MAT133 Study Guide

MAT133 in simpler terms

Ian Zhang

July 25, 2023

Disclaimer: I wrote this guide based off what my thought process would be when tackling each concept. It may be the same as yours, or it may be different (seeing that you're here you probably want to know my process anyways). The purpose of this guide is to help the reader better understand the concepts that will be tested in MAT133. I highly discourage memorizing each concept; rather, you should try and understand each concept. Also on a side note, I understand every concept learned in this class, so really don't be afraid to ask me questions.

Contents

1 Linear algebra extreme basics

1.1 Vector operations

A vector is essentially a point in n-dimensional space (denoted $\mathbb{R}^n$). Sounds complicated? Actually...it's really not. We've been using vectors ever since we were introduced to the Cartesian plane. First, let's define a couple vectors so we have something to work with. Let $u, v \in \mathbb{R}^n$ such that $u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$ and $v = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ Actually, to make it even simpler, let's confine ourselves to $\mathbb{R}^2$, so $u = \begin{bmatrix} u_1 \\ u_1 \end{bmatrix}$ and $v = \begin{bmatrix} v_2 \\ v_2 \end{bmatrix}$[1]. u and v are literally just points in the Cartesian plane that we are all familiar with. The point (2,1)? It's a vector. Is (0,0) a vector? You bet. What about (-40920192,1029102.10201)? Yup. What about 5? No. This is not a vector (higher level linear algebra classes will say otherwise but for MAT133 purposes it's NOT a vector) because it doesn't have direction and is just a number (this is called a **scalar**). See a pattern yet?

Next, let's go over some vector properties. Using the vectors u and v that we defined earlier, then

1. $u + v = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \end{bmatrix}$
2. $\alpha u = \begin{bmatrix} \alpha u_1 \\ \alpha u_1 \end{bmatrix}$ for some scalar $\alpha \in \mathbb{R}$

The **dot product** of two vectors is a scalar and is defined as

$$u \cdot v = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = u_1 v_1 + u_2 v_2$$

The **magnitude**, or length, of a vector is defined as

$$\|u\| = \sqrt{u_1^2 + u_1^2}$$

A **unit vector** is a vector with magnitude 1.

$$\|u\| = 1$$

Note that I used 2-dimensional vectors simply for ease of understanding, but all of these formulas and properties apply to vectors in $\mathbb{R}^n$.

[1] MAT133 may use $u = u_1 i + u_2 j$ instead. I highly discourage using this format while working with vectors, however, it is best to still write final answers in the format that is required by the course instructor(s)

1.2 Distance between 2 points

The distance between 2 points can be defined using vectors. Let's bring back vectors u and v. The distance between u and v is equal to $\|u - v\|$. That is,

$$\begin{aligned}\|u - v\| &= \sqrt{\|u\|^2 + \|v\|^2 - 2(u \cdot v)} \\ &= \sqrt{(u_1 - v_2)^2 + ... + (u_n - v_n)^2}\end{aligned}$$

1.3 Matrix operations

This is a matrix:

$$\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

Basically, a matrix is an $m \times n$ array of numbers. It has m rows and n columns, so the matrix shown above is a 2×3 matrix. A vector is actually a special type of matrix that's either $1 \times n$ or $n \times 1$. On that note, matrix operations are the exact same as vector operations. Let's take a look.

1.3.1 Matrix addition

Define 2 matrices A, B:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, B = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

Then

$$A + B = \begin{bmatrix} a + e & b + f \\ c + g & d + h \end{bmatrix}$$

Matrix addition is just adding corresponding components of 2 matrices. Notice how matrix addition would not hold if A and B were different sizes because we would have a case where we have to add a component of a matrix to nothing.

1.3.2 Matrix multiplication

Matrix multiplication is a bit different. Let A be a 2×3 matrix and B be a 3×2 matrix.

$$A = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}, B = \begin{bmatrix} g & h \\ i & j \\ l & m \end{bmatrix}$$

Then AB is equal to

$$AB = \begin{bmatrix} ag + bi + cl & ah + bj + cm \\ dg + ei + fl & dh + ej + fm \end{bmatrix}$$

Each component of the product of two matrices is just the dot product. Furthermore, if A is an $m \times n$ matrix, then B must have n rows, or else we would end up multiplying components of M with nothing or vice versa. Let A be an $m \times n$ matrix and B be a $n \times o$ matrix. Then the size of the matrix AB is $m \times o$ and the matrix BA would not exist if $o \neq m$. Unlike scalar multiplication, matrix multiplication has a lot to do with direction of multiplication so pay extra attention to that.

1.4 Determinant

The determinant of a 2×2 matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

is computed by subtracting the product $a_{12}a_{21}$ from the product $a_{11}a_{22}$. That is,

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}$$

1.5 System of Linear Equations

You've probably seen this high school:

$$\begin{cases} x + y = 2 \\ 2x + y = 13 \end{cases}$$

You probably even know how to solve this:

Solution. Using the equation $x + y = 2$, we obtain $x = 2 - y$. Substituting that into the second equation,

$$\begin{aligned} 2x + y &= 13 \\ 2(2 - y) + y &= 13 \\ 4 - y &= 13 \\ y &= -9 \end{aligned}$$

and finally $x = 11$. ■

It's very easy to substitute when you have 2 unknowns and 2 equations but what happens if you have 3 unknowns? Or 4? Or 500? Then it becomes very tedious to substitute! Instead, we can do this matrix operation called "row reduction". Observe:

Example 1. Using the same same example as above, solve the following system of equations

$$\begin{cases} x + y = 2 \\ 2x + y = 13 \end{cases}$$

Solution. Using the coefficients of the equation, we can obtain something like this:

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 2 & 1 & 13 \end{array}\right]$$

Using row reduction operations, we can row reduce this matrix.

$$\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 2 & 1 & 13 \end{array}\right] \xrightarrow{R_2 - 2R_1} \left[\begin{array}{cc|c} 1 & 1 & 2 \\ 0 & 1 & 9 \end{array}\right] \xrightarrow{R_1 + R_2} \left[\begin{array}{cc|c} 1 & 0 & 11 \\ 0 & -1 & 9 \end{array}\right] \xrightarrow{-R_2} \left[\begin{array}{cc|c} 1 & 0 & 11 \\ 0 & 1 & -9 \end{array}\right]$$

Therefore, $x = 11, y = -9$ ■

We reached the same solution! You might be sitting there right now thinking to yourself *How did he just do that?* Luckily, I'm about to show you.

1.5.1 Row reduction

Given a system of equations, we can bring the coefficients into a matrix, like we did in the above example. Row reduction involves reducing a matrix to one of two forms: Row Echelon Form (REF) or Reduced Row Echelon Form (RREF). Here's the definition of both (pulled from Wikipedia):

Definition (Row Echelon Form (or REF))**.** This form satisfies the following conditions:

- All the rows consisting of only zeroes at are the bottom
- The pivot of a nonzero row is always strictly to the right of pivot of the row above and must be equal to 1

Definition (Reduced Row Echelon Form (or RREF))**.** This is the same form as REF but with one extra condition: each column containing a pivot must have zeroes in all it's other entries.

The form I arrived at in the example above is an example of RREF. I encourage you to search up what REF and RREF forms look like to get a better idea (it will make more sense if you can see the matrices).

For row reducing, there are 3 types of row operations you can do (called elementary row operations):

1. Add two rows
2. Multiply a row by a scalar multiple
3. Swap rows with one another

There is a pretty good set of steps to reduce a matrix. I don't really want to explain how to do it as it is highly intuitive so I'm going to just include a link: `https://www.sparknotes.com/math/algebra2/matrices/section4/`. One thing to note is always keep all 0 rows at the bottom (follows definition of REF/RREF) if they exist.

In row reducing a matrix corresponding to a system of equations, we can tell how many solutions there are by looking at the last row. If the last row looks like this:

$$\left[\begin{array}{cccc|c} 0 & 0 & \cdots & 0 & 0 \end{array}\right]$$

there are **infinite solutions**.

If the last row looks like this:

$$\left[\begin{array}{ccc|c} 0 & \cdots & a & b \end{array}\right]$$

there is **1 solution**.

If the last row looks like this:

$$\left[\begin{array}{ccc|c} 0 & \cdots & 0 & b \end{array}\right]$$

there are **no solutions**.

2 Game Theory

2.1 Zero sum games and payoff matrices

A zero sum matrix game is defined as a game where the pair of payoffs for each entry in the payoff matrix game is equal to 0. To explain this section, I'm going to introduce our own game to use to aid with my explanations.

Example 2. Let's pretend we have two companies, A and B. If both invest in C, A loses 50 dollars. If A invests in C and B invests in D, then A gains 100 dollars from B. If A invests in D and B invests in C, then B gains 100 dollars from A. If both invest in D, then B loses 100 dollars and gives it to A.

The payoff matrix for company A would look like this:

$$\begin{bmatrix} -50 & 100 \\ -100 & 100 \end{bmatrix}$$

Clearly, we can see that the optimal strategy for A is to play C over D. Even though if A plays C and B also plays C, A loses -50 dollars but if A plays D and B plays C, then A would lose 100. Moreover, if A wins, A gets 100 dollars either way, so it's less risky to always play C. Therefore, the strategy A would play is to always play C.

Since A is always going to play C, then in turn B will want to play C as well. Why? Well if A is going to play C and then B plays D, B will end up losing 100 dollars to A! But if both play C, then B will instead gain 50 dollars! Therefore, letting p and q represent the strategies that each company should play,

$$p = \begin{bmatrix} 1 & 0 \end{bmatrix}, q = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

It should be noted that elements of p and q represent the probability that the row/column player will play a certain move, so all the elements of each strategy must sum up to 1.

This game is an example of a zero sum game. Why? Because, for example, if A loses 50 dollars, then B gains the 50 dollars so the net loss of both sides added together is 0. Another thing to note is that these two strategies involve 0 chance. This means that these are strategies that each side for play **for sure**. These strategies are called **pure strategies**. **Mixed strategies** are strategies that involve chance, so each player's strategy has probabilities of making each move.

2.2 Fundamental theorem of Zero-Sum Games

Theorem (Fundamental theorem of Zero-Sum Games)**.** Let p and q represent the strategies of the Row and Column players respectively and let $p*$ and $q*$ denote the the optimal strategies of each player. Then

$$E(p*, q) \geq E(p, q) \geq E(p, q*)$$

there E(x,y) represents the expected payoff of the game.

This theorem actually makes a lot of sense. Looking back at the example game shown above, if A picks C, then the best option for B is to also pick C. Therefore, from A's POV, the best outcome would be if A chose A and B chose D, the next best would be if both chose optimally, the third would be B chose the optimal strategy but A didn't. Depending on the probability of playing each choice, the three different expected values could be equal. Conversely, if we were looking from B's

POV, then the theorem would be flipped.

The expected value of a game is computed using the following formula:

$$E(p,q) = pAq$$

where A is the payoff matrix, p, q are the strategies of the row and column players respectively.

2.3 Strictly determined game

A game is **strictly determined** if both players have optimal strategies that are pure strategies. The example above is an example of a strictly determined game. Strictly determined games are interesting because there's a very easy way of computing the expected values: by finding saddle points.

A **saddle point** of a payoff matrix A is an element A such that it is simultaneously the lowest value in it's row but highest in it's column. Let's look at the payoff matrix from earlier.

$$\begin{bmatrix} -50 & 100 \\ -100 & 100 \end{bmatrix}$$

From this matrix, we can see that -50 is a saddle point because it is less than 100 but greater than -100. That means our expected value for this game -50. Computing using $E(p,q) = pAq$,

$$\begin{aligned} E(p,q) &= pAq \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} -50 & 100 \\ -100 & 100 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} -50 & 100 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= -50 \end{aligned}$$

If a matrix doesn't have a saddle point, then it is not a strictly determined payoff matrix.

2.4 Non-strictly determined games

A non-strictly determined game is a game where the strategies aren't pure strategies. That is, both strategies involve a chance of playing a certain move. Given a payoff matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

the expected payoffs

$$E(p*, q) = E(p, q) = E(p, q*)$$

and the strategies $p*$ and $q*$ are given by

$$p* = \begin{bmatrix} \frac{a_{22}-a_{21}}{a_{11}+a_{22}-a_{12}-a_{21}} & \frac{a_{11}-a_{12}}{a_{11}+a_{22}-a_{12}-a_{21}} \end{bmatrix}, q* = \begin{bmatrix} \frac{a_{22}-a_{12}}{a_{11}+a_{22}-a_{12}-a_{21}} \\ \frac{a_{11}-a_{21}}{a_{11}+a_{22}-a_{12}-a_{21}} \end{bmatrix}$$

And the value of the game, v is given by

$$\begin{aligned} v &= \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}} \\ &= \frac{\det(A)}{a_{11} + a_{22} - a_{12} - a_{21}} \end{aligned}$$

3 Rates of change of univariable functions

3.1 Average rate of change

Given a function $f(x)$, the average rate of change from point $(a, f(a))$ to point $(b, f(b))$ is computed using

$$\frac{\Delta y}{\Delta x} = \frac{f(b) - f(a)}{b - a}$$

3.2 Relative rate of change

The relative rate of change from x_0 to x_1 is computed using

$$\text{Relative rate of change in } P = \frac{\Delta x}{x_0} = \frac{x_1 - x_0}{x_0}$$

3.3 Instantaneous rate of change

The instantaneous rate of change is the rate of change at any given point on a curve. Geometrically, if we draw a tangent line at that point (ie: a line that touches a curve once and only once (and doesn't go through the function)), the instantaneous rate of change is the slope of that line. Another name for instantaneous rate of change is **derivative**.

3.3.1 Approximating the instantaneous rate of change at a point

Given a function $y = f(x)$, the instantaneous rate of change at a point (a,f(a)) is approximated by

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

for some point $(b, f(b))$ on $f(x)$.

4 Differentiation

Let's begin the biggest chunk of the course. First, let's introduce some new notation. Given a function $y = f(x)$, the derivative of the function will be given by

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

Another way to write this notation is to use Leibniz notation,

$$f'(x) = \frac{dy}{dx}$$

The tangential line at a point $(a, f(a))$ on $y = f(x)$ is can be calculated using

$$y - b = \frac{dy}{dx}|_a(x - a)$$

The $\frac{dy}{dx}|_a$ is the same as writing $f'(a)$.
Similarly, for a second derivative, the notation would be

$$f''(x)$$

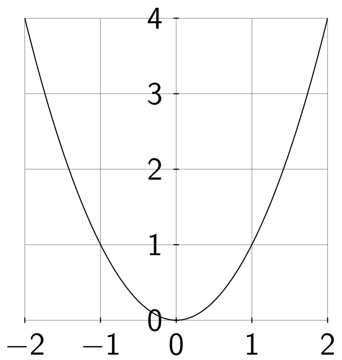
Leibniz is a bit more confusing:

$$\frac{d^y 2y}{dx^2}$$

Any derivative of higher degree than 3 would be written as

$$f^{(n)}(x) = \frac{d^n y}{dx^n}, n \geq 4$$

Let's take look at the graph $f(x) = x^2$.



Clearly, this graph has a minimum at $(0, 0)$. Visually, we can determine that the instantaneous rate of change at that minimum point is 0 (slope of the tangent line is 0). This means that $f'(0) = 0$

because the derivative represents the rate of change at each point on $f(x)$. Generally, we can use the **first derivative** to determine local maximum's and minimum's of a function. More about this topic will be given after we learn how to differentiate first.

Similarly, the second derivative gives us the concavity and points of inflection.

4.1 Rules of differentiation

There are many rules for differentiation, including rules for trigonometric functions, logarithmic functions, etc. Let's take a look at them.

4.1.1 Constant rule

The constant rule is the simplest. Take a function $f(x) = 1$. The rate of change throughout this function is clearly always 0, we can determine that the derivative of any constant is 0.

$$\frac{d}{dx}(c) = 0, c \in \mathbb{R}$$

4.1.2 Power rule

The power rule is given by

$$\frac{d}{dx}(x^n) = nx^{n-1}$$

Example 3. Find the derivative of the function

$$f(x) = x^2$$

Solution. From the rule, we can see that the derivative is computed by

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

■

4.1.3 Addition rule

The addition rule is given by

$$\frac{d}{dx}(f(x) + g(x)) = f'(x) + g'(x)$$

4.1.4 Product rule

The product rule is given by

$$\frac{d}{dx}(f(x)g(x)) = f'(x)g(x) + f(x)g'(x)$$

Example 4. Find the derivative of the following function:

$$f(x) = 3x^2$$

Solution. Letting $f(x) = 3, g(x) = x^2$,

$$\begin{aligned} f'(x) &= (0)x^2 + 3(2x) \\ &= 6x \end{aligned}$$

■

4.1.5 Quotient rule

The quotient rule is given by

$$\frac{d}{dx}\left(\frac{f(x)}{g(x)}\right) = \frac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$$

Example 5. Find the derivative of

$$h(x) = \frac{1}{x}$$

Solution. Let $f(x) = 1, g(x) = x$,

$$\begin{aligned} h'(x) &= \frac{0(x) - 1(1)}{x^2} \\ &= \frac{-1}{x^2} \end{aligned}$$

■

4.1.6 Chain rule

The chain rule is given by

$$\frac{d}{dx}(f(g(x)) = f'(g(x)) \cdot g'(x)$$

Example 6. Find the derivative of the following:

$$h(x) = (2x + 1)^2$$

Solution. Let $f(x) = x^2, g(x) = 2x + 1$.

$$\begin{aligned} h'(x) &= 2(2x+1) \cdot 2 \\ &= 4(2x+1) \\ &= 8x + 4 \end{aligned}$$

■

Now that we have gone over the most fundamental rules, let's go over some more complex derivatives.

4.2 Trigonometric derivatives

The rules for trigonometric derivatives needed for MAT133 are as follows:

$$\frac{d}{dx}\sin(x) = \cos(x) \qquad \frac{d}{dx}\cos(x) = -\sin(x)$$

Example 7. Using the rules you learned above, differentiate $\tan(x)$.

Solution.

$$\begin{aligned} (\tan(x))' &= \left(\frac{\sin(x)}{\cos(x)}\right)' \\ &= \frac{\cos(x)\cos(x) - \sin(x)(-\sin(x))}{\cos^2(x)} \\ &= \frac{1}{\cos^2(x)} = \sec^2(x) \end{aligned}$$

■

4.3 Transcendental derivatives

The derivatives of basic transcendental functions are

$$\frac{d}{dx}e^x = e^x \quad \frac{d}{dx}\ln(x) = \frac{1}{x} \quad \frac{d}{dx}\log_a x = \frac{1}{x\ln(a)} \quad \frac{d}{dx}a^x = a^x \ln a$$

5 Optimization

Optimization of a univariable function involves taking the first and second derivatives. The First Derivative Test allows us to find the maximum/minimum points while the Second Derivative Test allows us to determine the concavity of a function (which way the curve of a function is facing).

5.1 First Derivative Test

The First Derivative Test involves taking the derivative of a function and making it equal 0, then solving for x. This is calling finding **critical points**. Let's use an example to explain these tests. We're going to choose a basic function: $f(x) = x^3 + x^2$.
Taking the first derivative, we get

$$f'(x) = 3x^2 + 2x$$

To find the critical points, we're going to set this derivative to equal 0 and solve for x.

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

Critical points tell us one of 2 things: there's either a minimum or a maximum at the point on the curve. Looking at our function, this means there's either a maximum or minimum at $x = 0$ and the same follows for $x = \frac{-2}{3}$. Note that a point on the first derivative where x is undefined, such as if we have $f'(x) = \frac{1}{x} = 0$, the undefined point is also a critical point.
Now that we have two critical points, let's determine whether the are maximums or minimums by making a chart:

	$(-\infty, \frac{-2}{3})$	$(\frac{-2}{3}, 0)$	$(0, \infty)$
$f(x)$			
$f'(x)$			

At $x < \frac{-2}{3}$, we can see that $f'(x)$ is greater than 0, as shown below with $x = -1$

$$f'(-1) = 3(-1)^2 + 2(-1) = 1$$

A positive derivative means that $f(x)$ is increasing because there's a positive rate of change, so our table looks like this:

	$(-\infty, \frac{-2}{3})$	$(\frac{-2}{3}, 0)$	$(0, \infty)$
$f(x)$	increasing		
$f'(x)$	positive		

At $\frac{-2}{3} < x < 0$, we can see that $f'(x)$ is negative, as shown below with $x = \frac{-1}{3}$

$$f'(\frac{-1}{3}) = 3(\frac{-1}{3})^2 + 2(\frac{-1}{3}) = \frac{-1}{3}$$

A negative derivative means negative rate of change, so our table looks like this:

	$(-\infty, \frac{-2}{3})$	$(\frac{-2}{3}, 0)$	$(0, \infty)$
$f(x)$	increasing	decreasing	
$f'(x)$	positive	negative	

Doing the same for the last interval, we get this completed chart:

	$(-\infty, \frac{-2}{3})$	$(\frac{-2}{3}, 0)$	$(0, \infty)$
$f(x)$	increasing	decreasing	increasing
$f'(x)$	positive	negative	positive

Since the left hand side of $x = \frac{-2}{3}$ is increasing while the right hand side is decreasing, that means $x = \frac{-2}{3}$ must be a maximum. Following a similar logic for $x = 0$, then $x = 0$ must be a minimum. There are also a difference between a global and a local maximum/minimum but those I'll leave the reader to figure out.

5.2 Second Derivative Test

This test checks for concavity. Let's first find the second derivative:

$$f''(x) = 6x + 2$$

Making that equal 0:

$$\begin{aligned} f''(x) &= 0 \\ 6x + 2 &= 0 \\ x &= \frac{-1}{3} \end{aligned}$$

A positive second derivative means that the derivative is increasing on that interval: this means that the rate of change of the rate of change of a function is increasing, AKA concave up. Similarly, a negative second derivative means concave down. Let's make a table again:

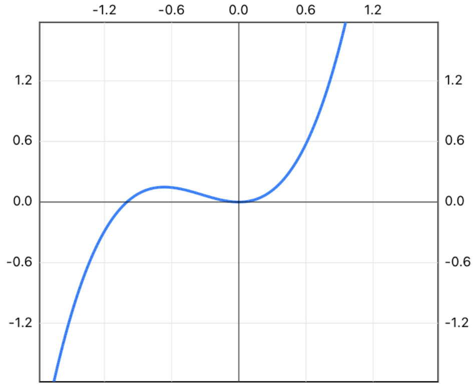
	$(-\infty, \frac{-1}{3})$	$(\frac{-1}{3}, \infty)$
$f(x)$		
$f''(x)$		

Using the same method as the first derivative test, let's fill in the table:

	$(-\infty, \frac{-1}{3})$	$(\frac{-1}{3}, \infty)$
$f(x)$	concave down	concave up
$f''(x)$	negative	positive

Since the left hand side of $x = \frac{-1}{3}$ is concave down but the right hand side is concave up, that means there's a point of inflection at $x = \frac{-1}{3}$, a point where the concavity changes.

Now that we have an idea of what the graph looks like, we can draw it and see that we are indeed correct.



6 Series

6.1 Compound interest

If we have a deposit P_0 and we put it into a bank account with an annual interest rate of r, then the balance after t years will be

$$P(t) = P_0 \left(1 + \frac{r}{n}\right)^{nt}$$

where n represents the number of times the deposit is compounded a year. If a deposit is compounded quarterly, that means it's compounded 4 time a year, so $n = 4$. Monthly means 12 times, daily means 365 times, etc. The more times we compound, the greater n gets until it reaches infinity. This can be modelled with the following limit:

$$P(t) = P_0 \lim_{n \to \infty} \left(1 + \frac{r}{n}\right)^{nt}$$

When n approaches infinity, we call this **continuous compounding**. Then the latter half of the formula, $\lim_{n \to \infty} \left(1 + \frac{r}{n}\right)^{nt}$ can be rewritten as e^{rt}, so for formula for **continuous compounding** is

$$P(t) = P_0 e^{rt}$$

6.2 Sequences of payments

Suppose that you have an account and you put 100 into it every at an interest rate of 7% per year for $n - 1$ years (meaning n number of payments). After your first year, you have 100(1.07) dollars.

After 2 years,

$$M = 100(1.07) + 100(1.07)^2$$

After the third year, you have

$$M = 100(1.07) + 100(1.07)^2 + 100(1.07)^3$$

and so on until you reach

$$M = 100(1.07) + \cdots + 100(1.07)^{n-1}$$

when you receive the nth payment. A way that we can represent this is using

$$M = \sum_{i=0}^{n-1} 100(1.07)^i$$

So the general formula for calculating the sum of a sequence of payments is

$$M = \sum_{i=0}^{n-1} P_0(1 + \frac{r}{n})^{ni}$$

or

$$M = \sum_{i=0}^{n-1} P_0 e^{ni}$$

Depending on how many years and the time at which you are calculating the amount of money, you may use $i = 1$ and/or n.

Similarly, to calculate the amount of money needed today to fund a sequence of payments that will go on for n years, we use the formula

$$M = \sum_{i=0}^{n-1} P_0(1 + \frac{r}{n})^{-ni}$$

or

$$M = \sum_{i=0}^{n-1} P_0 e^{-ni}$$

To calculate the present value of an infinite sequence, we can use the formula

$$M = \sum_{i=0}^{\infty} P_0 e^{-ni}$$

6.2.1 Computing infinite series

A geometric series is a series where each successive term is a multiple of the previous, so like

$$1 + 2 + 4 + 8$$

because $2 = 1(2)^1$, $4 = 2(2)^1 = 1(2)^2$ and so on. We can rewrite the above series as

$$\sum_{i=0}^{3} 2^i$$

The formula for calculating a sum of a geometric sequence is

$$S = \frac{a(1-r^n)}{1-r}$$

where each term of the sequence is computed as

$$t_n = ar^{n-1}$$

where a is the first term and r is the multiplying factor. For an infinite sum, our formula can be computed using

$$S_\infty = \lim_{n\to\infty} \frac{a(1-r^n)}{1-r} = \frac{a}{1-r}$$

which is only possible if $|r| < 1$ (we would have a diverging series otherwise).

7 Integration

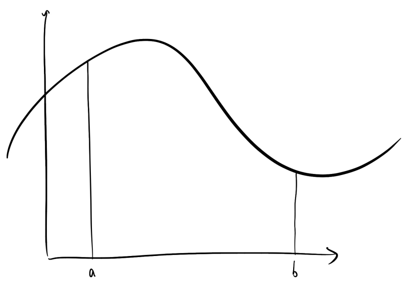
Integrals are just the opposites of derivatives (antiderivative). Think of them as the area under a curve. The notation for an definite integral, an integral with a value, is notated as

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

First, let's see how we can approximate them.

7.1 Left and Right hand sums

Given a function $f(x)$, the area under the curve between a and b can be approximated. If we break



the function to n strips, then the length of each strip is given by

$$\frac{b-a}{n}$$

The length of each strip would be

$$f(x_i) \text{ where } x_i = a + \frac{b-a}{n}i$$

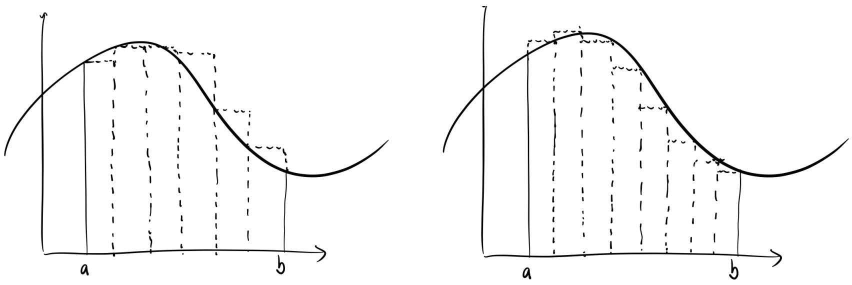
Then if we use a left hand sum, we use the formula

$$\int_a^b f(x)dx = \sum_{i=0}^{n-1} f(x_i)\frac{b-a}{n}$$

For the right hand, we use

$$\int_a^b f(x)dx = \sum_{i=1}^{n} f(x_i)\frac{b-a}{n}$$

The definite integral refers to the total change on a function between points a and b. Obviously,



there are much better ways of calculating this (Riemann sums are the worst and let's be honest, even a 5th grader can do them) but we have to first learn how to **integrate**. Also, on a side note, the area **above** the x-axis is positive and the area **below** the x-axis is negative.

7.2 Fundamental Theorem of Calculus

Before we move onto methods of integration, let me introduce a quick theorem.

Theorem (Fundamental Theorem of Calculus)**.** Let $f(x)$ be a function that is continuous on the interval $[a, b]$. Then the definite integral from a to b can be computed by

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

where $F(x)$ is the integral of $f(x)$.

This theorem opens up many new avenues for us, including calculating the average value of a function on the interval $[a, b]$, which is

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

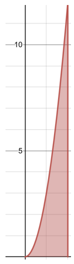
The average value of a function is the area under the graph between two points divided by the distance between the two points. You can probably see now that this value can also be easily estimated. The definite integral can also be used to tell us the value of $f(x)$ at a certain point by finding the area between 0 and the point on $f'(x)$. Let's do an example.

Example 8. Given $f(x) = x^3$, find the value of $f(4)$.

Solution. Since we are finding the point $f(2)$, then it makes sense for us to be integrating the derivative of $f(x)$, as the integral is the opposite of a derivative.

$$f'(x) = 3x^2$$

Let's graph this function.



We can see that this area is above the x-axis, so $f(2)$ must be positive. Let's do a Right Hand sum with 16 subdivisions.

$$\int_0^2 x^3 dx = \sum_{i=1}^{1} 6(3x_i))^2 \frac{2-0}{16}$$

Computing, this is equal to 8.765625. Looking back at the graph, using a Right Hand sum would have given us an overestimate, and it did because $f(2)$ is actually equal to 8. ■

8 Methods of integration

8.1 Integration basics

Taking integrals is a bit tricky, but nothing that isn't doable. Let's go over some basic integration rules.

8.1.1 Constant/coefficent rules

Given a function $f(x) = a$ where $a \in \mathbb{R}$,

$$\int a dx = ax + C$$

Given a function $g(x) = af(x)$ where $a \in \mathbb{R}$

$$\int g(x)dx = a \int f(x)dx + C$$

8.1.2 Power rule

$$\int x^n dx = \frac{1}{n+1}x^{n+1} + C$$

8.1.3 Integrals of trigonometric functions

$$\int \sin(x)dx = -\cos(x) + C, \int \cos(x)dx = \sin(x) + C$$

8.1.4 Integral of e^x

$$\int e^x dx = e^x + C$$

But wait. Why did I add $+C$ to every integral? Let's look at an example.

Example 9. Integrate $f(x) = 3x^2$

Solution.

$$\int 3x^2 dx = 3\frac{1}{2+1}x^{2+1} = x^3$$

■

Nothing wrong here right? But what if I take the derivative of $x^3 + 1$. That equals...$3x^2$. What if I take $x^3 + 4331232331$. That also equals...$3x^2$. Recall from differentiation that the derivative of any constant is 0. That means that infinitely many functions have the same derivative, so consequently, every function has infinitely many antiderivatives. That's why we use the $+C$ to denote an arbitrary constant.

$$\int 3x^2 dx = x^3 + C$$

NEVER EVER FORGET +C. YOUR PROFS WILL TAKE OFF 1 MARK IF YOU FORGET IT AND FOR GOOD REASON TOO BECAUSE YOUR INTEGRAL WILL BE TECHNICALLY WRONG

8.2 Integration by Substitution

Much like with derivatives, there are also integration techniques that we can use to quickly find the integral of functions. In differential calculus, we encountered the chain rule, and the integral calculus version of that rule is called "u-substitution". By subbing a function with u, we can simplify a function for us for integrate. Let's first derive the formula, then I'll show an example.

The chain rule is as follows:

$$\frac{d}{dx}(f \circ g(x)) = f'(g(x)) \cdot g'(x)$$

Integrating both sides, we get

$$\int (f(g(x)))'dx = \int f'(g(x))g'(x)dx$$
$$f(g(x)) = \int f'(g(x))g'(x)dx$$

By definition,

$$f(u) = \int f'(u)du$$

so if we let $g(x) = u$, then $g'(x)dx = du$ and so the formula for u-substitution is derived.

Example 10. Integrate $\frac{x}{x^2+1}$

Solution. Let $u = x^2 + 1$ and $du = 2xdx$. Then,

$$\int \frac{x}{x^2+1}dx = \int \frac{1}{2}\frac{2x}{x^2+1}dx = \frac{1}{2}\int \frac{1}{u}du = \frac{1}{2}\ln|u| = \frac{1}{2}\ln|x^2+1| + C$$

Note that I did not add $+C$ to $\frac{1}{2}\ln|u|$, as that was not the final answer. ■

8.3 Integration by Parts (IBP)

Integration by Parts, or IBP, is the reversal of the product rule from differential calculus. That is, we are reversing

$$\frac{d}{dx}(fg) = f'g + g'f$$

Rewriting this rule, we obtain

$$f'g = \frac{d}{dx}fg - g'f$$
$$\int f'gdx = \int \frac{d}{dx}fgdx - \int g'fdx$$
$$\int f'gdx = fg - \int g'fdx$$

If for some reason you want to use a shorter (but much stranger and more confusing) notation, then let $f(x) = u, g(x) = v, du = f'(x)dx, dv = g'(x)dx$, and

$$\int udv = uv - \int vdu$$

I cannot stress how much I dislike this notation simply because it is so counterintuitive and does not explicitly show the parts of the IBP formula, which is why I highly, highly encourage using the $f(x)g(x)$ notation.

Example 11. Integrate $\int x^3cos(x)dx$

Solution. Let $f'(x) = \cos(x), g(x) = x^3$. Then $f(x) = \sin(x), g'(x) = 3x^2$.

$$\begin{aligned}\int x^3\cos(x)dx &= x^3 sin(x) - \int 3x^2\sin(x)dx \\ &= x^3\sin(x) - 3\left(-x^2\cos(x) + \int 2x\cos(x)dx\right) \\ &= x^3\sin(x) + 3x^2\cos(x) - 3\left(2x\sin(x) - 2\int\sin(x)dx\right) \\ &= x^3\sin(x) + 3x^2\cos(x) - 6x\sin(x) - 6\cos(x) + C\end{aligned}$$

■

Example 12. Integrate $\int e^x\sin(x)dx$

Solution.

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

■

9 Continuous income streams

Let's define a function $S(t)$ as the rate of which income is streaming into a bank account. Additionally, let r represent the interest rate compound **continuously** for M years. Then

$$PV = \int_0^M S(t)e^{-rt}dt$$

and

$$FV = PV\cdot e^{rM} = \int_0^M S(t)e^{r(M-t)}dt$$

Recall that if we put an amount of money in a bank account and let it compound continuously for t years, the amount of money we'll have after M years is

$$P = P_0e^{rM}$$

Notice how $FV = PVe^{rM}$. FV is how much money we will have after M years while PV represents how much of an initial investment we need in order to have FV amount of money after M years.

10 Probability and statistics

10.1 Histograms

Histograms are a way of representing data graphically by splitting data into bins. A histogram graph shows the size of the bins (which should be equal) and the height of the bins.

If we want to compare different data sets, it's really helpful to divide the height of each bin by the total number of elements. In other words, we're comparing the proportion of data in each bin.

A **probability density function** (PDF), $p(x)$, represents the probability for each point in the data set. The area under $p(x)$ between a and b ($\int_a^b p(x)$) is the probability that a randomly chosen point in the data set is between a and b.

PDF must satisfy 2 conditions:

1. $p(x) \geq 0$ for every x
2. $\int_{-\infty}^{\infty} p(x)dx = 1$

This is because we a) can not have negative probability and b) the maximum probability of an event happening is 1, or 100%.

Example 13. Determine whether or not the following is a PDF.

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

Solution.

$$\begin{aligned} &\int_0^{\infty} 2e^{-2x} dx \\ =& \lim_{t\to\infty} \int_0^t 2e^{-2x} dx \end{aligned}$$

Using u-substitution, let $u = 2x, du = 2dx$. Then $x = 0 \Rightarrow u = 0, x = t \Rightarrow u = 2t$.

$$\begin{aligned} &\lim_{t\to\infty} \int_0^{2t} e^{-u} du \\ =& \lim_{t\to\infty} - e^{-u}\Big|_0^{2t} \\ =& \lim_{t\to\infty} \left(-e^{-2t} + e^{-0}\right) \\ =& \lim_{t\to\infty} \left(\frac{-1}{e^{2t}} + 1\right) \\ =& 1 \end{aligned}$$

showing that $p(x)$ is indeed a PDF. ■

10.2 Cumulative Distributive Functions

The cumulative distributive function, or CDF, is defined as the function showing the total probability of an event occurring at or before a certain data point. For example, pretend we have a CDF for a data set of people's salaries. At $200000, the corresponding value on the CDF is the total probability of someone making $200000 or less. In other words, it shows the total (cumulative) amount of people who's salaries are less than or equal to $200000.

Mathematically, the CDF is the integral of the PDF.

$$P(t) = \int_{-\infty}^{t} p(x)dx$$

10.3 Median

In elementary school, we learned that within a data set, the median is the middle-most value. In probability and calculus, the median is defined as being the point where the definite integral from the left boundary to that point is equal to 0.5, or 50%. In other words, if we let t represent the median,

$$0.5 = \int_{-\infty}^{t} p(x)dx = P(t)$$

10.4 Mean

Similarly to median, we were also introduced to mean in elementary school. We learned that the mean is the sum of all the values divided by the number of values. Let $E(x)$ denote the mean, or expected value. Using a right Riemann sum,

$$E(x) = \lim_{n\to\infty} \sum_{i=1}^{n} x_i p(x_i)\Delta x = \int_{-\infty}^{\infty} xp(x)dx$$

10.5 Normal distribution

Normal distribution is a special density function of the form

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

where μ is the mean and σ is the standard deviation.

11 Multivariable functions

11.1 Cross-sections

Assume we have a multivariable function $f(x, y)$. If we set one variable to be constant (equal to a number) and the other to be a variable, we can analyze our function in 2d. This is useful because we can more easily isolate a variable and analyze the function with respects to that variable.
Cross sections also help with graphing. By holding a variable constant, we can see the graph from different POVs and then combine them to form the multivariable graph.

11.1.1 Contours

Let's fix our multivariable function to a variable z, so $z = f(x, y)$. If we set z to be constant (think of a top view), we can see the curves for the xy-plane. Basically, using contours, we can view the corresponding z value for an (x, y) coordinate.

12 Multivariable differential calculus

12.1 Partial differentiation

Solving for partial derivatives involves treating one variable as a constant and differentiating the other. First, let's define a partial derivative. The partial derivative of $z = f(x, y)$ with respects to x, denoted f_x is

$$f_x = \lim_{h \to 0} \frac{f(x+h, y) - f(x, y)}{h} = \frac{\partial z}{\partial x}$$

Similarly, the partial derivative with respects to y, denoted f_y is

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

f_x and f_y can also be interpreted as the rate of change of z in the x and y direction respectively.

Example 14. Differentiate with respects to x and y: $z = f(x, y) = xy^2 + y^3 + ln(xy)$

Solution. Recall that the derivative of a constant is 0 and $(ax)'$ would be equal to a. Then

$$\begin{aligned} \frac{\partial z}{\partial x} = f_x &= y^2 + 0 + \frac{1}{xy} \cdot y \\ &= y^2 + \frac{1}{x} \end{aligned}$$

and

$$\frac{\partial z}{\partial y} = f_y = 2xy + 3y^2 + \frac{1}{xy} \cdot x$$
$$= 2xy + 3y^2 + \frac{1}{y}$$

■

12.2 Higher order partial derivatives

Much like regular derivatives, higher order partial derivatives also exist. Unfortunately, the notation for higher order derivatives isn't as clear. Let's first focus on one variable. The higher order derivatives with respect to x and y are

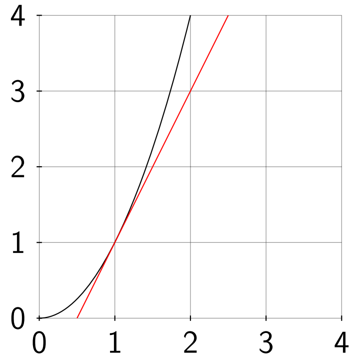
$$\frac{\partial^2 z}{\partial x^2} = f_{xx}$$
$$\frac{\partial^2 y}{\partial y^2} = f_{yy}$$

This was pretty simple but it gets a little more complex if we first derive with respects to x, then y, or vice versa. Observe:

$$f_{xy} = \frac{\partial z}{\partial y \partial x}$$
$$f_{yx} = \frac{\partial z}{\partial x \partial y}$$

12.3 Approximating a multivariable function

Previously, we encountered approximating a univariable function (y = f(x)). By using the slope of tangent lines, we could approximate a certain value on the function. Let's approximate $f(1.5)$ if $f(x) = x^2$



If we wanted to approximate the value of $f(1.5)$, we can use the information we now have. The slope of the tangent line at $f(1)$ is 2 and $\Delta x = 1.5 - 1 = 0.5$. Lastly, $f(1) = 1$. We now have all the information needed to approximate.

$$\begin{aligned} f(x+\Delta x) &= f(x) + f'(x)\Delta x \\ f(1.5) &= f(1) + f'(1)(0.5) \\ &= 1 + 2(0.5) = 2 \end{aligned}$$

$1.5^2 = 2.25$, so we can see that our approximation is pretty close.

Applying that to multivariable functions, we can use the partial derivatives and use a similar formula.

$$f(x+\Delta x, y+\Delta y) = f_x(x,y)\Delta x + f_y(x,y)\Delta y + f(x,y)$$

Depending on what we're approximating, this formula uses either a tangential line or plane.

By definition, a function is **differentiable** if we can zoom in on it's graph and eventually see that the function's contour diagram becomes equally spaced lines. Furthermore, in univariable calculus, we saw that for $y = f(x)$, $\Delta y = f'(x)\Delta y$. For multivariable calculus, we can apply the same to $z = f(x,y)$.

$$\Delta z = f_x \Delta x + f_y \Delta y$$

$$\partial z = f_x \partial x + f_y \partial y$$

This is called the **differential**.

Example 15. [2]Given the Cobb-Douglas $Q = \beta K^{\alpha} L^{1-\alpha}$ where $0 < \alpha < 1$, show that the function satisfies

$$K\frac{\partial Q}{\partial K} + L\frac{\partial Q}{\partial L} = Q$$

Solution.

$$\begin{aligned} \frac{\partial Q}{\partial K} &= \beta\alpha K^{\alpha-1} L^{1-\alpha} \\ \frac{\partial Q}{\partial L} &= \beta(1-\alpha) K^{\alpha} L^{-\alpha} \end{aligned}$$

$$\begin{aligned} K\frac{\partial Q}{\partial K} + L\frac{\partial Q}{\partial L} &= K(\beta\alpha K^{\alpha-1}L^{1-\alpha}) + L(\beta(1-\alpha)K^{\alpha}L^{-\alpha}) \\ &= (K^{\alpha}L^{1-\alpha})(\beta\alpha + (1-\alpha)\beta) \\ &= K^{\alpha}L^{1-\alpha}(\beta\alpha + \beta - \beta\alpha) \\ &= \beta K^{\alpha}L^{1-\alpha} \\ &= Q \end{aligned}$$

[2]This example was labelled "Do on your own" in the slides

■

12.4 Finding the equation to a tangential plane

You may remember from univariable calculus that there's a formula to follow for finding the equation of a tangential line to a curve.

Let $y = f(x)$. Then the equation of the tangent line at point (a, b) is

$$y - b = \frac{dy}{dx}|_a(x - a)$$

There's a similar formula for multivariable calculus.

$$z - f(a, b) = \frac{\partial z}{\partial x}|_{(a,b)}(x - a) + \frac{\partial z}{\partial y}|_{(a,b)}(y - b)$$

Or to make it slightly easier on your eyes:

$$z - f(a, b) = f_x(a, b)(x - a) + f_y(a, b)(y - b)$$

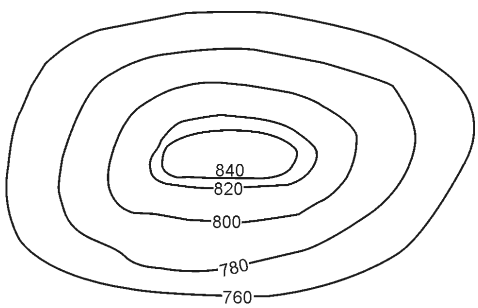
13 Vector calculus

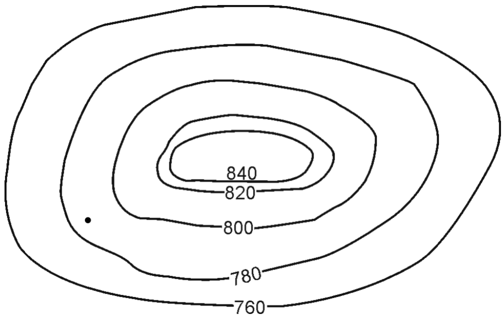
13.1 Directional derivatives

Now that we know how to interpret and take partial derivatives, let's move onto something that is deceptively difficult but is actually quite easy to understand. Let's picture a contour graph. Next, let's pick any point. Finally, let's pretend we're standing on the point and facing a certain direction. The **directional derivative** is the rate of change at that point in that direction. Sounds complicated? Let's work through some examples.

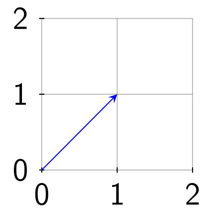
Example 16.

Here is a contour graph taken from the internet. Next, let's pick a random point on the graph.

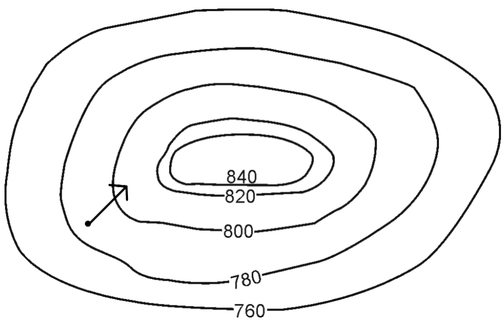




Now let's pick a direction. What does this mean? Well...it's easier just to show you. Let's pick a vector, say, (1,1).



We can see that the vector $(1, 1)$ points exactly 45° counterclockwise from the x-axis. We have now picked a direction. Let's apply this to our point that we picked. Now, to compute the directional



derivative, we compute the rate of change in the same direction as vector $(1, 1)$ from the point that we selected on our contour graph. I like to imagine that we are on a hill and are standing still at a random point on the hill. The direction is whichever way I am facing and I will be computing the rate of change in that exact direction in which I am facing. Does this make a bit more sense?

You may still be a bit confused, so I encourage you to go back to this example and work through the intuition. A vector is used because it has direction. Notice how we stayed in $\mathbb{R}^2$ when choosing the vector. That's because when we work with contours, we are looking at a function from a top view, so we are looking at the xy-plane. This just means that there's no need to over-complicate things with a vector in $\mathbb{R}^3$ and beyond!

13.1.1 Gradient vector

At any given point on a function $f(x, y)$, the **gradient vector** is the vector that points in the direction of the greatest rate of increase at that given point. For example, if we have a random function $f(x, y)$, the gradient vector at $(2, 5)$ might be equal to $(2, 4)$ because the greatest rate of change at $(2, 5)$ is in that direction. The gradient vector at any point (a, b) is denoted

$$\nabla f(a, b)$$

and the formula for calculating the vector is

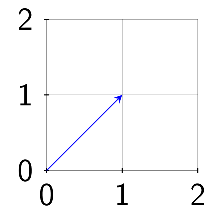
$$\nabla f(a, b) = \begin{bmatrix} \frac{\partial f}{\partial x}(a, b) \\ \frac{\partial f}{\partial y}(a, b) \end{bmatrix}$$

Basically, the x component of the gradient is the rate of change in the x-direction at (a, b) and the y component of the gradient is the rate of change in the y-direction at (a, b).

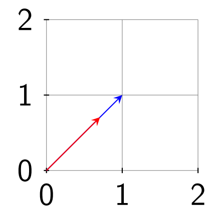
From the example earlier, the vector $(1, 1)$ is what's called a **direction vector** but here's the thing. We want to use a unit vector as a **direction vector** to eliminate the possibility of calculating a faulty derivative (can you imagine looking at a hill and thinking the rate of change in elevation is 5m/unit when it's actually 0.5?). Problem is, is $(1, 1)$ a unit vector though? Let's test it.

$$\left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\| = \sqrt{1^2 + 1^2} = \sqrt{2} \neq 1$$

Ok so we clearly have a problem here. Let's take a look at the vector once again.



If want the vector to be a unit vector, then it must be in the same direction as $(1, 1)$ but must also have a magnitude of 1. Now think to yourself: could shrinking the vector work? The answer is yes. Let's take a look.



Let's define a vector $\tilde{u}$ where $\tilde{u}$ is the unit vector in the direction of $(1, 1)$. Then

$$\tilde{u} = t \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1t \\ 1t \end{bmatrix}$$

where t is some scalar in $\mathbb{R}$. Then

$$\begin{aligned} \|\tilde{u}\| &= \left\| \begin{bmatrix} t \\ t \end{bmatrix} \right\| \\ 1 &= \sqrt{t^2 + t^2} \\ 1 &= \sqrt{2t^2} \\ t &= \frac{1}{\sqrt{2}} \end{aligned}$$

Solving for $\tilde{u}$ with our new value of t:

$$\begin{aligned} \tilde{u} &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \end{aligned}$$

If you really want to test it, you can try calculating the magnitude of $\tilde{u}$ and see that it equals 1. We've done it! Now we have a unit vector pointing in the same direction as (1,1). Going back to our definition of a directional derivative, a directional derivative is the instantaneous rate of change in $f(x, y)$ at the point (a, b) in the direction of $\tilde{u}$ and is denoted $f_{\tilde{u}}(a, b)$. Side note: the x component of $\tilde{u}$ is u_x and the y component is u_y

$$f_{\tilde{u}}(a, b) = \lim_{h \to 0} \frac{f(a + u_x, b + u_y) - f(a, b)}{h}$$

Thankfully, we don't have to use this bulky and gross formula. Instead, let's use a dot product.

$$f_{\tilde{u}}(a, b) = \tilde{u} \cdot \nabla f(a, b)$$

The directional derivative is equal to the dot product between the unit vector whose direction we are pointing in and the gradient. Let's do an example.

Example 17. Consider the function $f(x, y) = x^2 + y^2 - 4x + 6y$. Find the directional derivative in the direction of (3,1) at the point (2,1).

Solution. First, let's verify that our vector is indeed a unit vector.

$$\begin{aligned} \begin{bmatrix} 3 \\ 1 \end{bmatrix} &= \sqrt{3^2 + 1^2} \\ &= \sqrt{10} \neq 1 \end{aligned}$$

Clearly, this is not a unit vector, so let's find one.

Let $\tilde{u}$ be the unit vector in the direction (3,1). Then

$$\begin{aligned} \tilde{u} &= t \begin{bmatrix} 3 \\ 1 \end{bmatrix} \\ \|\tilde{u}\| &= \left\| \left[\begin{bmatrix} 3t \\ t \end{bmatrix} \right] \right\| \\ 1 &= \sqrt{9t^2 + t^2} \\ 1 &= \sqrt{10t^2} \\ t &= \frac{1}{\sqrt{10}} \\ \tilde{u} &= \begin{bmatrix} \frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{bmatrix} \end{aligned}$$

Next, let's find a gradient.

$$\begin{aligned} \nabla f(2,1) &= \begin{bmatrix} f_x(2,1) \\ f_y(2,1) \end{bmatrix} \\ &= \begin{bmatrix} 2x - 4|_{(2,1)} \\ 2y + 6|_{(2,1)} \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ 8 \end{bmatrix} \end{aligned}$$

Finally, let's use the formula.

$$\begin{aligned} f_{\tilde{u}}(2,1) &= \tilde{u} \cdot \nabla f(2,1) \\ &= \begin{bmatrix} \frac{3}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 8 \end{bmatrix} \\ &= \left(\frac{3}{\sqrt{10}}\right)(0) + \left(\frac{1}{\sqrt{10}}\right)(8) \\ &= \frac{8}{\sqrt{10}} \end{aligned}$$

■

14 Multivariable optimization

14.1 Determining critical points from contour graph

This is really easy, just observe the graph and you will be able to tell where the local minimums, maximums are. Pay attention to the values going in the x and y directions as they will help with finding the saddle points. A saddle point is a point where it is a local maximum in the x direction but a local minimum in the y direction or vice versa.

14.2 Optimization of multivariable function

Let's say we have a function $f(x, y) = x^2 + y^2 + 3x + 4y$ and now we want to find the critical points as well as define those points as local minimums, maximums, or saddle points. We follow a similar approach to univariable functions. Let's find some partial derivatives.

$$\frac{\partial f}{\partial x} = 2x + 3$$
$$\frac{\partial f}{\partial y} = 2y + 4$$

Now to find the critical points, we need to find points (x, y) such that the rate of change at those points equal 0. In other words, we need to find points where f_x and f_y both equal 0.

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

$$2y + 4 = 0 \Rightarrow y = -2$$

And thus we have determined that $f(x, y)$ has 1 critical point at $\left(\frac{-3}{2}, 2\right)$.

Another way of notating a critical point of $f(x, y)$ is to use the gradient. Recall that the gradient represents the maximum rate of change at a point so if we find points where the gradient equals 0 then we have found our critical points.

$$\begin{aligned}\nabla(x,y) &= \overrightarrow{0} \\ &= \begin{bmatrix} \frac{\partial f}{\partial x} = 0 \\ \frac{\partial f}{\partial y} = 0 \end{bmatrix}\end{aligned}$$

After we have our critical points, it's time to determine if they are local maximums, minimums, or saddle points. Luckily, there's a theorem for that. Suppose that we have found our critical points. Then define

$$\mathcal{D} = f_{xx}(a,b)f_{yy}(a,b) - (f_{xy}(a,b))^2$$

$$\text{If } \begin{cases} \mathcal{D} > 0, f_{xx}(a,b) > 0 \Rightarrow (a,b) \text{ is a local minimum} \\ \mathcal{D} > 0, f_{xx}(a,b) < 0 \Rightarrow (a,b) \text{ is a local maximum} \\ \mathcal{D} < 0 \Rightarrow (a,b) \text{ is a saddle point} \\ \mathcal{D} = 0 \Rightarrow \text{ inconclusive} \end{cases}$$

Going back to our example, let's compute the second order derivatives needed.

$$f_{xx} = 2, f_{yy} = 2, f_{xy} = 0$$

$$\begin{aligned}\mathcal{D} &= (2)(2) - (0)^2 \\ &= 4 > 0\end{aligned}$$

Since f_{xx} is also greater than 0, then our critical point $\left(\frac{-3}{2}, 2\right)$ is a local minimum.

This seems pretty simple, but it can get harder. Let's do another example where f_x and f_y aren't only in terms of x and y respectively.

Example 18. Find and classify all critical points of $f(x,y) = x^2 + 3y^2 + 4xy + 5y + 6x$

Solution. Let's first compute the first-order partial derivatives.

$$\begin{aligned}f_x &= 2x + 4y + 6 \\ f_y &= 6y + 4x + 5\end{aligned}$$

Now we make f_x and f_y equal 0 and obtain a system of equations for us to solve.

$$\begin{cases} 2x + 4y + 6 = 0 \\ 4x + 6y + 5 = 0 \end{cases}$$
$$= \begin{cases} 2x + 4y = -6 \\ 4x + 6y = -5 \end{cases}$$

Referring to what we learned in the first few weeks, we can solve this using row-reduction[3].

$$\left[\begin{array}{cc|c} 2 & 4 & -6 \\ 4 & 6 & -5 \end{array}\right] \xrightarrow{R_2-2R_1} \left[\begin{array}{cc|c} 2 & 4 & -6 \\ 0 & -2 & 7 \end{array}\right] \xrightarrow{\frac{1}{2}R_1} \left[\begin{array}{cc|c} 1 & 2 & -3 \\ 0 & -2 & 7 \end{array}\right] \xrightarrow{R_1+R_2} \left[\begin{array}{cc|c} 1 & 0 & 4 \\ 0 & -2 & 7 \end{array}\right] \xrightarrow{\frac{1}{2}R_2} \left[\begin{array}{cc|c} 1 & 0 & 4 \\ 0 & 1 & \frac{-7}{2} \end{array}\right]$$

We have now obtained a critical point: $\left(4, \frac{-7}{2}\right)$.

$$f_{xx} = 2, f_{yy} = 6, f_{xy} = 4$$

$$\mathcal{D} = (2)(6) - (4)^2 = 12 - 16 = -4 < 0$$

Since $\mathcal{D}$ is less than 0, then our critical point is a saddle point. ■

14.3 Constrained optimization

14.3.1 Lagrange multipliers

Now we constrain $f(x, y)$ within an equation $g(x, y) = c$ where $c \in \mathbb{R}$. We first start by setting up a system of equations:

$$\begin{cases} \nabla f = \lambda \nabla_g \\ g(x, y) = c \end{cases}$$

Solving for ∇f and ∇g, we get

$$\begin{cases} f_x = \lambda g_x \\ f_y = \lambda g_y \\ g(x, y) = c \end{cases}$$

How do we know if the points we find are maximums or minimums? Compare them to the endpoints of the constraint (if there exists any) and to any points where $\nabla g = 0$ (ie: direction where rate of change of g is equal to 0). We do this because the maximum/minimum(s) will occur at one of the 3 points mentioned above. In summary, max/min occur at

1. Solution to above system of equations

[3] If you don't remember row-reduction, you can also just straight up solve this system using substitution.

2. Endpoints of the constraint

3. Point(s) where $\nabla g = 0$

Let's do an example.

Example 19. This example is taken from the class slides.
An ant walks on a flat stove. The temperature in °C at a location $x$cm east and $y$cm north from the southwest corner of the stove is approximately

$$T(x,y) = 3y + 2x + 30$$

If she walks along the path $x^2 + 2y^2 = 34$, with $x \geq 0$ and $y \geq 0$, what are the warmest and coldest temperatures she experiences?

Solution. Let's set up the system of equations:

$$\begin{cases} T_x = \lambda y_x \\ T_x = \lambda g_y \\ g(x,y) = 34 \end{cases} \longrightarrow \begin{cases} 2 = \lambda 2x \\ 3 = \lambda 4y \\ x^2 + 2y^2 = 34 \end{cases}$$

Solving, we get

$$\frac{2}{3} = \frac{\lambda 2x}{\lambda 4x}$$
$$\frac{2}{3} = \frac{x}{2y}$$
$$x = \frac{4}{3}y$$

Plugging that into our third equation, we can solve for x and y.

$$\left(\frac{4}{3}y\right)^2 + 2y^2 = 34$$
$$\frac{34}{9}y^2 = 34$$
$$y = \pm 3 \rightarrow y \geq 0 \text{ so } y = 3$$
$$x = 4$$

For $x = 4, y = 3$, we have $T(4,3) = 47$°C. Let's compare this point to the endpoints and $\nabla g = 0$. First off, the end points are $g(x,0)$ or $g(0,y)$, so let's solve for those.

$$g(x,0) = 34 \rightarrow x^2 = 34 \rightarrow x = \sqrt{34}$$
$$g(0,y) = 34 \rightarrow 2y^2 = 34 \rightarrow y = \sqrt{17}$$

Plugging these points into $T(x, y)$, we get

$$T(\sqrt{34}, 0) = 42.4°\text{C}$$
$$T(0, \sqrt{17}) = 41.78°\text{C}$$

Let's calculate the gradient for $g(x, y)$.

$$\nabla g(x, y) = \begin{bmatrix} 2x \\ 4y \end{bmatrix} \rightarrow \begin{bmatrix} 2x = 0 \\ 4y = 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$T(0, 0) = 34°\text{C}$, so clearly our point $T(3, 4)$ is a **maximum.** ■

You may have noticed, what happened to λ? You see, even though λ was cancelled out, it's not actually useless because it represents something pretty important. If the constraint value c increases by 1 point, then the optimal value of f will increase by λ. This means that λ represents the rate of change of the optimal value of f WRT c.

15 Multivariable integral calculus

Multivariable integral calculus is interesting because we are integrating a function over a region (think of this as finding the volume of a function).

15.1 Multivariable Riemann Sums

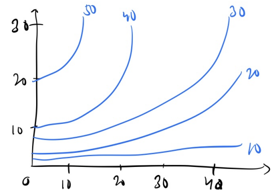
Let's say we have a function $f(x, y)$. Recall from single variable definite integrals, we found the area between two bounds. Since we are introducing 3 variables (ie: $\mathbb{R}^3$ space), then we have to adapt that definition. For a region $\mathcal{R} = [a, b] \times [c, d]$, we are finding the **volume** of $f(x, y)$ within that region. This is notated by

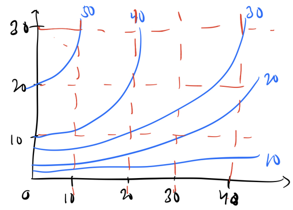
$$\int_{\mathcal{R}} f dA$$

where $\mathcal{R}$ is a region in the $xy-$plane (we're looking at contours again). If we break this region into small squares, we can approximate the volume of $f(x, y)$ in region $\mathcal{R}$. Much like with single variable calculus, if we add up the volumes of all the rectangles, we can approximate the definite integral over region $\mathcal{R}$. Let's show this with an example. Suppose we have a function over a region $[0, 40] \times [0, 30]$.

Let's make some rectangles.

Now we approximate the value of $f(x, y)$ for each region. The volume of each little square is approximately equal to $10 \times 10 \times f(x, y)$. Let's use the top right corner of each box for this





approximation. Adding them all up we get

$$53 \cdot 10 \cdot 10 + 50 \cdot 10 \cdot 10 + \cdots + 21 \cdot 10 \cdot 10) = 10^2(53 + \cdots + 21) = 470$$

So we know that

$$\int_{\mathcal{R}} f(x, y) dA \approx 470$$

Generalized, the approximation of the definite integral of a function $f(x, y)$ over the region $\mathcal{R}$ is defined by

$$\int_{\mathcal{R}} f(x, y) dA = \lim_{\Delta x \to 0, \Delta y \to 0} \sum_{i,j} f(u_{ij}, v_{ij}) \Delta x \Delta y$$

What this equation represents is that we're picking a point in each small rectangle of subdivision and computing the volume of each subdivision, then adding them up. The limit at the beginning of the equation represents us splitting $f(x, y)$ into infinitesimally small subdivisions (as we saw earlier with single variable integrals).

15.1.1 Average value

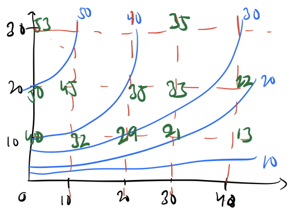
Recall from single variable calculus that average value of a function on the interval $[a, b]$ can be calculated as

$$Avg = \frac{\int_a^b f(x) dx}{b - a}$$

With multivariables, we are dividing the integral by the area of $\mathcal{R}$, so

$$Avg = \frac{\int_{\mathcal{R}} f(x, y) dA}{(b - a)(d - c)}$$

if $\mathcal{R} = [a, b] \times [c, d]$.



15.2 Iterated integrals

We can actually compute multivariable integrals by hand (called iterated integrals). Given a region $\mathcal{R} = [a, b] \times [c, d]$, we have

$$\int_R f(x, y)dA = \int_a^b \int_c^d f(x, y)dydx = \int_c^d \int_a^b f(x, y)dxdy$$

This type of integral is deceptively easy to calculate. First, we find the definite integral WRT to x or y, depending which of dx,dy comes first. Then we simply integrate the next integral. Recall that with multivariable calculus computations, we always hold one variable constant.

Example 20. This example is also from the class slides.

Compute the value of $\int_D xy^3 + x + 2dA$ where D is the region with $-1 \le y \le 0, 1 \le x \le 3$.

Solution. We know that $\mathcal{D} = [1, 3] \times [-1, 0]$. Then we can convert the integral to

$$\int_{-1}^0 \int_1^3 xy^3 + x + 2dxdy$$

Solving, we get

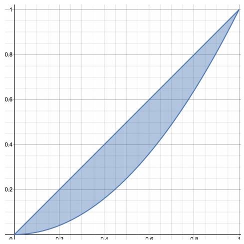
$$\begin{aligned}
\int_{-1}^0 \int_1^3 xy^3 + x + 2dxdy &= \int_{-1}^0 \left[\frac{x^2y^3}{2} + \frac{x^2}{2} + 2x\right]\Bigg|_1^3 dy \\
&= \int_{-1}^0 \left(\frac{9y^3}{2} + \frac{9}{2} + 6\right) - \left(\frac{y^3}{2} + \frac{1}{2} + 2\right) dy \\
&= \int_{-1}^0 4y^3 + 8dy \\
&= y^4 + 8y|_1^0 = (0 + 0) - (1 - 8) = 7
\end{aligned}$$

■

If we had solve with $dydx$ instead, we would have gotten the same answer (try it on your own!).

15.2.1 Iterated integral with non-rectangular domain

Let's say our region is between 2 curves. We can still find bounds. Let's say we want to integrate with $dxdy$ first. I'll show $dydx$ soon after. Looking from the dx perspective, the left hand function



is $y = x$ and the right hand is $y = x^2$. We have a problem. $dxdy$ means that we have to integrate with dx first, so we need our bounds have to be in terms of x. Then let's solve.

$$y = x \rightarrow x = y, y = x^2 \rightarrow x = \sqrt{y}$$

This means that the region from the x perspective is between the curves $x = y$ and $x = \sqrt{y}$. We have the first part of our integral.

$$\int_y^{\sqrt{y}} f(x, y)dx$$

Now we need to find the dy component. We've created bounds for the integral in the x direction. But if just put it between two functions, our volume will be infinite! That means we have to constrain our constraints to y values. From the graph, we clearly want to bound the constraints to be between $[0, 1]$ in the y-direction, so we have another constraint:

$$\int_0^1 f(x, y)dy$$

Combining these two integrals together

$$\int_0^1 \int_y^{\sqrt{y}} f(x, y)dxdy$$

Similarly, if we constrain y first, we get

$$\int_x^{x^2} f(x, y)dy$$

and then

$$\int_0^1 f(x, y)dx$$

so combining, we get

$$\int_0^1 \int_x^{x^2} f(x, y)dydx$$

16 Multivariable probability density functions

16.1 Joint PDF

This is going to be a short section. Joint PDFs for x and y if the probability that $a \leq\leq b$ and $c \leq y \leq d$ are given by

$$\int_a^b \int_c^d p(x, y) dx dy$$

Much like PDFs for single variable functions, joint PDFs share the same properties:

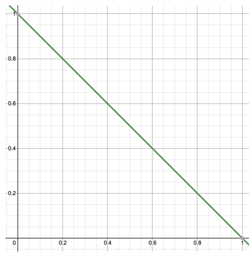
1. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(x, y) dx dy = 1$
2. The sign of $p(x, y)$ must be positive

Let's do an example from the worksheet.

Example 21. Pretend you have 2 chickens. One chicken lays good tasting eggs and the other chicken lays good smelling eggs. You have 1 dollars. What are the chances that you'll be able to afford investing in both chickens tomorrow if the PDF for the two chicken costs is

$$6e^{-2x-3y}$$

Solution. Our first step is to draw this function. Putting chicken 1 on the x-axis and chicken 2 on the y-axis, we get for all the possible combinations of x and y that equal 1. From this, we can



determine that our double integral is

$$\int_0^1 \int_0^{-x+1} 6e^{-2x-3y} dy dx$$

Solving, we get

$$\begin{aligned}
\int_0^1 \int_0^{-x+1} 6e^{-2x-3y} dydx &= 6 \int_0^1 \int_0^{-x+1} e^{-2x-3y} dydx \\
&= -\,2 \int_0^1 e^{-2x-3x} \Big|_0^{-x+1} dydx \\
&= -2 \int_0^1 e^{-2x+3x-3} - e^{-2x} dx \\
&= -2 \int_0^1 e^{x-3} - e^{-2x} dx \\
&= -2 \left(e^{x-3} + \frac{1}{2} e^{-2x} \Big|_0^1 \right) \\
&= -2(-0.346\ldots) = 69.36\%
\end{aligned}$$

■