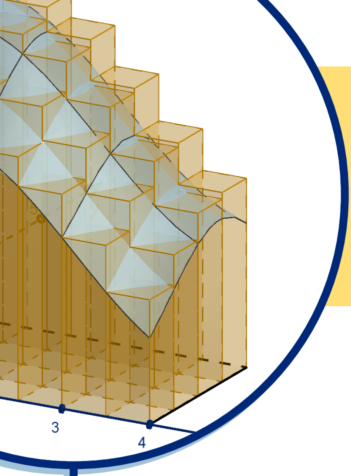
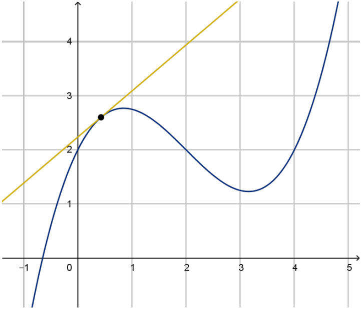


Introduction

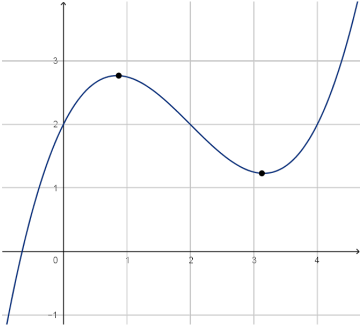
Advanced Calculus for Data Science

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

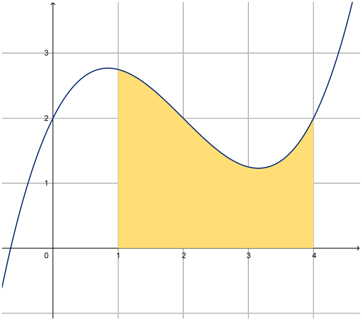
1. How quickly does the value of the function change as the input changes?
2. How do we estimate the value of the function near a point?



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 will develop methods to approximate integrals and to approximate functions.

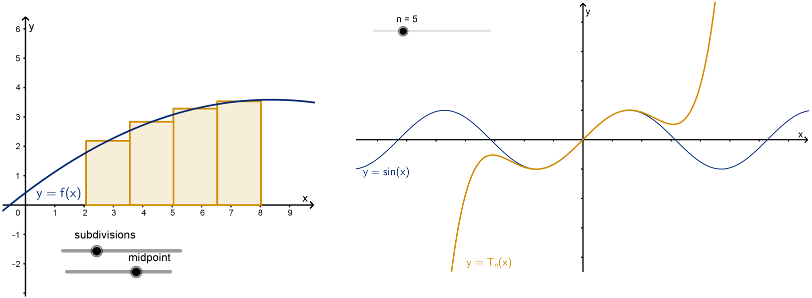


Figure: Approximations of an integral and of a function

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: $P = P_0 e^{rt}$

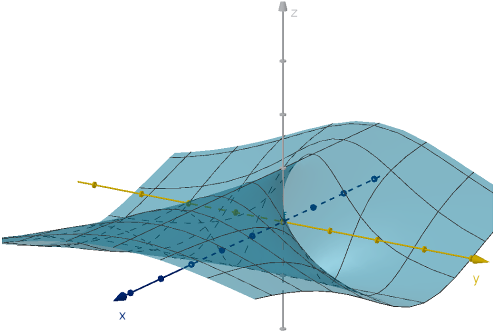


Figure: The graph of a two-variable function

Furthermore, real world data does not come prepackaged with a differentiable function to describe it.

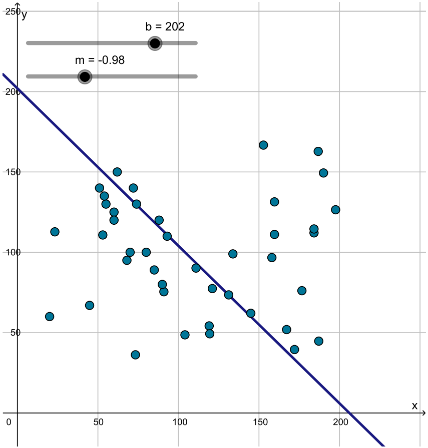


Figure: Fitting a line to a set of data points

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.

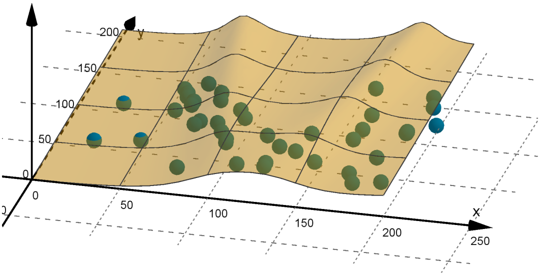
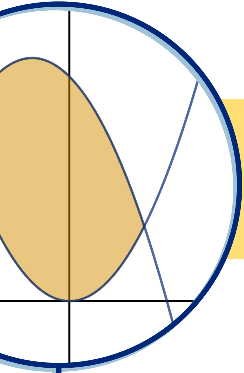


Figure: A function that models the outcomes of a random process

Course Aims

By the end of this course, you should be able to:

- Apply advanced methods to evaluate integrals.
- Measure areas and volumes with integrals.
- Implement code to compute integrals and derivatives and to visualize functions.
- Compute a probability using a continuous probability distribution.
- Approximate or manipulate a function using its Taylor Series.
- Produce or interpret a variety of visualizations of multivariable functions.
- Compute rates of change of multivariable functions.
- Find maximum and minimum values of a multivariable function, including with constraint.
- Integrate multivariable functions over a variety of regions.



Section 2.1

Area Between Curves

Goals:

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

How Is the Integral Related to Geometric Area?

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

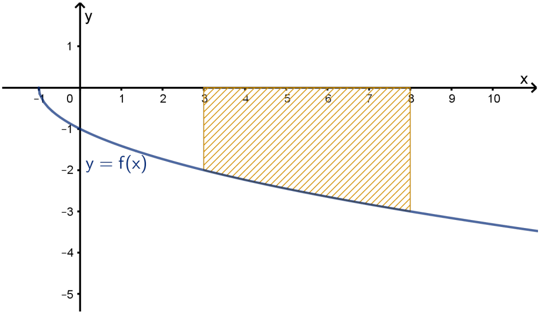


Figure: A region below the x-axis and above $y = f(x)$

Why does this happen?

Definition

The integral is computed by the following limit

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

When $f(x) < 0$, the product $f(x_i^*)\Delta x$ computes a negative "area" for each rectangle.

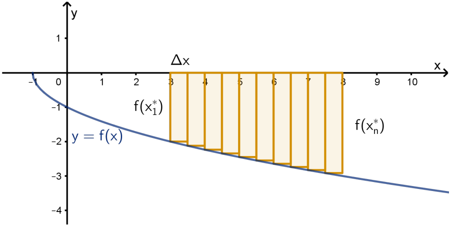


Figure: An approximation by rectangles of negative height

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.

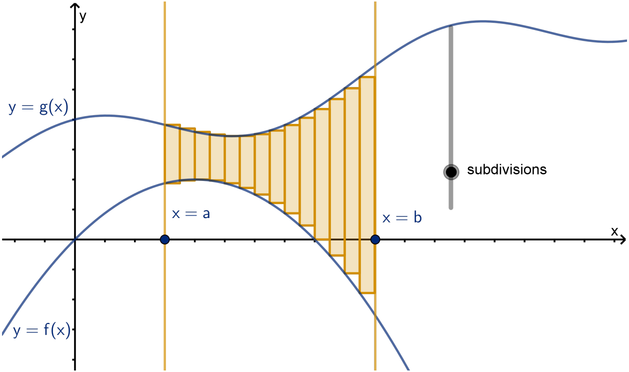
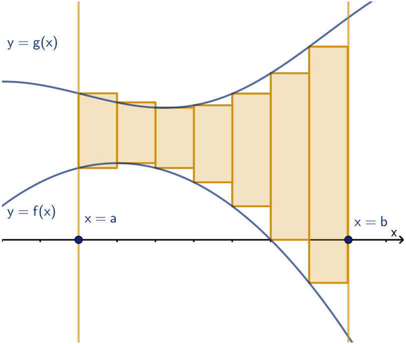


Figure: The region between $y = f(x)$ and $y = g(x)$, approximated by rectangles

Let's derive a formula for this rectangle approximation.



Main Idea

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

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

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?

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?

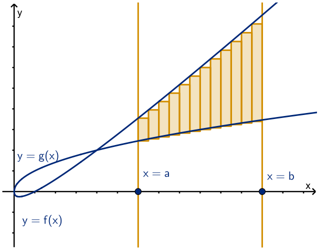


Figure: An approximation of the area between $y = x - \sqrt{x}$ and $y = \sqrt{x}$

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.

The Area Enclosed by Two Curves

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

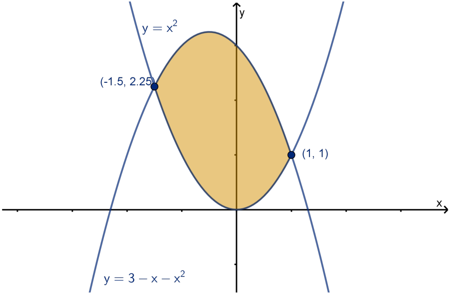


Figure: The area enclosed by two parabolas

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 $g(x) = 3x^2$.

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.

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.

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.

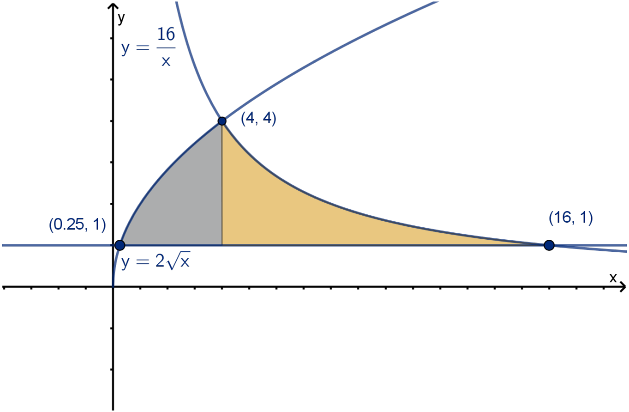
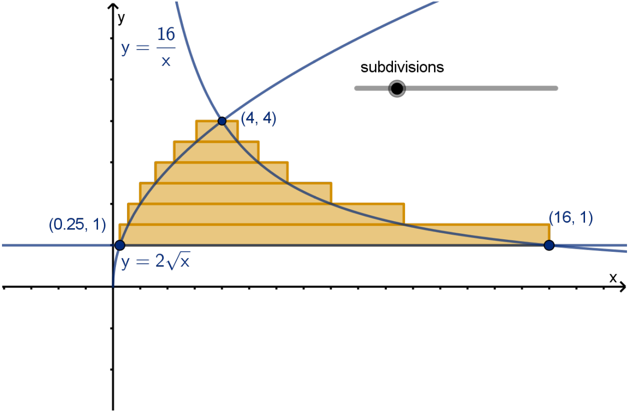


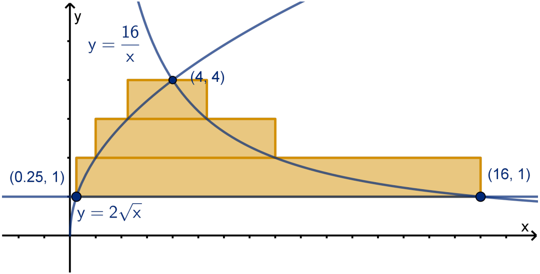
Figure: Two subregions whose areas can be expressed by integrals

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.



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?

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

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

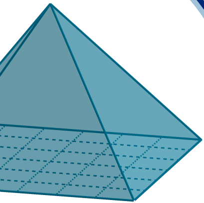
- a Who is correct?
- b How do you think the mistake could reasonably have happened, and how can you avoid it?

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.

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.

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.

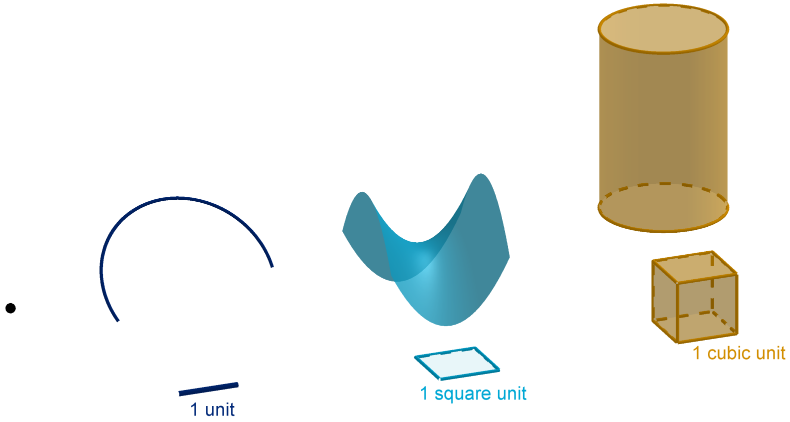


Figure: Objects of several dimensions and their units of measurement

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.

Formula for Volume of a Prism

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

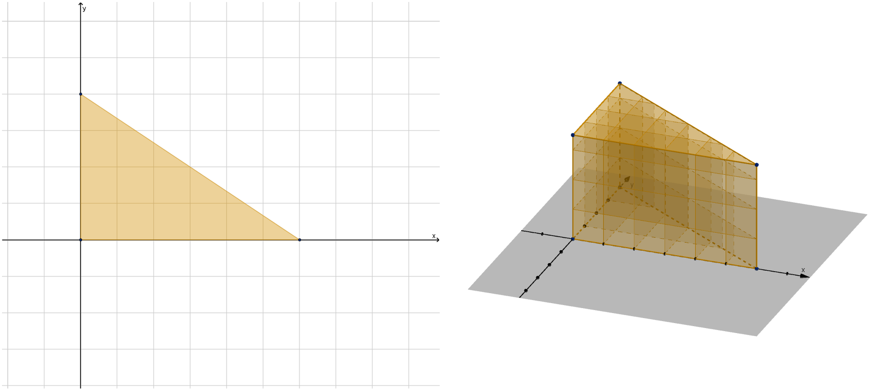


Figure: A prism divided into unit cubes and its base divided into unit squares.

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.

How Do We Visualize 3-Dimensional Solids?

Definition

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

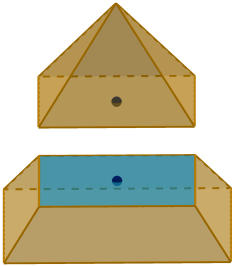


Figure: A cross section of a pyramid

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.

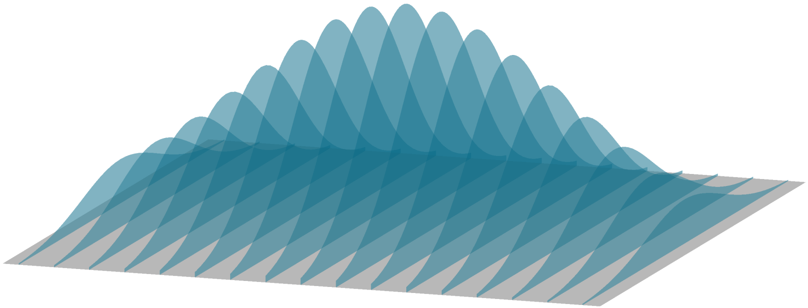


Figure: A set of parallel cross sections of a solid

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.

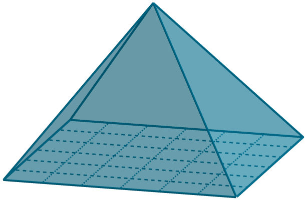


Figure: A pyramid with its base divided into unit squares

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

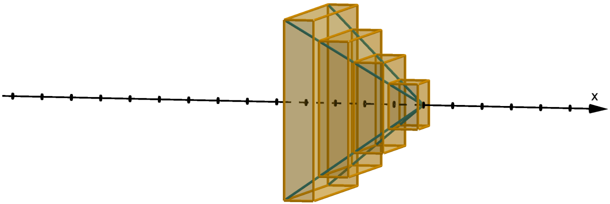


Figure: A pyramid approximated by prisms

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.

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.

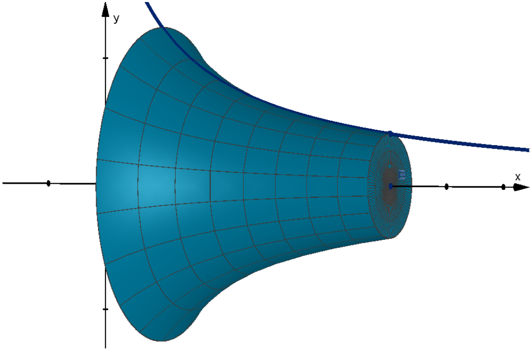


Figure: The solid obtained by rotating the region under $y = \frac{5}{x+1}$ about the x-axis

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 $\pi[f(x)]^2$.

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.

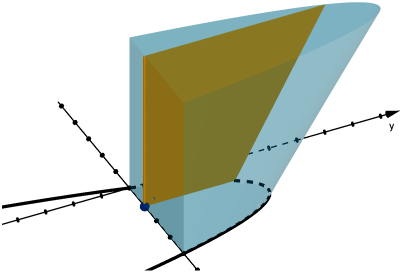


Figure: A solid with base between two graphs and trapezoidal cross-sections

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 $A(x)$.

A Solid Described by Measurements

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

Summary Questions

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

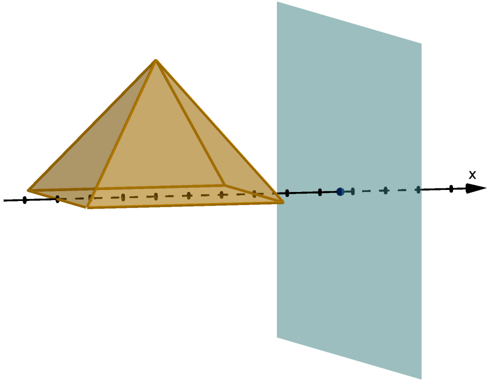
Q2 What is the significance of the function $A(x)$ in the formula for the volume of a solid?

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?

Q10

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



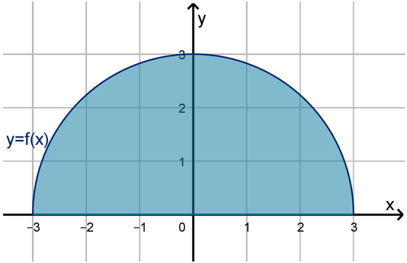
Q26

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.

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.



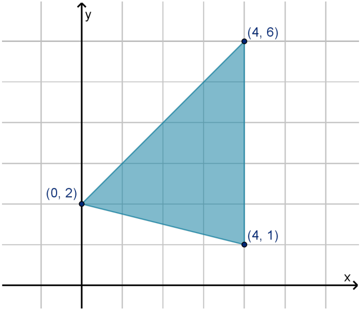
Q30

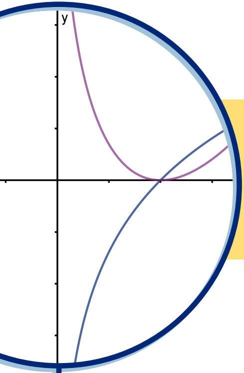
Consider the solid obtained by rotating the triangle below around the x-axis.

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.





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.

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**.

Example

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

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.

There is another differentiation rule that produces products.

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

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?

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

Theorem

Suppose an integral can be written $\int u \; dv$ 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:

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

Example 2.3.2

Computing an Anti-derivative Using Integration by Parts

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

How Do We Choose u and dv?

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

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

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.

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. $\displaystyle\int x^5 \ln x \; dx$
2. $\displaystyle\int x \sin x \; dx$
3. $\displaystyle\int x^2 \tan^{-1}(x) \; dx$

Using Integration by Parts More than Once

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

Change of Variables?

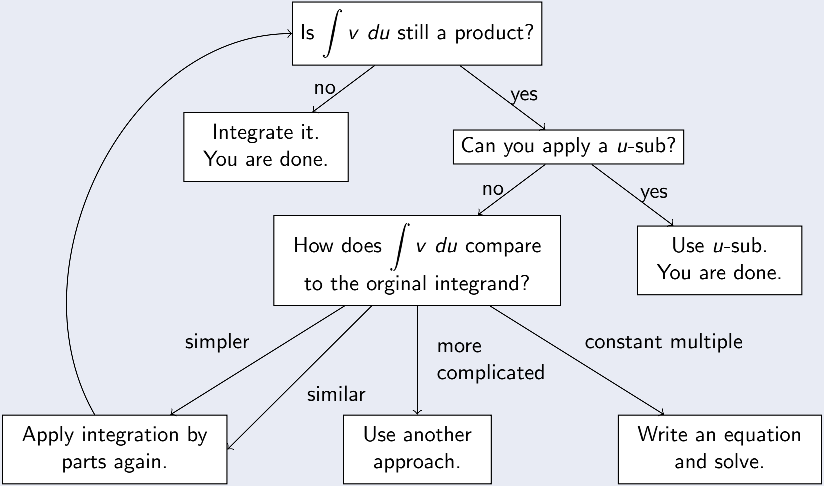
Notice that despite defining functions u and v, we continue to work in terms of the variable x. Contrast this with u-substitution where the variable x can be completely eliminated in a definite integral. That approach isn't possible here. We'd have to write v as a function of u. This would be complicated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

Using Integration by Parts to Produce an Equation

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

Main Idea

The success of integration by parts depends on the $\int v \, du$ term.



Summary Questions

- **Q1** What type of integrands are good candidates for integration by parts?
- **Q3** How is u handled differently in integration by parts than in u-substitution?
- **Q5** How is the acronym I.L.A.T.E. used?
- **Q7** Under what conditions would we want to apply integration by parts more than once?

Q6

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

$$\int e^x\,dx \quad \int xe^x\,dx \quad \int x^2e^x\,dx \quad \int x^3e^x\,dx$$

$$\int e^{x^2}\,dx \quad \int xe^{x^2}\,dx \quad \int x^2e^{x^2}\,dx \quad \int x^3e^{x^2}\,dx$$

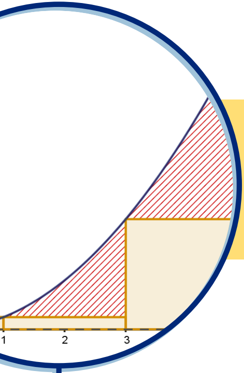
$$\int e^{x^3}\,dx \quad \int xe^{x^3}\,dx \quad \int x^2e^{x^3}\,dx \quad \int x^3e^{x^3}\,dx$$

$$\int e^{x^4}\,dx \quad \int xe^{x^4}\,dx \quad \int x^2e^{x^4}\,dx \quad \int x^3e^{x^4}\,dx$$

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.



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.

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^*.

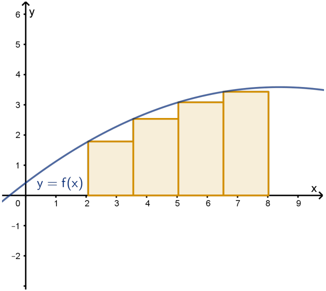
Notation

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^*.

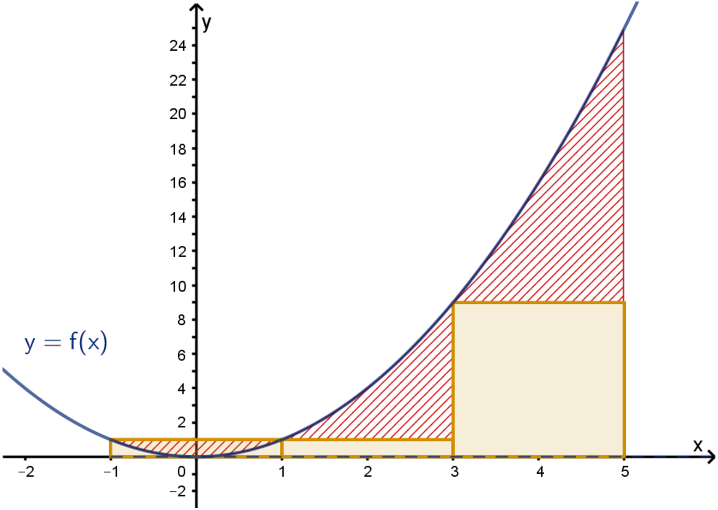


L_4 approximation

Computing an L_n Approximation

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

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



y
24
22
20
18
16
14
12
10
8
6
4
2
0
−2
y = f(x)
x
−2
−1
1
2
3
4
5

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.

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?

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?

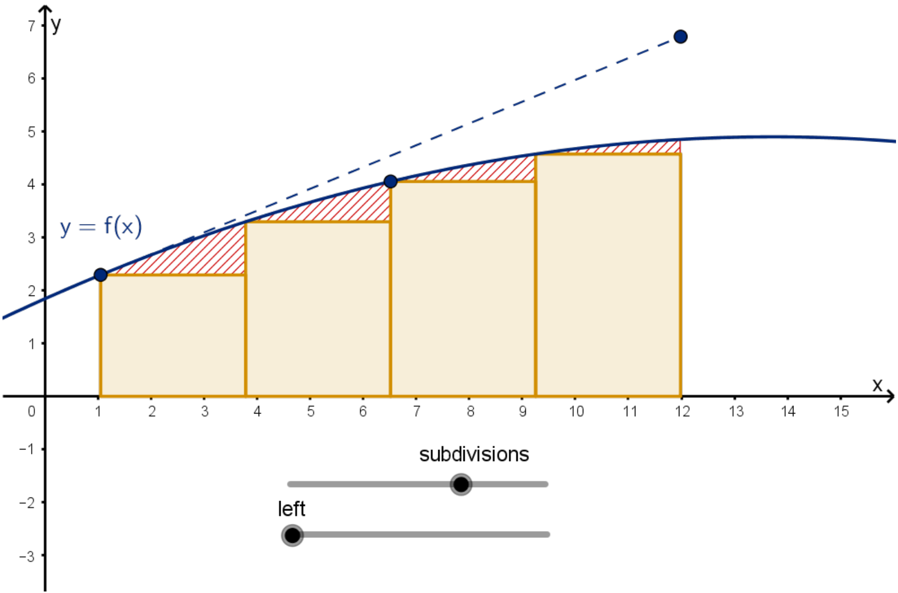


Figure: The error of an L_n approximation

Let's use the results of the exercise to formulate an error bound for L_n.

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

$$|E_L| \leq \frac{S(b-a)^2}{2n} \qquad \text{and} \qquad |E_R| \leq \frac{S(b-a)^2}{2n}$$

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?

How Can We Make our Approximation Less Sensitive to Slope?

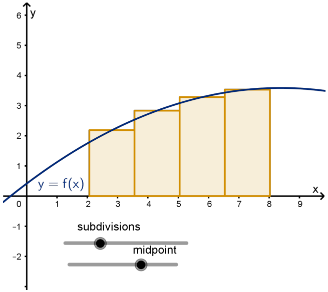
L_n and R_n have large errors when function is increasing or decreasing rapidly. We'll examine two approximations that are more resilient. The first is the midpoint approximation.

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

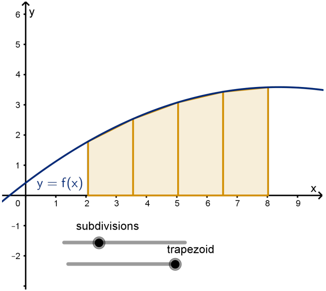
Our final approximation abandons rectangles entirely. Using trapezoids instead allows for shapes that reflect the value of the function at both the right and left endpoint.

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

A Midpoint Approximation

Calculate the M_3 approximation of $\int_{-1}^{5} x^2 \, dx$.

A Trapezoid Approximation Using a Table of Values

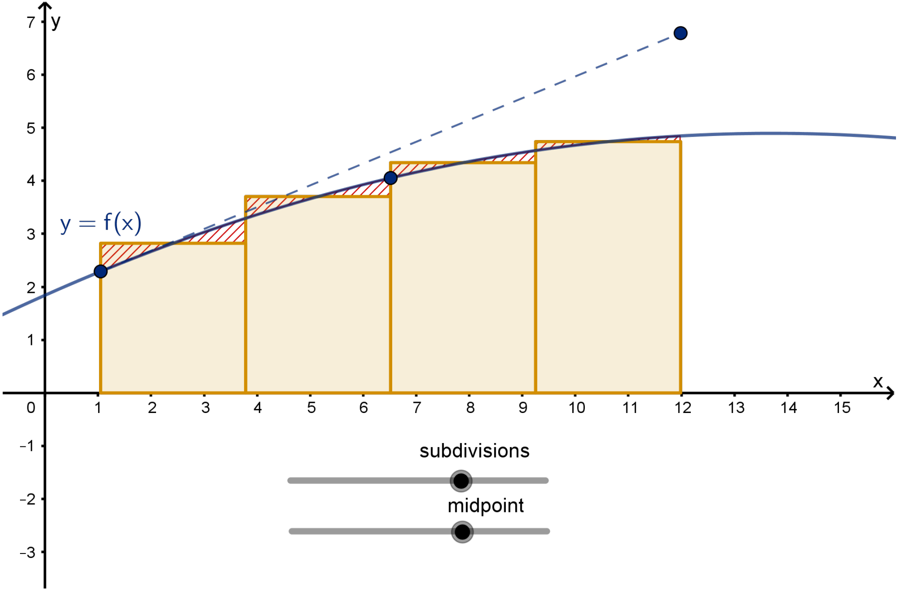
Suppose we have the following table of values for a function $f(x)$

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 $\displaystyle\int_2^{14} f(x)\, dx$.

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.

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 $\frac{1}{1000}$?

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 $\int_a^b f(x)\, dx$?

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?

Q28

Suppose we want to estimate $\int_4^{20} f(x)\, dx$ 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?

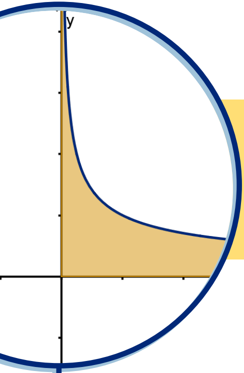
Q30

Suppose you are interested in the value of $\int_0^{25} f(x)\, dx$, 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 $\int_0^{25} f(x)\, dx$?

- **a** Propose one or two different approaches that you might use to approximate $\int_0^{25} f(x)\, dx$.
- **b** Compare your approaches with one or two people sitting near you. What are the strengths of each? Which one does your group think would be best, and why?



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.

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 $\tan^{-1}(\infty)$.

Exercise

Evaluate the following limits:

a $\displaystyle\lim_{x\to\infty} \frac{1}{x^2}$

b $\displaystyle\lim_{x\to\infty} \sqrt{x}$

c $\displaystyle\lim_{t\to-\infty} e^t$

d $\displaystyle\lim_{y\to\infty} \sin y$

e $\displaystyle\lim_{w\to\infty} \ln w$

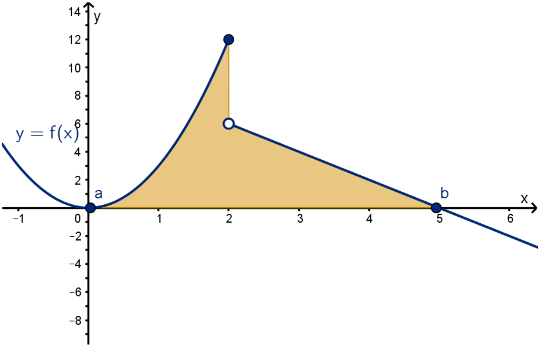
f $\displaystyle\lim_{x\to-\infty} \frac{3x^2+7}{x^2-5x}$

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 $\int_0^5 f(x)\, dx$?



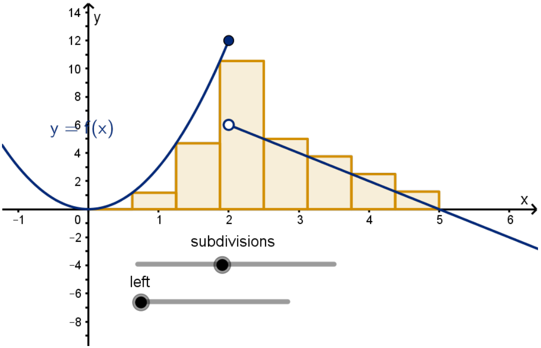


Figure: Rectangle approximations of the area beneath a discontinuous graph

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.

We don't have a form of $F(x)$ that we can evaluate, so how do we compute $\displaystyle\int_0^5 f(x)\,dx$?

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.

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.

Integrating a Function with a Vertical Asymptote

Definition

When $f(x)$ has a vertical asymptote at c in $[a, b]$ we call $\displaystyle\int_a^b f(x)\,dx$ an **improper integral**.

How can we compute $\displaystyle\int_0^4 \frac{1}{\sqrt{x}}\,dx$?

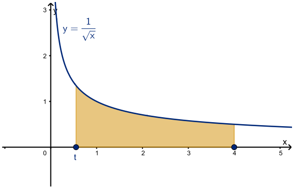


Figure: The area beneath a function with a vertical asymptote

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**.

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.

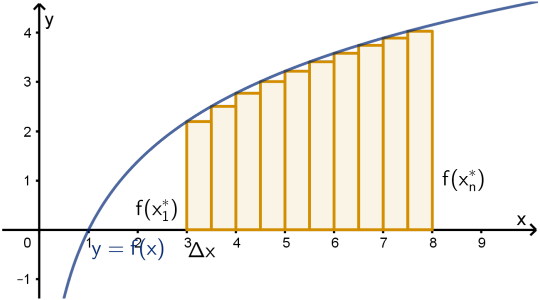


Figure: The area beneath a graph, approximated by rectangles

Consider how this approach would work with an unbounded interval: $a \leq x$.

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 $\int_{-\infty}^b f(x)\,dx = \lim_{t\to-\infty}\int_t^b f(x)\,dx$.

Evaluating an Improper Integral

Compute $\displaystyle\int_2^\infty \frac{32}{x^3}\,dx$.

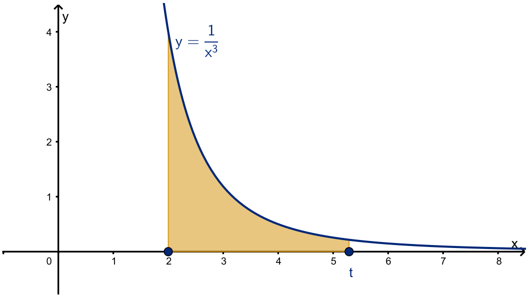


Figure: An integral over an unbounded domain

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 $\displaystyle\int_{-\infty}^{\infty} f(x)\ dx$.

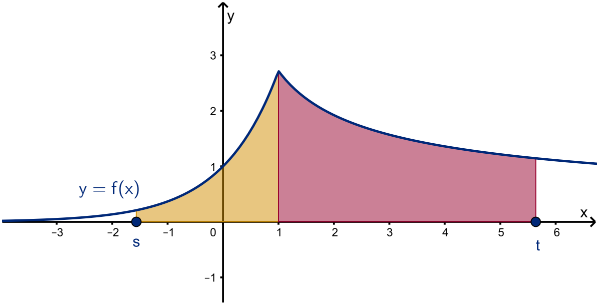


Figure: An integral over the real line, broken into two limits

Can We Take a Limit of $\int_{-t}^{t} f(x)\, dx$ Instead?

The integral $\int_{-\infty}^{\infty} x^3\, dx$ is a useful test case.

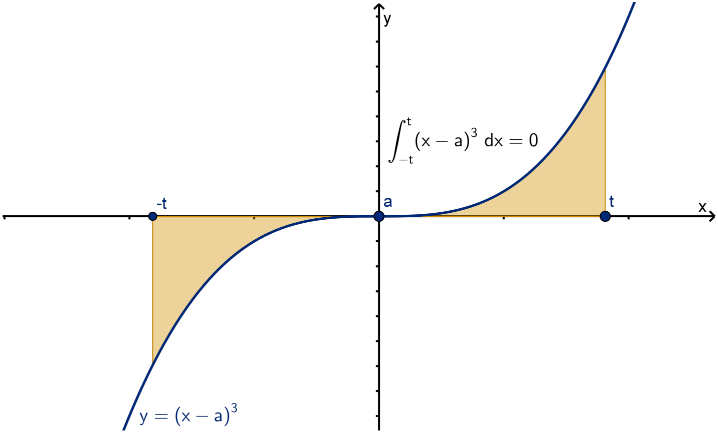


Figure: The area under a functions of the form $f(x) = (x-a)^3$

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.

A Comparison Test

Recall the following theorems

Theorem

If $f(x) \leq g(x)$ on $[a, b]$ then $\int_a^b f(x)\ dx \leq \int_a^b g(x)\ dx$.

Theorem

Let a be a real number or $\pm\infty$. If $F(x) \leq G(x)$ for all x near a, then $\lim_{x\to a} F(x) \leq \lim_{x\to a} G(x)$.

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 $\displaystyle\int_a^\infty f(x)\,dx$?

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?

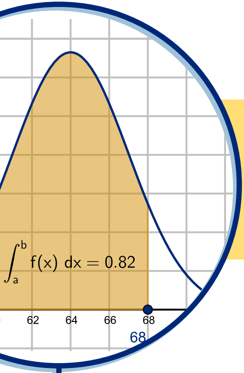
Q4 If we know that $\displaystyle\int_a^\infty g(x)\ dx$ converges, what condition on $f(x)$ would guarantee that $\displaystyle\int_a^\infty f(x)\ dx$ converges?

Q26

a 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.

b Compare with someone near you. Did you choose the same value of a when splitting the integral?

c Explain why a different choice of a would not give different values of the integral. Appeal to a specific integral rule if possible.



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.

What Is a Continuous Probability Distribution?

Why study probability?

The study of inferential statistics is dependent on the ability to compute probabilities. The basic model for statistical reasoning is:

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.

Probability and inferential statistics are explored thoroughly in later math courses.

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.

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

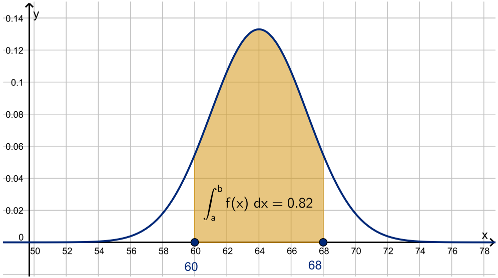
$$P(a \leq X \leq b) = \int_a^b f_X(x)\ dx.$$

$f_X(x)$ must satisfy

1. $f_X(x) \geq 0$ for all x.
2. $\displaystyle\int_{-\infty}^{\infty} f_X(x)\ dx = 1$

An integral is the natural way to measure probability.

$$P(a \leq X \leq c) + P(c \leq X \leq b) = P(a \leq X \leq b)$$
$$\int_a^c f_X(x)\ dx + \int_c^b f_X(x)\ dx = \int_a^b f_X(x)\ dx$$



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?

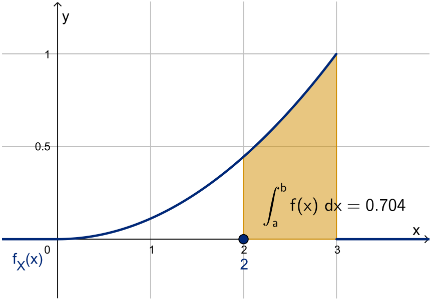


Figure: The density function of X and the area representing $P(X > 2)$

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.

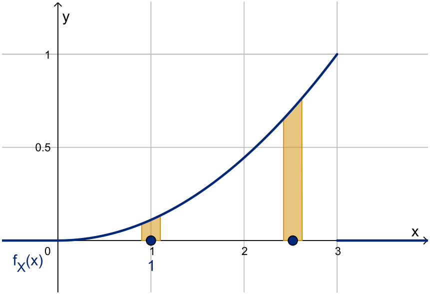


Figure: The density function of X and the areas that represent the likelihood of larger and smaller outcomes

What Density Functions Arise Naturally?

Definition

Given an interval $[a, b]$, the **uniform distribution** on $[a, b]$ is given by

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{if } x > b \text{ or } x < a \end{cases}$$

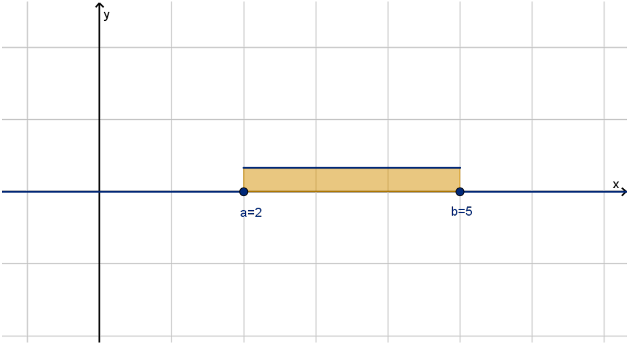


Figure: The density function of a uniform distribution

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.

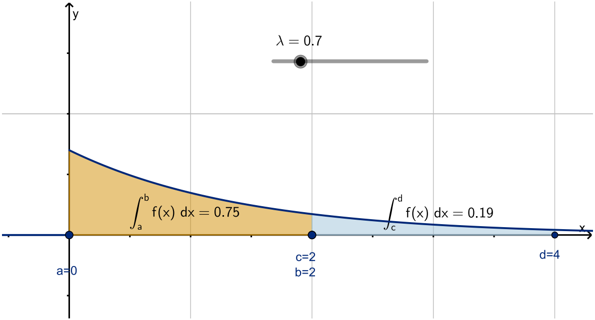


Figure: The density function of an exponential distribution

Definition

The **normal distribution** is sometimes called a **bell curve**. Many natural phenomena are normally distributed. The formula is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

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.

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.

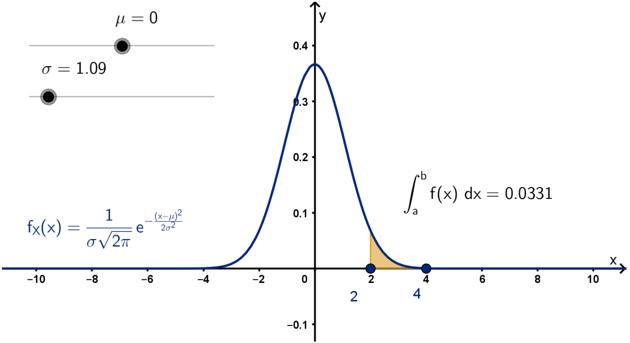


Figure: The density function (bell curve) of a normal distribution

What Is the Expected Value of a 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

$$E[X] = \sum_x x f_X(x)$$

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

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) \, dx$$

The Expected Value of a Uniform Random Variable

Compute the expected value of a uniform random variable on $[a, b]$.

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.

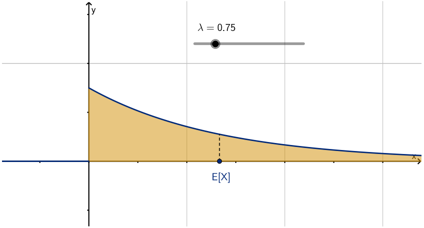


Figure: The expected value of a exponential random variable

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.

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.

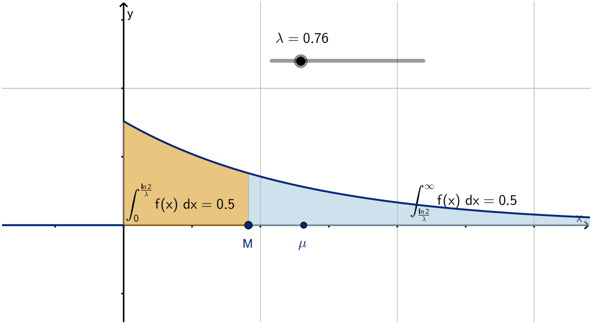


Figure: The median M and expected value μ of an exponential random variable

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.

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?

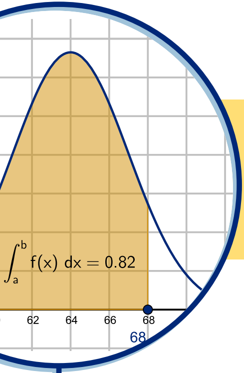
Q3 What must be true about a probability density function?

Q4 How do you compute the expected value of a random variable?

Q6

One of the following probability questions is different from the others. Explain why.

1. If you spin a prize wheel 3 times, what is the probability that my winnings add up to exactly \$80?
2. If you flip two weighted (unfair) coins, what is the probability that exactly one of them comes up tails?
3. If you pick a random person, what is the probability that her height is exactly 68 inches?
4. If I spin a wheel of names, what is the probability that it takes exactly 7 spins to land on my own name?



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.

What Is a Function of a Random Variable?

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 $f_Y(y)$.

$$f_X(x) = \begin{cases} 0.1 & \text{if } x = 0 \\ 0.2 & \text{if } x = 2 \\ 0.3 & \text{if } x = 3 \\ 0.4 & \text{if } x = -2 \\ 0 & \text{otherwise} \end{cases} \qquad f_Y(y) = \begin{cases} 0.1 & \text{if } y = 0 \\ 0.6 & \text{if } y = 4 \\ 0.3 & \text{if } y = 9 \\ 0 & \text{otherwise} \end{cases}$$

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

$$
\begin{aligned}
E[N] = {} & \underbrace{(1)\left(\frac{1}{100}\right) + \cdots + (1)\left(\frac{1}{100}\right)}_{\text{9 times}} \\
& + \underbrace{(2)\left(\frac{1}{100}\right) + \cdots + (2)\left(\frac{1}{100}\right)}_{\text{90 times}} \\
& + (3)\left(\frac{1}{100}\right) \\
= {} & 1.92
\end{aligned}
$$

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 $E[Y] = \int_{-\infty}^{\infty} y f_Y(y)\, dy$.

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

$$E[Y] = \int_{-\infty}^{\infty} g(x) f_X(x)\ dx$$

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?

The Average Value of a Function

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 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.

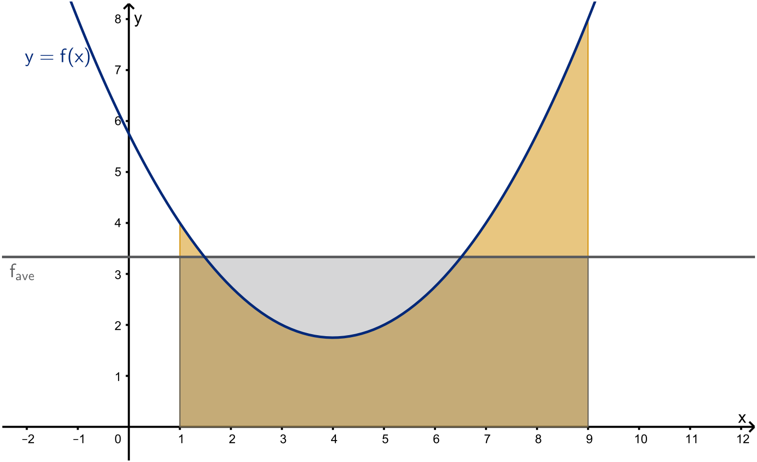


Figure: The graph of $y = f(x)$ and the constant function $y = f_{\text{ave}}$

Computing The Average Value of a Function

Compute the average value of $f(x) = xe^{x^2}$ between $x = 1$ and $x = 3$.

Variance

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.

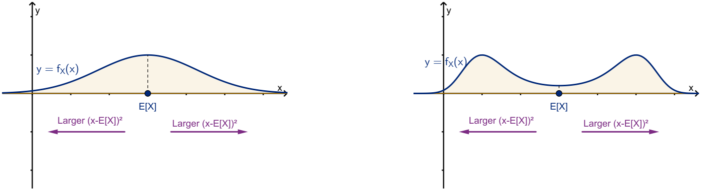


Figure: A density function with less variance and a density function with more variance

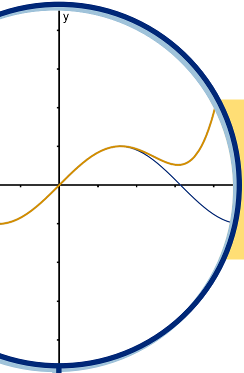
For example, we can compute the variance of X where X is a uniform random variable on $[0, 8]$.

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?



Section 3.1

Taylor Polynomials

Goals:

1. Approximate a function with a Taylor polynomial.
2. Compute error bounds for a Taylor polynomial.

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.

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.

What Is a Taylor Polynomial?

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 3.1.4

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

Example 3.1.5

A Taylor Polynomial in $\sum$ Notation

Write the 10th degree Taylor Polynomial for $f(x) = \frac{1}{x}$ centered at $x = 3$.

How Accurate Is the Taylor Polynomial?

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

$$f(x) - f(a) = \int_a^x f'(t)dt \quad g(x) - g(a) = \int_a^x g'(t)dt.$$

Larger functions have larger integrals.

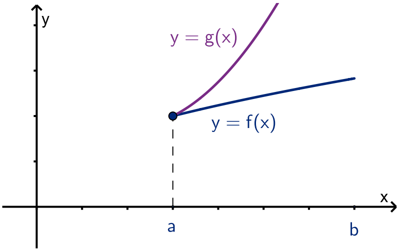


Figure: Two functions with a common value at a: $f(x)$ with a smaller derivative and $g(x)$ with a larger derivative.

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

$$f(x) = T_{n+1}(x) = T_n(x) + \frac{M}{(n+1)!}(x-a)^{n+1}.$$

Because derivatives and values of a Taylor polynomial match the function

$$\begin{tikzcd}
 & \boxed{\frac{d}{dx^{n+1}} f(x) = \frac{d}{dx^{n+1}} T_{n+1}(x) = M \text{ on } [a, b]} \arrow[d] \\
\boxed{\frac{d}{dx^{n}} f(a) = \frac{d}{dx^{n}} T_{n+1}(a)} \arrow[r] & \boxed{\frac{d}{dx^{n}} f(x) = \frac{d}{dx^{n}} T_{n+1}(x) \text{ on } [a, b]} \arrow[d] \\
\boxed{\frac{d}{dx^{n-1}} f(a) = \frac{d}{dx^{n-1}} T_{n+1}(a)} \arrow[r] & \boxed{\frac{d}{dx^{n-1}} f(x) = \frac{d}{dx^{n-1}} T_{n+1}(x) \text{ on } [a, b]} \arrow[d, dashed] \\
\boxed{\frac{d}{dx} f(a) = \frac{d}{dx} T_{n+1}(a)} \arrow[r] & \boxed{\frac{d}{dx} f(x) = \frac{d}{dx} T_{n+1}(x) \text{ on } [a, b]} \arrow[d] \\
\boxed{f(a) = T_{n+1}(a)} \arrow[r] & \boxed{f(x) = T_{n+1}(x) \text{ on } [a, b]}
\end{tikzcd}$$

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 $w(x) = T_n(x) + \frac{M}{(n+1)!}(x-a)^{n+1}$.

Because M is a bound on $\left|\frac{d}{dx^{n+1}} f(x)\right|$

$$\begin{tikzcd}
 & \frac{d}{dx^{n+1}} f(x) \leq \frac{d}{dx^{n+1}} w(a) = M \text{ on } [a,b] \arrow[d] \\
\frac{d}{dx^{n}} f(a) = \frac{d}{dx^{n}} w(a) \arrow[r] & \frac{d}{dx^{n}} f(x) \leq \frac{d}{dx^{n}} w(x) \text{ on } [a,b] \arrow[d] \\
\frac{d}{dx^{n-1}} f(a) = \frac{d}{dx^{n-1}} w(a) \arrow[r] & \frac{d}{dx^{n-1}} f(x) \leq \frac{d}{dx^{n-1}} w(x) \text{ on } [a,b] \arrow[d, dashed] \\
\frac{d}{dx} f(a) = \frac{d}{dx} w(a) \arrow[r] & \frac{d}{dx} f(x) \leq \frac{d}{dx} w(x) \text{ on } [a,b] \arrow[d] \\
f(a) = w(a) \arrow[r] & f(x) \leq w(x) \text{ on } [a,b]
\end{tikzcd}$$

Because derivatives and values of a Taylor polynomial match the function

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.

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 $f(x)$?

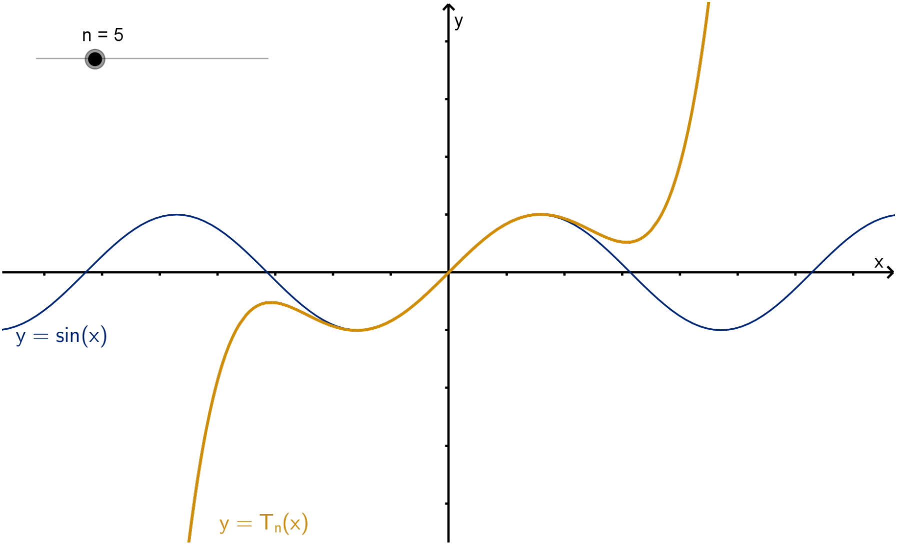


Figure: $f(x) = \sin x$ approximated by its Taylor polynomials, $T_n(x)$

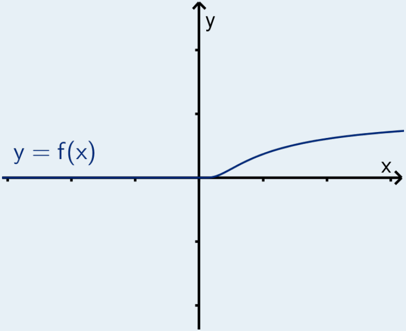
Main Ideas

- 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.

Some functions are not well estimated by their Taylor polynomial.

Example

$$f(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ e^{-\frac{1}{x}} & \text{if } x > 0 \end{cases}$$



$f^{(k)}(0) = 0$ 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.

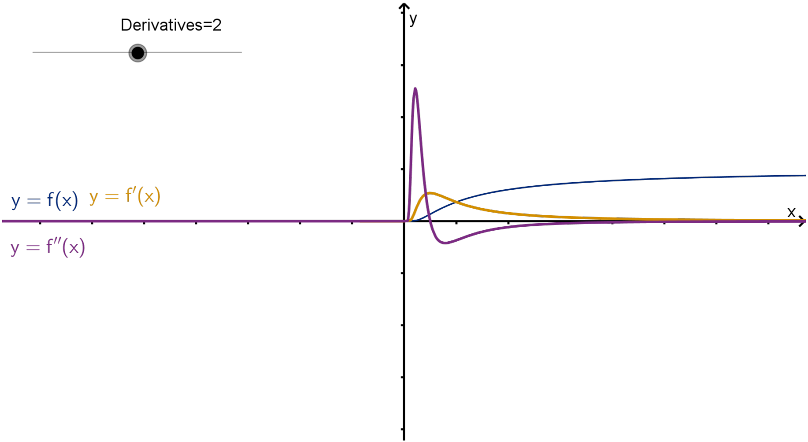


Figure: A function whose derivative bound grow factorially

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 $f(x)$?

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?

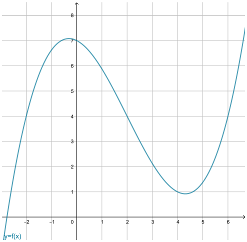
Q28

Let $f(x) = xe^x$.

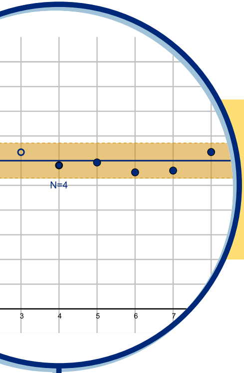
- a Compute the Taylor polynomial $T_3(x)$ for $f(x)$ centered at $x = 0$.
- b Compute the theoretical error bound for $T_3(2)$.
- c Explain the difficulties that would arise from this error bound, if your goal is to approximate $f(2)$ by hand. Can you resolve them?

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.



Section 3.2

Sequences

Goals:

1. Use notation to describe the terms of an infinite sequence.
2. Calculate the limit of an infinite sequence.

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.

$$\left\{\frac{1}{2},\frac{2}{3},\frac{3}{4},\frac{4}{5}\ldots\right\} \qquad \left\{\frac{n}{n+1}\right\}_{n=1}^{\infty} \qquad a_n=\frac{n}{n+1}$$

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 $\frac{1}{1000000}$.

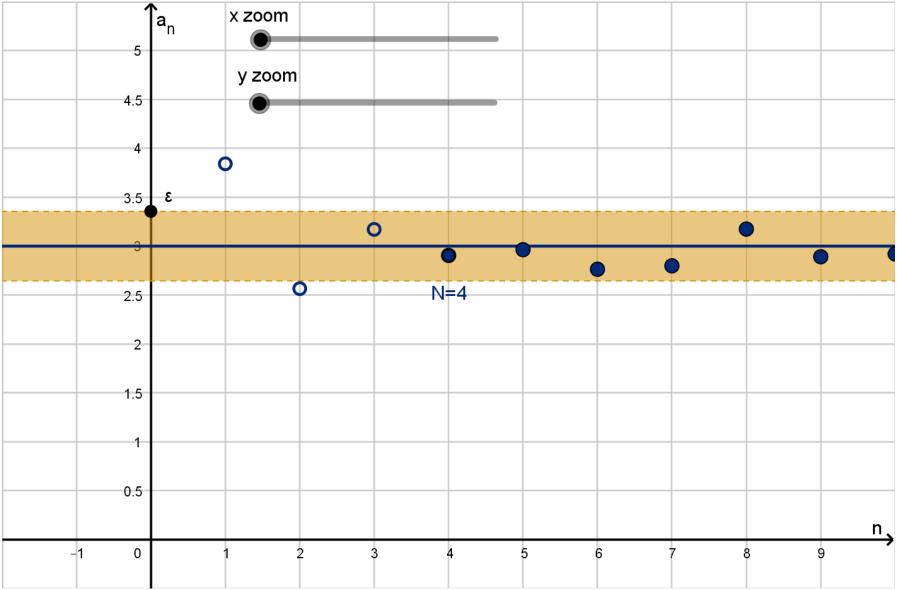


Figure: A sequence converging to $L = 3$

Computing a Limit

Calculate $\lim_{n\to\infty} \dfrac{n}{n+1}$

$$\left\{\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5} \dots\right\}$$

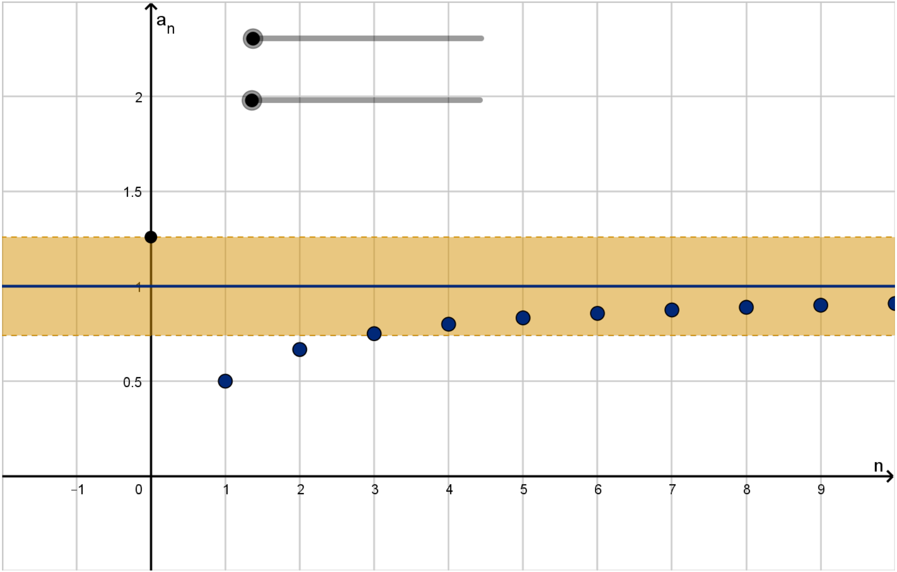


Figure: The sequence $\frac{n}{n+1}$ converges to $L = 1$.

How Are Limits of Sequences and Functions Related?

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

$$\lim_{n \to \infty} a_n = L.$$

Sequence Limits Using Functions

Find limits of the following sequences:

a $\lim_{n\to\infty} \frac{2n}{n+3}$

b $\lim_{n\to\infty} \frac{1}{n^3}$

c $\lim_{n\to\infty} e^{-n}$

d $\lim_{n\to\infty} \frac{n^2}{e^n}$

e $\lim_{n\to\infty} (-1)^n$

Theorems

All of the laws for limits of functions at infinity also apply to limits of sequences. For instance suppose $\lim_{n\to\infty} a_n = L$ and $\lim_{n\to\infty} b_n = M$. If $c_n = a_n + b_n$ then

$$\lim_{n\to\infty} c_n = L + M.$$

Indeterminate Forms with Factorials

Dominance

We say $f(x)$ **dominates** $g(x)$ if $\lim_{x \to \infty} \frac{f(x)}{g(x)} = \pm\infty$. 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:

$$\underset{\text{(larger base}>>\text{smaller base)}}{\text{exponential}} >> \underset{\text{(larger degree}>>\text{smaller)}}{\text{polynomial}} >> \underset{\text{(smaller power}>>\text{larger)}}{\text{root}}$$

Theorem

As $n \to \infty$, $n!$ will eventually **dominate** any exponential function (and thus any polynomial, root or logarithm).

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 $\lim_{n\to\infty} b_n$?

Q12

Consider the sequence $a_n = n\sin(\pi n)$

- a What is $\lim\limits_{x\to\infty} x\sin(\pi x)$?
- b Compute the first few values a_1, a_2, a_3, and a_4.
- c What is $\lim\limits_{n\to\infty} n\sin(\pi n)$?
- d Does this contradict one of our theorems? Explain.

Remark

Read theorems carefully. The hypothesis and conclusion are not interchangeable. Mixing them up can turn a true theorem into a false one.

Q24

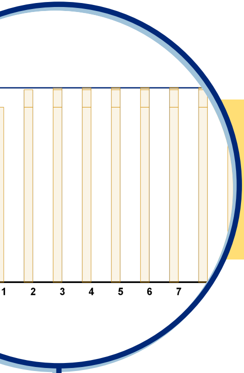
Let $T_n(x)$ be the nth Taylor polynomial of $f(x) = \ln x$ centered at $x = 1$.

- 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 ∞?

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.

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.

$$0.33333333\ldots \qquad 3.1415926\ldots$$

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 $\displaystyle\sum_{k=1}^{\infty} a_k$ is

$$s_n = a_1 + a_2 + a_3 + \cdots + a_n$$

A series $\displaystyle\sum_{k=1}^{\infty} a_k$ **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 $\displaystyle\sum_{k=1}^{\infty} \frac{3}{10^k}$.

a Compute the first few partial sums s_1, s_2, s_3 of this series.

b Compute $\displaystyle\lim_{n\to\infty} s_n$

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.

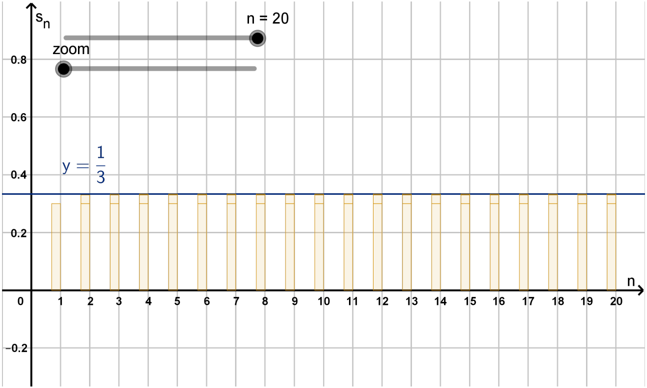


Figure: The partial sums s_n converging to $L = \frac{1}{3}$

The Harmonic Series

Consider a partial sum of the harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$.

$$\begin{aligned} s_8 =& \frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8} \\ >& \frac{1}{1}+\frac{1}{2}+\underbrace{\frac{1}{4}+\frac{1}{4}}_{\frac{1}{2}}+\underbrace{\frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}}_{\frac{1}{2}} \\ =& 1+\frac{1}{2}+\frac{1}{2}+\frac{1}{2} \end{aligned}$$

In general we can 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.

What Is a 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

- $\sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{k-1} = 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots$
- $\sum_{k=1}^{\infty} \frac{3}{10} \left(\frac{1}{10}\right)^{k-1} = \frac{3}{10} + \frac{3}{100} + \frac{3}{1000} + \cdots = \frac{1}{3}$

Evaluate $\displaystyle\sum_{k=1}^{\infty} ar^{k-1}$.

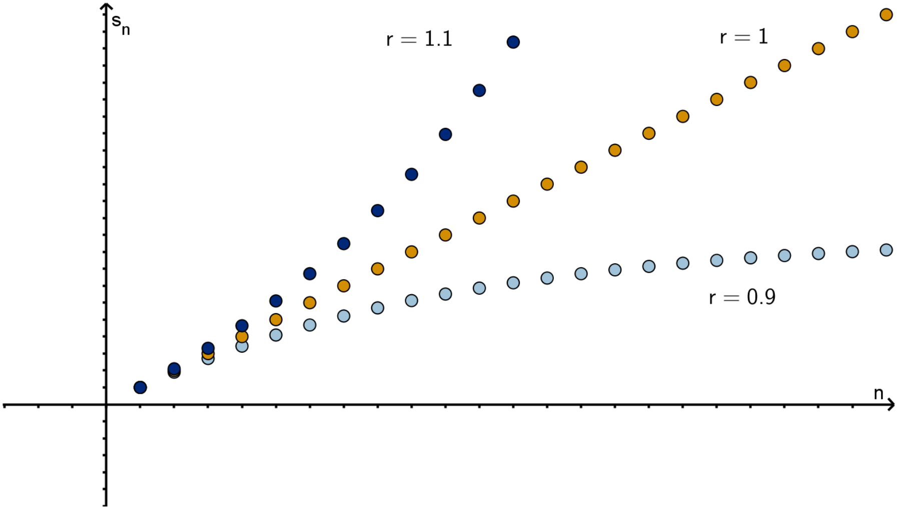


Figure: The partial sums of $\sum ar^{k-1}$ for various r

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 $0.999999\ldots$

Question 3.3.6

What Does the Size of a_k Tell Us About $\sum a_k$?

Theorem (The Divergence Test)

Let a_k be a sequence. If $\lim_{k\to\infty} a_k \neq 0$, then the series

$$\sum_{k=1}^{\infty} a_k$$

diverges.

Remark

The divergence test does not tell us anything, if $\lim_{k\to\infty} a_k = 0$. 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 $\displaystyle\sum_{k=2}^{\infty} \frac{k^2}{e^k}$

What Is the Ratio Test?

The convergence of a geometric series depends on the ratio r. Even when a series is not geometric, we can attempt to apply similar reasoning to determine whether it converges.

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.

Example 3.3.9

Applying the Ratio Test

a Does $\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k!}$ converge or diverge?

b Does $\displaystyle\sum_{k=1}^{\infty} \frac{2^k}{k^2}$ converge or diverge?

c Does $\displaystyle\sum_{k=1}^{\infty} k$ converge or diverge?

Main Ideas

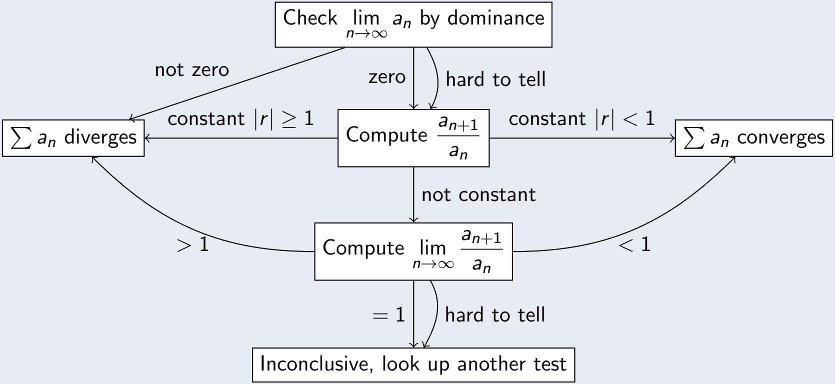
- When applying the ratio test, be sure to replace every k with $k+1$ for the a_{k+1} term.
- Familiarize yourself with the algebra rules that allow you to simplify ratios of exponentials and factorials.

Example 3.3.10

A Strategy for Series Tests

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

$$\sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Summary Questions

Q1 What is the difference between a sequence and a series?

Q2 How do we evaluate a series?

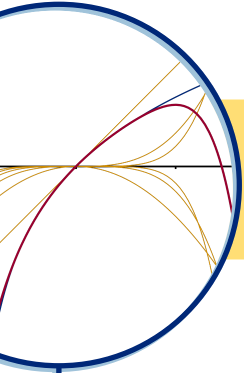
Q3 What is a geometric series. How do we evaluate one?

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 $\left|\frac{a_{k+1}}{a_k}\right|$?

- **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 $\sum ar^{k-1}$?



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.

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 $x = 0$.

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.

The Domain of a Power Series

What is the domain of $\displaystyle\sum_{k=1}^{\infty} \frac{k^2}{4^k}(x-5)^k$?

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.

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 $|x-a| > R$.

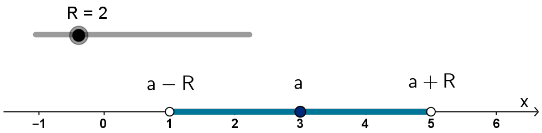


Figure: The domain $|x-a| < R$ of a power series.

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 $\displaystyle\sum_{k=0}^{\infty} c_k(x-a)^k$ 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. $\displaystyle f'(x) = \sum_{k=1}^{\infty} kc_k(x-a)^{k-1}$
2. $\displaystyle\int f(x)\,dx = C + \sum_{k=0}^{\infty} c_k\frac{(x-a)^{k+1}}{k+1}$

Both of these functions also have radius of convergence R.

Example

We have seen that $\dfrac{1}{1-x} = \displaystyle\sum_{k=0}^{\infty} x^k$ 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 $(-1, 1)$.

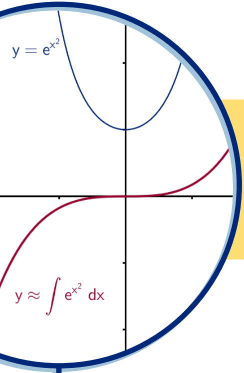
Summary Questions

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?



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.

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 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**.

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?

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 $T(x)$?

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 $T(x)$?

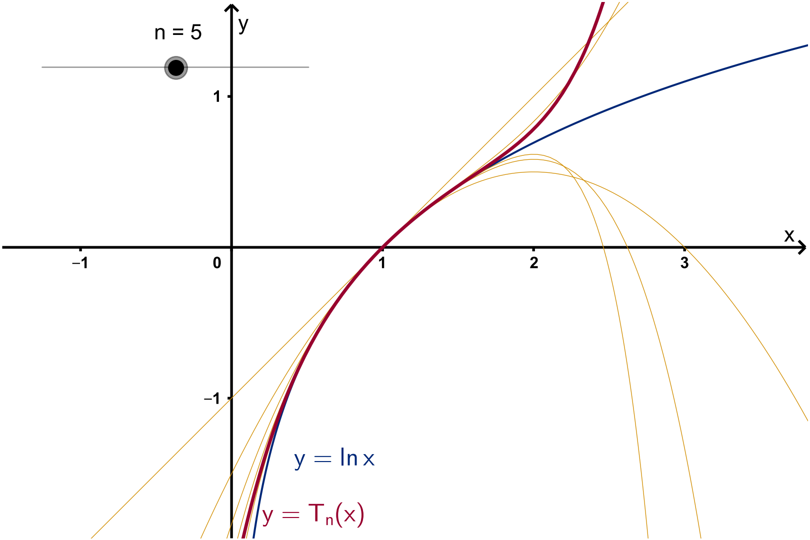


Figure: The Taylor polynomials approach $\ln x$ only on $(0, 2]$.

Mixing Taylor Series and Algebra

Let $f(x) = x^2 \sin x$. Compute a Taylor series for $f(x)$ centered at $x = 0$.

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.

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 $\int e^{x^2}\, dx$.

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 $f(x)$.

b Compute a Taylor series for $\int e^{x^2}\, dx$.

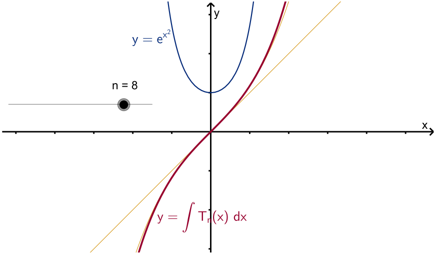


Figure: The graph of e^{x^2}, $\int e^{x^2}\,dx$, and the partial sums of its Taylor series.

Main Ideas

- Compositions of functions can be composed through Taylor series.
- Taylor series allow us to integrate functions that are otherwise impossible to integrate.

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?

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?

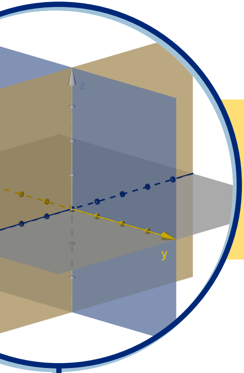
Q3 How can we produce the Taylor series for $x^n f(x)$ or $f(x^n)$? Where does the center need to be for the result to be a Taylor series?

Q4 What is a Maclaurin series?

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 $f(x) = T(x)$.



Section 4.1

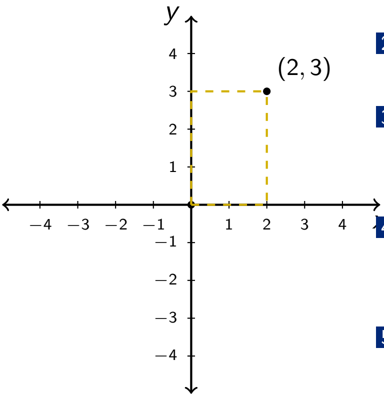
Three-Dimensional Coordinate Systems

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.

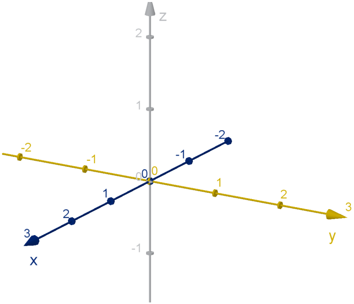
How Do Cartesian Coordinates Extend to Higher Dimensions?

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 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.

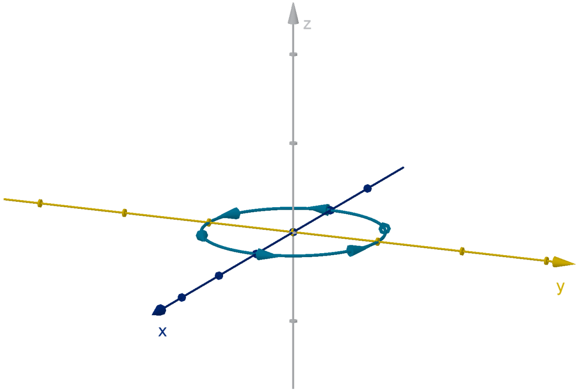


Figure: The counterclockwise unit circle in the xy-plane

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?

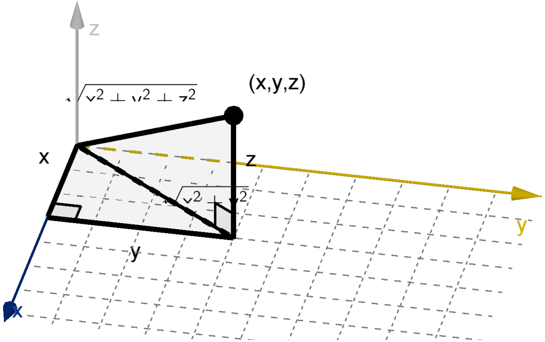
How can we draw a reasonable diagram of $(-5, 1, -4)$?

How Do We Measure Distance in Three-Space?

Theorem

The distance from the origin to the point (x, y, z) is given by the Pythagorean Theorem

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



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?

Definition

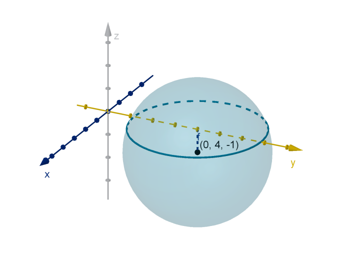
The **graph** of an implicit equation is the set of points whose coordinates satisfy that equation. In other words, the two sides are equal when we plug the coordinates in for x, y and z.

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 $y = 3$.

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.

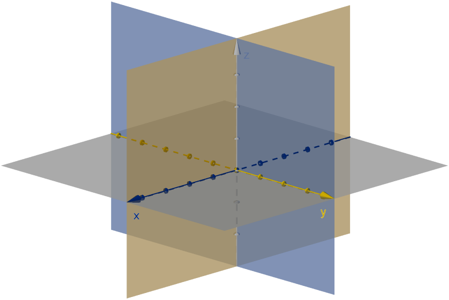


Figure: The coordinate planes in 3-dimensional space.

Graphing an Equation with One Free Variable

Sketch the graph of the equation $z = x^2 - 3$.

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).

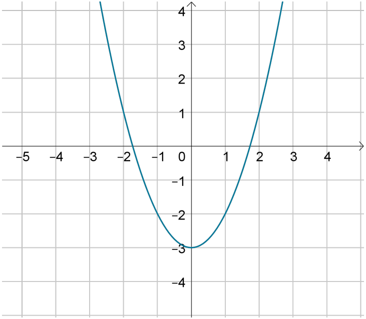


Figure: The curve $y = x^2 - 3$

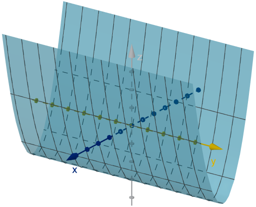


Figure: The surface $z = x^2 - 3$

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.

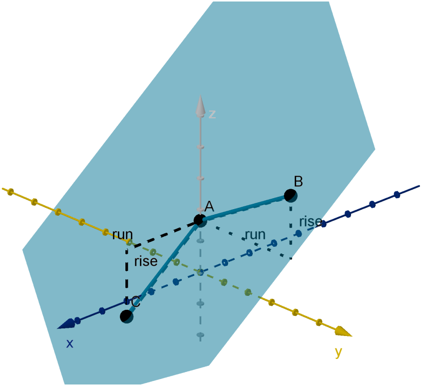


Figure: A plane with slopes in the x and y directions.

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 $(0,0,8)$.

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?

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.

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?

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?

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?

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.

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**.

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

$$f(1, 4) = \sqrt{36 - 4 \cdot 1^2 - 4^2} = 4.$$

The Domain of a Function

Identify the domain of $f(x, y) = \sqrt{36 - 4x^2 - y^2}$.

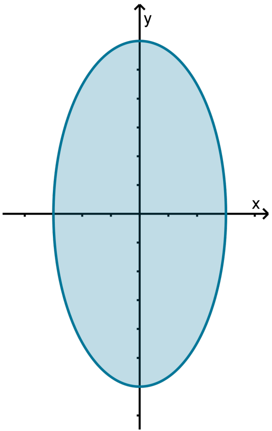


Figure: The domain of a function

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.

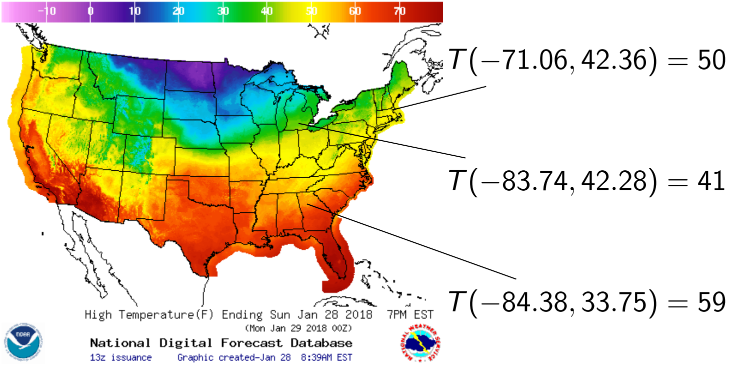


Figure: A temperature map

Digital Images

A digital image can be defined by a brightness function $B(x, y)$.

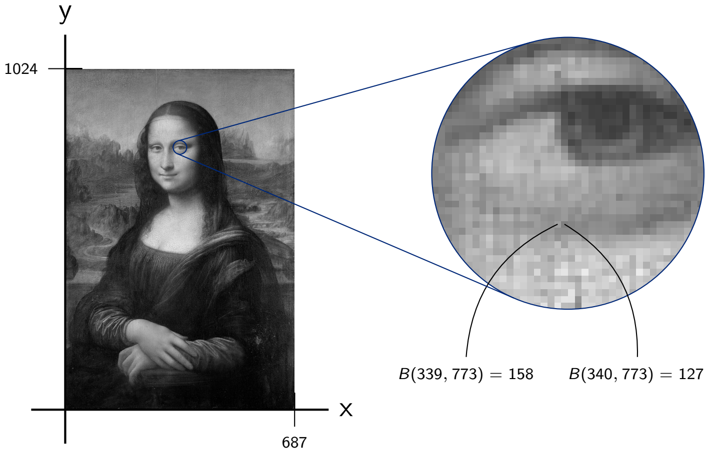


Figure: An image represented as a brightness function B on each pixel

What Is the Graph of a Two-Variable Function?

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 $(1,4,0)$.

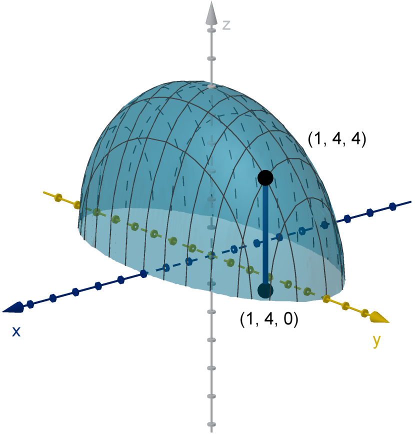


Figure: The graph $z = \sqrt{36 - 4x^2 - y^2}$

How Do We Visualize a Graph in Three-Space?

Definition

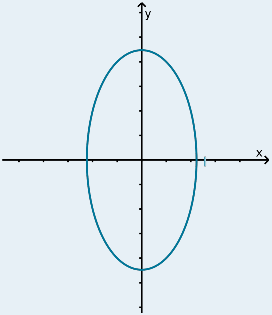
A **level set** of a function $f(x, y)$ is the graph of the equation $f(x, y) = c$ for some constant c. For a function of two variables this graph lies in the xy-plane and is called a **level curve**.

Example

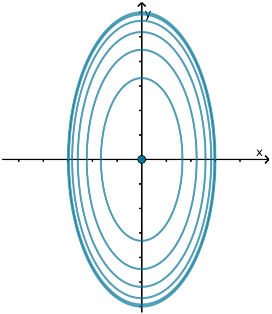
Consider the function

$$f(x, y) = \sqrt{36 - 4x^2 - y^2}.$$

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.



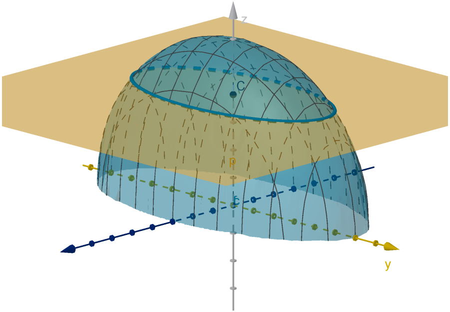


Figure: The graph $z = f(x, y)$, the planes $z = c$, and the level curves

Drawing Level Curves

Where are the level curves on this temperature map?

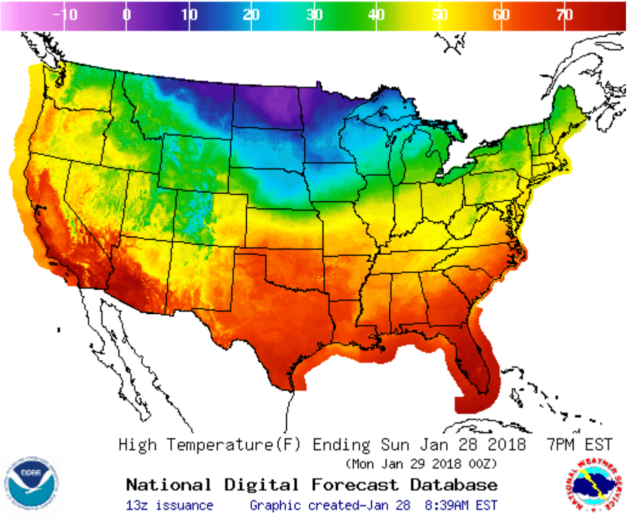


Figure: A temperature map

Using Level Curves to Describe a Graph

What features can we discern from the level curves of this topographical map?

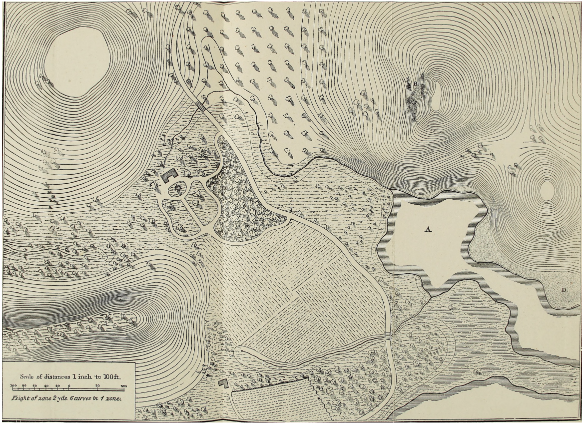


Figure: A topographical map

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 $y = 1$.

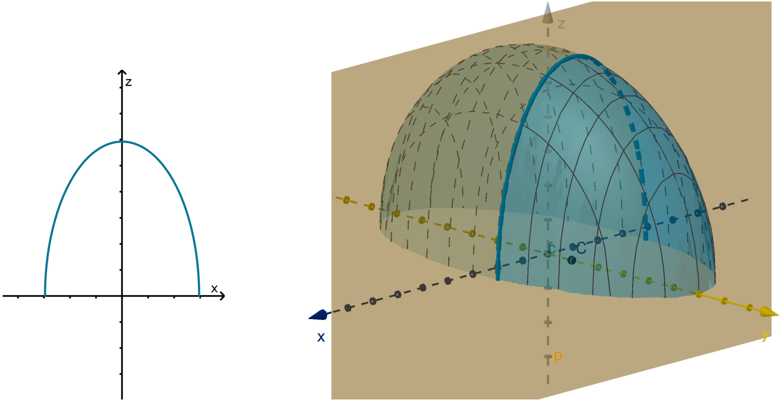


Figure: The $y = 1$ cross section of $z = \sqrt{36 - 4x^2 - y^2}$

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.

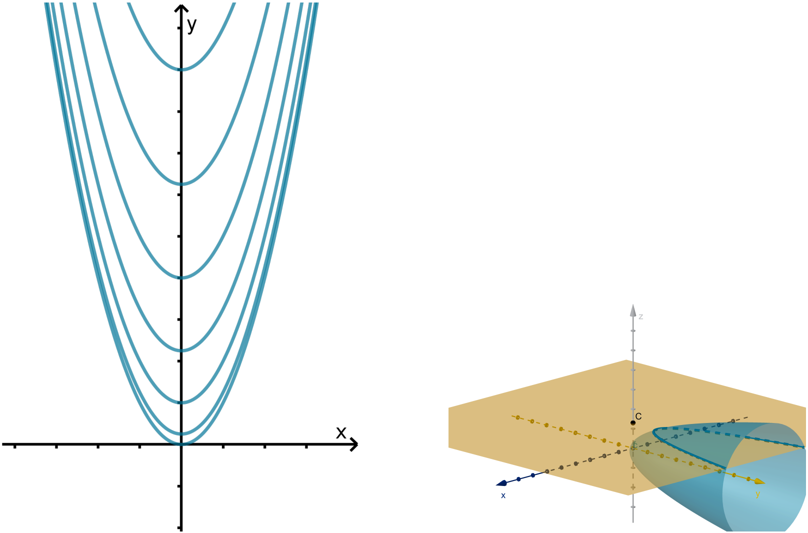


Figure: Level curves of $x^2 - y + z^2 = 0$

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	$f(x, y)$	$f(x, y, z)$	$f(x_1, \ldots,$
Domain	subset of $\mathbb{R}^2$	subset of $\mathbb{R}^3$	subset of
Graph	$z = f(x, y)$ in $\mathbb{R}^3$	$w = f(x, y, z)$ in $\mathbb{R}^4$	$x_{n+1} = f(x_1, \ldots,$
Level Sets	level curve in $\mathbb{R}^2$	**level surface** in $\mathbb{R}^3$	level set in

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).

$$\begin{array}{ccc} x^2+y^2+z^2=25 & \longrightarrow & f(x,y)=\pm\sqrt{25-x^2-y^2} \\ & \searrow & \\ & & \begin{array}{c} F(x,y,z)=x^2+y^2+z^2 \\ F(x,y,z)=25 \end{array} \end{array}$$

Both viewpoints will be useful in the future.

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?

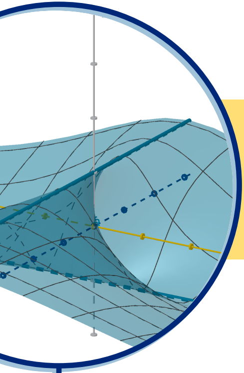
Q3 What are level sets in $\mathbb{R}^2$ and $\mathbb{R}^3$ called?

Q4 What is the difference between an implicit equation and explicit function?

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 $z = f(x, y)$?



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**.

What Is the Limit of a Function?

Definition

We write

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

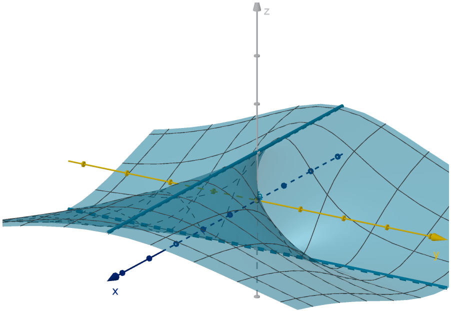
if we can make the values of f stay arbitrarily close to L by restricting to a sufficiently small neighborhood of (a, b).

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 $|f(x, y) - L| < \epsilon$.

Example 4.3.2

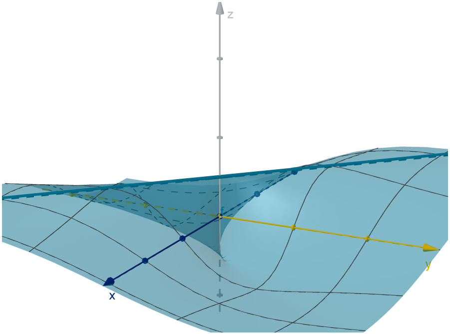
A Limit That Does Not Exist

Show that $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2 - y^2}{x^2 + y^2}$ does not exist.



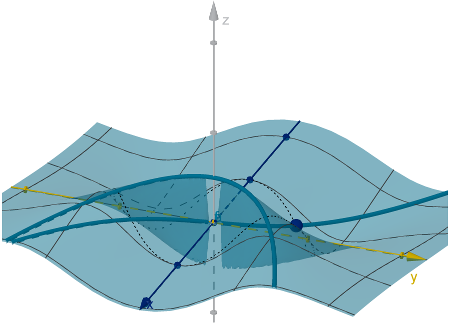
Another Limit That Does Not Exist

Show that $\lim_{(x,y)\to(0,0)} \dfrac{xy}{x^2+y^2}$ does not exist.



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.



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

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

$$\lim_{(x,y)\to(a,b)} f(x,y) = L.$$

What Is a Continuous Function?

Definition

We say $f(x, y)$ is **continuous** at (a, b) if

$$\lim_{(x,y)\to(a,b)} f(x, y) = f(a, b).$$

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.

In each of our examples, the function was a quotient of polynomials, but $(0, 0)$ was not in the domain.

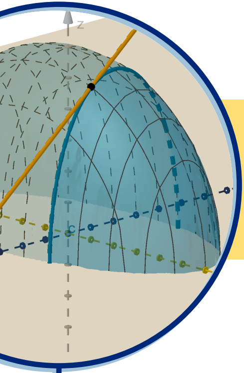
Remark

Limits, continuity and these theorems can all be extrapolated to functions of more variables.

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.

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?

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:

$$y - y_0 = f'(x_0)(x - x_0)$$

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

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

Notation

The partial derivative of a function can be denoted a variety of ways. Here are some equivalent notations

- f_x
- $\frac{\partial f}{\partial x}$
- $\frac{\partial z}{\partial x}$
- $\frac{\partial}{\partial x} f$
- $D_x f$

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.

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?

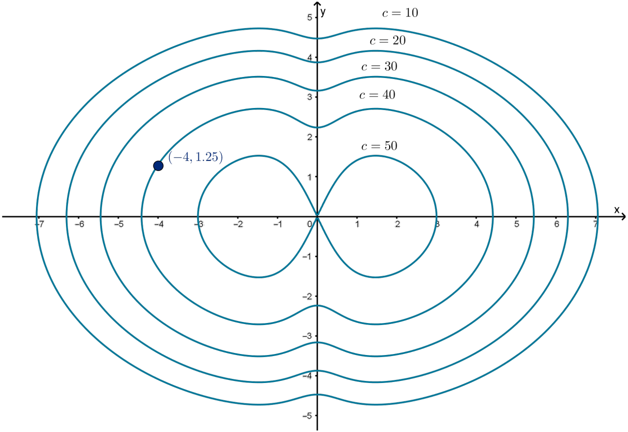


Figure: Some level curves of $f(x, y)$

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 $y = y_0$.

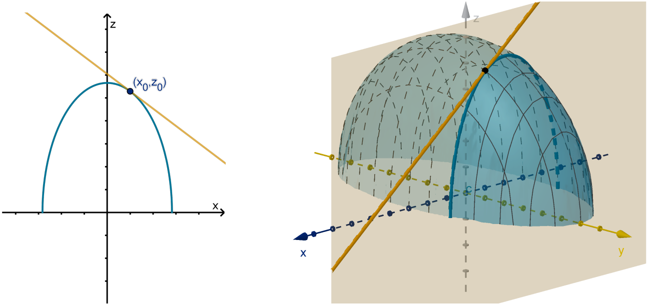


Figure: The tangent line to $z = f(x, y)$ in the x direction

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 $f = \sin(xy)$

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. For example if

$$f(x, y, z) = x^2 - xy + \cos(yz) - 5z^3$$

then we can calculate $\frac{\partial f}{\partial y}$:

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?

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

$$(f_x)_y = f_{xy} = \frac{\partial}{\partial y}\frac{\partial}{\partial x} f = \frac{\partial^2 f}{\partial y \partial x}$$

Example 4.4.9

A Higher Order Partial Derivative

If $f(x, y) = \sin(3x + x^2 y)$ calculate f_{xy}.

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 $f_{xyyz} = f_{zyxy}$.

Summary Questions

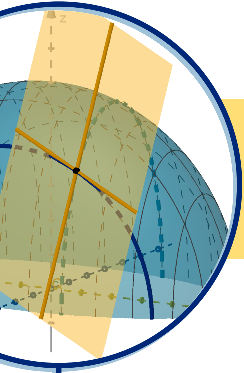
- **Q1** What is the role of each variable when we compute a partial derivative?
- **Q2** What does the partial derivative $f_y(a, b)$ mean geometrically?
- **Q3** Can you think of an example where the partial derivative does not accurately model the change in a function?
- **Q4** What is Clairaut's Theorem?

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?



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.

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.

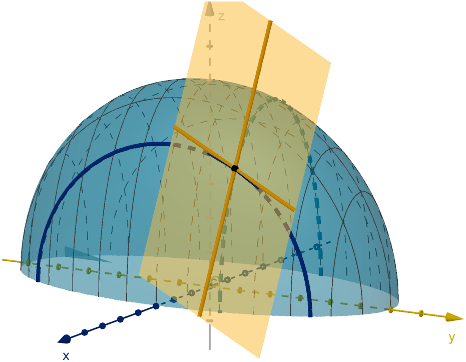


Figure: The tangent plane to $z = f(x, y)$ at a point

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.

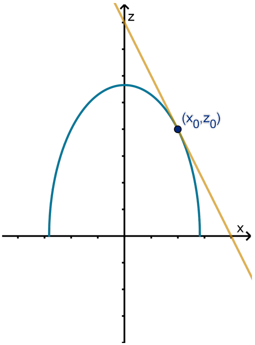
The cross sections of the tangent plane give the equation of the tangent lines we learned in single variable calculus.

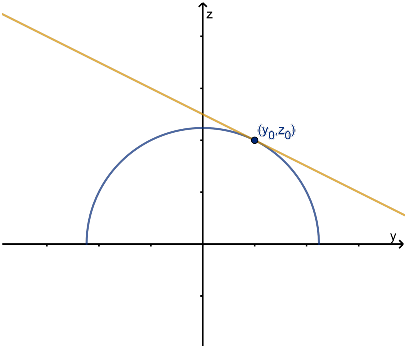
$$y = y_0$$

$$z - z_0 = f_x(x_0, y_0)(x - x_0) + 0$$

$$x = x_0$$

$$z - z_0 = 0 + f_y(x_0, y_0)(y - y_0)$$





Writing the Equation of a Tangent Plane

Give an equation of the tangent plane to $f(x, y) = \sqrt{xe^y}$ at $(4, 0)$

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 $\sqrt{4.02e^{0.05}}$.

How Does Differential Notation Work in More Variables?

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.

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?

Q22

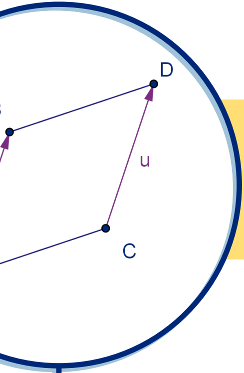
Suppose I decide to invest $\$10,000$ expecting a 6% annual rate of return for 12 years, after which I'll use it to purchase a house. The formula for compound interest

$$P = P_0 e^{rt}$$

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?



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.

Question 5.1.1

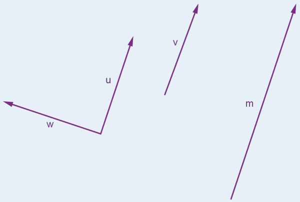
What is a Vector?

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

How Do We Denote Vectors?

Endpoint Notation

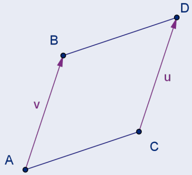
The vector $\vec{v}$ from point A to point B can be represented by the notation

$$\overrightarrow{AB}.$$

A is the **initial point** and B is the **terminal point**.

Theorem

$\overrightarrow{AB} = \overrightarrow{CD}$ if and only if $ABDC$ is a parallelogram (perhaps a squished one).



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

$$\overrightarrow{AB} = \langle 3, -2 \rangle$$

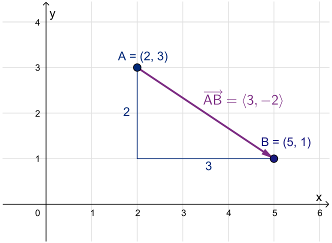


Figure: The x and y components of a vector

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 $\frac{\text{rise}}{\text{run}} = \frac{b}{a}$.

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.

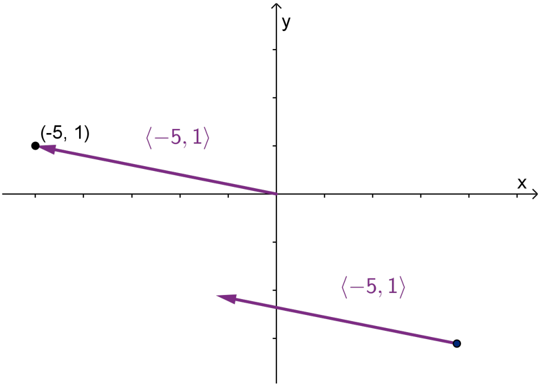
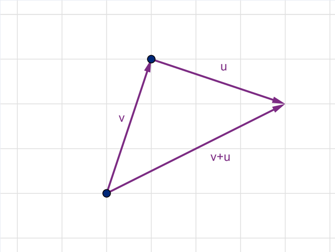


Figure: There is only one point equal to $(-5, 1)$, but there are many vectors equal to $\langle -5, 1 \rangle$.

What Arithmetic Can We Perform with Vectors?

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.

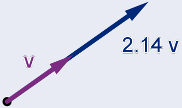


$$\begin{array}{rrr} & \langle \; 1, & 3\rangle \\ + & \langle \; 3, & -1\rangle \\ \hline & \langle \; 4, & 2\rangle \end{array}$$

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:

$$2.5\,\langle 6, 4\rangle = \langle 15, 10\rangle$$

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 $\vec{v} = \langle c, d\rangle$?

Question 5.1.5

What Is Standard Basis Notation?

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.

How Do We Measure the Length of a Vector?

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

$$|\vec{v}| = \sqrt{a^2 + b^2}$$

The Length of a Vector

If $\vec{v} = \langle 3, -5 \rangle$ calculate $|\vec{v}|$

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 $\vec{v}$.

How Do We Measure the Direction of a Vector?

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.

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.

$$|\langle a_1, a_2, a_3, \ldots, a_n\rangle| = \sqrt{a_1^2 + a_2^2 + a_3^2 + \cdots + a_n^2}$$

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.

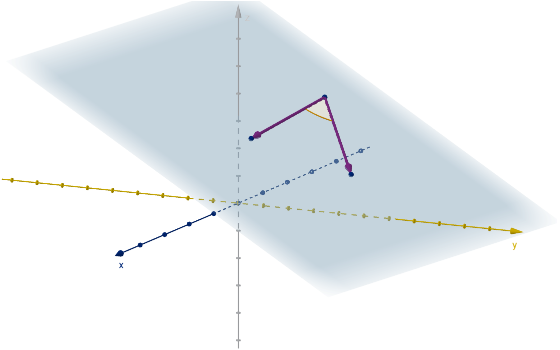


Figure: Two 3-vectors with a common initial point, the plane that contains them, and the angle between them

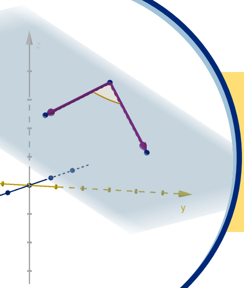
Summary Questions

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 $\overrightarrow{PQ}$?



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.

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 $(-2\vec{i} + 4\vec{k}) \cdot (\vec{i} + 2\vec{j} - \vec{k})$

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 $m\vec{u} \cdot n\vec{v} = mn(\vec{u} \cdot \vec{v})$

What Is the Geometric Significance of the Dot Product?

Theorem

If $\vec{u}$ and $\vec{v}$ are parallel then

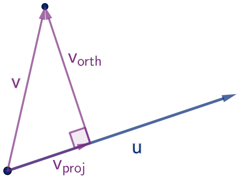
$$\vec{u} \cdot \vec{v} = \begin{cases} |\vec{u}||\vec{v}| & \text{if } \vec{u} \text{ and } \vec{v} \text{ have the same direction} \\ -|\vec{u}||\vec{v}| & \text{if } \vec{u} \text{ and } \vec{v} \text{ have opposite directions} \end{cases}$$

Theorem

If $\vec{u}$ and $\vec{v}$ are orthogonal then

$$\vec{u} \cdot \vec{v} = 0.$$

Two vectors need not be parallel or orthogonal, but given vectors $\vec{u}$ and $\vec{v}$ we can always write $\vec{v} = \vec{v}_{\text{proj}} + \vec{v}_{\text{orth}}$.



The properties of the dot product tell us that

$$\begin{aligned}\vec{u}\cdot\vec{v} &= \vec{u}\cdot\left(\vec{v}_{\text{proj}} + \vec{v}_{\text{orth}}\right)\\ &= \pm\,|\vec{u}||\vec{v}_{\text{proj}}| + 0\end{aligned}$$

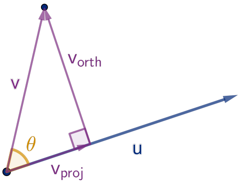
Definition

The number $\dfrac{\vec{u}\cdot\vec{v}}{|\vec{u}|}$ is called the **scalar projection** of $\vec{v}$ onto $\vec{u}$.

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 $\langle 1, 1, 0\rangle$?

Work

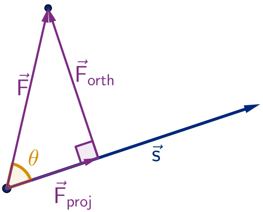
In physics, we say a force **works** on an object if it moves the object in the direction of the force. Given a force F and a displacement s, the formula for work is:

$$W = Fs$$

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.

$$W = \vec{F}_{\text{proj}} \cdot \vec{s} = \vec{F} \cdot \vec{s}$$



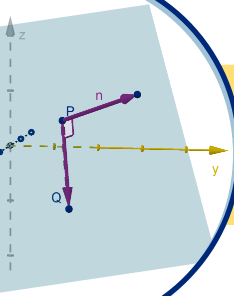
Summary Questions

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?



Section 5.3

Normal Equations of Planes

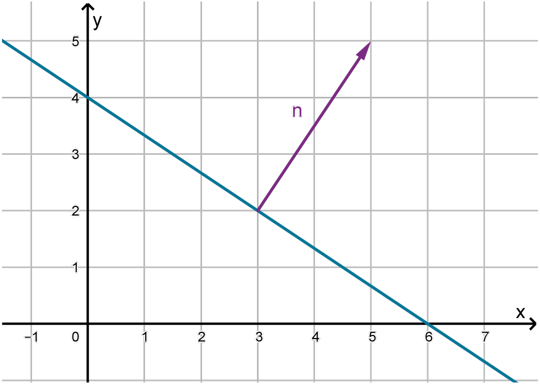
Goals:

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?



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.

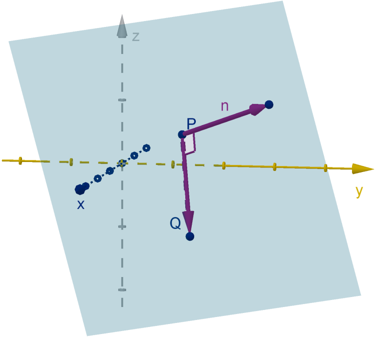


Figure: A plane, its normal vector $\vec{n}$, and a vector $\overrightarrow{PQ}$ in the plane

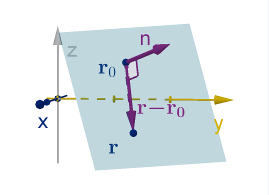
Theorem

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.

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**.

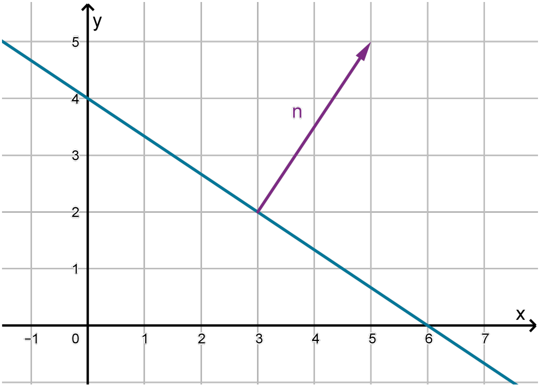
Example 5.3.2

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.

Using the Normal Vector to Compute Distance

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

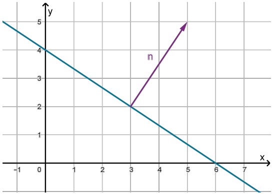
Let $P_1 = (7, 2)$ and $P_2 = (4, 0)$.

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?

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

$$\frac{L(q_1, \ldots, q_k)}{\vec{n}}.$$



The Distance from a Plane

Compute the geometric distance from the origin to the plane $6x + 8y + 3z - 24 = 0$.

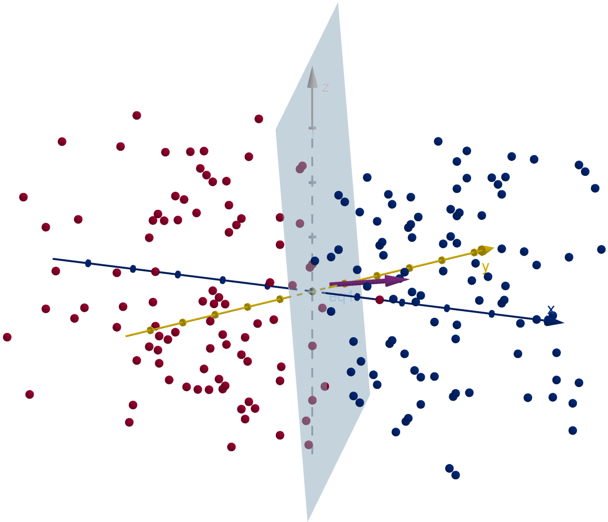
Support Vector Machines

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.


z
y
x
eq1

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?

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?

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**.

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?

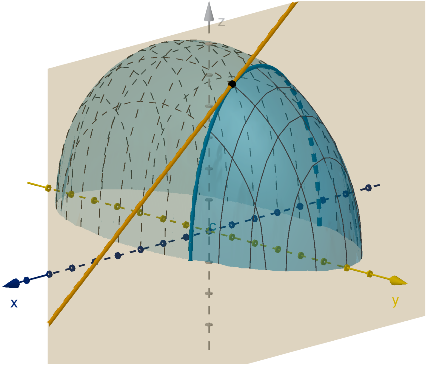


Figure: The tangent line to $z = f(x, y)$ in the x direction

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 $\vec{u}$.

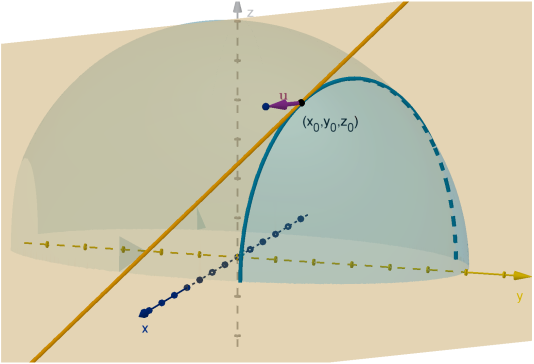


Figure: The tangent line to $f(x, y)$ in the direction of $\vec{u}$

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?

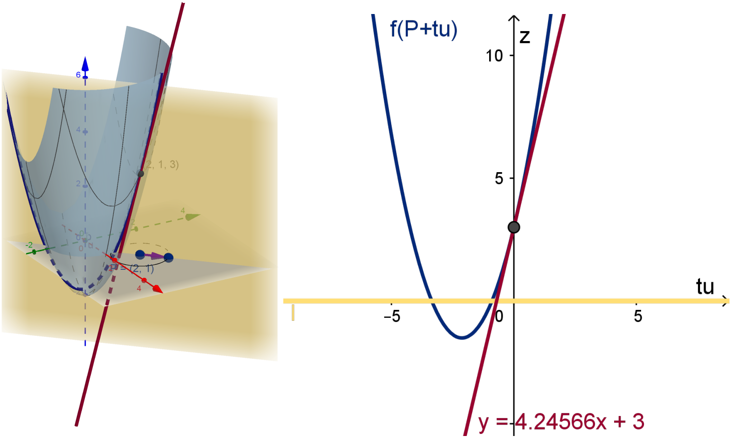


Figure: A cross section of $z = f(x, y)$ and a tangent line in the direction of $\vec{u}$

What Is the Gradient 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:

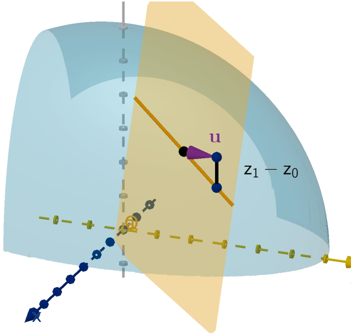
1. The gradient vector is a function of (x, y). Different points have different gradients.
2. $\vec{u}_{\text{max}}$, which maximizes $D_{\vec{u}} f$, points in the same direction as ∇f.
3. $\vec{u}_0$, which is tangent to the level curves, is orthogonal to ∇f.

How Do We Compute a Directional Derivative?

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 $\mathbb{R}^n$.
- $D_{\vec{u}}f = \nabla f \cdot \vec{u}$

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 $D_{\vec{u}}f(x, y) = 0$?

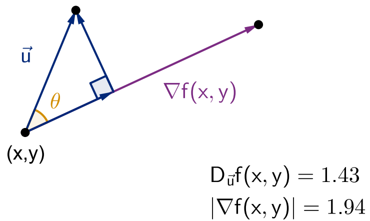


Figure: The angle between the gradient of f and a unit vector

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 $D_{\vec{u}} f(1, 2)$.

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?

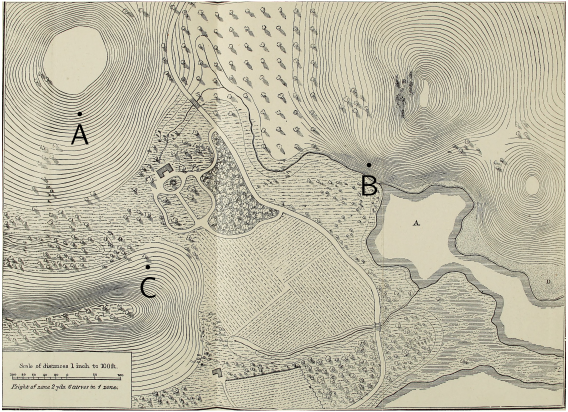


Figure: A topographical map

Application 5.4.7

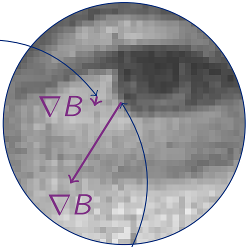
Edge Detection

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.

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 $(2, 1, -3)$.

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.

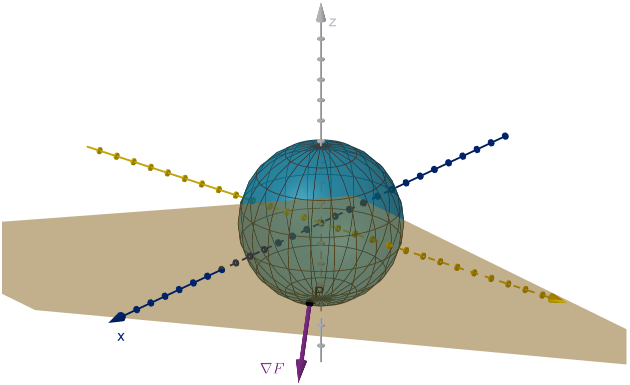


Figure: The level surface $x^2 + y^2 + z^2 = 14$, its tangent plane and ∇F.

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?

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?

How Can We Visualize a Composition with a Multivariable Function?

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.

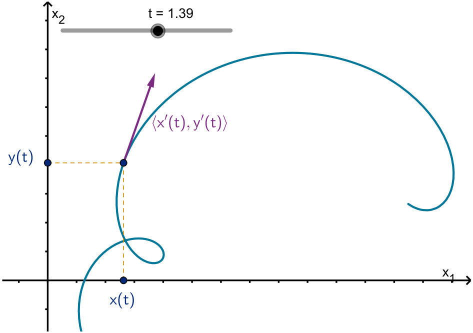


Figure: A particle whose position is defined by $x(t)$ and $y(t)$, the path it follows and its velocity vector

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 $y(t)$.

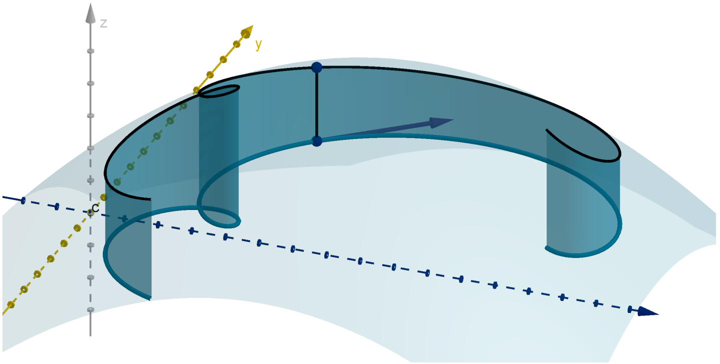


Figure: The composition $f(x(t), y(t))$, represented by the height of $z = f(x, y)$ over the path $(x(t), y(t))$

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.

Using the Chain Rule

If $P = R - C$ and we have $R = 100w$ and $C = 3000 + 70w - 0.1w^2$, calculate $\frac{dP}{dw}$.

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

$$\frac{df}{dt} = \nabla f(x, y, z) \cdot \langle x'(t), y'(t), z'(t)\rangle$$

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 $\frac{dP}{dT}$.

A Composition with Limited Information

Suppose $g(p, q, r) = re^{p^2 q}$. 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

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 $(3, 3)$?

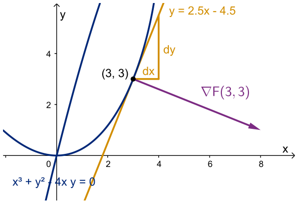


Figure: The graph of $F(x, y) = x^3 + y^2 - 4xy = 0$, its tangent line at $(3, 3)$, and the gradient of F

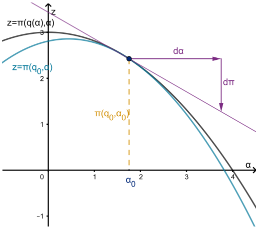
Main Ideas

- $\frac{dy}{dx}$ 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.

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?



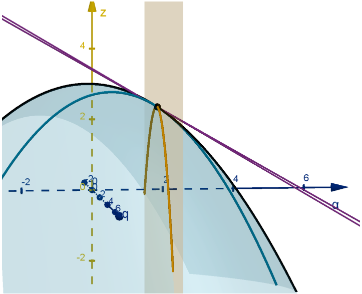


Figure: Two graphs of $z = \Pi(q, \alpha)$, one where q changes to be the optimal choice for each α and one where q is fixed at q_0, the optimal choice for α_0

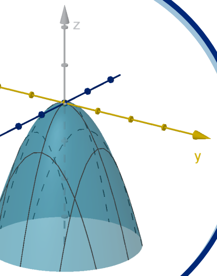
Summary Questions

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?



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.

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.

Where Do Local Extremes Lie?

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 horizontal.

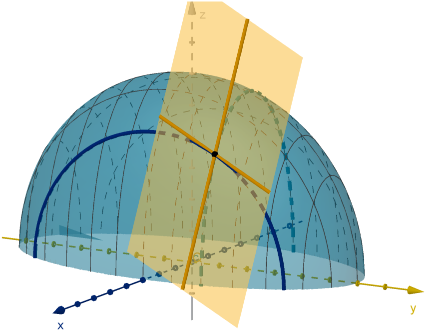


Figure: Tangent lines must have slope 0 at a local max.

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.

Finding Critical Points

The function $z = 2x^2 + 4x + y^2 - 6y + 13$ has a minimum value. Find it.

How Do We Identify Two-Variable Local Maximums and Minimums?

A critical point could be a local maximum. In this case f curves downward in every direction.

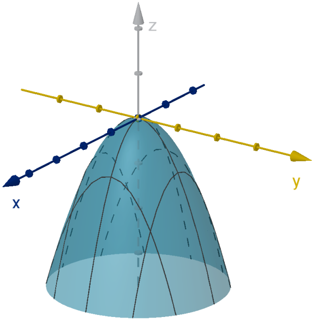


Figure: A local maximum at $(0, 0)$

A critical point could be a local minimum. In this case f curves upward in every direction.

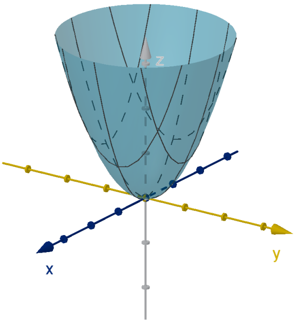


Figure: A local minimum at $(0, 0)$

A critical point could be neither. f curves upward in some directions but downward in others. This configuration is called a **saddle point**.

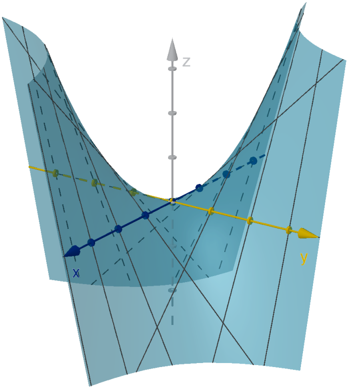


Figure: A saddle point at $(0, 0)$

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.

Classifying a Critical Point

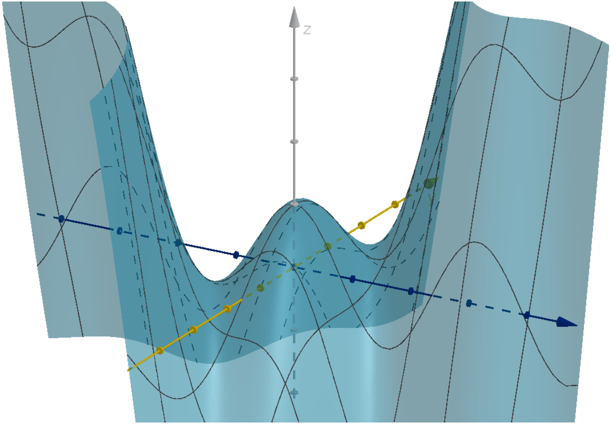


Figure: The graph $z = \cos(2x + y) + xy$ with a local maximum at $(0, 0)$

Question 5.6.6

How Do We Find Global Extremes?

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.

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

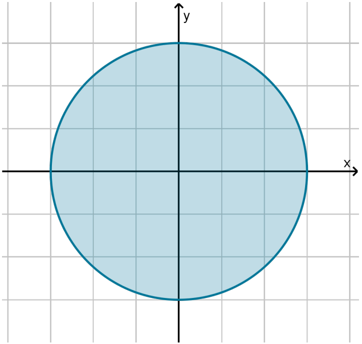


Figure: $x^2 + y^2 \leq 9$ is closed.

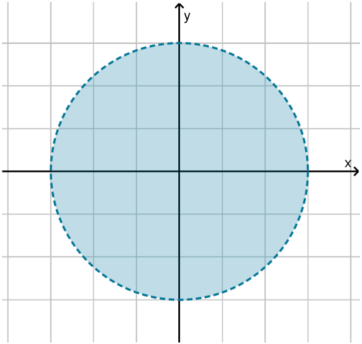


Figure: $x^2 + y^2 < 9$ is not closed.

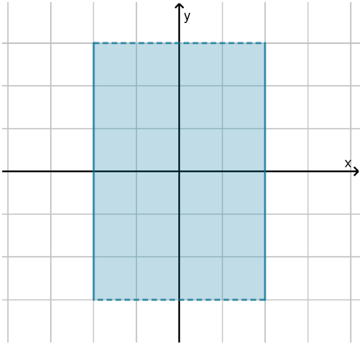


Figure: $-2 \leq x \leq 2$ and $-3 < y < 3$ is not closed.

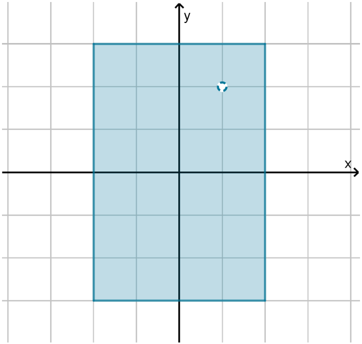


Figure: $-2 \leq x \leq 2$ and $-3 \leq y \leq 3$ and $(x, y) \neq (1, 2)$ is not closed.

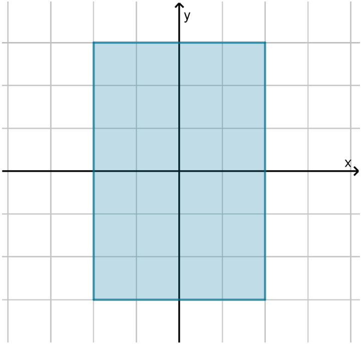


Figure: $-2 \leq x \leq 2$ and $-3 \leq y \leq 3$ is bounded.

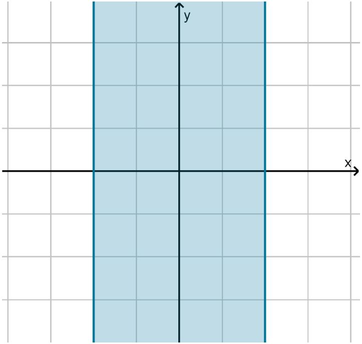


Figure: $-2 \leq x \leq 2$ is unbounded.

Example 5.6.7

Finding a Global Maximum

Consider the function $f(x, y) = x^2 + 2y^2 - x^2y$ on the domain

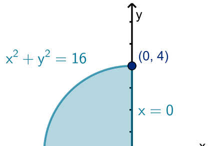
$$D = \{ \underbrace{(x, y)}_{\text{points in } \mathbb{R}^2} : \underbrace{x^2 + y^2 \leq 16, x \leq 0}_{\text{conditions}} \}$$

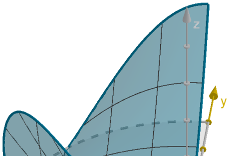
a Does f have a maximum value on D? How do we know?

b Find the critical points of f.

c Must one of the critical points be the maximum?

d Find the maximum of f.





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.

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?

Q32

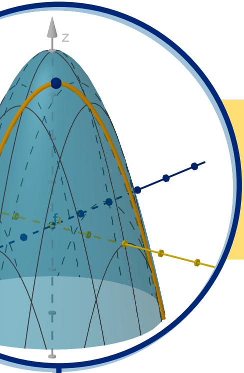
Let $f(x, y)$ be a differentiable function and let

$$D = \{(x, y) : y \geq x^2 - 4, x \geq 0, y \leq 5\}.$$

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.



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**.

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?

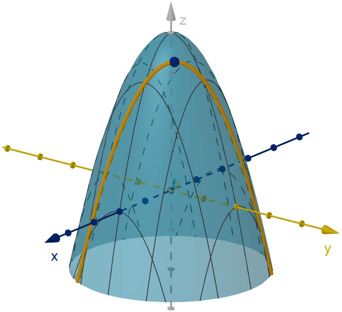


Figure: Maximizing f such that $x + y = 1$

How Do We Solve a Constrained Optimization?

The method of **Lagrange Multipliers** makes use of the following theorem.

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.

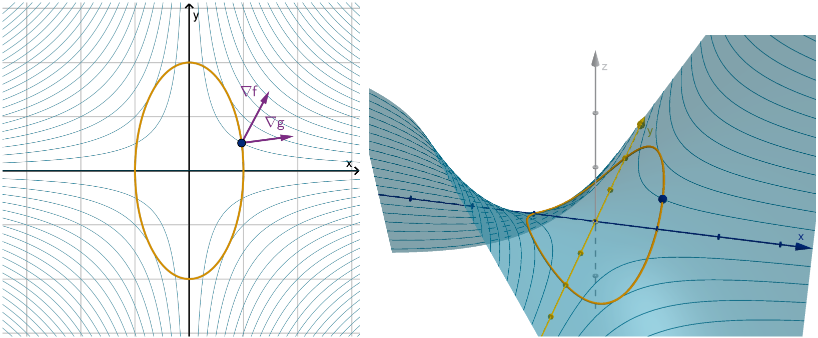


Figure: Where ∇f is not parallel to ∇g, we can travel along $g(x,y) = c$ and increase the value of f. This is because $D_{\vec{u}}f > 0$ for some $\vec{u}$ along the constraint.

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.

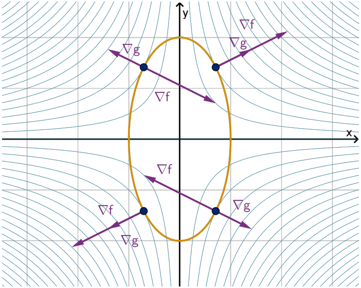


Figure: The four points that satisfy $\nabla f = \lambda \nabla g$ and $g(x, y) = c$.

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 $f(x, y, z) = x^4y^4z$ on the sphere $x^2 + y^2 + z^2 = 36$.

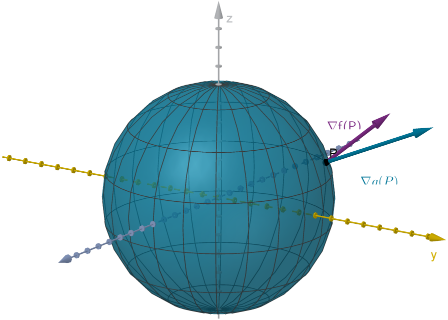
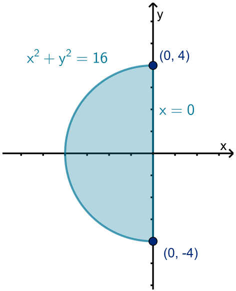


Figure: The gradient vector and level surface of a constraint function and the gradient vector of the objective function

Using the Extreme Value Theorem and Lagrange Multipliers

How can Lagrange multipliers help us find the maximum of $f(x, y) = x^2 + 2y^2 - x^2 y$ on the domain

$$D = \{(x, y) : x^2 + y^2 \leq 16, x \leq 0\}?$$



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.

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.

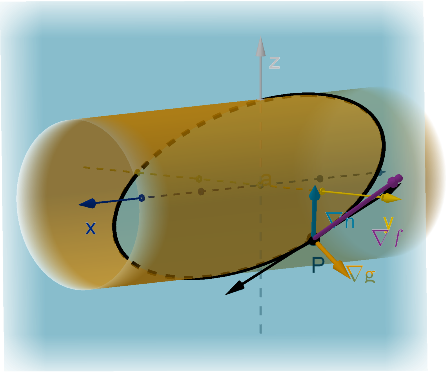


Figure: The intersection of the constraints $g(x, y, z) = c$ and $h(x, y, z) = d$

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 $\nabla h(P)$.

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:

$$\begin{aligned} f_x(x,y,z) &= \lambda g_x(x,y,z) + \mu h_x(x,y,z) & \qquad g(x,y,z) &= c \\ f_y(x,y,z) &= \lambda g_y(x,y,z) + \mu h_y(x,y,z) & \qquad h(x,y,z) &= d \\ f_z(x,y,z) &= \lambda g_z(x,y,z) + \mu h_z(x,y,z) \end{aligned}$$

We therefore generally expect this system to have a finite number of solutions, though there are plenty of counterexamples to this expectation.

Summary Questions

Q1 What is a constraint?

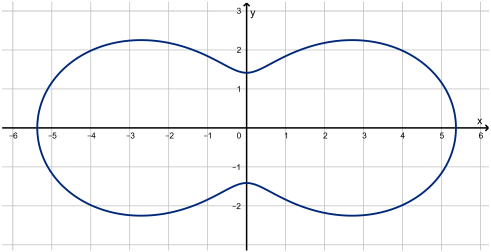
Q2 What equations do you write when you apply the method of Lagrange multipliers?

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?

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.



Q10

Show that $(3, 3)$ is not a local maximum of $f(x, y) = 2x^2 - 4xy + y^2 - 8x$ on the graph $x^3 + y^3 = 6xy$.

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

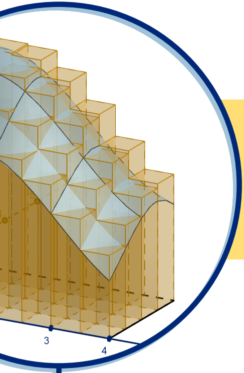
a How are the questions different?

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?

Q20

Consider the function $f(x, y) = x^2 + 6xy + 9y^2 + 5$. Find the maximum and minimum values of f on the domain
$D = \{(x, y) : y \geq x, x \geq 0, x^2 + y^2 \leq 10\}$



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.

How Do We Approximate the Volume Under $z = f(x, y)$?

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**.

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

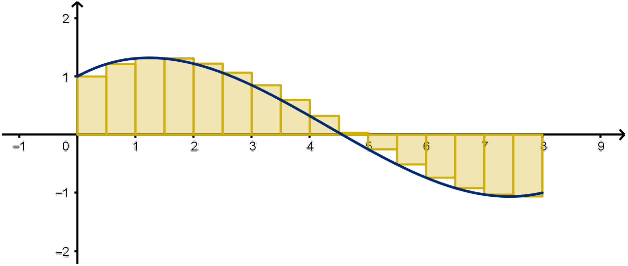
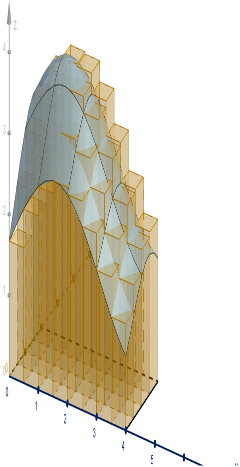


Figure: The area under $y = f(x)$ approximated by rectangles

A similar method approximates the signed volume under the graph $z = f(x, y)$ (where volume below the xy-plane counts as negative). We divide the domain



$$0 \leq x \leq 4$$
$$0 \leq y \leq 2$$

into subrectangles of area A. We draw a prism over each rectangle whose height is the value of the function over some test point (x_i^*, y_i^*).

$$\text{Volume} \approx \sum_{i=1}^{n} f(x_i^*, y_i^*)A.$$

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

$$\text{Volume} \approx \sum_{i=1}^{n} f(x_i^*, y_i^*) A_i.$$

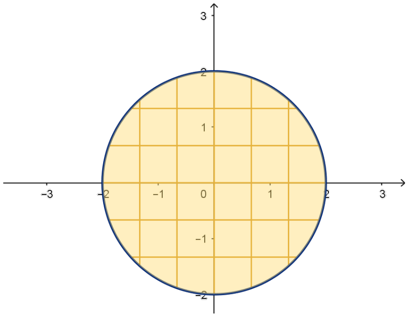


Figure: A domain subdivided into irregular subregions

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.

Definition

Let D be a domain in $\mathbb{R}^2$. For a given division of D into n subregions denote

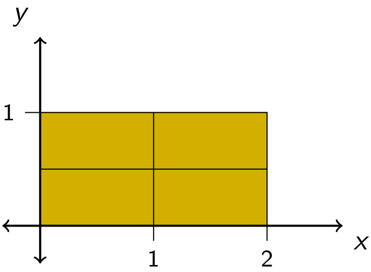
- A_i, the area of the i^{th} region.
- (x_i^*, y_i^*), any point in the i^{th} region
- $|A|$ is the diameter of the largest region.

We define the **double integral** of $f(x, y)$ to be a limit over all possible divisions of D.

$$\iint_D f(x, y) dA = \lim_{|A| \to 0} \sum_{i=1}^{n} f(x_i^*, y_i^*) A_i$$

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.



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.

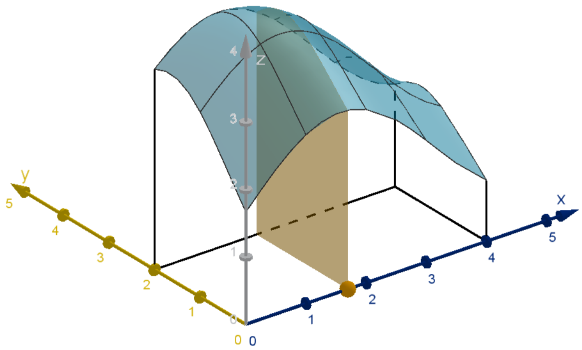


Figure: Cross sections of the region below the graph: $z = f(x, y)$

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

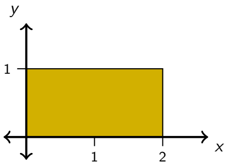
$$\iint_D f(x,y)\; dA = \int_c^d \left(\int_a^b f(x,y)\; dx \right)\; dy$$

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.

Example 6.1.4

Using Fubini's Theorem

Compute $\iint_D x^2 y \; dA$, where D is the region shown here:



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

$$\int_a^b \int_c^d f(x)g(y)dydx = \left(\int_a^b f(x)dx\right)\left(\int_c^d g(y)dy\right)$$

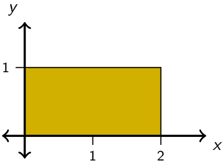
We won't be able to use this theorem all the time. It has two important requirements:

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:



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.

Integrating rainfall per square kilometer gives the total rain that fell in a watershed.

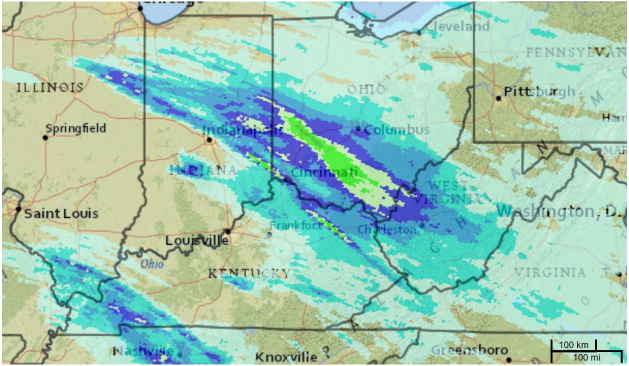


Figure: A rainfall density map

Integrating watts per square meter on a solar array gives the total energy generated.

Figure: Solar panels

By Jud McCranie - Own work, CC BY-SA 4.0

https://commons.wikimedia.org/w/index.php?curid=70132767

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.

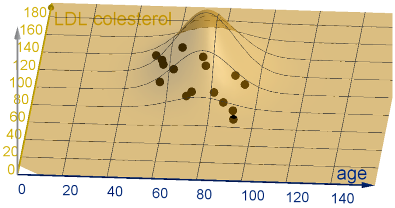


Figure: A highly correlated set of observations and an uncorrelated joint density function

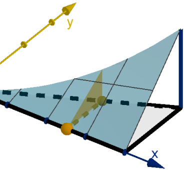
Summary Questions

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?



Section 6.2

Double Integrals over General Regions

Goals:

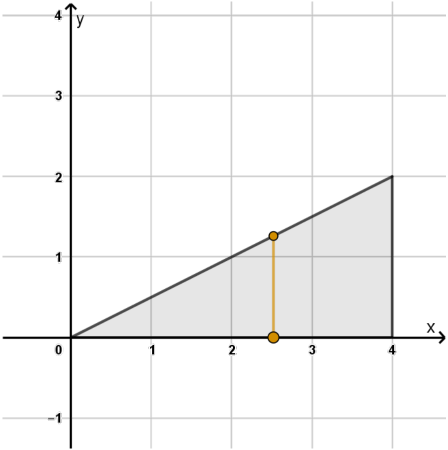
1. Set up double integrals over regions that are not rectangles.
2. Evaluate integrals where the bounds contain variables.
3. Decide when to make $\int dy$ the outer integral, and compute the change of bounds.

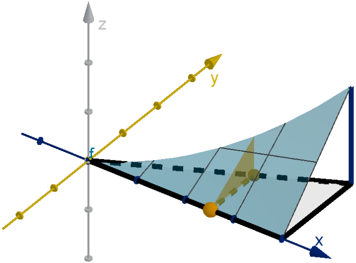
Example 6.2.1

Integrating Over a Polygon

Let D be the triangle with vertices $(0,0)$, $(4,0)$ and $(4,2)$. Calculate

$$\iint_D 4xy \; dA$$





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.

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:

$$\iint_D cf(x,y)dA = c\iint_D f(x,y)dA$$

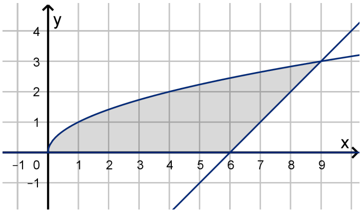
3. If D is the union of two non-overlapping subdomains D_1 and D_2 then

$$\iint_D f(x,y)dA = \iint_{D_1} f(x,y)dA + \iint_{D_2} f(x,y)dA$$

A Region Without a (Single) Bottom Curve

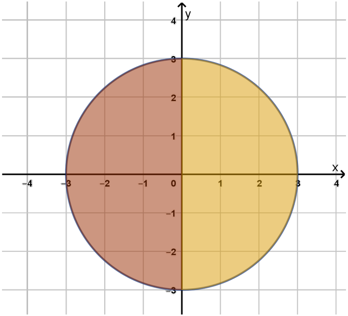
Let D be the region bounded by $y = \sqrt{x}$, $y = 0$ and $y = x - 6$. Calculate

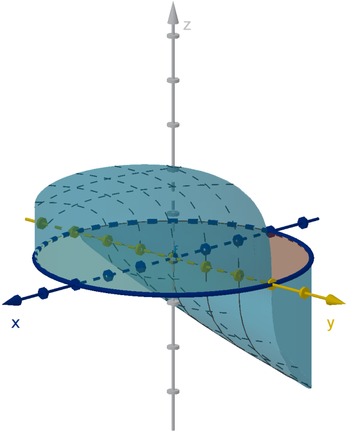
$$\iint_D (x + y)\ dA.$$



Using Anti-Symmetry

Let D be the region $x^2 + y^2 \leq 9$. Evaluate $\iint_D \sqrt[3]{x}\sqrt{y+3}\,dA$.





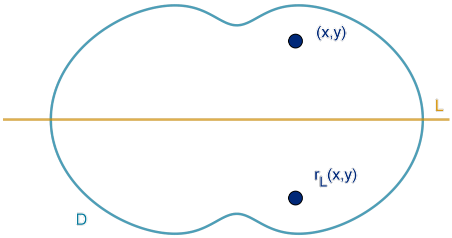
Main Idea

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:

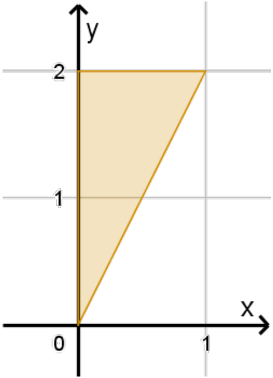
$$f(r_L(x,y)) = -f(x,y).$$



Using Order to Manipulate the Integrand

Let D be the triangle with vertices $(0,0)$, $(0,2)$ and $(1,2)$. Calculate

$$\iint_D e^{(y^2)}\ dA.$$



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.

Area of a Domain

Theorem

The area of a region D can be calculated:

$$\iint_D 1 \, dA.$$

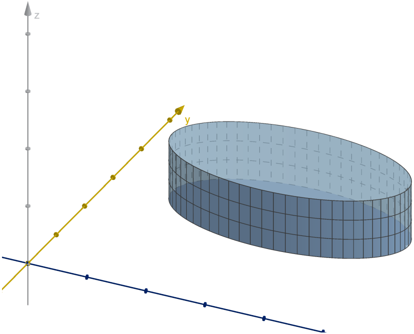


Figure: A solid of height 1 over a domain D

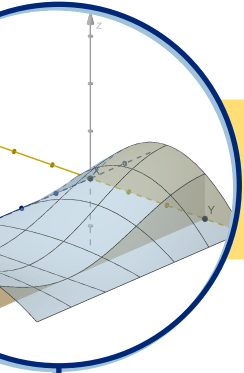
Summary Questions

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?



Section 6.3

Joint Probability Distributions

Goals:

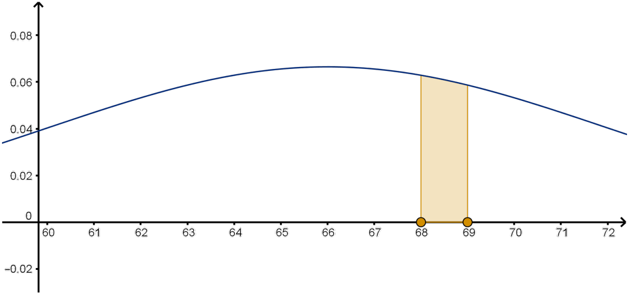
1. Integrate a joint density function to calculate a probability.
2. Recognize when random variables are independent.

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 $\int_a^b f(x)dx$.



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 $\mathbb{R}^2$

$$P((X, Y) \text{ lies in } D) = \iint_D f_{X,Y}(x, y)\, dA.$$

Using a Joint Density Function

Suppose the random variables X and Y have the joint density function

$$f_{X,Y}(x,y) = \begin{cases} x+y & \text{if } 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}.$$

Compute the probability that X is at least twice as large as Y.

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.

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).

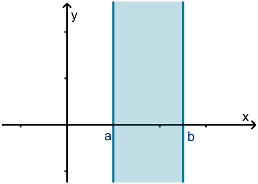
What Is a Marginal Density Function?

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 $P(a \leq X \leq b)$.

$$P(a \leq X \leq b) = \int_a^b \int_{-\infty}^{\infty} f_{X,Y}(x,y)\ dydx$$



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.

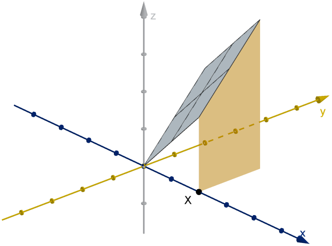


Figure: The marginal density function $f_X(x)$, represented as cross-sections under $z = f_{X,Y}(x, y)$

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

$$f_{X,Y}(x,y) = \begin{cases} \frac{12-0.012x}{1000}\left(\frac{y}{x^2} - \frac{y^2}{x^3}\right) & \text{if } 0 \le x \le 1000, 0 \le y \le x \\ 0 & \text{otherwise} \end{cases}$$

- 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?

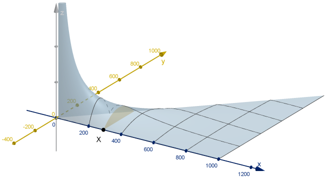


Figure: The marginal density function of X, represented as an area under the graph of $z = f_{X,Y}(x, y)$ (z-axis not to scale)

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.

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.

Why Do We Need Joint Distributions?

In some cases we don't.

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.

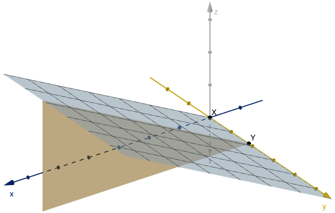


Figure: The density function for Darmok and Jalad's arrival times

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

$$f_{X,Y}(x,y) = \begin{cases} \frac{3\pi}{12\pi-8}\cos\left(\frac{\pi}{2}x\right)\left(2y-y^2\right) & \text{if } 0 \le x \le 6 \text{ and } 0 \le y \le 4 \\ 0 & \text{otherwise} \end{cases}$$

$f_{X,Y}(x,y)$ factors into the marginal density functions

$$f_X(x) = \begin{cases} \frac{\pi}{3\pi-2}\cos\left(\frac{\pi}{2}x\right) & \text{if } 0 \le x \le 6 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{3}{4}(2y-y^2) & \text{if } 0 \le y \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Thus we can conclude that X and Y are independent.

We can see independence in the cross sections of $z = f_{X,Y}(x, y)$.

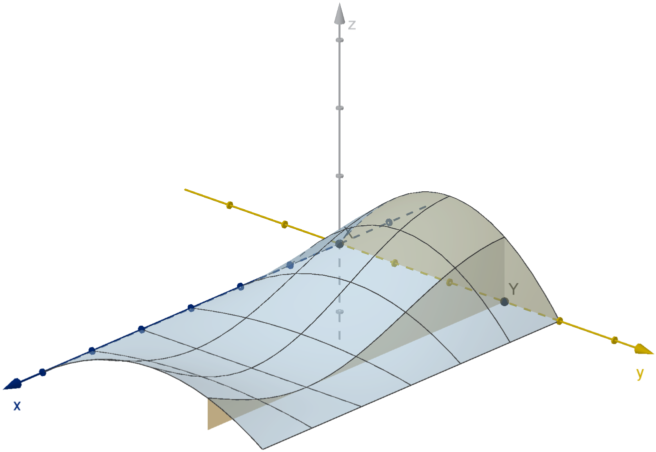


Figure: An independent joint density function and its cross sections

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

$$\begin{aligned} x =& \frac{1}{a} y^2 \\ x =& \frac{1}{b} y^2 \end{aligned}$$

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:

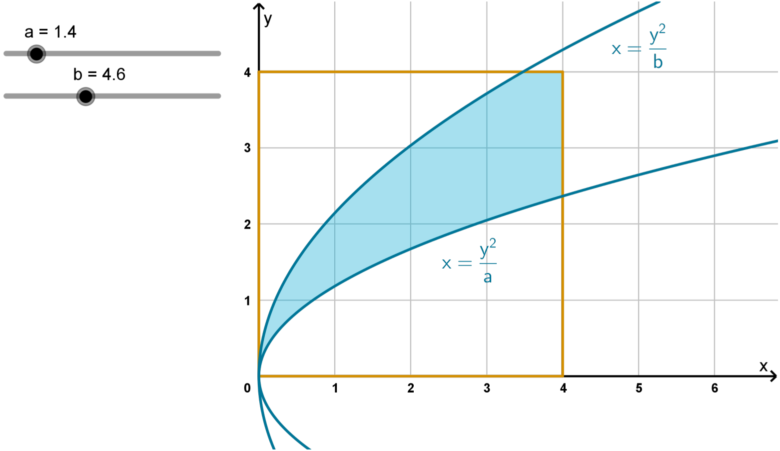


Figure: The region where $a \leq g(x, y) \leq b$

Computing expected values of functions by producing a new density function would be intractable. 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

$$f_{X,Y}(x, y) = \begin{cases} \frac{12-0.012x}{1000}\left(\frac{y}{x^2} - \frac{y^2}{x^3}\right) & \text{if } 0 \leq x \leq 1000, 0 \leq y \leq x \\ 0 & \text{otherwise} \end{cases}$$

Compute $E[X]$.

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.

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:

$$f_{X,Y} = \begin{cases} \frac{x}{32} & \text{if } 0 \leq x \leq 4 \text{ and } 0 \leq y \leq 4 \\ 0 & \text{otherwise} \end{cases}.$$

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

$$f_{ave} = \frac{\iint_D f(x, y)\ dA}{\iint_D 1\ dA}$$

Covariance and Correlation

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**.

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.

$$\begin{aligned}
\text{cov}(X, Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - E[X])(y - E[Y])f_{X,Y}(x, y)\ dydx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - E[X])(y - E[Y])g(x)h(y)\ dydx \\
&= \left(\int_{-\infty}^{\infty}(x - E[X])g(x)\ dx\right)\left(\int_{-\infty}^{\infty}(y - E[Y])h(y)\ dy\right) \\
&= \left(\int_{-\infty}^{\infty}xg(x)\ dx - E[X]\right)\left(\int_{-\infty}^{\infty}yh(y)\ dy - E[Y]\right) \\
&= (0)(0)
\end{aligned}$$

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.

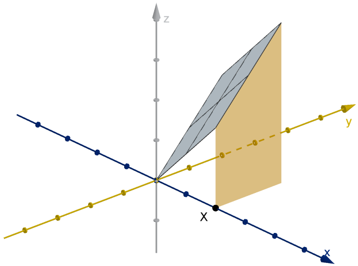
Summary Questions

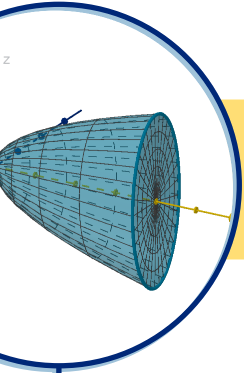
- **Q1** How do we use a joint density function to compute the probability of a certain set of outcomes?
- **Q2** What is a marginal density function and how do we compute it?
- **Q3** What does it mean for two random variables to be independent?
- **Q4** How can we tell from the graph of a joint density function that the two random variables are independent?

Q20

How does the distribution of Y change as X takes different values, given the following joint density function?

$$f_{X,Y}(x,y) = \begin{cases} x+y & \text{if } 0 \le x \le 1 \text{ and } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$





Section 6.4

Triple Integrals

Goals:

1. Set up triple integrals over three-dimensional domains.
2. Evaluate triple integrals.

How Do We Integrate a Three-Variable Function?

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.

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 $\iiint_R 3zy + x^2 \; dV$.

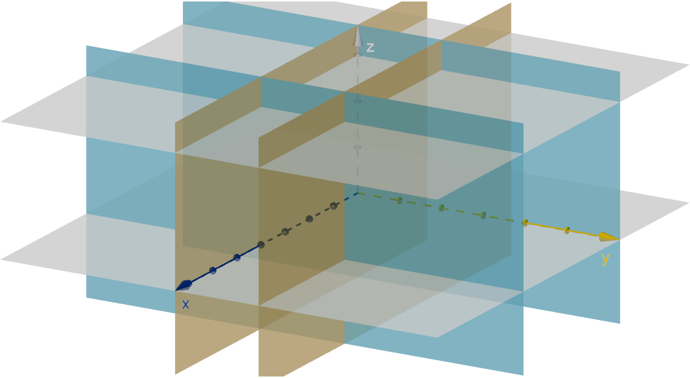


Figure: A Rectangular Prism

How Do We Interpret Triple Integrals Geometrically?

$\int_0^3 f(x, y, z)\, dz$ 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.

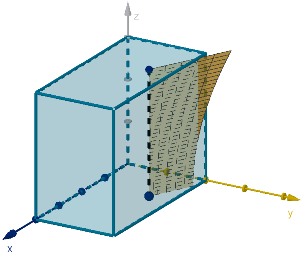


Figure: $\int_0^3 f(x, y, z)\, dz$, represented as an area in a zw-plane

$\int_0^2 \int_0^3 f(x, y, z)\, dzdy$ 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.

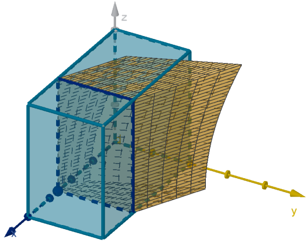


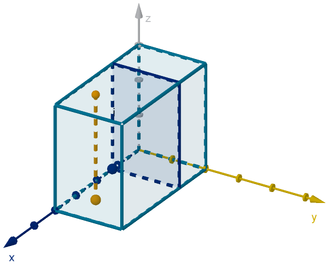
Figure: $\int_0^2 \int_0^3 f(x, y, z)\, dzdy$, represented as a volume in yzw-space

Triple Integrals in Math and Science

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.

Density lets us visualize a triple integral without referring to a fourth (geometric) dimension.

$\int_0^3 f(x, y, z)\ dz$ computes the density of the vertical segments at each (x, y).



$\int_0^2 \int_0^3 f(x, y, z)\ dzdy$ computes the density of the rectangle at each x.

$$\int_0^4 \int_0^2 \int_0^3 f(x, y, z)\ dzdydx$$ computes the total mass of the prism.

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 $\iiint_R 4yz\; dV$.

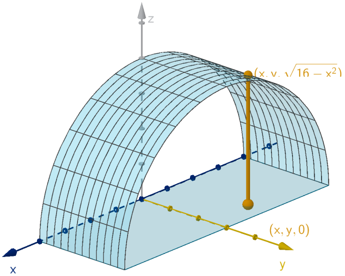


Figure: The region between $x^2+y^2=16$ and the xy-plane

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?

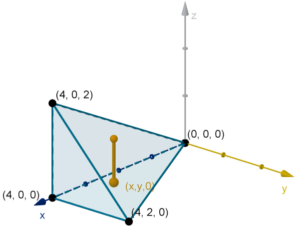


Figure: z bounds of T

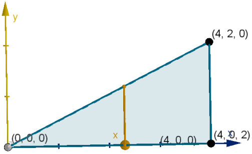


Figure: x, y bounds of T

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 $\iiint_D f(x, y, z)dV$.

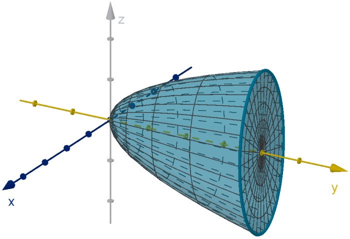


Figure: A region bounded by a paraboloid and a plane

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

$$\int_{x_1}^{x_2}\int_{y_1}^{y_2}\int_{z_1}^{z_2} f(x)g(y)h(z)\ dzdydx$$
$$= \left(\int_{x_1}^{x_2} f(x)\ dx\right)\left(\int_{y_1}^{y_2} g(y)\ dy\right)\left(\int_{z_1}^{z_2} h(z)\ dz\right)$$

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:

$$\begin{aligned}
&\int_0^4 \int_0^2 \int_0^3 3zy \; dzdydx + \int_0^4 \int_0^2 \int_0^3 x^2 \; dzdydx \\
=&3 \int_0^4 dx \int_0^2 y \; dy \int_0^3 z \; dz + \int_0^4 x^2 \; dx \int_0^2 dy \int_0^3 dz \\
=&3 \cdot 4 \int_0^2 y \; dy \int_0^3 z \; dz + 2 \cdot 3 \int_0^4 x^2 \; dx
\end{aligned}$$

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?

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 $\iiint_R g(x, y, z)dV$.

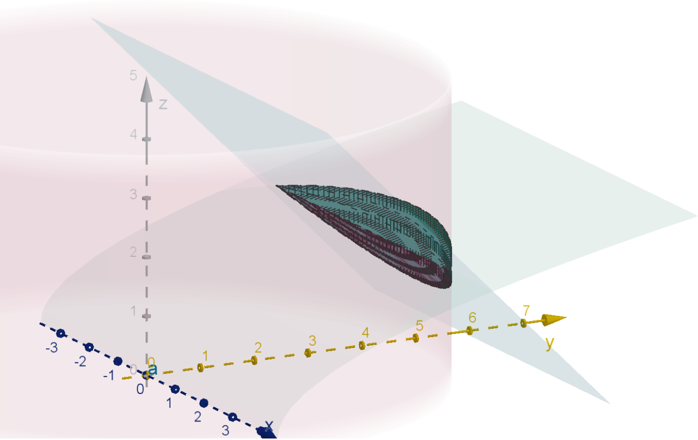


Figure: The region enclosed by $x^2 + y^2 = 25$, $z = 6 - y$, and $z = \sqrt{y}$

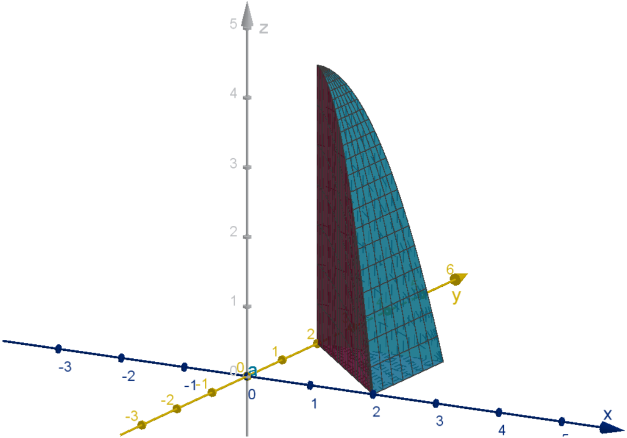
Q20

Cheng is integrating over R, the region given by $x^2 + y^2 + z^2 \leq 25$. He gives the following setup. Is this valid?

$$\int_{-\sqrt{25-y^2-z^2}}^{\sqrt{25-y^2-z^2}} \int_{-\sqrt{25-x^2-z^2}}^{\sqrt{25-x^2-z^2}} \int_{-\sqrt{25-x^2-y^2}}^{\sqrt{25-x^2-y^2}} f(x,y,z)\ dzdydx$$

Q36

Rewrite the integral $\int_0^2 \int_{2-x}^2 \int_0^{4-x^2} f(x, y, z)\ dzdydx$ as an integral with the differential $dxdzdy$.


z
x
y