Basic Concepts in Matrix Algebra

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

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

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

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

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

Basic Matrix Concepts (cont'd)

- Two vectors can be added if they have the same dimension. Addition is carried out elementwise.

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

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

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

Examples

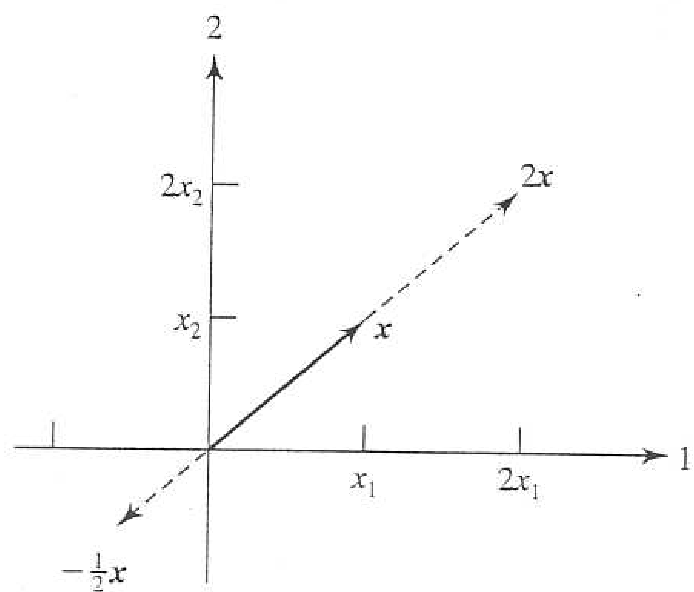
$$\mathbf{x} = \begin{bmatrix} 2 \\ 1 \\ -4 \end{bmatrix} \text{ and } \mathbf{x}' = \begin{bmatrix} 2 & 1 & -4 \end{bmatrix}$$

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

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

Basic Matrix Concepts (cont'd)

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



Basic Matrix Concepts (cont'd)

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

$$L_x = \sqrt{\sum_{j=1}^{p} x_j^2}$$

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

$$L_{cx} = \sqrt{\sum_{j=1}^{p} c^2 x_j^2} = |c| \sqrt{\sum_{j=1}^{p} x_j^2} = |c| L_x$$

- If $c = L_x^{-1}$, 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

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

Then

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

is a vector of unit length.

Angle Between Vectors

- Consider two vectors $\mathbf{x}$ and $\mathbf{y}$ in two dimensions. If θ_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

$$\cos(\theta_1) = \frac{x_1}{L_x} \qquad \cos(\theta_2) = \frac{y_1}{L_y}$$
$$\sin(\theta_1) = \frac{x_2}{L_x} \qquad \sin(\theta_2) = \frac{y_2}{L_y},$$

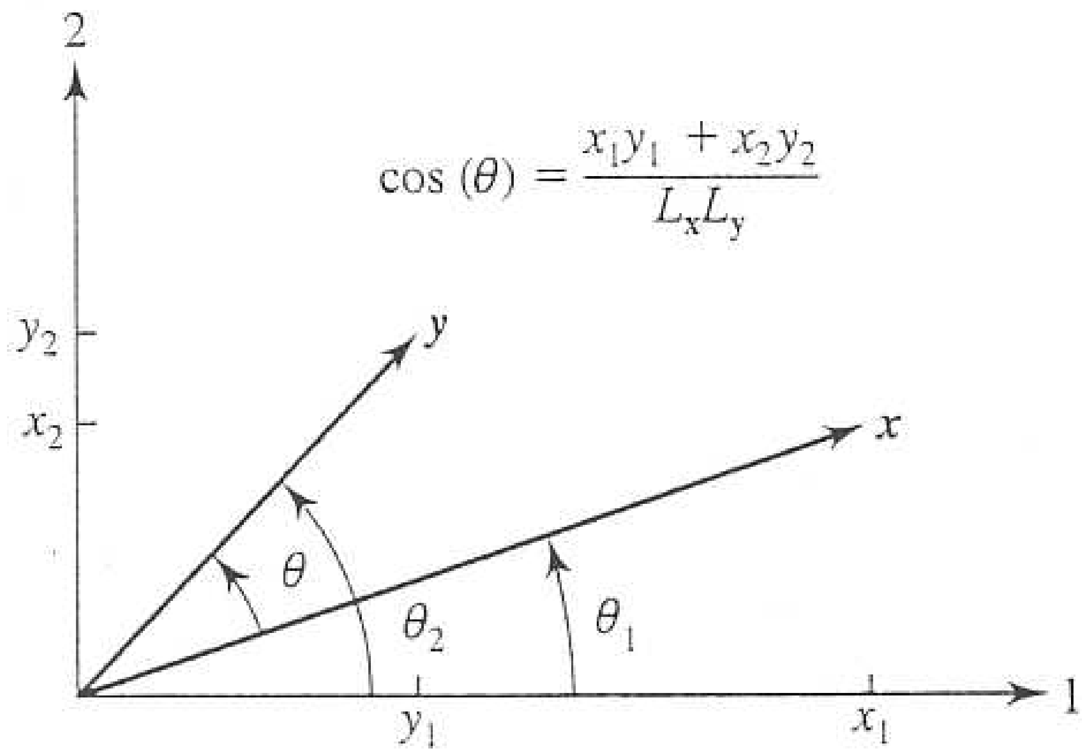
  If θ is the angle between $\mathbf{x}$ and $\mathbf{y}$, then

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

  Then

$$\cos(\theta) = \frac{x_1 y_1 + x_2 y_2}{L_x L_y}.$$

Angle Between Vectors (cont'd)



Inner Product

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

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

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

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

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

Linear Dependence

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

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

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

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

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

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

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

Linear Independence-example

Let

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

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

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

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

Projections

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

$$\text{Projection of } \mathbf{x} \text{ on } \mathbf{y} = \frac{\mathbf{x}'\mathbf{y}}{\mathbf{y}'\mathbf{y}}\mathbf{y} = \frac{\mathbf{x}'\mathbf{y}}{L_y}\frac{1}{L_y}\mathbf{y}.$$

- The length of the projection is

$$\text{Length of projection} = \frac{|\mathbf{x}'\mathbf{y}|}{L_y} = L_x\frac{|\mathbf{x}'\mathbf{y}|}{L_xL_y} = L_x|\cos(\theta)|,$$

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

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

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

Examples

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

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

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

Matrix Multiplication

- Multiplication of two matrices $A_{n\times p}$ 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.

Multiplication Examples

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

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

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

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,

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

- 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 $p \neq n$.

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

$$\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} \qquad \begin{bmatrix} 5 & 1 & -3 \\ 1 & 12 & -5 \\ -3 & -5 & 9 \end{bmatrix}$$

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 $A^{-1} = \text{diag}\{1/a_{ij}\}$.

Inverse Matrix

- For a 2×2 matrix A, the inverse is

$$A^{-1} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{\det(A)} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix},$$

where $\det(A) = (a_{11} \times a_{22}) - (a_{12} \times a_{21})$ denotes the *determinant* of A.

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 ($\mathbf{q}_j'\mathbf{q}_j = 1$) and are mutually perpendicular ($\mathbf{q}_j'\mathbf{q}_k = 0$ for any $j \neq k$).

Eigenvalues and Eigenvectors

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

$$A\mathbf{z} = \lambda\mathbf{z}$$

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

- A normalized eigenvector (of unit length) is denoted by $\mathbf{e}$.

- A $k \times k$ matrix A has k pairs of eigenvalues and eigenvectors

$$\lambda_1, \mathbf{e}_1 \qquad \lambda_2, \mathbf{e}_2 \qquad \ldots \qquad \lambda_k, \mathbf{e}_k$$

where $\mathbf{e}_i'\mathbf{e}_i = 1$, $\mathbf{e}_i'\mathbf{e}_j = 0$ 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 A is

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

Determinant and Trace

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

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

- 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

$$rank(A) = k$$

  i.e., there are no zero eigenvalues

Positive Definite Matrix

- For a $k \times k$ symmetric matrix A and a vector $\mathbf{x} = [x_1, x_2, ..., x_k]'$ the quantity $\mathbf{x}'A\mathbf{x}$ is called a *quadratic form*

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

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

- If $\mathbf{x}'A\mathbf{x} > 0$ for any vector $\mathbf{x} \neq \mathbf{0}$, 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 $A = 4\mathbf{e}_1\mathbf{e}_1' + \mathbf{e}_2\mathbf{e}_2'$. Write

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

and is zero only for $y_1 = y_2 = 0$.

Example 2.11 (cont'd)

- y_1, y_2 cannot be zero because

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

  with P' orthonormal so that $(P')^{-1} = P$. Then $\mathbf{x} = P\mathbf{y}$ and since $\mathbf{x} \neq 0$ it follows that $\mathbf{y} \neq 0$.

- 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 $\mathbf{x} = [x_1, x_2, ..., x_p]'$ and a $p \times p$ positive definite matrix A,

$$d^2 = \mathbf{x}' A \mathbf{x} > 0$$

  when $x \neq 0$. Thus, a positive definite quadratic form can be interpreted as a squared distance of $\mathbf{x}$ from the origin and vice versa.

- The squared distance from $\mathbf{x}$ to a fixed point $\boldsymbol{\mu}$ is given by the quadratic form

$$(\mathbf{x} - \boldsymbol{\mu})' A (\mathbf{x} - \boldsymbol{\mu}).$$

Distance and Quadratic Forms (cont'd)

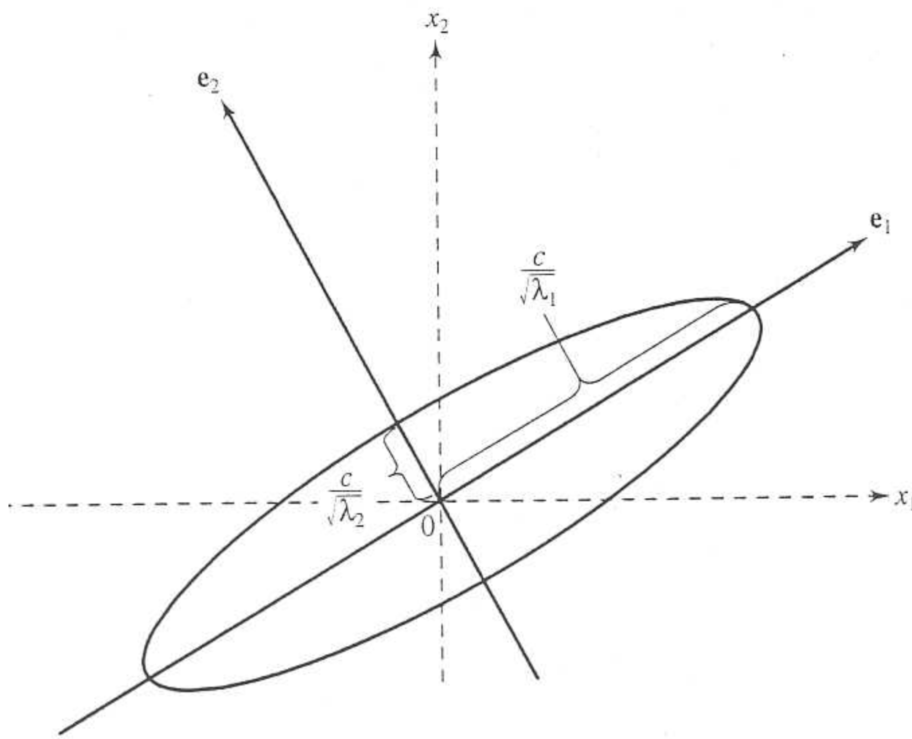
- We can interpret distance in terms of eigenvalues and eigenvectors of A as well. Any point $\mathbf{x}$ at constant distance c from the origin satisfies

$$\mathbf{x}'A\mathbf{x} = \mathbf{x}'(\sum_{j=1}^{p} \lambda_j \mathbf{e}_j \mathbf{e}_j')\mathbf{x} = \sum_{j=1}^{p} \lambda_j (\mathbf{x}'\mathbf{e}_j)^2 = c^2,$$

  the expression for an ellipsoid in p dimensions.

- Note that the point $\mathbf{x} = c\lambda_1^{-1/2}\mathbf{e}_1$ is at a distance c (in the direction of $\mathbf{e}_1$) from the origin because it satisfies $\mathbf{x}'A\mathbf{x} = c^2$. The same is true for points $\mathbf{x} = c\lambda_j^{-1/2}\mathbf{e}_j$, $j = 1, ..., p$. 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

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

with $\Lambda_{k \times k} = \text{diag}\{\lambda_j\}$, all $\lambda_j > 0$, and $P_{k \times k} = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \ldots \quad \mathbf{e}_p]$ an orthonormal matrix of eigenvectors. Then

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

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

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

Square-Root Matrices

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

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

  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.

Linear Combinations

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

$$\begin{aligned} E(X+Y) &= E(X)+E(Y) \\ E(AX) &= AE(X) \\ E(AXB) &= AE(X)B \\ E(AX+BY) &= AE(X)+BE(Y) \end{aligned}$$

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

$$E(cX) = cE(X).$$

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 μ_j and variance σ_{jj} defined in the usual way:

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

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

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

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

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

- If we carry out the multiplication (outer product)then Σ is equal to:

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

Mean Vectors and Covariance Matrices (cont'd)

- By taking expectations element-wise we find that

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

- Since $\sigma_{jk} = \sigma_{kj}$ for all $j \neq k$ we note that Σ is symmetric.

- Σ is also non-negative definite

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

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 Σ, 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:

$$E(X) = E \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$$

- Partitioned variance-covariance matrix:

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

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

Partitioning Covariance Matrices (cont'd)

- Σ_{11}, Σ_{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 Σ_{12}, only covariances. These covariancs 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 $\boldsymbol{\mu}$ and variance covariance matrix Σ, and let $\mathbf{c}$ be a $p \times 1$ vector of constants. Then the linear combination $\mathbf{c}'X$ has mean and variance:

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

- 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

$$\mu_Z = C\mu_X \text{ and } \Sigma_Z = C\Sigma_X C'.$$

Cauchy-Schwarz Inequality

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

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

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

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

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

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

since $\mathbf{b} - c\mathbf{d}$ must have positive length.

Cauchy-Schwarz Inequality

We can add and subtract $(\mathbf{b}'\mathbf{d})^2/(\mathbf{d}'\mathbf{d})$ to obtain

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

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

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

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

Extended Cauchy-Schwarz Inequality

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

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

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

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

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

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

Optimization

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

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

is attained when $\mathbf{x} = cB^{-1}\mathbf{d}$ for any constant $c \neq 0$.

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

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

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

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) $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_p$. Then

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

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

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

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

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