Lecture Notes, Math Workshop 2019

Daniel Kent
kent.249@osu.edu

Contents

1 Lecture 1: Introduction, Notation, Definitions, and Basic Mathematics

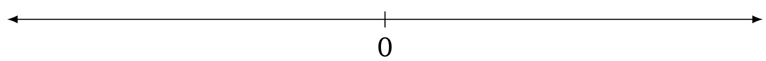
1.1 Course Overview

Although not all of the following material is drawn from Moore and Siegel (2013), the overlap is sizeable and intentional. Many thanks to the authors. Some of the examples are also drawn from Jason Morgan's lectures from when I took the same course at the start of graduate school. Many thanks also to Drew Rosenberg.

1.2 Variables and Constants

- **Variable**: A concept or a measure that takes different values in a given set.
  - E.g., GDP, polity score, party identification, etc.
- **Constant**: A concept or a measure that has a single value for a given set.

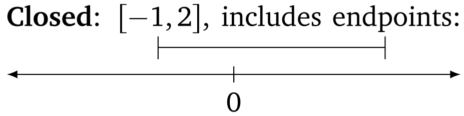
1.3 Number Types and Notation



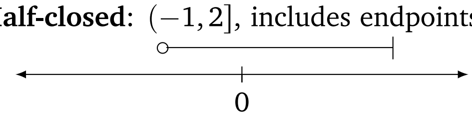
- **Real numbers**: ($\mathbb{R}$), can be placed anywhere on the line
- **Natural numbers**: ($\mathbb{N}$), positive with no decimal
- **Integer**: ($\mathbb{Z}$), non-decimal, both positive and negative
- **Rational number**: can be expressed as a ratio or fraction
- **Irrational number**: cannot be expressed as a fraction: π, e
- $\mathbb{R}^k$ is a k-dimensional space

1.4 Intervals

- **Open**: $(-1, 2)$, does not include endpoints, so greater than -1 and less than 2.
- **Closed**: $[-1, 2]$, includes endpoints: greater than or equal to -1 and less than or equal to 2.



- **Half-closed**: $(-1, 2]$, includes endpoints: greater than or equal to -1 and less than or equal to



2.

1.5 Levels of measurement

- **Nominal**: No mathematical relationship among the values
  - E.g., race, gender, country, party
- **Ordinal**: Ranking, but cannot do arithmetic because the distance between values is not equal.
  - E.g., K-12, age cohort, ideology (left-to-right)

- **Interval**: Fixed differences, but zero is arbitrary
  - E.g., temperature, date/time
- **Ratio**: Fixed differences with a true zero
  - E.g., age, length, income, votes

1.6 Sets

- A **set**, $S = \{\}$, is a collection of elements.
  - Order does not matter: $S = \{x, y, z\} = \{z, y, x\}$
    - $x \in S$, meaning x is an element of S
    - $a \notin S$, meaning a is not an element of S
- Other examples:
  - $\mathbb{Z} = \{..., -2, -1, 0, 1, 2, ...\}$ or $\{x|x \text{ is an integer}\}$
  - $T = \{x|8 \geqslant x \geqslant 7\}$ or $T = [7, 8]$
- Set notation:
  - $S = \{x, y, z\}$, $T = \{a, b, x\}$, $N = \{a, b\}$
    - Intersection: $T \cap S = \{x\}$
      - $N \cap S = \varnothing$: “empty set”
    - Union: $N \cup S = \{a, b, x, y, z\}$
    - Complement (not in the set): $(T \cap S)^c = \{a, b, y, z\}$
  - $\mathbb{U}$: universal set, all possible values
  - $A = \{x, y\}$
    - $A \subset S$: “A is a subset of S”
    - If $\not\subset$, then not a subset of. All subsets of A: $\{x\}$, $\{y\}$, $\{\varnothing\}$, $\{x, y\}$
- Transitivity:
  - If $Z \in Q$ and $Q \in R$, then $Z \in R$
- Disjoint:
  - No elements in common, more formally, two sets are disjoint if the intercept of sets is the null set: $N \cap S = \{\varnothing\}$

1.7 Independent and Dependent Variables

Let $y = f(x)$, where y is the outcome and x the input.

- **Independent variable**: the input – x
- **Dependent variable**: outcome – y

In a linear model, i.e. $y = \alpha + \beta x$, the dependent variable is y and the independent variable is x.

1.8 Functions

- $f(x) = y$
- **Constant function**: a function whose outcome is the same no matter the input. If $f(x) = c$, then no matter x the output is c.
- **Polynomial function**: $y = a + bx + cx^2$
- **Rational function**: can be defined by a fraction (ratio):

$$f(x, y) = \frac{a + bx^2}{1 + y}$$

1.9 Inequalities and Absolute Values

$$|x| = \sqrt{x^2} = \begin{cases} x & \text{if } x > 0 \\ -x & \text{if } x < 0 \\ 0 & \text{if } x = 0 \end{cases}$$

Useful properties:

- $|m| + |n| \geqslant |m + n|$
- $|m| \times |n| = |m \times n|$

1.10 Exponent Rules

- $x^n = x \cdot x \cdot x \cdot x \cdot \ldots \cdot x$ n times; $2^2 = 2 \cdot 2 = 4$
- $x^0 = 1$
- $x^m \cdot x^n = x^{m+n}$
- $\frac{x^m}{x^n} = x^{m-n}$
- $\frac{1}{x^m} = x^{-m}$
- $x^1 = x$
- $x^{\frac{1}{m}} = \sqrt[m]{x}$
- $(x^m)^n = x^{m \cdot n}$
- $x^m \cdot y^m = (xy)^m$

But:

- $x^m + y^n \neq (x + y)^{m+n}$

1.11 Commutative, Associative, and Distributive Laws

- **Associative property**: rewriting the parentheses does not change the outcome
  - $(a+b)+c = a+(b+c)$
  - $(a \cdot b) \cdot c = a \cdot (b \cdot c)$
- **Commutative property**: order of operands (inputs) does not change the outcome
  - $a+b = b+a$
  - $a \cdot b = b \cdot a$
- **Distributive property**:
  - $a(b+c) = ab+ac$

Also relevant:

- **Inverse property**:
  - For any x, there exists a $-x$ such that $-x+x=0$.
    - Formally[1]: $\exists(-x)$ s.t. $-x+x=0$
  - For any x, there exists a x^{-1} such that $x \cdot x^{-1} = 1$.
    - Formally: $\exists(x^{-1})$ s.t. $x^{-1} \cdot x = 1$
- **Identity property**:
  - $\exists(0)$ s.t. $x+0=x$
  - $\exists(1)$ s.t. $x \cdot 1 = x$
  - Commonly, we see: $\mathrm{I}(x) = x$

[1] $\exists$ = 'there exists'

2 Lecture 2: Basic Mathematics II

2.1 Summation Operator

- Consider a set $X = \{x_1, x_2, ..., x_n\}$
  - $\sum_{i=1}^{n} x_i = x_1 + x_2 + ... + x_n$
    - "The sum of x_i, over the range from $i = l$ through $i = n$."
- Let $Y = \{y_1, y_2, ..., y_n\}$
  - $\sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + ... + x_n y_n$
    - Here, we sum the outcome of $x_i y_i$ n times.
- But, what if there is more than one summation operator and subscript?

$$\begin{aligned}\Sigma_{j=1}^{m}\Sigma_{i=1}^{n} x_i y_j &= (y_1 x_1 + y_1 x_2 + ... + y_1 x_n) \\ &+ (y_2 x_1 + y_2 x_2 + ... + y_2 x_n) \\ &+ ... + (y_m x_1 + y_m x_2 + ... + y_m x_n)\end{aligned}$$

  We have m parentheses, with n x's in each parenthesis. We end up summing each combination of x_i and y_j.

- What if there is a constant, c? Drawing upon the *distributive property*:
  - $\sum_{i=1}^{n} c x_i = c x_1 + c x_2 + ... + c x_n = c \sum_{i=1}^{n} x_i$
- The *associative property* can also be applied to summation:
  - $\sum_{i=1}^{n} (x_i + y_i) = (x_1 + y_1) + (x_2 + y_2) + ... + (x_n + y_n) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$

2.2 Product Operator

- $X = \{x_1, x_2, ..., x_n\}$ and $Y = \{y_1, y_2, ..., y_n\}$
- $\prod_{i=1}^{n} x_i = x_1 \cdot x_2 \cdot ... \cdot x_n$
- $\prod_{i=1}^{n} (x_i + y_i) = (x_1 + y_1)(x_2 + y_2)...(x_n + y_n)$
  - We can't split $(x_i + y_i)$ like in summation. Instead, we multiply $(x_i + y_i)$ repeatedly.
- What about a constant?
  - $\prod_{i=1}^{n} c x_i = (c x_1)(c x_2) \cdot ... \cdot (c x_n) = c^n \prod_{i=1}^{n} x_i$

  We can move c to the front, but we have to exponentiate it to c^n.

2.3 Factorials, Permutations, and Combinations

Most of our quantitative coursework is about modeling probabilities, where:

$$\text{probability} = \frac{\#\text{occurences}}{\#\text{possibilities}}$$

For both the numerator and denominator we are dealing with *counting* the number of relevant outcomes. Factorials, permutations, and combinations are foundational concepts when it comes to counting.

- Some useful illustrations/properties:
  - $0! = 1$
  - $x! = x \cdot (x-1) \cdot (x-2) \cdot ... \cdot 0!$
  - $2! = 2 \cdot 1$
  - $3! = 3 \cdot 2 \cdot 1$
  - $10! = 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot ... \cdot 1$
  - If n objects in boxes of size x where *order matters*, then the number of *permutations* is:

    $$\frac{n!}{(n-x)}$$

  - If order does not matter, then the number of *combinations* is:

    $$\frac{n!}{(n-x)!x!} = nCx = \binom{n}{x}$$

    - Pronounced: 'n choose x'

2.4 Solving equations, inequalities, and for roots

- Solving an equation example:

$$\begin{aligned} 3x + 4y + 8 = 0 \\ 4y &= -(3x+8) \\ y &= \frac{-(3x+8)}{4} \\ y &= -\frac{3}{4}x - 2 \end{aligned}$$

We can also write this answer as:

$$\{(x, y) \in \mathbb{R} | y = -\frac{3}{4}x - 2\}$$

- Solving an inequality ($x >$, $x \geqslant y$, $y < x$, $y \leqslant x$) example:

$$\begin{aligned} -4y &> 2x + 12 \\ y &< -\frac{2x}{4} - \frac{12}{4} \\ y &< \frac{x}{2} - 3 \end{aligned}$$

  Note that dividing by a negative flips the sign

- Solving for a quadratic
  - Quadratic formula:

$$\begin{aligned} &ax^2 + bx + c \\ &x \in \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \end{aligned}$$

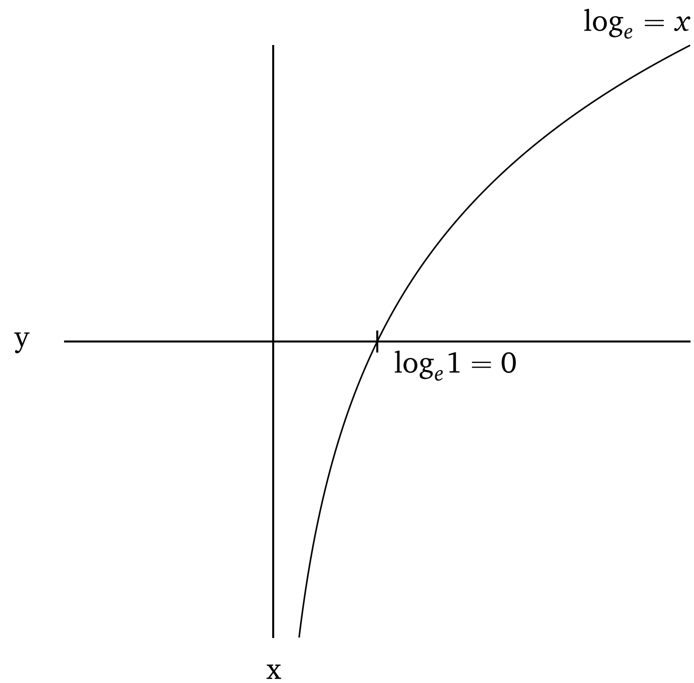
    Example:

$$\begin{aligned} &1.4x^2 + 3.7x + 1.1 = 0 \\ &a = 1.4, b = 3.7, c = 1.1 \\ &x = \frac{-3.7 \pm \sqrt{3.7^2 - 4 \times 1.4 \times 1.1}}{2.8} \\ &x = -0.341 \text{ or } x = -2.301 \end{aligned}$$

2.5 Logarithms

- If $y = a^x$, then we can rewrite as $\log_a y = x$
- Examples:
  - $\log_e e = 1$, because $e^1 = e$
  - Solving for x:
    - If $8 = 2^x$, then $\log_2 8 = x = 3$
- **Rules**
  - $\log(m + n) = \log(m) + \log(n)$
  - $\log(\frac{m}{n}) = \log(m) - \log(n)$
  - $\log(b^a) = a\log b$
  - $\log b^a = (\log_b e)(\log_e a)$
  - $\log b^a = \frac{1}{\log_a b}$

- **Visual**



- Note: $log_e(x) \equiv ln(x)$, pronounced the 'natural logarithm'

3 Lecture 3: Linear Algebra I

Most, if not all, algebra learned throughout K-12 educations deals with *scalar* algebra. Each variable only represents a single number. But, a variable can represent more than one element. Instead:

- *Scalar*: one element, x
- *Vector*: n elements
    - $\boldsymbol{x}$ includes the set of scalars $x_1, x_2, ..., x_i, ..., x_n$
- *Matrix*: $n \times m$ elements
    - $\boldsymbol{X}$ include the set of scalars $x_{11}, ..., x_{1m}, x_{21}, ..., x_{2m}, ..., x_{n1}, ..., x_{nm}$

There are considerable gains to be made through this notation. Beyond a more efficient notation, we're able to easily carry out all sorts of useful manipulations. We'll work up to these benefits in this lecture through various concepts.

3.1 Systems of equations

- *System of equations*: two or more equations with the same variables
    - To find a unique solution, we need as many equations as variables. E.g.,

$$\begin{aligned} 6x_1 - 3x_2 + 4x_3 &= -13 \\ 6x_1 &= -13 + 3x_2 - 4x_3 \\ x_1 &= -\frac{13}{6} + \frac{1}{2}x_2 - \frac{2}{3}x_3 \end{aligned}$$

    - We can use this single equation and solution to create a series of solutions, i.e. if $x_2 = 2$ and $x_3 = 3$, then $x_1 = -\frac{13}{6}$ and so on.

- Consider the following system of (linear) equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + ... + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + ... + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + ... + a_{mn}x_n &= b_m \end{aligned}$$

  Drawing on yesterday's lecture, we can also write this all as: $\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_j = b_i$ Or, bringing matrix notation in, as: $\boldsymbol{Ax} = \boldsymbol{b}$, where $\boldsymbol{A}$ is an $n \times m$ matrix.

3.1.1 Solving for a system of equation

Consider the following system of equations. Throughout r references row number. By setting all variables but one per equation, we are solving by *elimination*. We'll later introduce Gauss-Jordan elimination, which is the same process, but with an augmented matrix.

- First, the equations:

$$1x_1 + 1x_2 = 100,000$$
$$0.05x_1 + 0.09x_2 = 7,800$$

- Add $-0.05r1$ to $r2$

$$1x_1 + 1x_2 = 100,000$$
$$0x_1 + 0.04x_2 = 2,800$$

- $25r2$

$$1x_1 + 1x_2 = 100,000$$
$$0x_1 + 1x_2 = 70,000$$

- Subtract r2 from r1

$$1x_1 + 0x_2 = 30,000$$
$$0x_1 + 1x_2 = 70,000$$

- This gives us: $x_1 = 30,000$ and $x_2 = 70,000$

Now, let's try a lengthier example. Again, we're not substituting equations into each variable. We're eliminating all variables but one from each equation.

- First, the equations:

$$1x_1 + 2x_2 + 3x_3 = 6$$
$$2x_1 - 3x_2 + 2x_3 = 14$$
$$3x_3 + 1x_2 - 1x_3 = -2$$

- Subtract $-2r1$ from $r2$ and subtract $-3r1$ from $r3$

$$1x_1 + 2x_2 + 3x_3 = 6$$
$$0x_1 - 7x_2 - 4x_3 = 2$$
$$0x_3 - 5x_2 - 10x_3 = -20$$

- Multiply $r3$ by $-\frac{1}{5}$ and flip $r2$ and $r3$

$$1x_1 + 2x_2 + 3x_3 = 6$$
$$0x_1 + 1x_2 + 2x_3 = 4$$
$$0x_3 - 7x_2 - 4x_3 = 2$$

- Subtract $2r2$ from $r1$ and add $7r2$ to $r3$

$$1x_1 + 0x_2 - 1x_3 = -2$$
$$0x_1 + 1x_2 + 2x_3 = 4$$
$$0x_3 + 0x_2 + 10x_3 = 30$$

- Divide $r3$ by 10

$$1x_1 + 0x_2 - 1x_3 = -2$$
$$0x_1 + 1x_2 + 2x_3 = 4$$
$$0x_3 + 0x_2 + 1x_3 = 3$$

- Add $r3$ to $r1$ and subtract $2r3$ from $r2$

$$1x_1 + 0x_2 - 0x_3 = 1$$
$$0x_1 + 1x_2 + 0x_3 = -2$$
$$0x_3 + 0x_2 + 1x_3 = 3$$

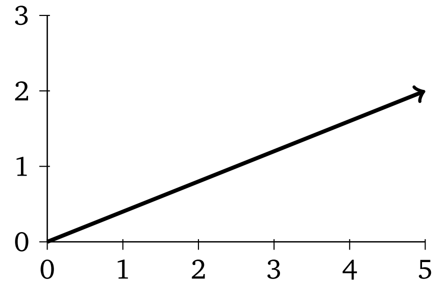
- $x_1 = 1, x_2 = -2, x_3 = 3$

Writing all of the $x's$ gets tedious though. So let's introduce vectors.

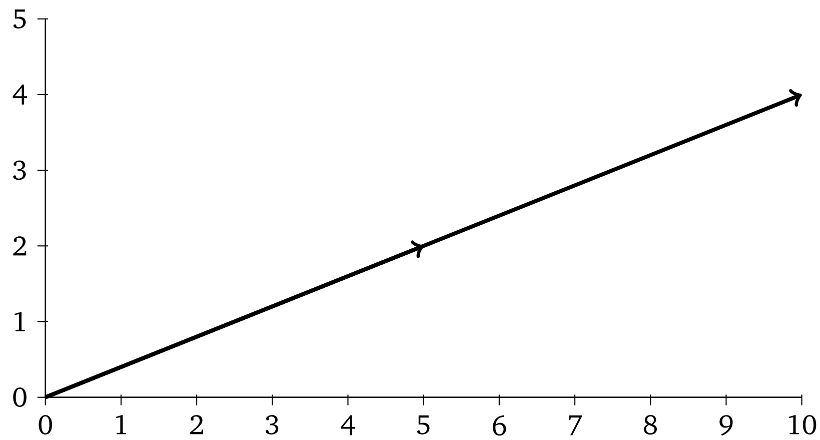
3.2 Vectors

- Examples:
  - $\boldsymbol{x} = [x_1, x_2, x_3, ..., x_n]$

- $\boldsymbol{x} = (5, 2)$
  - If, like in this case, there are two dimensions (number of components), then we can visually understand the vector as:



- If we multiply by a scalar: $a = 2$, $aX = [10, 4]$



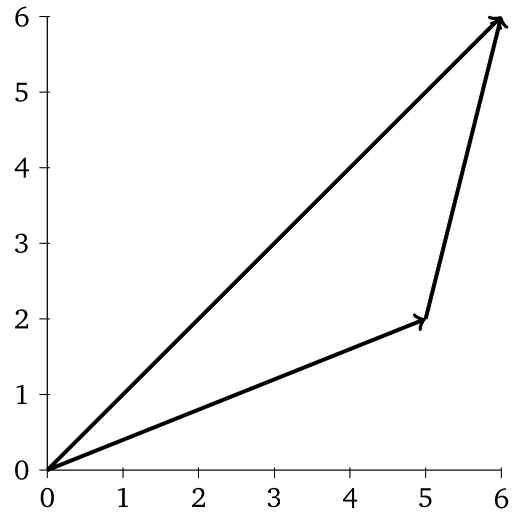
3.2.1 Vector length

- Also called the *norm*, length is not the same as *dimensions*. The formula is:

$$||x|| = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}$$

- Considering the visuals above, in two dimensions we are just using the Pythagorean theorem. But the formula is extendable to k-dimensions. For $\boldsymbol{x} = (5, 2)$, $||\boldsymbol{x}|| = \sqrt{25 + 4} = \sqrt{29}$.

- Another example
  - $x = (5, 2), y = (1, 4)$
    - $x + y = (6, 6)$
    - $||x + y|| = \sqrt{36 + 36} = \sqrt{72}$



3.2.2 Vector multiplication

- If c is a *scalar* and we multiply $a(x_1, x_2, ..., x_n)$, then we get $(ax_1, ax_2, ..., ax_n)$. Dividing by a scalar works the same way.

- But what about multiplying one vector by another vector? We use the **dot product**: $\boldsymbol{a} \cdot \boldsymbol{b}$. Another name for this operation is the **inner product**.[2]

  ◦ If $\boldsymbol{a}$ and $\boldsymbol{b}$ are both n-dimensional, then $\boldsymbol{a} \cdot \boldsymbol{b} = a_1 b_1 + a_2 b_2 + ... + a_n b_n = \sum_{i=1}^{n} a_i b_i$

$$\text{Ex: } \begin{bmatrix} 5 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 1 & 4 \end{bmatrix} = (5 \cdot 1) + (2 \cdot 4) = 5 + 8 = 13$$

  ◦ Note: the result of the dot product of vectors is a *scalar*.

- The **outer product** of two vectors instead produces a matrix:

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \cdot \begin{bmatrix} b_1 & b_2 \end{bmatrix} = \begin{bmatrix} a_1 b_1 & a_1 b_2 \\ a_2 b_1 & a_2 b_2 \end{bmatrix}$$

  ◦ The dimensions of this matrix are the two outer dimensions of the vectors multiplied together:

$$\underset{3\times1}{\begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}} \underset{1\times3}{\begin{bmatrix} 1 & 2 & 3 \end{bmatrix}} = \underset{3\times3}{\begin{bmatrix} 3 & 6 & 9 \\ 2 & 4 & 6 \\ 1 & 2 & 3 \end{bmatrix}}$$

[2] For a nice review of vector manipulation, see https://people.cs.clemson.edu/~dhouse/courses/401/notes/vectors.pdf

- ◦ *But* the inner dimensions must match up. See 1 and 1 above. If the first matrix's number of columns is not equal to the second matrix's number of rows, then cannot multiply.

3.3 Matrices

- A **matrix** is a rectangular table of numbers or variables arranged in a specific order in rows and columns. We express dimensions by rows, n, and columns, m. The dimensions of a matrix $A_{n\times m}$ are pronounced 'n by m'.

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} \in \mathbb{R}^{nm} = \mathbb{R}^{3\times 3}$$

- If $m = n$, then the matrix is symmetric/square.

- **Types of matrices**:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \text{zero matrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \text{diagonal matrix} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \text{identity matrix}$$

3.4 Matrix operators

- **Addition**

  - ◦ Must have the same number of elements

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 1 \\ 1 & 2 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \text{can't do}$$

$$\begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 3 & 4 \\ 6 & 8 \end{bmatrix}$$

- **Transposition**

  - ◦ Rotate so that the first column becomes the first row:

$$\boldsymbol{X} = \underset{3\times 2}{\begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 1 & 2 \end{bmatrix}}, \quad \boldsymbol{X}^T = \underset{2\times 3}{\begin{bmatrix} 1 & 0 & 1 \\ 2 & 0 & 2 \end{bmatrix}}$$

- **Multiplication**

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} 7 & 0 & -1 \\ 1 & 3 & 1 \end{bmatrix}$$

  - Because $\boldsymbol{A}$ is 2×2 and $\boldsymbol{B}$ is 2×3, we can multiply. But $\boldsymbol{AB}^T$ is undefined because $\boldsymbol{B}$ is 3×2.

$$\begin{aligned}
\boldsymbol{AB} &= \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 7 & 0 & -1 \\ 1 & 3 & 1 \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{A}_{11}\boldsymbol{B}_{11} + \boldsymbol{A}_{12}\boldsymbol{B}_{21} & \boldsymbol{A}_{11}\boldsymbol{B}_{12} + \boldsymbol{A}_{12}\boldsymbol{B}_{22} & \boldsymbol{A}_{11}\boldsymbol{B}_{13} + \boldsymbol{A}_{12}\boldsymbol{B}_{23} \\ \boldsymbol{A}_{21}\boldsymbol{B}_{11} + \boldsymbol{A}_{222}\boldsymbol{B}_{21} & \boldsymbol{A}_{21}\boldsymbol{B}_{12} + \boldsymbol{A}_{22}\boldsymbol{B}_{22} & \boldsymbol{A}_{21}\boldsymbol{B}_{13} + \boldsymbol{A}_{22}\boldsymbol{B}_{23} \end{bmatrix} \\
&= \begin{bmatrix} (1 \cdot 7) + (2 \cdot 1) & (1 \cdot 0) + (2 \cdot 3) & (1 \cdot -1) + (2 \cdot 1) \\ (3 \cdot 7) + (4 \cdot 1) & (3 \cdot 0) + (4 \cdot 3) & (3 \cdot -1) + (4 \cdot 1) \end{bmatrix} = \begin{bmatrix} 9 & 6 & 1 \\ 29 & 12 & 1 \end{bmatrix}
\end{aligned}$$

  - Ex with identity matrix:

$$\begin{aligned}
\boldsymbol{I}_{2\times 2}\boldsymbol{X} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \\
\text{b/c} & \begin{bmatrix} (1 \cdot 1) + (0 \cdot 3) & (0 \cdot 1) + (1 \cdot 2) \\ (0 \cdot 1) + (1 \cdot 3) & (0 \cdot 3) + (1 \cdot 4) \end{bmatrix}
\end{aligned}$$

4 Linear Algebra II

- Let's review matrix multiplication once before moving forward:

$$\underset{2\times 3}{A = \begin{bmatrix} 1 & 2 & -3 \\ 4 & 0 & -2 \end{bmatrix}}, \quad \underset{3\times 2}{B = \begin{bmatrix} 3 & 1 \\ 2 & 4 \\ -1 & 5 \end{bmatrix}}$$

- The dimensions of the product of these two matrices is the outside dimensions:

$$\begin{aligned} \boldsymbol{AB} &= (\mathbf{2} \times 3) \cdot (3 \times \mathbf{2}) = 2 \times 2 \\ \boldsymbol{BA} &= (\mathbf{3} \times 2) \cdot (2 \times \mathbf{3}) = 3 \times 3 \end{aligned}$$

- We know we can multiply the two together in either direction because the inner dimensions match in either order.

$$\begin{bmatrix} 1 & 2 & -3 \\ 4 & 0 & 2 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 2 & 4 \\ -1 & 5 \end{bmatrix} = \begin{bmatrix} 3+4+3 & 1+8-15 \\ 12+0-2 & 4+0+10 \end{bmatrix} = \begin{bmatrix} 10 & -6 \\ 10 & 14 \end{bmatrix}$$

- One more example:

$$\begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 7 & 8 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 4+14+1 & 3+16+2 \\ 0+7+1 & 0+8+2 \end{bmatrix} = \begin{bmatrix} 19 & 21 \\ 8 & 10 \end{bmatrix}$$

4.1 Matrix Representation of Systems of Equations

- Take this system of equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

- Or in matrix form:

$$\underset{m\times n}{\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}} \underset{n\times 1}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}} = \underset{m\times 1}{\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}}$$

- Ex:

$$\begin{aligned} 2x_1 + 3x_2 - 1x_3 &= 7 \\ 4x_1 + 5x_2 + 6x_3 &= 8 \\ -1x_1 + 2x_2 + 1x_3 &= 9 \end{aligned}$$

$$\Downarrow$$

$$\underbrace{\begin{bmatrix} 2 & 3 & -1 \\ 4 & 5 & 6 \\ -1 & 2 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}}_{X} = \underbrace{\begin{bmatrix} 7 \\ 8 \\ 9 \end{bmatrix}}_{b}$$

4.2 Linear Dependence and Independence

- The **span** of a vector is all of its linear combinations. I.e. $c \begin{bmatrix} 2 \\ 3 \end{bmatrix} \forall c$.
  - On linear combinations: the linear combination of

$$\begin{aligned} 2x \\ 3x \end{aligned}$$

    is $x \begin{bmatrix} 2 \\ 3 \end{bmatrix}$
- If one vector falls in the span of another vector then the two are **linearly dependent**. There is no new information in the second vector.
  - $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ and $\begin{bmatrix} 4 \\ 6 \end{bmatrix}$ are linearly dependent. The second is the first times 2.
- If one vector *does not* fall in the span of another vector, then the two are **linearly independent**.
  - $\begin{bmatrix} 7 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ -1 \end{bmatrix}$ are linearly independent. One cannot be represented as a linear combination of the other.
- If we were to draw each item in the vector set, then we are interested in whether or not the overall span of a set changes when a vector is added or removed.

4.3 Properties of Matrix Operators

- Matrix Addition
  - If $\boldsymbol{A}$ and $\boldsymbol{B}$ are both the same size ($m \times n$):
    1. $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$
    2. $\boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}) = (\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C}$

3. Additive Inverse: $\boldsymbol{A} + (-\boldsymbol{A}) = 0_{m \times n}$

- Matrix Multiplication:
  - If $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ are conformable:
    1. $\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}$
    2. $\boldsymbol{IA} = \boldsymbol{AI} = \boldsymbol{A}$
       - Where $\boldsymbol{I}$ is an identity matrix, e.g. $\boldsymbol{I}_{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$
    3. $\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC} \neq \boldsymbol{BA} + \boldsymbol{CA}$ because order matters

- Matrix Exponents
  1. $\boldsymbol{A}^p \cdot \boldsymbol{A}^q = \boldsymbol{A}^{p+q}$
  2. $(\boldsymbol{A}^p)^q = \boldsymbol{A}^{pq}$
  3. $(\boldsymbol{AB})^p \neq \boldsymbol{A}^p\boldsymbol{B}^p$ unless $\boldsymbol{AB} = \boldsymbol{BA}$

- Scalar Multiplication; $\boldsymbol{A}, \boldsymbol{B}$ are matrices and r, s are scalars:
  1. $r(s\boldsymbol{A}) = (rs\boldsymbol{A})$
  2. $(r + s)\boldsymbol{A} = r\boldsymbol{A} + s\boldsymbol{A}$
  3. $r(\boldsymbol{A} + \boldsymbol{B}) = r\boldsymbol{A} + r\boldsymbol{B}$
  4. $A(r\boldsymbol{B}) = r\boldsymbol{AB} = \boldsymbol{AB}r$

- Matrix Transposition
  1. $(\boldsymbol{A}^T)^T = \boldsymbol{A}$
  2. $(\boldsymbol{A} + \boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$
  3. $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$
  4. $(rA)^T = r(A^T)$

- Symmetric Matrix: A square matrix is symmetric if $\boldsymbol{A}^T = \boldsymbol{A}$
  - Ex: $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 1 \\ 3 & 1 & 5 \end{bmatrix}$

4.4 Idempotent Matrices

- A matrix, $\boldsymbol{A}$, is idempotent if $\boldsymbol{AA} = \boldsymbol{A}$. Multiplying the matrix by itself returns the original matrix.

4.5 Reduced Row/Row Echelon Form and Solving Linear Systems of Equations Gauss-Jordan Reduction/Elimination

- We can use a matrix to represent a system of equations:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n &= b_m \end{aligned}$$

- Rewritten as an **augmented matrix**:

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & ... & a_{1n} & b_1 \\ a_{21} & a_{22} & ... & a_{2n} & b_2 \\ \vdots & & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & ... & a_{mn} & b_m \end{array}\right]$$

- To solve for a system of equations, we often want to translate a matrix into **row echelon** or **reduced row echelon** form. What conditions describe a matrix in row echelon and reduced row echelon form. The conditions are:

  1. If any rows are zeros, then they are below nonzero rows (include at least one nonzero element):
  $\left[\begin{array}{ccc|c} 1 & 2 & 3 & 7 \\ 0 & 0 & 0 & 0 \end{array}\right]$
  2. The first non-zero entry in any row is 1 (except all zero row).
  $\left[\begin{array}{ccc|c} 0 & 1 & 3 & 7 \\ 0 & 0 & 0 & 0 \end{array}\right]$
  3. For each non-zero row the leading 1 is to the right of each 1 in the row above.
  $\left[\begin{array}{ccc|c} 1 & 0 & 3 & 7 \\ 0 & 1 & 4 & 0 \\ 0 & 0 & 1 & 2 \end{array}\right]$
  4. **Reduced row-echelon form**: makes the solution to a system of equations obvious. Below $x_1 = 7, x_2 = 0, x_3 = 2$
  $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 2 \end{array}\right]$

- Transforming a matrix into row echelon and reduced row echelon form is referred to as **Gauss-Jordan elimination**. We do so through **elementary row operators**:

- ◦ Multiplying a row by a constant
  - Remember, the matrix is a system of equations. So we're just multiplying both sides of an equation.
- ◦ Adding/subtracting rows
- ◦ Interchanging rows

- Example:

$$\begin{aligned} 1x_1 + 2x_2 + 4x_3 &= 3 \\ 2x_1 + 1x_2 + 3x_3 &= 2 \\ 1x_1 + 2x_2 + 2x_3 &= 3 \end{aligned}$$

- ◦ As an augmented matrix:

$$\left[\begin{array}{ccc|c} 1 & 2 & 4 & 3 \\ 2 & 1 & 3 & 2 \\ 1 & -2 & 2 & 3 \end{array}\right]$$

- ◦ Add $-2r1$ to $r2$ and $-1r1$ to $r3$

$$\left[\begin{array}{ccc|c} 1 & 2 & 4 & 3 \\ 0 & -3 & -5 & -4 \\ 0 & -4 & -2 & 0 \end{array}\right]$$

- ◦ $r3 \times -\frac{1}{4}$ and interchange $r2$ and $r3$

$$\left[\begin{array}{ccc|c} 1 & 2 & 4 & 3 \\ 0 & 1 & 1/2 & 0 \\ 0 & -3 & -5 & -4 \end{array}\right]$$

- ◦ Add $3r2$ to $r3$

$$\left[\begin{array}{ccc|c} 1 & 2 & 4 & 3 \\ 0 & 1 & 1/2 & 0 \\ 0 & 0 & -7/2 & -4 \end{array}\right]$$

- ◦ Multiply $r3$ by $-\frac{2}{7}$

$$\left[\begin{array}{ccc|c} 1 & 2 & 4 & 3 \\ 0 & 1 & 1/2 & 0 \\ 0 & 0 & 1 & 8/7 \end{array}\right]$$

- ◦ Subtract $4r3$ from $r1$ and subtract $\frac{1}{2}r3$ from $r2$

$$\left[\begin{array}{ccc|c} 1 & 2 & 0 & -11/7 \\ 0 & 1 & 0 & -4/7 \\ 0 & 0 & 1 & 8/7 \end{array}\right]$$

- Subtract $2r2$ from $r1$

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & -3/7 \\ 0 & 1 & 0 & -4/7 \\ 0 & 0 & 1 & 8/7 \end{array}\right]$$

- $x_1 = -\frac{3}{7}, x_2 = -\frac{4}{7}, x_3 = \frac{8}{7}$
- Or: $\mathrm{I}_3 X = \begin{bmatrix} -3/7 \\ -4/7 \\ 8/7 \end{bmatrix}$

5 Linear Algebra III

5.1 Matrix Inversion

- For a scalar a, the inverse: $a^{-1} = \frac{1}{a}$, where $a^{-1}a = 1$.
- We can also invert a matrix A: $A^{-1}A = AA^{-1} = \mathrm{I}$
  - Note: if A^{-1} does not exist, then A is singular/not invertible.
  - $A^T \neq A^{-1}$
- For a 2×2 matrix, we use the following steps. We'll cover larger matrices after, for which we'll use a lengthier, but more intuitive process.

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

Where $\frac{1}{ad-bc}$ is the *determinant*, which we discuss more later. An example:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$$
$$A^{-1} = \frac{1}{(1 \cdot 4) - (3 \cdot 2)} \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix}$$
$$= \frac{1}{-2} \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 3/2 & -\frac{1}{2} \end{bmatrix}$$

- How to check if $A^{-1}A = \mathrm{I}$?

$$\begin{bmatrix} -2 & 1 \\ 3/2 & -1/2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

5.2 Properties of Matrix Inversion

Singular vs nonsingular:

- Invertible = Nonsingular
- Noninvertible = Singular
- If A is nonsingular, then A^{-1} is nonsingular
- If A & B are nonsingular, then $(AB)^{-1} = B^{-1}A^{-1}$

- If A is nonsingular, then $(A^T)^{-1} = (A^{-1})^T$

Some conditions for nonsingularity (we can find A^{-1}, there are often more conditions listed, but they are technically covered by these three):

1. Rows and columns are linearly independent
   - No rows and/or columns add up to each other. Generally speaking if rows and columns are not linearly independent, then the matrix is invertible.
2. Matrix A is row equivalent to I (can we use row operators to turn A into the identity matrix?)
3. The determinant (we cover below) is not 0.

One intuitive and practical procedure for finding A^{-1}, regardless of the size:

1. Find $[A|\mathrm{I}]$ where both A and I are both $n \times n$
2. Find the reduced row echelon form for A (left side)
3. If step 2 gives us $[\mathrm{I}|C]$, then $C = A^{-1}$

A note on rank: If A is $m \times n$, then the rank(A) =

$$\min \begin{cases} \text{max \# of linearly ind. rows} \\ \text{max \# of linearly ind. cols} \end{cases}$$

- E.g.: $\begin{bmatrix} \mathbf{1} & \mathbf{2} & \mathbf{3} \\ \mathbf{1} & 2 & 3 \\ \mathbf{2} & 4 & 6 \end{bmatrix}$ Only the bold rows and columns are linearly ind. So, rank$(A) = \min(1,1) = 1$

- **Examples:**

$$\boldsymbol{A}\mathrm{I} = \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 2 & 3 & 0 & 1 & 0 \\ 5 & 5 & 1 & 0 & 0 & 1 \end{array}\right]$$

  - Subtract 1/2 r2 from r1, divide r2 by 2, subtract 5r1 from r3

$$= \left[\begin{array}{ccc|ccc} 1 & 0 & -1/2 & 1 & -1/2 & 0 \\ 0 & 1 & 3/2 & 0 & 1/2 & 0 \\ 0 & 0 & 4 & -5 & 0 & 1 \end{array}\right]$$

  - Subtract 1/8r3 from r1, add 3/8r3 to r2, divide r3 by -4

$$= \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 13/8 & -1/2 & -1/8 \\ 0 & 1 & 0 & 15/8 & 1/2 & 3/8 \\ 0 & 0 & 1 & 5/4 & 0 & -1/4 \end{array}\right]$$

  - Where the right-hand side matrix is A^{-1}

- Another example, but A^{-1} doesn't exist:

$$\left[\begin{array}{ccc|ccc} 1 & 2 & -3 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 & 1 & 0 \\ 5 & -2 & -3 & 0 & 0 & 1 \end{array}\right]$$

  - Subtract r1 from r2, subtract 5r1 from r3

$$\begin{bmatrix} 1 & 2 & 3 & \dots & \dots & \dots \\ 0 & -4 & 4 & \dots & \dots & \dots \\ 0 & -12 & 12 & \dots & \dots & \dots \end{bmatrix}$$

  - We aren't concerned with the right-hand side matrix because the rows of the left matrix are not independent. r3 is a linear combination of r2, meaning the matrix is singular...

5.3 Determinant

- Determinants convert a matrix into a scalar but can only be defined for a square matrix. Determinants are useful for checking if a matrix is invertible. They also can play a role in solving for systems of equations. The formula is straightforward for a 2×2 matrix, but less so for larger matrices.
- Let $A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$
  - The determinant of A is the two diagonal products differenced:
    $|A| = (a_{11} \cdot a_{22}) - (a_{21} \cdot a_{12})$
- Examples:
  - $\begin{bmatrix} 2 & 3 \\ 1 & 5 \end{bmatrix} \Rightarrow (2 \cdot 5) - (3 \cdot 1) = 7$
  - $\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \Rightarrow 2 - 2 = 0$
- For a 3×3 matrix we sum the products of all elements in any row or column, alternating signs, and the determinants of a specific 2×2 submatrix. An element's submatrix is the remaining elements when the elements from the relevant row and column are removed. I.e., for:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

The submatrix for a_{23} is $\begin{bmatrix} a_{11} & a_{12} \\ a_{31} & a_{32} \end{bmatrix}$

Taking the first column, the determinant of A is:

$$a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$$

- We can use any row or column. But it is best to use one with zeros, if available.

- The **minor** of an element is the determinant of its submatrix.
  - $M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = (a_{21} \cdot a_{33}) - (a_{31} \cdot a_{23})$

5.3.1 Determinants via Cofactor Expansion

- The **cofactor** of any element i, j: $C_{ij} = (-1)^{i+j} M_{ij}$, which is used for calculating the determinants of $n \times n$ matrices where $n > 2$. i is rows, j is columns.
  - $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$
  - Ex: $C_{11} = (-1)^{1+1}\ M_{11} = 1\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}$

- The determinant of a $n \times n$ matrix where $n > 2$ is the sum of the products of each element and its cofactor for any row or column. We just choose a single row or a single column – ideally one with as many zeros as possible.
  - Given row i:
  $$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij}$$
  - Or row j:
  $$\det(A) = \sum_{i=1}^{n} a_{ij} C_{ij}$$

- Let's show what this means:

- Choose the row or column with the most zeros:
  - Ex: $\begin{bmatrix} 1 & 2 & -3 & 4 \\ -4 & 2 & 1 & 3 \\ 0 & 0 & 0 & -3 \\ 2 & 0 & -2 & 3 \end{bmatrix}$
  - So let's take all elements in row 3, multiply each times its cofactor, and add it all together. Because of the zeros, we only have to find one cofactor!

$$
\begin{aligned}
|A| = \sum_{j=1}^{4} a_{3j}C_{3j} &= a_{31}C_{31} + a_{32}C_{32} + a_{33}C_{33} + a_{34}C_{34} \\
a_{34}C_{34} &= 0 + 0 + 0 + a_{34}C_{34} \\
a_{34}C_{34} &= -3\left[(-1)^{3+4}\begin{vmatrix} 1 & 2 & -3 \\ -4 & 2 & 1 \\ 2 & 0 & 2 \end{vmatrix}\right] \\
&= -3 \cdot -1 \left[2\begin{vmatrix} 2 & -3 \\ 2 & 1 \end{vmatrix} - 0\begin{bmatrix} 1 & -3 \\ -4 & 1 \end{bmatrix} + (-2)\begin{vmatrix} 1 & 2 \\ -4 & 2 \end{vmatrix}\right] \\
&= 3\left[2(2+6) - 2(2+8)\right] \\
&= 3[16 - 20] = 3(-4) = -12
\end{aligned}
$$

- To wrap up, some useful properties of determinants:
  1. $|A| = |A^T|$
  2. If a row or column of A is a linear combination of other rows or columns, then $\det(A) = 0$
  3. If A is diagonal, then $|A|$ is the product of the diagonals.
  4. $|AB| = |A| \cdot |B|$
  5. If A is non-singular, then $|A^{-1}| = \frac{1}{|A|}$

5.4 Adjoint Matrix

- **Adjoint matrix**: the transpose of a matrix of the cofactors of each element:

$$
\text{adj}(A) = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & \dots & \dots & c_{2n} \\ \vdots & & & \\ c_{n1} & \dots & \dots & c_{nn} \end{bmatrix}
$$

Where c is an element's cofactor. For example:

$$
A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \Rightarrow C = \begin{bmatrix} +\begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} & -\begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} & +\begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ -\begin{vmatrix} 2 & 3 \\ 8 & 9 \end{vmatrix} & +\begin{vmatrix} 1 & 3 \\ 7 & 9 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} \\ +\begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix} & +\begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} -3 & 6 & -3 \\ 6 & -12 & 6 \\ -3 & 6 & -3 \end{bmatrix}
$$

Where the last matrix is the adjoint matrix.

This information gives us another way to calculate the inverse of A with the following formula:

$$A^{-1} = \frac{\text{adj}^T}{|A|}$$

5.5 Trace

Let's close with something simpler. The **trace** of a $n \times n$ matrix is just the sum of the diagonal elements.

$$\text{Tr}(A) = \Sigma_{i=1}^{n} a_{11} = a_{11} + a_{22} + ... + a_{nn}$$

This is less frequently used, but worth being aware of.

6 Calculus I

Before we begin introducing mathematical formula, what is a derivative? A derivative calculates the rate of change of a function at any given point. In high school when we learned about the slope of a line – $y = mx + b$, where m is the slope – that slope is a derivative. In that context, the derivative is the same at all points because the line is straight. But we often encounter functions that are not a straight line and we care about calculating the slope across values of that function. We can use calculus for this.

6.1 Sequences and Limits

A sequence is an ordered list of numbers, e.g.

$$\{x_n\} = \{x_1, x_2, ..., x_n\}$$

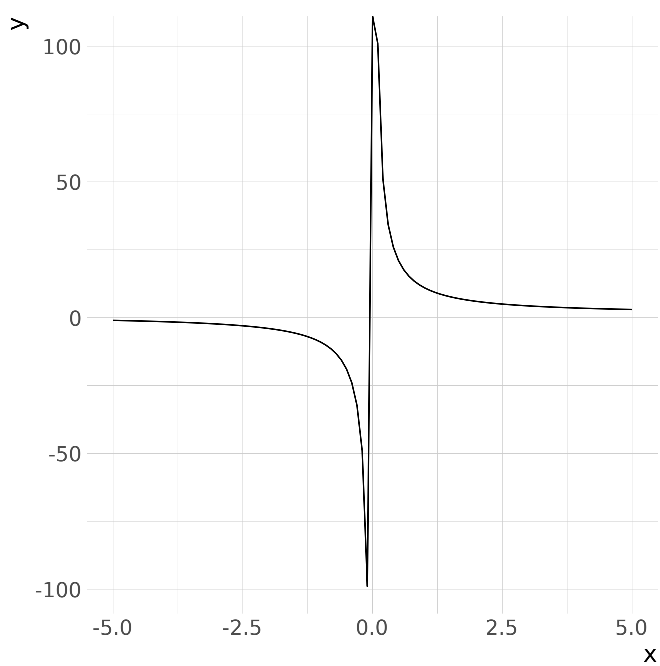
Where x_n is a real number that extends from x_1 to x_n. We usually encounter n extending to ∞. Another way to write the series is:

$$\{x_n\}_{n=1}^{\infty}$$

Central to calculus is the notion of a sequence "converging to a limit", generally where $n \to \infty$ or $n \to 0$. This is written as:

$$\lim_{n\to\infty} y_n = L$$

where L is the limit. Let's visualize for an arbitrary function, $f(x) = \frac{x+10}{x}$:



We can see that as $x \to \infty$ the value of $f(x)$ stabilizes. Indeed:

$$\lim_{x\to\infty} \frac{x+10}{x} = \lim_{x\to\infty} \left(1 + \frac{10}{x}\right) = \underbrace{\lim_{x\to\infty} 1}_{1} + \underbrace{\lim_{x\to\infty} \frac{10}{x}}_{0} = 1$$

This gives us a limit of 1.

6.2 Derivatives and the Difference Quotient

The derivative is the rate of change of $f(x)$ at any x. For a straight line, i.e $y = mx + b$, the derivative is constant at all points. But for a nonlinear function, i.e. $y = 2x^4$, the rate of change varies across x.

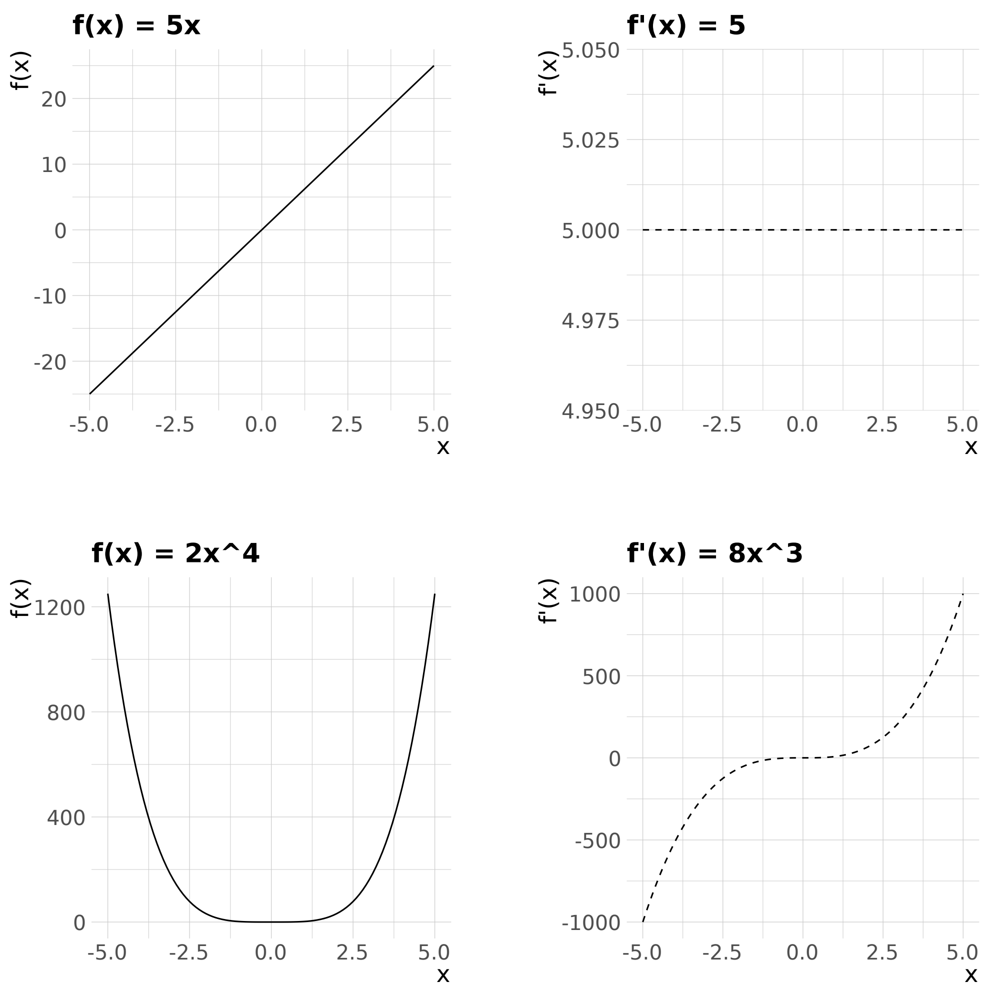


Figure 1: Functions and their derivatives

Now, let's introduce how we produce a derivative. Let f be a function with an open interval that contains x. Let h be the interval where $f(x)$ changes. Below we are simply calculating rise over run at each specified interval.

$$\frac{d}{dx}f(x) = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

The numerator represents the change in $f(x)$ as Δx approaches 0 and the denominator is the change in Δx – or the change in x. We generally see two notations for derivatives:

- Lagrange: $f'(x)$
- Leibniz: $\frac{d}{dx}y = \frac{dy}{dx}$
  - The change in y, given change in x.

Side-note, on the problem set, when we are asked about the difference quotient as a function of $x + \Delta x$, think about how a specified function would look if it were plugged in for $f(x)$ and $f(x + \Delta x)$.

6.3 Rules for Derivatives

- Power rule:
  - $y = f(x) = ax^n$, $f'(x) = nax^{n-1}$
- Constant multiplier rule:
  - $f(x) = ax$, $f'(x) = a$
- Constant rule:
  - $f(x) = a$, $f'(x) = 0$
- Summation rule:
  - $f(x) = g(x) + h(x)$, $f'(x) = g'(x)h'(x)$
- Product rule:
  - $f(x) = g(x)h(x)$
  - $f'(x) = g'(x)h(x) + g(x)h'(x)$
- Quotient rule:
  - $f(x) = g(x)/h(x)$
  - $f'(x) = \frac{g'(x)h(x) - g(x)h'(x)}{[h(x)]^2}$
- Chain rule:
  - $f(x) = h[g(x)]$
  - $f'(x) = h'[g(x)] \cdot g'(x)$
- Exponent rule:
  - $f(x) = a^x$

  - $f'(x) = a^x(ln(a))$
  - $f(x) = e^x$, $f'(x) = e^x$

- Logarithm rule:
  - $f(x) = \log_a x$
  - $f'(x) = \frac{1}{x \ln a}$
  - Natural log: $f(x) = ln(x)$, $f'(x) = \frac{1}{x}$

6.4 Derivative Examples

1. $f(x) = 40x^{400}$
   - $f'(x) = 400 \cdot 40x^{399} = 16000x^{399}$
2. $f(x) = 16x$
   - $f'(x) = 16$
3. $f(x) = 1000$
   - $f'(x) = 0$
4. $f(x) = 3x^{100} + 5x^2$
   - $f'(x) = 300x^{99} + 10x$
5. $f(x) = (3x + 5)(9x + 2)$
   - $f'(x) = 3(9x + 2) + 9(3x + 5)$
6. $f(x) = \frac{3x+5}{9x+2}$
   - $f'(x) = \frac{3(9x+2)-9(3x+5)}{(9x+2)^2}$
7. $f(x) = 20(x + 3)^{10}$
   - $f'(x) = 200(x + 3)^9 \cdot 1$
8. $f(x) = 20(x^3 + 3x)^{10}$
   - $f'(x) = 200(x^3 + 3x)^9 \cdot (3x^2 + 3)$
9. $f(x) = 20[(x + 3)^3 + 4x]^{10}$
   - $f'(x) = 200[(x + 3)^3 + 4x]^9 \cdot g'(x)$
   - $g'(x) = 3(x + 3)^2 \cdot 1 + 4$
   - $f'(x) = 200[(x + 3)^3 + 4x]^9 \cdot 3(x + 3)^2 + 4$

10. $f(x) = e^{\sqrt{x}}$, $f'(x) = e^{x^{\frac{1}{2}}} \cdot (\frac{1}{2}x^{-\frac{1}{2}})$

11. $f(x) = a^{\sqrt{x}}$, $f'(x) = a^{\sqrt{x}} ln(a) \cdot (\frac{1}{2}x^{-\frac{1}{2}})$

    - Keep in mind for the problem set.

12. $f(x) = ln(3x^2 + 3x + 3)$, $f'(x) = \frac{1}{3x^2+3x+3} \cdot 6x + 3 = \frac{6x+3}{3x^2+3x+3}$

7 Calculus II

7.1 The Chain Rule

Let's revisit the chain rule in more detail. Another way to write it out is:

$$\frac{dy}{dx} = \frac{dy}{du} \times \frac{du}{dx}$$

So we first specify one of the functions as u and then calculate $\frac{dy}{du}$. Then we calculate $\frac{du}{dx}$, which is the change in the function u, given a change in x. Last, we write out the product of the two derivatives. This is an incredibly common and powerful trick. One more example:

$$\begin{aligned} y &= \left[\frac{(x^2+1)(3x+9)}{(8x^2-9)}\right]^3 \\ u &= \frac{(x^2+1)(3x+9)}{(8x^2-9)} \\ \frac{dx}{dy} &= 3[u]^2 \cdot u' \end{aligned}$$

Where we solve for u' with the product and quotient rules.

7.2 Implicit Differentiation

Sometimes we cannot fully x and y when finding $\frac{dy}{dx}$. This means we cannot solve for y solely in terms of x. We use implicit differentiation for these situations. You'll notice that our answer to the following example includes both x and y. The trick is to write out all parts of our function in terms of $\frac{dy}{dx}$. Then we solve for $\frac{dy}{dx}$.

$$\begin{aligned} f(x,y) = y^5 + xy + x^2 \text{ at } f(x,y) &= 3 \\ \frac{d}{dx} f(x,y) &= \frac{d}{dx}(3) \\ \frac{d}{dx}(y^5 + xy + x^2) &= \frac{d}{dx}(3) \\ \frac{d}{dy}y^5 + \frac{d}{dx}xy + \frac{d}{dx}x^2 &= 0 \\ 5y^4 + \frac{dx}{dx}y + x\frac{dy}{dx} + 2x &= 0 \\ 5y^4 + y + \frac{dy}{dx}x + 2x &= 0 \\ \frac{dy}{dx} &= \frac{-5y^4 - y - 2x}{x} \end{aligned}$$

Now, we have a quantity for $\frac{dy}{dx}$, it just depends upon the values of $x \; and \; y$.

7.3 Partial Derivatives

What if we have a multivariate function (multiple variables are changing) and are only concerned with the change in y, given the change in *one* variable? The trick is to treat other variables as a constant, because we are only concerned with variation in one variable. So:

$$\begin{aligned} f(x,y) &= 5x^2y^3 \\ \frac{df}{dx} &= 10xy^3 \\ \frac{df}{dy} &= 15x^2y^2 \end{aligned}$$

Some additional notation: $\frac{df}{dx} = f_x$ or $\frac{df}{dy} = f_y$
Examples:

$$\begin{aligned} f(x,y) &= (x+4)(3x+2y) \\ f_x &= 1 \cdot (3x+2y) + (x+4)(3) \\ f_y &= 0(3x+2y) + (x+4)2 \\ &= 2(x+4) \end{aligned}$$

But:

$$\begin{aligned} f(x,y,z) &= (x+4)(3x+2y)(3z) \\ f_z &= 3(x+4)(3x+2y) \end{aligned}$$

This last example is simpler because the other parts of the function don't include z, so we treat them like constants and then derive $3z$.

7.4 Gradient

A vector with all of a function's patial derivatives is the **gradient**. Take a function $f(x)$, with n variables. The gradient – or $\nabla f(x)$ – is:

$$\nabla f(x) = \left[\frac{df(x)}{dx_1}, \frac{df(x)}{dx_2}, ..., \frac{df(x)}{dx_n}\right]$$

For a function $f(x,y) = \left[\frac{df}{dx}, \frac{df}{dy}\right] = [f_x, f_y]$

7.5 Second Derivatives and the Hessian

We can take higher-order derivatives, which are just the rate of change of the previous derivative. I.e. $f(x) = x^3$, $f'(x) = 3x^2$, $f''(x) = 6x$, and so on...
In your statistics courses you will encounter the **Hessian**, which is a matrix of all combinations of second derivatives:

$$\begin{bmatrix} \frac{d^2f}{dx_1dx_1} & \cdots & \frac{d^2f}{dx_ndx_1} \\ \frac{d^2f}{dx_2dx_1} & \frac{d^2f}{dx_2dx_2} & \vdots \\ \vdots & & \ddots \\ \frac{d^2f}{dx_ndx_1} & \cdots & \frac{d^2f}{dx_ndx_n} \end{bmatrix}$$

This gives us the change of the curvature of a function in all directions.

7.6 More on limits

We can also take 'one-sided' limits, which come from different directions. Consider $f(x) = \frac{1}{x}$:

$$\lim_{x \to 0^+} \frac{1}{x} = +\infty$$
$$\lim_{x \to 0^-} \frac{1}{x} = -\infty$$

We also sometimes need to simplify a function to find the limit. In this example, plugging 1 in gives us 0/0. But simplifying gives us 2.

$$\lim_{x \to 1} \frac{1 - x^2}{1 - x} = \frac{(1 - x)(1 + x)}{1 - x} = 1 + x = 2$$

7.7 L'Hôpital's Rule

One more trick for limits: If $\lim_{x \to c} f(x) = \lim_{x \to c} g(x) = 0$ or $= \pm\infty$, *and* $\lim_{x \to c} \frac{f'(x)}{g'(x)}$ exists, then:

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

This is useful for undefined limits. Ex:

$$\begin{aligned}
\lim_{x\to 1} \frac{x^2-1}{x^2+3x-4} &= \frac{1^2-1}{1^2+3(1)-4} = \frac{0}{0} \\
&= \lim_{x\to 1} \frac{2x}{2x+3} \\
&= \frac{2(1)}{2(1)+3} \\
&= \frac{2}{5}
\end{aligned}$$

8 Calculus III

8.1 Optimization

We often want to know where our functions are at their maximum or minimum. We do this in two steps. First, wherever the first derivative is at 0, means we are either at a minimum or maximum. Second, depending upon the direction of the second derivative, we can tell if we are at a maximum or minimum.

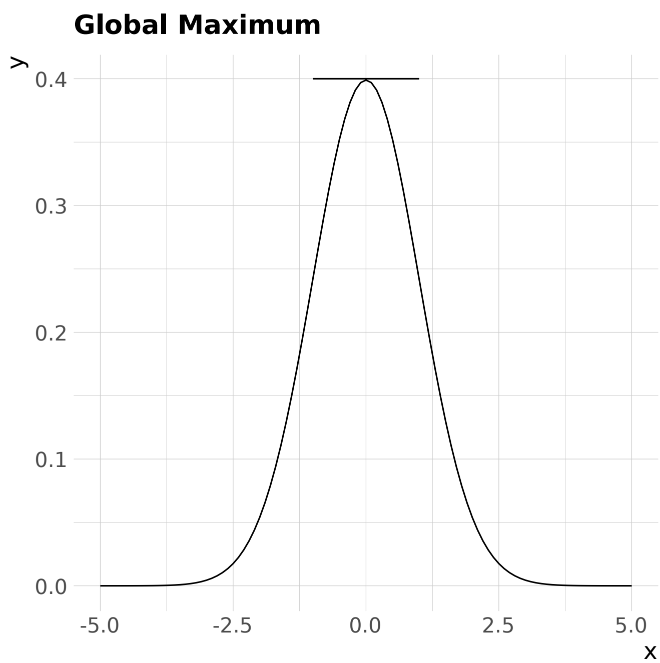


Figure 2: Visualizing how a derivative is equal to 0 at a function's maximum or minimum.

- **First-order condition**: $f'(x) = 0$
  - Maximum or minimum
  - This assumes that $f(x)$ is differentiable at all values – it is a continuous function.
- **Second-order condition**: $f''(x)$, where > 0 means minimum, < 0 means maximum, and $= 0$ an inflection point.

$$\begin{aligned} f(x) &= x^2 - 4x - 1 \\ \text{FOC: } f'(x) &= 2x - 4 \\ x &= 2 \\ \text{SOC: } f''(x) &= 2 \Rightarrow \text{minimum} \end{aligned}$$

What if we have multiple variables?

- Let $f(x) = -x_1^2 + x_1 x_2 - x_2^2$
  - FOC:
    - $f_{x_1} = -2x_1 + x_2 = 0$
    - $f_{x_2} = x_1 - 2x_2 = 0$

    $$\left[\begin{array}{cc|c} -2 & 1 & 0 \\ 1 & -2 & 0 \end{array}\right]$$

      - It turns out $x_1 = 0$, $x_2 = 0$
  - SOC: We need to build the hessian:

  $$\begin{bmatrix} \frac{d^2 f}{dx_1 x_1} & \frac{d^2}{dx_1 x_2} \\ \\ \frac{d^2 f}{dx_2 x_1} & \frac{d^2 f}{dx_2 x_2} \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$$

- Now, we ask whether these points are minima, maxima, indeterminate, or saddle points.
  - We calculate the determinants of each "principal minor".
  - $\text{PM}_1 = -2$, $D_1 = -2$
  - $\text{PM}_2 = \begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$, $D_2 = (-2 \cdot -2) - (1 \cdot 1) = 3$
- The hessian is:
  - Positive definite: $D_1 > 0$, $D_2 > 0$, strictly local minima
  - Negative definite: $D_1 < 0$, $D_2 > 0$, strictly local maxima
    - Start negative and then alternate signs
  - Positive semi-definite: $D_1 \geqslant 0$, $D_2 \geqslant 0$
  - Negative semi-definite: $D_1 \leqslant 0$, $D_2 \geqslant 0$
- For larger matrices?
  - $\begin{bmatrix} . \end{bmatrix}$, $\text{PM}_1 =$ determinant of 1×1
  - $\begin{bmatrix} . & . \\ . & . \end{bmatrix}$, $\text{PM}_2 =$ determinant of 2×2
  - $\begin{bmatrix} . & . & . \\ . & . & . \\ . & . & . \end{bmatrix}$, $\text{PM}_3 =$ determinant of 3×3

- If we have a 3×3 matrix:
  - PD: $D_1 > 0$, $D_2 > 0$, $D_3 > 0$
  - ND: $D_1 < 0$, $D_2 > 0$, $D_3 < 0$
  - PSD: $D_1 \geqslant 0$, $D_2 \geqslant 0$, $D_3 \geqslant 0$
  - NSD: $D_1 \leqslant 0$, $D_2 \geqslant 0$, $D_3 \leqslant 0$

8.2 Constrained Optimization – Lagrange Multiplier

Sometimes, we have to find the maxima or minima under set conditions.

$$\text{Max } f(x_1, x_2) \text{ s.t. } g(x_1, x_2) = 0$$
$$\text{Max } f(x_1, x_2) - \lambda g(x_1, x_2)$$

- λ is us creating a new variable.
- FOC:

$$\begin{aligned} f_{x_1} - \lambda g_{x_1} &= 0 \\ f_{x_2} - \lambda g_{x_1} &= 0 \\ g(x_1, x_2) &= 0 \end{aligned}$$

- Then:

$$\begin{aligned} f(x_1, x_2) =& 36 - x_1^2 - x_2^2, \ g(x_1, x_2) = x_1 + 7x_2 - 25 \\ f(x_1, x_2, \lambda) =& 36 - x_1^2 - x_2^2 - \lambda(x1 + 7x_2 - 25) \\ f_{x_1} &= -2x_1 - \lambda = 0, \Rightarrow x_1 = -\lambda/2 \\ f_{x_2} &= -2x^2 - 7\lambda = 0, \Rightarrow x_2 = -7\lambda/2 \\ g(x_1, x_2) &= -\lambda/2 + -7\lambda/2 - 25 = 0 \\ \lambda &= -1 \\ \text{Therefore: } x_1 &= \frac{1}{2}, x_2 = \frac{7}{2} \end{aligned}$$

8.3 Integration

- $f(x) = 40x$
- $F(x) = 20x^2 + c$

  - This is the 'antiderivative', $F'(x) = f(x)$, because if $F(X)$ is derived then it returns $f(x)$.
  - We add c because any constant will turn to 0 when the derivative is taken.

- Powerfully, antidifferentiation lets us solve for the value of y, given dy/dx.

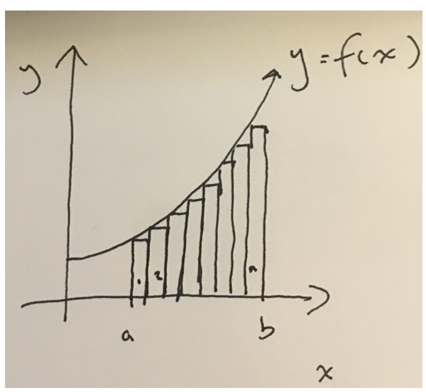
- **Rules of integration**

  - Power rule: $f(x) = x^n$, $F(x) = \frac{1}{n+1}x^{n+1} + c$
  - Exponent rule: $f(x) = e^x$, $F(x) = e^x + c$
    - But: $f(x) = e^a x$, $F(x) = \frac{1}{a}e^{ax} + c$
  - Logarithm rule: $f(x) = \frac{1}{x}$, $F(x) = ln(x) + c$
  - Chain rule: $f(x) = g'(x)e^{g(x)}$, $F(x) = e^{g(x)} + c$
  - Chain + log: $f(x) = \frac{g'(x)}{g(x)}$, $F(x) = ln|g(x)| + c$
  - Sum rule: $f(x) = g(x) + h(x)$, $F(x) = G(x) + H(x)$
  - Constant rule: $f(x) = kg(x)$, $F(x) = kG(x)$
  - Integral of a constant: $f(x) = k$, $F(x) = kx + c$

We integrate so that we can add continuously. The integral is the limit of the sum, where we add all slices under the curve until we reach infinity. We end up summing the values of infinitesimally small ranges. This is the idea behind Riemann sums.

8.4 Areas and Riemann Sums

Consider the following function



where $\sum_{u=1}^{n} f(x_i)\Delta x$ and $\Delta x = \frac{b-a}{n}$. This equals the sum of the area of all the rectangles under $f(x)$. The closer n gets to ∞, then the closer the sum of all the areas gets to the area under $f(x)$ from a to b.

It also turns out that

$$\lim_{n\to\infty} \sum_{i=1}^{n} f(x_i)\Delta x = \int_b^a f(x)dx$$

Because of this, we can use a definite integral to calculate the area under a function/curve between two points.

9 Calculus IV

9.1 Fundamental Theorem of Calculus

This brings us to the grandiose titled fundamental theorem of calculus:

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

Where the difference between the antiderivative at two points gives us the area under a curve between these two values. This is given such importance because it links derivatives and integrals.
Ex:

$$\int_1^3 x^2 dx = \frac{1}{3}x^3\Big|_1^3 = \frac{27}{3} - \frac{1}{3} = \frac{26}{3}$$

9.2 Rules for definite integrals

1. $$\int_a^b f(x)d(x) = -\int_b^a f(x)dx$$

2. $$\int_a^d f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx + \int_c^d f(x)dx \quad \text{where} \quad a \leqslant b \leqslant c \leqslant d$$

3. $$\int_a^b -f(x)dx = -\int_a^b f(x)dx$$

4. $$\int_a^b kf(x)dx = k\int_a^b f(x)dx$$

5. $$\int_a^b [f(x) \pm g(x)]\, dx = \int_a^b f(x)dx \pm \int_a^b g(x)dx$$

9.3 Integration by substitution

1. Define a new variable, $u = G(x)$ such that when written in terms of u the integrand is simpler.

2. Integrate the function, changing the limits of the integral to be from a and b to $g(a)$ or $g(b)$.

3. Rewrite in terms of x.

Ex:

$$\int_{x=a}^{x=b} 8x(x^2+1)^3 dx$$

- Let $u = x^2 + 1$
- Then, $\frac{du}{dx} = 2x$
- And: $\frac{1}{2x}du = dx$. So we can rewrite the above function as:

$$\int_a^b 8xu^3 \cdot \frac{1}{2x}du$$

We simplify to change the limits to reflect $g(u)$.

$$\begin{aligned} &\int_{a^2+1}^{b^2+1} 4u^3 du \\ &= \int_{a^2+1}^{b^2+1} \frac{4}{4}u^4 = u^4\Big|_{a^2+1}^{b^2+1} \end{aligned}$$

Then we rewrite in terms of x:

$$(x^2+1)^4\Big|_a^b$$

9.3.1 Some examples

- $$\begin{aligned} &\int_a^b 3x^2\sqrt{x^3+1}dx \\ &\text{Let } u = x^3+1 \\ &\frac{du}{dx} = 3x^2,\ \frac{1}{3x^2}du = dx \\ &= \int_{a^3+1}^{b^3+1} 3x^2\sqrt{u}\frac{1}{3x^2}du \\ &= \int_{a^3+1}^{b^3+1} \sqrt{u}\,du = \int_{a^3+1}^{b^3+1} u^{\frac{1}{2}}du \\ &= \frac{2}{3}u^{\frac{3}{2}}\Big|_{a^3+1}^{b^3+1} = \frac{2}{3}(x^3+1)^{\frac{3}{2}}\Big|_a^b \end{aligned}$$

- $$\begin{aligned} &\int_a^b x^3e^{x^3}dx \\ &\text{Let } u = x^3, \frac{du}{dx} = 3x^2, dx = \frac{du}{3x^2} \\ &\int_{a^3}^{b^3} x^2e^u\frac{du}{3x^2} = \frac{1}{3}\int_{a^3}^{b^3} e^u du \\ &= \frac{1}{3}e^u\Big|_{a^3}^{b^3} = \frac{1}{3}e^{x^3}\Big|_a^b \end{aligned}$$

•

$$\begin{aligned}
&\int_a^b \frac{2-x}{\sqrt{2x^2-8x+1}}dx \\
&\text{Let } u = 2x^2 - 8x + 1 \\
&\frac{du}{dx} = 4x - 8, dx = \frac{du}{4x-8} \\
&\int (2-x)u^{-\frac{1}{2}} \frac{du}{-4(2-x)} = -\frac{1}{4}\int u^{-\frac{1}{2}} du \\
&= -\frac{1}{4} \cdot \frac{u^{\frac{1}{2}}}{\frac{1}{2}} \\
&- \frac{1}{2}u^{\frac{1}{2}}\Big|_a^b
\end{aligned}$$

9.4 Integration by parts

Sometimes we get functions where integrating by substitution does not gain us any leverage. For these, we can *integrate by parts*. Say we have two functions u and v, which are differentiable. We can use the product rule to show that:

$$\frac{d}{dx}(uv) = v\frac{du}{dx} + u\frac{dv}{dx}$$

And rearrange:

$$u\frac{dv}{dx} = \frac{d}{dx}(uv) - v\frac{du}{dx}$$

Integrate with respect to x, and we get the formula for integration by parts:

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

An example:

$$\begin{aligned}
&\int_a^b x^2 \ln x dx \\
&u = \ln x \Rightarrow du = \frac{1}{x}dx \\
&dv = x^2 dx \text{ Integrate } \Rightarrow v = \frac{x^3}{3} \\
&\text{So the original equation equals:} \\
&\frac{1}{3}x^3 \ln x - \int \frac{1}{3}x^3 \cdot \frac{1}{x} dx \\
&= \frac{1}{3}x^3 \ln x - \frac{1}{3}\int x^2 dx = \left[\frac{1}{3}x^3 \ln x - \frac{1}{3} \cdot \frac{1}{3}x^3\right]_a^b
\end{aligned}$$

One more:

$$\int_1^3 xe^x dx$$
$$u = x, du = dx, dv = e^x dx, v = e^x$$
$$\int_1^3 xe^x dx = xe^x - \int e^x dx$$
$$= [xe^x - e^x]_1^3$$

9.5 Improper integrals

What if we integrate to ∞ instead of a integer? We take the limit when we get to plugging our bounds in.

$$\int_1^\infty \frac{1}{\lambda} e^{-\lambda x} dx$$
$$= \frac{1}{\lambda} \int_1^\infty -\frac{1}{\lambda} e^{\lambda x} = \left[-\frac{1}{\lambda^2} e^{-\lambda x} \right]_1^3$$
$$= \left[\lim_{x \to \infty} -\frac{1}{\lambda^2} e^{-\lambda x} \right] - \left[-\frac{1}{\lambda^2} e^{-\lambda 1} \right]$$
$$= 0 + \left[\frac{1}{\lambda^2} e^{-\lambda} \right]$$

10 Calculus V

10.1 Double Integrals

Sometimes we want to integrate the area of a function depending on multiple variables. In this case we can take a double integral. We just take the integral of the function with respect to one variable. And then we integrate the outcome with respect to another variable:

$$
\begin{aligned}
&\int_{x_2=0}^{1}\int_{x_1=-1}^{1}(x_1^2+x_2^2)dx_1dx_2\\
=&\int_0^1\left[\int_{-1}^{1}(x_1^2+x_2^2)dx_1\right]dx_2\\
=&\int_0^1\left[\frac{1}{3}x_1^3+x_1x_2^2\right]_{-1}^{1}dx_2\\
=&\int_0^1\left\{\left[\frac{1}{3}1^3+1x_2^2\right]-\left[\frac{1}{3}-1^3+(-1)x_2^2\right]\right\}dx_2\\
=&\int_0^1\left(\frac{2}{3}+2x_2^2\right)dx_2\\
=&\left[\frac{2}{3}x_2+\frac{2}{3}x_2^3\right]_0^1\\
=&\frac{4}{3}
\end{aligned}
$$

References

W. H. Moore and D. A. Siegel. *A mathematics course for political and social research*. Princeton University Press, 2013.