Matrix Algebra Basics

1 Matrix operations

- For a complex number $z = x + iy$, its conjugate is $\bar{z} = x - iy$.
- The transpose of a m-by-n matrix $A = [a_{ij}]_{mn}$ is $A' = A^T = [a_{ji}]_{nm}$.
- A is symmetric if $A' = A$.
- The conjugate transpose or the Hermitian conjugate of a complex matrix $A = [a_{ij}]_{mn}$ is $A^H = A^\dagger = [\bar{a}_{ji}]_{nm}$.
- A is Hermitian if $A^H = A$.
- When two matrices are of the same dimensions $m \times n$, their sum $A + B = [a_{ij}]_{mn} + [b_{ij}]_{mn} = [a_{ij} + b_{ij}]_{mn}$.
- The product of two matrices of dimension $m \times n$ and $n \times p$ is

$$C = AB = [a_{ik}]_{mn}[b_{kj}]_{np} = [c_{ij}]_{mp}, \quad \text{with } (i,j)th \text{ entry} \quad c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

  Consequently, the product of three matrices is

$$H = A_{mn} B_{np} D_{pq} = [h_{ij}]_{mq}$$

  and the (i,j)th entry of the product matrix H is

$$h_{ij} = \sum_{\ell=1}^{p} \left(\sum_{k=1}^{n} a_{ik} b_{k\ell} \right) d_{\ell j} = \sum_{k=1}^{n} a_{ik} \left(\sum_{\ell=1}^{p} b_{k\ell} d_{\ell j} \right).$$

- The transpose of a sum is

$$(A + B)^T = A^T + B^T$$

- The transpose of a product are

$$(AB)^T = B^T A^T.$$

- Suppose that A is an $n \times n$ square matrix. B is an inverse of A, write as $B = A^{-1}$, if

$$BA = AB = I.$$

  where $I = I_n$ is the $n \times n$ identity matrix with 1's on the diagonal and 0 elsewhere.

- The inverse of a product

  If $A = A_{nn}$ is invertible and $A = B_{nn} C_{nn}$, then B, C are also invertible, and

$$A^{-1} = (BC)^{-1} = C^{-1} B^{-1}$$

- $A = A_{nn}$ is an orthogonal matrix if

$$A^{-1} = A^T,$$

  or equivalently,

$$AA^T = A^T A = I_n$$

- The trace of a square matrix $A = [a_{ij}]_{n \times n}$ is the sum of the diagonal elements.

$$tr(A) = a_{11} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}$$

  – Linear property of matrix trace

$$tr(A + B) = tr(A) + tr(B)$$
$$tr(cA) = c\ tr(A)$$

  where $c \in \mathbb{R}$ is a scalar.

  – Trace of product
  For square matrices A and B,

$$tr(AB) = tr(BA)$$

  – Trace and Frobenius norm of matrix
  For $A = [a_{ij}]_{n \times m}$,

$$tr(AA^T) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2 = \|A\|_F^2$$

2 Eigenvalues and eigenvectors

$x \in \mathbb{R}^p \setminus \{0\}$ is an eigenvector of matrix $A = [a_{ij}]_{p \times p}$ with eigenvalue λ if

$$A\boldsymbol{x} = \lambda \boldsymbol{x}$$

- The eigenvector of an eigenvalue is not unique (since $A\boldsymbol{x}_c = \lambda \boldsymbol{x}_c$ for $\boldsymbol{x}_c = c\boldsymbol{x}$, for any scalar c).
- All eigenvectors of the same eigenvalue form a subspace of $\mathbb{R}^p$, which is an eigenspace.
- Sometimes normalization $\|\boldsymbol{x}\| = 1$ is used, where $\|\boldsymbol{x}\|$ is the norm or length of the vector $\boldsymbol{x} = (x_1, \ldots, x_p)'$. The most common norm is the Euclidean norm

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

- Eigenvalues of matrix A are the roots of the characteristic polynomial of A, defined as

$$P(\lambda) = det(\lambda I_p - A)$$

  where det means the determinant of the matrix. (Sometimes, $P(\lambda) = det(A - \lambda I_p)$ is used.)

- In other words, eigenvalues are the solutions of the characteristic equation

$$det(A - \lambda I_p) = 0$$

- The characteristic polynomial can be factored as

$$P(\lambda) = \prod_{i=1}^{p} (\lambda - \lambda_i)$$

  with possible repeats of the factors (algebraic multiplicity) .

- From the expansion of the characteristic polynomial, we can obtain the following useful properties.

  – The determinant of a square matrix $A = A_{p\times p}$ is equal to the product of its eigenvalues,

$$det(A) = \lambda_1 \lambda_2 \cdots \lambda_p = \prod_{i=1}^{p} \lambda_i$$

  – The trace of a square matrix $A = A_{p\times p}$ is equal to the sum of its eigenvalues,

$$tr(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} \lambda_i$$

3 Symmetry, positive definiteness, and the spectral decomposition

Symmetric matrices and symmetric positive matrices are common matrices with many useful properties.

3.1 Symmetric matrix

If $A_{p\times p}$ is symmetric (or complex conjugate symmetric = Hermitian), then

- All eigenvalues are real. $\lambda_j \in \mathbb{R}, j = 1, \cdots, p$.
- There exists an orthonormal basis consists of eigenvectors $\boldsymbol{e}_j$ of λ_j:

$$\mathbf{A}\boldsymbol{e}_j = \lambda_j \boldsymbol{e}_j, \qquad \boldsymbol{e}_i \boldsymbol{e}_j = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \qquad i, j = 1, \cdots, p.$$

- Arrange the eigenvectors $\boldsymbol{e}_i$ as columns of a matrix,

$$A[\boldsymbol{e}_1\ \boldsymbol{e}_2\ \cdots\ \boldsymbol{e}_p] = [\lambda_1 \boldsymbol{e}_1\ \lambda_2 \boldsymbol{e}_2\ \cdots\ \lambda_p \boldsymbol{e}_p] = [\boldsymbol{e}_1\ \boldsymbol{e}_2\ \cdots\ \boldsymbol{e}_p] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & 0 & \lambda_p \end{bmatrix}$$

Let P denote the $p \times p$ matrix with the orthonormal eigenvectors $\boldsymbol{e}_i$ as the columns,

$$P = [\boldsymbol{e}_1\ \boldsymbol{e}_2\ \cdots\ \boldsymbol{e}_p]$$

then P is an orthogonal matrix $PP' = P'P = I$,
Let Λ be the $p \times p$ diagonal matrix with λ_i as the ith entry on the diagonal,

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & 0 & \lambda_p \end{bmatrix}, \qquad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p.$$

The above definitions yield

$$AP = P\Lambda$$

which leads to the eigenvalue-eigenvector decomposition of matrix A as

$$A = P\Lambda P'$$

which is also called the spectral decomposition of A.

- When matrix A has a spectral decomposition, it is often written as a decomposition as

$$A = \sum_{i=1}^{p} \lambda_i P_i, \qquad where \quad P_i = \boldsymbol{e}_i \boldsymbol{e}_i'$$

  with the following properties:

  – $P_i = \boldsymbol{e}_i \boldsymbol{e}_i'$ is a $p \times p$ orthogonal projection matrix to the 1-dimensional eigenspace $span\{\boldsymbol{e}_j\}$, where the span of vectors $span\{\cdots\}$ is the vector space consists of linear combination of the vectors.

  – By the property of orthogonal matrix, P_i is idempotent: $P_i P_i = P_i$.

  – By the property of orthogonal matrix, P_i's complementary $I - P_i$ is a projection matrix to $(span\{\boldsymbol{e}_i\})^{\perp}$, the orthogonal complement subspace of $span\{\boldsymbol{e}_j\}$, defined by $W^{\perp} = \{v : v'w = 0, \forall w \in W\}$.

Proof. To prove $A = \sum_{i=1}^{p} \lambda_i P_i$, first define $M = \sum_{i=1}^{p} \lambda_i P_i = \sum_{i=1}^{p} \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i'$. We will show $A = M$.
Recall $A\boldsymbol{e}_j = \lambda_j \boldsymbol{e}_j, \qquad \boldsymbol{e}_i \boldsymbol{e}_j = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \qquad i, j = 1, \cdots, p.$

Since $\{\boldsymbol{e}_i\}_{i=1,\cdots,p}$ forms a full set of orthonormal basis in $\mathbb{R}^p$, to show $A = M$ it is sufficient to show that $M\boldsymbol{e}_k = A\boldsymbol{e}_k$ for $k = 1, \cdots, p$. Now

$$M\boldsymbol{e}_k = \left(\sum_{i=1}^{p} \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i' \right) \boldsymbol{e}_k = \sum_{i=1}^{p} \lambda_i \boldsymbol{e}_i (\boldsymbol{e}_i' \boldsymbol{e}_k) = \sum_{i=1}^{p} \lambda_i \boldsymbol{e}_i \delta_{ik} = \lambda_k e_k = Ae_k, \qquad \forall k = 1, \cdots, p.$$

Therefore $A = M$. We have proved that $A = \sum_{i=1}^{p} \lambda_i P_i$.

□

Remarks

Another intuitive way to show the spectral decomposition can be via eigenvalue-eigenvector decomposition,

$$AP = A[\boldsymbol{e}_1\ \boldsymbol{e}_2\ \cdots\ \boldsymbol{e}_p] = [\lambda_1\boldsymbol{e}_1\ \lambda_2\boldsymbol{e}_2\ \cdots\ \lambda_p\boldsymbol{e}_p]$$

Using $PP' = I_n$,

$$A = [\lambda_1\boldsymbol{e}_1\ \lambda_2\boldsymbol{e}_2\ \cdots\ \lambda_p\boldsymbol{e}_p]\begin{bmatrix}\boldsymbol{e}_1'\\ \cdots\\ \boldsymbol{e}_p'\end{bmatrix}$$

Applying block matrix multiplication,

$$A = \sum_i \lambda_i\boldsymbol{e}_i\boldsymbol{e}_i'$$

3.2 positive definite matrix

Definition: A $p\times p$ symmetric matrix A is positive definite (p.d.) if $\boldsymbol{x}'A\boldsymbol{x} > 0$ for any p-vector $\boldsymbol{x}\neq\boldsymbol{0}$ (the zero vector in $\mathbb{R}^p$).

If A is p.d., then

- All eigenvalues $\lambda_j > 0$.
- A has an inverse matrix A^{-1} such that $AA^{-1} = A^{-1}A = I$,
  - A^{-1} is also p.d.
  - A^{-1} has eigenvalues $1/\lambda_i$, $i = 1,\cdots,p$.
  - If A has an orthonormal eigenvalue-eigenvector decomposition

    $$A = P\Lambda P'$$

    then A^{-1} also has an orthonormal eigenvalue-eigenvector decomposition, which can have the form

    $$A^{-1} = P\Lambda^{-1}P'$$

  - If A has a spectral decomposition as

    $$A = \sum_{i=1}^{p}\lambda_i\boldsymbol{e}_i\boldsymbol{e}_i' = \sum_{i=1}^{p}\lambda_i P_i$$

    then A^{-1} has a spectral decomposition in the form

    $$A^{-1} = \sum_{i=1}^{n}\frac{1}{\lambda_i}\boldsymbol{e}_i\boldsymbol{e}_i' = \sum_{i=1}^{p}\frac{1}{\lambda_i}P_i$$

- If $A = \sum_{i=1}^{p}\lambda_i\boldsymbol{e}_i\boldsymbol{e}_i' = P\Lambda^{-1}P'$, then we can define a "square-root matrix" of A as

  $$R = \sum_{i=1}^{n}\sqrt{\lambda_i}\boldsymbol{e}_i\boldsymbol{e}_i' = P\Lambda^{1/2}P'$$

where $\Lambda^{1/2}$ is the notation for the square-root matrix defined by

$$\Lambda^{1/2} = \begin{bmatrix}\sqrt{\lambda_1} & 0 & \cdots & 0\\ 0 & \sqrt{\lambda_2} & \cdots & 0\\ \vdots & & & \vdots\\ 0 & \cdots & 0 & \sqrt{\lambda_p}\end{bmatrix}$$

The definition of the square-root matrix is valid, because the eigenvalue matrix

$$\Lambda = \begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \cdots & 0\\ \vdots & & & \vdots\\ 0 & \cdots & 0 & \lambda_p\end{bmatrix}$$

has positive diagonals,

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$$

thank to the positive-definiteness.

Then matrix $R = P\Lambda^{1/2}P'$ is also symmetric p.d. with eigenvalues $\sqrt{\lambda_i} > 0$. Recall $PP' = I$,

$$RR = P\Lambda^{1/2}P'P\Lambda^{1/2}P' = P\Lambda^{1/2}\Lambda^{1/2}P'P\Lambda P' = A$$

We may using the notation $A^{1/2}$ to denote such matrix R,

$$A^{1/2} = R$$

Then, the notation reflects the relationship

$$A^{1/2}A^{1/2} = A$$

If A is symmetric and $\boldsymbol{x}'A\boldsymbol{x} \geq 0,\ \forall\boldsymbol{x}\neq\boldsymbol{0}$, then A is positive semi-definite or non-negative definite with $\lambda_j \geq 0$.

A symmetric matrix A is negative definite if $\boldsymbol{x}'A\boldsymbol{x} < 0$ for any vector $\boldsymbol{x}\neq\boldsymbol{0}$.

4 Matrix inequalities and maximization

The results in this section are used throughout the course, for example, in the derivation of the principal components in PCA and in the derivation of simultaneous confidence intervals derived from Hotelling's T^2.

4.1 Cauchy-Schwarz Inequality

Let $\boldsymbol{v},\boldsymbol{w}$ be vectors in a p-dimensional vector space such as $\mathbb{R}^n$ or $\mathbb{C}^n$. Here we consider $\mathbb{R}^n$ for convenience.

The inner product of $\boldsymbol{v}$ and $\boldsymbol{w}$ is denoted as $\boldsymbol{v}'\boldsymbol{w}$, and the corresponding vector norm $\|\boldsymbol{v}\|$ is defined as $\|\boldsymbol{v}\| = \sqrt{\boldsymbol{v}'\boldsymbol{v}}$.

The **Cauchy-Schwarz Inequality** is a theorem, stated as

$$(\boldsymbol{v}'\boldsymbol{w})^2 \leq \|\boldsymbol{v}\|^2\|\boldsymbol{w}\|^2 = (\boldsymbol{v}'\boldsymbol{v})(\boldsymbol{w}'\boldsymbol{w})$$

The equality holds if and only if $\boldsymbol{v} = c\boldsymbol{w}$ for some constant c.

Proof. The inequality is true if $\boldsymbol{w} = \boldsymbol{0} \in \mathbb{R}^n$, where $\boldsymbol{0}$ is the zero vector.

For the non-trivial case $\|\boldsymbol{w}\| > 0$, consider

$$\left\|\boldsymbol{v} - \frac{\boldsymbol{v}'\boldsymbol{w}}{\|\boldsymbol{w}\|^2}\boldsymbol{w}\right\|^2 = \left(\boldsymbol{v} - \frac{\boldsymbol{v}'\boldsymbol{w}}{\|\boldsymbol{w}\|^2}\boldsymbol{w}\right)'\left(\boldsymbol{v} - \frac{\boldsymbol{v}'\boldsymbol{w}}{\|\boldsymbol{w}\|^2}\boldsymbol{w}\right) = \|\boldsymbol{v}\|^2 - \frac{(\boldsymbol{v}'\boldsymbol{w})^2}{\|\boldsymbol{w}\|^2} \geq 0 \quad \Rightarrow \quad \|\boldsymbol{v}\|^2\|\boldsymbol{w}\|^2 \geq (\boldsymbol{v}'\boldsymbol{w})^2$$

So the Cauchy-Schwarz inequality follows. The quality holds if and only if $\boldsymbol{v} = \frac{\boldsymbol{v}'\boldsymbol{w}}{\|\boldsymbol{w}\|}\boldsymbol{w}$, and true for any multiple of such $\boldsymbol{v}$ by the scaling-invariance of the equality. Therefore, for any $\boldsymbol{v} = c\boldsymbol{w}, c \in \mathbb{R}$,

$$|\boldsymbol{v}'\boldsymbol{w}| = |c\boldsymbol{w}'\boldsymbol{w}| = |c| \cdot \|\boldsymbol{w}\|^2 = \sqrt{c^2\|\boldsymbol{w}\|^2\|\boldsymbol{w}\|^2} = \sqrt{\|\boldsymbol{v}\|^2\|\boldsymbol{w}\|^2}$$

which implies the equality in the Cauchy-Schwarz inequality.

□

4.2 Extension of Cauchy-Schwarz Inequality

Let B be a $p \times p$ symmetric positive definite matrix, $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^p$. The Extension of Cauchy-Schwarz Inequality asserts that

$$(\boldsymbol{v}'\boldsymbol{w})^2 \leq (\boldsymbol{v}'B\boldsymbol{v})(\boldsymbol{w}'B^{-1}\boldsymbol{w})$$

The equality holds if and only if $\boldsymbol{v} = cB^{-1}\boldsymbol{w}$ for some constant c.

Proof. By the symmetry and positive definiteness of B, there exists symmetric positive definite matrix R such that $B = R^2 = R'R$, and $B^{-1} = R^{-1}R'^{-1}$. Write

$$\boldsymbol{v}'\boldsymbol{w} = \boldsymbol{v}'R'R'^{-1}\boldsymbol{w} = (R\boldsymbol{v})'(R'^{-1}\boldsymbol{w}).$$

Apply Cauchy-Schwarz Inequality to $R\boldsymbol{v}$ and $R'^{-1}\boldsymbol{w}$,

$$\begin{aligned}[(R\boldsymbol{v})'(R'^{-1}\boldsymbol{w})]^2 &\leq [(R\boldsymbol{v})'(R\boldsymbol{v})]\,[(R\boldsymbol{w})'(R\boldsymbol{w})] \\ &= (\boldsymbol{v}'R'R\boldsymbol{v})(\boldsymbol{w}'R^{-1}R'^{-1}\boldsymbol{w}) = (\boldsymbol{v}'B\boldsymbol{v})(\boldsymbol{w}'B^{-1}\boldsymbol{w})\end{aligned}$$

From the Cauchy-Schwarz Inequality, the equality holds if and only if $R\boldsymbol{v} = c\,R'^{-1}\boldsymbol{w}$, which holds if and only if $\boldsymbol{v} = c\,R^{-1}R'^{-1}\boldsymbol{w} = cB^{-1}\boldsymbol{w}$.

□

4.3 The Maximization Lemma

The extended C-S inequality can be stated in the context of optimization. Let B be a symmetric positive definite matrix of dimension $p \times p$, let $\boldsymbol{w} \in \mathbb{R}^p$ be a given vector. The Maximization Lemma claims that

$$\max_{\boldsymbol{v}\neq 0} \frac{(\boldsymbol{v}'\boldsymbol{w})^2}{\boldsymbol{v}'B\boldsymbol{v}} = \boldsymbol{w}'B^{-1}\boldsymbol{w}$$

The maximum attained if and only if $\boldsymbol{v} = cB^{-1}\boldsymbol{w}$ for some constant $c \neq 0$.

Proof. The inequality

$$\frac{(\boldsymbol{v}'\boldsymbol{w})^2}{\boldsymbol{v}'B\boldsymbol{v}} \leq \boldsymbol{w}'B^{-1}\boldsymbol{w}, \qquad \forall \boldsymbol{v} \in \mathbb{R}^p$$

is directly from the Extension of Cauchy-Schwarz Inequality, and equality hold if and only if $\boldsymbol{v} = cB^{-1}\boldsymbol{w}$, which is when the maximum is attained:

$$\max_{\boldsymbol{v}\neq 0} \frac{(\boldsymbol{v}'\boldsymbol{w})^2}{\boldsymbol{v}'B\boldsymbol{v}} = \frac{c^2(\boldsymbol{w}'B^{-1}\boldsymbol{w})^2}{c^2(\boldsymbol{w}'B'^{-1})B(B^{-1}\boldsymbol{w})} = \frac{(\boldsymbol{w}'B^{-1}\boldsymbol{w})^2}{\boldsymbol{w}'B'^{-1}\boldsymbol{w}} = \boldsymbol{w}'B'^{-1}\boldsymbol{w}.$$

□

4.4 Applications of the Maximization Lemma

The maximization lemma has been applied broadly, utilizing the properties of symmetric positive definite matrices.

Recall that if matrix $C \in \mathbb{R}^{p\times p}$ is positive definite, then there is a spectral decomposition $C = \sum_{i=1}^p \lambda_i \boldsymbol{e}_i\boldsymbol{e}_i'$ by its eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p > 0$ and corresponding orthonormal eigenvectors $\boldsymbol{e}_i$'s.

For positive matrix C, the Maximization Lemma leads to the following useful corollaries.

- $\max\limits_{\boldsymbol{v}\in\mathbb{R}^p} \dfrac{\boldsymbol{v}'C\boldsymbol{v}}{\boldsymbol{v}'\boldsymbol{v}} = \max\limits_{\|\boldsymbol{v}\|=1} \boldsymbol{v}'C\boldsymbol{v} = \lambda_{\max} = \lambda_1$, attained by $\boldsymbol{v} = \boldsymbol{e}_1$.
- $\min\limits_{\boldsymbol{v}\in\mathbb{R}^p} \dfrac{\boldsymbol{v}'C\boldsymbol{v}}{\boldsymbol{v}'\boldsymbol{v}} = \min\limits_{\|\boldsymbol{v}\|=1} \boldsymbol{v}'C\boldsymbol{v} = \lambda_{\min} = \lambda_p$, attained by $\boldsymbol{v} = \boldsymbol{e}_p$.
- $\max\limits_{\substack{\boldsymbol{v}\perp\boldsymbol{e}_1,\cdots,\boldsymbol{e}_\ell \\ \boldsymbol{v}\in\mathbb{R}^p}} \dfrac{\boldsymbol{v}'C\boldsymbol{v}}{\boldsymbol{v}'\boldsymbol{v}} = \max\limits_{\substack{\boldsymbol{v}\perp\boldsymbol{e}_1,\cdots,\boldsymbol{e}_\ell \\ \|\boldsymbol{v}\|=1}} \boldsymbol{v}'C\boldsymbol{v} = \lambda_{\ell+1}$, attained by $\boldsymbol{v} = \boldsymbol{e}_{\ell+1}$.

These are known as the maximization inequality of quadratic forms, also called Rayleigh quotients (where the matrix can be complex Hermitian).

These corollaries are used widely, such as in the application to Fisher's discriminants in this course.

We provide a proof below.

Proof. Let $v_i, i = 1, \cdots, p$ be orthonormal eigenvalues of matrix C corresponding to eigenvalues λ_i,

$$Cv_i = \lambda_i v_i, \qquad v_i'v_j = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \qquad i, j = 1, \cdots, p.$$

The eigenvalue-eigenvector decomposition of matrix C can be written as

$$C = V\Lambda V'$$

where $p \times p$ matrices

$$\Lambda = diag\{\lambda_1, \cdots, \lambda_p\}, \qquad V = [v_1 \cdots v_p], \qquad VV' = V'V = I_p.$$

For any $\boldsymbol{x} \in \mathbb{R}^p$, let

$$\boldsymbol{y} = V\boldsymbol{x} = \begin{bmatrix} y_1 \\ \vdots \\ y_p \end{bmatrix}$$

Note that $C = C', \Lambda = \Lambda'$ are symmetric matrices, and $VV' = V'V = I$. Then the ratio

$$\frac{\boldsymbol{x}'C\boldsymbol{x}}{\boldsymbol{x}'\boldsymbol{x}} = \frac{x'V'\Lambda Vx}{x'VV'x} = \frac{y'\Lambda y}{y'y} = \frac{\sum_{i=1}^p \lambda_i y_i^2}{\sum_{i=1}^p y_i^2}$$

Since

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$$

we have

$$\sum_{i=1}^p \lambda_p y_i^2 \;\leq\; \sum_{i=1}^p \lambda_i y_i^2 \;\leq\; \sum_{i=1}^p \lambda_1 y_i^2$$

Therefore,

$$\lambda_p = \frac{\sum_{i=1}^p \lambda_p y_i^2}{\sum_{i=1}^p y_i^2} \le \frac{\sum_{i=1}^p \lambda_i y_i^2}{\sum_{i=1}^p y_i^2} \le \frac{\sum_{i=1}^p \lambda_1 y_i^2}{\sum_{i=1}^p y_i^2} = \lambda_1$$

That is

$$\lambda_p \ \le\ \frac{x'Cx}{x'x} \ \le\ \lambda_1, \qquad \forall x \in \mathbb{R}^p.$$

The equalities are achieved at $x = v_1$ and v_p respectively,

$$\frac{v_1'Cv_1}{v_1'v_1} = \frac{v_1'\lambda_1 v_1}{v_1'v_1} = \lambda_1, \qquad \frac{v_p'Cv_p}{v_p'v_p} = \frac{v_p'\lambda_p v_p}{v_p'v_p} = \lambda_p.$$

which proves the first two results.

The last result can be obtained by applying the first result to the subspace spanned by the eigenvectors $\{e_{k+1}, \cdots, e_p\}$. (Exercise)

□

5 Some matrix calculus

Recall from multivariate calculus:

If $f(x) : \mathbb{R}^n \to \mathbb{R}$ is a function defined on $\mathbb{R}^n$, the derivative with respect to the n-vector x is an n-vector itself, often written in the row form and commonly denoted as ∇f.

$$\frac{\partial f}{\partial x} = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}\right) \triangleq \nabla f.$$

If the function f itself is a p-vector of functions of n-vector x,

$$f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{bmatrix} \in \mathbb{R}^p, \qquad x \in \mathbb{R}^n,$$

then its derivative

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x} \\ \vdots \\ \frac{\partial f_p}{\partial x} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_p}{\partial x_1} & \cdots & \frac{\partial f_p}{\partial x_n} \end{bmatrix} \ \in\ \mathbb{R}^{p\times n}$$

is a $p \times n$ matrix. In the spacial case $n = 1$,

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{df_1}{dx} \\ \vdots \\ \frac{df_p}{dx} \end{bmatrix} \ \in\ \mathbb{R}^p$$

is a p-variate vector.

An application

Let A be a real $n \times n$ matrix, x be a vector in $\mathbb{R}^n$. Consider the quadratic form

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

which is a function $\mathbb{R}^n \to \mathbb{R}$. Its derivative (with respect to $x \in \mathbb{R}^n$) as a column vector can be expressed in the following neat form.

$$\frac{\partial}{\partial x}\left(x^T Ax\right) \ =\ A^T x + Ax$$

Proof.

Note that $\frac{\partial}{\partial x}\left(x^T Ax\right)$ is a vector in $\mathbb{R}^n$, with the kth component $\frac{\partial}{\partial x_k}\left(x^T Ax\right)$, for $k = 1, \cdots, n$.

Write out $x^T Ax$ in terms of the summation of component variable x_i's, the kth element of the derivative can be written as

$$\begin{aligned}
\frac{\partial}{\partial x_k}\left(x^T Ax\right) &= \frac{\partial}{\partial x_k}\left(\sum_{i=1}^n x_i \left(\sum_{j=1}^n a_{ij}x_j\right)\right) \\
&= \frac{\partial}{\partial x_k}\left(a_{kk}x_k^2 + x_k \sum_{j\ne k} a_{kj}x_j + \left(\sum_{i\ne k} x_i a_{ik}\right) x_k\right) \\
&= 2a_{kk}x_k + \sum_{j\ne k} a_{kj}x_j + \sum_{i\ne k} x_i a_{ik} \\
&= \sum_{j=1}^n a_{kj}x_j + \sum_{i=1}^n x_i a_{ik} \\
&= \left(k^{th}\ row\ of\ A^T\right) x + \left(k^{th}\ row\ of\ A\right) x \\
&= k^{th}\ element\ of\ (A^T x) + k^{th}\ element\ of\ (Ax)
\end{aligned}$$

The above holds for any $k = 1, \cdots, n$. We have shown $\frac{\partial}{\partial x}\left(x^T Ax\right) = A^T x + Ax$ by proving that

$$k^{th}\ element\ of\ \frac{\partial}{\partial x}\left(x^T Ax\right) \ =\ k^{th}\ element\ of\ (A^T x + Ax), \quad k = 1, \cdots, n.$$

□

A similar but simpler proof can show that $\frac{\partial}{\partial x}(Ax) = A$ (exercise).

Consequently, for symmetric matrix,

$$A = A^T$$

The derivatives are simplified,

$$\frac{\partial}{\partial x}\left(x^T Ax\right) = 2Ax, \qquad \frac{\partial^2}{\partial x \partial x^T}\left(x^T Ax\right) = 2A$$