성균관대학교
SUNG KYUN KWAN UNIVERSITY

Linear Algebra Review

(from CS229 Lecture Notes)

Eunbyung Park

Assistant Professor

School of Electronic and Electrical Engineering

Eunbyung Park (silverbottlep.github.io)

Basic Concepts

$$\begin{aligned} 4x_1 - 5x_2 &= -13 \\ -2x_1 + 3x_2 &= 9 \end{aligned}$$

$$A = \begin{bmatrix} 4 & -5 \\ -2 & 3 \end{bmatrix}, \qquad b = \begin{bmatrix} -13 \\ 9 \end{bmatrix}$$

$$Ax = b$$

Basic Notation

$$A \in \mathbb{R}^{m \times n}$$

$$x \in \mathbb{R}^{n}$$

Inner Products

$$x, y \in \mathbb{R}^n$$

$$x^\top y \in \mathbb{R}$$

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{i=1}^{n} x_i y_i$$

Outer Products

$$x, y \in \mathbb{R}^n$$

$$xy^\top \in \mathbb{R}^{n \times n}$$

Matrix Vector Products

$A \in \mathbb{R}^{m \times n}$

$x \in \mathbb{R}^{n}$

$Ax \in$

Matrix Matrix Products

$$A \in \mathbb{R}^{m \times n}$$

$$B \in \mathbb{R}^{n \times p}$$

$$AB \in$$

Matrix Matrix Products

- Associative
  - $(AB)C = A(BC)$
- Distributive
  - $A(B + C) = AB + AC$
- Not commutative
  - $AB \neq BA$

Identity Matrix and Diagonal Matrices

$$AI = A = IA$$

$$D = \mathrm{diag}(d_1, d_2, \ldots, d_n) =$$

The Transpose

$$(A^\top)_{ij} = A_{ji}$$

$$(A^\top)^\top = A$$
$$(AB)^\top = B^\top A^\top$$
$$(A+B)^\top = A^\top + B^\top$$

Symmetric Matrices

$$A \in \mathbb{R}^{n\times n}$$

$A = A^\top$ (symmetric)

$A = -A^\top$ (anti-symmetric)

$$A = \frac{1}{2}(A + A^\top) + \frac{1}{2}(A - A^\top)$$

The Trace

$$A \in \mathbb{R}^{n \times n}$$

$$\mathrm{tr}A = \sum_{i=1}^{n} A_{ii}$$

$$\mathrm{tr}A = \mathrm{tr}A^{\top}$$

$$\mathrm{tr}(cA) = \mathrm{ctr}A$$

$$\mathrm{tr}AB = \mathrm{tr}BA$$

$$\mathrm{tr}ABC = \mathrm{tr}BCA = \mathrm{tr}CAB$$

Norms

$$\|x\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x^\top x}$$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad \|x\|_\infty = \max_i |x_i|$$

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \qquad \|A\|_F = \sqrt{\sum_{i=1}^{n} A_{ij}^2} = \sqrt{\mathrm{tr}(A^\top A)}$$

Linear Independence and Rank

- A set of vectors $\{x_1, x_2, \dots, x_n\} \subset \mathbb{R}^m$ is said to be *linearly independent* if no vector can be represented as a linear combination of the remaining vectors

$$x_n = \sum_{i=1}^{n-1} \alpha_i x_i \qquad \text{(linearly dependent)}$$

- Geometrical interpretation

Linear Independence and Rank

- The *column rank* of a matrix $A \in \mathbb{R}^{m \times n}$ is the largest number of *columns* that constitute a linearly independent set
- The *row rank* of a matrix $A \in \mathbb{R}^{m \times n}$ is the largest number of *rows* that constitute a linearly independent set
- For any matrix $A \in \mathbb{R}^{m \times n}$ the *column rank* is equal to the *row rank*, so both quantities are referred to collectively as the *rank of A*.
- For $\mathrm{A} \in \mathbb{R}^{m \times n}$, $\mathrm{rank}(A) \leq \min(m, n)$. If $\mathrm{rank}(A) = \min(m, n)$, then A is said to be *full rank*

The Inverse of a Square Matrix

- The inverse of a square matrix
  - Non-square matrices do not have inverses by definition
  - A^{-1} may not exist: non-invertiable or singular (not full rank)

$$A^{-1}A = I = AA^{-1}$$

$$(A^{-1})^{-1} = A$$
$$(AB)^{-1} = B^{-1}A^{-1}$$
$$(A^{-1})^{\top} = (A^{\top})^{-1} = A^{-\top}$$

- For standard linear system, $Ax = b, x = A^{-1}b$
- What if A is not square?

Orthogonal Matrices

- If all its columns are orthogonal to each other and are normalized

$$x^\top y = 0 \qquad \text{(orthogonal)}$$

$$\|x\|_2 = 1 \qquad \text{(normalized)}$$

$$U^\top U = I = UU^\top \qquad \text{(orthogonal)}$$

$$U^\top = U^{-1}$$

Range and Nullspace of a Matrix

- The *span* of a set of vectors $\{x_1, \dots, x_n\}$ is the set of all vectors that can be expressed as a linear combination of $\{x_1, \dots, x_n\}$

$$\mathrm{span}(\{x_1, \dots, x_n\}) = \left\{ v: v = \sum_{i=1}^{n} \alpha_i x_i \,, \alpha_i \in \mathbb{R} \right\}$$

- The *range* (aka the column space) of a matrix $A \in \mathbb{R}^{m \times n}$ is the *span* of columns of A.

$$R(A) = \{v: v = Ax, x \in \mathbb{R}^n\}$$

- The *nullspace* of a matrix $A \in \mathbb{R}^{m \times n}$ is the set of all vectors that equal to 0 when multiplied by A

$$N(A) = \{x \in \mathbb{R}^n: Ax = 0\}$$

Quadratic Forms

- Given a square matrix $A \in \mathbb{R}^{n\times n}$ and a vector $x \in \mathbb{R}^n$, the scalar value $x^\top Ax$ is a quadratic form

$$x^\top Ax = \sum_{i=1}^{n} x_i (Ax)_i = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} A_{ij} x_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

- Does anti-symmetric part matter?

$$A = \frac{1}{2}(A + A^\top) + \frac{1}{2}(A - A^\top) = A_{sym} + A_{asym}$$

$$A_{asym} = -A_{asym}^\top$$

$$x^\top A_{asym} x = \left(x^\top A_{asym} x\right)^\top = -x^\top A_{asym}^\top x$$

Positive Semidefinite Matrices

- A symmetric matrix $A \in \mathbb{S}^n$ is *positive definite (PD)*, a.k.a $\mathbb{S}^n_{++}$
  - $x^\top A x > 0,$ for all non-zero vectors $x \in \mathbb{R}^n$
- A symmetric matrix $A \in \mathbb{S}^n$ is *positive semidefinite (PSD)*, a.k.a $\mathbb{S}^n_{+}$
  - $x^\top A x \geq 0,$ for all non-zero vectors $x \in \mathbb{R}^n$
- A symmetric matrix $A \in \mathbb{S}^n$ is *negative definite (ND)*
  - $x^\top A x < 0,$ for all non-zero vectors $x \in \mathbb{R}^n$
- A symmetric matrix $A \in \mathbb{S}^n$ is *seminegative definite (ND)*
  - $x^\top A x \leq 0,$ for all non-zero vectors $x \in \mathbb{R}^n$
- A symmetric matrix $A \in \mathbb{S}^n$ is *indefinite*
  - If there exists $x, y \in \mathbb{R}^n$ such that $x^\top A x > 0$ and $y^\top A y \leq 0$

Positive Semidefinite Matrices

- Positive definite matrices (or negative definite) are always full rank, invertible
- Prove by contradiction

$$a_j = \sum_{i \neq \mathrm{j}} x_i a_i$$

(linearly dependent)

If $x_{\mathrm{j}} = -1$, then $Ax = 0$, so $x^\top Ax = 0$

Gram Matrix

- For any matrix $A \in \mathbb{R}^{m \times n}$, gram matrix is symmetric
- And, *always positive semidefinite*

$$G = A^\top A$$

$$x^\top G x = \sum_{i=1}^{n} \sum_{j=1}^{n} G_{ij} x_i x_j = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i^\top a_j x_i x_j = \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i a_i)^\top (x_j a_j)$$

$$= \left(\sum_{i=1}^{n} x_i a_i \right)^\top \left(\sum_{j=1}^{n} x_j a_j \right) = \left\| \sum_{i=1}^{n} x_i a_i \right\|^2 \geq 0$$

Eigenvalues and Eigenvectors

- Given a square matrix $A \in \mathbb{R}^{n \times n}$, $\lambda \in \mathbb{C}$ is an eigenvalue of A and $x \in \mathbb{C}^n$ is the corresponding eigenvector if

$$Ax = \lambda x, \qquad x \neq 0$$

$$(A - \lambda I)x = 0, \qquad x \neq 0$$

- In order to have a non-zero solution, $A - \lambda I$ should be singular

$$|A - \lambda I| = 0$$

Eigenvalues and Eigenvectors

- $\text{trace}(A) = \sum_{i=1}^{n} \lambda_i$
- $\det(A) = \prod_{i=1}^{n} \lambda_i$
- The rank of A is equal to the number of non-zero eigenvalues of A

Eigenvalues and Eigenvectors of Symmetric Matrices

- All eigenvalues of A are real numbers
- The eigenvectors corresponding to different eigenvalues are orthogonal

$$Ax = \lambda x, \qquad Ay = \mu y, \qquad \lambda \neq \mu$$

$$\mu x^\top y = x^\top A y = y^\top A x = \lambda y^\top x$$

$\mu \neq \lambda$ implies that $x^\top y = 0$

Eigendecomposition of Symmetric Matrices

$$U \Lambda U^\top = A$$

$$U = \begin{bmatrix} | & | & | & | \\ u_1 & u_2 & \vdots & u_n \\ | & | & | & | \end{bmatrix} \text{ is orthonormal}$$

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Eigenvalues as Optimization

$$\max_x x^\top A x, \qquad \text{subject to } \|x\|_2^2 = 1$$

$$L(x, \lambda) = x^\top A x - \lambda(x^\top x - 1)$$

$$\nabla_x L(x, \lambda) = 2Ax - 2\lambda x = 0$$

$$Ax = \lambda x$$