Multivariable Calculus

March 26, 2019

Contents

Part I

Review

Vectors

In this section, we review some basics about vectors. This includes the definition of a vector, basic vector operations, standard basis vectors, and notation.

Vectors

In linear algebra, we often worked with vectors. We begin by recalling the (algebraic) definition of a vector in $\mathbb{R}^n$.

Definition 1. *A* vector *in* $\mathbb{R}^n$ *is an ordered* n*-tuple of real numbers. That is, a vector* $\vec{v}$ *may be written as*

$$\vec{v} = (a_1, a_2, ..., a_n)$$

where $a_1, a_2, ..., a_n$ *are real numbers.*

We call the numbers a_i *the* components *or* entries *of the vector. We call* n *the* dimension *of the vector* $\vec{v}$*, and say that* $\vec{v}$ *is* n-dimensional*.*

We write the vector with an arrow above it, as $\vec{v}$, in order to make the distinction between vectors and *scalars*, which are just real numbers. Some other common notations for vectors are $\mathbf{v}$ and $\hat{v}$. It's important to make this distinction between vectors and scalars, so you should make use of one of these notations for vectors.

Example 1. $\vec{v} = (1, 3)$ *is a vector in* $\mathbb{R}^2$*.*

$\vec{w} = (-1, 5, 0)$ *is a vector in* $\mathbb{R}^3$*.*

$\vec{x} = (1, -2, 3)$ *is a vector in* $\mathbb{R}^3$*.*

$\vec{y} = (-6, \pi, 1/24, -0.5, 3)$ *is a vector in* $\mathbb{R}^5$*.*

It's sometimes convenient to write a vector as a column vector instead (particularly when working with linear transformations, which we'll review in a later

---

Learning outcomes:
Author(s):

section). We could write

$$\vec{v} = (a_1, a_2, ..., a_n) = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

or

$$\vec{v} = (a_1, a_2, ..., a_n) = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

The choice between square brackets and parentheses is just a difference in notation, they mean the same thing, and you should feel free to use either.

Example 2. *We write the following vectors as column vectors.*

$\vec{v} = (1, 3) = \begin{bmatrix} 1 \\ 3 \end{bmatrix}.$

$\vec{w} = (-1, 5, 0) = \begin{bmatrix} -1 \\ 5 \\ 0 \end{bmatrix}.$

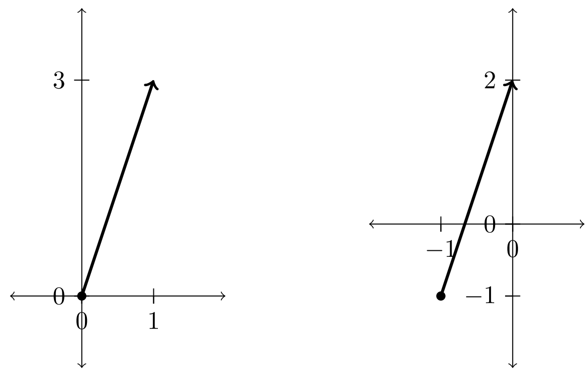
$\vec{x} = (1, -2, 3) = \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}.$

$\vec{y} = (-6, \pi, 1/24, -0.5, 3) = \begin{bmatrix} -6 \\ \pi \\ 1/24 \\ -0.5 \\ 3 \end{bmatrix}.$

The Geometric Perspective

We also can think of a vector geometrically, as giving a direction and magnitude, but without a fixed position.

In two or three dimensions, it is useful to visualize a vector as an arrow in $\mathbb{R}^n$. We might visualize a vector $\vec{v} = (1, 3)$ in $\mathbb{R}^2$ as the arrow starting at the origin and ending at the point $(1, 3)$, thus giving a direction and a magnitude. However, we typically don't think of a vector as having a set location. We could also visualize the vector $\vec{v}$ as starting at the point $(-1, -1)$ and ending at the point $(0, 2)$. Note that this arrow would have the same direction and magnitude as the one starting at the origin, thus they represent the same vector.



In four or higher dimensions, visualizing anything becomes very difficult. It can still be useful to think of a vector $(1, 2, 3, 4, 5)$ in $\mathbb{R}^5$ as starting at the origin and ending at the point $(1, 2, 3, 4, 5)$, but you probably won't be able to have a very clear picture of this in your head.

This concept will probably seem more useful once you think about a displacement vector.

Definition 2. *Given points $P_1 = (x_1, ..., x_n)$ and $P_2 = (y_1, ..., y_n)$ in $\mathbb{R}^n$, the* displacement vector *from P_1 to P_2 is*

$$\vec{P_1 P_2} = (y_1 - x_1, ..., y_n - x_n).$$

This is the vector that starts at P_1 and ends at P_2.

Notice that the notation $(a_1, ..., a_n)$ that we use for a vector in $\mathbb{R}^n$ is identical to the notation that we'd use for a point in $\mathbb{R}^n$. Since both vectors and points in $\mathbb{R}^n$ are defined as n-tuples of points, they are, in some sense, the same thing. The difference between the two comes when we consider the context and geometric significance of the vector or point that we're working with. As we move into multivariable calculus, we'll often blur the distinction between a vector and a point, and sometimes think of a vector as a point and vice versa. This will be greatly simplify notation, and we promise that it won't be as confusing as it sounds!

Vector Operations

Before defining some basic vector operations, we define what it means for two vectors to be equal. This is done by comparing the components of the vectors.

Definition 3. *Two vectors $(a_1, a_2, ..., a_n)$ and $(b_1, b_2, ..., b_n)$ in $\mathbb{R}^n$ are* equal *if their corresponding components are equal, so $a_1 = b_1$, $a_2 = b_2$, ... , $a_n = b_n$.*

Notice that, in order to be equal, two vectors must have the same dimension and the same entries in the same order. Thus, the vectors $(1,3)$ and $(1,3,0)$ are not equal.

We now define addition of two vectors of the same dimension, which is done componentwise.

Definition 4. *Let* $\vec{a} = (a_1, a_2, ..., a_n)$ *and* $\vec{b} = (b_1, b_2, ..., b_n)$ *be vectors in* $\mathbb{R}^n$. *We define* $\vec{a} + \vec{b}$ *to be the vector in* $\mathbb{R}^n$ *given by*

$$\vec{a} + \vec{b} = (a_1 + b_1, a_2 + b_2, ..., a_n + b_n).$$

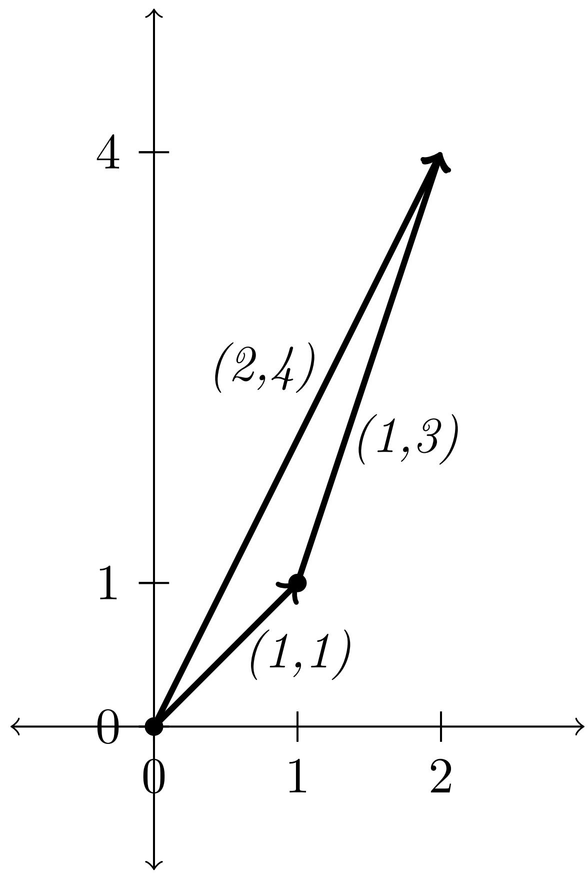
Note that we can only add two vectors if they have the same dimension.

Example 3. *Adding the vectors* $\vec{a} = (1, -8, 2)$ *and* $\vec{b} = (3, -1, -2)$*, we obtain*

$$\begin{aligned} \vec{a} + \vec{b} &= (1, -8, 2) + (3, -1, -2) \\ &= (1 + 3, -8 - 1, 2 - 2) \\ &= (4, -9, 0). \end{aligned}$$

Geometrically, we can add vectors by placing the start point of the second vector at the end point of the first vector, and drawing an arrow from the start point of the first vector to the end point of the second vector.

Example 4. *In this example, we add the vectors* $(1,1)$ *and* $(1,3)$*. Adding these vectors algebraically, we obtain* $(2,4)$*. We can also see this geometrically by placing the start point of the vector* $(1,3)$ *at the end of the vector* $(1,1)$ *(so at the point* $(1,1)$*), and drawing the vector from the origin to the end point of the vector* $(1,3)$*, which is now at* $(2,4)$*.*



Another vector operation is scalar multiplication. Here, we multiply a vector by a real number, possibly changing the length of the vector.

Definition 5. *Let $\vec{a} = (a_1, ..., a_n)$ be a vector in $\mathbb{R}^n$, and let r be a real number (also called a scalar). We define the* scalar product *$r\vec{a}$ to be*

$$r\vec{a} = (ra_1, ..., ra_n).$$

Thus, we see that scalar multiplication is defined by multiplying each component of the vector by the scalar r.

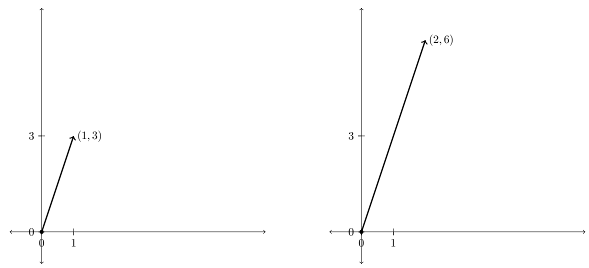
Example 5. $3(1, 5, -2) = (3, 15, -6)$

$-1(1, 1, 1) = (-1, -1, -1)$

$0(6, 2, 4) = (0, 0, 0)$

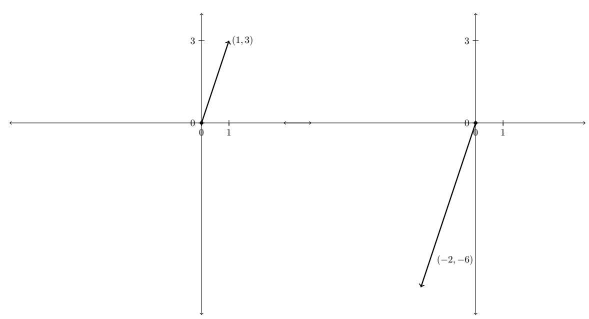
Now, let's look at what scalar multiplication does geometrically. Consider the vector $(1, 3)$.

When we multiply $(1,3)$ by 2, we obtain $(2,6)$, which is twice as long as $(1,3)$ and goes in the same direction.



When we multiply $(1,3)$ by $\dfrac{1}{2}$, we obtain $\left(\dfrac{1}{2}, \dfrac{3}{2}\right)$, which is half as long as $(1,3)$ and goes in the same direction.

If we multiply $(1,3)$ by -2, we obtain $(-2,-6)$, which is twice as long as $(1,3)$ and goes in the exact opposite direction.



Thus, we have seen that multiplying a vector by a scalar changes the length of a vector, but not the direction (except for reversing it, if the scalar is negative).

Properties

Now, let's recall some useful properties of vector addition and scalar multiplication.

Proposition 1. *Suppose $\vec{a}, \vec{b}, \vec{c}$ are vectors in $\mathbb{R}^n$ and k, l are real numbers. Then*

(a) $\vec{a} + \vec{b} = \vec{b} + \vec{a}$ *(vector addition is* commutative*);*

(b) $\vec{a} + (\vec{b} + \vec{c}) = (\vec{a} + \vec{b}) + \vec{c}$ *(vector addition is* associative*);*

(c) $\vec{a} + \vec{0} = \vec{0} + \vec{a} = \vec{a}$*, where* $\vec{0} = (0, ..., 0)$ *is the* zero vector *in* $\mathbb{R}^n$*;*

(d) $(k + l)\vec{a} = k\vec{a} + l\vec{a}$*;*

(e) $k(\vec{a} + \vec{b}) = k\vec{a} + k\vec{b}$ *(with the previous property, scalar multiplication is* distributive*);*

(f) $k(l\vec{a}) = (kl)\vec{a} = l(k\vec{a})$*;*

(g) $1\vec{a} = \vec{a}$*.*

These properties tell us different kinds of algebraic manipulations that we can do with vectors.

Standard Basis Vectors

It's often useful to write things in terms of the standard basis vectors for $\mathbb{R}^n$.

Definition 6. *The vectors* $\vec{e}_1 = (1, 0, ...0)$*,* $\vec{e}_2 = (0, 1, 0, ..., 0)$*, ...,* $\vec{e}_n = (0, ..., 0, 1)$ *in* $\mathbb{R}^n$ *are called the* standard basis vectors *for* $\mathbb{R}^n$*.*

Note that any vector in $\mathbb{R}^n$ can be written uniquely as a linear combination of the standard unit vectors. For example, in $\mathbb{R}^4$,

$$\begin{aligned}(1, 5, -3, 6) &= 1(1, 0, 0, 0) + 5(0, 1, 0, 0) - 3(0, 0, 1, 0) + 6(0, 0, 0, 1)\\ &= 1\vec{e}_1 + 5\vec{e}_2 - 3\vec{e}_3 + 6\vec{e}_4.\end{aligned}$$

In $\mathbb{R}^2$, we sometimes write the standard basis vectors as $\mathbf{i} = (1, 0)$ and $\mathbf{j} = (0, 1)$. This gives us a new notation for vectors, for example we could write

$$(3, 4) = 3\mathbf{i} + 4\mathbf{j}.$$

Similarly, in $\mathbb{R}^3$, we sometimes write the standard basis vectors as $\mathbf{i} = (1, 0, 0)$, $\mathbf{j} = (0, 1, 0)$, and $\mathbf{k} = (0, 0, 1)$. We can then write

$$(2, 3, 4) = 2\mathbf{i} + 3\mathbf{j} + 4\mathbf{k}.$$

Summary

In this section, we reviewed some basics about vectors, including the definition of a vector, basic vector operations, standard basis vectors, notation, and the geometric perspective.

The Dot Product

In this section we review the dot product on vectors. This also includes the angle between vectors and the projection of one vector onto another.

The Dot Product

We begin with the definition of the dot product.

Definition 7. *The* dot product *of two vectors* $\vec{v} = (v_1, v_2, ..., v_n)$ *and* $\vec{w} = (w_1, w_2, ..., w_n)$ *in* $\mathbb{R}^n$ *is*

$$\vec{v} \cdot \vec{w} = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.$$

Notice that the dot product takes two vectors and outputs a scalar.

Example 6. $(1, 6) \cdot (-3, -6) = -3 - 36 = -39$

$(1, 2, 3) \cdot (7, -2, 4) = 7 - 4 + 12 = 15$

$(1, 7, -3) \cdot (3, 0, 1) = 3 + 0 - 3 = 0$

We can also compute the dot product using the magnitude (or length) of the vectors and the angle in between them.

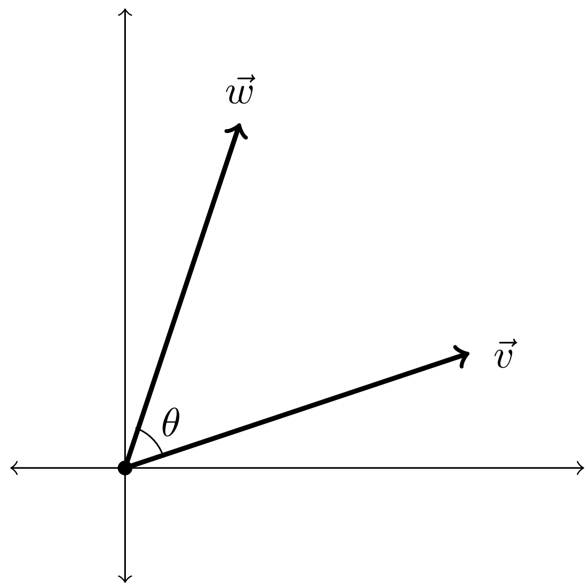
Proposition 2. *If* $\vec{v}$ *and* $\vec{w}$ *are vectors in* $\mathbb{R}^n$*, then*

$$\vec{v} \cdot \vec{w} = \|\vec{v}\| \, \|\vec{w}\| \cos\theta,$$

where $\|\vec{v}\|$ *and* $\|\vec{w}\|$ *are the lengths of the vectors* $\vec{v}$ *and* $\vec{w}$*, respectively, and* θ *is the angle between* $\vec{v}$ *and* $\vec{w}$*.*

This is illustrated in the picture below.

Learning outcomes:
Author(s):



This provides us with a geometric interpretation of the dot product: it gives us a measure of "how much" in the same direction two vectors are (taking their lengths into account). This also gives us a useful way to compute the angle between two vectors.

Example 7. *Consider the vectors* $(1,4)$ *and* $(-2,2)$*. We have*

$$\begin{aligned}(1,4)\cdot(-2,2) &= -2+8=6,\\ \|(1,4)\| &= \sqrt{1^2+4^2}=\sqrt{17},\\ \|(-2,2)\| &= \sqrt{(-2)^2+2^2}=\sqrt{8}.\end{aligned}$$

From $\vec{v}\cdot\vec{w}=\|\vec{v}\|\,\|\vec{w}\|\cos\theta$*, we then have*

$$6=\sqrt{17}\sqrt{8}\cos\theta.$$

Solving for θ*, we obtain the angle between the vectors as*

$$\theta=\arccos\left(\frac{6}{\sqrt{17}\sqrt{8}}\right)\approx 59.04^\circ$$

Furthermore, note that for nonzero vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^n$, their dot product is 0 if and only if $\cos(\theta)=0$. This means that θ would have to be 90° or 270°, meaning that $\vec{v}$ and $\vec{w}$ are perpendicular.

Proposition 3. *Two nonzero vectors* $\vec{v}$ *in* $\vec{w}$ *in* $\mathbb{R}^n$ *are perpendicular if and only if* $\vec{v}\cdot\vec{w}=0$.

This provides us with a very useful algebraic method for determining if two vectors are perpendicular.

Example 8. *The vectors* $(1, 7, -3)$ *and* $(3, 0, 1)$ *in* $\mathbb{R}^3$ *are perpendicular, since*

$$(1, 7, -3) \cdot (3, 0, 1) = 3 + 0 - 3 = 0.$$

By taking the dot product of a vector with itself, we get an important relationship between the dot product and the length of a vector.

Proposition 4. *Let* $\vec{v}$ *be a vector in* $\mathbb{R}^n$*. Then*

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

This can be shown directly, or using the fact that the angle between $\vec{v}$ and itself is 0.

Projection of one vector onto another

We can also use the dot product to define the projection of one vector onto another.

Definition 8. *For vectors* $\vec{a}$ *and* $\vec{b}$ *in* $\mathbb{R}^n$*, we define the* vector projection *of* $\vec{a}$ *onto* $\vec{b}$ *as*

$$proj_{\vec{b}}(\vec{a}) = \frac{\vec{a} \cdot \vec{b}}{\|\vec{b}\|^2}\vec{b} = \frac{\vec{a} \cdot \vec{b}}{\vec{b} \cdot \vec{b}}\vec{b}.$$

Example 9. *We can use this to find the projection of* $(2, 4, 3)$ *onto* $(1, -1, 1)$*.*

$$\begin{aligned} proj_{(1,-1,1)}(2, 4, 3) &= \frac{(2, 4, 3) \cdot (1, -1, 1)}{(1, -1, 1) \cdot (1, -1, 1)}(1, -1, 1) \\ &= \frac{2 - 4 + 3}{1 + 1 + 1}(1, -1, 1) \\ &= \frac{1}{3}(1, -1, 1) \\ &= \left(\frac{1}{3}, -\frac{1}{3}, \frac{1}{3}\right) \end{aligned}$$

Summary

In this section we reviewed the dot product on vectors, the angle between vectors, and the projection of one vector onto another.

The Cross Product

In this section, we review the vector cross product, including the geometric perspective of the cross product, the area of a parallelogram, and the volume of parallelepiped.

The Cross Product

The cross product is fundamentally different from the dot product in a couple of ways. The cross product is defined only on vectors in $\mathbb{R}^3$, while the dot product is defined in $\mathbb{R}^n$ for any positive integer n. Furthermore, the cross product takes two vectors and produces another vector, while the dot product takes two vectors and produces a scalar.

We now give the algebraic definition of the cross product.

Definition 9. *Let $\vec{a} = (a_1, a_2, a_3)$ and $\vec{b} = (b_1, b_2, b_3)$ be vectors in $\mathbb{R}^3$. The* cross product *of $\vec{a}$ and $\vec{b}$, denoted $\vec{a} \times \vec{b}$, is defined to be*

$$\vec{a} \times \vec{b} = \begin{pmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{pmatrix}.$$

Equivalently, we can compute the cross product as

$$\vec{a} \times \vec{b} = det \begin{pmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix},$$

where

$$\mathbf{i} = (1, 0, 0),$$
$$\mathbf{j} = (0, 1, 0),$$
$$\mathbf{k} = (0, 0, 1).$$

Learning outcomes:
Author(s):

Example 10.

$$\begin{aligned}
(3,2,-1)\times(9,0,2) &= \mathit{det}\begin{pmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 3 & 2 & -1 \\ 9 & 0 & 2 \end{pmatrix} \\
&= (2\cdot 2)\mathbf{i} - (0\cdot -1)\mathbf{i} + (-1\cdot 9)\mathbf{j} - (2\cdot 3)\mathbf{j} + (3\cdot 0)\mathbf{k} - (9\cdot -1)\mathbf{k} \\
&= 4\mathbf{i} - 15\mathbf{j} + 9\mathbf{k} \\
&= (4,-15,9)
\end{aligned}$$

The cross product has some nice algebraic properties, which can be very useful.

Proposition 5. *Let $\vec{a}$, $\vec{b}$, and $\vec{c}$ be vectors in $\mathbb{R}^3$, and let $k \in \mathbb{R}$ be a scalar. The cross product has the following properties:*

(a) $\vec{a}\times\vec{b} = -\vec{b}\times\vec{a}$ *(the cross product is* anticommutative*);*

(b) $\vec{a}\times(\vec{b}+\vec{c}) = \vec{a}\times\vec{b}+\vec{a}\times\vec{c}$*;*

(c) $(\vec{a}+\vec{b})\times\vec{c} = \vec{a}\times\vec{c}+\vec{b}\times\vec{c}$ *(with the previous property, the cross product is* distributive *over vector addition);*

(d) $k(\vec{a}\times\vec{b}) = (k\vec{a})\times\vec{b} = \vec{a}\times(k\vec{b})$.

In particular, it's important to remember that the cross product is *not* commutative, so the order of the vectors matters!

Geometry of the Cross Product

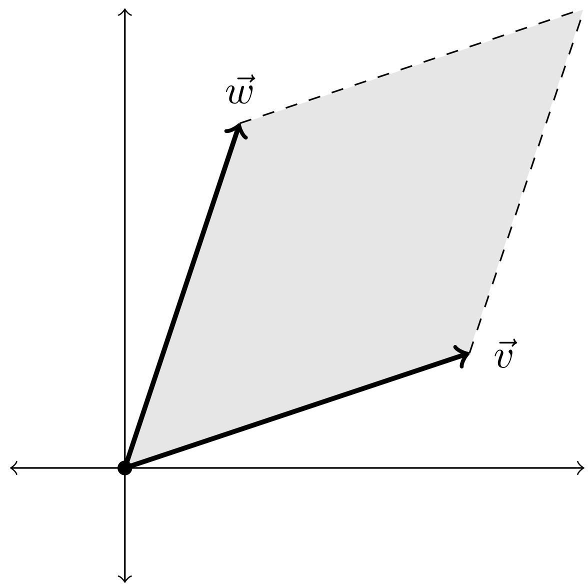
It's often easiest to compute cross products algebraically, but it's easier to understand their significance from a geometric perspective. We now discuss some of the geometric properties of the cross product.

Proposition 6. *Let $\vec{a}$ and $\vec{b}$ be vectors in $\mathbb{R}^3$, and consider their cross product $\vec{a}\times\vec{b}$.*

- *The* magnitude *of the vector $\vec{a}\times\vec{b}$ can be computed as*

  $$\|\vec{a}\times\vec{b}\| = \|\vec{a}\|\,\|\vec{b}\|\sin(\theta),$$

  where θ is the angle between the vectors $\vec{a}$ and $\vec{b}$. Furthermore, this magnitude is equal to the area of the parallelogram determined by $\vec{a}$ and $\vec{b}$.



- *The vector $\vec{a} \times \vec{b}$ is always perpendicular to both $\vec{a}$ and $\vec{b}$, and follows the* right-hand rule. *That is, if you take your right hand and orient it so you can curl your fingers from the vector $\vec{a}$ to the $\vec{b}$, your thumb will be pointing in the same direction as the cross product $\vec{a} \times \vec{b}$.*

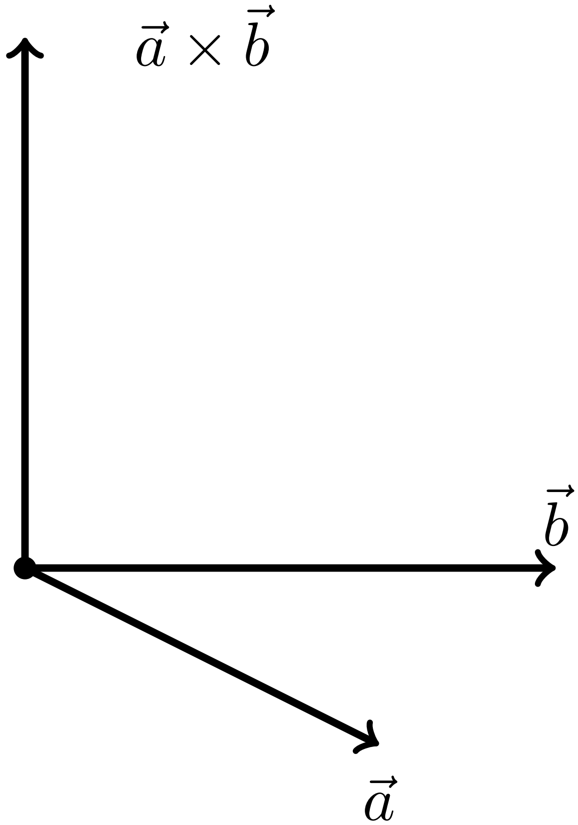
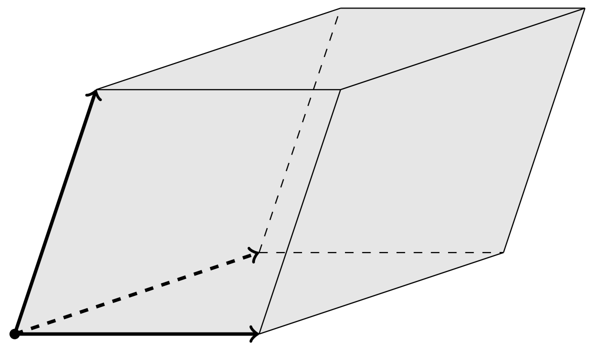


Image this image in $\mathbb{R}^3$, so that $\vec{a} \times \vec{b}$ is perpendicular to both $\vec{a}$ and $\vec{b}$.

Volume of a Parallelepiped

We can use the cross product and dot product together to compute the volume of a parallelepiped.



The volume of the parellelepiped can be computed as the area of the base times the height. We've seen that the area of the base can be computed as the magnitude of a cross product, $\|\vec{a} \times \vec{b}\|$. The height of the parallelepiped can be computed as $\|\vec{c}\| \, |\cos(\theta)|$, where θ is the angle between the vector $\vec{c}$ and a line perpendicular to the base. We then have that the volume is $\|\vec{a} \times \vec{b}\| \, \|\vec{c}\| \, |\cos(\theta)|$, which we can recognize as the absolute value of the dot product of the vectors $\vec{a} \times \vec{b}$ and $\vec{c}$. Thus we have the following proposition.

Proposition 7. *The volume of the parallelepiped determined by the vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$ can be computed as $|(\vec{a} \times \vec{b}) \cdot \vec{c}|$.*

Summary

We've reviewed the cross product, including its properties and geometric perspective, including its use in finding the area of parallelograms and volume of parallelepipeds.

Matrices

In this section, we review matrices, including the determinant and the linear transformation represented by a matrix.

Matrices

We begin with the definition of a matrix.

Definition 10. *An $m \times n$* matrix *A is a rectangular array of numbers a_{ij}, with m rows and n columns:*

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

where the a_{ij} are real numbers for i and j integers with $1 \leq i \leq m$ and $1 \leq j \leq n$.

The numbers a_{ij} are called the entries *of the matrix A.*

Note that for an entry a_{ij}, the subscript ij describes the location of a_{ij} in the matrix A: i gives the row, and j gives the column.

We can also think of a matrix as a "vector of vectors" in two different ways. If we imagine that the columns of A are vectors in $\mathbb{R}^n$, then the matrix of A can be viewed as a vector of column vectors. If we imagine that the rows of A are vectors in $\mathbb{R}^n$, then the matrix A can be viewed as a vector of row vectors.

Matrix Operations

Here, we'll define matrix addition and matrix multiplication.

In order to be able to add two matrices, they need to have the exact same dimensions. That is, they both need to be $m \times n$ matrices for some fixed values of m and n. When we have two matrices with the same dimensions, we define their sum component-wise or entry-wise.

Learning outcomes:
Author(s):

Definition 11. *Let A and B be two $m \times n$ matrices, with*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$
$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{pmatrix}.$$

Then we define the matrix sum *$A + B$ to be*

$$A + B = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn}. \end{pmatrix},$$

That is, $A + B$ is the $m \times n$ matrix obtained by adding the corresponding entries of A and B.

Example 11. *We can add the matrices $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ and $B = \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix}$ as follows:*

$$\begin{aligned} A + B &= \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} + \begin{pmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix} \\ &= \begin{pmatrix} 1+7 & 2+8 & 3+9 \\ 4+10 & 5+11 & 6+12 \end{pmatrix} \\ &= \begin{pmatrix} 8 & 10 & 12 \\ 14 & 16 & 18 \end{pmatrix} \end{aligned}$$

Example 12. *We cannot add the matrices $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ and $B = \begin{pmatrix} 7 & 8 \\ 10 & 11 \end{pmatrix}$, because their dimensions don't match.*

As you might expect, matrix addition has some nice properties which are inherited from addition of real numbers. We list some of them here.

Proposition 8. *Let A, B, and C be $m \times n$ matrices. Then we have*

(a) $A + B = B + A$ *(matrix addition is* commutative*);*

(b) $A + (B + C) = (A + B) + C$ *(matrix addition is* associative*).*

Furthermore, there is an $m \times n$ matrix O, called the zero matrix*, such that $A + O = A$ for any $m \times n$ matrix A. All of the entries of the zero matrix are the real number 0.*

We've seen that matrix addition works in a very natural way, and multiplying a matrix by a scalar (or real number) is similarly nice. We now define scalar multiplication for matrices.

Definition 12. *Let*

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

be an $m \times n$ matrix, and let $k \in \mathbb{R}$ be a scalar. Then the scalar product *of k and A, denoted kA, is*

$$kA = \begin{pmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \vdots & \vdots & & \vdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{pmatrix}.$$

That is, we obtain the scalar product by multiplying each entry in A by the scalar k.

Example 13. *We can compute the scalar product of 2 and the matrix $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$ as follows:*

$$\begin{aligned} 2A &= 2\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \\ &= \begin{pmatrix} 2\cdot 1 & 2\cdot 2 & 2\cdot 3 \\ 2\cdot 4 & 2\cdot 5 & 2\cdot 6 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 4 & 6 \\ 8 & 10 & 12 \end{pmatrix}. \end{aligned}$$

We now list some nice properties of scalar multiplication.

Proposition 9. *Let A and B be $m \times n$ matrices, and let k and l be scalars in $\mathbb{R}$. Then*

(a) $(k+l)A = kA + lA$ *(scalar multiplication is* distributive *over scalar addition);*

(b) $k(A+B) = kA + kB$ *(scalar multiplication is* distributive *over matrix addition);*

(c) $k(lA) = (kl)A = l(kA)$.

We'll now define matrix multiplication, which can be a bit trickier to work with than matrix addition or scalar multiplication. Here are some important things to remember about matrix multiplication:

- Not all matrices can be multiplied. In order to compute the product AB of two matrices A and B, the number of columns in A needs to be the same as the number of rows in B.

- Matrix multiplication is *not* commutative. In fact, its possible that the matrix product AB exists but the product BA does not.

Definition 13. *Let A be an $m \times n$ matrix, and let B be an $n \times p$ matrix, with*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$
$$B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{pmatrix}.$$

Note that we are assuming the number of columns in A is the same as the number of rows in B.

We define the matrix product *of A and B, denoted AB, to be the $m \times p$ matrix*

$$AB = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + \cdots + a_{1n}b_{n1} & a_{11}b_{12} + a_{12}b_{22} + \cdots + a_{1n}b_{n2} & \cdots & a_{11}b_{1p} + a_{12}b_{2p} + \cdots + a_{1n}b_{np} \\ a_{21}b_{11} + a_{22}b_{21} + \cdots + a_{2n}b_{n1} & a_{21}b_{12} + a_{22}b_{22} + \cdots + a_{2n}b_{n2} & \cdots & a_{21}b_{1p} + a_{22}b_{2p} + \cdots + a_{2n}b_{np} \\ \vdots & \vdots & & \vdots \\ a_{m1}b_{11} + a_{m2}b_{21} + \cdots + a_{mn}b_{n1} & a_{m1}b_{12} + a_{m2}b_{22} + \cdots + a_{mn}b_{n2} & \cdots & a_{m1}b_{1p} + a_{m2}b_{2p} + \cdots + a_{mn}b_{np} \end{pmatrix}$$

Equivalently, we could define the ijth entry of AB to be the dot product of the ith row of A with the jth column of B. This makes sense, since the number of columns in A is the same as the number of rows in B (both n), which ensures that the ith row of A and the jth column of B are both vectors in $\mathbb{R}^n$.

This definition can seem a bit convoluted, and it's easier to understand how matrix multiplication works by going through an example.

Example 14. *We can compute the product AB of the matrices $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$*

and $B = \begin{pmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{pmatrix}$ *as follows:*

$$\begin{aligned}
AB &= \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{pmatrix}, \\
&= \begin{pmatrix} 1\cdot 7 + 2\cdot 9 + 3\cdot 11 & 1\cdot 8 + 2\cdot 10 + 3\cdot 12 \\ 4\cdot 7 + 5\cdot 9 + 6\cdot 11 & 4\cdot 8 + 5\cdot 10 + 6\cdot 12 \end{pmatrix}, \\
&= \begin{pmatrix} 58 & 64 \\ 139 & 154 \end{pmatrix}.
\end{aligned}$$

Note that we multiplied a 2×3 *matrix by a* 3×2 *matrix, and we obtained a* 2×2 *matrix.*

We can also compute the product BA *for the same matrices as above.*

$$\begin{aligned}
BA &= \begin{pmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \\
&= \begin{pmatrix} 7\cdot 1 + 8\cdot 4 & 7\cdot 2 + 8\cdot 5 & 7\cdot 3 + 8\cdot 6 \\ 9\cdot 1 + 10\cdot 4 & 9\cdot 2 + 10\cdot 5 & 9\cdot 3 + 10\cdot 6 \\ 11\cdot 1 + 12\cdot 4 & 11\cdot 2 + 12\cdot 5 & 11\cdot 3 + 12\cdot 6 \end{pmatrix}, \\
&= \begin{pmatrix} 39 & 54 & 69 \\ 49 & 68 & 81 \\ 59 & 82 & 105 \end{pmatrix}.
\end{aligned}$$

Note that we multiplied a 3×2 *matrix by a* 2×3 *matrix, and we obtained a* 3×3 *matrix.*

Note that in this case $AB \neq BA$*; matrix multiplication is not commutative, so the order of the matrices matters!*

Although matrix multiplication is not commutative, it still has some nice algebraic properties. We list some of them here.

Proposition 10. *Let* A*,* B*, and* C *be matrices of dimensions such that the following operations are defined, and let* k *be a scalar. Then*

(a) $A(BC) = (AB)C$ *(matrix multiplication is* associative*);*

(b) $k(AB) = (kA)B = A(kB)$*;*

(c) $A(B + C) = AB + AC$*;*

(d) $(A + B)C = AC + BC$ *(with the previous property, matrix multiplication is* distributive *over matrix addition).*

Determinants

When we have a square matrix (meaning an $n \times n$ matrix, where the number of rows and number of columns are the same), we can compute an important number, called the determinant of the matrix. It turns out that this single number can tell us some important things about the matrix!

We begin by defining the determinant of a 2×2 matrix.

Definition 14. *Consider the 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We define the* determinant *of the matrix A to be*

$$det(A) = ad - bc.$$

We also sometimes use the notation $|A|$ for the determinant of the matrix A.

Note that the determinate of a 2×2 matrix is just a number, not a matrix. We compute the determinant in a couple of examples.

Example 15. *We'll compute the determinant of the matrix $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.*

$$\begin{aligned} det(A) &= 1 \cdot 4 - 2 \cdot 3 \\ &= -2. \end{aligned}$$

We've defined the determinant of 2×2 matrices, but we haven't defined the determinant of a larger square matrix yet. It turns out that the determinant is defined *inductively*. This means that the determinant of a 3×3 matrix is defined using determinants of 2×2 matrices, the determinant of a 4×4 matrix is defined using determinants of 3×3 matrices, the determinant of a 5×5 matrix is defined using determinants of 4×4 matrices, and so on. This means in order to compute the determinant of a large square matrix, we often need to compute the determinants of many smaller matrices.

We now give the definition of the determinant of an $n \times n$ matrix.

Definition 15. *Let A be an $n \times n$ matrix, with entries a_{ij}. We defined the* determinant *of A to be the number computed by*

$$det(A) = (-1)^{1+1} a_{11} det(A_{11}) + (-1)^{1+2} a_{12} det(A_{12}) + \cdots + (-1)^{1+n} a_{1n} det(A_{1n}),$$

where A_{ij} is the $(n-1) \times (n-1)$ submatrix of A which we obtain by deleting the ith row and $j$the column from A.

This definition is pretty confusing if you read through it without seeing an example, but this actually follows a nice pattern. This pattern is easier to see with an example.

Example 16. *We compute the determinant of the 4×4 matrix,*

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

Note that we begin by writing this in terms of determinants of 3×3 matrices. But in order to compute the determinant of each 3×3 matrix, we write it in terms of 2×2 matrices! This winds up being a lot of determinants to compute.

$$\begin{aligned} det(A) = (-1)^{1+1} 1 det \begin{pmatrix} 0 & -2 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + (-1)^{1+2} 4 det \begin{pmatrix} 0 & -2 & 1 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ + (-1)^{1+3} 2 det \begin{pmatrix} 0 & 0 & 1 \\ -3 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} + (-1)^{1+4}(-1) det \begin{pmatrix} 0 & 0 & -2 \\ -3 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \end{aligned}$$

We now compute the determinant of each of the 3×3 submatrices, which we *compute using determinants of* 2×2 *matrices.*

$$\begin{aligned}
det\begin{pmatrix} 0 & -2 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} &= (-1)^{1+1}0 det\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (-1)^{1+2}(-2) det\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \\
&\qquad + (-1)^{1+3}1 det\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \\
&= 1 \cdot 0 \cdot (1 \cdot 1 - 0 \cdot 0) + -1 \cdot (-2) \cdot (0 \cdot 1 - 0 \cdot 0) + 1 \cdot 1 \cdot (0 \cdot 0 - 1 \cdot 0) \\
&= 0 \\
det\begin{pmatrix} 0 & -2 & 1 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} &= (-1)^{1+1}0 det\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (-1)^{1+2}(-2) det\begin{pmatrix} -3 & 0 \\ 0 & 1 \end{pmatrix} \\
&\qquad + (-1)^{1+3}1 det\begin{pmatrix} -3 & 1 \\ 0 & 0 \end{pmatrix} \\
&= 1 \cdot 0 \cdot (1 \cdot 1 - 0 \cdot 0) + -1 \cdot (-2) \cdot (-3 \cdot 1 - 0 \cdot 0) + 1 \cdot 1 \cdot (-3 \cdot 0 - 1 \cdot 0) \\
&= 6 \\
det\begin{pmatrix} 0 & 0 & 1 \\ -3 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} &= (-1)^{1+1}0 det\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + (-1)^{1+2}0 det\begin{pmatrix} -3 & 0 \\ 0 & 1 \end{pmatrix} \\
&\qquad + (-1)^{1+3}1 det\begin{pmatrix} -3 & 0 \\ 0 & 0 \end{pmatrix} \\
&= 1 \cdot 0 \cdot (0 \cdot 1 - 0 \cdot 0) + -1 \cdot 0 \cdot (-3 \cdot 1 - 0 \cdot 0) + 1 \cdot 1 \cdot (-3 \cdot 0 - 0 \cdot 0) \\
&= 0 \\
det\begin{pmatrix} 0 & 0 & -2 \\ -3 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} &= (-1)^{1+1}0 det\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} + (-1)^{1+2}0 det\begin{pmatrix} -3 & 1 \\ 0 & 0 \end{pmatrix} \\
&\qquad + (-1)^{1+3}(-2) det\begin{pmatrix} -3 & 0 \\ 0 & 0 \end{pmatrix} \\
&= 1 \cdot 0 \cdot (0 \cdot 0 - 1 \cdot 0) + -1 \cdot 0 \cdot (-3 \cdot 0 - 1 \cdot 0) + 1 \cdot (-2) \cdot (-3 \cdot 0 - 0 \cdot 0) \\
&= 0
\end{aligned}$$

Substituting these in to our computation of the determinant of A*, we then obtain*

$$\begin{aligned}
det(A) &= 1 \cdot 1 \cdot 0 + (-1) \cdot 4 \cdot (6) + 1 \cdot 2 \cdot 0 + (-1) \cdot (-1) \cdot 0 \\
&= -24.
\end{aligned}$$

We sometimes call this method of computing a determinant as "expanding along the first row." This is because we can also compute the determinant of a matrix by similarly expanding along a different row, or even a column.

Proposition 11. *We can similarly compute the determinant of an* $n \times n$ *matrix*

A by expanding along any row or column. Expanding along the ith row, we have

$$det(A) = (-1)^{i+1} a_{i1}\, det(A_{i1}) + (-1)^{i+2} a_{i2}\, det(A_{i2}) + \cdots + (-1)^{i+n} a_{in}\, det(A_{in}).$$

Expanding along the jth column, we have

$$det(A) = (-1)^{1+j} a_{1j}\, det(A_{1j}) + (-1)^{2+j} a_{2j}\, det(A_{2j}) + \cdots + (-1)^{n+j} a_{nj}\, det(A_{nj}).$$

Once again, A_{ij} is the $(n-1) \times (n-1)$ submatrix obtained from A by deleting the ith row and jth column.

It can be useful to think about which row or column will be easiest to expand along. In particular, choosing a row or column with a lot of zeros greatly simplifies computation.

Example 17. *We'll once again compute the determinant of the 4×4 matrix*

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

this time by expanding along the second column. Note that this column is a good choice, since there's only one nonzero element. We have

$$det(A) = (-1)^{1+2}(4)\, det \begin{pmatrix} 0 & -2 & 1 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + (-1)^{2+2}(0)\, det \begin{pmatrix} 1 & 2 & -1 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + (-1)^{3+2}(0)\, det \begin{pmatrix} 1 & 2 & -1 \\ 0 & -2 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

We'll only compute the determinant of the submatrix $\begin{pmatrix} 0 & -2 & 1 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$; we won't bother computing the others since their determinants will be multiplied by 0.

$$\begin{aligned} det \begin{pmatrix} 0 & -2 & 1 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} &= (-1)^{3+1}(0)\, det \begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix} + (-1)^{3+2}(0)\, det \begin{pmatrix} 0 & 1 \\ -3 & 0 \end{pmatrix} + (-1)^{3+3}(1)\, det \begin{pmatrix} 0 & -2 \\ -3 & 1 \end{pmatrix} \\ &= 0 + 0 + (1)(1)(0 \cdot 1 - (-2) \cdot (-3)), \\ &= -6. \end{aligned}$$

Once again, we don't bother computing the determinants which will be multiplied by zero. Note that we chose to expand across the last row, since it had two zeroes. Expanding along the first column would also have been a reasonable choice.

Returning to our computation of the determinant of A, we have

$$\begin{aligned} det(A) &= (-1)^{1+2}(4)\, det \begin{pmatrix} 0 & -2 & 1 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + (-1)^{2+2}(0)\, det \begin{pmatrix} 1 & 2 & -1 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + (-1)^{3+2}(0)\, det \begin{pmatrix} 1 & 2 & -1 \\ 0 & -2 & 1 \\ 0 & 0 & 1 \end{pmatrix} \\ &= (-1)(4)(-6) + 0 + 0 + 0, \\ &= 24. \end{aligned}$$

Notice that this matching our previous computation, expanding along the first row.

One of the most powerful uses of the determinant is to tell us whether or not a matrix is invertible. Recall that an $n \times n$ matrix A is *invertible* if there is a matrix B such that $AB = BA = I_n$, where I_n is the $n \times n$ identity matrix.

Proposition 12. *An $n \times n$ matrix A is invertible if and only if its determinant is nonzero.*

This gives us a convenient way to test if a matrix is invertible, without needing to produce an explicit inverse.

Example 18. *We found that the determinant of the matrix*

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

is 24. *Since this is nonzero, the matrix A is invertible.*

On the other hand, you can verify that the determinant of the matrix

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

is 0. *Thus, the matrix B is not invertible.*

Linear Transformations

One of the most important uses of matrices is to represent linear transformations. Recall the definition of a linear transformation.

Definition 16. *A function T from $\mathbb{R}^n$ to $\mathbb{R}^n$ is a linear transformation if for all vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^n$ and all scalars $k \in \mathbb{R}$, we have*

(a) $T(\vec{v} + \vec{w}) = T(\vec{v}) + T(\vec{w})$;

(b) $T(k\vec{v}) = kT(\vec{v})$.

We can view an $m \times n$ matrix A as representing a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ as follows. We write vectors as column vectors, or, equivalently, $n \times 1$ or $m \times 1$ matrices. For an input column vector $\vec{v}$ in $\mathbb{R}^n$, we multiply $\vec{v}$ by A

on the left, using matrix multiplication. This produces an $m \times 1$ matrix, or, equivalently, a column vector in $\mathbb{R}^m$. Thus, we can define a function

$$T_A(\vec{v}) = A\vec{v}.$$

Using properties of matrix multiplication, we have that this is a linear transformation. Thus, we have the linear transformation associated to a matrix.

Example 19. *Consider the linear transformation T_A from $\mathbb{R}^3$ to $\mathbb{R}^2$ corresponding to the 2×3 matrix*

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

Let investigate the images of several vectors in $\mathbb{R}^3$ under the linear transformation T_A.

$$\begin{aligned} T_A((1,2,3)) &= \begin{pmatrix} 1 & -5 & 3 \\ 2 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot 1 + -5 \cdot 2 + 3 \cdot 3 \\ 2 \cdot 1 + 0 \cdot 2 + -1 \cdot 3 \end{pmatrix} \\ &= \begin{pmatrix} 0 \\ -1 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} T_A((1,-1,2)) &= \begin{pmatrix} 1 & -5 & 3 \\ 2 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot 1 + -5 \cdot -1 + 3 \cdot 1 \\ 2 \cdot 1 + 0 \cdot -1 + -1 \cdot 1 \end{pmatrix} \\ &= \begin{pmatrix} 9 \\ 1 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} T_A((1,0,0)) &= \begin{pmatrix} 1 & -5 & 3 \\ 2 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot 1 + -5 \cdot 0 + 3 \cdot 0 \\ 2 \cdot 1 + 0 \cdot 0 + -1 \cdot 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 2 \end{pmatrix} \end{aligned}$$

$$\begin{aligned}
T_A((0,1,0)) &= \begin{pmatrix} 1 & -5 & 3 \\ 2 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 \cdot 0 + -5 \cdot 1 + 3 \cdot 0 \\ 2 \cdot 0 + 0 \cdot 1 + -1 \cdot 0 \end{pmatrix} \\
&= \begin{pmatrix} -5 \\ 0 \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
T_A((0,0,1)) &= \begin{pmatrix} 1 & -5 & 3 \\ 2 & 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\
&= \begin{pmatrix} 1 \cdot 0 + -5 \cdot 0 + 3 \cdot 1 \\ 2 \cdot 0 + 0 \cdot 0 + -1 \cdot 1 \end{pmatrix} \\
&= \begin{pmatrix} 3 \\ -1 \end{pmatrix}
\end{aligned}$$

Notice that when we apply the linear transformation to the standard unit vectors $\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$, we obtain the columns of A as the output vector. This observation can be used to reconstruct a matrix from a given linear transformation.

Proposition 13. *Given any linear transformation T from $\mathbb{R}^n$ to $\mathbb{R}^m$, there is an $m \times n$ matrix such that $T = T_A$.*

Furthermore, the columns of A can be obtained by applying T to the standard unit vectors. More specifically, the jth column of A is given by $T(\vec{e}_j)$.

We can see how this is useful through an example.

Example 20. *Consider the linear transformation T from $\mathbb{R}^2$ to $\mathbb{R}^2$ that rotates a vector by 30° counterclockwise. We can see geometrically that, for the standard unit vectors $\vec{e}_1$ and $\vec{e}_2$ in $\mathbb{R}^2$, we have*

$$\begin{aligned}
T((1,0)) &= \left(\frac{\sqrt{3}}{2}, \frac{1}{2} \right), \\
T((0,1)) &= \left(-\frac{1}{2}, \frac{\sqrt{3}}{2} \right).
\end{aligned}$$

These tell us the columns of the matrix corresponding to the linear transformation, so we then know that the rotation can be represented by the matrix

$$A = \begin{pmatrix} \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}.$$

Summary

In this section, we reviewed matrix operations and properties, determinants, and linear transformations.

Although we've reviewed some of the most important concepts from linear algebra, there is still a lot of material that we weren't able to include here. Make sure you refer back to your linear algebra textbook if there's anything else you need to review!

Representations of Lines and Planes

In this section, we review the different ways we can represent lines and planes, including parametric representations.

Representations of Lines

When you think of describing a line algebraically, you might think of the standard form

$$y = mx + b,$$

where m is the slope and b is the y-intercept. This is often called *slope-intercept* form.

In addition to slope-intercept form, there are several other ways to represent lines. For example, you may remember using *point-slope* form in single variable calculus. We can describe a line of slope m going through a point (x_0, y_0) with the equation

$$y - y_0 = m(x - x_0).$$

It's important to note that there are many different possible choices for the point (x_0, y_0). Because of this, unlike slope-intercept form, point-slope form does not give a unique representation of a line.

In linear algebra, we saw that we could parametrize a line using a vector $\vec{v} = (v_1, v_2)$ giving the direction of the line, and a point (x_0, y_0) that the line passes through. We parametrize the line as

$$\begin{aligned} \vec{x}(t) &= (v_1, v_2)t + (x_0, y_0), \\ &= (v_1 t + x_0, v_2 t + y_0). \end{aligned}$$

Note that this representation works a bit differently from the previous two representations. In slope-intercept form and point-slope form, the line was the set of points (x, y) satisfying the given equation. However, in the parametrization, we plug in values for the parameter t in order to get points on the line.

Unlike slope-intercept form and point-slope form, the parametrization of a line can easily be generalized to three or more dimensions. That is, a line in $\mathbb{R}^n$ through the point $\vec{a}$ and in the direction of the vector $\vec{v}$ can be parametrized as

$$\vec{x}(t) = \vec{v}t + \vec{a},$$

Learning outcomes:
Author(s):

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

If we would like to describe a line in higher dimensions using equations (rather than a parametrization), we would need more than one equation. For example, in $\mathbb{R}^3$, we would require two equations to determine a line.

Representations of Planes

We also have multiple ways to represent planes. Here, we'll focus on planes in $\mathbb{R}^3$.

Recall that a plane can be determined by two vectors (giving the "direction" of the plane) and a point that the plane passes through. We can use this to give a parametrization for the plane through the point $\vec{a}$ and parallel to the vectors $\vec{v}$ and $\vec{w}$:

$$\vec{x}(s,t) = \vec{v}s + \vec{w}t + \vec{a},$$

for s and t in $\mathbb{R}$. Note that we require two parameters for the parametrization of the plane.

We can also describe a plane using a single linear equation in x, y, and z. For example,

$$2x + 4y - z = 9$$

defines a plane. A standard way to do this is using a point on the plane and a normal vector to the plane. Recall that a normal vector is perpendicular to every vector in the plane. If $\vec{n} = (n_1, n_2, n_3)$ is a normal vector to a plane passing through the point $\vec{a} = (a_1, a_2, a_3)$, the plane is defined by the equation

$$\vec{n} \cdot (\vec{x} - \vec{a}) = 0.$$

This can be rewritten as

$$n_1(x - a_1) + n_2(y - a_2) + n_3(z - a_3) = 0.$$

Summary

We reviewed various representations of lines and planes, including parametrizations.

Part II

Coordinate Systems and Functions

Review of Coordinate Systems

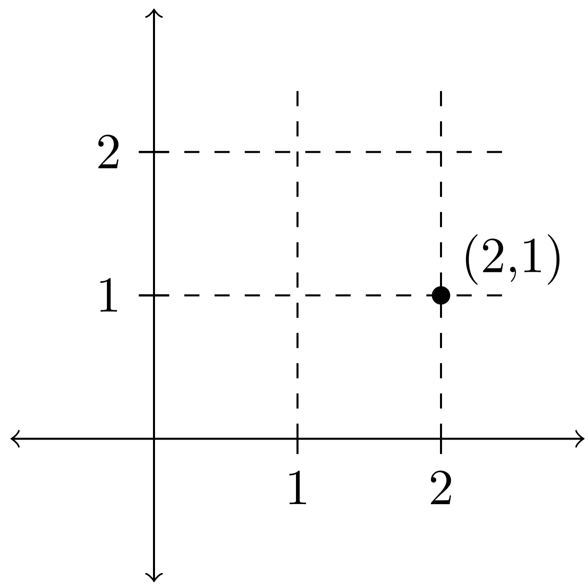
In this activity, we review coordinate systems that you've seen before, in preparation for introducing new coordinate systems in subsequent sections.

Cartesian Plane

The coordinates that you're probably most comfortable with are standard two-dimensional coordinates, also called Cartesian coordinate system on the plane.

In Cartesian coordinates, we describe a point using an x-coordinate and a y-coordinate. We write a point as (x, y), where the x-coordinate describes the horizontal displacement of the point, and the y-coordinate describes the vertical displacement of the point.

Learning outcomes:
Author(s):



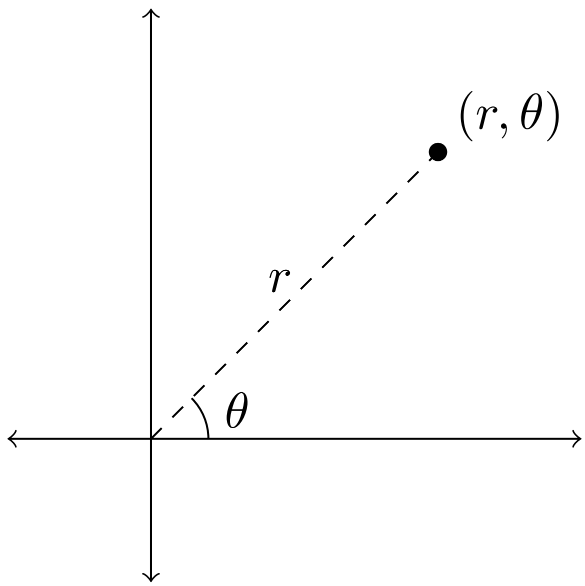
Polar Coordinates

You've also seen polar coordinates.

In polar coordinates, we describe a point with an r-coordinate and a θ-coordinate. The r coordinate gives the distance between the point and the origin, and the θ-coordinate gives the angle (in radians) between the positive x-axis and the segment connecting the origin and the point.

We can switch between cartesian and polar coordinates using the equations

$$x = r\cos\theta$$
$$y = r\sin\theta$$



Problem 1 *Write the point $(r, \theta) = (5, \pi/3)$ in cartesian coordinates.*

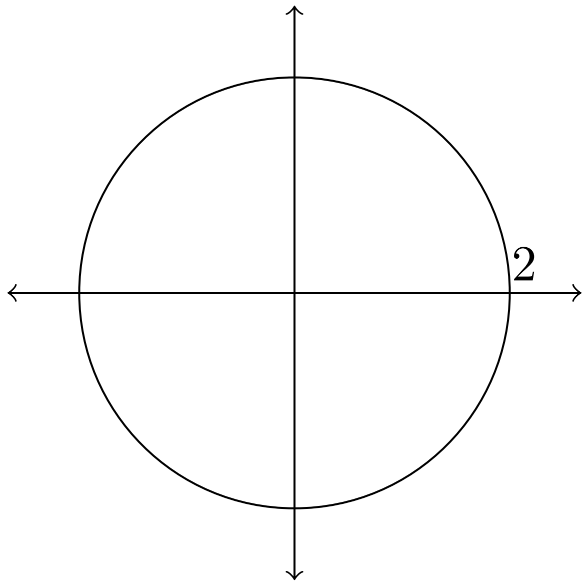
$$(x, y) = \boxed{(5/2, 5\sqrt{3}/2)}$$

Write the point $(x, y) = (-2, 2)$ in polar coordinates.

$$(r, \theta) = \boxed{(\sqrt{8}, 3\pi/4)}$$

Example 21. *Recall that we can describe a circle of radius 2 using Cartesian points as the set of points (x, y) satisfying*

$$x^2 + y^2 = 4.$$



We would like to write describe this circle using polar coordinates.

By definition, the circle of radius 2 *centered at the origin consists of the points which are distance* 2 *from the origin. Because of this, for any point on the circle, we have*

$$r = \boxed{2}.$$

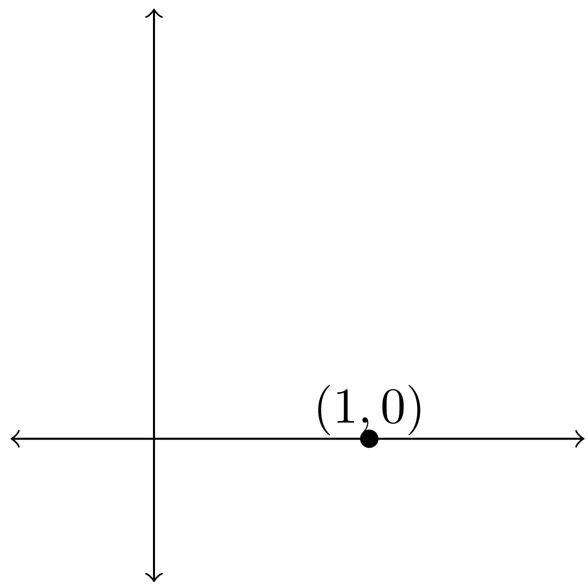
There are points on the circle making every possible angle with the positive x*-axis, so we don't need any restrictions on* θ*. If, however, we only wanted part of the circle, we would accomplish this by restricting* θ*. For example, to get the top half of the circle, we would make the restriction* $0 \leq \theta \leq \pi$*.*

Thus, in polar coordinates, the circle of radius 2 *centered at the origin can be described as the set of points* (r, θ) *such that*

$$r = 2.$$

There's an important difference between Cartesian coordinates and polar coordinates: Cartesian coordinates are *unique*, while polar coordinates are not. This means that, given a point P in the plane, there's only one way to describe this point as (x, y) using Cartesian coordinates. However, there are many ways to write the point as (r, θ), using polar coordinates.

Take, for example, the point $(1, 0)$, written in Cartesian coordinates.



This point is on the x-axis and is distance 1 from the origin. Thus, perhaps the most obvious way to represent this point in polar coordinates is as $(r, \theta) = (1, 0)$ (coincidentally, the same as in Cartesian coordinates). But we could also describe the angle as 2π, 4π, -2π, etc. So, we could also write the point in polar coordinates as $(r, \theta) = (1, 2\pi)$, and so on.

Perhaps more surprisingly, we can describe this point as $(-1, \pi)$. Imagine making an angle of π with the positive x-axis (so we're on the negative x-axis), then going backwards past the origin. This also gets you to our point. Using equivalent angles, we can also represent the point as $(-1, 3\pi)$, $(-1, -\pi)$, and so on.

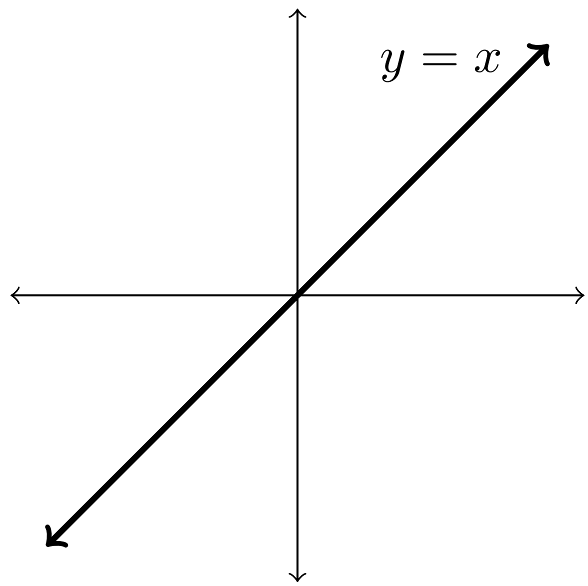
There are some times in working with polar coordinates when we would like to be able to represent points uniquely, and in these situations, we often make restrictions

$$0 \leq r,$$
$$0 \leq \theta < 2\pi.$$

However, even with these restrictions, there still is a point that has multiple representations! Namely, the origin can be written as $(r, \theta) = (0, \theta)$ for any angle θ.

Depending on the situation and context, different people may use different restrictions or conventions for their ranges for r and θ. For this reason, it's good to specify what values you're allowing, to avoid being misunderstood!

Example 22. *Let's consider the line described in Cartesian coordinates as the set of points (x, y) such that $y = x$. We'll figure out how to describe this line in polar coordinates.*



Let's restrict our polar coordinates to $0 \leq r$ *and* $0 \leq \theta < 2\pi$. *Perhaps your first guess is to describe the line as the points* (r, θ) *such that*

$$\theta = \pi/4.$$

Which shape does this describe?

Multiple Choice:

(a) *A point.*

(b) *Half of the line.* ✓

(c) *The whole line.*

(d) *A different line.*

(e) *A circle.*

Describing the line as $\theta = \pi/4$ *is a reasonable first guess, as we can see that many of the points make an angle* $\pi/4$ *with the positive* x*-axis. However, with the restriction that* $r \geq 0$, *this leaves out half of the line! In order to describe the entire line, we have a couple of options. One option would be to relax our restriction on* r, *and allow negative values as well. This would certainly give us the whole line. If, however, we would like to keep this restriction that* $r \geq 0$, *we could also include points with* $\theta = 5\pi/4$, *which will give us the other half of the line.*

Which of the following describe the line $y = x$ *in polar coordinates? Select all that work.*

Select All Correct Answers:

(a) *The points* (r, θ) *such that* $\theta = \pi/4$, *where* $r \geq 0$.

(b) *The points* (r, θ) *such that* $\theta = \pi/4$, *where* r *can be any real number.* ✓

(c) *The points* (r, θ) *such that* $\theta = \pi/4$ *or* $\theta = -\pi/4$, *where* $r \geq 0$.

(d) *The points* (r, θ) *such that* $\theta = \pi/4$ *or* $\theta = -\pi/4$, *where* r *can be any real number.*

(e) *The points* (r, θ) *such that* $\theta = \pi/4$ *or* $\theta = 5\pi/4$, *where* $r \geq 0$. ✓

(f) *The points* (r, θ) *such that* $\theta = \pi/4$ *or* $\theta = 5\pi/4$, *where* r *can be any real number.* ✓

Recall that the relationship between Cartesian and polar coordinates:

$$\begin{aligned} x &= \boxed{r\cos\theta}, \\ y &= \boxed{r\sin\theta}. \end{aligned}$$

Recall the following equations describing the relationship between Cartesian and polar coordinates, which can be useful for converting between these two coordinate systems.

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ r^2 &= x^2 + y^2 \\ \tan\theta &= \frac{y}{x} \end{aligned}$$

Example 23. *Consider the set of points* (r, θ) *such that* $r = 2\cos\theta$. *What does this set of points look like?*

It's not very clear from $r = 2\cos\theta$ *what shape this is describing, so let's try converting this to Cartesian coordinates, and see if we get something we recognize.*

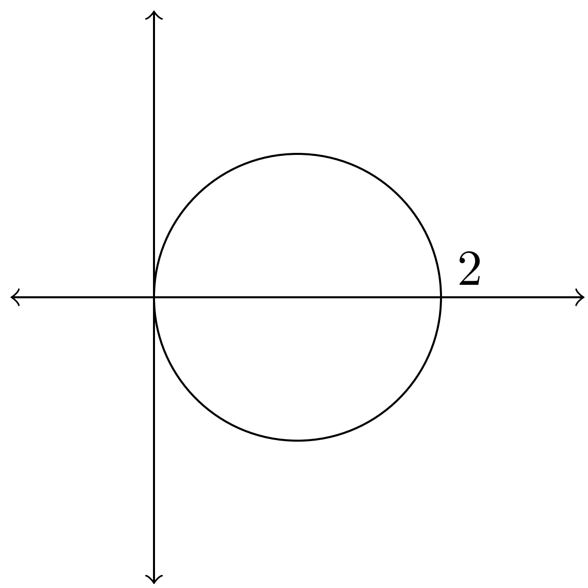
Using the conversion equations above, we have that $r^2 = \boxed{x^2 + y^2}$, *and* $\cos\theta = \frac{x}{r}$. *Making substitutions using these facts, we have:*

$$\begin{aligned} r &= 2\cos\theta \\ r &= 2\frac{x}{r} \\ r^2 &= 2x \\ x^2 + y^2 &= 2x \end{aligned}$$

We now have an equation solely in terms of x and y, but maybe it isn't quite recognizable yet. But if we do a bit more algebra...

$$\begin{aligned} x^2 + y^2 &= 2x \\ (x^2 - 2x + 1) + y^2 &= 1 \\ (x-1)^2 + y^2 &= 1 \end{aligned}$$

Now, we can see that this is a circle of radius $\boxed{1}$ *centered at* $\boxed{(1,0)}$.



Linear Change of Coordinates

In Linear Algebra, we saw how different coordinate systems arose through linear change of coordinates. You may remember this referred to as "slanty space."

When we write a point in Cartesian coordinates as (x, y), we can think of this as a linear combination of the standard basis vectors:

$$(x, y) = x(1, 0) + y(0, 1).$$

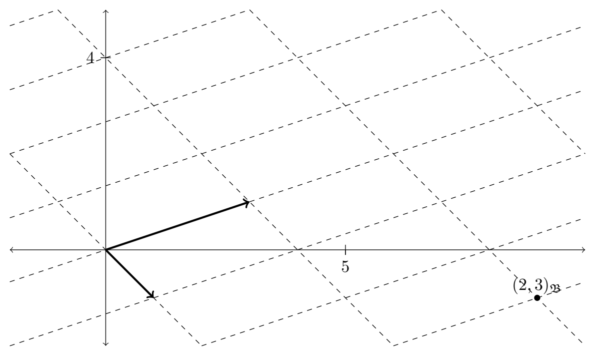
Of course, we can just as well write a point as a linear combination of vectors from a different basis, say $(3, 1)$ and $(1, -1)$. Let's call this basis $\mathfrak{B}$. For example, we can write the vector $(9, -1)$ as

$$(9, -1) = 2(3, 1) + 3(1, -1).$$

Taking the coefficients, in $\mathfrak{B}$-coordinates, we would write this point as

$$(2,3)_{\mathfrak{B}}.$$

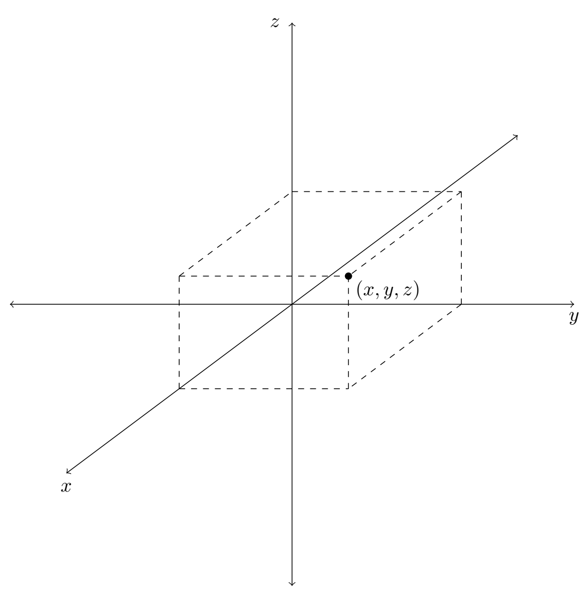
Note that we write $\mathfrak{B}$ in the subscript, in order to remind us that these are $\mathfrak{B}$-coordinates, rather than standard Cartesian coordinates.



With linear changes of coordinates, it's easy to make a mistake and forget which coordinates you're using. Make sure to keep careful track!

Three-Dimensional Coordinates

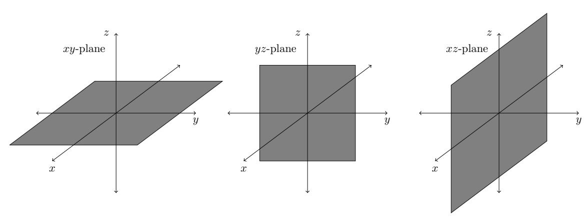
In Linear Algebra, we also worked in three-dimensional Cartesian coordinates, (x,y,z) in $\mathbb{R}^3$.



It's important to remember that the x, y, and z axes follow the right hand rule. That is, if you take your right hand, and point your pointer finger in the direction of the positive x-axis, point your middle finger in the direction of the positive y-axis, then your thumb points in the direction of the positive z-axis.

Another way to say this is that if you point the fingers of your right hand in the direction of the positive x-axis and curl them to point in the direction of the positive y-axis, your thumb points in the direction of the positive z-axis.

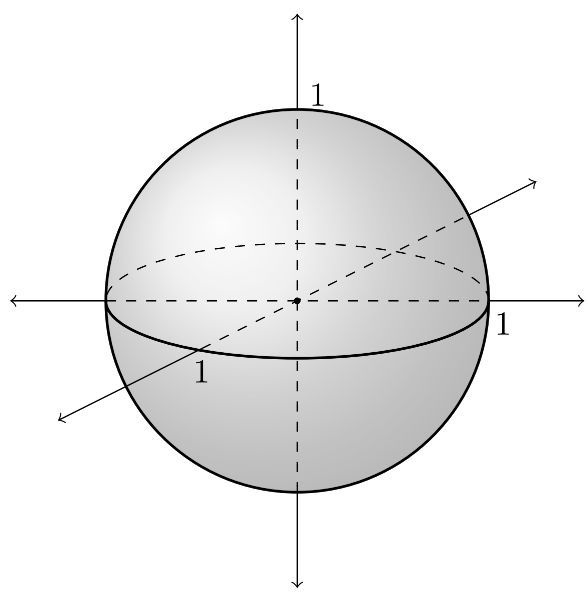
We'll often refer to the *coordinate planes* in $\mathbb{R}^3$. These are the three planes we obtain by setting each of the coordinates to be zero.



More precisely, the xy-plane is the set of points (x, y, z) such that $z = 0$, the yz-plane is the set of points such that $x = 0$, and the xz-plane is the set of points such that $y = 0$.

Similarly to in the plane, we can describe sets of points in $\mathbb{R}^3$ using equations.

Example 24. *The set of points (x, y, z) such that $x^2 + y^2 + z^2 = 1$ is the sphere of radius 1 centered at the origin in $\mathbb{R}^3$.*



Conclusion

In this activity, we reviewed coordinate systems that you've seen before: standard two-dimensional coordinates, polar coordinates, coordinates with respect to a given set of basis vectors, and three-dimensional coordinates.

Cylindrical Coordinates

In this activity, we introduce cylindrical coordinates, a new coordinate system on $\mathbb{R}^3$. We also discuss how to convert between cylindrical and Cartesian coordinates.

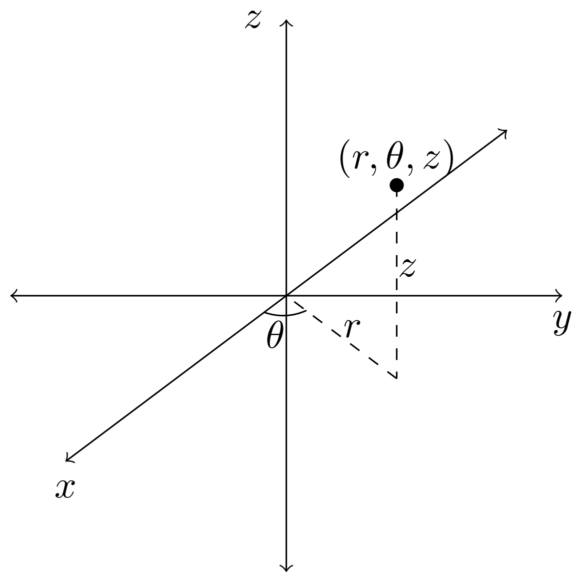
Cylindrical Coordinates

We've seen how points in $\mathbb{R}^2$ can be written using polar coordinates. Polar coordinates can be useful for describing shapes that are difficult to describe in Cartesian coordinates.

We'd now like to extend this idea to $\mathbb{R}^3$, using a coordinate system called *cylindrical coordinates*. Like polar coordinates, cylindrical coordinates will be useful for describing shapes in $\mathbb{R}^3$ that are difficult to describe using Cartesian coordinates. Later in the course, we will also see how cylindrical coordinates can be useful in multivariable Calculus, when differentiating or integrating in Cartesian coordinates is difficult or impossible.

Cylindrical coordinates are really just a simple extension of polar coordinates. For points in the xy-plane, we describe them using r and θ, where r is the distance from the origin and θ is the angle with the positive x-axis. We then tack on a z-coordinate, the exact same as the z-coordinate in Cartesian coordinates, which tells us the vertical displacement of the point.

Learning outcomes:
Author(s):



Example 25. *We'll convert the point $(x, y, z) = (1, 1, 1)$ to cylindrical coordinates.*

We can figure out r and θ by just considering the x- and y-coordinates of the point, $(1, 1)$. Then this becomes equivalent to representing the point in polar coordinates, so we have

$$(r, \theta) = \boxed{(\sqrt{2}, \pi/4)}.$$

For last coordinate, z, notice that this is telling us the height of the point, which is the exact same as the z-coordinate of the point written in Cartesian coordinates! So, our z coordinate is $\boxed{1}$, and the point $(x, y, z) = (1, 1, 1)$ can be written in cylinderical coordinates as

$$(r, \theta, z) = \boxed{(\sqrt{2}, \pi/4, 1)}.$$

You may use the applet below to experiment with how changing the different coordinates changes the point given in cylindrical coordinates.

MATH INSIGHT APPLET

Uniqueness

When we studied polar coordinates, we saw that there were many different ways to represent a point. For example, the point $(x, y) = (0, 1)$ could be written as $(r, \theta) = (1, \pi/2)$, $(1, 5\pi/2)$, or even $(-1, 3\pi/2)$. And the origin was especially devious, it could be written as $(0, \theta)$ for any angle θ.

Because of this and the relationship between polar and cylindrical coordinates, it's not surprisingly that cylindrical coordinates have similar issues with uniqueness. For example, the point $(0, 1, 1)$ can be written as $(r, \theta, z) = (1, \pi/2, 1)$, $(1, 5\pi/2, 1)$, $(-1, 3\pi/w, 1)$, and so on. Any point on the z-axis can be written as $(0, \theta, z)$, where z is its z-coordinate, and θ is any angle.

Problem 2 *Which of the following, written in cylindrical coordinates, is equivalent to the point $(x, y, z) = (1, 1, 1)$? Select all that apply.*

Select All Correct Answers:

(a) $(1, 1, 1)$

(b) $(1, \pi/4, 1)$

(c) $\sqrt{2}, \pi/4, 1)$ ✓

(d) $(-1, 3\pi/4, 1)$

(e) $-\sqrt{2}, \pi/4, 1)$

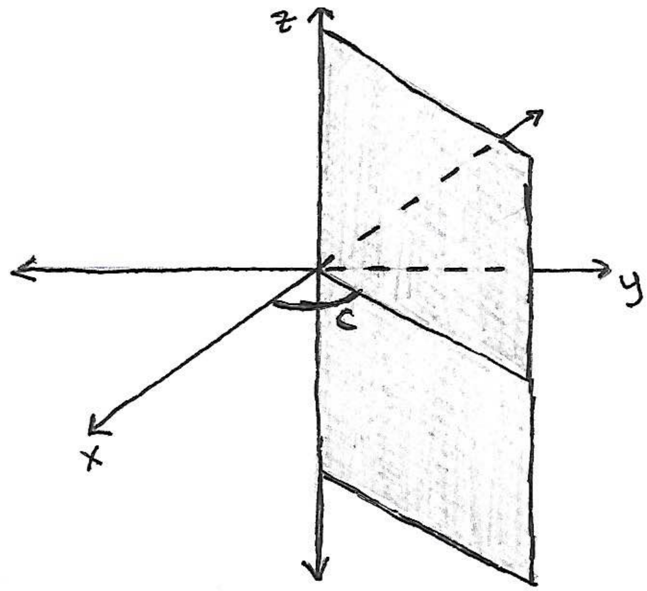
(f) $-\sqrt{2}, -3\pi/4, 1)$ ✓

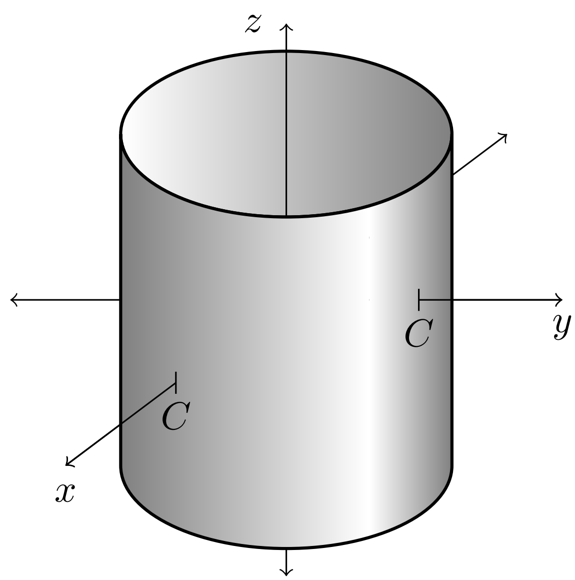
As with polar coordinates, in situations where uniqueness is important, we will often make the restrictions $r \geq 0$ and $0 \leq \theta < 2\pi$.

Constant-Coordinate Surfaces

Let's look at what happens in cylindrical coordinates when we set each of the coordinates r, θ, z to be constant, with the standard restrictions that $0 \leq r$ and $0 \leq \theta \leq \pi/2$. This can give us insight to how cylindrical coordinates behave.

We'll begin by examining the set of points (r, θ, z), where $r = C$ is a constant. We have that $r = C$ is constant, which means that the distance between any such point and the z axis is constant, C. Also, θ and z can be anything. This will give us the cylinder of radius C, centered at the z-axis.

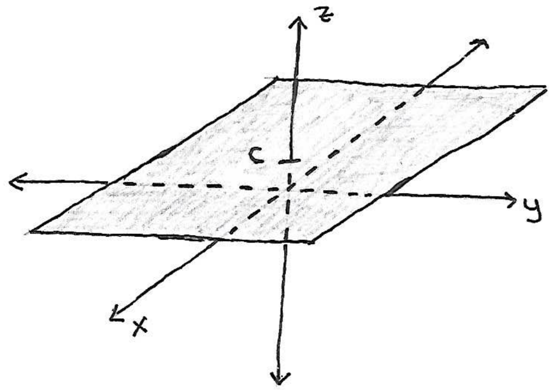




Next, we'll investigate the set of points (r, θ, z), where $\theta = C$ is constant. Let's consider the projection of this point onto the xy-plane. The projection will make an angle C with the positive x-axis, and have distance $r \geq 0$ from the origin. The height of the point can be any real number. From these observations, we conclude that the set of such points is the following half plane in $\mathbb{R}^3$.

Note that if we didn't have the restriction $r \geq 0$, we would get an entire plane rather than a half plane.

Finally, we'll consider the set of points (r, θ, z), where $z = C$ is constant. Since $z = C$, we will only have points at height C. Varying r and θ will then give us all points in the plane at height C parallel to the xy-plane, as below.



Converting between Cartesian and cylindrical coordinates

Perhaps not surprisingly, converting between Cartesian coordinates and cylindrical coordinates is very similar to how we converted between Cartesian coordinates and polar coordinates. That is, we can use the equations:

$$\begin{aligned} x &= r\cos\theta, \\ y &= r\sin\theta, \\ z &= z, \\ r^2 &= x^2 + y^2, \\ \tan\theta &= \frac{y}{x}. \end{aligned}$$

Example 26. *We'll convert $z = \sqrt{1 - r^2}$ to Cartesian coordinates.*

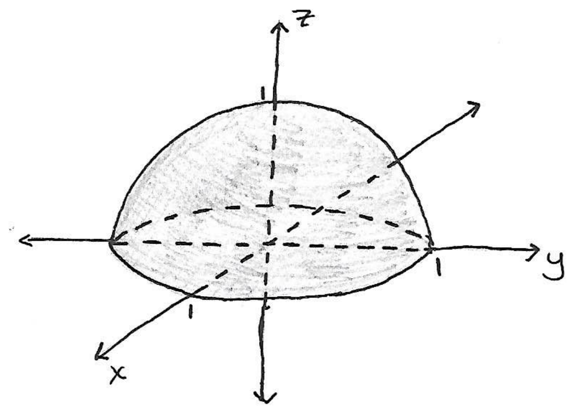
Using $r^2 = x^2 + y^2$, we have

$$z = \sqrt{1 - x^2 - y^2}.$$

You may recognize this as the top half of the sphere of radius 1 centered at the origin. You could also rewrite this as

$$x^2 + y^2 + z^2 = 1,$$

keeping in mind that $z \geq 0$*.*



Example 27. *We'll convert* $(x-2)^2 + y^2 = 1$ *(where* z *can be anything) to cylindrical coordinates. Note that this is the cylinder of radius 1, centered at the vertical line through* $(2, 0, 0)$*.*

Expanding the expression, we have

$$x^2 - 4x + 1 + y^2 = 1.$$

Substituting $r^2 = x^2 + y^2$ *and subtracting* 1 *from each side, we obtain*

$$r^2 - 4x = 0.$$

We then substitute $x = r\cos\theta$*.*

$$r^2 - 4r\cos\theta = 0.$$

Dividing both sides by r*, we have*

$$r - 4\cos\theta = 0,$$

or

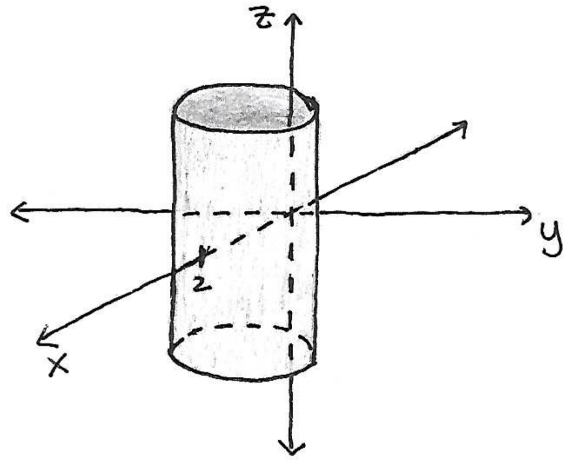
$$r = 4\cos\theta,$$

and z *can be anything.*

When we divided by r*, we implicitly assumed that* r *was not* 0*. This means that we might accidentally be omitting the origin, but if we take* $\theta = \pi/2$*, we have*

$$r = 4\cos(0) = 0,$$

so the origin is already included in the surface $r = 4\cos\theta$*.*



Conclusion

We introduced cylindrical coordinates and how to convert between cylindrical coordinates and Cartesian coordinates, and we discussed the uniqueness of cylindrical coordinates.

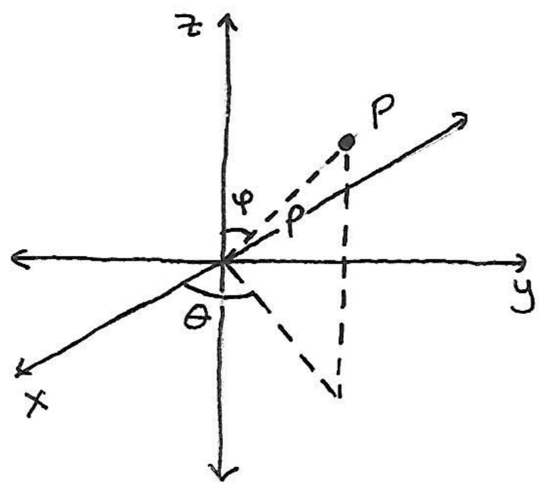
Spherical Coordinates

In this activity, we introduce spherical coordinates, a new coordinate system on $\mathbb{R}^3$. We also discuss how to convert between spherical and Cartesian coordinates.

Spherical Coordinates

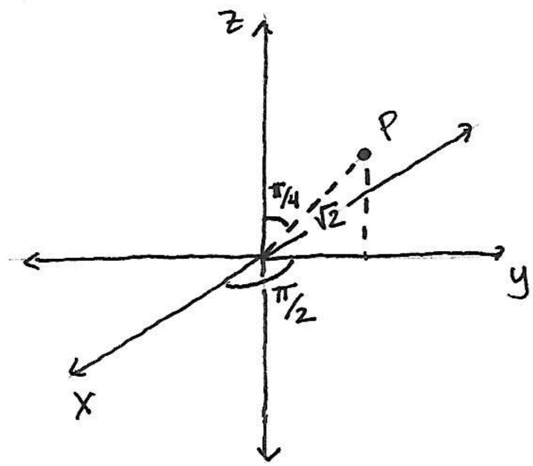
We've seen how to express points in $\mathbb{R}^3$ using Cartesian coordinates and using cylindrical coordinates. We'll now introduce a new coordinate system, called *spherical coordinates*.

Given a point P in $\mathbb{R}^3$, imagine drawing a line segment from the origin to P. In spherical coordinates, we write P as (ρ, θ, ϕ). Here, ρ is the length of the segment (also the distance between P and the origin). The second coordinate, θ, is angle between the positive x-axis and the projection of the segment onto the xy-plane. The third coordinate, ϕ, is the angle between the segment and the positive z-axis.



Example 28. *We'll write the point P in spherical coordinates, where P is given by $(x, y, z) = (0, 1, 1)$ in Cartesian coordinates.*

Learning outcomes:
Author(s):



The distance between P and the origin is

$$\sqrt{0^2 + 1^2 + 1^2} = \boxed{\sqrt{2}},$$

so $\rho = \boxed{\sqrt{2}}$.

The angle between the positive x-axis and the projection of P onto the xy-plane is $\boxed{\pi/2}$ *(in radians), so* $\theta = \boxed{\pi/2}$.

The angle between P and the positive z-axis is $\boxed{\pi/4}$ *(in radians), so* $\phi = \boxed{\pi/4}$.

Thus we can write P in spherical coordinates as $\boxed{(\sqrt{2}, \pi/2, \pi/4)}$.

Although we will be consistent with our definitions of θ and ϕ as above, it's important to know that some people reverse the roles of θ and ϕ. This is particularly common among physicists.

Uniqueness

As with polar and cylindrical coordinates, there are issues of uniqueness with spherical coordinates that we do not encounter in Cartesian coordinates.

Let's take for the example the point $(x, y, z) = (0, 1, 1)$, written in Cartesian coordinates. We've seen the canonical way to write this point in spherical coordinates, as $(\sqrt{2}, \pi/2, \pi/4)$. However, we could also write this as $(\sqrt{2}, 5\pi/2, \pi/4)$, $(\sqrt{2}, -3\pi/2, \pi/4)$, or even $(-\sqrt{2}, 3\pi/2, 5\pi/4)$.

Because of this issue, we'll commonly use the restrictions

$$
\begin{aligned}
0 \leq& \rho \\
0 \leq& \theta < 2\pi \\
0 \leq& \phi \leq \pi
\end{aligned}
$$

when working with spherical coordinates in order to improve the uniqueness situation. Unfortunately, there are still multiple ways to represent the origin in spherical coordinates.

Problem 3 *Which of the following represent the origin in spherical coordinates? Select all that apply.*

Select All Correct Answers:

(a) $(0, 0, 0)$ ✓

(b) $(0, \pi/2, 0)$ ✓

(c) $(0, 0, \pi/4)$ ✓

(d) $(0, \pi/2, \pi/4)$ ✓

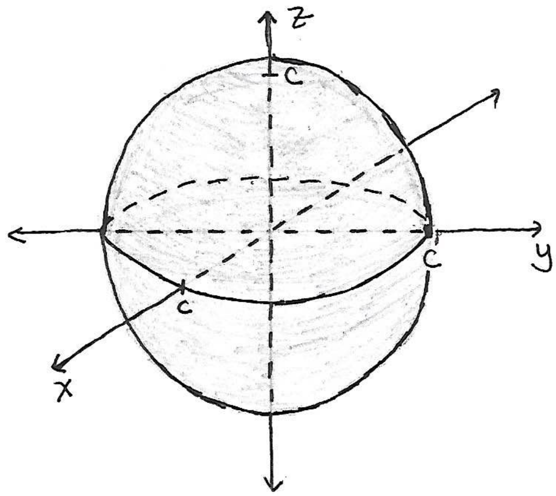
You may use the following applet to experiment with how the different coordinates change a point written in spherical coordinates.

MATH INSIGHT APPLET

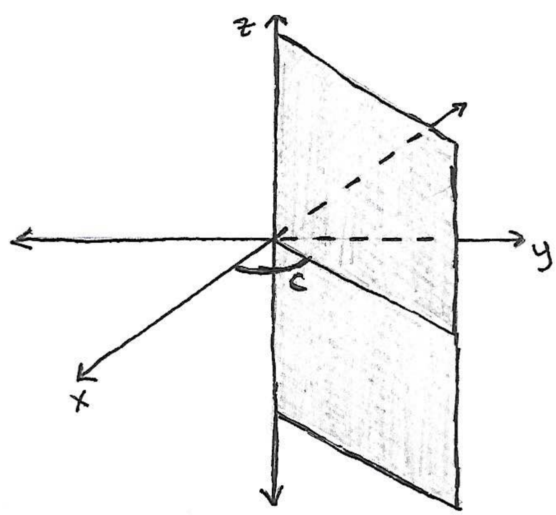
Constant-Coordinate Surfaces

As we did with cylindrical coordinates, we'll see what happens when we set each of the coordinates to be constant.

Consider the set of points (ρ, θ, ϕ), where $\rho = C$ is constant. This means that the distance between the origin and any such point is C. Varying the angles θ and ϕ gives us all such points, which make a sphere of radius C.



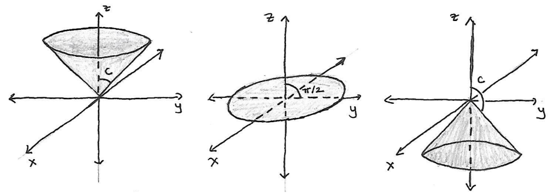
Now, consider the set of points (ρ, θ, ϕ), where $\theta = C$ is constant. This means that the projection of any such point onto the xy-plane will make an angle C with the positive x-axis. Varying ρ gives us points at various distances from the origin, and varying ϕ gives us points making various angles with the positive z-axis. With the restrictions $\rho \geq 0$ and $0 \leq \phi \leq \pi$, we obtain a half plane, as below.



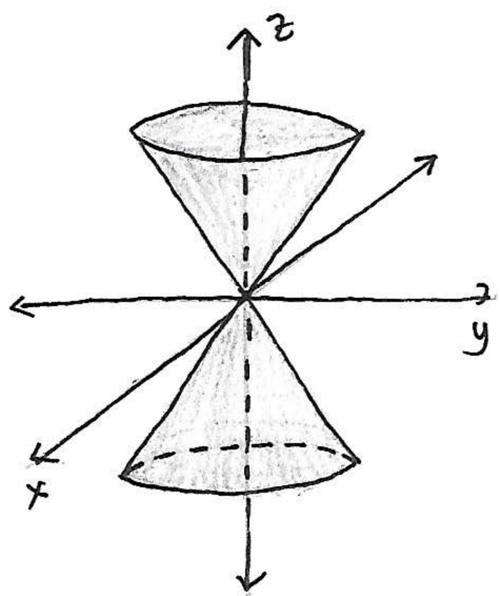
Notice that if we relaxed the restrictions on ρ and ϕ, we could obtain the entire plane.

Finally, we consider the set of points (ρ, θ, ϕ), where $\phi = C$ is constant. This means that every such point has an angle C with the positive z-axis. Varying

ρ and θ, with the restriction $\rho \geq 0$, we get the surfaces below, depending on if $C < \pi/2$, $C = pi/2$, or $C > pi/2$.



Looking at the surfaces when $C > \pi/2$ or $C < \pi/2$, we would commonly call these surfaces "cones." However, in most mathematics, "cone" is more commonly used to describe the surface below, which you might call a double cone.



Note that if you relax the restriction $\rho \geq 0$, you'll get cone (or double cone) above when $C \neq 0$.

It may seem strange that mathematicians prefer this double cone to the seemingly simpler cones that you're used to. However, it turns out that the double cone is easier to describe algebraically.

You can use the following applet to see what happens when you vary the value of the constant C for each of the constant-coordinate surfaces above:

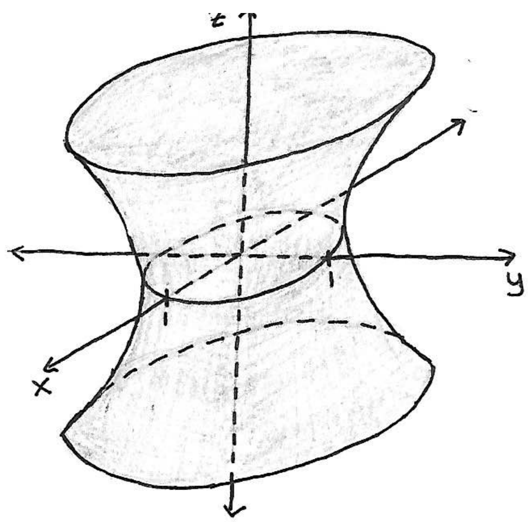
MATH INSIGHT APPLET

Converting Between Spherical and Cartesian Coordinates

When converting between spherical coordinates and Cartesian coordinates, it can be useful to use the following equations, which describe the relationship between the two coordinate systems.

$$\begin{aligned} x &= \rho\cos\theta\sin\phi \\ y &= \rho\sin\theta\sin\phi \\ z &= \rho\cos\phi \\ \rho^2 &= x^2+y^2+z^2 \end{aligned}$$

Example 29. *We'll convert $x^2+y^2-z^2=1$ from Cartesian coordinates to spherical coordinates. This surface is called an* elliptic hyperboloid*, and its graph is shown below. We'll learn how to identify this and other surfaces later in the course.*



Making the substitutions $x=\rho\cos\theta\sin\phi$, $y=\rho\sin\theta\sin\phi$, and $z=\rho\cos\phi$, we have

$$\rho^2\cos^2\theta\sin^2\phi+\rho^2\sin^2\theta\sin^2\phi-\rho^2\cos^2\phi=1.$$

We can factor $\rho^2\sin^2\phi$ out of the first two terms and obtain

$$\rho^2\sin^2\phi(\cos^2\theta+\sin^2\theta)-\rho^2\cos^2\phi=1.$$

Recalling that $\cos^2\theta+\sin^2\theta=1$, we can simplify this to

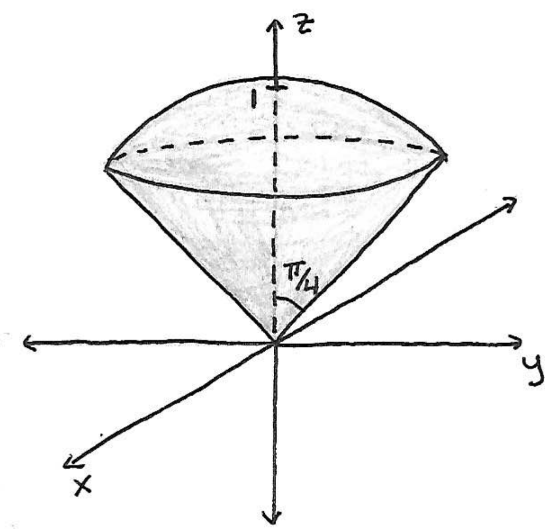
$$\rho^2\sin^2\phi-\rho^2\cos^2\phi=1.$$

Recalling the double angle formula $\cos(2\phi) = \cos^2(\phi) - \sin^2(\phi)$*, we can simplify this to*

$$\rho^2 \cos(2\phi) = 1.$$

Example 30. *Sketch the set of points* (ρ, θ, ϕ) *(in spherical coordinates) such that* $0 \leq \rho \leq 1$ *and* $0 \leq \phi \leq \pi/4$.

The condition $0 \leq \rho \leq 1$ *means that we'll have only points within distance* 1 *of the origin. The condition* $0 \leq \phi \leq \pi/4$ *means that we'll have only points within angle* $\pi/4$ *from the* z*-axis. Putting these conditions together, we have the solid "ice-cream cone" region sketched below.*



Spherical coordinates in $\mathbb{R}^n$ [OPTIONAL]

Since we've seen polar coordinates in $\mathbb{R}^2$, and cylindrical and spherical coordinates in $\mathbb{R}^3$, you might be wondering if there are similar coordinate systems in $\mathbb{R}^4$, $\mathbb{R}^5$, and so on.

It is possible to define spherical coordinates in $\mathbb{R}^n$ for any n, and you can find a description here.

Conclusion

We introduced spherical coordinates and how to convert between spherical coordinates and Cartesian coordinates, and we discussed the uniqueness of spherical coordinates.

Functions

In this activity, we cover the definition of a function. We also cover several important definitions and properties of functions, including domain, codomain, range, surjective and injective functions, and component functions.

Definition of a Function

You've certainly seen many functions before. For example, you've worked with linear functions, such as

$$f(x) = 3x + 2,$$

quadratic functions, such as

$$h(t) = -4.9t^2 + 20t + 5,$$

and more complicated functions such as

$$g(x) = e^{5\sin(x^2)} + \ln\cot x.$$

You've seen functions of more than one variable in the form of linear transformations, such as

$$T : \mathbb{R}^2 \to \mathbb{R}^3,$$
$$T(x, y) = \begin{pmatrix} 1 & 2 \\ 0 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Not surprisingly, in multivariable calculus, we'll be studying functions of more than one variable. Before starting to work with these functions, we now cover some of the fundamental definitions and properties related to functions in general, beginning with the definition of a function.

Definition 17. *For sets X and Y, a* function $f : X \to Y$ *from X to Y assigns an element of Y to each element of X.*

We call X the domain *of f, and Y the* codomain *of f.*

VIDEO

We commonly think of X as giving the set of inputs to a function, and Y as containing the outputs. Each input coming from the set X has to have some

Learning outcomes:
Author(s):

corresponding output, but some elements of Y might not actually occur as outputs of the function.

Problem 4 *Which of the following are functions? Select All that apply.*

Select All Correct Answers:

(a) $f : \mathbb{R} \to \mathbb{R}$ *defined by* $f(x) = x^2$ ✓

(b) $f : \mathbb{R}^2 \to \mathbb{R}$ *defined by* $f(a, b) = a - b$ ✓

(c) $f : \mathbb{R} \to \mathbb{R}$ *defined by* $f(x) = \pm x$

(d) $f : \mathbb{R} \to \mathbb{R}^2$ *defined by* $f(x) = (x, x)$ ✓

If we would like to refer to the elements in the codomain which actually do occur as outputs, we call this the range of f.

Definition 18. *The* range *of a function* $f : X \to Y$ *is the set of elements* $y \in Y$ *such that there is some* $x \in X$ *with* $f(x) = y$. *That is, there is some input* x *that has* y *as an output. In set notation, we write*

$$\textit{Range } f = \{y \in Y \,:\, y = f(x) \text{ for some } x \in X\}.$$

VIDEO

Problem 5 *What is the range of the function* $f : \mathbb{R} \to \mathbb{R}^2$ *defined by* $f(x) = (x, x)$*?*

Multiple Choice:

(a) $\mathbb{R}^2$

(b) $\mathbb{R}$

(c) $\{(a, b) \in \mathbb{R}^2 \,:\, a = b\}$ ✓

(d) $\{(a, b) \in \mathbb{R}^2 \,:\, a = b\}$

Sometimes we work with functions that aren't defined on all of $\mathbb{R}^n$. When the domain of f is a subset D of $\mathbb{R}^n$, we write

$$f : D \subset \mathbb{R}^n \to \mathbb{R}^m.$$

When we're working with functions on subsets of $\mathbb{R}^n$, we'll frequently want to work with the largest possible set that the function is defined on. We call this the *natural domain* of the function.

Problem 6 *What is the natural domain of the function* $f(x,y) = \dfrac{x}{x-y}$*?*

Multiple Choice:

(a) $\mathbb{R}^2$

(b) $\mathbb{R}^2 \setminus \{(0,0)\}$

(c) $\mathbb{R}^2 \setminus \{(a,b) \,:\, a=0 \text{ or } b=0\}$

(d) $\mathbb{R}^2 \setminus \{(x,y) \,:\, a=b\}$ ✓

Types of Functions

In some special situations, every element of Y really does appear as an output of the function f. In this case, we say that f is onto, or surjective.

Definition 19. *A function* $f : X \to Y$ *is* onto*, or* surjective*, if for every element* $y \in Y$*, there is some* $x \in X$ *such that* $f(x) = y$*. We can also write this condition as*

$$Y = \mathit{Range}\ f.$$

VIDEO

Problem 7 *Which of the following functions are onto? Select all that apply.*

Select All Correct Answers:

(a) $f : \mathbb{R} \to \mathbb{R}$ *defined by* $f(x) = x^2$

(b) $f : \mathbb{R}^2 \to \mathbb{R}$ *defined by* $f(a,b) = a - b$ ✓

(c) $f : \mathbb{R} \to \mathbb{R}^2$ *defined by* $f(x) = (x,x)$

(d) $f : \mathbb{R}^2 \to \mathbb{R}$ *defined by* $f(x,y) = x^2y^2$

Another important type of function is a one-to-one, or injective, function. For a one-to-one function, different inputs always go to different outputs.

Definition 20. *A function* $f : X \to Y$ *is* one-to-one*, or* injective*, if whenever* $f(x_1) = f(x_2)$ *for* $x_1, x_2 \in X$*, then we must have* $x_1 = x_2$*.*

Another way to say this is that whenever $x_1 \neq x_2$*, we have* $f(x_1) \neq f(x_2)$*.*

VIDEO

Problem 8 *Which of the following functions are injective? Select all that apply.*

Select All Correct Answers:

(a) $f : \mathbb{R} \to \mathbb{R}$ *defined by* $f(x) = x^2$

(b) $f : \mathbb{R}^2 \to \mathbb{R}$ *defined by* $f(a, b) = a - b$

(c) $f : \mathbb{R} \to \mathbb{R}^2$ *defined by* $f(x) = (x, x)$ ✓

(d) $f : \mathbb{R}^2 \to \mathbb{R}$ *defined by* $f(x, y) = x^2y^2$

Component Functions

When we're trying to understand the behavior of a function $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$, it can sometimes be helpful to split $\mathbb{R}^m$ into its components. From this, we get the component functions of f.

Definition 21. *Let $f : X \subset \mathbb{R}^n \to \mathbb{R}^m$ be a function. The* component functions *of f are scalar-valued functions $f_i : X \subset \mathbb{R}^n \to \mathbb{R}$ for $1 \leq i \leq m$ such that*

$$f(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), ..., f_m(\vec{x})).$$

Conclusion

We covered the definition of a function. We also covered several important definitions and properties of functions, including domain, codomain, range, surjective and injective functions, and component functions.

Graphing Functions

In this activity, we give the formal definition of the graph of a function $f : \mathbb{R}^2 \to \mathbb{R}$. We discuss strategies for figuring out the shapes of the graph of such functions, using contour curves, level curves, and sections.

Definition of the Graph of a Function

You might already have an intuitive idea of what the graph of a function $f : X \subset \mathbb{R}^2 \to \mathbb{R}$ should be, but perhaps don't know the formal definition, or how to figure out what the graph of an arbitrary function looks like. We'll begin with the definition of the graph, before discussing how to actually produce graphs.

Definition 22. *Let $f : X \subset \mathbb{R}^2 \to \mathbb{R}$ be a function. The* graph *of f is the set of points*

$$\textit{Graph } f = \{(\vec{x}, f(\vec{x}) : \vec{x} \in X\}$$

in $\mathbb{R}^3$.

We typically visualize a point in the graph as lying over the point $\vec{x}$ in the plane at a height $f(\vec{x})$.

Note that this is similar to the graph of a function from a subset of $\mathbb{R}$ to $\mathbb{R}$. The graph of a function $X \subset \mathbb{R}^n \to \mathbb{R}$ can be defined similarly, but this tends to be less useful once $n \geq 3$, since it's hard to visualize four or more dimensions!

Strategies for Graphing

It can be much trickier to sketch the graph of a function $f : X \subset \mathbb{R}^2 \to \mathbb{R}$ than it was to sketch the graphs of functions $\mathbb{R} \to \mathbb{R}$. One common strategy that people will initially try is plotting individual points to try to get a sense of the graph. However, for graphs in $\mathbb{R}^3$, you would need a lot of points to get a representative sample of the plane. For this reason, *plotting points alone is not an effective strategy.* However, plotting a single point here or there can be helpful.

We've now told you what doesn't work for graphing functions in $\mathbb{R}^3$, so now we should probably tell you what does work. The essential idea of all of these

Learning outcomes:
Author(s):

strategies is that we know you're pretty comfortable graphing in $\mathbb{R}^2$, so we're going to take advantage of that experience.

We'll begin with contour curves, which are obtained by setting the z-coordinate to be constant. Think of this as taking horizontal slices of the graph.

Definition 23. *Let $f : X \subset \mathbb{R}^2 \to \mathbb{R}$ be a function. The* contour curve *of the function f at height C is the set of points in $\mathbb{R}^3$ obtained by taking the intersection of the graph of f with the plane $z = C$.*

PICTURE/VIDEO EXAMPLE

We can also consider the level curves of a function, which are closely related to contour curves.

Definition 24. *Let $f : X \subset \mathbb{R}^2 \to \mathbb{R}$ be a function. The* level curve *of the function f at height C is the set of points in $\mathbb{R}^2$ satisfying $C = f(x, y)$.*

After reading this definition, you're probably thinking "hey, aren't contour curves and level curves the exact same thing?" They're certainly closely related. The key difference is that level curves exist in the plane, $\mathbb{R}^2$, while contour curves exist in three-space, $\mathbb{R}^3$. Since they're in the plane, level curves are usually easier to draw. However, contour curves are more useful for figuring out the shape of a graph. For these reasons, it can be useful to go back and forth between level curves and contour curves.

PICTURE/VIDEO EXAMPLE

We can think of contour curves as taking slices of the graph where z is constant. It can also be useful to take slices of the graph where x or y is constant. We call these slices sections of the graph.

Definition 25. *Let $f : X \subset \mathbb{R}^2 \to \mathbb{R}$ be a function, and let C be a constant.*

The section *of the graph of f by $x = C$ is the set of points*

$$\{(C, y, z) \in \mathbb{R}^3 \, : \, z = f(C, y)\}$$

The section *of the graph of f by $y = C$ is the set of points*

$$\{(x, C, z) \in \mathbb{R}^3 \, : \, z = f(x, C)\}$$

Note that, like contour curves, sections exist in $\mathbb{R}^3$.

PICTURE/VIDEO EXAMPLE

EXTRA EXAMPLE

Level Surfaces

So far, we have focused on graphing functions from subsets of $\mathbb{R}^2$ to $\mathbb{R}$, so the graphs are in $\mathbb{R}^3$.

We now turn our attention to the graphs of functions from subsets of $\mathbb{R}^3$ to $\mathbb{R}$. Note that the graph of such a function will exist in $\mathbb{R}^4$. Since the world we live in only has three physical dimensions, it can be very difficult to visualize a four dimensional object! Fortunately, there are various tricks that can be used to get some sense of what a four dimensional object looks like. We cover one of them here.

When we had a function $f : D \subset \mathbb{R}^2 \to \mathbb{R}$, we could get a sense of the graph by looking at its level curves, which were curves in the same plane.

For a function $f : D \subset \mathbb{R}^3 \to \mathbb{R}$, we can adopt a similar approach. We can once again consider the level sets, which are obtained by taking the output to be some constant:

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

In this case, the level sets will be level surfaces, which live in $\mathbb{R}^3$. By graphing several level surfaces, we can see what a slice of the graph of f looks like at various heights, giving us some sense of how the overall graph behaves. Of course, because this graph exists in four dimensions, we still probably won't be able to visualize this perfectly.

To see how this can help us visualize the four-dimensional graph of a function $f : \mathbb{R}^3 \to \mathbb{R}$, we give an example.

Example 31. *Consider the function*

$$f(x, y, z) = \sqrt{x^2 + y^2 + z^2}.$$

Find the level surfaces at heights -1, 0, 1, 2, *and* 3. *Use these level surfaces to describe the graph of* f.

We'll begin with the level surface at height -1. *This is the set of points* (x, y, z) *in* $\mathbb{R}^3$ *such that*

$$-1 = \sqrt{x^2 + y^2 + z^2}.$$

There are no points that satisfy this equation, so the level surface is empty.

Now we'll consider the level surface at height 0. *This is the set of points* (x, y, z) *in* $\mathbb{R}^3$ *such that*

$$0 = \sqrt{x^2 + y^2 + z^2}.$$

The only point which satisfies this equation is the origin, so the level "surface" is the single point $(0, 0, 0)$.

Let's look at the level surface at height 1. *This is the set of points* (x, y, z) *in* $\mathbb{R}^3$ *such that*

$$1 = \sqrt{x^2 + y^2 + z^2}.$$

Squaring both sides, we can rewriting this as

$$1 = x^2 + y^2 + z^2.$$

The graph of this equation is the sphere of radius 1 *centered at the origin, which is our level surface.*

Let's look at the level surface at height 2*. This is the set of points* (x, y, z) *in* $\mathbb{R}^3$ *such that*

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

Squaring both sides, we can rewriting this as

$$4 = x^2 + y^2 + z^2.$$

The graph of this equation is the sphere of radius 2 *centered at the origin, which is our level surface.*

Let's look at the level surface at height 3*. This is the set of points* (x, y, z) *in* $\mathbb{R}^3$ *such that*

$$3 = \sqrt{x^2 + y^2 + z^2}.$$

Squaring both sides, we can rewriting this as

$$9 = x^2 + y^2 + z^2.$$

The graph of this equation is the sphere of radius 3 *centered at the origin, which is our level surface.*

We graph our level surfaces below.

PICTURE

We can see that the level surfaces are spheres whose radii increase linearly with the height. So, we can describe the graph of f *as some sort of four-dimensional cone.*

Conclusion

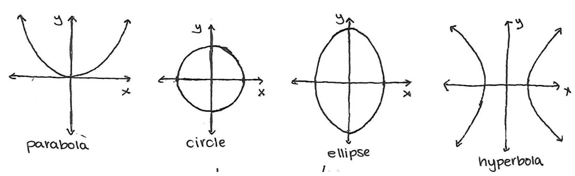
In this activity, we gave the formal definition of the graph of a function $f : \mathbb{R}^2 \to \mathbb{R}$. We discussed strategies for figuring out the shapes of the graph of such functions, using contour curves, level curves, and sections.

Quadric Surfaces

In this activity, we introduce and classify quadric surfaces, which form an important family of surfaces.

Definition of a Quadric Surface

You might remember studying conic sections, such as parabolas, circles, ellipses, and hyperbolas. These are curves in the plane that arise through polynomial equations of degree two in two variables.



Quadric Surfaces are the three dimensional analogue of conic sections. That is, a quadric surface is the set of points in $\mathbb{R}^3$ satisfying some polynomial of degree two in three variables.

Definition 26. *A* quadric surface *is the set of points* (x, y, z) *in* $\mathbb{R}^3$ *satisfying the equation*

$$Ax^2 + Bxy + Cxz + Dy^2 + Eyz + Fz^2 + Gx + Hy + Iz + J = 0,$$

where $A, B, C, D, E, F, G, H, I, J \in \mathbb{R}$ *are constants.*

Simple Forms

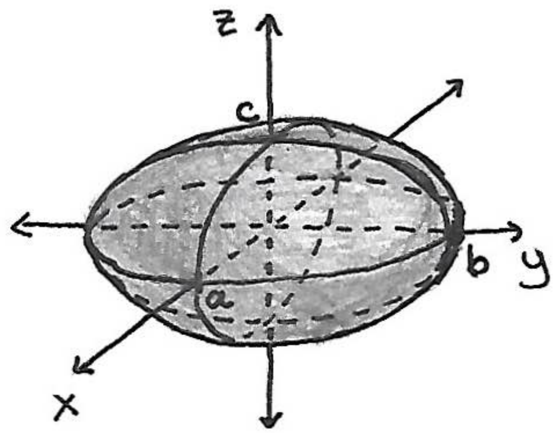
Dealing with quadric surfaces in general can be computationally cumbersome, so we'll focus on quadric surfaces in some simple forms.

Example 32. *The set of points satisfying*

$$x^2/a^2 + y^2/b^2 + z^2/c^2 = 1,$$

for some constants $a, b, c \in \mathbb{R}$*, is called an* ellipsoid.

Learning outcomes:
Author(s):



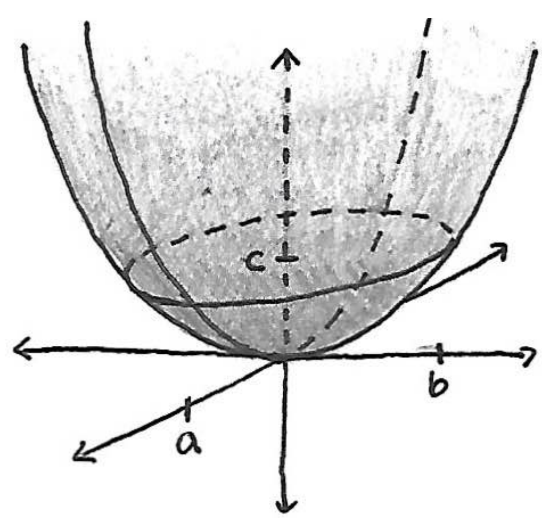
An ellipsoid is kind of like a three dimensional ellipse. In fact, the sections and contour curves of such an an ellipsoid are ellipses.

In the special case that $a = b = c$, this ellipsoid is a sphere of radius a.

Example 33. *The set of points satisfying*

$$z/c = x^2/a^2 + y^2/b^2,$$

for some constants $a, b, c \in \mathbb{R}$, is called an elliptic paraboloid.

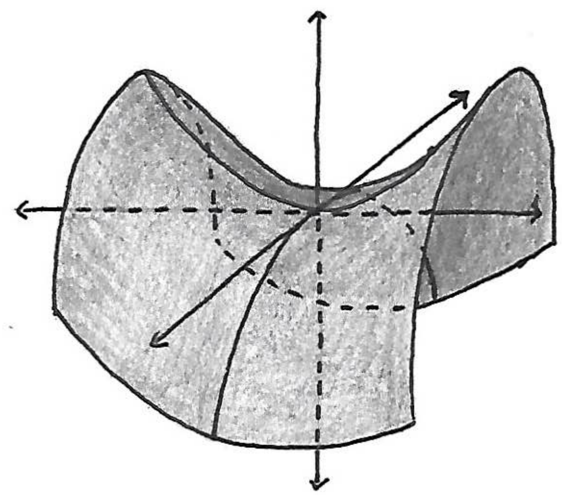


The contour curves of such an elliptic paraboloid are ellipses, however the sections are parabolas which all open in the same direction.

Example 34. *The set of points satisfying*

$$z/c = y^2/b^2 - x^2/a^2,$$

for some constants $a, b, c \in \mathbb{R}$, is called a hyperbolic paraboloid*.*

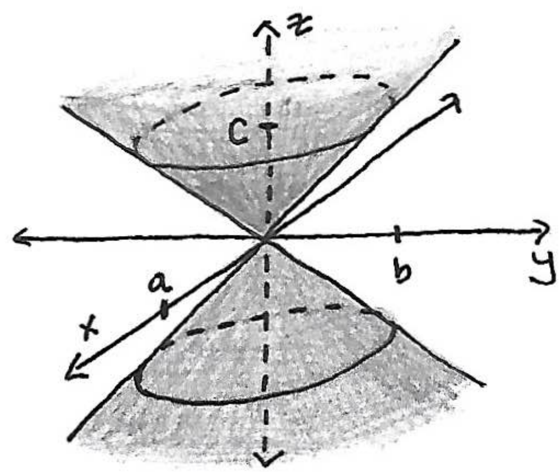


The contour curves of such a hyperbolic paraboloid are hyperbolas, and the sections are parabolas opening in opposite directions for x and y sections. This surface is often described as a "saddle".

Example 35. *The set of points satisfying*

$$z^2/c^2 = x^2/a^2 + y^2/b^2,$$

for some constants $a, b, c \in \mathbb{R}$, is called an elliptic cone*.*

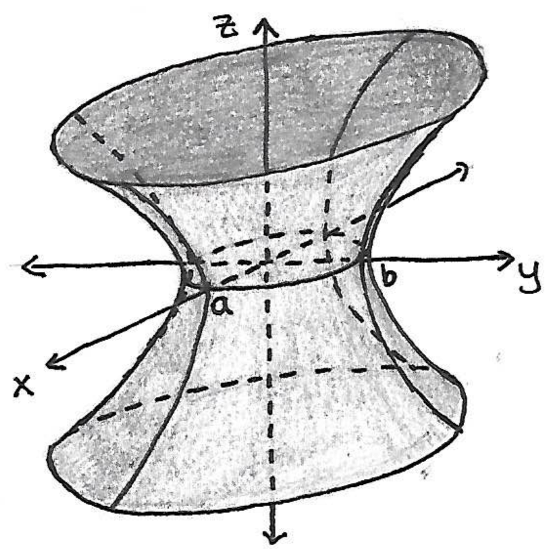


The contour curves of such an elliptic cone are ellipses, and the sections by $x = 0$ and $y = 0$ are pairs of intersecting lines.

Example 36. *The set of points satisfying*

$$x^2/a^2 + y^2/b^2 - z^2/c^2 = 1.$$

for some constants $a, b, c \in \mathbb{R}$, is called a hyperboloid of one sheet*.*

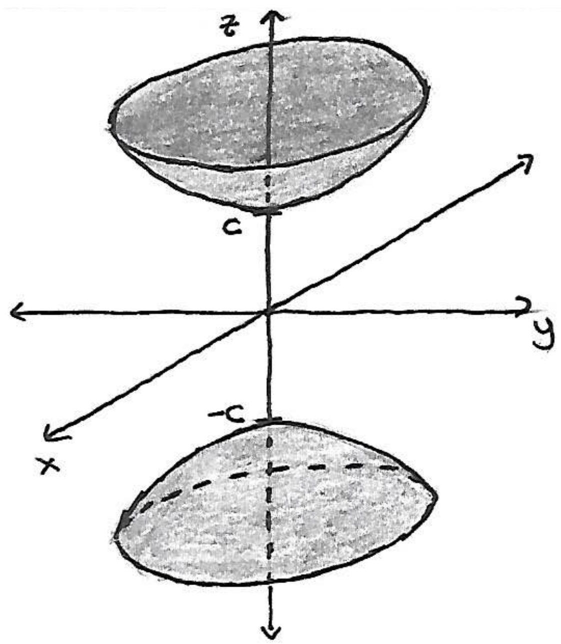


The contour curves of such a hyperboloid are ellipses, and the sections are hyperbolas.

Example 37. *The set of points satisfying*

$$z^2/c^2 - x^2/a^2 - y^2/b^2 = 1.$$

for some constants $a, b, c \in \mathbb{R}$, is called a hyperboloid of two sheets*.*



The contour curves of such a hyperboloid are ellipses, and the sections are hyperbolas. We describe this as the hyperboloid "of two sheets" since it has two disconnected pieces, as opposed to the hyperboloid of one sheet, which has only one.

APPLET/INTERACTIVE VARYING PARAMETERS

Some Other Forms

Although we won't really work with quadric surfaces in their most general form, we will consider quadric surfaces that are translations of the forms given above.

For example, the graph of the equation

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

is an ellipsoid centered at $(3, -2, 1)$.

PICTURE

However, equations describing quadric surfaces might not always be given to you in easily identifiable forms. In these cases, you might have to do some algebra in order to get the equation into a form where it can be identified as a particular quadric surface. These manipulations will frequently involve completing the square.

We now work through an example of identifying a quadric surface given in a non-standard form.

Example 38. *Identify the type of quadric surface determined by the equation*

$$-4x^2 + 2y^2 + z^2 + 8x + 4y + 4z = 2,$$

and sketch a graph of this surface.

Our strategy for writing this equation in a recognizable form will be to group terms involving x, group terms involving y, and group terms involving z. We'll then complete the square for each variable.

Grouping terms by variable, we have

$$(-4x^2 + 8x) + (2y^2 + 4y) + (z^2 + 4z) = 2.$$

For each of these grouping, we factor out the leading coefficient, obtaining

$$\boxed{-4}(x^2 - 2x) + \boxed{2}(y^2 + 2y) + (z^2 + 4z) = 2.$$

We now add or subtract as needed to make the quadratics into squares, getting

$$-4(x^2 - 2x + 1) + 2(y^2 + 2y + 1) + (z^2 + 4z + 2) = \boxed{4}.$$

We factor the quadratics to get

$$-4(\boxed{x-1})^2 + 2(\boxed{y+1})^2 + (\boxed{z+2})^2 = 4.$$

Finally, we divide by the constant on the right, to get the final form

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

We can see that this quadric surface is centered at $\boxed{(1, -1, -2)}$*, but maybe it still isn't apparent which quadric surface this determines.*

Notice that this form is similar to our standard form for a hyperboloid of one sheet, except here it's the x-term that's subtracted instead of the z-term. This is because this is, in fact, a hyperboloid of one sheet, it just happens to be "around" a line parallel to the x-axis, rather than a vertical line.

Let's look at a section, in order to help with our sketch. Taking the section $x = 1$, we have an ellipse parallel to the yz-plane, centered at $(1, -1, -2)$, with radii $\sqrt{2}$ and 2.

Combining our observations, we can sketch the graph of this hyperboloid as below.

PICTURE

Conclusion

In this activity, we introduced and classified quadric surfaces, which form an important family of surfaces.

Part III

Curves

Parametric Curves

In this activity, we parametrize curves in $\mathbb{R}^n$, focusing on the cases $n = 2$ and $n = 3$.

Review of Parametrizations in $\mathbb{R}^2$

We've dealt with several ways to describe curves in $\mathbb{R}^2$:

- As the graph of a function. For example, $f(x) = x^2$.
- As the set of points satisfying an equation. For example, the points (x, y) such that $x^2 + y^2 = 1$.
- As the set of points satisfying an equation in another coordinate system. For example, $r = \sin(\theta)$ in polar coordinates.

Another way that we can describe a curve is using *parametric equations*. In parametric equations, we define x and y in terms of a third variable, usually t, called the *parameter*. This gives us another way to describe curves in $\mathbb{R}^2$, and potentially describe some new and strange curves.

Example 39. *We can describe the unit circle in $\mathbb{R}^2$ with the parametric equations*

$$\begin{aligned} x &= \cos(t), \\ y &= \sin(t), \end{aligned}$$

for $0 \leq t \leq 2\pi$.

PICTURE

We can think of t as giving the angle that a point makes with the positive axis. It can also be helpful to imagine t as representing time, and the parametric equations tracing out the circle as time passes.

Learning outcomes:
Author(s):

Parametrizing Curves in $\mathbb{R}^n$

Consider the parametric equations for the unit circle in $\mathbb{R}^2$:

$$x = \cos(t),$$
$$y = \sin(t),$$

for $0 \leq t \leq 2\pi$.

We can combine these equations into a single vector,

$$\vec{x}(t) = (\cos(t), \sin(t)) \text{ for } 0 \leq t \leq 2\pi.$$

This defines a function $\vec{x}$ from the interval $[0, 2\pi] \subset \mathbb{R}$ to $\mathbb{R}^2$, and is the motivation behind our definition for paths.

Definition 27. *A* path *in $\mathbb{R}^n$ is a continuous function*

$$\vec{x} : I \subset \mathbb{R} \to \mathbb{R}^n,$$

where $I \subset \mathbb{R}$ is an interval.

This is also called a parametrized curve *or* parametric curve*.*

We'll focus on the cases $n = 2$ and $n = 3$ in this course.

We defined a path as a continuous function, however, we haven't said what it means for a multivariable function to be continuous. We'll come back to this later, and we'll give a rigorous definition for continuity. For now, this should fit with your intuition: you can draw the path without lifting your pencil from the paper.

Sometimes we can more about the image of a path than how the path is drawn out, and then we refer to a curve.

Definition 28. *A* curve *in $\mathbb{R}^n$ is the image of a path $\vec{x} : I \subset \mathbb{R} \to \mathbb{R}^n$.*

We say that $\vec{x}$ is a parametrization *for the curve.*

The difference between a curve and a path is largely a matter of perspective: when working with a curve, we pay attention to *what* is drawn; when working with a path, we care about *how* it is drawn.

Example 40. *There are many different parametrizations for a given curve.*

Consider again the unit circle C in $\mathbb{R}^2$. Which of the following are parametrizations for C?

Select All Correct Answers:

(a) $\vec{x}(t) = (\cos(t), \sin(t))$ *for* $0 \leq t \leq 2\pi$ ✓

(b) $\vec{x}(t) = (\sin(t), \cos(t))$ *for* $0 \leq t \leq \pi$

(c) $\vec{x}(t) = (t, \pm\sqrt{1-t^2})$ *for* $-1 \leq t \leq 1$

(d) $\vec{x}(t) = (\sin(2\pi t), \cos(2\pi t))$ *for* $0 \leq t \leq 1$ ✓

(e) $\vec{x}(t) = (\cos(t), \sin(t))$ *for* $-10 \leq t \leq 10$ ✓

Example 41. *In this example, we review how to parametrize the line through points $\vec{a}$ and $\vec{b}$ in $\mathbb{R}^n$.*

Given points $\vec{a}$ and $\vec{b}$ in $\mathbb{R}^n$, we obtain a vector starting at $\vec{a}$ and ending at $\vec{b}$ by taking $\vec{b} - \vec{a}$. This vector is parallel to the line through $\vec{a}$ and $\vec{b}$. Then, taking scalar multiples $t(\vec{b} - \vec{a})$ for $t \in \mathbb{R}$, we have a line parallel to the line through $\vec{a}$ and $\vec{b}$. Finally, we add one of the points, $\vec{a}$, to ensure that our line passes through these two points. Thus, we arrive at our parametrization,

$$\vec{l}(t) = \vec{a} + t(\vec{b} - \vec{a}) \text{ for } t \in \mathbb{R}.$$

PICTURE

Example 42. *In this example, we see how we can obtain new transformations from old ones, using linear algebra and simple transformations.*

Recall the parametrization for the unit circle in $\mathbb{R}^2$,

$$\vec{x}(t) = (\cos(t), \sin(t)) \text{ for } 0 \leq t \leq 2\pi.$$

Now, consider the ellipse below.

PICTURE

We can think of this ellipse as the result of stretching the unit circle horizontally by a factor of 3 and vertically by a factor of 2. That is, we are applying the linear transformation

$$\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}.$$

We can apply this to the parametrization for the unit circle, in order to parametrization for the ellipse.

$$\begin{aligned} \vec{(y)}(t) &= \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}, \\ &= (3\cos(t), 2\sin(t)). \end{aligned}$$

Thus, we have a parametrization for the ellipse given by

$$\vec{y}(t) = (3\cos(t), 2\sin(t)) \text{ for } 0 \leq t \leq 2\pi.$$

Next, consider the following ellipse.

PICTURE

We can obtain this from our previous ellipse by counterclockwise rotation of $\pi/4$. The matrix for this linear transformation is

$$\begin{pmatrix} \cos(\pi/4) & -\sin(\pi/4) \\ \sin(\pi/4) & \cos(\pi/4) \end{pmatrix} = \begin{pmatrix} \boxed{1/\sqrt{2}} & \boxed{-1/\sqrt{2}} \\ \boxed{1/\sqrt{2}} & \boxed{1/\sqrt{2}} \end{pmatrix}.$$

Applying this rotation to our parametrization for the previous ellipse, we obtain a parametrization for our new ellipse.

$$\vec{z}(t) = \boxed{(3/\sqrt{2}\cos(t) - 2/\sqrt{2}\sin(t), 3/\sqrt{2}\cos(t) + 2/\sqrt{2}\sin(t))} \text{ for } 0 \le t \le 2\pi$$

Finally, we consider an ellipse in $\mathbb{R}^3$, shown below.

PICTURE

This ellipse is parallel to the xy-plane, and will have constant z-coordinate. Note the similarity to the first ellipse we considered. A parametrization for this ellipse can be obtained by taking the parametrization $\vec{y}$ for our first ellipse in $\mathbb{R}^2$, and appending the constant z-coordinate.

$$\vec{a}(t) = \boxed{(3\cos(t), 2\sin(t))} \text{ for } 0 \le t \le 2\pi$$

Examples in $\mathbb{R}^3$

In this section, we give examples of parametrizations of a couple of more complicated curves in $\mathbb{R}^3$, taking advantage of our previous experience with cylindrical coordinates.

Example 43. *We'll parametrize the intersection of the cylinder $x^2 + y^2 = 4$ and the plane $z = 7 - 3x$ in $\mathbb{R}^3$, pictured below.*

PICTURE

Our x and y coordinates must satisfy $x^2 + y^2 = 4$, which would define a circle, if we were in $\mathbb{R}^2$. Recalling our parametrizations for circles, these coordinates can be written as

$$x(t) = 2\cos(t)$$
$$y(t) = 2\sin(t)$$

for $0 \le t \le 2\pi$.

It remains to write the z-coordinate in terms of the parameter t. Turning our attention to the equation for the plane, $z = 7 - 3x$, we have z expressed in terms *of x. Since we have expressed x in terms of t, we can make this substitution to describe z in terms of t,*

$$z(t) = \boxed{7 - 6\cos(t)}.$$

Putting all of this together, we have a parametrization for this intersection given by

$$\vec{x}(t) = \boxed{(2\cos(t), 2\sin(t), 7 - 6\cos(t))} \text{ for } 0 \leq t \leq 2\pi.$$

Geogebra link: `https://tube.geogebra.org/m/dxywtu7x`

Example 44. *Consider the curve below, which lies on the cone $z^2 = x^2 + y^2$, and makes five rotations around the z-axis as the height ranges from 0 to 1. We'll refer to this curve as a "tornado."*

PICTURE

We'll parametrize this curve by thinking about it in cylindrical coordinates, using the height as the parameter.

First, let's consider what's happening with the z-coordinate. Since the height of the tornado rangers from 0 to 1, so will z. We'll set $z = t$, with $0 \leq t \leq 1$, and express x and y in terms of t as well.

Now, we turn our attention to the angle θ. As the height ranges from 0 to 1, the tornado makes five revolutions, so θ should range from 0 to 10π. Thus, expressing θ in terms of t, we let $\theta = 10\pi t$.

Next, we consider the radius r. Since we are on the cone $z^2 = x^2 + y^2$, we have $z^2 = r^2$. Since $z \geq 0$, we have $z = r$. Thus, we can write r in terms of t as $r = t$.

Finally, putting all of this together with $x = r\cos\theta$ and $y = r\sin\theta$, we have a parametrization for the tornado given by

$$\vec{x}(t) = \boxed{(t\cos(10\pi\theta), t\sin(10\pi\theta), t)} \text{ for } t \in [0, 1].$$

Geogebra link: `https://tube.geogebra.org/m/tdmnpcg5`

Conclusion

In this activity, we parametrized curves in $\mathbb{R}^n$, focusing on the cases $n = 2$ and $n = 3$.

Velocity and Speed

In this activity, we learn how to find the velocity and speed of a parametrized curve in $\mathbb{R}^n$.

Derivatives

Consider a path $\vec{x} : I \subset \mathbb{R} \to \mathbb{R}^n$ which parametrizes a curve in $\mathbb{R}^n$. We often think about this as a particle tracing out the curve as time, given by t, passes. We would like to be able to understand and describe the motion of the particle on the curve, and find its velocity and speed, in particular. In order to do this, we need to figure out how to differentiate a path.

Before we define the derivative of a path, we quickly review the single variable definition of a derivative, given in Calculus I.

Given a single variable function $f(x)$, we found the instantaneous rate of change at x of this function by taking the derivative of f at x. The derivative also told us the slope of the tangent line at x. In order to compute this, we imagined finding the slope of secant lines getting closer and closer to the point. Taking a limit, we obtained the slope of the tangent line.

PICTURE

The slope of the secant line through the points $(x, f(x))$ and $(x+h, f(x+h))$ is given by $\dfrac{f(x+h)-f(x)}{h}$, so we defined the derivative of f at x to be

$$f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}.$$

We use the same idea for a path $\vec{x}$ in $\mathbb{R}^n$. We consider secant vectors from $\vec{x}(t)$ to $\vec{x}(t+h)$ as $h \to 0$.

PICTURE

Scaling these vectors to account for the change in the parameter and taking a limit, we arrive at the definition of the derivative.

Definition 29. *Let $\vec{x} : I \subset \mathbb{R} \to \mathbb{R}^n$ be a path in $\mathbb{R}^n$. We define the derivative of $\vec{x}$ at t to be*

$$\vec{x}'(t) = \lim_{h\to 0} \frac{\vec{x}(t+h)-\vec{x}(t)}{h},$$

Learning outcomes:
Author(s):

if the limit exists.

We also call $\vec{x}'(t)$ the velocity vector *of $\vec{x}$, and write it as $\vec{v}(t)$.*

When we first defined derivatives in Calculus I, we spent weeks figuring out how to compute them. We started computing using only the limit definition, then we introduced the power rule, the product rule, the chain rule, and so on. Fortunately, we don't need to repeat this process in Multivariable Calculus: we can take advantage of our previous experience computing derivatives. In order to see why this is the case, let's take another look at our definition for the derivative of a path.

We have $\vec{x}'(t) = \lim_{h\to 0} \dfrac{\vec{x}(t+h) - \vec{x}(t)}{h}$ for a path $\vec{x}$. We can write out the path $\vec{x}$ in terms of its components, so

$$\vec{x}(t) = (x_1(t), ..., x_n(t)).$$

Substituting this into the limit, we have

$$\begin{aligned}
\vec{x}'(t) &= \lim_{h\to 0} \frac{\vec{x}(t+h) - \vec{x}(t)}{h}, \\
&= \lim_{h\to 0} \frac{(x_1(t+h), ..., x_n(t+h)) - (x_1(t), ..., x_n(t))}{h}, \\
&= \lim_{h\to 0} \frac{(x_1(t+h) - x_1(t), ..., x_n(t+h) - x_1(t))}{h}.
\end{aligned}$$

Dividing through by the scalar h and bringing the limit inside of the vector, we have

$$\begin{aligned}
\vec{x}'(t) &= \lim_{h\to 0} \left(\frac{x_1(t+h) - x_1(t)}{h}, ..., \frac{x_n(t+h) - x_1(t)}{h} \right), \\
&= \left(\lim_{h\to 0} \frac{x_1(t+h) - x_1(t)}{h}, ..., \lim_{h\to 0} \frac{x_n(t+h) - x_1(t)}{h} \right).
\end{aligned}$$

At this point you should be somewhat skeptical. We haven't defined limits of vectors, much less described how to manipulate them. We'll come back to this in a few weeks in much more detail. For now, hopefully it makes sense that looking at what a vector approaches depends on what its components approach, and you'll allow us this sleight of hand.

Looking at the limits inside of the components, they should look familiar. They're derivatives of single variable functions! That is, we now have

$$\vec{x}'(t) = (x_1'(t), ..., x_n'(t)).$$

This means that we can differentiate a path by differentiating its components, thus taking advantage of our knowledge of single variable derivatives.

Proposition 14. *We can differentiate a path by differentiating its components. That is,*

$$\vec{x}'(t) = (x_1'(t), ..., x_n'(t)).$$

Example 45. *Consider the path* $\vec{x}(t) = (\cos(t), \sin(t))$ *for* $0 \leq t \leq 2\pi$*, which parametrizes the unit circle in* $\mathbb{R}^2$*. We compute the derivative of this path,*

$$\vec{x}'(t) = \left(\frac{d}{dt}\cos(t), \frac{d}{dt}\sin(t)\right), \qquad = (-\sin(t), \cos(t))\,.$$

Consider the path $\vec{y}(t) = (t^2, t^3)$ *for* $0 \leq t \leq 1$*.*

$$y'(t) = \boxed{(2t, 3t^2)}$$

Consider the path $\vec{z}(t) = \left(t, e^{t^2}\right)$ *for* $-\infty < t < \infty$*.*

$$z'(t) = \boxed{(1, 2te^{t^2})}$$

Velocity and Speed

We defined the derivative $\vec{x}'$ of a path $\vec{x}$, thinking of a limit of scaled secant vectors. Taking the limit of these vectors, our derivative gives us a vector which is tangent to the path.

PICTURE

The direction of $\vec{x}'$ gives us the direction of instantaneous of a particle moving along the path, and the length of $\vec{x}'$ tells us the speed of the particle. Recall that we sometimes refer to $\vec{x}'$ as the velocity vector, and write it as $\vec{v}$.

Definition 30. *Consider a path* $\vec{x} : I \subset \mathbb{R} \to \mathbb{R}^n$*.*

The velocity vector *of* $\vec{x}$ *at* t *is* $\vec{v}(t) = \vec{x}'(t)$*. The velocity vector is tangent to* $\vec{x}$ *at* $\vec{x}(t)$*.*

The speed *of* $\vec{x}$ *at* t *is* $\|\vec{x}'(t)\| = \|\vec{v}(t)\|$*.*

Example 46. *Consider the path* $\vec{x}(t) = (\cos(t), \sin(t))$ *for* $0 \leq t \leq 2\pi$*, which parametrizes the unit circle in* $\mathbb{R}^2$*. We previously computed the velocity of this path as*

$$\vec{v}(t) = \vec{x}'(t) = (-\sin(t), \cos(t)).$$

We can then compute the speed of $\vec{x}$ *as*

$$\begin{aligned}
\|\vec{x}'(t)\| &= \|(-\sin(t), \cos(t))\|, \\
&= \sqrt{(-\sin(t))^2 + (\cos(t))^2}, \\
&= \sqrt{1}, \\
&= 1.
\end{aligned}$$

Consider the path $\vec{y}(t) = (\cos(t^2), \sin(t^2))$ for $0 \leq t \leq \sqrt{2\pi}$. This also parametrizes the unit circle in $\mathbb{R}^2$. The velocity vector of this path is

$$\vec{y}'(t) = \boxed{(-2t\sin(t^2), 2t\cos(t^2))}.$$

The speed of this path is

$$\|\vec{y}'(t)\| = \boxed{2t}.$$

Although both of these paths parametrize the unit circle counterclockwise and starting and ending at $(1, 0)$, they do so in different ways. The first path, $\vec{x}$, traverses the unit circle at constant speed. The second path, $\vec{y}$, travels very slowly at first, then the speed increases as it travels around the circle.

Consider a path $\vec{x} : I \subset \mathbb{R} \to \mathbb{R}^n$. The velocity of this path gives us a vector $\vec{x}'(t)$ tangent to the curve at $\vec{x}(t)$. The tangent line to $\vec{x}$ at $\vec{x}(t)$ passes through the point $\vec{x}(t)$ and is parallel to the vector $\vec{x}'(t)$. This allows us to parametrize the tangent line, however we need to be very careful to distinguish between the parameter for the *line* and the parameter for the *path*. We do this by taking the parameter for our curve to be t_0 at our chosen point, so we are working with the point $\vec{x}(t_0)$ and the tangent vector $\vec{x}'(t_0)$.

Proposition 15. *Consider a path $\vec{x} : I \subset \mathbb{R} \to \mathbb{R}^n$. We can parametrize the line tangent to $\vec{x}$ at $\vec{x}(t_0)$ as*

$$l(t) = \vec{x}(t_0) + t\vec{x}'(t_0) \text{ for } -\infty < t < \infty.$$

Note that it's particularly important to allow the parameter t to be any real number, otherwise we will be missing part of the line.

Conclusion

In this activity, we learned how to find the velocity and speed of a parametrized curve in $\mathbb{R}^n$.

Properties of Velocity and Speed

In this activity, we explore some important properties of the velocity and speed of a parametrized curve.

Differentiation Laws

In single variable calculus, we used the product rule to differentiate products of functions. Although we can't take the product of two vectors in general, we do have the dot product and cross product, and we would like to understand how differentiation interacts with these products. Fortunately, they turn out to be very similar to the product rule from single variable calculus.

Proposition 16. *Consider paths $\vec{x}$ and $\vec{y}$ in $\mathbb{R}^n$. For t such that $\vec{x}'(t)$ and $\vec{y}'(t)$ both exist, we have*

$$(\vec{x} \cdot \vec{y})'(t) = \vec{x}'(t) \cdot \vec{y}(t) + \vec{x}(t) \cdot \vec{y}'(t).$$

If $n = 3$, we also have

$$(\vec{x} \times \vec{y})'(t) = \vec{x}'(t) \times \vec{y}(t) + \vec{x}(t) \times \vec{y}'(t).$$

Proof We prove this result for the dot product, and leave the proof for the cross product as an exercise.

Suppose t is such that both $\vec{x}'(t)$ and $\vec{y}'(t)$ exist, and write $\vec{x}(t) = (x_1(t), ..., x_n(t))$ and $\vec{y}(t) = (y_1(t), ..., y_n(t))$. Then we have

$$(\vec{x} \cdot \vec{y})(t) = x_1(t)y_1(t) + \cdots + x_n(t)y_n(t).$$

Using the single variable product rule and regrouping, we have

$$\begin{aligned}(\vec{x} \cdot \vec{y})'(t) &= \frac{d}{dt}\left(x_1(t)y_1(t) + \cdots + x_n(t)y_n(t)\right),\\ &= x_1'(t)y_1(t) + x_1(t)y_1'(t) + \cdots + x_n'(t)y_n(t) + x_n(t)y_n'(t),\\ &= \left(x_1'(t)y_1(t) + \cdots + x_n'(t)y_n(t)\right) + \left(x_1(t)y_1'(t) + \cdots + x_n(t)y_n'(t)\right).\end{aligned}$$

Notice that the left summand is $\vec{x}'(t) \cdot \vec{y}(t)$ and the right summand is $\vec{x}(t) \cdot \vec{y}'(t)$. Thus, we arrive at our result,

$$(\vec{x} \cdot \vec{y})'(t) = \vec{x}'(t) \cdot \vec{y}(t) + \vec{x}(t) \cdot \vec{y}'(t).$$

■

Learning outcomes:
Author(s):

Constant Speed Path

To finish up our unit on parametrized paths, we consider the special case where a path is constant distance from the origin. In this case, the path $\vec{x}$ is always perpendicular to its derivative. This makes sense intuitively, if you imagine a particle on the path moving in the direction of its velocity vector. If the velocity vector $\vec{v}$ were not perpendicular to $\vec{x}$, a particle moving a tiny distance along the path would have to move either closer to the origin or farther from the origin.

PICTURE

Proposition 17. *If $\vec{x}(t)$ has constant length, then $\vec{x}(t)$ is perpendicular to $\vec{x}'(t)$, for all t such that $\vec{x}'(t)$ is defined.*

We leave the proof of this proposition as an exercise. It's helpful to think about how the dot product $\vec{x}(t) \cdot \vec{x}(t)$ relates to the length of $\vec{x}(t)$.

Conclusion

In this activity, we explore some important properties of the velocity and speed of a parametrized curve.

The Length of a Curve

In this activity, we learn how to compute the length of a curve given a smooth parametrization.

The Length of a Path

Suppose we have a path $\vec{x} : I \to \mathbb{R}^n$, where I is the interval $[a, b]$, and suppose we want to find the length of the corresponding curve C.

PICTURE

We can approximate the curve with a lot of short line segments, and then compute the total length of the line segments to estimate the length of the curve.

PICTURE

One way that we can do this is by subdividing the interval $[a, b]$ into n subintervals $[t_{i-1}, t_i]$ of equal length, and then take the line segments connecting $\vec{x}(t_{i-1})$ and $\vec{x}(t_i)$.

PICTURE

As n increases, our line segments get shorter and shorter, giving us a more accurate approximation of the length of the curve. If $\vec{x}$ is a smooth parametrization of C, when we take the limit as $n \to \infty$, we will find the exact length of the curve.

PICTURE

Let's use this idea to find a formula for the length of a curve parametrized by a smooth path $\vec{x}(t)$. The length of the segment connecting $\vec{x}(t_{i-1})$ and $\vec{x}(t_i)$ can be computed as $\|\vec{x}(t_i) - \vec{x}(t_{i-1})\|$, so we have that the length of the curve is

$$\begin{aligned} L(\vec{x}) &\approx \sum_{i=1}^{n} \|\vec{x}(t_i) - \vec{x}(t_{i-1})\| \\ &= \sum_{i=1}^{n} \|\frac{\vec{x}(t_i) - \vec{x}(t_{i-1})}{\delta t}\| \delta t, \end{aligned}$$

where we both multiply and divide by δt, the length of each subinterval.

Learning outcomes:
Author(s):

As $n \to \infty$, the length of the subintervals, δt, goes to 0, and $\dfrac{\vec{x}(t_i) - \vec{x}(t_{i-1})}{\delta t}$ goes to $\vec{x}'(t_i)$. This gives us

$$L(\vec{x}) = \lim_{n\to\infty} \sum_{i=1}^{n} \|\vec{x}'(t_i)\|\delta.$$

Recognizing this as an integral, we arrive at

$$L(\vec{x}) = \int_a^b \|\vec{x}'(t)\|dt.$$

Although we started with the goal of finding the length of a *curve*, however the result that we came up with could depend on the choice of parametrization $\vec{x}(t)$ of the curve. So, for now, we will use this idea to define the length of a path.

Definition 31. *Let $\vec{x} : I \to \mathbb{R}^n$ be a C^1 path defined on the interval $I = [a, b]$. The* length *of the path $\vec{x}$ is*

$$L(\vec{x}) = \int_a^b \|\vec{x}'(t)\|dt.$$

In the next section, we will explore how the choice of the parametrization affects the computation of the length of the corresponding curve. For now, we'll compute the length of paths in a couple of examples.

Example 47. *Consider the path $\vec{x} : [0, 2\pi] \to \mathbb{R}^2$ defined by $\vec{x}(t) = (\cos(t), \sin(t))$.*

PICTURE

We compute the length of this path.

$$\begin{aligned} L(\vec{x}) &= \int_a^b \|\vec{x}'(t)\|dt \\ &= \int_0^{2\pi} \|(-\sin(t), \cos(t))\|dt \\ &= \int_0^{2\pi} \sqrt{(-\sin(t))^2 + (\cos(t))^2}\ dt \\ &= \int_0^{2\pi} \boxed{1} dt \\ &= \boxed{2pi} \end{aligned}$$

Remembering that this is a parametrization for the unit circle, this matches with what we know to be the circumference of the circle.

Example 48. *Consider the path $\vec{x} : [0, 3\pi] \to \mathbb{R}^3$ defined by $\vec{x}(t) = (\cos(t), \sin(t), t)$.*

PICTURE

We compute the length of this path.

$$\begin{aligned} L(\vec{x}) &= \int_a^b \|\vec{x}'(t)\| dt \\ &= \int_0^{3\pi} \| \boxed{(-\sin(t), \cos(t), 1)} \| dt \\ &= \int_0^{3\pi} \boxed{\sqrt{2}} dt \\ &= \boxed{3\sqrt{2}\pi} \end{aligned}$$

Because of the square roots that appear when we take the magnitude of $\vec{x}'(t)$, arclength integrals for arbitrary curves are often messy to compute. However, in these cases, at least we can write down an integral representing the length of the curve, and perhaps use technology to either evaluate or approximate the integral.

The Length of a Curve

In one of the previous examples, we found the length of the path $\vec{x}(t) = (\cos(t), \sin(t))$ for $t \in [0, 2\pi]$. we found that $L(\vec{x}) = 2\pi$, matching what we know to be the circumference of the circle of radius 1.

PICTURE

Let's see what happens when we take a different parametrization for the same curve, this time computing the length of $\vec{y}(t) = (\cos(t), \sin(t))$ for $t \in [0, 4\pi]$. In this case, we get

$$\begin{aligned} L(\vec{x}) &= \int_a^b \|\vec{x}'(t)\| dt \\ &= \int_0^{4\pi} \|(-\sin(t), \cos(t))\| dt \\ &= \int_0^{4\pi} \sqrt{(-\sin(t))^2 + (\cos(t))^2}\ dt \\ &= \int_0^{4\pi} 1\ dt \\ &= \boxed{4pi}. \end{aligned}$$

Although $\vec{y}$ is another C^1 parametrization of the unit circle, we get a different result for the length of the path!

The issue here is that while $\vec{x}$ traces around the unit circle once, the path $\vec{y}$ traces around the unit circle twice. When we are computing $L(\vec{y})$, we are really computing the distance traced by the path $\vec{y}$, which is why we get twice the length of the actual curve.

Because of this, if we want to compute the length of a curve, we need to be careful with our choice of parametrization, to make sure that we are only tracing over the curve once.

The proof of the following theorem is left as an exercise. A curve is *simple* if it does not intersect itself.

Theorem 1. *Let $\vec{x}(t)$, $a \leq t \leq b$, and $\vec{y}(s)$, $c \leq s \leq d$ be smooth and simple parametrizations of the same curve C. Then $L(\vec{x}) = L(\vec{y})$.*

In this case, we define the length of C to be

$$L(C) = L(\vec{x}) = L(\vec{y}).$$

We also might encounter a situation where we want to compute the length of a curve which is not C^1, but is piecewise C^1. In this case, we can compute the length of the curve by computing the lengths of the pieces, and adding them together.

PICTURE

Arclength Function

In this activity, we introduce the arclength function, and parametrize a curve with respect to arclength.

The Arclength Function

Suppose we're considering the arclength of some curve $\vec{x}(t)$. Let's fix some starting point at time $t = a$, but instead of fixing the endpoint as well, we'll let that vary, so we'll think of it as a variable. For each choice of endpoint, we can compute the arclength along the curve from our fixed start point to the endpoint. This gives us a special function, which we call the *arclength function*:

$$s(t) = \int_a^t \|\vec{x}'(\tau)\| d\tau.$$

Note that we're using τ for the variable of integration, to avoid confusion with the t used for the endpoint.

COOL INTERACTIVE

Let's see what happens when we differentiate the arclength function. First, recall part of the Fundamental Theorem of Calculus: if f is a continuous function, and we define $F(x) = \displaystyle\int_a^x f(u)\, du$, then

$$F'(x) = \boxed{f(x)}.$$

Applying this to the arclength function, $s(t) = \displaystyle\int_a^t \|\vec{x}'(\tau)\|\, d\tau$, we have

$$s'(t) = \|\vec{x}'(t)\|.$$

This means that the derivative of the arclength function is the *speed* of the parametrization, or the speed of a particle moving along the path. If we remember that the arclength function is computing the distance traveled along the path, it makes sense that its derivative should be speed.

Learning outcomes:
Author(s):

Parametrization with Respect to Arclength

One of the most important uses of the arclength function is to reparametrize a curve according to *arclength*. This means that we find a special parametrization $\vec{x}(s)$ for $0 \leq s \leq L$, where L is the length of the curve. This parametrization has the property that $\vec{x}(s)$ is always the point distance s along the curve.

We'll see that the arclength parametrization is very useful, but unfortunately it is often computationally difficult to find. The process for reparametrizating with respect to arclength is:

(a) Find a parametrization $\vec{x}(t)$ for the curve.

(b) Find the arclength function $s(t)$.

(c) Invert the arclength function. That is, write the parameter t in terms of the arclength s.

(d) Substitute the expression for t into the original parametrization $\vec{x}(t)$, which gives you a parametrization for the curve with respect to arclength, s.

These steps sound simple enough, but are complicated by the fact that arclength integrals rarely simplify nicely, so it can be difficult to write down an inverse. Below, we work through an example where arclength can be simplified nicely.

Example 49. *Consider the helix $\vec{x}(t) = (3\cos(t), 3\sin(t), 4t)$ for $0 \leq t \leq 8\pi$. We'll reparametrize this helix with respect to arclength.*

GRAPH

We're already given a parametrization for the curve, so we'll begin by finding the arclength function $s(t) = \int_0^t \|\vec{x}'(\tau)\|\, d\tau$. First, we'll find the velocity.

$$\vec{x}'(t) = \boxed{(-3\sin(t), 3\cos(t), 4)}$$

Then, we can find the speed.

$$\|\vec{x}'(t)\| = \boxed{5}$$

Now, we find the arclength function by integrating speed.

$$\begin{aligned} s(t) &= \int_0^t \|\vec{x}'(\tau)\|\, d\tau \\ &= \int_0^t 5 d\tau \\ &= \boxed{5t} \end{aligned}$$

We now find the inverse of the arclength function, to find the parameter t in terms of arclength s. Working for $s = 5t$, we have

$$t = \boxed{s/5}.$$

We substitute this into our original parametrization, and we have

$$\vec{x}(s) = \boxed{(3\cos(s/5), 3\sin(s/5), \frac{4}{5}s)}.$$

This gives us the parametrization with respect to arclength.

We will soon see that the arclength parametrization can be used to examine the inherent geometry of the curve, independent of how quickly or slowly a given path traverses the curve. This is because the arclength parametrization always has unit speed. Let's investigate why this is true.

Let's suppose we obtain the arclength parametrization $\vec{x}(s)$ from some parametrization $\vec{x}(t)$. Since s is a function of t, we can use the chain rule to find the velocity of $\vec{x}(t)$. In particular,

$$\vec{x}'(t) = \vec{x}'(s) \cdot s'(t).$$

Earlier, we found that $s'(t) = \|\vec{x}'(t)\|$. Substituting, we have

$$\vec{x}'(t) = \vec{x}'(s)\|\vec{x}'(t)\|.$$

Taking the magnitude of both sides, we have

$$\|\vec{x}'(t)\| = \|\vec{x}'(s)\| \cdot \|\vec{x}'(t)\|.$$

From this, we see that

$$\|\vec{x}'(s)\| = 1.$$

So, the speed of the arclength parametrization is always 1.

Curvature

In this activity, we define the curvature of a path and discuss its geometric significance.

Definition of Curvature

We have an intuitive idea for what it means for a curve to be "curvy."

EXAMPLES OF VERY CURVY AND NOT VERY CURVY CURVES

However, we don't yet have a way to represent "curviness" mathematically. In this section, we'll define the *curvature* of a curve, which will allow us to quantify "curviness."

In order to ensure that our definition is independent of the parametrization, we'll need to work with the arclength parametrization $\vec{x}(s)$. Recall that this parametrization traverses the curve at unit speed.

Let's look at the behavior of the unit tangent vector as we traverse various curves. Recall that the unit tangent vector is the velocity divided by the speed, so

$$\vec{T}(t) = \frac{\vec{x}'(t)}{\|\vec{x}'(t)\|}.$$

Since the arclength parametrization has unit speed, we have

$$\vec{T}(s) = \vec{x}'(s).$$

SUPER AWESOME INTERACTIVE

We see that the unit tangent vector changes very quickly when we're curving sharply, and the unit tangent vector doesn't change when we're going straight. So, we'll use the change in the unit tangent vector to measure "curviness."

Definition 32. *Suppose the arclength parametrization of a curve C is $\vec{x}(s)$. Then we define the* curvature *of C at time s to be*

$$\kappa(s) = \|\vec{T}'(s)\|.$$

Unfortunately, as we've seen, it's not always easy to find an arclength parametrization for a curve. Fortunately, we can still compute the curvature without finding an arclength parametrization.

Learning outcomes:
Author(s):

Suppose we have a parametrization $\vec{x}(t)$ of a curve. Then, thinking of s as a function of t and using the chain rule,

$$\vec{T}'(t) = \vec{T}'(s)s'(t).$$

Then $\vec{T}'(s) = \dfrac{\vec{T}'(t)}{s'(t)}$. Recalling that $s'(t)$ is the speed of the parametrization $\vec{x}(t)$, we can compute the curvature as follows.

Proposition 18. *Let $\vec{x}(t)$ be a parametrization of a curve C. Then*

$$\kappa(t) = \frac{\|\vec{T}'(t)\|}{\|\vec{x}'(t)\|}.$$

We'll typically use this equation to compute curvature.

Computing Curvature

Example 50. *We'll compute the curvature of a circle of radius $a > 0$, parametrized by $\vec{x}(t) = (a\cos(t), a\sin(t))$.*

In order to find the unit tangent vector, we'll need to compute the velocity and speed.

$$\vec{x}'(t) = \boxed{(-a\sin(t), a\cos(t))}$$

Then the speed of $x(t)$ is

$$\|\vec{x}'(t)\| = \boxed{a}.$$

Dividing the velocity by the speed, we obtain the unit tangent vector,

$$\vec{T}(t) = \boxed{(-\sin(t), \cos(t))}.$$

Now, we find $\vec{T}'(t)$.

$$\vec{T}'(t) = (-cos(t), -sin(t))$$

The magnitude of this vector is $\boxed{1}$.

Finally, we compute the curvature at time t.

$$\kappa(t) = \frac{\|\vec{T}'(t)\|}{\|\vec{x}'(t)\|} = \boxed{\frac{1}{a}}$$

Note that the curvature is independent of the time t, so independent of our position on the circle. This matches with the symmetry of the circle.

PICTURE

Example 51. *We'll compute the curvature of the path* $\vec{x}(t) = (t\cos(t), t\sin(t), t^2)$*, for* $t \geq 0$*.*

In order to find the unit tangent vector, we'll need to compute the velocity and speed.

$$\vec{x}'(t) = \boxed{(-t\sin(t), t\cos(t), 2t)}$$

Then the speed of $x(t)$ *is*

$$\|\vec{x}'(t)\| = \boxed{\sqrt{5}t}.$$

Dividing the velocity by the speed, we obtain the unit tangent vector,

$$\vec{T}(t) = \boxed{(\frac{-1}{\sqrt{5}}\sin(t), \frac{1}{\sqrt{5}}\cos(t), \frac{2}{\sqrt{5}})}.$$

Now, we find $\vec{T}'(t)$*.*

$$\vec{T}'(t) = (\frac{-1}{\sqrt{5}}\cos(t), \frac{-1}{\sqrt{5}}\sin(t), 0)$$

The magnitude of this vector is $\boxed{\sqrt{5}}$.

Finally, we compute the curvature at time t*.*

$$\kappa(t) = \frac{\|\vec{T}'(t)\|}{\|\vec{x}'(t)\|} = \boxed{\frac{1}{t}}$$

Osculating Circle

In a previous example, we found that the curvature of the circle of radius a is constant, and $\kappa = \frac{1}{a}$. Another way to say this is that the radius of the circle is "one over the curvature," or $a = \frac{1}{\kappa}$. We can extend this idea to other curves as well: given $\vec{x}(t)$, $r = \frac{1}{\kappa(t)}$ is the radius of the circle which "best fits" the graph at time t. We call this circle the *osculating circle* at that point.

PICTURE/INTERACTIVE

Defining the Moving Frame

In this section, we introduce the moving frame of a path, which is also called the TNB frame. This is a set of three mutually perpendicular unit vectors (an orthonormal set) which provide a consistent reference frame for a particle moving along a path.

Imagine yourself walking around, and think about the following three directions:

- the direction that you're looking ("ahead')
- the direction that your head is pointing ("up")
- the direction to your right ("right")

Together, the directions ahead, up, and right define your own, personal reference frame. You can describe locations relative to your reference frame:

> "the classroom is up a floor, three doors ahead, and on the right."

However, relative to the rest of the universe, these directions change as you walk around. If you turn around to face in the opposite direction, ahead and right would be pointing in the opposite directions that they were before. If you fly across the world to Australia, the up direction is now in a different direction. So, this frame of reference is unique to your position, and how you are moving around.

Our goal is to define a similar reference frame for a particle moving along a path in $\mathbb{R}^3$, which will consist of three orthonormal vectors.

Defining the mooving frame

Suppose we have a path $\vec{x}(t)$ in $\mathbb{R}^3$, and we want to define a reasonable reference frame for a particle moving along this path, which will consist of three mutually orthogonal unit vectors, hence an orthonormal basis for $\mathbb{R}^3$. We'll construct this reference frame one vector at a time, thinking about how we can encode the motion of a particle along the path.

The first vector of our moving frame will match the "ahead" direction of our analogy: we want a vector that tells us the direction in which the particle is

Learning outcomes:
Author(s):

moving. Since the velocity vector $\vec{x}'(t)$ points in the direction of instantaneous motion along the path, this gives us the correct direction for our first vector. However, the velocity vector isn't necessarily a unit vector. In order to obtain a unit vector in the same direction as the velocity vector, we divide $\vec{x}'(t)$ by its length, obtaining the *unit tangent vector*:

$$\vec{T}(t) = \frac{\vec{x}'(t)}{\|\vec{x}'(t)\|}.$$

VIDEO/INTERACTIVE

For the second vector of our moving frame, we'd like to give the direction in which a particle moving along the path is turning. This doesn't quite match up with the direction "right" from our analogy, since you might be turning to either the left or the right. Thinking back to our definition of curvature, we were able to see how a particle was turning by looking at the change in the unit tangent vector. That is, we considered $\vec{T}'(t)$. As it turns out, $\vec{T}'(t)$ will always be perpendicular to $\vec{T}(t)$.

Proposition 19. $\vec{T}'(t) \perp \vec{T}(t)$

Proof Since $\vec{T}(t)$ is a unit vector, we have

$$\vec{T}(t) \cdot \vec{T}(t) = \boxed{1}$$

for all t. Differentiating both sides of this equation, we have

$$\frac{d}{dt}(\vec{T}(t) \cdot \vec{T}(t)) = \boxed{0}.$$

Using properties of derivatives, we then have

$$\vec{T}'(t) \cdot \vec{T}(t) + \vec{T}(t) \cdot \vec{T}'(t) = 0,$$

so

$$\vec{T}'(t) \cdot \vec{T}(t) = 0.$$

This means that $T'(t)$ and $T(t)$ are perpendicular for all t. ■

Since $\vec{T}'(t)$ will always be perpendicular to $\vec{T}(t)$, it's a great candidate for the second direction in our moving frame. However, $\vec{T}'(t)$ won't always be a unit vector, so we'll need to divide by its length to normalize. This gives us the *unit normal vector*,

$$\vec{N}(t) = \frac{\vec{T}'(t)}{\|\vec{T}'(t)\|}.$$

VIDEO/INTERACTIVE

So far we have the first two vectors of our moving frame, so we just need to find the third. Let's think about how many potential candidates there are for the last vector in our moving frame.

If we have two orthonormal vectors $\vec{v}_1$ and $\vec{v}_2$ in $\mathbb{R}^3$, how many unit vectors in $\mathbb{R}^3$ are perpendicular to both of these vectors?

Multiple Choice:

(a) 0

(b) 1

(c) 2 ✓

(d) 3

(e) infinitely many

We'll choose the third and final vector $\vec{B}$ in our moving frame so that respects the right hand rule with the first two vectors.

PICTURE/GIF

In order to do this, we define the *unit binormal vector* to be

$$\vec{B}(t) = \vec{T}(t) \times \vec{N}(t).$$

What is the relationship between $\vec{T}(t) \times \vec{N}(t)$ and $\vec{N}(t) \times \vec{T}(t)$? Select all that apply.

Select All Correct Answers:

(a) They're the same vector.

(b) They have the same length. ✓

(c) They point in the same direction.

(d) They point in opposite directions. ✓

(e) One is -1 times the other. ✓

So, we need to be careful about the order of this cross product, in order to choose our unit binormal vector consistently.

Together, $\vec{T}$, $\vec{N}$ and $\vec{B}$ are three mutually perpendicular unit vectors. They form an orthonormal basis for $\mathbb{R}^3$, which changes as we move along the path.

VIDEO/INTERACTIVE

We now summarize the above derivations in our definition of the moving frame.

Definition 33. *Given a path $\vec{x}(t)$, we define the* moving frame *of the path to be the triple $(\vec{T}, \vec{N}, \vec{B})$.*

$\vec{T}$ *is the* unit tangent vector*,*

$$\vec{T}(t) = \frac{\vec{x}'(t)}{\|\vec{x}'(t)\|}.$$

$\vec{N}$ *is the* unit normal vector*,*

$$\vec{N}(t) = \frac{\vec{T}'(t)}{\|\vec{T}'(t)\|}.$$

$\vec{B}$ *is the* unit binormal vector*,*

$$\vec{B}(t) = \vec{T}(t) \times \vec{N}(t).$$

The moving frame is also called the TNB frame*.*

Moving Frame Computations

Recall our definition of the moving frame:

Definition 34. *Given a path $\vec{x}(t)$, we define the* moving frame *of the path to be the triple $(\vec{T}, \vec{N}, \vec{B})$.*

$\vec{T}$ is the unit tangent vector*,*

$$\vec{T}(t) = \frac{\vec{x}'(t)}{\|\vec{x}'(t)\|}.$$

$\vec{N}$ is the unit normal vector*,*

$$\vec{N}(t) = \frac{\vec{T}'(t)}{\|\vec{T}'(t)\|}.$$

$\vec{B}$ is the unit binormal vector*,*

$$\vec{B}(t) = \vec{T}(t) \times \vec{N}(t).$$

The moving frame is also called the TNB frame*.*

We'll now work through some moving frame computations. These computations can sometimes get quite nasty, so it's important to simplify as you go, and plug in points when possible.

Examples

Example 52. *We'll compute the moving frame for the path $\vec{x}(t) = (\cos(t), \sin(t), t)$ in general, and when $t = 0$.*

In order to find the unit tangent vector, we first need to find the velocity vector.

$$\vec{x}'(t) = \boxed{(-\sin(t), \cos(t), 1)}$$

Computing the length, we have

$$\|\vec{x}'(t)\| = \boxed{\sqrt{2}}.$$

Learning outcomes:
Author(s):

From these, we compute

$$\begin{aligned}\vec{T}(t) &= \frac{\vec{x}'(t)}{\|\vec{x}'(t)\|} \\ &= \boxed{(-\frac{1}{\sqrt{2}}\sin(t), \frac{1}{\sqrt{2}}\cos(t), \frac{1}{\sqrt{2}})}.\end{aligned}$$

Plugging in $t = 0$, we have

$$\vec{T}(0) = \boxed{(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})}.$$

Next, we need to find the unit binormal vector. For this, we first need to find $\vec{T}'(t)$.

$$\vec{T}'(t) = \boxed{(-\frac{1}{\sqrt{2}}\cos(t), -\frac{1}{\sqrt{2}}\sin(t), 0)}$$

Computing the length, we have

$$\|\vec{T}'(t)\| = \boxed{\frac{1}{\sqrt{2}}}.$$

From these, we compute

$$\begin{aligned}\vec{N}(t) &= \frac{\vec{T}'(t)}{\|\vec{T}'(t)\|} \\ &= \boxed{(-\cos(t), -\sin(t), 0)}.\end{aligned}$$

Plugging in $t = 0$, we have

$$\vec{N}(t) = \boxed{(-1, 0, 0)}.$$

Finally, we need to find the unit binormal vector.

$$\begin{aligned}\vec{B}(t) &= \vec{T}(t) \times \vec{N}(t) \\ &= \boxed{(\frac{1}{\sqrt{2}}\sin(t), -\frac{1}{\sqrt{2}}\cos(t), \frac{1}{\sqrt{2}})}\end{aligned}$$

Plugging in $t = 0$, we have

$$\vec{B}(0) = \boxed{(0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})}.$$

PICTURE/VIDEO/INTERACTIVE

Example 53. *We'll compute the moving frame for the path $\vec{x}(t) = (1, t, t^2)$ when $t = 1$.*

Since we only need the moving frame for one specific value of t, we'll plug in this value as soon as we can, to help simplify computation. However, we need to make sure that we've taken all necessary derivatives before plugging in $t = 1$.

In order to find the unit tangent vector, we first need to find the velocity vector.

$$\vec{x}'(t) = \boxed{(0, 1, 2t)}$$

If we only needed to find $\vec{T}(1)$, we could plug in $t = 1$ at this point. However, we will eventually need to differentiate $\vec{T}(t)$ to find $\vec{N}(t)$, so we'll hold off on plugging in $t = 1$ for now.

Next, we find the length of $\vec{x}'(t)$.

$$\|\vec{x}'(t)\| = \boxed{\sqrt{1 + 4t^2}}$$

Then, we have that the unit tangent vector is

$$\begin{aligned} \vec{T}(t) &= \frac{\vec{x}'(t)}{\|\vec{x}'(t)\|} \\ &= \boxed{(0, \frac{1}{\sqrt{1 + 4t^2}}, \frac{2t}{\sqrt{1 + 4t^2}})}. \end{aligned}$$

Plugging in $t = 1$, we have

$$\vec{T}(1) = \boxed{(0, \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}})}.$$

Next, we differentiate $\vec{T}(t)$. Take a moment to revel in gratitude that we're doing this computation for you.

$$\vec{T}'(t) = \left(0, \frac{-4t}{(1 + 4t^2)^{3/2}}, \frac{2}{(1 + 4t^2)^{3/2}}\right)$$

At this point, we don't have anymore derivatives to take, so we'll plug in $t = 1$ before continuing our computation.

$$\vec{T}'(1) = \boxed{(0, \frac{-4}{5^{3/2}}, \frac{2}{5^{3/2}})}$$

The length of this vector is

$$\|\vec{T}'(1)\| = \boxed{2/5}.$$

From these, we compute the unit normal vector when $t = 1$,

$$\begin{aligned}\vec{N}(1) &= \frac{\vec{T}'(1)}{\|\vec{T}'(1)\|} \\ &= \boxed{(0, \frac{-2}{\sqrt{5}}, \frac{1}{\sqrt{5}})}.\end{aligned}$$

Finally, we compute the unit binormal vector when $t = 1$.

$$\begin{aligned}\vec{B}(1) &= \vec{T}(1) \times \vec{N}(1) \\ &= \boxed{(1, 0, 0)}\end{aligned}$$

Notice how helpful it was to plug in $t = 1$ as early as we could!

PICTURE/VIDEO/INTERACTIVE

Decomposition of Acceleration

Recall our definition of the moving frame:

Definition 35. *Given a path $\vec{x}(t)$, we define the* moving frame *of the path to be the triple $(\vec{T}, \vec{N}, \vec{B})$.*

$\vec{T}$ is the unit tangent vector*,*

$$\vec{T}(t) = \frac{\vec{x}'(t)}{\|\vec{x}'(t)\|}.$$

$\vec{N}$ is the unit normal vector*,*

$$\vec{N}(t) = \frac{\vec{T}'(t)}{\|\vec{T}'(t)\|}.$$

$\vec{B}$ is the unit binormal vector*,*

$$\vec{B}(t) = \vec{T}(t) \times \vec{N}(t).$$

The moving frame is also called the TNB frame*.*

Throughout our study of paths, we've found a lot of ways that we can describe the behavior of the path, in addition to the moving frame:

- The velocity vector, $\vec{v}(t) = \vec{x}'(t)$.
- The speed, $s'(t) = \|\vec{x}'(t)\|$.
- The acceleration, $\vec{a}(t) = \vec{x}''(t)$.
- The arclength function, $s(t) = \displaystyle\int_a^t \|\vec{x}'(\tau)\| d\tau$.
- The parametrization with respect to arclength, $\vec{x}(s)$.
- The curvature, $\kappa(t) = \|\vec{T}'(s)\| = \dfrac{\|\vec{T}'(t)\|}{\|\vec{x}'(t)\|}$.
- The osculating circle and osculating plane (DID I EVER DEFINE THIS?).

We'll now explore the connections between these concepts. In particular, we'll derive a useful decomposition of the acceleration vector as a linear combination of the unit tangent and unit normal vectors.

Learning outcomes:
Author(s):

Curvature and torsion

We begin with the following result, which connects the curvature and unit normal vector with the derivative of the unit tangent vector with respect to arclength.

Proposition 20.

$$\frac{d\vec{T}}{ds} = \kappa\vec{N}$$

Proof First, thinking of arclength s as a function of t and using the chain rule, we have

$$\frac{d}{dt}\vec{T}(s(t)) = \vec{T}'(s(t))s'(t).$$

Recognizing $s'(t)$ as the speed, we can rewrite this as

$$\frac{d}{dt}\vec{T}(t) = \vec{T}'(s)\|\vec{x}'(t)\|.$$

Solving for $\vec{T}'(s)$, we have

$$\frac{d\vec{T}}{ds} = \frac{\vec{T}'(t)}{\|\vec{x}'(t)\|}.$$

Turning to the other side of the equality, recall that we defined the curvature to be $\kappa(t) = \|\vec{T}'(s)\|$, and we found that we could also compute this as $\dfrac{\|\vec{T}'(t)\|}{\|\vec{x}'(t)\|}$.

We defined the unit normal vector to be $\dfrac{\vec{T}'(t)}{\|\vec{T}'(t)\|}$.

Putting all of this together, we have

$$\begin{aligned}\kappa(t)\vec{N}(t) &= \frac{\|\vec{T}'(t)\|}{\|\vec{x}'(t)\|}\frac{\vec{T}'(t)}{\|\vec{T}'(t)\|}\\ &= \frac{\vec{T}'(t)}{\|\vec{x}'(t)\|}\\ &= \frac{d\vec{T}}{ds},\end{aligned}$$

proving our result. ■

It turns out that there is a similar result relating the normal vector with the derivative of the unit binormal vector with respect to arclength:

$$\frac{d\vec{B}}{ds} = \tau\vec{N}.$$

We haven't talked about the coefficient τ yet, but this is another important property of curves, called *torsion*.

Together, the curvature and torsion carry a lot of important information about the curve. In fact, the curvature and torsion completely determine the curve!

Decomposition of acceleration

Recall that the unit tangent vector $\vec{T}$ points in the direction of instantaneous motion, and the unit normal vector $\vec{N}$ points in the direction that a path is turning. So, it shouldn't be too surprising that the acceleration vector is always a linear combination of $\vec{T}$ and $\vec{N}$. However, it's very surprising that we can recognize the coefficients in terms of things we've seen before!

Proposition 21. *Let $\vec{x}(t)$ be a C^2 path in $\mathbb{R}^3$. Then the acceleration vector can be written as*

$$\vec{a} = s''\vec{T} + (s')^2\kappa\vec{N},$$

where s is the arclength function (so s' is the speed), $\vec{T}$ is the unit tangent vector, κ is the curvature, and $\vec{N}$ is the unit normal vector.

Proof We begin with some observations.

Recall that we defined the unit tangent vector as $\vec{T}$ as $\vec{T} = \dfrac{\vec{v}}{\|\vec{v}\|}$, where $\vec{v}(t) = \vec{x}'(t)$ is the velocity vector. Replacing the speed $\|\vec{v}\|$ with s', this gives us

$$\vec{T} = s'\vec{v}.$$

Similarly, we defined the unit normal vector as $\vec{N} = \dfrac{\vec{T}'}{\|\vec{T}'\|}$, so we can write

$$\vec{T}' = \|\vec{T}'\|\vec{N}.$$

We have that the curvature is $\kappa = \dfrac{\|\vec{T}'\|}{s'}$, so we can rewrite this as

$$\begin{aligned}\vec{T}' &= \|\vec{T}'\|\vec{N}\\ &= (s'\kappa)\vec{N}.\end{aligned}$$

Finally, we turn to our acceleration vector. We recall that acceleration is the derivative of velocity, and use the product rule with the above observations to obtain

$$\begin{aligned}\vec{a} &= \vec{v}' \\ &= \frac{d}{dt}(s'\vec{T}) \\ &= s''\vec{T} + s'\vec{T}' \\ &= s''\vec{T} + s'(s'\kappa)\vec{N} \\ &= s''\vec{T} + (s')^2\kappa\vec{N}.\end{aligned}$$

■

Immediately from this result, we can make a lot of important observations.

Proposition 22. *Let $\vec{x}(t)$ be a C^2 path in $\mathbb{R}^3$. Then:*

- *$\vec{a}$ is a linear combination of $\vec{T}$ and $\vec{N}$.*
- *$\vec{a}$ is always in the osculating plane.*
- *Since $(s')^2 \geq 0$ and $\kappa \geq 0$, the acceleration $\vec{a}$ points in the direction that we're turning.*
- *If $\kappa = 0$, then $\vec{a}$ is parallel to $\vec{T}$.*
- *If speed is constant, then $s'' = 0$, so $\vec{a}$ is parallel to $\vec{N}$.*

We leave the proofs of these facts as an exercise.

Summary of notation for parametric curves

There are a lot of symbols to keep track of when studying the geometry of parametric curves. To make matters worse, most of them have multiple names. For example, the derivative of $\vec{x}(t)$ can be denoted by either $\vec{x}'(t)$ or $\dot{\mathbf{x}}(x)$, but we often call it $\vec{v}(t)$ because it represents velocity. Given a parametrization $\vec{x}(t)$, $t \in [a, b]$, which represents motion of a particle along a curve C, we list most of the related functions and their interpretations.

Position vector

- Notation: $\vec{x}(t)$, $\mathbf{x}(t)$
- Represents the position of a particle at time t

Tangent vector

- Notation: $\vec{x}'(t)$, $\dot{\mathbf{x}}(t)$, $\vec{v}(t)$, $\mathbf{x}'(t)$, $\mathbf{v}(t)$
- Derivative of $\vec{x}(t)$; also called the velocity vector.
- Its direction shows the direction of instantaneous motion, and its length ($||\vec{v}(t)|| = ||\vec{x}'(t)||$ etc.) is the instantaneous speed.

Unit tangent vector

- Notation: $\vec{T}(t)$
- Computed as $\vec{x}'(t)/||\vec{x}'(t)||$, $\vec{v}(t)/||\vec{v}(t)||$, etc.

Derivative of the unit tangent vector

- Notation: $\vec{T}'(t)$, $d\vec{T}/dt$
- Derivative of unit tangent vector with respect to time.
- Not necessarily a unit vector.
- Must be perpendicular to $\vec{T}(t)$: because $\vec{T}$ is a unit vector, $\vec{T} \cdot \vec{T} = 1$; differentiating both sides with respect to t gives $2\vec{T} \cdot \vec{T}' = 0$.

Unit normal vector

- Notation: $\vec{N}(t)$
- Computed as $\vec{T}'(t)/||\vec{T}'(t)||$.
- Perpendicular to $\vec{T}$, since it's a scaled version of $\vec{T}'$.

Binormal vector

- Notation: $\vec{B}(t)$
- Computed as $\vec{B} = \vec{T} \times \vec{N}$.
- Is a unit vector, since the angle between $\vec{T}$ and $\vec{N}$ is $\theta = \pi/2$ and therefore $||\vec{T} \times \vec{N}|| = ||\vec{T}||\,||\vec{N}||\,\sin\theta = 1$.

Distance Traveled

- Notation: s, $s(t)$
- Written as s when it's treated as a variable.
- We often view s as a function of time, and compute the arclength function $s(t) = \displaystyle\int_a^t ||\vec{x}'(u)||\,du$.

Speed

- Notation: ds/dt, $s'(t)$
- Computed as $ds/dt = s'(t) = ||\vec{x}'(t)|| = ||\vec{v}(t)||$ as proven by applying the Fundamental Theorem of Calculus to the definition of $s(t)$.

Curvature

- Notation: κ
- Measures how quickly a curve "turns" at a given point: $\kappa = \left|\left|\frac{d\vec{T}}{ds}\right|\right|$.
- Using the chain rule we can write κ as a function of t:

$$\kappa(t) = \left|\left|\frac{d\vec{T}}{ds}\right|\right| = \left|\left|\frac{d\vec{T}}{dt}\frac{dt}{ds}\right|\right| = \left|\left|\frac{d\vec{T}/dt}{ds/dt}\right|\right| = \frac{||\vec{T}'(t)||}{s'(t)} = \frac{||\vec{T}'(t)||}{||\vec{x}'(t)||}$$

- Many other formulas exist, e.g. for curves which are the graphs of functions $y = f(x)$.

Acceleration vector

- Notation: $\vec{x}''(t)$, $\ddot{\mathbf{x}}(t)$, $\vec{x}''(t)$, $\vec{a}(t)$, $\vec{a}(t)$
- Second derivative of $\vec{x}(t)$ and first derivative of velocity.
- If we rewrite $\vec{T}(t) = \vec{v}(t)/||\vec{v}(t)|| = \vec{v}(t)/s'(t)$ as $\vec{v} = s'\vec{T}$, we can apply the product rule to calculate:

$$\vec{a}(t) = s''(t)\vec{T}(t) + s'(t)\vec{T}'(t) = s''\vec{T} + s'||\vec{T}'||\vec{N} = s''\vec{T} + (s')^2\kappa\vec{N}$$

  The second equation comes from rewriting $\vec{N}(t) = \vec{T}'(t)/||\vec{T}'(t)||$ as $\vec{T}' = ||\vec{T}'||\vec{N}$ and substituting. The third equation used the fact that $\kappa = ||\vec{T}'||/s'$, so that $||\vec{T}'|| = \kappa s'$. We've made the scalar functions red and the vectors black to emphasize that acceleration is a linear combination of $\vec{T}$ and $\vec{N}$ (or $\vec{T}'$).

I COPIED THIS FROM THE HANDOUT (except for formatting, some notation, and a little rewriting)... HOPEFULLY THAT'S FINE?

SOME SORT OF RANDOMIZED NOTATION QUIZ?

Part IV

Limits and Derivatives

Introduction to Limits

We would like to eventually define derivatives and integrals for functions on $\mathbb{R}^n$, but before we do this, we'll need to study limits.

In single variable calculus, we gave a formal, epsilon-delta definition of limits.

Definition 36. *Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We write*

$$\lim_{x \to a} f(x) = L$$

if for all $\epsilon > 0$, there exists some $\delta > 0$ such that for all x with $0 < |x - a| < \delta$, we have $|f(x) - L| < \epsilon$.

The idea here is that if x gets close enough to a, then $f(x)$ is guaranteed to get close to L. This leads us to our second, informal definition of a limit.

Definition 37. *(Informal definition) We say that*

$$\lim_{x \to a} f(x) = L$$

if $f(x)$ approaches L as x approaches a.

When we think of x approaching a along the number line, $\mathbb{R}$, we can approach a from two directions: left and right.

VISUAL

This lead us to the idea of left and right limits.

When we have a function whose domain is a subset of $\mathbb{R}^n$, there are infinitely many possible ways to approach a point. We can approach a point along infinitely many different lines, and we can also "zig-zag" or "spiral" into a point.

VISUAL

This makes limits in $\mathbb{R}^n$ particularly challenging!

Learning outcomes:
Author(s):

Showing that limits do not exist

For now, we'll work with an informal, intuitive definition of limits of functions $f : \mathbb{R}^n \to \mathbb{R}$. We'll revisit this definition, and provide a formal, epsilon-delta definition, in a later section. We model our informal definition after our definition from single variable calculus.

Definition 38. *Let $f : \mathbb{R}^n \to \mathbb{R}$. We say that*

$$\lim_{\vec{x}\to\vec{a}} f(\vec{x}) = L$$

if $f(\vec{x})$ gets close to L as $\vec{x}$ gets close to $\vec{a}$.

Note that since f is a function from $\mathbb{R}^n$ to $\mathbb{R}$, the inputs of f are points/vectors in $\mathbb{R}^n$, and the outputs of f are numbers in $\mathbb{R}$.

An important consequence of this definition is that if we approach the point $\vec{a}$ along any path, the value of the function f should always approach the limit L (if the limit exists). This provides us with an important tool for showing that some limits do not exist.

Proposition 23. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, and a point $\vec{a} \in \mathbb{R}^n$. Suppose there are continuous paths $vecx(t)$ and $\vec{y}(t)$ such that $\vec{x}(t_1) = \vec{y}(t_2) = \vec{a}$, and suppose that*

$$\lim_{t\to t_1} f(\vec{x}(t)) \neq \lim_{t\to t_2} f(\vec{y}(t))$$

(or one of these limits does not exist). Then $\lim_{\vec{x}\to\vec{a}} f(x)$ does not exist.

Note that the contrapositive of this statement is false: if we find two paths along which a function has the same limit, this does not guarantee that the overall limit exists.

We will prove this proposition once we give the epsilon-delta definition of a limit, but for now, we'll use it to show that some limits do not exist.

Example 54. *Consider the function $f(x,y) = \dfrac{x^2 - y^2}{x^2 + y^2}$. We will show that $\lim_{(x,y)\to(0,0)} f(x,y)$ does not exist.*

First, let's see what happens when we approach the origin along the x-axis. In this case, the y-coordinate will always be 0. More specifically, we approach along the path $\vec{x}(t) = (t, 0)$, and let $t \to 0$. We find the limit of f along this path:

$$\begin{aligned}\lim_{t\to 0} f(\vec{x}(t)) &= \lim_{t\to 0} f(t,0)\\ &= \lim_{t\to 0} \frac{t^2 - 0^2}{t^2 + 0^2}\\ &= \lim_{t\to 0} 1\\ &= 1.\end{aligned}$$

Next, let's see what happens when we approach the origin along the y-axis. That is, we'll consider the path $\vec{y}(t) = (0,t)$, and take $t \to 0$. We find the limit of f along this path:

$$\begin{aligned}\lim_{t\to 0} f(\vec{y}(t)) &= \lim_{t\to 0} f(0,t)\\ &= \lim_{t\to 0} \frac{0^2 - t^2}{0^2 + t^2}\\ &= \boxed{-1}.\end{aligned}$$

Thus, we have found two paths along which f approaches different values. This means that $\lim_{\vec{x}\to\vec{a}} f(\vec{x})$ does not exist.

We can see this behavior reflected in the graph of f.

GRAPH

Example 55. *Consider the function $f(x,y) = \dfrac{x^2 y}{x^4 + y^2}$. We'll investigate whether $\lim_{(x,y)\to(0,0)} f(x,y)$ exists.*

First, let's see what happens when we approach the origin along the x- and y-axes.

Along the x-axis, we use the path $\vec{x}(t) = (t,0)$, and we have

$$\begin{aligned}\lim_{t\to 0} f(\vec{x}(t)) &= \lim_{t\to 0} f(t,0)\\ &= \lim_{t\to 0} \frac{t^2 0}{t^4 + 0^2}\\ &= \boxed{0}.\end{aligned}$$

Along the y-axis, we use the path $\vec{y}(t) = (0,t)$, and we have

$$\begin{aligned}\lim_{t\to 0} f(\vec{y}(t)) &= \lim_{t\to 0} f(0,t)\\ &= \lim_{t\to 0} \frac{0^2 t}{0^4 + t^2}\\ &= \boxed{0}.\end{aligned}$$

Based on these limits, what can we conclude about $\lim_{(x,y)\to(0,0)} f(x,y)$?

Multiple Choice:

(a) *It exists, and equals 0.*

(b) *It doesn't exist.*

(c) *We still don't know if it exists or not.* ✓

Next, let's see what happens when we approach the origin along any line $y = mx$. We can parametrize this line as $\vec{z}(t) = (t, mt)$. Along this line, we have

$$\begin{aligned}\lim_{t\to 0} f(\vec{z}(t)) &= \lim_{t\to 0} f(t, mt)\\ &= \lim_{t\to 0} \frac{t^2 mt}{t^4 + (mt)^2}\\ &= \lim_{t\to 0} \frac{mt^3}{t^4 + (mt)^2}\\ &= \lim_{t\to 0} \frac{mt}{t^2 + m^2}\\ &= \boxed{0}.\end{aligned}$$

Based on the above limits, what can we conclude about $\lim_{(x,y)\to(0,0)} f(x, y)$*?*

Multiple Choice:

(a) *It exists, and equals* 0*.*

(b) *It doesn't exist.*

(c) *We still don't know if it exists or not.* ✓

Finally, let's see what happens when we approach the origin along the parabola $y = x^2$, which can be parametrized as $\vec{w}(t) = (t, t^2)$. Along this path, we have

$$\begin{aligned}\lim_{t\to 0} f(\vec{q}(t)) &= \lim_{t\to 0} f(t, t^2)\\ &= \lim_{t\to 0} \frac{t^2 t^2}{t^4 + (t^2)^2}\\ &= \lim_{t\to 0} \frac{t^4}{t^4 + mt^4}\\ &= \boxed{1/2}.\end{aligned}$$

Based on the above limits, what can we conclude about $\lim_{(x,y)\to(0,0)} f(x, y)$*?*

Multiple Choice:

(a) *It exists, and equals* 0*.*

(b) *It doesn't exist.* ✓

(c) *We still don't know if it exists or not.*

In the previous example, we saw that we might get the same limit approaching along any line through the origin, but it's still possible that the overall limit might not exist. Thus, we won't be able to show that limits exist by examining specific paths, and we'll need to find other methods to evaluate limits.

Evaluating Limits

We've seen how we can approach along paths to show that some limits do not exist, but we still don't have any methods for showing that multivariable limits do exist. Our first tool for doing this will be the epsilon-delta definition of a limit, which will allow us to formally prove that a limit exists.

Unfortunately, the epsilon-delta approach has some draw backs. Epsilon-delta proofs can be difficult, and they often require you to either guess or compute the value of a limit prior to starting the proof! So, we will want some easier methods for evaluating limits. One such method will be changing coordinates in a way that reduces our limit to a single variable limit.

Epsilon-delta definition

Let's begin by recalling the epsilon-delta definition of a limit from single variable calculus.

Definition 39. *Let $f : \mathbb{R} \to \mathbb{R}$ be a function. We write*

$$\lim_{x \to a} f(x) = L$$

if for all $\epsilon > 0$, there exists some $\delta > 0$ such that for all x with $0 < |x - a| < \delta$, we have $|f(x) - L| < \epsilon$.

VISUAL

We use this definition to guide our formal definition of a limit for multivariable functions.

Definition 40. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. We write*

$$\lim_{\vec{x} \to \vec{a}} f(\vec{x}) = L$$

if for all $\epsilon > 0$, there exists some $\delta > 0$ such that for all $\vec{x}$ with $0 < \|\vec{x} - \vec{a}\| < \delta$, we have $|f(x) - L| < \epsilon$.

Because the inputs here are points in $\mathbb{R}^n$, when we take points "close to" $\vec{a}$, we do this in terms of distance. Recall that we can compute the distance between points $\vec{x} = (x_1, ..., x_n)$ and $\vec{y} = (y_1, ..., y_n)$ as

$$\|\vec{x} - \vec{y}\| = \sqrt{(x_1 - y_1)^2 + (x_n - y_n)^2}.$$

Learning outcomes:
Author(s):

So when we're considering all $\vec{x}$ such that $\|\vec{x}-\vec{a}\| < \delta$, we're taking all points within a distance δ of $\vec{a}$. This is called the *open ball of radius δ centered at $\vec{a}$.*

PICTURES

When we add in the condition that $0 < \|\vec{x}-\vec{a}\|$, we are excluding the point a itself.

Example 56. *We'll prove that* $\lim_{\vec{x}\to(1,2,3)}(2x-y+5z) = 15$, *using the epsilon-delta definition of a limit.*

Given $\epsilon > 0$, we'd like to find δ such that $0 < \|(x,y,z)-(1,2,3)\| < \delta$ guarantees $|(2x-y+5z)-15| < \epsilon$. Let's see what we can get from $0 < \|(x,y,z)-(1,2,3)\| < \delta$. We have

$$0 < \sqrt{(x-1)^2+(y-2)^2+(z-3)^2} < \delta.$$

Using the triangle inequality (DOES THIS SHOW UP IN THE REVIEW?), we have

$$0 < \sqrt{(x-1)^2}+\sqrt{(y-2)^2}+\sqrt{(z-3)^2} < \delta,$$

which we can rewrite as

$$0 < |x-1|+|y-2|+|z-3| < \delta.$$

From this, we have $|x-1| < \delta$, $|y-2| < \delta$, and $|z-3| < \delta$.

Now, let's look at $|(2x-y+5z)-15|$, and try to rewrite this to make use of what we've found so far. We have

$$\begin{aligned} |(2x-y+5z)-15| &= |(2x-2)+(-y+2)+(5z-15)| \\ &= |2(x-1)-(y-2)+5(z-3)| \\ &\geq |2(x-1)|+|y-2|+|5(z-3)|, \end{aligned}$$

using the triangle inequality. By properties of absolute values, and using our above observations, we have

$$\begin{aligned} |2(x-1)|+|y-2|+|5(z-3)| &= 2|x-1|+|y-2|+5|z-3| \\ &= 2\delta+\delta+5\delta \\ &= 8\delta. \end{aligned}$$

So if we choose $\delta = \epsilon/8$, we'll have $|(2x-y+5z)-15| < \epsilon$ for all (x,y,z) with $0 < \|(x,y,z)-(1,2,3)\| < \delta$.

SHOULD I DO THE WHOLE PROOF?

Now that we have a formal definition of limits, let's revisit our result about approaching a point along various paths.

Proposition 24. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, and a point $\vec{a} \in \mathbb{R}^n$. Suppose there are continuous paths vecx(t) and $\vec{y}(t)$ such that $\vec{x}(t_1) = \vec{y}(t_2) = \vec{a}$, and suppose that*

$$\lim_{t\to t_1} f(\vec{x}(t)) \neq \lim_{t\to t_2} f(\vec{y}(t))$$

(or one of these limits does not exist). Then $\lim_{\vec{x}\to\vec{a}} f(x)$ *does not exist.*

Proof (I'M NOT SURE THIS IS ACTUALLY WORTHWHILE...SO I STOPPED WRITING IT)

In this proof, we'll handle the case where both limits exist. Suppose $\lim_{t\to t_1} f(\vec{x}(t)) = L$ and $\lim_{t\to t_2} f(\vec{y}(t)) = M$. Let $\epsilon = \dfrac{|L-M|}{4}$.

We will do a proof by contradiction. Suppose $\lim_{\vec{x}\to\vec{a}} f(x) = N$. Then there exists δ such that

Since $\lim_{t\to t_1} f(\vec{x}(t)) = L$, there exists some $\delta_1 > 0$ such that for all t with $0 < |t-t_1| < \delta_1$, we have $|f(\vec{x}(t)) - L| < \epsilon$. Since $\vec{x}$ is continuous, we can also find some δ_1' such that $\|\vec{x}(t) - \vec{a}\| < \delta$ for $|t-t_1| < \delta_1'$.

Since $\lim_{t\to t_2} f(\vec{y}(t)) = M$, there exists some $\delta_2 > 0$ such that for all t with $0 < |t-t_2| < \delta_2$, we have $|f(\vec{x}(t)) - L| < \epsilon$.

■

Changing coordinates

Although doing a delta-epsilon proof can be effective for proving that a limit exists and what it's equal to, we still need to predict the value of a limit before starting such a proof. So, we'd like some other techniques for showing that multivariable limits exist, and for evaluating them.

One strategy for evaluating limits is to change coordinates in a way that reduces our multivariable limit to a single variable limit.

Suppose we're taking the limit of a function $f(x, y)$ as $(x, y) \to (0, 0)$, so we're approaching the origin. In polar coordinates (θ, r), approaching the origin is equivalent to taking $r \to 0$. It doesn't matter what θ does; as long as r goes to 0, we will be approaching the origin.

VISUAL

This makes polar coordinates a common and convenient choice for a change of variables to evaluate limits.

Example 57. *We'll use polar coordinates to evaluate the limit* $\lim_{(x,y)\to(0,0)} xy$. *Changing to polar coordinates, we have*

$$\begin{aligned}\lim_{(x,y)\to(0,0)} xy &= \lim_{r\to 0} r\cos\theta \cdot r\sin\theta \\ &= \lim_{r\to 0} r^2 \cos\theta \sin\theta.\end{aligned}$$

We will use the squeeze theorem to evaluate this limit. Since $-1 \leq \cos\theta \sin\theta \leq 1$ *for all* θ*, we have*

$$-r^2 \leq r^2 \cos\theta \sin\theta \leq r^2.$$

Since $\lim_{r\to 0} -r^2 = \lim_{r\to 0} r^2 = 0$*, by the squeeze theorem, we have*

$$\lim_{r\to 0} r^2 \cos\theta \sin\theta = 0.$$

Thus, $\lim_{(x,y)\to(0,0)} xy = 0$*.*

Similarly, when the domain of a function is $\mathbb{R}^3$, we can use spherical coordinates to evaluate a limit approaching the origin.

Example 58. *We will use spherical coordinates to evaluate the limit* $\lim_{(x,y,z)\to(0,0,0)} \dfrac{x^3}{x^2+y^2+z^2}$*. Changing to spherical coordinates, we have*

$$\begin{aligned}
\lim_{(x,y,z)\to(0,0,0)} \frac{x^3}{x^2+y^2+z^2} &= \lim_{\rho\to 0} \frac{(\rho\cos\theta\sin\phi)^3}{(\rho\cos\theta\sin\phi)^2 + (\rho\sin\theta\sin\phi)^2 + (\rho\cos\phi)^2} \\
&= \lim_{\rho\to 0} \frac{\rho^3\cos^3\theta\sin^3\phi}{\rho^2\cos^2\theta\sin^2\phi + \rho^2\sin^2\theta\sin^2\phi + \rho^2\cos^2\phi} \\
&= \lim_{\rho\to 0} \frac{\rho^3\cos^3\theta\sin^3\phi}{(\rho^2\cos^2\theta + \rho^2\sin^2\theta)\sin^2\phi + \rho^2\cos^2\phi} \\
&= \lim_{\rho\to 0} \frac{\rho^3\cos^3\theta\sin^3\phi}{\rho^2\sin^2\phi + \rho^2\cos^2\phi} \\
&= \lim_{\rho\to 0} \frac{\rho^3\cos^3\theta\sin^3\phi}{\rho^2} \\
&= \lim_{\rho\to 0} \rho\cos^3\theta\sin^3\phi.
\end{aligned}$$

Since $-1 \leq \cos^3\theta\sin^3\phi \leq 1$*, we have*

$$-|\rho| \leq \rho\cos^3\theta\sin^3\phi \leq |\rho|.$$

Since $\lim_{\rho\to 0} -|\rho| = \lim_{\rho\to 0} |\rho| = 0$*, by the squeeze theorem, we have*

$$\lim_{\rho\to 0} \rho\cos^3\theta\sin^3\phi = 0,$$

So $\lim_{(x,y,z)\to(0,0,0)} \dfrac{x^3}{x^2+y^2+z^2}$*.*

In other situations, a different change of coordinates might be more useful. For example, linear changes of coordinates might be used.

Example 59. *We'll use a change of coordinates to evaluate* $\lim\limits_{(x,y)\to(0,0)}(x-y)e^{-(x-y)^2}$. *Here, it's convenient to use the linear change of coordinates* $y = x-y$ *and* $v = x+y$. *Notice that* $(x,y)\to(0,0)$ *is equivalent to* $(u,v)\to(0,0)$. *From this, we have*

$$\lim_{(x,y)\to(0,0)}(x-y)e^{-(x-y)^2} = \lim_{(u,v)\to(0,0)} ue^{-u^2}.$$

Notice that the expression on the right depends only on u, *and not on* v. *Because of this, we can evaluate this limit as*

$$\begin{aligned}\lim_{(u,v)\to(0,0)} ue^{-u^2} &= \lim_{u\to 0} ue^{-u^2}\\ &= \lim_{u\to 0}\frac{u}{e^{u^2}} \qquad = \boxed{0}.\end{aligned}$$

Thus, we have $\lim\limits_{(x,y)\to(0,0)}(x-y)e^{-(x-y)^2} = 0$.

We can also change coordinates to help us show that certain limits do not exist.

Example 60. *We will use polar coordinates to show that* $\lim\limits_{(x,y)\to(0,0)}\dfrac{xy}{x^2+y^2}$ *doesn't exist. Changing to polar coordinates, we have*

$$\begin{aligned}\lim_{(x,y)\to(0,0)}\frac{xy}{x^2+y^2} &= \lim_{r\to 0}\frac{r\cos\theta\cdot r\sin\theta}{r^2}\\ &= \lim_{r\to 0}\cos\theta\sin\theta.\end{aligned}$$

If we take $\theta = 0$, *this is equivalent to approaching the origin along the* x*-axis, and we have* $\cos\theta\sin\theta = 0$.

If we take $\theta = \pi/4$, *this is equivalent to approaching the origin along the line* $y = x$, *and we have* $\cos\theta\sin\theta = \dfrac{1}{2}$.

Since we get different values approaching the origin along different paths, we see that $\lim\limits_{(x,y)\to(0,0)}\dfrac{xy}{x^2+y^2}$ *does not exist.*

Note that the change of coordinates here wasn't absolutely necessary. We could have directly evaluated the limits approaching along the x*-axis and along the line* $y = x$, *and seen that the limit does not exist. However, sometimes it's easier to see that a limit doesn't exist by attempting a change to polar coordinates, and finding that we end up with a limit that depends on the value of* θ.

In all of the above cases, we considered limits approaching the origin. We can use similar techniques to evaluate limits approaching other points, often by translating coordinate systems.

Example 61. *Consider the limit* $\lim_{(x,y)\to(2,3)} (x-2)(y-3)$*. Here, it's convenient to use polar coordinates translated so they're centered at the point* $(2,3)$*. That is, we let* $x = 2 + r\cos\theta$ *and* $y = 3 + r\sin\theta$*, so that* $r = 0$ *when* $(x,y) = (2,3)$*. Then we can evaluate our limit as*

$$\lim_{(x,y)\to(2,3)} (x-2)(y-3) = \lim_{r\to 0} (2 + r\cos\theta - 2)(3 + r\sin\theta - 3)$$
$$= \lim_{r\to 0} r\cos\theta \cdot r\sin\theta \qquad = \lim_{r\to 0} r^2 \cos\theta\sin\theta$$

Notice that $-r^2 \le r^2\cos\theta\sin\theta \le r^2$*, and* $\lim_{r\to 0} -r^2 = \lim_{r\to 0} r^2 = 0$*. Then, by the squeeze theorem,* $\lim_{r\to 0} r^2\cos\theta\sin\theta = 0$*. Thus, we have*

$$\lim_{(x,y)\to(2,3)} (x-2)(y-3) = 0.$$

SHOULD THIS INCLUDE SOMETHING ABOUT THE REQUIREMENTS FOR A CHANGE OF COORDINATES?

Continuity and Limits in General

So far, we've seen how we can show that limits exist using a delta-epsilon proof, or by changing coordinates. In single variable calculus, we were often able to evaluate limits by direct substitution. For example, we could evaluate

$$\begin{aligned}\lim_{x\to 2} x^2 + x\sin(\pi x) &= 2^2 + 2\sin(\pi\cdot 2)\\ &= \boxed{4}.\end{aligned}$$

We are able to do this because the function $f(x) = x^2 + x\sin(\pi x)$ is continuous. Recall that a function $f : \mathbb{R} \to \mathbb{R}$ is *continuous* at $x = a$ if

$$\lim_{x\to a} f(x) = f(a).$$

For continuous functions, we can evaluate limits by simply plugging in the value.

Once we define continuity for multivariable functions, and determine which functions are continuous, we can use similar methods to evaluate multivariable limits.

Continuity

We define continuity similarly to how we did in single variable calculus.

Definition 41. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is* continuous *at $\vec{x} = \vec{a}$ in $\mathbb{R}^n$ if*

$$\lim_{\vec{x}\to\vec{a}} f(\vec{x}) = f(\vec{a}).$$

Also similar to single variable calculus, virtually all of the common functions that we work with are continuous on their domains. That is, anywhere that they're defined, they are continuous.

Theorem 2. *The following functions are continuous on their domains:*

- *polynomials*
- *root functions*
- *rational functions*

Learning outcomes:
Author(s):

- *trigonometric functions and inverse trigonometric functions*
- *exponential functions and logarithmic functions*

Furthermore, all of the ways that we'd like to combine continuous functions will result in another continuous function.

Theorem 3. *If f is continuous at $\vec{x} = \vec{a}$, and k is a real number, then kf is also continuous at $\vec{a}$.*

If f and g are continuous at $\vec{x} = \vec{a} \in \mathbb{R}^n$, then $f + g$ is also continuous at $\vec{a}$.

If f and g are continuous at $\vec{x} = \vec{a} \in \mathbb{R}^n$, then fg is also continuous at $\vec{a}$.

If g is continuous at $\vec{x} = \vec{a}$, and f is continuous at $g(\vec{a})$, then $f \circ g$ is continuous at $\vec{a}$.

Example 62. *Which of the following functions is continuous at $(0, 0)$? Select all that apply.*

Select All Correct Answers:

(a) $f(x, y) = 3x^3 + 2xy^2 + x + 1$ ✓

(b) $g(x, y) = \sin(x)\cos(x)$ ✓

(c) $h(x, y) = \dfrac{x^2 + y^2 + 1}{x + y + 1}$ ✓

(d) $i(x, y) = \dfrac{x^2 y}{x^2 + y^2}$

(e) $j(x, y) = \tan(xy)$ ✓

(f) $k(x, y) = e^{\sin(x+y)}$ ✓

(g) $l(x, y) = \ln(x^2 + y^2)$

(h) $m(x, y) = \dfrac{1}{\ln(x^2 + y^2 + 2)}$ ✓

Example 63. *Evaluate the following limits, or enter "DNE" if they do not* *exist.*

$$\lim_{(x,y)\to(0,0)} 3x^3 + 2xy^2 + x + 1 = \boxed{1}$$
$$\lim_{(x,y)\to(0,0)} \sin(x)\cos(x) = \boxed{0}$$
$$\lim_{(x,y)\to(0,0)} \frac{x^2+y^2+1}{x+y+1} = \boxed{1}$$
$$\lim_{(x,y)\to(0,0)} \frac{x^2 y}{x^2+y^2} = \boxed{0}$$
$$\lim_{(x,y)\to(0,0)} \tan(xy) = \boxed{0}$$
$$\lim_{(x,y)\to(0,0)} e^{\sin(x+y)} = \boxed{1}$$
$$\lim_{(x,y)\to(0,0)} \ln(x^2+y^2) = \boxed{DNE}$$
$$\lim_{(x,y)\to(0,0)} \frac{1}{\ln(x^2+y^2+2)} = \boxed{1/\ln(2)}$$

Note that even if a function is discontinuous at a point, it's still possible that the limit exists. In this case, you'll need to use a change of coordinates or other method to evaluate the limit.

Limits in General

So far, we've defined limits of scalar-valued functions, $\mathbb{R}^n \to \mathbb{R}$. We've seen how we can evaluate these limits, or show that they do not exist. However, we've yet to deal with more general multivariable functions, $\mathbb{R}^n \to \mathbb{R}^m$.

Fortunately, limits in the more general setting turn out to be an easy extension of the limits that we've already defined. That is, if we have a function $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$, we can write $\vec{f}$ in terms of its coordinate functions,

$$\vec{f}(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), ..., f_m(\vec{x})).$$

Then, we can use the limits of the coordinate functions to define a limit of f.

Definition 42. *Suppose we have a function $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$ with $\vec{f}(\vec{x}) = (f_1(\vec{x}), f_2(\vec{x}), ..., f_m(\vec{x}))$. Then we define*

$$\lim_{\vec{x}\to\vec{a}} \vec{f}(\vec{x}) = \left(\lim_{\vec{x}\to\vec{a}} f_1(\vec{x}), \lim_{\vec{x}\to\vec{a}} f_2(\vec{x}), ..., \lim_{\vec{x}\to\vec{a}} f_m(\vec{x})\right),$$

provided each of these limits exist.

So, we can evaluate the limit of a function $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$ by taking the limit of the component functions. Because of this, the results and methods that we've used for limits of scalar functions carry naturally over to this more general setting.

Example 64. *Evaluate the following limits, or enter "DNE" if they do not exist.*

$$\lim_{(x,y,z)\to(1,2,3)} (x^2 + y, z, xz) = \boxed{(3,3,3)}$$

$$\lim_{(x,y)\to(0,1)} \left(\sin(x+y), \frac{1}{\ln(y)}\right) = \boxed{DNE}$$

$$\lim_{(x,y)\to(0,0)} \left(\frac{x^3 y}{x^2+y^2}, \frac{xy^3}{x^2+y^2}\right) = \boxed{(0,0)}$$

Partial Derivatives

Now that we've defined limits of multivariable function, we're ready to begin to explore how multivariable functions change, using derivatives. Let's recall how we found derivatives in single variable calculus, where they gave us a way to compute the instantaneous rate of change of a function.

We defined the *derivative* of a function f at a to be

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

This definition arose out of geometric considerations. The instantaneous rate of change of a function at a point can be found as the slope of the tangent line to the graph of the function at that point. Our key insight was that, as we get very close to a, the slope of secant lines will approach the slope of the tangent line. This lead us to our limit definition for the derivative. The quotient $\dfrac{f(a+h) - f(a)}{h}$ gives us the slope of the secant line between the points $(a, f(a))$ and $(a+h, f(a+h))$. As we take $h \to 0$, these points get closer together, and the slope of the secant line approaches the slope of the tangent line.

VISUAL

We're beginning to study how multivariable functions change, and we'd like to do this using derivatives. With multivariable functions, it's not clear what this should mean. For single variable functions the question was simple: if we change x, what happens to y? But with multivariable functions, we have multiple inputs, and we could change them in a variety of ways.

For example, consider the function $f(x, y) = x^2 + y$ at $(x, y) = (1, 1)$. We could look at how this function changes if we increase x by a little bit, and leave $y = 1$. We could also look at how this function changes if we increase y by a little bit, and leave $x = 1$. We could also look at how this function changes if we increase x and y by the same amount, or increase y by twice as much as x, or infinitely many other ways.

Because of the breadth of possibilities, it's hard to decide what a multivariable derivative should be. We'll revisit this question later, but for now, we'll see how a function $f : \mathbb{R}^n \to \mathbb{R}$ changes with respect to one input variable at a time.

Learning outcomes:
Author(s):

Definition of Partial Derivatives

In order to study how a function $f : \mathbb{R}^n \to \mathbb{R}$ changes with respect to one input variable, we keep the other variables constant, and change only that variable. This leads us to our definition of partial derivatives. For clarity, we'll begin with the $n = 2$ case, before introducing more general partial derivatives.

Definition 43. *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$.*

We define the partial derivative of f with respect to x *to be*

$$f_x(x, y) = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h},$$

provided this limit exists.

We define the partial derivative of f with respect to y *to be*

$$f_y(x, y) = \lim_{h \to 0} \frac{f(x, y + h) - f(x, y)}{h},$$

provided this limit exists.

Note that these are single variable limits, so we haven't taken advantage of our new multivariable limits yet.

Let's look at how we can compute partial derivatives, similarly to how we computed single variable derivatives using limits.

Example 65. *We'll compute the partial derivatives of the function $f(x, y) = x^2 + xy + y$ at the point $(1, 2)$.*

$$\begin{aligned}
f_x(1, 2) &= \lim_{h \to 0} \frac{f(1 + h, 2) - f(1, 2)}{h} \\
&= \lim_{h \to 0} \frac{((1 + h)^2 + (1 + h) \cdot 2 + 2) - (1^2 + 1 \cdot 2 + 2)}{h} \\
&= \lim_{h \to 0} \frac{((1 + 2h + h^2) + (2 + 2h) + 2) - 5}{h} \\
&= \lim_{h \to 0} \frac{h^2 + 4h}{h} \\
&= \lim_{h \to 0} (h + 4) \\
&= 4
\end{aligned}$$

$$\begin{aligned} f_y(1,2) &= \lim_{h\to 0} \frac{f(1,2+h)-f(1,2)}{h} \\ &= \lim_{h\to 0} \frac{(1^2+1\cdot(2+h)+(2+h))-(1^2+1\cdot 2+2)}{h} \\ &= \lim_{h\to 0} \frac{2h}{h} \\ &= \lim_{h\to 0} 2 \\ &= 2 \end{aligned}$$

More generally, for functions $f:\mathbb{R}^n \to \mathbb{R}$, partial derivatives are defined similarly.

Definition 44. *Consider a function $f:\mathbb{R}^n \to \mathbb{R}$. For $1 \leq i \leq n$, we define the* partial derivative of f with respect to x_i *to be*

$$f_{x_i}(x_1,...,x_n) = \lim_{h\to 0} \frac{f(x_1,...,x_{i-1},x_i+h,x_{i+1},...,x_n)-f(x_1,...,x_n)}{h},$$

provided this limit exists.

ADD PIECEWISE EXAMPLE

Computing Partial Derivatives

When we compute partial derivatives, we're really just taking all but one of the input variables to be constant, and computing a single variable derivative with respect to the remaining variable. Because of this, all of the differentiation rules that we learned in single variable calculus will also apply to partial derivatives. This greatly simplifies computation of partial derivatives.

Example 66. *We'll compute the partial derivatives of $f(x,y) = x^2+xy+y$.*

Thinking of y as a constant and differentiating with respect to x, we obtain the partial derivative with respect to x:

$$f_x(x,y) = 2x+y.$$

Thinking of x as a constant and differentiating with respect to y, we obtain the partial derivative with respect to y:

$$f_y(x,y) = x+1.$$

Example 67. *For each of the following functions, compute the partial derivatives.*

$f(x, y) = x^2y^2 + x^2 + y^2$

$$f_x(x, y) = \boxed{2xy^2 + 2x}$$
$$f_y(x, y) = \boxed{2x^2y + 2y}$$

$g(x, y) = \sin(xy)$

$$g_x(x, y) = \boxed{y\cos(xy)}$$
$$g_y(x, y) = \boxed{x\cos(xy)}$$

$h(x, y, z) = xyz + xe^{xy}$

$$h_x(x, y, z) = \boxed{yz} + e^{xy} + xye^{xy}$$
$$h_y(x, y, z) = \boxed{xz} + x^2e^{xy}$$
$$h_z(x, y, z) = \boxed{xy}$$

Geometric Interpretation of Partial Derivatives

We've defined the partial derivatives of a function as follows.

Definition 45. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. For $1 \leq i \leq n$, we define the* partial derivative of f with respect to x_i *to be*

$$f_{x_i}(x_1, ..., x_n) = \lim_{h \to 0} \frac{f(x_1, ..., x_{i-1}, x_i + h, x_{i+1}, ..., x_n) - f(x_1, ..., x_n)}{h},$$

provided this limit exists.

In other words, if we treat all variables except for x_i as constants, and differentiate with respect to x_i, we get the partial derivative with respect to x_i.

When computing a partial derivative with respect to x_i, we're looking at the instantaneous rate of change of f with respect to x_i, if we keep the rest of the variables constant. Roughly speaking, we're asking: how does increasing x_i a tiny bit affect the value of f?

We can see the partial derivatives reflected in the shape of the graph of f. So that we can visualize the graph of f, we'll focus on a function $f : \mathbb{R}^2 \to \mathbb{R}$, so we're considering the partial derivative of f with respect to x, and with respect to y.

Suppose at the point $(1, 2)$, we have that $f_x(1, 2) > 0$ and $f_y(1, 2) > 0$. Then, around the point $(1, 2)$, if we move a tiny amount in the positive x direction, the value of f will increase. If we move a tiny amount in the positive y direction, the value of f will increase as well.

INTERACTIVE

Similarly, suppose at the point $(1, 2)$, we have that $f_x(1, 2) < 0$ and $f_y(1, 2) < 0$. Then, around the point $(1, 2)$, if we move a tiny amount in the positive x direction, the value of f will decrease. If we move a tiny amount in the positive y direction, the value of f will decrease as well.

INTERACTIVE

Now, let's consider the case where $f_x(1, 2) > 0$ and $f_y(1, 2) < 0$. Then, around the point $(1, 2)$, if we move a tiny amount in the positive x direction, the value of f will increase. But, if we move a tiny amount in the positive y direction, the value of f will decrease.

Learning outcomes:
Author(s):

INTERACTIVE

Next, let's suppose that $f_x(1,2) > 0$ and $f_y(1,2) = 0$. As expected, f increases as we move a tiny amount in the positive x direction. On the other hand, the graph of f has flattened out as we move in the y direction. However, this doesn't mean that it's constant! It's just the instantaneous rate of change that's 0 at that one point.

INTERACTIVE

Now, let's look at a case where $f_x(1,2) = 0$ and $f_y(1,2) = 0$. As before, this does not mean that f is constant. This just means that the rates of change are both instantaneously 0. Points with this property will be important later in the course, when we study optimization.

INTERACTIVE

Geometry of Differentiability

In single variable calculus, derivatives were closely related to the slope of the tangent line to a graph at a point. We used this idea of the slope of the tangent line to define derivatives as a limit of slopes of secant lines,

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

VISUAL

In the other direction, we were able to use differentiation rules to more easily find the equation for the tangent line to a graph at a point.

Example 68. *We'll find the equation for the tangent line to the graph of $f(x) = x^3 + 2x + 1$ at $x = 2$.*

We can find the slope of the tangent line by computing $f'(2)$. Using differentiation rules, we have

$$f'(x) = \boxed{3x^2 + 2}.$$

Plugging in $x = 2$, we have $f'(2) = \boxed{14}$.

Since the tangent line will have to pass through the point $(2, f(2))$, we compute

$$f(2) = \boxed{13}.$$

So, the tangent line will pass through the point $(2, 13)$, and will have slope 14. Writing the equation of the line in point-slope form, we have

$$y - 13 = \boxed{14(x-2)}.$$

When a single variable function is differentiable, we can use the above method to find an equation for the tangent line. In addition, the tangent line provides us with a good linear approximation for the function.

We would like to do something analogous for multi-variable functions, but this raises a few questions. What would be the equivalent of the tangent line? What does it mean for a function to be differentiable?

As we begin to explore these questions, we'll focus on functions $\mathbb{R}^2 \to \mathbb{R}$, so that we can visualize their graphs.

Learning outcomes:
Author(s):

Geometric Interpretation of Differentiability

In single variable calculus, we could get a good sense of whether a function was differentiable by looking at its graph.

Problem 9 *For each of the graphs, determine whether the given function is differentiable at $x = a$.*

GRAPH PARABOLA

Multiple Choice:

(a) *differentiable* ✓

(b) *not differentiable*

GRAPH JUMP DISCONTINUITY

Multiple Choice:

(a) *differentiable*

(b) *not differentiable* ✓

GRAPH ABSOLUTE VALUE

Multiple Choice:

(a) *differentiable*

(b) *not differentiable* ✓

If there is a discontinuity or some sort of corner or cusp in the graph at a point, then the function will not be differentiable at that point. Roughly speaking, if we "zoom in" on the graph of a function near a point, and the graph looks very close to a line, then the function will be differentiable at that point.

PARABOLA ZOOM

We'll extend this idea to make our first, informal definition of differentiability for a function from $\mathbb{R}^2$ to $\mathbb{R}$.

Definition 46. *(Informal Definition) Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$. Suppose for some point (a, b) in $\mathbb{R}^2$, if we zoom in on the graph of f near the point (a, b), the graph of f looks like a plane. Then f is differentiable at (a, b).*

Although this definition provides us with nice geometric intuition for determining if a function is differentiable, it's not at all precise or rigorous. Eventually, we'll need a more formal definition of differentiability, so we'll return to this concept later.

For now, let's use this informal definition to investigate differentiability for a couple of functions.

Example 69. *Consider the function $f(x, y) = xy + 2x + y$, graphed below.*

ZOOMABLE GRAPH

Is f differentiable at $(0, 0)$?

Multiple Choice:

(a) *Yes.* ✓

(b) *No.*

Example 70. *Consider the function*

$$f(x, y) = \begin{cases} 1 - |y| & \text{if } |y| \leq |x| \\ 1 - |x| & \text{if } |y| > |x| \end{cases},$$

graphed below.

ZOOMABLE GRAPH

Is f differentiable at $(0, 0)$?

Multiple Choice:

(a) *Yes.*

(b) *No.* ✓

The tangent plane

If a function $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable, when we zoom in, the graph will look like a plane. Because of this, there will be a plane that's a good linear approximation for the function near that point. We can use the partial derivatives with respect to x and y to find an equation for this plane, which we call the tangent plane.

Definition 47. *If $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at (a, b), then the* tangent plane *to the graph of f at (a, b) is defined by the equation*

$$z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

Let's revisit our previous examples, and see what happens with the tangent plane.

Example 71. *Consider the function $f(x,y) = xy + 2x + y$, which we found is differentiable at the point $(0,0)$. Let's find an equation for the tangent plane to the graph of f at $(0,0)$.*

We begin by finding the partial derivatives with respect to x and y.

$$\begin{aligned} f_x(x,y) &= \boxed{y+2} \\ f_y(x,y) &= \boxed{x+1} \end{aligned}$$

At $(0,0)$, we have

$$\begin{aligned} f_x(0,0) &= \boxed{2}, \\ f_y(0,0) &= \boxed{1}. \end{aligned}$$

Finding the value of f at $(0,0)$, we have

$$f(0,0) = \boxed{0}.$$

Putting all of this together, we obtain an equation for the tangent plane to the graph of f at $(0,0)$.

$$\begin{aligned} z &= f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) \\ &= \boxed{2x+y} \end{aligned}$$

Example 72. *Consider the function*

$$f(x,y) = \begin{cases} 1-|y| & \text{if } |y| \le |x| \\ 1-|x| & \text{if } |y| > |x| \end{cases},$$

which we found is not differentiable $(0,0)$. Even though this function is not differentiable, let's see what happens when we try to find an equation for the tangent plane.

To compute the partial derivatives with respect to x and y, we'll need to use the limit definition.

$$\begin{aligned}
f_x(0,0) &= \lim_{h\to 0} \frac{f(0+h,0)-f(0,0)}{h} \\
&= \lim_{h\to 0} \frac{(1-|0|)-(1-|0|)}{h} \\
&= \lim_{h\to 0} \frac{0}{h} \\
&= 0. f_y(0,0) \qquad\qquad = \lim_{h\to 0} \frac{f(0,0+h)-f(0,0)}{h} \\
&= \lim_{h\to 0} \frac{(1-|0|)-(1-|0|)}{h} \\
&= \lim_{h\to 0} \frac{0}{h} \\
&= 0.
\end{aligned}$$

We can also find $f(0,0)$*,*

$$f(0,0) = 1.$$

So, we have all of the necessary pieces to find an equation for the (nonexistent) "tangent plane":

$$\begin{aligned}
z &= f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) \\
&= 1.
\end{aligned}$$

However, we decided that the function wasn't differentiable at $(0,0)$*, so the graph does not have a tangent plane at the point.*

This example brings up a couple of important points.

- It's possible for the partial derivatives of a function to all exist, and yet the function is not differentiable.
- It's possible that we can find the equation $z = f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b)$, and yet f has no tangent plane at the point (a,b).

For these reasons, differentiability is a much more subtle concept in multivariable calculus than it was in single variable calculus, and our next task will be to find a formal definition for differentiability.

Differentiability of Functions of Two Variables

So far, we have an informal definition of differentiability for functions $f\mathbb{R}^2 \to \mathbb{R}$: if the graph of f "looks like" a plane near a point, then f is differentiable at that point.

Definition 48. *(Informal Definition) Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$. Suppose for some point (a, b) in $\mathbb{R}^2$, if we zoom in on the graph of f near the point (a, b), the graph of f looks like a plane. Then f is differentiable at (a, b).*

In the case where a function is differentiable at a point, we defined the tangent plane at that point.

Definition 49. *If $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable at (a, b), then the* tangent plane *to the graph of f at (a, b) is defined by the equation*

$$z = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

We would like a formal, precise definition of differentiability. The key idea behind this definition is that a function should be differentiable if the plane above is a "good" linear approximation. To see what this means, let's revisit the single variable case.

In single variable calculus, a function $f : \mathbb{R} \to \mathbb{R}$ is differentiable at $x = a$ if the following limit exists:

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

This limit exists if and only if

$$\lim_{x \to a} \left(\frac{f(x) - f(a)}{x - a} - f'(a) \right) = 0.$$

In turn, this is true if and only if

$$\lim_{x \to a} \frac{f(x) - f(a) - f'(a)(x - a)}{x - a} = 0.$$

If we let $L(x) = f(a) + f'(a)(x - a)$, this is equivalent to

$$\lim_{x \to a} \frac{f(x) - L(x)}{x - a} = 0.$$

Learning outcomes:
Author(s):

Recall that $L(x)$, as defined above, is the linear approximation to f at $x = a$. This is also a function whose graph is the tangent line to f at $x = a$. So, roughly speaking, we have shown that a single variable function is differentiable if and only the difference between $f(x)$ and its linear approximation goes to 0 quickly as x approaches a.

This idea will inform our definition for differentiability of multivariable functions: a function will be differentiable at a point if it has a good linear approximation, which will mean that the difference between the function and the linear approximation gets small quickly as we approach the point.

Formal definition of differentiability

We are now in position to give our formal definition of differentiability for a function $f : \mathbb{R}^2 \to \mathbb{R}$. We'll make our definition so that a function is differentiable at a point if the difference between the function and the linear approximation

$$h(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

gets small "quickly". Here, "quickly" is relative to how $\vec{x}$ is approaching $\vec{a}$, so relative to the distance $\|\vec{x} - \vec{a}\|$ between these points.

Notice that the function $h(x, y)$ matches the equation for the tangent plane, when the function f is differentiable.

Definition 50. *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, and suppose that the partial derivatives f_x and f_y are defined at the point $(x, y) = (a, b)$. Define the linear function*

$$h(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

We say that f is differentiable *at $(x, y) = (a, b)$ if*

$$\lim_{(x,y)\to(a,b)} \frac{f(x, y) - h(x, y)}{\|(x, y) - (a, b)\|} = 0.$$

If either of the partial derivatives $f_x(a, b)$ and $f_y(a, b)$ do not exist, or the above limit does not exist or is not 0, then f is not differentiable at (a, b).

We had previously used our informal definition of differentiability to determine that the function $f(x, y) = xy + 2x + y$ is differentiable at $(0, 0)$. Let's verify this using our new, formal definition of differentiability.

Example 73. *We'll show that the function $f(x, y) = xy+2x+y$ is differentiable at $(0, 0)$. In order to do this, we first need to find the function $h(x, y)$. This repeats earlier work, where we found the tangent plane to $f(x, y) = xy + 2x + y$ at $(0, 0)$.*

We begin by finding the partial derivatives with respect to x and y.

$$f_x(x,y) = \boxed{y+2}$$
$$f_y(x,y) = \boxed{x+1}$$

At $(0,0)$, we have

$$f_x(0,0) = \boxed{2},$$
$$f_y(0,0) = \boxed{1}.$$

Finding the value of f at $(0,0)$, we have

$$f(0,0) = \boxed{0}.$$

Putting all of this together, we obtain an equation for the function $h(x,y)$.

$$\begin{aligned} h(x,y) &= f(a,b) + f_x(a,b)(x-a) + f_y(a,b)(y-b) \\ &= \boxed{2x+y} \end{aligned}$$

Now, we show that f is differentiable at $(a,b) = (0,0)$, by evaluating the limit

$$\begin{aligned} \lim_{(x,y)\to(a,b)} \frac{f(x,y)-h(x,y)}{\|(x,y)-(a,b)\|} &= \lim_{(x,y)\to(0,0)} \frac{(xy+2x+y)-(2x+y)}{\sqrt{(x-0)^2+(y-0)^2}} \\ &= \lim_{(x,y)\to(0,0)} \frac{xy}{\sqrt{x^2+y^2}}. \end{aligned}$$

Switching to polar coordinates, we have

$$\begin{aligned} \lim_{(x,y)\to(0,0)} \frac{xy}{\sqrt{x^2+y^2}} &= \lim_{r\to 0} \frac{r\cos\theta \cdot r\sin\theta}{\sqrt{r^2}} \\ &= \lim_{r\to 0} \frac{r^2\cos\theta\sin\theta}{|r|}. \end{aligned}$$

Since $-1 \le \cos\theta\sin\theta \le 1$, we have

$$-|r| \le \frac{r^2\cos\theta\sin\theta}{|r|} \le |r|.$$

Since $\lim_{r} \to 0 - |r| = \lim_{r} \to 0|r| = 0$, by the squeeze theorem, we have

$$\lim_{r\to} \frac{r^2\cos\theta\sin\theta}{|r|} = 0.$$

Thus, we have shown that $\lim_{(x,y)\to(0,0)} \dfrac{f(x,y)-h(x,y)}{\|(x,y)-(0,0)\|} = 0$, showing that f is differentiable at $(0,0)$.

The Gradient

We've given a formal definition for differentiability of a function $f : \mathbb{R}^2 \to \mathbb{R}$,

Definition 51. *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, and suppose that the partial derivatives f_x and f_y are defined at the point $(x, y) = (a, b)$. Define the linear function*

$$h(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

We say that f is differentiable *at $(x, y) = (a, b)$ if*

$$\lim_{(x,y)\to(a,b)} \frac{f(x, y) - h(x, y)}{\|(x, y) - (a, b)\|} = 0.$$

If either of the partial derivatives $f_x(a, b)$ and $f_y(a, b)$ do not exist, or the above limit does not exist or is not 0, then f is not differentiable at (a, b).

The idea behind this definition is that $h(x, y)$ will be a "good" linear approximation to $f(x, y)$ near (a, b) if f is differentiable at (a, b).

We would now like to define differentiability for scalar-valued functions of more than two variables, so functions $\mathbb{R}^n \to \mathbb{R}$. This definition will closely resemble our definition above, which handles the case $n = 2$. For example, in the case $n = 3$, we will use the linear function

$$h(x, y, z) = f(a, b, c) + f_x(a, b, c)(x - a) + f_y(a, b, c)(y - b) + f_z(a, b, c)(z - c).$$

For larger n, we'll define a similar function h, but this notation will quickly become unwieldy! In order to simplify notation, we'll now introduce a new object to organize our partial derivatives: the gradient of a scalar-valued function.

The gradient

In order to organize our information about partial derivatives, and streamline our definition of differentiability for functions $\mathbb{R}^n \to \mathbb{R}$, we now define the gradient of a scalar-valued function.

Definition 52. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. The* gradient *of f is the function $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$, defined by*

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n} \right).$$

Learning outcomes:
Author(s):

The gradient vector will be a useful computation tool in general, not only for defining differentiability.

Example 74. *For $f(x, y, z) = x^2 + ye^z$, we can compute the partial derivatives*

$$\begin{aligned} f_x(x, y, z) &= 2x, \\ f_y(x, y, z) &= e^z, \\ f_z(x, y, z) &= ye^z. \end{aligned}$$

Then the gradient of f is

$$\nabla f = (2x, e^z ye^z).$$

Problem 10 *Find the gradient of each function.*

$f(x, y, z) = \sin(xyz)$

$$\nabla f(x, y, z) = \boxed{(yz\cos(xyz), xz\cos(xyz), xy\cos(xyz))}$$

$g(x, y) = x^2e^y + y$

$$\nabla g(x, y) = \boxed{(2xe^y, x^2e^y + 1)}$$

$h(x_1, x_2, x_3, x_4) = x_1^2x_2 + x_1x_3 + x_2x_4$

$$\nabla h(x_1, x_2, x_3, x_4) = \boxed{(2x_1x_2 + x_3, x_1^2 + x_4, x_1, x_2)}$$

Differentiability

Now that we've defined the gradient, let's revisit our definition of differentiability for a function from $\mathbb{R}^2$ to $\mathbb{R}$. We used the function

$$h(x, y) = f(a, b) + f_x(a, b)(x - a) + f_y(a, b)(y - b).$$

Looking at the terms $f_x(a, b)(x - a) + f_y(a, b)(y - b)$, we can rewrite this as a dot product of two vectors:

$$f_x(a, b)(x - a) + f_y(a, b)(y - b) = (f_x(a, b), f_y(a, b)) \cdot (x - a, y - b).$$

The first vector is the gradient of f evaluated at (a, b), so we can rewrite this as

$$(f_x(a, b), f_y(a, b)) \cdot (x - a, y - b) = \nabla f(a, b) \cdot (x - a, y - b).$$

If we take $\vec{x} = (x, y)$ and $\vec{a} = (a, b)$, we can write this as

$$\nabla f(\vec{a}) \cdot (\vec{x} - \vec{a}).$$

With these notational changes in mind, we now define differentiability for a function $f : \mathbb{R}^n \to \mathbb{R}$.

Definition 53. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. For a point $\vec{a} \in \mathbb{R}^n$, define*

$$h(\vec{x}) = f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a}).$$

We say that f is differentiable *at $\vec{a}$ if*

$$\lim_{\vec{x} \to \vec{a}} \frac{f(\vec{x}) - h(\vec{x})}{\|\vec{x} - \vec{a}\|} = 0.$$

If f is differentiable, we say that $h(\vec{x})$ is the tangent hyperplane *to f at $\vec{a}$.*

If any of the partial derivatives of f do not exist, or the above limit does not exist or is not 0, then f is not differentiable at $\vec{a}$.

Example 75. *We'll use this definition of differentiability to prove that the function $f(x, y, z) = xy + z$ is differentiable at $(1, 1, 1)$.*

First, we find the gradient of f.

$$\nabla f(x, y, z) = \boxed{(y, x, 1)}$$

At the point $(1, 1, 1)$, we have

$$\nabla f(1, 1, 1) = \boxed{(1, 1, 1)}.$$

From this, we find the formula for $h(x, y, z)$.

$$\begin{aligned} h(x, y, z) &= f(1, 1, 1) + \nabla f(1, 1, 1) \cdot ((x, y, z) - (1, 1, 1)) \\ &= 2 + (1, 1, 1) \cdot (x - 1, y - 1, z - 1) \\ &= \boxed{x + y + z - 1} \end{aligned}$$

Next, we evaluate the limit

$$\begin{aligned} \lim_{(x,y,z) \to (1,1,1)} \frac{f(x, y, z) - h(x, y, z)}{\|(x, y, z) - (1, 1, 1)\|} &= \lim_{(x,y,z) \to (1,1,1)} \frac{(xy + z) - (x + y + z - 1)}{\sqrt{(x-1)^2 + (y-1)^2 + (z-1)^2}} \\ &= \lim_{(x,y,z) \to (1,1,1)} \frac{xy - x - y + 1}{\sqrt{(x-1)^2 + (y-1)^2 + (z-1)^2}}. \end{aligned}$$

To evaluate this limit, we switch to translated spherical coordinates

$$\begin{aligned} x &= 1 + \rho \cos\theta \sin\phi, \\ y &= 1 + \rho \sin\theta \sin\phi, \\ z &= 1 + \rho \cos\phi. \end{aligned}$$

Making this change, we obtain

$$\begin{aligned} \lim_{(x,y,z) \to (1,1,1)} \frac{xy - x - y + 1}{\sqrt{(x-1)^2 + (y-1)^2 + (z-1)^2}} &= \lim_{\rho \to 0} \frac{(1 + \rho\cos\theta\sin\phi)(1 + \rho\sin\theta\sin\phi) - (1 + \rho\cos\theta\sin\phi) - (}{|\rho|} \\ &= \lim_{\rho \to 0} \frac{\rho^2 \cos\theta \sin\theta \sin^2\phi}{|\rho|}. \end{aligned}$$

Since $-|\rho| \leq \dfrac{\rho^2 \cos\theta \sin\theta \sin^2\phi}{|\rho|} \leq |\rho|$, *we use the squeeze theorem to obtain*

$$\lim_{\rho \to 0} \frac{\rho^2 \cos\theta \sin\theta \sin^2\phi}{|\rho|} = 0.$$

Thus, we have shown that $f(x, y, z) = xy + z$ *is differentiable at* $(1, 1, 1)$.

Differentiability in General

We've defined differentiability for scalar-valued functions, $\mathbb{R}^n \to \mathbb{R}$.

Definition 54. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. For a point $\vec{a} \in \mathbb{R}^n$, define*

$$h(\vec{x}) = f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a}).$$

We say that f is differentiable *at $\vec{a}$ if*

$$\lim_{\vec{x} \to \vec{a}} \frac{f(\vec{x}) - h(\vec{x})}{\|\vec{x} - \vec{a}\|} = 0.$$

If f is differentiable, we say that $h(\vec{x})$ is the tangent hyperplane *to f at $\vec{a}$.*

If any of the partial derivatives of f do not exist, or the above limit does not exist or is not 0, then f is not differentiable at $\vec{a}$.

We'd now like to define differentiability for vector-valued functions, $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$.

In order to define differentiability for scalar-valued functions $f : \mathbb{R}^n \to \mathbb{R}$, we organized our partial derivatives into a vector, the gradient of f.

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, ..., \frac{\partial f}{\partial x_n} \right)$$

We would like to do something similar for a vector-valued function $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$, and organize all of the partial derivatives into a single object. However, for a function $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$, we not only have partial derivatives with respect to all of the different variables, we have partial derivatives of all of the component functions with respect to all of the different variables! This leads us to the derivative matrix.

The derivative matrix

Definition 55. *Let $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$, and suppose all of the partial derivatives of $\vec{f}$ exist. Write $\vec{f}$ in terms of its component functions,*

$$\vec{f}(x_1, ..., x_n) = (f_1(x_1, .., x_n), ..., f_m(x_1, ..., x_n)).$$

Learning outcomes:
Author(s):

We define the derivative matrix *of $\vec{f}$ to be the $m \times n$ matrix with $\dfrac{\partial f_i}{\partial x_j}$ as the ijth entry. That is,*

$$D\vec{f}(x_1, ..., x_n) = \begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{pmatrix}.$$

It can be hard to remember whether the variable or the component changes across the rows or columns. Here are a couple of ways to remember which way it goes:

- The derivative matrix represents a linear transformation $\mathbb{R}^n \to \mathbb{R}^m$.
- The rows are gradients of the component functions.

Example 76. *We'll find the derivative matrix of the function $\vec{f}(x, y, z) = (x^2 + yz, xyz)$. Since $\vec{f} : \mathbb{R}^3 \to \mathbb{R}^2$, the derivative matrix will be a 2×3 matrix.*

$$\begin{aligned} D\vec{f}(x, y, z) &= \begin{pmatrix} \dfrac{\partial}{\partial x}(x^2 + yz) & \dfrac{\partial}{\partial y}(x^2 + yz) & \dfrac{\partial}{\partial z}(x^2 + yz) \\ \dfrac{\partial}{\partial x}(xyz) & \dfrac{\partial}{\partial y}(xyz) & \dfrac{\partial}{\partial z}(xyz) \end{pmatrix} \\ &= \begin{pmatrix} 2x & z & y \\ yz & xz & xy \end{pmatrix} \end{aligned}$$

Problem 11 *Find the derivative matrix of each of the following functions.*

$\vec{f}(x, y) = (x^2 + y^2, \sin(xy), e^{x+y})$

$$D\vec{f}(x, y) = \begin{pmatrix} \boxed{2x} & \boxed{2y} \\ \boxed{y\cos(xy)} & \boxed{x\cos(xy)} \\ \boxed{e^{x+y}} & \boxed{e^{x+y}} \end{pmatrix}$$

$\vec{g}(x, y, z) = (x^2 z + yz^2, x + y + z, x + y^2 + z)$

$$D\vec{g}(x, y, z) = \begin{pmatrix} \boxed{2xz} & \boxed{z^2} & \boxed{x^2 + 2yz} \\ \boxed{1} & \boxed{1} & \boxed{1} \\ \boxed{1} & \boxed{2y} & \boxed{1} \end{pmatrix}$$

Differentiability

We can now generalize our definition of differentiability for scalar-valued functions, by replacing the gradient with the derivative matrix.

Definition 56. *Consider a function $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$. For a point $\vec{a} \in \mathbb{R}^n$, define*

$$\vec{h}(\vec{x}) = \vec{f}(\vec{a}) + D\vec{f}(\vec{a})(\vec{x} - \vec{a}).$$

We say that $\vec{f}$ is differentiable *at $\vec{a}$ if*

$$\lim_{\vec{x} \to \vec{a}} \frac{\vec{f}(\vec{x}) - \vec{h}(\vec{x})}{\|\vec{x} - \vec{a}\|} = \vec{0}.$$

If any of the partial derivatives of $\vec{f}$ do not exist, or the above limit does not exist or is not 0, then $\vec{f}$ is not differentiable at $\vec{a}$.

Note one of the quirks of multivariable differentiation: if the derivative matrix exist, it's still possible for the function to not be differentiable.

For vector-valued functions, we can also reduce differentiability to differentiability of its component functions.

Theorem 4. *A function $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable if and only if its component functions are all differentiable.*

This theorem can quickly be proved from the definitions of differentiability in these two cases.

A criterion for differentiability

Checking differentiability using the limit definitions that we've found can be a huge pain! It would be much nicer if we could tell if a function is differentiable just by looking at the partial derivatives. Fortunately, this is possible in some cases.

Theorem 5. *Consider a function $\vec{f} : \mathbb{R}^n \to \mathbb{R}^m$. Suppose that all of the partial derivatives, $\partial f_i / \partial x_j$, exist and are continuous in a neighborhood of a point $\vec{a}$. Then $\vec{f}$ is differentiable at $\vec{a}$.*

An analogous result holds for scalar-valued functions. This theorem requires a very important note: its converse is *false*. That is, if one or more of the partial derivatives of a function is discontinuous, it's still possible that the function is differentiable. In this case, you would probably need to resort to the limit definition to determine differentiability.

Example 77. *We'll use the above theorem to show that* $\vec{g}(x, y, z) = (x^2z + yz^2, x + y + z, x + y^2 + z)$ *is differentiable at all points in* $\mathbb{R}^3$.

First, we find all of the partial derivatives of g.

$$\frac{\partial g_1}{\partial x} = \boxed{2xz}$$
$$\frac{\partial g_1}{\partial y} = \boxed{z^2}$$
$$\frac{\partial g_1}{\partial z} = \boxed{x^2 + 2yz}$$
$$\frac{\partial g_2}{\partial x} = \boxed{1}$$
$$\frac{\partial g_2}{\partial y} = \boxed{1}$$
$$\frac{\partial g_2}{\partial z} = \boxed{1}$$
$$\frac{\partial g_3}{\partial x} = \boxed{1}$$
$$\frac{\partial g_3}{\partial y} = \boxed{2y}$$
$$\frac{\partial g_2}{\partial z} = \boxed{1}$$

All of these functions are polynomials, hence continuous at all points $\mathbb{R}^3$. *Since the partial derivatives of* $\vec{g}$ *all exist and are continuous on* $\mathbb{R}^3$, *by the theorem above,* $\vec{g}$ *is differentiable at all points in* $\mathbb{R}^3$.

Higher Order Partial Derivatives

In this activity, we introduce higher order partial derivatives, and discuss their geometric meaning.

Higher Order Partials

Back in single variable Calculus, we were able to use the second derivative to get information about a function. For instance, the second derivative gave us valuable information about the shape of the graph. More specifically, we could use the second derivative to determine the concavity.

- If $f''(x) > 0$ on an interval, then the graph of f is concave up on that interval.
- If $f''(x) < 0$ on an interval, then the graph of f is concave down on that interval.
- If f'' changes signs at a point, then the graph of f has an inflection point at that point.

Furthermore, we were able to use second derivative in conjunction with roots of the first derivative to find local maxima and minima. We could also think about the second derivative as the rate-of-change of a rate-of-change, or describing some sort of acceleration or deceleration.

Since the second derivative contains so much useful information, we would like to come up with a way to define second derivatives for multivariable functions! At first, this might seem simple: just take the derivative of the derivative. But when we took the total derivative of a multivariable function $f : \mathbb{R}^m \to \mathbb{R}^n$, we got a matrix of partial derivatives, and it's not at all clear how we would differentiate a matrix.

For now, we'll settle for defining second order partial derivatives, and we'll have to wait until later in the course to define more general second order derivatives. Fortunately, second order partial derivatives work exactly like you'd expect: you simply take the partial derivative of a partial derivative.

Example 78. *Consider the function $f(x, y) = 2x^2 + 4xy - 7y^2$. We'll start by*

Learning outcomes:
Author(s):

computing the first order partial derivatives of f, with respect to x and y.

$$f_x(x,y) = \boxed{6x + 4y}$$
$$f_y(x,y) = \boxed{4x - 14y}$$

We can then compute the second order partial derivatives f_{xx} and f_{yy} by differentiating with respect to x again, and with respect to y again.

$$f_{xx}(x,y) = \boxed{6}$$
$$f_{yy}(x,y) = \boxed{-14}$$

However, this isn't the only way that we could take second order partial derivatives! We could differentiate with respect to x first, and with respect to y second, to get f_{xy}. We could also differentiate with respect to y first, and with respect to x second, to get f_{yx}.

$$f_{xy}(x,y) = \boxed{4}$$
$$f_{yx}(x,y) = \boxed{4}$$

Notice that we got the same result for f_{xy} and f_{yx}, so it didn't end up mattering what order we took these derivatives in. In turns out that this is not a coincidence, and it's a consequence of Clairaut's Theorem, which we'll talk about in the next section.

There are a few different ways that we can denote second order partials. We can denote the second order partial of f that we get by differentiating with respect to x twice as any of the following.

$$f_{xx}(x,y) \qquad \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right) \qquad \frac{\partial^2 f}{\partial x^2}$$

For the second order partial of f that we get by differentiating with respect to x first, then differentiating with respect to y, we denote this as below.

$$f_{xy}(x,y) \qquad \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right) \qquad \frac{\partial^2 f}{\partial y \partial x}$$

To remember what order to take these derivatives for all of these notations, start with the variable closest to f, and work your way out.

We can similarly define and compute third order partials, fourth order partials, and so on.

Example 79. *COMPUTE SOME PARTIAL DERIVATIVES*

Clairaut's Theorem

In previous examples, we've seen that it doesn't matter what order you use to take higher order partial derivatives, you seem to wind up with the same answer no matter what. This isn't an amazing coincidence where we randomly chose functions that happened to have this property; this turns out to be true for many functions. Clairaut's Theorem gives us this result.

Theorem 6. *(Clairaut's Theorem) Suppose we have a function $f : D \subset \mathbb{R}^n \to \mathbb{R}$. Suppose further that* all *of the second order mixed partial derivatives of $f(x_1, ..., x_n)$ exist and are continuous on an open disc around $\mathbf{a} \in D$. Then, for any x_i and x_j,*

$$f_{x_i x_j}(\mathbf{a}) = f_{x_j x_i}(\mathbf{a}).$$

Proof prove things ■

We can prove a similar result for even higher order partial derivatives. Before we do that, we'll introduce a new definition to make it easier to describe how "nice" functions are.

Definition 57. *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is of class $\mathcal{C}^k$ if all of the partial derivatives of f up to and including the kth order exist and are continuous.*

If f is of class $\mathcal{C}^k$ for all k, then we say f is of class $\mathcal{C}^\infty$.

If f is continuous, we say f is of class $\mathcal{C}^0$.

We'll often write this as $f \in \mathcal{C}^3$ and say "f is $\mathcal{C}$ three," for example, if f is of class $\mathcal{C}^3$.

Problem 12 *If $f(x, y) = x^2 + y^2 + 2xy$, which of the following are true? Select all that apply.*

Select All Correct Answers:

(a) f is $\mathcal{C}^0$. ✓

(b) f is $\mathcal{C}^1$. ✓

(c) f is $\mathcal{C}^2$. ✓

(d) f is $\mathcal{C}^3$. ✓

(e) f is $\mathcal{C}^\infty$. ✓

We can generalize Clairaut's Theorem to kth order derivatives for $\mathcal{C}^k$ functions.

Theorem 7. *If f is $\mathcal{C}^k$, then the kth order mixed partials can be computed in any order.*

For example, if f is $\mathcal{C}^{12}$, we have

$$\begin{aligned} f_{xyxyzyzzxyxz} &= f_{xxxxyyyyzzzz} \\ &= f_{xyzxyzxyzxyz}. \end{aligned}$$

Problem 13 *If f is $\mathcal{C}^5$, which of the following are guaranteed to be equal to f_{zzyzx}? Select all that apply,*

Select All Correct Answers:

(a) f_{xxzxy}

(b) f_{xyzzz} ✓

(c) f_{xzzzy} ✓

(d) f_{zzxzy} ✓

(e) f_{yyzyx}

Geometric Significance

Remembering back to single variable calculus, we could use the first and second derivatives of a function to figure out the shape of the graph.

More precisely, we used the first derivative to determine where a function was increasing and where it was decreasing. If $f'(x) > 0$ on some interval, then $f(x)$ is increasing on that interval. If $f'(x) < 0$ on some interval, then $f(x)$ is decreasing on that interval.

PICTURE

We used the second derivative to determine the concavity of a function. If $f''(x) > 0$ on an interval, then $f(x)$ is concave up on that interval. If $f''(x) < 0$ on an interval, then $f(x)$ is concave down on that interval.

PICTURE

Partial derivatives can give us similar information about the graph of a multi-variable function, although the situation is a bit more nuanced. Let's consider a function $f : \mathbb{R}^2 \to \mathbb{R}$, so we can visualize its graph. We've already seen that the partial derivative with respect to x tells us how f changes as x changes.

More specifically, suppose $f_x(x, y_0) > 0$ for x in some open interval containing x_0. Then $f(x, y)$ is increasing as we move *in the positive x-direction* from the point (x_0, y_0).

PICTURE

Similarly, suppose $f_x(x, y_0) < 0$ for x in some open interval constaining x_0. Then $f(x, y)$ is increasing as we move in the positive x-direction from the point (x_0, y_0).

PICTURE

Suppose $f_x(x, y_0) = 0$ for x in some open interval constaining x_0. Then $f(x, y)$ is constant as we move in the positive x-direction from the point (x_0, y_0).

PICTURE

The partial derivative with respect to y, f_y, can tell us where f is increasing, decreasing, or constant as we move in the positive y direction.

PICTURE

Now, let's look at what the second-order partial derivatives tell us about the graph of the function $f(x, y)$. It shouldn't be too surprising that the sign of f_{xx} tells us about the concavity of f as we move in the positive x direction.

PICTURE

Similarly, the sign of f_{yy} tells us about the concavity of f as we move in the positive y direction.

PICTURE

But what do the mixed partials, f_{xy} and f_{yx}, tell us about the graph of f? Let's consider f_{xy}, which is the partial derivative of f_x with respect to y. The partial derivative f_x tells us the rate of change of f as we move in the positive x direction. Then, the partial derivative of f_x with respect to y tells us how f_x changes as we move in the positive y direction. That is, we look at how the rate of change in the x direction changes as we move in the y direction. This is a lot to unravel!

PICTURE/VIDEO/INTERACTIVE

We can similarly think of f_{yx} as the change in f_y as we move in the positive x direction.

PICTURE/VIDEO/INTERACTIVE

Conclusion

In this activity, we considered higher order partial derivatives, and found that the order of differentiation doesn't matter for "nice" functions using Clairaut's theorem. We also considered the geometric information carried by second order partial derivatives.

Differentiation Properties

In this activity, we explore some of the properties of differentiation. This include how derivatives interact with addition and scalar multiplication, as well as product and quotient rules for scalar valued functions.

Linearity of the derivative

As you may recall from single variable calculus, "the derivative of the sum is the sum of the derivatives." That is, if we have differentiable functions $f(x)$ and $g(x)$, we compute the derivative of the sum $f(x)+g(x)$ by taking the sum of the derivatives $f'(x)$ and $g'(x)$. For example,

$$\begin{aligned}\frac{d}{dx}\left(\sin(x)+x^2\right) &= \frac{d}{dx}\left(\sin(x)\right)+\frac{d}{dx}\left(x^2\right)\\ &= \cos(x)+2x.\end{aligned}$$

An analogous result holds in multivariable calculus, for the derivative matrix.

Proposition 25. *Suppose $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{g} : Y \subset \mathbb{R}^n \to \mathbb{R}^m$ are differentiable functions on subsets X and Y of $\mathbb{R}^n$, respectively. Then $\mathbf{f}+\mathbf{g}$ is differentiable on $X \cap Y$, and $D(\mathbf{f}+\mathbf{g}) = D\mathbf{f}+D\mathbf{g}$ on $X \cap Y$.*

Proof We will begin by showing that $D(\mathbf{f}+\mathbf{g}) = D\mathbf{f}+D\mathbf{g}$ on $X \cap Y$. Since $X \cap Y$ is contained in the domains of both $\mathbf{f}$ and $\mathbf{g}$, $D\mathbf{f}$ and $D\mathbf{g}$ both exist on $X \cap Y$.

Write $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_m(\mathbf{x}))$ and $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), ..., g_m(\mathbf{x}))$ in terms of their component functions. Then, we have

$$\begin{aligned}(\mathbf{f}+\mathbf{g})(\mathbf{x}) &= \mathbf{f}(\mathbf{x})+\mathbf{g}(\mathbf{x})\\ &= (f_1(\mathbf{x}), ..., f_m(\mathbf{x}))+(g_1(\mathbf{x}), ..., g_m(\mathbf{x}))\\ &= (f_1(\mathbf{x})+g_1(\mathbf{x}), ..., f_m(\mathbf{x})+g_m(\mathbf{x}))\\ &= ((f_1+g_1)(\mathbf{x}), ..., (f_m+g_m)(\mathbf{x})),\end{aligned}$$

giving us the component functions of $\mathbf{f}+\mathbf{g}$.

Using the component functions found above, we take the derivative matrix of

Learning outcomes:
Author(s):

$\mathbf{f} + \mathbf{g}$, using the linearity of partial derivatives.

$$
\begin{aligned}
D(\mathbf{f} + \mathbf{g}) &= \begin{pmatrix} \frac{\partial(f_1 + g_1)}{\partial x_1} & \frac{\partial(f_1 + g_1)}{\partial x_2} & \cdots & \frac{\partial(f_1 + g_1)}{\partial x_n} \\ \frac{\partial(f_2 + g_2)}{\partial x_1} & \frac{\partial(f_2 + g_2)}{\partial x_2} & \cdots & \frac{\partial(f_2 + g_2)}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial(f_m + g_m)}{\partial x_1} & \frac{\partial(f_m + g_m)}{\partial x_2} & \cdots & \frac{\partial(f_m + g_m)}{\partial x_n} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial f_1}{\partial x_1} + \frac{\partial g_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} + \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} + \frac{\partial g_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} + \frac{\partial g_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} + \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} + \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} + \frac{\partial g_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} + \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} + \frac{\partial g_m}{\partial x_n} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix} + \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \frac{\partial g_1}{\partial x_2} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \frac{\partial g_2}{\partial x_1} & \frac{\partial g_2}{\partial x_2} & \cdots & \frac{\partial g_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_1} & \frac{\partial g_m}{\partial x_2} & \cdots & \frac{\partial g_m}{\partial x_n} \end{pmatrix} \\
&= D\mathbf{f} + D\mathbf{g}.
\end{aligned}
$$

Thus, we have $D(\mathbf{f} + \mathbf{g}) = D\mathbf{f} + D\mathbf{g}$ on $X \cap Y$.

Next, we need to show that $\mathbf{f} + \mathbf{g}$ is differentiable on $X \cap Y$. In order to show this, we will show that the following limit evaluates to 0, for $\mathbf{a} \in X \cap Y$. This is done by separating the $\mathbf{f}$ and $\mathbf{g}$ terms, using the triangle inequality, and using the fact that $\mathbf{f}$ and $\mathbf{g}$ are both differentiable on $X \cap Y$.

$$\begin{aligned}
&\lim_{\mathbf{x}\to\mathbf{a}} \frac{\|(\mathbf{f}+\mathbf{g})(\mathbf{x}) - ((\mathbf{f}+\mathbf{g})(\mathbf{a}) + D(\mathbf{f}+\mathbf{g})(\mathbf{a})(\mathbf{x}-\mathbf{a})\|}{\|\mathbf{x}-\mathbf{a}\|} \\
&\quad = \lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - \mathbf{g}(\mathbf{a}) - (D\mathbf{f}(\mathbf{a}) + D\mathbf{g}(\mathbf{a}))(\mathbf{x}-\mathbf{a})\|}{\|\mathbf{x}-\mathbf{a}\|} \\
&\quad = \lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) + \mathbf{g}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - \mathbf{g}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a}) - D\mathbf{g}(\mathbf{a})(\mathbf{x}-\mathbf{a})\|}{\|\mathbf{x}-\mathbf{a}\|} \\
&\quad = \lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a}) + \mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{a}) - D\mathbf{g}(\mathbf{a})(\mathbf{x}-\mathbf{a})\|}{\|\mathbf{x}-\mathbf{a}\|} \\
&\quad \le \lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})\| + \|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{a}) - D\mathbf{g}(\mathbf{a})(\mathbf{x}-\mathbf{a})\|}{\|\mathbf{x}-\mathbf{a}\|} \\
&\quad \le \lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}) - D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a})\|}{\|\mathbf{x}-\mathbf{a}\|} + \lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{a}) - D\mathbf{g}(\mathbf{a})(\mathbf{x}-\mathbf{a})\|}{\|\mathbf{x}-\mathbf{a}\|} \\
&\quad = \lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{f}(\mathbf{x}) - (\mathbf{f}(\mathbf{a}) + D\mathbf{f}(\mathbf{a})(\mathbf{x}-\mathbf{a}))\|}{\|\mathbf{x}-\mathbf{a}\|} + \lim_{\mathbf{x}\to\mathbf{a}} \frac{\|\mathbf{g}(\mathbf{x}) - (\mathbf{g}(\mathbf{a}) + D\mathbf{g}(\mathbf{a})(\mathbf{x}-\mathbf{a}))\|}{\|\mathbf{x}-\mathbf{a}\|} \\
&\quad = 0 + 0 \\
&\quad = 0
\end{aligned}$$

$\blacksquare$

You may also recall the constant multiple rule from single variable calculus. For example,

$$\begin{aligned}
\frac{d}{dx}(4x^2) &= 4\left(\frac{d}{dx}(x^2)\right) \\
&= 4(2x) \\
&= 8x.
\end{aligned}$$

We have an analogous result in multivariable calculus.

Proposition 26. *Suppose $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}^m$ is a differentiable function on a subset X of $\mathbb{R}^n$, and let k be any constant. Then $k\mathbf{f}$ is also differentiable on X, and $D(k\mathbf{f}) = kD\mathbf{f}$ on X.*

The proof of this proposition is somewhat similar to the previous theorem, and it is left as an exercise.

Although these are important results, they actually aren't particularly useful for computing derivative matrices. In practice, you'd compute the derivative matrix of a sum of functions by first adding the components, then differentiating. We include an example to demonstrate this.

Example 80. *Consider the functions $\mathbf{f}, \mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$ given by $\mathbf{f}(x, y) = (x^2, xy)$ and $\mathbf{g}(x, y) = (xy^2, y)$.*

We will compute the derivative matrix $D(\mathbf{f}+\mathbf{g})$ *in two ways: first directly, and then by using the sum rule proved above.*

The function $\mathbf{f}+\mathbf{g}$ *is given by*

$$(\mathbf{f}+\mathbf{g})(x,y) = \boxed{(x^2+xy^2, xy+y)}.$$

We can then compute the derivative matrix $D(\mathbf{f}+mathbf g)$ *directly as*

$$D(\mathbf{f}+\mathbf{g}) = \begin{pmatrix} \boxed{2x+y^2} & \boxed{2xy} \\ \boxed{y} & \boxed{x+1} \end{pmatrix}.$$

We will now use our multivariable sum rule. We have

$$D\mathbf{f}(x,y) = \begin{pmatrix} \boxed{2x} & \boxed{0} \\ \boxed{y} & \boxed{x} \end{pmatrix},$$

and

$$D\mathbf{g}(x,y) = \begin{pmatrix} \boxed{y^2} & \boxed{2xy} \\ \boxed{0} & \boxed{1} \end{pmatrix}.$$

Adding these matrices, we then have that

$$\begin{aligned} D(\mathbf{f}+\mathbf{g})(x,y) &= D\mathbf{f}(x,y) + D\mathbf{g}(x,y) \\ &= \begin{pmatrix} \boxed{2x+y^2} & \boxed{2xy} \\ \boxed{y} & \boxed{x+1} \end{pmatrix}. \end{aligned}$$

We get the correct answer using either method, but using the sum rule doesn't seem to provide much (if any) of a computational advantage over computing the derivative matrix directly. Nonetheless, it is mathematically important that the derivative seems to "distribute" over addition.

Product and Quotient laws

We will now try to find multi-variable analogs to the product and quotient rules from single variable calculus. Let's start by considering when these rules might make sense.

Suppose we have functions $\mathbf{f}(x,y) = (x^2, xy)$ and $\mathbf{g}(x,y) = (xy^2, y)$. If we wanted to define the product of these functions, what would that mean? The outputs of $\mathbf{f}$ and $\mathbf{g}$ are vectors, so we'd be trying to multiply two vectors - but there isn't really a clear multiplication on vectors,! We could try multiplying component-wise, taking the dot product, or taking the cross product, and in different settings, these all might be reasonable things to do. We could work on finding product rules for all of the different ways we could "multiply" two vectors (these can be found in the exercises), but we'll save some time, and focus on a case where we do have one clear choice for multiplication: *scalar-valued functions.*

Proposition 27. *Suppose* $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}$ *and* $\mathbf{g} : Y \subset \mathbb{R}^n \to \mathbb{R}$ *are differentiable scalar-valued functions on subsets* X *and* Y *of* $\mathbb{R}^n$*, respectively. Then* fg *is differentiable on* $X \cap Y$*, and*

$$D(fg)(\mathbf{a}) = g(\mathbf{a})Df(\mathbf{a}) + f(\mathbf{a})Dg(\mathbf{a})$$

for $\mathbf{a}$ *in* $X \cap Y$*.*

Similarly, we have a multi-variable quotient rule for scalar valued functions.

Proposition 28. *Suppose* $\mathbf{f} : X \subset \mathbb{R}^n \to \mathbb{R}$ *and* $\mathbf{g} : Y \subset \mathbb{R}^n \to \mathbb{R}$ *are differentiable scalar-valued functions on subsets* X *and* Y *of* $\mathbb{R}^n$*, respectively. Then* $\dfrac{f}{g}$ *is differentiable on* $X \cap Y$*, and*

$$D\left(\frac{f}{g}\right)(\mathbf{a}) = \frac{g(\mathbf{a})Df(\mathbf{a}) - f(\mathbf{a})Dg(\mathbf{a})}{(g(\mathbf{a}))^2}$$

for $\mathbf{a}$ *in* $X \cap Y$*.*

We'll leave the proofs of these results as exercises, as they are similar to the single variable proofs.

Chain Rule

In this activity, we introduce the multi-variable chain rule, and we use it to compute derivatives of compositions of functions.

The Chain Rule

Let's begin by recalling the chain rule from single variable calculus. If we have differentiable functions f and g, then we can compute the derivative of the composition as $(f \circ g)'(x) = f'(g(x))g'(x)$.

Example 81. *Let $f(x) = \sin(x)$ and $g(x) = x^2$. Then $f'(x) = \cos(x)$, $g'(x) = 2x$, and we can differentiate the composition $(f \circ g)(x) = \sin(x^2)$ using the chain rule:*

$$\begin{aligned}(f \circ g)'(x) &= f'(g(x))g'(x) \\ &= \cos(g(x)) \cdot 2x \\ &= \cos(x^2) \cdot 2x.\end{aligned}$$

We can also use the chain rule to differentiate the composition $(g \circ f)(x) = \sin^2(x)$.

$$\begin{aligned}(g \circ f)'(x) &= g'(f(x))f'(x) \\ &= 2(f(x))\cos(x) \\ &= 2\sin(x)\cos(x)\end{aligned}$$

The multi-variable chain rule is similar, with the derivative matrix taking the place of the single variable derivative, so that the chain rule will involve matrix multiplication. We also need to pay extra attention to whether the composition of functions is even defined.

Theorem 8. *Suppose $\mathbf{f} : Y \subset \mathbb{R}^p \to \mathbb{R}^n$ and $\mathbf{g} : X \subset \mathbb{R}^m \to \mathbb{R}^p$ are defined on open sets $Y \subset \mathbb{R}^p$ and $X \subset \mathbb{R}^m$, respectively. Suppose that $\mathbf{g}(X) \subset Y$, so the image of $\mathbf{g}$ is contained in the domain of $\mathbf{f}$. Suppose further that $\mathbf{g}$ is differentiable at some point $\mathbf{x}_0 \in X$, and that $\mathbf{f}$ is differentiable at $\mathbf{y}_0 = \mathbf{g}(\mathbf{x}_0) \in Y$.*

Then the composition $\mathbf{f} \circ \mathbf{g}$ is differentiable at $\mathbf{x}_0$, and

$$\begin{aligned}D(\mathbf{f} \circ \mathbf{g})(\mathbf{x}_0) &= D\mathbf{f}(\mathbf{y}_0)D\mathbf{g}(\mathbf{x}_0) \\ &= D\mathbf{f}(\mathbf{g}(\mathbf{x}_0))D\mathbf{g}(\mathbf{x}_0).\end{aligned}$$

Learning outcomes:
Author(s):

Although the conditions sound complicated, essentially they're just requiring that all of the derivatives mentioned actually exist. Note the similarities to the single variable chain rule.

Proof super great proof of chain rule ■

A Special Case

We'll now consider a special case of the chain rule, when we have a composition $f \circ g$ of functions $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{g} : \mathbb{R} \to \mathbb{R}^n$. Note that f is a scalar function, and we can think of $\mathbf{g}$ as a curve in $\mathbb{R}^n$.

Let's look at what the chain rule tells us in this case. For any $x_0 \in \mathbb{R}$, we have

$$D(f \circ \mathbf{g})(x_0) = Df(\mathbf{g}(x_0))D\mathbf{g}(x_0).$$

Writing $\mathbf{g}(x) = (g_1(x), ..., g_n(x))$ in terms of its components, we have

$$D\mathbf{g} = \begin{pmatrix} \partial g_1/\partial x \\ \vdots \\ \partial g_n/\partial x \end{pmatrix}.$$

Since $\mathbf{g}$ only has one input variable, we can rewrite this as

$$D\mathbf{g}(x) = \begin{pmatrix} g_1'(x) \\ \vdots \\ g_n'(x) \end{pmatrix}.$$

Now that we've sorted out $D\mathbf{g}$, let's consider Df. Since f is a scalar-valued function, Df will consist of only one row,

$$Df = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} & \cdots & \dfrac{\partial f}{\partial x_n} \end{pmatrix}.$$

For $Df(\mathbf{g}(x_0))$, we would evaluate these partial derivatives at $\mathbf{g}(x_0)$:

$$Df(\mathbf{g}(x_0)) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(\mathbf{g}(x_0)) & \cdots & \dfrac{\partial f}{\partial x_n}(\mathbf{g}(x_0)) \end{pmatrix}.$$

Now let's turn our attention back to the composition $f \circ \mathbf{g}$. Putting together our results from above, we have

$$\begin{aligned} D(f \circ \mathbf{g})(x_0) &= Df(\mathbf{g}(x_0))D\mathbf{g}(x_0) \\ &= \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(\mathbf{g}(x_0)) & \cdots & \dfrac{\partial f}{\partial x_n}(\mathbf{g}(x_0)) \end{pmatrix} \begin{pmatrix} g_1'(x_0) \\ \vdots \\ g_n'(x_0) \end{pmatrix} \\ &= \left(\left(\frac{\partial f}{\partial x_1}(\mathbf{g}(x_0)) \right) g_1'(x_0) + \cdots + \left(\frac{\partial f}{\partial x_n}(\mathbf{g}(x_0)) \right) g_n'(x_0) \right) \end{aligned}$$

Since $f \circ \mathbf{g} : \mathbb{R} \to \mathbb{R}$ is a single variable function, its derivative matrix at x_0 only has one entry, which is $\dfrac{df \circ \mathbf{g}}{dx}(x_0)$. So, we can rewrite the above as

$$\frac{df \circ \mathbf{g}}{dx}(x_0) = \left(\frac{\partial f}{\partial x_1}(\mathbf{g}(x_0))\right) g_1'(x_0) + \cdots + \left(\frac{\partial f}{\partial x_n}(\mathbf{g}(x_0))\right) g_n'(x_0).$$

This gives us a special case of the Chain Rule, that can be useful when we have a composition of functions $\mathbb{R} \to \mathbb{R}^n \to \mathbb{R}$.

Examples

Example 82. *Let's consider the functions* $\mathbf{f}(x, y) = (x^2 + y^2, xy)$ *and* $\mathbf{g}(x, y) = (3xy, x - y, 7y^2)$. *First, let's decide how we can compose these functions.*

Which composition(s) exist?

Multiple Choice:

(a) *Neither* $\mathbf{f} \circ \mathbf{g}$ *nor* $\mathbf{g} \circ \mathbf{f}$ *exists.*

(b) $\mathbf{f} \circ \mathbf{g}$ *exists, but* $\mathbf{g} \circ \mathbf{f}$ *does not.*

(c) $\mathbf{g} \circ \mathbf{f}$ *exists, but* $\mathbf{f} \circ \mathbf{g}$ *does not.* ✓

(d) $\mathbf{f} \circ \mathbf{g}$ *and* $\mathbf{g} \circ \mathbf{f}$ *both exist.*

We'll compute the derivative matrix $D(\mathbf{g} \circ \mathbf{f}$ *in two ways: using the chain rule, and directly.*

Let's begin by using the chain rule. We'll have

$$D(\mathbf{g} \circ \mathbf{f})(\mathbf{x}) = D\mathbf{g}(\mathbf{f}(\mathbf{x}))D\mathbf{f}(\mathbf{x}),$$

so we'll start by computing the derivative matrices $D\mathbf{f}$ *and* $D\mathbf{g}$.

$$D\mathbf{f}(x, y) = \begin{pmatrix} \boxed{2x} & \boxed{2y} \\ \boxed{y} & \boxed{x} \end{pmatrix}$$

$$D\mathbf{g}(x, y) = \begin{pmatrix} \boxed{3y} & \boxed{3x} \\ \boxed{1} & \boxed{-1} \\ \boxed{0} & \boxed{14y} \end{pmatrix}$$

Now, for $D\mathbf{g}(\mathbf{f}(\mathbf{x}))$, *we need to input* $\mathbf{f}(x, y)$ *into* $D\mathbf{g}$.

$$\begin{aligned} D\mathbf{g}(\mathbf{f}(x, y)) &= D\mathbf{g}(x^2 + y^2, xy) \\ &= \begin{pmatrix} \boxed{3(xy)} & \boxed{3(x^2 + y^2)} \\ \boxed{1} & \boxed{-1} \\ \boxed{0} & \boxed{14(xy)} \end{pmatrix} \end{aligned}$$

To compute $D(\mathbf{g} \circ \mathbf{f})(x,y)$, we multiply matrices, and obtain

$$D(\mathbf{g}\circ\mathbf{f})(x,y) = \begin{pmatrix} \boxed{3(xy)} & \boxed{3(x^2+y^2)} \\ \boxed{1} & \boxed{-1} \\ \boxed{0} & \boxed{14(xy)} \end{pmatrix} \begin{pmatrix} \boxed{2x} & \boxed{2y} \\ \boxed{y} & \boxed{x} \end{pmatrix}$$
$$= \begin{pmatrix} \boxed{9x^2y+3y^3} & \boxed{3x^3+9xy^2} \\ \boxed{2x-y} & \boxed{2y-x} \\ \boxed{14xy^2} & \boxed{14x^2y} \end{pmatrix}.$$

Let's verify our answer, by computing the $D(\mathbf{g} \circ \mathbf{f})(x,y)$ directly, without using the chain rule. We'll begin by finding $\mathbf{g} \circ \mathbf{f}$.

$$\begin{aligned}(\mathbf{g}\circ\mathbf{f})(x,y) &= \mathbf{g}(f(x,y)) \\ &= \mathbf{g}(x^2+y^2, xy) \\ &= \boxed{(3(x^2+y^2)(xy), (x^2+y^2)-(xy), 7(xy)^2)}.\end{aligned}$$

This simplifies to $(\mathbf{g} \circ \mathbf{f})(x,y) = (3x^3y + 3xy^3, x^2 + y^2 - xy, 7x^2y^2)$. We can then compute the derivative matrix.

$$(\mathbf{g}\circ\mathbf{f})(x,y) = \begin{pmatrix} \boxed{9x^2y+3y^3} & \boxed{3x^3+9xy^2} \\ \boxed{2x-y} & \boxed{2y-x} \\ \boxed{14xy^2} & \boxed{14x^2y} \end{pmatrix}$$

We see that this gives the same result as using the chain rule.

Example 83. *special case*

Example 84. *Consider the function $g : \mathbb{R}^2 \to \mathbb{R}^2$, given by $\mathbf{g}(r,\theta) = (r\cos\theta, r\sin\theta)$. We can think of this function as converting from polar coordinates to Cartesian coordinates. Suppose we have a function $f : \mathbb{R}^2 \to \mathbb{R}$, and we want to know how f changes with respect to polar coordinates. We could find this by computing*

$$D(f\circ\mathbf{g}) = \begin{pmatrix} \frac{\partial(f\circ\mathbf{g})}{\partial r} & \frac{\partial(f\circ\mathbf{g})}{\partial\theta} \end{pmatrix}.$$

For specific functions f, we could probably compute this derivative matrix directly. However, if we use the chain rule, we can find a general formula for the derivative matrix. The chain rule tells us

$$D(f\circ\mathbf{g})(r,\theta) = Df(\mathbf{g}(r,\theta))D\mathbf{g}(r,\theta).$$

We don't know precisely what Df will be, but we can write it in terms of its partial derivatives:

$$Df = \begin{pmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} \end{pmatrix}.$$

Given $\mathbf{g}(r,\theta) = (r\cos\theta, r\sin\theta)$*, we can compute the derivative matrix of* $\mathbf{g}$*.*

$$\begin{aligned} D\mathbf{g}(r,\theta) &= \begin{pmatrix} \partial g_1/\partial r & \partial g_1/\partial\theta \\ \partial g_2 \partial r & \partial g_2/\partial\theta \end{pmatrix} \\ &= \begin{pmatrix} \boxed{\cos\theta} & \boxed{-r\sin\theta} \\ \boxed{\sin\theta} & \boxed{r\cos\theta} \end{pmatrix} \end{aligned}$$

Then, from the chain rule, we have

$$\begin{aligned} D(f\circ\mathbf{g})(r,\theta) &= Df(\mathbf{g}(r,\theta))D\mathbf{g}(r,\theta) \\ &= \begin{pmatrix} \frac{\partial f}{\partial x}(\mathbf{g}(r,\theta)) & \frac{\partial f}{\partial y}(\mathbf{g}(r,\theta)) \end{pmatrix} \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta\frac{\partial f}{\partial x}(\mathbf{g}(r,\theta)) + \sin\theta\frac{\partial f}{\partial y}(\mathbf{g}(r,\theta)) & -r\sin\theta\frac{\partial f}{\partial x}(\mathbf{g}(r,\theta)) + r\cos\theta\frac{\partial f}{\partial y}(\mathbf{g}(r,\theta)) \end{pmatrix} \end{aligned}$$

Since

$$D(f\circ\mathbf{g}) = \begin{pmatrix} \frac{\partial(f\circ\mathbf{g})}{\partial r} & \frac{\partial(f\circ\mathbf{g})}{\partial\theta} \end{pmatrix},$$

we then have that

$$\begin{aligned} \frac{\partial(f\circ\mathbf{g})}{\partial r} &= \cos\theta\frac{\partial f}{\partial x}(\mathbf{g}(r,\theta)) + \sin\theta\frac{\partial f}{\partial y}(\mathbf{g}(r,\theta)) \\ &= \cos\theta\frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta) + \sin\theta\frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta), \end{aligned}$$

and

$$\begin{aligned} \frac{\partial(f\circ\mathbf{g})}{\partial\theta} &= -r\sin\theta\frac{\partial f}{\partial x}(\mathbf{g}(r,\theta)) + r\cos\theta\frac{\partial f}{\partial y}(\mathbf{g}(r,\theta)) \\ &= -r\sin\theta\frac{\partial f}{\partial x}(r\cos\theta, r\sin\theta) + r\cos\theta\frac{\partial f}{\partial y}(r\cos\theta, r\sin\theta). \end{aligned}$$

Conclusion

Directional Derivatives

In order to find how a function $f : \mathbb{R}^n \to \mathbb{R}$ changes with each of the input variables, we defined the partial derivatives of f. For example, when $n = 2$, we defined the partial derivative of f with respect to x to be

$$f_x(x,y) = \lim_{h \to 0} \frac{f(x+h,y) - f(x,y)}{h}.$$

Here, we thought of y as a constant, which made f only a function of x, and reduced us to a single variable derivative. This told us how a small change in x would after the value of f, if we kept y constant. In other words, the partial derivatives described how the function f was changing in the positive x-direction and in the positive y-direction.

PICTURE

But what if we want to find how f changes if we change both x and y? One possible way to do this would be to change x and y by the same amount, which would be equivalent to finding how f changes as we move along the line $y = x$.

PICTURE

Alternatively, we could change y by twice as much as x. This would be equivalent to finding how f changes as we move along the line $y = 2x$.

PICTURE

As you can see, there are many different ways that we can change x and y, corresponding to different directions in the xy-plane. In order to determine how f changes as we move in all of these different directions, we will now define directional derivatives.

Directional derivatives

We would like to compute the instantaneous rate of change of a function $f : \mathbb{R}^n \to \mathbb{R}$ as we move in some given direction in $\mathbb{R}^n$. We will model our definition after partial derivatives and single variable derivatives, and use a unit vector $\vec{v}$ to describe the direction.

Definition 58. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, a point $\vec{a} \in \mathbb{R}^n$, and a direction given by a unit vector $\vec{v} \in \mathbb{R}^n$. Then we define the* directional derivative *of*

Learning outcomes:
Author(s):

f at $\vec{a}$ in the direction of $\vec{v}$ to be

$$D_{\vec{v}} f(\vec{a}) = \lim_{h \to 0} \frac{f(\vec{a} + h\vec{v}) - f(\vec{a})}{h},$$

provided this limit exists.

Noticing that by looking at $f(\vec{a} + h\vec{v})$, we are finding the value of f when we move a small distance, h, in the direction of $\vec{v}$ from the point $\vec{a}$.

VISUAL

When computing directional derivatives, it's important to remember that the direction must be given by a *unit* vector. Otherwise, the length of the vector will change the value of the limit above. If you'd like to find a directional derivative in a direction given by a non-unit vector $\vec{w}$, you should normalize $\vec{w}$ to unit length.

Example 85. *We'll compute the directional derivative of $f(x, y) = x^2 y + y^2$ at $\vec{a} = (2, 0)$, in the direction of $(3, 4)$.*

Since $(3, 4)$ isn't a unit vector, we need to normalize it. Since $\|(3, 4)\| = \sqrt{3^2 + 4^2} = 5$, we'll use the vector $\vec{v} = \left(\frac{3}{5}, \frac{4}{5}\right)$ to compute our desired directional derivative.

$$\begin{aligned} D_{\vec{v}} f(\vec{a}) &= \lim_{h \to 0} \frac{f(\vec{a} + h\vec{v}) - f(\vec{a})}{h} \\ &= \lim_{h \to 0} \frac{f\left((2, 0) + h\left(\frac{3}{5}, \frac{4}{5}\right)\right) - f(2, 3)}{h} \\ &= \lim_{h \to 0} \frac{f\left(2 + \frac{3}{5}h, \frac{4}{5}h\right) - f(2, 0)}{h} \\ &= \lim_{h \to 0} \frac{\left(2 + \frac{3}{5}h\right)^2 \cdot \frac{4}{5}h + \left(\frac{4}{5}h\right)^2 - 0}{h} \\ &= \lim_{h \to 0} \left(2 + \frac{3}{5}h\right)^2 \cdot \frac{4}{5} + \left(\frac{4}{5}\right)^2 h \\ &= 4 \cdot \frac{4}{5} \\ &= \frac{16}{5}. \end{aligned}$$

Fortunately, we won't always need to resort to evaluating directional derivatives using the limit definition. We'll soon see how we can use the gradient to compute directional derivatives.

The Gradient and Level Sets

We've defined the directional derivatives of a function, which allow us to determine how a function is changing in various directions.

Definition 59. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, a point $\vec{a} \in \mathbb{R}^n$, and a direction given by a unit vector $\vec{v} \in \mathbb{R}^n$. Then we define the* directional derivative *of f at $\vec{a}$ in the direction of $\vec{v}$ to be*

$$D_{\vec{v}} f(\vec{a}) = \lim_{h \to 0} \frac{f(\vec{a} + h\vec{v}) - f(\vec{a})}{h},$$

provided this limit exists.

However, we would like an easier way to evaluate directional derivatives, that doesn't require the limit definition.

Such a method will require use of the gradient of the function. Recall our definition of the gradient of a function $f : \mathbb{R}^n \to \mathbb{R}$.

Definition 60. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. The* gradient *of f is the function $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$, defined by*

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, ..., \frac{\partial f}{\partial x_n} \right).$$

We'll see that this vector turns out to be closely related to directional derivatives.

The gradient and directional derivatives

Let's suppose we have a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, and consider our definition of the directional derivative a function f at $\vec{a}$ in the direction of a unit vector $\vec{v}$, which was

$$D_{\vec{v}} f(\vec{a}) = \lim_{h \to 0} \frac{f(\vec{a} + h\vec{v}) - f(\vec{a})}{h}.$$

We'll rewrite this definition by considering another function, $F(h) = f(\vec{a} + h\vec{v})$. Notice that F is a single variable function, and when we rewrite the directional

Learning outcomes:
Author(s):

derivative, we have

$$\begin{aligned}
D_{\vec{v}}f(\vec{a}) &= \lim_{h\to 0}\frac{f(\vec{a}+h\vec{v})-f(\vec{a})}{h}\\
&= \lim_{h\to 0}\frac{F(h)-F(0)}{h-0}\\
&= F'(0).
\end{aligned}$$

So, the directional derivative is just the derivative of this single variable function $F(h)$ evaluated at 0. Revisiting our definition of $F(h)$, we can use the chain rule to find the derivative of F.

$$\begin{aligned}
\frac{d}{dh}F(h) &= \nabla f(\vec{a}+h\vec{v})\cdot\frac{d}{dh}(\vec{a}+h\vec{v})\\
&= \nabla f(\vec{a}+h\vec{v})\cdot\vec{v}
\end{aligned}$$

Evaluating at $h=0$, we have

$$\begin{aligned}
D_{\vec{v}}f(\vec{a}) &= F'(0)\\
&= \nabla f(\vec{a})\cdot\vec{v}.
\end{aligned}$$

Thus, we have arrived at the following result.

Theorem 9. *Suppose $f:\mathbb{R}^n\to\mathbb{R}$ is differentiable at $\vec{a}\in\mathbb{R}^n$. Then $D_{\vec{v}}f(\vec{a})$ exists for all unit vectors $\vec{v}\in\mathbb{R}^n$, and*

$$D_{\vec{v}}f(\vec{a}) = \nabla f(\vec{a})\cdot\vec{v}.$$

Let's use this result to compute some directional derivatives.

Example 86. *We'll compute the directional derivative of $f(x,y)=x^2y+y^2$ at $\vec{a}=(2,0)$, in the direction of $(3,4)$. (We previously computed this directional derivative using the limit definition.)*

Since $(3,4)$ isn't a unit vector, we need to normalize it. Since $\|(3,4)\| = \sqrt{3^2+4^2}=5$, we'll use the vector $\vec{v}=\left(\frac{3}{5},\frac{4}{5}\right)$ to compute our desired directional derivative.

Next, we'll need the gradient of f.

$$\nabla f(x,y) = (2xy, x^2+2y)$$

Since the partial derivatives of f are polynomials, they are continuous, so f is differentiable. Thus, we can use the above theorem to compute the directional derivative.

Then, we can compute the directional derivative as

$$\begin{aligned} D_{\vec{v}} f(\vec{a}) &= \nabla f(\vec{a}) \cdot \vec{v} \\ &= \nabla f(2,0) \cdot \left(\frac{3}{5}, \frac{4}{5}\right) \\ &= (0,4) \cdot \left(\frac{3}{5}, \frac{4}{5}\right) \\ &= \frac{16}{5}. \end{aligned}$$

This matches what we had previously computed using the definition of directional derivatives.

Problem 14 *Compute the directional derivative of* $f(x, y, z) = 3xy + xz^2$ *at* $\vec{a} = (2, 0, 1)$*, in the direction of* $(2, 2, 1)$*.*

$$D_{\vec{v}} f(\vec{a}) = \boxed{6}$$

Compute the directional derivative of $f(x, y, z) = 3xy + xz^2$ *at* $\vec{a} = (2, 0, 1)$*, in the direction of* $(-2, 1, -1)$*.*

$$D_{\vec{v}} f(\vec{a}) = \boxed{0}$$

The gradient and level sets

We've shown that for a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, we can compute directional derivatives as

$$D_{\vec{v}} f(\vec{a}) = \nabla f(\vec{a}) \cdot \vec{v}.$$

What does this mean for the possible values for a directional derivative? Recall that the dot product can be computed as

$$\nabla f(\vec{a}) \cdot \vec{v} = \|\nabla f(\vec{a})\| \, \|\vec{v}\| \cos\theta,$$

Where θ is the angle between the two vectors. Since $\vec{v}$ is a unit vector, we have

$$\|\nabla f(\vec{a})\| \, \|\vec{v}\| \cos\theta = \|\nabla f(\vec{a})\| \cos\theta.$$

Since $-1 \leq \cos\theta \leq 1$, we have that

$$-\|\nabla f(\vec{a})\| \leq D_{\vec{v}} f(\vec{a}) \leq \|\nabla f(\vec{a})\|.$$

In particular, the largest that $D_{\vec{v}} f(\vec{a})$ can be is $\|\nabla f(\vec{a})\|$, and this occurs when $\vec{v}$ points in the same direction as $\nabla f(\vec{a})$ (so $\theta = 0$). Thus, the gradient points in the direction of greatest increase.

VISUAL

On the other hand, the minimum value that $D_{\vec{v}} f(\vec{a})$ can have is $-\|\nabla f(\vec{a})\|$, and this occurs when $\vec{v}$ points in the opposite direction from $\nabla f(\vec{a})$, in the direction of $-\nabla f(\vec{a})$. Thus, $-\nabla f(\vec{a})$ points in the direction of greatest decrease.

VISUAL

Additionally, from $D_{\vec{v}} f(\vec{a}) = \nabla f(\vec{a}) \cdot \vec{v}$, we can see that $\vec{v}$ is perpendicular to $\nabla f(\vec{a})$ if and only if $D_{\vec{v}} f(\vec{a}) = 0$. But what does it mean for $D_{\vec{v}} f(\vec{a})$? This means that there is no instantaneous change in f in the direction of $\vec{v}$, which means that $\vec{v}$ will be a tangent vector to a level curve.

VISUAL

We'll state this observation more formally, and prove that the gradient is perpendicular to the level curves.

Theorem 10. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, and suppose f is of class $\mathcal{C}^1$. For some constant c, consider the level set*

$$S = \{\vec{x} \in \mathbb{R}^n \ : \ f(\vec{x}) = c\}.$$

Then, for any point $\vec{x}_0$ in S, the gradient $\nabla f(\vec{x}_0)$ is perpendicular to S.

Proof Let $\vec{x}(t)$ be a parametrization of S, and let $\vec{x}_0 = \vec{x}(t_0)$. In order to show that $\nabla f(\vec{x}_0)$ is perpendicular to S, we will show that the gradient $\nabla f(\vec{x}_0)$ is perpendicular to the velocity vector $\vec{x}'(t_0)$. By the definition of S,

$$f(\vec{x}(t)) = c,$$

for all t. Differentiating both sides of this identity, and using the chain rule on the left side, we obtain

$$\nabla f(\vec{x}(t)) \cdot \vec{x}'(t) = 0.$$

Plugging in $t = t_0$, this gives us

$$\nabla f(\vec{x}(t_0)) \cdot \vec{x}'(t_0) = 0,$$

which we can rewrite as

$$\nabla f(\vec{x}_0) \cdot \vec{x}'(t_0) = 0.$$

Thus, we have shown that $\nabla f(\vec{x}_0)$ is perpendicular to the level set S.

THIS PROOF ONLY WORKS WHEN IT'S A LEVEL CURVE, NOT A LEVEL SURFACE ETC. ■

Implicit Curves and Surfaces

By studying directional derivatives and their relationship to the gradient, we observed that the gradient of a function is always perpendicular to its level curves. That is, we proved the following theorem.

Theorem 11. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$, and suppose f is of class $\mathcal{C}^1$. For some constant c, consider the level set*

$$S = \{\vec{x} \in \mathbb{R}^n \ : \ f(\vec{x}) = c\}.$$

Then, for any point $\vec{x}_0$ in S, the gradient $\nabla f(\vec{x}_0)$ is perpendicular to S.

Now, suppose we have a curve defined implicitly. For example, the *Folium of Descartes* is the curve defined by the equation

$$x^3 + y^3 = 3xy.$$

PICTURE

Suppose we wanted to find an equation for the tangent line to this curve at the point $\left(\frac{2}{3}, \frac{4}{3}\right)$. There are a couple ways we could approach this problem. We could try to solve for y in terms of x, and use the standard single variable methods to find an equation for the tangent line. We could also use implicit differentiation, and find an equation for the tangent line that way.

We will now introduce a new method for finding an equation for the tangent line to this curve, that will use our above observations about the gradient of a function and level sets.

Tangent lines

Example 87. *Consider the Folium of Descartes, C, defined by $x^3 + y^3 = 3xy$. We will find an equation for the tangent line to this curve at the point $\left(\frac{2}{3}, \frac{4}{3}\right)$.*

PICTURE

If we define a function $f(x, y) = x^3 + y^3 - 3xy$, then the curve C is a level set of f. In particular, it's the level set $f(x, y) = 0$. Then the gradient of f will always be perpendicular to C. This means that the tangent line to C at the point

Learning outcomes:
Author(s):

$\left(\frac{2}{3},\frac{4}{3}\right)$ *will be perpendicular to* $\nabla f\left(\frac{2}{3},\frac{4}{3}\right)$. *In order to take advantage of this fact, we'll compute the gradient of* f.

$$\nabla f(x,y) = (3x^2 - 3y, 3y^2 - 3x)$$

At the point $\left(\frac{2}{3},\frac{4}{3}\right)$,

$$\nabla f\left(\frac{2}{3},\frac{4}{3}\right) = \left(-\frac{8}{3},\frac{10}{3}\right).$$

Suppose we have a point $\vec{x} = (x,y)$ *in the tangent plane. Since* $\vec{a} = \left(\frac{2}{3},\frac{4}{3}\right)$ *is also in the tangent plane, the vector between these points,* $\vec{x}-\vec{a}$, *will be contained in the tangent plane.*

PICTURE

Thus, $\vec{x}-\vec{a}$ *must be perpendicular to* $\nabla f\left(\frac{2}{3},\frac{4}{3}\right)$. *That is, we must have*

$$\nabla f\left(\frac{2}{3},\frac{4}{3}\right)\cdot(\vec{x}-\vec{a}) = 0,$$

which we can rewrite as

$$\left(-\frac{8}{3},\frac{10}{3}\right)\cdot\left(x-\frac{2}{3},y-\frac{4}{3}\right) = 0.$$

Expanding the dot product, we obtain the equation,

$$\frac{8}{3}\left(x-\frac{2}{3}\right)+\frac{10}{3}\left(y-\frac{4}{3}\right) = 0,$$

which is an equation for the tangent line to C *at the point* $\left(\frac{2}{3},\frac{4}{3}\right)$.

Tangent planes

We can use a similar method to find equation for tangent planes to surfaces.

Example 88. *We'll find the tangent plane to the surface* S *defined by the equation* $z^2 + yz = x^2 + xy$ *in* $\mathbb{R}^3$, *at the point* $(1,1,1)$.

PICTURE

Notice that we can think of S *as a level set of the function*

$$f(x,y,z) = z^2 + yz - x^2 - xy.$$

In particular, S is the level set $f(x, y, z) = \boxed{0}$.

Knowing that the tangent plane will be perpendicular to the gradient of f, we compute

$$\nabla f(x, y, z) = \boxed{(-2x - y, z - x, 2z + y)}.$$

At the point $(1, 1, 1)$,

$$\nabla f(1, 1, 1) = \boxed{(-3, 0, 3)}.$$

So, the tangent plane will be the plane perpendicular to $(-3, 0, 3)$, *which passes through the point* $(1, 1, 1)$. *This plane is defined by the equation*

$$(-3, 0, 3) \cdot (x - 1, y - 1, z - 1) = 0,$$

Which can be expanded to

$$-3(x - 1) + 3(z - 1) = 0.$$

Part V

Behavior of Functions

Review of Taylor Polynomials

In single variable calculus, we were able to use derivatives to approximate functions through Taylor polynomials. We were able to do this because derivatives encode significant information about the behavior of a function. In the best situations, a function is completely determined (up to a constant) by its derivatives. Approximating functions with polynomial is incredibly useful for applications and computations, since computations involving polynomials are typically much simpler than computations involving arbitrary functions.

We would like to do something similar in multivariable calculus, and we will use derivatives to investigate the behavior of functions. Before we embark towards this goal, we'll begin by reviewing Taylor polynomials in single variable calculus, and see how we can immediately extend these to multivariable functions in some special cases.

Review of Taylor polynomials

In single variable calculus, we defined Taylor polynomials using the derivatives of a function f. The idea behind this definition is to define a polynomial that will have the same derivatives as f, up to the degree of the polynomial.

Definition 61. *Let $f : \mathbb{R} \to \mathbb{R}$ be a function that the derivatives $f'(a), f''(a), ..., f^{(n)}(a)$ exist for given $a \in \mathbb{R}$ (up to and including the nth derivative). Then we define the nth degree Taylor polynomial of f centered at a to be*

$$p_n(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n.$$

Note that the kth term of this polynomial is

$$\frac{f^{(k)}(a)}{k!}(x-a)^k.$$

We'll show that the first and second derivatives of $f(x)$ and $p_n(x)$ at a are the same for $n \geq 2$, and we'll leave this verification for higher derivatives as an

Learning outcomes:
Author(s):

exercise. Taking the first derivative of $p_n(x)$, we obtain

$$\begin{aligned}
p_n'(x) &= \frac{d}{dx}\left(f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n\right) \\
&= 0 + f'(a) + f''(a)(x-a) + \frac{f'''(a)}{2}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{(n-1)!}(x-a)^{n-1}.
\end{aligned}$$

Plugging in $x = a$, we have

$$\begin{aligned}
p_n'(a) &= f'(a) + f'(a)(a-a) + \frac{f'''(a)}{2}(a-a)^2 + \cdots + \frac{f^{(n)}(a)}{(n-1)!}(a-a)^{n-1} \\
&= f'(a),
\end{aligned}$$

so we see that $p_n'(a) = f'(a)$. Now let's check the second derivatives. Differentiating $p_n'(x)$, we obtain

$$p_n''(x) = 0 + f''(a) + f'''(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{(n-2)!}(x-a)^{n-2}.$$

Plugging in $x = a$, we have

$$\begin{aligned}
p_n''(a) &= f''(a) + f'''(a)(a-a) + \cdots + \frac{f^{(n)}(a)}{(n-2)!}(a-a)^{n-2} \\
&= f''(a).
\end{aligned}$$

So we see that the second derivatives match as well, and hopefully the pattern is clear enough to believe that higher order derivatives will match as well.

We'll do an example of computation a Taylor polynomial, and then review some important Taylor polynomials.

Example 89. *We'll find the fifth degree Taylor polynomial of $f(x) = \sin(x)$ centered at $x = 0$. First, we compute the first five derivatives of f.*

$$\begin{aligned}
f'(x) &= \boxed{\cos(x)} \\
f''(x) &= \boxed{-\sin(x)} \\
f'''(x) &= \boxed{-\cos(x)} \\
f^{(4)}(x) &= \boxed{\sin(x)} \\
f^{(5)}(x) &= \boxed{\cos(x)}
\end{aligned}$$

Plugging in $x = 0$, we have

$$\begin{aligned} f'(0) &= \boxed{1} \\ f''(0) &= \boxed{0} \\ f'''(0) &= \boxed{-1} \\ f^{(4)}(0) &= \boxed{0} \\ f^{(5)}(0) &= \boxed{1}. \end{aligned}$$

From this, we can find the fifth degree Taylor polynomial of $f(x)$ centered at 0 as

$$\begin{aligned} p_5(x) &= f(0) + f'(0)(x-0) + \frac{f''(0)}{2}(x-0)^2 + \frac{f'''(0)}{3!}(x-0)^3 + \frac{f^{(4)}(0)}{4!}(x-0)^4 + \frac{f^{(5)}(0)}{5!}(x-0)^5 \\ &= \boxed{x - \frac{1}{6}x^3 + \frac{1}{120}x^5}. \end{aligned}$$

Proposition 29. *The nth degree Taylor polynomial of $f(x) = \frac{1}{1-x}$ centered at 0 is*

$$\begin{aligned} p_n(x) &= 1 + x + x^2 + \cdots + x^n \\ &= \sum_{k=0}^{n} x^k. \end{aligned}$$

The nth degree Taylor polynomial of $f(x) = e^x$ centered at 0 is

$$\begin{aligned} p_n(x) &= 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} \\ &= \sum_{k=0}^{n} \frac{x^k}{k!}. \end{aligned}$$

The $2n$th degree Taylor polynomial of $f(x) = \cos(x)$ centered at 0 is

$$\begin{aligned} p_n(x) &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} \\ &= \sum_{k=0}^{n} (-1)^k \frac{x^{2k}}{(2k)!}. \end{aligned}$$

The $(2n-1)$th degree Taylor polynomial of $f(x) = \sin(x)$ centered at 0 is

$$\begin{aligned} p_n(x) &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^{n-1} \frac{x^{2n-1}}{(2n-1)!} \\ &= \sum_{k=0}^{n} (-1)^{k-1} \frac{x^{2k-1}}{(2k-1)!}. \end{aligned}$$

The nth degree Taylor polynomial of $f(x) = \ln(x+1)$ centered at 0 is

$$\begin{aligned} p_n(x) &= x - \frac{x^2}{2} + \frac{x^3}{3!} - \frac{x^4}{4} + \frac{x^5}{5} - \cdots + (-1)^{n-1}\frac{x^n}{n} \\ &= \sum_{k=0}^{n} (-1)^{k-1}\frac{x^k}{k}. \end{aligned}$$

The $(2n-1)$th degree Taylor polynomial of $f(x) = \arctan(x)$ centered at 0 is

$$\begin{aligned} p_n(x) &= x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \frac{x^9}{9} - \cdots + (-1)^{n-1}\frac{x^{2n-1}}{2n-1} \\ &= \sum_{k=0}^{n} (-1)^{k-1}\frac{x^{2k-1}}{2k-1}. \end{aligned}$$

Taylor polynomials of multivariable functions

We'll now turn our attention to multivariable functions. In order to define Taylor polynomials for multivariable functions, we need to be able to take higher order derivatives of multivariable functions. For the first derivative, we have the derivative matrix giving the total derivative of a multivariable function. However, we haven't defined any analogous "second order total derivative," much less higher order derivatives! We will need to do this before we can give any sort of definition of Taylor polynomials beyond degree one.

For now, we can think of the Taylor polynomial as giving the best polynomial approximation for a function, and use our knowledge about single variable Taylor polynomials in order to find Taylor polynomials of some special multivariable functions.

Example 90. *We will find the fifth degree Taylor polynomial of $f(x,y) = \sin(x+y)$ centered at the origin. Notice that if we let $u = x+y$, we can make use of the fifth degree Taylor polynomial of $g(u) = \sin(u)$, which we found to be*

$$p_5(u) = u - \frac{1}{6}u^3 + \frac{1}{120}u^5.$$

When we substitute $u = x+y$, we obtain

$$p_5(x,y) = (x+y) - \frac{1}{6}(x+y)^3 + \frac{1}{120}(x+y)^5.$$

This is a fifth degree polynomial in x and y, and it is the fifth degree Taylor polynomial of $f(x,y) = \sin(x+y)$.

Example 91. *We will find the fourth degree Taylor polynomial of $f(x,y) = e^{xy}$ centered at the origin. Letting $u = xy$, we'll use the fourth degree Taylor polynomial of e^u:*

$$p_4(u) = 1 + u + \frac{u^2}{2!} + \frac{u^3}{3!} + \frac{u^4}{4!}.$$

Substituting $u = xy$, we obtain

$$1 + xy + \frac{(xy)^2}{2!} + \frac{(xy)^3}{3!} + \frac{(xy)^4}{4!}.$$

However, this is not a fourth degree polynomial in x and y! Since $(xy)^4$ has total degree 8, we have terms of to large of a degree. Taking only terms of degree 4 or less, we find the fourth degree Taylor polynomial of $f(x, y) = e^{xy}$.

$$p_4(x, y) = 1 + xy + \frac{(xy)^2}{2!}$$

Quadratic Forms

In order to better understand the behavior of multivariable functions, we would like to define some sort of second derivative for multivariable functions. For first derivatives, we have the gradient and derivative matrix filling the roles of derivatives for scalar-valued and vector-valued functions, respectively. Before we define the second derivative, we will try to understand second-order behavior in multivariable functions specifically. That is, we'll consider polynomials in n-variables that only terms of degree 2, and determine how these polynomials behave.

Polynomials with the property are very important throughout mathematics, and they are called quadratic forms. You've frequently seen a quadratic form which arises from the length of a vector $\vec{x} = (x_1, ..., x_n)$,

$$\|\vec{x}\|^2 = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Quadratic Forms

Definition 62. *A* quadratic form *in $\mathbb{R}^n$ is a polynomial $p : \mathbb{R}^n \to \mathbb{R}$ in which each term has total degree 2. That is, it has the form*

$$\begin{aligned} p(x_1, ..., x_n) = c_{11}x_1^2 &+ c_{12}x_1x_2 + c_{13}x_1x_3 + \cdots + c_{1n}x_1x_n \\ &+ c_{22}x_2^2 \quad + c_{23}x_2x_3 + \cdots + c_{2n}x_2x_n \\ &\qquad\qquad + c_{33}x_3^2 \quad + \cdots + c_{3n}x_3x_n \\ &\qquad\qquad\qquad\qquad \ddots \quad \vdots \\ &\qquad\qquad\qquad\qquad\quad + c_{nn}x_n^2 \end{aligned}$$

Example 92. *For example, the polynomials $x^2 + 2xy + y^2$ and $xy + yz - xz$ are quadratic forms.*

On the other hand, $\sin(x^2)$ and $\dfrac{x^3}{y}$ are not quadratic forms, since they're not polynomials. Also, the polynomials $x^2 + y^2 - 1$ and $(x + 1)^2$ are not quadratic forms, since they include terms of degrees other than 2

Learning outcomes:
Author(s):

For any quadratic form

$$\begin{aligned} p(x_1, ..., x_n) = c_{11}x_1^2 &+ c_{12}x_1x_2 + c_{13}x_1x_3 + \cdots + c_{1n}x_1x_n \\ &+ c_{22}x_2^2 \quad + c_{23}x_2x_3 + \cdots + c_{2n}x_2x_n \\ &\qquad\qquad + c_{33}x_3^2 \quad + \cdots + c_{3n}x_3x_n \\ &\qquad\qquad\qquad\qquad \ddots \quad \vdots \\ &\qquad\qquad\qquad\qquad\quad + c_{nn}x_n^2, \end{aligned}$$

we can write

$$p(x_1, ..., x_n) = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \end{pmatrix} \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ 0 & c_{22} & \cdots & c_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & c_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Thus, we can represent the quadratic form with the matrix

$$\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ 0 & c_{22} & \cdots & c_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & c_{nn} \end{pmatrix}.$$

Example 93. *For example, the quadratic form corresponding to the matrix* $\begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}$ *is*

$$\begin{aligned} p(x, y) &= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} x + 2y \\ 3y \end{pmatrix} \qquad = x^2 + 2xy + 3y^2. \end{aligned}$$

The quadratic form corresponding to the matrix $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}$ *is*

$$\begin{aligned} p(x, y) &= \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= \boxed{x^2 + 4xy + 8xz + 5z^2 + 6yz + 7z^2}. \end{aligned}$$

The quadratic form corresponding to the matrix $\begin{pmatrix} 1 & 0 & 0 \\ 2 & 4 & 0 \\ 3 & 5 & 6 \end{pmatrix}$ *is*

$$\begin{aligned} p(x,y) &= \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 2 & 4 & 0 \\ 3 & 5 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= \boxed{x^2 + 4xy + 8xz + 5z^2 + 6yz + 7z^2}. \end{aligned}$$

Notice in the previous example, there were two different matrices that gave rise to the same quadratic form. In general, there will be many different matrices corresponding to the same quadratic form. However, if we add the condition that a matrix be *symmetric*, then we do have uniqueness.

Theorem 12. *For any quadratic form p, there is a unique symmetric matrix A such that*

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

In particular, the symmetric matrix

$$A = \begin{pmatrix} c_{11} & \frac{1}{2}c_{12} & \cdots & \frac{1}{2}c_{1n} \\ \frac{1}{2}c_{12} & c_{22} & \cdots & \frac{1}{2}c_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{1}{2}c_{1n} & \cdots & \frac{1}{2}c_{(n-1)n} & c_{nn} \end{pmatrix}$$

corresponds to the quadratic form

$$\begin{aligned} p(x_1, ..., x_n) = c_{11}x_1^2 &+ c_{12}x_1x_2 + c_{13}x_1x_3 + \cdots + c_{1n}x_1x_n \\ &+ c_{22}x_2^2 \quad + c_{23}x_2x_3 + \cdots + c_{2n}x_2x_n \\ &\qquad\qquad + c_{33}x_3^2 \quad + \cdots + c_{3n}x_3x_n \\ &\qquad\qquad\qquad\qquad \ddots \quad \vdots \\ &\qquad\qquad\qquad\qquad + c_{nn}x_n^2. \end{aligned}$$

Categorizing Quadratic Forms

We can categorize quadratic forms according their behavior. This behavior tells us about the shape of their graphs, and these observations will be important when we transition to studying the behavior of more general functions.

Definition 63. *A quadratic form $p(\vec{x})$ is...*

- positive definite *if $p(\vec{x}) > 0$ for all $\vec{x} \neq 0$.*

- negative definite *if* $p(\vec{x}) < 0$ *for all* $\vec{x} \neq 0$.
- positive semi-definite *if it is not positive definite, and* $p(\vec{x}) \geq 0$ *for all* $\vec{x}$*).*
- negative semi-definite *if it is not positive definite, and* $p(\vec{x}) \geq 0$ *for all* $\vec{x}$*).*
- indefinite *if it is none of the above. That is, there exist* $\vec{x}_1, \vec{x}_2 \neq \vec{0}$ *such that* $p(\vec{x}_1) < 0$ *and* $p(\vec{x}_2) > 0$.

Example 94. *The quadratic form* $p(x, y) = x^2 + y^2$ *is positive definite. Its graph is pictured below.*

PICTURE

The graphs of other positive definite quadratic forms look similar, though they may be stretched in various directions. Notice that for a positive definite quadratic form, there is always a strict minimum at the origin.

The quadratic form $p(x, y) = -x^2 - y^2$ *is negative definite. Its graph is pictured below.*

PICTURE

The graphs of other negative definite quadratic forms look similar, though they may be stretched in various directions. Notice that for a negative definite quadratic form, there is always a strict maximum at the origin.

The quadratic form $p(x, y) = x^2$ *is positive semi-definite. Its graph is pictured below.*

PICTURE

The graphs of other positive semi-definite quadratic forms look similar, though they may be stretched in various directions.

The quadratic form $p(x, y) = -x^2$ *is negative semi-definite. Its graph is pictured below.*

PICTURE

The graphs of other negative semi-definite quadratic forms look similar, though they may be stretched in various directions.

The quadratic form $p(x, y) = x^2 - y^2$ *is indefinite. Its graph is pictured below.*

PICTURE

The graphs of other indefinite quadratic forms look similar, though they may be stretched in various directions. Notice the behavior of the graph around the origin; because of its shape, this is called a saddle point.

Sylvester's Theorem

We can use the symmetric matrix representing a quadratic form to classify the quadratic form, by looking at a sequence of determinants.

Theorem 13. *(Sylvester's Theorem) Let $p(\vec{x}) = \vec{x}^T A\vec{x}$, where A is a symmetric matrix,*

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{12} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{13} & a_{23} & a_{33} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \cdots & a_{nn} \end{pmatrix}.$$

Consider the sequence of determinants

$$d_1 = a_{11}, d_2 = det\begin{pmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{pmatrix}, d_3 = det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix}, ..., d_n = det\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{12} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{13} & a_{23} & a_{33} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & a_{3n} & \cdots & a_{nn} \end{pmatrix}.$$

- *If all of the determinants are positive, so $d_i > 0$ for all i, then p is positive definite.*
- *If $d_1 < 0$, and the signed of the determinants alternate (so the odd numbered determinants are negative, and the even numbered determinants are positive), then p is negative definite.*
- *If all of the determinants are nonzero, but don't follow the above patterns, then p is indefinite.*

Notice that this theorem requires that A be a symmetric matrix. Also, in the case where one or more of the determinants is zero, we can't use this theorem to classify the quadratic form.

Example 95. *Consider the quadratic form $p(\vec{x}) = \vec{x}^T A\vec{x}$ for $A = \begin{pmatrix} -1 & 1 & -2 \\ 1 & -5 & 1 \\ -2 & 1 & -10 \end{pmatrix}$.*
Notice that A is symmetric, so we'll be able to use Sylvester's Theorem. Let's

find the sequence of determinants.

$$\begin{aligned} d_1 &= -1 \\ d_2 &= det \begin{pmatrix} -1 & 1 \\ 1 & -5 \end{pmatrix} \\ &= \boxed{3} \\ d_3 &= det \begin{pmatrix} -1 & 1 & -2 \\ 1 & -5 & 1 \\ -2 & 1 & -10 \end{pmatrix} \\ &= \boxed{-23} \end{aligned}$$

Based on this sequence of determinants, and using Sylvester's Theorem, we have

Multiple Choice:

(a) *p is positive definite.*

(b) *p is negative definite.* ✓

(c) *p is indefinite.*

(d) *We cannot use Sylvester's Theorem to categorize this quadratic form.*

Example 96. *Consider the quadratic form $p(\vec{x}) = \vec{x}^T A\vec{x}$ for $A = \begin{pmatrix} 1 & 11 \\ 11 & 121 \end{pmatrix}$. Notice that A is symmetric, so we'll be able to use Sylvester's Theorem. Let's find the sequence of determinants.*

$$\begin{aligned} d_1 &= \boxed{1} \\ d_2 &= det \begin{pmatrix} 1 & 11 \\ 11 & 121 \end{pmatrix} \\ &= \boxed{0} \end{aligned}$$

Based on this sequence of determinants, and using Sylvester's Theorem, we have

Multiple Choice:

(a) *p is positive definite.*

(b) *p is negative definite.*

(c) *p is indefinite.*

(d) *We cannot use Sylvester's Theorem to categorize this quadratic form.* ✓

Example 97. *Consider the quadratic form* $p(\vec{x}) = \vec{x}^T A\vec{x}$ *for* $A = \begin{pmatrix} -1 & 2 & 0 \\ 0 & -4 & 0 \\ -2 & 0 & 3 \end{pmatrix}$.
Notice that A is symmetric, so we won't immediately be able to use Sylvester's Theorem. First, we'll need to find the symmetric matrix representing p. Expanding p, we have

$$\begin{aligned} p(x,y,z) &= \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} -1 & 2 & 0 \\ 0 & -4 & 0 \\ -2 & 0 & 3 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ &= \boxed{-x^2 + 2xy - 4y^2 - 2xz + 3z^2}. \end{aligned}$$

From this, we can find the symmetric matrix representing p to be

$$\begin{pmatrix} -1 & 1 & -1 \\ 1 & -4 & 0 \\ -1 & 0 & 3 \end{pmatrix}.$$

We now find the sequence of determinants for the symmetric matrix above.

$$\begin{aligned} d_1 &= -1 \\ d_2 &= det \begin{pmatrix} -1 & 1 \\ 1 & -4 \end{pmatrix} \\ &= \boxed{3} \\ d_3 &= det \begin{pmatrix} -1 & 1 & -1 \\ 1 & -4 & 0 \\ -1 & 0 & 3 \end{pmatrix} \\ &= \boxed{-23} \end{aligned}$$

Based on this sequence of determinants, and using Sylvester's Theorem, we have

Multiple Choice:

(a) *p is positive definite.*

(b) *p is negative definite.*

(c) *p is indefinite.* ✓

(d) *We cannot use Sylvester's Theorem to categorize this quadratic form.*

The Hessian Matrix

We've been working towards defining some sort of "second derivative" for multivariable functions, which can tell us about the second-order behavior of functions. It will also enable us to define degree two Taylor polynomials, and we'll later see how it can be used to classify critical points in optimization.

The Hessian Matrix

Suppose we have a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$. We can take the gradient of this function.

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \frac{\partial f}{\partial x_3}, ..., \frac{\partial f}{\partial x_n} \right)$$

We can think of the gradient as a function $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$. Assuming all of the partial derivatives exist, we can then take the derivative matrix of ∇f. This gives us a square $n \times n$ matrix, which we call the Hessian of f.

Definition 64. *The* Hessian Matrix *of $f : \mathbb{R}^n \to \mathbb{R}$ is the $n \times n$ matrix*

$$Hf = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 x_n} & \frac{\partial^2 f}{\partial x_2 x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}.$$

Notice that if f has continuous first and second order partial derivatives, then the Hessian matrix will be symmetric by Clairaut's Theorem.

Example 98. *Consider the function $f(x, y) = x + 2xy + 3y^3$. We'll compute the Hessian of f. First, we find the gradient of f.*

$$\nabla f = \boxed{(1 + 2y, 2x + 9y^2)}$$

Taking the derivative matrix of the gradient, we obtain the Hessian of f.

$$Hf = \begin{pmatrix} \boxed{0} & \boxed{2} \\ \boxed{2} & \boxed{18y} \end{pmatrix}$$

Learning outcomes:
Author(s):

Consider the function $g(x, y, z) = e^{xyz}$. We'll compute the Hessian of g. First, we find the gradient of g.

$$\nabla g = \boxed{(yze^{xyz}, xze^{xyz}, xye^{xyz})}$$

Taking the derivative matrix of the gradient we obtain the Hessian of f.

$$Hf = \begin{pmatrix} \boxed{y^2z^2e^{xyz}} & \boxed{z(xyz+1)e^{xyz}} & \boxed{y(xyz+1)e^{xyz}} \\ \boxed{z(xyz+1)e^{xyz}} & \boxed{x^2z^2e^{xyz}} & \boxed{x(xyz+1)e^{xyz}} \\ \boxed{y(xyz+1)e^{xyz}} & \boxed{x(xyz+1)e^{xyz}} & \boxed{x^2y^2e^{xyz}} \end{pmatrix}$$

Taylor Polynomials

We're now in position to define the second-order Taylor polynomial of a function $f : \mathbb{R}^n \to \mathbb{R}$, using the Hessian matrix to find the degree two terms.

Definition 65. *Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. The degree two Taylor polynomial of f centered at $\vec{a}$ is*

$$p(\vec{x}) = f(\vec{a}) + \nabla f(\vec{a}) \cdot (\vec{x} - \vec{a}) + \frac{1}{2}(\vec{x} - \vec{a})^T Hf(\vec{a})(\vec{x} - \vec{a}).$$

Example 99. *We'll compute the degree two Taylor polynomial of the function $f(x, y) = x^2 + 2xy + 3y^3$ centered at $(2, 1)$. We previously found that the gradient and Hessian of this function were*

$$\nabla f(x, y) = \boxed{(1 + 2y, 2x + 9y^2)},$$
$$Hf(x, y) = \begin{pmatrix} \boxed{0} & \boxed{2} \\ \boxed{2} & \boxed{18y} \end{pmatrix}.$$

Plugging in the point $(2, 1)$, we have

$$\nabla f(2, 1) = \boxed{(3, 13)},$$
$$Hf(x, y) = \begin{pmatrix} \boxed{0} & \boxed{2} \\ \boxed{2} & \boxed{18} \end{pmatrix}.$$

Then, the degree 2 Taylor polynomial of f centered at $(2, 1)$ is

$$\begin{aligned} p_2(x, y) &= f(2, 1) + \nabla f(2, 1) \cdot (x - 2, y - 1) + \frac{1}{2}(x - 2, y - 1)^T Hf(2, 1)(x - 2, y - 1) \\ &= 11 + (3, 13) \cdot (x - 2, y - 1) + \frac{1}{2}\begin{pmatrix} x - 2 & y - 1 \end{pmatrix} \begin{pmatrix} 0 & 2 \\ 2 & 18 \end{pmatrix} \begin{pmatrix} x - 2 \\ y - 1 \end{pmatrix} \qquad = 11 + 3(x - 2) + 13(y \\ &= 5 + x - 9y + 2xy + 9y^2. \end{aligned}$$

Notice that the order two part of the Taylor polynomial,

$$\frac{1}{2}(\vec{x}-\vec{a})^T Hf(\vec{a})(\vec{x}-\vec{a}),$$

will be a quadratic form when $\vec{a}=\vec{0}$. We will soon make use of our classification of quadratic forms in order to use the Hessian matrix to determine the order two behavior of a function, which will be useful for optimizing multivariable functions.

Local Extrema

Now that we've defined the Hessian matrix and studied quadratic forms, we're in position to find local extrema of multivariable function. This process will closely resemble the process that we following in single variable calculus, so we'll begin by reviewing that case.

Recall that we say $f(x)$ has a local maximum at $x = a$ if $f(x) \leq f(a)$ for all x near a. Similarly, $f(x)$ has a local minimum at $x = a$ if $f(a) \leq f(x)$ for all x near a. We can state these definitions more precisely, as below.

Definition 66. *We say that $f(x)$ has a* local maximum *at $x = a$ if there exists $r > 0$ such that $|x - a| < r$ implies $f(x) \leq f(a)$.*

We say that $f(x)$ has a local minimum *at $x = a$ if there exists $r > 0$ such that $|x - a| < r$ implies $f(x) \geq f(a)$.*

In single variable calculus, we found local extrema by finding critical points, and then classifying them using the first or second derivative test.

Definition 67. *The point $x = a$ is a* critical point *of $f(x)$ if $f'(a) = 0$, or if $f'(a)$ does not exist.*

Proposition 30. *If $f(x)$ has a local minimum or maximum at $x = a$, then a is a critical point of $f(x)$.*

If a is a critical point of $f(x)$ and $f''(a) < 0$, then $f(x)$ has a local maximum at $x = a$.

If a is a critical point of $f(x)$ and $f''(a) > 0$, then $f(x)$ has a local minimum at $x = a$.

If a is a critical point of $f(x)$ and $f''(a) = 0$ or $f''(a)$ does not exist, then $f(x)$ could have a local maximum at a, a local minimum at a, or neither.

Example 100. *As an example, we'll find the local extrema of the function $f(x) = x^3 + 3x^2 - 9x + 1$. Differentiating, we have*

$$f'(x) = \boxed{3x^2 + 6x - 9}.$$

Solving $f'(x) = 0$ for x, we see that there are two critical points, $x = -3$ and $x = 1$.

We'll classify these critical points using the second derivative test. We find the second derivative of $f(x)$.

$$f''(x) = \boxed{6x + 6}$$

Learning outcomes:
Author(s):

Plugging in $x = -3$, we have

$$f''(-3) = \boxed{-12}.$$

This tells us that at $x = -3$, $f(x)$ has a

Multiple Choice:

(a) *local maximum.* ✓

(b) *local minimum.*

(c) *neither.*

Plugging in $x = 1$, we have

$$f''(1) = \boxed{12}.$$

This tells us that at $x = 1$, $f(x)$ has a

Multiple Choice:

(a) *local maximum.*

(b) *local minimum.* ✓

(c) *neither.*

Local extrema for multivariable functions

We begin by defining local minima and local maxima for multivariable functions. These follow the same idea as in the single variable case. For example, f has a local minimum at $\vec{x} = \vec{a}$ if $f(\vec{a}) \leq f(\vec{x})$ for $\vec{x}$ "near" a. Now, we need to decide what "near" means. This means that we can find some distance $r > 0$ such that for $\vec{x}$ within distance r from $\vec{a}$, we must have $f(\vec{a}) \leq f(\vec{x})$. We restate this more formally.

Definition 68. *We say that $f : \mathbb{R}^n \to \mathbb{R}$ has a* local minimum *at $\vec{x} = \vec{a}$ if there exists $r > 0$ such that for all $\vec{x}$ with $\|\vec{x} - \vec{a}\| < r$, we have $f(\vec{a}) \leq f(\vec{x})$.*

We say that $f : \mathbb{R}^n \to \mathbb{R}$ has a local maximum *at $\vec{x} = \vec{a}$ if there exists $r > 0$ such that for all $\vec{x}$ with $\|\vec{x} - \vec{a}\| < r$, we have $f(\vec{a}) \geq f(\vec{x})$.*

In some cases, we can determine local extrema using our knowledge about functions, and without using Calculus.

Example 101. *Consider the function $f(x, y) = (x-1)^2 + (y+2)^2$. Since the terms $(x-1)^2$ and $(y+2)^2$ are always nonnegative, we can see that $f(x, y)$ will always be nonnegative. Since $f(1, -2) = 0$, and $0 \leq f(x, y)$ for all (x, y), this* *means that $f(x, y)$ has a local minimum at $(1, -2)$. (This local minimum is also an absolute maximum.)*

Consider the function $g(x, y, z) = 1 - \sqrt{x^2 + y^2 + z^2}$. Since $\sqrt{x^2 + y^2 + z^2} \geq 0$, we'll always have $g(x, y, z) \leq 1$. Then, since $g(0, 0, 0) = 1$, g has a local maximum at $(0, 0, 0)$.

Critical points

Our definition of critical points of multivariable functions is also very similar to the definition from single variable calculus. Here, the gradient vector replaces the derivative.

Definition 69. *We say that $f : \mathbb{R}^n \to \mathbb{R}$ has a* critical point *at $\vec{x} = \vec{a}$ if $\nabla f(\vec{a}) = \vec{0}$, or $\nabla f(\vec{a})$ does not exist.*

For functions $f : \mathbb{R}^2 \to \mathbb{R}$, we can visualize critical points as points where the tangent plane is either horizontal, or undefined.

PICTURES

As in single variable calculus, if a function has local extrema, they will occur at critical points.

Proposition 31. *If $f : \mathbb{R}^n \to \mathbb{R}$ has a local maximum or local minimum at $\vec{x} = \vec{a}$, then $\vec{a}$ is a critical point of f.*

Example 102. *We'll find the critical points of the function $f(x, y) = x^3 + x^2y - y^2 - 4y$. The gradient of f is*

$$\nabla f(x, y) = (3x^2 + 2xy, x^2 - 2y - 4).$$

This is defined at all points in $\mathbb{R}^2$, so the critical points will satisfy $\nabla f(x, y) = (0, 0)$. In order to find the critical points, we solve the system of equations

$$\begin{cases} 3x^2 + 2xy = 0 \\ x^2 - 2y - 4 = 0 \end{cases}.$$

Factoring the first equation, we have $x(3x + 2y) = 0$, giving us the cases $x = 0$ or $3x + 2y = 0$.

If $x = 0$, the second equation gives us $y = -2$. So $(0, -2)$ is a critical point.

If $3x+2y = 0$, plugging $y = \dfrac{-3}{2}x$ into the second equation gives us $x^2+3x-4 = 0$, so $x = 1$ or $x = -4$. This gives us the critical points $(1, -3/2)$ and $(-4, 6)$.

Thus, the critical points of f are $(0, -2)$, $(1, -3/2)$, and $(-4, 6)$.

Classifying critical points

We can use the Hessian matrix to classify critical points in some cases.

Proposition 32. *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ has continuous second order partial derivatives (so f has $\mathcal{C}^2$) at and near a critical point $\vec{a} \in U$.*

- *If $Hf(\vec{a})$ is positive definite, then f has a* local minimum *at $\vec{x} = \vec{a}$.*
- *If $Hf(\vec{a})$ is negative definite, then f has a* local maximum *at $\vec{x} = \vec{a}$.*
- *If $Hf(\vec{a})$ is indefinite, then f has a saddle point at $\vec{x} = \vec{a}$.*

Example 103. *Consider again the function $f(x, y) = x^3+x^2y-y^2-4y$. Earlier, we found the critical points of this function, which are $(0, -2)$, $(1, -3/2)$, and $(-4, 6)$. In order to classify these critical points, we find the Hessian matrix of f.*

$$Hf(x, y) = \begin{pmatrix} \boxed{6x + 2y} & \boxed{2x} \\ \boxed{2x} & \boxed{-2} \end{pmatrix}$$

Plugging in the point $(0, -2)$, we have

$$Hf(0, -2) = \begin{pmatrix} \boxed{-4} & \boxed{0} \\ \boxed{0} & \boxed{-2} \end{pmatrix}.$$

Using Sylvester's Theorem, we can see that $Hf(0, -2)$ is

Multiple Choice:

(a) *positive definite.*

(b) *negative definite.* ✓

(c) *indefinite.*

(d) *we cannot determine this using Sylvester's Theorem.*

This means that, at $(0, -2)$, $f(x, y)$ has a

Multiple Choice:

(a) *local maximum.* ✓

(b) *local minimum.*

(c) *saddle point.*

(d) *none of the above.*

Next, we'll classify the critical point $(1, -3/2)$*. We find the Hessian matrix at this point,*

$$Hf(1, -3/2) = \begin{pmatrix} \boxed{3} & \boxed{2} \\ \boxed{2} & \boxed{-2} \end{pmatrix}.$$

Using Sylvester's Theorem, we can see that $Hf(1, -3/2)$ *is*

Multiple Choice:

(a) *positive definite.*

(b) *negative definite.*

(c) *indefinite.* ✓

(d) *we cannot determine this using Sylvester's Theorem.*

This means that, at $(1, -3/2)$*,* $f(x, y)$ *has a*

Multiple Choice:

(a) *local maximum.*

(b) *local minimum.*

(c) *saddle point.* ✓

(d) *none of the above.*

Finally, we'll classify the critical point $(-4, 6)$*. We find the Hessian matrix at this point,*

$$Hf(-4, 6) = \begin{pmatrix} \boxed{-12} & \boxed{-8} \\ \boxed{-8} & \boxed{-2} \end{pmatrix}.$$

Using Sylvester's Theorem, we can see that $Hf(-4, 6)$ *is*

Multiple Choice:

(a) *positive definite.*

(b) *negative definite.*

(c) *indefinite.* ✓

(d) *we cannot determine this using Sylvester's Theorem.*

This means that, at $(-4, 6)$*,* $f(x, y)$ *has a*

Multiple Choice:

(a) *local maximum.*

(b) *local minimum.*

(c) *saddle point.* ✓

(d) *none of the above.*

Absolute Extrema

We'll now turn our attention to absolute extrema, also called global extrema. We'll begin by reviewing the situation in single variable calculus, where we optimized over closed intervals. This is sometimes called the "closed interval method."

Definition 70. *Let S be a subset of $\mathbb{R}$, and let $f : S \to \mathbb{R}$. We say that f has an* absolute maximum *over S at $x = c$ if $f(x) \leq f(c)$ for all $x \in S$. If this is the case, then we say that $f(c)$ is the absolute maximum of f over S.*

Let S be a subset of $\mathbb{R}$, and let $f : S \to \mathbb{R}$. We say that f has an absolute minimum *over S at $x = c$ if $f(x) \geq f(c)$ for all $x \in S$. If this is the case, then we say that $f(c)$ is the absolute minimum of f over S.*

In general, a function is not guaranteed to have an absolute maximum or minimum.

Usually, we took f to be a continuous function and S to be a closed interval. We did this because, in this case, f is guaranteed to have an absolute maximum and absolute minimum, and these will either happen at critical points or at the endpoints of S. This is true by the Extreme Value Theorem.

Theorem 14. *Let $I \subset \mathbb{R}$ be a closed interval, and let $f : I \to \mathbb{R}$ be a continuous function. Then f has an absolute maximum and absolute minimum over I. Furthermore, if f has an absolute maximum or absolute minimum over I at $x = c$, then either c is a critical point of f, or c is an endpoint of the interval I.*

Using this theorem, we can find the absolute maximum and absolute minimum of a continuous function f over a closed interval $I = [a, b]$ by following the steps below.

(a) Find the critical points of f in the interval I, and find the value of f at each of these points.

(b) Find $f(a)$ and $f(b)$.

(c) Among the values of f at the critical points and at the endpoints, the largest value is the absolute maximum, and the smallest value is the absolute minimum.

We will be able to find an analogous theorem and process for finding the absolute maximum and absolute minimum of a multivariable function, but this raises

Learning outcomes:
Author(s):

the immediate question: in $\mathbb{R}^n$, what is the analogue of a closed interval? This brings us to compact sets.

Compact sets

In order to formulate a version of the Extreme Value Theorem for multivariable functions, we need to consider functions over compact sets.

Definition 71. *A subset $X \subset \mathbb{R}^n$ is* closed *if it includes its boundary. Roughly speaking, this means that the "edges" of X are solid lines, not dashed lines.*

A subset $X \subset \mathbb{R}^n$ is bounded *if there exists some R such that that $X \subset B_R(\vec{0}) = \{\vec{x} : \|\vec{x}\| < R\}$. Roughly speaking, this means that X doesn't stretch to infinity in any direction, so we can zoom out far enough that we can see all of X.*

A subset $X \subset \mathbb{R}^n$ is compact *if it is both closed and bounded.*

Example 104. *The set $\{(x, y) : x^2 + y^2 < 1\} \subset \mathbb{R}^2$ is bounded, but not closed, since it does not contain its boundary. Hence it is not compact.*

PICTURE

The set $\{(x, y) : x^2 + y^2 \leq 1\} \subset \mathbb{R}^2$ is both closed and bounded, hence compact.

PICTURE

The set $\{(x, y) : x^2 + y^2 \geq 1\} \subset \mathbb{R}^2$ is closed but not bounded. Hence it is not compact.

PICTURE

Problem 15 *Sketch each region, and determine if it is closed, bounded, and/or compact.*

$\{(x, y)\} : -1 \leq x \leq 1\}$

Select All Correct Answers:

(a) *closed* ✓

(b) *bounded*

(c) *compact*

$\{(x, y)\} : -1 \leq y \leq 1\}$

Select All Correct Answers:

(a) *closed* ✓

(b) *bounded*

(c) *compact*

$\{(x,y)\} \; : \; -1 \leq x \leq 1 \text{ and } -1 \leq y \leq 1\}$

Select All Correct Answers:

(a) *closed* ✓

(b) *bounded* ✓

(c) *compact* ✓

$\{(x,y)\} \; : \; x^2 - 1 \leq y \leq 1 - x^2\}$

Select All Correct Answers:

(a) *closed* ✓

(b) *bounded* ✓

(c) *compact* ✓

$\{(x,y)\} \; : \; -x^2 \leq y \leq x^2\}$

Select All Correct Answers:

(a) *closed* ✓

(b) *bounded*

(c) *compact*

Extreme Value Theorem

As in the single variable case, as long as we have a continuous function over a compact region, there is guaranteed to be an absolute maximum and absolute minimum. Furthermore, these will always occur either at critical points, or on the boundary.

Theorem 15. *(Extreme Value Theorem) Let $f : X \to \mathbb{R}$ be a continuous function on a compact region $X \subset \mathbb{R}^n$. Then f has both an absolute maximum and an absolute minimum over X.*

Furthermore, if f has an absolute maximum or absolute minimum at $\vec{a}$, then either $\vec{a}$ is a critical point of f in X, or $\vec{a}$ is on the boundary of X.

Because of this theorem, we can follow the steps below to optimize a continuous function f on a compact region X.

(a) Find the critical points of f in X, and find the value of f at each of these points.

(b) Find the absolute maximum and absolute minimum of f on the boundary of X.

(c) Of the values from steps 1 and 2, the largest value is the absolute maximum of f over X, and the smallest value is the absolute minimum of f over X.

Example 105. *We'll find the absolute maximum and minimum values of* $f(x, y) = x^2 - xy + y^2$ *on* $D = \{(x, y) \ : \ x^2 + y^2 \leq 1\} \subset \mathbb{R}^2$.

GRAPH D

First, note that f is continuous, and D is a compact region, so f has an absolute maximum and an absolute minimum over D, and each occurs either at a critical point or on the boundary of D.

We first find the critical points of f, by finding where the gradient is $\vec{0}$.

$$\nabla f(x, y) = \boxed{(2x - y, -x + 2y)}$$

Solving for where $\nabla f(x, y) = (0, 0)$, we obtain the only critical point,

$$(x, y) = \boxed{(0, 0)}.$$

The value of f at this critical point is $\boxed{0}$.

Next, we need to find the maximum and minimum values of f on the boundary of X, which is the unit circle. In order to do this, we parametrize the unit circle,

$$\vec{x}(t) = (\cos(t), \sin(t)),$$

for $t \in [0, 2\pi]$. Substituting this into f, we can find the maximum and minimum of f on the boundary of X by finding the maximum and minimum of the single variable function

$$\begin{aligned} g(t) &= \cos^2 t - \cos(t)\sin(t) + \sin^2 t \\ &= 1 - \cos(t)\sin(t) \end{aligned}$$

over $[0, 2\pi]$. To optimize this function, we follow the closed interval method for single variable functions. Differentiating g, we have

$$\begin{aligned} g'(t) &= \sin^2(t) - \cos^2(t) \\ &= -\cos(2t). \end{aligned}$$

Solving $g'(t) = 0$ *for* t*, the critical points in the interval* $[0, 2\pi]$ *are* $t = \frac{\pi}{4}$, $t = \frac{3\pi}{4}$, $t = \frac{5\pi}{4}$, *and* $t = \frac{7\pi}{4}$ *. The values of* g *at these points are*

$$\begin{aligned} g\left(\frac{\pi}{4}\right) &= \frac{1}{2}, \\ g\left(\frac{3\pi}{4}\right) &= \frac{3}{2}, \\ g\left(\frac{5\pi}{4}\right) &= \frac{1}{2}, \\ g\left(\frac{7\pi}{4}\right) &= \frac{3}{2}. \end{aligned}$$

The values of g *at the endpoints,* $t = 0$ *and* $t = 2\pi$*, are*

$$\begin{aligned} g(0) &= 1, \\ g(2\pi) &= 1. \end{aligned}$$

Comparing all of these values, we see that the absolute maximum of g *is* $\frac{3}{2}$*, and this occurs when* $t = \frac{3\pi}{4}$ *and when* $t = \frac{7\pi}{4}$*. These correspond to the points* $\left(\frac{-\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ *and* $\left(\frac{\sqrt{2}}{2}, \frac{-\sqrt{2}}{2}\right)$ *on the boundary of* D*. The absolute minimum of* g *is* $\frac{1}{2}$*, and this occurs when* $t = \frac{\pi}{4}$ *and when* $t = \frac{5\pi}{4}$*. These correspond to the points* $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ *and* $\left(\frac{-\sqrt{2}}{2}, \frac{-\sqrt{2}}{2}\right)$ *on the boundary of* D*.*

Now, considering the values of f *at the critical points and on the boundary of* X*, we see that the absolute maximum of* f *is* $\frac{3}{2}$ *and this occurs at* $\left(\frac{-\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ *and* $\left(\frac{\sqrt{2}}{2}, \frac{-\sqrt{2}}{2}\right)$*. The absolute minimum of* f *is* 0 *and this occurs at* $(0, 0)$*.*

GRAPH f OVER D

Optimization with Constraints

In this section, we will study optimization subject to constraints. For example, suppose we would like the find the maximum value of $f(x, y) = xy$ subject to the constraint $2x + 2y = 100$. We can think of this as restricting ourselves to the line $2x + 2y = 100$, and finding the maximum value of f over this line.

PICTURE

These types of problems often arise in real-world applications: you have some limitation on resources, which provide the constraint, and want to optimize some sort of performance. A classic problem from geometry and single variable calculus is to maximize the area of a rectangle, subject to the constraint that the perimeter is $100m$. Setting up the problem algebraically, this is equivalent to finding the maximum value of $f(x, y) = xy$ subject to the constraint $2x + 2y = 100$. In single variable calculus, we could solve for y in the second equation, and use this to reduce the objective function, $f(x, y)$, to a single variable function, which we would them optimize. That is, from $y = 50 - x$, we would have

$$\begin{aligned} g(x) &= f(x, 50 - x) \\ &= x(50 - x) \\ &= 50x - x^2. \end{aligned}$$

Differentiating g, we get

$$g'(x) = 50 - 2x.$$

So, we have $g'(x) = 0$ when $x = 25$. The second derivative of g is $g''(x) = -2 < 0$, so g has a local maximum at $x = 25$. Since g is continuous on all of $\mathbb{R}$ and has only one critical point, this means that g has an absolute maximum at $x = 25$. So, the maximum value of f is

$$f(25, 50 - 25) = g(25) = 15625.$$

For optimization subject to a constraint for multivariable functions in general, we'll follow a similar process: use the constraint to make a substitution, and then optimize the resulting function. However, we'll soon see that this can be much more complicated than it was in the single variable case, and it requires paying careful attention to constraints and domains.

Learning outcomes:
Author(s):

Optimization with Constraints

Let's consider a higher dimensional version of the problem from the introduction. That is, let's try to find the maximum value of $f(x, y, z) = xyz$ subject to the constraint that $x+y+z = 30$. Taking the constraint $x+y+z = 30$ and solving for z, we have $z = 30 - x - y$. Using substitution, we can reduce the function f to a function of two variables,

$$\begin{aligned} g(x, y) &= f(x, y, 30 - x - y) \\ &= xy(30 - x - y) \\ &= 30xy - x^2y - xy^2. \end{aligned}$$

So, we'd like to optimize g. In order to do this, we'll find the critical points of g. For this, we first find the gradient of g.

$$\nabla g(x, y) = (30y - 2xy - y^2, 30x - 2xy - x^2)$$

Solving $(30y - 2xy - y^2, 30x - 2xy - x^2) = (0, 0)$, we find two critical points: $(0, 0)$ and $(10, 10)$. Let's use the Hessian matrix of g to determine the behavior of g at these critical points.

$$Hg(x, y) = \begin{pmatrix} -2y & 30 - 2x - 2y \\ 30 - 2x - 2y & -2x \end{pmatrix}$$

At $(0, 0)$, the Hessian matrix is

$$Hg(x, y) = \begin{pmatrix} 0 & 30 \\ 30 & 0 \end{pmatrix}.$$

This corresponds to the quadratic form $p(x, y) = 60xy$, which is indefinite. So, g has a saddle point at $(0, 0)$.

At $(10, 10)$, the Hessian matrix is

$$Hg(x, y) = \begin{pmatrix} -20 & -10 \\ -10 & -20 \end{pmatrix}.$$

Using Sylvester's Theorem, this matrix is negative definite. So, g has a local maximum at $(10, 10)$. But does this mean that g has an absolute maximum at $(10, 10)$? Let's look at the graph of g to investigate.

PICTURE

We see that g actually doesn't have any absolute maximum (or minimum) - it's values get arbitrarily large, and arbitrarily small. So, here we see an example where the multivariable situation is more nuanced than the single variable situation. Often, it will be difficult to determine if multivariable functions have absolute extrema using the methods that we've covered so far. However, in the case where we're optimizing a continuous function over a compact region, we know that absolute extrema exist. Because of this fact, we will focus most of our attention on this special situation.

With that in mind, let's consider a revised version of our previous problem.

Example 106. *We'll find the maximum and minimum values of $f(x, y, z) = xyz$ subject to the constraints that $x + y + z = 30$, and x, y, and z must be* nonnegative.

If we consider the set of (x, y, z) such that $x + y + z = 30$, and $x, y, z \geq 0$, we see that this is a triangle in $\mathbb{R}^3$, and that it is a compact region. Since f is a continuous function, the Extreme Value Theorem guarantees that it will achieve and absolute maximum and minimum. We could optimize this function using our earlier methods for compact regions, but we'll instead optimize using substitution.

Substituting $z = 30 - x - y$, we reduce f to a function of two variables,

$$\begin{aligned} g(x, y) &= f(x, y, 30 - x - y) \\ &= xy(30 - x - y) \\ &= 30xy - x^2y - xy^2. \end{aligned}$$

Keeping our previous constraints in mind, we want to optimize g over the region $X = \{(x, y) \ : \ x \geq 0, y \geq 0, \text{ and } 30 - x - y \geq 0\}$. Graphing this region in $\mathbb{R}^2$, we see that X is a compact region, hence has an absolute maximum and absolute minimum.

PICTURE

We find the gradient of g to find the critical points in X,

$$\nabla g(x, y) = (30y - 2xy - y^2, 30x - 2xy - x^2),$$

and solving $(30y - 2xy - y^2, 30x - 2xy - x^2) = (0, 0)$, we find two critical points: $(0, 0)$ and $(10, 10)$. Evaluating g at these critical points, we find

$$\begin{aligned} g(0, 0) &= 0, \\ g(10, 10) &= 1000. \end{aligned}$$

Now, let's examine g on the boundary of X. The boundary of X consists of three line segments, along which we have $x = 0$, $y = 0$, or $30 - x - y = 0$. In each of these cases, $g(x, y) = 0$, so g is zero along the entire boundary of X.

Comparing the values that we've found, we see that g has an absolute maximum of 1000 at $(10, 10)$, and an absolute minimum of 0 which occurs along the line segments $x = 0$, $y = 0$, and $30 - x - y = 0$.

Translating this back to our original function f, the absolute maximum of f subject to the constraint $x + y + z = 30$ for nonnegative x, y, z is 1000, which occurs at the point $(10, 10, 10)$. The absolute minimum is 0, which occurs when $x = 0$, $y = 0$, or $z = 0$.

We'll now look at another example where optimizing a multivariable function subject to a constraint using substitution is deceptively difficult.

Example 107. *We'll find the absolute maximum and absolute minimum of* $f(x,y) = x^2 - y^2$ *subject to the constraint* $x^2 + y^2 = 1$.

PICTURE

Since $\{(x,y) \,:\, x^2 + y^2 = 1\}$ *is a compact region (it's the unit circle), and* f *is a continuous function, the Extreme Value Theorem tells us that* f *will have an absolute maximum and minimum.*

Let's rewrite the constraint as $x^2 = 1 - y^2$*, and substitute this into* f *to reduce* f *to a function of a single variable:*

$$\begin{aligned} g(y) &= (1 - y^2) - y^2 \\ &= 1 - 2y^2. \end{aligned}$$

Differentiating g*, we obtain*

$$g'(y) = -4y.$$

This has one critical point, $y = 0$*. Taking the second derivative of* g*, we have*

$$g''(y) = -4,$$

so g *has a local maximum at* $y = 0$*. Since* g *only has one critical point and it's a local maximum,* g *has an absolute maximum at* $y = 0$*. Furthermore,* g *has no absolute minimum.*

However, we said that f *must have an absolute minimum, so something must be wrong! Here, we've lost track of some of the information contained in our constraint. Consider again*

$$x^2 = 1 - y^2.$$

Since $x^2 \geq 0$ *for all* x*, we must have* $-1 \leq y \leq 1$*. So, we should really be optimizing* g *over the closed interval* $[-1, 1]$*. We've already found the critical point,* $y = 0$*, of* g*. Comparing the value of* g *at this points and the endpoints of the interval, we have*

$$\begin{aligned} g(0) &= 1, \\ g(-1) &= -1, \\ g(1) &= -1. \end{aligned}$$

So, we see that g *has an absolute maximum of* 1 *at* $y = 0$*, and an absolute minimum of* -1 *at* $y = -1$ *and* $y = 1$.

Translating this back to our original function, f *has an absolute maximum of* 1 *at* $(\pm 1, 0)$*, and an absolute minimum of* -1 *at* $(0, \pm 1)$.

We can see this reflected in the graph of f *over* $x^2 + y^2 = 1$.

PICTURE

The moral of this section is: weird things can happen with multivariable functions, and you need to be very careful with constraints.

Lagrange Multipliers

We've seen how we can optimize a function subject to a constraint using substitution, and we've seen that it can be very difficult to correctly handle the constraints! Fortunately, there is a tool that we can use to simplify this process. This tool is called "Lagrange multipliers."

Gradients

Suppose we wish to optimize a function $f : \mathbb{R}^n \to \mathbb{R}$ subject to some constraint $g(\vec{x}) = C$, where C is a constant, and $g : \mathbb{R}^n \to \mathbb{R}$ is a function. For now, we'll focus on the case $n = 2$, so we have a function $f : \mathbb{R}^2 \to \mathbb{R}$ and a constraint $g(x, y) = C$. The graph of f will be a surface in $\mathbb{R}^3$, and the graph of $g(x, y) = C$ is a curve in $\mathbb{R}^2$.

SOME PICTURES SOMEWHERE IN HERE, REDO THIS IN TERMS OF CRITICAL POINTS?

Now, suppose that the maximum value of $f(x, y)$ subject to the constraint $g(x, y) = C$ occurs at some point (a, b). Suppose we parametrize the curve $g(x, y) = C$ as $\vec{x}(t)$, with $(a, b) = \vec{x}(t_0)$. We can view $g(x, y) = C$ as a level curve of the function $g(x, y)$, and then the gradient of $g(x, y)$ will always be perpendicular to the curve $g(x, y) = C$. More precisely, for any point $(x, y) = \vec{x}(t)$ on the curve $g(x, y) = C$, we'll have $\nabla g(x, y)$ is perpendicular to $\vec{x}'(t)$. That is, $\nabla g(\vec{x}(t)) \perp \vec{x}'(t)$ for all t.

Next, let's turn our attention back to f. If f has an absolute maximum subject to $g(x, y) = C$, then $f(\vec{x}(t))$ has an absolute maximum at $t = t_0$. This means that $f(\vec{x}(t))$ has a critical point at $t = t_0$, so $\frac{d}{dt} f(\vec{x}(t))|_{t=t_0} = 0$. Using the chain rule, we can rewrite this as

$$\nabla f(\vec{x}(t_0)) \cdot \vec{x}'(t) = 0.$$

So, both $\nabla f(a, b)$ and $\nabla g(a, b)$ are perpendicular to $\vec{x}'(t)$. Since we are considering vectors in $\mathbb{R}^2$, this means that $\nabla f(a, b)$ and $\nabla g(a, b)$ are parallel, so we can write

$$\nabla f(a, b) = \lambda \nabla g(a, b)$$

for some constant λ.

Learning outcomes:
Author(s):

So, we can find candidate points for the absolute maximum (and similarly, the absolute minimum) of $f(x, y)$ subject to the constraint $g(x, y) = C$ by finding points where

$$\nabla f(a, b) = \lambda \nabla g(a, b).$$

This observation generalizes to $\mathbb{R}^n$.

Proposition 33. *Consider $\mathcal{C}^1$ functions $f, g : \mathbb{R}^n \to \mathbb{R}$, and let C be a constant. If f has an absolute maximum or absolute minimum at $\vec{a}$ subject to the constraint $g(\vec{x}) = C$, then there exists some scalar λ such that*

$$\nabla f(\vec{a}) = \lambda \nabla g(\vec{a}).$$

The constant λ is called a Lagrange Multiplier*.*

We can leverage this theorem into a method for finding absolute extrema of a function subject to a constraint, which we call the *method of Lagrange multipliers.*

To find the absolute extrema of a function $f(\vec{x})$ subject to a constraint $g(\vec{x}) = C$:

(a) Compute the gradients $\nabla f(\vec{x})$ and $\nabla g(\vec{x})$.

(b) Solve the system of equations

$$\begin{cases} \nabla f(\vec{x}) = \lambda \nabla g(\vec{x}) \\ g(\vec{x}) = C \end{cases}$$

for $\vec{x}$ and λ.

(c) The solutions to the system of equations in (2) are the critical points of $f(\vec{x})$ subject to $g(\vec{x}) = C$. Classify these critical points in order to determine the absolute extrema.

If $g(\vec{x}) = C$ is compact, we can determine the absolute extrema by comparing the values of f at the critical points. In this case, absolute extrema are guaranteed to exist by the Extreme Value Theorem.

We'll see how this process works in a couple of examples. First, we repeating an optimization problem which was previously done with substitution.

Example 108. *We'll find the absolute maximum and absolute minimum of $f(x, y) = x^2 - y^2$ subject to the constraint $x^2 + y^2 = 1$. Notice that $x^2 + y^2 = 1$ is a compact region, so absolute extrema will exist.*

If we let $g(x, y) = x^2 + y^2$, our constraint is $g(x, y) = 1$.

Computing the gradients of f and g, we have

$$\nabla f(x, y) = \boxed{(2x, -2y)},$$
$$\nabla g(x, y) = \boxed{(2x, 2y)}.$$

Setting up the system of equations

$$\begin{cases} \nabla f(\vec{x}) = \lambda \nabla g(\vec{x}) \\ g(\vec{x}) = 1 \end{cases},$$

we have

$$\begin{aligned} \nabla(2x, -2y) &= \lambda(2x, 2y) \\ x^2 + y^2 &= 1 \end{aligned}$$

We can rewrite this as a system of three equations,

$$\begin{cases} 2x = 2\lambda x \\ 2y = -2\lambda y \\ x^2 + y^2 = 1 \end{cases}.$$

From the first equation, we have either $x = 0$ *or* $\lambda = 1$.

If $x = 0$, *the third equation gives us* $y = \pm 1$. *Thus, we obtain the critical points* $(0, \pm 1)$.

If $\lambda = 1$, *the second equation gives us* $y = 0$. *Then, the third equation gives us* $x = \pm 1$. *This gives us the critical points* $(\pm 1, 0)$.

Since we know that f *will have an absolute maximum and minimum subject to our constraint, we will compare the values at the critical points to determine the absolute maximum and minimum.*

$$\begin{aligned} f(0, -1) &= -1 \\ f(0, 1) &= -1 \\ f(-1, 0) &= 1 \\ f(1, 0) &= 1 \end{aligned}$$

We see that the absolute maximum of f *subject to our constraint is* 1, *and this occurs at the points* $(\pm 1, 0)$. *The absolute minimum of* f *subject to our constraint is* -1, *and this occurs at the points* $(0, \pm 1)$.

Example 109. *Next, we'll attempt to optimize* $f(x, y) = 12 - x^2 - y^2$ *subject to the constraint* $y^2 = x + 5$. *Here, we'll need to be a bit more careful, since the curve defined by* $y^2 = x + 5$ *is not compact, as it is unbounded.*

PICTURE

However, Lagrange multipliers will still be helpful for finding critical points. We can rewrite our constaint as $y^2 - x = 5$, *and take* $g(x, y) = y^2 - x$. *So, we are optimizing* $f(x, y)$ *subject to the constraint* $g(x, y) = 5$.

We begin by finding the gradients of f *and* g.

$$\nabla f(x,y) = \boxed{(-2x,-2y)}$$
$$\nabla g(x,y) = \boxed{(-1,2y)}$$

Next, we solve the system

$$\begin{cases} \nabla f(\vec{x}) = \lambda \nabla g(\vec{x}) \\ g(\vec{x}) = 5 \end{cases},$$

which is

$$\begin{cases} (-2x,-2y) = \lambda(-1,2y) \\ y^2 - x = 5 \end{cases}.$$

We can rewrite this system as the three equations

$$\begin{cases} -2x = -\lambda \\ -2y = 2\lambda y \\ y^2 - x = 5 \end{cases}.$$

From the second equation, we have either $y = 0$, *or* $\lambda = -1$.

If $y = 0$, *the third equation gives us* $x = -5$. *So, we have a critical point* $(-5,0)$.

If $\lambda = -1$, *the first equation gives us* $x = -\frac{1}{2}$. *Then the third equation gives us* $y = \pm\frac{3}{\sqrt{2}}$. *So, we have the critical points* $\left(-\frac{1}{2}, \pm\frac{3}{\sqrt{2}}\right)$.

In this case, we can't determine the absolute maximum and absolute minimum by plugging in these points, since we aren't optimizing over a compact region.

However, looking at the graph of f *over the curve* $g(x,y) = 5$, *we can see that the absolute maximum occurs at the points* $\left(-\frac{1}{2}, \pm\frac{3}{\sqrt{2}}\right)$. *Although there is a local minimum at* $(-5,0)$, *this is not an absolute minimum, as there is no absolute minimum.*

GRAPH

Part VI

Homework

Homework 2: Graphing

Online Problems

Problem 16 *Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$f(x, y) = x^2 + 4y^2 - 2.$$

What is the domain of f?

Multiple Choice:

(a) $\mathbb{R}$

(b) $\mathbb{R} \setminus \{0\}$

(c) $[0, \infty)$

(d) $(0, \infty)$

(e) $\mathbb{R}^2$ ✓

(f) $\mathbb{R}^2 \setminus \{(0, 0)\}$

What is the range of f?

$$\text{Range } f = \boxed{[-2, \infty)}$$

Is f onto?

Multiple Choice:

(a) *yes*

(b) *no* ✓

Learning outcomes:
Author(s):

Problem 16.1 *We would like to restrict the codomain of the function f so that it becomes onto. We'll describe our new codomain as the set of numbers a in $\mathbb{R}$ such that some condition holds. Which condition gives us the largest possible codomain such that f is onto?*

Multiple Choice:

(a) $a \in \mathbb{R}$

(b) $a \geq 0$

(c) $a > 0$

(d) $a \neq 0$

(e) $a = 0$

(f) $a \geq 2$

(g) $a > 2$

(h) $a \neq 2$

(i) $a = 2$

(j) $a \geq -2$ ✓

(k) $a > -2$

(l) $a \neq -2$

(m) $a = -2$

Is f one-to-one?

Multiple Choice:

(a) *yes*

(b) *no* ✓

Problem 16.2 *We would like to restrict the domain of the function f, so that it becomes one-to-one. We'll describe our new domain as the set of points (x, y) in $\mathbb{R}^2$ such that some condition(s) hold. Which condition(s) give us the largest possible domain such that f is one-to-one?*

Select All Correct Answers:

(a) $x \neq 0$

(b) $x \geq 0$ ✓

(c) $x > 0$

(d) $y \neq 0$

(e) $y \geq 0$ ✓

(f) $y > 0$

Problem 17 *Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be the function defined by*

$$f(\vec{x}) = 3\vec{x} + \boldsymbol{i} - 2\boldsymbol{j}.$$

Find the component functions of f in terms of x, y, and z.

$$f_1(x, y, z) = \boxed{3x + 1}$$
$$f_2(x, y, z) = \boxed{3y - 2}$$
$$f_3(x, y, z) = \boxed{3z}$$

Problem 18 *Consider the linear function $f : \mathbb{R}^3 \to \mathbb{R}^2$ given by $f(\vec{x}) = A\vec{x}$, where*

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

and $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$.

(a) *Determine the component functions of f in terms of x_1, x_2, and x_3.*

$$f_1(x_1, x_2, x_3) = \boxed{x_1 + 5x_2 + 2x_3}$$
$$f_2(x_1, x_2, x_3) = \boxed{-2x_1 + x_3}$$

(b) *Is f one-to-one?*

Multiple Choice:

(i) *Yes*
(ii) *No* ✓

(c) *Is f onto?*

Multiple Choice:

(i) *Yes* ✓
(ii) *No*

Problem 19 *Consider the function*

$$f(x, y) = xy.$$

What is the shape of the level curve at height 0 *of* f*?*

Multiple Choice:

(a) *Empty*

(b) *A single line*

(c) *Two intersecting lines* ✓

(d) *Two parallel lines*

(e) *Circle*

(f) *Ellipse*

(g) *Parabola*

(h) *Hyperbola*

What is the shape of the level curve at height 1 *of* f*?*

Multiple Choice:

(a) *Empty*

(b) *A single line*

(c) *Two intersecting lines*

(d) *Two parallel lines*

(e) *Circle*

(f) *Ellipse*

(g) *Parabola*

(h) *Hyperbola* ✓

What is the shape of the level curve at height -1 of f?

Multiple Choice:

(a) *Empty*

(b) *A single line*

(c) *Two intersecting lines*

(d) *Two parallel lines*

(e) *Circle*

(f) *Ellipse*

(g) *Parabola*

(h) *Hyperbola* ✓

What is the shape of the level curve at height 2 of f?

Multiple Choice:

(a) *Empty*

(b) *A single line*

(c) *Two intersecting lines*

(d) *Two parallel lines*

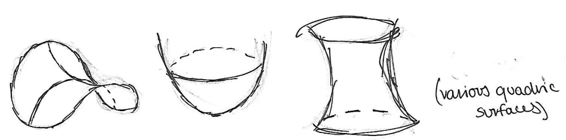
(e) *Circle*

(f) *Ellipse*

(g) *Parabola*

(h) *Hyperbola* ✓

Which of the following is the graph of f?



Problem 20 *Consider the function*

$$f(x, y) = |x|.$$

What is the shape of the level curve at height 0 *of* f*?*

Multiple Choice:

(a) *Empty*
(b) *A single line* ✓
(c) *Two intersecting lines*
(d) *Two parallel lines*
(e) *Circle*
(f) *Ellipse*
(g) *Parabola*
(h) *Hyperbola*

What is the shape of the level curve at height 1 *of* f*?*

Multiple Choice:

(a) *Empty*
(b) *A single line*
(c) *Two intersecting lines*
(d) *Two parallel lines* ✓
(e) *Circle*
(f) *Ellipse*
(g) *Parabola*
(h) *Hyperbola*

What is the shape of the level curve at height -1 *of* f*?*

Multiple Choice:

(a) *Empty* ✓

(b) *A single line*

(c) *Two intersecting lines*

(d) *Two parallel lines*

(e) *Circle*

(f) *Ellipse*

(g) *Parabola*

(h) *Hyperbola*

What is the shape of the level curve at height 2 of f?

Multiple Choice:

(a) *Empty*

(b) *A single line*

(c) *Two intersecting lines*

(d) *Two parallel lines* ✓

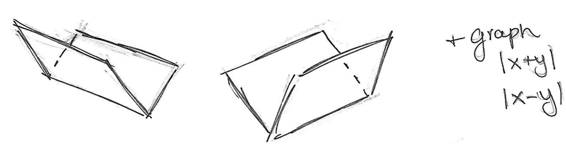
(e) *Circle*

(f) *Ellipse*

(g) *Parabola*

(h) *Hyperbola*

Which of the following is the graph of f?



Problem 21 *Which of the following is the graph of the ellipsoid*

$$\frac{x^2}{9} + y^2 + \frac{z^2}{4} = 1?$$

PICTURES

Is there a function $f(x, y)$ such that the graph of f is the ellipsoid above?

Multiple Choice:

(a) *Yes*

(b) *No* ✓

Problem 21.1 *Why is this impossible?*

Multiple Choice:

(a) *It wouldn't be one-to-one.*

(b) *It wouldn't be onto.*

(c) *There would be multiple inputs with the same output.*

(d) *A single input would need to have two outputs.* ✓

Problem 22 *Classify the quadric surface defined by the equation*

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

Multiple Choice:

(a) *Ellipsoid* ✓

(b) *Elliptic Paraboloid*

(c) *Hyperbolic Paraboloid*

(d) *Elliptic Cone*

(e) *Hyperboloid of One Sheet*

(f) *Hyperboloid of Two Sheets*

It is centered at $\boxed{(0, -1, 0)}$.

Problem 22.1 *Which of the following is the graph of the quadric surface given above?*

GRAPHS

Problem 23 *Classify the quadric surface defined by the equation*

$$2x^2 + 2y^2 - 8y - z + 4 = 0$$

Multiple Choice:

(a) *Ellipsoid*

(b) *Elliptic Paraboloid* ✓

(c) *Hyperbolic Paraboloid*

(d) *Elliptic Cone*

(e) *Hyperboloid of One Sheet*

(f) *Hyperboloid of Two Sheets*

It is centered at $\boxed{(0, 2, -4)}$.

Problem 23.1 *Which of the following is the graph of the quadric surface given above?*

GRAPHS

Written Problems

Problem 24 *Consider the function*

$$f(x, y, z) = \frac{4}{\sqrt{9 - x^2 - y^2 - z^2}}.$$

(a) *What is the domain of f? Describe this domain as a region in $\mathbb{R}^3$.*

(b) *What is the range of f?*

Problem 25 *Consider the function*

$$f(x) = x^2 + y^2 - 4.$$

(a) *Draw at least five level curves of f.*

(b) *Use these level curves to sketch the graph of f.*

Problem 26 *Draw the graph of the surface in $\mathbb{R}^3$ determined by the equation*

$$x = y^2/4 - z^2/9.$$

Use level curves and/or sections to justify why your drawing is correct.

Homework 3: Parametrized Curves

Online Problems

Problem 27 *Two ants are running on the top of a table. Their paths are described by*

$$\vec{x}(t) = (t^2 + 1, 2t - 1)$$

and

$$\vec{y}(t) = (\sqrt{t+3}, t),$$

with coordinates in inches, for $t \geq 0$ in seconds.

At what time do the ants collide?

$$t = \boxed{1}$$

Where do the ants collide?

$$(x, y) = \boxed{(2, 1)}$$

Problem 28 *Several parametrized curves are graphed below, and the arrow indicates the direction in which the parameter increases.*

PICTURE

Which is the graph of the path $\vec{x}(t) = (4 - t, 2t + 1)$, for $-1 \leq t \leq 1$?

Multiple Choice:

(a) *(a)*

(b) *(b)*

(c) *(c)*

(d) *(d)*

(e) *(e)*

Learning outcomes:
Author(s):

Problem 29 *Several parametrized curves are graphed below, and the arrow indicates the direction in which the parameter increases.*

PICTURE

Which is the graph of the path $\vec{x}(t) = (3\sin(t), 2\cos(t))$, for $0 \le t \le \pi$?

Multiple Choice:

(a) *(a)*

(b) *(b)*

(c) *(c)*

(d) *(d)*

(e) *(e)*

Problem 30 *Several parametrized curves are graphed below, and the arrow indicates the direction in which the parameter increases.*

PICTURE

Which is the graph of the path $\vec{x}(t) = (t^2, t^3)$, for $-1 \le t \le 1$?

Multiple Choice:

(a) *(a)*

(b) *(b)*

(c) *(c)*

(d) *(d)*

(e) *(e)*

Problem 31 *Several parametrized curves are graphed below, and the arrow indicates the direction in which the parameter increases.*

PICTURE

Which is the graph of the path $\vec{x}(t) = (e^t, t)$, for $-1 \le t \le 1$?

Multiple Choice:

(a) *(a)*

(b) *(b)*

(c) *(c)*

(d) *(d)*

(e) *(e)*

Problem 32 *Consider the curve below.*

PICTURE

Which of the following are parametrizations for the curve? Select all that apply.

Select All Correct Answers:

(a) $\vec{x}(t) = (\cos t, \sin t)$, *for* $0 \leq t \leq 2\pi$

(b) $\vec{x}(t) = (\sin t, \cos t)$, *for* $0 \leq t \leq \pi$

(c) $\vec{x}(t) = (\cos t, -\sin t)$, *for* $0 \leq t \leq \pi$ ✓

(d) $\vec{x}(t) = (-\cos t, \sin t)$, *for* $0 \leq t \leq \pi$

(e) $\vec{x}(t) = (\sin t, \cos t)$, *for* $\pi/2 \leq t \leq \pi/2$ ✓

(f) $\vec{x}(t) = (\cos t, \sin t)$, *for* $\pi \leq t \leq 2\pi$ ✓

(g) $\vec{x}(t) = (-\sqrt{1-t^2}, t)$ *for* $-1 \leq t \leq 1$

(h) $\vec{x}(t) = (\sqrt{1-t^2}, t)$ *for* $0 \leq t \leq 1$

(i) $\vec{x}(t) = (t, -\sqrt{1-t^2})$ *for* $-1 \leq t \leq 1$ ✓

Problem 33 *Consider the curve below.*

PICTURE

Which of the following are parametrizations for the curve? Select all that apply.

Select All Correct Answers:

(a) $\vec{x}(t) = (t, \sin t)$ *for* $0 \leq t \leq 2\pi$ ✓

(b) $\vec{x}(t) = (\arcsin t, t)$ *for* $-1 \leq t \leq 1$

(c) $\vec{x}(t) = (2t^2, \sin(2t^2))$ *for* $-1 \leq t \leq 1$ ✓

(d) $\vec{x}(t) = \left(t, \cos\left(\frac{\pi/2}{(}t - 1)\right)\right)$ *for* $0 \le t \le 2$

(e) $\vec{x}(t) = \left(\frac{\pi}{2}t, \cos\left(\frac{\pi/2}{t}\right)\right)$ *for* $0 \le t \le 2$

(f) $\vec{x}(t) = \left(\frac{\pi}{2}t, \cos\left(\frac{\pi/2}{(}t - 1)\right)\right)$ *for* $0 \le t \le 2$

(g) $\vec{x}(t) = \left(\frac{\pi}{2}t, \cos\left(\frac{\pi/2}{(}t - 1)\right)\right)$ *for* $0 \le t \le 4$ ✓

Problem 34 *Consider the path* $\vec{x}(t) = (3\cos(t), -2\sin(t))$, *for* $t \in \mathbb{R}$.

Compute the velocity of $\vec{x}$.

$$\vec{v}(t) = \boxed{(-3\sin(t), -2\cos(t))}$$

Compute the speed of $\vec{x}$.

$$\|\vec{x}'(t)\| = \boxed{\sqrt{13}}$$

Problem 35 *Consider the path* $\vec{x}(t) = (\cos(t^4), \sin(t^4), \frac{1}{2}t^4)$, *for* $t \ge 0$.

Compute the velocity.

$$\vec{v}(t) = \boxed{(-4t^3\sin(t^4), 4t^3\cos(t^2), 2t^3)}$$

Compute the speed.

$$\|\vec{x}'(t)\| = \boxed{\sqrt{20t^3}}$$

Problem 36 *Consider the curve* $\vec{x}(t) = (4t + 2, 1 - 3t)$ *for* $t \in \mathbb{R}$.

Compute the velocity.

$$\vec{v}(t) = \boxed{(4, -3)}$$

Compute the speed.

$$\|\vec{x}'(t)\| = \boxed{4}$$

Problem 37 *Consider the curve* $\vec{x}(t) = (2\cos t, 5\sin t, t^2)$ *for* $t \in \mathbb{R}$.

Find the velocity.

$$\vec{v}(t) = \boxed{(-2\sin t, 5\cos t, 2t)}$$

Find the velocity when $t = \pi$.

$$\vec{v}(\pi) = \boxed{(0, -5, 2\pi)}$$

Find a parametrization for the tangent line to $\vec{x}$ *at the point where* $t = \pi$, *so that* $L(0) = \vec{x}(\pi)$.

$$L(t) = \boxed{(-2, 5, \pi^2) + t(0, -5, 2\pi)}$$

Problem 38 *Consider the curve* $\vec{x}(t) = (t, t^2, t^3)$ *for* $t \in \mathbb{R}$.

Find the velocity.

$$\vec{v}(t) = \boxed{(1, 2t, 3t^2)}$$

Find the velocity when $t = 2$.

$$\vec{v}(\pi) = \boxed{(1, 4, 12)}$$

Find a parametrization for the tangent line to $\vec{x}$ *at the point where* $t = 2$, *so that* $L(0) = \vec{x}(2)$.

$$L(t) = \boxed{(2, 4, 8) + t(1, 4, 12)}$$

Problem 39 *Consider the curve* $\vec{x}(t) = (t, te^t, e^{t^2})$ *for* $t \in \mathbb{R}$.

Find the velocity.

$$\vec{v}(t) = \boxed{(1, e^t + te^t, 2te^{t^2})}$$

Find the velocity when $t = 0$.

$$\vec{v}(\pi) = \boxed{(1, 1, 0)}$$

Find a parametrization for the tangent line to $\vec{x}$ *at the point where* $t = 0$, *so that* $L(0) = \vec{x}(0)$.

$$L(t) = \boxed{(0, 0, 1) + t(1, 1, 0)}$$

Written Problems

Problem 40 (a) *Graph the surface* $z^2 = x^2 + y^2$ *and the curve* $\vec{x}(t) = (t\cos(t), t\sin(t), t)$ *for* $-5 \le t \le 5$.

(b) *Verify algebraically that the curve lies on the surface.*

Problem 41 (a) *Graph the surface* $1 = x^2 + y^2 + z^2$ *and the curve* $\vec{x}(t) = (\cos(8t)\sin(t), \sin(8t)\sin(t), \cos(t))$ *for* $0 \le t \le \pi$.

(b) *Verify algebraically that the curve lies on the surface.*

Problem 42 *Prove the following product rule for cross products.*

Let $\vec{x}$ *and* $\vec{y}$ *be paths in* $\mathbb{R}^3$, *then*

$$(\vec{x} \times \vec{y})'(t) = \vec{x}'(t) \times \vec{y}(t) + \vec{x} \times \vec{y}'(t),$$

for t *such that* $x'(t)$ *and* $y'(t)$ *exist.*

Problem 43 *Consider the path* $\vec{x}(t) = (3t - 3^3, 3t^2)$, *for* $t \in \mathbb{R}$.

(a) *Graph* $\vec{x}$.

(b) *Find the point* P *where* $\vec{x}$ *intersects itself.*

(c) *There are two tangent vectors* $\vec{x}$ *at* P, *one for each time the path passes through this point. Find the angle between these two vectors.*

Problem 44 *Consider the unit circle* $x^2 + y^2 = 1$ *in* $\mathbb{R}^2$, *and consider all lines* l_t *passing through the point* $(1, 0)$, *indexed by their slopes* t. *For the line of slope* t, *let* $\vec{x}(t)$ *be the point (other than* $(0, 1)$*) where the line* l_t *intersects the unit circle.*

(a) *Find* $\vec{x}(0)$, $\vec{x}(1)$, *and* $\vec{x}(-1)$.

(b) *Find an equation for* l_t.

(c) *Use your equation for* l_t *and the equation for the unit circle to find* $\vec{x}(t)$ *in terms of only* t.

(d) *Consider the path* $\vec{x}(t)$, *for* $t \in \mathbb{R}$, *given by your answer to (c). What curve does this path parametrize? Are there any "missing" points?*

Homework 8: Differentiability

Graded Problems

Problem 45 *Consider the linear transformation $\mathbf{T} : \mathbb{R}^3 \to \mathbb{R}^2$ given by*

$$T(x, y, z) = \begin{pmatrix} 1 & 0 & 2 \\ 4 & 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

(a) *Compute $D\mathbf{T}(0, 0, 0)$.*

(b) *Explain why your answer to (a) makes sense in the context of linear approximations.*

Problem 46 *Let $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ be the function given by*

$$\mathbf{f}(x, y) = (x^2 + y, 3y).$$

In this problem you will prove according to the limit definition that $\mathbf{f}$ is differentiable at the point $\mathbf{a} = (1, 2)$. In each part, do all calculations by hand and show appropriate amounts of work. Do not use technology.

(a) *Find the derivative matrix $D\mathbf{f}$ of $\mathbf{f}$.*

(b) *To prove $\mathbf{f}$ is differentiable, you must show it has a good linear approximation at the point $\mathbf{a} = (1, 2)$. Using your answer from part (a), determine the formula for this linear approximation. (No simplifications necessary.)*

(c) *Use Definition 3.8 to prove that $\mathbf{f}$ is differentiable at $\mathbf{a} = (1, 2)$. In this step it would be very useful to simplify the numerator of the "difference quotient" as much as possible, and then make some substitutions on the top and bottom to help you evaluate the limit.*

Learning outcomes:
Author(s):

Professional Problem

Problem 47 *The first draft of your project is due this week.*

Completion Packet

Problem 48 *Find the gradient of each function at the given point.*

(a) $f(x,y) = xe^{xy} + x^2y$, *at the point* $(1,1)$.

(b) $g(x,y) = \sin(x^2+y) + y\cos(x)$, *at the point* $(0,\pi)$.

Problem 49 *Find the matrix of partial derivatives of the following functions.*

(a) $f(x,y,z) = \ln(x^2y) + xyz$

(b) $\mathbf{f}(x,y,z) = \left(\dfrac{\sqrt{x^2+y^2}}{z}, z^2y^3\right)$

(c) $\mathbf{f}(x) = (\ln(x), xe^x, \sin(x))$

Problem 50 *Consider the function* $\mathbf{f}(x,y,z) = \left(\dfrac{1}{x^2+y^2}, \dfrac{xy}{z}\right)$.

(a) *Compute the partial derivatives of* $\mathbf{f}$.

(b) *Show that* $\mathbf{f}$ *is differentiable on its domain.*

Problem 51 *Find the point where the plane tangent to* $z = x^2+y^2$ *at* $(1,1,2)$ *intersects the* z*-axis.*

Problem 52 *Because the partial derivatives of a function are necessary to construct the limit in the definition of differentiability, a sort of converse to Theorem 3.10 is: If* $\mathbf{f} : D \subset \mathbb{R}^n \to \mathbb{R}^m$ *is differentiable at* $\mathbf{a} \in \mathbb{R}^n$*, then the partial derivatives of* $\mathbf{f}(\mathbf{x})$ *must all be defined at* $\mathbf{a}$*. Use this result to show the following functions are not differentiable at the indicated point.*

(a) $f(x, y, z) = \sqrt{x^2 + y^2 + z^2}$, *at the point* $(0, 0, 0)$.

(b) $g(x, y) = |x - y|$, *at the point* $(3, 3)$.

Homework 9: Properties of Derivatives

Graded Problems

Problem 53 (a) *Suppose $f : \mathbb{R}^3 \to \mathbb{R}$ is a smooth function. Prove that $f_{xyzx} = f_{zxxy}$, where these are fourth-order partial derivatives. You are free to use Theorem 4.3 for this problem, but not Theorem 4.5.*

(b) *Verify the result of (a) for the function $f(x, y, z) = x^3y^2z$.*

Problem 54 *Suppose we have differentiable functions $\mathbf{f} : \mathbb{R}^2 \to \mathbb{R}^2$ and $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$ such that*

$$\begin{aligned} \mathbf{g}(1,2) &= (3,0), \\ D\mathbf{f}(x,y) &= \begin{pmatrix} xe^y & x^2y \\ 0 & yx^2+1 \end{pmatrix}, \\ D(\mathbf{f} \circ \mathbf{g})(1,2) &= \begin{pmatrix} 1 & 4 \\ 2 & -1 \end{pmatrix}. \end{aligned}$$

Find $D\mathbf{g}(1,2)$.

Professional Problem

Problem 55 (a) *Consider differentiable functions $g : \mathbb{R} \to \mathbb{R}$ and $f : \mathbb{R}^2 \to \mathbb{R}$ such that $f(x, y) = 0$ when $y = g(x)$. Prove that if $\partial f/\partial y \neq 0$, then*

$$\frac{dg}{dx} = -\frac{\partial f/\partial x}{\partial f/\partial y}.$$

(b) *Consider the equation $\sin(x) + \cos(y) = 0$. Make a suitable choice of f, and use the result of (a) to compute $\frac{dy}{dx}$ in terms of x and y.*

(c) *Use implicit differentiation to verify your answer to (b).*

Learning outcomes:
Author(s):

Completion Packet

Problem 56 *Consider the functions* $\mathbf{f}(x, y) = (x^2 + y, xy - \sin(xy))$ *and* $\mathbf{g}(x, y) = (e^{xy}, x^2 y)$. *Compute* $D(\mathbf{f} + \mathbf{g})$ *in two ways:*

(a) *By computing the derivative of* $\mathbf{f} + \mathbf{g}$

(b) *By computing the derivatives of* $\mathbf{f}$ *and* $\mathbf{g}$, *and using the sum rule.*

Problem 57 *Consider the functions* $f(x, y, z) = xy^2z^3$ *and* $g(x, y, z) = xyz$.

(a) *Verify that*

$$D(fg)(\mathbf{a}) = g(\mathbf{a})Df(\mathbf{a}) + f(\mathbf{a})Dg(\mathbf{a}).$$

(b) *Verify that*

$$D(f/g)(\mathbf{a}) = \frac{g(\mathbf{a})Df(\mathbf{a}) - f(\mathbf{a})Dg(\mathbf{a})}{g(\mathbf{a})^2}.$$

Problem 58 *Find all second-order partial derivatives of the function* $f(x, y) = \ln(xy)$.

Problem 59 *Find all second-order partial derivatives of the function* $g(u, v) = e^{u^2+v^2}$.

Problem 60 *Find all second-order partial derivatives of the function* $h(x, y, z) = xy^2z^3$.

Problem 61 *Consider the functions* $y(s, t) = e^s + e^{st} + e^t$ *and* $\mathbf{x}(t) = (t, t^2)$. *Compute* $D(\mathbf{x} \circ y)$ *in two ways:*

(a) *Determining a formula for the composition* $\mathbf{x} \circ y$, *then computing the total derivative.*

(b) *Computing total derivatives of* $\mathbf{x}$ *and* y, *and using the chain rule.*

Problem 62 *Consider functions* $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^2$ *and* $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^3$ *such that* $\mathbf{f}(x, y, z) = (x^2y + e^z, e^x + y)$ *and* $D\mathbf{g}(x, y) = \begin{pmatrix} y & x \\ 0 & 1 \\ 0 & xy \end{pmatrix}$. *For each of the given total derivatives, either explain why they do not exist, compute them, or explain what additional information we would need to compute them.*

(a) $D(\mathbf{f} \circ \mathbf{g})$

(b) $D(\mathbf{g} \circ \mathbf{f})$

Problem 63 *Consider functions* $\mathbf{f} : \mathbb{R}^3 \to \mathbb{R}^2$ *and* $\mathbf{g} : \mathbb{R}^2 \to \mathbb{R}^2$ *such that* $\mathbf{f}(x, y, z) = (\sin(xyz), \cos(xyz))$ *and* $D\mathbf{g}(x, y) = \begin{pmatrix} e^{xy} & 0 \\ 0 & e^{xy} \end{pmatrix}$. *For each of the given total derivatives, either explain why they do not exist, compute them, or explain what additional information we would need to compute them.*

(a) $D(\mathbf{f} \circ \mathbf{g})$

(b) $D(\mathbf{g} \circ \mathbf{f})$

Problem 64 *Consider the function* $\mathbf{g} : \mathbb{R}^3 \to \mathbb{R}^3$ *given by* $\mathbf{g}(\rho, \theta, \phi) = (\rho \cos\theta \sin\phi, \rho \sin\theta \sin\phi, \rho \cos\phi)$, *which changes spherical coordinates to Cartesian coordinates. For any differentiable function* $f : \mathbb{R}^3 \to \mathbb{R}$, *compute* $\partial f/\partial\rho$, $\partial f/\partial\theta$, *and* $\partial f/\partial\phi$ *in terms of* $\partial f/\partial x$, $\partial f/\partial y$, *and* $\partial f/\partial z$.

Homework 10: Directional Derivatives

Graded Problems

Problem 65 *Consider the function*

$$f(x,y) = \begin{cases} \dfrac{2x^2y}{x^4+y^2} & \text{for } (x,y) \neq (0,0) \\ 0 & \text{for } (x,y) = (0,0). \end{cases}$$

(a) *For the vector* $\mathbf{u} = (a,0)$ *in* $\mathbb{R}^2$*, find* $D_{\mathbf{u}}f(0,0)$*.*

(b) *For any vector of the form* $\mathbf{u} = (a,b)$ *with* $b \neq 0$ *in* $\mathbb{R}^2$*, find* $D_{\mathbf{u}}f(0,0)$*.*

(c) *Is* f *continuous at* $(0,0)$*?*

Problem 66 *Consider the function* $f(x,y,z) = 9x^2 + 4y^2 + z^2$*. At the point* $(2,-1,1)$*, find the direction in which* f *is changing most rapidly.*

Professional Problem

Problem 67 *You must email me the revised draft of your project by Wednesday, 11/21. Your revised draft must be sent as a pdf, and it must include your pin number, but not your names.*

Completion Packet

Problem 68 *For each function, compute the directional derivative* $D_{\mathbf{u}}f(\mathbf{a})$*.*

(a) $f(x,y) = x - 2xy$, $\mathbf{u} = (-1,1)$, $\mathbf{a} = (2,1)$.

Learning outcomes:
Author(s):

(b) $f(x,y) = \dfrac{1}{x-y}$, $\mathbf{u} = (1,1)$, $\mathbf{a} = (1,2)$.

(c) $f(x,y,z) = e^{xyz} + \dfrac{1}{x-y+2z}$, $\mathbf{u} = (2,1,0)$, $\mathbf{a} = (1,1,1)$.

Problem 69 *The alien Pmytmu is building a temple to honor their god, Suluclac. The temple is in the shape of a hemisphere with height 10 Azzips. Pmytmu is standing on the temple, at a vertical height of 5 Azzips above the ground. They are directly East of the center of the temple.*

(a) *Determine the slope of Pmytmu's path if they move east on the surface of the temple.*

(b) *Determine the slope of Pmytmu's path if they move west on the surface of the temple.*

(c) *Determine the slope of Pmytmu's path if they move southwest on the surface of the temple.*

Problem 70 *Through a series of unfortunate events, the unlucky alien Pmytmu has left their planet, and they are now floating around in space. It's very cold in space, and Pmytmu is afraid they might not survive due to the low temperatures. Fortunately, Pmytmu has some old textbooks with them. Due to conservation of momentum, when Pmytmu throws a textbook in one direction, Pmytmu will float in the exact opposite direction. The temperature in space is given by the function*

$$T(x,y,z) = x^2 + yz - e^{xy},$$

and Pmytmu is currently at the point $(1,1,1)$. Pmytmu would like to move in the direction in which temperature is increasing the fastest. In which direction should Pmytmu throw their first textbook?

Problem 71 *Find an equation for the tangent plane to each given surface at the given point.*

(a) $e^{xyz}\sin(x) = 1$, *at the point* $(\pi/2, -1, 0)$.

(b) $x^2 - z^2 + 2xy - 3yz = 3$ *at the point* $(1,1,0)$.

Problem 72 *Let* $\mathbf{x} : \mathbb{R} \to \mathbb{R}^2$ *be a smooth curve which is a level set of the graph of a function* $f : \mathbb{R}^2 \to \mathbb{R}$ *at height* $z = c$*. For any point* $\mathbf{a} = \mathbf{x}(t_0)$ *in* $\mathbb{R}^2$*, what is the directional derivative of* f *at* $\mathbf{a}$ *in the direction of* $\mathbf{x}'(t_0)$*?*

Homework 11: Taylor's Theorem

Graded Problems

Problem 73 *Give a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ with $a > 0$ and $\det(A) > 0$ for which the quadratic form $p(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is **NOT** positive definite. Does this mean the theorem in problem 6 is incorrect?*

Problem 74 *Prove the following theorem: Let $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ be a symmetric 2×2 matrix. If $\det(A) < 0$, then the quadratic form $p(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is indefinite, regardless of the value of a.*

(Hint: think about the cases $a > 0$, $a = 0$ and $a < 0$. In two cases you can apply Sylvester's Theorem. In the third case, you'll have to do some work by hand to show $p(x, y)$ can have both positive and negative values.)

Professional Problem

Problem 75 *Complete the online peer review form posted on moodle.*

Completion Packet

Problem 76 *Find the symmetric matrix that represents each quadratic form.*

(a) $r(x_1, x_2, x_3, x_4) = x_3^2 - x_2x_3 + x_1x_4$

Learning outcomes:
Author(s):

(b) $t(\mathbf{x}) = \mathbf{x}^T \begin{bmatrix} 6 & 1 & 8 & -2 \\ 0 & 5 & 1 & 9 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{x}$

Problem 77 *Prove the following theorem* without *using Sylvester's theorem: Let* $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ *be a symmetric* 2×2 *matrix. If* $a > 0$ *and* $\det(A) > 0$*, then the quadratic form* $p(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ *is positive definite.*

(Hint: Write out p *in terms of the variables* x *and* y*, then complete the square with respect to* x *and collect the remaining terms.)*

Problem 78 (a) *Write the Taylor series for* $f(x) = \sin(x)$ *centered at* $x = 0$.

(b) *Find the second-order Taylor approximation for* $f(x, y) = \sin(xy)$ *centered at* $(0, 0)$*, using your answer to part (a).*

(c) *Verify your answer to part (b), by computing the second-order Taylor approximation for* $f(x, y) = \sin(xy)$ *directly.*

Problem 79 (a) *Write the Taylor series for* $f(x) = e^x$ *centered at* $x = 0$.

(b) *Find the second-order Taylor approximation for* $f(x, y) = e^{x^2+y^2}$ *centered at* $(0, 0)$*, using your answer to part (a).*

(c) *Verify your answer to part (b), by computing the second-order Taylor approximation for* $f(x, y) = e^{x^2+y^2}$ *directly.*

Problem 80 *Compute the Hessian matrix for each function at the given point.*

(a) $f(x, y) = \sqrt{xy}$ at $(1, 1)$

(b) $f(x, y) = \cos(x) + x^2 \sin(y)$ at $(0, \pi)$

(c) $f(x, y, z) = \dfrac{1}{x^2 + y^2 + z^2 + 1}$ at $(0, 0, 0)$

Problem 81 *Consider the function* $f(x,y) = \dfrac{x}{x+y}$ *and the point* $\mathbf{a} = (1,2)$.

(a) *Find the first-order Taylor polynomial of* f *at* $\mathbf{a}$.

(b) *Find the second-order Taylor polynomial of* f *at* $\mathbf{a}$.

(c) *Express the second-order Taylor polynomial using the derivative matrix and the Hessian matrix, as in formula (10) of section 4.1 of the textbook.*

Problem 82 *Consider the function* $f(x,y) = x^2 e^y$ *and the point* $\mathbf{a} = (1,0)$.

(a) *Find the first-order Taylor polynomial of* f *at* $\mathbf{a}$.

(b) *Find the second-order Taylor polynomial of* f *at* $\mathbf{a}$.

(c) *Express the second-order Taylor polynomial using the derivative matrix and the Hessian matrix, as in formula (10) of section 4.1 of the textbook.*

Homework 12: Extrema

Graded Problems

Problem 83 *Consider the function* $f(x, y) = (y^2 - x)(2y^2 - x)$.

(a) *Show that* $f(x, y)$ *has a critical point at the origin.*

(b) *Show that* $f(x, y)$ *has a local minimum along any line through the origin. That is, show that for constant* $(a, b) \neq (0, 0)$, *that the function* $g(t) = f(at, bt)$ *has a local minimum at* $t = 0$.

(c) *Show that* $f(x, y)$ *does not have a local minimum at the origin.*

Problem 84 (a) *Let* $x_1, x_2, ..., x_n$ *be nonnegative numbers such that their sum is constant. That is,* $x_1 + x_2 + \cdots + x_n = C$ *for some constant* C. *Show that the product* $x_1 x_2 \cdots x_n$ *is a maximum if and only if* $x_1 = x_2 = \cdots = x_n = C/n$.

(b) *Using the result of part (a), show that if* $x_1, x_2, ..., x_n$ *are nonnegative numbers such that* $x_1 + x_2 + \cdots + x_n = 1$, *then* $x_1 x_2 \cdots x_n \leq 1/n^n$.

Professional Problem

Problem 85 *Final draft of your project, worth two professional problems.*

Completion Packet

Problem 86 *Find and classify all critical points of each function. When using the Hessian fails to classify points, find another method.*

Learning outcomes:
Author(s):

(a) $f(x,y) = x^3 + y^3 + 12xy$

(b) $f(x,y) = x^2 + y^2 + \dfrac{1}{x^2y^2}$, for $xy \neq 0$

(c) $f(x,y) = (x+y)^2 + x^4$

(d) $f(x,y) = (x+y)e^{-xy}$

(e) $f(x,y,z) = 2x^2 + y^2 + z^2 - xz + xy$

(f) $f(x,y,z) = xy + xz$

Problem 87 *What are the conditions on a, b, c for $f(x,y) = ax^2 + bxy + cy^2$ to have a...*

(a) *...local minimum at the origin?*

(b) *...local maximum at the origin?*

(c) *...saddle point at the origin?*

Problem 88 *For nonzero constants a and b, consider the function $f(x,y) = ax^{-1} + by^{-1} + xy$.*

(a) *Find the (single) critical point of this function.*

(b) *What are the conditions on a and b for the critical point to be a local minimum? A local maximum? A saddle point?*

Problem 89 *Find the shortest distance between a point on the surface $(x-2)^2 + (y-3)^2 + z^2 = 1$ and the origin in $\mathbb{R}^3$.*

Problem 90 (a) *Let $x_1, x_2, ..., x_n$ be positive numbers such that their product is constant. That is, $x_1 x_2 \cdots x_n = C$ for some constant C. Show that the sum $x_1 + x_2 + \cdots + x_n$ is a* local *minimum if and only if $x_1 = x_2 = \cdots = x_n$.*

(b) *The local minimum which you found in part (a) is, in fact, an absolute minimum (you do not need to show this). Using this fact, show that if $x_1, x_2, ..., x_n$ are positive numbers such that $x_1 x_2 \cdots x_n = 1$, then $x_1 + x_2 + \cdots + x_n \geq n$.*

Winter Assignment: Substitution and Lagrange Multipliers

On the Winter Assignment there is no professional problem or designated graded problems. Staple all of your problems together in one packet. Some of the problems will be graded for correctness, and others will be graded for completion.

Problem 91 *Use substitution to find the absolute maximum and minimum values for each function and constraint.*

(a) $f(x, y) = xy$, *subject to* $y = x^2 - 1$, *with* $-2 \leq x \leq 2$.

(b) $f(x, y) = 2x - y$, *subject to* $g(x, y) = x^2 + y^2 = 1$.

(c) $f(x, y) = x^2 - 3y^2$, *subject to* $\dfrac{x^2}{16} + \dfrac{y^2}{4} = 1$.

Problem 92 *Use the method of Lagrange Multipliers to identify all critical points of each function subject to the constraints.*

(a) $f(x, y, z) = xyz$, *subject to* $x + y + z = C$ *for some fixed* $C > 0$.
(You've seen this problem before, so make sure to show your work and justify your answer!).

(b) $f(x, y) = x^2 + y^2$, *subject to* $x - y + \dfrac{1}{2}(x + y)^2 = 3$.

(c) $f(x, y, z) = x^2 + y^2 + z^2$, *subject to* $x - y + z = 2$

Learning outcomes:
Author(s):

Part VII

Practice Problems

Practice Problems: Review

Online Problems

Problem 93 *Compute the following:*

$$(1, 2, 3) + (8, 3, 6) = \boxed{(9, 5, 9)}$$

$$4(1, -2, 4) = \boxed{(4, -8, 16)}$$

$$-12((5, 2, 6) - (8, 2, 4)) = \boxed{(36, 0, 24)}$$

Problem 94 *Let h be a constant. Compute the following:*

$$(7, 2, -1) + (2h, 0, h) = \boxed{(7 + 2h, 2, h - 1)}$$

$$h(1, 8, 2) = \boxed{(h, 8h, 2h)}$$

Problem 95 *For each of the following, determine whether the quantity exists or does not exist.*

$(1, 8, 3, 7) + (-1, 7, 2, 7)$

Multiple Choice:

(a) *Exists.* ✓

Learning outcomes:
Author(s):

(b) *Does not exist.*

$(2, 8, 3) + (1, 7)$

Multiple Choice:

(a) *Exists.*

(b) *Does not exist.* ✓

$(2, 7, 3) + 1$

Multiple Choice:

(a) *Exists.*

(b) *Does not exist.* ✓

$(2, 8, 3)(1, 7, 3)$

Multiple Choice:

(a) *Exists.*

(b) *Does not exist.* ✓

$2(7, 2, 3, 7, 2)$

Multiple Choice:

(a) *Exists.* ✓

(b) *Does not exist.*

Problem 96 *For points $P_1 = (2, -3, 7, 1)$ and $P_2 = (-1, 7, 2, 1)$, compute the displacement vector $\vec{P_1P_2}$.*

$$\vec{P_1P_2} = \boxed{(-3, 10, -5, 0)}$$

Problem 97 *Write the vector $2\mathbf{i} - 5\mathbf{j} + 2\mathbf{k}$ in $\mathbb{R}^3$ in standard vector notation.*

$$2\mathbf{i} - 5\mathbf{j} + 2\mathbf{k} = \boxed{(2, -5, 2)}$$

Problem 98 *Compute the dot product.*

$$(1, 8, 3) \cdot (-2, 6, 0) = \boxed{46}$$

Problem 99 *Compute the dot product.*

$$(1, -5, 0, 2) \cdot (2, -1, 4, 1) = \boxed{9}$$

Problem 100 *Compute the dot product.*

$$(1, 8, 3) \cdot (-3, 0, 1) = \boxed{0}$$

What can you conclude about the vectors?

Multiple Choice:

(a) *They're perpendicular.* ✓

(b) *They aren't perpendicular.*

Problem 101 *Compute the dot product.*

$$(1, 2, 3) \cdot (-3, -2, -1) = \boxed{-10}$$

What can you conclude about the vectors?

Multiple Choice:

(a) *They're perpendicular.*

(b) *They aren't perpendicular.* ✓

Problem 102 *Compute the dot product.*

$$(1, 8, 3, 6) \cdot (3, -3, -1, 4) = \boxed{-0}$$

What can you conclude about the vectors?

Multiple Choice:

(a) *They're perpendicular.* ✓

(b) *They aren't perpendicular.*

Problem 103 *For each expression, determine whether it exists or does not exist.*

$(2, 8, 3) \cdot (1, 8, 2, 4)$

Multiple Choice:

(a) *Exists.*

(b) *Does not exist.* ✓

$(2, 8, 3, 1, 8) \cdot (1, 8, 2, 4, 2)$

Multiple Choice:

(a) *Exists.* ✓

(b) *Does not exist.*

Problem 104 *Compute the angle between the vectors $(2, 8)$ and $(-8, 2)$ in degrees.*

$$\theta = \boxed{90}^\circ$$

Problem 105 *Compute the angle between the vectors $(2, 1, 3, 1, 1)$ and $(-3, 1, 1, 1, -2)$ in degrees.*

$$\theta = \boxed{104.48}^\circ$$

(Give your answer as a positive number to two decimal places.

Problem 106 *Suppose you have vectors $\vec{v}$ and $\vec{w}$ such that $\|\vec{v}\| = 6$ and $\|\vec{w}\| = 2$, and the angle between $\vec{v}$ and $\vec{w}$ is $\dfrac{\pi}{4}$ radians. Compute the dot product of $\vec{v}$ and $\vec{w}$.*

$$\vec{v} \cdot \vec{w} = \boxed{6\sqrt{2}}$$

Problem 107 *Compute the projection of* $\vec{v} = (1, 7, 3)$ *onto* $\vec{w} = (-3, -2, 1)$.

$$\text{proj}_{\vec{w}}(\vec{v}) = \boxed{(7/2, 7/3, -7/6)}$$

Problem 108 *Compute the projection of the vector* $\vec{v} = (1, 2, 3)$ *onto the vector* $\vec{w} = (-3, 2, -1)$.

$$\text{proj}_{\vec{w}}(\vec{v}) = \boxed{(0, 0, 0)}$$

Problem 108.1 *Why does your answer make sense?*

Multiple Choice:

(a) *The vectors are parallel.*

(b) *The vectors are perpendicular.* ✓

(c) *They are the same length.*

Problem 109 *Compute the cross product.*

$$(1, 7, 3) \times (2, 8, -1) = \boxed{(-31, 7, -6)}$$

Problem 110 *Compute the cross product.*

$$(-1, 7, 2) \times (2, 8, 3) = \boxed{(5, 7, -22)}$$

Problem 111 *For each of the following, determine whether the expression exists or does not exist.*

$(-1, 7, 3) \times (1, 7, 2)$

Multiple Choice:

(a) *Exists.* ✓

(b) *Does not exist.*

$(-1, 7, 3, 8) \times (1, 7, 2, 0)$

Multiple Choice:

(a) *Exists.*

(b) *Does not exist.* ✓

$(2, 0, 0) \times (1, 7, 2, 0)$

Multiple Choice:

(a) *Exists.*

(b) *Does not exist.* ✓

$(0, 0, 0) \times (0, 0, 0)$

Multiple Choice:

(a) *Exists.* ✓

(b) *Does not exist.*

Problem 112 *Compute the area of the parallelogram determined by* $(1, 6)$ *and* $(1, 0)$.

$$\text{Area} = \boxed{6}$$

Problem 113 *Compute the volume of the parallelepiped determined by* $(1, 6, 2)$, $(-1, 2, 0)$, *and* $(0, 3, 1)$.

$$\text{Volume} = \boxed{2}$$

Problem 114 *Suppose $\vec{v}$ and $\vec{w}$ are unit vectors in the xy-plane, and we know that they are perpendicular. What is $\vec{v} \times \vec{w}$?*

Multiple Choice:

(a) $(0, 0, 1)$

(b) $(0, 0, -1)$

(c) *Not enough information.* ✓

Problem 115 *Find a parametrization $\vec{x}(t)$ of the line parallel to the vector $(1, 6, 3)$ and through the point $(1, 3, 2)$, such that $\vec{x}(0) = (1, 3, 2)$.* $\vec{x}(t) = \boxed{(t + 1, 6t + 3, 3t + 2)}$

Problem 116 *Find a parametrization $\vec{x}(s, t)$ of the plane containing vectors $(1, 6, 2)$ and $(1, 3, 2)$, and passing through the point $(1, 0, 0)$, such that $\vec{x}(0, 0) = (1, 0, 0)$ and $\vec{x}(1, 0) = (1, 6, 2)$.*

$$\vec{x}(s, t) = \boxed{(s + t + 1, 6s + 3t, 2s + 2t)}$$

Problem 117 *Give an equation which describes the plane perpendicular to the vector $(1, 7, 3)$ and through the point $(-2, 4, 1)$.*

$$0 = \boxed{(x + 2) + 7(y - 4) + 3(z - 1)}$$

Written Problems

Problem 118 *For any vector $\vec{v}$ in $\mathbb{R}^n$, prove that $\vec{v} \cdot \vec{v} = \|\vec{v}\|^2$.*

Problem 119 *Prove that vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^n$ are perpendicular if and only if $\text{proj}_{\vec{w}}(\vec{v})$ is the zero vector.*

Problem 120 *For any vector $\vec{v}$ in $\mathbb{R}^3$, prove that $\vec{v} \times \vec{v}$ is the zero vector.*

Practice Problems: Coordinate Systems

Online Problems

Problem 121 *Find the Cartesian coordinates of each point, which is given in polar coordinates.*

$(r, \theta) = (2, \pi/6)$

$$(x, y) = \boxed{(\sqrt{3}, 1)}$$

$(r, \theta) = (\sqrt{2}, 3\pi/4)$

$$(x, y) = \boxed{(-1, 1)}$$

$(r, \theta) = (1, 0)$

$$(x, y) = \boxed{(1, 0)}$$

$(r, \theta) = (3, \pi)$

$$(x, y) = \boxed{(-3, 0)}$$

Problem 122 *Find the polar coordinates of each point, which is given in Cartesian coordinates. Your answers should satify $0 \le r$ and $0 \le \theta < 2\pi$.*

$(x, y) = (\sqrt{3}, 0)$

$$(r, \theta) = \boxed{(\sqrt{3}, 0)}$$

$(x, y) = (0, 2)$

$$(r, \theta) = \boxed{(2, \pi/2)}$$

$(x, y) = (-1, -1)$

$$(r, \theta) = \boxed{(\sqrt{2}, 5\pi/4)}$$

$(x, y) = (1, \sqrt{3})$

$$(r, \theta) = \boxed{(2, \pi/3)}$$

Learning outcomes:
Author(s):

Problem 123 *Find the Cartesian coordinates of the point* $(\pi/2, \pi, 2)$, *given in cylindrical coordinates.*

$$(x, y, z) = \boxed{(0, -\pi/2, 2)}$$

Problem 124 *Find cylindrical coordinates for the point* $(0, -1, 3)$, *written in Cartesian coordinates. Your answer should satisfy* $0 \leq r$ *and* $0 \leq \theta < 2\pi$.

$$(r, \theta, z) = \boxed{(1, \pi, 3)}$$

Problem 125 *Consider the surface described in Cartesian coordinates by*

$$2z^2 = x^2 + y^2.$$

Describe this surface with an equation in cylindrical coordinates, of the form $0 = f(r, \theta, z)$.

$$0 = \boxed{r^2 - 2z^2}$$

FIGURE OUT HOW TO HANDLE THIS!!! What type of shape is this?

Multiple Choice:

(a) *Plane*

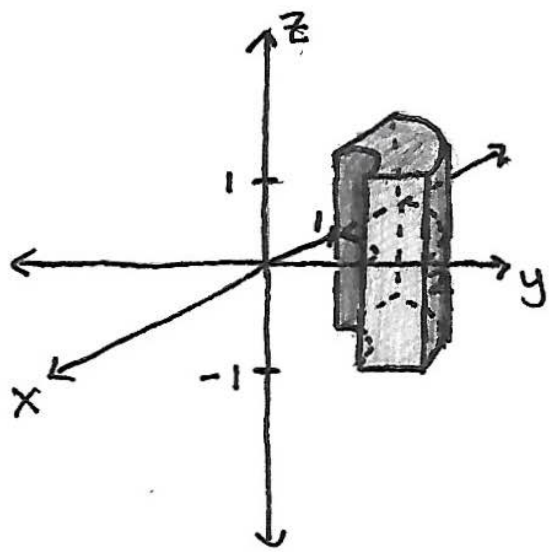
(b) *Cylinder*

(c) *Sphere*

(d) *Cone* ✓

(e) *Other*

Problem 126 *Consider the following region in* $\mathbb{R}^3$.



This region is the set of points (r, θ, z), in cylindrical coordinates, satisfying the inequalities

$$\boxed{1} \leq r \leq \boxed{2}$$
$$\boxed{\pi/2} \leq \theta \leq \boxed{\pi}$$
$$\boxed{-1} \leq z \leq \boxed{1}$$

Problem 127 *For each of the following equations in cylindrical coordinates, select the type of shape they define.*

FIGURE OUT CORRECT ANSWERS

$r = \cos\theta$

Multiple Choice:

(a) *plane*

(b) *cylinder*

(c) *sphere*

(d) *other*

$z = r\cos\theta$

Multiple Choice:

(a) *plane*

(b) *cylinder*

(c) *sphere*

(d) *other*

$z = -r$

Multiple Choice:

(a) *plane* ✓

(b) *cylinder*

(c) *sphere*

(d) *other*

Problem 128 *Find the Cartesian coordinates of each point, given in cylindrical coordinates.*

$(r, \theta, z) = (1, 1, 1)$

$$(x, y, z) = \boxed{(\cos(1), \sin(1), 1)}$$

$(r, \theta, z) = (\pi, \pi, \pi)$

$$(x, y, z) = \boxed{(-\pi, 0, \pi)}$$

$(r, \theta, z) = (2, 4\pi/3, -2)$

$$(x, y, z) = \boxed{(-\sqrt{3}, -1, -2)}$$

Problem 129 *Find the cylindrical coordinates of each point, given in Cartesian coordinates. Your answers should satisfy $r \geq 0$ and $0 \leq \theta < 2\pi$.*

$(x, y, z) = (1, 1, 1)$

$$(r, \theta, z) = \boxed{(\sqrt{2}, \pi/4, 1)}$$

$(x, y, z) = (\pi, \pi, \pi)$

$$(r, \theta, z) = \boxed{(\sqrt{2}\pi, \pi/4, \pi)}$$

$(x, y, z) = (2, 2\sqrt{3}, -2)$

$$(r, \theta, z) = \boxed{(4, \pi/6, -2)}$$

Problem 130 *Find the Cartesian coordinates of the point $(2, \pi, \pi/2)$, given in spherical coordinates.*

$$(x, y, z) = \boxed{(-2, 0, 0)}$$

Problem 131 *Find spherical coordinates for the point $\left(-\sqrt{2}, \sqrt{2}, 2\sqrt{3}\right)$, written in Cartesian coordinates. Your answer should satisfy $0 \leq \rho$, $0 \leq \theta \leq 2\pi$, and $0 \leq \phi \leq \phi$.*

$$(\rho, \theta, \phi) = \boxed{(4, 3\pi/4, \pi/6)}$$

Problem 132 *Consider the surface described in Cartesian coordinates by*

$$2z^2 = x^2 + y^2.$$

Describe this surface with an equation in spherical coordinates, of the form $0 = f(\rho, \theta, \phi)$.

$$0 = \boxed{\rho^2 \sin^2 \phi - 2 \cos^2 \phi}$$

FIGURE OUT HOW TO HANDLE THIS!!! What type of shape is this?

Multiple Choice:

(a) *Plane*

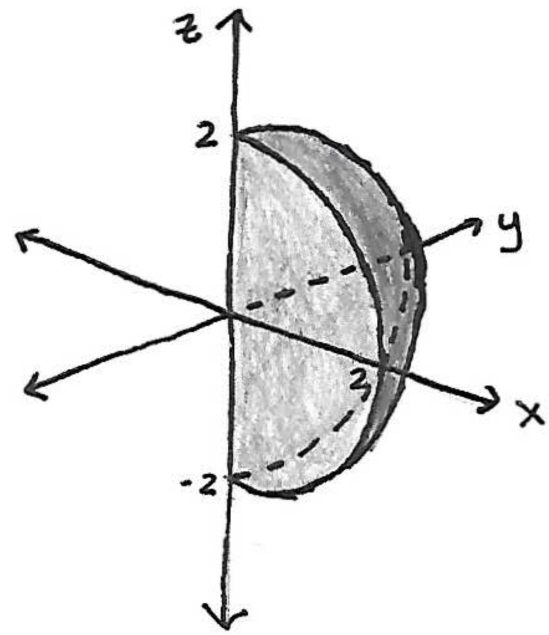
(b) *Cylinder*

(c) *Sphere*

(d) *Cone* ✓

(e) *Other*

Problem 133 *Consider the following region in $\mathbb{R}^3$.*



This region is the set of points (ρ, θ, ϕ), in spherical coordinates, satisfying the inequalities

$$\boxed{0} \le \rho \le \boxed{2}$$
$$\boxed{0} \le \theta \le \boxed{pi/2}$$
$$\boxed{0} \le \phi \le \boxed{pi}$$

Problem 134 *For each of the following equations in spherical coordinates, select the type of shape they define.*

FIGURE OUT CORRECT ANSWERS

$\rho = \cos\phi$

Multiple Choice:

(a) *plane*

(b) *cylinder*

(c) *sphere*

(d) *other*

$\rho = \sin\theta$

Multiple Choice:

(a) *plane*

(b) *cylinder*

(c) *sphere*

(d) *other*

$\rho\cos\theta\sin\phi = 1$

Multiple Choice:

(a) *plane* ✓

(b) *cylinder*

(c) *sphere*

(d) *other*

Problem 135 *Find the spherical coordinates of each point, given in Cartesian coordinates. Your answers should satisfy* $0 \le r$, $0 \le \theta < 2\pi$, *and* $0 \le \phi \le \pi$.

$(x, y, z) = (1, 1, 1)$

$$(\rho, \theta, \phi) = \boxed{(\sqrt{3}, \pi/4, \pi/4)}$$

$(x, y, z) = (1, -1, -1)$

$$(\rho, \theta, \phi) = \boxed{(\sqrt{3}, 3\pi/4, 3\pi/4)}$$

$(x, y, z) = (1, \sqrt{3}, 0)$

$$(\rho, \theta, \phi) = \boxed{(\sqrt{4}, \pi/6, \pi/2)}$$

Problem 136 *Find the Cartesian coordinates of each point, given in spherical coordinates.*

$(\rho, \theta, \phi) = (\pi, \pi, \pi)$

$$(x, y, z) = \boxed{(0, 0, -\pi)}$$

$(\rho, \theta, \phi) = (3, \pi/2, \pi/4)$

$$(x, y, z) = \boxed{(0, 3\sqrt{2}/2, 3\sqrt{2}/2)}$$

$(\rho, \theta, \phi) = (2, 7\pi/6, 3\pi/4)$

$$(x, y, z) = \boxed{(-\sqrt{6}/2, -\sqrt{2}/2, -\sqrt{2})}$$

Written Problems

Problem 137 *For several values of the constant a, sketch the graph of the curve in $\mathbb{R}^2$ given by the polar equation*

$$r = a \sin\theta.$$

What do you notice about these curves?

Problem 138 *Which points in $\mathbb{R}^2$ have the same coordinates when written in Cartesian and polar coordinates? (That is, for what points do we have $x = r$ and $y = \theta$?)*

Problem 139 *Consider the surface described by $(r-3)^2+z^2 = 1$ in cylindrical coordinates, with the restriction $r \geq 0$.*

(a) *Sketch the intersection of the surface with the half-plane $\theta = 0$.*

(b) *Sketch the intersection of the surface with the half-plane $\theta = \dfrac{\pi}{2}$.*

(c) *Sketch the intersection of the surface with the plane $z = 0$.*

(d) *Sketch the surface.*

Problem 140 *Sketch the region in $\mathbb{R}^3$ with cylindrical coordinates satisfying the inequality*

$$r \leq z \leq 4 - 2r$$

Problem 141 *Convert the following equation, given in cylindrical coordinates, into Cartesian coordinates.*

$$r = 0$$

Problem 142 *Sketch the region in* $\mathbb{R}^3$ *given by*

$$r \leq z \leq 3$$
$$0 \leq \theta \leq \pi/2$$

in cylindrical coordinates.

Problem 143 *Sketch the region in* $\mathbb{R}^3$ *given by*

$$r^2 - 2 \leq z \leq 2 - r^2$$

in cylindrical coordinates.

Problem 144 *Which points in* $\mathbb{R}^3$ *have the same coordinates when written in Cartesian and cylindrical coordinates?*

Problem 145

(a) *Given a function* f*, consider the graphs of the equations* $r = f(\theta)$ *and* $r = 2f(\theta)$*, in polar coordinates. How are these graphs related?*

(b) *Given a function* f*, consider the graphs of the equations* $\rho = f(\theta, \phi)$ *and* $\rho = 2f(\theta, \phi)$*, in spherical coordinates. How are these graphs related?*

(c) *Given a function* f*, consider the graphs of the equations* $r = f(\theta)$ *and* $r = -f(\theta)$*, in polar coordinates. How are these graphs related?*

(d) *Given a function* f*, consider the graphs of the equations* $\rho = f(\theta, \phi)$ *and* $\rho = -f(\theta, \phi)$*, in spherical coordinates. How are these graphs related?*

Problem 146 *Sketch the surface in* $\mathbb{R}^3$ *given by the equation*

$$1 - \cos\phi$$

in spherical coordinates.

Problem 147 *Consider the surface in $\mathbb{R}^3$ given by the equation*

$$\rho \sin\phi \sin\theta = 1$$

in spherical coordinates.

Convert this equation to Cartesian coordinates and cylindrical coordinates, and sketch the surface.

Problem 148 *Consider the region in $\mathbb{R}^3$ consisting of points whose spherical coordinates satisfy*

$$1 \leq \rho \leq 3$$

Sketch this region.

Problem 149 *Consider the region in $\mathbb{R}^3$ consisting of points whose spherical coordinates satisfy*

$$0 \leq \phi \leq \pi/2$$
$$0 \leq \rho \leq 1$$

Sketch this region.

Problem 150 *Consider the region in $\mathbb{R}^3$ consisting of points whose spherical coordinates satisfy*

$$\cos\phi \leq \rho \leq 2$$

Sketch this region.

Problem 151 *Which points in $\mathbb{R}^3$ have the same coordinates when written in Cartesian and spherical coordinates?*
