A

Linear algebra

In this appendix we review the main concepts of linear algebra. We stress the geometrical interpretation wherever possible and we do not shun loose expressions in order to appeal to intuition. For a thorough introduction to linear algebra the reader is referred to references such as Lang (1997).

A.1 Vector space

The natural environment of linear algebra are finite-dimensional vector spaces. A *vector space* is a set on whose elements we can perform certain operations. In practice, we focus our attention on the Euclidean space $\mathbb{R}^N$. We can represent geometrically the Euclidean space $\mathbb{R}^N$ as the space generated by N axes, as in the left portion of Figure A.1

A vector in $\mathbb{R}^N$ can be represented as a column of N real numbers

$$\mathbf{v} \equiv (v_1, \ldots, v_N)' , \tag{A.1}$$

where the symbol $'$ denotes transposition. Geometrically, it is natural to represent a vector as an arrow whose tail sits on the origin of the N axes that generate the space and whose tip is the N-tuple $(A.1)$.

Alternatively, it is useful to think of an analytical representation of a vector as a function that with each of the first N integers associates a real number, the "entry" on the respective axis:

$$\mathbf{v} : \; n \in \{1, \ldots, N\} \to v_n \in \mathbb{R}. \tag{A.2}$$

Refer again to Figure A.1 for an interpretation.

The set of such vectors is a vector space, since the following operations are properly defined on its elements.

The *sum* of two vectors is defined component-wise as follows:

$$[\mathbf{u} + \mathbf{v}]_n \equiv u_n + v_n. \tag{A.3}$$

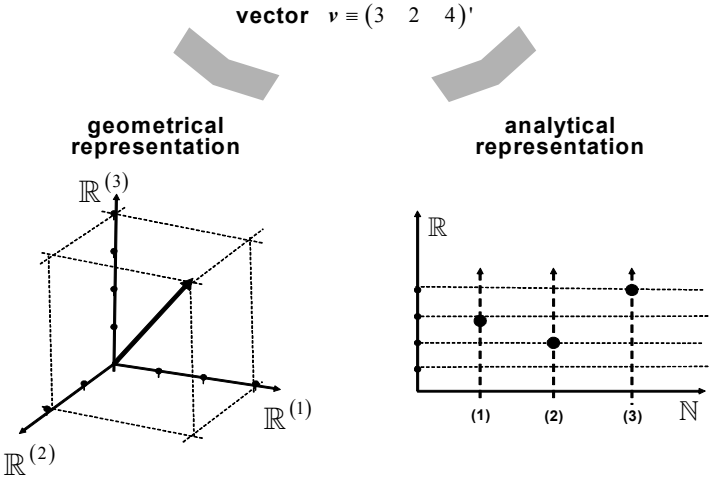

Fig. A.1. Representations of a vector

This is the *parallelogram rule*: the sum of two arrows stemming from the origin is the diagonal of the parallelogram spanned by the arrows.

The *multiplication by a scalar* is defined component-wise as follows:

$$[\alpha \mathbf{v}]_n \equiv \alpha v_n. \tag{A.4}$$

This is a stretch by a factor α in the direction of $\mathbf{v}$.

Combining sums and multiplications by a scalar we obtain *linear combinations* of vectors.

All possible linear combinations of an arbitrary set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_K\}$ in a vector space generates a *vector subspace* of that vector space. To visualize a subspace, consider the parallelotope described by the vertices of a set of vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_K\}$. The subspace generated by these vectors is the parallelotope obtained by stretching all the vertices to plus and minus infinity.

Vectors are *linearly independent* if the parallelotope they generate is non-degenerate. We see in Figure A.2 the case of three vectors, respectively linearly independent and linearly dependent.

The last important feature of the Euclidean space $\mathbb{R}^N$ is the existence of an *inner product*, an operation that allows to define useful concepts such as orthogonality and length. The inner product is defined as the sum of the entry-by-entry multiplication of two vectors:

$$\langle \mathbf{u}, \mathbf{v} \rangle \equiv \sum_{n=1}^{N} u_n v_n. \tag{A.5}$$

By means of the inner product we can define the length of a vector in $\mathbb{R}^N$, also called the *norm*:

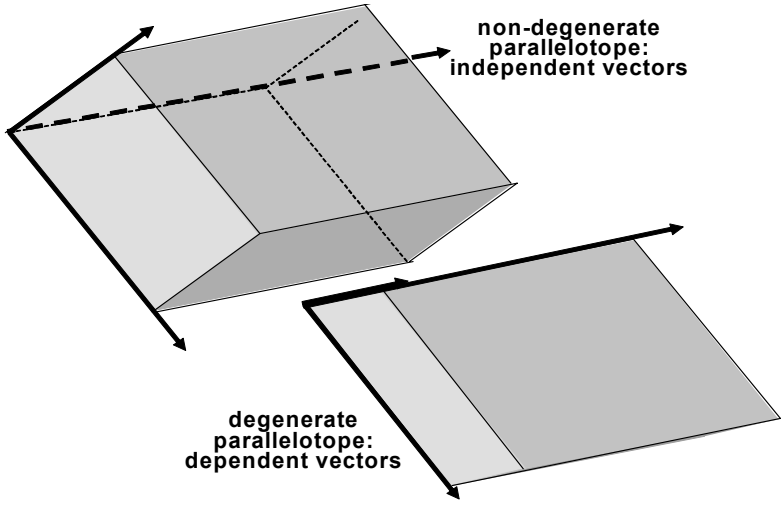

Fig. A.2. Linear (in)dependence among vectors

$$\|\mathbf{v}\| \equiv \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}. \tag{A.6}$$

The reader will recognize that the norm is indeed the length, as its definition can be interpreted in geometric terms as the *Pythagorean theorem*. Furthermore, the norm displays the following intuitive properties of a length:

$$\begin{aligned} \|\mathbf{v}\| &\geq 0 \\ \|\mathbf{v}\| = 0 &\Leftrightarrow \mathbf{v} = 0 \\ \|\alpha \mathbf{v}\| &= |\alpha| \, \|\mathbf{v}\| \\ \|\mathbf{u} + \mathbf{v}\| &\leq \|\mathbf{u}\| + \|\mathbf{v}\| \, . \end{aligned} \tag{A.7}$$

The last property is called *triangular inequality* and follows from the *Cauchy-Schwartz inequality*:

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \, \|\mathbf{v}\| \, , \tag{A.8}$$

in which the equality holds if and only if $\mathbf{u} \equiv \alpha \mathbf{v}$ for some scalar α. If the scalar α is positive:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \|\mathbf{u}\| \, \|\mathbf{v}\| \, ; \tag{A.9}$$

if the scalar α is negative:

$$\langle \mathbf{u}, \mathbf{v} \rangle = - \|\mathbf{u}\| \, \|\mathbf{v}\| \, . \tag{A.10}$$

We omit the (easy) proof.

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* if their inner product is null:

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0. \tag{A.11}$$

The *projection* of a vector $\mathbf{u}$ on a subspace V is the vector of that subspace that is closest to $\mathbf{u}$:

$$P(\mathbf{u}, V) \equiv \underset{\mathbf{v} \in V}{\operatorname{argmin}} \|\mathbf{u} - \mathbf{v}\|. \tag{A.12}$$

It is possible to check that if two vectors are orthogonal the projection of either one on the subspace generated by the other is zero: geometrically, this means that the two vectors are perpendicular. Therefore orthogonal vectors are linearly independent, since the parallelotope they generate is not skewed, and thus non-degenerate, see Figure A.2.

A.2 Basis

A *basis* for a vector space is a set of linearly independent elements of that space that can generate all the other vectors by means of linear combinations. The number of these elements is the *dimension* of that vector space. In the case of the Euclidean space $\mathbb{R}^N$, this number is N. Therefore, a basis is a set of vectors

$$\mathbf{e}^{(n)}, \quad n = 1, \ldots, N, \tag{A.13}$$

such that, for suitable scalars $\alpha_1, \ldots, \alpha_N$, any vector $\mathbf{v}$ of $\mathbb{R}^N$ can be expressed as a linear combination:

$$\mathbf{v} = \sum_{n=1}^{N} \alpha_n \mathbf{e}^{(n)}. \tag{A.14}$$

The *canonical basis* is the following set of vectors:

$$\begin{aligned} \boldsymbol{\delta}^{(1)} &\equiv (1, 0, \ldots, 0)' \\ &\vdots \\ \boldsymbol{\delta}^{(N)} &\equiv (0, 0, \ldots, 1)'. \end{aligned} \tag{A.15}$$

It is possible to check that the canonical basis is the only set of vectors such that the inner product of one of them, say $\boldsymbol{\delta}^{(n)}$, with a generic vector $\mathbf{v}$ in $\mathbb{R}^N$ yields the n-th entry of that vector:

$$\left\langle \mathbf{v}, \boldsymbol{\delta}^{(n)} \right\rangle = v_n. \tag{A.16}$$

The generic element $\boldsymbol{\delta}^{(n)}$ of this basis is called the *Kronecker delta centered in n*. This name stems from the analytical representation of the vector $\boldsymbol{\delta}^{(n)}$ as in the right portion of Figure A.1, which is a function peaked on the integer n.

A.3 Linear transformations

Consider a function A that maps vectors $\mathbf{v}$ of the Euclidean space $\mathbb{R}^N$ into vectors that belong to the same Euclidean space $\mathbb{R}^N$, or to another Euclidean space $\mathbb{R}^M$:

$$A: \ \mathbf{v} \in \mathbb{R}^N \mapsto \mathbf{u} \equiv A[\mathbf{v}] \in \mathbb{R}^M. \tag{A.17}$$

The function A is a *linear transformation*, or a *linear application*, if it preserves the sum and the multiplication by a scalar:

$$\begin{aligned} A[\mathbf{u}+\mathbf{v}] &= A[\mathbf{u}] + A[\mathbf{v}] \\ A[\alpha\mathbf{v}] &= \alpha A[\mathbf{v}]. \end{aligned} \tag{A.18}$$

In Figure A.3 we sketch the graphical meaning of a linear application. Consider the parallelotope $\mathfrak{P}$ described by the vertices of a set of K vectors $\{\mathbf{v}_1, \ldots, \mathbf{v}_K\}$. Now consider the parallelotope $\mathfrak{P}'$ described by the vertices of the set of vectors $\{A[\mathbf{v}_1], \ldots, A[\mathbf{v}_K]\}$. A transformation A is linear if $A(\mathfrak{P}) = \mathfrak{P}'$, i.e. if parallelotopes are mapped into parallelotopes: it is called a linear application because it does not bend straight lines. This interpretation

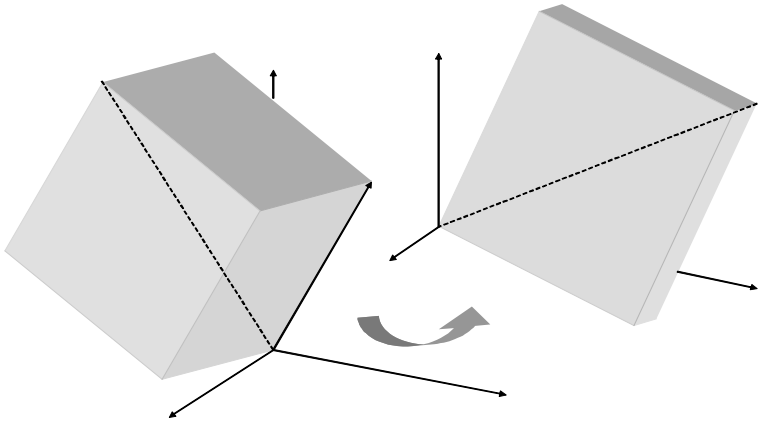

Fig. A.3. Geometrical representation of a linear transformation

makes it immediate to see that a sequence of two linear applications

$$(B \circ A)[\mathbf{v}] \equiv B[A[\mathbf{v}]] \tag{A.19}$$

is a linear application.

The inverse A^{-1} of a linear transformation A is the transformation that applied either before or after the linear transformation A cancels the effect of the transformation A. In other words, for all vectors $\mathbf{v}$ the inverse transformation A^{-1} satisfies:

$$\left(A^{-1} \circ A\right)[\mathbf{v}] = \mathbf{v} = \left(A \circ A^{-1}\right)[\mathbf{v}]. \tag{A.20}$$

The inverse of a linear application is not always defined: if a linear transformation A "squeezes" a parallelotope into a degenerate parallelotope it is not possible to recover univocally the vectors that generated the original parallelotope. In this case the dimension of the image space $A\left[\mathbb{R}^N\right]$ is less than the dimension N of the original space.

The dimension of the image space is called the *rank* of the application A:

$$\text{rank}(A) \equiv \dim\left(A\left[\mathbb{R}^N\right]\right). \tag{A.21}$$

Since a linear application can either squeeze a vector space or preserve its dimension, it follows from the definition (A.21) of rank that:

$$\text{rank}(B \circ A) \leq \min\left(\text{rank}(A), \text{rank}(B)\right). \tag{A.22}$$

A linear transformation is invertible if it is full-rank, i.e. if its rank is equal to the dimension of the original vector space. Therefore, a linear transformation is full-rank if it maps a basis into another basis.

If a linear transformation is full-rank, the inverse transformation A^{-1} exists and it is also a linear application, since in turn it maps parallelotopes in parallelotopes.

A.3.1 Matrix representation

Just like vectors can be identified with N-tuples of numbers as in (A.1), linear transformations can be identified with $M \times N$ matrices. Indeed, consider a generic linear transformation (A.17). A Taylor expansion around zero of the generic entry u_m as a function of the entries of $\mathbf{v}$ reads:

$$u_m = A_m + \sum_{n=1}^{N} A_{mn} v_n + \sum_{n,l=1}^{N} A_{mnl} v_n v_l + \cdots, \tag{A.23}$$

where $A_{\ldots}$ are suitable constant coefficients. In order for (A.18) to hold only the coefficients A_{mn} in the second term can contain non-zero elements. Collecting these terms in a matrix $\mathbf{A}$ we can represent the linear transformation (A.17) by means of its *matrix representation* as follows:

$$\mathbf{u} \equiv A[\mathbf{v}] \equiv \mathbf{A}\mathbf{v}, \tag{A.24}$$

where the product of a matrix by a vector is defined as:

$$[\mathbf{A}\mathbf{v}]_m \equiv \sum_{n=1}^{N} A_{mn} v_n. \tag{A.25}$$

For example, consider the *identity transformation* defined as follows:

$$I[\mathbf{v}] \equiv \mathbf{v}. \tag{A.26}$$

It is immediate to check that the identity transformation is represented by the *identity matrix*, defined as follows:

$$\mathbf{I}_N \equiv \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}. \tag{A.27}$$

From $(A.25)$ we also derive the "row-by-column" multiplication rule for matrices. Indeed, it is easy to check that the matrix representation $\mathbf{C}$ of the transformation $C \equiv B \circ A$ defined in $(A.19)$ reads:

$$C_{mn} = \sum_{l} B_{ml} A_{ln}. \tag{A.28}$$

Notice that a matrix can be seen as a function from the two-dimensional grid of integer coordinates to the real numbers:

$$\mathbf{A}: \ (m,n) \in \{1,\ldots,M\} \times \{1,\ldots,N\} \to A_{mn} \in \mathbb{R}. \tag{A.29}$$

This definition parallels the analytical definition $(A.2)$ of a vector.

A.3.2 Rotations

Rotations are special kinds of linear transformations. As intuition suggests, a linear transformation R is a *rotation* in the Euclidean space $\mathbb{R}^N$ if it does not alter the length[1], i.e. the norm $(A.6)$, of any vector in $\mathbb{R}^N$:

$$\|R[\mathbf{v}]\| = \|\mathbf{v}\|. \tag{A.30}$$

A rotation is always invertible, since it does not "squeeze" parallelotopes and therefore it does not make them degenerate. Moreover, the inverse of a rotation is a rotation.

From the definition of rotation $(A.30)$, the definition of norm $(A.6)$, the rule for the representation of the composition of two linear applications $(A.28)$

[1] More precisely, this is the definition of *isometries*, which include rotations, reflections and inversions.

and the representation of the identity ($A.27$), it is easy to derive the following result for the matrix representation $\mathbf{R}$ of the rotation R:

$$\mathbf{R}^{-1} = \mathbf{R}'. \tag{A.31}$$

In words, a linear transformation R is a rotation if and only if the representation of its inverse is the transpose of its representation.

For example, for any θ the matrix

$$\mathbf{R}_\theta \equiv \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{A.32}$$

satisfies ($A.31$), and thus it represents a rotation in $\mathbb{R}^2$. Indeed, it represents a counterclockwise rotation of an angle θ: this can be easily verified by checking the result of applying ($A.32$) to the two vectors of the canonical basis ($A.15$).

Furthermore, it can be proved that any rotation in $\mathbb{R}^2$ can be represented by a matrix of the form ($A.32$) for a suitable angle θ.

A.4 Invariants

Consider a generic linear transformation A from $\mathbb{R}^N$ to itself. Consider now another transformation $\widetilde{A}$ obtained in terms of the composition ($A.19$) with an invertible transformation B as follows:

$$\widetilde{A} \equiv B^{-1} \circ A \circ B. \tag{A.33}$$

We call the transformations A and $\widetilde{A}$ *equivalent linear transformations*. Indeed, $\widetilde{A}$ brings the original reference frame into an equivalent one by means of the invertible transformation B, then performs the same operation as A and finally brings the result back to the original reference frame by means of the inverse transformation B^{-1}.

Two equivalent transformations A and $\widetilde{A}$ must share many properties. Nevertheless, their matrix representations $\mathbf{A}$ and $\widetilde{\mathbf{A}}$ might be very different. Therefore, it can be hard to detect equivalent transformations from their representations. In this section we describe some features that are common to any representation of equivalent transformations.

A.4.1 Determinant

Consider the parallelotope $\mathfrak{P}$ described by the vertices of a set of independent vectors. We recall that the linear transformation A by definition maps this parallelotope into another parallelotope $\mathfrak{P}'$, see Figure A.3. In so doing, A stretches and turns $\mathfrak{P}$ and therefore modifies its volume by some factor. This

factor does not depend on the particular choice of $\mathfrak{P}$: the linearity of A implies that the volume of any parallelotope is modified by the same factor. We call this factor, modulo a sign, the determinant. In other words, the *determinant* of the transformation A is the number $\det(A)$ such that

$$\text{Vol}(\mathfrak{P}') = \pm \det(A) \text{Vol}(\mathfrak{P}), \tag{A.34}$$

where "Vol" denotes the volume and the sign is positive (negative) if the transformation includes an even (odd) number of reflections.

In particular, the transformation A is not invertible if and only if $\mathfrak{P}'$ is degenerate, i.e. if its volume is zero. Therefore, a transformation A is not invertible if and only if

$$\det(A) = 0. \tag{A.35}$$

Furthermore, we see that for the composite transformation $(A.19)$ the following rule holds:

$$\det(B \circ A) = \det(B) \det(A). \tag{A.36}$$

In particular, since the identity transformation $(A.26)$ does not alter the volumes:

$$1 = \det\left(B \circ B^{-1}\right) = \det(B) \det\left(B^{-1}\right). \tag{A.37}$$

Now we can prove that the determinant is indeed an invariant. If a linear transformation $\widetilde{A}$ is equivalent to a linear transformation A as in $(A.33)$, then:

$$\det\left(\widetilde{A}\right) = \det\left(B^{-1}\right) \det(A) \det(B) = \det(A). \tag{A.38}$$

It can be proved that the formula to compute explicitly the determinant of a linear transformation A in terms of its matrix representation $\mathbf{A}$ reads:

$$\det(A) \equiv |\mathbf{A}| = \sum_{\{i_1,\ldots,i_N\} \in \mathcal{P}} \pm A_{i_1 1} \cdots A_{i_N N}, \tag{A.39}$$

where the sum is taken over all the permutations $\mathcal{P}$ of the first N integers and the sign is positive for even permutations (i.e. obtained by a sequence of an even number of switches) and negative for odd permutations.

For example, the formula for the determinant of a generic 2×2 matrix

$$\mathbf{A} \equiv \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \tag{A.40}$$

is

$$|\mathbf{A}| = A_{11}A_{22} - A_{21}A_{12}. \tag{A.41}$$

In one situation the determinant is particularly easy to compute. Consider a diagonal matrix $\mathbf{D}$, i.e. a matrix where all elements D_{mn} for $m \neq n$ are zero. Geometrically, a diagonal matrix $\mathbf{D}$ represents a stretch by a factor D_{nn}

along the generic n-th axis. In this situation a parallelotope is stretched into a new parallelotope whose volume is multiplied by $D_{11} \cdots D_{NN}$. Therefore the determinant in this case reads

$$|\mathbf{D}| = \prod_{n=1}^{N} D_{nn}, \tag{A.42}$$

i.e., the determinant is the product of the diagonal elements. Notice that $(A.42)$ automatically accounts for the change in sign due to reflections, since a reflection is associated with a negative entry on the diagonal.

Since the determinant is an invariant, the result is the same for any equivalent representation $\widetilde{\mathbf{A}}$. Therefore, it is particularly convenient to find, if possible, equivalent representations $\underset{\sim}{\mathbf{A}}$ of a generic linear transformation A that are diagonal.

A.4.2 Trace

The *trace* of a generic linear transformation A from $\mathbb{R}^N$ to itself is defined in terms of its matrix representation $\mathbf{A}$ as the sum of the diagonal entries:

$$\mathrm{tr}\,(A) \equiv \mathrm{tr}\,(\mathbf{A}) \equiv \sum_{n=1}^{N} A_{nn}. \tag{A.43}$$

From this definition and the multiplication rule $(A.28)$ we obtain the *circular property* of the trace:

$$\mathrm{tr}\,(A \circ B \circ C) = \mathrm{tr}\,(B \circ C \circ A)\,. \tag{A.44}$$

Consider now two equivalent linear transformations A and $\widetilde{A}$ as in $(A.33)$. Then the following result holds:

$$\mathrm{tr}\left(\widetilde{A}\right) = \mathrm{tr}\left(B^{-1} \circ A \circ B\right) = \mathrm{tr}\left(B \circ B^{-1} \circ A\right) = \mathrm{tr}\,(A)\,. \tag{A.45}$$

This proves that the trace is indeed an invariant.

A.4.3 Eigenvalues

An *eigenvector* of a linear transformation A from $\mathbb{R}^N$ to $\mathbb{R}^N$ is a vector $\mathbf{v}$ that is not rotated by the transformation, i.e. such that for a suitable scalar λ the following holds:

$$A\,[\mathbf{v}] = \lambda \mathbf{v}. \tag{A.46}$$

The number λ is called the *eigenvalue* relative to the eigenvector $\mathbf{v}$. Notice that if $\mathbf{v}$ is an eigenvector of A, so is any multiple $\alpha\mathbf{v}$. In general, a linear transformation A does not admit eigenvalues. Nevertheless, if some eigenvalues exist, it becomes much easier to analyze the properties of A.

If they exist, eigenvalues are invariants (beware: eigenvectors are *not* invariants). Indeed, if there exists a pair $(\lambda, \mathbf{v})$ that satisfies (A.46) then, for any equivalent transformation $\widetilde{A}$ as in (A.33) we can see that $\mathbf{w} \equiv \mathbf{B}^{-1}\mathbf{v}$ is an eigenvector for the same eigenvalue:

$$\widetilde{A}[\mathbf{w}] = \mathbf{B}^{-1}\mathbf{A}\mathbf{B}\mathbf{w} = \mathbf{B}^{-1}\mathbf{A}\mathbf{v} = \lambda\mathbf{B}^{-1}\mathbf{v} = \lambda\mathbf{w}. \tag{A.47}$$

In order to compute the eigenvalues of A, or to realize that they do not exist, notice from the definition (A.46) that λ is an eigenvalue if and only if the linear application $A - \lambda I$, where I is the identity (A.26), "squeezes" a specific direction, i.e. the direction spanned by the eigenvector $\mathbf{v}$, into the zero vector. This can happen only if $A - \lambda I$ is not invertible. Therefore, from (A.35) an eigenvalue λ solves the equation

$$\det(A - \lambda I) = 0. \tag{A.48}$$

In general, this equation does not necessarily admit real solutions.

For example, consider a generic 2×2 matrix (A.40). Making use of (A.41) it is easy to check that (A.48) becomes:

$$0 = \lambda^2 - \lambda \operatorname{tr}(A) + \det(A). \tag{A.49}$$

The possible solutions read:

$$\lambda = \frac{1}{2}\left(\operatorname{tr}(A) \pm \sqrt{\operatorname{tr}(A)^2 - 4\det(A)}\right). \tag{A.50}$$

This shows that if $\operatorname{tr}(A)^2 < 4\det(A)$ there is no solution. Otherwise, the two solutions are invariants, as they only depend on trace and determinant, which are invariants.

A.5 Spectral theorem

The spectral theorem is an extremely useful result whose interpretation and application involve all the invariants described in Section A.4.

A.5.1 Analytical result

In general a linear transformation does not admit eigenvectors and eigenvalues. Nevertheless, in a special, yet very important, case it is possible to find a whole basis of orthogonal eigenvectors.

First we need two definitions. A linear application S is *symmetric* if its matrix representation is symmetric with respect to the diagonal, i.e. it is equal to its transpose:

$$\mathbf{S} = \mathbf{S}'. \tag{A.51}$$

A linear application S is *positive* if for any $\mathbf{v} \in \mathbb{R}^N$ its matrix representation satisfies the following inequality:[2]

$$\langle \mathbf{v}, \mathbf{S}\mathbf{v} \rangle \geq 0. \tag{A.52}$$

We stress that a positive matrix can have negative entries.

The *spectral theorem* states that a symmetric matrix admits an orthogonal basis of eigenvectors. In other words, if a square matrix $\mathbf{S}$ satisfies $(A.51)$, then there exist N numbers $(\lambda_1, \ldots, \lambda_N)$ and N vectors $\left(\mathbf{e}^{(1)}, \ldots, \mathbf{e}^{(N)}\right)$ such that

$$\mathbf{S}\mathbf{e}^{(n)} = \lambda_n \mathbf{e}^{(n)}, \tag{A.53}$$

and, if $m \neq n$,

$$\left\langle \mathbf{e}^{(m)}, \mathbf{e}^{(n)} \right\rangle = 0. \tag{A.54}$$

If in addition the matrix $\mathbf{S}$ is positive, due to $(A.52)$ all the eigenvalues must be positive. Furthermore, we can always rearrange the eigenvalues, and their respective eigenvectors, in such a way that:

$$\lambda_1 \geq \ldots \geq \lambda_N \geq 0. \tag{A.55}$$

Finally, we can always normalize the eigenvectors in such a way that their length is unitary:

$$\left\| \mathbf{e}^{(n)} \right\| = 1, \quad n = 1, \ldots, N. \tag{A.56}$$

Under the restrictions $(A.55)$ and $(A.56)$, and modulo a reflection of the eigenvectors, there exists only one such set of eigenvalue-eigenvector pairs $\left\{\lambda_n, \mathbf{e}^{(n)}\right\}$.

For example the matrix

$$\mathbf{S} \equiv \begin{pmatrix} \frac{9}{4} & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{11}{4} \end{pmatrix} \tag{A.57}$$

is symmetric and positive definite. Indeed, the eigenvalues can be computed as in $(A.50)$ and read:

$$\lambda_1 = 3, \quad \lambda_2 = 2. \tag{A.58}$$

Solving $(A.46)$ for the eigenvectors

$$\begin{pmatrix} \frac{9}{4} - 3 & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{11}{4} - 3 \end{pmatrix} \mathbf{e}^{(1)} = \mathbf{0}, \quad \begin{pmatrix} \frac{9}{4} - 2 & \frac{\sqrt{3}}{4} \\ \frac{\sqrt{3}}{4} & \frac{11}{4} - 2 \end{pmatrix} \mathbf{e}^{(2)} = \mathbf{0}, \tag{A.59}$$

[2] It is customary to define a matrix as *positive definite* if the inequality in $(A.52)$ is strict and *positive semi-definite* if that inequality is slack.

we obtain:

$$\mathbf{e}^{(1)} = \rho \begin{pmatrix} 1 \\ \sqrt{3} \end{pmatrix} = \widetilde{\rho} \begin{pmatrix} \cos\frac{\pi}{3} \\ \sin\frac{\pi}{3} \end{pmatrix} \tag{A.60}$$

$$\mathbf{e}^{(2)} = \kappa \begin{pmatrix} -\sqrt{3} \\ 1 \end{pmatrix} = \widetilde{\kappa} \begin{pmatrix} -\sin\frac{\pi}{3} \\ \cos\frac{\pi}{3} \end{pmatrix}. \tag{A.61}$$

In this expression $\rho, \widetilde{\rho}, \kappa$ and $\widetilde{\kappa}$ are arbitrary constants: imposing $(A.56)$ we obtain $\widetilde{\rho} \equiv \widetilde{\kappa} \equiv 1$.

Notice that $(A.54)$ and $(A.56)$ imply that the following matrix, defined as the juxtaposition of the eigenvectors:

$$\mathbf{E} \equiv \left(\mathbf{e}^{(1)}, \ldots, \mathbf{e}^{(N)}\right), \tag{A.62}$$

satisfies:

$$\mathbf{E}\mathbf{E}' = \mathbf{I}_N. \tag{A.63}$$

Comparing $(A.63)$ with $(A.31)$ we see that $\mathbf{E}$ represents a rotation in $\mathbb{R}^N$ and thus does not alter the norm of a vector:

$$\|\mathbf{E}\mathbf{v}\| = \|\mathbf{v}\| = \|\mathbf{E}'\mathbf{v}\|. \tag{A.64}$$

Defining:

$$\mathbf{\Lambda} \equiv \operatorname{diag}(\lambda_1, \ldots, \lambda_N), \tag{A.65}$$

we can restate the spectral theorem $(A.53)$ as follows:

$$\mathbf{S} = \mathbf{E}\mathbf{\Lambda}\mathbf{E}'. \tag{A.66}$$

From the invariance of the trace $(A.45)$.we obtain the following relation between the diagonal elements of $\mathbf{S}$ and the sum of its eigenvalues:

$$\sum_{n=1}^{N} S_{nn} \equiv \operatorname{tr}(\mathbf{S}) = \operatorname{tr}(\mathbf{\Lambda}) \equiv \sum_{n=1}^{N} \lambda_n, \tag{A.67}$$

Notice also that the first, largest eigenvalue of the symmetric and positive matrix $\mathbf{S}$ satisfies the following identity:

$$\begin{aligned} \lambda_1 &= \max_{\|\mathbf{u}\|=1} \{\mathbf{u}'\mathbf{\Lambda}\mathbf{u}\} = \max_{\|\mathbf{E}'\mathbf{z}\|=1} \left\{ (\mathbf{E}'\mathbf{z})' \mathbf{\Lambda} (\mathbf{E}'\mathbf{z}) \right\} \\ &= \max_{\|\mathbf{z}\|=1} \{\mathbf{z}'\mathbf{S}\mathbf{z}\} = \max_{\mathbf{z}} \left\{ \frac{\mathbf{z}'\mathbf{S}\mathbf{z}}{\mathbf{z}'\mathbf{z}} \right\}. \end{aligned} \tag{A.68}$$

Similarly, the last, smallest eigenvalue of $\mathbf{S}$ satisfies:

$$\lambda_N = \min_{\mathbf{z}} \frac{\mathbf{z}\mathbf{S}\mathbf{z}}{\mathbf{z}'\mathbf{z}}. \tag{A.69}$$

We conclude mentioning that if all the entries of a symmetric and positive matrix $\mathbf{S}$ are positive, the *Perron-Frobenius theorem* implies that the entries of the eigenvector relative to the largest eigenvalue are all positive, see Smirnov (1970). In other words, the first eigenvector points in the direction of the first orthant in the geometrical representation on the left of Figure A.1.

A.5.2 Geometrical interpretation

By means of the spectral theorem we can provide an intuitive geometrical representation of a symmetric and positive matrix. First of all, we write the spectral theorem $(A.66)$ as follows:

$$\mathbf{S} = \mathbf{E}\sqrt{\mathbf{\Lambda}}\sqrt{\mathbf{\Lambda}}\mathbf{E}', \tag{A.70}$$

where $\mathbf{\Lambda}$ is the diagonal matrix $(A.65)$ of the positive eigenvalues of $\mathbf{S}$ and $\mathbf{E}$ is the juxtaposition of the eigenvectors of $\mathbf{S}$ as defined in $(A.62)$.

In our example $(A.57)$ we have

$$\sqrt{\mathbf{\Lambda}} \equiv \text{diag}\left(\sqrt{3}, \sqrt{2}\right), \tag{A.71}$$

and

$$\mathbf{E} \equiv \begin{pmatrix} \cos\frac{\pi}{3} & -\sin\frac{\pi}{3} \\ \sin\frac{\pi}{3} & \cos\frac{\pi}{3} \end{pmatrix}. \tag{A.72}$$

Consider the following locus:

$$\mathcal{E}_{\mathbf{m},\mathbf{S}} \equiv \left\{\mathbf{x} \in \mathbb{R}^N \text{ such that } (\mathbf{x}-\mathbf{m})' \mathbf{S}^{-1} (\mathbf{x}-\mathbf{m}) \leq 1\right\}, \tag{A.73}$$

where $\mathbf{m}$ is any fixed vector in $\mathbb{R}^N$. This equation represents an ellipsoid. Indeed, consider a new set of coordinates $\mathbf{y}$ in $\mathbb{R}^N$, obtained by the following affine transformation:

$$\mathbf{y} \equiv \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{E}' (\mathbf{x}-\mathbf{m}). \tag{A.74}$$

Using $(A.63)$ we invert this relation as follows:

$$\mathbf{x} = \mathbf{m} + \mathbf{E}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{y}. \tag{A.75}$$

Substituting this expression in $(A.73)$ we see that $\mathcal{E}_{\mathbf{m},\mathbf{S}}$ is the equation of the unit sphere in the new coordinates:

$$\mathcal{E}_{\mathbf{m},\mathbf{S}} \equiv \left\{\mathbf{y} \in \mathbb{R}^N \text{ such that } y_1^2 + \cdots + y_N^2 \leq 1\right\}. \tag{A.76}$$

On the other hand, from $(A.75)$ it follows that the locus $(A.73)$ is obtained by first left-multiplying each point $\mathbf{y}$ on the unit sphere by the matrix $\mathbf{\Lambda}^{\frac{1}{2}}$; then by left-multiplying the outcome by the matrix $\mathbf{E}$; and finally by adding the vector $\mathbf{m}$.

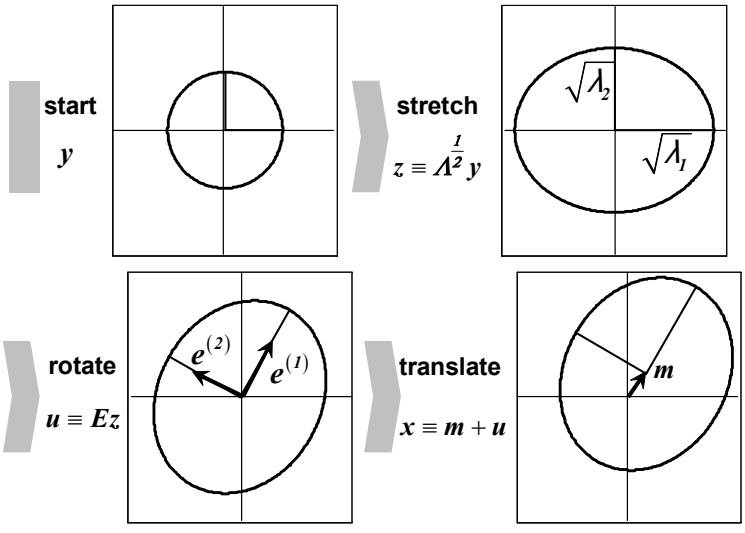

Fig. A.4. Representation of symmetric positive matrices as ellipsoids

Since the matrix $\mathbf{\Lambda}^{\frac{1}{2}}$ is diagonal, the first operation in $(A.75)$, namely the multiplication by $\mathbf{\Lambda}^{\frac{1}{2}}$, corresponds to stretching the unit sphere along each coordinate axis by an amount equal to the square root of the respective eigenvalue, see Figure A.4. Therefore the sphere becomes an ellipsoid whose principal axes are aligned with the reference axes and where, for each $n = 1, \ldots, N$, the length of the n-th principal axis is the square root of the n-th eigenvalue of $\mathbf{S}$. This step defines the *shape* of the ellipsoid. In particular, the volume of the ellipsoid is proportional to the product of the lenghts of the principal axes:

$$\text{Vol}\{\mathcal{E}_{\mathbf{m},\mathbf{S}}\} = \gamma_N \sqrt{\lambda_1} \cdots \sqrt{\lambda_N} = \gamma_N \sqrt{|\mathbf{\Lambda}|} = \gamma_N \sqrt{|\mathbf{S}|}, \tag{A.77}$$

In this expression the constant γ_N is the volume of the unit sphere in N dimensions:

$$\gamma_N \equiv \frac{\pi^{\frac{N}{2}}}{\Gamma\left(\frac{N}{2}+1\right)}, \tag{A.78}$$

where Γ is the gamma function $(B.80)$, see Fang, Kotz, and Ng (1990), p. 74.

In our example from $(A.71)$ the first reference axis is stretched by a factor $\sqrt{3}$ and the second reference axis is stretched by a factor $\sqrt{2}$. Thus the area of the ellipsoid is $\pi\sqrt{6}$.

As for the second operation in $(A.75)$, namely the multiplication by the rotation $\mathbf{E}$, from $(A.15)$ and $(A.62)$ the rotation $\mathbf{E}$ applied to the n-th element

of the canonical basis $\boldsymbol{\delta}^{(n)}$ satisfies:

$$\mathbf{E}\boldsymbol{\delta}^{(n)} = \left(\mathbf{e}^{(1)}, \cdots, \mathbf{e}^{(N)}\right)\boldsymbol{\delta}^{(n)} = \mathbf{e}^{(n)}. \tag{A.79}$$

Therefore $\mathbf{E}$ rotates $\boldsymbol{\delta}^{(n)}$, i.e. the direction of the n-th coordinate axis, into the direction defined by the n-th eigenvector of $\mathbf{S}$. In other words, the rotation $\mathbf{E}$ brings the principal axes of the ellipsoid, that originally were aligned with the reference axes, to be aligned with the direction of the eigenvectors, see Figure A.4. This step defines the *orientation* of the ellipsoid.

In our example, comparing $(A.72)$ with $(A.32)$ we see that $\mathbf{E}$ represents a counterclockwise rotation of a $\pi/3$ angle in the plane.

Finally the third operation in $(A.75)$, namely adding the vector $\mathbf{m}$, translates the center of the ellipsoid from the origin to the point $\mathbf{m}$, keeping the principal axes parallel to the eigenvectors. This step defines the *location* of the ellipsoid.

In our example we assumed:

$$\mathbf{m} \equiv (0.3, 0.4)'. \tag{A.80}$$

Therefore the ellipsoid is translated in such a way that $(A.80)$ becomes its center.

To summarize, the locus $\mathcal{E}_{\mathbf{m},\mathbf{S}}$ defined in $(A.73)$ is an ellipsoid. The principal axes of this ellipsoid are parallel to the eigenvectors of $\mathbf{S}$ and the lenghts of the principal axes are the square roots of the eigenvalues of $\mathbf{S}$. Hence, the orientation and the shape of the ellipsoid $\mathcal{E}_{\mathbf{m},\mathbf{S}}$ contain all the information about $\mathbf{S}$, namely the information about eigenvalues and eigenvectors: therefore the orientation and the shape of $\mathcal{E}_{\mathbf{m},\mathbf{S}}$ are a representation of $\mathbf{S}$. Similarly, the ellipsoid $\mathcal{E}_{\mathbf{m},\mathbf{S}}$ is centered in $\mathbf{m}$. Hence, the location of the ellipsoid $\mathcal{E}_{\mathbf{m},\mathbf{S}}$ contains all the information about $\mathbf{m}$ and thus the location of $\mathcal{E}_{\mathbf{m},\mathbf{S}}$ is a representation of $\mathbf{m}$.

A.6 Matrix operations

We present here some matrix operations that we apply in the main text to tackle financial problems. See Searle (1982), Magnus and Neudecker (1999), and references therein for more on this subject.

A.6.1 Useful identities

From $\mathbf{I} = \mathbf{A}\mathbf{A}^{-1}$ and $\mathbf{I} = \mathbf{I}'$ we obtain the following identity:

$$(\mathbf{A}')^{-1} = \left(\mathbf{A}^{-1}\right)'. \tag{A.81}$$

From $(A.36)$ we derive:

$$|\mathbf{BA}| = |\mathbf{B}|\,|\mathbf{A}|. \tag{A.82}$$

In particular, from $(A.37)$ we obtain:

$$\left|\mathbf{A}^{-1}\right| = \frac{1}{|\mathbf{A}|}. \tag{A.83}$$

Changing the matrix $\mathbf{A}$ into its transpose $\mathbf{A}'$ in the computation of the determinant $(A.39)$ does not affect the result, therefore:

$$|\mathbf{A}'| = |\mathbf{A}|. \tag{A.84}$$

From $(A.44)$ we obtain:

$$\mathrm{tr}\,(\mathbf{ABC}) = \mathrm{tr}\,(\mathbf{BCA}). \tag{A.85}$$

Finally, partition a generic $N \times N$ invertible matrix $\mathbf{M}$ as follows:

$$\mathbf{M} \equiv \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}, \tag{A.86}$$

where the $K \times K$ matrix $\mathbf{A}$ is invertible and so is the $(N-K) \times (N-K)$ matrix $\mathbf{D}$, the size of the remaining matrices being determined accordingly. Define the *Schur complements* of $\mathbf{A}$ and $\mathbf{D}$ respectively:

$$(\mathbf{M}|\mathbf{A}) \equiv \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}, \quad (\mathbf{M}|\mathbf{D}) \equiv \mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}; \tag{A.87}$$

and define:

$$(\mathbf{B}|\mathbf{M}) \equiv (\mathbf{M}|\mathbf{D})^{-1}\,\mathbf{BD}^{-1}, \quad (\mathbf{C}|\mathbf{M}) \equiv \mathbf{D}^{-1}\mathbf{C}\,(\mathbf{M}|\mathbf{D})^{-1}. \tag{A.88}$$

Then

$$\mathbf{M}^{-1} = \begin{pmatrix} (\mathbf{M}|\mathbf{D})^{-1} & -(\mathbf{B}|\mathbf{M}) \\ -(\mathbf{C}|\mathbf{M}) & (\mathbf{M}|\mathbf{A})^{-1} \end{pmatrix}. \tag{A.89}$$

In particular, some algebra shows that the following identity holds for any conformable matrices:

$$\left(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\left(\mathbf{CA}^{-1}\mathbf{B} - \mathbf{D}\right)^{-1}\mathbf{CA}^{-1}. \tag{A.90}$$

Also, the relation below follows:

$$|\mathbf{I}_J + \mathbf{CB}| = |\mathbf{I}_K + \mathbf{BC}|, \tag{A.91}$$

where $J \equiv N - K$ is the number of rows in $\mathbf{C}$, which is arbitrary, since N and K are arbitrary.

A.6.2 Tensors and Kronecker product

Loosely speaking, vectors can be considered as matrices with only one side. Matrices have two sides. Tensors are matrices with three or more sides. Tensors are the subject of multilinear analysis. A *tensor* of order p is a function from the p-dimensional grid of coordinates to $\mathbb{R}$:

$$T: \{1,\dots,N_1\}\times\cdots\times\{1,\dots,N_p\}\mapsto T_{n_1\cdots n_p}\in\mathbb{R}. \tag{A.92}$$

For example, from $(A.2)$ a vector is a tensor of order 1:

$$\mathbf{v}: \{1,\dots,N\}\mapsto v_n\in\mathbb{R}. \tag{A.93}$$

Similarly, from $(A.29)$ a matrix is a tensor of order 2:

$$\mathbf{A}: \{1,\dots,M\}\times\{1,\dots,N\}\mapsto A_{mn}\in\mathbb{R}. \tag{A.94}$$

The set of tensors of a given order is a vector space whose elements enjoy remarkable transformation properties. A less superficial discussion of this subject is beyond the scope of this book.

The *Kronecker product* is an operation defined between two generic matrices $\mathbf{A}$ and $\mathbf{B}$ of dimensions $M\times N$ and $P\times Q$ respectively. The result is a tensor of order four:

$$[\mathbf{A}\otimes\mathbf{B}]_{mnpq}\equiv A_{mn}B_{pq}. \tag{A.95}$$

Given the special structure of the tensor $(A.95)$, we can represent the Kronecker product equivalently as the following $MP\times NQ$ matrix:

$$\mathbf{A}\otimes\mathbf{B}\equiv\begin{pmatrix} A_{11}\mathbf{B} & \cdots & A_{1N}\mathbf{B} \\ \vdots & \ddots & \vdots \\ A_{M1}\mathbf{B} & \cdots & A_{MN}\mathbf{B}\end{pmatrix}. \tag{A.96}$$

We can check from the definition $(A.96)$ that the Kronecker product is distributive with respect to the sum and associative:

$$\begin{aligned}&\mathbf{A}\otimes(\mathbf{B}+\mathbf{C})=\mathbf{A}\otimes\mathbf{B}+\mathbf{A}\otimes\mathbf{C}\\ &(\mathbf{B}+\mathbf{C})\otimes\mathbf{A}=\mathbf{B}\otimes\mathbf{A}+\mathbf{C}\otimes\mathbf{A}\\ &\mathbf{A}\otimes(\mathbf{B}\otimes\mathbf{C})=(\mathbf{A}\otimes\mathbf{B})\otimes\mathbf{C}.\end{aligned} \tag{A.97}$$

Nevertheless, it is not commutative:

$$\mathbf{A}\otimes\mathbf{B}\neq\mathbf{B}\otimes\mathbf{A}. \tag{A.98}$$

Also, the Kronecker product satisfies:

$$(\mathbf{A}\otimes\mathbf{B})'=\mathbf{A}'\otimes\mathbf{B}' \tag{A.99}$$

and

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}. \tag{A.100}$$

If $\mathbf{A}$ is an $N \times N$ invertible matrix and $\mathbf{B}$ is a $K \times K$ invertible matrix, from $(A.100)$ it follows immediately:

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}. \tag{A.101}$$

Also, for the determinant of the Kronecker product it follows:

$$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^K |\mathbf{B}|^N, \tag{A.102}$$

and for the trace of the Kronecker product:

$$\mathrm{tr}(\mathbf{A} \otimes \mathbf{B}) = \mathrm{tr}(\mathbf{A})\,\mathrm{tr}(\mathbf{B}). \tag{A.103}$$

A.6.3 The "vec" and "vech" operators

The vec *operator* stacks the K columns of a generic $N \times K$ matrix $\mathbf{A} \equiv \left(\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(K)}\right)$ into an NK-dimensional column vector:

$$\mathrm{vec}[\mathbf{A}] \equiv \begin{pmatrix} \mathbf{a}^{(1)} \\ \vdots \\ \mathbf{a}^{(K)} \end{pmatrix}. \tag{A.104}$$

For instance, in the case $N \equiv 2$:

$$\mathrm{vec}\left[\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\right] \equiv \begin{pmatrix} a_{11} \\ a_{21} \\ a_{12} \\ a_{22} \end{pmatrix}. \tag{A.105}$$

A notable link between the vec operator and the Kronecker product is the following relation, that holds for any conformable matrices:

$$\mathrm{vec}[\mathbf{ABC}] = (\mathbf{C}' \otimes \mathbf{A})\,\mathrm{vec}[\mathbf{B}]. \tag{A.106}$$

Also notice the simple relation between the vec operator and the trace:

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{vec}[\mathbf{A}']'\,\mathrm{vec}[\mathbf{B}]. \tag{A.107}$$

If instead of stacking the columns of $\mathbf{A}$ we stacked the columns of its transpose $\mathbf{A}'$ we would obtain an NK-dimensional vector with the same entries, but in different order. The matrix $\mathbf{K}$ that transforms one vector into the other is called the *commutation matrix* and is thus defined by the following identity:

$$\text{vec}\left[\mathbf{A}\right] \equiv \mathbf{K}_{NK}\,\text{vec}\left[\mathbf{A}'\right]. \tag{A.108}$$

The commutation matrix satisfies:

$$\mathbf{K}'_{NK} = \mathbf{K}^{-1}_{NK} = \mathbf{K}_{KN}. \tag{A.109}$$

The explicit expression of the commutation matrix is given in terms of the canonical basis $(A.15)$ as follows:

$$\mathbf{K}_{NK} \equiv \sum_{n=1}^{N}\sum_{k=1}^{K}\left(\left[\boldsymbol{\delta}^{(n)}\right]\left[\boldsymbol{\delta}^{(k)}\right]' \otimes \left[\boldsymbol{\delta}^{(k)}\right]\left[\boldsymbol{\delta}^{(n)}\right]'\right). \tag{A.110}$$

For instance, in the case $N \equiv K \equiv 2$:

$$\mathbf{K}_{22} \equiv \begin{pmatrix} 1&0&0&0\\ 0&0&1&0\\ 0&1&0&0\\ 0&0&0&1 \end{pmatrix}. \tag{A.111}$$

Consider now a symmetric $N \times N$ square matrix $\boldsymbol{\Omega}$. To deal only with the non-redundant entries of $\boldsymbol{\Omega}$ we introduce the vech operator, which stacks the columns of a $\boldsymbol{\Omega}$ skipping the entries above the diagonal. The result is an $N(N+1)/2$-dimensional column vector.

For instance, in the case $N \equiv 2$:

$$\text{vech}\left[\begin{pmatrix} \omega_{11} & \omega_{21} \\ \omega_{21} & \omega_{22} \end{pmatrix}\right] \equiv \begin{pmatrix} \omega_{11} \\ \omega_{21} \\ \omega_{22} \end{pmatrix}. \tag{A.112}$$

Since $\text{vec}\left[\boldsymbol{\Omega}\right]$ contains the redundant entries of $\boldsymbol{\Omega}$, it can be obtained from $\text{vech}\left[\boldsymbol{\Omega}\right]$ by means of a suitable constant matrix $\mathbf{D}$, called the *duplication matrix*, which is defined by the following identity:

$$\text{vec}\left[\boldsymbol{\Omega}\right] \equiv \mathbf{D}_N\,\text{vech}\left[\boldsymbol{\Omega}\right]. \tag{A.113}$$

For instance in the case $N \equiv 2$:

$$\mathbf{D}_2 \equiv \begin{pmatrix} 1&0&0\\ 0&1&0\\ 0&1&0\\ 0&0&1 \end{pmatrix}. \tag{A.114}$$

A.6.4 Matrix calculus

We assume known the rules of calculus for smooth real-valued functions $f(\mathbf{x})$, where $\mathbf{x}$ is a vector in $\mathbb{R}^N$. Consider an $N \times K$ matrix of variables $\mathbf{X}$ and a smooth real-valued function $f(\mathbf{X})$. By means of the vec operator we can extend the rules of calculus to this new environment. Indeed, the function f can be seen equivalently as feeding on NK-dimensional vectors:

$$f(\mathbf{X}) \equiv f(\mathrm{vec}[\mathbf{X}]), \tag{A.115}$$

where vec is the operator $(A.104)$. In view of optimization problems, we are mainly interested in computing the gradient $\mathbf{g}$, which is an NK-dimensional vector

$$\mathbf{g} \equiv \frac{\partial f}{\partial \mathrm{vec}[\mathbf{X}]}, \tag{A.116}$$

and the Hessian $\mathbf{H}$, which is an $NK \times NK$ symmetric matrix:

$$\mathbf{H} \equiv \frac{\partial^2 f}{\partial \mathrm{vec}[\mathbf{X}]\, \partial \mathrm{vec}[\mathbf{X}]'}. \tag{A.117}$$

Since the direct computation of these quantities from the definition might be hard, we propose alternative routes to obtain the desired results, based on a Taylor expansion. Indeed, if we manage to express the first variation of the function f due to an infinitesimal change $d\mathbf{X}$ as follows:

$$df = \mathbf{g}' \mathrm{vec}[d\mathbf{X}], \tag{A.118}$$

then $\mathbf{g}$ is the gradient $(A.116)$. For instance the following result holds:

$$df = \mathrm{tr}(\mathbf{G} d\mathbf{X}) \Rightarrow \frac{\partial f}{\partial \mathrm{vec}[\mathbf{X}]} = \mathrm{vec}[\mathbf{G}'], \tag{A.119}$$

which follows from $(A.118)$ and the set of equalities:

$$\mathrm{tr}(\mathbf{G} d\mathbf{X}) = \sum_{m,n=1}^{N} [\mathbf{G}']_{nm}\, d\mathbf{X}_{nm} = \mathrm{vec}[\mathbf{G}']' \mathrm{vec}[d\mathbf{X}]. \tag{A.120}$$

Similarly, if we manage to express the second variation of the function f due to an infinitesimal change $d\mathbf{X}$ as follows:

$$d(df) = \mathrm{vec}[d\mathbf{X}]' \mathbf{H} \mathrm{vec}[d\mathbf{X}], \tag{A.121}$$

where $\mathbf{H}$ is symmetric, then $\mathbf{H}$ is the Hessian $(A.117)$.

As an application we derive the gradient and the Hessian of $\ln|\mathbf{X}|$, where $\mathbf{X}$ is a square $N \times N$ matrix. Consider first a matrix $\boldsymbol{\epsilon}$ of small elements. A direct computation of the determinant $(A.39)$ shows that:

$$|\mathbf{I}+\boldsymbol{\epsilon}| \approx 1+\mathrm{tr}(\boldsymbol{\epsilon})+\cdots, \tag{A.122}$$

where the dots contain products of two or more small terms ϵ_{mn} which are second-order with respect to the leading terms. Then:

$$\begin{aligned} d|\mathbf{X}| &\equiv |\mathbf{X}+d\mathbf{X}|-|\mathbf{X}| = |\mathbf{X}|\left(\left|\mathbf{I}+\mathbf{X}^{-1}d\mathbf{X}\right|-1\right) \\ &= |\mathbf{X}|\,\mathrm{tr}\left(\mathbf{X}^{-1}d\mathbf{X}\right), \end{aligned} \tag{A.123}$$

and thus:

$$d\ln|\mathbf{X}| = \mathrm{tr}\left(\mathbf{X}^{-1}d\mathbf{X}\right). \tag{A.124}$$

Applying the general rule (*A*.119) to this specific case we obtain:

$$\frac{\partial \ln|\mathbf{X}|}{\partial \mathrm{vec}[\mathbf{X}]} = \mathrm{vec}\left[\left(\mathbf{X}'\right)^{-1}\right]. \tag{A.125}$$

To compute the Hessian of $\ln|\mathbf{X}|$ first of all we differentiate $\mathbf{I}=\mathbf{X}\mathbf{X}^{-1}$ to obtain:

$$d\left(\mathbf{X}^{-1}\right) = -\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1}. \tag{A.126}$$

Computing the second differential from (*A*.124) we obtain:

$$d(d\ln|\mathbf{X}|) = \mathrm{tr}\left(d\left(\mathbf{X}^{-1}\right)d\mathbf{X}\right) = -\mathrm{tr}\left(\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1}d\mathbf{X}\right). \tag{A.127}$$

Using (*A*.106), (*A*.107) and (*A*.108) we arrive at the following expression:

$$\begin{aligned} d(d\ln|\mathbf{X}|) &= -\mathrm{vec}[d\mathbf{X}']'\,\mathrm{vec}\left[\mathbf{X}^{-1}(d\mathbf{X})\mathbf{X}^{-1}\right] \\ &= -\mathrm{vec}[d\mathbf{X}]'\,\mathbf{K}_{NN}\left(\left(\mathbf{X}'\right)^{-1}\otimes\mathbf{X}^{-1}\right)\mathrm{vec}[d\mathbf{X}]. \end{aligned} \tag{A.128}$$

Therefore from (*A*.121) we obtain:

$$\frac{\partial^2 \ln|\mathbf{X}|}{\partial \mathrm{vec}[\mathbf{X}]\,\partial \mathrm{vec}[\mathbf{X}]'} = -\mathbf{K}_{NN}\left(\left(\mathbf{X}'\right)^{-1}\otimes\mathbf{X}^{-1}\right). \tag{A.129}$$