Lecture 11: Quadratic Forms and Singular Value Decomposition

Baboo J. Cui, Yangang Cao

July 18, 2019

Contents

1 Symmetric and Skew Symmetric Matrices

$A \in \mathbb{R}^{n \times n}$ is **symmetric** if $A^T = A$.

Fact:
For a symmetric matrix A , all of its eigenvalues are real and all of its eigenvectors are orthogenal. • Symmetric A can be diagonalized by an orthogonal matrix Q:

$$Q^{-1}AQ = Q^T AQ = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$$

A is **skew symmetric** if $A^T = -A$.

Fact:
For a skew symmetric matrix A, all of its eigenvalues are purely imaginary.

- If n is odd, then A has at least a zero eigenvalue
- Skew symmetric A can be diagonalized (by a unitary matrix)5

2 Quadratic Forms

Quadratic form corresponding to a symmetric $A \in \mathbb{R}^{n \times n}$ is the function

$$f(x) = \langle x, Ax \rangle = x^T Ax, \forall x \in \mathbb{R}^n$$

- **Uniqueness of matrix representation**: For symmetric A, $\tilde{A} \in \mathbb{R}^{n \times n}$

$$x^T Ax = x^T \tilde{A}x, \forall x \in \mathbb{R}^n \quad \Leftrightarrow \quad A = \tilde{A}$$

- Why limiting to symmetric A?

3 Bounds of Quadratic Forms

Given $A = A^T \in \mathbb{R}^{n \times n}$, with sorted (real) eigenvalues $\lambda_{\min} \leq \cdots \leq \lambda_{\max}$

Fact:
$\lambda_{min}||x||^2 \leq x^T Ax \leq \lambda_{max}||x||^2$, $\forall x \in \mathbb{R}^n$

4 Positive/Negative Definite Matrices

A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is called

- **positive semidefinite** if $x^T Ax \geq 0$ for all $x \in \mathbb{R}^n$
  • Denoted $A \succeq 0$
- **positive definite** if $x^T Ax > 0$ for all $x \in \mathbb{R}^n, x \neq 0$
  • Denoted $A \succ 0$
- **negative semidefinite** if $-A \succeq 0$
  • Denoted $A \preceq 0$
- **negative definite** if $-A \succ 0$
  • Denoted $A \prec 0$

5 Characterizing Positive Definite Matrices

That $A = A^T$ is positive semidefinite ($A \succeq 0$) is equivalent to

- All eigenvalues of A are nonnegative
- All principal minors of A are nonnegative
- $A = BB^T$ for some $B \in \mathbb{R}^{n \times m}$

6 Comparison of Symmetric Matrices

Given two symmetric matrices A and $\tilde{A}$ of the same dimension

- Denote $A \succeq \tilde{A}$ if $A - \tilde{A} \succeq 0$
- Denote $A \succ \tilde{A}$ if $A - \tilde{A} \succ 0$
- Denote $A \preceq \tilde{A}$ if $A - \tilde{A} \preceq 0$
- Denote $A \prec \tilde{A}$ if $A - \tilde{A} \prec 0$
- $A \succeq \tilde{A}$ means the quadratic forms $x^T A x \geq x^T \tilde{A} x, \forall x$
- $A \succ \tilde{A}$ means the quadratic forms $x^T A x > x^T \tilde{A} x, \forall x \neq 0$
- These relations are transitive, e.g., $A \succeq B$ and $B \succeq C \Rightarrow A \succeq C$

7 Ellipsoids

Each $A \succ 0$ is represented by an ellipsoid in $\mathbb{R}^n$ centered at the origin:

$$\mathcal{E}_A := \left\{ x \in \mathbb{R}^n | x^T A x \leq 1 \right\}$$

- Eigenvectors $v_i, i = 1, \ldots, n$, determine directions of semi-axes
- Eigenvalues $\lambda_i, i = 1, \ldots, n$, determine lengths of semi-axes: semi-axis along v_i has length $\left(\frac{1}{\sqrt{\lambda_i}}\right)$

8 (Induced) Matrix Norm

Given $A \in \mathbb{R}^{m \times n}$ and the Euclidean vector norms $\|\cdot\|$ on $\mathbb{R}^n$ and $\mathbb{R}^m$

Definition (Induced Matrix Norm):
The norm of the matrix $A \in \mathbb{R}^{m \times n}$ induced from the vector norm $\|\cdot\|$ is:

$$\|A\| := \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|}{\|x\|}$$

• Induced matrix norm is called the spectrum norm (or L^2 norm)

9 Properties of Induced Matrix Norm

Induced matrix norm is a norm on the vector space $\mathbb{R}^{m\times n}$:

- $\|\alpha A\| = |\alpha|\|A\|, \forall \alpha \in \mathbb{R}$
- $\|A+B\| \leq \|A\| + \|B\|$ (Triangle Inequality)
- $\|A\| = 0$ if and only if $A = 0$

Moreover, it has the additional properties:

- $\|Ax\| \leq \|A\|\|x\|, \forall x \in \mathbb{R}^n$
- $\|AB\| \leq \|A\| \cdot \|B\|$ (assume the product makes sense)

10 Characterizing L^2 Matrix Norm

For $A \in \mathbb{R}^{m\times n}$, $\|A\|^2 = \sup_{x\neq 0} \frac{\|Ax\|^2}{\|x\|^2} = \sup_{x\neq 0} \frac{x^\top A^T A x}{x^T x} = \lambda_{\max}\left(A^T A\right)$

Fact:
The L^2 norm of matrix $A \in \mathbb{R}^{m\times n}$ is

$$\|A\| = \sqrt{\lambda_{\max}\left(A^T A\right)}$$

11 Singular Value Decomposition (SVD)

Any matrix $A \in \mathbb{R}^{m\times n}$ can be decomposed as

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

- $U = \begin{bmatrix} u_1 & \dots & u_m \end{bmatrix} \in \mathbb{R}^{m\times m}$ is orthogonal: $U^T U = U U^T = I_m$
- $V = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \in \mathbb{R}^{n\times n}$ is orthogonal: $V^T V = V V^T = I_n$
- $\Sigma = \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m\times n}$ where $\Sigma_+ = \operatorname{diag}\left(\sigma_1, \ldots, \sigma_r\right)$ with $\sigma_1 \geq \cdots \geq \sigma_r > 0$ and $r = \operatorname{rank}(A)$

12 Finding U and V in SVD

$$A^T A = V\Sigma^T \Sigma V^T, \quad AA^T = U\Sigma\Sigma^T U^T$$

Fact:
The L^2 norm of $A \in \mathbb{R}^{m\times n}$ is its largest singular value:

$$\|A\| = \sigma_{\max}(A) = \sqrt{\lambda_{\max}\left(AA^T\right)} = \sqrt{\lambda_{\max}\left(A^T A\right)}$$

13 Constructive Proof of SVD

14 Transformation Interpretation of SVD

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T = U_1 \Sigma_+ V_1^T = \sum_{i=1}^{r} \sigma_r u_i v_i^T$$

Considered as a linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$

- maps v_1 to $\sigma_1 u_1$ (most sensitive input/output direction)
- ...
- maps v_r to $\sigma_r u_r$
- maps $v_{r+1}, \ldots, v_n$ to 0
- The range space of A is $\mathcal{R}(A) = \mathcal{R}(U_1) = \operatorname{span}\{u_1, \ldots, u_r\}$
- The null space of A is $\mathcal{N}(A) = \mathcal{R}(V_2) = \operatorname{span}\{v_{r+1}, \ldots, v_n\}$

15 Pseudo-Inverse

For any $A \in \mathbb{R}^{m \times n}$ with the SVD

$$A = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1 & V_2 \end{bmatrix}^T = U_1 \Sigma_+ V_1^T$$

its **pseudo-inverse** (or **Moore-Penrose** inverse) is defined as

$$A^\dagger := \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} \Sigma_+^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U_1 & U_2 \end{bmatrix}^T = V_1 \Sigma_+^{-1} U_1^T \in \mathbb{R}^{n \times m}$$

Properties of pseudo-inverse:

- $\left(A^\dagger\right)^\dagger = A$
- $(\alpha A)^\dagger = \alpha^{-1} A^\dagger$ for $\alpha \neq 0$
- $\left(A^T\right)^\dagger = \left(A^\dagger\right)^T$
- $AA^\dagger$ is the projection matrix onto $\mathcal{R}(A)$
- $A^\dagger A$ is the projection matrix onto $\mathcal{N}(A)^\perp$
- $AA^\dagger A = A$, and $A^\dagger A A^\dagger = A^\dagger$
- For nonsingular square matrix $A, A^\dagger = A^{-1}$

16 Least Square Solutions of Linear Equations

For $A \in \mathbb{R}^{m\times n}$, consider the linear equation

$$Ax = y$$

Its **least square solution** is defined as

$$x^* = \text{argmin}_x \|Ax - y\|^2$$

Fact:
A least square solution is given by $x^* = A^\dagger y$

In the special case when A is one-to-one (tall, $m \geq n,$ and full rank, $r = n$)

$$A^\dagger = \left(A^T A\right)^{-1} A^T$$

- In this case, $A^\dagger$ is a left inverse of $A : \left(A^\dagger A\right) = I_n$
- x^* is the only least square solution

17 Least Norm Solutions of Linear Equations

For $A \in \mathbb{R}^{m\times n}$, consider the linear equation

$$Ax = y$$

Assume $y \in \mathcal{R}(A)$. The **least norm solution** is defined as

$$x^* = \text{argmin}_{x \text{ such that } Ax=y} \|x\|$$

Fact:
The least norm solution is given by $x^* = A^\dagger y$

If in particular A is onto, i.e., A is fat $(m \leq n)$ and full rank $(r = m)$

$$A^\dagger = A^T \left(AA^T\right)^{-1}$$

• In this case, $A^\dagger$ is a right inverse of $A : \left(AA^\dagger\right) = I_m$

18 Lower Rank Approximations

Suppose $A \in \mathbb{R}^{m\times n}, \text{rank}(A) = r$, has SVD

$$A = U\Sigma V^T = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

Problem: Find a lower rank matrix $\hat{A}, \text{rank}(\hat{A}) = p < r$, such that $\hat{A} \simeq A$ in the sense that $\hat{A}$ is the solution of the following optimization problem:

$$\min_{\hat{A}} \|A - \hat{A}\|$$

Solution: The optimal rank-p approximator of A is

$$\hat{A} = \sum_{i=1}^{p} \sigma_i u_i v_i^T$$

Interpretation of singular values: singular value $\sigma_i, i = 1, \ldots, r$, is the distance of A to the nearest rank-$(r-i)$ matrix:

$$\sigma_i = \min\{\|A - \hat{A}\| \mid \mathrm{rank}(\hat{A}) = r - i\}$$