Stat 501: Review of Matrix and Linear Algebra

Ranjan Maitra

2107 Snedecor Hall
Department of Statistics
Iowa State University.
Phone: 515-294-7757
maitra@iastate.edu

January 23, 2025

Basic Concepts in Matrix Algebra

- An column array of p elements is called a vector of dimension p and is written as

$$\boldsymbol{x}_{p\times 1} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}.$$

- The transpose of the column vector $\boldsymbol{x}_{p\times 1}$ is row vector

$$\boldsymbol{x}' = [x_1 \; x_2 \; \ldots \; x_p]$$

- A vector can be represented in p-space as a directed line with components along the p axes.
- Two vectors can be added if they have the same dimension. Addition is carried out elementwise.

$$\boldsymbol{x} + \boldsymbol{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_p \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_p + y_p \end{bmatrix}$$

- A vector can be contracted or expanded if multiplied by a constant c. Multiplication is also elementwise.

$$c\boldsymbol{x} = c\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix} = \begin{bmatrix} cx_1 \\ cx_2 \\ \vdots \\ cx_p \end{bmatrix}$$

Examples

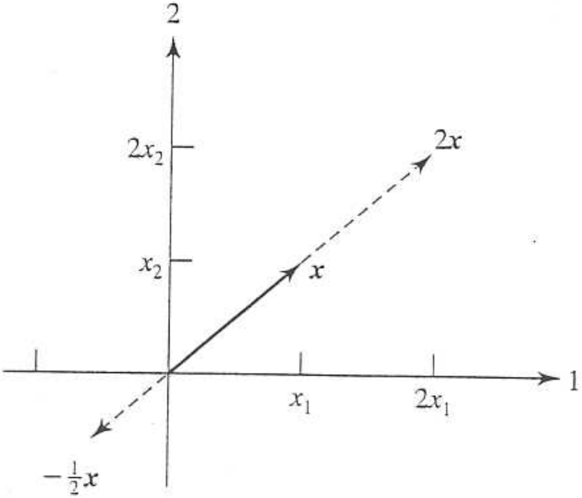
$$\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix} \text{ and } \boldsymbol{x}' = \begin{bmatrix} 2 & 1 & -4 \end{bmatrix}$$

$$6 \times \boldsymbol{x} = 6 \times \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix} = \begin{bmatrix} 6 \times 2 \\ 6 \times 1 \\ 6 \times (-4) \end{bmatrix} = \begin{bmatrix} 12 \\ 6 \\ -24 \end{bmatrix}$$

$$\boldsymbol{x} + \boldsymbol{y} = \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix} + \begin{bmatrix} 5 \\ -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2+5 \\ 1-2 \\ -4+0 \end{bmatrix} = \begin{bmatrix} 7 \\ -1 \\ -4 \end{bmatrix}$$

frtMultiplication by a Scalar

- Multiplication by $c > 0$ does not change the direction of $\boldsymbol{x}$.
  Direction is reversed if $c < 0$.



Length of a vector

- The *length* of a vector $\boldsymbol{x}$ is the Euclidean distance from the origin

$$L_{\boldsymbol{x}} = \sqrt{\sum_{j=1}^{p} x_j^2} = \|\boldsymbol{x}\|.$$

- Multiplication of a vector $\boldsymbol{x}$ by a constant c changes the length:

$$L_{c\boldsymbol{x}} = \sqrt{\sum_{j=1}^{p} c^2 x_j^2} = |\, c \,| \sqrt{\sum_{j=1}^{p} x_j^2} = |\, c \,|\; L_{\boldsymbol{x}}.$$

- If $c = L_{\boldsymbol{x}}^{-1}$, then $c\boldsymbol{x}$ is a vector of unit length.

Examples

The length of $\boldsymbol{x} = \begin{bmatrix} 2 \\ 1 \\ -4 \\ -2 \end{bmatrix}$ is

$$L_{\boldsymbol{x}} = \sqrt{(2)^2 + (1)^2 + (-4)^2 + (-2)^2} = \sqrt{25} = 5$$

Then

$$\boldsymbol{z} = \frac{1}{5} \times \begin{bmatrix} 2 \\ 1 \\ -4 \\ -2 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.2 \\ -0.8 \\ -0.4 \end{bmatrix}$$

is a vector of unit length.

Angle Between Vectors

- Consider two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in two dimensions. If θ_1 is the angle between $\boldsymbol{x}$ and the horizontal axis and $\theta_2 > \theta_1$ is the angle between $\boldsymbol{y}$ and the horizontal axis, then

$$\cos(\theta_1) = \frac{x_1}{L_{\boldsymbol{x}}} \qquad \cos(\theta_2) = \frac{y_1}{L_{\boldsymbol{y}}}$$
$$\sin(\theta_1) = \frac{x_2}{L_{\boldsymbol{x}}} \qquad \sin(\theta_2) = \frac{y_2}{L_{\boldsymbol{y}}},$$

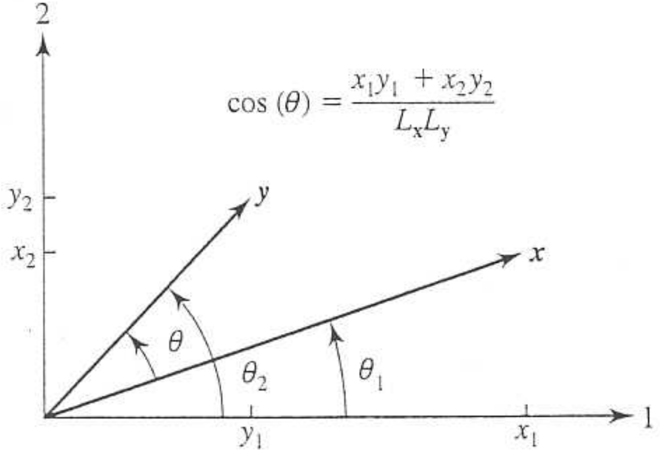
If θ is the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$, then

$$\cos(\theta) = \cos(\theta_2 - \theta_1) = \cos(\theta_2)\cos(\theta_1) + \sin(\theta_2)\sin(\theta_1).$$

Then

$$\cos(\theta) = \frac{x_1 y_1 + x_2 y_2}{L_{\boldsymbol{x}} L_{\boldsymbol{y}}} = \frac{\boldsymbol{x}'\boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}.$$

Angle Between Vectors (continued)



Inner Product Between Two Vectors

- The inner product between two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$\boldsymbol{x}'\boldsymbol{y} = \sum_{j=1}^{p} x_j y_j.$$

- Then $L_x = \sqrt{\boldsymbol{x}'\boldsymbol{x}}$, $L_y = \sqrt{\boldsymbol{y}'\boldsymbol{y}}$ and

$$\cos(\theta) = \frac{\boldsymbol{x}'\boldsymbol{y}}{\sqrt{(\boldsymbol{x}'\boldsymbol{x})}\sqrt{(\boldsymbol{y}'\boldsymbol{y})}}$$

- Since $\cos(\theta) = 0$ when $\boldsymbol{x}'\boldsymbol{y} = 0$ and $\cos(\theta) = 0$ for $\theta = 90$ or $\theta = 270$, then the vectors are perpendicular (orthogonal) when $\boldsymbol{x}'\boldsymbol{y} = 0$.

Linear Dependence

- Two vectors, $\boldsymbol{x}$ and $\boldsymbol{y}$, are *linearly dependent* if there exist two constants c_1 and c_2, not both zero, such that

$$c_1\boldsymbol{x} + c_2\boldsymbol{y} = 0$$

- If two vectors are linearly dependent, then one can be written as a linear combination of the other. From above:

$$\boldsymbol{x} = (c_2/c_1)\boldsymbol{y}$$

- k vectors, $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$, are linearly dependent if there exist constants $(c_1, c_2, ..., c_k)$ not all zero such that

$$\sum_{j=1}^{k} c_j\boldsymbol{x}_j = 0.$$

- Vectors of the same dimension that are not linearly dependent are said to be *linearly independent*.

Linear Independence-Example

Let

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad \boldsymbol{x}_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}$$

Then $c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + c_3\boldsymbol{x}_3 = 0$ if

$$\begin{array}{ccccccc} c_1 & + & c_2 & + & c_3 & = & 0 \\ 2c_1 & + & 0 & - & 2c_3 & = & 0 \\ c_1 & - & c_2 & + & c_3 & = & 0 \end{array}$$

The unique solution is $c_1 = c_2 = c_3 = 0$, so the vectors are linearly independent.

Projections

- The projection of $\boldsymbol{x}$ on $\boldsymbol{y}$ is defined by

$$\text{Projection of } \boldsymbol{x} \text{ on } \boldsymbol{y} = \frac{\boldsymbol{x}'\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}}\boldsymbol{y} = \frac{\boldsymbol{x}'\boldsymbol{y}}{L_{\boldsymbol{y}}}\frac{1}{L_{\boldsymbol{y}}}\boldsymbol{y}.$$

- The length of the projection is

$$\text{Length of projection} = \frac{|\boldsymbol{x}'\boldsymbol{y}|}{L_{\boldsymbol{y}}} = L_{\boldsymbol{x}}\frac{|\boldsymbol{x}'\boldsymbol{y}|}{L_{\boldsymbol{x}}L_{\boldsymbol{y}}} = L_{\boldsymbol{x}}\,|\cos(\theta)|,$$

where θ is the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$.

Matrix Algebra

A matrix $\boldsymbol{A}$ is an array of elements a_{ij} with n rows and p columns:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix}$$

The transpose $\boldsymbol{A}'$ has p rows and n columns. The j-th row of $\boldsymbol{A}'$ is the j-th column of $\boldsymbol{A}$

$$\boldsymbol{A}' = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{n1} \\ a_{12} & a_{22} & \cdots & a_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1p} & a_{2p} & \cdots & a_{np} \end{bmatrix}$$

- Multiplication of $\boldsymbol{A}$ by a constant c is carried out element by element.

$$c\boldsymbol{A} = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1p} \\ ca_{21} & ca_{22} & \cdots & ca_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{np} \end{bmatrix}$$

Matrix Addition

Two matrices $A_{n\times p} = \{a_{ij}\}$ and $B_{n\times p} = \{b_{ij}\}$ of the same dimensions can be added element by element. The resulting matrix is $C_{n\times p} = \{c_{ij}\} = \{a_{ij} + b_{ij}\}$

$$\begin{aligned}
\boldsymbol{C} &= \boldsymbol{A} + \boldsymbol{B} \\
&= \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & b_{22} & \cdots & b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{np} \end{bmatrix} \\
&= \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1p}+b_{1p} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2p}+b_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}+b_{n1} & a_{n2}+b_{n2} & \cdots & a_{np}+b_{np} \end{bmatrix}
\end{aligned}$$

Example:

$$\begin{bmatrix} 2 & 1 & -4 \\ 5 & 7 & 0 \end{bmatrix}' = \begin{bmatrix} 2 & 5 \\ 1 & 7 \\ -4 & 0 \end{bmatrix}$$

$$6 \times \begin{bmatrix} 2 & 1 & -4 \\ 5 & 7 & 0 \end{bmatrix} = \begin{bmatrix} 12 & 6 & -24 \\ 30 & 42 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 2 & -1 \\ 0 & 3 \end{bmatrix} + \begin{bmatrix} 2 & 1 \\ 5 & 7 \end{bmatrix} = \begin{bmatrix} 4 & 0 \\ 5 & 10 \end{bmatrix}$$

Matrix Multiplication

- Multiplication of two matrices $\boldsymbol{A}_{n\times p}$ and $\boldsymbol{B}_{m\times q}$ can be carried out only if the matrices are *compatible for multiplication*:
  - $\boldsymbol{A}_{n\times p} \times \boldsymbol{B}_{m\times q}$: compatible if $p = m$.
  - $\boldsymbol{B}_{m\times q} \times \boldsymbol{A}_{n\times p}$: compatible if $q = n$.

  The element in the i-th row and the j-th column of $\boldsymbol{A} \times \boldsymbol{B}$ is the inner product of the i-th row of $\boldsymbol{A}$ with the j-th column of $\boldsymbol{B}$.

$$\begin{bmatrix} 2 & 0 & 1 \\ 5 & 1 & 3 \end{bmatrix} \times \begin{bmatrix} 1 & 4 \\ -1 & 3 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 10 \\ 4 & 29 \end{bmatrix}$$

Example:

$$\begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix} \times \begin{bmatrix} 1 & 4 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 11 \\ 2 & 29 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 4 \\ -1 & 3 \end{bmatrix} \times \begin{bmatrix} 2 & 1 \\ 5 & 3 \end{bmatrix} = \begin{bmatrix} 22 & 13 \\ 13 & 8 \end{bmatrix}$$

The Identity Matrix and the Inverse

- An *identity matrix*, denoted by $\boldsymbol{I}$, is a square matrix with 1's along the main diagonal and 0's everywhere else. For example,

$$\boldsymbol{I}_{2\times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \boldsymbol{I}_{3\times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- If $\boldsymbol{A}$ is a square matrix, then $\boldsymbol{AI} = \boldsymbol{IA} = \boldsymbol{A}$.
- $\boldsymbol{I}_{n\times n}\boldsymbol{A}_{n\times p} = \boldsymbol{A}_{n\times p}$ but $\boldsymbol{A}_{n\times p}\boldsymbol{I}_{n\times n}$ is not defined for $p \neq n$.
- Consider two square matrices $\boldsymbol{A}_{k\times k}$ and $\boldsymbol{B}_{k\times k}$. If

$$\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$$

  then $\boldsymbol{B}$ is the *inverse* of $\boldsymbol{A}$, denoted $\boldsymbol{A}^{-1}$.
- The inverse of $\boldsymbol{A}$ exists only if the columns of $\boldsymbol{A}$ are linearly independent.
- If $\boldsymbol{A} = \text{diag}\{a_{ij}\}$ then $\boldsymbol{A}^{-1} = \text{diag}\{1/a_{ij}\}$.

Symmetric and Orthogonal Matrices

- A square matrix is *symmetric* if $\boldsymbol{A} = \boldsymbol{A}'$.
- If a square matrix $\boldsymbol{A}$ has elements $\{a_{ij}\}$, then $\boldsymbol{A}$ is symmetric if $a_{ij} = a_{ji}$.
- A square matrix $\boldsymbol{Q}$ is *orthogonal* if

$$\boldsymbol{Q}\boldsymbol{Q}' = \boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I},$$

  or $\boldsymbol{Q}' = \boldsymbol{Q}^{-1}$.
- If $\boldsymbol{Q}$ is orthogonal, its rows and columns have unit length ($\boldsymbol{q}_j'\boldsymbol{q}_j = 1$) and are mutually perpendicular ($\boldsymbol{q}_j'\boldsymbol{q}_k = 0$ for any $j \neq k$).

Eigenvalues and Eigenvectors

- A square matrix $\boldsymbol{A}$ has an eigenvalue λ with corresponding eigenvector $\boldsymbol{z} \neq \boldsymbol{0}$ if

$$\boldsymbol{Az} = \lambda \boldsymbol{z}$$

- The eigenvalues of $\boldsymbol{A}$ are the solution to $|A - \lambda I| = 0$.
- A normalized eigenvector (of unit length) is denoted by $\boldsymbol{e}$.
- A $k \times k$ matrix $\boldsymbol{A}$ has k pairs of eigenvalues and eigenvectors

$$(\lambda_1, \boldsymbol{e}_1), \quad (\lambda_2, \boldsymbol{e}_2) \quad \ldots \quad (\lambda_k, \boldsymbol{e}_k)$$

where $\boldsymbol{e}_i'\boldsymbol{e}_i = 1$, $\boldsymbol{e}_i'\boldsymbol{e}_j = 0$ and the eigenvectors are unique up to a change in sign unless two or more eigenvalues are equal.

Spectral Decomposition

- Eigenvalues and eigenvectors will play an important role in this course. For example, principal components are based on the eigenvalues and eigenvectors of sample covariance matrices.
- The *spectral decomposition* of a $k \times k$ symmetric matrix $\boldsymbol{A}$ is

$$
\begin{aligned}
\boldsymbol{A} &= \lambda_1 \boldsymbol{e}_1 \boldsymbol{e}_1' + \lambda_2 \boldsymbol{e}_2 \boldsymbol{e}_2' + \ldots + \lambda_k \boldsymbol{e}_k \boldsymbol{e}_k' \\
&= [\boldsymbol{e}_1 \ \boldsymbol{e}_2 \ \cdots \ \boldsymbol{e}_k] \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{bmatrix} [\boldsymbol{e}_1 \ \boldsymbol{e}_2 \ \cdots \ \boldsymbol{e}_k]' \\
&= \boldsymbol{P\Lambda P}'
\end{aligned}
$$

Determinant, Trace and Rank of a Matrix

- The *trace* of a $k \times k$ matrix $\boldsymbol{A}$ is the sum of the diagonal elements, i.e., $trace(\boldsymbol{A}) = \sum_{i=1}^{k} a_{ii}$
- The trace of a square, symmetric matrix $\boldsymbol{A}$ is the sum of the eigenvalues, i.e., $trace(\boldsymbol{A}) = \sum_{i=1}^{k} a_{ii} = \sum_{i=1}^{k} \lambda_i$
- The determinant of a square, symmetric matrix $\boldsymbol{A}$ is the product of the eigenvalues, i.e., $|\boldsymbol{A}| = \prod_{i=1}^{k} \lambda_i$
- The rank of a square matrix $\boldsymbol{A}$ is
  - The number of linearly independent rows
  - The number of linearly independent columns
  - The number of non-zero eigenvalues
- The inverse of a $k \times k$ matrix $\boldsymbol{A}$ exists, if and only if

$$rank(\boldsymbol{A}) = k$$

i.e., there are no zero eigenvalues

Positive-Definite Matrices

- For a $k \times k$ symmetric matrix $\boldsymbol{A}$ and a vector $\boldsymbol{x} = [x_1, x_2, ..., x_k]'$ the quantity $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ is called a *quadratic form*
- Note that $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij} x_i x_j$
- If $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} \geq 0 \quad$ for any vector $\boldsymbol{x}$, both $\boldsymbol{A}$ and the quadratic form are said to be *non-negative definite*.
- If $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} > 0 \quad$ for any vector $\boldsymbol{x} \neq \mathbf{0}$, both $\boldsymbol{A}$ and the quadratic form are said to be *positive definite*.

Example 2.11

- Show that the matrix of the quadratic form $3x_1^2 + 2x_2^2 - 2\sqrt{2}x_1x_2$ is positive definite.
- For

$$\boldsymbol{A} = \begin{bmatrix} 3 & -\sqrt{2} \\ -\sqrt{2} & 2 \end{bmatrix},$$

  the eigenvalues are $\lambda_1 = 4, \lambda_2 = 1$. Then $\boldsymbol{A} = 4\boldsymbol{e}_1\boldsymbol{e}_1' + \boldsymbol{e}_2\boldsymbol{e}_2'$. Write

$$\begin{aligned} \boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} &= 4\boldsymbol{x}'\boldsymbol{e}_1\boldsymbol{e}_1'\boldsymbol{x} + \boldsymbol{x}'\boldsymbol{e}_2\boldsymbol{e}_2'\boldsymbol{x} \\ &= 4y_1^2 + y_2^2 \geq 0, \end{aligned}$$

  and is zero only for $y_1 = y_2 = 0$.
- y_1, y_2 cannot be zero because

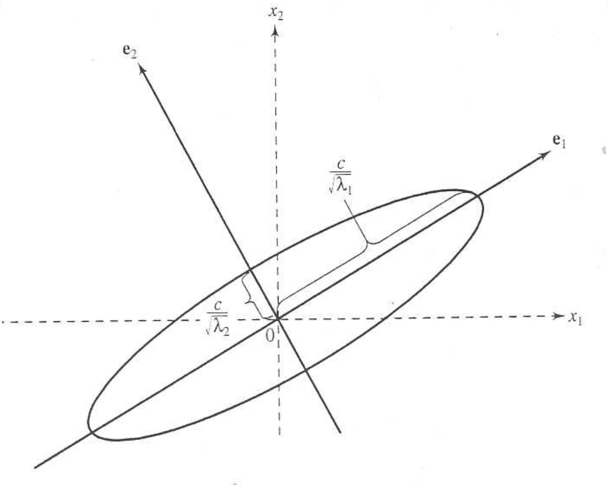
$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}_1' \\ \boldsymbol{e}_2' \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = P'_{2\times 2}\boldsymbol{x}_{2\times 1}$$

  with P' orthonormal so that $(P')^{-1} = P$. Then $\boldsymbol{x} = P\boldsymbol{y}$ and since $\boldsymbol{x} \neq 0$ it follows that $\boldsymbol{y} \neq 0$.
- Using the spectral decomposition, we can show that:
  - A is positive definite if all of its eigenvalues are positive.
  - A is non-negative definite if all of its eigenvalues are ≥ 0.

Distance and Quadratic Forms

- For $\boldsymbol{x} = [x_1, x_2, ..., x_p]'$ and a $p \times p$ positive definite matrix A,
$$d^2 = \boldsymbol{x}' A \boldsymbol{x} > 0$$
when $x \neq 0$. Thus, a positive definite quadratic form can be interpreted as a squared distance of $\boldsymbol{x}$ from the origin and vice versa.
- The squared distance from $\boldsymbol{x}$ to a fixed point $\boldsymbol{\mu}$ is given by the quadratic form
$$(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{A} (\boldsymbol{x} - \boldsymbol{\mu}).$$
- We can interpret distance in terms of eigenvalues and eigenvectors of A as well. Any point $\boldsymbol{x}$ at constant distance c from the origin satisfies
$$\boldsymbol{x}' A \boldsymbol{x} = \boldsymbol{x}' (\sum_{j=1}^{p} \lambda_j \boldsymbol{e}_j \boldsymbol{e}_j') \boldsymbol{x} = \sum_{j=1}^{p} \lambda_j (\boldsymbol{x}' \boldsymbol{e}_j)^2 = c^2,$$
the expression for an ellipsoid in p dimensions.
- Note that the point $\boldsymbol{x} = c\lambda_1^{-1/2} \boldsymbol{e}_1$ is at a distance c (in the direction of $\boldsymbol{e}_1$) from the origin because it satisfies $\boldsymbol{x}' A \boldsymbol{x} = c^2$. The same is true for points $\boldsymbol{x} = c\lambda_j^{-1/2} \boldsymbol{e}_j$, $j = 1, ..., p$. Thus, all points at distance c lie on an ellipsoid with axes in the directions of the eigenvectors and with lengths proportional to $\lambda_j^{-1/2}$.

Distance and Quadratic Forms (cont'd)



Square-Root Matrices

- Spectral decomposition of a positive definite matrix A yields

$$\boldsymbol{A} = \sum_{j=1}^{p} \lambda_j \boldsymbol{e}_j \boldsymbol{e}_j' = \boldsymbol{P} \Lambda \boldsymbol{P},$$

with $\boldsymbol{\Lambda}_{k\times k} = \text{diag}\{\lambda_j\}$, all $\lambda_j > 0$, and $\boldsymbol{P}_{k\times k} = [\boldsymbol{e}_1 \;\; \boldsymbol{e}_2 \;\; \ldots \;\; \boldsymbol{e}_p]$ an orthonormal matrix of eigenvectors. Then

$$\boldsymbol{A}^{-1} = \boldsymbol{P} \Lambda^{-1} \boldsymbol{P}' = \sum_{j=1}^{p} \frac{1}{\lambda_j} \boldsymbol{e}_j \boldsymbol{e}_j'$$

- With $\boldsymbol{\Lambda}^{1/2} = \text{diag}\{\lambda_j^{1/2}\}$, a square-root matrix is

$$\boldsymbol{A}^{1/2} = \boldsymbol{P} \Lambda^{1/2} \boldsymbol{P}' = \sum_{j=1}^{p} \sqrt{\lambda_j} \boldsymbol{e}_j \boldsymbol{e}_j'$$

- The square root of a positive definite matrix $\boldsymbol{A}$ has the following properties:
  1. Symmetry: $(\boldsymbol{A}^{1/2})' = \boldsymbol{A}^{1/2}$
  2. $\boldsymbol{A}^{1/2}\boldsymbol{A}^{1/2} = \boldsymbol{A}$
  3. $\boldsymbol{A}^{-1/2} = \sum_{j=1}^{p} \lambda_j^{-1/2} \boldsymbol{e}_j \boldsymbol{e}_j' = \boldsymbol{P}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{P}'$
  4. $\boldsymbol{A}^{1/2}\boldsymbol{A}^{-1/2} = \boldsymbol{A}^{-1/2}\boldsymbol{A}^{1/2} = \boldsymbol{I}$
  5. $\boldsymbol{A}^{-1/2}\boldsymbol{A}^{-1/2} = \boldsymbol{A}^{-1}$

  Note that there are other ways of defining the square root of a positive definite matrix: in the Cholesky decomposition $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}'$, with $\boldsymbol{L}$ a matrix of lower triangular form, $\boldsymbol{L}$ is also called a square root of $\boldsymbol{A}$.

Random Vectors and Matrices

- A random matrix (vector) is a matrix (vector) whose elements are random variables.
- If $X_{n \times p}$ is a random matrix, the *expected value of X* is the $n \times p$ matrix

$$E(\boldsymbol{X}) = \begin{bmatrix} E(X_{11}) & E(X_{12}) & \cdots & E(X_{1p}) \\ E(X_{21}) & E(X_{22}) & \cdots & E(X_{2p}) \\ \vdots & \vdots & \cdots & \vdots \\ E(X_{n1}) & E(X_{n2}) & \cdots & E(X_{np}) \end{bmatrix},$$

where

$$E(X_{ij}) = \int_{\infty}^{\infty} x_{ij} f_{ij}(x_{ij}) dx_{ij}$$

with $f_{ij}(x_{ij})$ the density function of the continuous random variable X_{ij}. If X is a discrete random variable, we compute its expectation as a sum rather than an integral.

Expectations of Linear Combinations

- The usual rules for expectations apply. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two random matrices and $\boldsymbol{A}$ and $\boldsymbol{B}$ are two constant matrices of the appropriate dimensions, then

$$
\begin{aligned}
E(\boldsymbol{X}+\boldsymbol{Y}) &= E(\boldsymbol{X})+E(\boldsymbol{Y}) \\
E(\boldsymbol{A}\boldsymbol{X}) &= \boldsymbol{A}E(\boldsymbol{X}) \\
E(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) &= \boldsymbol{A}E(\boldsymbol{X})\boldsymbol{B} \\
E(\boldsymbol{A}\boldsymbol{X}+\boldsymbol{B}\boldsymbol{Y}) &= \boldsymbol{A}E(\boldsymbol{X})+\boldsymbol{B}E(\boldsymbol{Y})
\end{aligned}
$$

- Further, if c is a scalar-valued constant then

$$E(c\boldsymbol{X}) = cE(\boldsymbol{X}).$$

Mean Vectors and Covariance Matrices

- Suppose that $\boldsymbol{X}$ is a $p \times 1$ (continuous) random vector drawn from some $p-$dimensional distribution.
- Each element of $\boldsymbol{X}$, say X_j has its own marginal distribution with marginal mean μ_j and variance σ_{jj} defined in the usual way:

$$\mu_j = \int_{-\infty}^{\infty} x_j f_j(x_j) dx_j$$

$$\sigma_{jj} = \int_{-\infty} (x_j - \mu_j)^2 f_j(x_j) dx_j$$

- To examine association between a pair of random variables we need to consider their joint distribution.
- A measure of the linear association between pairs of variables is given by the covariance

$$\begin{aligned} \sigma_{jk} &= E\left[(X_j - \mu_j)(X_k - \mu_k)\right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_j - \mu_j)(x_k - \mu_k) f_{jk}(x_j, x_k) dx_j dx_k. \end{aligned}$$

- If the joint density function $f_{jk}(x_j, x_k)$ can be written as the product of the two marginal densities, e.g.,

$$f_{jk}(x_j, x_k) = f_j(x_j) f_k(x_k),$$

  then X_j and X_k are independent.
- More generally, the $p-$dimensional random vector $\boldsymbol{X}$ has mutually independent elements if the $p-$dimensional joint density function can be written as the product of the p univariate marginal densities.
- If two random variables X_j and X_k are independent, then their covariance is equal to 0. [Converse is not always true.]

Mean Vectors and Covariance Matrices (continued)

- We use $\boldsymbol{\mu}$ to denote the $p \times 1$ vector of marginal population means and use $\boldsymbol{\Sigma}$ to denote the $p \times p$ population variance-covariance matrix:

$$\boldsymbol{\Sigma} = E\left[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'\right].$$

- If we carry out the multiplication (outer product)then $\boldsymbol{\Sigma}$ is equal to:

$$E\begin{bmatrix} (X_1-\mu_1)^2 & (X_1-\mu_1)(X_2-\mu_2) & \cdots & (X_1-\mu_1)(X_p-\mu_p) \\ (X_2-\mu_2)(X_1-\mu_1) & (X_2-\mu_2)^2 & \cdots & (X_2-\mu_2)(X_p-\mu_p) \\ \vdots & \vdots & \cdots & \vdots \\ (X_p-\mu_p)(X_1-\mu_1) & (X_p-\mu_p)(X_2-\mu_2) & \cdots & (X_p-\mu_p)^2 \end{bmatrix}.$$

- By taking expectations element-wise we find that

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{bmatrix}.$$

- Since $\sigma_{jk} = \sigma_{kj}$ for all $j \neq k$ we note that Σ is symmetric.
- $\boldsymbol{\Sigma}$ is also non-negative definite: this can be shown from the fact that for any vector $\boldsymbol{a}$, the variance of $\boldsymbol{a}'\boldsymbol{X}$ is nonnegative.

Correlation Matrix

- The population correlation matrix is the $p \times p$ matrix with off-diagonal elements equal to ρ_{jk} and diagonal elements equal to 1.

$$\begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1p} \\ \rho_{21} & 1 & \cdots & \rho_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{p1} & \rho_{p2} & \cdots & 1 \end{bmatrix}.$$

- Since $\rho_{ij} = \rho_{ji}$ the correlation matrix is symmetric
- The correlation matrix is also non-negative definite
- The $p \times p$ population *standard deviation* matrix $\boldsymbol{V}^{1/2}$ is a diagonal matrix with $\sqrt{\sigma_{jj}}$ along the diagonal and zeros in all off-diagonal positions. Then

$$\Sigma = \boldsymbol{V}^{1/2}\boldsymbol{P}\boldsymbol{V}^{1/2}$$

and the population correlation matrix is

$$(\boldsymbol{V}^{1/2})^{-1}\boldsymbol{\Sigma}(\boldsymbol{V}^{1/2})^{-1}$$

- Given $\boldsymbol{\Sigma}$, we can easily obtain the correlation matrix

Partitioning Random vectors

- If we partition the random $p \times 1$ vector $\boldsymbol{X}$ into two components $\boldsymbol{X}_1, \boldsymbol{X}_2$ of dimensions $q \times 1$ and $(p - q) \times 1$ respectively, then the mean vector and the variance-covariance matrix need to be partitioned accordingly.
- Partitioned mean vector:

$$E(\boldsymbol{X}) = E \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix} = \begin{bmatrix} E(\boldsymbol{X}_1) \\ E(\boldsymbol{X}_2) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$$

Partitioning Covariance Matrices

- Partitioned variance-covariance matrix:

$$\boldsymbol{\Sigma} = \left[\begin{array}{cc} Var(\boldsymbol{X}_1) & Cov(\boldsymbol{X}_1,\ \boldsymbol{X}_2) \\ Cov(\boldsymbol{X}_2, \boldsymbol{X}_1) & Var(\boldsymbol{X}_2) \end{array} \right] = \left[\begin{array}{cc} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22} \end{array} \right],$$

  where $\boldsymbol{\Sigma}_{11}$ is $q \times q$, $\boldsymbol{\Sigma}_{12}$ is $q \times (p-q)$ and $\boldsymbol{\Sigma}_{22}$ is $(p-q) \times (p-q)$.
- $\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22}$ are the variance-covariance matrices of the sub-vectors X_1, X_2, respectively. The off-diagonal elements in those two matrices reflect linear associations among elements *within* each sub-vector.
- There are no variances in $\boldsymbol{\Sigma}_{12}$, only covariances. These covariancs reflect linear associations between elements in the two different sub-vectors.

Linear Combinations of Random variables

- Let $\boldsymbol{X}$ be a $p \times 1$ vector with mean $\boldsymbol{\mu}$ and variance covariance matrix $\boldsymbol{\Sigma}$, and let $\boldsymbol{c}$ be a $p \times 1$ vector of constants. Then the linear combination $\boldsymbol{c}'\boldsymbol{X}$ has mean and variance:

$$E(\boldsymbol{c}'\boldsymbol{X}) = \boldsymbol{c}'\boldsymbol{\mu}, \qquad \text{and} \qquad Var(\boldsymbol{c}'\boldsymbol{X}) = \boldsymbol{c}'\boldsymbol{\Sigma}\boldsymbol{c}$$

- In general, the mean and variance of a $q \times 1$ vector of linear combinations $\boldsymbol{Z} = \boldsymbol{C}_{q\times p}\boldsymbol{X}_{p\times 1}$ are

$$\boldsymbol{\mu}_{\boldsymbol{Z}} = \boldsymbol{C}\boldsymbol{\mu}_{\boldsymbol{X}} \text{ and } \boldsymbol{\Sigma}_{\boldsymbol{Z}} = \boldsymbol{C}\boldsymbol{\Sigma}_{\boldsymbol{X}}\boldsymbol{C}'.$$

Cauchy-Schwarz Inequality

- We will need some of the results below to derive some maximization results later in the course.

Cauchy-Schwarz inequality Let $\boldsymbol{b}$ and $\boldsymbol{d}$ be any two $p \times 1$ vectors. Then,

$$(\boldsymbol{b}'\boldsymbol{d})^2 \leq (\boldsymbol{b}'\boldsymbol{b})(\boldsymbol{d}'\boldsymbol{d})$$

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

Proof: The equality is obvious for $\boldsymbol{b} = \mathbf{0}$ or $\boldsymbol{d} = \mathbf{0}$. For other cases, consider $\boldsymbol{b} - c\boldsymbol{d}$ for any constant $c \neq 0$. Then if $\boldsymbol{b} - c\boldsymbol{d} \neq \mathbf{0}$, we have

$$0 < (\boldsymbol{b} - c\boldsymbol{d})'(\boldsymbol{b} - c\boldsymbol{d}) = \boldsymbol{b}'\boldsymbol{b} - 2c(\boldsymbol{b}'\boldsymbol{d}) + c^2\boldsymbol{d}'\boldsymbol{d},$$

since $\boldsymbol{b} - c\boldsymbol{d}$ must have positive length.
We can add and subtract $(\boldsymbol{b}'\boldsymbol{d})^2/(\boldsymbol{d}'\boldsymbol{d})$ to obtain

$$0 < \boldsymbol{b}'\boldsymbol{b} - 2c(\boldsymbol{b}'\boldsymbol{d}) + c^2\boldsymbol{d}'\boldsymbol{d} - \frac{(\boldsymbol{b}'\boldsymbol{d})^2}{\boldsymbol{d}'\boldsymbol{d}} + \frac{(\boldsymbol{b}'\boldsymbol{d})^2}{\boldsymbol{d}'\boldsymbol{d}} = \boldsymbol{b}'\boldsymbol{b} - \frac{(\boldsymbol{b}'\boldsymbol{d})^2}{\boldsymbol{d}'\boldsymbol{d}} + (\boldsymbol{d}'\boldsymbol{d})\left(c - \frac{\boldsymbol{b}'\boldsymbol{d}}{\boldsymbol{d}'\boldsymbol{d}}\right)^2$$

Since c can be anything, we can choose $c = \boldsymbol{b}'\boldsymbol{d}/\boldsymbol{d}'\boldsymbol{d}$. Then,

$$0 < \boldsymbol{b}'\boldsymbol{b} - \frac{(\boldsymbol{b}'\boldsymbol{d})^2}{\boldsymbol{d}'\boldsymbol{d}} \qquad \Rightarrow \qquad (\boldsymbol{b}'\boldsymbol{d})^2 < (\boldsymbol{b}'\boldsymbol{b})(\boldsymbol{d}'\boldsymbol{d})$$

for $\boldsymbol{b} \neq c\boldsymbol{d}$ (otherwise, we have equality).

Extended Cauchy-Schwarz Inequality

If $\boldsymbol{b}$ and $\boldsymbol{d}$ are any two $p \times 1$ vectors and $\boldsymbol{B}$ is a $p \times p$ positive definite matrix, then

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

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

Proof: Consider $\boldsymbol{B}^{1/2} = \sum_{i=1}^{p} \sqrt{\lambda_i}\boldsymbol{e}_i\boldsymbol{e}_i'$, and $\boldsymbol{B}^{-1/2} = \sum_{i=1}^{p} \frac{1}{(\sqrt{\lambda_i})}\boldsymbol{e}_i\boldsymbol{e}_i'$. Then we can write

$$\boldsymbol{b}'\boldsymbol{d} = \boldsymbol{b}'\boldsymbol{I}\boldsymbol{d} = \boldsymbol{b}'\boldsymbol{B}^{1/2}\boldsymbol{B}^{-1/2}\boldsymbol{d} = (\boldsymbol{B}^{1/2}\boldsymbol{b})'(\boldsymbol{B}^{-1/2}\boldsymbol{d}) = \boldsymbol{b}^{*\prime}\boldsymbol{d}^{*}.$$

To complete the proof, simply apply the Cauchy-Schwarz inequality to the vectors $\boldsymbol{b}^*$ and $\boldsymbol{d}^*$.

Optimization

Let $\boldsymbol{B}$ be positive definite and let $\boldsymbol{d}$ be any $p \times 1$ vector. Then

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

is attained when $\boldsymbol{x} = c\boldsymbol{B}^{-1}\boldsymbol{d}$ for any constant $c \neq 0$. **Proof:** By the extended Cauchy-Schwartz inequality: $(\boldsymbol{x}'\boldsymbol{d})^2 \leq (\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x})(\boldsymbol{d}'\boldsymbol{B}^{-1}\boldsymbol{d})$. Since $\boldsymbol{x} \neq \mathbf{0}$ and $\boldsymbol{B}$ is positive definite, $\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x} > 0$ and we can divide both sides by $\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}$ to get an upper bound

$$\frac{(\boldsymbol{x}'\boldsymbol{d})^2}{\boldsymbol{x}'\boldsymbol{B}\boldsymbol{x}} \leq \boldsymbol{d}'\boldsymbol{B}^{-1}\boldsymbol{d}.$$

Differentiating the left side with respect to $\boldsymbol{x}$ shows that maximum is attained at $\boldsymbol{x} = cB^{-1}\boldsymbol{d}$.

Maximization of a Quadratic Form on a Unit Sphere

- $\boldsymbol{B}$ is positive definite with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and associated eigenvectors (normalized) $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_p$. Then

$$\max_{\boldsymbol{x} \neq 0} \frac{\boldsymbol{x}' B \boldsymbol{x}}{\boldsymbol{x}' \boldsymbol{x}} = \lambda_1, \qquad \text{attained when } \boldsymbol{x} = \boldsymbol{e}_1$$

$$\min_{\boldsymbol{x} \neq 0} \frac{\boldsymbol{x}' B \boldsymbol{x}}{\boldsymbol{x}' \boldsymbol{x}} = \lambda_p, \qquad \text{attained when } \boldsymbol{x} = \boldsymbol{e}_p.$$

- Furthermore, for $k = 1, 2, \cdots, p-1$,

$$\max_{\boldsymbol{x} \perp \boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_k} \frac{\boldsymbol{x}' B \boldsymbol{x}}{\boldsymbol{x}' \boldsymbol{x}} = \lambda_{k+1} \qquad \text{is attained when } \boldsymbol{x} = \boldsymbol{e}_{k+1}.$$

See proof at end of chapter 2 in the textbook (pages 80-81).

Determinant, Inverse of Symmetric Partitioned Matrix

- Let $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$ and $\boldsymbol{\Sigma}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}^{11} & \boldsymbol{\Sigma}^{12} \\ \boldsymbol{\Sigma}^{21} & \boldsymbol{\Sigma}^{22} \end{bmatrix}$
- The determinant and matrix inverses and the determinants have representations as follows:
  - Then the determinant of $\boldsymbol{\Sigma}$ can be written as:

$$|\boldsymbol{\Sigma}| = |\boldsymbol{\Sigma}_{11}||\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}|$$

  - The inverse $\boldsymbol{\Sigma}^{-1}$ has partitioned elements given by:

$$\begin{aligned} \boldsymbol{\Sigma}^{11} &= \boldsymbol{\Sigma}_{11}^{-1} + \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1}\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1} \\ \boldsymbol{\Sigma}^{12} &= -\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1} \\ \boldsymbol{\Sigma}^{22} &= (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})^{-1} \end{aligned}$$

Matrix derivatives: Some useful results

- The following results are often useful in statistics:
  1. $\frac{d\boldsymbol{\Sigma}'\boldsymbol{A}\boldsymbol{\Sigma}}{d\boldsymbol{\Sigma}} = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{\Sigma}$
  2. $\frac{d\boldsymbol{a}'\boldsymbol{\Sigma}}{d\boldsymbol{\Sigma}} = \boldsymbol{a}$
  3. $\frac{d\log(|\boldsymbol{\Sigma}|)}{d\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}^{-1}$ for symmetric $\boldsymbol{\Sigma}$
  4. $\frac{d\boldsymbol{a}'\boldsymbol{\Sigma}^{-1}\boldsymbol{b}}{d\boldsymbol{\Sigma}} = -\boldsymbol{\Sigma}^{-1}\boldsymbol{a}\boldsymbol{b}'\boldsymbol{\Sigma}^{-1}$ for symmetric $\boldsymbol{\Sigma}$
     - These results are useful in coming up with MLEs for the multivariate normal distribution.
- What should we do if we cannot find a formula that we need?
  - Use matrix differential calculus!

Matrix derivatives: preliminaries

- Suppose that $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_n)$ is $m \times n$ matrix
- *vec* operator is defined as $vec(\boldsymbol{\Sigma}) = (\boldsymbol{\sigma}'_1, \ldots, \boldsymbol{\sigma}'_n)'$.
  - Thus, $vec(\boldsymbol{\Sigma})$ is a vector of length mn
- If $\boldsymbol{\Sigma} = (\boldsymbol{\sigma}_1, \ldots, \boldsymbol{\sigma}_n)$ is a symmetric $m \times m$ matrix, then a *vech* operator is defined as $vech(\boldsymbol{\Sigma}) = (\sigma_{11}, \ldots, \sigma_{m1}, \sigma_{22}, \ldots, \sigma_{m2}, \ldots, \sigma_{mm})'$
  - Thus, $vech(\boldsymbol{\Sigma})$ is a vector of length $m(m+1)/2$
- Note that
  - $vec(a') = vec(a) = a$
  - $vec(ab') = b \otimes a$, where $\otimes$ denotes Kronecker's product
- Important property: $vec(\boldsymbol{abC}) = (\boldsymbol{C}' \otimes \boldsymbol{a})vec(\boldsymbol{b})$, where $\boldsymbol{a}$, $\boldsymbol{b}$ and $\boldsymbol{C}$ are matrices such that the product matrix $\boldsymbol{abC}$ exists (prove at home)
  - This is a major tool for taking derivatives with respect to matrices and vectors

Matrix derivatives: notation

- It is important to introduce "good" notation (see Magnus and Neudecker, 1999)

	scalar variable	vector variable	matrix variable
Scalar function	$f(\sigma)$	$f(\boldsymbol{\sigma})$	$f(\boldsymbol{\Sigma})$
Vector function	$\boldsymbol{f}(\sigma)$	$\boldsymbol{f}(\boldsymbol{\sigma})$	$\boldsymbol{f}(\boldsymbol{\Sigma})$
Matrix function	$\boldsymbol{F}(\sigma)$	$\boldsymbol{F}(\boldsymbol{\sigma})$	$\boldsymbol{F}(\boldsymbol{\Sigma})$

- Derivative $\boldsymbol{Df}(\boldsymbol{\sigma})$: $\mathrm{d}\boldsymbol{f}(\boldsymbol{\sigma}) = \boldsymbol{a}(\boldsymbol{\sigma})\mathrm{d}\boldsymbol{\sigma} \quad \Leftrightarrow \quad \boldsymbol{Df}(\boldsymbol{\sigma}) = \boldsymbol{a}(\boldsymbol{\sigma})$
  - $\boldsymbol{D}f(\boldsymbol{\sigma}) = \frac{\partial f(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}$
  - $\boldsymbol{Df}(\boldsymbol{\sigma}) = \frac{\partial \boldsymbol{f}(\boldsymbol{\sigma})}{\partial \boldsymbol{\sigma}'}$
  - In a more general case: $\boldsymbol{DF}(\boldsymbol{\Sigma}) = \frac{\partial vec(\boldsymbol{F}(\boldsymbol{\Sigma}))}{\partial vec(\boldsymbol{\Sigma})'}$

Matrix derivatives: examples

- **Example 1:**
  - Find the derivative of $\boldsymbol{F}(\boldsymbol{\Sigma}) = \boldsymbol{a\Sigma b}$
    1. $\mathrm{d}\boldsymbol{F}(\boldsymbol{\Sigma}) = \boldsymbol{a}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{b}$
    2. Therefore, $\mathrm{d}vec(\boldsymbol{F}(\boldsymbol{\Sigma})) = (\boldsymbol{b}' \otimes \boldsymbol{a})\mathrm{d}vec(\boldsymbol{\Sigma})$
    3. Hence, it follows that $\boldsymbol{DF}(\boldsymbol{\Sigma}) = \boldsymbol{b}' \otimes \boldsymbol{a}$
- **Example 2:**
  - Find the derivative of $\boldsymbol{f}(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma a}$
    1. $\mathrm{d}\boldsymbol{f}(\boldsymbol{\Sigma}) = (\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{a}$
    2. Then, $vec(\boldsymbol{I}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{a}) = (\boldsymbol{a}' \otimes \boldsymbol{I})\mathrm{d}vec(\boldsymbol{\Sigma})$
    3. Hence, it follows that $\boldsymbol{Df}(\boldsymbol{\Sigma}) = \boldsymbol{a}' \otimes \boldsymbol{I}$
- **Example 3:**
  - Find the derivative of $\boldsymbol{F}(\boldsymbol{\sigma}) = \boldsymbol{\sigma\sigma}'$
    1. $\mathrm{d}\boldsymbol{F}(\boldsymbol{\sigma}) = \mathrm{d}\boldsymbol{\sigma\sigma}' = (\mathrm{d}\boldsymbol{\sigma})\boldsymbol{\sigma}' + \boldsymbol{\sigma}(\mathrm{d}\boldsymbol{\sigma})'$
    2. Then, $\mathrm{d}vec(\boldsymbol{\sigma\sigma}') = (\boldsymbol{\sigma} \otimes \boldsymbol{I})\mathrm{d}vec(\boldsymbol{\sigma}) + (\boldsymbol{I} \otimes \boldsymbol{\sigma})\mathrm{d}vec(\boldsymbol{\sigma}')$
    3. Therefore, $\mathrm{d}vec(\boldsymbol{\sigma\sigma}') = (\boldsymbol{\sigma} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{\sigma})\mathrm{d}\boldsymbol{\sigma}$
    4. Hence, $\boldsymbol{DF}(\boldsymbol{\sigma}) = \boldsymbol{\sigma} \otimes \boldsymbol{I} + \boldsymbol{I} \otimes \boldsymbol{\sigma}$
- **Example 4:**
  - Find the derivative of $\boldsymbol{F}(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{-1}$, where $\boldsymbol{\Sigma}$ is nonsingular
    1. $\mathrm{d}\boldsymbol{F}(\boldsymbol{\Sigma}) = -\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}$
    2. Then, $\mathrm{d}vec(\boldsymbol{F}(\boldsymbol{\Sigma})) = -((\boldsymbol{\Sigma}')^{-1} \otimes \boldsymbol{\Sigma}^{-1})\mathrm{d}vec(\boldsymbol{\Sigma})$
    3. Hence, $\boldsymbol{DF}(\boldsymbol{\Sigma}) = -(\boldsymbol{\Sigma}')^{-1} \otimes \boldsymbol{\Sigma}^{-1}$

Matrix derivatives: summary of some results

- The following table provides some useful results

function	differential	derivative
$\boldsymbol{a}'\sigma$	$\boldsymbol{a}'\mathrm{d}\sigma$	$\boldsymbol{a}'$
$\sigma'\boldsymbol{a}\sigma$	$\sigma'(\boldsymbol{a}+\boldsymbol{a}')\mathrm{d}\sigma$	$\sigma'(\boldsymbol{a}+\boldsymbol{a}')$
$\mathrm{t}r\{\boldsymbol{a\Sigma}\}$	$\mathrm{t}r\{\boldsymbol{a}\mathrm{d}\boldsymbol{\Sigma}\}$	$\boldsymbol{a}'$
$\lvert\boldsymbol{\Sigma}\rvert$	$\lvert\boldsymbol{\Sigma}\rvert\mathrm{t}r\{\boldsymbol{\Sigma}^{-1}\}\mathrm{d}\boldsymbol{\Sigma}$	$\lvert\boldsymbol{\Sigma}\rvert(\boldsymbol{\Sigma}^{-1})'$
$\boldsymbol{\Sigma}^{-1}$	$-\boldsymbol{\Sigma}^{-1}(\mathrm{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}$	$-(\boldsymbol{\Sigma}')^{-1}\otimes\boldsymbol{\Sigma}^{-1}$

Matrix derivatives: derivatives wrt symmetric matrices

- Issue: some elements are identical
- How should we treat the derivative with respect to symmetric matrices?
  - Recall *vech* operator: it works similar to *vec* operator but chooses unique elements only
- Consider a symmetric matrix $\boldsymbol{\Sigma}$
- Then, the following relationships hold:
  - $vec(\boldsymbol{\Sigma}) = \boldsymbol{G}vech(\boldsymbol{\Sigma})$, where $\boldsymbol{G}$ is the unique matrix or zeroes and ones that provides this equality
  - $vech(\boldsymbol{\Sigma}) = \boldsymbol{H}vec(\boldsymbol{\Sigma})$, where $\boldsymbol{H}$ is generally a nonunique matrix

Matrix derivatives: G-matrix example

- Consider a 3×3 symmetric matrix $\boldsymbol{\Sigma}$
- Then we obtain $\boldsymbol{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$
- $\boldsymbol{G}$ is the unique 9×6 matrix that consists of 0 and 1 only

Matrix derivatives: H-matrix example

- $\boldsymbol{H}$ matrix is not unique!

$$\boldsymbol{H}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{H}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Matrix derivatives: G-matrix

- For taking derivatives wrt symmetric matrices, use the following strategy:
  - obtain a derivative in vector form ignoring the symmetric structure
  - adjust the obtained result pre-multipying it by $\boldsymbol{G}'$
  - the result is a vector consisting of unique derivatives