STAT31430 Applied Linear Algebra

Notes for Exam Preparation

Seung Chul Lee

A Note of Caution:
These notes are created solely for my personal use and do not accurately represent the pedagogy or the material covered for the course named above. I have taken this class during Fall 2022, which may or may not have identical structure in future quarters. All errors contained are my own.

Contents

1 Definitions

1.1 Matrix Basics

- Linearly independent:

$$\forall \alpha_1, \dots, \alpha_n \in \mathbb{K}, \;\; \alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_n x_n = 0 \Rightarrow \alpha_1 = \cdots = \alpha_n = 0$$

- Orthonormal:

$$\langle y_i, y_j \rangle = 0, \;\; \forall i \neq j, \;\; \|y_i\| = 1, \;\; \forall i$$

- Kernel:
For $A \in \mathcal{M}_{n,p}(\mathbb{K})$,

$$\ker(A) = \{x \in \mathbb{K}^p : Ax = 0\} \subset \mathbb{K}^p$$

- Image:
For $A \in \mathcal{M}_{n,p}(\mathbb{K})$,

$$\text{im}(A) = \{Ax : x \in \mathbb{K}^p\} \subset \mathbb{K}^n$$

- Dimension:
The number of elements in a spanning linearly independent set of vectors, i.e., a basis.

- Rank:
  $\operatorname{rank} A = \dim(\operatorname{im} A)$
- Trace:
  $A = (a_{ij})_{1\le i,j\le n}$, $\operatorname{tr}(A) = \sum_{i=1}^n a_{ii}$
- Permutation:
  $\sigma : \{1,\dots,n\} \to \{1,\dots,n\}$ such that it is both injective and surjective, i.e., bijective.
- Determinant:
  For $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{K})$,
  $$\det(A) = \sum_{\sigma\in S_n} \varepsilon(\sigma) \prod_{i=1}^n a_{i,\sigma(i)}$$
  where $\varepsilon(\sigma) = (-1)^{p(\sigma)}$, the signature of σ, and $p(\sigma) = \sum_{1\le i\le j\le n} \operatorname{Inv}_\sigma(i,j)$, the inversion counter.
- Adjoint/conjugate transpose/Hermitian transpose:
  For $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{C})$, $A^* \in \mathcal{M}_n(\mathbb{C})$ given by $A^* = \overline{A^\top} = (\overline{a_{ji}})$
- $A \in \mathcal{M}_n(\mathbb{C})$ is
  - self-adjoint (or Hermitian) if $A = A^*$.
  - unitary if $A^{-1} = A^*$, i.e., $AA^* = A^*A = I$.
  - normal if $AA^* = A^*A$.
- $A \in \mathcal{M}_n(\mathbb{R})$ is
  - symmetric (= self-adjoint) if $A = A^\top$.
  - orthogonal (= unitary) if $A^{-1} = A^\top$, i.e., $AA^\top = A^\top A = I$.
  - normal if $AA^\top = A^\top A$.

1.2 Spectral Theory

- Characteristic polynomial:
  For $A \in \mathcal{M}_n(\mathbb{C})$,
  $$P_A : \mathbb{C} \to \mathbb{C}, \quad P_A(\lambda) = \det(A - \lambda I)$$
- Eigenvalues:
  The roots of the characteristic polynomial, i.e.,
  $$\lambda \in \mathbb{C} \text{ s.t. } \det(A - \lambda I) = 0$$
- Spectrum:
  $$\sigma(A) = \{\lambda \in \mathbb{C} : \det(A - \lambda I) = 0\}$$
- Algebraic multiplicity: The largest k such that
  $$P_A(z) = (z - \lambda)^k Q(z)$$
- Eigenvector:
  A nonzero vector $x \in \mathbb{C}^n$ s.t. $Ax = \lambda x$ for some $\lambda \in \sigma(A)$.
- Spectral radius:
  For $A \in \mathcal{M}_n(\mathbb{C})$, the spectral radius of A is
  $$\rho(A) := \max_{\lambda\in\sigma(A)} |\lambda|$$

- Eigenspace:
  For $\lambda \in \sigma(A),\ A \in \mathcal{M}_n(\mathbb{C})$, the eigenspace of $\mathbb{C}^n$ associated to λ is
$$E_\lambda := \ker(A - \lambda I) = \{x \in \mathbb{C}^n : Ax = \lambda x\}$$
- Generalized eigenspace:
$$F_\lambda := \bigcup_{k \geq 1} \ker(A - \lambda I)^k$$
- Matrix polynomial:
  For polynomial $P \in \mathbb{C}[x] = \{a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d : a_1, \ldots, a_d \in \mathbb{C}, d \geq 0\}$ and $A \in \mathcal{M}_n(\mathbb{C})$, then $P : \mathcal{M}_n(\mathbb{C}) \to \mathcal{M}_n(\mathbb{C})$ determined by
$$P(A) = a_0 I + a_1 A + a_2 A^2 + \cdots + a_d A^d$$
  is the corresponding matrix polynomial.
- Direct sum:
  If $F_1, \ldots, F_p \subset \mathbb{C}^n$ are subspaces, we write
$$\mathbb{C}^n = \bigoplus_{i=1}^{p} F_i$$
  if any $x \in \mathbb{C}^n$ can be written uniquely as $x = \sum_{i=1}^p x_i,\ \ x_i \in F_i,\ \ 1 \leq i \leq p$.
- Reduction to triangular form:
  $A \in \mathcal{M}_n(\mathbb{C})$ can be reduced to upper (lower) triangular form if $\exists P \in \mathbb{M}_n(\mathbb{C})$ nonsingular and an upper (lower) triangular matrix T s.t. $A = PTP^{-1}$.
- Similar matrices:
  A and T are similar matrices if $\exists P$ invertible s.t. $A = PTP^{-1}$.
- Diagonalizability:
  A is said to be diagonalizable if $A = PDP^{-1}$ for suitable P and D diagonal.
- Rayleigh quotient:
  $A \in \mathcal{M}_n(\mathbb{C})$ self-adjoint (Hermitian). The Rayleigh quotient is the function $R_A : \mathbb{C}^n \setminus \{0\} \to \mathbb{R}$ defined by
$$R_A(x) = \frac{\langle Ax, x\rangle}{\langle x, x\rangle}$$

1.3 Singular Value Decomposition

- Positive definiteness:
  $A \in \mathcal{M}_n(\mathbb{C})$: Hermitian is positive definite if every eigenvalue $\lambda \in \sigma(A)$ satisfies $\lambda > 0$.
- Positive semidefiniteness:
  $A \in \mathcal{M}_n(\mathbb{C})$: Hermitian is positive definite if every eigenvalue $\lambda \in \sigma(A)$ satisfies $\lambda \geq 0$.
- Singular values:
  The singular values of $A \in \mathcal{M}_{m,n}(\mathbb{C})$ are the square roots of the eigenvalues of A^*A.
- Moore-Penrose pseudoinverse:
  Given a matrix $A \in \mathcal{M}_{m,n}(\mathbb{C})$ with SVD $A = V\tilde{\Sigma}U^*$, the pseudoinverse $A^\dagger \in \mathcal{M}_{n,m}(\mathbb{C})$ is the matrix
$$A^\dagger = U\tilde{\Sigma}^\dagger V^*, \ \ \tilde{\Sigma}^\dagger = \begin{bmatrix} \Sigma^{-1} & 0 \\ 0 & 0 \end{bmatrix} \in \mathcal{M}_{n,m}(\mathbb{R})$$

- Fundamental spaces of matrices:

$$A \in \mathcal{M}_{m,n}(\mathbb{R}) = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} \tilde{a}_1 \\ \vdots \\ \tilde{a}_m \end{bmatrix}$$

  - Column space: $\mathrm{col}(A) = \mathrm{span}\{a_1, \ldots, a_n\}$
  - Kernel or null space: $\ker(A) = \mathrm{null}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$
  - Row space: $\mathrm{row}(A) = \mathrm{span}\{\tilde{a}_1, \ldots, \tilde{a}_n\} = \mathrm{col}(A^\top)$
  - Left null space: $\ker(A^\top) = \{y \in \mathbb{R}^m : A^\top y = 0\}$

1.4 Matrix Norms

- Norm:
  A norm $\|\cdot\| : \mathbb{K}^d \to [0, \infty)$ is a function satisfying
  i. positive definiteness: $\|x\| \geq 0$ with $\|x\| = 0$ iff $x = 0$, $\forall x \in \mathbb{K}^d$.
  ii. homogeneity: $\|\lambda x\| = |\lambda|\|x\|$, $\forall x \in \mathbb{K}^d, \lambda \in \mathbb{K}$
  iii. triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$, $\forall x, y \in \mathbb{K}^d$
- Inner product:
  $\langle \cdot, \cdot, \rangle$ an inner product on $V \times V \to \mathbb{C}$ is a map satisfying
  i. $\langle v, v\rangle \geq 0$, $\forall v \in V$
  ii. $\langle \alpha_1 w_1 + \alpha_2 w_2, v\rangle = \alpha_1 \langle w_1, v\rangle + \alpha_2 \langle w_2, v\rangle$, $w_1, w_2, v \in V, \alpha_1, \alpha_2 \in \mathbb{C}$
  iii. $\langle v, v\rangle = 0 \iff v = 0 \in V$
  iv. $\langle v, w\rangle = \langle w, v\rangle$, $\forall v, w \in V$
- Euclidean norm:
$$\|x\|_2 = \left(\sum_{i=1}^{d} |x_i|^2\right)^{\frac{1}{2}}$$
- p-norm:
$$\|x\|_p = \left(\sum_{i=1}^{d} |x_i|^p\right)^{\frac{1}{p}}, \quad 1 \leq p \leq \infty$$
- Weighted p-norm:
$$\|x\|_{p,w} = \left(\sum_{i=1}^{d} w_i |x_i|^p\right)^{\frac{1}{p}}, \quad w = (w_1, \ldots, w_d), \quad w_i > 0, \quad \forall i = 1, \ldots, d$$
- Norm using matrix:
  For A: real, positive definite, symmetric matrix,
$$\|x\|_A = \left(x^\top A x\right)^{\frac{1}{2}} = \left(\sum_{i,j=1}^{n} a_{ij} x_i x_j\right)^{\frac{1}{2}}$$
  defines a norm.
- ∞-norm:
$$\|x\|_\infty = \max_{1 \leq i \leq d} |x_i| \left(= \lim_{p \to \infty} \|x\|_p\right)$$

- **Frobenius norm** (Euclidean, Schur norm):
  $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{K})$,
$$\|A\|_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^2 \right)^{\frac{1}{2}}$$
- (Hölder) q-norm ($q \geq 1$):
$$\|A\|_{\ell^q} = \|A\|_{H,q} = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij}|^q \right)^{\frac{1}{q}}$$
- Infinity norm (∞-norm):
$$\|A\|_{\ell^\infty} = \|A\|_{H,\infty} = \max_{1 \leq i,j \leq n} |a_{ij}|$$
- **Matrix norm**:
  A norm $\|\cdot\|$ on $\mathcal{M}_n(\mathbb{K})$ is a matrix norm if for all $A, B \in \mathcal{M}_n(\mathbb{K})$, $\|AB\| \leq \|A\|\|B\|$
- **Subordinate (induced) norm**:
  Let $\|\cdot\|_*$ be a vector norm on $\mathbb{K}^n$. Then, the norm
$$\|A\|_* = \sup_{x \in \mathbb{K}^n \setminus \{0\}} \frac{\|Ax\|_*}{\|x\|_*}$$
  is a matrix norm on $\mathcal{M}_n(\mathbb{K})$ which is said to be "subordinate" to the vector norm.
- Operator norm:
$$\|A\|_{a,b} = \sup_{x \neq 0} \frac{\|Ax\|_b}{\|x\|_a}$$
  for $\|\cdot\|_a$ norm on $\mathbb{C}^n$, $\|\cdot\|_b$ norm on $\mathbb{C}^m$, $A \in \mathrm{Lin}(\mathbb{C}^m, \mathbb{C}^n)$. (Not necessarily matrix norms.)
- Convergence for sequences of matrices:
  A sequence of matrices $(A_i)_{i \geq 1}$ converges to a limiting matrix $A \in \mathcal{M}_n(\mathbb{C})$ if, for some matrix norm $\|\cdot\|$,
$$\lim_{i \to \infty} \|A_i - A\| = 0$$
  and we write $\lim\limits_{i \to \infty} A_i = A$.
- Matrix power series:
  Given a sequence $(a_i)_{i \geq 1} \subset \mathbb{C}$, the associated matrix power series is $\sum_{i=0}^{\infty} a_i A^i$.
- Analytic functions of matrices:
  $f : \mathbb{C} \to \mathbb{C}$ analytic on $\{z \in \mathbb{C} : |z| < R\}, R > 0$, written as a power series
$$f(z) = \sum_{i=0}^{\infty} a_i z^i, \quad |z| < R.$$
  For $A \in \mathcal{M}_n(\mathbb{C})$ with $\rho(A) < R$, we define
$$f(A) = \sum_{i=0}^{\infty} a_i A^i.$$

1.5 Algorithms for Matrix Computation/Linear Systems of Equations

- Complexity:
  The complexity of a (matrix) algorithm is the number of multiplications/divisions requried to execute it. For a problem of size n, write
$$\mathrm{N}_{\mathrm{op}}(n) = \# \text{ of mults/divs required.}$$

- Asymptotic complexity:
  Let $\mathrm{N}_{\mathrm{op}}(n)$ denote the complexity of the best algorithm performing a matrix operation. The bound $\mathrm{N}_{\mathrm{op}}(n) \leq Cn^{\alpha}, n \geq 0$ with C, α independent of n, is the asymptotic complexity of the operation, and write
  $$\mathrm{N}_{\mathrm{op}}(n) = O(n^{\alpha})$$

- Condition number:
  The condition number of a matrix $A \in \mathcal{M}_n(\mathbb{K})$ relative to a subordinate matrix norm $\|\cdot\|$ is
  $$\mathrm{cond}(A) = \|A\|\|A^{-1}\|.$$

- Diagonal submatrices:
  $A \in \mathcal{M}_n(\mathbb{R})$, $A = (a_{ij})$, the diagonal submatrices of A are
  $$\Delta^k = \begin{bmatrix} a_{11} & \cdots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{k1} & \cdots & a_{kk} \end{bmatrix}, \quad k = 1, \ldots, n.$$

- Splitting:
  $A \in \mathcal{M}_n(\mathbb{R})$ nonsingular has a splitting (M, N) with $M, N \in \mathcal{M}_n(\mathbb{R})$ if M is nonsingular and $A = M - N$.

- Iterative method:
  The iterative method based on the splitting (M, N) is defined by fixing $x_0 \in \mathbb{R}^n$ and letting x_{k+1} solve
  $$Mx_{k+1} = Nx_k + b, \quad k \geq 0.$$

- Convergence of an iterative method:
  An iterative method converges if (x_k) converges to the exact solution x for any choice of the initial data x_0.

- Residual and error:
  For solving $Ax = b$ with an iterative method, $r_k := b - Ax_k$ is the residual at kth iteration and $\varepsilon_k := x_k - x$ is the error after k iterations.

- Krylov space:
  $r \in \mathbb{R}^n$. For $k \geq 0$, the Krylov space associated to r (and A) is the space
  $$\mathcal{K}_k = \mathrm{span}\{r, Ar, \ldots, A^k r\}.$$

2 Useful Facts

2.1 Matrix Basics

- Gram-Schmidt Orthogonalization:
  Let $\{x_1, \ldots, x_n\}$ be a linearly independent set of vectors in $\mathbb{K}^d$. Then, $\exists$ orthonormal family $\{y_1, \ldots, y_n\} \subset \mathbb{K}^d$ s.t. $\text{span}\{y_1, \ldots, y_p\} = \text{span}\{x_1, \ldots, x_p\}$, $\forall 1 \leq p \leq n$.
- Dimensionality result:
  Let $A \subset \mathbb{K}^d$ be a subspace. If $\{v_1, \ldots, v_k\}$, $\{w_1, \ldots, w_l\}$ are two sets of basis vectors for A, then $k = l$.
- For $A \in \mathcal{M}_n(\mathbb{K})$, TFAE
  i) A is invertible, i.e., $\exists B \in \mathcal{M}_n(\mathbb{K})$ s.t. $AB = BA = I$.
  ii) $\ker(A) = \{0\}$
  iii) $\text{im}(A) = \mathbb{K}^n$
  iv) $\exists B \in \mathcal{M}_n(\mathbb{K})$ s.t. $AB = I_n$ (left inverse)
  v) $\exists B \in \mathcal{M}_n(\mathbb{K})$ s.t. $BA = I_n$ (right inverse)
- Property of trace:
  $A, B \in \mathcal{M}_n(\mathbb{K})$, $\text{tr}(AB) = \text{tr}(BA)$.
- Properties of determinants:
  i) $A, B \in \mathcal{M}_n(\mathbb{K})$, $\det(AB) = \det(A)\det(B) = \det(BA)$.
  ii) $A \in \mathcal{M}_n(\mathbb{K})$, $\det(A) = \det\left(A^\top\right)$
  iii) $A \in \mathcal{M}_n(\mathbb{K})$ is invertible iff $\det(A) \neq 0$.
- Property of triangular matrices:
  i) $T \in \mathcal{M}_n(\mathbb{K})$ lower triangular. If T^{-1} exists, it is also a lower triangular matrix with diagonal entries given as reciprocals of diagonal entries of T.
  ii) If T' is lower triangular, TT' is also lower triangular with diagonal entries being products of diagonal entries of T and T'.
- Inner products and matrices:
  $x, y \in \mathbb{C}^d$,
  $$\langle Ax, y \rangle = \langle x, A^* y \rangle$$
- Block matrices:
  $A = (A_{I,J})$, $B = (B_{I,J})$ for some partition (n_I). Then, $C = AB$ also has block structure $(C_{I,J})$ with
  $$C_{I,J} = \sum_{k=1}^{P} A_{I,K} B_{K,J} \text{ for } 1 \leq I, J \leq P$$
- $\det \begin{bmatrix} A & C \\ 0 & B \end{bmatrix} = \det(A)\det(B)$

2.2 Spectral Theory

- $\lambda \in \sigma(A)$ implies $\exists$ eigenvector associated to λ, i.e., $\ker(A - \lambda I) \neq \{0\}$.
- If $\exists x \neq 0$ with $Ax = \lambda x$, then λ is an eigenvalue of A.

- **Invariance of eigenvalues**:
  Both the characteristic polynomial and eigenvalues are invariant under change of basis, i.e., for any $Q \in \mathcal{M}_n(\mathbb{C})$ invertible,
  $$P_{QAQ^{-1}} = P_A, \quad \sigma(QAQ^{-1}) = \sigma(A).$$
- If A: Hermitian, then all its eigenvalues are real.
- Lemma:
  If $x \in \mathbb{C}^d$ satisfies $Ax = \lambda x$ for some $\lambda \in \mathbb{C}$, then $P(A)x = P(\lambda)x$ for all polynomial $P \in \mathbb{C}[x]$. In particular, $\lambda \in \sigma(A) \Rightarrow P(\lambda) \in \sigma(P(A))$.
- Cayley-Hamilton Thm:
  Given $A \in \mathcal{M}_n(\mathbb{C})$. Let $P_A \in \mathbb{C}[x]$ be the characteristic polynomial of A. Then, $P_A(A) = 0$.
- **Spectral Decomposition (Spectral Thm)**:
  Suppose $A \in \mathcal{M}_n(\mathbb{C})$ has p distinct eigenvalues $\lambda_1, \ldots, \lambda_p$ with each λ_i having algebraic multiplicity n_i. Then, the generalized eigenspaces F_{λ_i} satisfy $\dim F_{\lambda_i} = n_i$.
- Proposition:
  Any matrix $A \in \mathcal{M}_n(\mathbb{C})$ can be reduced to (upper) triangular form.
- **Schur Factorization**:
  For all $A \in \mathcal{M}_n(\mathbb{C})$, $\exists U \in \mathcal{M}_n(\mathbb{C})$ unitary (i.e., $UU^* = U^*U = I$) s.t. $T = U^{-1}AU$ is triangular.
- Proposition:
  If $A \in \mathcal{M}_n(\mathbb{C})$ has p distinct eigenvalues $\lambda_1, \ldots, \lambda_p$, then A is diagonalizable.
- **Thm**:
  $A \in \mathcal{M}_n(\mathbb{C})$ is normal $\iff \exists U \in \mathcal{M}_n(\mathbb{C})$ unitary s.t. $A = U \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\} U^{-1}$.
- **Thm**:
  $A \in \mathcal{M}_n(\mathbb{C})$ is self-adjoint (Hermitian) $\iff$ A: diagonalizable w.r.t. an orthonormal basis and has real eigenvalues.
- Thm:
  $A \in \mathcal{M}_n(\mathbb{C})$ self-adjoint. The smallest eigenvalue λ_1 of A satisfies
  $$\lambda_1 = \min_{x \in \mathbb{C}^n \setminus \{0\}} R_A(x) = \min_{x \in \mathbb{C}^n, \|x\|=1} \langle Ax, x \rangle$$
  and the minimum value is attained for at least one eigenvector $x \neq 0$.
- Proposition:
  $A \in \mathcal{M}_n(\mathbb{C})$ self-adjoint with eigenvalues $\lambda_1, \ldots, \lambda_n$ in increasing order. Then, for $i = 2, \ldots, n$,
  $$\lambda_i = \min_{x \perp \operatorname{span}\{x_1, \ldots, x_{i-1}\}} R_A(x)$$
  where $\{x_1, \ldots, x_n\}$ are eigenvectors of A associated to eigenvalues $(\lambda_1, \ldots, \lambda_n)$, respectively.
- Courant-Fisher Thm: $A \in \mathcal{M}_n(\mathbb{C})$ self-adjoint with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. For all $i = 1, \ldots, n$,
  $$\lambda_i = \max_{\{a_1, \ldots, a_{i-1}\} \subset \mathbb{C}^n} \min_{x \perp \operatorname{span}\{a_1, \ldots, a_{i-1}\}} R_A(x)$$

2.3 Singular Value Decomposition

- **SVD Factorization**:
  Let $A \in \mathcal{M}_{m,n}(\mathbb{C})$ be a matrix having r positive singular values $\mu_1 \geq \mu_2 \geq \cdots \mu_r > 0$.
  Set $\Sigma = \operatorname{diag}\{\mu_1, \ldots, \mu_r\}$ and $\tilde{\Sigma} = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} \in \mathcal{M}_{m,n}(\mathbb{R})$.
  Then, there exist unitary matrices $U \in \mathcal{M}_n(\mathbb{C})$, $V \in \mathcal{M}_m(\mathbb{C})$ s.t.
  $$A = V\tilde{\Sigma}U^*$$

- **Properties of SVD**:
  - If $A = V\tilde{\Sigma}U^*$ is a SVD factorization and $\mu_1, \dots, \mu_r$ are nonzero singular values of A,

$$A = \sum_{i=1}^{r} \mu_i v_i u_i^*$$

  - Columns u_i of U are eigenvectors of A^*A, and columns v_i of V are eigenvectors of AA^*.
  - $\operatorname{rank} A = r \leq \min\{m, n\}$
- Properties of the pseudoinverse:
  - i) If $\operatorname{rank}(A) = n \leq m$

$$A^\dagger = (A^*A)^{-1}A^*$$

    so that if A is square and nonsingular, then $AA^\dagger = A^\dagger A = I$ and $A^\dagger = A^{-1}$.
  - ii) $A^\dagger$ is the unique matrix X s.t. all of the following hold
    1. $AXA = A$
    2. $XAX = X$
    3. $XA = (XA)^*$
    4. $AX = (AX)^*$
  - iii) Minimum length solution to $Ax = b \Rightarrow x^\dagger = A^\dagger b$.
- Properties of fundamental spaces:
  - $\dim(\ker A) = n - \operatorname{rank} A$ (rank-nullity thm)
  - $\dim(\operatorname{row} A) = \operatorname{rank} A \leq n$
  - $\dim(\ker A^\top) = m - \operatorname{rank} A$
  - $\ker A = \operatorname{row}(A)^\perp$
  - $\ker A^\top = \operatorname{col}(A)^\perp$
- Polar decomposition:
  For all $A \in \mathcal{M}_n(\mathbb{R})$, there exists orthogonal Q and $S \in \mathcal{M}_n(\mathbb{R})$ symmetric and positive semidefinite s.t. $A = QS$. If A is invertible, S is positive definite.

2.4 Matrix Norms

- Comparing norms:
  - For $p \geq 1$, $x \in \mathbb{K}^d$,

$$|x_i| \leq \left(\sum_{i=1}^{d} |x_i|^p\right)^{\frac{1}{p}}, \quad \forall i \quad \Rightarrow \|x\|_\infty \leq \|x\|_p$$

  - For $p \geq 1$, $x \in \mathbb{K}^d$,

$$\|x\|_p = \left(\sum_{i=1}^{d} |x_i|^p\right)^{\frac{1}{p}} \leq \left(\sum_{i=1}^{d} \|x\|_\infty^p\right)^{\frac{1}{p}} = \|x\|_\infty d^{\frac{1}{p}}$$

  - For $x \in \mathbb{K}^d$,

$$\|x\|_2 = \left(\sum_{i=1}^{d} |x_i|^2\right)^{\frac{1}{2}} \leq \sum_{i=1}^{d} \left(|x_i|^2\right)^{\frac{1}{2}} = \sum_{i=1}^{d} |x_i| = \|x\|_1$$

- Properties of vector norms:
  - $\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\|$ and $|\|x\| - \|y\|| \leq \|x - y\|$.
    In particular, $x \mapsto \|x\|$ is uniformly (Lipschitz) continuous.
  - On $\mathbb{R}^d$, Cauchy-Schwarz: $x \cdot y \leq \|x\|_2 \|y\|_2$
- **Equivalence of vector norms**:
  E: finite dimensional veector space. All norms on E are equivalent in the sense that for all norms $\|\cdot\|$, $\|\cdot\|'$, $\exists c, C > 0$ s.t. $c\|x\| \leq \|x\|' \leq C\|x\|$ for all $x \in E$.
- Frobenius norm is a matrix norm. $\|\cdot\|_{\ell^\infty}$ is not a matrix norm.
- Properties of subordinate norms:
  - All subordinate matrix norms are matrix norms. Not all matrix norms are subordinate to a vector norm. (e.g., Frobenius norm)
  - By homogeneity, for $A \in \mathcal{M}_n(\mathbb{K})$,
    $$\|A\|_* = \sup_{\substack{x \in \mathbb{K}^n \\ \|x\|_* = 1}} \|Ax\|_* = \sup_{\substack{x \in \mathbb{K}^n \\ \|x\|_* \leq 1}} \|Ax\|_*$$
  - $\|I_n\|_* = 1$ for all vector norms $\|\cdot\|_*$, generating a subordinate norm.
- Proposition:
  Let $\|\cdot\|$ be a subordinate matrix norm on $\mathcal{M}_n(\mathbb{K})$. Then, for $A \in \mathcal{M}_n(\mathbb{K})$, $\exists x_A \in \mathbb{K}^n \setminus \{0\}$ s.t.
  $$\|A\| = \frac{\|Ax_A\|}{\|x_A\|}$$
- $\tilde{x}_A = \frac{x_A}{\|x_A\|} \Rightarrow \exists x_{\max}$ with $\|x_{\max}\| = 1$ s.t. $\|Ax_{\max}\| = \|A\|$.
- **Property of 1-norm**:
  Let $A \mapsto \|A\|_1$ denote the matrix norm subordinate to $\|\cdot\|_1$ on $\mathbb{K}^n$. Then, for $A \in \mathcal{M}_n(\mathbb{K})$,
  $$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}|$$
  i.e., the largest column sum.
- **Property of ∞-norm**:
  Let $A \mapsto \|A\|_\infty$ denote the matrix norm subordinate to $\|\cdot\|_\infty$ on $\mathbb{K}^n$. Then, for $A \in \mathcal{M}_n(\mathbb{K})$,
  $$\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{i=1}^{n} |a_{ij}|$$
  i.e., the largest row sum.
- **Property of 2-norm**:
  Let $\|\cdot\|_2$ be the matrix norm subordinate to $\|\cdot\|_2$ for $A \in \mathcal{M}_n(\mathbb{K})$. This is also called the spectral norm. Then, $\forall A \in \mathcal{M}_n(\mathbb{K})$,
  $$\|A\|_2 = \|A^*\|_2 = \mu_1$$
  where $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_r > 0$ are nonzero singular values of A for $A \neq 0$.
- Lemma:
  If $U \in \mathcal{M}_n(\mathbb{C})$ is unitary ($UU^* = U^*U = I$), then for all $A \in \mathcal{M}_n(\mathbb{C})$,
  $$\|UA\|_2 = \|AU\|_2 = \|A\|_2$$
- Properties of spectral radius:

  - $A \mapsto \rho(A)$ is not a norm on $\mathbb{C}^{n\times n}$.
  - If $A \in \mathcal{M}_n(\mathbb{C})$ is a normal matrix, then $\|A\|_2 = \rho(A)$.
  - If $A \mapsto \|A\|$ is a matrix norm defined on $\mathcal{M}_n(\mathbb{C})$, then $\rho(A) \leq \|A\|$ for all $A \in \mathcal{M}_n(\mathbb{C})$.
  - Given $A \in \mathcal{M}_n(\mathbb{C})$ and $\varepsilon > 0$, there exists a subordinate matrix norm $B \mapsto \|B\|_{A,\varepsilon}$ s.t. $\|A\|_{A,\varepsilon} \leq \rho(A) + \varepsilon$.

- **Proposition**:
  Let $A = V\tilde{\Sigma}U^*$ be an SVD factorization of $A \in \mathcal{M}_{m,n}(\mathbb{C})$ with r nonzero singular values of A arranged in decreasing order.
  For each $1 \leq k \leq r$, the matrix $A_k = \sum_{i=1}^{k} \mu_i v_i u_i^*$ satisfies
  $$\|A - A_k\|_2 \leq \|A - X\|_2$$
  for all $X \in \mathcal{M}_{m,n}(\mathbb{C})$ with $\operatorname{rank} X = k$. Moreover, $\|A - A_k\|_2 = \mu_{k+1}$.

- Proposition:
  $A \in \mathcal{M}_n(\mathbb{C})$. Then, TFAE
  i) $A^i \to 0$ as $i \to \infty$.
  ii) $A^i x \to 0$ as $i \to \infty, \forall x \in \mathbb{C}^n$.
  iii) $\rho(A) < 1$.
  iv) There is a subordinate matrix norm $\|\cdot\|$ with $\|A\| < 1$.

- Theorem:
  Suppose $(a_i) \subset \mathbb{C}$ defines a power series in $\mathbb{C}$ with radius of convergence $R \to 0$, $\sum_{i=0}^{\infty} a_i z^i$. Then, for any $A \in \mathcal{M}_n(\mathbb{C})$ with $\rho(A) < R$, the series
  $$\sum_{i=0}^{\infty} a_i A^i$$
  converges in $\mathcal{M}_n(\mathbb{C})$.

- Proposition:
  $A \in \mathcal{M}_n(\mathbb{C})$, $\rho(A) < 1$. Then, $(I - A) \in \mathcal{M}_n(\mathbb{C})$ is nonsingular with
  $$(I - A)^{-1} = \sum_{i=0}^{\infty} A^i.$$

2.5 Algorithms for Matrix Computation/Linear Systems of Equations

- Strassen's algorithm:
  We can compute AB using 7 (block matrix) multiplications.

- Theorem:
  Each of the following has the same asymptotic complexity in the sense that if any has an algorithm computing with complexity $O(n^\alpha), \alpha \geq 2$, then so do the other three:
  (i) $A, B \rightsquigarrow C = AB$ (matrix multiplication)
  (ii) $A \rightsquigarrow A^{-1}$ (taking the inverse)
  (iii) $A \rightsquigarrow \det(A)$ (computing determinant)
  (iv) $A, b \rightsquigarrow x = A^{-1}b$ (solving linear system)

- Theorem:
  $A \in \mathcal{M}_n(\mathbb{C}), b \in \mathbb{C}^n$. $x \in \mathbb{C}^n$ s.t. $Ax = b$.
  - A: nonsingular $\iff \exists! x = A^{-1}b$.

  - A: singular $\Rightarrow$ either

    (1) $b \in \operatorname{im} A$, $\{x_0 + v : v \in \ker A\}$; or,

    (2) $b \notin \operatorname{im} A$, $Ax = b$ has no solutions.

- Cramer's formula:
  $A \in \mathcal{M}_n(\mathbb{R})$ nonsingular, with columns $a_1, \ldots, a_n$ and consider $Ax = b, b \in \mathbb{R}^n$. The solution $x = (x_1, \ldots, x_n)$ is given by

$$x_i = \frac{\det \begin{bmatrix} a_1 & a_2 & \cdots & a_{i-1} & b & a_{i+1} & \cdots & a_n \end{bmatrix}}{\det(A)}$$

  i.e., replacing ith column with b for the determinant in the numerator.

- Facts about simple matrices

  1. A: diagonal $\rightsquigarrow Ax = b$ requires $O(n)$.
  2. A: unitary $\rightsquigarrow Ax = b$ requires $O(n^2)$.
  3. A: lower triangular $\rightsquigarrow Ax = b$ requires $O(n^2)$ (via forward substitution).
  4. A: upper triangular $\rightsquigarrow Ax = b$ requires $O(n^2)$ (via backward substitution).

- Properties of condition number:

  1. $\operatorname{cond}(A) \geq 1$
  2. Perturbation bound:
     Suppose $Ax = b$. $A_\varepsilon = A + \varepsilon B, b_\varepsilon = b + \varepsilon\gamma$ for some $B \in \mathcal{M}_n(\mathbb{K}), \gamma \in \mathbb{K}^n$. Consider $A_\varepsilon x_\varepsilon = b_\varepsilon$. Then, the perturbation bound is

$$\frac{\|x_\varepsilon - x\|}{\|x\|} \leq \operatorname{cond}(A) \left(\frac{\|A_\varepsilon - A\|}{\|A\|} + \frac{\|b_\varepsilon - b\|}{\|b\|} \right) + O(\varepsilon^2)$$

  3. Proposition:
     $A \in \mathcal{M}_n(\mathbb{R}), b \in \mathbb{R}^n \setminus \{0\}, \delta_b \in \mathbb{R}^n$. If $Ax = b, A(x + \delta_x) = b + \delta_b$, then

$$\frac{\|\delta_x\|}{\|x\|} \leq \operatorname{cond}(A) \frac{\|\delta_b\|}{\|b\|}$$

  4. Equivalence of condition numbers:
     - $n^{-1} \operatorname{cond}_2(A) \leq \operatorname{cond}_1(A) \leq n \operatorname{cond}_2(A)$
     - $n^{-1} \operatorname{cond}_\infty(A) \leq \operatorname{cond}_2(A) \leq n \operatorname{cond}_\infty(A)$
     - $n^{-2} \operatorname{cond}_1(A) \leq \operatorname{cond}_\infty(A) \leq n^2 \operatorname{cond}_1(A)$
  5. A: nonsingular, $\operatorname{cond}(A) = \operatorname{cond}(A^{-1})$.
  6. $\alpha \in \mathbb{C} \setminus \{0\}$, $\operatorname{cond}(\alpha A) = \operatorname{cond}(A)$
  7. $\operatorname{cond}_2(A) = \frac{\mu_1(A)}{\mu_n(A)}$ where $\mu_1(A)$: the largest, $\mu_n(A)$: the smallest singular value
  8. A: normal, $\operatorname{cond}_2(A) = \rho(A)\rho(A^{-1}) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}$
  9. U: unitary, $\operatorname{cond}_2(A) = 1$
  10. $\rho(A)\rho(A^{-1}) \leq \operatorname{cond}(A)$
  11. A: normal, $\operatorname{cond}_2(A) \leq \operatorname{cond}(A)$ for any condition number

- Lemma:
  $A \in \mathcal{M}_n(\mathbb{C})$: nonsingular, then
  $$\frac{1}{\text{cond}_2(A)} = \inf\left\{\frac{\|A-B\|_2}{\|A\|_2} : B \in S_n(\mathbb{C})\right\}.$$
  where $S_n(\mathbb{C}) = \{B \in \mathcal{M}_n(\mathbb{C}) : B \text{ singular}\}$.

- Theorem (Gaussian elimination):
  $A \in \mathcal{M}_n(\mathbb{C})$. $\exists M \in \mathcal{M}_n(\mathbb{C})$ nonsingular s.t. $T = MA$ is upper triangular.

- Proposition (LU decomposition):
  $A \in \mathcal{M}_n(\mathbb{R})$ s.t. all diagonal submatrices $\Delta^k,\ k = 1, \ldots, n$ are nonsingular.
  Then, $\exists! L, U \in \mathcal{M}_n(\mathbb{R})$ s.t. $L = (\ell_{ij}), \ell_{ii} = 1, \forall 1 \le i \le n$ lower triangular, $U = (u_{ij})$ upper triangular with $A = LU$.
  (cf. Matrix analog of Gaussian elimination.)

- Theorem (Cholesky decomposition):
  $A \in \mathcal{M}_n(\mathbb{R})$ symmetric, p.d. Then, $\exists! B$ real lower triangular s.t. $A = BB^\top$ with diagonal entries strictly positive.

- Theorem (QR decomposition):
  $A \in \mathcal{M}_n(\mathbb{R})$ nonsingular. Then, $\exists!(Q, R)$ s.t. $Q \in \mathcal{M}_n(\mathbb{R})$ orthogonal, $R \in \mathcal{M}_n(\mathbb{R})$ upper triangular with $A = QR$.
  (cf. Matrix analog of Gram-Schmidt process.)

- Lemma:
  $A \in \mathcal{M}_{n,p}(\mathbb{R}), b \in \mathbb{R}^n$. Then, $x \in \mathbb{R}^p$ minimizes $\|b - Ax\|_2 \iff A^*Ax = A^*b$.

- Theorem:
  $A \in \mathcal{M}_{n,p}(\mathbb{R})$. Then, $\exists x \in \mathbb{R}^p$ s.t. $A^*Ax = A^*b$ (the normal equation).

- Proposition:
  $A^*Ax = A^*b$ has exactly one solution $\iff \ker A = \{0\}$.

- **Theorem**:
  The iterative method associated to the splitting (M, N) converges $\iff \rho(M^{-1}N) < 1$.

- Richardson's method/Gradient descent/Steepest descent:
  Splitting: $M = \alpha^{-1}I, N = \alpha^{-1}I - A$.
  Iteration matrix: $B_\alpha = M^{-1}N = I - \alpha A$.
  (cf. converges iff $0 < \alpha < \frac{2}{\rho(A)}$.)

- Jacobi method:
  Splitting: $M = D = \text{diag}(a_{11}, \ldots, a_{nn}), N = D - A$.
  Iteration matrix: $J = M^{-1}N = I - D^{-1}A$.
  (cf. well defined if $\det D = a_{11} \cdot a_{nn} \neq 0$.)

- Theorem:
  A: Hermitian, p.d. If (M, N): splitting of A, then $M^* + N$: Hermitian. If $(M^* + N)$: p.d., then $\rho(M^{-1}N) < 1$.

- Gauss-Seidel method:
  $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$, write $A = D - E - F$ with $D = \text{diag}(a_{11}, \ldots, a_{nn})$ $-E$: lower triangular part of A, and $-F$: upper triangular part of A.
  Splitting: $M = D - E, N = F$.
  Iteration matrix: $G = M^{-1}N = (D - E)^{-1}F$.

- Proposition:
  $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$ for $A \in \mathcal{M}_n(\mathbb{R})$ symmetric, $b \in \mathbb{R}^n$. Then, $(\nabla f)(x) = Ax - b$. Moreover, A: p.d. $\Rightarrow f$ admits a unique minimum x_0 solving $Ax_0 = b$.

- Proposition:
  A: real symmetric p.d. $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$. Then, if $F \subset \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$, then $\exists x_0 \in F$ s.t. $f(x_0) \le f(x), \forall x \in F$. Moreover, x_0 is the unique vector in F s.t. $\langle Ax_0 - b, y\rangle = 0, \forall y \in F$.

- Theorem:
  A: real symmetric p.d. $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$. Then, $x \in \mathbb{R}^n$ minimizes $f \iff (\nabla f)(x) = 0$ and if $x \in \mathbb{R}^n$ s.t. $(\nabla f)(x) \neq 0$ then $\forall \alpha \in \left(0, \frac{2}{\rho(A)}\right)$, $f(x - \alpha \nabla f(x)) < f(x)$.

- Properties of Krylov spaces:
  1. $\mathcal{K}_k \subset \mathcal{K}_{k+1}, \forall k \ge 0$.
  2. $\forall r_0 \in \mathbb{R}^n \setminus \{0\}, \exists k_0 \in \{0, \dots, n-1\}$, "Krylov critical dimension" s.t.
     $\dim \mathcal{K}_k = k + 1$ for $0 \le k \le k_0$, and $\dim \mathcal{K}_k = k_0 + 1$ for $k \ge k_0$.
  3. The gradient iteration and its residual $r_k = b - Ax_k$ satisfy
     (i) $r_k \in \mathcal{K}_k(r_0, A)$
     (ii) $x_{k+1} \in [x_0 + \mathcal{K}_k] = \{x : x - x_0 \in \mathcal{K}_k\}$ (the subspace).