MATH3841 Statistical Analysis of Dependent Data

Session 2, 2012

Review of Vectors and Matrices

School of Mathematics and Statistics
University of New South Wales

Vectors – basics

1. If $\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$, is an n-dimensional (column) vector, the **transpose** of $\boldsymbol{a}$ is a row vector
$\boldsymbol{a}^\top = (a_1, a_2, \cdots, a_n)$.
Also, if $\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$, $\boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$ then the sum is $\boldsymbol{a} + \boldsymbol{b} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}$.
The **inner product** of $\boldsymbol{a}$ and $\boldsymbol{b}$
$$\boldsymbol{a}^\top \boldsymbol{b} = \sum_{i=1}^{n} a_i b_i = \boldsymbol{b}^\top \boldsymbol{a}.$$
The inner product generalises the dot product for two vectors $X \cdot Y = |X||Y|\cos\theta$, where θ is the angle between X and Y. It is therefore measures

linear association. (Compare to correlation)

2. If c is a constant,
$$c\boldsymbol{a} = \begin{pmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_n \end{pmatrix}.$$
3. Let $\mathbf{0}$ denote a vector of zeros and $\mathbf{1}$ denote a vector of ones. This is sometimes written $\mathbf{0}_p$ and $\mathbf{1}_p$ to specify the length.

Matrices – basics

If A is a $p \times n$ matrix
$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{pmatrix} = (a_{ik})_{p \times n}$$
and c is a constant, then $cA = (ca_{ij})_{p \times n}$.

If B is $n \times m$, $AB = (\sum_{k=1}^{n} a_{ik} b_{kj})_{p \times m}$ and $AB \neq BA$ in general, even if of the same dimension.

Subject to dimension restrictions (Can you verify these?):

1. $A + B = B + A$
2. (A+B)+C= A+(B+C)

3. c(A+B)=cA+ cB
4. (AB)C= A(BC)
5. A(B+C)= AB + AC
6. (A+B)C= AC+ BC

Some useful notatation: Let $\boldsymbol{O_{p\times p}}$ denote a $p \times p$ matrix of zeros and $\boldsymbol{I_{p\times p}}$ denote a $p \times p$ matrix with diagonal elements all equal to 1 and all off-diagonal elements equal to 0.

$\boldsymbol{O_{p\times p}}$ is a **zero matrix** $\boldsymbol{I_p}$ is an **identity matrix:**
$\boldsymbol{O_{p\times p}}A$ is a $p \times n$ matrix of zeros
$\boldsymbol{I_p}A = A$ for any $p \times n$ matrix A.

D is a $p \times p$ **diagonal matrix** if

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_p \end{pmatrix} = \mathrm{Diag}(\lambda_1, \lambda_2, \cdots, \lambda_p).$$

The **transpose** of $A = (a_{ij})_{p\times n}$ is

$$A^\top = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{p1} \\ a_{12} & a_{22} & \cdots & a_{p2} \\ \vdots & & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{pn} \end{pmatrix}_{n\times p}.$$

Note (Can you show these?):

1. $(A^\top)^\top = A$
2. $(A+B)^\top = A^\top + B^\top$

3. $(AB)^\top = B^\top A^\top$

For example: Show $(A+B)^\top = A^\top + B^\top$.

Proof:
The $(i,j)^{th}$ element of $(A+B)^\top$ is $a_{ji} + b_{ji}$.
The $(i,j)^{th}$ element of $A^\top$ is a_{ji}, and of $B^\top$ is b_{ji}.
Hence the $(i,j)^{th}$ element of $A^\top + B^\top$ is $a_{ji} + b_{ji}$ as required.

If $A = (\boldsymbol{a_1}, \boldsymbol{a_2}, \cdots, \boldsymbol{a_n})$, then

$$A^\top = \begin{pmatrix} \boldsymbol{a_1}^\top \\ \boldsymbol{a_2}^\top \\ \vdots \\ \boldsymbol{a_n}^\top \end{pmatrix}.$$

Determinants of Matrices

The **determinant** of a square matrix A ($n \times n$ for some n) is

$$\begin{aligned} |A| &= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} |A^*_{ij}| \qquad \text{(for some } i\text{)} \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} |A^*_{ij}| \qquad \text{(for some } j\text{)} \end{aligned}$$

where A^*_{ij} is A, but with the i^{th} row and j^{th} column deleted, and if $B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $|B| = ad - bc$.
If $n = 1$, A is a scalar and $|A| = A$.

For example, choosing row $i = 1$:

$$\begin{vmatrix} 3 & -2 & 1 \\ 8 & 1 & -3 \\ -1 & 0 & 6 \end{vmatrix} = (-1)^2 \cdot 3 \begin{vmatrix} 1 & -3 \\ 0 & 6 \end{vmatrix} + (-1)^3 \cdot -2 \begin{vmatrix} 8 & -3 \\ -1 & 6 \end{vmatrix} + (-1)^4 \cdot 1 \begin{vmatrix} 8 & 1 \\ -1 & 0 \end{vmatrix}$$
$$= (3 \times 6) + (2 \times 45) + (1 \times 1) = 109.$$

Note (Can you show these?):

1. If A contains a row or column of zeros, $|A| = 0$.
2. $|A| = |A^\top|$.
3. If a row or column of A is multiplied by a constant c, $|A|$ is multiplied by c.
4. $|cA| = c^n|A|$.
5. $|\mathrm{Diag}(\lambda_1, \lambda_2, ..., \lambda_p)| = \prod_{i=1}^p \lambda_i$.
6. $|AB| = |A||B|$.

Trace of a Matrix

If $A = (a_{ij})_{n\times n}$, the **trace** of A is $\mathrm{tr}(A) = \sum_{i=1}^n a_{ii}$.

Note (Can you show these?):

1. $\mathrm{tr}(A) = \mathrm{tr}(A^\top)$.
2. $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$.
3. $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ if A is $m \times n$ and B is $n \times m$.
4. $\mathrm{tr}(A^\top A) = \sum_{j=1}^n \sum_{i=1}^n a_{ij}^2$.
5. $\mathrm{tr}(B^{-1}AB) = \mathrm{tr}(A)$
6. $\mathrm{tr}(cA) = c\mathrm{tr}(A)$.
7. $\mathrm{tr}(I_n) = n$.

Matrix Inverse

If A is $n \times n$ and $|A| \neq 0$, then A is **nonsingular**. There then exists a unique B such that $AB = I_n$. B is the **inverse** of A. Write $B = A^{-1}$.

Then (Can you show these?)

1. $AA^{-1} = A^{-1}A = I_n$.
2. $(A^{-1})^\top = (A^\top)^{-1}$.
3. $(AC)^{-1} = C^{-1}A^{-1}$ if both A and C are nonsingular and $n \times n$.
4. $|A^{-1}| = \frac{1}{|A|}$.
5. If $A = \mathrm{Diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$ with all $\lambda_i \neq 0$, then $A^{-1} = \mathrm{Diag}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \cdots, \frac{1}{\lambda_n}\right)$.

For Example: Show that $(AC)^{-1} = C^{-1}A^{-1}$ if both A and C are nonsingular and $n \times n$.

Proof: We note that

$$(AC)C^{-1}A^{-1} = A(CC^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n$$

as required.

If A is $n \times n$ and $A^\top = A$ ($a_{ij} = a_{ji}$ for all i, j), A is **symmetric**.

More on vectors

If $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, the **length** of $\boldsymbol{x}$ is $|\boldsymbol{x}| = \sqrt{x_1^2 + x_2^2}$.

If $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, $|\boldsymbol{x}| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}}$.

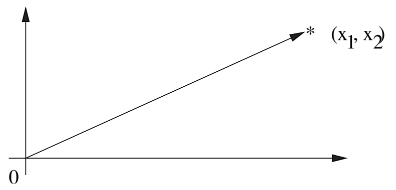

Figure 1: Length $|\boldsymbol{x}|$ of a vector $\boldsymbol{x}$.

12 If c is constant, $|c\boldsymbol{x}| = |c||\boldsymbol{x}|$.

$\cos\theta_1 = \frac{x_1}{|\boldsymbol{x}|}$, $\cos\theta_2 = \frac{y_1}{|\boldsymbol{y}|}$, $\sin\theta_1 = \frac{x_2}{|\boldsymbol{x}|}$, $\sin\theta_2 = \frac{y_2}{|\boldsymbol{y}|}$ and

$$\cos\theta = \cos(\theta_2 - \theta_1) = \cos\theta_2\cos\theta_1 + \sin\theta_2\sin\theta_1 = \frac{x_1y_1 + x_2y_2}{|\boldsymbol{x}||\boldsymbol{y}|}.$$

Therefore the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$ (or $\boldsymbol{x}^\top$ and $\boldsymbol{y}^\top$) is specified by

$$\cos\theta = \frac{\boldsymbol{x}^\top\boldsymbol{y}}{|\boldsymbol{x}||\boldsymbol{y}|} = \frac{\boldsymbol{x}^\top\boldsymbol{y}}{\sqrt{\boldsymbol{x}^\top\boldsymbol{x}}\sqrt{\boldsymbol{y}^\top\boldsymbol{y}}}.$$

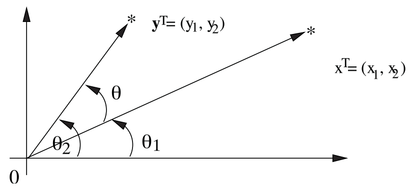

Figure 2: Angle relations.

13 The same formula applies if $\boldsymbol{x} = (x_1\ x_2\ \ldots x_n)^T$ and $\boldsymbol{y} = (y_1\ y_2\ \ldots y_n)^T$.

Do you see the connection between $\cos\theta$ and correlation?

$\boldsymbol{x}$ and $\boldsymbol{y}$ are **orthogonal** (that is, at right-angles to eachother) if $\boldsymbol{x}^\top\boldsymbol{y} = 0 \Longleftrightarrow \theta = \frac{\pi}{2}$.

The **projection** of $\boldsymbol{x}$ on $\boldsymbol{y}$ is a vector

$$\boldsymbol{z} = \left(\frac{\boldsymbol{x}^\top\boldsymbol{y}}{\boldsymbol{y}^\top\boldsymbol{y}}\right)\boldsymbol{y}.$$

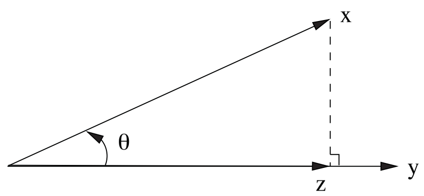

Figure 3: Vector projection

14 Now $|\boldsymbol{z}| = \cos\theta\,|\boldsymbol{x}| = \frac{\boldsymbol{x}^\top\boldsymbol{y}}{|\boldsymbol{y}|}$ and $\boldsymbol{z} = c\boldsymbol{y}$. Thus

$$|\boldsymbol{z}| = c|\boldsymbol{y}| \Rightarrow c = \frac{\boldsymbol{x}^\top\boldsymbol{y}}{|\boldsymbol{y}|^2} = \frac{\boldsymbol{x}^\top\boldsymbol{y}}{\boldsymbol{y}^\top\boldsymbol{y}}.$$

That is, the projection quantifies the degree of similarity between two vectors. Correlation!

A set of p-vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_k$ are **linearly dependent** if there exist constants $c_1, c_2, \cdots, c_k$, not all zero, such that

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_k\boldsymbol{x}_k = \boldsymbol{0}.$$

Otherwise the vectors are **linearly independent**.

Example 0.1 : $\boldsymbol{x}_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \boldsymbol{x}_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \boldsymbol{x}_3 = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$

15

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + c_3\boldsymbol{x}_3 = \boldsymbol{0} \implies \left.\begin{array}{lll} c_1 + c_2 & +c_3 = 0 \\ 2c_1 & -2c_3 = 0 \\ c_1 - c_2 & +c_3 = 0 \end{array}\right\} \implies c_1 = c_2 = c_3 = 0.$$

Hence, $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are linearly independent.

Note: Linear dependence means any one of the vectors can be written as a linear combination of the others.

More on Matrices

If A is an $n \times n$ matrix, then

1. A^{-1} exists if and only if the n columns of A are linearly independent (Do you see why?).
2. A^{-1} is **symmetric** if A is symmetric.
3. A is **orthogonal** if $AA^\top = A^\top A = I_n$. (Do you see why?)
   If $A = (\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_n)$ is orthogonal, then A^{-1} exists since $|A| \neq 0$ and
   (a) $A^{-1} = A^\top$.
   (b) $\boldsymbol{a}_i^\top \boldsymbol{a}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$
   (c) $|A| = 1$ or -1.
   (d) If B is also orthogonal, so is AB as:

$$AB(AB)^\top = ABB^\top A^\top = AA^\top = I_n.$$

Eigenvalues and Eigenvectors

The **eigenvalues** $\lambda_1, \lambda_1, \cdots, \lambda_n$ of A are the n roots of $|A - \lambda I_n| = 0$.
If $|A - \lambda_i I_n| = 0$, then $A - \lambda_i I_n$ is singular and hence from (1 on previous slide) the columns of $A - \lambda_i I_n$ are linearly dependent. Thus there exist constants (not all zero) $x_1, x_2, \cdots, x_n$ such that

$$(A - \lambda_i I_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \mathbf{0}$$

since if $B = (\boldsymbol{b}_1, \boldsymbol{b}_2, \cdots, \boldsymbol{b}_n)$, then

$$B \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 \boldsymbol{b}_1 + x_2 \boldsymbol{b}_2 + \cdots + x_n \boldsymbol{b}_n;$$

That is, multiplying a matrix on the right by a vector gives a linear combination of the columns of the matrix. Thus corresponding to each λ_i is a vector $\boldsymbol{x}_i$ for which $(A - \lambda_i I_n)\boldsymbol{x}_i = \mathbf{0}$ or $A\boldsymbol{x}_i = \lambda_i \boldsymbol{x}_i$ and $\boldsymbol{x}_i$ is an **eigenvector** corresponding to λ_i. If $\boldsymbol{e}_i = \boldsymbol{x}_i / |\boldsymbol{x}_i|$, then the $\boldsymbol{e}_i$'s are eigenvectors with unit length.

1. If A is symmetric, the λ_i's are all real (i.e. non-complex) and
   (a) Eigenvectors corresponding to distinct eigenvalues are orthogonal,
   (b) If $B = PAP^{-1}$ then A and B have the same eigenvalues,
   (c) There exists an orthogonal matrix Γ such that

$$\Gamma^\top A \Gamma = \mathrm{Diag}(\lambda_1, \lambda_2, \cdots, \lambda_n).$$

   Note that $A\Gamma = \Gamma \mathrm{Diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, since $(\Gamma^\top)^{-1} = \Gamma$. Hence if $\Gamma = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n)$, then $A\boldsymbol{\gamma}_i = \lambda_i \boldsymbol{\gamma}_i$ since if

$$A = \begin{pmatrix} \boldsymbol{a}_1^\top \\ \boldsymbol{a}_2^\top \\ \vdots \\ \boldsymbol{a}_n^\top \end{pmatrix}, \qquad A\Gamma = \begin{pmatrix} \boldsymbol{a}_1^\top \boldsymbol{\gamma}_1 & \cdots & \boldsymbol{a}_1^\top \boldsymbol{\gamma}_n \\ \vdots & & \vdots \\ \boldsymbol{a}_n^\top \boldsymbol{\gamma}_1 & \cdots & \boldsymbol{a}_n^\top \boldsymbol{\gamma}_n \end{pmatrix}$$

and the i^{th} column of $A\Gamma$ is

$$\begin{pmatrix} \boldsymbol{a}_1^\top \boldsymbol{\gamma}_i \\ \vdots \\ \boldsymbol{a}_n^\top \boldsymbol{\gamma}_i \end{pmatrix} = A\boldsymbol{\gamma}_i$$

and the i^{th} column of

$$\Gamma \mathrm{Diag}(\lambda_1, \cdots, \lambda_n) \text{ is } \Gamma \begin{pmatrix} 0 \\ \vdots \\ \lambda_i \\ \vdots \end{pmatrix} = \lambda_i \boldsymbol{\gamma}_i.$$

Thus $\boldsymbol{\gamma}_i$ is the unit length eigenvector corresponding to λ_i.

We also write

$$A = \Gamma\Lambda\Gamma^\top,$$

where $\Lambda = \mathrm{Diag}(\lambda_1, \cdots, \lambda_n)$.

(d) $|A| = \prod_{i=1}^n \lambda_i$.

(e) $\mathrm{tr}(A) = \sum_{i=1}^n \lambda_i$.

Example 0.2 : $A = \begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix}$.

$$|A - \lambda I| = \begin{vmatrix} 1-\lambda & -5 \\ -5 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 25$$

$$\Longrightarrow \lambda^2 - 2\lambda - 24 = 0 \Longrightarrow \text{eigenvalues } \lambda_1 = 6, \lambda_2 = -4.$$

$$A\boldsymbol{x} = \lambda_1 \boldsymbol{x} \Leftrightarrow \begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 6\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$\Leftrightarrow \left.\begin{array}{rrl} x_1 & -5x_2 & = 6x_1 \\ -5x_1 & +x_2 & = 6x_2 \end{array}\right\} \Longrightarrow x_1 = -x_2.$$

So every vector (except $\mathbf{0}$) of the form $\begin{pmatrix} a \\ -a \end{pmatrix}$ is an eigenvector corresponding to λ_1, so a unit length eigenvector is $\boldsymbol{e}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$.

Similarly, $\lambda_2 = -4$ and $\boldsymbol{e}_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$.

Note: $\boldsymbol{e}_1^\top \boldsymbol{e}_2 = 0$ and if $\Gamma = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$,

$$\Gamma^\top A \Gamma = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & -5 \\ -5 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 0 \\ 0 & -4 \end{pmatrix}.$$

Also, $|A| = -24 = \lambda_1 \times \lambda_2$ and $\mathrm{tr}(A) = 2 = \lambda_1 + \lambda_2$.

(f) The **spectral decomposition** of A is given by

$$A = \lambda_1 \boldsymbol{e}_1 \boldsymbol{e}_1^\top + \lambda_2 \boldsymbol{e}_2 \boldsymbol{e}_2^\top + \cdots + \lambda_n \boldsymbol{e}_n \boldsymbol{e}_n^\top$$

where $(\lambda_i, \boldsymbol{e}_i)$ are the eigenvalue, unit eigenvector pairs.

Proof: from 5(c), $A\boldsymbol{e}_i = \lambda_i \boldsymbol{e}_i$ and so $A\boldsymbol{e}_i\boldsymbol{e}_i^\top = \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^\top$ and hence

$$A\left(\sum_{i=1}^n \boldsymbol{e}_i \boldsymbol{e}_i^\top\right) = \sum_{i=1}^n \lambda_i \boldsymbol{e}_i \boldsymbol{e}_i^\top,$$

so we need to prove that $\sum_{i=1}^n \boldsymbol{e}_i \boldsymbol{e}_i^\top = I_n$. Now

$$\boldsymbol{e}_i \boldsymbol{e}_i^\top = \begin{pmatrix} e_{1i} \\ e_{2i} \\ \vdots \\ e_{ni} \end{pmatrix}(e_{1i}\, e_{2i} \cdots e_{ni}) = \begin{pmatrix} e_{1i}^2 & e_{1i}e_{2i} & \cdots & e_{1i}e_{ni} \\ e_{2i}e_{1i} & e_{2i}^2 & \cdots & e_{2i}e_{ni} \\ \vdots & & \ddots & \vdots \\ e_{ni}e_{1i} & e_{ni}e_{2i} & \cdots & e_{ni}^2 \end{pmatrix}$$

and

$$\Gamma = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n) = \begin{pmatrix} e_{11} & e_{12} & \cdots & e_{1n} \\ e_{21} & e_{22} & \cdots & e_{2n} \\ \vdots & & \ddots & \vdots \\ e_{n1} & e_{n2} & \cdots & e_{nn} \end{pmatrix}$$

and $\Gamma\Gamma^\top = I_n$. Now $(\Gamma\Gamma^\top)_{kk} = \sum_{i=1}^n e_{ki}^2$ and, for $k \neq j$, $(\Gamma\Gamma^\top)_{kj} = \sum_{i=1}^n e_{ki}e_{ji}$. Therefore

$$\sum_{i=1}^n \boldsymbol{e}_i \boldsymbol{e}_i^\top = \Gamma\Gamma^\top = I_n.$$

(g) If A is symmetric and the **quadratic form** $\boldsymbol{x}^\top A \boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$, then A is **positive definite**. If $\boldsymbol{x}^\top A \boldsymbol{x} \geq 0$ for all $\boldsymbol{x} \neq \mathbf{0}$, then A is **nonnegative definite**.

Example 0.3 : $A = \begin{pmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{pmatrix}$.

Let $\boldsymbol{x} = (x_1\, x_2)$, then $\boldsymbol{x}^\top A \boldsymbol{x} = 3x_1^2 + 2x_2^2 - 2\sqrt{2}x_1x_2$. We obtain $\lambda_1 = 4$, $\lambda_2 = 1$, so

$$A = 4\boldsymbol{e}_1\boldsymbol{e}_1^\top + \boldsymbol{e}_2\boldsymbol{e}_2^\top$$

and

$$\begin{aligned}\boldsymbol{x}^\top A\boldsymbol{x} &= 4\boldsymbol{x}^\top \boldsymbol{e}_1\boldsymbol{e}_1^\top\boldsymbol{x} + \boldsymbol{x}^\top\boldsymbol{e}_2\boldsymbol{e}_2^\top\boldsymbol{x} \\ &= 4\left(\boldsymbol{e}_1^\top\boldsymbol{x}\right)^2 + \left(\boldsymbol{e}_2^\top\boldsymbol{x}\right)^2 = 4y_1^2 + y_2^2,\end{aligned}$$

where $y_1 = \boldsymbol{e}_1^\top\boldsymbol{x}$ and $y_2 = \boldsymbol{e}_2^\top\boldsymbol{x}$. Thus, if y_1 and y_2 are not both zero, then $\boldsymbol{x}^\top A\boldsymbol{x} > 0$ for all $\boldsymbol{x} \neq \mathbf{0}$ and A is positive definite. Now

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{e}_1^\top \\ \boldsymbol{e}_2^\top \end{pmatrix}\boldsymbol{x} = \Gamma^\top\boldsymbol{x}$$

and if $\boldsymbol{x} \neq \mathbf{0}$ then $\boldsymbol{y} \neq \mathbf{0}$ since $\Gamma^\top$ is nonsingular. Thus, A is positive definite.

(h) If A is positive definite, $\lambda_i > 0$ for all i.

Proof: $A\boldsymbol{e}_i = \lambda_i\boldsymbol{e}_i$ implies

$$\boldsymbol{e}_i^\top A\boldsymbol{e}_i = \lambda_i\boldsymbol{e}_i^\top\boldsymbol{e}_i = \lambda_i > 0$$

since A is positive definite. Similarly, if A is nonnegative definite, $\lambda_i \geq 0$ for all i.

2. The square root of a matrix: if A is positive definite, then

$$A = \sum_{i=1}^{n} \lambda_i\boldsymbol{e}_i\boldsymbol{e}_i^\top = \Gamma\Lambda\Gamma^\top,$$

where $\Gamma = (\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n)$ and $\Lambda = \mathrm{Diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$.

Proof: The first equality is the spectral decomposition of A. The second equality follows since Γ diagonalizes A; that is, $\Gamma^\top A\Gamma = \Lambda$ and, since Γ is orthogonal, Γ^{-1} exists and $(\Gamma^\top)^{-1} = \Gamma$ and $\Gamma^{-1} = \Gamma^\top$.

Now $\lambda_i > 0$ for all i since A is positive definite, so Λ is nonsingular and $\Lambda^{-1} = \mathrm{Diag}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \cdots, \frac{1}{\lambda_n}\right)$. Thus $A^{-1} = \Gamma\Lambda^{-1}\Gamma^\top$ since

$$(\Gamma\Lambda^{-1}\Gamma^\top)(\Gamma\Lambda\Gamma^\top) = \Gamma\Lambda^{-1}\Lambda\Gamma^\top = \Gamma\Gamma^\top = I_n$$

and $A^{-1} = \sum_{i=1}^n \frac{1}{\lambda_i}\boldsymbol{e}_i\boldsymbol{e}_i^\top$ since $A = \Gamma\Lambda\Gamma^\top = \sum_{i=1}^n \lambda_i\boldsymbol{e}_i\boldsymbol{e}_i^\top$.

Let $\Lambda^{1/2} = \mathrm{Diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_n})$. Then

$$\sum_{i=1}^{n} \sqrt{\lambda_i}\boldsymbol{e}_i\boldsymbol{e}_i^\top = \Gamma\Lambda^{\frac{1}{2}}\Gamma^\top$$

is the **square root** of A and is denoted by $A^{1/2}$.

Properties:

(a) $\left(A^{1/2}\right)^\top = A^{1/2}$ ($A^{1/2}$ is symmetric).

(b) $A^{1/2}A^{1/2} = A$.

(c) $A^{-1/2} \equiv \left(A^{1/2}\right)^{-1} = \sum_{i=1}^n \frac{1}{\sqrt{\lambda_i}}\boldsymbol{e}_i\boldsymbol{e}_i^\top = \Gamma\Lambda^{-1/2}\Gamma^\top$, where $\Lambda^{-1/2} = \mathrm{Diag}(\frac{1}{\sqrt{\lambda_1}}, \frac{1}{\sqrt{\lambda_2}}, \cdots, \frac{1}{\sqrt{\lambda_n}})$.

(d) $A^{1/2}A^{-1/2} = A^{-1/2}A^{1/2} = I_n$, and $A^{-1/2}A^{-1/2} = A^{-1}$.

Partitioned matrices

1. If A, B are $p \times n$ matrices and

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \qquad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

are similarly partitioned, then

$$A + B = \begin{pmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{pmatrix}.$$

2. If A is $p \times n$:

$$A = \begin{array}{cc} \begin{array}{cc} q & n-q \end{array} & \\ \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} & \begin{array}{c} r \\ p-r \end{array} \end{array}$$

and C is $n \times m$:

$$C = \begin{array}{cc} \begin{array}{cc} s & m-s \end{array} & \\ \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} & \begin{array}{c} q \\ n-q \end{array} \end{array}$$

then AC is $p \times m$:

$$AC = \begin{array}{cc} s & m-s \\ \left(\begin{array}{c} A_{11}C_{11} + A_{12}C_{21} \\ A_{21}C_{11} + A_{22}C_{21} \end{array}\right. & \left.\begin{array}{c} A_{11}C_{12} + A_{12}C_{22} \\ A_{21}C_{12} + A_{22}C_{22} \end{array}\right) \end{array} \begin{array}{c} \\ r \\ p-r \end{array}$$

(as if A_{ij}, C_{ij} were scalars).

3. Suppose A is $n \times n$ and $|A| \neq 0$. Let A and A^{-1} be similarly partitioned:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad A^{-1} = \begin{pmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{pmatrix}$$

where A_{11}, A^{11}, A_{22} and A^{22} are square matrices. If $|A_{11}| \neq 0$ and $|A_{22}| \neq 0$ and we let

$$B_{11} = A_{11} - A_{12}A_{22}^{-1}A_{21},$$

and

$$B_{22} = A_{22} - A_{21}A_{11}^{-1}A_{12},$$

then

(a) $A^{11} = B_{11}^{-1}$ and $A^{12} = -A_{11}^{-1}A_{12}B_{22}^{-1}$,
$A^{21} = -A_{22}^{-1}A_{21}B_{11}^{-1}$ and $A^{22} = B_{22}^{-1}$.

Thus we can express A^{-1} in terms of the sub-matrices of A.

(b) $|A| = |B_{11}||A_{22}|$.

(c) $|A| = |A_{11}||B_{22}|$.

Matrix and Eigenvalue Inequalities

Lemma 0.1 *The Cauchy-Schwarz Inequality for $d \times 1$ vectors $\boldsymbol{b}$, $\boldsymbol{d}$ states that*

$$\left(\boldsymbol{b}^\top \boldsymbol{d}\right)^2 \leq (\boldsymbol{b}^\top \boldsymbol{b})(\boldsymbol{d}^\top \boldsymbol{d})$$

with equality if and only if $\boldsymbol{b} = c\boldsymbol{d}$ for some c.

Proof
If $\boldsymbol{b} - c\boldsymbol{d} \neq \boldsymbol{0}$ for any c, then for all c,

$$\begin{aligned} 0 < (\boldsymbol{b} - c\boldsymbol{d})^\top(\boldsymbol{b} - c\boldsymbol{d}) &= \boldsymbol{b}^\top\boldsymbol{b} - c\boldsymbol{b}^\top\boldsymbol{d} - c\boldsymbol{d}^\top\boldsymbol{b} + c^2\boldsymbol{d}^\top\boldsymbol{d} \\ &= \boldsymbol{b}^\top\boldsymbol{b} - 2c\boldsymbol{b}^\top\boldsymbol{d} + c^2\boldsymbol{d}^\top\boldsymbol{d} \\ &= \boldsymbol{d}^\top\boldsymbol{d}\left\{c^2 - 2c\frac{\boldsymbol{b}^\top\boldsymbol{d}}{\boldsymbol{d}^\top\boldsymbol{d}} + \left(\frac{\boldsymbol{b}^\top\boldsymbol{d}}{\boldsymbol{d}^\top\boldsymbol{d}}\right)^2\right\} - \frac{(\boldsymbol{b}^\top\boldsymbol{d})^2}{\boldsymbol{d}^\top\boldsymbol{d}} + \boldsymbol{b}^\top\boldsymbol{b} \\ &= \boldsymbol{d}^\top\boldsymbol{d}\left(c - \frac{\boldsymbol{b}^\top\boldsymbol{d}}{\boldsymbol{d}^\top\boldsymbol{d}}\right)^2 + \boldsymbol{b}^\top\boldsymbol{b} - \frac{(\boldsymbol{b}^\top\boldsymbol{d})^2}{\boldsymbol{d}^\top\boldsymbol{d}}. \end{aligned}$$

In particular, if we put $c = \frac{\boldsymbol{b}^\top\boldsymbol{d}}{\boldsymbol{d}^\top\boldsymbol{d}}$, we have

$$(\boldsymbol{b}^\top\boldsymbol{d})^2 < (\boldsymbol{b}^\top\boldsymbol{b})(\boldsymbol{d}^\top\boldsymbol{d}) \qquad \text{if } \boldsymbol{b} \neq c\boldsymbol{d} \text{ for any } c.$$

If $\boldsymbol{b} = c\boldsymbol{d}$ for some c, $\boldsymbol{b}^\top\boldsymbol{d} = c\boldsymbol{d}^\top\boldsymbol{d}$ and

$$0 = (\boldsymbol{b} - c\boldsymbol{d})^\top (\boldsymbol{b} - c\boldsymbol{d}) = \boldsymbol{b}^\top \boldsymbol{b} - \frac{(\boldsymbol{b}^\top \boldsymbol{d})^2}{\boldsymbol{d}^\top \boldsymbol{d}}$$

by the above argument, so $(\boldsymbol{b}^\top \boldsymbol{d})^2 = (\boldsymbol{b}^\top \boldsymbol{b})(\boldsymbol{d}^\top \boldsymbol{d})$. ■

Corollary 0.1 *Assume that B is positive definite. Under the assumptions of Lemma 0.1*

$$(\boldsymbol{b}^\top \boldsymbol{d})^2 \leq (\boldsymbol{b}^\top B \boldsymbol{b})(\boldsymbol{d}^\top B^{-1} \boldsymbol{d})$$

with equality if and only if $\boldsymbol{b} = cB^{-1}\boldsymbol{d}$ for some c.

32

Proof

If B has eigenvalue-eigenvector pairs $(\lambda_1, \boldsymbol{e}_1), \cdots, (\lambda_d, \boldsymbol{e}_d)$, then

$$B^{p/q} = \sum_{i=1}^{d} \lambda_i^{p/q} \boldsymbol{e}_i \boldsymbol{e}_i^\top$$

Now

$$\boldsymbol{b}^\top \boldsymbol{d} = \boldsymbol{b}^\top B^{1/2} B^{-1/2} \boldsymbol{d} = \left(B^{1/2}\boldsymbol{b}\right)^\top \left(B^{-1/2}\boldsymbol{d}\right) \text{ and}$$

hence, by the Cauchy-Schwarz inequality,

$$\begin{aligned} \left(\boldsymbol{b}^\top \boldsymbol{d}\right)^2 &= \left[\left(B^{1/2}\boldsymbol{b}\right)^\top \left(B^{-1/2}\boldsymbol{d}\right)\right]^2 \\ &\leq \left(B^{1/2}\boldsymbol{b}\right)^\top \left(B^{1/2}\boldsymbol{b}\right) \left(B^{-1/2}\boldsymbol{d}\right)^\top \left(B^{-1/2}\boldsymbol{d}\right) \\ &= \left(\boldsymbol{b}^\top B \boldsymbol{b}\right) \left(\boldsymbol{d}^\top B^{-1} \boldsymbol{d}\right) \end{aligned}$$

with equality if and only if $B^{1/2}\boldsymbol{b} = cB^{-1/2}\boldsymbol{d}$ or $\boldsymbol{b} = cB^{-1}\boldsymbol{d}$ for some c. ■

33

Lemma 0.2 *If B is positive definite $d \times d$ matrix and $\boldsymbol{d}$ is a d-vector, then*

$$\max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\left(\boldsymbol{x}^\top \boldsymbol{d}\right)^2}{\boldsymbol{x}^\top B \boldsymbol{x}} = \boldsymbol{d}^\top B^{-1} \boldsymbol{d}$$

and the maximum is attained when $\boldsymbol{x} = cB^{-1}\boldsymbol{d}$ for any $c \neq 0$.

Proof

By Corollary 0.1, for any $\boldsymbol{x}$,

$$\left(\boldsymbol{x}^\top \boldsymbol{d}\right)^2 \leq \left(\boldsymbol{x}^\top B \boldsymbol{x}\right) \left(\boldsymbol{d}^\top B^{-1} \boldsymbol{d}\right),$$

so

$$\frac{\left(\boldsymbol{x}^\top \boldsymbol{d}\right)^2}{\boldsymbol{x}^\top B \boldsymbol{x}} \leq \boldsymbol{d}^\top B^{-1} \boldsymbol{d} \quad \text{if } \boldsymbol{x} \neq 0$$

since B is positive definite, which implies that

34

$$\max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\left(\boldsymbol{x}^\top \boldsymbol{d}\right)^2}{\boldsymbol{x}^\top B \boldsymbol{x}} \leq \boldsymbol{d}^\top B^{-1} \boldsymbol{d}.$$

When $\boldsymbol{x} = cB^{-1}\boldsymbol{d}$ for any c,

$$\left(\boldsymbol{x}^\top \boldsymbol{d}\right)^2 = c^2 \left(\boldsymbol{d}^\top B^{-1} \boldsymbol{d}\right)^2$$

and

$$\begin{aligned} \left(\boldsymbol{x}^\top B \boldsymbol{x}\right) \left(\boldsymbol{d}^\top B^{-1} \boldsymbol{d}\right) &= c^2 \left(\boldsymbol{d}^\top B^{-1} B B^{-1} \boldsymbol{d}\right) \left(\boldsymbol{d}^\top B^{-1} \boldsymbol{d}\right) \\ &= c^2 \left(\boldsymbol{d}^\top B^{-1} \boldsymbol{d}\right)^2 = \left(\boldsymbol{x}^\top \boldsymbol{d}\right)^2 \end{aligned}$$

so

$$\frac{\left(\boldsymbol{x}^\top \boldsymbol{d}\right)^2}{\boldsymbol{x}^\top B \boldsymbol{x}} = \boldsymbol{d}^\top B^{-1} \boldsymbol{d}.$$

35

Thus

$$\max_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\left(\boldsymbol{x}^\top \boldsymbol{d}\right)^2}{\boldsymbol{x}^\top B \boldsymbol{x}} = \boldsymbol{d}^\top B^{-1} \boldsymbol{d}.$$

■

Proposition 0.2 *Suppose B is positive definite and with eigenvalue-eigenvector pairs $(\lambda_i, \boldsymbol{e}_i)$ and $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_d > 0$. The following hold:*

1.
$$\max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}^\top B\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}} = \lambda_1, \qquad \textit{attained when } \boldsymbol{x} = \boldsymbol{e}_1;$$

2.
$$\min_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}^\top B\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}} = \lambda_d, \qquad \textit{attained when } \boldsymbol{x} = \boldsymbol{e}_d.$$

3. *For* $k = 1, 2, \cdots, d-1$,
$$\max_{\boldsymbol{x}\neq\boldsymbol{0}} \frac{\boldsymbol{x}^\top B\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}} = \lambda_{k+1}, \qquad \textit{attained when } \boldsymbol{x} = \boldsymbol{e}_{k+1},$$
where the maximum is over all $\boldsymbol{x}$ *orthogonal to* $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_k$.

Proof
Let $\Gamma = (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d)$ and $\Lambda = \mathrm{Diag}(\lambda_1, \cdots, \lambda_d)$.
Then $B = \Gamma\Lambda\Gamma^\top$ and for $\boldsymbol{x} \neq \boldsymbol{0},\ \boldsymbol{y} = \Gamma^\top\boldsymbol{x} \Longrightarrow \boldsymbol{y} \neq \boldsymbol{0}$ (Γ is nonsingular) and

$$\begin{aligned}
\frac{\boldsymbol{x}^\top B\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}} &= \frac{\boldsymbol{x}^\top\Gamma\Lambda\Gamma^\top\boldsymbol{x}}{\boldsymbol{x}^\top\Gamma\Gamma^\top\boldsymbol{x}} \qquad (\Gamma\Gamma^\top = I_d)\\
&= \frac{\left(\Gamma^\top\boldsymbol{x}\right)^\top\Lambda\Gamma^\top\boldsymbol{x}}{\boldsymbol{x}^\top\Gamma\Gamma^\top\boldsymbol{x}}\\
&= \frac{\boldsymbol{y}^\top\Lambda\boldsymbol{y}}{\boldsymbol{y}^\top\boldsymbol{y}} = \frac{\sum_{i=1}^d \lambda_i y_i^2}{\sum_{i=1}^d y_i^2}\\
&\leq \frac{\lambda_1\sum_{i=1}^d y_i^2}{\sum_{i=1}^d y_i^2} = \lambda_1 \qquad \text{for any } \boldsymbol{x} \neq \boldsymbol{0}.
\end{aligned}$$

Putting $\boldsymbol{x} = \boldsymbol{e}_1$, $\boldsymbol{y} = \Gamma^\top\boldsymbol{e}_1 = (1\,0\ \ldots\ 0)^\top$ and
$$\frac{\boldsymbol{e}_1^\top B\boldsymbol{e}_1}{\boldsymbol{e}_1^\top\boldsymbol{e}_1} = \frac{\sum_{i=1}^d \lambda_i y_i^2}{\sum_{i=1}^d y_i^2} = \lambda_1,$$
so the maximum is attained at $\boldsymbol{x} = \boldsymbol{e}_1$.

Similarly for part (2) and $\boldsymbol{x} \neq \boldsymbol{0}$, we have

$$\frac{\boldsymbol{x}^\top B\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}} = \frac{\sum_{i=1}^d \lambda_i y_i^2}{\sum_{i=1}^d y_i^2} \geq \frac{\lambda_d\sum_{i=1}^d y_i^2}{\sum_{i=1}^d y_i^2} = \lambda_d$$

and putting $\boldsymbol{x} = \boldsymbol{e}_d$, $\boldsymbol{y} = \Gamma^\top\boldsymbol{e}_d = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$, so
$$\frac{\boldsymbol{e}_d^\top B\boldsymbol{e}_d}{\boldsymbol{e}_d^\top\boldsymbol{e}_d} = \frac{\sum_{i=1}^d \lambda_i y_i^2}{\sum_{i=1}^d y_i^2} = \lambda_d,$$
so the minimum is attained at $\boldsymbol{x} = \boldsymbol{e}_d$.

Finally for part (3) and any $\boldsymbol{x} \neq \boldsymbol{0}$, if $\boldsymbol{y} = \Gamma^\top\boldsymbol{x}$ then

$$\boldsymbol{x} = \Gamma\,\boldsymbol{y} = y_1\boldsymbol{e}_1 + y_2\boldsymbol{e}_2 + \cdots + y_p\boldsymbol{e}_d.$$
Thus if $\boldsymbol{x}$ is orthogonal to $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_k$,

$$\begin{aligned}
0 &= \boldsymbol{e}_i^\top\boldsymbol{x}\\
&= y_1\boldsymbol{e}_i^\top\boldsymbol{e}_1 + y_2\boldsymbol{e}_i^\top\boldsymbol{e}_2 + \cdots + y_p\boldsymbol{e}_i^\top\boldsymbol{e}_d\\
&= y_i \qquad \text{for } 1 \leq i \leq k
\end{aligned}$$
and so

$$\frac{\boldsymbol{x}^\top B\boldsymbol{x}}{\boldsymbol{x}^\top\boldsymbol{x}} = \frac{\sum_{i=k+1}^d \lambda_i y_i^2}{\sum_{i=k+1}^d y_i^2} \leq \frac{\lambda_{k+1}\sum_{i=k+1}^d y_i^2}{\sum_{i=k+1}^d y_i^2} = \lambda_{k+1}.$$

For $\boldsymbol{x} = \boldsymbol{e}_{k+1}$ (which is orthogonal to $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_k$)

$$\boldsymbol{y} = \Gamma^\top \boldsymbol{e}_{k+1} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}; \text{ that is, } y_{k+1} = 1$$

and

$$\frac{\boldsymbol{x}^\top B \boldsymbol{x}}{\boldsymbol{x}^\top \boldsymbol{x}} = \lambda_{k+1},$$

so the maximum is attained at $\boldsymbol{x} = \boldsymbol{e}_{k+1}$. ■