Basic Concepts in Matrix Algebra

- An column array of \( p \) elements is called a vector of dimension \( p \) and is written as
  \[
  x_{p \times 1} = \begin{bmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_p
  \end{bmatrix}.
  \]

- The transpose of the column vector \( x_{p \times 1} \) is row vector
  \[
  x' = [x_1 \ x_2 \ \ldots \ x_p]
  \]

- A vector can be represented in \( p \)-space as a directed line with components along \( p \) axes.

Basic Matrix Concepts (cont’d)

- Two vectors can be added if they have the same dimension. Addition is carried out elementwise.
  \[
  x + y = \begin{bmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_p
  \end{bmatrix} + \begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_p
  \end{bmatrix} = \begin{bmatrix}
  x_1 + y_1 \\
  x_2 + y_2 \\
  \vdots \\
  x_p + y_p
  \end{bmatrix}
  \]

- A vector can be contracted or expanded if multiplied by a constant \( c \). Multiplication of a matrix by a scalar:
  \[
  cx = c \begin{bmatrix}
  x_1 \\
  x_2 \\
  \vdots \\
  x_p
  \end{bmatrix} = \begin{bmatrix}
  cx_1 \\
  cx_2 \\
  \vdots \\
  cx_p
  \end{bmatrix}
  \]

Examples

\[
\begin{align*}
  x &= \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix} \quad \text{and} \quad x' = [2 \ 1 \ -4] \\
  6x &= 6 \times \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix} = \begin{bmatrix} 6 \times 2 \\ 6 \times 1 \\ 6 \times (-4) \end{bmatrix} = \begin{bmatrix} 12 \\ 6 \\ -24 \end{bmatrix} \\
  x + y &= \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix} + \begin{bmatrix} 5 \\ -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 + 5 \\ 1 - 2 \\ -4 + 0 \end{bmatrix} = \begin{bmatrix} 7 \\ -1 \\ -4 \end{bmatrix}
\end{align*}
\]
Basic Matrix Concepts (cont’d)

• The length of a vector $\mathbf{x}$ is the Euclidean distance from the origin
  \[ L_{\mathbf{x}} = \sqrt{\sum_{j=1}^{p} x_j^2} \]

• Multiplication of a vector $\mathbf{x}$ by a constant $c$ changes the length:
  \[ L_{c\mathbf{x}} = \sqrt{\sum_{j=1}^{p} c^2 x_j^2} = |c| \sqrt{\sum_{j=1}^{p} x_j^2} = |c| L_{\mathbf{x}} \]

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

Examples

The length of $\mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ -4 \\ -2 \end{bmatrix}$ is
  \[ L_{\mathbf{x}} = \sqrt{(2)^2 + (1)^2 + (-4)^2 + (-2)^2} = \sqrt{25} = 5 \]

Then
  \[ z = \frac{1}{5} \times \begin{bmatrix} 2 \\ 1 \\ -4 \\ -2 \end{bmatrix} = \begin{bmatrix} 0.4 \\ 0.2 \\ -0.8 \\ -0.4 \end{bmatrix} \]

is a vector of unit length.

Angle Between Vectors

• Consider two vectors $\mathbf{x}$ and $\mathbf{y}$ in two dimensions. If $\theta_1$ is the angle between $\mathbf{x}$ and the horizontal axis and $\theta_2 > \theta_1$ is the angle between $\mathbf{y}$ and the horizontal axis, then
  \[ \cos(\theta_1) = \frac{x_1}{L_{\mathbf{x}}} \quad \cos(\theta_2) = \frac{y_1}{L_{\mathbf{y}}} \]
  \[ \sin(\theta_1) = \frac{x_2}{L_{\mathbf{x}}} \quad \sin(\theta_2) = \frac{y_2}{L_{\mathbf{y}}} \]

If $\theta$ is the angle between $\mathbf{x}$ and $\mathbf{y}$, then
  \[ \cos(\theta) = \cos(\theta_2 - \theta_1) = \cos(\theta_2) \cos(\theta_1) + \sin(\theta_2) \sin(\theta_1) \]

Then
  \[ \cos(\theta) = \frac{x_1 y_1 + x_2 y_2}{L_{\mathbf{x}} L_{\mathbf{y}}} = \frac{\mathbf{x}^T \mathbf{y}}{\sqrt{\mathbf{x}^T \mathbf{x}} \sqrt{\mathbf{y}^T \mathbf{y}}} \]

Angle Between Vectors (cont’d)
**Inner Product**

- The inner product between two vectors $x$ and $y$ is
  \[ x'y = \sum_{j=1}^{p} x_j y_j. \]

- Then $L_x = \sqrt{x'x}$, $L_y = \sqrt{y'y}$ and
  \[ \cos(\theta) = \frac{x'y}{\sqrt{(x'x)(y'y)}}. \]

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

**Linear Dependence**

- Two vectors, $x$ and $y$, are *linearly dependent* if there exist two constants $c_1$ and $c_2$, not both zero, such that
  \[ c_1 x + c_2 y = 0 \]
- If two vectors are linearly dependent, then one can be written as a linear combination of the other. From above:
  \[ x = (c_2/c_1)y \]
- $k$ vectors, $x_1, x_2, \ldots, x_k$, are linearly dependent if there exist constants $(c_1, c_2, \ldots, c_k)$ not all zero such that
  \[ \sum_{j=1}^{k} c_j x_j = 0 \]
- Vectors of the same dimension that are not linearly dependent are said to be *linearly independent*.

**Linear Independence-example**

Let
\[ x_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \]

Then $c_1 x_1 + c_2 x_2 + c_3 x_3 = 0$ if
\[
\begin{align*}
    c_1 + c_2 + c_3 &= 0 \\
    2c_1 + 0 - 2c_3 &= 0 \\
    c_1 - c_2 + c_3 &= 0
\end{align*}
\]

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

**Projections**

- The projection of $x$ on $y$ is defined by
  \[ y = \frac{x'y}{y'y} y = \frac{x'y}{L_y} L_y y. \]
- The length of the projection is
  \[ \text{Length of projection} = \frac{|x'y|}{L_y} = \frac{L_x|x'y|}{L_x L_y} = L_x|\cos(\theta)|, \]
  where $\theta$ is the angle between $x$ and $y$. 

68 69 70 71
Matrices

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

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

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

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

Matrix Algebra

- Multiplication of $A$ by a constant $c$ is carried out element by element.

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

Matrix Addition

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

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

Examples

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

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

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

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

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

Identity Matrix

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

  $I_{2 \times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ and $I_{3 \times 3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

- If $A$ is a square matrix, then $AI = IA = A$.

- $I_{n \times n}A_{n \times p} = A_{n \times p}$ but $A_{n \times p}I_{n \times n}$ is not defined for $p \neq n$.

Multiplication Examples

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

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

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

Symmetric Matrices

- A square matrix is symmetric if $A = A'$.

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

- Examples

  \[
  \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} \quad \begin{bmatrix} 5 & 1 & -3 \\ 1 & 12 & -5 \\ -3 & -5 & 9 \end{bmatrix}
  \]
Inverse Matrix

- Consider two square matrices $A_{k \times k}$ and $B_{k \times k}$. If $AB = BA = I$ then $B$ is the inverse of $A$, denoted $A^{-1}$.

- The inverse of $A$ exists only if the columns of $A$ are linearly independent.

- If $A = \text{diag}(a_{ij})$ then $A^{-1} = \text{diag}(1/a_{ij})$.

Orthogonal Matrices

- A square matrix $Q$ is orthogonal if $QQ' = Q'Q = I$, or $Q' = Q^{-1}$.

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

Eigenvalues and Eigenvectors

- A square matrix $A$ has an eigenvalue $\lambda$ with corresponding eigenvector $z \neq 0$ if $Az = \lambda z$.

- The eigenvalues of $A$ are the solution to $|A - \lambda I| = 0$.

- A normalized eigenvector (of unit length) is denoted by $e$.

- A $k \times k$ matrix $A$ has $k$ pairs of eigenvalues and eigenvectors $\lambda_1, e_1, \lambda_2, e_2, \ldots, \lambda_k, e_k$ where $e'_i e_i = 1$, $e'_j e_j = 0$ and the eigenvectors are unique up to a change in sign unless two or more eigenvalues are equal.
Spectral Decomposition

- Eigenvectors 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 $A$ is

\[
A = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + \ldots + \lambda_k e_k e_k'
\]

\[
= [e_1 \ e_2 \ \ldots \ e_k] \begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_k
\end{bmatrix} [e_1 \ e_2 \ \ldots \ e_k]'
\]

\[= P \Lambda P'
\]

Determinant and Trace

- The trace of a $k \times k$ matrix $A$ is the sum of the diagonal elements, i.e., $\text{trace}(A) = \sum_{i=1}^{k} a_{ii}$

- The trace of a square, symmetric matrix $A$ is the sum of the eigenvalues, i.e., $\text{trace}(A) = \sum_{i=1}^{k} a_{ii} = \sum_{i=1}^{k} \lambda_i$

- The determinant of a square, symmetric matrix $A$ is the product of the eigenvalues, i.e., $|A| = \prod_{i=1}^{k} \lambda_i$

Rank of a Matrix

- The rank of a square matrix $A$ 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 $A$ exists, if and only if $\text{rank}(A) = k$

  i.e., there are no zero eigenvalues

Positive Definite Matrix

- For a $k \times k$ symmetric matrix $A$ and a vector $x = [x_1, x_2, \ldots, x_k]'$ the quantity $x'Ax$ is called a quadratic form

- Note that $x'Ax = \sum_{i=1}^{k} \sum_{j=1}^{k} a_{ij}x_ix_j$

- If $x'Ax \geq 0$ for any vector $x$, both $A$ and the quadratic form are said to be non-negative definite.

- If $x'Ax > 0$ for any vector $x \neq 0$, both $A$ 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
  \[
  A = \begin{bmatrix}
  3 & -\sqrt{2} \\
  -\sqrt{2} & 2
  \end{bmatrix},
  \]
  the eigenvalues are \(\lambda_1 = 4, \lambda_2 = 1\). Then \(A = 4e_1e_1' + e_2e_2'\).
  Write
  \[
  x'Ax = 4x'e_1e_1'x + x'e_2e_2'x = 4y_1^2 + y_2^2 \geq 0,
  \]
  and is zero only for \(y_1 = y_2 = 0\).

Distance and Quadratic Forms

• For \(x = [x_1, x_2, ..., x_p]'\) and a \(p \times p\) positive definite matrix \(A\),
  \[
  d^2 = x'Ax > 0
  \]
  when \(x \neq 0\). Thus, a positive definite quadratic form can be interpreted as a squared distance of \(x\) from the origin and vice versa.

• The squared distance from \(x\) to a fixed point \(\mu\) is given by the quadratic form
  \[
  (x - \mu)'A(x - \mu).
  \]

Example 2.11 (cont’d)

• \(y_1\) and \(y_2\) cannot be zero unless \(x_1\) and \(x_2\) are zero because
  \[
  \begin{bmatrix}
  y_1 \\
  y_2
  \end{bmatrix} = \begin{bmatrix}
  e_1' \\
  e_2'
  \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = P_{2 \times 2}x_{2 \times 1}
  \]
  with \(P\) orthonormal so that \((P)'^{-1} = P\). Then \(x = Py\) and since \(x \neq 0\) it follows that \(y \neq 0\).

• 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 \(\geq 0\).

Distance and Quadratic Forms (cont’d)

• We can interpret distance in terms of eigenvalues and eigenvectors of \(A\) as well. Any point \(x\) at constant distance \(c\) from the origin satisfies
  \[
  x'Ax = x'(\sum_{j=1}^{p} \lambda_j e_j e_j')x = \sum_{j=1}^{p} \lambda_j (x'e_j)^2 = c^2,
  \]
  the expression for an ellipsoid in \(p\) dimensions.

• Note that the point \(x = c\lambda_1^{-1/2}e_1\) is at a distance \(c\) (in the direction of \(e_1\)) from the origin because it satisfies \(x'Ax = c^2\). The same is true for points \(x = c\lambda_j^{-1/2}e_j, j = 1, ..., p\). 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 symmetric positive definite matrix $A$ yields

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

with $\Lambda_{k \times k} = \text{diag}(\lambda_j)$, all $\lambda_j > 0$, and $P_{k \times k} = [e_1 \ e_2 \ ... \ e_p]$ an orthonormal matrix of eigenvectors. Then

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

Square-Root Matrices

The square root of a symmetric positive definite matrix $A$ has the following properties:

1. Symmetry: $(A^{1/2})' = A^{1/2}$
2. $A^{1/2}A^{1/2} = A$
3. $A^{-1/2} = \sum_{j=1}^{p} \lambda_j^{-1/2} e_j e'_j = P\Lambda^{-1/2}P'$
4. $A^{1/2}A^{-1/2} = A^{-1/2}A^{1/2} = I$
5. $A^{-1/2}A^{-1/2} = A^{-1}$
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

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

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.

Mean Vectors and Covariance Matrices

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

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

Mean Vectors and Covariance Matrices (cont’d)

- 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

$$\sigma_{jk} = E[(X_j - \mu_j)(X_k - \mu_k)]$$
$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_j - \mu_j)(x_k - \mu_k)f_{jk}(x_j,x_k) dx_j dx_k.$$
Mean Vectors and Covariance Matrices (cont’d)

- 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 $X$ 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 (cont’d)

- We use $\mu$ to denote the $p \times 1$ vector of marginal population means and use $\Sigma$ to denote the $p \times p$ population variance-covariance matrix:
  \[ \Sigma = E[(X - \mu)(X - \mu)^\top]. \]

- If we carry out the multiplication (outer product) then $\Sigma$ is equal to:
  \[
  E \begin{bmatrix}
  (X_1 - \mu_1)^2 & (X_1 - \mu_1)(X_2 - \mu_2) & \cdots & (X_1 - \mu_1)(X_p - \mu_p) \\
  (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_p - \mu_p) \\
  \vdots & \vdots & \ddots & \vdots \\
  (X_p - \mu_p)(X_1 - \mu_1) & (X_p - \mu_p)(X_2 - \mu_2) & \cdots & (X_p - \mu_p)^2
  \end{bmatrix}.
  \]

Correlation Matrix

- The population correlation matrix is the $p \times p$ matrix with off-diagonal elements equal to $\rho_{jk}$ and diagonal elements equal to 1.
  \[
  \begin{bmatrix}
  1 & \rho_{12} & \cdots & \rho_{1p} \\
  \rho_{21} & 1 & \cdots & \rho_{2p} \\
  \vdots & \vdots & \ddots & \vdots \\
  \rho_{p1} & \rho_{p2} & \cdots & 1
  \end{bmatrix}.
  \]

- Since $\rho_{ij} = \rho_{ji}$ the correlation matrix is symmetric.

- The correlation matrix is also non-negative definite.
Correlation Matrix (cont’d)

- The $p \times p$ population standard deviation matrix $V^{1/2}$ is a diagonal matrix with $\sqrt{\sigma_{jj}}$ along the diagonal and zeros in all off-diagonal positions. Then

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

and the population correlation matrix is

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

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

Partitioning Random vectors

- If we partition the random $p \times 1$ vector $X$ into two components $X_1, X_2$ 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:

$$E(X) = E \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

- Partitioned variance-covariance matrix:

$$\Sigma = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{bmatrix},$$

where $\Sigma_{11}$ is $q \times q$, $\Sigma_{12}$ is $q \times (p-q)$ and $\Sigma_{22}$ is $(p-q) \times (p-q)$.

Partitioning Covariance Matrices (cont’d)

- $\Sigma_{11}, \Sigma_{22}$ 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 $\Sigma_{12}$, only covariances. These covariances reflect linear associations between elements in the two different sub-vectors.

Linear Combinations of Random variables

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

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

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

$$\mu_Z = C\mu_X \quad \text{and} \quad \Sigma_Z = C\Sigma_X C'.$$
Cauchy-Schwarz Inequality

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

**Cauchy-Schwarz inequality** Let \( b \) and \( d \) be any two \( p \times 1 \) vectors. Then,
\[
(b'd)^2 \leq (b'b)(d'd)
\]
with equality only if \( b = cd \) for some scalar constant \( c \).

**Proof:** The equality is obvious for \( b = 0 \) or \( d = 0 \). For other cases, consider \( b - cd \) for any constant \( c \neq 0 \). Then if \( b - cd \neq 0 \), we have
\[
0 < (b - cd)'(b - cd) = b'b - 2c(b'd) + c^2d'd,
\]
since \( b - cd \) must have positive length.

Extended Cauchy-Schwarz Inequality

If \( b \) and \( d \) are any two \( p \times 1 \) vectors and \( B \) is a \( p \times p \) positive definite matrix, then
\[
(b'd)^2 \leq (b'Bb)(d'B^{-1}d)
\]
with equality if and only if \( b = cB^{-1}d \) or \( d = cBb \) for some constant \( c \).

**Proof:** Consider \( B^{1/2} = \sum_{i=1}^{p} \sqrt{\lambda_i} e_i e_i' \), and \( B^{-1/2} = \sum_{i=1}^{p} \frac{1}{\sqrt{\lambda_i}} e_i e_i' \). Then we can write
\[
b'd = b'd = b' B^{1/2} B^{-1/2} d = (B^{1/2} b)' (B^{-1/2} d) = b^* d^*.
\]

To complete the proof, simply apply the Cauchy-Schwarz inequality to the vectors \( b^* \) and \( d^* \).

Cauchy-Schwarz Inequality

We can add and subtract \( (b'd)^2/(d'd) \) to obtain
\[
0 < b'b - 2c(b'd) + c^2d'd = b'b - (b'd)^2/d'd + (d'd)(c - b'd/d'd)^2
\]
Since \( c \) can be anything, we can choose \( c = b'd/d'd \). Then,
\[
0 < b'b - (b'd)^2/d'd \Rightarrow (b'd)^2 < (b'b)(d'd)
\]
for \( b \neq cd \) (otherwise, we have equality).

Optimization

Let \( B \) be positive definite and let \( d \) be any \( p \times 1 \) vector. Then
\[
\max_{x \neq 0} \frac{(x'd)^2}{x' B x} = d' B^{-1} d
\]
is attained when \( x = cB^{-1}d \) for any constant \( c \neq 0 \).

**Proof:** By the extended Cauchy-Schwarz inequality: \( (x'd)^2 \leq (x'Bx)(d'B^{-1}d) \). Since \( x \neq 0 \) and \( B \) is positive definite, \( x'Bx > 0 \) and we can divide both sides by \( x'Bx \) to get an upper bound
\[
\frac{(x'd)^2}{x'Bx} \leq d' B^{-1} d.
\]
Differentiating the left side with respect to \( x \) shows that maximum is attained at \( x = cB^{-1}d \).
Maximization of a Quadratic Form on a Unit Sphere

• $B$ is positive definite with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ and associated eigenvectors (normalized) $e_1, e_2, \cdots, e_p$. Then

max $x' B x$ is attained when $x = e_k$. 

Furthemore, for $k = 1, 2, \ldots, p - 1$,

max $x' B x$ is attained when $x \perp e_1, e_2, \cdots, e_k$. 

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