

MATH 20700 (Honors Analysis in \mathbb{R}^n I) Notes

Steven Labalme

November 20, 2021

Contents

I	Linear Algebra	1
1	Basic Notions	2
1.1	Notes	2
1.2	Chapter 1: Basic Notions	4
2	Systems of Linear Equations	5
2.1	Notes	5
2.2	Chapter 2: Systems of Linear Equations	6
3	Determinants	9
3.1	Notes	9
3.2	Chapter 3: Determinants	10
4	Introduction to Spectral Theory	11
4.1	Notes	11
4.2	Chapter 4: Introduction to Spectral Theory	13
5	Inner Product Spaces	14
5.1	Notes	14
5.2	Chapter 5: Inner Product Spaces	18
6	Structure of Operators on Inner Product Spaces	20
6.1	Notes	20
6.2	Chapter 6: Structure of Operators on Inner Product Spaces	24
7	Bilinear and Quadratic Forms	31
7.1	Notes	31
7.2	Chapter 7: Bilinear and Quadratic Forms	33
8	Dual Spaces and Tensors	36
8.1	Notes	36
8.2	Chapter 8: Dual Spaces and Tensors	37
9	Advanced Spectral Theory	39
9.1	Notes	39
9.2	Chapter 9: Advanced Spectral Theory	42
II	Point Set Topology of Metric Spaces	43
1	The Real and Complex Number Systems	44
1.1	Notes	44
1.2	Chapter 1: The Real and Complex Number Systems	44

2	Basic Topology	46
2.1	Notes	46
2.2	Chapter 2: Basic Topology	52
3	Numerical Sequences and Series	54
3.1	Notes	54
3.2	Chapter 3: Numerical Sequences and Series	54
4	Continuity	57
4.1	Notes	57
4.2	Chapter 4: Continuity	59
7	Sequences and Series of Functions	61
7.1	Notes	61
	References	65

List of Figures

3.1	Visualizing properties of determinants.	9
5.1	The unit ball of norms corresponding to $p = 1, 2, \infty$	15
6.1	Orientation in \mathbb{R}^2	29
2.1	2-cell.	50
2.2	k -cells are compact.	50
2.3	Nonempty perfect sets are uncountable.	51
2.4	Constructing the Cantor set.	51
4.1	Set theoretic definition of continuity.	57

List of Tables

2.1	Types of sets.	48
-----	------------------------	----

Part I

Linear Algebra

Chapter 1

Basic Notions

1.1 Notes

- 9/27:
- **Vector space:** Basically, a set for which you have an addition and multiplication.
 - \mathbb{F}^d is used for \mathbb{R}^d or \mathbb{C}^d in Treil (2017).
 - \mathbb{P}_n is the vector space of polynomials up to degree n .
 - $C([0, 1])$ is the set of continuous functions defined on $[0, 1]$, an infinite-dimensional vector space.
 - **Generating set:** A subset of a vector space, all linear combinations of which generate the vector space. *Also known as spanning set.*
 - Any element of VS is a linear comb. of elements of the generating set.
 - **Linearly independent** (list): A list of vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in V$ such that $\sum_{i=1}^k \alpha_i \mathbf{v}_i = 0$ implies $\alpha_i = 0$ for all i .
 - **Base:** A generating set consisting of linearly independent vectors.
 - Any element of a VS can be written as a *unique* linear combination of the vectors in a base.
 - If $\mathbf{x} = \sum_{i=1}^k \alpha_i \mathbf{v}_i = \sum_{i=1}^k \beta_i \mathbf{v}_i$, then $\alpha_i = \beta_i$ for all i .
 - **Linear transformation:** A function $T : X \rightarrow Y$, where X, Y are VSs, such that

$$T(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha T\mathbf{x} + \beta T\mathbf{y}$$

for all $\mathbf{x} \in X, \mathbf{y} \in Y$.

- Examples of linear transformations:
 - Consider \mathbb{P}_n . Let $Tp_n = p'_n$. This T is linear.
 - Rotation in \mathbb{R}^d .
 - Think graphically about two vectors \mathbf{x}, \mathbf{y} .
 - Rotating and summing them is the same as summing and rotating. Same for scaling.
 - Thus, rotation is actually linear!
 - Reflection as well.
- Consider $T : \mathbb{R} \rightarrow \mathbb{R}$.
 - Any linear map on the line is a line.
 - We must have $Tx = \alpha x$: $Tx = T(1x) = xT(1) = x\alpha$.

- Consider $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ linear.
 - Any linear map between \mathbb{R}^n and \mathbb{R}^m is linear.
 - Thus, $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$, where A is an $m \times n$ matrix.
- To find A , do the same calculation as for $Tx = \alpha x$ but more carefully:
 - Let $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ be a basis.
 - So $\mathbf{x} = \sum_{i=1}^n \alpha_i \mathbf{e}_i$.
 - Thus, $T\mathbf{x} = \sum_{i=1}^n \alpha_i T(\mathbf{e}_i)$.
 - Each $T(\mathbf{e}_i)$ is part of the matrix that we multiply by the column vector representing \mathbf{x} .
- Multiplication of matrices is equivalent to composition of linear maps.
- Consider $T_1 : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $T_2 : \mathbb{R}^m \rightarrow \mathbb{R}^r$.
 - $T_2 \circ T_1$ is equivalent to BA , if A represents T_1 and B represents T_2 . In other words, $(T_2 \circ T_1)(\mathbf{x}) = BA\mathbf{x}$ for all \mathbf{x} .
- Recall that if $A = (\alpha_{ij})$ and $B = (\beta_{ij})$, then $(BA)_{ij} = (\sum \beta_{ik} \alpha_{kj})$.
- Properties of multiplication:

$$(AB)C = A(BC)$$

$$A(B + C) = AB + AC$$

$$(A + B)C = AC + BC$$

- However, it is not true in general that $AB = BA$.

- **Trace** (of an $n \times n$ matrix A): The sum of the diagonal entries of A . Denoted by $\text{tr}(A)$. Given by

$$\text{tr}(A) = \sum \alpha_{ii}$$

- It is true that $\text{tr}(AB) = \text{tr}(BA)$.
 - Indeed, on the diagonals, multiplication is commutative; it's the other terms that mess you up in general.
- Invertibility of matrices.
 - In general, matrices are not invertible: Not every system of equations is solveable; $Ax = b$ does not always have a solution $x = A^{-1}b$.
- C is the inverse from the left: $CA = I$. B is the inverse from the right: $AB = I$. A matrix can have a left and a right inverse and still not be invertible. A matrix is invertible iff $C = B$.
- Any time we write “inverse,” we do so under the assumption that it exists.
- $(AB)^{-1} = B^{-1}A^{-1}$ — easy proof by multiplication.
- If $A = (a_{ij})$, $A^T = (a_{ji})$.
 - $(A^{-1})^T = (A^T)^{-1}$.
 - $(AB)^T = B^T A^T$.
- Let X, Y VS.
 - $X \cong Y^{[1]}$ if there exists a linear $T : X \rightarrow Y$ that is one-to-one and onto.
 - Check: $A(\text{basis of } X) = \text{basis of } Y$. Prove by definition and expression of elements as linear combinations.
- **Subspace**: A subset of a vector space which happens to be a vector space, itself.

¹“ X is isomorphic to Y .”

1.2 Chapter 1: Basic Notions

From Treil (2017).

- 10/24:
- **Coordinates** (of $\mathbf{v} \in V$ wrt. a basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ of V): The unique scalars $\alpha_1, \dots, \alpha_n \in \mathbb{F}$ such that $\mathbf{v} = \alpha_1 \mathbf{v}_1 + \dots + \alpha_n \mathbf{v}_n$.
 - **Spanning system**: A list of vectors that spans V . *Also known as **generating system**, **complete system**.*
 - **Trivial** (linear combination): A linear combination $\alpha_1 \mathbf{v}_1 + \dots + \alpha_n \mathbf{v}_n$ of vectors such that $\alpha_k = 0$ for each $k = 1, \dots, n$.
 - **Transformation**: A function $T : X \rightarrow Y$. *Also known as **transform**, **mapping**, **map**, **operator**, and **function**.*
 - The matrix of a linear transformation T is often denoted by $[T]$.
 - To compute the reflection of vectors over an arbitrary line through the origin in \mathbb{R}^2 , represent the overall transformation as a composition of rotating the line to be the x -axis, reflecting over the x -axis, and rotating back.
 - Theorem 1.5.1: If A is an $m \times n$ matrix and B is an $n \times m$ matrix, then

$$\text{tr}(AB) = \text{tr}(BA)$$

- Theorem 1.6.1: If a linear transformation is invertible, then its left and right inverses are unique and coincide.
- The column $(1, 1)^T$ is left-invertible, with one possible left inverse being $(1/2, 1/2)$.
 - Note that it is not right invertible since its left inverses are not unique (see Theorem 1.6.1).
- An invertible matrix must be square.
- **Isomorphic** (vector spaces): Two vectors spaces V, W such that there exists an isomorphism $A : V \rightarrow W$. *Denoted by $V \cong W$.*
- Theorem 1.6.8: $A : X \rightarrow Y$ is invertible if and only if for any right side $\mathbf{b} \in Y$, the equation

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

has a unique solution $\mathbf{x} \in X$.

- Corollary 1.6.9: An $m \times n$ matrix is invertible if and only if its columns form a basis in \mathbb{F}^m .
- **Linear span** (of $\mathbf{v}_1, \dots, \mathbf{v}_n \in V$): The set of all linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_n$. *Denoted by $\mathcal{L}\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, $\text{span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$.*

Chapter 2

Systems of Linear Equations

2.1 Notes

9/29:

- Row elimination:

- Let

$$A = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 3 & 1 & 2 & 7 \\ 2 & 1 & 2 & 1 \end{pmatrix}$$

- Then the **echelon form** matrix

$$A_e = \begin{pmatrix} 1 & 2 & 3 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 2 & -4 \end{pmatrix}$$

- Lastly, the **reduced echelon form** matrix

$$A_{re} = \begin{pmatrix} 1 & 0 & 0 & 7 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -2 \end{pmatrix}$$

- **echelon form:**

- All zero rows are below nonzero rows.

- For any nonzero row, its leading element is strictly to the left of the nonzero entry of the next row.

- **Reduced echelon form:**

- All pivots are 1.

- Used to solve systems of the form $Ax = b$.

- **Inconsistent** (system of equations): A system with no solution.

- If the last row is of the form $(0, \dots, 0, b)$ where $b \neq 0$, then there is no solution.

- Unique solution if A_e has a pivot in every column.

- There exists a solution for every b if there is a pivot in every row?

- Let $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be a matrix. Then $\ker A = \{x \in \mathbb{R}^n : Ax = 0\}$ (subspace of \mathbb{R}^n) and $\text{range } A = \{Ax : x \in \mathbb{R}^n\}$ (subspace of \mathbb{R}^m).

- Also consider $\ker(A^T)$ and $\text{range}(A^T)$, the basis of the kernel and range, and dimension.

- Finite-dimensional vector spaces:

- A basis is a generating set (so every element of V can be written uniquely as a linear combination of the basis) the length of which is equal to the dimension of V .
- All bases of finite-dimensional vector spaces have the same number of elements.

- Let v_1, v_2, v_3 and w_1, w_2 be two generating sets of V .
- Then

$$v_1 = \lambda_{11}w_1 + \lambda_{12}w_2$$

$$v_2 = \lambda_{21}w_1 + \lambda_{22}w_2$$

$$v_3 = \lambda_{31}w_1 + \lambda_{32}w_2$$

- Suppose the only solution to $\alpha_1v_1 + \alpha_2v_2 + \alpha_3v_3 = 0$ is $\alpha_1 = \alpha_2 = \alpha_3 = 0$.
- But this is not true, as we can find another one in terms of the λ s.
- If you have a list of linearly independent vectors, you can complete it into a basis.
 - If there exists a vector that can't be written as a linear combination of the list, add it to the list.
- If you find any particular solution to a system $Ax = b$, and you add to it any element of $\ker A$, you will obtain another solution.
 - $Ax_1 = b$ and $Ax_h = 0$ implies that $A(x_1 + x_h) = b$.
 - $Ax_1 = b$ and $Ax_2 = b$ imply that $A(x_1 - x_2) = 0$, i.e., that $x_1 - x_2 \in \ker A$.
- If $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $\dim \text{range } A = m$, then $Ax = b$ is solvable for all $b \in \mathbb{R}^m$.
- Let $\text{rank } A = \dim \text{range } A$.
- Rank theorem:
 - $\text{rank } A = \text{rank } A^T$.
 - Let $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$. We know that $\dim \ker A + \dim \text{range } A = n$.
 - $\dim \ker A^T + \text{rank } A^T = m$.
 - This theorem survives linear algebra and enters functional analysis under the name **Fredholm's alternative**.

- **Fredholm's alternative:** $Ax = b$ has a solution for all $b \in \mathbb{R}^n$ iff $\dim \ker A^T = 0$.

- $\dim \ker A^T = 0$ implies $\text{rank } A^T = m$ implies $\text{rank } A = m$ implies $\dim \text{range } A = m$, as desired.

- **Pivot column** (of A): A column of A where A_e has pivots.
- The **pivot columns** of A give a basis for $\text{range } A$.
- The pivot rows of A_e give a basis for $\text{range } A^T$.
- A basis for the kernel is enough to solve $Ax = 0$.
- If you take these three things as givens, you can prove the rank theorem.

2.2 Chapter 2: Systems of Linear Equations

From Treil (2017).

- 10/24:
- A system is inconsistent iff the echelon form of the augmented matrix has a row of the form $(0 \ \cdots \ 0 \ b)$.
 - A solution to $Ax = b$ is unique iff there are no free variables, i.e., iff there is a pivot in every column.
 - $Ax = b$ is consistent iff the echelon form of the coefficient matrix has a pivot in every row.

- $A\mathbf{x} = \mathbf{b}$ has a unique solution for any \mathbf{b} iff the echelon form of the coefficient matrix A has a pivot in every row and column.
- Proposition 2.3.1: Let $\mathbf{v}_1, \dots, \mathbf{v}_m \in \mathbb{F}^n$, and let $A = [\mathbf{v}_1 \ \cdots \ \mathbf{v}_m]$ be an $n \times m$ matrix with columns $\mathbf{v}_1, \dots, \mathbf{v}_m$. Then
 1. The system $\mathbf{v}_1, \dots, \mathbf{v}_m$ is linearly independent iff the echelon form of A has a pivot in every column.
 2. The system $\mathbf{v}_1, \dots, \mathbf{v}_m$ is complete iff the echelon form of A has a pivot in every row.
 3. The system $\mathbf{v}_1, \dots, \mathbf{v}_m$ is a basis of \mathbb{F}^n iff the echelon form of A has a pivot in every column and in every row.
- Proposition 2.3.6: A matrix A is invertible if and only if its echelon form has a pivot in every column and every row.
- Corollary 2.3.7: An invertible matrix must be square ($n \times n$).
- Proposition 2.3.8: If a square ($n \times n$) matrix is left invertible or if it is right invertible, then it is invertible. In other words, to check the invertibility of a square matrix A , it is sufficient to check only one of the conditions $AA^{-1} = I$, $A^{-1}A = I$.
- Any invertible matrix is row-equivalent to (can be row-reduced to) the identity matrix.
- **Homogeneous** (system of linear equations): A system of the form $A\mathbf{x} = \mathbf{0}$.
- Theorem 2.6.1: Let a vector \mathbf{x}_1 satisfy the equation $A\mathbf{x} = \mathbf{b}$. and let H be the set of all solutions of the associated homogeneous system $A\mathbf{x} = \mathbf{0}$. Then the set

$$\{\mathbf{x}_1 + \mathbf{x}_h : \mathbf{x}_h \in H\}$$

is the set of all solutions to the equation $A\mathbf{x} = \mathbf{b}$.

- The pivot columns are a basis of range A . The pivot rows are a basis of range A^T . The solutions to the equation $A\mathbf{x} = \mathbf{0}$ are a basis of $\ker A$.
- Theorem 2.7.3: Let A be an $m \times n$ matrix. Then the equation $A\mathbf{x} = \mathbf{b}$ has a solution for every $\mathbf{b} \in \mathbb{R}^m$ iff the dual equation $A^T\mathbf{x} = \mathbf{0}$ has a unique (only the trivial) solution.
 - Note that this is a corollary to the rank theorem.
- Change of coordinates formula:
 - Let $T : V \rightarrow W$ be a linear transformation, and let $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ and $\mathcal{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ be bases of V and W , respectively.
 - The $m \times n$ matrix of T with respect to these bases is $[T]_{\mathcal{W}\mathcal{V}}$, and relates the coordinates of $[T\mathbf{v}]_{\mathcal{W}}$ and $[\mathbf{v}]_{\mathcal{V}}$ via

$$[T\mathbf{v}]_{\mathcal{W}} = [T]_{\mathcal{W}\mathcal{V}}[\mathbf{v}]_{\mathcal{V}}$$

- Change of coordinates matrix: If \mathcal{A}, \mathcal{B} are two bases of V , then we can convert the coordinates of a vector in \mathcal{B} to its in \mathcal{A} with the identity matrix (with respect to the appropriate bases). In particular,

$$[\mathbf{v}]_{\mathcal{B}} = [I]_{\mathcal{B}\mathcal{A}}[\mathbf{v}]_{\mathcal{A}}$$

- Note that the k^{th} column of $[I]_{\mathcal{B}\mathcal{A}}$ is the coordinate representation in \mathcal{B} of \mathbf{a}_k , i.e., $[\mathbf{a}_k]_{\mathcal{B}}$.
- The change of coordinates matrix from a basis \mathcal{B} to the standard basis \mathcal{S} is easy to compute; by the above, it's just

$$[I]_{\mathcal{S}\mathcal{B}} = [\mathbf{b}_1 \ \cdots \ \mathbf{b}_n]$$

- It follows that $[I]_{\mathcal{B}\mathcal{S}} = ([I]_{\mathcal{S}\mathcal{B}})^{-1}$.

- This allows us to compute $[I]_{\mathcal{B}\mathcal{A}}$ as $[I]_{\mathcal{B}\mathcal{S}}[I]_{\mathcal{S}\mathcal{A}}$
- If $T : V \rightarrow W$, $\mathcal{A}, \tilde{\mathcal{A}}$ are bases of V , and $\mathcal{B}, \tilde{\mathcal{B}}$ are bases of W , and we have $[T]_{\mathcal{B}\mathcal{A}}$, then

$$[T]_{\tilde{\mathcal{B}}\tilde{\mathcal{A}}} = [I]_{\tilde{\mathcal{B}}\mathcal{B}}[T]_{\mathcal{B}\mathcal{A}}[I]_{\mathcal{A}\tilde{\mathcal{A}}}$$

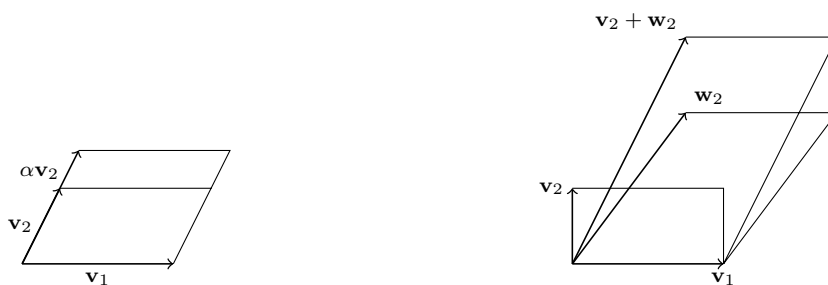
- Change of basis ends up at similarity; two operators are similar if we can change the basis of one into another.

Chapter 3

Determinants

3.1 Notes

- 9/29:
- The determinant, geometrically, is the volume of the object (in \mathbb{R}^3) you get when you take linear combinations of the vectors.
 - In 2D:
 - Let v_1, v_2 be two vectors. Put tail to tail and forming a parallelogram, the determinant of the matrix (v_1, v_2) is the area of said parallelogram.
 - Linearity 1: $D(av_1, v_2, \dots, v_n) = aD(v_1, \dots, v_n)$ is the same as saying that if you stretch one vector by a , you scale up the area by that much, too.
 - Linearity 2: $D(v_1, \dots, v_{k+} + v_{k-}, \dots, v_n) = D(-) + D(+)$.
 - Antisymmetry: $D(v_1, \dots, v_k, \dots, v_j, \dots, v_n) = -D(v_1, \dots, v_j, \dots, v_k, \dots, v_n)$. Interchanging columns flips the sign of the determinant.
 - Basis: $D(e_1, \dots, e_n) = 1$.
 - Determinant: Denoted by $D(v_1, \dots, v_n)$, where (v_1, \dots, v_n) is an $n \times n$ matrix.
- 10/1:
- Consider an $n \times n$ matrix A consisting of n columns containing vectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^n$.
 - $D(A)$ is the volume of the solid $V = \sum_{i=1}^n \alpha_i v_i$.
 - $D(\mathbf{e}_1, \dots, \mathbf{e}_n) = 1$.



(a) $D(\mathbf{v}_1, \alpha \mathbf{v}_2) = \alpha D(\mathbf{v}_1, \mathbf{v}_2)$.

(b) $D(\mathbf{v}_1, \mathbf{v}_2 + \mathbf{w}_2) = D(\mathbf{v}_1, \mathbf{v}_2) + D(\mathbf{v}_1, \mathbf{w}_2)$.

Figure 3.1: Visualizing properties of determinants.

- Basic properties of the determinant.
 - If A has a zero column, then $\det A = 0$: Scalar property.

- If A has two equal columns, then $\det A = 0$: Multiply one by minus and add.
- If A has a column which is a multiple of another, then $\det A = 0$: Pull out the multiple and then you have the previous one.
- If columns are linearly dependent, then $\det A = 0$: Decompose it into sums, split, add back up with previous properties.
- The determinant is preserved under column reduction.
- $\det A^T = \det A$: Put everything in rref.
- If A is not invertible, then $\det A = 0$ (not invertible implies linearly dependent columns, implies $\det A = 0$).
- $\det(AB) = \det A \det B$.
- Determinant of...
 - A diagonal matrix: The product of the diagonal entries (pull out the terms, and then note that the remaining identity matrix has determinant 1).
 - An upper triangular matrix: The product of the diagonal entries (column reduction to make it into a diagonal matrix, and then the property above).

3.2 Chapter 3: Determinants

From Treil (2017).

10/24: • Let $A_{j,k}$ denote the $(n-1) \times (n-1)$ matrix obtained from A by crossing out row j and column k and pushing it together.

- **Cofactors** (of A): The numbers $C_{j,k}$, one per entry, defined by

$$C_{j,k} = (-1)^{j+k} \det A_{j,k}$$

- **Cofactor matrix** (of A): The matrix

$$C = \{C_{j,k}\}_{j,k=1}^n$$

- Theorem 3.5.2: Let A be an invertible matrix and let C be its cofactor matrix. Then

$$A^{-1} = \frac{1}{\det A} C^T$$

- **Cramer's rule**: If A is invertible and $A\mathbf{x} = \mathbf{b}$, then

$$x_k = \frac{\det B_k}{\det A}$$

where B_k is obtained from A by replacing column k of A by the vector \mathbf{b} .

- **Minor** (of order k of A): The determinant of a $k \times k$ submatrix of A .
- Theorem 3.6.1: The rank of a nonzero matrix A is equal to the largest integer k such that there exists a nonzero minor of order k .

Chapter 4

Introduction to Spectral Theory

4.1 Notes

- 10/1:
- **Difference equation:** Like a differential equation, but instead of writing a differentials, you write differences.
 - Suppose we want to solve $x_{n+1} = Ax_n$ with x_0 given.
 - You will find that $x_n = A^n x_0$.
 - This gets hard to compute, so we want to find a way to simplify the computation.
 - Thus, we want to diagonalize the matrix, and this concept is inherently linked to eigenvalues and eigenvectors.
 - If you can decompose the x_0 into a linear combination of eigenvectors, then you can simplify the computation a lot:
$$x_n = \sum \alpha_i A^n v_i = \sum \alpha_i \lambda_i^n v_i$$
 - An $n \times n$ matrix will have n eigenvalues. You want n linearly independent eigenvectors, creating an eigenbasis.
 - To find eigenvalues and eigenvectors, we need to solve $Ax = \lambda x$, i.e., $(A - \lambda I)x = 0$. Thus, $\ker(A - \lambda I) \neq \{0\}$, so $\det(A - \lambda I) = 0$.
 - The eigenvalues of A are independent of the choice of basis of the domain of A or the range.
- 10/4:
- We need to know everything in Treil (2017).
 - We don't need to know the applications sections, but you should be interested.
 - **Spectral theory:** Decomposing a linear operator.
 - Let $A : V \rightarrow V$ be a linear operator. $\lambda \in \mathbb{C}$ is an eigenvalue if there exists $x \in V$ nonzero such that $Ax = \lambda x$.
 - Let A be an $n \times n$ matrix over \mathbb{C} or \mathbb{R} .
 - The eigenvalues are the roots of the polynomial $\det(A - \lambda I) = 0$ in λ .
 - Things we want to do:
 - Given A , find the eigenvalues and eigenvectors (solve $(A - \lambda I)x = 0$).

- In order to simplify A , make it a diagonal matrix:

$$A = S \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} S^{-1}$$

- Eigenvalues are independent of the choice of basis.

- From the book, we have that

$$[A]_{\mathcal{A}\mathcal{A}} = [S]_{\mathcal{A}\mathcal{B}}[B]_{\mathcal{B}\mathcal{B}}[S]_{\mathcal{A}\mathcal{B}}^{-1}$$

- It follows that

$$A - \lambda I = [S]_{\mathcal{A}\mathcal{B}}(B - \lambda I)[S]_{\mathcal{A}\mathcal{B}}^{-1}$$

so

$$\det(A - \lambda I) = \det([S]_{\mathcal{A}\mathcal{B}}(B - \lambda I)[S]_{\mathcal{A}\mathcal{B}}^{-1}) = \det([S]_{\mathcal{A}\mathcal{B}}[S]_{\mathcal{A}\mathcal{B}}^{-1}(B - \lambda I)) = \det(B - \lambda I)$$

- If $p(z) = (z - \lambda)^k q(z)$, then k is the **algebraic multiplicity** of λ . The **geometric multiplicity** of λ is $\dim \ker(A - \lambda I)$.

- These terms are not always the same, but they are related.

- Diagonalization:

- Given A that corresponds to $T : V \rightarrow V$, can we find a basis of V in which the operator is a diagonal matrix?
- $A = SDS^{-1}$ iff there exists a basis of V consisting of the eigenvectors of A .
- Proves $A^N = SD^N S^{-1}$ via $A^2 = SDS^{-1}SDS^{-1} = SDIDS^{-1} = SD^2 S^{-1}$.

- Let A be an $n \times n$ matrix over \mathbb{F} . If $\lambda_1, \dots, \lambda_r$ are distinct eigenvalues, then their eigenvectors are linearly independent.

- Prove with induction contradiction argument. Assume true for \mathbf{v}_{r-1} . Then

$$0 = (A - \lambda_r I)[\mathbf{v}_1 + \dots + \mathbf{v}_r] = (\lambda_1 - \lambda_r)\mathbf{v}_1 + \dots + (\lambda_{r-1} - \lambda_r)\mathbf{v}_{r-1}$$

- Implies $\lambda_r = \lambda_i$ for all $i \in [r-1]$, a contradiction.
- If A has n distinct eigenvalues, then A is diagonalizable.

- If $A : V \rightarrow V$ has n complex eigenvalues, then A is diagonalizable iff the algebraic multiplicity equals the geometric multiplicity for each eigenvalue.

- Goes through a sample diagonalization with $\begin{pmatrix} 1 & 2 \\ 8 & 1 \end{pmatrix}$.

- We have

$$A - \lambda I = \begin{pmatrix} 1 - \lambda & 2 \\ 8 & 1 - \lambda \end{pmatrix}$$

so

$$0 = \det(A - \lambda I) = (1 - \lambda)^2 - 16$$

- It follows that $\lambda = 5, -3$.
- This yields

$$\begin{pmatrix} 1 & 2 \\ 8 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 5 & 0 \\ 0 & -3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 2 & -2 \end{pmatrix}^{-1}$$

by inspection.

- As another example, consider $\begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}$.

- Here, we have $\lambda = 1 \pm 2i$.

4.2 Chapter 4: Introduction to Spectral Theory

From Treil (2017).

10/24:

- **Spectrum** (of A): The set of all eigenvalues of A . Denoted by $\sigma(A)$.
- Proposition 4.1.1: The geometric multiplicity of an eigenvalue cannot exceed its algebraic multiplicity.
- Theorem 4.2.1: A matrix A (with values in \mathbb{F}) admits a representation $A = SDS^{-1}$ where D is a diagonal matrix and S is invertible if and only if there exists a basis of \mathbb{F}^n of eigenvectors of A . Moreover, in this case diagonal entries of D are the eigenvalues of A and columns of S are the corresponding eigenvectors.
- Any operator on a complex vector space has n eigenvalues (counting multiplicities).
 - Think n necessary roots of the characteristic polynomial, or the necessary upper triangular representation.
- Theorem 4.2.8: Let an operator $A : V \rightarrow V$ have exactly $n = \dim V$ eigenvalues (counting multiplicities). Then A is diagonalizable if and only if for each eigenvalue λ , the dimension of the eigenspace $\ker(A - \lambda I)$ (i.e., the geometric multiplicity of λ) coincides with the algebraic multiplicity of λ .
- Theorem 4.2.9: A real $n \times n$ matrix A admits a real factorization (i.e., a real representation $A = SDS^{-1}$ where S and D are real matrices, D is diagonal, and S is invertible) if and only if it admits a complex factorization and all eigenvalues of A are real.
- Example of a nondiagonalizable matrix:

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

- $p(\lambda) = (1 - \lambda)^2$, so $\lambda = 1$ with algebraic multiplicity 2.
- However, $\dim \ker(A - I) = 1$ since $A - I$ has only one pivot, hence $2 - 1 = 1$ free variable.
- Thus, apply Theorem 4.2.8.

Chapter 5

Inner Product Spaces

5.1 Notes

10/6: • We define

$$\ell^2(\mathbb{R}) = \left\{ \{a_n\}_{n \geq 1} \subset \mathbb{R} : \sum_1^\infty |a_n|^2 < \infty \right\}$$

• **Inner product:** A map $V \times V \rightarrow \mathbb{F}$ that takes $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x} \cdot \mathbf{y}$. Denoted by $\cdot, (\cdot, \cdot), \langle \cdot, \cdot \rangle$.

• Properties of the inner product:

- $(\mathbf{x}, \mathbf{y}) = \overline{(\mathbf{y}, \mathbf{x})}$ (symmetry).
- $(\alpha \mathbf{x} + \beta \mathbf{y}, \mathbf{z}) = \alpha(\mathbf{x}, \mathbf{z}) + \beta(\mathbf{y}, \mathbf{z})$ (linearity).
- $(\mathbf{x}, \mathbf{x}) \geq 0$.
- $(\mathbf{x}, \mathbf{x}) = 0$ iff $\mathbf{x} = 0$.

• If $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i y_i$$

• If $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$, then

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n x_i \bar{y}_i$$

• If $f, g \in \mathbb{P}_n(t)$, then

$$(f, g) = \int_{-1}^1 f \bar{g} dt$$

- The conjugate of a polynomial is the polynomial with the conjugate of the coefficients of the original polynomial. Symbolically, if $f = \sum_{i=0}^n \alpha_i t^i$ is a polynomial, then $\bar{f} = \sum_{i=0}^n \bar{\alpha}_i t^i$.

• It is a fact that

$$\left| \sum_{n=1}^{\infty} a_n \bar{b}_n \right| \leq \|(a_n)_{n \geq 1}\| \|(b_n)_{n \geq 1}\|$$

• Suppose we want to define the inner product between two matrices.

- A common one is

$$(A, B) = \text{tr}(B^* A)$$

where $B^* = \bar{B}^T = \overline{B^T}$ is the conjugate transpose.

- We define the norm as a function $V \rightarrow [0, \infty)$ given by

$$\|\mathbf{x}\| = \sqrt{(\mathbf{x}, \mathbf{x})}$$

- Properties of the norm.

- $\|\alpha\mathbf{x}\| = |\alpha|\|\mathbf{x}\|$.
- $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.
- $\|\mathbf{x}\| = 0$ iff $\mathbf{x} = 0$.

- In \mathbb{R}^n ,

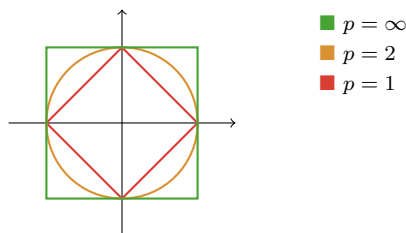


Figure 5.1: The unit ball of norms corresponding to $p = 1, 2, \infty$.

- The standard norm is

$$\|\mathbf{x}\| = \sqrt{\sum |x_i|^2}$$

- We can also define

$$\|\mathbf{x}\|_p = \sqrt[p]{\sum |x_i|^p}$$

- We can even define

$$\|\mathbf{x}\|_\infty = \max |x_i|$$

- And we can prove that all of these are valid norms.
- Only the norm corresponding to ℓ^2 is given by an inner product, but all the other quantities are still norms as defined by the properties (see Treil (2017)).
- Figure 5.1 shows the unit ball of each norm, i.e., the set of all points which have norm 1.

- The parallelogram rule:

$$\|\mathbf{x} + \mathbf{y}\|^2 + \|\mathbf{x} - \mathbf{y}\|^2 = 2(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2)$$

- Orthogonality: Given \mathbf{v}, \mathbf{w} , if $\mathbf{v} \perp \mathbf{w}$, then $(\mathbf{v}, \mathbf{w}) = 0$.

- In particular, if $\mathbf{v} \perp \mathbf{w}$, then

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$$

- Let E be a subspace of V . If $\mathbf{v} \perp E$, then $\mathbf{v} \perp \mathbf{e}$ for all $\mathbf{e} \in E$, i.e., $\mathbf{v} \perp$ a set of vectors spanning E .
- Any set of orthogonal vectors is linearly independent. Thus, if V is n dimensional, then $\mathbf{v}_1, \dots, \mathbf{v}_n$ orthogonal is a basis.
- Let E be a subspace of V . Take $\mathbf{v} \in V$. We want to define the projection $P_E \mathbf{v}$ of \mathbf{v} onto E .
 - We have that $P_E \mathbf{v} \in E$ and $\mathbf{v} - P_E \mathbf{v} \perp E$.
 - Additionally, we have that

$$\|\mathbf{v} - P_E \mathbf{v}\| \leq \|\mathbf{v} - \mathbf{e}\|$$

for all $\mathbf{e} \in E$.

- Lastly, we have that $P_E \mathbf{v}$ is unique.
- If we receive a basis of a vector space, how do we create out of that a basis that is orthogonal? The process of doing this is called **Gram-Schmidt orthogonalization**.
 - We keep \mathbf{v}_1 , subtract $P_{\mathbf{v}_1} \mathbf{v}_2$ from \mathbf{v}_2 , subtract $P_{\{\mathbf{v}_1, \mathbf{v}_2\}} \mathbf{v}_3$ from \mathbf{v}_3 , and on and on.
- If we are given a set of orthogonal vectors, we can normalize them by dividing each by its norm. This creates an orthonormal list. The standard basis is orthonormal.
- Let

$$E^\perp = \{v \in V : v \perp E\}$$

- It follows that $V = E \oplus E^\perp$.
- How close can we come to solving $A\mathbf{x} = \mathbf{b}$ if we cannot solve it exactly (i.e., if the columns are not linearly independent)?
 - Let A be an $m \times n$ matrix, and let $\mathbf{b} \in \mathbb{R}^m$.
 - Then the best solution is given by minimizing $\|A\mathbf{x} - \mathbf{b}\|$. We minimize this with projections. A special case of this is least squares regression! More details in Treil (2017).

10/8:

- Soug is gonna send us a hefty amount of reading for the weekend.
- Least square approximation:
 - If we want to minimize $\|A\mathbf{x} - \mathbf{b}\|$, the best we can do is project \mathbf{b} onto the range of A .
 - Let $\mathbf{v}_1, \dots, \mathbf{v}_k$ be an orthogonal basis of range A .
 - Then

$$\text{Proj}_{\text{range } A} \mathbf{b} = \sum_{k=1}^k \frac{(\mathbf{b}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2} \mathbf{v}_k$$

- Matrix equation form:

$$\text{Projection}_{\text{range } A} = A(A^*A)^{-1}A^*$$

if A^*A is invertible, where $A^* = \bar{A}^T$.

■ Soug never uses this though.

- The minimum is found when $\mathbf{b} - A\mathbf{x} \perp \text{range } A$. Implies that $\mathbf{b} - A\mathbf{x} \perp \mathbf{a}_k$ for all k . Implies $(\mathbf{b} - A\mathbf{x}, \mathbf{a}_k) = \bar{\mathbf{a}}_k^T (\mathbf{b} - A\mathbf{x}) = 0$.
- Note that we're letting $\bar{\mathbf{a}}_k^T$ be the row vector

$$\bar{\mathbf{a}}_k^T = (\bar{a}_{1,k} \quad \cdots \quad \bar{a}_{n,k})$$

- We also have $\bar{A}^T (\mathbf{b} - A\mathbf{x}) = 0$, from which it follows that $A^*A\mathbf{x} = A^*\mathbf{b}$, so $\mathbf{x} = (A^*A)^{-1}A^*\mathbf{b}$. Thus, $\text{Proj}_{\text{range } A} = Ax$, so $\text{Proj}_{\text{range } A} = A(A^*A)^{-1}A^*$.
- Adjoint of a linear map $T : V \rightarrow W$ is the A^* discussed above.
 - First, we'll do this for matrices. And then we'll do it for any finite-dimensional vector space.
 - Let A be an $m \times n$ matrix. We claim then that

$$(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, A^*\mathbf{y})$$

for all $\mathbf{x} \in \mathbb{C}^n, \mathbf{y} \in \mathbb{C}^m$. Proof:

$$\begin{aligned} (A\mathbf{x}, \mathbf{y}) &= \bar{\mathbf{y}}^T A\mathbf{x} \\ &= \mathbf{y}^* A\mathbf{x} \\ &= (A^*\mathbf{y})^* \mathbf{x} \\ &= (\mathbf{x}, A^*\mathbf{y}) \end{aligned}$$

- Properties of the adjoint:

$$(AB)^T = B^T A^T$$

$$(AB)^* = B^* A^*$$

$$(A^*)^* = A$$

- A^* is the unique matrix B such that $(A\mathbf{x}, \mathbf{y}) = (\mathbf{x}, B\mathbf{y})$.
- Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be a basis of V , and let $\mathbf{w}_1, \dots, \mathbf{w}_m$ be a basis of W .
- Definition of A^* : If $(A\mathbf{x}, \mathbf{y}) = (y, A^*\mathbf{x})$ for all $\mathbf{x} \in V$ and $\mathbf{y} \in W$.
- But it's not enough to define something; we have to check that it exists.
- If $[A]_{AB}$, then $[A^*]_{AB}$.
- More properties (give criteria for solving systems of equations):

$$\ker A^* = (\text{range } A)^\perp$$

$$\ker A = (\text{range } A^*)^\perp$$

$$\text{range } A = (\ker A^*)^\perp$$

$$\text{range } A^* = (\ker A)^\perp$$

■ Soug proves these.

• Isometries and unitary operators.

- $U : X \rightarrow Y$ is an isometry if $\|\mathbf{x}\| = \|U\mathbf{x}\|$ for all $\mathbf{x} \in X$. It is an isometry because it preserves the distance between points.
- It immediately follows that $\|\mathbf{x}_1 - \mathbf{x}_2\| = \|U\mathbf{x}_1 - U\mathbf{x}_2\| = \|U(\mathbf{x}_1 - \mathbf{x}_2)\|$.
- This definition is equivalent to an inner product one: $(\mathbf{x}, \mathbf{y}) = (U\mathbf{x}, U\mathbf{y})$. This follows from the definition of the norm.
- We have

$$(\mathbf{a}, \mathbf{b}) = \frac{1}{4} \sum_{\alpha=\pm 1, \pm i} \alpha \|\mathbf{a} + \alpha \mathbf{b}\|^2$$

■ $(a+b)^2 - (a-b)^2 = 4ab$ for any $a, b \in \mathbb{R}$, so $ab = \frac{1}{4}[(a+b)^2 - (a-b)^2]$. Thus, in a real inner product space,

$$(\mathbf{a}, \mathbf{b}) = \frac{1}{4} \left(\|\mathbf{a} + \mathbf{b}\|^2 - \|\mathbf{a} - \mathbf{b}\|^2 \right)$$

■ It follows that isometries preserve inner products.

- U is an isometry if and only if $U^*U = I$. Proof:

$$(\mathbf{x}, \mathbf{x}) = (U\mathbf{x}, U\mathbf{y}) = (U^*U\mathbf{x}, \mathbf{x})$$

$$(\mathbf{x}, \mathbf{y}) = (U\mathbf{x}, U\mathbf{y}) = (U^*U\mathbf{x}, \mathbf{y})$$

$$(\mathbf{x}, \mathbf{y}) = (U\mathbf{x}, U\mathbf{y}) = (\mathbf{x}, \mathbf{y})$$

for all \mathbf{y} .

- An isometry is unitary if it is invertible.

■ Thus, $U : X \rightarrow Y$ an isometry is unitary iff $\dim X = \dim Y$.

- Note that it follows that $U^* = U^{-1}$ for U an isometry.
- U unitary implies $|\det U| = 1$, so λ an eigenvalue of U implies that $|\lambda| = 1$.
- A is diagonalizable iff it has an orthogonal basis of eigenvectors.

5.2 Chapter 5: Inner Product Spaces

From Treil (2017).

- 10/24: • **Standard inner product** (on \mathbb{C}^n): The inner product (\mathbf{z}, \mathbf{w}) defined by

$$(\mathbf{z}, \mathbf{w}) = \mathbf{w}^* \mathbf{z}$$

- Corollary 5.1.5: Let \mathbf{x}, \mathbf{y} be vectors in an inner product space V . The equality $\mathbf{x} = \mathbf{y}$ holds if and only if

$$(\mathbf{x}, \mathbf{z}) = (\mathbf{y}, \mathbf{z})$$

for all $\mathbf{z} \in V$.

- Corollary 5.1.6: Suppose two operator $A, B : X \rightarrow Y$ satisfy

$$(A\mathbf{x}, \mathbf{y}) = (B\mathbf{x}, \mathbf{y})$$

for all $\mathbf{x} \in X$ and $\mathbf{y} \in Y$. Then $A = B$.

- **Normed space:** A vector space V equipped with a norm that satisfies properties of homogeneity, the triangle inequality, non-negativity, and non-degeneracy.
- Any inner product space is naturally a normed space.
- If $1 \leq p < \infty$, we can define a corresponding norm on \mathbb{R}^n or \mathbb{C}^n by

$$\|\mathbf{x}\|_p = \left(\sum_{k=1}^n |x_k|^p \right)^{1/p}$$

- We can also define the norm for $p = \infty$ by

$$\|\mathbf{x}\|_\infty = \max\{|x_k| : k = 1, \dots, n\}$$

- Note that the norm of this form for $p = 2$ is the usual norm.
- These norms are heavily associated with Figure 5.1.

- **Minkowski inequality:** One of the triangle inequalities for norms with $p \neq 2$.
- Theorem 5.1.11: A norm in a normed space is obtained from some inner product if and only if it satisfies the Parallelogram Identity

$$\|\mathbf{u} + \mathbf{v}\|^2 + \|\mathbf{u} - \mathbf{v}\|^2 = 2(\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2)$$

for all $\mathbf{u}, \mathbf{v} \in V$.

- It follows that norms with $p \neq 2$ do not have associated inner products, since such norms fail to satisfy the parallelogram identity.

- Lemma 5.2.5 (Generalized Pythagorean Identity): Let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be an orthogonal system. Then

$$\left\| \sum_{k=1}^n \alpha_k \mathbf{v}_k \right\|^2 = \sum_{k=1}^n |\alpha_k|^2 \|\mathbf{v}_k\|^2$$

- Proposition 5.3.3: Let $\mathbf{v}_1, \dots, \mathbf{v}_r$ be an orthogonal basis in E . Then the orthogonal projection $P_E \mathbf{v}$ of a vector \mathbf{v} is given by the formula

$$P_E \mathbf{v} = \sum_{k=1}^r \frac{(\mathbf{v}, \mathbf{v}_k)}{\|\mathbf{v}_k\|^2} \mathbf{v}_k$$

– It follows that

$$\begin{aligned} P_E \mathbf{v} &= \sum_{k=1}^r \frac{\mathbf{v}_k^* \mathbf{v}}{\|\mathbf{v}_k\|^2} \mathbf{v}_k \\ &= \sum_{k=1}^r \frac{1}{\|\mathbf{v}_k\|^2} \mathbf{v}_k \mathbf{v}_k^* \mathbf{v} \\ &= \left(\sum_{k=1}^r \frac{1}{\|\mathbf{v}_k\|^2} \mathbf{v}_k \mathbf{v}_k^* \right) \mathbf{v} \end{aligned}$$

– Thus, we have that

$$P_E = \sum_{k=1}^r \frac{1}{\|\mathbf{v}_k\|^2} \mathbf{v}_k \mathbf{v}_k^*$$

- **Gram-Schmidt orthogonalization:** Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be a linearly independent system of vectors to orthogonalize. Then $\mathbf{v}_1 = \mathbf{x}_1$, $\mathbf{v}_2 = \mathbf{x}_2 - P_{\text{span}\{\mathbf{v}_1\}} \mathbf{x}_2$, $\mathbf{v}_3 = \mathbf{x}_3 - P_{\text{span}\{\mathbf{v}_1, \mathbf{v}_2\}} \mathbf{x}_3$, and on and on.
- To find the least squares solution to $A\mathbf{x} = \mathbf{b}$, solve $A\mathbf{x} = P_{\text{range } A} \mathbf{b}$.
 - We can do this by finding an orthogonal basis of range A and then applying the projection formula.
 - Alternatively, we can use the following formula to speed things up if A^*A is invertible:

$$P_{\text{range } A} \mathbf{b} = A(A^*A)^{-1}A^*\mathbf{b}$$

- Theorem 5.4.1: For an $m \times n$ matrix A ,

$$\ker A = \ker(A^*A)$$

- Thus, A^*A is invertible iff A is invertible iff A is full rank. This gives us a condition on when we can use the projection formula.
- Theorem 5.6.1: An operator $U : X \rightarrow Y$ is an isometry if and only if it preserves the inner product, i.e., if and only if

$$(\mathbf{x}, \mathbf{y}) = (U\mathbf{x}, U\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y} \in X$.

- Lemma 5.6.2: An operator $U : X \rightarrow Y$ is an isometry if and only if $U^*U = I$.
- **Unitary** (operator): An invertible isometry.
- Proposition 5.6.3: An isometry $U : X \rightarrow Y$ is a unitary operator iff $\dim X = \dim Y$.
- **Orthogonal** (matrix): A unitary matrix with real entries.
- Unitary operator properties:
 1. $U^{-1} = U^*$.
 2. U unitary implies $U^* = U^{-1}$ unitary.
 3. If $\mathbf{v}_1, \dots, \mathbf{v}_n$ is orthonormal, $U\mathbf{v}_1, \dots, U\mathbf{v}_n$ is orthonormal.
 4. U, V unitary implies UV unitary.
- A matrix U is an isometry iff its columns form an orthonormal system.
- Proposition 5.6.4: Let U be a unitary matrix. Then
 1. $|\det U| = 1$. In particular, if U is orthogonal, then $\det U = \pm 1$.
 2. $|\lambda| = 1$ for every eigenvalue λ of U .
- Proposition 5.6.5: A matrix A is unitarily equivalent to a diagonal one iff it has an orthogonal (or-thonormal) basis of eigenvectors.

Chapter 6

Structure of Operators on Inner Product Spaces

6.1 Notes

10/11:

- Spectral decomposition of self-adjoint linear maps.
 - Can we write a map in term of the eigenvalues only?
 - Let $A : X \rightarrow X$ be linear and self-adjoint. Where $\dim X < \infty$.
 - Let A have eigenvalues $\lambda_1, \dots, \lambda_n$ and eigenvectors $\mathbf{v}_1, \dots, \mathbf{v}_n$. Then there is an orthonormal basis of X consisting of eigenvectors of A . An operator is self-adjoint if $A = A^*$.
 - If A is self-adjoint, then A can be written as diagonal with the eigenvalues on the diagonal with respect to some orthonormal basis of eigenvectors.
 - Let $\mathbb{F} = \mathbb{C}$.
- If there exists an orthonormal basis u_1, \dots, u_n of X such that A is triangular, then $A = UTU^*$ where U is unitary and T is upper triangular.
 - Proved with induction on $\dim X$.
 - $\dim X = 1$ is clear.
 - Assume for $\dim X = n - 1$, WTS for $\dim X = n$.
 - The subspace has a basis $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$ such that A has a diagonal form.
 - Let $u \in X$ be linearly independent of $\mathbf{v}_1, \dots, \mathbf{v}_{n-1}$.
 - Let λ be the remaining eigenvalue and u the corresponding eigenvector. Let $E = \text{span}(u)$. Then make the matrix λ in the upper left corner, and block diagonal with “ A_{n-1} ” in the bottom right corner, zeroes everywhere else.
- **Self-adjoint** (matrix A): A linear map $A : X \rightarrow X$ where $\dim X < \infty$ such that $A = A^*$.
 - Similarly, $(Ax, y) = (x, Ay)$.
 - A self-adjoint implies all eigenvalues are real, eigenvectors corresponding to different eigenvalues are orthogonal.
 - Soug proves this.
- **Strictly positive** (operator A): A self-adjoint operator $A : X \rightarrow X$ such that $(Ax, x) > 0$ for all $x \neq 0$. Also known as **positive definite**.
 - Implies that all eigenvalues are strictly positive.

- **Nonnegative** (operator A): A self-adjoint operator $A : X \rightarrow X$ such that $(Ax, x) \geq 0$ for all $x \neq 0$. Also known as **definite**.

- All eigenvalues are nonnegative.

- Suppose $A \geq 0$ is self-adjoint. Then there exists a unique self-adjoint $B \geq 0$ such that $B^2 = A$.

- A self-adjoint is diagonal (wrt. some basis).
- A positive means that all eigenvalues (diagonal entries) are positive.
- Thus, take

$$B = \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sqrt{\lambda_n} \end{pmatrix}$$

- Suppose $B^2 = A$, $C^2 = A$. Then we have an orthonormal basis corresponding to B and an orthonormal basis corresponding to C . It follows that $B^2 = C^2 = A$. Write B^2x and C^2x in terms of their bases; will necessitate that the bases are the same.

10/13:

- If we get yes/no questions, we don't have to justify.
- Cauchy-Schwarz inequality:

$$|(\mathbf{x}, \mathbf{y})| \leq \|\mathbf{x}\| \|\mathbf{y}\|$$

- Real spaces, V vs. (\cdot, \cdot) inner product.
- Proof:

$$\begin{aligned} 0 &\leq \|\mathbf{x} + t\mathbf{y}\|^2 \\ &= t^2 \|\mathbf{y}\|^2 + 2t(\mathbf{x}, \mathbf{y}) + \|\mathbf{x}\|^2 \end{aligned}$$

Thus, the discriminant must be less than zero (because the whole polynomial is positive, so the discriminant [the opposite of the x^0 term of the factored form of the polynomial] must be less than zero so the polynomial doesn't get dragged down to negative values):

$$(\mathbf{x}, \mathbf{y})^2 - \|\mathbf{x}\|^2 \|\mathbf{y}\|^2 \leq 0$$

Taking square roots of both sides proves the desired inequality.

- Recall that if $A^* = A$, then all eigenvalues are real and all eigenvectors of distinct eigenvalues are orthogonal to each other.
- **Normal** (matrix): A matrix N such that $N^*N = NN^*$.
 - Examples: Diagonal, self-adjoint, and unitary operators are all normal.
- Any normal operator in a complex vector space has an orthonormal set of eigenvectors, e.g., $N = UDU^*$.
 - Proof: N is upper triangular wrt. some basis (because all matrices are). WTS any normal upper triangular matrix is diagonal. Done by induction on the dimension of N from $n = 2$.
 - Assume the claim for every $(n - 1) \times (n - 1)$ normal upper triangular matrix.
 - Let

$$N = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & & & \\ 0 & & & \\ \vdots & & N_1 & \\ 0 & & & \end{pmatrix}$$

(we know every normal matrix can be written in this upper triangular form)

- Then just compute NN^* and N^*N . Knowing they have to be equal, we have that $a_{12} = \dots = a_{1n} = 0$.
- We can also prove from the above (block diagonal multiplication) that N_1 is normal. Thus, it's diagonal, too. Therefore, the whole thing is diagonal.
- N is normal if and only if $\|N\mathbf{x}\| = \|N^*\mathbf{x}\|$.
 - Proof: $(N\mathbf{x}, N\mathbf{y}) = (N^*N\mathbf{x}, \mathbf{y}) = (NN^*\mathbf{x}, \mathbf{y}) = (N^*\mathbf{x}, N^*\mathbf{y})$. This is equivalent to the desired condition.
- If A is nonnegative and $(A\mathbf{e}_k, \mathbf{e}_k) = a_{kk}$, then

$$\sum_{i,j=1}^n a_{ij} \mathbf{x}_i \mathbf{x}_j$$

- **Positive definite** (matrix): An $n \times n$ self-adjoint matrix such that $(A\mathbf{x}, \mathbf{x}) > 0$ for all $\mathbf{x} \in X$.
- Let $A : X \rightarrow Y$, $\dim X = \dim Y$. Then AA^* is positive semidefinite. And there exists a unique square root $R = \sqrt{A^*A}$.
 - Proof: $(A^*A\mathbf{x}, \mathbf{x}) = (A\mathbf{x}, A\mathbf{x}) = \|A\mathbf{x}\|^2 \geq 0$.
- **Modulus** (of A): The matrix $|A| = \sqrt{A^*A}$.
- Check $\| |A|\mathbf{x} \| = \|A\mathbf{x}\|$.

$$\| |A|\mathbf{x} \|^2 = (|A|\mathbf{x}, |A|\mathbf{x}) = (|A|^*|A|\mathbf{x}, \mathbf{x}) = (A^*A\mathbf{x}, \mathbf{x}) = (A\mathbf{x}, A\mathbf{x}) = \|A\mathbf{x}\|^2$$

- Let $A : X \rightarrow X$ be a linear operator. Then $A = U|A|$ where U is unitary.
- Look at singular matrices.

10/15:

- Recall that if $A : X \rightarrow Y$, we have that A^*A is semidefinite, positive, and self adjoint.
 - Thus, there exists a unique matrix $R = \sqrt{A^*A} \geq 0$, which we define to be $|A| = \sqrt{A^*A}$.
- Polar form of a matrix:

$$A = U|A|$$

- This may not be unique!
- Proof: Suppose $A\mathbf{x} = U(|A|\mathbf{x})$. $A\mathbf{x} \in \text{range } A$, and $|A|\mathbf{x} \in \text{range } (|A|)$. $\mathbf{x} \in \text{range } (|A|)$ implies that there exists $\mathbf{v} \in X$ such that $\mathbf{x} = |A|\mathbf{v}$.
- Define $U\mathbf{x} = A\mathbf{x}$. U is a well-defined linear map.
- $\|U\mathbf{x}\| = \|A\mathbf{x}\| = \||A|\mathbf{v}\| = \|\mathbf{x}\|$.
- U is an isometry.
- $\text{range } |A| \rightarrow X$.
- Use $\ker A = \ker |A| = (\text{range } A)^\perp$ to extend U_0 to U : $U = U_0 + U_1$.
- **Singular values** (of a matrix): The eigenvalues of $|A|$.
 - So if $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A^*A , the singular values of A are $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$.
- Let $A : X \rightarrow Y$ be a linear map.
 - Let $\sigma_1, \dots, \sigma_n$ be the singular values of A . Then $\sigma_1, \dots, \sigma_n > 0$.
 - Additionally, if $\mathbf{v}_1, \dots, \mathbf{v}_n$ is an orthonormal basis of eigenvectors of A^*A , then the list of n vectors $\mathbf{w}_1, \dots, \mathbf{w}_n$ defined by $\mathbf{w}_k = 1/\sigma_k A\mathbf{v}_k$ for each $k = 1, \dots, n$ is orthonormal.

■ Proof:

$$(\mathbf{w}_k, \mathbf{w}_j) = \frac{1}{\sigma_k \sigma_j} (A\mathbf{v}_k, A\mathbf{v}_j) = \frac{1}{\sigma_k \sigma_j} = \frac{1}{\sigma_k \sigma_j} (A^* A \mathbf{v}_k, \mathbf{v}_j) = \frac{\sigma_k^2}{\sigma_k \sigma_j} (\mathbf{v}_k, \mathbf{v}_j) = 0$$

and

$$\|\mathbf{w}_k\| = \frac{1}{\sigma_k} \|A\mathbf{v}_k\| = \frac{1}{\sigma_k} \||A|\mathbf{v}_k\| = 1$$

– Schmidt decomposition of A :

$$A\mathbf{x} = \sum_{k=1}^r \sigma_k (\mathbf{x}, \mathbf{v}_k) \mathbf{w}_k$$

■ This is because $\mathbf{x} = \sum (\mathbf{x}, \mathbf{v}_k) \mathbf{v}_k$, so by the above,

$$A\mathbf{x} = \sum_{k=1}^n (\mathbf{x}, \mathbf{v}_k) A\mathbf{v}_k = \sum_{k=1}^r \sigma_k (\mathbf{x}, \mathbf{v}_k) \mathbf{w}_k$$

• **Operator norm:** $\|A\| = \max\{\|A\mathbf{x}\| : \|\mathbf{x}\| \leq 1\}$.

• Properties of the operator norm:

- $\|A\mathbf{x}\| \leq \|A\| \|\mathbf{x}\|$.
- $\|\alpha A\| = |\alpha| \|A\|$.
- $\|A + B\| \leq \|A\| + \|B\|$.
- $\|A\| \geq 0$.
- $\|A\| = 0$ iff $A = 0$.

• **Frobenius norm:** The norm $\|A\|_2^2 = \text{tr}(A^* A)$.

• The operator norm is always less than or equal to the Frobenius norm.

• If $A : \mathbb{F}^n \rightarrow \mathbb{F}^n$, then $A = W\Sigma V^*$ where σ is a diagonal matrix of nonzero singular values.

• The operator norm of A is the largest of the singular values.

• An orthogonal matrix can be decomposed to a block-diagonal matrix of rotations.

10/18:

• Soug tests what he teaches and doesn't give super tricky questions.

• Structure of orthogonal matrices.

• **Orthogonal (matrix):** A unitary matrix U with all elements real and $|\det U| = 1$.

• Theorem: Let U be an orthogonal operator on \mathbb{R}^n such that $\det U = 1$. Then there exists an orthonormal basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ such that with respect to this basis,

$$U = \begin{pmatrix} R_{\phi_1} & & & \mathbf{0} \\ & \ddots & & \\ & & \mathbb{R}_{\phi_k} & \\ \mathbf{0} & & & I_{n-2k} \end{pmatrix}$$

where each R_{ϕ_i} is a 2×2 rotation matrix.

- If you are in \mathbb{R}^7 for example, you would be able to express U as a composition of at most 3 rotation maps and the identity map.
- Each rotation map acts on two orthonormal vectors.
- Proof: $P(\lambda)$ is the n -degree characteristic polynomial $\det(U - \lambda I) = 0$. The eigenvalues are the roots of it.

- $p(\lambda) = 0$ if and only if $p(\bar{\lambda}) = 0$.
 - $\lambda \in \mathbb{C}$ is an eigenvalue with eigenvector $\mathbf{u} \neq 0$ iff $U\mathbf{u} = \lambda\mathbf{u}$ and $U\bar{\mathbf{u}} = \bar{\lambda}\bar{\mathbf{u}}$.
- Recall that U unitary implies $|\lambda| = 1$.
 - Proof^[1]: $\|U\mathbf{x}\| = \|\mathbf{x}\|$ and $U\mathbf{x} = \lambda\mathbf{x}$. Thus,

$$\|U\mathbf{x}\| = \|\lambda\mathbf{x}\| = |\lambda|\|\mathbf{x}\| = \|\mathbf{x}\|$$

and since $\mathbf{x} \neq 0$, we can divide by $\|\mathbf{x}\|$, so $|\lambda| = 1$.

- Let $\mathbf{u} = \text{Re } \mathbf{u} + i \text{Im } \mathbf{u}$.
- It follows that we may define

$$\mathbf{x} = \text{Re } \mathbf{u} = \frac{\mathbf{u} + \bar{\mathbf{u}}}{2} \qquad \mathbf{y} = \text{Im } \mathbf{u} = \frac{\mathbf{u} - \bar{\mathbf{u}}}{2}$$

- Thus, $\mathbf{u} = \mathbf{x} + i\mathbf{y}$ and $\bar{\mathbf{u}} = \mathbf{x} - i\mathbf{y}$.
- Since $U\mathbf{x} = \frac{U\mathbf{u} + U\bar{\mathbf{u}}}{2} = \frac{\lambda\mathbf{u} + \bar{\lambda}\bar{\mathbf{u}}}{2}$, $U\mathbf{y} = \text{Im}(\lambda\mathbf{u}) = \text{Re}(\lambda\mathbf{u})$.
- Since $|\lambda| = 1$, $\lambda = e^{i\alpha}$ and $\bar{\lambda} = e^{-i\alpha}$.
- It follows that $U\mathbf{x} = (\cos \alpha)\mathbf{x} - (\sin \alpha)\mathbf{y}$ and $U\mathbf{y} = (\cos \alpha)\mathbf{y} + (\sin \alpha)\mathbf{x}$.
- Thus, since $U\mathbf{x} = \text{Re } \lambda\mathbf{u}$, we have that

$$\begin{aligned} \lambda\mathbf{u} &= (\cos \alpha + i \sin \alpha)(\mathbf{x} + i\mathbf{y}) \\ &= (\cos \alpha)\mathbf{x} - (\sin \alpha)\mathbf{y} + i[(\cos \alpha)\mathbf{y} + (\sin \alpha)\mathbf{x}] \end{aligned}$$

- If E_λ is a 2 dimensional space spanned by \mathbf{x} and \mathbf{y} and invariant by U . Thus, any block of the desired matrix leaves its desired sub-block invariant.
- We also know that the eigenvectors of a unitary matrix corresponding to different eigenvalues are orthogonal.
- Thus, $\|\mathbf{x}\| = \|\mathbf{y}\| = \sqrt{2}/2\|\mathbf{u}\|$, $\mathbf{x} \perp \mathbf{y}$.
- Let \mathbf{x}, \mathbf{y} complete the theorem to form a basis of \mathbb{R}^n .
- It will follow that

$$U = \begin{pmatrix} R_\alpha & \mathbf{0} \\ \mathbf{0} & U_1 \end{pmatrix}$$

where U_1 is orthogonal, and we may repeat the process.

6.2 Chapter 6: Structure of Operators on Inner Product Spaces

From Treil (2017).

10/24:

- Theorem 6.1.1: Let $A : X \rightarrow X$ be an operator acting in a complex inner product space. Then there exists an orthonormal basis $\mathbf{u}_1, \dots, \mathbf{u}_n$ of X such that the matrix of A in this basis is upper triangular. In other words, any $n \times n$ matrix A can be represented as $A = UTU^*$, where U is unitary and T is upper-triangular.
- Theorem 6.1.2: Let $A : X \rightarrow X$ be an operator acting on a real inner product space. Suppose that all eigenvalues of A are real. Then there exists an orthonormal basis $\mathbf{u}_1, \dots, \mathbf{u}_n$ in X such that the matrix of A in this basis is upper triangular. In other words, any real $n \times n$ matrix A with all real eigenvalues can be represented as $T = UTU^* = UTU^T$, where U is orthogonal and T is a real upper-triangular matrix.

¹This would be a good exam question.

- Theorem 6.2.1: Let $A = A^*$ be a self-adjoint operator in an inner product space X (the space can be complex or real). Then all eigenvalues of A are real and there exists an orthonormal basis of eigenvectors of A in X .

Equivalently (see Theorem 6.2.2), A can be represented as $A = UDU^*$ where U is a unitary matrix and D is a diagonal matrix with real entries. Moreover, if A is real, U can be chosen to be real, i.e., orthogonal.

- Proposition 6.2.3: Let $A = A^*$ be a self-adjoint operator and let $\lambda, \mathbf{u}, \mu, \mathbf{v}$ be such that $A\mathbf{u} = \lambda\mathbf{u}$ and $A\mathbf{v} = \mu\mathbf{v}$. Then if $\lambda \neq \mu$, $\mathbf{u} \perp \mathbf{v}$.
- Since complex multiplication is commutative,

$$D^*D = DD^*$$

for every diagonal matrix D .

– It follows that $A^*A = AA^*$ if the matrix of A in some orthonormal basis is diagonal.

- Theorem 6.2.4: Any normal operator N in a complex vector space has an orthonormal basis of eigenvectors.

Equivalently, any matrix N satisfying $N^*N = NN^*$ can be represented as $N = UDU^*$ where U is unitary and D is diagonal.

- Proposition 6.2.5: An operator $N : X \rightarrow X$ is normal iff

$$\|N\mathbf{x}\| = \|N^*\mathbf{x}\|$$

for all $\mathbf{x} \in X$.

- **Hermitian square** (of A): The matrix A^*A .
- **Modulus** (of A): The unique positive semidefinite square root $\sqrt{A^*A}$.
- Proposition 6.3.3: For a linear operator $A : X \rightarrow Y$,

$$\| |A|\mathbf{x} \| = \|A\mathbf{x}\|$$

- Corollary 6.3.4: $\ker A = \ker |A|$.
- Theorem 6.3.5: Let $A : X \rightarrow X$ be an operator (square matrix). Then A can be represented as

$$A = U|A|$$

where U is a unitary operator.

- **Singular value** (of A): An eigenvalue of $|A|$.
 - A positive square root of an operator of A^*A .
- Proposition 6.3.6: Let $\sigma_1, \dots, \sigma_n$ be the singular values of A , ordered such that $\sigma_1, \dots, \sigma_r$ are the nonzero singular values, and let $\mathbf{v}_1, \dots, \mathbf{v}_n$ be an orthonormal basis of eigenvectors of A^*A . Then the system

$$\mathbf{w}_k = \frac{1}{\sigma_k} A\mathbf{v}_k$$

for $k = 1, \dots, r$ is orthonormal.

- **Schmidt decomposition** (of A): The decompositions

$$A = \sum_{k=1}^r \sigma_k \mathbf{w}_k \mathbf{v}_k^*$$

and

$$A\mathbf{x} = \sum_{k=1}^r \sigma_k (\mathbf{x}, \mathbf{v}_k) \mathbf{w}_k$$

- Note that these can be verified by plugging $\mathbf{x} = \mathbf{v}_j$ for each $j = 1, \dots, n$ into the latter equation.

10/25:

- Lemma 6.3.7: A can be represented as the Schmidt decomposition

$$A = \sum_{k=1}^r \sigma_k \mathbf{w}_k \mathbf{v}_k^*$$

where $\sigma_k > 0$ for any orthonormal systems $\mathbf{v}_1, \dots, \mathbf{v}_r$ and $\mathbf{w}_1, \dots, \mathbf{w}_r$.

- Corollary 6.3.8: Let $A = \sum_{k=1}^r \sigma_k \mathbf{w}_k \mathbf{v}_k^*$ be a Schmidt decomposition of A . Then

$$A^* = \sum_{k=1}^r \sigma_k \mathbf{v}_k \mathbf{w}_k^*$$

is a Schmidt decomposition of A^* .

- **Reduced singular value decomposition** (of A): The decomposition

$$A = \tilde{W} \tilde{\Sigma} \tilde{V}^*$$

where $A : \mathbb{F}^n \rightarrow \mathbb{F}^m$ has the Schmidt decomposition $A = \sum_{k=1}^r \sigma_k \mathbf{w}_k \mathbf{v}_k^*$, $\tilde{\Sigma} = \text{diag}\{\sigma_1, \dots, \sigma_r\}$, and \tilde{V}, \tilde{W} are matrices with columns $\mathbf{v}_1, \dots, \mathbf{v}_r$ and $\mathbf{w}_1, \dots, \mathbf{w}_r$, respectively. *Also known as **compact singular value decomposition**.*

- Note that \tilde{V} is an $n \times r$ matrix, $\tilde{\Sigma}$ is an $r \times r$ matrix, and \tilde{W} is an $m \times r$ matrix.
- Since $\mathbf{v}_1, \dots, \mathbf{v}_r$ and $\mathbf{w}_1, \dots, \mathbf{w}_r$ are orthonormal, \tilde{V}, \tilde{W} are isometries.
- Note that $r = \text{rank } A$ (see Problem 6.3.1).
 - It follows that if A is invertible, then $m = n = r$, so \tilde{V}, \tilde{W} are unitary and $\tilde{\Sigma}$ is an invertible diagonal matrix.
- However, A need not be invertible for us to get a representation similar to $A = \tilde{W} \tilde{\Sigma} \tilde{V}^*$.
 - Complete $\mathbf{v}_1, \dots, \mathbf{v}_r$ and $\mathbf{w}_1, \dots, \mathbf{w}_r$ to bases of \mathbb{F}^n and \mathbb{F}^m , respectively.
 - Then we get the following.
- **Singular value decomposition** (of A): The decomposition

$$A = W \Sigma V^*$$

where $V \in M_{n \times n}^{\mathbb{F}}$ and $W \in M_{m \times m}^{\mathbb{F}}$ are unitary matrices with columns $\mathbf{v}_1, \dots, \mathbf{v}_n$ and $\mathbf{w}_1, \dots, \mathbf{w}_m$, respectively, and $\Sigma \in M_{m \times n}^{\mathbb{R}^+}$ is a “diagonal” matrix such that

$$\Sigma_{j,k} = \begin{cases} \sigma_k & j = k \leq r \\ 0 & \text{otherwise} \end{cases}$$

- Notice that if $A = W\Sigma V^*$, then

$$A^*A = (W\Sigma V^*)^*(W\Sigma V^*) = V\Sigma^*W^*W\Sigma V^* = V\Sigma^2V^*$$

proving that the singular values of A , squared, are the eigenvalues of A^*A .

- If A is invertible, the reduced SVD is the matrix form of the Schmidt decomposition is the SVD.
- If $A = W\Sigma V^*$ is $n \times n$, then

$$A = \underbrace{(WV^*)}_U \underbrace{(V\Sigma V^*)}_{|A|}$$

is a polar decomposition of A .

- Consider the unit ball $B = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq 1\}$.
 - We want to describe $A(B)$, i.e., the image of the unit ball under A .
 - Let $\mathbf{x} = (x_1, \dots, x_n)^T$ and let $\mathbf{y} = (y_1, \dots, y_n)^T$. If $A = \text{diag}\{\sigma_1, \dots, \sigma_n\}$, we have $\mathbf{y} \in A(B)$ iff $\mathbf{y} = A\mathbf{x}$ where $\mathbf{x} \in B$ iff

$$\sum_{k=1}^n \frac{y_k^2}{\sigma_k^2} = \sum_{k=1}^n x_k^2 = \|\mathbf{x}\|^2 \leq 1$$

- Thus, $A(B)$ is an ellipsoid with half-axes $\sigma_1, \dots, \sigma_n$.
- In the more general case, if $A = W\Sigma V^*$, then since V^* is unitary, $V^*(B) = B$. $\Sigma V^*(B) = \Sigma(B)$ is thus by the above an ellipsoid in range Σ with half-axes $\sigma_1, \dots, \sigma_r$. Thus, since isometries don't change geometry, $W(\Sigma(B))$ is also an ellipsoid with the same half-axes, but in range A .
- Conclusion: The image $A(B)$ of the closed unit ball B is an ellipsoid in range A with half-axes $\sigma_1, \dots, \sigma_r$, where r is the number of nonzero singular values, i.e., the rank of A .
- Finding the maximum of $\|A\mathbf{x}\|$ for $\mathbf{x} \in B$.
 - For a diagonal matrix Σ with nonnegative entries, the maximum is clearly the maximal diagonal entry: In this case if s_1 is the maximal diagonal entry, then since

$$\Sigma\mathbf{x} = \sum_{k=1}^r s_k x_k \mathbf{e}_k$$

we have that

$$\|A\mathbf{x}\|^2 = \sum_{k=1}^r s_k^2 |x_k|^2 \leq s_1^2 \sum_{k=1}^r |x_k|^2 = s_1^2 \cdot \|\mathbf{x}\|^2$$

- We get the following by a similar logic to before.
- Conclusion: The maximum of $\|A\mathbf{x}\|$ on the unit ball B is the maximal singular value of A .
- **Operator norm** (of A): The following quantity. Denoted by $\|A\|$. Given by

$$\|A\| = \max\{\|A\mathbf{x}\| : \mathbf{x} \in X, \|\mathbf{x}\| \leq 1\}$$

- $\|A\|$ clearly satisfies the four properties of a norm.
- Additionally,

$$\|A\mathbf{x}\| \leq \|A\| \cdot \|\mathbf{x}\|$$

- Alternate definition: The operator norm $\|A\|$ is the smallest number $C \geq 0$ such that $\|A\mathbf{x}\| \leq C\|\mathbf{x}\|$.

- **Frobenius norm:** The following norm. *Also known as Hilbert-Schmidt norm.* Denoted by $\|A\|_2$. Given by

$$\|A\|_2^2 = \text{tr}(A^*A)$$

- If we let s_1, \dots, s_n be the singular values of A and let s_1 be the largest value, then we have

$$\|A\|^2 = s_1^2 \leq \sum_{k=1}^n s_k^2 = \text{tr}(A^*A) = \|A\|_2^2$$

- Conclusion: The operator norm of a matrix cannot be more than its Frobenius norm.
- Suppose we want to solve $A\mathbf{x} = \mathbf{b}$ where A is invertible, but there is some (experimental) error $\Delta\mathbf{b}$ in \mathbf{b} . Then we are really solving for an approximate solution $\mathbf{x} + \Delta\mathbf{x}$ to the equation

$$A(\mathbf{x} + \Delta\mathbf{x}) = \mathbf{b} + \Delta\mathbf{b}$$

- It follows since A is invertible that $\mathbf{x} = A^{-1}\mathbf{b}$ and $\Delta\mathbf{x} = A^{-1}\Delta\mathbf{b}$.
- To estimate the relative error $\|\Delta\mathbf{x}\|/\|\mathbf{x}\|$ in the solution in comparison with the relative error $\|\Delta\mathbf{b}\|/\|\mathbf{b}\|$ in the data, use

$$\frac{\|\Delta\mathbf{x}\|}{\|\mathbf{x}\|} = \frac{\|A^{-1}\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \leq \frac{\|A^{-1}\| \cdot \|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \frac{\|A\| \cdot \|\mathbf{x}\|}{\|\mathbf{x}\|} = \|A^{-1}\| \cdot \|A\| \cdot \frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}$$

- **Condition number** (of A): The following quantity. *Given by*

$$\|A\| \cdot \|A^{-1}\|$$

- If s_1 is the largest singular value of A and s_n is the smallest, then

$$\|A\| \cdot \|A^{-1}\| = s_1 \cdot \frac{1}{s_n} = \frac{s_1}{s_n}$$

- **Well-conditioned** (matrix): A matrix the condition number of which is not “too big.”
- **Ill-conditioned** (matrix): A matrix that is not well-conditioned.
- Theorem 6.5.1: Let U be an orthogonal operator on \mathbb{R}^n and let $\det U = 1$. Then there exists an orthonormal basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ such that the matrix of U in this basis has the block diagonal form

$$\begin{pmatrix} R_{\varphi_1} & & & 0 \\ & \ddots & & \\ & & R_{\varphi_k} & \\ 0 & & & I_{n-2k} \end{pmatrix}$$

where each R_{φ_j} is a two-dimensional rotation

$$R_{\varphi_j} = \begin{pmatrix} \cos \varphi_j & -\sin \varphi_j \\ \sin \varphi_j & \cos \varphi_j \end{pmatrix}$$

and I_{n-2k} represents the $(n-2k) \times (n-2k)$ identity matrix.

- Alternate interpretation: Any rotation in \mathbb{R}^n can be represented as a composition of at most $n/2$ commuting planar rotations.

- Theorem 6.5.2: Let U be an orthogonal operator on \mathbb{R}^n and let $\det U = -1$. Then there exists an orthonormal basis $\mathbf{v}_1, \dots, \mathbf{v}_n$ such that the matrix of U in this basis has block diagonal form

$$\begin{pmatrix} R_{\varphi_1} & & & & 0 \\ & \ddots & & & \\ & & R_{\varphi_k} & & \\ & & & I_r & \\ 0 & & & & -1 \end{pmatrix}$$

where $r = n - 2k - 1$ and each R_{φ_j} is a two-dimensional rotation

$$R_{\varphi_j} = \begin{pmatrix} \cos \varphi_j & -\sin \varphi_j \\ \sin \varphi_j & \cos \varphi_j \end{pmatrix}$$

- Corollary: An orthogonal 2×2 matrix U with determinant -1 is always a reflection.
- Theorem 6.5.3: Any rotation U (i.e., any orthogonal transformation U with $\det U = 1$) can be represented as a product of at most $n(n-1)/2$ elementary rotations.
- Consider the following orthonormal bases of \mathbb{R}^2 .



Figure 6.1: Orientation in \mathbb{R}^2 .

- Notice that a rotation will get you from the standard basis (a) to basis (b), but not from the standard basis (a) to basis (c).
- This is the motivation for defining orientation.
- More formally, we know that there is a unique linear transformation U such that $U\mathbf{e}_k = \mathbf{v}_k$ for each $k = 1, 2$. In particular, the matrix of U with respect to the standard basis is orthogonal with columns $\mathbf{v}_1, \mathbf{v}_2$.
- By Theorems 6.5.1 and 6.5.2, if $\det U = 1$, then U is a rotation, and if $\det U = -1$, then U is not a rotation.
- **Similarly oriented** (bases \mathcal{A}, \mathcal{B}): Two bases \mathcal{A}, \mathcal{B} of a real vector space such that the change of coordinates matrix $[I]_{\mathcal{B}\mathcal{A}}$ has a positive determinant.
- **Differently oriented** (bases \mathcal{A}, \mathcal{B}): Two bases \mathcal{A}, \mathcal{B} of a real vector space that are not similarly oriented (i.e., $[I]_{\mathcal{B}\mathcal{A}}$ has a negative determinant).
- We usually let the standard basis of \mathbb{R}^n have a **positive orientation**.
 - In an abstract vector space, we need only fix a basis and declare its orientation to be positive.
- **Continuously transformable** (bases \mathcal{A}, \mathcal{B}): Two bases \mathcal{A}, \mathcal{B} such that $\mathcal{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ can be continuously transformed to a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$. In particular, there exists a **continuous family of bases** $\mathcal{V}(t) = \{\mathbf{v}_1(t), \dots, \mathbf{v}_n(t)\}$, $t \in [a, b]$, such that

$$\mathbf{v}_k(a) = \mathbf{a}_k \qquad \mathbf{v}_k(b) = \mathbf{b}_k$$

for each $k = 1, \dots, n$.

- **Continuous family of bases:** A family of bases $\mathcal{V}(t) = \{\mathbf{v}_1(t), \dots, \mathbf{v}_n(t)\}$, $t \in [a, b]$, such that the vector-functions $\mathbf{v}_k(t)$ are continuous (their coordinates in some bases are continuous functions) and the system $\mathbf{v}_1(t), \dots, \mathbf{v}_n(t)$ is a basis for all $t \in [a, b]$.
- Theorem 6.6.1: Two bases $\mathcal{A} = \{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ and $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ have the same orientation if and only if one of the bases can be continuously transformed to the other.

Chapter 7

Bilinear and Quadratic Forms

7.1 Notes

10/18: • **Bilinear form:** A function $L : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ such that

$$L(\alpha \mathbf{x}_1 + \beta \mathbf{x}_2, \mathbf{y}) = \alpha L(\mathbf{x}_1, \mathbf{y}) + \beta L(\mathbf{x}_2, \mathbf{y}) \quad L(\mathbf{x}, \alpha \mathbf{y}_1 + \beta \mathbf{y}_2) = \alpha L(\mathbf{x}, \mathbf{y}_1) + \beta L(\mathbf{x}, \mathbf{y}_2)$$

$$- L(\mathbf{x}, \mathbf{y}) = (A\mathbf{x}, \mathbf{y}).$$

• **Quadratic form:** A bilinear form $L(\mathbf{x}, \mathbf{x})$.

$$- (\mathbf{x}, \mathbf{x}) \text{ is a polynomial of degree 2 in } \mathbf{x}_1, \dots, \mathbf{x}_n:$$

$$L(\lambda \mathbf{x}, \lambda \mathbf{x}) = (\lambda \mathbf{x}, \lambda \mathbf{x}) = \lambda^2 (\mathbf{x}, \mathbf{x})$$

• We have that

$$(A\mathbf{x}, \mathbf{x}) = (A\lambda \mathbf{x}, \lambda \mathbf{x}) = \lambda^2 (A\mathbf{x}, \mathbf{x}) = \sum_{j,i=1}^n \alpha_{j,i} \mathbf{x}_i \mathbf{x}_j$$

• The general form of a quadratic form:

$$- \text{Can any quadratic form on } \mathbb{R}^n \text{ be written as } (A\mathbf{x}, \mathbf{x})?$$

10/20: • Bilinear forms are linear in each argument when keeping the other fixed.

• Quadratic forms $Q(\mathbf{x}) = L(\mathbf{x}, \mathbf{x})$ are quadratic polynomials in the coordinates of x .

$$- \text{In particular, } Q(\lambda \mathbf{x}) = |\lambda|^2 Q(\mathbf{x}).$$

• If Q quadratic is real, then $Q(\mathbf{x}) = (A\mathbf{x}, \mathbf{x})$ where A is some square matrix.

$$- \text{If } \mathbf{e}_1, \dots, \mathbf{e}_n \text{ is an orthonormal basis of } \mathbb{R}^n, \text{ then there exists a unique } A = A^* \text{ such that } (A)_{ij} = L(\mathbf{e}_i, \mathbf{e}_j).$$

$$- \text{Keeping } \mathbf{x} = \sum_{i=1}^n \mathbf{x}_i \mathbf{e}_i \text{ fixed, we have}$$

$$\begin{aligned} Q(\mathbf{x}) &= L(\mathbf{x}, \mathbf{x}) \\ &= L\left(\sum_{i=1}^n \mathbf{x}_i \mathbf{e}_i, \sum_{j=1}^n \mathbf{x}_j \mathbf{e}_j\right) \\ &= \sum_{i=1}^n \mathbf{x}_i L\left(\mathbf{e}_i, \sum_{j=1}^n \mathbf{x}_j \mathbf{e}_j\right) \\ &= \sum_{i,j=1}^n \mathbf{x}_i \mathbf{x}_j \underbrace{L(\mathbf{e}_i, \mathbf{e}_j)}_{A_{ij}} \end{aligned}$$

- We have that

$$\begin{aligned}(A\mathbf{x}, \mathbf{x}) &= (UDU^{-1}\mathbf{x}, \mathbf{x}) \\ &= (DU^{-1}\mathbf{x}, U^{-1}\mathbf{x}) \\ &= \sum_{i=1}^n \lambda_i \underbrace{(U^{-1}\mathbf{x})_i}_{\mathbf{y}_i} \underbrace{(U^{-1}\mathbf{x})_i}_{\mathbf{y}_i}\end{aligned}$$

- Can we characterize the set $\{\mathbf{x} : (A\mathbf{x}, \mathbf{x}) = 1\}$?
 - Note that this set is equivalent to $\{\mathbf{y} : (D\mathbf{y}, \mathbf{y}) = 1\}$ by the above. This set is a rotation of the previous one. Ellipse?
- Positive quadratic form:
 - Q is positive definite if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$ and Q is positive semidefinite if $Q(\mathbf{x}) \geq 0$ for all $\mathbf{x} \neq \mathbf{0}$.
 - Take a self-adjoint matrix $A = A^*$. It is positive definite if $Q(\mathbf{x}) = (A\mathbf{x}, \mathbf{x})$ is positive definite.
- Theorem: If $A = A^*$, then
 1. A is positive definite if and only if all eigenvalues of A are positive.
 2. A is positive semidefinite if and only if all eigenvalues of A are nonnegative.
 3. A is negative semidefinite if and only if all eigenvalues of A are nonpositive.
 4. A is negative definite if and only if all eigenvalues of A are negative.
 5. A is indefinite if and only if the eigenvalues of A have positive and negative values.
- Theorem: $A = A^*$ is positive definite iff $\det A_k > 0$ for all $k = 1, \dots, n$ where A_k is the upper left $k \times k$ submatrix.
- Minimax representation of eigenvalues of a self-adjoint A .
 - Let E be a subspace of X where $\dim X < \infty$. We define $\text{codim}(E) = \dim E^\perp$.
 - Thus, $\dim E + \text{codim } E = \dim X$.
 - Theorem: Let $A = A^*$, $\lambda_1 \geq \dots \geq \lambda_n$ eigenvalues of A . Then

$$\lambda_k = \max_{\substack{E \text{ subspace} \\ \dim E = k}} \min_{\substack{\mathbf{x} \in E \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}) = \min_{\substack{F \text{ subspace} \\ \text{codim } F = k-1}} \max_{\substack{\mathbf{x} \in F \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x})$$

- Proof: A diagonal equals $(\lambda_1, \dots, \lambda_n)$.
- An orthonormal basis of X such that $\dim E = k$, $\text{codim } F = k-1$, $\dim F = n-k+1$.
- There exists an $\mathbf{x}_0 \neq \mathbf{0}$ such that $\mathbf{x}_0 \in E \cap F$.
- Note that if $B = B^*$, then the max and min of $(B\mathbf{x}, \mathbf{x})$ over the unit sphere is the maximal and minimal eigenvalue of B .
- Thus,

$$\min_{\substack{\mathbf{x} \in E \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}) \leq (A\mathbf{x}_0, \mathbf{x}_0) \leq \max_{\substack{\mathbf{x} \in F \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x})$$

- This is true for any E, F subspaces. $\dim E = k$, $\text{codim } F = k-1$, $E_0 = \text{span}(\mathbf{e}_1, \dots, \mathbf{e}_k)$ and $F_0 = \text{span}(\mathbf{e}_k, \dots, \mathbf{e}_n)$.
- Thus,

$$\min_{\substack{E_0 \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x}) = \lambda_k = \max_{\substack{F_0 \\ \|\mathbf{x}\|=1}} (A\mathbf{x}, \mathbf{x})$$

■ Additionally,

$$\lambda_{k_1} \leq \max_{\dim E=k} \min_{\mathbf{x}} (A\mathbf{x}, \mathbf{x}) \leq \min_{\operatorname{codim} F=k-1} \max_{\mathbf{x}} (A\mathbf{x}, \mathbf{x}) \leq \lambda_k$$

- Corollary: Let $A = A^* = (a_{jk})_{1 \leq j, k \leq n}$ with eigenvalues $\lambda_1, \dots, \lambda_n$ listed in decreasing order. Let $\tilde{A} = (a_{j,k})_{1 \leq j, k \leq n-1}$ with eigenvalues μ_1, \dots, μ_{n-1} listed in decreasing order. Then $\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \dots \geq \mu_{n-1} \geq \lambda_n$.

– Consider $(A\mathbf{x}, \mathbf{x})$ on $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$, but then restrict yourself to $\mathbf{x} \in \mathbb{R}^{n-1}$ on $\{\mathbf{e}_1, \dots, \mathbf{e}_{n-1}\}$.

7.2 Chapter 7: Bilinear and Quadratic Forms

From Treil (2017).

10/25: • **Bilinear form** (on \mathbb{R}^n): A function $L(\mathbf{x}, \mathbf{y})$ of two arguments $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ that is linear in each argument.

– Linearity in each argument:

$$L(\alpha\mathbf{x}_1 + \beta\mathbf{x}_2, \mathbf{y}) = \alpha L(\mathbf{x}_1, \mathbf{y}) + \beta L(\mathbf{x}_2, \mathbf{y}) \quad L(\mathbf{x}, \alpha\mathbf{y}_1 + \beta\mathbf{y}_2) = \alpha L(\mathbf{x}, \mathbf{y}_1) + \beta L(\mathbf{x}, \mathbf{y}_2)$$

- If $\mathbf{x} = (x_1, \dots, x_n)^T$ and $\mathbf{y} = (y_1, \dots, y_n)^T$, then

$$\begin{aligned} L(\mathbf{x}, \mathbf{y}) &= \sum_{j,k=1}^n a_{j,k} x_k y_j \\ &= (A\mathbf{x}, \mathbf{y}) \\ &= \mathbf{y}^T A\mathbf{x} \end{aligned}$$

where

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix}$$

– A is uniquely determined by L .

- **Quadratic form** (on \mathbb{R}^n): The diagonal of a bilinear form L , i.e., a bilinear form $Q[\mathbf{x}] = L(\mathbf{x}, \mathbf{x}) = (A\mathbf{x}, \mathbf{x})$.

– Alternatively: A homogeneous polynomial of degree 2, i.e., a polynomial in x_1, \dots, x_n with only ax_k^2 and cx_jx_k terms.

- There are infinitely many ways to write a quadratic form as $(A\mathbf{x}, \mathbf{x})$.

– However, there is a unique representation $(A\mathbf{x}, \mathbf{x})$ where A is a (real) symmetric matrix.

- **Quadratic form** (on \mathbb{C}^n): A function of the form $Q[\mathbf{x}] = (A\mathbf{x}, \mathbf{x})$ where A is self-adjoint.

- Lemma 7.1.1: Let $(A\mathbf{x}, \mathbf{x})$ be real for all $\mathbf{x} \in \mathbb{C}^n$. Then $A = A^*$.

- To classify quadratic forms, consider the set of points $\mathbf{x} \in \mathbb{R}^n$ defined by $Q[\mathbf{x}] = 1$ for some quadratic form Q .

– If the matrix of Q is diagonal, i.e., $Q[\mathbf{x}] = a_1x_1^2 + \dots + a_nx_n^2$, then the set of points can easily be visualized.

- The standard method of diagonalizing a quadratic form is change of variables.
- Orthogonal diagonalization.

- Let $Q[\mathbf{x}] = (A\mathbf{x}, \mathbf{x})$ in \mathbb{F}^n .
- Suppose $\mathbf{y} = S^{-1}\mathbf{x}$ where S is an invertible $n \times n$ matrix. Then

$$Q[\mathbf{x}] = Q[S\mathbf{y}] = (AS\mathbf{y}, S\mathbf{y}) = (S^*AS\mathbf{y}, \mathbf{y})$$

so in the new variables \mathbf{y} , the quadratic form has matrix S^*AS .

- Thus, we can let $A = UDU^*$, choose $D = U^*AU$ as our new (diagonal) matrix, and let this matrix act on the variables $\mathbf{y} = U^*\mathbf{x}$.

- Non-orthogonal diagonalization:

- Completing the square:
 - Eliminate all $x_i x_j$ terms by completing the square. Then substitute in a y_k for each squared term.
- Row/column operations:
 - Augment $(A|I)$. Row reduce A to D . Then $I \rightarrow S^*$.

10/28:

- **Sylvester's Law of Inertia:** For a Hermitian matrix A (i.e., for a quadratic form $Q[\mathbf{x}] = (A\mathbf{x}, \mathbf{x})$) and any of its diagonalizations $D = S^*AS$, the number of positive, negative, and zero diagonal entries of D depends only on A , but not on a particular choice of diagonalization.
- **Positive** (subspace $E \subset \mathbb{F}^n$ corresponding to A): A subspace E such that $(A\mathbf{x}, \mathbf{x}) > 0$ for all nonzero $\mathbf{x} \in E$. Also known as **A-positive**.
- **Negative** (subspace $E \subset \mathbb{F}^n$ corresponding to A): A subspace E such that $(A\mathbf{x}, \mathbf{x}) < 0$ for all nonzero $\mathbf{x} \in E$. Also known as **A-negative**.
- **Neutral** (subspace $E \subset \mathbb{F}^n$ corresponding to A): A subspace E such that $(A\mathbf{x}, \mathbf{x}) = 0$ for all nonzero $\mathbf{x} \in E$. Also known as **A-neutral**.
- Theorem 7.3.1: Let A be an $n \times n$ Hermitian matrix, and let $D = S^*AS$ be its diagonalization by an invertible matrix S . Then the number of positive (resp. negative) diagonal entries of D coincides with the maximal dimension of an A -positive (resp. A -negative) subspace.
- Lemma 7.3.2: Let $D = \text{diag}\{\lambda_1, \dots, \lambda_n\}$. Then the number of positive (resp. negative) diagonal entries of D coincides with the maximal dimension of a D -positive (resp. D -negative) subspace.
- **Positive definite** (quadratic form Q): A quadratic form Q such that $Q[\mathbf{x}] > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- **Positive semidefinite** (quadratic form Q): A quadratic form Q such that $Q[\mathbf{x}] \geq 0$ for all \mathbf{x} .
- **Negative definite** (quadratic form Q): A quadratic form Q such that $Q[\mathbf{x}] < 0$ for all $\mathbf{x} \neq \mathbf{0}$.
- **Negative semidefinite** (quadratic form Q): A quadratic form Q such that $Q[\mathbf{x}] \leq 0$ for all \mathbf{x} .
- **Indefinite** (quadratic form Q): A quadratic form Q for which there exist $\mathbf{x}_1, \mathbf{x}_2$ such that $Q[\mathbf{x}_1] > 0$ and $Q[\mathbf{x}_2] < 0$.
- **Positive definite** (Hermitian matrix A): A matrix A for which the corresponding quadratic form $Q[\mathbf{x}] = (A\mathbf{x}, \mathbf{x})$ is positive definite.
 - Positive semidefinite, negative definite, negative semidefinite, and indefinite Hermitian matrices are defined similarly.
- Theorem 7.4.1: Let $A = A^*$. Then
 1. A is positive definite iff all eigenvalues of A are positive.
 2. A is positive semidefinite iff all eigenvalues of A are non-negative.

3. A is negative definite iff all eigenvalues of A are negative.
 4. A is negative semidefinite iff all eigenvalues of A are non-positive.
 5. A is indefinite iff it has both positive and negative eigenvalues.
- **Upper left submatrix** (of A): A $k \times k$ matrix A_k composed of all entries of A from row (column) 1 through k in the same arrangement.
 - Theorem 7.4.2 (Sylvester's Criterion of Positivity): A matrix $A = A^*$ is positive definite if and only if $\det A_k > 0$ for all $k = 1, \dots, n$.
 - To check if a matrix A is negative definite, check that the matrix $-A$ is positive definite.
 - Theorem 7.4.3 (Minimax characterization of eigenvalues): Let $A = A^*$ be an $n \times n$ matrix and let $\lambda_1 \geq \dots \geq \lambda_n$ be its eigenvalues taken in decreasing order. Then

$$\lambda_k = \max_{E: \dim E = k} \min_{\mathbf{x} \in E: \|\mathbf{x}\|=1} (A\mathbf{x}, \mathbf{x}) = \min_{F: \text{codim } F = k-1} \max_{\mathbf{x} \in F: \|\mathbf{x}\|=1} (A\mathbf{x}, \mathbf{x})$$

- Corollary 7.4.4 (Intertwining of eigenvalues): Let $A = A^* = \{a_{j,k}\}_{j,k=1}^n$ be a self-adjoint matrix and let $\tilde{A} = \{a_{j,k}\}_{j,k=1}^{n-1}$ be its submatrix of size $(n-1) \times (n-1)$. Let $\lambda_1, \dots, \lambda_n$ and μ_1, \dots, μ_{n-1} be the eigenvalues of A and \tilde{A} respectively, taken in decreasing order. Then

$$\lambda_1 \geq \mu_1 \geq \lambda_2 \geq \mu_2 \geq \dots \geq \lambda_{n-1} \geq \mu_{n-1} \geq \lambda_n$$

Chapter 8

Dual Spaces and Tensors

8.1 Notes

10/22: • **Functional:** A linear bounded map $L : H \rightarrow F$, where H is finite dimensional (equivalent to \mathbb{R}^n).

• **Dual space:** The set of bounded linear functionals on H . Denoted by H' , H^* .

• If $l \leq p < \infty$, then

$$l^p = \left\{ (a_n)_{n \in \mathbb{N}} : \sum_{n=1}^{\infty} |a_n|^p < \infty \right\}$$

• Back to finite dimensions, $H' \approx \mathbb{R}^n$.

• Let $\mathbf{a}_1, \dots, \mathbf{a}_n$ be a basis of H . Then $L\mathbf{x} = (L\mathbf{a}_1, \dots, L\mathbf{a}_n) \approx \mathbb{R}^n$.

• Let $L((a_n)_{n \in \mathbb{N}}) = \sum_{n=1}^{\infty} a_n b_n$. Then $L((a_n)_{n \in \mathbb{N}})$ will be bounded if and only if $(b_n)_{n \in \mathbb{N}} \in l^q$ where $1 < p < q$ where $\frac{1}{q} + \frac{1}{p} = 1$.

• **Young's inequality:** The statement

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q}$$

• We have $|\sum a_n b_n| \leq \|a_n\|_p \|b_n\|_p$.

• Conclusion:

$$\sum \frac{|a_n| |b_n|}{\|a_n\|_p \|b_n\|_q} = 1$$

• We can define H'' , too. This contains linear functionals on H' .

• We know that $L(x) = \langle x, L \rangle = x(L)$. $x \in H''$.

• Riesz representation theorem: Let H have an inner product. $L \in H'$ if and only if there exists a unique $y \in H$ such that $L(x) = (x, y)$.

– Gives us a way to identify all bounded linear functionals on H .

– In finite dimensions, $L(x)$, where $x = \sum_1^n \alpha_i a_i$ gives us $L(x) = \sum_1^n \alpha_i L(a_i)$.

8.2 Chapter 8: Dual Spaces and Tensors

10/28:

- Linear functionals are denoted by L .
 - L is given by a $1 \times n$ matrix denoted by $[L]$.
- The collection of all $[L]$ (the dual space) is isomorphic to \mathbb{R}^n via $[L] \mapsto [L]^T$.
 - However, the objects are different: Let $[I]_{\mathcal{B}\mathcal{A}}$ be the change of coordinates matrix in \mathbb{R}^n . We thus have that

$$[\mathbf{v}]_{\mathcal{B}} = [I]_{\mathcal{B}\mathcal{A}}[\mathbf{v}]_{\mathcal{A}}$$

but we also have that

$$[L]_{\mathcal{B}} = [L]_{\mathcal{A}}[I]_{\mathcal{A}\mathcal{B}}$$

so that

$$[L]_{\mathcal{B}}^T = ([L]_{\mathcal{A}}[I]_{\mathcal{A}\mathcal{B}})^T = [I]_{\mathcal{A}\mathcal{B}}^T [L]_{\mathcal{A}}^T$$

- Essentially, “if S is the change of coordinate matrix in $X \dots$ then the change of coordinate matrix in the dual space X' is $(S^{-1})^T$ ” (Treil, 2017, p. 219).
- Lemma 8.1.3: Let $\mathbf{v} \in V$. If $L(\mathbf{v}) = 0$ for all $L \in V'$, then $\mathbf{v} = \mathbf{0}$. As a corollary, if $L(\mathbf{v}_1) = L(\mathbf{v}_2)$ for all $L \in V'$, then $\mathbf{v}_1 = \mathbf{v}_2$.
- The second dual V'' is canonically (i.e., in a natural way) isomorphic to V .
- **Dual basis** (to $\mathbf{b}_1, \dots, \mathbf{b}_n \in V$): The system of vectors $\mathbf{b}'_1, \dots, \mathbf{b}'_n \in V'$ uniquely defined by the following equation. *Also known as biorthogonal basis.*

$$\mathbf{b}'_k(\mathbf{b}_j) = \delta_{kj}$$

- The k^{th} coordinate of a vector \mathbf{v} in a basis $\mathbf{b}_1, \dots, \mathbf{b}_n$ is $\mathbf{b}'_k(\mathbf{v})$.
 - This is a baby version of the **abstract non-orthogonal Fourier decomposition** of \mathbf{v} .
- Theorem 8.2.1 (Riesz representation theorem): Let H be an inner product space. Given a linear functional L on H , there exists a unique vector $\mathbf{y} \in H$ such that

$$L(\mathbf{v}) = (\mathbf{v}, \mathbf{y})$$

for all $\mathbf{v} \in H$.

- If V is a real inner product space, we can define an isomorphism from V to V' by $\mathbf{y} \mapsto L_{\mathbf{y}} = (\mathbf{v}, \mathbf{y})$.
 - If V is complex, this function is not linear since if α is complex,

$$L_{\alpha\mathbf{y}}(\mathbf{v}) = (\mathbf{v}, \alpha\mathbf{y}) = \bar{\alpha}(\mathbf{v}, \mathbf{y}) = \bar{\alpha}L_{\mathbf{y}}(\mathbf{v})$$

- It follows by such a mapping that $\mathbf{b}'_k = \mathbf{b}_k$ for each k .
- **Conjugate linear** (transformation): A transformation T such that

$$T(\alpha\mathbf{x} + \beta\mathbf{y}) = \bar{\alpha}T\mathbf{x} + \bar{\beta}T\mathbf{y}$$

- It is customary to write outputs of linear functionals $L(\mathbf{v})$ in the form $\langle \mathbf{v}, L \rangle$.
 - This expression is linear in both arguments, unlike the inner product.
- Defines the dual transformation as the unique transformation such that

$$\langle A\mathbf{x}, \mathbf{y}' \rangle = \langle \mathbf{x}, A'\mathbf{y} \rangle$$

for all $\mathbf{x} \in X, \mathbf{y}' \in Y'$.

- It's matrix in the standard bases equals A^T .
- Annihilators are denoted by E^\perp here.
- Proposition 8.3.6: The annihilator of the annihilator of E equals E .
- Let $A : X \rightarrow Y$ be an operator acting from one vector space to another. Then
 1. $\ker A' = (\text{range } A)^\perp$.
 2. $\ker A = (\text{range } A')^\perp$.
 3. $\text{range } A = (\ker A')^\perp$.
 4. $\text{range } A' = (\ker A)^\perp$.

Chapter 9

Advanced Spectral Theory

9.1 Notes

- 10/22:
- Let $p(z) = \sum_{i=0}^n a_i z^i$ be a polynomial. Let A be an $n \times n$ matrix. We let $p(A) = \sum_{i=0}^n a_i A^i$.
 - Theorem: If A is an $n \times n$ and $p(\lambda) = \det(A - \lambda I)$, then $p(A) = 0$.
 - We know that $p(\lambda) = a(z - \lambda_1) \cdots (z - \lambda_n)$ where $\lambda_1, \dots, \lambda_n$ are the eigenvalues.
 - Thus $p(A) = a(A - \lambda_1 I) \cdots (A - \lambda_n I)$.
 - If you are in \mathbb{R}^n and have this property, you can factorize your matrix.
 - Thus, $p(A)\mathbf{x} = \mathbf{0}$ since \mathbf{x} can be decomposed into a linear combination of eigenvectors of A , which will be taken to 0 one by one by the terms of $p(A)$.
 - $\sigma(B) = \{\text{eigenvalues of } B\}$ is known as the **spectrum** of B .
 - If p is an arbitrary polynomial and A is $n \times n$, then μ is an eigenvalue of $p(A)$ if and only if $\mu = p(\lambda)$ where λ is an eigenvalue of A . In essence, $\sigma(p(A)) = p(\sigma(A))$.
 - Chapter 9 will not be on the exam. We don't have to know the generalization to infinite dimensional spaces.
- 10/25:
- If A is an $n \times n$ square matrix and $p(\lambda) = \det(A - \lambda I)$, then $p(A) = 0$.
 - Proof: WLOG, let A be an upper triangular matrix with diagonal entries equal to the eigenvalues.
 - Think of $p(z) = (-1)^n (z - \lambda_1) \cdots (z - \lambda_n)$.
 - Thus, $p(A) = (-1)^n (A - \lambda_1 I) \cdots (A - \lambda_n I)$.
 - WTS: $p(A)\mathbf{x} = 0$ for all $\mathbf{x} \in V$.
 - Let $E_k = \text{span}(e_1, \dots, e_k)$ be the span of the first k eigenvectors of A , where e_1, \dots, e_n is a standard basis in \mathbb{C}^n .
 - A triangular implies $AE_k \subset E_k$. Thus, $(A - \lambda I)E_k \subset E_k$, so E_k is invariant under $A - \lambda I$ for all λ .
 - If we apply $A - \lambda_k I$ to a vector in E_k , we are left with a vector in E_{k-1} .
 - Thus, if we apply $\prod_{k=1}^n (A - \lambda_k I) = p(A)$ to any vector in $E_n = V$, we will kill it piece by piece down to zero.
 - Let A be a square $n \times n$ matrix. Then p an arbitrary polynomial implies $\sigma(p(A)) = p(\sigma(A))$. (Any eigenvalue μ of $p(A)$ is $\mu = p(\lambda)$, where λ is an eigenvalue of A).
 - Shows that polynomials of operators commute.

- Proof: Let λ be an eigenvalue of A . We want to show that $p(\lambda)$ is an eigenvalue of $p(A)$. This is obvious since $A\mathbf{x} = \lambda\mathbf{x}$ for some \mathbf{x} , so $A^k\mathbf{x} = \lambda^k\mathbf{x}$, so in particular, $p(A)\mathbf{x} = p(\lambda)\mathbf{x}$.
- On the other hand, if μ is an eigenvalue of $p(A)$, we want to show that there exists $\lambda \in \sigma(A)$ such that $\mu = p(\lambda)$.
- Consider $q(z) = p(z) - \mu$. Then $q(A) = p(A) - \mu I$. Since μ is an eigenvalue of $p(A)$, $q(A)$ is not invertible.
- Thus, $q(z) = (-1)^n(z - z_1) \cdots (z - z_n)$ and $q(A) = (-1)^n(A - z_1 I) \cdots (A - z_n I)$.
- But $q(A)$ is not invertible, so one of the $A - z_k I$ is not invertible. Take z_k such that $A - z_k I$ is not invertible. Then $z_k \in \sigma(A)$. It follows that $q(z_k) = p(z_k) - \mu = 0$.
- If A is $n \times n$, $\lambda_1, \dots, \lambda_n$ are its eigenvalues, p is a polynomial, then $p(A)$ is invertible if and only if $p(\lambda_k) \neq 0$ for each $k = 1, \dots, n$.
 - This is an immediate corollary to the previous result.
- We now build up to the **generalized eigenspace**, which is related to some “geometric” properties of the algebraic multiplicity of an eigenvalue.
- If $A : V \rightarrow V$ is a linear operator and $E \subset V$ is a subspace, E is A -invariant if $AE \subset E$.
- Facts:
 - If E is A -invariant, E is A^k -invariant.
 - Thus, E is $p(A)$ -invariant.
- Consider the restriction map $A|_E$.
- A has a block-diagonalized matrix where each block corresponds to the generalized eigenvectors of a generalized eigenvalue of A .
 - Let E_1, \dots, E_r be a **basis of invariant subspaces**.
 - Let $A_k = A|_{E_k}$. Then the A_k ’s act independently of each other.
- **Generalized eigenvector** (of A): A vector \mathbf{v} corresponding to an eigenvalue λ if there exists $k \geq 1$ such that $(A - \lambda I)^k \mathbf{v} = \mathbf{0}$.
- **Generalized eigenspace**: The set E_λ of all of the generalized eigenvectors of λ . *Given by*

$$E_\lambda = \bigcup_{k \geq 1} \ker(A - \lambda I)^k$$

- E_λ is a linear subspace of V .
- **Degree** (of λ): The smallest number k such that increasing k any more does not add further vectors to the generalized eigenspace. *Denoted by $d(\lambda)$. Also known as **depth**.*
 - Symbolically, $d(\lambda)$ is the smallest number such that

$$E_\lambda = \bigcup_{k=1}^{d(\lambda)} \ker(A - \lambda I)^k$$

- Start working through the first 25 problems of Rudin (1976) (his metric spaces problems).
- 10/27: • Jordan form.
- Reviews build up to generalized eigenvectors.

- Theorem: If $\sigma(A) = \{\lambda_1, \dots, \lambda_n\}$ and E_1, \dots, E_n are the corresponding generalized eigenspaces, then E_1, \dots, E_n is a basis of subspaces of U , i.e., $V = \oplus_k E_k$.
- Corollary: $A : V \rightarrow V$ can be represented as $A = D + N$ where D is diagonalizable and N is nilpotent and $ND = DN$.

– Proof: Consider the basis of generalized eigenspaces known to exist from the theorem. Then $A = \text{diag}\{A_1, \dots, A_r\}$.

– Let

$$N_k = A_k - \lambda_k I_{E_k}$$

This is nilpotent.

– Then let

$$D = \text{diag}\{\lambda_1 I_{E_1}, \dots, \lambda_n I_{E_n}\}$$

– These two matrices satisfy the necessary properties.

- Let $\dot{\mathbf{x}} = A\mathbf{x}$.

– Let $\mathbf{x}(t) = e^{tA}$, where

$$e^{tA} = \sum \frac{(tA)^k}{k!}$$

– $\|e^{tA}\| \leq \sum \frac{\|A^k\|}{k!} = \sum \frac{\|A\|^k}{k!}$.

– Let p be a polynomial of degree k . Then

$$p(a+x) = \sum_{k=0}^d \frac{p^{(k)}(a)}{k!} x^k$$

– If $A = D + N$, then...

- Nilpotent operators:

– Let $A = \text{diag}\{A_1, \dots, A_r\}$.

– We know that $A_k = \lambda_k I_{E_k} + N_k$ for each k .

– Every nilpotent N can be written in the form

$$\begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix}$$

10/28:

- The exam is long but not that hard. The only question is will you do good or very good.
- Revise the previous two homeworks, especially the last two.
- No justification for any of the true/false questions. Just circle T or F.
 - There are four problems. One is T/F (with multiple subparts); the other three are problem problems (with subparts, too).
 - Some of the questions will take you 2 seconds. Some you've already seen in the PSets.
 - The exam is supposed to be boring.
- Calculators?
 - No calculators needed. Calculators are for chem/physics exams.

- Not a lot of computation.
- 50 minutes.
- Chloe will be proctoring.
- Remember the determinant of “special” matrices.
 - $|\det U| = 1$ if U is unitary.
 - $\det A = \pm 1$ if A is orthogonal.
 - Make a list of matrix types that are automatically diagonalizable.
 - Determinant is the product of the eigenvalues.
 - Determinant of A is equal to the conjugate of the determinant of A^* .
- Most of the exercises use the inner product.
 - Whenever you had something to prove about eigenvalues or eigenbasis, you went through diagonalization or SVD or the inner product or polar decomposition.
 - Proving eigenvalues of self-adjoint matrices are real w/ the inner product.
- Eigenvalues/eigenvectors of a projection.
 - It’s implied that it’s asking you the multiplicities!!!
- Know useful facts but have an idea how to prove them as well.
- Recommends against shorthanding in the exams.
- Not grading on clarity (since the exam is long).
- Max and min are for when you’re sure something will be attained. Otherwise use sup and inf.

9.2 Chapter 9: Advanced Spectral Theory

- 10/28:
- Theorem 9.1.1 (Cayley-Hamilton): If p is the characteristic polynomial of A , $p(A) = 0$.
 - Theorem 9.2.1 (Spectral Mapping Theorem): For a square matrix A and an arbitrary polynomial p , $\sigma(p(A)) = p(\sigma(A))$. In other words, μ is an eigenvalue of $p(A)$ if and only if $\mu = p(\lambda)$ for some eigenvalue λ of A .
 - Corollary 9.2.2: Let A be a square matrix with eigenvalues $\lambda_1, \dots, \lambda_n$ and let p be a polynomial. Then $p(A)$ is invertible iff $p(\lambda_k) \neq 0$ for all $k = 1, \dots, n$.
 - Algebraic multiplicity is the dimension of the corresponding generalized eigenspace.

Part II

Point Set Topology of Metric Spaces

Chapter 1

The Real and Complex Number Systems

1.1 Notes

- 11/1:
- Spent a lot of time trying to cheer us up regarding the midterm.
 - There may be some true/false on linear algebra on the final.
 - Facts:
 1. $\sqrt{2}$ is irrational.
 2. Archimedes principle: If $x > 0$ and $y \in \mathbb{R}$, then there exists n such that $nx > y$.
 3. If $x > y$, then there exists $q \in \mathbb{Q}$ such that $x > q > y$.

1.2 Chapter 1: The Real and Complex Number Systems

From Rudin (1976).

- 11/6:
- Rudin (1976) presents several interesting proofs throughout this section that may be of interest later by means of their divergence from the ones with which I am familiar.
 - **Least-upper-bound property:** The property pertaining to a set S that if $E \subset S$, $E \neq \emptyset$, and E is bounded above, then $\sup E \in S$.
 - For example, \mathbb{Q} does not have the least-upper-bound property.
 - The **greatest-lower-bound property** is analogously defined.
 - Theorem: Suppose S is an ordered set with the least-upper-bound property, $B \subset S$ is nonempty, and B is bounded below. Let L be the set of all lower bounds of B . Then $\alpha = \sup L$ exists in S , and $\alpha = \inf B$. In particular, $\inf B$ exists in S .
 - Essentially, this theorem states that any set that satisfies the least-upper-bound property satisfies the greatest lower bound property.
 - **Existence theorem:** There exists an ordered field \mathbb{R} which has the least-upper-bound property. Moreover, \mathbb{R} contains \mathbb{Q} as a subfield.
 - The second statement implies that the operations of addition and multiplication on \mathbb{R} , when applied to \mathbb{Q} , coincide with the operations of addition and multiplication on \mathbb{Q} .

- **Archimedean property** (of \mathbb{R}): If $x \in \mathbb{R}$, $y \in \mathbb{R}$, and $x > 0$, then there is a positive integer n such that $nx > y$.
- Rudin (1976) proves several theorems about the real numbers from the least-upper-bound property as opposed to the traditional construction of the real numbers.
- Introduces the decimal system.
- **Finite real number system**: That which has been defined thus far.
- **Extended real number system**: The set $\mathbb{R} \cup \{+\infty, -\infty\}$ where $+\infty, -\infty$ obey the expected properties (supremum [resp. infimum] of every set, $x + \infty = \infty$, etc.).
- Defines the complex field axiomatically with complex numbers in the form (a, b) for $a, b \in \mathbb{R}$.
 - Notes that the real numbers form a subfield of the complex field.
 - Defines $i = (0, 1)$, proves $i^2 = -1$, proves $a + bi = (a, b)$.
- **Schwarz inequality**: If a_1, \dots, a_n and b_1, \dots, b_n are complex numbers, then

$$\left| \sum_{j=1}^n a_j \bar{b}_j \right|^2 \leq \sum_{j=1}^n |a_j|^2 \sum_{j=1}^n |b_j|^2$$

- **Euclidean k -space**: The vector space \mathbb{R}^k over the real field.

Chapter 2

Basic Topology

2.1 Notes

11/1:

- Equivalence relationships are denoted $A \sim B$.
 - These are...
 - Reflexive ($A \sim A$).
 - Symmetric ($A \sim B \iff B \sim A$).
 - Transitive ($A \sim B \ \& \ B \sim C \implies A \sim C$).
 - Equivalence relations give rise to equivalence classes.
- **Countable** (set A): A set A such that $A \sim \mathbb{N}$, in the sense that there exists a one-to-one and onto map from $\mathbb{N} \rightarrow A$.
 - Alternatively, A can be written in the form $A = \{f(n) : n \in \mathbb{N}\}$.
- **Finite countable** vs. **infinite countable** (see Rudin (1976)).
- \mathbb{N} denotes the natural numbers.
- \mathbb{N}_0 denotes the natural numbers including 0.
- \mathbb{Z} denotes the integers.
- We know that $\mathbb{N} \sim \mathbb{Z}$: Let $f : \mathbb{N} \rightarrow \mathbb{Z}$ be defined by

$$f(n) = \begin{cases} \frac{n}{2} & n \text{ even} \\ \frac{n-1}{2} & n \text{ odd} \end{cases}$$

- More facts.
 1. Every subset of a countable set is countable.
 2. Unions of countable sets are countable.
 - If the sets E_n for some finite list of numbers are countable, then $\bigcup_n E_n$ is countable.
 - Soug goes over the diagonalization method of counting.
 3. n -fold Cartesian products of countable sets are countable (we induct on n).
 - If A is countable and B is countable, then $A \times B$ is countable.
 - If A is finite and to each $\alpha \in A$ we assign a countable set E_α , $\otimes_{\alpha \in A} E_\alpha$ is countable.
- **Metric space**: A space X along with a metric $d : X \times X \rightarrow [0, \infty)$ such that

- $d(x, y) > 0$ iff $x \neq y$, and $d(x, x) = 0$ iff $x = 0$.
- $d(x, y) = d(y, x)$.
- $d(x, y) \leq d(x, z) + d(z, y)$.

- Example (\mathbb{R}^n):

- We may define d by

$$d(x, y) = \sqrt{\sum (x_i - y_i)^2}$$

- We can also define the p -metrics (recall normed spaces) with p where 2 is.

- Example ($X_p = \{f : Y \rightarrow \mathbb{R} : 1 \leq p < \infty, \int_Y |f|^p dy < \infty\}$):

- This is ℓ_p .
- Define

$$\|f - g\|_p = \left[\int_Y |f - g|^p dy \right]^{1/p}$$

- Convergence: $x_n \rightarrow x \iff d(x_n, x) \rightarrow 0$.

- **Neighborhood**: The set of all points a distance less than r away from p . Denoted by $N_r(p)$. Given by

$$N_r(p) = \{q \in X : d(p, q) < r\}$$

- **Limit point** (of E): A point p such that every neighborhood of p intersects E at a point other than p . Also known as **accumulation point**.

- Symbolically,

$$N_r(p) \cap (E \setminus \{p\}) \neq \emptyset$$

for all $r > 0$.

- **Isolated point** (of E): A point p such that $p \in E$ and p is not a limit point of E .

- **Closed** (set E): A set E that contains all of its limit points.

- **Interior** (point p): A point p such that there exists $N_r(p) \subset E$.

- **Open** (set E): A set E , all points of which are interior points.

- **Perfect** (set E): A set E that is closed and every point of E is a limit point of E .

- **Bounded** (set E): There exists a number M and a $y \in X$ such that $E \subset \{p : d(p, y) \leq M\}$.

- **Dense** (set E in X): A set E such that every point of X is a limit point of E or a point of E , itself.

11/3:

- Every neighborhood is an open set.

- If p is a limit point of E , every neighborhood of p contains infinitely many points of E .

- Thus, a finite set cannot have a limit point.

- Prove by contradiction: Suppose there is a neighborhood that contains only finitely many points of E . Then the neighborhood with radius smaller than the distance to the closest point does not contain any points of E , a contradiction.

- E is open iff $E^{c[1]}$ is closed.

- Assume E^c closed. If $p \in E$, then p is not a limit point of E^c . It follows that there exists a neighborhood of p that is entirely contained within E , so p is interior, as desired.

¹The complement of E .

- Suppose E is open. Let p be any limit point of E^c . Then $p \in E^c$.
- F is closed iff F^c is open.
- If $(G_\alpha)_{\alpha \in A}$ is a family of open sets in X , then the union is open.
 - Let $p \in \bigcup_{\alpha \in A} G_\alpha$. Then $p \in G_\alpha$ for some $\alpha \in A$. It follows that p is an interior point of G_α , so thus an interior point of the union of G_α with everything else.
- Finite intersections of open sets are open.
 - In the infinite case $\bigcap_{n \in \mathbb{N}} (-1/n, 1/n) = \{0\}$, an intersection of infinitely many open sets is closed.
 - However, in the finite case, just consider the neighborhood with the smallest radius and take this one.
- The intersection of closed sets is closed.
- The union of finitely many closed sets is closed.
 - These follow from the previous two by De Morgan's rule.
- Let $\bar{E} = E \cup E'$ where E' is the set of limit points of E .
- Let X be a metric space and $E \subset X$. Then
 1. \bar{E} is closed.
 - WTS: \bar{E}^c is open. Let $p \in \bar{E}^c$. Then p is neither in E nor is it a limit point of E . Thus, there exists a neighborhood of \bar{E}^c containing entirely points of \bar{E}^c . Therefore, \bar{E}^c is open, so \bar{E} is closed.
 2. $E = \bar{E}$ iff E is closed.
 - Think $p \in \bigcap G_\alpha$?
 3. $\bar{E} \subset F$ for any closed $F \supset E$.
 - If $E \subset F$, then any limit point of E will be a limit point of F . Thus, $E' \subset F'$. Then $\bar{E} = E \cup E' \subset F \cup F' = \bar{F} = F$ where the last equality holds because F is closed.
- Types of sets.

	Closed	Open	Perfect	Bounded
$\{z \in \mathbb{Q} : z < 1\}$	N	Y	N	Y
$\{z \in \mathbb{Q} : z \leq 1\}$	Y	N	Y	Y
Nonempty finite set	Y	N	N	Y
\mathbb{Z}	Y	N	N	N
$\{1/n : n \in \mathbb{N}\}$	N	N	N	Y
\mathbb{R}^2	Y	Y	Y	N
(a, b)	N	?	N	Y

Table 2.1: Types of sets.

- **Relatively open** (set E to Y): A set $E \subset Y \subset X$ such that if $p \in E$, then there exists a Y -neighborhood of E contained in E .
- Let $N_r^X(p) = \{y \in X : d(y, p) < r\}$ be a neighborhood of p in X , and let $N_r^Y(p) = \{y \in Y : d(y, p) < r\}$ be a neighborhood of p in Y . Then $N_r^Y(p) = N_r^X(p) \cap Y$.

- E is open relative to Y iff $E = G \cap Y$ where G is open relative to X .
- Introduces the supremum.
- If $E \subset \mathbb{R}$, $E \neq \emptyset$, and E is bounded above, $\sup E < \infty$.
- Let $y = \sup E$. Then $y \in \bar{E}$.
- There exists a sequence $a_n \in A$ such that $a_n \rightarrow x = \sup A$.
- A is compact iff any open cover of the set has a finite subcover.
- Study and *know* all of these proofs.

11/5:

- Compactness: Defines compactness in terms of open covers.
- Finite sets are compact.
- Compactness is “absolute” (i.e., it is not a relative property like openness).
 - If $K \subset Y \subset X$, then K is compact relative to X iff K is compact relative to Y .
 - V is open relative to Y iff $V = G \cap Y$ where G is open relative to X .
- Compact implies closed.
 - We will show K compact implies K^c open.
 - WTS: For all $p \in K^c$, there exists $N_r(p) \subset K^c$ such that $N_r(p) \cap K = \emptyset$.
- A closed subset of a compact set is compact.
 - Let K be compact and let $F \subset K$ be closed.
 - Take any open cover of F . Extend it to an open cover of K . Take the finite subcover of K . Naturally, this finite subcover is also a finite cover of $F \subset K$.
- F closed, K compact implies $F \cap K$ compact.
- If $(K_\alpha)_{\alpha \in A}$ is compact in X with finite intersection property (every intersection of any finite number of these sets is nonempty), then $\bigcap_{\alpha \in A} K_\alpha \neq \emptyset$.
 - Argue by contradiction.
 - Let $G_\alpha = K_\alpha^c$.
 - Assume the intersection is empty. Assume WLOG that no point of K_1 is in any of the other K_α ’s.
 - Then $\{G_\alpha\}_{\alpha \in A}$ be an open cover of K_1 .
 - K_1 compact implies there is a finite subcover $G_{\alpha_1}, \dots, G_{\alpha_n}$. Then $K \subset G_{\alpha_1} \cup \dots \cup G_{\alpha_n}$. This implies that $K_1 \cap K_{\alpha_1} \cap \dots \cap K_{\alpha_n} = \emptyset$, a contradiction.
- Let E be an infinite subset of a compact K . Then E has a limit point in K .
 - Argue by contradiction.
 - Suppose for all $p \in K$, there exists $N_r(p)$ such that $N_r(p) \cap E = \{p\}$.
 - Consider the set $\{N_r(p) : p \in K\}$. This is an open cover of K . Thus, there exists a finite subcover of it. But since $E \subset K \subset N_{r_1}(p_1) \cup \dots \cup N_{r_n}(p_n) = \{p_1\} \cup \dots \cup \{p_n\}$, E is finite, a contradiction.
- **2-cell** (in \mathbb{R}^2): A set that is the Cartesian product of two closed intervals.
 - Generalizes to **k-cells**.
- Let $I_n = [a_n, b_n] \subset \mathbb{R}$ such that $I_{n+1} \subset I_n$. Then $\bigcap_{n=1}^{\infty} I_n \neq \emptyset$.

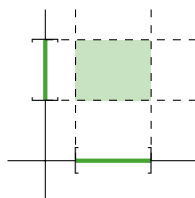
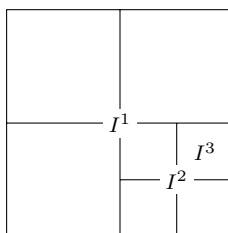


Figure 2.1: 2-cell.

- Let I_k be a k -cell in \mathbb{R}^k such that $I_k \supset I_{k+1}$. Then $\bigcap_k I_k \neq \emptyset$.
- We know that $a_m \leq a_{m+n} \leq b_{m+n} \leq b_m$, so $\sup a_n \in \bigcap I_n$.
- Every k -cell is compact.

Figure 2.2: k -cells are compact.

- Argue by contradiction.
- Consider an open cover of the k -cell I^1 . If it has a finite subcover, we're done. So suppose we have an open cover that doesn't have a finite subcover. Split the k -cell into 2^k chunks. At least one of the chunks I^2 must not have a finite subcover.
- Split that one into 2^k chunks. At least one of the chunks I^3 must not have a finite subcover.
- Continue.
- Thus, we have a decreasing family of k -cells, so by the previous result, their $\bigcap I^n \neq \emptyset$.
- Let $x \in \bigcap I^n$. Then the...
- Heine-Borel theorem: Let $E \subset \mathbb{R}^k$. Then TFAE^[2]
 1. E is closed and bounded.
 2. E is compact.
 3. Every infinite subset of E has a limit point in E .
 - $(1 \Rightarrow 2)$ E closed and bounded implies E is a closed subset of some I_k , so it's compact.
 - $(2 \Rightarrow 3)$ Already done.
 - $(3 \Rightarrow 1)$
 - Suppose E not bounded. Then there is an infinite sequence of points in E that never converges. Contradiction.
 - Suppose E is not closed. Then there exists a sequence of points in E which “converges” to an $x_0 \notin E$.

11/8: • Hewitt and Stromberg (1965) has harder analysis problems than Rudin (1976).

- Theorem: If P is a nonempty perfect subset of \mathbb{R}^k , then P is uncountable.

²The following are equivalent.

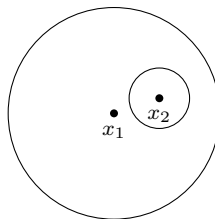


Figure 2.3: Nonempty perfect sets are uncountable.

- P perfect implies P infinite.
 - Suppose P is countable. Let $P = \{x_1, x_2, \dots\}$.
 - Start with x_1 . Take an open neighborhood V_1 of x_1 . Since x_1 is a limit point of P , there will be another point $x_2 \in P$ in V_1 . Choose V_2 to be a neighborhood of x_2 such that $\bar{V}_2 \subset V_1$.
 - Keep going — there is a point $x_3 \in P$ in V_2 , choose an appropriate neighborhood, etc.
 - Thus, we have a sequence of closed compact sets such that $\bar{V}_n \supset \bar{V}_{n+1}$ ($n \in \mathbb{N}$). It follows that $\bigcap \bar{V}_n \neq \emptyset$.
 - We also know that $V_n \cap P \neq \emptyset$ for each n .
 - Let $K_n = \bar{V}_n \cap P$. Each K_n is compact and $K_n \supset K_{n+1}$ for each n . Therefore, by compactness, $\bigcap K_n \neq \emptyset$. But the construction implies that $\bigcap K_n = \emptyset$ because we exhausted the whole sequence of possible points $x_i \in P$.
- Corollary: Any interval is uncountable.
 - The Cantor set:

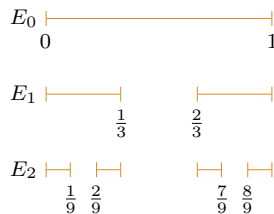


Figure 2.4: Constructing the Cantor set.

- Let $E_0 = [0, 1]$.
- Take out the middle third, so that $E_1 = [0, 1/3] \cup [2/3, 1]$.
- Take out the middle thirds of the remaining intervals and keep going.
- Thus, we are building a decreasing family of compact sets, so the overall intersection $E = \bigcap E_n$ of every set is nonempty.
- E^n is the union of 2^n closed intervals of length $n/3$. Thus, the overall length of E^n is $(2/3)^n$.
- Thus, we have a compact nonempty set with Lebesgue measure zero.
- E does not contain any segment of the form

$$\left(\frac{3k+1}{3^m}, \frac{3k+2}{3^m} \right)$$

for $k, m \in \mathbb{N}$.

- Therefore, no segment of the form (α, β) is contained in E (any segment of said form contains a segment of the above form).

- Moreover, E (the Cantor set) is perfect.
 - Let $x \in E$. WTS: For all segments S containing x , $S \cap (E \setminus \{x\}) \neq \emptyset$.
 - Let S be an arbitrary such segment...
- Consider the **Devil's staircase**.
 - $0 = \int_0^1 F'(x) dx = F(1) - F(0) = 1$. This function does not obey the fundamental theorem of calculus. A function satisfies the fundamental theorem of calculus if and only if it is absolutely continuous.
- Connected sets (motivation):
 - In a convex set, you can connect any two points with a straight line.
 - In a nonconvex connected set, there exist points that you must connect with a curve.
 - In a disconnected set, there exist points that cannot be connected via a line whose points lie wholly in the set.
- **Connected** (set E): A set E that is not the union of two **separated** sets.
- **Separated** (sets A, B): Two sets $A, B \subset X$ that are nonempty and such that $\bar{A} \cap B = \emptyset$, and $A \cap \bar{B} = \emptyset$.
- Theorem: $E \subset \mathbb{R}$ is connected iff $x, y \in E$ and $x < z < y$ implies $z \in E$.
 - If there is a $z \notin E$ between x, y , then $\{x \in E : x < z\}$ and $\{x \in E : z < y\}$ are separated sets, so E is not connected.

2.2 Chapter 2: Basic Topology

From Rudin (1976).

- 11/6:
- **Countable** (set A): A set A that is in bijective correspondence with the set of all positive integers. Also known as **enumerable**, **denumerable**.
 - **At most countable** (set A): A set A that is finite or countable.
 - An alternative definition of an **infinite** set would be a set that is equivalent to one of its proper subsets.
 - Theorem: Let A be the set of all sequences whose elements are the digits 0 and 1. This set A is uncountable.

Proof. Let $E = \{s_1, s_2, \dots\}$ be an arbitrary countable subset of A , where each s_j is a sequence whose elements are the digits 0 and 1. Let s be the sequence, the n^{th} term of which is the opposite of the n^{th} term of s_n (i.e., if the n^{th} term of s_n is 0, we set the n^{th} term of s equal to 1). This guarantees that s is distinct from each of the s_j , i.e., that $s \notin E$. It follows that $E \subsetneq A$, i.e., that every countable subset of A is a proper subset of A . Therefore, A must be uncountable (for otherwise A would be a proper subset of A , a contradiction). \square

- The idea of this proof is called **Cantor's diagonalization process**.
- Since every real number can be represented as a binary sequence of numbers, i.e., $A \sim \mathbb{R}$, the reals are uncountable.
- **Metric space**: A set X such that with any two points $p, q \in X$, there is associated a real number $d(p, q)$ such that
 1. $d(p, q) > 0$ if $p \neq q$; $d(p, p) = 0$.
 2. $d(p, q) = d(q, p)$.

3. $d(p, q) \leq d(p, r) + d(r, q)$ for any $r \in X$.

- **Distance** (from $p \in X$ to $q \in X$, X a metric space): The real number $d(p, q)$.
- **Distance function**: A function $d : X \times X \rightarrow \mathbb{R}$ that sends $(p, q) \mapsto d(p, q)$. Also known as **metric**.
- Every subset of a metric space is a metric space in its own right under the same distance function.
- **Segment** (from a to b): The set of all real numbers x such that $a < x < b$. Denoted by (a, b) .
- **Interval** (from a to b): The set of all real numbers x such that $a \leq x \leq b$. Denoted by $[a, b]$.
- **k -cell**: The set of all points $\mathbf{x} = (x_1, \dots, x_k) \in \mathbb{R}^k$ whose coordinates satisfy the inequalities $a_i \leq x_i \leq b_i$ where $a_i < b_i$ for each $1 \leq i \leq k$.

– Note that a 1-cell is an interval and a 2-cell is a rectangle.

- **Convex** (set E): A subset E of \mathbb{R}^k such that

$$\lambda \mathbf{x} + (1 - \lambda) \mathbf{y} \in E$$

for all $\mathbf{x}, \mathbf{y} \in E$ and $0 < \lambda < 1$.

– Balls and k -cells are both convex.

- The segment (a, b) is open as a subset of \mathbb{R}^1 , but not open as a subset of \mathbb{R}^2 .
- Since compactness is not relative, while it makes no sense to talk about *open* or *closed* metric spaces, it does make sense to talk about *compact* metric spaces.
- **Weierstrass theorem**: Every bounded infinite subset of \mathbb{R}^k has a limit point in \mathbb{R}^k .
- Theorem: Let P be a nonempty perfect set in \mathbb{R}^k . Then P is uncountable.

Proof. Since P is nonempty and perfect, there exists a limit point of P . It follows that P is infinite.

Now suppose for the sake of contradiction that P is countable, and denote the elements of P by $\mathbf{x}_1, \mathbf{x}_2, \dots$. We now construct a sequence $\{V_n\}$ of neighborhoods, as follows. Let $V_1 = N_r(\mathbf{x}_1)$. Clearly, $V_1 \subset P$ since $\mathbf{x}_1 \in P$. It follows that since V_1 is a neighborhood that V_1 contains infinitely many points of P . Now suppose inductively that V_n has been constructed. Thus, by analogous conditions to those on V_1 , we may let V_{n+1} be a neighborhood such that (i) $\bar{V}_{n+1} \subset V_n$, (ii) $\mathbf{x}_n \notin \bar{V}_{n+1}$, and (iii) $V_{n+1} \cap P$ is nonempty. By (iii), we can continue on to construct V_{n+2} , and so on and so forth.

Let $K_n = \bar{V}_n \cap P$. Since \bar{V}_n is closed and bounded, \bar{V}_n is compact. Additionally, since $\mathbf{x}_n \notin K_{n+1}$ for each n , no point of P lies in $\bigcap_1^\infty K_n$. Thus, since each $K_n \subset P$, $\bigcap_1^\infty K_n$ is empty. But this contradicts our previous result that since each K_n is nonempty, compact, and such that $K_n \supset K_{n+1}$, $\bigcap_1^\infty K_n$ is nonempty. \square

- Corollary: Every interval $[a, b]$ is uncountable. In particular, \mathbb{R} is uncountable.
- **Cantor set**: The set resulting from the following construction. Let $E_0 = [0, 1]$. Remove the segment $(1/3, 2/3)$, so that $E_1 = [0, 1/3] \cup [2/3, 1]$. Now remove the middle third of these two intervals to create E_2 . Continue on indefinitely.
 - This is a perfect set in \mathbb{R}^1 which contains no segment.
- **Separated** (sets A, B): Two subsets A, B of a metric space X such that $A \cap \bar{B}$ and $\bar{A} \cap B$ are empty.
- **Connected** (set E): A set E that is not the union of two nonempty separated sets.
- Separated sets are disjoint, but disjoint sets are not necessarily separated (consider $[0, 1]$ and $(1, 2)$).
- A subset E of \mathbb{R}^1 is connected if and only if it has the following property: If $x, y \in E$ and $x < z < y$, then $z \in E$.

Chapter 3

Numerical Sequences and Series

3.1 Notes

- 11/8: • Any bounded sequence in \mathbb{R}^k has a convergent subsequence.
- 11/10: • Read and understand the section about Cauchy sequences converging and the sup/inf.

3.2 Chapter 3: Numerical Sequences and Series

From Rudin (1976).

- 11/7: • Convergence of sequences is relative.
- For example, the sequence $1/n$ for $n = 1, 2, \dots$ converges in \mathbb{R} , but not in $(0, \infty)$.
- **Range** (of $\{p_n\}$): The set of all points p_n .
- This definition squares nicely with the formal definition of a sequence as a function p defined on \mathbb{N} .
- Theorem 3.6a: If $\{p_n\}$ is a sequence in a compact metric space X , then some subsequence of $\{p_n\}$ converges to a point of X .
- Theorem 3.7: The subsequential limits of a sequence $\{p_n\}$ in a metric space X form a closed subset of X .
- **Diameter** (of E): The supremum of the set

$$S = \{d(p, q) : p, q \in E\}$$

where E is a nonempty subset of a metric space X . Denoted by **diam** E .

- Theorem 3.10:
- (a) If \bar{E} is the closure of a set E in a metric space X , then
$$\text{diam } \bar{E} = \text{diam } E$$
 - (b) If K_n is a sequence of compact sets in X such that $K_n \supset K_{n+1}$ ($n = 1, 2, 3, \dots$) and if $\lim_{n \rightarrow \infty} \text{diam } K_n = 0$, then $\bigcap_1^\infty K_n$ consists of exactly one point.
- **Complete** (metric space): A metric space in which every Cauchy sequence converges.
- All compact metric spaces and all Euclidean spaces are complete.

- The metric space $(\mathbb{Q}, |x - y|)$ is not complete.
- **Monotonically increasing** (sequence $\{s_n\}$): A sequence $\{s_n\}$ of real numbers such that $s_n \leq s_{n+1}$ for each $n \in \mathbb{N}$.
- **Monotonically decreasing** (sequence $\{s_n\}$): A sequence $\{s_n\}$ of real numbers such that $s_n \geq s_{n+1}$ for each $n \in \mathbb{N}$.
- **Monotonic sequences**: The class of all sequences that are either monotonically increasing or monotonically decreasing.
- **Upper limit** (of $\{s_n\}$): The supremum of the set E of all subsequential limits of $\{s_n\}$. *Denoted by s^* , $\limsup_{n \rightarrow \infty} s_n$.*
- **Lower limit** (of $\{s_n\}$): The infimum of the set E of all subsequential limits of $\{s_n\}$. *Denoted by s_* , $\liminf_{n \rightarrow \infty} s_n$.*
- Theorem 3.17: Let $\{s_n\}$ be a sequence of real numbers. Then s^* has (and is the only number to have both of) the following two properties.
 - (a) $s^* \in E$.
 - (b) If $x > s^*$, then there is an integer N such that $n \geq N$ implies $s_n < x$.

An analogous result holds for s_* .

11/8:

- Series are defined in terms of sequences. Moreover, sequences can be defined in terms of series: Let $a_1 = s_1$, $a_n = s_n - s_{n-1}$ ($n \in \mathbb{N} + 1$). Thus, every theorem about sequences can be stated in terms of series and vice versa, but it is nevertheless useful to consider both concepts (Rudin, 1976, p. 59).
- Theorem 3.27: Suppose $\{a_n\}$ is a monotonically decreasing sequence of nonnegative terms. Then the series $\sum_{n=1}^{\infty} a_n$ converges if and only if the series

$$\sum_{k=0}^{\infty} 2^k a_{2^k} = a_1 + 2a_2 + 4a_4 + 8a_8 + \cdots$$

converges.

- Theorem 3.29: If $p > 1$,

$$\sum_{n=2}^{\infty} \frac{1}{n(\log n)^p}$$

converges; if $p \leq 1$, the series diverges.

- Note that $\log n = \ln n$.
- Note that we sum from $n = 2$ since $\log 1 = 0$.
- **e**: The number

$$e = \sum_{n=0}^{\infty} \frac{1}{n!}$$

- Theorem 3.31: $\lim_{n \rightarrow \infty} (1 + 1/n)^n = e$.
- Theorem 3.32: e is irrational.
- Theorem 3.39: Given the power series $\sum c_n z^n$, put

$$\alpha = \limsup_{n \rightarrow \infty} \sqrt[n]{|c_n|} \qquad R = \frac{1}{\alpha}$$

(If $\alpha = 0$, let $R = +\infty$; if $\alpha = +\infty$, let $R = 0$.) Then $\sum c_n z^n$ converges if $|z| < R$ and diverges if $|z| > R$.

- **Radius of convergence** (of a power series): The number R defined by Theorem 3.39.
- Theorem 3.41 (partial summation formula): Given two sequence $\{a_n\}, \{b_n\}$, put

$$A_n = \begin{cases} \sum_{k=0}^n a_k & n \geq 0 \\ 0 & n = -1 \end{cases}$$

Then if $0 \leq p \leq q$, we have

$$\sum_{n=p}^q a_n b_n = \sum_{n=p}^{q-1} A_n (b_n - b_{n+1}) + A_q b_q - A_{p-1} b_p$$

- **Product** (of $\sum a_n, \sum b_n$): The series $\sum c_n$ defined by

$$c_n = \sum_{k=0}^n a_k b_{n-k}$$

for each $n = 0, 1, 2, \dots$

- We motivate this definition by noting that if $\sum c_n$ is the product of $\sum a_n, \sum b_n$, then

$$\sum_{n=0}^{\infty} a_n z^n \cdot \sum_{n=0}^{\infty} b_n z^n = \sum_{n=0}^{\infty} c_n z^n$$

- Setting $z = 1$ then yields the given definition.

- The product of two convergent series may diverge. However...
- Theorem 3.50 (by Mertens): Suppose (a) $\sum_{n=0}^{\infty} a_n$ converges absolutely, (b) $\sum_{n=0}^{\infty} a_n = A$, (c) $\sum_{n=0}^{\infty} b_n = B$, and (d) $\sum_{n=0}^{\infty} c_n$ is the product of $\sum_{n=0}^{\infty} a_n$ and $\sum_{n=0}^{\infty} b_n$. Then

$$\sum_{n=0}^{\infty} c_n = AB$$

- Theorem 3.51 (by Abel): If $\sum a_n, \sum b_n, \sum c_n$ converge to A, B, C , respectively, and $\sum c_n$ is the product of $\sum a_n, \sum b_n$, then $C = AB$.
- **Rearrangement** (of $\sum a_n$): A series $\sum a'_n$ defined by $a'_n = a_{k_n}$ for each $n \in \mathbb{N}$, where $\{k_n\}$ is a sequence in which every positive integer appears once and only once (that is, $\{k_n\}$ is a 1-1 function from \mathbb{N} onto \mathbb{N}).
- Theorem 3.54: Let $\sum a_n$ be a series of real number which converges, but not absolutely. Suppose $-\infty \leq \alpha \leq \beta \leq \infty$. Then there exists a rearrangement $\sum a'_n$ with partial sums s'_n such that

$$\liminf_{n \rightarrow \infty} s'_n = \alpha \qquad \limsup_{n \rightarrow \infty} s'_n = \beta$$

- Theorem 3.55: If $\sum a_n$ is a series of complex numbers which converges absolutely, then every rearrangement of $\sum a_n$ converges, and they all converge to the same sum.

Chapter 4

Continuity

4.1 Notes

- 11/8:
- Consider a function $f : X \rightarrow Y$ whose domain and codomain are, respectively, the metric spaces (X, d_X) and (Y, d_Y) .
 - **Limit** (of f at p): A point $q \in Y$ such that for all $\epsilon > 0$, there exists δ such that $d_X(x, p) < \delta$ implies $d_Y(q, f(x)) < \epsilon$, where p is a limit point of X (otherwise, $x \nrightarrow p$).
 - **Continuous** (function f at p): A function f such that $\lim_{x \rightarrow p} f(x) = f(p)$.
 - f is continuous on X if it is continuous at every $p \in X$.
 - **Uniformly continuous** (function f): A function f such that for every $\epsilon > 0$, there exists a $\delta > 0$ such that $d_X(x, y) < \delta$ implies $d_Y(f(x), f(y)) < \epsilon$ for all $x, y \in X$.
- 11/10:
- f, g continuous implies $f + g, fg$, and f/g continuous, the latter where $g(x) \neq 0$.
 - If f, g continuous, then $h = g \circ f$ is continuous.
 - Theorem: $f : X \rightarrow Y$ is continuous iff $f^{-1}(V)$ is open in X for every $V \subset Y$ open.

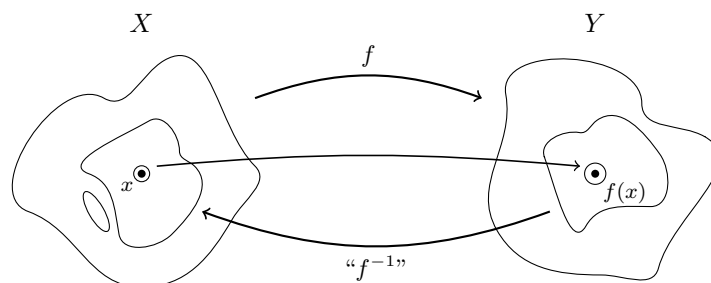


Figure 4.1: Set theoretic definition of continuity.

- This works in a general topological space, too, not just a metric space.
- Note that $f^{-1}(V)$ is not a function defined on V ; it's a specifically defined set $\{x \in X : f(x) \in V\}$.
- f being continuous means that open circular neighborhood of a point x in the domain maps to an area of the range encompassed by a circular neighborhood of $f(x)$.
- The other condition means that every open set surrounding $f(x)$ maps to an open set of the domain surrounding x . Indeed, going off of this definition, if an open set containing $f(x)$ maps to an open set containing x , then we can choose a neighborhood subset of the open set surrounding x and know that it will map into a neighborhood subset of the open set surrounding $f(x)$.

- Corollary: $f : X \rightarrow Y$ continuous iff $f^{-1}(C)$ closed for every $C \subset Y$ closed.
 - We use the property that $f^{-1}(X \subset C) = X \subset f^{-1}(C)$.
 - Let $f_1 : X_1 \rightarrow Y_1$ and $f_2 : X_2 \rightarrow Y_2$. Suppose $f : X_1 \times X_2 \rightarrow Y_1 \times Y_2$ is defined by $f(x_1, x_2) = (f_1(x_1), f_2(x_2))$. Then f is continuous iff f_1, f_2 are continuous, under appropriately defined metrics.
 - Continuity and compactness.
 - Theorem: $f : X \rightarrow Y$ continuous and X compact imply $f(X)$ compact.
 - Let $\{V_\alpha\}$ be an open cover of $f(X)$.
 - Then $\{f^{-1}(V_\alpha)\}$ is an open cover of X .
 - Choose a finite subcover of $\{f^{-1}(V_\alpha)\}$. Then the corresponding V_α 's form a finite subcover of $f(X)$.
 - If $f : X \rightarrow \mathbb{R}^k$ is continuous and X is compact, $f(X)$ is compact and closed/bounded.
 - If $f : X \rightarrow \mathbb{R}$ is continuous and X is compact, then $M = \sup_{x \in X} |f(x)| = |f(\bar{x})|$ and $m = \inf_{x \in X} |f(x)| = |f(\underline{x})|$ where $\bar{x}, \underline{x} \in X$.
 - There is a subsequence $\{x_m\}$ such that $|f(x_m)| \rightarrow M$. Since $f(X)$ is compact, the limit of this sequence is in $f(X)$.
 - If $f : X \rightarrow Y$ is continuous, bijective, and X, Y are compact, $f^{-1} : Y \rightarrow X$ is continuous.
 - Uniform continuity.
 - Examples.
 - Linear functions are uniformly continuous.
 - $f : (0, 1) \rightarrow \mathbb{R}$ defined by $f(x) = x^2$ is uniformly continuous.
 - $f : (a, \infty) \rightarrow \mathbb{R}$ defined by $f(x) = x^2$ is *not* uniformly continuous.
 - $f : (a, \infty) \rightarrow \mathbb{R}$ defined by $f(x) = 1/x$ is uniformly continuous if $a > 0$.
 - $f : (0, \infty) \rightarrow \mathbb{R}$ defined by $f(x) = 1/x$ is *not* uniformly continuous.
 - **Lipschitz continuous** (function f on $E \subset X$): A function such that $|f(x) - f(y)| \leq L|x - y|$ for each $x, y \in E$.
 - Theorem: $f : X \rightarrow Y$ continuous and X compact implies f is uniformly continuous.
 - Fix $\epsilon > 0$. There exists $\delta = \delta(p) > 0$.
 - Def. of continuity: $q \in N_{\delta(p)}(p)$ implies $f(q) \in N_\epsilon(f(p))$.
 - $\{N_{\delta(p)/2}(p) : p \in X\}$ is an open cover of X . Choose a finite subcover. Let $\delta = \min(\delta(p_1)/2, \dots, \delta(p_n)/2)$.
 - ...
- 11/12:
- $f : X \rightarrow Y$ continuous and $E \subset X$ connected implies $f(E)$ connected.
 - Suppose $f(E) = A \cup B$, A, B nonempty, separated subsets of Y .
 - Let $G = E \cap f^{-1}(A)$, $H = E \cap f^{-1}(B)$. It follows that $E = G \cup H$, where G, H nonempty.
 - $A \subset \bar{A}$ implies $G \subset f^{-1}(\bar{A})$ implies (inverse image def. of continuity) $\bar{G} \subset f^{-1}(A)$ implies $f(\bar{G}) \subset \bar{A}$.
 - $f(H) = B$ and $\bar{A} \cap B = \emptyset$ yields $\bar{G} \cap H = \emptyset$. Symmetrically, $G \cap \bar{H} = \emptyset$. This contradicts our assumption that E is connected.
 - Introduces monotone functions.
 - Theorem: If f is monotonic on (a, b) , then the set of points of (a, b) at which f is discontinuous is at most countable.
 - ...

4.2 Chapter 4: Continuity

From Rudin (1976).

- 11/8:
- **Limit** (of f at p): The point $q \in Y$, if it exists, such that for every $\epsilon > 0$, there exists a $\delta > 0$ such that $d_Y(f(x), q) < \epsilon$ for all points $x \in E$ for which $0 < d_X(x, p) < \delta$, where $(X, d_X), (Y, d_Y)$ are metric spaces, $E \subset X$, $f : E \rightarrow Y$, and $p \in E'$. Denoted by $\lim_{x \rightarrow p} f(x)$.
 - Note that we do not require that $p \in E$; only that some elements of the domain E approach p .
 - We also write $f(x) \rightarrow q$ as $x \rightarrow p$.
 - Theorem 4.2: Let X, Y, E, f , and p be as specified above. Then $\lim_{x \rightarrow p} f(x) = q$ iff $\lim_{n \rightarrow \infty} f(p_n) = q$ for every sequence $\{p_n\}$ in E such that $p_n \neq p$ for any n and $\lim_{n \rightarrow \infty} p_n = p$.
 - Rudin (1976) proves the sum, product, and quotient rules of limits from the analogous properties of series.
 - Continuity is defined.
 - Note that f *does* have to be defined at p to be continuous at p (in comparison to the fact that it can have a limit at a point p' at which it is not defined).
 - Thus, for proofs concerning continuity (as opposed to limits), we will consider functions f the domains of which are metric spaces, not *subsets* of metric spaces.
 - It follows from the definition that if $p \in E$ is isolated, then every possible f defined on E is continuous at p .
 - Theorem 4.7: Compositions of continuous functions are continuous.
 - Theorem 4.8: Preimage definition of continuity.
 - Theorem 4.9: If f, g are complex continuous functions on X , $f + g$, fg , and f/g are continuous on X .
 - Theorem 4.10: \mathbf{f} continuous implies f_1, \dots, f_k continuous. Also, $\mathbf{f}, \mathbf{g} : X \rightarrow \mathbb{R}^k$ continuous implies $\mathbf{f} + \mathbf{g}$ and $\mathbf{f} \cdot \mathbf{g}$ continuous.
- 11/9:
- Theorem 4.14: f continuous and X compact implies $f(X)$ compact.
 - Theorem 4.15: $\mathbf{f} : X \rightarrow \mathbb{R}^k$ continuous and X compact implies $f(X)$ closed and bounded.
 - Theorem 4.16: f continuous and X compact implies f attains its minimum and maximum.
 - Theorem 4.17: $f : X \rightarrow Y$ continuous, 1-1 for X, Y compact implies $f^{-1} : Y \rightarrow X$ continuous.
 - Theorem 4.19: f continuous and X compact implies f uniformly continuous.
 - Theorem 4.20: Compactness is a necessary condition in Theorems 4.14, 4.15, 4.16, and 4.19.
 - Theorem 4.22: $f : X \rightarrow Y$ continuous and $E \subset X$ connected implies $f(E)$ connected.
 - Theorem 4.23: Intermediate value theorem.
 - **Right-hand limit** (of f at x): Denoted by $f(x+)$.
 - **Left-hand limit** (of f at x): Denoted by $f(x-)$.
 - **Discontinuity of the first kind** (of f at x): A discontinuity of f at x such that $f(x+)$ and $f(x-)$ exist. Also known as **simple discontinuity**.
 - **Discontinuity of the second kind** (of f at x): A discontinuity of f at x that is not of the first kind (i.e., a discontinuity such that at least one of $f(x+)$ and $f(x-)$ does not exist).

- Theorem 4.29: If f is monotonic on (a, b) , then $f(x+), f(x-)$ exist at every $x \in (a, b)$.
- Corollary: Monotonic functions have no discontinuities of the second kind.
- Theorem 4.30: If f is monotonic on (a, b) , then the set of points of (a, b) at which f is discontinuous is at most countable.

11/20:

Proof. Suppose first that f is increasing. Let E be the set of points at which f is discontinuous. By Theorem 4.29, for every $x \in E$, $f(x-), f(x+)$ exist. Thus, we may pick a rational number $r(x)$ such that $f(x-) < r(x) < f(x+)$. Moreover, since $x_1 < x_2$ implies $f(x_1+) \leq f(x_2-)$, we have that $r(x_1) \neq r(x_2)$. Having established an injective function from E to the rationals \mathbb{Q} , we know that E is at most countable. The argument where f is decreasing is symmetric. \square

- Gives an example of a function with discontinuities that are not isolated.

Chapter 7

Sequences and Series of Functions

7.1 Notes

11/15:

- Soug will not test on differentiation/integration assuming that we know them already.
- **Pointwise convergent** (sequence $(f_n)_{n \in \mathbb{N}}$ of functions): A sequence of functions $f_n : E \rightarrow \mathbb{R}$ such that $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ for all $x \in E$.
- Can we interchange “limit” in the above definition with continuity, convergence of series, integration, differentiation, etc.?
- Examples with negative answer:

1. Interchanging limits: Let $S_{mn} = \frac{m}{m+n}$. $S_{mn} \rightarrow 1$ as $m \rightarrow \infty$, and $S_{mn} \rightarrow 0$ as $n \rightarrow \infty$.
2. $f_n(x) = x^2/(1+x)^n$. $f(x) = \sum_{n=1}^{\infty} f_n(x)$. If $x = 0$, then $f_n(x) = 0$ for all n and $f(x) = 0$. If $x \neq 0$, we have

$$f(x) = \sum_{n=1}^{\infty} \frac{x^2}{(1+x^2)^n} = x^2 \sum_{n=1}^{\infty} X^n = \frac{x^2}{1-X} = \frac{x^2}{1-(1/(1+x^2))} = 1+x^2$$

3. Consider $f_m(x) = \lim_{n \rightarrow \infty} (\cos(m\pi x))^{2n}$. $\lim_{m \rightarrow \infty} f_m(x)$ goes to 0 if $x \notin \mathbb{Q}$ and goes to 1 if $x \in \mathbb{Q}$. $f_m \rightarrow \chi_{\mathbb{Q}}$, where $\chi_{\mathbb{Q}}$ is the characteristic function of the rationals which is not Riemann integrable (partitions, upper and lower integrals, etc.).
 4. $f_n(x) = \sin nx/\sqrt{n} \rightarrow f(x) = 0$ for all x . However, $f'_n(x) = \sqrt{n} \cos nx \not\rightarrow 0$
 5. If $0 \leq x \leq 1$, define $f_n(x) = n^2 x(1-x^2)^n$. We know that $f_n(0) = 0$. $\lim_{n \rightarrow \infty} f_n(x) = 0$ for all $x \in (0, 1]$. We can show that $\int_0^1 x(1-x^2)^n dx = 1/(2n+2)$. Thus, $\int_0^1 f_n(x) dx = n^2/(2n+2)$. Limit of the functions is zero, but their integrals diverge to infinity.
- **Uniformly convergent** (sequence $(f_n)_{n \in \mathbb{N}}$ of functions on E): A sequence of functions $f_n : E \rightarrow \mathbb{R}$ such that for all $\epsilon > 0$, there exists N such that if $n \geq N$, then $|f_n(x) - f(x)| < \epsilon$ for all $x \in E$. Denoted by $f_n \rightrightarrows f$.
 - Theorem: $f_n \rightrightarrows f$ iff $(f_n)_{n \in \mathbb{N}}$ is uniformly Cauchy (i.e., for all $\epsilon > 0$, there exists N such that if $n, m \geq N$ then $|f_n(x) - f_m(x)| < \epsilon$ for all $x \in E$).
 - Let $M_n = \sup_{x \in E} |f_n(x) - f(x)|$. If $f_n \rightarrow f$ pointwise, then $f_n \rightrightarrows f$ if $M_n \rightarrow 0$.
 - Theorem: If $(f_n)_{n \in \mathbb{N}}$ and $|f_n(x)| \leq M_n$, then $\sum f_n \rightrightarrows f$ if $\sum M_n < \infty$.
 - Theorem: If E is a compact metric space, $f_n \rightrightarrows f$ in E , x is a limit point of E , and $\lim_{t \rightarrow x} f_n(t) = A_n$ exists, then $(A_n)_{n \in \mathbb{N}}$ converges and $\lim_{t \rightarrow x} f(t) = \lim_{n \rightarrow \infty} A_n$.

- Corollary: $\lim_{t \rightarrow x} \lim_{n \rightarrow \infty} f_n(t) = \lim_{n \rightarrow \infty} \lim_{t \rightarrow x} f_n(t)$.

11/16:

- Fix $\epsilon > 0$. Then $f_n \rightrightarrows f$ implies that there exists some N such that $n, m \geq N$ implies $|f_n(t) - f_m(t)| < \epsilon$ for all $t \in E$.
- x is a limit point of E and $t \rightarrow x$ implies $|A_n - A_m| < \epsilon$. Thus, $(A_n)_{n \in \mathbb{N}}$ is cauchy, so there exists A such that $A_n \rightarrow A$.
- WTS: $|f(t) - A| \leq |f(t) - f_n(t)| + |f_n(t) - A_n| + |A_n - A|$, so we WTS the three terms on the right are small.
- There exists n such that $|f(t) - f_n(t)| < \epsilon/3$ for all t since $f_n \rightrightarrows f$ by hypothesis.
- Since t is in a small neighborhood of x , there exists n such that $|A_n - A| < \epsilon/3$.
- We also have $|f_n(t) - A_n| < \epsilon/3$ by hypothesis.
- This is a very important proof to understand, because proofs like this pop up often.
- Corollary: f_n continuous and $f_n \rightrightarrows f$ implies f is continuous.
- Theorem: Let K be compact. Assume
 - (a) $(f_n)_{n \in \mathbb{N}} \subset C(K) = \{f : K \rightarrow \mathbb{R} \mid f \text{ continuous}\}$.
 - (b) $f_n \rightarrow f$ pointwise in K and $f \in C(K)$.
 - (c) $f_n(x) \geq f_{n+1}(x)$ for all $x \in K$.

Then $f_n \rightrightarrows f$.

- WLOG $f = 0$, $g_n = f_n - f \rightarrow 0$, $g_n \geq g_{n+1} \geq 0$.
- For all $\epsilon > 0$, there exists N such that $n \geq N$ and $0 \leq g_n(x) \leq \epsilon$ for all $x \in K$.
- $K_n = \{x \in K : g_n(x) \geq \epsilon\}$.
- g_n continuous implies K_n closed. This combined with K compact implies K_n is compact.
- g_n decreasing implies $K_n \supset K_{n+1}$. Thus, K_n is a nested family of compact sets, so $\bigcap K_n \neq \emptyset$.
- This implies that each K_n is nonempty, contradicting the fact that each $g_n \rightarrow 0$ for all x .
- Thus, there exists an N such that K_n is empty for all $n \geq N$. Thus $g_n(x) \leq \epsilon$ for all $x \in K$, $n \geq N$.
- Note that the compactness of K is important. If $f : (0, 1) \rightarrow \mathbb{R}$ is defined by $f(x) = 1/(nx + 1)$, then $f_n \rightarrow 0$, but $f_n \not\rightrightarrows f$.
- Let $C(X) = \{f : X \rightarrow \mathbb{R} \mid f \text{ continuous, bounded}\}$ for X a metric space.
- If we define $\|f\| = \sum_{x \in X} |f(x)|$, for $f, g \in C(X)$, we may define $d(f, g) = \|f - g\|$. This definition satisfies the properties of a distance function, and $\|\cdot\|$ is a norm.
 - Thus, $C(X)$ is a complete metric space, a normed space, or, specifically, a **Banach space**.
- Theorem: $(f_n)_{n \in \mathbb{N}} \subset C(X)$ such that $\|f_n - f_m\|_{n, m \rightarrow \infty} \rightarrow 0$. Then there exists $f \in C(X)$ such that $\|f_n - f\|_{n \rightarrow \infty} \rightarrow 0$.
 - We get such a strong statement using properties of the image, not properties of the domain.
 - For all $\epsilon > 0$, there exists N such that $n, m \geq N$.
 - $|f_n(x) - f_m(x)| \leq \|f_n - f_m\| < \epsilon$ for all x .
 - Then there exists f such that $f_m(x) \rightarrow f(x)$. It follows that $|f_n(x) - f_m(x)| < \epsilon$
- Uniform convergence and integration.
- Stieltjes integral.

- Define $\alpha : \mathbb{R} \rightarrow \mathbb{R}$ nondecreasing.
- If we sum over the minimums/maximums of a partition times $\alpha(x_{i+1}) - \alpha(x_i)$ instead of $x_{i+1} - x_i$, we obtain said integral as the upper/lower limits just like the Riemann integral.
- We write $\int_a^b f(x) d\alpha(x)$ where $d\alpha(x) = \alpha(x) dx$.
- Theorem: If α is nondecreasing on $[a, b]$, $f_n \in R(\alpha)$ such that $f_n \rightrightarrows f$ on $[a, b]$
 - We have

$$\begin{aligned}
 \left| \int f_n(x) d\alpha(x) - \int f(x) d\alpha(x) \right| &= \left| \int (f_n - f)(x) d\alpha(x) \right| \\
 &\leq \|f_n - f\|(\alpha(b) - \alpha(a)) \\
 &\leq \int |f_n - f| d\alpha(x) \\
 &\leq \int \|f_n - f\| d\alpha(x) \leq \|f_n - f\| \int_a^b d\alpha(x) = \|f_n - f\|(\alpha(b) - \alpha(a))
 \end{aligned}$$

11/19:

- Suppose $f_n \rightarrow f$ and $f'_n \rightarrow g$. When does $f' = g$?
- Theorem: If $f_n : [a, b] \rightarrow \mathbb{R}$ is differentiable, $f_n(x_0)$ converges for some $x_0 \in [a, b]$, and f'_n converges uniformly on $[a, b]$, then there exists f differentiable on $[a, b]$ such that $f_n \rightrightarrows f$ and $f'_n \rightrightarrows f'$.

- Assume the f'_n are continuous. Then $f_n(x) - f_n(x_0) = \int_{x_0}^x f'_n(y) dy$.
- Since $f'_n \rightrightarrows g$, $\int_{x_0}^x f'_n(y) dy \rightarrow \int_{x_0}^x g(y) dy$.
- It follows since $f_n(x_0) \rightarrow f(x_0)$ that $f_n \rightrightarrows f$.
- By the previous theorem, if

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

then

$$\lim_{n \rightarrow \infty} f'_n(x) = \lim_{h \rightarrow 0} \lim_{n \rightarrow \infty} \frac{f_n(x+h) - f_n(x)}{h} = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$

- Fix $\epsilon > 0$. Then there exists N such that $n, m \geq N$ such that $|f_n(x_0) - f_m(x_0)| < \epsilon/2$ and $|f'_n(t) - f'_m(t)| < \epsilon/2$ for all $t \in [a, b]$.
- We know that $f_n(t) - f_n(x_0) = \int_{x_0}^t f'_n(y) dy$ and $f_m(t) - f_m(x_0) = \int_{x_0}^t f'_m(y) dy$.
- Thus,

$$|f_n(t) - f_n(x_0)| \leq |f_n(t) - f_m(t)| + |f_m(t) - f_m(x_0)| + |f_m(t) - f_m(x_0)|$$
- Let $f_n(t) - f_n(x_0) = c_n(t - x_0)$ and $f_m(t) - f_m(x_0) = c_m(t - x_0)$.
- ...
- Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. What conditions on f imply that f' exists?
- Suppose f is Lipschitz continuous (equivalent to saying there exists $L > 0$ such that $|f(x) - f(y)| \leq L|x - y|$); then f' exists **almost everywhere**.
 - If f differentiable, this is equivalent to saying f bounded.
- **Almost everywhere**: Something happens almost everywhere if the set of places where it doesn't happen has measure zero.
- Suppose f is **Hölder continuous**, then f' does not exist?
- **Hölder continuous** (function f): There exists $L > 0$ such that $|f(y) - f(x)| < L|x - y|^\alpha$ where $\alpha \in (0, 1)$

- Suppose f exists such that f is Hölder continuous in a neighborhood of every point in the domain. This function is not anywhere differentiable. Such a function does indeed exist (and it's Brownian motion). The construction of such a function is the essence of Stochastic analysis.
 - Probabilistically: Has mean zero, distributed as a normal function like the Gaussian, and the increments are independent of each other.
 - Analytically: It's a function that is Hölder continuous at half plus ϵ for every ϵ and it is nowhere differentiable.
- Theorem: There exists $f : \mathbb{R} \rightarrow \mathbb{R}$ continuous but nowhere differentiable.
 - This theorem is due to Weierstrass and as such, such functions are typically called Weierstrass functions.
- A general class of functions that are nowhere differentiable (not in Rudin (1976); we don't have to prove this).
 - Example 1:

$$f(x) = \sum_{n=0}^{\infty} a^n \cos(b^n \pi x)$$

where $0 < a < 1$, b positive odd integer greater than 1, and $ab > 1 + \frac{3}{2}\pi$.

■ This function at every point oscillates more and more and more.

- Rudin (1976)'s simple example.
 - $\phi : [-1, 1] \rightarrow \mathbb{R}$ defined by $\phi(x) = |x|$ is not differentiable at zero.
 - Takes ϕ extends it periodically with period 2, creating a sawtooth function.
 - Repeat the behavior so that the nondifferentiability becomes more and more frequent to get

$$f(x) = \sum_1^{\infty} \left(\frac{3}{4}\right)^n \phi(4^n x)$$

- This is continuous.
- Fix any $x \in \mathbb{R}$, $m \in \mathbb{N}$. Then $\delta_m = \pm \frac{1}{2} \cdot 4^{-m}$.
- Then consider $4^m x$, $4^m(x + \delta_m)$.
- Rudin asserts

$$\left| \frac{f(x + \delta_m) - f(x)}{\delta_m} \right| \rightarrow \infty$$

as $m \rightarrow \infty$ for all x .

References

- Hewitt, E., & Stromberg, K. (1965). *Real and abstract analysis: A modern treatment of the theory of functions of a real variable*. Springer.
- Rudin, W. (1976). *Principles of mathematical analysis* (A. A. Arthur & S. L. Langman, Eds.; Third). McGraw-Hill.
- Treil, S. (2017). *Linear algebra done wrong* [http://www.math.brown.edu/streil/papers/LADW/LADW_2017-09-04.pdf].