Lecture 1:

Groups A set G with operation $+$ is a group if: (P1) $(a+b)+c = a+(b+c)$ (P2) $\exists e \in G: \ e+g = g+e = g$ (P3) $\forall a \in G, \exists b \in G: \ a+b = b+a = e$ Commutative (Abelian) if $(P4)$ $a+b = b+a$ Examples: $(\mathbb{R}^n, +)$, $(\mathbb{R}^*, \cdot)$, $(S_n, \circ)$, rotations of a square. **Fields** A set F with $+, \cdot$ is a field if: (P1) $(F, +)$ commutative group (id 0), (P2) $(F \setminus \{0\}, \cdot)$ commutative group (id 1), (P3) $a(b+c) = ab+ac$ Examples: $(\mathbb{R}, +, \cdot)$, $(\mathbb{C}, +, \cdot)$, $(\mathbb{Z}_p, +_p, \cdot_p)$ for prime p $(\mathbb{Z}_n, +_n, \cdot_n)$ is a field $\Leftrightarrow$ n prime. **Vector Spaces** V over field F: (P1) $(V, +)$ commutative group, (P2) $1 \cdot v = v$, (P3) $a(u+v) = au+av, (a+b)u = au+bu$ Examples: $\mathbb{R}^n, F(X, \mathbb{R}), C(X), C^r([a,b])$. **Subspaces** $U \subseteq V$ is subspace if $\forall \lambda, \mu \in F, u, v \in U : \lambda u + \mu v \in U$ Examples: $C(X) \subset F(X, \mathbb{R})$, symmetric matrices, lines through origin. **Linear Combination & Span** Linear combination: $\sum_{i=1}^{n} \lambda_i u_i$; $\text{span}\{u_1, \dots, u_n\} = \{\sum_{i=1}^{n} \lambda_i u_i \mid \lambda_i \in F\}$; generators: $\{u_1, \dots, u_n\}$. **Linear Independence** $\sum_{i=1}^{n} \lambda_i v_i = 0 \Rightarrow \lambda_i = 0 \ \forall i$ Examples: $(1,0,0)^T, (1,1,0)^T, (1,1,1)^T$ independent; $\sin x, \cos x$ independent; any $d+1$ vectors in $\mathbb{R}^d$ dependent. **Basis & Dimension** $B \subset V$ is basis if $\text{span}(B) = V$ and B linearly independent. If U spans V, can reduce to a basis. V finite-dimensional $\Leftrightarrow$ has finite basis. Any lin. indep. set can extend to a basis. All bases of finite-dim V have same length. $\dim V$ = length of a basis. **Sum & Direct Sum** $U_1 + U_2 = \{u_1 + u_2 \mid u_1 \in U_1, u_2 \in U_2\}$ Direct sum $(U_1 \oplus U_2)$ if each element unique. If V finite-dim, $\forall U \subset V, \exists W : V = U \oplus W$. **Rings** $R = (R, +, \cdot)$ satisfies: (P1) $(R, +)$ commutative group, (P2) $a \cdot b \in R, (a \cdot b) \cdot c = a \cdot (b \cdot c), a \cdot 1 = 1 \cdot a = a$, (P3) $a(b+c) = ab+ac, (a+b)c = ac+bc$. Example: $(\mathbb{R}^{m \times m}, +, \cdot)$. **Proper Subspaces** $U < V$ if U subspace of V, $U \neq V$. Every subspace contains 0. Setting $n-m$ components of $\mathbb{R}^n$ to 0 gives proper subspace $\cong \mathbb{R}^m$. **Orthogonality & Complements**

Lecture 2:

Linear Mappings Definition: $T : U \to V$ is linear if $\forall u_1, u_2 \in U, \lambda \in F : T(u_1 + u_2) = T(u_1) + T(u_2)$ and $T(\lambda u_1) = \lambda T(u_1)$. The set of all linear maps $U \to V$ is $L(U, V)$; if $U = V$, denote $L(U)$. Examples: zero map $0 : V \to W$, $v \mapsto 0$; identity $I : V \to V$, $Iv = v$; non-linear examples: e^x, $f(x) = x - 1$. **Kernel** $\ker(T) = \{u \in U \mid Tu = 0\}$; properties: (1) $\ker(T)$ is a subspace of U; (2) T injective $\Leftrightarrow$ $\ker(T) = \{0\}$. **Range** $\text{range}(T) = \{Tu \mid u \in U\}$; properties: (1) range is subspace of V; (2) T surjective $\Leftrightarrow$ $\text{range}(T) = V$. **Preimage** For $v' \subseteq V$, $T^{-1}(v') = \{u \in U \mid Tu \in v'\}$; if v' is subspace of V, then $T^{-1}(v')$ is subspace of U. **Rank–Nullity Theorem** For finite-dim V, $T \in L(V, W)$, $\dim(V) = \dim(\ker T) + \dim(\text{range}\, T)$. Example: $A \in \mathbb{R}^{3 \times 3}$ rank 2, nullity 1 $\Rightarrow$ $3 = 1 + 2$. **Injective $\Leftrightarrow$ Surjective $\Leftrightarrow$ Bijective** For finite-dim V, $T \in L(V, V)$: injective $\Leftrightarrow$ surjective $\Leftrightarrow$ bijective. **Applications** Range $\Rightarrow$ PCA subspace (projection); Preimage $\Rightarrow$ SVM feature space mapping. **Matrices and Linear Maps** Let $T \in L(V, W)$, bases $(v_1, \dots, v_n)$ of V, $(w_1, \dots, w_m)$ of W. For $v = \sum \lambda_i v_i$: $T(v) = \sum \lambda_i T(v_i)$, and $T(v_j) = \sum_i a_{ij} w_i$. The matrix $A = (a_{ij})$ has m rows, n columns. Denote matrix by $M(T, B, C)$. **Matrix Properties** Linearity: $M(S + T) = M(S) + M(T)$, $M(\lambda S) = \lambda M(S)$. For coordinates $\lambda = (\lambda_1, \dots, \lambda_n)^T$, $T(v) = M(T)\lambda$. Composition: $M(S \circ T) = M(S)\, M(T)$. **Matrix Algebra** Addition/scalar mult.: $A + B = B + A$, $A + (B + C) = (A + B) + C$, $c(A + B) = cA + cB$, $(c + d)A = cA + dA$, $1A = A$. Multiplication: $A(BC) = (AB)C$, $A(B + C) = AB + AC$, $(A + B)C = AC + BC$, $AB \neq BA$ in general, $AI = IA = A$. Transpose: $(A^T)^T = A$, $(A + B)^T = A^T + B^T$, $(cA)^T = cA^T$, $(AB)^T = B^T A^T$. **Invertible Maps and Matrices** $T \in L(V, W)$ invertible $\Leftrightarrow$ $\exists S \in L(W, V) : S \circ T = Id_V, T \circ S = Id_W$. S is T^{-1} (unique). Invertible $\Leftrightarrow$ bijective. **Example (Gauss Elimination)** $A = \begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix} \Rightarrow A^{-1} = \begin{bmatrix} 3/10 & -7/10 \\ -1/5 & 2/5 \end{bmatrix}$. **Inverse Matrices** $A \in F^{n \times n}$ invertible $\Leftrightarrow$ $\exists B \in F^{n \times n} : AB = BA = I$. $B = A^{-1}$. Matrix inverse represents inverse map: $M(T^{-1}) = (M(T))^{-1}$. Matrix invertible $\Leftrightarrow$ map invertible. **Inverse Properties** (1) $(A^{-1})^{-1} = A$ (2) $(AB)^{-1} = B^{-1}A^{-1}$ (3) $(A^T)^{-1} = (A^{-1})^T$ (4) A invertible $\Leftrightarrow$ $\text{rank}(A) = n$. Set of invertible matrices: $GL(n, F) = \{A \in F^{n \times n} \mid A \text{ invertible}\}$. If A invertible: $A^{-1}A = AA^{-1} = I$. Double inverse $(A^{-1})^{-1} = A$. For 2×2 matrix $\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$. Example: $\begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix}^{-1} = \frac{1}{10} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{bmatrix}$. Remark: for $n > 2$, use Gauss–Jordan elimination. **References** UCD Math 67 Notes; MIT 18.06 Handouts; Wikipedia Linear Map.

Lecture 3: Transposes, Change of Basis, Rank, Determinant

Transpose Given $A = (a_{ij}) \in F^{m \times n}$, $(A^T)_{kj} = a_{jk}$. Example $A = \begin{bmatrix} 1 & -9 & 3 \\ 1 & 2 & -2 \\ -2 & 1 & 1 \end{bmatrix} \Rightarrow A^T = \begin{bmatrix} 1 & 1 & -2 \\ -9 & 2 & 1 \\ 3 & -2 & 1 \end{bmatrix}$. If $F = \mathbb{C}$, conjugate transpose $(A^*)_{ij} = \overline{a_{ij}}$; adjoint of a matrix is A^* (preserves inner products). **Change of Basis** Let $I : V \to V$, bases $A = \{a_1, \dots, a_n\}$, $B = \{b_1, \dots, b_n\}$. Express $a_j = \sum_{i=1}^{n} t_{ij} b_i$. The change-of-basis matrix $M(I, A, B) = [t_{ij}] \in F^{n \times n}$. If bases equal, $M(I, A, A) = I_n$. Column j records coordinates of a_j in basis B. **Invertibility of Change** $M(I, A, B)$ and $M(I, B, A)$ are inverses: $T_{A \to B} = T_{B \to A}^{-1}$. **Conjugation of a Map under Basis Change** Let $T : V \to V$, $X = M(T, A, A)$, $A_0 = M(I, A, B)$, then $Y := A_0\, X\, A_0^{-1} = M(T, B, B)$. **Rank of a Matrix** Rank = maximal number of linearly independent rows or columns; equals dimensions of row/column spaces. For $A \in F^{m \times n}$, column rank = $\dim(\text{span}\{\text{cols}\})$, row rank = $\dim(\text{span}\{\text{rows}\})$. **Rank Properties** (1) Row rank = column rank. (2) $\text{rank}(A) = \text{rank}(A^T)$. (3) For $T \in L(V, W)$, any $M(T)$ has basis-independent rank. (4) $\text{rank}(M(T)) = \dim(\text{range}(T))$. **Determinant (as a Mapping)** A determinant is $d : F^{n \times n} \to F$ such that: (i) *multilinear in columns*: $\det(\dots, a_i' + a_i'', \dots) = \det(\dots, a_i', \dots) + \det(\dots, a_i'', \dots)$ and $\det(\dots, \lambda a_i, \dots) = \lambda \det(\dots, a_i, \dots)$; (ii) *alternating*: two equal columns $\Rightarrow$ $\det(A) = 0$; (iii) *normed*: $\det(I) = 1$. **Determinant Properties** Basis independent; exists and unique. Further: $\det(cA) = c^n \det(A)$; $\det(AB) = \det(A)\det(B)$; $\det(A^T) = \det(A)$; if A invertible, $\det(A^{-1}) = \det(A)^{-1} \neq 0$; if A is upper triangular, $\det(A)$ is product of diagonal entries; swapping two rows/cols flips sign; in general $\det(A + B) \neq \det(A) + \det(B)$. **Leibniz Formula** $\det(A) = \sum_{\sigma \in S_n} \text{sign}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}$. **Laplace (Cofactor) Expansion** Along row 1: $\det(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det(A_{1j})$; in general $\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$. **Special Cases** $n = 1$: $\det([a]) = a$. $n = 2$: $\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$. $n = 3$: $\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$. **Geometric Intuition** In 2D, columns a_1, a_2 form a parallelogram with area $|\det(A)|$; flipping column order flips sign. In 3D, $|\det(A)|$ is volume of parallelotope from columns. For diagonalizable case on new axes, $\det(A) = \prod_{i=1}^{n} \lambda_i$. **Change of Variables for Integrals** If $\Omega \subset \mathbb{R}^n$ open, $\sigma : \Omega \to \mathbb{R}^n$ differentiable, $f : \sigma(\Omega) \to \mathbb{R}$, then $\int_{\sigma(\Omega)} f(y)\, dy = \int_{\Omega} f(\sigma(x)) \left| \det(\sigma'(x)) \right| dx$. **Lecture 4:**

Eigenvalues, Characteristic Polynomials, and Trace

Definition. $T : V \to V$ is linear. $\lambda \in F$ is an eigenvalue if $\exists v \neq 0$ with $Tv = \lambda v$. Then v is an eigenvector. The eigenspace is $E(\lambda, T) = \ker(T - \lambda I)$.

Remarks. (1) $T(v) = \lambda v$ scales v without changing direction. (2) Over $\mathbb{R}$ some T (e.g. rotations) have no real eigenvectors, but over $\mathbb{C}$ every linear operator has at least one eigenvalue (algebraically closed field). (3) If v is eigenvector then any nonzero av also is. (4) Eigenvectors of distinct eigenvalues are linearly independent. (5) Vectors in the same eigenspace may be dependent. (6) We can pick a maximal independent subset of each eigenspace as a basis. (7) $E(\lambda, T)$ is a subspace.

Equivalent Characterizations. For finite-dim. V, λ is an eigenvalue $\Leftrightarrow$ $T - \lambda I$ is not injective $\Leftrightarrow$ not surjective $\Leftrightarrow$ not bijective.

Direct Sum of Eigenspaces. If $\lambda_1, \dots, \lambda_m$ distinct, then $E(\lambda_1, T) + \dots + E(\lambda_m, T)$ is direct and $\sum \dim E(\lambda_i, T) \leq \dim V$.

Existence. Every $T : V \to V$ over $\mathbb{C}$ has an eigenvalue. Proof: by linear dependence of $\{v, Tv, \dots, T^n v\}$ and factorization of a polynomial over $\mathbb{C}$.

Example. $A = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$ has eigenvalue $\lambda = 2$, $(A - 2I)x = 0$ gives $\ker(A - 2I) = \{(x_1, 0)^T\}$, not injective.

Characteristic Polynomial. $P_A(t) = \det(A - tI)$; roots are eigenvalues. Example: $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, $P_A(t) = t^2 - t(a_{11} + a_{22}) + (a_{11}a_{22} - a_{12}a_{21})$.

Properties. (1) $\deg P_A = n$. (2) $P_{UAU^{-1}}(t) = P_A(t)$. (3) Roots are eigenvalues. (4) A invertible $\Leftrightarrow$ 0 not eigenvalue. (5) λ eigenvalue $\Rightarrow$ λ^2 eigenvalue of A^2, and if A invertible, λ^{-1} eigenvalue of A^{-1}.

Multiplicity. Algebraic multiplicity = root multiplicity of P_A; geometric multiplicity = $\dim E(\lambda, A)$; geometric $\leq$ algebraic.

Example. $A = \begin{bmatrix} 4 & 2 \\ 1 & 3 \end{bmatrix}$, $P_A(t) = t^2 - 7t + 10 = (t - 2)(t - 5)$, eigenvalues $2, 5$.

Rotation Example. $R(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$, $P_{R(\theta)}(t) = t^2 - 2\cos\theta\, t + 1$, eigenvalues $t = \cos\theta \pm i \sin\theta$, not real when $\theta \notin \pi\mathbb{Z}$.

Trace. $\text{tr}(A) = \sum_i a_{ii}$. Properties: (1) linearity $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$, $\text{tr}(cA) = c\,\text{tr}(A)$. (2) Cyclic $\text{tr}(AB) = \text{tr}(BA)$. (3) Basis-independent. (4) For diagonalizable A, $\text{tr}(A) = \sum \lambda_i$, $\det(A) = \prod \lambda_i$. (5) If $P_A(t) = t^n + a_{n-1}t^{n-1} + \dots + a_0$, then $\text{tr}(A) = -a_{n-1}$.

Example. $R(\theta)$ as above: $\text{tr}(R(\theta)) = 2\cos\theta$, $P_{R(\theta)}(t) = t^2 - 2\cos\theta t + 1$, eigenvalues $\cos\theta \pm i \sin\theta$.

Applications.

(1) PCA: compute covariance $\Sigma = \frac{1}{m-1} X^T X$, eigenvectors with largest eigenvalues give principal components.

(2) Spectral Clustering: form Laplacian $L = D - W$, compute eigenvectors of smallest nonzero eigenvalues, cluster via k-means.

References. Wikipedia: Eigenvalues and Eigenvectors; PCA; Spectral Clustering.

Lecture 5: Diagonalization, Triangular Matrices, Metric Spaces, Norms

Diagonalization Definition: $T \in L(V)$ is diagonalizable if $\exists$ a basis of V with $M(T) = \text{diag}(\lambda_1, \dots, \lambda_n)$. Example: any diagonal matrix D is diagonalizable. Nice property: eigenvectors form the basis. Proposition (equivalences for finitedim V): (P1) A diagonalizable. (P2) char poly splits into linear factors and each root's algebraic multiplicity equals its geometric multiplicity. (P3) If $\lambda_1, \dots, \lambda_k$ are distinct then $V = E(A, \lambda_1) \oplus \dots \oplus E(A, \lambda_k)$. Symmetric matrices are diagonalizable. Not all matrices are diagonalizable; triangular form is the next best.

Triangular Matrices Upper triangular: zeros below main diagonal; lower triangular: zeros above. If all offdiagonal are zero $\Rightarrow$ diagonal. Proposition: For $T \in L(V)$ and basis $B = \{v_1, \dots, v_n\}$, $M(T, B)$ upper triangular $\Leftrightarrow$ $T(v_j) \in \text{span}\{v_1, \dots, v_j\}$ for each j. Consequence: eigenvalues of an upper triangular matrix are its diagonal entries. In a complex field, every matrix is similar to an upper triangular matrix. If T has upper triangular form in some basis, then those diagonal entries are precisely the eigenvalues.

Metric Spaces Definition: metric $d : X \times X \to \mathbb{R}$ satisfies (P1) $d(u, v) > 0$ if $u \neq v$ and $d(u, u) = 0$ (P2) symmetry $d(u, v) = d(v, u)$ (P3) triangle inequality $d(u, v) \leq d(u, w) + d(w, v)$.

Sequences Cauchy: (x_n) is Cauchy if $\forall \varepsilon > 0\, \exists N\, \forall n, m > N : d(x_n, x_m) < \varepsilon$. Every convergent sequence is Cauchy. Convergence: $x_n \to x$ if $\forall \varepsilon > 0\, \exists N\, \forall n > N : d(x_n, x) < \varepsilon$. Example: on $(0, 1)$ the sequence $x_n = 1/n$ is Cauchy but not convergent in that space; on $[0, 1]$ it converges to 0.

Complete A metric space is complete if every Cauchy sequence converges (e.g., $\mathbb{R}$ with standard metric).

Epsilon Ball $B_\varepsilon(u) = \{x \in X \mid d(x, u) < \varepsilon\}$; open sets are those for which every point contains some B_ε inside the set.

Closed & Open Closed: every Cauchy sequence in A converges to a point in A. Open: $\forall a \in A\, \exists \varepsilon > 0 : B_\varepsilon(a) \subset A$. A set can be neither, or both (clopen), depending on the space.

Interior/Closure/Boundary Interior A°: all interior points. Closure $\overline{A}$: points approximable by sequences in A; $\overline{A}$ is closed and equals $A \cup \partial A$. Boundary $\partial A = \overline{A} \setminus A^\circ$. Dense: A dense in X if $\forall x \in X, \forall \varepsilon > 0 : B_\varepsilon(x) \cap A \neq \emptyset$ (e.g., $\mathbb{Q}$ dense in $\mathbb{R}$). Bounded: $\exists D > 0$ s.t. $d(u, v) \leq D$ for all $u, v \in A$.

Norms A norm $\|\cdot\| : V \to \mathbb{R}$ satisfies (P1) $\|\lambda x\| = |\lambda| \|x\|$ (P2) $\|x + y\| \leq \|x\| + \|y\|$ (P3) $x = 0 \Rightarrow \|x\| = 0$ (P4) $\|x\| = 0 \Rightarrow x = 0$. Seminorm: satisfies (P1)(P3). Intuition: $\|x\|$ is the length of x. Examples on $\mathbb{R}^d$: Euclidean $\|x\|_2 = (\sum_i x_i^2)^{1/2}$; Manhattan $\|x\|_1 = \sum_i |x_i|$.

p-Norms For $1 \leq p < \infty$: $\|x\|_p = (\sum_i |x_i|^p)^{1/p}$ is a norm; unit ball $B_p = \{x \mid \|x\|_p \leq 1\}$. Infinity norm: $\|x\|_\infty = \max_i |x_i|$ is a norm. "Zeronorm": $\|x\|_0 = \#\{i : x_i \neq 0\}$ is not a norm (fails homogeneity).

Key Takeaways Diagonalizable $\Leftrightarrow$ eigenbasis exists; over $\mathbb{C}$ any matrix triangularizable; eigenvalues of triangular matrices are diagonal entries; completeness concerns Cauchy convergence; norms induce common metrics, with $\|\cdot\|_p$ families central in analysis and ML.

Lecture 6: Norm and Function Spaces (Feb 5, 2025)

Equivalence of Norms on $\mathbb{R}^n$ Theorem: any two norms $\|\cdot\|_a, \|\cdot\|_b$ on $\mathbb{R}^n$ are equivalent: $\exists\, \alpha, \beta > 0 : \ \alpha\|x\|_a \leq \|x\|_b \leq \beta\|x\|_a$, $\forall x$. Proof idea: compare any norm to $\|\cdot\|_\infty$. Lemma 1: $\exists C_1 > 0 : \ \|x\| \leq C_1\|x\|_\infty$ (expand $x = \sum x_i e_i$). Lemma 2: $\exists C_2 > 0 : \ \|x\|_\infty \leq C_2\|x\|$ (continuity of $f(x) = \|x\|$ on compact $S = \{x : \|x\|_\infty = 1\}$ gives $\min f > 0$).

Convex Sets as Unit Balls of Norms Definitions: convex C if $bx + (1 - b)y \in C$ for $b \in [0, 1]$; symmetric if $x \in C \Rightarrow -x \in C$. Gauge $p(x) = \inf\{t > 0 : \ x \in tC\}$. Theorem: if $C \subset \mathbb{R}^d$ is closed, symmetric, convex, nonempty interior then p is a seminorm; if C bounded then p is a norm and $C = \{x : \ p(x) \leq 1\}$. Conversely, the unit ball $\{x : \ \|x\| \leq 1\}$ of any norm is bounded, symmetric, closed, convex, with nonempty interior. Proof ideas: show $B_\varepsilon(0) \subset C$ then p well-defined; verify $p(0) = 0$, homogeneity $p(\alpha x) = |\alpha| p(x)$, triangle inequality by convexity; $p(x) = 0 \Rightarrow x = 0$ when C bounded.

Normed Function Spaces $C_b(T) = \{f : T \to \mathbb{R} \mid f \text{ continuous and bounded}\}$ with $\|f\|_\infty = \sup_{t \in T} |f(t)|$. Then $(C_b(T), \|\cdot\|_\infty)$ is a Banach space (complete metric comes from uniform convergence). Banach space: a normed space complete under the metric $d(x, y) = \|x - y\|$.

Tikhonov Regularization (ML) Least squares with noisy $b = \hat{b} + e$: solve $\min_x \|Ax - b\|^2 + \|L_\mu x\|^2$ as a stable proxy for $\min_{\hat{x}} \|A\hat{x} - \hat{b}\|^2$; μ controls regularization, common $L = I$. Imaging model $Y = Hf + n$; solve $\min_f \ \|Y - Hf\|_2^2 + \alpha\|Cf\|_2^2$ (e.g. C high-pass) for super-resolution / inverse problems.

Space of Differentiable Functions On $[a, b] \subset \mathbb{R}$ define $C^1([a, b]) = \{f : [a, b] \to \mathbb{R} \mid f \text{ cont.}, f' \text{ cont.}\}$. Useful norms: $\|f\| := \sup_{t \in [a,b]} \max\{|f(t)|, |f'(t)|\}$ or $\|f\| := \|f\|_\infty + \|f'\|_\infty$; with either, $C^1([a, b])$ is Banach.

Banach Spaces in ML (Insights) Feature spaces: $C([a, b])$ with sup-norm for function approximation. Stability: Lipschitz continuity via norms. Optimization: fixed-point/contraction arguments for convergence of iterative methods. Dual spaces & regularization: Lasso (ℓ_1), Ridge (ℓ_2).

Applications (Banach-Space View) RKBS: kernel methods beyond Hilbert setting. Neural nets: expressivity and universal approximation on function spaces. Metric learning/embeddings: distances induced by norms. Inverse problems: variational formulations in Banach spaces. Sparse learning & compressed sensing: ℓ_p $(0 < p \leq 1)$ promote sparsity.

Lecture 7: Computational Foundations of Artificial Intelligence

Inner Product and Hilbert Spaces

Definition: A mapping $\langle\cdot,\cdot\rangle : V\times V\to F$ is an inner product if it satisfies: (P1) $\langle x_1+x_2,y\rangle=\langle x_1,y\rangle+\langle x_2,y\rangle$; (P2) $\langle\lambda x,y\rangle=\lambda\langle x,y\rangle$; (P3) $\langle x,y\rangle=\overline{\langle y,x\rangle}$; (P4) $\langle x,x\rangle\ge 0$; (P5) $\langle x,x\rangle=0\Leftrightarrow x=0$.
Examples: Euclidean inner product on $\mathbb{R}^n$: $\langle x,y\rangle=\sum_{i=1}^n x_iy_i$; On $\mathbb{C}^n$: $\langle x,y\rangle=\sum x_i\bar{y}_i$; For $C([a,b])$: $\langle f,g\rangle=\int_a^b f(t)g(t)\,dt$.
A vector space with a norm is a normed space. If every Cauchy sequence converges, it is a Banach space. A vector space with an inner product is a pre-Hilbert space; if complete, it is a Hilbert space. Define $\|x\|=\sqrt{\langle x,x\rangle}$; then $\|\cdot\|$ is a norm induced by $\langle\cdot,\cdot\rangle$.
Conversely, if a norm is given, $d(x,y)=\|x-y\|$ defines a metric. Not every norm arises from an inner product.

Orthogonal Basis and Projection

Two vectors $v_1,v_2\in V$ are orthogonal if $\langle v_1,v_2\rangle=0$. Denote $v_1\perp v_2$. For subsets $V_1,V_2\subseteq V$, they are orthogonal if $\langle v_1,v_2\rangle=0$ for all $v_1\in V_1, v_2\in V_2$. A set $\{v_1,\dots,v_n\}$ is orthonormal if each pair is orthogonal and $\|v_i\|=1$. For $S\subseteq V$, define the orthogonal complement $S^\perp=\{v\in V\mid v\perp s,\forall s\in S\}$.

Orthogonal Projection

A map $A\in L(V)$ is a projection if $A^2=A$. For a finite-dimensional subspace $U\subset H$, there exists a projection $P_U:H\to U$ with $\ker(P_U)=U^\perp$. Define $P_U(w)=\sum_{i=1}^n\frac{\langle w,v_i\rangle}{\|v_i\|^2}v_i$ for orthogonal basis $\{v_i\}$. If $\{u_i\}$ is orthonormal, any $v\in V$ can be written $v=\sum_{i=1}^n\langle v,u_i\rangle u_i$.

Gram–Schmidt Orthogonalization

Transforms any basis $\{v_1,\dots,v_n\}$ into an orthonormal basis $\{u_1,\dots,u_n\}$.
Step 1: $u_1=\frac{v_1}{\|v_1\|}$, $U_1=\mathrm{span}\{u_1\}$.
Step k: Given $u_1,\dots,u_{k-1}$, compute $\tilde u_k=v_k-P_{U_{k-1}}(v_k)$, then normalize $u_k=\tilde u_k/\|\tilde u_k\|$.
In practice, use Householder reflections for numerical stability.

Orthogonal Matrices

$Q\in\mathbb{R}^{n\times n}$ is orthogonal if its columns are orthonormal ($Q^TQ=I$). If $Q\in\mathbb{C}^{n\times n}$ with orthonormal columns under the complex inner product, it is unitary ($Q^*Q=I$).
Examples: Identity $\begin{bmatrix}1&0\\0&1\end{bmatrix}$; Reflection $\begin{bmatrix}1&0\\0&-1\end{bmatrix}$; Permutation $\begin{bmatrix}0&1\\1&0\end{bmatrix}$; Rotation $\begin{bmatrix}\cos\theta&-\sin\theta\\\sin\theta&\cos\theta\end{bmatrix}$; 3D rotation $\begin{bmatrix}1&0&0\\0&\cos\theta&-\sin\theta\\0&\sin\theta&\cos\theta\end{bmatrix}$.
Properties: Columns orthogonal $\Leftrightarrow$ rows orthogonal; $Q^{-1}=Q^T$; $\|Qv\|=\|v\|$ (isometry); $\langle Qu,Qv\rangle=\langle u,v\rangle$; $|\det(Q)|=1$. Analogously, for unitary U, $U^{-1}=U^*$.

Isometries

For $S\in L(V)$: (1) S is an isometry $\Leftrightarrow\|Sv\|=\|v\|$ for all v. (2) There exists an orthonormal basis of V such that the matrix of S has block-diagonal form with each block B_i being either ±1 or a 2×2 rotation matrix.

End of Lecture 7

Lecture 8: Positive Definite Matrices, Variational Characterization of Eigenvalues

Symmetric / Hermitian

Definition: $A\in\mathbb{R}^{n\times n}$ symmetric if $A=A^T$; $A\in\mathbb{C}^{n\times n}$ Hermitian if $A=A^*$.
Proposition: If A is Hermitian, all eigenvalues are real and eigenvectors for distinct eigenvalues are orthogonal. Sketch: $Ax=\lambda x\Rightarrow\langle Ax,x\rangle=\langle x,Ax\rangle=\bar\lambda\langle x,x\rangle=\lambda\langle x,x\rangle\Rightarrow\lambda\in\mathbb{R}$; for $(\lambda_1,x_1),(\lambda_2,x_2)$, $\lambda_1\langle x_1,x_2\rangle=\lambda_2\langle x_1,x_2\rangle\Rightarrow(\lambda_1\ne\lambda_2)\Rightarrow\langle x_1,x_2\rangle=0$.
Self-adjoint operator: $T\in L(V)$ is self-adjoint if $\langle Tu,w\rangle=\langle u,Tw\rangle$. On $\mathbb{R}^n$: symmetric; on $\mathbb{C}^n$: Hermitian. A self-adjoint T has a real eigenvalue.

Spectral Theorem

Real case: For $A=A^T$, $\exists$ orthogonal Q and diagonal $D=\mathrm{diag}(\lambda_1,\dots,\lambda_n)$ with $A=QDQ^T=\sum_{i=1}^n\lambda_iq_iq_i^T$.
Complex case: For $A=A^*$, $\exists$ unitary U and real diagonal D with $A=UDU^*$. Columns of Q or U form an orthonormal eigenbasis.

Positive (Semi)Definite

Definitions: A PD if $x^TAx>0$ for all $x\ne0$; PSD if $x^TAx\ge0$ for all x. Gram matrix: $G=(\langle v_i,v_j\rangle)_{ij}$ is Hermitian (symmetric over $\mathbb{R}$) and PSD.
Equivalences for Hermitian A: A PSD $\Leftrightarrow$ all eigenvalues $\ge0\Leftrightarrow\langle x,x\rangle_A:=x^*Ax$ is positive semidefinite form (an inner product on a subspace) $\Leftrightarrow A$ is a Gram matrix. Over $\mathbb{C}$, PD $\Rightarrow$ Hermitian; over $\mathbb{R}$, there exist PD but non-symmetric matrices.

Square Root of PSD

If $A=A^T\succeq0$, then by spectral theorem $A=QDQ^T$ with $D\ge0$. Define $A^{1/2}=QD^{1/2}Q^T$ where $D^{1/2}=\mathrm{diag}(\sqrt{\lambda_i})$. Then $A^{1/2}\succeq0$ and $(A^{1/2})^2=A$.

Rayleigh Quotient & Extremal Eigenvalues

Rayleigh quotient: $R_A(x)=\frac{x^TAx}{x^Tx}$ for $x\ne0$. For symmetric A with eigenvalues $\lambda_1\le\cdots\le\lambda_n$ and orthonormal eigenvectors v_i: $\min_{\|x\|=1}R_A(x)=\lambda_1$ attained at v_1; $\max_{\|x\|=1}R_A(x)=\lambda_n$ attained at v_n. Constrained step: $\min_{\|x\|=1,x\perp v_1}R_A(x)=\lambda_2$ (and so on).

Courant–Fischer (Min–Max)

For symmetric A and $k=1,\dots,n$: $\lambda_k=\min_{\substack{U\subset\mathbb{R}^n\\\dim U=k}}\max_{x\in U\setminus\{0\}}R_A(x)=\max_{\substack{U\subset\mathbb{R}^n\\\dim U=n-k+1}}\min_{x\in U\setminus\{0\}}R_A(x)$.

Notes

Orthogonal rank-1 projectors $q_iq_i^T$ decompose symmetric matrices; PSD cones are closed under sums and congruences $A\mapsto S^TAS$; unitary/orthogonal similarity preserves spectrum, trace, determinant.

End of Lecture 8

Singular Value Decomposition

For $A\in\mathbb{R}^{m\times n}$ of rank r, there exist orthogonal $U\in\mathbb{R}^{m\times m}$, $V\in\mathbb{R}^{n\times n}$ and diagonal $\Sigma\in\mathbb{R}^{m\times n}$ with singular values $\sigma_1\ge\cdots\ge\sigma_r>0$ on the diagonal such that $A=U\Sigma V^T$. Columns u_i of U are left singular vectors; columns v_i of V are right singular vectors; $Av_i=\sigma_iu_i$, $A^Tu_i=\sigma_iv_i$.

Construction via A^TA

Let $B=A^TA\in\mathbb{R}^{n\times n}$. Then B is symmetric PSD; take an orthonormal eigenbasis $\{x_i\}$ with $Bx_i=\lambda_ix_i$, $\lambda_i\ge0$. Set $\sigma_i=\sqrt{\lambda_i}$, $v_i=x_i$, and $u_i=Ax_i/\sigma_i$ for $\sigma_i>0$; complete U,V to orthogonal matrices; then $A=U\Sigma V^T$.

Relationships

Eigenvalues: nonzero eigenvalues of A^TA and AA^T equal σ_i^2. Ranges: $\mathcal{R}(A)=\mathrm{span}\{u_1,\dots,u_r\}$; $\mathcal{R}(A^T)=\mathrm{span}\{v_1,\dots,v_r\}$. Null spaces: $\mathcal{N}(A)=\mathrm{span}\{v_{r+1},\dots,v_n\}$; $\mathcal{N}(A^T)=\mathrm{span}\{u_{r+1},\dots,u_m$

SVD vs. Eigen-Decomposition

SVD exists for any real matrix; singular values are real and nonnegative; U,V are orthogonal. If A is symmetric, its SVD aligns with eigen-decomposition: eigenpairs (λ_i,v_i) give singular pairs $(|\lambda_i|,v_i)$; for $A\succeq0$, $\sigma_i=\lambda_i$ and $U=V$.

Matrix Norms

Max-entry (infinity) norm: $\|A\|_{\max}=\max_{i,j}|a_{ij}|$. One norm: $\|A\|_1=\sum_{i,j}|a_{ij}|$. Frobenius: $\|A\|_F=\sqrt{\sum_{i,j}a_{ij}^2}=\sqrt{\mathrm{tr}(A^TA)}=\sqrt{\sum_i\sigma_i^2}$. Spectral (operator 2-) norm: $\|A\|_2=\sigma_{\max}(A)=\max_{x\ne0}\|Ax\|_2/\|x\|_2$.

Best Rank-k Approximation

Let $A=U\Sigma V^T$ with $\sigma_1\ge\cdots\ge\sigma_p$ where $p=\min\{m,n\}$. Define $A_k=\sum_{i=1}^k\sigma_iu_iv_i^T=U_k\Sigma_kV_k^T$. Then Eckart–Young–Mirsky: for any rank-k matrix B, $\|A-A_k\|_F\le\|A-B\|_F$ and $\|A-A_k\|_2\le\|A-B\|_2$; moreover $\|A-A_k\|_F=\sqrt{\sum_{i>k}\sigma_i^2}$, $\|A-A_k\|_2=\sigma_{k+1}$.

Moore–Penrose Pseudoinverse

For $A=U\Sigma V^T$, define $\Sigma^\dagger$ by $(\Sigma^\dagger)_{ii}=1/\sigma_i$ if $\sigma_i>0$ and 0 otherwise; then $A^\dagger=V\Sigma^\dagger U^T$. Characterization: $AA^\dagger A=A$, $A^\dagger AA^\dagger=A^\dagger$, $(AA^\dagger)^T=AA^\dagger$, $(A^\dagger A)^T=A^\dagger A$. Least-squares: $x^\star=A^\dagger b$ is the minimum-norm solution of $\min_x\|Ax-b\|_2$.

Quick Example (2-by-2)

If $A=\begin{bmatrix}3&0\\0&1\end{bmatrix}$, then $U=I$, $V=I$, $\Sigma=\mathrm{diag}(3,1)$; $A_1=3\,e_1e_1^T$, $\|A-A_1\|_2=1$, $\|A-A_1\|_F=1$.

Practice Reminders

Order singular values descending; drop small σ_i for compression/denoising; spectral norm equals largest singular value; condition number $\kappa_2(A)=\sigma_{\max}/\sigma_{\min}$ (for full column-rank square A).