MATRIX ANALYSIS WITH A FOCUS ON INEQUALITIES

Zhekai Pang and Yuhong Zhang

(Supervised by Michael Cavers)

November 10th, 2021

Table of Contents

Introduction

Definition (Symmetric Matrix)

A matrix A is a symmetric matrix if $A^T = A$.

Definition (Orthogonal Matrix)

A square matrix A is an orthogonal matrix if $A^T = A^{-1}$.

Theorem (Spectral Decomposition for Symmetric Matrices)

Let A be an $n \times n$ real symmetric matrix. Let $\lambda_1, \ldots, \lambda_n$ be its eigenvalues. Let $v_1, \ldots, v_n$ be corresponding eigenvectors. Then $A = P\Lambda P^T$, where $P = [v_1, \cdots, v_n]$ and $\Lambda = diag[\lambda_1, \cdots, \lambda_n]$.

Remark

The matrix P above is an orthogonal matrix.

Introduction

Definition (Non-Negative Definite Matrix)

An $n \times n$ real symmetric matrix A is called non-negative definite if $\mathbf{x}^T A \mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

Theorem

An $n \times n$ real symmetric matrix A is non-negative definite if and only if all its eigenvalues are non-negative.

Definition (Positive Definite Matrix)

An $n \times n$ real symmetric matrix A is called positive definite if $\mathbf{x}^T A \mathbf{x} > 0$ for all nonzero $\mathbf{x} \in \mathbb{R}^n$.

Theorem

An $n \times n$ real symmetric matrix A is positive definite if and only if all its eigenvalues are positive.

Introduction

Example

Let A and B be two $n \times n$ real symmetric positive definite matrices. Let $k_1, k_2 \in \mathbb{R}^+$. Let $C = k_1 A + k_2 B$. Then $AB^{-1}A$, C and C^{-1} are all symmetric positive definite matrices.

Proof.

- Since A is a real symmetric positive definite, we know A is non-singular.
- So it is full rank, which implies that A has a trivial null space.
- Since B is positive definite, B^{-1} is also.
- Then $x^T(AB^{-1}A)x = (Ax)B^{-1}(Ax) = 0$ if and only if $Ax = 0$, if and only if $x = 0$.
- Thus, the matrix $AB^{-1}A$ is positive definite.

□

Introduction

Definition (Matrix Square Root)

Let A be a positive definite matrix. Then the square root of matrix A is $A^{\frac{1}{2}} = P\Lambda^{\frac{1}{2}}P^{-1}$.

Example

Let $A = \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix}$. Then $\lambda_1 = 4$ with $v_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\lambda_2 = 16$ with $v_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

We now have $A = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 16 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}^{-1}$.

Thus, $A^{\frac{1}{2}} = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 16 \end{bmatrix}^{\frac{1}{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$.

Theorem

Let A be a positive definite matrix and let $A^{\frac{1}{2}}$ be its positive square root. Let $A^{-\frac{1}{2}}$ be the inverse of $A^{\frac{1}{2}}$. Then $A^{\frac{1}{2}}$ is symmetric, and

$$A^{\frac{1}{2}}A^{\frac{1}{2}} = A, \quad A^{\frac{1}{2}}A^{-\frac{1}{2}} = I, \quad A^{-\frac{1}{2}}A^{-\frac{1}{2}} = A^{-1}.$$

Introduction

Definition (Hermitian Matrix)

A square matrix A is a Hermitian (or self-adjoint) matrix if $A = A^*$, which means it is equal to its own conjugate transpose.

Example

Let $A = \begin{bmatrix} 1 & 2+i \\ 2-i & 2 \end{bmatrix}$.

Take the complex conjugate: $\begin{bmatrix} 1 & 2-i \\ 2+i & 2 \end{bmatrix}$.

Take the transpose: $\begin{bmatrix} 1 & 2+i \\ 2-i & 2 \end{bmatrix}$.

The matrix A is a Hermitian matrix since $A^* = A$.

If A has real entries, then A is Hermitian if and only if it is symmetric.

Matrix Equality

- We sometimes use rank-k correction when new data is added to a model.
- By using the Woodbury Matrix Identity, we can do a rank-k correction to the inverse of the original matrix in order to compute the inverse of a rank-k correction of some matrix.

Theorem (Woodbury Matrix Identity)

Let A, X, R, and Y be complex matrices with size $n \times n$, $n \times r$, $r \times r$, and $r \times n$, respectively. Suppose that A, R, and $A + XRY$ are invertible. Then

$$(A + XRY)^{-1} = A^{-1} - A^{-1}X(R^{-1} + YA^{-1}X)^{-1}YA^{-1}.$$

Proof.

$(A + XRY)[A^{-1} - A^{-1}X(R^{-1} + YA^{-1}X)^{-1}YA^{-1}]$
$= I + XRYA^{-1} - X(R^{-1} + YA^{-1}X)^{-1}YA^{-1} - XRYA^{-1}X(R^{-1} + YA^{-1}X)^{-1}YA^{-1}$
$= I + XRYA^{-1} - (X + XRYA^{-1}X)(R^{-1} + YA^{-1}X)^{-1}YA^{-1}$
$= I + XRYA^{-1} - XR(R^{-1} + YA^{-1}X)(R^{-1} + YA^{-1}X)^{-1}YA^{-1}$
$= I + XRYA^{-1} - XRYA^{-1}$
$= I$ □

Theorem (Woodbury Matrix Identity)

Let A, X, R, and Y be complex matrices with size $n \times n$, $n \times r$, $r \times r$, and $r \times n$, respectively. Suppose that A, R, and $A + XRY$ are invertible. Then $(A + XRY)^{-1} = A^{-1} - A^{-1}X(R^{-1} + YA^{-1}X)^{-1}YA^{-1}$.

Corollary

Consider the special case that $n = r$ and $X = Y = I$. Then

$$\begin{aligned}(A+R)^{-1} &= A^{-1} - A^{-1}(R^{-1} + A^{-1})^{-1}A^{-1} \\ &= A^{-1} - A^{-1}(AR^{-1} + I)^{-1} \\ &= A^{-1} - (AR^{-1}A + A)^{-1}\end{aligned}$$

Matrix Inequality

Definition (Gersgorin Disc)

Let $A = [a_{ij}]$ be a complex $n \times n$ matrix. For $i \in \{1, \dots, n\}$, let R_i be the sum of the absolute values of the non-diagonal entries in the ith row, that is,

$$R_i = \sum_{j=1, j \neq i}^{n} |a_{ij}|.$$

The Gersgorin Disc (centered at a_{ii} with radius R_i) is

$$D_i = \{z \in \mathbb{C} : |z - a_{ii}| \leq R_i\}.$$

Theorem (Gersgorin Disc Theorem)

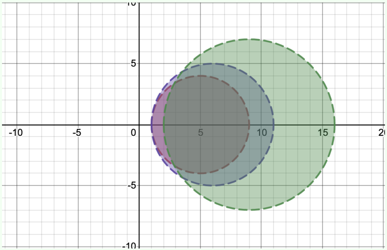
Every eigenvalue of a matrix lies within at least one Gershgorin disc.

Example

If $M = \begin{bmatrix} 5 & 1 & 3 \\ 3 & 6 & 2 \\ 4 & 3 & 9 \end{bmatrix}$ then the Gershgorin Discs are:

$D_1 = \{z \in \mathbb{C} : |z-5| \leq 4\},\ D_2 = \{z \in \mathbb{C} : |z-6| \leq 5\},\ D_3 = \{z \in \mathbb{C} : |z-9| \leq 7\}.$

We can draw the discs for M:



By the theorem, every eigenvalue lies within at least one discs.
The eigenvalues of M are $\{4, 2\left(4+\sqrt{5}\right), 2\left(4-\sqrt{5}\right)\}$.

Matrix Inequality

Theorem (Gersgorin Disc Theorem)

Every eigenvalue of a matrix lies within at least one Gershgorin disc.

Proof.

- Let λ be an eigenvalue of A with corresponding non-zero eigenvector x and suppose $|x_i| \geq |x_j|$ for all $j \in \{1, \ldots, n\}$.
- We know $Ax = \lambda x$. Then

$$Ax = \lambda x \implies \sum_{j=1}^{n} a_{ij} x_j = \lambda x_i \implies \sum_{j, j \neq i} a_{ij} x_j = (\lambda - a_{ii}) x_i.$$

- We divide both sides by x_i and take absolute value of previous expression:

$$|\lambda - a_{ii}| = \left| \sum_{j, j \neq i} a_{ij} \frac{x_j}{x_i} \right| \leq \sum_{j, j \neq i} |a_{ij}| \left| \frac{x_j}{x_i} \right| \leq \sum_{j, j \neq i} |a_{ij}| = R_i$$

□

Theorem

Let B be a $p \times p$ positive definite symmetric matrix and $b > 0$. Then

$$\frac{1}{(\det \Sigma)^b} e^{-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}B)} \leq \frac{1}{(\det B)^b} \left(\frac{2b}{e}\right)^{bp}$$

for all $p \times p$ positive definite matrix Σ. The equality holds only for $\Sigma = \frac{1}{2b}B$.

Proof.

- Let $B^{\frac{1}{2}}$ be the symmetric square root of B. Then $\mathrm{tr}(\Sigma^{-1}B) = \mathrm{tr}(\Sigma^{-1}B^{\frac{1}{2}}B^{\frac{1}{2}}) = \mathrm{tr}(B^{\frac{1}{2}}\Sigma^{-1}B^{\frac{1}{2}})$.
- Let λ_i be the eigenvalues of $B^{\frac{1}{2}}\Sigma^{-1}B^{\frac{1}{2}}$. Since the matrix is positive definite, $\lambda_i > 0$ for all i.
- $\sum_{i=1}^{p} \lambda_i = \mathrm{tr}(B^{\frac{1}{2}}\Sigma^{-1}B^{\frac{1}{2}}) = \mathrm{tr}(\Sigma^{-1}B)$
- $\prod_{i=1}^{p} \lambda_i = \det(B^{\frac{1}{2}}\Sigma^{-1}B^{\frac{1}{2}}) = \det(\Sigma^{-1}B) = \frac{\det(B)}{\det(\Sigma)}$.

  Thus, $\det(\Sigma) = \frac{\det(B)}{\prod_{i=1}^{p} \lambda_i}$.

Proof.

$$\begin{aligned}
\frac{1}{(\det\Sigma)^b} e^{-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}B)} &= \frac{\prod_{i=1}^{p}\lambda_i^b}{(\det B)^b} e^{-\frac{1}{2}\sum_{i=1}^{p}\lambda_i} \\
&= \frac{1}{(\det B)^b}\prod_{i=1}^{p}\lambda_i^b e^{-\frac{1}{2}\lambda_i} \\
&\le \frac{1}{(\det B)^b}\prod_{i=1}^{p}\sup_{\lambda_i\ge 0}\{\lambda_i^b e^{-\frac{1}{2}\lambda_i}\} \\
&= \frac{1}{(\det B)^b}\prod_{i=1}^{p}(2b)^b e^{-\frac{1}{2}\cdot 2b} \\
&= \frac{1}{(\det B)^b}(\frac{2b}{e})^{bp}.
\end{aligned}$$

The equality holds iff $\Sigma = \frac{1}{2b}B$. □

Remark

By applying the theorem, we can find the maximum likelihood estimators of multivariate normal distribution.

Matrix Inequality

- The Rayleigh–Ritz theorem is a numerical method of approximating eigenvalues and originated in the context of solving physical boundary value problems.

Definition (Rayleigh Quotient)

The Rayleigh quotient for a complex Hermitian matrix A and nonzero vector x is defined as

$$R(A, x) = \frac{x^* A x}{x^* x}.$$

Theorem (Rayleigh-Ritz Theorem)

Let A be a $n \times n$ symmetric matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. Then for $x \neq 0$,

$$\lambda_1 \leq R(A, x) \leq \lambda_n$$

$$\lambda_n = \max_{x \neq 0} \frac{x^T A x}{x^T x} = \max_{||x||=1} x^T A x.$$

$$\lambda_1 = \min_{x \neq 0} \frac{x^T A x}{x^T x} = \min_{||x||=1} x^T A x.$$

Example

- Let $A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$.
- Let $x_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, $x_2 = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$, and $x_3 = \begin{bmatrix} 4 \\ 6 \\ 8 \end{bmatrix}$.
- Then
$$R(A, x_1) = \frac{x_1^T A x_1}{x_1^T x_1} = \frac{70}{14} = 5,$$
- and
$$R(A, x_2) = \frac{x_2^T A x_2}{x_2^T x_2} = \frac{382}{77} \approx 4.961,$$
- and
$$R(A, x_3) = \frac{x_3^T A x_3}{x_3^T x_3} = \frac{584}{116} \approx 5.034.$$

These give lower bounds for the largest eigenvalue of A (note $\lambda_3 \approx 5.049$).

A generalization of the Rayleigh-Ritz Theorem is the Courant-Fischer Theorem.

Theorem (Courant-Fischer Theorem)

Let A be a symmetric $n \times n$ matrix. Let $\lambda_1 \leq \ldots \leq \lambda_n$ be its real eigenvalues and $v_1, \ldots, v_n$ be the corresponding eigenvectors. For $1 \leq k \leq n$, let $S_0 = \{0\}$, $S_k = \text{span}\{v_1, \cdots, v_k\}$ and $S_k^\perp$ be the orthogonal complement of S_k. Then

$$\lambda_k = \min_{||x||=1, x \in S_{k-1}^\perp} x^T A x = \min_{x \neq 0, x \in S_{k-1}^\perp} \frac{x^T A x}{x^T x}$$

Proof.

Since A is a symmetric matrix, we can let $A = Q\Lambda Q^T$ be the spectral decomposition, where Q is an orthogonal matrix. Thus $||Q^T x|| = ||x||$.
$x^T A x = x^T Q \Lambda Q^T x = (Q^T x)^T \Lambda (Q^T x)$
Thus, we just need to consider the case when A is a diagonal matrix.

Let $A = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$.

Then $x^T A x = \begin{bmatrix} x_1, \ldots, x_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \sum_{i=1}^{n} \lambda_i x_i^2$

Since A is a diagonal matrix, we know e_i is an eigenvector corresponding to λ_i.
If $x \in S_{k-1}^{\perp}$, then $x \perp e_i$ for $i \in \{1, \cdots, k-1\}$.
Thus $< x, e_i >= 0$ for $i \in \{1, \ldots, k-1\}$.
Thus $x_i =< x, e_i >= 0$ for $i \in \{1, \ldots, k-1\}$.

Matrix Inequality

Proof.

When $||x|| = 1$ and $x \in S_{k-1}^{\perp}$, we have

$$\begin{aligned} x^T A x &= \sum_{i=1}^{n} \lambda_i x_i^2 \\ &= \sum_{i=k}^{n} \lambda_i x_i^2 \text{ since } x_1 = \ldots = x_{k-1} = 0 \\ &\geq \sum_{i=k}^{n} \lambda_k x_i^2 \text{ since } \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n \\ &= \lambda_k \sum_{i=1}^{n} x_i^2 \\ &= \lambda_k \end{aligned}$$

Also, when $x = e_k$, $x^T A x = e_k^T A e_k = \lambda_k$.
Thus, $\lambda_k = \min_{||x||=1, x \in S_{k-1}^{\perp}} x^T A x$. Similarly, we know $\lambda_n = \max_{||x||=1} x^T A x$. □

Matrix Inequality

Theorem (Interlocking Eigenvalue Lemma)

Let A be a symmetric $n \times n$ matrix. Let $\lambda_1 \leq \ldots \leq \lambda_n$ be its real eigenvalues. Let $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_n$ be the eigenvalues of $A + bb^T$, where b is a vector in $\mathbb{R}^n$. Then

$$\lambda_1 \leq \mu_1 \leq \lambda_2 \leq \mu_2 \leq \cdots \leq \lambda_n \leq \mu_n.$$

The interlocking eigenvalue lemma compares the eigenvalues of the original matrix with the eigenvalues after adding a rank 1 matrix.

Example

Let $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Then $A + bb^T = \begin{bmatrix} 2 & 2 & 3 \\ 2 & 5 & 6 \\ 3 & 6 & 10 \end{bmatrix}$.

We have $\lambda_1 = \lambda_2 = \lambda_3 = 1$.

By the Interlocking Eigenvalue Lemma, $\mu_1 = \mu_2 = 1$.

Also,

$$\mu_3 = \mathrm{tr}(A + bb^T) - \mu_1 - \mu_2 = 15.$$

- In mathematics, an eigenvalue perturbation problem is that of finding the eigenvectors and eigenvalues of a system that is perturbed from one with known eigenvectors and eigenvalues. This is useful for studying how sensitive the original system's eigenvectors and eigenvalues are to changes.

Theorem (Weyl's Inequality)

Let A, B be $n \times n$ Hermitian matrices such that the eigenvalues of A, B and $A + B$ are $\lambda_i(A), \lambda_i(B)$ and $\lambda_i(A+B)$ arranged in increasing order, respectively. Then for each $k \in \{1, 2, \ldots, n\}$.

$$\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A+B) \leq \lambda_k(A) + \lambda_n(B).$$

When $k = 1$ we have:

$$\lambda_1(A) + \lambda_1(B) \leq \lambda_1(A+B) \leq \lambda_1(A) + \lambda_n(B).$$

When $k = n$ we have:

$$\lambda_n(A) + \lambda_1(B) \leq \lambda_n(A+B) \leq \lambda_n(A) + \lambda_n(B).$$

Matrix Inequality

Proof.

For any $0 \neq x \in \mathbb{C}^n$, by Rayleigh Quotient Theorem, $\lambda_1(B) \leq \frac{x^*Bx}{x^*x} \leq \lambda_n(B)$. Thus, for any $k \in \{1, 2, \dots, n\}$,

$$\begin{aligned}
\lambda_k(A+B) &= \min_{s_1, s_2, \dots, s_{n-k} \in \mathbb{C}^n} \max_{x \neq 0, x \perp s_1, s_2, \dots, s_{n-k}} \frac{x^*(A+B)x}{x^*x} \\
&= \min_{s_1, s_2, \dots, s_{n-k} \in \mathbb{C}^n} \max_{x \neq 0, x \perp s_1, s_2, \dots, s_{n-k}} \frac{x^*Ax}{x^*x} + \frac{x^*Bx}{x^*x} \\
&\geq \min_{s_1, s_2, \dots, s_{n-k} \in \mathbb{C}^n} \max_{x \neq 0, x \perp s_1, s_2, \dots, s_{n-k}} \frac{x^*Ax}{x^*x} + \lambda_1(B) \\
&= \lambda_k(A) + \lambda_1(B)
\end{aligned}$$

Matrix Inequality

Proof.

Similarly,

$$\begin{aligned}\lambda_k(A+B) &= \min_{s_1,s_2,\ldots,s_{n-k}\in\mathbb{C}^n} \max_{x\neq 0, x\perp s_1,s_2,\ldots,s_{n-k}} \frac{x^*(A+B)x}{x^*x} \\ &= \min_{s_1,s_2,\ldots,s_{n-k}\in\mathbb{C}^n} \max_{x\neq 0, x\perp s_1,s_2,\ldots,s_{n-k}} \frac{x^*Ax}{x^*x} + \frac{x^*Bx}{x^*x} \\ &\leq \min_{s_1,s_2,\ldots,s_{n-k}\in\mathbb{C}^n} \max_{x\neq 0, x\perp s_1,s_2,\ldots,s_{n-k}} \frac{x^*Ax}{x^*x} + \lambda_n(B) \\ &= \lambda_k(A) + \lambda_n(B)\end{aligned}$$

Thus,

$$\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A+B) \leq \lambda_k(A) + \lambda_n(B).$$

□

References

David G. Luenberger Yinyu Ye (2016)
Springer
Linear and Nonlinear Programming Fourth Edition.

Richard A. Johnson (2007)
Pearson Education.Inc
Applied Multivariate Statistical Analysis Sixth Edition.

Leslie Hogben (2007)
Chapman & Hall/ CRC
Handbook of Linear Algebra.

Frankel Theodore (2004)
Cambridge University Press
The Geometry of Physics: an introduction p. 652. ISBN 0-521-53927-7..

Leissa, A.W. (2005)
Journal of Sound and Vibration
The historical bases of the Rayleigh and Ritz methods,287 (4–5): 961–978.

Jonathan Kelner (2009)
MIT OpenCourseWare
18.409 Topics in Theoretical Computer Science: An Algorithmist's Toolkit.

Thank you!