Introduction to Matrix Algebra for Statistics

Contents

1 Definitions

Matrix is a rectangular array of numbers arranged into rows and columns. A general matrix consisting r rows and c columns is of the form

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1c} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2c} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3c} \\ \vdots & \vdots & \vdots & \ddots & \\ a_{r1} & a_{r2} & a_{r3} & \dots & a_{rc} \end{bmatrix},$$

where the numbers a_{ij} are called elements or entries of the matrix $\mathbf{A}$.Three dots indicate, for example in the first row, that the elements a_{11}, a_{12}, $a_{13,}$ continue in sequence up to a_{1c}. The elements are often denoted by lower case letters with a subscripts, e.g. a_{ij} refers to the element in row i and column j of the matrix $\mathbf{A}$. For example a_{13} is the element in the first row and the third column of the matrix $\mathbf{A}$.

Previous form of writing a matrix specifies its entries and also the number of rows and columns that is referred to as the dimension (or order or size) of the matrix. Dimension of the previous matrix $\mathbf{A}$ is $r \times c$ (i.e. matrix $\mathbf{A}$ has r rows and c columns). Sometimes it is useful to use subscript notation to denote the dimension of a matrix. For example $\mathbf{A}_{r\times c}$ is a matrix having r rows and c columns.

Example: R dataset trees provides measurements of the girth, height and volume of timber in 31 felled black cherry trees. The following table presents the first 7 rows of the data.

Girth	**Height**	**Volume**
8.3	70	10.3
8.6	65	10.3
8.8	63	10.2
10.5	72	16.4
10.7	81	18.8
10.8	83	19.7
11.0	66	15.6

This table can be presented using 7×3 matrix

$$\begin{bmatrix} 8.3 & 70 & 10.3 \\ 8.6 & 65 & 10.3 \\ 8.8 & 63 & 10.2 \\ 10.5 & 72 & 16.4 \\ 10.7 & 81 & 18.8 \\ 10.8 & 83 & 19.7 \\ 11.0 & 66 & 15.6 \end{bmatrix}$$

Note: The word *order* is sometimes used for other characteristics of a matrix, but in this handout order always refers to the number of rows and columns.

Example: Dimension of the matrix

$$\mathbf{B} = \{b_{ij}\}_{3\times 4} = \begin{bmatrix} 2 & 0 & 2.4 & 0 \\ -8 & 1 & 5 & 3 \\ 6 & 4 & -4 & 4 \end{bmatrix}$$

is 3×4 (read 3 by 4) and element at the row 1 and column 3 is $b_{13} = 2.4$.

Note: Bold capital letters are commonly used to denote matrices. For example $\mathbf{X}$ and $\mathbf{Y}$ refers to matrices $\mathbf{X}$ and $\mathbf{Y}$. In this handout elements are real numbers, unless otherwise stated. In practice data matrix might also have missing values.

Note: It is conventional to use round or square brackets to delineate matrices when they are written as rectangular arrays.

Matrix $\mathbf{A}$ can be presented in a more compact form as

$$\mathbf{A} = \{a_{ij}\}_{r\times c}$$

where a_{ij} indicates the element at the intersection of the i^{th} row and j^{th} column and subscript $r \times c$ indicates the dimension. The curly brackets indicate that element a_{ij} is a typical element.

The i^{th} row of the matrix $\mathbf{A}_{r\times c}$ is

$$\begin{bmatrix} a_{i1} & a_{i2} & a_{i3} & \ldots & a_{ic} \end{bmatrix}$$

and the j^{th} column of the matrix $\mathbf{A}_{r\times c}$ is

$$\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{rj} \end{bmatrix}.$$

Note: A matrix with single row is called a row vector and a matrix with single column is called a column vector. Vectors are denoted by bold lower-case letters and the elements of the vector are denoted by lower-case letters with single subscript indicating its position in the vector. The prime superscript $'$ or superscript ^{T} is often used to distinguish a row vector from a column vector i.e. vector $\mathbf{x}$ is a column vector and $\mathbf{x}'$ is a row vector. We will later learn that the prime superscript denotes the transpose of a vector/matrix.

Note: A vector with n elements is called a n-dimensional vector since we can think each observation as a point in the n-dimensional space and the whole data is tought of as being a cloud of points in that space. For example a scatter plot of two variables shows data as a cloud of points on a 2 dimensional plane.

Example: Let

$\mathbf{a}' = \begin{bmatrix} 2 & -1 & 5 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$.

Now vector $\mathbf{a}'$ is a row vector with 3 elements i.e. it is a 3 dimensional vector. The third element of the vector $\mathbf{a}'$ is 5.

Vector $\mathbf{b}$ is a column vector with 2 elements.

Note: A 1×1 matrix is called a scalar. In other words, a scalar is a single number. Scalars are often denoted by lower case letters.

R example Vectors are one of the basic data structures in R that are often needed. In R Vectors are defined using function c(). You can open function help page using command ?c or help("c") By default setting c() combines arguments to form a vector.

```r
# The easiest way to create a vector is to use the function c()
# In this example we create a numeric vector called x that contain values 1,0,-3 and 4.
x <- c(1,0,-3,4)
# In general "<-" is the assign operator that assign a value to the variable.
# In this example we assigned vector to the variable x
# print vector x
x
```

```
## [1]  1  0 -3  4
```

There are various ways to construct matrices in R. Small matrices can be entered directly using the command matrix().

Matrices in external files can be imported into R using the specific importing commands like scan(), read.table(), read.csv() or read.xls(). In this handot we don't deal with external datasets that should be imported into R. You can find information about importing data from the "An Introduction to R" manual (https://cran.r-project.org/doc/manuals/r-release/R-intro.pdf).

```r
# Define matrix B that we used previously
B <- matrix(c(2,0,2.4,0,-8,1,5,3,6,4,-4,4),byrow=TRUE,nrow=3,ncol=4)
# Here argument byrow=TRUE defines that elements are filled in a matrix row by row in
```

```r
# a given order.
# Elements are given in a vector form using the function c(), nrow=3 defines that matrix
# has 3 rows and ncol=4 defines that we want to create a matrix which has 4 columns.
#
# Print matrix B
B
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    2    0  2.4    0
## [2,]   -8    1  5.0    3
## [3,]    6    4 -4.0    4
```

```r
# Notice we could leave out nrow=3 or ncol=4 argument since the number of rows or
# the number of columns is enough to determine also the number of columns/rows
B <- matrix(c(2,0,2.4,0,-8,1,5,3,6,4,-4,4),byrow=TRUE,nrow=3)
B
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    2    0  2.4    0
## [2,]   -8    1  5.0    3
## [3,]    6    4 -4.0    4
```

```r
# Extract the element in row 1 and column 3
B[1,3]
```

```
## [1] 2.4
```

```r
# Extract the third row (column index is left empty)
B[3,]
```

```
## [1]  6  4 -4  4
```

```r
# and the second column (row index is left empty)
B[,2]
```

```
## [1] 0 1 4
```

```r
# Extract the second and the third elements from the first row
B[1,2:3]
```

```
## [1] 0.0 2.4
```

```r
# or
B[1,c(2,3)]
```

```
## [1] 0.0 2.4
```

```r
# Matrix is filled column by column if we set byrow=FALSE or don't set it at all.
C <- matrix(data=c(2,0,2.4,0,-8,1,5,3,6,4,-4,4),byrow=FALSE,nrow=3)
# Print matrix C
C
```

```
##      [,1] [,2] [,3] [,4]
## [1,]  2.0    0    5    4
## [2,]  0.0   -8    3   -4
## [3,]  2.4    1    6    4
```

```r
C <- matrix(data=c(2,0,2.4,0,-8,1,5,3,6,4,-4,4),nrow=3)
# Print matrix C
C
```

```
##      [,1] [,2] [,3] [,4]
## [1,]  2.0    0    5    4
## [2,]  0.0   -8    3   -4
## [3,]  2.4    1    6    4
```

```r
# We can use the function length() to determine the length of the vector and dim() to
# determine the dimension of the matrix
# Define vector having 5 elements
y <- c(0,8,-3,1,4)
# Length of the vector y is
length(y)
```

```
## [1] 5
```

```r
# Determine the dimension of the matrix C
# Print previously defined matrix C
C
```

```
##      [,1] [,2] [,3] [,4]
## [1,]  2.0    0    5    4
## [2,]  0.0   -8    3   -4
## [3,]  2.4    1    6    4
```

```r
# Dimension of the matrix C
dim(C)
```

```
## [1] 3 4
```

```r
# The function dim() returns a vector where the first element tells the number of rows
# and the second tells  the number of columns. For example matrix C has 3 rows
# and 4 columns.
```

A matrix with n rows and n columns ($n \times n$) is called a square matrix. In other words matrix is square matrix when the number of rows equals the number of columns. The main diagonal elements of a square matrix are elements where the row and the column indices are the same i.e. the elements a_{11}, a_{22}, a_{33},. . . , a_{nn} are said to lie on the main diagonal of the square matrix $\mathbf{A}_{n\times n}$. The main diagonal elements are marked in the following 3×3 matrix

$$\begin{bmatrix} \boxed{1} & 2 & 3 \\ 4 & \boxed{5} & 6 \\ 7 & 8 & \boxed{9} \end{bmatrix}.$$

In the 3×3 square matrix above, the diagonal elements are 1, 5 and 9.

```r
# Define matrix
A <- matrix(c(1,2,3,4,5,6,7,8,9),byrow=TRUE,nrow=3,ncol=3)
A
```

```
##      [,1] [,2] [,3]
## [1,]    1    2    3
## [2,]    4    5    6
## [3,]    7    8    9
```

```r
# or simply use operator : to define a vector which has integers from 1 to 9
1:9
```

```
## [1] 1 2 3 4 5 6 7 8 9
```

```r
A <- matrix(1:9,byrow=TRUE,nrow=3,ncol=3)
# Command diag() extracts the diagonal elements.
diag(A)
```

```
## [1] 1 5 9
```

```r
# We can assign diagonal elements to vector d and print them.
d <- diag(A); d
```

```
## [1] 1 5 9
```

```r
# Notice that we combined two different commands on the same line using operator ;
# Usually it is cleaner to write one commend per line
```

A square matrix is said to be a diagonal matrix, if all entries off the main diagonal are zero. The general form of $n \times n$ diagonal matrix is

$$\mathbf{A}_{n\times n} = \begin{bmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & & 0 \\ \vdots & & \ddots & 0 \\ 0 & 0 & 0 & a_{nn} \end{bmatrix}.$$

That is, if $\mathbf{A} = \{a_{ij}\}_{n\times n}$ is a square matrix and $a_{ij} = 0$ for all $i \neq j$, then $\mathbf{A}$ is a diagonal matrix. Diagonal matrices are sometimes denoted using notation $diag(a_1, a_2, ..., a_n)$ or $\mathbf{D}(a_1, a_2, ..., a_n)$ in which diagonal elements are $a_1, a_2, ..., a_n$.

Special type of diagonal matrix is the identity matrix in which all diagonal elements are 1 and all off diagonal elements are 0. The identity matrix is often denoted with a bold capital letter $\mathbf{I}$. Dimension of the identity matrix can be denoted with a subscript. For example $\mathbf{I}_n$ is a identity matrix having n rows and n columns.

Example: Consider matrices

$$\mathbf{A} = \begin{bmatrix} -1 & 0 \\ 0 & 4 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix} \text{ and } \mathbf{C} = \begin{bmatrix} -1 & 0 \\ 1 & 4 \end{bmatrix}$$

Matrices $\mathbf{A}$ and $\mathbf{B}$ are diagonal matrices but the matrix $\mathbf{C}$ isn't since the element at row 2 and column 2 isn't zero.

Matrix

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is a identity matrix.

```r
# In R we can define diagonal matrix using command diag()
diag(c(3,0,4))
```

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

```r
# Assign the diagonal matrix to variable A
A <- diag(c(1,-1,4))
A
```

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

```r
# Also identity matrices can be created using command diag()
# The number of dimensions is set inside the parentheses.
diag(3)
```

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

A square matrix is called a upper triangular matrix if all of the entries below the main diagonal are zero.

Matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \end{bmatrix}$$

is a upper triangular matrix.

Similarly, a square matrix is called a lower triangular matrix if all of the entries above the main diagonal are zero.

Matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 4 & 5 & 0 \end{bmatrix}$$

is a lower triangular matrix.

A zero matrix (or a null matrix) is a matrix with all its entries being zero. Zero matrices are denoted with **0** (bold zero). Dimension of the zero matrix can be denoted with a subscript e.g.

$$\mathbf{0}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

```r
# In R we can select upper or lower diagonal elements of a matrix using commands
# lower.tri() and upper.tri()
# Define matrix
A <- matrix(c(1,2,3,4,1,2,1,1,3),nrow=3,byrow=T)
A
```

```
##      [,1] [,2] [,3]
## [1,]    1    2    3
```

```
## [2,]    4    1    2
## [3,]    1    1    3
```

```r
lower.tri(A,diag=TRUE)
```

```
##      [,1]  [,2]  [,3]
## [1,] TRUE FALSE FALSE
## [2,] TRUE  TRUE FALSE
## [3,] TRUE  TRUE  TRUE
```

```r
# or
lower.tri(A,diag=FALSE)
```

```
##       [,1]  [,2]  [,3]
## [1,] FALSE FALSE FALSE
## [2,]  TRUE FALSE FALSE
## [3,]  TRUE  TRUE FALSE
```

```r
# Now we will set that all elements below the main diagonal are zero. We replace all
# corresponding elements that have logical value TRUE by 0
A[lower.tri(A,diag=FALSE)] <- 0
A
```

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

Matrices $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$ are said to be equal if matrices have the same dimension and all corresponding entries are equal i.e. $a_{ij} = b_{ij}$ for all i and j.

Example 1: Matrices $\mathbf{A} = \begin{bmatrix} 1 & 4 \\ -3 & c \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 4 \\ -3 & 0 \end{bmatrix}$ are equal if and only if $c = 0$.

Example 2: Matrices

$\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$ are not equal although they have the same entries.

```r
# We can use comparison operators to make comparisons for the elements of the matrix

# Define matrices A and B
A <- matrix(c(1,4,-3,2),nrow=2,byrow=T)
B <- matrix(c(1,4,-3,0),nrow=2,byrow=T)
# We can use function identical() see if the two matrices are identical
identical(A,B)
```

```
## [1] FALSE
```

```r
# The matrices are not equal so R returned the logical value FALSE. If the matrices are
# equal R will return logical value TRUE
# Define matrix X that is equal to A
X <- matrix(c(1,4,-3,2),nrow=2,byrow=T)
# Test equality
identical(A,X)
```

```
## [1] TRUE
```

```r
# Element wise comparisons can be done using comparison operators. For example we can
# compare which elements of matrix A are equal to corresponding elements of matrix B
A == B
```

```
##      [,1]  [,2]
## [1,] TRUE  TRUE
## [2,] TRUE FALSE
```

```r
# Notice equality is compared using operator == (and operator = is the assignment operator)
X
```

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

```r
X=3 # Assign value 3 to X
X
```

```
## [1] 3
```

```r
# Compare which elements of A are larger than corresponding elements of B
A > B
```

```
##       [,1]  [,2]
## [1,] FALSE FALSE
## [2,] FALSE  TRUE
```

2 Basic matrix operations

2.1 Addition and subtraction

Two matrices are said to be compatible if matrices have the same dimension i.e. both matrices have the same number of rows and the same number of columns.

If and only if two matrices are compatible they can be added or subtracted. The sum is a matrix of the same size as the operand matrices and it is calculated by adding the corresponding elements of the matrices.

Let $\mathbf{A} = \{a_{ij}\}_{r\times c}$ and $\mathbf{B} = \{b_{ij}\}_{r\times c}$ then the sum of the matrices $\mathbf{A}$ and $\mathbf{B}$ is

$\mathbf{A} + \mathbf{B} = \{a_{ij} + b_{ij}\}_{r\times c}$.

Example: Let $\mathbf{A} = \begin{bmatrix} -1 & 2 \\ 3 & 4 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0.5 & 8 \\ 3.5 & 1 \end{bmatrix}$ and $\mathbf{C} = \begin{bmatrix} 2 & 3 & 9 \\ 1 & 0 & -1 \end{bmatrix}$.

Now matrices $\mathbf{A}$ and $\mathbf{B}$ are compatible (2×2 matrices) so they can be added together

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} -1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 0.5 & 8 \\ 3.5 & 1 \end{bmatrix} = \begin{bmatrix} -1+0.5 & 2+8 \\ 3+3.5 & 4+1 \end{bmatrix} = \begin{bmatrix} -0.5 & 10 \\ 6.5 & 5 \end{bmatrix}.$$

Note: The additions $\mathbf{A} + \mathbf{C}$ or $\mathbf{B} + \mathbf{C}$ cannot be calculated, as matrices to be added are not of the same dimension. The dimension of the matrix $\mathbf{C}$ is 2×3 and the dimension of the matrices $\mathbf{A}$ and $\mathbf{B}$ is 2×2

```r
# Define matrices A, B and C
A <- matrix(c(-1,2,3,4),nrow=2,byrow=T)
B <- matrix(c(0.5,8,3.5,1),nrow=2,byrow=T)
C <- matrix(c(2,3,9,1,0,-1),nrow=3,byrow=T)
# Compute the addition A + B
A+B
```

```
##      [,1] [,2]
## [1,] -0.5   10
## [2,]  6.5    5
```

```r
# A+C is not defined. R will notify us about conformability if we try to
# calculate the sum
```

Subtraction is analogous to matrix addition. Matrices can be subtracted if each matrix to be subtracted has the same dimensions. Subtraction is computed by subtracting corresponding elements.

Let $\mathbf{A} = \{a_{ij}\}_{r \times c}$ and $\mathbf{B} = \{b_{ij}\}_{r \times c}$ then

$\mathbf{A} - \mathbf{B} = \{a_{ij} - b_{ij}\}_{r \times c}$.

Example: Let $\mathbf{A} = \begin{bmatrix} -1 & 2 \\ 3 & 4 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 0.5 & 8 \\ 3.5 & 1 \end{bmatrix}$.

Since matrices $\mathbf{A}$ and $\mathbf{B}$ are of the same dimension (2×2 matrices) we can subtract them

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} -1 & 2 \\ 3 & 4 \end{bmatrix} - \begin{bmatrix} 0.5 & 8 \\ 3.5 & 1 \end{bmatrix} = \begin{bmatrix} -1-0.5 & 2-8 \\ 3-3.5 & 4-1 \end{bmatrix} = \begin{bmatrix} -1.5 & -6 \\ -0.5 & 3 \end{bmatrix}$$

```r
# Define matrices A and B
A <- matrix(c(-1,2,3,4),nrow=2,byrow=T)
B <- matrix(c(0.5,8,3.5,1),nrow=2,byrow=T)
# Compute the subtraction A - B and assign the result to the matrix C
C <- A-B
# print matrix C
C
```

```
##      [,1] [,2]
## [1,] -1.5   -6
## [2,] -0.5    3
```

2.1.1 Properties of the matrix addition

Matrix addition (and subtraction) has similar properties as addition of numbers.

1. Commutative property of addition: $\mathbf{A}+\mathbf{B}=\mathbf{B}+\mathbf{A}$. Commutative property means that changing the order of the operands does not change the result.

2. Associative property of addition: $\mathbf{A}+(\mathbf{B}+\mathbf{C})=(\mathbf{A}+\mathbf{B})+\mathbf{C}$. Associative property says that both formulas will give the same result so the parenthesis are unnecessary.

3. Additive identity property: For any matrix $\mathbf{A}$ there is a unique matrix $\mathbf{0}$ such that $\mathbf{A}+\mathbf{0}=\mathbf{A}$

4. Additive inverse: For every matrix $\mathbf{A}$ there is a matrix $-\mathbf{A}$ such that $\mathbf{A}+(-\mathbf{A})=\mathbf{0}$, where $\mathbf{0}$ is a matrix with all its elements being zero. The additive inverse of matrix is the matrix obtained by changing the sign of every element of the matrix.

5. Closure property of addition: $\mathbf{A}+\mathbf{B}$ is a matrix of the same dimension as $\mathbf{A}$ and $\mathbf{B}$

2.2 Transpose

Transpose of any $m \times n$ matrix $\mathbf{A}$ is the $n \times m$ matrix $\mathbf{A}'$ that is obtained by interchanging rows and columns of $\mathbf{A}$. The transpose of a matrix $\mathbf{A}$ is often written as $\mathbf{A}'$ or $\mathbf{A}^T$ i.e. when there is a prime superscript it means that matrix or vector is transposed.

If $\mathbf{A}=\{a_{ij}\}$, then the transpose of $\mathbf{A}$ is defined as

$\mathbf{A}'=\{a_{ji}\}$. That is element in row i and column j in matrix $\mathbf{A}$ is in row j and column i in the transposed matrix.

Example: If $\mathbf{A}=\begin{bmatrix}1 & 2 & 3\\ 4 & 5 & 6\\ 7 & 8 & 9\\ 10 & 11 & 12\end{bmatrix}$ then the transpose of $\mathbf{A}$ is

$$\mathbf{A}' = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix}$$

```r
# R example of computing transpose
# In R we compute transpose using function t()
# For example define matrix A
A <- matrix(1:12,nrow=3,byrow=T)
A
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    2    3    4
## [2,]    5    6    7    8
## [3,]    9   10   11   12
```

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

```
##      [,1] [,2] [,3]
## [1,]    1    5    9
## [2,]    2    6   10
## [3,]    3    7   11
## [4,]    4    8   12
```

Note: Transpose does not change the diagonal elements (a_{ii}) of a matrix $\mathbf{A} = \{a_{ij}\}$.

Note: Matrix $\mathbf{A}$ is said to be symmetric if $\mathbf{A} = \mathbf{A}'$ i.e. transpose of the matrix is equal to itself. Symmetric matrices are often encountered in statistics. For example correlation and covariance matrices are symmetric.

Note: The transpose of a column vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_c \end{bmatrix}$$

is a row vector $\mathbf{x}' = \begin{bmatrix} x_1 & x_2 & \dots & x_c \end{bmatrix}$ and vice versa the transpose of a row vector is a column vector. It is commonly used convention that all vectors are column vectors unless otherwise stated and this is the explanation why we earlier denoted row vectors with the prime superscript.

```r
# R example of computing the transpose
# Define vector x having elements 3,7,4,1,0,-4
x <- c(3,7,4,1,0,-4)
# The transpose of x is arow vector
t(x)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    3    7    4    1    0   -4
```

```r
# and the transpose of the transpose is a column vector
t(t(x))
```

```
##      [,1]
## [1,]    3
## [2,]    7
## [3,]    4
## [4,]    1
## [5,]    0
## [6,]   -4
```

Note: If the matrix $\mathbf{A}$ is a $r \times c$ matrix then the transpose of $\mathbf{A}$ is a $c \times r$ matrix. If we use subscript notation for the dimension of a $\mathbf{A}$ then marking $\mathbf{A}'_{r\times c}$ is ambiguous since this could mean that the dimension of transposed matrix $\mathbf{A}$ is $r \times c$ or that the transpose has dimension $r \times c$. This ambiguity can be avoided using brackets for clarity i.e. $(\mathbf{A}_{r\times c})'$ if $\mathbf{A}$ is a $r \times c$ matrix or $(\mathbf{A}')_{r\times c}$ if the transpose is a $r \times c$ matrix.

2.2.1 Properties of the transpose

1) The transpose of a transposed matrix is the matrix itself i.e. $(\mathbf{A}')' = \mathbf{A}$
2) Scalar multiplication: $(c\mathbf{A})' = c\mathbf{A}'$, when c is a scalar (see below)
3) Transpose of the sum: The transpose of two added matrices is the same as the addition of the two transposed matrices i.e. $(\mathbf{A}+\mathbf{B})' = \mathbf{A}' + \mathbf{B}'$

2.3 Multiplication by scalar

In scalar multiplication, each element of the matrix are multiplied by the given scalar i.e. real number.

If λ is a scalar and $\mathbf{A} = \{a_{ij}\}$ is a $r \times c$ matrix, then the result of scalar multiplication is a $r \times c$ matrix

$\lambda\mathbf{A} = \lambda\{a_{ij}\}_{r\times c} = \{\lambda \cdot a_{ij}\}_{r\times c}$.

Example: Let $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$. Now

$$3\mathbf{A} = 3\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} 3\cdot 1 & 3\cdot 2 & 3\cdot 3 \\ 3\cdot 4 & 3\cdot 5 & 3\cdot 6 \end{bmatrix} = \begin{bmatrix} 3 & 6 & 9 \\ 12 & 15 & 18 \end{bmatrix}$$

```r
# R example of scalar multiplication
# Define matrix A
A <- matrix(1:6,nrow=2,byrow=T)
# Compute 3*A
3*A
```

```
##      [,1] [,2] [,3]
## [1,]    3    6    9
## [2,]   12   15   18
```

```r
# or
A*3
```

```
##      [,1] [,2] [,3]
## [1,]    3    6    9
## [2,]   12   15   18
```

Note: The negative of a matrix is formed by multiplying each element of the matrix by scalar -1. That is negation of a matrix $\mathbf{X}$ is $-\mathbf{X} = -1\mathbf{X}$.

2.3.1 Properties

Scalar multiplication also satisfies the usual properties of multiplication of the real numbers. If λ_1, λ_2 and λ_3 are scalars then

1) The associative property of multiplication: $(\lambda_1\lambda_2)\mathbf{A} = \lambda_1(\lambda_2\mathbf{A})$
2) The distributive property: $\lambda(\mathbf{A}+\mathbf{B}) = \lambda\mathbf{A} + \lambda\mathbf{B}$.
3) The multiplicative identity property: $1\mathbf{A} = \mathbf{A}$
4) Closure property of multiplication: $\lambda\mathbf{A}$ is a matrix of the same dimension as matrix $\mathbf{A}$.
5) Multiplicative properties of zero: $0\mathbf{A} = \mathbf{0}$, where $\mathbf{0}$ is a zero matrix of the same dimension as $\mathbf{A}$.

2.4 Inner product

Multiplication of a row vector $\mathbf{x}'$ by a column vector $\mathbf{y}$ is called an inner product ($\mathbf{x}'\mathbf{y}$). This multiplication is possible only if the vectors have the same number of elements.

The result of the inner product is a scalar that is computed as the sum of the products of the corresponding elements of the vectors. That is if

$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ so $\mathbf{x}' = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, then their inner product is defined as

$$\mathbf{x}'\mathbf{y} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1y_1 + x_2y_2 + x_3y_3 + ... + x_ny_n = \sum_{i=1}^{n} x_iy_i$$

Note: In inner product it doesn't matter in which order we do the multiplication: $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$

Note: Inner product is defined only if the two vectors are of the equal size.

Example: The local shop sells 3 types of products and the customer needs certain amount of each product.

Product 1 cost 30 euro each.
Product 2 cost 5.5 euro each.
Product 3 cost 10 euro each.

Customer needs 45 items of product 1, 100 items of product 2 and 50 items of product 3. We can present this as an table

	Price	Bought quantity
Product 1	30	45
Product 2	5.5	100
Product 3	10	50

This table can be written as a 3×2 matrix

$$\begin{bmatrix} 30 & 45 \\ 5.5 & 100 \\ 10 & 50 \end{bmatrix}$$

The row vector containing sales prices of the products is

$$\mathbf{a}' = \begin{bmatrix} 30 & 5.5 & 10 \end{bmatrix}$$

and row vector

$$\mathbf{x}' = \begin{bmatrix} 45 & 100 & 50 \end{bmatrix}$$

is a vector containing sold quantities. Now the total price is

$$\mathbf{a}'\mathbf{x} = \begin{bmatrix} 30 & 5.5 & 10 \end{bmatrix} \begin{bmatrix} 45 \\ 100 \\ 50 \end{bmatrix} = 30 \cdot 45 + 5.5 \cdot 100 + 10 \cdot 50 = 2400.$$

So the total price 2400 euro is calculated as the inner product of the prices and quantities.

Note: When vectors are not of the same order the inner product is undefined.

Example: Inner product is undefined for the vectors $\mathbf{x}' = \begin{bmatrix} 1 & -2 & 3 \end{bmatrix}$ and $\mathbf{y}' = \begin{bmatrix} 8 & 3 & 4 & 0 \end{bmatrix}$ since vectors are not of the same order.

```
# R example of the inner product
# Define vectors x and y
# Vector that contain selling prices
x <- c(30,5.5,10)
# Vector that contain bought quantities
y <- c(45,100,50)
# operator * does element-wise multiplication
x*y
```

```
## [1] 1350  550  500
```

```
# Matrix multiplication is done using the operator %*%
# The inner product between the vectors is
# (Note that the transpose of x does not need to be defined explicitly.)
x%*%y
```

```
##      [,1]
## [1,] 2400
```

```
# Notice that the inner product x%*%y is a 1 by 1 matrix
class(x%*%y)
```

```
## [1] "matrix"
```

```
# If we want to convert this to scalar we can use function drop() that will
# delete the dimensions of a 1 by 1 matrix.
drop(x%*%y)
```

```
## [1] 2400
```

```
class(drop(x%*%y))
```

```
## [1] "numeric"
```

```
# Notice that the inner product is the sum of element-wise products. That is
sum(x*y)
```

```
## [1] 2400
```

2.5 Multiplying matrix by a vector

In the previous example we calculated total cost of buying products from the local shop. Consider now that prices of the same products in the second shop were 29 euro, 6 euro and 9.5 euro respectively. We can present selling prices in the row vector $\begin{bmatrix} 29 & 6 & 9.5 \end{bmatrix}$.

Then purchasing the required products from the second shop would cost

$\begin{bmatrix} 29 & 6 & 9.5 \end{bmatrix} \begin{bmatrix} 45 \\ 100 \\ 50 \end{bmatrix} = 29 \cdot 45 + 6 \cdot 100 + 9.5 \cdot 50 = 2380$ (euro). This was calculated as an inner product.

If we combine two vectors containing selling prices in the matrix

$$\mathbf{A} = \begin{bmatrix} 30 & 5.5 & 10 \\ 29 & 6 & 9.5 \end{bmatrix},$$

we can present total costs of buing products from shops simultaneously as a single equation

$$\mathbf{Ax} = \begin{bmatrix} 30 & 5.5 & 10 \\ 29 & 6 & 9.5 \end{bmatrix} \begin{bmatrix} 45 \\ 100 \\ 50 \end{bmatrix} = \begin{bmatrix} 30 \cdot 45 + 5.5 \cdot 100 + 10 \cdot 50 \\ 29 \cdot 45 + 6 \cdot 100 + 9.5 \cdot 50 \end{bmatrix} = \begin{bmatrix} 2400 \\ 2380 \end{bmatrix}$$

The elements of the resulting vector are the inner products presented previously. That is, the elements of the result vector $\mathbf{Ax}$ are derived in the same way as the inner product $\mathbf{a}'\mathbf{x}$ was calculated earlier, using the rows of matrix $\mathbf{A}$ as the vector $\mathbf{a}'$. This was an example of procut $\mathbf{Ax}$ that is obtained by repetition of the inner product $\mathbf{a}'\mathbf{x}$ using the rows of $\mathbf{A}$ as the vector $\mathbf{a}'$ and writing the result as a column vector.

Note: In the previous examples we did only consider the case where the same number of products is bought from all shops.

```r
# R example of multiplying matrix by a vector
# Define vector a1 that contain selling prices of the local store
a1 <- c(30,5.5,10)
# Define vector a2 that contain selling prices of the second store
a2 <- c(29,6,9.5)
# Vector that contain bought quantities
x <- c(45,100,50)
# Functions rbind() and cbind: combine vectors into a matrix by row or column.
cbind(a1,a2)
```

```
##        a1   a2
## [1,] 30.0 29.0
```

```
## [2,]  5.5  6.0
## [3,] 10.0  9.5
```

```
A <- rbind(a1,a2)
A
```

```
##    [,1] [,2] [,3]
## a1   30  5.5 10.0
## a2   29  6.0  9.5
```

```
# Compute the product Ax using the operator %*%
A%*%x
```

```
##    [,1]
## a1 2400
## a2 2380
```

```
# Notice that A*x is the element-wise product
A*x
```

```
##    [,1] [,2] [,3]
## a1 1350  275 1000
## a2 2900  270  475
```

This procedure can be presented more generally. If matrix $\mathbf{A} = \{a_{ij}\}_{r \times c}$ is a $r \times c$ matrix i.e. $\mathbf{A}$ has r rows and j columns, then element at row i and column j is a_{ij}.

Let $\mathbf{A} = \{a_{ij}\}_{r \times c} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1j} & \cdots & a_{1c} \\ a_{21} & a_{22} & \cdots & a_{2j} & \cdots & a_{2c} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{i1} & a_{j2} & \cdots & a_{ij} & \cdots & a_{ic} \\ \vdots & \vdots & & \vdots & & \vdots \\ a_{r1} & a_{r2} & \cdots & a_{rj} & \cdots & a_{rc} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_r^T \end{bmatrix}$,

where $\mathbf{a}_i^T$ is i^{th} row of a matrix $\mathbf{A}$ and vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_c \end{bmatrix} = \{x_j\}_{c \times 1}$.

Now multiplication $\mathbf{Ax}$ gives

$$\mathbf{Ax} = \begin{bmatrix} \mathbf{a}_1^T\mathbf{x} \\ \mathbf{a}_2^T\mathbf{x} \\ \vdots \\ \mathbf{a}_r^T\mathbf{x} \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{c} a_{1j} \cdot x_j \\ \sum_{j=1}^{c} a_{2j} \cdot x_j \\ \vdots \\ \sum_{j=1}^{c} a_{rj} \cdot x_j \end{bmatrix} = \left\{ \sum_{j=1}^{c} a_{ij} x_j \right\}_{r \times 1}$$

Example: Let $\mathbf{A} = \begin{bmatrix} -1 & 2 \\ 0 & 3 \\ 2 & -1 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$.

Then multiplication $\mathbf{Ax}$ gives

$$\mathbf{Ax} = \begin{bmatrix} -1 & 2 \\ 0 & 3 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \cdot 4 + 2 \cdot 2 \\ 0 \cdot 4 + 3 \cdot 2 \\ 2 \cdot 4 + (-1) \cdot 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 6 \end{bmatrix}$$

2.6 Outer product

The outer product is a multiplication of a column vector by a row vector. Opposite to inner product that results scalar, outer product is a matrix.

Let $\mathbf{a}' = \begin{bmatrix} a_1 & a_2 & \dots & a_m \end{bmatrix}$ and $\mathbf{b}' = \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix}$ then the outer product between vectors $\mathbf{a}$ and $\mathbf{b}$ is defined by

$$\mathbf{ab}' = \begin{bmatrix} a_1 \\ a_2 \\ \ddots \\ a_m \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix} = \begin{bmatrix} a_1b_1 & a_1b_2 & ... & a_1b_n \\ a_2b_1 & a_2b_2 & ... & a_2b_n \\ \vdots & \vdots & ... & \vdots \\ a_mb_1 & a_mb_2 & ... & a_mb_n \end{bmatrix}$$

The element at the row i and column j of matrix $\mathbf{ab}'$ is a_ib_j. When $\mathbf{a}$ is a m dimensional vector and $\mathbf{b}$ is a n dimensional vector, outer product is a $m \times n$ matrix.

Example: For the vectors $\mathbf{a} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}$ and $\mathbf{b}' = \begin{bmatrix} -3 & 5 & 0 & 1 \end{bmatrix}$ the outer product is

$$\mathbf{ab}' = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix} \begin{bmatrix} -3 & 5 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 \cdot (-3) & 1 \cdot 5 & 1 \cdot 0 & 1 \cdot 1 \\ 0 \cdot (-3) & 0 \cdot 5 & 0 \cdot 0 & 0 \cdot 1 \\ -2 \cdot (-3) & -2 \cdot 5 & -2 \cdot 0 & -2 \cdot 1 \end{bmatrix} = \begin{bmatrix} -3 & 5 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 6 & -10 & 0 & -2 \end{bmatrix}$$

Note: The inner product is the trace of the outer product. The trace is the sum of the diagonal entries (see section 2.8).

```
# R example of outer product
# Define vectors a and b
a <- c(1,0,-2)
b <- c(-3,5,0,1)
# R will interpret vectors a and b as either a row or a column vector according
# to the context
# If we try to use operator %*% to vectors R interprets that we want to do
# a inner product

# Outer product can be computed using the function outer()
outer(a,b)
```

```
##      [,1] [,2] [,3] [,4]
## [1,]   -3    5    0    1
## [2,]    0    0    0    0
## [3,]    6  -10    0   -2
```

```
# or with function %o%
a%o%b
```

```
##      [,1] [,2] [,3] [,4]
## [1,]   -3    5    0    1
## [2,]    0    0    0    0
## [3,]    6  -10    0   -2
```

```
# We can define vectors using command matrix to avoid ambiguity that can be sometimes
# caused by operator c() since vector defined using operator c() does not have the
# dimension
# Define vectors using function matrix()
a1=matrix(c(1,0,-2),ncol=1)
b1=matrix(c(-3,5,0,1),nrow=1)
# Now we can use %*% operator to calculate the outer product
a1%*%b1
```

```
##      [,1] [,2] [,3] [,4]
## [1,]   -3    5    0    1
## [2,]    0    0    0    0
## [3,]    6  -10    0   -2
```

2.7 Matrix multiplication

In this section we consider the operations involved in computing the product of two matrices $\mathbf{A}_{r\times c}$ and $\mathbf{B}_{c\times s}$. A product of matrices $\mathbf{A}$ and $\mathbf{B}$ can be written as

$$\mathbf{C}_{r\times s} = \mathbf{A}_{r\times c}\mathbf{B}_{c\times s} = \left(\sum_{k=1}^{n} a_{ik}b_{kj}\right), \text{ for i=1,...,r and j=1,...,s.}$$

The element in row i and column j of the multiplication $\mathbf{AB}$ is calculated as the inner product of the i^{th} row of $\mathbf{A}$ and the j^{th} column of $\mathbf{B}$.

Example: If matrix $\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 4 & 3 \end{bmatrix}$ is a 2×3 matrix and $\mathbf{B} = \begin{bmatrix} 0 & 6 & 1 & 5 \\ 1 & 1 & 0 & 7 \\ 3 & 4 & 4 & 3 \end{bmatrix}$ is a 4×4 matrix then multiplication $\mathbf{AB}$ is

$$
\begin{aligned}
\mathbf{AB} &= \begin{bmatrix} 1 & 0 & 2 \\ -1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 0 & 6 & 1 & 5 \\ 1 & 1 & 0 & 7 \\ 3 & 4 & 4 & 3 \end{bmatrix} \\
&= \begin{bmatrix} 1 \cdot 0 + 0 \cdot 1 + 2 \cdot 3 & 1 \cdot 6 + 0 \cdot 1 + 2 \cdot 4 & 1 \cdot 1 + 0 \cdot 0 + 2 \cdot 4 & 1 \cdot 5 + 0 \cdot 7 + 2 \cdot 3 \\ -1 \cdot 0 + 4 \cdot 1 + 3 \cdot 3 & -1 \cdot 6 + 4 \cdot 1 + 3 \cdot 4 & -1 \cdot 1 + 4 \cdot 0 + 3 \cdot 4 & -1 \cdot 5 + 4 \cdot 7 + 3 \cdot 3 \end{bmatrix} \\
&= \begin{bmatrix} 6 & 14 & 9 & 11 \\ 13 & 10 & 11 & 32 \end{bmatrix}
\end{aligned}
$$

For example the element at the first row and the first column is given by the inner product

$$
\begin{bmatrix} 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix} = 1 \cdot 0 + 0 \cdot 1 + 2 \cdot 3 = 6
$$

and the element at row 2 and column 3 is given by the inner product

$$
\begin{bmatrix} -1 & 4 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix} = -1 \cdot 1 + 4 \cdot 0 + 3 \cdot 4 = 11.
$$

```r
# R example of matrix multiplication
# Define matrices A and B
A <- matrix(c(1,0,2,-1,4,3),byrow=T,ncol=3)
B <- matrix(c(0,6,1,5,1,1,0,7,3,4,4,3),byrow=T,ncol=4)
# Matrix multiplication AB
# Assign product to matrix C
C <- A%*%B
# Print the matrix C
C
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    6   14    9   11
## [2,]   13   10   11   32
```

```r
# Dimension of the matrix C
dim(C)
```

```
## [1] 2 4
```

```r
# Multiplication B%*%A is not defined
```

Note: If matrix $\mathbf{A}$ is of dimension $r \times c$ and matrix $\mathbf{B}$ is of dimension $c \times s$ then the multiplication $\mathbf{AB}$ is a $r \times s$ matrix.

$$\underset{r \times s}{\overset{\mathbf{C}}{\begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}}} = \underset{r \times c}{\overset{\mathbf{A}}{\begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}}}\underset{c \times s}{\overset{\mathbf{B}}{\begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix}}}$$

That is, the number of rows of $\mathbf{AB}$ equals to the number of rows of $\mathbf{A}$ and the number of columns of $\mathbf{AB}$ equals to the number of columns of $\mathbf{B}$.

Two matrices can be multiplied if the number of columns of the first matrix equals the number of rows of the second matrix. So if we can form product $\mathbf{AB}$ it does not mean that it is possible to form the product $\mathbf{BA}$.

Example: If matrix $\mathbf{X}$ is $n \times m$ matrix and matrix $\mathbf{Y}$ is $m \times q$ matrix, then the product $\mathbf{XY}$ is defined since the number of columns of $\mathbf{X}$ equals to number of rows of $\mathbf{Y}$. However multiplication $\mathbf{YX}$ is defined only if the number of columns of matrix $\mathbf{Y}$ equal to the number of rows of matrix $\mathbf{X}$ i.e. multiplication is possible if $q = n$.

Note: It is allways possible to multiply two square matrices of the same dimension in any order.

In the product $\mathbf{AB}$ it is said that matrix $\mathbf{B}$ is premultiplied by matrix $\mathbf{A}$ and $\mathbf{A}$ is postmultiplied by $\mathbf{B}$. This distinction on which matrix comes first in multiplication is important since even if both multiplications ($\mathbf{AB}$ and $\mathbf{BA}$) exists, they are generally not the same i.e. often $\mathbf{AB} \neq \mathbf{BA}$.

Example: If $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix}$ then matrices are comformable. Now

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1 \cdot 5 + 2 \cdot 7 & 1 \cdot 6 + 2 \cdot 8 \\ 3 \cdot 5 + 4 \cdot 7 & 3 \cdot 6 + 4 \cdot 8 \end{bmatrix} = \begin{bmatrix} 5 + 14 & 6 + 16 \\ 15 + 28 & 18 + 32 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$$

and

$$\mathbf{BA} = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 5 \cdot 1 + 6 \cdot 3 & 5 \cdot 2 + 6 \cdot 4 \\ 7 \cdot 1 + 8 \cdot 3 & 7 \cdot 2 + 8 \cdot 4 \end{bmatrix} = \begin{bmatrix} 5 + 18 & 10 + 24 \\ 7 + 24 & 14 + 32 \end{bmatrix} = \begin{bmatrix} 23 & 34 \\ 31 & 46 \end{bmatrix}$$

We notice that $\mathbf{AB} \neq \mathbf{BA}$.

```r
# R example of matrix multiplication
# Matrices A and B
A <- matrix(c(1,2,3,4),byrow=T,ncol=2)
B <- matrix(c(5,6,7,8),byrow=T,ncol=2)
C1 <- A%*%B
C1
```

```
##      [,1] [,2]
## [1,]   19   22
## [2,]   43   50
```

```r
C2 <- B%*%A
C2
```

```
##      [,1] [,2]
## [1,]   23   34
## [2,]   31   46
```

```r
# Multiplications AB and BA are not equal like we noticed
identical(C1,C2)
```

```
## [1] FALSE
```

Note: In general for matricess $\mathbf{A}$ and $\mathbf{B}$ matrix multiplication $\mathbf{AB}$ may but need not equal $\mathbf{BA}$ even if matrices are conformable. That is matrix multiplication is not commutative.

Note: Matrices $\mathbf{A}$ and $\mathbf{B}$ are said to *commute* if $\mathbf{AB} = \mathbf{BA}$.

When more than two matrices are multiplied all multiplications must be conformable that multiplication is defined. For example product of three matrices $\mathbf{ABC}$ requires that matrices $\mathbf{A}$ and $\mathbf{B}$ are conformable for multiplication $\mathbf{AB}$ and that also matrices $\mathbf{B}$ and $\mathbf{C}$ are conformable for multiplication $\mathbf{BC}$. The product $\mathbf{ABC}$ can be optained by computing $\mathbf{AB}$ and then post multiplying the reslut of $\mathbf{AB}$ by $\mathbf{C}$ or by computing $\mathbf{BC}$ first and then pre multiplying $\mathbf{BC}$ by $\mathbf{A}$.

$$\underset{r \times t}{\overset{\mathbf{D}}{\begin{bmatrix} \quad \end{bmatrix}}} = \underset{r \times c}{\overset{\mathbf{A}}{\begin{bmatrix} \quad \end{bmatrix}}}\underset{c \times s}{\overset{\mathbf{B}}{\begin{bmatrix} \quad \end{bmatrix}}}\underset{s \times t}{\overset{\mathbf{C}}{\begin{bmatrix} \quad \end{bmatrix}}}$$

Transpose of the product of the two matrices is the product of the transposes of each matrix in reverse order i.e.

The transpose of the matrix multiplication $\mathbf{AB}$ is $\mathbf{B}'\mathbf{A}'$ and the transpose of the matrix multiplication $\mathbf{ABC}$ is $\mathbf{C}'\mathbf{B}'\mathbf{A}'$

Note: If transpose $\mathbf{A}'$ is $n \times c$ matrix and transpose $\mathbf{B}'$ is $r \times n$ matrix then multiplication $\mathbf{A}'\mathbf{B}'$ is defined only if $c = r$ but multiplication $\mathbf{B}'\mathbf{A}'$ is defined.

Example (Markov chain) As an example application of matrix multiplication lets consider voting behavior. Assume that voters are distributed between three parties Democrats (D), Republicans (R) and Liberals (L).

In each election each voter can vote the same party they voted in the previous elections or vote other party. This process can be described by a transition matrix $\mathbf{P}$ that contain the probabilities of particular transitions. In transition matrix the entries of each column vector are positive and their sum is 1. Let's assume that Voters transition matrix is

$$\mathbf{P} = \begin{array}{c} \begin{array}{ccc} D & R & L \end{array} \\ \begin{bmatrix} 0.70 & 0.10 & 0.30 \\ 0.20 & 0.80 & 0.30 \\ 0.10 & 0.10 & 0.40 \end{bmatrix} \end{array} \begin{array}{c} \\ D \\ R \\ L \end{array} .$$

From the first column of the transition matrix we can read that in an upcoming elections of those who voted Democrats in the previous election, 70% will vote Democrats again, 20% will vote Republicans and 10% will vote Liberals

Second column tells the transition probabilities for those who voted Republicans. That is, 10% of those who voted Republicans in the previous election will vote Democrats, 80% will vote Republicans again, 10% will vote Liberals

Similarly third column of transition matrix contains the transition probabilities for Liberals

Assume that in the previous elections voters were distributed such that 55% voted democrats, 40% voted Republicans and 5% voted Liberals. Those percentages form a initial state vector $\mathbf{v}_0$ that is

$$\mathbf{v}_0 = \begin{bmatrix} 0.55 \\ 0.40 \\ 0.05 \end{bmatrix}.$$

Now predicted result of the next elections can be calculated using transition matrix $\mathbf{P}$ and the initial state vector $\mathbf{V}_0$ as

$$\mathbf{P}\mathbf{v}_1 = \begin{bmatrix} 0.70 & 0.10 & 0.30 \\ 0.20 & 0.80 & 0.30 \\ 0.10 & 0.10 & 0.40 \end{bmatrix} \begin{bmatrix} 0.55 \\ 0.40 \\ 0.05 \end{bmatrix} = \begin{bmatrix} 0.44 \\ 0.445 \\ 0.115 \end{bmatrix}.$$

That is we predict that in the next elections 44% of the voters will vote Democrats, 44.5% will vote Republicans and 11.5% will vote Liberals. In this model we assume that voters decision of which party they vote depends only about their voting decision in the previous elections i.e. transition probabilities doesn't change and the voting behaviour is memoryless.

Vector $\mathbf{v}_1$ is called a state vector. We can use it to calculate prediction for the result for the elections following the next elections. Prediction is

$$\mathbf{v}_2 = \mathbf{P}\mathbf{v}_1 = \begin{bmatrix} 0.70 & 0.10 & 0.30 \\ 0.20 & 0.80 & 0.30 \\ 0.10 & 0.10 & 0.40 \end{bmatrix} \begin{bmatrix} 0.44 \\ 0.445 \\ 0.115 \end{bmatrix} = \begin{bmatrix} 0.3870 \\ 0.4785 \\ 0.1345 \end{bmatrix}$$

Of particular interest is a vector $\mathbf{v}$ such that $\mathbf{P}\mathbf{v} = \mathbf{v}$. That is, if we multiply transition matrix by a vector $\mathbf{v}$ and get out the vector having exactly the same elements as $\mathbf{v}$, we have found a steady state vector $\mathbf{v}$.

In the election example steady state vector shows where party popularity will settle in a long term.

We can solve that $\mathbf{v}_{20} = \begin{bmatrix} 0.3214 \\ 0.5357 \\ 0.1429 \end{bmatrix}$ and $\mathbf{v}_{30} = \begin{bmatrix} 0.3214 \\ 0.5357 \\ 0.1429 \end{bmatrix}$

So in this example steady state vector is

$$\mathbf{v} = \begin{bmatrix} 0.3214 \\ 0.5357 \\ 0.1429 \end{bmatrix}.$$

That is in the long run Democrats support will settle to 38.7%, Republican support will settle to 47.85% and Liberals support will settle to 13.45%. Steady state could be solved using eigenvalue analysis (See Chapter 8). Notice that in this example transition matrix was imaginary and in reality models for predicting election results are much more complex.

The model we used is known as *a discrete time Markov chain.* Model undergoes transitions from one state to the another, with the probability distribution of the next state depending only of the current state. That is if the transition matrix is $\mathbf{P}$ and the initial state vector is $\mathbf{v}_0$ then

$\mathbf{v}_n = \mathbf{P}^n \mathbf{v}_0$.

Notice that we could calculate state vector iteratively or use matrix power. For example in the election example

$$\mathbf{v}_2 = \begin{bmatrix} 0.70 & 0.10 & 0.30 \\ 0.20 & 0.80 & 0.30 \\ 0.10 & 0.10 & 0.40 \end{bmatrix}^2 \begin{bmatrix} 0.55 \\ 0.40 \\ 0.05 \end{bmatrix} = \begin{bmatrix} 0.54 & 0.18 & 0.36 \\ 0.33 & 0.69 & 0.42 \\ 0.13 & 0.13 & 0.22 \end{bmatrix} \begin{bmatrix} 0.55 \\ 0.40 \\ 0.05 \end{bmatrix} = \begin{bmatrix} 0.3870 \\ 0.4785 \\ 0.1345 \end{bmatrix}$$

```r
# R example of the Markov chain
# Define the transition matrix P (this time elements are filled  column-wise)
P <- matrix(c(0.7,0.2,0.1,0.1,0.8,0.1,0.3,0.3,0.4),ncol=3)
# Define initial state vector
#v0 <- c(0.55,0.4,0.05)
# Calculate prediction of the first elections
#v1 <- P%*%v0
#v1
# Compute the state vector v2
#v2 <- P%*%V1
#v2

# Matrix powers can be calculated using operator %^% in library(expm)
# Install package expm if needed
# install.packages("expm")
# Load package expm
# library(expm)

# Calculate the second power of the transition matrix P
# P%^%2
# This is the same as PP
P%*%P
```

```
##      [,1] [,2] [,3]
## [1,] 0.54 0.18 0.36
## [2,] 0.33 0.69 0.42
## [3,] 0.13 0.13 0.22
```

```r
# The twentieth power
# P%^%20
# Prediction of the result of the twentieth upcoming elections
# P%^%20%*%v20
```

2.7.1 Comparison to scalar algebra

There are some differences between matrix algebra and scalar algebra that are good to keep in mind.

a) If x and y are real numbers and $xy = 0$, then $x = 0$ or $y = 0$. For matrices this doesn't hold generally. If $\mathbf{A}$ and $\mathbf{B}$ are matrices and $\mathbf{AB} = \mathbf{0}$ then it isn't mandatory that one of the matrices is a zero matrix.

Example If $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$, then

$$\mathbf{AB} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{0}$$

b) If scalar x is a real number (i.e. $x \in \mathbb{R}$) and $x^2 = 0$, then for scalars it holds that $x = 0$.

However if $\mathbf{A}$ is a matrix and $\mathbf{A}^2 = \mathbf{0}$, then $\mathbf{A}$ is not necessarily zero matrix.

Example If $\mathbf{A} = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \\ -1 & -2 & -5 \end{bmatrix}$, then $\mathbf{A}^2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

c) If scalar y and $y^2 = 1$, then it implies that $y = 1$ or $y = -1$

For matrices if $\mathbf{B}$ is a matrix and $\mathbf{B}^2 = \mathbf{I}$, then $\mathbf{B}$ is not necessarily identity matrix $\mathbf{I}$ or negative of identity matrix $-\mathbf{I}$.

Example

If $\mathbf{B} = \begin{bmatrix} 1 & 0 \\ 4 & -1 \end{bmatrix}$, then $\mathbf{B}^2 = \mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

d) If x, y and z are real numbers and $xy = xz$ then $y = z$.

For matrices, $\mathbf{BC} = \mathbf{AC}$ doesn't necessarily mean that $\mathbf{A} = \mathbf{B}$.

Example

Let $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$ and $\mathbf{C} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$. Now $\mathbf{AC} = \mathbf{BC} = \mathbf{0}$ even though $\mathbf{A} \neq \mathbf{B}$.

2.8 The trace of a matrix

The sum of the main diagonal elements of the square matrix $\mathbf{A}$ is called trace of the matrix. If matrix $\mathbf{A}$ is a $n \times n$ square matrix then the trace denoted as $tr(\mathbf{A})$ is defined by

$tr(\mathbf{A}) = a_{11} + a_{22} + a_{33} + ...a_{nn} = \sum_{i=1}^{n} a_{ii}$

Example: If $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}$, then

$$tr\left(\mathbf{A}\right) = tr\left(\begin{bmatrix} \boxed{1} & 2 & 3 \\ 4 & \boxed{5} & 6 \\ 7 & 8 & \boxed{9} \end{bmatrix}\right) = 1 + 5 + 9 = 15$$

Note: The trace of a scalar is the scalar itself since scalar can be treated as a 1×1 matrix. If λ is a scalar then $tr(\lambda) = \lambda$.

2.8.1 Properties of the trace

1) Linear property: Trace of the sum is the sum of the traces i.e. $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B})$
2) Trace of a scalar multiplication: If k is a scalar then $tr(k\mathbf{A}) = k \cdot tr(\mathbf{A})$
3) Trace of a matrix is equal to the trace of its transpose: $tr(\mathbf{A}) = tr(\mathbf{A}')$

Example: The trace of the transpose

$$tr\left(\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix}'\right) = tr\left(\begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}\right) = 1 + 5 + 9 = 15$$

```r
# R example of computing the trace
# Define matrix A
A <- matrix(c(1:9),ncol=3,byrow=TRUE)
# The trace can be computed as a sum of the diagonal elements
# The diagonal elements of A are
diag(A)
```

```
## [1] 1 5 9
```

```r
# compute the sum of the diagonal elements
sum(diag(A))
```

```
## [1] 15
```

```r
# Before you compute the trace using this method check that the matrix is
# square since this will compute sum of diagonal elements even though matrix
# is not square. For example
```

```r
B <- matrix(c(1:12),ncol=3,byrow=TRUE)
B
```

```
##      [,1] [,2] [,3]
## [1,]    1    2    3
## [2,]    4    5    6
## [3,]    7    8    9
## [4,]   10   11   12
```

```r
sum(diag(A))
```

```
## [1] 15
```

2.9 Exercises:

Exercise 1: Consider matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 3 & 4 & 4 \\ 7 & 9 & 8 & -2 & 3 \\ 1 & -3 & 4 & -6 & -7 \\ 3 & 5 & 2 & 1 & 5 \end{bmatrix}$$

a) What is the dimension of the matrix?

b) What is the entry at a_{43}?

c) What is the trace of the matrix

$$\mathbf{C} = \begin{bmatrix} 2 & 6 & 2 & 4 \\ 1 & -1 & 9 & 0 \\ 0 & 3 & 3 & -4 \\ -2 & -4 & 2 & 0 \end{bmatrix}?$$

Exercise 2: Find the transpose of the following matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & -1 \\ 3 & 0 & 2 \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} -2 & 1 & 0 & 4 \end{bmatrix}$$

Exercise 3: Consider vectors

$$\mathbf{x} = \begin{bmatrix} 1 \\ -2 \\ 2 \\ 3 \\ 0 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} -1 \\ 2 \\ 3 \\ 4 \\ 1 \end{bmatrix}$$

Compute the inner product $\mathbf{x}'\mathbf{y}$ by hand and using R.

Exercise 4: Consider vectors

$$\mathbf{x} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

Compte the outer product $\mathbf{x}\mathbf{y}'$ by hand and using R.

Exercise 5: Consider $\mathbf{X} = \begin{bmatrix} 1 & 3 & -1 \\ 3 & 0 & 2 \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 1 & -1 \end{bmatrix}$

Compute $\mathbf{XY}$ by hand and using R.

Exercise 6: Consider matrices $\mathbf{A}_{3\times 4}$, $\mathbf{B}_{2\times 4}$, $\mathbf{C}_{3\times 4}$ and $\mathbf{D}_{4\times 2}$

Find if the multiplications are defined and define the dimension of the multiplication result if it is defined$ a) $\mathbf{ADB}$ b) $\mathbf{DC}$ c) $\mathbf{A}'\mathbf{AD}$

Exercise 7: Consider $n \times n$ matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. What is the transpose of the matrix multiplication $\mathbf{A}'\mathbf{B}\mathbf{C}'$?

Exercise 8: Consider $n \times n$ matrices $\mathbf{X}$ and $\mathbf{Y}$. What is the transpose of the matrix multiplication $\mathbf{X}'\mathbf{XYX}'\mathbf{X}$?

Exercise 9: If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices. When does it hold that$(\mathbf{A}+\mathbf{B})(\mathbf{A}+\mathbf{B}) = \mathbf{A}^2 - \mathbf{B}^2$?

3 Special types of matrices

Recall that a square matrix is a matrix with an equal number of rows and columns.

Example: Let

$$\mathbf{A} = \begin{bmatrix} -1 & 2 \\ 3 & 4 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 3 & 2 & 5 \\ 2 & 0 & 7 \\ 1 & 0 & 4 \end{bmatrix}$$

Matrix $\mathbf{A}$ is a 2×2 square matrix and matrix $\mathbf{D}$ is a 3×3 square matrix.

Let $\mathbf{A} = \{a_{ij}\}_{n\times n}$ be a square matrix. Now elements a_{11}, a_{22},...,a_{nn} of the matrix $\mathbf{A}$ are called diagonal elements and all the other elements are called off-diagonal elements.

Diagonal matrix is a matrix with the only non-zero elements on the diagonal.

3.1 Symmetric matrices

Matrix $\mathbf{A}$ is symmetric if matrix is equal to its transpose i.e.

$\mathbf{A} = \mathbf{A}'$

Note: Even though matrices $\mathbf{A}$ and $\mathbf{B}$ are symmetric, multiplication $\mathbf{AB}$ isn't necessarily symmetric because $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}' = \mathbf{BA}$.

Square matrix $\mathbf{X}$ is skew-symmetric if $\mathbf{X}' = -\mathbf{X}$ so $x_{ij} = -x_{ji}$ for all i and j. From the definition we can say that all diagonal elements of skew-symmetric matrix are zero.

Example: For symmetric matrices $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 3 & 7 \\ 7 & 6 \end{bmatrix}$ matrix multiplication

$\mathbf{AB} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 3 & 7 \\ 7 & 6 \end{bmatrix} = \begin{bmatrix} 17 & 19 \\ 27 & 32 \end{bmatrix}$ is not symmetric.

Note: Matrices $\mathbf{AA}'$ and $\mathbf{A}'\mathbf{A}$ are always symmetric but usually $\mathbf{AA}' \neq \mathbf{A}'\mathbf{A}$

Note: Matrix $\mathbf{B} = \{b_{ij}\}_{n\times n}$ is skew symmetric if $\mathbf{B}' = -\mathbf{B}$.

Matrix multiplciation $\mathbf{AA}'$ is always defined, resulting symmetric $m \times m$ matrix if $\mathbf{A}$ is $m \times n$ matrix. Similarly multiplication $\mathbf{A}'\mathbf{A}$ is defined resulting symmetric $n \times n$ matrix. In statistics multiplication $\mathbf{A}'\mathbf{A}$ is used for example in linear regression analysis when we solve the least squares estimates for the coefficients.

3.2 Orthogonal and orthonormal vectors and matrices

3.2.1 Vector norm and unit vector

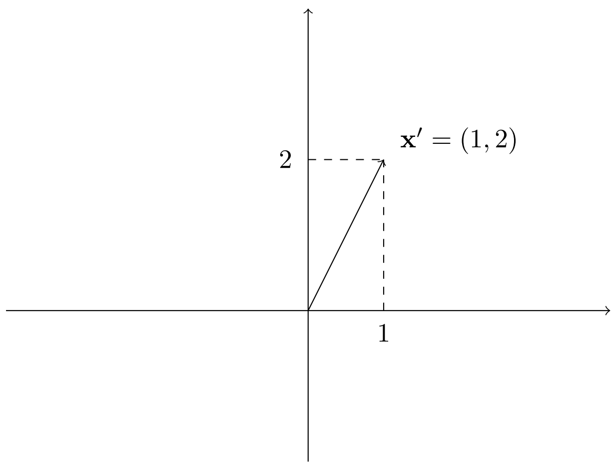
Vector norm (length, magnitude) is the most commonly defined as its Euclidean length. The norm of the vector $\mathbf{x}' = (x_1, ..., x_n)$ is

$||\mathbf{x}|| = \sqrt{\mathbf{x}'\mathbf{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}.$

Notice that the norm of the vector is denoted by two vertical bars || and this should not be confused to determinant which will be later denoted by a single vertical bars.

Example: We calculate the length of the vector $\mathbf{x}' = \begin{bmatrix} 1 & 2 \end{bmatrix}$

First we will draw a plot of this vector.



Norm of the vector $\mathbf{x}$ is

$||\mathbf{x}|| = \sqrt{1^2 + 2^2} = \sqrt{5}$

That is the length (magnitude) of the vector $\mathbf{x}$ is $\sqrt{5}$.

```r
# R example of calculating the length of the vector
# Define  vector x
x <- c(1,2)
# compute the inner product
x%*%x
```

```
##      [,1]
## [1,]    5
```

```r
# function sqrt() is used to take the square root of the inner product
sqrt(5)
```

```
## [1] 2.236068
```

```r
# or
sqrt(x%*%x)
```

```
##          [,1]
## [1,] 2.236068
```

```
# This is the length of the vector x
```

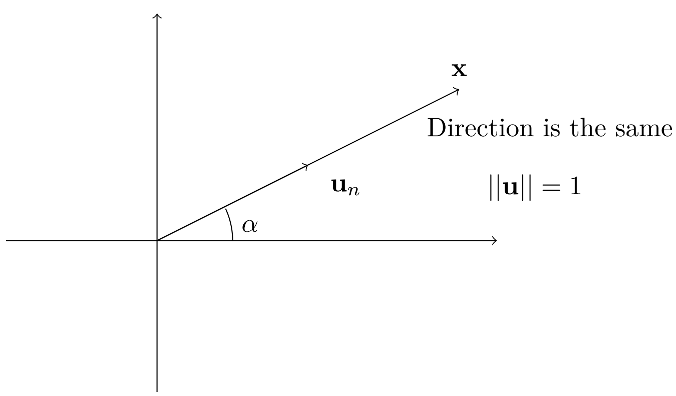
Note: Vector with norm 1 is called a unit vector.

Note: Every vector that has non-zero length ($\mathbf{x}' \neq \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$) can be transformed to unit vector. In other words norm of the vector

$$\mathbf{u} = \frac{\mathbf{x}}{||\mathbf{x}||} = \frac{\mathbf{x}}{\sqrt{\mathbf{x}'\mathbf{x}}}$$

is 1 (this is called normalized form of a vector $\mathbf{x}$), since $\mathbf{u}'\mathbf{u} = 1$.

When we change vector $\mathbf{x}$ to unit vector we change only the magnitude and not the direction (slope in the graph). If constant $c = \sqrt{\mathbf{x}'\mathbf{x}}$ then $\mathbf{x} = c \times \mathbf{u}$.



3.3 Ortogonal vectors

Let vector $\mathbf{x} \neq \mathbf{0}$ and vector $\mathbf{y} \neq \mathbf{0}$ i.e. $\mathbf{x}$ and $\mathbf{y}$ are non-zero vectors. Vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal, if $\mathbf{x}'\mathbf{y} = 0$. That is vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if inner product between the vectors is zero. That is vectors are orthogonal if they form a right angle.

Example: Lets define vectors $\mathbf{x}$ and $\mathbf{y}$ as

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \text{ and } \mathbf{x} = \begin{bmatrix} 7 \\ -4 \\ 3 \\ -2 \end{bmatrix}.$$

Vectors are orthogonal since

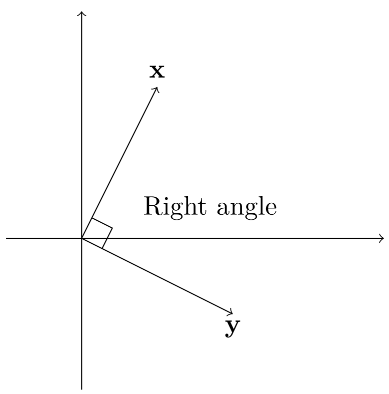
$$\mathbf{x}'\mathbf{y} = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 7 \\ -4 \\ 3 \\ -2 \end{bmatrix} = 7 - 8 + 9 - 8 = 0.$$

Example: Let
$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$.

Now inner product between the vectors $\mathbf{x}$ and $\mathbf{y}$ is

$\mathbf{x}'\mathbf{y} = 0$

This means that vectors form a right angle.



Note: Vectors $\mathbf{u}$ and $\mathbf{v}$ are called orthonormal, if vectors are unit vectors ($||\mathbf{u}|| = ||\mathbf{v}|| = 1$) and vectors are also orthogonal (i.e. inner product between the vectors is zero $\mathbf{u}'\mathbf{v} = 0$).

Example: Consider vectors

$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$

Next we will form normalized vectors $\mathbf{u}$ and $\mathbf{v}$.

Normalized vector of $\mathbf{x}$ is a vector in the same direction but with norm 1. Normalized vector is denoted as $\mathbf{u} = \dfrac{\mathbf{x}}{||\mathbf{x}||}$

$$\mathbf{u} = \frac{\mathbf{x}}{||\mathbf{x}||} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$$

and normalized vector of $\mathbf{y}$ is called as $\mathbf{v}$

$$\mathbf{v} = \frac{\mathbf{y}}{||\mathbf{y}||} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix} = \begin{bmatrix} \frac{2}{\sqrt{5}} \\ \frac{-1}{\sqrt{5}} \end{bmatrix}$$

Now vectors $\mathbf{u}$ and $\mathbf{v}$ are also orthogonal since $\mathbf{u}'\mathbf{v} = 0$. This is obvious since we noriced in a previous exmple that vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal and that normalization does not change direction. Since vectors $\mathbf{u}$ and $\mathbf{v}$ are unit vectors and orthogonal they are orthonormal.

```
# R example of orthonormal vector
# Define  vector x
x <- c(1,2)
# compute the length of the vector x
sqrt(x%*%x)
```

```
##          [,1]
## [1,] 2.236068
```

```
# Create normalized vector of x that is called as u
u <- x/sqrt(x%*%x)
```

```
## Warning in x/sqrt(x %*% x): Recycling array of length 1 in vector-array arithmetic is deprecated.
##   Use c() or as.vector() instead.
```

```
u
```

```
## [1] 0.4472136 0.8944272
```

```
# Define  vector y
y <- c(2,-1)
# Create normalized vector of y that is called as v
v <- y/sqrt(y%*%y)
```

```
## Warning in y/sqrt(y %*% y): Recycling array of length 1 in vector-array arithmetic is deprecated.
##   Use c() or as.vector() instead.
```

```
v
```

```
## [1]  0.8944272 -0.4472136
```

```
# Vectors u and v are orthonormal if inner product between the vectors is zero.
u%*%v
```

```
##      [,1]
## [1,]    0
```

Note: The set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p\}$ is called orthogonal set of vectors if all possible pairs of vectors are orthogonal. In other words

$\mathbf{u}_i'\mathbf{u}_j = 0 \;\; \forall \; i \neq j.$

Example: In this example we investigate if the set of vectors is orthogonal.

Consider the set of three vectors $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$, where

$$\mathbf{u}_1^T = \begin{bmatrix} 3 & 1 & 1 \end{bmatrix},$$

$$\mathbf{u}_2^T = \begin{bmatrix} -1 & 2 & 1 \end{bmatrix}$$

and

$$\mathbf{u}_3^T = \begin{bmatrix} -1/2 & -2 & 7/2 \end{bmatrix}$$

We investigate if the vector pairs are orthogonal.

Are vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ orthogonal?

$\mathbf{u}_1^T\mathbf{u}_2 = 3 \cdot (-1) + 1 \cdot 2 + 1 \cdot 1 = 0$
Yes they are, since inner product between the vectors is zero.

Pair $\mathbf{u}_1$ and $\mathbf{u}_3$ is also orthogonal since

$\mathbf{u}_1^T\mathbf{u}_3 = 3 \cdot (-1/2) + 1 \cdot (-2) + 1 \cdot 7/2 = 0.$
and the vectors $\mathbf{u}_2$ and $\mathbf{u}_3$ are also orthogonal since

$\mathbf{u}_2^T\mathbf{u}_3 = -1 \cdot (-1/2) + 2 \cdot (-2) + 1 \cdot 7/2 = 0.$

We confirmed that all possible pairs of vectors are orthogonal so the set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is orthogonal.

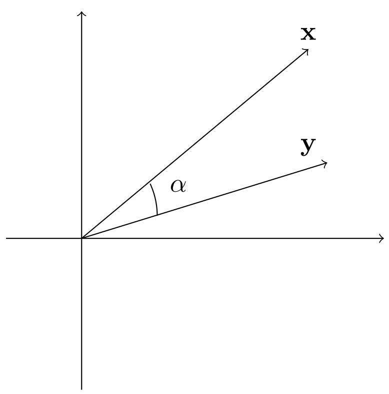
Note: The set of vectors is orthonormal if the set of vectors is orthogonal and all vectors are unit vectors. To be more precise the set of vectors $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_p\}$ is orthonormal if

$$\mathbf{u}_i^T\mathbf{u}_j = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

Note: It holds for the angle α between the vectors $\mathbf{x}$ and $\mathbf{y}$ that

$$\cos\alpha = \frac{\mathbf{x}'\mathbf{y}}{||\mathbf{x}||\,||\mathbf{y}||}$$

That is the inner product is directly related to angle between vectors and the lengths of the vectors. The angle between the vectors $\mathbf{x}$ and $\mathbf{y}$ is expressed in radians.



3.3.1 Orthogonal matrices

A square matrix $\mathbf{A}$ is said to be orthogonal if $\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}' = \mathbf{I}$. That is, the matrix multiplied by its transpose yields an identity matrix.

Note: Equation $\mathbf{A}\mathbf{A}' = \mathbf{A}'\mathbf{A} = \mathbf{I}$ implies that the matrix is orthogonal if it is square and has orthonormal rows and orthonormal columns.

The product of two orthogonal $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ is also orthogonal since $(\mathbf{AB})'\mathbf{AB} = \mathbf{B}'\mathbf{A}'\mathbf{AB} = \mathbf{B}'\mathbf{IB} = \mathbf{B}'\mathbf{B} = \mathbf{I}$ because $\mathbf{A}'\mathbf{A} = \mathbf{I}$ and $\mathbf{B}'\mathbf{B} = \mathbf{I}$

Note: In some sources the multiplication of the two equal size matrices $\mathbf{X}$ and $\mathbf{Y}$ are sometimes said to be orthogonal if $\mathbf{XY} = \mathbf{0}$ i.e. multiplication is a zero matrix. In this case the term orthogonal is used in the same sense as when we say that two vectors are orthogonal.

Note: When term orthogonal matrix is used in this handout, it is meant that the matrix multiplied by its transpose is identity matrix i.e. for example matrix $\mathbf{X}$ if orthogonal if $\mathbf{X}'\mathbf{X} = \mathbf{X}\mathbf{X}' = \mathbf{I}_n$

Example: Consider matrix

$$\mathbf{A} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{2} \\ \sqrt{3} & \sqrt{3} & 0 \\ 1 & 1 & -2 \end{bmatrix}.$$

Matrix matrix $\mathbf{A}$ is orthogonal if product $\mathbf{A}'\mathbf{A} = \mathbf{I}$ or $\mathbf{A}\mathbf{A}' = \mathbf{I}$. It is enough to show that one of the products

$\mathbf{A}'\mathbf{A}$ or $\mathbf{A}\mathbf{A}'$ is unit matrix if $\mathbf{A}$ is square.

We will calculate product $\mathbf{A}'\mathbf{A}$ that is

$$\mathbf{A}'\mathbf{A} = \left(\frac{1}{\sqrt{6}}\right)^2 \begin{bmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{2} \\ \sqrt{3} & \sqrt{3} & 0 \\ 1 & 1 & -2 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \mathbf{I}_3$$

Matrix $\mathbf{A}$ is orthogonal.

Note: The product of orthogonal matrices is also orthogonal.

```r
# R example
# Define  vectors x and y
x <- c(1,3)
y <- c(2,-1)
# compute the length of the vector x
sqrt(x%*%x)
```

```
##          [,1]
## [1,] 3.162278
```

```r
# Create normalized vector of x that is called u
u <- x/sqrt(x%*%x)
```

```
## Warning in x/sqrt(x %*% x): Recycling array of length 1 in vector-array arithmetic is deprecated.
##   Use c() or as.vector() instead.
```

```r
u
```

```
## [1] 0.3162278 0.9486833
```

```r
# Define  vector y

# Create normalized vector of y that is called v
v <- y/sqrt(y%*%y)
```

```
## Warning in y/sqrt(y %*% y): Recycling array of length 1 in vector-array arithmetic is deprecated.
##   Use c() or as.vector() instead.
```

```r
v
```

```
## [1]  0.8944272 -0.4472136
```

```r
# Vectors u and v are orthonormal if inner product between the vectors is zero.
u%*%v
```

```
##             [,1]
## [1,] -0.1414214
```

```
# Compute inner product between the vectors (x and y) and divide that by product of
# the lengths of the vectors
x%*%y/(sqrt(x%*%x)*sqrt(y%*%y))
```

```
##             [,1]
## [1,] -0.1414214
```

```
# The angle between the vectors x and y is expressed in radians
# Solve angle using the inverse cosine (acos in R)
alpha <- acos(x%*%y/(sqrt(x%*%x)*sqrt(y%*%y)))
alpha
```

```
##           [,1]
## [1,] 1.712693
```

```
# Angle in degrees is (degrees=radians*180/pi)
alpha*180/pi
```

```
##          [,1]
## [1,] 98.1301
```

3.4 Idempotent matrices

An important class of matrices are those which has the property that raising them to a power leaves them unchanged. This kind of Matrices are called idempotent. That is the square matrix $\mathbf{X}$ is idempotent if $\mathbf{X}^2 = \mathbf{X}$ i.e. the second power of $\mathbf{X}$ is equal to $\mathbf{X}$. This means that also higher integer powers of idempotent matrix is the matrix itself.

Idempotent matrices are square and they are not singular (see chapter 5). The requirement for idempotent matrice to be square follows from the definition of matrix product, since otherwise multiplication is not defined.

Example: Matrix

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 1 \\ -2 & 3 & -2 \\ -4 & 4 & -3 \end{bmatrix}$$

is idempotent since

$$\mathbf{AA} = \mathbf{A}.$$

```r
# R example of idempotent matrix
# Define matrix A that is idempotent
A <- matrix(c(2,-1,1,-2,3,-2,-4,4,-3),nrow=3,byrow=T)
# Compute multiplication AA
A%*%A
```

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

```r
# Also other powers equal the matrix itself since A is idempotent
# For example AAA
A%*%A%*%A
```

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

Note: If the matrix $\mathbf{K}$ is idempotent i.e. $\mathbf{K}^2 = \mathbf{K}$

Then also the matrix $(\mathbf{I} - \mathbf{K})^2 = \mathbf{I} - \mathbf{K}$ is idempotent, since

$(\mathbf{I} - \mathbf{K})^2 = (\mathbf{I} - \mathbf{K})(\mathbf{I} - \mathbf{K}) = \mathbf{I} - \mathbf{K} - \mathbf{K} + \mathbf{K}^2 = \mathbf{I} - 2\mathbf{K} + \mathbf{K} = \mathbf{I} - \mathbf{K}$

but the matrix $(\mathbf{K} - \mathbf{I})$ is not idempotent.

Note: Idempotent matrices that are symmetric are called projection matrices.

In statistics idempotent matrices are important since idempotent matrices are used as projection matrices. Idempotent matrices are encountered for example in the analysis of linear (or nonlinear) models.

3.5 Matrices having all elements equal

Vector whose each element is 1 is called as a summing vector. Summing vectors can be used to express sums in matrix notation using inner product.

The n-element summing vector is

$$\mathbf{1}'_n = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \end{bmatrix}.$$

Example We can calculate the sum of the elements of vector $\mathbf{x}' = \begin{bmatrix} -1 & 3 & 0 \end{bmatrix}$ using summing vector of order 3 that is $\mathbf{1}'_3 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$

$$\mathbf{1}_3'\mathbf{x} = \begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}-1\\3\\0\end{bmatrix} = 1\cdot -1 + 1\cdot 3 + 1\cdot 0 = 2$$

Or the same in the other order

$$\mathbf{x}'\mathbf{1}_3 = \begin{bmatrix}-1 & 3 & 0\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix} = 2$$

```r
# R example
# Define unit vector having 3 elements. It isn't possible to give variable name
# beginning with number.
# Function rep can be used to replicate given element multiple times
i1 <- rep(1,3) # replicate number one 3 times
i1
```

```
## [1] 1 1 1
```

```r
# Create vector x having elements -1, 3 and 0
x <- c(-1,3,0)
# Compute the sum of the elements of vector x
# easiest way is to use function sum
sum(x)
```

```
## [1] 2
```

```r
# but we could also use unit vector and compute the inner product between i1 and x
i1%*%x
```

```
##      [,1]
## [1,]    2
```

For matrices we can calculate row and column sums using summing vectors.

Example Consider

$$\mathbf{1}_3' = \begin{bmatrix}1 & 1 & 1\end{bmatrix},\ \mathbf{1}_2' = \begin{bmatrix}1 & 1\end{bmatrix} \text{ and } \mathbf{A} = \begin{bmatrix}1 & 4\\2 & 5\\3 & 6\end{bmatrix}$$

Now the column sums are

$$
\mathbf{1}_3'\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 15 \end{bmatrix}
$$

and the row sums are

$$
\mathbf{A}\mathbf{1}_2 = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ 9 \end{bmatrix}.
$$

```r
# R example
# Define unit vector having 3 elements.
i1 <- rep(1,3) # replicate number one 3 times
i1
```

```
## [1] 1 1 1
```

```r
# Create matrix A
A <- matrix(1:6,ncol=2)
A
```

```
##      [,1] [,2]
## [1,]    1    4
## [2,]    2    5
## [3,]    3    6
```

```r
# Compute the column sums
# Easiest way is to use function colSums()
colSums(A)
```

```
## [1]  6 15
```

```r
# or using unit vector
i1%*%A
```

```
##      [,1] [,2]
## [1,]    6   15
```

Note: The order of the summing vector is $\mathbf{1}_n'\mathbf{1}_n = n$.

Note: The outer product of the summing vectors is a matrix in which all elements are ones.

$$
\mathbf{1}_r\mathbf{1}_s' = \mathbf{J}_{r\times s}.
$$

Example: The outer product of the summing vectors $\mathbf{1}_3$ and $\mathbf{1}_2$ is

$$\mathbf{1}_3\mathbf{1}_2' = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} = \mathbf{J}_{3\times 2}$$

Note: Useful result for statistics is

$$\mathbf{C}_n = \mathbf{I} - \bar{\mathbf{J}}_n = \mathbf{I} - \frac{1}{n}\mathbf{J}_n.$$

Matrix $\mathbf{C}_n$ is known as a centering matrix that is symmetric.

It is easy to validate that

$$\mathbf{C}_n = \mathbf{C}_n' = \mathbf{C}_n^2,$$

$$\mathbf{C}_n\mathbf{1}_n = \mathbf{0}_n$$

and

$$\mathbf{C}_n\mathbf{J}_n = \mathbf{J}_n\mathbf{C}_n = \mathbf{0}_n$$

Examples:

For the data vector

$$\mathbf{x}' = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$$

The mean is

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\mathbf{1}_n'\mathbf{x} = \frac{1}{n}\mathbf{x}'\mathbf{1}_n$$

and deviations from the mean are

$$\mathbf{x}'\mathbf{C}_n = \mathbf{x}'(\mathbf{I}_n - \frac{1}{n}\mathbf{J}_n) = \mathbf{x}' - \frac{1}{n}\mathbf{x}'\mathbf{J}_n = \mathbf{x}' - \frac{1}{n}\mathbf{x}'\mathbf{1}_n\mathbf{1}_n' = \mathbf{x}' - \bar{x}\mathbf{1}_n' = \mathbf{x}' - \bar{x}\mathbf{1}_n'$$

$$= (x_1, ..., x_n) - (\bar{x}, ..., \bar{x}) = \begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \vdots \\ x_n - \bar{x} \end{bmatrix}.$$

This is the reason why $\mathbf{C}_n$ is called a centering matrix.

Postmultiplying $\mathbf{x}'\mathbf{C}_n$ by $\mathbf{x}$ gives

$$\mathbf{x}'\mathbf{C}_n\mathbf{x} = (\mathbf{x}' - \bar{x}\mathbf{1}_n')\mathbf{x} = \mathbf{x}'\mathbf{x} - \bar{x}(\mathbf{1}_n'\mathbf{x}) = \mathbf{x}'\mathbf{x} - n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

So the variance is

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{\mathbf{x}'\mathbf{C}\mathbf{x}}{n-1}$$

Note: $\mathbf{x}'\mathbf{C}_n\mathbf{x}$ is an example of a quadratic form. If sample $\mathbf{x}' = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$ is from the normal distribution $N(0,1)$, then $\mathbf{x}'\mathbf{C}_n\mathbf{x}$ is χ^2-distribted with n-1 degrees of freedom.

```r
# R example of computing the variance
# Define data vector x having elements 2,6,-1,0 and 8
x <- c(2,6,-1,0,8)
# Easy way to calculate the variance is to use function var()
var(x)
```

```
## [1] 15
```

```r
# But we can also use centering matrix C
# Define unit matrix J having 5 rows and 5 columns. Unit matrix has 5 rows
# and 5 columns since there are 5 observations in the vector x.
length(x)
```

```
## [1] 5
```

```r
J <- matrix(1,nrow=5,ncol=5)
# Define 5 x 5 identity matrix
I=diag(5)
# Compute the centering matrix C
C <- I-1/5*J
# Compute the variance
t(x)%*%C%*%x/(5-1)
```

```
##      [,1]
## [1,]   15
```

3.6 Exercises:

Exercise 10 Using R define vector the

$$\mathbf{x} = \begin{bmatrix} 6 \\ 19 \\ 10 \\ 13 \\ 31 \\ 34 \end{bmatrix}.$$

a) Compute the normalized vector $\mathbf{h} = \frac{\mathbf{h}}{||x||}$

b) Compute the matrix $\mathbf{H} = \mathbf{I}_6 - 2c \cdot \mathbf{h}\mathbf{h}'$

c) Is matrix $\mathbf{H}$ orthogonal?

Exercise 11: Consider matrix $\mathbf{V}$ that is

$$\mathbf{V} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'.$$

Compute $\mathbf{V}\mathbf{V}$. Is the matrix $\mathbf{V}$ idempotent?

4 Determinant

For every square $n \times n$ matrix $\mathbf{A}$ there is a value $|\mathbf{A}|$ that is called determinant of $\mathbf{A}$. Determinants have many roles in linear algebra and also in applied statistics. For example in statistical experimental design the determinant of the symmetric matrix $\mathbf{X}'\mathbf{X}$ plays a role in finding certain types of optimal designs.

The determinant is a scalar that can be calculated only for a square matrix. In practice, determinants are usually calculated using computer since calculation of the determinant can get very laborious task when the dimension of the matrix is not small. Although we present the procedure for calculating the determinant that works for any square matrix of any dimension, we mainly use computer to compute determinants and it is not necessary to memorize how to compute determinant by hand for matrices higher than 3×3.

Note: The determinant of a matrix $\mathbf{A}$ is denoted by $|\mathbf{A}|$ or $det(\mathbf{A})$.

4.1 2×2 matrix

The determinant of 1×1 matrix is the matrix sole element itself.

The determinant of a 2×2 matrix is not much more difficult to compute. If $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then the determinant of $\mathbf{A}$ is the product of the main diagonal elements (ad) minus the product of the elements off the main diagonal (bc). That is

$$|\mathbf{A}| = ad - bc.$$

Example:

$$\begin{vmatrix} 6 & 8 \\ 17 & 21 \end{vmatrix} = 6 \cdot 21 - 8 \cdot 17 = -10$$

4.2 Larger matrices

Computing the determinant of a higher dimensional matrix than 3×3 is done inductively. It can be computed either by expansion by rows or by columns.

If it is expoanded by elements of row i, the determinant of the $n \times n$ matrix $\mathbf{A}$ is obtained as follows.

$|\mathbf{A}| = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}|\mathbf{M}_{ij}|$, for any i.

If it is expanded by the elements of a column j, it is

$|\mathbf{A}| = \sum_{i=1}^{n} a_{ij}(-1)^{i+j}|\mathbf{M}_{ij}|$, for any j.

Expansion by any row or column gives the same value for the determinant.

Term a_{ij} is entry at row i column j of a matrix $\mathbf{A}$ and $|\mathbf{M}_{ij}|$ is called minor that is the determinant of a matrix that is obtained by removing the i^{th} row and the j^{th} column of the matrix $\mathbf{A}$.

Note: Product $(-1)^{i+j}|\mathbf{M}_{ij}|$ is called cofactor. The cofactor of the entry in the i^{th} row and j^{th} column of $\mathbf{A}$ is often denoted by C_{ij}.

4.3 Sarrus' rule for 3×3 matrices

Determinant of the 3×3 matrix can be calculated using the Sarrus' rule. That is we write out the first 2 columns of the matrix to the right of the 3rd column, such that we have 5 columns in a row. Then the determinant is the sum of the products along the diagonals (solid lines in the picture below) minus the sum of the products along the antidiagonals (dashed lines).

$$
\begin{array}{ccc|cc}
+ & + & + & & \\
a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\
a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\
a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \\
- & - & - & &
\end{array}
$$

If $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$, then according to Sarrus rule we have

$|\mathbf{A}| = a_{11} \cdot a_{22} \cdot a_{33} - a_{13} \cdot a_{22} \cdot a_{31} + a_{12} \cdot a_{23} \cdot a_{31} - a_{11} \cdot a_{23} \cdot a_{32} + a_{13} \cdot a_{21} \cdot a_{32} - a_{12} \cdot a_{21} \cdot a_{33}$

Example:

Consider matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{bmatrix}$

Now the determinant of **A** is calculated using the Sarrus rule

$$
\begin{array}{ccc|cc}
+ & + & + & & \\
1 & 2 & 3 & 1 & 2 \\
4 & 5 & 6 & 4 & 5 \\
7 & 8 & 10 & 7 & 8 \\
- & - & - & &
\end{array}
$$

$$
\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix} = 1 \cdot 5 \cdot 10 - 3 \cdot 5 \cdot 7 + 2 \cdot 6 \cdot 7 - 1 \cdot 6 \cdot 8 + 3 \cdot 4 \cdot 8 - 2 \cdot 4 \cdot 10 = -3
$$

Note: Sarrus rule can only be used on a 3×3 matrices.

```r
# R example of computing the determinant
# It is often easiest to compute the determinant using computer when determinant
# is needed
# In R determinant can be computed using function det()
# Define matrix A
A <- matrix(1:9,byrow=TRUE,ncol=3)
# Compute the determinant
det(A)
```

```
## [1] 6.661338e-16
```

```r
# Defne 2 x 2 matrix B
B <- matrix(c(6,17,8,21),ncol=2)
# Compute the determinant
det(B)
```

```
## [1] -10
```

4.4 Applications

Determinants are used for example in multivariate statistics. Determinant occurs also in the pdf of multivsraite normal distribution. It is also needed to compute matrix inverse.The determinant of a covariance matrix represents the generalized variance that is a single number that characterizes how much variability remains for the set of variables after removing the shared variance among the variables. Consept of the generalized variance was proposed by mathematician Samuel wilks who contributed hugely in the development of mathematical statistics in regard to practical applications.

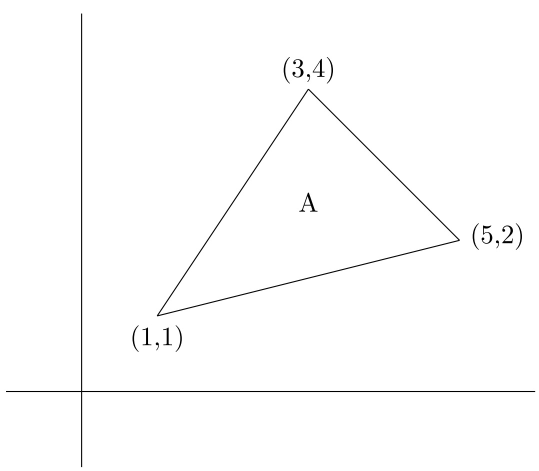
Example: Consider 2×2 correlation matrix

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix},$$

where r_{ij} is the correlation coefficient between the variables i and j.

The determinant of the correlation matrix is $|\mathbf{R}| = 1 - r_{12}r_{21} = 1 - r^2$. Since squared correlation coefficient represents the proportion of the variation shared, $1 - r^2$ represents the variation remaining after removing the shared variation among the variables. This consept holds also for the larger matrices where the generalized variance also describes the remaining variation in the variables after accounting for the associations among the variables. This appears for example in Wilks' Λ test statistic that uses the ratio of two determinants.

Determinant can be used to calculate area of a triangle.



Consider a triangle with vertices at (x_1, y_1), (x_2, y_2), and (x_3, y_3). Then the area of a triangle is $\frac{1}{2}|\mathbf{M}|$, where matrix

$$\mathbf{M} = \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}.$$

Example: The area of the triangle in the picture is

$$M = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 3 & 4 \\ 1 & 5 & 2 \end{vmatrix} = -10.$$

Since we ger a negative value for the determinant, we just drop the sign and make it positive. In this case the sign of the determinant is just determined by the order we put the coordinates of the points in.

The area of the triangle is $\dfrac{1}{2}10 = 5$.

```r
# The area of the triangle
# determine matrix M
M <- matrix(c(1,1,1,1,3,4,1,5,2),byrow=T,nrow=3)
# Determinant of M is
det(M)
```

```
## [1] -10
```

```r
# In R we can take absolute value using function abs()
abs(M)
```

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

```r
# Area of the triangle is
0.5*abs(det(M))
```

```
## [1] 5
```

Note The volume of a tetrahedron is given by $\dfrac{1}{3!}|\mathbf{M}|$, where 3! is a factorial $(3 \cdot 2 \cdot 1)$ and

$$\mathbf{M} = \begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix}$$

4.5 Properties

1) The determinant of the transpose of a matrix $\mathbf{A}$ equals the matrix itself. That is $|\mathbf{A}'| = |\mathbf{A}|$.
2) If two rows (or columns) of $\mathbf{A}$ are the same the determinant is zero.
3) Multiples of rows (or columns) can be added together without affecting the value of the determinant.
4) If matrices $\mathbf{A}$ and $\mathbf{B}$ are square and of the same size then $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$.
5) The determinant of a matrix having a row (or column) full of zeros is zero.

4.6 Exercises

Exercise 12: Let $\mathbf{A} = \begin{bmatrix} -1 & 4 \\ 0.5 & 2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} -1 & 4 & 3 \\ 0.5 & 2 & -2 \\ 4 & -6 & 0 \end{bmatrix}$.

Calculate the determinants by hand and using R

a) $|\mathbf{A}| = \begin{vmatrix} -1 & 4 \\ 0.5 & 2 \end{vmatrix}$

b) $|\mathbf{B}| = \begin{vmatrix} -1 & 4 & 3 \\ 0.5 & 2 & -2 \\ 4 & -6 & 0 \end{vmatrix}$

Exercise 13: Consider the vector $\mathbf{v} = \begin{bmatrix} 2 \\ -1 \\ 0 \\ 5 \\ 3 \end{bmatrix}$

Is the matrix $\mathbf{vv'} - \mathbf{v'vI}_5$ singular?

5 Inverse of a matrix

The inverse of a square matrix $\mathbf{A}$ is a matrix whose product with a matrix $\mathbf{A}$ is identity matrix. The inverse of matrix $\mathbf{A}$ is denoted by $\mathbf{A}^{-1}$ (The superscript -1 denotes the inverse of a matrix).

For scalars the inverse is reciprocal, for example for real number x the inverse is $\frac{1}{x}$ or x^{-1}. For real numbers multiplying by reciprocal is equivalent to dividing. For example if we consider real number 5 then $5 \cdot \frac{1}{5} = \frac{5}{5} = 1$.

For matrices, division by a matrix is not defined but the matrix inverse is equivalent to reciprocal for scalars.

In statistics, we often need to rescale random variables by their standard deviation. For random vectors, the corresponding operation requires the use of matrix inverse. Inverse is therefore an operation that is very often needed in statistics.

In general inverse of a matrix $\mathbf{A}$ is a matrix such that multiplications $\mathbf{A}^{-1}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^{-1}$ equals to an identity matrix $\mathbf{I}$. That is a matrix multiplied by its inverse is always an identity matrix.

Example: For $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ the inverse is $\mathbf{A}^{-1} = \begin{bmatrix} -2 & 1 \\ 1.5 & -0.5 \end{bmatrix}$.

Now the multiplication

$$\mathbf{A}\mathbf{A}^{-1} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 1.5 & -0.5 \end{bmatrix} = \begin{bmatrix} 1 \cdot (-2) + 2 \cdot 1.5 & 1 \cdot 1 + 2 \cdot (-0.5) \\ 3 \cdot (-2) + 4 \cdot 1.5 & 3 \cdot 1 + 4 \cdot (-0.5) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbf{I}_2.$$

Similarly one can verify that $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_2$.

Only square matrices can have the inverse, but all square matrices do not have an inverse matrix. If square matrix has an inverse, it is unique. Matrix is said to be invertible (or nonsingular) if it has an inverse. If matrix does not have an inverse, matrix is said to be singular.

The inverse of a square matrix $\mathbf{A}$ exists only if the determinant $|\mathbf{A}|$ is not zero. This follows from the property $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$. Since $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$ it follows that $|\mathbf{A}\mathbf{A}^{-1}| = |\mathbf{A}||\mathbf{A}| = 1$ since the determinant of an identity matrix is always 1. That is if $|\mathbf{A}| = 0$ the previous does not hold.

5.1 Calculating the matrix inverse

Since obtaining the inverse of a matrix is computationally demanding task we usually compute the inverse using computer when the inverse matrix is needed. In some special cases matrix inversion does not require extensive computing.

1) The inverse of a 2×2 matrix.

Let $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, where a, b, c and d are real numbers. Then the inverse of $\mathbf{A}$ is obtained by

$$\mathbf{A}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

That is to obtain inverse of a 2×2 matrix we interchange diagonal elements of matrix, change the sign of the off-diagonal elements and multiply the result by scalar $\frac{1}{ad - bc}$ (where $ad - bc$ is the determinant of the matrix $\mathbf{A}$).

Example: If $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ then the inverse of $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \frac{1}{1 \cdot 4 - 3 \cdot 2} \begin{bmatrix} 4 & -3 \\ -2 & 1 \end{bmatrix} = \frac{1}{-2} \begin{bmatrix} 4 & -2 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1.5 & -0.5 \end{bmatrix}.$$

```
# R example of calculating the inverse
# We can use function solve() to compute the inverse if the inverse exists
# Define matrix A
A <- matrix(c(1,2,3,4),nrow=2,byrow=T)
# Compute the inverse of A
solve(A)
```

```
##      [,1] [,2]
## [1,] -2.0  1.0
## [2,]  1.5 -0.5
```

```
# Compute the matrix multiplied by its inverse
A%*%solve(A)
```

```
##      [,1]         [,2]
## [1,]    1 1.110223e-16
## [2,]    0 1.000000e+00
```

```
# or
solve(A)%*%A
```

```
##      [,1]         [,2]
## [1,]    1 4.440892e-16
## [2,]    0 1.000000e+00
```

```
# Function solve()  can be used to compute the inverse to matrices of any dimension
```

2) The inverse of an identity matrix is the identity matrix itself i.e.

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

Note: The only idempotent matrices that are invertible are identity matrices.

3) The inverse of the diagonal matrix of any size is obtained by replacing each corresponding diagonal element by its reciprocal.

For example inverse of a diagonal matrix

$$\mathbf{D} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

is

$$\mathbf{D}^{-1} = \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1/3 \end{bmatrix}$$

Note: The determinant of the matrix $\mathbf{J_{n\times n}}$ is zero, so it does not have the inverse matrix.

```r
# R example of calculating the inverse
# Define 5 by 5 matrix J in which all the elements are the same
J <- matrix(1,nrow=5,ncol=5)
J
```

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

```r
# The inverse of the matrix J is not defined. Try command solve(J)
# The determinant of the matrix J
det(J)
```

```
## [1] 0
```

Note: If at least two rows (or columns) of a matrix are equal, then the determinant is 0.

4) Orthogonal matrix $\mathbf{P}$ is square and has nonzero determinant so the inverse $\mathbf{P}^{-1}$ exists. Matrix $\mathbf{P}$ is orthogonal if $\mathbf{PP'} = \mathbf{I}$ and $\mathbf{P'P} = \mathbf{I}$ so the inverse of $\mathbf{P}$ is the transpose of $\mathbf{P}$ i.e. $\mathbf{P}^{-1} = \mathbf{P'}$.

The inverse $\mathbf{A}^{-1}$ can exist only if matrix $\mathbf{A}$ is square and when the determinant of $\mathbf{A}$ is not zero.

In general the inverse of a nonsingular matrix $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} adj(\mathbf{A}),$$

where $adj(\mathbf{A})$ is the adjugate matrix of $\mathbf{A}$. The adjugate matrix of $\mathbf{A}$ is the transpose of the matrix of cofactors.

Consider $n \times n$ matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & ... & a_{1n} \\ a_{21} & a_{22} & ... & a_{2n} \\ \vdots & & \ddots & \\ a_{n1} & a_{n2} & ... & a_{nn} \end{bmatrix}$$

and the matrix of cofactors

$$cof(\mathbf{A}) = \begin{bmatrix} C_{11} & C_{12} & ... & C_{1n} \\ C_{21} & C_{22} & ... & C_{2n} \\ \vdots & & \ddots & \\ C_{n1} & C_{n2} & ... & C_{nn} \end{bmatrix}.$$

Now the inverse of $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \begin{bmatrix} C_{11} & C_{21} & ... & C_{n1} \\ C_{12} & C_{22} & ... & C_{n2} \\ \vdots & & \ddots & \\ C_{1n} & C_{2n} & ... & C_{nn} \end{bmatrix}$$

This form of derivating the inverse of a matrix is inconvenient in cases where dimension is not small. However, this form describes well how the elements of the inverse of a matrix are related to. The cofactors are essential to understanding the relationship between the elements of $\mathbf{A}$ and the inverse $\mathbf{A}^{-1}$.

Example: Let $\mathbf{A} = \begin{bmatrix} 6 & 6 & 1 \\ 1 & 3 & 4 \\ 1 & 2 & 2 \end{bmatrix}$ and $|\mathbf{A}| = -1$.

Next we will derive the cofactors.

The cofactors of the first column are:

$$C_{11} = (-1)^{1+1} \begin{vmatrix} 3 & 4 \\ 2 & 2 \end{vmatrix} = -2$$

$$C_{21} = (-1)^{2+1} \begin{vmatrix} 6 & 1 \\ 2 & 2 \end{vmatrix} = -10$$

$$C_{31} = (-1)^{3+1} \begin{vmatrix} 6 & 1 \\ 3 & 4 \end{vmatrix} = 21$$

Similarly the cofactors of elements of the second Column are:

$$\begin{aligned} C_{12} &= 2 \\ C_{22} &= 11 \\ C_{32} &= -23 \end{aligned}$$

and Column 3:

$$\begin{aligned} C_{13} &= -1 \\ C_{23} &= -6 \\ C_{33} &= 12 \end{aligned}$$

The matrix of cofactors is

$$cof(\mathbf{A}) = \begin{bmatrix} -2 & 2 & 1 \\ -10 & 11 & -6 \\ 21 & -23 & 12 \end{bmatrix}$$

and the adjugate of **A** is

$$adj(\mathbf{A}) = cof(\mathbf{A})' = \begin{bmatrix} -2 & -10 & 21 \\ 2 & 11 & -23 \\ -1 & -6 & 12 \end{bmatrix}$$

Now the inverse of **A** is

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} adj(\mathbf{A}) = \frac{1}{-1} \begin{bmatrix} -2 & -10 & 21 \\ 2 & 11 & -23 \\ -1 & -6 & 12 \end{bmatrix} = \begin{bmatrix} 2 & 10 & -21 \\ -2 & -11 & 23 \\ 1 & 6 & -12 \end{bmatrix}$$

5.2 Useful properties of the inverse

If matrix $\mathbf{A}$ is a square nonsingular matrix, its inverse $\mathbf{A}^{-1}$ has the following properties

1) The inverse commutes with $\mathbf{A}$. That is both multiplications $\mathbf{A}^{-1}\mathbf{A}$ and $\mathbf{A}\mathbf{A}^{-1}$ are identity matrices.

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

2) The inverse of matrix $\mathbf{A}$ is unique.

3) The inverse of an inverse is the original matrix i.e.

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}.$$

4) The determinant of the inverse of matrix $\mathbf{A}$ is the reciprocal of the determinant of the matrix $\mathbf{A}$. That is $|\mathbf{A}^{-1}| = \dfrac{1}{|\mathbf{A}|}$.

5) The inverse of the transpose is the transpose of the inverse i.e.

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$$

6) Inverse matrix of a symmetric matrix is also symmetric i.e if $\mathbf{A}' = \mathbf{A}$ then

$$(\mathbf{A}^{-1})' = \mathbf{A}^{-1}.$$

7) The inverse of a matrix multiplication is equal to the multiplication of the inverse matrices in the reverse order i.e.

$$(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}.$$

Note: In seventh property matrices need to be square and of the same dimension. Otherwise multiplication would not be possible. Matrices also have to be nonsingular so that all inverses exist.

5.3 Solving the system of linear equations using the inverse

Consider a system of linear equations having n equations and n variables, i.e. we can present the system of linear equations in a matrix form as $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is square matrix holding coefficients, vector $\mathbf{x}$ contains the unknown variables and vector $\mathbf{b}$ is the right hand side. If the matrix $\mathbf{A}$ is nonsingular i.e. matrix has an inverse then

$$\mathbf{Ax} = \mathbf{b}$$

can be solved as

$$\mathbf{Ax} = \mathbf{y}$$

$$\Leftrightarrow \mathbf{A}^{-1}\mathbf{Ax} = \mathbf{A}^{-1}\mathbf{y}$$

$$\Leftrightarrow \mathbf{Ix} = \mathbf{A}^{-1}\mathbf{y}$$

$$\Leftrightarrow \mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$$

That is if $\mathbf{A}$ is nonsingular $(det(\mathbf{A}) \neq 0))$, then $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ is a unique solution for the system of linear equations for any vector $\mathbf{b}$.

Note: If matrix $\mathbf{A}$ is singular i.e. $|\mathbf{A}| = 0$, then the system of linear equations $\mathbf{Ax} = \mathbf{y}$ doesn't have unique solution or the solution does not exist.

Example:

If we have equations $\begin{cases} x_1 + x_2 = 3 \\ 2x_1 - 0.5x_2 = 1 \end{cases}$

we can write this in a matrix form as

$$\begin{bmatrix} 1 & 1 \\ 2 & -0.5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

More generally this is

$$\mathbf{Ax} = \mathbf{b},$$

where

$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & -0.5 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$

We can solve unknown vector $\mathbf{x}$ using the inverse of $\mathbf{A}$

$$\mathbf{A}^{-1} = \frac{1}{1 \cdot (-1) - 1 \cdot 2} \begin{bmatrix} -0.5 & -1 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} 1/5 & 2/5 \\ 4/5 & -2/5 \end{bmatrix}.$$

Thus the vector x is

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \begin{bmatrix} 1/5 & 2/5 \\ 4/5 & -2/5 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

We can check the solution by putting $x_1 = 1$ and $x_2 = 2$ back to the original equation in place of x_1 and x_2.

We get

$$\begin{cases} 1 + 2 = 3 \\ 2 \times 1 - 0.5 \times 2 = 1 \end{cases}$$

this holds so the solution is valid.

```r
# R example of solving the system of linear equations
# In R we define equation in the matrix form
# Define coefficients matrix A
A <- matrix(c(1,1,2,-0.5),byrow=T,ncol=2)
# Define vector b
b=c(3,1)
# In R we solve system of linear equations using function solve that can be used
# if the inverse exists
# Function solve takes coefficients matrix and right hand side vector as an arguments.
solve(A,b)
```

```
## [1] 1 2
```

```r
# We also could have used the inverse matrix to solve x
solve(A)%*%b
```

```
##      [,1]
## [1,]    1
## [2,]    2
```

Note: If the inverse exists, the system of linear equations has unique solution that can be found for example using the inverse. This is useful property that is used for example in regression analysis.

5.4 Exercises

Exercise 14: Find the inverse of the diagonal matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$$

Exercise 15: Using R find the inverse of the matrix

$$\mathbf{Z} = \begin{bmatrix} 3 & 7 & 7 & 7 \\ 2 & 8 & 3 & 8 \\ 1 & 6 & 3 & 7 \\ 3 & 2 & 2 & 0 \end{bmatrix}$$

Exercise 16: Consider the system of linear equations

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 3 \\ -1x_1 + 3x_2 + 2x_3 = 5 \\ 2x_1 + 3x_3 = 8 \end{cases}$$

a) Write the equations in a matrix form as

$$\mathbf{Ax} = \mathbf{b}.$$

b) Solve the the system of linear equations using the inverse.

6 Linear dependence and matrix rank

6.1 Linear dependence and independence

The non-zero vector $\mathbf{x}_0$ is said to be a linear function (or linear combination) of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ if there exists constants $b_1, b_2, \ldots, b_k$ such that

$$\mathbf{x}_0 = b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \ldots + b_k\mathbf{x}_k$$

or more compactly

$$\mathbf{x}_0 = \mathbf{Xb}$$

where $\mathbf{b}' = \begin{bmatrix} b_1 & b_2 & \ldots & b_k \end{bmatrix}$ and $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_k \end{bmatrix}$.

Consider a set of vectors, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$. Sometimes it is important to know whether one of the vectors can be written as a linear combination of the other vectors. If that is the case, then the vectors are said to be linearly dependent. Eqivalently, a set of k vectors $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, ..., \mathbf{x}_k$ is said to be linearly dependent if there exists scalars a_1, a_2, a_3,..., a_k such that

$$a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + \ldots + a_k\mathbf{x}_k = \mathbf{0}$$

where at least one of the scalars a_1, a_2, a_3,..., a_k is not zero and none of the vectors $\mathbf{x}_i$ is null vector. Equivalently,

$$\mathbf{Xa} = \mathbf{0}$$

for some non-null vector $\mathbf{a}$, where $\mathbf{X} = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_k\end{bmatrix}$ where none of columns of $\mathbf{X}$ is a null vector. If a set of vectors is not linearly dependent then vectors are said to be linearly independent.

Example: Vectors $\mathbf{x}_1 = \begin{bmatrix}1\\2\\3\end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix}2\\-1\\0\end{bmatrix}$ and $\mathbf{x}_3 = \begin{bmatrix}5\\0\\3\end{bmatrix}$ are linearly dependent since when we select $a_1 = 1$, $a_2 = 2$ and $a_3 = -1$ we have

$$\begin{bmatrix}1\\2\\3\end{bmatrix} + 2\begin{bmatrix}2\\-1\\0\end{bmatrix} - \begin{bmatrix}5\\0\\3\end{bmatrix} = \begin{bmatrix}1+2\cdot 2-5\\2+2\cdot(-1)-0\\3+0-3\end{bmatrix} = \begin{bmatrix}0\\0\\0\end{bmatrix} = \mathbf{0}.$$

In this example for example third vector $\mathbf{x}_3$ is the following linear combination of the vectors $\mathbf{x}_1$ and $\mathbf{x}_2$: $\mathbf{x}_3 = -1 \cdot \mathbf{x}_1 + 2\mathbf{x}_2$. That is vectors are linearly dependent if at least one of the vectors can be presented as a linear combination of the other vectors.

Note Also $\mathbf{x}_2$ can be written as a linear combination of $\mathbf{x}_1$ and $\mathbf{x}_2$, and $\mathbf{x}_1$ can be written as a linear combination of $\mathbf{x}_2$ and $\mathbf{x}_3$. Find those linear combinations yourself.

Let us assume that p linearly dependent vectors of order p are used as columns of square matrix $\mathbf{A}$. Linear dependence of the columns implies that at least one column of the matrix can be expressed as a linear combination of the other columns. Therefore, the determinant $|\mathbf{A}|$ is zero. This provides a simple test for linear dependence of p vectors of order p: evaluate the determinant of the matrix formed from the vectors. If the determinant is zero then the columns vectors are linearly dependent. If the determinant is not zero, the vectors are linearly independent.

Note. If the columns of a square matrix $\mathbf{A}$ are linearly dependent, then also the columns of $\mathbf{A}'$ are linearly dependent. If the columns of a square matrix $\mathbf{A}$ are linearly independent, then also the columns of $\mathbf{A}'$ are linearly independent.

Example: Vectors $\mathbf{x}_1 = \begin{bmatrix}0\\1\\2\end{bmatrix}$, $\begin{bmatrix}2\\3\\1\end{bmatrix}$ and $\begin{bmatrix}0\\2\\4\end{bmatrix}$ are linearly dependent since

$$\begin{vmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3\end{vmatrix} = \begin{vmatrix}0 & 2 & 0\\1 & 3 & 2\\2 & 1 & 4\end{vmatrix} = 0$$

But vectors $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix}$ are linearly independent since

$$\begin{vmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 0 \\ 1 & 3 & 2 \\ 2 & 1 & 4 \end{vmatrix} = 10.$$

Note: If $\mathbf{Xa} = \mathbf{0}$ only for $\mathbf{a} = \mathbf{0}$, then the columns of $\mathbf{X}$ are linearly independent.

Example: Consider vectors $\mathbf{x}_1 = \begin{bmatrix} 3 \\ -6 \\ 9 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 5 \\ -5 \end{bmatrix}$, $\mathbf{x}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{x}_4 = \begin{bmatrix} -6 \\ 12 \\ -18 \end{bmatrix}$.

If $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 & \mathbf{x}_4 \end{bmatrix}$, then $\mathbf{Xa} = \mathbf{0}$ is

$$\begin{bmatrix} 3 & 0 & 2 & -6 \\ -6 & 5 & 1 & 12 \\ 9 & -5 & 1 & -18 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} 3a_1 + 2a_3 - 6a_4 = 0 \\ -6a_1 + 5a_2 + a_3 + 12a_4 = 0 \\ 9a_1 - 5a_2 + a_3 - 18a_4 = 0 \end{cases}$$

$$\mathbf{a} = \begin{bmatrix} a_4 \\ -3/2a_4 \\ 3/2a_4 \\ a_4 \end{bmatrix},$$

Any choice of a_4 will give $\mathbf{Xa} = \mathbf{0}$. Therefore the vectors are linearly dependent.

Note: A set of linearly independent vectors of order p cannot contain more than p vectors. Consider a set of p linearly independent vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$ of order p. Now it is impossible to find still another vector of length p, $\mathbf{x}_{p+1}$ that is not a linear combination of vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p$. Therefore, the columns of matrix $\mathbf{X}_{r \times c}$, are linearly dependent if $c > r$, i.e., if the number of columns is higher than the number of rows. Correspondingly, the columns of matrix $\mathbf{X}'$ (i.e., the rows of matrix $\mathbf{X}$) are linearly dependent if $r > c$.

6.2 Matrix rank

Earlier we learned that the determinant of a square matrix is zero when the columns (or equivalently the rows) are linearly dependent. This leads to more general question of how many linearly independent columns and rows the matrix has. This relationship is simple since the number of linearly independent rows in a matrix is the same as the number of linearly independent columns.

The rank of matrix $\mathbf{X}$ is the maximum number of linearly independent columns in $\mathbf{X}$. It can be conceptually illustrated as follows. Consider matrix $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_k \end{bmatrix}$. Let us construct all possible matrices $\mathbf{A}_j$ that can be constructed by dropping $0, 1, \ldots k-1$ columns from matrix $\mathbf{X}$. Therefore, we get a set of matrices where the number of columns varies between 1 and k. For each of these matrices, we can define whether the columns are linearly independent. The rank of matrix $\mathbf{X}$, denoted by $rank(\mathbf{X})$ or $r(\mathbf{X})$, is the maximum number of columns among such matrices $\mathbf{A}_j$ that has linearly independent columns.

Note: We could also define the rank using rows (i.e., using columns of the transpose $\mathbf{X}'$). That is, the maximum number of linearly independent rows equals to the maximum number of linearly independent columns.

Matrix $\mathbf{X}_{r \times c}$ is said to have full column rank if $r(\mathbf{X}) = c$.

Matrix $\mathbf{X}_{r \times c}$ is said to have full row rank if $r(\mathbf{X}) = r$.

A square matrix $\mathbf{X}_{p \times p}$ is said to have full rank if $r(\mathbf{X}) = p$.

Note: Rank is important characteristic of any matrix and is often encountered in linear algebra and statistics. Especially, it is important to understand whether a matrix has full rank, full column rank and full row rank. In linear model $\mathbf{y} = \mathbf{X}\beta + \mathbf{e}$, the model matrix should have full column rank. If that is not the case, the product matrix $\mathbf{X}'\mathbf{X}$ is singular and cannot be inverted to compute the parameter estimates using the unbiased estimator $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. In such case, we could find an estimate of β but it is not unique. This means that the model is not identifiable.

Example: Next we compute the rank of matrices $\mathbf{A} = \begin{bmatrix} 1 & 2 & 0 \\ 1 & 3 & 2 \\ 2 & 1 & 4 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 0 & 1 \end{bmatrix}$ using R.

In R we can compute the rank using function qr() so computing rank i.e. the number of linearly independent columns and rows is not necessary.

```
# R example of computing the rank of a matrix

# Define matrix A
A <- matrix(c(1,1,2,2,3,1,0,2,4),nrow=3)
A
```

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

```
# Compute the rank
qr(A)
```

```
## $qr
##            [,1]      [,2]       [,3]
## [1,] -2.4494897 -2.857738 -4.0824829
## [2,]  0.4082483 -2.415229  0.6900656
## [3,]  0.8164966 -0.752101  1.6903085
##
## $rank
```

```
## [1] 3
##
## $qraux
## [1] 1.408248 1.659048 1.690309
##
## $pivot
## [1] 1 2 3
##
## attr(,"class")
## [1] "qr"
```

```r
#or
qr(A)$rank
```

```
## [1] 3
```

```r
# The rank of A is 3 i.e. the square matrix A is of full rank. That is the matrix
# has 3 linearly independent rows and columns like we determined earlier by computing
# the determinant that was 10
det(A)
```

```
## [1] 10
```

```r
# Define matrix B
B <- matrix(c(1,2,0,2,4,1),nrow=3)
B
```

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

```r
# Compute the rank
qr(B)
```

```
## $qr
##            [,1]      [,2]
## [1,] -2.2360680 -4.472136
## [2,]  0.8944272 -1.000000
## [3,]  0.0000000  1.000000
##
## $rank
## [1] 2
##
## $qraux
## [1] 1.447214 1.000000
##
## $pivot
## [1] 1 2
##
## attr(,"class")
## [1] "qr"
```

```r
qr(B)$rank
```

```
## [1] 2
```

```r
# The rank is 2 i.e. this matrix has 2 independent rows and columns.
# Matrix B has full column rank.
```

Note: The rank does not tell which rows and columns of the matrix are linearly independent or linearly dependent.

6.3 Properties

1) Rank of the matrix is a positive integer except that the rank of null matrix is 0 i.e. $r(\mathbf{0}) = 0$.
2) The rank of $rank(\mathbf{A}_{r \times c}) \leq min(r, c)$. That is, the rank equals or is less than the smaller of the number of rows and columns.
3) When $\mathbf{A}$ is a $n \times n$ square matrix and $r(\mathbf{A}) = n$ then $\mathbf{A}$ is nonsingular that is the inverse $\mathbf{A}^{-1}$ exists.
4) When $\mathbf{A}$ is a $n \times n$ square matrix and $r(\mathbf{A}) < n$ then $\mathbf{A}$ is singular and the inverse $\mathbf{A}^{-1}$ does not exist.
5) When $\mathbf{A}$ is a $n \times n$ square matrix and $r(\mathbf{A}) = n$ then $\mathbf{A}$ is said to have full rank.
6) When $\mathbf{A}$ is a $r \times c$ square matrix and the $r(\mathbf{A}) = r < c$ then $\mathbf{A}$ is said to have full row rank. (Note that in $\mathbf{A}$ the number of rows r is less than the number of columns c)
7) When $\mathbf{A}$ is a $r \times c$ square matrix and the $r(\mathbf{A}) = c < r$ then $\mathbf{A}$ is said to have full column rank. (Note that in $\mathbf{A}$ the number of columns c is less than the number of rows r)

6.4 Exercises

Exercise 20: Determine if vectors $\mathbf{x}_1 = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}$ and $\mathbf{x}_3 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$ form a linearly independent set of vectors.

Exercise 21: Find the rank of the matrix $\mathbf{A} = \begin{bmatrix} 2 & 3 & 1 & 2 & 1 \\ 3 & 3 & 3 & 3 & 1 \\ 1 & 1 & 2 & 2 & 1 \\ 3 & 3 & 2 & 2 & 2 \end{bmatrix}$

7 Partitioned matrices (Block matrix)

A matrix can be partitioned into smaller matrices by inserting horizontal and/or vertical lines between selected rows and columns.

Example Consider matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 0 & 2 \\ 4 & 6 & 8 & 0 & 1 \\ 2 & 3 & 5 & 7 & 9 \end{bmatrix}$$

We can partition matrix $\mathbf{A}$ for example to four submatrices $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$ such that

$$\mathbf{A} = \left[\begin{array}{cc|ccc} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 0 & 2 \\ \hline 4 & 6 & 8 & 0 & 1 \\ 2 & 3 & 5 & 7 & 9 \end{array}\right].$$

By defining $\mathbf{A}_{11} = \begin{bmatrix} 1 & 2 \\ 6 & 7 \end{bmatrix}$, $\mathbf{A}_{12} = \begin{bmatrix} 3 & 4 & 5 \\ 8 & 0 & 2 \end{bmatrix}$, $\mathbf{A}_{21} = \begin{bmatrix} 4 & 6 \\ 2 & 3 \end{bmatrix}$ ja $\mathbf{A}_{22} = \begin{bmatrix} 8 & 0 & 1 \\ 5 & 7 & 9 \end{bmatrix}$,

matrix $\mathbf{A}$ acn be written as

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

This procedure is called a partitioning of matrix $\mathbf{A}$ and matrix $\mathbf{A}$ that is presented using submatrices is called a partitioned matrix. Since we use vertical and horizontal lines in partitioning, matrices $\mathbf{A}_{11}$ and $\mathbf{A}_{12}$ have same number of columns as do matrices $\mathbf{A}_{12}$ and $\mathbf{A}_{22}$. Matrices $\mathbf{A}_{12}$ and $\mathbf{A}_{22}$ have the same number of rows as do matrices $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$.

```r
# R example of partitioned matrices
# Define matrix A
A <- matrix(c(1,2,3,4,5,6,7,8,0,2,4,6,8,0,1,2,3,5,7,9),byrow=TRUE,ncol=5)
A
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    2    3    4    5
## [2,]    6    7    8    0    2
## [3,]    4    6    8    0    1
## [4,]    2    3    5    7    9
```

```r
# Pick submatrices
A11 <- A[1:2,1:2]
A11
```

```
##      [,1] [,2]
## [1,]    1    2
## [2,]    6    7
```

```r
A12 <- A[1:2,3:5]
A12
```

```
##      [,1] [,2] [,3]
## [1,]    3    4    5
## [2,]    8    0    2
```

```r
A21 <- A[c(3,4),c(1,2)]
A21
```

```
##      [,1] [,2]
## [1,]    4    6
## [2,]    2    3
```

```r
A22 <- A[c(3,4),c(3,4,5)]
A22
```

```
##      [,1] [,2] [,3]
## [1,]    8    0    1
## [2,]    5    7    9
```

```r
# Form the original matrix using the submatrices
# Combine submatrices A11 and A12 using the column bind (cbind)
AA1 <- cbind(A11,A12)
AA1
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    2    3    4    5
## [2,]    6    7    8    0    2
```

```r
# Combine submatrices A21 and A22
AA2 <- cbind(A21,A22)
AA2
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    4    6    8    0    1
## [2,]    2    3    5    7    9
```

```r
# Combine AA1 and AA2 by rows
AA <- rbind(AA1,AA2)
AA
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1    2    3    4    5
## [2,]    6    7    8    0    2
## [3,]    4    6    8    0    1
## [4,]    2    3    5    7    9
```

```r
# Check if matrices A and AA are equal
identical(A,AA)
```

```
## [1] TRUE
```

There is no restriction that matrix have to be partitioned into four submatrices. For example matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 0 & 2 \\ 4 & 6 & 8 & 0 & 1 \\ 2 & 3 & 5 & 7 & 9 \end{bmatrix}$$

can be written in partitioned form as

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \mathbf{a}_3 & \mathbf{a}_4 & \mathbf{a}_5 \end{bmatrix}$$

such that submatrices are columns of matrix $\mathbf{A}$. In this case matrix is said to be partitioned by columns and submatrices are

$$\mathbf{a}_1 = \begin{bmatrix} 1 \\ 6 \\ 4 \\ 2 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 2 \\ 7 \\ 6 \\ 3 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} 3 \\ 8 \\ 8 \\ 5 \end{bmatrix}, \mathbf{a}_4 = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 7 \end{bmatrix} \text{ and } \mathbf{a}_5 = \begin{bmatrix} 5 \\ 2 \\ 1 \\ 9 \end{bmatrix}.$$

```r
# R example of partitioned matrices
# Define column vectors a1, a2, a3, a4 and a5
a1 <- c(1,6,4,2)
a2 <- c(2,7,6,3)
a3 <- c(3,8,8,5)
a4 <- c(4,0,0,7)
a5 <- c(5,2,1,9)

# Combine vectors to a matrix
A <- cbind(a1,a2,a3,a4,a5)
A
```

```
##      a1 a2 a3 a4 a5
## [1,]  1  2  3  4  5
## [2,]  6  7  8  0  2
## [3,]  4  6  8  0  1
## [4,]  2  3  5  7  9
```

```r
# or combining by rows
B <- rbind(a1,a2,a3,a4,a5)
B
```

```
##    [,1] [,2] [,3] [,4]
## a1    1    6    4    2
## a2    2    7    6    3
## a3    3    8    8    5
## a4    4    0    0    7
## a5    5    2    1    9
```

Note: The submatrices are also called blocks.

It is possible to think that matrix partition is done by "drawing" lines between certain rows and/or columns and each array of numbers separated by lines is a submatrix. The same matrix can be partitioned in many ways but each line must go the full breadth of the matrix since matrix partitioning in any staggered manner is not allowed. For example partitioning

$$\left(\begin{array}{cc|ccc} 1 & \multicolumn{1}{c}{2} & 3 & 4 & 5 \\ \cline{1-2} 6 & 7 & 8 & 0 & 2 \\ 4 & 6 & 8 & 0 & 1 \\ \cline{1-2} 2 & \multicolumn{1}{c}{3} & 5 & 7 & 9 \end{array}\right)$$

is not possible.

A block diagonal matrix is a partitioned matrix of the form

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{33} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{nn} \end{bmatrix},$$

where matrices $\mathbf{A}_{ii}$ are square matrices of any size and the zero matrices are of conformable size.

Example: Consider matrices $\mathbf{A}_{11} = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}$, $\mathbf{A}_{22} = \begin{bmatrix} -3 \end{bmatrix}$ and $\mathbf{A}_{33} = \begin{bmatrix} 1 & -4 & 1 \\ 8 & 2 & 0 \\ 3 & 3 & -8 \end{bmatrix}$.

Matrices $\mathbf{A}_{11}$, $\mathbf{A}_{22}$ and $\mathbf{A}_{33}$ are all square matrices but unequal size. From those matrices we can form a block diagonal matrix as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{33} \end{bmatrix} = \begin{bmatrix} 1 & 3 & 0 & 0 & 0 & 0 \\ 2 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -4 & 1 \\ 0 & 0 & 0 & 8 & 2 & 0 \\ 0 & 0 & 0 & 3 & 3 & -8 \end{bmatrix}$$

```r
# R example of constructing the same block diagonal matrix as in the previous example
# We use library magic to create matrix
# (install package if needed using the install.package("magic") command)
library(magic)
```

```
## Loading required package: abind
```

```r
# Define submatrices
X1 <- matrix(1:4,nrow=2)
X2 <- -3
X3 <- matrix(c(1,8,3,-4,2,3,1,0,-8),nrow=3)
# Construct the block diagonal matrix using adiag() command. Notice that matrix X2
# is a scalar Matrices are given in the same order that they are in the block
# diagonal matrix.
D <- adiag(X1,X2,X3)
D
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    1    3    0    0    0    0
## [2,]    2    4    0    0    0    0
## [3,]    0    0   -3    0    0    0
## [4,]    0    0    0    1   -4    1
## [5,]    0    0    0    8    2    0
## [6,]    0    0    0    3    3   -8
```

7.1 Multiplication of partitioned matrices

Matrix multiplication of two partitioned matrices $\mathbf{A}$ and $\mathbf{B}$ can be expressed in partitioned form if submatrices of multiplied matrices are appropriately conformable for multiplication. If matrices $\mathbf{A}$ and $\mathbf{B}$ are partitioned such that

$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$

Now the multiplication $\mathbf{AB}$ is

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} = \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix}.$$

So if partitioning of $\mathbf{A}$ along columns is the same as partitioning of $\mathbf{B}$ along its rows, then $\mathbf{A}_{11}$ (and $\mathbf{A}_{12}$) has the same number of columns as $\mathbf{B}_{11}$ (and $\mathbf{B}_{12}$) has rows, and similarly $\mathbf{A}_{21}$ (and $\mathbf{A}_{22}$) has the same number of columns as $\mathbf{B}_{21}$ (and $\mathbf{B}_{22}$) has rows.

7.2 Exercises

Exercise 17: Partition matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 & 6 \\ -1 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix}$ as

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

such that submatrix $\mathbf{A}_{21}$ have dimension 1×2. write down matrices $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$.

8 Eigenvalues and eigenvectors

Eigenvalues and eigenvectors are important in multiple areas of statistics. For example principal components analysis can be obtained from the eigenvalue decomposition. Principal components analysis is often used for dimensionality reduction, since method transforms the data into new variables called principal components such that principal components are linearly uncorrelated.

When matrix $\mathbf{A}$ is multiplied by a vector $\mathbf{u}$ and the result is the same vector ($\mathbf{u}$) multiplied by a scalar λ i.e.

$\mathbf{Au} = \lambda\mathbf{u}$,

we say that scalar λ is an eigenvalue of $\mathbf{A}$ and $\mathbf{u}$ is a eigenvector of $\mathbf{A}$ corresponding to λ.

Note: In practice we would solve eigenvalues and eigenvectors using computer.

Eigenvalues of the matrix $\mathbf{A}$ can be solved from the characteristic equation that is

$$|\mathbf{A} - \lambda\mathbf{I}| = 0.$$

Eigenvalues are the roots of the characteristic equation. There are a total of n roots when $\mathbf{A}$ is a $n \times n$ matrix, but some roots may be zero. Eigenvalues are denoted by λ_1, λ_2, λ_3,$\ldots$,λ_n. Corresponding to each eigenvalue λ_i there is a vector $\mathbf{u}_i$ satisfying equation

$$\mathbf{Au} = \lambda\mathbf{u}$$

Those vectors $\mathbf{u}_1$, $\mathbf{u}_2$,$\ldots$,$\mathbf{u}_n$ are called eigenvectors.

Example: Matrix $\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}$

has characteristic equation defined by

$$|\mathbf{A} - \lambda\mathbf{I}| = \left| \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = \begin{vmatrix} 1-\lambda & 4 \\ 9 & 1-\lambda \end{vmatrix} = 0.$$

By expandng the determinant we get that

$\Rightarrow (1-\lambda)^2 - 36 = 0$

$\Rightarrow (1-\lambda)^2 = 36$

$\Rightarrow 1 - \lambda = \pm 6$

$\Rightarrow \lambda_1 = 7$ or $\lambda_2 = -5$

That is the eigenvalues of the matrix $\mathbf{A}$ are $\lambda_1 = 7$ and $\lambda_2 = -5$

It can be shown that

$\begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = -5 \begin{bmatrix} 2 \\ -3 \end{bmatrix}$ and $\begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 7 \begin{bmatrix} 2 \\ 3 \end{bmatrix}$. That is the vector $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$ is the eigenvector corresponding to the eigenvalue -5 and $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ is the eigenvector corresponding to the eigenvalue 7.

Note: The determinant of the $n \times n$ matrix can be calculated as a product of the eigenvalues. If $\mathbf{A}$ is $n \times n$ matrix then the determinant of $\mathbf{A}$ is

$|\mathbf{A}| = \lambda_1 \lambda_2 ... \lambda_n = \prod_{i=1}^{n} \lambda_i$

Example: Consider matrix $\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}$ then

$$|\mathbf{A}| = \lambda_1 \lambda_2 = 7 \cdot (-5) = -35$$

Note: The trace of a matrix is the sum of the eigenvalues. $tr(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$

Examle: Let $\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}$

Then the trace of the matrix $\mathbf{A}$ is

$$tr(\mathbf{A}) = |\mathbf{A}| = \lambda_1 + \lambda_2 = 7 - 5 = 2$$

```r
# R example of computing the egenvalues
# In R eigenvalues and eigenvectors can be computed using function eigen()
# Define matrix A
A <- matrix(c(1,4,9,1),byrow=TRUE,ncol=2)
# Compute the eigenvalues and eigenvectors and save them to variable eigA
```

```r
eigA <- eigen(A)
eigA
```

```
## eigen() decomposition
## $values
## [1]  7 -5
##
## $vectors
##           [,1]       [,2]
## [1,] 0.5547002 -0.5547002
## [2,] 0.8320503  0.8320503
```

```r
# Function eigen() returns the list that contain the eigenvalues and the eigenvectors
# Pick the eigenvalues
eigA$values
```

```
## [1]  7 -5
```

```r
# R will return eigenvectors corresponding to the eigenvalues in a matrix form
eigA$vectors
```

```
##           [,1]       [,2]
## [1,] 0.5547002 -0.5547002
## [2,] 0.8320503  0.8320503
```

```r
# The first column corresponds to the eigenvalue 7
eigA$vectors[,1]
```

```
## [1] 0.5547002 0.8320503
```

```r
# and the second column corresponds to the eigenvalue -5
eigA$vectors[,2]
```

```
## [1] -0.5547002  0.8320503
```

```r
# compute the product of the eigenvalues
prod(eigA$values)
```

```
## [1] -35
```

```r
# This equals to the determinant that is
det(A)
```

```
## [1] -35
```

```r
# Compte the sum of the eigenvalues
sum(eigA$values)
```

```
## [1] 2
```

```r
# This equals to the trace of the matrix A that is
sum(diag(A))
```

```
## [1] 2
```

8.1 Exercises

Exercise 18: Using R generate matrices **A** and **B** using commands A <- matrix(c(4,5,3,6,3,7,2,7,6,3,1,5,4,7,3,6),byrow=T,nrow=4) B <- (A+t(A))/5

a) Find the eigenvalues λ_i of **B**

b) What is the eigenvector corresponding to the largest eigenvalue of **B**?

c) Show that $\sum_{i=1}^{8} = tr(\mathbf{B})$.

d) Show that $|\mathbf{B}| = \prod_{i=1}^{8} \lambda_i$

e) Compute eigenvectors of **A**

9 Generalized inverse

The inverse of a square matrix $\mathbf{A}$, $\mathbf{A}^{-1}$, was defined to be a matrix that produce identity matrix when multiplied with the original matrix. However matrix inverse exists only if determinant of the matrix is nonzero. If it exists, it is unique.

Generalized inverse is generalization of the inverse matrix. It is defined for any matrix $\mathbf{A}$ and denoted by $\mathbf{G}$ (sometimes by $\mathbf{A}^{-}$) so that $\mathbf{AGA} = \mathbf{A}$. The generalized inverse is not unique. The most commonly used generalized inverse is the Moore-Penrose pseudoinverse.

Generalized inverse is used for example to compute least squares solution to a system of linear equations that does not have unique solution. It is also used in Restricted Maximum Likelihood estimation of mixed-effect models.

If $\mathbf{A}$ is any matrix, then a Moore-Penrose inverse of a matrix $\mathbf{A}$ is a unique matrix $\mathbf{G}$ that satisfy four conditions

1) $\mathbf{AGA} = \mathbf{A}$

2) $\mathbf{GAG} = \mathbf{G}$

3) $(\mathbf{AG})' = \mathbf{AG}$

4) $(\mathbf{GA})' = \mathbf{GA}$

The Moore-Penrose inverse ($\mathbf{G}$) is uniquely defined for by these four conditions and this matrix always exists for any real valued matrix.

Example: To demonstrate how generalized inverse works when we solve system of linear equations we use simple example of linear model where the matrix $\mathbf{X}'\mathbf{X}$ is not invertible. That is we cannot solve coefficients using inverse but we can get solution using the generalized inverse altought it should be kept in mind that this solution is not unique. Assume that we have dataset with three variables and four measurements.

The regression model equations are defined by

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \beta_3 x_{13} + e_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \beta_3 x_{23} + e_2 \\ y_3 = \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} + \beta_3 x_{33} + e_3 \\ y_4 = \beta_0 + \beta_1 x_{41} + \beta_2 x_{42} + \beta_3 x_{43} + e_4 \end{cases},$$

where y_i is the value of the response for the i^{th} observation, x_{i1}, x_{i2} and x_{i3} are the predictor values for the i:th observation, β_0, β_1, β_2 and β_0 are the coefficients to be estimated and e_i is the residual error.

This models can be written shorter as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$$

or in a matrix form that offers a compact way to present the model. In a matrix form model is

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} \\ 1 & x_{21} & x_{22} & x_{23} \\ 1 & x_{31} & x_{32} & x_{33} \\ 1 & x_{41} & x_{42} & x_{43} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$$

That is simply

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

where we write observations in a column vector $\mathbf{y}$, matrix $\mathbf{X}$ is called as the design matrix, the coefficients are written into vector $\boldsymbol{\beta}$ and the residual errors are written into a column vector $\mathbf{e}$.

Assume that our design matrix is

$$\mathbf{X} = \begin{bmatrix} 1 & 3 & 3 & 1 \\ 1 & 2 & 2 & 1 \\ 1 & 4 & 4 & 3 \\ 1 & 2 & 2 & 2 \end{bmatrix} \text{ and the measured response values are } \mathbf{y} = \begin{bmatrix} 6 \\ 7 \\ -1 \\ -2 \end{bmatrix}.$$

Now our task is to solve coefficients vector $\boldsymbol{\beta}$. Without going to details solution is given by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

however now matrix $\mathbf{X}'\mathbf{X}$ in noninvertible so we cannot solve the equation using the inverse $(\mathbf{X}'\mathbf{X})^{-1}$. However we can solve this using the Moore-Penrose generalized inverse of the matrix $\mathbf{X}'\mathbf{X}$.

If the Moore-Penrose generalized inverse is denoted by $\mathbf{G}$ then the estimated coefficients are given by

$\hat{\boldsymbol{\beta}} = \mathbf{G}\mathbf{X}'\mathbf{y}$ that is an unbised estimator of $\boldsymbol{\beta}$.

Fitted values are given by

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{G}\mathbf{X}'\mathbf{y}$$

Since computing Moore-Penrose generalized inverse by hand is a laborious task we solve this using R.

```
# First we have to download library MASS that contain function which we will
# use to calculate the Moore-Penrose inverse
# Note: you have to install package MASS only once so if you have installed
# it earlier you can skip this step. Use command install.packages("MASS") to
# download the package  You can select any server from the list that opens.
# Load package
library(MASS)

# Define the design matrix X
X <- matrix(c(1,3,3,1,1,2,2,1,1,4,4,3,1,2,2,2),byrow=T,nrow=4)

# print design matrix
X
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    3    3    1
## [2,]    1    2    2    1
## [3,]    1    4    4    3
## [4,]    1    2    2    2
```

```
# Define vector Y
Y=c(6,7,-1,-2)

# Calculate X'X and assign it to variable XTX
XTX <- t(X)%*%X
```

```r
# You can try to solve the inverse of X'X using command solve(XTX)

# Since XTX is not invertible we use Moore-Penrose generalized inverse
# Solve the Moore-Penrose inverse of XTX
G <- ginv(XTX)
# Print G
G
```

```
##               [,1]       [,2]       [,3]          [,4]
## [1,]  3.000000e+00 -0.5000000 -0.5000000  2.170486e-14
## [2,] -5.000000e-01  0.1527778  0.1527778 -1.944444e-01
## [3,] -5.000000e-01  0.1527778  0.1527778 -1.944444e-01
## [4,]  6.383782e-16 -0.1944444 -0.1944444  6.111111e-01
```

```r
# and finaly solve coefficients
beta <- G%*%t(X)%*%Y

# Print estimated coefficients
beta
```

```
##           [,1]
## [1,]  6.000000
## [2,]  1.166667
## [3,]  1.166667
## [4,] -5.666667
```

```r
# Fitted values are
Yhat <- X%*%G%*%t(X)%*%Y
Yhat
```

```
##            [,1]
## [1,]  7.3333333
## [2,]  5.0000000
## [3,] -1.6666667
## [4,] -0.6666667
```

The Moore-Penrose generalized inverse is (approximately since due to rounding this isn't exact)

$$\mathbf{G} = \begin{bmatrix} 3.000000e+00 & -0.5000000 & -0.5000000 & 2.170486e-14 \\ -5.000000e-01 & 0.1527778 & 0.1527778 & -1.944444e-01 \\ -5.000000e-01 & 0.1527778 & 0.1527778 & -1.944444e-01 \\ 6.383782e-16 & -0.1944444 & -0.1944444 & 6.111111e-01 \end{bmatrix}$$

The unbised estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 6.000000 \\ 1.166667 \\ 1.166667 \\ -5.666667 \end{bmatrix}.$$

The fitted values are

$$\hat{\mathbf{y}} = \begin{bmatrix} 7.3333333 \\ 5.0000000 \\ -1.6666667 \\ -0.6666667 \end{bmatrix}$$

9.1 Exercises:

Exercise 19:

Consider the system of linear equation $\mathbf{Ax} = \mathbf{y}$,

where $\mathbf{A} = \begin{bmatrix} 2 & 3 & 1 & 2 & 1 \\ 3 & 3 & 3 & 3 & 1 \\ 1 & 1 & 2 & 2 & 1 \\ 3 & 3 & 2 & 2 & 2 \end{bmatrix}$ and $y = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$.

a) Compute the Moore-Penrose generalized inverse of the matrix A?
b) Solve the system of linear equations $\mathbf{Ax} = \mathbf{y}$ using the Moore-Penrose generalized inverse.
c) Verify that the solution satisfy the linear equation $\mathbf{Ax} = \mathbf{y}$.

10 solutions

Exercise 1: Consider matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 3 & 4 & 4 \\ 7 & 9 & 8 & -2 & 3 \\ 1 & -3 & 4 & -6 & -7 \\ 3 & 5 & 2 & 1 & 5 \end{bmatrix}$$

a) What is the dimension of the matrix?

b) What is the entry at a_{43}?

c) c) What is the trace of the matrix

$$\mathbf{C} = \begin{bmatrix} 2 & 6 & 2 & 4 \\ 1 & -1 & 9 & 0 \\ 0 & 3 & 3 & -4 \\ -2 & -4 & 2 & 0 \end{bmatrix}?$$

solution:

a) The dimension of the matrix $\mathbf{A}$ is 4×5. The matrix $\mathbf{A}$ has 4 rows and 5 columns.

b) Entry at row 4 and column 3 is $a_{43} = 2$

c) The trace of $\mathbf{C}$ is the sum of the diagonal elements. Trace is

$$tr\,(\mathbf{C}) = 2 + (-1) + 3 + 0 = 4.$$

Exercise 2: Find the transpose of the following matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & -1 \\ 3 & 0 & 2 \end{bmatrix}$$

$$\mathbf{x} = \begin{bmatrix} -2 & 1 & 0 & 4 \end{bmatrix}$$

solution:

$$\mathbf{A}' = \begin{bmatrix} 1 & 3 \\ 3 & 0 \\ -1 & 2 \end{bmatrix}$$

$$\mathbf{x}' = \begin{bmatrix} -2 \\ 1 \\ 0 \\ 4 \end{bmatrix}$$

Exercise 3: Consider vectors

$$\mathbf{x} = \begin{bmatrix} 1 \\ -2 \\ 2 \\ 3 \\ 0 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} -1 \\ 2 \\ 3 \\ 4 \\ 1 \end{bmatrix}$$

Compute the inner product $\mathbf{x}'\mathbf{y}$ by hand and using R.

solution:

The inner product $\mathbf{x} = \begin{bmatrix} 1 & -2 & 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \\ 3 \\ 4 \\ 1 \end{bmatrix} = 1 \cdot (-1) - 2 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + 0 \cdot 1 = -1 - 4 + 6 + 12 = 13$

```r
# Define vectors x and y
x <- c(1,-2,2,3,0)
y <- c(-1,2,3,4,1)
# The inner product is
x%*%y
```

```
##      [,1]
## [1,]   13
```

Exercise 4 Consider vectors

$$\mathbf{x} = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}$$

Compte the outer product $\mathbf{x}\mathbf{y}'$ by hand and using R.

solution:

The outer product: $\mathbf{x}\mathbf{y}' = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 0 \cdot (-1) & 0 \cdot 2 & 0 \cdot 1 \\ 2 \cdot (-1) & 2 \cdot 2 & 2 \cdot 1 \\ 1 \cdot (-1) & 1 \cdot 2 & 1 \cdot 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -2 & 4 & 2 \\ -1 & 2 & 1 \end{bmatrix}$

```r
# Define vectors x and y
x <- c(0,2,1)
y <- c(-1,2,1)
# The outer product is
x%*%t(y)
```

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

Exercise 5: Consider matrices $\mathbf{X} = \begin{bmatrix} 1 & 3 & -1 \\ 3 & 0 & 2 \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 1 & -1 \end{bmatrix}$

Compute $\mathbf{XY}$ by hand and using R.

solution:

$$\mathbf{XY} = \begin{bmatrix} 1 & 3 & -1 \\ 4 & 0 & 2 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 0 & 3 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1\cdot 2 + 3\cdot 0 + (-1)\cdot 1 & 1\cdot 1 + 3\cdot 0 + (-1)\cdot(-1) \\ 4\cdot 2 + 0\cdot 0 + 2\cdot 1 & 4\cdot 1 + 0\cdot 0 + 2\cdot(-1) \end{bmatrix} = \begin{bmatrix} 1 & 11 \\ 10 & 2 \end{bmatrix}$$

```r
# Define matrices X and Y
X <- matrix(c(1,3,-1,4,0,2),byrow=T,nrow=2)
Y <- matrix(c(2,1,0,3,1,-1),byrow=T,nrow=3)

# Compute multiplication XY
X%*%Y
```

```
##      [,1] [,2]
## [1,]    1   11
## [2,]   10    2
```

Exercise 6: Consider matrices $\mathbf{A}_{3\times 4}$, $\mathbf{B}_{2\times 4}$ $\mathbf{C}_{3\times 4}$ and $\mathbf{D}_{4\times 2}$

Find if the multiplications are defined and define the dimension of the multiplication result if it is defined

a) $\mathbf{ADB}$
b) $\mathbf{DC}$
c) $\mathbf{A'AD}$

solution:

a) Multiplication $\mathbf{ADB}$ is 3×4 matrix.
b) Multiplication $\mathbf{DC}$ is not defined.

c) Multiplication $\mathbf{A}'\mathbf{AD}$ is 4×2 matrix.

Exercise 7: Consider $n \times n$ matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. What is the transpose of the matrix multiplication $\mathbf{A}'\mathbf{BC}'$?

solution:

Use the property that $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$ and $(\mathbf{A}')' = \mathbf{A}$.

$(\mathbf{A}'\mathbf{BC}')' = (\mathbf{C}')'\mathbf{B}'(\mathbf{A}')' = \mathbf{CB}'\mathbf{A}$

Exercise 8: Consider $n \times n$ matrices $\mathbf{X}$ and $\mathbf{Y}$. What is the transpose of the matrix multiplication $\mathbf{X}'\mathbf{XYX}'\mathbf{X}$?

solution: Use the property that $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$.

$(\mathbf{X}'\mathbf{XYX}'\mathbf{X})' = \mathbf{X}'(\mathbf{X}')'\mathbf{Y}'\mathbf{X}'(\mathbf{X}')' = \mathbf{X}'\mathbf{XY}'\mathbf{X}'\mathbf{X}$

Exercise 9: If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices. When does it hold that$(\mathbf{A}+\mathbf{B})(\mathbf{A}-\mathbf{B}) = \mathbf{A}^2 - \mathbf{B}^2$?

solution:

$(\mathbf{A}+\mathbf{B})(\mathbf{A}-\mathbf{B})$ is equal to $\mathbf{A}^2 - \mathbf{B}^2$ when $\mathbf{AB} + \mathbf{BA} = \mathbf{0}_{n \times n} \rightarrow \mathbf{AB} = \mathbf{BA}$, since

$(\mathbf{A}+\mathbf{B})(\mathbf{A}-\mathbf{B}) = \mathbf{AA} + \mathbf{AB} - \mathbf{BA} - \mathbf{BB} = \mathbf{A}^2 + \mathbf{AB} - \mathbf{BA} - \mathbf{B}^2$. This is $\mathbf{A}^2 - \mathbf{B}^2$ only when $\mathbf{AB} = \mathbf{BA}$.

Exercise 10: Using R define vector $\mathbf{x} = \begin{bmatrix} 6 \\ 19 \\ 10 \\ 13 \\ 31 \\ 34 \end{bmatrix}$.

a) Compute the normalized vector $\mathbf{h} = \dfrac{\mathbf{x}}{||x||}$

b) Compute the matrix $\mathbf{H} = \mathbf{I}_6 - 2 \cdot \mathbf{hh}'$

c) Is matrix $\mathbf{H}$ orthogonal?

solution:

```r
# Define vector x
x <- c(6,19,10,13,151,3)
# a)
# Create normalized vector h
x=sample(0:15,6,rep=T)
h <- x/sqrt(t(x)%*%x)
```

```
## Warning in x/sqrt(t(x) %*% x): Recycling array of length 1 in vector-array arithmetic is deprecated.
##   Use c() or as.vector() instead.
```

```r
h
```

```
## [1] 0.35220822 0.35220822 0.05031546 0.00000000 0.50315461 0.70441645
```

```r
# We can confirm that this is a unit vector by calculating the length of the vector h
sqrt(h%*%h)
```

```
##      [,1]
## [1,]    1
```

```r
# Since the length of the vector h is 1, this is normalized vector of x.
# b) Define matrix H
H <- diag(6)-2*h%*%t(h)
H
```

```
##             [,1]        [,2]        [,3] [,4]        [,5]         [,6]
## [1,]  0.75189873 -0.24810127 -0.03544304    0 -0.35443038 -0.496202532
## [2,] -0.24810127  0.75189873 -0.03544304    0 -0.35443038 -0.496202532
## [3,] -0.03544304 -0.03544304  0.99493671    0 -0.05063291 -0.070886076
## [4,]  0.00000000  0.00000000  0.00000000    1  0.00000000  0.000000000
## [5,] -0.35443038 -0.35443038 -0.05063291    0  0.49367089 -0.708860759
## [6,] -0.49620253 -0.49620253 -0.07088608    0 -0.70886076  0.007594937
```

```r
# c) Matrix H is orthogonal if H'H=I and HH'=I, where I is a identity matrix and H' is
# the transpose of H
# pre- and postmultiply H by its transpose.
t(H)%*%H
```

```
##               [,1]          [,2]          [,3] [,4]          [,5]
## [1,]  1.000000e+00 -2.775558e-17 -6.938894e-18    0  0.000000e+00
## [2,] -2.775558e-17  1.000000e+00 -6.938894e-18    0  0.000000e+00
## [3,] -6.938894e-18 -6.938894e-18  1.000000e+00    0  6.938894e-18
## [4,]  0.000000e+00  0.000000e+00  0.000000e+00    1  0.000000e+00
## [5,]  0.000000e+00  0.000000e+00  6.938894e-18    0  1.000000e+00
## [6,] -7.719519e-17 -7.719519e-17 -1.311885e-17    0 -1.032160e-16
##               [,6]
## [1,] -7.719519e-17
## [2,] -7.719519e-17
## [3,] -1.311885e-17
## [4,]  0.000000e+00
## [5,] -1.032160e-16
## [6,]  1.000000e+00
```

```r
H%*%t(H)
```

```
##               [,1]          [,2]          [,3] [,4]          [,5]
## [1,]  1.000000e+00 -2.775558e-17 -6.938894e-18    0  0.000000e+00
## [2,] -2.775558e-17  1.000000e+00 -6.938894e-18    0  0.000000e+00
## [3,] -6.938894e-18 -6.938894e-18  1.000000e+00    0  6.938894e-18
## [4,]  0.000000e+00  0.000000e+00  0.000000e+00    1  0.000000e+00
## [5,]  0.000000e+00  0.000000e+00  6.938894e-18    0  1.000000e+00
## [6,] -7.719519e-17 -7.719519e-17 -1.311885e-17    0 -1.032160e-16
##               [,6]
## [1,] -7.719519e-17
## [2,] -7.719519e-17
## [3,] -1.311885e-17
## [4,]  0.000000e+00
## [5,] -1.032160e-16
## [6,]  1.000000e+00
```

```r
# Matrix H is orthogonal. Due to the rounding error all off-diagonal elements aren't
# exactly zero.
# Round the multiplications to 10 decimal places.
round(t(H)%*%H,10)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    1    0    0    0    0    0
## [2,]    0    1    0    0    0    0
## [3,]    0    0    1    0    0    0
## [4,]    0    0    0    1    0    0
## [5,]    0    0    0    0    1    0
## [6,]    0    0    0    0    0    1
```

```r
round(H%*%t(H),10)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]    1    0    0    0    0    0
## [2,]    0    1    0    0    0    0
## [3,]    0    0    1    0    0    0
## [4,]    0    0    0    1    0    0
## [5,]    0    0    0    0    1    0
## [6,]    0    0    0    0    0    1
```

Exercise 11: Consider matrix $\mathbf{V}$ that is

$$\mathbf{V} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'.$$

Compute $\mathbf{VV}$. Is the matrix $\mathbf{V}$ idempotent?

solution:

$$\mathbf{VV} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}' = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A},$$

since $(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{A} = \mathbf{I}$.

Matrix $\mathbf{V}$ is idempotent.

Exercise 12: Consider matrices $\mathbf{A} = \begin{bmatrix} -1 & 4 \\ 0.5 & 2 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} -1 & 4 & 3 \\ 0.5 & 2 & -2 \\ 4 & -6 & 0 \end{bmatrix}$.

Calculate the determinants by hand and using R

a) $|\mathbf{A}| = \begin{vmatrix} -1 & 4 \\ 0.5 & 2 \end{vmatrix}$

b) $|\mathbf{B}| = \begin{vmatrix} -1 & 4 & 3 \\ 0.5 & 2 & -2 \\ 4 & -6 & 0 \end{vmatrix}$

solution:

a) $|\mathbf{A}| = \begin{vmatrix} -1 & 4 \\ 0.5 & 2 \end{vmatrix} = -1 \cdot 2 - 4 \cdot 0.5 = -2 - 2 = -4$

```r
# Define matrix A
A <- matrix(c(-1,4,0.5,2),byrow=T,nrow=2)

# The determinant of A is
det(A)
```

```
## [1] -4
```

b)

$$|\mathbf{B}| = \begin{vmatrix} -1 & 4 & 3 \\ 0.5 & 2 & -2 \\ 4 & -6 & 0 \end{vmatrix} = -1 \cdot 2 \cdot 0 + 4 \cdot -2 \cdot 4 + 3 \cdot 0.5 \cdot -6 - 4 \cdot 2 \cdot 3 - (-6) \cdot (-2) \cdot (-1) - 0 \cdot 0.5 \cdot 4$$

$$= -32 - 9 - 24 + 12 = -53$$

```r
# Define matrix A
B <- matrix(c(-1,4,3,0.5,2,-2,4,-6,0),byrow=T,nrow=3)

# The determinant of B is
det(B)
```

```
## [1] -53
```

Exercise 13: Consider vector $\mathbf{v} = \begin{bmatrix} 2 \\ -1 \\ 0 \\ 5 \\ 3 \end{bmatrix}$

Is the matrix $\mathbf{v}\mathbf{v}' - \mathbf{v}'\mathbf{v}\mathbf{I}_5$ singular?

solution:

The matrix is singular if the determinant of the matrix is zero. A singular matrix does not have the inverse matrix.

```r
# Define vector v
v <- c(1,-1,0,5,3)
# compute the matrix
v%*%t(v)-c(t(v)%*%v)*diag(5)
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]  -35   -1    0    5    3
## [2,]   -1  -35    0   -5   -3
## [3,]    0    0  -36    0    0
## [4,]    5   -5    0  -11   15
## [5,]    3   -3    0   15  -27
```

```r
# The determinant is
det(v%*%t(v)-c(t(v)%*%v)*diag(5))
```

```
## [1] 0
```

```r
# Matrix is singular since the determinant is zero
```

$$\mathbf{v}\mathbf{v}' - \mathbf{v}'\mathbf{v}\mathbf{I}_5 = \begin{bmatrix} -35 & -1 & 0 & 5 & 3 \\ -1 & -35 & 0 & -5 & -3 \\ 0 & 0 & -36 & 0 & 0 \\ 5 & -5 & 0 & -11 & 15 \\ 3 & -3 & 0 & 15 & -27 \end{bmatrix}$$

$$\begin{vmatrix} -35 & -1 & 0 & 5 & 3 \\ -1 & -35 & 0 & -5 & -3 \\ 0 & 0 & -36 & 0 & 0 \\ 5 & -5 & 0 & -11 & 15 \\ 3 & -3 & 0 & 15 & -27 \end{vmatrix} = 0$$

Exercise 14: Find the inverse of the diagonal matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 0 \\ 0 & 4 \end{bmatrix}$$

solution:

The inverse of a diagonal matrix is obtained by replacing each entry in the diagonal with its reciprocal.

The inverse of $\mathbf{A}$ is $\mathbf{A}^{-1} = \begin{bmatrix} 1/3 & 0 \\ 0 & 1/4 \end{bmatrix}$

Exercise 15: Using R find the inverse of the matrix

$$\mathbf{Z} = \begin{bmatrix} 3 & 7 & 7 & 7 \\ 2 & 8 & 3 & 8 \\ 1 & 6 & 3 & 7 \\ 3 & 2 & 2 & 0 \end{bmatrix}$$

solution:

The inverse of $\mathbf{Z}$

```
# Define matrix Z
Z <- matrix(c(3,7,7,7,2,8,3,8,1,6,3,7,3,2,2,0),byrow=T,nrow=4)

# The inverse of Z is
solve(Z)
```

```
##       [,1]  [,2] [,3] [,4]
## [1,] -0.88 -1.68  2.8  1.4
## [2,]  0.84  2.24 -3.4 -1.2
## [3,]  0.48  0.28 -0.8 -0.4
## [4,] -0.80 -1.80  3.0  1.0
```

The inverse of $\mathbf{Z}$ is $\mathbf{Z}^{-1} = \begin{bmatrix} -0.88 & -1.68 & 2.80 & 1.40 \\ 0.84 & 2.24 & -3.40 & -1.20 \\ 0.48 & 0.28 & -0.80 & -0.40 \\ -0.80 & -1.80 & 3.00 & 1.00 \end{bmatrix}$

Exercise 16: Consider the system of linear equations

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 3 \\ -1x_1 + 3x_2 + 2x_3 = 5 \\ 2x_1 + 3x_3 = 8 \end{cases}$$

a) Write the equations in a matrix form as $\mathbf{Ax} = \mathbf{b}$.
b) Solve the equations using the inverse.

solution:

a) $\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ -1 & 3 & 2 \\ 2 & 0 & 3 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 3 \\ 5 \\ 8 \end{bmatrix}$

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 3 & 2 \\ 2 & 0 & 3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 8 \end{bmatrix}$$

b) The inverse of $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \begin{bmatrix} 1.8 & -1.2 & -1 \\ 1.4 & -0.6 & -1 \\ 1.2 & 0.8 & 1 \end{bmatrix}$$

Vector $\mathbf{x}$ is given by $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$

$$\mathbf{x} = \begin{bmatrix} 1.8 & -1.2 & -1 \\ 1.4 & -0.6 & -1 \\ 1.2 & 0.8 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 8 \end{bmatrix} = \begin{bmatrix} -8.6 \\ -6.8 \\ 8.4 \end{bmatrix}$$

Solve the system of linear equations using R

```r
# Define coefficients matrix A
A <- matrix(c(1,2,3,-1,3,2,2,0,3),byrow=T,nrow=3)
# Define vector b
b <- c(3,5,8)
# The inverse of A is
solve(A)
```

```
##      [,1] [,2] [,3]
## [1,]  1.8 -1.2   -1
## [2,]  1.4 -0.6   -1
## [3,] -1.2  0.8    1
```

```r
# Vector x is
x <- solve(A,b)
x
```

```
## [1] -8.6 -6.8  8.4
```

```r
# or
solve(A)%*%b
```

```
##      [,1]
## [1,] -8.6
## [2,] -6.8
## [3,]  8.4
```

Exercise 17: Partition matrix $\mathbf{A} = \begin{bmatrix} 1 & 2 & 6 \\ -1 & 1 & 0 \\ 3 & 4 & 1 \end{bmatrix}$ as

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

such that $\mathbf{A}_{21}$ have dimension 1×2. write down matrices $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$.

solution:

Partitioned matrices $\mathbf{A}_{21}$ and $\mathbf{A}_{22}$ must have same number of rows as do the matrices $\mathbf{A}_{11}$ and $\mathbf{A}_{12}$. Also matrices $\mathbf{A}_{11}$ and $\mathbf{A}_{21}$ must have the same number of columns as do $\mathbf{A}_{12}$ and $\mathbf{A}_{22}$

$$\mathbf{A} = \left[\begin{array}{cc|c} 1 & 2 & 6 \\ -1 & 1 & 0 \\ \hline 3 & 4 & 1 \end{array}\right]$$

Submatrices are $\mathbf{A}_{11} = \begin{bmatrix} 1 & 2 \\ -1 & 1 \end{bmatrix}$, $\mathbf{A}_{12} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$, $\mathbf{A}_{21} = \begin{bmatrix} 3 & 4 \end{bmatrix}$ and $\mathbf{A}_{22} = \begin{bmatrix} 1 \end{bmatrix}$.

Exercise 18:

Using R generate matrices $\mathbf{A}$ and $\mathbf{B}$ using commands

```
A <- matrix(c(4,5,3,6,3,7,2,7,6,3,1,5,4,7,3,6),byrow=T,nrow=4)
```

```
B <- (A+t(A))/5
```

a) Define the eigenvalues λ_i of $\mathbf{B}$

b) What is the eigenvector corresponding to the largest eigenvalue of $\mathbf{B}$?

c) Show that $\sum_{i=1}^{8} = tr(\mathbf{B})$.

d) Show that $|\mathbf{B}| = \prod_{i=1}^{8} \lambda_i$

e) Compute eigenvectors of $\mathbf{A}$

solution:

```r
# a)
# Define matrices A and B
A <- matrix(c(4,5,3,6,3,7,2,7,6,3,1,5,4,7,3,6),byrow=T,nrow=4)
B <- (A+t(A))/5
# Eigenvalues
eigen <- eigen(B)
eigen$values
```

```
## [1]  7.5302549  0.9879674 -0.3672831 -0.9509391
```

```r
# and the corresponding eigenvectors
eigen$vectors
```

```
##            [,1]        [,2]        [,3]       [,4]
## [1,] -0.4555322 -0.59148544 -0.43937342  0.4995863
## [2,] -0.5754017  0.64814816 -0.46903014 -0.1697868
## [3,] -0.3290431 -0.47167848  0.04749966 -0.8166969
## [4,] -0.5942508  0.08699628  0.76466039  0.2336496
```

```r
# b)
# The eigenvector corresponding to the largest eigenvalue 7.530 is
eigen$vectors[,1]
```

```
## [1] -0.4555322 -0.5754017 -0.3290431 -0.5942508
```

```r
# and the

# c)
# Sum of the eigenvalues
sum(eigen$values)
```

```
## [1] 7.2
```

```r
# The trace of the matrix
sum(diag(B))
```

```
## [1] 7.2
```

```r
# d)
# Product of the eigenvalues
prod(eigen$values)
```

```
## [1] 2.5984
```

```r
# The determinant of the matrix
det(B)
```

```
## [1] 2.5984
```

Exercise 19:

Consider the system of linear equation $\mathbf{Ax} = \mathbf{y}$,

where $\mathbf{A} = \begin{bmatrix} 2 & 3 & 1 & 2 & 1 \\ 3 & 3 & 3 & 3 & 1 \\ 1 & 1 & 2 & 2 & 1 \\ 3 & 3 & 2 & 2 & 2 \end{bmatrix}$ and $y = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$.

a) Compute the Moore-Penrose generalized inverse of the matrix A?
b) Solve the system of linear equations $\mathbf{Ax} = \mathbf{y}$ using the Moore-Penrose generalized inverse.
c) Verify that the solution satisfy the linear equation $\mathbf{Ax} = \mathbf{y}$.

solution:

```
# Define matrix A
A <- matrix(c(2,3,1,2,1,3,3,3,3,1,1,1,2,2,1,3,3,2,2,2),byrow=T,nrow=4)
# Define vector b
b <- 1:4
# a)
# Solve the Moore-Penrose inverse
library(MASS)
ginv=ginv(A)
# Print the Moore-Penrose generalized inverse
ginv
```

```
##               [,1]  [,2]  [,3]  [,4]
## [1,] -5.000000e-01  0.45 -0.65  0.35
## [2,]  5.000000e-01 -0.05 -0.15 -0.15
## [3,] -5.000000e-01  0.35  0.05  0.05
## [4,]  5.000000e-01 -0.15  0.55 -0.45
## [5,] -5.551115e-16 -0.80  0.60  0.60
```

```
# b)
# Solve vector x
x <- ginv%*%b
x
```

```
##       [,1]
## [1,] -0.15
## [2,] -0.65
## [3,]  0.55
## [4,]  0.05
## [5,]  2.60
```

```
# c) Verify the solution by computing Ax
A%*%x
```

```
##      [,1]
```

```
## [1,]    1
## [2,]    2
## [3,]    3
## [4,]    4
```

```r
# The solution is valid since this is the same as vector b.
```

Solution vector $\mathbf{x} = \begin{bmatrix} -0.15 \\ -0.65 \\ 0.55 \\ 0.05 \\ 2.60 \end{bmatrix}$

Exercise 20: Determine if vectors $\mathbf{x}_1 = \begin{bmatrix} 2 \\ 4 \\ 6 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}$ and $\mathbf{x}_3 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix}$ form a linearly independent set of vectors.

solution:

```r
# define vectors
x1 <- c(2,4,6)
x2 <- c(1,1,4)
x3 <- c(1,3,1)
# Combine vectors into matrix using cbind()
A <- cbind(x1,x2,x3)
A
```

```
##      x1 x2 x3
## [1,]  2  1  1
## [2,]  4  1  3
## [3,]  6  4  1
```

```r
# Now vectors are linearly dependent if the determinant of A is 0
det(A)
```

```
## [1] 2
```

```r
# Vectors are linearly independent. Other ther way to check this is to compute
# the rank of A.
qr(rbind(A))$rank
```

```
## [1] 3
```

```r
# The rank of A is 3.
# The matrix A is a full rank matrix i.e. all rows and columns are linearly independent.
```

Exercise 21: Find the rank of the matrix $\mathbf{A} = \begin{bmatrix} 2 & 3 & 1 & 2 & 1 \\ 3 & 3 & 3 & 3 & 1 \\ 1 & 1 & 2 & 2 & 1 \\ 3 & 3 & 2 & 2 & 2 \end{bmatrix}$

solution:

```r
# Define matrix A
A <- matrix(c(2,3,1,2,1,3,3,3,3,1,1,1,2,2,1,3,3,2,2,2),byrow=T,nrow=4)
# Rank of A is
qr(A)$rank
```

```
## [1] 4
```

```r
# The rank is 4. Matrix A has full row rank.
```