Introduction to linear algebra in R

Søren Højsgaard,
Department of Mathematical Sciences,
Aalborg University, Denmark
https://people.math.aau.dk/~sorenh/

February 15, 2021

Contents

1 Introduction

The first version of these notes were written in 2005. These notes/slides have two aims: 1) Introducing linear algebra (vectors and matrices) and 2) showing how to work with these concepts in R. They were written in an attempt to give a specific group of students a "feeling" for what matrices, vectors etc. are all about. Hence the notes/slides are not suitable for a course in linear algebra.

2 Vectors

2.1 Vectors

A column vector is a list of numbers stacked on top of each other, e.g.

$$a = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

A row vector is a list of numbers written one after the other, e.g.

$$b = (2, 1, 3)$$

In both cases, the list is ordered, i.e.

$$(2, 1, 3) \neq (1, 2, 3).$$

We make the following convention:

- In what follows all vectors are column vectors unless otherwise stated.
- However, writing column vectors takes up more space than row vectors. Therefore we shall frequently write vectors as row vectors, but with the understanding that it really is a column vector.

A general n–vector has the form

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

where the a_is are numbers, and this vector shall be written $a = (a_1, \dots, a_n)$.

A graphical representation of 2–vectors is shown Figure 1. Note that row and column vectors

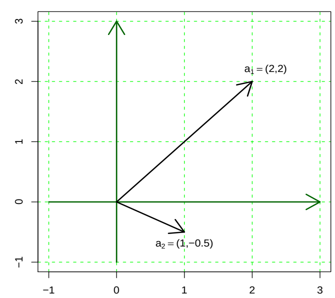

Figure 1: Two 2-vectors

are drawn the same way.

```
> a <- c(1, 3, 2)
> a

[1] 1 3 2
```

The vector `a` is in R printed "in row format" but can really be regarded as a column vector, cfr. the convention above.

2.2 Transpose of vectors

Transposing a vector means turning a column (row) vector into a row (column) vector. The transpose is denoted by "$^\top$".

Example 1

$$\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}^\top = [1, 3, 2] \quad \texttt{og} \quad [1, 3, 2]^\top = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix}$$

□

Hence transposing twice takes us back to where we started:

$$a = (a^\top)^\top$$

```
> t(a)

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

2.3 Multiplying a vector by a number

If a is a vector and α is a number then αa is the vector

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

See Figure 2.

Example 2

$$7 \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 7 \\ 21 \\ 14 \end{bmatrix}$$

□

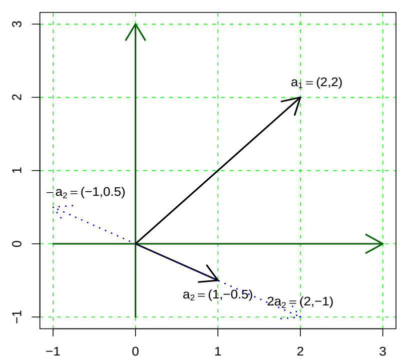

Figure 2: Multiplication of a vector by a number

```
> 7 * a
[1]  7 21 14
```

2.4 Sum of vectors

Let a and b be n–vectors. The sum $a + b$ is the n–vector

$$a + b = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{bmatrix} = b + a$$

See Figure 3 and 4. Only vectors of the same dimension can be added.

Example 3

$$\begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 8 \\ 9 \end{bmatrix} = \begin{bmatrix} 1+2 \\ 3+8 \\ 2+9 \end{bmatrix} = \begin{bmatrix} 3 \\ 11 \\ 11 \end{bmatrix}$$

□

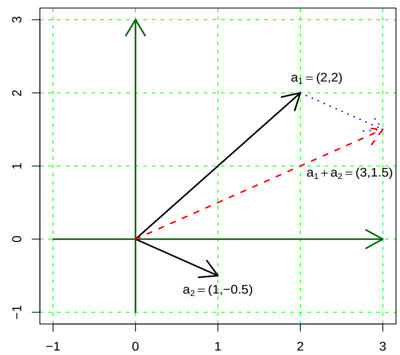

Figure 3: Addition of vectors

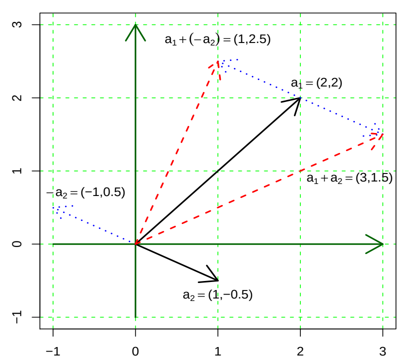

Figure 4: Addition of vectors and multiplication by a number

```
> a <- c(1, 3, 2)
> b <- c(2, 8, 9)
> a+b

[1]  3 11 11
```

2.5 Inner product of vectors

Let $a = (a_1, \dots, a_n)$ and $b = (b_1, \dots, b_n)$. The inner product of a and b is

$$a \cdot b = a_1 b_1 + \cdots + a_n b_n$$

Note, that the inner product is a number – not a vector.

```
> sum(a * b)
```

```
[1] 44
```

2.6 The length (norm) of a vector

The length (or norm) of a vector a is

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

```
> sqrt(sum(a * a))
```

```
[1] 3.741657
```

2.7 The 0–vector and 1–vector

The 0-vector (1–vector) is a vector with 0 (1) on all entries. The 0–vector (1–vector) is frequently written simply as 0 (1) or as 0_n (1_n) to emphasize that its length n.

```
> rep(0, 5)
```

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

```
> rep(1, 5)
```

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

2.8 Orthogonal (perpendicular) vectors

Two vectors v_1 and v_2 are orthogonal if and only if their inner product is zero, written

$$v_1 \perp v_2 \Leftrightarrow v_1 \cdot v_2 = 0$$

Note that any vector is orthogonal to the 0–vector.

```
> v1 <- c(1, 1)
> v2 <- c(-1, 1)
> sum(v1 * v2)
```

```
[1] 0
```

3 Matrices

3.1 Matrices

An $r \times c$ matrix A (reads "an r times c matrix") is a table with r rows og c columns

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1c} \\ a_{21} & a_{22} & \dots & a_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \dots & a_{rc} \end{bmatrix}$$

Note that one can regard A as consisting of c columns vectors put after each other:

$$A = [a_1 : a_2 : \cdots : a_c]$$

Likewise one can regard A as consisting of r row vectors stacked on to of each other.

```
> A <- matrix(c(1, 3, 2, 2, 8, 9), ncol=3)
> A

     [,1] [,2] [,3]
[1,]    1    2    8
[2,]    3    2    9
```

Note that the numbers $1, 3, 2, 2, 8, 9$ are read into the matrix column–by–column. To get the numbers read in row–by–row do

```
> A2 <- matrix(c(1, 3, 2, 2, 8, 9), ncol=3, byrow=T)
> A2

     [,1] [,2] [,3]
[1,]    1    3    2
[2,]    2    8    9
```

3.2 Multiplying a matrix with a number

For a number α and a matrix A, the product αA is the matrix obtained by multiplying each element in A by α.

Example 4

$$7 \begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 9 \end{bmatrix} = \begin{bmatrix} 7 & 14 \\ 21 & 56 \\ 14 & 63 \end{bmatrix}$$

□

```
> 7 * A

     [,1] [,2] [,3]
[1,]    7   14   56
[2,]   21   14   63
```

3.3 Transpose of matrices

A matrix is transposed by interchanging rows and columns and is denoted by "$^\top$".

Example 5

$$\begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 9 \end{bmatrix}^\top = \begin{bmatrix} 1 & 3 & 2 \\ 2 & 8 & 9 \end{bmatrix}$$

□

Note that if A is an $r \times c$ matrix then $A^\top$ is a $c \times r$ matrix.

```
> t(A)

     [,1] [,2]
[1,]    1    3
[2,]    2    2
[3,]    8    9
```

3.4 Sum of matrices

Let A and B be $r \times c$ matrices. The sum $A + B$ is the $r \times c$ matrix obtained by adding A and B elementwise.

Only matrices with the same dimensions can be added.

Example 6

$$\begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 9 \end{bmatrix} + \begin{bmatrix} 5 & 4 \\ 8 & 2 \\ 3 & 7 \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 11 & 10 \\ 5 & 16 \end{bmatrix}$$

□

```
> B <- matrix(c(5, 8, 3, 4, 2, 7), ncol=3, byrow=T)
> A + B

     [,1] [,2] [,3]
[1,]    6   10   11
[2,]    7    4   16
```

3.5 Multiplication of a matrix and a vector

Let A be an $r \times c$ matrix and let b be a c-dimensional column vector. The product Ab is the $r \times 1$ matrix

$$Ab = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1c} \\ a_{21} & a_{22} & \dots & a_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ a_{r1} & a_{r2} & \dots & a_{rc} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_c \end{bmatrix} = \begin{bmatrix} a_{11}b_1 + a_{12}b_2 + \cdots + a_{1c}b_c \\ a_{21}b_1 + a_{22}b_2 + \cdots + a_{2c}b_c \\ \vdots \\ a_{r1}b_1 + a_{r2}b_2 + \cdots + a_{rc}b_c \end{bmatrix}$$

Example 7

$$\begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 9 \end{bmatrix} \begin{bmatrix} 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 1 \cdot 5 + 2 \cdot 8 \\ 3 \cdot 5 + 8 \cdot 8 \\ 2 \cdot 5 + 9 \cdot 8 \end{bmatrix} = \begin{bmatrix} 21 \\ 79 \\ 82 \end{bmatrix}$$

□

```
> A %*% a

     [,1]
[1,]   23
[2,]   27
```

Note the difference to

```
> A * a

     [,1] [,2] [,3]
[1,]    1    4   24
[2,]    9    2   18
```

Please figure out yourself what goes on!

3.6 Multiplication of matrices

Let A be an $r \times c$ matrix and B a $c \times t$ matrix, i.e. $B = [b_1 : b_2 : \cdots : b_t]$. The product AB is the $r \times t$ matrix given by:

$$AB = A[b_1 : b_2 : \cdots : b_t] = [Ab_1 : Ab_2 : \cdots : Ab_t]$$

Example 8

$$\begin{aligned} \begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 9 \end{bmatrix} \begin{bmatrix} 5 & 4 \\ 8 & 2 \end{bmatrix} &= \left[\begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 9 \end{bmatrix} \begin{bmatrix} 5 \\ 8 \end{bmatrix} : \begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 9 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} \right] \\ &= \begin{bmatrix} 1 \cdot 5 + 2 \cdot 8 & 1 \cdot 4 + 2 \cdot 2 \\ 3 \cdot 5 + 8 \cdot 8 & 3 \cdot 4 + 8 \cdot 2 \\ 2 \cdot 5 + 9 \cdot 8 & 2 \cdot 4 + 9 \cdot 2 \end{bmatrix} = \begin{bmatrix} 21 & 8 \\ 79 & 28 \\ 82 & 26 \end{bmatrix} \end{aligned}$$

□

Note that the product AB can only be formed if the number of rows in B and the number of columns in A are the same. In that case, A and B are said to be conforme.

In general AB and BA are not identical.

A *mnemonic for matrix multiplication* is :

$$\begin{bmatrix} 1 & 2 \\ 3 & 8 \\ 2 & 9 \end{bmatrix} \begin{bmatrix} 5 & 4 \\ 8 & 2 \end{bmatrix} = \begin{array}{cc|cc} & & 5 & 4 \\ & & 8 & 2 \\ \hline 1 & 2 & 1 \cdot 5 + 2 \cdot 8 & 1 \cdot 4 + 2 \cdot 2 \\ 3 & 8 & 3 \cdot 5 + 8 \cdot 8 & 3 \cdot 4 + 8 \cdot 2 \\ 2 & 9 & 2 \cdot 5 + 9 \cdot 8 & 2 \cdot 4 + 9 \cdot 2 \end{array} = \begin{bmatrix} 21 & 8 \\ 79 & 28 \\ 82 & 26 \end{bmatrix}$$

```
> A <- matrix(c(1, 3, 2, 2, 8, 9), ncol=2)
> B <- matrix(c(5, 8, 4, 2), ncol=2)
> A %*% B

     [,1] [,2]
[1,]   21    8
[2,]   79   28
[3,]   82   26
```

3.7 Vectors as matrices

One can regard a column vector of length r as an $r \times 1$ matrix and a row vector of length c as a $1 \times c$ matrix.

3.8 Some special matrices

- An $n \times n$ matrix is a *square matrix*
- A matrix A is *symmetric* if $A = A^\top$.
- A matrix with 0 on all entries is the *0–matrix* and is often written simply as 0.
- A matrix consisting of 1s in all entries is often written J.
- A square matrix with 0 on all off–diagonal entries and elements $d_1, d_2, \dots, d_n$ on the diagonal a *diagonal matrix* and is often written $diag\{d_1, d_2, \dots, d_n\}$
- A diagonal matrix with 1s on the diagonal is called the *identity matrix* and is denoted I. The identity matrix satisfies that $IA = AI = A$. Likewise, if x is a vector then $Ix = x$.

- 0-matrix and 1-matrix

```
> matrix(0, nrow=2, ncol=3)

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

> matrix(1, nrow=2, ncol=3)

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

- Diagonal matrix and identity matrix

```
> diag(c(1, 2, 3))

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

> diag(1, 3)
```

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

Note what happens when `diag` is applied to a matrix:

```
> diag(diag(c(1, 2, 3)))

[1] 1 2 3

> diag(A)

[1] 1 8
```

3.9 Inverse of matrices

In general, the inverse of an $n \times n$ matrix A is the matrix B (which is also $n \times n$) which when multiplied with A gives the identity matrix I. That is,

$$AB = BA = I.$$

One says that B is A's inverse and writes $B = A^{-1}$. Likewise, A is Bs inverse.

Example 9 Let

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \quad B = \begin{bmatrix} -2 & 1.5 \\ 1 & -0.5 \end{bmatrix}$$

Now $AB = BA = I$ so $B = A^{-1}$. □

Example 10 If A is a 1×1 matrix, i.e. a number, for example $A = 4$, then $A^{-1} = 1/4$. □

Some facts about inverse matrices are:

- Only square matrices can have an inverse, but not all square matrices have an inverse.
- When the inverse exists, it is unique.
- Finding the inverse of a large matrix A is numerically complicated (but computers do it for us).

Finding the inverse of a matrix in R is done using the `solve()` function:

```
> A <- matrix(c(1, 3, 2, 4), ncol=2,byrow=T)
> A

     [,1] [,2]
[1,]    1    3
[2,]    2    4

> #M2 <- matrix(c(-2,1.5,1,-0.5),ncol=2,byrow=T)
> B <- solve(A)
> B
```

```
     [,1] [,2]
[1,]   -2  1.5
[2,]    1 -0.5

> A %*% B

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

3.10 Solving systems of linear equations

Example 11 Matrices are closely related to systems of linear equations. Consider the two equations

$$\begin{aligned} x_1 + 3x_2 &= 7 \\ 2x_1 + 4x_2 &= 10 \end{aligned}$$

The system can be written in matrix form

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 10 \end{bmatrix} \quad \text{i.e. } Ax = b$$

Since $A^{-1}A = I$ and since $Ix = x$ we have

$$x = A^{-1}b = \begin{bmatrix} -2 & 1.5 \\ 1 & -0.5 \end{bmatrix} \begin{bmatrix} 7 \\ 10 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

A geometrical approach to solving these equations is as follows: Isolate x_2 in the equations:

$$x_2 = \frac{7}{3} - \frac{1}{3}x_1 \quad x_2 = \frac{1}{0}4 - \frac{2}{4}x_1$$

These two lines are shown in Figure 5 from which it can be seen that the solution is $x_1 = 1, x_2 = 2$.

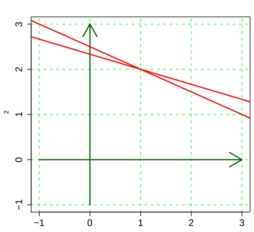

Figure 5: Solving two equations with two unknowns.

From the Figure it follows that there are 3 possible cases of solutions to the system

1. Exactly one solution – when the lines intersect in one point
2. No solutions – when the lines are parallel but not identical
3. Infinitely many solutions – when the lines coincide.

$\square$

```
> A <- matrix(c(1, 2, 3, 4), ncol=2)
> b <- c(7, 10)
> x <- solve(A) %*% b
> x

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

3.11 Some additional rules for matrix operations

For matrices A, B and C whose dimension match appropriately: the following rules apply

$$(A+B)^\top = A^\top + B^\top$$

$$(AB)^\top = B^\top A^\top$$

$$A(B+C) = AB + AC$$

$$AB = AC \not\Rightarrow B = C$$

In genereal $AB \neq BA$

$$AI = IA = A$$

If α is a number then $\alpha AB = A(\alpha B)$

3.12 Details on inverse matrices*

3.12.1 Inverse of a 2×2 matrix*

It is easy find the inverse for a 2×2 matrix. When

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then the inverse is

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

under the assumption that $ab - bc \neq 0$. The number $ab - bc$ is called the determinant of A, sometimes written $|A|$ or $det(A)$. A matrix A has an inverse if and only if $|A| \neq 0$.

3.12.2 Inverse of diagonal matrices*

Finding the inverse of a diagonal matrix is easy: Let

$$A = diag(a_1, a_2, \dots, a_n)$$

where all $a_i \neq 0$. Then the inverse is

$$A^{-1} = diag(\frac{1}{a_1}, \frac{1}{a_2}, \dots, \frac{1}{a_n})$$

If one $a_i = 0$ then A^{-1} does not exist.

3.12.3 Generalized inverse*

Not all square matrices have an inverse. However all square matrices have an infinite number of generalized inverses. A generalized inverse of a square matrix A is a matrix G satisfying that

$$AGA = A.$$

For many practical problems it suffice to find a generalized inverse.

```
> A <- matrix(c(1, 2, 3, 2, 3, 4, 3, 5, 7), ncol=3)
> A # 3rd column is sum of the two first; the inverse does not exist
```

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

```
> det(A)
```

```
[1] 0
```

```
> G <- MASS::ginv(A)
> A %*% G        # Not identity
```

```
           [,1]      [,2]       [,3]
[1,]  0.8333333 0.3333333 -0.1666667
[2,]  0.3333333 0.3333333  0.3333333
[3,] -0.1666667 0.3333333  0.8333333
```

```
> A %*% G %*% A  # This is A
```

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

3.12.4 Inverting an $n \times n$ matrix*

In the following we will illustrate one frequently applied methopd for matrix inversion. The method is called Gauss–Seidels method and many computer programs use variants of the method for finding the inverse of an $n \times n$ matrix.

Consider the matrix A:

```
> A <- matrix(c(2, 2, 3, 3, 5, 9, 5, 6, 7), ncol=3)
> A
```

```
     [,1] [,2] [,3]
[1,]    2    3    5
[2,]    2    5    6
[3,]    3    9    7
```

We want to find the matrix $B = A^{-1}$. To start, we append to A the identity matrix and call the result AB:

```
> AB <- cbind(A, diag(c(1, 1, 1)))
> AB

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

On a matrix we allow ourselves to do the following three operations (sometimes called elementary operations) as often as we want:

1. Multiply a row by a (non–zero) constant.
2. Multiply a row by a (non–zero) constant and add the result to another row.
3. Interchange two rows.

The aim is to perform such operations on `AB` in a way such that one ends up with a 3×6 matrix which has the identity matrix in the three leftmost columns. The three rightmost columns will then contain $B = A^{-1}$.

Recall that writing e.g. `AB[1,]` extracts the enire first row of AB.

- First, we make sure that `AB[1,1]=1`. Then we subtract a constant times the first row from the second to obtain that `AB[2,1]=0`, and similarly for the third row:

```
> AB[1,] <- AB[1,] / AB[1,1]
> AB[2,] <- AB[2,] - 2 * AB[1,]
> AB[3,] <- AB[3,] - 3 * AB[1,]
> AB

     [,1] [,2] [,3] [,4] [,5] [,6]
[1,]    1  1.5  2.5  0.5    0    0
[2,]    0  2.0  1.0 -1.0    1    0
[3,]    0  4.5 -0.5 -1.5    0    1
```

- Next we ensure that `AB[2,2]=1`. Afterwards we subtract a constant times the second row from the third to obtain that `AB[3,2]=0`:

```
> AB[2,] <- AB[2,] / AB[2,2]
> AB[3,] <- AB[3,] - 4.5 * AB[2,]
```

- Now we rescale the third row such that `AB[3,3]=1`:

```
> AB[3,] <- AB[3,] / AB[3,3]
> AB

     [,1] [,2] [,3]       [,4]      [,5]       [,6]
[1,]    1  1.5  2.5  0.5000000 0.0000000  0.0000000
[2,]    0  1.0  0.5 -0.5000000 0.5000000  0.0000000
[3,]    0  0.0  1.0 -0.2727273 0.8181818 -0.3636364
```

  Then `AB` has zeros below the main diagonal.

- We then work our way up to obtain that `AB` has zeros above the main diagonal:

```
> AB[2,] <- AB[2,] - 0.5 * AB[3,]
> AB[1,] <- AB[1,] - 2.5 * AB[3,]
> AB

     [,1] [,2] [,3]       [,4]        [,5]       [,6]
[1,]    1  1.5    0  1.1818182 -2.04545455  0.9090909
[2,]    0  1.0    0 -0.3636364  0.09090909  0.1818182
[3,]    0  0.0    1 -0.2727273  0.81818182 -0.3636364

> AB[1,] <- AB[1,] - 1.5 * AB[2,]
> AB

     [,1] [,2] [,3]       [,4]        [,5]       [,6]
[1,]    1    0    0  1.7272727 -2.18181818  0.6363636
[2,]    0    1    0 -0.3636364  0.09090909  0.1818182
[3,]    0    0    1 -0.2727273  0.81818182 -0.3636364
```

Now we extract the three rightmost columns of AB into the matrix B. We claim that B is the inverse of A, and this can be verified by a simple matrix multiplication

```
> B <- AB[,4:6]
> A%*% B

              [,1]         [,2]         [,3]
[1,]  1.000000e+00 3.330669e-16 1.110223e-16
[2,] -4.440892e-16 1.000000e+00 2.220446e-16
[3,] -2.220446e-16 9.992007e-16 1.000000e+00
```

So, apart from rounding errors, the product is the identity matrix, and hence $B = A^{-1}$. This example illustrates that numerical precision and rounding errors is an important issue when making computer programs.

4 Least squares

Consider the table of pairs (x_i, y_i) below.

x	1.00	2.00	3.00	4.00	5.00
y	3.70	4.20	4.90	5.70	6.00

A plot of y_i against x_i is shown in Figure 6.

The plot in Figure 6 suggests an approximately linear relationship between y and x, i.e.

$$y_i = \beta_0 + \beta_1 x_i \text{ for } i = 1, \dots, 5$$

Writing this in matrix form gives

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_5 \end{bmatrix} \approx \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_5 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \boldsymbol{X\beta}$$

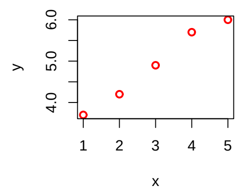

Figure 6: Regression

The first question is: Can we find a vector β such that $y = X\beta$? The answer is clearly no, because that would require the points to lie exactly on a straight line.

A more modest question is: Can we find a vector $\hat{\beta}$ such that $X\hat{\beta}$ is in a sense "as close to y as possible". The answer is yes. The task is to find $\hat{\beta}$ such that the length of the vector

$$e = y - X\beta$$

is as small as possible. This leads to the so–called system of normal equations

$$(X^\top X)\beta = X^\top y$$

If $(X^\top X)$ is invertible, the (unique) solution is

$$\hat{\beta} = (X^\top X)^{-1} X^\top y$$

```
> y

[1] 3.7 4.2 4.9 5.7 6.0

> X


       x
[1,] 1 1
[2,] 1 2
[3,] 1 3
[4,] 1 4
[5,] 1 5

> beta.hat <- solve(t(X) %*% X) %*% t(X) %*% y
> beta.hat

  [,1]
  3.07
x 0.61
```

The fitted values are

```
> as.numeric(X %*% beta.hat)

[1] 3.68 4.29 4.90 5.51 6.12
```

4.1 Least squares with generalized inverse

Expand the setting above: Let X_2 be X with an extra column added: The sum of the columns of X.

```
> X2 <- cbind(X, rowSums(X))
> X2
```

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

Then $(X_2^\top X_2)$ is not invertible. There are infinitely many solutions to the normal equations. One is:

```
> G <- MASS::ginv(t(X2) %*% X2)
> G
```

```
           [,1]       [,2]        [,3]
[1,]  0.6333333 -0.4333333  0.20000000
[2,] -0.4333333  0.3000000 -0.13333333
[3,]  0.2000000 -0.1333333  0.06666667
```

```
> beta2.hat <- G %*% t(X2) %*% y
> beta2.hat
```

```
           [,1]
[1,]  1.8433333
[2,] -0.6166667
[3,]  1.2266667
```

Another solution is (why?)

```
> beta22.hat <- c(beta.hat, 0)
```

The fitted values are the same:

```
> as.numeric(X2 %*% beta2.hat)
```

```
[1] 3.68 4.29 4.90 5.51 6.12
```

```
> as.numeric(X2 %*% beta22.hat)
```

```
[1] 3.68 4.29 4.90 5.51 6.12
```

5 A neat little exercise – from a bird's perspective

On a sunny day, two tables are standing in an English country garden. On each table birds of unknown species are sitting having the time of their lives.

A bird from the first table says to those on the second table: "Hi – if one of you come to our table then there will be the same number of us on each table". "Yeah, right", says a bird from the second table, "but if one of you comes to our table, then we will be twice as many on our table as on yours".

Question: How many birds are on each table? More specifically,

- Write up two equations with two unknowns.
- Solve these equations using the methods you have learned from linear algebra.
- Simply finding the solution by trial–and–error is considered cheating.