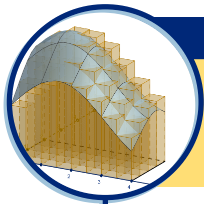
Advanced Calculus For Data Science

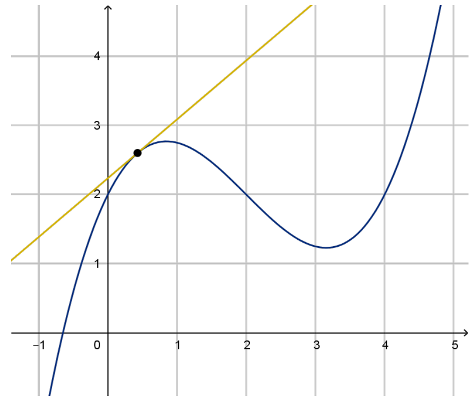
Mike Carr

Contents

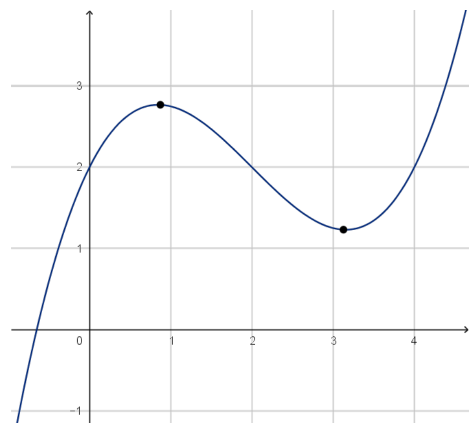


Introduction

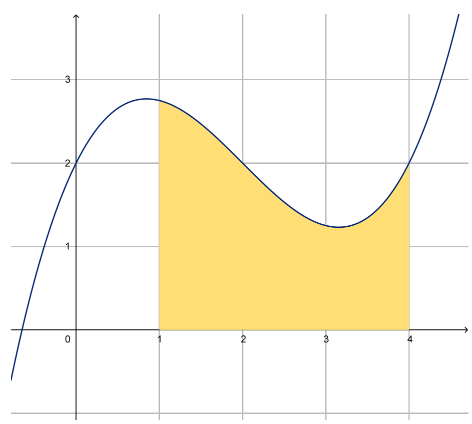
So far in calculus you have developed the tools to answer the following questions about a function of one variable:



1. How quickly does the value of the function change as the input changes?
2. How do we estimate the value of the function near a point?



3. What are the maximum and minimum values of the function?



4. What is the area under the graph of the function? What does it mean?

These are all useful tools, but they don't necessarily apply to the types of data that we encounter in the world.

Data generally takes the form of a set of observations, rather than an algebraic function. How do we perform calculus with such a set? We cannot integrate it without an antiderivative. In some cases, the best functions to model our data are difficult to work with. We take for granted that $\sin x$ is a useful function, but how do we even evaluate a quantity like $\sin(7.52)$? In all these circumstances, the best we can do is approximate. We will develop methods to approximate integrals and to approximate functions.

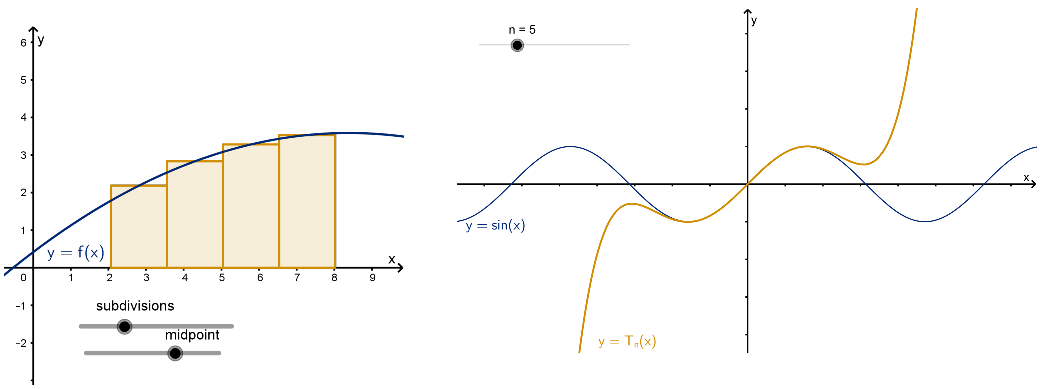


Figure: Approximations of an integral and of a function

Many measurable quantities can be found to depend on the value of multiple inputs. These are multivariable functions like $z = F(x, y)$, where z is a function of two independent variables. Examples appear in all the sciences

1. Chemistry: $V = \dfrac{nrt}{P}$
2. Physics: $F = \dfrac{GMm}{r^2}$
3. Economics: $P = P_0 e^{rt}$

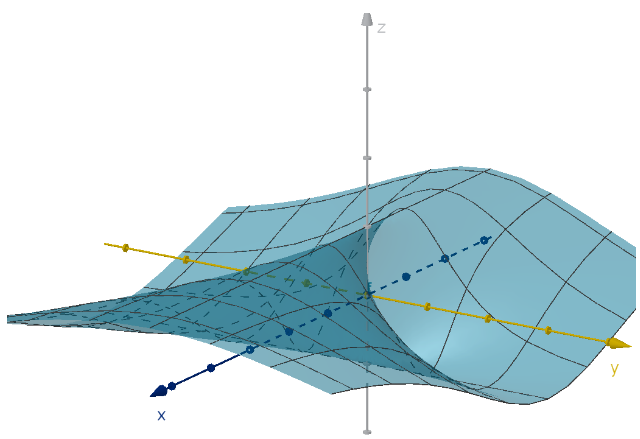


Figure: The graph of a two-variable function

We want to understand how to measure rates of change of these functions, and what these measurements can do for us.

Furthermore, real world data does not come prepackaged with a differentiable function to describe it. One approach is to find a line of best fit. Doing so requires optimizing two variables at once (slope and intercept) to find the best fit.

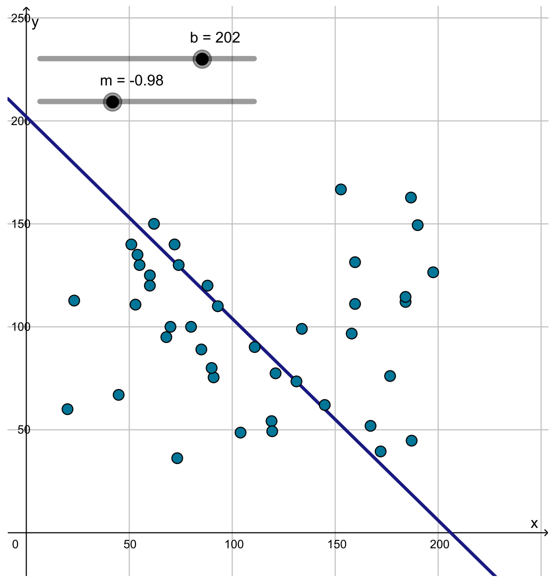


Figure: Fitting a line to a set of data points

The values of y may not be a function of x at all. Another view point is to see (x, y) as a randomly chosen point in the plane. To model such random choices, we use a two-variable density function. Volumes under its graph (computed by integrals) tell us where these random points are likely to lie.

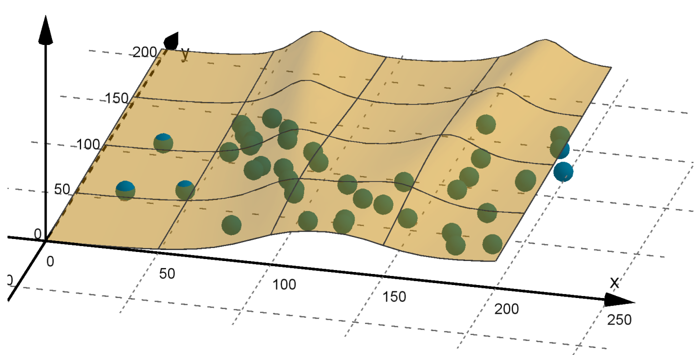


Figure: A function that models the outcomes of a random process

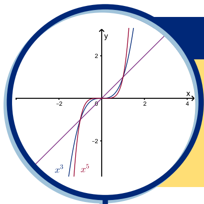
These approaches will requires us to use derivatives and integrals of multivariable functions.

Chapter 1

Review of Algebra and Calculus

This chapter reviews the most important information about functions, limits, derivatives, and integrals. It is not meant to teach this material to a first-time learner, but can serve as a reference or reminder.

Contents



Section 1.1

Graphs of Functions

Goals:

1. Graph algebraic and trigonometric functions.
2. Solve equations using inverse functions.
3. Solve equations containing quotients.
4. Graph transformations of functions.

Definition

The **graph** of an equation is the set of ordered pairs (x, y) that satisfy the equation. These are the points that, when their coordinates are plugged in for x and y, the two sides of the equation are equal.

Linear Functions

Linear functions can be written in **slope-intercept form**:

$$f(x) = mx + b.$$

- The graph $y = mx + b$ of a linear function is a line.
- m is the slope, which is the change in y over the change in x between any two points on the line.
- $(0, b)$ is the y intercept.
- If we have the slope and a known point (x_0, y_0) on a line. We can write its equation in **point-slope** form.
$$y - y_0 = m(x - x_0)$$
- If we have both the x- and y-intercepts of the line, it is convenient to write it in **normal form**
$$ax + by + c = 0$$

Monomials

A **monomial** is a function of the form:

$$f(x) = x^n$$

where n is an integer greater than 0.

- For $n \geq 2$ the graph $y = x^n$ curves upward over the positive values of x.
- Higher values of n have lower values when $0 < x < 1$ but higher values after $x > 1$.
- For even values of n the graph is symmetric across the y-axis, curving up when x is negative.
- For odd values of n the graph curves down when x is negative. It is **anti-symmetric** across the $x = 0$.

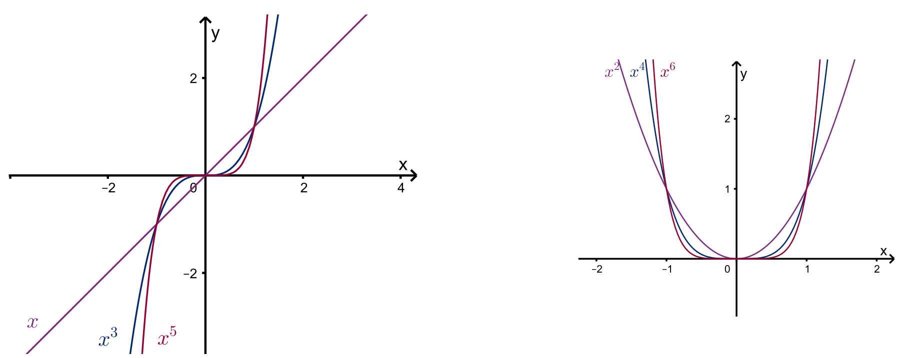


Figure: Graphs of monomials of odd and even powers

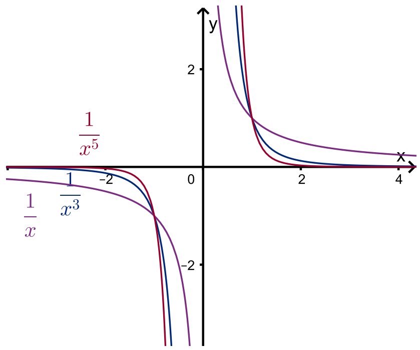
Monomials of Negative Power

Monomials of negative power have the form $f(x) = x^{-n}$. They are also commonly written

$$f(x) = \frac{1}{x^n}$$

.

- The graph $y = \frac{1}{x^n}$ has a vertical asymptote at $x = 0$.
- The graph approaches the x-axis, $y = 0$ as x gets large.
- For even values of n, the graph is above the x-axis.
- For odd values of n, the graph is above the x-axis for positive x and below it for negative x.
- A larger choice of n makes the function approach the x-axis more quickly.



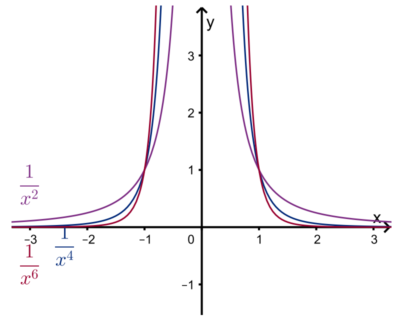


Figure: Graphs of monomials of negative odd and even powers

Roots

A **root functiom** is a function of the form:

$$f(x) = \sqrt[n]{x}$$

where n is an integer greater than 0.

- The domain of $\sqrt[n]{x}$ is $[0, \infty)$ if n is even and all real numbers if n is odd.
- The x and y intercept of $y = \sqrt[n]{x}$ is at $(0, 0)$.
- Root functions are increasing. At $x = 0$, they travel straight up.

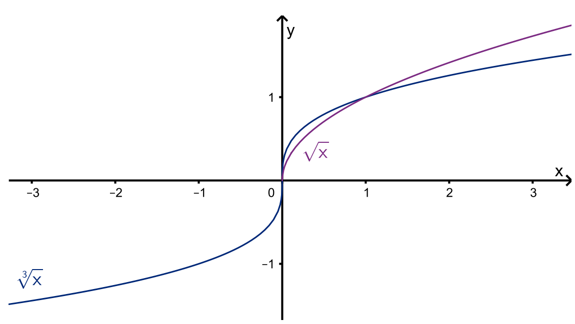


Figure: The graphs of $y = \sqrt{x}$ and $y = \sqrt[3]{x}$

Exponential Functions

An **exponential** function has the form:

$$f(x) = a^x$$

where a is a number greater than 0.

- a is called the **base** of the exponential function.
- The graph $y = a^x$ passes through $(0, 1)$.
- If $a > 1$ then $f(x)$ increases quickly as x takes on positive values. Higher values of a give a steeper increase. $f(x)$ approaches 0 as x goes to $-\infty$. Higher values of a give a faster approach. The graph does not touch or cross the x-axis.
- If $a < 1$, then the above is reversed.
- e is a commonly used base. e is approximately 2.718.

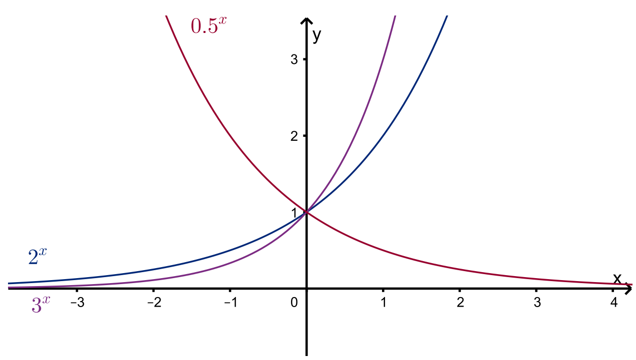


Figure: The graphs of exponential functions

Logarithms

A **logarithmic function** have the form:

$$f(x) = \log_a^x$$

where a is a number greater than 1. $\log_a x$ is the number b such that $a^b = x$.

- a^b can never be 0 or less. The domain of $f(x) = \log_a^x$ is $(0, \infty)$.
- As x goes to 0, $\log_a x$ goes to $-\infty$.
- $y = \log_a x$ has an x intercept at $(1, 0)$.

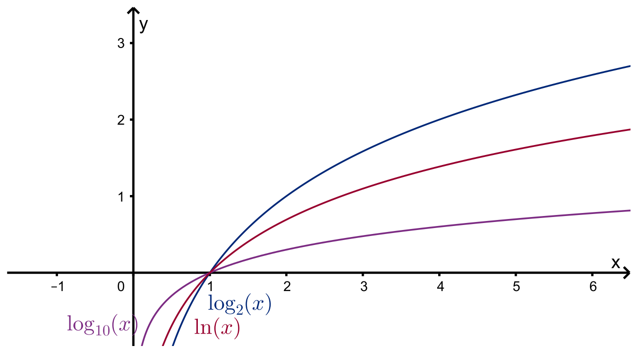


Figure: The graphs of logarithm functions

Logarithms and exponents are inverse functions. We solve exponential equations by applying a logarithm to both sides. We solve logarithm equations by exponentiating both sides.

$$a^x = c \longrightarrow x = \log_a c$$
$$\log_a x = c \longrightarrow x = a^c$$

Trigonometric Functions

$f(x) = \sin x$ and $f(x) = \cos x$ are periodic functions.

- $\sin x$ and $\cos x$ have a range of $[-1, 1]$.
- These functions are periodic. This means that for all x, $f(x + 2\pi) = f(x)$.

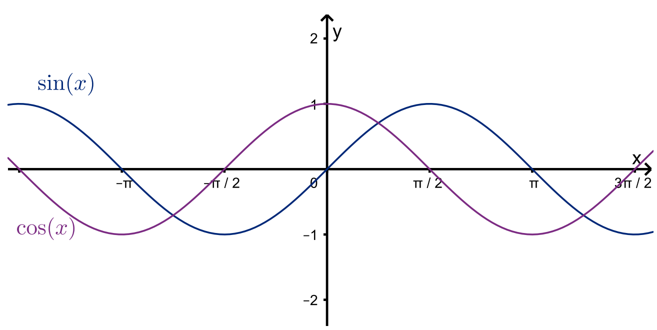


Figure: The graphs of $y = \sin x$ and $y = \cos x$

The other trigonometric functions can be written in terms of sine and cosine.

$$\tan x = \frac{\sin x}{\cos x}$$
$$\cot x = \frac{\cos x}{\sin x}$$
$$\sec x = \frac{1}{\cos x}$$
$$\csc x = \frac{1}{\sin x}$$

Since trigonometric functions obtain the same values infinitely many times, the do not technically have inverse functions. However, we define inverse trigonometric functions on a restricted range.

$$-\frac{\pi}{2} \leq \sin^{-1} x \leq \frac{\pi}{2}$$
$$0 \leq \cos^{-1} x \leq \pi$$
$$-\frac{\pi}{2} \leq \tan^{-1} x \leq \frac{\pi}{2}$$

These functions provide one solution to a trigonometric equation. We can obtain the others by using the periodic behavior of trognometric funtions.

$$\sin x = c \longrightarrow x = \sin^{-1} c + 2\pi n \text{ or } \pi - \sin^{-1} c + 2\pi n$$
$$\cos x = c \longrightarrow x = \cos^{-1} c + 2\pi n \text{ or } -\cos^{-1} c + 2\pi n$$
$$\tan x = c \longrightarrow x = \tan^{-1} c + \pi n$$

Where n can be any integer.

Question 1.1.1

How Do Transformations Affect the Graph of a Function?

Transformations

Suppose we would like to transform the graph $y = f(x)$. Here are four ways we can.

- The graph of $y = af(x)$ is stretched by a factor of a in the y direction.
- The graph of $y = f(x) + b$ is shifted by b in the positive y direction.
- The graph of $y = f(cx)$ is compressed by a factor of c in the x direction.
- The graph of $y = f(x + d)$ is shifted by d in the negative x direction.

Question 1.1.1 How Do Transformations Affect the Graph of a Function?

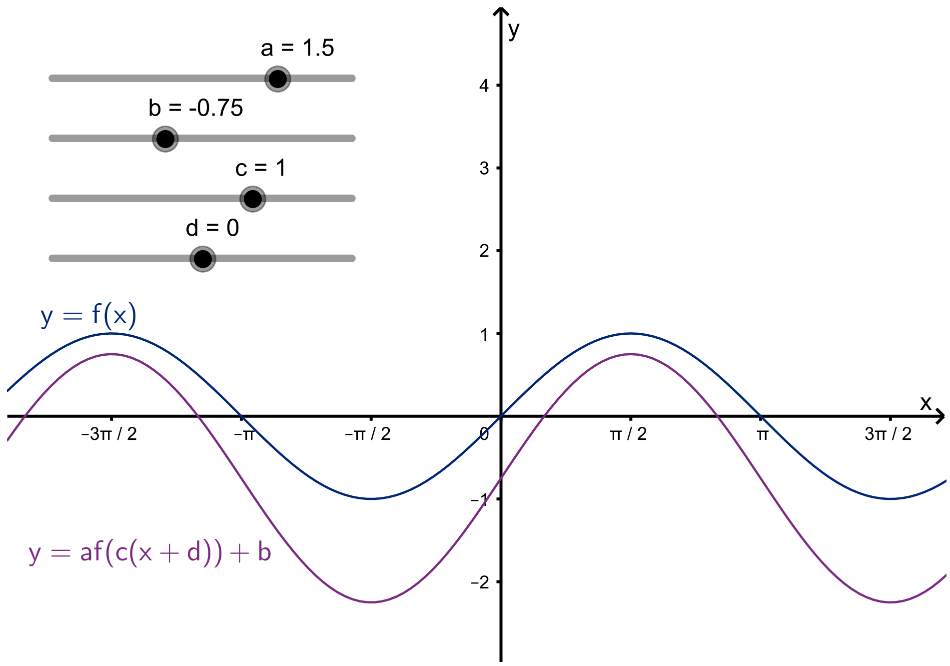


Figure: The graphs of $y = f(x)$ and it transformation

Example 1.1.2
A Equation with Quotients

An equation of the form $\frac{f(x)}{g(x)} = 0$ is satisfied whenever $f(x) = 0$ but $g(x) \neq 0$.

Example

Solve

$$\frac{2x^2 - 3x - 5}{x^2 + 3x + 2} = 0$$

Solution

$$\begin{aligned} 2x^2 - 3x - 5 &= 0 && \text{set numerator} = 0 \\ (2x-5)(x+1) &= 0 && \text{factor} \end{aligned}$$

$$x = \frac{5}{2} \text{ or } x = -1$$

Then we must check that neither of these causes the denominator to be 0.

$$\left(\frac{5}{2}\right)^2 + 3\left(\frac{5}{2}\right) + 2 = \frac{63}{4} \qquad (-1)^2 + 3(-1) + 2 = 0$$

So $x = \frac{5}{2}$ is the only solution.

If there are terms besides the quotient, move them all to the same side of the equation and use a common denominator to combine them.

Example

Solve

$$2 + \frac{x+3}{x+1} = \frac{4}{x}$$

Solution

$$\begin{aligned} 2 + \frac{x+3}{x+1} - \frac{4}{x} &= 0 && \text{move to one side} \\ \frac{2x^2+2x}{x^2+x} + \frac{x^2+3x}{x^2+x} - \frac{4x+4}{x^2+x} &= 0 && \text{common denominator} \\ \frac{3x^2+x-4}{x^2+x} &= 0 && \text{combine} \\ \text{set } 3x^2 + x - 4 &= 0 \\ (3x+4)(x-1) &= 0 && \text{factor} \end{aligned}$$

$$x = -\frac{4}{3} \text{ or } x = 1$$

Then we must check that neither of these causes the denominator to be 0.

$$\left(-\frac{4}{3}\right)^2 + \left(-\frac{4}{3}\right) = \frac{4}{9} \qquad 1^2 + 1 = 2$$

Both solutions are valid. $x = -\frac{4}{3}$ or $x = 1$.

1.1

Q1 Simplify $\left(5^2 5^4\right)^3$

Q2 Simplify $e^5(e^4)^3$

Q3 Compress $2\log_5 x + \log_5 y - 3\log_5 z$ into a single logarithm.

Q4 Compress $3\ln(x+y) - \ln(x^2+2xy+y^2)$ into a single logarithm.

Q5 Solve $2e^x - 7 = 22$

Q6 Solve $4\cos(2x) = 1$

Q7 Solve $2\sin^2 x - 1 = 0$

Q8 Solve $2\ln(x-5) = 16$

Q9 Solve $4^{3x-2} = 15$

Q10 Solve $\log_7(x^2+5) - 3 = 11$

1.1.1

Q11 Graph $y = 3\sin(2x)$.

Q12 Graph $y = -\ln x + 5$.

Q13 Graph $y = e^x - 4$.

Q14 Graph $y = \sqrt[3]{x+3}$.

Q15 Graph $y = \dfrac{1}{(x-2)^2}$.

Q16 Graph $y = -2\sqrt{x+1} + 4$.

1.1.2

Q17 Solve for x: $\dfrac{x^2+5x-6}{x-1} = 0$

Q18 Solve for x: $\dfrac{e^x-2}{x^2+2x-3} = 0$

Q19 Solve for x: $\dfrac{3x^2-5}{2e^x-7} = 0$

Q20 Solve for t: $\dfrac{\ln t-4}{3-t} = 0$

Q21 Solve for x: $\dfrac{\ln x-4}{3-x} = 0$

Q22 Solve for x: $\dfrac{3}{x+2} = \dfrac{7}{x+4}$

Q23 Solve for u: $\dfrac{5}{(u+1)^2} = \dfrac{u}{u+1}$

Section 1.2

Limits and Derivatives

Goals:

1. Compute limits of functions.
2. Verify that a function is continuous.
3. Compute derivatives.
4. Use derivatives to understand graphs and vice versa.

Question 1.2.1

What Is a Limit?

The Limit of a Function

- If we can make $f(x)$ arbitrarily close to some number L by considering only x in a small interval $(a, a+\delta)$ then we say the **limit** of f as x approaches a **from the right** is L. We write:

$$\lim_{x \to a^+} f(x) = L$$

  If $f(x)$ cannot be made arbitrarily close to any number, then this limit **does not exist**.

- Similarly, if we can make $f(x)$ arbitrarily close to some number L by considering only x in a small interval $(a-\delta, a)$ then we say the **limit** of f as x approaches a **from the left** is L. We write:

$$\lim_{x \to a^-} f(x) = L$$

  If $f(x)$ cannot be made arbitrarily close to any number, then this limit **does not exist**.

- If both $\lim\limits_{x \to a^+} f(x) = L$ and $\lim\limits_{x \to a^-} f(x) = L$, we say the **two-sided limit** or just **limit** of f as x approaches a is L. We write

$$\lim_{x \to a} f(x) = L$$

  If the either the limit from the left or the limit from the right does not exist, or if they do exist but are not equal to each other, then the two sided limit does not exist.

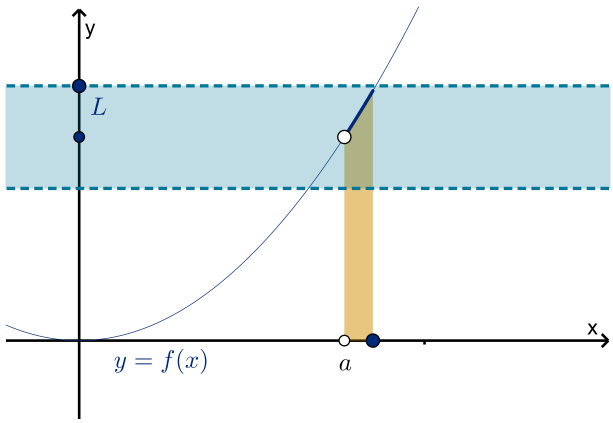


Figure: An interval of x values that produce values in a small neighborhood of L when plugged into $f(x)$.

Infinite Limits

If $f(x)$ can be made arbitrarily large by considering only x in a small interval $(a, a+\delta)$ then we say the limit of f as x approaches a from the right is ∞.

$$\lim_{x \to a^+} f(x) = \infty$$

This is a way of representing growth without bound. Infinite limits from the left are defined analogously. Also analogous is our treatment of a function then decreases without bound. We say these functions limit to $-\infty$. If either one-sided limit at $x = a$ is infinite, then the line $x = a$ is a **vertical asymptote** of $y = f(x)$.

Example

Let $f(x) = \frac{1}{x}$.

- $\lim_{x \to 0^+} f(x) = \infty$
- $\lim_{x \to 0^-} f(x) = -\infty$

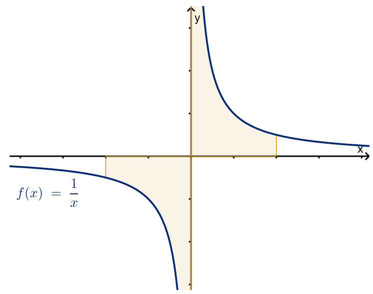


Figure: The graph of $y = \frac{1}{x}$

Vertical Asymptotes

There are only two common algebraic constructions that produce infinite limits.

- A function of the form $\frac{f(x)}{g(x)}$ where $\lim_{x \to a} g(x) = 0$ and $\lim_{x \to a} f(x) \neq 0$.
- $\lim_{x \to 0^+} \log_a x = -\infty$.

Remark

∞ is not a number, so if $\lim_{x \to a^+} f(x) = \infty$ we would still say that $\lim_{x \to a^+} f(x)$ does not exist.

There are several limit laws that allow us to compute limits of combinations of simpler functions.

Theorem [Limit Laws]

The following hold limits, provided that $\lim_{x \to a} f(x)$ and $\lim_{x \to a} g(x)$ exist.

- $\lim_{x \to a} (f(x) + g(x)) = \lim_{x \to a} f(x) + \lim_{x \to a} g(x)$
- $\lim_{x \to a} (cf(x)) = c \lim_{x \to a} f(x)$
- $\lim_{x \to a} (f(x)g(x)) = \left(\lim_{x \to a} f(x)\right)\left(\lim_{x \to a} g(x)\right)$
- $\lim_{x \to a} \left(\frac{f(x)}{g(x)}\right) = \left(\frac{\lim_{x \to a} f(x)}{\lim_{x \to a} g(x)}\right)$ provided that $\lim_{x \to a} g(x) \neq 0$
- $\lim_{x \to a} f(g(x)) = \lim_{x \to b} f(x)$ provided that $\lim_{x \to a} g(x) = b$

We can write similar statements for one-sided limits, though we need to be careful about directions in the composition rule.

Question 1.2.2

What is Continuity?

Definition

A function $f(x)$ is **continuous** at a, if

$$\lim_{x \to a} f(x) = f(a)$$

Remark

This definition is useful, if we already know we are dealing with a continuous function. For example $f(x) = \sin x$ is continuous so

$$\lim_{x \to \frac{\pi}{6}} \sin x = \sin \frac{\pi}{6} = \frac{1}{2}$$

Fortunately, many familiar functions are continuous.

Theorem

The following functions are continuous on their domains

1. Constant functions
2. Linear functions
3. Polynomials
4. Roots
5. Exponential functions
6. Logarithms
7. Trigonometric functions
8. $f(x) = |x|$

More complex functions made from continuous functions are also continuous.

Question 1.2.2 What is Continuity?

Theorem

If $f(x)$ and $g(x)$ are continuous on their domains, and c is a constant, then the following are also continuous on their domains

1. $f(x) + g(x)$
2. $f(x) - g(x)$
3. $f(x)g(x)$
4. $\frac{f(x)}{g(x)}$ (note that any x where $g(x) = 0$ is not in the domain)
5. $f(x)^{g(x)}$ as long as $f(x) > 0$
6. $f(g(x))$

Remark

Putting the above theorems together, we see that just about any function we can write using algebraic and trigonometric expressions is continuous on its domain. This does not mean it is continuous everywhere. $f(x) = \frac{1}{x}$ is not continuous at $x = 0$, for example.

Example 1.2.3

Computing a Limit

How do we compute $\lim\limits_{x\to 3} \dfrac{x^2 - 7x + 12}{x - 3}$?

Solution

$f(x) = \frac{x^2-7x+12}{x-3}$ is continuous on its domain, but $x = 3$ is not in the domain. However, let $g(x) = x-4$. We know $\frac{x^2-7x+12}{x-3} = x - 4$ for every x except $x = 3$. Specifically, in any neighborhood around $x = 3$, $f(x) = g(x)$ so they have the same limit.

$$\begin{aligned}
\lim_{x\to 3} \frac{x^2 - 7x + 12}{x - 3} &= \lim_{x\to 3} x - 4 && \text{because they agree around } x = 3 \\
&= 3 - 4 && \text{because } g(x) = x - 4 \text{ is continuous at } x = 3 \\
&= -1
\end{aligned}$$

Question 1.2.4

What Is the Intermediate Value Theorem?

One early intuition for continuity is that the graph of the function can be drawn without any breaks. There are many ways to formalize this idea. One of the most important is the following theorem.

Theorem [The Intermediate Value Theorem]

If f is a continuous function on $[a, b]$ and K is a number between $f(a)$ and $f(b)$, then there is some number c between a and b such that $f(c) = K$.

This theorem essentially states that a continuous graph cannot get from one side of the line $y = K$ to the other without intersecting $y = K$. Notice that this theorem does not say exactly where this intersection must occur, only that it must occur somewhere in the interval (a, b). It also does not rule out the possibility of more than one such c existing.

Example

Show that $f(x) = e^x - 3x$ has a root between 0 and 1.

Solution

A root is a number c such that $f(c) = 0$. To prove such a root exists, we check the conditions of the IVT.

- $f(x)$ is a sum of continuous functions, so it is continuous on its domain.
- $f(0) = 1$
- $f(1) = e - 3 < 0$
- 0 is between $f(0)$ and $f(1)$

We conclude there is some c between 0 and 1 such that $f(c) = 0$.

Question 1.2.5
What Is a Limit at Infinity?

Definition

- If we can make $f(x)$ arbitrarily close to some number L by considering only x in some interval (n, ∞) then we say the **limit** of f as x approaches ∞ is L. We write:

$$\lim_{x\to\infty} f(x) = L$$

  If $f(x)$ cannot be made arbitrarily close to any number, then this limit **does not exist**.

- Similarly if we can $f(x)$ arbitrarily close to L by considering only x in some interval $(-\infty, n)$ then we say the **limit** of f as x approaches $-\infty$ is L. We write:

$$\lim_{x\to-\infty} f(x) = L$$

- If either $\lim_{x\to\infty} f(x) = L$ or $\lim_{x\to-\infty} f(x) = L$, then $y = L$ is a **horizontal aysmptote** of the graph $y = f(x)$.

By observing graphs or using arithmetic intuition, we arrive at the following limits at infinity.

$f(x)$	$\lim_{x\to\infty} f(x)$	$\lim_{x\to-\infty} f(x)$	Comments
x^n (n odd)	∞	$-\infty$	$n > 0$
x^n (n even)	∞	∞	$n > 0$
$\sqrt[n]{x}$ (n odd)	∞	DNE	domain is $x \geq 0$
$\sqrt[n]{x}$ (n even)	∞	$-\infty$	
$\frac{1}{x^n}$	0	0	$n > 0$
a^x ($a > 1$)	∞	0	
a^x ($0 < a < 1$)	0	∞	
$\log_a x$	∞	DNE	$a > 1$, domain is $x > 0$
$\sin x$	DNE	DNE	oscillates
$\tan^{-1} x$	$\frac{\pi}{2}$	$-\frac{\pi}{2}$	

Question 1.2.6

How Do We Measure the Change in a Function?

Definition

The **average rate of change** of a function $f(x)$ between $x = a$ and $x = b$ is

$$\frac{f(b) - f(a)}{b - a}$$

This is also the slope of the **secant line** from $(a, f(a))$ to $(b, f(b))$ on the graph $y = f(x)$.

Knowing the average rate of change over a range of inputs (or times) doesn't tell us the rate of change at a specific point (or moment). Geometrically the is the slope of the **tangent line** to $y = f(x)$ at a particular point $(a, f(a))$

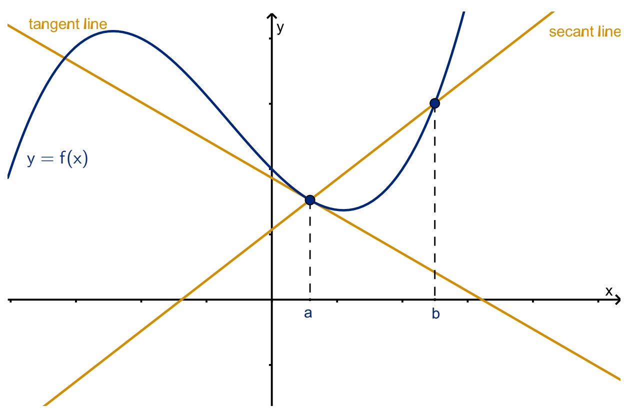


Figure: A secant line and a tangent line

The secant lines get closer and closer to the tangent line (in slope) as b gets closer to a. This suggests that we could take the limit of these approaching values to get the actual slope.

Definition

The **instantaneous rate of change** or **derivative** of a function $f(x)$ at $x = a$ is

$$\lim_{h \to 0} \frac{f(a + h) - f(a)}{h}$$

provided that this limit exists. This is also the slope of the tangent line to $y = f(x)$ at $(a, f(a))$. Two common notations for the derivative are

- Prime notation: $f'(a)$
- Leibniz notation: $\left.\frac{df}{dx}\right|_{x=a}$

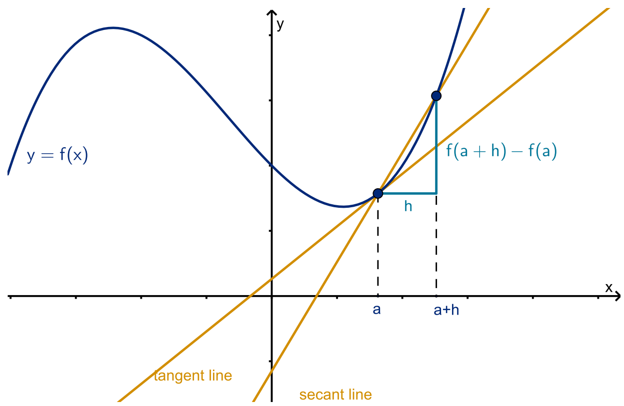


Figure: A limit of the slopes of secant lines

We can attempt to compute the derivative at any point a. We can put these values together to create a function $f'(x)$.

Definition

The **derivative function** of $f(x)$ is the function that takes the value

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

at each x.

We can denote the derivative function as $f'(x)$ or $\frac{df}{dx}$. The second can be rewritten $\frac{d}{dx} f$ to emphasize that we are applying the differentiation operation to the function f.

Example

If $f(x) = x^2 + 2x$, compute $f'(x)$.

Solution

$$
\begin{aligned}
f'(x) &= \lim_{h\to 0} \frac{f(x+h)-f(x)}{h} && \text{definition of derivative}\\
&= \lim_{h\to 0} \frac{(x+h)^2+2(x+h)-x^2-2x}{h} && \text{plug in } x \text{ and } x+h\\
&= \lim_{h\to 0} \frac{x^2+2xh+h^2+2x+2h-x^2-2x}{h} && \text{distribute}\\
&= \lim_{h\to 0} \frac{2xh+h^2+2h}{h} && \text{cancel}\\
&= \lim_{h\to 0} 2x+h+2 && \text{functions agree except at } h=0 \text{ so limits are equal}\\
&= 2x+0+2 && \text{limit } = \text{ value on a continuous function}\\
&= 2x+2
\end{aligned}
$$

Theorem

If $f'(x) > 0$ for all x in some interval $[a, b]$ then $f(x)$ is **increasing** on $[a, b]$.
If $f'(x) < 0$ for all x on $[a, b]$ then $f(x)$ is **decreasing** on $[a, b]$.

We can take **higher order derivatives** by taking derivatives of derivatives. The derivative function of f in this context is called the **first derivative**. Its derivative function is the **second derivative**. The second derivative's derivative function is the **third derivative** and so on.

Notation

The following notations are used for higher order derivatives

name	prime notation	Leibniz notation
first derivative	$f'(x)$	$\frac{df}{dx}$
second derivative	$f''(x)$	$\frac{d^2f}{dx^2}$
third derivative	$f'''(x)$	$\frac{d^3f}{dx^3}$
fourth derivative	$f^{(4)}(x)$	$\frac{d^4f}{dx^4}$
fifth derivative	$f^{(5)}(x)$	$\frac{d^5f}{dx^5}$

Question 1.2.6 How Do We Measure the Change in a Function?

The sign of a higher order derivative tells us how the derivative of one order lower is changing. For example if $\dfrac{d^5f}{dx^5} < 0$, then $\dfrac{d^4f}{dx^4}$ is decreasing. The sign of higher order derivatives is difficult to discern from the shape of $y = f(x)$, with the exeption of the second derivative.

Theorem

If $f''(x) > 0$ on some interval, then $y = f(x)$ is **concave up** on that interval. If $f''(x) < 0$, then $y = f(x)$ is **concave down**.

Definition

A point a such that $f(x)$ is concave up to one side of a and concave down to the other side is called an **inflection point**.

Question 1.2.7 How Do We Compute Derivatives

The limit definition of a derivative is too unwieldy to use every time. A better approach is to learn the derivatives of some simple functions, and then use theorems to compute derivatives when those functions are combined.

Derivatives of Simple Functions

- $\frac{d}{dx}c = 0$ (derivative of a constant is 0)
- $\frac{d}{dx}x^n = nx^{n-1}$ for any $n \neq 0$ (The Power Rule)
- $\frac{d}{dx}\sin x = \cos x$
- $\frac{d}{dx}\cos x = -\sin x$
- $\frac{d}{dx}e^x = e^x$
- $\frac{d}{dx}a^x = a^x \ln a$ for $a > 0$
- $\frac{d}{dx}\ln x = \frac{1}{x}$

Theorem

The following rules allow us to differentiate functions made of simpler functions whose derivative we know.

Sum Rule $(f(x)+g(x))' = f'(x)+g'(x)$

Constant Multiple Rule $(cf(x))' = cf'(x)$

Product Rule $(f(x)g(x))' = f'(x)g(x)+g'(x)f(x)$

Quotient Rule $\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)g(x)-g'(x)f(x)}{(g(x))^2}$ unless $g(x)=0$

Chain Rule $(f(g(x))' = f'(g(x))g'(x)$

Example

Compute $\dfrac{d}{dx}\tan(x)$

Solution

$\tan x = \dfrac{\sin x}{\cos x}$. We apply the quotient rule

$$
\begin{aligned}
(\tan x)' &= \frac{(\sin x)'\cos x - (\cos x)'\sin x}{\cos^2 x} && \text{quotient rule} \\
&= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} \\
&= \frac{1}{\cos^2 x} && \text{Pythagorean identity} \\
&= \sec^2 x
\end{aligned}
$$

Application 1.2.8
The Shape of a Graph

What can the first and second derivative of $f(x) = 8x^3 - x^4$ tell us about the shape of its graph?

Solution

We will compute the first and second derivative using the power rule. Factoring them will allow us to perform a sign analysis.

$$f'(x) = 24x^2 - 4x^3 = 4x^2(6-x) \qquad f''(x) = 48x - 12x^2 = 12x(4-x)$$

$4x^2$	$+$		$+$		$+$
$(6-x)$	$+$		$+$		$-$
$f'(x)$	$+$		$+$		$-$
		0		6	

$12x$	$-$		$+$		$+$
$(4-x)$	$+$		$+$		$-$
$f''(x)$	$-$		$+$		$-$
		0		4	

From the sign of $f'(x)$ we conclude f is increasing on $(-\infty, 0)$ and $(0, 6)$ but decreasing on $(6, \infty)$. From the sign of $f''(x)$ we conclude that f is concave down on $(-\infty, 0)$ and $(4, \infty)$, but concave up on $(0, 4)$.

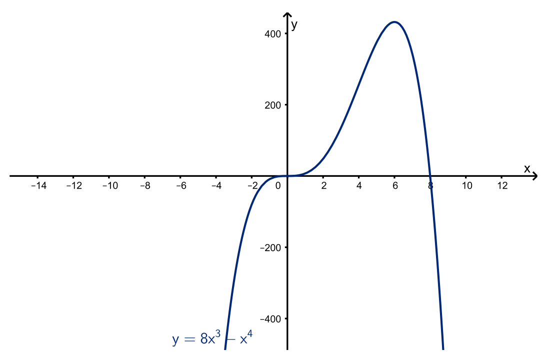
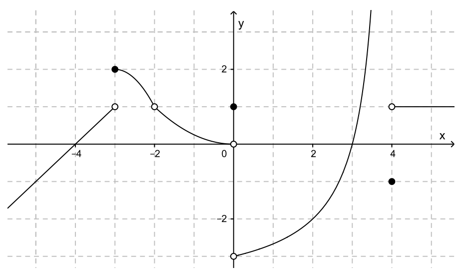


Figure: The graph of $y = 8x^3 - x^4$

1.2.1

Q1 Given the graph of $y = f(x)$ here, give the value of each of the following limits (if they exist).



a $\lim_{x\to -3^-} f(x)$

b $\lim_{x\to -3^+} f(x)$

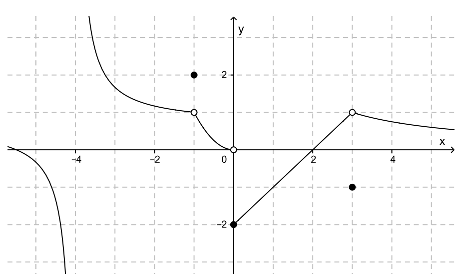
c $\lim_{x\to -2} f(x)$

d $\lim_{x\to 0} f(x)$

e $\lim_{x\to 4^-} f(x)$

f $\lim_{x\to 4^+} f(x)$

Q2 Given the graph of $y = g(x)$ here, give the value of each of the following limits (if they exist).



a $\lim_{x\to 0^-} g(x)$

b $\lim_{x\to 0^+} g(x)$

c $\lim_{x\to 0} g(x)$

d $\lim_{x\to 3^-} g(x)$

e $\lim_{x\to 3^+} g(x)$

f $\lim_{x\to 3} g(x)$

g $\lim_{x\to -4^-} g(x)$

h $\lim_{x\to -4^+} g(x)$

i $\lim_{x\to -1} g(x)$

1.2.2

Q3 Explain why $f(x) = \frac{e^x}{x^2+3}$ is continuous on $\mathbb{R}$.

Q4 Explain why $f(x) = \sqrt{\sin(3x^2)}$ is continuous on its domain.

Q5 Is

$$f(x) = \begin{cases} \sin(2x) & \text{if } x < 0 \\ 4 & \text{if } x = 0 \\ -x^2 & \text{if } x > 0 \end{cases}$$

continuous at $x = 0$? Justify your answer.

Q6 Is

$$f(x) = \begin{cases} x^3 - 2x + 1 & \text{if } x < 0 \\ e^x & \text{if } x \geq 0 \end{cases}$$

continuous at $x = 0$? Justify your answer.

Q7 Is

$$f(x) = \begin{cases} x + 5 & \text{if } x < 1 \\ 6 & \text{if } x = 1 \\ x^2 + 4x + 1 & \text{if } x > 1 \end{cases}$$

continuous at $x = 1$? Justify your answer.

Q8 Where is

$$f(x) = \begin{cases} \cos(\pi x) & \text{if } x < 4 \\ 1 & \text{if } x = 4 \\ \sqrt{x - 3} & \text{if } x > 4 \end{cases}$$

continuous?

1.2.3

Q9 Compute $\lim_{x\to 3} \dfrac{x-3}{x^2-9}$

Q10 Compute $\lim_{x\to 1} \dfrac{x^2-4x+3}{x-1}$

Q11 Compute $\lim_{x\to 9} \dfrac{2x-18}{\sqrt{x}-3}$

Q12 Compute $\lim_{x\to 4} \dfrac{\frac{1}{x^2}-\frac{1}{16}}{x-4}$

1.2.4

Q13 Explain why $\sin x = 2x-1$ has a solution in $[0,1]$.

Q14 Explain why $\sqrt[3]{x} = \log_2 x$ has a solution in $[0,8]$.

Q15 What does the Intermediate Value Theorem say about whether $f(x) = \frac{1}{x} - \frac{1}{2}$ has a root in $[-1,1]$?

Q16 Consider the equation $\sin x = \dfrac{3}{4}$. Gloria computes $\sin\dfrac{\pi}{3} = \dfrac{\sqrt{3}}{2}$ and $\sin\dfrac{5\pi}{6} = \dfrac{1}{2}$. Since $\dfrac{3}{4}$ is not between $\dfrac{1}{2}$ and $\dfrac{\sqrt{3}}{2}$, she concludes that $\sin x = \dfrac{3}{4}$ has no roots in $\left[\frac{\pi}{3}, \frac{5\pi}{6}\right]$. What do you think of Gloria's reasoning?

1.2.5

Q17 Compute $\lim_{x\to\infty} \dfrac{x^2+2x-9}{3x-6}$.

Q18 Compute $\lim_{x\to\infty} \dfrac{4x^2-7x+9}{2x^2+11}$.

Q19 Compute $\lim_{x\to\infty} \sqrt{e^{1/x}}$.

Q20 Compute $\lim_{x\to\infty} \dfrac{1}{\ln x}$.

Q21 Compute $\lim_{x\to-\infty} e^{e^x}$.

Q22 Compute $\lim_{x\to\infty} \sin(\ln x)$.

1.2.6

Q23 Let $f(x) = x^3$.

- **a** Compute the average rate of change of f from $x = 2$ to $x = 5$.
- **b** Give the equation of the secant line that meets $y = f(x)$ at $x = 2$ and $x = 5$.
- **c** Use the limit definition of the derivative to compute $f'(2)$.

Q24 Let $f(x) = \sqrt{x}$ Compute the average rate of change of f between $x = 4$ and $x = 9$. Based on the graph of $y = f(x)$, is the instantaneous rate of change at $x = 4$ greater or less than this average?

Q25 Let $f(x) = 3x^2 - 7$. Compute $f'(6)$ using the limit definition of the derivative.

Q26 Let $f(x) = \frac{1}{x+2}$. Compute $f'(1)$ using the limit definition of the derivative.

Q27 Let $f(x) = \frac{1}{x^2}$. Compute $f'(x)$ using the limit definition of the derivative.

Q28 Let $f(x) = \sqrt{x}$. Compute $f'(x)$ using the limit definition of the derivative.

1.2.7

Q29 Use derivative rules to differentiate each of the following functions.

a. $5x^7 - 3x^2 + \dfrac{5}{x^2}$

b. $\dfrac{4x^5 - 2x^2 + 3x + 4}{x}$

c. $(x^2 + 2x)\sin x$

d. $\dfrac{e^x}{x^2}$

e. $\sqrt{x - 5}$

f. $\cos(4x)$

g. $\sin(e^x)$

h. $(x^2 + 5x + 4)^{60}$

i. $e^{x^2 \sin x}$

j. $\dfrac{\ln(x^2 + 2)}{x^2 + 3x}$

Q30 Use derivative rules to differentiate each of the following functions.

a. $\dfrac{3}{x} + \dfrac{7}{x^3}$

b. $\dfrac{5x^4 + 3x^3 - 8x^2}{x^2}$

c. $\dfrac{\ln x}{x}$

d. $4^x \sin(x)$

e. $\tan(2x + 7)$

f. e^{3x+2}

g. $\cos(x^3 + 2x)$

h. $\frac{5}{(\cos x)^3}$

i. $e^{x^2} \sin^3 x$

j. $\ln(\sqrt{x}\sin x)$

Q31 Let $f(x) = \sin(3x)$. Compute $f'''(x)$.

Q32 Let $f(x) = e^{x^3}$. Compute $f''(x)$.

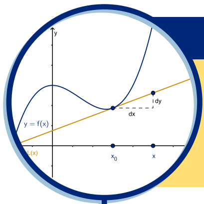
1.2.8

Q33 Where in its domain is the function $f(x) = x^3 - x^2$ increasing?

Q34 Where in its domain is $f(x) = e^x - x^2$ concave up?

Q35 Where in its domain is $f(x) = 1024\sqrt{x} - x^4$ increasing?

Q36 Find the inflection point(s) of $x^4 - 8x^3$.



Section 1.3

Applications of Derivatives

Goals:

1. Write the equation of a tangent line.
2. Identify local maximums and minimums.
3. Use the Extreme Value Theorem to find minimums and maximums.
4. Use l'Hôpital's rule to compute limits.

This section reviews the most important applications of the derivative.

Application 1.3.1

The Tangent Line to a Graph

Given a function $f(x)$, the derivative $f'(a)$ is the slope of the line tangent to $y = f(x)$ at $(a, f(a))$.

Formula

The equation of the tangent line to $y = f(x)$ at $x = a$ in point-slope form is:

$$y - f(a) = f'(a)(x - a)$$

We can rewrite the tangent line as a function of x. We call this a linearization, because this function is linear, but it approximates the value of $f(x)$ for x near a.

Formula

The linearization of $y = f(x)$ at $x = a$ is the function:

$$L(x) = f(a) + f'(a)(x - a)$$

If we want to emphasize the change in x and y instead of their actual values we can use differential notation:

Application 1.3.1 The Tangent Line to a Graph

Notation

If $y = f(x)$ is approximated by a tangent line at $x = a$ then we let

- $dx = x - a$ represents a change in x from a. Since x is an independent variable, so is dx.
- $dy = f'(a)\ dx$ is equal to the change in y corresponding to dx, if we travel along the tangent line. This approximates the actual change in $f(x)$ if x increases by dx.

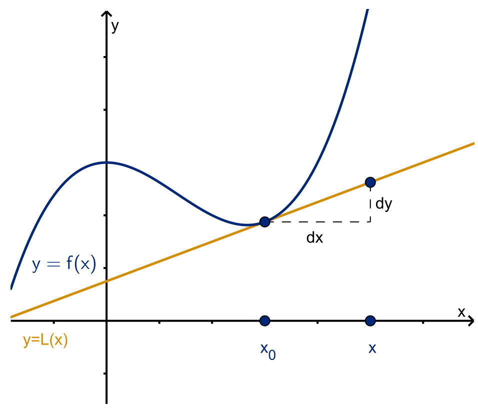


Figure: The differentials dx and dy on the tangent line to $y = f(x)$

Application 1.3.2 Maximum and Minimum Values of a Function

Definition

A number a is a **maximum** of a function $f(x)$ if $f(a) \geq f(x)$ for all x in the domain of f.
a is a **minimum** if $f(a) \leq f(x)$ for all x in the domain of f.

Definition

A number a is a **local maximum** of a function $f(x)$ if $f(a) \geq f(b)$ for all b in some neighborhood of a. a is a **local minimum** if $f(a) \leq f(b)$ for all b in some neighborhood of a.

To distinguish ordinary maximums from the local variety, we sometimes call them **global maximums** or **absolute maximums**. Every global maximum is a local maximum, but local maximums need not be global maximums. If $f'(a) > 0$ then there are larger values of $f(a)$ to the right of a and lower values to the left. Thus a cannot be a local maximum or minimum. The same argument applies if $f'(a) < 0$.

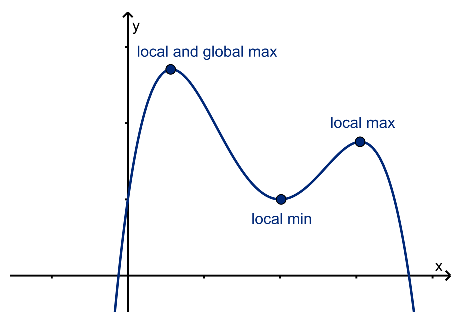


Figure: Maximum and minimum values of $f(x)$

Definition

A **critical point** of $f(x)$ is a value a in the domain of f such that either $f'(a) = 0$ or $f'(a)$ does not exist.

Theorem [The First Derivative Test]

Local maximums and minimums of $f(x)$ can only occur at critical points.

We can use concavity as a way to classify critical points. Knowing whether a graph is concave up or concave down at a point where $f'(x) = 0$ allows us to visualize a small neighborhood of that point.

Theorem [The Second Derivative Test]

Let a be a critical point of f.

- If $f''(a) < 0$ then a is a local maximum.
- If $f''(a) > 0$ then a is a local minimum.
- If $f''(a) = 0$ or does not exist, then the test is inconclusive. a could be a local maximum, a local minimum, or neither.

Example

What does the second derivative test tell you about the critical points of $f(x) = 8x^3 - x^4$?

Solution

First we compute the critical points.

$$\begin{aligned} f'(x) &= 24x^2 - 4x^3 && \text{compute first derivative} \\ 0 &= 24x^2 - 4x^3 && \text{set equal to } 0 \\ 0 &= 4x^2(6 - x) && \text{factor} \\ x &= 0 \text{ or } x = 6 \end{aligned}$$

Now we compute the second derivative and evaluate it at each critical point.

$$\begin{aligned} f''(x) &= 48x - 12x^2 \\ f''(0) &= 0 \\ f''(6) &= (48)(6) - (12)(36) = -144 \end{aligned}$$

$f''(6) < 0$ so $x = 6$ is a local maximum. $f''(0) = 0$ so the second derivative test cannot tell whether $x = 0$ is a local maximum or local minimum (in fact it is neither).

Question 1.3.3

Does a Function Always Have a Maximum?

No. Many functions don't have maximums, because as x gets larger and larger the values of $f(x)$ increase or decrease without bound. However, if we restrict the domain, we can sometimes guarantee a maximum

Theorem [The Extreme Value Theorem]

If $f(x)$ is a continuous function on a closed domain $[a, b]$ then f has an absolute maximum and an absolute minimum on $[a, b]$.

Remark

When the EVT applies, we can find the absolute maximum and minimum by process of elimination. A maximum exists, so it must occur at a critical point. We can find the critical points and evaluate f at each of them. Whichever has the greatest value is the maximum.

Note that a and b are always critical points because the derivative does not exist there. There is no limit from the left at a because those points are outside the domain of f. Similarly, there is no limit from the right at b.

Example

Compute the maximum and minimum value of $f(x) = 8x^3 - x^4$ on the domain $[2, 8]$, if they exist.

Solution

$f(x)$ is continuous and $[2, 8]$ is closed, so the EVT guarantees that a maximum and minimum exist. The first derivative test says that they can only occur at critical points.

$$\begin{aligned} f'(x) &= 24x^2 - 4x^3 && \text{compute first derivative} \\ 0 &= 24x^2 - 4x^3 && \text{set equal to } 0 \\ 0 &= 4x^2(6 - x) && \text{factor} \\ x &= 0 \text{ or } x = 6 \end{aligned}$$

$x = 0$ is not in the domain, so we discard it. On the other hand $x = 2$ and $x = 8$ are also critical points because the derivative does not exist there. To find which critical point is the maximum and which is the minimum, we plug each into f and compare.

$$\begin{aligned} f(2) &= (8)(8) - 16 = 48 \\ f(6) &= (8)(216) - 1296 = 436 && \text{(maximum)} \\ f(8) &= (8)(512) - 4096 = 0 && \text{(minimum)} \end{aligned}$$

Application 1.3.4

L'Hôpital's Rule

The limit rules tell us how to take limits of quotients, products, sums and differences. What happens if one of the functions being divided goes to ∞, or if the denominator of a quotient goes to 0? In some cases we can reason this out using our intuition of arithmetic.

Example

Consider $\displaystyle\lim_{x\to\infty} \frac{\tan^{-1}(x)}{\ln x}$.

- $\displaystyle\lim_{x\to\infty} \tan^{-1} x = \frac{\pi}{2}$
- $\displaystyle\lim_{x\to\infty} \ln x = \infty$

Since the numerators are approaching $\pi/2$ and the denominators are increasing without bound, we conclude that this ratio get smaller and smaller and will limit to 0.

In other cases, intuition cannot help us.

Definition

A limit of the form $\displaystyle\lim_{x\to a} \frac{f(x)}{g(x)}$ is of **indeterminate form** if either

- $\displaystyle\lim_{x\to a} f(x) = \pm\infty$ and $\displaystyle\lim_{x\to a} g(x) = \pm\infty$ or
- $\displaystyle\lim_{x\to a} f(x) = 0$ and $\displaystyle\lim_{x\to a} g(x) = 0$

This definition also applies to one-sided limits or to limits at $\pm\infty$.

Limits of products and sums can sometimes be rewritten as quotients of indeterminate form as well.

Theorem [L'Hôpital's Rule]

If $\displaystyle\lim_{x\to a} \frac{f(x)}{g(x)}$ is of indeterminate form, then it is equal to

$$\lim_{x\to a} \frac{f'(x)}{g'(x)}$$

assuming this limit exists.

Often L'Hôpital's Rule converts a limit of indeterminate form to one we can evaluate through intuition or direct computation. Sometimes, we need to apply L'Hôpital's Rule more than once.

Warning

If a limit is not of indeterminate form, then L'Hôpital's Rule does not apply. Attempting to apply it will usually give an incorrect value for the limit.

Example 1.3.5
A Limit of Indeterminate Form

Evaluate $\lim\limits_{x\to 0} \dfrac{e^x - x - 1}{x^2}$

Solution

$$\begin{aligned}
&\lim_{x\to 0} \frac{e^x - x - 1}{x^2} && \frac{0}{0} \text{ form} \\
&= \lim_{x\to 0} \frac{e^x - 1}{2x} && \text{L'Hôpital's Rule, still } \frac{0}{0} \text{ form} \\
&= \lim_{x\to 0} \frac{e^x}{2} && \text{L'Hôpital's Rule again} \\
&= \frac{1}{2}
\end{aligned}$$

Section 1.3
Exercises

1.3.1

Q1 Write the equation of the tangent line to $y = \sqrt{x}$ at $(4, 2)$.

Q2 Write the equation of the tangent line to $y = \frac{1}{x^2}$ at $\left(5, \frac{1}{25}\right)$.

Q3 Let $f(x) = \sin(x)$

- **a** Write the equation of the linearization $y = f(x)$ at $x = \frac{\pi}{3}$.
- **b** If we wanted to use **a** to approximate $\sin(1)$ by hand, what number(s) would we need decimal approximations of?
- **c** Use a calculator to get decimal approximations of those numbers, then show how to approximate $\sin(1)$.

Q4 Write a linearization of $f(x) = \dfrac{1}{x}$ at $x = 3$ and use it to approximate $\dfrac{1}{2.93}$.

Q5 A baterical culture has mass $3g$ after $t = 5$ hours of growth. At that time, its instantaneous rate of growth is $0.2g/hr$.

a Write a linear function to approximate $m(t)$ the mass of the culture at hour t.

b Approximate the mass at time 8 hours.

c Given that $m''(t) > 0$, is your answer to **b** is overestiamte or an underestimate?

Q6 A space capsule is descending from orbit. After 90 seconds, it is $10,000$m above sea level and falling at 400m per second.

a Write a linearization for $h(t)$, the height of the capsule at time t.

b Use **a** to predict when the capsule will splash down into the ocean.

c Do you expect that your answer to **b** is an overestimate or underestimate? Explain.

1.3.2

Q7 Find the critical points of $f(x) = 12x^{2/3} - x$.

Q8 Find the critical points of $g(x) = x^4 - 18x^2 + 5$. Apply the second derivative test to each.

Q9 Find the critical points of $f(x) = x^3 - 75x$. Apply the second derivative test to each.

Q10 Find the critical points of $g(x) = e^x - 2x$. Apply the second derivative test to each.

1.3.3

Q11 Find the maximum and minimum values of $f(x) = x^{2/3}$ on $[-8, 1]$.

Q12 Find the maximum and minimum values of $f(x) = x^3 - 75x$ on $[-10, 10]$.

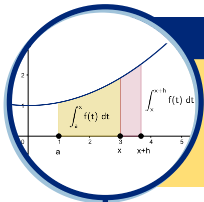
1.3.4

Q13 Evaluate $\displaystyle\lim_{x\to 0^+} \frac{x\cos(x-\pi)}{e^x-1}$.

Q14 Evaluate $\displaystyle\lim_{x\to 0^+} \frac{e^{-3x}+3x-1}{\sin(x^2)}$.

Q15 Evaluate $\displaystyle\lim_{x\to\infty} \frac{x\ln x}{x^{5/2}+3}$.

Q16 Evaluate $\displaystyle\lim_{x\to-\infty} e^x x^2$.



Section 1.4

Definite Integrals

Goals:

1. Express areas under a graph and antiderivatives using integral notation.
2. Derive antiderivatives from known derivatives.
3. Compute general antiderivatives.
4. Compute definite integrals using the Fundamental Theorem of Calculus.
5. Use u-substitution to compute integrals where necessary.

By definition, integrals compute area under a graph. The Fundamental Theorem of Calculus connects integrals to antiderivatives, meaning that integrals can also be used to compute total change, given a rate of change function.

Question 1.4.1

What Is an Antiderivative?

Definition

$F(x)$ is **antiderivative** of a function $f(x)$, if $F'(x) = f(x)$.

Every derivative we know also tells us an antiderivative.

Example

$\frac{d}{dx}\left(\frac{x^2}{2} + 5\right) = x$ so $F(x) = \frac{x^2}{2} + 5$ is an antiderivative of $f(x) = x$.

Notice that $\frac{x^2}{2} + 2$, $\frac{x^2}{2} - 6$, and $\frac{x^2}{2}$ are also antiderivatives of $f(x) = x$.

Functions have infinitely many antiderivatives. Adding a constant to one antiderivative produces another, since the derivative of a constant is 0. In fact, this is the only relationship between antiderivatives.

Theorem

If $F(x)$ and $G(x)$ are antideriavatives of $f(x)$, then there is a constant c such that

$$F(x) = G(x) + c.$$

Since the antiderivatives are related this way, it is easy to express all of the antiderivatives of a function at once.

Definition

If $F(x)$ is an antiderivative of $f(x)$, then the **general antiderivative** of $f(x)$ is the family of functions:

$$F(x) + c$$

where c can be any constant.

Here is a table of antiderivatives that we can compute just by reverse engineering the derivatives we already know.

$f(x)$	general antiderivative of $f(x)$
x^n	$\frac{x^{n+1}}{n+1} + c$
e^x	$e^x + c$
a^x	$\frac{a^x}{\ln a} + c$
$\frac{1}{x}$	$\ln x + c$
$\sin x$	$-\cos x$
$\cos x$	$\sin x$

Remark

Many familiar functions are missing from this list. This is because we just haven't come across them as derivatives of some other function. For instance, we do not yet know a function $F(x)$ whose derivative is $\ln x$ or $\tan x$.

Question 1.4.2

How Do We Compactly Denote a Sum of Many Terms

Defining the definite integral requires us to add up many numbers. The problem is not just that the number of summands is large. We need to be flexible about how many terms are in the sum. The notation that gives us this flexibility is Σ notation.

Notation

Σ ('sigma') notation allows us to sum many different values of an expression using an **index variable**. The index variable will be replaced by each integer between an initial and final value, and the resulting outputs are added together.

$$\sum_{k=1}^{n} f(k) = f(1) + f(2) + f(3) + \cdots + f(n)$$

We may choose any variable as the index variable. The index variable could also have a different initial value, if that is more convenient.

Question 1.4.2 How Do We Compactly Denote a Sum of Many Terms

Example

$$\sum_{j=3}^{7} \frac{j^2}{j+1} = \frac{9}{4} + \frac{16}{5} + \frac{25}{6} + \frac{36}{7} + \frac{49}{8}$$

Part of the challenge of writing a sum in $\sum$ notation is choosing an f that will produce all the terms of your sum.

Example 1.4.3 Writing a Sum in Σ Notation

Write each of the following sums in Σ notation.

a $4 + 7 + 10 + 13 + 16 + 19 + 22$

b $2 + 6 + 18 + 54 + 162 + 486$

c $-3 + 4 - 5 + 6 - 7 + 8 - 9 + 10$

d $\frac{1}{4} + \frac{\sqrt{2}}{9} + \frac{\sqrt{3}}{16} + \frac{2}{25} + \frac{\sqrt{5}}{36}$

Solution

a The terms increase by 3 each time. Repeated addition is multiplication, in this case $3k$ plus some starting value. Starting with index $k = 0$ is convenient, because $3(0) = 0$ at the starting value.

$$4 + 7 + 10 + 13 + 16 + 19 + 22 = \sum_{k=0}^{6} 4 + 3k$$

b The terms are multiplied by 3 each time. Repeated multiplication is exponentiation, in this case 3^k times some starting value. Starting with index $k = 0$ is convenient, because $3^0 = 1$ at the starting value.

$$2 + 6 + 18 + 54 + 162 + 486 = \sum_{k=0}^{5} (2)(3^k)$$

c The absolute values of this sum could just be the values of the index variable. To create an alternating $+$ and $-$ pattern, we can multiply by $(-1)^k$.

$$-3+4-5+6-7+8-9+10=\sum_{k=3}^{10}(-1)^k k$$

d In a fraction, we can model the numerator and denominator separately.

$$\frac{1}{4}+\frac{\sqrt{2}}{9}+\frac{\sqrt{3}}{16}+\frac{2}{25}+\frac{\sqrt{5}}{36}=\sum_{k=1}^{5}\frac{\sqrt{k}}{(k+1)^2}$$

Question 1.4.4
How Do We Compute the Area Under a Graph?

Suppose we would like to know the area below the graph $y=f(x)$ between $x=a$ and $x=b$. We approximate this area by rectangles. We can improve these approximations and take a limit of such improvements to compute the actual area. Here is the procedure.

1. Divide $[a,b]$ into n subintervals, of lengths Δx_i.
2. Pick a point x_i^* in each subinterval.
3. Evaluate $f(x_i^*)$, which is the height of the graph above x_i^*.
4. Produce a rectangle of height $f(x_i^*)$ and width Δx_i over each subinterval.
5. Sum the areas of these rectangles. This is an approximation of the actual area.
6. Take a limit of such approximations as $|\Delta x|$, the largest of the Δx_i goes to 0.

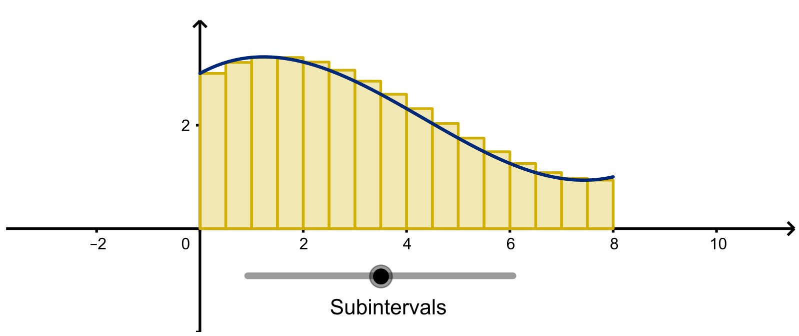


Figure: The area under $y=f(x)$ approximated by rectangles

Defintion

We define the **definite integral** of $f(x)$ over $[a,b]$ to be

$$\int_a^b f(x)\ dx = \lim_{|\Delta x|\to 0} \sum f(x_i^*)\Delta x_i$$

where the limit is taken over all divisions of $[a,b]$, Δx_i is the length of the ith subinterval, x_i^* is a point in the ith subinterval and $|\Delta x|$ is the largest Δx_i.

Notice there is no requirement that the subintervals be the same length. Because of this, we don't take a limit as n approaches ∞. For instance, using a large number of rectangles from $\left[a, \frac{a+b}{2}\right]$ and only a single rectangle over $\left[\frac{a+b}{2}, b\right]$ will not give us a good approximation, no matter how many rectangles we use. Instead we take a limit as the largest Δx_i approaches 0.

In practice, we get the same limit whether the subintervals are equal length or not not. It is common to use the same $\Delta x = \frac{b-a}{n}$ for each subinterval.

The definite integral almost solves our area problem, but wherever $f(x) < 0$, the product $f(x_i^*)\Delta x_i$ will be negative.

Theorem

If $f(x) > 0$ on $[a,b]$ then $\displaystyle\int_a^b f(x)\ dx$ computes the area under $y = f(x)$ over $[a,b]$. In general $\displaystyle\int_a^b f(x)\ dx$ computes the **signed area** between $y = f(x)$ and the x-axis, where area above the axis counts as positive, and area below the axis counts a negative.

Since integrals are limits, they inherit two laws from limits. The third can be taken from geometry, setting the area of a region equal to the sum of the areas of two subregions.

Integral Laws

- $\displaystyle\int_a^b f(x) + g(x)\ dx = \int_a^b f(x)\ dx + \int_a^b g(x)\ dx$ (Sum Rule)
- $\displaystyle\int_a^b cf(x)\ dx = c\int_a^b f(x)\ dx$ (Constant Multiple Rule)
- $\displaystyle\int_a^b f(x)\ dx = \int_a^c f(x)\ dx + \int_c^b f(x)\ dx$ (Union Rule)

Question 1.4.5

How Do We Evaluate an Integral?

The limit form of an integral is usually impossible to evaluate directly. Instead we use a powerful pair of theorems.

Theorem [The First Fundamental Theorem of Calculus]

Given a function $f(x)$, let $g(x) = \displaystyle\int_a^x f(t)\ dt$. At any x where f is continuous, $g'(x) = f(x)$.

To prove this, we use the definition of a derivative.

$$\begin{aligned} g'(x) &= \lim_{h\to 0} \frac{g(x+h) - g(x)}{h} \\ &= \lim_{h\to 0} \frac{\int_a^{x+h} f(t)\ dt - \int_a^x f(t)\ dt}{h} \\ &= \lim_{h\to 0} \frac{\int_x^{x+h} f(t)\ dt}{h} && \text{union rule} \end{aligned}$$

As the interval $[x, x+h]$ shrinks, the values of f over that interval can be made arbitrarily close to $f(x)$, since f is continuous. Thus $\int_x^{x+h} f(t)\ dt$ approaches the area of a rectangle of height $f(x)$ and width h. Thus

$$\lim_{h\to 0} \frac{\int_x^{x+h} f(t)\ dt}{h} = f(x)$$

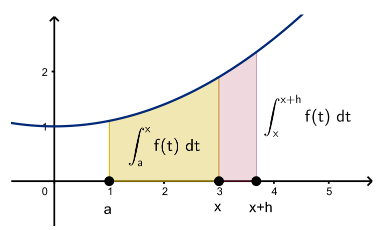


Figure: $g(x+h) - g(x)$ represented as an area

The main use of the First Fundamental Theorem of Calculus is to prove the Second Fundamental Theorem of Calculus.

Theorem [The Second Fundamental Theorem of Calculus]

Let $f(x)$ be a continuous function on $[a, b]$. If $F(x)$ an antiderivative of $f(x)$, then

$$\int_a^b f(x)\ dx = F(b) - F(a)$$

This follows immediately from the First Fundamental Theorem. If we continue to define $g(x) = \int_a^x f(t)\ dt$, then

$$\begin{aligned}\int_a^b f(x)\ dx &= \int_a^b f(x)\ dx - \int_a^a f(x)\ dx \\ &= g(b) - g(a)\end{aligned}$$

We know that $g(x)$ is an antiderivative of $f(x)$. If we instead pick a different antiderivative $F(x)$, then $F(x) = g(x) + c$, and

$$\begin{aligned}F(b) - F(a) &= g(b) + c - (g(a) + c) \\ &= g(b) - g(a) \\ &= \int_a^b f(x)\ dx\end{aligned}$$

Because we will be computing $F(b) - F(a)$ frequently, we will develop the following shorthand.

Notation

The quantity $F(b) - F(a)$ can be denoted

$$F(x)\Big|_a^b$$

This relationship between integrals and antiderivatives motivates the following vocabulary.

Notation

The general antiderivative of $f(x)$ is also called an **indefinite integral** and is denoted

$$\int f(x)\ dx.$$

Example 1.4.6

A Definite Integral

Compute $\int_2^5 x^2\ dx$

Solution

$$\begin{aligned}\int_2^5 x^2\ dx &= \left.\frac{x^3}{3}\right|_2^5 \\ &= \frac{5^3}{3} - \frac{2^3}{3} \\ &= \frac{125-8}{3} \\ &= 39\end{aligned}$$

Question 1.4.7

How Do We Apply the Chain Rule in an Antiderivative?

The chain rule states that

$$(f(g(x)))' = f'(g(x))g'(x).$$

The key insight here is to rewrite think of g as a variable, in addition to being a function of x. Typically we rename it with a letter closer to the end of the alphabet, like u. The following substituion theorem uses the chain rule to say that we can integrate with respect to u instead of x.

Theorem

If $u(x)$ is a function of x, then

$$\int_a^b f(u(x))u'(x)\ dx = \int_{u(a)}^{u(b)} f(u)\ du$$

This allows us to replace a complicated integrand in x with a simpler one in u. To correctly rewrite the integral, the bounds must be updated to the corresponding values of u.

We can also apply this to indefinite integrals. If F is an antiderivative of f, then

$$\begin{aligned}\int f(u(x))u'(x)\ dx &= \int f(u)\ du \\ &= F(u) + c \\ &= F(u(x)) + c\end{aligned}$$

The most common u substitutions are linear, where $u = ax$.

Question 1.4.7 How Do We Apply the Chain Rule in an Antiderivative?

Example

Compute $\int \sin 3x \; dx$

Solution

We will perform a u substitution, using $u = 3x$.

$$\begin{aligned} \int \sin(3x)\; dx &= \int \frac{1}{3} \sin u \; du \\ &= -\frac{1}{3} \cos u + c \\ &= -\frac{1}{3} \cos(3x) + c \end{aligned}$$

u-substitution

$$\begin{aligned} u &= 3x \\ du &= 3\; dx \\ \tfrac{1}{3}\; du &= dx \end{aligned}$$

Note that we should express our antiderivatives in terms of the original variable (often x), not in terms of u.

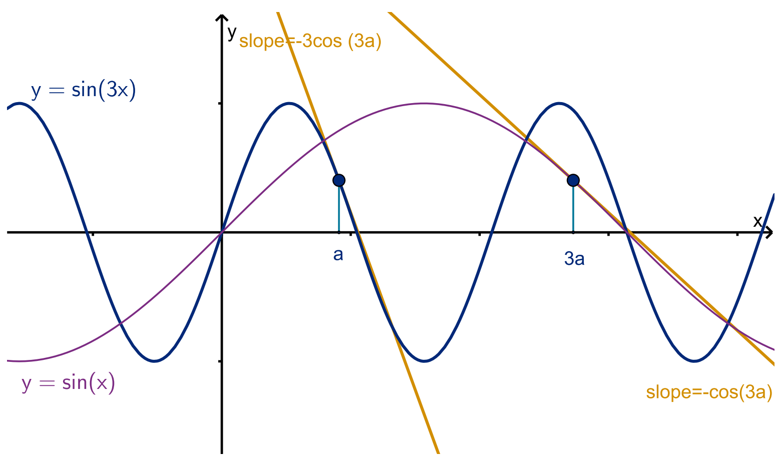


Figure: The graphs $y = \sin x$, $y = \sin 3x$ their related tangent lines

Example 1.4.8

A u-substitution

Compute the integral:

$$\int_0^3 x e^{x^2} \; dx$$

Solution

We start by looking for a candidate for $u(x)$. Since we want the integrand to be $f(u(x))u'(x)$, we note $u(x)$ should be the inner function in some composition. x^2 is the natural target. We attempt the substituion, and hope that the remaining factors in the integrand can be expressed in terms of $u'(x)$. We see that our $u'(x)\ dx$ is $2x\ dx$. Since we only have an $x\ dx$ in our integrand, we divide by 2.

$$\begin{aligned}\int_0^3 xe^{x^2}\ dx &= \int_0^9 \frac{1}{2}e^u\ du \\ &= \frac{1}{2}e^u\Big|_0^9 \\ &= \frac{1}{2}(e^9-1)\end{aligned}$$

u-substitution	
$u = x^2$	$x = 0 \Rightarrow u = 0$
$du = 2x\ dx$	$x = 3 \Rightarrow u = 9$
$\frac{1}{2}\ du = x\ dx$	

Section 1.4

Exercises

1.4.1

Q1 Write two different anti-derivatives of $f(x) = x + 5$.

Q2 Write two different anti-derivatives of $f(x) = x^3 - 6x^2 + \frac{2}{x}$.

Q3 Write a general antiderivative of $f(x) = 4\cos x + 6x^2$.

Q4 Suppose $x^4 - \sin(x^3)$ is an antiderivative of $f(x)$. Write three other antiderivatives of $f(x)$. You should do this without computing what f is.

Q5 If $F(x)$ and $G(x)$ are both antiderviatives of $f(x)$, find the value b such that $3F(x) - bG(x)$ is also an antiderivative of $f(x)$.

Q6 Suppose F and G are both antiderivatives of $f(x)$. Suppose further that $\mathcal{F}$ is an antiderivative of F and $\mathcal{G}$ is an antiderivative of G. Describe the possible values of $\mathcal{F}(x) - \mathcal{G}(x)$.

1.4.2

Q7 Evaluate $\sum\limits_{k=2}^{5} 3k-2$

Q8 Evaluate $\sum\limits_{j=-1}^{4} j^2-j$

Q9 Write a formula for the value of $\sum\limits_{k=a}^{b} c$.

Q10 We do not need to write a constant multiple rule for Σ notation because we already have one.

Explain what rules of mathematics tell us that $\sum\limits_{k=a}^{b} cf(k) = c\sum\limits_{k=a}^{b} f(k)$.

Q11 Explain what's wrong with the following notation:

$$\sum_{k=1}^{k} 3k^2 + \frac{1}{k}$$

Q12 Consider the sum $\sum\limits_{k=1}^{n} \frac{1}{2^k}$ for a few different values of n. Can you conjecture a formula for this sum (it will depend on n).

1.4.3

Q13 Write the following sums in Σ notation.

a $3+7+11+15+19$

b $6+12+24+48+96+192$

c $\frac{3}{4}-\frac{4}{5}+\frac{5}{6}-\frac{6}{7}+\frac{7}{8}-\frac{8}{9}$.

Q14 Write the following sums in Σ notation.

a $5-15+25-35+45-55+65-75+85-95$

b $\frac{1}{4} + \frac{4}{16} + \frac{9}{64} + \frac{16}{256} + \frac{25}{1024}$

c $\sqrt{2} + \sqrt{6} + \sqrt{12} + \sqrt{20} + \sqrt{30} + \sqrt{42} + \sqrt{56}$.

1.4.4

Q15 Does $\int_{1/2}^{1} \ln x \ dx$ compute the area under $y = \ln x$ over $\left[\frac{1}{2}, 1\right]$? Explain.

Q16 Suppose $\int_a^b f(x) \ dx < 0$. What does this tell you about the graph $y = f(x)$? Be specific.

Q17 Draw a careful graph of $y = \sqrt{x}$. Use 5 subintervals of $[1, 11]$ to estimate the area beneath the graph over $[1, 11]$. Use the left endpoints of each subinterval as the test points x_i^*.

Q18 Draw a careful graph of $y = 3x$. Use 3 subintervals of $[2, 8]$ to estimate the area beneath the graph, with the test points x_i^* being the left endpoints of each subinterval.

Q19 Draw the graph of $y = 7$. Use geometry to evaluate $\int_3 87 \ dx$.

Q20 Draw the graph of $y = \frac{x}{3} + 1$. Use geometry to evaluate $\int_{-3} 9\frac{x}{3} + 1 \ dx$.

1.4.5

Q21 Let $g(x) = \int_5^x f(t) \ dt$. What is $g'(8)$?

Q22 Let $g(x) = \int_2^x \cos t \ dt$. Is $g(x)$ increasing or decreasing at $x = 3$? Explain.

Q23 Suppose $f(x)$ is an increasing function. Is $\int_{22}^{31} f'(x) \ dx$ positive or negative?

Q24 Suppose $F(x)$ and $G(x)$ are both antiderivatives of $f(x)$. Given the following incomplete table of values, compute $\int_1^4 f(x) \ dx$.

x	1	2	3	4	5	6
$F(x)$	–	7	–	13	–	9
$G(x)$	3	–	9	–	10	5

Q25 Explain the difference between $\int f(x)\ dx$ and $\int_a^b f(x)\ dx$ in a few sentences.

Q26 Compute $\int_0^{\pi} \cos(x)\ dx$. Explain the geometric meaning of your answer in a sentence or two.

1.4.6

Q27 Compute $\int_1^8 x - \frac{3}{x}\ dx$.

Q28 Compute $\int_1^4 \frac{1}{t^{3/2}}\ dt$.

Q29 Compute $\int e^x - 6x^2\ dx$.

Q30 Compute $\int_t^0 \frac{1}{3}e^x + 5\ dx$.

Q31 Compute $\int \sqrt{t}\ dt$.

Q32 Compute $\int_{10}^2 \frac{x^2+2}{5x}\ dx$.

Q33 Compute $\int \frac{3}{5}\sin y\ dy$.

Q34 Compute $\int_0^2 x^4 - 3x + 2\ dx$.

Q35 Compute $\int_{\pi/6}^{3\pi/4} 2\cos v\ dv$.

Q36 Compute $\int_0^{\pi} 2\sin t + \cos t\ dt$.

1.4.7

Q37 Write some general rules. Suppose $F(x) + c$ is the antiderivative of $f(x)$

a. What is the antiderivative of $f(x + a)$?

b. What is the antiderivative of $f(ax)$?

Q38 Assuming that $\int_a^b f(x)\,dx$ exists, argue that it is equal to $\int_{2a}^{2b} \frac{1}{2} f\left(\frac{x}{2}\right)\,dx$, in the following two ways:

a. By appealing to an integration rule.

b. By describing the relationship between the graphs of $y = f(x)$ and $y = f\left(\frac{x}{2}\right)$. A picture might help.

1.4.8

Q39 Compute $\int e^{7x}\,dx$.

Q40 Compute $\int \sqrt{5x+3}\,dx$.

Q41 Compute $\int \cos\left(\frac{\theta}{3}\right)\,d\theta$.

Q42 Compute $\int (t-2)^6\,dt$.

Q43 Compute $\int_0^{1/4} \sin(\pi t)\,dt$.

Q44 Compute $\int_0^3 x^2 e^{x^3}\,dx$.

Q45 Compute $\int (x^5 - 2x)(5x^4 - 2)\,dx$.

Q46 Compute $\int_{\pi/4}^{3\pi/4} \cos(x)\frac{1}{\sin^2 x}\,dx$.

Section 1.4 Exercises

Chapter 2

Advanced Integration and Applications

This chapter covers a variety of methods and applications for single-variable integrals. The first two sections lay the groundwork for multivariable integration by exploring the connections between integration and geometry. One section touches on approximation methods for integrals. Other sections prepare us for our goal: applying integration to probability and statistics.

Contents

Section 2.1

Area Between Curves

Goals:

1. Use integrals to calculate the geometric area of a region.

The Fundamental Theorem of Calculus relates the change in a function to the area under a curve. Modern scientists have seized upon integration as a way to study change, whether they are measuring a chemical reaction, the position of a particle, or economic activity. The geometric applications are irrelevant to most consumers of calculus.

Historically, these methods were exciting to scholars who had been limited to area formulas for circles and triangles. Now any shape that was defined by an algebraic function was fair game. In this section we push integration beyond areas under a curve to areas bounded by two or more curves. This gives us the ability to measure a wide variety of shapes, but geometry is not our end goal. Instead the goal is to study how integration works on these oddly shaped regions. We will find that the methods of this section return to relevance when it is time to integrate functions of more than one variable.

Question 2.1.1

How Is the Integral Related to Geometric Area?

When we defined the definite integral, we were attempting to compute the area under a curve. However, our methods introduced a glitch. Consider the following example.

This region has an area of $\dfrac{38}{3}$, but $\displaystyle\int_3^8 f(x)\,dx = -\frac{38}{3}$.

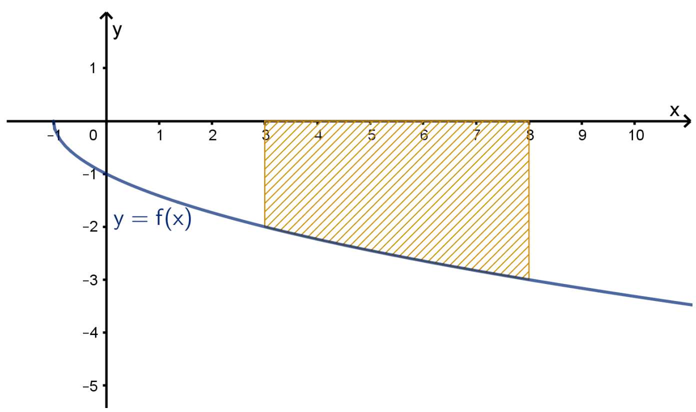


Figure: A region below the x-axis and above $y = f(x)$

We were taught that the integral does not measure geometric area, but instead **signed area**. Area below the x axis counts as negative.

Why does this happen? Recall the definition of the definite integral.

Definition

The integral is computed by the following limit

$$\int_a^b f(x)\ dx = \lim_{\Delta x \to 0} \sum_i f(x_i^*)\Delta x$$

This limit takes better and better approximations of the area. The approximation is a sum of rectangles, whose area is height $\times$ width. All the rectangles have width Δx, but their heights vary, and we used the height of the graph $y = f(x)$ to measure them. This works fine when $f(x)$ is positive. When $f(x) < 0$, the product $f(x_i^*)\Delta x$ computes a negative "area" for each rectangle.

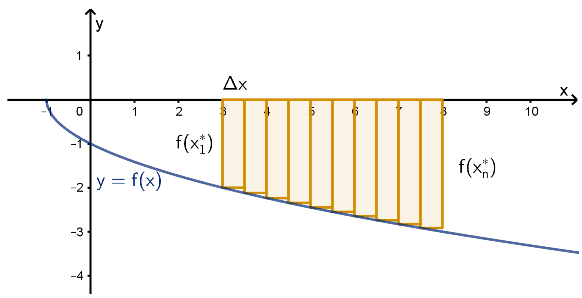


Figure: An approximation by rectangles of negative height

In this example the resolution of this glitch is straightforward. Eliminating the negative sign, we obtain the correct area. However, we can imagine a region that requires a more sophisticated approach.

Question 2.1.2
What Integral Computes the Geometric Area Between Two Graphs?

Suppose we want to know the area between the graphs $y = f(x)$ and $y = g(x)$ for some interval $a \leq x \leq b$. We can approximate this by rectangles. As the number of rectangles increases, the approximation becomes more accurate.

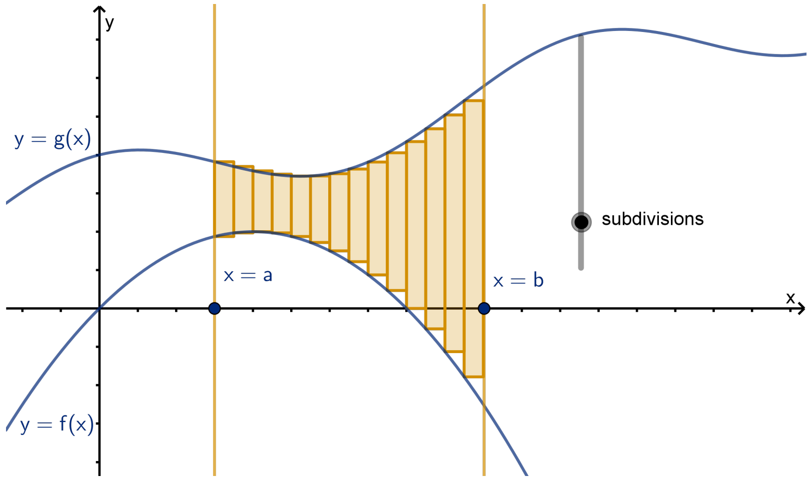
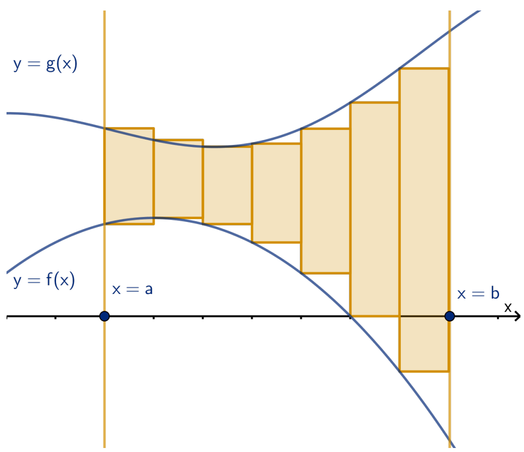


Figure: The region between $y = f(x)$ and $y = g(x)$, approximated by rectangles

Let's derive a formula for this rectangle approximation.



We let x_i^* denote the left endpoint of each subinterval. The rectangles have width Δx and height $g(x_i^*) - f(x_i^*)$. We compute:

$$\text{Area} = \lim_{\Delta x \to 0} \sum_i (g(x_i^*) - f(x_i^*))\Delta x$$

This limit exactly matches the definition of a definite integral. The function being integrated is $g(x) - f(x)$. Thus we can compute the area below $y = g(x)$ and above $y = f(x)$ by integrating $g(x) - f(x)$ from a to b.

Main Idea

The area above $y = f(x)$ and below $y = g(x)$ from $x = a$ to $x = b$ is computed

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

Example 2.1.3

The Area Between Two Curves

Suppose we want to compute the area between $y = \sqrt{x}$ and $y = x - \sqrt{x}$ from $x = 6$ to $x = 12$. How do we know which graph is on top and which is on the bottom?

The height of a graph is the value of the function. We can evaluate the function at some x in the interval $[6, 12]$. The most convenient x is $x = 9$.

$$\sqrt{9} = 3 \qquad 9 - \sqrt{9} = 6$$

So at $x = 9$, $y = x - \sqrt{x}$ is above $y = \sqrt{x}$.

Exercise

We've established that at $x = 9$, $y = x - \sqrt{x}$ is above $y = \sqrt{x}$. Unfortunately there are infinitely many points between $x = 6$ and $x = 12$. How can we decide which graph is on top at each of them?

1. Does the graph of $y = \sqrt{x}$ intersect the graph of $y = x - \sqrt{x}$ between $x = 6$ and $x = 12$?
2. What theorem could we use to argue that if $y = \sqrt{x}$ is ever above $y = x - \sqrt{x}$ then the graphs must have intersected?

Solution

1. To test where the graphs intersect, we set the functions equal to each other.

$$\begin{aligned}
\sqrt{x} &= x - \sqrt{x} \\
0 &= x - 2\sqrt{x} \\
0 &= \sqrt{x}(\sqrt{x} - 2) \qquad \text{(factor)} \\
x &= 0 \text{ or } \sqrt{x} - 2 = 0 \\
x &= 0 \text{ or } 4
\end{aligned}$$

   Neither of these is in $[6, 12]$.

2. The Intermediate Value Theorem tells us that these functions cannot switch places without intersecting. Switching places means that the difference $(x - \sqrt{x}) - (\sqrt{x})$ would change from positive to negative. As this is a continuous function, the Intermediate Value Theorem says there must be some point along the way where $(x - \sqrt{x}) - (\sqrt{x}) = 0$. We've already shown that all those points lie outside the interval, so we can conclude that $y = x - \sqrt{x}$ is above $y = \sqrt{x}$ over the entire interval $[6, 12]$.

The figure below confirms that $y = x - \sqrt{x}$ is on top for all x in $[6, 12]$.

Example 2.1.3 The Area Between Two Curves

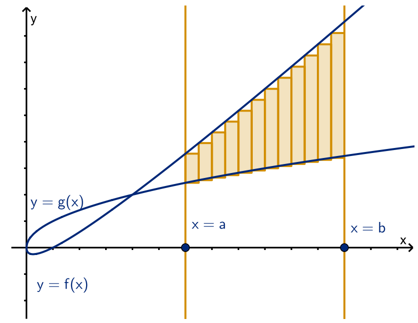


Figure: An approximation of the area between $y = x - \sqrt{x}$ and $y = \sqrt{x}$

Main Ideas

- Plugging a test point into $f(x)$ and $g(x)$ tells us which graph is above the other.
- If the functions are continuous, then solving $f(x) = g(x)$ computes the only points where the graphs can change positions.

Example 2.1.4 The Area Enclosed by Two Curves

Set up an integral that computes the area enclosed between the curves $y = x^2$ and $y = 3 - x - x^2$.

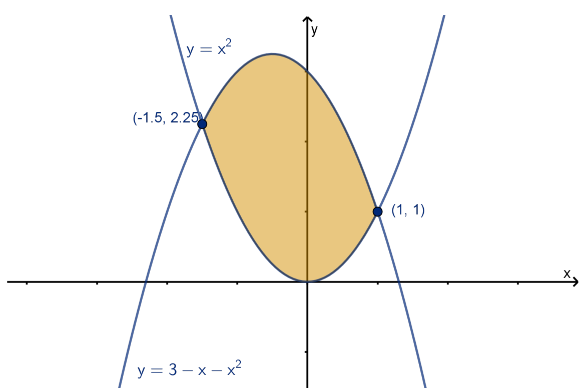


Figure: The area enclosed by two parabolas

Solution

These are parabolas. If they enclose any area, the downward facing parabola must lie above the upward facing parabola. This tells us we are integrating

$$\int_a^b 3 - x - x^2 - x^2 \, dx$$

But what are the bounds of integration? To know this we must find the points where the graphs intersect.

$$\begin{aligned} 3 - x - x^2 = x^2 \\ 0 &= 2x^2 + x - 3 \\ 0 &= (2x+3)(x-1) \\ x &= -\frac{3}{2} \text{ or } 1 \end{aligned}$$

The area is computed

$$\text{Area} = \int_{-3/2}^{1} 3 - x - x^2 - x^2 \, dx$$

Main Ideas

- To determine the range of x values that define an enclosed region, solve for the intersection points between the graphs.
- Sketching the graphs can be a time-saver and a reality check for your answer.

Example 2.1.5

The Area Enclosed by Two Curves that Intersect More than Twice

Compute the area enclosed by $f(x) = x^3 - 10x$ and $g(x) = 3x^2$.

Example 2.1.5 The Area Enclosed by Two Curves that Intersect More than Twice

Solution

To find the intersections we set $f(x) = g(x)$ and solve:

$$\begin{aligned} x^3 - 10x &= 3x^2 \\ x^3 - 3x^2 - 10x &= 0 \\ x(x-5)(x+2) &= 0 \\ x &= 0, 5, \text{ or } -2 \end{aligned}$$

Our region is bounded between $x = -2$ and $x = 5$, but one graph does not need to be above the other for the entire region. The graphs intersect at $x = 0$ so one graph might be on top for $[-2, 0]$, while the other is on top for $[0, 5]$. To find out which is which we could evaluate at test points (we would need two). Alternately, since we've already factored $f(x) - g(x) = x(x-5)(x+2)$ we can perform a sign analysis:

		-2		0		5	
x	$-$		$-$		$+$		$+$
$(x-5)$	$-$		$-$		$-$		$+$
$(x+2)$	$-$		$+$		$+$		$+$
$f(x) - g(x)$	$-$		$+$		$-$		$+$

Thus $x^3 - 10x > 3x^2$ on $[-2, 0]$ and $x^3 - 10x < 3x^2$ on $[0, 5]$. The enclosed area is computed by:

$$\begin{aligned} \text{Area} &= \int_{-2}^{0} x^3 - 10x - 3x^2 \, dx + \int_0^5 3x^2 - x^3 + 10x \, dx \\ &= \frac{x^4}{4} - 5x^2 - x^3 \Big|_{-2}^{0} + x^3 - \frac{x^4}{4} + 5x^2 \Big|_0^5 \\ &= (0 - 0 - 0 - 4 + 20 - 8) + \left(125 - \frac{625}{4} + 125 - 0 + 0 - 0\right) \\ &= \frac{407}{4} \end{aligned}$$

Main Ideas

- With more intersections, we must check the region between each pair of intersections to see which graph is on top.
- It can be more efficient to make a sign analysis chart.
- Sketching the graphs may be more difficult. If you can do it, it will corroborate (or correct) your calculations.

Example 2.1.6

A Region without a Single Top Curve

Compute the area enclosed by the curves $y = 1$, $y = \dfrac{16}{x}$ and $y = 2\sqrt{x}$.

We should start by drawing this region and finding the coordinates of the intersections. There are three intersections to solve for, one using each pair of equations.

$$\frac{16}{x} = 2\sqrt{x} \qquad\qquad \frac{16}{x} = 1 \qquad\qquad 2\sqrt{x} = 1$$

$$16 = 2x^{\frac{3}{2}} \qquad\qquad 16 = x \qquad\qquad \sqrt{x} = \frac{1}{2}$$

$$8 = x^{\frac{3}{2}}$$

$$x = 4 \qquad\qquad x = 16 \qquad\qquad x = \frac{1}{4}$$

If we write this area as an integral $\int_{\frac{1}{4}}^{16} g(x) - f(x)\, dx$, the top function would need to be piece-wise:

$$g(x) = \begin{cases} 2\sqrt{x} & \text{if } \frac{1}{4} \le x \le 4 \\ \frac{16}{x} & \text{if } 4 \le x \le 16 \end{cases}.$$

We don't know the anti-derivative of a piece-wise function. Instead, we consider a few different approaches. Since the upper boundary is defined by a different function for different values of x, one approach is to break the region into two integrals.

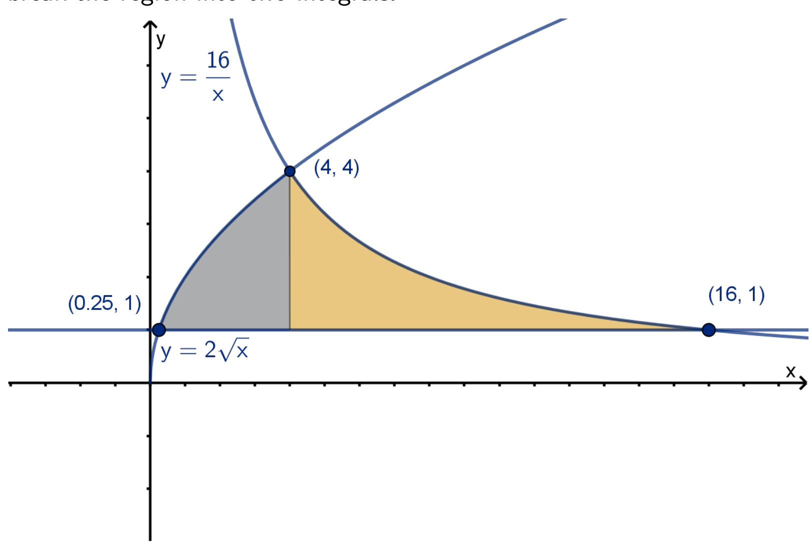


Figure: Two subregions whose areas can be expressed by integrals

The area of the region on the left is $\int_{\frac{1}{4}}^{4} 2\sqrt{x} - 1\, dx$. The are of the region on the right is $\int_4^{16} \frac{16}{x} - 1\, dx$. Adding these together gives the total enclosed area.

Another approach would be to obtain the area by subtraction. Find the following two areas on the diagram:

$$\int_{\frac{1}{4}}^{16} 2\sqrt{x} - 1\, dx \qquad \int_4^{16} 2\sqrt{x} - \frac{16}{x}\, dx$$

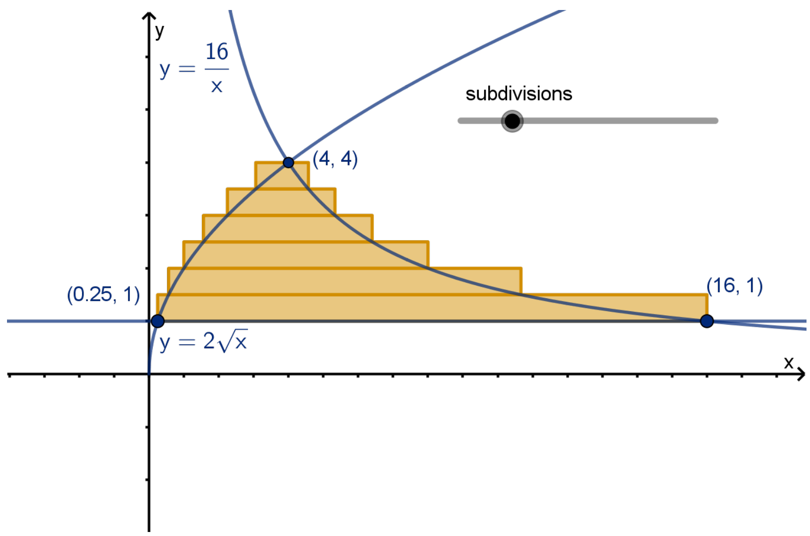
Example 2.1.6 A Region without a Single Top Curve

You should be able to convince yourself that

$$\text{Enclosed Area} = \int_{\frac{1}{4}}^{16} 2\sqrt{x} - 1 \; dx - \int_{4}^{16} 2\sqrt{x} - \frac{16}{x} \; dx$$

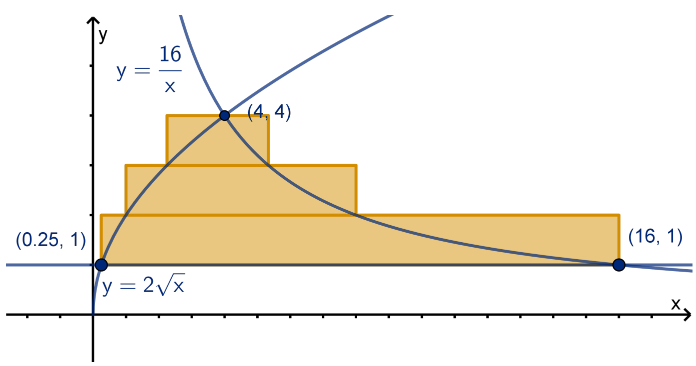
Both of these approaches require us to evaluate two integrals. That is unavoidable because our integrals are limits of an approximation by rectangles of different heights, and those heights are determined by different enclosing graphs, depending on which x value we measure at. For this particular region, there is a way to avoid this.

Instead we can approximate the region by rectangles of different widths.



Notice the left endpoint always lies on $y = 2\sqrt{x}$ and the right endpoint always lies on $y = \dfrac{16}{x}$. As the height of the rectangles goes to 0, the approximation becomes exact.

Let's derive a formula for this rectangle approximation and compute the exact area.



Let Δy be the height of each rectangle. The widths are given by the horizontal distance between the graph $y = 2\sqrt{x}$ and $y = \dfrac{16}{x}$ at the heights y_i^* corresponding to the bottom of each rectangle. Horizontal distance is the difference in x values. What x values correspond to y_i^*? We can plug in y_i^* and solve for x.

$$\begin{aligned} y_i^* &= 2\sqrt{x} & \qquad\qquad y_i^* &= \frac{16}{x} \\ \frac{y_i^*}{2} &= \sqrt{x} & xy_i^* &= 16 \\ \frac{(y_i^*)^2}{4} &= x & x &= \frac{16}{y_i^*} \end{aligned}$$

These computations should be familiar. Finding x in terms of y is called finding the **inverse function**. These inverse functions give the left and right bounds of our region. To find the area, we take a sum of the areas of these rectangles of different widths. Then we take a limit. Notice that to make the width positive we subtract the smaller x value from the larger x value. Geometrically, this is the right endpoint $\left(\frac{16}{y_i^*}\right)$ minus the left endpoint $\left(\frac{(y_i^*)^2}{4}\right)$.

$$\lim_{\Delta y \to 0} \sum_i \underbrace{\left(\frac{16}{y_i^*} - \frac{(y_i^*)^2}{4}\right)}_{\text{width}} \underbrace{\Delta y}_{\text{height}} = \int_1^4 \left(\frac{16}{y} - \frac{y^2}{4}\right)\, dy$$

This limit is an integral, but the variable of integration is y, not x. The bounds of integration are the set of y values in the region. The lowest point in the region is at $y = 1$. The highest is at $y = 4$. We evaluate the integral using the Fundamental Theorem of Calculus, but with y instead of x.

$$\begin{aligned} \text{Area Enclosed} &= \int_1^4 \left(\frac{16}{y} - \frac{y^2}{4}\right)\, dy \\ &= 16\ln|y| - \frac{y^3}{12}\bigg|_1^4 \\ &= \left(16\ln 4 - \frac{64}{12}\right) - \left(16\ln 1 - \frac{1}{12}\right) \\ &= 16\ln 4 - \frac{63}{12} \end{aligned}$$

Main Idea

The area to the right of $x = f^{-1}(y)$ and to the left of $x = g^{-1}(y)$ for y from a to b can be computed

$$\int_a^b g^{-1}(y) - f^{-1}(y)\, dy.$$

Strategy

Changing an integral to dy may be more work than breaking it into two or more parts. When solving an area problem, consider both methods and use the one that seems more promising. If you run into problems with your chosen approach, give the other method a try.

Summary Questions

Q1 What is the geometric significance of $f(x)-g(x)$ in the formula for the area between two graphs?

Q2 How do we determine which curve is the top of a region and which is the bottom? Describe the difficulties that can arise.

Q3 How do we use boundaries of the form $y = g(x)$ and $y = f(x)$ in an dy-integral to compute geometric area?

Q4 When setting up a dy-integral, how can we visually identify which graph's function will be subtracted from which?

Q5 An integral can be positive or negative. If we are solving for area (which may not be negative) describe the steps we take to guarantee our area is positive.

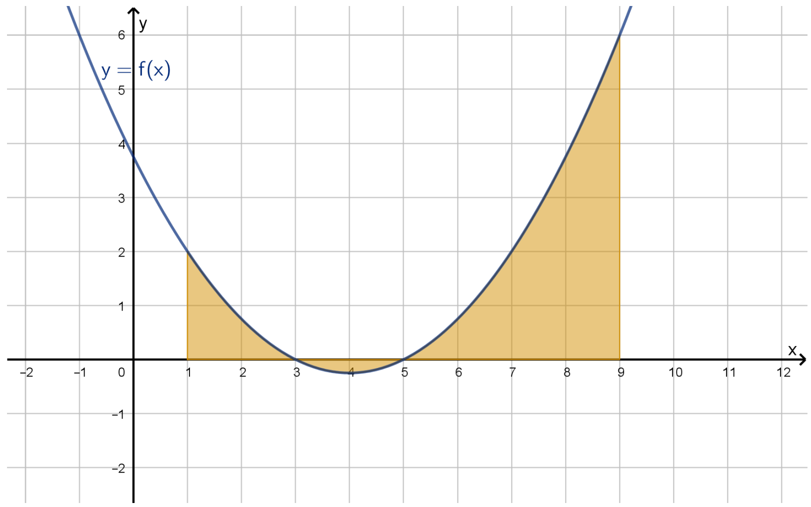
Q6 Explain the difference between "The region enclosed by $y = f(x)$ and $y = g(x)$" and "The region $f(x) \leq y \leq g(x)$."

2.1.1

Q7 Suppose the graph $y = f(x)$ is above the x-axis.

- **a** How much would the geometric area between $y = f(x)$ and the x-axis for $a \leq x \leq b$ increase if the graph were shifted up by k units. Try to argue geometrically or with a visual.
- **b** Would shifting the graph down by k instead decrease the area by the same amount? Draw a graph for which it wouldn't.

Q8 How would we use integrals to calculate the geometric area of the shaded region below?



Q9 The expressions

$$\int_a^b |f(x)|\, dx \qquad \text{and} \qquad \left| \int_a^b f(x)\, dx \right|$$

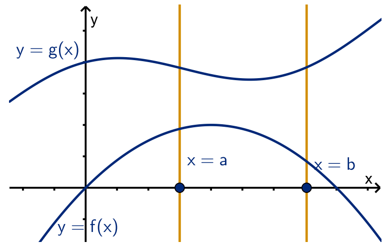
are not equivalent. Explain why, and draw the graph of a function on which these expressions disagree.

Q10 Given a differentiable function $f(x)$, the signed area between the graph $y = f'(x)$ and the x-axis from $x = a$ to $x = b$ is denoted $\int_a^b f'(x)\, dx$ and is equal to the change in $f(x)$ from $x = a$ to $x = b$. In what sense does the geometric area between the graph of $y = f'(x)$ and the x-axis represent a change in $f(x)$?

2.1.2

Q11 Suppose $y = f(x)$ and $y = g(x)$ are below the x-axis. What integral computes the geometric area between them. How does this compare to the situation when they are above the x-axis?

Q12 Here is another way to derive the formula for the area between curves. Consider the functions graphed here:



a Indicate on the graph what areas are denoted by $\int_a^b f(x)\ dx$ and $\int_a^b g(x)\ dx$. How are they related to the region between $y = f(x)$ and $y = g(x)$.

b Is $\int_a^b g(x)\ dx - \int_a^b f(x)\ dx$ equivalent to the expression for area we derived in 2.1.2? What integral rule(s) would you apply to justify this?

c If $y = f(x)$ is below the x-axis, how does this change the meaning of $\int_a^b f(x)\ dx$? Does the formula from b still work? Explain.

2.1.3

Q13 Compute the area between $y = 4x$ and $y = x^3$ from $x = 3$ to $x = 5$

Q14 Compute the area between $y = e^x$ and $y = \sin(\pi x)$ from $x = -1$ to $x = 0$

2.1.4

Q15 Compute the area enclosed by $y = \sqrt{x}$ and $y = x^2$.

Q16 Compute the area enclosed by $y = x^2 - 5$ and $y = 4x$.

Q17 Compute the area enclosed by $y = x^2$, $y = 2x - 1$ and $x = -3$.

Q18 Compute the area enclosed by $y = x + 2$ and $y = 3\sqrt{x}$.

2.1.5

Q19 Compute the area between $y = \sin x$ and $y = \cos x$ over the interval $[0, 2\pi]$.

Q20 Erica and Carter were asked to compute the area enclosed by $y = 4x$ and $y = x^3$. They agree that $4x = x^3$ when $x = -2$ and when $x = 2$. Erica thinks the area is

$$\int_{-2}^{2} 4x - x^3 \, dx$$

Carter thinks it is

$$\int_{-2}^{2} x^3 - 4x \, dx$$

a Who is correct?

b How do you think the mistake could reasonably have happened, and how can you avoid it?

Q21 Compute the area enclosed by $y = xe^{x^2}$, and $y = ex$.

Q22 Set up an integral or integrals to compute the region enclosed by the curves $f(x) = x^2(x^2 - 4)$ and $g(x) = x^4(x^2 - 4)$.

Q23 Often the top curve of an enclosed region alternates between $f(x)$ and $g(x)$ at each intersection. Can you explain what about the previous problem caused this pattern to fail?

Q24 Suppose $y = f(x)$ and $y = g(x)$ intersect multiple times, with $x = a$ their leftmost intersection and $x = b$ their rightmost. We can express the area enclosed between them by $\int_a^b |g(x) - f(x)| \, dx$.

a Explain why this formula works.

b Explain why this formula isn't partcilaularly helpful.

2.1.6

Q25 Compute the area enclosed by $y = 6$, $y = \sqrt{x}$ and $y = -2x$

Q26 Compute the area enclosed by $y = e^x$, $y = e^{4-x}$, and $y = 1$.

Q27 You have been taught at least three ways to set up an expression that will compute the area enclosed by (all of) $y = 3$, $y = 3x$, $y = 9$ and $x + y = -5$. Set up all the methods you know that will do this. You do not need to evaluate them.

Q28 Write the area in the first quadrant enclosed by $y = \sqrt{3}x$, $y = 0$, and $x^2 + y^2 = 4$ as a single integral.

Q29 Write the area enclosed by $y = \sqrt{x}$ and $y = x^2$ as

a. an integral in x

b. an integral in y

Q30 Write the area in the first quadrant enclosed by $y = x^2$, $y = 3x^2$, and $y = 18 - 3x$ as

a. a sum of integrals in x

b. a sum of integrals in y

Extension and Synthesis

Q31 Suppose you've found that $y = f(x)$ and $y = g(x)$ intersect at $x = a$ (along with perhaps other places). What could knowing the values of $f'(a)$ and $g'(a)$ tell you about where each graph is above the other? Be as specific as possible.

Q32 Suppose you are given that for all x:

- $f'(x) > 0$
- $g'(x) < 0$

We approximate area between $y = f(x)$ and $y = g(x)$ from $x = a$ to $x = b$ by rectangles, letting the x_i^* be the right endpoints of each subinterval. What can we say about whether the approximation will overestimate or underestimate the true area?

Section 2.2

Volumes

Goals:

1. Recognize cross sections of a solid object.
2. Write the area of each cross section as a function.
3. Compute the volume of a solid.
4. Visualize and compute the volume of a solid of revolution.

The motivation for the definite integral was computing an area. However, the definition turns out to be more useful than that. With the correct setup, we can express a volume as an integral as well.

Question 2.2.1

What Is Volume?

Dimension

In mathematics, we define the **dimension** of an object. Dimension measures the number of degrees of freedom available to a point traveling in the object.

The definition may not match your intuition for dimension. For example, you only encounter a parabola in two (or more)-dimensional space. However, the parabola itself is one-dimensional. If you imagine that you are an insect crawling on the parabola, you can only travel forward or backward, not side to side. If you were small enough, the parabola would seem indistinguishable from a line.

Example

1. A plane is two dimensional. You can travel left/right or up/down.
2. A circle is one dimensional. You can only travel clockwise/counterclockwise.
3. A point is zero dimensional. There is nowhere to travel within it.

We measure objects of different dimensions differently. In all cases, measuring is counting how many units of measurement fit inside the object. A 6 unit by 3 unit rectangle has area 18 square units, because 18 unit squares can fit inside it. For less regular objects we need to consider parts of square units. This requires a lot of work to do formally, but the intuition should be straightforward.

Question 2.2.1 What Is Volume?

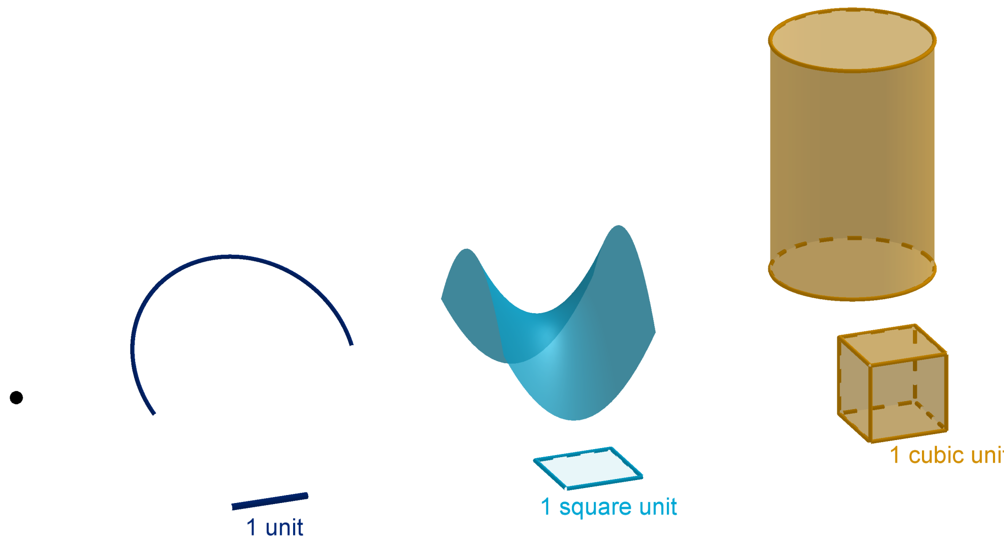


Figure: Objects of several dimensions and their units of measurement

We use different names to describe objects and their measurements in different dimensions:

Dimension	Names	Measurement
0	point	none
1	line, circle, curve	length
2	square, polygon, disc, sphere, surface	area
3	cube, polyhedron, ball, solid	volume

Vocabulary Check

It doesn't make sense to talk about the volume of a surface. No unit cubes will fit inside it.

Similarly it doesn't make sense to talk about the area of a solid. Infinitely many unit squares will fit in any solid. However, solids have boundary surfaces, and we do sometimes measure their areas.

The simplest solid to measure is a (right) prism. If a prism has height h, we can see that each unit square (or part thereof) in the base has h unit cubes stacked above it. Thus we have

Formula for Volume of a Prism

$$\text{volume} = \text{area of base} \times \text{height}$$

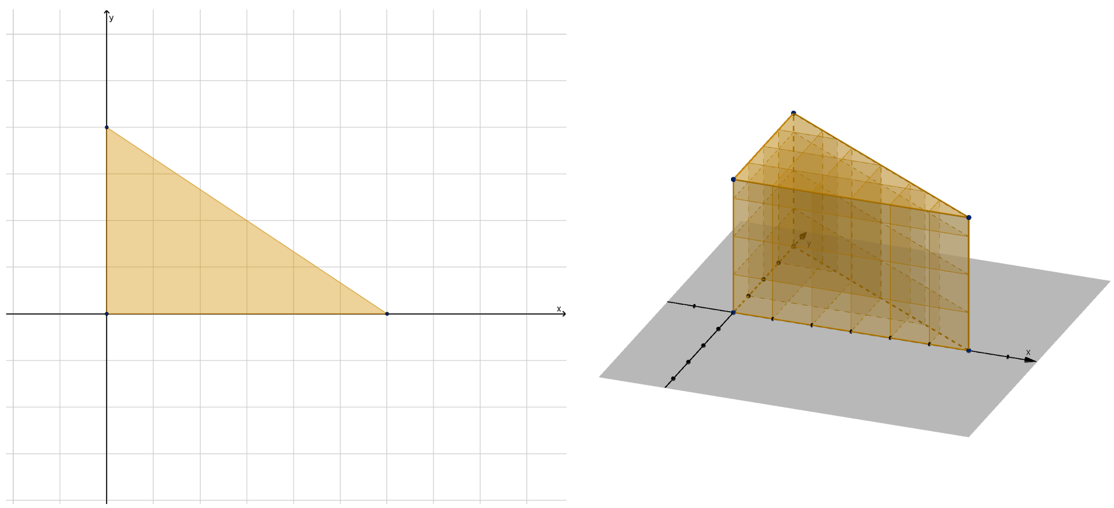


Figure: A prism divided into unit cubes and its base divided into unit squares.

Here we see the base of the prism and the square units (or parts thereof) that it contains. The prism has height 3.5. We can see there are 3.5 cubic units above each square unit in the base.

You may be questioning the relevance of studying areas and volumes in the 21st century. Few people need to compute geometric measurements in their careers. However, geometry is not the end goal of this investigation.

Remark

Our motivation for studying solids is not to solve geometry problems. Recall that the definite integral allowed us to express total change as an area:

$$\text{total change} = \text{rate of change} \times \text{time}$$

$$f(b) - f(a) = \int_a^b f'(t)\ dt$$

This allowed us to use our geometric intuition of areas to better understand rates of change. Similarly, volume will allow us to use geometry understand different types of rates later on.

Question 2.2.2

How Do We Visualize 3-Dimensional Solids?

Without computer graphics, it can be difficult to visualize anything but the simplest solids. Taking an arbitrary solid like a lamp or a sculpture, computing its volume by filling it with cubes is a hopeless endeavor (though a computer could make a decent estimate using small enough cubes). In the absence of a computer rendering, how do we give our brains a visual reference, and how can we leverage this to make measurements? We use cross sections.

Definition

A **cross section** of a solid object is its intersection with some transversal plane.

Transversal means the plane cuts across the solid. In the case of this square-based pyramid, a transversal plane parallel to the base intersects the pyramid in a square. If it intersects at a different height, the intersection would be larger or smaller. If it intersects at a different angle, it wouldn't produce a square at all.

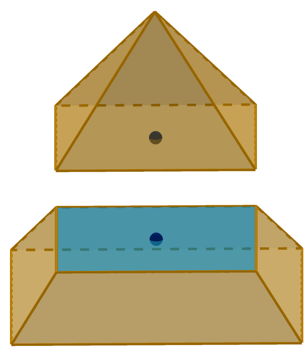


Figure: A cross section of a pyramid

A solid can be reassembled from its cross sections. This is valuable because cross sections are two-dimensional, making them easier to draw or visualize. If you have a set of parallel cross sections, you can imagine them side by side and infer the shape of the original solid.

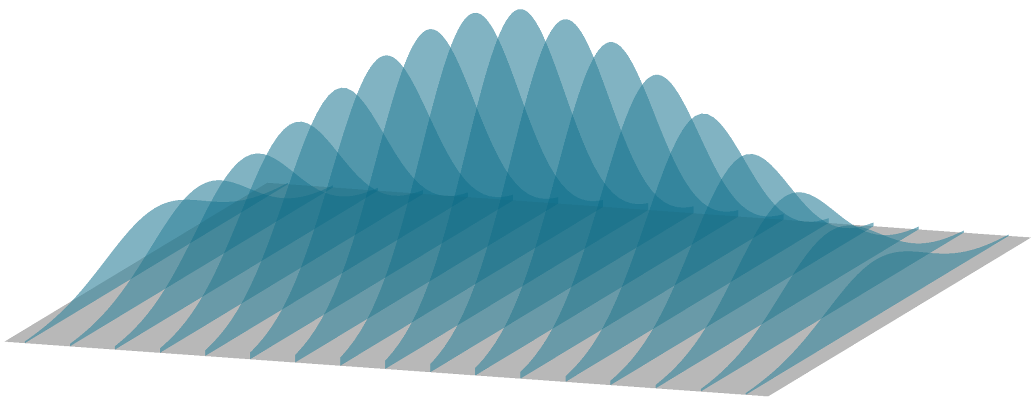


Figure: A set of parallel cross sections of a solid

Question 2.2.3
How Can We Approximate or Compute the Volume of a Non-Prism Solid?

Suppose we want to find the volume of a pyramid. Different square units of the base have a different number of cubic units above them. Thus we need a more robust approach than counting cubes.

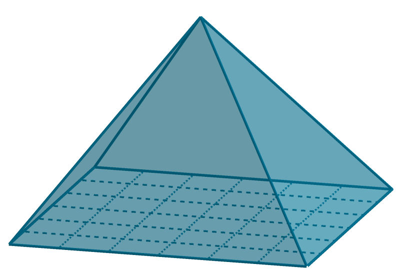


Figure: A pyramid with its base divided into unit squares

We will approximate the pyramid by prisms, whose bases are cross sections.

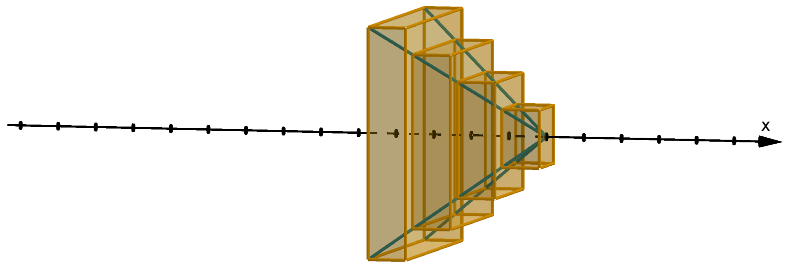


Figure: A pyramid approximated by prisms

The key insight is to represent the different heights of these cross sections by the variable x. We can imagine the x-axis running through the solid in the direction of its height. The bases of the prisms are cross sections. We let x_i^* denote the height at which the i^{th} prism's base lies. The distance between the heights x_i^* is denoted Δx, which is also the height of each prism. At different heights, we have different cross sections with different areas. Area is what we really care about, since we want to compute the volume of these prisms. We write cross sectional area as a function.

$$A(x) = \text{Area of the cross section at height } x$$

The sum of the volumes of these prisms can be written:

$$\sum_i A(x_i^*)\Delta x.$$

Taking a limit gives the exact volume of the solid:

$$\text{Volume} = \lim_{\Delta x \to 0} \sum_i A(x_i^*)\Delta x$$

Notice that this is fits the definition of a definite integral, where $A(x)$ is the function being integrated. That is excellent news for us. Instead of having to learn a new way of evaluating this limit, we can use the tools of integration that we already know.

Theorem

If the cross section of a solid, perpendicular to the x-axis, has area $A(x)$ at each x, then the volume of the solid is

$$\int_a^b A(x)\, dx$$

where a and b are the values of x at the bottom and top of the solid.

Example 2.2.4

A Solid with Its Cross-Sections Given

Suppose a solid S extends from $x = 2$ to $x = 6$ and the cross section at each x is a right triangle of height $\frac{1}{x}$ and base x^2. Compute the volume of S.

Solution

We will let the x direction be the height of our solid. Then the cross sectional area at each x is the area of the triangle at that x.

$$A(x) = \frac{1}{2}bh = \frac{1}{2}x^2\frac{1}{x} = \frac{1}{2}x$$

Integrating this from $x = 2$ to $x = 6$ gives the volume.

$$\begin{aligned} \text{Volume} &= \int_2^6 A(x)\ dx \\ &= \int_2^6 \frac{1}{2}x\ dx \\ &= \frac{1}{4}x^2\bigg|_2^6 \\ &= \frac{1}{4}36 - \frac{1}{4}4 \\ &= 8 \end{aligned}$$

The volume is 8 cubic units.

Example 2.2.5

A Solid Obtained by Rotation

Suppose the region under the graph $y = \frac{5}{x+1}$ from $x = 1$ to $x = 4$ is rotated around the x-axis. Compute the volume of the resulting solid.

Example 2.2.5 A Solid Obtained by Rotation

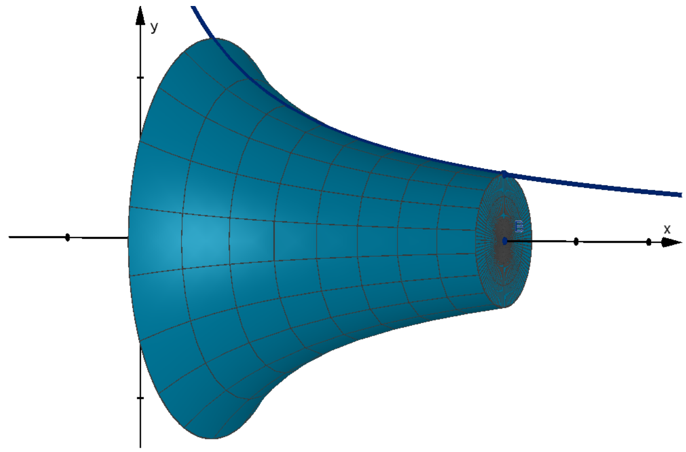


Figure: The solid obtained by rotating the region under $y = \frac{5}{x+1}$ about the x-axis

Solution

When we cut the region under the graph perpendicular to the x-axis, we obtain a line segment whose height is the value of the function. When that line segment is rotated around the axis, it sweeps out a circle, with the line segment as the radius. We can use the formula for the area of a circle.

$$A(x) = \pi r^2 = \pi \left(\frac{5}{x+1}\right)^2 = \frac{25\pi}{(x+1)^2}$$

We apply our volume formula.

$$\begin{aligned}
\text{Volume} &= \int_1^4 A(x)\, dx \\
&= \int_1^5 \frac{25\pi}{(x+1)^2}\, dx \\
&= \int_2^6 \frac{25\pi}{u^2}\, du \\
&= -\frac{25\pi}{u}\Bigg|_2^6 \\
&= -\frac{25\pi}{6} + \frac{25\pi}{2} \\
&= \frac{25\pi}{3}
\end{aligned}$$

u-substitution	
$u = x + 1$	$x = 1 \Rightarrow u = 2$
$du = dx$	$x = 5 \Rightarrow u = 6$

The volume of the solid is $\frac{25\pi}{3}$ cubic units.

Main Idea

When the region under a graph $y = f(x)$ is rotated around the x-axis, the cross sections are discs of radius $f(x)$. Their areas are $\pi[f(x)]^2$.

Example 2.2.6
A Solid Defined by Its Base

Suppose we have a solid S with the following properties:

- The base of S is the region enclosed by $y = 0$ and $y = 4x - x^2$.
- The cross-sections of S perpendicular to the x-axis are trapezoids which have one base in the base of S, another base twice as long, and whose heights are 6 units.

Compute the volume of S.

Solution

We find the x-bounds of S by computing the x-bounds of the base. We solve

$$\begin{aligned} 0 &= 4x - x^2 \\ 0 &= x(4 - x) \\ x &- 0 \text{ or } 4 \end{aligned}$$

So x ranges from 0 to 4. The base of the trapezoid at each x is the height from $y = 0$ to $y = 4x - x^2$. Note $4x - x^2 > 0$ when $0 < x < 4$. Thus the base $b_1 = 4x - x^2$. The other base is twice as long, so it is $8x - 2x^2$. The height is 6, regardless of x.

$$\begin{aligned} A(x) &= \frac{1}{2}(b_1 + b_2)h && \text{area of a trapezoid} \\ &= \frac{1}{2}(4x - x^2 + 8x - 2x^2))6 \\ &= 36x - 9x^2 \\ \text{Volume} &= \int_0^4 36x - 9x^2 \, dx \\ &= 18x^2 - 3x^3 \Big|_0^4 \\ &= 96 \end{aligned}$$

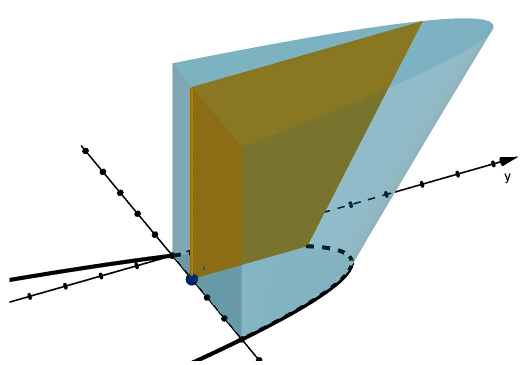


Figure: A solid with base between two graphs and trapezoidal cross-sections

Example 2.2.6 A Solid Defined by Its Base

Main Idea

The cross section of the base of a solid is a segment. If we know what role this segment plays in the cross section of the solid, we can use the expression for the length of this segment to derive an expression for $A(x)$.

Remark

Notice it is not necessary to be able to visualize the solid to compute its volume from cross sections. It is not even necessary to know what the cross-sections look like precisely. For instance, our trapezoids may or may not have a right angle. As long as we can compute the area, the exact shape is irrelevant.

Example 2.2.7 A Solid Described by Measurements

Compute the volume of a pyramid with a square base of side length s and a height of h.

Solution

Let $x = 0$ be the base of the pyramid and $x = h$ be the vertex. The cross sections are squares. Since the edges of the pyramid are straight, the squares shrink linearly from s at $x = 0$ to 0 at $x = h$. The line that goes through these two points is

$$\text{Side length} = -\frac{s}{h}x + s$$

The cross sections have area

$$A(x) = (\text{Side length})^2 = \left(-\frac{s}{h}x + s\right)^2 = s^2\left(\frac{1}{h^2}x^2 - \frac{2}{h}x + 1\right)$$

We can plug this into the formula for volume.

$$\begin{aligned}
\text{Volume} &= \int_0^h s^2\left(\frac{1}{h^2}x^2 - \frac{2}{h}x + 1\right)\,dx \\
&= s^2\left(\frac{1}{3h^2}x^3 - \frac{1}{h}x^2 + x\right)\Big|_0^h \\
&= s^2\left(\frac{1}{3h^2}h^3 - \frac{1}{h}h^2 + h - 0\right) \\
&= s^2\left(\frac{1}{3} - 1 + 1\right)h \\
&= \frac{1}{3}s^2h
\end{aligned}$$

The volume of the pyramid in cubic units is $V = \frac{1}{3}s^2h$.

Summary Questions

Q1 Describe how a cross section of a solid is produced.

Q2 What is the significance of the function $A(x)$ in the formula for the volume of a solid?

Q3 What shapes do we use to approximate the volume of a solid? Why do we choose that shape?

Q4 When we rotate the region under $y = f(x)$ around the x axis, how do we compute the area of each cross-section?

2.2.1

Q5 Which of the following shapes have (nonzero) volume?

- a square
- a ball
- a sphere
- a cube
- a cone
- a triangle

Q6 Suppose I have a solid S. I tried to fit a unit cube into S but I couldn't do it, no matter where I placed the cube or how I rotated it. I conclude that the volume of S is less than 1 unit cube. What do you think of my conclusion?

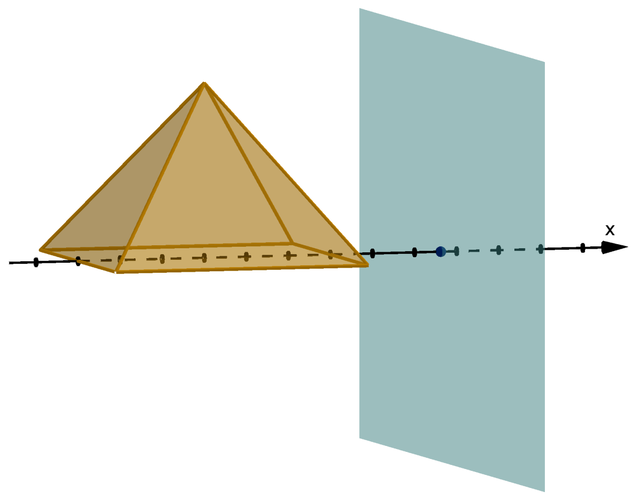
Q7 Will the volume of an object be greater is measured in cubic centimeters or cubic inches? Explain using the definition of how we measure volume.

Q8 Suppose I create a solid by stacking a cone on top of a cylinder. How is the volume of my new solid related to the volume of the cone and the volume of the cylinder? Explain using the definition of how we measure volume.

2.2.2

Q9 Let S be a sphere of radius 5 centered at the origin. What are the cross sections, perpendicular to the x-axis? How do they change as you travel along the axis from -5 to 5?

Q10 Describe or draw the cross sections of the pyramid below when it is cut by planes parallel to the one pictured.



Q11 Suppose all of the cross sections of a solid S, perpendicular to the height, are identical (same shape and same size). What kind of solid is S?

Q12 Describe the cross sections of a cube

a) perpendicular to an edge.

b) perpendicular to the line connecting the midpoints to two opposite edges.

c) perpendicular to the diagonal that connects two opposite vertices.

2.2.3

Q13 Suppose I'm trying to approximate the volume of a solid S of height 12 using four prisms of equal height. Supoose those prisms have volumes 5.1, 6, 7.2 and 9.6

a What is the approximate volume of S?

b What are the areas of the cross sections I used to produce each prism?

Q14 Suppose I'm trying to approximate the volume of the half-ball below by prisms. I subdivide the height into n subheights and use the cross section at the left hand side of each as the base of each prism. Will I overestimate or underestimate the volume? Explain how you know in a sentence or two.

Q15 Produce an approximation of the volume of a pyramid with height 9 and square base of side length 6 using 3 prisms. There are multiple correct answer to this, corresponding to different choices of where to take the cross sections.

Q16 Suppose a solid S has height 16. Suppose all of its cross-sections perpendicular to the height have a different shape, but all of those shapes have area 5.

a What is the volume of S?

b Do you really need calculus to solve **a**? Discuss.

2.2.4

Q17 Compute the volume of the solid between $x = 0$ and $x = 3$ whose cross sections at each x are squares of side length e^x.

Q18 Compute the volume of the solid between $x = 0$ and $x = 2$ whose cross sections at each x are trapezoids of bases $x + 1$ and $x + 3$ and height x^2.

Q19 Compute the volume of the solid whose cross sections, perpendicular to the x-axis, are triangles whose bases lie between $y = 3x$ and $y = x^2$ from $x = 0$ to $x = 3$ and whose heights are equal to the length of their bases.

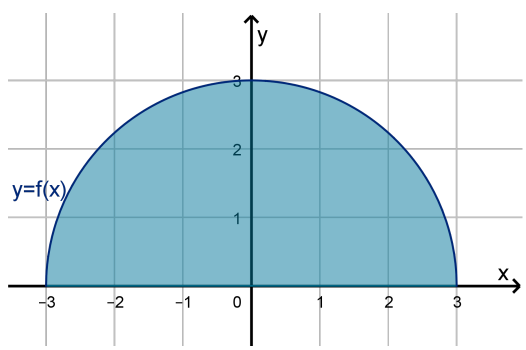
Q20 Compute the volume of a solid between $x = 1$ and $x = e^2$ whose cross sections perpendicular to the x-axis are rectangles of base $\ln x$ and height $\frac{\ln x}{x}$.

2.2.5

Q21 Compute the volume of the solid created by rotating the region under $y = \sqrt{x}$ from $x = 0$ to $x = 9$ around the x-axis.

Q22 Consider the semidisk of radius 3 below:

- **a** Write a function $y = f(x)$ that defines the boundary of this semidisk.
- **b** Suppose this semidisk is rotated around the x-axis. Describe the resulting solid.
- **c** Compute $A(x)$, the area of the cross section at each value of x.
- **d** Write and evaluate an integral that computes the volume the solid of rotation.



Q23 Compute the volume of the solid created by rotating the region $y = 4 - x^2$ from $x = -2$ to $x = 2$ about the x-axis.

Q24 Compute the volume of the solid created by rotating a trapezoid with vertices $(2, 0)$, $(5, 0)$, $(5, 8)$ and $(2, 2)$ around the x-axis.

2.2.6

Q25 Compute the volume of a solid whose base is the triangle under $y = -\frac{1}{2}x + 3$ in the first quadrant and whose cross sections, perpendicular to the x-axis are triangles of height 8.

Q26 Compute the volume of a solid whose base is the region enclosed by $y = \sqrt{x}$ and $y = \frac{x}{2}$ and whose cross sections, perpendicular to the x-axis are squares.

Q27 Compute the volume of a solid whose base is a right triangle with legs 4 and 3 and whose cross sections, perpendicular to the leg of length 4, are semicircles with their diameter in the base.

Q28 Compute the volume of a solid S whose base is the unit disc and whose cross sections perpendicular to the x-axis are isosceles right triangles, with one leg in the base.

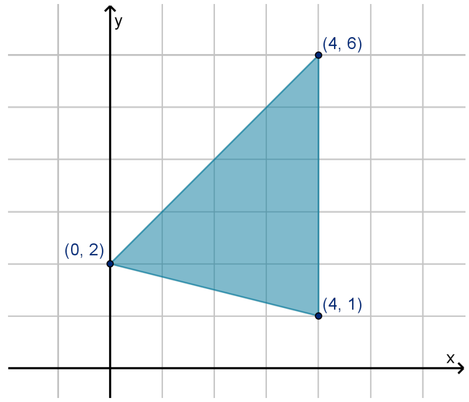
Extension and Synthesis

Q29 Let D be the region enclosed by $y = x^2 - 6x$ and the x-axis.

- **a** Set up an integral that will compute the geometric area of D. You do not need to evaluate it.
- **b** Let S be a solid whose base is D and whose cross sections perpendicular to the x-axis are semicircles with their diameter in D. Set up an integral that will compute the volume of S. You do not need to evaluate it.

Q30 Consider the solid obtained by rotating the triangle below around the x-axis.

- **a** Describe the shape of the cross sections. Which measurements of this shape depend on x?
- **b** Compute a formula for $A(x)$, the area of the cross section at each value of x.
- **c** Compute the volume of the solid.



Q31 A solid S of height 12 has the following cross sections areas $A(x)$ at height x. How would you approximate the volume?

x	$A(x)$
1	10
5	12
7	11
10	7
12	2

Section 2.3

Integration by Parts

Goals:

1. Use the integration by parts formula to find anti-derivatives and definite integrals.
2. Choose appropriate decompositions for integrating by parts.
3. Recognize when applying the formula multiple times will be fruitful.

The product rule gives us a reliable method for computing derivatives of products. If you can differentiate each factor in a product, you can differentiate the entire product. This is not the case for integration. In this section we add another tool to our limited tool set for integrating a product of two functions. Even with this method, many problems will be permanently out of reach.

Question 2.3.1

How Do We Compute an Anti-Derivative of a Product of Two Functions?

We reversed the chain rule (which computes derivatives) to compute anti-derivatives of certain functions. This method is called **u-substitution**. The du term means that we often end up integrating a product of functions with this method.

Example

Compute the integral: $\displaystyle\int_0^3 xe^{x^2}\, dx$

Solution

$$\begin{aligned}\int_0^3 xe^{x^2}\, dx &= \int_0^9 \frac{1}{2}e^u\, du \\ &= \frac{1}{2}e^u\Big|_0^9 \\ &= \frac{1}{2}(e^9 - 1)\end{aligned}$$

u-substitution	
$u = x^2$	$x = 0 \Rightarrow u = 0$
$du = 2x\, dx$	$x = 3 \Rightarrow u = 9$

Main Idea

u-substitution is extremely fragile. Our example relies on the fact that the factor x is a constant multiple of the derivative of the inner function, x^2.

Since the chain rule can only produce certain products, we should look for other differentiation rules that could produce other products. The product rule is the obvious candidate.

Question 2.3.1 How Do We Compute an Anti-Derivative of a Product of Two Functions?

Reminder

The **Product Rule** states that if $f(x)$ and $g(x)$ are differentiable, then

$$[f(x)g(x)]' = f'(x)g(x) + g'(x)f(x).$$

Example

Compute $\int x^2 \cos x + 2x \sin x \; dx$

Solution

This integrand looks like it might be the output of the product rule. If we write

$$f'(x)g(x) + g'(x)f(x) = x^2 \cos x + 2x \sin x$$

we can match up the factors as

$$f(x) = \sin x \qquad\qquad f'(x) = \cos x$$

$$g(x) = x^2 \qquad\qquad g'(x) = 2x$$

Since $\frac{d}{dx}(\sin(x)x^2) = x^2 \cos x + 2x \sin x$ we can conclude

$$\int x^2 \cos x + 2x \sin x \; dx = \sin(x)x^2 + c$$

If anything, this is more fragile than u-substitution. It requires a sum of compatible products. How can we make the formula $[f(x)g(x)]' = f'(x)g(x) + g'(x)f(x)$ more useful?

A formula that applies to a single product instead of a sum of two products would be much more useful. We can obtain it by subtracting.

$$f'(x)g(x) + g'(x)f(x) = [f(x)g(x)]' \qquad \text{product rule}$$

$$\int f'(x)g(x) + g'(x)f(x) \; dx = f(x)g(x) + c \qquad \text{integrate both sides}$$

$$\int f'(x)g(x) \; dx + \int g'(x)f(x) \; dx = f(x)g(x) + c \qquad \text{sum rule of integrals}$$

$$\int g'(x)f(x) \; dx = f(x)g(x) - \int f'(x)g(x) \qquad \text{subtract from both sides}$$

Notice we don't need the "$+c$" anymore. Both sides contain an indefinite integral so the possible constant of difference is built in on both sides. We can make one further move to simplify the equation. Since $g'(x)dx$ is the differential of $g(x)$ and $f'(x)dx$ is the differential of $f(x)$, it is convenient to represent these functions with variables. u and v are the traditional choices here.

This method is called **integration by parts**. Here is the formal statement.

Theorem

Suppose an integral can be written $\displaystyle\int u\ dv$ where

- u is a function (more precisely $u(x)$),
- and dv is a differential (more precisely $v'(x)dx$).

We can apply the following formula:

$$\int u\ dv = uv - \int v\ du$$

The integration by parts formula was not difficult to derive. The more pressing question is whether it is useful. It replaces the problem of evaluating $\int u\ dv$ with a new problem: evaluating $\int v\ du$. We need to see some examples to determine whether it is ever any help at all.

Example 2.3.2

Computing an Anti-derivative Using Integration by Parts

Compute $\displaystyle\int xe^x\ dx$.

Solution

To use integration by parts, we need to look at the integrand xe^x and decide which part is u and which part is dv. Let's try letting $u = x$ and $dv = e^x dx$. The formula says

$$\int u\ dv = uv - \int v\ du.$$

We can replace $\displaystyle\int xe^x\ dx$ by the right hand side, but we need to know what du and v are. We find du by taking the differential of u. We find v by taking the antiderivative of dv.

$$\begin{aligned} u = x \quad &\Longrightarrow \quad du = dx \\ dv = e^x\ dx \quad &\Longrightarrow \quad v = e^x \end{aligned}$$

Now we can apply the integration by parts formula.

$$\int xe^x\ dx = xe^x - \int e^x\ dx$$

Notice the integrand vdu is not a product. It is a function whose antiderivative we know. Thus integration by parts allowed us to replace a product we couldn't integrate with something we could. Evaluating the integral, we obtain:

$$\int xe^x\ dx = xe^x - e^x + c$$

Example 2.3.2 Computing an Anti-derivative Using Integration by Parts

We can always verify our antiderivatives by differentiating them. In this case

$$\frac{d}{dx}(xe^x - e^x + c) = \underbrace{xe^x + e^x(1)}_{\text{product rule}} - e^x$$
$$= xe^x$$

This verifies that we have found the correct antiderivative of xe^x.

Remark

The most general antiderivative of $dv = e^x\ dx$ would be $v = e^x + c$. However, we can get away with using a specific antiderivative instead. To convince yourself of this, try redoing the problem with $v = e^x + c$, and see that the c cancels out of your answer.

Question 2.3.3

How Do We Choose u and dv?

What would happen if we again solved $\displaystyle\int xe^x\ dx$ by parts, but set

- $u = e^x$
- $dv = x\ dx$?

In this case we compute

$$\int xe^x\ dx$$
$$= \frac{1}{2}e^x x^2 - \int \frac{1}{2}x^2 e^x\ dx$$

by parts	
$u = e^x$	$dv = x\ dx$
$du = e^x\ dx$	$v = \frac{1}{2}x^2$

This is no less correct than our previous application of the formula. It is, however, much less useful. To evaluate this we need to know an anti-derivative of $\frac{1}{2}x^2e^x$, which seems like an even harder problem than the one we started with. As we can see, the choice of u and dv can determine the success or failure of integration by parts. So what makes a good choice of u and dv?

In integration by parts, u is going to be differentiated. This usually makes functions simpler if anything. dv is going to be integrated. This could make $\displaystyle\int v\ du$ difficult to compute. The following mnemonic helps us decide which factor to choose as u and which as v.

I.L.A.T.E.

When deciding which factor of a product should be u and which should be dv, put them into the chart below.

better u's $\longleftrightarrow$ better dv's

Inverse functions	Logarithms	Algebraic expressions (polyniomials)	Trig functions	Exponential functions

Let's apply I.L.A.T.E to the following products:

1. $\displaystyle\int x^5 \ln x \, dx$

   x^5 is algebraic. $\ln x$ is a logarithm. We should let $u = \ln x$ and $dv = x^5 \, dx$.

2. $\displaystyle\int x \sin x \, dx$

   x is algebraic. $\sin x$ is trigonometric. We should let $u = x$ and $dv = \sin x \, dx$.

3. $\displaystyle\int x^2 \tan^{-1}(x) \, dx$

   x^2 is algebraic. $\tan^{-1}(x)$ is an inverse function. We should let $u = \tan^{-1}(x)$ and $dv = x^2 \, dx$.

$$
\begin{aligned}
&\int x^2 \tan^{-1}(x) \, dx \\
&= \frac{1}{3}x^3 \tan^{-1}(x) - \int \frac{1}{3}x^3 \frac{1}{1+x^2} \, dx \\
&= \frac{1}{3}x^3 \tan^{-1}(x) - \int \frac{1}{3}x^3 \frac{1}{1+x^2} \, dx \\
&= \frac{1}{3}x^3 \tan^{-1}(x) - \int \frac{1}{6}\frac{x^2}{1+x^2} 2x \, dx \\
&= \frac{1}{3}x^3 \tan^{-1}(x) - \int \frac{1}{6}\frac{u-1}{u} \, du \\
&= \frac{1}{3}x^3 \tan^{-1}(x) - \frac{1}{6}\int 1 - \frac{1}{u} \, du \\
&= \frac{1}{3}x^3 \tan^{-1}(x) - \frac{1}{6}(u - \ln|u|) + c \\
&= \frac{1}{3}x^3 \tan^{-1}(x) - \frac{1}{6}(1 + x^2 - \ln|1+x^2|) + c
\end{aligned}
$$

by parts

$u = \tan^{-1}(x)$ $\quad dv = x^2 \, dx$

$du = \frac{1}{1+x^2} \, dx$ $\quad v = \frac{1}{3}x^3$

u-substitution

$u = 1 + x^2$

$du = 2x \, dx$

Example 2.3.4

Using Integration by Parts More than Once

Compute $\displaystyle\int_0^\pi x^2 \cos x \; dx$

Solution

I.L.A.T.E. suggests $u = x^2$ and $dv = \cos x \; dx$. When we apply integration by parts to a definite integral, the $\int v \; du$ maintains the same bounds of integration. The uv is evaluated at those bounds, because it is part of the antiderivative.

$$\begin{aligned} &\int_0^\pi x^2 \cos x \; dx \\ &= x^2 \sin x \Big|_0^\pi - \int_0^\pi 2x \sin x \; dx \end{aligned}$$

by parts	
$u = x^2$	$dv = \cos x \; dx$
$du = 2x \; dx$	$v = \sin x$

Unfortunately, we don't know the anti-derivative of $2x \sin x$. It is still a product. We can try applying integration by parts again to replace $\int_0^\pi 2x \sin x$ with something we can evaluate.

$$\begin{aligned} &\int_0^\pi x^2 \cos x \; dx \\ &= x^2 \sin x \Big|_0^\pi - \int_0^\pi 2x \sin x \; dx \\ &= x^2 \sin x \Big|_0^\pi - \left[-2x \cos x \Big|_0^\pi - \int_0^\pi -2 \cos x \; dx \right] \\ &= x^2 \sin x \Big|_0^\pi + 2x \cos x \Big|_0^\pi - 2 \sin x \Big|_0^\pi \\ &= (\pi^2)(0) - (0)(0) + (2\pi)(-1) - (0)(1) - (0) + (0) \\ &= -2\pi \end{aligned}$$

by parts (again)	
$u = 2x$	$dv = \sin x \; dx$
$du = 2 \; dx$	$v = -\cos x$

Change of Variables?

Notice that despite defining functions u and v, we continue to work in terms of the variable x. Contrast this with u-substitution where the variable x can be completely eliminated in a definite integral. That approach isn't possible here. We'd have to write v as a function of u. This would be complicated or impossible.

Example 2.3.5

Using Integration by Parts to Produce an Equation

Compute $\displaystyle\int e^{2x}\cos x\ dx$

Solution

I.L.A.T.E. suggests $u=\cos x$ and $dv=e^{2x}\ dx$. To integrate dv we use a u-substitution. We apply the integration by parts formula, factoring the $-\frac{1}{2}$ from the integrand:

$$\begin{aligned}&\int e^{2x}\cos x\ dx\\ &=\frac{1}{2}e^{2x}\cos x-\int -\frac{1}{2}e^{2x}\sin x\ dx\\ &=\frac{1}{2}e^{2x}\cos x+\frac{1}{2}\int e^{2x}\sin x\ dx\end{aligned}$$

by parts	
$u=\cos x$	$dv=e^{2x}\ dx$
$du=-\sin x\ dx$	$v=\frac{1}{2}e^{2x}$

Did this help? We don't know the antiderivative of $e^{2x}\sin x$. Even worse, it doesn't seem to have improved in any way. It is just as complicated as what we started with. Our intuition might be to give up and try another approach. Perhaps I.L.A.T.E. has done us wrong and we should choose a different u and dv. In this case, however, we should reject that intuition and continue. We'll apply integration by parts again.

$$\begin{aligned}&\int e^{2x}\cos x\ dx\\ &=\frac{1}{2}e^{2x}\cos x+\frac{1}{2}\int e^{2x}\sin x\ dx\\ &=\frac{1}{2}e^{2x}\cos x+\frac{1}{2}\left[\frac{1}{2}e^{2x}\sin x-\frac{1}{2}\int e^{2x}\cos x\ dx\right]\\ &=\frac{1}{2}e^{2x}\cos x+\frac{1}{4}e^{2x}\sin x-\frac{1}{4}\int e^{2x}\cos x\ dx\end{aligned}$$

by parts again	
$u=\sin x$	$dv=e^{2x}\ dx$
$du=\cos x\ dx$	$v=\frac{1}{2}e^{2x}$

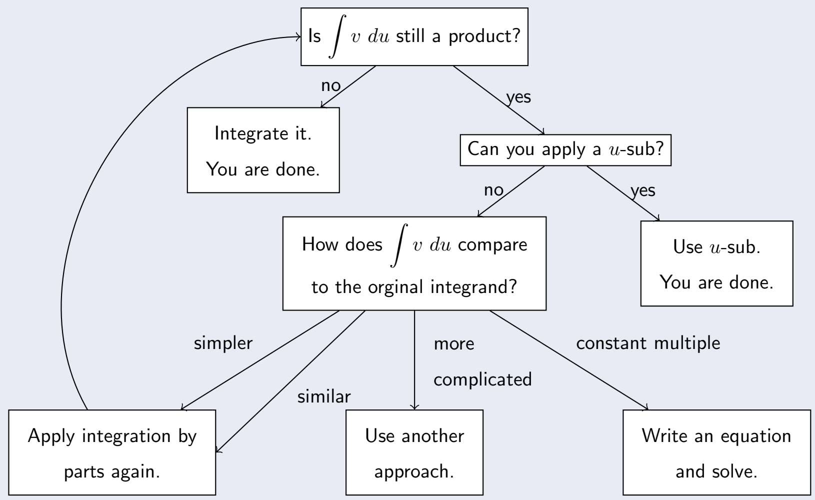
Does this help? Again the integrand does not seem to have improved, until we notice that the integrand is exactly what we began with. We could add $\frac{1}{4}\int e^{2x}\cos x\ dx$ to both sides of the equation, and we could solve for $\int e^{2x}\cos x\ dx$ algebraically.

$$\begin{aligned}\int e^{2x}\cos x\ dx&=\frac{1}{2}e^{2x}\cos x+\frac{1}{4}e^{2x}\sin x-\frac{1}{4}\int e^{2x}\cos x\ dx\\ \frac{5}{4}\int e^{2x}\cos x\ dx&=\frac{1}{2}e^{2x}\cos x+\frac{1}{4}e^{2x}\sin x+c\\ \int e^{2x}\cos x\ dx&=\frac{4}{5}\left[\frac{1}{2}e^{2x}\cos x+\frac{1}{4}e^{2x}\sin x\right]+c\\ \int e^{2x}\cos x\ dx&=\frac{2}{5}e^{2x}\cos x+\frac{1}{5}e^{2x}\sin x+c\end{aligned}$$

Example 2.3.5 Using Integration by Parts to Produce an Equation

Main Idea

We've seen a variety of techniques to apply when integration by parts does not give us an immediate answer. The success of integration by parts depends on the $\int v\, du$ term. You might use the following flow chart to decide how to proceed once you have applied integration by parts.



Section 2.3 Exercises

Summary Questions

Q1 What type of integrands are good candidates for integration by parts?

Q2 How is u handled differently in integration by parts than in u-substitution?

Q3 How is the acronym I.L.A.T.E. used?

Q4 Under what conditions would we want to apply integration by parts more than once?

2.3.1

Q5 Compute $\displaystyle\int \frac{\sin x}{1+x^2} + \cos x \tan^{-1} x \; dx$

Q6 Which of the following can be integrated using u-substitution?

$$\begin{array}{cccc}
\int e^x \; dx & \int xe^x \; dx & \int x^2e^x \; dx & \int x^3e^x \; dx \\[2ex]
\int e^{x^2} \; dx & \int xe^{x^2} \; dx & \int x^2e^{x^2} \; dx & \int x^3e^{x^2} \; dx \\[2ex]
\int e^{x^3} \; dx & \int xe^{x^3} \; dx & \int x^2e^{x^3} \; dx & \int x^3e^{x^3} \; dx \\[2ex]
\int e^{x^4} \; dx & \int xe^{x^4} \; dx & \int x^2e^{x^4} \; dx & \int x^3e^{x^4} \; dx
\end{array}$$

2.3.3

Q7 Evaluate $\displaystyle\int \frac{\ln x}{x^3} \; dx$.

Q8 Evaluate $\displaystyle\int x \sin x \; dx$.

Q9 Use integration by parts to compute $\displaystyle\int \tan^{-1} x \; dx$. Note that $\frac{d}{dx} \tan^{-1} x = \frac{1}{1+x^2}$

Q10 We can write $\displaystyle\int \ln x \; dx$ as a product: $\displaystyle\int (1)(\ln x) \; dx$.

- **a** How does I.L.A.T.E. suggest we proceed?
- **b** Use integration by parts to compute the antiderivative.

Q11 Compute $\int \sin^{-1} x \; dx$.

Q12 Compute $\int_0^{\pi/4} \tan^{-1} x \; dx$.

2.3.4

Q13 Compute $\displaystyle\int x^2 \cos(x+2)\ dx$.

Q14 Compute $\displaystyle\int_0^1 x^3 e^x\ dx$.

Q15 Compute $\displaystyle\int x^{-7} \sin(x^{-2})\ dx$. Hint: The easiest way to split this is not the correct way. You'll need some factors of x to find an antiderivative of your trig function.

Q16 Compute $\displaystyle\int_0^{\pi} x \sin x\ dx$.

2.3.5

Q17 Compute $\displaystyle\int e^{3x} \sin x\ dx$.

Q18 Compute $\displaystyle\int e^{-x} \cos 2x\ dx$.

Extension and Synthesis

Q19 Compute $\displaystyle\int x^3 e^{x^2}\ dx$. Choose your dv carefully. You want something that you can integrate.

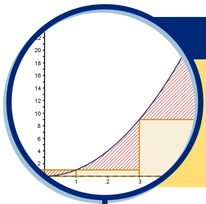
Q20 Compute $\displaystyle\int \sin(\ln x)\ dx$. Perform a u-substitution before trying by parts.

Q21 Compute the area enclosed by $y = xe^x$ and $y = ex$.

Q22 Let S be a solid between $x = 0$ and $x = 3$ whose cross-sections perpendicular to the x-axis are triangles of base x and height e^x. Compute the volume of S.

Q23 Let S be the solid obtained by rotating the region below $y = \ln x$ from $x = 1$ to $x = 5$ about the x-axis. Compute the volume of S.

Q24 Suppose that S is a solid between $x = 1$ and $x = 5$ whose cross sections (perpendicular to the x-axis) are triangles of height x^2 and base $\ln x$ at each x. Compute the volume of S.



Section 2.4

Approximate Integration

Goals:

1. Use several methods to approximate definite integrals.
2. Assess the accuracy of an approximation.
3. Approximate integrals given incomplete information.

One of the first applications of integration is to measure total change. If $v(t)$ is our velocity, $\int_a^b f(t)\,dt$ computes the total displacement between the times a and b. In practice, to evaluate such an integral, we need to know the antiderivative of f. Can we realistically expect to do this? Except in theoretical situations (say a physics experiment), we cannot. A person driving a car will not produce a velocity function that can be expressed in terms of algebra or trigonometry. While every continuous function has an antiderivative, it doesn't help us if we don't know what it is or how to evaluate it.

Our best option in these situations is to approximate the integral. For instance, if we measure velocity once per second, we could multiply each velocity by one second to approximate the distance traveled in that second. Adding these up would approximate the total displacement. What we've done is approximated the integral by rectangles of width 1. The natural question to ask is: how accurate is such an approximation? How can we make it more accurate? These are the questions we'll need to address whenever we want to apply calculus to data sets instead of abstract functions.

Question 2.4.1

What x_i^* Can We Use when Approximating an Integral?

Recall the following

Definition

The **definite integral** is given by the formula

$$\int_a^b f(x)\,dx = \lim_{\Delta x \to 0} \sum_{i=1}^{n} f(x_i^*)\Delta x$$

where Δx are the lengths of the subintervals of $[a, b]$, and x_i^* is a number in the i^{th} subinterval.

Without the limit (which is difficult or impossible to compute anyway) the sums on the right are approximations of the integral. Once we choose an x_i^* for each i, we can evaluate this approximation.

The simplest idea is to just use the left endpoint of each subinterval as x_i^*.

Question 2.4.1 What x_i^* Can We Use when Approximating an Integral?

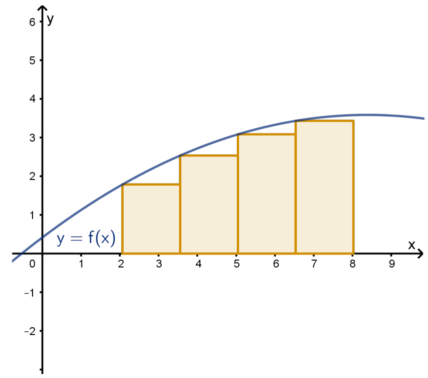
Notation

The notation L_n refers to the approximation of $\int_a^b f(x)\,dx$ by n rectangles,

$$\sum_{i=1}^{n} f(x_i^*)\Delta x,$$

where the x_i^* are the left endpoints of each subinterval.

Similarly R_n refers to the approximation using the right endpoints for x_i^*.



L_4 approximation

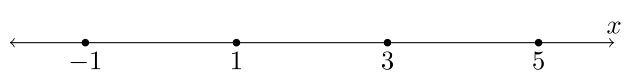
Example 2.4.2 Computing an L_n Approximation

a Compute an L_3 approximation of $\int_{-1}^{5} x^2\,dx$.

b Does L_3 over or underestimate the actual value of $\int_{-1}^{5} x^2\,dx$?

Solution

a Let $f(x) = x^2$. The interval $[-1, 5]$ has length $5 - (-1) = 6$. Three rectangles means that $\Delta x = \frac{6}{3} = 2$. We can divide up the interval to find all three subintervals. A diagram is a good way to avoid mistakes.



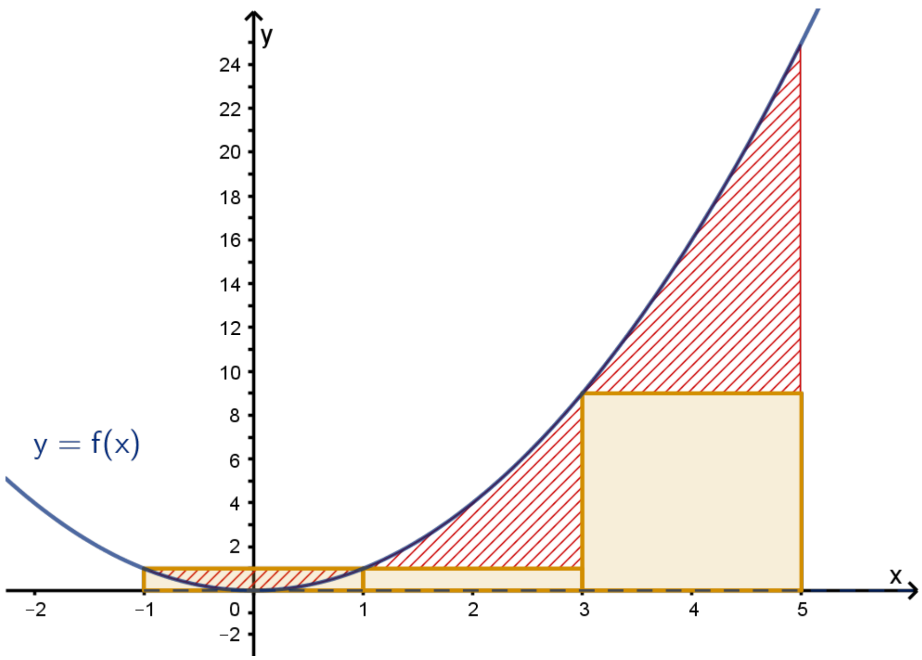
The left endpoints are -1, 1 and 3. Our approximation is

$$\begin{aligned}
L_3 &= \sum_{i=1}^{3} f(x_i^*)\Delta x \\
&= f(x_1^*)\Delta x + f(x_2^*)\Delta x + f(x_3^*)\Delta x \\
&= \Delta x(f(x_1^*) + f(x_2^*) + f(x_3^*)) \\
&= 2((-1)^2 + 1^2 + 3^2) \\
&= 22
\end{aligned}$$

b When the function increases, it has more signed area beneath it than then left-endpoint rectangles.

When it decreases it has less. $f(x) = x^2$ increases and decreases, but on the interval $[-1, 5]$, it spends much more time increasing than decreasing. Thus we expect that L_3 underestimates the true integral. We can verify our intuition with a computation.

$$\int_{-1}^{5} x^2 \, dx = \left. \frac{x^3}{3} \right|_{-1}^{5} = \frac{126}{3} > 22$$



Question 2.4.3
How Accurate is an L_n or R_n Approximation?

An approximation is much more useful, if we have some idea of how accurate (or inaccurate) it might be. The way we quantify this inaccuracy is error.

Question 2.4.3 How Accurate is an L_n or R_n Approximation?

Definitions

The **error** in an approximation is given by

$$\text{error} = \text{approximated value} - \text{actual value}$$

In a real world approximation, we do not know the exact error (why?). We will settle for putting a **bound** on error. This is a number N such that we are sure that

$$|\text{error}| \leq N.$$

Determining error bounds can be difficult. Here are some questions to ask.

1. In what circumstances is the approximation exact?
2. What property or measurement seems to correspond to the amount of error?
3. Is there a "worst case scenario" associated to that property or measurement?

The following exercise explores these questions.

Exercise

a. Draw a function for which L_n is always an overestimate.

b. Draw a function for which L_n is always an underestimate.

c. What has to be true of a function for L_n to always be exact?

d. What familiar calculus measurement appears to measure whether you are in the situations you described in a - c?

Solution

a A decreasing function will be overestimated by L_n.

b An increasing function will be underestimated by L_n.

c If L_n is always exact, then $f(x)$ is a constant function.

d Functions can be classified as increasing, decreasing or constant by their first derivative. $f'(x)$ seems to determine the sign (and maybe size) of the error.

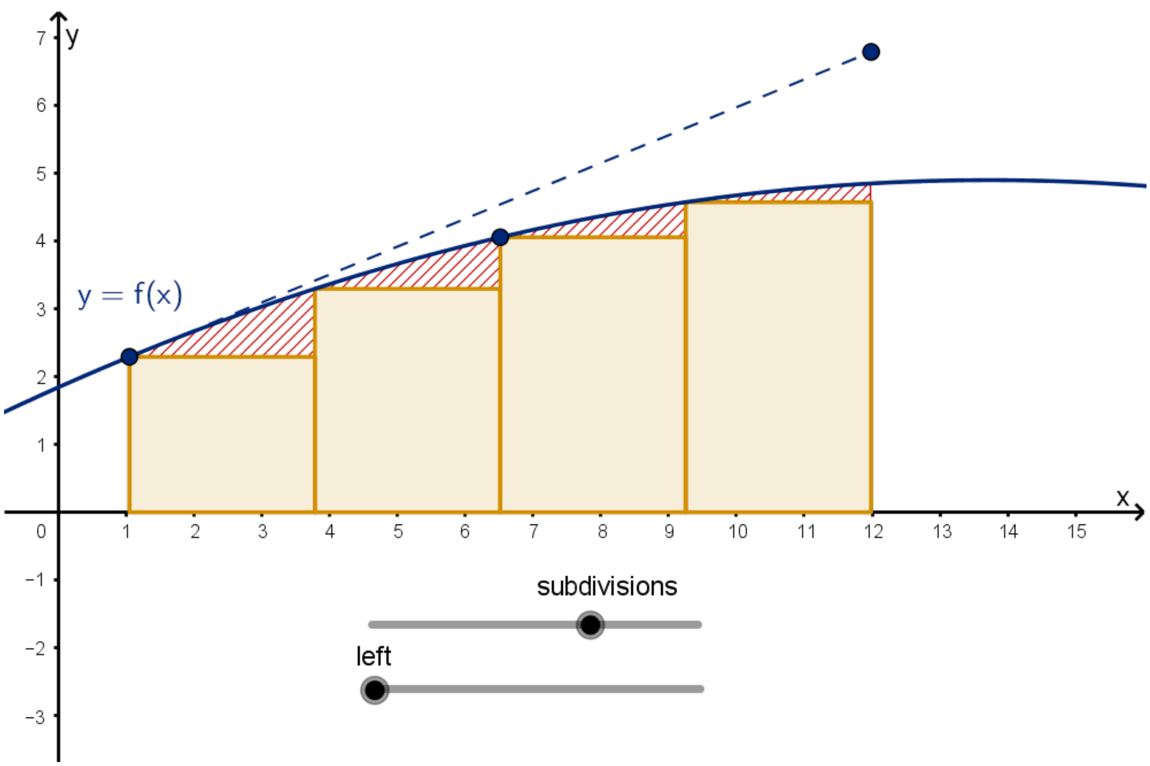


Figure: The error of an L_n approximation

Let's use the results of the exercise to formulate an error bound for L_n.

Higher derivatives seem to produce more negative errors. If we allow for steeper and steeper slopes, there is no limit to how large the error could be. So let's put a bound on how big the derivative is. Suppose we know that $f'(x) \leq S$ on $[a, b]$. Over each interval $[x_i, x_{i+1}]$ we know that $f(x)$ lies below the line of slope S through $(x_i, f(x_i))$:

$$f(x) \leq S(x - x_i) + f(x_i)$$

Question 2.4.3 How Accurate is an L_n or R_n Approximation?

The region below the graph $y = f(x)$ and above the i^{th} rectangle is smaller than the region below the line and above the rectangle, but we can compute the area of the larger region. It is a triangle. Its base is $\Delta x = \frac{b-a}{n}$. Its height can be determined by the slope of the line.

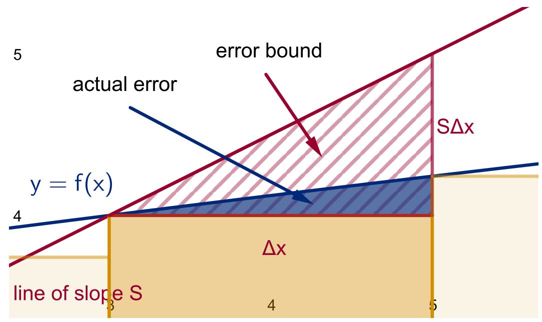


Figure: The error and the error bound over one rectangle of an L_n approximation

$$\frac{\text{height}}{\text{base}} = \frac{\text{rise}}{\text{run}} = S \qquad \text{area} = \frac{1}{2}(\text{base})(\text{height})$$

$$\frac{\text{height}}{\Delta x} = S \qquad = \frac{1}{2}S\Delta x^2$$

$$\text{height} = S\Delta x \qquad = \frac{1}{2}S\left(\frac{b-a}{n}\right)^2$$

So the error over each subinterval can be no larger than $\frac{1}{2}S\left(\frac{b-a}{n}\right)^2$. There are n subintervals, so the total L_n approximation underestimates $\int_a^b f(x)\,dx$ by no more than $\frac{S(b-a)^2}{2n}$.

We can make a similar argument that if $f'(x) \geq -S$ then L_n overestimates $\int_a^b f(x)\,dx$ by no more than $\frac{S(b-a)^2}{2n}$. We can combine these two statements into one by using absolute values. $-S \leq f'(x) \leq S$ is rewritten $|f'(x)| \leq S$.

We could make the same argument for the R_n approximation. We'd only need to swapping the overestimate with the underestimate. The error bounds it produces are the same. Our result can be stated as a theorem:

Theorem

If E_L and E_R are the errors in an L_n and R_n approximations of $\displaystyle\int_a^b f(x)\,dx$ and $|f'(x)| \leq S$ on $[a,b]$ then

$$|E_L| \leq \frac{S(b-a)^2}{2n} \qquad \text{and} \qquad |E_R| \leq \frac{S(b-a)^2}{2n}$$

Remark

The argument that the line of slope S is the "worst case" scenario is a useful heuristic, but you may be unsatisfied with its lack of rigor. A formal argument relies on the following ideas:

- Larger functions have larger integrals. If $f(x) \leq g(x)$, then $\int_a^b f(x)\ dx \leq \int_a^b g(x)\ dx$ as long as $a \leq b$.
- The Fundamental Theorem of Calculus tells us we can write $f(x) = f(x_i) + \int_{x_i}^{x} f'(t)dt$.

The line of slope S would be $L(x) = f(x_i) + \int_{x_i}^{x} S\ dt$. Over the interval $[x_i, x_{i+1}]$, comparing these integrals shows that $f(x) \leq L(x)$. Thus $\int_{x_i}^{x_{i+1}} f(x)\ dx \leq \int_{x_i}^{x_{i+1}} L(x)\ dx$. This tells us that there is more error, and thus a larger underestimate in the left hand approximation of $L(x)$ than there is in the left hand approximation of $f(x)$.

Example 2.4.4
Computing an E_L Bound

Suppose we want to understand the error of an L_n approximation of $\displaystyle\int_1^{16} \sqrt{x}\ dx$.

a What bounds can we put on $|f'(x)|$ for our error calculation?

b What bound can we put on the error of the L_5 approximation?

c What n would we need in order to guarantee that the L_n approximation has error at most $\dfrac{1}{100}$.

d What problem would result, if we tried to bound the error of an L_n approximation of $\displaystyle\int_0^{16} \sqrt{x}\ dx$? How might you resolve this?

Solution

a $f'(x) = \frac{1}{2\sqrt{x}}$. This is always positive, and it decreases as x increases. The largest value of $f'(x)$ on $[1, 16]$ occurs when $x = 1$. If we let $S = f'(1) = \frac{1}{2}$, we are guaranteed that for all x in $[1, 16]$, $|f'(x)| < \frac{1}{2}$.

Example 2.4.4 Computing an E_L Bound

b By our theorem

$$\begin{aligned}|E_L| &\le \frac{S(b-a)^2}{2n}\\ &= \frac{\frac{1}{2}(16-1)^2}{2(5)}\\ &= \frac{45}{4}\end{aligned}$$

So the error lies between $-\frac{45}{4}$ and $\frac{45}{4}$.

c We can set our error bound (with n as a variable) to be less than $\frac{1}{100}$ and solve for n.

$$\begin{aligned}|E_L| \le \frac{\frac{1}{2}(16-1)^2}{2n} &\le \frac{1}{100}\\ \frac{225}{4n} &\le \frac{1}{100}\\ (225)(100) &\le 4n\\ (225)(25) &\le n\\ 5625 &\le n\end{aligned}$$

We conclude that the error will be less than $\frac{1}{100}$ as long as n is at least 5625. Note that since this is an error bound, the actual error may shrink below $\frac{1}{100}$ with fewer rectangles. We would need a different method to verify that, though.

d If we want apply our theorem to $\displaystyle\int_0^{16} \sqrt{x}\,dx$, we need an S such that $|f'(x)| \le S$. This derivative is $f'(x) = \frac{1}{2\sqrt{x}}$, which increases without bound as $x \to 0^+$. Thus there is no S, and we cannot apply the error bound theorem.

To get around this problem we could break the interval into two parts and bound them by different methods. We can bound the error on rectangles 2 through n over the interval $[\Delta x, 16]$ using the theorem as above. In this case $S = \frac{1}{2\sqrt{\Delta x}}$ will work. To bound the error over the first rectangle $[0, \Delta x]$, note that $f(x)$ is increasing. The first rectangle of L_n will underestimate the integral, while the first rectangle of R_n will overestimate it. Thus the actual error can be no bigger than the difference between them, which is $\sqrt{\Delta x}\Delta x - 0\Delta x$. The total error can be no larger than the sum of the error bound over $[0, \Delta x]$ and the error bound over $[\Delta x, 16]$.

Question 2.4.5

How Can We Make our Approximation Less Sensitive to Slope?

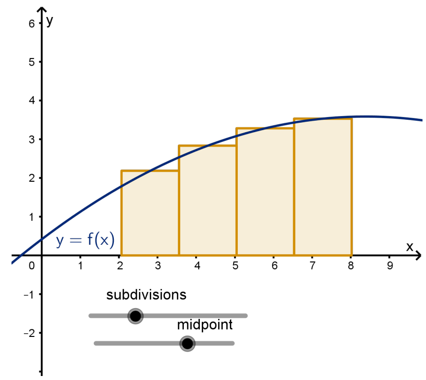
L_n and R_n have large errors when function is increasing or decreasing rapidly. We'll examine two approximations that are more resilient. The first is the midpoint approximation.

Notation

The M_n approximation of $\int_a^b f(x)\ dx$ is calculated by summing:

$$\sum_{i=1}^{n} f(x_i^*)\Delta x$$

where the x_i^* are the midpoints of each subinterval.



M_4

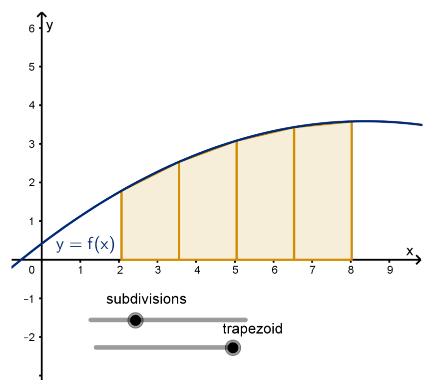
Our final approximation abandons rectangles entirely. Using trapezoids instead allows for shapes that reflect the value of the function at both the right and left endpoint. In this construction, the trapezoids are sideways from the way you may be used to looking at them when you learned their area formula $A = \frac{1}{2}(b_1 + b_2)h$. The parallel bases are vertical. The height is along the x-axis.

Notation

The T_n approximation of $\int_a^b f(x)\ dx$ is calculated by summing:

$$\sum_{i=1}^{n} \frac{1}{2}(f(x_i) + f(x_{i+1}))\Delta x$$

where x_i and x_{i+1} and the two endpoints of the i^{th} subinterval.
T_n can also be calculated as $\frac{1}{2}(L_n + R_n)$.



T_4

Example 2.4.6

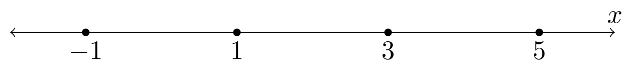
A Midpoint Approximation

Calculate the M_3 approximation of $\int_{-1}^{5} x^2\ dx$.

Solution

$\Delta x = \frac{5-(-1)}{3} = 2$. We can sketch the intervals:

Example 2.4.6 A Midpoint Approximation



The midpoints are $x_1^* = 0$, $x_2^* = 2$ and $x_3^* = 4$.

$$
\begin{aligned}
M_3 &= \sum_{i=1}^{n} f(x_i^*)\Delta x \\
&= \Delta x(f(x_1^*) + f(x_2^*) + f(x_3^*)) \\
&= 2(0^2 + 2^2 + 4^2) \\
&= 40
\end{aligned}
$$

Example 2.4.7

A Trapezoid Approximation Using a Table of Values

Approximation has no practical use for algebraic functions. We would rather get the exact answer by taking an antiderivative and applying the Fundamental Theorem of Calculus. In many real-world applications, our data about a function consists of a finite number of measurements. In this case, we don't even have an expression for the function, let alone its antiderivative. Here is an example where approximation is the best we can do.

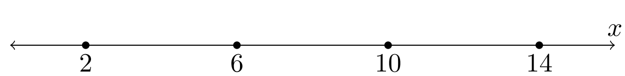
Suppose we have the following table of values for a function $f(x)$

x	0	2	4	6	8	10	12	14	16
$f(x)$	2	5	3	4	7	8	5	4	1

Calculate the T_3 approximation of $\displaystyle\int_2^{14} f(x)\ dx$.

Solution

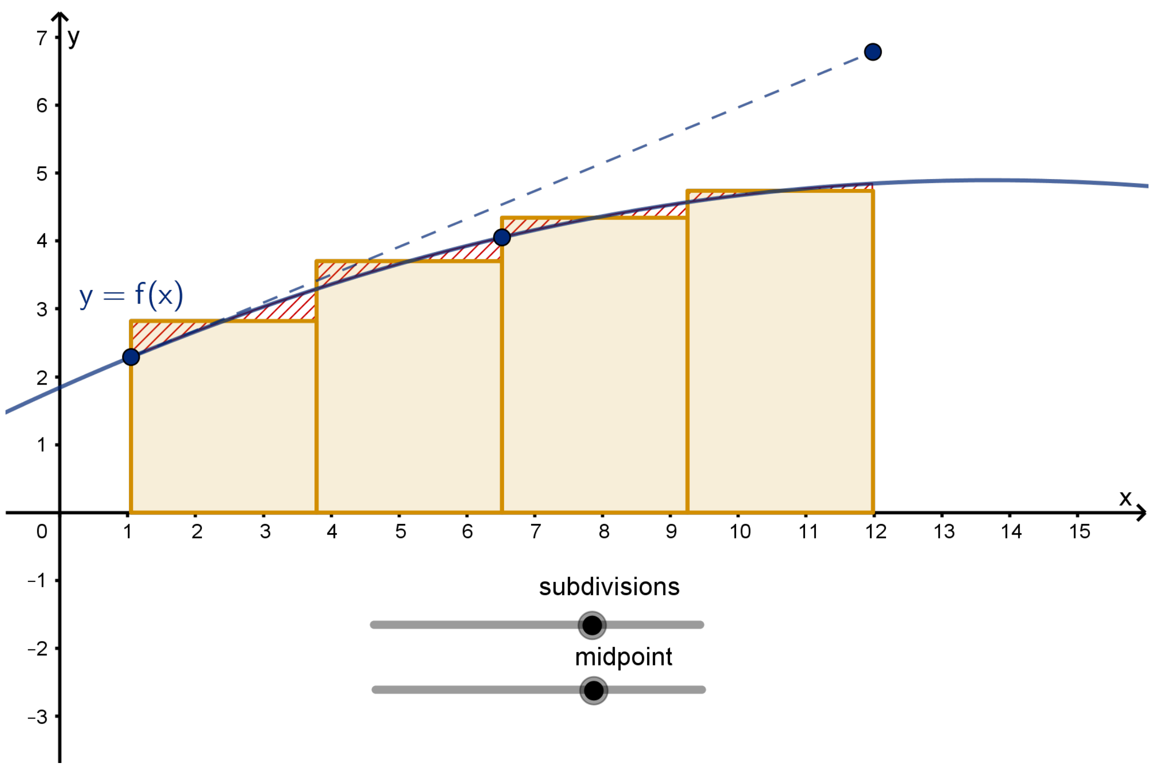
$\Delta x = \frac{14-2}{3} = 4$. We can sketch the intervals:



$$
\begin{aligned}
T_3 &= \sum_{i=1}^{3} \frac{1}{2}(f(x_i) + f(x_{i+1}))\Delta x \\
&= \frac{1}{2}\Delta x(f(x_1) + f(x_2) + f(x_2) + f(x_3) + f(x_3) + f(x_4)) \\
&= \frac{1}{2}\Delta x(f(2) + f(6) + f(6) + f(10) + f(10) + f(14)) \\
&= \frac{1}{2}(4)(5 + 4 + 4 + 8 + 8 + 4) \\
&= 66
\end{aligned}
$$

Question 2.4.8

How Do the Error Bounds of the Approximations Compare?



T_n and M_n have zero error when $f(x)$ is a straight line, regardless of slope. Larger errors result from high rates of curvature. You can see this by using a small number of rectangles/trapezoids and increasing the curvature of the function. Proving an error bound involves using a quadratic as a "worst case scenario." Any function with second derivative smaller than the quadratic will have a smaller error. Here is the result.

Question 2.4.8 How Do the Error Bounds of the Approximations Compare?

Theorem

Suppose $|f''(x)| \leq K$ for $a \leq x \leq b$. If E_T and E_M are the error in the trapezoid and midpoint approximations of $\int_a^b f(x)\,dx$ then

$$|E_T| \leq \frac{K(b-a)^3}{12n^2} \qquad \text{and} \qquad |E_M| \leq \frac{K(b-a)^3}{24n^2}$$

Remarks

1. The maximum error is smaller when the function has less curvature.
2. The error is also reduced by increasing n, the number of subintervals.
3. These formulas indicate that we can usually expect M_n to have half as much error as T_n.
4. As n increases, the error bounds for M_n and T_n approach 0 much more quickly than L_n and R_n.

Example 2.4.9

Choosing n to Meet an Error Target

Suppose we wish to approximate $\int_1^{16} \sqrt{x}\,dx$ by a midpoint approximation. How many rectangles must we use to guarantee that the error is smaller than $\frac{1}{1000}$?

Solution

The midpoint error formula requires use to have a bound K on $|f''(x)|$ on $[1, 16]$.

$$f'(x) = \frac{1}{2\sqrt{x}}$$

$$f''(x) = -\frac{1}{4x^{3/2}}$$

As x gets larger, the denominator of $f''(x)$ gets larger, meaning $|f''(x)|$ gets smaller (we could also verify this by checking the sign of $f'''(x)$). Thus it will be largest at $x = 1$. We can safely use the value there as our K

$$|f''(x)| \leq |f''(1)| = \frac{1}{4} = K$$

We can now apply the error bound formula, leaving n as a variable. We will set the error bound to be less than $\frac{1}{1000}$ and solve for n.

$$\begin{aligned}
|E_M| \leq \left|\frac{K(b-a)^3}{24n^2}\right| &\leq \frac{1}{1000} \\
\left|\frac{\frac{1}{4}(16-1)^3}{24n^2}\right| &\leq \frac{1}{1000} \\
\frac{\frac{1}{4}(16-1)^3}{24n^2} &\leq \frac{1}{1000} && \text{all factors are postive} \\
\frac{(1000)(15)^3}{(4)(24)} &\leq n^2 && \text{isolate } n^2 \\
\frac{140,625}{4} &\leq n^2 \\
\frac{375}{2} &\leq n && \text{square root of both sides}
\end{aligned}$$

Thus any n bigger than $375/2$, will work. We need to use at least 188 rectangles to guarantee that the error is less than $\frac{1}{1000}$. Note that we might achieve a sufficiently small error with fewer rectangles, but our error bound theorem can not guarantee it.

Section 2.4

Exercises

Summary Questions

Q1 How is the error in an approximation defined?

Q2 What does the first derivative of $f(x)$ tell you about the error in the right-hand approximation of $\displaystyle\int_a^b f(x)\,dx$?

Q3 As the number of subintervals gets large, which approximation(s) converge most quickly to the actual value?

Q4 Under what situation is a midpoint approximation preferable to a trapezoid approximation? When would trapezoid be preferable?

2.4.1

Q5 Seong-ju and Anthony are both approximating $\displaystyle\int_{-4}^{4} x^2\, dx$ with 4 rectangles. They know that they can use any combination of test points in their rectangles. What is the maximum difference between their approximations?

Q6 **a** What Δx and x_i^*'s would you use for the L_4 approximation of

$$\int_3^{23} f(x)\, dx?$$

b Can you write a general expression for Δx and the x_i^*'s for

$$\int_a^b f(x)\, dx?$$

2.4.2

Q7 Compute the L_5 approximation of $\displaystyle\int_1^{16} x^{3/2}\, dx$.

Q8 Compute the R_3 approximation of $\displaystyle\int_2^8 x\sin\left(\frac{\pi x}{12}\right)\, dx$.

Q9 Compute the L_4 approximation of $\displaystyle\int_0^2 x^3 e^x$.

Q10 Compute the L_5 approximation of $\displaystyle\int_3^{18} \frac{3^x}{x}\, dx$.

2.4.3

Q11 Compute the theoretical error bound on the L_{14} approximation of $\displaystyle\int_1^8 \sqrt[3]{x}\ dx$.

Q12 Compute the theoretical error bound on the R_5 approximation of $\displaystyle\int_0^{15} \frac{1}{x^2+1}\ dx$.

Q13 How large would n need to be to guarantee that the L_n approximation of $\displaystyle\int_2^8 \log_2 x\ dx$ is within $\frac{1}{10000}$ of the actual value?

Q14 How large would n need to be to guarantee that the R_n approximation of $\displaystyle\int_{-1}^2 x^3\ dx$ is within $\frac{1}{1000}$ of the actual value?

2.4.4

Q15 Suppose we make the following approximations of $\displaystyle\int_{15}^{30} 4x+7\ dx$. Without computing them, put them in order from least to greatest (some may be equal).

- L_4
- L_8
- R_4
- R_8
- M_4
- M_8
- The actual value

Q16 Yiming has a great idea. He approximates $\int_a^b f(x)\ dx$ by 12 rectangles. In order to mitigate the error of left and right hand approximations, he takes the right endpoint of the first subinterval as a test point, but the left endpoint of the second subinterval. He continues to alternate for all 12 subintervals. What is another name for the approximation Yiming has produced?

2.4.5

Q17 Compute the T_3 approximation of $\displaystyle\int_1^{16} x^2 - x\ dx$.

Q18 Compute the M_3 approximation of $\displaystyle\int_1^{16} x^2 - x\ dx$.

Q19 Compute the M_4 approximation of $\displaystyle\int_1^{9} \cos\left(\frac{\pi x^2}{12}\right)\ dx$.

Q20 Compute the T_2 approximation of $\displaystyle\int_0^{6} e^{x^2+2x}$.

2.4.6

Q21 Given the following table of values of $f(x)$

x	0	3	6	9	12	15	18	21
$f(x)$	10	13	11	15	13	11	9	12

a Compute the M_2 approximation of $\displaystyle\int_3^{15} f(x)\ dx$.

b Compute the T_3 approximation of $\displaystyle\int_0^{18} f(x)\ dx$.

Q22 Given the following table of values of $h(x)$

x	1	2	3	4	5	6	7	8	9
$h(x)$	2	-1	3	4	2	1	-3	5	4

a Compute the T_3 approximation of $\displaystyle\int_1^{9} h(x)\ dx$.

b Compute the M_3 approximation of $\displaystyle\int_2^{8} h(x)\ dx$.

Q23 Let $f(x) = \frac{1}{x^3}$. If you wanted to use a midpoint approximation with n rectangles to approximate $\int_3^5 f(x)\,dx$. How large must n be to guarantee your approximation had an error of no more than $\frac{1}{10000}$? Your answer should have the form $n \geq \ldots$, but you do not need to simplify any arithmetic.

Q24 Suppose we want to approximate $\int_1^9 \sqrt{x}\,dx$.

a Produce the T_4 approximation. Don't bother simplifying the arithmetic.

b Solve for a value n such that T_n has an error of at most $\frac{1}{1000000}$. Don't simplify the arithmetic.

Q25 Consider the following data about an unknown function $g(x)$.

x	0	2	4	6	8	10	12	14
$g(x)$	3	5	8	9	7	4	3	1

a Compute a M_3 approximation of $\int_0^{12} g(x)\,dx$.

b If you are given that $|g''(x)| < \frac{1}{4}$, what bound can you put on the error of the previous approximation?

Q26 Sasha is trying to bound the error of her M_{10} approximation of $\int_0^{\pi} \sin x\,dx$. She computes $f''(0) = 0$ and $f''(\pi) = 0$ and so decides to use $K = 0$.

a What does her choice of K imply about the accuracy of her calculation.

b Explain what is wrong with Sasha's reasoning.

c Compute the actual error bound for the M_{10} approximation.

Extension and Synthesis

Q27 Give an example of a function for which L_4 and R_4 are both overestimates on some interval. You may want to express your function by drawing its graph.

Q28 Suppose we want to estimate $\displaystyle\int_4^{20} f(x)\, dx$ and have the following table of values

x	4	6	8	10	12	14	16	18	20
$f(x)$	3	5	4	2	-1	6	2	5	8

a. What estimates are possible with this data?

b. Would you expect the M_4 or the T_8 approximation to give you a better estimate?

Q29 Consider T_3, the trapezoid approximation of $\displaystyle\int_2^8 x^3\, dx$.

a. Produce this approximation. Do not simplify the arithmetic.

b. Compute the theoretical error bound for this approximation.

c. Explain in a couple sentences how you can tell whether the error is positive or negative. You can include a diagram, if you'd like to.

Q30 Suppose you are interested in the value of $\displaystyle\int_0^{25} f(x)\, dx$, but you have only the following data.

x	1	2	6	8	13	14	20	23	25
$f(x)$	12	19	20	20	28	34	50	57	66

How might you approximate $\displaystyle\int_0^{25} f(x)\, dx$?

Q31 Suppose you invent your own approximation for a definite integral. You name it the "ultimate approximation" and denote it U_n. Its formula is

$$U_n = \frac{L_n + R_n + M_n + T_n}{4}.$$

Will U_n overestimate or underestimate the integral of a linear function? Justify your answer.

Q32 Suppose we compute an L_5 approximation of $\int_{-7}^{13} f(x)\, dx$.

a What formula that we learned would give a bound on the error of this approximation? Fill in all the information you can, and indicate the information that you would need to complete the calculation. Be as specific as possible.

b Suppose that, instead of the information you need for the formula, you were only given that f is an increasing function on $[-7, 13]$. How could you compute an error bound in this case? Justify your answer.

Section 2.5

Improper Integrals

Goals:

1. Integrate a function that has a discontinuity.
2. Recognize when an integral is improper.
3. Determine whether an improper integral converges or diverges.
4. Compute the value of an improper integral.
5. Use comparison to determine convergence.

So far we have been content to evaluate integrals of continuous functions over bounded integrals. Not all functions are continuous. We may be interested in the area under a discontinuous function, even one with a vertical asymptote. We may be interested in the area under the entire graph of a function, not just over some subset. In many cases these areas will be infinite, but in some cases they are not. We will need to develop the methods to determine which case is which.

Question 2.5.1

What Is Infinity?

In this section we'll be revisiting ideas about infinity.

Notation

The symbol ∞ implies that a variable or function is increasing without bound. It eventually gets bigger than every number.

∞ is not a number. We cannot evaluate $\dfrac{1}{\infty}$ or $\infty \cdot 0$ or $\tan^{-1}(\infty)$.

The main way that we've encountered this notation is with limits. Limits at infinity will also be relevant to improper integrals, so you may want to review them.

Exercise

Evaluate the following limits:

a $\lim_{x \to \infty} \frac{1}{x^2}$

b $\lim_{x \to \infty} \sqrt{x}$

c $\lim_{t \to -\infty} e^t$

d $\lim_{y \to \infty} \sin y$

e $\lim_{w \to \infty} \ln w$

f $\lim_{x \to -\infty} \frac{3x^2 + 7}{x^2 - 5x}$

Solution

a $\lim_{x \to \infty} \frac{1}{x^2} = 0.$

b $\lim_{x \to \infty} \sqrt{x} = \infty.$

c $\lim_{t \to -\infty} e^t = 0.$

d $\lim_{y \to \infty} \sin y$ does not exist.

e $\lim_{w \to \infty} \ln w = \infty.$

f $\lim_{x \to -\infty} \frac{3x^2 + 7}{x^2 - 5x} = 3.$

Question 2.5.2

How Do We Integrate a Discontinuous Function?

Consider the function

$$f(x) = \begin{cases} 3x^2 & \text{if } x \leq 2 \\ 10 - 2x & \text{if } x > 2 \end{cases}$$

What is $\int_0^5 f(x)\, dx$?

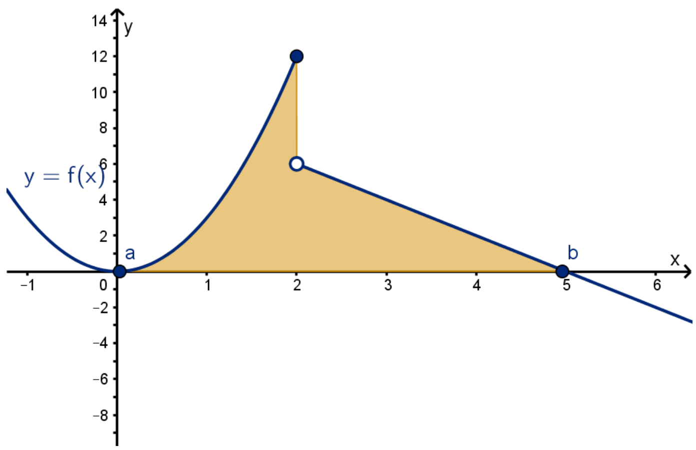


Figure: The area beneath a discontinuous graph

$\int_0^5 f(x)\, dx$ is the signed area under $f(x)$ from $x = 0$ to $x = 5$. It is equal to a limit

$$\int_0^5 f(x)\, dx = \lim_{\Delta x \to 0} \sum_{i=1}^{n} f(x_i^*)\Delta x$$

If we look at the rectangle approximations in this equation, we see that they can badly estimate the function near the point of discontinuity.

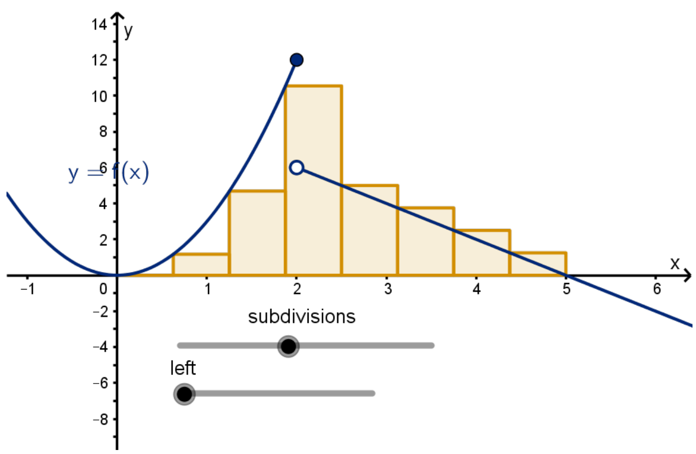


Figure: Rectangle approximations of the area beneath a discontinuous graph

Remarks

- We might worry that the approximations are so bad, that the limit $\lim_{\Delta x \to 0} \sum_{i=1}^{n} f(x_i^*)\Delta x$ does not exist. Fortunately, it does, as long as there are only finitely many discontinuities..
- $f(x)$ almost has an antiderivative function. $F(x) = \int_0^x f(t)\ dt$ has derivative $f(x)$ at all x, except perhaps at the points of discontinuity.

While it may be comforting to know that an antiderivative function exists, it doesn't help us evaluate the integral. We don't know what number to assign to $F(x)$ for many values of x. So how do we compute $\int_0^5 f(x)\ dx$? Instead of dealing with a a function whose antiderivative we don't know, we break this into two integrals that we do know.

$$\begin{aligned}\int_0^5 f(x)\ dx &= \int_0^2 f(x)\ dx + \int_2^5 f(x)\ dx \\ &= \int_0^2 3x^2\ dx + \int_2^5 f(x)\ dx\end{aligned}$$

Why can't we replace $\int_2^5 f(x)\ dx$ with $\int_2^5 10 - 2x\ dx$? At $x = 2$, $f(x) = 3x^2$, not $10 - 2x$. This is unfortunate, because for any number $t > 2$ we could replace $\int_t^5 f(x)\ dx$ with $\int_t^5 10 - 2x\ dx$. We will need to break our integral down further.

$$\begin{aligned}\int_0^5 f(x)\ dx &= \int_0^2 f(x)\ dx + \int_2^t f(x)\ dx + \int_t^5 f(x)\ dx \\ &= \int_0^2 3x^2\ dx + \int_2^t f(x)\ dx + \int_t^5 10 - 2x\ dx\end{aligned}$$

We still don't know the value of the middle integral, but we know that as t approaches 2, the domain of integration shrinks to 0. We can take advantage of this by taking a limit.

$$\begin{aligned}\int_0^5 f(x)\ dx &= \lim_{t \to 2^+} \int_0^2 3x^2\ dx + \int_2^t f(x)\ dx + \int_t^5 10 - 2x\ dx \\ &= \lim_{t \to 2^+} x^3\Big|_0^2\ dx + \int_2^t f(x)\ dx + 10x - x^2\Big|_t^5 \\ &= \lim_{t \to 2^+} 8 - 0 + \int_2^t f(x)\ dx + (50 - 25) - (10t - t^2) \\ &= \lim_{t \to 2^+} 33 - 10t + t^2 + \int_2^t f(x)\ dx \\ &= 33 - 10(2) + 2^2 + \int_2^2 f(x)\ dx\end{aligned}$$

$$= 17$$

Notice that we had to evaluate an integral with the variable t as a bound. Once we had applied the Fundamental Theorem of Calculus and plugged in t, this integral became a continuous function and we could evaluate the limit.

Notice also the strange role the limit played in this computation. Usually we take limits to see what value a changing function approaches. Our function has the same value for any choice of t (make sure you see why), so technically we were taking the limit of a constant function. The limit was a purely computational tool.

Remark

The discontinuity at $x = 2$ meant that we were stuck with an integral $\int_2^t f(x)\ dx$. With a less well-behaved function we might have also needed an integral on the left side of 2, like $\int_s^2 f(x)\ dx$. However, these two integrals can always be sent to zero by a limit, so when solving integrals of discontinuous functions, we can leave these out of our calculations.

We can summarize the method as follows:

Integrating discontinuous functions

If $f(x)$ is discontinuous at $x = c$ and $a \leq c \leq b$, then

$$\int_a^b f(x)\ dx = \lim_{t \to c^-} \int_a^t f(x)\ dx + \lim_{s \to c^+} \int_s^b f(x)\ dx$$

provided that both of these limits exist.

A removable discontinuity should not slow us down even this much. The area under a single point of discontinuity is zero. We can use the following theorem for a function with any finite number of removable discontinuities.

Theorem

If $f(x)$ and $g(x)$ are equal on $[a, b]$ except at a finite number of points, then

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

This theorem eliminates the need to use limits in our example

$$\int_0^5 f(x)\ dx = \int_0^2 \underbrace{f(x)}_{=3x^2}\ dx + \int_2^5 \underbrace{f(x)}_{\substack{=10-2x \\ \text{except at } x=2}}\ dx$$

$$= \int_0^2 3x^2\ dx + \int_2^5 10 - 2x\ dx$$

Most discontinuities can be handled this way, but there is one type that will still require limits.

Example 2.5.3

Integrating a Function with a Vertical Asymptote

Definition

When $f(x)$ has a vertical asymptote at c in $[a, b]$ we call $\int_a^b f(x)\, dx$ an **improper integral**.

How can we compute $\int_0^4 \frac{1}{\sqrt{x}}\, dx$?

In this case, breaking this integral into 2 doesn't help.

$$\int_0^4 \frac{1}{\sqrt{x}}\, dx = \lim_{t \to 0^+} \int_0^t \frac{1}{\sqrt{x}}\, dx + \int_t^4 \frac{1}{\sqrt{x}}\, dx$$

We cannot take for granted that $\lim_{t \to 0^+} \int_0^t \frac{1}{\sqrt{x}}\, dx$ goes to 0. The interval is getting smaller, but the values of the function may be so large that its rectangle approximations stay arbitrarily large and do not limit to 0. If there were an unbounded amount of area in $\lim_{t \to 0^+} \int_0^t \frac{1}{\sqrt{x}}\, dx$, then as $t \to 0^+$, $\int_t^4 \frac{1}{\sqrt{x}}\, dx$ would absorb more and more of that area and tend to ∞. Thus if (and only if) $\lim_{t \to 0^+} \int_t^4 \frac{1}{\sqrt{x}}\, dx$ exists, we can assume that the remaining piece $\int_0^t \frac{1}{\sqrt{x}}\, dx$ limits to 0 and can be ignored.

Solution

$$\begin{aligned}
\int_0^4 \frac{1}{\sqrt{x}}\, dx &= \lim_{t \to 0^+} \int_t^4 \frac{1}{\sqrt{x}}\, dx \\
&= \lim_{t \to 0^+} 2\sqrt{x}\Big|_t^4 \\
&= \lim_{t \to 0^+} 2\sqrt{4} - 2\sqrt{t} \\
&= 4 - 0
\end{aligned}$$

Since $\lim_{t \to 0^+} \int_t^4 \frac{1}{\sqrt{x}}\, dx$ exists, we conclude that

$$\int_0^4 \frac{1}{\sqrt{x}}\, dx = \lim_{t \to 0^+} \int_t^4 \frac{1}{\sqrt{x}}\, dx = 4$$

Example 2.5.3 Integrating a Function with a Vertical Asymptote

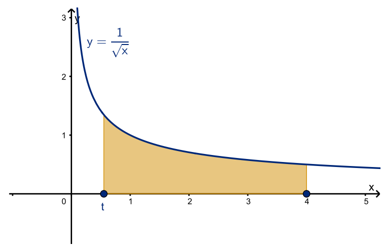


Figure: The area beneath a function with a vertical asymptote

Main Idea

To compute an improper integral, we introduce a dummy variable t and take limit(s) as $t \to c$. If the limit(s) exist, we say the integral **converges**. If any do not, we say it **diverges**.

Remark

Convergent and divergent are the terms that describe whether the limit which defines an integral approaches a single, finite numerical value. They perform a similar role to "exists" and "does not exist" for limits or "defined" and "undefined" for arithmetic.

Question 2.5.4

How Can We Compute an Integral over an Unbounded Region?

So far we have been interested in integrals over bounded intervals: $a \leq x \leq b$. We approximated these with rectangles.

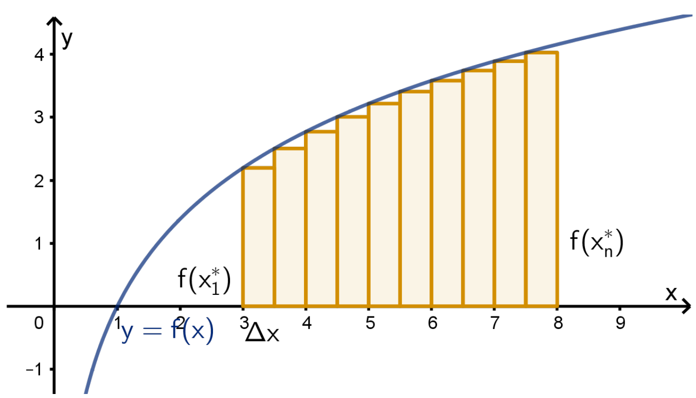


Figure: The area beneath a graph, approximated by rectangles

Consider how this approach would work with an unbounded interval: $a \le x$.

Rectangles will not approximate the area we want, but we can compute any finite subsection of it: $\int_a^t f(x)\,dx$. Like with a discontinuity, we'll take a limit.

Definition

An integral of the form $\int_a^\infty f(x)\,dx$ is also called an **improper integral**. We evaluate it by computing

$$\int_a^\infty f(x)\,dx = \lim_{t\to\infty} \int_a^t f(x)\,dx$$

assuming this limit exists. If the limit exists we say the improper integral **converges**. Otherwise we say it **diverges**.

Similarly, we can compute $\int_{-\infty}^b f(x)\,dx = \lim_{t\to-\infty} \int_t^b f(x)\,dx$.

Example 2.5.5

Evaluating an Improper Integral

Compute $\int_2^\infty \frac{32}{x^3}\,dx$.

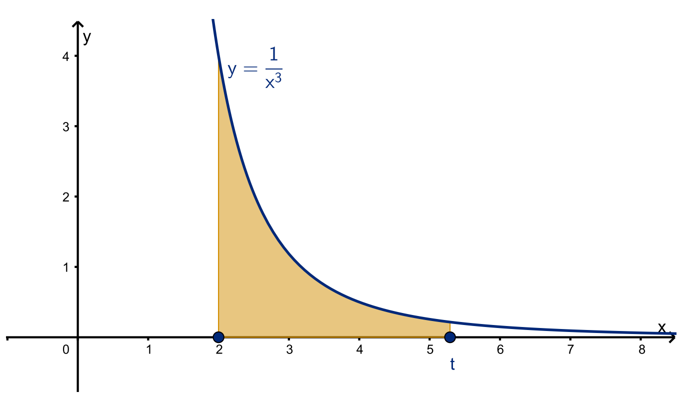


Figure: An integral over an unbounded domain

Example 2.5.5 Evaluating an Improper Integral

Solution

We'll compute the limit.

$$\begin{aligned}\lim_{t\to\infty}\int_2^\infty \frac{32}{x^3}\,dx &= \lim_{t\to\infty} -\frac{16}{x^2}\Big|_2^t \\ &= \lim_{t\to\infty} -\frac{16}{t^2}+4 \\ &= 4\end{aligned}$$

Since the limit exists, it is the value of the improper integral. $\displaystyle\int_2^\infty \frac{32}{x^3}\,dx = 4.$

Example 2.5.6
An Integral over the Entire Real Line

So far we have looked at intervals unbounded in one direction. If the interval is $(-\infty,\infty)$, the entire real line, then we use the following definition.

Definition

The improper integral $\displaystyle\int_{-\infty}^\infty f(x)\,dx$ is computed:

$$\int_{-\infty}^\infty f(x)\,dx = \int_{-\infty}^a f(x)\,dx + \int_a^\infty f(x)\,dx$$

for any number a, so long as both integrals on the right converge. If either integral diverges, then we say $\displaystyle\int_{-\infty}^\infty f(x)\,dx$ diverges as well.

Let

$$f(x) = \begin{cases} e^x & \text{if } x < 1 \\ \frac{e}{\sqrt{x}} & \text{if } x \geq 1 \end{cases}.$$

Compute $\displaystyle\int_{-\infty}^\infty f(x)\,dx$.

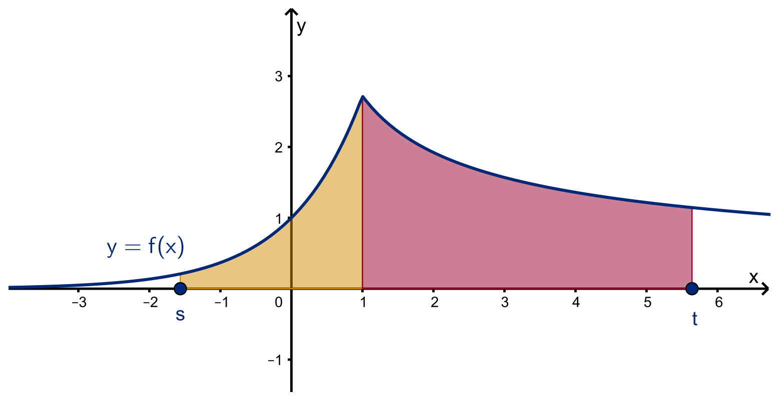


Figure: An integral over the real line, broken into two limits

Solution

We break this integral into two limits. The natural breaking point is $a = 1$ since that is where the function changes branches anyway. Both limits must converge for the integral to converge.

$$\lim_{s\to-\infty}\int_s^1 f(x)\,dx \qquad\qquad \lim_{t\to\infty}\int_1^t f(x)\,dx$$

$$\lim_{s\to-\infty}\int_s^1 e^x\,dx \qquad\qquad \lim_{t\to\infty}\int_1^t \frac{e}{\sqrt{x}}\,dx$$

$$= \lim_{s\to-\infty} e^x\Big|_s^1 \qquad\qquad = \lim_{t\to\infty} 2e\sqrt{x}\Big|_1^t$$

$$= \lim_{s\to-\infty} e - e^s \qquad\qquad = \lim_{t\to\infty} 2e\sqrt{t} - 2e$$

$$= e \qquad\qquad = \infty \text{ (diverges)}$$

One limit converges to e. The other diverges. This means that $\int_{-\infty}^{\infty} f(x)\,dx$ diverges.

Question 2.5.7

Can We Take a Limit of $\int_{-t}^{t} f(x)\,dx$ Instead?

We might wonder whether we need to break an integral $\int_{-\infty}^{\infty} f(x)\,dx$ into two integrals. Instead of two dummy variables, one going to $-\infty$ and one going to ∞, could we replace them by one? The integral $\displaystyle\int_{-\infty}^{\infty} x^3\,dx$ is a useful test case. We can certainly compute

$$\begin{aligned}\lim_{t\to\infty}\int_{-t}^{t} x^3\,dx &= \lim_{t\to\infty} \left.\frac{x^4}{4}\right|_{-t}^{t} \\ &= \lim_{t\to\infty} \frac{t^4}{4}-\frac{t^4}{4} \\ &= \lim_{t\to\infty} 0 \\ &= 0\end{aligned}$$

This might even seem right because the area above the axis seems to cancel out the area below the axis. However, intuitively, we expect that the area of a region should be preserved if we shift it in some direction. Let's shift this graph one unit to the left.

$$\begin{aligned}\lim_{t\to\infty}\int_{-t}^{t} (x+1)^3\,dx &= \lim_{t\to\infty} \left.\frac{(x+1)^4}{4}\right|_{-t}^{t} \\ &= \lim_{t\to\infty} \frac{(t+1)^4}{4}-\frac{(-t+1)^4}{4} \\ &= \lim_{t\to\infty} \frac{t^4+4t^3+6t^2+4t+1}{4}-\frac{t^4-4t^3+6t^2-4t+1}{4} \\ &= \lim_{t\to\infty} -2t^3-2t \\ &= -\infty\end{aligned}$$

We can see that, for any choice of t, there will be more area below the graph than above, and the difference grows quickly as t increases. If the area of a region changes when we shift it to the side, then that area was not well defined to begin with. We thus say that these integrals diverge, not because they go to ∞ or $-\infty$, but because they are not defined at all. The formal definition above handles this example correctly. $\displaystyle\int_{-\infty}^{0} x^3\,dx$ diverges, so $\displaystyle\int_{-\infty}^{\infty} x^3\,dx$ also diverges.

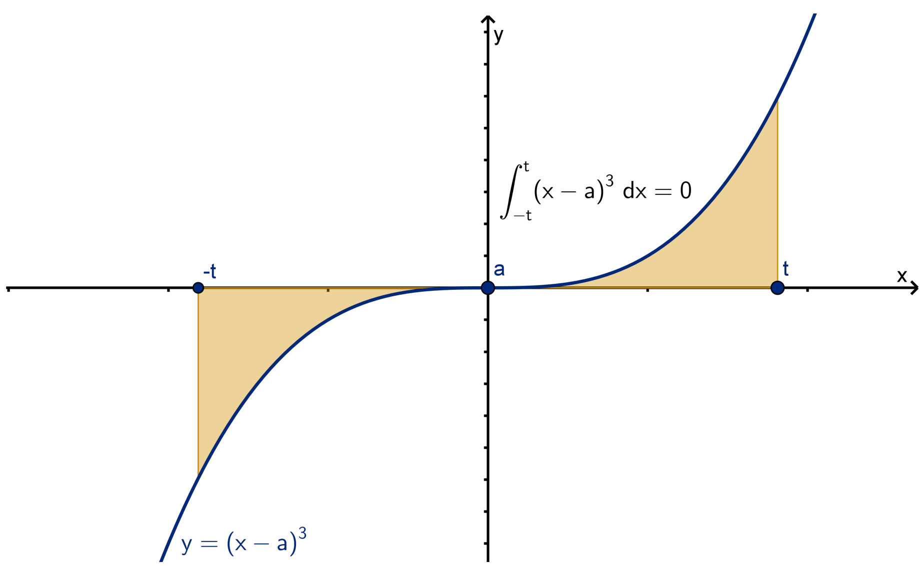


Figure: The area under a functions of the form $f(x) = (x-a)^3$

Main Idea

Do not replace the correct definition:

$$\lim_{t\to-\infty}\int_t^a f(x)\ dx + \lim_{t\to\infty}\int_a^t f(x)\ dx$$

with the "shortcut:"

$$\lim_{t\to\infty}\int_{-t}^t f(x)\ dx$$

The "shortcut" can suggest that the integral converges, when in fact it diverges.

Synthesis 2.5.8

A Comparison Test

Recall the following theorems

Theorem

If $f(x) \le g(x)$ on $[a, b]$ then $\displaystyle\int_a^b f(x)\ dx \le \int_a^b g(x)\ dx$.

Theorem

Let a be a real number or $\pm\infty$. If $F(x) \le G(x)$ for all x near a, then $\displaystyle\lim_{x\to a} F(x) \le \lim_{x\to a} G(x)$.

Suppose we have a function $f(x)$ whose anti-derivative we don't know, and a function $g(x)$ whose anti-derivative we do know. What can the divergence or convergence of $\displaystyle\int_a^\infty g(x)\ dx$ tell us about $\displaystyle\int_a^\infty f(x)\ dx$?

Solution

If we know that $f(x) \le g(x)$ then for all $t \ge a$, $\int_a^t f(x)\ dx \le \int_a^t g(x)\ dx$. This allows us to also compare their limits, which are the improper integrals: $\int_a^\infty f(x)\ dx$ and $\int_a^\infty g(x)\ dx$. This could be useful in a couple ways.

- If $\lim_{t\to\infty} \int_a^t g(x)\ dx = -\infty$ then $\lim_{t\to\infty} \int_a^t f(x)\ dx = -\infty$ as well, meaning $\int_a^\infty f(x)\ dx$ diverges.
- If on the other hand $f(x) \ge g(x)$ and $\lim_{t\to\infty} \int_a^t g(x)\ dx = \infty$ then $\lim_{t\to\infty} \int_a^t f(x)\ dx = \infty$ as well, which also means $\int_a^\infty f(x)\ dx$ diverges.
- We might like to reverse these and say that if $\int_a^\infty g(x)\ dx$ converges, $\int_a^\infty f(x)\ dx$ must as well, but $\int_a^\infty f(x)\ dx$ can diverge without going to infinity. $f(x)$ could oscillate between positive and negative so that $\int_a^t f(x)\ dx$ increases and decreases and does not have a limit as $t \to \infty$.

We can actually solve the last issue adding the assumption that $f(x)$ is non-negative. The result is not easy to prove, but it is useful.

Theorem

Suppose $0 \le f(x) \le g(x)$ for all x.

- If $\int_a^\infty f(x)\ dx$ diverges, $\int_a^\infty g(x)\ dx$ diverges.
- If $\int_a^\infty g(x)\ dx$ converges, then $\int_a^\infty f(x)\ dx$ converges.

There are similar versions of this theorem for integrals to $-\infty$ or for functions that are non-positive.

Summary Questions

Q1 What is an improper integral?

Q2 Under what conditions were we able to conclude that $\int_a^b f(x)\ dx = \int_a^b g(x)\ dx$?

Q3 What does it mean for an improper integral to converge or diverge?

Q4 If we know that $\int_a^\infty g(x)\ dx$ converges, what condition on $f(x)$ would guarantee that $\int_a^\infty f(x)\ dx$ converges?

2.5.1

Q5 In the expressions below, which of the boxes can legally be replaced by an ∞ symbol?

$$\lim_{x\to\boxed{1}} x+\boxed{2}=\boxed{3} \qquad \int_0^{\boxed{4}} f(x)\ dx = e^{\boxed{5}}+\frac{1}{\boxed{6}} \qquad x^2+2x-\log_{\boxed{7}}|x|\Big|_1^{\boxed{8}}$$

Q6 Evaluate $\displaystyle\lim_{x\to\infty}\sqrt[4]{x^3-2x+1}$.

Q7 Evaluate the following limits:

a $\displaystyle\lim_{x\to\infty}\frac{x^2+3x+5}{e^x}$

b $\displaystyle\lim_{x\to-\infty}\frac{x^2+3x+5}{e^x}$

Q8 Evaluate $\displaystyle\lim_{w\to\infty}\ln\left(\frac{1}{w}\right)$.

2.5.2

Q9 Evaluate $\int_0^3 \frac{x^2}{x}\, dx$. Explain how you dealt with any discontinuities.

Q10 Let

$$f(x) = \begin{cases} 4 & x = 1, 4, \text{ or } 6 \\ 2 & \text{otherwise} \end{cases}.$$

a. Sketch the graph $y = f(x)$.

b. Evaluate $\displaystyle\int_0^5 f(x)\, dx$. State what tool you used to deal with any discontinuities.

Q11 Let

$$g(x) = \begin{cases} \sqrt{x} & \text{if } 0 \le x \le 4 \\ 3 & \text{if } 4 < x < 6 \\ \frac{1}{x^2} & \text{if } 6 \le x \end{cases}.$$

Compute $\displaystyle\int_1^8 g(x)\, dx$.

Q12 The sign function has the form

$$\sigma(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \end{cases}.$$

Write a formula (in terms of a and b) for $\displaystyle\int_a^b \sigma(x)\, dx$. Your answer will be a piecewise expression.

2.5.3

Q13 Consider the integral $\displaystyle\int_{-2}^2 \frac{1}{x}\, dx$.

a. Sketch the graph of $y = \frac{1}{x}$.

b. Set up the limits that would compute this integral.

c. Do these limits exist?

Q14 Evaluate $\displaystyle\int_0^1 \ln x \; dx$.

Q15 Evaluate $\displaystyle\int_0^4 \frac{1}{\sqrt{x}} + \frac{1}{\sqrt{4-x}} \; dx$.

Q16 Evaluate $\displaystyle\int_0^3 \frac{2}{w^2} \; dw$.

2.5.4

Q17 How large will the base (Δx) of each rectangle be, if we want to approximate:

- **a** The area over the interval $[4, 16]$ with 3 rectangles?
- **b** The area over the interval $[a, b]$ with n rectangles?
- **c** The area over the interval $[a, \infty)$ with n rectangles?

Q18 Compute $\displaystyle\int_3^\infty \frac{2}{x} \; dx$.

Q19 Compute $\displaystyle\int_{-\infty}^0 e^x \; dx$.

Q20 Evaluate $\displaystyle\int_0^\infty e^{-2x} \; dx$.

Q21 Evaluate $\displaystyle\int_0^1 \ln x \; dx$. You may need l'Hôpital's rule.

Q22 Compute $\displaystyle\int_3^\infty \frac{1}{x^3} \; dx$, showing all necessary steps.

2.5.5

Q23 Compute

$$\int_{-\infty}^{\infty} xe^{-x^2}\,dx.$$

Q24 Show how to evaluate $\displaystyle\int_{-\infty}^{\infty} x^{1/3}\,dx$ or show that it diverges.

Q25 Let

$$f(x)\begin{cases} \frac{1}{x^3} & \text{if } x<-2 \\ \frac{1}{(x+4)^2} & \text{if } x\geq -2 \end{cases}.$$

Evaluate $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx$.

Q26 How would you write $\displaystyle\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx$ as a sum of two limits? You might recall that $\displaystyle\int \frac{1}{1+x^2}\,dx = \tan^{-1}x + c$. Use this to evaluate the integral.

Extension and Synthesis

Q27 Let

$$f(x)\begin{cases} \sqrt[3]{x} & \text{if } x<8 \\ 10-x & \text{if } x\geq 8 \end{cases}.$$

a Is $f(x)$ continuous? Justify your answer with a calculation

b What is the area enclosed by $y=f(x)$ and $y=0$?

Q28 Let

$$f(x)\begin{cases} x^{-4/3} & \text{if } x<-8 \\ \frac{1}{\sqrt[3]{x}} & \text{if } -8\leq x<0 \\ e^{-x} & \text{if } x\geq 0 \end{cases}.$$

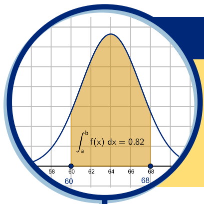
Evaluate $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx$.

Q29 Consider the region R below $y = \frac{1}{x}$, above $y = 0$ and to the right of $x = 1$.

- **a** Try to compute the area of R using an integral.
- **b** Suppose R is rotated around the x-axis to create a solid S. Compute the volume of S.
- **c** How annoying are the conclusions of **a** and **b**?

Q30 Consider the region in the first quadrant whose boundary is the curves $y = \frac{3}{x}$, $y = 2x - 1$ and $y = 0$.

- **a** Write the area of this region as an integral in the variable y. Do not evaluate.
- **b** Suppose this region is rotated around the x-axis. Write the resulting volume using one or more integrals. Do not evaluate.



Section 2.6

Probability

Goals:

1. Test the properties of a probability density function.
2. Use probability density function to describe the underlying random variable.
3. Use the uniform, exponential, and normal distributions.
4. Compute probabilities and expected values.

The main problem facing every planner is uncertainty. When will the next epidemic strike? Will the stock market go up or down? How many rare particles will flow through a detection device? These outcomes cannot be known ahead of time, but they can be modeled as probabilities. Knowing when the epidemic is likely to happen can guide our decision of how much to invest in mitigation. Knowing how many particles are likely to pass through an area can inform us how sensitive our detection device needs to be.

On the other hand, probabilities can also help us understand what has already happened. Probabilities tell us whether the results of an experiment are likely to be a coincidence. Is an apparent pattern just the variation inherent in random sampling, or is it likely to be present if the procedure is repeated? This is in fact the basic model for statistical reasoning:

1. Assume that the type of pattern you're looking for does not exist (a null hypothesis).
2. Collect observations.
3. Compute the probability of seeing those observations, given your assumption.
4. If the probability is very low, then the assumption is probably false.

Such reasoning allows us to conclude that survey is representative of the population as a whole. It allows us understand what outcome will occur on average, or how much outcomes are likely to vary. Such statistics help us understand the way the world works. We can design our next experiment or plan our future behavior around that understanding. For example, on average, the stock market goes up. This is one of the most powerful financial facts available to long-term investors, and it can be grounded in a probabilistic study of past performance.

Question 2.6.1

What Is a Continuous Probability Distribution?

Definition

A **random variable** encodes the possible outcomes of a random selection. We use the notation $P(\text{outcome})$ to denote the probability that a particular outcome occurs. If an outcome is impossible, we write $P(\text{outcome}) = 0$. If it is certain we write $P(\text{outcome}) = 1$.

Example

Our outcome can be any expression concerning the random variable, for instance:

- If S is the sum of the rolls of two six-sided dice, then

$$P(S = 8) = \frac{5}{36}.$$

- If T is the number of tails when two coins are flipped then

$$P(T \geq 1) = \frac{3}{4}.$$

We can encode these probabilities with a **distribution function**. The value of the function at each number a is the probability that the outcome is a.

Example

If T is the number of tails obtained from two fair coins then

$$f_T(t) = \begin{cases} \frac{1}{4} & \text{if } t = 0 \\ \frac{1}{2} & \text{if } t = 1 \\ \frac{1}{4} & \text{if } t = 2 \\ 0 & \text{if } t = \text{anything else} \end{cases}$$

Notice

- The sum of the probabilities adds to 1.
- There are only finitely many values of T that are possible.

What if we wanted to model height with a random variable? No one is exactly 68 inches tall. Even people who say they are "five feet eight inches" are slightly taller or shorter. A distribution function like we made for coins is unsuitable. It would have the property $f_H(h) = 0$ for all h. To handle this situation, we need to define a different kind of random variable with a different relationship to a defining function.

Question 2.6.1 What Is a Continuous Probability Distribution?

Definition

A **continuous random variable** X is a random variable whose outcomes are real numbers, and whose probability is modeled by a **probability density function** $f_X(x)$ such that

$$P(a \le X \le b) = \int_a^b f_X(x)\ dx.$$

$f_X(x)$ must satisfy

1. $f_X(x) \ge 0$ for all x.
2. $\displaystyle\int_{-\infty}^{\infty} f_X(x)\ dx = 1$

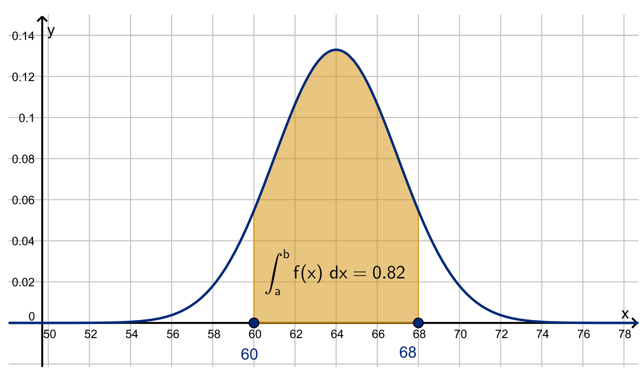
Remark

The term **density** should give us a hint about how to think about these functions. Density is a rate. The value of a probability density function tells you the rate of likelihood per unit of length on the real number line. Integrating this rate over an interval gives the total likelihood of lying on that interval, much like integrating a rate of change over an interval computes the total change.

An integral is the natural way to measure probability. The rules of integration are compatible with our intuition of probability. Suppose we have an interval $[a, b]$ broken into two or more subintervals. The total probability of X having an outcome in $[a, b]$ is equal to the sum of the probabilities of the outcome lying in each subinterval. Similarly, the area above $[a, b]$ and below the graph $y = f(x)$ is equal to the sum of the areas above each subinterval. In equations, these are the laws:

$$P(a \le X \le c) + P(c \le X \le b) = P(a \le X \le b)$$

$$\int_a^c f_X(x)\ dx + \int_c^b f_X(x)\ dx = \int_a^b f_X(x)\ dx$$



Example 2.6.2

Describing a Random Variable from its Density Function

Consider the function

$$f_X(x) = \begin{cases} \frac{1}{9}x^2 & \text{if } 0 \le x \le 3 \\ 0 & \text{if } x > 3 \text{ or } x < 0 \end{cases}$$

a Verify that f_X is a probability density function.

b If f_X is the density function of X, compute $P(X \ge 2)$.

c What does f_X tell us about the likely values of X?

Solution

a We need to check that $f_X(x)$ is never negative and $\displaystyle\int_{-\infty}^{\infty} f_X(x)\ dx = 1$

- $f_X(x)$ is never negative, because it is either a square or 0.
- $$\begin{aligned}\int_{-\infty}^{\infty} f_X(x)\ dx &= \int_{-\infty}^{0} f_X(x)\ dx + \int_{0}^{3} f_X(x)\ dx + \int_{3}^{\infty} f_X(x)\ dx \\ &= \int_{-\infty}^{0} 0\ dx + \int_{0}^{3} \frac{1}{9}x^2\ dx + \int_{3}^{\infty} 0\ dx \\ &= \frac{1}{27}x^3\Big|_0^3 \\ &= \frac{1}{27}(27-0) \\ &= 1\end{aligned}$$

b

$$\begin{aligned}P(x \ge 2) &= \int_{2}^{\infty} f_X(x)\ dx \\ &= \int_{2}^{3} f_X(x)\ dx + \int_{3}^{\infty} f_X(x)\ dx \\ &= \int_{2}^{3} \frac{1}{9}x^2\ dx + \int_{3}^{\infty} 0\ dx \\ &= \frac{1}{27}x^3\Big|_2^3\end{aligned}$$

$$= \frac{1}{27}(27 - 8)$$
$$= \frac{19}{27}$$

c Outcomes outside of $[0, 3]$ are impossible. Among the outcomes in $[0, 3]$, outcomes closer to 3 are more likely than outcomes closer to 0, because the density function has a greater value there.

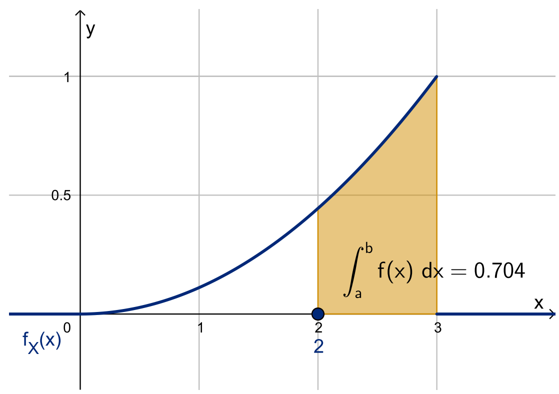


Figure: The density function of X and the area representing $P(X > 2)$

Main Ideas

- To verify that a function is a probability density function, we need to check that it is never negative and that it integrates, over the entire real line, to 1.
- We compute the probability that X has an outcome in an interval by integrating $f_X(x)$ over that interval.
- Outcomes of X where $f_X(x)$ is large are more likely than outcomes where $f_X(x)$ is small.

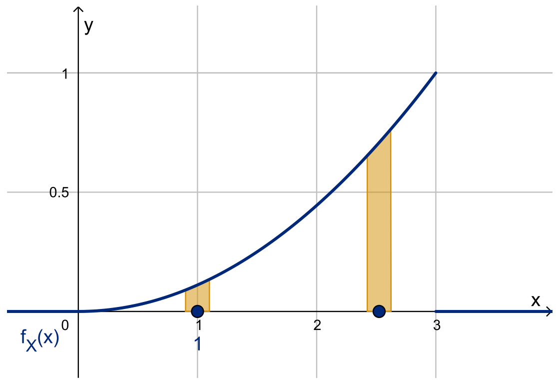


Figure: The density function of X and the areas that represent the likelihood of larger and smaller outcomes

Question 2.6.3
What Density Functions Arise Naturally?

The requirements to be a probability density function are not very strict. The vast majority of probability density functions do not model a real life phenomenon or even an intriguing thought experiment. What follows are three families of density functions that are especially useful. The first is the simplest. When we lack data to suggest otherwise, it is a common choice when creating a model with some randomness.

Definition

Given an interval $[a, b]$, the **uniform distribution** on $[a, b]$ is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{if } x > b \text{ or } x < a \end{cases}$$

Notice that the shorter the interval $[a, b]$ is, the higher density is required to integrate to a total probability of 1.

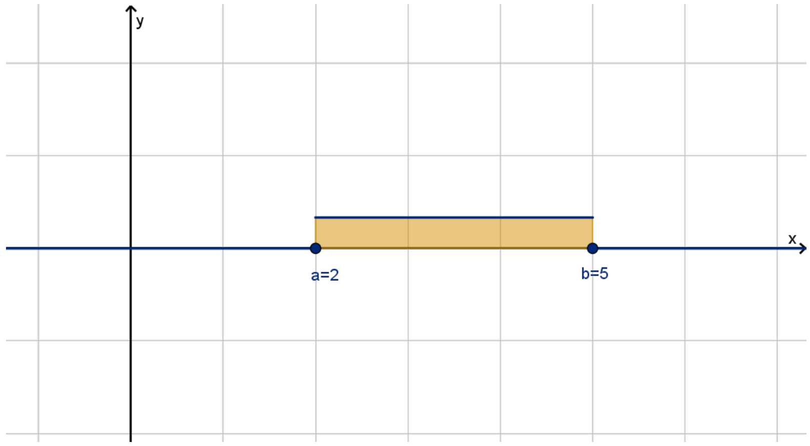


Figure: The density function of a uniform distribution

An intuitive but imprecise way to describe a random variable with a uniform distribution is to say that all outcomes in $[a, b]$ are equally likely. Since every outcome of a continuous random variable occurs with probability 0, this is unhelpful. X is remarkable, because all outcomes in $[a, b]$ have equal probability density. To connect this to actual probabilities, we might say that all subintervals of $[a, b]$ are equally likely to contain the outcome of X, but this is incorrect. X is 3 times as likely to have an outcome in an interval of length 6 as an interval of length 2. A precise statement would be: the likelihood of the outcome of X occurring in each subinterval of $[a, b]$ is proportional to the length of the subinterval.

Our second family of random variables naturally measures waiting time. This answer questions like: when will the next customer come in? When will this device next detect a certain type of ambient particle? Here is the formal definition.

Definition

Suppose an event happens randomly and uniformly at an average rate of λ times per unit of time (x). Then the amount of time until it next occurs is given by the **exponential distribution**:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } 0 \leq x \\ 0 & \text{if } x < 0 \end{cases}$$

Observe the following

1. Higher λ means that X is likely to be smaller, as the event occurs sooner.
2. The probability of the event occurring in given interval, given that it did not occur before that interval, depends only on the length of the interval.

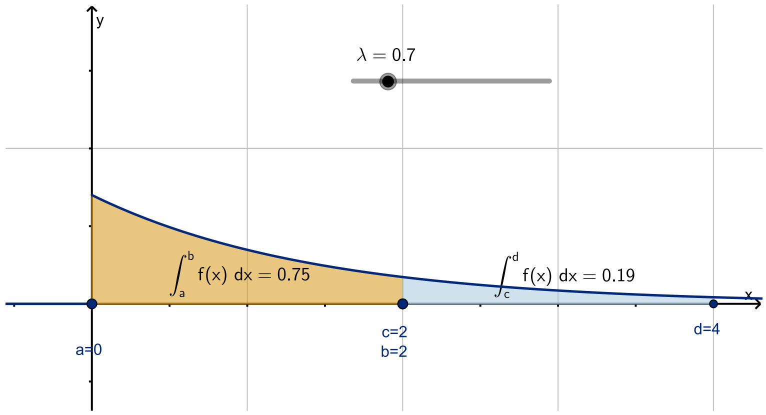


Figure: The density function of an exponential distribution

The second point is best illustrated with a concrete example.

Example

Gravitational waves large enough to detect pass through the earth from time to time. Suppose we switch on a gravitational wave detector, and the time (in days) until the first detection is modeled by the exponential random variable X with density function $0.7e^{-0.7x}$.

- The probability that the first detection occurs within two days is 0.75.
- If the first detection does not occur in the first two days, then the probability that it occurs in the following two days is 0.75
- If the first detection does not occur in the first four days, then the probability that it occurs in the following two days is 0.75
- And so on

From this we can compute

$$\begin{aligned} P(2 \leq X \leq 4) &= \underbrace{(1 - P(X \leq 2))}_{\substack{\text{X is not in} \\ \text{the first two days}}}(0.75) \\ &= (0.25)(0.75) \\ &= 0.1875 \end{aligned}$$

Our final family is the most famous, because it is the most generally applicable.

Definition

The **normal distribution** is sometimes called a **bell curve**. Many natural phenomena are normally distributed. The formula is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Question 2.6.3 What Density Functions Arise Naturally?

The anti-derivative of this density function cannot be expressed with functions that we can evaluate. Instead we can look up values in a table. The normal distribution has a special role in statistics:

Theorem [The Central Limit Theorem]

The average of any n independent identically distributed random variables (for instance performing the same experiment n times) will converge to a normal distribution as n gets large.

This theorem helps explain why many natural measurements are approximated by bell curves. For example, human height is affected by hundreds of factors, including individual genes, nutrition and environment. If we view human height as an average of these factors, scaled with appropriate units, then we expect human heights to be modeled by a normal random variable. Viewing a histogram of human height statistics shows the expected bell curve.

The parameters in f_X can be interpreted as follows:

- μ is the average value of X. It corresponds to the peak of the bell curve.
- σ is the standard deviation of X. Larger σ means that X has a larger probability of being far from μ.

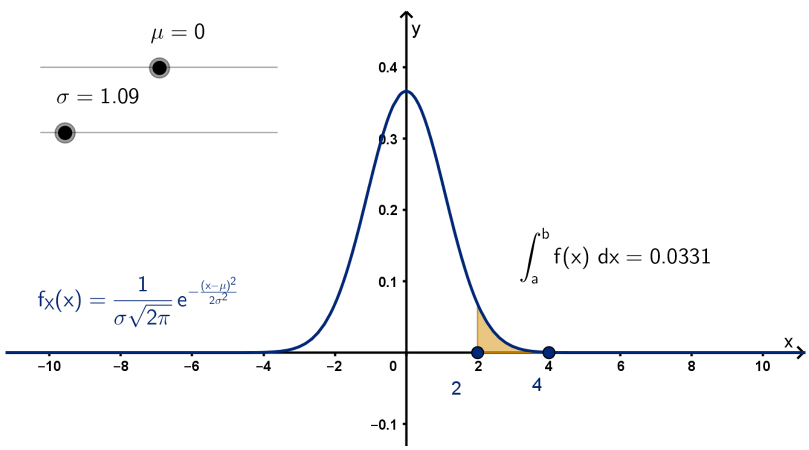


Figure: The density function (bell curve) of a normal distribution

Question 2.6.4 What Is the Expected Value of a Random Variable?

Expected value will be the first **statistic** we can compute for a random variable. Statistics of a data set tell us something about the numbers in the data set. Statistics of a random variable should tell us something about the outcomes of the random variable.

The **expected value** or **average value** of X describes what the average result will be, if you let X take a value at random many times. It is typically denoted $E[X]$ or with the letter μ.

Example

Suppose we average our rolls of a six-sided die. As the number of rolls n gets large, we'll roll each number close to $\frac{n}{6}$ times. The sum of the rolls will be approximately

$$1\left(\frac{n}{6}\right)+2\left(\frac{n}{6}\right)+3\left(\frac{n}{6}\right)+4\left(\frac{n}{6}\right)+5\left(\frac{n}{6}\right)+6\left(\frac{n}{6}\right)$$

to compute the average, we divide by n. Fortunately, every term already has an n.

$$\mu=1\left(\frac{1}{6}\right)+2\left(\frac{1}{6}\right)+3\left(\frac{1}{6}\right)+4\left(\frac{1}{6}\right)+5\left(\frac{1}{6}\right)+6\left(\frac{1}{6}\right)=3.5$$

In general dividing the number of occurrences of the result a in n evaluations of X will be $nf_X(a)$. When we divide out n, we obtain the following weighted average:

Formula

The expected value of a (discrete) random variable X with probability distribution function f_X is

$$E[X]=\sum_x xf_X(x)$$

where x is summed over all possible outcomes of X.

To produce the corresponding formula for a continuous random variable, instead of multiplying each outcome by its probability and summing, we multiply each output by its density and integrate

Formula

The expected value of a continuous random variable X with probability density function f_X is

$$E[X]=\int_{-\infty}^{\infty} xf_X(x)\ dx$$

Example 2.6.5

The Expected Value of a Uniform Random Variable

Compute the expected value of a uniform random variable on $[a, b]$.

Solution

We'll apply the formula. Since $f_X(x)$ has discontinuities at a and b, we will break it into three parts.

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f_X(x)\ dx \\
&= \int_{-\infty}^{a} x(0)\ dx + \int_{a}^{b} x \frac{1}{b-a}\ dx + \int_{b}^{\infty} x(0)\ dx \\
&= \frac{1}{2(b-a)} x^2 \Big|_a^b \\
&= \frac{1}{2(b-a)} b^2 - \frac{1}{2(b-a)} a^2 \\
&= \frac{b^2 - a^2}{2(b-a)} \\
&= \frac{(b-a)(b+a)}{2(b-a)} \\
&= \frac{b+a}{2}
\end{aligned}
$$

Notice that this is the midpoint of the interval $[a, b]$. Since X is uniformly distributed across the interval, we'd expect the average value to occur at the midpoint.

Main Ideas

$E[X]$ is typically occurs somewhere in the middle of the possible outcomes of X. With symmetric density functions, it is the midpoint.

Example 2.6.6

The Expected Value of an Exponential Random Variable

a. Compute the expected value of a exponential random variable.

b. Explain why the role of λ in the answer to a makes sense.

Solution

a We will use the formula. Even after removing the region of 0 density, we are left with an improper integral. We therefore will compute a limit.

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x f_X(x)\ dx \\
&= \int_{-\infty}^{0} x(0)\ dx + \int_{0}^{\infty} x\lambda e^{-\lambda x}\ dx \\
&= \lim_{t\to\infty} \int_{0}^{t} x\lambda e^{-\lambda x}\ dx \\
&= \lim_{t\to\infty}\ -xe^{-\lambda x}\Big|_0^t - \int_0^t -e^{-\lambda x}\ dx \\
&= \lim_{t\to\infty}\ -xe^{-\lambda x} - \frac{1}{\lambda}e^{-\lambda x}\Big|_0^t \\
&= \lim_{t\to\infty} -te^{-\lambda t} - e^{-\lambda t} + 0e^0 + \frac{1}{\lambda}e^0 \\
&= \lim_{t\to\infty} -te^{-\lambda t} - 0 + 0 + \frac{1}{\lambda} \\
&= \frac{1}{\lambda} + \lim_{t\to\infty} -\frac{t}{e^{\lambda t}} && \left(\frac{\infty}{\infty} \text{ form}\right) \\
&= \frac{1}{\lambda} + \lim_{t\to\infty} -\frac{1}{\lambda e^{\lambda t}} && \text{(l'Hôpital's rule)} \\
&= \frac{1}{\lambda} + 0
\end{aligned}$$

by parts

$u = x$	$dv = \lambda e^{-\lambda x}\ dx$
$du = dx$	$v = -e^{-\lambda x}$

Our final answer is

$$E[X] = \frac{1}{\lambda}$$

b X measures the time until an event with average frequency λ occurs. Thus on average, we expect to wait $\frac{1}{\lambda}$ for it. For example, if an event occurs three times per hour, we would expect to wait about 20 minutes for it to occur.

Example 2.6.6 The Expected Value of an Exponential Random Variable

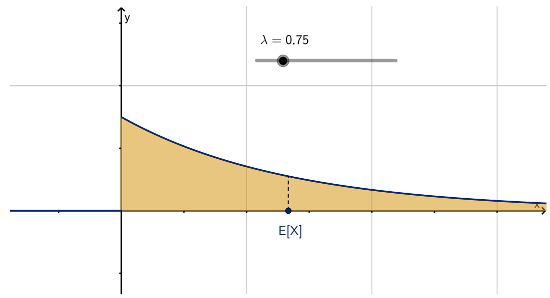


Figure: The expected value of a exponential random variable

Main Idea

For asymmetric density functions, $E[X]$ will not be in the middle of the range of values. It will be pulled toward regions of higher likelihood.

Synthesis 2.6.7
Median Wait Time

Suppose that an exponential random variable models the wait time of a random caller to a call center.

a What is the median wait time?

b Explain graphically why the median wait time less than the expected wait time.

Solution

a The median is the number m such that half the outcomes are larger than m and half are smaller.

We can write this as the following equation and solve for m.

$$\begin{aligned}
P(X \leq m) &= 0.5 \\
\int_{-\infty}^{m} f_X(x)\ dx &= 0.5 \\
\int_{-\infty}^{0} f_X(x)\ dx + \int_{0}^{m} f_X(x)\ dx &= 0.5 && \text{(presumably } m > 0\text{)} \\
\int_{-\infty}^{0} 0\ dx + \int_{0}^{m} \lambda e^{-\lambda x}\ dx &= 0.5 \\
-e^{-\lambda x}\Big|_{0}^{m} &= 0.5 \\
-e^{-\lambda m} + e^{0} &= 0.5 \\
-e^{-\lambda m} &= -0.5 \\
-\lambda m &= \ln 0.5 \\
m &= \frac{1}{\lambda} \ln 2
\end{aligned}$$

b The median is the point such that half the area under $y = f_X(x)$ lies on either side. The expected value is weighted. A few outcomes far to one side can balance many outcomes slightly to the other side. The outcomes of X extends to ∞ on the right but only to 0 on the left. These distant outcomes pull the average to the right, but their distant position has no effect on the median.

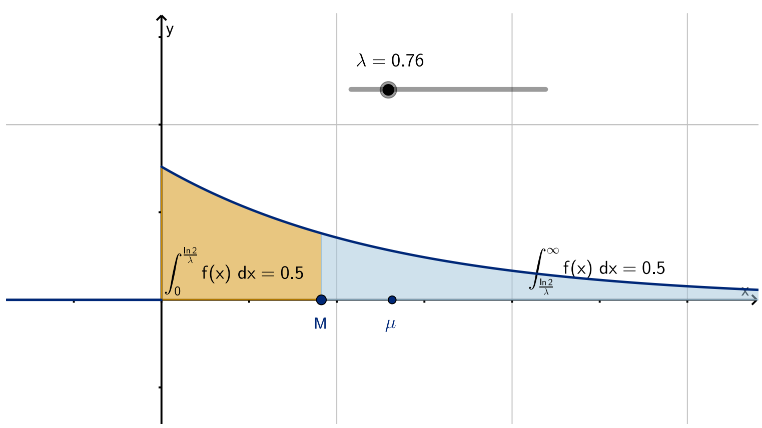


Figure: The median M and expected value μ of an exponential random variable

Synthesis 2.6.7 Median Wait Time

Main Idea

- The median is the value m such that half the area under $y = f_X(x)$ lies on either side of $x = m$.
- We compute the median by setting $P(X \leq m) = 0.5$ and solving for m.
- Median is not the same as expected value. $y = f_X(x)$ may have more area on one side of $E[X]$ than the other, if the smaller side's area is farther from the middle.

Section 2.6 Exercises

Summary Questions

Q1 Describe the difference between a continuous random variable and a non-continuous (discrete) one.

Q2 How do we use a probability density function to compute the probability of an outcome?

Q3 What must be true about a probability density function?

Q4 How do you compute the expected value of a random variable?

2.6.1

Q5 How many possible outcomes does a continuous random variable have?

Q6 One of the following probability questions is different from the others. Explain why.

i. If you spin a prize wheel 3 times, what is the probability that my winnings add up to exactly $\$80$?
ii. If you flip two weighted (unfair) coins, what is the probability that exactly one of them comes up tails?
iii. If you pick a random person, what is the probability that her height is exactly 68 inches?

iv. If I spin a wheel of names, what is the probability that it takes exactly 7 spins to land on my own name?

Q7 Let X be a continuous random variable. Compute $P(X = 13)$.

Q8 Another book might teach you that $P(a < X < b) = \int_a^b f_X(x)\, dx$, instead of $P(a \le X \le b) = \int_a^b f_X(x)\, dx$. Why shouldn't this bother you?

Q9 Let $f_T(t)$ be a probability density function of a random variable T. What quantity is represented by $\int_{-\infty}^{5} f_T(t)\, dt$?

Q10 Let $f_X(x)$ be a probability density function of a random variable X. What quantity is represented by $\int_2^{\infty} f_X(x)\, dx$?

Q11 Given a density function $f_U(u)$ for a random variable U, write an integral or integrals to compute $P(4 \le U^2 \le 9)$.

Q12 Suppose the height of a mature sunflower is given by the random variable H with density function $f_H(h)$. If you friend tells you that her sunflower is in the top quintile in height, explain how you could use f_H to determine a range that the height of her sunflower must lie in.

2.6.2

Q13 Let W be a random variable with density function

$$f_W(w) = \begin{cases} \frac{36-w^2}{144} & \text{if } 0 \le w \le 6 \\ 0 & \text{otherwise} \end{cases}$$

Compute $P(2 \le W \le 9)$

Q14 Let T be a random variable with density function

$$f_T(t) = \begin{cases} \frac{3\sqrt{t}}{2} & \text{if } 0 \le t \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute $(0 \le T \le \frac{1}{4})$

2.6.3

Q15 If U is a uniform random variable on $[4, 7.5]$, compute is the probability that $U \leq 5.5$.

Q16 If X is a uniform random variable on $[2, c]$ and $P(0 \leq X \leq 4) = 0.25$, what is c?

Q17 If W is an exponential random variable such that $P(W \geq 1) = \frac{2}{7}$, then compute the value of the parameter λ in its density function f_W.

Q18 Juan looks at the density function of an exponential random variable X and says "X is more likely to have the value 1 than 5." "That's silly," replies Neha, "X has exactly zero probability of being either of those. They are equally likely." What do you think of their argument?

2.6.4

Q19 Let $f(x) = \begin{cases} bx^{-3} & x \geq 2 \\ 0 & x < 2 \end{cases}$.

- a Compute a number b so that f is a probability density function.
- b If f is the density function for some random variable Z, compute $E[Z]$.

Q20 Suppose X is a random variable with density function $f_X(x)$. Suppose $f_X(x)$ is 0 outside $[3, 11]$ and decreasing on $[3, 11]$. Is $E[X]$ greater or less than 7? Explain.

Q21 Suppose X is a continuous random variable with probability density function

$$f_X(x) = \begin{cases} \frac{3\sqrt{x}}{16} & \text{if } 0 \leq x \leq 4 \\ 0 & \text{if } x > 4 \text{ or } x < 0 \end{cases}$$

- a In a sentence or two, state what you would need to check to ensure that $f_X(x)$ is a valid probability density function. You do not need to actually perform the calculations.
- b Compute $E[X]$.

Q22 Explain how you can use the graph of a normal random variable to identify the expected value. Then compute that value using the expected value formula.

2.6.5

Q23 Give the expected value of a uniform random variable on $[5.2, 9.4]$.

Q24 If the uniform random variable on $[a, b]$ has expected value 7, and $a = 3$, what is b?

Q25 In this example, we divided by $(b - a)$. What would happen if $b - a = 0$?

Q26 If you know the expected value μ of a uniform random variable X, what is the probability that $\geq \mu$? Is this problem answerable without the assumption that X is uniform? Explain.

2.6.6

Q27 Suppose X and Y are two different exponential random variables modeling events that occur on average p and $2p$ times per day respectively. How are their expected values related?

Q28 Does our expected value formula result sense if $\lambda < 0$? Why should this not bother us.

Q29 On bus route 70, 3 buses come per hour, on average.

- **a** Write a probability density function for X, the amount of time until the next bus arrives.
- **b** What is the expected amount of time until the next bus comes?
- **c** How likely is it that you will wait more than an hour for the bus?

Q30 If X is an exponential random variable, what is the probability that $X \leq E[X]$.

2.6.7

Q31 Compute the median value of a uniform random variable on $[a, b]$.

Q32 Let W be a random variable with density function

$$f_W(w) = \begin{cases} \frac{36-w^2}{144} & \text{if } 0 \leq w \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

Compute the median value of W.

Q33 Let T be a random variable with density function

$$f_T(t) = \begin{cases} \frac{3\sqrt{t}}{2} & \text{if } 0 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the median value of T.

Q34 Examine the graph of the density function of a normal random variable X. What is the median of X? Explain how you can see this in the graph.

Extension and Synthesis

Q35 Suppose X is a uniform random variable on $[a, b]$ and $P(3 \leq X \leq 4) = \frac{1}{2}$. Describe all possible values of a and b.

Q36 Suppose the random variable W has the density function

$$f_W(w) = \begin{cases} k(7-w) & \text{if } 1 \leq w \leq 7 \\ 0 & \text{if } w > 7 \text{ or } w < 1 \end{cases}$$

a. What values of W are possible?

b. What can you say about which values of W are more likely than others?

c. Given that f_W is a density function, what is the value of the constant k?

d What is the average value of W?

e Can you compute the median value of W? This might be easier with geometry than with calculus.

Q37 Suppose that $g(x)$ is a probability distribution for a random variable X and $g(x) = 0$ for all $x \geq 0$.

a What is the value of $\int_{-\infty}^{0} g(x)\, dx$? Justify your answer with a sentence or computation.

b Give a formula for $E[X]$. Is it positive or negative? Justify your answer in a sentence or two.

Q38 Recall that an even function $f(x)$ has the property that $f(x) = f(-x)$ for all x. If the density function of a random variable is even, what does that say about the expected value and median of X? Explain your answer.

Section 2.7

Functions of Random Variables

Goals:

1. Compute expected values of functions of a random variable.
2. Compute the average value of a function.
3. Compute the variance of a random variable.

Sometimes the quantity modeled by a random variable is not the quantity we actually care about. For example, while we might have a model for how many people will contract a disease, what we actually would like to predict is how many healthcare resources they will require. The number of patients determines the required resources, so mathematically, resources is a function of patients. Expected values of such functions turn out to be straightforward to compute. A natural way to generate statistics about a random variable is to write a function that measures something interesting and compute its expected value.

Question 2.7.1

What Is a Function of a Random Variable?

When we write a function $g(X)$ of a random variable X, then the output Y of this function is itself a random variable. These functions are most intuitive with a discrete random variable. In this case we can compute Y's probability distribution function by applying g to each outcome of X and summing the probabilities that produce each output.

Example

Let X be a discrete random variable with probability distribution function $f_X(x)$. If $Y = g(X) = X^2$ then Y is a random variable and we can compute its probability distribution function $f_Y(y)$.

$$f_X(x) = \begin{cases} 0.1 & \text{if } x = 0 \\ 0.2 & \text{if } x = 2 \\ 0.3 & \text{if } x = 3 \\ 0.4 & \text{if } x = -2 \\ 0 & \text{otherwise} \end{cases} \qquad f_Y(y) = \begin{cases} 0.1 & \text{if } y = 0 \\ 0.6 & \text{if } y = 4 \\ 0.3 & \text{if } y = 9 \\ 0 & \text{otherwise} \end{cases}$$

Since $X = 2$ and $X = -2$ both produce $Y = 4$, we added their probabilities together.

The function g does not need to be algebraically defined.

Example

Let X be a discrete random variable whose outputs are integers from 1 to 100, uniformly distributed (meaning each occurs with probability $\frac{1}{100}$). Let N give the number of digits of X. Then N has distribution function.

$$f_N(n) = \begin{cases} \frac{9}{100} & \text{if } n = 1 \\ \frac{90}{100} & \text{if } n = 2 \\ \frac{1}{100} & \text{if } n = 3 \\ 0 & \text{otherwise} \end{cases}$$

Question 2.7.2

How Do We Compute Expected Value of a Function?

In the case of a discreet random variable, we can compute expected value directly from the distribution function.

Example

Let X be a discrete random variable whose outputs are integers from 1 to 100, uniformly distributed. Let N give the number of digits of X.

$$E[N] = (1)\left(\frac{9}{100}\right) + (2)\left(\frac{90}{100}\right) + (3)\left(\frac{1}{100}\right) = 1.92$$

Alternately, we could avoid using f_N by directly applying the digits function to each outcome X and taking a weighted average.

Example

$$\begin{aligned} E[N] = {} & \underbrace{(1)\left(\frac{1}{100}\right) + \cdots + (1)\left(\frac{1}{100}\right)}_{9 \text{ times}} \\ & + \underbrace{(2)\left(\frac{1}{100}\right) + \cdots + (2)\left(\frac{1}{100}\right)}_{90 \text{ times}} \\ & + (3)\left(\frac{1}{100}\right) \\ = {} & 1.92 \end{aligned}$$

Question 2.7.2 How Do We Compute Expected Value of a Function?

In general this gives us two ways to compute the expected value of a function.

Formulas

If $Y = g[X]$ then we can compute $E[Y]$ from f_X or from f_Y.

$$E[Y] = \sum_{\text{outcomes } y_i} y_i f_Y(y_i)$$

$$E[Y] = \sum_{\text{outcomes } x_i} g(x_i) f_X(x_i)$$

Remarks

- We can equate these formulas by substituting

$$f_Y(y_i) = \sum_{g(x_j)=y_i} f_X(x_j).$$

  All that remains is to distribute the y_i.

- Both formulas will get us to the answer, but one of them skips the step of finding a distribution function for Y.

In the case of a continuous random variable X, we might find it difficult to find the expected value of $Y = g(X)$ directly. We would need to

- Find a density function $f_Y(y)$ such that

$$\int_a^b f_Y(y)\ dy = P(a \leq g(X) \leq b)$$

  for all a and b

- Integrate $E[Y] = \displaystyle\int_{-\infty}^{\infty} y f_Y(y)\ dy$.

The first step is difficult for any but the simplest functions.

Fortunately, there is an integration analogue of substitution and distributive argument for discrete variables. This allows us to compute the average outcome of Y as a weighted average of the probabilities of X.

Theorem

If $Y = g(X)$ is a function of a continuous random variable X with density function $f_X(x)$, then

$$E[Y] = \int_{-\infty}^{\infty} g(x) f_X(x)\ dx$$

Notice that the expected value of X is a special case of this theorem. In this case, we are computing the expected value of the function $g(X) = X$.

Example 2.7.3
Computing the Expected Value of a Function

Consider the random variable X with density function

$$f_X(x) = \begin{cases} \frac{1}{9}x^2 & \text{if } 0 \leq x \leq 3 \\ 0 & \text{if } x > 3 \text{ or } x < 0 \end{cases}$$

What is the expected value of e^X?

Solution

Since we want $E[e^X]$, our function is $g(x) = e^x$.

$$\begin{aligned} E[e^X] &= \int_{-\infty}^{\infty} e^x f_X(x)\ dx \\ &= \int_0^3 \frac{1}{9}x^2 e^x\ dx \\ &= \frac{1}{9}x^2 e^x \Big|_0^3 - \int_0^3 \frac{2}{9}xe^x\ dx \\ &= \frac{1}{9}x^2 e^x \Big|_0^3 - \frac{2}{9}xe^x \Big|_0^3 + \int_0^3 \frac{2}{9}e^x\ dx \\ &= \frac{1}{9}x^2e^x - \frac{2}{9}xe^x + \frac{2}{9}e^x \Big|_0^3 \\ &= \frac{5e^3 - 2}{9} \end{aligned}$$

by parts

$$u = \tfrac{1}{9}x^2 \qquad dv = e^x\ dx$$
$$du = \tfrac{2}{9}x\ dx \qquad v = e^x$$

by parts again

$$u = \tfrac{2}{9}x \qquad dv = e^x\ dx$$
$$du = \tfrac{2}{9}\ dx \qquad v = e^x$$

We can check whether our answer is reasonable. Since X has outcomes between 0 and 3, e^X should have outcomes between 1 and e^3. Our expected value should also fall in that range, and it does.

Application 2.7.4

The Average Value of a Function

Sometimes people refer to the average value of a function without any reference to a random variable. In this case, we understand the input variable to be uniformly distributed.

Definition

The **average value of a function** from $x = a$ to $x = b$ is the expected value of $f(X)$, where X is a uniform random variable on $[a, b]$. The density function is a constant, so we can factor it out of the integral. We obtain the formula:

$$f_{\text{ave}} = \frac{1}{b-a} \int_a^b f(x)\ dx.$$

The number f_{ave} has geometric significance as well. The signed area under the graph $y = f(x)$ from $x = a$ to $x = b$ is

$$\text{Area} = \int_a^b f(x)\ dx.$$

The region under the horizontal line $y = f_{\text{ave}}$ is a rectangle with equal signed area:

$$\text{Area} = \text{width} \times \text{height} = (b-a)\left(\frac{1}{b-a}\int_a^b f(x)\ dx\right).$$

In other words, if we flattened the area under f into a rectangle, f_{ave} would be its height.

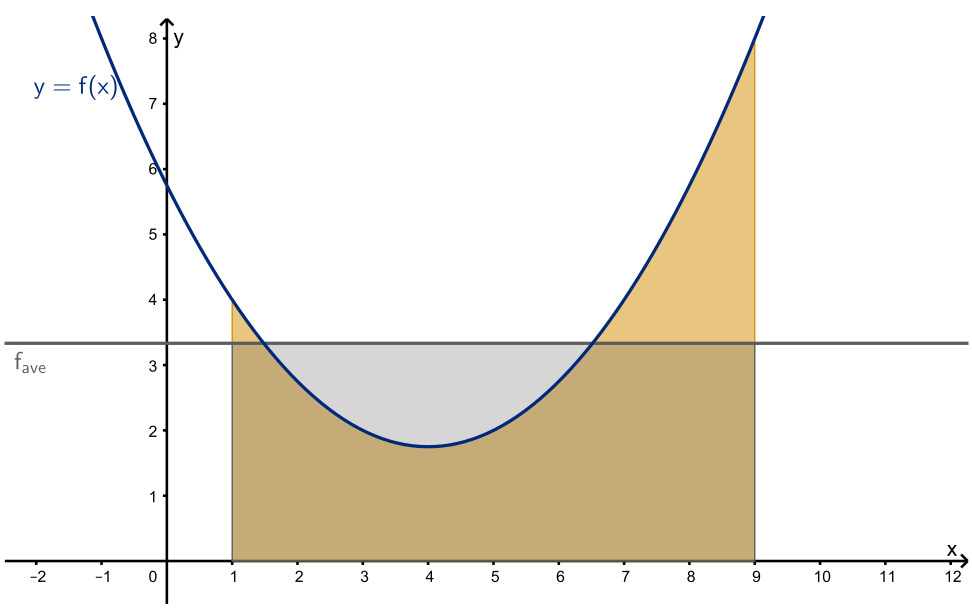


Figure: The graph of $y = f(x)$ and the constant function $y = f_{\text{ave}}$

Example 2.7.5
Computing The Average Value of a Function

Compute the average value of $f(x) = xe^{x^2}$ between $x = 1$ and $x = 3$.

Solution

$$
\begin{aligned}
f_{\text{ave}} &= \frac{1}{3-1}\int_1^3 xe^{x^2}\,dx \\
&= \frac{1}{2}\int_1^9 \frac{1}{2}e^u\,du \\
&= \frac{1}{4}e^u\Big|_1^9 \\
&= \frac{1}{4}(e^9 - e)
\end{aligned}
$$

u-substitution

$$
\begin{array}{rl}
u = x^2 & x = 1 \Rightarrow u = 1 \\
du = 2x\,dx & x = 3 \Rightarrow u = 9 \\
\frac{1}{4}\,du = \frac{y}{2}\,dy &
\end{array}
$$

Application 2.7.6
Variance

Suppose we wanted to plan ahead for the outcome of some random variable X. We might choose to prepare for the circumstance in which X takes on the value $E[X]$. This is most likely to be a good bet, but how much effort should we expend preparing for outcomes far from $E[X]$? It would help to know how likely X is to be far from $E[X]$. We can model this with a distance function (actually we'll use distance squared) and compute the expected value of the distance function.

Definition

The **variance** of a random variable X is the expected value of $(X - E[X])^2$. If X is continuous with density function $f_X(x)$, we obtain the formula

$$\int_{-\infty}^{\infty} (x - E[X])^2 f_X(x)\,dx$$

The square root of variance is the **standard deviation**. Standard deviation is often denoted by σ, and variance is often denoted by σ^2.

If the expected value of $(x - E[X])^2$ is larger, then X is more likely to be far from its expected value.

Application 2.7.6 Variance

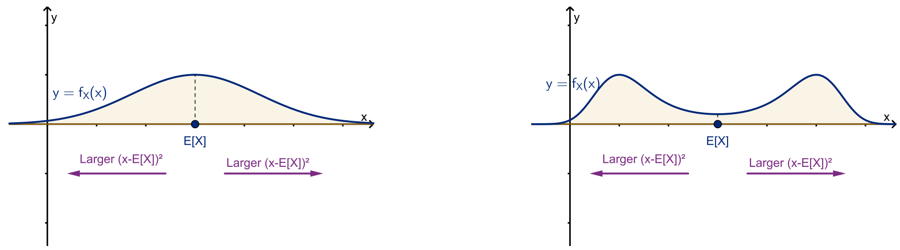


Figure: A density function with less variance and a density function with more variance

For example, we can compute the variance of X where X is a uniform random variable on $[0, 8]$.

Solution

Variance is the expected value of $(X - E[X])^2$, so first we need to know the number $E[X]$. We showed earlier that for a uniform random variable, $E[X]$ is the midpoint of the interval. In this case that is $\frac{8+0}{2} = 4$. Armed with this value, we can compute the variance.

$$\begin{aligned}
E\left[(X-4)^2\right] &= \int_{-\infty}^{\infty} (x-4)^2 f_X(x)\ dx \\
&= \int_0^8 (x-4)^2 \frac{1}{8-0}\ dx && \text{because } f_X(x) = 0 \text{ outside } [0,8] \\
&= \frac{1}{8}\int_0^8 x^2 - 8x + 16\ dx && \text{factor out } \frac{1}{8} \\
&= \frac{1}{8}\left(\frac{x^3}{3} - 4x^2 + 16x\right)\Big|_0^8 \\
&= \frac{1}{8}\left(\frac{512}{3} - (4)(64) + (16)(8) - 0 + 0 - 0\right) \\
&= \left(\frac{1}{8}\right)\left(\frac{128}{3}\right) \\
&= \frac{16}{3}
\end{aligned}$$

Remarks

- In order to solve for variance, we need to know the expected value. We may have to compute $E[X] = \int_{-\infty}^{\infty} x f_X(x)\ dx$.
- Variance is larger when the area under $y = f_X(x)$ is spread farther to both sides, away from $E[X]$.

Summary Questions

Q1 What kind of object is a function of a random variable?

Q2 How do we compute the expected value of a random variable?

Q3 If someone mentions the "average value" of a function without mentioning what random variable to use, what do you assume?

Q4 What function's expected value is the variance?

2.7.1

Q5 Let X be a random variable that indicates how long from now an event will occur (in hours). How could a random variable indicating how long until the event happens in minutes be defined in terms of X?

Q6 Suppose the radius of a circle R is a random variable. How could we define a random variable to express the area of the circle?

Q7 Dominic buys 200 shares of a stock for $\$60$ each. At the end of the day, the stock is worth $\$V$ per share, where V is a random variable. How could you express Dominic's profit or loss from his stock purchase with a random variable?

Q8 Suppose X is a random variable with outcomes in the range $[2, 7]$. What is the range of outcomes of the random variable $Y = \frac{3}{X^2}$?

2.7.2

Q9 Suppose X is a random variable and $Y = cX$ for some number c. Explain using one or more rules of integration why $E[Y] = cE[X]$.

Q10 Suppose X is a random variable and $Y = X + d$ for some number d. Explain using one or more rules of integration why $E[Y] = E[X] + d$.

Q11 Let X be a uniform random variable on $[2, 5]$ with density function f_X. Write a density function f_Y for $Y = 10X$. Explain how your density function differs from f_X.

Q12 Let X be a uniform random variable on $[0, 3]$. Is $Y = X^2$ a uniform random variable on $[0, 9]$? Provide evidence for your answer.

2.7.3

Q13 Let W be a random variable with density function

$$f_W(w) = \begin{cases} \frac{36-w^2}{144} & \text{if } 0 \leq w \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

Compute $E\left[\frac{1}{W}\right]$

Q14 Let T be a random variable with density function

$$f_T(t) = \begin{cases} \frac{2\sqrt{t}}{3} & \text{if } 0 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute $E[T^3]$.

Q15 Let X be an exponential random variable. Compute $E[X^2]$.

Q16 Let $g(x) = c$ be a constant function. Let X be a random variable. Compute $E[g(X)]$.

2.7.4

Q17 Suppose that you are told that the average value of $f(x)$ from $x = a$ to $x = b$ is 0.

- **a** What geometric information does this give you about the graph $y = f(x)$. Be specific.
- **b** Suppose you are told that $f(x)$ is non-negative for all x. How does that affect your answer to **a**?

Q18 Suppose you know that $f(x) = \sqrt[3]{x}$ has a positive average value over $[a, b]$. What does this tell you about a and b?

2.7.5

Q19 Compute the average value of $f(x) = x^2$ over $[0, 3]$.

Q20 Compute the average value of $g(x) = x \sin x$ over $[0, \pi]$.

Q21 Compute the average value of $f(x) = x^2 e^{3x}$ over $[0, 2]$

Q22 What happens if we try to compute the average value of $h(x) = \frac{1}{x^2}$ over $[-2, 2]$?

2.7.6

Q23 Compute the variance of an exponential random variable X. Note that you may already know some components of this computation from earlier examples and exercises.

Q24 Compute the variance of a uniform random variable on $[2, 7]$.

Q25 Let W be a random variable with density function

$$f_W(w) = \begin{cases} \frac{36-w^2}{144} & \text{if } 0 \leq w \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

Compute the variance of W. I'd suggest using a computer to help with the algebra.

Q26 Let T be a random variable with density function

$$f_T(t) = \begin{cases} \frac{2\sqrt{t}}{3} & \text{if } 0 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the variance of T.

Synthesis and Extension

Q27 Let X be a random variable with density function f_X. Let $Y = cX$ for some number c. Write a formula for f_Y

Q28 Compute the value b such that the average value of $f(x) = x^2$ over $[0, b]$ is 1.

Q29 Some people memorize compute variance using the formula $\sigma^2 = E[X^2] - E[X]^2$. Explain why this formula is equivalent to the one we gave. (This is a famous calculation, so if you can't figure it out, look it up and try to explain each step).

Chapter 3

Series

This chapter introduces the Taylor polynomial, which is a useful tool for approximating functions that cannot be evaluated with arithmetic. Like with the derivative and integral before it, we would like to send the error in these approximations to 0. This requires us to take a new kind of limit called a series. We will develop the tools to work with series, with the ultimate goal of defining and utilizing Taylor series.

Contents

Section 3.1

Taylor Polynomials

Goals:

1. Approximate a function with a Taylor polynomial.
2. Compute error bounds for a Taylor polynomial.

When learning algebra and trigonometry, we learn to use exact values like $\sqrt{7}$ instead of decimal approximations, like 2.646. This prevents us from introducing errors into our calculations. However, there are also advantages to approximation. Decimal approximations give us a much better sense of the size of a number than $\ln 873$ or $e^{\frac{9}{5}}$. (Which of these is larger?)

Unfortunately arithmetic does not give us methods for approximating many quantities. Ideally, we would like a method of approximation whose accuracy is limited only by how much time we wish to spend computing. An example of this is long division. We can compute as many decimal places of $\frac{32}{13}$ as we want, getting closer and closer to the exact value. Of course, long division can only approximate fractions.

The method we will develop in this section is called a Taylor polynomial. It gives us a way to approximate otherwise incomputable functions. The beginning point is the tangent line. The tangent line was the motivation for developing the derivative, but its greatest benefit is not geometric. The tangent line approximates the values of a function near the point of tangency. While the function may be difficult to evaluate, the equation of the tangent line is linear. We can evaluate it by hand.

Question 3.1.1

How Can We Improve on a Linearization?

Formula

The **linearization** or **tangent line** to a function $f(x)$ at a has the equation.

$$L(x) = f(a) + f'(a)(x-a)$$

By design f and L have

1. Equal values at a.
2. Equal first derivatives at a.

This means that for values of x near a, $L(x)$ and $f(x)$ will have similar values. $L(x)$, which is easy to compute, can be used as an approximation of $f(x)$. As x travels away from a and $y = f(x)$ curves away from its tangent line, this method will lose accuracy. We could make a better approximation, if we could match second, third, fourth derivatives of $f(x)$. A line cannot do that, but a polynomial can.

Question 3.1.2
What Is a Taylor Polynomial?

A polynomial that mimics the first n derivatives of a function is called a Taylor polynomial. Here is the formal definition.

Definition

The n^{th} **Taylor polynomial** of $f(x)$ at $x = a$ is a degree n polynomial that shares the value and first n derivatives of f at $x = a$. Its formula is

$$T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

Remarks

- The variable is x. $f^{(k)}(a)$ is not a function but a number.
- $f^{(0)}$ is the zeroth derivative, meaning $f^{(0)}(a) = f(a)$.
- $0!$ is defined to be 1.

Example 3.1.3
Computing a Taylor Polynomial

a. Find the degree 3 Taylor polynomial of $y = \sqrt{x}$ at $x = 4$.

b. Use it to estimate $\sqrt{5}$.

Example 3.1.3 Computing a Taylor Polynomial

Solution

a We will apply the equation of the Taylor polynomial where $a = 4$ and $n = 3$. Examining the formula shows we need to know the value of first three derivatives of $f(x)$ at $a = 4$.

$$\begin{aligned} f(x) &= x^{1/2} & f(4) &= 2 \\ f'(x) &= \frac{1}{2}x^{-1/2} & f'(4) &= \frac{1}{4} \\ f''(x) &= -\frac{1}{4}x^{-3/2} & f''(4) &= -\frac{1}{32} \\ f'''(x) &= \frac{3}{8}x^{-5/2} & f'''(4) &= \frac{3}{256} \end{aligned}$$

We can plug these into the summation formula:

$$\begin{aligned} T_3(x) &= \sum_{k=0}^{3} \frac{f^{(k)}(4)}{k!}(x-4)^k \\ &= \frac{f(4)}{0!}(1) + \frac{f'(4)}{1!}(x-4) + \frac{f''(4)}{2!}(x-4)^2 + \frac{f'''(4)}{3!}(x-4)^3 \\ &= \frac{1}{2} + \frac{\frac{1}{4}}{1}(x-4) - \frac{\frac{1}{32}}{2}(x-4)^2 + \frac{\frac{3}{256}}{6}(x-4)^3 \\ &= \frac{1}{2} + \frac{1}{4}(x-4) - \frac{1}{64}(x-4)^2 + \frac{1}{512}(x-4)^3 \end{aligned}$$

b To approximate $\sqrt{5}$, notice $\sqrt{5} = f(5)$ and $f(5) \approx T_3(5)$.

$$\begin{aligned} T_3(5) &= \frac{1}{2} + \frac{1}{4}(5-4) - \frac{1}{64}(5-4)^2 + \frac{1}{512}(5-4)^3 \\ &= \frac{1}{2} + \frac{1}{4}(1) - \frac{1}{64}(1) + \frac{1}{512}(1) \\ &= \frac{256}{512} + \frac{128}{512} - \frac{8}{512} + \frac{1}{512} \\ &= \frac{377}{512} \end{aligned}$$

Example 3.1.4

Writing a Sum in Σ Notation

As our Taylor polynomials get longer, we would like to condense them into $\sum$ notation. Part of the challenge is choosing an expression that will produce all the terms of our sum. Write each of the following sums in Σ notation.

a $4+7+10+13+16+19+22$

b $2+6+18+54+162+486$

c $-3+4-5+6-7+8-9+10$

d $\frac{1}{4}+\frac{\sqrt{2}}{9}+\frac{\sqrt{3}}{16}+\frac{2}{25}+\frac{\sqrt{5}}{36}$

Solution

a The terms increase by 3 each time. Repeated addition is multiplication, in this case $3k$ plus some starting value. Starting with index $k=0$ is convenient, because $3(0)=0$ at the starting value.

$$4+7+10+13+16+19+22=\sum_{k=0}^{6}4+3k$$

b The terms are multiplied by 3 each time. Repeated multiplication is exponentiation, in this case 3^k times some starting value. Starting with index $k=0$ is convenient, because $3^0=1$ at the starting value.

$$2+6+18+54+162+486=\sum_{k=0}^{5}(2)(3^k)$$

c The absolute values of this sum could just be the values of the index variable. To create an alternating $+$ and $-$ pattern, we can multiply by $(-1)^k$.

$$-3+4-5+6-7+8-9+10=\sum_{k=3}^{10}(-1)^k k$$

d In a fraction, we can model the numerator and denominator separately.

$$\frac{1}{4}+\frac{\sqrt{2}}{9}+\frac{\sqrt{3}}{16}+\frac{2}{25}+\frac{\sqrt{5}}{36}=\sum_{k=1}^{5}\frac{\sqrt{k}}{(k+1)^2}$$

Example 3.1.5

A Taylor Polynomial in $\sum$ Notation

Write the 10th degree Taylor Polynomial for $f(x) = \frac{1}{x}$ centered at $x = 3$.

Solution

Computing 10 derivatives seems excessive, so we will compute 4 and try to find a pattern. We'll write $f(x) = x^{-2}$ and apply the power rule.

$$\begin{aligned} f(x) &= x^{-2} \\ f'(x) &= -2x^{-3} \\ f''(x) &= 6x^{-4} \\ f'''(x) &= -24x^{-5} \\ f^{(4)}(x) &= 120x^{-6} \end{aligned}$$

We observe

- The sign of these derivatives is alternating, which we can model with a $(-1)^k$.
- The coefficients look like a factorial pattern, but offset. For example when $k = 2$ we obtain $3!$. We model this with $(k+1)!$.
- The exponent of x decreases by the same amount each step. We model it with $-2-k$.

This suggests a general formula for the kth derivative.

$$f^{(k)}(x) = (-1)^k (k+1)! x^{-2-k}$$

We plug $x = 3$ into $f^{(k)}(x)$ and assemble the Taylor Polynomial:

$$T_{10}(x) = \sum_{k=0}^{10} \frac{(-1)^k (k+1)! 3^{-2-k}}{k!} (x-3)^k$$

Question 3.1.6

How Accurate Is the Taylor Polynomial?

An approximation is much more useful, if we can put a bound on its error. We will present an error bound theorem called "Taylor's Inequality." Taylor polynomials are effective approximations because they try to match the values and rates of change of the original function. In order to make a careful argument, we begin with the basic principal that we can compare functions using the values of their derivatives.

Theorem

Let f and g be differentiable functions. Consider an interval $[a, b]$, and suppose $f(a) = g(a)$.

1. If $f'(x) = g'(x)$ on $[a, b]$, then $f(x) = g(x)$ on $[a, b]$
2. If $f'(x) < g'(x)$ on $[a, b]$, then $f(x) < g(x)$ on $(a, b]$

Reasoning

Intuitive If two functions start at the same value at a, then the one that grows faster will have a higher value at b.

Formal The Fundamental Theorem of Calculus says

$$f(x) - f(a) = \int_a^x f'(t)dt \qquad\qquad g(x) - g(a) = \int_a^x g'(t)dt.$$

Larger functions have larger integrals.

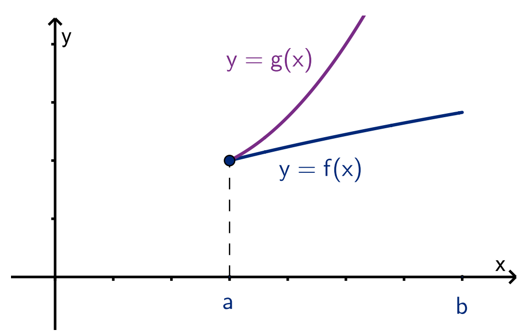


Figure: Two functions with a common value at a: $f(x)$ with a smaller derivative and $g(x)$ with a larger derivative.

Notation

Given a function $f(x)$ and its nth Taylor polynomial $T_n(x)$ centered at a, the **remainder** at x is

$$R_n(x) = f(x) - T_n(x)$$

If we are using $T_n(x)$ to approximate $f(x)$,

$$R_n(x) = -\text{error of } T_n(x).$$

We should be very interested in knowing the value of $R_n(x)$. We will use our derivative comparison theorem to make two arguments

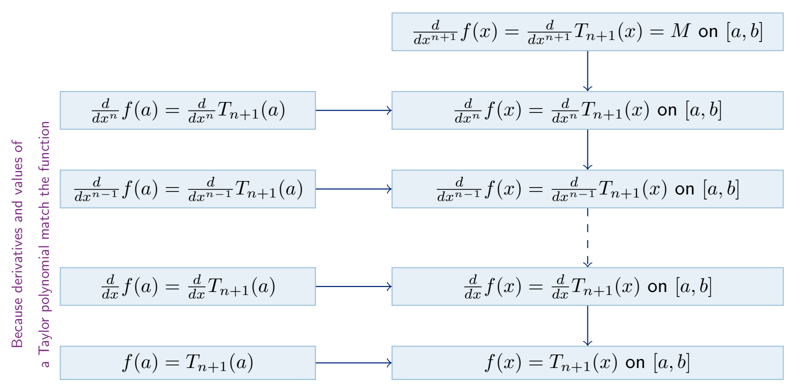
1. If $f^{(n+1)}(x)$ is a constant M, then we can compute $R_n(x)$ exactly.
2. If $|f^{(n+1)}(x)| \leq M$ then the error in 1 is the worst-case scenario.

Theorem

If $f^{(n+1)}(x)$ is a constant M on $[a, b]$, then

$$f(x) = T_{n+1}(x) = T_n(x) + \frac{M}{(n+1)!}(x-a)^{n+1}.$$

Beginning with our assumption about the $(n+1)$th derivatives and the equality of the nth derivatives at a, we can use our derivative comparison theorem to equate the nth derivatives on $[a.b]$. We can use that equality to equate the $(n-1)$th derivatives on $[a, b]$. We continue this reasoning until we conclude that the functions are equal.



Remark

This theorem tells us that when $f^{(n+1)}(x)$ is a constant M, $R_n(x) = f(x) - T_n(x) = \frac{M}{(n+1)!}(x-a)^{n+1}$

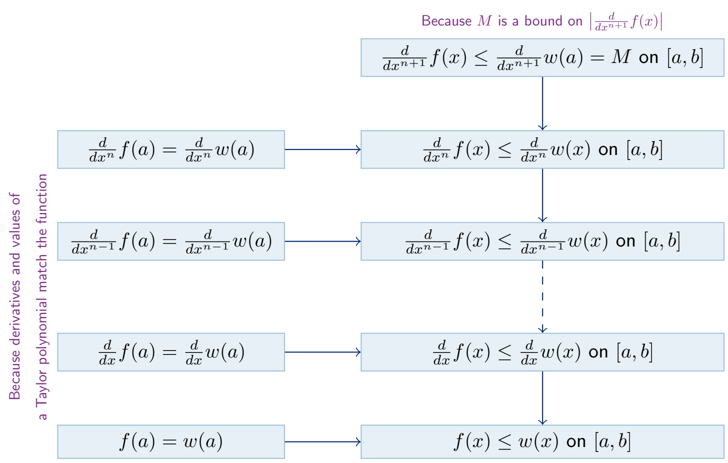
But what if $f^{(n+1)}(x)$ is not a constant? In this case we will settle for a bound on $f^{(n+1)}(x)$.

Theorem [Taylor's Inequality]

If $\left|f^{(n+1)}(t)\right| \leq M$ for all x between a and b, then for all x between a and b,

$$|R_n(x)| \leq \left|\frac{M}{(n+1)!}(x-a)^{n+1}\right|$$

To prove Taylor's Inequality, we compare the derivatives of $f(x)$ with the worst-case scenario $w(x) = T_n(x) + \frac{M}{(n+1)!}(x-a)^{n+1}$. The derivatives $w^{(k)}(a)$ are the same as $T_n^{(k)}(a)$ and $f^{(k)}(a)$ for $0 \leq k \leq n$, and $\frac{d}{dx^{n+1}}w(x) = M$.



To finish the argument we need to

1. Produce a lower bound for f using $w(x) = T_n(x) - \frac{M}{(n+1)!}(x-a)^{n+1}$.

2. Solve the inequality bounds for $R_n(x)$.

$$T_n(x) - \frac{M}{(n+1)!}(x-a)^{n+1} \leq f(x) \leq T_n(x) + \frac{M}{(n+1)!}(x-a)^{n+1}$$
$$-\frac{M}{(n+1)!}(x-a)^{n+1} \leq R_n(x) \leq \frac{M}{(n+1)!}(x-a)^{n+1}$$

3. Repeat for intervals of the form $[b,a]$. These work the same way with a sign reversed.

Example 3.1.7

A Taylor Approximation Error Bound

Let $f(x) = \sin x$.

a. Give a general form for the n^{th} Taylor polynomial for f at $x = 0$.

b. Find a bound on $f^{(n)}(x)$ for each n.

c. What happens to the error bound as x increases but n stays the same?

d. What happens to the error bound as n increases but x stays the same?

e. What does this tell us about the relationship between the $T_n(x)$ approximations and $f(x)$?

Solution

a. For the Taylor polynomial formula, we need to compute the derivatives of $f(x)$.

$$\begin{aligned} f(x) &= \sin x & f(0) &= 0 \\ f'(x) &= \cos x & f'(0) &= 1 \\ f''(x) &= -\sin x & f''(0) &= 0 \\ f'''(x) &= -\cos x & f'''(0) &= -1 \\ f^{(4)}(x) &= \sin x & f^{(4)}(0) &= 0 \\ f^{(5)}(x) &= \cos x & f^{(5)}(0) &= 1 \\ &\vdots & &\vdots \end{aligned}$$

In order to write a general Taylor polynomial, we would need a general expression for $f^{(k)}(0)$. The pattern is obvious, but trying to express it as a formula is much more difficult. The solution is a trick worth remembering:

Since the even derivatives are zero, those terms do not appear in our Taylor polynomials. Since we want to only have odd terms in our summation, we can let our index variable be k, but our exponents in each term be $2k+1$. Thus as k goes from 0 to n, the summation will include only the odd terms x^1 through x^{2n+1}. We can produce the following chart to work out our coefficients:

k	$f^{(2k+1)}(0)$
0	1
1	-1
2	1
3	-1
$\vdots$	$\vdots$

This is an easier pattern to express:

$$f^{(2k+1)}(0) = (-1)^k$$

Now we are ready to write a formula. Since we intend to sum from $k = 0$ to $k = n$, we are actually producing the $(2n + 1)$th Taylor polynomial.

$$\begin{aligned} T_{2n+1}(x) &= \sum_{k=0}^{n} \frac{f^{(2k+1)}(0)}{(2k+1)!} x^{2k+1} \\ &= \sum_{k=0}^{n} \frac{(-1)^k}{(2k+1)!} x^{2k+1} \end{aligned}$$

These are the odd degree Taylor polynomials, but what about the even numbered ones? Since $T_{2n}(x)$ is just $T_{2n-1}(x)$ plus the $2n$th term, and the $2n$th term is zero, we can write

$$T_{2n}(x) = \sum_{k=0}^{n-1} \frac{(-1)^k}{(2k+1)!} x^{2k+1}$$

b Given the chart above, we can see that the derivatives are sines and cosines. These are bounded above by 1 and below by -1. Since Taylor's inequality requires a bound of the form $|f^{(n+1)}(x)| \le M$, we write

$$|f^{(n+1)}(x)| \le 1$$

And luckily, thus works for all x and all n.

c Taylor's Inequality says that $|R_n(x)| \le \left|\frac{1}{(n+1)!} x^{n+1}\right|$. As x goes to ∞, this bound goes to ∞ as well. This makes sense, since $T_n(x)$ is polynomial, while the function it is approximating stays between -1 and 1.

d When n increases x^{n+1} increases by a factor of x. On the other hand, $(n + 1)!$ increases by a factor of $n + 2$. As n increases without bound, $(n + 1)!$ grows faster than x^{n+1} and their ratio approaches 0.

e Any $T_n(x)$ will eventually become inaccurate outside a certain distance from 0. On the other hand, if we want to approximate $\sin(x)$ for a particular x, we can make $T_n(x)$ have as small an error as we want by choosing sufficiently large n.

Example 3.1.7 A Taylor Approximation Error Bound

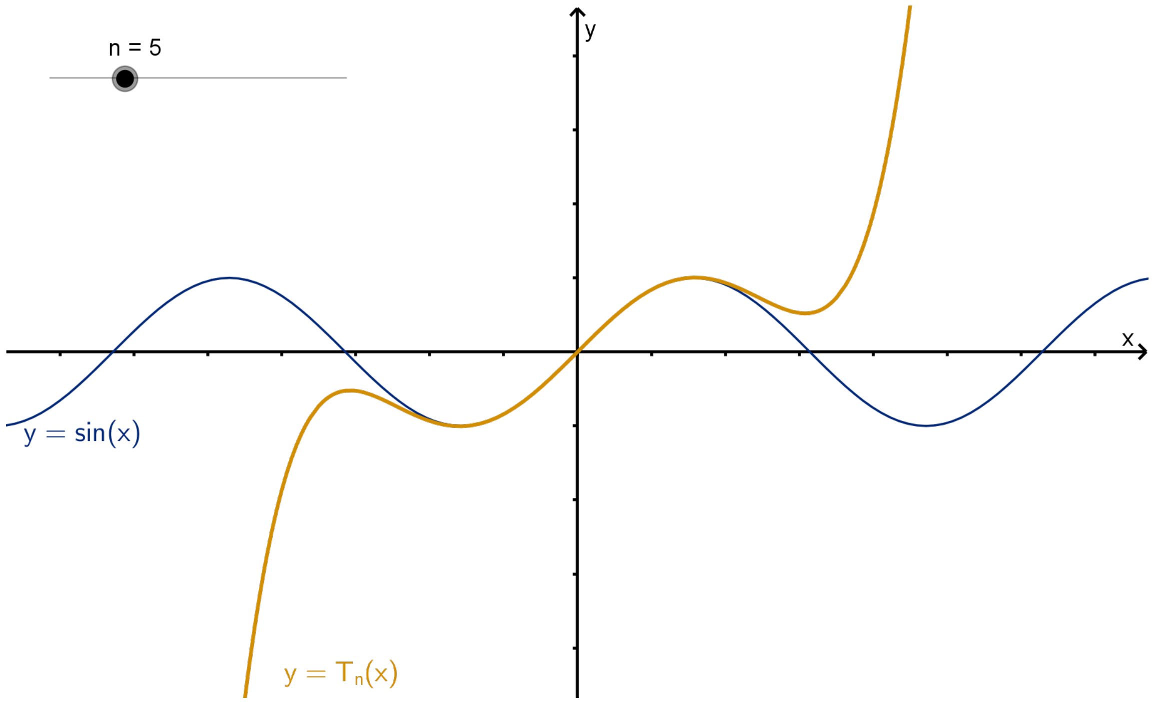


Figure: $f(x) = \sin x$ approximated by its Taylor polynomials, $T_n(x)$

Main Ideas

- In order to understand how the error changes as n increases, we need to have an expression for $f^{(n)}(x)$.
- We can choose M to be the largest value of $|f^{(n+1)}|$ on the interval $[a, x]$. This may not be the value of $|f^{(n+1)}(a)|$.
- In general, Taylor polynomials will become less accurate the farther you get from a.
- We can often mitigate this inaccuracy by choosing larger values of n.

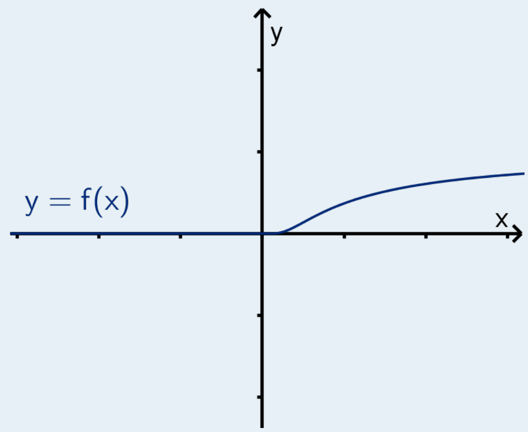
The $(n+1)!$ in Taylor's Inequality might suggest that as n increases, the error in the nth Taylor polynomial must shrink toward 0. However, this is not the case. Some functions are not well estimated by their Taylor polynomial.

Example

$$f(x) = \begin{cases} 0 & \text{if } x \le 0 \\ e^{-\frac{1}{x}} & \text{if } x > 0 \end{cases}$$

$f^{(k)}(0) = 0$ for all k. So the Taylor polynomial at $x = 0$ is

$$T_n(x) = \sum_{k=0}^{n} 0x^k.$$



No matter how large n gets, $T_n(x)$ will not get any closer to $f(x)$ for any $x > 0$.

How can this happen, given Taylor's Inequality? The derivatives of f get bigger and bigger. M grows so fast that the error $R_n(x)$ gets no smaller even with an $(n+1)!$ in the denominator of Taylor's Inequality.

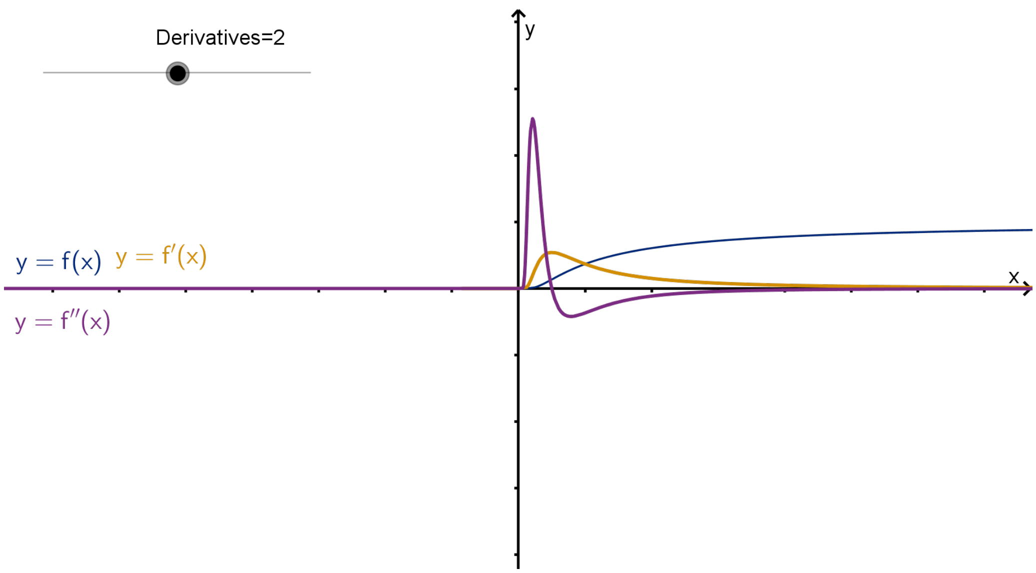


Figure: A function whose derivative bounds grow factorially

Despite examples like this, it turns out that Taylor polynomials often do a good job of approximating functions. For numerical computations, an approximation is good enough. For more theoretical situations, we would like to let n go to ∞ so that the error goes to 0 and we can use the polynomial as an exact replacement of the function. Unfortunately, with infinitely many terms, we no longer have a polynomial at all. Instead we have an object that we will call a Taylor series. We will develop the tools to define and work with Taylor series over the course of this chapter.

Section 3.1

Exercises

Summary Questions

Q1 Why do we use Taylor polynomials?

Q2 Why is there a denominator of $k!$ in the formula for a Taylor polynomial?

Q3 Explain why we'd always rather center a Taylor polynomial for $y = \ln x$ at $x = 1$.

Q4 What properties make a Taylor polynomial $T_n(x)$ a better approximation of $f(x)$?

3.1.1

Q5 Suppose we use the linearization of $f(x) = \sqrt[3]{x}$ at $x = 8$ to approximate $\sqrt[3]{6}$.

- **a** What is the relationship between $f(x) = \sqrt[3]{x}$ and $\sqrt[3]{6}$?
- **b** Suppose $L(x)$ is the the linearization of $f(x)$ at $x = 8$. Would you expect $L(6)$ to overestimate or underestimate $\sqrt[3]{6}$? Explain in a sentence or two.

Q6 Suppose you were locked in a room with only a pencil and paper and asked to compute the first ten decimal places of the following numbers:

$$\frac{4}{17} \qquad \sqrt{7} \qquad e$$

- Which could you compute?
- For the ones you can compute, how would you do it?

3.1.2

Q7 Is a tangent line a Taylor polynomial?

Q8 Suppose $T_4(x)$ is the Taylor polynomial for $f(x)$ centered at $x = 10$. List what information $T_4(x)$ and $f(x)$ have in common, being as specific as possible.

Q9 If $f(x)$ is a decreasing function, what can you say about the coefficients of any Taylor polynomial of $f(x)$?

Q10 Suppose $f(x)$ has a Taylor polynomial

$$T_4(x) = 5 + 3(x-2) - \frac{1}{6}(x-2)^2 + 2(x-2)^4$$

- **a** What is $f(2)$?
- **b** Is f increasing or decreasing at $x = 2$?
- **c** Is f concave up or concave down at $x = 2$?

3.1.3

Q11 Let $f(x) = e^x$.

- **a** Find the degree 8 Taylor polynomial of $y = f(x)$ centered at $x = 0$.
- **b** How could you use this to estimate the value of e?
- **c** Can you use sigma notation to write a general form for the degree n Taylor polynomial of $y = e^x$?

Q12 Let $f(x) = \ln x$

- **a** Write the 5th Taylor polynomial of $f(x)$ at $x = 1$.
- **b** Use your polynomial to approximate $\ln 2$.

Q13 Write the 10th Taylor polynomial for $f(x) = \cos x$ centered at $x = \pi$.

Q14 Write the 4th Taylor polynomial for $f(x) = \frac{1}{x^2}$ centered at $x = 5$.

3.1.4

Q15 Write each of the following sums in Σ notation.

- **a** $15 - 45 + 105 - 315 + 945$
- **b** $24 + 19 + 14 + 9 + 4 - 1 - 6$
- **c** $\frac{1}{8} + \frac{1}{18} + \frac{1}{50} + \frac{1}{72} + \frac{1}{98}$

Q16 Write each of the following sums in Σ notation.

- **a** $11 - 13 + 15 - 17 + 19 - 21 + 23$
- **b** $384 + 192 + 96 + 48 + 24 + 12 + 6$
- **c** $\frac{2}{10} + \frac{3}{100} + \frac{4}{1000} + \frac{5}{10000}$

3.1.5

Q17 Write an expression in Σ notation for the 53rd Taylor polynomial of $f(x) = \ln x$ centered at $x = 1$

Q18 Write an expression in Σ notation for the 15th Taylor polynomial of $f(x) = e^x$ centered at $x = 0$

Q19 Write an expression in Σ notation for the 100th Taylor polynomial of $f(x) = \cos x$ centered at $x = 0$

Q20 Write an expression in Σ notation for the 71st Taylor polynomial of $f(x) = \frac{1}{x^2}$ centered at $x = 10$

3.1.6

Q21 Why don't we have any theorems for a lower bound for error? Give your answer in a few sentences.

Q22 Suppose you are using Taylor polynomials of $f(x)$ centered at $x = 0$ to approximate $f(-3)$. However, for each k, the best bound you can put on $f^{(k)}(x)$ on $[-3, 0]$ is $\frac{k!}{4^k}$. Will you be able to guarantee a good approximation of $f(-3)$ this way? Explain.

Q23 Suppose the fourth derivative of $f(x)$ is $f^{(4)}(x) = e^{x^3}$. Suppose we have written $T_4(x)$, the degree 4 Taylor polynomial of $f(x)$ centered at $x = 1$. What can you say about the difference between $T_4(5)$ and $f(5)$? Be specific and justify your answer with a computation. You do not need to simplify any arithmetic in your calculations.

Q24 Sketch a graph of $y = e^x$ and several tangent lines. On which part of the graph do the tangent lines appear to approximate the function better? Does Taylor's Inequality confirm this observation? Explain.

3.1.7

Q25 Here is the degree 3 Taylor polynomial of $f(x) = \sqrt{x}$ centered at $x = 4$:

$$T_3(x) = 2 + \frac{1}{4}(x-4) - \frac{1}{64}(x-4)^2 + \frac{1}{512}(x-4)^3$$

a Which derivative will let you bound the error of this approximation?

b Can you put a bound on this derivative that holds for all x?

c Can you put a bound on this derivative that holds for x in the interval $[4, 5]$?

d What error bound does this suggest for using $T_3(5)$ to approximate $\sqrt{5}$?

Q26 Let $f(x) = \sqrt[3]{x}$.

a Write the degree 2 Taylor polynomial of f centered at $x = 8$.

b If you wanted to use the Taylor polynomial to approximate $\sqrt[3]{10}$, how would you do that?

c What bound could you place on the error in the approximation in b?

Q27 Let $f(x) = e^x$.

a Write the degree 5 Taylor polynomial of f centered at 0.

b How could we use this polynomial to approximate $\frac{1}{\sqrt{e}}$?

c Produce an error bound for your approximation in b.

Q28 Let $f(x) = xe^x$.

a Compute the Taylor polynomial $T_3(x)$ for $f(x)$ centered at $x = 0$.

b Compute the theoretical error bound for $T_3(2)$.

c Explain the difficulties that would arise from this error bound, if your goal is to approximate $f(2)$ by hand. Can you resolve them?

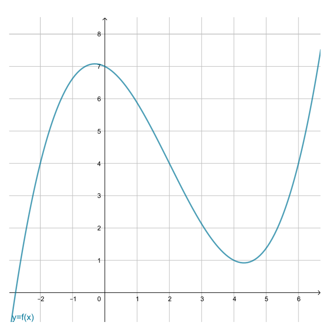
Q29 Let $f(x) = \cos 3x$

a Write the degree 4 Taylor polynomial of f centered at $x = 0$.

b How would you use that Taylor polynomial to approximate the value of $\cos\frac{3\pi}{4}$?

c What bound can you place on the error of such an approximation?

Q30 Consider the graph of $y = f(x)$ below.



a Suppose you wanted to produce the second degree Taylor polynomial of f centered at $a = -1$. Indicate whether the constant term and each coefficient would be positive or negative. Provide evidence for your answer.

b Would $T_2(4)$ underestimate or overestimate $f(4)$? Explain.

Synthesis and Extension

Q31 Let $f(x) = x^3 - 3x + 5$.

a Write an expression for $T_3(x)$, the Taylor polynomial centered at $x = 2$.

b What can you say about th error $R_3(x)$ for any x?

c What relationship does this suggest between $f(x)$ and $T_3(x)$?

d Can you verify this relationship algebraically?

e Conjecture a general relationship between polynomial functions and certain Taylor polynomials. Can you use Taylor's inequality to justify your conjecture?

Section 3.2

Sequences

Goals:

1. Use notation to describe the terms of an infinite sequence.
2. Calculate the limit of an infinite sequence.

Sequences are the first step in our development of Taylor series. While they appear to have little in common with polynomials of infinite degree, they are the scaffolding on which such objects are built.

Question 3.2.1

What Is a Sequence?

A **sequence** is an ordered set of numbers. If this set is infinite, we can most rigorously define it by giving a general formula for the n^{th} term for some index variable n. Here are three different notations for the same sequence.

$$\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5} \cdots\right\} \qquad \left\{\frac{n}{n+1}\right\}_{n=1}^{\infty} \qquad a_n = \frac{n}{n+1}$$

Example

The first three terms of $\left\{\frac{n^2}{2^n}\right\}_{n=0}^{\infty}$ are

$$\frac{0^2}{2^0} = 0 \qquad \frac{1^2}{2^1} = \frac{1}{2} \qquad \frac{2^2}{2^2} = 1$$

Question 3.2.2

What Is the Limit of a Sequence?

Definition

If we can make the elements of a sequence a_n arbitrarily close to some number L by considering only n above a certain number, then we write

$$\lim_{n \to \infty} a_n = L$$

and we say the sequence **converges** to L. If a_n does not converge to any such L then we say it **diverges**.

Remarks

- The first few or even the first thousand terms of a sequence have no bearing on the limit. We only care that we can eventually get close to L.
- "Arbitrarily close" means any level of closeness than anyone could ask for. Eventually the sequence must be within $\frac{1}{100}$ of L, and $\frac{1}{1000}$ and $\frac{1}{1000000}$.

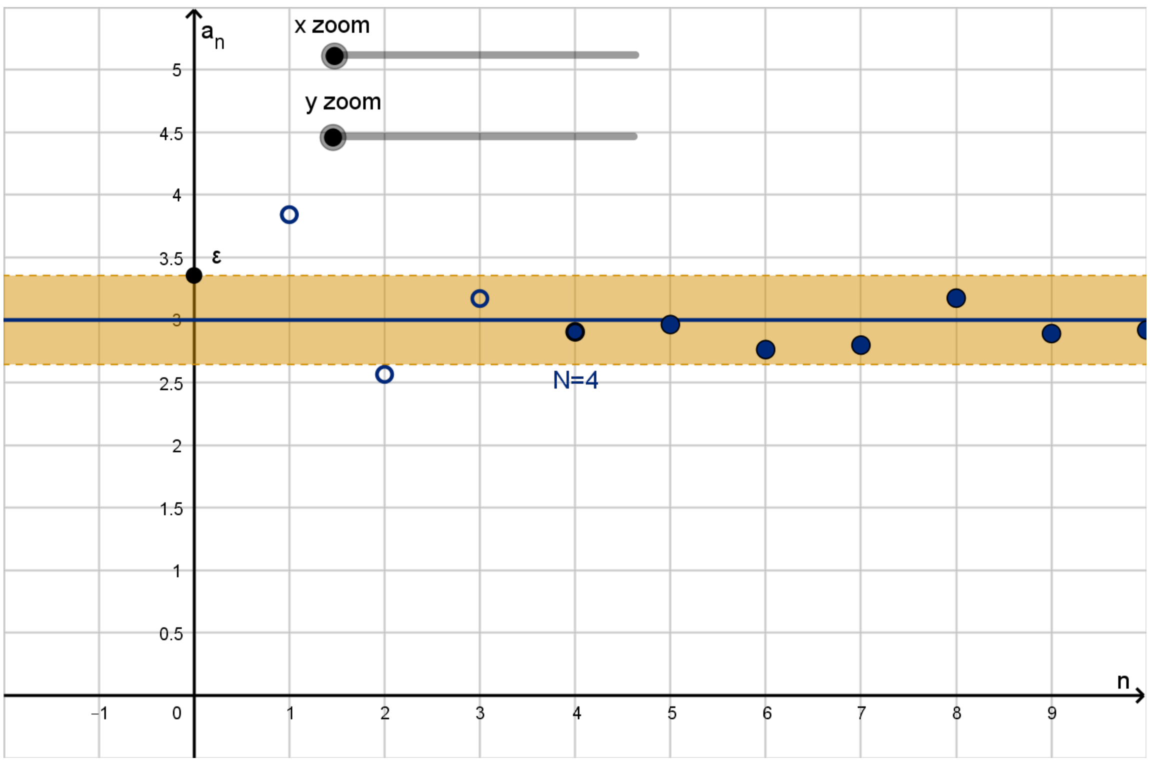


Figure: A sequence converging to $L = 3$

Example 3.2.3

Computing a Limit

Calculate $\lim_{n\to\infty} \dfrac{n}{n+1}$

$\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5} \cdots\}$

Example 3.2.3 Computing a Limit

Solution

Writing the first few terms suggests that this sequence approaches 1. To see that, we can measure the distance to 1:

$$1 - a_n = 1 - \frac{n}{n+1} = \frac{1}{n+1}$$

We can make this smaller than any positive number. For instance to make a_n within $\frac{1}{1000}$ of 1, we can consider only $n > 1000$. We conclude $\lim_{n\to\infty} \frac{n}{n+1} = 1$

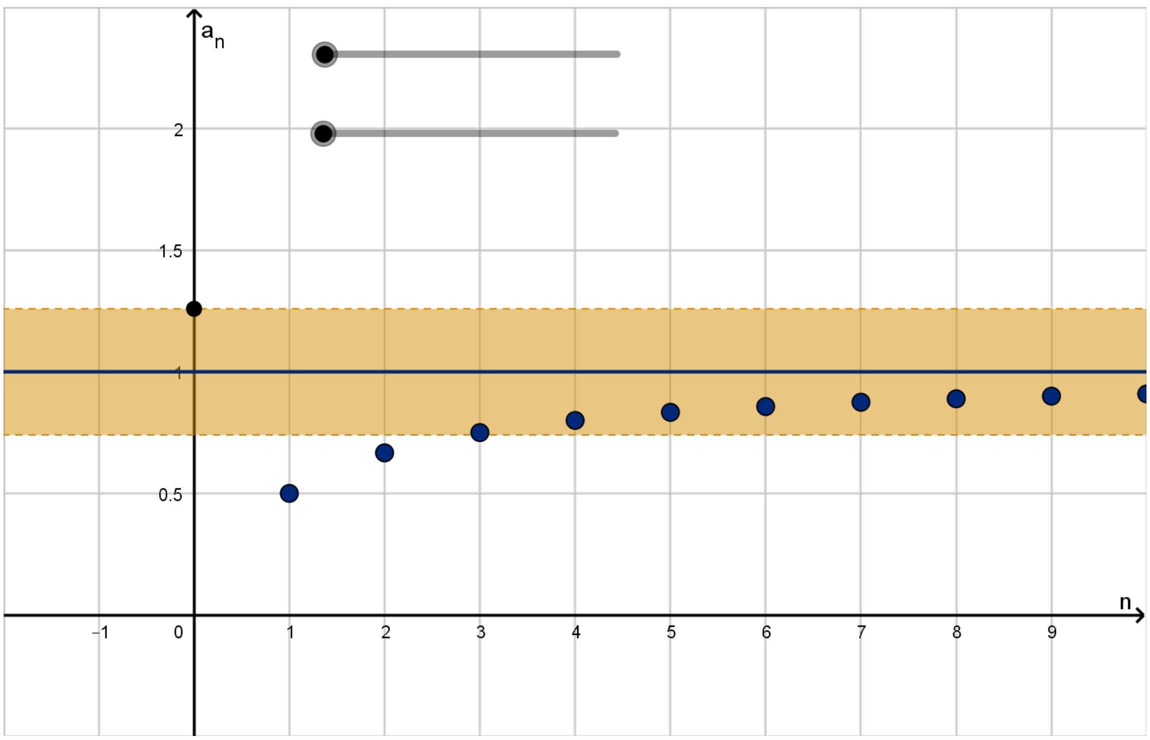


Figure: The sequence $\frac{n}{n+1}$ converges to $L = 1$.

Question 3.2.4 How Are Limits of Sequences and Functions Related?

The definition of $\lim_{n\to\infty} a_n$ should look familiar. The definition of the limit of a function is similar. In fact, the limit of a $f(x)$ as $x \to \infty$ has a nearly identical construction, except that n must be an integer, while x can be any real number. The following theorem lets us use that connection to evaluate limits.

Theorem

Suppose for a sequence a_n, there is a function $f(x)$ such that $f(n) = a_n$ for all n (or at least all n sufficiently large). If

$$\lim_{x\to\infty} f(x) = L$$

we can conclude that

$$\lim_{n\to\infty} a_n = L.$$

Example 3.2.5

Sequence Limits Using Functions

Find limits of the following sequences:

a $\lim\limits_{n\to\infty} \dfrac{2n}{n+3}$

b $\lim\limits_{n\to\infty} \dfrac{1}{n^3}$

c $\lim\limits_{n\to\infty} e^{-n}$

d $\lim\limits_{n\to\infty} \dfrac{n^2}{e^n}$

e $\lim\limits_{n\to\infty} (-1)^n$

Solution

We will use x to denote a real number variable and n to denote natural numbers.

a $\lim\limits_{x\to\infty} \dfrac{2x}{x+3} = 2$, so $\lim\limits_{n\to\infty} \dfrac{2n}{n+3} = 2$.

b $\lim\limits_{x\to\infty} \dfrac{1}{x^3} = 0$, so $\lim\limits_{n\to\infty} \dfrac{1}{n^3} = 0$.

c $\lim\limits_{x\to\infty} e^{-x} = 0$ so $\lim\limits_{n\to\infty} e^{-n} = 0$.

Example 3.2.5 Sequence Limits Using Functions

d $\lim_{x\to\infty} \frac{x^2}{e^x}$ can be evaluated with L'hôpital's rule.

$$\begin{aligned}\lim_{x\to\infty} \frac{x^2}{e^x} &= \lim_{x\to\infty} \frac{2x}{e^x} && \left(\frac{\infty}{\infty}\text{ form, L'hôpital's again}\right)\\ &= \lim_{x\to\infty} \frac{2}{e^x}\\ &= 0\text{so } \lim_{n\to\infty} \frac{n^2}{e^n} && = 0\end{aligned}$$

e $f(x) = (-1)^x$ is not well defined for real numbers so we can't use its limit. Instead examine the sequence directly. The sequence has the form

$$-1, 1, -1, 1, -1, 1, -1, 1, \ldots$$

This does not approach arbitrarily close to any number. No matter how many early terms we disregard, there will always be terms remaining that are not close to 1, or not close to -1 or not close to any other number. Thus $a_n = (-1)^n$ diverges.

The following limit laws for sequences should look familiar. They mirror the laws for limits of functions.

Theorem [Limit Laws]

If $\lim_{n\to\infty} a_n = K$ and $\lim_{n\to\infty} b_n = L$ then the following sequences converge with the following limits:

- $\lim_{n\to\infty} (a_n + b_n) = K + L$
- $\lim_{n\to\infty} (a_n - b_n) = K - L$
- $\lim_{n\to\infty} (a_n b_n) = KL$
- If $L \neq 0$, then $\lim_{n\to\infty} \frac{a_n}{b_n} = \frac{K}{L}$
- For any constant c, $\lim_{n\to\infty} ca_n = cK$

We will encounter sequences of the form $a_n = \frac{b_n}{c_n}$. If b_n or c_n both go to 0 or $\pm\infty$, then any attempt to use

$$\lim_{n\to\infty} a_n = \lim_{x\to\infty} f(x)$$

would require l'Hôpital's rule.

Dominance

We say $f(x)$ **dominates** $g(x)$ if $\lim_{x\to\infty} \dfrac{f(x)}{g(x)} = \pm\infty$. We write

$$f(x) >> g(x)$$

Even if you include a constant multiple or add multiple functions together, the dominant function will outgrow any combination of dominated ones. We have already established an order of dominance using l'Hôpital's rule:

$$\underset{\text{(larger base}>>\text{smaller base)}}{\text{exponential}} >> \underset{\text{(larger degree}>>\text{smaller)}}{\text{polynomial}} >> \underset{\text{(smaller power}>>\text{larger)}}{\text{root}} >> \underset{\text{(smaller base}>>\text{larger)}}{\text{logarithm}}$$

But $n!$ is not a differentiable function. We cannot analyze it using l'Hôpital's rule. Where does it fit in the domincance pecking order?

Theorem

As $n \to \infty$, $n!$ will eventually **dominate** any exponential function (and thus any polynomial, root or logarithm).

We will not provide a formal proof, but here is a useful thought experiment. Suppose we compare $n!$ to 63^n. At first 63^n grows faster, multiplying by 63 every time we increase n. However, when n is greater than 63, $n!$ is multiplying by a higher number. When n reaches one billion, 63^n increases by a factor of 63 every step, while $n!$ increases by a factor of $1,000,000,000$. By this point $n!$ is much larger and growing much faster.

Section 3.2

Exercises

Summary Questions

Q1 Why do we use n instead of x as an index for a sequence?

Q2 Describe three different ways of denoting a sequence.

Q3 When is the limit of a sequence equal to the limit of a function?

Q4 If $a_n = b_n + 1000$ for $1 \leq n \leq 2000000$, what does that tell us about the limits $\lim_{n\to\infty} a_n$ and $\lim_{n\to\infty} b_n$?

3.2.1

Q5 Find a general expression for a_n, the n^{th} term of the following sequences. Use this to write the sequences using **both** other types of notation.

a $\{2, 5, 10, 17, 26, 37, 50, \ldots\}$

b $\left\{\frac{3}{2}, -\frac{3}{4}, \frac{3}{8}, -\frac{3}{16}, \frac{3}{32}, \ldots\right\}$

c $\left\{\frac{1}{2}, \frac{1}{6}, \frac{1}{12}, \frac{1}{20}, \frac{1}{30}, \ldots\right\}$

Q6 What is the fourth term in the sequence $\{n^3 - 5n\}_{n=3}^{\infty}$?

3.2.2

Q7 Show using the definition of the limit of a sequence that $\lim_{n\to\infty} \frac{\sin n}{n^2} = 0$.

Q8 Show using the definition of the limit of a sequence that $\lim_{n\to\infty} \frac{2^n - 1}{2^n} = 1$.

Q9 A sequence is increasing if every term is larger than the previous term. Must an increasing sequence always diverge? Explain.

Q10 A sequence is alternating if its terms alternate between positive and negative values. Is it possible that the limit of an alternating sequence exists? What would its value have to be?

3.2.3

Q11 Consider the sequence $a_n = 2^n$.

- **a** What function could we write such that $f(n) = a_n$.
- **b** Does $\lim_{x\to\infty} f(x)$ converge?
- **c** Does the theorem equating limits of functions and sequences apply to this function?
- **d** Can we argue that $\lim_{n\to\infty} 2^n$ diverges anyway?

Q12 Consider the sequence $a_n = n\sin(\pi n)$

- **a** What is $\lim_{x\to\infty} x\sin(\pi x)$?
- **b** Compute the first few values a_1, a_2, a_3, and a_4.
- **c** What is $\lim_{n\to\infty} n\sin(\pi n)$?
- **d** Does this contradict one of our theorems? Explain.

3.2.4

Q13 Compute $\lim\limits_{n\to\infty} \dfrac{\log n}{3n}$.

Q14 Compute $\lim\limits_{n\to\infty} \dfrac{n}{2^n}$.

Q15 Compute $\lim\limits_{n\to\infty} \dfrac{n^3+3}{4n^3-9}$.

Q16 Compute $\lim\limits_{n\to\infty} \dfrac{\sin n}{\log n}$.

Q17 Compute $\lim\limits_{n\to\infty} \dfrac{e^n}{\sqrt{n}}$.

Q18 Compute $\lim\limits_{n\to\infty} \tan^{-1} n$.

3.2.5

Q19 Compute $\lim\limits_{n\to\infty} \dfrac{n!}{5^n}$.

Q20 Compute $\lim\limits_{n\to\infty} \dfrac{n^4+3n+1}{n!}$.

Q21 Does n^n grow faster or slower than $n!$? Explain.

Q22 Yuran knows that $\lim\limits_{n\to\infty} \dfrac{n!}{5^n} = \infty$ because $n!$ growns faster than a^n. However, he thinks he can make the denominator grown faster than the numerator if he uses a product like $\frac{n!}{5^n 6^n}$ or $\frac{n!}{5^n 6^n 7^n}$. Will he eventually obtain a non-infinite limit by this method? Explain how you know.

Synthesis & Extension

Q23 Suppose we have a sequence $a_n = \begin{cases} f(n) & \text{if } n \leq 342 \\ g(n) & \text{if } n > 342 \end{cases}$. Which of the following could help us evaluate $\lim_{n\to\infty} a_n$?

- $\lim_{x\to\infty} f(x)$
- $\lim_{x\to\infty} g(x)$

Q24 Let $T_n(x)$ be the nth Taylor polynomial of $f(x) = \ln x$ centered at $x = 1$.

a Write an expression for $T_n(x)$ using Σ notation.

b Write an expression for the error bound of $T_n(x)$ for some x between 0 and 1.

c For what values of x will the error bound shrink to 0 as n goes to ∞?

Section 3.3

Series

Goals:

1. Identify partial sums of a series.
2. Recognize harmonic and alternating harmonic series.
3. Apply the divergence test.
4. Evaluate geometric series.
5. Apply the ratio test.

The first step in understanding a Taylor polynomial of infinite degree is understanding how to add up infinitely many of anything. This proposition is mechanically absurd. Addition is an operation for two numbers at a time. Adding three or four numbers requires us to add two or three times. Adding infinitely many requires us to add infinitely many times, something no one has time to do.

Yet there are some intuitive exercises we could perform. Suppose we lay a length of $\frac{1}{2}$m next to $\frac{1}{4}$m next to $\frac{1}{8}$m. If we continued indefinitely, we could imagine these lengths extending an entire meter.

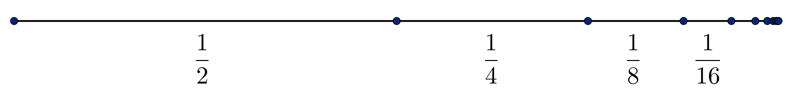


Figure: One meter expressed as a sum of infinitely many smaller lengths

What reasoning could we use to make this exercise rigorous? How could we add up lengths or numbers where the pattern is not so intuitive? The formal object that does this is called a series. A series is the first step on our way to push the Taylor polynomial to infinite degree. It is also the most general. While we are concerned with one specific (and very useful) type of series, there are other applications worth exploring as well.

Question 3.3.1

What Is a Series?

You have been encountering series since you first learned about decimals. You likely have not seen a rigorous description of what they mean.

$$0.33333333\ldots \qquad 3.1415926...$$

We can write

$$0.3333\ldots = \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \frac{3}{10000} + \cdots$$

or

$$3.1415\ldots = 3 + \frac{1}{10} + \frac{4}{100} + \frac{1}{1000} + \frac{5}{10000} + \cdots$$

You may have an intuitive sense of what these quantities are, but what does it mean to add up infinitely many numbers?

Definition

A **series** is a sum of the form $\sum_{k=1}^{\infty} a_k$ where a_k is an infinite sequence. If it is more convenient, we can give k a different initial value. If the context is clear, we can write $\sum a_k$ as a shorthand.

Example

- $0.33333\ldots = \sum_{k=1}^{\infty} \frac{3}{10^k}$
- The **harmonic series** is $\sum_{k=1}^{\infty} \frac{1}{k}$

This tells us what a series is but not how to evaluate it. How do we know that, for example

$$0.333\ldots = \frac{1}{3}?$$

We evaluate a series by associating it with a sequence of partial sums.

Definition

The n^{th} **partial sum** of the series $\sum_{k=1}^{\infty} a_k$ is

$$s_n = a_1 + a_2 + a_3 + \cdots + a_n$$

A series $\sum_{k=1}^{\infty} a_k$ **converges** to L if

$$\lim_{n\to\infty} s_n = L.$$

A series that does not converge to any L **diverges**.

Vocabulary Note

Do not confuse a sequence with a series. One is a list of numbers. The other is the sum of a list of numbers.

Example 3.3.2

Computing Partial Sums

Consider $\sum_{k=1}^{\infty} \frac{3}{10^k}$.

a Compute the first few partial sums s_1, s_2, s_3 of this series.

b Compute $\lim_{n\to\infty} s_n$

Solution

a

$$\begin{aligned}
s_1 &= \frac{3}{10} \\
s_2 &= \frac{3}{10} + \frac{3}{100} = \frac{33}{100} \\
s_3 &= \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} = \frac{333}{1000} \\
s_4 &= \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \frac{3}{10000} = \frac{3333}{10000}
\end{aligned}$$

b In order to use our usual methods of limits, we would need an algebraic expression for s_n. It isn't immediately clear how to produce one. Given our knowledge of decimals, we expect the answer to be $\frac{1}{3}$. We will use this as a hint. We expect $\frac{1}{3} - s_n$ to approach 0.

$$\begin{aligned}
\frac{1}{3} - s_1 &= \frac{1}{30} \\
\frac{1}{3} - s_2 &= \frac{1}{300} \\
\frac{1}{3} - s_3 &= \frac{1}{3000} \\
\frac{1}{3} - s_4 &= \frac{1}{30000} \\
\text{extrapolating suggests } \frac{1}{3} - s_n &= \frac{1}{3(10)^n}
\end{aligned}$$

Assuming this pattern holds, we have

$$\lim_{n\to\infty} \frac{1}{3} - s_n = \lim_{n\to\infty} \frac{1}{3(10)^n} = 0$$

and we conclude that $\sum_{k=1}^{\infty} \frac{3}{10^k} = \frac{1}{3}$.

Main Idea

Often, we can show that $\sum_{k=1}^{\infty} a_k = L$ by computing $L - s_n$ and seeing that it converges to 0.

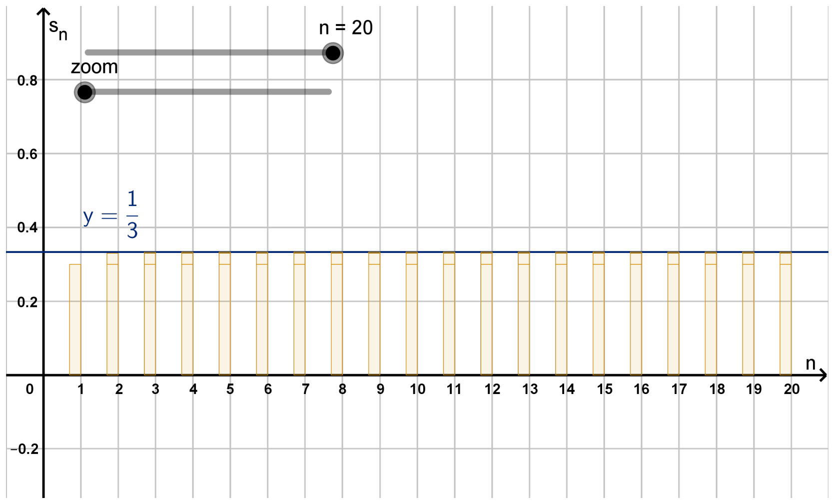


Figure: The partial sums s_n converging to $L = \frac{1}{3}$

Example 3.3.3

The Harmonic Series

We have seen examples of series in which the terms approach 0 as $k \to \infty$. These have allowed us to add infinitely many terms and obtain a finite sum. Does this always work? No. A series can have its terms approach 0, and yet the partial sums go to ∞. The most famous example of this is the harmonic series: $\sum_{k=1}^{\infty} \frac{1}{k}$. Rather than computing the partial sums directly (which would be a lot of computation) we will compare the partial sums to an expression that is easier to calculate. We will replace each term by a fraction with a power of 2 in the denominator. Here's what we'll do with s_8.

$$\begin{aligned}
s_8 =& \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} \\
>& \frac{1}{1} + \frac{1}{2} + \underbrace{\frac{1}{4} + \frac{1}{4}}_{\frac{1}{2}} + \underbrace{\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}}_{\frac{1}{2}} \\
=& 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}
\end{aligned}$$

Since we replaced each term with something smaller and obtained a sum of $\frac{5}{2}$, we can conclude that $s_8 > \frac{5}{2}$. Continuing this pattern, the terms $\frac{1}{9}$ to $\frac{1}{16}$ sum to more than $\frac{1}{2}$ so $s_{16} > \frac{6}{2}$. In general we can

Example 3.3.3 The Harmonic Series

make s_n bigger than any integer c by setting $n = 2^m$ where

$$1 + \frac{1}{2}m > c.$$

This tells us that the harmonic series diverges.

Question 3.3.4

What Is a Geometric Series?

The two series so far that we have been able to evaluate belonged to a larger family. These are the geometric series.

Definition

A **geometric series** is a series of the form $\sum_{k=1}^{\infty} ar^{k-1}$.

a is the initial term. r is the common ratio between terms.

Example

- $\sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{k-1} = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$
- $\sum_{k=1}^{\infty} \frac{3}{10}\left(\frac{1}{10}\right)^{k-1} = \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \cdots = \frac{1}{3}$

Unlike many other series, geometric series are simple enough that we can write a formula for their sum. We can get a convenient expression for s_n by performing a cute algebra trick. We'll multiply s_n by r and subtract rs_n from s_n. Most of the terms cancel and we obtain an equation that we can solve for s_n.

$$\begin{aligned} s_n &= a + ar + ar^2 + \cdots + ar^{n-1} \\ -rs_n &= -ar - ar^2 - ar^3 - \cdots - ar^n \\ (1-r)s_n &= a - ar^n \\ s_n &= \frac{a(1-r^n)}{1-r} \end{aligned}$$

The last step requires that $1 - r \neq 0$, since we cannot divide by 0. As long as $r \neq 1$, we can evaluate the series by taking a limit.

$$\sum_{k=1}^{\infty} ar^{k-1} = \lim_{n\to\infty} \frac{a(1-r^n)}{1-r}$$

To evaluate this limit, we need to understand the behavior of r^n as $n \to \infty$

- If $-1 < r < 1$ then higher powers of r get smaller and smaller and $r^n \to 0$.
- If $r > 1$ then higher powers of r get larger and larger and $r^n \to \infty$.
- If $r < -1$ then higher powers of r get larger but alternate signs. $\lim\limits_{n\to\infty} r^n$ does not exist.
- If $r = 1$ then the series is $a+a+a+a+a+\cdots$, then $s_n = an$ which diverges to $\pm\infty$, depending on the sign of a.
- If $r = -1$ then the series is $a - a + a - a + a - \cdots$, then s_n alternates between a and 0. This sequence does not converge.

We can apply the above to completely solve the problem of evaluating a geometric series. Our result is the following theorem:

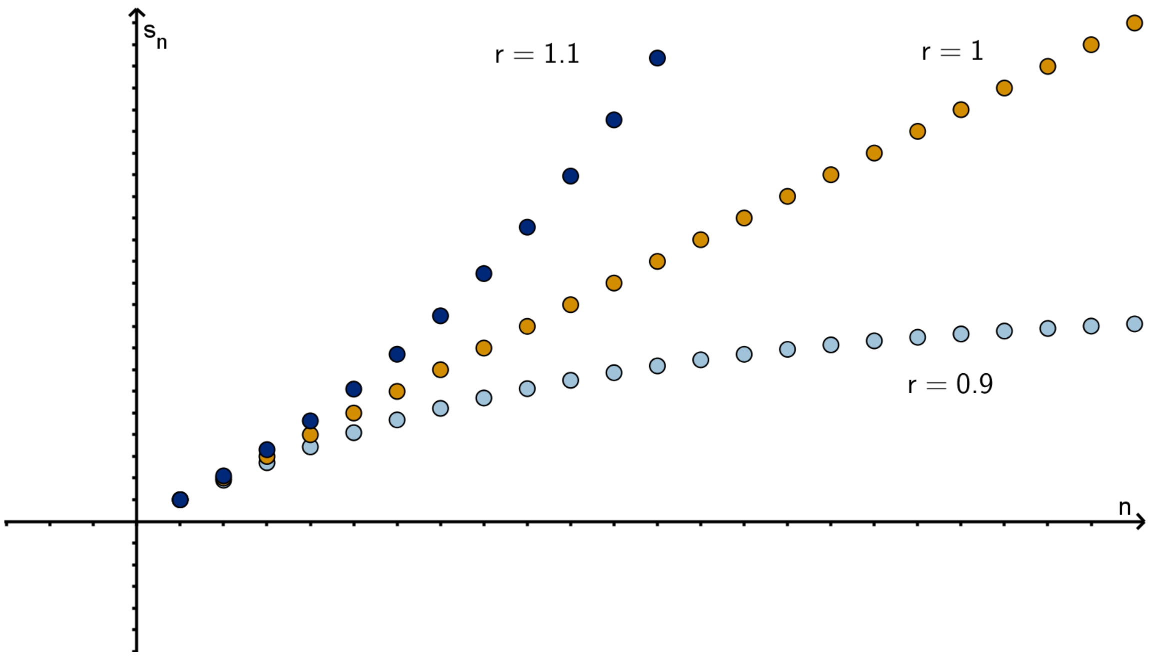


Figure: The partial sums of $\sum ar^{k-1}$ for various r

Theorem

Geometric series have the following partial sums

$$s_n = \sum_{k=1}^{n} ar^{k-1} = \begin{cases} \dfrac{a(1-r^n)}{1-r} & \text{if } r \neq 1 \\ an & \text{if } r = 1 \end{cases}$$

These converge to $\dfrac{a}{1-r}$ when $|r| < 1$ and diverge when $|r| \geq 1$.

Example 3.3.5

Evaluating Geometric Series

Identify a and r in the following geometric series. Then evaluate the series.

a $\frac{2}{3} + \frac{4}{15} + \frac{8}{75} + \cdots$

b $\displaystyle\sum_{n=2}^{\infty} 3^n$

c $0.999999\ldots$

Solution

a a is the initial term, which is $\frac{2}{3}$. The common ratio is the ratio between any two terms. $\frac{4/15}{2/3} = \frac{2}{5}$.

Since $|r| < 1$, the sum of the series is

$$\sum_{k=1}^{\infty} \left(\frac{2}{3}\right)\left(\frac{2}{5}\right)^{k-1} = \frac{\frac{2}{3}}{1-\frac{2}{5}} = \frac{\frac{2}{3}}{\frac{3}{5}} = \frac{10}{9}$$

b The initial term of this series is 9. The common ratio is 3. Since $|3| \geq 1$, $\displaystyle\sum_{n=2}^{\infty} 3^n$ diverges.

c $0.999999\ldots = \frac{9}{10} + \frac{9}{100} + \frac{9}{1000} + \cdots$. This has an initial term of $\frac{9}{10}$ and a common ratio of $\frac{1}{10}$.

$|r| < 1$ so

$$0.999999\ldots = \frac{\frac{9}{10}}{1-\frac{1}{10}} = \frac{\frac{9}{10}}{\frac{9}{10}} = 1$$

Question 3.3.6

What Does the Size of a_k Tell Us About $\sum a_k$?

The discussion of the geometric series suggests that certain properties of a series make convergence impossible. Specifically, in the cases in which the terms were not shrinking to 0, the partial sums were growing without bound or oscillating. This intuition can be formalized in the following theorem, which applies to more than just geometric series.

Theorem [The Divergence Test]

Let a_k be a sequence. If $\lim\limits_{k\to\infty} a_k \neq 0$, then the series

$$\sum_{k=1}^{\infty} a_k$$

diverges.

Remark

The divergence test does not tell us anything, if $\lim\limits_{k\to\infty} a_k = 0$. The series might converge, and it might not. In this case we say the test is **inconclusive**.

Example 3.3.7

Applying the Divergence Test

What does the divergence test tell us about each of the following series?

a $\displaystyle\sum_{k=2}^{\infty} 3^k$

b $\displaystyle\sum_{k=2}^{\infty} \frac{1}{k}$

c $\displaystyle\sum_{k=2}^{\infty} \frac{k^2-1}{3k^2+7}$

d $\displaystyle\sum_{k=2}^{\infty} \frac{k^2}{e^k}$

Example 3.3.7 Applying the Divergence Test

Solution

a The sequence is $a_k = 3^k$. $\lim_{k\to\infty} 3^k = \infty$. This limit is not 0, so by the divergence test, the series diverges.

b The sequence is $a_k = \frac{1}{k}$. $\lim_{k\to\infty} \frac{1}{k} = 0$. The divergence test is inconclusive. It cannot tell us whether this series diverges or converges. By our earlier work, we happen to know this series diverges.

c The sequence is $a_k = \frac{k^2-1}{3k^2+7}$. $\lim_{k\to\infty} \frac{k^2-1}{3k^2+7} = \frac{1}{3}$. This limit is not 0, so by the divergence test, the series diverges.

d The sequence is $a_k = \frac{k^2}{e^k}$. We need L'Hôptial's rule to evaluate the limit.

$$\begin{aligned}\lim_{k\to\infty} \frac{k^2}{e^k} &= \lim_{k\to\infty} \frac{2k}{e^k} && \left(\text{still } \frac{\infty}{\infty} \text{ form}\right)\\ &= \lim_{k\to\infty} \frac{2}{e^k}\\ &= 0\end{aligned}$$

The divergence test is inconclusive. It cannot tell us whether this series diverges or converges. It turns out that this series converges, but we do not have a method to verify that yet.

Question 3.3.8

What Is the Ratio Test?

So far we have two tests to determine the convergence of a series. One test is very specific, applying only to geometric series. The other is very imprecise. The divergence test is often inconclusive. It does not help us to evaluate a series at all, only recognizing some series that diverge. Unfortunately, these shortcoming are typical of series tests. A rigorous study of infinite series requires learning almost a dozen tests. On a randomly chosen series, most of these tests will be inconclusive, and none of them will give a numerical value, even if the series happens to converge. Because we are interested in extending Taylor polynomials to have infinitely many terms, some of these tests are much more useful than others. The most useful is the ratio test, though it is still no help in evaluating a series and is still sometimes inconclusive.

In the case of a geometric series, $\sum ar^{k-1}$, the common ratio between terms determines whether this series grows out of control, or whether the terms shrink quickly enough that the partial sums converge. Even when a series is not geometric, we can attempt to apply similar reasoning to determine whether it converges. A non-geometric series does not have a constant ratio. The ratio between successive terms will change as we progress through them. We will instead compute the limit of these ratios.

Theorem [The Ratio Test]

If $\lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right| = L < 1$, then $\sum a_k$ **converges absolutely**.

If $\lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right| = L > 1$ or is infinite, then $\sum a_k$ is divergent.

If $\lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right| = 1$, then the ratio test is inconclusive.

Remark

Converges absolutely is a term for series with both positive and negative terms. It means the series would converge, even if the signs of all the terms were all positive. The alternative is **conditional convergence**, meaning the series's convergence may require the positive and negative terms partially canceling each other out.

Example

The series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \cdots$$

converges (we won't prove this). If we made all the terms positive, it would be the harmonic series, which diverges. This series converges conditionally, not absolutely.

Absolute versus conditional convergence can be interesting to play with. You may see references to it in other math books, but we won't have any further use for it.

Example 3.3.9

Applying the Ratio Test

a Does $\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!}$ converge or diverge?

b Does $\sum_{k=1}^{\infty} \frac{2^k}{k^2}$ converge or diverge?

c Does $\sum_{k=1}^{\infty} k$ converge or diverge?

Example 3.3.9 Applying the Ratio Test

Solution

a First we will compute and simplify the ratio. Then we will take its limit and draw a conclusion.

$$\begin{aligned}
\left|\frac{a_{k+1}}{a_k}\right| &= \left|\frac{\frac{(-1)^k}{(k+1)!}}{\frac{(-1)^{k-1}}{k!}}\right| \\
&= \left|\frac{(-1)^k k!}{(-1)^{k-1}(k+1)!}\right| \\
&= \left|\frac{(-1)^k (1)(2)(3)\cdots(k)}{(-1)^{k-1}(1)(2)(3)\cdots(k)(k+1)}\right| && \text{(expand the factorials)} \\
&= \left|\frac{(-1)^k}{(-1)^{k-1}(k+1)}\right| && \text{(cancel the matching factors)} \\
&= \left|\frac{-1}{k+1}\right| && \text{(cancel } k-1 \text{ powers of } -1\text{)} \\
&= \frac{1}{k+1} && \text{(absolute value of a negative number is its negatve)}
\end{aligned}$$

Now we take the limit

$$\lim_{k\to\infty} \frac{1}{k+1} = 0$$

$0 < 1$ so by the ratio test, $\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!}$ converges.

b We will apply the ratio test. First we compute the ratio, and then we take a limit.

$$\begin{aligned}
\left|\frac{a_{k+1}}{a_k}\right| &= \left|\frac{\frac{2^{k+1}}{(k+1)^2}}{\frac{2^k}{k^2}}\right| \\
&= \left|\frac{2^{k+1}k^2}{2^k(k^2+2k+1)}\right| \\
&= \left|\frac{2k^2}{k^2+2k+1}\right| && \text{(cancel the 2s)} \\
&= \frac{2k^2}{k^2+2k+1}
\end{aligned}$$

$$\lim_{k\to\infty} \frac{2k^2}{k^2+2k+1} = 2$$

$2 > 1$ so by the ratio test, this series diverges.

c We will apply the ratio test. First we compute the ratio, and then we take a limit.

$$\begin{aligned}\left|\frac{a_{k+1}}{a_k}\right| &= \left|\frac{k+1}{k}\right| \\ &= \frac{k+1}{k} \lim_{k\to\infty} \frac{k+1}{k} \qquad\qquad = 1\end{aligned}$$

Here the ratio test is inconclusive. It cannot tell whether this series converges or diverges. However, we can probably figure this out another way. The terms of this series are increasing, which means the partial sums will grow faster and faster. This was the reasoning behind the divergence test.

$$\lim_{k\to\infty} k = \infty$$

Since $\lim_{k\to\infty} k \neq 0$, the divergence test concludes that the series diverges.

Main Ideas

- When applying the ratio test, be sure to replace every k with $k+1$ for the a_{k+1} term.
- Familiarize yourself with the algebra rules that allow you to simplify ratios of exponentials and factorials.

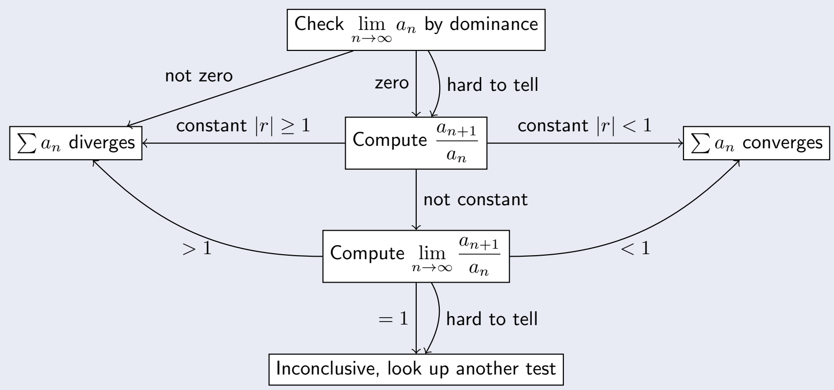
Example 3.3.10

A Strategy for Series Tests

Example 3.3.10 A Strategy for Series Tests

Strategy

Given the three ways we have to test for divergence and convergence and the relative ease of applying each, here is a reasonable approach to testing a series.



Let's apply our strategy to see what we can tell about

$$\sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Solution

First we'll check that the terms go to zero. If they don't we quickly classify this as a divergent series.

$$\lim_{n\to\infty} \frac{1}{n^2} = 0$$

They do, so we need another check. Now we'll compute the ratio between terms.

$$\frac{a_{n+1}}{a_n} = \frac{\frac{1}{(n+1)^2}}{\frac{1}{n^2}} = \frac{n^2}{n^2+2n+1}$$

This is not a constant; it depends on n. Thus a_n is not a geometric series. We'll try the ratio test.

$$\begin{aligned}
\lim_{n\to\infty} \left|\frac{a_{n+1}}{a_n}\right| &= \lim_{n\to\infty} \left|\frac{n^2}{n^2+2n+1}\right| \\
&= \lim_{n\to\infty} \frac{n^2}{n^2+2n+1} \\
&= 1
\end{aligned}$$

This means that the ratio test is inconclusive. We do not know whether this series converges or diverges. We have exhausted all our tests. If we want the answer, we need to look up another test.

Summary Questions

Q1 What is the difference between a sequence and a series?

Q2 How do we evaluate a series?

Q3 What is a geometric series. How do we evaluate one?

Q4 What does it mean to say that a series test is inconclusive?

Q5 How do each of the following factors behave in the ratio $\left|\dfrac{a_{k+1}}{a_k}\right|$?

- **a** k^p (p a constant)
- **b** c^k (c a constant)
- **c** $k!$

Q6 How would the ratio test apply to a geometric series $\sum ar^{k-1}$?

3.3.1

Q7 Give a more common name for each of the following series.

- **a** $2 + \dfrac{7}{10} + \dfrac{1}{100} + \dfrac{8}{1000} + \dfrac{2}{10000} + \dfrac{8}{100000} + \cdots$
- **b** $\dfrac{6}{10} + \dfrac{6}{100} + \dfrac{6}{1000} + \dfrac{6}{10000} + \cdots$

Q8 Use a calculator to get a decimal approximation of $\frac{25}{33}$ and write it as a series of fractions with powers of 10 as denominators.

3.3.2

Q9 Consider the series

$$\sum_{k=1}^{\infty} \frac{1}{k(k+1)}$$

a Compute the first four elements in the series.

b Compute the partial sums: s_1, s_2, s_3, s_4.

c What do the partial sums appear to be converging to?

d Can you use algebra to generalize your answer to **2** to s_n?

Q10 Compute the first 3 partial sums of $\sum_{k=1}^{\infty} \frac{k+1}{k^2}$. Don't simplify the arithmetic.

Q11 Compute the first four partial sums of $\sum_{k=1}^{\infty} (-1)^k$. What do you think this suggests about the sum of the series?

Q12 Compute the first five partial sums of $\sum_{k=0}^{\infty} \frac{1}{(-2)^k}$. Use them to make a prediction about the value of the series.

3.3.3

Q13 Give an example of an n such that you know the nth partial sum of the harmonic series is greater than 20.

Q14 Modify our argument for the harmonic series to show that $\sum_{k=0}^{\infty} \frac{1}{\sqrt{k}}$ diverges?

3.3.4

Q15 Is $\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\frac{1}{8}+\cdots$ a geometric series? How can you tell?

Q16 Is $1+4+9+16+25+\cdots$ a geometric series? How can you tell?

Q17 The first two terms of a geometric series are 5 and 7.5. What is the third term?

Q18 The fifth term of a geometric series is 17. The eigth term is 51. What is the sixth term?

3.3.5

Q19 Evaluate $\displaystyle\sum_{k=0}^{\infty} 5(0.3)^k$

Q20 Evaluate $\displaystyle\sum_{k=0}^{\infty} \frac{1}{4}\left(\frac{4}{3}\right)^k$.

Q21 Evaluate $\displaystyle\sum_{j=3}^{\infty} \frac{15}{5^j}$.

Q22 Evaluate $\displaystyle\sum_{k=1}^{\infty} 0.8^k$.

Q23 Evaluate $\displaystyle\sum_{k=4}^{\infty} \frac{3^k}{2^k(18)}$.

Q24 Evaluate $\displaystyle\sum_{k=1}^{\infty} \frac{37}{100^k}$. What decimal does this represent?

Q25 For what values of z does $\displaystyle\sum_{k=0}^{\infty} \frac{3^k}{z^k}$ converge?

Q26 For what values of p does $\displaystyle\sum_{k=3}^{\infty} \frac{12p^{2k}}{16^k}$ converge?

3.3.6

Q27 If $a_k > \frac{1}{100}$ for all k, then what can you say about the value of $s_n = \sum_{k=1}^{n} a_k$?

Q28 If $\lim_{k\to\infty} a_k = \frac{1}{100}$, use the definition of a limit and the reasoning in the previous exercise to show that $\sum_{k=1}^{\infty} a_n$ diverges.

3.3.7

Q29 What does the divergence test say about $\sum_{k=1}^{\infty} \frac{1}{k^3}$?

Q30 What does the divergence test say about $\sum_{k=1}^{\infty} \frac{k^2+1}{5k^2+3k}$?

Q31 What does the divergence test say about $\sum_{k=2}^{\infty} \ln k$?

Q32 What does the divergence test say about $\sum_{k=2}^{\infty} \frac{1}{\ln k}$?

3.3.8

Q33 Will the divergence test detect every series that "fails" the ratio test $(L > 1)$? Explain.

Q34 If $\lim_{n\to\infty} \left|\frac{an+1}{a_n}\right|$ does not exist, the ratio test is inconclusive. Give examples of two series where this limit does not exist, one series that diverges and one that converges.

3.3.9

Q35 Apply the ratio test to $\sum_{k=1}^{\infty} \frac{k!}{4^k}$. What can you conclude?

Q36 Apply the ratio test to $\sum_{k=1}^{\infty} \frac{k5^k}{(k+1)!}$. What can you conclude?

Q37 Apply the ratio test to $\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^2}$. What can you conclude?

Q38 Apply the ratio test to $\sum_{k=1}^{\infty} \frac{(-8)^k}{k^2 5^k}$. What can you conclude?

Q39 Apply the ratio test to $\sum_{k=1}^{\infty} \frac{k^2}{4^k}$. What can you conclude?

Q40 Apply the ratio test to $\sum_{k=3}^{\infty} \frac{k!}{5k^3+4k-2}$. What can you conclude?

Q41 Apply the ratio test to $\sum_{k=1}^{\infty} \frac{\sqrt{k+1}}{k^2}$ What can you conclude?

Q42 Apply the ratio test to $\sum_{k=1}^{\infty} \sqrt{k}e^{-k}$ What can you conclude?

3.3.10

Q43 Use one of the tests from this section to deterine whether $\sum_{k=1}^{\infty} \frac{k+1}{k}$ converges.

Q44 Use one of the tests from this section to deterine whether $\sum_{k=1}^{\infty} \frac{3(4^k)}{7^k}$ converges.

Q45 Use one of the tests from this section to deterine whether $\sum_{k=1}^{\infty} \frac{ke^k}{4^{k+1}}$ converges.

Q46 Use one of the tests from this section to deterine whether $\sum_{k=1}^{\infty} \frac{7k9^k}{k3^{2k+1}}$ converges.

Synthesis & Extension

Q47 In a paragraph or two, explain: How is evaluating an improper integral similar to evaluating an infinite series. How are they different?

Q48 Suppose we have a sequence a_n such that $\lim_{n\to\infty} a_n = 30$. Suppose we then increase the values of the first few terms of a_n by $10,000$ each.

a Explain how this will affect the value of $\lim_{n\to\infty} a_n$.

b Explain how this will affect the value of $\sum_{n=1}^{\infty} a_n$.

Q49 Suppose we wanted to approximate $\int_0^\infty \frac{1}{e^x}\,dx$ by rectangles of length $\Delta x = 1$, with heights measured at the left endpoints.

a What are the areas of the first 5 rectangles, starting from $x = 0$?

b How many rectangles will you need in total?

c Express the sum of the areas of these rectangles as a series.

d Does this series converge? To what value?

e Does your series over- or underestimate the true value of the integral?

Q50 Suppose we wanted to approximate $\int_1^\infty \frac{1}{x^2}\,dx$ by rectangles of length $\Delta x = 1$, with heights measured at the right endpoints.

a What are the areas of the first 5 rectangles, starting from $x = 1$?

b Express the sum of the areas of all the the rectangles you'll need as a series.

c Does your series over- or underestimate the true value of the integral?

d What is the true value of the integral? What does this suggest about whether your series converges or diverges?

Q51 Suppose that a discrete random variable X has distribution function

$$f_X(x) = \begin{cases} \frac{1}{2^x} & \text{if } x \text{ is a positive integer} \\ 0 & \text{otherwise} \end{cases}$$

a. Verify that $f_X(x)$ is a valid probability distribution function.

b. Compute $P(X > 4)$.

c. Compute $E[X]$ (this is difficult).

Q52 Suppose that a discrete random variable X has distribution function

$$f_X(x) = \begin{cases} \frac{1}{x} - \frac{1}{x+1} & \text{if } x \text{ is a positive integer} \\ 0 & \text{otherwise} \end{cases}$$

a. Verify that $f_X(x)$ is a valid probability distribution function.

b. Compute $P(3 \leq X \leq 5)$.

c. Explain why you can't compute $E[X]$.

Section 3.4

Power Series

Goals:

1. Use series tests to determine for what values of x a power series converges.
2. Identify the radius of convergence of a power series.
3. Recognize functions that can be rewritten as a power series.

The infinite degree polynomials we seek to define are series. The tools we've developed so far provide the foundation for understanding the objects we want to construct, but there is more to do. A polynomial also contains a variable. In this section we deal with the ramifications of including a variable in an infinite series.

Question 3.4.1

What Is a Power Series?

So far we have studied infinite series of numbers. If instead of just numbers, our terms include variables, then we've created a function. Plugging in different values for the variable gives us a different series of numbers.

Example

The expression

$$1 + x + x^2 + x^3 + \cdots$$

becomes

$$1 + 2 + 4 + 8 + \cdots$$

when we evaluate it at $x = 2$. It becomes

$$1 - \frac{1}{3} + \frac{1}{9} - \frac{1}{27} + \cdots$$

when we evaluate it at $x = -\frac{1}{3}$.

Definition

An infinite series of the form

$$\sum_{k=0}^{\infty} c_k (x-a)^k$$

is called a **power series** centered at a.

It is a function of x whose domain is all values of x that make the series converge.

For the purposes of this definition, we define $x^0 = 1$ even when $x = 0$.

Example 3.4.2

A Geometric Series as a Power Series

Use the geometric series formula to write $f(x) = \dfrac{1}{1-x}$ as a power series and find its domain.

Solution

$\dfrac{1}{1-x}$ is the sum of a geometric series. In this case, the initial term $a = 1$ and the common ratio r is x. If we write out the first few terms we obtain $1 + x + x^2 + x^3 + \cdots$. We see this is a power series centered at 0. The coefficients c_k are all equal to 1. We could write it as $\sum_{k=0}^{\infty} x^k$.

The domain of a power series is the values of x that make it converge. We know that this geometric series converges if and only if the common ratio x has absolute value less than 1. Those values of x, the open interval $(-1, 1)$, are the domain of f.

Example 3.4.3

The Domain of a Power Series

What is the domain of $\displaystyle\sum_{k=1}^{\infty} \frac{k^2}{4^k}(x-5)^k$?

Solution

The domain is the set of x values that make the series converge. The ratio test will be helpful here. The ratio between terms is

$$\begin{aligned}
\left|\frac{a_{k+1}}{a_k}\right| &= \left|\frac{\frac{(k+1)^2}{4^{k+1}}(x-5)^{k+1}}{\frac{k^2}{4^k}(x-5)^k}\right| \\
&= \left|\frac{(k+1)^2 4^k (x-5)^{k+1}}{k^2 4^{k+1}(x-5)^k}\right| \\
&= \left|\frac{(k^2+2k+1)(x-5)}{4k^2}\right|
\end{aligned}$$

Notice this entire computation is invalid if $x = 5$, because we cannot divide by 0. We can examine this case directly. If $x = 5$ then every term of the series is 0, and the series converges. For the rest of the real numbers, we compute the limit as $k \to \infty$, but x will remain in the result.

$$\lim_{k\to\infty}\left|\frac{(k^2+2k+1)(x-5)}{4k^2}\right| = \left|\frac{(x-5)}{4}\right| \lim_{k\to\infty}\left|\frac{k^2+2k+1}{k^2}\right| \qquad = \left|\frac{(x-5)}{4}\right|$$

Example 3.4.3 The Domain of a Power Series

The ratio test can tell us whether the series converges for some values of x. If $\left|\frac{(x-5)}{4}\right| < 1$ the series converges. We can solve for x

$$\begin{aligned}
\left|\frac{(x-5)}{4}\right| &< 1 \\
\left|x-5\right| &< 4 && \text{(since } 4 > 0\text{)} \\
-4 < x-5 &< 4 \\
1 < x &< 9 && \text{(add } 5 \text{ to all three expressions)}
\end{aligned}$$

On the other hand, if $\left|\frac{(x-5)}{4}\right| > 1$ the series diverges. Solving for x follows a similar procedure.

$$\begin{aligned}
\left|\frac{(x-5)}{4}\right| &> 1 \\
\left|x-5\right| &> 4 && \text{(since } 4 > 0\text{)} \\
x-5 < -4 \text{ or } x-5 &> 4 \\
x < 1 \text{ or } x &> 9
\end{aligned}$$

What about when $x = 1$ or $x = 9$? $\left|\frac{(x-5)}{4}\right| = 1$ so the ratio test is indeterminate. We would need another test to resolve these points. In this case, we are lucky. If $x = 9$ the series becomes $\sum_{k=1}^{\infty}\left(\frac{k^2}{4}\right)(4)$. The divergence test is useful here: $\lim\limits_{k\to\infty} k^2 = \infty$. Since the terms do not approach 0, the series diverges. A similar argument works for $k = 1$.

Main Idea

The ratio test is usually successful in finding where a power series converges. Generally it is inconclusive at only two points. We will not always have a test that can tell us whether the series converges at these points.

You may notice a pattern in the types of domains we have computed for power series. That pattern is formalized in the theorem below, which tells us that the domain of a power series must take a very particular form.

Theorem

Given a power series $\sum_{k=0}^{\infty} c_k(x-a)^k$ centered at a, one of the following is true.

1. The series converges only when $x = a$.
2. The series converges when x is any real number.
3. There is a **radius of convergence** R such that
   a. The series converges when $|x-a| < R$, and
   b. The series diverges when $|x-a| > R$.

In case 3, the inequality $|x-a| < R$ solves to $a - R < x < a + R$, which means the domain is an interval centered at a and extending a distance R to either side. The theorem does not state whether this is a closed, open or half open interval. This reasoning extends intuitively, if not formally, to the other cases. 1 can the thought of as a (closed) interval extending distance 0 on either side. 2 would then be an interval extending infinitely on either side.

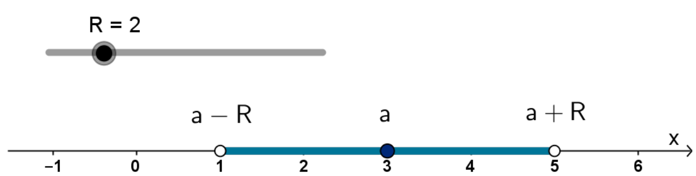


Figure: The domain $|x-a| < R$ of a power series.

Remark

The main consequence of this theorem is that when solving for the domain of a power series, we can simplify our use of the ratio test. The interval of convergence will always be the solution to $\lim_{k\to\infty} \left|\frac{ak+1}{a_k}\right| < 1$. The endpoints may or may not lie in the domain. The points beyond the endpoints will never be part of the domain.

Question 3.4.4
Can We Integrate or Differentiate a Power Series?

When $f(x)$ is a polynomial, we can find the derivative and anti-derivative of $f(x)$ by computing the (anti-)derivative of each term. The following theorem says that we can do this for a power series too.

Theorem

If $f(x)$ is the power series $\sum_{k=0}^{\infty} c_k(x-a)^k$ and $f(x)$ has radius of convergence $R > 0$ then $f(x)$ is differentiable and continuous on the interval $(a-R, a+R)$, and

1. $f'(x) = \sum_{k=1}^{\infty} kc_k(x-a)^{k-1}$
2. $\int f(x)\ dx = C + \sum_{k=0}^{\infty} c_k \frac{(x-a)^{k+1}}{k+1}$

Both of these functions also have radius of convergence R.

Remark

Notice that we remove the $k = 0$ term from the derivative. The derivative of that term is 0, but $0c_0(x-a)^{-1}$ is undefined at $x = a$.

Example

We have seen that $\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$ on the interval $(-1, 1)$. From that we can compute:

$$\frac{d}{dx}\sum_{k=0}^{\infty} x^k = \sum_{k=1}^{\infty} kx^{k-1}$$

$$\int \sum_{k=0}^{\infty} x^k\ dx = \sum_{k=0}^{\infty} \frac{x^{k+1}}{k+1} + c$$

Both have domain $(-1, 1)$.

Summary Questions

Q1 What is the difference between a polynomial and a power series?

Q2 What test is useful for establishing the domain of a power series? What form can this domain have?

Q3 How can we integrate or differentiate a power series?

Q4 How does differentiation affect the radius of convergence of a power series?

3.4.1

Q5 Use Σ notation to express the following series

a $10 + 15x + 20x^2 + 25x^3 + 30x^4 + \cdots$

b $\frac{1}{2} - \frac{1}{4}x^2 + \frac{1}{8}x^4 - \frac{1}{16}x^6 + \frac{1}{32}x^8 - \cdots$

Q6 Use Σ notation to express the following series

a $1 - \frac{x^2}{2} + \frac{x^4}{24} - \frac{x^6}{720} + \frac{x^8}{40640} - \cdots$

b $x^3 + 4x^4 + 9x^5 + 16x^6 + 25x^7 + \cdot$

3.4.2

Q7 Consider $f(x) = \frac{1}{1-4x^2}$.

- **a** If f is the sum of a geometric series, what is r?
- **b** Write $f(x)$ as a geometric series centered at 0.
- **c** What is the domain of your answer in **b**?

Q8 Write $\dfrac{5}{1-3(x-2)}$ as a power series centered at $x = 2$.

Q9 Can the power series $p(x) = \displaystyle\sum_{k=1}^{\infty} \frac{k^3}{4^k}(x+7)^k$ be evaluated using the sum of a geometric series formula? Explain.

Q10 Evaluate $f(x) = \displaystyle\sum_{k=3}^{\infty} \frac{1}{5^k}(x-2)^k$ at $x = 6$ using the formula for the sum of a geometric series.

3.4.3

Q11 What is the domain of $\displaystyle\sum_{k=1}^{\infty} 2^k (x-3)^k$?

Q12 Compute the domain of $\displaystyle\sum_{k=0}^{\infty} \frac{(x+2)^k}{k^3}$.

Q13 Compute the domain of $\displaystyle\sum_{k=0}^{\infty} \frac{1}{4^k}(x-6)^k$.

Q14 Compute the domain of $\displaystyle\sum_{k=0}^{\infty} \frac{x^k}{k!}$.

Q15 Compute the radius of convergence of $\displaystyle\sum_{k=0}^{\infty} k(x+3)^k$. What interval does this guarantee the series converges on?

Q16 Compute the radius of convergence of $\displaystyle\sum_{k=0}^{\infty} k!x^k$. What interval does this guarantee the series converges on?

Q17 Compute the radius of convergence of $\sum_{k=1}^{\infty} \frac{4k}{3^k}(x-5)^k$. What interval does this guarantee the series converges on?

Q18 Suppose you are told that a given power series $p(x)$ centered at $x=a$ converges at $x=-4$ and diverges at $x=-7$.

a If $a=1$, what can you say about the domain of $p(x)$?

b What are all of the the possible values of a? Explain your reasoning (briefly).

3.4.4

Q19 Compute the antiderivative of $\sum_{k=0}^{\infty} 2^k(x-3)^k$.

Q20 Compute the derivative of $\sum_{k=0}^{\infty} \frac{(x+2)^k}{k^3}$. What is its domain?

Q21 Compute the derivative of $\sum_{k=0}^{\infty} \frac{1}{4^k}(x-6)^k$. What is its domain?

Q22 Compute the antiderivative of $\sum_{k=0}^{\infty} \frac{x^k}{k!}$.

Q23 What is the domain of the fifth deriative of $\sum_{k=0}^{\infty} k(x+3)^k$?

Q24 Compute the radius of convergece of the antiderivative of $\sum_{k=4}^{\infty} \frac{4k}{3^k}(x-5)^k$.

Synthesis & Extension

Q25 Consider the power series

$$p(x) = \sum_{k=0}^{\infty} \frac{k^2 + k}{5^k}(x+3)^k.$$

a Compute the domain of P. You do not need to check any endpoints of your answer.

b Write an expression for

$$\int p(x)\, dx.$$

Q26 Consider the series $S = \sum_{k=1}^{\infty} \frac{k}{2^k}$.

a How is S related to the power series $p(x) = \sum_{k=1}^{\infty} \frac{kx^{k-1}}{2^k}$.

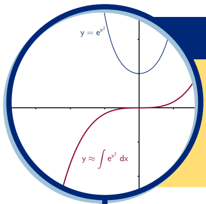
b Compute the an avtiderivative $P(x)$ of $p(x)$.

c Write $P(x)$ as ratio $F(x)$, using the sum of a geometric series formula.

d Compute $F'(1)$. What is the significance of this value?

Q27 Write a power series for $f(x) = \tan^{-1} x$ by

- Diffrentiating $f(x)$
- Writing $f'(x)$ as a geometric series
- Taking an antiderivative of the geometric series



Section 3.5

Taylor Series

Goals:

1. Use a combination of power series and algebra to work with functions.
2. Integrate and differentiate power series.

Our goal has been to understand how to extend a Taylor polynomial to have infinite degree. We are now ready to define the object rigorously. In general we will not know how to evaluate Taylor series. If all we want to do is approximate values, they offer no advantages over Taylor polynomials. The applications of Taylor series are more abstract. After defining these objects, we collect some tricks and applications for working with them.

Question 3.5.1

What Is a Taylor Series?

Definition

The **Taylor series** of $f(x)$ at $x = a$ is

$$T(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

The Taylor series's notation simply swaps an n for an ∞ in the expression of a Taylor polynomial. If we wanted to describe the mathematical relationship precisely, we would say its partial sums s_n are the Taylor polynomials $T_n(x)$ of f at $x = a$.

Remark

Several mathematicians contributed to the discovery of Taylor series. Taylor series centered at $x = 0$ were popularized by Colin Maclaurin, and so are often called **Maclaurin series**.

This definition is built upon a stack of more general definitions, and the methods we have for working with those apply here.

- A Taylor series is a type of power series.
- A power series is a type of series
- A series is equivalent to a sequence of partial sums.

This list should make us feel better about our hard work over the last few sections. It also gives us information that helps us understand Taylor series better. For example, since Taylor series are power series, their domains are also intervals of radius R centered at a.

Question 3.5.1 What Is a Taylor Series?

Limitations of Taylor Series

Taylor polynomials were designed to approximate $f(x)$. We might hope that $T(x)$ would be the perfect approximation, that $T(x)$ and $f(x)$ are equal. Unfortunately, there are obstacles to this.

- The Taylor series might not converge for all x.
- The Taylor polynomials might not approximate $f(x)$ very well at all. Recall our example

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ e^{-\frac{1}{x}} & \text{if } x > 0 \end{cases}$$

  For this function $T(x) = 0$.

Example 3.5.2

Writing a Taylor series

Let $f(x) = e^x$

a Find the Taylor series for $f(x)$ centered at $x = 0$.

b On what interval does it converge?

Solution

a We have seen previously that $f^{(k)}(x) = e^x$ for all k and thus $f^{(k)}(0) = 1$. We plug this into the Taylor series formula.

$$T(x) = \sum_{k=0}^{\infty} \frac{1}{k!} x^k$$

b A Taylor series is a power series. We will use the ratio test to identify the interval of convergence.

The ratio of successive terms is

$$\begin{aligned}
\left|\frac{a_{k+1}}{a_k}\right| &= \left|\frac{\frac{1}{(k+1)!}x^{k+1}}{\frac{1}{k!}x^k}\right| \\
&= \left|\frac{k!x^{k+1}}{(k+1)!x^k}\right| \\
&= \left|\frac{x}{k+1}\right|
\end{aligned}$$

$$\lim_{k\to\infty}\left|\frac{x}{k+1}\right| = 0$$

This limit is zero no matter what value of x we choose. Since $0 < 1$, the ratio test concludes that this series converges for any value of x. In other words, the domain is all real numbers.

Synthesis 3.5.3

Is a Taylor Series Equal to the Function it Approximates?

Let $f(x) = \ln x$

a Find a pattern in the derivatives and write a general expression for the kth derivative: $f^{(k)}(x)$.

b Use your answer to a to write expressions for the Taylor polynomials $T_n(x)$ and the Taylor series $T(x)$ of $\ln x$ centered at 1. Simplify the coefficients if possible.

c What does the ratio test tell you about where $T(x)$ converges?

d If we wanted to apply Taylor's inequality to $T_n(x)$, we would need to know where the derivative is largest (in absolute value). Where is the $(n+1)$th derivative largest on the interval $[x, 1]$? (Here $0 < x < 1$).

e Where is the $(n+1)$th derivative largest on the interval $[1, x]$? (Here $x > 1$).

f What does Taylor's inequality say about where $R_n(x) \to 0$ as $n \to \infty$?

g What does our answer to the previous question tell us about $T(x)$?

Solution

a Let's compute some derivatives and see if we can find an expression for $f^{(k)}(x)$

$$\begin{aligned} f(x) &= \ln(x) & f(1) &= 0 \\ f'(x) &= x^{-1} & f'(1) &= 1 \\ f''(x) &= -x^{-2} & f''(1) &= -1 \\ f'''(x) &= 2x^{-3} & f'''(1) &= 2 \\ f^{(4)}(x) &= -6x^{-4} & f^{(4)}(1) &= -6 \\ f^{(5)}(x) &= 24x^{-5} & f^{(5)}(1) &= 24 \end{aligned}$$

These answers look like factorials, but they're shifted by 1. They're also alternating signs, which we can model with $(-1)^k$, except that the even powers are negative. The power of x is $-k$. One way to model this is $f^{(k)}(x) = (-1)^{k+1}(k-1)!x^{-k}$.

b Plugging in $x = 1$ gives $f^{(k)}(1) = (-1)^{k+1}(k-1)!$ except at $k = 0$. For that case we compute $\ln 1 = 0$. This means we can leave it out of the summation. The form for the remaining terms allows for some nice simplification.

$$\begin{aligned} T(x) &= \sum_{k=1}^{\infty} \frac{(-1)^{k+1}(k-1)!}{k!}(x-1)^k \\ &= \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}(x-1)^k \end{aligned}$$

c We'll apply the ratio test

$$\begin{aligned} \left|\frac{a_{k+1}}{a_k}\right| &= \left|\frac{\frac{(-1)^{k+2}}{(k+1)}(x-1)^{k+1}}{\frac{(-1)^{k+1}}{k}(x-1)^k}\right| \\ &= \left|\frac{(-1)^{k+2}k(x-1)^{k+1}}{(-1)^{k+1}(k+1)(x-1)^k}\right| \\ &= \left|\frac{-k(x-1)}{k+1}\right| \\ &= \frac{k|x-1|}{k+1} \end{aligned}$$

Now we'll solve for when the limit of this ratio is less than 1.

$$\begin{aligned}\lim_{k\to\infty}\frac{k|x-1|}{k+1} &< 1\\ |x-1|\lim_{k\to\infty}\frac{k}{k+1} &< 1\\ |x-1| &< 1\\ -1 < x-1 &< 1\\ 0 < x &< 2\end{aligned}$$

The Taylor series converges on the interval $(0, 2)$.

d To apply Taylor's inequality to bound $|R_n(x)|$. We need a bound on $|f^{(n+1)}(x)|$ on the interval from 1 to x. Looking back at our earlier computation, we obtain $f^{(n+1)}(x) = (-1)^{n+2}n!x^{-n-1}$. In this case that $x > 1$, the derivative $f^{(n+1)}$ decreases in magnitude from x to 1 so it is largest at x. We can use $M = n!x^{-n-1}$.

e In this case, $f^{(n+1)}$ decreases in magnitude from 1 to x so it is largest at 1. We can use $M = n!$.

f The easier case is $x \geq 1$. In this case Taylor's inequality states

$$\begin{aligned}|R_n(x)| &\leq \left|\frac{n!}{(n+1)!}(x-1)^{n+1}\right|\\ &\leq \frac{1}{n+1}(x-1)^{n+1}\end{aligned}$$

As n approaches infinity, this bound goes to infinity if $x - 1 > 1$ and to 0 if $x - 1 \leq 1$. In the case that $0 < x < 1$, we need some clever algebra to write this as a multiple of an exponential.

$$\begin{aligned}|R_n(x)| &\leq \left|\frac{n!x^{-n-1}}{(n+1)!}(x-1)^{n+1}\right|\\ &\leq \frac{1}{n+1}\left|\left(\frac{x-1}{x}\right)^{n+1}\right|\end{aligned}$$

This goes to 0 if $\frac{x-1}{x} \leq 1$ and infinity otherwise. Solving this (and assuming $x > 0$) gives $x \geq \frac{1}{2}$. Putting these together, we can state that the error bound from Taylor's inequality approaches 0 as we takes higher degree Taylor polynomials, as long as $\frac{1}{2} \leq x \leq 2$.

g The answer to the previous question tells us that $T(x)$ converges to $\ln x$ on $\left(\frac{1}{2}, 2\right)$, since the error bound and hence the error goes to 0. On the other hand, outside this interval, the error might still go to 0 on $\left(0, \frac{1}{2}\right)$, even though the error bound does not. The series diverges outside $(0, 2)$ so it cannot converge to $\ln x$ there.

Remark

It turns out that $T(x) = \ln x$ on $(0, 2]$, which is a larger interval than we were able to establish using Taylor's inequality. This should not bother us. Taylor's inequality produces a bound on the error. The fact that the bound on the error is going to infinity, doesn't mean the actual error does. In this case, for x between 0 and $\frac{1}{2}$, the actual error approaches 0.

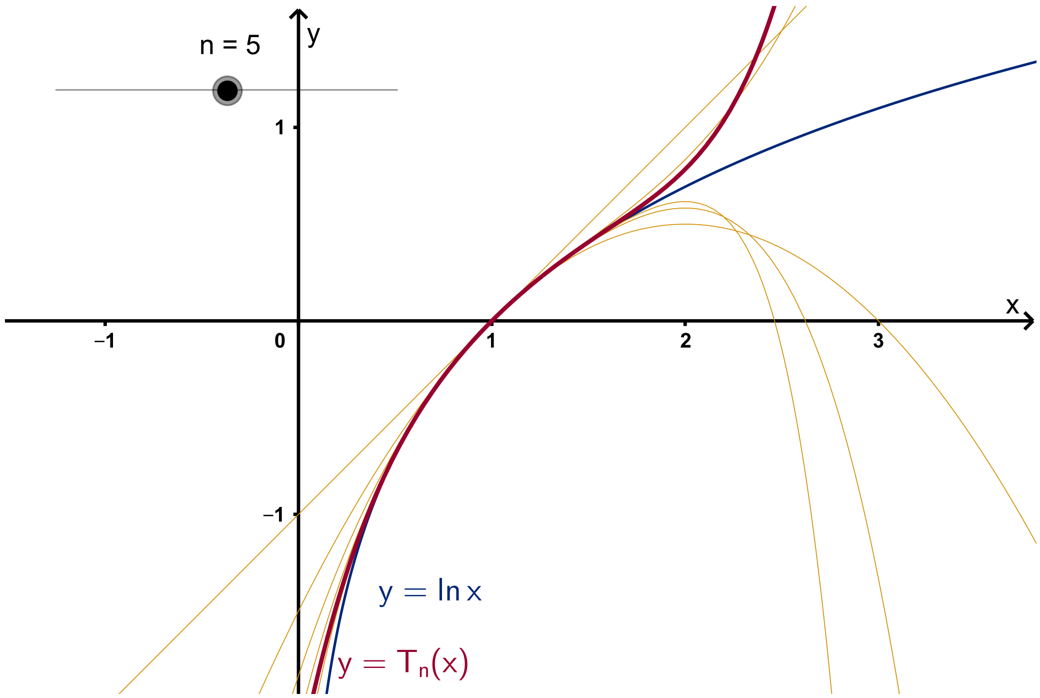


Figure: The Taylor polynomials approach $\ln x$ only on $(0, 2]$.

Example 3.5.4

Mixing Taylor Series and Algebra

Let $f(x) = x^2 \sin x$. Compute a Taylor series for $f(x)$ centered at $x = 0$.

Solution

We could try to work out a pattern in the derivatives of f, but even evaluating at $x = 0$ the computations become intractable.

$$f'(x) = 2x \sin x + x^2 \cos x$$

$$f''(x) = 2 \sin x + 4x \cos x - x^2 \sin x$$

$$f'''(x) = 6 \cos x - 6x \sin x - x^2 \cos x$$

$$f^{(4)}(x) = -12 \sin x - 8x \cos x + x^2 \sin x$$

Instead we can write the Taylor series for $\sin x$. Our earlier work gave us an expression for the Taylor polynomials and showed that their error goes to 0 as the degree goes to infinity.

$$\sin x = \sum_{k=0}^{\infty} \frac{(-1)^k)}{(2k+1)!} x^{2k+1}$$

We can obtain an expression for $x^2 \sin x$ by multiplying both sides by x^2. Since we're only multiplying by a power of x, the resulting series will still be a power series centered at 0.

$$\begin{aligned} x^2 \sin x &= x^2 \sum_{k=0}^{\infty} \frac{(-1)^k)}{(2k+1)!} x^{2k+1} \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k)}{(2k+1)!} x^{2k+3} \end{aligned}$$

Main Idea

When constructing a Taylor series for $f(x) = x^k g(x)$ centered at 0, construct the Taylor series of $g(x)$, and then distribute the x^k.

Example 3.5.5

Integrating a Taylor Series

Let $f(x) = e^{x^2}$.

a Write a Taylor polynomial $T_4(x)$ for $f(x)$ at $x = 0$.

b Find a better way to produce the Taylor series for $f(x)$.

c Compute a Taylor series for $\displaystyle\int e^{x^2}\, dx$.

Example 3.5.5 Integrating a Taylor Series

Solution

a We will compute the first four derivatives of $f(x)$. We will need the chain rule and later the product rule.

$$
\begin{aligned}
f(x) &= e^{x^2} & f(0) &= 1 \\
f'(x) &= 2xe^{x^2} & f'(0) &= 0 \\
f''(x) &= 2e^{x^2} + 4x^2e^{x^2} & f''(0) &= 2 \\
f'''(x) &= 12xe^{x^2} + 8x^3e^{x^2} & f'''(0) &= 0 \\
f^{(4)}(x) &= 12e^{x^2} + 48x^2e^{x^2} + 16x^4e^{x^2} & f^{(4)}(0) &= 12
\end{aligned}
$$

We can plug these values into our $T_4(x)$ formula.

$$
\begin{aligned}
T_4(x) &= \frac{1}{0!}x^0 + \frac{0}{1!}x^1 + \frac{2}{2!}x^2 + \frac{0}{3!}x^3 + \frac{12}{4!}x^4 \\
&= 1 + x^2 + \frac{1}{2}x^4
\end{aligned}
$$

We can see that our derivative calculations would quickly get out of hand as we take higher order derivatives. Even if there is a discernible pattern, it might take more computation to determine it.

b A better approach is to start with a simpler Taylor series that we know.

$$e^x = \sum_{k=0}^{\infty} \frac{1}{k!}x^k$$

e^{x^2} is a composition of e^x and x^2, so we will plug in x^2 for x in our e^x Taylor series.

$$
\begin{aligned}
e^{x^2} &= \sum_{k=0}^{\infty} \frac{1}{k!}(x^2)^k \\
&= \sum_{k=0}^{\infty} \frac{1}{k!}x^{2k}
\end{aligned}
$$

c Taylor series are also power saeries. By our theorem on power series, we can integrate term by term.

$$\int e^{x^2}\, dx = \sum_{k=0}^{\infty} = \frac{1}{k!(2k+1)}x^{2k+1} + c$$

Note that $\int e^{x^2}\, dx$ is not a function we can express algebraically or compute. A Taylor series gives us some way to represent this function, but we shouldn't be too satisfied. If we actually wanted to evaluate it, the best we could do is approximate it with a partial sum.

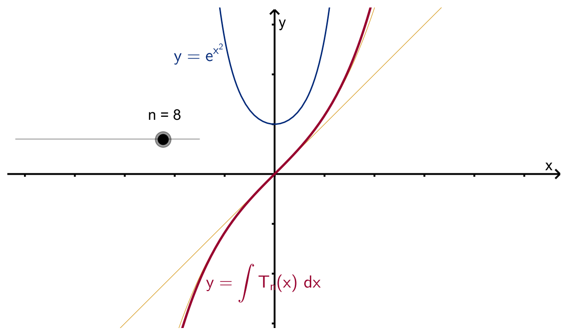


Figure: The graph of e^{x^2}, $\int e^{x^2}\,dx$, and the partial sums of its Taylor series.

Main Ideas

- Compositions of functions can be composed through Taylor series.
- Taylor series allow us to integrate functions that are otherwise impossible to integrate.

Application 3.5.6
Euler's Formula

Recall i is an imaginary number that satisfies $i^2 = -1$.

a Find an expression for $f(x) = e^{ix}$.

b Write your answer in terms of the Taylor series for $\sin x$ and $\cos x$.

c Write two different expressions for e^{i2x}. How is this equation useful?

Application 3.5.6 Euler's Formula

Solution

a We can express e^{ix} by replacing x by ix in the Taylor series for e^x.

$$T(x) = \sum_{k=0}^{\infty} \frac{1}{k!}(ix)^k$$

b To make much sense of this, we should try to simplify i^k.

$$\begin{aligned} i^0 &= 1 & i^4 &= 1 \\ i^1 &= i & i^5 &= i \\ i^2 &= -1 & &\vdots \\ i^3 &= -i \end{aligned}$$

We can write out the terms of $T(x)$ as follows:

$$T(x) = 1 + ix - \frac{1}{2}x^2 - \frac{1}{3!}ix^3 + \frac{1}{4!}x^4 + \frac{1}{5!}ix^5 - \frac{1}{6!}x^6 - \frac{1}{7!}ix^7$$

The terms with a factor of i are the Taylor series for $\sin x$ multiplied by i. The terms without a factor of i are the Taylor series for $\cos x$. We can write

$$e^{ix} = \cos x + i \sin x$$

c One way to write this would be to substitute $2x$ for x:

$$e^{i2x} = \cos 2x + i \sin 2x$$

Another way would be to square our original formula.

$$\begin{aligned} e^{i2x} &= \left(e^{ix}\right)^2 \\ &= (\cos x + i \sin x)^2 \\ &= \cos^2 x + 2i \cos x \sin x - \sin^2 x && (i^2 = -1) \end{aligned}$$

Setting these equal to each other, we note that for two complex numbers to be equal, their real parts must be equal and their imaginary parts must be equal.

$$\cos 2x + i \sin 2x = \cos^2 x + 2i \cos x \sin x - \sin^2 x$$

$$\begin{aligned} \cos 2x &= \cos^2 x - \sin^2 x \\ \text{and } \sin 2x &= 2 \cos x \sin x \end{aligned}$$

These are the double angle formulas for sine and cosine.

We can take higher powers of e^{ix} to produce triple or quadruple angle formulas. This converts a difficult geometry problem into something a high school algebra student could compute.

Remark

You would expect a relationship like this to be very famous, and it is. $e^{ix} = \cos x + i \sin x$ is called **Euler's Formula**. In addition to trigonometric formulas, it gives us insight into the complex numbers. This connection between an exponential and a periodic function is so powerful that it is used in such concrete applications as electrical engineering and signal processing.

Section 3.5

Exercises

Summary Questions

Q1 How can we be sure that a Taylor series converges to the function it is approximating?

Q2 What is the domain of a Taylor series?

Q3 How can we produce the Taylor series for $x^n f(x)$ or $f(x^n)$? Where does the center need to be for the result to be a Taylor series?

Q4 What is a Maclaurin series?

3.5.1

Q5 If we wanted to compute a decimal approximation of $\ln(1.25)$ by hand, would the Taylor polynomial or the Taylor series be more useful?

Q6 If $T(x)$ is a Taylor series centered at $x = a$, what are the possible forms that the domain of $T(x)$ could take?

3.5.2

Q7 How would the Taylor series of $f(x) = e^x$ change if we centered it at $x = 1$ instead of $x = 0$?

Q8 Let $T(x)$ be the Taylor series of $f(x) = e^x$ centered at 0. Verify that $T'(x) = T(x)$.

Q9 Write a Taylor series of $f(x) = \frac{1}{x}$ centered at 4.

Q10 Write a Taylor series of $f(x) = \frac{1}{x^2}$ centered at -5.

Q11 Write a Taylor series of $f(x) = \cos x$ centered at 0.

Q12 Write a Taylor series of $f(x) = \sin x$ centered at 0.

3.5.3

Q13 Show that the Taylor series of $f(x) = e^x$ centered at $x = 0$ is equal to $f(x)$ for all real numbers x.

Q14 Show that the Taylor series of $f(x) = \sin x$ centered at $x = 0$ is equal to $f(x)$ for all real numbers x.

Q15 Show that the Taylor series of $f(x) = \frac{1}{x}$ centered at 4 is equal to $f(x)$ for all x in the interval $(2, 6)$.

Q16 Suppose for a function f we are able to place a bound of $\frac{3^k}{k!}$ on the kth derivative of f over any interval. For what values of x can we conclude that $T(x)$, the Taylor series centered at 2, is equal to $f(x)$?

Q17 We didn't have a series test to determine whether $\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k}$ converges. How does our analysis of the Taylor series of $\ln x$ allows us to conclude that this series converges? Hint: what is $T(2)$?

Q18 For a general function f and its Taylor polynomials and series, how are the following sets of points related? Does every number belonging to one of these sets belong to one of the others?

- The set of numbers x where $T(x)$ converges.
- The set of numbers x where $|R_n(x)| \to 0$ as $n \to \infty$.
- The set of numbers where $f(x) = T(x)$.

3.5.4

Q19 Write a Taylor series for $f(x) = x^5 \cos x$ centered at $x = 0$.

Q20 Write a Taylor series for $f(x) = x^3 e^x$ centered at $x = 0$.

Q21 Can we use our Taylor series for $f(x) = \ln x$ centered at 1 to write a Taylor series for $g(x) = x^2 \ln x$? Explain.

Q22 Write a Taylor series for $f(x) = \frac{(x+5)^3}{x^2}$ centered at -5.

3.5.5

Q23 Let $g(x)$ be an antiderivative of e^{x^3}. Write the Taylor series for $g(x)$ centered at $x = 0$.

Q24 Let $g(x)$ be an antiderivative of $\cos(x^2)$. Write the Taylor series for $g(x)$ centered at $x = 0$.

Q25 Let $f(x) = \cos x$. Let $T(x)$ be the Taylor series of f centered at $x = 0$. Compute $T''(x)$. Why does your answer make sense?

Q26 Write the Taylor series for $f(x) = \frac{1}{x}$ centered at 1. Verify that one of its antiderivatives is a Taylor series for $\ln x$.

3.5.6

Q27 Rewrite our formula for $\cos(2x)$ to be entirely in terms of $\cos x$.

Q28 Use Euler's formula to compute a formula for $\cos 3x$ in terms of $\cos x$ and $\sin x$.

Q29 According to Euler's formula, what is the value of $e^{2\pi i}$?

Q30 Use the Taylor series of $\ln x$ centered at $x = 1$ to compute $\ln(1 + i)$. Do you think this series converges?

Synthesis & Extension

Q31 Let $h(x) = \frac{1}{x^2}$.

a Compute the Taylor polynomial T_3 centered at $x = 4$.

b If you wanted to use your Taylor polynomial from **a** to approximate $\frac{1}{2.5^2}$, what bound would Taylor's inequality put on the error? Don't simplify the arithmetic.

c What does the ratio test tell you about the domain of the Taylor series of $h(x)$ centered at $x = 4$?

Q32 Let X be a normal random variable with mean 0 and standard deviation 1. Write a series whose value is $P(0 \leq X \leq 1)$.

Q33 Suppose we produce the Taylor series $T(x)$ for some $f(x)$ centered at $x = 10$.

a If the Taylor series converges at $x = 5$, must it also converge at $x = 7$? Explain.

b If the errors of the Taylor polynomials $T_n(2)$ converge 0 as n goes to ∞ for some x, must $T(2)$ converge? If $T(2)$ converges, must the errors converge to 0?

c If you wanted to approximate $f(7)$ as accurately as possible, which would be more useful, a Taylor polynomial or a Taylor series?

Q34 Suppose we have a function $f(x)$ and two different numbers a and b. Suppose further that the Taylor series for $f(x)$ centered at a is equal to the Taylor series for $f(x)$ centered at b. What can you say about the domain of this Taylor series?

Chapter 4

Multivariable Functions

This chapter introduces functions of more than one variable. We construct the higher dimensional spaces needed for their domains, we produce tools to visualize them, and we compute their rates of change.

Contents

Section 4.1

Three-Dimensional Coordinate Systems

Goals:

1. Plot points in a three-dimensional coordinate system.
2. Use the distance formula.
3. Recognize the equation of a sphere and find its radius and center.
4. Graph an implicit function with a free variable.

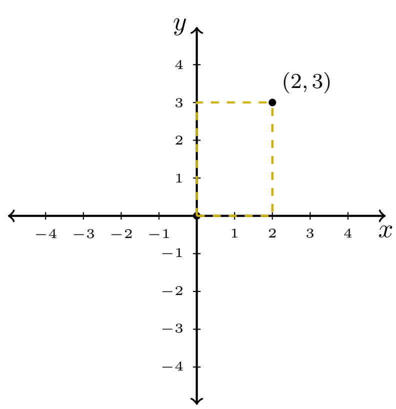
Suppose we wanted to understand the growth rate of a species of bacteria. We could grow several dozen cultures and take a series of measurements of size s at times t of each. Each measurement is an ordered pair (t, s). We can plot these pairs in a coordinate plane to get a visual sense of how growth occurs over time. We might even fit a function that approximates s as a function of t. What if we wanted to understand the role of some other measurement, like temperature, light, or the availability of various food sources? We could grow many cultures in different conditions. Now a single measurement has three or more pieces of information. While we could strip these out and plot our data on a temperature/size coordinate plane, we risk missing important relationships with the other variables. In order to take advantage of the visual and computational benefits of a coordinate system, we must be prepared to work with a coordinate system of more than two variables.

Question 4.1.1

How Do Cartesian Coordinates Extend to Higher Dimensions?

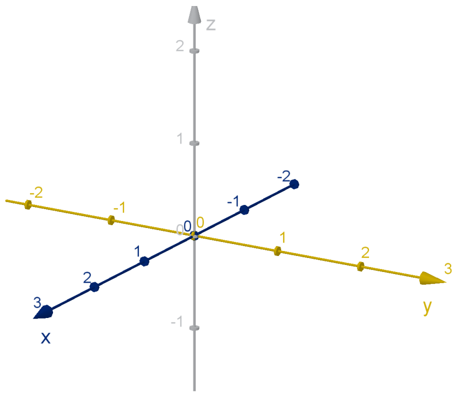
The best way to define a higher-dimensional coordinate system is to extrapolate from the coordinate plane. This way we don't need to remember a set of novel and arbitrary rules, and our two-dimensional experience will be a guide to us in dimensions where we have no visual intuition.

Recall how we constructed the Cartesian plane.



1. Assign origin and two directions (x, y).
2. y is 90 degrees anticlockwise from x.
3. Axes consist of the points displaced in only one direction.
4. Coordinates refer to displacement from the origin in each direction.
5. Either displacement can happen first.
6. Each point has exactly one ordered pair that refers to it.

In a three-dimensional Cartesian coordinate system. We can extrapolate from two dimensions.



1. Assign origin and three directions (x, y, z).
2. Each axis makes a 90 degree angle with the other two.
3. The z direction is determined by the right-hand rule.

Question 4.1.2
How Do We Establish Which Direction Is Positive in Each Axis?

The choice of which direction is positive is arbitrary. However, it is important that we all make the same choice, or our visualizations will be incompatible. In two dimensions, we agree that the positive y-axis is counterclockwise from the positive x-axis. This will not work in three dimensions. Suppose the positive y-axis is counterclockwise from the positive x-axis in three-space. If you rotate your point of view to see the axes from the other side, the positive y-axis is now clockwise from the positive x-axis. Thus the relative orientation of the positive x and y directions does not matter. You could pick a different orientation, and just be looking at three-space from a different viewpoint.

The z direction is different. Once we've chosen a positive x and y direction, there are two equally valid possible directions for positive z, pointing in opposite directions from each other. The choice here matters, but it will be arbitrary. We agree to define the positive z direction by the **right hand rule**.

The right hand rule says that if you make the fingers of your right hand follow the (counterclockwise) unit circle in the xy-plane, then your thumb indicates the direction of the positive z-axis.

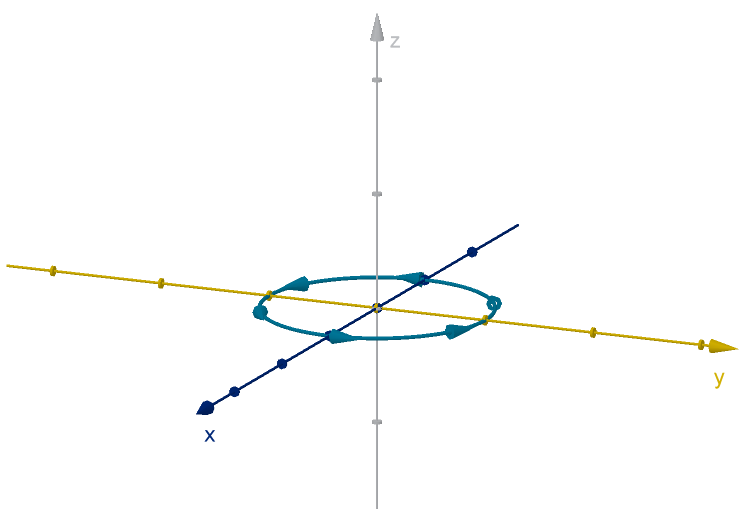


Figure: The counterclockwise unit circle in the xy-plane

Example 4.1.3

Drawing a Location in Three-Dimensional Coordinates

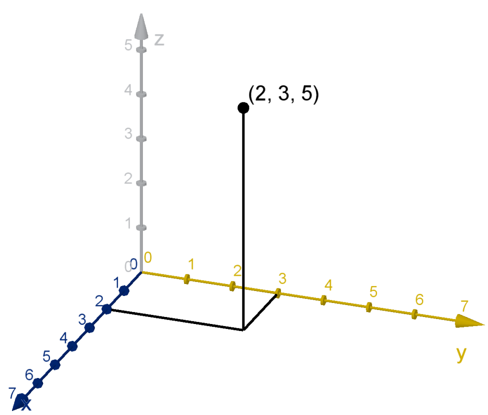
The point $(2, 3, 5)$ is the point displaced from the origin by

- 2 in the x direction
- 3 in the y direction
- 5 in the z direction.

How do we draw a reasonable diagram of where this point lies?

Solution

We can begin by finding the points $(2, 0, 0)$ which lies on the x-axis two units from the origin and $(0, 3, 0)$ which lies on the y-axis three units from the origin. Along with the origin itself, these points and $(2, 3, 0)$ form a parallelogram. Now we need a displacement of 5 in the z direction. We can copy the length and direction of this displacementof the segment from $(0, 0, 0)$ to $(0, 0, 5)$ on the z-axis. We draw a segment of that length and direction from $(2, 3, 0)$. The top of this segment is $(2, 3, 5)$.



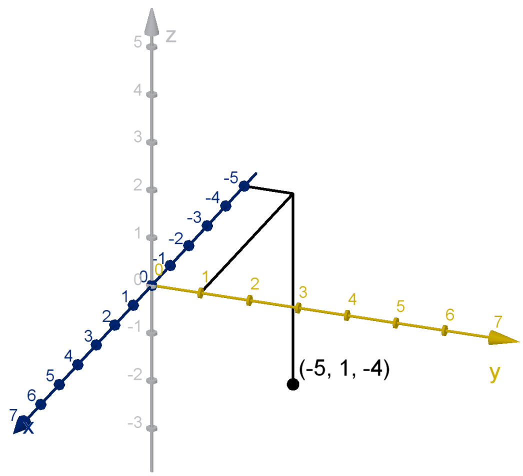
Remark

The extra lines we used to construct $(2, 3, 5)$ are not just useful for guaranteeing accuracy, they also help our audience to correctly visualize the location we mean to plot. When we project three-space onto a flat page, each point on the page represents infinitely many points stretching into the background. If we only draw a isolated point, which of these are we representing? Lines like the ones we produced in this example trick a viewers brain into visualizing correct three-dimensional location in our flat diagram.

How can we draw a reasonable diagram of $(-5, 1, -4)$?

Solution

The procedure here is the same, except that the displacements in the x and z directions are negative. Thus when producing these displacements, we travel backward along their axes.



Question 4.1.4

How Do We Measure Distance in Three-Space?

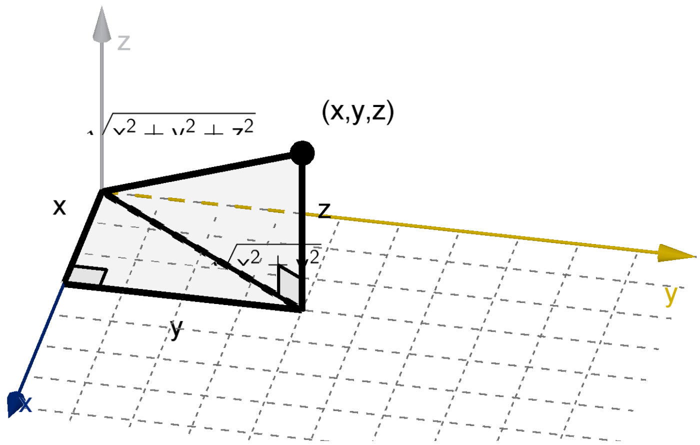
Since coordinate displacements in two-space are perpendicular, we compute the distance to a point using the Pythagorean theorem. This reasoning extends to higher dimensions, but we need to build the correct length using two or more right triangles.

Theorem

The distance from the origin to the point (x, y, z) is given by the Pythagorean Theorem

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

Question 4.1.4 How Do We Measure Distance in Three-Space?



We first compute the distance from the origin to $(x, y, 0)$ using a right triangle in the xy-plane. The right triangle with the vertices $(0, 0, 0)$, $(x, y, 0)$ and (x, y, z) allows us to apply the Pythagorean theorem again.

$$D^2 = \left(\sqrt{x^2 + y^2}\right)^2 + z^2$$

If neither of the points is the origin, we can compute the displacements by subtraction. This is a natural extension of the two-space distance formula.

Theorem

The distance from the point (x_1, y_1, z_1) to the point (x_2, y_2, z_2) is given by

$$D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

Question 4.1.5 What Is a Graph?

A well-prepared calculus student has learned to understand the graphs of many equations: lines, circles, parabolas. The definition of a graph, on the other hand, is often discarded after a few exercises of plotting points by hand. The definition is worth recalling. It applies to a space of any dimension.

Definition

The **graph** of an implicit equation is the set of points whose coordinates satisfy that equation. In other words, the two sides are equal when we plug the coordinates in for x, y and z.

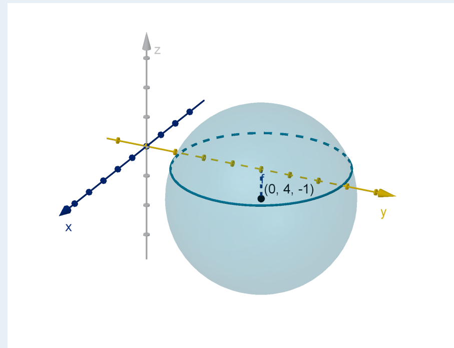
This definition allows us to immediately understand the graphs of some equations. The graph of the following equation consists of the points that, when plugged into a specific distance formula and squared, give a result of 9. This is a sphere.

Example

The graph of

$$x^2 + (y-4)^2 + (z+1)^2 = 9$$

is the set of points that are distance 3 from the point $(0, 4, -1)$



Example 4.1.6

Graphing an Equation with Two Free Variables

Sketch the graph of the equation $y = 3$.

Solution

The naive approach would have us seek out the point marked with 3 on the y-axis. However, in two-space, we know that the graph would be a horizontal line, not just the point $(0, 3)$. Why is this? Any point of the form $(x, 3)$ satisfies the equation $y = 3$. Similarly, any point of the form $(x, 3, z)$ in three-space satisfies $y = 3$. These are all the points that can be reached from $(0, 3, 0)$ by displacements in the x and z directions. They create a plane through $(0, 3, 0)$ parallel to the x and z axes.

Much as lines are the simplest and most fundamental one-dimensional objects, planes are the simplest and most fundamental two-dimensional objects. In addition to coordinate axes, 3-dimensional space has 3 **coordinate planes**.

1. The graph of $z = 0$ is the xy-plane.
2. The graph of $x = 0$ is the yz-plane.
3. The graph of $y = 0$ is the xz-plane.

Example 4.1.6 Graphing an Equation with Two Free Variables

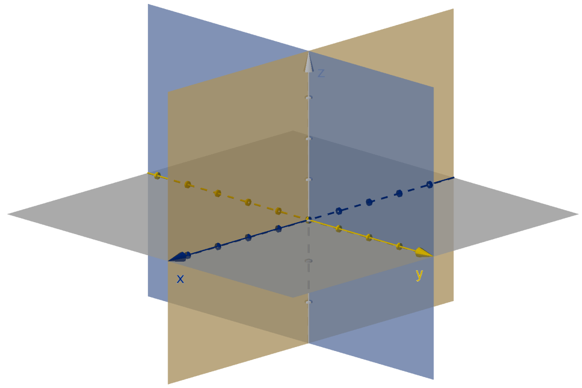


Figure: The coordinate planes in 3-dimensional space.

Remark

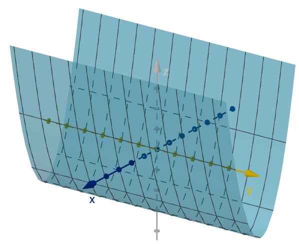
Planes extend forever but our pictures of them cannot. Notice that graphing software cuts them off parallel to the axes they contain. The resulting images are parallelograms. This is a good practice when drawing planes by hand too. It suggests the proper orientation to the viewer, despite the limitations of a flat visualization.

Example 4.1.7 Graphing an Equation with One Free Variable

Sketch the graph of the equation $z = x^2 - 3$.

Solution

We should recognize this as the equation of a parabola. If we ignore the variable y, we can graph this equation in the xz plane. What does the absence of absence of y in the equation mean? If we follow the definition of a graph, the value of y has no effect on whether a point lies on the graph or not. We can take the parabola in the xz plane, and project it in the y direction to obtain a surface called a **parabolic cylinder**.



$z = x^2 - 3$

Question 4.1.8
What Do the Graphs of Implicit Equations Look Like Generally?

Notice that the graph of an implicit equation in the plane is generally one-dimensional (a curve), whereas the graph of an implicit equation in three-space is generally two-dimensional (a surface).

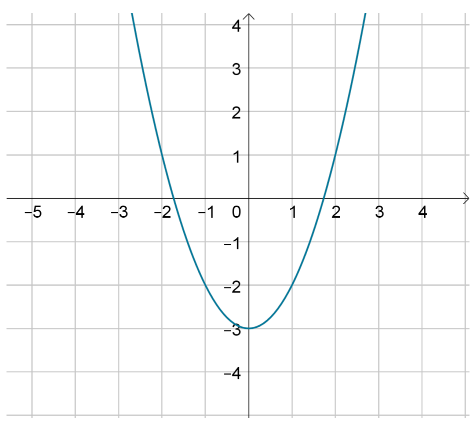


Figure: The curve $y = x^2 - 3$

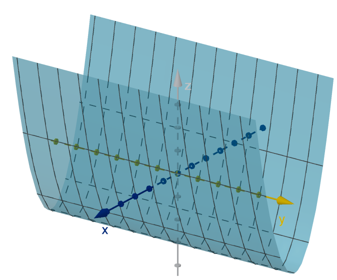


Figure: The surface $z = x^2 - 3$

Question 4.1.9
What Is the Slope-Intercept Equation of a Plane?

Unlike a line, a non-vertical plane has two slopes. One measures rise over run in the x-direction, the other in the y-direction.

Question 4.1.9 What Is the Slope-Intercept Equation of a Plane?

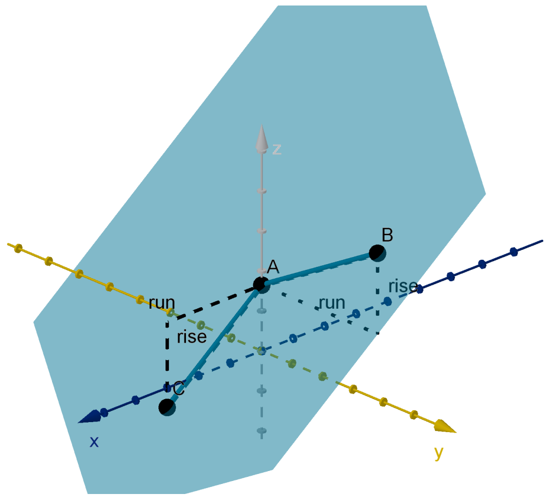


Figure: A plane with slopes in the x and y directions.

Equation

A plane with z intercept $(0,0,b)$ and slopes m_x and m_y in the x and y directions has equation

$$z = m_x x + m_y y + b.$$

Example 4.1.10
Writing the Equation of a Plane

Write the equation of a plane with intercepts $(4,0,0)$, $(0,6,0)$ and $(0,0,8)$.

Solution

From the point $(4,0,0)$ to the point $(0,0,8)$, the plane rises by 8 while x is reduced by 4. This gives a slope in the x direction.

$$m_x = \frac{8-0}{0-4} = -2.$$

Similarly,

$$m_y = \frac{8-0}{0-6} = -\frac{4}{3}.$$

The point $(0,0,8)$ is on the z-axis, and so indicates that the z-intercept is 8. Combining these, we conclude the plane has equation:

$$z = -2x - \frac{4}{3}y + 8$$

Main Idea

Given three points in a plane $A = (x_1, y_1, z_1)$, $B = (x_2, y_2, z_2)$ and $C = (x_3, y_3, z_3)$

1. If two points share an x-coordinate, we can directly compute m_y and vice versa.
2. Failing that, we can set up a system of equations and solve for m_x, m_y and b.

Question 4.1.11

How Do We Extrapolate to Even Higher Dimensions?

The measurements we take of each observation, the more dimensions we need to plot the data we have produced. Extrapolating from three-space to even higher dimensions introduces no new difficulties, except that we cannot visualize the result. We can use a coordinate system to describe a space with more than 3 dimensions. k-dimensional space can be defined as the set of points of the form

$$P = (x_1, x_2, \ldots, x_k).$$

Theorem

The distance from the origin to $P = (x_1, x_2, \ldots, x_k)$ in k-space is

$$\sqrt{x_1^2 + x_2^2 + \cdots + x_k^2}$$

There is no right hand rule for higher dimensions, because we can't draw these spaces anyway.

Section 4.1
Exercises

Summary Questions

Q1 What displacements are represented by the notation (a, b, c)?

Q2 What is the right hand rule and what does it tell you about a three-dimensional coordinate system?

Q3 In three-space, what is the y-axis? What are the coordinates of a general point on it?

Q4 In three space, what is the xz-plane? What are the coordinates of a general point on it? What is its equation?

Q5 How do we use a free variable to sketch a graph?

Q6 How do we recognize the equation of a sphere?

4.1.1

Q7 Suppose that instead of denoting each point $P = (x, y)$ in $\mathbb{R}^2$ by its displacements from the origin in the x- and y-directions, we denote it by $P = (d, m)$ where d is its distance from the origin, and m is the slope of the line through P and the origin. What problems could arise from adopting this convention?

Q8 Suppose the x and y axes were not perpendicular. Could we still assign coordinates to each point by its x and y displacements from the origin? Demonstrate with a diagram.

4.1.2

Q9 Which of the following depictions of the xy-plane are consistent with the usual orientation, and which are backwards?

- **a** The positive x axis points up, and the positive y-axis points left.
- **b** The positive x axis points down, and the negative y-axis points right.
- **c** The positive x axis points left, and the positive y-axis points up.
- **d** The negative x axis points right, and the positive y-axis points down.
- **e** The positive x axis points up and to the right, and the positive y-axis points down and to the right.

Q10 Suppose we draw the xy plane on our paper in the standard way, and our paper is lying on a table. Does the z-axis point down into the table or up out of the table?

4.1.3

Q11 Draw diagrams of points with the following coordinates.

- **a** $(6, 1, 2)$
- **b** $(-3, 0, 0)$
- **c** $(2, -1, 4)$
- **d** $(0, 3, 5)$

Q12 Draw diagrams of points with the following coordinates.

- **a** $(-4, 0, 0)$
- **b** $(3, -2, 0)$

c $(4, 5, -3)$

d $(-1, 3, 4)$

4.1.4

Q13 Compute the distance between $(3, 6, 2)$ and $(7, 3, -10)$.

Q14 Compute the distance between $(0, 3, 2)$ and $(5, 1, 0)$.

Q15 Compute the distance between $(10, 12, 109)$ and $(11, 9, 105)$.

Q16 Compute the distance between $(53, 42, 9)$ and $(43, 78, 2)$.

4.1.5

Q17 Does the point $(4, 3, 8)$ lie on the graph of $z = x^2 - 2$? Explain how you know.

Q18 Does $(2, 2, 1)$ lie on the graph of $x^2 + y^2 + z^2 = 9$? Explain how you know.

Q19 What is the graph of $y^2 + z^2 = -1$? Explain your reasoning.

Q20 The point $(2, 3, 4)$ lies on the graph $ax + ay - z = 26$. What is the value of the number a?

Q21 Olivia says that the graph of $(x-2)(y-3) = 0$ in the xy-plane is the point $(2, 3)$. Do you agree? How would you explain it?

Q22 How is the graph of $f(x, y, z)g(x, y, z) = 0$ related to the graphs of $f(x, y, z) = 0$ and $g(x, y, z) = 0$?

4.1.6

Q23 Does the graph of $z = 4$ intersect the graph of $z = 6$? Explain both using geometry and algebra.

Q24 Does the graph of $x = 2$ intersect the graph of $z = 1$? Explain.

4.1.7

Q25 Sketch the graph of each equation.

- **a** $x = -4$
- **b** $x^2 + y^2 = 9$
- **c** $x^2 + 4x + y^2 + z^2 - 2z = 4$

Q26 Sketch a graph of $z = -2$.

Q27 Sketch a graph of $y = -z^2$.

Q28 Sketch a graph of $x^2 + z^2 = 25$.

4.1.8

Q29 What dimension you we expect the graph of an equation to be in 6-dimensional space?

Q30 What is the graph of $x^2 + y^2 = 0$ in the xy-plane? Is this an exception to our intuition about the dimension of a graph?

Q31 Zoe and Muhammad both sketch the graph of $y = x^2$. Zoe's graph is a curve. Muhammad's is a surface. Has one of them drawn the wrong graph? Explain.

Q32 In $\mathbb{R}^3$, what is the dimension of the intersection of the graphs $x^2 + y^2 = 25$ and $z = 1$? Can you explain this in terms of our intuition about the dimension of a graph.

4.1.9

Q33 Suppose that y is a free variable in the equation of a plane. What does that tell us about m_x and m_y?

Q34 Gabby is trying to find the equation of a plane P, but she doesn't know any points on the xz-plane or yz-plane. Instead she knows that P contains the points:

$$A = (1, 3, 6) \qquad B = (5, 3, 4) \qquad C = (7, 5, 10)$$

Using points A and B, she decides that $m_x = \frac{4-6}{5-1} = -\frac{1}{2}$. Using points A and C, she decides that $m_y = \frac{10-6}{5-3} = 2$.

a Which of Gabby's conclusions do you agree with and which do you disagree with? Why?

b How could you fix the one that is wrong?

Q35 Supoose you intend to write the equation of the plane through A, B and C in slope-intercept form. If $A = (3, 5, 7)$ and $B = (3, 2, 4)$, what value(s) of the y coordinate of C would make it easiest to compute m_x?

Q36 Recall that we can write the equation of a line in $\mathbb{R}^2$ in **point-slope** form:

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

where m is the slope and (x_0, y_0) is a known point. This was especially useful in single-variable calculus for writing equations of tangent lines.

a How would you expect to write the equation of the plane P through $(2, 4, -6)$ with slopes $m_x = \frac{1}{2}$ and $m_y = -3$?

b Does your answer to **a** actually pass through $(2, 4, -6)$? How do you know?

c Is your answer to **a** actually the equation of a plane? How do you know? Does it have the correct slopes?

d Write a general expression for point-slope form for a plane.

Q37 The plane P has slopes $m_x = 3$ and $m_y = -1$ and passes through $(2, 5, -1)$.

a Write the equation of P is point-slope form.

b What is the z-intercept of P.

Q38 Given a plane with $m_x = 5$ and $m_y = 2$, we can conclude that the plane is steeper in the x-direction than the y-direction. Is the x-direction the steepest direction we could travel in? If not, what is?

4.1.10

Q39 Write the equation of a plane through $(3, 0, 0)$, $(0, 7, 0)$, and $(0, 0, -1)$.

Q40 Write the equation of a plane with intercepts $(2, 0, 0)$, $(0, -2, 0)$, and $(0, 0, 4)$.

Q41 Write the equation of a plane through $(6, 4, 1)$, $(6, 7, -2)$, and $(8, 7, 1)$.

Q42 Write the equation of a plane through $(2, 2, 1)$, $(4, 2, 9)$, and $(2, 0, 0)$.

Q43 Write the equation of a plane through $(3, 4, 2)$, $(5, 5, 6)$, and $(7, 4, 6)$.

Q44 Write the equation of a plane through $(1, 5, 2)$, $(11, 5, 4)$, and $(6, 3, -3)$.

4.1.11

Q45 Assuming you could draw in 4 dimensions, describe how you might construct the graph of $x_1^2 + x_3^2 + x_4^2 = 25$ in $\mathbb{R}^4$.

Q46 Assuming you could draw in 4 dimensions, describe how you might construct the graph of $x_2 = x_3^2$ in $\mathbb{R}^4$.

Q47 What equation(s) would describe the x_2x_4-plane in $\mathbb{R}^4$?

Q48 What would you call the object in $\mathbb{R}^4$ defined by $x_1 = 0$?

Extension and Synthesis

Q49 The points $(1, 0, 3)$ and $(1, 4, 0)$ are both on the sphere S. What are the possible values for the radius of S?

Q50 The graph of $x^2 + y^2 = 0$ in $\mathbb{R}^2$ is a point, not a curve. Use this idea to write an equation for the intersection of the graphs $f(x, y, z) = c$ and $g(x, y, z) = d$. What do you expect the dimension of this intersection to be?

Q51 Suppose the x and y axes in $\mathbb{R}^2$ were not perpendicular. Would the distance formula still hold? Demonstrate.

Section 4.2

Functions of Several Variables

Goals:

1. Convert an implicit function to an explicit function.
2. Calculate the domain of a multivariable function.
3. Calculate **level curves** and **cross sections**.

If we want to understand the relationship between variables, a function is the gold standard. For example, when we can write y as a function of x, then at each value of x, we simply need plug in the value and simplify the arithmetic. There is no chance that algebraic manipulation will lead us to multiple values of y, or to an equation we cannot solve. Naturally, we want to understand this type of relationship between more than two variables. Much like our investigation of n-space, we'll begin by adding one variable. After this initial step, extrapolating to more variables will be straightforward.

Question 4.2.1

What Is a Function of More than One Variable?

Definition

A function of two variables is a rule that assigns a number (the **output**) to each ordered pair of real numbers (x, y) in its **domain**. The output is denoted $f(x, y)$.

Some functions can be defined algebraically. If $f(x, y) = \sqrt{36 - 4x^2 - y^2}$ then

$$f(1, 4) = \sqrt{36 - 4 \cdot 1^2 - 4^2} = 4.$$

Example 4.2.2

The Domain of a Function

Identify the domain of $f(x, y) = \sqrt{36 - 4x^2 - y^2}$.

Example 4.2.2 The Domain of a Function

Solution

The only obstacle to evaluating this function is that the value under the square root might be negative. We can write an inequality to express this and solve.

$$\begin{aligned} 36 - 4x^2 - y^2 &\geq 0 \\ 36 &\geq 4x^2 + y^2 \\ 1 &\geq \frac{x^2}{9} + \frac{y^2}{36} \end{aligned}$$

These are the points inside an ellipse whose intercepts are $(\pm 3, 0)$ and $(0, \pm 6)$.

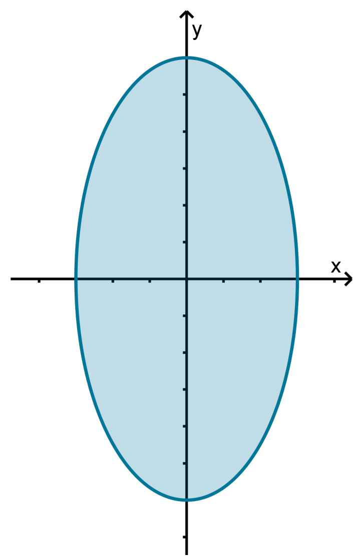


Figure: The domain of a function

Main Idea

When solving for the domain of an algebraic function, we look for the same obstacles to evaluating the function that we do for one-variable functions.

- Expressions in a denominator cannot be 0 (including built-in fractions like $\tan x = \frac{\sin x}{\cos x}$)
- Expressions in a square root must be greater than or equal to 0.
- Expressions in a logarithm must be greater than 0.

The conditions these produce with a two-variable function may be harder to visualize or simplify than with a function of one variable.

Application 4.2.3
Temperature Maps

Many useful functions cannot be defined algebraically. There is a function $T(x, y)$ which gives the temperature at each latitude and longitude (x, y) on earth. No pair (x, y) has more than one temperature, and no pair fails to have a temperature. Still there is no hope of producing an expression that computes T for any x and y. Mathematically (though perhaps not meteorologically) this function is arbitrary.

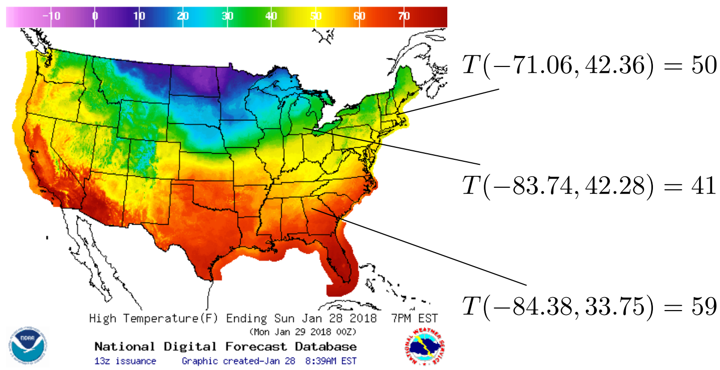


Figure: A temperature map

This function is represented graphically by using color to portray the value of T at each point.

Application 4.2.4
Digital Images

A digital image is made up of pixels, each with a different color. In many modern images, these pixels are too small to see. The color of each pixel is a function of that pixel's location. Since colors are harder to define numerically, we can consider the simpler case: where each pixel is a different shade of gray. In this case we have a brightness function $B(x, y)$ where the output is a number that represents the brightness of the pixel at the coordinates (x, y).

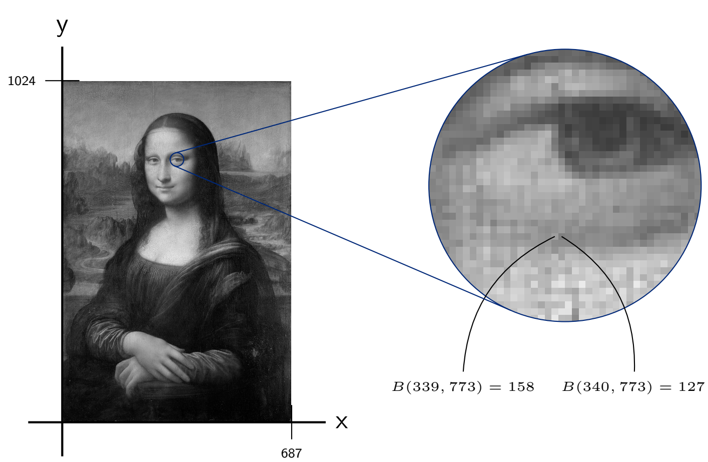


Figure: An image represented as a brightness function B on each pixel

Application 4.2.4 Digital Images

Remark

The brightness function differs from other functions we've studied in one key way. It is only defined for (x, y) where x and y are integers. Other examples can take any real numbers as coordinates. This makes our usual calculus methods impossible. We cannot get arbitrarily close to a point in order to compute a limit. All other points are at least 1 unit away. However, if we are willing to settle for approximations, we can apply calculus and get useful results.

Question 4.2.5 What Is the Graph of a Two-Variable Function?

A graph is our most important way to visualize a function. The graph of a one variable functions is an object in two-space. One dimension measures the input variable. The other measures the output. For a two variable function, the graph lies in three-space.

Definition

The graph of a function $f(x, y)$ is the set of all points (x, y, z) that satisfy

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

The height z above a point (x, y) represents the value of the function at (x, y). In this figure, $f(1, 4)$ is equal to the height of the graph above $(1, 4, 0)$.

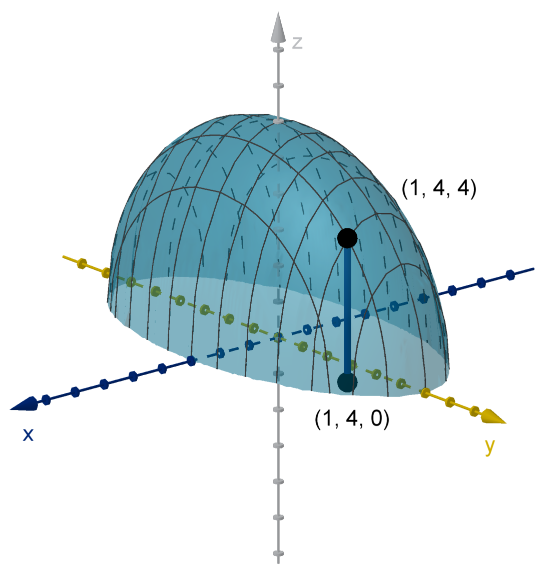


Figure: The graph $z = \sqrt{36 - 4x^2 - y^2}$

Question 4.2.6

How Do We Visualize a Graph in Three-Space?

Three-space is harder to visualize than two-space. What's more, plotting points is more arduous with two dimensions of inputs. In the absence of computer graphics, mathematicians have used a variety of visualization tools.

Definition

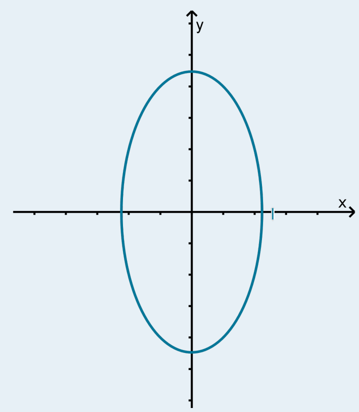
A **level set** of a function $f(x, y)$ is the graph of the equation $f(x, y) = c$ for some constant c. For a function of two variables this graph lies in the xy-plane and is called a **level curve**.

Example

Consider the function

$$f(x, y) = \sqrt{36 - 4x^2 - y^2}.$$

The level curve $\sqrt{36 - 4x^2 - y^2} = 4$ simplifies to $4x^2 + y^2 = 20$. This is an ellipse.
Other level curves have the form $\sqrt{36 - 4x^2 - y^2} = c$ or $4x^2+y^2 = 36 - c^2$. These are larger or smaller ellipses.



Level curves take their shape from the intersection of $z = f(x, y)$ and $z = c$. Seeing many level curves at once can help us visualize the shape of the graph.

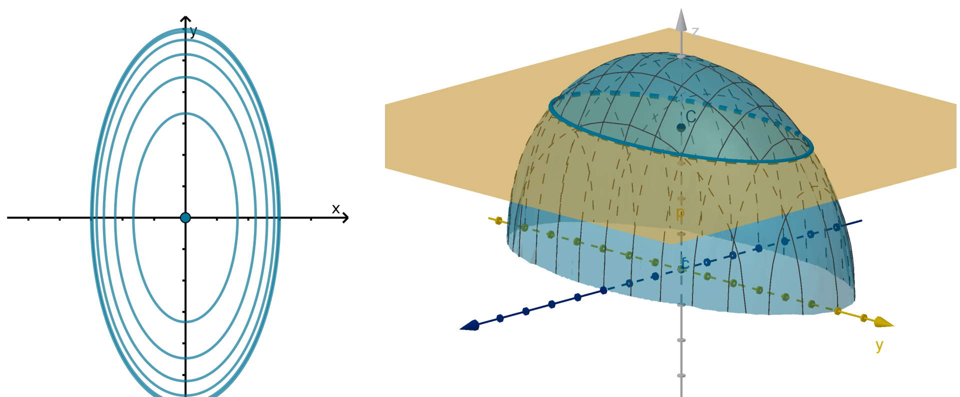


Figure: The graph $z = f(x, y)$, the planes $z = c$, and the level curves

Example 4.2.7

Drawing Level Curves

Where are the level curves on this temperature map?

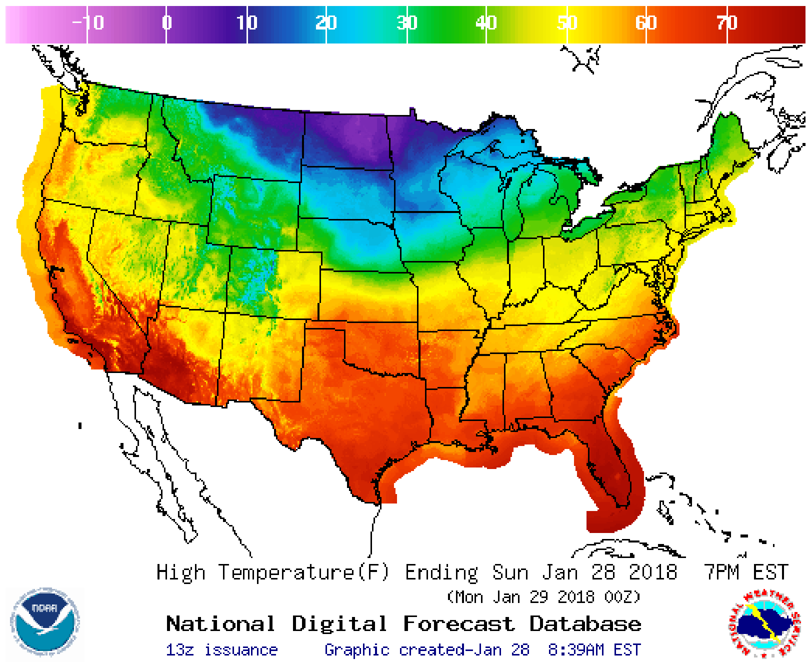
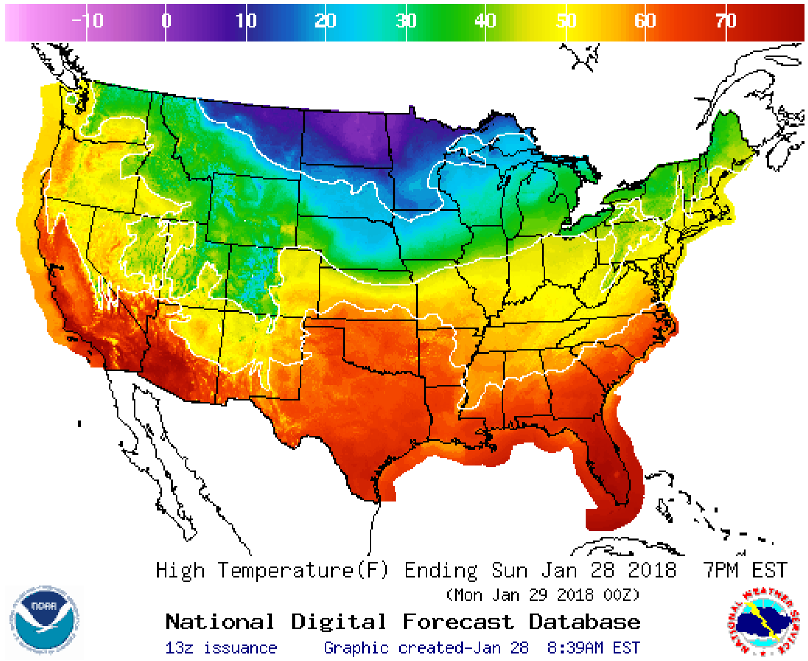


Figure: A temperature map

Solution

The level sets are the points where the temperature has a certain value. Since the colors represent ranges of temperatures, it's difficult to pick out the level sets within that range. However, at the transition from one color to the next, we know that the temperature is equal to the cutoff temperature between those ranges. The picture below shows a reasonable attempt to sketch three level curves in white. Notice that the level curves (especially the one between green and yellow) are not connected, and that drawing them in perfect detail is beyond the ability of a human.



Example 4.2.8
Using Level Curves to Describe a Graph

What features can we discern from the level curves of this topographical map?

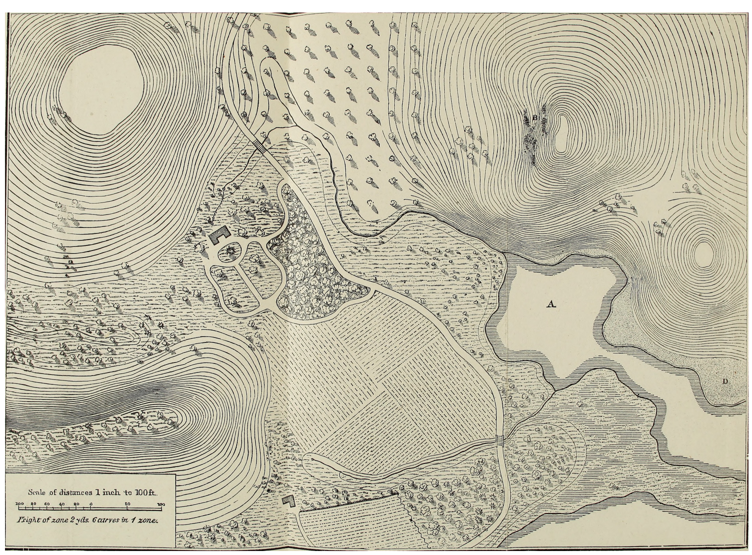
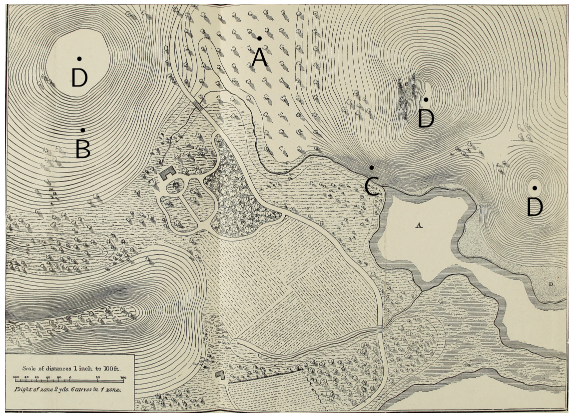


Figure: A topographical map

Example 4.2.8 Using Level Curves to Describe a Graph

Solution



There are many features we could describe. Here is a sample.

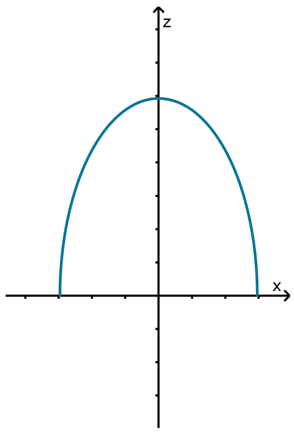
- The point A is surrounded by relatively flat terrain. There are not many level curves here, which means the altitude is not increasing or decreasing to higher or lower levels.
- The points B and C are on slopes. If we travel north and south we cross level curves, meaning our altitude is increasing or decreasing. The slope is steeper at B than at C, because traveling north from B we cross more level curves than traveling north from C
- The points marked D are in the middle of a series of rings of level curves. These are either the tops of hills or (less likely given the context) the bottoms of valleys.

Example 4.2.9 A Cross Section

Definition

The intersection of a plane with a graph is a **cross section**. A level curve is a type of cross section, but not all cross sections are level curves.

Find the cross section of $z = \sqrt{36 - 4x^2 - y^2}$ at the plane $y = 1$.



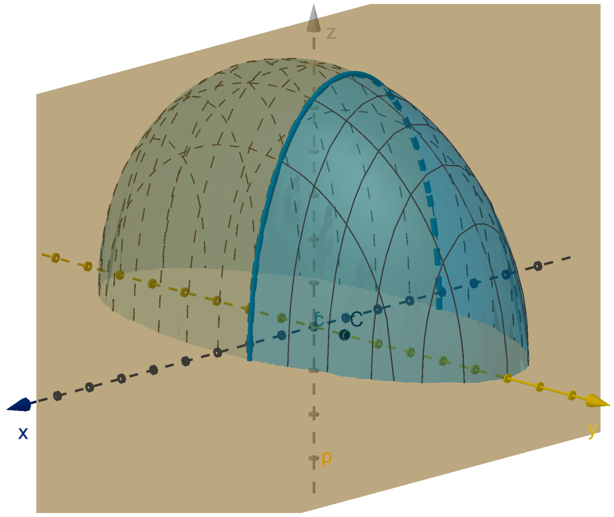


Figure: The $y = 1$ cross section of $z = \sqrt{36 - 4x^2 - y^2}$

Example 4.2.10

Converting an Implicit Equation to a Function

Definition

We sometimes call an equation in x, y and z an **implicit equation**. Often in order to graph these, we convert them to **explicit functions** of the form $z = f(x, y)$

Write the equation of a paraboloid $x^2 - y + z^2 = 0$ as one or more explicit functions so it can be graphed. Then find the level curves.

Example 4.2.10 Converting an Implicit Equation to a Function

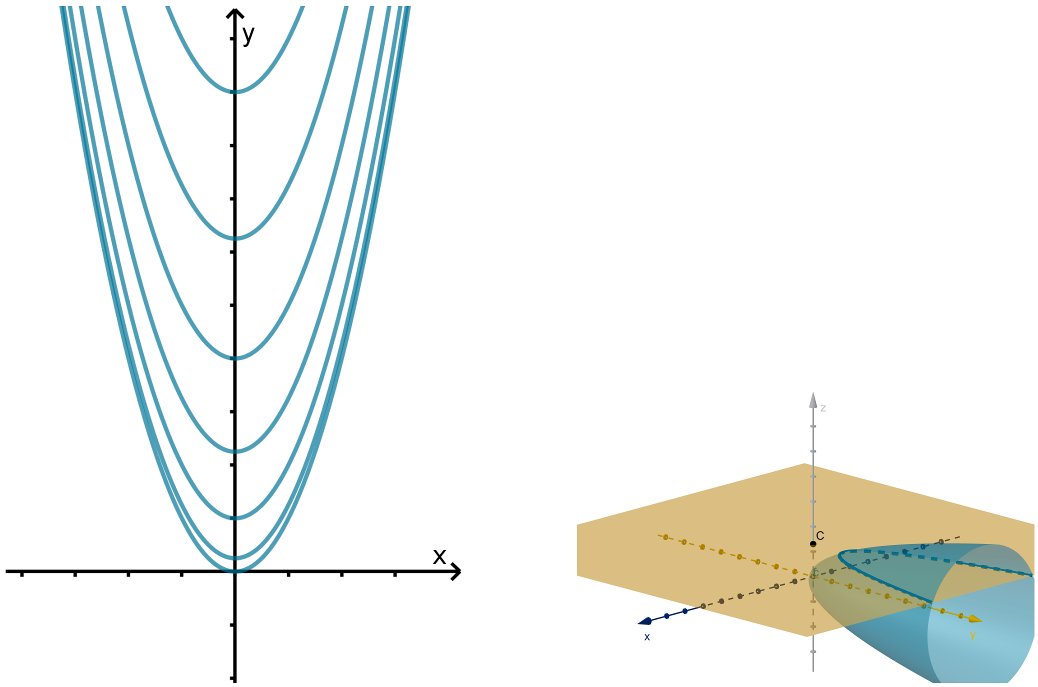


Figure: Level curves of $x^2 - y + z^2 = 0$

Question 4.2.11 How Does this Apply to Functions of More Variables?

We can define functions of three variables as well. Denoting them $f(x, y, z)$. For even more variables, we use x_1 through x_n. The definitions of this section can be extrapolated as follows.

Variables	2	3	n
Function	$f(x, y)$	$f(x, y, z)$	$f(x_1, \ldots, x_n)$
Domain	subset of $\mathbb{R}^2$	subset of $\mathbb{R}^3$	subset of $\mathbb{R}^n$
Graph	$z = f(x, y)$ in $\mathbb{R}^3$	$w = f(x, y, z)$ in $\mathbb{R}^4$	$x_{n+1} = f(x_1, \ldots, x_n)$ in $\mathbb{R}^{n+1}$
Level Sets	level curve in $\mathbb{R}^2$	**level surface** in $\mathbb{R}^3$	level set in $\mathbb{R}^n$

Observation

We might hope to solve an implicit equation of n variables to obtain an explicit function of $n-1$ variables. However, we can also treat it as a level set of an explicit function of n variables (whose graph lives in $n+1$ dimensional space).

$$x^2 + y^2 + z^2 = 25 \longrightarrow f(x,y) = \pm\sqrt{25 - x^2 - y^2}$$

$$F(x,y,z) = x^2 + y^2 + z^2$$
$$F(x,y,z) = 25$$

Both viewpoints will be useful in the future.

Section 4.2 Exercises

Summary Questions

Q1 What does the height of the graph $z = f(x,y)$ represent?

Q2 What is the distinction between a level set and a cross section?

Q3 What are level sets in $\mathbb{R}^2$ and $\mathbb{R}^3$ called?

Q4 What is the difference between an implicit equation and explicit function?

4.2.1

Q5 If $f(x,y) = 13x + \frac{y}{x}$, compute $f(2,-8)$.

Q6 If $f(x,y) = \cos(\pi xy)$, compute $f\left(4, \frac{1}{3}\right)$.

Q7 Is $f(x,y) = \pm\sqrt{4x-y}$ a function? Explain.

Q8 Is the following a function? Explain.

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

4.2.2

Q9 Compute the domain of $f(x,y) = \frac{1}{x+y}$.

Q10 What is the domain of $f(x,y) = \frac{1}{x^2+y^2}$?

Q11 What is the domain of $g(x,y) = x^3 + \sqrt{y^2 - 25}$?

Q12 What is the domain of $g(x,y) = 15 + \ln(y - 2x)$?

Q13 What is the domain of $f(x,y) = \frac{\sqrt{x+3}}{y^2-x}$?

Q14 Compute the domain of $h(x,y) = \frac{4x}{y - \ln x}$

4.2.3

Q15 On the temperature map, we saw $T(-84.38, 33.75) = 59$. Is $T(-84.38, 35.75)$ greater than or less than 59?

Q16 On the temperature map, we saw $T(-83.74, 42.28) = 41$. Is $T(-93.74, 42.28)$ greater than or less than 41?

Q17 What range of temperatures are found in South Dakota? In which parts of the state are the extreme temperatures found?

Q18 Can you use this diagram to approximate $T(-61.06, 42.36)$? Explain.

4.2.4

Q19 In our image of Mona Lisa, what is the domain of B?

Q20 In our blow-up of the digital image, we see Mona Lisa's eye is near the coordiante $(369, 800)$. Where is her other eye?

4.2.5

Q21 Can the points $(1, 3, 5)$ and $(1, 3, 7)$ both be on the graph of $z = f(x, y)$? Explain.

Q22 If the graph $z = f(x, y)$ is below the xy-plane, what does that tell us about $f(x, y)$?

Q23 If $f(x, y)$ has a z-intercept of c, what does that tell us about f?

Q24 What is the significance of the points where the graph $z = f(x, y)$ intersects the xy-plane?

4.2.6

Q25 Describe the level curves of $f(x, y) = (x-2)^2 + (y+1)^2$.

Q26 Describe the level curves of $f(x, y) = x^2 - 3y + 5$.

Q27 Describe the level curves of $\frac{x^2}{y}$.

Q28 Describe the level curves of $g(x, y) = \frac{y}{e^x}$.

Q29 Give the equation of the level curve of $f(x, y) = x^3 + y^3$ that passes through $(4, 2)$.

Q30 Give the equation of the level curve of $g(x, y) = 17x^2 - 3xy + y^3$ that contains the point $(1, 2)$.

Q31 Given a function $f(x, y)$, how many level curves might pass through $(3, 7)$?

Q32 If the points (x_1, y_1) and (x_2, y_2) lie on the same level curve of $h(x, y)$, what are the possible values of the expression $h((x_1, y_1) - h(x_2, y_2)$?

4.2.7

Q33 In our level curves on the temperature map, what physical meaning can we take from the fact that the green-yellow and red-orange level curves are closer together in Kansas than they are farther east?

Q34 Explain why it makes sense physically that level curves of a temperature function would be complicated and disconnected.

4.2.8

Q35 In the topographical map, what can we deduce from the fact that no level curves cross the farm fields in the lower center of the map?

Q36 Explain why it makes physical sense that there are level curves alongside the creeks in this map.

4.2.9

Q37 Give an equation for the $y = 2$ cross-section of the graph $z = f(x, y)$ where $f(x, y) = x^3 + y^3$.

Q38 Consider the plane P whose equation is $f(x, y) = 3x - 5y + 7$.

i. Give the equation of the $y = 0$ cross section of P. What is this graph? What is the significance of the various parts of its equation?
ii. Give the equation of the $x = 0$ cross section of P. What is the significance of the various parts of its equation?
iii. Give the equation and describe the set of all level curves of f.

Q39 If the cross sections of $z = f(x, y)$ in the planes $y = b$ are identical for all values of b, what does that tell us about f?

Q40 If $f(x, y)$ is a function that satisfies $f(x, y) = f(x, -y)$ for all x and y, how will this be refelected in the cross sections of $z = f(x, y)$?

4.2.10

Q41 Rewrite $y = x^2 + z^2$ as one or more explicit functions $z = f(x, y)$.

Q42 Rewrite $\ln x + \ln y + \ln z = 0$ as one or more explicit functions $z = f(x, y)$.

Q43 Rewrite $x^2 + y^2 + z^2 + xyz = 20$ as one or more explicit functions $z = f(x, y)$.

Q44 Explain why it would be difficult to write $\frac{\ln y}{z} - \sqrt{xz} = 5 + x$ as an explicit function of the form $z = f(x, z)$. Choose a better dependent and variable and write that variable as a function of the other two.

4.2.11

Q45 Consider the function $f(x_1, x_2, x_3, x_4, x_5)$.

- **a** What space does the graph of f lie in?
- **b** What space does a level set of f lie in?

Q46 Write $xyz = 1$ as

- **a** A level set of a function
- **b** An explicit function $z = f(x, y)$

Q47 Consider a one-variable function $f(x)$.

- **a** What space does the graph of $f(x)$ lie in?
- **b** Where does a level set of f lie in? What does a typical level set look like?

Q48 Show how the graph of an explicit function $x_{n+1} = f(x_1, x_2, \ldots, x_n)$ can be converted to the level set of an $n + 1$-variable function.

Synthesis & Extension

Q49 Let $f(x, y) = x^2$. Sketch the graph of $z = f(x, y)$. What is the role of y in this graph?

Q50 Consider the implicit equation $zx = y$

- **a** Rewrite this equation as an explicit function $z = f(x, y)$.
- **b** What is the domain of f?
- **c** Solve for and sketch a few level sets of f.
- **d** What do the level sets tell you about the graph $z = f(x, y)$?

Q51 Consider the implicit equation: $x = \sin z$.

a Sketch a graph of the equation.

b Describe (in words) what the cross section of the graph in the $x = \frac{1}{2}$ plane looks like.

Section 4.3

Limits and Continuity

Goals:

1. Understand the definition of a **limit** of a multivariable function.
2. Use the Squeeze Theorem
3. Apply the definition of **continuity**.

Limits of multivariable functions are conceptually similar to one-variable functions. However, even though the requirement is the same, it is a much harder to satisfy. Since there are so many more ways to approach a given point in a higher dimensional space, there are more nearby points to check to see whether the function is actually approaching the proposed limit.

Question 4.3.1

What Is the Limit of a Function?

Definition

We write

$$\lim_{(x,y)\to(a,b)} f(x,y) = L$$

if we can make the values of f stay arbitrarily close to L by restricting to a sufficiently small neighborhood of (a,b).

Proving a limit exists requires a formula or rule. For any amount of closeness required (ϵ), you must be able to produce a radius δ around (a,b) sufficiently small to keep $|f(x,y) - L| < \epsilon$. For this reason, we will not prove that any limits exist. We will present three examples of functions whose limit does not exist.

Example 4.3.2

A Limit That Does Not Exist

Show that $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2-y^2}{x^2+y^2}$ does not exist.

Solution

Let's define $f(x,y) = \frac{x^2-y^2}{x^2+y^2}$. We will approach the point $(0,0)$ from two different directions. If we approach along the x-axis, then the points on our path have the form $(x,0)$. When we plug these into the function, the value is $f(x,0) = \frac{x^2-0}{x^2+0}$. This is equal to 1 for all values of x except 0, so as x approaches 0, the values of f are arbitrarily close (in fact exactly equal) to 1.

On the other hand, if we approach 1 along the y-axis, then the points have the form $(0,y)$. When we plug these into the function, the value is $f(0,y) = \frac{0-y^2}{0+y^2}$. This is equal to -1 for all values of y except 0, so as y approaches 0, the values of f are arbitrarily close (in fact exactly equal) to -1.

What does this say about the limit of f? The $\lim\limits_{(x,y)\to(0,0)} f(x,y) \neq 1$ because there are points on the y-axis do not give values close to 1, but any neighborhood of $(0,0)$ includes some points on the y-axis. Similarly, $\lim\limits_{(x,y)\to(0,0)} f(x,y) \neq -1$. If we tried to argue that the limit had any other value, the x-axis and y-axis would both present a problem. This this limit does not exist.

We can identify the problem behavior in the graph of $z = f(x,y)$. As the graph approaches the origin, there are points of all heights between -1 and 1. Specifically we can see the line above the x-axis and below the y-axis. No amount of closeness can exclude this range of values.

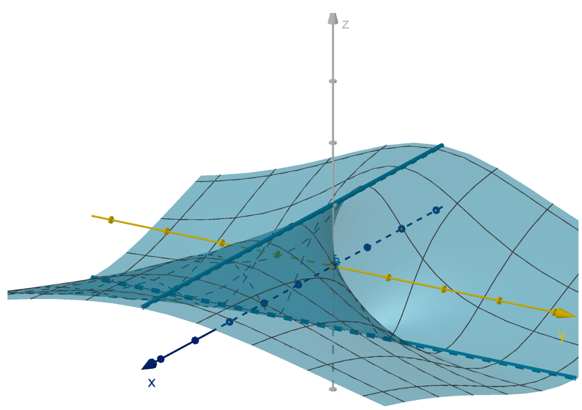


Figure: A function with no limit at $(0,0)$

We might take away the idea that checking limits of two-variable functions requires checking in both the x-direction and the y-direction. Unfortunately, even that is not sufficient.

Example 4.3.3

Another Limit That Does Not Exist

Show that $\lim\limits_{(x,y)\to(0,0)} \dfrac{xy}{x^2+y^2}$ does not exist.

Example 4.3.3 Another Limit That Does Not Exist

Solution

Let $f(x, y) = \frac{xy}{x^2+y^2}$. We can check the values of this function on the x- and y-axes. Except at $(0, 0)$, $f(x, 0) = 0$ and $f(0, y) = 0$. However, not all the points close to $(0, 0)$ lie on an axis. Suppose we work with the points on another line: $y = mx$. These points have the form (x, mx). We can evaluate f on this line.

$$\begin{aligned} f(x, xm) &= \frac{(x)(mx)}{x^2 + (mx)^2} \\ &= \frac{mx^2}{(m^2 + 1)x^2} \\ &= \frac{m}{m^2 + 1} && \text{(except at } (0, 0)\text{)} \end{aligned}$$

Thus there are point arbitrarily close to $(0, 0)$ on which f is valued as low as -0.5 $(m = -1)$ and as high as 0.5 $(m = 1)$. The limit does not exist.

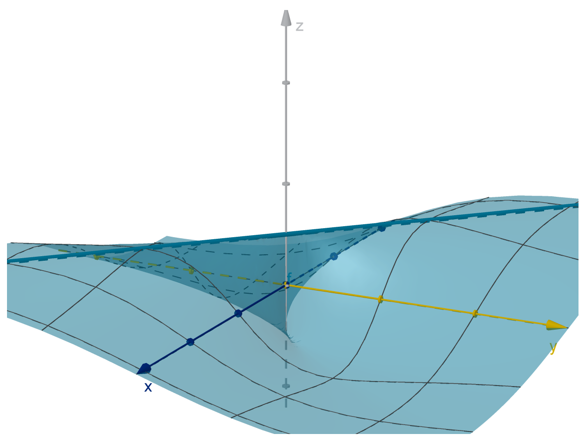


Figure: The graph $z = \frac{xy}{x^2+y^2}$ and the line of height $\frac{1}{2}$ over $x = y$.

We might take away the idea that checking limits of two-variable functions requires checking along each line through the point in question. Unfortunately, even that is not sufficient.

Example 4.3.4 Yet Another Limit That Does Not Exist

Show that $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy^2}{x^2 + y^4}$ does not exist.

Solution

Let $f(x, y) = \frac{xy^2}{x^2+y^4}$. We can check the values of this function on the x- and y-axes. Except at $(0, 0)$,

$f(x,0)=0$ and $f(0,y)=0$. We can also check the values along a line of the form $y=mx$.

$$\begin{aligned} f(x,xm) &= \frac{(x)(mx)^2}{x^2+(mx)^4} \\ &= \frac{m^2x^3}{x^2(1+m^4x^2)} \\ \lim_{x\to 0} f(x,xm) &= \lim_{x\to 0} \frac{m^2x^3}{x^2(1+m^4x^2)} \\ &= \lim_{x\to 0} \frac{m^2x}{1+m^4x^2} \\ &= 0 \end{aligned}$$

Thus along each line, the values of f approach 0 as we approach the origin. However, we have not considered paths that are not line. Consider the parabola $x=y^2$. Points on this parabola have the form (y^2,y). We compute the values on this parabola.

$$\begin{aligned} f(y^2,y) &= \frac{(y^2)(y)^2}{(y^2)^2+y^4} \\ &= \frac{y^4}{2y^4} \end{aligned}$$

For any point on this parabola except the origin f has a value of $\frac{1}{2}$. Thus f takes values of $\frac{1}{2}$ and 0 in any neighborhood of $(0,0)$, meaning the limit does not exist.

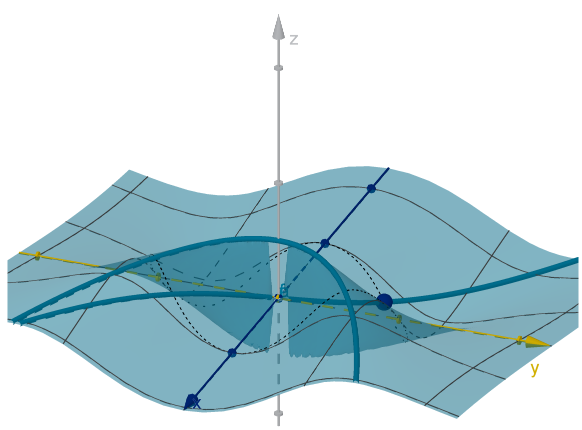


Figure: The graph $z=\frac{xy^2}{x^2+y^4}$, which limits to 0 along any line through the origin, but has height $\frac{1}{2}$ over the parabola $x=y^2$

We take away from these exercises that establishing the value of a multi-variable limit cannot be reduced to computing a single-variable limit, or even a family of single-variable limits. The formal arguments that establish multi-variable limits are more advanced and beyond the scope of this text.

Question 4.3.5

What Tools Apply to Multi-Variable Limits?

The limit laws from single-variable limits transfer comfortably to multi-variable functions.

1. Sum/Difference Rule
2. Constant Multiple Rule
3. Product/Quotient Rule

These rules allow us to compute limits of complicated functions from simpler ones. How do we come by those simpler limits in the first place? We can apply the kind of advanced arguments we alluded to earlier. Another tool is the squeeze theorem.

The Squeeze Theorem

If $g < f < h$ in some neighborhood of (a, b) and

$$\lim_{(x,y)\to(a,b)} g(x, y) = \lim_{(x,y)\to(a,b)} h(x, y) = L,$$

then

$$\lim_{(x,y)\to(a,b)} f(x, y) = L.$$

Question 4.3.6

What Is a Continuous Function?

Definition

We say $f(x, y)$ is **continuous** at (a, b) if

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b).$$

In a rigorous development of calculus, we compute limits and use them to show that functions are continuous. Given that evaluating limits is beyond our current means, we will reverse the process. Rather than worrying about how to prove the following theorem, we will assume it is true and use it to evaluate limits.

Theorem

- Polynomials, roots, trig functions, exponential functions and logarithms are continuous on their domains.
- Sums, differences, products, quotients and compositions of continuous functions are continuous on their domains.

The limit of a continuous function is equal to the value of the function. When we need to compute a limit of these functions, we'll just evaluate them instead. Why didn't this work in our examples? In each of our examples, the function was a quotient of polynomials, but $(0,0)$ was not in the domain.

Remark

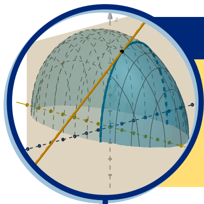
Limits, continuity and these theorems can all be extrapolated to functions of more variables.

Section 4.3 Exercises

Summary Questions

Q1 Why is it harder to verify a limit of a multivariable function?

Q2 What do you need to check in order to determine whether a function is continuous?



Section 4.4

Partial Derivatives

Goals:

1. Calculate **partial derivatives**.
2. Realize when not to calculate partial derivatives.

The first task in developing calculus is to understand rates of change. In the single-variable case, we ask how the dependent variable changes per unit of increase in the independent variable. With more than one independent variable we must ask: what kind of increase do we mean? There is more than one possible answer. Partial derivatives are the simplest and most intuitive rate of change.

Question 4.4.1

What Is the Rate of Change of a Multivariable Function?

Motivational Example

The force due to gravity between two objects depends on their masses and on the distance between them. Suppose at a distance of $8,000$km the force between two particular objects is 100 newtons and at a distance of $10,000$km, the force is 64 newtons.

How much do we expect the force between these objects to increase or decrease per kilometer of distance?

Solution

The change in force divided by the change in distance is

$$\frac{64\text{N} - 100\text{N}}{10,000\text{km} - 8,000\text{km}} = -0.018\frac{N}{km}$$

Notice that the change in force is entirely attributable to the change in distance. That is because the masses of the objects did not change. The only change in the dependent variables is the $2,000$km increase in distance.

Our goals in understanding multi-variable rates of change are guided by what we accomplished with one variable. Derivatives of a single-variable function were a way of measuring the change in a function. Recall the following facts about $f'(x)$.

1. Average rate of change is realized as the slope of a secant line:

$$\frac{f(x) - f(x_0)}{x - x_0}$$

2 The derivative $f'(x)$ is defined as a limit of slopes:

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

3 The derivative is the instantaneous rate of change of f at x.

4 The derivative $f'(x_0)$ is realized geometrically as the slope of the tangent line to $y = f(x)$ at x_0.

5 The equation of that tangent line can be written in point-slope form:

$$y - y_0 = f'(x_0)(x - x_0)$$

In the physics example above, the rate of change was easier to understand because only one independent variable is changing. That was an average rate of change, taken between two points. We now develop a corresponding instantaneous rate of change. A partial derivative measures the rate of change of a multivariable function as one variable changes, but the others remain constant.

Definition

The **partial derivatives** of a two-variable function $f(x, y)$ are the functions

$$f_x(x, y) = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h}$$

and

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

We can see the idea of each partial derivative in the formula. f_x compares the values of f at $(x+h, y)$ and (x, y). The x values change between these two points, but the y values remain constant. The opposite is true in the formula for f_y.

Notation

The partial derivative of a function can be denoted a variety of ways. Here are some equivalent notations

- f_x
- $\frac{\partial f}{\partial x}$
- $\frac{\partial z}{\partial x}$
- $\frac{\partial}{\partial x} f$
- $D_x f$

Example 4.4.2

Computing a Partial Derivative

Find $\frac{\partial}{\partial y}(y^2 - x^2 + 3x\sin y)$.

Main Idea

To compute a partial derivative f_y, perform single-variable differentiation. Treat y as the independent variable and x as a constant.

Solution

We take an ordinary derivative, treating y as the variable and x as a constant. The familiar rules of derivatives apply. The sum rule means we can differentiate term-by-term.

- $\frac{\partial}{\partial y} y^2 = 2y$
- $\frac{\partial}{\partial y} x^2 = 0$, since the x^2 term is treated as constant.
- $\frac{\partial}{\partial y} 3x\sin y = 3x\cos y$, since $3x$ is treated as constant multiple of the function $\sin y$.

Together this gives the partial derivative

$$\frac{\partial}{\partial y}(y^2 - x^2 + 3x\sin y) = 2y + 3x\cos y.$$

Synthesis 4.4.3

Interpreting Derivatives from Level Sets

Below are the level curves $f(x,y) = c$ for some values of c. Can we tell whether $f_x(-4, 1.25)$ and $f_y(-4, 1.25)$ are positive or negative?

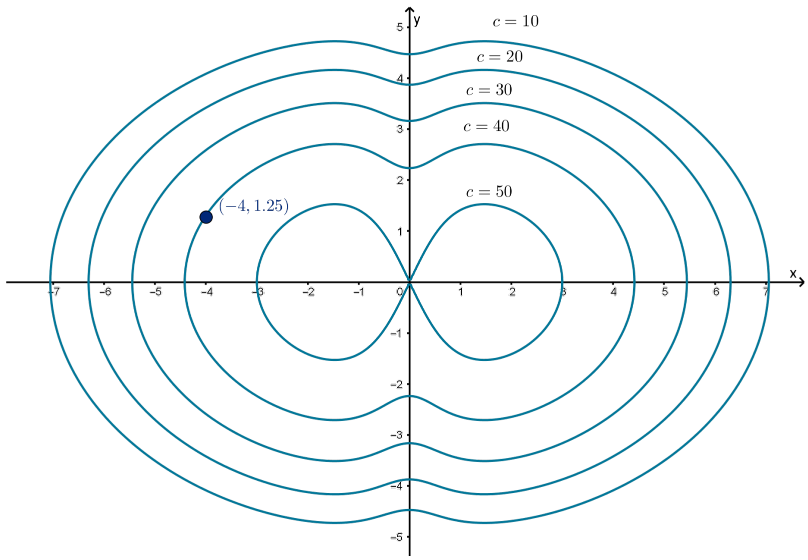


Figure: Some level curves of $f(x, y)$

Solution

As x increases and y remains constant, we travel to the right in the coordinate plane. Based on the labeling of the level curves, this takes f from the value 40 to values between 40 and 50, meaning f increases. Thus $f_x > 0$.

Similarly, as y increases and x remains constant, we travel upwards in the coordinate plane. This takes f from the value 40 to values between 30 and 40, meaning f decreases. Thus $f_y < 0$.

Question 4.4.4

What Is the Geometric Significance of a Partial Derivative?

The partial derivative $f_x(x_0, y_0)$ is realized geometrically as the slope of the line tangent to $z = f(x, y)$ at (x_0, y_0, z_0) and traveling in the x direction. Since y is held constant, this tangent line lives in $y = y_0$, a plane perpendicular to the y-axis. The line is tangent to the cross section of the graph with that plane.

Question 4.4.4 What Is the Geometric Significance of a Partial Derivative?

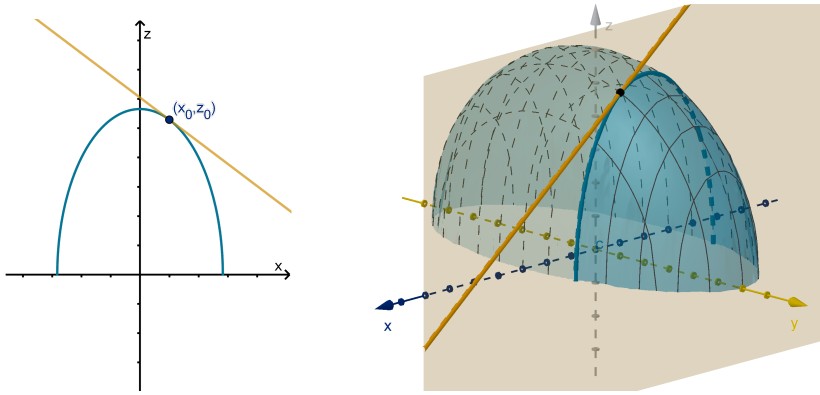


Figure: The tangent line to $z = f(x, y)$ in the x direction

Example 4.4.5
Derivative Rules and Partial Derivatives

Find f_x for the following functions $f(x, y)$:

a $f = \sqrt{xy}$ (on the domain $x > 0, y > 0$)

b $f = \frac{y}{x}$

c $f = \sqrt{x + y}$

d $f = \sin(xy)$

Solution

a We can rewrite this as $f(x, y) = \sqrt{x}\sqrt{y}$. In this setting, $\sqrt{y}$ is a constant multiple. Thus $f_x(x, y) = \frac{1}{2\sqrt{x}}\sqrt{y}$

b We can rewrite this as $f(x, y) = \frac{1}{x}y$. We treat y as a constant multiple. $f_x(x, y) = -\frac{1}{x^2}y$.

c We cannot rewrite this as $f(x, y)\sqrt{x} + \sqrt{y}$, because that is not a valid algebraic manipulation. Instead we use the chain rule.

- The outer function is $\sqrt{x}$. Its derivative is $\frac{1}{2\sqrt{x}}$.

- The inner function is $x+y$. Its derivative is 1.
- By the chain rule

$$\frac{\partial}{\partial x}\sqrt{x+y} = \frac{1}{2\sqrt{x+y}}(1) = \frac{1}{2\sqrt{x+y}}$$

d We do not have an easy trig rule to break up products. We'll use the chain rule again.

- The outer function is $\sin x$. Its derivative is $\cos x$.
- The inner function is xy. Its derivative is y.
- By the chain rule

$$\frac{\partial}{\partial x}\sin(xy) = \cos(xy)y$$

Main Idea

Sometimes we can detach the variable held constant from the changing variable using the rules of algebra. When we can't, we'll often need a differentiation rule (usually the chain rule).

Question 4.4.6
What If We Have More than Two Variables?

We can also calculate partial derivatives of functions of more variables. All variables but one are held to be constants. :

Example

If

$$f(x,y,z) = x^2 - xy + \cos(yz) - 5z^3,$$

then

$$\begin{aligned}\frac{\partial f}{\partial y} &= 0 - x - \sin(yz)z - 0 \\ &= -x - z\sin(yz)\end{aligned}$$

Example 4.4.7
A Function of Three Variables

For an ideal gas, we have the law $P = \frac{nRT}{V}$, where P is pressure, n is the number of moles of gas molecules, T is the temperature, and V is the volume.

a. Calculate $\frac{\partial P}{\partial V}$.

b. Calculate $\frac{\partial P}{\partial T}$.

c. (Science Question) Suppose we're heating a sealed gas contained in a glass container. Does $\frac{\partial P}{\partial T}$ tell us how quickly the pressure is increasing per degree of temperature increase?

Solution

a. We can write $P = nrT\frac{1}{V}$ and treat nrT as a constant multiple. Then $\frac{\partial P}{\partial V} = nrT\left(-\frac{1}{V^2}\right)$.

b. In this case, $nr\frac{1}{V}$ is a constant multiple. $\frac{\partial P}{\partial T} = nr\frac{1}{V}(1)$.

c. No. $\frac{\partial P}{\partial T}$ assumes n and V are constant, but glass expands as it heats. The volume of both the container and the gas is increasing, not constant.

Question 4.4.8
How Do Higher Order Derivatives Work?

Taking a partial derivative of a partial derivative gives us a higher order partial derivative. We use the following notation.

Notation

$$(f_x)_x = f_{xx} = \frac{\partial^2 f}{\partial x^2}$$

We need not use the same variable each time

Notation

$$(f_x)_y = f_{xy} = \frac{\partial}{\partial y}\frac{\partial}{\partial x} f = \frac{\partial^2 f}{\partial y \partial x}$$

Remark

Notice the subscript notation and the ∂ notation express higher order derivatives in opposite order. Subscripts are added to the right of f, which the differential operation is applied on the left of f.

Example 4.4.9

A Higher Order Partial Derivative

If $f(x, y) = \sin(3x + x^2 y)$ calculate f_{xy}.

Solution

First we compute f_x. We'll need the chain rule.

- The outer function is $\sin x$. Its derivative is $\cos x$.
- The inner function is $3x + x^2 y$. Its derivative is $3 + 2xy$.
- $f_x = \cos(3x + x^2 y)(3 + 2xy)$.

Computing $(f_x)_y$ will require the product rule. $\frac{\partial}{\partial y} \cos(3x + x^2 y)$ requires the chain rule.

- The outer function is $\cos x$. Its derivative is $-\sin x$.
- The inner function is $3x + x^2 y$. Its derivative is x^2.
- $\frac{\partial}{\partial y} \cos(3x + x^2 y) = -\sin(3x + x^2 y)(x^2)$.

Now we apply the product rule.

$$\begin{aligned}\frac{\partial}{\partial y} \cos(3x + x^2 y)(3 + 2xy) &= \frac{\partial}{\partial y}\left(\cos(3x + x^2 y)\right)(3 + 2xy) + \cos(3x + x^2 y)\frac{\partial}{\partial y}(3 + 2xy) \\ &= -\sin(3x + x^2 y)(x^2)(3 + 2xy) + \cos(3x + x^2 y)(2x)\end{aligned}$$

Question 4.4.10

Does Differentiation Order Matter?

No. Specifically, the following is due to Clairaut:

Theorem

If f is defined on a neighborhood of (a, b) and the functions f_{xy} and f_{yx} are both continuous on that neighborhood, then $f_{xy}(a, b) = f_{yx}(a, b)$.

This readily generalizes to larger numbers of variables, and higher order derivatives. For example $f_{xyyz} = f_{zyxy}$.

Section 4.4

Exercises

Summary Questions

Q1 What is the role of each variable when we compute a partial derivative?

Q2 What does the partial derivative $f_y(a, b)$ mean geometrically?

Q3 Can you think of an example where the partial derivative does not accurately model the change in a function?

Q4 What is Clairaut's Theorem?

4.4.1

Q5 Give the equation of the line that lies in the plane $x = 2$ and is tangent to the graph $z = xe^{3xy} + x$ at the point $(2, 0, 4)$. You may give your equation in any notation that works in 2 dimensions.

Q6 Alexander performs an experiment with his wireless networking router. At each level of power output (in miliwatts) and distance from his computer (in meters), he measures $T(p,d)$, the maximum transfer speed of data (in megabits per second). Here is a table of his observations.

	$0mW$	$100mW$	$200mW$	$300mW$	$400mW$	$500mW$	$600mW$	$700mW$
$10m$	6	15	40	100	300	800	800	800
$20m$	0	2	15	30	90	300	800	800
$30m$	0	0	2	10	50	100	400	800
$40m$	0	0	0	0	5	20	50	100
$50m$	0	0	0	0	2	5	20	45

a Use this data to approximate $T_p(300,20)$. Show what values you used. There is more than one reasonable way to do this.

b What does the derivative in **a** mean in physical terms? Be precise and include units.

c Use this data to approximate $T_d(500,30)$. Show what values you used. There is more than one reasonable way to do this.

d What appears to be true about the sign of $T_d(p,d)$? What does this mean in physical terms, and why does it make sense?

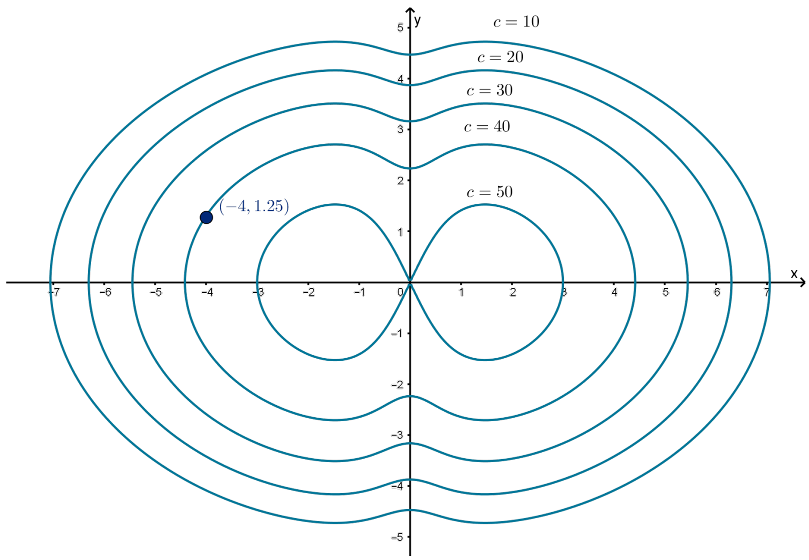
4.4.2

Q7 Let $f(x,y) = 7x^2 + 5y\cos x + e^y$. Compute $f_x(x,y)$. Explain the role of y in each term where it is present.

Q8 Let $f(x,y) = \sin x \sin y$. Show how to compute $f_y(x,y)$ using the product rule, then suggest a more efficient approach.

4.4.3

Q9 In the diagram from this example, is $f_x(3,0)$ positive or negative? Explain.



Q10 In the diagram from this example, use a point on the $c = 30$ level set to approximate $f_y(4, -1.25)$.

Q11 In the diagram from this example, use a point on the $c = 50$ level set to approximate $f_x(4, -1.25)$.

Q12 In the diagram from this example, what is $f_y(0,0)$? Explain your reasoning.

4.4.4

Q13 Find f_x and f_y for the following functions $f(x,y)$

a $f(x,y) = x^2 - y^2$

b $f(x,y) = \sqrt{\frac{y}{x}}$ (assume $x > 0$ and $y > 0$)

c $f(x,y) = ye^{xy}$

Q14 Find $g_x(x,y)$ and $g_y(x,y)$ for the following functions $g(x,y)$

a $g(x,y) = e^{x^2+y^2}$

b $g(x,y) = y\ln(y-x)$

c $g(x,y) = \frac{3x^2+4x-2}{e^{(y^3)}}$

4.4.5

Q15 Extrapolate from the limit defintion of $f_x(x,y)$ to give a limit definition of $f_x(x,y,z)$. Explain why this limit represents a change in f where only x is changing.

Q16 Let $f(x,y,z) = e^{3x}y + \sqrt[3]{yz} + x^3z^7$. Compute $\frac{\partial f}{\partial z}$.

Q17 Let $g(u,v,w) = e^{uv+w^2}$. Compute $\frac{\partial g}{\partial v}$.

Q18 Let $p(r,s,t) = \frac{e^r+e^s+e^t}{rst}$. Compute $\frac{\partial p}{\partial r}$.

4.4.6

Q19 In this example, does the fact that glass expands as it is heated suggest that $\frac{\partial P}{\partial T}$ overstates or understates the actual rate of pressure increase as T increases?

Q20 Suppose Jinteki Corporation makes widgets which is sells for $\$100$ each. It commands a small enough portion of the market that its production level does not affect the demand (price) for its products. If W is the number of widgets produced and C is their operating cost, Jinteki's profit is modeled by

$$P = 100W - C.$$

Since $\frac{\partial P}{\partial W} = 100$ does this mean that increasing production can be expected to increase profit at a rate of $\$100$ per widget?

4.4.7

Q21 Suppose $g(s,t)$ is the partial derivative of $f(s,t)$ with respect to t, and $h(s,t)$ is the partial derivative for $g(s,t)$ with respect to s. Write h in terms of f using both subscript and ∂ notation.

Q22 Physicists note that velocity is the derivative of position with respect to time, and acceleration is the derivative of velocity with respect to time. If $s(t,f)$ is the position of a rocket with f kilograms of fuel after t seconds, what is the physical meaning of $\frac{\partial^3 s}{\partial^2 t \partial f}$?

4.4.8

Q23 If $f(x,y) = \sin(3x + x^2 y)$ calculate f_{yx}. Verify that you get the same answer that we did for f_{xy}.

Q24 Let $f(x,y) = \ln(x^2 + y)$. Compute $f_{xy}(x,y)$.

Q25 Let $g(x,y,z) = 2x^3 z + ye^{xy^2}$.

a Compute $\frac{\partial g}{\partial y}$.

b Compute $\frac{\partial^2 g}{\partial x^2}$.

Q26 Compute the following partial derivatives of

$$g(x,y,z) = \frac{x^3 \sin(xz)}{y}$$

a $\dfrac{\partial g}{\partial y}$

b $\dfrac{\partial^2 g}{\partial z^2}$

c $\dfrac{\partial^2 g}{\partial z \partial x}$

4.4.9

Q27 If $f(x, y, z)$ is a smooth function, which of the following are equavalent to f_{xyyzy}?

i. f_{xzzyz}
ii. f_{zyyxy}
iii. f_{yyyzx}
iv. f_{xxxyz}
v. f_{xyzy}
vi. f_{xyz}
vii. f_{yxxzx}

Q28 How many third partial derivatives does a two-variable function have? Assuming these derivatives are continuous, which of them are equal according to Clairaut's theorem?

Synthesis & Extension

Q29 Let $f(x, y) = \frac{e^{xy}}{x+y}$. Is $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y}$? If so, why? If not, how are they related?

Q30 The function $f(x, y) = e^{x+y}$ has the strange property that $f_x x, y = f_y(x, y)$ at every point (x, y). What does this mean geometrically about the function f?

Q31 Do we know that $f_x(x, y)$ is in fact a function? What fact about limits is relevant to this question?

Section 4.5

Linear Approximations

Goals:

1. Calculate the equation of a **tangent plane**.
2. Rewrite the tangent plane formula as a **linearization** or **differential**.
3. Use linearizations to estimate values of a function.
4. Use a differential to estimate the error in a calculation.

In single-variable calculus, the tangent line was one of the great applications of the derivative. It solves a difficult geometry problem, but it also gives a method of approximating a difficult to compute function. The height of the tangent line is close to the height of the graph near the point of tangency. This means the value of the tangent line function approximates the value of the function, close to the point of tangency. The two-variable analogue of a tangent line is a tangent plane.

Question 4.5.1

What Is a Tangent Plane?

Definition

A **tangent plane** at a point $P = (x_0, y_0, z_0)$ on a surface is a plane containing the tangent lines to the surface through P.

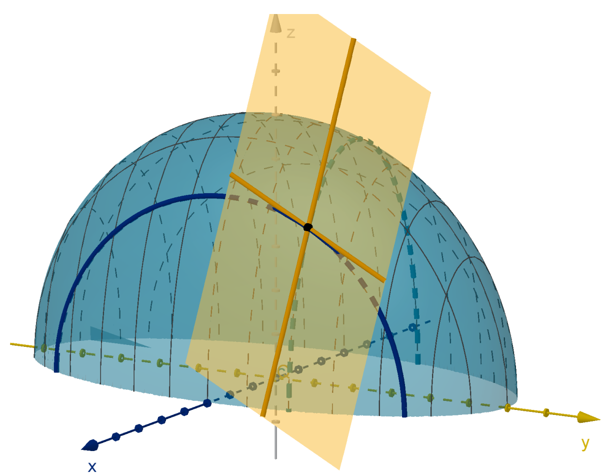


Figure: The tangent plane to $z = f(x, y)$ at a point

Equation

If the graph $z = f(x, y)$ has a tangent plane at (x_0, y_0), then it has the equation:

$$z - z_0 = f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0).$$

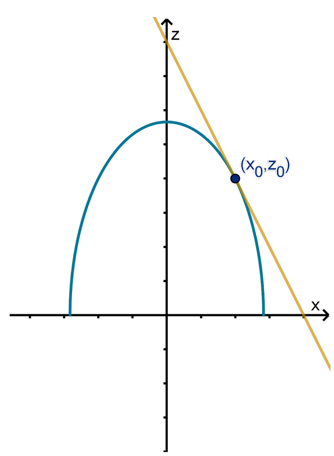
Remarks

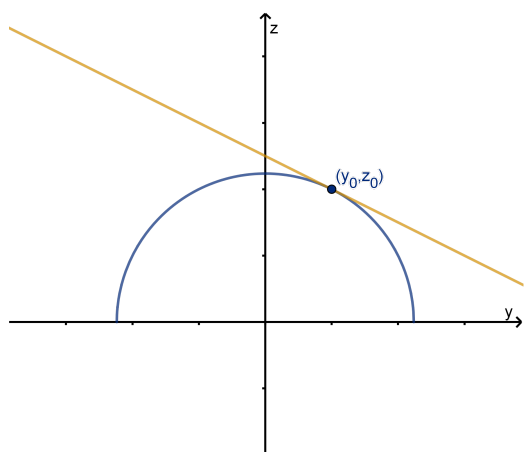
1. This is the point-slope form of the equation of a plane. $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ are the slopes.
2. x_0 and y_0 are numbers, so $f_x(x_0, y_0)$ and $f_y(x_0, y_0)$ are numbers. The variables in this equation are x, y and z.

The cross sections of the tangent plane give the equation of the tangent lines we learned in single variable calculus.

$$y = y_0 \qquad\qquad x = x_0$$

$$z - z_0 = f_x(x_0, y_0)(x - x_0) + 0 \qquad\qquad z - z_0 = 0 + f_y(x_0, y_0)(y - y_0)$$





This shows that the tangent plane does contain these two tangent lines.

Example 4.5.2

Writing the Equation of a Tangent Plane

Give an equation of the tangent plane to $f(x, y) = \sqrt{xe^y}$ at $(4, 0)$

Solution

Writing the formula requires us to fill in 5 values.

1. $x_0 = 4$ is given.
2. $y_0 = 0$ is given.
3. z_0 is the height of the graph at $(4, 0)$ which is $\sqrt{4e^0} = 2$.
4. To compute $f_x(x_0, y_0)$ we compute the partial derivative function

$$f_x(x, y) = \frac{1}{2\sqrt{x}}\sqrt{e^y}.$$

   Then we evaluate at $(4, 0)$.

$$f_x(4, 0) = \frac{1}{2\sqrt{4}}\sqrt{e^0} = \frac{1}{4}.$$

5. $f_y(x_0, y_0)$ is similar though we will use the chain rule.

$$f_y(x, y) = \sqrt{x}\frac{1}{2\sqrt{e^y}}e^y$$

$$f_y(4, 0) = \sqrt{4}\frac{1}{2\sqrt{e^0}}e^0 = 1$$

We plug these values into the tangent plane formula.

$$z - 2 = \frac{1}{4}(x - 4) + 1(y - 0)$$

which simplifies to

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

Question 4.5.3

How Do We Rewrite a Tangent Plane as a Function?

Definition

If we write z as a function $L(x, y)$, we obtain the **linearization** of f at (x_0, y_0).

$$L(x, y) = f(x_0, y_0) + f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

If the graph $z = f(x, y)$ has a tangent plane, then $L(x, y)$ approximates the values of f near (x_0, y_0).

Notice $f(x_0, y_0)$ just calculates the value of z_0. This formula is equivalent to the tangent plane equation after we solve for z by adding z_0 to both sides.

Example 4.5.4

Approximating a Function

Use a linearization to approximate the value of $\sqrt{4.02e^{0.05}}$.

Solution

We don't know $\sqrt{4.02e^{0.05}}$, but we can think of this as the value of the function $f(x, y) = \sqrt{xe^y}$. We don't know the value of this function at $(4.02, 0.05)$, but the point $(4, 0)$ is nearby, and we can evaluate it there. This is where we'll produce our linearization. We already produced the equation of the tangent plane in Example .4.5.2.

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

We write z as the function $L(x, y)$ and solve for it:

$$L(x, y) = 2 + \frac{1}{4}(x - 4) + y$$

For points near $(4, 0)$, $L(x, y)$ is close to $f(x, y)$. This is the basis of our approximation.

$$\begin{aligned}
\sqrt{4.02e^{0.05}} = f(4.02, 0.05) &\approx L(4.02, 0.05) \\
&\approx 2 + \frac{1}{4}(4.02 - 4) + 0.05 \\
&\approx 2 + 0.005 + 0.05 \\
&\approx 2.055
\end{aligned}$$

Question 4.5.5

How Does Differential Notation Work in More Variables?

The one-variable differential is a shorthand way to express change in the linearization of a function. The differential dx is an independent variable. It can take on any value. The differential dy depends on both x_0 and dx.

$$dy = f'(x_0)dx$$

Once we've chosen x_0 and dx, dy is the amount that the tangent line to $y = f(x)$ at x_0 rises when we increase x by dx.

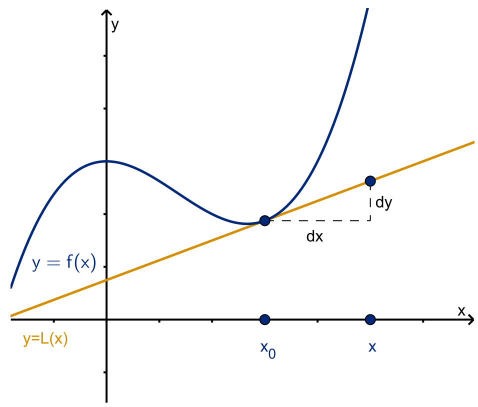


Figure: The differentials dx and dy on the tangent line to $y = f(x)$

The differential dz measures the change in the linearization of $f(x, y)$ given particular changes in the inputs: dx and dy. It is a useful shorthand when one is estimating the error in an initial computation.

Definition

For $z = f(x, y)$, the **differential** or **total differential** dz is a function of a point (x_0, y_0) and two independent variables dx and dy.

$$\begin{aligned} dz &= f_x(x_0, y_0)dx + f_y(x_0, y_0)dy \\ &= \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy \end{aligned}$$

Remark

The differential formula is just the tangent plane formula with

$$dz = z - z_0 \qquad dx = x - x_0 \qquad dy = y - y_0.$$

An old trigonometry application is to measure the height of a pole by standing at some distance. We then measure the angle θ of incline to the top, as well as the distance b to the base. The height is $h = b \tan\theta$.

a If the distance to the base is $13m$ and the angle of incline is $\frac{\pi}{6}$, what is the height of the pole?

b Human measurement is never perfect. If our measurement of b is off by at most $0.1m$ and our measurement of θ is off by at most $\frac{\pi}{120}$, use a differential to approximate the maximum possible error in our h.

Solution

a The height is $13\tan\frac{\pi}{6} = \frac{13}{\sqrt{3}}$.

b To compute the differential, we need to know the partial derivatives of h:

$$\frac{\partial h}{\partial b} = \tan\theta \qquad \frac{\partial h}{\partial \theta} = b\sec^2\theta$$

$$\left.\frac{\partial h}{\partial b}\right|_{\left(13,\frac{\pi}{6}\right)} = \frac{1}{\sqrt{3}} \qquad \left.\frac{\partial h}{\partial \theta}\right|_{\left(13,\frac{\pi}{6}\right)} = \frac{56}{3}$$

We can now compute the differential.

$$\begin{aligned} dh &= \frac{\partial h}{\partial b}db + \frac{\partial h}{\partial \theta}d\theta \\ &= \frac{1}{\sqrt{3}}db + \frac{56}{3}d\theta \end{aligned}$$

dh is largest when $db = 0.1$ and $d\theta$ is $\frac{\pi}{120}$.

$$\begin{aligned} \max dh &= \frac{1}{\sqrt{3}}(0.1) + \frac{56}{3}\frac{\pi}{120} \\ &= \frac{1}{10\sqrt{3}} + \frac{13\pi}{90} \end{aligned}$$

Summary Questions

Q1 What do you need to compute in order to write the equation of a tangent plane to $z = f(x, y)$ at (x_0, y_0, z_0)?

Q2 For what kinds of functions are linear approximations useful?

Q3 How are the tangent plane and the linearization related?

Q4 How is the differential defined for a two variable function? What does each variable in the formula mean?

4.5.1

Q5 Let $p(x, y) = 3x + 5y - 2$.

- **a** What is the graph $z = p(x, y)$? What is the significance of 3, 5 and -2?
- **b** Give the equation of the tangent plane to $z = p(x, y)$ at $(1, 4, 21)$
- **c** How is the tangent plane equation related to $z = p(x, y)$? Why does this make sense?

Q6 Olivia computes the tangent plane of $z = x^2 + y^2$ at $(4, 3, 25)$. Her answer is $z - 25 = 2x(x - 4) + 2y(y - 3)$.

- **a** Is this the equation of a plane? Explain.
- **b** What does Olivia need to do to fix her answer?

Q7 If the equation of the tangent plane of $z = f(x, y)$ does not have a y in it, does that mean that y is a free variable of f? Explain.

Q8 Can our tangent plane formula ever give us a plane parallel to the xy-plane? The xz-plane? The zy-plane? Explain.

4.5.2

Q9 Compute the equation of the tangent plane to $z=\sqrt{36-4x^2-y^2}$ at $(2,2,4)$.

Q10 Let $g(x,y)=\frac{3x^2+4x-2}{e^{(y^3)}}$. Write the equation of the tangent plane to $z=g(x,y)$ at $(0,1)$.

Q11 Let $f(x,y)=\sqrt{\frac{y}{x}}$. Write the equation of the tangent plane to $z=f(x,y)$ at $(4,36,3)$.

Q12 Let $f(x,y)=\ln(x^2+y)$. Write the equation of the tagent plane to $z=f(x,y)$ at $(e^3,0,6)$.

4.5.3

Q13 Write a linearization of $f(x,y)=ye^{xy}$ at $(3,2)$.

Q14 Write a linearization of $g(x,y)=e^{x^2+y^2}$ at $(3,-4)$.

4.5.4

Q15 Suppose you want to approximate $\sqrt{5.5e^{0.3}}$ by hand. Would using the linearization of $f(x,y)=\sqrt{xe^y}$ at $(5,0)$ be a good strategy? Explain.

Q16 Show how to use an appropriate linearization to approximate $\frac{1}{5.12}\sin\left(\frac{31\pi}{30}\right)$.

Q17 Let $g(x,y)=\frac{x^2}{y}$. Suppose you don't remember how to divide decimals. Show how you can use a linearization of g to approximate $\frac{3.97^2}{1.05}$.

Q18 Show how to use a linearization to approximate the value of $\sqrt{(4.02)^2+\sqrt{80.93}}$ by hand.

4.5.5

Q19 Let $f(x,y) = \frac{y}{x^2+y^2}$. Write the differential of f at $(4,3)$.

Q20 Let $g(p,q) = p\ln q$. Write the differential of g at $(3,e^2)$.

Q21 Boris is measuring the area of a rectangular field, so he can decide how much grass seed to buy.

According to his measurements, the field is $30m$ by $50m$, giving an area of $1500m^2$. If we accept that each of his measurements has an error no larger than $0.2m$, use a differential to approximate the maximum error in his area computation.

Q22 Suppose I decide to invest $\$10,000$ expecting a 6% annual rate of return for 12 years, after which I'll use it to purchase a house. The formula for compound interest

$$P = P_0 e^{rt}$$

indicates that when I want to buy a house, I will have $P = 10,000e^{0.72}$.

I accept that my expected rate of return might have an error of up to $dr = 2\%$. Also, I may decide to buy a house up to $dt = 3$ years before or after I expected.

a Write the formula for the differential dP at $(r_0, t_0) = (0.06, 12)$.

b Given my assumptions, what is the maximum estimated error dP in my initial calculation?

c What is the actual maximum error in P?

Q23 Let $z = 2x - y^3$. At the point $(x,y) = (5,2)$, what is the maximum value of the differential dz?

Q24 Let $f(x,y)$ be a function. What differential and what inputs into that differential would you use to approximate $f(5.5, 3.2) - f(4.7, 3.8)$.

Synthesis & Extension

Q25 Let $L(x,y)$ be the linearization of $f(x,y)$ at $(3,2)$. If $f_{yy}(x,y) < 0$ for all (x,y), at which points can we guarantee that $L(x,y)$ either under or overestimates the value of $f(x,y)$? Explain.

Q26 Let $f(x,y) = 25 - (x+1)^2 - y(y-3)^2$. Describe the set of points (a,b) such that the tangent plane to $z = f(x,y)$ at $(a,b,f(a,b))$ passes through the origin.

Q27 Here is a table of selected values for a function $f(x, y)$

		x					
		0	2	4	6	8	10
y	0	2	5	8	10	11	11
	2	6	9	12	14	15	15
	4	9	12	15	17	18	18
	6	12	15	18	20	21	21
	8	14	17	20	22	23	23
	10	17	20	23	25	23	23

a Using any reasonable approximation method, show how to produce a linearization of $f(x, y)$ at $(4, 2)$.

b Does your linearization over or underestimate $f(10, 2)$? Explain what that suggests about one or more derivatives of $f(x, y)$.

Q28 **a** Give an equation of the plane that passes through the points $(3, 4, 2)$, $(5, 5, 1)$ and $(6, 2, 6)$.

b Suppose there is a function $f(x, y)$ and the plane in part **a** is tangent to the graph $z = f(x, y)$ at $(3, 4, 2)$. What partial derivatives of f can you compute exactly (be specific)? Compute them.

Chapter 5

Vectors in Calculus

This chapter introduces vectors and their applications to calculus. We will use them to compute directional derivatives, to differentiate compositions of functions, and to find minimum and maximum values of a function.

Contents

Section 5.1

Vectors

Goals:

1. Distinguish vectors from scalars (real numbers) and points.
2. Add and subtract vectors, multiply by scalars.
3. Express real world vectors in terms of their components.

Calculus is the study of change. We defined the partial derivative to be instantaneous rate of change of a multi-variable function when one variable changed but the other stayed constant. If we want to describe a more complicated change, we will need new notations and vocabulary to describe them. We will need vectors.

Question 5.1.1

What is a Vector?

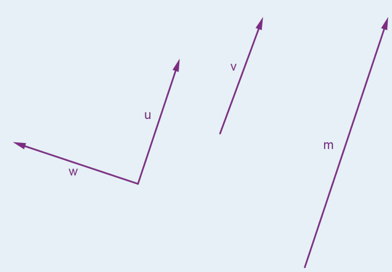
A vector is a way of describing a change in position in n-space. To keep things simple, we'll start with vectors in the plane. We need two pieces of information to identify a vector.

Definition

A **vector** in 2-space consists of a magnitude (length) and a direction. Two vectors with the same magnitude and the same direction are **equal**.

Example

Here are four vectors in 2-space (the plane) represented by arrows. Two of these vectors are equal.



Here are some vectors

- 3 miles south

- The force that a magnetic field applies to a charged particle
- The velocity of an airplane

Here are some non-vectors

- 17
- The mass of an automobile
- 3:15 PM
- Atlanta, GA

Question 5.1.2

How Do We Denote Vectors?

When defining a new type of object, we need to agree on a notation. This allows us to communicate clearly which vector we are referring to. One way of denoting a vector is by its endpoints.

Endpoint Notation

The vector $\vec{v}$ from point A to point B can be represented by the notation

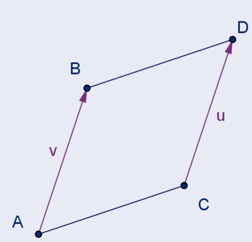
$$\overrightarrow{AB}.$$

A is the **initial point** and B is the **terminal point**.

How does this notation interact with the idea of equal vectors?

Theorem

$\overrightarrow{AB} = \overrightarrow{CD}$ if and only if $ABDC$ is a parallelogram (perhaps a squished one).



The plane has a coordinate system. We can take advantage of this to produce a more quantitative notation for vectors.

Question 5.1.2 How Do We Denote Vectors?

Coordinate Notation

We can represent a vector in the Cartesian plane by the x and y components of its displacement. If $A = (2, 3)$ and $B = (5, 1)$, then $\overrightarrow{AB}$ increases x by $5 - 2 = 3$ and y by $1 - 3 = -2$. We can represent

$$\overrightarrow{AB} = \langle 3, -2 \rangle$$

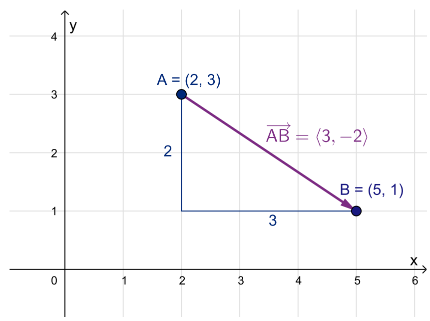


Figure: The x and y components of a vector

We can use coordinate notation to quickly test whether two vectors are equal.

Theorem

$\vec{v} = \vec{u}$ if and only if their coordinate representations match in each component.

We can also measure slope using the coordinate notation. For the vector $\vec{v} = \langle a, b \rangle$:

- b represents the displacement in the y-direction (rise).
- a represents the displacement in the x-direction (run).
- The slope of $\vec{v}$ is $\frac{\text{rise}}{\text{run}} = \frac{b}{a}$.

Vectors are not points, but their coordinate notations look awfully similar. We can connect them more formally. Every point in a Cartesian coordinate system has a **position vector**, which gives the displacement of that point from the origin. The components of the vector are the coordinates of the point.

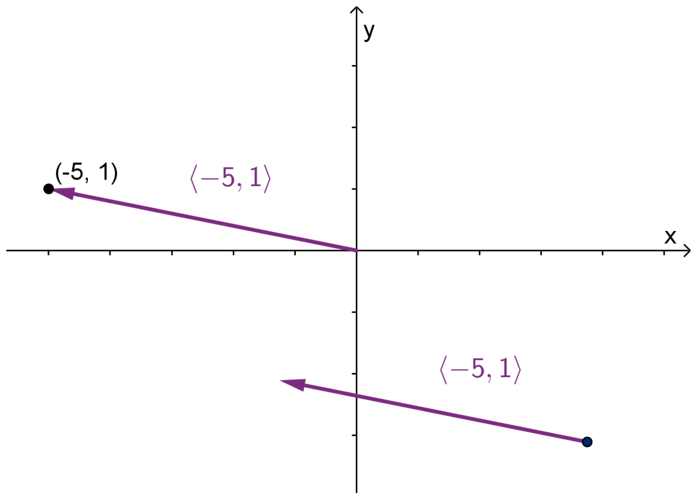


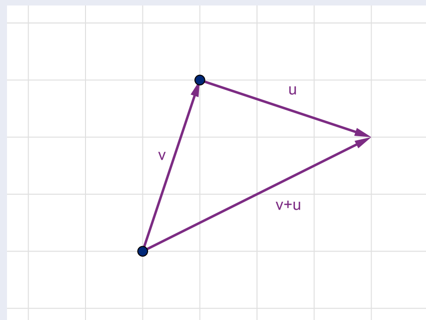
Figure: There is only one point equal to $(-5, 1)$, but there are many vectors equal to $\langle -5, 1 \rangle$.

Question 5.1.3
What Arithmetic Can We Perform with Vectors?

Unlike locations (points), displacements (vectors) can be added and multiplied. This arithmetic allows unlocks a variety of computations and measurements, specifically it will allow us to do calculus. Since we have multiple ways of representing vectors, we will want to understand how to perform these operations with each of those representations.

Vector Sums

The **sum** of two vectors $\vec{v}+\vec{u}$ is calculated by positioning $\vec{v}$ and $\vec{u}$ head to tail. The sum is the vector from the initial point of one to the terminal point of the other. In coordinate notation, we just add each component numerically.



$$\begin{array}{r} \langle\ 1,\quad 3\rangle \\ +\langle\ 3,\ -1\rangle \\ \hline \langle\ 4,\quad 2\rangle \end{array}$$

Scalar Multiples

Given a number (called a scalar) λ and a vector $\vec{v}$ we can produce the **scalar multiple** $\lambda\vec{v}$, which is the vector in the same direction as $\vec{v}$ but λ times as long.

If λ is negative then $\lambda\vec{v}$ extends in the opposite direction. Either way, we say $\lambda\vec{v}$ is **parallel** to $\vec{v}$.

In coordinates scalar multiplication is distributed to each component. For example:

$$2.5\,\langle 6,4\rangle = \langle 15,10\rangle$$

Example 5.1.4

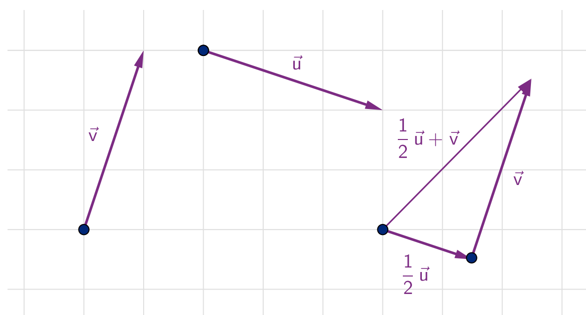
Performing Vector Arithmetic

Given diagrams of two vectors $\vec{u}$ and $\vec{v}$, how would we calculate $\frac{1}{2}\vec{u}+\vec{v}$?

What if we are instead given the components $\vec{u}=\langle a,b\rangle$ and $\vec{v}=\langle c,d\rangle$?

Solution

After drawing a random $\vec{u}$ and a random $\vec{v}$, we draw $\frac{1}{2}\vec{u}$ in the same direction as $\vec{u}$ but is half as long. We place it head to tail with $\vec{v}$, and $\frac{1}{2}\vec{u}+\vec{v}$ completes the triangle.



In coordinates the computation is as follows.

$$\begin{aligned}\frac{1}{2}\vec{u}+\vec{v} &= \frac{1}{2}\langle a,b\rangle+\langle c,d\rangle\\ &= \left\langle \frac{1}{2}a,\frac{1}{2}b\right\rangle+\langle c,d\rangle\\ &= \left\langle \frac{1}{2}a+c,\frac{1}{2}b+d\right\rangle\end{aligned}$$

Question 5.1.5

What Is Standard Basis Notation?

Vector arithmetic gives us another notation that takes advantage of our algebraic intuition. We can represent any vector in the plane as a sum of scalar multiples of the following **standard basis vectors**.

Standard Basis Vectors

The emphstandard basis vectors in $\mathbb{R}^2$ are

$$\vec{i} = \langle 1, 0 \rangle$$

$$\vec{j} = \langle 0, 1 \rangle$$

For example, the vector $\langle 3, -5 \rangle$ can be written as $3\vec{i} - 5\vec{j}$. You can check yourself that the sum on the right gives the correct vector.

Question 5.1.6

How Do We Measure the Length of a Vector?

A vector consists of two pieces of information: magnitude and direction. How do we measure these? Length is the distance between the endpoints. We already have a method for measuring distance in the plane.

Definition

The **length** or **magnitude** of a vector is calculated using the distance formula and notated $|\vec{v}|$. If $\vec{v} = a\vec{i} + b\vec{j}$, then

$$|\vec{v}| = \sqrt{a^2 + b^2}$$

Example 5.1.7

The Length of a Vector

If $\vec{v} = \langle 3, -5 \rangle$ calculate $|\vec{v}|$

Solution

$|\vec{v}| = \sqrt{3^2 + (-5)^2} = \sqrt{34}$

Definition

A **unit vector** is a vector of length 1. Given a vector $\vec{v}$ the scalar multiple

$$\frac{1}{|\vec{v}|}\vec{v}$$

is a unit vector in the same direction as $\vec{v}$.

Question 5.1.8

How Do We Measure the Direction of a Vector?

Direction cannot be described as clearly as length. How do we even measure it? A partial answer is to measure the difference in direction between two vectors.

Angles are a good way of comparing directions. In general, two vectors will not intersect to form an angle, so we use the following definition:

Definition

The angle between two vectors is the angle they make when they are placed so their initial points are the same.

If they make a right angle, we call them **orthogonal**. If they make an angle of 0 or π, they are parallel.

Question 5.1.9

How Do We Denote Vectors in Higher Dimensions?

Higher dimensional vectors represent displacements in higher dimensional spaces. We can call a vector in n-space an n-vector. We can still denote and n-vector by its endpoints. We can also denote it in coordinate notation, but we need more components.

Example

If $A = (2, 4, 1)$ and $B = (5, -1, 3)$ then

$$\overrightarrow{AB} = \langle 3, -5, 2 \rangle .$$

In three space, we add another standard basis vector $\vec{k}$.

Standard basis for 3-vectors

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

Example

$$\langle 3, -5, 2 \rangle = 3\vec{i} - 5\vec{j} + 2\vec{k}$$

Higher dimensions still have a standard basis, but at this point the naming conventions are less standard. $\{\vec{e}_1, \vec{e}_2, \vec{e}_3, \ldots, \vec{e}_n\}$ is common for n-vectors.

Length of a Vector

The length of an n-vector derives from the distance formula in n-space.

$$|\langle a_1, a_2, a_3, \ldots, a_n \rangle| = \sqrt{a_1^2 + a_2^2 + a_3^2 + \cdots + a_n^2}$$

We might be concerned that direction becomes an even more difficult concept to work with as the dimension increases. However, angles are a valid a way of comparing directions any dimension (though they may be more difficult to compute).

Angles Between Vectors

Any two vectors with the same initial point lie in a plane. Their angle is a two-dimensional measurement. However there is no good way to measure clockwise in 3 or more dimensions. The angle between two vectors is never negative, nor more than π.

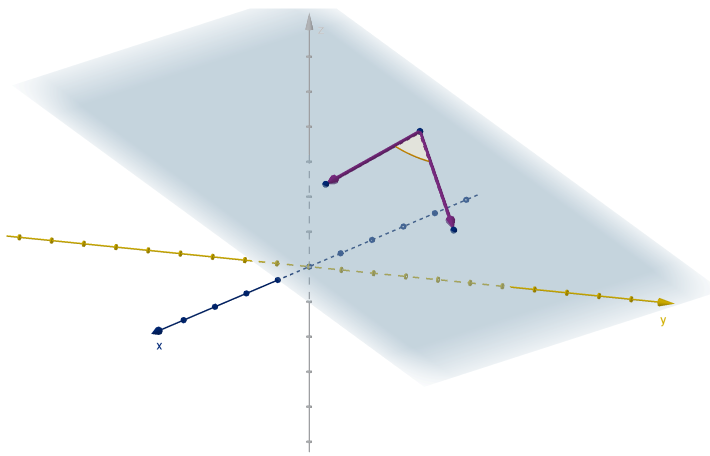


Figure: Two 3-vectors with a common initial point, the plane that contains them, and the angle between them

Section 5.1 Exercises

Summary Questions

Q1 How is a vector similar to a point? To a number?

Q2 How is a vector different from a point? From a number?

Q3 How can you tell if two vectors point in the same direction? Opposite directions?

Q4 If $\vec{u}$ and $\vec{v}$ are position vectors of the points P and Q, how are $\vec{u}$ and $\vec{v}$ related to $\overrightarrow{PQ}$?

5.1.1

Q5 Which of the following are vectors?

i. The reading on a speedometer.
ii. The intersection of two lines.
iii. Five miles toward Atlanta.
iv. The length of a string.
v. The velocity of a projectile.

Q6 Which of the following are vectors?

i. The displacement of a key on a keyboard, when pressed.
ii. The speed of light.
iii. The center of the earth.
iv. The force applied by a rocket engine.
v. The mass of five hippopotamuses.

Q7 If $\overrightarrow{AB} = \overrightarrow{AC}$, what does that tell us about the points B and C? Explain.

Q8 If $\overrightarrow{AB} = \overrightarrow{BA}$, what does that tell us about the points A and B? Explain.

5.1.2

Q9 If $A = (8, 7, 11)$ and $B = (2, 3, 15)$ write the vector $\overrightarrow{AB}$

a. in terms of its components

b. in standard basis notation

Q10 If $P = (-2, 3, 5)$ and $Q = (-2, 0, -4)$ write the vector $\overrightarrow{PQ}$

a. in terms of its components

b. in standard basis notation

Q11 What is the slope of the vector $-4\vec{i} + 10\vec{j}$?

Q12 Give three different vectors of slope $\frac{3}{7}$.

Q13 Suppose two different vectors have the equal slopes. How are they related?

Q14 Given a number m, give two different vectors with slope m.

5.1.3

Q15 Let $\vec{u}$ be a vector. How are the magnitude and direction of $\vec{u}$ and $2\vec{u}$ related?

Q16 How is the direction and magnitude of $\vec{u}$ related to the direction and magnitude of $-\vec{u}$?

Q17 Given diagrams of two vectors $\vec{u}$ and $\vec{v}$, how would we draw $\vec{u}-\vec{v}$? What it its significance?

Q18 If $\vec{u}$ is a vector and $\vec{2u}=\vec{u}$, what does that tell us about $\vec{u}$? Explain.

Q19 If $\vec{u}=\overrightarrow{AB}$, $\vec{v}=\overrightarrow{AC}$, and $\frac{1}{2}\vec{u}+\frac{1}{2}\vec{v}=\overrightarrow{AD}$, where is D?

Q20 If $\vec{u}=\overrightarrow{AB}$, $\vec{v}=\overrightarrow{AC}$, and $\frac{1}{5}\vec{u}+\frac{4}{5}\vec{v}=\overrightarrow{AD}$, where is D?

5.1.4

Q21 Let $\vec{u}=4\vec{i}+3\vec{j}$ and $\vec{v}=5\vec{i}-2\vec{j}$. Compute $\vec{u}+\vec{v}$.

Q22 Let $\vec{w}=\langle 5,-1\rangle$ and $\vec{v}=\langle 12,10\rangle$. Compute $\vec{w}-\vec{v}$.

Q23 For Lindsey to get from her house to Sam's house, she travels $5mi$ north and $3mi$ west. To get to Russel's house, she travels $2mi$ due south. What displacement would get her from Sam's house to Russel's house?

Q24 One can get from Atlanta to Decatur by travelling $8km$ east and $2km$ north. To get from Decatur to Covington, one can travel $43km$ east and $20km$ south. Describe how to get directly from Atlanta to Covington.

Q25 Using the diagram below, describe each vector in terms of $\vec{u}$ and $\vec{v}$ using vector addition and scalar multiplication. Use the fact that $ACDB$ and $ACBE$ are parallelograms.

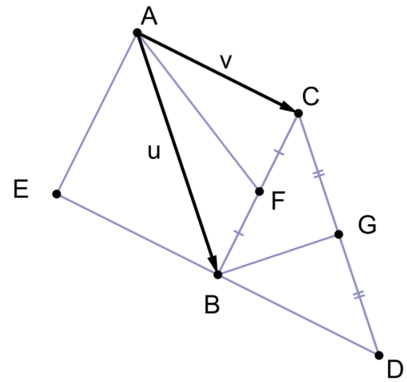
a $\overrightarrow{EB}$

b $\overrightarrow{CG}$

c $\overrightarrow{BC}$

d $\overrightarrow{AF}$

e $\overrightarrow{GB}$



Q26 Using the diagram below, describe each vector in terms of $\vec{u}$ and $\vec{v}$ using vector addition and scalar multiplication. Use the fact that $ACBD$ is a parallelogram, and the marked segments are congruent.

a $\overrightarrow{BD}$

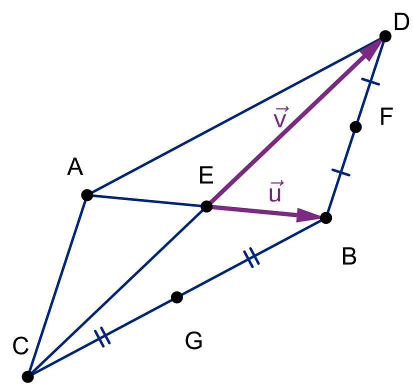
b $\overrightarrow{EA}$

c $\overrightarrow{DC}$

d $\overrightarrow{BG}$

e $\overrightarrow{AG}$

f $\overrightarrow{CF}$



5.1.5

Q27 Write $\langle 5, 2\rangle$ in standard basis notation.

Q28 For any numbers a and b, use the definition of $\vec{i}$ and $\vec{j}$ to show that $a\vec{i} + b\vec{j} = \langle a, b\rangle$.

5.1.6

Q29 Compute the length of $\vec{u} = \langle -5, 12 \rangle$.

Q30 Given a nonzero vector $\vec{u}$, many vectors of length 5 are parallel to $\vec{u}$? Explain.

Q31 Find a unit vector in the direction of $3\vec{i} - \vec{j}$.

Q32 Find a unit vector in the direction of $\langle 12, -16 \rangle$.

5.1.7

Q33 If $\vec{u}$ and $\vec{v}$ are vectors in $\mathbb{R}^2$ whose components are all positive, what is the largest possible angle between $\vec{u}$ and $\vec{v}$?

Q34 Explain the difference between the terms "perpendicular" and "orthogonal."

Q35 Suppose two vectors do not have the same inital point, but when we represent them by arrows, the arrows happen to cross. Is the angle made in the crossing equal to the angle between the vectors (as we defined it)?

Q36 Describe all the vectors that make an angle of $\frac{\pi}{4}$ with $\vec{v} = -\vec{j}$.

5.1.8

Q37 If $\vec{u} = \langle 2, 0, 3 \rangle$ and $\vec{v} = \langle 5, 6, 0 \rangle$, compute $3\vec{u} - 4\vec{v}$.

Q38 If $\vec{a} = 10\vec{i} - 25\vec{k}$ and $\vec{b} = 8\vec{i} - 4\vec{j} + 10\vec{k}$, compute $\frac{3}{5}\vec{a} + \frac{1}{2}\vec{b}$.

Q39 Compute the magnitude of $\vec{v} = 2\vec{i} - 7\vec{j} + 6\vec{k}$.

Q40 Compute two unit vectors parallel to $\vec{v} = \langle 4, -4, 2 \rangle$.

Q41 **a** How many different (nonequal) unit vectors are orthogonal to a given vector in $\mathbb{R}^2$? How are they related to each other?

b How many different (nonequal) unit vectors are orthogonal to a given vector in $\mathbb{R}^3$? How are they related to each other?

Q42 Let $\vec{u}$ and $\vec{v}$ be non-parallel vectors in $\mathbb{R}^3$. How many unit vectors in $\mathbb{R}^3$ are orthogonal to both $\vec{u}$ and $\vec{v}$?

Synthesis and Extension

Q43 Is the vector $\vec{v} = 2\vec{i} + 3\vec{j} + 8\vec{k}$ parallel to the plane p whose slope-intercept equation is $z = x + 2y - 7$?

Q44 For a two-variable function $f(x, y)$, $f_x(x_0, y_0)$ is the slope of the line tangent to $z = f(x, y)$ at $(x_0, y_0, f(x_0, y_0))$ in the x-direction. Write a vector $\vec{v}$ that is parallel to this line.

Q45 If $\vec{u} = \overrightarrow{AB}$ and $\vec{v} = \overrightarrow{AC}$, show that for any scalar t, $t\vec{u} + (1-t)\vec{v} = \overline{AD}$ where D is a point on the line through B and C.

Q46 If $\vec{u}$, $\vec{v}$ and $\vec{w}$ are position vectors of the three vertices A, B and C of a triangle, then $\frac{1}{3}(\vec{u}+\vec{v}+\vec{w})$ is the position vector of K, the center of mass of the triangle. Verify this by showing that K lies on the line between A and the midpoint of the side $\overline{BC}$.

Q47 Suppose we become interested in studying vectors of infinite dimension (yes this is something mathematicians actually do).

a Explain what trouble we might run computing the length of the vector $\langle 1, 1, 1, 1, 1, \ldots \rangle$.

b What would the length of the vector $\langle 1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots \rangle$ be?

Section 5.2

The Dot Product

Goals:

1. Calculate the dot product of two vectors.
2. Determine the geometric relationship between two vectors based on their dot product.
3. Calculate vector and scalar projections of one vector onto another.

The arithmetic of vectors appears to have room for expansion. While we can add and subtract vectors, we only defined how to multiply them by scalars, not by other vectors. There are in fact products of two vectors. The simplest and most useful is the dot product. The dot product takes two n-vectors and outputs a single number. Despite this apparent loss of information, the dot product is the key tool in computing the angle between vectors, the work done by a force, or the illumination in a digital scene.

Question 5.2.1
What Is the Dot Product?

Definition

The **dot product** of two vectors is a number.

For two dimensional vectors $\vec{v} = \langle v_1, v_2 \rangle$ and $\vec{u} = \langle u_1, u_2 \rangle$ we define

$$\vec{v} \cdot \vec{u} = v_1 u_1 + v_2 u_2$$

For three dimensional vectors $\vec{v} = \langle v_1, v_2, v_3 \rangle$ and $\vec{u} = \langle u_1, u_2, u_3 \rangle$ we define

$$\vec{v} \cdot \vec{u} = v_1 u_1 + v_2 u_2 + v_3 u_3$$

This pattern can be extended to any dimension.

Example 5.2.2
Computing a Dot Product

a. Calculate $\langle 2, 3, -1 \rangle \cdot \langle 4, 1, 5 \rangle$

b. Calculate $(-2\vec{i} + 4\vec{k}) \cdot (\vec{i} + 2\vec{j} - \vec{k})$

Example 5.2.2 Computing a Dot Product

Solution

a $\langle 2,3,-1\rangle \cdot \langle 4,1,5\rangle = (2)(4)+(3)(1)+(-1)(5) = 6$

b $(-2\vec{i}+4\vec{k})\cdot(\vec{i}+2\vec{j}-\vec{k}) = (-2)(1)+(0)(2)+(4)(-1) = -6$

Question 5.2.3

What Are the Algebraic Properties of the Dot Product?

Theorem

The following algebraic properties hold for any vectors $\vec{u}, \vec{v}$ and $\vec{w}$ and scalars m and n.

Commutative $\vec{u}\cdot\vec{v} = \vec{v}\cdot\vec{u}$

Distributive $\vec{u}\cdot(\vec{v}+\vec{w}) = \vec{u}\cdot\vec{v}+\vec{u}\cdot\vec{w}$

Associative $m\vec{u}\cdot n\vec{v} = mn(\vec{u}\cdot\vec{v})$

Question 5.2.4

What Is the Geometric Significance of the Dot Product?

$\vec{u}\cdot\vec{v}$ encodes key information about the magnitude and direction of $\vec{u}$ and $\vec{v}$. This geometric relationship can be derived from the algebraic properties we've established. We begin with the idea that $\vec{u}\cdot\vec{u} = |\vec{u}|^2$. This doesn't tell us the value of every dot product, but we can extend the reasoning to any pair of parallel vectors.

Theorem

If $\vec{u}$ and $\vec{v}$ are parallel then

$$\vec{u} \cdot \vec{v} = \begin{cases} |\vec{u}||\vec{v}| & \text{if } \vec{u} \text{ and } \vec{v} \text{ have the same direction} \\ -|\vec{u}||\vec{v}| & \text{if } \vec{u} \text{ and } \vec{v} \text{ have opposite directions} \end{cases}$$

Since $\vec{u}$ and $\vec{v}$ are parallel, we can write $\vec{v} = m\vec{u}$ for some scalar m. $\vec{v}$ is m times as long as $\vec{u}$. Both lengths are positive, so this means if $m > 0$ then $|\vec{v}| = m|\vec{u}|$, but if $m < 0$, then $|\vec{v}| = -m|\vec{u}|$

$$\begin{aligned} \vec{u} \cdot \vec{v} &= \vec{u} \cdot (m\vec{u}) \\ &= m\vec{u} \cdot \vec{u} \\ &= m|\vec{u}|^2 \\ &= |\vec{u}|m|\vec{u}| \\ &= \begin{cases} |\vec{u}||\vec{v}| & \text{if } \vec{u} \text{ and } \vec{v} \text{ have the same direction} \\ -|\vec{u}||\vec{v}| & \text{if } \vec{u} \text{ and } \vec{v} \text{ have opposite directions} \end{cases} \end{aligned}$$

We can establish the dot product in another special case: when the vectors are orthogonal.

Theorem

If $\vec{u}$ and $\vec{v}$ are orthogonal then

$$\vec{u} \cdot \vec{v} = 0.$$

In this case, we place $\vec{u}$ and $\vec{v}$ head to tail and draw $\vec{u} + \vec{v}$. Since $\vec{u}$ and $\vec{v}$ make a right angle, these three vectors make a right triangle. The Pythagorean theorem applies to the lengths of the vectors.

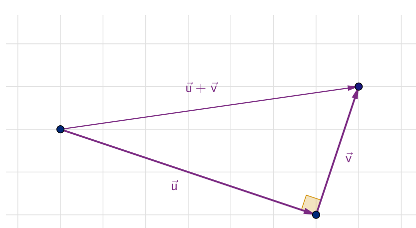
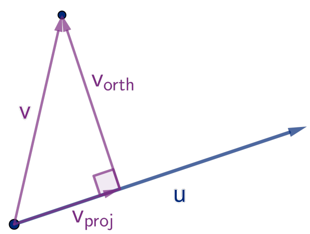


Figure: Orthogonal vectors and their sum making a right triangle

$$\begin{aligned} |\vec{u} + \vec{v}|^2 &= |\vec{u}|^2 + |\vec{v}|^2 && \text{(Pythagorean theorem)} \\ (\vec{u} + \vec{v}) \cdot (\vec{u} + \vec{v}) &= \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} \\ \vec{u} \cdot \vec{u} + \vec{u} \cdot \vec{v} + \vec{v} \cdot \vec{u} + \vec{v} \cdot \vec{v} &= \vec{u} \cdot \vec{u} + \vec{v} \cdot \vec{v} && \text{(distributive property)} \\ \vec{u} \cdot \vec{v} + \vec{v} \cdot \vec{u} &= 0 \\ 2\vec{u} \cdot \vec{v} &= 0 && \text{(commutative property)} \\ \vec{u} \cdot \vec{v} &= 0 \end{aligned}$$

Question 5.2.4 What Is the Geometric Significance of the Dot Product?

Two vectors need not be parallel or orthogonal, but given vectors $\vec{u}$ and $\vec{v}$ we can always write $\vec{v} = \vec{v}_{\text{proj}} + \vec{v}_{\text{orth}}$. We choose $\vec{v}_{\text{proj}}$ to be parallel to $\vec{u}$ and $\vec{v}_{\text{orth}}$ to be orthogonal to $\vec{u}$.



The properties of the dot product tell us that

$$\begin{aligned}\vec{u}\cdot\vec{v} =& \vec{u}\cdot(\vec{v}_{\text{proj}} + \vec{v}_{\text{orth}})\\ =& \pm|\vec{u}||\vec{v}_{\text{proj}}| + 0\end{aligned}$$

Definition

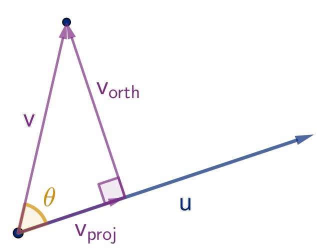
The number $\dfrac{\vec{u}\cdot\vec{v}}{|\vec{u}|}$ is called the **scalar projection** of $\vec{v}$ onto $\vec{u}$.

The scalar projection is equal to the length of $\vec{v}_{\text{proj}}$ if $\vec{v}_{\text{proj}}$ is in the same direction as $\vec{u}$. Otherwise, it is the negative of the length.

Theorem

Let $\vec{u}$ and $\vec{v}$ have the same initial point and meet at angle θ. The following formula holds in any dimension:

$$\vec{u}\cdot\vec{v} = |\vec{u}||\vec{v}|\cos\theta$$



Recall that $\cos\theta$ is

- positive when $\theta < \pi/2$
- negative when $\theta > \pi/2$
- zero when $\theta = \pi/2$.

So the sign of $\vec{u}\cdot\vec{v}$ tells us whether θ is acute, obtuse or right.

Example 5.2.5 Using the Cosine Formula

What is the angle between $\langle 1, 0, 1\rangle$ and $\langle 1, 1, 0\rangle$?

Solution

We'll apply the cosine formula, compute all of the components besides θ and solve.

$$\begin{aligned}
\langle 1,0,1\rangle \cdot \langle 1,1,0\rangle &= |\,\langle 1,0,1\rangle\,||\,\langle 1,1,0\rangle\,|\cos\theta \\
(1)(1)+(0)(1)+(1)(0) &= \sqrt{1^2+0^2+1^2}\sqrt{1^2+1^2+0^2}\cos\theta \\
1 &= \sqrt{2}\sqrt{2}\cos\theta \\
\frac{1}{2} &= \cos\theta \\
\cos^{-1}\left(\frac{1}{2}\right) &= \theta \\
\frac{\pi}{3} &= \theta
\end{aligned}$$

We can verify this by noting that these vectors are diagonals in a unit cube. We could connect them with a third diagonal to make an equilateral triangle. We may recall that an equilateral triangle has angles of $\frac{\pi}{3}$.

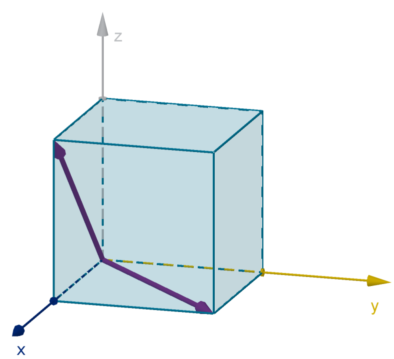


Figure: Two vectors in a unit cube

Application 5.2.6
Work

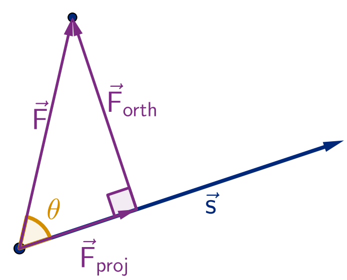
In physics, we say a force **works** on an object if it moves the object in the direction of the force. Given a force F and a displacement s, the formula for work is:

$$W = Fs$$

Application 5.2.6 Work

In higher dimensions, displacement and force are vectors. If the force and the displacement are not in the same direction, then only $\vec{F}_{\text{proj}}$ contributes to work.

$$W = \vec{F}_{\text{proj}} \cdot \vec{s} = \vec{F} \cdot \vec{s}$$



Section 5.2 Exercises

Summary Questions

Q1 What algebraic properties does a dot product share with real number multiplication?

Q2 What is the significance of the dot product of two parallel vectors?

Q3 How is the angle between two vectors related to their dot product?

Q4 What is a scalar projection, and how do you compute it?

5.2.1

Q5 What do $\vec{v} \cdot \vec{i}$ and $\vec{v} \cdot \vec{j}$ measure about $\vec{v}$?

Q6 Elaine computes $\vec{u} \cdot \vec{v}$ and gets $\langle 15, 4 \rangle$. How can you tell that Elaine got the wrong answer without even knowing what $\vec{u}$ and $\vec{v}$ are?

5.2.2

Q7 Compute the following dot products.

- a $\langle 4, 5 \rangle \cdot \langle -1, -2 \rangle$
- b $(5\vec{i} + 6\vec{j}) \cdot (\vec{i} - 2\vec{j})$
- c $\langle 2, 4, -10 \rangle \cdot \langle 0, -1, -2 \rangle$

Q8 Compute the following dot products.

- a $\langle 4, 5 \rangle \cdot \langle -1, -2 \rangle$
- b $(5\vec{i} + 6\vec{j}) \cdot (\vec{i} - 2\vec{j})$
- c $(2\vec{i} - 3\vec{k}) \cdot (7\vec{j} - \vec{k})$

5.2.3

Q9 Let $\vec{u} = \langle 2, 3 \rangle$, $\vec{v} = \langle 4, -1 \rangle$ and $\vec{w} = \langle -5, 2 \rangle$.

- a Compute $\vec{u} \cdot \vec{u}$ and $\vec{u} \cdot \vec{v}$ and $\vec{u} \cdot \vec{w}$.
- b Compute $\vec{v} \cdot \vec{u}$. How does it compare to $\vec{u} \cdot \vec{v}$?

c How is $\vec{u} \cdot \vec{u}$ related to $|\vec{u}|$?

d Compute $3\vec{u}$ and $3\vec{v}$ then take their dot product. How is it related to $\vec{u} \cdot \vec{v}$?

e Compute $\vec{v} + \vec{w}$ then compute $\vec{u} \cdot (\vec{v} + \vec{w})$. How is it related to $\vec{u} \cdot \vec{v}$ and $\vec{u} \cdot \vec{w}$?

f Why do you think we call this operation a "dot product" and not a "dot sum?"

g If you wanted to prove that relationships your noticed in b - e work for all possible vectors, how would you do that?

Q10 Expand the parentheses $2\vec{u} \cdot (3\vec{v} - \vec{w})$.

Q11 Expand the parentheses $(\vec{a} - 3\vec{b}) \cdot (5\vec{c} + 2\vec{d})$.

Q12 Factor $\vec{a} \cdot \vec{a} + 6\vec{a} \cdot \vec{b} + 9\vec{b} \cdot \vec{b}$.

5.2.4

Q13 Suppose we know that $\vec{u}$ and $\vec{v}$ are parallel, that $|\vec{v}| = 4$ and that $\vec{u} \cdot \vec{v} = -28$.

a What is the length of $\vec{u}$?

b What can you say about the directions of $\vec{u}$ and $\vec{v}$?

Q14 If $|\vec{u}| = 12$, $|\vec{v}| = 9$, and $\vec{u} \cdot \vec{v} = 0$, what is the magnitude of the vector $\vec{w} = \vec{u} + \vec{v}$?

Q15 If $|\vec{u}| = 5$ and $\vec{u} \cdot \vec{v} = 15$, what are the possible values of $|\vec{v}|$?

Q16 If $|\vec{u}| = 6$ and $|\vec{v}| = 10$ what are the greatest and least possible values of $\vec{u} \cdot \vec{v}$?

Q17 Let $\vec{v} = 7\vec{i} - 2\vec{j} + \vec{k}$, what unit vector $\vec{u}$ produces the largest possible dot product $\vec{u} \cdot \vec{v}$?

Q18 Argue that $\vec{u} \cdot \vec{v}$ cannot be any larger than $|\vec{u}||\vec{v}|$.

5.2.5

Q19 Compute the angle between $\langle 6, 1, 4\rangle$ and $\langle 7, 0, 2\rangle$.

Q20 Compute the angle between $\langle 0, 3, -5\rangle$ and $\langle 3, -4, 3\rangle$.

Q21 Let A be the vertex of a cube. Let B the a vertex closest to A and C be the vertex farthest from A. Compute the angle between $\overrightarrow{AB}$ and $\overrightarrow{AC}$.

Q22 Let A be the vertex of a cube, and B and C be any two other points on the cube. Use a dot product to explain why the angle between $\overrightarrow{AB}$ and $\overrightarrow{AC}$ cannot be larger than $\frac{\pi}{2}$. (Hint, put A at $(0, 0, 0)$.)

Synthesis and Extension

Q23 How could you use the dot product to determine whether two vectors are parallel? How does this compare with the methods we already have?

Q24 Use dot products to find at least one vector that is orthogonal to both $\langle 5, -1, 2\rangle$ and $\langle 4, 4, 1\rangle$

Q25 "Think of a vector $\vec{v}$" says Raphael, "tell me its dot product with the vector of my choice, and I'll tell you what your vector was."

- **a** Is there any mathematical way to make such a trick work? Explain.
- **b** How many dot products would you need to ask for to uniquely identify an unknown vector? What dot products would you ask for?

Section 5.3

Normal Equations of Planes

Goals:

1. Give equations of planes in both vector and normal forms.
2. Use normal vectors to measure the distance to a plane.

Question 5.3.1

What is a Normal Vector to a Plane?

In algebra, you learned the **normal equation** of a line: e.g. $2x + 3y - 12 = 0$. Why is it called this?

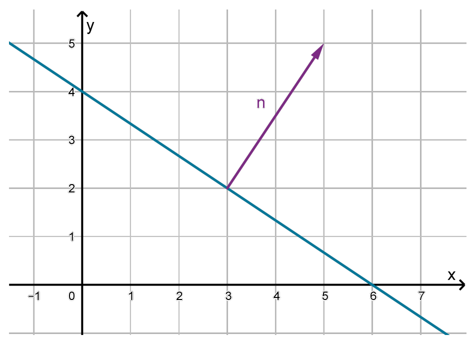


Figure: A line and one of its normal vectors

The vector $\langle 2, 3\rangle$ is a normal vector to the line, meaning it is orthogonal to any vector contained in the line. We can extend this definition to planes in 3-space. A **normal vector** to a plane is orthogonal to every vector in the plane.

Theorem

In three-dimensional space, every plane has normal vectors. They are all parallel to each other.

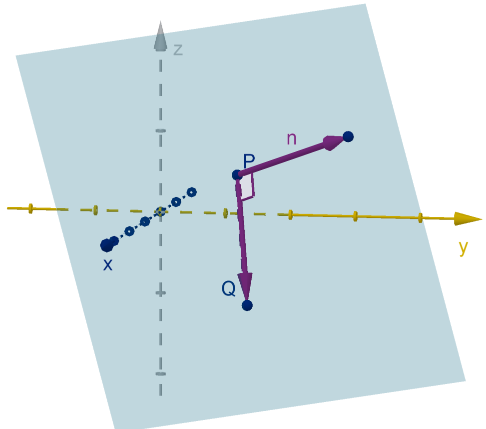


Figure: A plane, its normal vector $\vec{n}$, and a vector $\overrightarrow{PQ}$ in the plane

This gives us an avenue to test whether a point Q lies on the plane or not. If $\overrightarrow{PQ}$ is orthogonal to $\vec{n}$, then Q lies on the plane. If $\overrightarrow{PQ}$ and $\vec{n}$ make a different angle, then Q is not on the plane.

We'd like to rewrite this relationship terms of the coordinates of Q. If $\vec{r}_0$ is the position vector of P and $\vec{r}$ is the position vector of Q, then $\overrightarrow{PQ} = \vec{r} - \vec{r}_0$. The dot product gives us a simple test to see whether this vector is orthogonal to $\vec{n}$.

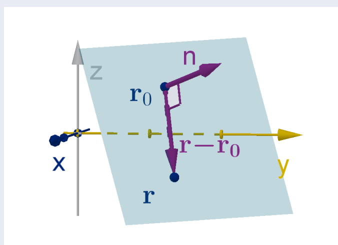
Theorem

If $\vec{r}_0 = \langle x_0, y_0, z_0 \rangle$ describes an **known point** on a plane, and $\vec{n} = \langle a, b, c \rangle$ is a normal vector. Then the **normal equation** of the plane is

$$(\vec{r} - \vec{r}_0) \cdot \vec{n} = 0$$

or

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$



Notice that since x_0, y_0 and z_0 are constants, we can distribute and collect them into a single term: d.

$$\begin{aligned} ax + by + cz - ax_0 - by_0 - cz_0 &= 0 \\ ax + by + cz + d &= 0 \end{aligned}$$

This reasoning works in any dimension to define a set of points whose displacement from a known point is orthogonal to some normal vector.

Question 5.3.1 What is a Normal Vector to a Plane?

Example

- $a(x - x_0) + b(y - y_0) = 0$ defines a line.
- $a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$ defines a plane.
- $a_1(x_1 - c_1) + a_2(x_2 - c_2) + \cdots + a_n(x_n - c_n) = 0$ defines a **hyperplane**.

Example 5.3.2
Computing a Normal Vector

Find the normal equation of the plane with intercepts $(4, 0, 0)$, $(0, 3, 0)$ and $(0, 0, 8)$. Compute a normal vector.

Solution

The normal equation of a plane has the form $ax + by + cz + d = 0$. Each of these points must satisfy this equation. We will plug them in and see what they tell me about the coefficients.

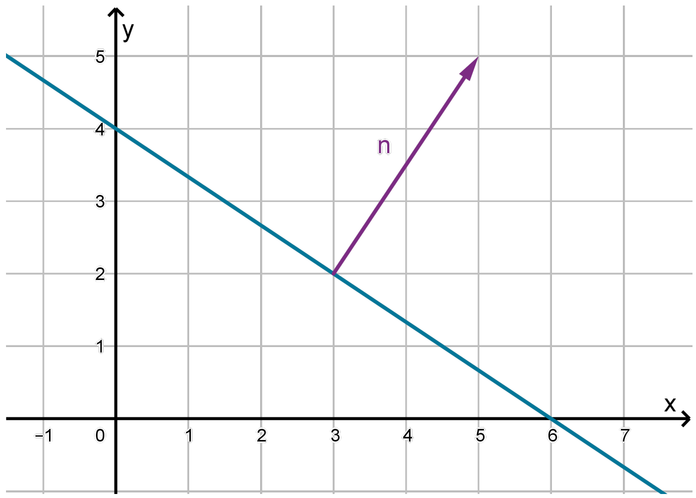
$$\begin{aligned} a(4) + b(0) + c(0) + d &= 0 & 4a + d &= 0 \\ & & d &= -4a \\ a(0) + b(3) + c(0) + d &= 0 & 3b + d &= 0 \\ & & d &= -3b \\ a(0) + b(0) + c(8) + d &= 0 & 8c + d &= 0 \\ & & d &= -8c \end{aligned}$$

There are infinitely many solutions to this system of equations. This makes sense, because there are infinitely many normal vectors to a plane. Different choices of d give $\vec{n}$'s that are scalar multiples of each other. A convenient choice for d is -24, but any nonzero value will work. $d = -24$ gives

$$6x + 8y + 3z - 24 = 0$$

The normal vector is $\langle 6, 8, 3 \rangle$.

Consider the line $2x + 3y - 12 = 0$.



This is the line with normal vector $\vec{n} = \langle 2, 3 \rangle$ and known point $P = (3, 2)$.

Example

Let $P_1 = (7, 2)$ and $P_2 = (4, 0)$.

1. Draw the vectors $\overrightarrow{PP_1}$ and $\overrightarrow{PP_2}$.
2. If you didn't have a picture, how could you use the values of $\vec{n} \cdot \overrightarrow{PP_1}$ and $\vec{n} \cdot \overrightarrow{PP_2}$ to determine which side of the line P_1 and P_2 lie on?

Solution

Since $\vec{n}$ is a normal vector, its angle with any vector in the line is $\frac{\pi}{2}$. The vectors on the same side of the line as $\vec{n}$ make an acute angle with $\vec{n}$. The vectors on the far side make an obtuse angle. Thus when $\vec{n} \cdot \overrightarrow{PP_i} < 0$, P_i lies on the far side of the line from $\vec{n}$. When $\vec{n} \cdot \overrightarrow{PP_i} > 0$, P_i lies on the same side as $\vec{n}$.

We can get more detailed information than just the sign of the dot product. We can actually compute a distance.

Theorem

Given a line, plane, or hyperplane with normal equation $L(x_1, \ldots, x_k) = 0$ and corresponding normal vector $\vec{n}$, the signed distance from the hyperplane to the point $Q = (q_1, \ldots, q_k)$ is

$$\frac{L(q_1, \ldots, q_k)}{\vec{n}}.$$

Let P be a known point on the hyperplane. The scalar projection of $\overrightarrow{PQ}$ onto $\vec{n}$ is equal to the signed distance from the hyperplane to Q.

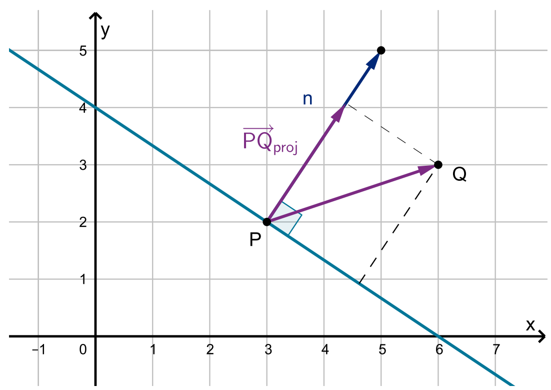


Figure: The scalar projection of $\overrightarrow{PQ}$ onto the normal vector of a line

$$\begin{aligned} \text{Distance} &= \frac{\overrightarrow{PQ} \cdot \vec{n}}{|\vec{n}|} && \text{(formula for scalar projection)} \\ &= \frac{L(q_1, \ldots, q_k)}{|\vec{n}|} && \text{(normal equation of the plane)} \end{aligned}$$

This formula is especially powerful because we do not need to know a point on the hyperplane. The equations

$$\begin{aligned} a(x - x_0) + b(y - y_0) + c(z - z_0) &= 0 \\ ax + by + cz + d &= 0 \end{aligned}$$

are equivalent, and correspond to the same normal vector. We can use whichever one we happen to have in our signed distance formula.

Example 5.3.4
The Distance from a Plane

Compute the geometric distance from the origin to the plane $6x + 8y + 3z - 24 = 0$.

Solution

$\vec{n} = \langle 6, 8, 3 \rangle$. The signed distance from the plane to the origin is

$$\begin{aligned}\frac{L(0,0,0)}{|\vec{n}|} &= \frac{(6)(0) + (8)(0) + (3)(0) - 24}{\sqrt{36 + 64 + 9}} \\ &= -\frac{24}{\sqrt{109}}\end{aligned}$$

Geometric distance cannot be negative, so it is $\frac{24}{\sqrt{109}}$.

Application 5.3.5
Support Vector Machines

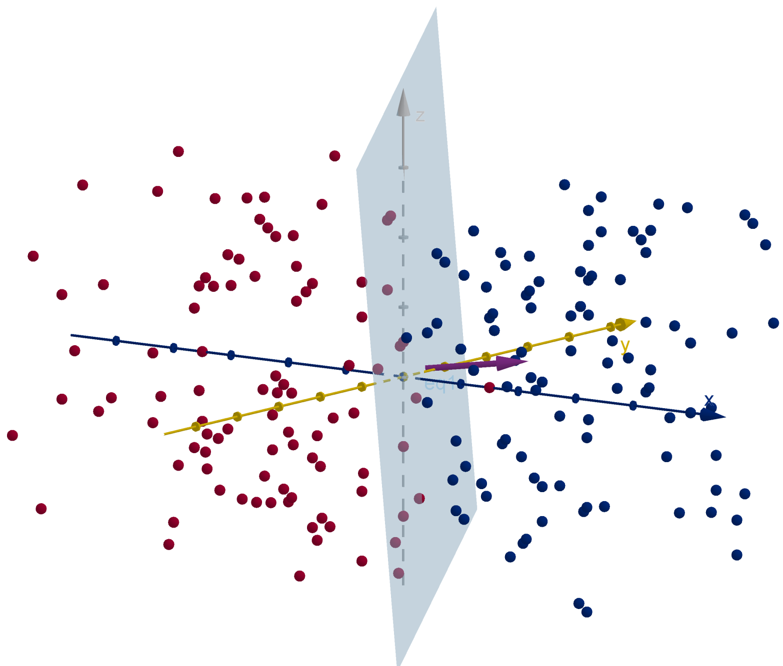
One type of machine learning involves training a computer to distinguish between two states. For example, a computer might be trained to distinguish between a cancerous tumor and a benign one.

To do this the computer is given a large set of cases. Each case is measured by numerical data, such as:

- The size of the tumor
- The location of the tumor
- The age of the patient
- Results of blood tests
- The brightness of each pixel in a CT scan or MRI

Each data type is a dimension, and each case is a point in a (probably very high) dimensional space. The computer would like a simple test to divide these cases into cancerous and benign. The test will be which side of a hyperplane they lie on. It is unlikely that any such hyperplane exists initially, so the computer attempts a sequence of transformations of the data until they are separated by a hyperplane with some degree of reliability.

Application 5.3.5 Support Vector Machines



Section 5.3 Exercises

Summary Questions

Q1 What information do you need in order to write the normal equation of a plane?

Q2 How are the normal vectors of a plane related to each other?

Q3 What is the significance of the coefficients in the normal equation of a plane?

Q4 How do we compute the signed distance from a point to a plane?

5.3.1

Q5 Is $\vec{v} = \langle 8, -3, -10 \rangle$ parallel to the plane $6x + 6y + 3z + 11 = 0$? Explain.

Q6 Is $\vec{v} = 9\vec{\imath} - 15\vec{\jmath} + 6\vec{k}$ normal to the plane $-6x + 10y - 4z + 23 = 0$? Explain.

Q7 Name a normal vector to the following planes:

i. $3x - 8y + 10z - 4 = 0$
ii. $z - 2 = 4(x + 7) - 5(y + 1)$

Q8 Suppose that $\vec{n}$ is a normal vector to $6x - 3y + 2z - 4 = 0$, that happens to also be a unit vector. Give all possible values of $\vec{n}$.

Q9 Write a normal equation of a plane parallel to $7x - 11y + 8z + 15 = 0$ that passes through the origin.

Q10 Write a normal equation of a plane parallel to $10x - 11y + z + 20 = 0$ that passes through $(2, 3, 5)$.

Q11 Given that the plane $ax + by + cz + d = 0$ passes through the origin, what can you say about a, b, c, and d?

Q12 Given that plane $ax + by + cz + d = 0$ contains the x-axis, what can you say about a, b, c, and d?

Q13 Are the planes $4x + 6y + 8z + 15 = 0$ and $10x + 15y + 20z - 7 = 0$ parallel? Explain how you know.

Q14 Suppose we know the planes $12x + 18y + 6z - 15 = 0$ and $ax + by + 4z + d = 0$ are parallel. What can you say about the values of a, b and d?

Q15 The equations $3x - y + 4z + 10 = 0$ and $-6x + 2y - 8z + k = 0$ describe the same plane. What is the value of k?

Q16 Consider the plane with normal equation $7x + y - 2z = 5$.

a Give two other normal equations of this plane.

b What are the normal vectors corresponding to the orginal equation and your two equations in **a**?

c How are these vectors in b related to each other?

5.3.2

Q17 Give a normal equation of the plane with intercepts $(10,0,0)$, $(0,-5,0)$ and $(0,0,2)$.

Q18 Give a normal equation of the plane with intercepts $(-18,0,0)$, $(0,9,0)$ and $(0,0,-4)$.

Q19 Give a normal equation of the plane through $(4,3,0)$, $(5,1,1)$ and $(-2,5,2)$.

Q20 Give a normal equation of the plane through $(1,1,1)$, $(8,1,4)$ and $(0,0,4)$.

5.3.3

Q21 Katie is computing the distance from the point $(6,3)$ to the line $2x+3y-12=0$. She notices that $(6,0)$ is the x-intercept of the line. Since $(6,3)$ is 3 units away from $(6,0)$ she concludes the distance from the point to the line is 3. What do you think of Katie's reasoning?

Q22 Consider the line L with normal equation $2x+3y-12=0$ and the point $Q=(6,3)$.

a What is the slope of L?

b What would be the slope of a line perpendicular to L?

c Write an equation (in any form you'd like) of a line K that passes through Q and is perpendicular to L.

d Compute the intersection point of P of L and K.

e What is the distance from P to Q?

f Check that your answer to e matches the distance formula we derived. Which method do you like better?

5.3.4

Q23 How far is $(5,2,1)$ from $3x+2y-5z+10=0$?

Q24 How far is $(0,0,1)$ from $3x+12y-4z+20=0$?

Q25 Are $(6,7,1)$ and $(5,-3,-4)$ on the same or different sides of $3x-10y+9z+46=0$?

Q26 The point $(x,4,5)$ lies on the same side of the plane $2x+y-2z+10=0$ as the origin does. What does that tell you about the value of x?

5.3.5

Q27 We have six images of dogs and cats. We measure four things about each, and have collected the data below. We would like to use the hyperplane $2x_1+5x_2-4x_3+10x_4+k=0$ to separate the images of dogs from the images of cats.

Type	Measurements
Cat	$(5,1,3,6)$
Dog	$(7,3,7,2)$
Dog	$(7,2,6,4)$
Dog	$(9,1,8,5)$
Cat	$(6,4,5,5)$
Cat	$(9,2,7,6)$

a. What values of k would cause the hyperplane to correctly separate the dog images from the cat images?

b. If you intended to use the hyperplane to guess whether a future image was a dog or cat, what k would you choose? Why?

Q28 Suppose we have a hyperplane that we would like to separate two sets of points, but it doesn't quite work. We measure the error of this separation by taking the sum of the geometric distances from the hyperplane of each point that is on the wrong side of the hyperplane. Suppose we were hoping that the line $2x+3y-12=0$ would separate the points of type T from the points of type S.

Type	Coordinates
T	$(6,2)$
T	$(2,1)$
T	$(5,3)$
T	$(4,4)$
S	$(1,5)$
S	$(1,1)$
S	$(4,0)$
S	$(4,2)$

a Create a diagram of these points (labelled or colored by type) and the line.

b We did not specify which side of the line should be T and which should be S. Use your diagram to decide which choice of sides will give less error.

c Compute the error in this method of separation.

d Suppose we were trying to find a better line of the form $ax+by+c=0$. When $a=2$, $b=3$ and $c=-12$, would increasing a increase or decrease the error? Justify your answer with a derivative.

Synthesis and Extension

Q29 Write the equation of a plane that contains all the points equidistant from $A=(1,-2,7)$ and $B=(7,0,5)$

Q30 Two planes are perpendicular if their normal vectors are orthogonal.

a Are $4x-7y+z-3=0$ and $5x+y+13z+25=0$ perpendicular?

b If two planes are perpendicular, is every vector in the first plane orthogonal to every vector in the second plane?

Q31 Write the normal equation of a plane that contains the x and z axes. Where have we seen this plane before?

Q32 What trouble do you run into if you try to write the equation of the plane through $(6,0,0)$, $(0,8,0)$ and $(3,4,0)$? Explain geometrically why this makes sense.

Section 5.4

The Gradient Vector

Goals:

1. Calculate the **gradient vector** of a function.
2. Relate the gradient vector to the shape of a graph and its level curves.
3. Compute **directional derivatives**.

Armed with ideas about vectors, we have the vocabulary to discuss more complex changes in the variables of a function. Rather than having one variable change and the other stay constant, we can indicate a change in both variables with a vector. When exploring these computations, we will construct one of the most important tools for multivariable calculus.

Question 5.4.1

How Do We Compute Rates of Change in Another Direction?

The partial derivatives of $f(x, y)$ give the instantaneous rate of change in the x and y directions. This is realized geometrically as the slope of the tangent line. What if we want to travel in a different direction?

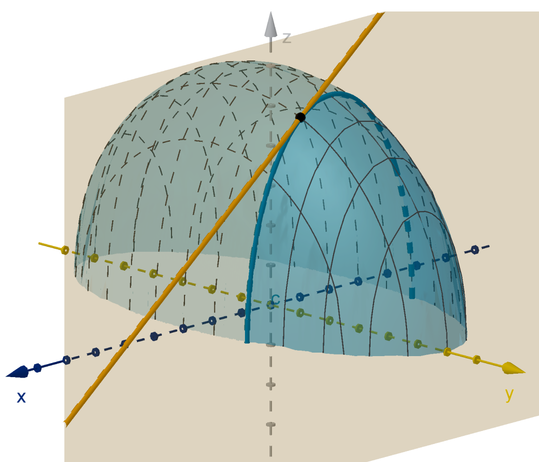


Figure: The tangent line to $z = f(x, y)$ in the x direction

Definition

Let $f(x, y)$ be a function and $\vec{u}$ be a unit vector in $\mathbb{R}^2$. The **directional derivative**, denoted $D_{\vec{u}}f$, is the instantaneous rate of change of f as we move in the $\vec{u}$ direction. This is also the slope of the tangent line to $y = f(x, y)$ in the direction of $\vec{u}$.

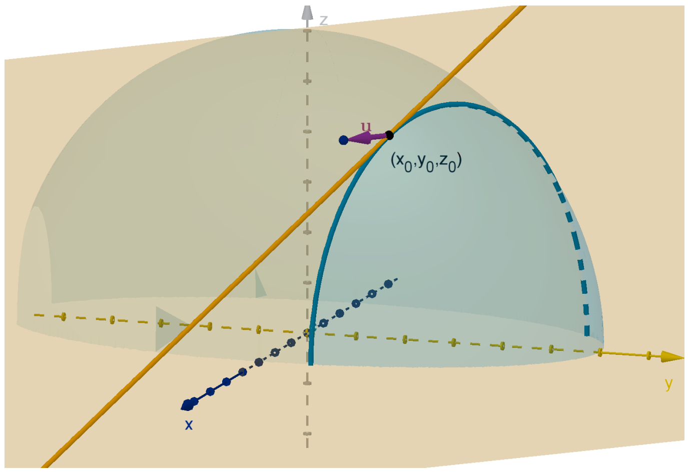


Figure: The tangent line to $f(x, y)$ in the direction of $\vec{u}$

Recall that we compute $D_x f$ by comparing the values of f at (x, y) to the value at $(x + h, y)$, a displacement of h in the x-direction.

$$D_x f(x, y) = \lim_{h \to 0} \frac{f(x + h, y) - f(x, y)}{h}$$

To compute $D_{\vec{u}} f$ for $\vec{u} = a\vec{i} + b\vec{j}$, we compare the value of f at (x, y) to the value at $(x + ta, y + tb)$, a displacement of t in the $\vec{u}$-direction.

Limit Formula

$$D_{\vec{u}} f(x, y) = \lim_{t \to 0} \frac{f(x + ta, y + tb) - f(x, y)}{t}$$

Questions:

1. What direction produces the greatest directional derivative? The smallest?
2. How are these directions related to the geometry (specifically the level curves) of the graph?
3. How these directions related to the partial derivatives?

We can explore these questions with an applet in the Other Cross Sections activity.

Question 5.4.1 How Do We Compute Rates of Change in Another Direction?

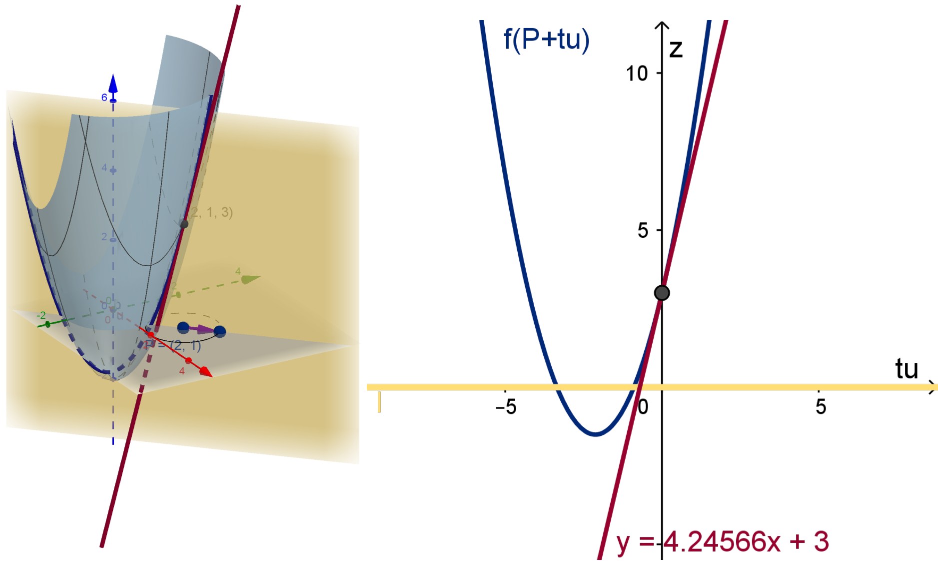


Figure: A cross section of $z = f(x, y)$ and a tangent line in the direction of $\vec{u}$

Question 5.4.2 What Is the Gradient Vector?

The relationship between the direction of maximum increase and the partial derivatives suggest that we could treat the partial derivatives like components of a vector.

Definition

The **gradient vector** of f at (x, y) is

$$\nabla f(x, y) = \langle f_x(x, y), f_y(x, y)\rangle$$

Remarks:

1. The gradient vector is a function of (x, y). Different points have different gradients.
2. $\vec{u}_{\text{max}}$, which maximizes $D_{\vec{u}}f$, points in the same direction as ∇f.
3. $\vec{u}_0$, which is tangent to the level curves, is orthogonal to ∇f.

Remark

Students often wonder: what is the geometric intuition behind the gradient vector and its properties? The answer is often disappointing, but important. The gradient vector does not have a geometric motivation. We artificially created the gradient vector because it has convenient algebraic properties. If that were the end of the story, we wouldn't bother learning about it. However, the gradient turns out to be so useful that we will study it intensely, despite its uncompelling origins.

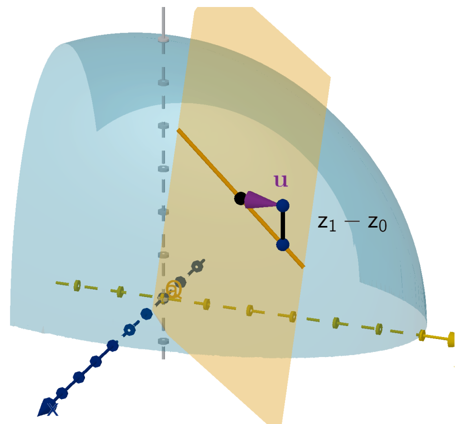
Question 5.4.3

How Do We Compute a Directional Derivative?

There are several ways to derive a formula for the directional derivative. One approach is to apply algebra and limit laws to the limit definition. A more geometric method is to exploit our previous work with the tangent plane. The directional derivative is the slope of a tangent line. The tangent lines live in the tangent plane. We can compute their slope by rise over run.

Let $\vec{u}$ be a unit vector from (x_0, y_0) to (x_1, y_1). Let the associated z values in the tangent plane be z_0 and z_1 respectively.

$$D_{\vec{u}}f(x_0, y_0) = \frac{\text{rise}}{\text{run}} = \frac{z_1 - z_0}{|\vec{u}|}$$
$$= f_x(x_0, y_0)(x_1 - x_0) + f_y(x_0, y_0)(y_1 - y_0)$$
$$= \nabla f(x_0, y_0) \cdot \vec{u}.$$



Functions of More Variables

We can also define directional derivatives of higher variable functions with analogous results.

- $f(x_1, \ldots, x_n)$ is a differentiable function.
- $\vec{u}$ is a unit vector in $\mathbb{R}^n$.
- $D_{\vec{u}}f$ denotes the directional derivative in the direction of $\vec{u}$.
- $\nabla f = \langle f_{x_1}, \ldots, f_{x_n} \rangle$ is an n-dimensional vector function on $\mathbb{R}^n$.
- $D_{\vec{u}}f = \nabla f \cdot \vec{u}$

Synthesis 5.4.4

Directional Derivative and the Cosine Formula

Now that we have a formula for directional derivatives, we can verify our observations from earlier. Suppose $f(x, y)$ is a differentiable function and we can choose any unit vector $\vec{u}$.

a Write $D_{\vec{u}}f(x, y)$ in terms of the length of a vector and an angle.

b In what direction $\vec{u}$ will f increase fastest?

c What will be the value of $D_{\vec{u}}f(x,y)$ in that direction?

d In what direction $\vec{u}$ will $D_{\vec{u}}f(x,y) = 0$?

Solution

a Since the directional derivative is a dot product, we can apply our formula that relates the dot product to the lengths of the vectors and the angle between them.

$$\begin{aligned} D_{\vec{u}}f(x,y) &= \nabla f(x,y) \cdot \vec{u} && \text{dot product formula} \\ &= |\nabla f(x,y)||\vec{u}| \cos\theta && \text{cosine formula} \\ &= |\nabla f(x,y)| \cos\theta && \vec{u} \text{ is a unit vector} \end{aligned}$$

b Given a particular (x,y), $|\nabla f(x,y)| \cos\theta$ is largest when $\theta = 0$ This means that $D_{\vec{u}}f(x,y)$ is maximized when $\vec{u}$ is in the direction of $\nabla f(x,y)$. The formula for a unit vector in the direction of the gradient is

$$\vec{u} = \frac{1}{|\nabla f(x,y)|} \nabla f(x,y)$$

c In this direction, $\cos\theta = 1$ so $D_{\vec{u}}f(x,y) = |\nabla f(x,y)|$.

d We can solve for θ

$$\begin{aligned} D_{\vec{u}}f(x,y) &= 0 \\ |\nabla f(x,y)| \cos\theta &= 0 \text{by part (a)} \\ \cos\theta &= 0 && \text{as long as } \nabla f(x,y) \neq \vec{0} \\ \theta &= \frac{\pi}{2} \end{aligned}$$

We conclude that $\vec{u}$ must be orthogonal to $\nabla f(x,y)$.

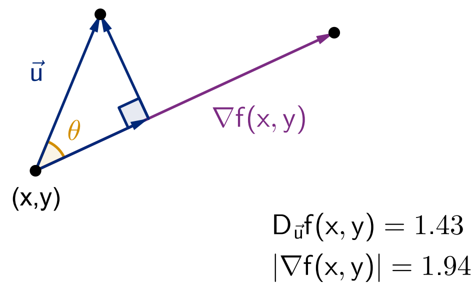


Figure: The angle between the gradient of f and a unit vector

Main Ideas

- The cosine formula for the dot product lets us relate the directional derivative to an angle.
- f increases fastest in the direction of $\nabla f(x, y)$.
- $D_{\vec{u}} f(x, y) = 0$ when $\nabla f(x, y)$ and $\vec{u}$ are orthogonal.

Example 5.4.5
A Directional Derivative

Let $f(x, y) = \sqrt{9 - x^2 - y^2}$ and let $\vec{u} = \langle 0.6, -0.8 \rangle$.

a What are the level curves of f?

b What direction does $\nabla f(1, 2)$ point?

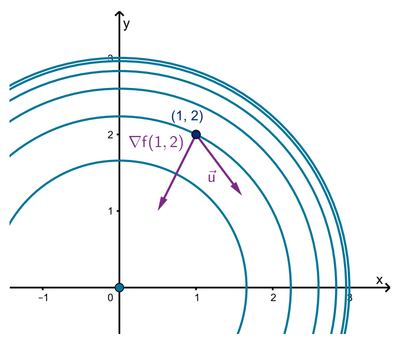
c Without calculating, is $D_{\vec{u}} f(1, 2)$ positive or negative?

d Calculate $\nabla f(1, 2)$ and $D_{\vec{u}} f(1, 2)$.

Solution

a The level curves have the equations $\sqrt{9-x^2-y^2}=c$. These solve to $x^2+y^2=9-c^2$. As c increases from 0 to 3 these are circles starting at radius 3 and shrinking to the origin. For c outside this range, the level curve has no points.

b ∇f points in the direction of increase and normal to the level curves. Since higher level curves are smaller circles, closer to the origin, $\nabla f(1,2)$ points toward the origin.



c $D_{\vec{u}}f(1,2)=\nabla f(1,2)\cdot\vec{u}$. Since $\vec{u}$ appears to make an acute angle with $\nabla f(1,2)$, we expect this dot product to be positive.

d First we need to compute $\nabla f(1,2)$.

$$\begin{aligned}\nabla f(x,y)&=\langle f_x(x,y),f_y(x,y)\rangle\\&=\left\langle\frac{1}{2\sqrt{9-x^2-y^2}}(-2x),\frac{1}{2\sqrt{9-x^2-y^2}}(-2y)\right\rangle\qquad\text{(chain rule)}\\\nabla f(1,2)&=\left\langle\frac{1}{2\sqrt{9-1^2-2^2}}(-2)(1),\frac{1}{2\sqrt{9-1^2-2^2}}(-2)(2)\right\rangle\\&=\left\langle-\frac{1}{2},-1\right\rangle\end{aligned}$$

Now we use the dot product formula to compute $D_{\vec{u}}f(1,2)$.

$$\begin{aligned}D_{\vec{u}}f(1,2)&=\nabla f(1,2)\cdot\vec{u}\\&=\left\langle-\frac{1}{2},-1\right\rangle\cdot\langle 0.6,-0.8\rangle\end{aligned}$$

$$= -0.3 + 0.8$$
$$= 0.5$$

This confirms our intuition that $D_{\vec{u}} f(1,2)$ is positive.

Example 5.4.6
Drawing the Gradient

Let $h(x, y)$ give the altitude at longitude x and latitude y. Assuming h is differentiable, draw the direction of $\nabla h(x, y)$ at each of the points labeled below. Which gradient is the longest?

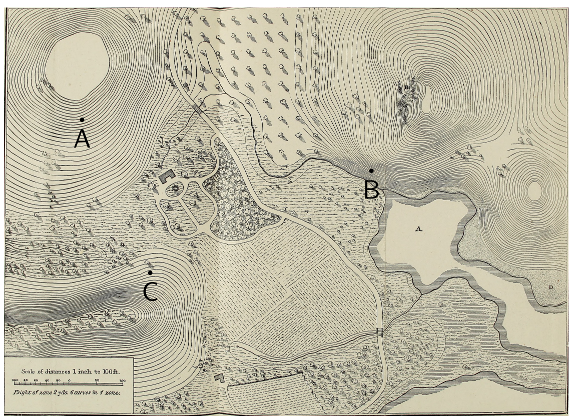
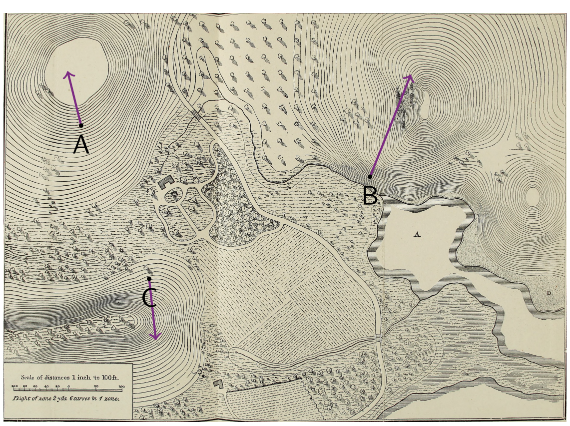


Figure: A topographical map

Solution

The gradient vector at each point is normal to the level curves, pointing uphill. The hill is steepest at B, because the level curves are closer together. This tells us that the partial derivatives are larger. Thus $\nabla h(B)$ is longer than $\nabla h(A)$ and $\nabla h(C)$.



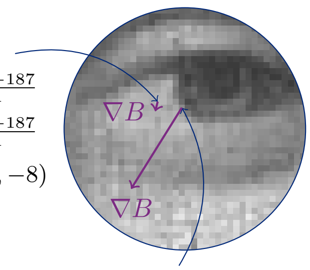
Application 5.4.7
Edge Detection

Representing an image by defining a brightness (or color) function on the pixels is simple enough, but can a computer be taught to make sense of what it sees? Image recognition is an exciting field that promises to automate and improve tasks from medical diagnosis to driving a vehicle.

The problem is daunting. What algorithm can possibly take a set of pixels and locate a tumor or a pedestrian? The first step is to identify the objects in the image. The first step of object identification is edge detection, determining where one object ends and another begins. We can do this by approximating the partial derivatives at each pixel. We compare each pixel to nearby pixels and compute rise over run (how these are chosen and averaged can significantly affect the accuracy of the algorithm).

The length of the gradient of a brightness function detects the edges in a picture, where the brightness is changing quickly.

$$\frac{\partial B}{\partial x}(336,785) \approx \frac{185-187}{1}$$
$$\frac{\partial B}{\partial y}(336,785) \approx \frac{179-187}{1}$$
$$\nabla B(336,785) \approx (-2,-8)$$



$$\frac{\partial B}{\partial x}(340,784) \approx \frac{97-139}{1}$$
$$\frac{\partial B}{\partial y}(340,784) \approx \frac{72-139}{1}$$
$$\nabla B(340,784) \approx (-42,-67)$$

Figure: A long gradient vector indicates a swift change in brightness. Its direction suggests the shape of the edges.

Notice that the gradient is long near the edge of the iris in Mona Lisa's eye. It is much shorter at a point in the white of her eye. Moreover, the gradient at the edge of the iris is approximately normal to the edge of her iris, because gradients are normal to level curves. This information can be used by an algorithm to detect not only the location of the edges, but also their direction.

Application 5.4.8
Tangent Planes to a Level Surface

Use a gradient vector to find the equation of the tangent plane to the graph $x^2 + y^2 + z^2 = 14$ at the point $(2, 1, -3)$.

There are two solutions worth comparing here.

Solution 1

We can write z as a function of x and y and apply the tangent plane formula.

$$\begin{aligned} x^2+y^2+z^2 &= 14 \\ z^2 &= 14-x^2-y^2 \\ z &= -\sqrt{14-x^2-y^2} && (z=-3 \text{ is on the negative branch of the function}) \\ f_x(x,y) &= -\frac{1}{2\sqrt{14-x^2-y^2}}(-2x) && f_x(2,1)=\frac{2}{3} \\ f_y(x,y) &= -\frac{1}{2\sqrt{14-x^2-y^2}}(-2y) && f_y(2,1)=\frac{1}{3} \end{aligned}$$

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

Solution 2

Define $F(x,y,z)=x^2+y^2+z^2$. The graph $x^2+y^2+z^2=14$ is a level surface of F. $\nabla F(2,1,-3)$ is normal to the level surface, meaning it is also a normal vector for the tangent plane.

$$\begin{aligned} \nabla F(x,y,z) &= \langle 2x,2y,2z\rangle \\ \nabla F(2,1,-3) &= \langle 4,2,-6\rangle \end{aligned}$$

We now have a normal vector $\vec{n}=\nabla F(2,1,-3)$. Our known point is $(x_0,y_0,z_0)=(2,1,-3)$. The normal equation of the plane is

$$4(x-2)+2(y-1)-6(z+3)=0.$$

Solution 2 requires more conceptual reasoning, but is computationally much easier. In fact, in some cases we cannot use Solution 1 at all because we do not know how to solve for z. Once we are comfortable with the concepts involved, the second method is generally superior for graphs of implicit equations.

Application 5.4.8 Tangent Planes to a Level Surface

Main Idea

The graph of an implicit equation can be written as a level set of a function. The gradient of that function is a normal vector to the level set and also to its tangent line/plane/hyperplane.

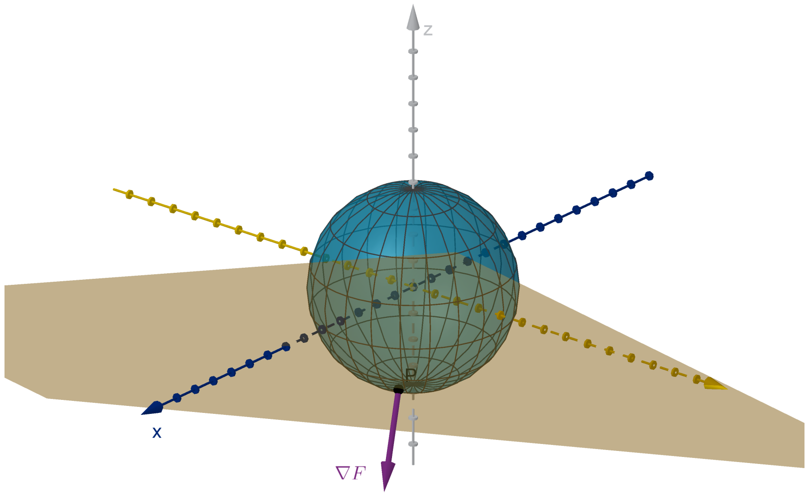


Figure: The level surface $x^2 + y^2 + z^2 = 14$, its tangent plane and ∇F.

Section 5.4 Exercises

Summary Questions

Q1 What does the direction of the gradient vector tell you?

Q2 What does the directional derivative mean geometrically?

Q3 How do you compute a directional derivative?

Q4 How is the gradient vector related to a level set?

5.4.1

Q5 Suppose that $f(3,7) = 12$ and $f(7,4) = 10$.

- **a** What is the distance from $(3,7)$ to $(7,4)$?
- **b** Approximate the rate of change of f at $(3,7)$ travelling toward $(7,4)$

Q6 Suppose $g(0,2) = 15$ and $g(4,1) = 17$.

- **a** What is the distance from $(0,2)$ to $(4,1)$?
- **b** Approximate the rate of change of g at $(0,2)$ travelling toward $(4,1)$.
- **c** If you wanted to express the previous rate of change as an approximation of $D_{\vec{u}}g(0,2)$, what would the unit vector $\vec{u}$ be?

5.4.2

Q7 If $f(x,y) = x^2\sin(xe^y)$, what is $\nabla f(x,y)$?

Q8 If $g(x,y) = \sqrt{6x^2+5y^4}$, what is $\nabla g(x,y)$?

Q9 If $\nabla f(x_0,y_0)$ is orthogonal to $\nabla g(x_0,y_0)$, what can we say about the level curves of f and g? Be specific.

Q10 Harriet says "The gradient vector of f is tangent to the graph of $z = f(x,y)$."
"No," says Marcus, "it is normal to the graph of $z = f(x,y)$." Who is correct?

5.4.3

Q11 Consider our computation of the directional derivative as a dot product.

a. Where did we use the fact that $\vec{u}$ is a unit vector?

b. If $\vec{u}$ were not a unit vector, then $\nabla f \cdot \vec{u}$ would no longer represent rise over run. What would it represent instead?

Q12 Suppose the linearization of $f(x, y)$ at $(-3, 9)$ has the equation

$$L(x, y) = 4 + 2(x + 3) - \frac{1}{3}(y - 9).$$

What is the slope of L from $(-3, 9)$ to $(5, 3)$?

5.4.4

Q13 Given a function $f(x, y)$ and a point (x, y), in what direction $\vec{u}$ is f decreasing fastest? Compute an expression for $\vec{u}$.

Q14 If $D_{\vec{u}}f(x, y) < 0$, what can you say about the directions of $\nabla f(x, y)$ and $\vec{u}$?

Q15 If $f_x(3, 5) = f_y(3, 5)$ in what direction(s) from $(3, 5)$ could f increase most quickly?

Q16 Explain why it makes sense that if $D_{\vec{u}}f(a, b, c) = 0$, then $\vec{u}$ is tangent to the level surface of f through (a, b, c).

Q17 If $f(x, y, z) = 3xy + z^2$, find the unit vector $\vec{u}$ that maximizes $D_{\vec{u}}f(2, 1, -4)$. What is the value of $D_{\vec{u}}f(2, 1, -4)$ for this $\vec{u}$?

Q18 Let $f(x, y) = 2x^2y - 10x - y^2$.

a. What unit vector $\vec{u}$ maximizes the quantity $D_{\vec{u}}f(-1, 3)$?

b. Compute $D_{\vec{u}}f(-1, 3)$ for the $\vec{u}$ you found in part a.

5.4.5

Q19 If $\vec{u} = \left\langle \frac{2}{3}, -\frac{1}{3}, -\frac{2}{3} \right\rangle$ and $f(x,y,z) = xe^{yz}$, compute $D_{\vec{u}} f(3,0,4)$.

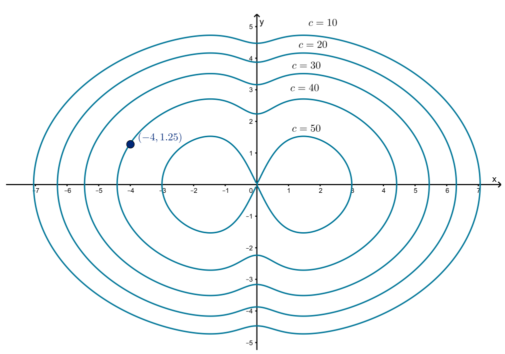
Q20 If $\vec{u} = \left\langle \frac{3}{7}, \frac{6}{7}, -\frac{2}{7} \right\rangle$ and $f(x,y,z) = xy + yz + zx$, compute $D_{\vec{u}} f(7,-7,14)$.

Q21 If $\vec{u}$ is a unit vector in the direction of $\langle 2, 3 \rangle$ and $f(x,y) = x^2 + 3xy + 2$, calculate $D_{\vec{u}} f(-1,4)$.

Q22 Compute the directional derivative of $g(x,y) = e^{x^2 - y}$ at $(3,7)$ in the direction of $\langle -12, 5 \rangle$.

5.4.6

Q23 In this diagram, we have several level sets of $f(x,y)$.

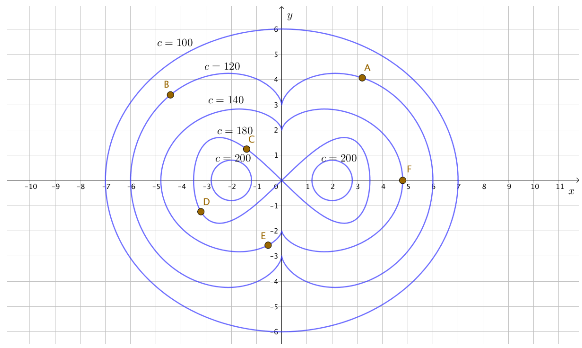


a Which way does $\nabla f(-4, 1.25)$ point?

b Mark all the points (x,y) that satisfy

- $f(x,y) = 30$
- $\nabla f(x,y)$ points in the positive y-direction

Q24 Some level curves of f are drawn below. Indicate the direction of the gradient of f at each labelled point.



5.4.7

Q25 If $\nabla B(x_0, y_0) = \langle 13, -17 \rangle$, would you expect the pixels above (x_0, y_0) to be brighter or dimmer than (x_0, y_0)? Explain.

Q26 The brightness function on the Mona Lisa image ranges from 0 to 255. If we use adjacent points to apporixmate the gradient as in the example, what is the longest gradient vector we could theoretically produce?

5.4.8

Q27 Calculate a normal equation of a tangent line to $x^3 + 8y^3 - 12xy = 0$ at $(3, 1.5)$.

Q28 Let P be a point on the circle $x^2 + y^2 = r^2$. Show that the position vector of P is normal to the tangent line to the circle at P.

Q29 Produce an equation of the tangent plane to $z^3 - xz^2 - yx^2 = 24$ at $(4, -2, 2)$.

Q30 Give an equation of the tangent plane to the graph $z^2x + 2yz - x^2y^2 = 59$ at $(3, 2, 5)$.

Synthesis and Extension

Q31 Suppose $f(x,y)$ is a differentiable function, and we know that for $\vec{u} = \langle -0.6, 0.8\rangle$, $D_{\vec{u}}f(5,-1) = 4$ and for $\vec{v} = \langle 0,-1\rangle$ we know that $D_{\vec{v}}f(5,-1) = -2$. What is $\nabla f(5,-1)$?

Q32 Suppose the point $P = (x_0, y_0, z_0)$ lies on the graph $z = f(x,y)$.

- **a** Give the formula for tangent plane to this graph at P.
- **b** $z = f(x,y)$ is a level surface of $F(x,y,z) = f(x,y) - z$. Use the gradient of F to write the equation of the tangent plane to $F(x,y,z) = 0$ at P.
- **c** Are these equations equivalent? Justify your answer with algebra.

Q33 How could you use the gradient of f to rewrite the formula for the linearization $L(x,y)$ of $f(x,y)$ at (x_0, y_0)?

Q34 Suppose $f(x,y)$ is a differentiable function and $\nabla f(a,b)$ is not the zero vector. How many unit vectors $\vec{u}$ exist such that $D_{\vec{u}}f(a,b) = 0$. How are they related geometrically?

Q35 Suppose $f(x,y,z)$ is a differentiable function and $\nabla f(a,b,c)$ is not the zero vector. How many unit vectors $\vec{u}$ exist such that $D_{\vec{u}}f(a,b,c) = 0$. How are they related geometrically?

Q36 Suppose that $f(x,y,z)$ is a differentiable function, and $f(3,5,-2) = 13$. Suppose further that the vectors $\langle 3,1,0\rangle$ and $\langle 0,2,5\rangle$ both lie in the tangent plane to the surface $f(x,y,z) = 13$ at $(3,5,-2)$. If the maximum value of $D_{\vec{u}}f(3,5,-2)$ is 20, find all possible values of $\nabla f(3,5,-2)$.

Q37 Consider the function $h(x,y) = x^2 + 2x + 4y^{3/2}$

- **a** Compute all possible unit vectors $\vec{u}$ such that $D_{\vec{u}}h(2,3) = 6$
- **b** What angle do these vectors $\vec{u}$ make with the tangent line to the level curve $h(x,y) = 8 + 12\sqrt{3}$ at $(2,3)$.

Q38 Let $f(x,y) = x^4y + 3x - y^3$.

- **a** Give an equation of the level curve of f through the point $(-1,2)$.
- **b** Give an equation of the tangent line to the level curve of f at $(-1,2)$. Write your equation in normal form.

c Give an expression for the linearization of f at $(-1, 2)$.

Section 5.5

The Chain Rule

Goals:

1. Use the **chain rule** to compute derivatives of compositions of functions.
2. Perform implicit differentiation using the chain rule.

Motivational Example

Suppose Jinteki Corporation makes widgets which is sells for $\$100$ each. It commands a small enough portion of the market that its production level does not affect the demand (price) for its products. If W is the number of widgets produced and C is their operating cost, Jinteki's profit is modeled by

$$P = 100W - C$$

The partial derivative $\frac{\partial P}{\partial W} = 100$ does not correctly calculate the effect of increasing production on profit. How can we calculate this correctly?

Question 5.5.1

How Can We Visualize a Composition with a Multivariable Function?

You may recall **parametric equations** from high school algebra. A parametric equation actually consists of two or more equations. Each expresses a variable in our coordinate system in terms of a parameter t.

We can visualize a parametric equation as particle traveling through space.

- The variable t represents time.
- $x(t)$ and $y(t)$ represent the coordinates of the position at time t.
- The vector $\langle x'(t), y'(t)\rangle$ represents **velocity**. It points in the direction of travel.

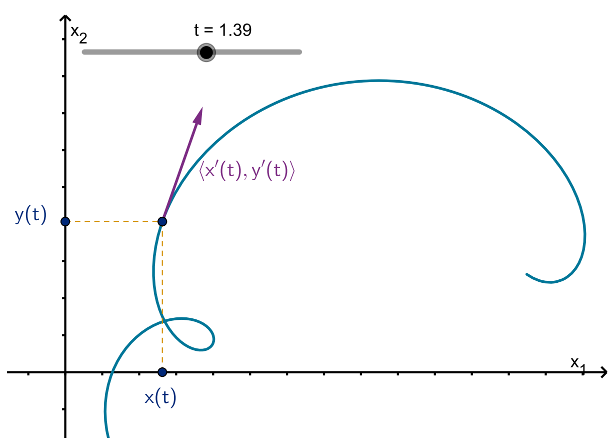


Figure: A particle whose position is defined by $x(t)$ and $y(t)$, the path it follows and its velocity vector

Question 5.5.1 How Can We Visualize a Composition with a Multivariable Function?

Given a function $f(x, y)$ where $x = x(t)$ and $y = y(t)$, we can ask how f changes as t changes. We can visualize this change by drawing the graph $z = f(x, y)$ over the path given by the parametric equations $x(t)$ and $y(t)$.

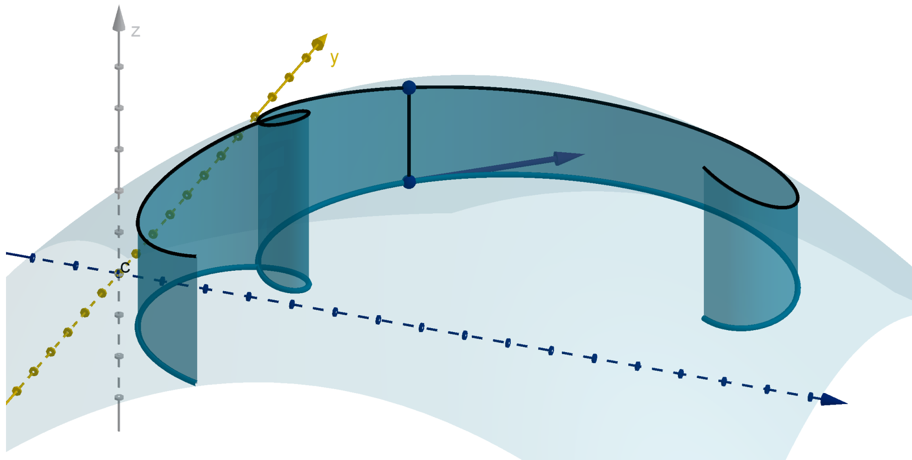


Figure: The composition $f(x(t), y(t))$, represented by the height of $z = f(x, y)$ over the path $(x(t), y(t))$

Question 5.5.2 How Do We Compute the Derivative of a Composition of Functions?

Theorem [The Chain Rule]

Consider a differentiable function $f(x, y)$. If we define $x = x(t)$ and $y = y(t)$, both differential functions, we have

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt}$$

or

$$\frac{df}{dt} = \nabla f(x, y) \cdot \langle x'(t), y'(t) \rangle$$

Remarks

- $f(x(t), y(t))$ is a function (only) of t. Because of this, $\frac{df}{dt}$ is an ordinary derivative, not a partial derivative.
- $\frac{df}{dt}$ is not the slope of the composition graph.

$$\text{slope} = \frac{\text{rise in } z}{\text{run in } xy\text{-plane}}$$

$$\frac{df}{dt} = \frac{\text{rise in } z}{\text{change in } t}$$

- The chain rule is easy to remember because of its similarity to the differential:

$$dz = \frac{\partial z}{\partial x}dx + \frac{\partial z}{\partial y}dy.$$

The proof is more complicated than just sticking a dt under each term.

Example 5.5.3

Using the Chain Rule

If $P = R - C$ and we have $R = 100w$ and $C = 3000 + 70w - 0.1w^2$, calculate $\frac{dP}{dw}$.

Solution

The chain rule says

$$\frac{dP}{dw} = \frac{\partial P}{\partial R}\frac{dR}{dw} + \frac{\partial P}{\partial C}\frac{dC}{dw}$$

We compute the required partial derivatives:

$$\frac{\partial P}{\partial R} = 1 \qquad \frac{\partial P}{\partial C} = -1$$

$$\frac{dR}{dw} = 100 \qquad \frac{dC}{dw} = 70 - 0.2w$$

We plug these into the formula to get

$$\begin{aligned}\frac{dP}{dw} &= (1)(100) + (-1)(70 - 0.2w)\\ &= 30 + 0.2w\end{aligned}$$

Example 5.5.3 Using the Chain Rule

Remark

Notice we don't need the chain rule when we have expressions for each function. We can write the composition ourselves and take an ordinary derivative. In this example we could just differentiate $P = 100w - (3000 + 70w - 0.1w^2)$.

Question 5.5.4
What If We Have More Variables?

The chain rule works just as well if x and y are functions of more than one variable. In this case it computes partial derivatives.

Theorem

If $f(x, y)$, $x(s, t)$ and $y(s, t)$, are all differentiable, then

$$\frac{\partial f}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s}$$

or

$$\frac{\partial f}{\partial s} = \nabla f(x, y) \cdot \left\langle \frac{\partial x}{\partial s}, \frac{\partial y}{\partial s} \right\rangle$$

We can also modify it for functions of more than two variables.

Theorem

Given $f(x, y, z)$, $x(t)$, $y(t)$ and $z(t)$, all differentiable, we have

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt}$$

or

$$\frac{df}{dt} = \nabla f(x, y, z) \cdot \langle x'(t), y'(t), z'(t) \rangle$$

Example 5.5.5

A Composition with More Variables

Recall that for an ideal gas $P(n,T,V) = \frac{nRT}{V}$. R is a constant. n is the number of molecules of gas. T is the temperature in Celsius. V is the volume in meters. Suppose we want to understand the rate at which the pressure changes as an air-tight glass container of gas is heated.

a Apply the chain rule to get an expression for $\frac{dP}{dT}$.

b What is $\frac{dn}{dT}$?

c What is $\frac{dT}{dT}$?

d Suppose that $\frac{dV}{dT} = (5.9 \times 10^{-6})V$. Calculate and simplify the expression you got for $\frac{dP}{dT}$.

Solution

a $\frac{dP}{dT} = \frac{\partial P}{\partial T}\frac{dT}{dT} + \frac{\partial P}{\partial n}\frac{dn}{dT} + \frac{\partial P}{\partial V}\frac{dV}{dT}$

b The container is sealed so no molecules are getting in or out. $\frac{dn}{dT} = 0$.

c If we write T as a function of T, we get $T = T$. $\frac{dT}{dT} = 1$.

d We'll compute the partial derivatives and then plug them into our chain rule expression.

$$\begin{aligned}
\frac{\partial P}{\partial T} &= \frac{nR}{V} \\
\frac{\partial P}{\partial V} &= -\frac{nRT}{V^2} \\
\frac{dP}{dT} &= \frac{nR}{V}(1) + 0 - \frac{nRT}{V^2}(5.9)(10^{-6})V \\
&= \frac{nR(1 - 0.0000059T)}{V}
\end{aligned}$$

Example 5.5.6

A Composition with Limited Information

Suppose $g(p,q,r) = re^{p^2q}$. Given that p, q, r are all differentiable functions of x with the values in the following table, compute $\frac{dg}{dx}$ when $x = 2$.

x	0	1	2	3
$p(x)$	3	1	5	10
$p'(x)$	-3	2	3	4
$q(x)$	6	2	-2	3
$q'(x)$	-1	-5	2	3
$r(x)$	10	11	7	3
$r'(x)$	1	0	-1	-3

Solution

The chain rule says

$$\frac{dg}{dx} = \frac{\partial g}{\partial p}\frac{dp}{dx} + \frac{\partial g}{\partial q}\frac{dq}{dx} + \frac{\partial g}{\partial r}\frac{dr}{dx}$$

We require the partial derivatives of g

$$\begin{aligned}\frac{\partial g}{\partial p} &= 2pqre^{p^2q}\\ \frac{\partial g}{\partial q} &= p^2re^{p^2q}\\ \frac{\partial g}{\partial r} &= e^{p^2q}\end{aligned}$$

Now we plug in the partial derivatives, along with the derivatives of p, q and r from the table.

$$\frac{dg}{dx} = 2pqre^{p^2q}(3) + p^2re^{p^2q}(2) + e^{p^2q}(-1)$$

This is correct, but not sufficiently simplified. We have left p's, q's and r's in the expression, but the table tells us what value these have when $x = 2$. We can make these subsitutions:

$$\begin{aligned}\frac{dg}{dx} &= 2(5)(-2)(7)e^{(5)^2(-2)}(3) + (5)^2(7)e^{(5)^2(-2)}(2) + e^{(5)^2(-2)}(-1)\\ &= -420e^{-50} + 350e^{-50} - e^{-50}\\ &= -71e^{-50}\end{aligned}$$

Application 5.5.7

Implicit Differentiation

Recall that an implicit equation on n variables is a level curve of a n-variable function. Consider the graph $x^3 + y^2 - 4xy = 0$. How can we use this to calculate $\frac{dy}{dx}$ at the point $(3, 3)$?

Solution

First, note that $(3, 3)$ does lie on the graph. When we plug $x = 3$ and $y = 3$ into our equation, we get $27 + 9 - 36 = 0$, which is true. Now suppose that for every x near 3, we can define $y(x)$ to be the y coordinate on the graph $x^3 + y^2 - 4xy = 0$.

Define $F(x, y) = x^3 + y^2 - 4xy$. The points $(x, y(x))$ lie on the graph $F(x, y) = 0$. We can use this equation to obtain an expression for $\frac{dy}{dx}$. When we differentiate $F(x, y(x))$, both components change as x changes, so we cannot use a partial derivative. We need the chain rule.

$$\begin{aligned}
F(x, y(x)) &= 0 \\
\frac{d}{dx}F(x, y(x)) &= \frac{d}{dx}0 && \text{differentiate both sides} \\
\frac{\partial F}{\partial x}\frac{dx}{dx} + \frac{\partial F}{\partial y}\frac{dy}{dx} &= 0 && \text{apply chain rule} \\
\frac{\partial F}{\partial x} + \frac{\partial F}{\partial y}\frac{dy}{dx} &= 0 && \frac{dx}{dx} = 1 \\
\frac{\partial F}{\partial y}\frac{dy}{dx} &= -\frac{\partial F}{\partial x} \text{solve for } \frac{dy}{dx} \\
\frac{dy}{dx} &= -\frac{\frac{\partial F}{\partial x}}{\frac{\partial F}{\partial y}}
\end{aligned}$$

We compute the partial derivatives at $(3, 3)$, then plug them into the formula we derived.

$$\begin{aligned}
F_x(x, y) &= 3x^2 - 4y && F_x(3, 3) = 15 \\
F_y(x, y) &= 2y - 4x && F_y(3, 3) = -6 \\
\frac{dy}{dx} &= -\frac{15}{-6} \\
&= \frac{5}{2}
\end{aligned}$$

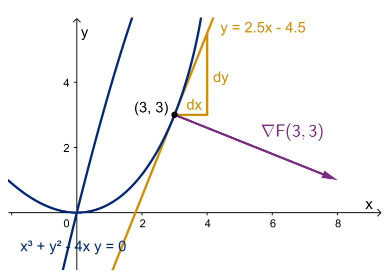


Figure: The graph of $F(x, y) = x^3 + y^2 - 4xy = 0$, its tangent line at $(3, 3)$, and the gradient of F

Application 5.5.7 Implicit Differentiation

Main Ideas

- $\frac{dy}{dx}$ is the slope of the tangent line to $F(x, y) = c$.
- The chain rule allows us to derive $\frac{dy}{dx} = -\frac{F_x}{F_y}$
- $-\frac{F_x}{F_y}$ is the negative reciprocal of $\frac{F_y}{F_x}$, which is the slope of ∇F.

In order to solve for $\frac{dy}{dx}$ we had to assume that y was a differentiable function of x. How do we know that's even true? There is an advanced and powerful theorem that tells us when we can write one variable in an implicit equation as a function of the others. Here is the two-variable version.

Theorem [The Implicit Function Theorem]

Suppose we have a point (x_0, y_0) on the graph of $F(x, y) = c$. Suppose that

1. The partial derivatives of F exist and are continuous at (x_0, y_0)
2. $F_y(x_0, y_0) \neq 0$

Then there is a function $y = f(x)$ that agrees with the graph of $F(x, y) = c$ in some neighborhood around (x_0, y_0). Furthermore

1. f is continuous
2. f is differentiable
3. $f'(x_0) = -\dfrac{F_x(x_0, y_0)}{F_y(x_0, y_0)}$

In the case of our example, the partial derivatives in question are polynomials. As long as $F_y(x_0, y_0) \neq 0$, we are guaranteed that our graph has a tangent line at (x_0, y_0), and its slope is $-\dfrac{F_x(x_0, y_0)}{F_y(x_0, y_0)}$.

Application 5.5.8 Indirect Profit Functions

Suppose a firm chooses how much quantity q to produce, but their profit $\Pi(q, \alpha)$ depends on some parameter α outside their control (maybe a tax or a measure of regulatory burden). The firm, once it knows the value of α, will choose the q that maximizes profit. How will their profit change as α changes?

Solution

The change in the firms profit is $\frac{d\Pi}{d\alpha}$. Since q is also a function of α we will need the chain rule.

$$\frac{d\Pi}{d\alpha} = \frac{\partial \Pi}{\partial q}\frac{dq}{d\alpha} + \frac{\partial \Pi}{\partial \alpha}\frac{d\alpha}{d\alpha}$$

We can substitute $\frac{d\alpha}{d\alpha} = 1$. We can also argue that $\frac{\partial \Pi}{\partial q} = 0$. Why? Because q is the choice that maximizes profit, and maximums occur at critical points. If $\frac{\partial \Pi}{\partial q} > 0$ then the firm could increase q to increase profit (without changing α, which it has no control over). Similarly, If $\frac{\partial \Pi}{\partial q} < 0$ then reducing production would increase profit.

Performing these substitutions gives:

$$\frac{d\Pi}{d\alpha} = \frac{\partial \Pi}{\partial \alpha}$$

This suggests that in this case, the total derivative is equal to the partial derivative.

We can verify this equality graphically as well. Pick a particular α_0 and let $q_0 = q(\alpha_0)$. Notice:

- The graph $\pi(q_0, \alpha)$ is never above $\pi(q(\alpha), \alpha)$ for any α, since $q(\alpha)$ is the optimal choice of q.
- The graphs $\pi(q_0, \alpha)$ and $\pi(q(\alpha), \alpha)$ meet at α_0, since $q_0 = q(\alpha_0)$.
- If two graphs meet but one stays below the other, they are tangent. They have the same tangent line and thus the same derivative.

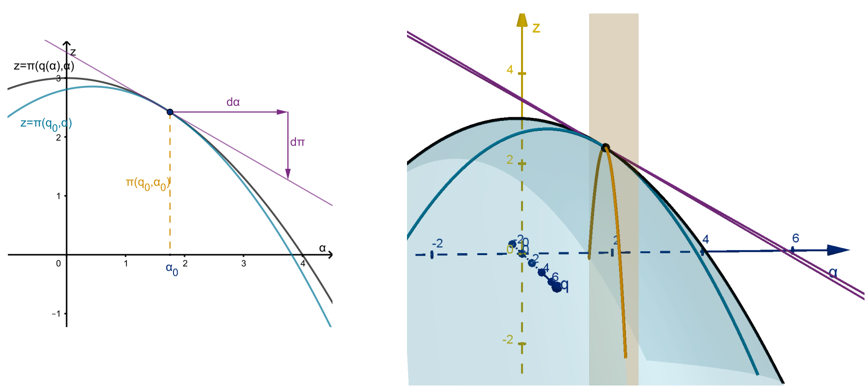


Figure: Two graphs of $z = \Pi(q, \alpha)$, one where q changes to be the optimal choice for each α and one where q is fixed at q_0, the optimal choice for α_0

Application 5.5.8 Indirect Profit Functions

Remark

If we had an expression for $q(\alpha)$ and an expression for Π, we could substitute and use ordinary differentiation. Since we did not, we needed the chain rule. Even with such an expression, to find $\frac{d\Pi}{d\alpha}$ directly we would need to

1. Solve for q as a function of α
2. Substitute $q(\alpha)$ into $\Pi(q, \alpha)$
3. Differentiate the result

Taking a partial derivative is less work. Our result (which economists call the **envelope theorem**) is both a useful abstraction and a computational shortcut.

Section 5.5 Exercises

Summary Questions

Q1 How can we visualize $f(x, y)$, when x and y are functions of t?

Q2 Explain why $\frac{df}{dt}$ cannot be interpreted as a slope of f over the xy-plane.

Q3 What is the difference between $\frac{dz}{dx}$ and $\frac{\partial z}{\partial x}$? How is the first one computed?

Q4 How do you use the chain rule to differentiate implicit functions?

5.5.1

Q5 Plug in a few different t values and plot the corresponding points of

$$x(t) = 3 + 5t$$
$$y(t) = -2 + 4t$$

What is the resulting curve? What is the significance of the t coefficients?

Q6 Consider the curve defined by

$$x(t) = t$$
$$y(t) = e^t$$

a Plot a few points on the curve by plugging in different values of t.

b In general, what curve does

$$x(t) = t$$
$$y(t) = f(t)$$

seem to produce?

Q7 A particle is travelling according to the parametric equations

$$x(t) = 2\cos t$$
$$y(t) = 3\sin t$$

What is the speed (magnitude of velocity) at $t = \frac{\pi}{3}$?

Q8 Produce a tangent vector to the curve defined by

$$x(t) = t^3$$
$$y(t) = t^2$$

at the point $(-27, 9)$.

Q9 Is the graph of

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

the graph of a function? How can you tell without graphing it?

Q10 How are the graphs of the following two parametric equations related? Can you generalize your answer to similar pairs of parametric equations?

$$x(t) = \cos t \qquad\qquad x(t) = \cos(t^3)$$
$$y(t) = \ln t \qquad\qquad y(t) = \ln(t^3)$$

5.5.2

Q11 Let $f(x, y)$ be a funtion. Under what conditions is $\frac{df}{dt}$ equal to the directional derivative of f in the direction of the tangent vector $\langle x'(t), y'(t)\rangle$?

Q12 Liam says "If f is a function of x and y and x and y are increasing, then f is increasing." We all know Liam is incorrect. How could we use the chain rule to refute him?

5.5.3

Q13 The angular speed of an object is given by $\omega = \frac{v}{r}$ where r is the distance from the center of rotation and v is the linear speed. Suppose an object is orbiting earth at a radius of $8400000m$ and a speed of $6900m/s$. If the radius is increasing at a rate of $100m/s$ and the linear speed is decreasing by $60m/s^2$, how quickly is the angular speed changing?

Q14 Let $x = t^2$ and $y = \sin t$. Let $f(x, y) = xy$.

a Compute $\frac{df}{dt}$ using the multivariable chain rule.

b Compute $\frac{df}{dt}$ by substituting and using single-variable differentiation.

c What earlier rule of differentiation can we recover by applying the chain rule to $f(x, y) = xy$?

5.5.4

Q15 Suppose $h(x_1, x_2, x_3, x_4)$ is a four-variable function and each $x_i(x, t)$ is a function of parameters s and t. How would the multivariable chain rule compute $\frac{\partial h}{\partial t}$?

Q16 Suppose $k(x)$ is a function and $x(r, s, t)$ is a function of paramters r, s, and t. How does the multivariable chain rule say we should compute $\frac{\partial k}{\partial r}$?

5.5.5

Q17 Agular momemtum is given by $L = rmv$ where r is the radius of roatation, m is the mass of the object, and v is its linear speed. At a certain time t_0, r is 42 million meters and increasing at $80,000$ meters per second, m is $6000kg$ and not changing, and v is $3100m/s$ and increasing at $20m/s^2$. How quickly is angular momentum increasing?

Q18 Let $f(x, y) = x^2 - y^2$. If $x(r, \theta) = r\cos\theta$ and $y(r, \theta) = r\sin\theta$, compute $\frac{\partial f}{\partial \theta}$ at $(r, \theta) = \left(4, \frac{\pi}{6}\right)$.

5.5.6

Q19 Suppose $x(t)$ and $y(t)$ are differentiable functions of t such that

$$x(2) = 3 \qquad x'(2) = 2 \qquad y(2) = -5 \qquad y'(2) = 10$$

If $f(x, y) = ye^{(x^2y)}$, show how to compute $\frac{df}{dt}$ at $t = 2$.

Q20 Suppose that x and y are functions of t such that when $t = 2$:

$$x = 3 \qquad y = 1$$

$$\frac{dx}{dt} = 5 \quad \frac{dy}{dt} = 2$$

If $g(x, y) = 3xy^2 - x^2 + 2y$, compute $\left.\frac{dg}{dt}\right|_{t=2}$.

5.5.7

Q21 Compute $\frac{dy}{dx}$ at $(4, 2)$, if x and y satisfy $y^3 - xy + x^2 - 4 = 0$

Q22 Compute $\frac{dy}{dx}$ at $(3, 0)$, if x and y satisfy $xe^{xy} = 3$

Q23 What is the slope of the tangent line to $x - y^2 = 9$ at $(18, -3)$?

Q24 Compute the slope of the tangent line to $x^3 = y^2$ at $(4, -8)$.

Q25 Angular momentum is given by $L = rmv$. One law of physics states that angular momentum of an object is conversed (unchanged) unless the a force (besides gravity) acts to speed up or slow down the object. Use the chain rule to derive an expression for $\frac{dv}{dr}$, the amount of linear speed an object gains or loses per unit that its radius of rotation increases. What do you notice about the role of mass in your answer?

Q26 Another principle in physics is the conservation of energy. Kenetic energy is given by $E = \frac{1}{2}mv^2$, where m is the mass and v is the linear speed of the object. Suppose that we have a rock drifiting through space. Suppose it impacts stationary rocks and the combined mass sticks together (without releasing any energy as heat, light or sound). Thus the mass of the total travelling object increases, while the total energy stays the same. Derive an expression for how speed changes per unit of increase in mass.

5.5.8

Q27 Suppose that x is a function of t and that when $t = 9$, we have $x = 7$ and $\frac{dx}{dt} = -3$. Define $f(x, t) = \sqrt{x + t}$.

- **a** Compute the partial derivate $\dfrac{\partial f}{\partial t}(7, 9)$.
- **b** Compute the total derivative $\dfrac{df}{dt}(7, 9)$.
- **c** In a few sentences, explain what these two quantities compute and why they are different from each other.

Q28 A firm with a monopoly produces gets to set the price of its products and decide how much to produce. There is a demand function p such that if the firm produces q units, it must set its price at $p(q)$ to get consumer to buy all of its production. Each unit costs c to produce. The profit function of the firm is

$$\pi(q, c) = p(q)q - cq$$

We can assume that once the firm has worked out what c is, it chooses the q to maximize profit. How much will the firm's actual profit change per unit of increase in c?

Synthesis and Extension

Q29 Find the slope of the tangent line to $x^2 + 2x - y^2 = 8$ at $(5, -3)$ using each of the following two methods.

a Using a gradient vector to write the normal equation of the line and solving for the slope.

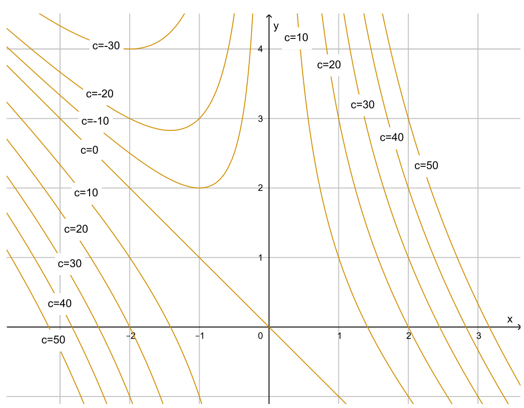
b Using implicit differentiation.

Q30 Suppose the position of a particle at time t is given by

$$x(t) = t^2$$
$$y(t) = 3 - t$$
$$z(t) = \sqrt{t}$$

At $t = 4$, how quickly is particle travelling away from the plane $x + 2y - 2z = 10$?

Q31 Here is a diagram of the level curves of $h(x, y)$ for certain values of c.



a Is $h_y(2, 1)$ positive or negative? Explain in a sentence or two.

b Add a vector to the diagram that indicates the direction of greatest increase of h at $(-2, 0)$.

c Suppose $x = 4 - 5t$ and $y = 3t^2$. Determine, with the aid of a relevant calculation, whether $\frac{dh}{dt}$ is positive or negative at $t = 1$.

Q32 Let $f(x, y) = x^5 + 20xy + 5y^2$.

a Give an equation of the level curve of f through the point $(1, -1)$.

b Give an equation of the tangent plane to $z = f(x, y)$ at the point $(1, -1, -14)$.

c Use the differential of f to estimate how much the z value of $z = f(x, y)$ would change from $(1, -1, -14)$, if x increased by 3 and y decreased by 1. If you don't remember differential notation, you may use another notation for partial credit.

Section 5.6

Maximum and Minimum Values

Goals:

1. Find **critical points** of a function.
2. Test critical points to find local maximums and minimums.
3. Use the **Extreme Value Theorem** to find the global maximum and global minimum of a function over a closed set.

Functions can be used to model a variety of real-world quantities. A company's profit, a disease's infection rate, or the impact of a government program. In these cases, the most pressing question is: what choice of independent variables will maximize or minimize the value of the function? Answering this question was one of the headline applications of single-variable calculus. In this section we will generalize those methods to functions of multiple variables.

Question 5.6.1
What Are Local Extremes?

The **local extremes** of a function are the local minimums and maximums.

Definition

Given an n-variable function $f(x_1, x_2, \ldots, x_n)$ we say that a point P in n-space is

1. a **local maximum** if $f(P) \geq f(Q)$ for all Q in some neighborhood around P.
2. a **local minimum** if $f(P) \leq f(Q)$ for all Q in some neighborhood around P.

Question 5.6.2
Where Do Local Extremes Lie?

At a local maximum (or minimum) $D_{\vec{u}}f$ cannot be positive (or negative) in any direction. Thus at a local extreme, $\nabla f(P) = \vec{0}$, the zero vector. In other words, all the partial derivatives of f are 0 at P.

In the case of a two-variable function, we can visualize this condition. If $f_x(P) \neq 0$, then we could travel in the x direction to increase or decrease f. If $f_x(P) \neq 0$, then we could travel in the y direction to increase or decrease f. Thus at a local maximum or local minimum, the tangent plane must be

Question 5.6.2 Where Do Local Extremes Lie?

horizontal.

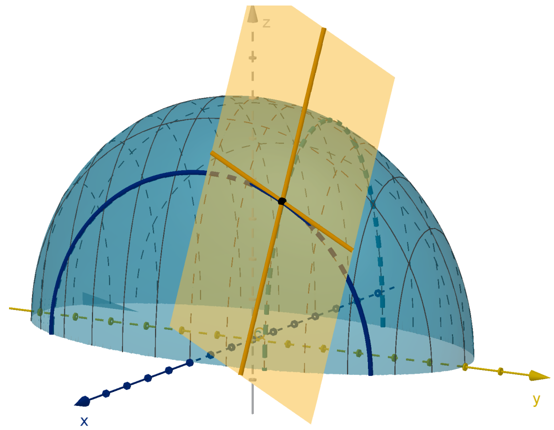


Figure: Tangent lines must have slope 0 at a local max.

This argument works anywhere that ∇f exists. That motivates the following definition:

Definition

We say P is a **critical point** of f if either

1. $\nabla f(P) = \vec{0}$ or
2. $\nabla f(P)$ does not exist (because one of the partial derivatives does not exist).

Theorem

The local maximums and minimums of a function can only occur at critical points.

Example 5.6.3
Finding Critical Points

The function $z = 2x^2 + 4x + y^2 - 6y + 13$ has a minimum value. Find it.

Solution

We know the minimum value exists, so it must lie at a critical point. We compute

$$\nabla f(x,y) = \langle 4x+4, 2y-6 \rangle$$

One type of critical point is where this is undefined, but no value of (x,y) makes these expressions undefined. The other type of critical point occurs when these components are 0. We can solve that system of equations.

$$\begin{aligned} 4x+4 &= 0 & \qquad 2y-6 &= 0 \\ x &= -1 & y &= 3 \end{aligned}$$

The only point that satisfies this requirement is $(-1,3)$. Since there is only one critical point, and the promised minimum lies at a critical point, $(-1,3)$ must be that point. The minimum value is

$$z = (2)(-1)^2 + (4)(-1) + 3^2 - (6)(3) + 13 = 2$$

¿

Question 5.6.4

How Do We Identify Two-Variable Local Maximums and Minimums?

Once we have found a critical point, how do we know whether it is a local minimum, a local maximum or neither? Consider a function $f(x,y)$ and a critical point P. There are two possibilities for $\nabla f(P)$. In the case that $\nabla f(P)$ does not exist, calculus can be no further use to us. If $\nabla f(P) = \langle 0,0 \rangle$, there are a few different shapes the graph could take. Since we are working with two-variables, we can visualize these shapes.

A critical point could be a local maximum. In this case f curves downward in every direction.

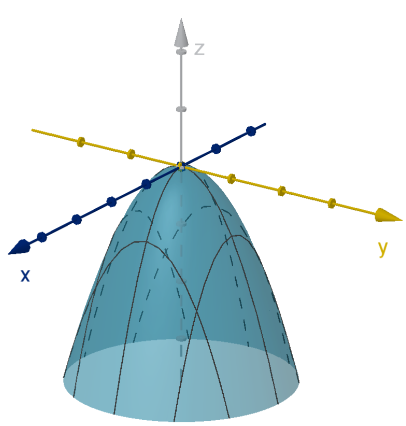


Figure: A local maximum at $(0,0)$

A critical point could be a local minimum. In this case f curves upward in every direction.

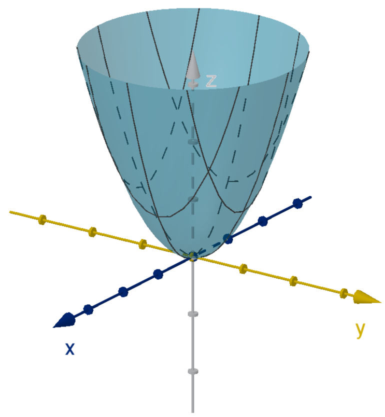


Figure: A local minimum at $(0, 0)$

A critical point could be neither. f curves upward in some directions but downward in others. This configuration is called a **saddle point**.

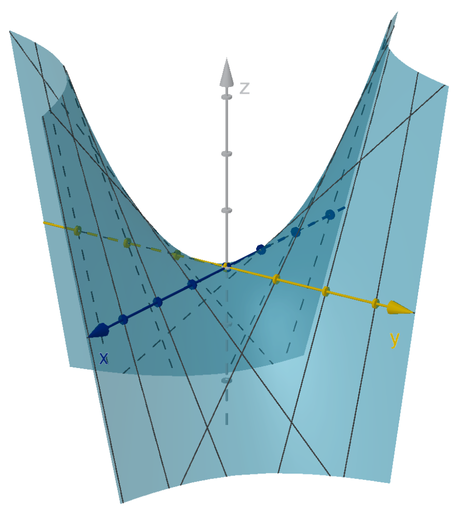


Figure: A saddle point at $(0, 0)$

Curvature is measured by the second derivatives. This matches our experience with single-variable critical points, where the second derivative test classifies critical points as local maximums or local minimums. We have a similar test for two-variable functions, though the computation is more involved.

Theorem [The Second Derivatives Test]

Suppose f is differentiable at (P) and $f_x(P) = f_y(P) = 0$. Then we can compute

$$D = f_{xx}(P)f_{yy}(P) - [f_{xy}(P)]^2$$

1. If $D > 0$ and $f_{xx}(P) > 0$ then P is a local minimum.
2. If $D > 0$ and $f_{xx}(P) < 0$ then P is a local maximum.
3. If $D < 0$ then P is a saddle point.

Unfortunately, if $D = 0$, this test gives no information.

Definition

The quantity D in the second derivatives test is actually the determinant of a matrix called the **Hessian** of f.

$$f_{xx}(P)f_{yy}(P) - [f_{xy}(P)]^2 = \det \underbrace{\begin{bmatrix} f_{xx}(P) & f_{xy}(P) \\ f_{yx}(P) & f_{yy}(P) \end{bmatrix}}_{Hf(P)}$$

Hf follows a logical pattern and can be a useful mnemonic for the second derivatives test.

Example 5.6.5
Classifying a Critical Point

Let $f(x, y) = \cos(2x + y) + xy$

a. Verify that $\nabla f(0, 0) = \langle 0, 0\rangle$.

b. Is $(0, 0)$ a local minimum, a local maximum, or neither?

Example 5.6.5 Classifying a Critical Point

Solution

a

$$\begin{aligned} f_x(x,y) &= -\sin(2x+y)(2)+y && \text{(chain rule)} \\ f_x(0,0) &= -\sin((2)(0)+0)(2)+0=0 \\ f_y(x,y) &= -\sin(2x+y)(1)+x && \text{(chain rule)} \\ f_y(0,0) &= -\sin((2)(0)+0)(1)+0=0 \\ \nabla f(0,0) &= \langle 0,0\rangle \end{aligned}$$

b For the second derivatives test, we need to compute f_{xx}, f_{xy} and f_{yy} at $(0,0)$.

$$\begin{aligned} f_{xx}(x,y) &= -2\cos(2x+y)(2) && \text{(chain rule)} \\ f_{xx}(0,0) &= -2\cos((2)(0)+(0))(2)=-4 \\ f_{xy}(x,y) &= -2\cos(2x+y)(1)+1 && \text{(chain rule)} \\ f_{xy}(0,0) &= -2\cos((2)(0)+(0))(1)+1=-1 \\ f_{yy}(x,y) &= -\cos(2x+y)(1) && \text{(chain rule)} \\ f_{yy}(0,0) &= -\cos((2)(0)+(0))(1)=-1 \end{aligned}$$

$$\begin{aligned} D &= f_{xx}(0,0)f_{yy}(0,0)-[f_{xy}(0,0)]^2 \\ &= (-4)(-1)-(-1)^2 \\ &= 3 \end{aligned}$$

Since $D>0$ and $f_{xx}<0$, $(0,0)$ is a local maximum of f.

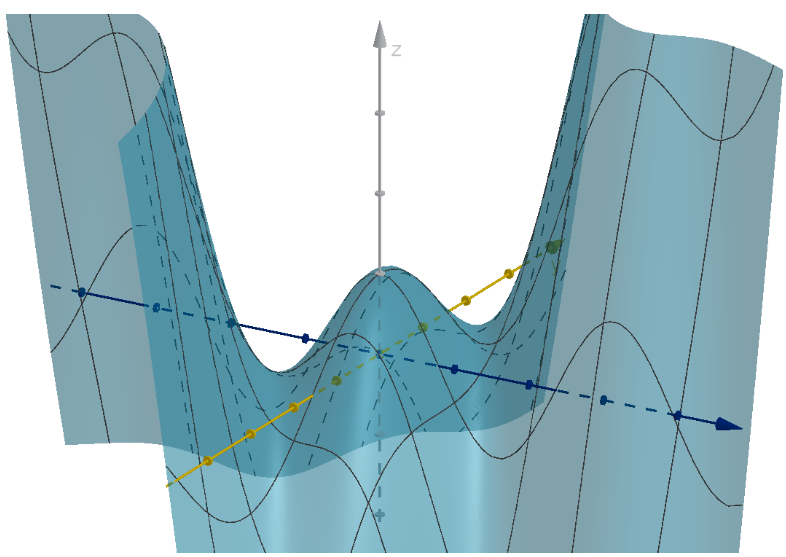


Figure: The graph $z=\cos(2x+y)+xy$ with a local maximum at $(0,0)$

Remark

Why does the final determination between maximum and minimum rely on $f_{xx}(P)$ instead of $f_{yy}(P)$? Actually it doesn't matter which we test. In order for D to be positive, $f_{xx}(P)$ and $f_{yy}(P)$ must have the same sign.

Question 5.6.6
How Do We Find Global Extremes?

The second derivatives test can categorize local extremes, but what about a global extreme?

Definition

Given an n-variable function $f(x_1, x_2, \ldots, x_n)$ we say that a point P in n-space is

1. a **local maximum** if $f(P) \geq f(Q)$ for all Q in the domain of f.
2. a **local minimum** if $f(P) \leq f(Q)$ for all Q in the domain of f.

In a real-world application, we are much more interested in finding global extremes than local ones. Many abstract functions do not even have global extremes. $y = e^x$ has no global maximum. It increases without bound. $y = \frac{1}{x^2}$ has no global minimum. It approaches 0 but never reaches it. The following theorem guarantees that certain functions will have global extremes for us to try to find.

Theorem [The Extreme Value Theorem]

A continuous function f on a closed and bounded domain D has a global maximum and a global minimum somewhere in D.

Two of the words in this theorem have not been defined yet. Here are their definitions.

Definition

Let D be a subset of n-space.

- D is **closed** if it contains all of the points on its boundary.
- D is **bounded** if there is some upper limit to how far its points get from the origin (or any other fixed point). If there are points of D arbitrarily far from the origin, then D is **unbounded**.

For one-variable functions. The EVT requires that the domain be a union of finite, closed intervals (and maybe finitely many isolated points).

Figure: A union of finite, closed intervals

In 2-space, we can get a better sense of what these requirements mean. The boundary of D is the set of points from which you can find points in D and points outside D arbitrarily close by. The boundary of a disc is a circle. If the disc includes the circle, it is closed. If it does not include the circle, it is not closed.

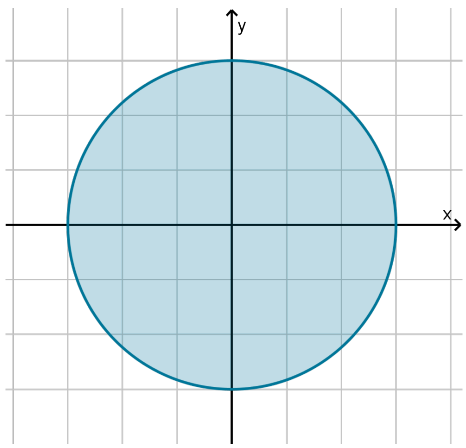


Figure: $x^2 + y^2 \leq 9$ is closed.

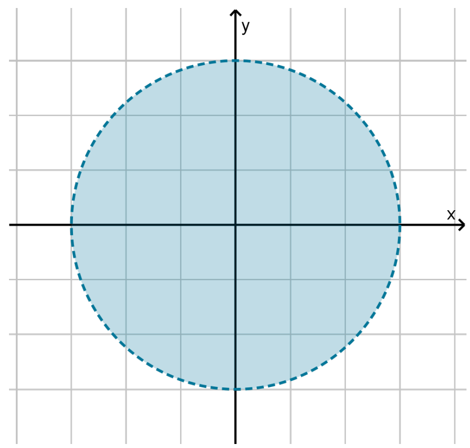


Figure: $x^2 + y^2 < 9$ is not closed.

Containing part of the boundary is not enough. Any missing point means that D is not closed. Even removing an isolated point from the interior of D is a problem. That point is arbitrarily close to points in D. It is also arbitrarily close to a point outside D, itself. Thus it is a boundary point not contained in D, and D is not closed.

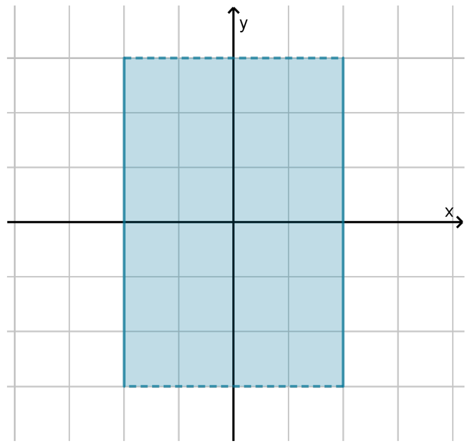


Figure: $-2 \leq x \leq 2$ and $-3 < y < 3$ is not closed.

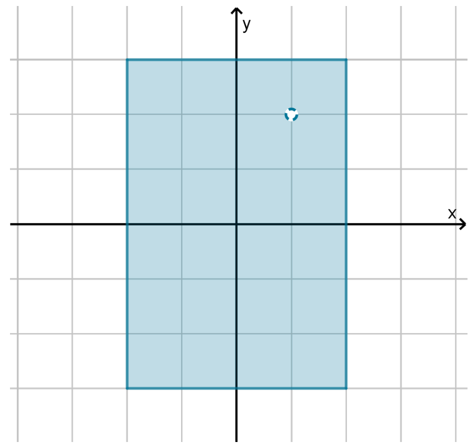


Figure: $-2 \leq x \leq 2$ and $-3 \leq y \leq 3$ and $(x, y) \neq (1, 2)$ is not closed.

Bounded regions are easier to understand. If we can enclose the region in a sufficiently large circle, it is bounded. If it stretches outside any circle we would draw around it, then it is unbounded.

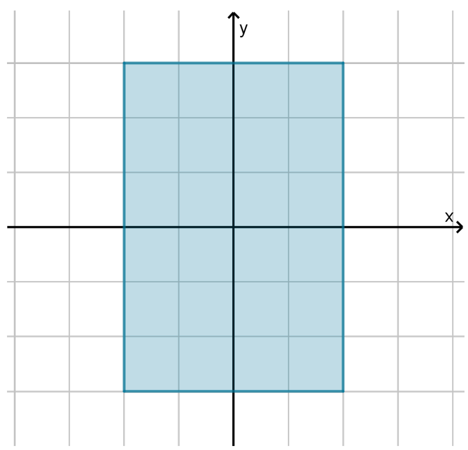


Figure: $-2 \leq x \leq 2$ and $-3 \leq y \leq 3$ is bounded.

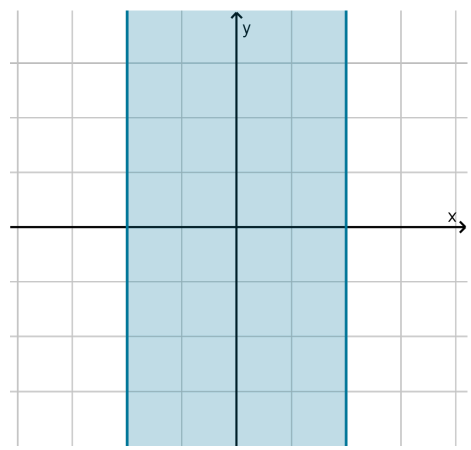


Figure: $-2 \leq x \leq 2$ is unbounded.

Example 5.6.7

Finding a Global Maximum

Consider the function $f(x, y) = x^2 + 2y^2 - x^2y$ on the domain

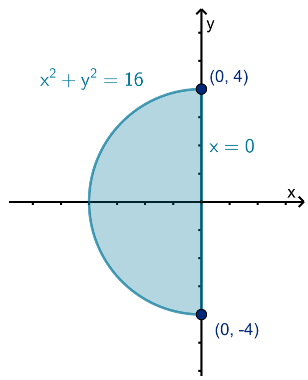
$$D = \{ \underbrace{(x, y)}_{\text{points in } \mathbb{R}^2} : \underbrace{x^2 + y^2 \leq 16, x \leq 0}_{\text{conditions}} \}$$

a Does f have a maximum value on D? How do we know?

b Find the critical points of f.

c Must one of the critical points be the maximum?

d Find the maximum of f.



Remark

The set notation

$$\{\text{type of objects in the set} \ : \ \text{conditions that thoise objects must satisfy}\}$$

is used throughout mathematics, because it is so flexible. It can denote sets of numbers, points, functions, vectors or any other objects.

Solution

a f is a polynomial, so it is continuous. D is a semi-disc that includes its boundary, so it is closed and bounded. The extreme value theorem guarantees that f has a global maximum on D.

b We begin by computing the gradient of f.

$$f_x(x, y) = 2x - 2xy \qquad\qquad f_y(x, y) = 4y - x^2$$

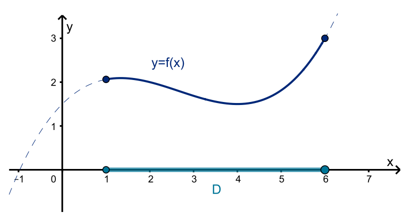
These are never undefined, so there are no critical points of that type. The only critical points will be where both partial derivatives are 0.

$$
\begin{array}{lll}
0 = 2x - 2xy & 0 = 4y - x^2 & \\
0 = 2x(1-y) & & \text{(factor } 2x - 2xy\text{)} \\
x = 0 & \text{or } y = 1 & \\
& & \\
0 = 4y - 0^2 & 0 = 4(1) - x^2 & \text{(examine each case seperately)} \\
0 = y & x = \pm 2 &
\end{array}
$$

We should be careful not to lose track of the logic. The $x = \pm 2$ solution goes with the $y = 1$ case. The $y = 0$ solution goes with the $x = 0$ case. Mixing these up will give invalid solutions. You can always plug in pair of (x, y) to verify they satisfy the system of equations.

We conclude that $(0, 0)$, $(2, 1)$ and $(-2, 1)$ are the critical points, but $(2, 1)$ is not in the domain, so we discard it.

C No. Recall our method for maximizing single variable functions on a closed interval. The maximum can occur at the endpoint of the interval without being detected by the derivative.



The same is true here. If the maximum is on the boundary of D, the gradient need not be 0. In the single-variable case, we only need to test the endpoints (by evaluating f there). There are infinitely many points on the boundary of D. Evaluating f on all of them is not an option. With graphing software we can see that the maximum occurs on the boundary somewhere in the third quadrant, but how can we solve for it exactly?

Example 5.6.7 Finding a Global Maximum

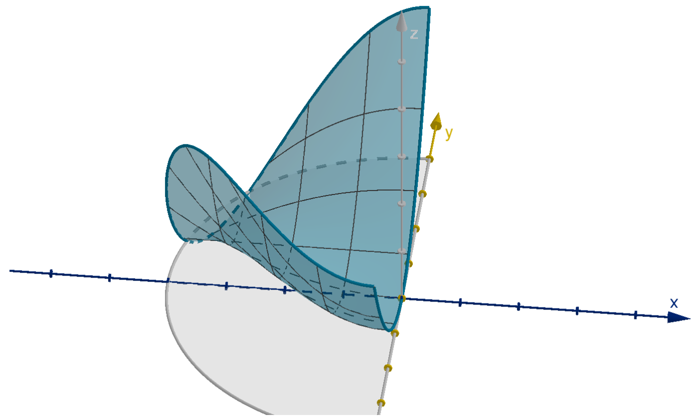


Figure: The graph of $y = f(x, y)$ over the domain D

d To narrow down the search for a maximum on the boundary of D, we will use the boundary equations to write an expression for f that is valid only on the boundary. We can find the critical points of this expression, and rule out any point that is not a critical point.

- Suppose the maximum lies on $x = 0$. The function on $x = 0$ is $f(0, y) = 0^2 + 2y^2 - 0^2 y = 2y^2$. This function only has one variable, so we can find potential maximums by looking for its critical points.

$$f'(y) = 4y$$

This is never undefined. It is 0 at $y = 0$. The only critical point of $f(y)$ on $x = 0$ is $(0, 0)$. However, not all of $x = 0$ is the boundary of D. This component of the boundary ends at $(0, 4)$ and $(0, -4)$. Like with a closed interval, the derivative of $f(y)$ cannot detect a maximum at those endpoints.

- Suppose the maximum lies on $x^2 + y^2 = 16$. On this graph, we can similarly reduce $f(x, y)$ to a function of one variable, but the substitution is more complicated. We solve

$$\begin{aligned}
x^2 + y^2 &= 16 \\
x^2 &= 16 - y^2 \\
f(y) &= (16 - y^2) + 2y^2 - (16 - y^2)y && \text{(substitute for } x^2\text{)} \\
&= y^3 + y^2 - 16y + 16 \\
f'(y) &= 3y^2 + 2y - 16 \\
0 &= 3y^2 + 2y - 16 && \text{(solve for critical points)} \\
0 &= (3y + 8)(y - 2)
\end{aligned}$$

$$\begin{aligned}
y &= -\frac{8}{3} & y &= 2 \\
x^2 + \left(-\frac{8}{3}\right)^2 &= 16 & x^2 + 2^2 &= 16 && \text{(substituue into } x^2 + y^2 = 16\text{)} \\
x^2 &= 16 - \frac{64}{9} & x^2 &= 16 - 4
\end{aligned}$$

$$x = -\sqrt{\frac{80}{9}} \qquad\qquad x = -\sqrt{12} \qquad\qquad (+\text{ solutions are not in } D)$$

Our critical points are $\left(-\sqrt{\frac{80}{9}}, -\frac{8}{3}\right)$ and $\left(-\sqrt{12}, 2\right)$. This component of the boundary also ends at $(0,4)$ and $(0,-4)$, so the maximum might lie there.

We can now argue that one of the points we have found is the maximum.

- If the maximum is not on the boundary, it lies at $(-2,1)$.
- If the maximum is on $x = 0$, then it lies at $(0,0)$, $(0,4)$ or $(0,-4)$.
- If the maximum is on $x^2 + y^2 = 16$, then it lies at $\left(-\sqrt{\frac{80}{9}}, -\frac{8}{3}\right)$, $\left(-\sqrt{12}, 2\right)$, $(0,4)$ or $(0,-4)$.

One of these must be the case. To figure out which it is, we can evaluate f at each point and see which produces the largest value.

- $f(-2,1) = (-2)^2 + 2(1)^2 - (-2)^2(1) = 2$
- $f(0,0) = (0)^2 + 2(0)^2 - (0)^2(0) = 0$
- $f(0,4) = (0)^2 + 2(4)^2 - (0)^2(4) = 32$
- $f(0,-4) = (0)^2 + 2(-4)^2 - (0)^2(-4) = 32$
- $f\left(-\sqrt{\frac{80}{9}}, -\frac{8}{3}\right) = \left(-\sqrt{\frac{80}{9}}\right)^2 + 2\left(-\frac{8}{3}\right)^2 - \left(-\sqrt{\frac{80}{9}}\right)^2\left(-\frac{8}{3}\right) = \frac{1264}{27}$ (maximum)
- $f\left(-\sqrt{12}, 2\right) = (-\sqrt{12})^2 + 2(2)^2 - (-\sqrt{12})^2(2) = -4$

Main Ideas

- If the Extreme Value Theorem applies, then all we need to do is find the critical points and evaluate f at each. One is guaranteed to be the maximum, and one is guaranteed to be the minimum.
- $\nabla f = \vec{0}$ will detect critical points on the interior, but not on the boundary.
- We can rewrite the function on a boundary component using substitution. Set the derivative equal to 0 to find critical points.
- Derivatives will not detect maximums at the endpoints of a boundary curve. These must be included in your set of critical points.

Section 5.6

Exercises

Summary Questions

Q1 Where must the local maximums and minimums of a function occur? Why does this make sense?

Q2 What does the second derivatives test tell us?

Q3 What hypotheses does the Extreme Value Theorem require? What does it tell us?

Q4 Assuming a maximum and minimum exist, where must you look in a domain to be sure you find them?

5.6.1

Q5 Raina claims that $(0,0)$ is the maximum of $f(x,y) = x^2 - y^2 - 10xy$. Disprove her claim without using calculus.

Q6 Is a global maximum also a local maximum? Explain.

Q7 Suppose $g(x,y) = e^{f(x,y)}$. If (a,b) is a local minimum of $f(x,y)$, is it also a local minimum of $g(x,y)$? Explain.

Q8 Does a constant function have any local maximums? Justify your answer with the definition of local maximum.

5.6.2

Q9 Suppose $\nabla f(4,2) = \langle -5, 11 \rangle$. Where would you travel from $(4,2)$ to find higher values of f?

Q10 The function $f(x,y) = |x| + |y|$ has its global minimum at $(0,0)$. Is this a critical point? Explain.

Q11 If (a,b) produces the minimum value of $|\nabla f(x,y)|$, must $(0,0)$ must be a critical point? Explain.

Q12 Suppose $f(x)$ is a function of x with critical points $x = a$ and $x = b$. Suppose $g(y)$ is a function of y with critical points $y = c$ and $y = d$. What are the critical points of $h(x,y) = f(x) + g(y)$?

5.6.3

Q13 Find the critical points of $f(x,y) = x^4 + 4xy + y^4$.

Q14 Find the critical points of $g(x,y) = x^2 + y^2 - 3xy - 13x + 12y$.

5.6.4

Q15 If (x_0, y_0) is critical point and $f_(xx)(x_0, y_0) = 0$, can (x_0, y_0) be a local maximum of f? What must be the value of $f_{xy}(x_0, y_0)$ if so?

Q16 For what values of a does $f(x,y) = x^2 + y^2 + axy$ have a local minimum at the origin?

5.6.5

Q17 Find the critical points of $h(x,y) = x^2y - x^2 - 2y^2$. Classify each as a local maximum, local minimum, or saddle point.

Q18 Find all critical points of $f(x,y) = \frac{1}{3}x^3 - 4xy + 2y^2$. Classify them as local maximums, local minimums, or saddle points.

Q19 Compute the critical points of $f(x,y) = 2x^3 - 12xy + 3y^2$ and classify each as a local maximum, local minimum, or saddle point.

Q20 Let $h(x,y) = x^2 + y^3 + 3xy$. Find the critical points of h, and classify each as a local maximum, local minimum or saddle point.

Q21 Let $f(x,y) = x^3 - 15x^2 - 9x + 12xy - 3y^2 - 18y$. Find the critical points of f and classify each one as local maximum, local minimum or saddle point.

Q22 Let $f(x,y) = x^5 + 20xy + 5y^2$. Find the critical points of f and classify each one as local maximum, local minimum or saddle point.

Q23 Find the critical points of $g(x,y) = e^{x^3+y^2-12x+10y}$. Classify each one as local maximum, local minimum or saddle point.

Q24 Find the critical points of $f(x,y) = \frac{1}{x^4-x^2y+y^2+10}$. Classify each one as local maximum, local minimum or saddle point.

5.6.6

Q25 Draw a sketch of $D = \{(x,y) \ : \ y \geq x^2, y \leq x^3\}$. State whether D is closed and whether D is bounded.

Q26 Draw a sketch of $D = \{(x,y) \ : \ y \geq x, y \leq 2x, xy < 1\}$. State whether D is closed and whether D is bounded.

Q27 Draw a sketch of $D = \{(x,y) \ : \ x > 0, y \geq x^4\}$. State whether D is closed and whether D is bounded.

Q28 Draw a sketch of $D = \{(x, y)\ :\ -1 < x^2 + y^2 \leq 16\}$. State whether D is closed and whether D is bounded.

Q29 Let $D = \{(x, y)\ :\ y \geq x^2\}$. Can the Extreme Value Theorem guarantee that f has a maximum on D? Explain.

Q30 Does the function $f(x, y) = \frac{1}{x^2+y^2}$ have a maximum and minimum value on the domain $D = \{(x, y) : -3 \leq x \leq 3, -4 \leq y \leq 4\}$? If yes, find them. If not, explain why the extreme value theorem does not apply.

5.6.7

Q31 Draw a careful diagram of $D = \{(x, y)\ :\ y \geq x^2, x^2 + y^2 \leq 20\}$. Where would you need to check to guarantee you'd find the maximum value of a continuous function f on D?

Q32 Let $f(x, y)$ be a differentiable function and let

$$D = \{(x, y) : y \geq x^2 - 4, x \geq 0, y \leq 5\}.$$

- **a** Sketch the domain D.
- **b** Does the Extreme Value Theorem guarantee that f has an absolute minimum on D? Explain.
- **c** List all the places you would need to check in order to locate the minimum.

Q33 Find the maximum and minimum value of $f(x, y) = e^{x+3y}$ in the triangle with vertices $(0, 0)$, $(6, 0)$ and $(0, 3)$.

Q34 Find the maximum and minimum value of $f(x, y) = 3x + y$ on D, the closed region bounded by $y = x^2$ and $y = 16$.

Q35 Find the global max and min of $f(x, y) = x^3 - 12x + y^3 - 3y$ on the rectangle $0 \leq x \leq 4$ and $-2 \leq y \leq 2$.

Q36 Consider the function $g(x, y) = \frac{x^4-2x^2+2}{y^2-2y+2}$ on the rectangle $-2 \leq x \leq 2$ and $0 \leq y \leq 3$.

a Does the extreme value theorem apply to this function? Why might you be concerned, and what would you have to check?

b Find the min and max of g.

Synthesis and Extension

Q37 Consider the function $f(x, y) = x^2 - 4xy + 4y^2$.

a Find the critical point(s) of f.

b What does the second derivatives test say about the critical points of f?

c Can you classify the critical points using algebra instead? Explain.

Q38 If $g(x)$ is an increasing function, explain why the local maximums and minimums of any $f(x, y)$ are the same as the maximums and minimums of $g(f(x, y))$.

Section 5.7

Lagrange Multipliers

Goals:

1. Find minimum and maximum values of a function subject to a **constraint**.
2. If necessary, use **Lagrange multipliers**.

Many of the functions we studied do not have maximum values. Polynomials and exponential functions increase without bound. Yet in the real world, we never see corporations producing infinite quantities of goods. We never see infinite populations of animals. Does this mean that polyonomials and exponentials have no real-world applications? On the contrary, they are ubiquitous, but the corporations and populations that opperate under these models also have constraints on their inputs.

Corporations do not have infinite money to invest. Animals do not have infinite food sources. In this section we develop the tools to find maximum and minimum values of a function, when our inputs are constrained.

Question 5.7.1

What Is a Constraint?

Sometimes we aren't interested in the maximum value of $f(x,y)$ over the whole domain, we want to restrict to only those points that satisfy a certain **constraint** equation.

The maximum on the constraint is unlikely to be the same as the unconstrained maximum (where $\nabla f = 0$). Can we still use ∇f to find the maximum on the constraint?

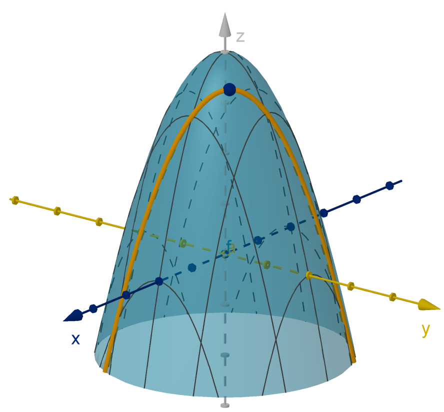


Figure: Maximizing f such that $x+y=1$

We explore this question in the **Maximums on a Constraint** activity.

Question 5.7.2

How Do We Solve a Constrained Optimization?

The method of **Lagrange Multipliers** makes use of the following theorem.

Question 5.7.2 How Do We Solve a Constrained Optimization?

Theorem

Suppose an **objective function** $f(x, y)$ and a **constraint function** $g(x, y)$ are differentiable. The local extremes of $f(x, y)$ given the constraint $g(x, y) = c$ occur where

$$\nabla f = \lambda \nabla g$$

for some number λ, or else where $\nabla g = 0$. The number λ is called a **Lagrange Multiplier**.

This theorem generalizes to functions of more variables.

We can justify the theorem visually by examining the relationship ∇f, ∇g and the constraint. The constraint $g(x, y) = c$ is by definition a level curve of g. It is normal to ∇g.

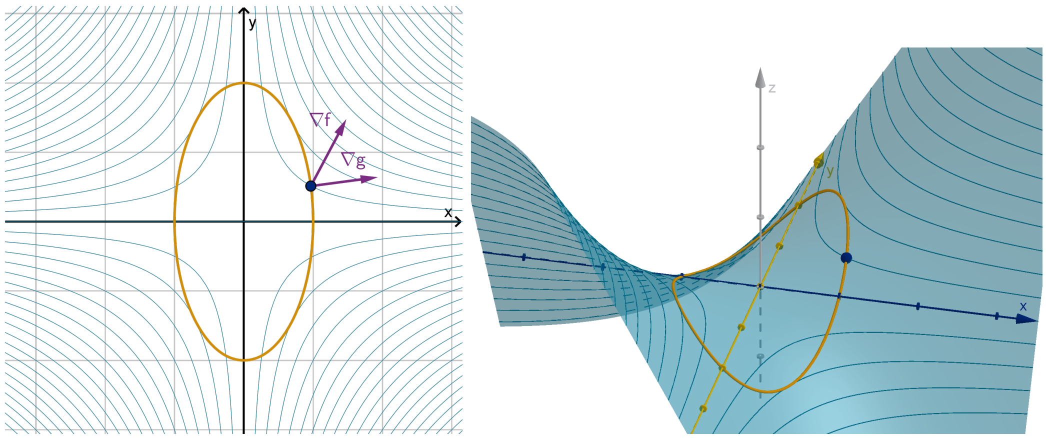


Figure: Where ∇f is not parallel to ∇g, we can travel along $g(x, y) = c$ and increase the value of f. This is because $D_{\vec{u}} f > 0$ for some $\vec{u}$ along the constraint.

By this argument, the only place a maximum or minimum of the objective function can lie of the contraint is where $D_{\vec{u}} f$ would have to be 0, because ∇f is parallel to ∇g.

Remark

When $\nabla f(P)$ is parallel to $\nabla g(P)$ (and neither of these vectors is $\vec{0}$), the level curves of f through P is tangent to the level curve $g(x, y) = c$. If we can draw the level curves of f, this gives us a visual method of identifying the potential maximums and minimums.

Example 5.7.3
The Maximum on a Curve

Find the point(s) on the ellipse $4x^2 + y^2 = 4$ on which the function $f(x, y) = xy$ is maximized.

The EVT and constraints

Are we guaranteed that a maximum exists at all? The Extreme Value Theorem can still be applied to constraints. Here are a few ways we can identify that a constraint is closed:

1. A curve is closed if it includes its endpoints (or none exist).
2. A surface is closed if it includes its boundary (or none exists).
3. The level set of a continuous function is always closed.

Even armed with these, we still need to check that the domain is bounded.

Solution

We'll check the conditions of the Extreme Value Theorem

1. $4x^2 + y^2 = 4$ is a curve with no endpoints, so it is closed.
2. $4x^2 + y^2 = 4$ is an ellipse. It stays within a bounded distance from the origin.
3. f is continuous.

By the Extreme Value Theorem, we know that a maximum exists. We will use Lagrange multipliers to narrow down our search to the possible maximums. We set $g(x, y) = 4x^2 + y^2$ and compute the gradient vectors of f and g.

$$\nabla f(x, y) = \langle y, x\rangle \qquad \nabla g(x, y) = \langle 8x, 2y\rangle$$

The theorem allows two possibilities at a maximum.

1. $\nabla g(x, y) = \langle 0, 0\rangle$. The only (x, y) that satisfies this is $(0, 0)$. But $(0, 0)$ is not on the constraint, so it is not a valid solution.
2. $\nabla f = \lambda \nabla g$. We can factor the λ across each component of the vectors, but that gives us two equations and three variables (x ,y and λ). We need another equation, and fortunately we have one. x and y must satisfy $4x^2 + y^2 = 4$ as well. Here is one (but not the only) way to solve this system of equations.

Example 5.7.3 The Maximum on a Curve

$$y = \lambda 8x \qquad x = \lambda 2y \qquad 4x^2 + y^2 = 4$$

$$\begin{aligned}
y &= \lambda 8(\lambda 2y) \\
0 &= \lambda^2 16y - y \\
0 &= y(4\lambda - 1)(4\lambda + 1)
\end{aligned}$$

either $0 = y$ $\qquad x = \lambda 2(0) \qquad 4(0)^2 + 0^2 = 4$ (no solution)

or $\lambda = \pm\frac{1}{4}$

$$\begin{aligned}
y &= \left(\pm\frac{1}{4}\right) 8x \\
y &= \pm 2x
\end{aligned}$$

$$\begin{aligned}
4x^2 + (\pm 2x)^2 &= 4 \\
8x^2 &= 4 \\
x^2 &= \frac{1}{2} \\
x &= \pm\frac{1}{\sqrt{2}}
\end{aligned}$$

$$\begin{aligned}
y &= \pm 2\left(\frac{1}{\sqrt{2}}\right) \\
y &= \pm\sqrt{2}
\end{aligned}$$

This tells us the only possible locations for the maximum are:

$$(x, y) = \left(\pm\frac{1}{\sqrt{2}}, \pm\sqrt{2}\right)$$

We identify the maximum by evaluating f at each point.

$$f\left(\frac{1}{\sqrt{2}}, \sqrt{2}\right) = 1 \qquad f\left(-\frac{1}{\sqrt{2}}, \sqrt{2}\right) = -1$$

$$f\left(-\frac{1}{\sqrt{2}}, -\sqrt{2}\right) = 1 \qquad f\left(\frac{1}{\sqrt{2}}, -\sqrt{2}\right) = -1$$

We conclude that the maximum occurs at $\left(\frac{1}{\sqrt{2}}, \sqrt{2}\right)$ and $\left(-\frac{1}{\sqrt{2}}, -\sqrt{2}\right)$.

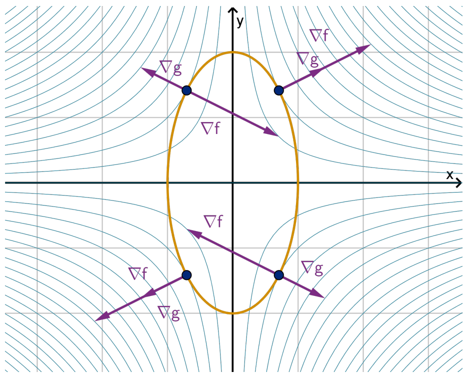


Figure: The four points that satisfy $\nabla f = \lambda \nabla g$ and $g(x, y) = c$.

Main Idea

The level set of a continuous (constraint) function is always closed. If it is also bounded and the objective function is differentiable, then one of the points produced by Lagrange multipliers will be the global maximum and one will be the global minimum of the constrained optimization.

Example 5.7.4
The Maximum on a Surface

Find the maximum value of the function $f(x, y, z) = x^4 y^4 z$ on the sphere $x^2 + y^2 + z^2 = 36$.

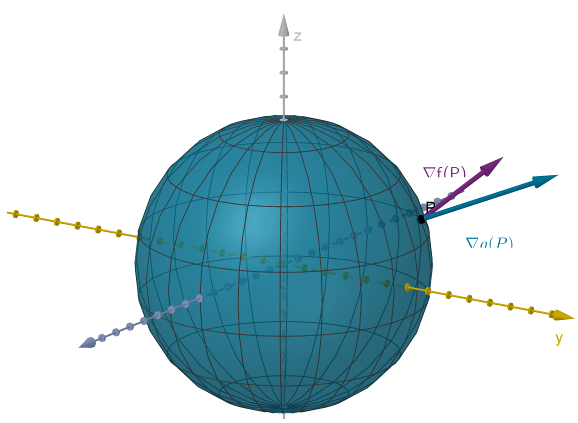


Figure: The gradient vector and level surface of a constraint function and the gradient vector of the objective function

Example 5.7.4 The Maximum on a Surface

Solution

First note that the EVT applies, since a sphere is closed and bounded and f is continuous. To identify potential maximums, we appeal to Lagrange multipliers.

Set $g(x, y, z) = x^2 + y^2 + z^2$. Then $\nabla g(x, y, z) = \langle 2x, 2y, 2z \rangle$. The case $\nabla g(x, y, z) = \vec{0}$ only occurs at the origin, which is not on the sphere. The critical points must be only the points where $\nabla f = \lambda \nabla g$.

$$\nabla f(x, y, z) = \left\langle 4x^3y^4z, 4x^4y^3z, x^4y^4 \right\rangle .$$

Equating each coordinate gives us three equations, and the constraint is a fourth. We thus have a system of four equations and four variables.

$$4x^3y^4z = \lambda 2x \qquad 4x^4y^3z = \lambda 2y \qquad x^4y^4z = \lambda 2z \qquad x^2 + y^2 + z^2 = 36$$

The most obvious way to solve this algebraically is to solve for λ, but this requires us to divide by x, y and z. We would need to remember that another possible solution is that x, y or z is 0. We can avoid this by multiplying and factoring instead.

$$\begin{aligned} 4x^3y^4z &= \lambda 2x & 4x^4y^3z &= \lambda 2y & x^4y^4 &= \lambda 2z \\ 4x^3y^5z^2 &= \lambda 2xyz & 4x^5y^3z^2 &= \lambda 2xyz & x^5y^5 &= \lambda 2xyz \\ 4x^3y^5z^2 &= 4x^5y^3z^2 & & & x^5y^5 &= 4x^5y^3z^2 \end{aligned}$$

$$\begin{aligned} 4x^3y^5z^2 - 4x^5y^3z^2 &= 0 & x^5y^5 - 4x^5y^3z^2 &= 0 \\ 4x^3y^3z^2(y - x)(y + x) &= 0 & x^5y^3(y - 2z)(y + 2z) &= 0 \end{aligned}$$

$$\begin{aligned} \text{either } x &= 0 \\ \text{or } y &= 0 \\ \text{or } y &= \pm x \qquad \text{and} \qquad y = \pm 2z \\ \pm 2z &= x \end{aligned}$$

$$\begin{aligned} x^2 + y^2 + z^2 &= 36 \\ (\pm 2z)^2 + (\pm 2z)^2 + z^2 &= 36 \\ 9z^2 &= 36 \\ z &= \pm 2 \end{aligned}$$

$$\begin{aligned} (\pm 2)(\pm 2) &= x & y &= (\pm 2)(\pm 2) \\ \pm 4 &= x & y &= \pm 4 \end{aligned}$$

This gives us 8 critical points: $(\pm 4, \pm 4, \pm 2)$. In addition every point in the $x = 0$ cross section of the sphere is a critical point, as is every point in the $y = 0$ cross-section. This is infinitely many points to evaluate, but fortunately the algebra of our objective function allows us to evaluate these points in large batches.

$$\begin{aligned} \text{if } x = 0 \quad f(x, y, z) &= 0^4y^4z = 0 \\ \text{if } y = 0 \quad f(x, y, z) &= x^40^4z = 0 \\ f(\pm 4, \pm 4, 2) &= (\pm 4)^4(\pm 4)^4(2) = 2^{17} \\ f(\pm 4, \pm 4, -2) &= (\pm 4)^4(\pm 4)^4(-2) = -2^{17} \end{aligned}$$

Thus the maximum value is 2^{17}. It occurs at the four points $(\pm 4, \pm 4, 2)$.

Remark

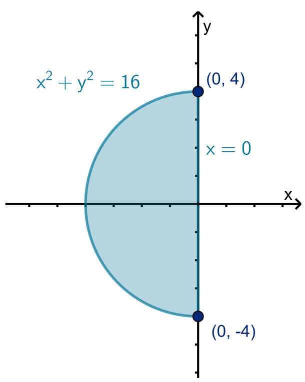
If we hadn't seen how to avoid dividing by x, y and z, we could have gone ahead and done the division. Remember that when you divide while solving an equation, you obtain an extra solution where the divisor is 0. This would lead us to check $x = 0$, $y = 0$ and $z = 0$ as we did in the factoring solution.

Synthesis 5.7.5

Using the Extreme Value Theorem and Lagrange Multipliers

How can Lagrange multipliers help us find the maximum of $f(x, y) = x^2 + 2y^2 - x^2y$ on the domain

$$D = \{(x, y) : x^2 + y^2 \leq 16, x \leq 0\}?$$



Solution

We can continue Example 7. After finding the critical points of f at $(0, 0)$ and $(-2, 1)$, we turn to the boundaries. The boundaries are level curves.

- For $x^2 + y^2 = 16$, set $g(x, y) = x^2 + y^2 = 16$. We have

$$\nabla f(x, y) = \left\langle 2x - 2xy, 4y - x^2 \right\rangle \qquad \nabla g(x, y) = \langle 2x, 2y \rangle$$

  $\nabla g(x, y) = \vec{0}$ only at the origin, which isn't on the constraint. So we solve $\nabla f(x, y) = \lambda \nabla g(x, y)$ and $g(x, y) = 4$.

$$2x - 2xy = \lambda 2x \qquad 4y - x^2 = \lambda 2y \qquad x^2 + y^2 = 16$$

$$2x - 2xy - 2\lambda x = 0$$

$$2x(1 - y - \lambda) = 0$$

$$\text{if } x = 0 \qquad 0^2 + y^2 = 16$$

$$y = \pm 4$$

$$\text{if } 1 - y - \lambda = 0$$

$$\lambda = 1 - y \qquad 4y - x^2 = (1-y)2y$$

$$2y^2 + 2y = x^2$$

$$(2y^2 + 2y) + y^2 = 16$$

$$3y^2 + 2y - 16 = 0$$

$$(3y + 8)(y - 2) = 0$$

$$\text{if } y = -\frac{8}{3} \qquad \text{if } y = 2$$

$$x^2 + \left(-\frac{8}{3}\right)^2 = 16 \qquad x^2 + 2^2 = 16$$

$$x^2 + \frac{64}{9} = \frac{144}{9} \qquad x^2 = 12$$

$$x^2 = \frac{80}{9} \qquad = x = \pm\sqrt{12}$$

$$x = \pm\sqrt{\frac{80}{9}}$$

The critical points are $(0, \pm 4)$, $\left(-\sqrt{12}, 2\right)$ and $\left(-\sqrt{\frac{80}{9}}, -\frac{8}{3}\right)$. The solutions with positive x are not in D.

- On $x = 0$, substitution is probably the easier choice, but Lagrange multipliers are still possible. $x = 0$ is a level set of the function $g(x, y) = x$.

$$\nabla g(x, y) = \langle 1, 0 \rangle$$

$\nabla g \neq \vec{0}$ so we solve $\nabla f(x, y) = \lambda \nabla g(x, y)$.

$$2x - 2xy = \lambda \qquad 4y - x^2 = 0 \qquad x = 0$$

$$4y = 0$$

This is the same equation we obtained by substituting $x = 0$ into f and differentiating.

Main Idea

To find the absolute minimum and maximum of a differentiable function $f(x, y)$ over a closed and bounded domain D:

1. Compute ∇f and find the critical points inside D.
2. Identify the boundary components. Find the critical points on each using substitution **or Lagrange multipliers**.
3. Identify the endpoints (intersections) of the boundary components.
4. Evaluate $f(x, y)$ at all of the above. The minimum is the lowest number, the maximum is the highest.

Synthesis 5.7.6
The Gradient on the Boundary

Suppose P is a critical point of f on a boundary component of a domain D. What does the direction of $\nabla f(P)$ tell us about whether P is a maximum or minimum?

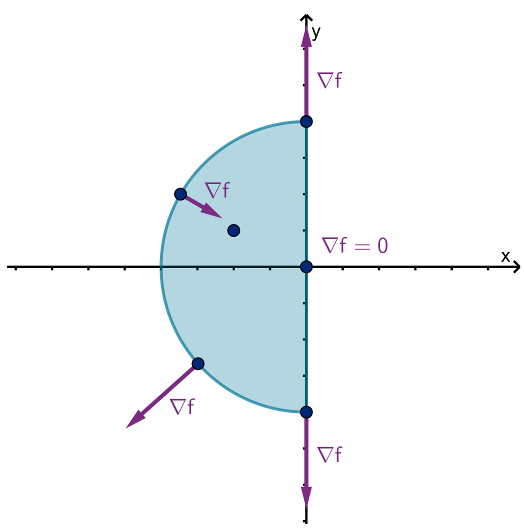


Figure: The critical points and gradient vectors of $f(x, y)$ on a closed and bounded domain

Solution

First suppose $\nabla f(P)$ points into D. Then f increases as we travel into D. Thus P cannot be a local maximum.

Synthesis 5.7.6 The Gradient on the Boundary

P may be a local minimum but may not be. The directional derivative along the boundary is 0, so f could curve upward or downward along the boundary. If f curves downward we could find lower values of f nearby and P would not be a minimum. If f curves upward, then P would be a minimum. We could compute this curvature by taking the substituted version of f that we used to solve for P and computing its second derivative at P.

On the other hand, if we suppose that $\nabla f(P)$ points out of D, then D decreases as we travel into D, and P cannot be a local minimum. It may or may not be a local maximum.

Question 5.7.7
Can This Lagrange Apply to More Than One Constraint?

If we have two constraints in three-space, $g(x, y, z) = c$ and $h(x, y, z) = d$, then their intersection is generally a curve.

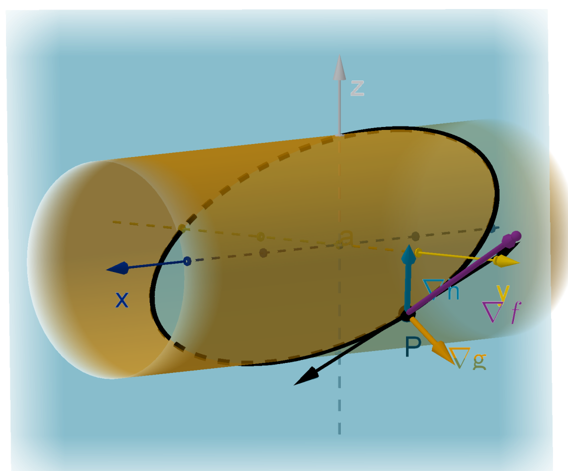


Figure: The intersection of the constraints $g(x, y, z) = c$ and $h(x, y, z) = d$

According to our earlier argument about directional derivatives, at a maximum P on the constraint, $\nabla f(P)$ must be normal to the constraint. There are more ways for this to happen with two constraint equations.

1. $\nabla f(P)$ could be parallel to $\nabla g(P)$.
2. $\nabla f(P)$ could be parallel to $\nabla h(P)$.
3. $\nabla f(P)$ could be the vector sum of a vector parallel to $\nabla g(P)$ and a vector parallel to $\nabla h(P)$.

You should look at Figure 380 to convince yourself that these $\nabla f(P)$ would all be normal to the constraint. We can express this condition algebraically

Theorem

If $f(x,y,z)$ is a differentiable function and $g(x,y,z) = c$ and $h(x,y,z) = d$ are two constraints. If P is a maximum of $f(x,y,z)$ among the points that satisfy these constraints then either

$$\nabla f(P) = \lambda \nabla g(P) + \mu \nabla h(P)$$

for some scalars λ and μ, or $\nabla g(P)$ and $\nabla h(P)$ are parallel.

This system of equations is usually difficult to solve by hand.

Remark

You can check the reasonableness of this method by noting that it gives us a system of 5 variables, x, y, z, λ, μ, and five equations:

$$\begin{aligned} f_x(x,y,z) &= \lambda g_x(x,y,z) + \mu h_x(x,y,z) \qquad & g(x,y,z) &= c \\ f_y(x,y,z) &= \lambda g_y(x,y,z) + \mu h_y(x,y,z) & h(x,y,z) &= d \\ f_z(x,y,z) &= \lambda g_z(x,y,z) + \mu h_z(x,y,z) \end{aligned}$$

We therefore generally expect this system to have a finite number of solutions, though there are plenty of counterexamples to this expectation.

Section 5.7 Exercises

Summary Questions

Q1 What is a constraint?

Q2 What equations do you write when you apply the method of Lagrange multipliers?

Q3 Is the set of points that satisfies a constraint closed and bounded? Explain.

Q4 How does a constraint arise when finding the maximum over a closed and bounded domain?

5.7.1

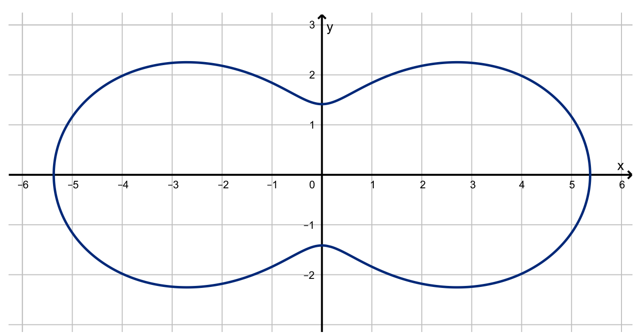
Q5 Suppose we have $\$230$ to spend on three goods. Good 1 costs $\$13$ per unit. Good 2 costs $\$22$ per unit. Good 3 costs $\$11$ per unit. Write a budget constraint that expresses what purchases (x, y, z) of good 1, good 2 and good 3 are possible, if you spend you budget.

Q6 Suppose he maximum value of $f(x, y)$ occurs at $(3, -4)$. Where is the maximum value of $f(x, y)$ that satisfies the constraint $x^2 + y^2 = 25$? Explain.

5.7.2

Q7 Suppose $f(x, y, z)$ is a smooth function. Suppose the maximum value of f on the sphere $x^2 + y^2 + z^2 = 25$ occurs at P. What can you say about $\nabla f(P)$ and the tangent plane to the sphere at P?

Q8 Suppose the curve below is the graph of $g(x, y) = k$. Use methods from calculus to find and mark the approximate location of the point that maximizes the function $f(x, y) = 3y - x$ subject to the constraint $g(x, y) = k$. Justify your reasoning in a few sentences.



Q9 Suppose that (a, b) is a local maximum of the smooth function $f(x, y)$ which also happens to satisfy the constraint $g(a, b) = k$.

a Is (a, b) also a local maximum of f among the points on the constraint? Explain.

b If we used Lagrange multipliers to detect (a, b), what would we expect λ to be equal to at that point?

Q10 Show that $(3,3)$ is not a local maximum of $f(x,y) = 2x^2 - 4xy + y^2 - 8x$ on the graph $x^3 + y^3 = 6xy$.

5.7.3

Q11 Compute the maximum value of $y - x^2$ on the constraint $x^2 + y^2 = 4$.

Q12 Refer to your "Maximums on a Constraint" worksheet.

- **a** What system of equations would you set up to find the critical points of f on the constraint $p(x,y) = c$?
- **b** Can you solve it?
- **c** Which was easier, using Lagrange or using substitution?

5.7.4

Q13 Find the maximum value of $f(x,y,z) = xyz$ on the sphere $x^2 + y^2 + z^2 = 36$.

Q14 Find the maximum value of $f(x,y,z) = xz$ on the sphere $x^2 + y^2 + z^2 = 36$.

Q15 Find the maximum value of $f(x,y,z) = 3y + 2z$ on the ellipsoid $25x^2 + y^2 + 4z^2 = 100$.

Q16 Find the minimum value of $h(x,y,z) = x^2 + y^2 + z^2$ on the plane $3x + 5y - 2z = 30$.

5.7.5

Q17 Suppose $f(x, y)$ is differentiable but has no critical points. Will the method of Lagrange multipliers detect the maximum value of f in $D = \{(x, y) \ : \ x^2 + y^2 \leq 49\}$? Explain.

Q18 Consider the following two questions:

- Find the maximum value of $f(x, y)$ that satisfies $x^2 + y^2 \leq 9$.
- Find the maximum value of $f(x, y)$ that satisfies $x^2 + y^2 = 9$.

a How are the questions different?

b Which question takes less work to solve? Explain how you know.

c Do solutions exist to both questions? What additional information would guarantee that they do?

Q19 Let $D = \{(x, y) \ : \ x^2 + y^2 \leq 1, x \geq 0, y \leq 0\}$. Find the maximum and minimum values of $f(x, y) = x^2 - y$ on D.

Q20 Consider the function $f(x, y) = x^2 + 6xy + 9y^2 + 5$. Find the maximum and minimum values of f on the domain $D = \{(x, y) : y \geq x, x \geq 0, x^2 + y^2 \leq 10\}$

Q21 Let $D = \{(x, y) \ : \ x^2 + y^2 \leq 20, y \geq -x\}$. Find the maximum and minimum values of $f(x, y) = x^4 y$ on D.

Q22 Let $D = \{(x, y) \ : \ x^2 + y^2 \leq 25, y \geq x + 1, y \geq 0\}$. Find the maximum and minimum values of $f(x, y) = x^3 y^2$ on D.

Q23 Let $D = \{(x, y) \ : \ x^2 + y^2 \leq 20, y \geq -x\}$. Find the maximum and minimum values of $f(x, y) = x^4 y$ on D.

Q24 Let $D = \left\{(x, y) \ : \ \dfrac{x^2}{16} + \dfrac{y^2}{64} \leq 1, x \geq 0\right\}$. Find the points in D that obtain the maximum and minimum values of $f(x, y) = 2x + 3y$.

5.7.6

Q25 Suppose the maximum of $f(x, y)$ on

$$D = \{(x, y) \mid g(x, y) \leq c\}$$

occurs at P on the boundary of D. We know that $\nabla f(P)$ points out of D. What does this tell us about the sign of λ?

Q26 Explain why knowing which way ∇f points is not useful for ruling out potential maximums given a domain of the form $g(x, y) = c$.

5.7.7

Q27 How does the method of Lagrange multipliers suggest we solve for the maximum value of $f(x, y)$ on the constraints $x + y = 1$ and $x - y = 0$? Do we need to know what f is to solve this? Why shouldn't that bother us?

Q28 Write a system of equations that one would solve to find the maximum and minimum values of $f(x, y, z) = x$ on the two constraints $y^2 + z^2 = 25$ and $x + y + z = 1$.

Synthesis and Extension

Q29 Consider the plane p with normal equation $7x + 6y - 3z - 42 = 0$

- **a** Use Lagrange multipliers to find the point A on p that s closest to the origin O.
- **b** Show that $\overrightarrow{OA}$ is a normal vector to p.
- **c** Show how you can use the observation in **b** to solve for the closest point (A) without using calculus.

Q30 Determine the smallest rectangle (parallel to the x and y axes) that contains the ellipse $x^2 + 3xy + 4y^2 - 4x - 13y + 4 = 0$.

Q31 An aquarium with an open top has volume $20m^3$. Its rectangular base is made of slate, and its sides are made of glass. Slate costs five times as much (per unit area) as glass. Set up and solve a constrained onstrained optimization problem to find the dimensions (ℓ, w, h) of the aquarium that will minimize the cost of materials.

Q32 Let D be the region enclosed by $2x + y = 8$, $y = 8$ and $x = 4$. Consider the function

$$f(x, y) = xy - 3y - 6x.$$

a Does f have a maximum and minimum value on D? What tool can you use to verify this? What did you need to check before applying this tool?

b Find the maximum and minimum values of f on D. Demonstrate in your work that you've checked all the relevant places for potential maximums.

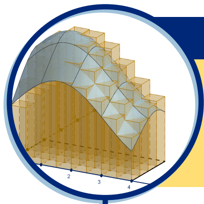
Q33 Find the maximum and minimum values of $f(x, y) = 2x^2 + 2xy + 5y^2$ on the ellipse $x^2 + 4y^2 = 106$.

Chapter 6

Multivariable Integration

This chapter introduces integration of functions of more than one variable. It also introduces joint probability distributions as an application.

Contents



Section 6.1

Double Integrals

Goals:

1. Approximate the volume under a graph by adding prisms.
2. Calculate the volume under a graph using a double integral.

In single-variable calculus, the definite integral computes a total change from a rate of change. Moreover, it also solves a geometry problem. This connection means that we can use geometric intuition to understand integrals better. Integrals of multi-variable functions also allow us to aggregate rates into totals. To build our intuition, we begin with the geometric problem that they solve.

Question 6.1.1

How Do We Approximate the Volume Under $z = f(x, y)$?

Begin by remembering our construction of the single-variable definite integral. We approximated the area under the graph $y = f(x)$ by rectangles. Smaller rectangles give a better approximation, and we defined the limit of these approximations to be the **definite integral**.

$$\int_a^b f(x)dx = \lim_{\Delta x \to 0} \sum_{i=1}^{n} f(x_i^*)\Delta x$$

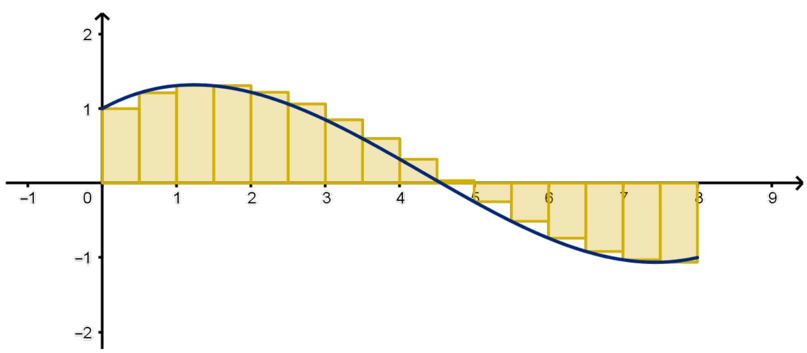


Figure: The area under $y = f(x)$ approximated by rectangles

Consider a function $f(x, y)$ and a rectangle $D = \{(x, y) \ : \ 0 \leq x \leq 4, 0 \leq y \leq 2\}$. We would like to compute the volume under the part of the graph $z = f(x, y)$ that lies above D. This will need to be a signed volume, where volume below the xy-plane counts as negative. We approximate this volume with prisms, because we have a formula for the volume of a prism.

We subdivide D into subrectangles. Over each subrectangle we place a prism. The height of each prism is the height of the graph above a test point.

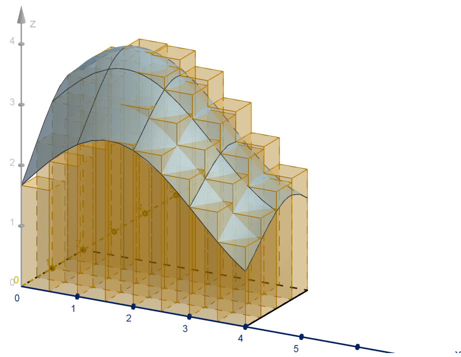


Figure: The volume under $z = f(x, y)$ and the prisms that approximate it

If A is the area of each subrectangle, and (x_i^*, y_i^*) is the test point in the i^{th} subrectangle, then our approximation is

$$\text{Volume} \approx \sum_{i=1}^{n} f(x_i^*, y_i^*)A.$$

If our domain is not a rectangle, we may not be able to divide it into subrectangles. Luckily, the formula for volume of a prism works for any shape base. We can still compute

$$\text{Volume} \approx \sum_{i=1}^{n} f(x_i^*, y_i^*)A_i.$$

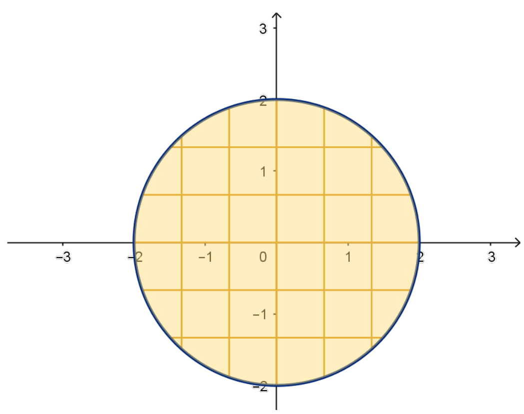


Figure: A domain subdivided into irregular subregions

Notice that instead of a single variable A for the area of all subregions, we need a different area for each. For each i, A_i denotes the area of the i^{th} subregion.

For a reasonably well-behaved function $f(x, y)$, the actual volume can be computed by taking a limit of these approximations. We call this limit the double-integral.

Question 6.1.1 How Do We Approximate the Volume Under $z = f(x,y)$?

Definition

Let D be a domain in $\mathbb{R}^2$. For a given division of D into n subregions denote

- A_i, the area of the i^{th} region.
- (x_i^*, y_i^*), any point in the i^{th} region
- $|A|$ is the diameter of the largest region.

We define the **double integral** of $f(x,y)$ to be a limit over all possible divisions of D.

$$\iint_D f(x,y)dA = \lim_{|A|\to 0} \sum_{i=1}^{n} f(x_i^*, y_i^*)A_i$$

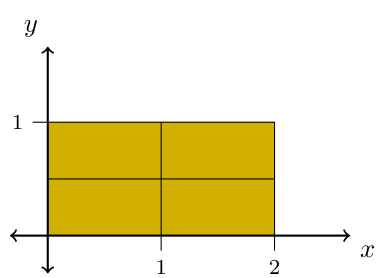
Remark

The **diameter** of a region is the distance between its two most distant points. Sending the largest diameter to 0 ensures that all of the regions' diameters shrink to 0.

Notice that we do not take the limit as the area goes to 0. If only the areas approach 0, the regions could become long and thin. The test points could all be chosen from one end of the domain which is unrepresentative of the whole.

Example 6.1.2 Approximating a Double Integral

Consider $\iint_D x^2 y dA$, where D is the region shown here. Approximate the integral using the division of D shown, and evaluating $f(x,y)$ at the midpoint of each rectangle.



Solution

The value of A is the area of each rectangle. In this case that is

$$A = (1)(0.5) = 0.5.$$

The test points are the midpoints of each rectangle:

$$(x_1^*, y_1^*) = (0.5, 0.25) \qquad (x_3^*, y_3^*) = (1.5, 0.25)$$
$$(x_2^*, y_2^*) = (0.5, 0.75) \qquad (x_4^*, y_4^*) = (1.5, 0.75)$$

We can expand the sum and evaluate:

$$\begin{aligned}
\text{Volume} &\approx \sum_{i=1}^{4} f(x_i^*, y_i^*)A \\
&\approx A\sum_{i=1}^{4} f(x_i^*, y_i^*) \\
&\approx A\big(f(0.5,0.25) + f(0.5,0.75) + f(1.5,0.25) + f(1.5,0.75)\big) \\
&\approx 0.5\big((0.5)^2(0.25) + (0.5)^2(0.75) + (1.5)^2(0.25) + (1.5)^2(0.75)\big)
\end{aligned}$$

Question 6.1.3

How Do We Evaluate Double Integrals?

We already know another way of computing a volume. We can compute the area of the cross sections perpendicular to the x-axis. Let the function $A(x)$ denote this area at each x. Then

$$\text{Volume} = \int_a^b A(x)\ dx$$

$A(x)$ is itself the area under a curve. In a particular cross section, x is constant, and $f(x, y)$ is a function of y. The area below this graph is the integral

$$A(x) = \int_c^d f(x, y)\ dy$$

We can put these together to obtain an **iterated integral**, an integral whose integrand is itself an integral.

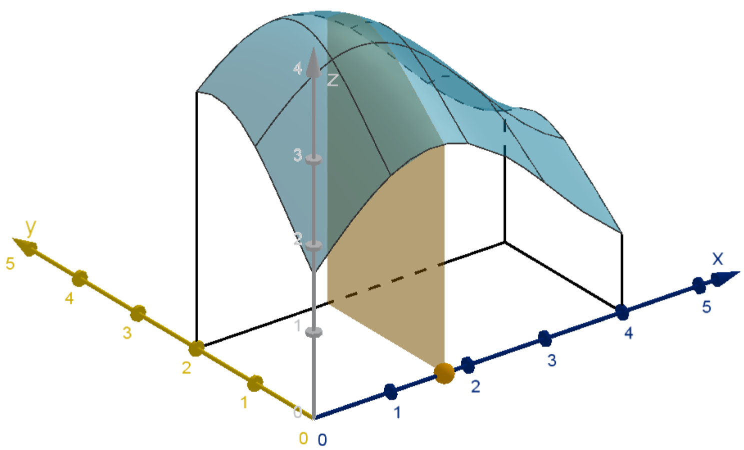


Figure: Cross sections of the region below the graph: $z = f(x, y)$

This method computes the same signed volume as the double integral we defined. The formal argument that they are equivalent is called Fubini's theorem.

Theorem [Fubini's Theorem]

For any domain D we have

$$\iint_D f(x,y)\ dA = \int_a^b \left(\int_c^d f(x,y)\ dy \right)\ dx$$

where a and b are the x bounds of D, and c and d are the y bounds of the cross section at each x. Alternately, we can write

$$\iint_D f(x,y)\ dA = \int_c^d \left(\int_a^b f(x,y)\ dx \right)\ dy$$

where c and d are the y bounds of D, and a and b are the x bounds of the cross section at each y.

Notation

We will generally omit the parentheses and write

$$\int_a^b \int_c^d f(x,y)\ dydx.$$

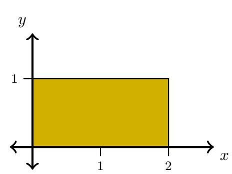
In some cases, rather than figuring out what a, b, c and d are, we will use a hybrid notation. It indicates a particular order of integration but does not go into details about the bounds of x and y.

$$\iint_D f(x,y)dydx.$$

Example 6.1.4

Using Fubini's Theorem

Compute $\iint_D x^2y\ dA$, where D is the region shown here:



Solution

The x bounds of this region are $0 \le x \le 2$. The y bounds are $0 \le y \le 1$. We rewrite this as an integrated integral and solve:

$$\begin{aligned}
\iint_D x^2y\ dA &= \int_0^2 \int_0^1 x^2y\ dydx && \text{(Fubini's theorem)}\\
&= \int_0^2 \left.\frac{x^2y^2}{2}\right|_{y=0}^{1} dx && \text{(FTC on the inner integral)}\\
&= \int_0^2 \frac{x^2 1^2}{2} - \frac{x^2 0^2}{2}\ dx && \text{(plug in } y \text{ values)}\\
&= \int_0^2 \frac{x^2}{2}\ dx\\
&= \left.\frac{x^3}{6}\right|_{x=0}^{2}\\
&= \frac{8}{6} - \frac{0}{6}
\end{aligned}$$

Example 6.1.4 Using Fubini's Theorem

$$= \frac{4}{3}$$

Question 6.1.5

Can We Break a Double Integral into a Product of Single Integrals?

In general, we can't expect to factor out the inner integral of $\iint_D f(x, y)dydx$ (using the constant multiple rule). The y-bounds may depend on x, and the y terms may not factor out of the integrand. However, for certain functions and domains, this factoring is possible.

Theorem

$$\int_a^b \int_c^d f(x)g(y)dydx = \left(\int_a^b f(x)dx\right)\left(\int_c^d g(y)dy\right)$$

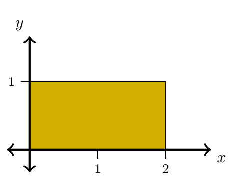
We won't be able to use this theorem all the time. It has two important requirements:

1. The bounds of integration (a, b, c, d) are constants. We'll see integrals soon where this is not the case.
2. The integrand can be factored into a function of x times a function of y. Most two-variable functions cannot.

Example 6.1.6

Integrating a Product

Use a product decomposition to compute $\iint_D x^2 y dA$, where D is the region shown here:



Solution

$\iint_D x^2y\ dA = \int_0^2 \int_0^1 x^2y\ dydx$ has constant bounds and the integrand can factor as $(x^2)(y)$. The product theorem applies:

$$\begin{aligned}
\int_0^2 \int_0^1 x^2y\ dydx &= \left(\int_0^2 x^2 dx\right)\left(\int_0^1 ydy\right) \\
&= \left(\left.\frac{x^3}{3}\right|_0^2\right)\left(\left.\frac{y^2}{2}\right|_0^1\right) \\
&= \left(\frac{2^3}{3} - \frac{0^3}{3}\right)\left(\frac{1^2}{2} - \frac{0^2}{2}\right) \\
&= \left(\frac{8}{3}\right)\left(\frac{1}{2}\right) \\
&= \frac{4}{3}
\end{aligned}$$

This matches our computation from Example 4.

Remark

The product decomposition does not save us much work in most cases, but it can help us avoid mixing up the variables.

Application 6.1.7
Rates (per Area)

Single integrals can compute total change given a rate of change.

- meters traveled per second $\longrightarrow$ total meters traveled.
- GDP growth per year $\longrightarrow$ total GDP growth.
- mass of a chemical produced per second $\longrightarrow$ total mass produced.

Double integrals can compute a total from a rate per unit of area. Integrating rainfall per square kilometer gives the total rain that fell in a watershed.

Application 6.1.7 Rates (per Area)

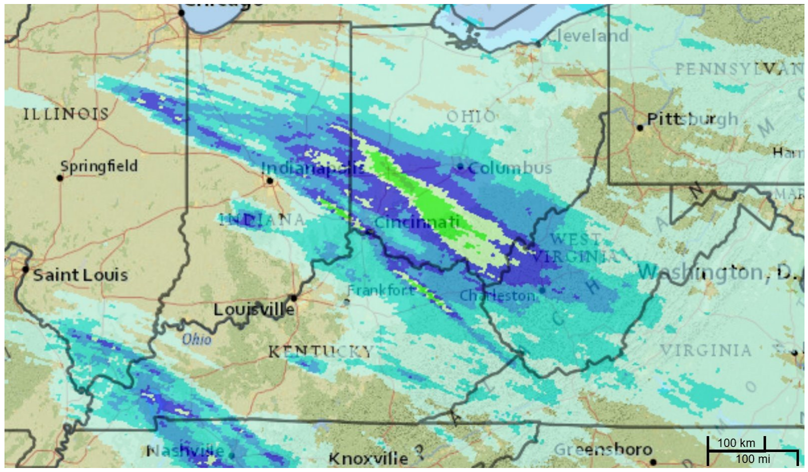


Figure: A rainfall density map

Integrating watts per square meter on a solar array gives the total energy generated.

Figure: Solar panels

Application 6.1.8 Probability

If we generate a data set in which we have measured two variables, then the probability that a random data point lies in a given region is the double integral of a **joint density function** over that area.

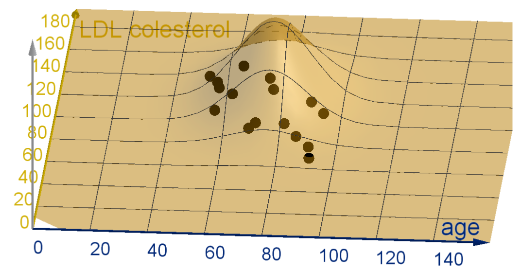


Figure: A highly correlated set of observations and an uncorrelated joint density function

Section 6.1
Exercises

Summary Questions

Q1 What shape do we use to approximate volume under a surface?

Q2 What formula do we use to compute the exact volume under a graph $z = f(x, y)$?

Q3 What does Fubini's Theorem tell us?

Q4 What conditions do you need in order to write a double integral as a product of single integrals?

6.1.1

Q5 Suppose that we are approximating the volume under $z = f(x, y)$ over T, the triangle with vertices $(0, 0)$, $(2, 0)$ and $(0, 1)$. We'd like to use subregions about 0.25 units long per side. Here are two options:

- Cover as much of T with square prisms as possible, use triangluar prisms in the remaining spots.

- Cover as much of T with square prisms as possible, and just forget about the remaining space.

a Draw a diagram of where the squares and triangles could reasonably be placed.

b Suppose the side length of the squares shrinks to be arbitrarilty small. Explain why it does not matter which of the two options we use in these approximations.

Q6 Let S be the unit square:

$$S = \{(x, y) \ : \ 0 \le x \le 1, 0 \le y \le 1\}$$

Suppose we approximate $\displaystyle\iint_S x \; dA$ by n prisms whose bases are rectangles of length 1 in the x-direction and width $\frac{1}{n}$ in the y-direction.

a How could you pick test points in each rectangle to ensure that the value of this approximation is 0, no matter what n is?

b How could you pick test points in each rectangle to ensure that the value of this approximation is 1, no matter what n is?

c Does the fact that both of these approximations are possible no matter how many rectangles we use mean that $\displaystyle\iint_S x \; dA$ does not exist? Explain.

6.1.2

Q7 Show how to approximate the integral $\int_0^6 \int_3^{12} xy \; dydx$ using six 3 unit by 3 unit squares and using their lower right corners as test points. You do not need to simplify the arithmetic.

Q8 Approximate the value of $\int_0^4 \int_{-2}^4 \sin^2(\pi xy) dydx$ by dividing the domain into an array of 4 rectangles (2×2), and evaluating the function at the midpoint of each.

Q9 Consider the integral

$$\int_0^6 \int_5^9 \frac{x}{y} \; dydx.$$

a Show how to approximate the integral using six 2 unit by 2 unit squares and using their lower right corners as test points. You do not need to simplify the arithmetic.

b Explain how you can tell whether your approximation in **a** is an overestimate or underestimate without computing the actual value of the integral.

Q10 Let T be the triangle with vertices $(0,0)$, $(1,0)$ and $(0,2)$. Show how to approximate $\iint_T e^{x+y}\,dA$ by dividing T into four right triangles with legs of length 1 and $\frac{1}{2}$. Use the midpoint of the hypotenuses as the test points.

6.1.3

Q11 Let R be the rectangle

$$R = \{(x,y)\ :\ 0 \le x \le 5, 0 \le y \le 3\}.$$

Let S be the solid region above R and below the graph $z = y^2 \sin \pi x + 9$. What is the area of the $y = 2$ cross-section of S?

Q12 Let R be the rectangle

$$R = \{(x,y)\ :\ -2 \le x \le 2, -1 \le y \le 1\}.$$

Let S be the solid region above R and below the graph $z = x^2y + xy^2$. Write a function $A(x)$ which gives the area of the cross section of S perpendicular to the x-axis at each value of x.

6.1.4

Q13 Let R be the rectangle

$$R = \{(x,y)\ :\ 0 \le x \le 5, 0 \le y \le 3\}.$$

Compute $\displaystyle\iint_R y^2 \sin \pi x + 9\ dA$

Q14 Let R be the rectangle

$$R = \{(x,y)\ :\ -2 \le x \le 2, -1 \le y \le 1\}.$$

Compute $\displaystyle\iint_R x^2y + xy^2\ dA.$

Q15 Evaluate $\int_4^5 \int_0^3 ye^x \; dydx$.

Q16 Evaluate $\int_0^{10} \int_2^4 y^3 - x \; dydx$.

6.1.5

Q17 Let R be the rectangle

$$R = \{(x,y) \; : \; -a \leq x \leq b, c \leq y \leq d\}.$$

Let S be the solid region above R and below the graph $z = f(x)g(y)$. Write a function $A(x)$ which gives the area of the cross section of S perpendicular to the x-axis at each value of x. Explain why you can factor the $f(x)$ out of this integral.

Q18 Let R be the rectangle

$$R = \{(x,y) \; : \; -2 \leq x \leq 2, -1 \leq y \leq 1\}.$$

Explain why the product decomposition theorem does not apply to $\iint_R x^2y + xy^2 \; dA$.

6.1.6

Q19 Let R be the rectangle

$$R = \{(x,y) \; : \; 0 \leq x \leq 5, 0 \leq y \leq 3\}.$$

Write $\iint_R y^2 \sin \pi x \; dA$ as a product of two single-variable integrals.

Q20 Write $\int_{-3}^{3} \int_2^5 \frac{1}{y^2} \; dydx$ as a product of two single-variable integrals.

6.1.7

Q21 A corrugated metal sheet has density of $dx, y = 3 + \sin 2x \ kg/m^2$. What is the mass of the rectangular sheet $R = \{(x, y) \ : \ 0 \leq x \leq 4\pi, 0 \leq y \leq 10\}$?

Q22 The shadow of a tree passes over part of a solar panel each day, covering the bottom of the panel more of the day than the top. The rate of daily energy generation per unit of area at the point (x, y) is given by $p(x, y) = 8 \sin\left(y + \frac{\pi}{3}\right)$ kilowatt hours per square meter. Compute the total power generated per day by the panel whose bounds (in meters) are given by $0 \leq x \leq 1$ and $0 \leq y \leq \frac{\pi}{6}$.

Synthesis & Extension

Q23 Suppose we wanted to compute the volume above $z = f(x, y)$ and below $z = g(x, y)$ over the rectangle

$$R = \{(x, y) \ : \ a \leq x \leq b, c \leq y \leq d\}.$$

What double integral would compute this volume?

Q24 Suppose you want to approximate

$$\int_a^b \int_c^d f(x, y) dy dx$$

by rectangles sampled from either upper-left, upper-right, lower-left or lower-right corners. If you are told that $f_x(x, y) < 0$ at all points (x, y), what does that tell you about which approximations are larger than which?

Section 6.2

Double Integrals over General Regions

Goals:

1. Set up double integrals over regions that are not rectangles.
2. Evaluate integrals where the bounds contain variables.
3. Decide when to make $\int dy$ the outer integral, and compute the change of bounds.

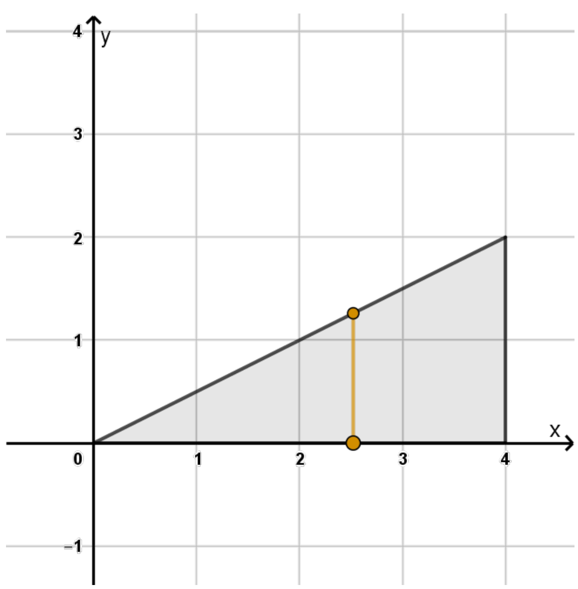
So far, we have computed double integrals over rectangular domains. In this section, we consider double integrals over more complicated domains.

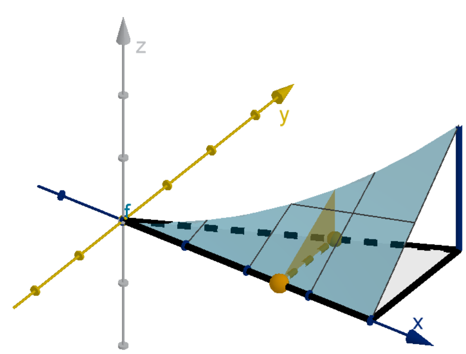
Example 6.2.1

Integrating Over a Polygon

Let D be the triangle with vertices $(0,0)$, $(4,0)$ and $(4,2)$. Calculate

$$\iint_D 4xy\ dA$$





Solution

The naive approach would be to use the x and y bounds of D to write the integral

$$\iint_D 4xy\ dA = \int_0^4 \int_0^2 4xy\ dydx$$

There are a couple reasons to distrust this approach

1. These are the same bounds we would use for a rectangle, and D is not a rectangle.
2. The y bounds are supposed to be the bounds of the cross section, and not every cross section extends from $y = 0$ to $y = 2$.

In fact, the y bounds of the cross section depend on which cross section we're looking at. At $x = 1$ the cross section has area $A(1) = \int_0^{0.5} 4xy\ dy$. At $x = 3$ the area is $A(3) = \int_0^{1.5} 4xy\ dy$. Another way to say this is that the y bounds are a function of x. No matter what x we choose, the lower y bound appears to be 0. The upper bound always lies on the line from $(0, 0)$ to $(4, 2)$. We can express the y-values of this line as a function of x by writing its equation: $y = \frac{1}{2}x$. The correct iterated integral is

$$\iint_D 4xy\ dA = \int_0^4 \int_0^{\frac{1}{2}x} 4xy\ dydx$$

This may appear harder to solve, but it isn't. The only difference is that when we apply the fundamental theorem of calculus to the inner integral, we plug in an expression instead of a number.

$$\begin{aligned}
\iint_D 4xy\ dA &= \int_0^4 \int_0^{\frac{1}{2}x} 4xy\ dydx && \text{(Fubini's theorem)} \\
&= \int_0^4 2xy^2 \Big|_0^{\frac{1}{2}x}\ dx && \text{(FTC)} \\
&= \int_0^4 2x\left(\frac{1}{2}x\right)^2 - 2x(0)^2\ dx \\
&= \int_0^4 \frac{x^3}{2}\ dx \\
&= \frac{x^4}{8}\Big|_{x=0}^{4} && \text{(FTC again)} \\
&= \frac{4^4}{8} - \frac{0^4}{8} \\
&= 32
\end{aligned}$$

Main Idea

To find the bounds of a double integral

1. Find the x value where the domain begins and ends. These numbers are the bounds of the outer integral.
2. Find the functions (of the form $y = g(x)$) which define the top and bottom of the domain. These functions are the bounds of the inner integral.

Question 6.2.2

What Are the Integral Laws for Double Integrals?

Some single variable integral laws apply to double integrals as well (provided the integrals exist).

1. The sum rule:

$$\iint_D f(x,y) + g(x,y)dA = \iint_D f(x,y)dA + \iint_D g(x,y)dA$$

2. The constant multiple rule:

$$\iint_D cf(x,y)dA = c\iint_D f(x,y)dA$$

3. If D is the union of two non-overlapping subdomains D_1 and D_2 then

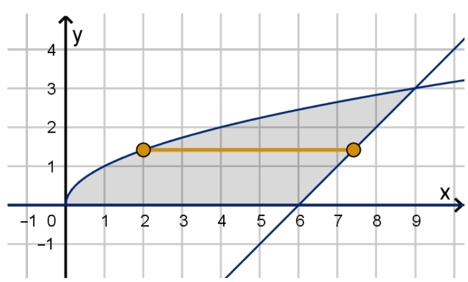
$$\iint_D f(x,y)dA = \iint_{D_1} f(x,y)dA + \iint_{D_2} f(x,y)dA$$

Example 6.2.3

A Region Without a (Single) Bottom Curve

Let D be the region bounded by $y = \sqrt{x}$, $y = 0$ and $y = x - 6$. Calculate

$$\iint_D (x+y)\ dA.$$



We begin by finding the intersections of these graphs. There are three pairs of graphs to solve for.

$$\begin{aligned} \sqrt{x} &= x - 6 \\ x &= (x-6)^2 \\ 0 &= x^2 - 13x - 36 \\ 0 &= (x-4)(x-9) \\ x &= 4 \text{ or } x = 9 \end{aligned} \qquad \begin{aligned} 0 &= \sqrt{x} \\ 0 &= x \end{aligned} \qquad \begin{aligned} \vec{0} &= x - 6 \\ 6 &= x \end{aligned}$$

When we square both sides of an equation we have to check our solutions. $x = 4$ does not satisfy $\sqrt{x} = x - 6$ but $x = 9$ does. Look at the graph of these functions. There is not a single y lower bound that applies to all cross sections of this region. For some values of x, the lower bound lies on $y = 0$. For others it lies on $y = x - 6$. We will present three solutions to this problem. We'll only evaluate the last one.

Solution 1

Using the third integral law, we break up D into two subdomains, each of which has a single bottom curve. The break happens at $x = 6$ since that is where $y = 0$ meets $y = x - 6$.

$$\int_0^6 \int_0^{\sqrt{x}} (x+y)\ dA + \int_6^9 \int_{x-6}^{\sqrt{x}} (x+y)\ dydx$$

Solution 2

D can be written as the region between $y = 0$ and $y = \sqrt{x}$ with a triangle removed. We can use this to write $\iint_D (x+y)\ dA$ as a difference of two integrals.

$$\int_0^9 \int_0^{\sqrt{x}} (x+y)\ dA - \int_6^9 \int_0^{x-6} (x+y)\ dydx$$

Solution 3

Instead of taking cross sections perpendicular to the x-axis we can take cross sections perpendicular to the y-axis. In this case, we need to know the x bounds of each cross section (as a function of y). Drawing the horizontal line segments through D at each y, we see that the upper x-bound lies on $y = x - 6$ and the lower x bound lies on $y = \sqrt{x}$. We need to write these x values as functions of y so we solve them for y:

$$y = \sqrt{x} \qquad\qquad y = x - 6$$

$$y^2 = x \qquad\qquad y + 6 = x$$

The lower y bound for the region is $y = 0$. The upper y bound is the intersection of $y = \sqrt{x}$ and $y = x - 6$, where $x = 9$ and $y = 3$. Thus we can write

$$\begin{aligned}
\iint_D (x+y)\ dA &= \int_0^3 \int_{y^2}^{y+6} (x+y)\ dxdy \\
&= \int_0^3 \left. \frac{x^2}{2} + xy \right|_{y^2}^{y+6} dy \\
&= \int_0^3 \frac{y^2+12y+36}{2} + y^2 + 6y - \frac{y^4}{2} - y^3\ dy \\
&= \int_0^3 -\frac{1}{2}y^4 - y^3 + \frac{3}{2}y^2 + 12y + 18\ dy \\
&= \left. -\frac{1}{10}y^5 - \frac{1}{4}y^4 + \frac{1}{2}y^3 + 6y^2 + 18y \right|_0^3
\end{aligned}$$

Example 6.2.3 A Region Without a (Single) Bottom Curve

$$= -\frac{243}{10} - \frac{81}{4} + \frac{27}{2} + 54 + 54$$
$$= \frac{1341}{20}$$

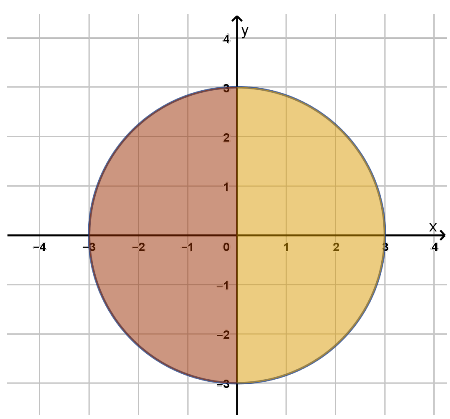
Main Idea

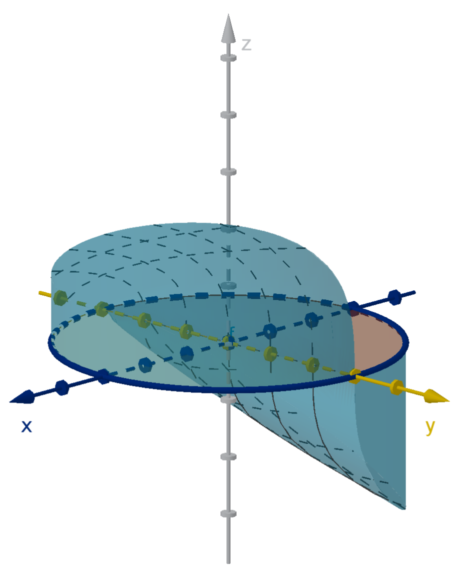
For a region without a single upper or lower curve, the strategies for integrating a function are the same as the strategies for computing the area.

1. Break the region into two or more pieces, each of which has a single top curve and a single bottom curve.
2. See if the region has a single left curve (lower x bound) and a single right curve (upper x bound). If so, solve the bounds for x and change the order of integration.

Example 6.2.4
Using Anti-Symmetry

Let D be the region $x^2 + y^2 \leq 9$. Evaluate $\iint_D \sqrt[3]{x}\sqrt{y+3}dA$.





Solution

The function f and the domain D both have a particular type of symmetry. D is symmetric about the y-axis. We can flip the right side of D over onto the left side of D and they match up perfectly. We can express this transformation in algebra by

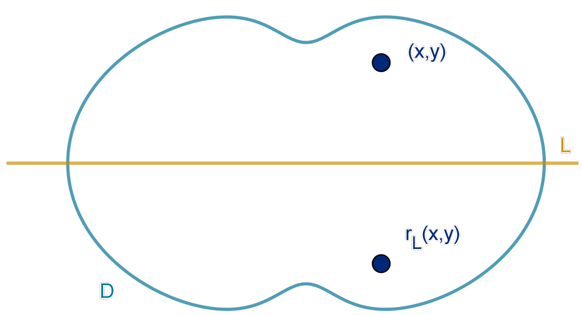
$$(x, y) \to (-x, y)$$

Furthermore, $f(x, y) = \sqrt[3]{x}\sqrt{y+3}$ and $f(-x, y) = \sqrt[3]{-x}\sqrt{y+3}$ are opposites (they sum to 0). Thus the height of the graph $z = f(x, y)$ above the left half of D is equal to the depth of the graph below the right half of D. These two regions have opposite signed volumes. Their sum, which is the integral over all of D, is 0.

Main Idea

We can argue that an integral $\iint_D f(x, y)dA$ is equal to zero when

1. D is **symmetric** about some line L. If we folded it over L, one side of D would lie exactly on the other side.
2. f is **antisymmetric** about L. For each point (x, y) in D the image of (x, y) across L, denoted $r_L(x, y)$ has the property:
$$f(r_L(x, y)) = -f(x, y).$$

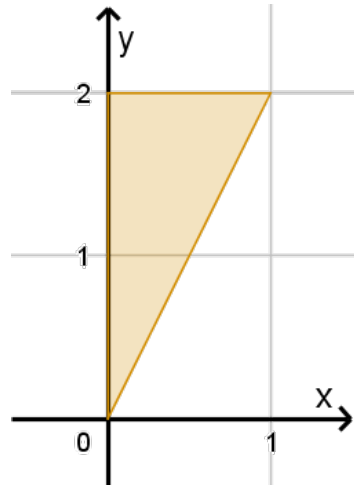


Example 6.2.5

Using Order to Manipulate the Integrand

Let D be the triangle with vertices $(0,0)$, $(0,2)$ and $(1,2)$. Calculate

$$\iint_D e^{(y^2)}\ dA.$$



Solution

D is the region above $y = 2x$ and below $y = 2$ so we can write the integral

$$\iint_D e^{(y^2)}\ dA = \int_0^1 \int_{2x}^2 e^{(y^2)}\ dydx$$

The next step is to integrate with respect to y, but $e^{(y^2)}$ does not have an antiderivative that we can evaluate precisely. The trick in this case is to change the order of integration. The lower x bound is $x = 0$ the upper x bound is $x = \frac{y}{2}$.

$$\begin{aligned}
\iint_D e^{(y^2)}\ dA &= \int_0^2 \int_0^{\frac{y}{2}} e^{(y^2)}\ dxdy \\
&= \int_0^2 e^{(y^2)} x \Big|_0^{\frac{y}{2}}\ dy \\
&= \int_0^2 e^{(y^2)} \left(\frac{y}{2}\right)\ dy \\
&= \int_0^4 \frac{1}{4} e^u\ du \\
&= \frac{1}{4}\ e^u \Big|_0^4 \\
&= \frac{e^4}{4} - \frac{1}{4}
\end{aligned}$$

u-substitution

$$\begin{aligned}
u &= y^2 & y = 0 &\Rightarrow u = 0 \\
du &= 2y\ dy & y = 2 &\Rightarrow u = 4 \\
\tfrac{1}{4}\ du &= \tfrac{y}{2}\ dy
\end{aligned}$$

Main Idea

If we don't know the anti-derivative of an integrand with respect to one variable, try switching the order of integration. Remember to change the bounds too.

Application 6.2.6

Area of a Domain

We can use a double integral of f to measure the domain of integration, or compute statistics about f. Here are two examples.

Theorem

The area of a region D can be calculated:

$$\iint_D 1 \; dA.$$

This theorem may seem counter-intuitive at first, because a double integral computes a volume, not an area. However, the volume under a graph of height 1 is equal to 1 times the area of the base. As long as we change from cubic units to square units, the integral will be numerically equal to the area.

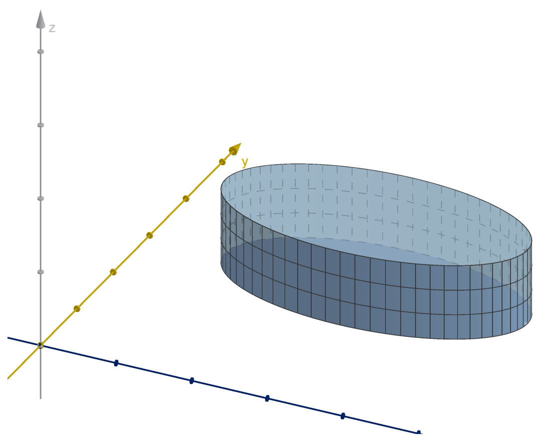


Figure: A solid of height 1 over a domain D

Section 6.2

Exercises

Summary Questions

Q1 What are the steps for writing a double integral over a general region?

Q2 How do you decide whether dx or dy is the inner variable?

Q3 What is antisymmetry, and how can we use it to evaluate integrals?

Q4 How can we use a double integral to compute the area of a region?

6.2.1

Q5 If D is the triangle with vertices $(0, -2)$, $(4, 0)$ and $(0, 8)$ calculate $\iint_D x^2 y \, dA$

Q6 Integrate the function $f(x, y) = y$ over the region enclosed by the lines $y = 5x$, $y = 6 - x$ and $y = x$.

Q7 Let $f(x, y)$ be a function and D be the trapezoid with vertices $(3, 1), (3, 6), (6, 5)$ and $(6, 4)$. Draw D and set up the bounds of $\iint_D f(x, y) dA$.

Q8 Let D be the parallelogram with vertices $(0, 1)$, $(0, 4)$, $(5, 3)$ and $(5, 6)$. Let $f(x, y)$ be a continuous function.

a Set up the bounds of integration of $\displaystyle\iint_D f(x, y) \, dA$.

b Could we save time by computing $\int_0^5 \int_1^4 f(x, y) \, dydx$ instead? Explain.

Q9 Let D be the region enclosed by $y = 6 - x^2$ and $y = x$. Evaluate $\displaystyle\iint_D xe^y \, dA$.

Q10 If D is the region bounded by $y = x^2$ and $y = 8 - x^2$, set up and calculate $\iint_D x^3 dA$.

6.2.2

Q11 Let T be the triangle with vertices $(0,3)$, $(7,10)$ and $(9,0)$. Set up the bounds for two intgrals whose sum is $\iint_T f(x,y)\ dA$.

Q12 Let P be the pentagon with vertices $(0,0)$, $(0,2)$, $(4,3)$, $(4,1)$ and $(3,0)$.

- **a** Set up the bounds for two integrals whose sum is $\iint_P f(x,y)\ dA$.
- **b** Set up the bounds for two integrals whose difference is $\iint_P f(x,y)\ dA$.

6.2.3

Q13 Let D be the region enclosed by $y = \ln x$, $x = 1$ and $y = 4 - \ln x$. Set up the integral

$$\iint_D f(x,y)\ dA$$

in two different ways, using both orders of dx and dy. Do not evaluate either.

Q14 Let $D = \{(x,y)\ :\ x^2 + y^2 \leq 9, x \geq 0\}$. Draw D and set up $\iint_D f(x,y)\ dA$ in two different ways.

Q15 Consider the region D enclosed by $y = \sqrt{x}$, $y = 27\sqrt{x}$, and $y = 90 - x$.

- **a** Rewrite $\iint_D f(x,y)\ dA$ as one or more integrals with differential $dydx$. Do not evaluate.
- **b** Rewrite $\iint_D f(x,y)\ dA$ as one or more integrals with differential $dxdy$. Do not evaluate.

Q16 Let $D = \{(x,y)\ :\ y \leq 12 - x^2, y \geq x, y \geq -x\}$.

- **a** Rewrite $\iint_D f(x,y)\ dA$ as one or more integrals with differential $dydx$. Do not evaluate.
- **b** Rewrite $\iint_D f(x,y)\ dA$ as one or more integrals with differential $dxdy$. Do not evaluate.

Q17 Draw the domain of the integral $\displaystyle\int_1^5 \int_0^{10-2x} f(x,y)\ dydx$. Then rewrite the integral in the order $dxdy$.

Q18 Consider the integral $\displaystyle\int_{-6}^{6}\int_{-\sqrt{36-y^2}}^{0} x^2\,dxdy$. Write this integral in the order $dydx$.

6.2.4

Q19 Let $f(x,y) = \sqrt[3]{\cos x \sin y}$. Argue that $\displaystyle\int_{-8}^{8}\int_{-\sqrt{64-x^2}}^{\sqrt{64-x^2}} f(x,y)\,dydx = 0$.

Q20 Let $g(x,y) = x^3 e^{y^2}$. Argue that $\displaystyle\int_{-4}^{4}\int_{-3}^{3} g(x,y)\,dydx = 0$.

Q21 Let R be the kite with vertices

$$\begin{array}{cc} (1,1) & (5,7) \\ (7,7) & (7,5) \end{array}$$

Suppose you wanted to argue that $\iint_R f(x,y)dA = 0$ by a symmetry argument. Describe with a diagram or formula what would need to be true about $f(x,y)$ for such an argument to work.

Q22 Let D be the trapezoid with vertices $(0,5)$, $(6,5)$, $(2,0)$ and $(4,0)$. Let $g(x,y)$ be some continuous function.

a. Sketch D and set up the bounds of integration for $\iint_D g(x,y)\,dA$ such that you obtain one integral (not a sum or difference of integrals).

b. If you wanted to use an antisymmetry argument to show that $\iint_D g(x,y)\,dA = 0$ what would need to be true about $g(x,y)$? Express your answer as a formula.

Q23 Let $h(x)$ be a one-variable function that takes only positive values. Let $f(x,y)$ be a two-variable function. Describe the antisymmetry of f that would allow us to conclude that $\displaystyle\int_a^b\int_{-h(x)}^{h(x)} f(x,y)\,dydx = 0$.

Q24 Suppose you are given that $f(x,y) = -f(-y,-x)$. Over what domains D can we argue by symmetry that $\displaystyle\iint_D f(x,y)\,dA = 0$? Draw an example of one.

6.2.5

Q25 Would the method in this example still work, if we instead defined D to have vertices $(0,0)$, $(1,0)$, and $(0,2)$? Explain.

Q26 Suggest a domain D over which it would be possible to evaluate $\displaystyle\iint_D e^{y^3}\ dA$.

Q27 Evaluate $\displaystyle\int_0^2\int_0^3 ye^{xy}\ dydx$.

Q28 Evaluate $\displaystyle\int_0^3\int_x^3 \sin(\pi y^2)\ dydx$.

6.2.6

Q29 Use geometry to evaluate $\displaystyle\int_0^{10}\int_0^{\sqrt{100-x^2}}\ dydx$.

Q30 Use geomtery to evaluate $\displaystyle\int_0^{8}\int_0^{4-\frac{1}{2}x}\ dydx$.

Synthesis & Extension

Q31 What is the geometric significance of the inner integral in a double integral of the form

$$\int_a^b\int_{g(y)}^{h(y)} f(x,y)\ dxdy?$$

Q32 Consider the integral

$$\int_{-4}^4\int_0^6 x^3\sqrt{y}dydx$$

a Show how to approximate the value of this integral, dividing the domain into sub-rectangles of length 2 units and width 3 units and using the lower right corners as test points. You should evaluate any functions that appear in your estimate, but you do not need to simplify the arithmetic.

b Explain in a sentence or two how you can determine the exact value of this integral without calculating any anti-derivatives.

c Discuss what test point you could have picked in a, such that your approximation would have computed the exact value of the integral. Note: There are several relevant observations to make in response to this question.

Section 6.3

Joint Probability Distributions

Goals:

1. Integrate a joint density function to calculate a probability.
2. Recognize when random variables are independent.

Some of the most compelling statistical conclusions do not rely on one measurement but on many, and the relationship between them. Suppose we test a drug by randomly giving different doses to different participants, then measuring their symptoms. Knowing the likelihood of each level of symptoms doesn't tell you whether the drug is effective. Adding in the knowledge of what percentage of test subjects receive each dosage does not help. Instead you need to know how likely certain pairs of dose and outcomes are:

(low dose, low symptoms) (low dose, high symptoms)

(medium dose, low symptoms) (no dose, medium symptoms)

If (no dose, high symptoms) and (high dose, low symptoms) are likely enough, then there is a correlation which points to efficacy of the drug. Individual random variables with individual density functions cannot model this behavior. We need two-variable density functions and double integrals.

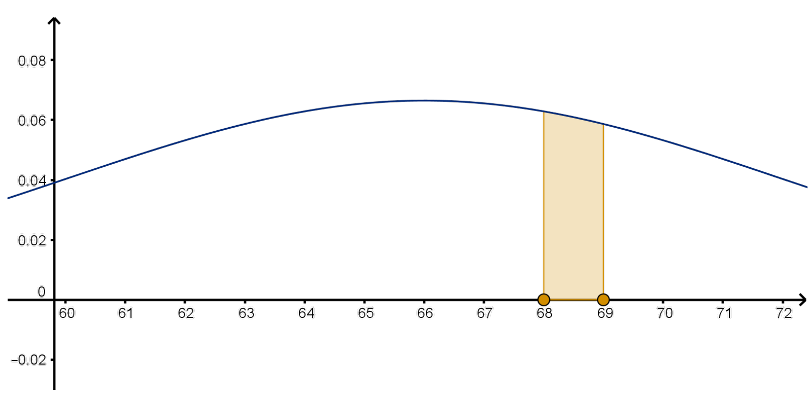
Question 6.3.1

How Do We Use Double Integrals to Compute Probabilities?

Recall how we modeled continuous random variables.

Definition

A function f is a **probability density function** for a random variable X, if the chance of an outcome $a < X < b$ is $\int_a^b f(x)dx$.



Question 6.3.1 How Do We Use Double Integrals to Compute Probabilities?

Definition

A pair (or more) of random variables X and Y, along with the likelihood of various outcomes (X, Y) is called a joint distribution. If the space of outcomes is continuous, the distribution is modeled by a **joint probability density function** $f_{X,Y}(x, y)$ as follows:

$$P(a \leq X \leq b \text{ and } c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y)\, dydx$$

More generally, for any region D in $\mathbb{R}^2$

$$P((X, Y) \text{ lies in } D) = \iint_D f_{X,Y}(x, y)\, dA.$$

Example 6.3.2

Using a Joint Density Function

Suppose the random variables X and Y have the joint density function

$$f_{X,Y}(x, y) = \begin{cases} x + y & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Compute the probability that X is at least twice as large as Y.

Solution

We can write "X is at least twice as large as Y" with the inequality $X \geq 2Y$. This is everything below the line $y = \frac{1}{2}x$ Call this region H. We'll integrate f over this region. This may seem daunting, but $f(x, y) = 0$ outside the unit square. We can break H into two subregions, one that lies inside the square and one that lies outside. A diagram will make it easier to find the bounds.

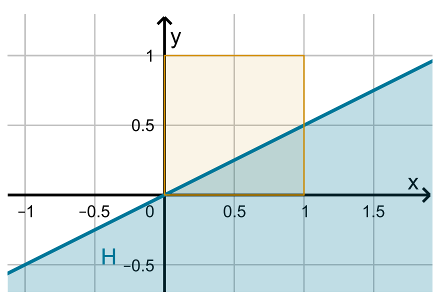


Figure: The target region H and the unit square of possible outcomes

$$
\begin{aligned}
P\left(Y \leq \frac{1}{2}X\right) &= \iint_H x + y \; dA \\
&= \int_0^1 \int_0^{\frac{1}{2}x} x + y \; dydx + \iint_{\text{the rest of } H} 0 \; dA \\
&= \int_0^1 xy + \frac{y^2}{2}\Big|_0^{\frac{1}{2}x} \; dx \\
&= \int_0^1 \frac{1}{2}x^2 + \frac{1}{8}x^2 \; dx \\
&= \int_0^1 \frac{5}{8}x^2 \; dx \\
&= \frac{5}{24}x^3\Big|_0^1 \\
&= \frac{5}{24}
\end{aligned}
$$

Warning

The region of integration in this example has one fourth of the area of the total region of possibilities, yet the answer was $\frac{5}{24}$ not $\frac{1}{4}$. Do not confuse area with probability. Not all outcomes are equally likely to occur.

Since we got a low probability, relative to area, we can deduce that the probability density in the region we examined is lower than at some other parts of the domain. That makes sense. The joint density function $x + y$ is largest in the upper right corner and lowest in the lower left. More of our triangle was near the lower left than the upper right.

Example 6.3.2 Using a Joint Density Function

Exercise

Darmok and Jalad each travel to the island of Tanagra and arrive between noon and 4 PM. Let (X, Y) represent their respective arrival times in hours after noon. Suppose their joint density function is

$$f_{X,Y}(x,y) = \begin{cases} \frac{x}{32} & \text{if } 0 \leq x \leq 4 \text{ and } 0 \leq y \leq 4 \\ 0 & \text{otherwise} \end{cases}.$$

1. What is the value of $\int_0^4 \int_0^4 f_{X,Y}(x,y)dydx$?
2. Calculate the probability that Darmok arrives after 2PM.
3. Calculate the probability that Darmok arrives before Jalad.
4. What does the distribution say about when Darmok is likely to arrive? What about Jalad?
5. Write an integral that computes the probability that they arrive within an hour of each other (set it up, don't evaluate).

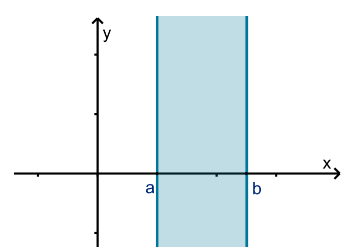
Question 6.3.3
What Is a Marginal Density Function?

Suppose we have a joint density function $f_{X,Y}(x,y)$. What if we are only interested in the values of X? Perhaps we want to compute the expected value. Recall that a density function $f_X(x)$ of X satisfies the property

$$P(a \leq X \leq b) = \int_a^b f_X(x)\, dx$$

How can we get this function from the joint density function? We can compute $P(a \leq X \leq b)$.

$$P(a \leq X \leq b) = \int_a^b \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dydx$$



Compare this to the definition of a probability density function. Both compute the same probability. Both integrate over the same range of x-values. The only way for this to be true for all values of a and b is if the integrand is the same. This means that the inner integral $\displaystyle\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$ is equal to $f_X(x)$, the probability density function of X.

When we obtain a density function of one random variable from a joint distribution, we call it a **marginal density function**.

Theorem

Given a joint distribution X, Y with joint density function $f_{X,Y}$, the individual variables have marginal density functions:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dy$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\ dx$$

For each x-value x_0, the inner integral $\int_{-\infty}^{\infty} f_{X,Y}(x_0, y)\ dy$ is the area of the $x = x_0$ cross-section under $z = f_{X,Y}(x, y)$. In this figure, we see that larger values of X are more likely, because their cross-sections have more area.

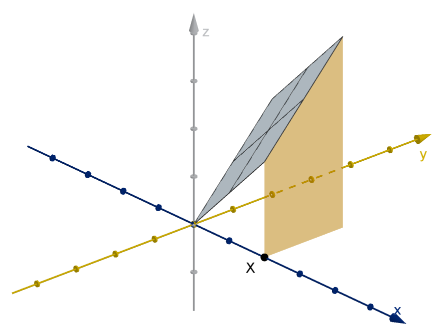


Figure: The marginal density function $f_X(x)$, represented as cross-sections under $z = f_{X,Y}(x, y)$

Example 6.3.4
Computing Marginal Density Functions

Students at schools around the world compete in a rocketry contest. Rockets are scored based on the altitude they reach (in meters). Suppose the first and second place altitudes at a randomly chosen school are modeled by X and Y, which have joint density function

$$f_{X,Y}(x, y) = \begin{cases} \frac{12-0.012x}{1000}\left(\frac{y}{x^2} - \frac{y^2}{x^3}\right) & \text{if } 0 \le x \le 1000, 0 \le y \le x \\ 0 & \text{otherwise} \end{cases}$$

Example 6.3.4 Computing Marginal Density Functions

a What can we infer about the possible altitudes of student rockets from this joint density function?

b Compute the marginal density function of X, the altitude of the first place rocket.

c What can we conclude about what values of X are more or less likely?

Solution

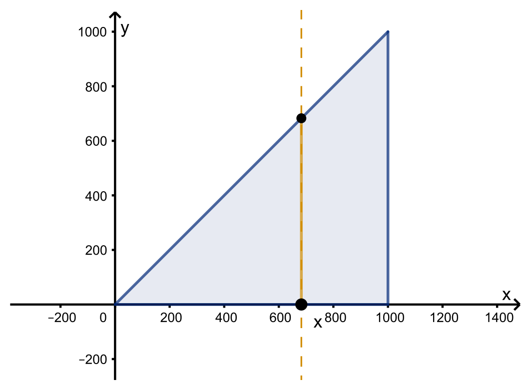


Figure: The possible outcomes of (X, Y), and the possible outcomes of Y for each X

a The maximum altitude of a rocket is $1000m$. The second-place rocket always has a lower altitude than the first-place rocket, which makes sense.

b For $x > 1000$ or $x < 0$, the function $f_{X,Y}(x, y) = 0$ for any choice of Y. For $0 \leq x \leq 1000$, the function $f_{X,Y}(x, y)$ is piecewise function of y. We can see this in the figure above, $f_{X,Y}$ is only nonzero when $0 \leq y \leq x$.

$$
\begin{aligned}
f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy \\
&= 0 && \text{if } x < 0 \text{ or } x > 1000 \\
&= \int_{-\infty}^{0} f_{X,Y}(x, y)\, dy + \int_{0}^{x} f_{X,Y}(x, y)\, dy + \int_{x}^{\infty} f_{X,Y}(x, y)\, dy && \text{if } 0 \leq x \leq 1000 \\
&= 0 + \int_0^x \frac{12 - 0.012x}{1000} \left(\frac{y}{x^2} - \frac{y^2}{x^3} \right) dy + 0 \\
&= \frac{12 - 0.012x}{1000} \int_0^x \left(\frac{y}{x^2} - \frac{y^2}{x^3} \right) dy && \text{constant multiple rule} \\
&= \frac{12 - 0.012x}{1000} \left(\frac{y^2}{2x^2} - \frac{y^3}{3x^3} \right) \Bigg|_0^x
\end{aligned}
$$

$$= \frac{12 - 0.012x}{1000}\left(\frac{x^2}{2x^2} - \frac{x^3}{3x^3} - 0 + 0\right)$$

$$= \frac{12 - 0.012x}{1000}\left(\frac{1}{2} - \frac{1}{3}\right)$$

$$= \frac{2 - 0.002x}{1000}$$

$$f_X(x) = \begin{cases} \frac{2-0.002x}{1000} & \text{if } 0 \le x \le 1000 \\ 0 & \text{otherwise} \end{cases}$$

c $f_X(x)$ has its largest value at $x = 0$ and shrinks to 0 as x increases to 1000. This indicates that lower altitudes are much more likely than higher altitudes.

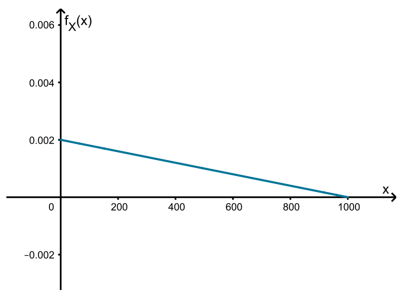


Figure: The marginal density function of X

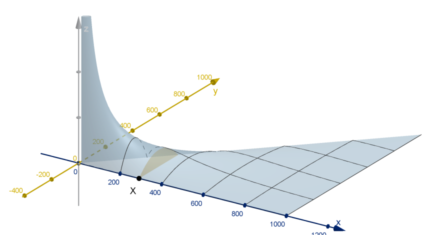


Figure: The marginal density function of X, represented as an area under the graph of $z = f_{X,Y}(x, y)$ (z-axis not to scale)

Remark

Even though the range of possible outcomes is greater for larger X, the probability of achieving that X is smaller. We can see this in the cross sections on the joint-density function. Larger values of X have longer cross sections, but it is the area under the graph $z = f_{X,Y}(x, y)$ that matters.

Example 6.3.4 Computing Marginal Density Functions

Main Idea

If the range of possible outcomes is limited, then computing $f_X(x)$ requires us to:

1. make different computations for different ranges of X and
2. within each computation, divide the integral into pieces depending on which values of Y are possible.

Question 6.3.5

Why Do We Need Joint Distributions?

If we want to communicate about the possible outcomes of X and Y, do we need to give an expression for $f_{X,Y}$? Maybe the marginal density functions f_X and f_Y tell us everything we need to know about what outcomes are likely. In fact it does not. Marginal density functions cannot tell us how the likely outcomes of Y change as the outcome of X changes, or vice versa. For instance, perhaps a patient given a random amount of medicine X is likelier to have smaller symptoms Y when X is larger.

However, in some cases, there is no change at all. No matter what the outcome of X is, the likelihood of each Y outcome is the same. In this case, the marginal density functions tell us everything we need to know. This situation is useful to recognize when it occurs, and we have a name for it.

Definition

If the outcomes of Y don't depend on the outcome of X and vice versa, we say X and Y are **independent**. In this case

$$P(a \leq X \leq b \text{ and } c \leq Y \leq d) = \int_a^b f_X(x)\, dx \int_c^d f_Y(y)\, dy$$

Example

Suppose Darmok and Jalad's arrival times have the joint density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{x}{32} & \text{if } 0 \leq x \leq 4 \text{ and } 0 \leq y \leq 4 \\ 0 & \text{otherwise} \end{cases}.$$

Jalad's arrival time is uniformly distributed. Darmok's is **triangular**. Neither distribution depends on the arrival time of the other.

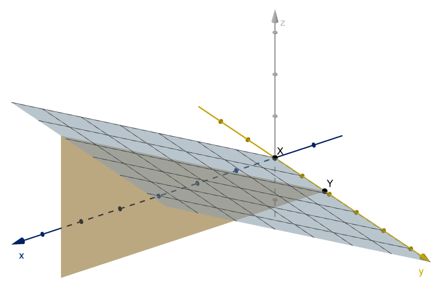


Figure: The density function for Darmok and Jalad's arrival times

Independence is straightforward to check, and it is closely related to the product decomposition of a double integral.

Theorem

X and Y are independent, if and only if their joint density function can be written $f_{X,Y}(x, y) = g(x)h(y)$, where

- $g(x)$ is a function only of x
- $h(y)$ is a function only of y

Remark

$g(x)$ and $h(y)$ can be chosen to be the marginal density functions of X and Y, but they don't need to be. As long as a factorization exists, the variables are independent.

Example

Suppose

$$f_{X,Y}(x, y) = \begin{cases} \frac{3\pi}{12\pi-8} \cos\left(\frac{\pi}{2}x\right)(2y - y^2) & \text{if } 0 \leq x \leq 6 \text{ and } 0 \leq y \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

$f_{X,Y}(x, y)$ factors into the marginal density functions

$$f_X(x) = \begin{cases} \frac{\pi}{3\pi-2} \cos\left(\frac{\pi}{2}x\right) & \text{if } 0 \leq x \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{3}{4}(2y - y^2) & \text{if } 0 \leq y \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

Thus we can conclude that X and Y are independent.

Question 6.3.5 Why Do We Need Joint Distributions?

We can see independence in the cross sections of $z = f_{X,Y}(x, y)$.

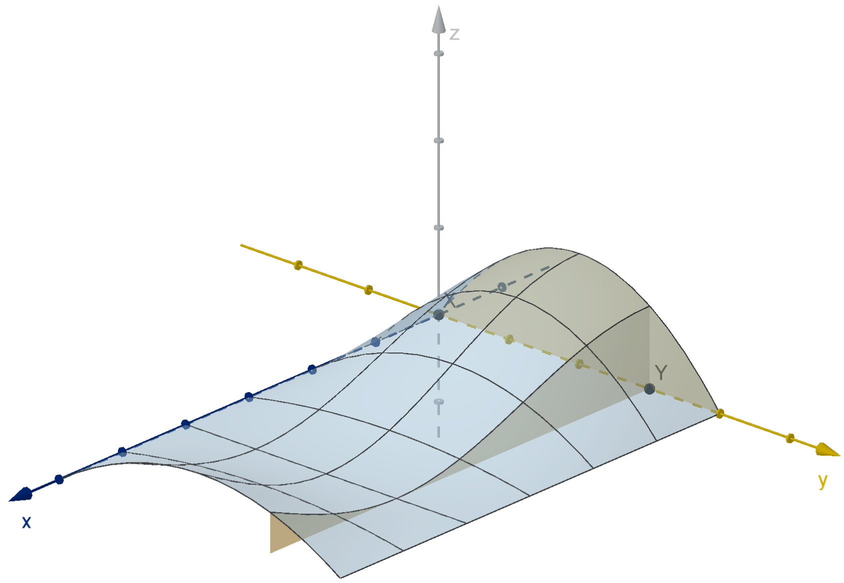


Figure: An independent joint density function and its cross sections

The area of a $y = y_0$ cross section is $f_Y(y_0)$ the likelihood that Y is near y_0. The shape of the cross section indicates what X values are likely for that choice of Y. For independent variables, the X values are distributed the same way no matter what Y value we choose. Mathematically, the cross section functions are constant multiples of each other. Multiplying by a constant does not change what portion of the total area lies over a given range of X values.

Question 6.3.6 What Is the Expected Value of a Function of X and Y?

What if we wanted to know the expected value the function $g(X, Y) = \frac{Y^2}{X}$? By definition, this is very hard. We would need to write a density function $h(t)$ such that

$$\int_a^b h(t)\, dt = P\left(a \leq \frac{Y^2}{X} \leq b\right)$$

Notice $g(x, y) = a$ and $g(x, y) = b$ are level curves of g. In this case they solve to

$$x = \frac{1}{a} y^2$$
$$x = \frac{1}{b} y^2$$

In the case of Darmok and Jalad, the probabilities that $h(t)$ produces would have to integrate to give the probability that (X, Y) lies between the level curves:

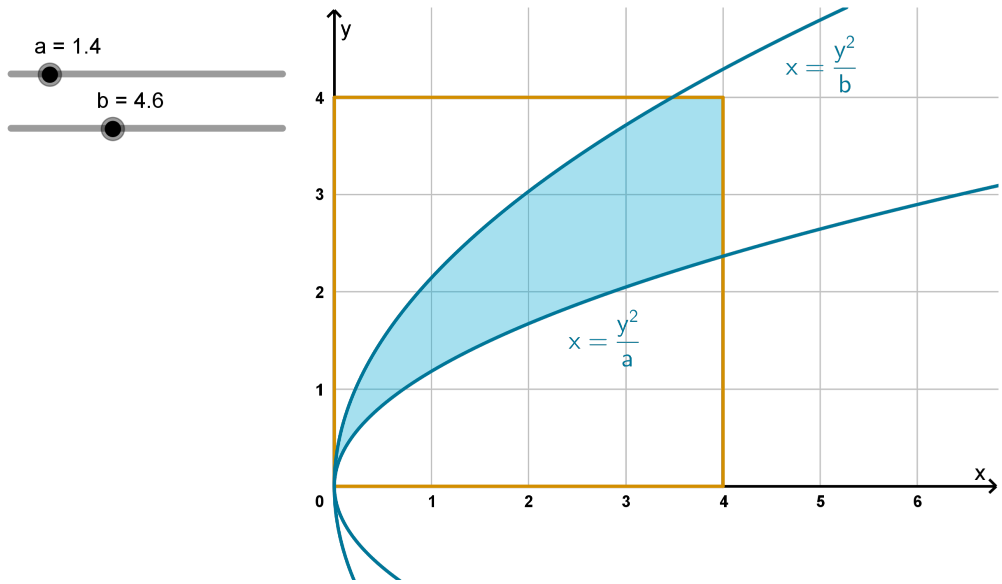


Figure: The region where $a \leq g(x, y) \leq b$

Even if you did work through the steps to describe the bounds of such a region, you'd need to

1. Write the bounds as a function of a and b, which will be piecewise depending on whether the level curves exit through the top or the side of the square.
2. Evaluate the integral of $f_{X,Y}(x, y)$ over such a region to compute $P(a \leq \frac{Y^2}{X} \leq b)$.
3. Use the Fundamental Theorem of Calculus to write an integrand $h(t)$ that integrates to the probability you found.
4. Integrate $\displaystyle\int_{-\infty}^{\infty} th(t)\ dt$.

Only then would you know the expected value of g.

Fortunately there is a multivariable analogue to the expected value theorem from single variable density functions.

Theorem

The expected value of a function $g(X, Y)$ of two continuous random variables X and Y with joint density function $f_{X,Y}(x, y)$ can be computed:

$$E[g(X)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y)\ dydx.$$

Example 6.3.7

Expected Value of a Random Variable

A special case of the expected value formula is to compute the expected values of $g(x,y)=x$ or $g(x,y)=y$. Suppose X and Y have joint density function

$$f_{X,Y}(x,y)=\begin{cases}\frac{12-0.012x}{1000}\left(\frac{y}{x^2}-\frac{y^2}{x^3}\right) & \text{if } 0\le x\le 1000, 0\le y\le x\\ 0 & \text{otherwise}\end{cases}$$

Compute $E[X]$.

Solution

$E[X]=E[g(X,Y)]$ where $g(x,y)=x$. We apply the expected value formula

$$\begin{aligned}
E[g(X,Y)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)\;dydx\\
&= \int_0^{1000}\int_0^x xf_{X,Y}(x,y)\;dydx+\iint_{\text{everywhere else}} xf_{X,Y}(x,y)\;dydx\\
&= \int_0^{1000}\int_0^x x\frac{12-0.012x}{1000}\left(\frac{y}{x^2}-\frac{y^2}{x^3}\right)\;dydx+0\\
&= \int_0^{1000}\frac{x(12-0.012x)}{1000}\int_0^x\left(\frac{y}{x^2}-\frac{y^2}{x^3}\right)\;dydx\\
&= \int_0^{1000}\frac{x(12-0.012x)}{1000}\left(\frac{y^2}{2x^2}-\frac{y^3}{3x^3}\right)\Bigg|_0^x\;dx\\
&= \int_0^{1000}\frac{x(12-0.012x)}{1000}\left(\frac{x^2}{2x^2}-\frac{x^3}{3x^3}-0+0\right)\;dx\\
&= \int_0^{1000}\frac{x(12-0.012x)}{1000}\left(\frac{1}{2}-\frac{1}{3}\right)\;dx\\
&= \int_0^{1000}\frac{x(2-0.002x)}{1000}\;dx\\
&= \int_0^{1000}\frac{2x-0.002x^2}{1000}\;dx\\
&= \frac{x^2}{1000}-\frac{0.002x^3}{3000}\Bigg|_0^{1000}\\
&= 1000-\frac{2000}{3}\\
&= \frac{1000}{3}
\end{aligned}$$

Main Ideas

- We can compute $E[X]$ or $E[Y]$ by integrating

$$E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)\ dydx$$
$$E[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)\ dydx$$

- If we already have the marginal density function $f_X(x)$ (or $f_Y(y)$), we can use the single-variable expected value formula:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\ dx$$

In fact, we saw this integral partway through our solution. Computing the marginal density function is nearly equivalent to computing the inner integral in the two-variable expected value formula.

Example 6.3.8

Expected Value of a Function

Compute the expected value of $\frac{Y^2}{X}$ where X is Darmok's arrival time and Y is Jalad's arrival time. Assume that X and Y have joint density function:

$$f_{X,Y} = \begin{cases} \frac{x}{32} & \text{if } 0 \le x \le 4 \text{ and } 0 \le y \le 4 \\ 0 & \text{otherwise} \end{cases}.$$

Solution

The expected value is given by

$$\begin{aligned} E[Y^2/X] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{y^2}{x} f_{X,Y}(x,y)\ dydx \\ &= \int_0^4\int_0^4 \frac{y^2}{x}\frac{x}{32}\ dydx && \text{because } f_{X,Y}=0 \text{ outside } 0\le x\le 4, 0\le y\le 4 \\ &= \int_0^4\int_0^4 \frac{y^2}{32}\ dydx \\ &= \int_0^4 \left.\frac{y^3}{96}\right|_0^4\ dx \\ &= \int_0^4 \frac{2}{3}\ dx \end{aligned}$$

Example 6.3.8 Expected Value of a Function

$$= \frac{2}{3}x\Big|_0^4$$

$$= \frac{8}{3}$$

Application 6.3.9

Average Value of a Function

Definition

The **uniform distribution** over a region D in $\mathbb{R}^2$ has the joint density function

$$f_{X,Y} = \begin{cases} \frac{1}{\text{area of } D} & \text{if } (x,y) \text{ is inside } D \\ 0 & \text{if } (x,y) \text{ is outside } D \end{cases}$$

Like with single variable function, we default to the uniform distribution whenever we average a function and no specific random variable is specified.

Definition

The **average value** of a function f over a region D is defined to be the expected value of $f(X,Y)$ where X,Y are uniformly distributed over D.

$$f_{ave} = \frac{1}{\text{Area of } D} \iint_D f(x,y)\ dA$$

Since we can also compute the area of D using a double integral, we can also write

$$f_{ave} = \frac{\iint_D f(x,y)\ dA}{\iint_D 1\ dA}$$

One of the most useful things to know about a pair of random variables is whether they are correlated, whether high values of one tend to correspond to high values (or low values) of the other. We can measure this by examining the expected value of a specific function, which is positive when X and Y are both above average or both below average, and negative for pairs when one is above and the other is below.

Definition

The average value of $(X - E[X])(Y - E[Y])$ is called the **covariance** of X and Y, denoted $\text{cov}(X, Y)$.

1. If $\text{cov}(X, Y) > 0$, higher values of X tend to be correlated with higher values of Y.
2. If $\text{cov}(X, Y) < 0$, higher values of X tend to be correlated with lower values of Y.
3. If $\text{cov}(X, Y) = 0$, X and Y are **uncorrelated**.

To test this, we can look at a type of joint distribution whose correlation we already understand. Suppose X and Y are independent. Then outcomes of X should not depend on outcomes of Y. The joint density function can be written $f(x, y) = g(x)h(y)$. We can use our integral rules to see that covariance is always 0, matching our intuition.

$$\begin{aligned}
\text{cov}(X, Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E[X])(y - E[Y]) f_{X,Y}(x, y)\ dydx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E[X])(y - E[Y]) g(x)h(y)\ dydx \\
&= \left(\int_{-\infty}^{\infty} (x - E[X]) g(x)\ dx \right) \left(\int_{-\infty}^{\infty} (y - E[Y]) h(y)\ dy \right) \\
&= \left(\int_{-\infty}^{\infty} x g(x)\ dx - E[X] \right) \left(\int_{-\infty}^{\infty} y h(y)\ dy - E[Y] \right) \\
&= (0)(0)
\end{aligned}$$

Covariance on its own does not allow us to compare whether one joint distribution is better correlated than another. A joint distribution could have a large covariance because the variables are consistently correlated, or because X (or Y) has high variance (meaning X is generally farther from $E[X]$). To control for the latter effect we often compute:

Pearson's Correlation

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

Where the σs are standard deviations.

ρ returns a value between -1 and 1 which is one measure of how well-correlated two random variables are.

Section 6.3

Exercises

Summary Questions

Q1 How do we use a joint density function to compute the probability of a certain set of outcomes?

Q2 What is a marginal density function and how do we compute it?

Q3 What does it mean for two random variables to be independent?

Q4 How can we tell from the graph of a joint density function that the two random variables are independent?

6.3.1

Q5 Given a joint density function $f_{X,Y}(x, y)$, what does

$$\int_0^1 \int_0^x f_{X,Y}(x, y)\, dydx$$

compute?

Q6 Suppose X and Y have the joint density function

$$f_{X,Y}(x, y) = \begin{cases} ax & \text{if } 0 \le x \le 4 \text{ and } 0 \le y \le 5 \\ 0 & \text{otherwise} \end{cases}$$

What is the value of the number a?

6.3.2

Q7 Suppose X and Y have the joint density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{y^2}{18} & \text{if } 0 \leq x \leq 2 \text{ and } 0 \leq y \leq 3 \\ 0 & \text{otherwise} \end{cases}$$

Compute the probability that $X + Y$ is greater than 3.

Q8 Suppose X and Y have the joint density function

$$f_{X,Y}(x,y) = \begin{cases} x+y & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the probability that X and Y differ by at least $\frac{1}{2}$.

Q9 Suppose X and Y have the joint density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{4-x+y}{32\pi} & \text{if } x^2+y^2 \leq 8 \\ 0 & \text{otherwise} \end{cases}$$

a What values of (X,Y) are possible?

b Among the possible values of (X,Y), describe which are more or less likely than others.

c Set up an integral or integrals that would compute the probability that $Y > X$. You don't need to evaluate it.

Q10 Suppose we perform an experiment in which a pair of strangers find an amount of money on the ground. Suppose X and Y are continuous random variables that model the portion of the money ($0 =$none, while $1 =$ all) that each person keeps. Any money not kept is turned into the authorities. Suppose the joint density function of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 24xy & \text{if } x \geq 0, y \geq 0, \text{ and } x+y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

a In a few sentences, interpret what this density function says about which outcomes are likely and which are not. Feel free to include any comments on human nature that you need to get off your chest.

b Set up an integral (or integrals) that computes the probability that each person takes at most twice as much as the other. Do not evaluate.

6.3.3

Q11 Let T be the triangle with vertices $(1,2)$, $(4,0)$ and $(3,5)$. If X and Y are a joint distribution with a density function $f_{X,Y}$ that is nonzero on T and zero everywhere else. For what values of x is the marginal density function $f_X(x)$ nonzero? Illustrate with a diagram.

Q12 Let D be the region between $y = x^2$ and $y = 2x + 15$. If X and Y are a joint distribution with a density function $f_{X,Y}$ that is nonzero on D and zero everywhere else. For what values of y is the marginal density function $f_Y(y)$ nonzero? Illustrate with a diagram.

Q13 Suppose that X and Y are a joint distribution whose density function $f_{X,Y}$ is nonzero in the disk $x^2 + y^2 \leq 1$ and nowhere else. If the marginal density function of X is the density function of a uniform random variable, what does this tell you about where the function $f_{X,Y}(x,y)$ is higher and lower?

Q14 Suppose X and Y have joint density function

$$f_{X,Y} = \begin{cases} g(y) & \text{if } a \leq x \leq b \text{ and } c \leq y \leq d \\ 0 & \text{otherwise} \end{cases}$$

where g is a function only of y. What is the marginal density funtion of X? Justify your answer, preferably without actually evaluating any integrals.

6.3.4

Q15 Suppose the random variables X and Y have the joint density function

$$f_{X,Y}(x,y) = \begin{cases} x + y & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Compute the marginal density function of X.

Q16 Suppse X and Y have joint density function

$$f_{X,Y}(x,y) = \begin{cases} 4xy - 2x - 2y + 2 & \text{if } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

a Compute the marginal density function $f_X(x)$.

b Compute the marginal density function $f_y(y)$.

c What familiar kind of random variables are X and Y?

Q17 Let T be the triangle with vertices $(0,0)$, $(1,0)$ and $(0,1)$. Let X and Y have joint density function

$$f_{X,Y}(x,y) = \begin{cases} 6x & \text{if } (x,y) \text{ is in } T \\ 0 & \text{otherwise} \end{cases}$$

Compute the marginal density function of X.

Q18 Suppose X and Y have joint density function

$$f_{X,Y}(x,y) = \begin{cases} 15y & \text{if } x^2 \leq y \leq x \\ 0 & \text{otherwise} \end{cases}$$

a Draw the region of possible outcomes of (X,Y) in $\mathbb{R}^2$.

b Compute the marginal density function of X.

6.3.5

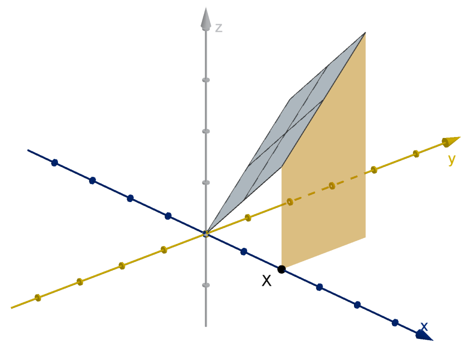
Q19 Suppose X and Y are independent. Their joint density function $f_{X,Y}(x,y)$ has the values

$$f_{X,Y}(3,7) = 0.1 \qquad f_{X,Y}(5,7) = 0.15 f_{X,Y}(5,2) \qquad = 0.21$$

What is $f_{X,Y}(3,2)$?

Q20 How does the distribution of Y change as X takes different values, given the following joint density function?

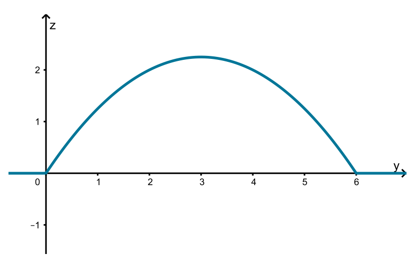
$$f_{X,Y}(x,y) = \begin{cases} x+y & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$



Are X and Y independent?

Q21 Suppose X and Y are independent random variables. If their joint density function $f_{X,Y}(x,y)$ is 0 except on D, what can we say about the shape of D?

Q22 $f_{X,Y}(x,y)$ is a joint density function for a pair of independent variables X and Y. Here is a picture of the $x = 2$ cross section of $z = f_{X,Y}(x,y)$.



a Describe what values of Y are more or less likely when $X = 2$.

b Assume $f_{X,Y}(x,y)$ is not always 0 at $x = 5$. Describe what values of Y are more or less likely when $X = 5$.

c How is the shape of the $x = 2$ cross section of $z = f_{X,Y}(x,y)$ related to the $x = 5$ cross section of $z = f_{X,Y}(x,y)$?

6.3.6

Q23 Let X and Y be random variables with joint density function $f_{X,Y}(x,y)$ and let D be the distance from (X,Y) to the origin. What region would we need to integrate over to compute $P(1 \le D \le 2)$?

Q24 Let X and Y be random variables with joint density function $f_{X,Y}(x,y)$ and let Z be the difference $X - Y$. What region would we need to integrate over to compute $P(0 \le Z \le 5)$?

Q25 Use the expected value formula to show that if Z_1 and Z_2 are both functions of X and Y, then $E[Z_1 + Z_2] = E[Z_1] + E[Z_2]$.

Q26 Let T be the triangle with vertices $(0,0)$, $(4,0)$ and $(0,4)$. Suppose X and Y are random variables with joint density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{8} & \text{if } (x,y) \text{ is in } T \\ 0 & \text{otherwise} \end{cases}$$

Let $Z = X + Y$.

a Write a function $G(z)$ which gives the probability that $Z < z$.

b Compute $g(z) = \frac{dG}{dz}$. Explain why $P(a \le Z \le b) = \int_a^b g(z)\, dz$.

c Use g to directly compute the expected value of Z.

d Compute the expected value of Z instead using our multivariable expected value of a function formula.

6.3.7

Q27 Suppose X and Y have joint density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{12-0.012x}{1000}\left(\frac{y}{x^2} - \frac{y^2}{x^3}\right) & \text{if } 0 \le x \le 1000, 0 \le y \le x \\ 0 & \text{otherwise} \end{cases}$$

Compute $E[Y]$

Q28 Suppose the random variables X and Y have the joint density function

$$f_{X,Y}(x,y) = \begin{cases} x+y & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Compute $E[Y]$.

Q29 Darmok and Jalad's arrival times X and Y have the joint density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{x}{32} & \text{if } 0 \leq x \leq 4 \text{ and } 0 \leq y \leq 4 \\ 0 & \text{otherwise} \end{cases}.$$

What is the expected arrival time of Darmok?

Q30 Suppose the joint density function of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 24xy & \text{if } x \geq 0, y \geq 0, \text{ and } x+y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

. What is the expected value of X?

6.3.8

Q31 Suppose the random variables X and Y have the joint density function

$$f_{X,Y}(x,y) = \begin{cases} x+y & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Compute the expected value of XY.

Q32 Darmok and Jalad's arrival times X and Y have the joint density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{x}{32} & \text{if } 0 \leq x \leq 4 \text{ and } 0 \leq y \leq 4 \\ 0 & \text{otherwise} \end{cases}.$$

Darmok is trying to break his habit of arriving late. He has agreed to donate 120 credits to a local charity for each hour Jalad has to wait for him (prorated across partial hours). Assuming that this incentive has no effect on their arrival times, what is the expected donation?

Q33 Suppse X and Y have joint density function

$$f_{X,Y}(x,y) = \begin{cases} 4xy - 2x - 2y + 2 & \text{if } 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the expected value of X^2Y^2.

Q34 The longitude and latitude of a meteorite landing are random variables X degrees and Y degrees with joint density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{8100-y^2}{349920000} & \text{if } -180 \le x \le 180 \text{ and } -90 \le y \le 90 \\ 0 & \text{otherwise} \end{cases}$$

a. Write an integral that computes the probability that a meteorite lands within 20 degrees longitude of the prime meridian ($x = 0$). Do not evaluate it.

b. What does this density function say about where a meteorite is likely or unlikely to strike? Answer in a few sentences.

c. Suppose a perverse lottery is established that pays out 30 dollars minus the distance in degrees from the south pole ($y = -90$), if the meteorite strikes within 30 degrees of the south pole. Otherwise it pays out nothing. Set up an integral that computes the average payout from this lottery. Do not evaluate.

6.3.9

Q35 Compute the average value of the function $f(x,y) = 2y$ on the unit disc $x^2 + y^2 \le 1$.

Q36 Compute the average value of the function $f(x,y) = y$ on the region enclosed by $y = x^2$ and $y = 16$.

Q37 Compute the average value of the function $f(x,y) = xy$ on the triangle with vertices $(0,0)$, $(4,0)$ and $(0,8)$.

Q38 Compute the average value of the function $f(x,y) = x^2$ on the triangle with vertices $(-2,0)$, $(2,0)$ and $(0,2)$.

6.3.10

Q39 Recall our friends Darmok and Jalad arriving in Tanagra between noon and 4 PM. The joint density function of their respective arrival times (X, Y) is

$$f_{X,Y} = \begin{cases} \frac{x}{32} & \text{if } 0 \leq x \leq 4 \text{ and } 0 \leq y \leq 4 \\ 0 & \text{otherwise} \end{cases}.$$

We found that $E[X] = \frac{8}{3}$ and $E[Y] = 2$. Consider the function $g(X, Y) = \left(X - \frac{8}{3}\right)(Y - 2)$.

- **a** Draw the domain of possible values of (X, Y) At what points in this domain is g positive? Where is it negative?
- **b** Could you argue, using the laws of integrals instead of a computation, that $E[g(X)] = 0$?

Q40 Suppose X and Y have joint density function

$$f_{X,Y}(x, y) = \begin{cases} x + y & \text{if } 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- **a** Compute $E[X]$ and $E[Y]$.
- **b** Compute the covariance of X and Y.
- **c** What does your answer to **b** suggest about how X and Y are correlated?

Synthesis & Extension

Q41 Suppose D is the region enclosed by $6x - x^2$ and the x-axis. Let X and Y are the uniform desntiy function over D. Let $f_X(x)$ be the marginal density function of X.

- **a** What values of X are possible outcomes? What values are impossible?
- **b** What values of X and more likely and what values are less likely? Justify your answer.

Q42 If X and Y have a uniform joint distribution over some region D, can X and Y be correlated? Explain or demonstrate.

Q43 Suppose on a trip to the movies, the number of minutes you wait in line for tickets (X) and the number of minutes you wait in line for snacks (Y) are random variables with joint density function:

$$f_{X,Y}(x,y) = \begin{cases} \frac{12x - x^2 + 10y - y^2}{4880} & \text{if } 0 \leq x \leq 12 \text{ and } 0 \leq y \leq 10 \\ 0 & \text{otherwise} \end{cases}$$

a Are X and Y independent? Justify your answer in a sentence or two.

b Compute the probability that the ticket line takes less than 5 minutes. You don't need to simplify the arithmetic.

c You decide to pay a friend 25 cents per minute to wait in line for snacks while you wait for the tickets. If you're still in line when she gets the snacks, she brings them to you and you pay her. If she's still in line when you get tickets, you pay her and take her place. Write an integral or integrals that compute the expected (average) amount you will pay her. Do not evaluate.

Q44 When you go to the movies, you have to wait in line for tickets, and then to buy snacks. You model the ticket wait (in minutes) with the random variable X. You model the snack wait with the random variable Y. Suppose X and Y have the joint density function

$$f_{X,Y}(x,y) = \begin{cases} 50e^{-5x-10y} & \text{if } 0 \leq x \text{ and } 0 \leq y \\ 0 & \text{otherwise} \end{cases}$$

.

a Compute the probability that you wait a total of no more than 15 minutes in both lines.

b Are the X and Y in this problem indepedent?

c Is the independence of X and Y a reasonable assumpton? Explain.

Q45 Darmok and Jalad have agreed to meet up again at Tanagra. Darmok's arrival time (in hours) after noon is denoted by the random variable X, while Jalad's is denoted by the random variable Y. X and Y have the joint density function

$$f_{X,Y}(x,y) = \begin{cases} \frac{y}{6x^2} & \text{if } 1 \leq x \leq 4 \text{ and } 0 \leq y \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

a Describe the possible arrival times of Darmok and the possible arrival times of Jalad.

b Compute the probability that Darmok arrives at least two hours after Jalad.

c Darmok and Jalad leave Tanagra at exactly 6PM. Write an integral or integrals that compute the average amount of time they spend together at Tanagra. Do not evaluate your integral(s), but your integrand(s) should be functions whose antiderivative(s) are well known.

Q46 Let

$$D = \{(x,y)\ :\ x^2+y^2 \leq 4, y \geq 0\}$$

Suppose X and Y have joint density function

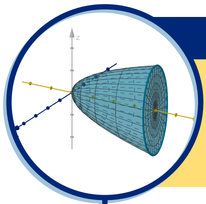
$$f_{X,Y}(x,y) = \begin{cases} \frac{3y}{16} & \text{if } (x,y) \text{ is in } D \\ 0 & \text{otherwise} \end{cases}$$

a Compute the marginal density function of Y

b What integral would compute the expected value of X? How do you know the value of this integral without computing it?

Q47 Suppose we wish to approximate $\int_0^6 x^2\, dx$ by dividing the domain into two equal subintervals. Suppose the test points for each subinterval are independently chosen, uniformly distributed random variables on their respective subintervals. Produce an integral that computes the probability that this approximation overestimates the actual value of the integral.

Q48 Suppose X and Y are independent, and their joint density function is written as a product $f_{X,Y}(x,y) = g(x)h(y)$. How is the marginal density function $f_X(x)$ related to $g(x)$?



Section 6.4

Triple Integrals

Goals:

1. Set up triple integrals over three-dimensional domains.
2. Evaluate triple integrals.

The theory of integrating a two-variable function extends without much trouble to functions of more variables. Visualizing the domains and writing bounds of integration is a much greater challenge. Any function whose domain is a piece of the real world needs (at least) three variables. Joint density functions can also relate any number of random variables. In both cases, a triple integral allows us to aggregate a rate (per unit of volume) to compute a total over the domain in question.

Question 6.4.1

How Do We Integrate a Three-Variable Function?

A triple integral is a natural extension of the double integral. A good exercise is to compare the two definitions, point by point.

Definition

Given a domain D in three dimension space, and a function $f(x, y, z)$. We can subdivide D into regions

- V_i is the volume of the i^{th} region.
- (x_i^*, y_i^*, z_i^*) is a point in the i^{th} region.
- V is the diameter of the largest region.

We define the **triple integral** of f over D to be the following limit over all possible divisions of D:

$$\iiint_D f(x, y, z)\, dV = \lim_{V \to 0} \sum_{i=1}^{n} f(x_i^*, y_i^*, z_i^*) V_i$$

Fubini's theorem applies to triple integrals as well. We write them as interacted integrals.

Question 6.4.1 How Do We Integrate a Three-Variable Function?

Theorem

$$\iiint_D f(x,y,z)dV = \int_{x_1}^{x_2}\int_{y_1}^{y_2}\int_{z_1}^{z_2} Df(x,y,z)\ dzdydx$$

where

- z_1 and z_2 are the bounds of z, which may be functions of x and y.
- y_1 and y_2 are the bounds of y, which may be functions of x.
- x_1 and x_2 are the bounds of x. They are numbers.

The variables of can also be reordered, with the bounds defined analogously.

Example 6.4.2
Integrating Over a Prism

Let $R = \{(x,y,z) : 0 \le x \le 4, 0 \le y \le 2, 0 \le z \le 3\}$. Compute $\iiint_R 3zy + x^2\ dV$.

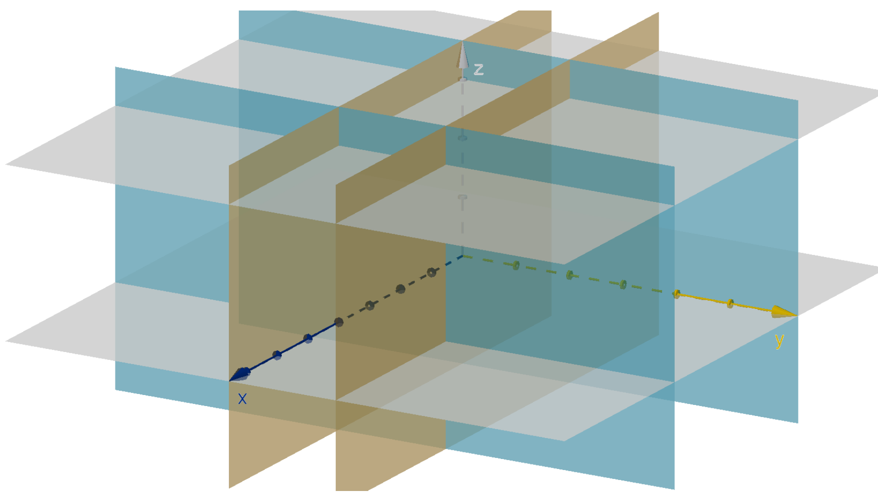


Figure: A Rectangular Prism

Solution

We will set this up as an integral of the form $dzdydx$. The inner integral is dz. No matter where we are in R, we can travel in the positive z direction until we hit $z=3$ or the negative z direction until we hit $z=0$. Thus the inner integral is

$$\int_0^3 3zy + x^2\ dz.$$

Different choices of (x,y) will give different values of this inner integral. The points in R corresponding to a choice of (x,y) are a vertical segment ranging from $z=0$ to $z=3$. These segments exist for any

(x, y) in the rectangle $0 \le x \le 4, 0 \le y \le 2$. We can set up the x and y bounds over this rectangle as we would over a normal double integral. The integrand is the dz integral above. Together we have an iterated integral.

$$\int_0^4 \int_0^2 \int_0^3 3zy + x^2 \; dzdydx$$

We evaluate the inner integral, then the middle, and finally the outer.

$$\begin{aligned}
\int_0^4 \int_0^2 \int_0^3 3zy + x^2 \; dzdydx &= \int_0^4 \int_0^2 \left. \frac{3}{2}yz^2 + x^2 z \right|_0^3 \; dydx \\
&= \int_0^4 \int_0^2 \frac{27}{2}y + 2x^2 \; dydx \\
&= \int_0^4 \left. \frac{27}{4}y^2 + 2x^2 y \right|_0^2 \; dx \\
&= \int_0^4 27 + 4x^2 \; dx \\
&= \left. 27x + \frac{4}{3}x^3 \right|_0^4 \\
&= 108 + \frac{256}{3} \\
&= \frac{580}{3}
\end{aligned}$$

Like a rectangle for double integrals, the right rectangular prism has constant bounds for triple integrals. This is because the bounds of the inner variables remain the same, no matter what values the outer variables take.

Question 6.4.3

How Do We Interpret Triple Integrals Geometrically?

The double integral is the volume under the graph of a two-variable function. This graph lives in three-space. The triple integral is thus a fourth-dimensional volume "under" the graph of a three-variable function. This graph lives in four-space and is thus more problematic to visualize. However we can flatten the fourth dimension into three-space, much like we can flatten the three directions of three-space onto a two-dimensional page. Such a representation loses some information, but can be a useful heuristic. We can examine the role of each iteration in an iterated triple integral through this construction.

$\displaystyle\int_0^3 f(x, y, z) \; dz$ computes the area under the graph $w = f(x, y, z)$ over each vertical segment of the form $(x, y) = (x_0, y_0)$ in the domain. It is a function of x and y.

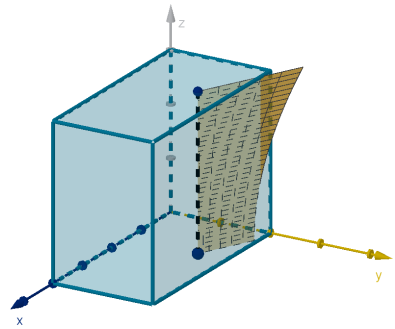


Figure: $\int_0^3 f(x, y, z)\ dz$, represented as an area in a zw-plane

$\int_0^2 \int_0^3 f(x, y, z)\ dzdy$ computes the volume under the graph $w = f(x, y, z)$ over each $x = x_0$ cross-section of the domain. It is a function of x.

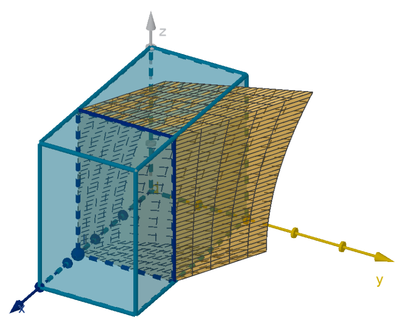


Figure: $\int_0^2 \int_0^3 f(x, y, z)\ dzdy$, represented as a volume in yzw-space

The final integral would require us to represent a fourth-dimensional analogue of volume, which would severely overlap in this visualization.

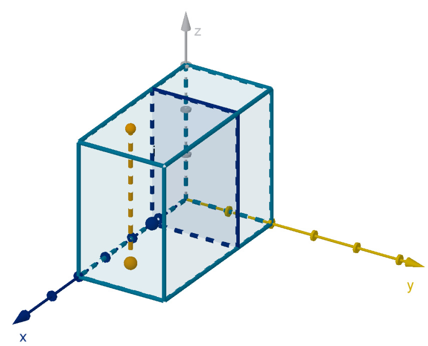
Triple integrals have a variety of applications, largely in physics which tried to model our three-dimensional world.

1. Integrating a function $\rho(x, y, z)$, which gives the density of an object at each point, gives the total mass of the object.
2. Integrating $x\rho(x, y, z)$, $y\rho(x, y, z)$ and $z\rho(x, y, z)$ gives the center of mass of the object.
3. Integrating a three-dimensional probability distribution over a region gives the probability that the triple (X, Y, Z) lies in that region.
4. Integrating $1\ dV$ over a region gives the volume of that region.

Even if we aren't interested in physics, this connection provides us with another visual model for integration. Density lets us visualize a triple integral without referring to a fourth (geometric) dimension.

$\displaystyle\int_0^3 f(x, y, z)\ dz$ computes the density of the vertical segments at each (x, y).

$\displaystyle\int_0^2 \int_0^3 f(x, y, z)\ dzdy$ computes the density of the rectangle at each x.



$\displaystyle\int_0^4 \int_0^2 \int_0^3 f(x, y, z)\ dzdydx$ computes the total mass of the prism.

Remark

Technically the density at each step is a different kind of rate.

- $f(x, y, z)$ represents mass per unit of volume.
- $\displaystyle\int_0^3 f(x, y, z)\ dz$ represents mass per unit of area, since you would need all the segments above some area to produce a volume.
- $\displaystyle\int_0^2 \int_0^3 f(x, y, z)\ dzdy$ represents mass per unit of length, since you would need to stack a segment worth of rectangles to produce a volume.

Example 6.4.5

Integrating Over an Irregular Region

Let R be the region above the xy plane, below the cylinder $x^2+z^2=16$ and between $y=0$ and $y=3$. Compute $\iiint_R 4yz\; dV$.

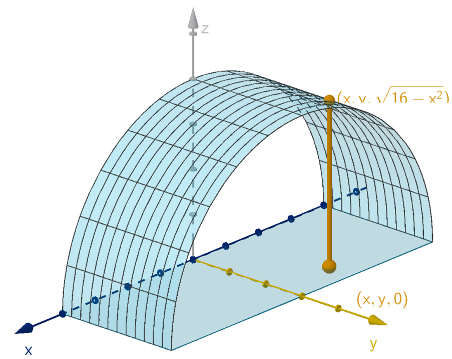


Figure: The region between $x^2+y^2=16$ and the xy-plane

Solution

The words "above" and "below" are useful hints here. "Above the xy-plane" indicates that for each (x, y) the lower bound of z will be on the xy-plane, where $z=0$. "Below the cylinder" indicates that the upper bound of z will satisfy $x^2+z^2=16$ which solves to $z=\pm\sqrt{16-x^2}$. Since these z values are above the xy-plane, the positive branch must be the upper bound. Thus our inner integral is

$$\int_0^{\sqrt{16-x^2}} 4yz\; dz$$

To complete the middle and outer bounds we consider what x and y values lie in R. The lines $y=0$ and $y=3$ suggest bounds for y, but they do not enclose any region. Where else can we get information? Since R is bounded above and below by $y=0$ and above by $y=\sqrt{16-x^2}$, then it is also bounded by where these graphs meet. We can solve for that intersection:

$$\begin{aligned}
0 &= \sqrt{16-x^2}\\
0 &= 16-x^2\\
0 &= (4+x)(4-x)\\
x &= -4 \text{ or } x=4
\end{aligned}$$

Putting this together with the bounds we already have, we see that our x and y bounds are rectangular. We set them up as we would in a double integral and put the inner integral as an integrand:

$$\int_{-4}^{4}\int_0^3\int_0^{\sqrt{16-x^2}} 4yz\; dzdydx$$

We now turn to evaluating the integral. Having a function of x in our z-bounds should be familiar from double integrals.

$$\begin{aligned}
\int_{-4}^{4}\int_{0}^{3}\int_{0}^{\sqrt{16-x^2}} 4yz\ dzdydx &= \int_{-4}^{4}\int_{0}^{3} 2yz^2\Big|_{0}^{\sqrt{16-x^2}}\ dydx \\
&= \int_{-4}^{4}\int_{0}^{3} 2y(\sqrt{16-x^2})^2\ dydx \\
&= \int_{-4}^{4}\int_{0}^{3} 32y-2yx^2\ dydx \\
&= \int_{-4}^{4} 16y^2-x^2y^2\Big|_{0}^{3}\ dx \\
&= \int_{-4}^{4} 144-9x^2\ dx \\
&= 144x-3x^3\Big|_{-4}^{4} \\
&= 576-192-(-576+192) \\
&= 768
\end{aligned}$$

Main Idea

The following approach will produce the bounds of a region with a top surface and a bottom surface.

1. The z bounds are given by the equations $z=f(x,y)$ and $z=g(x,y)$ of the top and bottom surface.
2. The intersection of the top and bottom surface can produce relevant bounds on x and y. We can graph these, along with any given bounds involving x and y.
3. After drawing the bounded region in the xy-plane, the x and y bounds are computed as for a double integral.

Like with double integrals, we will want to break the region into smaller pieces in some cases. In other cases, we may want to change the order of integration.

Example 6.4.6
A Solid Given by Vertices

Suppose we want to integrate over T, the tetrahedron (pyramid) with vertices $(0,0,0)$, $(4,0,0)$, $(4,2,0)$ and $(4,0,2)$. How would we set up the bounds of integration?

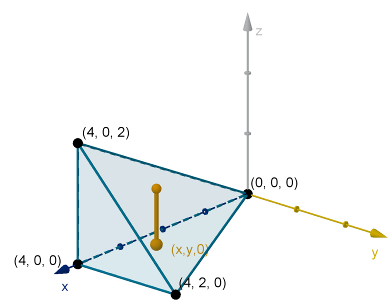


Figure: z bounds of T

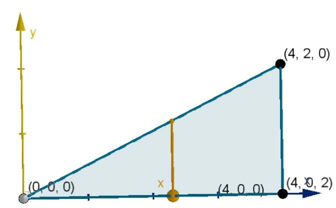


Figure: x, y bounds of T

Solution

In this case, it is helpful to draw a diagram of the tetrahedron in three-space. First we examine the inner integral. The bounds of z are functions of (x,y). Visually, we want to imagine the vertical segments lying in different parts of T and ask where their upper and lower endpoints lie. No matter which veritcal segment we pick, its lower endpoint in on the xy-plane and its upper endpoint is on the triangle with vertices $(0,0,0)$, $(4,2,0)$ and $(4,0,2)$.

The xy-plane gives us a lower bound $z=0$. The upper bound triangle also lies in a plane. Every upper endpoint lies in this plane, so its z coordinates must satisfy the equation of that plane. This plane has a z-intercept of 0 since $(0,0,0)$ is a vertex of the triangle. We can solve for the slopes and write the equation.

$$\begin{aligned} m_x &= \frac{2-0}{4-0} = \frac{1}{2} \\ m_y &= \frac{2-0}{0-2} = -1 \\ z &= \frac{1}{2}x - 1y + 0 \\ z &= \frac{1}{2}x - y \end{aligned}$$

Our inner integral is

$$\int_0^{\frac{1}{2}x-y} f(x,y,z)\;dz$$

To find the outer bounds, we ask what values of (x,y) lie in T? Every point in T lies directly above the triangle with vertices $(0,0,0)$, $(4,2,0)$ and $(4,0,0)$. Thus its (x,y) coordinates match those of a point in the triangle. We can draw this triangle in the xy-plane and set up the bounds of a double integral over it. The result is

$$\int_0^4 \int_0^{\frac{1}{2}x} \int_0^{\frac{1}{2}x-y} f(x,y,z)\ dzdydx$$

Main Idea

In the case of a polyhedron given by vertices, we generally need to plot the vertices and draw the faces to discern the upper and lower z bounds. The equations of these bounds are planes. We can then draw the set of possible (x,y) in two-space and proceed as in a double integral.

Example 6.4.7

Changing the Order of Integration

Suppose D is the bounded region enclosed between the graph of $y = 4x^2 + z^2$ and the plane $y = 4$. Set up the bounds of the integral $\iiint_D f(x,y,z)dV$.

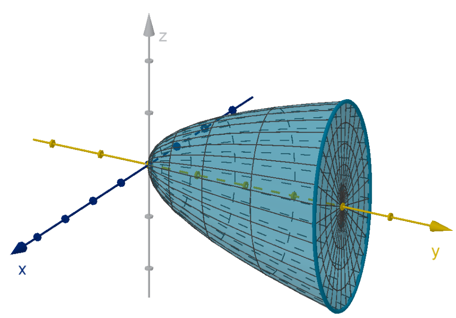


Figure: A region bounded by a paraboloid and a plane

Example 6.4.7 Changing the Order of Integration

Solution 1

We can begin by finding bounds for the z, the inner variable. The plane $y = 4$ does not have a z. z is a free variable and thus the plane extends in the z direction, and cannot be the top or bottom of a vertical segment. On the other hand $y = 4x^2 + z^2$ solves to $z = \pm\sqrt{y - 4x^2}$. Since this gives a plus and a minus branch, it can provide both the upper and lower bound of z. The inner integral is

$$\int_{-\sqrt{y-4x^2}}^{\sqrt{y-4x^2}} f(x, y, z)\ dz$$

For xy-bounds we have equation $y = 4$, but this does not bound any region. We can search for additional bounds by seeing where the top surface meets the bottom. We'll use the fact that a square root can only equal a negative square root if both are 0.

$$\begin{aligned} -\sqrt{y - 4x^2} &= \sqrt{y - 4x^2} \\ y - 4x^2 &= 0 \\ y &= 4x^2 \end{aligned}$$

We can add this parabola to a graph. We set up our xy bounds using our usual method for double integrals. The graphs $y = 4x^2$ and $y = 4$ intersect at $x = \pm 1$. Between $x = -1$ and $x = 1$, $y = 4$ is greater than $y = 4x^2$. Here are the bounds.

$$\int_{-1}^{1} \int_{4x^2}^{4} \int_{-\sqrt{y-4x^2}}^{\sqrt{y-4x^2}} f(x, y, z)\ dzdydx$$

The bounds in this solution look difficult to work with. For example, in the first step, we'll plug $\sqrt{y - 4x^2}$ in for z in the antiderivative of f. The resulting integrand would be even more difficult to work with. We can improve this situation somewhat by choosing a different variable for our inner integral.

Solution 2

Since both bounds are already solved for y, we will use y as our inner variable. We can test which is the upper and which is the lower bound with a test point, but we don't yet know which x and z values lie in the region. We do not have any x or z bounds that don't involve y, so we set the y-bounds equal to each other.

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

We may recognize this as an ellipse. Even if we do not, we can proceed at usual for a double integral, except that our variables are x and z. We will use z is the inner variable and solve the bound for z.

$$\begin{aligned} 4x^2 + z^2 &= 4 \\ z^2 &= 4 - 4x^2 \\ z &= \pm\sqrt{4 - 4x^2} \end{aligned}$$

These give upper and lower bounds for z. To find x bounds, we solve for where the z-bounds intersect.

$$-\sqrt{4-4x^2} = \sqrt{4-4x^2}$$

$$4-4x^2 = 0 \qquad \text{a sqaure root equals its nagative only at } 0$$

$$4(1-x)(1+x) = 0$$

$$x = -1 \text{ or } x = 1$$

This gives us the bounds of the outer integrals:

$$\int_{-1}^{1} \int_{-\sqrt{4-4x^2}}^{\sqrt{4-4x^2}} \ldots dz dx$$

We still need a test point for the y bounds. We can choose $x = 0$ since that is between -1 and 1. We can choose $z = 0$ since that is between $-\sqrt{4-4(0)^2}$ and $\sqrt{4-4(0)^2}$. We plug them into both y bounds and see that $y = 4$ is the upper bound.

$$y = (4)(0)^2 + 0^2 \qquad \text{vs} \qquad y = 4$$

Our final integral is

$$\int_{-1}^{1} \int_{-\sqrt{4-4x^2}}^{\sqrt{4-4x^2}} \int_{4x^2+z^2}^{4} f(x,y,z)\ dydzdx.$$

We still have difficult z bounds under this method, but we delay plugging them in until the second step, which means they may cause less trouble for us.

Main Idea

When setting up a triple integral bounded by graphs, it may be more convenient to use an inner variable that has a less complicated relationship with the bounding equations.

Question 6.4.8

When Does a Triple Integral Decompose as a Product?

The product theorem from double integrals also works here:

Theorem

If y_1, y_2, z_1 and z_2 are constants, then

$$\int_{x_1}^{x_2} \int_{y_1}^{y_2} \int_{z_1}^{z_2} f(x)g(y)h(z)\ dzdydx$$

$$= \left(\int_{x_1}^{x_2} f(x)\ dx\right) \left(\int_{y_1}^{y_2} g(y)\ dy\right) \left(\int_{z_1}^{z_2} h(z)\ dz\right)$$

Question 6.4.8 When Does a Triple Integral Decompose as a Product?

Example

Along with the sum and constant multiple rules we can simplify

$$\int_0^4 \int_0^2 \int_0^3 3zy + x^2 \; dzdydx$$

to obtain the following:

$$\begin{aligned}
&\int_0^4 \int_0^2 \int_0^3 3zy \; dzdydx + \int_0^4 \int_0^2 \int_0^3 x^2 \; dzdydx \\
=&3 \int_0^4 dx \int_0^2 y \; dy \int_0^3 z \; dz + \int_0^4 x^2 \; dx \int_0^2 dy \int_0^3 dz \\
=&3 \cdot 4 \int_0^2 y \; dy \int_0^3 z \; dz + 2 \cdot 3 \int_0^4 x^2 \; dx
\end{aligned}$$

Section 6.4 Exercises

Summary Questions

Q1 What does Fubini's theorem say about integrals with dV?

Q2 How is density used to understand triple integrals. Why wasn't it necessary or appropriate for double integrals?

Q3 How do you find the bounds of the inner variable in a triple integral?

Q4 How to you find the bounds of the other two variables?

6.4.1

Q5 Suppose we want to approximate $\int_0^3 \int_2^{10} \int_{-2}^{2} f(x,y,z)\ dzdydx$ by subdividing the domain of integration into 12 sub-prisms of equal volume. What will V be?

Q6 Let R be a cube of side length 4, with edges parallel to the x-, y- and z-axes, and with vertices $(0,0,0)$ and $(4,4,4)$. Suppose we want to approximate $\iiint_R xyzdV$ using a subdivsion of R into 8 identical cubes.

- **a** What will V be?
- **b** What test points would you use to make your approximation as large as possible.
- **c** Produce the smallest possible approximation using this subdivision.

6.4.2

Q7 Is $\int_0^4 \int_0^7 \int_0^2 f(x,y,z)\ dzdydx = \int_0^7 \int_0^2 \int_0^4 f(x,y,z)\ dzdydx$? Explain.

Q8 Set up the bounds of integration of a function $f(x,y,z)$ over the a general prism

$$P = \{(x,y,z)\ :\ x_0 \leq x \leq x_1, y_0 \leq y \leq y_1, z_0 \leq z \leq z_1\}$$

Q9 Evaluate $\int_0^2 \int_0^2 \int_0^3 (x+y)z\ dzdydx$.

Q10 Evaluate $\int_0^5 \int_0^{11} \int_{-1}^{1} ye^{2x+z}\ dzdydx$.

6.4.3

Q11 In a triple integral, the inner integral $\int_{z_0}^{z_1} f(x,y,z)\,dz$ is a function of x and y, while $\int_{y_0}^{y_1}\int_{z_0}^{z_1} f(x,y,z)\,dzdy$ is a function of only x.

a Explain why this occurs algebraically.

b Explain why this makes sense given the context of an iterated triple integral.

Q12 In each of the following questions, assume x, y, z, and w are the variables of four-space.

a What is the dimension of the set of points that satisfy $x = x_0$?

b What is the dimension of the set of points that satisfy both $x = x_0$ and $y = y_0$?

Q13 Give the area of the $x = 4$ and $y = 1$ cross-section of the region "under" the graph of $w = x + ye^z$ and "above" the prism

$$P = \{(x,y,z) \;:\; 0 \le x \le 6, 0 \le y \le 4, 0 \le z \le 3\}$$

Q14 Give the volume of the $x = 2$ cross-section of the region "under" the graph of $w = \frac{z^2\sqrt{13-x^2}}{y}$ and "above" the prism

$$P = \{(x,y,z) \;:\; 0 \le x \le 3, 1 \le y \le 2, -3 \le z \le 3\}$$

6.4.4

Q15 A prism of length ℓ, width w and height h can be defined by the inequalities

$$0 \le x \le \ell$$
$$0 \le y \le w$$
$$0 \le z \le h$$

Set up a triple integral to compute the volume of this prism. Verify that the value of this integral matches the well-known volume formula, $V = \ell wh$.

Q16 Denser matter tends to sink to the bottom of a container. After sitting undisturbed for several days, the density of a soil sample in the box

$$P = \{(x, y, z) \ : \ 0 \le x \le 5, 0 \le y \le 4, 0 \le z \le 2\}$$

is given by $\rho(x, y, z) = e^{-z/10}$. Find the total mass of the soil in the box.

Q17 Xavier, Yolanda and Zoe's respective arrival times (in hours after noon) at a restaurant are given by joint random variables X, Y and Z. The joint density function of X, Y and Z is

$$f_{X,Y,Z}(x, y, z) = \begin{cases} \frac{12}{11}(1 - x^2yz) & \text{if } 0 \le x \le 1, 0 \le y \le 1, 0 \le z \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Compute the probability that they all arive by $12 : 15$.

Q18 Random variables X, Y and Z are **uniform** if their density function has the form

$$f_{X,Y,Z}(x, y, z) = \begin{cases} \frac{1}{V} & \text{if } (x, y, z) \text{ is in } R \\ 0 & \text{otherwise} \end{cases}$$

where V is the volume of R. If X, Y and Z are uniform on

$$R = \{(x, y, z) \ : \ 0 \le x \le 10, 0 \le y \le 10, 0 \le z \le 10\},$$

compute $P(X \le 4 \text{ and } Z \ge 3)$.

6.4.5

Q19 Let R be the region given by $x^2 + y^2 + z^2 \le 25$.

- **a** Describe R geometrically.
- **b** Set up the bounds of integration for $\iiint_R f(x, y, z)\ dV$.
- **c** If we plug in the function $f(x, y, z) = 1$ do you happen to know the value of this integral?

Q20 Cheng is integrating over R, the region given by $x^2 + y^2 + z^2 \le 25$. He gives the following setup. Is this valid?

$$\int_{-\sqrt{25-y^2-z^2}}^{\sqrt{25-y^2-z^2}} \int_{-\sqrt{25-x^2-z^2}}^{\sqrt{25-x^2-z^2}} \int_{-\sqrt{25-x^2-y^2}}^{\sqrt{25-x^2-y^2}} f(x, y, z)\ dzdydx$$

Q21 Consider the domain

$$D : \{(x, y, z) \ : \ y \geq 0, y \leq -x, z \geq 9, z \leq 25 - x^2 - y^2\}$$

a Set up the bounds of $\iiint_D x \ dV$. Do not evaluate.

b Do you expect the integral in a to be positive, negative, or zero? In a sentence or two, explain how you know without computing it.

Q22 Let $R = \{(x, y, z) \ : \ z \leq 2x - y, \ z \geq 0, \ y \geq x^2\}$. Compute

$$\iiint_R xz \ dV.$$

Q23 Let R be the region enclosed by the graphs $z = x^2 + y^2$ and $2y - z = 0$. Set up the bounds for $\iiint_R (y - 1) \ dV$. Do not evaluate.

Q24 Set up a triple integral that will compute the volume enclosed by the planes $x = 0$, $x = 5$, $y = 0$, $z = 2y$ and $z = 6$. Do not evaluate.

Q25 Let R be the region enclosed by $z = x^2$, $z = 16$, $y = 2$ and $y = 6$. Set up and evaluate $\displaystyle\iiint_R x + z \ dV$.

Q26 Let R be the region enclosed by $y = \sqrt{25 - x^2}$, $z = 6 - y$ and $z = \sqrt{y}$. Set up the bounds of $\iiint_R g(x, y, z) dV$.

6.4.6

Q27 Let P be a square pyramid with vertices $(0, 0, 0)$, $(2, 0, 0)$, $(2, 2, 0)$, $(0, 2, 0)$ and $(0, 0, 4)$.

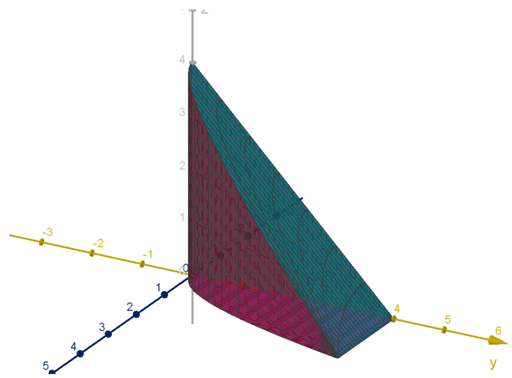
a Explain why it might not be a good idea to use z as the inner variable when setting up the bounds of $\displaystyle\iiint_P f(x, y, z) \ dV$.

b Set up the bounds using a different inner variable.

Q28 Set up the bounds of integration of $\iiint_T f(x,y,z)\ dV$, where T is a tetrahedron with vertices $(0,0,0)$, $(8,0,0)$, $(0,6,0)$ and $(0,0,3)$.

6.4.7

Q29 Let R be the region over the first quadrant enclosed by $y = x^2$, $x = 0$, $z = 0$ and $z = 4 - y$.



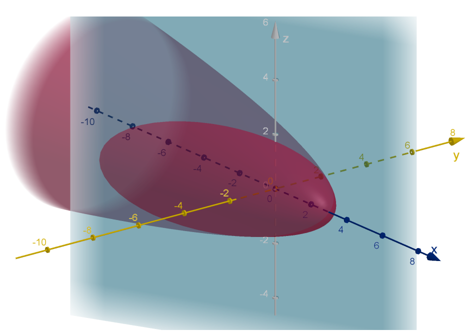
Set up the integral $\iiint_R f(x,y,z)\ dV$

a with z as the inner variable

b with y as the inner variable

c with x as the inner variable

Q30 Let R be the region enclosed by the paraboloid $x = 3 - y^2 - z^2$ and the plane $y = \frac{1}{2}x$.



a Set up the integral $\iiint_R f(x,y,z)\ dV$ with z as the inner variable.

b Set up the integral $\iiint_R f(x,y,z)\, dV$ with x as the inner variable.

c Explain why it would be difficult to set up $\iiint_R f(x,y,z)\, dV$ with y as the inner variable.

Q31 Let P be the prism whose base has vertices $(0,0,0)$, $(0,5,0)$ and $(0,0,-2)$ and whose height is 4 units in the direction of the positive x axis. Set up a triple integral $\iiint_P g(x,y,z)\, dV$ in three different ways, using three different inner variables.

Q32 Let P be the trapezoidal prism with vertices $(0,0,0)$, $(0,6,0)$, $(0,4,2)$, $(0,0,2)$, $(5,0,0)$, $(5,6,0)$, $(5,4,2)$, and $(5,0,2)$. Set up the bounds of integration of $\displaystyle\iiint_P h(x,y,z)\, dV$ without writing it as a sum or difference of multiple integrals.

Q33 Consider the tetrahedron T whose vertices are $(0,0,0)$, $(0,0,4)$, $(0,6,3)$, $(2,6,3)$. Which variable(s) could you use as the inner variable of a triple integral over T without having to break the domain into two or more pieces.

Q34 Set up (but do not evaluate) one or more integrals of $f(x,y,z)$ over the region

$$R = \{(x,y,z)\ :\ z \ge 0,\ x \ge y^2+z^2,\ x+2z \le 8\}$$

Use $dxdydz$ as your order of integration.

Q35 Rewrite the integral $\displaystyle\int_0^1 \int_{x^2}^{x} \int_0^{x-y} f(x,y,z)\, dzdydx$ as an integral with the differential $dxdzdy$.

Q36 Rewrite the integral $\displaystyle\int_0^2 \int_{2-x}^{2} \int_0^{4-x^2} f(x,y,z)\, dzdydx$ as an integral with the differential $dxdzdy$.

6.4.8

Q37 Use product and sum rules to decompose $\displaystyle\int_3^4 \int_0^8 \int_{-1}^1 y^2\sin x - e^{y+z}\, dzdydx$ into an expression containing only single integrals.

Q38 Let $S = \{(x,y,z)\ :\ x^2+y^2+z^2 \le 25\}$. Explain why $\displaystyle\iiint_S x^3y^4\cos\pi z\, dV$ cannot be decomposed as a product.

Synthesis & Extension

Q39 Consider the domain

$$D : \{(x, y, z) \ : x - 16 \leq y \leq 2, x^2 + y^2 \leq z \leq x^2 + x + 4\}$$

a Set up the bounds of $\iiint_D xyz \, dV$. You may use one or more integrals to do so. Do not evaluate.

b Does the function $f(x, y, z) = xyz$ have a maximum value on D? Justify your answer with a theorem, and verify that the theorem does or does not apply.

Q40 Let S be the region above $z = 0$ and below the graph $z = f(x, y)$ over the rectangle

$$R = \{(x, y) \ : \ a \leq x \leq b, c \leq y \leq d\}.$$

a Write the volume of S as a double integral.

b Write the volume of S as a triple integral.

c Show that if you evaluated your answer to **b**, your answer to **a** would be one of the step of this computation.

Q41 Suppose that R is the solid obtained by rotating the region under $y = f(x)$ from $x = a$ to $x = b$ around the x-axis. Write a triple integral that computes the volume of R.