A 2020 Vision of Linear Algebra

https://life-is-too-short-to-for-slides.app

Monday 21 April 2025

Contents

Lecture 1

Important Matrix Decompositions

Many computations in linear algebra rely on writing a matrix A in special forms:

- **Column–Row factorization:** $A = C\,R$, where columns of C are independent.
- **LU factorization:** $A = L\,U$, with L lower–triangular, U upper–triangular.
- **QR factorization:** $A = Q\,R$, where Q has orthonormal columns.
- **Spectral decomposition:** $S = Q\Lambda Q^T$, $Q^T = Q^{-1}$, for a symmetric S.
- **Eigen–decomposition:** $A = X\Lambda X^{-1}$, satisfying $Ax = \lambda x$.
- **SVD:** $A = U\Sigma V^T$, singular values $\sigma_i = \sqrt{\lambda_i(A^T A)}$.

Column Space

The **column space** of A, denoted $C(A)$, is

$$C(A) = \{Ax : x \in^n\} = \{\text{columns of } A\}.$$

In ^{3}, three columns may lie in a plane or a line. If only two columns are independent, $C(A)$ is a plane.

Example: A = C R with Rank 2

Let

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 3 & 2 & 5 \\ 2 & 1 & 3 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 \\ 3 & 2 \\ 2 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Then $A = C\,R$. Here $(A) = 2$, since $C \in^{3\times 2}$ has two independent columns, and the third column of A is $_3(A) =_1 (C) + 2\,_2(C)$.

Key Facts on Rank and Bases

From $A = C\,R$ with $r = (A)$:

1. The r columns of C are independent and form a basis of $C(A)$.
2. Every column of A is a combination of those r columns.
3. The r rows of R are independent and form a basis of the row space of A.
4. Every row of A is a combination of those r rows.

Hence **column rank** $=$ **row rank** $= r$.

Rank via Column–Row Factorization

A=CR implies **column rank**=**row rank**=r:

1. C's r columns are independent.
2. A's columns are combinations of C's.
3. R's r rows are independent (contains an I_r block).
4. A's rows are combinations of R's.

Thus $\dim C(A) = \dim \text{Row}(A) = r$ and $\dim N(A) = n - r$ (e.g. $\{(1, 1, -1)\}$ when $n = 3, r = 2$).

Rank-1 Case and CR Trade-offs

If $(A) = 1$ then $A = v\,w^T$ so all columns (and rows) are multiples of one vector. Benefits: C uses original columns, R is RREF, rank equality transparent. Drawbacks: C,R may be ill-conditioned; if A invertible then $C = A, R = I$ (no simplification).

Orthogonal Subspaces (Big Picture)

For $A \in^{m\times n}$, the four fundamental subspaces satisfy

$$N(A) \perp \mathrm{Row}(A), \quad N(A^T) \perp C(A),$$

with $\dim C(A) = \dim \mathrm{Row}(A) = r,\ \dim N(A) = n - r,\ \dim N(A^T) = m - r.$

Outer Products and Factorizations

Any $A = \sum_{i=1}^{n} b_i\, c_i^T$ splits into rank-1 matrices. Key factorizations: CR, LU, QR, Spectral, Eigen, SVD.

LU via Gaussian Elimination

Gaussian elimination yields $A = L\,U$ with L unit lower-triangular, U upper-triangular. Example:

$$\begin{pmatrix} 2 & 3 \\ 4 & 7 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & 1 \end{pmatrix}.$$

With row exchanges one obtains $P\,A = L\,U$.

LU Decomposition via Elimination

To solve $Ax = b$ we factor $A = L\,U$ by Gaussian elimination, where L is unit lower-triangular and U is upper-triangular. Concretely:

1. For $i = 2, \dots, n$, subtract $\ell_{i1} = \frac{a_{i1}}{a_{11}}$ times row 1 from row i.
2. Repeat on the remaining $(n-1) \times (n-1)$ submatrix to introduce zeros below each pivot.

The multipliers ℓ_{ij} fill L, and the final upper form is U, so $A = L\,U$.

Orthogonal Matrices and Length Preservation

A matrix $Q \in^{m\times n}$ has **orthonormal** columns $q_1, \dots, q_n$ iff

$$Q^T Q = I_n.$$

If Q is square then also $QQ^T = I$, so $Q^{-1} = Q^T$. Key properties:

- $\langle Qx,\, Qy \rangle = x^T Q^T Q\, y = x^T y$: inner products preserved.
- $\|Qx\|_2 = \|x\|_2$: lengths unchanged.
- Eigenvalues satisfy $|\lambda| = 1$. E.g. a planar rotation $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

QR Factorization via Gram–Schmidt

Given $A = [a_1, \dots, a_n]$ with independent columns, we construct $A = Q\,R$ where $Q = [q_1, \dots, q_n]$ has orthonormal columns and R is upper-triangular:

$$\text{For } k = 1, \dots, n: \quad v_k = a_k - \sum_{j=1}^{k-1} (q_j^T a_k)\, q_j,$$
$$r_{kk} = \|v_k\|_2, \quad q_k = v_k / r_{kk}, \quad r_{jk} = q_j^T a_k \quad (j < k).$$

Then $Q^T A = R$ and $Q^T Q = I$.

Least Squares via QR

For an overdetermined system $m > n$, the best fit $\hat{x}$ minimizes $\|b - Ax\|_2$. One solves

$$A^T A\, x = A^T b \quad \Longrightarrow \quad R\, x = Q^T b$$

if $A = Q\,R$. The projection $p = A\hat{x} = Q\,Q^T b$ lies in the column space $C(A)$, and the residual $b - p$ is orthogonal to $C(A)$.

Least Squares via QR

Given $A \in^{m\times n}$ with $m > n$, the **least squares** solution $\hat{x}$ minimizes $\|b - Ax\|_2$. The normal equations

$$A^T(b - A\hat{x}) = 0 \quad \Longrightarrow \quad A^T A\,\hat{x} = A^T b.$$

If $A = Q\,R$ with $Q^TQ = I$, then

$$R\,\hat{x} = Q^T b.$$

The projection $p = A\hat{x}$ lies in $C(A)$ and the residual $e = b - p$ is orthogonal to $C(A)$.

Symmetric Matrices: Orthogonal Eigenvectors

Let $S \in^{n\times n}$ satisfy $S = S^T$. Then all eigenvalues λ are real, and eigenvectors corresponding to distinct eigenvalues are orthogonal.

> *Proof of Orthogonality.* Suppose $Sx = \lambda x$ and $Sy = \alpha y$ with $\lambda \neq \alpha$. Then
>
> $$x^T Sy = \lambda\, x^T y, \qquad x^T Sy = (Sx)^T y = \alpha\, x^T y \quad \Longrightarrow \quad (\lambda - \alpha)x^T y = 0.$$
>
> Hence $x^T y = 0$.

Spectral Decomposition and Connection to SVD

Orthonormal eigenvectors q_i form $Q = [q_1, \ldots, q_n]$ with $Q^TQ = I$ and

$$SQ = Q\Lambda, \quad S = Q\,\Lambda\,Q^T, \quad S = \sum_{i=1}^{n} \lambda_i\, q_i q_i^T.$$

For $S = A^TA$, set $\sigma_i = \sqrt{\lambda_i}$ to obtain the **SVD** $A = U\,\Sigma\,V^T = \sum_{i=1}^{r} \sigma_i\, u_i v_i^T$.

Eigen-decomposition of General Matrices

If $A \in^{n\times n}$ has n independent eigenvectors, write $AX = X\Lambda$. Then

$$A = X\,\Lambda\,X^{-1}, \quad A^k = X\,\Lambda^k\,X^{-1},$$

so the same eigenvectors diagonalize all powers. Moreover $A^k \to 0$ as $k \to \infty$ if $\max_i |\lambda_i| < 1$.

Positive Semidefiniteness of A^TA

The matrix $A^TA \in^{n\times n}$ is

1. *Square and symmetric:* $(A^TA)^T = A^TA$.
2. *Nonnegative definite:* for any x,
$$x^T(A^TA)x = \|Ax\|_2^2 \geq 0.$$

By the eigenvalue test, all eigenvalues of A^TA satisfy $\lambda \geq 0$.

Positive Semidefiniteness of A^TA

The matrix $A^TA \in^{n\times n}$ is:

1. *Square:* product of $(n \times m)(m \times n)$.
2. *Symmetric:* $(A^TA)^T = A^TA$.
3. *Nonneg. definite:* for any $x \in^n$, $x^T(A^TA)x = \|Ax\|_2^2 \geq 0$.

Hence all eigenvalues satisfy $\lambda \geq 0$.

Symmetry and Semidefiniteness of AA^T

$AA^T \in^{m\times m}$ is symmetric positive semidefinite and shares each nonzero eigenvalue of A^TA because

$$AA^T(A\,v) \;=\; A(A^TA)\,v \;=\; \lambda\,A\,v$$

whenever $A^TA\,v = \lambda v$.

Singular Value Decomposition (SVD)

Any $A \in^{m\times n}$ of rank r factors as $A = U\,\Sigma\,V^T$, where U and V are orthogonal and

$$\Sigma = (\sigma_1, \dots, \sigma_r, 0, \dots, 0), \quad \text{the } \textbf{singular values} \text{ satisfy } \sigma_1 \geq \cdots \geq \sigma_r > 0.$$

Columns v_i of V solve $A^TA\,v_i = \sigma_i^2 v_i$; setting $u_i = A\,v_i/\sigma_i$ gives $AA^T\,u_i = \sigma_i^2 u_i$. Extending $\{v_i\}$ to a basis of $\ker(A)$ and $\{u_i\}$ to $\ker(A^T)$ completes V and U, so that

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

Low-rank Approximation via SVD

Starting from the SVD of rank r,

$$A \;=\; \sum_{i=1}^{r} \sigma_i\,u_i v_i^T, \quad \sigma_1 \geq \cdots \geq \sigma_r > 0,$$

the **best rank-k approximation** (in any unitarily invariant norm) is

$$A_k \;=\; \sum_{i=1}^{k} \sigma_i\,u_i v_i^T, \quad \|A - A_k\| \;\leq\; \|A - B_k\|\,\forall\,(B_k) \leq k.$$

Key induced norms of A are

$$\|A\|_2 = \sigma_1, \quad \|A\|_F = \Big(\sum_{i=1}^{r} \sigma_i^2\Big)^{1/2}, \quad \|A\|_* = \sum_{i=1}^{r} \sigma_i.$$

Randomized Numerical Linear Algebra

Randomization enables fast approximations for huge matrices.

> *Column–Row Sampling.* One samples columns of A and rows of B via a sparse matrix S (and its transpose $T = S^T$). For example, with
>
> $$A = [a_1, a_2, a_3], \quad B^T = [b_1^T, b_2^T, b_3^T], \quad S = \begin{pmatrix} s_{11} & 0 & 0 \\ 0 & 0 & s_{32} \end{pmatrix},$$
>
> we get
>
> $$AS = [\,s_{11}a_1,\; s_{32}a_3\,], \quad STB = \begin{pmatrix} s_{11}b_1^T \\ s_{32}b_3^T \end{pmatrix}.$$
>
> Although $SS^T \neq I$, one can choose sampling probabilities (e.g. $p_i \propto \|a_i\|\|b_{i:}\|$) so that
>
> $$\mathbb{E}[SS^T] = I, \quad \mathbb{E}[ASTB] = AB,$$
>
> minimizing the variance of the estimator.

Additional Resources

- MIT OpenCourseWare: `https://ocw.mit.edu` – courses *18.06 Introduction to Linear Algebra, 18.065 Linear Algebra and Learning from Data.*
- Gilbert Strang, "A 2020 Vision of Linear Algebra" (OCW resource).
- New text: *Linear Algebra for Everyone* (expected 2021).