Matrix Algebra Basics - a review

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 $A = A_{mn}$ is $A^H = A^\dagger = [\bar{a}_{ji}]_{nm}$.

A is Hermitian if $A^H = A$.

When two matrices are of the same dimension, their sum $A_{mn} + B_{mn} = [a_{ij}]_{mn} + [b_{ij}]_{mn} = [a_{ij} + b_{ij}]_{mn}$.

The product of two matrices of dimension $m \times n$ and $n \times p$ is

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

Consequently, the product of three matrices is

$$H = A_{mn}B_{np}D_{pq} = [h_{ij}]_{mq}, \qquad h_{ij} = \sum_{\ell=1}^{p}\left(\sum_{k=1}^{n} a_{ik}b_{k\ell}\right)d_{\ell j} = \sum_{k=1}^{n} a_{ik}\left(\sum_{\ell=1}^{p} b_{k\ell}d_{\ell j}\right).$$

The transpose of a sum is

$$(A + B)' = A' + B'$$

The transpose of a product are

$$(AB)' = B'A'.$$

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

$$BA = AB = I.$$

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

$$A^{-1} = A', \quad \text{or equivalently,} \quad AA' = A'A = I_n$$

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

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

with the linear property

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

where c is a scalar, and

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

For $A = [a_{ij}]_{n\times n}$,

$$tr(AA') = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2$$

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 $\boldsymbol{x}_c = cx$, any 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 A's characteristic polynomial

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

(sometimes, $P(\lambda) = det(A - \lambda I_p)$).

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

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

where det means the determinant of the matrix.

- The characteristic polynomial can 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 matrix $A = A_{p\times p}$ is the product of its eigenvalues,

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

The trace of a matrix $A = A_{p\times p}$ is 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

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 <u>eigenvalue-eigenvector decomposition</u> of matrix A as

$$A = P\Lambda P'$$

- $\mathbf{A}$ has a <u>spectral decomposition</u>

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

  with the following properties.

  - $P_i = \boldsymbol{e}_i\boldsymbol{e}_i'$ is a $p \times p$ <u>orthogonal projection matrix</u> (or a projector) to the 1-dimensional eigenspace $\{\boldsymbol{e}_j\}$;
  - P_i is idempotent: $P_iP_i = P_i$. Consequently,
  - P_i's complementary $I - P_i$ is a projector to $\{\boldsymbol{e}_i\}^{\perp}$.

  Proof. To show $A = \sum_{i=1}^{p} \lambda_i P_i$.
  Let $M = \sum_{i=1}^{p} \lambda_i P_i = \sum_{i=1}^{p} \lambda_i \boldsymbol{e}_i\boldsymbol{e}_i'$. Since $\{\boldsymbol{e}_i\}_{i=1,\cdots,p}$ form an 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 $k = 1, \cdots, p$.

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

  Therefore $A = M$. □

  Remarks: Another intuitive way to show the spectral decomposition can be

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

  using $PP' = I$ and block matrix multiplication.

Definition: A symmetric matrix A is <u>positive definite</u> (p.d.) if $\boldsymbol{x}'A\boldsymbol{x} > 0$ for any p-vector $\boldsymbol{x} \neq \mathbf{0}$ (the zero vector).

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 the eigenvalue-eigenvector decomposition $A = P\Lambda P'$ then $A = P\Lambda^{-1}P'$.
  - If A has spectral decomposition

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

    then

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

- 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

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

  Then R is also symmetric p.d. with eigenvalues $\sqrt{\lambda_i} > 0$, and $R^2 = A$.

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

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

4 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 $f(x) = \begin{bmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{bmatrix}$, $\mathbb{R}^n \to \mathbb{R}^p$ is a p-vector of functions of n-vector x, then $\frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f_1}{\partial x} \\ \vdots \\ \frac{\partial f_p}{\partial x} \end{bmatrix}$ is a $p \times n$ matrix.

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

Notice 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 the summation $x^T Ax$ in terms of x_k, the kth element of the derivative can be written as

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

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,

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

5 Matrix inequalities and maximization

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

5.1 Cauchy-Schwarz Inequality

Let $\boldsymbol{v}, \boldsymbol{w}$ be vectors in a p-dimensional vector space with inner product $\boldsymbol{v}'\boldsymbol{w}$ and norms $\|\boldsymbol{v}\|, \|\boldsymbol{w}\|$. Then

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

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

5.2 Extension of Cauchy-Schwarz Inequality

Let B be a $p \times p$ symmetric positive definite matrix. Then

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

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 $B = R'R$. Then $B^{-1} = R^{-1}R'^{-1}$. Write

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

By Cauchy-Schwarz Inequality,

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

□

5.3 Maximization Lemma

Let B be a symmetric positive definite matrix of dimension $p \times p$, let $\boldsymbol{w} \in \mathbb{R}^p$ be a given vector, then

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

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

Proof. The inequality with $\leq$ is directly from the Extension of Cauchy-Schwarz Inequality. When $\boldsymbol{v} = cB^{-1}\boldsymbol{w}$,

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

□

5.4 Applications of the Maximization Lemma

Recall that if B is positive definite then there is a spectral decomposition $B = \sum_{i=1}^{p} \lambda_i \boldsymbol{e}_i\boldsymbol{e}_i'$ by its eigenvalues $\lambda_1 \geq \cdots \geq \lambda_p > 0$ and orthonormal eigenvectors $\boldsymbol{e}_i$. The Maximization Lemma leads to the following useful results.

- $\max_{\|\boldsymbol{v}\|=1} \boldsymbol{v}'B\boldsymbol{v} = \lambda_{\max} = \lambda_1$, attained by $\boldsymbol{v} = \boldsymbol{e}_1$.
- $\min_{\|\boldsymbol{v}\|=1} \boldsymbol{v}'B\boldsymbol{v} = \lambda_{\min} = \lambda_p$, attained by $\boldsymbol{v} = \boldsymbol{e}_p$.
- $\max_{\substack{\boldsymbol{v}\perp\boldsymbol{e}_1,\cdots,\boldsymbol{e}_\ell \\ \|\boldsymbol{x}\|=1}} \boldsymbol{x}'B\boldsymbol{x} = \lambda_{\ell+1}$, attained by $\boldsymbol{v} = \boldsymbol{e}_{\ell+1}$.