Crash Course in Matrix Algebra

Sven Otto

September 7, 2023

Table of contents

Welcome

Due to the multivariate character of many econometric topics, matrix algebra is a commonly used tool in modern econometrics. It provides a powerful and efficient framework for representing and manipulating systems of linear equations. This short lecture note series provides a brief introduction to the most relevant matrix algebra concepts for econometricians and their implementation in R.

To learn R or refresh your skills, please check out my tutorial **Getting Started With R**.

Accompanying R scripts

All R codes of the different sections can be found here:

- matrix-sec1.R
- matrix-sec2.R
- matrix-sec3.R
- matrix-sec4.R

Comments

Comments are welcome. Please let me know if you find any typing errors, too. Please report it on GitHub or contact me via e-mail: sven.otto@uni-koeln.de

1 Basic definitions

Let's start with some basic definitions and specific examples.

1.1 Scalar, vector, and matrix

A **scalar** a is a single real number. We write $a \in \mathbb{R}$.

A **vector** $\boldsymbol{a}$ of length k is a $k \times 1$ list of real numbers

$$\boldsymbol{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix}.$$

By default, when we refer to a vector, we always mean a column vector. We write $\boldsymbol{a} \in \mathbb{R}^k$. The value a_i is called i-th entry or i-th component of $\boldsymbol{a}$. A scalar is a vector of length 1. A row vector of length k is written as $\boldsymbol{b} = (b_1, \dots, b_k)$.

A **matrix** $\boldsymbol{A}$ of order $k \times m$ is a rectangular array of real numbers

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{km} \end{pmatrix}$$

with k rows and m columns. We write $\boldsymbol{A} \in \mathbb{R}^{k \times m}$. The value a_{ij} is called (i,j)-th entry or (i,j)-th component of $\boldsymbol{A}$. We also use the notation $(\boldsymbol{A})_{i,j}$ to denote the (i,j)-th entry. A vector of length k is a $k \times 1$ matrix. A row vector of length k is a $1 \times k$ matrix. A scalar is a matrix of order 1×1.

We may describe a matrix $\boldsymbol{A}$ by its column or row vectors as

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_m \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_k \end{pmatrix},$$

where

$$\boldsymbol{a}_i = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{ki} \end{pmatrix}$$

is the i-th column of $\boldsymbol{A}$ and $\boldsymbol{\alpha}_i = (a_{i1}, \dots, a_{im})$ is the i-th row.

1.2 Some specific matrices

A matrix is called **square matrix** if the numbers of rows and columns coincide (i.e., $k = m$).

$$\boldsymbol{B} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

is a square matrix. A square matrix is called **diagonal matrix** if all off-diagonal elements are zero.

$$\boldsymbol{C} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}$$

is a diagonal matrix. We also write $\boldsymbol{C} = \mathrm{diag}(1, 4)$. A square matrix is called **upper triangular** if all elements below the main diagonal are zero, and **lower triangular** if all elements above the main diagonal are zero. Examples of an upper triangular matrix $\boldsymbol{D}$ and a lower triangular matrix $\boldsymbol{E}$ are

$$\boldsymbol{D} = \begin{pmatrix} 1 & 2 \\ 0 & 4 \end{pmatrix}, \quad \boldsymbol{E} = \begin{pmatrix} 1 & 0 \\ 3 & 4 \end{pmatrix}.$$

The $k \times k$ diagonal matrix

$$\boldsymbol{I}_k = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} = \mathrm{diag}(1, \dots, 1)$$

is called **identity matrix** of order k. The $k \times m$ matrix

$$\mathbf{0}_{k \times m} = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}$$

is called **zero matrix**. The **zero vector** of length k is

$$\mathbf{0}_k = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

If the order becomes clear from the context, we omit the indices and write $\boldsymbol{I}$ for the identity matrix and $\mathbf{0}$ for the zero matrix or zero vector.

1.3 Transposition

The **transpose** $\boldsymbol{A}'$ of the matrix $\boldsymbol{A}$ is obtained by flipping rows and columns on the main diagonal:

$$\boldsymbol{A}' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{k1} \\ a_{12} & a_{22} & \cdots & a_{k2} \\ \vdots & \vdots & & \vdots \\ a_{1m} & a_{2m} & \cdots & a_{km} \end{pmatrix}.$$

If $\boldsymbol{A}$ is a matrix of order $k \times m$, then $\boldsymbol{A}'$ is a matrix of order $m \times k$. *Example:*

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{A}' = \begin{pmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \end{pmatrix}$$

The definition implies that transposing twice produces the original matrix:

$$(\boldsymbol{A}')' = \boldsymbol{A}.$$

The transpose of a (column) vector is a row vector:

$$\boldsymbol{a}' = (a_1, \ldots, a_k)$$

A **symmetric matrix** is a square matrix $\boldsymbol{A}$ with $\boldsymbol{A}' = \boldsymbol{A}$. An example of a symmetric matrix is

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.$$

1.4 Matrices in R

Let's define some matrices in R:

```r
A = matrix(c(1,4,7,2,5,8), nrow = 3, ncol = 2)
A
```

```
     [,1] [,2]
[1,]    1    2
[2,]    4    5
[3,]    7    8
```

```r
t(A) #transpose of A
```

```
     [,1] [,2] [,3]
[1,]    1    4    7
[2,]    2    5    8
```

```r
A[3,2] #the (3,2)-entry of A
```

```
[1] 8
```

```r
B = matrix(c(1,2,2,4), nrow = 2, ncol = 2) # another matrix
all(B == t(B)) #check whether B is symmetric
```

```
[1] TRUE
```

```r
diag(c(1,4)) #diagonal matrix
```

```
     [,1] [,2]
[1,]    1    0
[2,]    0    4
```

```r
diag(1, nrow = 3) #identity matrix
```

```
     [,1] [,2] [,3]
[1,]    1    0    0
[2,]    0    1    0
[3,]    0    0    1
```

```r
matrix(0, nrow=2, ncol=5) #matrix of zeros
```

```
     [,1] [,2] [,3] [,4] [,5]
[1,]    0    0    0    0    0
[2,]    0    0    0    0    0
```

```r
dim(A) #number of rows and columns
```

```
[1] 3 2
```

2 Sums and Products

2.1 Matrix summation

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ both be matrices of order $k \times m$. Their sum is defined componentwise:

$$\boldsymbol{A}+\boldsymbol{B} = \begin{pmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1m}+b_{1m} \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2m}+b_{2m} \\ \vdots & \vdots & & \vdots \\ a_{k1}+b_{k1} & a_{k2}+b_{k2} & \cdots & a_{km}+b_{km} \end{pmatrix}.$$

Only two matrices of the same order can be added. *Example:*

$$\boldsymbol{A} = \begin{pmatrix} 2 & 0 \\ 1 & 5 \\ 3 & 2 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} -1 & 1 \\ 7 & 1 \\ -5 & 2 \end{pmatrix}, \quad \boldsymbol{A}+\boldsymbol{B} = \begin{pmatrix} 1 & 1 \\ 8 & 6 \\ -2 & 4 \end{pmatrix}.$$

The matrix summation satisfies the following rules:

$$\begin{array}{rcr} \text{(i)} & \boldsymbol{A}+\boldsymbol{B} = \boldsymbol{B}+\boldsymbol{A} & \text{(commutativity)} \\ \text{(ii)} & (\boldsymbol{A}+\boldsymbol{B})+\boldsymbol{C} = \boldsymbol{A}+(\boldsymbol{B}+\boldsymbol{C}) & \text{(associativity)} \\ \text{(iii)} & \boldsymbol{A}+\boldsymbol{0} = \boldsymbol{A} & \text{(identity element)} \\ \text{(iv)} & (\boldsymbol{A}+\boldsymbol{B})' = \boldsymbol{A}'+\boldsymbol{B}' & \text{(transposition)} \end{array}$$

2.2 Scalar-matrix multiplication

The product of a $k \times m$ matrix $\boldsymbol{A}$ with a scalar $\lambda \in \mathbb{R}$ is defined componentwise:

$$\lambda\boldsymbol{A} = \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2n} \\ \vdots & \vdots & & \vdots \\ \lambda a_{m1} & \lambda a_{m2} & \cdots & \lambda a_{mn} \end{pmatrix}.$$

Example:

$$\lambda = 2, \quad \boldsymbol{A} = \begin{pmatrix} 2 & 0 \\ 1 & 5 \\ 3 & 2 \end{pmatrix}, \quad \lambda\boldsymbol{A} = \begin{pmatrix} 4 & 0 \\ 2 & 10 \\ 6 & 4 \end{pmatrix}.$$

Scalar-matrix multiplication satisfies the distributivity law:

$$\begin{array}{rl}\text{(i)} & \lambda(\boldsymbol{A}+\boldsymbol{B}) = \lambda\boldsymbol{A}+\lambda\boldsymbol{B} \\ \text{(ii)} & (\lambda+\mu)\boldsymbol{A} = \lambda\boldsymbol{A}+\mu\boldsymbol{A}\end{array}$$

2.3 Element-by-element operations in R

Basic arithmetic operations work on an element-by-element basis in R:

```
A = matrix(c(2,1,3,0,5,2), ncol=2)
B = matrix(c(-1,7,-5,1,1,2), ncol=2)
A+B #matrix summation
```

```
     [,1] [,2]
[1,]    1    1
[2,]    8    6
[3,]   -2    4
```

```
A-B #matrix subtraction
```

```
     [,1] [,2]
[1,]    3   -1
[2,]   -6    4
[3,]    8    0
```

```
2*A #scalar-matrix product
```

```
     [,1] [,2]
[1,]    4    0
[2,]    2   10
[3,]    6    4
```

```
A/2 #division of entries by 2
```

```
     [,1] [,2]
[1,]  1.0  0.0
[2,]  0.5  2.5
[3,]  1.5  1.0
```

```
A*B #element-wise multiplication
```

```
     [,1] [,2]
[1,]   -2    0
[2,]    7    5
[3,]  -15    4
```

2.4 Vector-vector multiplication

2.4.1 Inner product

The **inner product** (also known as dot product) of two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^k$ is

$$\boldsymbol{a}'\boldsymbol{b} = a_1 b_1 + a_2 b_2 + ... + a_k b_k = \sum_{i=1}^{k} a_i b_i \in \mathbb{R}.$$

Example:

$$\boldsymbol{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \boldsymbol{b} = \begin{pmatrix} -2 \\ 0 \\ 2 \end{pmatrix}, \quad \boldsymbol{a}'\boldsymbol{b} = 1 \cdot (-2) + 2 \cdot 0 + 3 \cdot 2 = 4.$$

The inner product is commutative:

$$\boldsymbol{a}'\boldsymbol{b} = \boldsymbol{b}'\boldsymbol{a}.$$

Two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are called **orthogonal** if $\boldsymbol{a}'\boldsymbol{b} = 0$. The vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are called **orthonormal** if, in addition to $\boldsymbol{a}'\boldsymbol{b}$, we have $\boldsymbol{a}'\boldsymbol{a} = 1$ and $\boldsymbol{b}'\boldsymbol{b} = 1$.

2.4.2 Outer product

The outer product (also known as dyadic product) of two vectors $\boldsymbol{x} \in \mathbb{R}^k$ and $\boldsymbol{y} \in \mathbb{R}^m$ is

$$\boldsymbol{x}\boldsymbol{y}' = \begin{pmatrix} x_1 y_1 & x_1 y_2 & ... & x_1 y_m \\ x_2 y_1 & x_2 y_2 & ... & x_2 y_m \\ \vdots & \vdots & & \vdots \\ x_k y_1 & x_k y_2 & ... & x_k y_m \end{pmatrix} \in \mathbb{R}^{k \times m}.$$

Example:

$$\boldsymbol{x} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \boldsymbol{y} = \begin{pmatrix} -2 \\ 0 \\ 2 \end{pmatrix}, \quad \boldsymbol{x}\boldsymbol{y}' = \begin{pmatrix} -2 & 0 & 2 \\ -4 & 0 & 4 \end{pmatrix}.$$

2.4.3 Vector multiplication in `R`

For vector multiplication in `R`, we use the operator `%*%` (recall that `*` is already reserved for element-wise multiplication). Let's implement some multiplications.

```
y = c(2,7,4,1) #y is treated as a column vector
t(y) %*% y #the inner product of y with itself
```

```
     [,1]
[1,]   70
```

```
y %*% t(y) #the outer product of y with itself
```

```
     [,1] [,2] [,3] [,4]
[1,]    4   14    8    2
[2,]   14   49   28    7
[3,]    8   28   16    4
[4,]    2    7    4    1
```

```
c(1,2) %*% t(c(-2,0,2)) #the example from above
```

```
     [,1] [,2] [,3]
[1,]   -2    0    2
[2,]   -4    0    4
```

2.5 Matrix-matrix multiplication

The **matrix product** of a $k \times m$ matrix $\boldsymbol{A}$ and a $m \times n$ matrix $\boldsymbol{B}$ is the $k \times n$ matrix $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B}$ with the components

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{im}b_{mj} = \sum_{l=1}^{m} a_{il}b_{lj} = \boldsymbol{a}_i'\boldsymbol{b}_j,$$

where $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{im})'$ is the i-th row of $\boldsymbol{A}$ written as a column vector, and $\boldsymbol{b}_j = (b_{1j}, \ldots, b_{mj})'$ is the j-th column of $\boldsymbol{B}$. The full matrix product can be written as

$$\boldsymbol{A}\boldsymbol{B} = \begin{pmatrix} \boldsymbol{a}_1' \\ \vdots \\ \boldsymbol{a}_k' \end{pmatrix} \begin{pmatrix} \boldsymbol{b}_1 & \ldots & \boldsymbol{b}_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{a}_1'\boldsymbol{b}_1 & \ldots & \boldsymbol{a}_1'\boldsymbol{b}_n \\ \vdots & & \vdots \\ \boldsymbol{a}_k'\boldsymbol{b}_1 & \ldots & \boldsymbol{a}_k'\boldsymbol{b}_n \end{pmatrix}.$$

The matrix product is only defined if the number of columns of the first matrix equals the number of rows of the second matrix. Therefore, we say that the $k \times m$ matrix $\boldsymbol{A}$ and the $m \times n$ matrix $\boldsymbol{B}$ are **conformable for matrix multiplication**.

Example: Let

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} -1 & 2 \\ -3 & 0 \end{pmatrix}.$$

Their matrix product is

$$\begin{aligned} \boldsymbol{AB} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} -1 & 2 \\ -3 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 \cdot (-1) + 0 \cdot (-3) & 1 \cdot 2 + 0 \cdot 0 \\ 0 \cdot (-1) + 1 \cdot (-3) & 0 \cdot 2 + 1 \cdot 0 \\ 2 \cdot (-1) + 1 \cdot (-3) & 2 \cdot 2 + 1 \cdot 0 \end{pmatrix} = \begin{pmatrix} -1 & 2 \\ -3 & 0 \\ -5 & 4 \end{pmatrix}. \end{aligned}$$

The `%*%` operator is used in `R` for matrix-matrix multiplications:

```r
A = matrix(c(1,0,2,0,1,1), ncol=2)
B = matrix(c(-1,-3,2,0), ncol=2)
A %*% B
```

```
     [,1] [,2]
[1,]   -1    2
[2,]   -3    0
[3,]   -5    4
```

Matrix multiplication is **not commutative**. In general, we have $\boldsymbol{AB} \neq \boldsymbol{BA}$. *Example:*

$$\begin{aligned} \boldsymbol{AB} &= \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 5 \\ 7 & 11 \end{pmatrix}, \\ \boldsymbol{BA} &= \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 4 & 6 \\ 7 & 10 \end{pmatrix}. \end{aligned}$$

Even if neither of the two matrices contains zeros, the matrix product can give the zero matrix:

$$\boldsymbol{AB} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 2 & -4 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = \mathbf{0}.$$

The following rules of calculation apply (provided the matrices are conformable):

$$
\begin{array}{rrcll}
\text{(i)} & \boldsymbol{A}(\boldsymbol{BC}) & = & (\boldsymbol{AB})\boldsymbol{C} & \text{(associativity)} \\
\text{(ii)} & \boldsymbol{A}(\boldsymbol{B}+\boldsymbol{D}) & = & \boldsymbol{AB}+\boldsymbol{AD} & \text{(distributivity)} \\
\text{(iii)} & (\boldsymbol{B}+\boldsymbol{D})\boldsymbol{C} & = & \boldsymbol{BC}+\boldsymbol{DC} & \text{(distributivity)} \\
\text{(iv)} & \boldsymbol{A}(\lambda\boldsymbol{B}) & = & \lambda(\boldsymbol{AB}) & \text{(scalar commutativity)} \\
\text{(v)} & \boldsymbol{A}\boldsymbol{I}_n & = & \boldsymbol{A}, & \text{(identity element)} \\
\text{(vi)} & \boldsymbol{I}_m\boldsymbol{A} & = & \boldsymbol{A} & \text{(identity element)} \\
\text{(vii)} & (\boldsymbol{AB})' & = & \boldsymbol{B}'\boldsymbol{A}' & \text{(product transposition)} \\
\text{(viii)} & (\boldsymbol{ABC})' & = & \boldsymbol{C}'\boldsymbol{B}'\boldsymbol{A}' & \text{(product transposition)}
\end{array}
$$

3 Rank and inverse

3.1 Linear combination

Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ be vectors of the same order, and let $\lambda_1, \dots, \lambda_n$ be scalars. The vector

$$\lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2 + \dots + \lambda_n \boldsymbol{x}_n$$

is called **linear combination** of $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$. A linear combination can also be written as a matrix-vector product. Let $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{x}_1 & \dots & \boldsymbol{x}_n \end{pmatrix}$ be the matrix with columns $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$, and let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)'$. Then,

$$\lambda_1 \boldsymbol{x}_1 + \lambda_2 \boldsymbol{x}_2 + \dots + \lambda_n \boldsymbol{x}_n = \boldsymbol{X}\boldsymbol{\lambda}.$$

The vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are called **linearly dependent** if at least one can be written as a linear combination of the others. That is, there exists a nonzero vector $\boldsymbol{\lambda}$ with

$$\boldsymbol{X}\boldsymbol{\lambda} = \lambda_1 \boldsymbol{x}_1 + \dots + \lambda_n \boldsymbol{x}_n = \boldsymbol{0}.$$

The vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ are called **linearly independent** if

$$\boldsymbol{X}\boldsymbol{\lambda} = \lambda_1 \boldsymbol{x}_1 + \dots + \lambda_n \boldsymbol{x}_n \neq \boldsymbol{0}$$

for all nonzero vectors $\boldsymbol{\lambda}$.

To check whether the vectors are linearly independent, we can solve the system of equations $\boldsymbol{X}\boldsymbol{\lambda} = \boldsymbol{0}$ by Gaussian elimination. If $\boldsymbol{\lambda} = \boldsymbol{0}$ is the only solution, then the columns of $\boldsymbol{X}$ are linearly independent. If there is a solution $\boldsymbol{\lambda}$ with $\boldsymbol{\lambda} \neq \boldsymbol{0}$, then the columns of $\boldsymbol{X}$ are linearly dependent.

3.2 Column rank

The **rank** of a $k \times m$ matrix $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{a}_1 & \dots & \boldsymbol{a}_m \end{pmatrix}$, written as $\mathrm{rank}(\boldsymbol{A})$, is the number of linearly independent columns $\boldsymbol{a}_i$. We say that $\boldsymbol{A}$ has **full column rank** if $\mathrm{rank}(\boldsymbol{X}) = m$.

The identity matrix $\boldsymbol{I}_k$ has full column rank (i.e., $\mathrm{rank}(\boldsymbol{I}_n) = k$). As another example, consider

$$\boldsymbol{X} = \begin{pmatrix} 2 & 1 & 4 \\ 0 & 1 & 2 \end{pmatrix},$$

which has linearly dependent columns since the third column is a linear combination of the first two columns:

$$\begin{pmatrix}4\\2\end{pmatrix} = 1\begin{pmatrix}2\\0\end{pmatrix} + 2\begin{pmatrix}1\\1\end{pmatrix}.$$

The first two columns are linearly independent since $\lambda_1 = 0$ and $\lambda_2 = 0$ are the only solutions to the equation

$$\lambda_1\begin{pmatrix}2\\0\end{pmatrix} + \lambda_2\begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}.$$

Therefore, we have $\mathrm{rank}(\boldsymbol{X}) = 2$, i.e., $\boldsymbol{X}$ does not have a full column rank.

Some useful properties are

i) $\mathrm{rank}(\boldsymbol{A}) \leq \min(k, m)$
ii) $\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A}')$
iii) $\mathrm{rank}(\boldsymbol{AB}) = \min(\mathrm{rank}(\boldsymbol{A}), \mathrm{rank}(\boldsymbol{B}))$
iv) $\mathrm{rank}(\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{A}'\boldsymbol{A}) = \mathrm{rank}(\boldsymbol{AA}')$.

We can use the `qr()` function to extract the rank in `R`. Let's compute the rank of the matrices

$$\boldsymbol{A} = \begin{pmatrix}1 & 2 & 3\\3 & 9 & 1\\0 & 11 & 5\end{pmatrix},$$

$\boldsymbol{B} = \boldsymbol{I}_3$, and $\boldsymbol{X}$ from the example above:

```r
A = matrix(c(1,3,0,2,9,11,3,1,5), nrow=3)
qr(A)$rank
```

```
[1] 3
```

```r
B = matrix(c(1,1,1,1,1,1,1,1,1), nrow=3)
qr(B)$rank
```

```
[1] 1
```

```r
X = matrix(c(2,0,1,1,4,2), ncol=3)
qr(X)$rank
```

```
[1] 2
```

3.3 Nonsingular matrix

A square $k \times k$ matrix $\boldsymbol{A}$ is called **nonsingular** if it has full rank, i.e., $\operatorname{rank}(\boldsymbol{A}) = k$. Conversely, $\boldsymbol{A}$ is called **singular** if it does not have full rank, i.e., $\operatorname{rank}(\boldsymbol{A}) < k$.

3.4 Determinant

Consider a square $k \times k$ matrix $\boldsymbol{A}$. The determinant $\det(\boldsymbol{A})$ is a measure of the volume of the geometric object formed by the columns of $\boldsymbol{A}$ (a parallelogram for $k = 2$, a parallelepiped for $k = 3$, a hyper-parallelepiped for $k > 3$). For 2×2 matrices, the determinant is easy to calculate:

$$\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \det(\boldsymbol{A}) = ad - bc.$$

If $\boldsymbol{A}$ is triangular (upper or lower), the determinant is the product of the diagonal entries, i.e., $\det(\boldsymbol{A}) = \prod_{i=1}^{k} a_{ii}$. Hence, Gaussian elimination can be used to compute the determinant by transforming the matrix to a triangular one. The exact definition of the determinant is technical and of little importance to us. A useful relation is the following:

$$\begin{aligned} \det(\boldsymbol{A}) = 0 \quad &\Leftrightarrow \quad \boldsymbol{A} \text{ is singular} \\ \det(\boldsymbol{A}) \neq 0 \quad &\Leftrightarrow \quad \boldsymbol{A} \text{ is nonsingular.} \end{aligned}$$

In `R`, we have the `det()` function to compute the determinant:

```
det(A)
```

```
[1] 103
```

```
det(B)
```

```
[1] 0
```

Since $\det(\boldsymbol{A}) \neq 0$ and $\det(\boldsymbol{B}) = 0$, we conclude that $\boldsymbol{A}$ is nonsingular and $\boldsymbol{B}$ is singular.

3.5 Inverse matrix

The **inverse** $\boldsymbol{A}^{-1}$ of a square $k \times k$ matrix $\boldsymbol{A}$ is defined by the property

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_k.$$

When multiplied from the left or the right, the inverse matrix produces the identity matrix. The inverse exists if and only if $\boldsymbol{A}$ is nonsingular, i.e., $\det(\boldsymbol{A}) \neq 0$. Therefore, a nonsingular matrix is also called **invertible matrix**. Note that only square matrices can be inverted.

For 2×2 matrices, there exists a simple formula:

$$\boldsymbol{A}^{-1} = \frac{1}{\det(\boldsymbol{A})} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

where $\det(\boldsymbol{A}) = ad - bc$. We swap the main diagonal elements, reverse the sign of the off-diagonal elements, and divide all entries by the determinant. *Example:*

$$\boldsymbol{A} = \begin{pmatrix} 5 & 6 \\ 1 & 2 \end{pmatrix}$$

We have $\det(\boldsymbol{A}) = ad - bc = 5 \cdot 2 - 6 \cdot 1 = 4$, and

$$\boldsymbol{A}^{-1} = \frac{1}{4} \cdot \begin{pmatrix} 2 & -6 \\ -1 & 5 \end{pmatrix}.$$

Indeed, $\boldsymbol{A}^{-1}$ is the inverse of $\boldsymbol{A}$ since

$$\boldsymbol{A}\boldsymbol{A}^{-1} = \begin{pmatrix} 5 & 6 \\ 1 & 2 \end{pmatrix} \cdot \frac{1}{4} \cdot \begin{pmatrix} 2 & -6 \\ -1 & 5 \end{pmatrix} = \frac{1}{4} \cdot \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \boldsymbol{I}_2.$$

One way to calculate the inverse of higher order square matrices is to solve equation $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$ with Gaussian elimination. `R` can compute the inverse matrix quickly using the function `solve()`:

```
solve(A) #inverse if A
```

```
           [,1]        [,2]        [,3]
[1,]  0.3300971  0.22330097 -0.24271845
[2,] -0.1456311  0.04854369  0.07766990
[3,]  0.3203883 -0.10679612  0.02912621
```

We have the following relationship between invertibility, rank, and determinant of a square matrix $\boldsymbol{A}$:

$$
\begin{aligned}
& \boldsymbol{A} \text{ is nonsingular} \\
\Leftrightarrow\ & \text{all columns of } \boldsymbol{A} \text{ are linearly independent} \\
\Leftrightarrow\ & \boldsymbol{A} \text{ has full column rank} \\
\Leftrightarrow\ & \text{the determinant is nonzero } (\det(\boldsymbol{A}) \neq 0).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& \boldsymbol{A} \text{ is singular} \\
\Leftrightarrow\ & \boldsymbol{A} \text{ has linearly dependent columns} \\
\Leftrightarrow\ & \boldsymbol{A} \text{ does not have full rank} \\
\Leftrightarrow\ & \text{the determinant is zero } (\det(\boldsymbol{A}) = 0).
\end{aligned}
$$

Below you will find some important properties for nonsingular matrices:

i) $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$
ii) $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$
iii) $(\lambda\boldsymbol{A})^{-1} = \frac{1}{\lambda}\boldsymbol{A}^{-1}$ for any $\lambda \neq 0$
iv) $\det(\boldsymbol{A}^{-1}) = \frac{1}{\det(\boldsymbol{A})}$
v) $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
vi) $(\boldsymbol{ABC})^{-1} = \boldsymbol{C}^{-1}\boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
vii) If $\boldsymbol{A}$ is symmetric, then $\boldsymbol{A}^{-1}$ is symmetric.

4 Advanced concepts

4.1 Trace

The **trace** of a $k \times k$ square matrix $\boldsymbol{A}$ is the sum of the diagonal entries:

$$\operatorname{tr}(\boldsymbol{A}) = \sum_{i=1}^{n} a_{ii}$$

Example:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 9 & 1 \\ 0 & 11 & 5 \end{pmatrix} \quad \Rightarrow \quad \operatorname{tr}(\boldsymbol{A}) = 1 + 9 + 5 = 15$$

In `R`we have

```
A = matrix(c(1,3,0,2,9,11,3,1,5), nrow=3)
sum(diag(A))  #trace = sum of diagonal entries
```

```
[1] 15
```

The following properties hold for square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and scalars λ:

i) $\operatorname{tr}(\lambda \boldsymbol{A}) = \lambda \operatorname{tr}(\boldsymbol{A})$
ii) $\operatorname{tr}(\boldsymbol{A} + \boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{B})$
iii) $\operatorname{tr}(\boldsymbol{A}') = \operatorname{tr}(\boldsymbol{A})$
iv) $\operatorname{tr}(\boldsymbol{I}_k) = k$

For $\boldsymbol{A} \in \mathbb{R}^{k \times m}$ and $\boldsymbol{B} \in \mathbb{R}^{m \times k}$ we have

$$\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{B}\boldsymbol{A}).$$

4.2 Idempotent matrx

The square matrix $\boldsymbol{A}$ is called **idempotent** if $\boldsymbol{A}\boldsymbol{A} = \boldsymbol{A}$. The identity matrix is idempotent: $\boldsymbol{I}_n \boldsymbol{I}_n = \boldsymbol{I}_n$. Another example is the matrix

$$\boldsymbol{A} = \begin{pmatrix} 4 & -1 \\ 12 & -3 \end{pmatrix}.$$

We have

$$\begin{aligned} \boldsymbol{A}\boldsymbol{A} &= \begin{pmatrix} 4 & -1 \\ 12 & -3 \end{pmatrix} \begin{pmatrix} 4 & -1 \\ 12 & -3 \end{pmatrix} \\ &= \begin{pmatrix} 16-12 & -4+3 \\ 48-36 & -12+9 \end{pmatrix} \\ &= \begin{pmatrix} 4 & -1 \\ 12 & -3 \end{pmatrix} = \boldsymbol{A}. \end{aligned}$$

4.3 Eigendecomposition

4.3.1 Eigenvalues

An **eigenvalue** λ of a $k \times k$ square matrix is a solution to the equation

$$\det(\lambda \boldsymbol{I}_k - \boldsymbol{A}) = 0.$$

The function $f(\lambda) = \det(\lambda \boldsymbol{I}_k - \boldsymbol{A})$ has exactly k roots so that $\det(\lambda \boldsymbol{I}_k - \boldsymbol{A}) = 0$ has exactly k solutions. The solutions $\lambda_1, \ldots, \lambda_k$ are the k eigenvalues of $\boldsymbol{A}$.

Most applications of eigenvalues in econometrics concern symmetric matrices. In this case, all eigenvalues are real-valued. In the case of non-symmetric matrices, some eigenvalues may be complex-valued.

Useful properties of the eigenvalues of a symmetric $k \times k$ matrix are:

i) $\det(\boldsymbol{A}) = \lambda_1 \cdot \ldots \cdot \lambda_k$
ii) $\mathrm{tr}(\boldsymbol{A}) = \lambda_1 + \ldots + \lambda_k$
iii) $\boldsymbol{A}$ is nonsingular if and only if all eigenvalues are nonzero
iv) $\boldsymbol{A}\boldsymbol{B}$ and $\boldsymbol{B}\boldsymbol{A}$ have the same eigenvalues.

4.3.2 Eigenvectors

If λ_i is an eigenvalue of $\boldsymbol{A}$, then $\lambda_i \boldsymbol{I}_k - \boldsymbol{A}$ is singular, which implies that there exists a linear combination vector $\boldsymbol{v}_i$ with $(\lambda_i \boldsymbol{I}_k - \boldsymbol{A})\boldsymbol{v}_i = \boldsymbol{0}$. Equivalently,

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i,$$

which can be solved by Gaussian elimination. It is convenient to normalize any solution such that $\boldsymbol{v}_i'\boldsymbol{v}_i = 1$. The solutions $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ are called eigenvectors of $\boldsymbol{A}$ to corresponding eigenvalues $\lambda_1, \ldots, \lambda_k$.

4.3.3 Spectral decomposition

If $\boldsymbol{A}$ is symmetric, then $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ are pairwise orthogonal (i.e., $\boldsymbol{v}_i'\boldsymbol{v}_j = 0$ for $i \neq j$). Let $\boldsymbol{V} = \begin{pmatrix} \boldsymbol{v}_1 & \ldots & \boldsymbol{v}_k \end{pmatrix}$ be the $k \times k$ matrix of eigenvectors and let $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_k)$ be the $k \times k$ diagonal matrix with the eigenvalues on the main diagonal. Then, we can write

$$\boldsymbol{A} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}',$$

which is called the **spectral decomposition** of $\boldsymbol{A}$. The matrix of eigenvalues can be written as $\boldsymbol{\Lambda} = \boldsymbol{V}'\boldsymbol{A}\boldsymbol{V}$.

4.3.4 Eigendecomposition in `R`

The function `eigen()` computes the eigenvalues and corresponding eigenvectors.

```
B=t(A)%*%A
B #A'A is symmetric
```

```
     [,1] [,2] [,3]
[1,]   10   29    6
[2,]   29  206   70
[3,]    6   70   35
```

```
eigen(B) #eigenvalues and eigenvector matrix
```

```
eigen() decomposition
$values
[1] 234.827160  12.582227   3.590613

$vectors
           [,1]       [,2]       [,3]
[1,] -0.1293953 -0.5312592  0.8372697
[2,] -0.9346164 -0.2167553 -0.2819739
[3,] -0.3312839  0.8190121  0.4684764
```

4.4 Definite matrix

The $k \times k$ square matrix $\boldsymbol{A}$ is called **positive definite** if

$$\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} > 0$$

holds for all nonzero vectors $\boldsymbol{c} \in \mathbb{R}^k$. If

$$\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} \geq 0$$

for all vectors $\boldsymbol{c} \in \mathbb{R}^k$, the matrix is called **positive semi-definite**. Analogously, $\boldsymbol{A}$ is called **negative definite** if $\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} < 0$ and **negative semi-definite** if $\boldsymbol{c}'\boldsymbol{A}\boldsymbol{c} \leq 0$ for all nonzero vectors $\boldsymbol{c} \in \mathbb{R}^k$. A matrix that is neither positive semi-definite nor negative semi-definite is called **indefinite**

The definiteness property of a symmetric matrix $\boldsymbol{A}$ can be determined using its eigenvalues:

i) $\boldsymbol{A}$ is positive definite $\Leftrightarrow$ all eigenvalues of $\boldsymbol{A}$ are strictly positive
ii) $\boldsymbol{A}$ is negative definite $\Leftrightarrow$ all eigenvalues of $\boldsymbol{A}$ are strictly negative
iii) $\boldsymbol{A}$ is positive semi-definite $\Leftrightarrow$ all eigenvalues of $\boldsymbol{A}$ are non-negative
iv) $\boldsymbol{A}$ is negative semi-definite $\Leftrightarrow$ all eigenvalues of $\boldsymbol{A}$ are non-positive

```r
eigen(B)$values #B is positive definite (all eigenvalues positive)
```

```
[1] 234.827160  12.582227   3.590613
```

The matrix analog of a positive or negative number (scalar) is a positive definite or negative definite matrix. Therefore, we use the notation

i) $\boldsymbol{A} > 0$ if $\boldsymbol{A}$ is positive definite
ii) $\boldsymbol{A} < 0$ if $\boldsymbol{A}$ is negative definite

iii) $\boldsymbol{A} \geq 0$ if $\boldsymbol{A}$ is positive semi-definite
iv) $\boldsymbol{A} \leq 0$ if $\boldsymbol{A}$ is negative semi-definite

The notation $\boldsymbol{A} > \boldsymbol{B}$ means that the matrix $\boldsymbol{A} - \boldsymbol{B}$ is positive definite.

4.5 Cholesky decomposition

Any positive definite and symmetric matrix $\boldsymbol{B}$ can be written as

$$\boldsymbol{B} = \boldsymbol{P}\boldsymbol{P}',$$

where P is a lower triangular matrix with strictly positive diagonal entries $p_{jj} > 0$. This representation is called **Cholesky decomposition**. The matrix $\boldsymbol{P}$ is unique. For a 2×2 matrix $\boldsymbol{B}$ we have

$$\begin{aligned}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} &= \begin{pmatrix} p_{11} & 0 \\ p_{21} & p_{22} \end{pmatrix} \begin{pmatrix} p_{11} & p_{21} \\ 0 & p_{22} \end{pmatrix} \\ &= \begin{pmatrix} p_{11}^2 & p_{11}p_{21} \\ p_{11}p_{21} & p_{21}^2 + p_{22}^2 \end{pmatrix},\end{aligned}$$

which implies $p_{11} = \sqrt{b_{11}}$, $p_{21} = b_{21}/p_{11}$, and $p_{22} = \sqrt{b_{22} - p_{21}^2}$. For a 3×3 matrix we obtain

$$\begin{aligned}\begin{pmatrix} b_{11} & b_{12} & b_{31} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} &= \begin{pmatrix} p_{11} & 0 & 0 \\ p_{21} & p_{22} & 0 \\ p_{31} & p_{32} & p_{33} \end{pmatrix} \begin{pmatrix} p_{11} & p_{21} & p_{31} \\ 0 & p_{22} & p_{32} \\ 0 & 0 & p_{33} \end{pmatrix} \\ &= \begin{pmatrix} p_{11}^2 & p_{11}p_{21} & p_{11}p_{31} \\ p_{11}p_{21} & p_{21}^2 + p_{22}^2 & p_{21}p_{31} + p_{22}p_{32} \\ p_{11}p_{31} & p_{21}p_{31} + p_{22}p_{32} & p_{31}^2 + p_{32}^2 + p_{33}^2 \end{pmatrix},\end{aligned}$$

which implies

$$\begin{gathered} p_{11} = \sqrt{b_{11}}, \quad p_{21} = \frac{b_{21}}{p_{11}}, \quad p_{31} = \frac{b_{31}}{p_{11}}, \quad p_{22} = \sqrt{b_{22} - p_{21}^2}, \\ p_{32} = \frac{b_{32} - p_{21}p_{31}}{p_{22}}, \quad p_{33} = \sqrt{b_{33} - p_{31}^2 - p_{32}^2}. \end{gathered}$$

Let's compute the Cholesky decomposition of

$$\boldsymbol{B} = \begin{pmatrix} 1 & -0.5 & 0.6 \\ -0.5 & 1 & 0.25 \\ 0.6 & 0.25 & 1 \end{pmatrix}$$

using the `R` function `chol()`:

```r
B = matrix(c(1, -0.5, 0.6, -0.5, 1, 0.25, 0.6, 0.25, 1), ncol=3)
chol(B)
```

```
     [,1]       [,2]      [,3]
[1,]    1 -0.5000000 0.6000000
[2,]    0  0.8660254 0.6350853
[3,]    0  0.0000000 0.4864840
```

4.6 Vectorization

The **vectorization operator** vec() stacks the matrix entries column-wise into a large vector. The vectorized $k \times m$ matrix $\boldsymbol{A}$ is the $km \times 1$ vector

$$\mathrm{vec}(\boldsymbol{A}) = (a_{11}, \dots, a_{k1}, a_{12}, \dots, a_{k2}, \dots, a_{1m}, \dots, a_{km})'.$$

```r
c(A) #vectorize the matrix A
```

```
[1]  1  3  0  2  9 11  3  1  5
```

4.7 Kronecker product

The **Kronecker product** $\otimes$ multiplies each element of the left-hand side matrix with the entire matrix on the right-hand side. For a $k \times m$ matrix $\boldsymbol{A}$ and a $r \times s$ matrix $\boldsymbol{B}$, we get the $kr \times ms$ matrix

$$A \otimes B = \begin{pmatrix} a_{11}\boldsymbol{B} & \dots & a_{1m}\boldsymbol{B} \\ \vdots & & \vdots \\ a_{k1}\boldsymbol{B} & \dots & a_{km}\boldsymbol{B} \end{pmatrix},$$

where each entry $a_{ij}\boldsymbol{B}$ is a $r \times s$ matrix.

```r
A %x% B #Kronecker product in R
```

```
      [,1]  [,2] [,3] [,4]  [,5]  [,6] [,7]  [,8] [,9]
[1,]   1.0 -0.50 0.60  2.0 -1.00  1.20  3.0 -1.50 1.80
[2,]  -0.5  1.00 0.25 -1.0  2.00  0.50 -1.5  3.00 0.75
[3,]   0.6  0.25 1.00  1.2  0.50  2.00  1.8  0.75 3.00
[4,]   3.0 -1.50 1.80  9.0 -4.50  5.40  1.0 -0.50 0.60
```

```
[5,] -1.5  3.00 0.75 -4.5  9.00  2.25 -0.5  1.00 0.25
[6,]  1.8  0.75 3.00  5.4  2.25  9.00  0.6  0.25 1.00
[7,]  0.0  0.00 0.00 11.0 -5.50  6.60  5.0 -2.50 3.00
[8,]  0.0  0.00 0.00 -5.5 11.00  2.75 -2.5  5.00 1.25
[9,]  0.0  0.00 0.00  6.6  2.75 11.00  3.0  1.25 5.00
```

4.8 Vector and matrix norm

A norm $\|\cdot\|$ of a vector or a matrix is a measure of distance from the origin. The most commonly used norms are the Euclidean vector norm

$$\|\boldsymbol{a}\| = \sqrt{\boldsymbol{a}'\boldsymbol{a}} = \sqrt{\sum_{i=1}^{k} a_i^2}$$

for $\boldsymbol{a} \in \mathbb{R}^k$, and the Frobenius matrix norm

$$\|\boldsymbol{A}\| = \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{m} a_{ij}^2}$$

for $\boldsymbol{A} \in \mathbb{R}^{k\times m}$.

A norm satisfies the following properties:

i) $\|\lambda\boldsymbol{A}\| = |\lambda|\|\boldsymbol{A}\|$ for any scalar λ (absolute homogeneity)
ii) $\|\boldsymbol{A}+\boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\|$ (triangle inequality)
iii) $\|\boldsymbol{A}\| = 0$ implies $\boldsymbol{A} = \boldsymbol{0}$ (definiteness)

5 Matrix calculus

Let $f(\beta_1, \dots, \beta_k) = f(\boldsymbol{\beta})$ be a twice-differential real-valued function that depends on some vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)'$. Examples that frequently appear in econometrics are functions of the inner product form $f(\boldsymbol{\beta}) = \boldsymbol{a}'\boldsymbol{\beta}$, where $\boldsymbol{a} \in \mathbb{R}^k$, and functions of the sandwich form $f(\boldsymbol{\beta}) = \boldsymbol{\beta}'\boldsymbol{A}\boldsymbol{\beta}$, where $\boldsymbol{A} \in \mathbb{R}^{k \times k}$.

5.1 Gradient

The **first derivatives vector** or **gradient** is

$$\frac{\partial f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_1} \\ \vdots \\ \frac{\partial f(\boldsymbol{\beta})}{\partial \beta_k} \end{pmatrix}$$

If the gradient is evaluated at some particular value $\boldsymbol{\beta} = \boldsymbol{b}$, we write

$$\frac{\partial f}{\partial \boldsymbol{\beta}}(\boldsymbol{b})$$

Useful properties for inner product and sandwich forms are

$$(i) \qquad \frac{\partial(\boldsymbol{a}'\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \boldsymbol{a}$$

$$(ii) \qquad \frac{\partial(\boldsymbol{\beta}'\boldsymbol{A}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = (\boldsymbol{A} + \boldsymbol{A}')\boldsymbol{\beta}.$$

5.2 Hessian

The **second derivatives matrix** or **Hessian** is the $k \times k$ matrix

$$\frac{\partial^2 f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \begin{pmatrix} \frac{\partial^2 f(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_1} & \dots & \frac{\partial^2 f(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_1} \\ \vdots & & \vdots \\ \frac{\partial^2 f(\boldsymbol{\beta})}{\partial \beta_1 \partial \beta_k} & \dots & \frac{\partial^2 f(\boldsymbol{\beta})}{\partial \beta_k \partial \beta_k} \end{pmatrix}.$$

If the Hessian is evaluated at some particular value $\boldsymbol{\beta} = \boldsymbol{b}$, we write

$$\frac{\partial^2 f}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}(\boldsymbol{b})$$

The Hessian is symmetric. Each column of the Hessian is the derivative of the components of the gradient for the corresponding variable in $\boldsymbol{\beta}'$:

$$\begin{aligned}
\frac{\partial^2 f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} &= \frac{\partial(\partial f(\boldsymbol{\beta})/\partial \boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \\
&= \begin{bmatrix} \frac{\partial(\partial f(\boldsymbol{\beta})/\partial \boldsymbol{\beta})}{\partial \beta_1} & \frac{\partial(\partial f(\boldsymbol{\beta})/\partial \boldsymbol{\beta})}{\partial \beta_2} & \dots & \frac{\partial(\partial f(\boldsymbol{\beta})/\partial \boldsymbol{\beta})}{\partial \beta_n} \end{bmatrix}
\end{aligned}$$

The Hessian of a sandwich form function is

$$\frac{\partial^2 (\boldsymbol{\beta}' \boldsymbol{A} \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \boldsymbol{A} + \boldsymbol{A}'.$$

5.3 Optimization

Recall the *first-order* (necessary) and *second-order* (sufficient) conditions for optimum (maximum or minimum) in the univariate case:

- **First-order condition**: the first derivative evaluated at the optimum is zero.
- **Second-order condition**: the second derivative at the optimum is negative for a maximum and positive for a minimum.

Similarly, we formulate first and second-order conditions for a function $f(\boldsymbol{\beta})$. The **first-order condition** for an optimum (maximum or minimum) at $\boldsymbol{b}$ is

$$\frac{\partial f}{\partial \boldsymbol{\beta}}(\boldsymbol{b}) = \boldsymbol{0}.$$

The **second-order condition** is

$$\begin{aligned}
\frac{\partial^2 f}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}(\boldsymbol{b}) > 0 \quad \text{for a minimum at } \boldsymbol{b}, \\
\frac{\partial^2 f}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}(\boldsymbol{b}) < 0 \quad \text{for a maximum at } \boldsymbol{b}.
\end{aligned}$$

Recall that, in the context of matrices, the notation "> 0" means positive definite, and "< 0" means negative definite.

6 Problems

Problem 1

Consider the matrix

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

a) Determine $\boldsymbol{A}'$. Is $\boldsymbol{A}$ symmetric?
b) Is $\boldsymbol{A}$ idempotent?
c) Compute the determinant and the rank. Is $\boldsymbol{A}$ nonsingular?
d) Compute the inverse.
e) Compute the trace.

Problem 2

a) Let $\boldsymbol{AB} = \boldsymbol{C}$, where

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 2 & 2 \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 1 & 3 \\ 2 & 5 \end{pmatrix}.$$

Compute $\boldsymbol{B}$.
b) $\boldsymbol{\delta}$ and $\boldsymbol{\gamma}$ are $c \times 1$ vectors, $\boldsymbol{X}$ is a $d \times c$ matrix, and $\boldsymbol{Y}$ is a $c \times d$ matrix. Determine the orders of $\boldsymbol{XY}$, $\boldsymbol{YX}$, $\boldsymbol{\gamma}'\boldsymbol{\gamma}$, $\boldsymbol{\gamma}\boldsymbol{\gamma}'$, and $\boldsymbol{\delta}'\boldsymbol{YX}\boldsymbol{\gamma}$. Under which conditions do the expressions $\boldsymbol{Y}^{-1}$ and $\boldsymbol{\delta}'\boldsymbol{YX} + \boldsymbol{\gamma}'\boldsymbol{\gamma}$ exist?
c) Compute $\operatorname{tr}(\lambda \boldsymbol{R}'\boldsymbol{R})$ for $\lambda \in \mathbb{R}$ and

$$\boldsymbol{R} = \begin{pmatrix} \frac{1}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{3}{4} \end{pmatrix}.$$

Problem 3

Let $\boldsymbol{A}$ be nonsingular. Simplify the expression

$$\left(\frac{1}{\sqrt{2}}\boldsymbol{A}^{-1}\left(\frac{1}{\sqrt{2}}\boldsymbol{A}'' + \frac{\sqrt{2}}{2}\boldsymbol{A}\right)\right).$$

Problem 4

Consider the $n \times k$ matrix $\boldsymbol{X}$ with $\operatorname{rank}(\boldsymbol{X}) = k$. Moreover, let $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, and let $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{P}$

a) Determine the order of the following matrices: $\boldsymbol{I}_n$, $\boldsymbol{X}'\boldsymbol{X}$, $\boldsymbol{P}$, $\boldsymbol{M}$
b) Which matrices from a) are symmetric?
c) Which matrices from a) are idempotent?
d) Compute the trace of $\boldsymbol{I}_n$ and $\boldsymbol{P}$.

Problem 5

Let $\boldsymbol{X}$ be a $n \times k$ matrix. Show that $\boldsymbol{X}'\boldsymbol{X}$ is positive semi-definite. Under which condition is $\boldsymbol{X}'\boldsymbol{X}$ positive definite?

Problem 6

Let $\boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{X}$ be a $n \times k$ matrix, and $\boldsymbol{\beta} \in \mathbb{R}^k$. Compute the derivatives

$$\frac{\partial f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}, \quad \frac{\partial^2 f(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'},$$

for the function $f(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$.

6.1 Solutions

Solutions to the problems will be added soon.