5

Matrices

Natasha Maurits

> After reading this chapter you know:
>
> - what matrices are and how they can be used,
> - how to perform addition, subtraction and multiplication of matrices,
> - that matrices represent linear transformations,
> - the most common special matrices,
> - how to calculate the determinant, inverse and eigendecomposition of a matrix, and
> - what the decomposition methods SVD, PCA and ICA are, how they are related and how they can be applied.

5.1 What Are Matrices and How Are They Used?

Matrices, in mathematics, are rectangular arrays of (usually) numbers. Their entries are called *elements* and can also be symbols or even expressions. Here, we discuss matrices of numbers. Of course, these numbers can be of any type, such as integer, real or complex (see Sect. 1.2). For most practical applications, the matrix elements have specific meanings, such as the distance between cities, demographic measures such as survival probabilities (represented in a *Leslie matrix*) or the absence or presence of a path between nodes in a *graph* (represented in an *adjacency matrix*), which is applied in *network theory*. Network theory has seen a surge of interest in recent years because of its wide applicability in the study of e.g. social communities, the world wide web, the brain, regulatory relationships between genes, metabolic pathways and logistics. We here first consider the simple example of roads between the four cities in Fig. 5.1.

N. Maurits (✉)

Department of Neurology, University Medical Center Groningen, Groningen, The Netherlands
e-mail: n.m.maurits@umcg.nl

A matrix **M** describing the distances between the cities is given by:

$$M = \begin{bmatrix} 0 & 22 & 14 & 10 \\ 22 & 0 & 8 & 12 \\ 14 & 8 & 0 & 13 \\ 10 & 12 & 13 & 0 \end{bmatrix}$$

Here, each row corresponds to a 'departure' city A–D and each column to an 'arrival' city A–D. For example, the distance from city B to city C (second row, third column) is 8, as is the distance from city C to city B (third row, second column). Cities A and B are not connected directly, but can be reached through cities C or D. In both cases, the distance between cities A and B amounts to 22 (14 + 8 or 12 + 10; first row, second column and second row, first column).

One of the advantages of using a matrix instead of a table is that they can be much more easily manipulated by computers in large-scale calculations. For example, matrices are used to store all relevant data for weather predictions and for predictions of air flow patterns around newly designed *airfoils*.

Historically, matrices, which were first known as 'arrays', have been used to solve systems of linear equations (see Chap. 2 and Sect. 5.3.1) for centuries, even dating back to 1000 years BC in China. The name 'matrix', however, wasn't introduced until the nineteenth century, when James Joseph Sylvester thought of a matrix as an entity that can give rise to smaller matrices by removing columns and/or rows, as if they are 'born' from the womb of a common parent. Note that the word 'matrix' is related to the Latin word for mother: 'mater'.

By removing columns or rows and considering just one row or one column of a matrix, we obtain so-called row- or column-matrices, which are simply row- or column-vectors, of
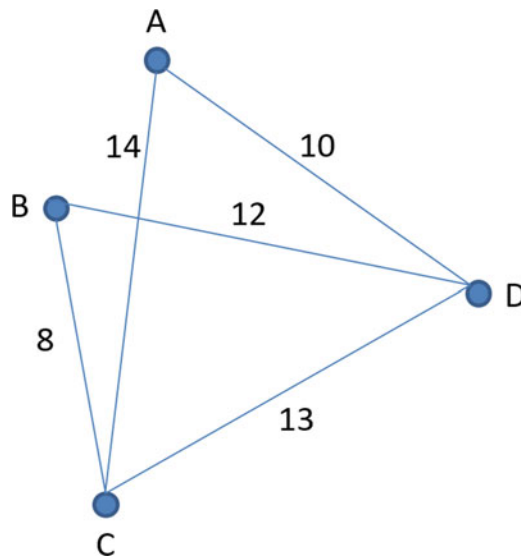

Fig. 5.1 Four cities A, B, C and D and the distances between them. In this example city A cannot be reached from city B directly, but can be reached via city C or city D.

course. Hence, vectors (see Chap. 4) are just special forms of matrices. Typically, matrices are indicated by capital letters, often in bold (as in this book) and sometimes, when writing manually, by underlined or overlined capital letters. The *order* or size of a matrix is always indicated by first mentioning the number of rows and then the number of columns. Thus, an $m \times n$ matrix has m rows and n columns. For example, $A = \begin{pmatrix} 1 & 6 & -3 \\ 0.5 & 7 & 4 \end{pmatrix}$ is a 2×3 matrix. When the number of rows equals the number of columns, as in the distance example above, the matrix is called *square*. An element of a matrix $\mathbf{A}$ in its ith row and j^{th} column can be indicated by a_{ij}, $a_{i,j}$ or $a(i,j)$. Thus, in the example matrix $\mathbf{A}$ above $a_{1,2} = 6$ and $a_{2,1} = 0.5$. This is very similar to how vector elements are indicated, except that vector elements have one index instead of two.

5.2   Matrix Operations

All common mathematical operations, such as addition, subtraction and multiplication, can be executed on matrices, very similar to how they are executed on vectors (see Sect. 4.2.1).

5.2.1 Matrix Addition, Subtraction and Scalar Multiplication

Since the concepts of vector addition, subtraction and scalar multiplication should by now be familiar to you, explaining the same operations for matrices becomes quite easy. The geometrical definition that could be used for vectors is not available for matrices, so we here limit the definition of these operations to the algebraic one. The notation for matrix elements that was just introduced in the previous section helps to define these basic operations on matrices. For example, addition of matrices $\mathbf{A}$ and $\mathbf{B}$ (of the same size) is defined as:

$$(A + B)_{ij} = a_{ij} + b_{ij}$$

Thus, the element at position (i,j) in the sum matrix is obtained by adding the elements at the same position in the matrices $\mathbf{A}$ and $\mathbf{B}$. Subtraction of matrices $\mathbf{A}$ and $\mathbf{B}$ is defined similarly as:

$$(A - B)_{ij} = a_{ij} - b_{ij},$$

and multiplication of a matrix $\mathbf{A}$ by a scalar s is defined as:

$$(sA)_{ij} = sa_{ij}$$

Exercises

5.1. In the figure below three cities A, B and C are indicated with two roads between them.

There is no direct connection between cities A and C. The distance matrix between the three cities is:

$$\begin{pmatrix} 0 & 12 & 21 \\ 12 & 0 & 9 \\ 21 & 9 & 0 \end{pmatrix}$$

(a) Copy the figure and add the distances to the roads.
(b) A direct (straight) road between cities A and C is built: how does the distance matrix change?
(c) There are also three public parks D, E and F in the area. The (shortest) distance matrix between the cities (rows) and the parks (columns) is given by:

$$\begin{pmatrix} 15 & 24 & 0 \\ 9 & 12 & 12 \\ 0 & 15 & 15 \end{pmatrix}$$

Indicate the location of the parks in the figure, assuming that the new road between A and C has been built.

5.2. Add and subtract (i.e. first-second matrix) the following pairs of matrices:

(a) $\begin{pmatrix} 3 & 4 \\ -1 & 8 \end{pmatrix}$ and $\begin{pmatrix} 2 & -2 \\ 3 & 7 \end{pmatrix}$

(b) $\begin{pmatrix} 3 & -7 & 4 \\ -2 & 6 & 5 \\ 1 & -2 & -9 \end{pmatrix}$ and $\begin{pmatrix} 4 & 3 & 2 \\ 1 & -2 & -4 \\ -5 & 8 & 11 \end{pmatrix}$

(c) $\begin{pmatrix} 1.2 & 3.2 & -1.5 \\ 3.4 & 2.3 & -3.2 \end{pmatrix}$ and $\begin{pmatrix} 0.8 & -1.6 & 0.5 \\ 1.7 & -1.3 & 1.2 \end{pmatrix}$

5.3. Calculate

(a) $\begin{pmatrix} 1 & 2 \\ 7 & 1 \\ 1 & 4 \end{pmatrix} + 3\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} - 2\begin{pmatrix} -1 & 1 \\ 0 & 4 \\ -2 & -3 \end{pmatrix}$

(b) $-\begin{pmatrix} 0.5 & 3.1 \\ 6.7 & 2.4 \end{pmatrix} - 2\begin{pmatrix} 1 & 0.7 \\ 1.2 & 0.7 \end{pmatrix} + 1.5\begin{pmatrix} 4 & 3 \\ 8 & 3 \end{pmatrix}$

5.2.2 Matrix Multiplication and Matrices as Transformations

Matrix addition, subtraction and scalar multiplication are quite straightforward generalizations of the same operations on vectors. For multiplication, the story is a little different, although there is a close relation between matrix multiplication and the vector inner product. A matrix product is only well defined if the number of columns in the first matrix equals the number of rows in the second matrix. Suppose $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{B}$ is an $n \times p$ matrix, then the product $\mathbf{AB}$ is defined by:

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

If you look closer at this definition, multiplying two matrices turns out to entail repeated vector inner product calculations, e.g. the element in the third row and second column of the product is the inner product of the third row of the first matrix with the second column of the second matrix. Let's work this out for one abstract and for one concrete example:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix}$$

$$\begin{pmatrix} 3 & 4 & 2 \\ -2 & -1 & 0 \\ 2 & -3 & 7 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix} = \begin{pmatrix} 3 \cdot 1 + 4 \cdot 3 + 2 \cdot 5 & 3 \cdot 2 + 4 \cdot 4 + 2 \cdot 6 \\ -2 \cdot 1 - 1 \cdot 3 + 0 \cdot 5 & -2 \cdot 2 - 1 \cdot 4 + 0 \cdot 6 \\ 2 \cdot 1 - 3 \cdot 3 + 7 \cdot 5 & 2 \cdot 2 - 3 \cdot 4 + 7 \cdot 6 \end{pmatrix} = \begin{pmatrix} 25 & 34 \\ -5 & -8 \\ 28 & 34 \end{pmatrix}$$

Exercises

5.4. Which pairs of the following matrices could theoretically be multiplied with each other and in which order? Write down all possible products.
A with order 2×3
B with order 3×4
C with order 3×3
D with order 4×2

5.5. What is the order (size) of the resulting matrix when multiplying two matrices of orders

(a) 2×3 and 3×7
(b) 2×2 and 2×1
(c) 1×9 and 9×1

5.6. Calculate whichever product is possible: **AB**, **BA**, none or both.

(a) $\mathbf{A} = \begin{pmatrix} 3 & 4 \\ -1 & 8 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 2 & -2 \\ 3 & 7 \end{pmatrix}$

(b) $\mathbf{A} = \begin{pmatrix} 3 & -7 \\ -2 & 6 \\ 1 & -2 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 4 & 3 & 2 \\ 1 & -2 & -4 \\ -5 & 8 & 11 \end{pmatrix}$

(c) $\mathbf{A} = \begin{pmatrix} 1.2 & 3.2 & -1.5 \\ 3.4 & 2.3 & -3.2 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 0.8 & -1.6 & 0.5 \\ 1.7 & -1.3 & 1.2 \end{pmatrix}$
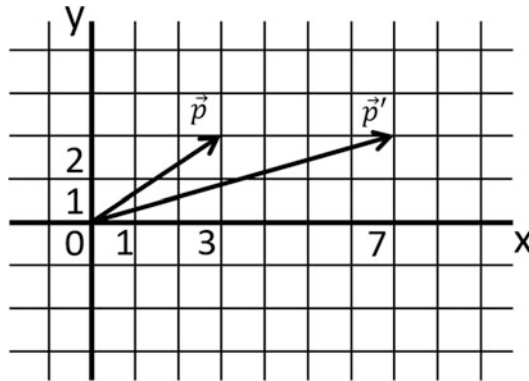
Fig. 5.2 Illustration of how vector $\vec{p}$ is transformed into vector $\vec{p}\,'$ by matrix M (see main text).

So far, we discussed matrix operations as operations on abstract entities. However, matrices are actually commonly used to represent (linear) *transformations*. What this entails can be easily demonstrated by showing the effect of the application of matrices on vectors. Application here means that the matrix and vector are multiplied. Let's first illustrate this by an example. Consider a vector $\vec{p} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}$, a matrix $M = \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix}$ and its product $\vec{p}\,' = M\vec{p} = \begin{pmatrix} 1 & 2 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 7 \\ 2 \end{pmatrix}$ (see Fig. 5.2).

First, observe that when we calculate $M\vec{p}$ another 2×1 vector results, since we multiply a 2×2 matrix and a 2×1 vector. Apparently, judging from Fig. 5.2, applying the matrix **M** to the vector $\vec{p}$ transforms the vector; it rotates it and changes its length. To understand this transformation, let's see how this matrix transforms the two basis vectors of 2D space; $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. These two vectors are called *basis vectors*, because any vector $\begin{pmatrix} a \\ b \end{pmatrix}$ in 2D space can be built from them by a linear combination as follows: $\begin{pmatrix} a \\ b \end{pmatrix} = a\begin{pmatrix} 1 \\ 0 \end{pmatrix} + b\begin{pmatrix} 0 \\ 1 \end{pmatrix}$. The matrix **M** transforms the first basis vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ to $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and the second basis vector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ to $\begin{pmatrix} 2 \\ -2 \end{pmatrix}$, i.e. the two columns of the matrix **M** show how the transform affects the basis vectors. The effect of **M** on any vector $\begin{pmatrix} a \\ b \end{pmatrix}$ is thus equal to $M\begin{pmatrix} a \\ b \end{pmatrix} = a\begin{pmatrix} 1 \\ 2 \end{pmatrix} + b\begin{pmatrix} 2 \\ -2 \end{pmatrix}$. For the example in Fig. 5.2 this results indeed in $M\begin{pmatrix} 3 \\ 2 \end{pmatrix} = 3\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 2\begin{pmatrix} 2 \\ -2 \end{pmatrix} = \begin{pmatrix} 7 \\ 2 \end{pmatrix}$.

There are some special geometric transformation matrices, both in 2D, as well as 3D. One that is often encountered is the rotation matrix, that leaves the length of a vector unaltered, but rotates it around the origin in 2D or 3D. The rotation matrix has applications in e.g., the study of rigid body movements and in the manipulation of images, as in the preprocessing of

functional magnetic resonance imaging (fMRI) scans. The transformation matrix that rotates a vector around the origin (in 2D) over an angle θ (counter clockwise) is given by $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, as illustrated in Fig. 5.3 (cf. Chap. 3 on the geometrical definition of sine and cosine).

Another common geometric transformation matrix is the *shearing* matrix, that transforms a square in 2D into a parallelogram (see Fig. 5.4). Applying the matrix transformation $\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ results in shearing along the *x*-axis (*y*-coordinate remains unchanged), whereas applying the matrix transformation $\begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}$ results in shearing along the *y*-axis (*x*-coordinate remains unchanged). This transformation matrix is also applied in e.g., the preprocessing of fMRI scans.
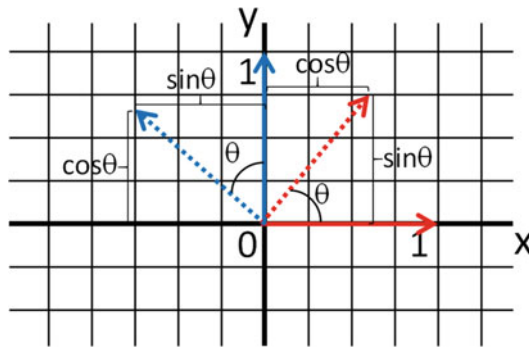

Fig. 5.3 Illustration of how the 2D basis vectors, $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ in *red* and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ in *blue*, are affected by a counter-clockwise rotation over an angle θ. The vectors before rotation are indicated by *drawn arrows*, the vectors after rotation are indicated by *dotted arrows*.
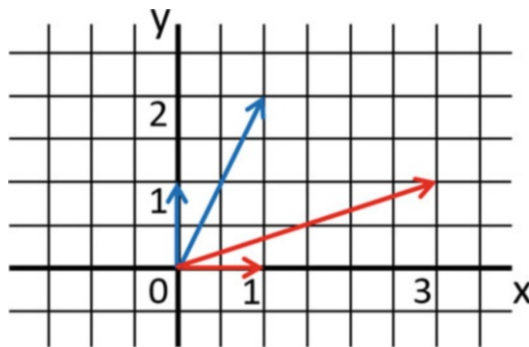

Fig. 5.4 Illustration of how the two 2D basis vectors are affected by shearing along the *x*- and *y*-axes. The applied transformation matrix for shearing along the *x*-axis is $\begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}$ (results in *red*) and the applied transformation matrix for shearing along the *y*-axis is $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ (results in *blue*).

5.2.3 Alternative Matrix Multiplication

Just as there are different ways of multiplying vectors, there are also different ways of multiplying matrices, that are less common, however. For example, matrices can be multiplied element-wise; this product is also referred to as the *Hadamard product*, Schur product or pointwise product:

$$(A \circ B)_{ij} = a_{ij} b_{ij}$$

This only works when the two matrices have the same size. Another matrix product is the *Kronecker product*, which is a generalization of the tensor product (or dyadic product) for vectors. For vectors this product is equal to an $n \times 1$ column vector being multiplied by a $1 \times n$ row vector, which results in an $n \times n$ matrix when following the standard multiplication rules for matrices:

$$\vec{a} \otimes \vec{b} = \vec{a}\vec{b}^{T} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \begin{pmatrix} b_1 & b_2 & \cdots & b_n \end{pmatrix} = \begin{pmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n b_1 & a_n b_2 & \cdots & a_n b_n \end{pmatrix}$$

For two matrices **A** $(m \times n)$ and **B** $(p \times q)$, the Kronecker product **A**⊗**B** is defined by:

$$A \otimes B = \begin{pmatrix} a_{11} B & a_{12} B & \cdots & a_{1n} B \\ a_{21} B & a_{22} B & \cdots & a_{2n} B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} B & a_{m2} B & \cdots & a_{mn} B \end{pmatrix}$$

Note that the size of the matrices **A** and **B** does not need to match for the Kronecker product. The Kronecker product has proven its use in the study of matrix theory (linear algebra), e.g. in solving matrix equations such as the Sylvester equation $\mathbf{AX} + \mathbf{XB} = \mathbf{C}$ for general **A**, **B** and **C** (see e.g. Horn and Johnson 1994).

Exercises

5.7. Calculate the Hadamard or pointwise product of

(a) $\begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}$ and $\begin{pmatrix} 2 & 3 \\ 4 & 1 \end{pmatrix}$

(b) $\begin{pmatrix} 2 & -0.3 & 1 \\ 1.5 & 7 & -0.4 \end{pmatrix}$ and $\begin{pmatrix} 1.4 & 9 & 0.5 \\ 8 & -0.1 & 10 \end{pmatrix}$

5.8. Calculate the Kronecker product of

(a) $\begin{pmatrix} 1 & 2 \\ -1 & 1 \end{pmatrix}$ and $\begin{pmatrix} 3 & -1 & 4 \end{pmatrix}$

(b) $\begin{pmatrix} -2 & -3 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{pmatrix}$

5.2.4 Special Matrices and Other Basic Matrix Operations

There are some special forms of matrices and other basic operations on matrices that should be known before we can explain more interesting examples and applications of matrices.

The *identity* or *unit matrix* of size n is a square matrix of size $n \times n$ with ones on the diagonal and zeroes elsewhere. It is often referred to as $\mathbf{I}$ or, if necessary to explicitly mention its size, as $\mathbf{I_n}$. The identity matrix is a special *diagonal matrix*. More generally, a diagonal matrix only has non-zero elements on the diagonal and zeroes elsewhere. If a diagonal matrix is extended with non-zero elements only above or only below the diagonal, we speak about an upper or lower *triangular matrix*, respectively. A matrix that is symmetric around the diagonal, i.e. for which $a_{ij} = a_{ji}$, is called *symmetric*. An example of a symmetric matrix is a distance matrix, such as encountered in the very first example in this chapter. A matrix is *skew-symmetric* if $a_{ij} = -a_{ji}$. A matrix is called *sparse* if most of its elements are zero. In contrast, a matrix whose elements are almost all non-zero is called *dense*. A *logical* matrix (or binary or Boolean) matrix only contains zeroes and ones.

One operation that was already encountered in the previous chapter on vectors is the *transpose*. The transpose of an n $\times$ m matrix $\mathbf{A}$ is an m $\times$ n matrix of which the elements are defined as:

$$\left(A^T\right)_{ij} = a_{ji}$$

A generalization of the transpose to complex-valued elements is the *conjugate transpose*, for which the elements are defined by:

$$\left(A^*\right)_{ij} = \bar{a}_{ji},$$

where the overbar indicates the complex conjugate (see Sect. 1.2.4.1). For real matrices, the transpose is equal to the conjugate transpose. In quantum mechanics the conjugate transpose is indicated by † (dagger) instead of *. Notice that $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$ and $(\mathbf{AB})^* = \mathbf{B}^*\mathbf{A}^*$ for matrices $\mathbf{A}$ and $\mathbf{B}$ for which the matrix product $\mathbf{AB}$ is possible.

Now that we have introduced the conjugate transpose, we can also introduce a few more special (complex) matrices. For a *Hermitian* matrix: $\mathbf{A} = \mathbf{A}^*$, for a *normal* matrix: $\mathbf{A}^*\mathbf{A} = \mathbf{AA}^*$ and for a *unitary* matrix: $\mathbf{AA}^* = \mathbf{I}$. Note that all unitary and Hermitian matrices are normal,

but the reverse is not true. The unitary matrices will return when we explain singular value decomposition in Sect. 5.3.3.

All special matrices that were introduced in this section are summarized in Table 5.1 with their abstract definition and a concrete example.

Table 5.1 Special matrices with their abstract definition and an example

	Definition	Example
Identity/unit	$a_{ij} = 0$ if $i \neq j$ $a_{ij} = 1$ if $i = j$	$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$
Diagonal	$a_{ij} = 0$ if $i \neq j$ $a_{ij} \neq 0$ if $i = j$	$\begin{pmatrix} 2 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 7 \end{pmatrix}$
Lower-triangular	$a_{ij} = 0$ if $i < j$ $a_{ij} \neq 0$ if $i \geq j$	$\begin{pmatrix} 1 & 0 & 0 \\ 3 & -2 & 0 \\ -5 & 6 & -4 \end{pmatrix}$
Upper-triangular	$a_{ij} = 0$ if $i > j$ $a_{ij} \neq 0$ if $i \leq j$	$\begin{pmatrix} 1 & 3 & 19 \\ 0 & -2 & -8 \\ 0 & 0 & -4 \end{pmatrix}$
Symmetric	$a_{ij} = a_{ji}$	$\begin{pmatrix} 0 & 12 & 21 \\ 12 & 0 & -9 \\ 21 & -9 & 0 \end{pmatrix}$
Skew-symmetric	$a_{ij} = -a_{ji}$	$\begin{pmatrix} 0 & -12 & -21 \\ 12 & 0 & -9 \\ 21 & 9 & 0 \end{pmatrix}$
Sparse	Most elements are zero	$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & -3 & 0 \\ 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ -1 & 0 & 0 & 0 & 7 \end{pmatrix}$
Dense	Most elements are non-zero	$\begin{pmatrix} 3 & 7 & -5 & 8 & -1 \\ 0 & 2 & 4 & -3 & 0 \\ 6 & 4 & -1 & 34 & 7 \\ 2 & 0 & 6 & -11 & 3 \\ -1 & -3 & 5 & 6 & 7 \end{pmatrix}$
Logical	$a_{ij} \in \{0,1\}$	$\begin{pmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix}$
Hermitian	$\mathbf{A} = \mathbf{A}^*$	$\begin{pmatrix} 3 & 1+i & 8 \\ 1-i & 7 & -i \\ 8 & i & -2 \end{pmatrix}$
Normal	$\mathbf{A}^*\mathbf{A} = \mathbf{A}\mathbf{A}^*$	$\begin{pmatrix} 3 & 1+i & 8 \\ 1-i & 7 & -i \\ 8 & i & -2 \end{pmatrix}$
Unitary	$\mathbf{A}\mathbf{A}^* = \mathbf{I}$	$\frac{1}{2}\begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix}$

Exercises

5.9. In what sense (according to the definitions in Table 5.1) are the following matrices special? Mention at least one property.

(a) $\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

(b) $\begin{pmatrix} 1 & 0 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

(c) $\begin{pmatrix} 0 & 4 & 3 \\ -4 & 0 & -7 \\ -3 & 7 & 0 \end{pmatrix}$

(d) $\begin{pmatrix} 1 & 4 & 3 \\ 0 & 1 & -7 \\ 0 & 0 & 1 \end{pmatrix}$

(e) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 23 & 0 \\ 0 & 0 & -7 \end{pmatrix}$

(f) $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

5.10. Determine the conjugate transpose of

(a) $\begin{pmatrix} 1 & 2 & 3 \\ -i & 1 & -3-2i \\ 5 & 4+5i & 3 \end{pmatrix}$

(b) $\begin{pmatrix} 1 & 2 & 3 \\ -1 & 1 & -3 \\ 5 & 4 & 0 \end{pmatrix}$

(c) $\begin{pmatrix} 4 & 0 & 3-2i \\ 19+i & -3 & -3 \\ -8i & -11-i & 17 \end{pmatrix}$

5.3   More Advanced Matrix Operations and Their Applications

Now that the basics of matrix addition, subtraction, scalar multiplication and multiplication hold no more secrets for you, it is time to get into more interesting mathematical concepts that rest on matrix calculus and to illustrate their applications.

5.3.1 Inverse and Determinant

One of the oldest applications of matrices is to solve systems of linear equations that were introduced to you in Chap. 2. Consider the following system of linear equations with three unknowns x, y and z:

$$3x + 4y - 2z = 5$$
$$-2x - 2y + z = -3$$
$$x + y - 7z = -18$$

This system can also be written as a matrix equation:

$$\begin{pmatrix} 3 & 4 & -2 \\ -2 & -2 & 1 \\ 1 & 1 & -7 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \\ -18 \end{pmatrix} \text{ or as}$$

$$M \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 5 \\ -3 \\ -18 \end{pmatrix} \text{ where } M = \begin{pmatrix} 3 & 4 & -2 \\ -2 & -2 & 1 \\ 1 & 1 & -7 \end{pmatrix}$$

Such a system and more generally, systems of n equations in n unknowns, can be solved by using *determinants*, which is actually similar to using the *inverse* of a matrix to calculate the solution to a system of linear equations. The inverse of a square matrix $\mathbf{A}$ is the matrix $\mathbf{A}^{-1}$ such that $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$. The inverse of a matrix $\mathbf{A}$ does not always exist; if it does $\mathbf{A}$ is called invertable. Notice that $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ for square, invertable matrices $\mathbf{A}$ and $\mathbf{B}$. Now suppose that the matrix $\mathbf{M}$ in the matrix equation above has an inverse $\mathbf{M}^{-1}$. In that case, the solution to the equation would be given by:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = I \begin{pmatrix} x \\ y \\ z \end{pmatrix} = M^{-1}M \begin{pmatrix} x \\ y \\ z \end{pmatrix} = M^{-1} \begin{pmatrix} 5 \\ -3 \\ -18 \end{pmatrix}$$

Mathematically, the inverse of 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is given by $\frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ where $ad - bc$ is the determinant $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$ of the matrix. This illustrates why a square matrix has an inverse if and only if (iff) its determinant is not zero, as division by zero would result in infinity. For a 3×3 matrix or higher-dimensional matrices the inverse can still be calculated by hand, but it quickly becomes cumbersome. In general, for a matrix $\mathbf{A}$:

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

where $adj(\mathbf{A})$ is the *adjoint* of $\mathbf{A}$. The adjoint of a matrix $\mathbf{A}$ is the transpose of the *cofactor matrix*. Each (i,j)-element of a cofactor matrix is given by the determinant of the matrix that remains when the i-th row and j-th column are removed, multiplied by -1 if $i + j$ is odd. In Box 5.1 the inverse is calculated of the matrix $\mathbf{M}$ given in the example that we started this section with.

Box 5.1 Example of calculating the inverse of a matrix

To be able to calculate $\mathbf{M}^{-1}$ where $M = \begin{pmatrix} 3 & 4 & -2 \\ -2 & -2 & 1 \\ 1 & 1 & -7 \end{pmatrix}$ we first need to know how to calculate

the determinant of a 3×3 matrix. This is done by first choosing a reference row or column and calculating the cofactors for that row or column. Then the determinant is equal to the sum of the products of the elements of that row or column with its cofactors. This sounds rather abstract so let's calculate det($\mathbf{M}$) by taking its first row as a reference.

$$\det(M) = \det \begin{pmatrix} 3 & 4 & -2 \\ -2 & -2 & 1 \\ 1 & 1 & -7 \end{pmatrix}$$

$$= 3 \begin{vmatrix} -2 & 1 \\ 1 & -7 \end{vmatrix} - 4 \begin{vmatrix} -2 & 1 \\ 1 & -7 \end{vmatrix} - 2 \begin{vmatrix} -2 & -2 \\ 1 & 1 \end{vmatrix}$$

$$= 3 \cdot 13 - 4 \cdot 13 - 2 \cdot 0 = -13$$

For the matrix $\mathbf{M}$ its cofactor matrix is given by

$$\begin{pmatrix} \begin{vmatrix} -2 & 1 \\ 1 & -7 \end{vmatrix} & -\begin{vmatrix} -2 & 1 \\ 1 & -7 \end{vmatrix} & \begin{vmatrix} -2 & -2 \\ 1 & 1 \end{vmatrix} \\ -\begin{vmatrix} 4 & -2 \\ 1 & -7 \end{vmatrix} & \begin{vmatrix} 3 & -2 \\ 1 & -7 \end{vmatrix} & -\begin{vmatrix} 3 & 4 \\ 1 & 1 \end{vmatrix} \\ \begin{vmatrix} 4 & -2 \\ -2 & 1 \end{vmatrix} & -\begin{vmatrix} 3 & -2 \\ -2 & 1 \end{vmatrix} & \begin{vmatrix} 3 & 4 \\ -2 & -2 \end{vmatrix} \end{pmatrix} = \begin{pmatrix} 13 & -13 & 0 \\ 26 & -19 & 1 \\ 0 & 1 & 2 \end{pmatrix}.$$

Hence, the adjoint matrix of $\mathbf{M}$ is its transpose $\begin{pmatrix} 13 & 26 & 0 \\ -13 & -19 & 1 \\ 0 & 1 & 2 \end{pmatrix}$. Thus, the inverse of $\mathbf{M}$ is given by:

$$M^{-1} = \frac{1}{-13} \begin{pmatrix} 13 & 26 & 0 \\ -13 & -19 & 1 \\ 0 & 1 & 2 \end{pmatrix}$$

To verify that the inverse that we calculated in Box 5.1 is correct, it suffices to verify that the matrix multiplied by its inverse equals the identity matrix. Thus, for matrix $\mathbf{M}$ we verify that $\mathbf{M}^{-1}\mathbf{M} = \mathbf{I}$:

$$M^{-1}M = \frac{1}{-13} \begin{pmatrix} 13 & 26 & 0 \\ -13 & -19 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 3 & 4 & -2 \\ -2 & -2 & 1 \\ 1 & 1 & -7 \end{pmatrix}$$

$$= -\frac{1}{13} \begin{pmatrix} -13 & 0 & 0 \\ 0 & -13 & 0 \\ 0 & 0 & -13 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = I$$

Finally, now that we have found $\mathbf{M}^{-1}$, we can find the solution to the system of equations that we started with:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = M^{-1} \begin{pmatrix} 5 \\ -3 \\ -18 \end{pmatrix} = -\frac{1}{13} \begin{pmatrix} 13 & 26 & 0 \\ -13 & -19 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ -3 \\ -18 \end{pmatrix} = -\frac{1}{13} \begin{pmatrix} -13 \\ -26 \\ -39 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

Finally, this solution can again be verified by inserting the solution in the system of equations that was given at the beginning of this section.

Cramer's rule is an explicit formulation of using determinants to solve systems of linear equations. We first formulate it for a system of three linear equations in three unknowns

$$a_1x + b_1y + c_1z = d_1$$
$$a_2x + b_2y + c_2z = d_2$$
$$a_3x + b_3y + c_3z = d_3$$

Its associated determinant is:

$$D = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}.$$

We can also define the determinant

$$D_x = \begin{vmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{vmatrix},$$

which is the determinant of the system's associated matrix with its first column replaced by the vector of constants and similarly

$$D_y = \begin{vmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{vmatrix} \text{ and } D_z = \begin{vmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{vmatrix}.$$

Then, x, y and z can be calculated as:

$$x = \frac{D_x}{D}, y = \frac{D_y}{D}, z = \frac{D_z}{D}.$$

Similarly, the solution of a system of n linear equations in n unknowns:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \text{ or } Ax = b,$$

according to Cramer's rule is given by:

$x_1 = \frac{D_{x_1}}{D}, x_2 = \frac{D_{x_2}}{D}, \quad \ldots \quad , x_n = \frac{D_{x_n}}{D}$, where D_{x_i} is the determinant of the matrix formed

by replacing the i-th column of $\mathbf{A}$ by the column vector $\vec{b}$. Note that Cramer's rule only applies when $D \neq 0$. Unfortunately, Cramer's rule is computationally very inefficient for larger systems and thus not often used in practice.

Exercises

5.11. Show that using Cramer's rule to find the solution to a system of two linear equations in two unknowns $ax + by = c$ and $dx + ey = f$ is the same as applying the inverse of the matrix $\begin{pmatrix} a & b \\ d & e \end{pmatrix}$ to the vector $\begin{pmatrix} c \\ f \end{pmatrix}$.

5.12. Find the solution to the following systems of linear equations using Cramer's rule:

(a) (Example 2.6) $3x+5=5y \wedge 2x-5y=6$
(b) $4x-2y-2z=10 \wedge 2x+8y+4z=32 \wedge 30x+12y-4z=24$

5.13. Find the solution to the following systems of linear equations using the matrix inverse:

(a) (Exercise 2.4a) $x - 2y = 4 \wedge \frac{x}{3} - y = \frac{4}{3}$
(b) (Example 2.7) $2x+y+z=4 \wedge x-7-2y=-3z \wedge 2y+10-2z=3x$

Now that you know how to calculate the determinant of a matrix, it is easy to recognize that the algebraic definition of the cross-product introduced in Sect. 4.2.2.2 is similar to calculating the determinant of a very special matrix:

$$\vec{a} \times \vec{b} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \times \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} = \begin{pmatrix} a_2b_3 - a_3b_2 \\ a_3b_1 - a_1b_3 \\ a_1b_2 - a_2b_1 \end{pmatrix} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

$$= \begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix} \vec{i} - \begin{vmatrix} a_1 & a_3 \\ b_1 & b_3 \end{vmatrix} \vec{j} + \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \vec{k}$$

Finding the inverse of larger square matrices and thus finding the solution to larger systems of linear equations may also be accomplished by calculating the inverse matrix by hand. However, as you will surely have noticed when doing the exercises, finding the solution of a system of three linear equations with three unknowns calculating the inverse matrix by hand is already rather cumbersome, tedious and error-prone. Computers do a much better job than we at this sort of task which is why (numerical) mathematicians have developed clever, fast computer algorithms to determine matrix inverses. There is even a whole branch of numerical mathematics that focuses on solving systems of equations that can be represented by sparse matrices, as fast as possible (Saad 2003). The relevance of sparse matrices is that they naturally appear in many scientific or engineering applications whenever *partial differential equations* (see Box 5.2 and Chap. 6) are numerically solved on a grid. Typically, only local

physical interactions play a role in such models of reality and thus only neighboring grid cells interact, resulting in sparse matrices. Examples are when heat dissipation around an underground gas tube or turbulence in the wake of an airfoil needs to be simulated. In Box 5.2 a simple example is worked out to explain how a *discretized* partial differential equation can result in a sparse matrix.

Box 5.2 How discretizing a partial differential equation can yield a sparse matrix

We here consider the *Laplace equation* for a *scalar function u* in two dimensions (indicated by *x* and *y*), which is e.g. encountered in the fields of electromagnetism and fluid dynamics:

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$$

It can describe airflow around airfoils or water waves, under specific conditions. In Sect. 6.11 partial derivatives are explained in more detail. To solve this equation on a rectangle, a grid or lattice of evenly spaced points (with distance *h*) can be overlaid as in Fig. 5.5.

One method to discretize the Laplace equation on this grid (see for details e.g. Press et al. n.d.) is:

$$\frac{u_{i-1,j} - 2u_{i,j} + u_{i+1,j}}{h^2} + \frac{u_{i,j-1} - 2u_{i,j} + u_{i,j+1}}{h^2} = 0$$

or

$$u_{i,j} = \frac{1}{4}\left(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}\right) \tag{5.1}$$

Here, *i* runs over the grid cells in *x*-direction and *j* runs over the grid cells in *y*-direction. This discretized equation already shows that the solution in a point (*i,j*) (in red in the figure) is only determined by the values in its local neighbours, i.e. the four points (*i* + 1,*j*), (*i* − 1,*j*), (*i, j* + 1) and (*i, j* − 1) (in blue in the figure) directly surrounding it. There are alternative discretizations that use variable spacings in the two grid directions and/or fewer or more points; the choice is determined by several problem parameters such as the given *boundary conditions*.

Here, I only want to illustrate how solving this problem involves a sparse matrix. Typically, iterative solution methods are used, meaning that the solution is approximated step-by-step starting from an initial solution and adapting the solution at every iteration until the solution hardly changes anymore. The solution at iteration step *n* + *1* (indicated by a superscript) can then be derived from the solution at time step *n* e.g. as follows (cf. Eq. 5.1):

$$u_{i,j}^{n+1} = \frac{1}{4}\left(u_{i-1,j}^n + u_{i+1,j}^n + u_{i,j-1}^n + u_{i,j+1}^n\right)$$

Here, we can recognize a (sparse) matrix transformation as I will explain now. Suppose we are dealing with the 11 × 7 grid in the figure. We can then arrange the values of *u* at iteration step *n* in a vector e.g. by row: $\vec{u}^n = \left(u_{1,1}^n \quad \cdots \quad u_{1,11}^n \quad u_{2,1}^n \quad \cdots \quad u_{2,11}^n \quad u_{3,1}^n \quad \cdots \quad u_{7,11}^n \right)^T$ and similarly for the values of *u* at iteration step *n* + *1*. The 77 × 77 matrix transforming $\vec{u}^n$ into $\vec{u}^{n+1}$ then has only four non-zero entries and 73 zeroes in every row (with some exceptions for the treatment of the boundary grid points) and is thus sparse. It should be noted that this particular iterative solution method is slow and that much faster methods are available.
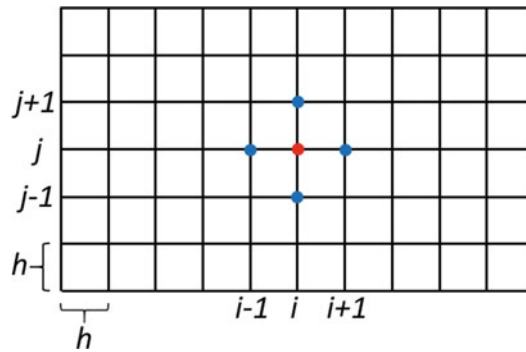
Fig. 5.5 Grid with equal spacing *h* for numerical solution of the Laplace equation in two dimensions. To solve for the point (*i,j*) (in *red*), values in the surrounding points (in *blue*) can be used.

5.3.2 Eigenvectors and Eigenvalues

A first mathematical concept that rests on matrix calculus and that is part of other interesting matrix decomposition methods (see Sect. 5.3.3) is that of *eigenvectors* and *eigenvalues*. An eigenvector $\vec{v}$ of a square matrix **M** is a non-zero column vector that does not change its orientation (although it may change its length by a factor λ) as a result of the transformation represented by the matrix. In mathematical terms:

$$M\vec{v} = \lambda\vec{v},$$

where λ is a scalar known as the eigenvalue. Eigenvectors and eigenvalues have many applications, of which we will encounter a few in the next section. For example, they allow to determine the principal axes of rotational movements of rigid bodies (dating back to eighteenth century Euler), to find common features in images as well as statistical data reduction.

To find the eigenvectors of a matrix **M**, the following system of linear equations has to be solved:

$$(M - \lambda I)\vec{v} = 0.$$

It is known that this equation has a solution iff the determinant of $M-\lambda I$ is zero. Thus, the eigenvalues of **M** are those values of λ that satisfy $|M-\lambda I|=0$. This determinant is a polynomial in λ of which the highest power is the order of the matrix **M**. The eigenvalues can be found by finding the *roots* of this polynomial, known as the characteristic polynomial of **M**. There are as many roots as the order (highest power) of the polynomial. Once the eigenvalues are known, the related eigenvectors can be found by solving for $\vec{v}$ in $M\vec{v} = \lambda\vec{v}$. This probably sounds quite abstract, so a concrete example is given in Box 5.3.

Box 5.3 Example of calculating the eigenvalues and eigenvectors of a matrix

To find the eigenvalues of the matrix $\begin{pmatrix} 3 & 1 \\ 2 & 4 \end{pmatrix}$, we first determine its characteristic equation as the determinant

$$\left| \begin{pmatrix} 3 & 1 \\ 2 & 4 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = \left| \begin{pmatrix} 3-\lambda & 1 \\ 2 & 4-\lambda \end{pmatrix} \right| = (3-\lambda)(4-\lambda) - 2 = \lambda^2 - 7\lambda + 10$$

Since $\lambda^2 - 7\lambda + 10 = (\lambda-2)(\lambda-5)$, the roots of this polynomial are given by $\lambda=2$ and $\lambda=5$. The eigenvector for $\lambda=2$ follows from:

$$\begin{pmatrix} 3 & 1 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 2 \begin{pmatrix} x \\ y \end{pmatrix} \Rightarrow \begin{pmatrix} 3x+y \\ 2x+4y \end{pmatrix} = \begin{pmatrix} 2x \\ 2y \end{pmatrix}$$
$$\Rightarrow 3x+y = 2x \wedge 2x+4y = 2y$$
$$\Rightarrow x+y = 0 \wedge 2x+2y = 0$$
$$\Rightarrow x = -y$$
$$\Rightarrow \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Notice that $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ is the eigenvector in this case, because for any multiple of this vector it is true that $x = -y$.

Similarly, the eigenvector for $\lambda=5$ follows from:

$$\begin{pmatrix} 3 & 1 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 5 \begin{pmatrix} x \\ y \end{pmatrix} \Rightarrow \begin{pmatrix} 3x+y \\ 2x+4y \end{pmatrix} = \begin{pmatrix} 5x \\ 5y \end{pmatrix}$$
$$\Rightarrow 3x+y = 5x \wedge 2x+4y = 5y$$
$$\Rightarrow -2x+y = 0 \wedge 2x-y = 0$$
$$\Rightarrow x = \frac{1}{2}y$$
$$\Rightarrow \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Notice that $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ is the eigenvector in this case, because for any multiple of this vector it is true that $2x = y$.

Hence, the matrix $\begin{pmatrix} 3 & 1 \\ 2 & 4 \end{pmatrix}$ has eigenvectors $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ with eigenvalues 2 and 5, respectively.

Exercises

5.14. Consider the shearing matrix $\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$. Without calculation, which are the eigenvectors of this matrix and why?

5.15. Calculate the eigenvectors and eigenvalues of

(a) $\begin{pmatrix} 7 & 0 & 0 \\ 0 & -19 & 0 \\ 0 & 0 & 2 \end{pmatrix}$

(b) $\begin{pmatrix} 2 & 1 \\ -1 & 4 \end{pmatrix}$

(c) $\begin{pmatrix} 2 & 2 & -1 \\ 1 & 3 & -1 \\ 1 & 4 & -2 \end{pmatrix}$

(d) $\begin{pmatrix} 5 & 1 \\ 4 & 5 \end{pmatrix}$

5.3.3 Diagonalization, Singular Value Decomposition, Principal Component Analysis and Independent Component Analysis

In the previous section I already alluded to several applications using eigenvalues and eigenvectors. In this section I will discuss some of the most well-known and most often used methods that employ *eigendecomposition* and that underlie these applications.

The simplest is *diagonalization*. A matrix **M** can be diagonalized if it can be written as

$$M = VDV^{-1} \tag{5.2}$$

where **V** is an invertible matrix and **D** is a diagonal matrix. First note that **V** has to be square to be invertible, so that **M** also has to be square to be diagonalizable. Now let's find out how eigenvalues and eigenvectors play a role in Eq. 5.2. We can rewrite Eq. 5.2 by right multiplication with **V** to:

$$MV = VD$$

or similarly, when we indicate the columns of **V** by $\vec{v}_i$ and the elements of D by d_i, $i = 1, \ldots, n$, then:

$$M\begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{pmatrix} = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{pmatrix}D = \begin{pmatrix} d_1\,\vec{v}_1 & d_2\,\vec{v}_2 & \cdots & d_n\,\vec{v}_n \end{pmatrix}$$

This shows that in Eq. 5.2 the columns of **V** must be the eigenvectors of **M** and the diagonal elements of **D** must be the eigenvalues of **M**. We just noted that only square matrices can potentially be diagonalized. What *singular value decomposition* does is generalize the concept of diagonalization to any matrix, making it a very powerful method.

For an intuitive understanding of SVD you can think of a matrix as a large table of data, e.g. describing which books from a collection of 500 a 1000 people have read. In that case the matrix could contains ones for books (in 500 columns) that a particular reader (in 1000 rows) read and zeroes for the books that weren't read. SVD can help summarize

the data in this large matrix. When SVD is applied to this particular matrix it will help identify specific sets of books that are often read by the same readers. For example, SVD may find that all thrillers in the collection are often read together, or all science fiction novels, or all romantic books; the composition of these subcollections will be expressed in the *singular vectors*, and their importance in the *singular values*. Now each reader's reading behavior can be expressed in a much more compact way than by just tabulating all books that he or she read. Instead, SVD allows to say e.g., that a reader is mostly interested in science fiction novels, or maybe that in addition, there is some interest in popular science books. Let's see how this compression of data can be achieved by SVD by first looking at its mathematics.

In *singular value decomposition* (SVD) an $m \times n$ rectangular matrix **M** is decomposed into a product of three matrices as follows:

$$M = U\Sigma V^*$$

where **U** is a unitary (see Table 5.1) $m \times m$ matrix, **Σ** an $m \times n$ diagonal matrix with non-negative real entries and **V** another unitary $n \times n$ matrix. To determine these matrices you have to calculate the sets of *orthonormal* eigenvectors of $\mathbf{MM}^*$ and $\mathbf{M}^*\mathbf{M}$. This can be done, as $\mathbf{MM}^*$ and $\mathbf{M}^*\mathbf{M}$ are square. The orthonormal eigenvectors of the former are the columns of **U** and the orthonormal eigenvectors of the latter are the columns of **V**. The square roots of the non-zero eigenvalues of $\mathbf{MM}^*$ or $\mathbf{M}^*\mathbf{M}$ (which do not differ) are the so-called *singular values* of M and form the diagonal elements of **Σ** in decreasing order, completed by zeroes if necessary. The columns of **U** are referred to as the *left-singular vectors* of **M** and the columns of **V** are referred to as the *right-singular vectors* of **M**. For an example which can also be intuitively understood, see Box 5.4.

Box 5.4 Example of SVD of a real square matrix and its intuitive understanding

To determine the SVD of the matrix $M = \begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix}$, we first determine the eigenvalues and eigenvectors of the matrix $\mathbf{M}^*\mathbf{M}$ (which is equal to $\mathbf{M}^\mathsf{T}\mathbf{M}$ in this real case) to get the singular values and right-singular vectors of **M**. Thus, we determine its characteristic equation as the determinant

$$\left| \begin{pmatrix} 1 & 2 \\ -2 & -1 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = \left| \begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| =$$

$$\left| \begin{pmatrix} 5 - \lambda & -4 \\ -4 & 5 - \lambda \end{pmatrix} \right| = (5 - \lambda)^2 - 16 = \lambda^2 - 10\lambda + 9 = (\lambda - 1)(\lambda - 9)$$

The singular values (in decreasing order) then are $\sigma_1 = \sqrt{9} = 3$, $\sigma_2 = \sqrt{1} = 1$ and $\Sigma = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$.

The eigenvector belonging to the first eigenvalue of $\mathbf{M}^\mathsf{T}\mathbf{M}$ follows from:

Box 5.4 (continued)

$$\begin{pmatrix} 5 & -4 \\ -4 & 5 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 9\begin{pmatrix} x \\ y \end{pmatrix} \Rightarrow \begin{pmatrix} 5x - 4y \\ -4x + 5y \end{pmatrix} = \begin{pmatrix} 9x \\ 9y \end{pmatrix}$$

$$\Rightarrow 5x - 4y = 9x \wedge -4x + 5y = 9y$$

$$\Rightarrow -4x - 4y = 0 \wedge -4x - 4y = 0$$

$$\Rightarrow x = -y$$

$$\Rightarrow \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

To determine the first column of **V**, this eigenvector must be normalized (divided by its length; see Sect. 4.2.2.1): $\vec{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Similarly, the eigenvector belonging to the second eigenvalue of $\mathbf{M^T M}$ can be derived to be $\vec{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, making $V = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$. In this case, $\vec{v}_1$ and $\vec{v}_2$ are already orthogonal (see Sect. 4.2.2.1, making further adaptations to arrive at an orthonormal set of eigenvectors unnecessary. In case orthogonalization is necessary, Gram-Schmidt orthogonalization could be used (see Sect. 4.3.2).

In practice, to now determine U for this real 2 × 2 matrix **M**, it is most convenient to use that when $M = U\Sigma V^*(=U\Sigma V^T)$, $MV = U\Sigma$ or $MV\Sigma^{-1} = U$ (using that $V^*V = V^T V = I$ and Σ is real). Thus, the first column of U is equal to $\vec{u}_1 = \frac{1}{\sigma_1}M\vec{v}_1 = \frac{1}{3}\begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and the second column of U is equal to $\vec{u}_2 = \frac{1}{\sigma_2}M\vec{v}_2 = \frac{1}{1}\begin{pmatrix} 1 & -2 \\ 2 & -1 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$, making $U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. One can now verify that indeed $M = U\Sigma V^T$.

As promised, I will now discuss how SVD can be understood by virtue of this specific example. So let's see what V^T, Σ and U (in this order) do to some vectors on the unit circle: $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ in blue, $\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ in yellow , $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ in red and $\frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \end{pmatrix}$ in green, as illustrated in Fig. 5.6.

V^T rotates these vectors over 45°. Σ subsequently scales the resulting vectors by factors of 3 in the x-direction and 1 in the y-direction, after which U performs another rotation over 45°. You can verify in Fig. 5.6 that the result of these three matrices applied successively is exactly the same as the result of applying **M** directly. Thus, given that **U** and **V** are rotation matrices and Σ is a scaling matrix, the intuitive understanding of SVD for real square matrices is that any such matrix transformation can be expressed as a rotation followed by a scaling followed by another rotation. It should be noted that for this intuitive understanding of SVD rotation should be understood to include *improper rotation*. Improper rotation combines proper rotation with reflection (an example is given in Exercise 5.16(a)).

Exercises

5.16. Calculate the SVD of

(a) $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

(b) $\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$

5.17. For the SVD of **M** such that $M=U\Sigma V^*$ prove that the columns of **U** can be determined from the eigenvectors of **MM***.

Now that the mathematics of SVD and its intuitive meaning have been explained it is much easier to explain how SVD can be used to achieve data compression. When you've calculated the SVD of a matrix, representing e.g., the book reading behavior mentioned before, or the pixel values of a black-and-white image, you can compress the information by maintaining only the largest L singular values in Σ, setting all other singular values to zero, resulting in a sparser matrix Σ'. When you then calculate $M'=U\Sigma' V^*$, you'll find that only the first L columns of **U** and **V** remain relevant, as all other columns will be multiplied by zero. So instead of keeping the entire matrix **M** in memory, you only need to store a limited number of columns of **U** and **V**. The fraction of information that is maintained in this manner is determined by the fraction of the sum of the eigenvalues that are maintained divided by the sum of all eigenvalues. In the reference list at the end of this chapter you'll find links to an online SVD calculator and examples of data compression using SVD, allowing you to explore this application of complex matrix operations further.
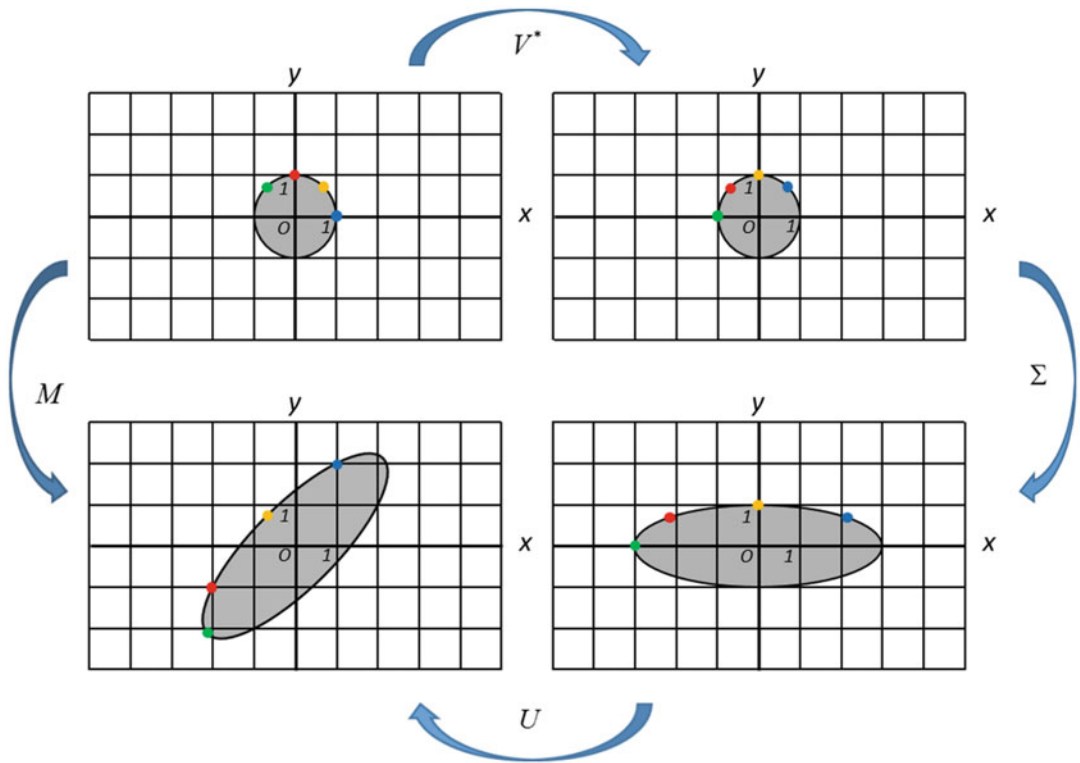

Fig. 5.6 Visualization of SVD for real square matrices to gain an intuitive understanding.
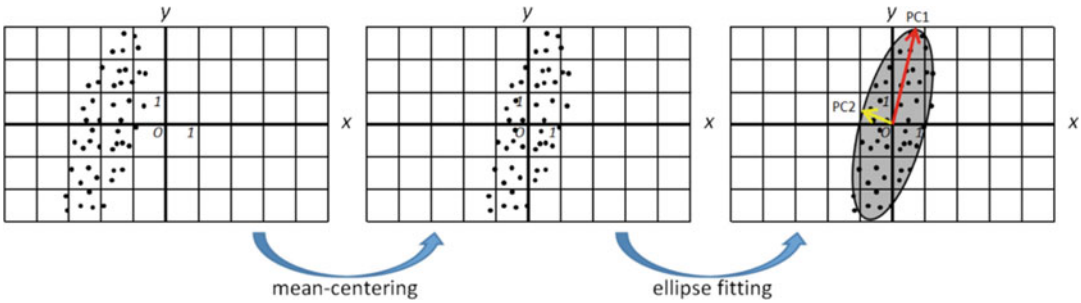
Fig. 5.7 Illustration of PCA in two dimensions. *Left*: example data distribution. *Middle*: mean-centered data distribution. *Right*: fitted ellipse and principal components PC1 (*red*) and PC2 (*yellow*).

Another method that is basically the same as SVD, that is also used for data compression but is formulated a bit differently is *principal component analysis* or PCA. PCA is typically described as a method that transforms data to a new coordinate system such that the largest variance is found along the first new coordinate (first principal component or PC), the next largest variance is found along the second new coordinate (second PC) etcetera. Let me explain this visually by an example in two dimensions. Suppose that the data is distributed as illustrated in Fig. 5.7 (left).

Here, we can order the data in a matrix $\mathbf{X}$ of dimensions $n \times 2$, where each row contains a data point (x_i, y_i), $i = 1 \ldots n$. In a p-dimensional space, our data matrix would have p columns. To perform PCA of the data (PC decomposition of this matrix), first the data is mean-centered (Fig. 5.7 (middle)), such that the mean in every dimension becomes zero. The next step of PCA can be thought of as fitting an ellipsoid (or ellipse in two dimensions) to the data (Fig. 5.7 (right)). The main axes of this ellipsoid represent the PCs; the longest axis the first PC, the next longest axis the second PC etcetera. Thus, long axes represent directions in the data with a lot of variance, short axes represent directions of little variance. The axes of the ellipsoid are represented by the orthonormalized eigenvectors of the *covariance matrix* of the data. The covariance matrix is proportional to $\mathbf{X}^*\mathbf{X}$ (or $\mathbf{X}^\mathbf{T}\mathbf{X}$ for real matrices). The proportion of the variance explained by a PC is equal to the eigenvalue belonging to the corresponding eigenvector divided by the sum of all eigenvalues. This must sound familiar: compare it to the explanation of the information maintained during data compression using SVD in the previous paragraph. Mathematically, the PC decomposition of the $n \times p$ matrix $\mathbf{X}$ is given by $\mathbf{T} = \mathbf{XW}$, where $\mathbf{W}$ (for 'weights') is a $p \times p$ matrix whose columns are the eigenvectors of $\mathbf{X}^*\mathbf{X}$ and $\mathbf{T}$ is an $n \times p$ matrix containing the component scores.

That PCA and SVD are basically the same can be understood as follows. We know now that the SVD of a matrix $\mathbf{M}$ is obtained by calculating the eigenvectors of $\mathbf{M}^*\mathbf{M}$ or $\mathbf{M}^\mathbf{T}\mathbf{M}$. And indeed, when the SVD of $\mathbf{X}$ is given by $X = U\Sigma W^T$, then $X^TX = (U\Sigma W^T)^T U\Sigma W^T = W\Sigma U^T U\Sigma W^T = W\Sigma^2 W^T$. Thus, the eigenvectors of $\mathbf{X}^\mathbf{T}\mathbf{X}$ are the columns of W, i.e. the right-singular vectors of $\mathbf{X}$. Or, in terms of the PC decomposition of $\mathbf{X}$: $T = XW = U\Sigma W^T W = U\Sigma$. Hence, each column of T is equal to the left singular vector of X times the corresponding singular value. Also notice the resemblance between Figs. 5.6 and 5.7.

As SVD, PCA is mostly used for data or dimensionality reduction (compression). By keeping only the first L components that explain most of the variance, high-dimensional

data may become easier to store, visualize and understand. This is achieved by, again, truncating the matrices **T** and **W** to their first L columns. An example of dimensionality reduction from work in my own group (collaboration with Dr. O.E. Martinez Manzanera) is to use PCA to derive simple quantitative, objective descriptors of movement to aid clinicians in obtaining a diagnosis of specific movement disorders. In this work we recorded and analyzed fingertip movement of *ataxia* patients performing the so-called finger-to-nose test, in which they are asked to repeatedly move their index finger from their nose to the fingertip of an examiner. In healthy people, the fingertip describes a smooth curve during the finger-to-nose task, while in ataxia patients, the fingertip trajectory is much more irregular. To describe the extent of irregularity, we performed PCA on the coordinates of the fingertip trajectory, assuming that the trajectory of healthy participants would be mostly in a plane, implying that for them two PCs would explain almost all of the variance in the data. Hence, as an application of dimensionality reduction using PCA, the variance explained by the first two PCs provides a compact descriptor of the regularity of movement during this task. This descriptor, together with other movement features, was subsequently used in a classifier to distinguish patients with ataxia from patients with a milder coordination problem and from healthy people.

Finally, I would like to briefly introduce the method of *independent component analysis* or ICA, here. For many it is confusing what the difference is between PCA and ICA and when to use one or the other. As explained before, PCA minimizes the covariance (second order moment) of the data by rotating the basis vectors, thereby employing *Gaussianity* or normality of the data. This means that the data have to be normally distributed for PCA to work optimally. ICA, on the other hand, can determine independent components for non-Gaussian signals by minimizing higher order moments of the data (such as *skewness* and *kurtosis*) that describe how non-Gaussian their distribution is. Here, independence means that knowing the values of one component does not give any information about another component. Thus, if the data is not well characterized by its variance then ICA may work better than PCA. Or, vice versa, when the data are Gaussian, linear and *stationary*, PCA will probably work. In practice, when sensors measure signals of several sources at the same time, ICA is typically the method of choice. Examples of signals that are best analyzed by ICA are simultaneous sound signals (such as speech) that are picked up by several receivers (such as microphones) or electrical brain activity recorded by multiple *EEG* electrodes. The ICA process is also referred to as blind source separation. Actually, our brain does a very good job at separating sources, when we identify our own name being mentioned in another conversation amidst the buzz of a party (the so-called 'cocktail party effect'). There are many applications of ICA in science, such as identifying and removing noise signals from EEG measurements (see e.g. Islam et al. 2016) and identifying brain functional connectivity networks in functional MRI measurements (see e.g. Calhoun and Adali 2012). In the latter case PCA and ICA are actually sometimes used successively, when PCA is first used for dimensionality reduction followed by ICA for identifying connectivity networks.

Glossary

Adjacency matrix Matrix with binary entries (i,j) describing the presence (1) or absence (0) of a path between nodes i and j.

Adjoint Transpose of the cofactor matrix.

Airfoil Shape of an airplane wing, propeller blade or sail.

Ataxia A movement disorder or symptom involving loss of coordination.

Basis vector A set of (N-dimensional) basis vectors is linearly independent and any vector in N-dimensional space can be built as a linear combination of these basis vectors.

Boundary conditions Constraints for a solution to an equation on the boundary of its domain.

Cofactor matrix The (i,j)-element of this matrix is given by the determinant of the matrix that remains when the i-th row and j-th column are removed from the original matrix, multiplied by -1 if $i + j$ is odd.

Conjugate transpose Generalization of transpose; a transformation of a matrix **A** indicated by $\mathbf{A}^*$ with elements defined by $(A^*)_{ij} = \bar{a}_{ji}$.

Dense matrix A matrix whose elements are almost all non-zero.

Determinant Can be seen as a scaling factor when calculating the inverse of a matrix.

Diagonalization Decomposition of a matrix **M** such that it can be written as $M = VDV^{-1}$ where **V** is an invertible matrix and **D** is a diagonal matrix.

Diagonal matrix A matrix with only non-zero elements on the diagonal and zeroes elsewhere.

Discretize To represent an equation on a grid.

EEG Electroencephalography; a measurement of electrical brain activity.

Eigendecomposition To determine the eigenvalues and eigenvectors of a matrix.

Element As in 'matrix element': one of the entries in a matrix.

Gaussian Normally distributed.

Graph A collection of nodes or vertices with paths or edges between them whenever the nodes are related in some way.

Hadamard product Element-wise matrix product.

Identity matrix A square matrix with ones on the diagonal and zeroes elsewhere, often referred to as **I**. The identity matrix is a special diagonal matrix.

Independent component analysis A method to determine independent components of non-Gaussian signals by minimizing higher order moments of the data.

Inverse The matrix $\mathbf{A}^{-1}$ such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.

Invertable A matrix that has an inverse.

Kronecker product Generalization of the outer product (or tensor product or dyadic product) for vectors to matrices.

Kurtosis Fourth-order moment of data, describing how much of the data variance is in the tail of its distribution.

Laplace equation Partial differential equation describing the behavior of potential fields.

Left-singular vector Columns of **U** in the SVD of **M**: $M = U\Sigma V^*$.

Leslie matrix Matrix with probabilities to transfer from one age class to the next in a population ecological model of population growth.

Logical matrix A matrix that only contains zeroes and ones (also: binary or Boolean matrix).

Matrix A rectangular array of (usually) numbers.

Network theory The study of graphs as representing relations between different entities, such as in a social network, brain network, gene network etcetera.

Order Size of a matrix.

Orthonormal Orthogonal vectors of length 1.

Partial differential equation An equation that contains functions of multiple variables and their partial derivatives (see also Chap. 6).

Principal component analysis Method that transforms data to a new coordinate system such that the largest variance is found along the first new coordinate (first PC), the then largest variance is found along the second new coordinate (second PC) etcetera.

Right-singular vector Columns of $\mathbf{V}$ in the SVD of $\mathbf{M}$: $M = U\Sigma V^*$

Root Here: a value of λ that makes the characteristic polynomial $|M - \lambda I|$ of the matrix $\mathbf{M}$ equal to zero.

Scalar function Function with scalar values.

Shearing To shift along one axis.

Singular value Diagonal elements of $\mathbf{\Sigma}$ in the SVD of $\mathbf{M}$: $M = U\Sigma V^*$.

Singular value decomposition The decomposition of an m × n rectangular matrix $\mathbf{M}$ into a product of three matrices such that $M = U\Sigma V^*$ where $\mathbf{U}$ is a unitary $m \times m$ matrix, $\mathbf{\Sigma}$ an $m \times n$ diagonal matrix with non-negative real entries and $\mathbf{V}$ another unitary $n \times n$ matrix.

Skewness Third-order moment of data, describing asymmetry of its distribution.

Skew-symmetric matrix A matrix $\mathbf{A}$ for which $a_{ij} = -a_{ji}$.

Sparse matrix A matrix with most of its elements equal to zero.

Stationary Time-dependent data for which the most important statistical properties (such as mean and variance) do not change over time.

Symmetric matrix A matrix $\mathbf{A}$ that is symmetric around the diagonal, i.e. for which $a_{ij} = a_{ji}$.

Transformation Here: linear transformation as represented by matrices. A function mapping a set onto itself (e.g. 2D space onto 2D space).

Transpose A transformation of a matrix $\mathbf{A}$ indicated by $\mathbf{A}^{\mathrm{T}}$ with elements defined by $(A^T)_{ij} = a_{ji}$.

Triangular matrix A diagonal matrix extended with non-zero elements only above or only below the diagonal.

Unit matrix Identity matrix.

Symbols Used in This Chapter (in Order of Their Appearance)

$\mathbf{M}$ or M	Matrix (bold and capital letter in text, italic and capital letter in equations)
$(\cdot)_{ij}$	Element at position (i,j) in a matrix
$\sum_{k=1}^{n}$	Sum over k, from 1 to n
$\vec{\cdot}$	Vector
θ	Angle
$\circ$	Hadamard product, Schur product or pointwise matrix product
$\otimes$	Kronecker matrix product
$\cdot^T$	(Matrix or vector) transpose
$\cdot^*$	(Matrix or vector) conjugate transpose
†	Used instead of * to indicate conjugate transpose in quantum mechanics
$\cdot^{-1}$	(Matrix) inverse
$|\cdot|$	(Matrix) determinant

Overview of Equations, Rules and Theorems for Easy Reference

Addition, subtraction and scalar multiplication of matrices
Addition of matrices **A** and **B** (of the same size):

$$(A + B)_{ij} = a_{ij} + b_{ij}$$

Subtraction of matrices **A** and **B** (of the same size):

$$(A - B)_{ij} = a_{ij} - b_{ij}$$

Multiplication of a matrix **A** by a scalar s:

$$(sA)_{ij} = sa_{ij}$$

Basis vector principle
Any vector $\begin{pmatrix} a \\ b \end{pmatrix}$ (in 2D space) can be built from the basis vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ by a linear
combination as follows: $\begin{pmatrix} a \\ b \end{pmatrix} = a \begin{pmatrix} 1 \\ 0 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.
The same principle holds for vectors in higher dimensions.

Rotation matrix (2D)
The transformation matrix that rotates a vector around the origin (in 2D) over an angle θ
(counter clockwise) is given by $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

Shearing matrix (2D)
$\begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$: shearing along the x-axis (y-coordinate remains unchanged)
$\begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}$: shearing along the y-axis (x-coordinate remains unchanged)

Matrix product
Multiplication **AB** of an $m \times n$ matrix **A** with an $n \times p$ matrix **B**:

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$

Hadamard product, Schur product or pointwise product:

$$(A \circ B)_{ij} = a_{ij} b_{ij}$$

Kronecker product:

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mn}B \end{pmatrix}$$

Special matrices
Hermitian matrix: $\mathbf{A} = \mathbf{A}^*$
normal matrix: $\mathbf{A}^*\mathbf{A} = \mathbf{AA}^*$
unitary matrix: $\mathbf{AA}^* = \mathbf{I}$

where $(A^*)_{ij} = \bar{a}_{ji}$ defines the conjugate transpose of $\mathbf{A}$.

Matrix inverse
For a square matrix $\mathbf{A}$ the inverse $A^{-1} = \frac{1}{\det(A)}\,adj(A)$, where $\det(\mathbf{A})$ is the determinant of $\mathbf{A}$ and adj($\mathbf{A}$) is the adjoint of $\mathbf{A}$ (see Sect. 5.3.1).

Eigendecomposition
An eigenvector $\vec{v}$ of a square matrix $\mathbf{M}$ is determined by:

$$M\vec{v} = \lambda\vec{v},$$

where λ is a scalar known as the eigenvalue

Diagonalization
Decomposition of a square matrix $\mathbf{M}$ such that:

$$M = VDV^{-1}$$

where $\mathbf{V}$ is an invertible matrix and $\mathbf{D}$ is a diagonal matrix
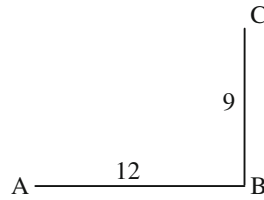
Singular value decomposition
Decomposition of an m × n rectangular matrix $\mathbf{M}$ such that:

$$M = U\Sigma V^*$$

where $\mathbf{U}$ is a unitary $m \times m$ matrix, $\boldsymbol{\Sigma}$ an $m \times n$ diagonal matrix with non-negative real entries and $\mathbf{V}$ another unitary $n \times n$ matrix.
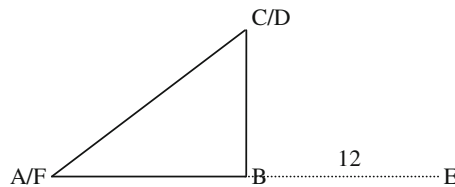
Answers to Exercises

5.1. (a)

(b) The direct distance between cities A and C can be calculated according to Pythagoras' theorem as $\sqrt{12^2 + 9^2} = \sqrt{144 + 81} = \sqrt{225} = 15$. Hence, the distance matrix becomes $\begin{pmatrix} 0 & 12 & 15 \\ 12 & 0 & 9 \\ 15 & 9 & 0 \end{pmatrix}$.

(c)

5.2. The sum and difference of the pairs of matrices are:

(a) $\begin{pmatrix} 5 & 2 \\ 2 & 15 \end{pmatrix}$ and $\begin{pmatrix} 1 & 6 \\ -4 & 1 \end{pmatrix}$

(b) $\begin{pmatrix} 7 & -4 & 6 \\ -1 & 4 & 1 \\ -4 & 6 & 2 \end{pmatrix}$ and $\begin{pmatrix} -1 & -10 & 2 \\ -3 & 8 & 9 \\ 6 & -10 & -20 \end{pmatrix}$

(c) $\begin{pmatrix} 2 & 1.6 & -1 \\ 5.1 & 1 & -2 \end{pmatrix}$ and $\begin{pmatrix} 0.4 & 4.8 & -2 \\ 1.7 & 3.6 & -4.4 \end{pmatrix}$

5.3. (a) $\begin{pmatrix} 6 & 0 \\ 7 & -4 \\ 8 & 10 \end{pmatrix}$

(b) $\begin{pmatrix} 3.5 & 0 \\ 2.9 & 0.7 \end{pmatrix}$

5.4. Possibilities for multiplication are **AB**, **AC**, **BD**, **CB** and **DA**.

5.5. (a) 2×7
(b) 2×1
(c) 1×1

5.6. (a) $\mathbf{AB} = \begin{pmatrix} 18 & 22 \\ 22 & 58 \end{pmatrix}$, $\mathbf{BA} = \begin{pmatrix} 8 & -8 \\ 2 & 68 \end{pmatrix}$

(b) $\mathbf{BA} = \begin{pmatrix} 8 & -14 \\ 3 & -11 \\ -20 & 61 \end{pmatrix}$

(c) no matrix product possible

5.7. (a) $\begin{pmatrix} 2 & 6 \\ -4 & 1 \end{pmatrix}$

(b) $\begin{pmatrix} 2.8 & -2.7 & 0.5 \\ 12 & -0.7 & -4 \end{pmatrix}$

5.8. (a) $\begin{pmatrix} 3 & -1 & 4 & 6 & -2 & 8 \\ -3 & 1 & -4 & 3 & -1 & 4 \end{pmatrix}$

(b) $\begin{pmatrix} 0 & -2 & -4 & 0 & -3 & -6 \\ -6 & -8 & -10 & -9 & -12 & -15 \\ -12 & -14 & -16 & -18 & -21 & -24 \end{pmatrix}$

5.9. (a) symmetric, logical
(b) sparse, upper-triangular
(c) skew-symmetric
(d) upper-triangular
(e) diagonal, sparse
(f) identity, diagonal, logical, sparse

5.10. (a) $\begin{pmatrix} 1 & i & 5 \\ 2 & 1 & 4-5i \\ 3 & -3+2i & 3 \end{pmatrix}$

(b) $\begin{pmatrix} 1 & -1 & 5 \\ 2 & 1 & 4 \\ 3 & -3 & 0 \end{pmatrix}$

(c) $\begin{pmatrix} 4 & 19-i & 8i \\ 0 & -3 & -11+i \\ 3+2i & -3 & 17 \end{pmatrix}$

5.11. Using Cramer's rule we find that $x = \dfrac{D_x}{D} = \begin{vmatrix} c & b \\ f & e \end{vmatrix} / \begin{vmatrix} a & b \\ d & e \end{vmatrix} = \dfrac{ce-bf}{ae-bd}$ and

$y = \dfrac{D_y}{D} = \begin{vmatrix} a & c \\ d & f \end{vmatrix} / \begin{vmatrix} a & b \\ d & e \end{vmatrix} = \dfrac{af-cd}{ae-bd}$. From Sect. 5.3.1 we obtain that the inverse of the

matrix $\begin{pmatrix} a & b \\ d & e \end{pmatrix}$ is equal to $\dfrac{1}{ae-bd} \begin{pmatrix} e & -b \\ -d & a \end{pmatrix}$ and the solution to the system of linear

equations is $\dfrac{1}{ae-bd}\begin{pmatrix} e & -b \\ -d & a \end{pmatrix}\begin{pmatrix} c \\ f \end{pmatrix} = \begin{pmatrix} (ce-bf)/(ae-bd) \\ (af-cd)/(ae-bd) \end{pmatrix}$ which is the same as the solution obtained using Cramer's rule.

5.12. (a) $x=-11, y=-5\frac{3}{5}$

(b) $D = \begin{vmatrix} 4 & -2 & -2 \\ 2 & 8 & 4 \\ 30 & 12 & -4 \end{vmatrix} = 4(8\cdot-4-4\cdot12)-\cdot-2(2\cdot-4-30\cdot4)+\cdot-2(2\cdot12$
$-30\cdot8) = -144$

$$D_x = \begin{vmatrix} 10 & -2 & -2 \\ 32 & 8 & 4 \\ 24 & 12 & -4 \end{vmatrix} = -1632 \quad D_y = \begin{vmatrix} 4 & 10 & -2 \\ 2 & 32 & 4 \\ 30 & 24 & -4 \end{vmatrix} = 2208$$

$$D_z = \begin{vmatrix} 4 & -2 & 10 \\ 2 & 8 & 32 \\ 30 & 12 & 24 \end{vmatrix} = -4752$$

Thus, $\quad x = \dfrac{D_x}{D} = \dfrac{-1632}{-144} = 11\dfrac{1}{3}, \qquad y = \dfrac{D_y}{D} = \dfrac{2208}{-144} = -15\dfrac{1}{3} \qquad$ and

$z = \dfrac{D_z}{D} = \dfrac{-4752}{-144} = 33$

5.13. (a) $x=4, y=0$

(b) $x=2, y=-1, z=1$

5.14. This shearing matrix shears along the x-axis and leaves y-coordinates unchanged (see Sect. 5.2.2). Hence, all vectors along the x-axis remain unchanged due to this transformation. The eigenvector is thus $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ with eigenvalue 1 (since the length of the eigenvector is unchanged due to the transformation).

5.15. (a) $\lambda_1=7$ with eigenvector $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ (x = x, y = 0, z = 0), $\lambda_2=-19$ with eigenvector

$\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ (x = 0, y = y, z = 0) and $\lambda_3=2$ with eigenvector $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ (x = 0, y = 0, z = z).

(b) $\lambda=3$ (double) with eigenvector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ (y = x).

(c) $\lambda_1=1$ with eigenvector $\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$ (x = −z, y = z), $\lambda_2=-1$ with eigenvector $\begin{pmatrix} 1 \\ 1 \\ 5 \end{pmatrix}$

(5x = z, 5y = z) and $\lambda_3=3$ with eigenvector $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ (x = y = z).

(d) $\lambda_1 = 3$ with eigenvector $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ (y = −2x) and $\lambda_2 = 7$ with eigenvector $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ (y = 2x).

5.16. (a) To determine the SVD, we first determine the eigenvalues and eigenvectors of $\mathbf{M}^\mathrm{T}\mathbf{M}$ to get the singular values and right-singular vectors of $\mathbf{M}$. Thus, we determine its characteristic equation as the determinant

$$\left| \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - \lambda\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| = \left| \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix} - \lambda\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right| =$$

$$\left| \begin{pmatrix} 5 - \lambda & 4 \\ 4 & 5 - \lambda \end{pmatrix} \right| = (5 - \lambda)^2 - 16 = \lambda^2 - 10\lambda + 9 = (\lambda - 1)(\lambda - 9)$$

Thus, the singular values are $\sigma_1 = \sqrt{9} = 3$ and $\sigma_1 = \sqrt{1} = 1$ and $\Sigma = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$.

The eigenvector belonging to the first eigenvalue of $\mathbf{M}^\mathrm{T}\mathbf{M}$ follows from:

$$\begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = 9\begin{pmatrix} x \\ y \end{pmatrix} \Rightarrow \begin{pmatrix} 5x + 4y \\ 4x + 5y \end{pmatrix} = \begin{pmatrix} 9x \\ 9y \end{pmatrix}$$
$$\Rightarrow 5x + 4y = 9x \wedge 4x + 5y = 9y$$
$$\Rightarrow -4x + 4y = 0 \wedge 4x - 4y = 0$$
$$\Rightarrow x = y$$
$$\Rightarrow \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

To determine the first column of $\mathbf{V}$, this eigenvector must be normalized (divided by its length; see Sect. 4.2.2.1) and thus $\vec{v}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Similarly, the eigenvector belonging to the second eigenvalue of $\mathbf{M}^\mathbf{T}\mathbf{M}$ can be derived to be $\vec{v}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, making $V = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. In this case, $\vec{v}_1$ and $\vec{v}_2$ are already orthogonal (see Sect. 4.2.2.1), making further adaptations to arrive at an orthonormal set of eigenvectors unnecessary.

To determine U we use that $\vec{u}_1 = \frac{1}{\sigma_1}M\vec{v}_1 = \frac{1}{3}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\vec{u}_2 = \frac{1}{\sigma_2}M\vec{v}_2 = \frac{1}{1}\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, making $U = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

You can now verify yourself that indeed $M = U\Sigma V^*$.

(b) Taking a similar approach, we find that

$$U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \Sigma = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } V = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

5.17. If $M = U\Sigma V^*$ and using that both **U** and **V** are unitary (see Table 5.1), then $MM^* = U\Sigma V^*(U\Sigma V^*)^* = U\Sigma V^* V\Sigma U^* = U\Sigma^2 U^*$. Right-multiplying both sides of this equation with U and then using that Σ^2 is diagonal, yields $MM^* U = U\Sigma^2 = \Sigma^2 U$. Hence, the columns of U are eigenvectors of **MM*** (with eigenvalues equal to the diagonal elements of Σ^2).

References

Online Sources of Information: History

http://www-history.mcs.st-and.ac.uk/history/HistTopics/Matrices_and_determinants.html#86

Online Sources of Information: Methods

https://en.wikipedia.org/wiki/Matrix_(mathematics)
https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors
https://en.wikipedia.org/wiki/Singular_value_decomposition
https://en.wikipedia.org/wiki/Rotation_matrix
https://www.cs.helsinki.fi/u/ahyvarin/whatisica.shtml
http://comnuan.com/cmnn01004/ (one of many online matrix calculators)
http://andrew.gibiansky.com/blog/mathematics/cool-linear-algebra-singular-value-decomposition/
   (examples of information compression using SVD)

Books

R.A. Horn, C.R. Johnson, *Topics in Matrix Analysis* (Cambridge University Press, Cambridge, 1994)
W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, *Numerical Recipes, the Art of Scientific Computing. (Multiple Versions for Different Programming Languages)* (Cambridge University Press, Cambridge)
Y. Saad. Iterative Methods for Sparse Linear Systems, SIAM, New Delhi 2003. http://www-users.cs.umn.edu/~saad/IterMethBook_2ndEd.pdf

Papers

V.D. Calhoun, T. Adali, Multisubject independent component analysis of fMRI: a decade of intrinsic networks, default mode, and neurodiagnostic discovery. IEEE Rev. Biomed. Eng. **5**, 60–73 (2012)
M.K. Islam, A. Rastegarnia, Z. Yang, Methods for artifact detection and removal from scalp EEG: a review. Neurophysiol. Clin. **46**, 287–305 (2016)