**Instituto Superior de Economia e Gestão**
**Universidade de Lisboa**
Mestrado em Econometria Aplicada e Previsão
(MSc in Applied Econometrics and Forecasting)
Econometrics
Paulo M.D.C. Parente
**Elements of Matrix Algebra**

# Vectors

A vector is an ordered set of numbers either organised usually in a column. Suppose $\mathbf{a}$ is a $(n \times 1)$ vector (also written $n - vector$) with typical element $a_i$, then we write

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

- A scalar is defined as a $(1 \times 1)$ vector.
- Define $\mathbf{0}_n$ as being the $(n \times 1)$ vector of zeros (every element of the vector equals zero). This is called the null vector.
- Write $\mathbf{a} \neq \mathbf{0}_n$ if $\mathbf{a}$ has at least one non-zero element.
- The transpose of $\mathbf{a}$ is denoted $\mathbf{a}'$, the $(1 \times n)$ vector corresponding to a row of $n$ numbers

  $$\mathbf{a}' = \begin{bmatrix} a_1 & a_2 & ... & a_n \end{bmatrix}$$

  Thus $\mathbf{a}'$ is a row vector. Sometimes a row vector is written as $\mathbf{a}' = (a_1, a_2, ..., a_n)$, thus the column vector is $\mathbf{a} = (a_1, a_2, ..., a_n)'$.
- Scalar, dot or inner product: If $\mathbf{a}$ and $\mathbf{b}$ are $(n \times 1)$ vectors:

  $$\mathbf{a}'\mathbf{b} = [a_1 \;\; a_2 \;\; \cdots \;\; a_n] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \sum_{i=1}^{n} a_i b_i = \mathbf{b}'\mathbf{a}$$

- The vectors $\mathbf{a}$ and $\mathbf{b}$ are orthogonal if $\mathbf{a}'\mathbf{b} = \mathbf{0}$
- For any $(n \times 1)$ vector $\mathbf{a}$, except for the null vector $\mathbf{a}'\mathbf{a} > \mathbf{0}$
- For the $(n \times 1)$ vectors $\mathbf{a}$ and $\mathbf{b}$, we write $\mathbf{a} > \mathbf{b}$ if $a_i > b_i$ for all $i = 1, ..., n$.
- The length or Euclidean norm of $\mathbf{a}$; $(n \times 1)$ is defined as

  $$\|\mathbf{a}\| = \sqrt{\sum_{i=1}^{n} a_i^2}.$$

  If $\mathbf{b}$ is also a $(n \times 1)$ vector the distance between $\mathbf{a}$ and $\mathbf{b}$ is

  $$\|\mathbf{a} - \mathbf{b}\| = \sqrt{\sum_{i=1}^{n} (a_i - b_i)^2}.$$

## Matrices

A rectangular array of numbers is called a *matrix*. A matrix with $n$ rows and $m$ columns is referred to as an $n \times m$ matrix. Let $\mathbf{A}$ be a $(n \times m)$ matrix with typical element $a_{ij}$ in row $i = 1, ..., n$; column $j = 1, ..., m$, then

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1,m-1} & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2,m-1} & a_{2m} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{n,m-1} & a_{nm} \end{bmatrix},$$

For the matrix $\mathbf{A}$; each of its rows can be considered as a row vector, and analogously, each of its columns can be seen as a column vector. The matrix $\mathbf{A}$ can be partitioned in terms of its rows; such that $\mathbf{A}$ is a vector of row vectors

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}'_{1\bullet} \\ \mathbf{a}'_{2\bullet} \\ \vdots \\ \mathbf{a}'_{n\bullet} \end{bmatrix},$$

where $\mathbf{a}'_{i\bullet} = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{i,m-1} & a_{im} \end{bmatrix}$ for $i = 1, ..., n$. Alternatively we could write

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_{\bullet 1} & \mathbf{a}_{\bullet 2} & \cdots & \mathbf{a}_{\bullet m-1} & \mathbf{a}_{\bullet m} \end{bmatrix},$$

where $\mathbf{a}_{\bullet 1}$ $(n \times 1)$, $j = 1, ..., m$ are the $m$ columns of $\mathbf{A}$, each column having $n$ elements,

$$\mathbf{a}_{\bullet j} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{bmatrix}, j = 1, ..., m.$$

- A matrix is a square matrix of order $n$ if $n = m$.

### Diagonal matrix

It is a square matrix with the elements off the leading diagonal equal zero.

$$\mathbf{A} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n \end{bmatrix}$$

Sometimes written $\mathbf{A} = diag\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$.

### Basic Rules

1. Multiplication **by a scalar:** If $\mathbf{A}$ is $n \times m$, and $c$ is $1 \times 1$ (a scalar),

$$c\mathbf{A} = \mathbf{A}c$$

   is $n \times m$ with elements

$$\{ca_{ij}\}.$$

2. **A matrix multiplied by a matrix**: If $\mathbf{A}$ is $n \times m$ and if $\mathbf{B}$ is $p \times q$, then

$$\mathbf{C} = \mathbf{AB}$$

   is defined if

$$m = p,$$

and $\mathbf{C}$ is $n \times q$, with:

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj}, \qquad i = 1, \ldots, n; \;\; j = 1, \ldots, q.$$

**Example:** If

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix},$$

then

$$\mathbf{C} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}.$$

Typically $\mathbf{AB} \neq \mathbf{BA}$.

3. Let $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ well defined matrices. Then $\mathbf{ABC} = \mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$

4. The $(i, j)^{th}$ element of matrix

$$\mathbf{C} = \mathbf{A} + \mathbf{B},$$

where $\mathbf{A}$ and $\mathbf{B}$ are both $(n \times m)$, is

$$\begin{aligned} c_{ij} &= a_{ij} + b_{ij}, \\ i &= 1, ..., n \\ j &= 1, ..., m \end{aligned}$$

5. Let $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ well defined matrices. Then $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.

6. The $(n \times n)$ identity matrix $\mathbf{I}_n$ is defined as $\mathbf{I}_n = diag\,(1, 1, \ldots, 1)$ and satisfies $\mathbf{AI}_n = \mathbf{I}_n\mathbf{A} = \mathbf{A}$.

## Transpose

The transpose of the $(n \times m)$ matrix $\mathbf{A}$ is denoted $\mathbf{A}'$ and is a $(m \times n)$ matrix. The rows which make up $\mathbf{A}'$ are the columns of $\mathbf{A}$; thus

$$\mathbf{A}' = \begin{bmatrix} \mathbf{a}'_{\bullet 1} \\ \mathbf{a}'_{\bullet 2} \\ \vdots \\ \mathbf{a}'_{\bullet m} \end{bmatrix}$$

The first row becomes the first column, the second row the second column etc.

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \mathbf{A}' = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Properties

1. $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$.
2. $(\mathbf{A}')' = \mathbf{A}$.
3. $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$
4. $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$.
5. If $\mathbf{A}$ is symmetric then $\mathbf{A}' = \mathbf{A}$, i.e. $a_{ij} = a_{ji}$ for all $i, j$.

6. If $\mathbf{A}$ is $(n \times m)$, then the matrix product $\mathbf{AB}$ is defined only if $\mathbf{B}$ is $(m \times p)$, i.e. the number of rows in $\mathbf{B}$ must equal the number of columns in $\mathbf{A}$: Writing

$$\underset{(n\times m)}{\mathbf{A}} = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_n' \end{bmatrix} \text{ and } \mathbf{B} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & ... & \mathbf{b}_p \end{bmatrix}$$

where $\mathbf{a}_i' = \begin{bmatrix} a_{i1} & a_{i2} & ... & a_{im} \end{bmatrix}$ $(1 \times m)$ and $\mathbf{b}_j = \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{mj} \end{bmatrix}$ $(m \times 1)$, $j = 1, ..., p$

$$\mathbf{AB} = \begin{bmatrix} \mathbf{a}_1'\mathbf{b}_1 & \mathbf{a}_1'\mathbf{b}_2 & \dots & \mathbf{a}_1'\mathbf{b}_p \\ \mathbf{a}_2'\mathbf{b}_1 & \mathbf{a}_2'\mathbf{b}_2 & \dots & \mathbf{a}_2'\mathbf{b}_p \\ \vdots & \vdots & & \vdots \\ \mathbf{a}_n'\mathbf{b}_1 & \mathbf{a}_n'\mathbf{b}_2 & \dots & \mathbf{a}_n'\mathbf{b}_p \end{bmatrix}$$

## Outer product

$$\mathbf{ab}' = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} [b_1 \ \ b_2 \ \cdots \ b_n] = \begin{bmatrix} a_1b_1 & a_1b_2 & \dots & a_1b_n \\ a_2b_1 & a_2b_2 & \dots & a_2b_n \\ \vdots & & & \vdots \\ a_nb_1 & \dots & \dots & a_nb_n \end{bmatrix}$$

## Linear Dependence and Rank

### Definitions

- If $\mathbf{A}$ is a $(n \times m)$ matrix and $\mathbf{b}$ is a $(m \times 1)$ vector, then the $(n \times 1)$ vector

  $$\mathbf{y} = \mathbf{Ab}$$

  can be expressed as a **linear combination** of the columns of $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & ... & \mathbf{a}_m \end{bmatrix}$ thus

  $$y = \sum_{j=1}^{m} \mathbf{a}_j b_j$$

  where $b_j$ is the $j$ th element of the vector $\mathbf{b}$.

- The set of $(n \times 1)$ vectors $\mathbf{a_j}$ , $j = 1, ..., m$ is **linearly dependent** if any of the vectors can be written as a linear combination of the others. That is, there exists values for $b_1, b_2, ..., b_m$ not all zero, such that

  $$b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + ... + b_m\mathbf{a}_m = 0$$

- The set of m-vectors $\mathbf{a_j}$ is **linearly independent** if the only solution to

  $$b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + ... + b_m\mathbf{a}_m = 0$$

  is $b_1 = b_2 = ... = b_m = 0$ ( the solution to $\mathbf{Ab} = \mathbf{0}$ is $\mathbf{b} = \mathbf{0}$).

- Let $\mathbf{A}$ be $(n \times m)$ matrix that can be viewed as a set of column vectors. The number of columns in the matrix equals the number of vectors in the set. The collection of all the vectors $\mathbf{b}$ such that $\mathbf{Ab} = \mathbf{0}$ constitutes a vector space called the null space of $\mathbf{A}$ or the kernel of $\mathbf{A}$.

- Definition: The rank of the matrix $\mathbf{A}$; denoted $\rho(\mathbf{A})$ ; is the maximum number of linearly independent columns (or rows). If $\rho(\mathbf{A}) = m$, then we say that the matrix $\mathbf{A}$ has full column rank, whereas if $\rho(\mathbf{A}) = n$ the matrix $\mathbf{A}$ has full row rank. If $\mathbf{A}$ is an $(n \times m)$ matrix then $\rho\,(\mathbf{A})$ is the maximum number of linearly independent rows or columns.

**Properties:**

1. Let $\mathbf{A}$ be $(n \times m)$ matrix, then $\rho(\mathbf{A}) \leq \min(m, n)$.
2. $\rho(\mathbf{A}) = \rho(\mathbf{A}')$.
3. $\rho(\mathbf{A}'\mathbf{A}) = \rho(\mathbf{A}\mathbf{A}') = \rho(\mathbf{A}\mathbf{A})$.
4. Let $\mathbf{A}$ be $(n \times m)$ matrix and $B$ be a $(m \times p)$ matrix, then $\rho(\mathbf{AB}) \leq \min[\rho(\mathbf{A}), \rho(\mathbf{B})]$.
5. Let $\mathbf{A}$ be $(n \times m)$ with $n \leq m$, matrix and $B$ be a $(m \times m)$ matrix with $\rho(\mathbf{B}) = m$, then $\rho(\mathbf{AB}) = \rho(\mathbf{A})$.

**Computing the rank of a Matrix**

To compute the rank of a matrix in practice it is necessary to introduce some definitions.

*Definition:* A row of a matrix is said to have $k$ leading zeros if the first $k$ elements of the row are all zeros and the $(k+1)$th element of the row is not zero. With this terminology, a matrix is in row echelon form if each row has more leading zeros than the row preceding it.

*Example:* The following matrices are in the row echelon form.

$$\begin{bmatrix} 1 & 2 & 3 \\ 0 & 0 & 4 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 3 & 4 \\ 0 & 0 & 4 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

We can reduce any matrix to a matrix in the row echelon form using the following elementary row operations:

1. interchange two rows of a matrix.
2. change a row by adding to it a multiple of another row. and
3. multiply each element in a row by the same number,

A simple way to compute the rank of a matrix is reduce a matrix to a row echelon form and the rank of a matrix is the number of nonzero rows in its row echelon form.

**Example:** Consider the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & -6 & 3 \\ -1 & 2 & -1 \end{bmatrix}$$

To find the rank of to write the matrices in the echelon form

$$\left(\tfrac{1}{3}\right) \begin{bmatrix} 3 & -6 & 3 \\ -1 & 2 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 3 & -6 & 3 \\ 0 & 0 & 0 \end{bmatrix}$$

Therefore $\rho(\mathbf{A}) = 1$.

## Inverse of a Matrix

If square matrix $\mathbf{A}$ $(n \times n)$ has full rank, that is $\rho(\mathbf{A}) = n$, then $\mathbf{A}$ is said to be non-singular and there is a unique matrix called inverse matrix and denoted as $\mathbf{A}^{-1}$ that satisfies

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n.$$

If $\rho(\mathbf{A}) < n$, $\mathbf{A}$ is said to be a singular matrix and the inverse $\mathbf{A}^{-1}$ does not exist.

Properties:

1. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
2. $\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}$.
3. $(\mathbf{A}')^{-1} = \left(\mathbf{A}^{-1}\right)'$.

4. If $\mathbf{A}=diag\left\{a_{11},a_{22},\ldots,a_{nn}\right\}$ then $\mathbf{A}^{-1}=diag\left\{a_{11}^{-1},a_{22}^{-1},\ldots,a_{nn}^{-1}\right\}$.
5. If $\mathbf{A}$ is an upper triangular matrix, ie

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1,n-1} & a_{1n} \\ 0 & a_{22} & \ldots & a_{2,n-1} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & 0 & \ldots & 0 & a_{nn} \end{bmatrix},$$

then $\mathbf{A}^{-1}$ is an upper triangular matrix. If $\mathbf{A}$ is a lower triangular matrix, then $\mathbf{A}^{-1}$ is a lower triangular matrix.

6. If $\mathbf{A}$ is a lower triangular matrix, ie,

$$\mathbf{A}=\begin{bmatrix} a_{11} & 0 & \ldots & 0 & 0 \\ a_{21} & a_{22} & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n-1,1} & a_{n-1,2} & & a_{n-1,n-1} & 0 \\ a_{n1} & a_{n,2} & \ldots & a_{n,n-1} & a_{nn} \end{bmatrix},$$

then $\mathbf{A}^{-1}$ is a lower triangular matrix.

## Determinant of a matrix

The determinant of a matrix is a scalar that is computed from a square matrix of numbers by a rule of combining products of the matrix entries. This number tell us something about the behaviour of a matrix. For example, the determinant is zero if and only the matrix $\mathbf{A}$ is singular (no inverse exists).

If $n=2$,

$$\mathbf{A}=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

and the determinant is $|\mathbf{A}|=a_{11}a_{22}-a_{12}a_{21}$.

The Laplace formula of the determinant is

$$\begin{aligned} |\mathbf{A}| &= \sum_{i=1}^{n}(-1)^{i+j}a_{i,j}M_{i,j} \text{ (for some chosen } j) \\ &= \sum_{j=1}^{n}(-1)^{i+j}a_{i,j}M_{i,j} \text{ (for some chosen } i) \end{aligned}$$

where $M_{i,j}$, is the determinant of the matrix that results from $\mathbf{A}$ by removing the $i$-th row and the $j$-th column, and $n$ is the length of the matrix. $M_{i,j}$ is termed the minor for entry $a_{i,j}$.
**Remark:** We obtain the minors inductively, working back to the $2\times 2$ case.

Other properties

- $\left|\mathbf{A}^{-1}\right|=\dfrac{1}{|\mathbf{A}|}$
- $|\mathbf{A}'|=|\mathbf{A}|$
- $|\mathbf{AB}|=|\mathbf{A}|\,|\mathbf{B}|$

Remember: the determinant can only be calculated for square matrices.

- If $\mathbf{A}=diag\left\{a_{11},a_{22},\ldots,a_{nn}\right\}$ then $|\mathbf{A}|=\prod_{i=1}^{n}a_{ii}$.
- If $\mathbf{A}$ is upper triangular or lower triangular, then $|\mathbf{A}|=\prod_{i=1}^{n}a_{ii}$.

## Characteristic roots and vectors

Eigenvalues are a special set of scalars associated with a linear system of equations (i.e., a matrix equation) that are sometimes also known as characteristic roots, characteristic values , proper values, or latent roots.

Let $\mathbf{A}$ be a $(n \times n)$ matrix. A scalar $\lambda$ is called an eigenvalue of $\mathbf{A}$ if there exist a nonzero column vector $x$ such that

$$\mathbf{A}x = \lambda x, \tag{1}$$

such vector is called an eigenvector or characteristic vectors belonging to $\lambda$.

This means that the vector $x$ has the property that its direction is not changed by the transformation $\mathbf{A}$, but that it is only scaled by a factor of $\lambda$.

Notice that the eigenvalues and eigenvectors satisfy

$$(\mathbf{A} - \lambda I_n)x = 0 \tag{2}$$

This is known as the eigenvalue problem.

If the matrix $(\mathbf{A} - \lambda I_n)$ is non-singular then the solution to the set of equations is $x = 0$. A nontrivial solution exists if $(\mathbf{A} - \lambda I_n)$ is singular and hence has a zero determinant. Thus

$$|\mathbf{A} - \lambda I_n| = 0$$

which is known as the characteristic equation of $\mathbf{A}$.

- The rank of $\mathbf{A}$ equals the number of non-zero eigenvalues.
- An *idempotent matrix* is a matrix that satisfies the condition $\mathbf{A}^2 = \mathbf{A}$. The eigenvalues of an idempotent matrix are equal to 0 or 1.

**Example:** Suppose that

$$\mathbf{A} = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix},$$

then the characteristic equation of $\mathbf{A}$ is

$$|\mathbf{A} - \lambda I_n| = \begin{vmatrix} 4-\lambda & 2 \\ 2 & 1-\lambda \end{vmatrix} = 0$$

or $(4-\lambda)(1-\lambda) - 4 = 0$ or $\lambda^2 - 5\lambda = 0$ and thus $\lambda = 0$ or $\lambda = 5$.

- The eigenvalues of a symmetric real matrix are real numbers.

To each eigenvalue $\lambda$ it corresponds a $(2 \times 1)$ eigenvector (characteristic vector). Now the characteristic vectors $x_1$ and $x_2$ with $x_i = \begin{bmatrix} x_{1,i} \\ x_{2,i} \end{bmatrix}, i = 1, 2$ can be derived from (2) .

When $\lambda = 0$, $\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x_{1,1} \\ x_{2,1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow x_{2,1} = -2x_{1,1}$

When $\lambda = 5$, $\begin{bmatrix} -1 & 2 \\ 2 & -4 \end{bmatrix} \begin{bmatrix} x_{1,2} \\ x_{2,2} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow x_{1,2} = 2x_{2,2}$.

Mathematicians love to normalize eigenvectors in terms of their Euclidean Distance , so all vectors are unit length, so we impose the normalising condition $x_{1,i}^2 + x_{2,i}^2 = 1$ that will allow us to achieve uniqueness. Thus the corresponding eigenvector are

$$\begin{aligned} \text{for } \lambda &= 0,\ x_1 = \begin{bmatrix} 1/\sqrt{5} \\ -2/\sqrt{5} \end{bmatrix} \\ \text{for } \lambda &= 5,\ x_2 = \begin{bmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix} \end{aligned}$$

## Eigen Decomposition of a symmetric matrix

Let $\mathbf{A}$ be a symmetric matrix of order $n$:

- The $n$ eigenvectors of a symmetric matrix are orthogonal, i.e. $x_i'x_j = 0$ for $i \neq j, i, j = 1, ..., n$
- We can collect the $n$ eigenvectors in a matrix $\mathbf{C} = \begin{bmatrix} x_1 & x_2 & ... & x_n \end{bmatrix}$ and the eigenvalues in a diagonal matrix $\mathbf{\Lambda} = diag\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. Since by construction $x_i'x_i = 1$ and $x_i'x_j = 0$, $\mathbf{C}$ has the property that $\mathbf{C}'\mathbf{C} = \mathbf{C}\mathbf{C}' = I_n$, and hence $\mathbf{C}' = \mathbf{C}^{-1}$. A matrix with this property is called an orthogonal matrix or orthonormal matrix.
- The set of equations in (1) can be written as

$$\mathbf{AC} = \mathbf{C\Lambda}.$$

  Therefore solving for $\mathbf{A}$ we obtain

$$\mathbf{A} = \mathbf{C\Lambda C}^{-1} = \mathbf{C\Lambda C}^{'}.$$

  and $|\mathbf{A}| = |\mathbf{C}|\,|\mathbf{\Lambda}|\,\left|\mathbf{C}^{-1}\right| = |\mathbf{\Lambda}| = \lambda_1 \times ... \times \lambda_n$ (as $\left|\mathbf{C}^{-1}\right| = 1/\left|\mathbf{C}\right|$).

## Trace

- Defined for square matrices as:

$$tr\left(\mathbf{A}\right) = \sum_{i=1}^{n} a_{ii}$$

- Provided $\mathbf{AB}$ and $\mathbf{BA}$ are both square matrices (e.g., $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times m$):

$$tr\left(\mathbf{AB}\right) = tr\left(\mathbf{BA}\right)$$

- For any matrix $\mathbf{A}$, its Euclidean norm is defined as

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

- The sum of eigenvalues of a square matrix $\mathbf{A}$ is the trace of the matrix

$$\begin{aligned} tr(\mathbf{A}) &= tr(\mathbf{C\Lambda C}^{-1}) \\ &= tr(\mathbf{\Lambda C}^{-1}\mathbf{C}) \\ &= tr(\mathbf{\Lambda}) \\ &= \sum_{i=1}^{n} \lambda_i. \end{aligned}$$

- If $\mathbf{A}^2 = \mathbf{A}$, then $tr(\mathbf{A}) = \rho(\mathbf{A})$.(A matrix that satisfies $\mathbf{A}^2 = \mathbf{A}$ is called an idempotent matrix).

## Quadratic forms

A quadratic form is defined as $z'\mathbf{A}z$ for an $(n \times n)$ symmetric matrix A

Classification of Quadratic Forms:

- A $(n \times n)$ symmetric matrix $\mathbf{A}$ is positive definite if $z'\mathbf{A}z > 0$ for all non-zero vectors $z$.
- A $(n \times \ n)$ symmetric matrix $\mathbf{A}$ is negative definite if $z\prime\mathbf{A}z < 0$ for all non-zero vectors $z$.
- A $(n \times \ n)$ symmetric matrix $\mathbf{A}$ is positive semidefinite if $z\prime\mathbf{A}z \geq 0$ for all vectors $z$ and $z\prime\mathbf{A}z = 0$ for some $z \neq 0$.
- A $(n \times \ n)$ symmetric matrix $\mathbf{A}$ is negative semidefinite if $z\prime\mathbf{A}z \leq 0$ for all vectors $z$ and $z\prime\mathbf{A}z = 0$ for some $z \neq 0$.
- A $(n \times n)$ symmetric matrix $\mathbf{A}$ is indefinite if $z'\mathbf{A}z > 0$ for some $z$ and $z'\mathbf{A}z < 0$ for other $z$.

**Classification using eigenvalues**

Denote $\lambda_1, ..., \lambda_n$ the ordered (from the larger to the smaller) eigenvalues of the symmetric matrix $A$

- A $(n \times n)$ symmetric matrix $\mathbf{A}$ is positive definite if and only if all its eigenvalues are positive
- A $(n \times n)$ symmetric matrix $\mathbf{A}$ is negative definite if and only if all its eigenvalues are negative.
- A $(n \times \ n)$ symmetric matrix $\mathbf{A}$ is positive semidefinite if

$$\begin{aligned}\lambda_i &> 0, i = 1, ..., \rho(\mathbf{A}) \\ \lambda_i &= 0, i = \rho(\mathbf{A}) + 1, ..., n\end{aligned}$$

- A $(n \times \ n)$ symmetric matrix $\mathbf{A}$ is negative semidefinite if

$$\begin{aligned}\lambda_i &= 0, i = 1, ..., n - \rho(\mathbf{A}) \\ \lambda_i &< 0, i = n - \rho(\mathbf{A}) + 1, ..., n\end{aligned}$$

- A $(n \times n)$ symmetric matrix $\mathbf{A}$ is indefinite if there are at least two eigenvalues with the opposite signs

**Remarks:**

- If $\mathbf{A}$ is positive definite so is $\mathbf{A}^{-1}$.
- Any symmetric positive definite matrix can be written as $\mathbf{A} = \mathbf{Q}'\mathbf{Q}$.

**Proof:** Notice that $\mathbf{A} = \mathbf{C\Lambda C}^{'}$, and since all the eigenvalues are positive we can define $\mathbf{\Lambda}^{1/2} = diag\left\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n}\right\}$. Thus $\mathbf{A} = \mathbf{C\Lambda C}^{'} = \mathbf{C}\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}\mathbf{C}^{'}$. The result follows from denoting $\mathbf{Q} = \mathbf{\Lambda}^{1/2}\mathbf{C}^{'}$.

- If $\mathbf{B}$ is a $m \times n$ matrix, the symmetric $n \times n$ matrix $\mathbf{B}'\mathbf{B}$ is
  - positive semi-definite
  - positive definite if and only if $\rho(\mathbf{B}) = n$
- If $\mathbf{A}, \mathbf{B}$ and $\mathbf{A} - \mathbf{B}$ are all positive definite, then $\mathbf{B}^{-1} - \mathbf{A}^{-1}$ is positive definite.

# Calculus and Matrix Algebra

Let $s(\cdot) : \Re^p \to \Re$ be a real valued function of the $p$-vector $\theta$. Then $\frac{\partial s(\theta)}{\partial \theta}$ is organized as a $p$-vector,

$$\frac{\partial s(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial s(\theta)}{\partial \theta_1} \\ \frac{\partial s(\theta)}{\partial \theta_2} \\ \vdots \\ \frac{\partial s(\theta)}{\partial \theta_p} \end{bmatrix}$$

Following this convention, its transpose $\frac{\partial s(\theta)}{\partial \theta'}$is a $1 \times p$ vector, and $\frac{\partial^2 s(\theta)}{\partial \theta \partial \theta'}$ is a $p \times p$ matrix. Also, the second derivatives matrix or Hessean is

$$\begin{aligned}\mathcal{H}(\theta) &= \frac{\partial^2 s(\theta)}{\partial \theta \partial \theta'} = \frac{\partial}{\partial \theta}\left(\frac{\partial s(\theta)}{\partial \theta'}\right) = \frac{\partial}{\partial \theta'}\left(\frac{\partial s(\theta)}{\partial \theta}\right) \\ &= \begin{bmatrix} \frac{\partial^2 s(\theta)}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 s(\theta)}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 s(\theta)}{\partial \theta_1 \partial \theta_p} \\ \frac{\partial^2 s(\theta)}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 s(\theta)}{\partial \theta_2 \partial \theta_2} & \cdots & \frac{\partial^2 s(\theta)}{\partial \theta_2 \partial \theta_p} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 s(\theta)}{\partial \theta_p \partial \theta_1} & \frac{\partial^2 s(\theta)}{\partial \theta_p \partial \theta_2} & \cdots & \frac{\partial^2 s(\theta)}{\partial \theta_p \partial \theta_p} \end{bmatrix}.\end{aligned}$$

**Exercise 1** *For $a$ and $x$ both $p$-vectors, show that $\frac{\partial a'x}{\partial x} = a$.*

Let $f(\theta)$:$\Re^p \to \Re^n$ be a $n$-vector valued function of the $p$-vector $\theta$. Let $f(\theta)'$ be the $1 \times n$ valued transpose of $f$ . Then $\left(\frac{\partial}{\partial\theta} f(\theta)'\right)' = \frac{\partial}{\partial\theta'} f(\theta)$ [notice that $\frac{\partial}{\partial\theta} f(\theta)'$ is a $p \times n$ matrix, thus $\frac{\partial}{\partial\theta'} f(\theta)$ is a $n \times p$ matrix].

- *Product rule:* Let $f(\theta)$:$\Re^p \to \Re^n$ and $h(\theta)$:$\Re^p \to \Re^n$ be $n$-vector valued functions of the $p$-vector $\theta$. Then
$$\frac{\partial}{\partial\theta'} h(\theta)' f(\theta) = h(\theta)' \left(\frac{\partial}{\partial\theta'} f(\theta)\right) + f(\theta)' \left(\frac{\partial}{\partial\theta'} h(\theta)\right)$$
has dimension $1 \times p$. Applying the transposition rule we get
$$\frac{\partial}{\partial\theta} h(\theta)' f(\theta) = \left(\frac{\partial}{\partial\theta} f(\theta)'\right) h(\theta) + \left(\frac{\partial}{\partial\theta} h(\theta)'\right) f(\theta),$$
which has dimension $p \times 1$.

**Exercise 2** *For $A$ a $p \times p$ matrix and $x$ a $p \times 1$ vector, show that $\frac{\partial x'Ax}{\partial x} = A + A'$.*

- *Chain rule*: Let $f(\cdot)$:$\Re^p \to \Re^n$ a $n$-vector valued function of a $p$-vector argument, and let $g()$:$\Re^r \to \Re^p$ be a $p$-vector valued function of an $r$-vector valued argument $\rho$. Then
$$\frac{\partial}{\partial\rho'} f\left[g\left(\rho\right)\right] = \left.\frac{\partial}{\partial\theta'} f(\theta)\right|_{\theta=g(\rho)} \frac{\partial}{\partial\rho'} g(\rho)$$
has dimension $n \times r$.

**Exercise 3** *For $x$ and $\beta$ both $p \times 1$ vectors, show that $\frac{\partial \exp(x'\beta)}{\partial\beta} = \exp(x'\beta)x$.*

1. Consider $\mathbf{a}'\mathbf{x} = \sum_{i=1}^{n} a_i x_i$; then $\frac{\partial\left(\mathbf{a}'\mathbf{x}\right)}{\partial\mathbf{x}} = \mathbf{a}$.
2. Consider $\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j$; then $\frac{\partial\left(\mathbf{x}'\mathbf{A}\mathbf{x}\right)}{\partial\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{A}'\mathbf{x}$.
3. $\frac{\partial\left(\mathbf{A}\mathbf{x}\right)}{\partial\mathbf{x}'} = \mathbf{A}$.

- **Taylor Theorem:** Let $s(\cdot) : \Re^p \to \Re$ be a real valued function of the $p$-vector $\theta$, if $s(\theta)$ is once continuously differentiable on $\Theta$ and $\theta_0 \in \Theta$
$$s(\theta) = s(\theta_0) + \frac{\partial s(\bar\theta)}{\partial\theta'}(\theta - \theta_0)$$
for some $\bar\theta = \lambda\theta_0 + (1-\lambda)\theta$ where $0 \le \lambda \le 1$.

- If s is twice continuously differentiable
$$s(\theta) = s(\theta_0) + \frac{\partial s(\bar\theta)}{\partial\theta'}(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)'\mathcal{H}\left(\bar\theta\right)(\theta - \theta_0)$$

- Let $f(\theta)$:$\Re^p \to \Re^n$ be a $n$-vector valued function of the $p$-vector $\theta$ if $s(\theta)$ is once continuously differentiable on $\Theta$
$$f(\theta) = f(\theta_0) + \frac{\partial f(\bar\theta)}{\partial\theta'}(\theta - \theta_0)$$

## Optimization

For maximizing or minimizing a function os several variables, say $s(\theta)$ the first order conditions are

$$\frac{\partial s(\theta)}{\partial \theta} = 0.$$

Let $\theta^*$ be the solution of this system of equations.

The second order conditions for an optimum are:

- $\mathcal{H}(\theta)$ is positive definite for a minimum
- $\mathcal{H}(\theta)$ is negative definite for a maximum

## Constrained Optimization

For maximizing or minimizing a function os several variables, say $s(\theta)$ subjects to the constraints $c(\theta) = 0$, where $c(\theta)$ is a vector function. One can use the Lagrangean approach. First we have to construct the Lagrangean:

$$L(\theta, \lambda) = s(\theta) + c(\theta)'\lambda.$$

the second step is to find the stationary points, i.e. the points at which the derivatives of the Lagrangean are zero

$$\begin{aligned} \frac{\partial L(\theta, \lambda)}{\partial \theta} &= \frac{\partial s(\theta)}{\partial \theta} + \left[\frac{\partial c(\theta)'}{\partial \theta}\right] \lambda = 0 \\ \frac{\partial L(\theta, \lambda)}{\partial \lambda} &= c(\theta) = 0 \end{aligned}$$

**Remarks:**

1. The constrained solution must be inferior to the unconstrained solution because we are restricting the parameter set.
2. If the Lagrangean multipliers are zero, the constrained solution is equal to the unconstrained solution.
3. The above second order conditions for unconstrained optimization do not apply here. The second order conditions for the constrained case will not be stated but can be seen in page 136 of Magnus and Neudecker (1999), *Matrix differential calculus with applications in statistics and econometrics*, New York: John Wiley & Sons.