Linear Algebra: Formulae and Theorems

]Reference : Howard Anton, Gil Strang, Golub and Von Loan

1. Preliminaries and Notation

Vectors: $x \in \mathbb{R}^n$ is a column vector. Standard inner product

$$\langle x, y \rangle = x^T y, \qquad \|x\|_2 = \sqrt{x^T x}.$$

General inner product space $(V, \langle \cdot, \cdot \rangle)$, induced norm $\|v\| = \sqrt{\langle v, v \rangle}$. Matrices: $A \in \mathbb{R}^{m \times n}$. Identity I_n. Spectrum $\mathrm{spec}(A)$. Transpose A^T, conjugate transpose A^* for complex matrices. Rank $\mathrm{rank}(A)$, nullspace $\ker(A) = \{x : Ax = 0\}$, column space $\mathrm{Col}(A)$, row space $\mathrm{Row}(A) = \mathrm{Col}(A^T)$. Notation: $\sigma_i(A)$ singular values, ordered $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$.

2. Vector Operations & Properties

2.1. Linear combinations, span, independence

$$\mathrm{span}\{v_1, \ldots, v_k\} = \left\{ \sum_{i=1}^{k} \alpha_i v_i : \ \alpha_i \in \mathbb{R} \right\}.$$

Linear independence: $\sum_i \alpha_i v_i = 0 \Rightarrow \alpha_i = 0$ for all i. Basis: linearly independent spanning set. Dimension equals cardinality of basis.

2.2. Orthogonality

$$x \perp y \iff \langle x, y \rangle = 0.$$

Orthogonal complement: $W^\perp = \{v : \langle v, w \rangle = 0, \ \forall w \in W\}$. Projection onto subspace spanned by orthonormal $Q = [q_1 \cdots q_k]$:

$$P = QQ^T, \qquad Py = \sum_{i=1}^{k} \langle y, q_i \rangle q_i.$$

3. Linear Systems

Solve $Ax = b$, $A \in \mathbb{R}^{m \times n}$.

- **Consistent** iff $b \in \mathrm{Col}(A)$.
- General solution: $x = x_p + x_h$, where x_p is any particular solution and $x_h \in \ker(A)$.
- Rank-nullity theorem:
$$\mathrm{rank}(A) + \dim \ker(A) = n.$$
- Square invertible case ($n = m$): unique solution $x = A^{-1} b$.
- Cramer's rule (if A invertible):
$$x_i = \frac{\det(A_i)}{\det(A)}, \quad A_i = [\, a_1 \ \cdots \ b \ \cdots \ a_n \,].$$

3.1. Least squares (overdetermined)

Minimize $\|Ax - b\|_2^2$. Normal equations:

$$A^T A x = A^T b.$$

If A full column rank ($\mathrm{rank} = n$):

$$x^\star = (A^T A)^{-1} A^T b = A^+ b.$$

Equivalently, compute QR with $A = QR$ and solve $Rx = Q^T b$.

4. Matrix Algebra and Identities

4.1. Basic identities

$$(AB)^T = B^T A^T, \quad (AB)^{-1} = B^{-1} A^{-1} \text{ (if invertible)}.$$

Trace properties:

$$\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B), \quad \mathrm{tr}(AB) = \mathrm{tr}(BA).$$

Determinant:

$$\det(AB) = \det A \det B, \quad \det(A^T) = \det A.$$

4.2. Matrix norms

Operator norm (induced 2-norm):

$$\|A\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \sigma_1(A).$$

Frobenius norm:

$$\|A\|_F = \sqrt{\sum_{i,j} a_{ij}^2} = \sqrt{\mathrm{tr}(A^T A)} = \sqrt{\sum_i \sigma_i^2}.$$

subordinate norms: $\|A\|_p = \max_{\|x\|_p=1} \|Ax\|_p$.

4.3. Condition number

For invertible A,

$$\kappa_2(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}.$$

Interpretation: relative error amplification in solving linear systems.

4.4. Useful lemmas

Matrix determinant lemma:

$$\det(A + uv^T) = \det(A)\,(1 + v^T A^{-1} u) \quad (A \text{ invertible}).$$

Sherman-Morrison formula:

$$(A + uv^T)^{-1} = A^{-1} - \frac{A^{-1} uv^T A^{-1}}{1 + v^T A^{-1} u}.$$

5. Determinants: properties and expansions

5.1. Cofactor expansion

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij}) \quad (\text{expansion along row } i).$$

Multilinearity: determinant linear in each row. Swapping two rows flips sign.

5.2. Volume interpretation

Absolute value $|\det(A)|$ equals volume scaling of the unit cube under linear map A. Zero determinant $\iff$ matrix singular.

6. Vector Spaces and Subspaces

6.1. Definitions

A subspace $W \subseteq V$ must satisfy $0 \in W$, closed under addition and scalar multiplication. Row space, column space, nullspace are examples.

6.2. Direct sums and complements

If $V = W \oplus U$ then every v decomposes uniquely $v = w + u$ with $w \in W$, $u \in U$ and $W \cap U = \{0\}$.

7. Inner Product Spaces, Orthogonality, and Projections

7.1. Properties and inequalities

Cauchy–Schwarz:

$$|\langle x, y \rangle| \leq \|x\| \|y\|.$$

Triangle inequality and parallelogram law:

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

7.2. Orthogonal projection formula

Projection onto $\mathrm{Col}(A)$ when A has full column rank:

$$P = A(A^TA)^{-1}A^T, \qquad P^T = P,\ P^2 = P.$$

Projection error orthogonality: $Ax^\star - b \perp \mathrm{Col}(A)$.

8. Orthogonalization: Gram–Schmidt and Householder

8.1. Classical Gram–Schmidt

Given $a_1, \dots, a_n$,

$$u_1 = a_1, \quad q_1 = \frac{u_1}{\|u_1\|},$$

$$u_k = a_k - \sum_{j=1}^{k-1} \langle a_k, q_j \rangle q_j, \quad q_k = \frac{u_k}{\|u_k\|}.$$

Produces Q with orthonormal columns and upper triangular R with $A = QR$.

8.2. Modified Gram–Schmidt (numerically stable)

Equivalent algebraically but reduces loss of orthogonality. Prefer using Householder reflections for stability and efficient QR.

8.3. Householder reflections

For vector x, Householder matrix

$$H = I - 2\frac{vv^T}{v^Tv}, \qquad v = x \pm \|x\|e_1,$$

is orthogonal, $H^TH = I$, and can zero subdiagonal elements to form QR with fewer orthogonality errors.

9. Rank Factorizations and Column–Row (CR) Decomposition

If $\mathrm{rank}(A) = r$, pick r independent columns to form $C \in \mathbb{R}^{m\times r}$ and select $R \in \mathbb{R}^{r\times n}$ such that

$$A = CR.$$

Equivalently express as sum of rank-1 outer products:

$$A = \sum_{i=1}^{r} u_i v_i^T.$$

If $A = U\Sigma V^T$ (SVD) then choose $u_i = \sigma_i u_i$ and $v_i = v_i$ to express outer-product form:

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T.$$

10. LU and Cholesky Decompositions

10.1. LU (Doolittle / Crout)

If A admits LU without pivoting,

$$A = LU, \qquad L \text{ unit lower triangular, } U \text{ upper triangular.}$$

With partial pivoting:

$$PA = LU,$$

where P is a permutation matrix.

10.2. Cholesky (symmetric positive definite)

If A symmetric positive definite (SPD),

$$A = LL^T, \qquad L \text{ lower triangular with positive diagonal.}$$

Algorithm: successive elimination yields $l_{ii} = \sqrt{a_{ii} - \sum_{k<i} l_{ik}^2}$.

11. QR Decomposition and Applications

$$A = QR, \qquad Q^T Q = I, \ R \text{ upper triangular.}$$

Thin (economy) QR when $m \geq n$: $A = Q_{m\times n} R_{n\times n}$. Use cases: least squares solution, orthonormal basis of column space, numerical stability.

12. Eigenvalues, Eigenvectors and Diagonalization

12.1. Definitions

Eigenpair (λ, v) satisfies $Av = \lambda v$ with $v \neq 0$. Characteristic polynomial

$$p_A(\lambda) = \det(A - \lambda I).$$

Algebraic multiplicity = multiplicity as root of p_A. Geometric multiplicity = $\dim \ker(A - \lambda I)$.

12.2. Diagonalization

If A has n linearly independent eigenvectors then

$$A = X \Lambda X^{-1}, \quad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n).$$

Special case: symmetric A admits orthogonal diagonalization

$$A = Q \Lambda Q^T, \quad Q^T Q = I.$$

12.3. Spectral theorem (real symmetric)

For real symmetric A, eigenvalues real, eigenvectors orthonormal. Functional calculus: $f(A) = Q f(\Lambda) Q^T$.

12.4. Rayleigh quotient and variational characterizations

Rayleigh quotient for symmetric A:

$$R_A(x) = \frac{x^T A x}{x^T x}.$$

Courant–Fischer:

$$\lambda_k = \min_{\dim S = k} \max_{x \in S, \ x \neq 0} R_A(x) = \max_{\dim T = n-k+1} \min_{x \in T, \ x \neq 0} R_A(x).$$

12.5. Gershgorin disk theorem (eigenvalue bounds)

Each eigenvalue λ lies in at least one disk

$$D(a_{ii}, R_i), \qquad R_i = \sum_{j \neq i} |a_{ij}|.$$

13. Jordan Canonical Form

If A not diagonalizable, over $\mathbb{C}$ there exists X such that

$$A = X J X^{-1}, \quad J = \bigoplus J_k(\lambda),$$

where each Jordan block $J_k(\lambda)$ has λ on diagonal and ones on superdiagonal. Jordan form classifies nilpotent structure.

14. Singular Value Decomposition (SVD)

14.1. Existence and form

For $A \in \mathbb{R}^{m\times n}$ there exist orthogonal $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ and diagonal $\Sigma \in \mathbb{R}^{m\times n}$ with nonnegative diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ such that

$$A = U \Sigma V^T.$$

Rank r equals number of nonzero singular values. Compact (thin) SVD:

$$A = U_r \Sigma_r V_r^T, \quad U_r \in \mathbb{R}^{m\times r}, \ V_r \in \mathbb{R}^{n\times r}.$$

14.2. Relations

$$A^T A = V\Sigma^T\Sigma V^T, \qquad AA^T = U\Sigma\Sigma^T U^T.$$

Thus σ_i^2 are eigenvalues of $A^T A$ and AA^T.

14.3. Pseudo-inverse via SVD

Define Σ^+ by inverting nonzero singular values. Then

$$A^+ = V\Sigma^+ U^T.$$

Properties:

$$AA^+A = A, \quad A^+AA^+ = A^+, \quad (AA^+)^T = AA^+, \quad (A^+A)^T = A^+A.$$

14.4. Eckart–Young theorem (best low-rank approximation)

The best approximation of A in the Frobenius or 2-norm by a matrix of rank $\leq k$ is

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T,$$

and

$$\min_{\text{rank}(B)\leq k} \|A - B\|_2 = \sigma_{k+1}, \qquad \min_{\text{rank}(B)\leq k} \|A - B\|_F = \sqrt{\sum_{i=k+1}^{r} \sigma_i^2}.$$

15. Principal Component Analysis (PCA)

Given data matrix $X \in \mathbb{R}^{N\times p}$ with rows observations. Center: $\tilde{X} = X - \mathbf{1}\bar{x}^T$ where $\bar{x} = \frac{1}{N}\sum_i x_i$. Sample covariance matrix:

$$S = \frac{1}{N-1}\tilde{X}^T\tilde{X}.$$

PCA by eigen-decomposition: $S = V\Lambda V^T$. Principal axes are columns of V, explained variances are λ_i. Alternatively compute SVD: $\tilde{X} = U\Sigma V^T$, then

$$S = \frac{1}{N-1}V\Sigma^T\Sigma V^T, \qquad \lambda_i = \frac{\sigma_i^2}{N-1}.$$

Scores (projections of data onto principal components): $U\Sigma$ (columns correspond to PC scores).

15.1. Whitening

Whitened data:

$$X_{\text{white}} = \tilde{X}V\Lambda^{-1/2} = U,$$

so covariance of X_{white} is identity.

16. Singular and Generalized Inverses

16.1. Moore–Penrose pseudoinverse

Unique A^+ satisfying

$$AA^+A = A, \quad A^+AA^+ = A^+, \quad (AA^+)^T = AA^+, \quad (A^+A)^T = A^+A.$$

Computed from SVD as $A^+ = V\Sigma^+U^T$.

16.2. Minimum-norm least squares

The minimum-norm solution to $\min \|x\|_2$ subject to $Ax = b$ (consistent) is $x = A^+b$.

17. Matrix Functions and Exponentials

For diagonalizable $A = X\Lambda X^{-1}$ and scalar function f,

$$f(A) = Xf(\Lambda)X^{-1},$$

with $f(\Lambda) = \mathrm{diag}(f(\lambda_i))$. Matrix exponential important for linear ODEs:

$$e^{At} = \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}.$$

If A diagonalizable $e^{At} = Xe^{\Lambda t}X^{-1}$.

18. Schur Decomposition and Schur Complement

18.1. Schur decomposition

Every square A admits orthogonal Schur decomposition

$$A = QTQ^T,$$

with Q orthogonal and T upper quasi-triangular (real Schur) or upper triangular (complex Schur).

18.2. Schur complement

For block matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

if A invertible, Schur complement of A is $D - CA^{-1}B$. Determinant identity:

$$\det(M) = \det(A)\det(D - CA^{-1}B).$$

19. Interlacing and Perturbation Results

19.1. Interlacing (Hermitian matrices)

If A Hermitian, and B is principal submatrix, eigenvalues interlace:

$$\lambda_i(A) \geq \lambda_i(B) \geq \lambda_{i+1}(A).$$

19.2. Weyl's inequalities (eigenvalue perturbation)

If A and E Hermitian, then for eigenvalues ordered,

$$|\lambda_i(A + E) - \lambda_i(A)| \leq \|E\|_2.$$

20. Selected Proof Sketches and Short Derivations

20.1. Rank(A) = rank(A^TA)

$$x \in \ker(A^TA) \iff x^TA^TAx = \|Ax\|^2 = 0 \iff Ax = 0.$$

Thus nullspaces equal so ranks equal.

20.2. SVD existence (sketch)

A^TA is symmetric p.s.d. so diagonalize $A^TA = V\Lambda V^T$. Let $\sigma_i = \sqrt{\lambda_i}$. Define $u_i = (1/\sigma_i)Av_i$ for $\sigma_i > 0$. Assemble U, Σ, V.

20.3. Eckart–Young idea

Error in Frobenius norm minimized by truncating SVD because singular values give orthogonal energy decomposition:

$$\|A - \sum_{i=1}^{k} \sigma_i u_i v_i^T\|_F^2 = \sum_{i=k+1}^{r} \sigma_i^2.$$

21. Computational Recipes and Algorithms

21.1. Solve $Ax = b$

1. If A square and well-conditioned: LU with partial pivoting $PA = LU$, solve $Ly = Pb$, then $Ux = y$.
2. If A tall ($m > n$) and full column rank: QR ($A = QR$) then solve $Rx = Q^T b$.
3. If rank-deficient or ill-conditioned: compute SVD and use pseudoinverse $x = A^+ b$.

21.2. Compute SVD numerically

Use bidiagonalization via Householder reflections then compute SVD of bidiagonal matrix by iterative methods (Golub–Kahan algorithm).

21.3. Compute eigenvalues

For symmetric matrices use QR algorithm with shifts. For non-symmetric use Francis double-shift QR or Hessenberg reduction then QR.

22. Summary Tables and Quick Reference

Object / Task	**Formula / Remarks**
Projection onto Col(A)	$P = A(A^T A)^{-1} A^T$ if A full column rank.
Least squares solution	$x^\star = (A^T A)^{-1} A^T b$ or via QR.
SVD	$A = U\Sigma V^T$, best rank-k approx $A_k = \sum_{i=1}^k \sigma_i u_i v_i^T$.
Pseudoinverse	$A^+ = V\Sigma^+ U^T$.
LU w/ pivoting	$PA = LU$.
Cholesky	$A = LL^T$ for SPD A.
Eigen decomposition	$A = X\Lambda X^{-1}$ (if diagonalizable).
Spectral theorem	$A = Q\Lambda Q^T$ for real symmetric A.
Rayleigh quotient	$R_A(x) = \dfrac{x^T A x}{x^T x}$.
Condition number	$\kappa_2(A) = \dfrac{\sigma_1}{\sigma_n}$.

23. Notes, Numerical Stability and Best Practices

- Use partial pivoting in LU to avoid numerical instability.
- Use Householder reflections for QR rather than classical Gram–Schmidt for better orthogonality.
- Use SVD when stability and rank determination are critical.
- Center data before PCA. Scale features if desirable (standardize).
- For large sparse problems use iterative solvers (CG, GMRES) and preconditioning.