1.2.12. *Let V be a vector space, and let $A = \{\boldsymbol{v_1}, \cdots, \boldsymbol{v_n}\}$ be a set of vectors in V. Suppose W is a vector subspace containing A. Then W contains* $\operatorname{span}(A)$.

Proof. proof □

Bases

We have figured out how to construct a vector subspace $\operatorname{span}(A)$ from a given subset A. One natural question to ask is the converse:

Given a vector subspace W, can we construct a subset B such that $W = \operatorname{span}(B)$? Moreover, is it possible to find a smallest such subset B?

Let us first consider an example:

Consider the vector subspace
$$W = \{(x, y, 0) \in \mathbb{R}^3\}$$
We can write W as the span of several different subsets:

(i) $W = \text{span}\{\langle 1,0,0\rangle, \langle 0,1,0\rangle\}$

(ii) $W = \text{span}\{\langle 1,0,0\rangle, \langle 0,1,0\rangle, \langle 2,0,0\rangle\}$

(iii) $W = \text{span}\{\langle 1,0,0\rangle, \langle 0,1,0\rangle, \langle 1,1,0\rangle\}$

(iv) $W = \text{span}\{\langle 1,0,0\rangle, \langle 0,1,0\rangle, \langle 0,0,0\rangle\}$

(v) $W = \text{span}\{\langle 1,0,0\rangle, \langle 0,2,0\rangle\}$

(vi) $W = \text{span}\{\langle 1,1,0\rangle, \langle 1,2,0\rangle\}$

From this collection of examples, we can make a few observations. Firstly, observe that the sets in examples *(ii)* to *(iv)* all contain the set in example *(i)*, yet they all span the subspace W. What this means is that in terms of linear combinations, examples *(ii)* to *(iv)* contain extraneous information:

- example *(ii)* contains two vectors that are scalar multiples of each other;
- example *(iii)* contains a vector that is a linear combination of the other two vectors;
- example *(iv)* also contains a vector that is a linear combination of the other two vectors (see exercise 1.35).

We can make this idea of "extraneous information" rigorous with the following definition:

Definition 1.2.13. *A set of vectors $A \subset V$ is said to be **linearly dependent** if for every nonempty finite subset of vectors $\{\boldsymbol{v_1}, \cdots, \boldsymbol{v_k}\} \subset A$, there exist scalars a_i, not all zero, such that*
$$a_1\boldsymbol{v_1} + \cdots + a_k\boldsymbol{v_k} = 0$$
*Otherwise, the set A is said to be **linearly independent**.*

This condition is precisely the same as our linear combination criterion as discussed above.

Proposition 1.2.14. *A set of vectors $A \subset V$ is linearly dependent if and only if one of the vectors is a linear combination of the others.*

In some sense, the notion of linear independence gives us a partial answer to finding a "smallest subset" that spans W:

Definition 1.2.15. *An ordered set of vectors $\mathcal{B} \subset V$ is said to be a **basis** of V if*

- $\text{span}(\mathcal{B}) = V$, *and*
- $\mathcal{B}$ *is linearly independent.*

*That is, a **basis** is a linearly independent spanning set.*

Consider the vector subspace
$$W = \{(x, y, 0) \in \mathbb{R}^3\}$$

The following subsets are bases of W:

(i) $\{(1,0,0),(0,1,0)\}$

(ii) $\{(0,1,0),(1,0,0)\}$

(iii) $\{(1,0,0),(0,2,0)\}$

(iv) $\{(1,1,0),(1,2,0)\}$

Observe that examples (i) and (ii) are different bases, because the vectors are listed are listed in a different order!

Examples (iii) and (iv) demonstrate the reason that we used quotes in talking about the "smallest subset". There is no meaningful relationship between these two examples and example (i) - neither of these subsets contain the other. This shows us that the set of bases of a subspace are not well-ordered - that is, there is no good notion of "smallest" when talking about bases. So for an arbitrary vector space, there is no reason to prefer any of the bases over the other.

However, specifically when talking about the vector space $\mathbb{R}^n$, there **is** a preferred basis, called the standard basis:

Definition 1.2.16. *The* ***standard basis*** *of* $\mathbb{R}^n$ *is the collection of vectors*

$$\{\boldsymbol{e_i} \mid i = 1, \cdots, n\},$$

where $\boldsymbol{e_i}$ *denotes the vector with a 1 in the i-th coordinate and 0 elsewhere.*

For example, the standard basis of $\mathbb{R}^3$ is the ordered set $\{(1,0,0),(0,1,0),(0,0,1)\}$.

Why is having a basis for a vector space useful?

Theorem 1.2.17. *Given a basis* $\mathscr{B} = \{\boldsymbol{v_1}, \cdots, \boldsymbol{v_k}\}$ *of* V*, the linear combination* $\boldsymbol{v} = \sum c_i \boldsymbol{v_i}$ *is* ***unique****.*

See

Definition 1.2.18. *Let* $\mathscr{B} = \{\boldsymbol{v_1}, \cdots, \boldsymbol{v_k}\}$ *be a basis of* V*. The* ***coordinates of a vector*** $\boldsymbol{v} = \sum c_i \boldsymbol{v_i}$ *is defined as the ordered set*

$$[\boldsymbol{v}]_{\mathscr{B}} := \{c_1, \cdots, c_n\}$$

Consider the vector $\boldsymbol{v} = \langle x, y, z\rangle \in \mathbb{R}^3$, and consider the standard basis in $\mathbb{R}^3$. Then we can write $\boldsymbol{v}$ as

$$\boldsymbol{v} = x\boldsymbol{e_1} + y\boldsymbol{e_2} + z\boldsymbol{e_3}$$

The coordinates of $\boldsymbol{v}$ in the standard basis is the ordered set (x, y, z).

Consider the vector subspace $W = \{(x, y, 0) \in \mathbb{R}^3\}$ with basis $\{(1,1,0),(1,2,0)\}$. This

To find the coordinates of an arbitrary vector $\boldsymbol{v} = \langle x, y, 0\rangle \in W$, we first observe that W is a two dimensional vector space, so the coordinates will be an ordered set of two elements, say

(a, b).

We must solve the linear system of equations

$$a(1,1,0) + b(1,2,0) = (x, y, 0)$$

So we see have two equations, $a + b = x$, and $a + 2b = y$, thus subtracting the first from the second, we see that $b = y - x$. This in turn tells us that $a + (y - x) = x$, so $a = 2x - y$.

Thus the coordinates for $\boldsymbol{v} = \langle x, y, 0\rangle \in W$ is the ordered set $(2x - y, y - x)$.

The concept of a basis also helps us distinguish vector spaces.

Theorem 1.2.19. *If $\{\boldsymbol{v_1}, \cdots, \boldsymbol{v_n}\}$ is a basis for V, then any other basis has cardinality n.*

Lemma 1.2.20. *Suppose that the set $\{\boldsymbol{v_1}, \cdots, \boldsymbol{v_n}\}$ is a basis for a vector space V. Then any subset of V that contains more than n vectors is linearly dependent.*

exposition

Definition 1.2.21. *If V has a basis of n elements, then the **dimension** of V is n.*

The dimension of $\mathbb{R}^n$ is n.

Theorem 1.2.19 tells us that the dimension of a vector space V is an invariant of vector spaces. That is, if we have two isomorphic vector space $V \cong W$, and we know that $\dim(V) = n$, then it must follow that $\dim(W) = n$. You will prove this in exercise 1.50.

Corollary 1.2.22. *Let V be a vector space of dimension n, and W is a vector space of dimension m. Then if $n \neq m$, then V and W cannot be the same vector space.*

In fact, an even stronger statement is true (which you will prove in exercise 1.51:

Theorem 1.2.23. *If V is a vector space of dimension n, then $V \cong \mathbb{R}^n$*

Thus, we completely understand all of the finite dimensional vector spaces.

Proposition 1.2.24. *Let V be a vector space of dimension n. Then any set of n linearly independent vectors $\{\boldsymbol{x_1}, \cdots, \boldsymbol{x_n}\} \subset V$ forms a basis for V.*

Proposition 1.2.25. *Let V be a vector space of dimension n. Then any set of n vectors $\{\boldsymbol{y_1}, \cdots, \boldsymbol{y_k}\} \subset V$ that spans V also forms a basis for V.*

Exercises

1.18 How is scalar multiplication a generalization of multiplication in $\mathbb{R}$?
(**Hint:** Describe scalar multiplication for the vector space $V = \mathbb{R}$).

1.19 Let $\boldsymbol{u} = \frac{1}{\sqrt{2}}\langle 1, 1\rangle$ and $\boldsymbol{v} = \frac{1}{\sqrt{2}}\langle 1, -1\rangle$ be vectors in $\mathbb{R}^2$. Write $\boldsymbol{w} = \langle 20, 4\rangle \in \mathbb{R}^2$ as a linear combination of $\boldsymbol{u}$ and $\boldsymbol{v}$.

1.20 Consider the set of complex numbers

$$\mathbb{C} := \{a + bi \mid a, b \in \mathbb{R}\}$$

a. Show that $\mathbb{C}$ is a vector space with the usual addition and scalar multiplication.

b. Find a basis for $\mathbb{C}$ as a vector space.

1.21 Prove that for any given vector space V, its distinguished element $\mathbf{0}$ must be unique.

1.22 Given a vector space V, prove that for any vector $\boldsymbol{v} \in V$, the scalar multiple of $0\boldsymbol{v}$ is the zero vector. That is,

$$0\boldsymbol{v} = \mathbf{0}$$

1.23 Given a vector space V, prove that the set $\{\mathbf{0}\}$ is a vector subspace of V.

1.24 What is the span of the set $\{\mathbf{0}\} \in \mathbb{R}^n$?

1.25 For a fixed vector $\langle a_1, \cdots, a_n \rangle \in \mathbb{R}^n$, is the set $W = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid \sum_i^n a_i x_i = 0\}$ always a subspace of $\mathbb{R}^n$?

1.26 For a fixed vector $\langle a_1, \cdots, a_n \rangle \in \mathbb{R}^n$, is the set $W = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid \sum_i^n a_i x_i = 1\}$ always a subspace of $\mathbb{R}^n$?

1.27 For a fixed vector $\langle a_1, \cdots, a_n \rangle \in \mathbb{R}^n$, is the set

$$W = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid \sum_i^n a_i (x_i)^2 = 0\}$$

always a subspace of $\mathbb{R}^n$?

1.28 Suppose that U_1 and U_2 are subspaces of a vector space V. Prove that their intersection,

$$U_1 \cap U_2 := \{\boldsymbol{v} \in V \mid \boldsymbol{v} \in U_1 \text{ and } \boldsymbol{v} \in U_2\}$$

is also a subspace of of V.

1.29 Prove that the set $\mathscr{F}(\mathbb{R})$ of functions $f : \mathbb{R} \to \mathbb{R}$ is a vector space with pointwise addition and pointwise scalar multiplication.

$$(f+g)(x) := f(x) + g(x) \qquad (\lambda f)(x) := \lambda(f(x))$$

1.30 Prove that the set $\mathscr{C}(\mathbb{R})$ of *continuous* functions $f : \mathbb{R} \to \mathbb{R}$ is a vector subspace of $\mathscr{F}(\mathbb{R})$. You can use any facts about continuous function that you know from single variable calculus.

1.31 Is the set $\{(1,0,1), (2,1,0)\}$ a basis of $\mathbb{R}^3$?

1.32 Is the set $\{(1,0,1), (2,1,0), (1,1,-1)\}$ a basis of $\mathbb{R}^3$?

1.33 Is the set $\{(1,0,1), (2,1,0), (1,1,1)\}$ a basis of $\mathbb{R}^3$?

1.34 Is the set $\{(1,0,0), (0,1,0), (0,0,1), (0,0,0)\}$ a basis of $\mathbb{R}^3$?

1.35 Prove that any subset of a vector space that contains the zero vector is linearly dependent.

1.36 Show that the collection P_n of all polynomials (with real coefficients) of degree less than or equal to n is a **subspace** of the vector space $\mathscr{F}(\mathbb{R})$, and that it has $\{1, x, x^2, \cdots, x^n\}$ as a basis.

1.37 Given a basis $\mathscr{B} = \{\boldsymbol{v_1}, \cdots, \boldsymbol{v_k}\}$ of V, prove that the linear combination $\boldsymbol{v} = \sum c_i \boldsymbol{v_i}$ is unique.

1.3 Maps of vector spaces

Now that we have a good grasp on what vector spaces are, we can now study how they relate to each other. In other words, we would like to study certain kinds of maps of vector spaces. For example, in this section, we will see what it means for two vector spaces to be the same!

Let us first consider two examples of maps:

Fix a vector $\boldsymbol{a} = \langle a, b, c\rangle$, and consider the map $f : \mathbb{R}^3 \to \mathbb{R}^3$ that sends a vector $\boldsymbol{v} = \langle x, y, z\rangle$ to the vector $\boldsymbol{v} + \boldsymbol{a} = \langle x + a, y + b, z + c\rangle$.

Observe that the map f satisfies the following properties:

(i) For any two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$, we have that $f(\boldsymbol{u} + \boldsymbol{v}) = f(\boldsymbol{u}) + f(\boldsymbol{v})$.

(ii) For any scalar $\lambda \in \mathbb{R}$, and for any vector $\boldsymbol{v} \in V$, we have that $f(\lambda\boldsymbol{u}) = \lambda f(\boldsymbol{u}))$.

Consider the function $g : \mathbb{R}^3 \to \mathbb{R}^3$ that sends a vector $\langle x, y, z\rangle$ to the vector $\langle x^2, y^2, z^2\rangle$.

Observe that the map g satisfies neither of the two properties

(i) For any two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$, we have that $f(\boldsymbol{u} + \boldsymbol{v}) = f(\boldsymbol{u}) + f(\boldsymbol{v})$.

(ii) For any scalar $\lambda \in \mathbb{R}$, and for any vector $\boldsymbol{v} \in \mathbb{R}^3$, we have that $f(\lambda\boldsymbol{u}) = \lambda f(\boldsymbol{u}))$.

See the problems scalarpreserving for examples of maps that preserve one, but not the other.

Why are properties (i) and (ii) important?

Observe that if a map f satisfies properties (i) and (ii), then it preserves the vector space structure on $\mathbb{R}^n$. That is, it preserves vector space addition and scalar multiplication.

Definition 1.3.1. *A map of vector spaces $T : V \to W$ is called a* ***linear map*** *(or sometimes a* ***linear transformation*** *if*

$$T\left(\sum_i^k \alpha_i x_i\right) = \sum_i^k \alpha_i T(x_i)$$

for all $k \in \mathbb{N}$, for all $\alpha_i \in \mathbb{R}$, and all vectors $x_i \in V$.

Proposition 1.3.2. *A map of vector spaces $f : V \to W$ is a linear map if and only if*

(i) For any two vectors $\boldsymbol{u}, \boldsymbol{v} \in V$, we have that $f(\boldsymbol{u} + \boldsymbol{v}) = f(\boldsymbol{u}) + f(\boldsymbol{v})$.

(ii) For any scalar $\lambda \in \mathbb{R}$, and for any vector $\boldsymbol{v} \in V$, we have that $f(\lambda\boldsymbol{u}) = \lambda f(\boldsymbol{u}))$.

Let $f : \mathbb{R} \to \mathbb{R}$ be the map that sends x to ax. Observe that f is linear.

Show that the map $i : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $i(x, y) = (x, y, 0)$ is linear.

Show that the map $p : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $p(x, y, z) = (x, y)$ is linear.

We can prove a lot of properties about linear maps of vector spaces simply using definition 1.3.1. Just as when we studied vector spaces, whenever you are are trying to determine whether or not something is true about linear maps of vector spaces, you should try to see if it is true for linear maps in $\mathbb{R}^n$. Then, you should try to translate the proof to the general case, purely using the definition of a linear map.

The reason that we rely on definition 1.3.1 to prove things about linear maps is because it saves us time from having to prove a property for each specific linear map that we are interested. By using the axioms, we can prove things to be true for *all* linear maps.

Proposition 1.3.3. *Let $T : V \to W$ be a linear map. Then T sends the 0 object in V to the 0 object in W. That is,*

$$T(\mathbf{0}) = \mathbf{0}$$

Proposition 1.3.4. *Let $T_1, T_2 : V \to W$ be two linear maps. Then the sum of two linear maps $T_1 + T_2$ is a linear map.*

Proposition 1.3.5. *Let $T : V \to W$ be a linear map. Then for any scalar $\lambda \in \mathbb{R}$, the map λT is a linear map.*

reference matrices later

Proposition 1.3.6. *Given two linear maps $S : V \to W$ and $T : U \to V$, prove that the composite $S \circ T : U \to W$ is also linear.*

Properties of linear maps

Let us now investigate some properties of linear maps. The properties are important not just in linear algebra, but in set theory, abstract algebra and topology as well.

Definition 1.3.7. *Let $f : V \to W$ be a map of vector spaces. We say that f is **injective** or **one-to-one** (or sometimes, f is an **injection**) if the following holds:*

$$\text{For all } \boldsymbol{v_1}, \boldsymbol{v_2} \in V, \text{ if } f(\boldsymbol{v_1}) = f(\boldsymbol{v_2}), \text{ then } \boldsymbol{v_1} = \boldsymbol{v_2}.$$

That is, a map f is injective if any element in the codomain of f is the image of **at most** one element in its domain.

Let $f : \mathbb{R} \to \mathbb{R}$ be the map that sends x to ax. Show that f is injective.

Show that the map $i : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $i(x, y) = (x, y, 0)$ is injective.

Equivalently, the contrapositive statement of injectivity states that f is **injective** (**one-to-one**) if the following holds:

$$\text{For all } \boldsymbol{v_1}, \boldsymbol{v_2} \in V, \text{ if } \boldsymbol{v_1} \neq \boldsymbol{v_2}, \text{ then } f(\boldsymbol{v_1}) \neq f(\boldsymbol{v_2}).$$

Definition 1.3.8. *Let $f : V \to W$ be a map of vector spaces. We say that f is **surjective** or **onto** (or sometimes, f is a **surjection**) if the following holds:*

$$\text{For all } \boldsymbol{w} \in W, \text{ there exists a } \boldsymbol{v} \in V \text{ such that } f(\boldsymbol{v}) = \boldsymbol{w}.$$

*That is, any element in the codomain of f is the image of **at least** one element in its domain.*

Show that the map $p : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $p(x, y, z) = (x, y)$ is surjective.

Definition 1.3.9. *Let $f : V \to W$ be a map of vector spaces. We say that f is **bijective** (or sometimes, f is a **bijection**)f f is both injective and surjective.*

That is, any element in the codomain of f is the image of ***exactly*** *one element in its domain.*

Characterising linear maps

Can we describe characterize linear maps $T : \mathbb{R}^n \to \mathbb{R}^m$?

Let us first start with the simplest case, where $n = m = 1$.

Proposition 1.3.10. *Let $T : \mathbb{R} \to \mathbb{R}$ be a linear map. Observe that if $T(1) = a$, then $T(x) = ax$.*

Thus we have shown that linear maps from $\mathbb{R} \to \mathbb{R}$ must take the form of multiplication by a. From proposition 1.3, we saw that the converse holds as well. Thus, we have classified all linear maps from $\mathbb{R} \to \mathbb{R}$!

Theorem 1.3.11. *A map $T : \mathbb{R} \to \mathbb{R}$ is a linear map if and only if*

$$T(x) = ax$$

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

In other words, the graph of a linear map $\mathbb{R} \to \mathbb{R}$ can be thought of as a line in $\mathbb{R}^2$, passing through the origin.
picture

How can we generalize this result?

Observe that the set $\{1\}$ is a basis for the vector space $\mathbb{R}$. Proposition 1.3.10 tells us that a linear map $T : \mathbb{R} \to \mathbb{R}$ is determined completely by where T sends the basis vector 1.
Recall also that given a basis $\mathscr{B} = \{\boldsymbol{v_1}, \cdots, \boldsymbol{v_k}\}$ of a vector space V, we can uniquely describe any vector $\boldsymbol{v} \in V$ as a linear combination of basis vectors (Theorem 1.2.17). That is,

$$\boldsymbol{v} = \sum c_i \boldsymbol{v_i}$$

In other words, in terms of its coordinates.

Let us consider the vector space $\mathbb{R}^3$, with the standard basis. Then, as we saw before, we can write the vector $\boldsymbol{v} = \langle x, y, z \rangle \in \mathbb{R}^3$ in terms of the standard basis:

$$\boldsymbol{v} = x\boldsymbol{e_1} + y\boldsymbol{e_2} + z\boldsymbol{e_3}$$

Let us consider a linear map $T : \mathbb{R}^3 \to W$, where W is an arbitrary vector space. Observe that using linearity, we have that

$$\begin{aligned} T(\boldsymbol{v}) &= T(x\boldsymbol{e_1} + y\boldsymbol{e_2} + z\boldsymbol{e_3}) \\ &= xT(\boldsymbol{e_1}) + yT(\boldsymbol{e_2}) + zT(\boldsymbol{e_3}) \end{aligned}$$

Thus, if we know the values of $T(\boldsymbol{e_1})$, $T(\boldsymbol{e_2})$ and $T(\boldsymbol{e_3})$ (e.g. how T acts on the basis vectors), then we can use the above formula to calculate $T(\boldsymbol{v})$ for any vector $\boldsymbol{v} \in \mathbb{R}^3$.

This idea generalizes to the following theorem:

Theorem 1.3.12. *Let V be a vector space with basis $\{\boldsymbol{v_1}, \cdots, \boldsymbol{v_n}\}$, and let W be an arbitrary vector space. A linear map $T : V \to W$ is determined by what it does on basis vectors.*

Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear mapping such that $T\langle 1,0,0\rangle = \langle 1,0,1\rangle$, $T\langle 0,1,0\rangle = \langle 3,0,2\rangle$, and $T\langle 0,0,1\rangle = \langle 4,2,1\rangle$.

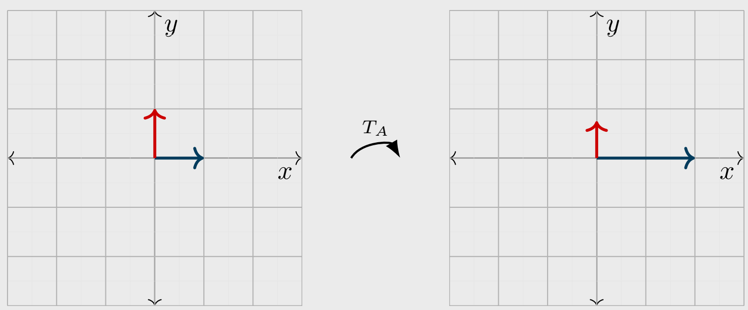
Compute $T(3,3,2)$.

How can we geometrically think about linear maps $\mathbb{R}^n \to \mathbb{R}^m$?

In other words, what do graphs of linear maps from $\mathbb{R}^n \to \mathbb{R}^m$ look like?

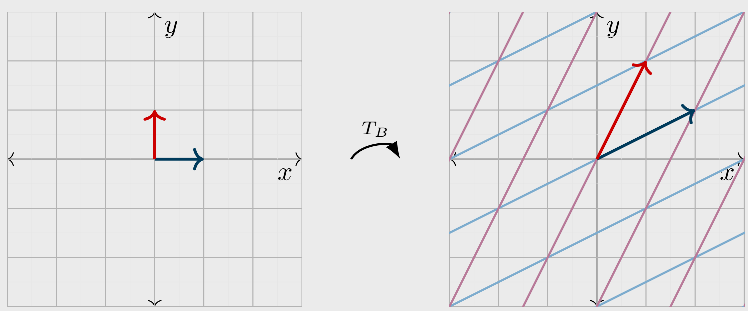
By theorem 1.3.12, we know that a linear map $\mathbb{R}^n \to \mathbb{R}^m$ is determined by what it does on basis vectors. Let us consider a few linear maps $\mathbb{R}^2 \to \mathbb{R}^2$.

Consider the linear map T_A sends the vector $\boldsymbol{e_1} = \langle 1,0\rangle$ to the vector $T_A(\boldsymbol{e_1}) = \langle 2,0\rangle$, and the vector $\boldsymbol{e_2} = \langle 0,1\rangle$ to the vector $T_A(\boldsymbol{e_2}) = \langle 0,\frac{3}{4}\rangle$. Geometrically, the map looks like this:



Not every linear map is as simple, however!

Let us consider the linear map T_B that sends the vector $\boldsymbol{e_1}$ to the vector $T_B(\boldsymbol{e_1}) = \langle 2,1\rangle$, and the vector $\boldsymbol{e_2}$ to the vector $T_B(\boldsymbol{e_2}) = \langle 1,2\rangle$. Geometrically, the map looks like this:



We observe that the linear map T_B sends the unit square spanned by $\boldsymbol{e_1}$ and $\boldsymbol{e_2}$ to the parallelogram spanned by $T_B(\boldsymbol{e_1})$ and $T_B(\boldsymbol{e_2})$. By linearity, T_B then sends the grid lines to a grid of congruent parallelograms.

In general, T_B is a good mental image for the geometric interpretation of a linear map $T : \mathbb{R}^2 \to \mathbb{R}^2$. This is why linear maps are sometimes called linear transformations - as they transform the vectors in the domain into (potentially different) vectors in the range.

Matrices

Is there a more efficient way to describe how a linear transformation acts on a basis?

Let us first set up some machinery:

Definition 1.3.13. *An $m \times n$* ***(real) matrix*** *A is an array of elements $a_{i,j} \in \mathbb{R}$ with m rows and n columns:*

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

The set of $m \times n$ (real) matrices is denoted $M_{m \times n}(\mathbb{R})$.

Theorem 1.3.14. *The set $M_{m \times n}(\mathbb{R})$, with entry-wise addition and scalar multiplication, is a vector space.*

$$A_{m,1} = \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{bmatrix}$$

Observe that we can write vectors in $\mathbb{R}^m$ as $m \times 1$ matrices.

An $m \times n$ matrix can act on a vector in $\mathbb{R}^m$ in the following way:

Definition 1.3.15 (Matrix by vector multiplication)**.** *Let $A \in M_{m \times n}(\mathbb{R})$, and let $\boldsymbol{x} \in \mathbb{R}^n$. The product $A\boldsymbol{x} \in \mathbb{R}^m$ is defined the vector whose i-th entry is sum $\sum_j a_{i,j} x_j$. That is,*

$$A\boldsymbol{x} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{1,1}x_1 + \cdots + a_{1,n}x_n \\ a_{2,1}x_1 + \cdots + a_{2,n}x_n \\ \vdots \\ a_{m,1}x_1 + \cdots + a_{m,n}x_n \end{bmatrix}$$

Compute $A\boldsymbol{x}$, where $\boldsymbol{x} = \langle 7, 8, 9 \rangle$, and

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

Proposition 1.3.16. *Observe that the matrix $A \in M_{m \times n}(\mathbb{R})$ determines a map $T : \mathbb{R}^n \to \mathbb{R}^m$.*

Thus $m \times n$ matrices are not the same as $n \times m$ matrices, unless $m = n$. If $m = n$, we call $A \in M_{n \times n}(\mathbb{R})$ a **square matrix**.

Theorem 1.3.17. *Let $A \in M_{m \times n}(\mathbb{R})$. The map $T : \mathbb{R}^n \to \mathbb{R}^m$ defined by $T(\boldsymbol{x}) = A\boldsymbol{x}$ is linear.*

Now, we will show the converse:

Definition 1.3.18. *Given a basis $\mathcal{B} = \{\boldsymbol{e_1}, \cdot \boldsymbol{e_n}\}$ for $\mathbb{R}^n$, the* ***standard matrix*** *A of a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ is given by*

$$[A] = \begin{bmatrix} | & | & & | \\ T(\boldsymbol{e_1}) & T(\boldsymbol{e_2}) & \cdots & T(\boldsymbol{e_n}) \\ | & | & & | \end{bmatrix} \in M_{m \times n}(\mathbb{R})$$

Thus, we have precisely characterized the linear maps from $\mathbb{R}^n \to \mathbb{R}^m$ - they correspond exactly to $n \times m$ matrices!

Theorem 1.3.19. *A linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ determined exactly by a matrix A such that*

$$T(\boldsymbol{x}) = A\boldsymbol{x}$$

We call A the ***standard matrix*** *of T.*

$$A_{1,n} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \end{bmatrix}$$

Observe that the coordinate vector in $\mathbb{R}^n$ is an $1 \times n$ matrix.

Definition 1.3.20. *If $A = [a_{i,j}]$, the* ***transpose matrix*** *is the matrix $A^T = [a_{j,i}]$.*

We previously saw in proposition 1.3.6 that the composition of linear maps is again linear. What is the standard matrix of the composite?

Let $S : \mathbb{R}^n \to \mathbb{R}^m$ and $T : \mathbb{R}^p \to \mathbb{R}^n$ be linear maps, with standard matrices $A \in M_{m \times n}(\mathbb{R})$ and $B \in M_{n \times p}(\mathbb{R})$.

Then $S \circ T$ is a map from $\mathbb{R}^p \to \mathbb{R}^m$, and should have a standard matrix $C \in M_{m \times p}(\mathbb{R})$

Definition 1.3.21 (Matrix multiplication)**.** *Let $A \in M_{m \times n}(\mathbb{R})$ and $B \in M_{n \times p}(\mathbb{R})$ be two matrices. Denote the columns of B by the vectors $\boldsymbol{b_1}, \cdots, \boldsymbol{b_p}$. Then the product of A and B is the matrix*

$$AB = \begin{bmatrix} | & | & & | \\ A(\boldsymbol{b_1}) & A(\boldsymbol{b_2}) & \cdots & A(\boldsymbol{b_n}) \\ | & | & & | \end{bmatrix} \in M_{m \times p}(\mathbb{R})$$

In terms of entries, if $A = [a_{i,j}]$ and $B = [b_{i,j}]$, then

$$AB = \left[\sum_k a_{i,k} b_{k,j} \right]$$

Find the matrix AB, where

$$A = \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \qquad \text{and} \qquad B = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 2 & 1 \end{bmatrix}$$

Theorem 1.3.22 (Properties of matrix multiplication)**.** a. *$A(BC) = (AB)C$ for $A \in M_{m \times n}(\mathbb{R})$, $B \in M_{n \times p}(\mathbb{R})$, $C \in M_{p \times q}(\mathbb{R})$ (associativity).*

b. $A(B+C) = AB + AC$ *for* $A \in M_{m\times n}(\mathbb{R})$, $B, C \in M_{n\times p}(\mathbb{R})$ *(distribution).*

c. $(A+B)C = AC + BC$ *for* $A, B \in M_{m\times n}(\mathbb{R})$, $C \in M_{n\times p}(\mathbb{R})$ *(distribution).*

d. $\lambda(AB) = (\lambda A)B = A(\lambda B)$ *(scalars).*

Warning: Matrix multiplication is **not** commutative.

Definition 1.3.23. *The* $n \times n$ ***identity matrix*** I_n *is the standard matrix of the identity transformation* $id : \mathbb{R}^n \to \mathbb{R}^n$*, which sends a vector* $\boldsymbol{v}$ *to itself.*

The $n \times n$*identity matrix has the form*

$$I_n = [\delta_{ij}]$$

That is, the entry in the i*-th row and* j*-th column is described by the* ***Kronecker delta*** *function, which is defined as*

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

The 3×3 identity matrix is of the form

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Proposition 1.3.24. *The identity matrix* $I_n \in M_{n\times n}(\mathbb{R})$ *satisfies the property that for all* $A \in M_{m\times n}(\mathbb{R})$,

$$I_m A = A = A I_n$$

Proposition 1.3.25. *The identity matrix* $I_n \in M_{n\times n}(\mathbb{R})$ *corresponds to the linear transformation* $Id_n : \mathbb{R}^n \to \mathbb{R}^n$*, determined by* $Id_n(\boldsymbol{v}) = \boldsymbol{v}$*.*

Definition 1.3.26. *A matrix* $A \in M_{n\times n}(\mathbb{R})$ *is* ***invertible*** *if there exists a matrix* $B \in M_{n\times n}(\mathbb{R})$ *such that* $AB = BA = I_n$*.*

B *is called the* ***inverse*** *of* A*.*

Show that the matrix

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

has inverse

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

The matrix

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

is **not** invertible. (**Hint:** use the fact that $A^2 = 0$ to obtain a contradiction)

What does invertibility look like for linear transformations?

Definition 1.3.27. *A linear transformation $T : V \to W$ is* ***invertible*** *if there exists a linear transformation $S : W \to V$ such that $S \circ T = id_V$ and $T \circ S = id_W$.*

If this is true, we also call T an ***isomorphism of vector spaces****, and V and W are* ***isomorphic****.*

Proposition 1.3.28. *If $T : V \to W$ is linear and invertible, then $T^{-1} : W \to V$ is also linear and invertible.*

The map $[]_{\mathcal{S}} : \mathbb{R}^n \to M_{n\times 1}$ that sends a point to its coordinate vector (with respect to the standard basis) is an isomorphism.

The vector space $M_{m\times n}(\mathbb{R})$ is isomorphic to the vector space $\mathbb{R}^{mn}$.

Note that an isomorphism of vector spaces does not preserve anything except the vector space structure.

Proposition 1.3.29. *A linear map $T : V \to W$ is an isomorphism if and only if it is both injective and surjective.*

How do we determine when a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$ is an isomorphism? Equivalently, when is a matrix $A \in M_{m\times n}(\mathbb{R})$ invertible?

From our discussion about dimension earlier, the dimension of a vector space V (that is, the cardinality of any basis of V) is an invariant of V). For example, the dimension of $\mathbb{R}^n$ is n. reference

Thus we restrict our attention to linear maps $T : \mathbb{R}^n \to \mathbb{R}^n$, or equivalently, square matrices $A \in M_{n\times n}(\mathbb{R})$. For a square matrix, we can develop the theory of **determinants**:

Definition 1.3.30. *The* ***determinant*** *of a 2×2 matrix is denoted and defined as follows:*

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

Find the determinant of the matrix

$$A = \begin{bmatrix} 3 & 2 \\ 5 & 4 \end{bmatrix}$$

Definition 1.3.31. *The* ***determinant*** *of a 3×3 matrix is denoted and defined as follows:*

$$\det(A) = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = a_1 \begin{vmatrix} b_2 & c_2 \\ b_3 & c_3 \end{vmatrix} - b_1 \begin{vmatrix} a_2 & c_2 \\ a_3 & c_3 \end{vmatrix} + c_1 \begin{vmatrix} a_2 & b_2 \\ a_3 & b_3 \end{vmatrix}$$

Find the determinant of the matrix

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

To see what the determinant measures, see section 1.4.

How can we generalize the determinant to $n \times n$ matrices?

Theorem 1.3.32 (Expansion along the i-th row)**.** *Let A be an $n \times n$ matrix, and let A_{ij} denote the $(n-1) \times (n-1)$ matrix obtained by deleting row i and column j from A.*

$$\det(A) = (-1)^{i+1} a_{i,1} \det(A_{i1}) + \cdots + (-1)^{i+n} a_{i,n} \det(A_{in})$$

This theorem tells us that given an $n \times n$ matrix A, we can expand along **any row**, and reduce the problem to computing determinants of $(n-1) \times (n-1)$. We can iterate this algorithm, and reduce computing the determinant of an $n \times n$ matrix to a problem of computing determinants of 2×2 matrices, which we know how to do.

Proposition 1.3.33. *Given a matrix A, swapping any two rows or any two columns changes the sign of the determinant.*

Where does this formula come from? See problem reference
Properties of the determinant

Remark 1.3.34. *Let $A, B \in M_{n\times n}(\mathbb{R})$. In general, $\det(A+B) \neq \det(A) + \det(B)$.*

Theorem 1.3.35. *Let $A, B \in M_{n\times n}(\mathbb{R})$. Then $\det(AB) = \det(A)\det(B)$.*

Definition 1.3.36. *Given a matrix $A = [a_{i,j}] \in M_{n\times n}(\mathbb{R})$, the **transpose matrix** is the matrix A^T defined by*

$$A^T = [b_{i,j}] \in M_{n\times n}(\mathbb{R})$$

where $b_{i,j} := a_{j,i}$.

Theorem 1.3.37. *Given a matrix $A = [a_{i,j}] \in M_{n\times n}(\mathbb{R})$, we have that*

$$\det(A^T) = \det(A)$$

Theorem 1.3.38. *A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ with standard matrix A is an isomorphism if and only if $\det(A) \neq 0$.*

Thus, the determinant allows us to detect isomorphisms.
Bucketlist of injective, surjective

Exercises

1.38 Consider the map $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} x, & \text{if } y = 0 \\ y, & \text{if } y \neq 0 \end{cases}$$

Show that f preserves scalar multiplication, but not vector addition.

1.39 Prove that a map of vector spaces $f : V \to W$ is a linear map if and only if for all vectors $\boldsymbol{v}, \boldsymbol{w} \in V$, and for any scalar $\lambda \in \mathbb{R}$, we have that

$$f(\lambda \boldsymbol{u} + \boldsymbol{v}) = \lambda f(\boldsymbol{u})) + f(\boldsymbol{v})$$

1.40 Given two linear maps $S : V \to W$ and $T : U \to V$, prove that the composite $S \circ T : U \to W$ is also linear.

1.41 Let $T : V \to W$ be a linear map. Is the set

$$\operatorname{Ker}(T) := \{\boldsymbol{v} \in V \mid T(\boldsymbol{v}) = \boldsymbol{0}\}$$

a subspace of V?

1.42 Let $T : V \to W$ be a linear map. Is the set

$$\operatorname{Im}(T) := \{T(\boldsymbol{v}) \mid \boldsymbol{v} \in V\}$$

a subspace of W?

1.43 Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear mapping such that $T\langle 0, 1, 1\rangle = \langle 1, 0, 1\rangle$, $T\langle 1, 0, 1\rangle = \langle 3, 0, 2\rangle$, and $T\langle 1, 1, 0\rangle = \langle 4, 2, 1\rangle$.

Compute $T(3, 3, 2)$.

1.44 et $T : \mathbb{R}^3 \to \mathbb{R}$ be the linear mapping such that $T(0, 1, 1) = 1$, $T(1, 0, 1) = 2$, and $T(1, 1, 0) = 3$. Find the standard matrix of T.

1.45 A subset $A \subset \mathbb{R}^n$ is said to be a **convex subset of** $\mathbb{R}^n$ if it contains the line segment joining any two points of A. That is, for all $\boldsymbol{a}, \boldsymbol{b} \in A$, and for all $t \in [0, 1]$, then $\boldsymbol{a} + t(\boldsymbol{b} - \boldsymbol{a}) \in A$.

Let $U \subset \mathbb{R}^n$ be a convex subset of $\mathbb{R}^n$. Show that if $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then $T(U)$ is a convex subset.

1.46 Show that the rotation matrix

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

has inverse

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

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

1.48 how that the matrix

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

is **not** invertible. (**Hint:** use the fact that $A^2 = 0$ to obtain a contradiction)

1.49 Let V and W be vector spaces such that $\dim(V) = n = \dim(W)$. Is a linear map $T : V \to W$ necessarily an isomorphism?

1.50 Show that an isomorphism of vector spaces must preserve dimension. That is, an isomorphism sends a basis of n elements to a basis of n elements.

1.51 Prove that if V is a vector space of dimension n, then $V \cong \mathbb{R}^n$.

(**Hint:** Linear maps are determined by their actions on basis vectors).

1.52 Show that if A is invertible, then $\det(A) \neq 0$.

1.53 Prove that if a map $T : V \to W$ is an isomorphism, then T is surjective.

1.54 Prove that if a map $T : V \to W$ is an isomorphism, then T is injective.

1.55 Suppose that $S : \mathbb{R}^3 \to \mathbb{R}^3$ is a linear map such that $S(\boldsymbol{e_1}) = S(\boldsymbol{e_2})$. Is S an isomorphism?

1.56 If a map $T : \mathbb{R}^n \to \mathbb{R}^n$ is both surjective and injective (bijective), what is the standard matrix of T^{-1}?

1.57 Find the determinant of the matrix

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

Is the linear map associated to A invertible?

1.58 Find the determinant of the matrix

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

Is the linear map associated to A invertible?

1.59 Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the linear mapping such that $T\langle 0, 1, 1\rangle = \langle 1, 0, 1\rangle$, $T\langle 1, 0, 1\rangle = \langle 3, 0, 2\rangle$, and $T\langle 1, 1, 0\rangle = \langle 4, 2, 1\rangle$. Is T invertible?

In this set of problems, we will see an abstract way to define the determinant map.
First, we need to define a property of maps of vector spaces:

Definition 1.3.39. *A map $M : V_1 \times \cdots \times V_k \to W$ is said to be* ***multilinear*** *if it is linear in each component. That is, for all $1 \leq i \leq k$, and for any fixed $\boldsymbol{v_\ell} \in V_\ell, \ell \neq i$, the map $V_i \to W$ given by*

$$\boldsymbol{v} \mapsto M(\boldsymbol{v_1}, \cdots, \boldsymbol{v_{i-1}}, \boldsymbol{v}, \boldsymbol{v_{i+1}}, \cdots, \boldsymbol{v_k})$$

is linear.

If $k = 2$, then we say M is ***bilinear****, instead of multilinear.*

1.60 how that the map $\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ defined by

$$\langle u_1, \cdots, u_n\rangle \cdot \langle v_1, \cdots, v_n\rangle = \sum u_i v_i$$

is bilinear.

(This map is called the **dot product**, which we will discuss in the next section).

1.61 Show that if we think of a matrix $A = [a_{i,j}] \in M_{m\times n}(\mathbb{R})$ as an array of n column vectors, then the determinant map for 2×2 and 3×3 matrices is multilinear in the columns of A.

Definition 1.3.40. *A multilinear map M is said to be* ***alternating*** *or* ***skew-symmetric*** *if*

$$M(\boldsymbol{v_1}, \cdots, \boldsymbol{v_i}, \cdots, \boldsymbol{v_j}, \cdots, \boldsymbol{v_k}) = -M(\boldsymbol{v_1}, \cdots, \boldsymbol{v_j}, \cdots, \boldsymbol{v_i}, \cdots, \boldsymbol{v_k})$$

for all $1 \leq i < j \leq k$ and for any $\boldsymbol{v_\ell} \in V_\ell$.

1.62 Show that the determinant map for 2×2 and 3×3 matrices is alternating.

1.63 Show that the dot product is **not** alternating. It is instead symmetric (commutative).

Definition 1.3.41 (The determinant). *There is exists a* ***unique*** *map,* $\det : \mathbb{R}^n \times \cdots \times \mathbb{R}^n \to \mathbb{R}$ *satisfying the following:*

a. $\det$ *is multilinear*

b. $\det$ *is alternating (skew-symmetric)*

c. $\det(\boldsymbol{e_1}, \cdots, \boldsymbol{e_n}) = 1$.

This is an abstract way to determine the determinant of an $n \times n$ *matrix.*

1.4 Operations on vector spaces

So far, we have discussed vector spaces and linear maps of vector spaces. Now we will turn to studying certain operations on $\mathbb{R}^n$ (and $\mathbb{R}^3$), which will tell us about the geometry of $\mathbb{R}^n$ (and $\mathbb{R}^3$)

Dot product

Definition 1.4.1. *Given two vectors* $\boldsymbol{u} = \langle u_1, \cdots, u_n \rangle$ *and* $\boldsymbol{v} = \langle v_1, \cdots, v_n \rangle$ *in* $\mathbb{R}^n$*, their (**dot product**) is the* <u>*scalar*</u> *defined by*

$$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \boldsymbol{u} \cdot \boldsymbol{v} = \sum u_i v_i$$

This defines a function $\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.

Remark 1.4.2. *Observe that we can identify* $\boldsymbol{u} \cdot \boldsymbol{v}$ *as the matrix product* $\boldsymbol{u}^T \boldsymbol{v}$.

Theorem 1.4.3 (Axioms of the dot product). a. $\boldsymbol{u} \cdot \boldsymbol{v} = \boldsymbol{v} \cdot \boldsymbol{u}$ *(commutativity).*

b. $\lambda(\boldsymbol{u} \cdot \boldsymbol{v}) = (\lambda \boldsymbol{u}) \cdot \boldsymbol{v} = \boldsymbol{u} \cdot \lambda(\boldsymbol{v})$ *(scalars).*

c. $\boldsymbol{u} \cdot (\boldsymbol{v} + \boldsymbol{w}) = \boldsymbol{u} \cdot \boldsymbol{v} + \boldsymbol{u} \cdot \boldsymbol{w}$ *(distribution).*

d. $\boldsymbol{v} \cdot \boldsymbol{v} \geq 0$, *with equality only when* $\boldsymbol{v} = \boldsymbol{0}$. *(positive definite).*

Proposition 1.4.4. *If* $\boldsymbol{v} \cdot \boldsymbol{w} = 0$ *for all* $\boldsymbol{v} \in V$*, then* $\boldsymbol{w} = \boldsymbol{0}$.

Corollary 1.4.5. *If* $\boldsymbol{v} \cdot \boldsymbol{w} = \boldsymbol{u} \cdot \boldsymbol{w}$ *for all* $\boldsymbol{w} \in V$*, then* $\boldsymbol{v} = \boldsymbol{u}$.

What is the geometric meaning behind the dot product in $\mathbb{R}^n$?

Proposition 1.4.6. *Observe that for all* $\boldsymbol{v} \in \mathbb{R}^n$*, we have that* $\boldsymbol{v} \cdot \boldsymbol{v} = ||\boldsymbol{v}||^2$.

We see that the dot product allows us to talk about the notion of <u>distance</u> in $\mathbb{R}^n$. Thus the dot product has a geometric interpretation!

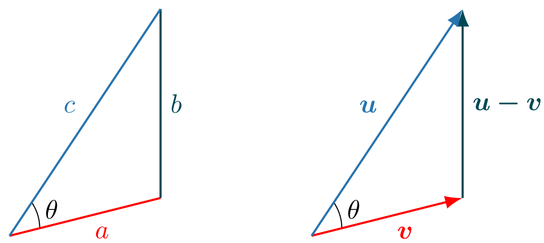
Proposition 1.4.7. *Let* θ *be the angle between two non-zero position vectors* $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *in* $\mathbb{R}^n$*. Then*

$$\cos(\theta) = \frac{\boldsymbol{u} \cdot \boldsymbol{v}}{||\boldsymbol{u}|| \; ||\boldsymbol{v}||}$$

Proof. This proposition follows from the Law of cosines: Given a triangle with side lengths a, b, c, where θ is the angle between the sides of length a and b, we have that

$$c^2 + a^2 + b^2 - 2ab\cos(\theta)$$

Given two non-zero vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, we can apply the Law of cosines to the triangle **spanned by $\boldsymbol{u}$ and $\boldsymbol{v}$** pictured below:



The law of cosines tells us that for this triangle,

$$||\boldsymbol{u} - \boldsymbol{v}||^2 = ||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2 - 2||\boldsymbol{u}||\ ||\boldsymbol{v}||\cos(\theta)$$

However, on the other hand, the properties of dot products tell us that

$$\begin{aligned} ||\boldsymbol{u} - \boldsymbol{v}||^2 &= (\boldsymbol{u} - \boldsymbol{v}) \cdot (\boldsymbol{u} - \boldsymbol{v}) \\ &= \boldsymbol{u} \cdot (\boldsymbol{u} - \boldsymbol{v}) - \boldsymbol{v} \cdot (\boldsymbol{u} - \boldsymbol{v}) \\ &= \boldsymbol{u} \cdot \boldsymbol{u} - \boldsymbol{u} \cdot \boldsymbol{v} - \boldsymbol{v} \cdot \boldsymbol{u} + \boldsymbol{v} \cdot \boldsymbol{v} \\ &= ||\boldsymbol{u}||^2 + ||\boldsymbol{v}||^2 - 2(\boldsymbol{u} \cdot \boldsymbol{v}) \end{aligned}$$

Given these two ways of evaluating $||\boldsymbol{u} - \boldsymbol{v}||^2$, we thus conclude that

$$||\boldsymbol{u}||\ ||\boldsymbol{v}||\cos(\theta) = \boldsymbol{u} \cdot \boldsymbol{v}$$

as desired.

□

Thus, we can use the dot product in $\mathbb{R}^n$ to determine whether the angle between the vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is acute, obtuse, or if $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal (that is, if $\theta = \frac{\pi}{2}$).

Proposition 1.4.8. *The angle θ between two non-zero position vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathbb{R}^n$ is:*

- *Acute* $(0 < \theta < \frac{\pi}{2})$ *if and only if* $\boldsymbol{u} \cdot \boldsymbol{v} > 0$,
- *Orthogonal* $(\theta = \frac{\pi}{2})$ *if and only if* $\boldsymbol{u} \cdot \boldsymbol{v} = 0$,
- *Obtuse* $(\frac{\pi}{2} < \theta < \pi)$ *if and only if* $\boldsymbol{u} \cdot \boldsymbol{v} < 0$.

The standard basis vectors $\{\boldsymbol{e_i}\}$ in $\mathbb{R}^n$ are all orthogonal to each other.

Definition 1.4.9. *A subset of vectors $S = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_k}\} \subseteq \mathbb{R}^n$ is said to be **orthogonal** if $\boldsymbol{v_i} \cdot \boldsymbol{v_j} = 0$ for all $i \neq j$.*

*Furthermore, if $||\boldsymbol{v_i}|| = 1$ for all $1 \leq i \leq k$, we say that the subset $S = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_k}\} \subseteq \mathbb{R}^n$ is **orthonormal**.*

Proposition 1.4.10. *Equivalently, a subset* $S = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_k}\} \subseteq V$ *is orthonormal if*

$$\boldsymbol{v_i} \cdot \boldsymbol{v_j} = \delta_{ij}$$

where δ_{ij} *is the* ***Kronecker delta****, which is the symbol*

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Orthonormal bases (such as the standard basis) are nice, because it turns out that we can compute the coordinate vectors for an orthonormal basis easily:

Proposition 1.4.11. *Recall that if* $\mathscr{B} = \{\boldsymbol{b_1}, \boldsymbol{b_2}, \cdots, \boldsymbol{b_n}\}$ *is an orthonormal basis of* V*, then for any* $\boldsymbol{v} \in V$*,*

$$\boldsymbol{v} = (\boldsymbol{v} \cdot \boldsymbol{b_1})\boldsymbol{b_1} + \cdots (\boldsymbol{v} \cdot \boldsymbol{b_n})\boldsymbol{b_n}$$

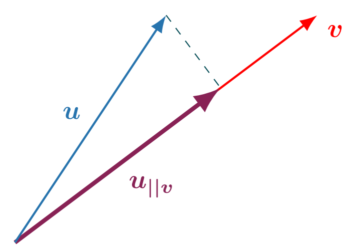
That is, the coordinates of a vector in an orthonormal basis $\{\boldsymbol{b_i}\}$ of $\mathbb{R}^n$ are exactly given by the dot products $\boldsymbol{v} \cdot \boldsymbol{b_1}$.

Definition 1.4.12. *Assume* $\boldsymbol{v} \neq 0$*. The* ***projection of u along v*** *is the vector*

$$\boldsymbol{u}_{||\boldsymbol{v}} = \left(\frac{\boldsymbol{u} \cdot \boldsymbol{v}}{\boldsymbol{v} \cdot \boldsymbol{v}}\right)\boldsymbol{v} = \left(\frac{\boldsymbol{u} \cdot \boldsymbol{v}}{||\boldsymbol{v}||^2}\right)\boldsymbol{v} = \left(\frac{\boldsymbol{u} \cdot \boldsymbol{v}}{||\boldsymbol{v}||}\right)\boldsymbol{e_v}$$

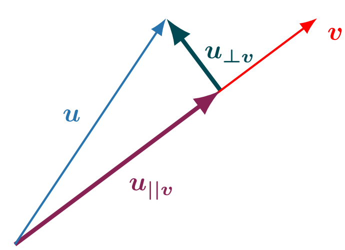
This is sometimes denoted $\mathrm{proj}_{\boldsymbol{v}}\boldsymbol{u}$*.*

The scalar $\frac{\boldsymbol{u} \cdot \boldsymbol{v}}{||\boldsymbol{v}||}$ *is called the* ***scalar component*** *of* ***u along v****.*



Definition 1.4.13. *The* ***orthogonal projection*** *of* ***u along v*** *is the vector*

$$\boldsymbol{u}_{\perp \boldsymbol{v}} = \boldsymbol{u} - \boldsymbol{u}_{||\boldsymbol{v}}$$



Theorem 1.4.14 (Further properties of the dot product)**.** *The following properties of the dot product tell us about the geometry for* $\mathbb{R}^n$*:*

a. *Cauchy-Schwarz inequality:* $|\boldsymbol{u} \cdot \boldsymbol{v}| \leq ||\boldsymbol{u}||\ ||\boldsymbol{v}||$

b. *Triangle inequality:* $||\boldsymbol{u}+\boldsymbol{v}|| \leq ||\boldsymbol{u}|| + ||\boldsymbol{v}||$

Geometric interpretation

Observe that the dot product on $\mathbb{R}^n$ is a function $\cdot : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$. Suppose we fix a vector $\boldsymbol{n} \in \mathbb{R}^n$. We can then consider the function

$$- \cdot \boldsymbol{n} : \mathbb{R}^n \to \mathbb{R}$$

defined by

$$\boldsymbol{v} \mapsto \boldsymbol{v} \cdot \boldsymbol{n}$$

Given a vector $\boldsymbol{n} \in \mathbb{R}^n$, can we describe the set of vectors $\boldsymbol{v} \in \mathbb{R}^n$ such that $\boldsymbol{v} \cdot \boldsymbol{n} = 0$?

We know that $\mathbf{0}$ is always in this set. We also know from proposition 1.4.7 that this is equivalent to describing all of the non-zero vectors that are *orthogonal* to $\boldsymbol{n}$.

Proposition 1.4.15. *For a fixed vector $\boldsymbol{n} \in \mathbb{R}^n$, the set*

$$\{\boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{v} \cdot \boldsymbol{n} = 0\}$$

is a subspace of $\mathbb{R}^n$.

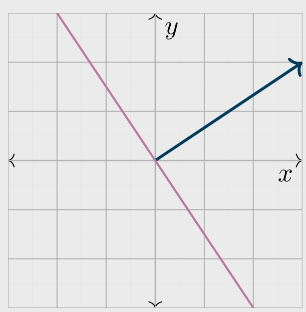
Let us examine a few examples.

Let us fix a vector $\boldsymbol{n} = \langle a, b\rangle \in \mathbb{R}^2$. Then we can write the subspace $\{\boldsymbol{v} \in \mathbb{R}^2 \mid \boldsymbol{v} \cdot \boldsymbol{n} = 0\}$ as

$$\{\boldsymbol{v} = \langle x, y\rangle \in \mathbb{R}^2 \mid ax + by = 0\}$$

In other words if $b \neq 0$, this subspace is precisely the line $y = \frac{-a}{b}x$. If $b = 0$, then this subspace is the line $x = 0$.

Let us take for example $\boldsymbol{n} = \langle 3, 2\rangle \in \mathbb{R}^2$ in blue. Then the subspace is the line in red below, passing through the origin, in the direction of the vector $\langle -2, 3\rangle$.



Thus, we see that the subspace of vectors in $\mathbb{R}^2$ that are orthogonal to a vector $\boldsymbol{n} = \langle a, b\rangle$ is a line passing through the origin, in the direction of the vector $\langle -b, a\rangle$.

In other words,

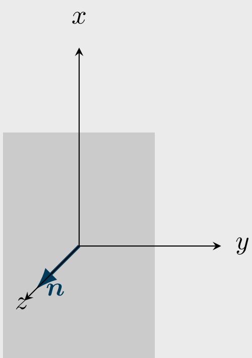
$$\{\boldsymbol{v} = \langle x, y\rangle \in \mathbb{R}^2 \mid ax + by = 0\} = \text{span}\{\langle -b, a\rangle\}$$

A vector parametrization of this line is $\boldsymbol{r}(t) = t\boldsymbol{v}$.

Let us fix a vector $\boldsymbol{n} = \langle a, b, c\rangle \in \mathbb{R}^3$. Then we can write the subspace $\{\boldsymbol{v} \in \mathbb{R}^3 \mid \boldsymbol{v} \cdot \boldsymbol{n} = 0\}$ as

$$\{\boldsymbol{v} = \langle x, y, z\rangle \in \mathbb{R}^3 \mid ax + by + cz = 0\}$$

Let us take for example $\boldsymbol{n} = \langle 3, 2, 1\rangle \in \mathbb{R}^3$ in blue.



105-30

We again observe that the following two subspaces are the same!

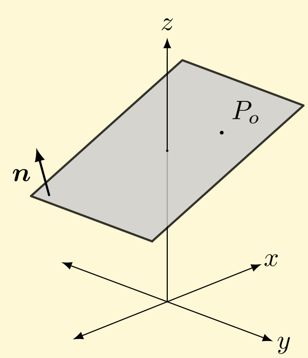
$$\{\boldsymbol{v} = \langle x, y, z\rangle \in \mathbb{R}^3 \mid 3x + 2y + 1z = 0\} = \operatorname{span}\{\langle 1, 0, -3\rangle, \langle 0, 1, -2\rangle\}$$

That is, the subspace of vectors in $\mathbb{R}^3$ orthogonal to $\boldsymbol{n} = \langle a, b, c\rangle$ is a plane in $\mathbb{R}^3$ passing through the origin.

In order to define a plane P in $\mathbb{R}^3$ (not necessarily passing through the origin), we simply need to describe a point on the plane P, and a normal vector $\boldsymbol{n} \in \mathbb{R}^3$.

Definition 1.4.16. *A plane P in $\mathbb{R}^3$ is determined by a point $P_0 = (x_0, y_0, z_0)$ and a **normal vector** $\boldsymbol{n} = \langle a, b, c\rangle$.*

The points in the plane P consist of all the vectors that originate from P_0 and are orthogonal to $\boldsymbol{n}$.



Theorem 1.4.17. *The plane P in $\mathbb{R}^3$ determined by a point $P_0 = (x_0, y_0, z_0)$ and a **normal vector** $\boldsymbol{n} = \langle a, b, c\rangle$ is described by the equations:*

$$\begin{array}{lll} \text{Vector form:} & \boldsymbol{n} \cdot \langle x, y, z\rangle & = d \\ \text{Scalar forms:} & a(x - x_0) + b(y - y_0) + c(z - z_0) & = 0 \\ & ax + by + cz & = d \end{array}$$

Where we set $d = ax_0 + by_0 + cz_0$.

Proposition 1.4.18. *Given a vector $\boldsymbol{n} \in \mathbb{R}^n$, the set*

$$\{\boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{v} \cdot \boldsymbol{n} = 0\}$$

is a vector subspace of $\mathbb{R}^n$ *of dimension* $n-1$.

From the previous proposition, we can now generalize our examples from $\mathbb{R}^2$ and $\mathbb{R}^3$:

Definition 1.4.19. *Let* $\boldsymbol{n} \in V$*, with* $\boldsymbol{n} \neq \boldsymbol{0}$*. The* ***hyperplane*** W *normal to* $\boldsymbol{n}$ *(passing through the origin) is the subspace*

$$W = \{\boldsymbol{v} \in V \mid \boldsymbol{n} \cdot \boldsymbol{v} = 0\}$$

We say that $\boldsymbol{n}$ *is a normal vector of* W*.*

Observe that we can also use vector addition to describe arbitrary planes in $\mathbb{R}^n$ - that is, in order to define a plane P in $\mathbb{R}^n$ (not necessarily passing through the origin), we simply need to describe a point on the plane P, and two non-parallel vectors. That is, we are describing planes as two-dimensional objects in the following way:

Definition 1.4.20. *The plane* $\mathscr{P}$ *through the point* $P = (x_1, \cdots, x_n)$ *and determined by two non-parallel vectors* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$*, can be described by the vector function* $\boldsymbol{r}(s,t) : \mathbb{R}^2 \to \mathbb{R}^n$ *defined by*

$$\boldsymbol{r}(s,t) = \boldsymbol{r_0} + s\boldsymbol{u} + t\boldsymbol{v}$$

where $\boldsymbol{r_0}$ *is the vector* $\boldsymbol{r_0} = \boldsymbol{OP} = \langle x_1, \cdots, x_n \rangle$*.*

We call $\boldsymbol{r}(s,t)$ *the* ***parametrization of*** $\mathscr{P}$*.*

That is, the plane is the set of all linear combinations of $\boldsymbol{u}$ and $\boldsymbol{v}$ that originate from P.

Remark 1.4.21. *In other words, we are describing a plane as the image of of a linear map, whereas we can also describe the plane as the kernel of a linear map.*

How can we reconcile these two notions?

This leads us to the notion of the cross product of vectors in $\mathbb{R}^3$.

Cross product

We saw previously that two non-zero vectors in $\mathbb{R}^3$ span a plane P in $\mathbb{R}^3$. Is there a way to determine the normal vector $\boldsymbol{n}$ associated to this plane?

In other words, given two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$, can we construct a vector $\boldsymbol{w} \in \mathbb{R}^3$ that is orthogonal to both $\boldsymbol{u}$ and $\boldsymbol{v}$?

The answer is yes (at least, in $\mathbb{R}^3$). We will define the cross product of two vectors in $\mathbb{R}^3$ in a few different ways (see definition 1.4.23, definition 1.4.24, and theorem 1.4.26). Each of these ways has its strengths.

We will first define the cross product in terms of direction and magnitude.

It turns out that in order to describe the direction of the cross product, we will need to make an arbitrary choice. This choice is called the right hand rule:

Definition 1.4.22. *We say that an ordered list of three orthogonal vectors* $\boldsymbol{u}$*,* $\boldsymbol{v}$ *and* $\boldsymbol{w}$ *in* $\mathbb{R}^3$ *satisfy the* ***Right Hand Rule*** *if the following is true.*

a. *Point the fingers of your right hand in the direction of* $\boldsymbol{u}$

b. *Curl your fingers towards* $\boldsymbol{v}$ *(maybe rotating your hand!)*

c. *If your thumb points in the direction of* $\boldsymbol{w}$*, then* $\boldsymbol{u}$*,* $\boldsymbol{v}$ *and* $\boldsymbol{w}$ *satisfy the right hand rule.*

Note that the order matters! If the ordered list $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ satisfies the right hand rule, then the ordered list $\boldsymbol{v}$, $\boldsymbol{u}$ and $\boldsymbol{w}$ does not satisfy the right hand rule.

For example, consider the standard basis vectors of $\mathbb{R}^3$: $\boldsymbol{i} = \langle 1,0,0\rangle$, $\boldsymbol{j} = \langle 0,1,0\rangle$, and $\boldsymbol{k} = \langle 0,0,1\rangle$.

We see that the ordered list $\boldsymbol{i}$, $\boldsymbol{j}$, and $\boldsymbol{k}$ does satisfy the right hand rule, while the ordered list $\boldsymbol{i}$, $\boldsymbol{k}$, and $\boldsymbol{j}$ doesn't.

We can now define the cross product in terms of direction and magnitude.

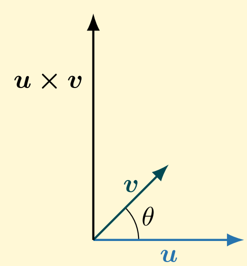
Definition 1.4.23 (Magnitude and Direction)**.** *Given two vectors* $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$*, with an angle* θ *between them, their* ***cross product*** *is the* unique *vector* $\boldsymbol{u} \times \boldsymbol{v} \in \mathbb{R}^3$ *with the following properties:*

(i) $\boldsymbol{u} \times \boldsymbol{v}$ *is orthogonal to both* $\boldsymbol{u}$ *and* $\boldsymbol{v}$*.*

(ii) $\boldsymbol{u}$*,* $\boldsymbol{v}$*, and* $\boldsymbol{u} \times \boldsymbol{v}$ *satisfies the Right Hand Rule.*

(iii) $||\boldsymbol{u} \times \boldsymbol{v}|| = ||\boldsymbol{u}||\ ||\boldsymbol{v}|| \sin(\theta)$

Geometrically, we can draw the cross product as follows:



Note that together, (i) and (ii) specify a unique direction for the cross product - in $\mathbb{R}^3$, the set of vectors orthogonal to both $\boldsymbol{u}$ and $\boldsymbol{v}$ spans a one-dimensional subspace (a.k.a. a line). Thus, to specify one out of the two directions in this line, we must make an arbitrary choice via the right hand rule. Specifying a magnitude via (3) thus gives us a unique vector. We will discuss later the choice of magnitude later in 1.4.28.

What is the cross product $\boldsymbol{i} \times \boldsymbol{j}$?

What is the cross product $\boldsymbol{j} \times \boldsymbol{k}$?

How can we compute the cross product algebraically? In other words, what are the components of the cross product?

We now give a second, algebraic definition of the cross product. This will lead to a formula for the components of the cross product (theorem 1.4.26).

Definition 1.4.24 (Algebraic characterization)**.** *Given two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$, their* ***cross product*** *is the* unique *vector $\boldsymbol{u} \times \boldsymbol{v} \in \mathbb{R}^3$ defined by the property:*

$$(\boldsymbol{u} \times \boldsymbol{v}) \cdot \boldsymbol{w} = \det \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \\ \boldsymbol{w} \end{bmatrix} \qquad \text{for all } \boldsymbol{w} \in \mathbb{R}^3$$

This is an interesting perspective on defining a vector in $\mathbb{R}^3$: it is determined by how it relates to all other vectors in $\mathbb{R}^3$. This kind of idea shows up a lot in higher level mathematics!

First of all, using this algebraic perspective, we can prove the following properties of the cross product:

Theorem 1.4.25. *Given two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$,*

a. $\boldsymbol{u} \times \boldsymbol{v} = -\boldsymbol{v} \times \boldsymbol{u}$ *(anti-commutativity).*

b. $\boldsymbol{u} \times \boldsymbol{v}$ *is orthogonal to $\boldsymbol{u}$ and $\boldsymbol{v}$.*

c. *The cross product $\times : \mathbb{R}^3 \to \mathbb{R}^3 \to \mathbb{R}^3$ is bilinear.*

d. $\boldsymbol{u} \times \boldsymbol{v} = \boldsymbol{0}$ *if and only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are parallel.*

Moreover, we can use the algebraic characterization of the cross product to find the components of the cross product. Recall from proposition 1.4.11, if we consider the standard basis for $\mathbb{R}^3$, $\mathscr{B} = \{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}\}$, we have that

$$\begin{aligned} \boldsymbol{u} \times \boldsymbol{v} &= (\boldsymbol{u} \times \boldsymbol{v} \cdot \boldsymbol{e_1})\boldsymbol{e_1} + (\boldsymbol{u} \times \boldsymbol{v} \cdot \boldsymbol{e_2})\boldsymbol{e_2} + \langle(\boldsymbol{u} \times \boldsymbol{v} \cdot \boldsymbol{e_3})\rangle\boldsymbol{e_3} \\ &= \det \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \\ \boldsymbol{e_1} \end{bmatrix} \boldsymbol{e_1} + \det \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \\ \boldsymbol{e_2} \end{bmatrix} \boldsymbol{e_2} + \det \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \\ \boldsymbol{e_3} \end{bmatrix} \boldsymbol{e_3} \\ &= \left\langle \det \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \\ \boldsymbol{e_1} \end{bmatrix}, \det \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \\ \boldsymbol{e_2} \end{bmatrix}, \det \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \\ \boldsymbol{e_3} \end{bmatrix} \right\rangle \end{aligned}$$

Thus, given the components of two vectors two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$, we can compute the components of the cross product $\boldsymbol{u} \times \boldsymbol{v}$.

Theorem 1.4.26. *Given two vectors $\boldsymbol{u} = \langle x_1, y_1, z_1\rangle$ and $\boldsymbol{v} = \langle x_2, y_2, z_2\rangle$, their* ***cross product*** $\boldsymbol{u} \times \boldsymbol{v}$ *can described as:*

$$\begin{aligned} \boldsymbol{u} \times \boldsymbol{v} &= \left\langle \det \begin{bmatrix} y_1 & z_1 \\ y_2 & z_2 \end{bmatrix}, -\det \begin{bmatrix} x_1 & z_1 \\ x_2 & z_2 \end{bmatrix}, \det \begin{bmatrix} x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \right\rangle \\ &= \langle y_1 z_1 - y_2 z_1,\ x_2 z_1 - x_1 z_2,\ x_1 y_2 - x_2 y_1 \rangle \end{aligned}$$

We saw previously that two non-zero, non-parallel vectors in $\mathbb{R}^3$ (or equivalently, 3 non-collinear points) span a plane P in $\mathbb{R}^3$, and we can now determine the normal vector $\boldsymbol{n}$ associated to this plane:

Find the equation for the plane P determined by the points $P = (1, 0, -1)$, $Q = (2, 2, 1)$, $R = (4, 1, 2)$.

Observe that $\boldsymbol{PQ} \times \boldsymbol{PR}$ is orthogonal to the plane, and we can use any of P, Q, or R to determine the plane.

How are these two definitions of the cross product related to each other?

In other words, (1) what does the right hand rule have to do with determinants? (2) why did we choose the magnitude of $\boldsymbol{u} \times \boldsymbol{v}$ to be $||\boldsymbol{u}||\ ||\boldsymbol{v}|| \sin(\theta)$?

More generally, what is the geometry of the determinant?

Proposition 1.4.27. *The right hand rule corresponds to having a positive determinant.*

explain (relaxed right-hand rule, adding scalar multiples does not change right-handedness - turn into diagonal matrix).

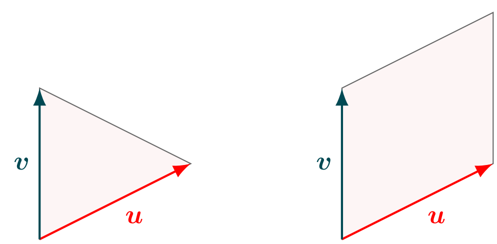
The standard basis $\{\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$ satisfies the right hand rule, and det $\begin{bmatrix} \boldsymbol{e_1} \\ \boldsymbol{e_2} \\ \boldsymbol{e_3} \end{bmatrix}$ has positive determinant.

We now turn to question 2: Why did we choose the magnitude of $\boldsymbol{u} \times \boldsymbol{v}$ to be $||\boldsymbol{u}||\ ||\boldsymbol{v}|| \sin(\theta)$?

Proposition 1.4.28. *Given two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$, with an angle θ between them, then*

$$||\boldsymbol{u} \times \boldsymbol{v}|| = ||\boldsymbol{u}||\ ||\boldsymbol{v}|| \sin(\theta)$$

First, observe that if two non-parallel vectors (say, $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$) start at the same point, then they uniquely determine a triangle, and a parallelogram.



Theorem 1.4.29. *Let P be the parallelogram spanned by $\boldsymbol{u}$ and $\boldsymbol{v}$, and T be the triangle spanned by $\boldsymbol{u}$ and $\boldsymbol{v}$. Then*

$$\text{area}(P) = ||\boldsymbol{u} \times \boldsymbol{v}|| \qquad \text{and} \qquad \text{area}(T) = \frac{1}{2}||\boldsymbol{u} \times \boldsymbol{v}||$$

Observe that by geometry, the area of P is twice the area of the triangle T. Furthermore, if the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ is θ, we can calculate the area of the parallelogram as begin base times height, e,g,

$$(||\boldsymbol{u}||)(||\boldsymbol{v}|| \sin(\theta)) = ||\boldsymbol{u} \times \boldsymbol{v}||$$

However, here's another way to think about this parallelogram: Without loss of generality, we can rotate our vectors so that they lie in the xy-plane. (Equivalently, we can choose a basis $\mathcal{B}$ so that the coordinates of $\boldsymbol{u} = \langle a, b, 0\rangle$, and that $\boldsymbol{v} = \langle c, d, 0\rangle$).

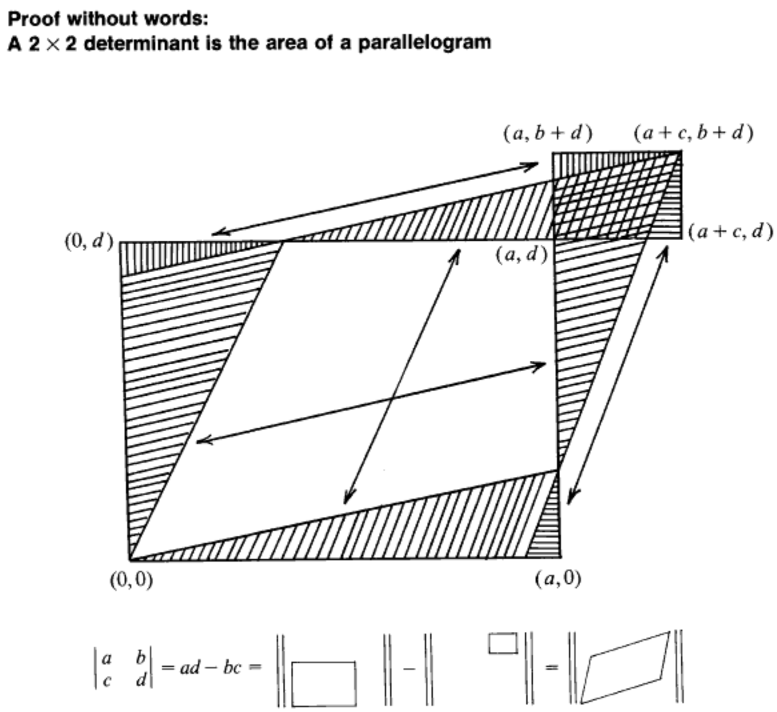


Figure 1.1: A Proof without Words, by Solomon W. Golomb (Mathematics Magazine, March 1985)

Observe that the figure above shows that

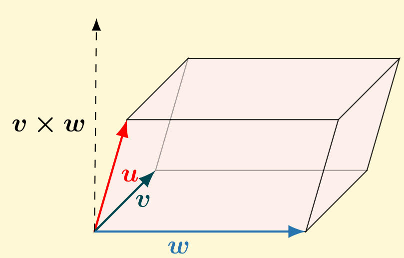
$$\left|\det \begin{bmatrix} a & b \\ c & d \end{bmatrix}\right| = \text{area}(P)$$

That is, the absolute value of a 2×2 determinant equals the area of the parallelograms spanned by the rows.

In other words, the determinant of a 2×2 matrix A measures the area of the parallelogram spanned by $A\boldsymbol{e_1} = \langle a, c\rangle$ and $A\boldsymbol{e_2} = \langle b, d\rangle$. That is, it measures the area of the image of the unit square under the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\boldsymbol{v}) = A\boldsymbol{v}$.

Does this generalize to linear transformations in higher dimensions?

Definition 1.4.30. *Given three vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ in $\mathbb{R}^3$, they span a **parallelpiped**. That is, a 3-dimensional solid, such that each face is a parallelogram.*



We can again use geometry to compute the volume of this parallelpiped:

Theorem 1.4.31. *Let D be the parallelpiped spanned by $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$, and let θ be the angle between $\boldsymbol{u}$ and $\boldsymbol{v} \times \boldsymbol{w}$. Then*

$$\text{volume}(D) = ||\boldsymbol{v} \times \boldsymbol{w}||\ ||\boldsymbol{u}||\ |\cos(\theta)|$$

Observe that we can write this formula in terms of a dot product.

Definition 1.4.32. *The scalar $\boldsymbol{u} \cdot (\boldsymbol{v} \times \boldsymbol{w})$ is called the* ***scalar triple product****.*

Let D be the parallelpiped spanned by $\boldsymbol{u} = \langle 1, 3, 2\rangle$, $\boldsymbol{v} = \langle 6, 5, 4\rangle$, and $\boldsymbol{w} = \langle 2, 2, 2\rangle$. Compute the volume of D.

This computation should seem familiar to you: see example 1.3

Theorem 1.4.33. *Let D be the parallelpiped spanned by $\boldsymbol{u} = \langle u_1, u_2, u_3\rangle$, $\boldsymbol{v} = \langle v_1, v_2, v_3\rangle$, and $\boldsymbol{w} = \langle w_1, w_2, w_3\rangle$. Then*

$$\text{volume}(D) = |\boldsymbol{u} \cdot (\boldsymbol{v} \times \boldsymbol{w})| = |\det \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \\ \boldsymbol{w} \end{bmatrix}|$$

Definition 1.4.34. *We say that three vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^n$ are* ***co-planar*** *if they all lie in the same plane.*

Show that if $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^3$ are co-planar, then $\text{volume}(D) = 0$. What is the geometric interpretation of this fact?

Just as we saw in the 2×2 case, we see again that the determinant of a 3×3 matrix A measures the volume of the parallelpiped spanned by $A\boldsymbol{e_1}$, $A\boldsymbol{e_2}$, and $A\boldsymbol{e_3}$. That is, it measures the volume of the image of the unit cube under the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $T(\boldsymbol{v}) = A\boldsymbol{v}$.

This generalizes to $n \times n$ matrices and their associated linear transformations $T : \mathbb{R}^n \to \mathbb{R}^n$. The determinant will measure the n-dimensional volume of the n-dimensional parallelpiped spanned by the image of the unit n-cube. In other words, the absolute value of $\det A$ is the factor by which T magnifies volume; and $\det A$ is zero if and only if T is not invertible.

> “The determinant is astonishing." — Jerry Shurman, Calculus and Analysis in Euclidean Space.

Inner products

How can we generalize the dot product in $\mathbb{R}^n$ to an arbitrary vector space V?

For $\mathbb{R}^n$, we saw that the dot product satisfies the property $\boldsymbol{v} \cdot \boldsymbol{v} = ||\boldsymbol{v}||^2$. Thus, our generalization should allow us to define a notion of distance in an abstract vector space.

Definition 1.4.35. *Given two vectors $\boldsymbol{u}, \boldsymbol{v} \in V$, an* ***inner product*** *on V is a map $\langle -, - \rangle : V \times V \to \mathbb{R}$ that is:*

a. $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = \langle \boldsymbol{v}, \boldsymbol{u}\rangle$ *(symmetric).*

b. $\langle -, - \rangle$ *is bilinear.*

c. $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$, *with equality only when* $\boldsymbol{v} = \mathbf{0}$. *(positive definite).*

Definition 1.4.36. *Given an innner product*$\langle -, - \rangle$ *on* V, *the* ***norm*** *(or* ***modulus****) of a vector* $\boldsymbol{v} \in V$ *is defined as*

$$||\boldsymbol{v}|| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v} \rangle}$$

Given two vectors $\boldsymbol{u} = \langle u_1, \cdots, u_n \rangle$ and $\boldsymbol{v} = \langle v_1, \cdots, v_n \rangle$ in $\mathbb{R}^n$, their **dot product** is an inner product.

$$\boldsymbol{u} \cdot \boldsymbol{v} = \sum u_i v_i$$

Theorem 1.4.37 (Properties of an inner product)**.** a. *Cauchy-Schwarz inequality:* $|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq |\boldsymbol{u}|\ |\boldsymbol{v}|$

b. *Triangle inequality:* $|\boldsymbol{u} + \boldsymbol{v}| \leq |\boldsymbol{u}| + |\boldsymbol{v}|$

What is the geometric interpretation of the inner product for an arbitrary vector space?

Definition 1.4.38. *A subset* $S = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_k}\} \subseteq V$ *is said to be* ***orthogonal*** *if* $\langle \boldsymbol{v_i}, \boldsymbol{v_j} \rangle = 0$ *for all* $i \neq j$.

Furthermore, if $|\boldsymbol{v_i}| = 1$ *for all* $1 \leq i \leq k$, *we say that the subset* $S = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_k}\} \subseteq V$ *is* ***orthonormal****.*

Proposition 1.4.39. *Equivalently, a set subset* $S = \{\boldsymbol{v_1}, \boldsymbol{v_2}, \cdots, \boldsymbol{v_k}\} \subseteq V$ *is orthonormal if*

$$\langle \boldsymbol{v_i}, \boldsymbol{v_j} \rangle = \delta_{ij}$$

where δ_{ij} *is the* ***Kronecker delta****, which is the symbol*

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Exercises

1.64 Find all values of b such that the vectors $\boldsymbol{u} = \langle b, 3, 2 \rangle$ and $\boldsymbol{v} = \langle 1, b, 1 \rangle$ are orthogonal.

1.65 Given that $||\boldsymbol{u}|| = 1$, $||\boldsymbol{v}|| = 3$, and $\boldsymbol{u} \cdot \boldsymbol{v} = 2$, evaluate the expression $2\boldsymbol{u} \cdot (3\boldsymbol{u} - \boldsymbol{v})$.

1.66 Given that $||\boldsymbol{u}|| = 2$, $||\boldsymbol{v}|| = 3$, and the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ is $120°$, calculate $||3\boldsymbol{u} - \boldsymbol{v}||^2$.

1.67 Consider the points $P = (0, 8, 3)$, $Q = (1, 3, 0)$, and $R = (1, 5, -2)$. Does the triangle PQR have (A) an obtuse angle (and two acute angles); (B) a right angle (and two acute angles); or (C) three acute angles?

1.68 Consider the points $P = (0, 8, 3)$, $Q = (1, 3, 0)$, and $R = (1, 5, -2)$, find the projection of $\boldsymbol{PQ}$ in the direction of $\boldsymbol{PR}$.

1.69 Find the projection of $\boldsymbol{u} = \langle 5, 7, -4 \rangle$ along $\boldsymbol{v} = \langle 0, 0, 1 \rangle$.

1.70 Find the projection of $\boldsymbol{u} = \langle a, a, b \rangle$ along $\boldsymbol{v} = \boldsymbol{i} - \boldsymbol{j}$.

1.71 Given $\boldsymbol{u} = \langle 4, -1, 2 \rangle$, and $\boldsymbol{v} = \langle 1, 0, 1 \rangle$

a. Find the vectors $\boldsymbol{u}_{||\boldsymbol{v}}$ and $\boldsymbol{u}_{\perp\boldsymbol{v}}$

b. Find the vectors $\boldsymbol{v}_{||\boldsymbol{u}}$ and $\boldsymbol{v}_{\perp\boldsymbol{u}}$

1.72 Write the equation of the plane in $\mathbb{R}^3$ with normal vector $\boldsymbol{n} = \langle -1, 2, 1\rangle$ passing through the point $(4, 1, 5)$.

1.73 Find the equation of the plane in $\mathbb{R}^3$ passing through the points $P = (5, 1, 1)$, $Q = (1, 1, 2)$, $R = (2, 1, 1)$.

1.74 Find the equation of the plane in $\mathbb{R}^3$ that contains the lines

$$\boldsymbol{r_1}(t) = \langle 2 + t, 2 + 3t, 3 + t\rangle$$

$$\boldsymbol{r_2}(s) = \langle 5 + s, 15 + 4s, 10 + 2s\rangle$$

1.75 Consider the plane P in $\mathbb{R}^3$ given by the equation $3x + 5y - 2z = 29$. Find the equation of the plane R that is parallel to P and passes through the point $(3, -1, 1)$.

Definition 1.4.40. *Let A be a nonempty subset of an inner product space V. The* ***orthogonal complement*** *of A is the subspace*

$$A^{\perp} := \{\boldsymbol{v} \in V \mid \langle \boldsymbol{a}, \boldsymbol{v}\rangle = 0 \textit{ for every } \boldsymbol{a} \in A\}$$

1.76 Prove that $A^{\perp}$ is a vector subspace.

Let $A = \{\boldsymbol{u_1}, \cdots \boldsymbol{u_k}\}$ be k linearly independent vectors in a vector space V of dimension n. Prove that $A^{\perp}$ is a subspace of dimension $n - k$.

Describe $A^{\perp}$ is the intersection of k hyperplanes.

1.77 Calculate the cross product of $\boldsymbol{u} = \langle 1, 1, 0\rangle$ and $\boldsymbol{v} = \langle 0, 1, 1\rangle$

1.78 Consider the points $P = (0, 8, 3)$, $Q = (1, 3, 0)$, and $R = (1, 5, -2)$, find the area of the parallelogram spanned by $\boldsymbol{PQ}$ and $\boldsymbol{PR}$.

1.79 Find the area of the parallelogram determined by the vectors $\langle a, 0, 0\rangle$ and $\langle 0, b, c\rangle$

1.80 Consider the points $P = (0, 8, 3)$, $Q = (1, 3, 0)$, and $R = (1, 5, -2)$, find the area of the triangle spanned by $\boldsymbol{PQ}$ and $\boldsymbol{PR}$.

1.81 Consider the points $P = (-1, 8, 3)$, $Q = (2, 3, 0)$, $R = (2, 5, -2)$, and $S = (3, 5, 2)$. Find the volume of the parallelpiped that has vertices P, Q, R, and S.

1.82 Find the volume of the parallelpiped in $\mathbb{R}^3$ spanned by the vectors $\boldsymbol{u} = \langle 1, 0, 4\rangle$, $\boldsymbol{v} = \langle 1, 3, 1\rangle$, $\boldsymbol{w} = \langle -4, 2, 6\rangle$.

1.83 Let P be the parallelpiped contained in the first octant (that is, all points in P satisfy $x \geq 0$, $y \geq 0$, $z \geq 0$) determined by the vertices $A = (2, 4, 1)$, $B = (5, 2, 5)$, $C = (1, 2, 4)$, and $D = (1, 1, 1)$. Find the point X in the parallelpiped P that is **furthest** from the origin. (**Hint:** Draw a picture).

1.84 Do the points $P = (0, 1, 1)$, $Q = (1, 1, 2)$, $R = (3, 3, 3)$ determine a plane or a line? Find the equation of either the plane or the line.

1.85 Let P be the plane given by the equation $-3x - 4y + 2z = -10$. The plane P intersects the plane $z = 0$ at some **acute** angle θ. Determine the value of $\cos(\theta)$.

1.86 The intersection of the plane $-3x - 4y + 2z = -10$ and the plane $z = 0$ is a line L. Find a vector parametrization of the line L.

1.87 The line $\boldsymbol{r}(t) = \langle 2, 2, 8\rangle + t\langle 0, 1, 2\rangle$ intersects the sphere $x^2 + (y + 3)^2 + (z + 2)^2 = 9$ in two points. Find these two points.

Chapter 2

Analysis on $\mathbb{R}^n$

Having studied the vector space structure of $\mathbb{R}^n$ and the geometric behavior of linear maps, we will now turn our focus to studying **non-linear** functions from $\mathbb{R}^n$ to $\mathbb{R}^m$. In particular, this will lead us to study the geometric and topological properties of $\mathbb{R}^n$.

In section 2.1, we will first explore functions $\mathbb{R} \to \mathbb{R}^n$ (vector-valued functions), as well as functions $\mathbb{R}^n \to \mathbb{R}$ (multivariable functions). These turn out to be enough to capture the behavior of general multivariable functions $\mathbb{R}^n \to \mathbb{R}^m$.

We will then study the notions of limits (section 2.2) and continuity (section 2.4) for these classes of functions, and see how it differs from limits and continuity in the single-variable case.

2.1 Multivariable functions

Vector-valued functions

We will begin by studying vector-valued functions. That is, maps of the form $f : \mathbb{R} \to \mathbb{R}^n$. Note that the domain of f is $\mathbb{R}$, so that the input of this function is a scalar. On the other hand, the codomain of f is $\mathbb{R}^n$, which means that the output is a point (position vector), hence the name vector-valued function.

Definition 2.1.1. *A **vector-valued function** is a function $\boldsymbol{r}(t) : \mathbb{R} \to \mathbb{R}^n$, given by*

$$\boldsymbol{r}(t) = \langle x_1(t), x_2(t), \cdots x_n(t) \rangle = x_1(t)\boldsymbol{e_1} + x_2(t)\boldsymbol{e_2} + \cdots + x_n(t)\boldsymbol{e_n}$$

where all of the $x_i(t) : \mathbb{R} \to \mathbb{R}$ are single-variable functions.

*$\boldsymbol{r}(t)$ is also called a **parametrization** (with **parameter** t).*

In other words, vector-valued functions are precisely described by their component functions $x_i(t)$. We will see that this will cause vector-valued functions $\boldsymbol{r}(t) : \mathbb{R} \to \mathbb{R}^n$ to behave extremely similarly to scalar functions $f : \mathbb{R} \to \mathbb{R}$. In some sense, a vector-valued function $\boldsymbol{r}(t) : \mathbb{R} \to \mathbb{R}^n$ is simply a collection of n single-variable scalar functions.

How should we geometrically interpret vector-valued functions?

We should think of a vector-valued function $\boldsymbol{r}(t)$ as describing the motion of a particle in $\mathbb{R}^n$. For example, let us think of the units of t as begin seconds. Then we can interpret the vector $\boldsymbol{r}(0)$ as describing the position of a particle at time $t = 0$. We can then think of the vector $\boldsymbol{r}(1)$ as describing the position of the particle 1 second after $t = 0$; and similarly, the vector $\boldsymbol{r}(-1)$ describes the position of the particle 1 second **before** $t = 0$.

Let us turn to some examples:

A scalar function $f : \mathbb{R} \to \mathbb{R}$ is a vector-valued function for $n = 1$.

Sketch the vector-valued function $\boldsymbol{c}(t) = \langle -\sin(t), \cos(t)\rangle$ in $\mathbb{R}^2$.

Sketch the vector-valued function $\boldsymbol{h}(t) = \langle -\sin(t), \cos(t), t\rangle$ in $\mathbb{R}^3$.

Sketch the vector-valued function $\boldsymbol{r}(s) = \langle -\sin(5s), \cos(5s), 5s\rangle$ in $\mathbb{R}^3$.

How do the two examples $\boldsymbol{h}(t)$ and $\boldsymbol{r}(s)$ differ from each other?

Definition 2.1.2. *Given a parametrization $\boldsymbol{r}(t) : \mathbb{R} \to \mathbb{R}^n$, we say that its* ***parametric curve*** *is the set of points*

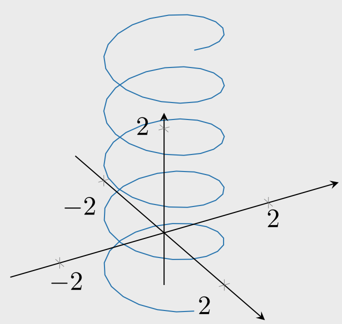
$$\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} = \boldsymbol{r}(t)\} \subset \mathbb{R}^n$$

That is, we can geometrically interpret the parametric curve as being the path traced out by the parametrization $\boldsymbol{r}(t)$.

The parametric curve of $\boldsymbol{c}(t) = \langle -\sin(t), \cos(t)\rangle$ is the circle of radius 1, centered at the origin. That is, it is the set of points $\{(x, y) \mid x^2 + y^2 = 1\}$

Remark 2.1.3. *The two examples $\boldsymbol{h}(t)$ and $\boldsymbol{r}(s)$ show us that there are multiple possible parametrizations of the same curve.*

The parametric curve of $\boldsymbol{h}(t) = \langle -\sin(t), \cos(t), t\rangle$ and $\boldsymbol{r}(s) = \langle -\sin(5s), \cos(5s), 5s\rangle$ is called a helix.



In other words, the parametrization describes how fast and in what direction you walk the path, whereas the parametric curve is the path itself.

How can we sketch parametrizations and parametric curves?

Plotting points is something that a computer can do pretty easily; however, plotting points by hand is often a very inefficient way to study parametric curves. We will discuss two tools that can help us more easily sketch and understand vector-valued functions in $\mathbb{R}^3$.

Projections

Given a vector-valued function $\boldsymbol{r}(t) = \langle x(t), y(t), z(t)\rangle$, if we isolate and study the behavior of any two components, then we can gain information about the parametrization.

Definition 2.1.4. *Given a vector-valued function $\boldsymbol{r}(t) = \langle x(t), y(t), z(t)\rangle$, we can study the following **projections**:*

- *The projection onto the xy-plane, which corresponds to $\boldsymbol{r}(t) = \langle x(t), y(t), 0\rangle$*
- *The projection onto the xz-plane, which corresponds to $\boldsymbol{r}(t) = \langle x(t), 0, z(t)\rangle$*
- *The projection onto the yz-plane, which corresponds to $\boldsymbol{r}(t) = \langle 0, y(t), z(t)\rangle$*

That is, we can think of these as shadows of the vector-valued function $\boldsymbol{r}(t)$.

Consider the helix parametrization $\boldsymbol{h}(t) = \langle -\sin(t), \cos(t), t\rangle$ in $\mathbb{R}^3$.

The projection onto the xy-plane corresponds to $\boldsymbol{r}(t) = \langle -\sin(t), \cos(t), 0\rangle$. This describes counterclockwise motion around the unit circle centered at the origin.

The projection onto the xz-plane corresponds to $\boldsymbol{r}(t) = \langle -\sin(t), 0, t\rangle$. This corresponds to a sine curve in the xz-plane.

The projection onto the yz-plane corresponds to $\boldsymbol{r}(t) = \langle 0, \cos(t), t\rangle$. This corresponds to a sine curve in the yz-plane.

picture

Quadric surfaces in $\mathbb{R}^3$

Another tool we can use to graph and sketch vector-valued functions in $\mathbb{R}^3$ are called quadric surfaces: These are certain well-studied surfaces in $\mathbb{R}^3$, that have been completely classified.

Definition 2.1.5. *A **quadric surface** in $\mathbb{R}^3$ is defined as the set of points*

$$\{(x, y, z) \mid Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + ax + by + cz + d = 0\}$$

A plane of the form $ax + by + cz + d = 0$ is a quadric surface.

Thus, studying these quadric surfaces is a generalization of our discussion of planes. We are now allowing for non-linear behavior!

say more

Let us investigate a few non-linear examples:

Let us consider the quadric surface C given by

$$C := \{(x, y, z) \mid x^2 + y^2 = 1\}$$

Observe that if we restrict our attention to the points that lie in both C and in the plane $z = 0$, we see that these points can be described precisely as a unit circle centered at the origin, in the $z = 0$ plane.

$$\{(x, y, 0) \mid x^2 + y^2 = 1\}$$

Furthermore, this same picture holds if we instead consider any plane of the form $z = k$! We again see a unit circle centered at the origin, but now in the $z = k$ plane.

$$\{(x, y, k) \mid x^2 + y^2 = 1\}$$

Thus, the surface C consists of a collection (one for each $k \in \mathbb{R}$) of unit circles centered at the origin, stacked on top of each other. In other words, C is an infinite cylinder of radius 1.

define cylinders

We can use this technique in general to study the solutions to an arbitrary equation in $\mathbb{R}^3$ ref later

Definition 2.1.6. *The* ***trace in the plane*** P *of a subset* $S \subset \mathbb{R}^3$ *is the intersection of* S *with* P*. That is,*

$$S \cap P = \{\boldsymbol{x} \in \mathbb{R}^3 \mid \boldsymbol{x} \in S \text{and} \boldsymbol{x} \in P\}$$

We typically study x*-traces,* y*-traces, and* z*-traces - that is, traces in the planes* $x = k$*,* $y = k$*, and* $z = k$ *respectively.*

We can think of traces as being slices of the graph. say more

Theorem 2.1.7. *Up to rotation and translation, there are 3 types of quadratic cylinders:*

- *Parabolic cylinders, which have standard form*

$$y = \left(\frac{x}{a}\right)^2$$

- *Elliptic cylinders, which have standard form*

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1$$

- *Hyperbolic cylinders , which have standard form*

$$\left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2 = 1$$

Theorem 2.1.8. *Up to rotation and translation, there are 8 types of quadric surfaces:*

- *Planes, which have standard form*

$$ax + by + cz = d$$

- *Quadratic cylinders (described in theorem 2.1.7).*
- *Elliptic paraboloids, which have standard form*

$$z = \left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2$$

- *Hyperbolic paraboloids, which have standard form*

$$z = \left(\frac{x}{a}\right)^2 - \left(\frac{y}{b}\right)^2$$

- *Ellipsoids, which have standard form*

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 + \left(\frac{z}{c}\right)^2 = 1$$

- *Elliptic cones, which have standard form*

$$\left(\frac{z}{c}\right)^2 = \left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2$$

- *Hyperboloids of one sheet, which have standard form*

$$\left(\frac{z}{c}\right)^2 + 1 = \left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = \left(\frac{z}{c}\right)^2 + 1$$

- *Hyperboloids of two sheets, which have standard form*

$$\left(\frac{z}{c}\right)^2 - 1 = \left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2$$

That is, given any quadric surface, corresponding to a quadratic equation $Ax^2+By^2+Cz^2+Dxy+Exz+Fyz+ax+by+cz+d=0$, one can rotate it and translate it so that the quadratic equation can be written in the standard form above.

The unit sphere of radius 1 can be described as the set of vectors $\{\boldsymbol{v} \in \mathbb{R}^3 \mid ||\boldsymbol{v}|| = 1\}$. Equivalently, the unit sphere is the set of points

$$\{(x,y,z) \; | x^2+y^2+z^2 = 1\}$$

Thus, the unit sphere of radius 1 is an example of an ellipsoid!

Sketch and describe the surface

$$x = \left(\frac{y-1}{4}\right)^2 + \left(\frac{z}{3}\right)^2$$

Describe and sketch the the quadric surface S given by

$$\frac{9(x-1)^2}{4} + 4z^2 = 9y^2 + 36$$

How can we use this to sketch vector-valued functions?

Given a vector-valued function $\boldsymbol{r}(t) = \langle x(t), y(t), z(t)\rangle$, if we can find a relationship between the components of the coordinate vectors, then we know that $\boldsymbol{r}(t)$ must lie on the quadric surface.

Consider the helix parametrization $\boldsymbol{h}(t) = \langle -\sin(t), \cos(t), t\rangle$ in $\mathbb{R}^3$.

Observe that the components of $\boldsymbol{h}(t)$ satisfy the equation $x(t)^2 + y(t)^2 = 1$. This tells us indeed that the helix lies on the cylinder $x^2+y^2=1$.

draw picture

Multivariable functions

We now to turn to studying functions of the form $f : \mathbb{R}^n \to \mathbb{R}$:

Definition 2.1.9. *A* ***multivariable function*** *is a function $f : \mathbb{R}^n \to \mathbb{R}$. We sometimes say that f is a* ***function of*** n ***variables****.*

A scalar function $f : \mathbb{R} \to \mathbb{R}$ is a function of 1 variable.

The distance between the origin and a point $P = (x_1, \cdots, x_n)$ is a multivariable function $d : \mathbb{R}^n \to \mathbb{R}$ defined by

$$d(x_1, \cdots, x_n) = \sqrt{\sum_{i=1}^{n} x_i^2}$$

Definition 2.1.10. *Consider a subset $D \subset \mathbb{R}^n$. The* ***indicator function*** 1_D *is the piecewise function defined by*

$$1_D(\boldsymbol{x}) := \begin{cases} 1 & \textit{if } \boldsymbol{x} \in D \\ 0 & \textit{if } \boldsymbol{x} \notin D \end{cases}$$

Just like with single-variable functions, we might be interested in functions that are defined only on a subset $D \subseteq \mathbb{R}^n$. Nevertheless, the following construction shows us that it is sufficient to study the theory of multivariable functions $f : \mathbb{R}^n \to \mathbb{R}$.

Let $D \subset \mathbb{R}^n$. Then a function $f : D \to \mathbb{R}$ can be extended to a function of n variables $\tilde{f} : \mathbb{R}^n \to \mathbb{R}$ in the following way:

$$\tilde{f}(\boldsymbol{x}) := \begin{cases} f(\boldsymbol{x}) & \text{if } \boldsymbol{x} \in D \\ 0 & \text{if } \boldsymbol{x} \notin D \end{cases}$$

That is, to every point $\boldsymbol{x} = (x_1, \cdots, x_n)$ in the domain D, we assign the real number $f(\boldsymbol{x}) \in \mathbb{R}$. To every other point $\boldsymbol{x} \notin D$, we assign the value 0. In other words, $\tilde{f}(\boldsymbol{x}) = f(\boldsymbol{x})1_D(\boldsymbol{x})$.

How should we geometrically interpret multivariable functions?

Recall that given a single variable function $f : \mathbb{R} \to \mathbb{R}$, we can draw its graph as the set of points

$$\{(x, f(x))\} \subseteq \mathbb{R}^2$$

We can generalize this to graph multivariable functions $f : \mathbb{R}^n \to \mathbb{R}$:

Definition 2.1.11. *Given a function $f : \mathbb{R}^n \to \mathbb{R}$, its* ***graph*** *is the following subset of $\mathbb{R}^{n+1}$:*

$$\Gamma_f := \{(x_1, \cdots, x_n, f(x_1, \cdots, x_n))\} \subset \mathbb{R}^{n+1}$$

In other words, the graph is given by the equation

$$x_{n+1} = f(x_1, \cdots, x_n)$$

in $\mathbb{R}^{n+1}$.

We can easily sketch and visualize graphs of functions with 2 variables:

Definition 2.1.12. *The **graph** of a function of two variables $f(x, y)$ is the surface in $\mathbb{R}^3$ consisting of the set of points (x, y, z) that are solutions to the equation*

$$z = f(x, y)$$

How can we sketch the graph of a function of two variables?

To sketch graphs in $\mathbb{R}^3$, we can use traces!

Definition 2.1.13. *The **trace in the plane** P of a graph $\Gamma \subset \mathbb{R}^3$ is the intersection of Γ with P.*

We typically study x-traces, y-traces, and z-traces - that is, traces in the planes $x = k$, $y = k$, and $z = k$ respectively.

We can think of traces as being slices of the graph. say more

Definition 2.1.14. *The **level curves** (isoclines, contour map) of a function of two variables $f(x, y)$ are the z-traces of the graph $z = f(x, y)$.*

picture

Sketch the graph of $f(x, y) = x^2 + y^2$.

Sketch the graph of $f(x, y) = \sqrt{9 - x^2 - y^2}$.

Implicit functions

From the previous section, we observed that some surfaces, such as the sphere in $\mathbb{R}^3$, are not graphs of multivariable functions. That is, the sphere cannot be described as $z = f(x, y)$ for $f(x, y)$ a function of 2 variables. (As we will see later, they are graphs of multivariable functions **locally**).

Nevertheless, we can still describe them using multivariable functions implicitly!

Definition 2.1.15. *Given a multivariable function $G(x_1, \cdots, x_n) : \mathbb{R}^n \to \mathbb{R}$, its **vanishing locus** is the set of points*

$$\{(x_1, \cdots, x_n) \mid G(x_1, \cdots, x_n) = 0\}$$

The unit sphere in $\mathbb{R}^3$ is the vanishing locus of the function

$$G(x, y, z) = x^2 + y^2 + z^2 - 1$$

All quadric surfaces are the vanishing loci of the general quadratic equation

$$Q(x, y, z) = Ax^2 + By^2 + Cz^2 + Dxy + Exz + Fyz + ax + by + cz + d$$

We can also describe graphs of functions as vanishing loci!

The graph of a multivariable function $x_{n+1} = f(x_1, \cdots, x_n)$ is the vanishing locus of the function

$$G(x_1, \cdots, x_{n+1}) = f(x_1, \cdots, x_n) - x_{n+1}$$

Note that the vanishing locus of $G : \mathbb{R}^n \to \mathbb{R}$ is a subset of $\mathbb{R}^n$, whereas the graph of a function $f : \mathbb{R}^n \to \mathbb{R}$ is a subset of $\mathbb{R}^{n+1}$! Again, this is analogous to the difference between the kernel of a linear map, versus the image of a linear map.

General Multivariable functions

We now to turn to the general case:

Definition 2.1.16. *A* ***general multivariable function*** *is a function $f : \mathbb{R}^n \to \mathbb{R}^m$, defined by*

$$f(x_1, \cdots, x_n) = \big\langle f_1(x_1, \cdots, x_n), f_2(x_1, \cdots, x_n), \cdots, f_m(x_1, \cdots, x_n)\big\rangle$$

That is, a general multivariable function is a vector-valued function, whose components are multi-variable functions!

A vector-valued function is a general multivariable function $f : \mathbb{R} \to \mathbb{R}^m$.

A multivariable function is a general multivariable function $f : \mathbb{R}^n \to \mathbb{R}$.

Exercises

2.1 Consider the vector-valued function $\boldsymbol{r}(t) = \langle t - \sin(t), 1 - \cos(t)\rangle$, which parametrizes a curve called a **cycloid**. Sketch $\boldsymbol{r}(t)$ from 0 to 2π.

2.2 Sketch and describe the vector-valued function $\boldsymbol{r}(t) = \langle e^{-t}\sin(t), e^{-t}\cos(t), e^{-t}\rangle$.

a. How many times does $\boldsymbol{r}(t)$ intersect a sphere of radius 1?

2.3 Sketch and describe the vector-valued function $\boldsymbol{r}(t) = \langle t\sin(t), t\cos(t), t\rangle$.

2.4 Find a parametrization of the intersection of the cylinder $x^2 + y^2 = 1$ and the plane $x + y + z = 1$. Sketch and describe the parametric curve.

2.5 Find a parametrization of the intersection of the cylinder $x^2 + y^2 = 1$ and the parabolic cylinder $z = 4x^2$
Sketch and describe the parametric curve.

2.6 Vistani's curve is the intersection of the surfaces $x^2 + y^2 = z^2$ and $y = z^2$. Find a parametrization of Vistani's curve.

2.7 Show that Vistani's curve lies on the unit sphere of radius 1, centered at $(0, 1, 0)$.

2.8 Show that the vector-valued function $\boldsymbol{r}(t) : \langle t^2 - 1, t - 2t^2, 4 - 6t\rangle$ lies in a plane. What is the equation of the plane?

We say that two lines $\boldsymbol{r_1}(t)$ and $\boldsymbol{r_2}(s)$ **intersect** if there is a point P lying on both curves. We say that $\boldsymbol{r_1}(t)$ and $\boldsymbol{r_2}(s)$ **collide** if there exists some t_0 such that $\boldsymbol{r_1}(t_0) = \boldsymbol{r_2}(t_0)$.

2.9 What is the difference between a point of intersection and a point of collision?

2.10 Consider the vector-valued functions

$$\boldsymbol{r_1}(t) = \langle t^2 - 3, -1, t^2 \rangle \qquad \boldsymbol{r_2}(s) = \langle 2s, 2s+1, 2s+3 \rangle$$

a. Do $\boldsymbol{r_1}(t)$ and $\boldsymbol{r_2}(s)$ **collide**? If so, find the point(s) of collision.

b. Find all values of t and s where the curves of $\boldsymbol{r_1}(t)$ and $\boldsymbol{r_2}(s)$ **intersect**.

2.11 Sketch traces of the surface $x^2 + (\frac{y}{4})^2 + z^2 = 1$.

2.12 Classify and sketch the quadric surface S given by

$$\left(\frac{x}{2}\right)^2 + \left(\frac{y}{5}\right)^2 = 5z^2 - 1$$

2.13 Classify and sketch the quadric surface S given by

$$9y^2 + 4\left(z-1\right)^2 = \frac{9x^2}{4} + 36$$

2.14 Sketch traces of the function $f(x,y) = \frac{y}{x^2}$.

2.15 Sketch traces of the function $f(x,y) = \frac{1}{x^2+y^2+1}$.

2.16 Sketch traces of the function $f(x,y) = \cos(x-y)$.

2.17 Consider the following equations:

$$\begin{array}{ll} \text{(i)} & \left(\frac{y}{2}\right)^2 = \left(\frac{x}{4}\right)^2 + \frac{8}{9} \\ \text{(ii)} & \left(\frac{y}{2}\right)^2 = \left(\frac{x}{4}\right)^2 + 1 \\ \text{(iii)} & \left(\frac{y}{2}\right)^2 + \left(\frac{z-1}{3}\right)^2 = 1 \end{array}$$

a. If a quadric surface S has (i) and (ii) as traces, then can we determine S?

b. If a quadric surface S has (i), (ii), and (iii) as traces, then can we determine S?

c. If (i) is a trace of a quadric surface $9y^2 + 4\left(z-1\right)^2 = \frac{9x^2}{4} + 36$, what plane could the trace be in?

d. If (iii) is a trace of a quadric surface $9y^2 + 4(z-1)^2 + 9(x-1)^2 = 36$, what plane could the trace be in?

2.18 Sketch the level curves of the multivariable function $f(x,y) = |x| + |y|$

2.2 Limits of sequences

Now that we have a good understanding of multivariable functions, let us turn to rigorously studying limits and continuity in multivariable calculus. We will first study limits of sequences (of vectors), and then turn to limits of functions (both vector-valued and multivariable functions).

First, recall the notion of a limit of sequences of real numbers:

Definition 2.2.1. *Let $\{a_k\}$ be a sequence of real numbers. We say that the sequence* ***converges*** *if the following holds:*

For all $\varepsilon > 0$, there exists M such that for all $m > M$,

$$|a_m - L| < \varepsilon$$

*We say L is the **limit** of the sequence $\{a_k\}$. If no such L exists, we say $\{a_k\}$ **diverges**.*

In other words, the limit of a sequence $\{a_k\}$ is L if for every small value ε, we can find an M such that a_m is ε-close to L for $m > M$.

Consider the sequence $\{\frac{1}{k^3+1}\}$. Then $\lim_{k\to\infty} \frac{1}{k^3+1} = 0$.

Proof. Let $\varepsilon > 0$. We wish to find M such that for $m > M$, $|a_m - 0| = a_m < \varepsilon$.
Let us choose $M > (\frac{1}{\varepsilon} - 1)^{\frac{1}{3}}$. Then if $m > M$, we have that

$$m^3 > M^3 > (\frac{1}{\varepsilon} - 1)$$

Thus,

$$m^3 + 1 > \frac{1}{\varepsilon}$$

We can then deduce that for $m > M$,

$$\varepsilon > \frac{1}{m^3+1} = a_m$$

That is, we have shown that $\lim_{k\to\infty} \frac{1}{k^3+1} = 0$. $\square$

Why do we need to use ε when talking about limits?

Ideally, what we would like to say is that the sequence $\{a_k\}$ will eventually reach its limit, L. However, if we consider the sequence

$$\left\{\frac{1}{k^3+1}\right\}$$

We can observe that $\lim_{k\to\infty} \frac{1}{k^3+1} = 0$, but $\frac{1}{k^3+1} > 0$.

Thus, since the sequence $\{a_k\}$ might not ever attain its limit L, we need a different way to describe the limit:

Proposition 2.2.2. *Let $x \in \mathbb{R}_{\geq 0}$. Then the following are equivalent:*

a. *For all $\varepsilon > 0$, $x < \varepsilon$*

b. $x = 0$.

As a consequence, to show that $a = b$, it is equivalent to show that there is no number between a and b!

Consider the sequence $\{\sum_{i=1}^{k} \frac{9}{10^i}\} = \{0.9, 0.99, 0.999, \cdots\}$. Then $\lim_{k\to\infty} a_k = 1$.
In other words, the notion of limits explains why $0.\overline{999} = 1$.

Observe that in our definition, we use the term "**the** limit", which should imply that the limit of a sequence of real numbers is unique. We verify that this is true:

Proposition 2.2.3 (Limits are unique)**.** *Suppose $\{a_k\}$ is a sequence of real numbers that converges to both a and b. Then $a = b$.*

Proof. By proposition 2.2.2, it is equivalent to show that for all $\varepsilon > 0$, $|a-b| < \varepsilon$. Thus, let $\varepsilon > 0$.

Since $\lim_{k\to\infty} a_k = a$, we know that there exists M_a such that for all $m > M_a$, $|a_m - a| < \frac{\varepsilon}{2}$. Similarly, since $\lim_{k\to\infty} a_k = b$, we know that there exists M_b such that for all $m > M_b$, $|a_m - b| < \frac{\varepsilon}{2}$. Set $M = \max\{M_A, M_B\}$

Observe that $|a-b| = |a - a_m + a_m - b|$, and we can apply the triangle inequality to observe that for all $m > M$,

$$\begin{aligned}|a-b| &= |a - a_m + a_m - b| \\ &\le |a - a_m| + |b - a_m| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\end{aligned}$$

Thus, we have proven that for all $\varepsilon > 0$, $|a-b| < \varepsilon$. Therefore, $a = b$, as desired. □

We can now rigorously prove the limit properties that we've seen in single variable calculus.

Theorem 2.2.4 (Properties of Limits in $\mathbb{R}$)**.** *Let $\{a_k\}$ and $\{b_k\}$ be sequences in $\mathbb{R}$ that converge to a and b, respectively. Then*

a. $\lim_{k\to\infty}(a_k + b_k) = a + b$

b. $\lim_{k\to\infty}(ca_k) = ca$ *for all* $c \in \mathbb{R}$.

c. $\lim_{k\to\infty}(a_k b_k) = ab$.

d. *If* $\{b_k\}, b \neq 0$, *then* $\lim_{k\to\infty} \frac{1}{b_k} = \frac{1}{b}$.

e. *If* $\{b_k\}, b \neq 0$, *then* $\lim_{k\to\infty} \frac{a_k}{b_k} = \frac{a}{b}$.

Theorem 2.2.5 (Squeeze theorem for sequences)**.** *Let $\{a_k\}$, $\{b_k\}$, and $\{c_k\}$ be sequences in $\mathbb{R}$ such that there exists an M such that $a_m \le b_m \le c_m$ for all $m > M$.*

Then if $\lim_{k\to\infty} a_k = \lim_{k\to\infty} c_k = L$, then $\lim_{k\to\infty} b_k = L$.

Corollary 2.2.6. *If $\lim_{k\to\infty} |a_k| = 0$, then $\lim_{k\to\infty} a_k = 0$.*

Limits of sequences of vectors

How do we generalize the notion of sequences in $\mathbb{R}$ to sequences of vectors in $\mathbb{R}^k$?

Recall that the idea of a limit of a sequence of real numbers was the following: L is the limit of a sequence $\{a_n\}$ if for every small value ε, we can find an N such that a_n is ε-close to L for $n > N$.

The same idea should hold true for vectors in $\mathbb{R}^n$ - we We say that two points $\boldsymbol{P}, \boldsymbol{Q} \in \mathbb{R}^n$ are close to each other if the distance $||\boldsymbol{P} - \boldsymbol{Q}||$ between them is small!

Definition 2.2.7. *Let $\{\boldsymbol{a_n}\}$ be a sequence of vectors in $\mathbb{R}^k$. We say that the sequence $\{\boldsymbol{a_n}\}$* ***converges*** *to the vector $\boldsymbol{L} \in \mathbb{R}^k$ if:*

For all $\varepsilon > 0$, there exists M such that for all $m > M$,

$$||\boldsymbol{a_m} - \boldsymbol{L}|| < \varepsilon$$

We say $\boldsymbol{L}$ is the limit of the sequence $\{\boldsymbol{a_n}\}$. If no such $\boldsymbol{L}$ exists, we say that $\{\boldsymbol{a_n}\}$ diverges.

Another way to state the definition of limits is through topology:

Definition 2.2.8. *Let $\boldsymbol{P} \in \mathbb{R}^n$. **The open ball of radius** ε **around** $\boldsymbol{P}$, denoted $B_\varepsilon(\boldsymbol{P})$, is the set of points*

$$B_\varepsilon(\boldsymbol{P}) := \{\boldsymbol{x} \in \mathbb{R}^n \mid ||\boldsymbol{x} - \boldsymbol{P}|| < \varepsilon\}$$

That is, $\lim_{n\to\infty} \boldsymbol{a_n} = \boldsymbol{L}$ if for every $\varepsilon > 0$, there is some M such that $\boldsymbol{a_m} \in B_\varepsilon(\boldsymbol{L})$.

picture

Once again, we should confirm that the terminology "**the** limit" is correct. The key idea of the proof is the same as in proposition 2.2.3.

Proposition 2.2.9 (Limits are unique)**.** *Suppose $\{\boldsymbol{a_n}\}$ converges to both $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathbb{R}^n$. Then $\boldsymbol{a} = \boldsymbol{b}$.*

Proof. By proposition 2.2.2, it is equivalent to show that for all $\varepsilon > 0$, $||\boldsymbol{a} - \boldsymbol{b}|| < \varepsilon$. Thus, let $\varepsilon > 0$.

Since $\lim_{k\to\infty} \boldsymbol{a_k} = \boldsymbol{a}$, we know that there exists M_a such that for all $m > M_a$, $||\boldsymbol{a_m} - \boldsymbol{a}|| < \frac{\varepsilon}{2}$. Similarly, since $\lim_{k\to\infty} \boldsymbol{a_k} = \boldsymbol{b}$, we know that there exists M_b such that for all $m > M_b$, $||\boldsymbol{a_m} - \boldsymbol{b}|| < \frac{\varepsilon}{2}$. Set $M = \max\{M_A, M_B\}$

Observe that $||\boldsymbol{a} - \boldsymbol{b}|| = ||\boldsymbol{a} - \boldsymbol{a_m} + \boldsymbol{a_m} - \boldsymbol{b}||$, and we can apply the triangle inequality to observe that for all $m > M$,

$$\begin{aligned} ||\boldsymbol{a} - \boldsymbol{b}|| &= ||\boldsymbol{a} - \boldsymbol{a_m} + \boldsymbol{a_m} - \boldsymbol{b}|| \\ &\leq ||\boldsymbol{a} - \boldsymbol{a_m}|| + ||\boldsymbol{a_m} - \boldsymbol{b}|| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

Thus, we have proven that for all $\varepsilon > 0$, $||\boldsymbol{a} - \boldsymbol{b}|| < \varepsilon$. Therefore, $\boldsymbol{a} = \boldsymbol{b}$, as desired. $\square$

Theorem 2.2.10 (Limits in $\mathbb{R}^m$ are determined componentwise)**.** *Let $\{\boldsymbol{a_n}\}$ be a sequence in $\mathbb{R}^m$, where $\boldsymbol{a_n} = \langle a_{n,1},\ a_{n,2},\ \cdots\ a_{n,m}\rangle$.*

Then $\{\boldsymbol{a_n}\}$ converges to $\boldsymbol{a} = \langle a_1, \cdots, a_m\rangle$ if and only if $\{a_{n,i}\}$ converges to a_i for all $1 \leq i \leq m$.

Proof. The key idea is the following inequalities $|a_{n,i} - a_i| \leq ||\boldsymbol{a_{n,i}} - \boldsymbol{a}|| \leq \sum_{i=1}^m |a_{n,i} - a_i|$. $\square$

Corollary 2.2.11. *Let $\{\boldsymbol{a_n}\}$ and $\{\boldsymbol{b_n}\}$ be sequences in $\mathbb{R}$ that converge to $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. Then*

a. $\lim_{n\to\infty}(\boldsymbol{a_n} + \boldsymbol{b_n}) = \boldsymbol{a} + \boldsymbol{b}$

b. $\lim_{n\to\infty}(c\boldsymbol{a_n}) = c\boldsymbol{a}$ *for all $c \in \mathbb{R}$.*

Divergent sequences

Do limits of sequences always exist?

There are two main types of sequences that diverge:

The sequence $\{(-1)^n\}$ in $\mathbb{R}$ diverges.

The sequence $\{n\}$ in $\mathbb{R}$ diverges.

We will investigate the behavior of both of the sequences.

Definition 2.2.12. *Let $\{\boldsymbol{a_n}\}$ be a sequence of vectors in $\mathbb{R}^k$. A* ***subsequence*** *of $\{\boldsymbol{a_n}\}$ is a sequence $\{\boldsymbol{b_i}\}$, where*

$$\boldsymbol{b_i} = \boldsymbol{a_{n_i}}$$

such that $n_1 < n_2 < \cdots < n_i < \cdots$.

Theorem 2.2.13. *Let $\{\boldsymbol{a_n}\}$ be a sequence of vectors in $\mathbb{R}^k$ such that $\{\boldsymbol{a_n}\}$ has two different subsequences $\{\boldsymbol{a_{n_i}}\}$ and $\{\boldsymbol{a_{n_j}}\}$ that converge to two different limits.*
Then $\{\boldsymbol{a_n}\}$ diverges.

The sequence $\{(-1)^n\}$ has the subsequences

$$\{(-1)^{2n}\} \text{ and } \{(-1)^{2n+1}\}$$

which converge to 1 and -1, hence $\{(-1)^n\}$ diverges.

Theorem 2.2.14 (Convergent sequences are bounded)**.** *Let $\{\boldsymbol{a_n}\}$ be a sequence in $\mathbb{R}^k$ such that $\lim_{n\to\infty} \boldsymbol{a_n} = \boldsymbol{L}$.*

Then the sequence $\{\boldsymbol{a_n}\}$ is ***bounded****. That is, there exists a number $A \in \mathbb{R}$ such that $||\boldsymbol{a_n}|| \leq A$ for all n.*

Proof. picture □

Corollary 2.2.15. *If $\{\boldsymbol{a_n}\}$ be a sequence in $\mathbb{R}^k$ that is unbounded, then $\{\boldsymbol{a_n}\}$ diverges.*

The sequence $\{n\}$ in $\mathbb{R}$ is unbounded, hence it diverges.

Theorem 2.2.16. *Let $\{\boldsymbol{a_n}\}$ be a sequence of vectors in $\mathbb{R}^k$. If $\{\boldsymbol{a_n}\}$ has a subsequence $\{\boldsymbol{a_{n_i}}\}$ that diverges, then $\{\boldsymbol{a_n}\}$ diverges.*

Exercises

2.19 Let $\{\boldsymbol{a_n}\}$ and $\{\boldsymbol{b_n}\}$ be sequences in $\mathbb{R}^n$ that converge to $\boldsymbol{a}$ and $\boldsymbol{b}$, respectively. Show that

a. $\lim_{n\to\infty}(\boldsymbol{a_n} \cdot \boldsymbol{b_n}) = \boldsymbol{a} \cdot \boldsymbol{b}$

b. $\lim_{n\to\infty}(||\boldsymbol{a_n}||) = ||\boldsymbol{a}||$

2.20 Let $\{\boldsymbol{a_n}\}$ be a sequence in $\mathbb{R}^n$ that converges to $\boldsymbol{a}$. Show that

a. $\lim_{n\to\infty}(||\boldsymbol{a_n}||) = ||\boldsymbol{a}||$

2.21 Why don't we have product and quotient rules for limits of two sequences $\{\boldsymbol{a_n}\}$, $\{\boldsymbol{b_n}\}$ in $\mathbb{R}^n$?

2.22 Conjecture and prove product and quotient rules for a sequence $\{\boldsymbol{a_n}\}$ in $\mathbb{R}^n$ and a sequence $\{b_n\}$ in $\mathbb{R}$.

2.23 Does the limit of the sequence $\{\frac{(-1)^n}{n}\}$ exist? Prove your answer.

2.24 Does the limit of the sequence $\{\frac{n}{n+1}\}$ exist? Prove your answer.

2.25 Let $\{\boldsymbol{a_n}\}$ be a sequence of real numbers. The sequence of averages is the sequence $\{b_n\}$ defined by $b_n = \frac{a_1+\cdots+a_n}{n}$.
Prove that if $\lim_{n\to\infty} a_n = 0$, then the sequence of averages $\{b_n\}$ satisfies $\lim_{n\to\infty} b_n = 0$.

Definition 2.2.17. *Let S be a non-empty subset of $\mathbb{R}$. A real number x is called an **upper bound** of S if $x \geq s$ for all $s \in S$.*

Definition 2.2.18. *Let S be a non-empty subset of $\mathbb{R}$. A real number x is called the **least upper bound** (alternatively, **supremum**) of S if*

a. *x is an upper bound for S*

b. *If y is an upper bound for S, then $x \leq y$*

We often write the least upper bound as $\sup(S)$

It turns out that if S is a non-empty subset of $\mathbb{R}$, then if S has an upper bound, then S has a least upper bound.

Definition 2.2.19. *Let S be a non-empty subset of $\mathbb{R}$. A real number x is called a **lower bound** of S if $x \leq s$ for all $s \in S$.*

Definition 2.2.20. *Let S be a non-empty subset of $\mathbb{R}$. A real number x is called the **greatest lower bound** (alternatively, **infimum**) of S if*

a. *x is a lower bound for S*

b. *If y is a lower bound for S, then $x \geq y$*

We often write the greatest lower bound as $\inf(S)$

It turns out that if S is a non-empty subset of $\mathbb{R}$, then if S has a lower bound, then S has a greatest lower bound.

Definition 2.2.21. *A sequence of real numbers $\{a_n\}$ is said to be **increasing** (resp. **decreasing**) if $a_n \leq a_{n+1}$ (resp. $a_n \geq a_{n+1}$) for all n.*

2.26 Prove that if a sequence of real numbers $\{a_n\}$ is increasing and bounded above, then $\lim_{n\to\infty} a_n = \sup(\{a_n\})$

2.27 Prove that if a sequence of real numbers $\{a_n\}$ is decreasing and bounded below, then $\lim_{n\to\infty} a_n = \inf(\{a_n\})$

2.28 Prove that if a sequence of real numbers $\{a_n\}$ is decreasing and bounded below, then $\lim_{n\to\infty} a_n = \inf(\{a_n\})$

2.29 Prove the following theorem:
Let $\{a_n\}$ be an increasing or a decreasing sequence. Then $\{a_n\}$ converges if and only if $\{a_n\}$ is bounded.

2.3 Limits of functions

Now let us turn to limits of functions! Let us first recall the notion of limits of single-variable functions:

Definition 2.3.1. *A function $f : \mathbb{R} \to \mathbb{R}$ has the **limit** b at a if*

For all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $\boldsymbol{x} \in \mathbb{R}$,

$$0 < |x - a| < \delta \qquad \text{implies} \qquad |f(x) - b| < \varepsilon$$

We write $\lim_{x \to a} f(x) = b$.

Since $\mathbb{R}$ is a well-ordered field (in other words, since), we can turn this into a statement about left limits and right limits.

left and right picture

Definition 2.3.2 (Limit of a general multivariable function). *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ has the **limit $\boldsymbol{b}$** at $\boldsymbol{a}$ if*

For all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $\boldsymbol{x} \in \mathbb{R}^n$,

$$0 < ||\boldsymbol{x} - \boldsymbol{a}|| < \delta \qquad \text{implies} \qquad ||f(\boldsymbol{x}) - \boldsymbol{b}|| < \varepsilon$$

We write $\lim_{\boldsymbol{x} \to \boldsymbol{a}} f(\boldsymbol{x}) = \boldsymbol{b}$.

Proposition 2.3.3 (Limits of functions are unique). *Let $f : \mathbb{R}^n \to \mathbb{R}^m$, and suppose $\lim_{\boldsymbol{x} \to \boldsymbol{a}} f(x) = \boldsymbol{b_1}$ and $\lim_{\boldsymbol{x} \to \boldsymbol{a}} f(x) = \boldsymbol{b_2}$. Then $\boldsymbol{b_1} = \boldsymbol{b_2}$.*

We will study this for the two different classes of functions (vector-valued functions $\mathbb{R} \to \mathbb{R}^n$ and multivariable functions $\mathbb{R}^n \to \mathbb{R}$) we've seen. This turns out to be enough, as we see in section

Limits of vector valued functions

Definition 2.3.4 (Limit of a vector-valued function). *A function $f : \mathbb{R} \to \mathbb{R}^m$ has the **limit $\boldsymbol{b}$** at a if*

For all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $x \in \mathbb{R}$,

$$0 < |x - a| < \delta \qquad \text{implies} \qquad ||f(x) - \boldsymbol{b}|| < \varepsilon$$

Theorem 2.3.5 (Limits of vector-valued functions are determined componentwise). *For any vector valued function $\boldsymbol{r}(t) = \langle x_1(t), \cdots, x_n(t) \rangle$,*

$$\lim_{t \to a} \boldsymbol{r}(t) = \langle x_1, \cdots, x_n \rangle$$

if and only if for all $1 \le i \le n$,

$$\lim_{t \to a} x_i(t) = x_i$$

Proof. The proof is essentially the same as the proof of theorem 2.2.10:
The key idea is again the following inequalities $|x_i(t) - x_i| \le ||\boldsymbol{r}(t) - \langle x_1, \cdots, x_n \rangle|| \le \sum_{i=1}^m |x_i(t) - x_i|$. $\square$

picture

Theorem 2.3.6. *Let $\boldsymbol{r}, \boldsymbol{s} : \mathbb{R} \to \mathbb{R}^m$ be vector-valued functions. Suppose that $\lim_{t \to a} \boldsymbol{r}(t)$ and $\lim_{t \to a} \boldsymbol{s}(t)$ exist. Then*

a. ***Sum Law:***

$$\lim_{t\to a}(\boldsymbol{r}(t)+\boldsymbol{s}(t)) = \lim_{t\to a}\boldsymbol{r}(t) + \lim_{t\to a}\boldsymbol{s}(t)$$

b. ***Scalar Multiple Law:***

$$\lim_{t\to a}\lambda(\boldsymbol{r}(t)) = \lambda\left(\lim_{t\to a}\boldsymbol{r}(t)\right)$$

Limits of multivariable functions

Definition 2.3.7 (Limit of a multivariable function)**.** *A function $f:\mathbb{R}^n\to\mathbb{R}$ has the* ***limit*** b *at* $\boldsymbol{a}$ *if*

For all $\varepsilon>0$, there exists $\delta>0$ such that for all $\boldsymbol{x}\in\mathbb{R}^n$,

$$0<||\boldsymbol{x}-\boldsymbol{a}||<\delta \qquad \text{implies} \qquad |f(\boldsymbol{x})-b|<\varepsilon$$

Theorem 2.3.8. *Let $f,g:\mathbb{R}^n\to\mathbb{R}$ be functions of n variables. Suppose that $\lim_{\boldsymbol{x}\to\boldsymbol{P}} f(\boldsymbol{x})$ and $\lim_{\boldsymbol{x}\to\boldsymbol{P}} g(\boldsymbol{x})$ exist. Then*

a. ***Sum Law:***

$$\lim_{\boldsymbol{x}\to\boldsymbol{P}}(f(\boldsymbol{x})+g(\boldsymbol{x})) = \lim_{\boldsymbol{x}\to\boldsymbol{P}} f(\boldsymbol{x}) + \lim_{\boldsymbol{x}\to\boldsymbol{P}} g(\boldsymbol{x})$$

b. ***Scalar Multiple Law:***

$$\lim_{\boldsymbol{x}\to\boldsymbol{P}}\lambda(f(\boldsymbol{x})) = \lambda\left(\lim_{\boldsymbol{x}\to\boldsymbol{P}} f(\boldsymbol{x})\right)$$

c. ***Product Law:***

$$\lim_{\boldsymbol{x}\to\boldsymbol{P}}(f(\boldsymbol{x})g(\boldsymbol{x})) = \left(\lim_{\boldsymbol{x}\to\boldsymbol{P}} f(\boldsymbol{x})\right)\left(\lim_{\boldsymbol{x}\to\boldsymbol{P}} g(\boldsymbol{x})\right)$$

d. ***Quotient Law:*** *If* $\lim_{\boldsymbol{x}\to\boldsymbol{P}} g(\boldsymbol{x})\neq 0$,

$$\lim_{\boldsymbol{x}\to\boldsymbol{P}}\frac{f(\boldsymbol{x})}{g(\boldsymbol{x})} = \frac{\lim_{\boldsymbol{x}\to\boldsymbol{P}} f(\boldsymbol{x})}{\lim_{\boldsymbol{x}\to\boldsymbol{P}} g(\boldsymbol{x})}$$

This theorem allows us to reduce to studying limits of single-variable functions.

Let $f(x,y)=x$. Then

$$\lim_{(x,y)\to(a,b)} f(x,y) = a$$

Proof. Let $\varepsilon>0$. We wish to show that there exists $\delta>0$ such that if $0<||(x,y)-(a,b)||<\delta$, then $|x-a|<\varepsilon$.

Observe that

$$|x-a| = \sqrt{(x-a)^2} \le \sqrt{(x-a)^2+(y-b)^2} = ||(x,y)-(a,b)||$$

Therefore, choosing $\delta=\varepsilon$, we see that if $||(x,y)-(a,b)||<\varepsilon$, then $|x-a|<\varepsilon$. $\square$

Let $f(x,y)=\frac{x\sin(y)}{e^x+1}$. Then

$$\lim_{(x,y)\to(a,b)} f(x,y) = \frac{a\sin(b)}{e^a+1}$$

Observe that theorem 2.3.8 allows us to evaluate the limits, except when we possibly divide by zero.

When do limits of multivariable functions not exist?

Definition 2.3.9. *Let $X \subset \mathbb{R}^n$. We say that a point $\boldsymbol{p} \in \mathbb{R}^n$ is a **limit point of** X if there is a sequence $\{\boldsymbol{a_n}\}$ contained inside X such that $\{\boldsymbol{a_n}\}$ converges to $\boldsymbol{p}$.*

Theorem 2.3.10. *Let $X \subset \mathbb{R}^n$, $f : X \to \mathbb{R}^m$ a function, and $\boldsymbol{a}$ a limit point of X. Then the following statements are equivalent:*

a. $\lim_{\boldsymbol{x}\to\boldsymbol{a}} f(x) = \boldsymbol{b}$

b. *For every sequence $\{\boldsymbol{a_n}\}$ converging to $\boldsymbol{a}$ (with $\boldsymbol{a_n} \neq \boldsymbol{a}$), the sequence $\{f(\boldsymbol{a_n})\}$ converges to $\boldsymbol{b}$.*

In other words, in order for a limit of a multivariable function to exist, it must yield the same value **along all possible approaches**.

Show that $\lim_{(x,y)\to(0,0)} \frac{x^2}{x^2+y^2}$ does not exist.
Consider the paths $\boldsymbol{r_1}(t) = \langle 0, t\rangle$, and $\boldsymbol{r_2}(t) = \langle t, t\rangle$.

Theorem 2.3.11 (Squeeze Theorem)**.** *Let $f(\boldsymbol{x})$, $g(\boldsymbol{x})$, and $h(\boldsymbol{x})$ be functions of n variables such that*

$$\lim_{\boldsymbol{x}\to\boldsymbol{P}} f(\boldsymbol{x}) = L = \lim_{\boldsymbol{x}\to\boldsymbol{P}} h(\boldsymbol{x})$$

If there exists $\delta > 0$ such that for all $\boldsymbol{x} \in B_\delta(\boldsymbol{P}) - \{\boldsymbol{P}\}$, we have that

$$f(\boldsymbol{x}) \leq g(\boldsymbol{x}) \leq h(\boldsymbol{x})$$

Then $\lim_{\boldsymbol{x}\to\boldsymbol{P}} g(\boldsymbol{x}) = L$.

Determine if the following limit exists:

$$\lim_{(x,y)\to(0,0)} x\sin\left(\frac{1}{x^2+y^2}\right)$$

Limits of general multivariable functions

Now that we have studied limits of both vector-valued functions $\mathbb{R} \to \mathbb{R}^n$ and multivariable functions $\mathbb{R}^n \to \mathbb{R}$, this turns out to capture the behavior of limits of general multivariable functions $\mathbb{R}^n \to \mathbb{R}^m$!

The key idea is the following theorem:

Theorem 2.3.12 (Limits of general multivariable functions are determined componentwise)**.** *For any general multivariable function $f : \mathbb{R}^n \to \mathbb{R}^m$, defined by*

$$f(x_1, \cdots, x_n) = \big\langle f_1(x_1, \cdots, x_n), f_2(x_1, \cdots, x_n), \cdots, f_m(x_1, \cdots, x_n)\big\rangle$$

Then

$$\lim_{\boldsymbol{x}\to\boldsymbol{P}} f(\boldsymbol{x}) = \langle b_1, \cdots, b_n\rangle$$

if and only if for all $1 \leq i \leq m$,

$$\lim_{\boldsymbol{x}\to\boldsymbol{P}} f_i(\boldsymbol{x}) = b_i$$

Proof. The proof idea is exactly the same as theorem 2.3.5. $\square$

Thus, it is indeed enough to study limits of multivariable functions $\mathbb{R}^n \to \mathbb{R}$.

Exercises

2.30 Show that $\lim_{(x,y)\to(0,0)} x\sin(\frac{1}{x^2+y^2}) = 0$.

2.31 Show that $\lim_{(x,y)\to(0,0)} \frac{y^2}{x^2+y^2}$ does not exist.

2.32 Use the squeeze theorem to compute $\lim_{(x,y)\to(0,0)} \frac{5xy^2}{x^2+y^2}$

2.33 Show that the following limit does not exist.

$$\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2+3y^4}$$

2.34 $\lim_{(x,y)\to(1,2)} x^2+y$ exist? Prove or disprove your claim.

2.35 Does $\lim_{(x,y)\to(0,0)} \frac{xy}{x^2+y^2}$ exist? Prove or disprove your claim.

2.36 Does $\lim_{(x,y,z)\to(0,0,0)} \frac{x+y+z}{x^2+y^2+z^2}$ exist? Prove or disprove your claim.

2.37 Does $\lim_{(x,y,z)\to(0,0,0)} \frac{xyz}{x^3+y^3+z^3}$ exist? Prove or disprove your claim.

2.38 Consider the function $f(x,y) = \begin{cases} \cos(x) & \text{if } x^2+y^2 = \frac{1}{n} \text{ for some } n = 1,2,3,\ldots \\ 0 & \text{otherwise} \end{cases}$
Compute $\lim_{(x,y)\to(0,0)} f(x,y)$ or show that it does not exist.

2.4 Continuity

Let us now turn to continuity. Let us first recall the notion of continuity fora single-variable functions:

Definition 2.4.1. *A function $f : \mathbb{R} \to \mathbb{R}$ is continuous at $a \in \mathbb{R}$ if $\lim_{x\to a} f(x) = f(a)$.*

That is, the limit of the function at a is precisely the value of the function at a.
picture
The idea for general multivariable functions is identical:

Definition 2.4.2. *A general multivariable function $f : \mathbb{R}^m \to \mathbb{R}^n$ is **continuous** at a point $\boldsymbol{P} \in \mathbb{R}^m$ if*

$$\lim_{\boldsymbol{x}\to\boldsymbol{P}} f(\boldsymbol{x}) = f(\boldsymbol{P})$$

f is said to be continuous on a subset $X \subset \mathbb{R}^m$ if it is continuous for every $\boldsymbol{P} \in X$.

That is, the limit of the function at $\boldsymbol{P}$ is precisely the value of the function at $\boldsymbol{P}$.

Thus, questions about continuity are really questions about limits: To show that a function is continuous, you need to show that the limit exists, and then check that the limit matches the value of the function.

Thus, all of our theorems about limits immediately yield theorems about continuity.

Theorem 2.4.3. *Suppose $f : \mathbb{R}^m \to \mathbb{R}^n$ and $g : \mathbb{R}^m \to \mathbb{R}^n$ are general multivariable functions that are **continuous** at a point $\boldsymbol{P}$. Then*

a. ***Sum Law:*** *$f + g$ is continuous at $\boldsymbol{P}$.*

b. ***Scalar Multiple Law:*** *λf is continuous at $\boldsymbol{P}$.*

Vector-valued functions

Proposition 2.4.4. *For any vector valued function*

$$\boldsymbol{r}(t) = \langle x_1(t), \cdots, x_n(t) \rangle$$

$\boldsymbol{r}(t)$ is continuous at P if and only if for all $1 \leq i \leq n$, $x_i(t)$ is continuous at P.

Multivariable functions

Theorem 2.4.5 (Properties of continuity of multivariable functions)**.** *Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be functions of n variables. Suppose that $f(\boldsymbol{x})$ and $g(\boldsymbol{x})$ are continuous at $\boldsymbol{P}$.*

a. ***Product Law:*** *$f(\boldsymbol{x})g(\boldsymbol{x})$ is continuous at $\boldsymbol{P}$.*

b. ***Quotient Law:*** *If $g(\boldsymbol{P}) \neq 0$, $\frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}$ is continuous at $\boldsymbol{P}$.*

The function $f : \mathbb{R}^2 \to \mathbb{R}$ given below is continuous for all $(x, y) \in \mathbb{R}^2$.

$$f(x,y) = \begin{cases} \frac{xy^2}{x^2+y^2} & \text{if } (x,y) \neq \mathbf{0} \\ 0 & \text{if } (x,y) = \mathbf{0} \end{cases}$$

Proof. By the quotient law, f is continuous for all $(x, y) \neq 0$. We thus only need to show that $\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2+y^2} = 0$

We will show this directly, but we could have also used the squeeze theorem, or the polar coordinates method.

Let $\varepsilon > 0$. We need to find $\delta > 0$ such that if $0 < ||\langle x, y\rangle|| < \delta$, then

$$|\frac{xy^2}{x^2+y^2} - 0| < \varepsilon$$

Observe that since $x^2 > 0$, then

$$|\frac{xy^2}{x^2+y^2}| \leq |\frac{xy^2}{y^2}| = |x|$$

Thus, if we choose $\delta = \varepsilon$, then we know that $|x| \leq ||\langle x, y\rangle|| < \varepsilon$. Hence $|\frac{xy^2}{x^2+y^2}| \leq |\frac{xy^2}{y^2}| = |x| < \varepsilon$, as desired. □

Theorem 2.4.6 (The composition of continuous functions is continuous.)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$, and let $g : \mathbb{R}^m \to \mathbb{R}^k$ be multivariable functions such that f is continuous at $\boldsymbol{P} \in \mathbb{R}^n$, and g is continuous at $f(\boldsymbol{P}) \in \mathbb{R}^m$.*

Then $g \circ f : \mathbb{R}^n \to \mathbb{R}^k$ is continuous at at $\boldsymbol{P} \in \mathbb{R}^n$.

Proof. do proof □

Theorem 2.4.7. $\det : M_{n\times n}(\mathbb{R}) \cong \mathbb{R}^{n^2} \to \mathbb{R}$ *is continuous.*

Corollary 2.4.8. *The cross product $\times : \mathbb{R}^3 \to \mathbb{R}^3 \to \mathbb{R}^3$ is continuous.*

Theorem 2.4.9. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. Then T is continuous (in fact, T is uniformly continuous).*

Definition 2.4.10. *A multivariable function $f : \mathbb{R}^n \to \mathbb{R}^m$ is **uniformly continuous** if*

For all $\varepsilon > 0$, there exists $\delta > 0$ such that for all $\boldsymbol{a}, \boldsymbol{b} \in X$,

$$0 < ||\boldsymbol{a} - \boldsymbol{b}|| < \delta \qquad \text{implies} \qquad ||f(\boldsymbol{a}) - f(\boldsymbol{b})|| < \varepsilon$$

Exercises

2.39 Consider the function $f(x,y) = \begin{cases} y^2 \sin(\frac{1}{(x-4)^2+y^2}) & \text{if } (x,y) \neq (4,0) \\ c & \text{otherwise} \end{cases}$

Either show that there is a c that will make f continuous at $(4,0)$, or show that f will not be continuous for any choice of c.

2.40 Consider the function $f(x,y) = \begin{cases} \frac{x^2+\sin(y)}{2x^2+y+e^{xy}-1} & \text{if } (x,y) \neq (0,0) \\ 1 & \text{otherwise} \end{cases}$

Is f continuous at $(0,0)$?

2.41 Consider the function $f(x,y) = \begin{cases} \frac{x^2+\sin(xy)+y^2}{2x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ c & \text{otherwise} \end{cases}$

Either show that there is a c that will make f continuous, or show that f will not be continuous for any choice of c.

2.42 Consider the function $f(x,y) = \begin{cases} \frac{y^5}{x^4+y^4} & \text{if } (x,y) \neq (0,0) \\ 1 & \text{otherwise} \end{cases}$

Is f continuous at $(0,0)$?

2.43 Consider the function $g(x,y) = \begin{cases} x^2 \cos(\frac{1}{x^2+y^2}) & \text{if } (x,y) \neq (0,0) \\ c & \text{otherwise} \end{cases}$

Either show that there is a c that will make g continuous, or show that g will not be continuous for any choice of c.

2.44 Consider the function $f(x,y) = \begin{cases} \frac{x^2}{x^2+|y|^3} & \text{if } (x,y) \neq (0,0) \\ c & \text{otherwise} \end{cases}$

Either show that there is a c that will make f continuous at $(0,0)$, or show that f will not be continuous for any choice of c.

2.45 Consider the function $f(x,y) = \begin{cases} x \sin(\frac{1}{x^2+y^2}) & \text{if } (x,y) \neq (0,0) \\ c & \text{otherwise} \end{cases}$

Either show that there is a c that will make f continuous at $(0,0)$, or show that f will not be continuous for any choice of c.

2.46 Consider the function $g(x,y) = \begin{cases} \frac{x^4+2x^2y^2+y^4}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ c & \text{otherwise} \end{cases}$

Either show that there is a c that will make g continuous at $(0,0)$, or show that g will not be continuous for any choice of c.

Definition 2.4.11. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be **Lipschitz** it there exists a constant $K \geq 0$ such that for all $\boldsymbol{x_1}, \boldsymbol{x_2} \in \mathbb{R}^n$,*

$$||f(\boldsymbol{x_1}) - f(\boldsymbol{x_2})|| \leq K||\boldsymbol{x_1} - \boldsymbol{x_2}||$$

2.47 Prove that if a function $f : \mathbb{R}^m \to \mathbb{R}^n$ is Lipschitz, then f is continuous.

Chapter 3

The Multivariable Derivative

Now that we have understood the notions of limits and continuity, we can turn to defining the notion of the multivariable derivative.

Let us first recall the notion of the derivative of a single-variable function:

Definition 3.0.1. *A function $f : D \subset \mathbb{R} \to \mathbb{R}$ is* ***differentiable*** *at a point $x_0 \in D$ if*

- *There exists δ such that $B_\delta(x_0) \subseteq D$*
- *The following limit exists:*

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h} = L$$

If f is differentiable at x_0, we say that L is ***the derivative of f at*** *x_0, and we write*

$$f'(x_0) := L$$

(sometimes written $\frac{df}{dx}(x_0) = L$)

The notion of the derivative in single-variable calculus has many interpretations:

The derivative $f'(a)$ tells us the slope of the tangent line to the graph $y = f(x)$ at the point $x = a$.

The derivative $f'(a)$ tells us the instantaneous rate of change of $f(x)$ at the point $x = a$.

The derivative $\frac{d}{dx}$ is an operator, whose input is a differentiable function $f : \mathbb{R} \to \mathbb{R}$, and whose output is a function $f' : \mathbb{R} \to \mathbb{R}$.

The derivative depends precisely on the local behavior of $f(x)$ near x_0. That is, we look precisely at the behavior of $f(x)$ on $B_\varepsilon(x_0)$.

3.1 The Multivariable Derivative

Bad definition:

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{f(\boldsymbol{x_0} + \boldsymbol{h}) - f(\boldsymbol{x_0})}{||\boldsymbol{h}||}$$

Bad example 1: $f(x, y) = x$. Limit does not exist taking the sequences $(\frac{1}{n}, 0)$, $(-\frac{1}{n}, 0)$, and $(0, \frac{1}{n})$

Bad example 2: Set $k = 1$. Does this agree with the usual definition of the limit?

$$\lim_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{|h|}$$

consider the function $f(x) = |x|$

Paradigm shift:

The derivative of a single variable function $f(x)$ at a point x_0 is the **approximation** of $f(x)$ by a linear transformation at x_0.

Definition 3.1.1. *A single variable function is* ***differentiable at*** x_0 *if there exists a linear transformation* $T : \mathbb{R} \to \mathbb{R}$ *such that*

$$\lim_{h\to 0}\left|\frac{f(x_0+h)-f(x_0)-T(h)}{h}\right| = 0$$

Observe that by our characterization of linear transformations $T : \mathbb{R} \to \mathbb{R}$,

$$T(x) = mx$$

for some m. *We define the derivative of* f *at* x_0 *to be* m. *That is,*

$$f'(x_0) := m$$

Good definition:

Definition 3.1.2. *A multivariable function* $f : A \subset R^m \to \mathbb{R}^n$ *is* ***differentiable at*** *an interior point* $\boldsymbol{x_0}$ *of* A *if there exists a linear transformation* $T : \mathbb{R}^m \to \mathbb{R}^n$ *such that*

$$\lim_{\boldsymbol{h}\to \boldsymbol{0}}\frac{||f(\boldsymbol{x_0}+\boldsymbol{h})-f(\boldsymbol{x_0})-T(\boldsymbol{h})||}{||\boldsymbol{h}||} = 0$$

The ***derivative*** *of* f *at* x_0 *is the linear transformation*

$$Df(\boldsymbol{x_0}) := T$$

Remark 3.1.3. *The derivative of a function* $f : \mathbb{R}^m \to \mathbb{R}^n$ *at a point* x_0 *is the* ***approximation*** *of* f *by a* ***linear transformation*** *at* x_0.

Proposition 3.1.4 (Uniqueness)**.** *Suppose* $f : A \subset \mathbb{R}^m \to \mathbb{R}^n$ *is differentiable at* $\boldsymbol{x_0}$.

Then the derivative $Df(\boldsymbol{x_0}) : \mathbb{R}^m \to \mathbb{R}^n$ *is unique.*

The constant map $c : \mathbb{R}^m \to \mathbb{R}^n$ defined by $c(\boldsymbol{x}) = \boldsymbol{a}$ is differentiable everywhere, and $Dc(\boldsymbol{x_0}) = 0$.

A linear map $T : \mathbb{R}^m \to \mathbb{R}^n$ is differentiable everywhere, and $DT(\boldsymbol{x_0}) = T$.

Proposition 3.1.5. *Suppose that* $f, g : A \to \mathbb{R}^n$ *are differentiable at* $x_0 \in A^\circ$. *Then* $f+g$ *and* λf *are differentiable at* $x_0 \in A^\circ$.

Proposition 3.1.6 (Differentiability implies continuity)**.** *Suppose* $f : A \subset \mathbb{R}^m \to \mathbb{R}^n$ *is differentiable at* $\boldsymbol{x_0}$.

Then f *is continuous at* $\boldsymbol{x_0}$.

Theorem 3.1.7 (The Chain rule)**.** *Suppose that* $g : \mathbb{R}^k \to \mathbb{R}^m$ *is differentiable at* $x_0 \in \mathbb{R}^k$, *and* $f : \mathbb{R}^m \to \mathbb{R}^n$ *is differentiable at* $g(\boldsymbol{x_0}) \in \mathbb{R}^m$.

Then $f \circ g : \mathbb{R}^k \to \mathbb{R}^n$ *is differentiable at* $x_0 \in \mathbb{R}^k$, *and*

$$D(f\circ g)(\boldsymbol{x_0}) = D(f)(g(\boldsymbol{x_0})) \circ Dg(\boldsymbol{x_0})$$

The derivative of a vector valued function

Theorem 3.1.8. *Let $\boldsymbol{r}(t) = \langle x_1(t), \cdots, x_n(t)\rangle$ be a vector-valued function.*

*The **derivative** of $\boldsymbol{r}(t) : \mathbb{R} \to \mathbb{R}^n$ at an interior point $\mathbf{t_0}$ is the linear transformation $T : \mathbb{R} \to \mathbb{R}^n$ given by*

$$T = \begin{bmatrix} x_1'(t_0) \\ \vdots \\ x_n'(t_0) \end{bmatrix}$$

We sometimes write this derivative as $\boldsymbol{r}'(t) : \mathbb{R} \to \mathbb{R}^n$.

Theorem 3.1.9. *A vector-valued function $\boldsymbol{r}(t) = \langle x_1(t), \cdots, x_n(t)\rangle$ is differentiable if and only if the functions $x_i(t)$ are differentiable.*

$$\boldsymbol{r}'(t) = \langle x_1'(t), \cdots, x_n'(t)\rangle$$

We can think of the derivative $\boldsymbol{r}'(t)$ as a tangent vector to the parametric curve of $\boldsymbol{r}(t)$.

Definition 3.1.10. *The **tangent line** at $\boldsymbol{r}(t_0)$ is the line determined by the direction vector $\boldsymbol{r}'(t_0)$ and the point $\boldsymbol{r}(t_0)$. That is, the line can be parametrized as*

$$\boldsymbol{L}(t) = \boldsymbol{r}(t_0) + t\boldsymbol{r}'(t_0)$$

picture

Theorem 3.1.11 (Differentiation Rules)**.** *Assume that $\boldsymbol{r}(t)$ and $\boldsymbol{s}(t)$ are differentiable vector-valued functions $\mathbb{R} \to \mathbb{R}^n$..*

a. $\frac{d}{dt}(\boldsymbol{r}(t) + \boldsymbol{s}(t)) = \boldsymbol{r}'(t) + \boldsymbol{s}'(t)$

b. ***Scalar product rule****. For any differentiable scalar-valued function $f(t)$,*

$$\frac{d}{dt}(f(t)\boldsymbol{r}(t)) = f(t)\boldsymbol{r}'(t) + f'(t)\boldsymbol{r}(t)$$

c. ***Chain rule****. For any differentiable scalar-valued function $f(t)$,*

$$\frac{d}{dt}(\boldsymbol{r}(f(t))) = \boldsymbol{r}'(f(t))f'(t)$$

Theorem 3.1.12 (Dot product rule)**.** *Assume that $\boldsymbol{r}(t)$ and $\boldsymbol{s}(t)$ are differentiable vector-valued functions $\mathbb{R} \to \mathbb{R}^n$. Then*

$$\frac{d}{dt}(\boldsymbol{r}(t) \cdot \boldsymbol{s}(t)) = \boldsymbol{r}'(t) \cdot \boldsymbol{s}(t) + \boldsymbol{r}(t) \cdot \boldsymbol{s}'(t)$$

Theorem 3.1.13 (Cross product rule)**.** *Assume that $\boldsymbol{r}(t)$ and $\boldsymbol{s}(t)$ are differentiable vector-valued functions $\mathbb{R} \to \mathbb{R}^3$. Then*

$$\frac{d}{dt}(\boldsymbol{r}(t) \times \boldsymbol{s}(t)) = \boldsymbol{r}'(t) \times \boldsymbol{s}(t) + \boldsymbol{r}(t) \times \boldsymbol{s}'(t)$$

The derivative of multivariable functions

How should we think of the derivative of a multivariable function $f : \mathbb{R}^n \to \mathbb{R}$?

Given a function $g : \mathbb{R}^n \to \mathbb{R}$, we can graph it as a surface $x_{n+1} = g(x_1, \cdots, x_n)$ in $\mathbb{R}^{n+1}$, which has points

$$(x_1, \cdots, x_n, g(x_1, \cdots, x_n))$$

If T is the derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$, then $T : \mathbb{R}^n \to \mathbb{R}$ is a linear map, which corresponds to a $n \times 1$ matrix.

$$T(x_1, \cdots, x_n) = A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_i^n a_i x_i$$

Proposition 3.1.14. *The graph of $T : \mathbb{R}^n \to \mathbb{R}$ is the hyperplane $x_{n+1} = \sum_i^n a_i x_i$.*

Definition 3.1.15. *The **i-th partial derivative** $D_i f$ of a multivariable function $f : A \subset \mathbb{R}^m \to \mathbb{R}$ for $\boldsymbol{x_0} \in A^\circ$ is defined as the limit*

$$D_i f(\boldsymbol{x_0}) = \lim_{t \to 0} \frac{f(\boldsymbol{x_0} + t\boldsymbol{e_i}) - f(\boldsymbol{x_0})}{t}, \qquad i = 1, \cdots, n$$

if it exists.

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = xe^{xy}$. Then the partial derivatives of f are

$$\frac{\partial f}{\partial x} = xye^{xy} + e^{xy} \qquad \frac{\partial f}{\partial y} = x^2 e^{xy}$$

Proposition 3.1.16. *Given a graph of $z = f(x_1, \cdots, x_n)$, the tangent (hyperplane) in $\mathbb{R}^{n+1}$ is spanned by vectors of the form $\boldsymbol{e_i} + \frac{\partial f}{\partial x_i} \boldsymbol{e_{n+1}}$*

picture

Proposition 3.1.17. *We can take the normal vector of the tangent hyperplane in $\mathbb{R}^{n+1}$ to the graph of a multivariable function $f : \mathbb{R}^n \to \mathbb{R}$ to be*

$$\boldsymbol{n} = \sum_i^n \left(\frac{\partial f}{\partial x_i} \boldsymbol{e_i} \right) - \boldsymbol{e_{n+1}} = \left\langle \frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}, -1 \right\rangle$$

Theorem 3.1.18. *Let $\boldsymbol{a} = \langle a_1, \cdots, a_n \rangle \in \mathbb{R}^n$. The equation of the tangent hyperplane in $\mathbb{R}^{n+1}$ to the graph of a multivariable function $f(\boldsymbol{a}) : \mathbb{R}^n \to \mathbb{R}$ at a point*

$$\overline{\boldsymbol{a}} = (a_1, \cdots, a_n, f(\boldsymbol{a}))$$

is given by

$$\left\langle \frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}, -1 \right\rangle \cdot (\boldsymbol{x} - \overline{\boldsymbol{a}}) = 0$$

Equivalently,

$$x_{n+1} = f(\boldsymbol{a}) + \begin{bmatrix} D_1 f(\boldsymbol{a}) & D_2 f(\boldsymbol{a}) & \cdots & D_n f(\boldsymbol{a}) \end{bmatrix} (\boldsymbol{x} - \boldsymbol{a})$$

We can use the multivariable derivative to approximate multivariable functions!

Theorem 3.1.19 (Linear approximation)**.** *If $f : A \subset \mathbb{R}^m \to \mathbb{R}$ is differentiable at a point $\boldsymbol{a} = (a_1, \cdots, a_n)$, and $\boldsymbol{x} = (x_1, \cdots, x_n)$ is close to $\boldsymbol{a}$, then*

$$\begin{aligned} f(\boldsymbol{x}) &\approx f(\boldsymbol{a}) + \begin{bmatrix} D_1 f(\boldsymbol{a}) & D_2 f(\boldsymbol{a}) & \cdots & D_n f(\boldsymbol{a}) \end{bmatrix} (\boldsymbol{x} - \boldsymbol{a}) \\ &= f(\boldsymbol{a}) + \sum_i^n \left(\frac{\partial f}{\partial x_i}(\boldsymbol{a}) \right) (x_i - a_i) \end{aligned}$$

Equivalently, one can say that the change in f near $\boldsymbol{a}$ can be approximated by the partial derivatives and the change in x_i.

$$\Delta f \approx \sum_i^n \left(\frac{\partial f}{\partial x_i}(\boldsymbol{a}) \right) \Delta x_i$$

The multivariable derivative in coordinates

How do we calculate the linear transformation $Df(\boldsymbol{x_0})$ for general multivariable functions $f : A \subset \mathbb{R}^m \to \mathbb{R}^n$?

Definition 3.1.20. *Let $f : A \subset \mathbb{R}^m \to \mathbb{R}^n$ be a multivariable function defined by $f_i : A \subset \mathbb{R}^m \to \mathbb{R}$:*

$$f(\boldsymbol{x}) = \begin{bmatrix} f^1(\boldsymbol{x}) \\ \vdots \\ f^n(\boldsymbol{x}) \end{bmatrix}$$

*The **Jacobian matrix** of f at $\boldsymbol{x_0}$ is*

$$[J_f(\boldsymbol{x_0})] = \begin{bmatrix} D_1 f^1(\boldsymbol{x_0}) & D_2 f^1(\boldsymbol{x_0}) & \cdots & D_m f^1(\boldsymbol{x_0}) \\ D_1 f^2(\boldsymbol{x_0}) & D_2 f^2(\boldsymbol{x_0}) & \cdots & D_m f^2(\boldsymbol{x_0}) \\ \vdots & \vdots & \vdots & \vdots \\ D_1 f^n(\boldsymbol{x_0}) & D_2 f^n(\boldsymbol{x_0}) & \cdots & D_m f^n \boldsymbol{x_0}) \end{bmatrix}$$

if the partial derivatives exist.

Theorem 3.1.21 (The derivative in coordinates)**.** *Let $f : A \subset \mathbb{R}^m \to \mathbb{R}^n$ be a multivariable function. If f is **differentiable at $\boldsymbol{x_0}$**, then all the partial derivatives $D_i f^j \boldsymbol{x_0})$ exist, and the standard matrix of $Df(\boldsymbol{x_0})$ is $[J_f(\boldsymbol{x_0})]$. That is,*

$$Df(\boldsymbol{x_0})(\boldsymbol{h}) = [J_f(\boldsymbol{x_0})]\boldsymbol{h}$$

Consider the transformation from polar coordinates to rectangular coordinates, which is the function $f(r, \theta) : (0, \infty) \times [0, 2\pi) \subset \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$f(r, \theta) = \langle r \cos(\theta), r \sin(\theta) \rangle$$

Then the Jacobian of f is the matrix

$$[J_f(r, \theta)] = \begin{bmatrix} \cos(\theta), -r \sin(\theta) \\ \sin(\theta), r \cos(\theta) \end{bmatrix}$$

Remark 3.1.22. *The converse does not hold - there exists functions f such that all the partial derivatives exist at some point $\boldsymbol{x_0}$, but f is not differentiable at $\boldsymbol{x_0}$.*

Consider the function $f(x,y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$

Then the partial derivatives of f exist at $(0,0)$, but f is not continuous at $(0,0)$!

picture

Consider the function $f(x,y) = \begin{cases} \frac{2xy(x+y)}{x^2+y^2} & \text{if } (x,y) \neq (0,0) \\ 0 & \text{otherwise} \end{cases}$

Then the partial derivatives exist at $(0,0)$, but f is not differentiable at $(0,0)$. You will prove this in the exercises below.

picture

However, if we strengthen the hypotheses, then we do have the following result:

Theorem 3.1.23. *Let $f : A \subset \mathbb{R}^m \to \mathbb{R}^n$ be a multivariable function. If all the partial derivatives $D_i f^j \boldsymbol{x_0}$) exist and are continuous in some open ball $B_\varepsilon(\boldsymbol{x_0})$, then f is differentiable at $\boldsymbol{x_0}$.*

The chain rule

Let us now revisit the previous theorems, now that we can compute general multivariable derivatives!

Theorem 3.1.24 (The Chain rule)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}^m$, and let $g : \mathbb{R}^m \to \mathbb{R}^k$ be multivariable functions such that f is differentiable at $\boldsymbol{x_0} \in \mathbb{R}^n$, and g is differentiable at $f(\boldsymbol{x_0}) \in \mathbb{R}^m$.*

Then $g \circ f : \mathbb{R}^n \to \mathbb{R}^k$ is differentiable at $\boldsymbol{x_0} \in \mathbb{R}^n$, and

$$D(g \circ f)(\boldsymbol{x_0}) = Dg(f(\boldsymbol{x_0})) \circ Df(\boldsymbol{x_0})$$

We can prove this using the definition of derivative.

However, since we know that the derivative can be computed in terms of the Jacobian, we equivalently have

Theorem 3.1.25 (The Chain rule in coordinates)**.** *Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\boldsymbol{x_0} \in \mathbb{R}^n$, and $g : \mathbb{R}^m \to \mathbb{R}^k$ is differentiable at $f(\boldsymbol{x_0}) \in \mathbb{R}^m$. Then*

$$[J_{g\circ f}(\boldsymbol{x_0})] = [J_g(f(\boldsymbol{x_0}))]\,[J_f(\boldsymbol{x_0})]$$

We can spell this out in terms of coordinates:

[The Chain rule in coordinates] Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\boldsymbol{x_0} \in \mathbb{R}^n$, and $g : \mathbb{R}^m \to \mathbb{R}^k$ is differentiable at $f(\boldsymbol{x_0}) \in \mathbb{R}^m$. Then

$$\begin{bmatrix} D_1(g\circ f)^1(\boldsymbol{x_0}) & \cdots & D_n(g\circ f)^1(\boldsymbol{x_0}) \\ \vdots & \vdots & \vdots \\ D_1(g\circ f)^k(\boldsymbol{x_0}) & \cdots & D_n(g\circ f)^k(\boldsymbol{x_0}) \end{bmatrix} = \begin{bmatrix} D_1 g^1(f(\boldsymbol{x_0})) & \cdots & D_m g^1(f(\boldsymbol{x_0})) \\ \vdots & \vdots & \vdots \\ D_1 g^k(f(\boldsymbol{x_0})) & \cdots & D_m g^k(f(\boldsymbol{x_0})) \end{bmatrix} \begin{bmatrix} D_1 f^1(\boldsymbol{x_0}) & \cdots & D_n f^1(\boldsymbol{x_0}) \\ \vdots & \vdots & \vdots \\ D_1 f^m(\boldsymbol{x_0}) & \cdots & D_n f^m(\boldsymbol{x_0}) \end{bmatrix}$$

How can we compute $D(g \circ f)(\boldsymbol{x_0})$? Equivalently, what are the entries of the matrix $[J_{g\circ f}(\boldsymbol{x_0})]$?

To answer this question, let us look at the chain rule in a specific case:

[The Chain rule for $\mathbb{R}^n \to \mathbb{R}^m \to \mathbb{R}$] Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\boldsymbol{x_0} \in \mathbb{R}^n$, and $g : \mathbb{R}^m \to \mathbb{R}$ is differentiable at $f(\boldsymbol{x_0}) \in \mathbb{R}^m$. Then

$$\begin{bmatrix} D_1(g\circ f)(\boldsymbol{x_0}) & \cdots & D_n(g\circ f)(\boldsymbol{x_0}) \end{bmatrix} = \begin{bmatrix} D_1 g(f(\boldsymbol{x_0})) & \cdots & D_m g(f(\boldsymbol{x_0})) \end{bmatrix} \begin{bmatrix} D_1 f^1(\boldsymbol{x_0}) & \cdots & D_n f^1(\boldsymbol{x_0}) \\ \vdots & \vdots & \vdots \\ D_1 f^m(\boldsymbol{x_0}) & \cdots & D_n f^m(\boldsymbol{x_0}) \end{bmatrix}$$

Therefore,

$$\mathcal{D}_i(g \circ f) = \sum_k D_i f^k \; D_k g$$

Corollary 3.1.26. *Suppose that $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $\boldsymbol{x_0} \in \mathbb{R}^n$, and $g : \mathbb{R}^m \to \mathbb{R}$ is differentiable at $f(\boldsymbol{x_0}) \in \mathbb{R}^m$. Then*

$$[J_{g\circ f}(\boldsymbol{x_0})] = \left[D_i(g\circ f)^j\right] = \left[\sum_k D_i f^k \; D_k g^j\right]$$

Let $f(x,y,z) = xy + z$. Calculate $\frac{\partial f}{\partial s}$, where $x = s^2$, $y = st$, $z = t^2$.

Let us now turn to another specific case of the chain rule:

Proposition 3.1.27 (Chain rule for paths)**.** *Let $f(x_1, \cdots x_n) : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function, and let $\boldsymbol{r}(t) = \langle x_1(t), \cdots, x_n(t)\rangle : \mathbb{R} \to \mathbb{R}^n$ be a vector-valued function. Then $f(\boldsymbol{r}(t)) : \mathbb{R} \to \mathbb{R}$ is a single variable function, and*

$$\begin{aligned} \frac{d}{dt} f(\boldsymbol{r}(t_0)) &= \begin{bmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{r}(t_0)) & \cdots & \frac{\partial f}{\partial x_n}(\boldsymbol{r}(t_0)) \end{bmatrix} \begin{bmatrix} x_1'(t_0) \\ \vdots \\ x_n'(t_0) \end{bmatrix} \\ &= \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(\boldsymbol{r}(t_0))\right) x_i'(t_0) \end{aligned}$$

This measures the rate of change of f along the path $\boldsymbol{r}(t)$.
PICTURE

Consider the linear path through $\boldsymbol{x_0} \in \mathbb{R}^n$ in the direction of a vector $\boldsymbol{v}$, say

$$\boldsymbol{r}(t) = \boldsymbol{x_0} + t\boldsymbol{v}$$

Observe that the chain rule depends on the magnitude of $\boldsymbol{v}$

Let us make a definition that is independent of $||\boldsymbol{V}||$

Definition 3.1.28. *If $\boldsymbol{u} = \langle u_1, \cdots, u_n\rangle$ is a unit vector in $\mathbb{R}^n$, then the* ***directional derivative*** *in the direction of $\boldsymbol{u}$ at the point $\boldsymbol{x_0} \in \mathbb{R}^n$ is defined as*

$$D_{\boldsymbol{u}} f(\boldsymbol{x_0}) = u_1 \frac{\partial f}{\partial x_1}(\boldsymbol{x_0}) + \cdots + u_n \frac{\partial f}{\partial x_n}(\boldsymbol{x_0})$$

The directional derivative measures the rate of change of f in the direction of $\boldsymbol{u}$. That is,

$$D_{\boldsymbol{u}} f(\boldsymbol{x_0}) = \frac{d}{dt} f(\boldsymbol{r}(0))$$

for $\boldsymbol{r}(t) = \boldsymbol{x_0} + t\boldsymbol{u}$.

PICTURE

The gradient

Definition 3.1.29. *If $f(x_1, \cdots, x_n)$ is a function of n variables, then the* ***gradient*** *of f is the vector-valued function*

$$\nabla f = \langle \frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n} \rangle$$

That is,

$$\nabla f = [Df(\boldsymbol{x_0})]^\top$$

transpose meaning

Proposition 3.1.30. *The chain rule for paths can be rewritten as*

$$\frac{d}{dt} f(\boldsymbol{r}(t_0)) = \nabla f(\boldsymbol{r}(t_0)) \cdot \boldsymbol{r}'(t_0)$$

Corollary 3.1.31. *If $\boldsymbol{u} \in \mathbb{R}^n$ is a unit vector, then the directional derivative in the direction of $\boldsymbol{u}$ at the point $\boldsymbol{P} \in \mathbb{R}^n$ can be computed as*

$$D_{\boldsymbol{u}} f(\boldsymbol{P}) = \nabla f(\boldsymbol{P}) \cdot \boldsymbol{u}$$

This helps us geometrically interpret the gradient:

Proposition 3.1.32. *The directional derivative $D_{\boldsymbol{u}} f$ is* ***maximized*** *when $\theta = 0$, so when $\boldsymbol{u} = \boldsymbol{e}_{\nabla f}$. The maximum value of $D_{\boldsymbol{u}} f$ is $||\nabla f||$.*

The directional derivative $D_{\boldsymbol{u}} f$ is ***minimized*** *when $\theta = \pi$, so when $\boldsymbol{u} = -\boldsymbol{e}_{\nabla f}$. The minimum value of $D_{\boldsymbol{u}} f$ is $-||\nabla f||$.*

Thinking of z as the height of $z = f(x, y)$, the gradient ∇f points in the direction of **steepest ascent**.

The opposite of the gradient, $-\nabla f$, points in the direction of **steepest descent**.

Furthermore, if $\boldsymbol{r}(t)$ parameterizes a level curve $f(x_1, \cdots, x_n) = k$, recall that means that for all t,

$$f(\boldsymbol{r}(t)) = k$$

Proposition 3.1.33. *The gradient ∇f is* ***orthogonal*** *to the (tangent lines of) the level curves.*

PICTURE

Corollary 3.1.34. *Let $f(x_1, \cdots, x_n) : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function at a point $\boldsymbol{a} = (a_1, \cdots, a_n)$. Moreover, suppose that $f(a_1, \cdots, a_n) = k$.*

Then $\nabla f = \boldsymbol{0}$, or ∇f is ***orthogonal*** *to the surface*

$$f(x_1, \cdots, x_n) = k$$

Corollary 3.1.35. *The **tangent hyperplane** to the surface $f(x_1, \cdots, x_n) = k$ in $\mathbb{R}^n$ at the point $\boldsymbol{a} = (a_1, \cdots, a_n)$ is given by*

$$\nabla f(\boldsymbol{a}) \cdot (\boldsymbol{x} - \boldsymbol{a}) = 0$$

How does this relate to previous notions of tangent hyperplane?

Theorem 3.1.36. *The graph of a multivariable function $g(x_1, \cdots, x_n) : \mathbb{R}^n \to \mathbb{R}$ can be described as*

$$x_{n+1} = g(x_1, \cdots, x_n)$$

in $\mathbb{R}^{n+1}$. The equation of the tangent hyperplane in $\mathbb{R}^{n+1}$ at a point

$$\overline{\boldsymbol{a}} = (a_1, \cdots, a_n, g(a_1, \cdots, a_n))$$

is given by

$$\left\langle \frac{\partial g}{\partial x_1}, \cdots, \frac{\partial g}{\partial x_n}, -1 \right\rangle \cdot (\overline{\boldsymbol{x}} - \overline{\boldsymbol{a}}) = 0$$

Observe that if we set $f(x_1, \cdots, x_{n+1}) = g(x_1, \cdots, x_n) - x_{n+1}$, then

$$\nabla f = \left\langle \frac{\partial g}{\partial x_1}, \cdots, \frac{\partial g}{\partial x_n}, -1 \right\rangle$$

Exercises

3.1 Find a function $f(x) : \mathbb{R} \to \mathbb{R}$ that shows that defining the derivative of a multivariable function as

$$f'(\boldsymbol{x_0}) = \lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{f(\boldsymbol{x_0} + \boldsymbol{h}) - f(\boldsymbol{x_0})}{||\boldsymbol{h}||}$$

does not generalize the single variable derivative.

3.2 Consider the function $f(x, y) = x$. Show that

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} \frac{f(\boldsymbol{x_0} + \boldsymbol{h}) - f(\boldsymbol{x_0})}{||\boldsymbol{h}||}$$

does not exist at $(0, 0)$. (Hence, this is also not a good definition of the multivariable derivative).

3.3 Use linear approximation to approximate $\frac{8.01}{\sqrt{(1.99)(2.01)}}$.

3.4 Use linear approximation to approximate $(2.92)^2\sqrt{4.08}$.

3.5 Consider the function $f(x, y) = \begin{cases} \frac{xy}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{otherwise} \end{cases}$ Use the limit definition of the partial derivative to compute $f_x(0, 0)$ and $f_y(0, 0)$.

3.6 Consider the function $f(x, y) = \begin{cases} \frac{x^3}{x^2+y^2} & \text{if } (x, y) \neq (0, 0) \\ 0 & \text{otherwise} \end{cases}$

Use the limit definition of the partial derivative to compute $f_x(0, 0)$ and $f_y(0, 0)$.

3.7 Let $f(u, v) = \tan(uv^3)$. Find the partial derivatives $f_u(u, v)$ and $f_v(u, v)$.

3.8 Let $f(x, y, z) = xy + z$. Calculate $\frac{\partial f}{\partial t}$, where $x = s^2$, $y = st$, $z = t^2$.

3.9 Let $f(x,y) = \tan(xy^2)$, and consider the path $\boldsymbol{r}(t) = \langle t, e^t \rangle$. Compute $\frac{d}{dt} f(\boldsymbol{r}(t))$ at the point $(2, e^2)$.

3.10 alculate the gradient of $g(x,y,z) = x\ln(y+z)$.

3.11 ind the maximum rate of change of $f(x,y) = e^{xy-y^2}$at the point $(1,1)$.

3.12 Let $f(x,y) = e^{xy-y^2}$. Compute the directional derivative in the direction of $\boldsymbol{u} = \langle 5, 12 \rangle$ at the point $P = (1,1)$.

3.13 Find the directional derivative of $g(x,y,z) = xy + z^2$ at the point $P = (3,2,-1)$, in the direction pointing to the origin.

3.14 Find the directional derivative of $g(x,y,z) = x\ln(y+z)$ in the direction of the vector $\boldsymbol{v} = \langle 2, -1, 1 \rangle$ at the point $P = (2, e, e)$.

3.15 Find an equation of the plane tangent to the graph of $f(x,y) = xy^3 + x^2$ at the point $(2, -2, -12)$

3.2 Optimizing Multivariable functions

Recall that in single-variable calculus, the derivative is also useful in optimising single-variable functions. That is, finding local maxima and minima, as well as global maxima and minima.

Definition 3.2.1. *A function $f : \mathbb{R} \to \mathbb{R}$ has a **local maximum** at $P \in \mathbb{R}$ if there exists some $r > 0$ such that*

$$f(P) \geq f(Q)$$

for all points $Q \in B_r(P)$.

Definition 3.2.2. *A function $f : \mathbb{R} \to \mathbb{R}$ has a **local mainmum** at $P \in \mathbb{R}$ if there exists some $r > 0$ such that*

$$f(P) \leq f(Q)$$

for all points $Q \in B_r(P)$.

Global

How can we find local maxima and minima?

Definition 3.2.3. *A point $P \in \mathbb{R}$ in the domain of $f(x)$ is a **critical point** if*

a. $f'(P) = 0$, *OR*

b. $f'(P)$ *does not exist.*

Theorem 3.2.4 (The single-variable first derivative test)**.** *If $f(x)$ has a local maximum or minimum at P, then P is a critical point of $f(x)$.*

We will see how to generalize these ideas to multivariable calculus:

Local optimization

First, let us define the notion of a local extremum for a multivariable function $f : \mathbb{R}^n \to \mathbb{R}$:

Definition 3.2.5. *A function $f : \mathbb{R}^n \to \mathbb{R}$ has a* ***local maximum*** *at $\boldsymbol{P} \in \mathbb{R}^n$ if there exists some $r > 0$ such that*

$$f(\boldsymbol{P}) \geq f(\boldsymbol{Q})$$

for all points $\boldsymbol{Q} \in B_r(\boldsymbol{P})$.

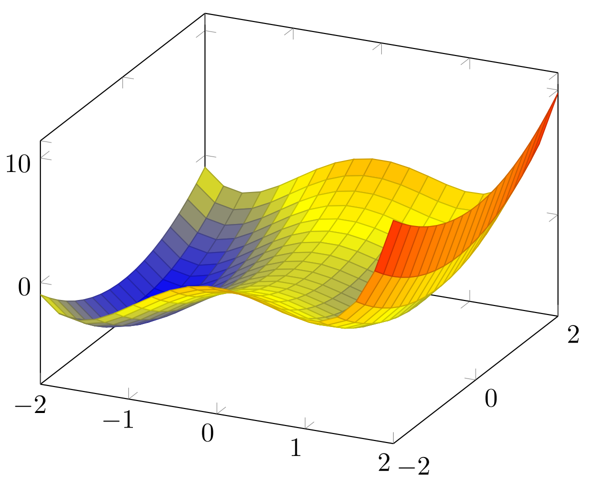
Definition 3.2.6. *A function $f : \mathbb{R}^n \to \mathbb{R}$ has a* ***local minimum*** *at $\boldsymbol{P} \in \mathbb{R}^n$ if there exists some $r > 0$ such that*

$$f(\boldsymbol{P}) \leq f(\boldsymbol{Q})$$

for all points $\boldsymbol{Q} \in B_r(\boldsymbol{P})$.

Note: local extrema only make sense for functions $f : \mathbb{R}^n \to \mathbb{R}$.

Remark 3.2.7. *Observe that* ***local*** *maxima/minima are not necessarily* ***global*** *maxima/minima.*



What is the multivariable analogue of the first derivative test?

Remark 3.2.8. *Observe that if $\boldsymbol{P} \in \mathbb{R}^n$ is a local maxima or minima of $f : \mathbb{R}^n \to \mathbb{R}$, and if f is differentiable at $\boldsymbol{P}$, then $Df(\boldsymbol{P})$ is the zero linear transformation (equivalently, $\nabla f = \boldsymbol{0}$).*

That is, the tangent plane at $\boldsymbol{P}$ is parallel to the plane $x_{n+1} = 0$.

Definition 3.2.9. *A point $\boldsymbol{P} \in \mathbb{R}^n$ is said to be a* ***critical point*** *of a function $f : \mathbb{R}^n \to \mathbb{R}$ if either*

a. *$Df(\boldsymbol{P}) = 0$, OR*

b. *$Df(\boldsymbol{P})$ does not exist.*

Proposition 3.2.10. *Equivalently, $\boldsymbol{P}$ is a critical point of $f : \mathbb{R}^n \to \mathbb{R}$ if*

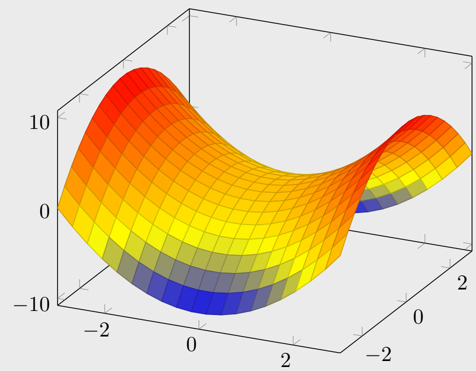
a. *$\nabla f(\boldsymbol{P}) = \boldsymbol{0}$, OR*

b. *$\nabla f(\boldsymbol{P})$ does not exist.*

Theorem 3.2.11 (The multivariable first derivative test). *If $f : \mathbb{R}^n \to \mathbb{R}$ has a local maximum or minimum at $\boldsymbol{P}$, then $\boldsymbol{P}$ is a critical point of f.*

Remark 3.2.12. *The converse of the first derivative test does not hold:*
If $\boldsymbol{P}$ is a critical point of f, then $\boldsymbol{P}$ is not necessarily a local maximum or minimum.

Consider the function $f(x, y) = x^2 - y^2$. Observe that $(0, 0)$ is a critical point, but it is neither a local maximum or a minimum.



The Hessian

How can we algebraically classify the behavior of critical points?

Recall that in single-variable calculus, we can do so using the **second derivative test**:

Theorem 3.2.13 (The single-variable second derivative test). *Let $f : \mathbb{R} \to \mathbb{R}$ be a single variable function.*

a. *If $f'(P) = 0$ and $f''(P) > 0$, then there is a **local minimum** at $x = P$.*

b. *If $f'(P) = 0$ and $f''(P) < 0$, then there is a **local maximum** at $x = P$.*

c. *If $f'(P) = 0$ and $f''(P) = 0$, then the test is **inconclusive**.*

*If $f''(P)$ does not exist, then the test is **inconclusive**.*

What is the multivariable analogue of the second derivative test?

First, we need to figure out what we mean by the second derivative of a multivariable function. Recall from Reference that the derivative of a linear transformation T is T itself. Thus, we can

Remark 3.2.14. *We have defined the derivative $Df(\boldsymbol{x_0})$ to be the linear approximation of f* at a point $\boldsymbol{x_0} \in \mathbb{R}^n$.

This is analogous to finding the slope of the tangent line of a single variable function at a point $x \in \mathbb{R}$.

Therefore, we wish to find the derivative of Df, not of $Df(\boldsymbol{x_0})$, where

$$Df : \mathbb{R}^n \to M_{1\times n}(\mathbb{R})$$

However, observe that we have not defined the derivative of a function that outputs matrices. However, since we know that

$$M_{1\times n}(\mathbb{R}) \cong \mathbb{R}^n$$

via the transpose map $(-)^\intercal$

$$M \mapsto M^\intercal$$

Let us consider the map

$$(Df)^\intercal : \mathbb{R}^n \to M_{1\times n}(\mathbb{R}) \to \mathbb{R}^n$$

which sends a vector

$$\boldsymbol{x_0} \mapsto \begin{bmatrix} D_1 f(\boldsymbol{x_0}) \\ \vdots \\ D_n f(\boldsymbol{x_0}) \end{bmatrix}$$

Observe that this is the same thing as the gradient of f! Thus, we can make the following definition:

Definition 3.2.15. *A multivariable function $f : \mathbb{R}^n \to \mathbb{R}$ is **twice-differentiable** if $f : \mathbb{R}^n \to \mathbb{R}$ and $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ are both differentiable.*

*If $f : \mathbb{R}^n \to \mathbb{R}$ is twice-differentiable, then the **second derivative of f at $\boldsymbol{x_0}$** is defined as*

$$D^2 f(\boldsymbol{x_0}) := D(\nabla f)(\boldsymbol{x_0})$$

Definition 3.2.16. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice-differentiable function. Then the standard matrix of $D^2 f(\boldsymbol{x_0})$ is called the **Hessian matrix**,*

$$[H_f(\boldsymbol{x_0})]$$

Proposition 3.2.17. *The **Hessian matrix** of $f : \mathbb{R}^n \to \mathbb{R}$ at $\boldsymbol{x_0}$ is*

$$[H_f(\boldsymbol{x_0})] = \begin{bmatrix} D_1D_1 f(\boldsymbol{x_0}) & D_2D_1 f(\boldsymbol{x_0}) & \cdots & D_nD_1 f(\boldsymbol{x_0}) \\ D_1D_2 f(\boldsymbol{x_0}) & D_2D_2 f(\boldsymbol{x_0}) & \cdots & D_nD_2 f(\boldsymbol{x_0}) \\ \vdots & \vdots & \vdots & \vdots \\ D_1D_n f(\boldsymbol{x_0}) & D_2D_n f(\boldsymbol{x_0}) & \cdots & D_nD_n f(\boldsymbol{x_0}) \end{bmatrix}$$

In other words, the Hessian matrix is the $n \times n$ matrix of all second-order partial derivatives of f,

$$f_{x_j x_i} = D_i D_j f = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)$$

Let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x, y) = xe^{xy}$. Compute the partial derivatives, the Jacobian matrix, ∇f, and the Hessian matrix of f at $(1, 1)$.

Definition 3.2.18. *A function $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is said to be of class C^2 (writing $f \in C^2(U)$) if all second-order partial derivatives exist and are continuous on U.*

Definition 3.2.19. *A function $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is said to be of class C^k (writing $f \in C^k(U)$) if all kth-order partial derivatives exist and are continuous on U.*

Definition 3.2.20. *A function $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is said to be of class C^∞ (writing $f \in C^\infty(U)$) if all partial derivatives exist and are continuous on U.*

Theorem 3.2.21 (Clairaut's theorem)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$. Suppose that $D_i f$, $D_j f$, and $D_i D_j f$ exist and are continuous on an open disk $D \subset \mathbb{R}^n$. Then $D_j D_i f$ exists on D, and moreover*

$$D_i D_j = D_j D_i \text{ on the disk } D$$

Clairaut's theorem is more commonly stated as the following corollary:

Corollary 3.2.22. *Let $f \in C^2(U)$. Then*

$$D_i D_j = D_j D_i \text{ on the region } U$$

We can use Clairaut's theorem to compute the partial derivative g_{zzwx} for

$$g(x, y, w, z) = x^2 w^2 z^2 + \sin\left(\frac{xy}{z^2}\right)$$

Note that as long as $z \neq 0$, we have that all partial derivatives of g exist.
Thus, $g_{zzwx} = g_{wxzz}$ as long as $z \neq 0$.
Observe that $g_w = 2x^2 w z^2$. Thus, $g_{zzwx} = g_{wxzz} = 8wx$, as long as $z \neq 0$

Corollary 3.2.23. *If $f \in C^2(U)$, then*

$$[H_f(\boldsymbol{x_0})]^\intercal = [H_f(\boldsymbol{x_0})]$$

With the second derivative defined, we can now state the two-variable second derivative test:

Theorem 3.2.24 (The two-variable second derivative test)**.** *Let $\boldsymbol{x_0} \subset U$ be a critical point of $f(x, y) : U \to \mathbb{R}$, and suppose that $f \in C^2(U)$. Let us write $D = \det[H_f(\boldsymbol{x_0})]$.*

a. *If $D > 0$ and $f_{xx}(\boldsymbol{x_0}) > 0$, then there is a **local minimum** at $\boldsymbol{x_0}$.*

b. *If $D > 0$ and $f_{xx}(\boldsymbol{x_0}) < 0$, then there is a **local maximum** at $\boldsymbol{x_0}$.*

c. *If $D < 0$, then f has a **saddle point** at $\boldsymbol{x_0}$.*

d. *If $D = 0$, then the test is **inconclusive**.*

*If D does not exist, then the test is **inconclusive**.*

Remark 3.2.25. *At a critical point $P = (a, b)$, we can apply the single-variable second derivative test to*

- *the trace $f(a, y) = g(y)$ in the plane $x = a$*
- *the trace $f(x, b) = h(x)$ in the plane $y = b$*

For the former, observe that

$$f_{yy}(a, b) = g''(b) \text{ is } \begin{cases} > 0 & \text{if } f(a, y) = g(y) \text{ has a local min at } b \\ < 0 & \text{if } f(a, y) = g(y) \text{ has a local max at } b \end{cases}$$

For the latter, observe that

$$f_{xx}(a, b) = h''(a) \text{ is } \begin{cases} > 0 & \text{if } f(x, b) = h(x) \text{ has a local min at } a \\ < 0 & \text{if } f(x, b) = h(x) \text{ has a local max at } a \end{cases}$$

How does this generalize to multiple variables?

Theorem 3.2.26 (Second order Taylor approximation). *If $f : U \subset \mathbb{R}^n \to \mathbb{R}$ is differentiable at a point $\boldsymbol{a} \in \mathbb{R}^n$, and $\boldsymbol{x}$ is close to $\boldsymbol{a}$, then*

$$f(\boldsymbol{x}) \approx f(\boldsymbol{a}) + [Df(\boldsymbol{a})](\boldsymbol{x}-\boldsymbol{a}) + \frac{1}{2}\left([H_f(\boldsymbol{a})](\boldsymbol{x}-\boldsymbol{a})\right)\cdot(\boldsymbol{x}-\boldsymbol{a})$$

Definition 3.2.27. *Let $A \in M_{n\times n}(\mathbb{R})$ be a symmetric matrix (that is, $A^\intercal = A$). Then A is said to be*

a. ***positive definite*** , *if for all $\boldsymbol{x} \neq \boldsymbol{0}$, then $(A\boldsymbol{x})\cdot\boldsymbol{x} > 0$,*

b. ***negative definite*** *if for all $\boldsymbol{x} \neq \boldsymbol{0}$, then $(A\boldsymbol{x})\cdot\boldsymbol{x} < 0$,*

c. ***indefinite*** *otherwise.*

Theorem 3.2.28 (The multivariable second derivative test). *Let $\boldsymbol{x_0} \in U \subset \mathbb{R}^n$ be a critical point of a function $f : U \subset \mathbb{R}^n \to \mathbb{R}$.*

Suppose that $f \in C^2(U)$, and that the Hessian $[H_f(\boldsymbol{x_0})]$ is invertible.

a. *If $[H_f(\boldsymbol{x_0})]$ is positive definite, then there is a **local minimum** at $\boldsymbol{x_0}$.*

b. *If $[H_f(\boldsymbol{x_0})]$ is negative definite, then there is a **local maximum** at $\boldsymbol{x_0}$.*

c. *If $[H_f(\boldsymbol{x_0})]$ is indefinite, then f has a **saddle point** at $\boldsymbol{x_0}$.*

Exercises

3.16 onsider the function

$$f(x,y,z,w) = \frac{3x^2 + e^y z + x}{3y^2 + 2e^{w^2}}$$

Compute the fourth order partial derivative f_{wyzx}.

3.17 Consider the function

$$f(x,y) = x^3 - xy + y^3$$

Find the critical points of f, and use the second derivative test to classify the critical points of f.

3.18 Consider the function

$$f(x,y) = x^2 + y^3 - 2x - 6y$$

Find the critical points of f, and use the second derivative test to classify the critical points of f.

3.19 onsider the function

$$f(x,y) = \ln(x) + 2\ln(y) - x - 4y$$

Find the critical points of f, and use the second derivative test to classify the critical points of f.

Global Optimization

We saw from the previous section that the gradient and the second derivative test allow us to find and classify **local** maxima and minima.

(On a given domain), how can we find **global** maxima and minima?

In single-variable calculus, we have the following theorem:

Theorem 3.2.29 (Single-variable global minima and maxima)**.** *If $f(x)$ is continuous on a closed and bounded interval $[a, b]$, then f has a global maximum and a minimum on $[a, b]$.*

That is, there exists an $M \in [a, b]$ and an $m \in [a, b]$ such that $f(m) \leq f(x) \leq f(M)$ for all $x \in [a, b]$.

Theorem 3.2.30. *The global maxima and minima of $f(x)$ on $[a, b]$ either occur at the critical points of f in $[a, b]$, or on the boundary of $[a, b]$.*

Remark 3.2.31. *Observe that if $A \subset \mathbb{R}$ is not closed, or not bounded, then global maxima/minima are not guaranteed to exist.*

Consider the function $f(x) = x$ on the interval $A = [0, 1)$.

Observe that this has function has no global maximum on A, but has a global minimum at 0.

Consider the function $f(x) = x$ on the interval $A = \{x \mid x \geq 0\}$.

Observe that this has function has no global maximum on A, but has a global minimum at 0.

Thus, we have a few questions to consider:

How do we generalize the notion of global maxima/minima to multivariable functions?

How do we generalize the notion of a bounded and closed interval $[a, b] \subset \mathbb{R}$ to a subset $A \subset \mathbb{R}^n$?

Can we find the possible points where global maxima/minima occur?

The first question is not that difficult to answer:

Definition 3.2.32. *A function $f : A \subset \mathbb{R}^n \to \mathbb{R}$ has a* ***global maximum*** *at $\boldsymbol{x_0} \in A$ if*

$$f(\boldsymbol{x_0}) \geq f(\boldsymbol{x})$$

for all $\boldsymbol{x} \in A$.

Definition 3.2.33. *A function $f : A \subset \mathbb{R}^n \to \mathbb{R}$ has a* ***global minimum*** *at $\boldsymbol{x_0} \in A$ if*

$$f(\boldsymbol{x_0}) \leq f(\boldsymbol{x})$$

for all $\boldsymbol{x} \in A$.

Let us now turn to the notions of closed and boundedness:

Definition 3.2.34. *A subset $D \subset \mathbb{R}^n$ is* ***bounded*** *if there exists some $r > 0$ such that $D \subset B_r(\mathbf{0})$.*

Definition 3.2.35. *A point $\boldsymbol{x_0} \in \mathbb{R}^n$ is a* ***boundary point*** *of $D \subset \mathbb{R}^n$ if: for all $\varepsilon > 0$,*

a. *$B_\varepsilon(\boldsymbol{x_0}) \cap D$ is non-empty,* ***and***

b. *$B_\varepsilon(\boldsymbol{x_0}) \cap D^c$ is non-empty.*

Definition 3.2.36. *A subset $D \subset \mathbb{R}^n$ is* ***closed*** *if it contains all of its boundary points.*

The set

$$D = \{(x, y) \in \mathbb{R}^2 \mid x \geq 0, y \geq 0\}$$

is closed, but not bounded.

Its boundary is the set $\{(x, 0) \in \mathbb{R}^2 \mid x \geq 0\} \bigcup \{(0, y) \in \mathbb{R}^2 \mid y \geq 0\}$

The set

$$D = \{(x, y) \in \mathbb{R}^2 \mid 1 < (x - a)^2 + (y - b)^2 \leq 3\}$$

is bounded, but not closed.

Its boundary is the set $\{(x, y) \in \mathbb{R}^2 \mid (x-a)^2+(y-b)^2 = 1\} \bigcup \{(x, y) \in \mathbb{R}^2 \mid (x-a)^2+(y-b)^2 = 3\}$

Theorem 3.2.37 (Multivariable global minima and maxima)**.** *If D is a closed and bounded subset of $\mathbb{R}^n$, and f is a continuous function on D, then f has a global maximum and a global minimum in D.*

That is, there exists an $\boldsymbol{M} \in D$ and an $\boldsymbol{m} \in D$ such that $f(\boldsymbol{m}) \leq f(\boldsymbol{x}) \leq f(\boldsymbol{M})$ for all $\boldsymbol{x} \in D$.

Theorem 3.2.38. *The global maxima and minima of f on D either occur at the critical points of f in D, or on the boundary of D.*

We know how to use the gradient and the second derivative test allow us to find and classify **local** maxima and minima. The question we turn to now is how to find maxima and minima on the boundary of a region D

Constrained optimization

Given a region D, how can we find the maxima and minima on D?

Remark 3.2.39. *The boundary of $D \subset \mathbb{R}^2$ can be thought of as:*

- *the parametric curve corresponding to $\boldsymbol{r}(t) = \langle x(t), y(t) \rangle$.*
- *The $z = 0$ trace of a graph $z = g(x, y)$*

To find local maxima and minima of $f(x, y)$ on the boundary of D, we are trying to find the local maxima and minima of $f(x, y)$ subject to the condition $g(x, y) = 0$.

In other words, we are trying to optimize $f(x, y)$ given the constraint $g(x, y) = 0$. We can do so using the method of Lagrange multipliers:

Theorem 3.2.40 (The method of Lagrange Multipliers)**.** *Assume that $f(x, y)$ and $g(x, y)$ are differentiable functions. If*

a. *$f(x, y)$ has a local maximum or minimum subject to the constraint $g(x, y) = 0$ at a point (a, b),* ***AND***

b. *if $\nabla g(a, b) \neq 0$*

then there is a scalar λ such that

$$\nabla f(a, b) = \lambda \nabla g(a, b)$$

Corollary 3.2.41. *We can use the* ***Lagrange equations***

$$f_x(a, b) = \lambda g_x(a, b) \qquad \text{and} \qquad f_y(a, b) = \lambda g_y(a, b)$$

to determine the critical points on the boundary.

Definition 3.2.42. *Points $P = (a, b)$ that satisfy the Lagrange equations for the optimization problem of $f(x, y)$ with constraint $g(x, y) = 0$ are called* ***critical points for optimization with constraint****.*

Remark 3.2.43. *A constrained critical point is not necessarily a local max or a local min.*

Maximize the area of a rectangular fence subject to the constraint that the perimeter is equal to 20.

Maximize $f(x, y) = xy$ subject to the constraint $4x^2 + 9y^2 = 32$.

Exercises

3.20 Is the set

$$D = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x + y \leq 1\}$$

closed? bounded? What is it's boundary?

3.21 Find the critical points of $f(x, y) = x^4 + y^4$ subject to the constraint $x^2 + y^2 = 1$.

3.22 Determine the coordinates of the global maxima and minima of $f(x, y) = x^4 + y^4$ on the disk $D = \{(x, y) \mid x^2 + y^2 \leq 1\}$.

3.23 Find the critical points of $f(x, y) = -x^2 + 2y^2 + 6x$ subject to the constraint $x^2 + y^2 = 1$.

3.24 Determine the coordinates of the global maxima and minima of $f(x, y) = -x^2 + 2y^2 + 6x$ on the disk $D = \{(x, y) \mid x^2 + y^2 \leq 1\}$.

3.25 Let $f(x, y) = (x^3 - 3x) + (y^3 - 3y)$. Find the constrained critical points of $f(x, y)$ on the boundary of the region $S = \{(x, y) \mid -1 \leq x \leq 1, -1 \leq y \leq 1\}$

3.26 Let $f(x,y) = (x^3 - 3x) + (y^3 - 3y)$. Find the global maxima and minima of $f(x,y)$ on the region $S = \{(x,y) \mid -1 \le x \le 1, -1 \le y \le 1\}$

3.27 ind a continuous function that does not have a global maximum on the domain

$$D = \{(x,y) \in \mathbb{R}^2 \mid 0 \le x + y \le 1\}$$

3.28 Find the critical points of $f(x,y) = 4x^2 + 9y^2$ subject to the constraint $xy = 4$.

3.29 Find the point (a,b) on the line $4x + 9y = 12$ that is closest to the origin.

3.30 Find the point $\boldsymbol{x}$ on the plane $ax + by + cz = d$ that is closest to the origin. What is the distance from $\boldsymbol{x}$ to the origin?

Chapter 4

Integration on $\mathbb{R}^n$

Chapter 5

Integration on manifolds

Chapter 6

Divergence theorem

Hints and Solutions

1.1. Position vectors.

Index