

Linear Algebra for Machine Learning: SVD, Eigenvalues & Matrix Calculus

Spectral Theorem, Singular Value Decomposition (SVD), PCA, LoRA low-rank adaptation, and Jacobian/Hessian identities.

EXECUTIVE SUMMARY

Essential mathematical reference connecting matrix factorizations (Eigen-decomposition, SVD, Cholesky) to Principal Component Analysis (PCA), Low-Rank Adaptation (LoRA), and gradient backpropagation.

KEY EQUATIONS & INVARIANTS

* Singular Value Decomposition (SVD):

$$\text{LaTeX: } A = U \Sigma V^{\top} = \sum_{i=1}^r \sigma_i u_i v_i^{\top}$$

Note: Every $m \times n$ matrix A decomposes into orthogonal U ($m \times m$), diagonal singular values $\Sigma \geq 0$, and orthogonal V ($n \times n$).

* Eckart-Young-Mirsky Best Rank-k Approximation:

$$\text{LaTeX: } A_k = \arg\min_{\text{rank}(B)=k} \|A - B\|_F = \sum_{i=1}^k \sigma_i u_i v_i^{\top}$$

Note: Truncating SVD to top-k singular values minimizes Frobenius and spectral reconstruction error.

* LoRA (Low-Rank Adaptation) Weight Update:

$$\text{LaTeX: } W = W_0 + \Delta W = W_0 + \frac{\alpha}{r} B A, \quad B \in \mathbb{R}^{d \times r}, A \in \mathbb{R}^{r \times k}, r \ll \min(d, k)$$

Note: Freezes pretrained W_0 and trains rank-r matrices A and B , cutting trainable parameters by >99%.

1. Matrix Decompositions & ML Applications

For centered data matrix $X \in \mathbb{R}^{n \times d}$, the sample covariance matrix is $C = \frac{1}{n-1} X^{\top} X$. Computing PCA via the thin SVD $X = U \Sigma V^{\top}$ avoids forming $X^{\top} X$ explicitly, squaring the condition number $\kappa(X^{\top} X) = \kappa(X)^2$.

Decomposition | Formula | Matrix Requirement | Primary ML Use Case

Spectral (Eigen) | $A = Q \Lambda Q^{\top}$ | Real symmetric ($A = A^{\top}$) | Covariance matrices, Graph Laplacians

SVD | $A = U \Sigma V^{\top}$ | Any $m \times n$ rectangular matrix | PCA, Moore-Penrose Pseudoinverse, Recommender Systems

Cholesky | $A = L L^{\top}$ | Symmetric Positive Definite (SPD) | Gaussian Processes, Multivariate Normal sampling

QR Factorization | $A = Q R$ | Full column rank $m \geq n$ | Numerically stable Ordinary Least Squares (OLS)

COMMON PRODUCTION & INTERVIEW PITFALLS

- Computing `np.linalg.inv(A) @ b` instead of `np.linalg.solve(A, b)` is both slower and numerically unstable.

FREQUENTLY ASKED QUESTIONS

Q: How are singular values of X related to the eigenvalues of $X^{\top} X$?

A: The non-zero singular values σ_i of X are the exact square roots of the non-zero eigenvalues λ_i of $X^{\top} X$ ($\sigma_i = \sqrt{\lambda_i}$), and the right singular vectors V are the eigenvectors of $X^{\top} X$.