

Chapter 18

Special Classes of Endomorphisms

In this chapter we discuss some classes of endomorphisms (or square matrices) whose eigenvalues and eigenvectors have special properties. Such properties only exist under further assumptions, and in this chapter our assumptions concern the relationship between the given endomorphism and its adjoint endomorphism. Thus, we focus on Euclidean or unitary vector spaces. This leads to the classes of normal, orthogonal, unitary and selfadjoint endomorphisms. Each of these classes has a natural counterpart in the set of square (real or complex) matrices.

18.1 Normal Endomorphisms

We start with the definition of a normal¹ endomorphism or matrix.

Definition 18.1 Let \mathcal{V} be a finite dimensional Euclidean or unitary vector space. An endomorphism $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ is called *normal* if $f \circ f^{ad} = f^{ad} \circ f$. A matrix $A \in \mathbb{R}^{n,n}$ or $A \in \mathbb{C}^{n,n}$ is called *normal* if $A^T A = A A^T$ or $A^H A = A A^H$, respectively.

For all $z \in \mathbb{C}$ we have $\bar{z}z = |z|^2 = z\bar{z}$. The property of normality can therefore be interpreted as a generalization of this property of complex numbers.

We will first study the properties of normal endomorphisms on a finite dimensional unitary vector space \mathcal{V} . Recall the following results:

- (1) If B is an orthonormal basis of \mathcal{V} and if $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$, then $([f]_{B,B})^H = [f^{ad}]_{B,B}$ (cp. Theorem 13.12).
- (2) Every $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ can be unitarily triangulated (cp. Corollary 14.20, Schur's theorem). This does not hold in general in the Euclidean case, since not every real polynomial decomposes into linear factors over \mathbb{R} .

¹This term was introduced by Otto Toeplitz (1881–1940) in 1918 in the context of bilinear forms.

Using these results we obtain the following characterization of normal endomorphisms on a unitary vector space.

Theorem 18.2 *If \mathcal{V} is a finite dimensional unitary vector space, then $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ is normal if and only if there exists an orthonormal basis B of \mathcal{V} such that $[f]_{B,B}$ is a diagonal matrix, i.e., f is unitarily diagonalizable.*

Proof Let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ be normal and let B be an orthonormal basis of \mathcal{V} such that $R := [f]_{B,B}$ is an upper triangular matrix. Then $R^H = [f^{ad}]_{B,B}$, and from $f \circ f^{ad} = f^{ad} \circ f$ we obtain

$$RR^H = [f \circ f^{ad}]_{B,B} = [f^{ad} \circ f]_{B,B} = R^H R.$$

We now show by induction on $n = \dim(\mathcal{V})$ that R is diagonal. This is obvious for $n = 1$.

Let the assertion hold for an $n \geq 1$, and let $R \in \mathbb{C}^{n+1,n+1}$ be upper triangular with $RR^H = R^H R$. We write R as

$$R = \begin{bmatrix} R_1 & r_1 \\ 0 & \alpha_1 \end{bmatrix},$$

where $R_1 \in \mathbb{C}^{n,n}$ is upper triangular, $r_1 \in \mathbb{C}^{n,1}$, and $\alpha_1 \in \mathbb{C}$. Then

$$\begin{bmatrix} R_1 R_1^H + r_1 r_1^H & \bar{\alpha}_1 r_1 \\ \alpha_1 r_1^H & |\alpha_1|^2 \end{bmatrix} = RR^H = R^H R = \begin{bmatrix} R_1^H R_1 & R_1^H r_1 \\ r_1^H R_1 & r_1^H r_1 + |\alpha_1|^2 \end{bmatrix}.$$

From $|\alpha_1|^2 = r_1^H r_1 + |\alpha_1|^2$ we obtain $r_1^H r_1 = 0$, hence $r_1 = 0$ and $R_1 R_1^H = R_1^H R_1$. By the induction hypothesis, $R_1 \in \mathbb{C}^{n,n}$ is diagonal, and therefore

$$R = \begin{bmatrix} R_1 & 0 \\ 0 & \alpha_1 \end{bmatrix}$$

is diagonal as well.

Conversely, suppose that there exists orthonormal basis B of \mathcal{V} such that $[f]_{B,B}$ is diagonal. Then $[f^{ad}]_{B,B} = ([f]_{B,B})^H$ is diagonal and, since diagonal matrices commute, we have

$$[f \circ f^{ad}]_{B,B} = [f]_{B,B} [f^{ad}]_{B,B} = [f^{ad}]_{B,B} [f]_{B,B} = [f^{ad} \circ f]_{B,B},$$

which implies $f \circ f^{ad} = f^{ad} \circ f$, and hence f is normal. □

The application of this theorem to the unitary vector space $\mathcal{V} = \mathbb{C}^{n,1}$ with the standard scalar product and a matrix $A \in \mathbb{C}^{n,n}$ viewed as element of $\mathcal{L}(\mathcal{V}, \mathcal{V})$ yields the following “matrix version”.

Corollary 18.3 *A matrix $A \in \mathbb{C}^{n,n}$ is normal if and only if there exists an orthonormal basis of $\mathbb{C}^{n,1}$ consisting of eigenvectors of A , i.e., A is unitarily diagonalizable.*

The following theorem presents another characterization of normal endomorphisms on a unitary vector space.

Theorem 18.4 *If \mathcal{V} is a finite dimensional unitary vector space, then $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ is normal if and only if there exists a polynomial $p \in \mathbb{C}[t]$ with $p(f) = f^{ad}$.*

Proof If $p(f) = f^{ad}$ for a polynomial $p \in \mathbb{C}[t]$, then

$$f \circ f^{ad} = f \circ p(f) = p(f) \circ f = f^{ad} \circ f,$$

and hence f is normal.

Conversely, if f is normal, then there exists an orthonormal basis B of \mathcal{V} , such that $[f]_{B,B} = \text{diag}(\lambda_1, \dots, \lambda_n)$. Furthermore,

$$[f^{ad}]_{B,B} = ([f]_{B,B})^H = \text{diag}(\bar{\lambda}_1, \dots, \bar{\lambda}_n).$$

Let $p \in \mathbb{C}[t]$ be a polynomial with $p(\lambda_j) = \bar{\lambda}_j$ for $j = 1, \dots, n$. Such a polynomial can be explicitly constructed using the Lagrange basis of $\mathbb{C}[t]_{\leq n-1}$ (cp. Exercise 10.12). Then

$$\begin{aligned} [f^{ad}]_{B,B} &= \text{diag}(\bar{\lambda}_1, \dots, \bar{\lambda}_n) = \text{diag}(p(\lambda_1), \dots, p(\lambda_n)) = p(\text{diag}(\lambda_1, \dots, \lambda_n)) \\ &= p([f]_{B,B}) = [p(f)]_{B,B}, \end{aligned}$$

and hence also $f^{ad} = p(f)$. □

Several other characterizations of normal endomorphisms on a finite dimensional unitary vector space and of normal matrices $A \in \mathbb{C}^{n,n}$ can be found in the article [HorJ12] (see also Exercise 18.8).

We now consider the Euclidean case, where we focus on real square matrices. All the results can be formulated analogously for normal endomorphisms on a finite dimensional Euclidean vector space.

Let $A \in \mathbb{R}^{n,n}$ be normal, i.e., $A^T A = A A^T$. Then A also satisfies $A^H A = A A^H$ and when A is considered as an element of $\mathbb{C}^{n,n}$, it is unitarily diagonalizable, i.e., $A = SDS^H$ holds for a unitary matrix $S \in \mathbb{C}^{n,n}$ and a diagonal matrix $D \in \mathbb{C}^{n,n}$. Despite the fact that A has real entries, neither S nor D will be real in general, since A as an element of $\mathbb{R}^{n,n}$ may not be diagonalizable. For instance,

$$A = \begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} \in \mathbb{R}^{2,2}$$

is a normal matrix that is not diagonalizable (over \mathbb{R}). Considered as element of $\mathbb{C}^{2,2}$, it has the eigenvalues $1 + 2i$ and $1 - 2i$ and it is unitarily diagonalizable.

To discuss the case of real normal matrices in more detail, we first prove a “real version” of Schur’s theorem.

Theorem 18.5 For every matrix $A \in \mathbb{R}^{n,n}$ there exists an orthogonal matrix $U \in \mathbb{R}^{n,n}$ with

$$U^T A U = R = \begin{bmatrix} R_{11} & \dots & R_{1m} \\ & \ddots & \vdots \\ & & R_{mm} \end{bmatrix} \in \mathbb{R}^{n,n},$$

where for every $j = 1, \dots, m$ either $R_{jj} \in \mathbb{R}^{1,1}$ or

$$R_{jj} = \begin{bmatrix} r_1^{(j)} & r_2^{(j)} \\ r_3^{(j)} & r_4^{(j)} \end{bmatrix} \in \mathbb{R}^{2,2} \text{ with } r_3^{(j)} \neq 0.$$

In the second case R_{jj} has, considered as complex matrix, a pair of complex conjugate eigenvalues of the form $\alpha_j \pm i\beta_j$ with $\alpha_j \in \mathbb{R}$ and $\beta_j \in \mathbb{R} \setminus \{0\}$. The matrix R is called a real Schur form of A .

Proof We proceed via induction on n . For $n = 1$ we have $A = [a_{11}] = R$ and $U = [1]$.

Suppose that the assertion holds for some $n \geq 1$ and let $A \in \mathbb{R}^{n+1,n+1}$ be given. We consider A as an element of $\mathbb{C}^{n+1,n+1}$. Then A has an eigenvalue $\lambda = \alpha + i\beta \in \mathbb{C}$, $\alpha, \beta \in \mathbb{R}$, corresponding to the eigenvector $v = x + iy \in \mathbb{C}^{n+1,1}$, $x, y \in \mathbb{R}^{n+1,1}$, and we have $Av = \lambda v$. Dividing this equation into its real and imaginary parts, we obtain the two real equations

$$Ax = \alpha x - \beta y \quad \text{and} \quad Ay = \beta x + \alpha y. \tag{18.1}$$

We have two cases:

Case 1: $\beta = 0$. Then the two equations in (18.1) are $Ax = \alpha x$ and $Ay = \alpha y$. Thus at least one of the real vectors x or y is an eigenvector corresponding to the real eigenvalue α of A . Without loss of generality we assume that this is the vector x and that $\|x\|_2 = 1$. We extend x by the vectors w_2, \dots, w_{n+1} to an orthonormal basis of $\mathbb{R}^{n+1,1}$ with respect to the standard scalar product. The matrix $U_1 := [x, w_2, \dots, w_{n+1}] \in \mathbb{R}^{n+1,n+1}$ then is orthogonal and satisfies

$$U_1^T A U_1 = \begin{bmatrix} \alpha & \star \\ 0 & A_1 \end{bmatrix}$$

for a matrix $A_1 \in \mathbb{R}^{n,n}$. By the induction hypothesis there exists an orthogonal matrix $U_2 \in \mathbb{R}^{n,n}$ such that $R_1 := U_2^T A_1 U_2$ has the desired form. The matrix

$$U := U_1 \begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix}$$

is orthogonal and satisfies

$$U^T A U = \begin{bmatrix} 1 & 0 \\ 0 & U_2^T \end{bmatrix} U_1^T A U_1 \begin{bmatrix} 1 & 0 \\ 0 & U_2 \end{bmatrix} = \begin{bmatrix} \alpha & \star \\ 0 & R_1 \end{bmatrix} =: R,$$

where R has the desired form.

Case 2: $\beta \neq 0$. We first show that x, y are linearly independent. If $x = 0$, then using $\beta \neq 0$ in the first equation in (18.1) implies that also $y = 0$. This is not possible, since the eigenvector $v = x + \mathbf{i}y$ must be nonzero. Thus, $x \neq 0$, and using $\beta \neq 0$ in the second equation in (18.1) implies that also $y \neq 0$. If $x, y \in \mathbb{R}^{n+1} \setminus \{0\}$ are linearly dependent, then there exists a $\mu \in \mathbb{R} \setminus \{0\}$ with $x = \mu y$. The two equations in (18.1) then can be written as

$$Ax = (\alpha - \beta\mu)x \quad \text{and} \quad Ax = \frac{1}{\mu}(\beta + \alpha\mu)x,$$

which implies that $\beta(1 + \mu^2) = 0$. Since $1 + \mu^2 \neq 0$ for all $\mu \in \mathbb{R}$, this implies $\beta = 0$, which contradicts the assumption that $\beta \neq 0$. Consequently, x, y are linearly independent.

We can combine the two equations in (18.1) to the system

$$A[x, y] = [x, y] \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix},$$

where $\text{rank}([x, y]) = 2$. Applying the Gram-Schmidt method with respect to the standard scalar product of $\mathbb{R}^{n+1,1}$ to the matrix $[x, y] \in \mathbb{R}^{n+1,2}$ yields

$$[x, y] = [q_1, q_2] \begin{bmatrix} r_{11} & r_{12} \\ 0 & r_{22} \end{bmatrix} =: Q R_1,$$

with $Q^T Q = I_2$ and $R_1 \in GL_2(\mathbb{R})$. It then follows that

$$A Q = A[x, y] R_1^{-1} = [x, y] \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} R_1^{-1} = Q R_1 \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} R_1^{-1}.$$

The real matrix

$$R_2 := R_1 \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} R_1^{-1}$$

has, considered as element of $\mathbb{C}^{2,2}$, the pair of complex conjugate eigenvalues $\alpha \pm \mathbf{i}\beta$ with $\beta \neq 0$. In particular, the $(2, 1)$ -entry of R_2 is nonzero, since otherwise R_2 would have two real eigenvalues.

We again extend q_1, q_2 by vectors w_3, \dots, w_{n+1} to an orthonormal basis of $\mathbb{R}^{n+1,1}$ with respect to the standard scalar product. (For $n = 1$ the list w_3, \dots, w_{n+1} is empty.) Then $U_1 := [Q, w_3, \dots, w_{n+1}] \in \mathbb{R}^{n+1, n+1}$ is orthogonal and we have

$$U_1^T A U_1 = U_1^T [A Q, A[w_3, \dots, w_{n+1}]] = U_1^T [Q R_2, A[w_3, \dots, w_{n+1}]] = \begin{bmatrix} R_2 & \star \\ 0 & A_1 \end{bmatrix}$$

for a matrix $A_1 \in \mathbb{R}^{n-1, n-1}$. Analogously to the first case, an application of the induction hypothesis to this matrix yields the desired matrices R and U . \square

Theorem 18.5 implies the following result for real normal matrices.

Corollary 18.6 *A matrix $A \in \mathbb{R}^{n, n}$ is normal if and only if there exists an orthogonal matrix $U \in \mathbb{R}^{n, n}$ with*

$$U^T A U = \text{diag}(R_1, \dots, R_m),$$

where, for every $j = 1, \dots, m$ either $R_j \in \mathbb{R}^{1,1}$ or

$$R_j = \begin{bmatrix} \alpha_j & \beta_j \\ -\beta_j & \alpha_j \end{bmatrix} \in \mathbb{R}^{2,2} \quad \text{with} \quad \beta_j \neq 0.$$

In the second case the matrix R_j has, considered as complex matrix, a pair of complex conjugate eigenvalues of the form $\alpha_j \pm \mathbf{i}\beta_j$.

Proof Exercise. \square

Example 18.7 The matrix

$$A = \frac{1}{2} \begin{bmatrix} 0 & \sqrt{2} & -\sqrt{2} \\ -\sqrt{2} & 1 & 1 \\ \sqrt{2} & 1 & 1 \end{bmatrix} \in \mathbb{R}^{3,3}$$

has, considered as a complex matrix, the eigenvalues $1, \mathbf{i}, -\mathbf{i}$. It is therefore neither diagonalizable nor can it be triangulated over \mathbb{R} . For the orthogonal matrix

$$U = \frac{1}{2} \begin{bmatrix} 0 & 2 & 0 \\ -\sqrt{2} & 0 & \sqrt{2} \\ \sqrt{2} & 0 & \sqrt{2} \end{bmatrix} \in \mathbb{R}^{3,3}$$

the transformed matrix

$$U^T A U = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

is in real Schur form.

18.2 Orthogonal and Unitary Endomorphisms

In this section we extend the concept of orthogonal and unitary matrices to endomorphisms.

Definition 18.8 Let \mathcal{V} be a finite dimensional Euclidean or unitary vector space. An endomorphism $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ is called *orthogonal* or *unitary*, respectively, if $f^{ad} \circ f = \text{Id}_{\mathcal{V}}$.

If $f^{ad} \circ f = \text{Id}_{\mathcal{V}}$, then $f^{ad} \circ f$ is bijective and hence f is injective (cp. Exercise 2.7). Corollary 10.11 implies that f is bijective. Hence f^{ad} is the unique inverse of f , and we also have $f \circ f^{ad} = \text{Id}_{\mathcal{V}}$ (cp. our remarks following Definition 2.21).

Note that an orthogonal or unitary endomorphism f is normal, and therefore all results from the previous section also apply to f .

Lemma 18.9 *Let \mathcal{V} be a finite dimensional Euclidean or unitary vector space and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ be orthogonal or unitary, respectively. If B is an orthonormal basis of \mathcal{V} , then $[f]_{B,B}$ is an orthogonal or unitary matrix, respectively.*

Proof Let $\dim(\mathcal{V}) = n$. For every orthonormal basis B of \mathcal{V} we have

$$I_n = [\text{Id}_{\mathcal{V}}]_{B,B} = [f^{ad} \circ f]_{B,B} = [f^{ad}]_{B,B} [f]_{B,B} = ([f]_{B,B})^H [f]_{B,B},$$

and thus $[f]_{B,B}$ is orthogonal or unitary, respectively. (In the Euclidean case $([f]_{B,B})^H = ([f]_{B,B})^T$.) \square

In the following theorem we show that an orthogonal or unitary endomorphism is characterized by the fact that it does not change the scalar product of arbitrary vectors.

Lemma 18.10 *Let \mathcal{V} be a finite dimensional Euclidean or unitary vector space with the scalar product $\langle \cdot, \cdot \rangle$. Then $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ is orthogonal or unitary, respectively, if and only if $\langle f(v), f(w) \rangle = \langle v, w \rangle$ for all $v, w \in \mathcal{V}$.*

Proof If f is orthogonal or unitary and if $v, w \in \mathcal{V}$, then

$$\langle v, w \rangle = \langle \text{Id}_{\mathcal{V}}(v), w \rangle = \langle (f^{ad} \circ f)(v), w \rangle = \langle f(v), f(w) \rangle.$$

On the other hand, suppose that $\langle v, w \rangle = \langle f(v), f(w) \rangle$ for all $v, w \in \mathcal{V}$. Then

$$\begin{aligned} 0 &= \langle v, w \rangle - \langle f(v), f(w) \rangle = \langle v, w \rangle - \langle v, (f^{ad} \circ f)(w) \rangle \\ &= \langle v, (\text{Id}_{\mathcal{V}} - f^{ad} \circ f)(w) \rangle. \end{aligned}$$

Since the scalar product is non-degenerate and v can be chosen arbitrarily, we have $(\text{Id}_{\mathcal{V}} - f^{ad} \circ f)(w) = 0$ for all $w \in \mathcal{V}$, and hence $\text{Id}_{\mathcal{V}} = f^{ad} \circ f$. \square

We have the following corollary (cp. Lemma 12.13).

Corollary 18.11 *If \mathcal{V} is a finite dimensional Euclidean or unitary vector space with the scalar product $\langle \cdot, \cdot \rangle$, $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ is orthogonal or unitary, respectively, and $\| \cdot \| = \langle \cdot, \cdot \rangle^{1/2}$ is the norm induced by the scalar product, then $\| f(v) \| = \| v \|$ for all $v \in \mathcal{V}$.*

For the vector space $\mathcal{V} = \mathbb{C}^{n,1}$ with the standard scalar product and induced norm $\|v\|_2 = (v^H v)^{1/2}$ as well as a unitary matrix $A \in \mathbb{C}^{n,n}$, we have $\|Av\|_2 = \|v\|_2$ for all $v \in \mathbb{C}^{n,1}$. Thus,

$$\|A\|_2 = \sup_{v \in \mathbb{C}^{n,1} \setminus \{0\}} \frac{\|Av\|_2}{\|v\|_2} = 1$$

(cp. (6) in Example 12.4). This holds analogously for orthogonal matrices $A \in \mathbb{R}^{n,n}$.

We now study the eigenvalues and eigenvectors of orthogonal and unitary endomorphisms.

Lemma 18.12 *Let \mathcal{V} be a finite dimensional Euclidean or unitary vector space and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ be orthogonal or unitary, respectively. If λ is an eigenvalue of f , then $|\lambda| = 1$.*

Proof Let $\langle \cdot, \cdot \rangle$ be the scalar product on \mathcal{V} . If $f(v) = \lambda v$ with $v \neq 0$, then

$$\langle v, v \rangle = \langle \text{Id}_{\mathcal{V}}(v), v \rangle = \langle (f^{ad} \circ f)(v), v \rangle = \langle f(v), f(v) \rangle = \langle \lambda v, \lambda v \rangle = |\lambda|^2 \langle v, v \rangle,$$

and $\langle v, v \rangle \neq 0$ implies that $|\lambda| = 1$. □

The statement of Lemma 18.12 holds, in particular, for unitary and orthogonal matrices. However, one should keep in mind that an orthogonal matrix (or an orthogonal endomorphism) may not have an eigenvalue. For example, the orthogonal matrix

$$A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in \mathbb{R}^{2,2}$$

has the characteristic polynomial $P_A = t^2 + 1$, which has no real roots. If considered as an element of $\mathbb{C}^{2,2}$, the matrix A has the eigenvalues \mathbf{i} and $-\mathbf{i}$.

Theorem 18.13

(1) *If $A \in \mathbb{C}^{n,n}$ is unitary, then there exists a unitary matrix $U \in \mathbb{C}^{n,n}$ with*

$$U^H A U = \text{diag}(\lambda_1, \dots, \lambda_n)$$

and $|\lambda_j| = 1$ for $j = 1, \dots, n$.

(2) *If $A \in \mathbb{R}^{n,n}$ is orthogonal, then there exists an orthogonal matrix $U \in \mathbb{R}^{n,n}$ with*

$$U^T A U = \text{diag}(R_1, \dots, R_m),$$

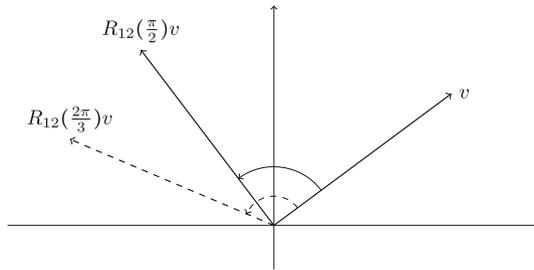
where for every $j = 1, \dots, m$ either $R_j = [\lambda_j] \in \mathbb{R}^{1,1}$ with $\lambda_j = \pm 1$ or

$$R_j = \begin{bmatrix} c_j & s_j \\ -s_j & c_j \end{bmatrix} \in \mathbb{R}^{2,2} \quad \text{with } s_j \neq 0 \quad \text{and } c_j^2 + s_j^2 = 1.$$

which satisfies

$$\begin{aligned} R_{12}(\alpha)^T R_{12}(\alpha) &= \begin{bmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{bmatrix} \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix} \\ &= \begin{bmatrix} \cos^2(\alpha) + \sin^2(\alpha) & 0 \\ 0 & \cos^2(\alpha) + \sin^2(\alpha) \end{bmatrix} \\ &= I_2 = R_{12}(\alpha)R_{12}(\alpha)^T. \end{aligned}$$

One easily sees that each of the matrices $R_{ij}(\alpha) \in \mathbb{R}^{n,n}$ is orthogonal. The multiplication of a vector $v \in \mathbb{R}^{n,1}$ with the matrix $R_{ij}(\alpha)$ results in a (counterclockwise) rotation of v by the angle α in the (i, j) -coordinate plane. In Numerical Mathematics, the matrices $R_{ij}(\alpha)$ are called *Givens rotations*.² This is illustrated in the figure below for the vector $v = [1.0, 0.75]^T \in \mathbb{R}^{2,1}$ and the matrices $R_{12}(\pi/2)$ and $R_{12}(\frac{2\pi}{3})$, which represent rotations by 90 and 120 degrees, respectively.



Example 18.15 For $u \in \mathbb{R}^{n,1} \setminus \{0\}$ we define the *Householder matrix*

$$H(u) := I_n - \frac{2}{u^T u} uu^T \in \mathbb{R}^{n,n}, \quad (18.2)$$

and for $u = 0$ we set $H(0) := I_n$. For every $u \in \mathbb{R}^{n,1}$ then $H(u)$ is an orthogonal matrix (cp. Exercise 12.17). The multiplication of a vector $v \in \mathbb{R}^{n,1}$ with the matrix $H(u)$ describes a reflection of v at the hyperplane

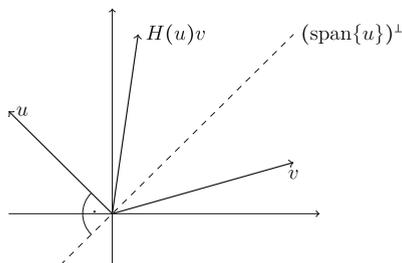
$$(\text{span}\{u\})^\perp = \{y \in \mathbb{R}^{n,1} \mid u^T y = 0\},$$

i.e., the hyperplane of vectors that are orthogonal to u with respect to the standard scalar product. This is illustrated in the figure below for the vector $v = [1.75, 0.5]^T \in \mathbb{R}^{2,1}$ and the Householder matrix

$$H(u) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

which corresponds to $u = [-1, 1]^T \in \mathbb{R}^{2,1}$.

²Wallace Givens (1910–1993), pioneer of Numerical Linear Algebra.



MATLAB-Minute.

Let $u = [5, 3, 1]^T \in \mathbb{R}^{3,1}$. Apply the command `norm(u)` to compute the Euclidean norm of u and form the Householder matrix $H = \text{eye}(3) - (2/(u' * u)) * (u * u')$. Check the orthogonality of H via the computation of `norm(H' * H - eye(3))`. Form the vector $v = H * u$ and compare the Euclidean norms of u and v .

18.3 Selfadjoint Endomorphisms

We have already studied selfadjoint endomorphisms f on a finite dimensional Euclidean or unitary vector space. The defining property for this class of endomorphisms is $f = f^{ad}$ (cp. Definition 13.13).

Obviously, selfadjoint endomorphisms are normal and hence the results of Sect. 18.1 hold. We now strengthen some of these results.

Lemma 18.16 *For a finite dimensional Euclidean or unitary vector space \mathcal{V} and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$, the following statements are equivalent:*

- (1) f is selfadjoint.
- (2) For every orthonormal basis B of \mathcal{V} we have $[f]_{B,B} = ([f]_{B,B})^H$.
- (3) There exists an orthonormal basis B of \mathcal{V} with $[f]_{B,B} = ([f]_{B,B})^H$.

(In the Euclidean case $([f]_{B,B})^H = ([f]_{B,B})^T$.)

Proof In Corollary 13.14 we have already shown that (1) implies (2), and obviously (2) implies (3). If (3) holds, then $[f]_{B,B} = ([f]_{B,B})^H = [f^{ad}]_{B,B}$ (cp. Theorem 13.12), and hence $f = f^{ad}$, so that (1) holds. □

We have the following strong result on the diagonalizability of selfadjoint endomorphisms in both the Euclidean and the unitary case.

Theorem 18.17 *If \mathcal{V} is a finite dimensional Euclidean or unitary vector space and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ is selfadjoint, then there exists an orthonormal basis B of \mathcal{V} such that $[f]_{B,B}$ is a real diagonal matrix.*

Proof Consider first the unitary case. If f is selfadjoint, then f is normal and hence unitarily diagonalizable (cp. Theorem 18.2). Let B be an orthonormal basis of \mathcal{V} so that $[f]_{B,B}$ is a diagonal matrix. Then $[f]_{B,B} = [f^{ad}]_{B,B} = ([f]_{B,B})^H$ implies that the diagonal entries of $[f]_{B,B}$, which are the eigenvalues of \underline{f} , are real.

Let \mathcal{V} be an n -dimensional Euclidean vector space. If $\tilde{B} = \{v_1, \dots, v_n\}$ is an orthonormal basis of \mathcal{V} , then $[f]_{\tilde{B},\tilde{B}}$ is symmetric and in particular normal. By Corollary 18.6, there exists an orthogonal matrix $U = [u_{ij}] \in \mathbb{R}^{n,n}$ with

$$U^T [f]_{\tilde{B},\tilde{B}} U = \text{diag}(R_1, \dots, R_m),$$

where for $j = 1, \dots, m$ either $R_j \in \mathbb{R}^{1,1}$ or

$$R_j = \begin{bmatrix} \alpha_j & \beta_j \\ -\beta_j & \alpha_j \end{bmatrix} \in \mathbb{R}^{2,2} \quad \text{with } \beta_j \neq 0.$$

Since $U^T [f]_{\tilde{B},\tilde{B}} U$ is symmetric, a 2×2 block R_j with $\beta_j \neq 0$ cannot occur. Thus, $U^T [f]_{\tilde{B},\tilde{B}} U$ is a real diagonal matrix.

We define the basis $B = \{w_1, \dots, w_n\}$ of \mathcal{V} by

$$(w_1, \dots, w_n) = (v_1, \dots, v_n)U.$$

Then, by construction, $U = [\text{Id}_{\mathcal{V}}]_{B,\tilde{B}}$ and hence $U^T = U^{-1} = [\text{Id}_{\mathcal{V}}]_{\tilde{B},B}$. Therefore, $U^T [f]_{\tilde{B},\tilde{B}} U = [f]_{B,B}$. If $\langle \cdot, \cdot \rangle$ is the scalar product on \mathcal{V} , then $\langle v_i, v_j \rangle = \delta_{ij}$, $i, j = 1, \dots, n$. With $U^T U = I_n$ we get

$$\langle w_i, w_j \rangle = \left\langle \sum_{k=1}^n u_{ki} v_k, \sum_{\ell=1}^n u_{\ell j} v_\ell \right\rangle = \sum_{k=1}^n \sum_{\ell=1}^n u_{ki} u_{\ell j} \langle v_k, v_\ell \rangle = \sum_{k=1}^n u_{ki} u_{kj} = \delta_{ij}.$$

Hence B is an orthonormal basis of \mathcal{V} . □

This theorem has the following “matrix version”.

Corollary 18.18

- (1) If $A \in \mathbb{R}^{n,n}$ is symmetric, then there exist an orthogonal matrix $U \in \mathbb{R}^{n,n}$ and a diagonal matrix $D \in \mathbb{R}^{n,n}$ with $A = UDU^T$.
- (2) If $A \in \mathbb{C}^{n,n}$ is Hermitian, then there exist a unitary matrix $U \in \mathbb{C}^{n,n}$ and a diagonal matrix $D \in \mathbb{R}^{n,n}$ with $A = UDU^H$.

The statement (1) in this corollary is known as the *principal axes transformation*. We will briefly discuss the background of this name from the theory of bilinear forms and their applications in geometry. A symmetric matrix $A = [a_{ij}] \in \mathbb{R}^{n,n}$ defines a symmetric bilinear form on $\mathbb{R}^{n,1}$ via

$$\beta_A : \mathbb{R}^{n,1} \times \mathbb{R}^{n,1} \rightarrow \mathbb{R}, \quad (x, y) \mapsto y^T A x = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i y_j.$$

The map

$$q_A : \mathbb{R}^{n,1} \rightarrow \mathbb{R}, \quad x \mapsto \beta_A(x, x) = x^T A x,$$

is called the *quadratic form* associated with this symmetric bilinear form.

Since A is symmetric, there exists an orthogonal matrix $U = [u_1, \dots, u_n]$ such that $U^T A U = D$ is a real diagonal matrix. If $B_1 = \{e_1, \dots, e_n\}$, then $[\beta_A]_{B_1 \times B_1} = A$. The set $B_2 = \{u_1, \dots, u_n\}$ forms an orthonormal basis of $\mathbb{R}^{n,1}$ with respect to the standard scalar product, and $[u_1, \dots, u_n] = [e_1, \dots, e_n]U$, hence $U = [\text{Id}_{\mathbb{R}^{n,1}}]_{B_2, B_1}$. For the change of bases from B_1 to B_2 we obtain

$$[\beta_A]_{B_2 \times B_2} = ([\text{Id}_{\mathbb{R}^{n,1}}]_{B_2, B_1})^T [\beta_A]_{B_1 \times B_1} [\text{Id}_{\mathbb{R}^{n,1}}]_{B_2, B_1} = U^T A U = D$$

(cp. Theorem 11.14). Thus, the real diagonal matrix D represents the bilinear form β_A defined by A with respect to the basis B_2 .

The quadratic form q_A associated with β_A is also transformed to a simpler form by this change of bases, since analogously

$$q_A(x) = x^T A x = x^T U D U^T x = y^T D y = \sum_{i=1}^n \lambda_i y_i^2 = q_D(y), \quad y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} := U^T x.$$

Thus, the quadratic form q_A is turned into a “sum of squares”, defined by the quadratic form q_D .

The *principal axes transformation* is given by the change of bases from the canonical basis of $\mathbb{R}^{n,1}$ to the basis given by the pairwise orthonormal eigenvectors of A in $\mathbb{R}^{n,1}$. The n pairwise orthogonal subspaces $\text{span}\{u_j\}$, $j = 1, \dots, n$, form the n *principal axes*. The geometric interpretation of this term is illustrated in the following example.

Example 18.19 For the symmetric matrix

$$A = \begin{bmatrix} 4 & 1 \\ 1 & 2 \end{bmatrix} \in \mathbb{R}^{2,2}$$

we have

$$U^T A U = \begin{bmatrix} 3 + \sqrt{2} & 0 \\ 0 & 3 - \sqrt{2} \end{bmatrix} = D,$$

with the orthogonal matrix $U = [u_1, u_2] \in \mathbb{R}^{2,2}$ and

$$u_1 = \begin{bmatrix} c \\ s \end{bmatrix}, \quad u_2 = \begin{bmatrix} -s \\ c \end{bmatrix}, \quad \text{where}$$

$$c = \frac{1 + \sqrt{2}}{\sqrt{(1 + \sqrt{2})^2 + 1}} = 0.9239, \quad s = \frac{1}{\sqrt{(1 + \sqrt{2})^2 + 1}} = 0.3827.$$

(The numbers here are rounded to the fourth significant digit.) With the associated quadratic form $q_A(x) = 4x_1^2 + 2x_1x_2 + 2x_2^2$, we define the set

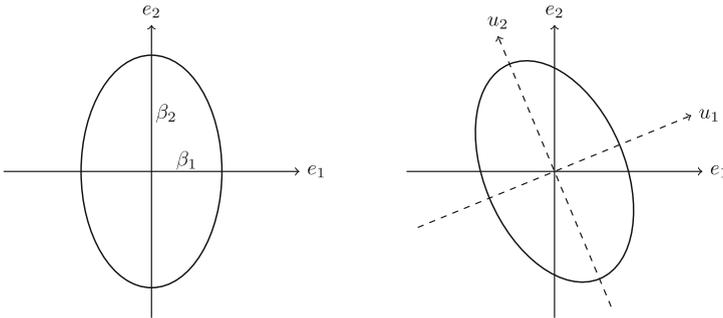
$$E_A = \{x \in \mathbb{R}^{2,1} \mid q_A(x) - 1 = 0\}.$$

As described above, the principal axes transformation consists in the transformation from the canonical coordinate system to a coordinate system given by an orthonormal basis of eigenvectors of A . If we carry out this transformation and replace q_A by the quadratic form q_D , we get the set

$$E_D = \{y \in \mathbb{R}^{2,1} \mid q_D(y) - 1 = 0\} = \left\{ [y_1, y_2]^T \in \mathbb{R}^{2,1} \mid \frac{y_1^2}{\beta_1^2} + \frac{y_2^2}{\beta_2^2} - 1 = 0 \right\},$$

where $\beta_1 = \sqrt{\frac{1}{3 + \sqrt{2}}} = 0.4760, \quad \beta_2 = \sqrt{\frac{1}{3 - \sqrt{2}}} = 0.7941.$

This set forms the ellipse centered at the origin of the two dimensional cartesian coordinate system (spanned by the canonical basis vectors e_1, e_2) with axes of lengths β_1 and β_2 , which is illustrated on the left part of the following figure:



The elements $x \in E_A$ are given by $x = Uy$ for $y \in E_D$. The orthogonal matrix

$$U = \begin{bmatrix} c & -s \\ s & c \end{bmatrix}$$

is a Givens rotation that rotates the ellipse E_D counterclockwise by the angle $\cos^{-1}(c) = 0.3926$ (approximately 22.5 degrees). Hence E_A is just a “rotated version” of E_D . The right part of the figure above shows the ellipse E_A in the cartesian coordinate system. The dashed lines indicate the respective spans of the vectors u_1 and u_2 , which are the eigenvectors of A and the *principal axes* of the ellipse E_A .

Let $A \in \mathbb{R}^{n,n}$ be symmetric. For a given vector $v \in \mathbb{R}^{n,1}$ and a scalar $\alpha \in \mathbb{R}$,

$$Q(x) = x^T A x + v^T x + \alpha, \quad x \in \mathbb{R}^{n,1}$$

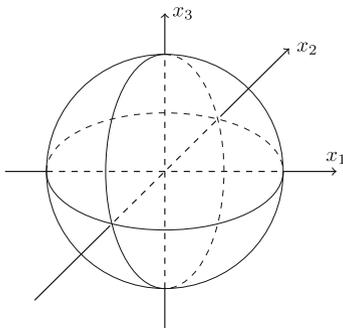
is a quadratic function in n variables (the entries of the vector x). The set of zeros of this function, i.e., the set $\{x \in \mathbb{R}^{n,1} \mid Q(x) = 0\}$, is called a *hypersurface of degree 2* or a *quadric*. In Example 18.19 we have already seen quadrics in the case $n = 2$ and with $v = 0$. We next give some further examples.

Example 18.20

(1) Let $n = 3$, $A = I_3$, $v = [0, 0, 0]^T$ and $\alpha = -1$. The corresponding quadric

$$\{[x_1, x_2, x_3]^T \in \mathbb{R}^{3,1} \mid x_1^2 + x_2^2 + x_3^2 - 1 = 0\}$$

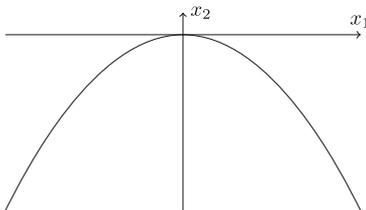
is the surface of the ball with radius 1 around the origin:



(2) Let $n = 2$, $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, $v = [0, 2]^T$ and $\alpha = 0$. The corresponding quadric

$$\{[x_1, x_2]^T \in \mathbb{R}^{2,1} \mid x_1^2 + 2x_2 = 0\}$$

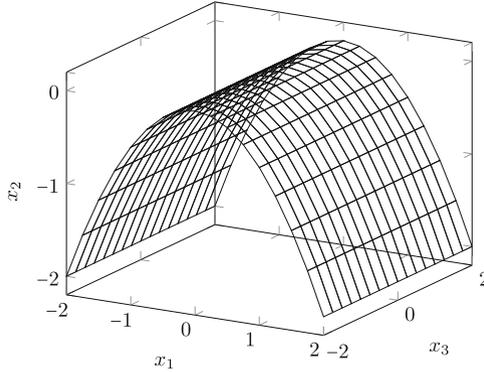
is a parabola:



(3) Let $n = 3$, $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $v = [0, 2, 0]^T$ and $\alpha = 0$. The corresponding quadric

$$\{[x_1, x_2, x_3]^T \in \mathbb{R}^{3,1} \mid x_1^2 + 2x_2 = 0\}$$

is a parabolic cylinder:



Corollary 18.18 motivates the following definition.

Definition 18.21 If $A \in \mathbb{R}^{n,n}$ is symmetric or $A \in \mathbb{C}^{n,n}$ is Hermitian with n_+ positive, n_- negative and n_0 zero eigenvalues (counted with their corresponding multiplicities), then the triple (n_+, n_-, n_0) is called the *inertia* of A .

Let us first consider, for simplicity, only the case of real symmetric matrices.

Lemma 18.22 If $A \in \mathbb{R}^{n,n}$ symmetric has the inertia (n_+, n_-, n_0) , then A and $S_A = \text{diag}(I_{n_+}, -I_{n_-}, 0_{n_0})$ are congruent.

Proof Let $A \in \mathbb{R}^{n,n}$ be symmetric and let $A = U\Lambda U^T$ with an orthogonal matrix $U \in \mathbb{R}^{n,n}$ and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^{n,n}$. If A has the inertia (n_+, n_-, n_0) , then we can assume without loss of generality that

$$\Lambda = \begin{bmatrix} \Lambda_{n_+} & & \\ & \Lambda_{n_-} & \\ & & 0_{n_0} \end{bmatrix} = \text{diag}(\Lambda_{n_+}, \Lambda_{n_-}, 0_{n_0}),$$

where the diagonal matrices Λ_{n_+} and Λ_{n_-} contain the positive and negative eigenvalues of A , respectively, and $0_{n_0} \in \mathbb{R}^{n_0, n_0}$. We have $\Lambda = \Delta S_A \Delta$, where

$$S_A := \text{diag}(I_{n_+}, -I_{n_-}, 0_{n_0}) \in \mathbb{R}^{n,n},$$

$$\Delta := \text{diag}((\Lambda_{n_+})^{1/2}, (-\Lambda_{n_-})^{1/2}, I_{n_0}) \in GL_n(\mathbb{R}).$$

Here $(\text{diag}(\mu_1, \dots, \mu_m))^{1/2} = \text{diag}(\sqrt{\mu_1}, \dots, \sqrt{\mu_m})$ and thus

$$A = U \Lambda U^T = U \Delta S_A \Delta U^T = (U \Delta) S_A (U \Delta)^T. \quad \square$$

This result will be used in the proof of *Sylvester’s law of inertia*.³

Theorem 18.23 *The inertia of a symmetric matrix $A \in \mathbb{R}^{n \times n}$ is invariant under congruence, i.e., for every matrix $G \in GL_n(\mathbb{R})$ the matrices A and $G^T A G$ have the same inertia.*

Proof The assertion is trivial for $A = 0$. Let $A \neq 0$ have the inertia (n_+, n_-, n_0) , then not both n_+ and n_- can be equal to zero. We assume without loss of generality that $n_+ > 0$. (If $n_+ = 0$, then the following argument can be applied for $n_- > 0$.)

By Lemma 18.22 there exist $G_1 \in GL_n(\mathbb{R})$ and $S_A = \text{diag}(I_{n_+}, -I_{n_-}, 0_{n_0})$ with $A = G_1^T S_A G_1$. Let $G_2 \in GL_n(\mathbb{R})$ be arbitrary and set $B := G_2^T A G_2$. Then B is symmetric and has an inertia $(\tilde{n}_+, \tilde{n}_-, \tilde{n}_0)$. Therefore, $B = G_3^T S_B G_3$ for $S_B = \text{diag}(I_{\tilde{n}_+}, -I_{\tilde{n}_-}, 0_{\tilde{n}_0})$ and a matrix $G_3 \in GL_n(\mathbb{R})$. If we show that $n_+ = \tilde{n}_+$ and $n_0 = \tilde{n}_0$, then also $n_- = \tilde{n}_-$.

We have

$$A = (G_2^{-1})^T B G_2^{-1} = (G_2^{-1})^T G_3^T S_B G_3 G_2^{-1} = G_4^T S_B G_4, \quad G_4 := G_3 G_2^{-1},$$

and $G_4 \in GL_n(\mathbb{R})$ implies that $\text{rank}(A) = \text{rank}(S_B) = \text{rank}(B)$, hence $n_0 = \tilde{n}_0$.

We set

$$\begin{aligned} G_1^{-1} &= [u_1, \dots, u_{n_+}, v_1, \dots, v_{n_-}, w_1, \dots, w_{n_0}] \quad \text{and} \\ G_4^{-1} &= [\tilde{u}_1, \dots, \tilde{u}_{\tilde{n}_+}, \tilde{v}_1, \dots, \tilde{v}_{\tilde{n}_-}, \tilde{w}_1, \dots, \tilde{w}_{\tilde{n}_0}]. \end{aligned}$$

Let $\mathcal{V}_1 := \text{span}\{u_1, \dots, u_{n_+}\}$ and $\mathcal{V}_2 := \text{span}\{\tilde{v}_1, \dots, \tilde{v}_{\tilde{n}_-}, \tilde{w}_1, \dots, \tilde{w}_{\tilde{n}_0}\}$. Since $n_+ > 0$, we have $\dim(\mathcal{V}_1) \geq 1$. If $x \in \mathcal{V}_1 \setminus \{0\}$, then

$$x = \sum_{j=1}^{n_+} \alpha_j u_j = G_1^{-1} [\alpha_1, \dots, \alpha_{n_+}, 0, \dots, 0]^T$$

for some $\alpha_1, \dots, \alpha_{n_+} \in \mathbb{R}$ that are not all zero. This implies

$$x^T A x = \sum_{j=1}^{n_+} \alpha_j^2 > 0.$$

³James Joseph Sylvester (1814–1897) proved this result for quadratic forms in 1852. He also coined the name *law of inertia* which according to him is “expressing the fact of the existence of an invariable number inseparably attached to such [bilinear] forms”.

If, on the other hand, $x \in \mathcal{V}_2$, then an analogous argument shows that $x^T A x \leq 0$. Hence $\mathcal{V}_1 \cap \mathcal{V}_2 = \{0\}$, and the dimension formula for subspaces (cp. Theorem 9.29) yields

$$\underbrace{\dim(\mathcal{V}_1)}_{=n_+} + \underbrace{\dim(\mathcal{V}_2)}_{=n-\tilde{n}_+} - \underbrace{\dim(\mathcal{V}_1 \cap \mathcal{V}_2)}_{=0} = \dim(\mathcal{V}_1 + \mathcal{V}_2) \leq \dim(\mathbb{R}^{n,1}) = n,$$

and thus $n_+ \leq \tilde{n}_+$. If we repeat the same construction by interchanging the roles of n_+ and \tilde{n}_+ , then $\tilde{n}_+ \leq n_+$. Thus, $n_+ = \tilde{n}_+$ and the proof is complete. \square

In the following result we transfer Lemma 18.22 and Theorem 18.23 to complex Hermitian matrices.

Theorem 18.24 *Let $A \in \mathbb{C}^{n,n}$ be Hermitian with the inertia (n_+, n_-, n_0) . Then there exists a matrix $G \in GL_n(\mathbb{C})$ with*

$$A = G^H \operatorname{diag}(I_{n_+}, I_{n_-}, 0_{n_0}) G.$$

Moreover, for every matrix $G \in GL_n(\mathbb{C})$ the matrices A and $G^H A G$ have the same inertia.

Proof Exercise. \square

Finally, we discuss a special class of symmetric and Hermitian matrices.

Definition 18.25 A real symmetric or complex Hermitian $n \times n$ matrix A is called

- (1) *positive semidefinite*, if $v^H A v \geq 0$ for all $v \in \mathbb{R}^{n,1}$ resp. $v \in \mathbb{C}^{n,1}$,
- (2) *positive definite*, if $v^H A v > 0$ for all $v \in \mathbb{R}^{n,1} \setminus \{0\}$ resp. $v \in \mathbb{C}^{n,1} \setminus \{0\}$.

If in (1) or (2) the reverse inequality holds, then the corresponding matrices are called *negative semidefinite* or *negative definite*, respectively.

For selfadjoint endomorphisms we define analogously: If \mathcal{V} is a finite dimensional Euclidean or unitary vector space with the scalar product $\langle \cdot, \cdot \rangle$ and if $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ is selfadjoint, then f is called *positive semidefinite* or *positive definite*, if $\langle f(v), v \rangle \geq 0$ for all $v \in \mathcal{V}$ resp. $\langle f(v), v \rangle > 0$ for all $v \in \mathcal{V} \setminus \{0\}$.

The following theorem characterizes symmetric positive definite matrices; see Exercise 18.19 and Exercise 18.20 for the transfer of the results to positive semidefinite matrices resp. positive definite endomorphisms.

Theorem 18.26 *If $A \in \mathbb{R}^{n,n}$ is symmetric, then the following statements are equivalent:*

- (1) *A is positive definite.*
- (2) *All eigenvalues of A are real and positive.*
- (3) *There exists a lower triangular matrix $L \in GL_n(\mathbb{R})$ with $A = LL^T$.*

Proof

- (1) \Rightarrow (2): The symmetric matrix A is diagonalizable with real eigenvalues (cp. (1) in Corollary 18.18). If λ is an eigenvalue with associated eigenvector v , i.e., $Av = \lambda v$, then $\lambda v^T v = v^T Av > 0$ and $v^T v > 0$ implies that $\lambda > 0$.
- (2) \Rightarrow (1): Let $A = U^T \text{diag}(\lambda_1, \dots, \lambda_n) U$ be a diagonalization A with an orthogonal matrix $U \in \mathbb{R}^{n,n}$ (cp. (1) in Corollary 18.18) and $\lambda_j > 0$, $j = 1, \dots, n$. Let $v \in \mathbb{R}^{n,1} \setminus \{0\}$ be arbitrary and let $w := Uv$. Then $w \neq 0$ and $v = U^T w$, so that

$$\begin{aligned} v^T Av &= (U^T w)^T U^T \text{diag}(\lambda_1, \dots, \lambda_n) U (U^T w) = w^T \text{diag}(\lambda_1, \dots, \lambda_n) w \\ &= \sum_{j=1}^n \lambda_j w_j^2 > 0. \end{aligned}$$

- (3) \Rightarrow (1): If $A = LL^T$ with $L \in GL_n(\mathbb{R})$, then for every $v \in \mathbb{C}^{n,1} \setminus \{0\}$ we have

$$v^T Av = v^T LL^T v = \|L^T v\|_2^2 > 0,$$

since L^T is invertible. (Note that here we do not need that L is lower triangular.)

- (1) \Rightarrow (3): Let $A = U^T \text{diag}(\lambda_1, \dots, \lambda_n) U$ be a diagonalization of A with an orthogonal matrix $U \in \mathbb{R}^{n,n}$ (cp. (1) in Corollary 18.18). Since A is positive definite, we know from (2) that $\lambda_j > 0$, $j = 1, \dots, n$. We set

$$\Lambda^{1/2} := \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}),$$

and then have $A = (U\Lambda^{1/2})(\Lambda^{1/2}U^T) =: B^T B$. Let $B = QR$ be a QR -decomposition of the invertible matrix B (cp. Corollary 12.12), where $Q \in \mathbb{R}^{n,n}$ is orthogonal and $R \in \mathbb{R}^{n,n}$ is an invertible upper triangular matrix. Then $A = B^T B = (QR)^T (QR) = LL^T$, where $L := R^T$. \square

One easily sees that an analogous result holds for complex Hermitian matrices $A \in \mathbb{C}^{n,n}$. In this case in assertion (3) the lower triangular matrix is $L \in GL_n(\mathbb{C})$ with $A = LL^H$.

The factorization $A = LL^T$ in (3) is called a *Cholesky factorization*⁴ of A . It is special case of the LU -decomposition in Theorem 5.4. In fact, Theorem 18.26 shows that an LU -decomposition of a (real) symmetric positive definite matrix can be computed without row permutations.

In order to compute the Cholesky factorization of the symmetric positive definite matrix $A = [a_{ij}] \in \mathbb{R}^{n,n}$, we consider the equation

⁴André-Louis Cholesky (1875–1918).

$$A = LL^T = \begin{bmatrix} l_{11} & & \\ \vdots & \ddots & \\ l_{n1} & \cdots & l_{nn} \end{bmatrix} \begin{bmatrix} l_{11} & \cdots & l_{n1} \\ & \ddots & \vdots \\ & & l_{nn} \end{bmatrix}.$$

For the first row of A we obtain

$$a_{11} = l_{11}^2 \implies l_{11} = \sqrt{a_{11}}, \tag{18.3}$$

$$a_{1j} = l_{11}l_{j1} \implies l_{j1} = \frac{a_{1j}}{l_{11}}, \quad j = 2, \dots, n. \tag{18.4}$$

Analogously, for the rows $i = 2, \dots, n$ of A we obtain

$$a_{ii} = \sum_{j=1}^i l_{ij}l_{ij} \implies l_{ii} = \left(a_{ii} - \sum_{j=1}^{i-1} l_{ij}^2 \right)^{1/2}, \tag{18.5}$$

$$\begin{aligned} a_{ij} &= \sum_{k=1}^n l_{ik}l_{jk} = \sum_{k=1}^i l_{ik}l_{jk} = \sum_{k=1}^{i-1} l_{ik}l_{jk} + l_{ii}l_{ji} \\ &\implies l_{ji} = \frac{1}{l_{ii}} \left(a_{ij} - \sum_{k=1}^{i-1} l_{ik}l_{jk} \right), \quad \text{for } j > i. \end{aligned} \tag{18.6}$$

The symmetric or Hermitian positive definite matrices are closely related to the positive definite bilinear forms on Euclidian or unitary vector spaces.

Theorem 18.27 *If \mathcal{V} is a finite dimensional Euclidian or unitary vector space and if β is a symmetric or Hermitian bilinear form on \mathcal{V} , respectively, then the following statements are equivalent:*

- (1) β is positive definite, i.e., $\beta(v, v) > 0$ for all $v \in \mathcal{V} \setminus \{0\}$.
- (2) For every basis B of \mathcal{V} the matrix representation $[\beta]_{B \times B}$ is (symmetric or Hermitian) positive definite.
- (3) There exists a basis B of \mathcal{V} such that the matrix representation $[\beta]_{B \times B}$ is (symmetric or Hermitian) positive definite.

Proof Exercise. □

Exercises

- 18.1 Let $A \in \mathbb{R}^{n,n}$ be normal. Show that αA for every $\alpha \in \mathbb{R}$, A^k for every $k \in \mathbb{N}_0$, and $p(A)$ for every $p \in \mathbb{R}[t]$ are normal.
- 18.2 Let $A, B \in \mathbb{R}^{n,n}$ be normal. Are $A + B$ and AB then normal as well?
- 18.3 Let $A \in \mathbb{R}^{2,2}$ be normal but not symmetric. Show that then

$$A = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}$$

for some $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R} \setminus \{0\}$.

- 18.4 Prove Corollary 18.6 using Theorem 18.5.
- 18.5 Show that real skew-symmetric matrices (i. e., matrices with $A = -A^T \in \mathbb{R}^{n,n}$) and complex skew-Hermitian matrices (i. e., matrices with $A = -A^H \in \mathbb{C}^{n,n}$) are normal.
- 18.6 Let \mathcal{V} be a finite dimensional unitary vector space and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ be normal. Show the following assertions:
- If $f = f^2$, then f is selfadjoint.
 - If $f^2 = f^3$, then $f = f^2$.
 - If f is nilpotent, then $f = 0$.
- 18.7 Let \mathcal{V} be a finite dimensional real or complex vector space and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ be diagonalizable. Show that there exists a scalar product on \mathcal{V} such that f is normal with respect to this scalar products.
- 18.8 Let $A \in \mathbb{C}^{n,n}$. Show the following assertions:
- A is normal if and only if there exists a normal matrix B with n distinct eigenvalues that commutes with A .
 - A is normal if and only if $A + aI$ is normal for every $a \in \mathbb{C}$.
 - Let $H(A) := \frac{1}{2}(A + A^H)$ be the Hermitian and $S(A) := \frac{1}{2}(A - A^H)$ the skew-Hermitian part of A . Show that $A = H(A) + S(A)$, $H(A)^H = H(A)$ and $S(A)^H = -S(A)$. Show, furthermore, that A is normal if and only if $H(A)$ and $S(A)$ commute.
- 18.9 Show that if $A \in \mathbb{C}^{n,n}$ is normal and if $f(z) = \frac{az+b}{cz+d}$ with $ad - bc \neq 0$ is defined on the spectrum of A , then $f(A) = (aA + bI)(cA + dI)^{-1}$.
(The map $f(z)$ is called a *Möbius transformation*.⁵ Such transformations play an important role in Function Theory and in many other areas of Mathematics.)
- 18.10 Let \mathcal{V} be a finite dimensional Euclidian or unitary vector space and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ be orthogonal or unitary, respectively. Show that f^{-1} exists and is again orthogonal or unitary, respectively.
- 18.11 Let $u \in \mathbb{R}^{n,1}$ and let the Householder matrix $H(u)$ be defined as in (18.2). Show the following assertions:
- For $u \neq 0$ the matrices $H(u)$ and $[-e_1, e_2, \dots, e_n]$ are orthogonally similar, i.e., there exists an orthogonal matrix $Q \in \mathbb{R}^{n,n}$ with

$$Q^T H(u) Q = [-e_1, e_2, \dots, e_n].$$

(This implies that $H(u)$ only has the eigenvalues 1 and -1 with the algebraic multiplicities $n - 1$ and 1, respectively.)
 - Every orthogonal matrix $A \in \mathbb{R}^{n,n}$ can be written as product of n Householder matrices, i.e., there exist $u_1, \dots, u_n \in \mathbb{R}^{n,1}$ with $A = H(u_1) \dots H(u_n)$.

⁵August Ferdinand Möbius (1790–1868).

- 18.12 Let $v \in \mathbb{R}^{n,1}$ satisfy $v^T v = 1$. Show that there exists an orthogonal matrix $U \in \mathbb{R}^{n,n}$ with $Uv = e_1$.
- 18.13 Transfer the proofs of Lemma 18.22 and Theorem 18.23 to complex Hermitian matrices and thus show Theorem 18.24.
- 18.14 Determine for the symmetric matrix

$$A = \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix} \in \mathbb{R}^{2,2}$$

an orthogonal matrix $U \in \mathbb{R}^{2,2}$ such that $U^T A U$ is diagonal. Is A positive (semi-)definite?

- 18.15 Let $K \in \{\mathbb{R}, \mathbb{C}\}$ and let $\{v_1, \dots, v_n\}$ be a basis of $K^{n,1}$. Prove or disprove: A matrix $A = A^H \in K^{n,n}$ is positive definite if and only if $v_j^H A v_j > 0$ for all $j = 1, \dots, n$.
- 18.16 Use Definition 18.25 to test whether the symmetric matrices

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \in \mathbb{R}^{2,2}$$

are positive (semi-)definite. Determine in all cases the inertia.

- 18.17 Let

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix} \in \mathbb{R}^{n,n}$$

with $A_{11} = A_{11}^T \in GL_m(\mathbb{R})$, $A_{12} \in \mathbb{R}^{m,n-m}$ and $A_{22} = A_{22}^T \in \mathbb{R}^{n-m,n-m}$. The matrix $S := A_{22} - A_{12}^T A_{11}^{-1} A_{12} \in \mathbb{R}^{m,m}$ is called the *Schur complement*⁶ of A_{11} in A . Show that A is positive definite if A_{11} and S are positive definite. (For the Schur complement, see also Exercise 4.17.)

- 18.18 Show that $A \in \mathbb{C}^{n,n}$ is Hermitian positive definite if and only if $\langle x, y \rangle = y^H A x$ defines a scalar product on $\mathbb{C}^{n,1}$.
- 18.19 Prove the following version of Theorem 18.26 for positive semidefinite matrices.

If $A \in \mathbb{R}^{n,n}$ is symmetric, then the following statements are equivalent:

- (1) A is positive semidefinite.
- (2) All eigenvalues of A are real and nonnegative.
- (3) There exists an upper triangular matrix $L \in \mathbb{R}^{n,n}$ with $A = LL^T$.

- 18.20 Let \mathcal{V} be a finite dimensional Euclidian or unitary vector space and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ be selfadjoint. Show that f is positive definite if and only if all eigenvalues of f are real and positive.
- 18.21 Let $A \in \mathbb{R}^{n,n}$. A matrix $X \in \mathbb{R}^{n,n}$ with $X^2 = A$ is called a square root of A (cp. Sect. 17.1).

⁶Issai Schur (1875–1941).

- (a) Show that a symmetric positive definite matrix $A \in \mathbb{R}^{n,n}$ has a symmetric positive definite square root.
- (b) Show that the matrix

$$A = \begin{bmatrix} 33 & 6 & 6 \\ 6 & 24 & -12 \\ 6 & -12 & 24 \end{bmatrix}$$

is symmetric positive definite and compute a symmetric positive definite square root of A .

- (c) Show that the matrix $A = J_n(0)$, $n \geq 2$, does not have a square root.

18.22 Show that the matrix

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} \in \mathbb{R}^{3,3}$$

is positive definite and compute a Cholesky factorization of A using (18.3)–(18.6).

18.23 Let $A, B \in \mathbb{C}^{n,n}$ be Hermitian and let B be furthermore positive definite. Show that the polynomial $\det(tB - A) \in \mathbb{C}[t]_{\leq n}$ has exactly n real roots.

18.24 Prove Theorem 18.27.