

Chapter 2

Elements of Linear Algebra

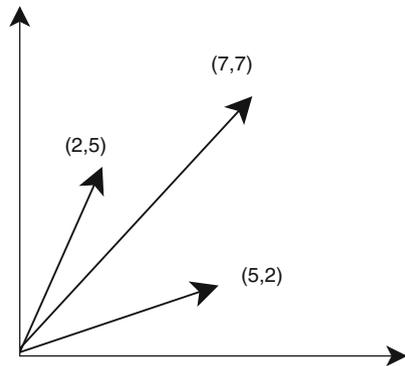
2.1 Introduction

This book will use some basic concepts of linear algebra. In this chapter we will briefly recall the essential elements that will be necessary in the rest of the book to understand the various tools devoted to the analysis of variance. Readers that are interested in the more detailed treatment of the subject are directed, for instance, to Golub and Van Loan (1996), Horn and Johnson (1991), Meyer (2000).

2.2 Elementary Vectors

The typical geometric definition introduces a vector as a segment emanating from the origin, with an arrow at the second extreme, indicating a “pointing” direction, or orientation (cf. Fig. 2.1), showing that a vector may be characterized by two properties: length and direction. Although this definition is usually employed on the plane, the same characterization can be used in higher dimension, that is on hyper-planes (e.g. space). While magnitude and direction, for instance, would be sufficient to uniquely identify a vector on the plane, this is not so in higher dimensions. In high dimensions, it is thus more appropriate to characterize vectors by means of their “components”. In Cartesian coordinates, these are the orthogonal projections of the vector on each Cartesian axis. Rigorously speaking, a vector is given by an *ordered* n -uple of real or complex numbers, that is, $\mathbf{b} = (b_1, \dots, b_n)$ is a (row) vector with n components, where each b_i is a real or complex number. Note that the order of the components is important, so that, e.g., the vector $\mathbf{a} = (1, 3)$ is different from the vector $\mathbf{b} = (3, 1)$. The ensemble of all possible vectors is then identified by the ensemble of all possible n -uples of numbers that can be formed with real or complex numbers. A vector with only one component is called a scalar. The vector $\mathbf{0} = (0, \dots, 0)$ is the zero vector. It is customary to identify the whole of the real and complex numbers with the symbols \mathbb{R} and \mathbb{C} , respectively. It thus follows, for instance, that the set of all possible couples is denoted by the symbol $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$. In general, \mathbb{R}^n is the set of vectors having n components.

Fig. 2.1 Elementary vectors on the plane



The addition between two vectors *with the same number of components*, is defined as the vector whose components are the sum of the corresponding vector components. If $\mathbf{a} = (a_1, a_2, \dots, a_n)$ and $\mathbf{b} = (b_1, b_2, \dots, b_n)$, then $\mathbf{c} := \mathbf{a} + \mathbf{b}$ as $\mathbf{c} = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n)$. Note that \mathbf{c} is a vector of n components. By again acting at the component level, we can stretch a vector by multiplying it with a scalar k : we can define $\mathbf{c} = \mathbf{a}k$ where $\mathbf{c} = (ka_1, ka_2, \dots, ka_n)$, meaning that each component is multiplied by the factor k . These are the basic operations that allow us to generate a key space for our analysis. In particular, \mathbb{R}^n is closed with respect to the sum and with respect to multiplication by a real scalar, which means that the result of these operations is still an element of \mathbb{R}^n . A real vector space is a set that is closed with respect to the addition and multiplication by a real scalar. Therefore, \mathbb{R}^n is a real vector space. There are more complex instances of vector spaces, but for the moment we will content ourselves with this fundamental example. An immediate generalization is given by the definition of a real vector subspace, which is a subset of a real vector space.

2.3 Scalar Product

We next introduce an operation between two vectors that provides the main tool for a geometric interpretation of vector spaces. Given two real vectors $\mathbf{a} = (a_1, a_2, \dots, a_n)$ and $\mathbf{b} = (b_1, b_2, \dots, b_n)$, we define the scalar product (or inner product) the operation $\langle \mathbf{a}, \mathbf{b} \rangle = a_1b_1 + a_2b_2 + \dots + a_nb_n$. Note that the operation is between vectors, whereas the result is a real scalar. We remark that if \mathbf{a} and \mathbf{b} were complex vectors, that is vectors with complex components, then a natural inner product would be defined in a different way, and in general, the result would be a complex number (see end of section). The real inner product inherits many useful properties from the product and sum of real numbers. In particular, for any vector $\mathbf{a}, \mathbf{b}, \mathbf{c}$ with n real components and for any real scalar k , it holds

1. Commutative property: $\langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{a} \rangle$
2. Distributive property: $\langle (\mathbf{a} + \mathbf{c}), \mathbf{b} \rangle = \langle \mathbf{a}, \mathbf{b} \rangle + \langle \mathbf{c}, \mathbf{b} \rangle$
3. Multiplication by scalar: $\langle (k\mathbf{a}), \mathbf{b} \rangle = k \langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{a}, (k\mathbf{b}) \rangle$

The scalar product between a vector and itself is of great interest, that is

$$\langle \mathbf{a}, \mathbf{a} \rangle = a_1^2 + a_2^2 + \cdots + a_n^2.$$

Note that $\langle \mathbf{a}, \mathbf{a} \rangle$ is always non-negative, since it is the sum of non-negative numbers. For $n = 2$ it is easily seen from Fig. 3.1 that this is the square of the length of a vector. More generally, we define the Euclidean norm (or simply norm) as

$$\|\mathbf{a}\| = \sqrt{\langle \mathbf{a}, \mathbf{a} \rangle}.$$

A versor is a vector of unit norm. Given a non-zero vector \mathbf{x} , it is always possible to determine a versor \mathbf{x}' by dividing \mathbf{x} by its norm, that is $\mathbf{x}' = \mathbf{x}/\|\mathbf{x}\|$. This is a standard form of normalization, ensuring that the resulting vector has norm one. Other normalizations may require to satisfy different criteria, such as, e.g., the first component equal to unity. If not explicitly mentioned, we shall always refer to normalization to obtain unit norm vectors. Given a norm, in our case the Euclidean norm, the distance associated with this norm is

$$d(\mathbf{a}, \mathbf{b}) = \|\mathbf{a} - \mathbf{b}\| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \cdots + (a_n - b_n)^2}.$$

Scalar products and the induced distance can be defined in several ways; here we are showing only what we shall mostly use in this text. Any function can be used as a norm as long as it satisfies three basic relations: (i) Non-negativity: $\|\mathbf{a}\| \geq 0$ and $\|\mathbf{a}\| = 0$ if and only if $\mathbf{a} = \mathbf{0}$; (ii) Commutative property: $d(\mathbf{a}, \mathbf{b}) = d(\mathbf{b}, \mathbf{a})$; (iii) Triangular inequality: $d(\mathbf{a}, \mathbf{b}) \leq d(\mathbf{a}, \mathbf{c}) + d(\mathbf{c}, \mathbf{b})$. Using norms we can distinguish between close vectors and far away vectors, in other words we can introduce a topology in the given vector space. In particular, property (i) above ensures that identical vectors ($\mathbf{a} = \mathbf{b}$) have a zero distance. As an example of the new possibility offered by vector spaces, we can go back to Fig. 2.1 and consider the angles α and β that the vectors \mathbf{a} and \mathbf{b} in \mathbb{R}^2 make with the reference axes. These angles can be easily expressed in terms of the components of the vectors,

$$\begin{aligned} \cos \beta &= \frac{b_1}{\sqrt{b_1^2 + b_2^2}} & \sin \beta &= \frac{b_2}{\sqrt{b_1^2 + b_2^2}}, \\ \cos \alpha &= \frac{a_1}{\sqrt{a_1^2 + a_2^2}} & \sin \alpha &= \frac{a_2}{\sqrt{a_1^2 + a_2^2}}, \end{aligned}$$

and also the angle between the two vectors, $\cos(\beta - \alpha)$,

$$\begin{aligned} \cos(\beta - \alpha) &= \cos \beta \cos \alpha + \sin \beta \sin \alpha \\ &= \frac{b_1}{\sqrt{b_1^2 + b_2^2}} \frac{a_1}{\sqrt{a_1^2 + a_2^2}} + \frac{b_2}{\sqrt{b_1^2 + b_2^2}} \frac{a_2}{\sqrt{a_1^2 + a_2^2}}, \end{aligned} \quad (2.1)$$

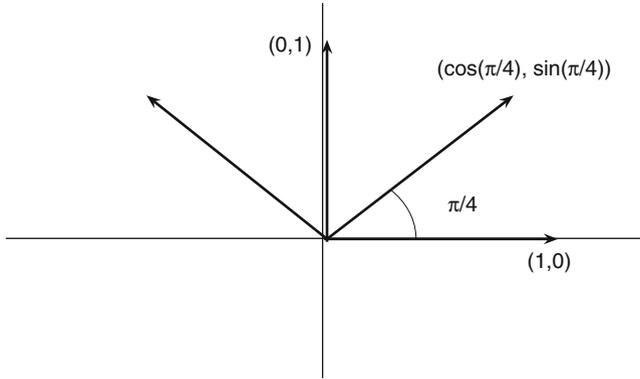


Fig. 2.2 Angles between vectors

or, equivalently,

$$\cos(\beta - \alpha) = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\| \|\mathbf{b}\|}. \quad (2.2)$$

We just proved that this relation holds in \mathbb{R}^2 . In higher dimension, the cosine of the angle of two vectors is defined as the ratio between their inner product and their norms, which is nothing but (2.2).

We explicitly observe that the scalar product of two vectors gives directly the cosine of the angle between them. By means of this new notion of angle, relative direction of vectors can now be expressed in terms of the scalar product. We say that two vectors are orthogonal if their inner product is zero. Formula (2.2) provides a geometric justification for this definition, which can be explicitly derived in \mathbb{R}^2 , where orthogonality means that the angles between the two vectors is $\pi/2$ radians (90°); cf. Fig. 2.2. If in addition the two vectors are in fact versors, they are said to be orthonormal.

We can also introduce another geometric interpretation of scalar products that follows from (2.2). The scalar product is also the projection of the vector \mathbf{a} on \mathbf{b} : from Fig. 2.3 and from the definition of the cosine the projection of \mathbf{a} onto the direction of \mathbf{b} is $Proj_{\mathbf{b}} \mathbf{a} = \|\mathbf{a}\| \cos \phi$. Analogously, the projection of \mathbf{b} onto the direction of \mathbf{a} is $Proj_{\mathbf{a}} \mathbf{b} = \|\mathbf{b}\| \cos \phi$. For normalized vectors the norm disappears and the scalar product gives directly the projections, that are obviously the same in both cases (bottom panel in Fig. 2.3). We close this section with the definition of inner product in the case of complex vectors. Let \mathbf{x}, \mathbf{y} be vectors in \mathbb{C}^n . Then $\langle \mathbf{x}, \mathbf{y} \rangle = \bar{x}_1 y_1 + \bar{x}_2 y_2 + \cdots + \bar{x}_n y_n$, where $\bar{x} = a - ib$ denotes the complex conjugate of $x = a + ib$, $i = \sqrt{-1}$. With this definition, the norm of a complex vector is defined as $\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle = |x_1|^2 + \cdots + |x_n|^2$.

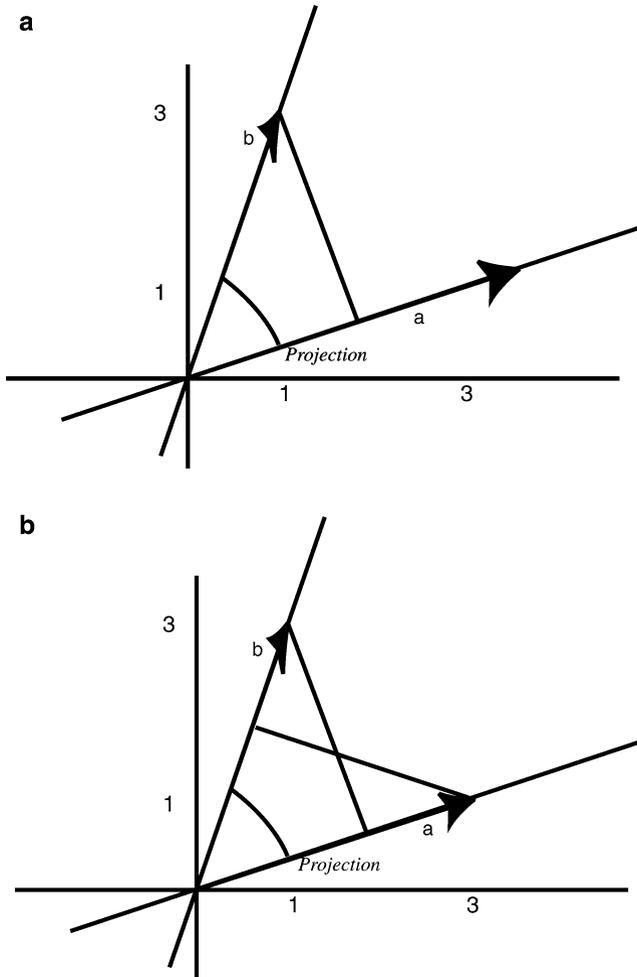


Fig. 2.3 Angles between vectors

Exercises and Problems

- Given the two vectors $\mathbf{a} = (1, -2, 0)$, $\mathbf{b} = (-3, -1, 4)$, compute $\mathbf{a} + \mathbf{b}$, $\mathbf{a} - \mathbf{b}$, $\mathbf{a} + 2\mathbf{b}$ and $\langle \mathbf{a}, \mathbf{b} \rangle$.
*We have $\mathbf{a} + \mathbf{b} = (1 - 3, -2 - 1, 0 + 4) = (-2, -3, 4)$, $\mathbf{a} - \mathbf{b} = (1 + 3, -2 + 1, 0 - 4) = (4, -1, -4)$ and $\mathbf{a} + 2\mathbf{b} = (1 + 2(-3), -2 + 2(-1), 0 + 2(4)) = (-4, -4, 8)$.
 Finally, we have $\langle \mathbf{a}, \mathbf{b} \rangle = 1(-3) + (-2)(-1) + 0(4) = -3 + 2 + 0 = -1$.*
- Given the two complex vectors $\mathbf{x} = (1 + i, -2 + 3i)$, $\mathbf{y} = (-5 + i, 4i)$, compute $\mathbf{x} + \mathbf{y}$ and $\langle \mathbf{x}, \mathbf{y} \rangle$.

We have $\mathbf{x} + \mathbf{y} = (1 - 5 + (1 + 1)i, -2 + (3 + 4)i) = (-4 + 2i, -2 + 7i)$.
 Moreover, using the definition of the product between complex numbers, $\langle \mathbf{x}, \mathbf{y} \rangle = (1 - i)(-5 + i) + (-2 - 3i)(4i) = 8 - 2i$.

3. Given the vectors $\mathbf{a} = (1, -2, 1)$ and $\mathbf{b} = (0, 2, -3)$, compute $\|\mathbf{a}\|$, $\|\mathbf{b}\|$ and $\|\mathbf{a} - \mathbf{b}\|$. Moreover, normalize \mathbf{a} so as to have unit norm.

We have $\|\mathbf{a}\| = (1^2 + (-2)^2 + 1^2)^{1/2} = 6^{1/2}$ and $\|\mathbf{b}\| = (0^2 + 2^2 + (-3)^2)^{1/2} = 13^{1/2}$. Moreover, $\|\mathbf{a} - \mathbf{b}\| = ((1 - 0)^2 + (-2 - 2)^2 + (1 - 3)^2)^{1/2} = 21^{1/2}$.
 Finally, $\mathbf{a}_i = \mathbf{a}/\|\mathbf{a}\| = (1/6)^{1/2}(1, -2, 1)$.

4. Check whether the following operations or results are admissible: (i) $\mathbf{x} + \mathbf{y}$, with $\mathbf{x} = (1, -1)$, $\mathbf{y} = (1, 2, 0)$; (ii) $\langle \mathbf{x}, \mathbf{y} \rangle$ with \mathbf{x} and \mathbf{y} as in (i); (iii) $\|\mathbf{a}\| = -1$; (iv) $\langle \mathbf{a}, \mathbf{b} \rangle = -\langle \mathbf{b}, \mathbf{a} \rangle$, with \mathbf{a}, \mathbf{b} real vectors of equal dimension; (v) $d(\mathbf{c}, \mathbf{d}) = -1.5$.
None of the statement above is correct. (i) \mathbf{x} and \mathbf{y} have a different number of components hence the two vectors cannot be added. (ii) Same as in (i). (iii) The norm of any vector is non-negative, therefore it cannot be equal to -1. (iv) The inner product of real vectors is commutative, therefore $\langle \mathbf{a}, \mathbf{b} \rangle = \langle \mathbf{b}, \mathbf{a} \rangle$. (v) Same as in (iii).

5. Compute the cosine of the angle between the vectors $\mathbf{a} = (-1, 2)$ and $\mathbf{b} = (-3, 0)$.

We first compute $\langle \mathbf{a}, \mathbf{b} \rangle = -1(-3) + 2(0) = 3$, $\|\mathbf{a}\| = \sqrt{5}$ and $\|\mathbf{b}\| = 3$, from which we obtain $\cos \phi = \langle \mathbf{a}, \mathbf{b} \rangle / (\|\mathbf{a}\| \|\mathbf{b}\|) = \frac{1}{\sqrt{5}}$.

2.4 Linear Independence and Basis

Some vectors can be combined and stretched by scalars, hence they can be obtained one from the other. For instance, the vector $(4, 4, 4)$ can be obtained as $(1, 1, 1) \cdot 4$ in such a way that all vectors of the form (k, k, k) are really different stretched versions of the same vector $(1, 1, 1)$. Vectors that cannot be reached with a simple stretching can be obtained with a combination, for instance the vector $(5, 2)$ can be written as $2 \cdot (1, 1) + 3 \cdot (1, 0)$. With this simple example we see that we can choose some particularly convenient vectors to represent all other vectors in the given space. Given r nonzero vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r$, we say that a vector \mathbf{x} is a linear combination of these r vectors if there exist r scalars $\alpha_1, \dots, \alpha_r$, not all equal to zero, such that

$$\mathbf{x} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \dots + \alpha_r \mathbf{x}_r.$$

This definition is used to distinguish between linearly dependent and independent vectors. In particular, $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r$ are said to be linearly dependent if there exist r scalars, not all equal to zero, such that $\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \dots + \alpha_r \mathbf{x}_r = \mathbf{0}$. In other words, they are linearly dependent if one of the vectors can be expressed as a linear combination of the other vectors. We are thus ready to define linearly independent vectors, and the associated concept of a basis of a vector space. We say that r vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_r$ are linearly independent if the only linear combination that gives the zero vector is obtained by setting all scalars equal to zero, that is if

the relation $\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \cdots + \alpha_r \mathbf{x}_r = \mathbf{0}$ implies $\alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$. The maximum number of linearly independent vectors is called the dimensionality of the vector space, maybe is not surprising that for \mathbb{R}^n this number turns out to be n . Given n linearly independent vectors in \mathbb{R}^n , any other vector can be obtained as a linear combination of these n vectors. For this reason, n linearly independent vectors of \mathbb{R}^n , are called a basis of \mathbb{R}^n . Clearly, a basis is not uniquely determined, since any group of n linearly independent vectors represents a basis. However, choices that are particularly convenient are given by sets of normalized and mutually orthogonal and thus independent vectors, namely, we select an “orthonormal basis”. In the case of \mathbb{R}^2 , orthonormal bases are for instance $(1, 0)$, $(0, 1)$, and also $(1/\sqrt{2}, 1/\sqrt{2})$, $(1/\sqrt{2}, -1/\sqrt{2})$. In fact, the latter can be obtained from the first one with a rotation, as it is shown in Fig. 2.2.

The orthonormal basis $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$, where $\mathbf{e}_k = (0, 0, \dots, 1, 0, \dots, 0)$, that is, all components are zero except the unit k th component, is called the canonical basis of \mathbb{R}^n . Note that it is very simple to obtain the coefficients in the linear combination of a vector of \mathbb{R}^n in terms of the canonical basis: these coefficients are simply the components of the vector (see Exercise 3 below). The choice of a particular basis is mainly dictated by either computational convenience or by ease of interpretation. Given two vectors \mathbf{x}, \mathbf{y} in \mathbb{R}^n , it is always possible to generate a vector from \mathbf{x} , that is orthogonal to \mathbf{y} . This goes as follows: we first define the vector $\mathbf{y}' = \mathbf{y}/\|\mathbf{y}\|$ and the scalar $t = \langle \mathbf{y}', \mathbf{x} \rangle$, with which we form $\mathbf{x}' = \mathbf{x} - \mathbf{y}'t$. The computed \mathbf{x}' is thus orthogonal to \mathbf{y} . Indeed, using the properties of the inner product, $\langle \mathbf{y}', \mathbf{x}' \rangle = \langle \mathbf{y}', \mathbf{x} - \mathbf{y}'t \rangle = \langle \mathbf{y}', \mathbf{x} \rangle - t \langle \mathbf{y}', \mathbf{y}' \rangle = t - t = 0$.

Determining an orthogonal basis of a given space is a major task. In \mathbb{R}^2 this is easy: given any vector $\mathbf{a} = (a_1, a_2)$, the vector $\mathbf{b} = (-a_2, a_1)$ (or $\mathbf{c} = -\mathbf{b} = (a_2, -a_1)$) is orthogonal to \mathbf{a} , therefore the vectors \mathbf{a}, \mathbf{b} readily define an orthogonal basis. In \mathbb{R}^n the process is far less trivial. A stable way to proceed is to take n linearly independent vectors $\mathbf{u}_1, \dots, \mathbf{u}_n$ of \mathbb{R}^n , and then orthonormalize them in a sequential manner. More precisely, we first normalize \mathbf{u}_1 to get \mathbf{v}_1 ; we take \mathbf{u}_2 , we orthogonalize it against \mathbf{v}_1 and then normalize it to get \mathbf{v}_2 . We thus continue with \mathbf{u}_3 , orthogonalize it against \mathbf{v}_1 and \mathbf{v}_2 and get \mathbf{v}_3 after normalization, and so on. This iterative procedure is the famous Gram-Schmidt process.

Exercises and Problems

- Given the two vectors $\mathbf{a} = (-1, -2)$ and $\mathbf{b} = (-3, -1)$: (i) verify that \mathbf{a} and \mathbf{b} are linearly independent. (ii) Compute a vector orthogonal to \mathbf{a} . (iii) If possible, determine a scalar k such that $\mathbf{c} = k\mathbf{a}$ and \mathbf{a} are linearly independent.
 - In \mathbb{R}^2 vectors are either multiple of each other or they are independent. Since \mathbf{b} is not a multiple of \mathbf{a} , we have that \mathbf{a} and \mathbf{b} are linearly independent. (ii) The vector $\mathbf{d} = (2, -1)$ is orthogonal to \mathbf{a} , indeed $\langle \mathbf{a}, \mathbf{d} \rangle = (-1)(2) + (-2)(-1) = 0$.
 - From the answer to (i), it follows that there is no such \mathbf{c} .
- Obtain an orthonormal set from the two linearly independent vectors: $\mathbf{a} = (2, 3)$ and $\mathbf{b} = (1, 1)$.

We use the Gram-Schmidt process. First, $\mathbf{a}' = \mathbf{a}/\|\mathbf{a}\| = \frac{1}{\sqrt{13}}(2, 3)$. Then, we compute $\langle \mathbf{a}', \mathbf{b} \rangle = \frac{5}{\sqrt{13}}$, so that $\mathbf{b} = \mathbf{b} - \mathbf{a}'\langle \mathbf{a}', \mathbf{b} \rangle = (1, 1) - \frac{5}{\sqrt{13}}\mathbf{a}' = (1, 1) - \frac{5}{13}(2, 3) = \frac{1}{13}(3, -2)$, from which $\mathbf{b}' = \mathbf{b}/\|\mathbf{b}\| = \frac{1}{2\sqrt{13}}(3, -2)$. Hence, \mathbf{a}', \mathbf{b}' is the sought after set. Not surprisingly (cf. text), \mathbf{b}' is the normalized version of $(-3, 2)$, easily obtainable directly from \mathbf{a} .

3. Given the vector $\mathbf{x} = (3, -2, 4)$, determine the coefficients in the canonical basis. We simply have $\mathbf{x} = 3\mathbf{e}_1 - 2\mathbf{e}_2 + 4\mathbf{e}_3$.

4. By simple inspection, determine a vector that is orthogonal to each of the following vectors: $\mathbf{a} = (1, 0, 0, 1)$, $\mathbf{b} = (4, 3, 1, 0, 0)$, $\mathbf{c} = (0.543, 1.456, 1, 1)$.

It can be easily verified that any of the vectors $(0, \alpha, \beta, 0)$, $(-1, \alpha, \beta, 1)$, $(1, \alpha, \beta, -1)$, with α, β scalars, are orthogonal to \mathbf{a} . Analogously, $(0, 0, 0, \alpha, \beta)$ are orthogonal to \mathbf{b} , together with $(-1, 1, 1, \alpha, \beta)$, $(1, -1, -1, \alpha, \beta)$.

For \mathbf{c} , simple choices are $(0, 0, -1, 1)$ and $(0, 0, 1, -1)$.

2.5 Matrices

A matrix is an $n \times m$ rectangular array of scalars, real or complex numbers, with n rows and m columns. When $m = n$ the matrix is “square” and n is its dimension. In this book, we will use capital bold letters to indicate matrices, whereas roman small case letters in bold are used to denote vectors; Greek letters will commonly denote scalars. The following are examples of matrices of different dimensions,

$$\mathbf{A} = \begin{pmatrix} 0 & -1 & 4 \\ \frac{1}{2} & 2 & 1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 1 \\ i & 0 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 1 & 0 \\ 0 & 1+2i \\ 0.05 & -1 \\ 1.4+5i & 2 \\ 0 & 3 \end{pmatrix}. \quad (2.3)$$

Matrix \mathbf{A} is 2×3 , \mathbf{B} is 2×2 and \mathbf{C} is 5×2 . Note that \mathbf{B} and \mathbf{C} have complex entries. The components of a matrix \mathbf{A} are denoted by $a_{i,j}$, where i corresponds to the i th row and j to the j th column, that is at the (i, j) position in the array. In the following we shall use either parentheses or brackets to denote matrices. The $n \times m$ matrix with all zero entries is called the zero matrix. The square matrix with ones at the (i, i) entries, $i = 1, \dots, n$ and zero elsewhere, is called the identity matrix and is denoted by \mathbf{I} . If the order is not clear from the context, we shall use \mathbf{I}_n . The position of the scalars within the array is important: matrices with the same elements, but in a different order, are distinct matrices. Of particular interest is the transpose matrix, i.e. the matrix $b\mathbf{A}^T$ obtained by exchanging rows and columns of the matrix \mathbf{A} . For instance, for the matrices in (2.3),

$$\mathbf{A}^T = \begin{pmatrix} 0 & \frac{1}{2} \\ -1 & 2 \\ 4 & 1 \end{pmatrix}, \quad \mathbf{B}^T = \begin{pmatrix} 0 & i \\ 1 & 0 \end{pmatrix},$$

$$\mathbf{C}^T = \begin{pmatrix} 1 & 0 & 0.05 & 1.4 + 5i & 0 \\ 0 & 1 + 2i & -1 & 2 & 3 \end{pmatrix},$$

are the transpose matrices of the previous example. In the case of matrices with complex entries, we can also define the complex transposition, indicated by the subscript ‘*’, obtained by taking the complex conjugate of each element of the transpose, $\mathbf{B}^* := \bar{\mathbf{B}}^T$, so that

$$\mathbf{B}^* = \begin{pmatrix} 0 & -i \\ 1 & 0 \end{pmatrix}, \quad \mathbf{C}^* = \begin{pmatrix} 1 & 0 & 0.05 & 1.4 - 5i & 0 \\ 0 & 1 - 2i & -1 & 2 & 3 \end{pmatrix}.$$

Clearly, for real matrices the *Hermitian adjoint* \mathbf{B}^* coincides with the transpose matrix. Transposition and Hermitian adjoint share the reverse order law, i.e. $(\mathbf{AB})^* = \mathbf{B}^* \mathbf{A}^*$ and $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$, where \mathbf{A} and \mathbf{B} have conforming dimensions. See later for the definition of matrix-matrix products. Matrices that satisfy $\mathbf{A}^* \mathbf{A} = \mathbf{A} \mathbf{A}^*$ are called *normal*. A real square matrix \mathbf{X} such that $\mathbf{X}^T \mathbf{X} = \mathbf{I}$ and $\mathbf{X} \mathbf{X}^T = \mathbf{I}$ is said to be an *orthogonal matrix*. A square complex matrix \mathbf{X} such that $\mathbf{X}^* \mathbf{X} = \mathbf{I}$ and $\mathbf{X} \mathbf{X}^* = \mathbf{I}$ is said to be unitary.

An $n \times n$ matrix \mathbf{A} is *invertible* if there exists a matrix \mathbf{B} such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$. If such a matrix \mathbf{B} exists, it is unique, and it is called the *inverse* of \mathbf{A} , and it is denoted by \mathbf{A}^{-1} . An invertible matrix is also called nonsingular. Therefore, a singular matrix is a matrix that is not invertible. Recalling the definition of orthogonal matrices, we can immediately see that an orthogonal matrix is always invertible and more precisely, we have that its inverse coincides with its transpose, that is $\mathbf{X}^T = \mathbf{X}^{-1}$ (for a unitary matrix \mathbf{X} , it is $\mathbf{X}^* = \mathbf{X}^{-1}$). Matrices with special structures are given specific names. For instance,

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\mathbf{U} = \begin{pmatrix} 1 & 1 + 2i & 4 + 2i & 3 \\ 0 & 2 & 1 & 3 \\ 0 & 0 & 5 & 4i \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{L} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 - 2i & 2 & 0 & 0 \\ 4 - 2i & 1 & 5 & 0 \\ 3 & 3 & 4i & 1 \end{pmatrix}. \quad (2.4)$$

Note also that in this example, $\mathbf{L} = \mathbf{U}^*$. Matrices like \mathbf{D} are called diagonal (zero entries everywhere but on the main “diagonal”), whereas matrices like \mathbf{U} and \mathbf{L} are called upper triangular and lower triangular, respectively, since only the upper (resp. lower) part of the matrix is not identically zero. Note that it is very easy to check whether a diagonal matrix is invertible. Indeed, it can be easily verified that a diagonal matrix with diagonal components the inverses of the original diagonal entries is the sought after inverse. For the example above, we have $\mathbf{D}^{-1} = \text{diag}(1, \frac{1}{2}, \frac{1}{5}, 1)$ (we have used here a short-hand notation, with obvious meaning). Therefore, if all diagonal entries of \mathbf{D} are nonzero, \mathbf{D} is invertible, and vice versa. Similar considerations can be applied for (upper or lower) triangular matrices, which are nonsingular if and only if their diagonal elements are nonzero. Explicitly determining the inverse of less structured matrices is a much more difficult task. Fortunately, in most applications this problem can be circumvented.

If we look closely at the definition of a matrix we can see that there are several analogies to the definition of vectors we have used in the preceding sections. In fact we can think of each column as a vector, for instance the first column of the matrix \mathbf{C} is the vector $(1, 0, 0.05, 1.4 + 5i, 0)$ a vector of the four-dimensional vector space \mathbb{C}^4 . More generally, any row or column of \mathbf{C} can be viewed as a single vector. In the following we will need both kinds of vectors, but we will follow the convention that we will use the name “vector” for the column orientation, i.e. matrices with dimension $n \times 1$. Since a vector is just a skinny matrix, we can go from a column to a row vector via a transposition:

$$\mathbf{u} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \quad \mathbf{u}^T = (1 \ 2 \ 3).$$

From now on, the use of row vectors will be explicitly indicated by means of the transposition operation.

It can be shown that matrices are representations of linear transformations connecting vector spaces. In other words, an $m \times n$ matrix \mathbf{M} is an application that maps a vector \mathbf{u} of an n -dimensional vector space \mathcal{U} onto an element \mathbf{v} of an m -dimensional vector space \mathcal{V} , that is

$$\mathbf{v} = \mathbf{M}\mathbf{u}.$$

The vector space \mathcal{U} is known as the *domain* of \mathbf{M} and the vector space \mathcal{V} is called the *range*. Another important vector space associated with a matrix \mathbf{M} is the null space, i.e. the subset of the domain such that for all \mathbf{u} in this space, it holds $\mathbf{M}\mathbf{u} = \mathbf{0}$.

The product of a matrix $\mathbf{A} = (a_{i,j})$ on a vector is defined as another vector $\mathbf{v} = \mathbf{A}\mathbf{u}$ whose components are obtained by the row-by-column multiplication rule

$$v_i = \sum_{j=1}^m a_{i,j}u_j, \quad i = 1, \dots, n.$$

It can be noticed that the i th component of the resulting vector \mathbf{v} is the scalar product of the i th row of \mathbf{A} with the given vector \mathbf{u} . The matrix–vector product rule can be used to define a matrix–matrix multiplication rule by repeatedly applying the rule to each column of the second matrix, such that the element of the product matrix $\mathbf{C} = \mathbf{AB}$ of the $q \times m$ matrix \mathbf{A} with the $m \times p$ matrix \mathbf{B} is given by

$$c_{i,j} = \sum_{k=1}^m a_{i,k} b_{k,j}, \quad i = 1, \dots, q, \quad j = 1, \dots, p.$$

Note that the resulting matrix \mathbf{C} has dimension $q \times p$. For the product to be correctly defined, the number of columns of the first matrix, in this case \mathbf{A} , must be equal to the number of rows of the second matrix, \mathbf{B} . Note that this operation is not commutative, that is, in general $\mathbf{AB} \neq \mathbf{BA}$, even if both products are well defined. On the other hand, matrices that do satisfy the commutative property are said to commute with each other. Subsets of matrices that commute with each other have special properties that will appear in the following. A simple class of commuting matrices is given by the diagonal matrices: if \mathbf{D}_1 and \mathbf{D}_2 are diagonal matrices, then it can be verified that it always holds that $\mathbf{D}_1 \mathbf{D}_2 = \mathbf{D}_2 \mathbf{D}_1$.

Our last basic fact concerning matrices is related to the generalization to matrices of the vector notion of norm. In particular, we will use the *Frobenius norm*, which is natural generalization to matrices of the Euclidean vector norm. More precisely, given $\mathbf{A} \in \mathbb{R}^{n \times m}$,

$$\|\mathbf{A}\|_F^2 := \sum_{j=1}^m \sum_{i=1}^n a_{i,j}^2, \quad (2.5)$$

which can be equivalently written as $\|\mathbf{A}\|_F^2 = \sum_{j=1}^m \|\mathbf{a}_j\|^2$, where \mathbf{a}_j is the j th column of \mathbf{A} (a corresponding relation holds for the rows). In particular, it holds that $\|\mathbf{A}\|^2 = \text{trace}(\mathbf{A}^T \mathbf{A})$, where the trace of a matrix is the sum of its diagonal elements. The definition naturally generalizes to complex matrices. It is also interesting that the Frobenius norm is invariant under rotations, that is the norm remains unchanged whenever we multiply an orthonormal matrix by the given matrix. In other words, for any orthonormal matrix \mathbf{Q} it holds that $\|\mathbf{A}\|_F = \|\mathbf{QA}\|_F$.

Exercises and Problems

1. Given the matrices $\mathbf{A} = \begin{pmatrix} -1 & -3 \\ -1 & 2 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 0 & -1 \\ -3 & 1 \end{pmatrix}$, compute \mathbf{AB} .

We have

$$\mathbf{AB} = \begin{pmatrix} -1 & -3 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ -3 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} -1(0) - 3(-3) & -1(1) - 3(1) \\ -1(0) + 2(-3) & -1(-1) + 2(1) \end{pmatrix} = \begin{pmatrix} 9 & -4 \\ -6 & 3 \end{pmatrix}.$$

2. Given the matrix \mathbf{A} above, compute \mathbf{A}^T .

$$\text{We have } \mathbf{A}^T = \begin{pmatrix} -1 & -1 \\ -3 & 2 \end{pmatrix}.$$

3. Given the matrix \mathbf{A} above, and the row vector $\mathbf{x}^T = (-1, 5)$, compute $\mathbf{x}^T \mathbf{A}$.

$$\text{We have } \mathbf{x}^T \mathbf{A} = (-1, 5) \begin{pmatrix} -1 & -1 \\ -3 & 2 \end{pmatrix} = ((-1)(-1) + 5(-3), (-1)(-1) + 5(2)) = (-14, 11).$$

4. Compute \mathbf{xy}^T with $\mathbf{x}^T = (1, -3)$ and $\mathbf{y}^T = (2, -2)$.

The result of this computation is the 2×2 matrix given by

$$\mathbf{xy}^T = \begin{pmatrix} 1 \\ -3 \end{pmatrix} (2, -2) = \begin{pmatrix} 2 & -2 \\ -6 & 6 \end{pmatrix}.$$

(In the computation, the vectors \mathbf{x}, \mathbf{y}^T are viewed as 2×1 and 1×2 matrices, respectively)

2.6 Rank, Singularity and Inverses

The maximum number of columns or rows that are linearly independent in a matrix \mathbf{A} is called rank, denoted in the following by $\text{rank}(\mathbf{A})$. For a given $m \times n$ matrix \mathbf{A} , clearly $\text{rank}(\mathbf{A}) \leq \min\{n, m\}$. The rank can be used very efficiently to characterize the existence of the solution of a linear system of equations. In matrix terms, a linear system can be written as

$$\mathbf{Ax} = \mathbf{b}, \tag{2.6}$$

where \mathbf{x} represents the vector of the unknown variables, the entries of \mathbf{A} the system's coefficients, and the components of \mathbf{b} are the given right-hand sides of each equation. The system in (2.6) can either have no solution, one solution or infinite solutions. Let us write $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$, where $\mathbf{a}_i, i = 1, \dots, n$ are the columns of \mathbf{A} . By reading (2.6) backwards, we look for $\mathbf{x} = (x_1, \dots, x_n)^T$ such that $\mathbf{b} = \mathbf{Ax}$, that is, we seek the coefficients x_1, \dots, x_n , such that $\mathbf{b} = \mathbf{a}_1 x_1 + \dots + \mathbf{a}_n x_n$. In other words, the solution vector \mathbf{x} yields the coefficients that allow us to write \mathbf{b} as a linear combination of the columns of \mathbf{A} . At least one solution exists if $\text{rank}(\mathbf{A}) = \text{rank}((\mathbf{A}, \mathbf{b}))$, where (\mathbf{A}, \mathbf{b}) is the matrix obtained by adding the vector \mathbf{b} as a column besides \mathbf{A} . This corresponds to saying that the $n+1$ vectors $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n, \mathbf{b}\}$ are linearly dependent. The condition on the rank also shows that the existence of solutions to the system is related to the rank of the coefficient matrix \mathbf{A} . For square matrices, using the definition of inverse, $\mathbf{Ax} = \mathbf{b}$ is equivalent to $\mathbf{A}^{-1} \mathbf{Ax} = \mathbf{A}^{-1} \mathbf{b}$,

that is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$. Hence, assuming that \mathbf{b} is a nonzero vector, a unique solution \mathbf{x} exists if and only if \mathbf{A} is nonsingular. A crucial result of linear algebra is the following: an $n \times n$ (square) matrix \mathbf{A} is invertible if and only if $\text{rank}(\mathbf{A}) = n$. In particular, this result implies that if \mathbf{A} is singular, the columns of \mathbf{A} are linearly dependent ($\text{rank}(\mathbf{A}) < n$), or equivalently, there exists a vector \mathbf{x} , not identically zero, such that $\mathbf{A}\mathbf{x} = \mathbf{0}$. We have thus found that a singular matrix is characterized by a non-empty null space (cf. Sect. 3.5).

2.7 Decomposition of Matrices: Eigenvalues and Eigenvectors

A complex scalar λ and a nonzero complex vector \mathbf{x} are said to be an eigenvalue and an eigenvector of a square matrix \mathbf{A} , respectively, if they satisfy

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}. \quad (2.7)$$

A vector satisfying (2.7) has the special property that multiplication by \mathbf{A} does not change its direction, but only its length. In the case of Hermitian \mathbf{A} (i.e. $\mathbf{A} = \mathbf{A}^*$), it can be shown that such vectors arise in the problem of maximizing $(\mathbf{x}, \mathbf{A}\mathbf{x})$, over all vectors \mathbf{x} such that $\|\mathbf{x}\| = 1$. It is then found that the solution must satisfy the equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, where λ is a scalar. The pair (λ, \mathbf{x}) is called an eigenpair of \mathbf{A} . The set of all eigenvalues of \mathbf{A} is called the *spectrum* of \mathbf{A} . It is important to notice that eigenvectors are not uniquely determined. For instance, if \mathbf{x} is an eigenvector associated with λ , then $\alpha\mathbf{x}$ with $\alpha \neq 0$ is also an eigenvector associated with λ . Finally, we observe that if \mathbf{A} is singular, then there exists a vector \mathbf{x} such that $\mathbf{A}\mathbf{x} = \mathbf{0} = 0\mathbf{x}$, that is, $\lambda = 0$ is an eigenvalue of \mathbf{A} and \mathbf{x} is the corresponding eigenvector.

A fundamental result is that each square matrix \mathbf{A} of dimension n has exactly n complex eigenvalues, not necessarily all distinct. In case of multiple copies of the same eigenvalue, such a number of copies is called the multiplicity of that eigenvalue.¹ On the one hand, there can be *at most* n linearly independent eigenvectors. If an eigenvalue has multiplicity m larger than one, then there may be at most m linearly independent eigenvectors associated with that eigenvalue. On the other hand, eigenvectors corresponding to different eigenvalues are always linearly independent. Therefore, for a general matrix \mathbf{A} , the only case when there may not be a full set of independent eigenvectors is when there are multiple eigenvalues.

The case of Hermitian matrices is particularly fortunate, since in this case, there always exists a set of n linearly independent, and even mutually orthonormal, eigenvectors, regardless of the eigenvalue multiplicity. For a general square matrix \mathbf{A} , if there exist n linearly independent eigenvectors, \mathbf{A} can be written as

$$\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}, \quad (2.8)$$

¹ To be more precise, this number is the *algebraic* multiplicity of the eigenvalue.

where $\mathbf{\Lambda}$ is a diagonal matrix having the eigenvalues of \mathbf{A} as diagonal entries, while $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ is a matrix formed by normalized eigenvectors. The inverse of \mathbf{X} exists in this case because we are assuming that the eigenvectors are linearly independent, namely \mathbf{X} has rank n . If a form as in (2.8) can be written, we say that \mathbf{A} is diagonalizable. In the important case of Hermitian matrices, thanks to the orthogonality of the eigenvectors, we can write

$$\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^*$$

where \mathbf{X} is the matrix of the eigenvectors, normalized so as to have unit norm. Therefore, for Hermitian matrices, no inversion is required, as $\mathbf{X}^* = \mathbf{X}^{-1}$. If \mathbf{A} is real and symmetric, then the eigenpairs are real.

It can be shown that the eigenvalues can be found by solving the following scalar equation as a function of λ ,

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0, \quad (2.9)$$

whose left-hand side is a polynomial (the *characteristic polynomial*) of degree n in λ . Afterwards, the eigenvectors are obtained by solving the singular system

$$(\mathbf{A} - \lambda_i\mathbf{I})\mathbf{x}_i = 0, \quad i = 1, \dots, k,$$

where the index i runs over all k distinct eigenvalues found from solving (2.9). From the theory of polynomials, it follows that if \mathbf{A} is real, then its eigenvalues are real or, they appear as complex conjugates, that is, if λ is a complex eigenvalue of \mathbf{A} , then $\bar{\lambda}$ is also an eigenvalue of \mathbf{A} . Eigenvectors corresponding to real eigenvalues of a real matrix \mathbf{A} , can be taken to be real. Finally, Hermitian matrices have only real eigenvalues.

Nondiagonalizable matrices cannot be written in the form (2.8) with $\mathbf{\Lambda}$ diagonal. In particular, a nondiagonalizable matrix of dimension n does not have n linearly independent eigenvectors. This situation may only occur in the presence of multiple eigenvalues (see Exercises 4 and 5 at the end of this chapter).

The transformation indicated by (2.8) is an example of a class of transformations known as similarity transformations. Two matrices \mathbf{A} and \mathbf{B} are said to be similar if they can be obtained from each other by a similarity transformation via a nonsingular matrix \mathbf{S} , that is

$$\mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}. \quad (2.10)$$

Similar matrices share important properties, for instance, they have the same set of eigenvalues. The similarity transformation is equivalent to a change of basis in the representation of the matrix, in fact it can be shown that the transformation (2.10) is equivalent to changing the basis of the column vectors of the matrix \mathbf{B} , resulting in different coordinates.

2.8 The Singular Value Decomposition

Square as well as rectangular matrices can always be *diagonalized* if we allow the usage of two transformation matrices instead of one. Any $m \times n$ matrix \mathbf{A} with $m \geq n$, can be decomposed as:

$$\mathbf{A} = \mathbf{U} \begin{pmatrix} \boldsymbol{\Sigma} \\ 0 \end{pmatrix} \mathbf{V}^*, \quad (2.11)$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n]$ are square, unitary matrices of dimension m and n , respectively. Matrix $\boldsymbol{\Sigma}$ is diagonal and real, $\boldsymbol{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$, with $\sigma_{i+1} \leq \sigma_i$, $i = 1, \dots, n-1$, and $\sigma_i \geq 0$, $i = 1, \dots, n$. A completely analogous decomposition holds for $n \geq m$. The decomposition in (2.11) is called *singular value decomposition* (SVD); the columns of \mathbf{U} and \mathbf{V} are left and right singular vectors, respectively; the real numbers $\sigma_1, \sigma_2, \dots, \sigma_n$ are called singular values. The following relations can be derived,

$$\mathbf{A}\mathbf{A}^* = \mathbf{U}\boldsymbol{\Sigma}^2\mathbf{U}^*, \quad \mathbf{A}^*\mathbf{A} = \mathbf{V}\boldsymbol{\Sigma}^2\mathbf{V}^*,$$

indicating that the columns of the matrix \mathbf{U} are the eigenvectors of the matrix $\mathbf{A}\mathbf{A}^*$, while the columns of \mathbf{V} are the eigenvectors of the transpose matrix $\mathbf{A}^*\mathbf{A}$. Using the orthogonality of \mathbf{U} and \mathbf{V} in (2.11), we can write

$$\mathbf{A}\mathbf{V} = \mathbf{U} \begin{pmatrix} \boldsymbol{\Sigma} \\ 0 \end{pmatrix}, \quad \mathbf{A}^*\mathbf{U} = \mathbf{V}(\boldsymbol{\Sigma}, 0).$$

If \mathbf{A} is real, then all matrices have real entries. A series of very important results links the SVD with the determination of the rank of matrices. It can be shown that the rank, i.e. the number of linearly independent columns or rows in a matrix, is given by the number of non-zero singular values. The problem of finding the rank of a matrix can therefore be reduced to the problem of finding the number of nonzero singular values. Full rank square matrices of dimension n , have therefore exactly n strictly positive singular values. Comparing (2.8) with (2.11) we can see that the singular values decomposition extends the diagonalization property of the eigenvalues to more general matrices, including rectangular ones. The eigenvalue decomposition looks for a similarity transformation to a diagonal form, whereas in the singular value decomposition, we look for two, in general different, unitary transformations to a diagonal form.

We next briefly discuss the tight connection between the SVD and certain matrix norms that are *induced* by a vector norm. Let \mathbf{A} be an $m \times n$ matrix. Using the Euclidean norm we can define

$$\|\mathbf{A}\|_2 = \max_{\mathbf{x} \neq \mathbf{0}, \mathbf{x} \in \mathbb{C}^n} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}.$$

It can be shown that the vector \mathbf{x} that achieves this maximum is the first right singular vector, \mathbf{v}_1 , so that $\|\mathbf{A}\|_2 = \|\mathbf{A}\mathbf{v}_1\|_2 = \sigma_1$. The SVD allows us to also determine the matrix of low rank that is closest to the original matrix \mathbf{A} in the 2-norm. More precisely, let

$$\mathbf{A}_k = (\mathbf{u}_1, \dots, \mathbf{u}_k) \begin{pmatrix} \sigma_1 & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_k \end{pmatrix} (\mathbf{v}_1, \dots, \mathbf{v}_k)^*$$

be the matrix formed by the first k singular triplets. In other words, \mathbf{A}_k is the matrix obtained by a *truncated* SVD of rank k . Then it holds

$$\min_{\mathbf{B} \in \mathbb{C}^{m \times n}, \text{rank}(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_2 = \|\mathbf{A} - \mathbf{A}_k\|_2 = \sigma_{k+1}.$$

The relation above says that \mathbf{A}_k is the rank- k matrix that is closest to \mathbf{A} when using the 2-norm. Moreover, it provides an explicit value for the error of such approximation, which is given by the first neglected singular value, σ_{k+1} .

The SVD can also be employed for computing the Frobenius norm of a matrix; see (2.5). Indeed, it holds that

$$\|\mathbf{A}\|_F^2 = \sum_{j=1}^{\min\{n,m\}} \sigma_j^2,$$

where σ_j 's are the singular values of the $n \times m$ matrix \mathbf{A} .

The singular value decomposition provides a formidable tool to replace the inverse of a singular or rectangular matrix. Assume that an $m \times n$ matrix \mathbf{A} with $m \geq n$ is decomposed as in (2.11), where $\mathbf{\Sigma}$ is nonsingular. Then the *Penrose* pseudo-inverse of \mathbf{A} (cf., e.g., [Golub and Van Loan 1996](#)) is defined as²

$$\mathbf{A}^- := \mathbf{V} \begin{pmatrix} \mathbf{\Sigma}^{-1} & 0 \end{pmatrix} \mathbf{U}^*. \quad (2.12)$$

Note that \mathbf{V} and \mathbf{U} are unitary, so that

$$\mathbf{A}\mathbf{A}^- = \mathbf{U} \begin{pmatrix} \mathbf{I} & 0 \\ 0 & 0 \end{pmatrix} \mathbf{U}^*.$$

Note that in general, $\mathbf{A}\mathbf{A}^- \neq \mathbf{I}$, unless \mathbf{A} is square and nonsingular.

The definition above can be generalized to any singular square matrix.

Finally, we make a simple connection between eigenvalues, singular values and singularity of a square matrix. Using the SVD of a given matrix \mathbf{A} , we can say that \mathbf{A}

² Common notations for the pseudo-inverse also include \mathbf{A}^\dagger and \mathbf{A}^+ .

is nonsingular if and only if Σ has nonzero diagonal elements, indeed \mathbf{A}^{-1} exists if and only if Σ^{-1} exists. A similar consideration holds with respect to the eigenvalue decomposition.

2.9 Functions of Matrices

It is possible to define functions of matrices in analogy to the familiar function on the real and complex numbers; see, e.g., [Horn and Johnson \(1991\)](#) for a more detailed treatment of this topic. For a given square matrix \mathbf{A} , a matrix polynomial of degree k is defined as

$$p(\mathbf{A}) = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{A} + \alpha_2 \mathbf{A}^2 + \dots + \alpha_k \mathbf{A}^k, \quad (2.13)$$

where the scalar coefficients $\alpha_0, \dots, \alpha_k$ can be real or complex. The polynomial $p(\mathbf{A})$ is a matrix and there is no ambiguity in its construction, as long as matrix powers are carried out with the matrix product rule. If \mathbf{A} is a diagonal matrix, that is $\mathbf{A} = \text{diag}(a_{1,1}, \dots, a_{n,n})$, then it can be easily verified that $p(\mathbf{A}) = \text{diag}(p(a_{1,1}), \dots, p(a_{n,n}))$, that is, the polynomial is applied to the single diagonal entries (cf. Exercise 6). We stress that this is only true for diagonal matrices, when their dimension is greater than one. If \mathbf{A} is diagonalizable, that is $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}$, then it is possible to write

$$p(\mathbf{A}) = p(\mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}) = \mathbf{X} \begin{bmatrix} p(\lambda_1) & \dots & 0 \\ \dots & \ddots & 0 \\ 0 & \dots & p(\lambda_n) \end{bmatrix} \mathbf{X}^{-1} = \mathbf{X}p(\mathbf{\Lambda})\mathbf{X}^{-1},$$

where we have used the property that $p(\mathbf{X}\mathbf{A}\mathbf{X}^{-1}) = \mathbf{X}p(\mathbf{A})\mathbf{X}^{-1}$ (this can be easily deduced first for \mathbf{A}^k , for any $k > 0$, and then for $p(\mathbf{A})$ using (2.13); see also Exercise 6). The calculation is rather interesting if we replace the polynomial p with a more general function f , such as $\exp(x)$, $\ln(x)$, \sqrt{x} , etc. Assume that f is a smooth function at the eigenvalues of \mathbf{A} . Then, as before, for diagonalizable \mathbf{A} we can write

$$f(\mathbf{A}) = f(\mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}) = \mathbf{X} \begin{bmatrix} f(\lambda_1) & \dots & 0 \\ \dots & \ddots & 0 \\ 0 & \dots & f(\lambda_n) \end{bmatrix} \mathbf{X}^{-1} = \mathbf{X}f(\mathbf{\Lambda})\mathbf{X}^{-1}.$$

In general the definition of a function of a matrix can be made rigorous without resorting to the diagonalization of \mathbf{A} , so that the matrix is not needed to be diagonalizable. We will assume that the function and the matrix we will use are all sufficiently well-behaved that the above definition can be used without special care.

Exercises and Problems

1. Given the matrix $\mathbf{A} = \begin{pmatrix} 2 & -4 \\ 1 & 0 \end{pmatrix}$, and the vector $\mathbf{b}^T = (-2, 1)$, verify that the vector $\mathbf{x}^T = [1, 1]$ is the (unique) solution to the system $\mathbf{A}\mathbf{x} = \mathbf{b}$.
We need to check that the definition is satisfied. Indeed, we have

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} 2(1) - 4(1) \\ 1(1) + 0(1) \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \mathbf{b}.$$

Note that \mathbf{A} is nonsingular, since the first row of the matrix is not a multiple of the second row (this is a sufficient consideration only in \mathbb{R}^2). Therefore the system solution is unique.

2. Given the matrix $\mathbf{A} = \begin{pmatrix} 3 & -4 \\ -1 & 1 \end{pmatrix}$, verify that $\mathbf{x}^T = (-1 - \sqrt{5}, 1)$ and $\lambda = 2 + \sqrt{5}$ are respectively an eigenvector and the associated eigenvalue of \mathbf{A} .
We need to check that the definition is satisfied. Indeed, we have

$$\mathbf{A}\mathbf{x} = \begin{pmatrix} 3(-1 - \sqrt{5}) - 4(1) \\ -1(-1 - \sqrt{5}) + 1 \end{pmatrix} = \begin{pmatrix} -7 - 3\sqrt{5} \\ 2 + \sqrt{5} \end{pmatrix} \text{ and } \lambda\mathbf{x} = \begin{pmatrix} -7 - 3\sqrt{5} \\ 2 + \sqrt{5} \end{pmatrix}.$$

3. Show that the eigenvalues of an $n \times n$ real triangular matrix \mathbf{A} coincide with its diagonal entries.

This can be checked by explicitly writing $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. Indeed, we have $\det(\mathbf{A} - \lambda\mathbf{I}) = (\lambda - a_{1,1})(\lambda - a_{2,2}) \cdots (\lambda - a_{n,n}) = 0$, which is satisfied for $\lambda = a_{i,i}$, $i = 1, \dots, n$.

4. Show that the matrix $\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$ only has one linearly independent eigenvector.

The matrix is triangular, therefore the eigenvalues are the diagonal elements (see exercise above). Hence, $\lambda_1 = \lambda_2 = 2$. Using the definition $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, eigenvectors of \mathbf{A} are obtained by solving the singular system $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ with $\lambda = 2$. We have

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

whose solution is $\mathbf{x} = (x_1, 0)^T$, $x_1 \in \mathbb{R}$. No other linearly independent solutions exist.

5. Show that the matrix $\mathbf{A} = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ has two linearly independent eigenvectors.

Proceeding as above, one finds that $\lambda_1 = \lambda_2 = \lambda_3 = 2$, and there are two linearly independent eigenvectors, $\mathbf{x} = (x_1, 0, 0)^T$ and $\mathbf{y} = (0, 0, y_3)^T$, $x_1, y_3 \in \mathbb{R}$.

6. Show that if \mathbf{A} is diagonal, $\mathbf{A} = \text{diag}(a_{1,1}, \dots, a_{n,n})$, then $p(\mathbf{A}) = \text{diag}(p(a_{1,1}), \dots, p(a_{n,n}))$ for any polynomial p .

The result follows from observing that for any $k \geq 0$, $\mathbf{A}^k = \text{diag}(a_{1,1}^k, \dots, a_{n,n}^k)$.

7. Given a square diagonalizable matrix $\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}$, show that $p(\mathbf{A}) = \mathbf{X}p(\mathbf{\Lambda})\mathbf{X}^{-1}$.

We write $p(\mathbf{A}) = \alpha_0\mathbf{I} + \alpha_1\mathbf{A} + \dots + \alpha_k\mathbf{A}^k$. We have $\mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1}\mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1} = \mathbf{X}\mathbf{\Lambda}^2\mathbf{X}^{-1}$. This in fact holds for any j , that is $\mathbf{A}^j = \mathbf{X}\mathbf{\Lambda}^j\mathbf{X}^{-1}$. Therefore,

$$p(\mathbf{A}) = \alpha_0\mathbf{X}\mathbf{X}^{-1} + \alpha_1\mathbf{X}\mathbf{\Lambda}\mathbf{X}^{-1} + \dots + \alpha_k\mathbf{X}\mathbf{\Lambda}^k\mathbf{X}^{-1} = \mathbf{X}p(\mathbf{\Lambda})\mathbf{X}^{-1},$$

where in the last equality the matrices \mathbf{X} and \mathbf{X}^{-1} have been collected on both sides.



<http://www.springer.com/978-90-481-3701-5>

A Guide to Empirical Orthogonal Functions for Climate
Data Analysis

Navarra, A.; Simoncini, V.

2010, VI, 151 p. With online files/update., Hardcover

ISBN: 978-90-481-3701-5