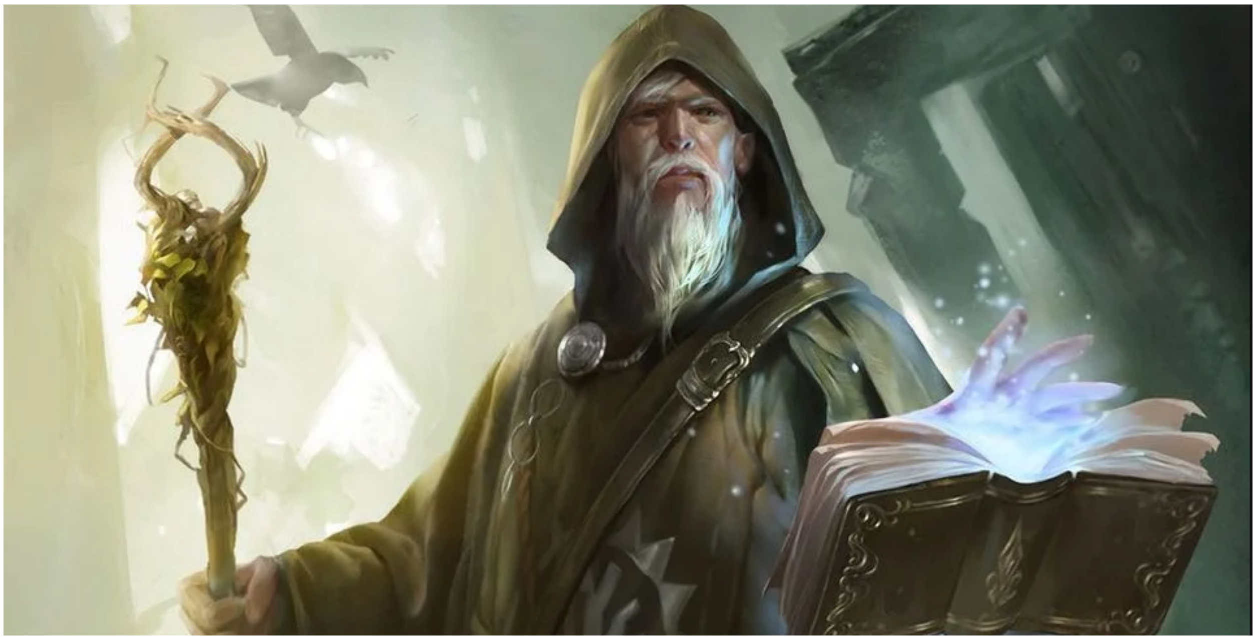

Notes: These are notes I took while studying for the Mathematics Subject GRE. There are likely a lot of errors and mistakes, please let me know if you find any!

Undergraduate Compendium

D. Zack Garza
University of Georgia
dzackgarza@gmail.com
Last updated: 2020-11-01

Contents

0.1 Differential Calculus

0.2 Big Theorems / Tools:

Proposition 0.2.1 ***(Fundamental Theorem of Calculus I).***

$$\frac{\partial}{\partial x}\int_a^x f(t)dt = f(x)$$

Proposition 0.2.2 ***(Generalized Fundamental Theorem of Calculus).***

$$\frac{\partial}{\partial x}\int_{a(x)}^{b(x)} f(x,t)dt - \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t)dt = f(x,t)\cdot \frac{\partial}{\partial x}(t)\Big|_{t=a(x)}^{t=b(x)}$$

$$= f(x,b(x))\cdot b'(x) - f(x,a(x))\cdot a'(x)$$

If $f(x,t) = f(t)$ doesn't depend on x, then $\frac{\partial f}{\partial x} = 0$ and the second integral vanishes:

$$\frac{\partial}{\partial x}\int_{a(x)}^{b(x)} f(t)dt = f(b(x))\cdot b'(x) - f(a(x))\cdot a'(x)$$

Find examples.

Remark 0.2.1.
Note that you can recover the original FTC by taking

$$\begin{aligned} a(x) &= c \\ b(x) &= x \\ f(x,t) &= f(t). \end{aligned}$$

Corollary 0.2.1 ***(?).***

$$\frac{\partial}{\partial x}\int_1^x f(x,t)dt = \int_1^x \frac{\partial}{\partial x} f(x,t)dt + f(x,x)$$

Proposition 0.2.3 ***(Extreme Value Theorem).***
Todo

Todo

Proposition 0.2.4 ***(Mean Value Theorem).***

$$\begin{aligned} f\in C^0(I) &\implies \exists p\in I: f(b)-f(a) = f'(p)(b-a) \\ &\implies \exists p\in I: \int_a^b f(x)\ dx = f(p)(b-a). \end{aligned}$$

Proposition 0.2.5 ***(Rolle's Theorem).***

todo

Proposition 0.2.6 ***(L'Hopital's Rule).***
If

- $f(x)$ and $g(x)$ are differentiable on $I - \{\mathrm{pt}\}$, and

$$\lim_{x \to \{\mathrm{pt}\}} f(x) = \lim_{x \to \{\mathrm{pt}\}} g(x) \in \{0, \pm\infty\}, \qquad \forall x \in I, g'(x) \neq 0, \qquad \lim_{x \to \{\mathrm{pt}\}} \frac{f'(x)}{g'(x)} \text{ exists,}$$

Then it is necessarily the case that

$$\lim_{x \to \{\mathrm{pt}\}} \frac{f(x)}{g(x)} = \lim_{x \to \{\mathrm{pt}\}} \frac{f'(x)}{g'(x)}.$$

Remark 0.2.2.
Note that this includes the following indeterminate forms:

$$\frac{0}{0}, \quad \frac{\infty}{\infty}, \quad 0 \cdot \infty, \quad 0^0, \quad \infty^0, \quad 1^\infty, \quad \infty - \infty.$$

For $0 \cdot \infty$, can rewrite as $\frac{0}{\frac{1}{\infty}} = \frac{0}{0}$, or alternatively $\frac{\infty}{\frac{1}{0}} = \frac{\infty}{\infty}$.

For $1^\infty, \infty^0$, and 0^0, set

$$L \coloneqq \lim f^g \implies \ln L = \lim g \ln(f)$$

to recover $\infty \cdot 0, 0 \cdot \infty$, or $0 \cdot 0$.

Proposition 0.2.7 ***(Taylor Expansion).***

$$\begin{aligned} T(a, x) &= \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n \\ &= f(a) + f'(a)(x-a) + \frac{1}{2} f''(a)(x-a)^2 \\ &\quad + \frac{1}{6} f'''(a)(x-a)^3 + \frac{1}{24} f^{(4)}(a)(x-a)^4 + \cdots \end{aligned}$$

There is a bound on the error:

$$|f(x) - T_k(a, x)| \leq \left| \frac{f^{(k+1)}(a)}{(k+1)!} \right|$$

where $T_k(a, x) = \sum_{n=0}^{k} \frac{f^{(n)}(a)}{n!}(x-a)^n$ is the kth truncation.

Remark 0.2.3.
Approximating change: $\Delta y \approx f'(x) \Delta x$

0.3 Differential

0.3.1 Limits

0.3.2 Tools for finding limits

Examples

How to find $\lim_{x\to a} f(x)$in order of difficulty:

- Plug in: if f is continuous, $\lim_{x\to a} f(x) = f(a)$.
- Check for indeterminate forms and apply L'Hopital's Rule.
- Algebraic rules
- Squeeze theorem
- Expand in Taylor series at a
- Monotonic + bounded
- One-sided limits: $\lim_{x\to a^-} f(x) = \lim_{\varepsilon\to 0} f(a-\varepsilon)$
- Limits at zero or infinity:

$$\lim_{x\to\infty} f(x) = \lim_{\frac{1}{x}\to 0} f\left(\frac{1}{x}\right) \text{ and } \lim_{x\to 0} f(x) = \lim_{x\to\infty} f\left(\frac{1}{x}\right)$$

  - Also useful: if $p(x) = p_n x^n + \cdots$ and $q(x) = q_n x^m + \cdots$,

$$\lim_{x\to\infty} \frac{p(x)}{q(x)} = \begin{cases} 0 & \deg p < \deg q \\ \infty & \deg p > \deg q \\ \frac{p_n}{q_n} & \deg p = \deg q \end{cases}$$

⚠ **Warning 0.1:** Be careful: limits may not exist!! Example : $\lim_{x\to 0} \frac{1}{x} \neq 0$.

0.3.3 Asymptotes

- Vertical asymptotes: at values $x = p$ where $\lim_{x\to p} = \pm\infty$
- Horizontal asymptotes: given by points $y = L$ where $L \lim_{x\to\pm\infty} f(x) < \infty$
- Oblique asymptotes: for rational functions, divide - terms without denominators yield equation of asymptote (i.e. look at the asymptotic order or "limiting behavior").
  - Concretely:

$$f(x) = \frac{p(x)}{q(x)} = r(x) + \frac{s(x)}{t(x)} \sim r(x)$$

0.3.4 Recurrences

- Limit of a recurrence: $x_n = f(x_{n-1}, x_{n-2}, \cdots)$
  - If the limit exists, it is a solution to $x = f(x)$

0.3.5 Derivatives

Proposition 0.3.1 ***(Chain Rule).***

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

Proposition 0.3.2 ***(Product Rule).***

$$\frac{\partial}{\partial x} f \cdot g = f' \cdot g + g' \cdot f$$

Proposition 0.3.3 ***(Quotient Rule).***

$$\frac{\partial}{\partial x} \frac{f(x)}{g(x)} = \frac{f'g - g'f}{g^2}$$

Mnemonic: Low d-high minus high d-low

Proposition 0.3.4 ***(Inverse Rule).***

$$\frac{\partial f^{-1}}{\partial x}(f(x_0)) = \left(\frac{\partial f}{\partial x}\right)^{-1}(x_0) = 1/f'(x_0)$$

0.3.6 Implicit Differentiation

$$(f(x))' = f'(x)\ dx, (f(y))' = f'(y)\ dy$$

- Often able to solve for $\frac{\partial y}{\partial x}$ this way.

- Obtaining derivatives of inverse functions: if $y = f^{-1}(x)$ then write $f(y) = x$ and implicitly differentiate.

0.3.7 Related Rates

General series of steps: want to know some unknown rate y_t

- Lay out known relation that involves y
- Take derivative implicitly (say w.r.t t) to obtain a relation between y_t and other stuff.
- Isolate $y_t =$ known stuff

Example 0.3.1 (?).
Example: ladder sliding down wall

- Setup: l, x_t and $x(t)$ are known for a given t, want y_t.

$$x(t)^2 + y(t)^2 = l^2 \implies 2xx_t + 2yy_t = 2ll_t = 0$$

(noting that l is constant)
  - So $y_t = -\frac{x(t)}{y(t)} x_t$
  - $x(t)$ is known, so obtain $y(t) = \sqrt{l^2 - x(t)^2}$ and solve.

0.4 Integral

0.4.1 Average Values

Proposition 0.4.1 ***(Integral formula for average value).***

$$\mu_f = \frac{1}{b-a} \int_a^b f(t)dt$$

Proof (?).
Apply MVT to $F(x)$. ∎

0.4.2 Area Between Curves

Area in polar coordinates:

$$A = \int_{r_1}^{r_2} \frac{1}{2} r^2(\theta)\ d\theta$$

0.4.3 Solids of Revolution

$$\begin{aligned} &\text{Disks} && A = \int \pi r(t)^2\ dt \\ &\text{Cylinders} && A = \int 2\pi r(t)h(t)\ dt. \end{aligned}$$

0.4.4 Arc Lengths

$$\begin{aligned} L &= \int\ ds & ds &= \sqrt{dx^2 + dy^2} \\ &= \int_{x_0}^{x_1} \sqrt{1 + \frac{\partial y}{\partial x}}\ dx \\ &= \int_{y_0}^{y_1} \sqrt{\frac{\partial x}{\partial y} + 1}\ dy \end{aligned}$$

$$SA = \int 2\pi r(x)\ ds$$

0.4.5 Center of Mass

Given a density $\rho(\mathbf{x})$ of an object R, the x_i coordinate is given by

$$x_i = \frac{\displaystyle\int_R x_i \rho(x)\ dx}{\displaystyle\int_R \rho(x)\ dx}$$

0.4.6 Big List of Integration Techniques

Given $f(x)$, we want to find an antiderivative $F(x) = \displaystyle\int f$ satisfying $\dfrac{\partial}{\partial x}F(x) = f(x)$

- Guess and check: look for a function that differentiates to f.
- u- substitution
  - More generally, any change of variables

$$x = g(u) \implies \int_a^b f(x)\ dx = \int_{g^{-1}(a)}^{g^{-1}(b)} (f \circ g)(x)\ g'(x)\ dx$$

Integration by Parts: The standard form:

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

- A more general form for repeated applications: let $v^{-1} = \displaystyle\int v$, $v^{-2} = \displaystyle\int\int v$, etc.

$$\begin{aligned}
\int_a^b uv &= uv^{-1}\Big|_a^b - \int_a^b u^1 v^{-1} \\
&= uv^{-1} - u^1 v^{-2}\Big|_a^b + \int_a^b u^2 v^{-2} \\
&= uv^{-1} - u^1 v^{-2} + u^2 v^{-3}\Big|_a^b - \int_a^b u^3 v^{-3} \\
&\ \ \vdots \\
\implies \int_a^b uv &= \sum_{k=1}^{n} (-1)^k u^{k-1} v^{-k}\Big|_a^b + (-1)^n \int_a^b u^n v^{-n}
\end{aligned}$$

- Generally useful when one term's nth derivative is a constant.

Shoelace Method

- Note: you can choose u or v equal to 1! Useful if you know the derivative of the integrand.

Derivatives	Integrals	Signs	Result
u	v	NA	NA
u'	$\int v$	$+$	$u \int v$
u''	$\int\int v$	$-$	$-u' \int\int v$
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Fill out until one column is zero (alternate signs). Get the result column by multiplying diagonally, then sum down the column.

Differentiating under the integral

$$\begin{aligned}\frac{\partial}{\partial x}\int_{a(x)}^{b(x)} f(x,t)dt - \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t)dt &= f(x,\cdot)\frac{\partial}{\partial x}(\cdot)\Big|_{a(x)}^{b(x)} \\ &= f(x,b(x))\ b'(x) - f(x,a(x))\ a'(x)\end{aligned}$$

Proof (?).
Let $F(x)$ be an antiderivative and compute $F'(x)$ using the chain rule.

For constants, this should allow differentiating under the integral when f, f_x are "jointly continuous"

- LIPET: Log, Inverse trig, Polynomial, Exponential, Trig: generally let u be whichever one comes first.
- The ridiculous trig sub: for any integrand containing only trig terms
  - Transforms *any* such integrand into a rational function of x
  - Let $u = 2\tan^{-1} x,\ du = \dfrac{2}{x^2+1}$, then

$$\int_a^b f(x)\ dx = \int_{\tan\frac{a}{2}}^{\tan\frac{b}{2}} f(u)\ du$$

■

Example 0.4.1 (?).

$$\int_0^{\pi/2} \frac{1}{\sin\theta}\ d\theta = 1/2$$

- Trigonometric Substitution

$$\begin{aligned}\sqrt{a^2-x^2} &\Rightarrow & x &= a\sin(\theta) & dx &= a\cos(\theta)\ d\theta \\ \sqrt{a^2+x^2} &\Rightarrow & x &= a\tan(\theta) & dx &= a\sec^2(\theta)\ d\theta \\ \sqrt{x^2-a^2} &\Rightarrow & x &= a\sec(\theta) & dx &= a\sec(\theta)\tan(\theta)\ d\theta\end{aligned}$$

Partial Fractions

Trigonometric Substitution

Completing the square

- Trig Formulas

$$\begin{aligned} \sin^2(x) &= \frac{1}{2}(1 - 2\cos x) \\ &= \\ &= \\ &= \\ &= \end{aligned}$$

Trig functions, double angle formulas.

- Products of trig functions
  - Setup: $\int \sin^a(x)\cos^b(x)\ dx$
    - ♢ Both a, b even: $\sin(x)\cos(x) = \frac{1}{2}\sin(x)$
    - ♢ a odd: $\sin^2 = 1 - \cos^2,\ u = \cos(x)$
    - ♢ b odd: $\cos^2 = 1 - \sin^2,\ u = \sin(x)$
  - Setup: $\int \tan^a(x)\sec^b(x)\ dx$
    - ♢ a odd: $\tan^2 = \sec^2 - 1,\ u = \sec(x)$
    - ♢ b even: $\sec^2 = \tan^2 - 1, u = \tan(x)$

Other small but useful facts:

$$\int_0^{2\pi} \sin\theta\ d\theta = \int_0^{2\pi} \cos\theta\ d\theta = 0.$$

0.4.7 Optimization

- Critical points: boundary points and wherever $f'(x) = 0$
- Second derivative test:
  - $f''(p) > 0 \implies p$ is a min
  - $f''(p) < 0 \implies p$ is a max
- Inflection points of h occur where the *tangent* of h' changes sign. (Note that this is where h' itself changes sign.)
- Inverse function theorem: The slope of the inverse is reciprocal of the original slope
- If two equations are equal at exactly one real point, they are tangent to each other there - therefore their derivatives are equal. Find the x that satisfies this; it can be used in the original equation.

- Fundamental theorem of Calculus: If

$$\int f(x)dx = F(b) - F(a) \implies F'(x) = f(x).$$

- Min/maxing - either derivatives of Lagranage multipliers!
- Distance from origin to plane: equation of a plane

$$P : ax + by + cz = d.$$

  - You can always just read off the normal vector $\mathbf{n} = (a, b, c)$. So we have $\mathbf{nx} = d$.
  - Since $\lambda\mathbf{n}$ is normal to P for all λ, solve $\mathbf{n}\lambda\mathbf{n} = d$, which is $\lambda = \dfrac{d}{\|\mathbf{n}\|^2}$
- A plane can be constructed from a point p and a normal n by the equation $np = 0$.
- In a sine wave $f(x) = \sin(\omega x)$, the period is given by $2\pi/\omega$. If $\omega > 1$, then the wave makes exactly ω full oscillations in the interval $[0, 2\pi]$.
- The directional derivative is the gradient dotted against a *unit vector* in the direction of interest
- Related rates problems can often be solved via implicit differentiation of some constraint function
- The second derivative of a parametric equation is not exactly what you'd intuitively think!
- For the love of god, remember the FTC!

$$\frac{\partial}{\partial x}\int_0^x f(y)dy = f(x)$$

- Technique for asymptotic inequalities: WTS $f < g$, so show $f(x_0) < g(x_0)$ at a point and then show $\forall x > x_0, f'(x) < g'(x)$. Good for big-O style problems too.
- Inflection points of h occur where the *tangent* of h' changes sign. (Note that this is where h' itself changes sign.)
- Inverse function theorem: The slope of the inverse is reciprocal of the original slope
- If two equations are equal at exactly one real point, they are tangent to each other there - therefore their derivatives are equal. Find the x that satisfies this; it can be used in the original equation.

- Fundamental theorem of Calculus: If

$$\int f(x)dx = F(b) - F(a) \implies F'(x) = f(x).$$

- Min/maxing - either derivatives of Lagranage multipliers!
- Distance from origin to plane: equation of a plane

$$P : ax + by + cz = d.$$

1 Sequences

Notation: $\{a_n\}_{n\in\mathbb{N}}$ is a **sequence**, $\sum\limits_{i\in\mathbb{N}} a_i$ is a **series**.

1.1 Known Examples

- Known sequences: let c be a constant.

$$c, c^2, c^3, \ldots = \{c^n\}_{n=1}^{\infty} \to 0 \qquad \forall |c| < 1$$

$$\frac{1}{c}, \frac{1}{c^2}, \frac{1}{c^3}, \ldots = \left\{\frac{1}{c^n}\right\}_{n=1}^{\infty} \to 0 \qquad \forall |c| > 1$$

$$1, \frac{1}{2^c}, \frac{1}{3^c}, \ldots = \left\{\frac{1}{n^c}\right\}_{n=1}^{\infty} \to 0 \qquad \forall c > 0$$

1.2 Convergence

Definition 1.2.1 (Convergence of a Sequence)**.**
A sequence $\{x_j\}$ **converges** to L iff

$$\forall \varepsilon > 0, \exists N > 0 \text{ such that } \quad n \geq N \implies |x_n - L| < \varepsilon.$$

Theorem 1.2.1 ***(Squeeze Theorem).***

$$b_n \leq a_n \leq c_n \text{ and } b_n, c_n \to L \implies a_n \to L$$

Theorem 1.2.2 ***(Monotone Convergence Theorem for Sequences).***
If $\{a_j\}$ monotone and bounded, then $a_j \to L = \limsup a_i < \infty$.

Theorem 1.2.3 ***(Cauchy Criteria).***
$|a_m - a_n| \to 0 \in \mathbb{R} \implies \{a_i\}$ converges.

1.2.1 Checklist

- Is the sequence bounded?
  - $\{a_i\}$ not bounded $\implies$ not convergent
  - If bounded, is it monotone?
    - ◇ $\{a_i\}$ bounded and monotone $\implies$ convergent
- Use algebraic properties of limits
- Epsilon-delta definition

- Algebraic properties and manipulation:
  - Limits commute with $\pm, \times$, Div and $\lim C = C$ for constants.
    - ♢ E.g. Divide all terms by n before taking limit
    - ♢ Clear denominators

2 | Sums ("Series")

Definition 2.0.1 (Series)**.**
A **series** is an function of the form

$$f(x) = \sum_{j=1}^{\infty} c_j x^j.$$

2.1 Known Examples

2.1.1 Conditionally Convergent

$$\begin{aligned}
\sum_{k=1}^{\infty} k^p < \infty & \qquad \iff p \leq 1 \\
\sum_{k=1}^{\infty} \frac{1}{k^p} < \infty & \qquad \iff p > 1 \\
\sum_{k=1}^{\infty} \frac{1}{k} = \infty &
\end{aligned}$$

2.1.2 Convergent

$$\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$
$$\sum_{n=1}^{\infty} \frac{1}{n^3} < \infty$$
$$\sum_{n=1}^{\infty} \frac{1}{n^{\frac{3}{2}}} < \infty$$
$$\sum_{n=1}^{\infty} \frac{1}{n!} = e$$
$$\sum_{n=1}^{\infty} \frac{1}{c^n} = \frac{c}{c-1}$$
$$\sum_{n=1}^{\infty} (-1)^n \frac{1}{c^n} = \frac{c}{c+1}$$
$$\sum_{n=1}^{\infty} (-1)^n \frac{1}{n} = \ln 2$$

2.1.3 Divergent

$$\sum_{n=1}^{\infty} \frac{1}{n} = \infty$$
$$\sum_{n=1}^{\infty} \frac{1}{\sqrt{n}} = \infty$$

2.2 Convergence

Useful reference: `http://math.hawaii.edu/~ralph/Classes/242/SeriesConvTests.pdf`

Definition 2.2.1 (Absolutely Convergent)**.**

todo

Remark 2.2.1.
$a_n \to 0$ does not imply $\sum a_n < \infty$. Counterexample: the harmonic series.

Proposition 2.2.1 ***(?).***
Absolute convergence $\implies$ convergence

Proposition 2.2.2 ***(The Cauchy Criterion).***

$$\limsup a_i \to 0 \implies \sum a_i \text{ converges}$$

2.2.1 The Big Tests

Theorem 2.2.1 ***(Comparison Test).***

- $a_n < b_n \sum b_n < \infty \implies \sum a_n < \infty$
- $b_n < a_n \sum b_n = \infty \implies \sum a_n = \infty$

Theorem 2.2.2 ***(Ratio Test).***

$$R = \lim_{n\to\infty} \left| \frac{a_{n+1}}{a_n} \right|$$

- $R < 1$: absolutely convergent
- $R > 1$: divergent
- $R = 1$: inconclusive

Theorem 2.2.3 ***(Root Test).***

$$R = \limsup_{n\to\infty} \sqrt[n]{|a_n|}$$

- $R < 1$: convergent - $R > 1$: divergent - $R = 1$: inconclusive

Theorem 2.2.4 ***(Integral Test).***

$$f(n) = a_n \implies \sum a_n < \infty \iff \int_1^\infty f(x)dx < \infty$$

Theorem 2.2.5 ***(Limit Test).***

$$\lim_{n\to\infty} \frac{a_n}{b_n} = L < \infty \implies \sum a_n < \infty \iff \sum b_n < \infty$$

Theorem 2.2.6 ***(Alternating Series Test).***

$$a_n \downarrow 0 \implies \sum (-1)^n a_n < \infty$$

Theorem 2.2.7 ***(Weierstrass M-Test).***

$$\sum_{n=1}^{\infty} \|f_n\|_\infty < \infty \implies \exists f \text{ such that } \left\| \sum_{n=1}^{\infty} f_n - f \right\|_\infty \to 0$$

In other words, the series converges uniformly.

Slogan: Convergence of the sup norms implies uniform convergence"

Remark 2.2.2.
The M in the name comes from defining $\sup\{f_k(x)\} \coloneqq M_n$ and requiring $\sum |M_n| < \infty$.

2.2.2 Checklist

- Do the terms tend to zero?
  - $a_i \not\to 0 \implies \sum a_i = \infty$.
    - ♢ Can check with L'Hopital's rule
- There are exactly 6 tests at our disposal:
  - Comparison, root, ratio, integral, limit, alternating
- Is the series alternating?
  - If so, does $a_n \downarrow 0$?
    - ♢ If so, **convergent**
- Is this series bounded above by a known convergent series?
  - p series with $p > 1$, i.e. : $\sum a_n \leq \sum \frac{1}{n^p} < \infty$
  - Geometric series with $|x| < 1$, i.e. $\sum a_n \leq \sum x^n$
- Is this series bounded below by a known divergent series?
  - p series with $p \leq 1$, i.e. $\infty = \sum \frac{1}{n^p} \leq \sum a_i$
- Are the ratios strictly less than or greater than 1?
  - $< 1 \implies$ **convergent**
  - $> 1 \implies$ **convergent**
- Does the integral analogue converge?
  - Integral converges $\iff$ sum converges
- Try the root test
  - $< 1 \implies$ **convergent**
  - $> 1 \implies$ **convergent**
- Try the limit test
  - Attempt to divide each term to obtain a known convergent/divergent series

Some Pattern Recognition:

- (stuff)!: Ratio test (only test that will work with factorials!!)
- (stuff)^n: Root test or ratio test
- Replace a_n with an $f(x)$ that's easy to integrate - integral test
- $p(x)$ or $\sqrt{p(x)}$: comparison or limit test

2.3 Radius of Convergence

Proposition 2.3.1 ***(Finding the radius of convergence).***
Use the fact that

$$\lim_{k\to\infty}\left|\frac{a_{k+1}x^{k+1}}{a_k x^k}\right| = |x| \lim_{k\to\infty}\left|\frac{a_{k+1}}{a_k}\right| < 1 \implies \sum a_k x^k < \infty,$$

so take $L := \lim\limits_{k\to\infty}\dfrac{a_{k+1}}{a_k}$ and then obtain the radius as

$$R = \frac{1}{L} = \lim_{k\to\infty}\frac{a_k}{a_{k+1}}$$

Remark 2.3.1.

- Note $L = 0 \implies$ absolutely convergent everywhere
- $L = \infty \implies$ convergent only at $x = 0$.
- Also need to check endpoints $R, -R$ manually.

3 | Vector Calculus

Need lots of pictures

Notation:

$$\begin{aligned}
\mathbf{v}, \mathbf{a}, \cdots & \qquad \text{vectors in } \mathbb{R}^n \\
\mathbf{R}, \mathbf{A}, \cdots & \qquad \text{matrices} \\
\mathbf{r}(t) & \qquad \text{A parameterized curve } \mathbf{r} : \mathbb{R} \to \mathbb{R}^n \\
\widehat{\mathbf{v}} & \qquad \frac{\mathbf{v}}{\|\mathbf{v}\|}.
\end{aligned}$$

3.1 Plane Geometry

Proposition 3.1.1 ***(Slope of a vector in $\mathbb{R}^2$).***

$$\mathbf{v} = [x, y] \in \mathbb{R}^2 \implies m = \frac{y}{x}.$$

Proposition 3.1.2 ***(Rotation matrices in $\mathbb{R}^2$).***

$$\mathbf{R}_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \implies \mathbf{R}_{\frac{\pi}{2}} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Corollary 3.1.1 *(?).*

$$\mathbf{R}_{\frac{\pi}{2}}\mathbf{x} \coloneqq \mathbf{R}_{\frac{\pi}{2}}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -y \\ x \end{bmatrix} \in \mathbb{R}\mathbf{x}^{\perp}.$$

Thus if a planar line is defined by the span of $[x, y]$ and a slope of $m = y/x$, a normal vector is given by the span of $[-y, x]$ of slope $-\frac{1}{m} = -x/y$.

Example 3.1.1 (?).
Given $\mathbf{v}$, the rotated vector $\mathbf{R}_{\frac{\pi}{2}}\mathbf{v}$ is orthogonal to $\mathbf{v}$, so this can be used to obtain normals and other orthogonal vectors in the plane.

Proposition 3.1.3.
There is a direct way to come up with one orthogonal vector to any given vector:

$$\mathbf{v} = [a, b, c] \implies \mathbf{y} \coloneqq \begin{cases} [-(b+c), a, a] & \mathbf{v} = [-1, -1, 0], \\ [c, c, -(a+b)] & \text{else} \end{cases} \in \mathbb{R}\mathbf{v}^{\perp}.$$

3.2 Projections

For a subspace given by a single vector $\mathbf{a}$:

$$\operatorname{proj}_{\mathbf{a}}(\mathbf{x}) = \langle \mathbf{x}, \ \mathbf{a} \rangle \hat{\mathbf{a}} \qquad \operatorname{proj}_{\mathbf{a}}^{\perp}(\mathbf{x}) = \mathbf{x} - \operatorname{proj}_{\mathbf{a}}(\mathbf{x}) = \mathbf{x} - \langle \mathbf{x}, \ \mathbf{a} \rangle \hat{\mathbf{a}}$$

In general, for a subspace $\operatorname{colspace}(A) = \{\mathbf{a}_1, \cdots \mathbf{a}_n\}$,

$$\operatorname{proj}_A(\mathbf{x}) = \sum_{i=1}^{n} \langle \mathbf{x}, \ \mathbf{a}_i \rangle \hat{\mathbf{a}}_{\mathbf{i}} = A(A^T A)^{-1} A^T \mathbf{x}$$

3.3 Lines

General Equation	$Ax + By + C = 0$
Parametric Equation	$\mathbf{r}(t) = t\mathbf{x} + \mathbf{b}.$

Characterized by an equation in inner products:

$$\mathbf{y} \in L \iff \langle \mathbf{y}, \ \mathbf{n} \rangle = 0$$

Proposition 3.3.1 ***(Equation for a line between two points).***
Given $\mathbf{p}_0, \mathbf{p}_1$, take $\mathbf{x} = \mathbf{p}_1 - \mathbf{p}_0$ and $\mathbf{b} = \mathbf{p}_i$ for either i:

$$\mathbf{r}(t) = t(\mathbf{p}_1 - \mathbf{p}_0) + \mathbf{p}_0 \qquad = t\mathbf{p}_1 + (1-t)\mathbf{p}_0.$$

Proposition 3.3.2 ***(Symmetric equation of a line).***
If a line L is given by

$$\mathbf{r}(t) = t[x_1, x_2, x_3] + [p_1, p_2, p_3],$$

then

$$(x, y, z) \in L \iff \frac{x - p_1}{x_1} = \frac{y - p_2}{x_2} = \frac{z - p_3}{x_3}.$$

Example 3.3.1 (?)**.**
The symmetric equation of the line through $[2, 1, -3]$ and $[1, 4, -3]$ is given by

$$\frac{x-2}{1} = \frac{y+1}{-5} = \frac{z-3}{6}.$$

3.3.1 Tangent Lines / Planes

Key idea: just need a point and a normal vector, and the gradient is normal to level sets.

Theorem 3.3.1 ***(The Tangent Plane Equation).***
For any locus $f(\mathbf{x}) = 0$, we have

$$\mathbf{x} \in T_f(\mathbf{p}) \implies \langle \nabla f(\mathbf{p}),\ \mathbf{x} - \mathbf{p} \rangle = 0.$$

3.3.2 Normal Lines

Key idea: the gradient is normal.

To find a normal line, you just need a single point $\mathbf{p}$ and a normal vector $\mathbf{n}$; then

$$L = \left\{ \mathbf{x} \;\middle|\; \mathbf{x} = \mathbf{p} + t\mathbf{v} \right\}.$$

3.4 Planes

$$\begin{aligned} \text{General Equation} \qquad & Ax + By + Cz + D = 0 \\ \text{Parametric Equation} \qquad & \mathbf{y}(t, s) = t\mathbf{x}_1 + s\mathbf{x}_2 + \mathbf{b} \end{aligned}$$

.

Characterized by an equation in inner products:

$$\mathbf{y} \in P \iff \langle \mathbf{y} - \mathbf{p}_0,\ \mathbf{n} \rangle = 0$$

Proposition 3.4.1 ***(Writing equation from a point and a normal).***
Determined by a point $\mathbf{p}_0$ and a normal vector $\mathbf{n}$

Proposition 3.4.2 ***(Writing equation from two vectors).***
Given $\mathbf{v}_0, \mathbf{v}_1$, set $\mathbf{n} = \mathbf{v}_0 \times \mathbf{v}_1$.

3.4.1 Finding a Normal Vector

- **Normal vector to a plane**
  - Can read normal off of equation: $\mathbf{n} = [a, b, c]$
- Computing D:
  - $D = \langle \mathbf{p}_0,\ \mathbf{n} \rangle = p_1 n_1 + p_2 n_2 + p_3 n_3$
  - Useful trick: once you have $\mathbf{n}$, you can let $\mathbf{p}_0$ be *any* point in the plane (don't necessarily need to use the one you started with, so pick any point that's convenient to calculate)

3.4.2 Distance from origin to plane

- Given by $D/\|\mathbf{n}\| = \langle \mathbf{p}_0,\ \widehat{\mathbf{n}} \rangle$. Gives a signed distance.

Distance from origin to plane.

3.4.3 Distance from point to plane

- Given by $\langle \cdot,\ \widehat{\mathbf{n}} \rangle$
- Finding vectors in the plane
- Given $P = [A, B, C] \cdot [x, y, z] = 0$, can take $\left[-\frac{B}{A}, 1, 0\right], \left[-\frac{C}{A}, 0, 1\right]$

Distance from point to plane.

3.5 Curves

$$\mathbf{r}(t) = [x(t), y(t), z(t)].$$

3.5.1 Tangent line to a curve

We have an equation for the tangent vector at each point:

$$\widehat{\mathbf{T}}(t) = \widehat{\mathbf{r}'}(t),$$

so we can write

$$\mathbf{L}_T(t) = \mathbf{r}(t_0) + t\widehat{\mathbf{T}}(t_0) := \mathbf{r}(t_0) + t\widehat{\mathbf{r}'}(t_0).$$

3.5.2 Normal line to a curve

- Use the fact that $\mathbf{r}''(t) \in \mathbb{R}\mathbf{r}'(t)^{\perp}$, so we have an equation for a normal vector at each point:

$$\widehat{\mathbf{N}}(t) = \widehat{\mathbf{r}''}(t).$$

Thus we can write

$$\mathbf{L}_N(t) = \mathbf{r}(t_0) + t\widehat{\mathbf{N}}(t_0) = \mathbf{r}(t_0) + t\widehat{\mathbf{r}''}(t_0).$$

Special case: planar graphs of functions Suppose $y = f(x)$. Set $g(x, y) = f(x) - y$, then

$$\nabla g = [f_x(x), -1] \implies m = -\frac{1}{f_x(x)}$$

3.6 Minimal Distances

Fix a point $\mathbf{p}$. Key idea: find a subspace and project onto it.

Key equations: projection and orthogonal projection of $\mathbf{b}$ onto $\mathbf{a}$:

$$\mathrm{proj}_{\mathbf{a}}(\mathbf{b}) = \langle \mathbf{b},\ \mathbf{a} \rangle \widehat{\mathbf{a}} \qquad \mathrm{proj}_{\mathbf{a}}^{\perp}(\mathbf{b}) = \mathbf{b} - \mathrm{proj}_{\mathbf{a}}(\mathbf{b}) = \mathbf{b} - \langle \mathbf{b},\ \mathbf{a} \rangle \widehat{\mathbf{a}}$$

3.6.1 Point to plane

- Given a point $\mathbf{p}$ and a plane $S = \left\{ \mathbf{x} \in \mathbb{R}^3 \;\middle|\; n_0 x + n_1 y + n_2 z = d \right\}$, let $\mathbf{n} = [n_1, n_2, n_3]$, find any point $\mathbf{q} \in S$, and project $\mathbf{q} - \mathbf{p}$ onto $S^{\perp} = \mathrm{Span}(\mathbf{n})$ using

$$d = \|\mathrm{proj}_{\mathbf{n}}(\mathbf{q} - \mathbf{p})\| = \|\langle \mathbf{q} - \mathbf{p},\ \mathbf{n} \rangle \widehat{\mathbf{n}}\| = \langle \mathbf{q} - \mathbf{p},\ \mathbf{n} \rangle.$$

- Given just two vectors $\mathbf{u}, \mathbf{v}$: manufacture a normal vector $\mathbf{n} = \mathbf{u} \times \mathbf{v}$ and continue as above.

Origin to plane Special case: if $\mathbf{p} = \mathbf{0}$,

$$d = \|\mathrm{proj}_{\mathbf{n}}(\mathbf{q})\| = \|\langle \mathbf{p},\ \mathbf{n} \rangle \widehat{\mathbf{n}}\| = \langle \mathbf{p},\ \mathbf{n} \rangle..$$

3.6.2 Point to line

- Given a line $L : \mathbf{x}(t) = t\mathbf{v}$ for some fixed $\mathbf{v}$, use

$$d = \left\| \mathrm{proj}_{\mathbf{v}}^{\perp}(\mathbf{p}) \right\| = \|\mathbf{p} - \langle \mathbf{p},\ \mathbf{v} \rangle \widehat{\mathbf{v}}\|.$$

- Given a line $L : \mathbf{x}(t) = \mathbf{w}_0 + t\mathbf{w}$, let $\mathbf{v} = \mathbf{x}(1) - \mathbf{x}(0)$ and proceed as above.

3.6.3 Point to curve

todo

3.6.4 Line to line

Given $\mathbf{r}_1(t) = \mathbf{p}_1 + t\mathbf{v}_2$ and $\mathbf{r}_2(t) = \mathbf{p}_2 + t\mathbf{v}_2$, let d be the desired distance.

- Let $\widehat{\mathbf{n}} = \widehat{\mathbf{v}_1 \times \mathbf{v}_2}$, which is orthogonal to both lines.
- Then project the vector connecting the two fixed points $\mathbf{p}_i$ onto this subspace and take its norm:

$$\begin{aligned} d &= \|\mathrm{proj}_{\mathbf{n}}(\mathbf{p}_2 - \mathbf{p}_1)\| \\ &= \|\langle \mathbf{p}_2 - \mathbf{p}_1,\ \mathbf{n}\rangle \widehat{\mathbf{n}}\| \\ &= \langle \mathbf{p}_2 - \mathbf{p}_1,\ \mathbf{n}\rangle \\ &:= \langle \mathbf{p}_2 - \mathbf{p}_1,\ \mathbf{v}_1 \times \mathbf{v}_2\rangle. \end{aligned}$$

3.7 Surfaces

$$S = \left\{(x, y, z) \,\middle|\, f(x, y, z) = 0\right\} \qquad\qquad z = f(x, y)$$

3.7.1 Tangent plane to a surface

- Need a point $\mathbf{p}$ and a normal $\mathbf{n}$. By cases:
- $f(x, y, z) = 0$
  - ∇f is a normal vector.
  - Write the tangent plane equation $\langle \mathbf{n},\ \mathbf{x} - \mathbf{p}_0\rangle$, done.
- $z = g(x, y)$:
  - Let $f(x, y, z) = g(x, y) - z$, then $\mathbf{p} \in S \iff \mathbf{p}$ is in a level set of f.
  - ∇f is normal to level sets (and thus the surface), so compute $\nabla f = [g_x, g_y, -1]$
  - Proceed as in previous case

3.7.2 Surfaces of revolution

- Given $f(x_1, x_2) = 0$, can be revolved around either the x_1 or x_2 axis.
  - $f(x, y)$ around the x axis yields $f(x, \pm\sqrt{y^2 + z^2}) = 0$
  - $f(x, y)$ around the y axis yields $f(\pm\sqrt{x^2 + z^2}, y) = 0$
  - Remaining cases proceed similarly - leave the axis variable alone, replace other variable with square root involving missing axis.
- Equations of lines tangent to an intersection of surfaces $f(x, y, z) = g(x, y, z)$:
  - Find two normal vectors and take their cross product, e.g. $n = \nabla f \times \nabla g$, then

$$L = \left\{\mathbf{x} \,\middle|\, \mathbf{x} = \mathbf{p} + t\mathbf{n}\right\}$$

- Level curves:
  - Given a surface $f(x, y, z) = 0$, the level curves are obtained by looking at e.g. $f(x, y, c) = 0$.

4 Multivariable Calculus

Theorem 4.0.1 ***(Key Theorem).***
Given a function $f : \mathbb{R}^n \to \mathbb{R}$, let $S_k := \left\{ \mathbf{p} \in \mathbb{R}^n \;\middle|\; f(\mathbf{p}) = k \right\}$ denote the level set for $k \in \mathbb{R}$. Then

$$\nabla f(\mathbf{p}) \in S_k^{\perp}.$$

4.1 Notation

$$\mathbf{v} = [v_1, v_2, \cdots] \qquad \text{a vector}$$

$$\mathbf{e}_i = [0, 0, \cdots, \overbrace{1}^{i\text{th term}}, \cdots, 0] \qquad \text{the } i\text{th standard basis vector}$$

$$\begin{aligned} \varphi &: \mathbb{R}^n \to \mathbb{R} \\ \varphi(x_1, x_2, \cdots) &= \cdots \end{aligned} \qquad \text{a functional on } \mathbb{R}^n$$

$$\begin{aligned} \mathbf{F} &: \mathbb{R}^n \to \mathbb{R}^n \\ \mathbf{F}(x_1, x_2, \cdots) &= [\mathbf{F}_1(x_1, x_2, \cdots), \mathbf{F}_2(x_1, x_2, \cdots), \cdots, \mathbf{F}_n(x_1, x_2, \cdots)] \end{aligned} \qquad \text{a multivariable function}$$

4.2 Partial Derivatives

Definition 4.2.1 (Partial Derivative)**.**
For a functional $f : \mathbb{R}^n \to \mathbb{R}$, the **partial derivative** of f with respect to x_i is

$$\frac{\partial f}{\partial x_i}(\mathbf{p}) := \lim_{h \to 0} \frac{f(\mathbf{p} + h\mathbf{e}_i) - f(\mathbf{p})}{h}$$

Example 4.2.1 $(n = 2)$**.**

$$\begin{aligned} f &: \mathbb{R}^2 \to \mathbb{R} \\ \frac{\partial f}{\partial x}(x_0, y_0) &= \lim_{h \to 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h} \end{aligned}$$

4.3 General Derivatives

Definition 4.3.1 (General definition of differentiability)**.**
A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is **differentiable** iff there exists a linear transformation $D_f : \mathbb{R}^n \to \mathbb{R}^m$

such that the following limit exists

$$\lim_{\mathbf{x}\to\mathbf{p}} \frac{\|f(\mathbf{x}) - f(\mathbf{p}) - D_f(\mathbf{x}-\mathbf{p})\|}{\|\mathbf{x}-\mathbf{p}\|} = 0.$$

Remark 4.3.1.
D_f is the "best linear approximation" to f.

Definition 4.3.2 (Jacobian)**.**
When f is differentiable, D_f can be given in coordinates by

$$(D_f)_{ij} = \frac{\partial f_i}{\partial x_j}$$

This yields the **Jacobian** of f:

$$D_f(p) \begin{bmatrix} | & | & & | \\ \nabla f_1(\mathbf{p}) & \nabla f_2(\mathbf{p}) & \cdots & \nabla f_m(\mathbf{p}) \\ | & | & & | \end{bmatrix}^T = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{p}) & \frac{\partial f_1}{\partial x_2}(\mathbf{p}) & \cdots & \frac{\partial f_1}{\partial x_n}(\mathbf{p}) \\ \frac{\partial f_2}{\partial x_1}(\mathbf{p}) & \frac{\partial f_2}{\partial x_2}(\mathbf{p}) & \cdots & \frac{\partial f_2}{\partial x_n}(\mathbf{p}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{p}) & \frac{\partial f_m}{\partial x_2}(\mathbf{p}) & \cdots & \frac{\partial f_m}{\partial x_n}(\mathbf{p}) \end{bmatrix}.$$

Remark 4.3.2.
This is equivalent to

- Taking the gradient of each component f_i of f,
- Evaluating ∇f_i at $\mathbf{p}$,
- Forming a matrix using these as the columns, and
- Transposing the resulting matrix.

Definition 4.3.3 (Hessian)**.**
For a function $f : \mathbb{R}^n \to \mathbb{R}$, the **Hessian** is a generalization of the second derivative, and is given in coordinates by

$$(H_f)_{ij} = \frac{\partial^2 f}{\partial x_i x_j}$$

Explicitly, we have

$$H_f(\mathbf{p}) = \begin{bmatrix} | & | & & | \\ D\nabla f_1(\mathbf{p}) & D\nabla f_2(\mathbf{p}) & \cdots & D\nabla f_m(\mathbf{p}) \\ | & | & & | \end{bmatrix}^T = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{a}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{a}) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n}(\mathbf{a}) \end{bmatrix}.$$

Remark 4.3.3.
Mnemonic: make matrix with ∇f as the columns, and then differentiate variables left to right.

4.4 The Chain Rule

Example 4.4.1 (How to expand a partial derivative)**.**
Write out tree of dependent variables:

$$
\begin{tikzcd}
 & u \arrow[r] \arrow[dd] \arrow[ddr] & x \\
z \arrow[ur] \arrow[urr] \arrow[dr] \arrow[drr] & & \\
 & v \arrow[r] \arrow[uur] & y
\end{tikzcd}
$$

Then sum each possible path.

Let subscripts denote which variables are held constant, then

$$
\begin{aligned}
\left(\frac{\partial z}{\partial x}\right)_y = &\left(\frac{\partial z}{\partial x}\right)_{u,y,v} \\
&+ \left(\frac{\partial z}{\partial v}\right)_{x,y,u}\left(\frac{\partial v}{\partial x}\right)_y \\
&+ \left(\frac{\partial z}{\partial u}\right)_{x,y,v}\left(\frac{\partial u}{\partial x}\right)_{v,y} \\
&+ \left(\frac{\partial z}{\partial u}\right)_{x,y,v}\left(\frac{\partial u}{\partial v}\right)_{x,y}\left(\frac{\partial v}{\partial x}\right)_y
\end{aligned}
$$

4.5 Approximation

Let $z = f(x, y)$, then to approximate near $\mathbf{p}_0 = [x_0, y_0]$,

$$
\begin{aligned}
f(\mathbf{x}) &\approx f(\mathbf{p}) + \nabla f(\mathbf{x} - \mathbf{p}_0) \\
\implies f(x, y) &\approx f(\mathbf{p}) + f_x(\mathbf{p})(x - x_0) + f_y(\mathbf{p})(y - y_0)
\end{aligned}
$$

.

4.6 Optimization

4.6.1 Classifying Critical Points

Definition 4.6.1 (Critical Points)**.**
Critical points of f given by points $\mathbf{p}$ such that the derivative vanishes:

$$\text{crit}(f) = \left\{ \mathbf{p} \in \mathbb{R}^n \;\middle|\; D_f(\mathbf{p}) = 0 \right\}$$

Proposition 4.6.1 ***(Second Derivative Test).***

1. Compute

$$|H_f(\mathbf{p})| := \begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix} (\mathbf{p})$$

2. Check by cases:
   - $|H(\mathbf{p})| = 0$: No conclusion
   - $|H(\mathbf{p})| < 0$: Saddle point
   - $|H(\mathbf{p})| > 0$:
     - $f_{xx}(\mathbf{p}) > 0 \implies$ local min
     - $f_{xx}(\mathbf{p}) < 0 \implies$ local max

Remark 4.6.1.
What's really going on?

- Eigenvalues have same sign $\iff$ positive definite or negative definite
  - Positive definite $\implies$ convex $\implies$ local min
  - Negative definite $\implies$ concave $\implies$ local max
- Extrema occur on boundaries, so parameterize each boundary to obtain a function in one less variable and apply standard optimization techniques to yield critical points. Test all critical points to find extrema.
- If possible, use constraint to just reduce equation to one dimension and optimze like single-variable case.

Add examples

4.6.2 Lagrange Multipliers

The setup:

$$\begin{aligned} &\text{Optimize } f(\mathbf{x}) \quad \text{subject to } g(\mathbf{x}) = c \\ &\implies \nabla f = \lambda \nabla g \end{aligned}$$

1. Use this formula to obtain a system of equations in the components of x and the parameter λ.
2. Use λ to obtain a relation involving only components of $\mathbf{x}$.
3. Substitute relations **back into constraint** to obtain a collection of critical points.

4. Evaluate f at critical points to find max/min.

Add examples

4.7 Change of Variables

For any $f : \mathbb{R}^n \to \mathbb{R}^n$ and region R,

$$\int_{g(R)} f(\mathbf{x})\ dV = \int_R (f \circ g)(\mathbf{x}) \cdot |D_g(\mathbf{x})|\ dV$$

4.8 Notation

R is a region, S is a surface, V is a solid.

$$\oint_{\partial S} \mathbf{F} \cdot d\mathbf{r} = \oint_{\partial S} [\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3] \cdot [dx, dy, dz] = \oint_{\partial S} \mathbf{F}_1 dx + \mathbf{F}_2 dy + \mathbf{F}_3 dz$$

The main vector operators

$$\begin{aligned} \nabla : (\mathbb{R}^n \to \mathbb{R}) &\to (\mathbb{R}^n \to \mathbb{R}^n) \\ \varphi &\mapsto \nabla\varphi := \sum_{i=1}^{n} \frac{\partial \varphi}{\partial x_i}\ \mathbf{e}_i \end{aligned}$$

$$\begin{aligned} \operatorname{div}(\mathbf{F}) : (\mathbb{R}^n \to \mathbb{R}^n) &\to (\mathbb{R}^n \to \mathbb{R}) \\ \mathbf{F} &\mapsto \nabla \cdot \mathbf{F} := \sum_{i=1}^{n} \frac{\partial \mathbf{F}_i}{\partial x_i} \end{aligned}$$

$$\begin{aligned} \operatorname{curl}(\mathbf{F}) : (\mathbb{R}^3 \to \mathbb{R}^3) &\to (\mathbb{R}^3 \to \mathbb{R}^3) \\ \mathbf{F} &\mapsto \nabla \times \mathbf{F} \end{aligned}$$

Some terminology:

Scalar Field	$\varphi : X \to \mathbb{R}$
Vector Field	$\mathbf{F} : X \to \mathbb{R}^n$
Gradient Field	$\mathbf{F} : X \to \mathbb{R}^n \Big\vert \exists\varphi : X \to \mathbb{R} \Big\vert \nabla\varphi = F$

- The Gradient: lifts scalar fields on $\mathbb{R}^n$ to vector fields on $\mathbb{R}^n$
- Divergence: drops vector fields on $\mathbb{R}^n$ to scalar fields on $\mathbb{R}^n$
- Curl: takes vector fields on $\mathbb{R}^3$ to vector fields on $\mathbb{R}^3$

$$\mathbf{x} \cdot \mathbf{y} = \langle \mathbf{x},\ \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots \qquad \text{inner/dot product}$$

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x},\ \mathbf{x} \rangle} = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x_1^2 + x_2^2 + \cdots} \qquad \text{norm}$$

$$\mathbf{a} \times \mathbf{b} = \hat{\mathbf{n}} \|\mathbf{a}\| \|\mathbf{b}\| \sin \theta_{\mathbf{a},\mathbf{b}} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} \qquad \text{cross product}$$

$$D_{\mathbf{u}}(\varphi) = \nabla \varphi \cdot \hat{\mathbf{u}} \qquad \text{directional derivative}$$

$$\nabla \coloneqq \sum_{i=1}^{n} \frac{\partial}{\partial x_i} \mathbf{e}_i = \left[\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \cdots, \frac{\partial}{\partial x_n} \right] \qquad \text{del operator}$$

$$\nabla \varphi \coloneqq \sum_{i=1}^{n} \frac{\partial \varphi}{\partial x_i}\ \mathbf{e}_i = \left[\frac{\partial \varphi}{\partial x_1}, \frac{\partial \varphi}{\partial x_2}, \cdots, \frac{\partial \varphi}{\partial x_n} \right] \qquad \text{gradient}$$

$$\Delta \varphi \coloneqq \nabla \cdot \nabla \varphi \coloneqq \sum_{i=1}^{n} \frac{\partial^2 \varphi}{\partial x_i^2} = \frac{\partial^2 \varphi}{\partial x_1^2} + \frac{\partial^2 \varphi}{\partial x_2} + \cdots + \frac{\partial^2 \varphi}{\partial x_n^2} \qquad \text{Laplacian}$$

$$\nabla \cdot \mathbf{F} \coloneqq \sum_{i=1}^{n} \frac{\partial \mathbf{F}_i}{\partial x_i} = \frac{\partial \mathbf{F}_1}{\partial x_1} + \frac{\partial \mathbf{F}_2}{\partial x_2} + \cdots + \frac{\partial \mathbf{F}_n}{\partial x_n} \qquad \text{divergence}$$

$$\nabla \times \mathbf{F} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ \dfrac{\partial}{\partial x} & \dfrac{\partial}{\partial y} & \dfrac{\partial}{\partial z} \\ \mathbf{F}_1 & \mathbf{F}_2 & \mathbf{F}_3 \end{vmatrix} = [\mathbf{F}_{3y} - \mathbf{F}_{2z}, \mathbf{F}_{1z} - \mathbf{F}_{3x}, \mathbf{F}_{2x} - \mathbf{F}_{1y}] \qquad \text{curl}$$

$$\iint_S (\nabla \times \mathbf{F}) \cdot d\mathbf{S} = \iint_S (\nabla \times \mathbf{F}) \cdot \mathbf{n}\ dS \qquad \text{surface integral}$$

4.9 Big Theorems

4.9.1 Stokes' and Consequences

Theorem 4.9.1 ***(Stokes' Theorem).***

$$\oint_{\partial S} \mathbf{F} \cdot d\mathbf{r} = \iint_S (\nabla \times \mathbf{F}) \cdot d\mathbf{S}.$$

Remark 4.9.1.
Note that if S is a closed surface, so $\partial S = \emptyset$, this integral vanishes.

Corollary 4.9.1 ***(Green's Theorem).***

$$\oint_{\partial R} (L\ dx + M\ dy) = \iint_R \left(\frac{\partial M}{\partial x} - \frac{\partial L}{\partial y} \right) dxdy.$$

Proof (?).
Recovering Green's Theorem from Stokes' Theorem:
Let $\mathbf{F} = [L, M, 0]$, then $\nabla \times \mathbf{F} = [0, 0, \dfrac{\partial M}{\partial x} - \dfrac{\partial L}{\partial y}]$

■

Corollary 4.9.2 ***(Divergence Theorem).***

$$\iint_{\partial V} \mathbf{F} \cdot d\mathbf{S} = \iiint_V (\nabla \cdot \mathbf{F})\ dV.$$

Remark 4.9.2.

- $\nabla \times (\nabla \varphi) = 0$
- $\nabla \cdot (\nabla \times \mathbf{F}) = 0$

4.9.2 Directional Derivatives

Definition 4.9.1 (Directional Derivative)**.**

$$D_{\mathbf{v}} f(\mathbf{p}) \coloneqq \frac{\partial f}{\partial t} (\mathbf{p} + t\mathbf{v})\Big|_{t=0}.$$

Remark 4.9.3.
Note that the directional derivative uses a normalized direction vector!

Theorem 4.9.2 ***(Dot product expression of directional derivative).***
Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$. Then

$$D_{\mathbf{v}} f(\mathbf{p}) = \langle \nabla f(\mathbf{p}),\ \mathbf{v} \rangle.$$

Proof (?).

We first use the fact that we can find L, the best linear approximation to f:

$$L(\mathbf{x}) := f(\mathbf{p}) + D_f(\mathbf{p})(\mathbf{x} - \mathbf{p})$$

$$\begin{aligned}
D_{\mathbf{v}} f(\mathbf{p}) &= D_{\mathbf{v}} L(\mathbf{p}) \\
&= \lim_{t\to 0} \frac{L(\mathbf{p} + t\mathbf{v}) - L(\mathbf{p})}{t} \\
&= \lim_{t\to 0} \frac{f(\mathbf{p}) + D_f(\mathbf{p})(\mathbf{p} + t\mathbf{v} - \mathbf{p}) - (f(\mathbf{p}) + D_f(\mathbf{p})(\mathbf{p} - \mathbf{p}))}{t} \\
&= \lim_{t\to 0} \frac{D_f(\mathbf{p})(t\mathbf{v})}{t} \\
&= D_f(\mathbf{p})\mathbf{v} \\
&:= \nabla f(\mathbf{p}) \cdot \mathbf{v}.
\end{aligned}$$

Need a better proof, not clear that this works.

■

4.10 Computing Integrals

4.10.1 Changing Coordinates

Multivariable Chain Rule

todo

Polar and Cylindrical Coordinates

$$\begin{aligned}
x &= r\cos\theta \\
y &= r\sin\theta \\
dV &\mapsto r \quad dr\ d\theta
\end{aligned}$$

Spherical Coordinates

$$\begin{aligned}
x &= r\cos\theta = \rho\sin\varphi\cos\theta \\
y &= r\sin\theta = \rho\sin\varphi\sin\theta \\
dV &\mapsto r^2\sin\varphi \quad dr\ d\varphi\ d\theta
\end{aligned}$$

4.10.2 Line Integrals

Curves

- Parametrize the path C as $\{\mathbf{r}(t) : t \in [a, b]\}$, then

$$\begin{aligned}\int_C f\ ds &\coloneqq \int_a^b (f \circ \mathbf{r})(t)\ \|\mathbf{r}'(t)\|\ dt \\ &= \int_a^b f(x(t), y(t), z(t))\sqrt{x_t^2 + y_t^2 + z_t^2}\ dt\end{aligned}$$

Vector Fields

- If exact:

$$\frac{\partial}{\partial y}\mathbf{F_1} = \frac{\partial}{\partial x}\mathbf{F_2} \implies \int \mathbf{F_1}\ dx + \mathbf{F_2}\ dy = \varphi(\mathbf{p_1}) - \varphi(\mathbf{p_0})$$

The function φ can be found using the same method from ODEs.

- Parametrize the path C as $\{\mathbf{r}(t) : t \in [a, b]\}$, then

$$\begin{aligned}\int_C \mathbf{F} \cdot d\mathbf{r} &\coloneqq \int_a^b (\mathbf{F} \circ \mathbf{r})(t) \cdot \mathbf{r}'(t)\ dt \\ &= \int_a^b [\mathbf{F}_1(x(t), y(t), \cdots), \mathbf{F}_2(x(t), y(t), \cdots)] \cdot [x_t, y_t, \cdots]\ dt \\ &= \int_a^b \mathbf{F}_1(x(t), y(t) \cdots)x_t + \mathbf{F}_2(x(t), y(t), \cdots)y_t + \cdots\ dt\end{aligned}$$

- Equivalently written:

$$\int_a^b \mathbf{F}_1\ dx + \mathbf{F}_2\ dy + \cdots \coloneqq \int_C \mathbf{F} \cdot d\mathbf{r}$$

in which case $[dx, dy, \cdots] \coloneqq [x_t, y_t, \cdots] = \mathbf{r}'(t)$.

- *Remember to substitute dx back into the integrand!!*

4.10.3 Flux

$$\iint_S \mathbf{F} \cdot d\mathbf{S} = \iint_S \mathbf{F} \cdot \widehat{\mathbf{n}}\ dS.$$

4.10.4 Area

Proposition 4.10.1 ***(Areas can be computed with Green's Theorem).***
Given R and $f(x, y) = 0$,

$$A(R) = \oint_{\partial R} x\ dy = -\oint_{\partial R} y\ dx = \frac{1}{2}\oint_{\partial R} -y\ dx + x\ dy.$$

Proof (?).
Compute

$$\begin{aligned} \oint_{\partial R} x\ dy = -\oint_{\partial R} y\ dx \\ = \frac{1}{2}\oint_{\partial R} -y\ dx + x\ dy = \frac{1}{2}\iint_R 1-(-1)\ dA = \iint_R 1\ dA \end{aligned}$$

■

4.10.5 Surface Integrals

- For a paramterization $\mathbf{r}(s,t) : U \to S$ of a surface S and any function $f : \mathbb{R}^n \to \mathbb{R}$,

$$\iint_S f\ dA = \iint_U (f \circ \mathbf{r})(s,t)\ \|\mathbf{n}\|\ dA$$

- Can obtain a normal vector $\mathbf{n} = T_u \times T_v$

4.11 Other Results

Example 4.11.1 (?)**.**
$\nabla \cdot \mathbf{F} = 0 \not\Longrightarrow \exists G : \ \mathbf{F} = \nabla \times G$. A counterexample

$$\begin{aligned} \mathbf{F}(x,y,z) = \frac{1}{\sqrt{x^2+y^2+z^2}}[x,y,z]\ , \quad S = S^2 \subset \mathbb{R}^3 \\ \implies \nabla\mathbf{F} = 0 \text{ but } \iint_{S^2} \mathbf{F} \cdot d\mathbf{S} = 4\pi \neq 0 \end{aligned}$$

Where by Stokes' theorem,

$$\begin{aligned} \mathbf{F} = \nabla \times \mathbf{G} \implies \iint_{S^2} \mathbf{F} &= \iint_{S^2} \nabla \times \mathbf{G} \\ &= \oint_{\partial S^2} \mathbf{G}\ d\mathbf{r} && \text{by Stokes} \\ &= 0 \end{aligned}$$

since $\partial S^2 = \emptyset$.

Proposition 4.11.1*(Sufficient Conditions).*
Sufficient condition: if $\mathbf{F}$ is everywhere C^1,

$$\exists \mathbf{G} : \ \mathbf{F} = \nabla \times \mathbf{G} \iff \iint_S \mathbf{F} \cdot d\mathbf{S} = 0 \text{ for all closed surfaces } S.$$

5 Linear Algebra

Remark 5.0.1.
The underlying field will be assumed to be $\mathbb{R}$ for this section.

5.1 Notation

$\mathrm{Mat}(m,n)$	the space of all $m \times n$ matrices
T	a linear map $\mathbb{R}^n \to \mathbb{R}^m$
$A \in \mathrm{Mat}(m,n)$	an $m \times n$ matrix representing T
$A^t \in \mathrm{Mat}(n,m)$	an $n \times m$ transposed matrix
$\mathbf{a}$	a $1 \times n$ column vector
$\mathbf{a}^t$	an $n \times 1$ row vector
$A = [\mathbf{a}_1, \cdots, \mathbf{a}_n]$	a matrix formed with $\mathbf{a}_i$ as the columns
V, W	vector spaces
$\lvert V\rvert, \dim(W)$	dimensions of vector spaces
$\det(A)$	the determinant of A
$\left[A \mid \mathbf{b}\right] := [\mathbf{a}_1, \mathbf{a}_2, \cdots \mathbf{a}_n, \mathbf{b}]$	augmented matrices
$\left[A \mid B\right] := [\mathbf{a}_1, \cdots \mathbf{a}_n, \mathbf{b}_1, \cdots, \mathbf{b}_m]$	block matrices
$\mathrm{Spec}(A)$	the multiset of eigenvalues of A
$A\mathbf{x} = \mathbf{b}$	a system of linear equations
$r := \mathrm{rank}(A)$	the rank of A
$r_b = \mathrm{rank}\left(\left[A \mid \mathbf{b}\right]\right)$	the rank of A augmented by $\mathbf{b}$.

5.2 Big Theorems

Theorem 5.2.1 ***(Rank-Nullity).***

$$|\ker(A)| + |\mathrm{im}(A)| = |\mathrm{dom}(A)|.$$

Generalization: the following sequence is always exact:

$$0 \to \ker(A) \xhookrightarrow{\mathrm{id}} \mathrm{dom}(A) \xrightarrow{A} \mathrm{im}(A) \to 0.$$

Moreover, it always splits, so $\mathrm{dom}\, A = \ker A \oplus \mathrm{im}\, A$ and thus $|\mathrm{dom}(A)| = |\ker(A)| + |\mathrm{im}(A)|$.

5.3 Big List of Equivalent Properties

Let A be an $m \times n$ matrix. TFAE: - A is invertible and has a unique inverse A^{-1} - A^T is invertible - $\det(A) \neq 0$ - The linear system $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $b \in \mathbb{R}^m$ -

The homogeneous system $A\mathbf{x} = 0$ has only the trivial solution $\mathbf{x} = 0$ - $\text{rank}(A) = n$ - i.e. A is full rank - $\text{nullity}(A) \coloneqq \dim \text{nullspace}(A) = 0$ - $A = \prod_{i=1}^{k} E_i$ for some finite k, where each E_i is an elementary matrix. - A is row-equivalent to the identity matrix I_n - A has exactly n pivots - The columns of A are a basis for $\mathbb{R}^n$ - i.e. $\text{colspace}(A) = \mathbb{R}^n$ - The rows of A are a basis for $\mathbb{R}^m$ - i.e. $\text{rowspace}(A) = \mathbb{R}^m$ - $(\text{colspace}(A))^\perp = (\text{rowspace}A)^\perp = \{\mathbf{0}\}$ - Zero is not an eigenvalue of A. - A has n linearly independent eigenvectors - The rows of A are coplanar.

Similarly, by taking negations, TFAE:

- A is not invertible
- A is singular
- A^T is not invertible
- $\det A = 0$
- The linear system $A\mathbf{x} = \mathbf{b}$ has either no solution or infinitely many solutions.
- The homogeneous system $A\mathbf{x} = \mathbf{0}$ has nontrivial solutions
- $\text{rank}\, A < n$
- $\dim \text{nullspace}\ A > 0$
- At least one row of A is a linear combination of the others
- The *RREF* of A has a row of all zeros.

Reformulated in terms of linear maps T, TFAE: - $T^{-1} : \mathbb{R}^m \to \mathbb{R}^n$ exists - $\text{im}(T) = \mathbb{R}^n$ - $\ker(T) = 0$ - T is injective - T is surjective - T is an isomorphism - The system $A\mathbf{x} = 0$ has infinitely many solutions

5.4 Vector Spaces

Proposition 5.4.1 ***(Two-step vector subspace test).***
If $V \subseteq W$, then V is a subspace of W if the following hold:

$$\mathbf{0} \in V \tag{1}$$

$$\mathbf{a}, \mathbf{b} \in V \implies t\mathbf{a} + \mathbf{b} \in V. \tag{2}$$

5.4.1 Linear Independence

Proposition 5.4.2 ***(?).***
Any set of two vectors $\{\mathbf{v}, \mathbf{w}\}$ is linearly **dependent** $\iff \exists \lambda :\ \mathbf{v} = \lambda \mathbf{w}$, i.e. one is not a scalar multiple of the other.

5.4.2 The Inner Product

The point of this section is to show how an inner product can induce a notion of “angle”, which agrees with our intuition in Euclidean spaces such as $\mathbb{R}^n$, but can be extended to much less intuitive things, like spaces of functions.

Definition 5.4.1 (The standard inner product)**.**
The **Euclidean inner product** is defined as

$$\langle \mathbf{a},\ \mathbf{b} \rangle = \sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n.$$

Also sometimes written as $\mathbf{a}^T \mathbf{b}$ or $\mathbf{a} \cdot \mathbf{b}$.

Proposition 5.4.3*(Inner products induce norms and angles).*
Yields a norm

$$\|\mathbf{x}\| \coloneqq \sqrt{\langle \mathbf{x},\ \mathbf{x} \rangle}$$

which has a useful alternative formulation

$$\langle \mathbf{x},\ \mathbf{x} \rangle = \|\mathbf{x}\|^2.$$

This leads to a notion of angle:

$$\langle \mathbf{x},\ \mathbf{y} \rangle = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta_{x,y} \implies \cos \theta_{x,y} \coloneqq \frac{\langle \mathbf{x},\ \mathbf{y} \rangle}{\|\mathbf{x}\| \|\mathbf{y}\|} = \langle \widehat{\mathbf{x}},\ \widehat{\mathbf{y}} \rangle$$

where $\theta_{x,y}$ denotes the angle between the vectors $\mathbf{x}$ and $\mathbf{y}$.

Remark 5.4.1.
Since $\cos \theta = 0$ exactly when $\theta = \pm \frac{\pi}{2}$, we can can declare two vectors to be **orthogonal** exactly in this case:

$$\mathbf{x} \in \mathbf{y}^\perp \iff \langle \mathbf{x},\ \mathbf{y} \rangle = 0.$$

Note that this makes the zero vector orthogonal to everything.

Definition 5.4.2 (Orthogonal Complement/Perp)**.**
Given a subspace $S \subseteq V$, we define its **orthogonal complement**

$$S^\perp = \left\{ \mathbf{v} \in V \;\middle|\; \forall \mathbf{s} \in S,\ \langle \mathbf{v},\ \mathbf{s} \rangle = 0 \right\}.$$

Remark 5.4.2.
Any choice of subspace $S \subseteq V$ yields a decomposition $V = S \oplus S^\perp$.

Proposition 5.4.4*(Formula expanding a norm and 'Pythagorean theorem').*
A useful formula is

$$\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\langle \mathbf{x},\ \mathbf{y} \rangle + \|\mathbf{y}\|^2, .$$

When $\mathbf{x} \in \mathbf{y}^{\perp}$, this reduces to

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2.$$

Proposition 5.4.5 ***(Properties of the inner product).***

1. **Bilinearity:**
$$\left\langle \sum_j \alpha_j \mathbf{a}_j,\ \sum_k \beta_k \mathbf{b}_k \right\rangle = \sum_j \sum_i \alpha_j \beta_i \langle \mathbf{a}_j,\ \mathbf{b}_i \rangle.$$

2. **Symmetry**:
$$\langle \mathbf{a},\ \mathbf{b}\rangle = \langle \mathbf{b},\ \mathbf{a}\rangle$$

3. **Positivity**:
$$\mathbf{a} \neq \mathbf{0} \implies \langle \mathbf{a},\ \mathbf{a}\rangle > 0$$

4. **Nondegeneracy**:
$$\mathbf{a} = \mathbf{0} \iff \langle \mathbf{a},\ \mathbf{a}\rangle = 0$$

Proof of Cauchy-Schwarz: See Goode page 346.

5.4.3 Gram-Schmidt Process

Extending a basis $\{\mathbf{x}_i\}$ to an orthonormal basis $\{\mathbf{u}_i\}$

$$\begin{aligned}
\mathbf{u}_1 &= N(\mathbf{x_1}) \\
\mathbf{u}_2 &= N(\mathbf{x}_2 - \langle \mathbf{x}_2,\ \mathbf{u}_1\rangle \mathbf{u}_1) \\
\mathbf{u}_3 &= N(\mathbf{x}_3 - \langle \mathbf{x}_3,\ \mathbf{u}_1\rangle \mathbf{u}_1 - \langle \mathbf{x}_3,\ \mathbf{u}_2\rangle \mathbf{u}_2) \\
\vdots\ & \quad \vdots \\
\mathbf{u}_k &= N(\mathbf{x}_k - \sum_{i=1}^{k-1} \langle \mathbf{x}_k,\ \mathbf{u}_i\rangle \mathbf{u}_i)
\end{aligned}$$

where N denotes normalizing the result.

In more detail The general setup here is that we are given an orthogonal basis $\{\mathbf{x}_i\}_{i=1}^n$ and we want to produce an **orthonormal** basis from them.

Why would we want such a thing? Recall that we often wanted to change from the standard basis $\mathcal{E}$ to some different basis $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \cdots\}$. We could form the change of basis matrix $B = [\mathbf{b}_1, \mathbf{b}_2, \cdots]$ acts on vectors in the $\mathcal{B}$ basis according to

$$B[\mathbf{x}]_{\mathcal{B}} = [\mathbf{x}]_{\mathcal{E}}.$$

But to change from $\mathcal{E}$ to $\mathcal{B}$ requires computing B^{-1}, which acts on vectors in the standard basis according to

$$B^{-1}[\mathbf{x}]_{\mathcal{E}} = [\mathbf{x}]_{\mathcal{B}}.$$

If, on the other hand, the $\mathbf{b}_i$ are orthonormal, then $B^{-1} = B^T$, which is much easier to compute. We also obtain a rather simple formula for the coordinates of $\mathbf{x}$ with respect to $\mathcal{B}$. This follows because we can write

$$\mathbf{x} = \sum_{i=1}^{n} \langle \mathbf{x},\ \mathbf{b}_i \rangle \mathbf{b}_i \coloneqq \sum_{i=1}^{n} c_i \mathbf{b}_i, .$$

and we find that

$$[\mathbf{x}]_{\mathcal{B}} = \mathbf{c} \coloneqq [c_1, c_2, \cdots, c_n]^T..$$

This also allows us to simplify projection matrices. Supposing that A has orthonormal columns and letting S be the column space of A, recall that the projection onto S is defined by

$$P_S = Q(Q^TQ)^{-1}Q^T..$$

Since Q has orthogonal columns and satisfies $Q^TQ = I$, this simplifies to

$$P_S = QQ^T..$$

The Algorithm Given the orthogonal basis $\{\mathbf{x}_i\}$, we form an orthonormal basis $\{\mathbf{u}_i\}$ iteratively as follows.

First define

$$\begin{aligned} N : \mathbb{R}^n &\to S^{n-1} \\ \mathbf{x} &\mapsto \widehat{\mathbf{x}} \coloneqq \frac{\mathbf{x}}{\|\mathbf{x}\|} \end{aligned}$$

which projects a vector onto the unit sphere in $\mathbb{R}^n$ by normalizing. Then,

$$\begin{aligned} \mathbf{u}_1 &= N(\mathbf{x_1}) \\ \mathbf{u}_2 &= N(\mathbf{x}_2 - \langle \mathbf{x}_2,\ \mathbf{u}_1 \rangle \mathbf{u}_1) \\ \mathbf{u}_3 &= N(\mathbf{x}_3 - \langle \mathbf{x}_3,\ \mathbf{u}_1 \rangle \mathbf{u}_1 - \langle \mathbf{x}_3,\ \mathbf{u}_2 \rangle \mathbf{u}_2) \\ \vdots\ &\quad \vdots \\ \mathbf{u}_k &= N(\mathbf{x}_k - \sum_{i=1}^{k-1} \langle \mathbf{x}_k,\ \mathbf{u}_i \rangle \mathbf{u}_i) \end{aligned}$$

In words, at each stage, we take one of the original vectors $\mathbf{x}_i$, then subtract off its projections onto all of the $\mathbf{u}_i$ we've created up until that point. This leaves us with only the component of $\mathbf{x}_i$ that is orthogonal to the span of the previous $\mathbf{u}_i$ we already have, and we then normalize each $\mathbf{u}_i$ we obtain this way.

5.4.4 The Fundamental Subspaces Theorem

Given a matrix $A \in \text{Mat}(m, n)$, and noting that

$$A : \mathbb{R}^n \to \mathbb{R}^m, \\ A^T : \mathbb{R}^m \to \mathbb{R}^n$$

We have the following decompositions:

$$\begin{aligned} \mathbb{R}^n &\cong \ker A \oplus \operatorname{im} A^T &&\cong \text{nullspace}(A) \oplus \text{colspace}(A^T) \\ \mathbb{R}^m &\cong \operatorname{im} A \oplus \ker A^T &&\cong \text{colspace}(A) \oplus \text{nullspace}(A^T) \end{aligned}$$

5.4.5 Computing change of basis matrices

todo

5.5 Matrices

Remark 5.5.1.
An $m \times n$ matrix is a map from n-dimensional space to m-dimensional space. The number of *rows* tells you the dimension of the codomain, the number of *columns* tells you the dimension of the *domain*.

⚠**Warning 5.1:** The space of matrices is not an integral domain! Counterexample: if A is singular and nonzero, there is some nonzero $\mathbf{v}$ such that $A\mathbf{v} = \mathbf{0}$. Then setting $B = [\mathbf{v}, \mathbf{v}, \cdots]$ yields $AB = 0$ with $A \neq 0, B \neq 0$.

Definition 5.5.1 (Rank of a matrix)**.**
The **rank** of a matrix A representing a linear transformation T is $\dim \text{colspace}(A)$, or equivalently $\dim \operatorname{im} T$.

Proposition 5.5.1 ***(?).***
$\text{rank}(A)$ is equal to the number of nonzero rows in $\text{RREF}(A)$.

Definition 5.5.2 (Trace of a Matrix)**.**

$$\text{Trace}(A) = \sum_{i=1}^{m} A_{ii}$$

Definition 5.5.3 (Elementary Row Operations)**.**
The following are **elementary row operations** on a matrix:

- Permute rows
- Multiple a row by a scalar
- Add any row to another

Proposition 5.5.2 ***(Formula for matrix multiplication).***
If $A = [\mathbf{a}_1, \mathbf{a}_2, \cdots] \in \text{Mat}(m, n)$ and $B = [\mathbf{b}_1, \mathbf{b}_2, \cdots] \in \text{Mat}(n, p)$, then

$$C \coloneqq AB \implies c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = \langle \mathbf{a_i},\ \mathbf{b_j} \rangle$$

where $1 \leq i \leq m$ and $1 \leq j \leq p$. In words, each entry c_{ij} is obtained by dotting *row* i of A against *column* j of B.

5.5.1 Systems of Linear Equations

Definition 5.5.4 (Consistent and inconsistent)**.**
A system of linear equations is **consistent** when it has at least one solution. The system is **inconsistent** when it has no solutions.

Definition 5.5.5 (Homogeneous Systems)**.**
?

Remark 5.5.2.
Homogeneous systems are always consistent, i.e. there is always at least one solution.

Remark 5.5.3.

- Tall matrices: more equations than unknowns, *overdetermined*
- Wide matrices: more unknowns than equations, *underdetermined*

Proposition 5.5.3 ***(Characterizing solutions to a system of linear equations).***
There are three possibilities for a system of linear equations:

1. No solutions (inconsistent)
2. One unique solution (consistent, square or tall matrices)
3. Infinitely many solutions (consistent, underdetermined, square or wide matrices)

These possibilities can be check by considering $r \coloneqq \text{rank}(A)$:

- $r < r_b$: case 1, no solutions.
- $r = r_b$: case 1 or 2, at least one solution.
  - $r_b = n$: case 2, a unique solution.
  - $r_b < n$: case 3, infinitely many solutions.

5.5.2 Determinants

Proposition 5.5.4 ***(?).***

$$\det (A \mod p) \mod p \equiv (\det A) \mod p$$

Proposition 5.5.5 ***(Inverse of a 2×2 matrix).***
For 2×2 matrices,

$$A^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

In words, swap the main diagonal entries, and flip the signs on the off-diagonal.

Proposition 5.5.6 ***(Properties of the determinant).***
Let $A \in \mathrm{Mat}(m, n)$, then there is a function

$$\begin{aligned} \det : \mathrm{Mat}(m, m) &\to \mathbb{R} \\ A &\mapsto \det(A) \end{aligned}$$

satisfying the following properties:

- det is a group homomorphism onto $(\mathbb{R}, \cdot)$:

$$\det(AB) = \det(A)\det(B)$$

  - Some corollaries:

$$\begin{aligned} \det A^k &= k \det A \\ \det(A^{-1}) &= (\det A)^{-1} \det(A^t) &= \det(A). \end{aligned}$$

- Invariance under adding scalar multiples of any row to another:

$$\det \begin{bmatrix} & \vdots & \\ — & \mathbf{a}_i & — \\ & \vdots & \end{bmatrix} = \det \begin{bmatrix} & \vdots & \\ — & \mathbf{a}_i + t\mathbf{a_j} & — \\ & \vdots & \end{bmatrix}$$

- Sign change under row permutation:

$$\det \begin{bmatrix} & \vdots & \\ — & \mathbf{a}_i & — \\ & \vdots & \\ — & \mathbf{a}_j & — \\ & \vdots & \end{bmatrix} = (-1) \det \begin{bmatrix} & \vdots & \\ — & \mathbf{a}_j & — \\ & \vdots & \\ — & \mathbf{a}_i & — \\ & \vdots & \end{bmatrix}$$

  - More generally, for a permutation $\sigma \in S_n$,

$$\det \begin{bmatrix} & \vdots & \\ — & \mathbf{a}_i & — \\ & \vdots & \\ — & \mathbf{a}_j & — \\ & \vdots & \end{bmatrix} = (-1)^{\mathrm{sgn}(\sigma)} \det \begin{bmatrix} & \vdots & \\ — & \mathbf{a}_{\sigma(j)} & — \\ & \vdots & \\ — & \mathbf{a}_{\sigma(i)} & — \\ & \vdots & \end{bmatrix}$$

- Multilinearity in rows:

$$\det\begin{bmatrix} & \vdots & \\ — & t\mathbf{a}_i & — \\ & \vdots & \end{bmatrix} = t\det\begin{bmatrix} & \vdots & \\ — & \mathbf{a}_i & — \\ & \vdots & \end{bmatrix}$$

$$\det\begin{bmatrix} — & t\mathbf{a}_1 & — \\ — & t\mathbf{a}_2 & — \\ & \vdots & \\ — & t\mathbf{a}_m & — \end{bmatrix} = t^m\det\begin{bmatrix} — & \mathbf{a}_1 & — \\ — & \mathbf{a}_2 & — \\ & \vdots & \\ — & \mathbf{a}_m & — \end{bmatrix}$$

$$\det\begin{bmatrix} — & t_1\mathbf{a}_1 & — \\ — & t_2\mathbf{a}_2 & — \\ & \vdots & \\ — & t_m\mathbf{a}_m & — \end{bmatrix} = \prod_{i=1}^{m} t_i\det\begin{bmatrix} — & \mathbf{a}_1 & — \\ — & \mathbf{a}_2 & — \\ & \vdots & \\ — & \mathbf{a}_m & — \end{bmatrix}.$$

- Linearity in each row:

$$\det\begin{bmatrix} & \vdots & \\ — & \mathbf{a}_i + \mathbf{a}_j & — \\ & \vdots & \end{bmatrix} = \det\begin{bmatrix} & \vdots & \\ — & \mathbf{a}_i & — \\ & \vdots & \end{bmatrix} + \det\begin{bmatrix} & \vdots & \\ — & \mathbf{a}_j & — \\ & \vdots & \end{bmatrix}.$$

- $\det(A)$ is the volume of the parallelepiped spanned by the columns of A.
- If any row of A is all zeros, $\det(A) = 0$.

Proposition 5.5.7 ***(Characterizing singular matrices).***
TFAE:

- $\det(A) = 0$
- A is singular.

5.5.3 Computing Determinants

Useful shortcuts:

- If A is upper or lower triangular, $\det(A) = \prod_i a_{ii}$.

Definition 5.5.6 (Minors)**.**
The **minor** M_{ij} of $A \in \mathrm{Mat}(n, n)$ is the *determinant* of the $(n-1)\times(n-1)$ matrix obtained by deleting the ith row and jth column from A.

Definition 5.5.7 (Cofactors)**.**
The **cofactor** C_{ij} is the scalar defined by

$$C_{ij} \coloneqq (-1)^{i+j}M_{ij}.$$

Proposition 5.5.8 ***(Laplace/Cofactor Expansion).***
For any fixed i, there is a formula

$$\det(A) = \sum_{j=1}^{n} a_{ij} C_{ij}.$$

Example 5.5.1 (?).
Let

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}.$$

Then

$$\det A = 1 \cdot \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \cdot \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} + 3 \cdot \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} = 1 \cdot (-3) - 2 \cdot (-6) + 3 \cdot (-3) = 0.$$

Proposition 5.5.9 ***(Computing determinant from RREF).***
$\det(A)$ can be computed by reducing A to $\mathrm{RREF}(A)$ (which is upper triangular) and keeping track of the following effects:

- $R_i \leftarrow R_i \pm tR_j$: no effect.
- $R_i \rightleftharpoons R_j$: multiply by (-1).
- $R_i \leftarrow tR_i$: multiply by t.

5.5.4 Inverting a Matrix

Proposition 5.5.10 ***(Cramer's Rule).***
Given a linear system $A\mathbf{x} = \mathbf{b}$, writing $\mathbf{x} = [x_1, \cdots, x_n]$, there is a formula

$$x_i = \frac{\det(B_i)}{\det(A)}$$

where B_i is A with the ith column deleted and replaced by $\mathbf{b}$.

Proposition 5.5.11 ***(Gauss-Jordan Method for inverting a matrix).***
Under the equivalence relation of elementary row operations, there is an equivalence of augmented matrices:

$$\left[A \mid I \right] \sim \left[I \mid A^{-1} \right]$$

where I is the $n \times n$ identity matrix.

Proposition 5.5.12 ***(Cofactor formula for inverse).***

$$A^{-1} = \frac{1}{\det(A)}[C_{ij}]^t.$$

where C_{ij} is the *cofactor*(Definition 5.5.7) at position i, j.[a]

[a]Note that the matrix appearing here is sometimes called the *adjugate*.

Example 5.5.2 (Inverting a 2×2 matrix)**.**

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \quad \text{where } ad - bc \neq 0$$

What's the pattern?

1. Always divide by determinant
2. Swap the diagonals
3. Hadamard product with checkerboard

Example 5.5.3 (Inverting a 3×3 matrix)**.**

$$\begin{bmatrix} + & - \\ - & + \end{bmatrix}$$

$$A^{-1} := \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}^{-1} = \frac{1}{\det A}\begin{bmatrix} ei - fh & -(bi - ch) & bf - ce \\ -(di - fg) & ai - cg & -(af - cd) \\ dh - eg & -(ah - bg) & ae - bd \end{bmatrix}.$$

The pattern:

1. Divide by determinant
2. Each entry is determinant of submatrix of A with corresponding col/row deleted
3. Hadamard product with checkerboard

$$\begin{bmatrix} + & - & + \\ - & + & - \\ + & - & + \end{bmatrix}$$

4. Transpose at the end!!

5.5.5 Bases for Spaces of a Matrix

Let $A \in \mathrm{Mat}(m, n)$ represent a map $T : \mathbb{R}^n \to \mathbb{R}^m$.

Add examples.

Definition 5.5.8 (Pivot)**.**
?

todo

Proposition 5.5.13.

$$\dim \operatorname{rowspace}(A) = \dim \operatorname{colspace}(A).$$

The row space

$$\operatorname{im}(T)^\vee = \operatorname{rowspace}(A) \subset \mathbb{R}^n.$$

Reduce to RREF, and take nonzero rows of $\operatorname{RREF}(A)$.

The column space

$$\operatorname{im}(T) = \operatorname{colspace}(A) \subseteq \mathbb{R}^m$$

Reduce to RREF, and take columns with pivots from original A.

Remark 5.5.4.
Not enough pivots implies columns don't span the entire target domain

The nullspace

$$\ker(T) = \operatorname{nullspace}(A) \subseteq \mathbb{R}^n$$

Reduce to RREF, zero rows are free variables, convert back to equations and pull free variables out as scalar multipliers.

Eigenspaces For each $\lambda \in \operatorname{Spec}(A)$, compute a basis for $\ker(A - \lambda I)$.

5.5.6 Eigenvalues and Eigenvectors

Definition 5.5.9 (Eigenvalues, eigenvectors, eigenspaces)**.**
A vector $\mathbf{v}$ is said to be an **eigenvector** of A with **eigenvalue** $\lambda \in \operatorname{Spec}(A)$ iff

$$A\mathbf{v} = \lambda \mathbf{v}$$

For a fixed λ, the corresponding **eigenspace** E_λ is the span of all such vectors.

Remark 5.5.5.

- Similar matrices have identical eigenvalues and multiplicities.
- Eigenvectors corresponding to distinct eigenvalues are **always** linearly independent
- A has n distinct eigenvalues $\implies$ A has n linearly independent eigenvectors.
- A matrix A is diagonalizable $\iff$ A has n linearly independent eigenvectors.

Proposition 5.5.14 ***(How to find eigenvectors).***
For $\lambda \in \mathrm{Spec}(A)$,

$$\mathbf{v} \in E_\lambda \iff \mathbf{v} \in \ker(A - I\lambda).$$

Remark 5.5.6.
Some miscellaneous useful facts:

- $\lambda \in \mathrm{Spec}(A) \implies \lambda^2 \in \mathrm{Spec}(A^2)$ with the same eigenvector.
- $\prod \lambda_i = \det A$
- $\sum \lambda_i = \mathrm{Tr}\ A$

Finding generalized eigenvectors

todo

Diagonalizability

Remark 5.5.7.
An $n \times n$ matrix P is diagonalizable iff its eigenspace is all of $\mathbb{R}^n$ (i.e. there are n linearly independent eigenvectors, so they span the space.)

Remark 5.5.8.
A is diagonalizable if there is a basis of eigenvectors for the range of P.

5.5.7 Useful Counterexamples

$$A := \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \implies A^n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}, \qquad \mathrm{Spec}(A) = [1, 1]$$

$$A := \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \implies A^2 = I_2, \qquad \mathrm{Spec}(A) = [1, -1]$$

5.6 Advanced Topics

5.6.1 Changing Basis

Proposition 5.6.1 ***(Changing to the standard basis).***
The transition matrix from a given basis $\mathcal{B} = \{\mathbf{b}_i\}_{i=1}^n$ to the standard basis is given by

$$A := \begin{bmatrix} | & | & & | \\ \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_n \\ | & | & & | \end{bmatrix},$$

and the transition matrix from the standard basis to $\mathcal{B}$ is A^{-1}.

5.6.2 Orthogonal Matrices

Given a notion of orthogonality for vectors, we can extend this to matrices. A square matrix is said to be orthogonal iff $QQ^T = Q^TQ = I$. For rectangular matrices, we have the following characterizations:

$$\begin{aligned} QQ^T = I &\implies \text{The rows of } Q \text{ are orthogonal,} \\ Q^TQ = I &\implies \text{The columns of } Q \text{ are orthogonal.} \end{aligned}$$

To remember which condition is which, just recall that matrix multiplication AB takes the inner product between the **rows** of A and the **columns** of B. So if, for example, we want to inspect whether or not the columns of Q are orthogonal, we should let $B = Q$ in the above formulation – then we just note that the rows of Q^T are indeed the columns of Q, so Q^TQ computes the inner products between all pairs of the columns of Q and stores them in a matrix.

5.6.3 Projections

Remark 5.6.1.
A projection P induces a decomposition

$$\text{dom}(P) = \ker(P) \oplus \ker(P)^{\perp}.$$

Check! Domain or range..?

Distance from a point $\mathbf{p}$ to a line $\mathbf{a} + t\mathbf{b}$: let $\mathbf{w} = \mathbf{p} - \mathbf{a}$, then: $\|\mathbf{w} - P(\mathbf{w}, \mathbf{v})\|$

Proposition 5.6.2 ***(Projection onto range).***

$$\text{Proj}_{\text{range}(A)}(\mathbf{x}) = A(A^tA)^{-1}A^t\mathbf{x}.$$

Mnemonic:

$$P \approx \frac{A^tA}{AA^t}.$$

With an inner product in hand and a notion of orthogonality, we can define a notion of **orthogonal projection** of one vector onto another, and more generally of a vector onto a subspace spanned by multiple vectors.

Projection Onto a Vector Say we have two vectors $\mathbf{x}$ and $\mathbf{y}$, and we want to define "the component of $\mathbf{x}$ that lies along $\mathbf{y}$", which we'll call $\mathbf{p}$. We can work out what the formula should be using a simple model:

We notice that whatever p is, it will in the direction of $\mathbf{y}$, and thus $\mathbf{p} = \lambda\widehat{\mathbf{y}}$ for some scalar λ, where in fact $\lambda = \|\mathbf{p}\|$ since $\|\widehat{\mathbf{y}}\| = 1$. We will find that $\lambda = \langle \mathbf{x},\ \widehat{\mathbf{y}} \rangle$, and so

$$\mathbf{p} = \langle \mathbf{x},\ \widehat{\mathbf{y}} \rangle \widehat{\mathbf{y}} = \frac{\langle \mathbf{x},\ \mathbf{y} \rangle}{\langle \mathbf{y},\ \mathbf{y} \rangle}\mathbf{y}.$$

Notice that we can then form a "residual" vector $\mathbf{r} = \mathbf{x} - \mathbf{p}$, which should satisfy $\mathbf{r}^{\perp}\mathbf{p}$. If we were to let λ vary as a function of a parameter t (making $\mathbf{r}$ a function of t as well) we would find that this particular choice minimizes $\|\mathbf{r}(t)\|$.

Projection Onto a Subspace In general, supposing one has a subspace $S = \operatorname{span}\{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_n\}$ and (importantly!) the $\mathbf{y}_i$ are orthogonal, then the projection of $\mathbf{p}$ of x onto S is given by the sum of the projections onto each basis vector, yielding

$$\mathbf{p} = \sum_{i=1}^{n} \frac{\langle \mathbf{x},\ \mathbf{y}_i \rangle}{\langle \mathbf{y}_i,\ \mathbf{y}_i \rangle}\mathbf{y}_i = \sum_{i=1}^{n} \langle \mathbf{x},\ \mathbf{y}_i \rangle \widehat{\mathbf{y}_i}.$$

Note: this is part of why having an orthogonal basis is desirable!

Letting $A = [\mathbf{y}_1, \mathbf{y}_2, \cdots]$, then the following matrix projects vectors onto S, expressing them in terms of the basis $\mathbf{y}_i$[1]:

$$\tilde{P}_A = (AA^T)^{-1}A^T,$$

while this matrix performs the projection and expresses it in terms of the standard basis:

$$P_A = A(AA^T)^{-1}A^T.$$

Equation of a plane: given a point $\mathbf{p}_0$ on a plane and a normal vector $\mathbf{n}$, any vector $\mathbf{x}$ on the plane satisfies

$$\langle \mathbf{x} - \mathbf{p}_0,\ \mathbf{n} \rangle = 0$$

To find the distance between a point $\mathbf{a}$ and a plane, we need only project $\mathbf{a}$ onto the subspace spanned by the normal $\mathbf{n}$:

$$d = \langle \mathbf{a},\ \mathbf{n} \rangle.$$

[1]For a derivation of this formula, see the section on least-squares approximations.

One important property of projections is that for any vector $\mathbf{v}$ and for any subspace S, we have $\mathbf{v} - P_S(\mathbf{v}) \in S^\perp$. Moreover, if $\mathbf{v} \in S^\perp$, then $P_s(\mathbf{v})$ must be zero. This follows by noting that in equation **??**, every inner product appearing in the sum vanishes, by definition of $\mathbf{v} \in S^\perp$, and so the projection is zero.

Least Squares

Proposition 5.6.3 ***(Normal Equations).***
$\mathbf{x}$ is a least squares solution to $A\mathbf{x} = \mathbf{b}$ iff

$$A^t A\mathbf{x} = A^t\mathbf{b}$$

Derivation of normal equations.

The general setup here is that we would like to solve $A\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$, where $\mathbf{b}$ is not in fact in the range of A. We thus settle for a unique "best" solution $\tilde{\mathbf{x}}$ such that the error $\|A\tilde{\mathbf{x}} - \mathbf{b}\|$ is minimized.

Geometrically, the solution is given by projecting $\mathbf{b}$ onto the column space of A. To see why this is the case, define the residual vector $\mathbf{r} = A\tilde{\mathbf{x}} - \mathbf{b}$. We then seek to minimize $\|\mathbf{r}\|$, which happens exactly when $\mathbf{r}^\perp \operatorname{im} A$. But this happens exactly when $\mathbf{r} \in (\operatorname{im} A)^\perp$, which by the fundamental subspaces theorem, is equivalent to $\mathbf{r} \in \ker A^T$.

From this, we get the equation

$$\begin{aligned} A^T\mathbf{r} &= \mathbf{0} \\ \implies A^T(A\tilde{\mathbf{x}} - \mathbf{b}) &= \mathbf{0} \\ \implies A^T A\tilde{\mathbf{x}} &= A^T\mathbf{b}, \end{aligned}$$

where the last line is described as the **normal equations**.

If A is an $m \times n$ matrix and is of full rank, so it has n linearly independent columns, then one can show that $A^T A$ is nonsingular, and we thus arrive at the least-squares solution

$$\tilde{\mathbf{x}} = (A^T A)^{-1} A^T\mathbf{b}\blacksquare$$

These equations can also be derived explicitly using Calculus applied to matrices, vectors, and inner products. This requires the use of the following formulas:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{x}} \langle \mathbf{x},\ \mathbf{a} \rangle &= \mathbf{a} \\ \frac{\partial}{\partial \mathbf{x}} \langle \mathbf{x},\ \mathbf{A}\mathbf{x} \rangle &= (A + A^T)\mathbf{x} \end{aligned}$$

as well as the adjoint formula

$$\langle A\mathbf{x},\ \mathbf{x} \rangle = \left\langle \mathbf{x},\ A^T\mathbf{x} \right\rangle..$$

From these, by letting $A = I$ we can derive

$$\frac{\partial}{\partial \mathbf{x}} \|\mathbf{x}\|^2 = \frac{\partial}{\partial \mathbf{x}} \langle \mathbf{x},\ \mathbf{x} \rangle = 2\mathbf{x}$$

.

The derivation proceeds by solving the equation

$$\frac{\partial}{\partial \mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|^2 = \mathbf{0}..$$

5.6.4 Normal Forms

Remark 5.6.2.
Every square matrix is similar to a matrix in Jordan canonical form.

5.6.5 Decompositions

The QR Decomposition Gram-Schmidt is often computed to find an orthonormal basis for, say, the range of some matrix A. With a small modification to this algorithm, we can write $A = QR$ where R is upper triangular and Q has orthogonal columns.

Why is this useful? One reason is that this also allows for a particularly simple expression of least-squares solutions. If $A = QR$, then R will be invertible, and a bit of algebraic manipulation will show that

$$\tilde{\mathbf{x}} = R^{-1}Q^T\mathbf{b}..$$

How does it work? You simply perform Gram-Schmidt to obtain $\{\mathbf{u}_i\}$, then

$$Q = [\mathbf{u}_1, \mathbf{u}_2, \cdots].$$

The matrix R can then be written as

$$r_{ij} = \begin{cases} \langle \mathbf{u}_i,\ \mathbf{x}_j \rangle, & i \leq j, \\ 0, & \text{else.} \end{cases}$$

Explicitly, this yields the matrix

$$R = \begin{bmatrix} \langle \mathbf{u}_1,\ \mathbf{x}_1 \rangle & \langle \mathbf{u}_1,\ \mathbf{x}_2 \rangle & \langle \mathbf{u}_1,\ \mathbf{x}_3 \rangle & \cdots \\ 0 & \langle \mathbf{u}_2,\ \mathbf{x}_2 \rangle & \langle \mathbf{u}_2,\ \mathbf{x}_3 \rangle & \cdots \\ 0 & 0 & \langle \mathbf{u}_3,\ \mathbf{x}_3 \rangle & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

Explain shortcut for diagonal.

5.7 Appendix: Lists of things to know

Textbook: Leon, *Linear Algebra with Applications*

5.7.1 Topics

- 1.6: Partition Matrices
- 3.5: Change of Basis
- 4.1: Linear Transformations
- 4.2: Matrix Representations
- 4.3: Similarity
  - *Exam 1*

- 5.1: Scalar Product in $\mathbb{R}^n$
- 5.2: Orthogonal Subspaces
- 5.3: Least Squares
- 5.4: Inner Product Spaces
- 5.5: Orthonormal Sets
- 5.6: Gram-Schmidt
- 6.1: Eigenvalues and Eigenvectors
  - *Exam 2*
- 6.2: Systems of Linear Differential Equations
- 6.3: Diagonalization
- 6.6: Quadratic Forms
- 6.7: Positive Definite Matrices
- 6.5: Singular Value Decomposition
- 7.7: The Moore-Penrose Pseudo-Inverse
  - *Final Exam*

5.7.2 Definitions

- System of equations
- Homogeneous system
- Consistent/inconsistent system
- Matrix
- Matrix (i.e. $A\mathbf{x} = \mathbf{b}$)
- Inverse matrix
- Singular matrix
- Determinant
- Trace
- Rank
- Elementary row operation
- Row equivalence
- Pivot
- Row Echelon Form
- Reduced Row Echelon Form
- Gaussian elimination
- Block matrix
- Vector space
- Vector subspace
- Linear transformation
- Span

- Linear independence
- Basis
- Change of basis
- Dimension
- Row space
- Column space
- Image
- Null space
- Kernel
- Direct sum
- Projection
- Orthogonal subspaces
- Orthogonal complement
- Normal equations
- Least-squares solution
- Orthonormal
- Eigenvalue
- Eigenvector
- Characteristic polynomial
- Similarity
- Diagonalizable
- Inner product
- Bilinearity
- Multilinearity
- Defective
- Singular value decomposition
- QR factorization
- Gram-Schmidt process
- Spectral theorem
- Symmetric matrix
- Orthogonal matrix
- Positive-definite
- Quadratic form

5.7.3 Lower-division review

- Systems of linear equations
  - Consistent vs. Inconsistent
  - Possibilities for solutions
  - Geometric interpretation
- Matrix Inverses
  - Detecting if a matrix is singular
  - Computing the inverse
    - ◇ Formula for 2x2 case
    - ◇ Augment with the identity
    - ◇ Cramer's Rule
- Vector Spaces

  - Definition in terms of closures
  - Span
  - Linear Independence
  - Subspace and the subspace test
  - Basis
- Common Computations
  - Reduction to RREF
  - Eigenvalues and eigenvectors
  - Basis for the column space
  - Basis for the nullspace
  - Basis for the eigenspace
  - Construct matrix from a given linear map
  - Construct change of basis matrix
  - Construct matrix projection onto subspace
  - Convert a basis to an orthonormal basis

5.7.4 Things to compute

- Construct a matrix representing a linear map
  - With respect to the standard basis in both domain and range
  - With respect to a nonstandard basis in the range
  - With respect to a nonstandard basis in the domain
  - With respect to nonstandard bases in both the domain and range
- Construct a change of basis matrix
- Check that a map is a linear transformation
- Compute the following spaces of a matrix and their orthogonal complements:
  - Row space
  - Column space
  - Null space
- Compute the shortest distance between a point and a plane
- Compute the least squares solution to linear system
- Prove that something is a vector space
- Prove that a map is an inner product
- Compute determinants
- Compute the RREF of a matrix
- Compute characteristic polynomials, eigenvalues, and eigenvectors
- Diagonalize a matrix
- Solve a system of ODEs resulting arising from tank mixing
- Compute the singular value decomposition of a matrix
- Compute the rank and nullity of a matrix
- Convert a set of vectors to a basis
- Convert a basis to an orthonormal basis
- Determine if a matrix is diagonalizable
- Compute the matrix for a projection onto a subspace
- Find the QR factorization of a matrix

5.7.5 Things to prove

- Prove facts about block matrices
- Prove facts about injective linear maps
- Prove facts about similar matrices
- Prove facts about orthogonal spaces and orthogonal complements
- Prove facts about inner products
- Prove facts about orthonormal sets
- Prove facts about eigenvalues/eigenvectors
- Understand when a matrix can be diagonalized
- Prove facts about diagonalizable matrices
- Prove facts about the orthogonal decomposition theorem

5.8 Techniques Overview

$$p(y)y' = q(x) \qquad \text{separable}$$

$$y' + p(x)y = q(x) \qquad \text{integrating factor}$$

$$y' = f(x,y), f(tx,ty) = f(x,y) \qquad y = xV(x) \text{ COV reduces to separable}$$

$$y' + p(x)y = q(x)y^n \qquad \text{Bernoulli, divide by } y^n \text{ and COV } u = y^{1-n}$$

$$M(x,y)dx + N(x,y)dy = 0 \qquad M_y = N_x : \varphi(x,y) = c(\varphi_x = M, \varphi_y = N)$$

$$P(D)y = f(x,y) \qquad x^k e^{rx} \text{ for each root}$$

Where e^{zx} *yields* $e^{ax}\cos bx, e^{ax}\sin bx$

5.9 Ordinary Differential Equations

- Separable equations:

$$p(y)\frac{dy}{dx} - q(x) = 0 \implies \int p(y)dy = \int q(x)dx + C$$

$$\frac{dy}{dx} = f(x)g(y) \implies \int \frac{1}{g(y)}dy = \int f(x)dx + C$$

  - Population growth:

$$\frac{dP}{dt} = kP \implies \qquad P = P_0 e^{kt}$$

  - Logistic growth:
    - ♢ General form: $\dfrac{dP}{dt} = (B(t) - D(t))P(t)$

- ♢ Assume birth rate is constant $B(t) = B_0$ and death rate is proportional to instantaneous population $D(t) = D_0P(t)$. Then let $r = B_0, C = B_0/D_0$ be the *carrying capacity*:

$$\frac{dP}{dt} = r\left(1 - \frac{P}{C}\right)P \implies \qquad P(t) = \frac{P_0}{\frac{P_0}{C} + e^{-rt}(1 - \frac{P_0}{C})}$$

- First order linear:

$$\frac{dy}{dx} + p(x)y = q(x) \implies I(x) = e^{\int p(x)dx}, \qquad y(x) = \frac{1}{I(x)}\left(\int q(x)I(x)dx + C\right)$$

- Exact:
  - $M(x,y)dx + N(x,y)dy = 0$ is exact $\iff \exists\varphi : \dfrac{\partial\varphi}{\partial x} = M(x,y),\ \dfrac{\partial\varphi}{\partial y} = N(x,y)$ $\iff \dfrac{\partial M}{\partial y} = \dfrac{\partial N}{x}$
  - General solution:

$$\varphi(x,y) = \int^x M(s,y)ds + \int^y N(x,t)dt - \int^y \frac{\partial}{\partial t}\left(\int^x M(s,t)ds\right)dt$$

(where $\int^x f(t)dt$ means take the antiderivative of f and consider it a function of x)

- Cauchy Euler: #todo
- Bernoulli: todo

5.10 Linear Homogeneous

General form:

$$\begin{aligned} y^{(n)} + c_{n-1}y^{(n-1)} + \cdots + c_2y'' + cy' + cy = 0 \\ p(D)y = \prod(D - r_i)^{m_i}y = 0 \end{aligned}$$

where p is a polynomial in the differential operator D with roots r_i:

- Real roots: contribute m_i solutions of the form

$$e^{rx}, xe^{rx}, \cdots, x^{m_i-1}e^{rx}$$

- Complex conjugate roots: for $r = a + bi$, contribute $2m_i$ solutions of the form

$$\begin{aligned} e^{(a\pm bi)x}, xe^{(a\pm bi)x},\ \cdots,\ x^{m_i-1}e^{(a\pm bi)x} \\ = e^{ax}\cos(bx), e^{ax}\sin(bx),\ xe^{ax}\cos(bx), xe^{ax}\sin(bx),\ \cdots, \end{aligned}$$

Example: by cases, second order equation of the form

$$ay'' + by' + cy = 0$$

- Two distinct roots: $c_1e^{r_1x} + c_2e^{r_2x}$ - One real root: $c_1e^{rx} + c_2xe^{rx}$ - Complex conjugates $\alpha \pm i\beta$: $e^{\alpha x}(c_1\cos\beta x + c_2\sin\beta x)$

5.11 Linear Inhomogeneous

General form:

$$y^{(n)} + c_{n-1}y^{(n-1)} + \cdots + c_2 y'' + cy' + cy = F(x)$$
$$p(D)y = \prod (D - r_i)^{m_i} y = 0$$

Then solutions are of the form $y_c + y_p$, where y_c is the solution to the associated homogeneous system and y_p is a particular solution.

Methods of obtaining particular solutions

5.11.1 Undetermined Coefficients

- Find an operator $p(D)$ the annihilates $F(x)$ (so $q(D)F = 0$)
- Find solution of $q(D)p(D) = 0$, subtract of known solutions from homogeneous part to obtain the form of the trial solution $A_0 f(x)$, where A_0 is the undetermined coefficient
- Substitute trial solution into original equation to determine A_0

Useful Annihilators:

$$\begin{aligned}
&F(x) = p(x) : && D^{\deg(p)+1} \\
&F(x) = p(x)e^{ax} : && (D-a)^{\deg(p)+1} \\
&F(x) = \cos(ax) + \sin(ax) : && D^2 + a^2 \\
&F(x) = e^{ax}(a_0 \cos(bx) + b_0 \sin(bx)) : && (D-z)(D-\overline{z}) = D^2 - 2aD + a^2 + b^2 \\
&F(x) = p(x)e^{ax}\cos(bx) + p(x)e^{ax}\cos(bx) : && ((D-z)(D-\overline{z}))^{\max(\deg(p),\deg(q))+1}
\end{aligned}$$

5.11.2 Variation of Parameters

todo

5.11.3 Reduction of Order

todo

5.12 Systems of Differential Equations

General form:

$$\frac{\partial \mathbf{x}(t)}{\partial t} = A\mathbf{x}(t) + \mathbf{b}(t) \iff \mathbf{x}'(t) = A\mathbf{x}(t) + \mathbf{b}(t)$$

General solution to homogeneous equation:

$$c_1\mathbf{x_1}(t) + c_2\mathbf{x_2}(t) + \cdots + c_n\mathbf{x_n}(t) = \mathbf{X}(t)\mathbf{c}$$

If A is a matrix of constants: $\mathbf{x}(t) = e^{\lambda_i t}\,\mathbf{v}_i$ is a solution for each eigenvalue/eigenvector pair $(\lambda_i, \mathbf{v}_i)$ - If A is defective, you'll need generalized eigenvectors.

Inhomogeneous Equation: particular solutions given by

$$\mathbf{x}_p(t) = \mathbf{X}(t) \int^t \mathbf{X}^{-1}(s)\mathbf{b}(s)\ ds$$

5.13 Laplace Transforms

Definitions:

$$H_a(t) = \begin{cases} 0, & 0 \le t < a \\ 1, & t \ge a \end{cases}$$

$$\delta(t) : \int_{\mathbb{R}} \delta(t-a) f(t)\ dt = f(a), \quad \int_{\mathbb{R}} \delta(t-a)\ dt = 1$$

$$(f * g)(t) = \int_0^t f(t-s) g(s)\ ds$$

Useful property: for $a \le b$, $H_a(t) - H_b(t) = \mathbb{1}\,[[a,b]]$.

$$\begin{aligned}
t^n, n \in \mathbb{N} &\quad \Longleftrightarrow \quad n!\frac{1}{s^{n+1}}, \quad s > 0 \\
t^{-\frac{1}{2}} &\quad \Longleftrightarrow \quad \sqrt{\pi} s^{-\frac{1}{2}} \quad s > 0 \\
e^{at} &\quad \Longleftrightarrow \quad \frac{1}{s-a}, \quad s > a \\
\cos(bt) &\quad \Longleftrightarrow \quad \frac{s}{s^2+b^2}, \quad s > 0 \\
\sin(bt) &\quad \Longleftrightarrow \quad \frac{b}{s^2+b^2}, \quad s > 0 \\
\delta(t-a) &\quad \Longleftrightarrow \quad e^{-as} \\
H_a(t) &\quad \Longleftrightarrow \quad s^{-1} e^{-as} \\
e^{at} f(t) &\quad \Longleftrightarrow \quad F(s-a) \\
H_a(t) f(t-a) &\quad \Longleftrightarrow \quad e^{-as} F(s) \\
f'(t) &\quad \Longleftrightarrow \quad sL(f) - f(0) \\
f''(t) &\quad \Longleftrightarrow \quad s^2 L(f) - s f(0) - f'(0) \\
f^{(n)}(t) &\quad \Longleftrightarrow \quad s^n L(f) - \sum_{i=0}^{n-1} s^{n-1-i} f^{(i)}(0) \\
f(t) g(t) &\quad \Longleftrightarrow \quad F(s) * G(s)
\end{aligned}$$

- For f periodic with period T, $L(f) = \dfrac{1}{1+e^{-sT}} \displaystyle\int_0^T e^{-st} f(t)\ dt$

$p(y)y' = q(x)$	separable
$y' + p(x)y = q(x)$	integrating factor
$y' = f(x,y), f(tx,ty) = f(x,y)$	$y = xV(x)$ COV reduces to separable
$y' + p(x)y = q(x)y^n$	Bernoulli, divide by y^n and COV $u = y^{1-n}$
$M(x,y)dx + N(x,y)dy = 0$	$M_y = N_x : \varphi(x,y) = c(\varphi_x = M, \varphi_y = N)$
$P(D)y = f(x,y)$	$x^k e^{rx}$ for each root

5.14 Systems of Differential Equations

Definition 5.14.1 (Wronksian)**.**
For a collection of n functions $f_i : \mathbb{R}^n \to \mathbb{R}$, define the $n \times 1$ column vector

$$W(f_i)(\mathbf{p}) \coloneqq \begin{bmatrix} f_i(\mathbf{p}) \\ f_i'(\mathbf{p}) \\ f_i''(\mathbf{p}) \\ \vdots \\ f^{(n-1)}(\mathbf{p}) \end{bmatrix}.$$

The **Wronskian** of this collection is defined as

$$W(f_1, \cdots, f_n)(\mathbf{p}) \coloneqq \det \begin{bmatrix} | & | & & | \\ W(f_1)(\mathbf{p}) & W(f_2)(\mathbf{p}) & \cdots & W(f_n)(\mathbf{p}) \\ | & | & & | \end{bmatrix}.$$

Proposition 5.14.1 ***(Wronskian detects linear dependence of functions).***
A set of functions $\{f_i\}$ is linearly independent on $I \iff \exists x_0 \in I : W(x_0) \neq 0$.

⚠**Warning 5.2:** $W \equiv 0$ on I does *not* imply that $\{f_i\}$ is linearly dependent! Counterexample: $\left\{x, x+x^2, 2x-x^2\right\}$ where $W \equiv 0$ but $x + x^2 = 3(x) + (2x - x^2)$ is a linear combination of the other two functions.

Sufficient condition: each f_i is the solution to a linear homogeneous ODE $L(y) = 0$.

5.15 To Sort

- Burnside's Lemma
- Cauchy's Theorem

- If $|G| = n = \prod p_i^{k_i}$, then for each i there exists a subgroup H of order p_i.

- The Sylow Theorems
  - If $|G| = n = \prod p_i^{k_i}$, for each ii and each $1 \leq k_j \leq k_i$ then there exists a subgroup H of order $p_i^{k_j}$.
- Galois Theory
- More terms: `http://mathroughguides.wikidot.com/glossary:abstract-algebra`
- Order p: One, Z_p
- Order p^2: Two abelian groups, Z_{p^2}, Z_p^2
- Order p^3:
  - 3 abelian $Z_{p^3}, Z_p \times Z_{p^2}.Z_p^3$,
  - 2 others $Z_p \rtimes Z_{p^2}$.
    - ◇ The other is the quaternion group for p = 2 and a group of exponent p for p > 2.
- Order pq:
  - $p \,\Big|\, q-1$: Two groups, Z_{pq} and $Z_q \rtimes Z_p$
  - Else cyclic, Z_{pq}
- Every element in a permutation group is a product of disjoint cycles, and the order is the lcm of the order of the cycles.
- The product ideal IJ is *not* just elements of the form ij, it is all sums of elements of this form! The product alone isn't enough.
- The intersection of any number of ideals is also an ideal

5.16 Big List of Notation

$$
\begin{array}{rrrr}
C(x) = & \left\{g \in G : gxg^{-1} = x\right\} & \subseteq G & \text{Centralizer} \\
C_G(x) = & \left\{gxg^{-1} : g \in G\right\} & \subseteq G & \text{Conjugacy Class} \\
G_x = & \{g.x : x \in X\} & \subseteq X & \text{Orbit} \\
x_0 = & \{g \in G : g.x = x\} & \subseteq G & \text{Stabilizer} \\
Z(G) = & \left\{x \in G : \forall g \in G,\ gxg^{-1} = x\right\} & \subseteq G & \text{Center} \\
\mathrm{Inn}(G) = & \left\{\varphi_g(x) = gxg^{-1}\right\} & \subseteq \mathrm{Aut}(G) & \text{Inner Aut.} \\
\mathrm{Out}(G) = & \mathrm{Aut}(G)/\mathrm{Inn}(G) & \hookrightarrow \mathrm{Aut}(G) & \text{Outer Aut.} \\
N(H) = & \left\{g \in G : gHg^{-1} = H\right\} & \subseteq G & \text{Normalizer}
\end{array}
$$

5.17 Group Theory

Notation: $H < G$ a subgroup, $N < G$ a normal subgroup, concatenation is a generic group operation.

- $\mathbb{Z}_n$ the unique cyclic group of order n
- $\mathbf{Q}$ the quaternion group
- $G^n = G \times G \times \cdots G$
- $Z(G)$ the center of G
- $o(G)$ the order of a group
- S_n the symmetric group
- A_n the alternating group
- D_n the dihedral group of order $2n$
- Group Axioms
  - Closure: $a, b \in G \implies ab \in G$
  - Identity: $\exists e \in G \mid a \in G \implies ae = ea = a$
  - Associativity: $a, b, c \in G \implies (ab)c = a(bc)$
  - Inverses: $a \in G \implies \exists b \in G \mid ab = ba = e$
- Definitions:
  - Order
    - ◊ Of a group: $o(G) = |G|$, the cardinality of G
    - ◊ Of an element: $o(g) = \min \{n \in \mathbb{N} : g^n = e\}$
  - Index
  - Center: the elements that commute with everything
  - Centralizer: all elements that commute with a given element/subgroup.
  - Group Action: a function $f : X \times G \to G$ satisfying
    - ◊ $x \in X, g_1, g_2 \in G \implies g_1.(g_2.x) = (g_1 g_2).x$
    - ◊ $x \in X \implies e.x = x$
  - Orbits partition any set
  - Transitive Action
  - Conjugacy Class: $C \subset G$ is a conjugacy class $\iff$
    - ◊ $x \in C, g \in G \implies gxg^{-1} \in C$
    - ◊ $x, y \in C \implies \exists g \in G : gxg^{-1} = y$
    - ◊ i.e. subsets that are closed under G acting on itself by conjugation and on which the action is transitive
    - ◊ i.e. orbits under the conjugation action
    - ◊ The order of any conjugacy class divides the order of G
  - p-group: Any group of order p^n.
  - Simple Group: no nontrivial normal subgroups
  - Normal Series: $0 \trianglelefteq H_0 \trianglelefteq H_1 \cdots \trianglelefteq G$
  - Composition Series: The successive quotients of the normal series
  - Solvable: G is solvable $\iff$ G has an abelian composition series.

- One step subgroup test:

$$a, b \in H \implies ab^{-1} \in H$$

.

- Useful isomorphism invariants:
  - Order profile of elements: n_1 elements of order p_1, n_2 elements of order p_2, etc
    - ♢ Useful to look at elements of order 2!
  - Order profile of subgroups
  - $Z(A) \cong Z(B)$
  - Number of generators (generators are sent to generators)
  - Number and size of conjugacy classes
  - Number of Sylow-p subgroups.
  - Commutativity
  - "Being cyclic"
  - Automorphism Groups
  - Solvability
  - Nilpotency
- Useful homomorphism invariants
  - $\varphi(e) = e$
  - $|g| = m < \infty \implies |\varphi(g)| = m$
  - Inverses, i.e. $\varphi(a)^{-1} = \varphi(a^{-1})$
  - $H < G \implies \varphi(H) < G'$
    - ♢ $H' < G' \implies \varphi^{-1}(H') < G$
  - $|G| < \infty \implies \varphi(G)$ divides $|G|, |G'|$

5.18 Big Theorems

- Classification of Abelian Groups

$$G \cong \mathbb{Z}_{p_1^{k_1}} \oplus \mathbb{Z}_{p_2^{k_2}} \oplus \cdots \oplus \mathbb{Z}_{p_n^{k_n}},$$

  where (p_i, k_i) are the set of elementary divisors of G.

- Isomorphism Theorems

$$\begin{aligned} \varphi : G \to G' \implies & \quad \frac{G}{\ker \varphi} \cong \varphi(G) \\ H \trianglelefteq G,\ K < G \implies & \quad \frac{K}{H \cap K} \cong \frac{HK}{H} \\ H, K \trianglelefteq G,\ K < H \implies & \quad \frac{G/K}{H/K} \cong \frac{G}{H} \end{aligned}$$

- Lagrange's Theorem: $H < G \implies o(H) \,\Big|\, o(G)$
  - Converse is false: $o(A_4) = 12$ but has no order 6 subgroup.

- The GZ Theorem: $G/Z(G)$ cyclic implies that $G \in \mathbf{Ab}$.
- Orbit Stabilizer Theorem: $G/x_0 \cong Gx$
- The Class Equation
  - Let $G \curvearrowright X$ and $\mathcal{O}_i \subseteq X$ be the nontrivial orbits, then

$$|X| = |X_0| + \sum_{[x_i] \in X/G} |Gx|.$$

  - The right hand side is the number of fixed points, plus a sum over all of the orbits of size greater than 1, where any representative within the orbit is chosen and we look at the index of its stabilizer in G.
  - Let $G \curvearrowright G$ and for each nontrivial conjugacy class C_G choose a representative $[x_i] = C_G = C_G(x_i)$ to obtain

$$|G| = |Z(G)| + \sum_{[x_i] = C_G(x_i)} [G : [x_i]] \,.$$

- Useful facts:
  - $H < G \in \mathbf{Ab} \implies H \trianglelefteq G$
    - ♢ Converse doesn't hold, even if all subgroups are normal. Counterexample: $\mathbf{Q}$
  - $G/Z(G) \cong \mathrm{Inn}(G)$
  - $H, K < G$ with $H \cong K \;\not\!\!\!\implies\; G/H \cong G/K$
    - ♢ Counterexample: $G = \mathbb{Z}_4 \times \mathbb{Z}_2, H =< (0,1) >, K =< (2,0) >$. Then $G/H \cong \mathbb{Z}_4 \not\cong \mathbb{Z}_2^2 \cong G/K$
  - $G \in \mathbf{Ab} \implies$ for each p dividing $o(G)$, there is an element of order p
  - Any surjective homomorphism $\varphi : A \twoheadrightarrow B$ where $o(A) = o(B)$ is an isomorphism
  - If G is abelian, for each $d \,\Big|\, |G|$ there is exactly one subgroup of order d.
- Sylow Subgroups:
  - Todo
- Big List of Interesting Groups
  - $\mathbb{Z}_4, \mathbb{Z}_2^2$
  - D_4
  - $Q = \langle a, b | a^4 = 1, a^2 = b^2, ab = ba^3 \rangle$ the quaternion group
  - S^3, the smallest nonabelian group
- Chinese Remainder Theorem:

$$\mathbb{Z}_{pq} \cong \mathbb{Z}_p \oplus \mathbb{Z}_q \iff (p, q) = 1$$

  - Fundamental Theorem of Finitely Generated Abelian Groups:
  - $G = \mathbb{Z}^n \oplus \bigoplus \mathbb{Z}_{q_i}$
- Finding all of the unique groups of a given order: #todo

5.18.1 Cyclic Groups

- Generated by ?
- For each d dividing $o(G)$, there exists a subgroup H of order d.
  - If $G =< a >$, then take $H =< a^{\frac{n}{d}} >$

5.18.2 The Symmetric Group

- Generated by:
  - Transpositions
  - #todo
- Cycle types: characterized by the number of elements in the cycle.
  - Two elements are in the same conjugacy class $\iff$ they have the same cycle type.
- Inversions: given $\tau = (p_1 \cdots p_n)$, a pair p_i, p_j is *inverted* iff $i < j$ but $p_j < p_i$
- Can count inversions $N(\tau)$
  - Equal to minimum number of transpositions to obtain non-decreasing permutation
- Sign of a permutation: $\sigma(\tau) = (-1)^{N(\tau)}$
- Parity of permutations $\cong (\mathbb{Z}, +)$
  - even $\circ$ even = even
  - odd $\circ$ odd = even
  - even $\circ$ odd = odd

5.19 Ring Theory

Ring Axioms

- Examples:
- Non-Examples:
- Definition of an Ideal
- Definitions of types of rings:
  - Field
  - Unique Factorization Domain (UFD)
  - Principal Ideal Domain (PID)
  - Euclidean Domain:
  - Integral Domain
  - Division Ring

$$\text{field} \implies \text{Euclidean Domain} \implies \text{PID} \implies \text{UFD} \implies \text{integral domain.}$$

- Counterexamples to inclusions are strict:
  - An ED that is not a field:
  - A PID that is not an ED: $\mathbb{Q}[\sqrt{19}]$
  - A UFD that is not a PID:
  - An integral domain that is not a UFD:
- Integral Domains
- Unique Factorization Domains
- Prime Elements
- Prime Ideals
- Field Extensions

- The Chinese Remainder Theorem for Rings
- Polynomial Rings
  - Irreducible Polynomials
    - ♢ Over $\mathbb{Z}_2$:

$$x,\ x+1,\ x^2+x+1,\ x^3+x+1,\ x^3+x^2+1.$$

    - ♢ Eisenstein's Criterion
- Gauss' Lemma

When is $\mathbb{Q}(\sqrt{d})$ a field?

6 Number Theory

6.1 Notation and Basic Definitions

$$\begin{aligned} (a,b) &:= \gcd(a,b) && \text{the greatest common divisor} \\ \mathbb{Z}_n & && \text{the ring of integers} \mod n \\ \mathbb{Z}_n^\times & && \text{the group of units} \mod n. \end{aligned}$$

Definition 6.1.1 (Multiplicative Functions)**.**
A function $f : \mathbb{Z} \to \mathbb{Z}$ is said to be **multiplicative** iff

$$(a,b) = 1 \implies f(ab) = f(a)f(b).$$

6.2 Big Theorems

Link to theorems.

6.3 Primes

Theorem 6.3.1 ***(The fundamental theorem of arithmetic).***
Every $n \in \mathbb{Z}$ can be written uniquely as

$$n = \prod_{i=1}^m p_i^{k_i}$$

where the p_i are the m distinct prime divisors of n.

Remark 6.3.1.
Note that the number of distinct prime factors is m, while the total number of factors is $\prod_{i=1}^m (k_i + 1)$.

6.4 Divisibility

Definition 6.4.1 (Divisibility)**.**

$$a \mid b \iff b \equiv 0 \mod a \iff \exists k \text{ such that } ak = b$$

6.4.1 $\mathrm{gcd}, \mathrm{lcm}$

Remark 6.4.1.
$\gcd(a, b)$ can be computed by taking prime factorizations of a and b, intersecting the primes occurring, and taking the lowest exponent that appears. Dually, $\mathrm{lcm}(a, b)$ can be computed by taking the *union* and the *highest* exponent.

Check

Proposition 6.4.1 ***(Relationship between*** $\gcd$ ***and*** lcm***).***

$$xy = \gcd(x, y)\ \mathrm{lcm}(x, y)$$

Proposition 6.4.2 ***(?).***
If $d \mid x$ and $d \mid y$, then

$$\gcd(x, y) = d \cdot \gcd\left(\frac{x}{d}, \frac{y}{d}\right)$$
$$\mathrm{lcm}(x, y) = d \cdot \mathrm{lcm}\left(\frac{x}{d}, \frac{y}{d}\right)$$

Check

Proposition 6.4.3 ***(Useful properties of*** $\gcd$***).***

$$\gcd(x, y, z) = \gcd(\gcd(x, y), z)$$
$$\gcd(x, y) = \gcd(x \bmod y, y)$$
$$\gcd(x, y) = \gcd(x - y, y).$$

6.4.2 The Euclidean Algorithm

$\gcd(a, b)$ can be computed via the Euclidean algorithm, taking the final bottom-right coefficient.

Example of Euclidean algorithm,

6.5 Modular Arithmetic

Generally concerned with the multiplicative group $(\mathbb{Z}_n, \times)$.

Theorem 6.5.1 ***(The Chinese Remainder Theorem).***
The system

$$\begin{aligned} x &\equiv a_1 \mod m_1 \\ x &\equiv a_2 \mod m_2 \\ &\vdots \\ x &\equiv a_r \mod m_r \end{aligned}$$

has a unique solution $x \mod \prod m_i \iff (m_i, m_j) = 1$ for each pair i, j, given by

$$x = \sum_{j=1}^{r} a_j \frac{\prod_i m_i}{m_j} \left[\frac{\prod_i m_i}{m_j}\right]^{-1}_{\mod m_j}.$$

Theorem 6.5.2 ***(Euler's Theorem).***

$$a^{\varphi(p)} \equiv 1 \mod n.$$

Theorem 6.5.3 ***(Fermat's Little Theorem).***

$$\begin{aligned} x^p &\equiv x \mod p \\ x^{p-1} &\equiv 1 \mod p \quad \text{if } p \nmid a \end{aligned}$$

6.5.1 Diophantine Equations

Proposition 6.5.1 ***(Solutions to linear Diophantine equations).***
Consider $ax + by = c$. This has solutions iff $c = 0 \mod (a, b) \iff \gcd(a, b)$ divides c.

How to obtain solutions.

6.5.2 Computations

Proposition 6.5.2 ***(?).***
If $x \equiv 0 \mod n$, then $x \equiv 0 \mod p^k$ for all p^k appearing in the prime factorization of n.

Remark 6.5.1.
If there are factors of the modulus in the equation, peel them off with addition, using the fact that $nk \equiv 0 \mod n$.

$$\begin{aligned} x &\equiv nk + r \mod n \\ &\equiv r \mod n \end{aligned}$$

So take $x = 463, n = 4$, then use $463 = 4 \cdot 115 + 4$ to write

$$\begin{aligned} 463 &\equiv y \mod 4 \\ \implies 4 \cdot 115 + 3 &\equiv y \mod 4 \\ \implies 3 &\equiv y \mod 4. \end{aligned}$$

Proposition 6.5.3 ***(Repeated square/fast exponentiation).***
For any n,

$$x^k \mod n \equiv (x^{k/d} \bmod n)^d \mod n.$$

Example 6.5.1 (?).

$$\begin{aligned} 2^{25} &\equiv (2^5 \mod 5)^5 \mod 5 \\ &\equiv 2^5 \mod 5 \\ &\equiv 2 \mod 5 \end{aligned}$$

Remark 6.5.2.
Make things easier with negatives! For example, $\mod 5$,

$$\begin{aligned} 4^{25} &\equiv (-1)^{25} \mod 5 \\ &\equiv (-1) \mod 5 \\ &\equiv 4 \mod 5 \end{aligned}$$

6.5.3 Invertibility

Proposition 6.5.4 ***(Reduction of modulus).***

$$xa = xb \mod n \implies a = b \mod \frac{n}{(x,n)}.$$

Proposition 6.5.5 ***(Characterization of invertibility).***

$$x \in \mathbb{Z}_n^\times \iff (x,n) = 1,$$

and thus

$$\mathbb{Z}_n^\times = \{1 \le x \le n : (x,n) = 1\}$$

and $|\mathbb{Z}_n^\times| = \varphi(n)$.

Example 6.5.2 (Using invertibility).
One can reduce equations by dividing through by a unit. Pick any x such that $x \mid a$ and $x \mid b$ with $(x, n) = 1$, then

$$a = b \mod n \implies \frac{a}{x} = \frac{b}{x} \mod n.$$

6.6 The Totient Function

Definition 6.6.1 (Euler's Totient Function).

$$\varphi(n) = |\{1 \leq x \leq n : (x, n) = 1\}|$$

Example 6.6.1 (?).

$$\begin{aligned}
\varphi(1) &= |\{1\}| = 1 \\
\varphi(2) &= |\{1\}| = 1 \\
\varphi(3) &= |\{1, 2\}| = 2 \\
\varphi(4) &= |\{1, 3\}| = 2 \\
\varphi(5) &= |\{1, 2, 3, 4\}| = 4
\end{aligned}$$

Proposition 6.6.1 ***(Formulas involving the totient).***

$$\begin{aligned}
\varphi(p) &= p - 1 \\
\varphi(p^k) &= p^{k-1}(p - 1) \\
\varphi(n) &= n \prod_{i=1}^{?} \left(1 - \frac{1}{p_i}\right) \\
n &= \sum_{d \mid n} \varphi(d)
\end{aligned}$$

Proof (?).
All numbers less than p are coprime to p; there are p^k numbers less than p^k and the only numbers *not* coprime to p^k are multiples of p, i.e. $\left\{p, p^2, \cdots p^{k-1}\right\}$ of which there are $k - 1$, yielding $p^k - p^{k-1}$
Along with the fact that φ is multiplicative, so $(p, q) = 1 \implies \varphi(pq) = \varphi(p)\varphi(q)$, compute this for any n by taking the prime factorization.

With these properties, one can compute:

$$\begin{aligned}\varphi(n) &= \varphi\left(\prod_i p_i^{k_i}\right) \\ &= \prod_i p_i^{k_i-1}(p_i-1) \\ &= n\left(\frac{\prod_i(p_i-1)}{\prod_i p_i}\right) \\ &= n\prod_i\left(1-\frac{1}{p_i}\right)\end{aligned}$$

\todo[inline]{Check and explain}

■

6.7 Quadratic Residues

Definition 6.7.1 (Quadratic Residue)**.**
x is a **quadratic residue** mod$n$ iff there exists an a such that $a^2 = x \mod n$.

Proposition 6.7.1*(?).*
In $\mathbb{Z}_p$, exactly half of the elements (even powers of generator) are quadratic residues.

Proposition 6.7.2*(?).*

$$-1 \text{ is a quadratic residue in } \mathbb{Z}_p \iff p = 1 \mod 4.$$

Definition 6.7.2 (The Jacobi Symbols)**.**

todo

Definition 6.7.3 (The Legendre Symbol)**.**

todo

6.8 Primality Tests

Proposition 6.8.1*(Fermat Primality Test).*
If n is prime, then

$$a^{n-1} = 1 \mod n$$

Proposition 6.8.2 ***(Miller-Rabin Primality Test).***
n is prime iff

$$x^2 = 1 \mod n \implies x = \pm 1$$

7 Real Analysis

7.1 Notation

Definition 7.1.1 (Continuously Differentiable)**.**
A function is **continuously differentiable** iff f is differentiable and f' is continuous.
Conventions:

- *Integrable* means *Riemann integrable.*

f	a functional $\mathbb{R}^n \to \mathbb{R}$
$\mathbf{f}$	a function $\mathbb{R}^n \to \mathbb{R}^m$
A, E, U, V	open sets
A'	the limit points of A
$\overline{A}$	the closure of A
$A^\circ := A \setminus A'$	the interior of A
K	a compact set
$\mathcal{R}_A$	the space of Riemann integral functions on A
$C^j(A)$	the space of j times continuously differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$
$\{f_n\}$	a sequence of functions
$\{x_n\}$	a sequence of real numbers
$f_n \to f$	pointwise convergence
$f_n \rightrightarrows f$	uniform convergence
$x_n \nearrow x$	$x_i \leq x_j$ and x_j converges to x
$x_n \searrow x$	$x_i \geq x_j$ and x_j converges to x
$\sum_{k \in \mathbb{N}} f_k$	a series
$D(f)$	the set of discontinuities of f.

7.2 Big Ideas

Summary for GRE:

- Limits,
- Continuity,

- Boundedness,
- Compactness,
- Definitions of topological spaces,
- Lipschitz continuity
- Sequences and series of functions.
- Know the interactions between the following major operations:
  - Continuity (pointwise limits)
  - Differentiability
  - Integrability
  - Limits of sequences
  - Limits of series/sums
- The derivative of a continuous function need not be continuous
- A continuous function need not be differentiable
- A uniform limit of differentiable functions need not be differentiable
- A limit of integrable functions need not be integrable
- An integrable function need not be continuous
- An integrable function need not be differentiable

Theorem 7.2.1 ***(Generalized Mean Value Theore).***

$$f, g \text{ differentiable on } [a,b] \implies \exists c \in [a,b] : [f(b) - f(a)]\, g'(c) = [g(b) - g(a)]\, f'(c)$$

Corollary 7.2.1 ***(Mean Value Theorem).***

?

todo

7.3 Important Examples

7.4 Commuting Limits

- Suppose $f_n \to f$ (pointwise, not necessarily uniformly)
- Let $F(x) = \displaystyle\int f(t)$ be an antiderivative of f
- Let $f'(x) = \dfrac{\partial f}{\partial x}(x)$ be the derivative of f.

Then consider the following possible ways to commute various limiting operations:

Does taking the derivative of the integral of a function always return the original function?

$$[\frac{\partial}{\partial x}, \int dx] : \qquad \frac{\partial}{\partial x} \int f(x,t) dt =_? \int \frac{\partial}{\partial x} f(x,t) dt$$

Answer: Sort of (but possibly not).

Counterexample:

$$f(x) = \begin{cases} 1 & x > 0 \\ -1 & x \leq 0 \end{cases} \implies \int f \approx |x|,$$

which is not differentiable. (This is remedied by the so-called "weak derivative")

Sufficient Condition: If f is continuous, then both are always equal to $f(x)$ by the FTC.

Is the derivative of a continuous function always continuous?

$$[\frac{\partial}{\partial x}, \lim_{x_i \to x}] : \qquad \lim_{x_i \to x} f'(x_n) =_? f'(\lim_{x_i \to x} x)$$

Answer: No.

Counterexample:

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases} \implies f'(x) = \begin{cases} 2x \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

which is discontinuous at zero.

Sufficient Condition: There doesn't seem to be a general one (which is perhaps why we study C^k functions).

Is the limit of a sequence of differentiable functions differentiable **and** the derivative of the limit?

$$[\frac{\partial}{\partial x}, \lim_{f_n \to f}] : \qquad \lim_{f_n \to f} \frac{\partial}{\partial x} f_n(x) =_? \frac{\partial}{\partial x} \lim_{f_n \to f} f_n(x)$$

Answer: *Super* no – even the uniform limit of differentiable functions need not be differentiable!

Counterexample: $f_n(x) = \dfrac{\sin(nx)}{\sqrt{n}} \rightrightarrows f = 0$ but $f_n' \not\to f' = 0$

Sufficient Condition: $f_n \rightrightarrows f$ and $f_n \in C^1$.

Is the limit of a sequence of integrable functions integrable **and** the integral of the limit?

$$[\int dx, \lim_{f_n \to f}](f) : \qquad \lim_{f_n \to f} \int f_n(x) dx =_? \int \lim_{f_n \to f} f_n(x) dx$$

Answer: No.

Counterexample: Order $\mathbb{Q} \cap [0,1]$ as $\{q_i\}_{i\in\mathbb{N}}$, then take

$$f_n(x) = \sum_{i=1}^{n} \mathbb{1}\left[q_n\right] \to \mathbb{1}\left[\mathbb{Q} \cap [0,1]\right]$$

where each f_n integrates to zero (only finitely many discontinuities) but f is not Riemann-integrable.

Sufficient Condition: - $f_n \rightrightarrows f$, or - f integrable and $\exists M : \forall n, |f_n| < M$ (f_n uniformly bounded)

Is the integral of a continuous function also continuous?

$$[\int dx, \lim_{x_i \to x}] : \qquad \lim_{x_i \to x} F(x_i) =_? F(\lim_{x_i \to x} x_i)$$

Answer: Yes.

Proof: $|f(x)| < M$ on I, so given c pick a sequence $x \to c$. Then

$$|f(x)| < M \implies \left|\int_c^x f(t)dt\right| < \int_c^x M dt \implies |F(x) - F(c)| < M(b-a) \to 0$$

Is the limit of a sequence of continuous functions also continuous?

$$[\lim_{x_i \to x}, \lim_{f_n \to f}] : \qquad \lim_{f_n \to f} \lim_{x_i \to x} f(x_i) =_? \lim_{x_i \to x} \lim_{f_n \to f} f_n(x_i)$$

Answer: No.

Counterexample: $f_n(x) = x^n \to \delta(1)$

Sufficient Condition: $f_n \rightrightarrows f$

Does a sum of differentiable functions necessarily converge to a differentiable function?

$$\left[\frac{\partial}{\partial x}, \sum_{f_n}\right] : \qquad \frac{\partial}{\partial x} \sum_{k=1}^{\infty} f_k =_? \sum_{k=1}^{\infty} \frac{\partial}{\partial x} f_k$$

Answer: No.

Counterexample: $f_n(x) = \dfrac{\sin(nx)}{\sqrt{n}} \rightrightarrows 0 := f$, but $f_n' = \sqrt{n}\cos(nx) \not\to 0 = f'$ (at, say, $x = 0$)

Sufficient Condition: When $f_n \in C^1, \exists x_0 : f_n(x_0) \to f(x_0)$ and $\sum \|f_n'\|_\infty < \infty$ (continuously differentiable, converges at a point, and the derivatives absolutely converge)

7.5 Continuity

Definition 7.5.1 (Limit definition of continuity)**.**

$$f \text{ continuous} \iff \lim_{x \to p} f(x) = f(p)$$

Definition 7.5.2 (ε-δ definition of continuity)**.**

$$f : (X, d_X) \to (Y, d_Y) \text{ continuous} \iff \forall \varepsilon,\ \exists \delta \;\middle|\; d_X(x, y) < \delta \implies d_Y(f(x), f(y)) < \varepsilon$$

Example 7.5.1 (A nonobviously discontinuous function)**.**

$$f(x) = \sin\left(\frac{1}{x}\right) \implies 0 \in D(f)$$

Proof (?).

todo

■

Example 7.5.2 (The Dirichlet function)**.**
The Dirichlet function is nowhere continuous:

$$f(x) = \mathbb{1}[\mathbb{Q}]$$

Proposition 7.5.1 ***(Thomae's function: the set of points of continuity and of discontinuity can both be infinite).***
The following function continuous at infinitely many points and discontinuous at infinitely many points:

$$f(x) = \begin{cases} 0 & x \in \mathbb{R} \setminus \mathbb{Q} \\ \dfrac{1}{q} & x = \dfrac{p}{q} \in \mathbb{Q} \end{cases}$$

Then f is discontinuous on $\mathbb{Q}$ and continuous on $\mathbb{R} \setminus \mathbb{Q}$.

Proof (?).
f is continuous on $\mathbb{Q}$:

- Fix ε, let $x_0 \in \mathbb{R} - \mathbb{Q}$, choose $n : \frac{1}{n} < \varepsilon$ using Archimedean property.
  - Define $S = \left\{ x \in \mathbb{Q} : x \in (0,1), x = \frac{m}{n'}, n' < n \right\}$
  - Then $|S| \leq 1 + 2 + \cdots (n-1)$, so choose $\delta = \min_{s \in S} |s - x_0|$
  - Then

$$x \in N_\delta(x_0) \implies f(x) < \frac{1}{n} < \varepsilon.$$

f is discontinuous on $\mathbb{R} \setminus \mathbb{Q}$:

- Let $x_0 = \frac{p}{q} \in \mathbb{Q}$ and $\{x_n\} = \left\{ x - \frac{1}{n\sqrt{2}} \right\}$. Then

$$x_n \uparrow x_0 \text{ but } f(x_n) = 0 \to 0 \neq \frac{1}{q} = f(x_0)$$

■

Remark 7.5.1.
There are no functions that are continuous on $\mathbb{Q}$ but discontinuous on $\mathbb{R} - \mathbb{Q}$

Definition 7.5.3 (Uniform Continuity)**.**

todo

Definition 7.5.4 (Absolute Continuity)**.**

Theorem 7.5.1 ***(Extreme Value Theorem).***
A continuous function on a compact space attains its extrema.

7.6 Differentiability

$$f'(p) := \frac{\partial f}{\partial x}(p) = \lim_{x \to p} \frac{f(x) - f(p)}{x - p}$$

- For multivariable functions: existence **and continuity** of $\frac{\partial \mathbf{f}}{\partial x_i} \forall i \implies \mathbf{f}$ differentiable
  - Necessity of continuity: example of a continuous functions with all partial and directional derivatives that is not differentiable:

$$f(x,y) = \begin{cases} \frac{y^3}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & \text{else} \end{cases}.$$

7.6.1 Properties, strongest to weakest

$$C^\infty \subsetneq C^k \subsetneq \text{ differentiable } \subsetneq C^0 \subset \mathcal{R}_K.$$

- Example showing $f \in C^0 \implies\!\!\!\!\!\!/\ \ f$ is differentiable **and** f not differentiable $\not\Longrightarrow f \notin C^0$.
  - Take $f(x) = |x|$ at $x = 0$.
- Example showing that f differentiable $\not\Longrightarrow f \in C^1$:
  - Take

$$f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases} \implies f'(x) = \begin{cases} -\cos\left(\frac{1}{x}\right) + 2x\sin\left(\frac{1}{x}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}$$

but $\lim_{x\to 0} f'(x)$ does not exist and thus f' is not continuous at zero.

Proof that f differentiable $\implies f \in C^0$:

$$f(x) - f(p) = \frac{f(x) - f(p)}{x - p}(x - p) \overset{\text{hypothesis}}{=} f'(p)(x - p) \overset{x \to p}{\rightrightarrows} 0$$

7.7 Giant Table of Relations

Bold are assumed hypothesis, regular text is the strongest conclusion you can reach, strikeout denotes implications that aren't necessarily true.

f'	f	$\therefore f$	F
~~exists~~	**continuous**	K-integrable	exists
~~continuous~~	**differentiable**	continuous	exists
~~exists~~	**integrable**	~~continuous~~	differentiable

Explanation of items in table:

- K-integrable: compactly integrable.
- f integrable $\implies F$ differentiable $\implies F \in C_0$
  - By definition and FTC, and differentiability $\implies$ continuity
- f differentiable and K compact $\implies f$ integrable on K.
  - In general, f differentiable $\not\Longrightarrow f$ integrable. Necessity of compactness:

$$f(x) = e^x \in C^\infty(\mathbb{R}) \text{ but } \int_{\mathbb{R}} e^x dx \to \infty.$$

- f integrable $\not\Longrightarrow f$ differentiable
  - An integrable function that is not differentiable: $f(x) = |x|$ on $\mathbb{R}$
- f differentiable $\implies f$ continuous a.e.

7.8 Integrability

- Sufficient criteria for Riemann integrability:
  - f continuous
  - f bounded and continuous almost everywhere, or
  - f uniformly continuous
- f integrable $\iff$ bounded and continuous a.e.

Theorem 7.8.1 ***(FTC for the Riemann Integral).***
If F is a differentiable function on the interval $[a, b]$, and F' is bounded and continuous a.e., then $F' \in L_R([a, b])$ and

$$\forall x \in [a, b] : \int_a^x F'(t)\, dt = F(x) - F(a)$$

Suppose f bounded and continuous a.e. on $[a, b]$, and define

$$F(x) := \int_a^x f(t)\, dt$$

Then F is absolutely continuous on $[a, b]$, and for $p \in [a, b]$,

$$f \in C^0(p) \implies F \text{ differentiable at } p,\ F'(p) = f(p),\ \text{and } F' \overset{\text{a.e}}{=} f.$$

Proposition 7.8.1.
The Dirichlet function is Lebesgue integrable but not Riemann integrable:

$$f(x) = \mathbb{1}\left[x \in \mathbb{Q}\right]$$

Proof (?).

todo

∎

7.9 List of Free Conclusions:

- f integrable on $U \implies$:
  - f is bounded
  - f is continuous a.e. (finitely many discontinuities)
  - $\displaystyle\int f$ is continuous
  - $\displaystyle\int f$ is differentiable
- f continuous on U:
  - f is integrable on compact subsets of U
  - f is bounded
  - f is integrable
- f differentiable at a point p:
  - f is continuous

- f is differentiable in U
  - f is continuous a.e.
- Defining the Riemann integral: #todo

7.10 Convergence

7.10.1 Sequences and Series of Functions

Definition 7.10.1 (Convergence of an infinite series)**.**
Define

$$s_n(x) := \sum_{k=1}^{n} f_k(x)$$

and

$$\sum_{k=1}^{\infty} f_k(x) := \lim_{n\to\infty} s_n(x),$$

which can converge pointwise, absolutely, uniformly, or not all.

Proposition 7.10.1 *(?)***.**
If $\limsup_{k\in\mathbb{N}} |f_k(x)| \neq 0$ then f_k is not convergent.

Proposition 7.10.2 *(?)***.**
If f is injective, then f' is nonzero in some neighborhood of ???

7.10.2 Pointwise convergence

$$f_n \to f = \lim_{n\to\infty} f_n.$$

Summary:

$$\lim_{f_n\to f} \lim_{x_i\to x} f_n(x_i) \neq \lim_{x_i\to x} \lim_{f_n\to f} f_n(x_i).$$

$$\lim_{f_n\to f} \int_I f_n \neq \int_I \lim_{f_n\to f} f_n.$$

Proposition 7.10.3 *(?)***.**
Pointwise convergence is strictly weaker than uniform convergence.

Proof (?).
$f_n(x) = x^n$ on $[0,1]$ converges pointwise but not uniformly.

- Towards a contradiction let $\varepsilon = \frac{1}{2}$.
- Let $n = N\left(\frac{1}{2}\right)$ and $x = \left(\frac{3}{4}\right)^{\frac{1}{n}}$.
- Then $f(x) = 0$ but

$$|f_n(x) - f(x)| = x^n = \frac{3}{4} > \frac{1}{2}$$

■

Proposition 7.10.4*(A pointwise limit of continuous functions is not necessarily continuous.).*

$$f_n \text{ continuous } \not\Longrightarrow f := \lim_n f_n \text{ is continuous.}$$

Proof (?).
Take

$$f_n(x) = x^n, \quad f_n(x) \to \mathbb{1}\,[[]\, x = 1].$$

■

Proposition 7.10.5*(The limit of derivatives need not equal the derivative of the limit).*

$$\begin{aligned} f_n \text{ differentiable } &\not\Longrightarrow f_n' \text{ converges} \\ f_n' \text{ converges } &\not\Longrightarrow \lim f_n' = f'. \end{aligned}$$

Proof (?).
Take

$$f_n(x) = \frac{1}{n}\sin(n^2 x) \to 0, \qquad \text{but } f_n' = n\cos(n^2 x) \text{ does not converge.}$$

■

Proposition 7.10.6*(?).*

$$f_n \in \mathcal{R}_I \not\Longrightarrow \lim_{f_n \to f} \int_I f_n \neq \int_I \lim_{f_n \to f} f_n.$$

Proof (?).

May fail to converge to same value, take

$$f_n(x) = \frac{2n^2x}{(1+n^2x^2)^2} \to 0 \qquad \text{but } \int_0^1 f_n = 1 - \frac{1}{n^2+1} \to 1 \neq 0.$$

■

7.10.3 Uniform Convergence

Notation:

$$f_n \rightrightarrows f = \lim_{n\to\infty} f_n \text{ and } \sum_{n=1}^{\infty} f_n \rightrightarrows S.$$

Summary:

$$\lim_{x_i\to x}\lim_{f_n\to f} f_n(x_i) = \lim_{f_n\to f}\lim_{x_i\to x} f_n(x_i) = \lim_{f_n\to f} f_n(\lim_{x_i\to x} x_i).$$

$$\lim_{f_n\to f}\int_I f_n = \int_I \lim_{f_n\to f} f_n.$$

$$\sum_{n=1}^{\infty}\int_I f_n = \int_I \sum_{n=1}^{\infty} f_n.$$

"The uniform limit of a(n) x function is x", for $x \in \{\text{continuous, bounded}\}$

- Equivalent to convergence in the uniform metric on the metric space of bounded functions on X:

$$f_n \rightrightarrows f \iff \sup_{x\in X} |f_n(x) - f(x)| \to 0.$$

  - $(B(X,Y), \|\|_\infty)$ is a metric space and $f_n \rightrightarrows f \iff \|f_n - f\|_\infty \to 0$ (where $B(X,Y)$ are bounded functions from X to Y and $\|f\|_\infty = \sup_{x\in I}\{f(x)\}$
- $f_n \rightrightarrows f \implies f_n \to f$ pointwise
- f_n continuous $\implies f$ continuous
  - i.e. "the uniform limit of continuous functions is continuous"
- $f_n \in C^1$, $\exists x_0 : f_n(x_0) \to f(x_0)$, and $f_n' \rightrightarrows g \implies f$ differentiable and $f' = g$ (i.e. $f_n' \to f'$)
  - Necessity of C^1 – look at failures of f_n' to be continuous:
    - ◊ Take $f_n(x) = \sqrt{\frac{1}{n^2} + x^2} \rightrightarrows |x|$, not differentiable
    - ◊ Take $f_n(x) = n^{-\frac{1}{2}}\sin(nx) \rightrightarrows 0$ but $f_n' \not\to f' = 0$ and $f' \neq g$

- f_n integrable $\implies f$ integrable and $\int f_n \to \int f$
- f_n bounded $\implies f$ bounded
- $f_n \rightrightarrows f_n \not\Rightarrow f_n'$ converges
  - Says nothing about it general
- $f_n' \rightrightarrows f' \not\Rightarrow f_n \rightrightarrows f$
  - Unless f converges at one or more points.

7.11 Sequences and Metric Spaces

Theorem 7.11.1 ***(Bolzano-Weierstrass).***
Every bounded sequence has a convergent subsequence.

Theorem 7.11.2 ***(Heine-Borel).***
In $\mathbb{R}^n, X$ is compact $\iff X$ is closed and bounded.

Remark 7.11.1.
Necessity of $\mathbb{R}^n$: $X = (\mathbb{Z}, d(x, y) = 1)$ is closed, complete, bounded, but not compact since $\{1, 2, \cdots\}$ has no convergent subsequence

Proposition 7.11.1 ***(Converse of Heine-Borel).***
Converse holds iff bounded is replaced with totally bounded

Definition 7.11.1 (Sequential Compactness)**.**
A topological space X is **sequentially compact** iff every sequence $\{x_n\}$ has a subsequence converging to a point in X.

Proposition 7.11.2 ***(Compactness and sequential compactness).***
If X is a metric space, X is compact iff X is sequentially compact.

Remark 7.11.2.
Note that in general, neither form of compactness implies the other.

Proposition 7.11.3 ***(All subsequences of a convergent sequence share a limit).***
$\{x_i\} \to p \implies$ every subsequence also converges to p.

Definition 7.11.2 (Cachy Sequence)**.**

todo

Proposition 7.11.4*(?)*.
Every convergent sequence in X is a Cauchy sequence.

Remark 7.11.3.
The converse need not hold in general, but if X is complete, every Cauchy sequence converges. An example of a Cauchy sequence that doesn't converge: take $X = \mathbb{Q}$ and set $x_i = \pi$ truncated to i decimal places.

Remark 7.11.4.
If any subsequence of a Cauchy sequence converges, the entire sequence converges.

Definition 7.11.3 (Metric)**.**

$$
\begin{aligned}
& d(x,y) \geq 0 && \text{Positive} \\
& d(x,y) = 0 \iff x = y && \text{Nondegenerate} \\
& d(x,y) = d(y,x) && \text{Symmetric} \\
& d(x,y) \leq d(x,p) + d(p,y) \quad \forall p && \text{Triangle Inequality.}
\end{aligned}
$$

Definition 7.11.4 (Complete)**.**
?
todo

Definition 7.11.5 (Bounded)**.**
?
todo

7.12 Topology

Definition 7.12.1 (Axioms for a Topology)**.**
Open Set Characterization: Arbitrary unions and finite intersections of open sets are open.
Closed Set Characterization: Arbitrary intersections and finite unions of closed sets are closed.

Remark 7.12.1.
The best source of examples and counterexamples is the open/closed unit interval in $\mathbb{R}$. Always test against these first!

Remark 7.12.2.
If f is a continuous function. the preimage of every open set is open and the preimage of every closed set is closed.

Proposition 7.12.1 *(?).*
In $\mathbb{R}$, singleton sets and finite discrete sets are closed.

Proof (?).
A singleton set can be written

$$\{p_0\} = (-\infty, p) \cup (p, \infty).$$

A finite discrete set $\{p_0\}$, which wlog (by relabeling) can be assumed to satisfy $p_0 < p_1 < \cdots$, can be written

$$\{p_0, p_1, \cdots, p_n\} = (-\infty, p_0) \cup (p_0, p_1) \cup \cdots \cup (p_n, \infty).$$

■

Proposition 7.12.2 *(?).*
This yields a good way to produce counterexamples to continuity.

In $\mathbb{R}$, singletons are closed. This means any finite subset is closed, as a finite union of singleton sets!

Proposition 7.12.3 *(?).*
If X is a compact metric space, then X is complete and bounded.

Proposition 7.12.4 *(?).*
If X complete and $X \subset Y$, then X closed in Y.

Remark 7.12.3.
The converse generally does not hold, and completeness is a necessary condition. Counterexample: $\mathbb{Q} \subset \mathbb{Q}$ is closed but $\mathbb{Q} \subset \mathbb{R}$ is not.

Proposition 7.12.5 *(?).*
If X is compact, then $Y \subset X \implies Y$ is compact $\iff Y$ closed.

7.13 Limits

7.14 Continuity

7.14.1 Lipschitz Continuity

7.15 Integrability

8 Topology

8.1 Definitions

Bring in Rudin's list

- Epsilon-neighborhood
  - $N_r(p) = \left\{q \;\middle|\; d_X(p,q) < r\right\}$
- Limit Point
  - p is a limit point of E iff $\forall N_r(p),\ \exists q \neq p \;\middle|\; q \in N_r(p)$
  - Equivalently, $\forall N_r(p),\ N_r(p) \cap E \neq \emptyset$
  - Let $L(E)$ be the set of limit points of E.
  - Example: $E = (0,1) \implies 0 \in L(E)$
- Isolated Point
  - p is an isolated point of E iff p is not a limit point of E
  - Equivalently, $\exists N_r(p) \;\middle|\; N_r(p) \cap E = \emptyset$
  - Equivalently, $E - L(E)$
- Perfect
  - E is perfect iff E is closed and $E \subseteq L(E)$
  - Equivalently, $L(E) = E$
- Interior
  - p is an interior point of E iff $\exists N_r(p) \;\middle|\; N_r(p) \subsetneq E$
  - Denote the interior of E by E°
- Exterior
- Closed sets
  - E is closed iff p a limit point of $E \implies p \in E$
  - Equivalently if $L(E) \subseteq E$
  - Closed under finite unions, arbitrary intersections
- Open sets
  - E is open iff $p \in E \implies p \in E^\circ$
  - Equivalently, if $E \subseteq E^\circ$
  - Closed under arbitrary unions, finite intersections
- Boundary
- Closure
- Dense
  - E is dense in X iff $X \subseteq E \cup L(E)$
- Connected
  - Space of connected sets closed under union, product, closures
  - Convex $\implies$ connected
- Disconnected
- Path Connected

  - $\forall x, y \in X \exists f : I \to X \bigm| f(0) = x, f(1) = y$
  - Path connected $\implies$ connected

- Simply Connected
- Totally Disconnected
- Hausdorff
- Compact
  - Every covering has a finite subcovering.
  - X compact and $U \subset X$: (U closed $\implies$ U compact)
    $\diamondsuit$ U compact $\implies$ U closed iff X is Hausdorff
  - Closed under products

Example 8.1.1 (?).
The space $\left\{\frac{1}{n}\right\}_{n\in\mathbb{N}}$.

List of properties preserved by continuous maps:

- Connectedness
- Compactness

Checking if a map is homeomorphism:

- f continuous, X compact and Hausdorff $\implies$ f is a homeomorphism.

8.2 Definitions

$$
\begin{aligned}
L^2(X) &= \left\{f : X \to \mathbb{R} : \int_{\mathbb{R}} f(x)\ dx < \infty\right\} && \text{square integrable functions} \\
\langle g,\ f\rangle_2 &= \int_{\mathbb{R}} g(x)f(x)\ dx && \text{the } L^2 \text{ inner product} \\
\|f\|_2^2 = \langle f,\ f\rangle &= \int_{\mathbb{R}} f(x)^2\ dx && \text{norm} \\
E[\,\cdot\,] &= \langle \cdot\,,\ f\rangle && \text{expectation} \\
(\tau_p f)(x) &= f(p - x) && \text{translation} \\
(f * g)(p) &= \int_{\mathbb{R}} f(t)g(p-t)\ dt = \int_{\mathbb{R}} f(t)(T_p g)(t)\ dt = \langle T_p g,\ f\rangle && \text{convolution}
\end{aligned}
$$

Definition 8.2.1 (Random Variable).
For (Σ, E, μ) a probability space with sample space Σ and probability measure μ, a random variable is a function $X : \Sigma \to \mathbb{R}$

Definition 8.2.2 (Probability Density Function (PDF)).

For any $U \subset \mathbb{R}$, given by the relation

$$P(X \in U) = \int_U f(x)\ dx$$
$$\implies P(a \leq X \leq b) = \int_a^b f(x)\ dx$$

Definition 8.2.3 (Cumulative Distribution Function (CDF))**.**
The antiderivative of the PDF

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\ dx.$$

Yields $\dfrac{\partial F}{\partial x} = f(x)$

Definition 8.2.4 (Mean/Expected Value)**.**

$$E[X] := \langle \mathrm{id},\ f \rangle = \int_{\mathbb{R}} x f(x)\ dx.$$

Also denoted μ_X.

Proposition 8.2.1 ***(Linearity of Expectation).***

$$E\left[\sum_{i \in \mathbb{N}} a_i X_i\right] = \sum_{i \in \mathbb{N}} a_i E[X_i].$$

Does not matter whether or not the X_i are independent.

Definition 8.2.5 (Variance)**.**

$$\begin{aligned}\mathrm{Var}(X) &= E[(X - E[X])^2] \\ &= \int (x - E[X])^2 f(x)\ dx \\ &= E[X^2] - E[X]^2 \\ &:= \sigma^2(X)\end{aligned}$$

where σ is the standard deviation. Can also defined as $\left\langle (\mathrm{id} - \langle \mathrm{id},\ f \rangle)^2,\ f \right\rangle$ Take the portion of the id function in the orthogonal complement of f, squared, and project it back onto f?

???

Proposition 8.2.2 ***(Properties of Variance).***

$$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$$
$$\mathrm{Var}\left(\sum_{\mathbb{N}} X_i\right) = \sum_i \mathrm{Var}(X_i) + 2\sum_{i<j}\mathrm{Cov}(X_i, X_j).$$

Definition 8.2.6 (Covariance)**.**

$$\begin{aligned}\mathrm{Cov}(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

Proposition 8.2.3 ***(Properties of Covariance).***

$$\begin{aligned}\mathrm{Cov}(X,X) &= \mathrm{Var}(X) \\ \mathrm{Cov}(aX,Y) &= a\mathrm{Cov}(X,Y) \\ \mathrm{Cov}(\sum_{\mathbb{N}} X_i, \sum_{\mathbb{N}} Y_j) &= \sum_i \sum_j \mathrm{Cov}(X_i, Y_j)\end{aligned}$$

.

Proposition 8.2.4 ***(Stirling's Approximation).***

$$k! \sim k^{\frac{k+1}{2}} e^{-k}\sqrt{2\pi}.$$

Proposition 8.2.5 ***(Markov Inequality).***

$$P(X \geq a) \leq \frac{1}{a}E[X]$$

One-sided Markov:

$$P(X \in N_\varepsilon(\mu)) = 2\frac{\sigma^2}{\sigma^2 + a^2}.$$

Proposition 8.2.6 ***(Chebyshev's Inequality).***

$$P(|X-\mu| \geq a) \leq \left(\frac{\sigma}{k}\right)^2$$

Proof (?).
Apply Markov to the variable $(X-\mu)^2$ and $a = k^2$

■

Theorem 8.2.1 ***(Central Limit Theorem).***
For X_i i.i.d.,

$$\lim_n \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \sim N(0,1).$$

Theorem 8.2.2 ***(Strong Large of Large Numbers).***

$$P(\frac{1}{n}\sum X_i \to \mu) = 1.$$

Proposition 8.2.7 ***(Chernoff Bounds).***
For all $t > 0$,

$$P(X \in N_\varepsilon(a)^c) \leq 2e^{-at}M_X(t)$$

.

Proposition 8.2.8 ***(Jensen's Inequality).***

$$E[f(X)] \geq f(E[X])$$

Definition 8.2.7 (Entropy).

$$H(X) = -\sum p_i \ln p_i$$

8.3 Theory and Background

Definition 8.3.1 (Axioms for a Probability Space).
Given a sample space Σ with events S, 1. $\mu(\Sigma) = 1$ 1. Yields $S \in \Sigma \implies 0 \leq P(S) \leq 1$ 2. For mutually exclusive events, $P(\cup_{\mathbb{N}} S_i) = \sum_{\mathbb{N}} P(S_i)$ 1. Yields $P(\emptyset) = 0$

Proposition 8.3.1 ***(Properties that follow from axioms).***
- $P(S^c) = 1 - P(S)$
- $E \subseteq F \implies P(E) \leq P(F)$
- Proof: $E \subseteq F \implies F = E \cup (E^c \cap F)$, which are disjoint, so $P(E) \leq P(E) + P(E^c \cap F) = P(F)$.
- $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

Definition 8.3.2 (Conditional Probability).

$$P(F)P(E \mid F) = P(E \cap F) = P(E)P(F \mid E)$$

Generalization:

$$P(\cap_{\mathbb{N}} E_i) = P(E_1)P(E_2 \mid E_1)P(E_3 \mid E_1 \cap E_2)\cdots$$

Theorem 8.3.1 ***(Bayes' Rule).***

$$P(E) = P(F)P(E \mid F) + P(F^c)P(E \mid F^c)$$
$$P(E) = \sum_i P(A_i)P(E \mid A_i)$$

Generalization: for $\coprod_{i=1}^{n} A_i = \Sigma$ and $A = A_i$ for some i,

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{\sum_{j=1}^{n} P(B \mid A_j)}.$$

The LHS: the posterior probability, while $P(A_i)$ are the priors.

Definition 8.3.3 (Odds)**.**

$$P(A)/P(A^c)$$

Conditional odds:

$$\frac{P(A \mid E)}{P(A^c \mid E)} = \frac{P(A)}{P(A^c)} \frac{P(E \mid A)}{P(E \mid A^c)}$$

Definition 8.3.4 (Independence)**.**

$$P(A \cap B) = P(A)P(B)$$

Proposition 8.3.2 ***(Change of Variables for PDFs).***
If g is differentiable and monotonic and $Y = g(X)$, then

$$f_Y(y) = \begin{cases} (f_X \circ g^{-1})(y) \left| \frac{\partial}{\partial y} g^{-1}(y) \right| & y \in \operatorname{im}(g) \\ 0 & \text{else} \end{cases}$$

Proposition 8.3.3 ***(PDF for a sum of independent random variables).***

$$f_{X+Y} = (F_X * f_y)$$

8.4 Distributions

Let X be a random variable, and f be its probability density function satisfying $f(k) = P(X = k)$

8.4.1 Uniform

- Consider an event with n mutually exclusive outcomes of equal probability, and let $X \in \{1, 2, \ldots, n\}$ denote which outcome occurs. Then,

$$\begin{aligned} f(k) = & \quad \frac{1}{n} \\ \mu = & \frac{n}{2} \\ \sigma^2 = & a \end{aligned}$$

- Examples:
  - Dice rolls where $n = 6$.
  - Fair coin toss where $n = 2$.
- Continuous: $\mu = (1/2)(b + a), \sigma^2 = (1/12)(b - a)^2$

8.4.2 Bernoulli

- Consider a trial with either a positive or negative outcome, and let $X \in \{0, 1\}$ where 1 denotes a success with probability p. Then,

$$\begin{aligned} f(k) = & \begin{cases} 1 - p, & k = 0 \\ p, & k = 1 \end{cases} \\ \mu = & \quad p \\ \sigma^2 = & \quad p(1 - p) \end{aligned}$$

- Examples: - A weighted coin with $P(\text{Heads}) = p$

8.4.3 Binomial

- Consider a sequence of independent Bernoulli trials, let $X \in \{1, \ldots, n\}$ denote the number of successes occurring in n trials. Then,

$$\begin{aligned} f(k) = & \quad \binom{n}{k} p^k (1 - p)^{n-k} \\ \mu = & \quad np \\ \sigma^2 = & \quad np(1 - p) \end{aligned}$$

- Examples:
  - A sequence of coin flips and the numbers of total heads occurring.

8.4.4 Poisson

- Given a parameter $\lambda > 0$ that denotes the rate per unit time of an event occurring and X the number of times the event occurs in one unit of time,

$$\begin{aligned} f(k) &= \frac{\lambda^k}{k!} e^{-\lambda} \\ \mu &= \lambda \\ \sigma^2 &= \lambda^2 \end{aligned}$$

- Approximates binomial when $n >> 1$ and $p << 1$ by using $\lambda = np$

8.4.5 Negative Binomial

- $B^-(r, p)$: similar to binomial, where X is the number of trials it takes to accumulate r successes

$$\begin{aligned} f(k) &= \binom{k-1}{r-1} p^r (1-p)^{k-r} \\ \mu &= \frac{r}{p} \\ \sigma^2 &= \frac{r(1-p)}{p^2} \end{aligned}$$

8.4.6 Geometric

- Consider a sequence of independent Bernoulli trials, let $X \in \{1, \dots, n\}$ where $X = k$ denotes the first success happening on the k-th trial. Then,

$$\begin{aligned} f(k) &= (1-p)^{k-1} p \\ \mu &= \frac{1}{p} \\ \sigma^2 &= \frac{1-p}{p^2} \end{aligned}$$

- Note this is equal to $B^-(1, p)$
- Examples:
  - A sequence of coin flips and the number of flips before the first heads appears.

8.4.7 Hypergeometric

- $H(n, m, s)$ An urn filled with n balls, where m are white and $n - m$ are black; pick a sample of size s and let X denote the number of white balls:

$$
\begin{aligned}
f(k) = &\quad \binom{m}{k}\binom{n-m}{s-k}\binom{n}{s}^{-1} \\
\mu = &\quad \frac{ms}{n} \\
\sigma^2 = &\quad \frac{ms}{n}(1 - \frac{m}{n})\left(1 - \frac{s-1}{n-1}\right)
\end{aligned}
$$

8.4.8 Normal

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

z	$\Phi(z)$
0	0.5
1	0.69
1.5	0.84
2	0.93
2.5	0.97
> 3	0.99

8.5 Common Problems

- Birthday Paradox
- Coupon Collectors
  - Given $X = \{1, \cdots n\}$, what is the expected number of draws until all n outcomes are seen?

8.6 Notes, Shortcuts, Misc

- When computing expected values, variation, etc, just insert a parameter k and compute the moments $E[X^k]$. Then with a solution in terms of k, let $k = 1, 2$ etc.
- Neat property of pdfs: $P(X \in N_\varepsilon(a)) \approx \varepsilon f(a)$

Definition 8.6.1 (The Gamma Function)**.**

$$\Gamma(x + 1) = \int_{\mathbb{R}>0} e^{-t} t^x \ dt.$$

Integrate by parts to obtain functional relation $\Gamma(x + 1) = x\Gamma(x)$

Proposition 8.6.1 ***(Boole's Inequality).***

$$P(\cup_{\mathbb{N}} A_i) \leq \sum_{\mathbb{N}} P(A_i)$$

- For any function $f : X \to \mathbb{R}$, there is a lower bound: $\max_{x \in X} f(x) \geq E[f(x)]$

Definition 8.6.2 (Moment Generating Functions)**.**
Define

$$M(t) = E[e^{Xt}]$$

- Then $M^{(n)}(0)$ is the n-th moment, i.e. $M'(0) = E[X], M''(0) = E[X^2]$, etc.
- $M_{X+Y}(t) = M_X(t) M_Y(t)$ (if independent)
- $M_{aX+b}(t) = e^{bt} M_X(at)$
- $f_X = \mathcal{F}^{-1}(M_X(it))$, denoting the inverse Fourier transform,

8.7 Table of Distribution Info

Table: let $q = 1 - p$.

Distribution	$f(x)$	μ	σ^2	$M(t)$
$B(n,p)$	$\binom{n}{x} p^x q^{n-x}$	np	npq	$(pe^t + q)^n$
$P(\lambda)$	$\frac{\lambda^x}{x!} e^{-\lambda}$	λ	λ	$e^{\lambda(e^t - 1)}$
$G(p)$	$q^{x-1} p$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^t}{1 - qe^t}$
$B^-(r,p)$	$\binom{n-1}{r-1} p^r q^{n-r}$	$\frac{r}{p}$	$\frac{rq}{p^2}$	$\left(\frac{pe^t}{1 - qe^t}\right)^r$
$U(a,b)$	$\mathbb{1}\left[a \leq x \leq b\right] \frac{1}{b-a}$	$\frac{1}{2}(a+b)$	$\frac{1}{12}(b-a)^2$	$\frac{e^{tb} - e^{ta}}{t(b-a)}$
$Exp(\lambda)$	$\mathbb{1}\left[0 \leq x\right] \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda - t}$
$\Gamma(s,\lambda)$	$\mathbb{1}\left[0 \leq x\right] \frac{\lambda e^{-\lambda x} (\lambda x)^{s-1}}{\Gamma(s)}$	$\frac{s}{\lambda}$	$\frac{s}{\lambda^2}$	$\left(\frac{\lambda}{\lambda - t}\right)^s$
$N(\mu,\sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{\mu t + \frac{1}{2}\sigma^2 t^2}$

- Why you need the Stieltjes Integral: let $X \sim B(n, \frac{1}{2}), Y \sim U(0,1)$, and

$$Z = \begin{cases} X, & \text{X} = 1 \\ Y, & else \end{cases}$$

then $|Z| = |\mathbb{R}|$ so Z is not discrete, but $P(X = 1) = \frac{1}{2} \neq 0$ so Z is not continuous. Definition:

$$\int_a^b g(x)\ dF(x) = \lim \sum_{i=1}^{n} g\left(x_i\right)\left(F\left(x_i\right) - F\left(x_{i-1}\right)\right).$$

8.8 Notation

$S_n = \{1, 2, \dots n\}$	the symmetric group
$\binom{n}{k} = \frac{n!}{k!(n-k)!}$	binomial coefficient
$n^{\underline{k}} = n(n-1)\cdots(n-k+1) = k!\binom{n}{k}$	falling factorial
$n^{\overline{k}} = n(n+1)\cdots(n+k-1) = k!\binom{n+n-1}{n}$	rising factorial
$\binom{n}{m_1, m_2, \cdots m_k} = \frac{n!}{\prod_{i=1}^{k} m_i!}$	multinomial coefficient

Note that the rising and falling factorials always have exactly k terms.

Multinomial: A set of n items divided into k distinct, disjoint subsets of sizes $m_1 \cdots m_k$. Multinomial theorem:

$$(x_1 + x_2 + \cdots x_k)^n = \sum_{\substack{m_1, m_2, \cdots, m_k \\ \sum m_i = n}} \binom{n}{m_1, m_2, \cdots, m_k} x_1^{m_1} x_2^{m_2} \cdots x_k^{m_k}$$

which contains $\binom{n+r-1}{r-1}$ terms.

Inclusion-Exclusion:

$$\begin{aligned} |\cup_{i=1}^{n} A_i| &= \sum_i |A|_i - \sum_{i_1 < i_2}^{\binom{n}{2^2}} |A_{i_1} \cap A_{i_2}| + \sum_{i_1 < i_2 < i_3}^{\binom{n}{2^3}} |A_{i_1} \cap A_{i_2} \cap A_{i_3}| + \cdots + (-1)^{n+1} |\cap_{i=1}^{n} A_i| \\ &= \sum_{k=1}^{n} \sum_{i_1 < \cdots < i_k} (-1)^{k+1} \left| \cap_{j=1}^{k} A_{i_j} \right| \end{aligned}$$

9 | The Important Numbers

- Binomial Coefficients

- The Binomial Theorem:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

- Choosing: $\binom{n}{k}$
- Choosing with repetition allowed: $\binom{n+k-1}{k}$

- Signed Stirling Numbers of the First Kind: $s(n,k)$
  - Count the number of permutations of n elements with k disjoint cycles, i.e. the number of elements elements in S_n that are the product of k disjoint cycles (including trivial cycles that fix a point).
  - Recurrence relation:

$$s(n,k) = s(n-1,k-1) + ks(n-1,k).$$

  - Relation to falling factorial: $x^{\underline{n}} = \sum_{k=1}^{n} s(n,k)x^k$

- Stirling Numbers of the Second Kind: $\left\{ {n \atop k} \right\}$
  - Counts the number of ways to partition a set N into k non-empty subsets S_i (i.e. such that $S_i \cap S_j = \emptyset$, $\coprod_{i=1}^{k} S_i = N$)
  - Recurrence relation:

$$\left\{ {n+1 \atop k} \right\} = k \left\{ {n \atop k} \right\} + \left\{ {n \atop k-1} \right\}$$

$$\left\{ {0 \atop 0} \right\} = 1, \quad \left\{ {n \atop 0} \right\} = \left\{ {0 \atop n} \right\} = 0$$

  - Explicit formula: $\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{i=0}^{k} (-1)^{k-i} \binom{k}{i} i^n$
  - $B_n = \sum_{i=0}^{n} \left\{ {n \atop i} \right\}$

- Betti Numbers
- Bell Numbers
- Compositions
  - A composition of n is a way of writing n as a sum of strictly positive integers, ie. $k_1 + k_2 + \cdots k_i = n$ where each $0 < k_i \leq n$, where order matters (and distinct orders count as distinct compositions).
  - Weak compositions: identical, but some terms are allowed to be zero.
  - Number of compositions of n into k parts: $\binom{n-1}{k-1}$
  - Number of *weak* compositions of n into k parts: $\binom{n+k-1}{n}$
  - Total number of compositions of n (into any number of parts): 2^{n-1}
- Partitions
  - A partition of n is a composition of n quotiented by permutations of the ordering of terms.

- ◇ Example: 2 compositions of 5 involving 1 and 4, given by $4+1$ and $1+4$, whereas there is only one such partition of 5 given by $4+1$.
  - Visualize with Young diagrams

9.1 Common Problems

- Stars and Bars
  - No two bars adjacent: $\binom{n-1}{k-1}$
  - Allowing adjacent bars: $\binom{n+k-1}{k-1}$

Coupon Collectors Problem

9.2 The Twelvefold Way

Consider a function $f : N \to K$ where $|N| = n, |K| = k$.

A number of valid interpretations: - f labels elements of N by elements of K - For each element of N, f chooses an element of K - f partitions N into classes that are mapped to the same element of K - Throw each of N balls into some of K boxes

Dictionary: - No restrictions: - Assign n labels, repetition allowed - Form a multiset of K of size n - Injectivity - Assign n labels without repetition - Select n distinct elements from K - Number of n-combinations of k elements - No more than one ball per box - Surjectivity: - Use every label at least once - Every element of K is selected at least once - "Indistinguishable" - Quotient by the action of S_n or S_k - n-permutations = injective functions - n-combinations = injective functions / S_n - n-multisets = all functions / S^n - Partitions of N into k subsets = surjective functions / S_k - Compositions of n into k parts = surjective functions / S_n

Permutations Restrictions	$N \xrightarrow{f} K$	$N \hookrightarrow K$	$N \twoheadrightarrow K$
f	k^n	$k^{\underline{n}}$	$k!\left\{ {n \atop k} \right\}$
$f \circ \sigma_N$	$\binom{n+k-1}{n}$	$\binom{k}{n}$	$\binom{n-1}{n-k}$
$\sigma_X \circ f$	$\sum_{i=0}^{k} \left\{ {n \atop i} \right\}$	$\mathbb{1}[n \leq k]$	$\left\{ {n \atop k} \right\}$
$\sigma_X \circ f \circ \sigma_N$	$p_k(n+k)$	$\mathbb{1}[n \leq k]$	$p_k(n)$

In words (todo: explain)

Perm. / Rest.	—	Injective	Surjective
—	A sequence of any n elements of K	Sequences of n *distinct* elements of K	Compositions of N with exactly k subsets
Permutations of N	Multisets of K with n elements	An n-element subset of K	Compositions of n with k terms

Perm. / Rest.	—	Injective	Surjective
Permutations of X	Partitions of N into $\leq k$ subsets	?	Partitions of N into exactly k nonempty subsets
Both	Partitions of n into $\leq k$ parts	?	Partitions of n into exactly k parts

Proofs/Explanations: todo

- Counting non-isomorphic things: Pick a systematic way. Can descend my maximum vertex degree, or ascend by adding nodes/leaves.

10 | Complex Analysis

- $\lim_{z \to z_0} f(z) = x_0 + iy_0$ iff the component functions limit to x_0 and y_0 respectively. Moreover, both ways are equal!

Notation: $z = a + ib, f(z) = u(x, y) + iv(x, y)$

10.1 Useful Equations and Definitions

$$\begin{aligned}
|z| &= \sqrt{a^2 + b^2} \\
|z|^2 = z\overline{z} &= a^2 + b^2 \\
\frac{z\overline{z}}{|z|^2} &= \frac{(a + ib)(a - ib)}{a^2 + b^2} = 1 \\
\frac{1}{z} &= \frac{\overline{z}}{|z|^2} = \frac{a - ib}{a^2 + b^2} \\
e^{zx} = e^{(a+ib)x} &= e^{ax}(\cos(bx) + i\sin(bx)) \\
x^z &:= e^{z \ln x} \\
\mathrm{Log}(z) &= \ln|z| + i\ \mathrm{Arg}(z) \\
\cos z &= \frac{1}{2}(e^{iz} + e^{-iz}) \\
\sin z &= \frac{1}{2i}(e^{iz} - e^{-iz}) \\
(x - z)(x - \overline{z}) &= x^2 - 2\mathcal{R}\rceil(z)x + (a^2 + b^2) \\
\frac{\partial}{\partial z} &= \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) \\
\frac{\partial}{\partial \overline{z}} &= \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)
\end{aligned}$$

10.2 Complex Arithmetic and Calculus

- n-th roots:

$$e^{\frac{ki}{2\pi n}}, \qquad k = 1, 2, \cdots n-1$$

10.2.1 Complex Differentiability

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

- A complex function that is not differentiable at a point: $f(z) = z/\bar{z}$ at $z = 0$

- Cauchy-Riemann Equations

$$u_x = v_y \qquad\qquad u_y = -v_x$$

- Alternatively:
  - $\frac{\partial[}{\partial f}]\bar{z} = 0$
  - $\langle \nabla u,\ \nabla v \rangle = 0$
  - $\Delta u = \Delta v = 0$ (both components are harmonic)

10.3 Complex Integrals

The main theorem:

$$\oint_C f(z)\ dz = 2\pi i \sum_k \operatorname{Res}(f, z_k)$$

Computing residues:

$$\operatorname{Res}(f, c) = \frac{1}{(n-1)!} \lim_{z\to c} \frac{d^{n-1}}{dz^{n-1}} \left((z-c)^n f(z)\right)$$

$$f(z) = \frac{g(z)}{h(z)} \implies \operatorname{Res}(f, c) = \frac{g(c)}{h'(c)}$$

Definitions

- Analytic: differentiable everywhere
- Entire
- Holomorphic
- Meromorphic

Complex Analytic $\implies$ smooth and all derivatives are analytic

> Not true in real case, take the everywhere differentiable but not C^1 function
>
> $$f(x) = \begin{cases} -\frac{1}{2}x^2 & x < 0 \\ \frac{1}{2}x^2 & x \geq 0 \end{cases}$$

11 | My Common Mistakes

$$-x^{-2} \neq \int x^{-1} = \int \frac{1}{x} = \ln x$$
$$\frac{1}{x} \neq \int \ln x = x \ln x - x$$
$$\int x^{-k} = \frac{1}{-k+1} x^{-k+1} \neq \frac{1}{-(k+1)} x^{-(k+1)}$$
$$\text{e.g. } \int x^{-2} = -x^{-1} \neq -\frac{1}{3} x^{-3} \lim_{n\to\infty} \frac{n}{n+1} = 1 \neq 0$$
$$\frac{\partial}{\partial x} a^x = \frac{\partial}{\partial x} e^{x \ln a} = e^{x \ln a} \ln a = a^x \ln a.$$

Exponentials: when in doubt, write $a^b = e^{b \ln a}$

$$\frac{\partial}{\partial x} x^{f(x)} = ?$$

$$\sum x^k = \frac{1}{1-x} \neq \frac{1}{1+x} = \sum (-1)^k x^k$$

11.1 Definitions

$$e^x = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n = \lim_{n\to\infty} \left(\frac{n+1}{n}\right)^n$$

11.2 Neat Tricks

- Commuting differentials and integrals:

$$\frac{d}{dx} \int_{a(x)}^{b(x)} f(x,t) dt = f(x, b(x)) \frac{d}{dx} b(x) - f(x, a(x)) \frac{d}{dx} a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x,t) dt$$

- Need $f, df dx f, \dfrac{df}{dx}$ to be continuous in both variables. Also need $a(x), b(x) \in C1 a(x), b(x) \in C_1$.
- If a, b are constant, boundary terms vanish.
- Recover the fundamental theorem with $a(x) = a, b(x) = ba(x) = a, b(x) = b$, and $f(x,t) = f(t) f(x,t) = f(t)$.

11.3 Big Derivative / Integral Table

$\frac{\partial f}{\partial x} \Leftarrow$	f	$\Rightarrow \int f dx$
$\frac{1}{2\sqrt{x}}$	$\sqrt{x}$	$\frac{2}{3}x^{\frac{3}{2}}$
nx^{n-1}	$x^n, n \neq -1$	$\frac{1}{n+1}x^{n+1}$
$-nx^{-(n+1)}$	$\frac{1}{x^n}, n \neq 1$	$-\frac{1}{n-1}x^{-(n-1)}$
$\frac{1}{x}$	$\ln(x)$	$x\ln(x) - x$
$a^x \ln(a)$	a^x	$\frac{a^x}{\ln a}$
$\cos(x)$	$\sin(x)$	$-\cos(x)$
$-\csc(x)\cot(x)$	$\csc(x)$	$\ln\lvert\csc(x) - \cot(x)\rvert$
$-\sin(x)$	$\cos(x)$	$\sin(x)$
$\sec(x)\tan(x)$	$\sec(x)$	$\ln\lvert\sec(x) + \tan(x)\rvert$
$\sec^2(x)$	$\tan(x)$	$\ln\left\lvert\frac{1}{\cos x}\right\rvert$
$-\csc^2(x)$	$\cot(x)$	$\ln\lvert\sin x\rvert$
$\frac{1}{1+x^2}$	$\tan^{-1}(x)$	$x\tan^{-1}x - \frac{1}{2}\ln(1+x^2)$
$\frac{1}{\sqrt{1-x^2}}$	$\sin^{-1}(x)$	$x\sin^{-1}x + \sqrt{1-x^2}$
$-\frac{1}{\sqrt{1-x^2}}$	$\cos^{-1}(x)$	$x\cos^{-1}x - \sqrt{1-x^2}$
$\frac{1}{\sqrt{x^2+a}}$	$\ln\left\lvert x + \sqrt{x^2+a}\right\rvert$	.
$2\sin x\cos x$	$\sin^2(x)$	$\frac{1}{2}(x - \sin x\cos x)$
$-2\sin x\cos x$	$\cos^2(x)$	$\frac{1}{2}(x + \sin x\cos x)$
$2\csc^2(x)\cot(x)$	$\csc^2(x)$	$-\cot(x)$
$2\sec^2(x)\tan(x)$	$\sec^2(x)$	$\tan(x)$
?	?	?
?	?	?
?	?	?
?	?	?
?	?	?
?	?	?
?	?	?
$(ax+1)e^{ax}$	xe^{ax}	$\frac{1}{a^2}(ax-1)e^{ax}$
?	$e^{ax}\sin(bx)$	$\frac{1}{a^2+b^2}e^{ax}(a\sin bx - b\cos bx)$
?	$e^{ax}\cos(bx)$	$\frac{1}{a^2+b^2}e^{ax}(a\sin bx + b\cos bx)$
?	?	?

11.4 Useful Series and Sequences

Notation: $\uparrow, \downarrow$: monotonically converges from below/above.

- Taylor Series:

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!} (x - x_0)^n$$

- Cauchy Product:

$$\left(\sum_{k=0}^{\infty} a_k x^k \right) \left(\sum_{k=0}^{\infty} b_i x^n \right) = \sum_{k=0}^{\infty} \left(\sum_{i=0}^{k} a_n b_n \right) x^k$$

- Differentiation:

$$\frac{\partial}{\partial x} \sum_{k=i}^{\infty} a_k x^k = \sum_{k=i+1}^{\infty} k \, a_k x^{k-1}$$

- Common Series

$$\sum_{k=0}^{N} x^k = \frac{1-x^{N+1}}{1-x}$$

$$\sum_{k=1}^{\infty} x^k = \frac{1}{1-x} \quad \text{for } |x| < 1$$

$$\sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2} \quad \text{for } |x| < 1$$

$$\sum_{k=2}^{\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3} \quad \text{for } |x| < 1$$

$$\sum_{k=3}^{\infty} k(k-1)(k-2)x^{k-3} = \frac{6}{(1-x)^4} \quad \text{for } |x| < 1$$

$$\sum_{k=1}^{\infty} \binom{n}{k} x^k y^{n-k} = (x+y)^n$$

$$\sum_{k=1}^{\infty} \frac{x^k}{k} = -\log(1-x)$$

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x$$

$$\sum_{n=0}^{\infty} \frac{(-1)^k}{(2n+1)!} x^{2k+1} = x - \frac{x^3}{3!} + \frac{x^5}{5!} = \sin(x)$$

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{(2n)!} x^{2k} = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} = \cos(x)$$

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{2n+1} x^{2k+1} = x - \frac{x^3}{3} + \frac{x^5}{5} = \arctan(x)$$

$$\sum_{k=0}^{\infty} \frac{1}{(2k+1)!} x^{2n+1} = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots = \sinh(x)$$

$$\sum_{k=0}^{\infty} \frac{1}{(2k)!} x^{2k} = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots = \cosh(x)$$

$$\sum_{k=0}^{\infty} \frac{x^{2k+1}}{2k+1} = \operatorname{arctanh} x$$

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

$$\sum_{k=1}^{\infty} (-1)^k \frac{1}{k} = \ln(2)$$

$$\sum_{k=1}^{N} \frac{1}{k} =_{\approx} \ln(N) + \gamma + \frac{1}{2N}$$

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

11.5 Partial Fraction Decomposition

Given $R(x) = \dfrac{p(x)}{q(x)}$, factor $q(x)$ into $\prod q_i(x)$.

- Linear factors of the form $q_i(x) = (ax+b)^n$ contribute

$$r_i(x) = \sum_{k=1}^{n} \frac{A_k}{(ax+b)^k} = \frac{A_1}{ax+b} + \frac{A_2}{(ax+b)^2} + \cdots$$

- Irreducible quadratics of the form $q_i(x) = (ax^2+bx+c)^n$ contribute

$$r_i(x) = \sum_{k=1}^{n} \frac{A_k x + B_k}{(ax^2+bx+c)^k} = \frac{A_1 x + B_1}{ax^2+bx+c} + \frac{A_2 x + B_2}{(ax^2+bx+c)^2} + \cdots$$

  - Note: ax^2+bx+c is irreducible $\iff b^2 < 4ac$

- Write $R(x) = \dfrac{p(x)}{\prod q_i(x)} = \sum r_i(x)$, then solve for the unknown coefficients A_k, B_k.
  - IMPORTANT SHORTCUT: don't try to solve the resulting linear system: for each $q_i(x)$, multiply through by that factor and evaluate at its root to zero out many terms!
  - For linear terms $q_i(x) = (ax+b)^n$, define $P(x) = (ax+b)^n R(x)$; then

$$A_k = \frac{1}{(n-k)!} P^{(n-k)}(a), \quad k = 1, 2, \cdots n$$
$$\implies A_n = P(a),\ A_{n-1} = P'(a),\ \cdots,\ A_1 = \frac{1}{(n-1)!} P^{(n-1)}(A)$$

  - Note: #todo check, might need to evaluate at $-b/a$ instead, extend to quadratics.

11.6 Properties of Norms

$$\begin{aligned}
\|t\mathbf{x}\| &= |t|\|\mathbf{x}\| \\
|\langle \mathbf{x},\ \mathbf{y} \rangle| &\leq \|\mathbf{x}\|\|\mathbf{y}\| \\
\|\mathbf{x}+\mathbf{y}\| &\leq \|\mathbf{x}\| + \|\mathbf{y}\| \\
\|\mathbf{x}-\mathbf{z}\| &\leq \|\mathbf{x}-\mathbf{y}\| + \|\mathbf{y}-\mathbf{z}\|
\end{aligned}$$

11.7 Logic Identities

- $P \implies Q \iff Q$ or $\neg P$
- $P \implies Q \iff \neg Q \implies \neg P$
- P or (Q and S) $\iff$ (P or Q) and (P or S)
- P and (Q or S) $\iff$ (P and Q) or (P and S)
- $\neg$(P and Q) $\iff \neg P$ or $\neg Q$
- $\neg$(P or Q) $\iff \neg P$ and $\neg Q$

11.8 Set Identities

$$
\begin{array}{rcrcl}
A \cup B & = & A \cup (A^c \cap B) & & \\
A & = & (B \cap A) \cup (B^c \cap A) & & \\
(\cup_{\mathbb{N}} A_i)^c & = & \cap_{\mathbb{N}} A_i^c & & \\
(\cap_{\mathbb{N}} A_i)^c & = & \cup_{\mathbb{N}} A_i^c & & \\
A - B & = & A \cap B^c & & \\
(A - B)^c & = & A^c \cup B & & \\
(A \cup B) - C & = & (A - C) \cup (B - C) & & \\
(A \cap B) - C & = & (A - C) \cap (B - C) & & \\
A - (B \cup C) & = & (A - B) \cap (A - C) & & \\
A - (B \cap C) & = & (A - B) \cup (A - C) & & \\
A - (B - C) & = & (A - B) \cup (A \cap C) & & \\
(A - B) \cap C & = & (A \cap C) - B & = & A \cap (C - B) \\
(A - B) \cup C & = & (A \cup C) - (B - C) & & \\
A \cup (B \cap C) & = & (A \cup B) \cap (A \cup C) & & \\
A \cap (B \cup C) & = & (A \cap B) \cup (A \cap C) & & \\
A \subseteq C \text{ and } B \subseteq C & \implies & A \cup B \subseteq C & & \\
C \subseteq A \text{ and } C \subseteq B & \implies & C \subseteq A \cup B & & \\
A_k \text{ countable} & \implies & \prod_{k=1}^{n} A_k,\ \cup_{k=1}^{\infty} A_k \quad \text{countable} & &
\end{array}
$$

11.9 Preimage Identities

Summary

- Injectivity: left cancellation
- Surjectivity: right cancellation
- Everything commutes with unions
- Preimage commutes with everything
- Image generally only results in an inequality

Preimage Equations

- $A \subseteq B \implies f(A) \subseteq f(B)$ or $f^{-1}(A) \subseteq f^{-1}(B)$
- $f^{-1}(\cup_{i \in I} A_i) = \cup_{i \in I} f^{-1}(A_i)$
  - Also holds for $f(\cup_{i \in I} A_i) = \cup_{i \in I} f(A_i)$
- $f^{-1}(\cap_{i \in I} A_i) = \cap_{i \in I} f^{-1}(A_i)$
  - Also holds for $f(\cap_{i \in I} A_i) = \cap_{i \in I} f(A_i)$
- $f^{-1}(A) - f^{-1}(B) = f^{-1}(A - B)$
  - BUT $f(A) - f(B) \subseteq f(A - B)$
- For $X \subset A, Y \subset B$:
  - $(f|_X)^{-1} = X \cap f^{-1}(Y)$
  - $(f \circ f^{-1})(Y) = Y \cap f(A)$
- Summary: preimage commutes with:

  - Union
  - Intersection
  - Complements
  - Difference
  - Symmetric Difference

Image Equations

- $A \subset B \implies f(A) \subset f(B)$
- $f(\cup A_i) = \cup f(A_i)$
- $f(\cap A_i) \subset \cap f(A_i)$
- $f(A - B) \supset f(A) - f(B)$
- $f(A^c) = \text{im}(f) - f(A)$

Equations Involving Both

- $A \subseteq f^{-1}(f(A))$
  - Equal $\iff$ f is injective
- $f(f^{-1}(A)) \subseteq A$
  - Equal $\iff$ f is surjective

11.10 Pascal's Triangle:

n	Sequence
3	$1, 2, 1$
4	$1, 3, 3, 1$
5	$1, 4, 6, 4, 1$
6	$1, 5, 10, 10, 5, 1$
7	$1, 6, 15, 20, 15, 16, 1$
8	$1, 7, 21, 35, 35, 21, 7, 1$

Obtain new entries by adding in L pattern rotated by π (e.g. 7 = 1+6, 12 = 6 + 15, etc). Note that $\binom{n}{i}$ is given by the entry in the n-th row, i-th column.

11.11 Table of Small Factorials

n	$n!$
2	2
3	6
4	24
5	120
6	720
7	5040
8	40320
9	362880
10	3628800

$\pi \approx 3.1415926535$ $e \approx 2.71828$ $\sqrt{2} \approx 1.4142135$

11.12 Primes Under 100:

$$
\begin{aligned}
&2, 3, 5, 7,\\
&11, 13, 17, 19,\\
&23, 29,\\
&31, 37,\\
&41, 43, 47,\\
&53, 59,\\
&61, 67,\\
&71, 73, 79,\\
&83, 89,\\
&97,\\
&101
\end{aligned}
$$

11.13 Checking Divisibility by Small Numbers

Note that $n \mod 10^k$ yields the last k digits. Let d_i denote the i-th digit of n.

The recursive prime procedure (RPP): for each prime p, there exists a k such recursive application of this procedure to n yields the same remainder mod p as n itself.

- Write $n_0 = 10x + y$ where $y = 0 \ldots 9$
- Let $n_1 = x + ky$, repeat until $n_i < 10$.

p	$p \mid n \iff$	Mnemonic
2	$n \equiv 2, 4, 6, 8 \mod 10$	Last digit is even
3	$\sum d_i \equiv 0 \mod 3$	3 divides the sum of digits (apply recursively)
4	$n \equiv 4k \mod 10^2$	Last two digits are divisible by 4
5	$n \equiv 0, 5 \mod 10$	Last digit is 0 or 5
6	$n \equiv 0 \mod 2$ and $n \equiv 0 \mod 3$	Reduce to 2, 3 case
7	RPP, $k = -2$	$-20 \equiv 1 \mod 7 \implies 10x + y \equiv 10(x - 2y) \mod 7$
8	$n \equiv 8k \mod 10^3$	Manually divide the last 3 digits by 8 (or peel off factors of 2)
9	$\sum d_i \equiv 0 \mod 9$	9 divides the sum of digits (apply recursively)
10	$n \equiv 0 \mod 10$	Last digit is 0
11	$\sum (-1)^i d_i \equiv 0 \mod 11$ or	11 divides alternating sum

p	$p \mid n \iff$	Mnemonic
13	RPP, $k = 4$	$40 \equiv 1 \mod 13 \implies$ $10x + y \equiv 10(x + 4y) \mod 13$
17	RPP, $k = -5$	$-50 \equiv 1 \mod 17 \implies$ $10x + y \equiv 10(x - 5y) \mod 19$
19	RPP, $k = 2$	$20 \equiv 1 \mod 19 \implies$ $10x + y \equiv 10(x + 2y) \mod 19$

11.14 Hyperbolic Functions

$$\begin{aligned}
\cosh(x) &= \frac{1}{2}(e^x + e^{-x}) \\
\sinh(x) &= \frac{1}{2}(e^x - e^{-x}) \\
\cos(iz) &= \cosh z \\
\cosh(iz) &= \cos z \\
\sin(iz) &= \sinh z \\
\sinh(iz) &= \sin z \\
\sinh^{-1} x &= ? \quad = \ln(x + \sqrt{x^2 + 1}) \\
\cosh^{-1} x &= ? \quad = \ln(x + \sqrt{x^2 - 1}) \\
\tanh^{-1} x &= \frac{1}{2} \ln(\frac{1 + x}{1 - x})
\end{aligned}$$

11.15 Integral Tables

$\frac{\partial f}{\partial x} \Leftarrow$	f	$\Rightarrow \int f dx$
$\frac{1}{2\sqrt{x}}$	$\sqrt{x}$	$\frac{2}{3}x^{\frac{3}{2}}$
nx^{n-1}	$x^n, n \neq -1$	$\frac{1}{n+1}x^{n+1}$
$\frac{1}{x}$	$\ln(x)$	$x\ln(x) - x$
$a^x \ln(a)$	a^x	$\frac{a^x}{\ln a}$
$\cos(x)$	$\sin(x)$	$-\cos(x)$
$-\sin(x)$	$\cos(x)$	$\sin(x)$
$2\sec^2(x)\tan(x)$	$\sec^2(x)$	$\tan(x)$
$2\csc^2(x)\cot(x)$	$\csc^2(x)$	$-\cot(x)$
$\sec^2(x)$	$\tan(x)$	$\ln\|\sec(x)\|$
$\sec(x)\tan(x)$	$\sec(x)$	$\ln\|\sec(x) + \tan(x)\|$
$-\csc(x)\cot(x)$	$\csc(x)$	$\ln\|\csc(x) - \cot(x)\|$
$\frac{1}{1+x^2}$	$\tan^{-1}(x)$	$x\tan^{-1}x - \frac{1}{2}\ln(1+x^2)$
$\frac{1}{\sqrt{1-x^2}}$	$\sin^{-1}(x)$	$x\sin^{-1}x + \sqrt{1-x^2}$
$-\frac{1}{\sqrt{1-x^2}}$	$\cos^{-1}(x)$	$x\cos^{-1}x - \sqrt{1-x^2}$
$\frac{1}{\sqrt{x^2+a}}$	$\ln\left\|x + \sqrt{x^2+a}\right\|$	$\cdot$
$-\csc^2(x)$	$\cot(x)$	?
?	$\cos^2(x)$	?
?	$\sin^2(x)$	?
?	xe^{ax}	$\frac{1}{a^2}(ax-1)e^{ax}$
?	$e^{ax}\sin(bx)$	$\frac{1}{a^2+b^2}e^{ax}(a\sin bx - b\cos bx)$
?	$e^{ax}\cos(bx)$	$\frac{1}{a^2+b^2}e^{ax}(a\sin bx + b\cos bx)$
?	?	?

12 Definitions

12.1 Set Theory

- Injectivity

$$\begin{aligned} f : X \to Y \text{ injective } &\iff \forall x_1, x_2 \in X, \quad f(x_1) = f(x_2) \implies x_1 = x_2 \\ &\iff \forall x_1, x_2 \in X, \quad x_1 \neq x_2 \implies f(x_1) \neq f(x_2) \end{aligned}$$

- Surjectivity

$$f : X \to Y \text{ surjective } \iff \forall y \in Y,\ \exists x \in X : f(x) = y$$

- Preimage

$$f : X \to Y, U \subseteq Y \implies f^{-1}(U) = \{x \in X : f(x) \in U\}$$

12.2 Calculus

- Limit

$$\begin{aligned} \lim_{x \to p} f(x) = L \iff \forall \varepsilon,\ \exists \delta : \\ d(x,p) < \delta \implies d(f(x), L) < \varepsilon \end{aligned}$$

- Continuity
  - Epsilon-delta definition:

$$\begin{aligned} f : X \to Y \text{ continuous at } p \iff \forall \varepsilon,\ \exists \delta : \\ d_X(x,p) < \delta \implies d_Y(f(x), f(p)) < \varepsilon \end{aligned}$$

  - Limit/Sequential definition:

$$\begin{aligned} f : X \to Y \text{ continuous at } p \iff \forall \{x_i\}_{i \in \mathbb{N}} \subseteq X : \{x_i\} \to p, \\ \lim_{i \to \infty} f(x_i) = f(\lim_{i \to \infty} x_i) = f(p) \end{aligned}$$

  - Topological Definition:

$$f : X \to Y \text{ continuous } \iff U \text{ open in } \operatorname{im}(f) \subseteq Y \implies f^{-1}(U) \text{ open in } X$$

- Differentiability and the Derivative
  - For single variable functions:

$$\begin{aligned} f : \mathbb{R} \to \mathbb{R} \text{ differentiable at } p \iff \forall \{x_i\}_{i \in \mathbb{N}} \to p, \\ f'(p) \coloneqq \lim_{i \to \infty} \frac{f(x_i) - f(p)}{x_i - p} < \infty \end{aligned}$$

- For multivariable functions:

$$\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^m \text{ differentiable at } \mathbf{p} \iff \exists \text{ a linear map } \mathbf{J} : \mathbb{R}^n \to \mathbb{R}^m \text{ such that:}$$
$$\lim_{\mathbf{h}\to 0} \frac{\|\mathbf{f}(\mathbf{p}+\mathbf{h}) - \mathbf{f}(\mathbf{p}) - \mathbf{J}(\mathbf{h})\|_{\mathbb{R}^n}}{\|\mathbf{h}\|_{\mathbb{R}^m}} = 0$$

- Gradient

$$\nabla f = [f_x, f_y, f_z]$$

- Divergence
- Curl
- Taylor Series (at a point a)
  - Single Variable $\mathbb{R} \to \mathbb{R}$

$$T_a(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \cdots$$
$$\implies T_a(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!}(x-a)^n$$

  - Multivariable $\mathbb{R}^n \to \mathbb{R}$:

$$T_a(\mathbf{x}) = f(\mathbf{a}) + (\mathbf{x}-\mathbf{a})^T \nabla f(\mathbf{a})$$

  - Multivariable $\mathbb{R}^n \to \mathbb{R}^m$:

$$T_{(a,b)}(x,y) = f(a,b) + (x-a)f_x(a,b) + (y-b)f_y(a,b) +$$
$$\frac{1}{2!}\Big((x-a)^2 f_{xx}(a,b) + 2(x-a)(y-b)f_{yy}(a,b) + (y-b)^2 f_{yx}(a,b)\Big) + \cdots$$

$$T_a(\mathbf{x}) = f(\mathbf{a}) + (\mathbf{x}-\mathbf{a})^{\mathrm{T}}\mathbf{J}(\mathbf{a}) + \frac{1}{2!}(\mathbf{x}-\mathbf{a})^{\mathrm{T}}\mathbf{H}(\mathbf{a})(\mathbf{x}-\mathbf{a}) + \cdots$$
$$\implies T_a(\mathbf{x}) = \sum_{|\alpha|\geq 0} \frac{(\mathbf{x}-\mathbf{a})^\alpha}{\alpha!}\left(\partial^\alpha f\right)(\mathbf{a})$$

12.3 Analysis

- Archimedean Property: $x \in \mathbb{R} \implies \exists n \in \mathbb{N} : x < n$ and $x > 0 \implies \exists n : \dfrac{1}{n} < x$
- Upper Bound (for S $\subseteq \mathbb{R}$)

$$\alpha \text{ is an upper bound for } S \iff s \in S \implies s < \alpha$$

- Triangle Inequality
  - $|a+b| \leq |a| + |b|$
  - $|a-b| \leq |a| + |b|$

- Reverse Triangle Inequality
  - $||a| - |b|| \le |a - b|$
- Least Upper Bound / Supremum (for S $\subseteq \mathbb{R}$)
$$\alpha \text{ is a LUB for } S \iff s \in S \implies s < \alpha \text{ and } \forall t : (s \in S \implies s < t),\ \alpha < t$$
- Greatest Lower Bound / Infimum (for S $\subseteq \mathbb{R}$)
$$\alpha \text{ is a GLB for } S \iff s \in S \implies \alpha < s \text{ and } \forall t : (s \in S \implies t < s),\ t < \alpha$$
- Open Set
- Closed Set
- Limit Point
- Interior Point
- Closure of a Set
- Boundary
- Metric
- Cauchy Sequence:
$$\{a_i\} \text{ is a cauchy sequence } \iff \forall \varepsilon\ \exists N \in \mathbb{N} : \quad m, n > N \implies d(x_m, x_n) < \varepsilon$$
- Connected: S is connected $\iff \nexists U, V \subset S$ nonempty, open, disjoint such that $S = U \cup V$
- Compact: Every open cover has a finite subcover:
$$X \subseteq \cup_{j \in J} V_j \implies \exists I \subseteq J : |I| < \infty \text{ and } X \subseteq \cup_{i \in I} V_i$$
- Sequential Compactness Every sequence has a convergent subsequence:
$$\{x_i\}_{i \in I} \subseteq X \implies \exists J \subseteq I,\ \exists p \in X : \quad \{x_j\}_{j \in J} \to p$$
- Bounded (sequences, subsets, metric spaces)
$$U \subseteq X \text{ is bounded } \iff \exists x \in X, \exists M \in \mathbb{R} : \quad u \in U \implies d(x, u) < M$$
- Totally Bounded asdsa#todo
- Pointwise Convergence
$$\text{For } \{f_n : X \to Y\}_{n \in \mathbb{N}},$$
$$f_n \to f \iff \forall \varepsilon > 0,\ \forall x \in X,\ \exists N(x, \varepsilon) \in \mathbb{N} : \quad n > N \implies d_Y\left(f_n(x), f(x)\right) < \varepsilon$$
- Uniform Convergence
$$\text{For } \{f_n : X \to Y\}_{n \in \mathbb{N}},$$
$$f_n \rightrightarrows f \iff \forall \varepsilon > 0,\ \exists N(\varepsilon) \in \mathbb{N} : \quad \forall x \in X,\ n > N \implies d_Y(f_n(x), f(x)) < \varepsilon$$
- Generalized Mean Value Theorem
$$(f(b) - f(a))g'(c) = (g(b) - g(a))f'(c)$$

12.4 Linear Algebra

Convention: always over a field k, and $T : k^n \to k^m$ is a generic linear map (or $m \times n$ matrix).

- Consistent

  A system of linear equations is *consistent* when it has at least one solution.
- Inconsistent

  A system of linear equations is *inconsistent* when it has no solutions.
- Rank

  The number of nonzero rows in RREF
- Elementary Matrix
-
- Row Equivalent
-
- Pivot
-
- Cofactor

$$\text{cofactor}(A)_{i,j} = (-1)^{i+j} M_{i,j}$$

where $M_{i,j}$ is the minor obtained by deleting the i-th row and j-th column of A.

- Adjugate

$$\text{adjugate}(A) = \text{cofactor}(A)^T = (-1)^{i+j} M_{j,i}$$

- Vector Space Axioms
  - Let k be a field and $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $r, s, t \in k$. A vector space V over k satisfies:
    1. Closure under addition: $\mathbf{v} + \mathbf{w} \in V$
    2. Closure under scalar multiplication: $r\mathbf{v} \in V$
    3. Commutativity of addition: $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$
    4. Associativity of addition: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$
    5. Existence of an additive zero $\mathbf{0}$ satisfying $\mathbf{v} + 0 = 0 + \mathbf{v} = \mathbf{v}$
    6. Existence of additive inverse $-\mathbf{v}$ satisfying $v + (-\mathbf{v}) = 0$
    7. Unit property: $1\mathbf{v} = \mathbf{v}$
    8. Associativity of scalar multiplication: $(rs)\mathbf{v} = r(s\mathbf{v})$
    9. Distribution of scalars multiplication over vector addition: $r(\mathbf{v} + \mathbf{w}) = r\mathbf{v} + r\mathbf{w}$
    10. Distribution of scalar multiplication over scalar addition: $(r + s)\mathbf{v} = r\mathbf{v} + s\mathbf{v}$
- Subspace

  - A nonempty subset $W \subseteq V$ that is a vector space and satisfies

$$\left\{ \sum_i c_i \mathbf{x}_i \;\middle|\; c_i \in \mathbb{F},\; x_i \in W \right\} \subseteq W$$

  - Quick counter-check: find $\mathbf{x}, \mathbf{y}$ such that $a\mathbf{x} + b\mathbf{y} \notin W$
- Span Given a set of n vectors $S = \{\mathbf{x}_i\}_{i=1}^n$, defined as

$$\mathrm{Span}(S) = \left\{ \sum_{i=1}^n c_i \mathbf{x}_i \;\middle|\; c_i \in k \right\}$$

- Row Space
  - The range of the linear map T.
  - Given $T = \begin{bmatrix} \mathbf{x}_1 \rightarrow \\ \mathbf{x}_2 \rightarrow \\ \vdots \\ \mathbf{x}_m \rightarrow \end{bmatrix}$, defined as

$$\mathrm{Span}(\{\mathbf{x}_i\}_{i=1}^m) \subseteq k^m$$

  - $\mathrm{rowspace}(T)^\perp = \mathrm{null}(T)$
  - $|\mathrm{rowspace}(T)| = \mathrm{Rank}(T)$
- Column Space
- Null Space
  - Defined as $\mathrm{null}(T) = \left\{ \mathbf{x} \in k^n \;\middle|\; T(\mathbf{x}) = 0 \in k^m \right\}$
  - $\mathrm{null}(T)^\perp = \mathrm{rowspace}(T)$
- Eigenvalue
  - A value λ such that $Ax = \lambda x$
  - Invariant under similarity.
- Eigenspace
  - For a linear map T with eigenvalue λ, defined as $E_\lambda = \left\{ \mathbf{x} \in k^n \;\middle|\; T(\mathbf{x}) = \lambda \mathbf{x} \right\}$
- Dimension
  - The cardinality of a basis of V
- Basis
  - A linearly independent set of vectors $S = \{\mathbf{x}_i\} \subset V$ such that $\mathrm{Span}(S) = V$
- Linear independence
  - A set of vectors $\{\mathbf{x}_i\}_{i=1}^n$ is linearly independent $\iff \sum_{i=1}^n c_i \mathbf{x}_i = 0 \implies c_i = 0$ for all i.

  - Can be detected by considering the matrix

$$T = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n]^T$$

  (linearly independent iff T is singular)

- Rank
  - Dimension of rowspace
- Rank-Nullity Theorem
  - $|\text{Nullspace}(A)| + |\text{Rank}(A)| = |\text{Codomain}(A)|$
- Nullspace
  - $\text{nullspace}(A) = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\}$
- Singular
  - A square $n \times n$ matrix T is singular iff $\text{Rank}(T) < n$
- Similarity
  - Two matrices A, B are similar iff there exists an invertible matrix S such that $B = SAS^{-1}$
- Diagonalizable
  - A matrix X is diagonalizable if it can be written $X = EDE^{-1}$ where D is diagonal.
  - If X is $n \times n$ and has n linearly independent eigenvectors λ_i, then $D_{ii} = \lambda_i$, and $E = \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \downarrow & \downarrow & \cdots & \downarrow \end{bmatrix}$
- Positive Definite
  - A matrix A is positive definite iff $\forall \mathbf{x} \in k^n$, we have the scalar inequality $\mathbf{x}^T A\mathbf{x} > 0$
- Projection
  - The projection of a vector $\mathbf{v}$ onto $\mathbf{u}$ is given by $P_{\mathbf{u}}(\mathbf{v}) = \langle \mathbf{u}, \mathbf{v} \rangle \, \widehat{u}$
  - The projection of a vector $\mathbf{v}$ onto a space $U = \text{Span}(\{\mathbf{u}_i\})$ is given by

$$P_U(\mathbf{v}) = \sum_i P_{\mathbf{u}_i}(\mathbf{v}) = \sum_i \langle \mathbf{u}_i, \mathbf{v} \rangle \, \widehat{u}_i$$

- Orthogonal Complement
  - Given a subspace $U \subseteq V$, defined as $U^\perp = \left\{ \mathbf{v} \in V \;\middle|\; \forall \mathbf{u} \in U, \langle \mathbf{u}, \mathbf{v} \rangle = 0 \right\}$
- Determinant

$$\det(A) = \sum_{\tau \in S^n} \prod_{i=1}^{n} \sigma(\tau) a_{i,\tau(i)}$$

- Trace

$$\text{Tr}(A) = \sum_{i=1}^{n} A_{ii}$$

- Characteristic Polynomial
  - $p_A(x) = \det(xI - A)$
  - Roots of p_A are eigenvalues of A
- Symmetric: $A = A^T$
- Skew-Symmetric: $A = -A^T$
- Inner Product
  - $\langle \mathbf{x},\ \mathbf{x} \rangle \geq 0$
  - $\langle \mathbf{x},\ \mathbf{x} \rangle = 0 \iff \mathbf{x} = \mathbf{0}$
  - $\langle \mathbf{x},\ \mathbf{y} \rangle = \overline{\langle \mathbf{y},\ \mathbf{x} \rangle}$
  - $\langle [,\ k \rangle \mathbf{x}]\mathbf{y} = k\langle \mathbf{x},\ \mathbf{y} \rangle = \langle \mathbf{x},\ k\mathbf{y} \rangle$
  - $\langle \mathbf{x} + \mathbf{y},\ \mathbf{z} \rangle = \langle \mathbf{x},\ \mathbf{z} \rangle + \langle [,\ y \rangle]\mathbf{z}$
  - $\langle [,\ a \rangle \mathbf{x}]b\mathbf{y} = \langle \mathbf{x},\ \mathbf{x} \rangle + \langle a\mathbf{x},\ y \rangle + \langle \mathbf{x},\ b\mathbf{y} \rangle + \langle \mathbf{y},\ \mathbf{y} \rangle$
  - Defines a norm: $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x},\ \mathbf{x} \rangle} \implies \|\mathbf{x}\|^2 = \langle \mathbf{x},\ \mathbf{x} \rangle$
- Cauchy-Schwarz Inequality: $|\langle \mathbf{x},\ \mathbf{y} \rangle| \leq \|\mathbf{x}\|\|\mathbf{y}\|$
- Orthogonality:
  - For vectors: $\mathbf{x}^{\perp}\mathbf{y} \iff \langle \mathbf{x},\ \mathbf{y} \rangle = 0$
  - For matrices: A is orthogonal $\iff A^{-1} = A^T$
- Orthogonal Projection of $\mathbf{x}$ onto $\mathbf{y}$:

$$P(\mathbf{x}, \mathbf{y}) = \langle \mathbf{x},\ \mathbf{y} \rangle \widehat{y} = \langle \mathbf{x},\ \mathbf{y} \rangle \frac{\mathbf{y}}{\|\mathbf{y}\|^2}$$

  - Note $\|P(\mathbf{x}, \mathbf{y})\| = \|\mathbf{x}\| \cos \theta_{x,y}$
- Defective: An $n \times n$ matrix A is defective $\iff$ the number of linearly independent eigenvectors of A is less than n.

12.5 Differential Equations

- Homogeneous

$$f(x, y) \text{ homogeneous of degree } n \iff \exists n \in \mathbb{N} : f(tx, ty) = t^n f(x, y).$$

- Separable

$$p(y)\frac{dy}{dx} - q(x) = 0$$

- Wronskian:

$$W\left[f_1, f_2, \ldots, f_k\right](x) = \begin{vmatrix} f_1(x) & f_2(x) & \ldots & f_k(x) \\ f_1'(x) & f_2'(x) & \ldots & f_k'(x) \\ \vdots & \vdots & & \vdots \\ f_1^{(k-1)}(x) & f_2^{(k-1)}(x) & \ldots & f_k^{(k-1)}(x) \end{vmatrix}$$

- Laplace Transform:

$$L_f(s) = \int_0^\infty e^{-st} f(t)\ dt$$

12.6 Algebra

- Ring
- Group
- Subgroup
  - Two step subgroups test:
- Integral Domain
- Division Ring
- Principal Ideal Domain
- Tensor Product: #todo insert construction

12.7 Complex Analysis

- Analytic
- Harmonic
- Cauchy-Euler Equations
- Holomorphic
- The Complex Derivative
- Meromorphic
- The Gamma Function: Satisfies $\Gamma(p+1) + p\Gamma(p)$ and $\Gamma(1) = 1$, defined as

$$\Gamma(p) = \int_0^\infty t^{p-1} e^{-t} dt, \quad p > 0$$

12.8 Algebra

- Looking at real roots:
  - Let p be number of sign changes in $f(x)$;
  - Let q be number of sign changes in $f(-x)$;
  - Let n be the degree of f.
  - Then p gives the maximum number of positive real roots, q gives the maximum number of negative real roots, and $n - p - q$ gives the *minimum* number of complex roots.
  - Rational Roots Theorem: If $p(x) = ax^n + \cdots + c$ and $r = \frac{p}{q}$ where $p(r) = 0$, then $p \mid c$ and $q \mid a$.
- Properties of logs:
  - $\ln(\prod) = \sum \ln$ but $\prod \ln \neq \ln \sum$
  - $\log_b x = \frac{\ln x}{\ln b}$

Be careful! $\frac{\ln x}{\ln y} \neq \ln \frac{x}{y} = \ln x - \ln y$

- Completing the square:
  - $p(x) = ax^2 + bx + c \implies p(x) = a(x + \frac{b}{2a})^2 + -\frac{1}{2}\left(\frac{b^2 - 4ac}{2a}\right)$

12.9 Geometry

- Generic Conic Sections

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$$

$$\frac{(x - x_0)^2}{w_0} \pm \frac{(y - y_0)^2}{h_0} = c$$

- Circles:

$$Ax^2 + By^2 + C = 0 \qquad (x - x_0)^2 + (y - y_0)^2 = r^2$$

  - Defining trait: locus of points at a constant distance from the **center**
  - **Center** at (x_0, y_0)
- Parabolas:

$$Ax^2 + Bx + Cy + D = 0 \qquad y = ax^2$$

  - Defining Trait:
    - Locus of points equidistant from the **focus** (a point) and the **directrix** (a line)
    - #todo add image
  - **Focus** at $(0, \frac{1}{4a})$
  - **Directrix** at the line $y = -\frac{1}{4a}$
    - For an arbitrary quadratic: complete the square to write in the form $y = a(x - w_0)^2 + h_0$, and translate points of interest by by $(x + w_0, y + h_0)$
- Ellipses:

$$\frac{x^2}{w^2} + \frac{y^2}{h^2} = 1$$

  - Defining trait:
    - The locus of points where the *sum* of distances to two **focii** are constant.
  - **Center** at $(0, 0)$ (can translate easily)
  - **Vertices** at $(\pm w, 0)$ and $(0, \pm h)$
  - **Focii** at $F_1 = (\sqrt{w^2 - h^2}, 0), F_2 = (-\sqrt{w^2 - h^2}, 0)$
  - Another useful shortcut form:
- Hyperbolas:

$$\frac{x^2}{w^2} - \frac{y^2}{h^2} = 1$$

  - Defining trait:
    - ♢ Locus of points where the *difference* between the distances to two **focii** are constant.
  - **Vertices** at $(0, \pm h)$ and $(\pm w, 0)$
  - **Focii** at $F_1 = (\sqrt{w^2 + h^2}, 0), F_2 = (-\sqrt{w^2 + h^2}, 0)$

- Summary of Traits:
  - One point p:
    - ♢ Distance to p is constant: circle
  - Two points a, b:
    - ♢ Distance to a equal to distance to b equals a constant: a line bisecting the midpoint of the line connecting them
    - ♢ Difference of distances constant: ellipse
    - ♢ Sum of differences constant: hyperbola
  - Point p and a line l:
    - ♢ Distance to p equals distance to l equals a constant: parabola

- Areas of certain figures:

13 Indices

List of Todos

Definitions

Theorems

Exercises

List of Figures