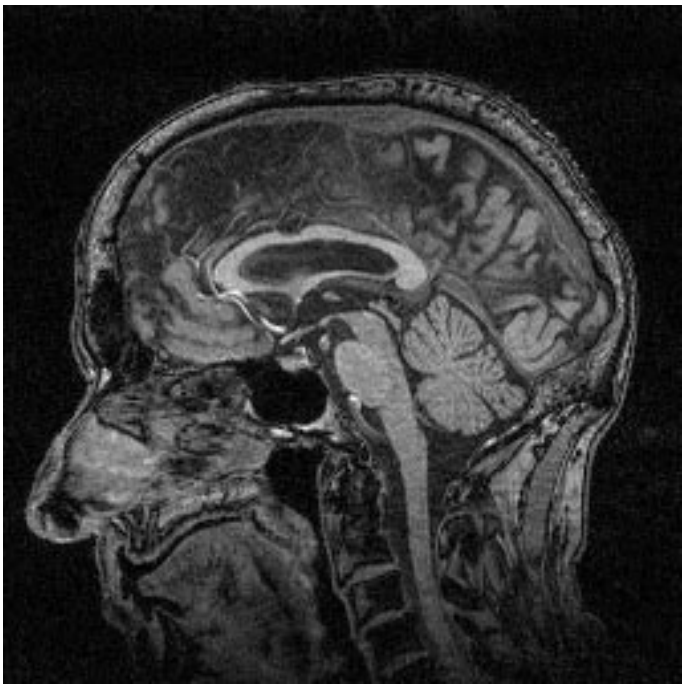
DTU Informatics
Department of Informatics and Mathematical Modelling

Hans Bruun Nielsen

LINEAR ALGEBRA
for
IT & HEALTH



September 2010

DTU Informatics – IMM
Richard Petersens Plads
DTU – building 321
DK-2800 Kgs. Lyngby

Phone: +45 4525 3351
www.imm.dtu.dk

Preface

This lecture note has been written for the teaching in the course *Modelling and Programming.* The aim has been to give an easy, but sufficient introduction to the subject of linear algebra, leaving out the parts that are not strictly necessary.

I wish to thank Jeppe Revall Frisvad for his careful reading of the note and many useful suggestions. Also, I appreciate the fruitful discussions with Peter Dalgaard and Rasmus Reinhold Paulsen.

Hans Bruun Nielsen

The front page illustration is discussed in Example 1.14 on page 8.

Contents

1. Introduction

1.1. Vectors and matrices

A *vector* $\boldsymbol{x} \in \mathbb{R}^n$ (also called an n-vector) is a list of n elements, $x_1, x_2, \ldots, x_n$, all of which are real numbers: $x_i \in \mathbb{R}$.

Example 1.1. In the usual Cartesian system a *geometric vector* $\overrightarrow{v}$ can be identified by the coordinates (v_1, v_2), cf the figure. This pair of coordinates is a vector in $\mathbb{R}^2$.

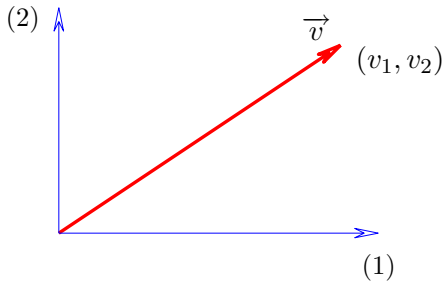


Figure 1.1. *Geometric vector.*

Conversely, one may visualize a vector $\boldsymbol{x} \in \mathbb{R}^2$ by a geometric vector $\overrightarrow{x}$ in the plane, having the coordinates (x_1, x_2).

It may help intuition to generalize this idea: One may think of an n-vector as an "arrow" in the n-dimensional space. ∎

Example 1.2. During the development of a new drug you need to perform clinical experiments to investigate how fast the drug is excreted from the human body. This investigation may consist in measuring the drug concentration in the kidneys at certain times after taking the drug. The results may be represented by two vectors, $\boldsymbol{t}$ and $\boldsymbol{y}$, where y_i is the concentration at time t_i.

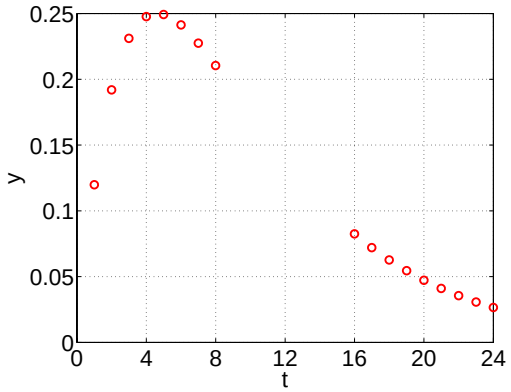


Figure 1.2. *Concentration y as function of time t.*

The figure is a MATLAB plot of the results of such an experiment. Time is given in hours after the drug was taken. It seems that there was an 8 hour break during the experiment.

We shall return to this problem several times, first in Example 1.5. In Chapter 3 we discuss a mathematical model for the behaviour shown in Figure 1.2. ∎

The *zero vector* $\mathbf{0}$ has all its elements equal to zero. We say that $\boldsymbol{x}$ is nonzero if it has at least one element different from zero.

A *matrix* $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ (also called an $m \times n$ matrix) is a list of $m \cdot n$ real numbers, organized in m rows, each of which has n elements. Alternatively we can think of $\boldsymbol{A}$ as organized in n columns, with m elements in each. The element in position(i, j) (ie the j^{th} element in the i^{th} row) is denoted a_{ij} or $(\boldsymbol{A})_{ij}$. The *diagonal* of $\boldsymbol{A}$ consists of the elements $a_{ii},\ i = 1, 2, \ldots, \min\{m, n\}$.

Example 1.3. Consider the 2×3 matrix

$$\boldsymbol{A} \;=\; \begin{pmatrix} 1.2 & 3.4 & 5.6 \\ 7.8 & 9.1 & -2.3 \end{pmatrix} .$$

We have, for instance, $a_{12} = 3.4$ and $a_{23} = -2.3$. The diagonal elements are $a_{11} = 1.2$ and $a_{22} = 9.1$. ∎

Note that $\mathbb{R}^{1 \times n}$ and $\mathbb{R}^{m \times 1}$ contain matrices with only one row and one column, respectively. Such degenerated matrices are equivalent to vectors, and we talk about a *row vector* if $\boldsymbol{x} \in \mathbb{R}^{1 \times n}$ and a *column vector* if $\boldsymbol{y} \in \mathbb{R}^{m \times 1}$. If nothing else is specified, $\boldsymbol{x} \in \mathbb{R}^n$ is considered to be a column vector.

The elements in the entire i^{th} row of the the matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a row vector, which we denote $\boldsymbol{A}_{i,:}$. Similarly, $\boldsymbol{A}_{:,j} \in \mathbb{R}^{m \times 1}$ consists of the elements in the j^{th} column of $\boldsymbol{A}$.

Example 1.4. For the matrix from Example 1.3, we see that

$$\boldsymbol{A}_{1,:} \;=\; \begin{pmatrix} 1.2 & 3.4 & 5.6 \end{pmatrix} , \qquad \boldsymbol{A}_{:,3} \;=\; \begin{pmatrix} 5.6 \\ -2.3 \end{pmatrix} .$$ ∎

Example 1.5. The results from the clinical experiment in Example 1.2 may be stored in a matrix $\boldsymbol{A}$, with the times and the measured concentrations in the first and second column, respectively: $\boldsymbol{A}_{:,1} = \boldsymbol{t},\ \boldsymbol{A}_{:,2} = \boldsymbol{y}$.

If the experiment involves 4 patients that are measured simultaneously, then the results may be stored in a matrix $\boldsymbol{A}$ with 5 columns, where $\boldsymbol{A}_{:,j+1}$ holds the y_i-values for the j^{th} person. ∎

A *square matrix* has the same number of rows and columns, $m = n$. A *diagonal matrix* is a square matrix where all the elements outside the diagonal are zero. The *identity matrix* $\boldsymbol{I}$ is a diagonal matrix with all diagonal elements equal to 1. A *lower triangular matrix* is a square matrix where all the elements above the diagonal are zero. Similarly, all the elements below the diagonal are zero in an *upper triangular matrix*.

Example 1.6. In the case $m = n = 3$ we have

$$\boldsymbol{I} \;=\; \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

As examples of a diagonal matrix and a lower and an upper triangular matrix we give

$$\boldsymbol{D} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad \boldsymbol{L} = \begin{pmatrix} 2 & 0 & 0 \\ 5 & -1 & 0 \\ 2 & 7 & 3 \end{pmatrix}, \quad \boldsymbol{U} = \begin{pmatrix} 2 & 3 & -4 \\ 0 & -1 & 7 \\ 0 & 0 & 3 \end{pmatrix}.$$

For the sake of better visual impression we often omit the zeros outside the diagonal in a diagonal matrix and above/below the diagonal in a lower/upper triangular matrix, for instance

$$\boldsymbol{I} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix}, \quad \boldsymbol{U} = \begin{pmatrix} 2 & 3 & -4 \\ & -1 & 7 \\ & & 3 \end{pmatrix}.$$

■

Example 1.7. MATLAB is an acronym for MATrix LABoratory. The basic storage element is a matrix. Let `A` be a MATLAB variable, then the command `[m,n] = size(A)` returns `m`, the number of rows in `A`, and `n`, the number of columns in `A`.
If $\mathtt{m} = 0$ or $\mathtt{n} = 0$, then `A` is empty.
If $\mathtt{m} = \mathtt{n} = 1$, then `A` is a scalar.
If $\mathtt{m} = 1$ and $\mathtt{n} > 1$, then `A` is a row vector.
If $\mathtt{m} > 1$ and $\mathtt{n} = 1$, then `A` is a column vector.
If $\mathtt{m} > 1$ and $\mathtt{n} > 1$, then `A` is a proper matrix. ■

1.2. Basic vector arithmetic

1° *Transpose.* The transpose of the vector $\boldsymbol{x}$ is denoted $\boldsymbol{x}^T$. If $\boldsymbol{x}$ is a column vector, then $\boldsymbol{x}^T$ is the row vector with de same elements, and vice versa.

2° *Multiplication by a scalar.* Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Then

$$\boldsymbol{z} = \alpha \boldsymbol{x}$$

is the vector with elements $z_i = \alpha x_i,\ i = 1, 2, \ldots, n$.

3° *Addition.* Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$, and consider the sum

$$\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{y}\ .$$

This is meaningful only if $\boldsymbol{x}$ and $\boldsymbol{y}$ have the same number of elements, ie $m = n$. If one distinguishes between row and column vectors, the two vectors must also be of the same type, ie both of them must be either row vectors or column vectors. When these conditions are satisfied, $\boldsymbol{z}$ is a vector of the same type with elements $z_i = x_i + y_i,\ i = 1, 2, \ldots, n$.

4° *Multiplication of two vectors.* Let $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$ be column vectors. They can be multiplied in three different ways:

4*a*° *Inner product* (also called *scalar product* or *dot product*): This is defined only if the two vectors have the same number of elements. The inner product of two n-vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$\boldsymbol{x}^T \boldsymbol{y} = x_1 y_2 + x_2 y_2 + \cdots + x_n y_n\ . \tag{1.1}$$

Thus, the inner product is a real number, a *scalar*. It follows immediately from the definition that swapping the two vectors does not change their scalar product:

$$\boldsymbol{y}^T\boldsymbol{x} \;=\; \boldsymbol{x}^T\boldsymbol{y} \;.$$

The vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are said to be *orthogonal* if their scalar product is zero, $\boldsymbol{x}^T\boldsymbol{y} = 0$.

4*b*° The *outer product* of the column vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is written

$$\boldsymbol{x}\,\boldsymbol{y}^T \;.$$

This is a matrix with the number of rows given by the number of elements in $\boldsymbol{x}$, n, and the number of columns is given by the number of elements in $\boldsymbol{y}$, m. The elements in the matrix are

$$\left(\boldsymbol{x}\,\boldsymbol{y}^T\right)_{ij} \;=\; x_i y_j \quad \begin{cases} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, n \end{cases} \;.$$

In general $\boldsymbol{y}\,\boldsymbol{x}^T \neq \boldsymbol{x}\,\boldsymbol{y}^T$. The outer product may also be expressed row-wise or column-wise:

$$\begin{aligned} \left(\boldsymbol{x}\,\boldsymbol{y}^T\right)_{i,:} &= x_i\boldsymbol{y}^T \;, \qquad i = 1, 2, \ldots, n \;, \\ \left(\boldsymbol{x}\,\boldsymbol{y}^T\right)_{:,j} &= y_j\boldsymbol{x} \;, \qquad j = 1, 2, \ldots, m \;. \end{aligned}$$

4*c*° *Element-wise multiplication.* In some cases it is relevant to compute the vector defined by element-wise multiplication of two vectors with the same number of elements, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. We use the notation

$$\boldsymbol{z} \;=\; \boldsymbol{x} \otimes \boldsymbol{y} \;.$$

This is also a vector with n elements, given by $z_i = x_i y_i,\ i = 1, 2, \ldots, n$.

Example 1.8. Let

$$\boldsymbol{x} \;=\; \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \;, \qquad \boldsymbol{y} \;=\; \begin{pmatrix} 4 \\ -5 \\ 6 \end{pmatrix} \;.$$

Then

$$2\boldsymbol{x}^T \;=\; \begin{pmatrix} 2 & 4 & 6 \end{pmatrix} \;, \quad \boldsymbol{x}^T + \boldsymbol{y}^T \;=\; \begin{pmatrix} 5 & -3 & 9 \end{pmatrix} \;,$$

$$\boldsymbol{x}^T\boldsymbol{y} \;=\; 12 \;, \quad \boldsymbol{y}\,\boldsymbol{x}^T \;=\; \begin{pmatrix} 4 & 8 & 12 \\ -5 & -10 & -15 \\ 6 & 12 & 18 \end{pmatrix} \;, \quad \boldsymbol{x} \otimes \boldsymbol{y} \;=\; \begin{pmatrix} 4 \\ -10 \\ 18 \end{pmatrix} \;.$$

In MATLAB the inner, outer, and element-wise products of the vectors `x` and `y` are obtained by the commands

```
ip = x' * y,   op = x * y',   ep = x .* y
```

■

5° *Norm.* In many applications we are interested in being able to talk about the "length" of a vector, and we introduce the norm for that purpose. The scalar product of an n-vector $\boldsymbol{x}$ and itself is

$$\boldsymbol{x}^T\boldsymbol{x} \;=\; x_1^2 + x_2^2 + \cdots + x_n^2 \;.$$

This is a real, nonnegative number, and we define the *norm* of $\boldsymbol{x}$, $\|\boldsymbol{x}\|$, as

$$\|\boldsymbol{x}\| \;=\; \sqrt{\boldsymbol{x}^T\boldsymbol{x}}\;. \tag{1.2}$$

It is fairly easy to verify that $\|\boldsymbol{x}\|$ satisfies the following so-called *norm conditions*

$5a^\circ$ $\quad \|\boldsymbol{x}\| \;\geq 0 \quad$ for all $\boldsymbol{x}$,

$5b^\circ$ $\quad \|\boldsymbol{x}\| \;=\; 0 \quad\Leftrightarrow\quad \boldsymbol{x} \;=\boldsymbol{0}$,

$5c^\circ$ $\quad \|\alpha\boldsymbol{x}\| \;=\; |\alpha|\cdot\|\boldsymbol{x}\| \quad$ for all $\alpha\in\mathbb{R}$,

$5d^\circ$ $\quad \|\boldsymbol{x}+\boldsymbol{y}\| \;\leq\; \|\boldsymbol{x}\|+\|\boldsymbol{y}\| \quad$ for all $\boldsymbol{y}\in\mathbb{R}^n$.

Example 1.9. Let $\overrightarrow{x}$ and $\overrightarrow{y}$ be geometric vectors with Cartesian coordinates (x_1, x_2) and (y_1, y_2). Their *dot product* is

$$\overrightarrow{x}\cdot\overrightarrow{y} \;=\; x_1y_1 + x_2y_2\;,$$

and the length of $\overrightarrow{x}$ is

$$|\overrightarrow{x}| \;=\; \sqrt{x_1^2+x_2^2}\;.$$

If $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^2$ are the column vectors with the coordinates, we see that $\boldsymbol{x}^T\boldsymbol{y} = \overrightarrow{x}\cdot\overrightarrow{y}$ and $\|\boldsymbol{x}\| = |\overrightarrow{x}|$.

The norm $\|\boldsymbol{x}\|$ generalizes this concept of "length" to vectors with any number of elements.

The condition $5d^\circ$ is known is the *triangle inequality*. The reason is that if a triangle in the plane has two sides given by $\overrightarrow{x}$ and $\overrightarrow{y}$, then the length of the third side is $|\overrightarrow{x}+\overrightarrow{y}|$, and this cannot be larger than the sum of the lengths of the two other sides. ∎

1.3. Basic matrix arithmetic

1° *Transpose.* The transpose of the matrix $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ is the matrix $\boldsymbol{A}^T\in\mathbb{R}^{n\times m}$. It is obtained from $\boldsymbol{A}$ by interchanging rows and columns: $(\boldsymbol{A}^T)_{ij} = (\boldsymbol{A})_{ji}$.

2° *Multiplication by a scalar.* Let $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ and $\alpha\in\mathbb{R}$. Then

$$\boldsymbol{C} \;=\; \alpha\boldsymbol{A}$$

is the matrix with elements $c_{ij} = \alpha a_{ij},\ i = 1,2,\ldots,m;\ j = 1,2,\ldots,n$.

3° *Addition of two matrices.* This makes sense only, if the two matrices are of the same type, ie if they have the same number of rows and the same number of columns. If $\boldsymbol{A},\boldsymbol{B}\in\mathbb{R}^{m\times n}$, then

$$\boldsymbol{C} \;=\; \boldsymbol{A}+\boldsymbol{B}$$

is the matrix with $c_{ij} = a_{ij}+b_{ij},\ i = 1,2,\ldots,m;\ j = 1,2,\ldots,n$.

4° *Matrix–vector multiplication.* Let $\boldsymbol{A}\in\mathbb{R}^{m\times n}$ and $\boldsymbol{x}\in\mathbb{R}^{n\times 1}$, and consider the product

$$\boldsymbol{y} \;=\; \boldsymbol{A}\,\boldsymbol{x}\;.$$

The result is defined by the scalar products

$$y_i \;=\; \boldsymbol{A}_{i,:}\boldsymbol{x} \quad i = 1, 2, \ldots, m \;,$$

so the matrix–vector product is

$$\boldsymbol{A}\,\boldsymbol{x} \;=\; \begin{pmatrix} a_{11}x_1 \;+\; a_{12}x_2 \;+\; \cdots \;+\; a_{1n}x_n \\ a_{21}x_1 \;+\; a_{22}x_2 \;+\; \cdots \;+\; a_{2n}x_n \\ \vdots \\ a_{m1}x_1 \;+\; a_{m2}x_2 \;+\; \cdots \;+\; a_{mn}x_n \end{pmatrix} \;. \tag{1.3}$$

Example 1.10. With the matrix from Example 1.3 we get

$$\boldsymbol{A} \;=\; \begin{pmatrix} 1.2 & 3.4 & 5.6 \\ 7.8 & 9.1 & -2.3 \end{pmatrix} \;, \qquad \boldsymbol{A}^T \;=\; \begin{pmatrix} 1.2 & 7.8 \\ 3.4 & 9.1 \\ 5.6 & -2.3 \end{pmatrix} \;,$$

$$\boldsymbol{A}\begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix} \;=\; \begin{pmatrix} 1.2\cdot 2 + 3.4\cdot(-2) + 5.6\cdot 3 \\ 7.8\cdot 2 + 9.1\cdot(-2) - 2.3\cdot 3 \end{pmatrix} \;=\; \begin{pmatrix} 12.4 \\ -9.5 \end{pmatrix} \;,$$

$$\boldsymbol{A}^T\begin{pmatrix} -1 \\ 1 \end{pmatrix} \;=\; \begin{pmatrix} 1.2 - 7.8 \\ 3.4 - 9.1 \\ 5.6 + 2.3 \end{pmatrix} \;=\; \begin{pmatrix} 6.6 \\ 5.7 \\ -7.9 \end{pmatrix} \;.$$

If $\boldsymbol{D}$ is a diagonal matrix, it follows immediately from Eq. (1.3) that the vector $\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x}$ has the elements $y_i = d_{ii}x_i$. In particular, $\boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}$. ∎

Example 1.11. Consider the parabola

$$p(t) \;=\; c_1 t^2 + c_2 t + c_3 \;,$$

where c_1, c_2 and c_3 are known coefficients. From basic vector arithmetic it follows that we can express $p(t)$ by

$$p(t) \;=\; \begin{pmatrix} t^2 & t & 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \;.$$

Now, let $y_i = p(t_i)$, $i = 1, 2, 3$, with $t_i = i$. Then it follows that

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \;=\; \begin{pmatrix} t_1^2 & t_1 & 1 \\ t_2^2 & t_2 & 1 \\ t_3^2 & t_3 & 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \;=\; \begin{pmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} \;.$$

The figure illustrates the problem for a specific choice of the coefficients.

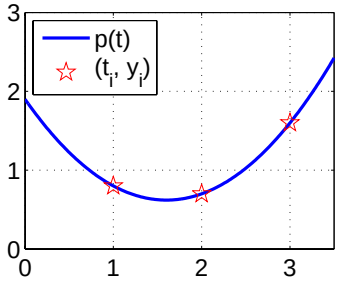


Figure 1.3. $p(t) = 0.5t^2 - 1.6t + 1.9$.

We return to this problem in Examples 2.1 and 2.6. ∎

5° *Matrix–matrix multiplication.* Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{B} \in \mathbb{R}^{p\times q}$, and consider the product

$$\boldsymbol{C} \;=\; \boldsymbol{A}\,\boldsymbol{B}\;.$$

This is defined by

$$(\boldsymbol{C})_{ij} \;=\; \boldsymbol{A}_{i,:}\boldsymbol{B}_{:,j}\;, \quad \begin{cases} i = 1, 2, \ldots, m \\ j = 1, 2, \ldots, q \end{cases} \quad . \tag{1.4}$$

In words: the $(i,j)^{\text{th}}$ element i $\boldsymbol{C}$ is the inner product of the i^{th} row in $\boldsymbol{A}$ and the j^{th} column in $\boldsymbol{B}$. The product exists only if these two vectors have the same number of elements, ie the number of columns in $\boldsymbol{A}$ must be equal to the number of rows i $\boldsymbol{B}$, $n = p$. The result can be expressed column-wise:

$$\boldsymbol{C}_{:,j} \;=\; \boldsymbol{A}\boldsymbol{B}_{:,j}, \quad j = 1, 2, \ldots, q\;, \tag{1.5}$$

ie the j^{th} column in $\boldsymbol{C}$ is the matrix-vector product of $\boldsymbol{A}$ and the j^{th} column in $\boldsymbol{B}$.

We shall sometimes use the following identity,

$$(\boldsymbol{A}\,\boldsymbol{B})^T \;=\; \boldsymbol{B}^T\boldsymbol{A}^T\;. \tag{1.6}$$

We already saw this relation in the case where the two matrices degenerate to a row and column matrix, respective, see Eq. (1.1). For general matrices it is easy to prove the identity by using the definition of the matrix–matrix product.

Example 1.12. If both $\boldsymbol{A}$ and $\boldsymbol{B}$ are square and have the same number of elements, then both $\boldsymbol{AB}$ and $\boldsymbol{BA}$ are defined; for instance

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

gives

$$\boldsymbol{A}\,\boldsymbol{B} \;=\; \begin{pmatrix} 9 & 22 \\ 13 & 50 \end{pmatrix}\;, \qquad \boldsymbol{B}\,\boldsymbol{A} \;=\; \begin{pmatrix} 13 & 14 \\ 31 & 46 \end{pmatrix}\;.$$

This illustrates, in general $\boldsymbol{AB} \neq \boldsymbol{BA}$. To illustrate Eq. (1.6) we see that

$$\boldsymbol{A}^T\boldsymbol{B}^T \;=\; \begin{pmatrix} 1 & 3 \\ 2 & 7 \end{pmatrix}\begin{pmatrix} -5 & 7 \\ 6 & 8 \end{pmatrix} \;=\; \begin{pmatrix} 13 & 31 \\ 14 & 46 \end{pmatrix}\;,$$

which is equal to $(\boldsymbol{B}\,\boldsymbol{A})^T$.

With the rectangular matrix from Example 1.3,

$$\boldsymbol{M} \;=\; \begin{pmatrix} 1.2 & 3.4 & 5.6 \\ 7.8 & 9.1 & -2.3 \end{pmatrix}\;,$$

the matrix products $\boldsymbol{AM}$ and $\boldsymbol{M}^T\boldsymbol{A}^T$ are defined, while $\boldsymbol{MA}$ is not defined.

$$\boldsymbol{AM} \;=\; \begin{pmatrix} 16.8 & 21.6 & 1 \\ 34.8 & 46.6 & 7.6 \end{pmatrix}\;, \quad \boldsymbol{M}^T\boldsymbol{A}^T \;=\; \begin{pmatrix} 16.8 & 34.8 \\ 21.6 & 46.6 \\ 1 & 7.6 \end{pmatrix}\;.$$

Again the result agrees with Eq. (1.6). ∎

Exercise 1.13. Let $\boldsymbol{A}$ be an $n\times n$ matrix and let $\boldsymbol{D}$ be an $n\times n$ diagonal matrix. Show that $\boldsymbol{B} = \boldsymbol{DA}$ has the rows

$$\boldsymbol{B}_{i,:} \;=\; d_{ii}\boldsymbol{A}_{i,:}, \quad i = 1, 2, \ldots, n \;.$$

Hint: Use Eq. (1.5) and Example 1.10.

Next, show that $\boldsymbol{C} = \boldsymbol{AD}$ has the columns

$$\boldsymbol{C}_{:,j} \;=\; d_{jj}\boldsymbol{A}_{:,j}, \quad j = 1, 2, \ldots, n \;.$$

Hint: Use Eq. (1.6) and the result for $\boldsymbol{DA}$. ∎

1.4. Linear mapping

Let $\boldsymbol{A}$ be an $m \times n$ matrix and let $\boldsymbol{x}$ be an n-vector. The m-vector $\boldsymbol{y} = \boldsymbol{Ax}$ is said to be obtained by a *linear mapping* from $\mathbb{R}^n$ into $\mathbb{R}^m$, cf Figure 1.4 below. We say

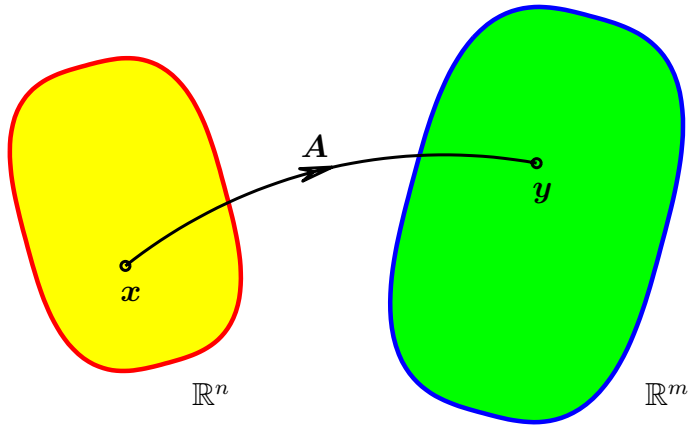


Figure 1.4. *Linear mapping,* $\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x}$.

that $\boldsymbol{y}$ is the *image* of $\boldsymbol{x}$ and that $\boldsymbol{A}$ is the *mapping matrix*.

Example 1.14. Linear mappings are widely used in mathematical models. In *CT scanning* or *MR scanning*, for instance, you get a vector $\boldsymbol{b}$ of measurements, which do not directly make sense. However, you can construct a matrix $\boldsymbol{A}$ such that for a given distribution of mass density $\boldsymbol{d}$ in the scanned object, you would measure $\boldsymbol{y} = \boldsymbol{Ad}$. The problem is to find the distribution $\boldsymbol{x}$ such that $\boldsymbol{Ax} = \boldsymbol{b}$. On the front page we show such a picture with $\boldsymbol{b}$ obtained by MR scanning.

The solution of the problem $\boldsymbol{Ax} = \boldsymbol{b}$ is the subject of the next chapter. First, however, we need to discuss some properties of linear mappings. ∎

Using the definition (1.3) of the matrix–vector product and the rules for vector arithmetic, we see that the relation

$$\boldsymbol{A}(\alpha\boldsymbol{u} + \beta\boldsymbol{v}) \;=\; \alpha\big(\boldsymbol{A}\,\boldsymbol{u}\big) + \beta\big(\boldsymbol{A}\,\boldsymbol{v}\big) \tag{1.7}$$

is true for all choices of n-vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and scalars α, β. This relation is known as the *linearity condition*.

In the special case where $\alpha = \beta = 0$, it follows that

$$\boldsymbol{A}\,\boldsymbol{0}_{[n]} \;=\; \boldsymbol{0}_{[m]} \;, \tag{1.8}$$

where the indices $[n]$ and $[m]$ are used to indicate that the zero vectors are in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. In words: a linear mapping maps the zero vector into the zero vector.

Example 1.15. Let

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \qquad \boldsymbol{z} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \qquad \boldsymbol{w} = \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}.$$

Using Eq. (1.3) we get

$$\boldsymbol{A}\,\boldsymbol{z} = \begin{pmatrix} 8 \\ 20 \\ 32 \end{pmatrix}, \qquad \boldsymbol{A}\,\boldsymbol{w} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

■

This example shows that for some linear mappings there are nonzero vectors $\boldsymbol{x}$ for which $\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{0}$. In order to get a better understanding of this we look at the definition (1.3) of the *matrix–vector product*,

$$\boldsymbol{A}\,\boldsymbol{x} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}.$$

From the rules for vector arithmetic it follows that an equivalent formulation of the right-hand side is

$$\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{x} = x_1\boldsymbol{A}_{:,1} + x_2\boldsymbol{A}_{:,2} + \cdots + x_n\boldsymbol{A}_{:,n}\,. \tag{1.9}$$

We say that $\boldsymbol{y}$ is a *linear combination* of the vectors $\boldsymbol{A}_{:,1}, \boldsymbol{A}_{:,2}, \ldots, \boldsymbol{A}_{:,n}$.

If $\boldsymbol{x} = \boldsymbol{0}$, we naturally get $\boldsymbol{y} = \boldsymbol{0}$, and if $\boldsymbol{y} \neq \boldsymbol{0}$ for all nonzero $\boldsymbol{x}$ (ie a vector with at least one nonzero element), then the vectors $\boldsymbol{A}_{:,1}, \boldsymbol{A}_{:,2}, \ldots, \boldsymbol{A}_{:,n}$ (the columns in $\boldsymbol{A}$) are said to be *linearly independent*. If there exists a nonzero $\boldsymbol{x}$ such that the image $\boldsymbol{y} = \boldsymbol{0}$, then the columns in $\boldsymbol{A}$ are said to be *linearly dependent*. Assume, for instance, that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ for a vector $\boldsymbol{x}$ with $x_1 \neq 0$. Then Eq. (1.9) implies that

$$\boldsymbol{A}_{:,1} = -\frac{x_2}{x_1}\boldsymbol{A}_{:,2} - \cdots - \frac{x_n}{x_1}\boldsymbol{A}_{:,n}\,.$$

This means that the first column in $\boldsymbol{A}$ is a linear combination of the other columns. In general, if the columns in $\boldsymbol{A}$ are linearly dependent, then one or more of the columns can be expressed as a linear combination of the other columns. The selection of which column to express in terms of the others is normally not unique.

Example 1.16. If, for instance, $\boldsymbol{A}_{:,k} = \boldsymbol{0}$, then the columns in $\boldsymbol{A}$ are linearly dependent: we may take $x_k = 1$ and all other $x_j = 0$.

Two nonzero vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent if and only if they are proportional, ie there is a real number β such that $\boldsymbol{u} = \beta\boldsymbol{v}$.

For the matrix from Example 1.15,

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix},$$

we can select two columns in three different ways (neglecting the ordering):

$$\begin{pmatrix} 1 & 2 \\ 4 & 5 \\ 7 & 8 \end{pmatrix} , \qquad \begin{pmatrix} 1 & 3 \\ 4 & 6 \\ 7 & 9 \end{pmatrix} , \qquad \begin{pmatrix} 2 & 3 \\ 5 & 6 \\ 8 & 9 \end{pmatrix} .$$

In all three cases the two columns are linearly independent.

In Example 1.15 we saw that

$$\boldsymbol{A}_{:,1} - 2\boldsymbol{A}_{:,2} + \boldsymbol{A}_{:,3} \;=\; \boldsymbol{0} \; .$$

This means that

$$\boldsymbol{A}_{:,1} = 2\boldsymbol{A}_{:,2} + \boldsymbol{A}_{:,3} \; , \quad \boldsymbol{A}_{:,2} = \tfrac{1}{2}\boldsymbol{A}_{:,1} + \tfrac{1}{2}\boldsymbol{A}_{:,3} \; , \quad \boldsymbol{A}_{:,3} = 2\boldsymbol{A}_{:,2} - \boldsymbol{A}_{:,1} \; .$$

These relations are easily verified. ∎

A linear mapping from $\mathbb{R}^n$ into $\mathbb{R}^m$ is sometimes confused with an *affine mapping* from $\mathbb{R}^n$ into $\mathbb{R}^m$. In this case the image $\boldsymbol{y}$ is given by

$$\boldsymbol{y} \;=\; \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \; , \tag{1.10}$$

where the m-vector $\boldsymbol{b}$ may be zero or nonzero. Obviously, the linearity condition Eq. (1.7) is not satisfied by an affine mapping with a nonzero $\boldsymbol{b}$.

1.5. Basis and Coordinates

Example 1.17. Let $\overrightarrow{b}_1$ and $\overrightarrow{b}_2$ be two geometric vectors in the plane. If they are linearly independent, ie they are not parallel, then we can write any $\overrightarrow{x}$ in the plane as a linear combination of $\overrightarrow{b}_1$ and $\overrightarrow{b}_2$,

$$\overrightarrow{x} \;=\; \tilde{x}_1 \overrightarrow{b}_1 + \tilde{x}_2 \overrightarrow{b}_2 \; .$$

The figure illustrates such a linear combination.

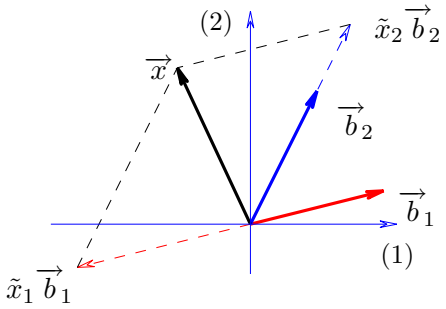


Figure 1.5. $\overrightarrow{x} = -1.3\,\overrightarrow{b}_1 + 1.5\,\overrightarrow{b}_2$.

∎

The observation in this example generalizes. If we know n linearly independent vectors in $\mathbb{R}^n$, $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n$, then any n-vector $\boldsymbol{x}$ can be written as a linear combination of these vectors:

$$\boldsymbol{x} \;=\; \tilde{x}_1\boldsymbol{b}_1 + \tilde{x}_2\boldsymbol{b}_2 + \cdots + \tilde{x}_n\boldsymbol{b}_n \; . \tag{1.11}$$

The vectors $\{\boldsymbol{b}_j\}$ are said to form a *basis* for $\mathbb{R}^n$, and the $\{\tilde{x}_j\}$ are the *coordinates* of $\boldsymbol{x}$ with respect to this basis.

The *usual basis* in $\mathbb{R}^n$ is the vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n$, where $\boldsymbol{e}_j = \boldsymbol{I}_{:,j}$, ie its j^{th} element is 1 and all the other elements are zero. The coordinates of a vector $\boldsymbol{x}$ with respect to this basis are simply the elements in $\boldsymbol{x}$.

Example 1.18. The usual basis in $\mathbb{R}^3$ is

$$\boldsymbol{e}_1 \;=\; \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \;, \quad \boldsymbol{e}_2 \;=\; \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \;, \quad \boldsymbol{e}_3 \;=\; \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \;.$$

The vector with coordinates 2, 3 and 4 is

$$\boldsymbol{x} \;=\; 2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \;+\; 3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \;+\; 4 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \;=\; \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} \;. \qquad \blacksquare$$

By comparison with Eq. (1.9) we see that Eq. (1.11) is equivalent to

$$\boldsymbol{x} \;=\; \boldsymbol{B}\,\tilde{\boldsymbol{x}} \;, \tag{1.12}$$

where $\tilde{\boldsymbol{x}}$ is the vector of coordinates with respect to the basis $\{\boldsymbol{b}_j\}$, and the matrix $\boldsymbol{B}$ has $\boldsymbol{b}_j$ as its j^{th} column, $\boldsymbol{B}_{:,j} = \boldsymbol{b}_j$.

Thus, given the coordinates, the vector $\boldsymbol{x} \in \mathbb{R}^n$ is obtained by a linear mapping of $\tilde{\boldsymbol{x}} \in \mathbb{R}^n$ with the mapping matrix $\boldsymbol{B}$. The *inverse problem* is: given $\boldsymbol{x}$, find $\tilde{\boldsymbol{x}}$. How to do that in the general case is the subject of the next chapter. In special cases, however, the inverse problem is simple:

If the basis vectors are *orthogonal*, ie

$$\boldsymbol{b}_i^T \boldsymbol{b}_j \;=\; 0 \quad \text{for all } i \neq j \;,$$

then it follows from Eq. (1.11) that

$$\begin{aligned} \boldsymbol{b}_k^T \boldsymbol{x} \;&=\; \tilde{x}_1 \left(\boldsymbol{b}_k^T \boldsymbol{b}_1\right) + \cdots + \tilde{x}_k \left(\boldsymbol{b}_k^T \boldsymbol{b}_k\right) + \cdots + \tilde{x}_n \left(\boldsymbol{b}_k^T \boldsymbol{b}_n\right) \\ &=\; \tilde{x}_k \left(\boldsymbol{b}_k^T \boldsymbol{b}_k\right) \;, \end{aligned}$$

showing that

$$\tilde{x}_k \;=\; \frac{\boldsymbol{b}_k^T \boldsymbol{x}}{\boldsymbol{b}_k^T \boldsymbol{b}_k} \;, \quad k = 1, 2, \ldots, n \;.$$

The computation is even easier if the basis is *orthonormal*. This means that the basis vectors are not only orthogonal, but also each of them has norm $\|\boldsymbol{b}_j\| = 1$, cf Eq. (1.2),

$$\boldsymbol{b}_i^T \boldsymbol{b}_j \;=\; \begin{cases} 0 & \text{if } \; i \neq j \;, \\ 1 & \text{if } \; i = j \;. \end{cases} \tag{1.13}$$

In this case the computation simplifies to

$$\tilde{x} \;=\; \boldsymbol{b}_k^T \boldsymbol{x} \;, \quad k = 1, 2, \ldots, n \;. \tag{1.14}$$

Example 1.19. It is simple to verify that the usual basis is orthonormal. Another example of an orthonormal basis in $\mathbb{R}^2$ is

$$\boldsymbol{b}_1 \;=\; \begin{pmatrix} \cos v \\ \sin v \end{pmatrix} \;, \quad \boldsymbol{b}_2 \;=\; \begin{pmatrix} -\sin v \\ \cos v \end{pmatrix} \;,$$

for any angle v. We easily see that the two vectors are nonzero and not proportional, so they form a basis in $\mathbb{R}^2$ for any choice of v. Next,

$$\boldsymbol{b}_1^T\boldsymbol{b}_2 \;=\; \cos v\cdot\sin v - \sin v\cdot\cos v \;=\; 0\ ,$$
$$\boldsymbol{b}_1^T\boldsymbol{b}_1 \;=\; \boldsymbol{b}_2^T\boldsymbol{b}_2 \;=\; \cos^2 v + \sin^2 v \;=\; 1\ .$$

Thus, the two vectors satisfy Eq. (1.13).

The matrix

$$\boldsymbol{B} \;=\; \begin{pmatrix} \cos v & -\sin v \\ \sin v & \cos v \end{pmatrix}$$

is known as a *rotation matrix*, cf the next example. ∎

Example 1.20. We shall demonstrate that a change of basis may simplify a problem: We want to find the points P: (x_1, x_2) that satisfy the quadratic equation

$$73x_1^2 - 72x_1x_2 + 52x_2^2 \;=\; 100\ .$$

The vectors

$$\boldsymbol{b}_1 \;=\; \begin{pmatrix} 0.6 \\ 0.8 \end{pmatrix}\ , \qquad \boldsymbol{b}_12 \;=\; \begin{pmatrix} -0.8 \\ 0.6 \end{pmatrix}$$

form an orthonormal basis in $\mathbb{R}^2$, and we introduce the coordinates $(\tilde{x}_1, \tilde{x}_2)$ with respect to this basis:

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \;=\; \begin{pmatrix} 0.6 & -0.8 \\ 0.8 & 0.6 \end{pmatrix}\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix}\ .$$

In accordance with this we use $0.6\tilde{x}_1 - 0.8\tilde{x}_2$ and $0.8\tilde{x}_1 + 0.6\tilde{x}_2$ to replace x_1 and x_2, respectively, in the quadratic equation:

$$\begin{aligned} 100 \;&=\; 73\,(0.6\tilde{x}_1 - 0.8\tilde{x}_2)^2 \\ &\quad -72\,(0.6\tilde{x}_1 - 0.8\tilde{x}_2)\,(0.8\tilde{x}_1 + 0.6\tilde{x}_2) \\ &\quad +52\,(0.8\tilde{x}_1 + 0.6\tilde{x}_2)^2 \\ &=\; 73\left(0.36\tilde{x}_1^2 - 0.96\tilde{x}_1\tilde{x}_2 + 0.64\tilde{x}_2^2\right) \\ &\quad -72\left(0.48\tilde{x}_1^2 - 0.28\tilde{x}_1\tilde{x}_2 - 0.48\tilde{x}_2^2\right) \\ &\quad +52\left(0.64\tilde{x}_1^2 + 0.96\tilde{x}_1\tilde{x}_2 - 0.36\tilde{x}_2^2\right) \\ &=\; 25\tilde{x}_1^2 + 100\tilde{x}_2^2\ . \end{aligned}$$

So, we got rid of the the term involving the product of the two elements in the vector. After dividing by 100 on both sides of this equation we get

$$\left(\frac{\tilde{x}_1}{2}\right)^2 + \left(\frac{\tilde{x}_2}{1}\right)^2 \;=\; 1\ .$$

This is the equation (with respect to the basis $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$) for points on an ellipse with half axes 2 and 1. In order to draw this curve we exploit that a point on the ellipse can be expressed by

$$\begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \end{pmatrix} \;=\; \begin{pmatrix} 2\cos u \\ \sin u \end{pmatrix}\ ,$$

for some value of the angle u. By letting u run though the interval $0 \le u \le 2\pi$ we can generate all points on the ellipse. This is the background for the following MATLAB code, which was used to plot the ellipse in the figure.

```
u = linspace(0,2*pi,201);
xt = [2*cos(u); sin(u)];
x = [0.6 -0.8; 0.8 0.6] * xt; % original coordinates
plot(x(1,:),x(2,:))
```

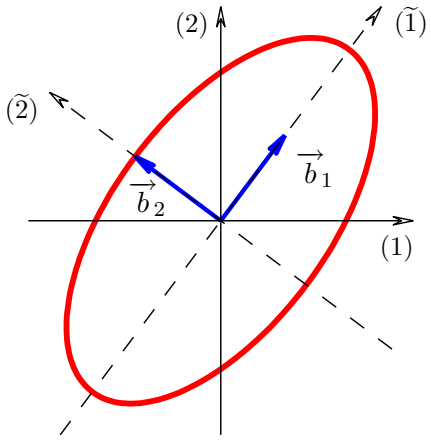


Figure 1.6. *Rotated ellipse.*

The matrix used in the relation between $\boldsymbol{x}$ and $\tilde{\boldsymbol{x}}$ is a rotation matrix

$$\boldsymbol{A} \;=\; \begin{pmatrix} \cos v & -\sin v \\ \sin v & \cos v \end{pmatrix} \;,$$

cf Example 1.19. The angle is defined by

$$\cos v \;=\; 0.6 \;, \quad \sin v \;=\; \sqrt{1-\cos^2 v} \;=\; 0.8 \;.$$

This corresponds to $v \simeq 53$ degrees, which is the angle between (1) and $\overrightarrow{b}_1$ in Figure 1.6.

In conclusion, the points satisfying the given quadratic are on an ellipse, which is rotated around its centre. The rotation angle is $v \simeq 53$ degrees.

In Example 3.9 we show how we found out that the specific choice of basis vectors $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ would simplify the quadratic equation. Also we show how we can avoid the tiresome calculations in the reformulation from $73x_1^2 - 72x_1x_2 + 52x_2^2 = 100$ to $25\tilde{x}_1^2 + 100\tilde{x}_2^2 = 100$. ∎

2. Linear Systems of Equations

A linear system of equations has the form

$$\begin{aligned} a_{11}x_1 \;+\; a_{12}x_2 \;+\cdots+\; a_{1n}x_n \;&=\; b_1 \\ a_{21}x_1 \;+\; a_{22}x_2 \;+\cdots+\; a_{2n}x_n \;&=\; b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \;&=\; b_m \end{aligned} \quad , \tag{2.1}$$

where the coefficients $\{a_{ij}\}$ and right-hand sides $\{b_i\}$ are given numbers. We seek the solution $x_1, x_2, \ldots, x_n$.

From the discussion of matrix–vector products in Section 1.3 it follows that we can write the system in the compact form

$$\boldsymbol{A}\boldsymbol{x} \;=\; \boldsymbol{b}\;, \tag{2.2}$$

where the *coefficient matrix* $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ has the elements a_{ij}, and the right-hand side vector $\boldsymbol{b}$ and the solution vector $\boldsymbol{x}$ contain respectively $\{b_i\}$ and $\{x_j\}$.

Example 2.1. Figure 2.1 shows three points (t_i, y_i) in the plane: $(1, 0.8)$, $(2, 0.7)$ and $(3, 1.6)$.

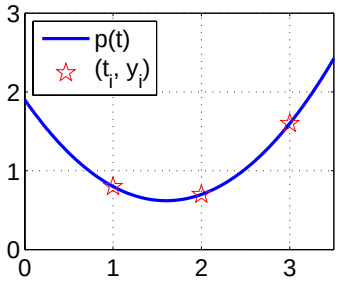


Figure 2.1. *Three given points on a parabola.*

We use a model which says that the points are on a parabola, $y_i = p(t_i)$, where

$$p(t) \;=\; x_1t^2 + x_2t + x_3\;.$$

However, we do not know the coefficients $x_1,\ x_2$ and x_3.

Using the same arguments as in Example 1.11, we see that the following equations must be satisfied,

$$t_i^2x_1 + t_ix_2 + x_3 \;=\; y_i\;,\quad i = 1, 2, 3\;.$$

This is three linear equations with the three unknowns $x_1,\ x_2$ and x_3. We can express the three equations in the form

$$\begin{aligned} t_1^2x_1 + t_1x_2 + x_3 \;&=\; y_1 \\ t_2^2x_1 + t_2x_2 + x_3 \;&=\; y_2 \\ t_3^2x_1 + t_3x_2 + x_3 \;&=\; y_3\;, \end{aligned}$$

or as $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$, where

$$\boldsymbol{A} = \begin{pmatrix} t_1^2 & t_1 & 1 \\ t_2^2 & t_2 & 1 \\ t_3^2 & t_3 & 1 \end{pmatrix}, \qquad \boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.$$

By inserting the given values for (t_i, y_i) we get the system

$$\begin{pmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0.8 \\ 0.7 \\ 1.6 \end{pmatrix}.$$

■

Example 2.2. In many textbooks the "determinant method" is used to solve a system of two linear equations with two unknowns,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 , \\ a_{12}x_1 + a_{22}x_2 &= b_2 . \end{aligned}$$

The *determinant* of the coefficient matrix $\boldsymbol{A}$ is defined by

$$\det \boldsymbol{A} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12} .$$

This is zero if and only if there is a scalar β such that

$$\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix} = \beta \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix},$$

ie, the two columns in $\boldsymbol{A}$ are linearly dependent, cf page 9. Such a matrix is said to be *singular*. This case is discussed in Section 2.3.

If $\det \boldsymbol{A} \neq 0$, we say that $\boldsymbol{A}$ is *nonsingular*, and the solution is

$$x_1 = \frac{a_{22}b_1 - a_{12}b_2}{\det \boldsymbol{A}}, \qquad x_2 = \frac{a_{11}b_2 - a_{21}b_1}{\det \boldsymbol{A}}.$$

This can be verified by inserting these expressions in the above system.

As a specific example, consider the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ with

$$\boldsymbol{A} = \begin{pmatrix} 2 & -4 \\ 2.5 & -2 \end{pmatrix}, \qquad \boldsymbol{b} = \begin{pmatrix} 4 \\ 6.5 \end{pmatrix}.$$

We get

$$\det \boldsymbol{A} = 2 \cdot (-2) - 2.5 \cdot (-4) = 6 ,$$

This is nonzero, so the matrix is nonsingular, and the solution is

$$x_1 = \frac{(-2) \cdot 4 - (-4) \cdot 6.5}{6} = 3 , \quad x_2 = \frac{2 \cdot 6.5 - 2.5 \cdot 4}{6} = 0.5 .$$

The determinant method may be generalized to larger systems, but as n grows it becomes very inefficient. ■

The solution of linear systems of equations occurs so often in scientific computation that it probably is the type of task that uses most computer effort. There are many efficient programs for the solution of the problem, and in surroundings like MATLAB or R there are simple commands that call such a program. In the next sections we give a brief introduction to the mathematical background of these programs. We first discuss some particularly simple cases, and then proceed to a generally applicable solution method.

2.1. Simple systems

The simplest case of a linear system of equations is when the coefficient matrix is *diagonal*, ie when the equations have the form

$$a_{ii}x_i \;=\; b_i \;, \quad i = 1, 2, \ldots, n \;.$$

Assuming that all the a_{ii} are nonzero, the solution is

$$x_i \;=\; b_i/a_{ii} \;, \quad i = 1, 2, \ldots, n \;.$$

Another simple case is when the system has the form

$$\boldsymbol{Q}\,\boldsymbol{x} \;=\; \boldsymbol{b} \;, \tag{2.3}$$

where $\boldsymbol{Q}$ is an *orthogonal matrix*. This means that the columns in $\boldsymbol{Q}$ are *orthonormal*:

$$\boldsymbol{Q}_{:,i}^T\boldsymbol{Q}_{:,j} \;=\; \begin{cases} 0 & \text{if } \; i \neq j \;, \\ 1 & \text{if } \; i = j \;. \end{cases} \;,$$

cf (1.13). The system (2.3) is equivalent to

$$x_1\boldsymbol{Q}_{:,1} + x_2\boldsymbol{Q}_{:,2} + \cdots + x_n\boldsymbol{Q}_{:,n} \;= \boldsymbol{b} \;,$$

and as we did in Section 1.5, we see that

$$x_k \;=\; \boldsymbol{Q}_{:,k}^T\boldsymbol{b} \;, \quad k = 1, 2, \ldots, n \;.$$

It is easy to verify that this means that the solution to the orthogonal system in Eq. (2.3) can be expressed as

$$\boldsymbol{x} \;=\; \boldsymbol{Q}^T\boldsymbol{b} \;, \tag{2.4}$$

ie the solution is found simply by multiplying the right-hand side $\boldsymbol{b}$ by the transposed of the coefficient matrix $\boldsymbol{Q}$. We shall use this several times in the remainder of the note.

Example 2.3. In Example 1.19 we showed that a rotation matrix

$$\boldsymbol{A} \;=\; \begin{pmatrix} \cos v & -\sin v \\ \sin v & \cos v \end{pmatrix} \;.$$

has orthonormal columns. This means that the vectors in the two columns can be used as basis in $\mathbb{R}^2$, and letting $\tilde{\boldsymbol{x}}$ denote the coordinates with respect to this basis, we see that

$$\boldsymbol{x} \;=\; \boldsymbol{A}\,\tilde{\boldsymbol{x}} \quad \Leftrightarrow \quad \tilde{\boldsymbol{x}} \;=\; \boldsymbol{A}^T\boldsymbol{x} \;.$$

∎

Finally, we consider *triangular systems*. This means that the coefficient matrix is triangular. We start by looking at $\boldsymbol{L}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{L}$ is a lower triangular matrix, so the system of equations has the form

$$\begin{aligned} \ell_{11}x_1 &= b_1 \\ \ell_{21}x_1 + \ell_{22}x_2 &= b_2 \\ &\;\;\vdots \\ \ell_{n1}x_1 + \ell_{n2}x_2 + \cdots + \ell_{nn}x_n &= b_n \end{aligned} \;.$$

The first equation has x_1 as the only unknown, and assuming that the coefficient $\ell_{11} \neq 0$, we get $x_1 = b_1/\ell_{11}$. We insert this value in the second equation and if $\ell_{22} \neq 0$, we get $x_2 = (b_2 - \ell_{21}x_1)/\ell_{22}$ The known values for x_1 and x_2 are inserted in the third equation, etc. We can summarize this in the form

$$\begin{aligned} & x_1 = b_1/\ell_{11} \\ & \textbf{for } i = 2, 3, \ldots, n \\ & \quad x_i = (b_i - \ell_{i1}x_1 - \cdots - \ell_{i,i-1}x_{i-1})/\ell_{ii} \\ & \textbf{end} \end{aligned} \tag{2.5}$$

This *algorithm* is called *forward substitution*. It can be used only if all $\ell_{ii} \neq 0$.

It is equally simple to solve the problem $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{c}$, where $\boldsymbol{U}$ is an upper triangular matrix,

$$\begin{aligned} u_{11}x_1 + u_{12}x_2 + \cdots + u_{1n}x_n &= c_1 \\ u_{22}x_2 + \cdots + u_{2n}x_n &= c_2 \\ &\vdots \\ u_{nn}x_n &= c_n \end{aligned} \quad .$$

We solve the last equation first and move up. Assuming that all $u_{ii} \neq 0$, we get

$$\begin{aligned} & x_n = c_n/u_{nn} \\ & \textbf{for } i = n{-}1, \ldots, 2, 1 \\ & \quad x_i = (c_i - u_{i,i+1}x_{i+1} - \cdots - u_{i,n}x_n)/u_{ii} \\ & \textbf{end} \end{aligned} \tag{2.6}$$

This algorithm is known as *back substitution*.

Example 2.4. Consider the upper triangular system

$$\begin{pmatrix} 9 & 3 & 1 \\ & \frac{2}{3} & \frac{5}{9} \\ & & \frac{3}{9} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 1.6 \\ \frac{-0.1}{9} \\ \frac{5.7}{9} \end{pmatrix} .$$

By means of Algorithm (2.6) we get

$$\begin{aligned} x_3 &= \tfrac{5.7}{9}/\tfrac{3}{9} = 1.9 , \\ x_2 &= (\tfrac{-0.1}{9} - \tfrac{5}{9} \cdot 1.9)/\tfrac{2}{3} = -1.6 , \\ x_1 &= (1.6 - 3 \cdot (-1.6) - 1 \cdot 1.9)/9 = 0.5 . \end{aligned}$$

■

2.2. Gaussian Elimination

The idea is to transform the given general, square system of equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ to an upper triangular system $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{c}$, which can then be solved by back substitution. The transformation must be made so that the solution is not changed, and this is ensured if we use a sequence of one or both of the two elementary operations

1° Interchange two equations.

2° Modify an equation by subtracting a multiple of one of the other equations.

Example 2.5. Given the system from Example 2.2

$$\begin{aligned} 2x_1 - 4x_2 &= 4 \\ 2.5x_1 - 2x_2 &= 6.5 \end{aligned}$$

We modify the second equation by subtracting 1.25 times the first equation:

$$2.5x_1 - 2x_2 - 1.25\,(2x_1 - 4x_2) \;=\; 6.5 - 1.25 \cdot 4\ .$$

This reduces to $3x_2 = 1.5$, so the original system is equivalent to the upper triangular system

$$\begin{aligned} 2x_1 - 4x_2 &= 4 \\ 3x_2 &= 1.5 \end{aligned}$$

By means of back substitution we get the same solution as in Example 2.2: $x_2 = 0.5$, $x_1 = 3$. Note that the *elimination factor* $1.25 = 2.5/2$ was chosen in order to eliminate x_2 from the modified equation.

Figure 2.2 gives a geometric interpretation: The two equations correspond to the straight lines $L_1:\ 2x_1 - 4x_2 = 4$ and $L_2:\ 2.5x_1 - 2x_2 = 6.5$, and the solution is the coordinates of the point where they intersect. The modification of the system corresponds to replacing L_2 by $\hat{L}_2:\ x_2 = 0.5$.

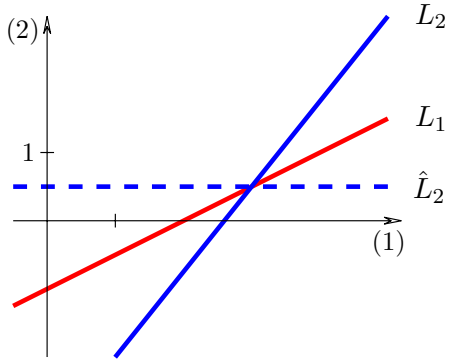


Figure 2.2. *Gaussian. elimination.*

Any linear equation with two unknowns and at least one nonzero coefficient can be interpreted as the equation for a straight line, and the solution to a 2×2 system is the the point of intersection between the two lines. The figure suggests possible complications:

L_1 may be parallel to the (1)-axis. This happens if $a_{11} = 0$, and this problem might be cured simply by interchanging the two equations, corresponding to renumbering the two lines.

The two lines may be parallel, in which case the system has no solution.

The two line may be coinciding, in which case there is infinitely many solutions: all points on the line satisfy both equations.

We will return to these problems in Section 2.3. ∎

Let us now consider a general $n\times n$ system $\boldsymbol{Ax} = \boldsymbol{b}$. It is convenient for the presentation (and for hand calculation) to introduce the so-called *augmented matrix* $\boldsymbol{T}$, which is the $n\times(n+1)$ matrix

$$\boldsymbol{T} \;=\; \left(\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} & b_n \end{array}\right) \;. \tag{2.7}$$

The vertical line is put there only to separate the coefficients and the elements of the right-hand side. Note that the i^{th} row $\boldsymbol{T}_{i,:}$ holds all information about the i^{th} equation. During computation the elements in $\boldsymbol{T}$ change, but in order to avoid a very complicated notation, we let a_{ij} denote the current value of the coefficient of x_j in the i^{th} equation, and b_i is the current value of the i^{th} right-hand side. Accordingly we use "$\leftarrow$", for instance $a \leftarrow a + b$, to indicate that the old value of a is replaced by $a+b$.

The first step in Gaussian elimination is to keep x_1 in one equation and eliminate it from all the others. This is done by first finding p_1, the number of the equation with the largest contribution from x_1,

$$|a_{p_1,1}| \;=\; \max_{i=1,2,\ldots,n} |a_{i1}| \tag{2.8}$$

If $a_{p_1,1} = 0$, then x_1 does not appear in any of the equations, and we proceed to the next step. Otherwise, if $p_1 \neq 1$, then interchange the 1st and the p_1th equation,

$$\boldsymbol{T}_{1,:} \;\leftrightarrow\; \boldsymbol{T}_{p_1,:} \;,$$

and modify equations

$$\left.\begin{array}{l} \ell_{i1} \;=\; a_{i1}/a_{11} \\ \boldsymbol{T}_{i,:} \;\leftarrow\; \boldsymbol{T}_{i,:} - \ell_{i1}\boldsymbol{T}_{1,:} \end{array}\right\} \quad i = 2, 3, \ldots, n \;.$$

After this step the original system has been changed to

$$\boldsymbol{T} \;=\; \left(\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ 0 & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & a_{n2} & \ldots & a_{nn} & b_n \end{array}\right) \;,$$

where all the a_{ij} and b_i may have values different from the original values. Note that the *active part* of of the system: rows $2,\ldots,n$, columns $2,\ldots,n,n+1$, has the same structure as the original system, but it is smaller.

The second step is similar to the first: Keep x_2 in one of the active equations and eliminate it from all the other active equations. This continues. For $k = 1,2,\ldots,n-1$ the active part of the system consists of rows $k,\ldots,n$ and columns $k,\ldots,n+1$. We keep x_k in one of the active equations (if possible) and delete it

from the others. The entire elimination algorithm can be expressed by

$$
\begin{array}{l}
\textbf{for } k = 1, 2, \ldots, n{-}1 \\
\quad \text{find } p_k : \quad |a_{p_k,k}| \;=\; \max\limits_{i=k+1,\ldots,n} |a_{i,k}| \\
\quad \textbf{if } \; a_{p_k,k} \neq 0 \; \textbf{ then} \\
\qquad \textbf{if } \; p_k \neq k \; \textbf{ then } \; \boldsymbol{T}_{k,:} \leftrightarrow \boldsymbol{T}_{p_k,:} \; \textbf{ end} \\
\qquad \textbf{for } i = k{+}1, \ldots, n \\
\qquad\quad \ell_{ik} = a_{ik}/a_{kk} \\
\qquad\quad \boldsymbol{T}_{i,:} \leftarrow \boldsymbol{T}_{i,:} - \ell_{ik}\boldsymbol{T}_{k,:} \\
\qquad \textbf{end} \\
\quad \textbf{end} \\
\textbf{end}
\end{array}
\tag{2.9}
$$

After executing this algorithm the augmented matrix has the structure

$$\boldsymbol{T} \;=\; \left(\boldsymbol{U} \;\middle|\; \boldsymbol{c}\right) \;,$$

where $\boldsymbol{U}$ is upper triangular and $\boldsymbol{c}$ contains the final values of the elements in the right-hand side vector.

Example 2.6. Consider the parabola problem from Example 2.1,

$$\begin{pmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \;=\; \begin{pmatrix} 0.8 \\ 0.7 \\ 1.6 \end{pmatrix} \;.$$

We use the augmented matrix notation, and indicate the operations that lead from one version of $\boldsymbol{T}$ to the next:

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0.8 \\ 4 & 2 & 1 & 0.7 \\ 9 & 3 & 1 & 1.6 \end{array}\right) \quad \begin{array}{c} \sim \\ p_1 = 3 \end{array} \quad \left(\begin{array}{ccc|c} 9 & 3 & 1 & 1.6 \\ 4 & 2 & 1 & 0.7 \\ 1 & 1 & 1 & 0.8 \end{array}\right) \quad \begin{array}{c} \sim \\ \ell_{21} = \frac{4}{9} \\ \ell_{31} = \frac{1}{9} \end{array}$$

$$\left(\begin{array}{ccc|c} 9 & 3 & 1 & 1.6 \\ 0 & \frac{2}{3} & \frac{5}{9} & -\frac{0.1}{9} \\ 0 & \frac{2}{3} & \frac{8}{9} & \frac{5.6}{9} \end{array}\right) \quad \begin{array}{c} \sim \\ p_2 = 2 \\ \ell_{3,2} = 1 \end{array} \quad \left(\begin{array}{ccc|c} 9 & 3 & 1 & 1.6 \\ 0 & \frac{2}{3} & \frac{5}{9} & \frac{-0.1}{9} \\ 0 & 0 & \frac{3}{9} & \frac{5.7}{9} \end{array}\right) \quad .$$

Thus, the solution $\boldsymbol{x}$ satisfies the upper triangular system

$$\begin{pmatrix} 9 & 3 & 1 \\ & \frac{2}{3} & \frac{5}{9} \\ & & \frac{3}{9} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \;=\; \begin{pmatrix} 1.6 \\ \frac{-0.1}{9} \\ \frac{5.7}{9} \end{pmatrix} \;.$$

We solved this system in Example 2.4. The solution is

$$\boldsymbol{x} \;=\; \begin{pmatrix} 0.5 & -1.6 & 1.9 \end{pmatrix}^T \;.$$

Note that this agrees with the coefficients in the parabola given in Figure 1.3, so the three given points identified the correct parabola. ∎

The term *Gaussian elimination* is used to cover the entire solution method, involving both the elimination algorithm and back substitution in the transformed, upper triangular system.

As presented, the algorithm involves *pivoting*. This means that each diagonal element is determined as the largest element in the leading column of the active part of the system. In hand calculation this part is needed only, if the current diagonal

element is zero. When the calculations are made on a computer, the use of it is very important. The reason is that a computer has a limited accuracy, and if an elimination factor ℓ_{ik} is very large, then there is a risk that the result of

$$\boldsymbol{T}_{i,:} \leftarrow \boldsymbol{T}_{i,:} - \ell_{ik}\boldsymbol{T}_{k,:}$$

is so dominated by the contribution $\ell_{ik}\boldsymbol{T}_{k,:}$, that the "old" information in $\boldsymbol{T}_{i,:}$ is lost. With pivoting all the elimination factors satisfy $|\ell_{ik}| \leq 1$, and this reduces the risk of loss of information.

Example 2.7. Consider the matrix $\boldsymbol{A}$ from Example 1.15 with a right-side $\boldsymbol{b} = \boldsymbol{A}\boldsymbol{z}$:

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 8 \\ 4 & 5 & 6 & 20 \\ 7 & 8 & 9 & 32 \end{array}\right) \underset{p_1 = 3}{\sim} \left(\begin{array}{ccc|c} 7 & 8 & 9 & 32 \\ 4 & 5 & 6 & 20 \\ 1 & 2 & 3 & 8 \end{array}\right) \underset{\substack{\ell_{21} = \frac{4}{7} \\ \ell_{31} = \frac{1}{7}}}{\sim}$$

$$\left(\begin{array}{ccc|c} 7 & 8 & 9 & 32 \\ 0 & \frac{3}{7} & \frac{6}{7} & \frac{12}{7} \\ 0 & \frac{6}{7} & \frac{12}{7} & \frac{24}{7} \end{array}\right) \underset{\substack{p_2 = 3 \\ \ell_{3,2} = \frac{1}{2}}}{\sim} \left(\begin{array}{ccc|c} 7 & 8 & 9 & 32 \\ 0 & \frac{6}{7} & \frac{12}{7} & \frac{24}{7} \\ 0 & 0 & 0 & 0 \end{array}\right) .$$

We see that $u_{33} = 0$, so there is a problem that we need to discuss. ∎

2.3. Complete Solution

The elimination algorithm from the previous section transforms the given system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ to an upper triangular system $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{c}$,

$$\begin{aligned} u_{11}x_1 + u_{12}x_2 + \cdots + u_{1n}x_n &= c_1 \\ u_{22}x_2 + \cdots + u_{2n}x_n &= c_2 \\ &\vdots \\ u_{nn}x_n &= c_n \end{aligned} \quad , \tag{2.10}$$

without changing the solution $\boldsymbol{x}$.

If all the $u_{ii} \neq 0$, then the solution is found simply by back substitution. Note that in this case the solution $\boldsymbol{x}$ is unique.

The case where one or more $u_{ii} = 0$ needs further investigation. We say that the matrix $\boldsymbol{A}$ is *singular* (and that it is *nonsingular* if all the $u_{ii} \neq 0$). We shall only give a detailed discussion of the simplest case, where $u_{nn} = 0$ is the only zero element on the diagonal of $\boldsymbol{U}$. Then the last equation in (2.10) has the form

$$0 \cdot x_n = c_n .$$

If $c_n \neq 0$, then this equation, and therefore the original system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, has no solution. If $c_n = 0$, we say that the system is *consistent*. Then the last equation is satisfied by $x_n = \alpha$, where α is any real number. In order to find the other elements in $\boldsymbol{x}$ we see from Eq. (2.10) that they must satisfy the reduced system

$$\begin{aligned} u_{11}x_1 + u_{12}x_2 + \cdots + u_{1,n-1}x_{n-1} &= c_1 - u_{1,n}\alpha \\ u_{22}x_2 + \cdots + u_{2,n-1}x_{n-1} &= c_2 - u_{2,n}\alpha \\ &\vdots \\ u_{n-1,n-1}x_{n-1} &= c_{n-1} - u_{n-1,n}\alpha \end{aligned} \quad . \tag{2.11}$$

All the diagonal coefficients in this system are nonzero, so we can use back substitution and get

$$x_i \;=\; z_i + \alpha v_i \;, \quad i = n{-}1, \ldots, 2, 1 \;,$$

where

$$z_{n-1} \;=\; c_{n-1}/u_{n-1,n-1} \;, \quad v_{n-1} \;=\; -u_{i,n-1}/u_{n-1,n-1} \;,$$

and for $i = n{-}2, \ldots, 2, 1$:

$$\begin{aligned} z_i &= \left(c_i - u_{i,i+1}z_{i+1} - \cdots - u_{i,n-1}z_{n-1}\right)/u_{ii} \;, \\ v_i &= \left(-u_{i,n} - u_{i,i+1}v_{i+1} - \cdots - u_{i,n-1}v_{n-1}\right)/u_{ii} \;. \end{aligned}$$

So, in this case the *complete solution* to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is

$$\boldsymbol{x} \;=\; \boldsymbol{x}_b + \alpha \boldsymbol{w} \;, \quad \alpha \in \mathbb{R} \;, \tag{2.12}$$

where

$$\boldsymbol{x}_b \;=\; \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{n-1} \\ 0 \end{pmatrix} \;, \qquad \boldsymbol{w} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_{n-1} \\ 1 \end{pmatrix} \;.$$

Example 2.8. In Example 2.7 we saw that

$$\left(\begin{array}{ccc|c} 1 & 2 & 3 & 8 \\ 4 & 5 & 6 & 20 \\ 7 & 8 & 9 & 32 \end{array}\right) \;\sim\; \left(\begin{array}{ccc|c} 7 & 8 & 9 & 32 \\ 0 & \frac{6}{7} & \frac{12}{7} & \frac{24}{7} \\ 0 & 0 & 0 & 0 \end{array}\right) \;.$$

Using the method described above we get the complete solution

$$\boldsymbol{x} \;=\; \boldsymbol{x}_b + \alpha \boldsymbol{w} \;=\; \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \;, \quad \alpha \in \mathbb{R} \;.$$

We recognize the vector $\boldsymbol{w}$ from Example 1.15. The vector $\boldsymbol{z}$ from the same example is obtained for $\alpha = 1$. ∎

The solution $\boldsymbol{x} = \boldsymbol{x}_b + \alpha \boldsymbol{w}$ was constructed so that

$$\boldsymbol{A}(\boldsymbol{x}_b + \alpha \boldsymbol{w}) \;=\; \boldsymbol{b}$$

for any choice of α. By use of the linearity condition of a linear mapping, Eq. (1.7), we see that $\boldsymbol{A}(\boldsymbol{x}_b + \alpha \boldsymbol{w}) = \boldsymbol{b}$ is equivalent to the condition

$$\boldsymbol{A}\,\boldsymbol{x}_b + \alpha \boldsymbol{A}\,\boldsymbol{w} \;=\; \boldsymbol{b}$$

for any choice of α. This can be satisfied only, if

$$\boldsymbol{A}\,\boldsymbol{x}_b \;=\; \boldsymbol{b} \qquad \text{and} \qquad \boldsymbol{A}\,\boldsymbol{w} = \; \boldsymbol{0} \;. \tag{2.13}$$

We say that $\boldsymbol{w}$ is in the *nullspace* (or *kernel*) of $\boldsymbol{A}$. The system

$$\boldsymbol{A}\,\boldsymbol{x} \;=\; \boldsymbol{0}$$

is said to be a *homogeneous system*. The zero vector $\boldsymbol{x} = \boldsymbol{0}$ satisfies this system for any $\boldsymbol{A}$, and if $\boldsymbol{A}$ is nonsingular, then $\boldsymbol{0}$ is the only solution. If $\boldsymbol{A}$ is singular, however, then it has a proper nullspace that contains nonzero vectors, for instance $\boldsymbol{w}$.

By rewriting the condition $\boldsymbol{A}\boldsymbol{w} = \boldsymbol{0}$ we see that

$$w_1\boldsymbol{A}_{:,1} + \cdots + w_{n-1}\boldsymbol{A}_{:,n-1} + \boldsymbol{A}_{:,n} \;=\; \boldsymbol{0}\ .$$

This shows (cf page 9) that the columns in a singular matrix are linearly dependent. We can write

$$\boldsymbol{A}_{:,n} \;=\; -w_1\boldsymbol{A}_{:,1} - \cdots - w_{n-1}\boldsymbol{A}_{:,n-1}\ ,$$

and this implies that the image of any $\boldsymbol{x} \in \mathbb{R}^n$ can be expressed as

$$\begin{aligned} \boldsymbol{A}\,\boldsymbol{x} \;&=\; x_1\boldsymbol{A}_{:,1} + \cdots + x_{n-1}\boldsymbol{A}_{:,n-1} + x_n\boldsymbol{A}_{:,n} \\ &=\; \hat{x}_1\boldsymbol{A}_{:,1} + \cdots + \hat{x}_{n-1}\boldsymbol{A}_{:,n-1}\ , \end{aligned} \tag{2.14}$$

where $\hat{x}_j = x_j - w_j x_n$. In other words, the image is a linear combination of the first $n-1$ columns in $\boldsymbol{A}$. This is a subspace of $\mathbb{R}^n$, called the *column range* of $\boldsymbol{A}$.

The condition for $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ having a solution is that $\boldsymbol{b}$ is "reachable", ie that $\boldsymbol{b}$ is in the column range of $\boldsymbol{A}$. If the $n\times n$ matrix is nonsingular, then the nullspace of $\boldsymbol{A}$ has $\boldsymbol{0}$ as the only element, and the column range of $\boldsymbol{A}$ is the entire $\mathbb{R}^n$.

Example 2.9. We shall briefly discuss a problem, $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \;\sim\; \boldsymbol{U}\boldsymbol{x} = \boldsymbol{c}$, where $\boldsymbol{U}$ has more than one $u_{ii} = 0$ or the zero diagonal element is not in the last row.

In all of these cases the matrix $\boldsymbol{A}$ is singular, and the solution – if it exists – has at least one free parameter. Intuitively, the number of free parameters is equal to the number of zero diagonal elements in $\boldsymbol{U}$, but this guess may be wrong! To illustrate that, consider the two examples

$$\left(\begin{array}{ccccc|c} 4 & 2 & 10 & 0 & 4 & 50 \\ 2 & 4 & 8 & 6 & 14 & 64 \\ 1 & 0 & 2 & 3 & 3 & 10 \\ 3 & 2 & 8 & 9 & 13 & 52 \\ 0 & 3 & 3 & 3 & 9 & 36 \end{array}\right) \;\sim\; \left(\begin{array}{ccccc|c} 4 & 2 & 10 & 0 & 4 & 50 \\ 0 & 3 & 3 & 6 & 12 & 39 \\ 0 & 0 & 0 & 4 & 4 & 4 \\ 0 & 0 & 0 & 8 & 8 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)\ ,$$

$$\left(\begin{array}{ccccc|c} 4 & 6 & 8 & 8 & 4 & 4 \\ 2 & 13 & 10 & 0 & 6 & 36 \\ 1 & -1 & 0.5 & 11 & 6 & -9.5 \\ 3 & 7 & 7.5 & 9 & 6 & 5.5 \\ 0 & -5 & -3 & 4 & -1 & -20 \end{array}\right) \;\sim\; \left(\begin{array}{ccccc|c} 4 & 6 & 8 & 8 & 4 & 4 \\ 0 & 10 & 6 & -4 & 4 & 34 \\ 0 & 0 & 0 & 8 & 6 & -2 \\ 0 & 0 & 0 & 4 & 2 & -6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)\ ,$$

In both cases $\boldsymbol{U}$ has two zero diagonals, and for each $\boldsymbol{A}$ the right-hand side was generated as $\boldsymbol{b} = \boldsymbol{A}\breve{\boldsymbol{x}}$, where

$$\breve{\boldsymbol{x}} \;=\; \begin{pmatrix} 1 & -2 & 3 & -4 & 5 \end{pmatrix}^T,$$

so both systems have a solution. For the first system the complete solution is

$$\boldsymbol{x} \;=\; \begin{pmatrix} 7 \\ 11 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -2 & 0 \\ -1 & -2 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix} \boldsymbol{\alpha}\ , \quad \boldsymbol{\alpha} \in \mathbb{R}^2\ ,$$

so the solution involves two free parameters. The vector $\breve{\boldsymbol{x}}$ is obtained for $\alpha_1 = -13$, $\alpha_2 = 5$.

The complete solution to the second system is

$$\boldsymbol{x} \;=\; \begin{pmatrix} 4.3 \\ -0.2 \\ 0 \\ -4 \\ 5 \end{pmatrix} + \begin{pmatrix} -1.1 \\ -0.6 \\ 1 \\ 0 \\ 0 \end{pmatrix} \alpha \;, \quad \alpha \in \mathbb{R} \;,$$

so for this problem there is only one free parameter. The vector $\breve{\boldsymbol{x}}$ is obtained for $\alpha = 3$.

It is outside the scope of this lecture note to describe how we computed the complete solution. In Section 4.3 we discuss a method, which is better than Gaussian elimination for computing nullspace, column range and complete solution. ∎

2.4. Matrix inverse

Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times q}$, and consider the *matrix equation*

$$\boldsymbol{A}\,\boldsymbol{X} \;=\; \boldsymbol{B} \;.$$

If $\boldsymbol{A}$ is nonsingular, then there is a unique solution $\boldsymbol{X} \in \mathbb{R}^{n \times q}$, whose columns satisfy the linear system of equations

$$\boldsymbol{A}\,\boldsymbol{X}_{:,j} \;=\; \boldsymbol{B}_{:,j} \;, \quad j = 1, 2, \ldots, q \;.$$

The algorithms in Gaussian elimination are simple to adjust to cope with matrix equations: If we let the augmented matrix include all the columns of $\boldsymbol{B}$,

$$\boldsymbol{T} \;=\; \left(\boldsymbol{A} \;\middle|\; \boldsymbol{B} \right) \;,$$

then the elimination algorithm (2.9) needs not be changed, and it leads to

$$\boldsymbol{T} \;\sim\; \left(\boldsymbol{U} \;\middle|\; \boldsymbol{C} \right) \;.$$

If there are no zeros on the diagonal of $\boldsymbol{U}$, then the solution can be found by the following modified version of the back substitution algorithm (2.6),

$$\begin{array}{l} \boldsymbol{X}_{n,:} \;=\; \boldsymbol{C}_{:,j} / u_{nn} \\ \textbf{for } i = n{-}1, \ldots, 2, 1 \\ \quad \boldsymbol{X}_{i,:} \;=\; \left(\boldsymbol{C}_{i,:} - u_{i,i+1} \boldsymbol{X}_{i+1,:} - \cdots - u_{i,n} \boldsymbol{X}_{n,:} \right) / u_{ii} \\ \textbf{end} \end{array}$$

Example 2.10. We would like to solve the matrix equation

$$\begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} \boldsymbol{X} \;=\; \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \;.$$

With the notation from Example 2.6 we get

$$\left(\begin{array}{cc|cc} 2 & 3 & 1 & 0 \\ 4 & 5 & 0 & 1 \end{array}\right) \underset{p_1=2}{\sim} \left(\begin{array}{cc|cc} 4 & 5 & 0 & 1 \\ 2 & 3 & 1 & 0 \end{array}\right) \underset{\ell_{21}=\frac{1}{2}}{\sim} \left(\begin{array}{cc|cc} 4 & 5 & 0 & 1 \\ 0 & \frac{1}{2} & 1 & -\frac{1}{2} \end{array}\right),$$

and back substitution gives

$$\begin{aligned} \boldsymbol{X}_{2,:} &= \begin{pmatrix} 1 & -\frac{1}{2} \end{pmatrix} / \tfrac{1}{2} = \begin{pmatrix} 2 & -1 \end{pmatrix}, \\ \boldsymbol{X}_{1,:} &= \left\{ \begin{pmatrix} 0 & 1 \end{pmatrix} - 5 \begin{pmatrix} 2 & -1 \end{pmatrix} \right\} / 4 = \begin{pmatrix} \frac{-5}{2} & \frac{3}{2} \end{pmatrix}. \end{aligned}$$

This means that the solution is $\boldsymbol{X} = \frac{1}{2} \begin{pmatrix} -5 & 3 \\ 4 & -2 \end{pmatrix}$. ∎

The matrix equation with $\boldsymbol{B} = \boldsymbol{I}$, the identity matrix, is of special interest. Its solution is called "$\boldsymbol{A}$ *inverse*" and is written $\boldsymbol{A}^{-1}$,

$$\boldsymbol{A}\,\boldsymbol{A}^{-1} = \boldsymbol{I}\,. \tag{2.15}$$

The matrix inverse is defined only for nonsingular matrices.

Now, let $\boldsymbol{A}$ be a nonsingular $n \times n$ matrix, $\boldsymbol{b}$ an n-vector, and define $\boldsymbol{y} = \boldsymbol{A}^{-1}\boldsymbol{b}$. Then

$$\boldsymbol{A}\,\boldsymbol{y} = \boldsymbol{A}\,\boldsymbol{A}^{-1}\boldsymbol{b} = \boldsymbol{I}\,\boldsymbol{b} = \boldsymbol{b}\,.$$

This shows that $\boldsymbol{y} = \boldsymbol{x}$, the unique solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, ie, if $\boldsymbol{A}$ is nonsingular, then

$$\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{b} \quad \Leftrightarrow \quad \boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}\,. \tag{2.16}$$

If we multiply first equation by $\boldsymbol{A}^{-1}$ from the left, we get

$$\boldsymbol{A}^{-1}\boldsymbol{A}\,\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}\,,$$

and combining this with Eqs. (2.15) and (2.16) we see that the matrix inverse satisfies

$$\boldsymbol{A}\,\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}\,. \tag{2.17}$$

This is one of the rare examples, where the interchange of the factors in a matrix-matrix product does not change the result.

Example 2.11. In Example 2.10 we actually computed a matrix inverse:

$$\boldsymbol{A} = \begin{pmatrix} 2 & 3 \\ 4 & 5 \end{pmatrix} \quad \Leftrightarrow \quad \boldsymbol{A}^{-1} = \frac{1}{2} \begin{pmatrix} -5 & 3 \\ 4 & -2 \end{pmatrix}.$$

Now consider the system $\boldsymbol{A}\boldsymbol{x} = \begin{pmatrix} 4 \\ 6 \end{pmatrix}$. By means of Eq. (2.16) we can find the solution:

$$\boldsymbol{x} = \frac{1}{2} \begin{pmatrix} -5 & 3 \\ 4 & -2 \end{pmatrix} \begin{pmatrix} 4 \\ 6 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}.$$ ∎

There are matrices for which it is almost embarrassingly simple to compute the inverse. We shall mention diagonal and orthogonal matrices:

$$\boldsymbol{D} = \begin{pmatrix} d_{11} & & & \\ & d_{22} & & \\ & & \ddots & \\ & & & d_{nn} \end{pmatrix} \quad \Leftrightarrow \quad \boldsymbol{D}^{-1} = \begin{pmatrix} d_{11}^{-1} & & & \\ & d_{22}^{-1} & & \\ & & \ddots & \\ & & & d_{nn}^{-1} \end{pmatrix}.$$

Example 2.12.

$$\boldsymbol{D} \;=\; \begin{pmatrix} 1 & & \\ & \frac{1}{2} & \\ & & 3 \end{pmatrix} \quad\Leftrightarrow\quad \boldsymbol{D}^{-1} \;=\; \begin{pmatrix} 1 & & \\ & 2 & \\ & & \frac{1}{3} \end{pmatrix} .$$

■

An orthogonal matrix $\boldsymbol{Q}$ satisfies

$$\boldsymbol{Q}^T\boldsymbol{Q} \;=\; \boldsymbol{I} \;, \tag{2.18}$$

cf the discussion at the start of Section 2.1. According to Eq. (2.17) this implies that

$$\boldsymbol{Q}^{-1} \;=\; \boldsymbol{Q}^T \;,$$

and that also $\boldsymbol{Q}^T$ is orthogonal

$$\boldsymbol{Q}\,\boldsymbol{Q}^T \;=\; \boldsymbol{I} \;.$$

It is wasteful to use Gaussian elimination to solve a system $\boldsymbol{Q}\boldsymbol{x} = \boldsymbol{y}$ with an orthogonal matrix. It is both faster and slightly more accurate to exploit that

$$\boldsymbol{Q}\,\boldsymbol{x} \;=\; \boldsymbol{b} \quad\Leftrightarrow\quad \boldsymbol{x} \;=\; \boldsymbol{Q}^T\boldsymbol{b} \;. \tag{2.19}$$

Example 2.13. We actually exploited Eq. (2.19) in connection with Eq. (1.14). The general equation for basis shift is $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{\alpha}$, where the columns in $\boldsymbol{B}$ are the coordinates of the "new" basis vectors. If they are orthonormal, then $\boldsymbol{B}$ is orthogonal, and the "new" coordinates are $\boldsymbol{\alpha} = \boldsymbol{B}^T\boldsymbol{x}$. ■

For a general square, nonsingular matrix $\boldsymbol{A}$ it must be said that Eq. (2.16) is a convenient notation when discussing a mathematical model, but "$\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$" should be read as shorthand for "find $\boldsymbol{x}$ as the solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$". The matrix inverse itself is only needed in special cases, and the solution via Gaussian elimination is both faster and usually also less affected by rounding errors than the solution via the matrix inverse.

Exercise 2.14. We can use this comment on the matrix inverse to derive the following result: Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be square and nonsingular. Show that

$$(\boldsymbol{A}\boldsymbol{B})^{-1} \;=\; \boldsymbol{B}^{-1}\boldsymbol{A}^{-1} \;.$$

Hint: Consider the matrix equation $\boldsymbol{A}\boldsymbol{B}\boldsymbol{X} = \boldsymbol{I}$ and introduce the auxiliary matrix $\boldsymbol{C} = \boldsymbol{B}\boldsymbol{X}$. ■

Example 2.15. It sometimes happens that one has to solve a series of problems

$$\boldsymbol{A}\boldsymbol{x}_k \;=\; \boldsymbol{b}_k \;, \quad k = 1, 2, \ldots, K \;,$$

with the same matrix but different right-hand sides. It is, of course, wasteful to start from scratch with each problem, since the operations on the matrix elements are the same.

If all $\boldsymbol{b}_k$ are known, then the simplest method is to solve the problem as a matrix equation $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{B}$, where the $n \times K$ matrix $\boldsymbol{B}$ has $\boldsymbol{b}_k$ as its k^{th} column, $\boldsymbol{B}_{:,k} = \boldsymbol{b}_k$.

In some cases, however, $\boldsymbol{b}_k$ is a function of $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{k-1}$, so it is not known from the start, and we cannot use the matrix equation approach. One might argue that

the use of $\boldsymbol{A}^{-1}$ would be appropriate here, since the inverse only has to be computed once, and then each $\boldsymbol{x}_k$ is computed by the matrix-vector product $\boldsymbol{A}^{-1}\boldsymbol{b}_k$. In the next section we describe a better method for solving this problem. ∎

2.5. LU factorization

Example 2.16. In Example 2.7 we saw that

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \quad \begin{matrix} \\ p_1 = 3 \\ \ell_{21} = \frac{4}{7} \\ \ell_{31} = \frac{1}{7} \end{matrix} \quad \begin{pmatrix} 7 & 8 & 9 \\ 0 & \frac{3}{7} & \frac{6}{7} \\ 0 & \frac{6}{7} & \frac{12}{7} \end{pmatrix}$$
$$\begin{matrix} \\ p_2 = 3 \\ \ell_{32} = \frac{1}{2} \end{matrix} \quad \begin{pmatrix} 7 & 8 & 9 \\ 0 & \frac{6}{7} & \frac{12}{7} \\ 0 & 0 & 0 \end{pmatrix} = \boldsymbol{U} .$$

We introduce a lower triangular matrix $\boldsymbol{L}$ with ones on the diagonal and the elimination factors stored column-wise in the strictly lower triangle. The storage is affected by row interchanges: if $p_k \neq k$, we use the row interchange $\boldsymbol{T}_{k,:} \leftrightarrow \boldsymbol{T}_{p_k,:}$ in Algorithm (2.9), and the same interchange should be applied to the "preceding" elimination factors,

$$(\boldsymbol{L})_{kj} \leftrightarrow (\boldsymbol{L})_{p_k j} , \quad j = 1, \ldots, k-1 .$$

With the above matrix we have $p_2 = 3$, so ℓ_{21} and ℓ_{31} end up in $(\boldsymbol{L})_{31}$ and $(\boldsymbol{L})_{21}$, respectively, while $(\boldsymbol{L})_{32} = \ell_{32}$:

$$\boldsymbol{L} = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{7} & 1 & 0 \\ \frac{4}{7} & \frac{1}{2} & 1 \end{pmatrix} .$$

It is easily verified that

$$\boldsymbol{L}\boldsymbol{U} = \begin{pmatrix} 7 & 8 & 9 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} .$$

This product is equal to the matrix obtained after the row interchange $\boldsymbol{A}_{1,:} \leftrightarrow \boldsymbol{A}_{3,:}$, followed by $\boldsymbol{A}_{2,:} \leftrightarrow \boldsymbol{A}_{3,:}$. ∎

The result in this example is no coincidence. We used the elimination algorithm Eq. (2.9), and it can be shown that the effect on the coefficient matrix can be expressed by

$$\boldsymbol{P}\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U} . \tag{2.20}$$

This is the so-called *LU factorization* of $\boldsymbol{A}$. The matrix $\boldsymbol{P}$ is a *permutation matrix*, characterized by each row and each column having one element equal to 1, and all the other elements are zero. $\boldsymbol{L}$ is a lower triangular matrix with ones on the diagonal and the elimination factors in the strictly lower triangle, stored as outlined in the example. Finally, $\boldsymbol{U}$ is the upper triangular matrix resulting from the elimination algorithm.

Example 2.17. The permutation matrix $\boldsymbol{P}$ can be obtained from the identity matrix $\boldsymbol{I}$ by applying the row interchanges defined by the elimination algorithm. For the

problem in Example 2.16 we get

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \overset{p_1 = 3}{\sim} \quad \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \quad \overset{p_2 = 3}{\sim} \quad \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \;=\; \boldsymbol{P} \,.$$

Note that the columns in $\boldsymbol{P}$ are orthogonal. This is a general property of a permutation matrix, and implies that $\boldsymbol{P}^{-1} = \boldsymbol{P}^T$.

Let j_i denote the position of the 1 in the i^{th} row of $\boldsymbol{P}$. Then the result of $\boldsymbol{PA}$ is that the rows of $\boldsymbol{A}$ come in the sequence $j_1, j_2, \ldots, j_n$. In particular, for the problem of Example 2.16 the above $\boldsymbol{P}$ gives $j_1 = 3$, $j_2 = 1$, $j_3 = 2$, and

$$\boldsymbol{P}\,\boldsymbol{A} \;=\; \begin{pmatrix} 7 & 8 & 9 \\ 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \,.$$

This is equal to the result of the product $\boldsymbol{LU}$ in Example 2.16. ∎

Now, assume that we know the LU factorization of $\boldsymbol{A}$ and want to solve the system $\boldsymbol{Ax} = \boldsymbol{b}$. This is obviously equivalent to $\boldsymbol{PAx} = \boldsymbol{Pb}$, or

$$\boldsymbol{L}\,\boldsymbol{U}\,\boldsymbol{x} \;=\; \boldsymbol{P}\,\boldsymbol{b} \,.$$

We introduce $\boldsymbol{c} = \boldsymbol{Ux}$, and it follows that the original problem can be solved in the two steps

$$\begin{aligned} &1^\circ \qquad \text{solve} \quad \boldsymbol{L}\,\boldsymbol{c} \;=\; \boldsymbol{Pb} \,, \\ &2^\circ \qquad \text{solve} \quad \boldsymbol{U}\,\boldsymbol{x} \;=\; \boldsymbol{c} \,. \end{aligned} \tag{2.21}$$

The two systems involve a lower and an upper triangular matrix, so their solution is found via forward and back substitution, respectively. The ones on the diagonal ensure that $\boldsymbol{L}$ is nonsingular, so $\boldsymbol{c}$ is unique. A singular system is reflected in zero diagonal elements in $\boldsymbol{U}$.

Example 2.18. The nonsingular matrix in Example 2.6 has the LU factorization

$$\boldsymbol{P} \;=\; \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \,, \quad \boldsymbol{L} \;=\; \begin{pmatrix} 1 & & \\ \frac{4}{9} & 1 & \\ \frac{1}{9} & 1 & 1 \end{pmatrix} \,, \quad \boldsymbol{U} \;=\; \begin{pmatrix} 9 & 3 & 1 \\ & \frac{6}{9} & \frac{2}{9} \\ & & \frac{3}{9} \end{pmatrix} \,.$$

The vector $\boldsymbol{Pb}$ is obtained by interchanging the first and third element in $\boldsymbol{b}$. We combine this with the forward substitution algorithm, Eq. (2.5), and exploit that all the diagonal elements are one, so there is no division involved:

$$\begin{aligned} c_1 \;&=\; 1.6 \,, \\ c_2 \;&=\; 0.7 - \tfrac{4}{9} \;=\; \tfrac{-0.1}{9} \,, \\ c_3 \;&=\; 0.8 - \tfrac{1}{9} \cdot 1.6 - 1 \cdot \tfrac{-0.1}{9} \;=\; \tfrac{5.7}{9} \,. \end{aligned}$$

This $\boldsymbol{c}$ is identical to the modified right-hand side in Example 2.6, which is no wonder: The calculations involved in step 1° are the same as were used to get from $\left(\boldsymbol{A} \;\middle|\; \boldsymbol{b}\right)$ to $\left(\boldsymbol{U} \;\middle|\; \boldsymbol{c}\right)$, except that they are done in a different order. Step 2° was performed already in Example 2.4. ∎

Example 2.19. Algorithm 2.21 can be interpreted as an application of the relation treated in Exercise 2.14:

$$\boldsymbol{LUx} \;=\; \boldsymbol{Pb} \quad \Leftrightarrow \quad \boldsymbol{x} \;=\; (\boldsymbol{LU})^{-1}\boldsymbol{Pb} \;=\; \boldsymbol{U}^{-1}\boldsymbol{L}^{-1}\boldsymbol{Pb} \;=\; \boldsymbol{U}^{-1}\boldsymbol{c} \,.$$

∎

Example 2.20. MATLAB has a built-in function `lu` for computing the LU factorization. Let the MATLAB variable `A` contain the matrix $\boldsymbol{A}$, then the command

```
[L, U, P] = lu(A)
```

returns $\boldsymbol{L}$, $\boldsymbol{U}$ and $\boldsymbol{P}$ in `L`, `U` and `P`, respectively. The command

```
[L, U] = lu(A)
```

returns $\boldsymbol{P}^T\boldsymbol{L}$ in `L` and $\boldsymbol{U}$ in `U`. Thus, `L` contains a row permuted lower triangular matrix. We return to this aspect in Example 2.28 at the end of the the next section. ∎

Example 2.21. It is relevant to ask the question: If $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{b} \in \mathbb{R}^n$, how much work is needed to solve the system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$? The elimination algorithm Eq. (2.9) involves pivot searches, row interchanges and a number of "flops". A *flop* (short for floating point operation) is a simple arithmetic operation (addition, subtraction, multiplication or division) between two real numbers. Traditionally, only the flops are counted, and it can be shown that the transformation from $\boldsymbol{A}$ to $\boldsymbol{U}$ involves

$$\left(\tfrac{2}{3}\,n^3 - \tfrac{1}{2}\,n^2 - \tfrac{1}{6}\,n\right) \text{ flops} .$$

For large n the first term dominates, and we say that the *"cost"* of the transformation from $\boldsymbol{A}$ to $\boldsymbol{U}$ is $\frac{2}{3}n^3$ flops. Note that this is also the cost of computing the LU factorization. The cost of solving each of the triangular systems $\boldsymbol{L}\boldsymbol{c} = \boldsymbol{P}\boldsymbol{b}$ and $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{c}$ is n^2 flops. For comparison, the matrix-vector multiplication $\boldsymbol{A}\boldsymbol{z}$ costs $2n^2$ flops, ie the same as the solution of both triangular systems.

For a matrix equation $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{B}$ we need $\frac{2}{3}\,n^3$ flops to transform $\boldsymbol{A}$ to $\boldsymbol{U}$ and $2n^2$ flops per column in $\boldsymbol{B}$. If there are q columns, the solution cost is $\left(\frac{2}{3}\,n^3 + 2q\cdot n^2\right)$ flops. If $q = n$, then the cost is $\frac{8}{3}\,n^3$ flops. In the special case where $\boldsymbol{B} = \boldsymbol{I}$ it is possible to exploit the zeros in the off-diagonal positions, and the cost of computing $\boldsymbol{A}^{-1}$ reduces to $2n^3$ flops.

For the problem in Example 2.15,

$$\boldsymbol{A}\boldsymbol{x}_k \;=\; \boldsymbol{b}_k \;, \quad k = 1, 2, \ldots, K \;,$$

it follows that the solution via the matrix inverse costs $\left(2n^3 + 2Kn^2\right)$ flops. Instead we should compute the LU factorization once and then use Eq. (2.21) for each $\boldsymbol{b}_k$. With this method the cost reduces to $\left(\frac{2}{3}n^3 + 2Kn^2\right)$ flops, so we save $\frac{4}{3}n^3$ flops. ∎

2.6. SPD systems

In an SPD system the $n\times n$ coefficient matrix $\boldsymbol{A}$ is *symmetric* and *positive definite* (the matrix is an SPD matrix). This means that $\boldsymbol{A}^T = \boldsymbol{A}$ and

$$\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x} \;>\; 0 \quad \text{for all nonzero } \boldsymbol{x} \in \mathbb{R}^n \;. \tag{2.22}$$

Note that the left hand side is the scalar product between the vectors $\boldsymbol{x}$ and $\boldsymbol{A}\boldsymbol{x}$, so it is a real number, and this is positive. Note also, that that this implies that $\boldsymbol{A}\boldsymbol{x} \neq \boldsymbol{0}$ for all nonzero $\boldsymbol{x}$, so the columns in $\boldsymbol{A}$ are linearly independent and $\boldsymbol{A}$ is nonsingular.

If $\boldsymbol{A}$ is symmetric and

$$\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x} \;\geq\; 0 \quad \text{for all nonzero } \boldsymbol{x} \in \mathbb{R}^n \;,$$

then $\boldsymbol{A}$ is an SPSD (symmetric, positive semidefinite) matrix.

Example 2.22. In Chapter 4 we introduce the *normal equations* matrix

$$\boldsymbol{A} \;= \boldsymbol{F}^T\boldsymbol{F} \;,$$

where $\boldsymbol{F}$ is an $m\times n$ matrix with $m \geq n$. We immediately see that $\boldsymbol{A}$ is an $n\times n$ matrix, and using Eq. (1.6) we get

$$\boldsymbol{A}^T \;=\; \boldsymbol{F}^T(\boldsymbol{F}^T)^T \;=\; \boldsymbol{F}^T\boldsymbol{F} \;=\; \boldsymbol{A} \;,$$

so $\boldsymbol{A}$ is symmetric. Next, we let $\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x}$ and get

$$\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x} \;=\; \boldsymbol{x}^T\boldsymbol{F}^T\boldsymbol{F}\,\boldsymbol{x} \;=\; \boldsymbol{y}^T\boldsymbol{y} \;=\; y_1^2 + \cdots + y_m^2 \;\geq\; 0 \;.$$

This shows that $\boldsymbol{A}$ is an SPSD matrix.

If the columns in $\boldsymbol{F}$ are linearly independent, then $\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} \neq \boldsymbol{0}$ for all nonzero $\boldsymbol{x}$, and it follows that

$$\boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x} \;=\; y_1^2 + \cdots + y_m^2 \;>\; 0 \;.$$

So in this case $\boldsymbol{A}$ is an SPD matrix. ∎

It can be shown that pivoting does not improve the accuracy of Gaussian elimination applied to an SPD system. If we omit this, there are no row interchanges, and formula (2.20) simplifies to

$$\boldsymbol{A} \;=\; \boldsymbol{L}\,\boldsymbol{U} \;.$$

Further, one can show that the symmetry is preserved in the active part of the matrix during the transformation, and this has the effect that $\boldsymbol{U} = \boldsymbol{D}\boldsymbol{L}^T$, where $\boldsymbol{D}$ is the diagonal matrix given by the diagonal elements in $\boldsymbol{U}$. This means that Gaussian elimination applied to an SPD system is equivalent to the so-called *LDL factorization*

$$\boldsymbol{A} \;=\; \boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^T \;. \tag{2.23}$$

Example 2.23. With the notation from for instance Example 2.6 we get

$$\begin{pmatrix} 4 & 12 & -8 & 16 \\ 12 & 45 & -33 & 66 \\ -8 & -33 & 41 & 30 \\ 16 & 66 & 30 & 502 \end{pmatrix} \begin{array}{c} \sim \\ \ell_{21}=3 \\ \ell_{31}=-2 \\ \ell_{41}=4 \end{array} \begin{pmatrix} 4 & 12 & -8 & 16 \\ 0 & 9 & -9 & 18 \\ 0 & -9 & 25 & 62 \\ 0 & 18 & 62 & 438 \end{pmatrix} \begin{array}{c} \sim \\ \ell_{32}=-1 \\ \ell_{42}=2 \end{array}$$

$$\begin{pmatrix} 4 & 12 & -8 & 16 \\ 0 & 9 & -9 & 18 \\ 0 & 0 & 16 & 80 \\ 0 & 0 & 80 & 402 \end{pmatrix} \begin{array}{c} \sim \\ \ell_{43}=5 \end{array} \begin{pmatrix} 4 & 12 & -8 & 16 \\ 0 & 9 & -9 & 18 \\ 0 & 0 & 16 & 80 \\ 0 & 0 & 0 & 2 \end{pmatrix} \;.$$

We see that the active parts of the matrix are symmetric. Collecting the elimination factors in $\boldsymbol{L}$ we get

$$\boldsymbol{L} \;=\; \begin{pmatrix} 1 & & & \\ 3 & 1 & & \\ -2 & -1 & 1 & \\ 4 & 2 & 5 & 1 \end{pmatrix} \;,$$

and $\boldsymbol{U}$ is the resulting upper triangular matrix,

$$\boldsymbol{U} = \begin{pmatrix} 4 & 12 & -8 & 16 \\ & 9 & -9 & 18 \\ & & 16 & 80 \\ & & & 2 \end{pmatrix} = \begin{pmatrix} 4 & & & \\ & 9 & & \\ & & 16 & \\ & & & 2 \end{pmatrix} \begin{pmatrix} 1 & 3 & -2 & 4 \\ & 1 & -1 & 2 \\ & & 1 & 5 \\ & & & 1 \end{pmatrix} .$$

We see that $\boldsymbol{U} = \boldsymbol{D}\boldsymbol{L}^T$. ∎

The symmetry can be exploited: We only need to compute the modified elements on the diagonal and in either the strictly lower or the strictly upper triangle of the matrix. Then we also know the modified elements in the other triangle. This means that the cost of computing the LDL factorization is only half the cost of computing the LU factorization, $\frac{1}{3}n^3$ flops.

The LDL factorization can be computed for any symmetric matrix. However, if the matrix is not positive definite, we should use pivoting in order to avoid excessive effects of rounding errors. Fortunately, it is simple to check the condition Eq. (2.22) during the computation of the factorization:

$$\begin{aligned} \boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x} &= \boldsymbol{x}^T\boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^T\boldsymbol{x} \\ &= \boldsymbol{y}^T\boldsymbol{D}\,\boldsymbol{y} = d_{11}y_1^2 + d_{22}y_2^2 + \cdots + d_{nn}y_n^2 . \end{aligned}$$

Here, we introduced $\boldsymbol{y} = \boldsymbol{L}^T\boldsymbol{x}$. The ones along the diagonal in $\boldsymbol{L}$ ensure that this matrix in nonsingular, and therefore $\boldsymbol{y} \neq \boldsymbol{0}$ for any nonzero $\boldsymbol{x}$. The right-hand side must be positive for any nonzero $\boldsymbol{y}$, and this is clearly the case if and only if all the $d_{ii} > 0$. This means that if we meet a $d_{kk} \leq 0$ during the computation of the factorization, then we know that the matrix is not SPD, and can stop.

Example 2.24. We change the second diagonal element in the matrix from Example 2.23 to $a_{22} = 35$ and get

$$\begin{pmatrix} 4 & 12 & -8 & 16 \\ 12 & 35 & -33 & 66 \\ -8 & -33 & 41 & 30 \\ 16 & 66 & 30 & 502 \end{pmatrix} \begin{matrix} \sim \\ \ell_{21} = 3 \\ \ell_{31} = \star \\ \ell_{41} = \star \end{matrix} \begin{pmatrix} 4 & 12 & -8 & 16 \\ 0 & -1 & \star & \star \\ \star & \star & \star & \star \\ \star & \star & \star & \star \end{pmatrix}$$

The $\star$ indicates values that we do not need to compute. $d_{22} = -1 \leq 0$ tells us that the matrix is not positive definite, and we stop the computation. A system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ with this matrix should be solved by Gaussian elimination with pivoting. ∎

An SPD system is often solved via the *Cholesky factorization* of the matrix,

$$\boldsymbol{A} = \boldsymbol{C}\,\boldsymbol{C}^T , \tag{2.24}$$

where $\boldsymbol{C}$ is a lower triangular matrix. We can derive this from the LDL factorization,

$$\boldsymbol{A} = \boldsymbol{L}\,\boldsymbol{D}\,\boldsymbol{L}^T = \left(\boldsymbol{L}\,\boldsymbol{D}^{1/2}\right)\left(\boldsymbol{D}^{1/2}\boldsymbol{L}^T\right) = \boldsymbol{C}\,\boldsymbol{C}^T .$$

Here, we have introduced the matrix

$$\boldsymbol{D}^{1/2} = \begin{pmatrix} \sqrt{d_{11}} & & & \\ & \sqrt{d_{22}} & & \\ & & \ddots & \\ & & & \sqrt{d_{nn}} \end{pmatrix} .$$

Since the d_{ii} are positive, the square roots exist.

Example 2.25. For the matrix from Example 2.23 the Cholesky factor is

$$\boldsymbol{C} = \begin{pmatrix} 1 & & & \\ 3 & 1 & & \\ -2 & -1 & 1 & \\ 4 & 2 & 5 & 1 \end{pmatrix} \begin{pmatrix} 2 & & & \\ & 3 & & \\ & & 4 & \\ & & & \sqrt{2} \end{pmatrix} = \begin{pmatrix} 2 & & & \\ 6 & 3 & & \\ -4 & -3 & 4 & \\ 8 & 6 & 20 & \sqrt{2} \end{pmatrix} .$$

■

The Cholesky factor may be computed directly, ie without the detour around the LDL factorization.

Once the Cholesky factorization is known, the SPD system

$$\boldsymbol{A}\,\boldsymbol{x} \;=\; \boldsymbol{b} \quad \Leftrightarrow \quad \boldsymbol{C}\,\boldsymbol{C}^T\boldsymbol{x} \;=\; \boldsymbol{b}$$

can be solved in the two steps

$$\begin{aligned} &1^\circ \qquad \text{solve} \quad \boldsymbol{C}\,\boldsymbol{z} \;=\; \boldsymbol{b}\,, \\ &2^\circ \qquad \text{solve} \quad \boldsymbol{C}^T\boldsymbol{x} \;=\; \boldsymbol{z}\,. \end{aligned} \tag{2.25}$$

The solution of these two systems is found via forward and back substitution, respectively.

Example 2.26. Consider the system

$$\begin{pmatrix} 4 & 12 & -8 & 16 \\ 12 & 45 & -33 & 66 \\ -8 & -33 & 41 & 30 \\ 16 & 66 & 30 & 502 \end{pmatrix} \boldsymbol{x} \;=\; \begin{pmatrix} 108 \\ 441 \\ -61 \\ 2034 \end{pmatrix} .$$

In Examples 2.23 and 2.25 we found that the symmetric coefficient matrix is positive definite and computed its Cholesky factor $\boldsymbol{C}$. The solution to the system is found in the two steps

$$\begin{pmatrix} 2 & & & \\ 6 & 3 & & \\ -4 & -3 & 4 & \\ 8 & 6 & 20 & \sqrt{2} \end{pmatrix} \boldsymbol{z} \;=\; \begin{pmatrix} 108 \\ 441 \\ -61 \\ 2034 \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{z} \;=\; \begin{pmatrix} 54 \\ 39 \\ 684\sqrt{2} \end{pmatrix} ,$$

$$\begin{pmatrix} 2 & 6 & -4 & 8 \\ & 3 & -3 & 6 \\ & & 4 & 20 \\ & & & \sqrt{2} \end{pmatrix} \boldsymbol{x} \;=\; \begin{pmatrix} 54 \\ 39 \\ 684\sqrt{2} \end{pmatrix} \quad \Rightarrow \quad \boldsymbol{x} \;=\; \begin{pmatrix} -1 \\ 2 \\ -3 \\ 4 \end{pmatrix} .$$

■

Example 2.27. Traditionally the Cholesky factorization is defined as in Eq. (2.24), $\boldsymbol{A} = \boldsymbol{C}\,\boldsymbol{C}^T$, where $\boldsymbol{C}$ is lower triangular. We might as well define it as

$$\boldsymbol{A} \;=\; \boldsymbol{R}^T\boldsymbol{R}\,,$$

where $\boldsymbol{R} = \boldsymbol{C}^T$ is upper triangular. This is the definition used in MATLAB. The built-in function `chol` can be used to compute the Cholesky factor. Let `A` contain a square matrix, then

```
R = chol(A)
```

attempts to compute the Cholesky factor for the matrix. The matrix is assumed to be symmetric and only the elements in the upper triangle are considered, for instance

```
R = chol([4 2;0 10])     R = chol([4 2;2 10])
```

both give `R = [2  1;0  3]`. If the matrix is not positive definite, you get an error return. ∎

Example 2.28. Let the MATLAB variables `A` and `b` contain respectively an $n \times n$ matrix $\boldsymbol{A}$ and column vector $\boldsymbol{b}$ with n elements. Then the MATLAB command

```
x = A \ b
```

returns the solution to the system $\boldsymbol{Ax} = \boldsymbol{b}$. The method depends on possible special properties of $\boldsymbol{A}$:

1° If $\boldsymbol{A}$ is lower triangular (or a permuted lower triangular matrix, cf Example 2.20), then the system is solved via (permuted) forward substitution. Similarly a (permuted) upper triangular system is solved by (permuted) back substitution.

2° If $\boldsymbol{A}$ is symmetric, then a Cholesky factorization of it is attempted. If it succeeds, then it is used in the two step algorithm Eq. (2.25) to compute the solution. Otherwise the general method is employed.

3° General method: Compute the LU factorization of $\boldsymbol{A}$, and use Eq. (2.21) to solve the system.

∎

3. Eigenvalues and Eigenvectors

Example 3.1. Let λ be a constant. The simple differential equation

$$\frac{dy}{dt} \;=\; \lambda\, y$$

has the solution

$$y(t) \;=\; c\, e^{\lambda t}\ ,$$

where c is an arbitrary constant. If $y(0) = \beta$ is known, then $c = \beta$ gives the desired solution.

A linear system of differential equations has the form

$$\frac{d\boldsymbol{y}}{dt} \;=\; \boldsymbol{A}\,\boldsymbol{y}\ ,$$

where each element in $\boldsymbol{y}$ is a function of time and the elements in the square matrix $\boldsymbol{A}$ are constant. It is tempting to postulate that the solution to this system is

$$\boldsymbol{y}(t) \;=\; e^{\boldsymbol{A}\,t}\,\boldsymbol{c}\ ,$$

where $\boldsymbol{c}$ is an arbitrary vector. Is there an interpretation of the "*matrix exponential*" $e^{\boldsymbol{A}\,t}$ that makes this postulate true ?

Some other interesting questions are

- How did we find a good change of basis for the ellipse problem in Example 1.20 ?
- Why do the most interesting links always come out close to the top in a *Google* search ?
- For a given $n\times n$ mapping matrix $\boldsymbol{A}$, is there a nonzero vector $\boldsymbol{v}$ whose image $\boldsymbol{A}\boldsymbol{v}$ is proportional to $\boldsymbol{v}$?

■

Surprisingly, the key question is the last one. This is the so-called *eigenproblem*

$$\boldsymbol{A}\,\boldsymbol{v} \;=\; \lambda\,\boldsymbol{v}\ , \tag{3.1}$$

where the scalar λ is the *eigenvalue* and the nonzero $\boldsymbol{v}$ is a corresponding *eigenvector*. We say that the pair $(\lambda, \boldsymbol{v})$ is an *eigensolution*.

3.1. Properties of eigensolutions

The definition $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$ is equivalent to

$$(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v} \;=\; \boldsymbol{0}\ . \tag{3.2}$$

We look for a nonzero solution to this homogeneous system of equations, and it follows from the discussion in Section 2.3 that the matrix $\boldsymbol{A}-\lambda\boldsymbol{I}$ must be singular. Especially, if $\lambda=0$ is an eigenvalue for $\boldsymbol{A}$, then the matrix itself is singular.

Example 3.2. For a 2×2 matrix $\boldsymbol{A}$ we can use the *determinant* defined in Example 2.2 to express that $\boldsymbol{A}-\lambda\boldsymbol{I}$ must be singular:

$$\begin{aligned} 0 \;=\; \det(\boldsymbol{A}-\lambda\boldsymbol{I}) \;&=\; \begin{vmatrix} a_{11}-\lambda & a_{12} \\ a_{21} & a_{22}-\lambda \end{vmatrix} \\ &=\; (a_{11}-\lambda)(a_{22}-\lambda) - a_{21}a_{12} \\ &=\; \lambda^2 - (a_{11}+a_{22})\lambda + a_{11}a_{22} - a_{21}a_{12}\ . \end{aligned}$$

This means that the eigenvalues are the roots of the so-called *characteristic polynomial*

$$P_A(\lambda) \;=\; \lambda^2 - (a_{11}+a_{22})\lambda + a_{11}a_{22} - a_{21}a_{12}\ .$$

We shall consider three specific examples,

$$\boldsymbol{B} \;=\; \begin{pmatrix} 73 & -36 \\ -36 & 52 \end{pmatrix}\ , \quad \boldsymbol{C} \;=\; \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\ , \quad \boldsymbol{H} \;=\; \begin{pmatrix} 73 & -36 \\ 36 & 52 \end{pmatrix}\ .$$

With the first matrix we get

$$P_B(\lambda) \;=\; \lambda^2 - 125\lambda + 2500\ .$$

Using a well known technique for solving the quadratic equation $P_B(\lambda)=0$, we first find the discriminant

$$D \;=\; (-125)^2 - 4\cdot1\cdot2500 \;=\; 5625\ ,$$

and the roots are

$$\lambda_1 \;=\; \frac{125+\sqrt{D}}{2} \;=\; 100\ , \quad \lambda_2 \;=\; \frac{125-\sqrt{D}}{2} \;=\; 25\ .$$

In order to find an eigenvector corresponding to λ_2 we solve the singular system $(\boldsymbol{B}-25\boldsymbol{I})\boldsymbol{v}_2=\boldsymbol{0}$. With the notation from Section 2.2 we get

$$\left(\begin{array}{cc|c} 48 & -36 & 0 \\ -36 & 27 & 0 \end{array}\right) \quad \begin{array}{c} \sim \\ p_1=1 \\ \ell_{21}=\frac{-36}{48} \end{array} \quad \left(\begin{array}{cc|c} 48 & -36 & 0 \\ 0 & 0 & 0 \end{array}\right)\ .$$

The reduced system has $u_{22}=0$, confirming that $\boldsymbol{B}-25\boldsymbol{I}$ is singular. Proceeding as in Section 2.3 we find the complete solution

$$\boldsymbol{v}_2 \;=\; \alpha\begin{pmatrix} 3 \\ 4 \end{pmatrix}\ , \quad \alpha\in\mathbb{R}\ .$$

Similarly we find that

$$\boldsymbol{v}_1 \;=\; \alpha\begin{pmatrix} -4 \\ 3 \end{pmatrix}\ , \quad \alpha\in\mathbb{R}\ ,$$

are the eigenvectors corresponding to $\lambda_1=100$.

Next, we find

$$P_C(\lambda) \;=\; (1-\lambda)^2\ .$$

The only root is $\lambda=1$. We say that this eigenvalue has *algebraic multiplicity* 2. The eigenvectors are $\begin{pmatrix}\alpha & 0\end{pmatrix}^T$ for any $\alpha\in\mathbb{R}$.

Finally,

$$P_H(\lambda) \;=\; \lambda^2 - 125\lambda + 5092 \;.$$

The discriminant is

$$D \;=\; (-125)^2 - 4 \cdot 1 \cdot 5092 \;=\; -4743 \;.$$

This is negative, so $\boldsymbol{H}$ has no eigenvalue $\lambda \in \mathbb{R}$. ∎

Example 3.3. The condition that $\boldsymbol{A} - \lambda \boldsymbol{I}$ be singular has the immediate consequence that the diagonal elements in a triangular (especially a diagonal) matrix are eigenvalues of that matrix. For instance

$$\boldsymbol{A} \;=\; \begin{pmatrix} 1 & 2 & 3 \\ & 4 & 5 \\ & & 6 \end{pmatrix}$$

has the eigenvalues $\lambda_1 = 1$, $\lambda_2 = 4$ and $\lambda_3 = 6$.

The $n \times n$ identity matrix $\boldsymbol{I}$ gives $\boldsymbol{I}\boldsymbol{v} = \boldsymbol{v}$ for any $\boldsymbol{v} \in \mathbb{R}^n$, cf Example 1.10. This means that $\boldsymbol{I}$ has the eigenvalue $\lambda = 1$ with arithmetic multiplicity n, and any $\boldsymbol{v} \in \mathbb{R}^n$ is an eigenvector. ∎

These examples show that if $\boldsymbol{A}$ has an eigensolution, then the eigenvector is not unique: From the linearity conditions in Eq. (1.7) and the above definition of the eigenproblem it follows that

$$\boldsymbol{A}(\alpha \boldsymbol{v}) \;=\; \alpha(\boldsymbol{A}\,\boldsymbol{v}) \;=\; \alpha(\lambda \boldsymbol{v}) \;=\; \lambda(\alpha \boldsymbol{v}) \;,$$

so $\alpha \boldsymbol{v}$ is also an eigenvector for any choice of the scalar α. Therefore it may be useful to think of an eigenvector as a direction in $\mathbb{R}^n$, cf the geometric view in Example 1.1. When we talk about "the eigenvector" in the following, we assume that it is normalized so that $\|\boldsymbol{v}\| = 1$. This makes $\boldsymbol{v}$ unique, except for sign.

We saw in Examples 3.2 and 3.3 that a matrix $\boldsymbol{A}$ may have several different eigenvalues λ_j with corresponding eigenvectors $\boldsymbol{v}_j$,

$$\boldsymbol{A}\,\boldsymbol{v}_j \;=\; \lambda_j \boldsymbol{v}_j \;, \quad j = 1, 2, \ldots, p \;. \tag{3.3}$$

Further, it may happen (as for instance with $\boldsymbol{I}$) that there are linearly independent eigenvectors corresponding to the same eigenvalue. This case is also covered by the formulation (3.3). If $\lambda_i = \lambda_j$, then

$$\boldsymbol{A}\,(\alpha \boldsymbol{v}_i + \beta \boldsymbol{v}_j) \;=\; \alpha(\lambda_i \boldsymbol{v}_i) + \beta(\lambda_j \boldsymbol{v}_j) \;=\; \lambda_i\,(\alpha \boldsymbol{v}_i + \beta \boldsymbol{v}_j)$$

for any choice of scalars α and β. This shows that any linear combination of eigenvectors corresponding to a *multiple eigenvalue* is also an eigenvector.

Without proof we give the following facts about the eigenproblem for an $n \times n$ matrix $\boldsymbol{A}$. We shall use 5° in Section 3.2.3.

1° $\boldsymbol{A}$ has at most n different eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p \;, \quad p \le n \;.$

2° If $\lambda_1, \lambda_2, \ldots, \lambda_p$ are mutually different, then the corresponding eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_p$ are linearly independent.

3° If $\boldsymbol{A}$ is symmetric and $\lambda_i \neq \lambda_j$ then $\boldsymbol{v}_i$ and $\boldsymbol{v}_j$ are orthogonal.

4° If $\boldsymbol{A}$ is symmetric, then it is possible to choose an orthonormal set of n eigenvectors for $\boldsymbol{A}$.

5° If all the elements in $\boldsymbol{A}$ are positive and the sum of the elements in each column is one, then $\lambda = 1$ is an eigenvalue for $\boldsymbol{A}$.

Example 3.4. Let

$$\boldsymbol{A} = \begin{pmatrix} 2 & -1 & -2 \\ -1 & 2 & -2 \\ -2 & -2 & -1 \end{pmatrix} , \qquad \boldsymbol{B} = \begin{pmatrix} 15 & -6 & -12 \\ 12 & -3 & -12 \\ 12 & -6 & -9 \end{pmatrix} .$$

Both matrices have the eigenvalues $\lambda_1 = -3$ and $\lambda_2 = \lambda_3 = 3$. For the symmetric matrix $\boldsymbol{A}$ we may choose the orthonormal eigenvectors

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} , \quad \boldsymbol{v}_2 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix} , \quad \boldsymbol{v}_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} .$$

The normalized eigenvectors for $\boldsymbol{B}$ may be chosen as

$$\boldsymbol{v}_1 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} , \quad \boldsymbol{v}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} , \quad \boldsymbol{v}_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} .$$

■

Now, assume that $\boldsymbol{A}$ has n linearly independent eigenvectors, and make a matrix $\boldsymbol{V}$ with $\boldsymbol{v}_j$ as its j^{th} column,

$$\boldsymbol{V} = \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_n \end{pmatrix} . \tag{3.4}$$

Then

$$\boldsymbol{A}\,\boldsymbol{V} = \begin{pmatrix} \lambda_1\boldsymbol{v}_1 & \lambda_2\boldsymbol{v}_2 & \cdots & \lambda_n\boldsymbol{v}_n \end{pmatrix} = \boldsymbol{V}\,\boldsymbol{\Lambda} , \tag{3.5}$$

where $\boldsymbol{\Lambda}$ is the diagonal matrix with the eigenvalues,

$$\boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} .$$

In Eq. (3.5) we used the column-wise formulation of a matrix–matrix product, Eq. (1.5), and the result from Exercise 1.13 about the product of a general and a diagonal matrix.

The assumption about the columns in $\boldsymbol{V}$ being linearly independent, implies that $\boldsymbol{V}$ is nonsingular. Therefore $\boldsymbol{V}^{-1}$ exists, and if we multiply with this matrix from the right on both sides in Eq. (3.5), we get

$$\boldsymbol{A} = \boldsymbol{V}\,\boldsymbol{\Lambda}\,\boldsymbol{V}^{-1} . \tag{3.6}$$

This *eigen factorization* of $\boldsymbol{A}$ is essential for the applications that we give in the next section.

If $\boldsymbol{A}$ is symmetric, then the assumption about n linearly independent eigenvectors is satisfied, and if we choose them to be orthonormal, cf fact 4° above, then $\boldsymbol{V}$ is an orthogonal matrix, and Eq. (3.6) simplifies to

$$\boldsymbol{A} = \boldsymbol{V}\,\boldsymbol{\Lambda}\,\boldsymbol{V}^{T} . \tag{3.7}$$

Example 3.5. We return to the question from Section 2.6: Let $\boldsymbol{A}$ be symmetric; is it also *positive definite* ?

To answer that question we may use the factorization in Eq. (3.7).

$$\begin{aligned} \boldsymbol{x}^T \boldsymbol{A}\, \boldsymbol{x} &= \boldsymbol{x}^T \boldsymbol{V}\, \boldsymbol{\Lambda}\, \boldsymbol{V}^T \boldsymbol{x} \\ &= \boldsymbol{y}^T \boldsymbol{\Lambda}\, \boldsymbol{y} \;=\; \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \ . \end{aligned}$$

Here, we introduced $\boldsymbol{y} = \boldsymbol{V}^T \boldsymbol{x}$. Since $\boldsymbol{V}$ is nonsingular, there exists a nonzero $\boldsymbol{y}$ corresponding to any nonzero $\boldsymbol{x}$. The right-hand side must be positive for any nonzero $\boldsymbol{y}$, and this is clearly the case if and only if all the $\lambda_i > 0$.

Therefore, if all the eigenvalues for a symmetric matrix are strictly positive, then it is an SPD matrix. It should be said, that it is computationally cheaper to check via the LDL^T or Cholesky factorization, as described in Section 2.6. ∎

Example 3.6. MATLAB has a built-in function `eig` for computing eigensolutions. Let `A` be a square matrix, then the command

```
la = eig(A)
```

returns a vector `la` with the eigenvalues of $\boldsymbol{A}$. The command

```
[V, La] = eig(A)
```

returns the arrays `V` and `La`, representing respectively $\boldsymbol{V}$ and $\boldsymbol{\Lambda}$ in Eq. (3.6), or Eq. (3.7) if $\boldsymbol{A}$ is symmetric.

The algorithm used is based on the following relation: If $\boldsymbol{S}$ is a nonsingular matrix, then

$$\boldsymbol{A}\, \boldsymbol{v} \;=\; \lambda\, \boldsymbol{v} \quad \Leftrightarrow \quad \boldsymbol{S}\, \boldsymbol{A}\, \boldsymbol{S}^{-1} \boldsymbol{S}\, \boldsymbol{v} \;=\; \lambda\, \boldsymbol{S}\, \boldsymbol{v} \ .$$

The transformation from $\boldsymbol{A}$ to $\tilde{\boldsymbol{A}} = \boldsymbol{S}\boldsymbol{A}\boldsymbol{S}^{-1}$ is called a *similarity transformation*. We see that it preserves the eigenvalues, but changes the eigenvectors. If we know the eigensolutions and if the eigenvectors are linearly independent, then it follows from Eq. (3.6) that if we take $\boldsymbol{S} = \boldsymbol{V}^{-1}$, then $\tilde{\boldsymbol{A}} = \boldsymbol{\Lambda}$.

We do not know the eigensolutions, but it is possible to compute a sequence of similarity transformations in such a way that the sequence of $\tilde{\boldsymbol{A}}$-matrices converge to an upper triangular matrix $\boldsymbol{U}$. According to Example 3.3 the diagonal elements in $\boldsymbol{U}$ are the eigenvalues, and it is fairly easy to compute the eigenvectors of $\boldsymbol{U}$ and transform them back to eigenvectors of $\boldsymbol{A}$.

The algorithm used in MATLAB uses orthogonal matrices for the similarity transformations, ie $\tilde{\boldsymbol{A}} = \boldsymbol{Q}\boldsymbol{A}\boldsymbol{Q}^T$. In case $\boldsymbol{A}$ is symmetric, the symmetry is preserved, and the sequence converges to a diagonal matrix. ∎

3.2. Applications of eigensolutions

In this section we shall answer the first three questions raised in Example 3.1.

3.2.1. Differential equations

First, we look at a model that involves n functions of t, $y_1(t)$, $y_2(t)$, ..., $y_n(t)$. We

do not know closed form expressions for the functions, but know their values at $t = 0$,

$$y_i(0) \;=\; \beta_i \;, \quad i = 1, 2, \ldots, n \;.$$

Further, we know that for $t > 0$ the functions satisfy a linear system of coupled differential equations,

$$\begin{aligned} y_1' &= a_{11}y_1 + a_{12}y_2 + \cdots + a_{1n}y_n \\ y_2' &= a_{21}y_1 + a_{22}y_2 + \cdots + a_{2n}y_n \\ &= \quad \vdots \\ y_1' &= a_{n1}y_1 + a_{n2}y_2 + \cdots + a_{nn}y_n \end{aligned} \;.$$

We introduce the vector function $\boldsymbol{y}(t)$ and the vector $\boldsymbol{\beta}$ with the i^{th} elements $y_i(t)$ and β_i, respectively. Further, we let the $n \times n$ matrix $\boldsymbol{A}$ hold the coefficients a_{ij}, and see that the problem can be written in the form

$$\boldsymbol{y}'(t) \;=\; \boldsymbol{A}\,\boldsymbol{y}(t) \quad \text{for } t > 0 \;, \qquad \boldsymbol{y}(0) \;=\; \boldsymbol{\beta} \;. \tag{3.8}$$

We want to solve this *initial value problem* for a *linear system of differential equations.*

The solution is quite easy to find with the help of the eigensolutions of $\boldsymbol{A}$, provided that there are n linearly independent eigenvectors, $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, …, $\boldsymbol{v}_n$. In that case we can use them as basis in $\mathbb{R}^n$, ie we can write

$$\boldsymbol{y}(t) \;=\; \tilde{y}_1(t)\boldsymbol{v}_1 + \tilde{y}_2(t)\boldsymbol{v}_2 + \cdots + \tilde{y}_n(t)\boldsymbol{v}_n \;=\; \boldsymbol{V}\,\tilde{\boldsymbol{y}}(t) \;, \tag{3.9}$$

where $\boldsymbol{V}$ is the matrix with the eigenvectors as columns, cf Eq. (3.4), and the vector $\tilde{\boldsymbol{y}}$ has $\tilde{y}_i$ as the i^{th} element. By differentiation we get

$$\boldsymbol{y}'(t) \;=\; \tilde{y}_1'(t)\boldsymbol{v}_1 + \tilde{y}_2'(t)\boldsymbol{v}_2 + \cdots + \tilde{y}_n'(t)\boldsymbol{v}_n \;=\; \boldsymbol{V}\,\tilde{\boldsymbol{y}}'(t) \;.$$

We insert these relations in the differential equation and use the factorization Eq. (3.6) for $\boldsymbol{A}$,

$$\boldsymbol{V}\,\tilde{\boldsymbol{y}}'(t) \;=\; \boldsymbol{A}\,\boldsymbol{V}\,\tilde{\boldsymbol{y}}(t) \;=\; \left(\boldsymbol{V}\,\boldsymbol{\Lambda}\,\boldsymbol{V}^{-1}\right)\boldsymbol{V}\,\tilde{\boldsymbol{y}}(t) \;.$$

So the differential equation is equivalent to

$$\tilde{\boldsymbol{y}}'(t) \;=\; \boldsymbol{\Lambda}\,\tilde{\boldsymbol{y}}(t) \;.$$

The starting values are given by $\boldsymbol{V}\tilde{\boldsymbol{y}}(0) = \boldsymbol{\beta}$, or

$$\tilde{\boldsymbol{y}}(0) \;=\; \tilde{\boldsymbol{\beta}} \quad \text{with} \quad \tilde{\boldsymbol{\beta}} \;=\; \boldsymbol{V}^{-1}\boldsymbol{\beta} \;.$$

We see that this change of variables transforms the original system of coupled differential equations to a system of scalar differential equations,

$$\begin{aligned} &\textbf{for } j = 1, 2, \ldots, n \\ &\quad \tilde{y}_j'(t) \;=\; \lambda_j\,\tilde{y}_j(t) \quad \text{for } t > 0 \;, \qquad \tilde{y}_j(0) \;=\; \tilde{\beta}_j \;. \end{aligned}$$

The solution is

$$\tilde{y}_j(t) \;=\; \tilde{\beta}_j e^{\lambda_j t} \;, \quad j = 1, 2, \ldots, n \;.$$

In vector form we can express this as

$$\tilde{\boldsymbol{y}}(t) \;=\; [\tilde{\boldsymbol{\beta}}]\boldsymbol{\varphi}(t)\;, \tag{3.10}$$

where

$$\boldsymbol{\varphi}(t) \;=\; \begin{pmatrix} e^{\lambda_1 t} \\ e^{\lambda_2 t} \\ \vdots \\ e^{\lambda_n t} \end{pmatrix}$$

and $[\boldsymbol{x}]$ denotes the diagonal matrix with diagonal elements given by the vector $\boldsymbol{x}$,

$$[\boldsymbol{x}] \;=\; \begin{pmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_n \end{pmatrix}\;. \tag{3.11}$$

Finally, by combining Eqs. (3.9) – (3.11) we see that the components of the solution can be expressed in the form

$$y_i(t) \;=\; \alpha_{i1}e^{\lambda_1 t} + \alpha_{i2}e^{\lambda_2 t} + \cdots + \alpha_{in}e^{\lambda_n t}\;, \quad \text{with } \alpha_{ij} \;=\; v_{ij}\tilde{\beta}_j\;. \tag{3.12}$$

Note that every component in both $\tilde{\boldsymbol{y}}(t)$ and $\boldsymbol{y}(t)$ is a linear combination of the exponential functions $e^{\lambda_j t}$, where the λ_j are the eigenvalues of $\boldsymbol{A}$. The coefficients in the linear combinations depend on the initial values and on the eigenvectors of $\boldsymbol{A}$.

Example 3.7. The drug concentration problem in Example 1.2 can be modelled by a *compartment model*. This has the form of Eq. (3.8), with n being the number of compartments.

As an example consider the initial value problem

$$\boldsymbol{y}'(t) \;=\; \begin{pmatrix} -0.18 & 0.15 \\ 0.024 & -0.27 \end{pmatrix}\boldsymbol{y}(t) \quad \text{for } t > 0\;, \qquad \boldsymbol{y}(0) \;=\; \begin{pmatrix} 0 \\ 1 \end{pmatrix}\;.$$

The eigensolutions of the matrix are

$$\lambda_1 \;=\; -0.15\;,\;\; \boldsymbol{v}_1 \;=\; \begin{pmatrix} 0.981 \\ 0.196 \end{pmatrix}\;, \qquad \lambda_2 \;=\; -0.30\;,\;\; \boldsymbol{v}_2 \;=\; \begin{pmatrix} -0.781 \\ 0.625 \end{pmatrix}\;.$$

We used the MATLAB function `eig` from Example 3.6 to compute the eigensolutions, and display the results with at most three decimals. Going through the steps of computing $\tilde{\boldsymbol{\beta}}$ and the coefficients α_{ij}, we find that the two components of $\boldsymbol{y}(t)$ are

$$y_1(t) \;=\; e^{-0.15t} - e^{-0.30t}\;, \qquad y_2(t) \;=\; 0.2e^{-0.15t} + 0.8e^{-0.30t}\;.$$

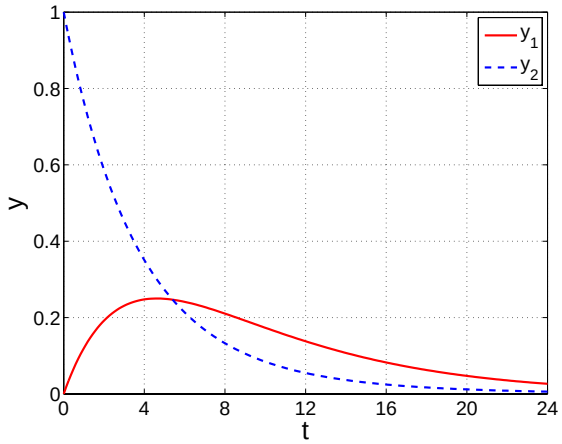


Figure 3.1. *Two components of* $\boldsymbol{y}(t)$.

The points in figure 1.2 have coordinates $(t_r, y_1(t_r))$ for 17 different values of t_r. ∎

Example 3.8. In Example 3.1 we asked the question: If we want to write the solution to $\boldsymbol{y}' = \boldsymbol{A}\boldsymbol{y}$ as $\boldsymbol{y}(t) = e^{\boldsymbol{A}t}\boldsymbol{c}$, how shall we interpret $e^{\boldsymbol{A}t}$?

The answer is not easy to see from Eq. (3.12). However, we can replace Eq. (3.10) by the equivalent formula

$$\tilde{\boldsymbol{y}}(t) \;=\; [\boldsymbol{\varphi}(t)]\,\tilde{\boldsymbol{\beta}} \;=\; [\boldsymbol{\varphi}(t)]\,\boldsymbol{V}^{-1}\boldsymbol{\beta}\;.$$

Then $\boldsymbol{y}(t) = \boldsymbol{V}\tilde{\boldsymbol{y}}(t)$ leads to $\boldsymbol{y}(t) = \boldsymbol{V}\,[\boldsymbol{\varphi}(t)]\,\boldsymbol{V}^{-1}\boldsymbol{\beta}$. This implies that a valid interpretation is

$$e^{\boldsymbol{A}t} \;=\; \boldsymbol{V}\,[\boldsymbol{\varphi}(t)]\,\boldsymbol{V}^{-1}\;,$$

where $[\boldsymbol{\varphi}(t)]$ is the diagonal matrix with $([\boldsymbol{\varphi}(t)])_{ii} = e^{\lambda_i t}$ and $\boldsymbol{V}$ is the matrix with $\boldsymbol{V}_{:,i} = \boldsymbol{v}_i$. ∎

3.2.2. Simplification of quadratics

This application also relies on using eigenvectors as basis. In Example 1.20 we asked a question of the form: which points (x_1, x_2) satisfy the equation $q(\boldsymbol{x}) = 0$, where $q(\boldsymbol{x})$ is the quadratic

$$q(\boldsymbol{x}) \;=\; c_1x_1^2 + c_2x_2^2 + c_3x_1x_2 + c_4x_1 + c_5x_2 + c_6\;.$$

It simplifies the problem if we can get rid of the mixed term $c_3x_1x_2$ involving the product of different elements in $\boldsymbol{x}$. To achieve that we introduce the matrix $\boldsymbol{A}$ and vector $\boldsymbol{b}$,

$$\boldsymbol{A} \;=\; \begin{pmatrix} c_1 & \frac{1}{2}c_3 \\ \frac{1}{2}c_3 & c_2 \end{pmatrix}\;, \quad \boldsymbol{b} \;=\; \begin{pmatrix} c_4 \\ c_5 \end{pmatrix}\;. \tag{3.13}$$

It follows that

$$\begin{aligned} \boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x} \;&=\; \boldsymbol{x}^T\begin{pmatrix} c_1x_1 + \frac{1}{2}c_3x_2 \\ \frac{1}{2}c_3x_1 + c_2x_2 \end{pmatrix} \;=\; c_1x_1^2 + c_2x_2^2 + c_3x_1x_2\;, \\ \boldsymbol{b}^T\boldsymbol{x} \;&=\; c_4x_1 + c_5x_2\;, \end{aligned}$$

so that

$$\begin{aligned} q(\boldsymbol{x}) \;&=\; \boldsymbol{x}^T\boldsymbol{A}\,\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c_6 \\ &=\; \boldsymbol{x}^T\boldsymbol{V}\,\boldsymbol{\Lambda}\,\boldsymbol{V}^T\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c_6 \\ &=\; \tilde{\boldsymbol{x}}^T\boldsymbol{\Lambda}\,\tilde{\boldsymbol{x}} + \tilde{\boldsymbol{b}}^T\tilde{\boldsymbol{x}} + c_6 \\ &=\; \lambda_1\tilde{x}_1^2 + \lambda_2\tilde{x}_2^2 + \tilde{b}_1\tilde{x}_1 + \tilde{b}_2\tilde{x}_2 + c_6\;. \end{aligned} \tag{3.14}$$

Here we used the factorization Eq. (3.7) for the symmetric matrix $\boldsymbol{A}$, and introduced

$$\tilde{\boldsymbol{x}} \;=\; \boldsymbol{V}^T\boldsymbol{x} \;\;\Leftrightarrow\;\; \boldsymbol{x} \;=\; \boldsymbol{V}\,\tilde{\boldsymbol{x}}\;, \qquad \tilde{\boldsymbol{b}} \;=\; \boldsymbol{V}^T\boldsymbol{b} \;\;\Leftrightarrow\;\; \boldsymbol{b} \;=\; \boldsymbol{V}\,\tilde{\boldsymbol{b}}\;,$$

so the elements in $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{b}}$ are the coordinates of $\boldsymbol{x}$ and $\boldsymbol{b}$ with respect to the basis given by the eigenvectors of $\boldsymbol{A}$. The change of variables makes it easier to discuss the solution to the equation $q(\boldsymbol{x}) = 0$. We shall illustrate that with two examples.

Example 3.9. The quadratic in Example 1.20 has the coefficients

$$\boldsymbol{c} \;=\; \begin{pmatrix} 73 & 52 & -72 & 0 & 0 & -100 \end{pmatrix}^T \;.$$

The corresponding matrix is

$$\boldsymbol{A} \;=\; \begin{pmatrix} 73 & -36 \\ -36 & 52 \end{pmatrix} \;,$$

which has the eigensolutions

$$\lambda_1 = 25 \;, \quad \boldsymbol{v}_1 = \begin{pmatrix} 0.6 \\ 0.8 \end{pmatrix} \;, \qquad \lambda_2 = 100 \;, \quad \boldsymbol{v}_2 = \begin{pmatrix} -0.8 \\ 0.6 \end{pmatrix} \;.$$

Therefore, Eq. (3.14) leads to

$$q(\boldsymbol{x}) \;=\; 25\tilde{x}_1^2 + 100\tilde{x}_2^2 - 100 \;=\; 0 \;.$$

This agrees with the equation that we derived in Example 1.20, using some rather tedious calculations. ∎

Example 3.10. For the quadratic

$$q(\boldsymbol{x}) \;=\; 18x_1^2 + 32x_2^2 + 48x_1x_2 - 40x_1 - 95x_2 + 37.5$$

we use Eq. (3.13) to get

$$\boldsymbol{A} \;=\; \begin{pmatrix} 18 & 24 \\ 24 & 32 \end{pmatrix} \;, \quad \boldsymbol{b} \;=\; \begin{pmatrix} -40 \\ -95 \end{pmatrix} \;.$$

The eigensolutions of $\boldsymbol{A}$ are (presented in the form defined by Eq. (3.5))

$$\boldsymbol{V} \;=\; \begin{pmatrix} 0.6 & -0.8 \\ 0.8 & 0.6 \end{pmatrix} \;, \quad \boldsymbol{\Lambda} \;=\; \begin{pmatrix} 50 & \\ & 0 \end{pmatrix} \;.$$

So $\lambda_1 = 50$, $\lambda_2 = 0$, and the eigenvectors are the same as in the previous example. Further,

$$\tilde{\boldsymbol{b}} \;=\; \boldsymbol{V}^T\boldsymbol{b} \;=\; \begin{pmatrix} -100 \\ -25 \end{pmatrix} \;,$$

and Eq. (3.14) takes the form

$$q(\boldsymbol{x}) \;=\; 50\tilde{x}_1^2 - 100\tilde{x}_1 - 25\tilde{x}_2 + 37.5 \;.$$

Therefore, the equation $q(\boldsymbol{x}) = 0$ is equivalent to

$$\tilde{x}_2 \;=\; 2x_1^2 - 4\tilde{x}_1 + 1.5 \;=\; 2\,(\tilde{x}_1 - 1)^2 - 0.5 \;.$$

This is the equation for the parabola shown in Figure 3.2 below.

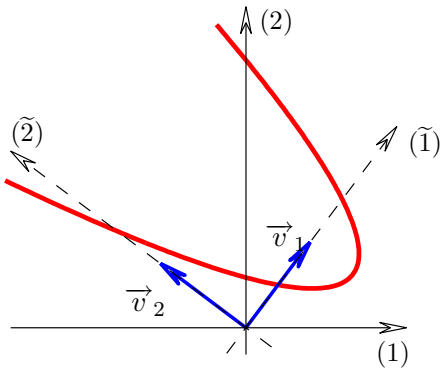


Figure 3.2. *Rotated parabola.* ∎

3.2.3. Google's PageRank

Finally, we give a simplified description of the PageRank principle used in Google. Figure 3.3 shows a very small web with only $n = 6$ pages and 10 *links*, indicated by arrows. For instance there are two links from page 4: to pages 5 and 6.

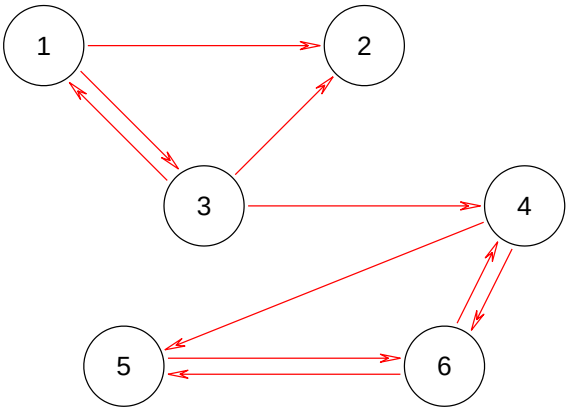


Figure 3.3. *Web with $n = 6$ pages.*

Let page i have the PageRank p_i. This number should reflect the importance of that page in such a way that it is better for the page to be pointed at from a page with a large PageRank than from a less important page. In order to quantify the PageRank we set up a *model* that simulates a web-surfer.

Suppose that the surfer is at page j, which has $q = q_j$ *out-links*. The model says that the surfer may either follow one of these links or choose a new page at random. In both cases it is assumed that all possible candidates for being the next page have the same probability to be chosen. In the follow-a-link case this means that each of the q target pages "inherits" the PageRank contribution $\frac{1}{q}p_j$. This can be represented by an $n\times n$ matrix $\boldsymbol{H}$, where all elements are zero, except for $h_{ij} = \frac{1}{q}$ for the pages pointed at from the j^{th} page, which has q out-links. For instance

$$\boldsymbol{H} = \begin{pmatrix} 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 1 & 0 \end{pmatrix}$$

is the $\boldsymbol{H}$-matrix for the web in Figure 3.3.

The web may contain "*dangling pages*", as for instance page 2 in Figure 3.3, from which there are no out-links. From such a page the surfer chooses any of the n pages at random. This corresponds to a modified matrix $\boldsymbol{S}$, which is equal to $\boldsymbol{H}$ except for the columns where all elements are zero. These columns are replaced by $\frac{1}{n}\boldsymbol{e}$, where $\boldsymbol{e}$ is the vector of all ones. The $\boldsymbol{S}$-matrix corresponding to the above $\boldsymbol{H}$ is

$$\boldsymbol{S} = \begin{pmatrix} 0 & \frac{1}{6} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{6} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{6} & \frac{1}{3} & 0 & 0 & \frac{1}{2} \\ 0 & \frac{1}{6} & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{6} & 0 & \frac{1}{2} & 1 & 0 \end{pmatrix}.$$

Finally, the surfer may occasionally want to pick a new page at random. If this was done from every page, the corresponding $\boldsymbol{S}$-matrix would be $\frac{1}{n}\boldsymbol{E}$, where $\boldsymbol{E}$ is the matrix of all ones. However, let us assume this random choice has probability $1-\alpha$. Then the matrix to use is

$$\boldsymbol{G} \;=\; \alpha\,\boldsymbol{S} + \frac{1-\alpha}{n}\,\boldsymbol{E}\;. \tag{3.15}$$

This is known as the *Google matrix*.

Returning to the PageRank, $g_{ij}p_j$ is the contribution from page j to the PageRank of page i, and we demand that

$$\begin{gathered}\textbf{for } i = 1, 2, \ldots, n \;:\\ g_{i1}p_1 + g_{i2}p_2 + 0\cdots + g_{in}p_n \;=\; p_i\;,\end{gathered}$$

or, in matrix–vector notation:

$$\boldsymbol{G}\,\boldsymbol{p} \;=\; \boldsymbol{p}\;. \tag{3.16}$$

We recognize this as an eigenproblem with $\lambda = 1$ and $\boldsymbol{p}$, the vector of PageRanks, is a corresponding eigenvector. This may seem a bit magical, but the matrix $\boldsymbol{G}$ satisfies the conditions of 5° on page 38, so $\lambda = 1$ is indeed one of its eigenvalues. The corresponding eigenvector depends on the choice of α, as shown below for the 6×6 matrix $\boldsymbol{G}$ corresponding to the web in Figure 3.3.

α	0.5	0.6	0.7	0.8	0.9
	0.116	0.102	0.085	0.064	0.037
	0.145	0.133	0.115	0.090	0.054
$\boldsymbol{p}$	0.124	0.111	0.093	0.071	0.042
	0.176	0.181	0.187	0.195	0.206
	0.199	0.213	0.230	0.254	0.286
	0.239	0.262	0.290	0.326	0.375

The eigenvector is normalized such that the sum of the PageRanks is one. We see that all these choices of α agree that the most important pages are 6, 5 and 4 in that order. It is believed that $\alpha = 0.85$ is used by Google. In Section 4.3 we describe how the relevant pages are found.

Example 3.11. The number n of pages registered by Google is not known, except that $n > 10^{10}$, so the matrices are huge, and it is impossible, just to store $\boldsymbol{G}$. However, $\boldsymbol{H}$ is very *sparse*, ie most of the elements are zero, and this can be exploited. The size of the matrix prevents the use of the method mentioned in Example 3.6. Instead a very simple algorithm is employed. This is the so-called *power method*, which has the form

Choose $\boldsymbol{p}^{[0]}$ with $p_1^{[0]} + \cdots + p_n^{[0]} = 1$ (for instance $\boldsymbol{p}^{[0]} = \frac{1}{n}\boldsymbol{e}$)
for $k = 1, 2, \ldots, K$:
$\quad \boldsymbol{p}^{[k]} \;=\; \boldsymbol{G}\,\boldsymbol{p}^{[k-1]}$
end

It can be shown that not only does $\boldsymbol{G}$ have the eigenvalue $\lambda = 1$, but also the other eigenvalues satisfy $|\lambda_j| < 1$, and that these properties imply that the sequence $\boldsymbol{p}^{[1]}, \boldsymbol{p}^{[2]}, \ldots$ converges to $\boldsymbol{p}$. For our 6×6 example with $\alpha = 0.85$ we get

$\boldsymbol{p}^{[0]}$	$\boldsymbol{p}^{[1]}$	$\boldsymbol{p}^{[2]}$	$\boldsymbol{p}^{[3]}$	$\boldsymbol{p}^{[4]}$	$\boldsymbol{p}^{[5]}$
0.167	0.096	0.082	0.068	0.062	0.057
0.167	0.167	0.123	0.103	0.090	0.083
0.167	0.119	0.089	0.077	0.068	0.064
0.167	0.167	0.193	0.187	0.198	0.196
0.167	0.190	0.230	0.244	0.255	0.261
0.167	0.261	0.281	0.321	0.327	0.339

Already $\boldsymbol{p}^{[2]}$ shows that most important pages are 6, 5 and 4 in that order.

In the computation of $\boldsymbol{G}\boldsymbol{p}^{[k-1]}$ we can exploit the sparsity of $\boldsymbol{H}$ when computing of the contribution $\alpha\boldsymbol{H}\boldsymbol{p}^{[k-1]}$. The contribution from a dangling page (no j) is to add $\alpha/n \cdot p_j^{[k-1]}$ to every component of $\boldsymbol{p}^{[k]}$. Finally, the condition on the sum of the elements in $\boldsymbol{p}^{[0]}$ is preserved:

$$\boldsymbol{e}^T\boldsymbol{p}^{[k-1]} = p_1^{[k-1]} + \cdots + p_n^{[k-1]} = 1\ , \quad k = 1, 2, \ldots\ .$$

This is useful when computing the contribution involving $\boldsymbol{E}\boldsymbol{p}^{[k-1]}$: Exploiting the outer product defined on page 3 we get

$$\boldsymbol{E}\,\boldsymbol{p}^{[k-1]} = \boldsymbol{e}\,\boldsymbol{e}^T\boldsymbol{p}^{[k-1]} = \boldsymbol{e}\ .$$

Therefore, the contribution $\frac{1-\alpha}{n}\,\boldsymbol{E}\boldsymbol{p}^{[k-1]}$ consists simply in adding $\frac{1-\alpha}{n}$ to each element in $\alpha\boldsymbol{S}\boldsymbol{p}^{[k-1]}$. ∎

4. Linear Least Squares Problems

Example 4.1. We have applied different loads p_i to an elastic spring, and measured the resulting lengths ℓ_i.

i	p_i	ℓ_i
1	0.8	7.97
2	1.6	10.2
3	2.4	14.2
4	3.2	16.0
5	4.0	21.2

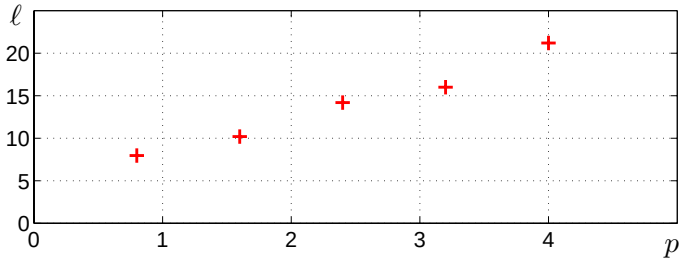


Figure 4.1. *Measured length (ℓ) as function of load (p).*

The figure shows the points (p_i, ℓ_i). They seem to be close to a straight line with the equation

$$\ell \;=\; x_1 + x_2 p \;,$$

for some values of the *parameters* x_1 and x_2. In physics this mathematical model is known as *Hooke's law* of elasticity. The first parameter, x_1, is the undisturbed length of the spring, and $1/x_2$ is the so-called *spring constant*. ∎

The problem of determining (a good approximation to) the parameters in Hooke's law is an example of a *linear data fitting problem*: Given *measurements* (t_i, y_i), $i = 1, 2, \ldots, m$, where y_i is assumed to be a perturbed value from some "background" function Y,

$$y_i \;=\; Y(t_i) + \varepsilon_i \;.$$

The ε_i are measurement errors. Further, we are given a *fitting model*

$$M(\boldsymbol{x}, t) \;=\; x_1 f_1(t) + x_2 f_2(t) + \cdots + x_n f_n(t) \;, \tag{4.1}$$

where the f_j are given functions of t. We wish to choose the parameters $\boldsymbol{x}$ such that $M(\boldsymbol{x}, t)$ is close to $Y(t)$ in a certain range of t-values, which includes the given t_i. In other words, we wish to find $\boldsymbol{x}$ such that the error function

$$e(\boldsymbol{x}, t) \;=\; Y(t) - M(\boldsymbol{x}, t)$$

is close to zero for all t in the desired range. This task, however, is impossible: We only have discrete information about Y, and furthermore, this information, the y_i, is inflicted with errors. The best we can do is to find $\boldsymbol{x}$ such that the *residuals*

$$r_i(\boldsymbol{x}) \;=\; y_i - M(\boldsymbol{x}, t_i) \tag{4.2}$$

are close to zero for all $i = 1, 2, \ldots, m$. Traditionally, this goal is interpreted as: Find $\hat{\boldsymbol{x}}$, the set of parameters that minimizes the function

$$\varphi(\boldsymbol{x}) \;=\; r_1(\boldsymbol{x})^2 + r_2(\boldsymbol{x})^2 + \cdots + r_m(\boldsymbol{x})^2 \;. \tag{4.3}$$

The minimizer $\hat{\boldsymbol{x}}$ is the so-called *least squares solution*. In the next sections we shall describe how it is computed.

4.1. Overdetermined systems

Combining (4.2) and (4.1) we see that the i^{th} residual is

$$r_i(\boldsymbol{x}) \;=\; y_i - (x_1 f_1(t_i) + x_2 f_2(t_i) + \cdots + x_n f_n(t_i)) \;.$$

This means that the *residual vector* $\boldsymbol{r} = \boldsymbol{r}(\boldsymbol{x})$ is

$$\boldsymbol{r}(\boldsymbol{x}) \;=\; \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} - \begin{pmatrix} f_1(t_1)x_1 \;+\; f_2(t_1)x_2 \;+\cdots+\; f_n(t_1)x_n \\ f_1(t_2)x_1 \;+\; f_2(t_2)x_2 \;+\cdots+\; f_n(t_2)x_n \\ \vdots \\ f_1(t_m)x_1 + f_2(t_m)x_2 + \cdots + f_n(t_m)x_n \end{pmatrix} \;.$$

Further, by comparison with Eq. (1.3) we see that

$$\boldsymbol{r}(\boldsymbol{x}) \;=\; \boldsymbol{y} - \boldsymbol{F}\boldsymbol{x} \;, \tag{4.4}$$

where $\boldsymbol{F}$ is the $m \times n$ matrix

$$\boldsymbol{F} \;=\; \begin{pmatrix} f_1(t_1) & f_2(t_1) & \cdots & f_n(t_1) \\ f_1(t_2) & f_2(t_2) & \cdots & f_n(t_2) \\ \vdots & & & \\ f_1(t_m) & f_2(t_m) & \cdots & f_n(t_m) \end{pmatrix} \;.$$

Example 4.2. For the problem with the elastic spring in Example 4.1 the free variable is denoted p instead of t and the observations are called ℓ_i instead of y_i. Taking this into account, the fitting model has the form Eq. (4.1) with $f_1(t) = 1$, $f_2(t) = t$ and we see that

$$\boldsymbol{y} \;=\; \begin{pmatrix} 7.97 \\ 10.2 \\ 14.2 \\ 16.0 \\ 21.2 \end{pmatrix} \;, \qquad \boldsymbol{F} \;=\; \begin{pmatrix} 1 & 0.8 \\ 1 & 1.6 \\ 1 & 2.4 \\ 1 & 3.2 \\ 1 & 4.0 \end{pmatrix} \;.$$

■

Ideally, we would like to find $\boldsymbol{x}$ so that $\boldsymbol{r}(\boldsymbol{x}) = \boldsymbol{0}$, ie so that

$$\boldsymbol{F}\boldsymbol{x} \;=\; \boldsymbol{y} \;. \tag{4.5}$$

In data fitting problems it is always the case that $m > n$, ie the system involves more equations than unknowns. We say that Eq. (4.5) is an *overdetermined system of equations.*

Example 4.3. Let $m = 3$ and $n = 2$. Figure 4.2 is a geometric illustration in the 3-dimensional space.

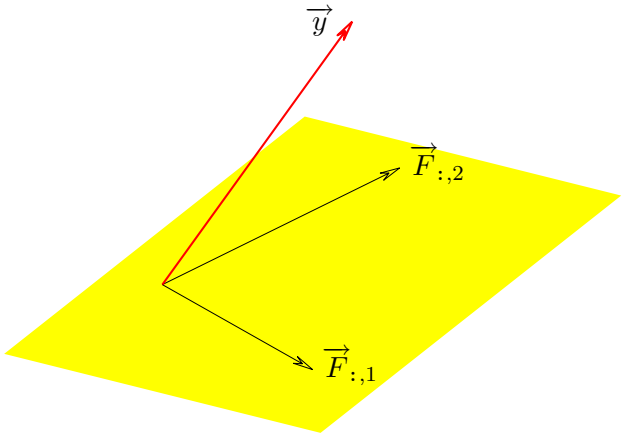


Figure 4.2. *Geometric illustration of an overdetermined system.*

As discussed on page 24, a vector $\boldsymbol{z} = \boldsymbol{F}\boldsymbol{x}$ is in the column range of $\boldsymbol{F}$. This is the shaded plane in the figure, spanned by the columns of $\boldsymbol{F}$. The vector $\overrightarrow{y}$ in the figure is outside this plane, and therefore the system (4.5) has no solution in this case. We shall use the somewhat sloppy notation, that $\hat{\boldsymbol{x}}$ is the least squares solution to

$$\boldsymbol{F}\boldsymbol{x} \simeq \boldsymbol{y} .$$

■

Combining Eqs. (4.3) and (1.2) we see that

$$\varphi(\boldsymbol{x}) = \|\boldsymbol{r}(\boldsymbol{x})\|^2 .$$

The least squares solution $\hat{\boldsymbol{x}}$ minimizes this, and it follows, that an equivalent definition is that $\hat{\boldsymbol{x}}$ minimizes $\|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{x}\|$. In words: find the vector $\boldsymbol{F}\hat{\boldsymbol{x}}$ in the column range of $\boldsymbol{F}$, which is closest to the right-hand side vector $\boldsymbol{y}$. The corresponding residual is

$$\hat{\boldsymbol{r}} = \boldsymbol{y} - \boldsymbol{F}\hat{\boldsymbol{x}} . \tag{4.6}$$

Example 4.4. Proceeding with the $m = 3$, $n = 2$ case from Example 4.3, we recall from Example 1.9 that $\|\boldsymbol{r}\| = |\overrightarrow{r}|$, the Euclidean length of the residual vector. Intuitively this is smallest when $\boldsymbol{r}$ is orthogonal to the column range, cf Figure 4.3.

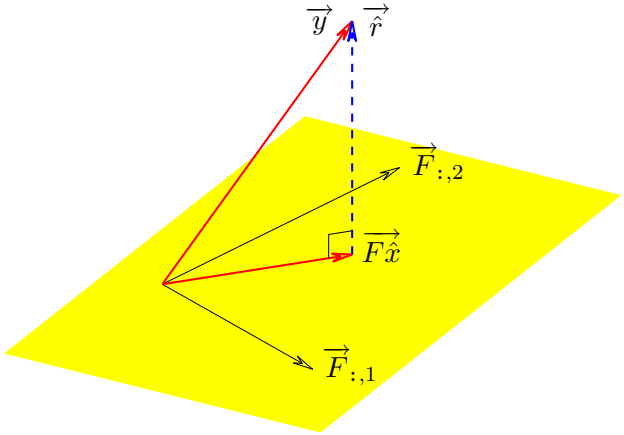


Figure 4.3. *Geometric illustration of least squares solution.*

■

A vector in the column range has the form $\boldsymbol{z} = \boldsymbol{F}\boldsymbol{x}$, where $\boldsymbol{x}$ is an arbitrary n-vector. Generalizing from the example, the vectors $\boldsymbol{z}$ and $\hat{\boldsymbol{r}}$ should be orthogonal, so their scalar product should be zero:

$$0 \;=\; (\boldsymbol{F}\boldsymbol{x})^T\,\hat{\boldsymbol{r}} \;=\; \boldsymbol{x}^T\boldsymbol{F}^T\hat{\boldsymbol{r}}\ .$$

This is satisfied for all $\boldsymbol{x}$ if

$$\boldsymbol{0} \;=\; \boldsymbol{F}^T\hat{\boldsymbol{r}} \;=\; \boldsymbol{F}^T(\boldsymbol{y}-\boldsymbol{F}\hat{\boldsymbol{x}})\ .$$

By rearranging the terms we see the least squares solution $\hat{\boldsymbol{x}}$ satisfies

$$\left(\boldsymbol{F}^T\boldsymbol{F}\right)\hat{\boldsymbol{x}} \;=\; \boldsymbol{F}^T\boldsymbol{y}\ . \tag{4.7}$$

This system of linear equations is known as the *normal equations*.

The matrix

$$\boldsymbol{A} \;=\; \boldsymbol{F}^T\boldsymbol{F}$$

is $n\times n$, and in Example 2.22 we saw that it is symmetric and positive definite, provided that the columns in $\boldsymbol{F}$ are linearly independent. We shall assume that this is the case. Then we can show that $\hat{\boldsymbol{x}}$ is indeed the minimizer of $\|\boldsymbol{r}(\boldsymbol{x})\|$: If we change $\boldsymbol{x}$ from $\hat{\boldsymbol{x}}$ to $\hat{\boldsymbol{x}}+\boldsymbol{u}$, then the residual changes to

$$\boldsymbol{r} \;=\; \boldsymbol{y}-\boldsymbol{F}\,(\hat{\boldsymbol{x}}+\boldsymbol{u}) \;=\; \hat{\boldsymbol{r}}-\boldsymbol{F}\boldsymbol{u}\ ,$$

and, exploiting Eq. (1.2),

$$\begin{aligned}\|\boldsymbol{r}\|^2 \;=\; \boldsymbol{r}^T\boldsymbol{r} \;&=\; (\hat{\boldsymbol{r}}-\boldsymbol{F}\boldsymbol{u})^T\,(\hat{\boldsymbol{r}}-\boldsymbol{F}\boldsymbol{u})\\ &=\; \hat{\boldsymbol{r}}^T\hat{\boldsymbol{r}}-\hat{\boldsymbol{r}}^T\boldsymbol{F}\,\boldsymbol{u}-\boldsymbol{u}^T\boldsymbol{F}^T\hat{\boldsymbol{r}}+\boldsymbol{u}^T\boldsymbol{F}^T\boldsymbol{F}\,\boldsymbol{u}\\ &=\; \hat{\boldsymbol{r}}^T\hat{\boldsymbol{r}}+\boldsymbol{u}^T\boldsymbol{A}\,\boldsymbol{u}\;>\hat{\boldsymbol{r}}^T\hat{\boldsymbol{r}}\qquad\text{if}\quad\boldsymbol{u}\neq\boldsymbol{0}\ .\end{aligned}$$

Summarizing: If the columns in $\boldsymbol{F}$ are linearly independent, then the least squares solution $\hat{\boldsymbol{x}}$ can be found by solving the normal equations (4.7). Since $\boldsymbol{A} = \boldsymbol{F}^T\boldsymbol{F}$ is an spd matrix, the least squares solution is unique, and it may be found via the Cholesky factorization of $\boldsymbol{A}$, cf Section 2.6.

Example 4.5. For the problem with the elastic spring in Example 4.1 we saw in Example 4.1 that

$$\boldsymbol{y} \;=\; \begin{pmatrix}7.97\\10.2\\14.2\\16.0\\21.2\end{pmatrix}\ ,\qquad \boldsymbol{F} \;=\; \begin{pmatrix}1&0.8\\1&1.6\\1&2.4\\1&3.2\\1&4.0\end{pmatrix}\ .$$

The columns in $\boldsymbol{F}$ are linearly independent. The normal equations and their solution are

$$\begin{pmatrix}5&12\\12&35.2\end{pmatrix}\begin{pmatrix}\hat{x}_1\\\hat{x}_2\end{pmatrix} \;=\; \begin{pmatrix}69.57\\192.776\end{pmatrix}\ ,\qquad \hat{\boldsymbol{x}} \;=\; \begin{pmatrix}4.2360\\4.0325\end{pmatrix}\ .$$

This shows, cf Example 4.1, that the undisturbed length of the spring and the spring constant are estimated to be 4.2360 and $1/4.0325 \simeq 0.2480$, respectively.

The data and the fit $M(\hat{\boldsymbol{x}}, p) = \hat{x}_1 + \hat{x}_2 p$ is shown in Figure 4.4.

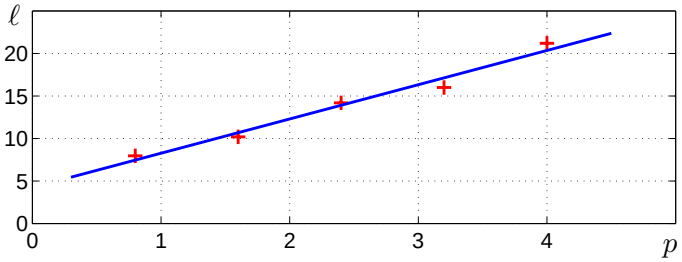


Figure 4.4. *Spring data and least squares fit.*

■

Example 4.6. We consider the problem from Example 1.2: We have measured the drug concentration y_i in the kidneys at times t_i after the drug was taken. The points shown in Figure 1.2 are $(t_i, Y(t_i))$, where Y is the background function. In practice the measurements have errors, and Figure 4.5 gives an example of results from a "practical" experiment.

The phenomenon may be modelled by a compartment model with two compartments, and according to Eq. (3.12) the background function has the form

$$Y(t) \;=\; c_1 e^{\lambda_1 t} + c_2 e^{\lambda_2 t}$$

for some values of $c_1,\ c_2, \lambda_1,\ \lambda_2$. This should be exploited in the choice of fitting model. If λ_1 and λ_2 are known, then

$$M(\boldsymbol{x}, t) \;=\; x_1 e^{\lambda_1 t} + x_2 e^{\lambda_2 t}$$

is appropriate. With the data in Figure 4.5 and $\lambda_1 = -0.15, \lambda_2 = -0.30$, cf Example 3.7, we get $\hat{x}_1 = 1.006$, $\hat{x}_2 = -1.011$

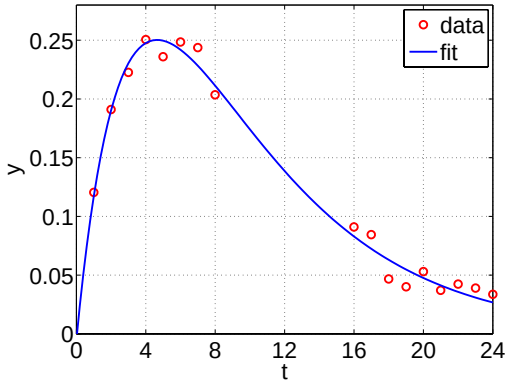


Figure 4.5. *Data and fit.* $M(\hat{\boldsymbol{x}}, t) = 1.006e^{-0.15t} - 1.011e^{-0.30t}$.

If we do not know λ_1 and λ_2, then it is appropriate to use a fitting model with four parameters

$$M_{[4]}(\boldsymbol{x}, t) \;=\; x_1 e^{x_3 t} + x_2 e^{x_4 t} \ .$$

In this model x_3 and x_4 appear nonlinearly, and we cannot use the approach leading to the linear system known as the normal equations. (The matrix $\boldsymbol{F}$ would depend on the unknowns x_3 and x_4). There is software, for instance in MATLAB and R, for finding the least squares solution also in this case, but it is outside the scope of this lecture note to discuss it. We only mention that for the data in Figure 4.5 and the above 4-parameter model the least squares solution is

$$\hat{\boldsymbol{x}}_{[4]} \;=\; \begin{pmatrix} 1.4291 & -1.4196 & -0.16157 & -0.25906 \end{pmatrix}^T \ .$$

The corresponding sum of the squared residuals is $\varphi(\hat{\boldsymbol{x}}_4) = 0.001458$. This is indeed

smaller than the result from the 2-parameter model, $\varphi(\hat{\boldsymbol{x}}) = 0.001512$. If we plotted the graph of $M_{[4]}(\hat{\boldsymbol{x}}_{[4]}, t)$, it would be almost indistinguishable from the graph of the least squares 2-parameter model $M(\hat{\boldsymbol{x}}, t)$. ∎

4.2. QR factorization

Let $\boldsymbol{F}$ be an $m\times n$ matrix with $m \geq n$. A *QR factorization* of $\boldsymbol{F}$ is a relation of the form

$$\boldsymbol{F} \;=\; \boldsymbol{Q}\begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{pmatrix}\;, \tag{4.8}$$

where $\boldsymbol{Q}$ is an $m\times m$ orthogonal matrix and the $n\times n$ matrix $\boldsymbol{R}$ is upper triangular. The $\boldsymbol{0}$ indicates an $(m-n)\times n$ matrix with all elements equal to zero. According to the rules of matrix–matrix products (see Eq. (1.5) and Eq. (1.9)) these zeros imply that the last $m-n$ columns in $\boldsymbol{Q}$ do not affect the product in Eq. (4.8). Let $\breve{\boldsymbol{Q}}$ denote this part of $\boldsymbol{Q}$ and let $\hat{\boldsymbol{Q}}$ be the $m\times n$ matrix consisting of the first n columns in $\boldsymbol{Q}$. It follows that Eq. (4.8) is equivalent to the so-called "*economy-size*" (or "*skinny*") QR factorization,

$$\boldsymbol{F} \;=\; \breve{\boldsymbol{Q}}\,\boldsymbol{R}\;, \tag{4.9}$$

This is illustrated in Figure 4.6

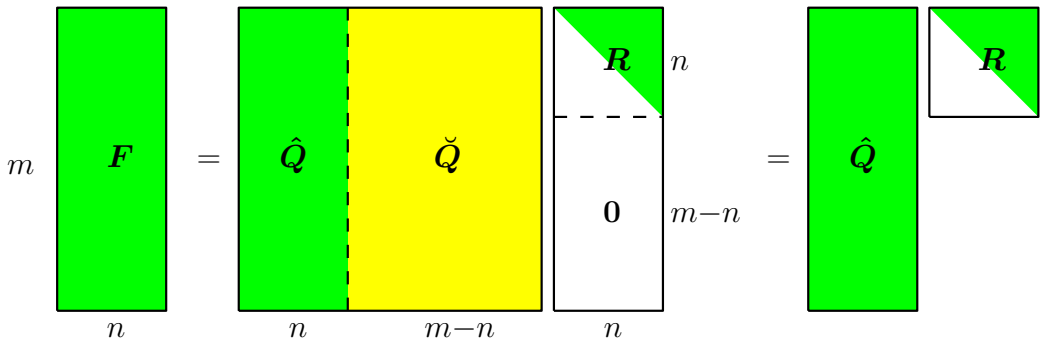


Figure 4.6. *QR factorization.*

It is outside the scope of this lecture note to show that any $\boldsymbol{F}$ has a QR factorization and to discuss methods for computing it. Instead we shall demonstrate how it can be used to solve the problem

$$\text{minimize } \|\boldsymbol{y} - \boldsymbol{F}\boldsymbol{x}\| \;.$$

First, we let $\boldsymbol{Q}$ be an arbitrary, orthogonal matrix, and use Eq. (1.2): $\|\boldsymbol{r}\|^2 = \boldsymbol{r}^T\boldsymbol{r}$, Eq. (1.6): $(\boldsymbol{Q}\boldsymbol{r})^T = \boldsymbol{r}^T\boldsymbol{Q}^T$, Eq. (2.18): $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}$ and Example 1.10: $\boldsymbol{I}\boldsymbol{r} = \boldsymbol{r}$:

$$\|\boldsymbol{Q}\boldsymbol{r}\|^2 \;=\; (\boldsymbol{Q}\boldsymbol{r})^T(\boldsymbol{Q}\boldsymbol{r}) \;=\; \boldsymbol{r}^T\boldsymbol{Q}^T\boldsymbol{Q}\,\boldsymbol{r} \;=\; \boldsymbol{r}^T\boldsymbol{I}\,\boldsymbol{r} \;=\; \boldsymbol{r}^T\boldsymbol{r} \;=\; \|\boldsymbol{r}\|^2\;.$$

This shows that $\boldsymbol{Q}\boldsymbol{r}$ has the same norm as $\boldsymbol{r}$, so a minimizer of $\|\boldsymbol{Q}\boldsymbol{r}(\boldsymbol{x})\|$ is also a minimizer for $\|\boldsymbol{r}(\boldsymbol{x})\|$. This holds for any orthogonal matrix, and if we use $\boldsymbol{Q}^T$ with

$\boldsymbol{Q}$ from the QR factorization, we get

$$\boldsymbol{Q}^T\boldsymbol{r}(\boldsymbol{x}) \;=\; \boldsymbol{Q}^T(\boldsymbol{y}-\boldsymbol{F}\boldsymbol{x}) \;=\; \boldsymbol{Q}^T\boldsymbol{y} - \begin{pmatrix}\boldsymbol{R}\\ \boldsymbol{0}\end{pmatrix}\boldsymbol{x} \;=\; \begin{pmatrix}\hat{\boldsymbol{Q}}^T\boldsymbol{y}-\boldsymbol{R}\boldsymbol{x}\\ \breve{\boldsymbol{Q}}^T\boldsymbol{y}\end{pmatrix} .$$

Finally, the definition of the norm, Eq. (1.2), implies that

$$\left\| \begin{pmatrix}\hat{\boldsymbol{Q}}^T\boldsymbol{y}-\boldsymbol{R}\boldsymbol{x}\\ \breve{\boldsymbol{Q}}^T\boldsymbol{y}\end{pmatrix} \right\|^2 \;=\; \|\hat{\boldsymbol{Q}}^T\boldsymbol{y}-\boldsymbol{R}\boldsymbol{x}\|^2 + \|\breve{\boldsymbol{Q}}^T\boldsymbol{y}\|^2 .$$

This is clearly minimized when $\boldsymbol{x}$ is chosen such that $\hat{\boldsymbol{Q}}^T\boldsymbol{y}-\boldsymbol{R}\boldsymbol{x}=\boldsymbol{0}$.

Summarizing, we have derive a method for computing the least squares solution $\hat{\boldsymbol{x}}$. This method consists in computing the economy-size factorization $\boldsymbol{F}=\hat{\boldsymbol{Q}}\boldsymbol{R}$ followed by solving the quadratic system

$$\boldsymbol{R}\hat{\boldsymbol{x}} \;=\; \hat{\boldsymbol{Q}}^T\boldsymbol{y} . \tag{4.10}$$

Since $\boldsymbol{R}$ is upper triangular, we can use back substitution to solve this system.

Example 4.7. In Example 2.28 we described what happens in MATLAB when we issue the command `x = A \ b`, with a square `A`. If `F` and `y` are the MATLAB representations of $\boldsymbol{F}\in\mathbb{R}^{m\times n}$ and $\boldsymbol{y}\in\mathbb{R}^m$, then

```
x = F \ y
```

returns `x`, the MATLAB representation for $\hat{\boldsymbol{x}}$, the least squares solution to $\boldsymbol{F}\boldsymbol{x}\simeq\boldsymbol{y}$. The solution is found via the QR factorization, because this method is less affected by rounding errors than the solution via the normal equations.

MATLAB has a built-in function `qr` for computing the QR factorization. The command

```
[Q, R] = qr(F,0)
```

returns the economy-size QR factorization. ∎

4.3. Singular value decomposition

Let $\boldsymbol{F}$ be an $m\times n$ matrix with $m\geq n$. The *SVD*: *singular value decomposition* of $\boldsymbol{F}$ is the factorization

$$\boldsymbol{F} \;=\; \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^T , \tag{4.11}$$

where the matrices $\boldsymbol{U}\in\mathbb{R}^{m\times n}$ and $\boldsymbol{V}\in\mathbb{R}^{n\times n}$ have orthonormal columns and $\boldsymbol{\Sigma}$ is the diagonal matrix

$$\boldsymbol{\Sigma} \;=\; \begin{pmatrix}\sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_n\end{pmatrix} ,$$

The σ_j are the *singular values*. They are nonnegative and ordered so that

$$\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_p>0 , \quad \sigma_{p+1}=\cdots=\sigma_n=0 . \tag{4.12}$$

If the columns in $\boldsymbol{F}$ are linearly independent, then all the σ_j are strictly positive, ie $p = n$.

Let $\boldsymbol{u}_j = \boldsymbol{U}_{:,j}$, the j^{th} column in $\boldsymbol{U}$. This is known as the j^{th} *left singular vector*. Similarly, $\boldsymbol{v}_j = \boldsymbol{V}_{:,j}$, $j = 1, \ldots, n$ are the *right singular vectors*.

SVD is a powerful tool for analyzing linear mappings: First, we see that the $\{\boldsymbol{v}_j\}$ form an orthonormal basis for $\mathbb{R}^n$. This means that any n-vector $\boldsymbol{x}$ can be expressed as

$$\boldsymbol{x} \;=\; \boldsymbol{V}\,\tilde{\boldsymbol{x}}\;, \tag{4.13}$$

where $\tilde{\boldsymbol{x}}$ is the vector of coordinates with respect to the basis $\{\boldsymbol{v}_j\}$, cf Section 1.5. The matrix $\boldsymbol{V}$ is orthogonal, and therefore Eq. (4.13) is equivalent to

$$\tilde{\boldsymbol{x}} \;=\; \boldsymbol{V}^T\boldsymbol{x}\;. \tag{4.14}$$

Next, by means of Equations (4.11) and (4.14) we get

$$\begin{aligned}\boldsymbol{F}\,\boldsymbol{x} \;&=\; \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^T\boldsymbol{x}\\ &=\; \boldsymbol{U}\,\boldsymbol{\Sigma}\,\tilde{\boldsymbol{x}}\\ &=\; \sigma_1\tilde{x}_1\boldsymbol{u}_1 + \sigma_2\tilde{x}_2\boldsymbol{u}_2 + \cdots + \sigma_p\tilde{x}_p\boldsymbol{u}_p\;.\end{aligned} \tag{4.15}$$

In the last reformulation we exploited that $(\boldsymbol{\Sigma}\tilde{\boldsymbol{x}})_j = \sigma_j\tilde{x}_j$, Eq. (1.9) and that $\sigma_j = 0$ for $j > p$. Eq. (4.15) shows that vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$ are in the *column range* of $\boldsymbol{F}$, cf page 24. They actually form an orthonormal basis for this subspace of $\mathbb{R}^m$. Further, if $p < n$, then the vectors $\boldsymbol{v}_{p+1}, \ldots, \boldsymbol{v}_n$ form an orthonormal basis for the *nullspace* of $\boldsymbol{F}$.

Now we look at the *inverse problem*: Given $\boldsymbol{F}$ and $\boldsymbol{y}$, find $\boldsymbol{x}$ such that the norm of the difference $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{F}\boldsymbol{x}$ is as small as possible. From the discussion in Section 4.1 we know that

$$\boldsymbol{F}\,\hat{\boldsymbol{x}} \;=\; \hat{\boldsymbol{y}}\;, \qquad \hat{\boldsymbol{r}} \;=\; \boldsymbol{y} - \boldsymbol{F}\,\hat{\boldsymbol{x}}\;,$$

where $\hat{\boldsymbol{y}}$ is in the range of $\boldsymbol{F}$ and $\hat{\boldsymbol{r}}$ is orthogonal to this subspace of $\mathbb{R}^m$. This condition implies that

$$\begin{aligned}0 \;&=\; \boldsymbol{u}_j^T(\boldsymbol{y} - \boldsymbol{F}\,\hat{\boldsymbol{x}})\\ &=\; \boldsymbol{u}_j^T(\boldsymbol{y} - \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^T\hat{\boldsymbol{x}})\\ &=\; \boldsymbol{u}_j^T(\boldsymbol{y} - \boldsymbol{U}\,\boldsymbol{\Sigma}\,\tilde{\hat{\boldsymbol{x}}})\\ &=\; \boldsymbol{u}_j^T\boldsymbol{y} - \sigma_j\,\tilde{\hat{x}}_j\;, \quad j = 1, 2, \ldots, p\;.\end{aligned}$$

The relation $\boldsymbol{u}_j^T\boldsymbol{U}\boldsymbol{\Sigma}\tilde{\hat{\boldsymbol{x}}} = \sigma_j\tilde{\hat{x}}_j$ follows from the orthonormality of the columns in $\boldsymbol{U}$ and the result of Exercise 1.13. Since $\boldsymbol{F}\boldsymbol{v}_j = \boldsymbol{0}$ for $j > p$, it follows that $\hat{\boldsymbol{x}}$ can have arbitrary coordinates $\tilde{\hat{x}}_j$, $j = p{+}1, \ldots, n$. Summarizing, we have

$$\hat{\boldsymbol{x}} \;= \boldsymbol{V}\,\tilde{\hat{\boldsymbol{x}}} \quad \text{with} \quad \tilde{\hat{x}}_j \;=\; \begin{cases}\boldsymbol{u}_j^T\boldsymbol{y}/\sigma_j & j = 1, 2, \ldots, p\;,\\ \alpha_j & j = p{+}1, \ldots, n\;,\end{cases} \tag{4.16}$$

where the α_j are arbitrary.

Example 4.8. In Example 2.7 we looked at the problem $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$ with

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \qquad \boldsymbol{y} = \begin{pmatrix} 8 \\ 20 \\ 32 \end{pmatrix}.$$

The SVD of the matrix is $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ with (rounded to three decimals)

$$\boldsymbol{U} = \begin{pmatrix} 0.215 & 0.887 & 0.408 \\ 0.521 & 0.250 & -0.816 \\ 0.826 & -0.388 & 0.408 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} 16.848 & & \\ & 1.068 & \\ & & 0 \end{pmatrix},$$

$$\boldsymbol{V} = \begin{pmatrix} 0.480 & -0.777 & -0.408 \\ 0.572 & -0.076 & 0.816 \\ 0.665 & 0.625 & -0.408 \end{pmatrix}.$$

We seed that $p = 2$ and that the nullspace of $\boldsymbol{A}$ is spanned by the last column in $\boldsymbol{V}$. This is equivalent to the nullspace basis $\begin{pmatrix} 1 & -2 & 1 \end{pmatrix}^T$ given in Example 1.15.

The coordinates of $\boldsymbol{y}$ with respect to the basis given by the columns in $\boldsymbol{U}$ are

$$\tilde{\boldsymbol{y}} = \boldsymbol{U}^T\boldsymbol{y} = \begin{pmatrix} 38.573 \\ -0.323 \\ 0 \end{pmatrix}.$$

$\tilde{y}_3 = 0$ shows that $\boldsymbol{y}$ has no component outside the range of $\boldsymbol{A}$, ie the system is consistent. ∎

Example 4.9. There is a close relation between the SVD for a matrix $\boldsymbol{F}$ and the eigensolutions of the corresponding normal equations matrix $\boldsymbol{A} = \boldsymbol{F}^T\boldsymbol{F}$. If we insert Eq. (4.11) and exploit that $\boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$, we get

$$\boldsymbol{A} = \boldsymbol{V}\,\boldsymbol{\Sigma}^T\,\boldsymbol{U}^T\boldsymbol{U}\,\boldsymbol{\Sigma}\boldsymbol{V}^T = \boldsymbol{V}\,\boldsymbol{\Sigma}^2\boldsymbol{V}^T,$$

where

$$\boldsymbol{\Sigma}^2 = \begin{pmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{pmatrix}.$$

Comparing this with Eq. (3.7) we see that the right singular vectors are eigenvectors of $\boldsymbol{A}$, and that $\sigma_j = \sqrt{\lambda_j}$. We know that the eigenvalues are nonnegative, and assume that they are numbered in decreasing order. The left singular vectors may afterwards be computed by means of Eq. (4.15):

$$\boldsymbol{u}_j = \frac{1}{\sigma_j}\boldsymbol{F}\boldsymbol{v}_j, \quad j = 1, 2, \ldots, p.$$

∎

Example 4.10. MATLAB has a built-in function `svd` for computing SVD. The command

```
s = svd(F)
```

returns a vector `s` with the singular values of $\boldsymbol{F}$. The command

```
[U,S,V] = svd(F,0)
```

returns the matrices $\boldsymbol{U}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{V}$. Instead of the method outlined in the previous example `svd` uses a special purpose algorithm that avoids the possible loss of information that may result from forming the normal equations matrix. ∎

We shall mention another application of SVD, which is closely related to a statistical method known as *PCA*: *principal component analysis*. It is based on an alternative formulation of the decomposition. Using the rules from Section 1.3 we see that $\boldsymbol{F} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$ is equivalent to

$$\boldsymbol{F} \;=\; \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T + \cdots + \sigma_n \boldsymbol{u}_n \boldsymbol{v}_n^T \;. \tag{4.17}$$

In words: Any matrix can be expressed as a weighted sum of outer products. The weights are the singular values, and they reflect the importance of each contribution. It sometimes happens that most of the information in $\boldsymbol{F}$ is accounted for by the approximation

$$\boldsymbol{F}_k \;=\; \sigma_1 \boldsymbol{u}_1 \boldsymbol{v}_1^T + \sigma_2 \boldsymbol{u}_2 \boldsymbol{v}_2^T + \cdots + \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^T \;,$$

where k is maybe much smaller than n. We can write this approximation in the form

$$\boldsymbol{F}_k \;=\; \boldsymbol{U}_k \, \boldsymbol{\Sigma}_k \, \boldsymbol{V}_k^T \;, \tag{4.18}$$

where $\boldsymbol{U}_k$ and $\boldsymbol{V}_k$ consist of the first k columns in $\boldsymbol{U}$ and $\boldsymbol{V}$, respectively, and $\boldsymbol{\Sigma}_k$ is the upper left $k \times k$ corner of $\boldsymbol{\Sigma}$.

Example 4.11. PCA is widely used in, for instance, medicine to identify certain combinations of parameters that can be used to explain most of the variation in the observed data.

As a specific example, however, we shall give a simplified description of a search machine like Google. We set up an $m \times n$ matrix, where m is the number of pages and n is the number of selected *keywords*. The element f_{ij} is a measure of the importance of the j^{th} keyword in the i^{th} page. As mentioned in Example 3.11, $m > 10^{10}$, and n is maybe 50 000, so $\boldsymbol{F}$ is a huge matrix (but it is very sparse).

Now, given a query, we set up a vector $\boldsymbol{q}$, according to the keywords, and compute its coordinates with respect to the columns in $\boldsymbol{V}_k$,

$$\tilde{\boldsymbol{q}} \;=\; \boldsymbol{V}_k^T \boldsymbol{q} \;.$$

Let J denote the set of indices for which $\tilde{q}_j \geq \tau_1 \, \|\tilde{\boldsymbol{q}}\|$ where τ_1 is a certain threshold, for instance $\tau_1 = 0.05$. If J has only one element, then $\boldsymbol{y} = \boldsymbol{U}_{:,J}^{[k]}$ is the corresponding vector in $\mathbb{R}^m$. If J has more than one element, then

$$\boldsymbol{y} \;=\; \boldsymbol{U}_J^{[k]} \, \boldsymbol{\Sigma}_J^{[k]} \, \tilde{\boldsymbol{q}}_J \;,$$

where index J signifies that we only include the contributions corresponding to the indices from J.

The *hits* are those pages for which $y_i \geq \tau_2 \, \|\boldsymbol{y}\|$, where for instance $\tau_2 = 0.01$. The ranking of page i is some combination of the value y_i and the PageRank p_i, cf Section 3.2.3.

If k in Eq. (4.18) is too small, we may both overlook interesting pages and get hits that we really do not want. For larger k these misclassifications are reduced, but the computational cost increases. See for instance the book by Eldén for a discussion of this. ∎

Literature

There are many books on linear algebra. For those readers that want to know more about the subject we can recommend

G. Strang, *Linear Algebra and its Application,* 4th edition, Cengage, 2006

J. Eising, *Lineær algebra,* Matematisk Institut, DTU, 1993. (In Danish)

The following book focuses on numerical aspects applications

L.N. Trefethen and D. Bau, *Numerical linear algebra,* SIAM 1997,

while

L. Eldén, *Matrix Methods In Data Mining and Pattern Recognition,* SIAM, 2007

treats both numerical linear algebra and applications. This includes the mathematics involved in Google.

Further, we would like to recommend

E.S. Allman and J.A. Rhodes, *Mathematical Models in Biology,* Cambridge University Press, 2004

M.W. Berry and M. Browne, *Understanding Search Engines,* SIAM, 1999

A.N. Langville and C.D. Meyer, *Google's PageRank and Beyond,* Princeton University Press, 2006

Notation

Vectors are normally written as lower case bold letters, and matrices are upper case, bold letters.

$\boldsymbol{A}$	Typical name for a matrix.
$\boldsymbol{A}^T$	The transpose of matrix $\boldsymbol{A}$.
$\boldsymbol{A}^{-1}$	Matrix $\boldsymbol{A}$ inverse.
$a_{ij},\ (\boldsymbol{A})_{ij}$	The element in position (i, j) in the matrix $\boldsymbol{A}$.
$\boldsymbol{A}_{i,:}$	The row vector with the elements in the i^{th} row of $\boldsymbol{A}$.
$\boldsymbol{A}_{:,j}$	The column vector with the elements in the j^{th} column of $\boldsymbol{A}$.
$m,\ n$	Positive integers.
$\mathbb{R}$	The set of real numbers.
$\mathbb{R}^n$	The set of vectors with n real elements.
$\mathbb{R}^{m\times n}$	The set of matrices with m rows and n columns, where all elements are real.
$\overrightarrow{v}$	Typical name for a geometric vector.
$\boldsymbol{x}$	Typical name for a vector. Unless otherwise stated, $\boldsymbol{x}$ is a column vector (cf page 2).
x_i	i^{th} element in the vector $\boldsymbol{x}$. i is the *index* of the element.
$\alpha,\ \beta$	Typical names for a scalar.
λ	Eigenvalue.
$\boldsymbol{\Lambda}$	Diagonal matrix with $(\boldsymbol{\Lambda})_{jj} = \lambda_j$, the j^{th} eigenvalue.

In MATLAB examples we use typewriter font to indicate the MATLAB representation of a variable. For instance `A` is the MATLAB variable that represents matrix $\boldsymbol{A}$.

List of Danish Words

I mange tilfælde er det danske ord for et begreb identisk med det engelske på nær en mindre ændring i stavemåde, fx *vector* $\sim$ *vektor, orthogonal* $\sim$ *ortogonal*. I andre tilfælde fremkommer den danske betegnelse ved direkte oversættelse af den engelske. Nedenfor giver vi en liste over de danske betegnelser i de tilfælde, hvor oversættelsen ikke er umiddelbar.

English	*Dansk*
augmented matrix	totalmatrix
back substitution	tilbage-løsning
column	søjle
column range	billedrum
complete solution	fuldstændig løsning
eigenvalue	egenværdi
eigenvector	egenvektor
forward substitution	fremad-løsning
Gaussian elimination	Gauss-elimination
identity matrix	enhedsmatrix
kernel	kærne
mapping	afbildning
matrix	matrix. Én matrix, matricen, flere matricer. En *matrice* er en støbeform
nonsingular	regulær (om matrix)
quadratic	2. grads polynomium
rank	rang
row	række
singular	singulær (om matrix)
square	kvadratisk eller kvadratformet (om matrix)
transpose	transponeret
triangular matrix	trekantmatrix

Index