

Chapter 8

Symmetric and Orthogonal Matrices

8.1 Real Orthogonal Matrices

In what follows $\mathbf{1}_n$ will denote the $n \times n$ identity matrix. It is defined by the equality $\mathbf{1}_n(i, j) = \delta_{ij}$.

8.1.1 Definition. A matrix $A \in \mathbb{R}^{n \times n}$ is called orthogonal if and only if it is invertible and its inverse coincides with its transpose, that is, if and only if:

$$A^{-1} = A^T \iff A \cdot A^T = A^T \cdot A = \mathbf{1}_n. \quad (8.1)$$

The identity matrix is trivially orthogonal: indeed it coincides both with its inverse and with its transpose:

$$\mathbf{1}_n^T = \mathbf{1}_n^{-1} = \mathbf{1}_n.$$

Let

$$O(n) := \{A \in Gl_n(\mathbb{R}) \mid A^T \cdot A = A \cdot A^T = \mathbf{1}_n\},$$

be the set of all $n \times n$ orthogonal matrices.

8.1.2 Proposition. The set $O(n)$ is a subgroup, called the orthogonal group, of the linear group of invertible $n \times n$ matrices.

Proof. The statement of the proposition is equivalent to saying, by the definition of group, that the product of orthogonal matrices is an orthogonal matrix, that the product is associative, that it has a neutral element, and that the inverse of an orthogonal matrix is orthogonal. Now, if $A, B \in O(n)$,

$$(AB)^{-1} = B^{-1}A^{-1} = B^T \cdot A^T = (AB)^T, \quad (8.2)$$

where in the first equality we used the rule for inverting the product of two matrices, in the second the fact that A and B are orthogonal, and in the last the relation (??) between the transpose of a product of matrices and the product of the transposed matrices. The equality between the first and last members of (8.2) proves that the product AB is therefore orthogonal. The product of orthogonal matrices is associative, since matrix multiplication of square matrices is associative in general. The neutral element of $O(n)$ is evidently the identity matrix $\mathbf{1}_n$, which is orthogonal. Finally, if $A \in O(n)$:

$$(A^{-1})^{-1} = (A^T)^{-1} = (A^{-1})^T$$

where in the first equality we used the orthogonality of A and in the second the equality (??) according to which the transpose of the inverse is equal to the inverse of the transpose. Thus $O(n)$ is a group, called the *orthogonal group*. ■

8.1.3 Proposition. *The determinant of an orthogonal matrix is 1 or -1 .*

Proof. Since $\det(A^T) = \det(A)$ for every $A \in \mathbb{R}^{n \times n}$, if A is orthogonal:

$$1 = \det(\mathbf{1}_n) = \det(A^T A) = \det(A^T) \det(A) = \det(A)^2,$$

where in the first equality we used the fact that the identity matrix has determinant 1, in the second the definition of orthogonal matrix, in the third the equality ??, ii) between the determinant of a product of two matrices and the product of the determinants, and in the fourth the fact that the determinant of a matrix coincides with that of its transpose. From the equality $\det(A)^2 = 1$ the assertion follows, that is $A \in O(n) \Rightarrow \det(A) = \pm 1$. ■

The set

$$SO(n) := \{A \in O(n) \mid \det(A) = 1\}.$$

is a *subgroup*¹ of $O(n)$, called the *special orthogonal group*. A matrix $A \in SO(n)$ is called a *special orthogonal matrix*.

8.1.4 Esempio. A point of the unit circle $S^1 := \{(x, y) \mid x^2 + y^2 = 1\}$ is called an *angle*. If $\theta \in S^1$, the abscissa of the point θ is called $\cos \theta$ and the ordinate $\sin \theta$. Then

$$R_\theta := \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$$

is a special orthogonal matrix. Indeed $R_\theta \cdot R_\theta^T = \mathbf{1}_2$. Moreover $\det(R_\theta) = \cos^2 \theta + \sin^2 \theta = 1$. The matrix R_θ represents a *rotation* of the vector plane by the angle θ . Note that the

¹That is, it is a subset of $O(n)$ which is a group with respect to the matrix product inherited from $O(n)$.

columns of R_θ form an orthonormal basis of $\mathbb{R}^2$, corresponding to the rotation of the canonical basis by an angle θ . If θ_1 and θ_2 are two angles, the reader is invited to compute the product $R_{\theta_1} \cdot R_{\theta_2}$.

8.1.5 Recall that if $\mathbb{K}$ is any field, the row-by-column product associates to every pair $(\alpha, \mathbf{u}) \in (\mathbb{K}^n)^\vee \times \mathbb{K}^n$ the scalar:

$$\alpha \cdot \mathbf{u} = \sum_{i=1}^n \alpha(i) \mathbf{u}(i).$$

This product induces a Euclidean structure on $(\mathbb{R}^n)^\vee$, the vector space of rows with n real entries, and on $\mathbb{R}^n$, the vector space of columns with n real entries, by setting, for every $\alpha, \beta \in (\mathbb{R}^n)^\vee$ and every $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$:

$$\langle \alpha, \beta \rangle = \alpha \cdot \beta^T \quad \text{and} \quad \langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \cdot \mathbf{v}. \quad (8.3)$$

The transpositions $^T : \mathbb{R}^n \rightarrow (\mathbb{R}^n)^\vee$ and $^T : (\mathbb{R}^n)^\vee \rightarrow \mathbb{R}^n$ are isomorphisms of Euclidean spaces, in the sense that

$$\langle \alpha^T, \beta^T \rangle = \langle \alpha, \beta \rangle \quad \text{and} \quad \langle \mathbf{u}^T, \mathbf{v}^T \rangle = \langle \mathbf{u}, \mathbf{v} \rangle,$$

for every pair $(\alpha, \beta) \in (\mathbb{R}^n)^\vee \times (\mathbb{R}^n)^\vee$ and every pair $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^n \times \mathbb{R}^n$. By abuse of notation the scalar products in the two vector spaces, distinct although isomorphic, of rows and columns have been denoted by the same symbol. The canonical bases of the space of n -tuples of real numbers written as rows and as columns:

$$E_n^T := (\epsilon_1, \dots, \epsilon_n) \quad \text{and} \quad E_n := (\mathbf{e}_1, \dots, \mathbf{e}_n)$$

are *orthonormal* bases with respect to the scalar products (8.3), in the sense that for every

$$1 \leq i, j \leq n$$

$$\langle \epsilon_i, \epsilon_j \rangle = \delta_{ij} \quad \text{and} \quad \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij},$$

where, as usual, δ_{ij} is the Kronecker indicator, equal to 0 for all pairs (i, j) except those for which $i = j$, in which case it is equal to 1.

8.1.6 Proposition. *A matrix $A \in \mathbb{R}^{n \times n}$ is orthogonal if and only if it is an isometry of $\mathbb{R}^n$ with respect to the scalar product $\langle, \rangle$ defined by the second of (8.3), that is, if and only if*

$$\langle A\mathbf{u}, A\mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle$$

for every pair of vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.

Proof. If $A \in O(n)$, then for every pair of vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$:

$$\begin{aligned}
 \langle A\mathbf{u}, A\mathbf{v} \rangle &= (\mathbf{Au})^T \cdot A\mathbf{v} = && \text{(definition (8.3) of the scalar product of columns)} \\
 &= (\mathbf{u}^T A^T) \cdot A\mathbf{v} = && \text{(transposition of the product } A\mathbf{u}) \\
 &= \mathbf{u}^T \cdot (A^T A) \cdot \mathbf{v} = && \text{(associativity of matrix multiplication)} \\
 &= \mathbf{u}^T \cdot \mathbf{1}_n \cdot \mathbf{v} = && \text{(orthogonality of the matrix } A) \\
 &= \mathbf{u}^T \mathbf{v} = && \text{(multiplication by the identity)} \\
 &= \langle \mathbf{u}, \mathbf{v} \rangle, && \text{(definition (8.3) of the scalar product of columns)}
 \end{aligned}$$

and thus A is an isometry. Conversely, suppose that A is an isometry. Then A is invertible. Indeed, if $\mathbf{u} \in \ker(A)$:

$$|\mathbf{u}|^2 = \langle \mathbf{u}, \mathbf{u} \rangle = \langle A\mathbf{u}, A\mathbf{u} \rangle = \langle \mathbf{0}, \mathbf{0} \rangle = 0,$$

whence $\mathbf{u} = \mathbf{0}$. Thus A is invertible because the homogeneous linear system $AX = \mathbf{0}$ has only the zero solution (Cf. Section ??). Moreover:

$$\begin{aligned}
 (A^T A)(i, j) &= R_i(A^T)C_j(A) = && \text{(matrix multiplication, rows by columns)} \\
 &= C_i(A)^T C_j(A) = && \text{(} i\text{-th row of the transpose of a matrix)} \\
 &= \langle C_i(A), C_j(A) \rangle = && \text{(definition of scalar product of columns)} \\
 &= \langle A\mathbf{e}_i, A\mathbf{e}_j \rangle = && \text{(expression for the } i\text{-th column of a matrix)} \\
 &= \langle \mathbf{e}_i, \mathbf{e}_j \rangle = && \text{(hypothesis that } A \text{ is an isometry)} \\
 &= \delta_{ij} = \mathbf{1}_n(i, j). && \text{(orthonormality of the canonical basis).}
 \end{aligned}$$

From the equality of the first and last members for every $1 \leq i, j \leq n$ follows the matrix equality $A^T \cdot A = \mathbf{1}_n$. To prove that $A \cdot A^T = \mathbf{1}_n$, one uses the fact that A is

invertible and therefore:

$$A \cdot A^T = (A \cdot A^T)AA^{-1} = A \cdot (A^T A)A^{-1} = A\mathbf{1}_n A^{-1} = \mathbf{1}_n,$$

and the proof is complete. $\blacksquare$

8.1.7 Corollary. *All eigenvalues of orthogonal matrices, real or complex, have modulus 1.*

Proof. Let A be an arbitrary orthogonal matrix and let $\lambda \in \text{Spec}(A)$. Certainly λ is nonzero because orthogonal matrices are invertible. Let $\mathbf{u}$ be a nonzero eigenvector corresponding to λ . Then

$$|\mathbf{u}| = |A\mathbf{u}| = |\lambda\mathbf{u}| = |\lambda||\mathbf{u}|$$

having used the fact that A is an isometry, by Proposition 8.1.6. Hence necessarily $|\lambda| = 1$. $\blacksquare$

8.1.8 Proposition. *Let $A \in \mathbb{R}^{n \times n}$. Then the following propositions are equivalent: (i) A is orthogonal; (ii) The columns of A are an orthonormal basis of $\mathbb{R}^n$; (iii) the rows of A form an orthonormal basis of $(\mathbb{R}^n)^\vee$.*

Proof. We prove that (i) $\iff$ (ii). If A is any matrix, then for every $1 \leq i, j \leq n$ one has

$$\langle C_i(A), C_j(A) \rangle = C_i(A)^T \cdot C_j(A) = R_i(A^T) \cdot C_j(A) = (A^T A)(i, j).$$

Thus if the columns form an orthonormal basis, $A^T A = \mathbf{1}_n$. Conversely, if A is orthogonal, $(A^T A)(i, j) = \delta_{ij}$ and the columns are then pairwise orthogonal and have modulus 1. In the same way one proves (i) $\iff$ (iii). Indeed

$$\langle R_i(A), R_j(A) \rangle = R_i(A)R_j(A)^T = R_i(A)C_j(A^T) = (A \cdot A^T)(i, j)$$

and therefore A is orthogonal if and only if the rows are an orthonormal basis of $(\mathbb{R}^n)^\vee$. $\blacksquare$

8.1.9 Esempio. All matrices of $SO(2)$ are of the form

$$R_\theta := \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

Indeed, it has already been seen in the example that the R_θ are special orthogonal. Conversely, let

$$A := \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SO(2).$$

Then $|C_1(A)| = 1$ implies that $\theta := (a, b) \in S^1$, that is, $a = \cos \theta$ and $b = \sin \theta$. Since $\langle C_1(A), C_2(A) \rangle = 0$, one has

$$b \cos \theta + d \sin \theta = 0$$

and the condition $C_1(A) \wedge C_2(A) = 1$ implies $b = -\sin \theta$ and $d = \cos \theta$.

8.1.10 Remark. Naturally the matrix

$$\begin{pmatrix} \sin \theta & -\cos \theta \\ \cos \theta & \sin \theta \end{pmatrix}$$

also belongs to $SO(2)$ and is in fact of the form R_ϕ with $\phi = \widehat{\delta} - \theta$, where $\widehat{\delta}$ is the right angle $(0, 1)$, that is, the equivalence class modulo 2π of the real number $\pi/2$.

8.2 The group $O(3)$ and $SO(3)$.

8.2.1 With respect to the notation introduced in paragraph 8.1, the group $O(3)$ is nothing but that of all 3×3 orthogonal matrices with respect to matrix multiplication. All eigenvalues of $A \in O(3)$ have modulus 1 by Corollary 8.1.7. Since the characteristic polynomial of a 3×3 matrix has degree 3, it follows that at least one eigenvalue is real. Every matrix $A \in O(3)$ therefore has either 1 or -1 as an eigenvalue. If 1 and -1 were the only eigenvalues, then A would be conjugate to the matrices

$$D_+ := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 \end{pmatrix} \quad \text{or} \quad D_- := \begin{pmatrix} -1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 \end{pmatrix} \quad (8.4)$$

up to permutations on the diagonal. It follows that if A has two real eigenvalues, it necessarily coincides either with $D_+ \in SO(3)$ or with $D_- \in O_-(3)$, the set of 3×3 orthogonal matrices with determinant -1 , also called *reflections*. Consider then a special orthogonal matrix which is not of type (8.4). Every such matrix has an eigenvalue equal to 1 and two other conjugate complex eigenvalues (necessarily of the form $\cos \alpha \pm \sqrt{-1} \sin \alpha$). Let $[\mathbf{u}]$ be the eigendirection of A corresponding to 1. It is called the *axis* of the rotation: it is a direction of fixed vectors (that is, $A\mathbf{u} = \mathbf{u}$ for every $\mathbf{v} \in [\mathbf{u}]$).

8.2.2 At this point we wish to learn how to determine all rotations of $\mathbb{R}^3$ that have as axis an arbitrary vector line $[\mathbf{u}]$ of $\mathbb{R}^3$. If the direction were, for example, $[\mathbf{e}_1]$, an

immediate verification shows that these should all be of the form

$$R_x(\alpha) := \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{pmatrix}.$$

Evidently one eigenvalue is equal to 1 and the other two, conjugate complex eigenvalues, are $\cos \alpha \pm i \sin \alpha$. Let now $[\mathbf{u}]$ be an arbitrary direction and denote by $\mathbf{b}_1$ a unit vector in it. It is orthogonal to a unique vector plane, which can be constructed easily by finding any vector $\mathbf{b}_2$ of modulus 1 orthogonal to $\mathbf{b}_1$. In that case the vector

$$\mathbf{b}_3 := \mathbf{b}_1 \times \mathbf{b}_2$$

is orthogonal both to $\mathbf{b}_1$ and to $\mathbf{b}_2$. Up to replacing $\mathbf{b}_3$ by its opposite, one may assume that $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ is a special orthogonal matrix, that is $\mathbf{b}_1 \wedge \mathbf{b}_2 \wedge \mathbf{b}_3 = 1$. The rotation matrix $R_{\mathbf{u}}(\alpha)$ with respect to the basis $\mathcal{B}$ is of the form $R_x(\alpha)$. Therefore:

$$R_{\mathbf{u}}(\alpha) = P^{E, \mathcal{B}} R_x(\alpha) P^{\mathcal{B}, E} = (\mathbf{b}_1, \cos \alpha \mathbf{b}_2 + \sin \alpha \mathbf{b}_3, \cos \alpha \mathbf{b}_2 - \sin \alpha \mathbf{b}_3) \begin{pmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \mathbf{b}_3^T \end{pmatrix}.$$

Note that if $\mathbf{u} = \mathbf{e}_1$ one obtains precisely the matrix $R_x(\alpha)$.

8.2.3 Example. Let $\mathbf{u} = \begin{pmatrix} 2 \\ 2 \\ 1 \end{pmatrix}$. Following the steps of 8.2.2, we determine all rotation matrices having the vector $\mathbf{u}$ as axis. It is convenient to begin by normalizing $\mathbf{u}$, obtaining:

$$\mathbf{b}_1 := \frac{\mathbf{u}}{|\mathbf{u}|} = \begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{pmatrix}.$$

Choose arbitrarily a vector orthogonal to $\mathbf{b}_1$, for example

$$\mathbf{v} := \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$$

and determine the vector $\mathbf{w} = \mathbf{u} \times \mathbf{v}$, orthogonal both to $\mathbf{u}$ and to $\mathbf{v}$.

$$\mathbf{w} = \begin{pmatrix} 1 \\ 1 \\ -4 \end{pmatrix}.$$

Normalizing the three vectors $\mathbf{u}, \mathbf{v}$ and the constructed $\mathbf{w}$,

$$(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3) = \left(\frac{\mathbf{u}}{|\mathbf{u}|}, \frac{\mathbf{v}}{|\mathbf{v}|}, \frac{\mathbf{w}}{|\mathbf{w}|} \right)$$

one obtains the orthogonal matrix

$$P^{E,\mathcal{B}} := \begin{pmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} \\ \frac{2}{3} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} \\ \frac{1}{3} & 0 & -\frac{2\sqrt{2}}{3} \end{pmatrix}$$

which we interpret as the transition matrix from the canonical basis $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ of $\mathbb{R}^3$ to the orthonormal basis $(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$. The matrix $P^{E,\mathcal{B}}$ is special orthogonal, because $\mathbf{b}_1 \wedge \mathbf{b}_2 \wedge \mathbf{b}_3 = 1$. Construct the matrix

$$R_{\mathbf{u}}(\alpha) := P^{E,\mathcal{B}} R_x(\alpha) P^{\mathcal{B},E}$$

which represents a generic rotation around the axis $[\mathbf{u}]$. Since the matrix is orthogonal,

$$P^{\mathcal{B},E} = (P^{E,\mathcal{B}})^{-1} = (P^{E,\mathcal{B}})^T = \begin{pmatrix} \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{6} & \frac{\sqrt{2}}{6} & -\frac{2\sqrt{2}}{3} \end{pmatrix}$$

and therefore finally

$$\begin{aligned}
 R_{\mathbf{u}}(\alpha) &= \begin{pmatrix} \frac{2}{3} & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} \\ \frac{2}{3} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{6} \\ \frac{1}{3} & 0 & -\frac{2\sqrt{2}}{3} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} & 0 \\ \frac{\sqrt{2}}{6} & \frac{\sqrt{2}}{6} & -\frac{2\sqrt{2}}{3} \end{pmatrix} \\
 &= \begin{pmatrix} \frac{1}{9}(4 + 5 \cos \alpha) & \frac{1}{9}(4 - 4 \cos \alpha - 3 \sin \alpha) & \frac{2}{9}(1 - \cos \alpha + 3 \sin \alpha) \\ \frac{1}{9}(4 - 4 \cos \alpha + 3 \sin \alpha) & \frac{1}{9}(4 + 5 \cos \alpha) & \frac{2}{9}(1 - \cos \alpha - 3 \sin \alpha) \\ \frac{2}{9}(1 - \cos \alpha - 3 \sin \alpha) & \frac{2}{9}(1 - \cos \alpha + 3 \sin \alpha) & \frac{1}{9}(1 + 8 \cos \alpha) \end{pmatrix}.
 \end{aligned}$$

The verification that $R_{\mathbf{u}}(\alpha) \cdot \mathbf{u} = \mathbf{u}$ is left to the reader.

8.3 Hermitian Matrices

Hermitian matrices are a useful generalization of symmetric matrices. Recall that if $A \in \mathbb{C}^{n \times n}$, the conjugate matrix of A is the unique matrix $\overline{A}$ whose entries are the conjugates of the entries of A :

$$\overline{A}(i, j) = \overline{A(i, j)}.$$

8.3.1 Definition. A matrix $A \in \mathbb{C}^{n \times n}$ is called Hermitian if and only if

$$A^T = \overline{A}.$$

Clearly a symmetric matrix $A \in \mathbb{R}^{n \times n}$ can be viewed as a Hermitian matrix, thinking of the real entries as complex entries. The unitary scalar product of $\mathbb{C}^n$ is the positive-definite *Hermitian sesquilinear form* that associates to every pair $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$ the complex scalar:

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \cdot \overline{\mathbf{v}}.$$

8.3.2 Proposition. *A Hermitian matrix $A \in \mathbb{C}^{n \times n}$ is self-adjoint, that is, the equality holds:*

$$\langle A\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, A\mathbf{v} \rangle \quad \forall \mathbf{u}, \mathbf{v} \in \mathbb{C}^n.$$

Proof. Indeed

$$\langle A\mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u})^T \bar{\mathbf{v}} = \mathbf{u}^T A^T \bar{\mathbf{v}} = \mathbf{u}^T \overline{A\mathbf{v}} = \mathbf{u}^T \overline{A\mathbf{v}} = \langle \mathbf{u}, A\mathbf{v} \rangle. \quad \blacksquare$$

One also says that A is *self-adjoint*.

8.3.3 Proposition. *All eigenvalues of a Hermitian matrix are real. Moreover, eigenvectors corresponding to distinct eigenvalues are orthogonal with respect to the unitary scalar product.*

Proof. Let A be Hermitian and let λ be one of its nonzero eigenvalues. Then there exists a vector $\mathbf{u} \neq \mathbf{0}$ such that $A\mathbf{u} = \lambda\mathbf{u}$. If $\lambda = 0$ it is evidently real. Suppose that $\lambda \neq 0$. Then:

$$\begin{aligned} \lambda|\mathbf{u}|^2 &= \lambda\langle \mathbf{u}, \mathbf{u} \rangle = && \text{(Definition of squared modulus of a vector)} \\ &= \langle \lambda\mathbf{u}, \mathbf{u} \rangle = && \text{(linearity in the first argument)} \\ &= \langle A\mathbf{u}, \mathbf{u} \rangle = && \text{(definition of eigenvalue)} \\ &= \langle \mathbf{u}, A\mathbf{u} \rangle = && \text{(Proposition 8.3.2)} \\ &= \langle \mathbf{u}, \lambda\mathbf{u} \rangle = && \text{(definition of eigenvalue)} \\ &= \bar{\lambda}\langle \mathbf{u}, \mathbf{u} \rangle = && \text{(antilinearity of } \langle \cdot, \cdot \rangle \text{ with respect to the second argument)} \\ &= \bar{\lambda}|\mathbf{u}|^2 && \text{(Definition of squared modulus of a vector)} \end{aligned}$$

From the equality of the first and last members, and from the fact that $|\mathbf{u}|^2 \neq 0$, follows the equality $\lambda = \bar{\lambda}$. Therefore $\lambda \in \mathbb{R}$, since it coincides with its conjugate.

Now consider $\lambda, \mu \in \text{Spec}(A)$ with $\lambda \neq \mu$. It is not restrictive to suppose $\lambda \neq 0$. Let $\mathbf{u}, \mathbf{v}$ be eigenvectors corresponding to λ and μ , respectively. Then:

$$\begin{aligned}
\lambda \langle \mathbf{u}, \mathbf{v} \rangle &= \langle \lambda \mathbf{u}, \mathbf{v} \rangle = && \text{(linearity of } \langle \cdot, \cdot \rangle \text{ with respect to the first argument)} \\
&= \langle A\mathbf{u}, \mathbf{v} \rangle = && \text{(hypothesis on } \lambda) \\
&= \langle \mathbf{u}, A\mathbf{v} \rangle = && \text{(Proposition 8.3.2)} \\
&= \langle \mathbf{u}, \mu \mathbf{v} \rangle = && \text{(hypothesis on } \mu) \\
&= \mu \langle \mathbf{u}, \mathbf{v} \rangle = && \text{(antilinearity and reality of } \mu, \text{ proved in the first part of the proposition)}
\end{aligned}$$

Comparing the first and last members, one obtains the equality

$$(\lambda - \mu) \langle \mathbf{u}, \mathbf{v} \rangle = 0,$$

whence, since $\lambda - \mu \neq 0$, one obtains $\langle \mathbf{u}, \mathbf{v} \rangle = 0$, as asserted. $\blacksquare$

8.3.4 Theorem (the Spectral Theorem). *Every Hermitian matrix is diagonalizable, that is, the geometric multiplicity of each of its eigenvalues (which is real) coincides with its algebraic multiplicity.*



-- The spectrum of a Hermitian matrix is real, not imaginary! --

Proof. The proof will be by induction on the order of the matrix. If A is a 1×1 Hermitian matrix, it is obviously diagonalizable, since the matrix is diagonal. Suppose then that every Hermitian $(n-1) \times (n-1)$ matrix is diagonalizable and prove that this implies the diagonalizability of every Hermitian $n \times n$ matrix. Let $A \in \mathbb{C}^{n \times n}$ be Hermitian. It has at least one real eigenvalue λ and, therefore, an

eigenvector $\mathbf{u} \neq \mathbf{0}$ such that $A\mathbf{u} = \lambda\mathbf{u}$. Let V be the vector subspace of $\mathbb{C}^n$ of all vectors orthogonal to $\mathbf{u}$ (with respect to the Hermitian scalar product):

$$V := \mathbf{u}^\perp := \{\mathbf{v} \in \mathbb{C}^n \mid \langle \mathbf{u}, \mathbf{v} \rangle = 0\}.$$

Naturally $\mathbb{C}^n = \mathbf{u} \oplus V$, since every vector $\mathbf{v} \in \mathbb{C}^n$ can be written as

$$\mathbf{v} := \frac{\langle \mathbf{v}, \mathbf{u} \rangle \mathbf{u}}{|\mathbf{u}|^2} + \left(\mathbf{v} - \frac{\langle \mathbf{v}, \mathbf{u} \rangle \mathbf{u}}{|\mathbf{u}|^2} \right)$$

and the second summand is orthogonal to $\mathbf{u}$. Moreover, the only vector parallel and orthogonal to $\mathbf{u}$ is the zero vector. In other words, $\mathbb{C}^n$ is the direct sum of a one-dimensional subspace of eigenvectors and the space orthogonal to it. We claim that V is invariant under A , that is, for every $\mathbf{v} \in V$, $A\mathbf{v} \in V$. Indeed, since A is self-adjoint (Proposition 8.3.2),

$$\langle A\mathbf{v}, \mathbf{u} \rangle = \langle \mathbf{v}, A\mathbf{u} \rangle = \langle \mathbf{v}, \lambda\mathbf{u} \rangle = \lambda \langle \mathbf{v}, \mathbf{u} \rangle = 0$$

where in the first equality we used Proposition 8.3.2, in the second the fact that $\mathbf{u} \in \text{Spec}(A)$, in the third linearity with respect to the second argument, and in the last the orthogonality of $\mathbf{u}$ and $\mathbf{v}$. Thus

$$A(V) \subseteq V.$$

Then A induces a linear map:

$$A|_V : V \rightarrow V.$$

Let us show that it is self-adjoint. For every $\mathbf{v}, \mathbf{w} \in V$ one has

$$\langle A|_V \mathbf{v}, \mathbf{w} \rangle = \langle A\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, A\mathbf{w} \rangle.$$

By the induction hypothesis $A|_V$ decomposes V into the direct sum of eigenspaces. Therefore $\mathbb{C}^n$ is the direct sum of eigenspaces of A and hence A is diagonalizable. ■

8.3.5 Corollary. *The eigenvalues of a real symmetric matrix are all real and the matrix is diagonalizable. Moreover, it can be diagonalized by means of an orthogonal matrix P .*

Proof. Every symmetric matrix is in particular Hermitian. Therefore it has all eigenvalues real and is diagonalizable. Since it is diagonalizable, it decomposes $\mathbb{R}^n$ into the direct sum of eigenspaces of A :

$$\mathbb{R}^n = V_{\lambda_1}(A) \oplus \dots \oplus V_{\lambda_r}(A)$$

where $\lambda_1, \dots, \lambda_r$ are the distinct eigenvalues of A and $V_{\lambda_i}(A) = \ker(A - \lambda_i \mathbf{1})$. Now, by Proposition 8.3.3, eigenspaces corresponding to distinct eigenvalues are pairwise orthogonal. Finally, every basis of $V_{\lambda_i}(A)$ can be orthonormalized by Gram–Schmidt (Cf. ??). Thus one obtains an orthonormal basis $(\mathbf{u}_1, \dots, \mathbf{u}_n)$ of $\mathbb{R}^n$ (by joining all the orthonormal bases of the individual eigenspaces) made of eigenvectors of A . Therefore the matrix $P := (\mathbf{u}_1, \dots, \mathbf{u}_n)$ is orthogonal and diagonalizes A . ■

8.3.6 Esempio. Let the matrix be given:

$$Q := \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

Being symmetric, it has all eigenvalues real and, by the spectral theorem 8.3.4, it is diagonalizable. The characteristic polynomial is

$$P_Q(T) = |T\mathbf{1}_3 - Q| = (T - 1)^2(T - 4).$$

Thus there is an eigenvalue equal to 1 with algebraic multiplicity 2, which coincides with the geometric multiplicity, given the a priori known diagonalizability of Q . The other eigenvalue, simple, is 4. Since

$$Q - \mathbf{1}_3 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix},$$

the eigenvectors corresponding to 1 have components x, y, z satisfying the system (of one equation and three unknowns, because the other two equations are equal to the first):

$$x + y + z = 0.$$

Therefore

$$\mathbf{u} := \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{v} := \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}$$

form a basis of the eigenspace corresponding to 1. To find the third eigenvector it is enough to remember that eigenspaces corresponding to distinct eigenvalues of a symmetric matrix are orthogonal. Then the cross product

$$\mathbf{u} \times \mathbf{v} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix},$$

is an eigenvector corresponding to the eigenvalue 4.

Thus a matrix $P \in Gl_3(\mathbb{R})$ that diagonalizes Q is:

$$P := (\mathbf{u}, \mathbf{v}, \mathbf{u} \times \mathbf{v}) = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix}.$$

However, P is not orthogonal. For example, the first two columns are not orthogonal. One observes, however, that

$$(\mathbf{u} \times \mathbf{v}) \times \mathbf{u} = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

is orthogonal both to $\mathbf{u} \times \mathbf{v}$ and to $\mathbf{u}$, by the properties of the cross product. Thus

$$P' := \left(\frac{\mathbf{u}}{|\mathbf{u}|}, \frac{(\mathbf{u} \times \mathbf{v}) \times \mathbf{u}}{|(\mathbf{u} \times \mathbf{v}) \times \mathbf{u}|}, \frac{\mathbf{u} \times \mathbf{v}}{|\mathbf{u} \times \mathbf{v}|} \right) = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ 0 & -\frac{\sqrt{6}}{3} & \frac{\sqrt{3}}{3} \end{pmatrix}$$

is an orthogonal matrix that diagonalizes Q .

8.4 Symmetric matrices and self-adjoint endomorphisms

8.4.1 A symmetric matrix $A \in \mathbb{R}^{n \times n}$, being Hermitian, is self-adjoint in the sense that

$$\langle A\mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, A\mathbf{v} \rangle, \quad (8.5)$$

for every $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. More generally:

8.4.2 Definition. An endomorphism ϕ of $\mathbb{R}^n$ is called self-adjoint if

$$\langle \phi(\mathbf{u}), \mathbf{v} \rangle = \langle \mathbf{u}, \phi(\mathbf{v}) \rangle.$$

where, as usual, the scalar product of $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is defined by $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \cdot \mathbf{v}$.

8.4.3 Proposition. Let $\mathcal{B} := (\mathbf{b}_1, \dots, \mathbf{b}_n)$ be an orthonormal basis of $\mathbb{R}^n$. Then $A := M_{\phi}^{\mathcal{B}, \mathcal{B}}$ is a symmetric matrix. Conversely, if A is a symmetric matrix, the endomorphism defined by

$$(\phi_A(\mathbf{u}))_{\mathcal{B}} = A \cdot (\mathbf{u})_{\mathcal{B}} \quad (8.6)$$

is a self-adjoint endomorphism.

Proof. First observe that the j -th component of the column $\phi(\mathbf{b}_i)_{\mathcal{B}}$ of the vector $\phi(\mathbf{b}_i)$ with respect to the basis $\mathcal{B}$ coincides with the component of the vector $\phi(\mathbf{b}_i)$ with respect to the basis $\mathcal{B}$ and is therefore equal to

$$R_j(\phi(\mathbf{b}_i)_{\mathcal{B}}) = \langle \mathbf{b}_j, \phi(\mathbf{b}_i) \rangle, \quad (8.7)$$

in accordance with equality (??) concerning the components of a vector with respect to an orthonormal basis. Then

$$\begin{aligned} A(j, i) &= R_j(C_i(M_{\phi}^{\mathcal{B}, \mathcal{B}})) = && \text{(equality (??))} \\ &= R_j(M_{\phi}^{\mathcal{B}, \mathcal{B}} \cdot \mathbf{e}_i) = && (i\text{-th column of the matrix } M_{\phi}^{\mathcal{B}, \mathcal{B}}) \\ &= \langle \mathbf{b}_j, \phi(\mathbf{b}_i) \rangle = && \text{(formula (8.7))} \\ &= \langle \phi(\mathbf{b}_j), \mathbf{b}_i \rangle = && (\phi \text{ is self-adjoint}) \\ &= \langle \mathbf{b}_i, \phi(\mathbf{b}_j) \rangle = && \text{(symmetry of the scalar product)} \\ &= R_i(C_j(M_{\phi}^{\mathcal{B}, \mathcal{B}})) = && (i\text{-th row of the } j\text{-th} \\ &&& \text{column of } M_{\phi}^{\mathcal{B}, \mathcal{B}}) \\ &= A(i, j) && \text{(definition of } A) \end{aligned}$$

and hence the matrix A is symmetric, since $A = A^T$. Conversely, if ϕ_A is defined as in (8.6), one has:

$$\begin{aligned} \langle \phi(\mathbf{u}), \mathbf{v} \rangle &= \langle \phi(\mathbf{u})_{\mathcal{B}}, (\mathbf{v})_{\mathcal{B}} \rangle = && \text{(by Proposition (??))} \\ &= \langle A \cdot (\mathbf{u})_{\mathcal{B}}, (\mathbf{v})_{\mathcal{B}} \rangle = && \text{(definition (8.6) of } \phi) \\ &= \langle (\mathbf{u})_{\mathcal{B}}, A \cdot (\mathbf{v})_{\mathcal{B}} \rangle = && \text{(property (8.5) of a symmetric} \\ &&& \text{matrix)} \\ &= \langle (\mathbf{u})_{\mathcal{B}}, \phi(\mathbf{v})_{\mathcal{B}} \rangle && \text{(definition (8.6) of } \phi) \\ &= \langle \mathbf{u}, \phi(\mathbf{v}) \rangle && \text{(again by Proposition (??))} \end{aligned}$$

and therefore ϕ is self-adjoint. ■

8.4.4 A notable example: the inertia tensor. Consider the set

$$\mathcal{S} := \{(P_1, m_1), \dots, (P_h, m_h)\} \subseteq \mathcal{E}^3 \times \mathbb{R}_+.$$

where $\mathbb{R}_+ := \{a \in \mathbb{R} \mid a \geq 0\}$. The positive real number m_i is called the *mass* of the point P_i . Consider the map $\sigma_O : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ defined by:

$$\sigma_O(\mathbf{u}) = \sum_{i=1}^h (m_i \overrightarrow{OP_i} \times \mathbf{u}) \times \overrightarrow{OP_i}. \quad (8.8)$$

where $O \in \mathcal{E}^3$. The map σ_O is undoubtedly linear, by the bilinearity properties ?? of the cross product, and in fact it is a *self-adjoint* endomorphism, as shown by the following chain of equalities.

$$\begin{aligned} \langle \sigma_O(\mathbf{u}), \mathbf{v} \rangle &= \langle \sum_{i=1}^h (m_i \overrightarrow{OP_i} \times \mathbf{u}) \times \overrightarrow{OP_i}, \mathbf{v} \rangle = && \text{(definition of } \sigma_O) \\ &= \langle \sum_{i=1}^h m_i \overrightarrow{OP_i} \times \mathbf{u}, \overrightarrow{OP_i} \times \mathbf{v} \rangle = && \text{(equality (??))} \\ &= \sum_{i=1}^h m_i \langle \overrightarrow{OP_i} \times \mathbf{u}, \overrightarrow{OP_i} \times \mathbf{v} \rangle = && \text{(linearity of } \langle \cdot, \cdot \rangle \text{ with respect} \\ &&& \text{to the first argument)} \\ &= \sum_{i=1}^h m_i \langle \overrightarrow{OP_i} \times \mathbf{v}, \overrightarrow{OP_i} \times \mathbf{u} \rangle = && \text{(symmetry of } \langle \cdot, \cdot \rangle) \\ &= \sum_{i=1}^h m_i \langle (\overrightarrow{OP_i} \times \mathbf{v}) \times \overrightarrow{OP_i}, \mathbf{u} \rangle = && \text{(again equality (??))} \\ &= \sum_{i=1}^h \langle m_i (\overrightarrow{OP_i} \times \mathbf{v}) \times \overrightarrow{OP_i}, \mathbf{u} \rangle = && \text{(linearity of } \langle \cdot, \cdot \rangle \text{ with respect} \\ &&& \text{to the first argument)} \\ &= \langle \sigma_O(\mathbf{v}), \mathbf{u} \rangle = && \text{(definition of } \sigma_O(\mathbf{v})) \\ &= \langle \mathbf{u}, \sigma_O(\mathbf{v}) \rangle && \text{(symmetry of the scalar} \\ &&& \text{product)} \end{aligned}$$

The endomorphism σ_O is called the *inertia tensor* of the system of *material points* $\mathcal{S} := \{(P_i, m_i)\}$ with respect to the point $O \in \mathcal{E}^3$, called the *pole*. Being self-adjoint, all its eigenvalues are real. The eigenvalues of σ_O indeed coincide with the eigenvalues of a matrix associated with σ_O with respect to any basis. If the basis is an orthonormal basis, this matrix is symmetric, and hence has all eigenvalues real, which

are called the *principal moments of inertia*. Let $\mathcal{B} := (\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)$ be an orthonormal basis of eigenvectors, that is $\sigma_O(\mathbf{b}_j) = I_j \mathbf{b}_j$, where (I_1, I_2, I_3) are the principal moments of inertia. It is easy to see that if the system $\mathcal{S}$ has at least 2 distinct points P_1 and P_2 such that O , P_1 and P_2 are not aligned, the eigenvalues I_1, I_2, I_3 of σ_O are all positive. Indeed, consider the *quadratic form of inertia*:

$$I_O(\mathbf{u}) = \langle \sigma(\mathbf{u}), \mathbf{u} \rangle.$$

Putting $u_i := \langle \mathbf{u}, \mathbf{b}_i \rangle$, one has

$$I_O(\mathbf{u}) = I_1(u_1)^2 + I_2(u_2)^2 + I_3(u_3)^2.$$

Then I_1, I_2, I_3 are all positive if and only if the quadratic form I_O is positive definite (see Definition 8.5.10 and Section 8.7 below), that is, if and only if

$$I_O(\mathbf{u}) > 0$$

for every $\mathbf{u} \neq \mathbf{0}$. Now:

$$\begin{aligned} I_O(\mathbf{u}) &= \sum_{i=1}^n m_i \langle (\overrightarrow{OP_i} \times \mathbf{u}) \times \overrightarrow{OP_i}, \mathbf{u} \rangle = && \text{(definition of } I_O \text{ and } \sigma_O) \\ &= \sum_{i=1}^n m_i \langle \overrightarrow{OP_i} \times \mathbf{u}, \overrightarrow{OP_i} \times \mathbf{u} \rangle = && \text{(property (??) of the cross product)} \\ &= \sum_{i=1}^n m_i |\overrightarrow{OP_i} \times \mathbf{u}|^2 \end{aligned}$$

Now $\sum_{i=1}^n m_i |\overrightarrow{OP_i} \times \mathbf{u}|^2 > 0$ for every $\mathbf{u} \neq \mathbf{0}$ by the hypothesis made on $\mathcal{S}$, because the squared modulus of any vector is positive unless it is zero. Even if $\mathbf{u}$ had the same direction as $\overrightarrow{OP_1}$, by the hypothesis made on $\mathcal{S}$ it would not have the same direction as $\overrightarrow{OP_2}$, and therefore $\sum_{i=1}^n m_i |\overrightarrow{OP_i} \times \mathbf{u}|^2 > 0$ if $\mathbf{u} \neq \mathbf{0}$. This proves that I_1, I_2, I_3 are positive.

If I_1, I_2, I_3 are pairwise distinct, the corresponding eigendirections are pairwise orthogonal. The lines passing through O parallel to the eigendirections are called *principal axes of inertia*. Moreover, if $\mathbf{u} = a_1 \mathbf{b}_1 + a_2 \mathbf{b}_2 + a_3 \mathbf{b}_3$ is a unit vector (and therefore $a_1^2 + a_2^2 + a_3^2 = 1$), one has:

$$I_{\mathbf{u}} := \langle \sigma_O(\mathbf{u}), \mathbf{u} \rangle = a_1 I_1 + a_2 I_2 + a_3 I_3$$

which is called the *moment of inertia* with respect to the axis passing through O and parallel to $\mathbf{u}$.

8.5 Real Symmetric Matrices and Congruence

8.5.1 Let Q be a symmetric $n \times n$ matrix and $P \in \mathbb{R}^{n \times n}$ any square matrix. Then the matrix $P^T \cdot Q \cdot P$ is symmetric. Indeed:

$$(P^T Q P)^T = P^T Q^T P^{TT},$$

by formula (??) expressing the transpose of a product of two matrices. Since $Q^T = Q$, because Q is symmetric, and $P^{TT} = P$, one obtains precisely:

$$(P^T Q P)^T = P^T \cdot Q \cdot P$$

as desired.

8.5.2 Definition. Let $Q_1, Q_2 \in \mathbb{R}^{n \times n}$ be two real symmetric matrices. The matrix Q_2 is called congruent to Q_1 if and only if there exists $P \in Gl_n(\mathbb{R})$ such that

$$Q_2 = P^T \cdot Q_1 \cdot P. \quad (8.9)$$

8.5.3 Remarks. The definition of congruent matrices resembles that of similar matrices, but the two notions are nevertheless clearly distinct. Consider, for example, the matrices

$$Q_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Q_2 := \begin{pmatrix} 2 & 3 \\ 3 & 4 \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}.$$

Then Q_2 is congruent to Q_1 since $Q_2 = P^T Q_1 P$ (verify it!) but Q_2 is not similar to Q_1 , as can be seen, for example, by noting that the trace of Q_2 is different from that of Q_1 . If Q_1 and Q_2 were similar, indeed, they should have the same characteristic polynomial and therefore also the same trace. Also note that two congruent symmetric matrices have the same rank. Indeed, multiplication by an invertible matrix is an isomorphism of $\mathbb{R}^n$, and therefore transforms vector subspaces into vector subspaces of the same dimension.

8.5.4 Proposition. The congruence relation is an equivalence relation.

Proof. It is enough to verify that the relation:

$$Q_1 \sim Q_2 \text{ if and only if } Q_2 \text{ is congruent to } Q_1$$

is reflexive, symmetric and transitive. The relation is reflexive. Indeed, every symmetric matrix Q is congruent to itself, because $Q = \mathbf{1}_n^T \cdot Q \cdot \mathbf{1}_n$. The relation is symmetric. Indeed, if Q_2 is congruent to Q_1 , there exists $P \in Gl_n(\mathbb{R})$ such that

$Q_2 = P^T \cdot Q_1 \cdot P$. Then $Q_1 = (P^T)^{-1} \cdot Q_2 \cdot P^{-1}$, whence, invoking (??) according to which the inverse of the transpose coincides with the transpose of the inverse:

$$Q_1 = (P^{-1})^T \cdot Q_2 \cdot P^{-1},$$

so Q_1 is congruent to Q_2 . As for transitivity, let Q_1, Q_2, Q_3 be such that $Q_1 \sim Q_2$ and $Q_2 \sim Q_3$. Then there exist $P_1, P_2 \in Gl_n(\mathbb{R})$ such that

$$Q_2 = P_1^T \cdot Q_1 \cdot P_1 \quad \text{and} \quad Q_3 = P_2^T \cdot Q_2 \cdot P_2,$$

hence:

$$Q_3 = (P_2)^T \cdot (P_1^T \cdot Q_1 \cdot P_1) \cdot P_2 = (P_1 P_2)^T \cdot Q_1 \cdot (P_1 P_2),$$

which proves that Q_3 is congruent to Q_1 . The congruence relation is therefore an equivalence relation. ■

8.5.5 Proposition. *Two symmetric matrices Q_1, Q_2 are congruent if and only if there exists a basis $\mathcal{B} := (\mathbf{b}_1, \dots, \mathbf{b}_n)$ of $\mathbb{R}^n$ such that*

$$Q_2(i, j) = \mathbf{b}_i^T \cdot Q_1 \cdot \mathbf{b}_j. \quad (8.10)$$

Proof. Suppose that Q_2 is congruent to Q_1 . Then there exists $P \in Gl_n(\mathbb{R})$ satisfying equality (8.9). Therefore:

$$\begin{aligned} Q_2(i, j) &= \mathbf{e}_i \cdot Q_2 \mathbf{e}_j = && \text{(entry } (i, j) \text{ of } Q_2) \\ &= (\mathbf{e}_i)^T \cdot (P^T \cdot Q_1 \cdot P) \cdot \mathbf{e}_j = && \text{(by (8.9))} \\ &= (P\mathbf{e}_i)^T \cdot Q_1 \cdot (P \cdot \mathbf{e}_j) = && \text{(formula (??) on the} \\ &&& \text{transpose of a product)} \\ &= \mathbf{b}_i^T \cdot Q_1 \cdot \mathbf{b}_j = && (\mathbf{b}_i = P\mathbf{e}_i) \end{aligned}$$

Clearly $\mathcal{B} := (\mathbf{b}_1, \dots, \mathbf{b}_n)$, where $\mathbf{b}_i = P\mathbf{e}_i$, is a basis of $\mathbb{R}^n$, because it is the image of the canonical basis of $\mathbb{R}^n$ through the invertible matrix P . Thus one of the two implications has been proved. Conversely, suppose there exists a basis $\mathcal{B}$ as in the statement, satisfying equality (8.10). Let $P := P^{\mathcal{B}, E_n}$ be the transition matrix from the canonical basis E_n to the basis $\mathcal{B}$, that is, such that $\mathbf{b}_i = P\mathbf{e}_i$ for every $1 \leq i \leq n$. Then one can write:

$$\begin{aligned}
Q_2(i, j) &= (P\mathbf{e}_i)^T \cdot Q_1(P\mathbf{e}_j) = && \text{(entry } (i, j) \text{ of } Q_2) \\
&= \mathbf{e}_i^T (P^T \cdot Q_1 \cdot P) \mathbf{e}_j = && \text{(formula (??) on the} \\
&&& \text{transpose of a product)} \\
&= (P^T \cdot Q_1 \cdot P)(i, j) && \text{(entry } (i, j) \text{ of } P^T Q_1 P).
\end{aligned}$$

Since two matrices are equal if and only if all their entries are equal, the previous chain of equalities implies that $Q_2 = P^T Q_1 P$, that is, Q_2 is congruent to Q_1 , as required. ■

8.5.6 Definition. A symmetric matrix $Q \in \mathbb{R}^{n \times n}$ is called nondegenerate if the vanishing of the product $\mathbf{u}^T Q \mathbf{v}$ for every $\mathbf{u} \in \mathbb{R}^n$ implies $\mathbf{v} = \mathbf{0}$.

8.5.7 Proposition. A symmetric matrix Q is nondegenerate if and only if $\det(Q) \neq 0$.

Proof. Let $\mathbf{v} \in \mathbb{R}^n$ be such that $\mathbf{u}^T Q \mathbf{v} = \mathbf{0}$ for every $\mathbf{u} \in \mathbb{R}^n$. By the arbitrariness of $\mathbf{u}$, one then has

$$\mathbf{e}_i^T Q \cdot \mathbf{v} = (\boldsymbol{\epsilon}_i \cdot Q) \mathbf{v} = \mathbf{0}$$

for every $1 \leq i \leq n$. Thus one obtains the n linear equations:

$$R_i(Q) \mathbf{v} = \mathbf{0}, \quad 1 \leq i \leq n, \quad (8.11)$$

which can be summarized in the homogeneous linear system $Q \mathbf{v} = \mathbf{0}$. Therefore Q is nondegenerate if and only if the only solution of the system (8.11) is $\mathbf{v} = \mathbf{0}$, that is, by the Rouché–Capelli theorem, if and only if $\det(Q) \neq 0$. ■

8.5.8 Definition. If Q is a symmetric matrix, two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ are called Q -orthogonal if and only if $\mathbf{u}^T \cdot Q \cdot \mathbf{v} = 0$.

If $\mathbf{u} \in \mathbb{R}^n$, one sets

$$\mathbf{u}_Q^\perp := \{\mathbf{v} \mid \mathbf{u}^T Q \mathbf{v} = 0\},$$

which is the set of all vectors Q -orthogonal to $\mathbf{u}$. Note that if $\mathbf{u} = \mathbf{0}$, then $\mathbf{u}_Q^\perp = V$. It is easy to prove, using the bilinearity properties of matrix multiplication, that $\mathbf{u}_Q^\perp$ is a vector subspace of $\mathbb{R}^n$, called the *vector subspace of vectors Q -orthogonal to $\mathbf{u}$* . The nondegeneracy of a symmetric matrix Q can be translated by saying that the only vector Q -orthogonal to all vectors of $\mathbb{R}^n$ is the zero vector.

8.5.9 Proposition. Let Q be symmetric and let $\mathbf{u} \in \mathbb{R}^n$ be such that $\mathbf{u}^T Q \mathbf{u} \neq 0$. Then

$$\mathbb{R}^n = [\mathbf{u}] \oplus \mathbf{u}_Q^\perp.$$

Proof. Clearly $[\mathbf{u}] \cap \mathbf{u}_Q^\perp = \mathbf{0}$, by the hypothesis. It is therefore enough to show that every $\mathbf{v} \in \mathbb{R}^n$ can be written as the sum of a vector parallel to $\mathbf{u}$ and a vector belonging to its Q -orthogonal complement. Now, if $\mathbf{v} \in \mathbb{R}^n$, one has:

$$\mathbf{v} = \frac{\mathbf{u}^T Q \mathbf{v}}{\mathbf{u}^T Q \mathbf{u}} \cdot \mathbf{u} + \left(\mathbf{v} - \frac{\mathbf{u}^T Q \mathbf{v}}{\mathbf{u}^T Q \mathbf{u}} \cdot \mathbf{u} \right).$$

The first summand clearly has the same direction as $\mathbf{u}$, and the verification that the second summand belongs to $\mathbf{u}_Q^\perp$ is left to the reader. ■

8.5.10 Definition. A real symmetric matrix Q is called

1. *positive semidefinite* if $\mathbf{u}^T Q \mathbf{u} \geq 0$ for every $\mathbf{u} \in \mathbb{R}^n$;
2. *positive definite* if it is positive semidefinite and $\mathbf{u}^T Q \mathbf{u} = 0$ implies $\mathbf{u} = \mathbf{0}$;
3. *(semi)negative definite* if $-Q$ is (semi)positive definite;
4. *degenerate* if $\det(Q) = 0$;
5. *indefinite* if there exist $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ such that $\mathbf{u}^T Q \mathbf{u} > 0$ and $\mathbf{v}^T Q \mathbf{v} < 0$;

8.5.11 Comments on Definition 8.5.10. Clearly, every definite symmetric matrix (positive or negative) is also semidefinite. If a matrix is indefinite, a simple application of the existence theorem for zeros of continuous functions proves that there exists a nonzero vector $\mathbf{w}$ that is Q -orthogonal to itself, that is, such that $\mathbf{w}^T Q \mathbf{w} = 0$. A definite matrix (positive or negative) is necessarily nondegenerate. Indeed, if $\mathbf{u}^T Q \mathbf{v} = 0$ for every $\mathbf{u} \in \mathbb{R}^n$, then, in particular with $\mathbf{u} = \mathbf{v}$, one has $\mathbf{v}^T Q \mathbf{v} = 0$. Since Q is positive definite, it follows that $\mathbf{v} = \mathbf{0}$.

8.5.12 Every diagonal matrix, of the form $D := (a_1 \mathbf{e}_1, \dots, a_n \mathbf{e}_n)$, is obviously symmetric. The classification of diagonal symmetric matrices is particularly simple. Indeed it is nondegenerate if and only if no element on the main diagonal is zero. Moreover it is positive definite if and only if the elements on the diagonal are all positive, negative definite if and only if the diagonal elements are all negative, and indefinite if both positive and negative elements appear on the diagonal.

8.5.13 Esempio. Consider the following diagonal matrices, viewed as symmetric matrices:

$$D_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad D_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}; \quad D_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad D_4 = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

D_1 corresponds to the Euclidean scalar product on $\mathbb{R}^4$; the matrix D_2 is indefinite and defines the Minkowski scalar product of the theory of special relativity; the matrix D_3 is positive semidefinite. Indeed $\mathbf{u}^T Q \mathbf{u} \geq 0$, but $\mathbf{e}_2^T Q \mathbf{e}_2 = 0$. Note that it is degenerate; finally D_4 is negative semidefinite.

8.5.14 Proposition. *Every symmetric matrix $Q \in \mathbb{R}^{n \times n}$ is congruent to at least one diagonal matrix.*

It is enough to prove that there exists a basis $\mathcal{B} = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ such that the product $\mathbf{b}_i^T \cdot Q \cdot \mathbf{b}_j$ is zero if $i \neq j$. Observe that if $n = 1$ the property is obviously verified. We now show that if every symmetric $(n-1) \times (n-1)$ matrix is congruent to a diagonal matrix, then every symmetric $n \times n$ matrix is as well. Let then Q be a symmetric $n \times n$ matrix. If the matrix is zero, it is already in diagonal form and there is nothing to prove. If $Q \neq \mathbf{0}_{n \times n}$, then there exists $\mathbf{b}_1 \in \mathbb{R}^n$ such that $\mathbf{b}_1^T Q \mathbf{b}_1 \neq 0$. Consider $\mathbf{b}_1^\perp_Q$, the subspace of $\mathbb{R}^{n-1}$ of vectors Q -orthogonal to $\mathbf{b}_1$. Then by Proposition 8.5.9:

$$\mathbb{R}^n = [\mathbf{b}_1] \oplus [\mathbf{b}_1]^\perp_Q.$$

Let $(\mathbf{b}_2, \dots, \mathbf{b}_n)$ be a basis of $\mathbf{b}_1^\perp_Q$ and let Q_1 be the symmetric matrix defined by

$$Q_1(i, j) = \mathbf{b}_i^T Q \mathbf{b}_j, \quad 1 \leq i, j \leq n.$$

Note that Q_1 is congruent to Q and that $Q_1(1, j) = Q_1(i, 1) = 0$ if i and j are different from 1. Let now Q_2 be the square $(n-1) \times (n-1)$ matrix defined by:

$$Q_2(i, j) = \mathbf{b}_i^T Q_1 \mathbf{b}_j = \mathbf{b}_i^T Q \mathbf{b}_j, \quad 2 \leq i \leq n.$$

This matrix is clearly symmetric. By induction, it is congruent to a diagonal matrix, that is, there exists a basis $(\mathbf{c}_2, \dots, \mathbf{c}_n)$ of $[\mathbf{b}_1]^\perp_Q$ such that $\mathbf{c}_i^T Q_2 \mathbf{c}_j = 0$ if $i \neq j$. Putting $P = (\mathbf{b}_1, \mathbf{c}_2, \dots, \mathbf{c}_n)$, it follows that $P^T Q_1 P$ is diagonal and congruent to Q_1 , which in turn is congruent to Q . Therefore Q is congruent to a diagonal matrix.

■

8.5.15 Theorem (Sylvester's law of inertia). *Let Q be congruent to two diagonal matrices D_1 and D_2 . Then D_1 and D_2 have the same number of positive and negative diagonal entries (and therefore also the same number of zero elements).*

Proof. Let $(\mathbf{b}_1, \dots, \mathbf{b}_n)$ and $(\mathbf{c}_1, \dots, \mathbf{c}_n)$ be bases such that

$$D_1(i, j) = \mathbf{b}_i^T Q \mathbf{b}_j \quad \text{and} \quad D_2(i, j) = \mathbf{c}_i^T Q \mathbf{c}_j.$$

Suppose that D_1 has p positive diagonal entries and n negative diagonal entries and that D_2 has p' positive diagonal entries and m' negative entries. Clearly $p + m =$

$p' + m'$ is equal to the rank of the symmetric matrix Q , since congruent matrices have the same rank. It is not restrictive to suppose that the bases are chosen so that

$$D_1(i, i) > 0 \quad \text{for} \quad 1 \leq i \leq p$$

and

$$D_2(i, i) > 0 \quad \text{for} \quad 1 \leq i \leq p'.$$

Let $W_1 := [\mathbf{b}_1, \dots, \mathbf{b}_p]$ and $W_2 := [\mathbf{c}_{p'+1}, \dots, \mathbf{c}_n]$. Then $W_1 \cap W_2 = \mathbf{0}$. Hence:

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 = p + n - p' \leq n$$

whence $p - p' \leq 0$. Exchanging the roles of p' and p , one obtains $p' \leq p$, and hence $p = p'$, which is what was to be proved. ■

8.5.16 Definition. *The signature of a real symmetric matrix is the pair (p, q) , where p is the number of positive entries and q is the number of negative entries of any diagonal matrix congruent to it.*

8.5.17 Proposition. *Let D be a diagonal matrix congruent to a symmetric matrix Q . Then Q is positive or negative (semi)definite, or indefinite, if and only if D is.*

Proof. We will show that if D is positive definite, then so is Q . The other cases are analogous and are left to the reader. Let $\mathbf{u} \in \mathbb{R}^n$ be arbitrary. By hypothesis there exists a matrix $P \in GL_n(\mathbb{R})$ such that $Q = P^T D P$. Thus

$$\mathbf{u}^T Q \mathbf{u} = \mathbf{u}^T (P^T D P) \mathbf{u} = (P \mathbf{u})^T D \cdot (P \mathbf{u}) \geq 0$$

because D is positive semidefinite. Moreover $\mathbf{u}^T Q \mathbf{u} = 0$ if and only if $(P \mathbf{u})^T D \cdot (P \mathbf{u}) = 0$, that is, if and only if $P \mathbf{u} = \mathbf{0}$, because D is positive definite, that is, if and only if $\mathbf{u} = \mathbf{0}$, since P is invertible. ■

8.5.18 At this point the problem of classifying real symmetric matrices is solved. The spectral theorem ensures that every symmetric matrix is diagonalizable over the reals by means of an orthogonal matrix P . In other words, there exists $P \in SO(n)$ such that

$$D = P^{-1} Q P = P^T Q P.$$

Thus Q is congruent to the diagonal matrix having the eigenvalues of Q on the main diagonal. The signature of a symmetric matrix therefore coincides with the pair (p, q) , where p is the number of positive eigenvalues and q the number of negative eigenvalues.

8.6 Bilinear forms

8.6.1 Let V be a $\mathbb{K}$ -vector space. A bilinear form defined on V is a function

$$\phi : V \times V \rightarrow \mathbb{K},$$

which is linear with respect to the first and second argument, that is, such that for every choice of $\lambda, \mu \in \mathbb{K}$ and $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{K}^n$ one has, respectively:

$$\phi(\lambda \mathbf{u} + \mu \mathbf{v}, \mathbf{w}) = \lambda \phi(\mathbf{u}, \mathbf{w}) + \mu \phi(\mathbf{v}, \mathbf{w}), \quad (8.12)$$

and

$$\phi(\mathbf{u}, \lambda \mathbf{v} + \mu \mathbf{w}) = \lambda \phi(\mathbf{u}, \mathbf{v}) + \mu \phi(\mathbf{u}, \mathbf{w}). \quad (8.13)$$

8.6.2 Esempio. Let $V = \mathbb{K}^n$ and $A \in \mathbb{K}^{n \times n}$. The function:

$$\phi_A(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T A \mathbf{v}$$

is a bilinear form. This is an obvious consequence of the bilinear and associative properties of matrix multiplication. If $\mathbf{u}(i) = u_i$ and $\mathbf{v}(j) = v_j$ for all $1 \leq i, j \leq n$, one has

$$\phi_A(\mathbf{u}, \mathbf{v}) = \left(\sum_{i=1}^n u_i \mathbf{e}_i \right)^T \cdot A \left(\sum_{j=1}^n v_j \mathbf{e}_j \right) = \sum_{i=1}^n \sum_{j=1}^n u_i (\mathbf{e}_i \cdot A \cdot \mathbf{e}_j) v_j = \sum_{i=1}^n \sum_{j=1}^n u_i \cdot A(i, j) \cdot v_j.$$

In other words: a bilinear form is a homogeneous polynomial expression of second degree in the components (u_i) and (v_j) of two vectors.

8.6.3 Esempio. Let $V = \mathbb{R}^3$. Then

$$\phi(\mathbf{u}, \mathbf{v}) = 3u_1v_1 + 2u_1v_3 + 2u_2v_3 + 3u_3v_1 + 5u_3v_3$$

is a bilinear form because it is a homogeneous polynomial expression of second degree, of first degree in the variables u (holding the variables v constant) and of first degree in the variables v (holding the variables u constant). It can be written as:

$$\phi(\mathbf{u}, \mathbf{v}) = (u_1, u_2, u_3) \cdot A \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

where

$$A = \begin{pmatrix} 3 & 0 & 2 \\ 0 & 0 & 2 \\ 3 & 0 & 5 \end{pmatrix}.$$

Note that $\phi(\mathbf{u}, \mathbf{v}) \neq \phi(\mathbf{v}, \mathbf{u})$.

8.6.4 Esempio. Let $V = \mathbb{C}^2$. The function

$$\phi(\mathbf{u}, \mathbf{v}) = (2 + i)\mathbf{u}(1)\mathbf{v}(2) - 4\mathbf{u}(2)\mathbf{v}(1) + 3\mathbf{u}(2)\mathbf{v}(2)$$

is a bilinear form over the complexes, in the sense that (8.12) and (8.13) hold for every $\lambda, \mu \in \mathbb{C}$. Instead, the function

$$\phi(\mathbf{u}, \mathbf{v}) = (2 + i)\mathbf{u}(1)\overline{\mathbf{v}(2)} - 4\mathbf{u}(2)\overline{\mathbf{v}(1)} + 3\mathbf{u}(2)\overline{\mathbf{v}(2)} \quad (8.14)$$

is $\mathbb{C}$ -linear with respect to the first argument but not with respect to the second. The reader can verify, in this case, that

$$\phi(\mathbf{u}, \lambda\mathbf{v} + \mu\mathbf{w}) = \bar{\lambda}\phi(\mathbf{u}, \mathbf{v}) + \bar{\mu}\phi(\mathbf{u}, \mathbf{w}). \quad (8.15)$$

One calls (8.15) *antilinearity with respect to the second argument*. However, (8.14) is $\mathbb{R}$ -linear, that is, (8.12) and (8.13) are verified for every $\lambda, \mu \in \mathbb{R}$.

8.6.5 Esempio. The map $\det : \mathbb{K}^2 \times \mathbb{K}^2 \rightarrow \mathbb{K}$ defined by

$$\det(\mathbf{u}, \mathbf{v}) = \mathbf{u} \wedge \mathbf{v} = \mathbf{u}(1)\mathbf{v}(2) - \mathbf{u}(2)\mathbf{v}(1),$$

is the determinant of ordered pairs of vectors of $\mathbb{K}^2$, an antisymmetric bilinear form, in the sense that $\det(\mathbf{u}, \mathbf{v}) = -\det(\mathbf{v}, \mathbf{u})$.

8.6.6 The set $\text{Bil}_{\mathbb{K}}(V)$ of all $\mathbb{K}$ -bilinear forms is a vector space, since every linear combination

$$\lambda\phi + \mu\psi$$

of bilinear forms is a bilinear form (verify it, checking that (8.12) and (8.13) hold).

8.6.7 A bilinear form $\phi : V \times V \rightarrow \mathbb{K}$ is called symmetric if and only if

$$\phi(\mathbf{u}, \mathbf{v}) = \phi(\mathbf{v}, \mathbf{u}),$$

for every pair of vectors $(\mathbf{u}, \mathbf{v}) \in V \times V$.

8.6.8 Esempio. Let Q be a symmetric $n \times n$ matrix and let $\phi_Q : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ be defined by:

$$\phi_Q(\mathbf{u}, \mathbf{v}) = \mathbf{u}^T Q \mathbf{v},$$

for every pair of vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. Then the function ϕ_Q is symmetric bilinear. Indeed, ϕ_Q is linear with respect to the first argument, since for every choice of $\mathbf{u}_1, \mathbf{u}_2, \mathbf{v} \in \mathbb{R}^n$ and every $\lambda, \mu \in \mathbb{R}$ one has:

$$\begin{aligned}
\phi_Q(\lambda \mathbf{u}_1 + \mu \mathbf{u}_2, \mathbf{v}) &= (\lambda \mathbf{u}_1 + \mu \mathbf{u}_2)^T \cdot Q \cdot \mathbf{v} = && \text{(definition of } \phi_Q) \\
&= (\lambda \mathbf{u}_1^T Q + \mu \mathbf{u}_2^T Q) \cdot \mathbf{v} = && \text{(linearity (??) of transposition and} \\
&&& \text{bilinearity (??) of the product} \\
&&& \text{of matrices)} \\
&= \lambda \mathbf{u}_1^T Q \mathbf{v} + \mu \mathbf{u}_2^T Q \mathbf{v} = && \text{(bilinearity (??) of the product} \\
&&& \text{of matrices)} \\
&= \lambda \phi_Q(\mathbf{u}_1, \mathbf{v}) + \mu \phi_Q(\mathbf{u}_2, \mathbf{v}) = && \text{(definition of } \phi_Q)
\end{aligned}$$

Moreover ϕ_Q is symmetric because

$$\begin{aligned}
\phi_Q(\mathbf{v}, \mathbf{u}) &= \mathbf{v}^T Q \mathbf{u} = && \text{(definition of } \phi_Q) \\
&= (\mathbf{v}^T Q \mathbf{u})^T = && \text{(transpose of a } 1 \times 1 \text{ matrix)} \\
&= \mathbf{u}^T Q \mathbf{v}^{TT} = && \text{(transpose (??) of a product of matrices)} \\
&= \mathbf{u}^T Q \mathbf{v} = && \text{(using the fact that } \mathbf{v}^{TT} = \mathbf{v}). \\
&= \phi_Q(\mathbf{u}, \mathbf{v}) && \text{(definition of } \phi_Q)
\end{aligned}$$

From linearity with respect to the first argument and symmetry follows linearity with respect to the second argument:

$$\begin{aligned}
\phi_Q(\mathbf{u}, \lambda \mathbf{v}_1 + \mu \mathbf{v}_2) &= \phi_Q(\lambda \mathbf{v}_1 + \mu \mathbf{v}_2, \mathbf{u}) && \text{(symmetry)} \\
&= \lambda \phi_Q(\mathbf{v}_1, \mathbf{u}) + \mu \phi_Q(\mathbf{v}_2, \mathbf{u}) = && \text{(linearity with respect to the} \\
&&& \text{first argument)} \\
&= \lambda \phi_Q(\mathbf{u}, \mathbf{v}_1) + \mu \phi_Q(\mathbf{u}, \mathbf{v}_2) && \text{(symmetry)}
\end{aligned}$$

The symmetric bilinear form ϕ_Q is said to be *associated* with the symmetric matrix Q . The associated quadratic form q_ϕ is called nondegenerate if Q is nondegenerate. It is called positive or negative (semi)definite if Q is positive or negative (semi)definite. It is called indefinite if Q is.

Let ϕ be a symmetric bilinear form and let

$$q_\phi : V \rightarrow \mathbb{K}$$

be the associated quadratic form defined by $q_\phi(\mathbf{u}) = \phi(\mathbf{u}, \mathbf{u})$.

8.6.9 Proposition. *The polarization formula holds:*

$$q_\phi(\lambda \mathbf{u} + \mu \mathbf{v}) = \lambda^2 q_\phi(\mathbf{u}) + 2\lambda\mu\phi(\mathbf{u}, \mathbf{v}) + \mu^2 q_\phi(\mathbf{v}), \quad (8.16)$$

for every $\mathbf{u}, \mathbf{v} \in V$ and $\lambda, \mu \in \mathbb{K}$.

Proof. The proof is purely formal:

$$\begin{aligned} q_\phi(\lambda \mathbf{u} + \mu \mathbf{v}) &= \phi(\lambda \mathbf{u} + \mu \mathbf{v}, \lambda \mathbf{u} + \mu \mathbf{v}) = && \text{(definition of } q_\phi) \\ &= \lambda\phi(\mathbf{u}, \lambda \mathbf{u} + \mu \mathbf{v}) + \mu\phi(\mathbf{v}, \lambda \mathbf{u} + \mu \mathbf{v}) = && \begin{array}{l} \text{(linearity of } \phi \\ \text{with respect to the first} \\ \text{argument)} \end{array} \\ &= \lambda^2\phi(\mathbf{u}, \mathbf{u}) + \lambda\mu\phi(\mathbf{u}, \mathbf{v}) + \lambda\mu\phi(\mathbf{v}, \mathbf{u}) + \mu^2\phi(\mathbf{v}, \mathbf{v}) = && \begin{array}{l} \text{(linearity of } \phi \\ \text{with respect to the second} \\ \text{argument)} \end{array} \\ &= \lambda^2\phi(\mathbf{u}, \mathbf{u}) + 2\lambda\mu\phi(\mathbf{u}, \mathbf{v}) + \mu^2\phi(\mathbf{v}, \mathbf{v}) = && \text{(symmetry of } \phi) \\ &= \lambda^2 q_\phi(\mathbf{u}) + 2\lambda\mu\phi(\mathbf{u}, \mathbf{v}) + \mu^2 q_\phi(\mathbf{v}). && \text{(definition of } q_\phi) \end{aligned}$$

and equation (8.16) is thereby proved. $\blacksquare$

The polarization formula is often used to give a definition of quadratic form on V .

8.6.10 Definition. *A function $q : V \rightarrow \mathbb{K}$ is called a quadratic form on V if and only if*

$$q(\lambda \mathbf{u}) = \lambda^2 q(\mathbf{u}) \quad \forall \lambda \in \mathbb{K} \quad \forall \mathbf{u} \in V; \quad (8.17)$$

and the function

$$\phi_q(\mathbf{u}, \mathbf{v}) = \frac{1}{2} (q(\mathbf{u} + \mathbf{v}) - q(\mathbf{u}) - q(\mathbf{v})) \quad (8.18)$$

is a symmetric bilinear form.

8.6.11 Assigning a symmetric bilinear form ϕ on V and a basis $\mathcal{B} = (\mathbf{b}_1, \dots, \mathbf{b}_n)$ of V determines a symmetric square matrix $M_\phi^\mathcal{B}$, associated with ϕ and $\mathcal{B}$, defined by:

$$M_\phi^\mathcal{B}(i, j) = \phi(\mathbf{b}_i, \mathbf{b}_j).$$

Thus the symmetric bilinear form ϕ can be written as

$$\phi(\mathbf{u}, \mathbf{v}) = (\mathbf{u})_\mathcal{B}^T M_\phi^\mathcal{B} (\mathbf{v})_\mathcal{B}.$$

The symmetric matrix $M_\phi^{\mathcal{B}}$ is called the symmetric matrix associated with the symmetric bilinear form ϕ , or equivalently with the associated quadratic form q_ϕ . The reader should verify that if $M_\phi^{\mathcal{C}}$ is the matrix associated with ϕ in another basis $\mathcal{C}$ of V , it is congruent to $M_\phi^{\mathcal{B}}$, since:

$$M_\phi^{\mathcal{C}} = (P^{\mathcal{C},\mathcal{B}})^T M_\phi^{\mathcal{B}} P^{\mathcal{C},\mathcal{B}}.$$

8.7 Real quadratic forms and their classification

Let now V be an arbitrary real vector space of dimension n and let $q : V \rightarrow \mathbb{R}$ be a real quadratic form. Given a basis $\mathcal{B}$, it can be written in the form

$$q(\mathbf{u}) = (\mathbf{u})_{\mathcal{B}}^T \cdot Q \cdot (\mathbf{u})_{\mathcal{B}}$$

where $Q = M_{\phi_q}^{\mathcal{B}}$. Then q is called *(semi)definite* (positive or negative) if and only if Q is a real symmetric *(semi)definite* matrix (positive or negative). It is called *degenerate* if and only if Q has zero determinant. It is called *indefinite* if Q is an indefinite matrix. The *signature* of ϕ is the signature of Q . In particular, the number of negative eigenvalues is called the *index*, and is extremely important in Morse theory [?]. These notions do not depend on the basis $\mathcal{B}$ used to obtain the matrix Q , because changing basis changes the matrix Q into a congruent symmetric matrix which, by Sylvester's theorem ?? on the inertia of a quadratic form, has the same signature, determined precisely by the number of positive or negative eigenvalues of any symmetric matrix Q representing the quadratic form q .

8.7.1 Esempio. To establish whether the quadratic form $q(x, y, z) = x^2 - 2xz + 2yz$ is positive or negative definite (or semidefinite), or indefinite, one writes any matrix associated with q . Indeed q can be written as:

$$q(x, y, z) = (x, y, z) \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 1 \\ -1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

The matrix appearing in (8.7.1) has characteristic polynomial

$$P_Q(T) = T^3 - T^2 - 2T + 1.$$

Since the polynomial has two sign changes and one sign preservation among its coefficients, there are two positive eigenvalues and one negative eigenvalue. The quadratic form is indefinite.

8.8 An interlude: the Singular Value Decomposition (SVD)

8.8.1 For uniformity of exposition, in this section we shall limit ourselves to working over the real field, thinking of matrices with real entries. Orthogonal matrices will be real orthogonal matrices: these are the ingredients we need in order to introduce, very briefly, leaving greater burdens to the parallel Numerical Analysis course, the notion of *singular value decomposition* of a rectangular matrix. We warn the reader, however, that nothing would change substantially if we replaced the real field $\mathbb{R}$ with the complex field $\mathbb{C}$, symmetric matrices with Hermitian ones and orthogonal transformations with unitary ones. Indeed, we propose to readers the exercise of rewriting everything that follows using complex numbers instead of real numbers.

8.8.2 If $A \in \mathbb{K}^{n \times n}$ is a square matrix with entries in $\mathbb{K}$, its rank evidently coincides with the number of its nonzero eigenvalues, counting each with its own multiplicity. Even if the matrix were not diagonalizable, an alternative canonical form could be obtained, for example, by using a triangular one. The SVD somehow responds to the need and desire to want/be able to associate to a nonsquare matrix a sequence of scalars capable of giving information about the matrix itself, just as eigenvalues do for square matrices. Let then m and n be fixed integers and let $A \in \mathbb{K}^{m \times n}$. As seen in Example ??, the matrices $AA^T \in \mathbb{R}^{m \times m}$ and $A^T A \in \mathbb{R}^{n \times n}$ are symmetric and positive semidefinite.

8.8.3 Proposition. *Let $S \in \mathbb{K}^{m \times n}$ be a matrix such that $i \neq j \implies S(i, j) = 0$. Then SS^T and $S^T S$ are diagonal matrices.*

Proof. Indeed $S^T S$ and SS^T are square matrices. To show that they are diagonal, it is sufficient to show that $i \neq j$ implies $(SS^T)(i, j) = 0$ and $(S^T S)(i, j) = 0$. We prove this in the first case, leaving the second, entirely analogous, to the care of the reader (better if both male and female readers). Using (??) one has:

$$(SS^T)(i, j) = R_i(S)C_j(S^T) = R_i(S)R_j(S)^T.$$

Now the i -th row of S has all entries zero, except possibly the i -th one. The same holds for the j -th row. Therefore the product of the i -th row by the transpose of the j -th is certainly zero if $i \neq j$. This proves that SS^T (and, analogously, $S^T S$) are diagonal matrices. ■

Clearly, being diagonal and positive semidefinite, the eigenvalues will all be non-negative. To emphasize this fact they will be denoted as squares

$$\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_p^2 \quad (p = m \text{ or } p = n)$$

listed in decreasing order.

8.8.4 Proposition Suppose that there exist three matrices $P \in O(m)$, $Q \in O(n)$ and $S_A \in \mathbb{R}^{m \times n}$ such that

$$A = P^T \cdot S_A \cdot Q. \quad (8.19)$$

Then the columns of P and Q are eigenvectors of $A^T A$ and AA^T , respectively.

Proof. Indeed, if (8.19) holds, then:

$$A^T A = Q^T S_A^T P \cdot P^T S_A Q = Q^T S_A^T S_A Q$$

hence

$$Q(A^T A)Q^T = S_A^T S_A.$$

Since $S_A^T S_A$ is diagonal, it follows that Q^T must be a matrix of eigenvectors of $A^T A$. Similarly

$$AA^T = P^T \cdot S_A \cdot QQ^T S_A^T P = P^T S_A S_A^T P$$

that is,

$$PAA^T P^T = S_A S_A^T$$

and P is thus a matrix of eigenvectors of AA^T . ■

If it exists, expression (8.19) is called a *singular value decomposition* of A . The rows of the matrix P are called *left singular vectors*, the columns of the matrix Q *right singular vectors*, and the square roots of the nonzero eigenvalues of AA^T or $A^T A$ are called singular values. They can also be characterized as follows.

8.8.5 Let $A \in \mathbb{R}^{m \times n}$. A scalar σ is called a singular value of A if there exists a pair of nonzero vectors $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$A\mathbf{u} = \sigma\mathbf{v} \quad \text{and} \quad \mathbf{v}^T A = \sigma\mathbf{u}^T. \quad (8.20)$$

In that case $\mathbf{v}$ is called a *left singular vector*, while $\mathbf{u}$ is called a *right singular vector*. Note that the second of (8.20) can be written as

$$A^T \mathbf{v} = \sigma\mathbf{u}$$

and hence, applying A to both sides:

$$AA^T \mathbf{v} = \sigma A\mathbf{u} = \sigma^2 \mathbf{v}$$

from which one deduces that $\mathbf{v}$ is an eigenvector of AA^T with respect to the eigenvalue σ^2 . Similarly, from the equation $A\mathbf{u} = \sigma\mathbf{v}$ one obtains the equality $\mathbf{u}^T A^T = \sigma\mathbf{v}^T$. Applying A to both sides:

$$\mathbf{u}^T A^T A = \sigma\mathbf{v}^T A = \sigma^2 \mathbf{u}^T$$

which is the same as saying that $\mathbf{u}$ is an eigenvector corresponding to $A^T A$ with respect to σ^2 . Thus one sees that the *singular values* of A are square roots of the eigenvalues of AA^T or of $A^T A$.

8.8.6 Proposition. *Every rectangular matrix $A \in \mathbb{R}^{m \times n}$ admits a singular value decomposition.*

Proof. We must prove the existence of two orthogonal matrices $P \in O(m)$, $Q \in O(n)$ and a diagonal $S_A \in \mathbb{R}^{m \times n}$ (in the sense that $S_A(i, j) = 0$ if $i \neq j$) satisfying equality (8.19). It is not restrictive to suppose that $m \leq n$. Indeed, if $m \geq n$, then $A^T \in \mathbb{R}^{n \times m}$ with $n \leq m$, and if there existed a decomposition (8.19) of the type $A^T = \tilde{P}^T \tilde{S}_A \tilde{Q}$, then

$$A = \tilde{Q}^T \tilde{S}_A^T \tilde{P}$$

would be an SVD for A . Thus we shall simply suppose that $m \leq n$. Hence AA^T is an $m \times m$ square matrix. Being symmetric, there exists $P \in O(m)$ such that

$$PAA^T P^T$$

is a diagonal matrix, that is, for all $1 \leq i, j \leq m$

$$(\epsilon_i P A) \cdot (A^T P^T e_j) = \sigma_i^2 \delta_{ij}.$$

In particular, $\mathbf{u}_j := \frac{A^T P^T e_j}{\sigma_j}$ are mutually orthogonal columns of $\mathbb{R}^n$ of modulus 1.

Let Q^T be the matrix $(\mathbf{u}_1, \dots, \mathbf{u}_r, \mathbf{u}_{r+1}, \dots, \mathbf{u}_n)$, whose columns complete the vectors $(\mathbf{u}_1, \dots, \mathbf{u}_r)$ to an orthonormal basis of $\mathbb{R}^n$. Then, if $1 \leq i \leq m$,

$$\epsilon_i P A Q^T e_j = \epsilon_i P A \cdot \mathbf{u}_j = \delta_{ij}$$

for all $j \in \{1, \dots, n\}$. And this proves that

$$A = P^T S_A Q$$

is an SVD, where we have put $S_A = P A Q^T$. ■

The diagonal elements of S_A are the square roots of the eigenvalues of AA^T . They are called the *singular values* of A .

8.8.7 Esempio. Let $A := (1, 2)$. We construct the singular value decomposition. First of all compute

$$AA^T = (1, 2) \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 5.$$

An orthogonal matrix that diagonalizes (5) is the matrix (1). Therefore $Q = (\mathbf{u}_1, \mathbf{u}_2)$, where

$$\mathbf{u}_1 = \begin{pmatrix} \frac{\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} \end{pmatrix}$$

and $\mathbf{u}_2$ is any vector of modulus 1 orthogonal to $\mathbf{u}_1$. Thus, for example,

$$(1, 2) = (\sqrt{5}, 0) \begin{pmatrix} \frac{\sqrt{5}}{5} & \frac{2\sqrt{5}}{5} \\ \frac{2\sqrt{5}}{5} & -\frac{\sqrt{5}}{5} \end{pmatrix}.$$

8.8.8 The proof of Proposition (8.8.6) in fact suggests an algorithm for obtaining the SVD of a matrix (which is not unique, because the choice of P and Q is not unique). In practice one starts from the matrix A , from which one constructs AA^T and $A^T A$. These are symmetric matrices with nonnegative nonzero eigenvalues. The square roots are the singular values. Thus

$$A = P^T S_A Q$$

where on the diagonal $i = j$ of the matrix S_A one reads the singular values (zero or nonzero), while P and Q are the orthogonal matrices that diagonalize, respectively, the matrices AA^T and $A^T A$.

8.8.9 Esempio. Let

$$A := \begin{pmatrix} 2 & 1 & 5 \\ 1 & 1 & 2 \end{pmatrix}.$$

We will find the SVD with the help of the library functions of MATLAB[®]. For example, compute AA^T and $A^T A$. One has

$$AA^T = \begin{pmatrix} 30 & 13 \\ 13 & 6 \end{pmatrix} \quad \text{and} \quad A^T A = \begin{pmatrix} 5 & 3 & 12 \\ 3 & 2 & 7 \\ 12 & 7 & 29 \end{pmatrix}.$$

With the instruction `eig` one finds:

$$\text{eig}(AA') = (0, 3082, \quad 35, 6918)$$

which are the approximate values of the eigenvalues of AA^T . Similarly,

$$\text{eig}(A'A) = (0, \quad 0, 3082, \quad 35, 6918).$$

It is clear that S_A will be given by:

$$S_A = \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \end{pmatrix}$$

where σ_1 and σ_2 are the square roots of the values found above. Naturally we can use the function $\mathbf{s} = \text{svd}(A)$. To determine P , S and Q it is enough to set the instruction

$$[P, S, Q] = \text{svd}(A)$$

obtaining the following output

$$P = \begin{pmatrix} -0.9160 & -0.4011 \\ -0.4011 & 0.9160 \end{pmatrix}$$

$$S = \begin{pmatrix} 5.9743 & 0 & 0 \\ 0 & 0.5552 & 0 \end{pmatrix}$$

$$Q = \begin{pmatrix} -0.3738 & 0.2052 & -0.9045 \\ -0.2205 & 0.9276 & 0.3015 \end{pmatrix}$$

Readers should verify the validity of the above calculations by opening a MATLAB® sheet themselves.

Out of curiosity for the reader, we recall that the SVD is often used in *data analysis* with important applications to sociology, psychometrics and the study of electoral flows. For an exposition aimed at, and inspired by, this type of problem, see the intriguing text by Luca Ricolfi [?]. And now, in the next section, a topic that seems to have nothing to do with what has been presented so far, except for the fact that in mathematics everything has to do with everything. And yet it does have

something to do with it, because the SVD solves an *optimization* problem, which we will not address, namely that of finding a vector of unit norm such that its product by a given matrix has minimum modulus. Enjoy the reading!

8.9 Exercises

8.9.1 To become familiar with complex numbers in MATLAB[®], try using the instructions

$[P, D] = \text{eig}(A)$ to diagonalize the matrix

$$\mathbb{I} := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

Compute its singular value decomposition.

Soluzione: [?, 11.8.1]

8.9.2 Prove that the matrix

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

is diagonalizable and diagonalize it by means of an orthogonal matrix P . Also compute its singular value decomposition and compare the results.

Soluzione: [?, 11.8.2]

8.9.3 Let the matrices be given:

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} \quad \text{and} \quad B := \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ -1 & -2 & -3 \end{pmatrix}.$$

Which of the two is diagonalizable? And why? For the non-diagonalizable matrix, determine the singular value decomposition.

Soluzione: [?, 11.8.3]

8.9.4 Determine the sign of the quadratic form

$$q(x, y, z) = x^2 + 4xy + 6y^2 - 10yz + 8z^2.$$

Is the quadratic form q degenerate?

Soluzione: [?, 11.9.1]

8.9.5 Determine the sign of the quadratic form

$$q(x, y, z) = (2x - y + z)(x + y - 3z).$$

Is the quadratic form q degenerate?

Soluzione: [?, 11.9.1]