Chapter 14
Eigenvalues of Endomorphisms

In previous chapters we have already studied eigenvalues and eigenvectors of matrices. In this chapter we generalize these concepts to endomorphisms, and we investigate when endomorphisms on finite dimensional vector spaces can be represented by diagonal matrices or (upper) triangular matrices. From such representations we easily can read off important information about the endomorphism, in particular its eigenvalues.

14.1 Basic Definitions and Properties

We first consider an arbitrary vector space and then concentrate on the finite dimensional case.

Definition 14.1 Let $\mathcal{V}$ be a K-vector space and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$. If $\lambda \in K$ and $v \in \mathcal{V} \setminus \{0\}$ satisfy

$$f(v) = \lambda v,$$

then λ is called an *eigenvalue* of f, and v is called an *eigenvector* of f corresponding to λ.

By definition, $v = 0$ cannot be an eigenvector, but an eigenvalue $\lambda = 0$ may occur (cp. the example following Definition 8.7).

The equation $f(v) = \lambda v$ can be written as

$$0 = \lambda v - f(v) = (\lambda \mathrm{Id}_\mathcal{V} - f)(v).$$

Hence, $\lambda \in K$ is an eigenvalue of f if and only if

$$\ker(\lambda \mathrm{Id}_\mathcal{V} - f) \neq \{0\}.$$

We already know that the kernel of an endomorphism on $\mathcal{V}$ forms a subspace of $\mathcal{V}$ (cp. Lemma 10.7). This holds, in particular, for $\ker(\lambda \mathrm{Id}_{\mathcal{V}} - f)$.

Definition 14.2 If $\mathcal{V}$ is a K-vector space and $\lambda \in K$ is an eigenvalue of $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$, then the subspace

$$\mathcal{V}_f(\lambda) := \ker(\lambda \mathrm{Id}_{\mathcal{V}} - f)$$

is called the *eigenspace* of f corresponding to λ and

$$g(\lambda, f) := \dim(\mathcal{V}_f(\lambda))$$

is called the *geometric multiplicity* of the eigenvalue λ.

By definition, the eigenspace $\mathcal{V}_f(\lambda)$ is spanned by all eigenvectors of f corresponding to the eigenvalue λ. If $\mathcal{V}_f(\lambda)$ is finite dimensional, then $g(\lambda, f) = \dim(\mathcal{V}_f(\lambda))$ is equal to the maximal number of linearly independent eigenvectors of f corresponding to λ.

Definition 14.3 Let $\mathcal{V}$ be a K-vector space, let $\mathcal{U} \subseteq \mathcal{V}$ be a subspace, and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$. If $f(\mathcal{U}) \subseteq \mathcal{U}$, i.e., if $f(u) \in \mathcal{U}$ holds for all $u \in \mathcal{U}$, then $\mathcal{U}$ is called an f-*invariant subspace* of $\mathcal{V}$.

An important example of f-invariant subspaces are the eigenspaces of f.

Lemma 14.4 *If $\mathcal{V}$ is a K-vector space and $\lambda \in K$ is an eigenvalue of $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$, then $\mathcal{V}_f(\lambda)$ is an f-invariant subspace of $\mathcal{V}$.*

Proof For every $v \in \mathcal{V}_f(\lambda)$ we have $f(v) = \lambda v \in \mathcal{V}_f(\lambda)$. $\qquad\qquad\square$

We now consider finite dimensional vector spaces and discuss the relationship between the eigenvalues of f and the eigenvalues of a matrix representation of f with respect to a given basis.

Lemma 14.5 *If $\mathcal{V}$ is a finite dimensional K-vector space and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$, then the following statements are equivalent:*

(1) $\lambda \in K$ is an eigenvalue of f.
(2) $\lambda \in K$ is an eigenvalue of the matrix $[f]_{B,B}$ for every basis B of $\mathcal{V}$.

Proof Let $\lambda \in K$ be an eigenvalue of f and let $B = \{v_1, \ldots, v_n\}$ be an arbitrary basis of $\mathcal{V}$. If $v \in \mathcal{V}$ is an eigenvector of f corresponding to the eigenvalue λ, then $f(v) = \lambda v$ and there exist (unique) coordinates $\mu_1, \ldots, \mu_n \in K$, not all equal to zero, with $v = \sum_{j=1}^{n} \mu_j v_j$. Using (10.4) we obtain

$$[f]_{B,B} \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} = \Phi_B(f(v)) = \Phi_B(\lambda v) = \lambda \Phi_B(v) = \lambda \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix},$$

and thus λ is an eigenvalue of $[f]_{B,B}$.

If, on the other hand, $[f]_{B,B}[\mu_1, \ldots, \mu_n]^T = \lambda[\mu_1, \ldots, \mu_n]^T$ with $[\mu_1, \ldots, \mu_n]^T \neq 0$ for a given (arbitrary) basis $B = \{v_1, \ldots, v_n\}$ of $\mathcal{V}$, then we set $v := \sum_{j=1}^{n} \mu_j v_j$. Then $v \neq 0$ and

$$f(v) = \sum_{j=1}^{n} \mu_j f(v_j) = (f(v_1), \ldots, f(v_n)) \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} = ((v_1, \ldots, v_n)[f]_{B,B}) \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

$$= (v_1, \ldots, v_n) \left(\lambda \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \right) = \lambda v,$$

i.e., λ is an eigenvalue of f. $\square$

Lemma 14.5 implies that the eigenvalues of f are the roots of the characteristic polynomial of the matrix $[f]_{B,B}$ (cp. Theorem 8.8). This, however, does *not* hold in general for a matrix representation of the form $[f]_{B,\widetilde{B}}$, where B and $\widetilde{B}$ are two *different* bases of $\mathcal{V}$. In general, the two matrices

$$[f]_{B,\widetilde{B}} = [\mathrm{Id}_{\mathcal{V}}]_{B,\widetilde{B}} [f]_{B,B} \quad \text{and} \quad [f]_{B,B}$$

do not have the same eigenvalues.

Example 14.6 Consider the vector space $\mathbb{R}^{2,1}$ with the bases

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, \quad \widetilde{B} = \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}.$$

Then the endomorphism

$$f : \mathbb{R}^{2,1} \to \mathbb{R}^{2,1}, \quad v \mapsto Fv, \quad \text{where} \quad F = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

has the matrix representations

$$[f]_{B,B} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad [f]_{B,\widetilde{B}} = \frac{1}{2} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}.$$

We have $\det(t I_2 - [f]_{B,B}) = t^2 - 1$, and thus f has the eigenvalues -1 and 1. On the other hand, the characteristic polynomial of $[f]_{B,\widetilde{B}}$ is $t^2 - \frac{1}{2}$, so that this matrix has the eigenvalues $-1/\sqrt{2}$ and $1/\sqrt{2}$.

For two different bases B and $\widetilde{B}$ of $\mathcal{V}$ the matrices $[f]_{B,B}$ and $[f]_{\widetilde{B},\widetilde{B}}$ are similar (cp. the discussion following Corollary 10.20). In Theorem 8.12 we have shown that

similar matrices have the same characteristic polynomial. This justifies the following definition.

Definition 14.7 If $n \in \mathbb{N}$, $\mathcal{V}$ is an n-dimensional K-vector space with the basis B, and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$, then

$$P_f := \det(t I_n - [f]_{B,B}) \in K[t]$$

is called the *characteristic polynomial* of f.

The characteristic polynomial P_f is always a monic polynomial with

$$\deg(P_f) = n = \dim(\mathcal{V}).$$

As we have discussed before, P_f is independent of the choice of the basis of $\mathcal{V}$. A scalar $\lambda \in K$ is an eigenvalue of f if and only if λ is a root of P_f, i.e., $P_f(\lambda) = 0$. As shown in Example 8.9, in real vector spaces with dimensions at least two, there exist endomorphisms that do not have eigenvalues.

If λ is a root of P_f, then $P_f = (t - \lambda) \cdot q$ for a monic polynomial $q \in K[t]$, i.e., the *linear factor $t - \lambda$ divides the polynomial P_f*; we will show this formally in Corollary 15.5 below. If also $q(\lambda) = 0$, then $q = (t - \lambda) \cdot \widetilde{q}$ for a monic polynomial $\widetilde{q} \in K[t]$, and thus $P_f = (t - \lambda)^2 \cdot \widetilde{q}$. We can continue until $P_f = (t - \lambda)^d \cdot g$ for a $g \in K[t]$ with $g(\lambda) \neq 0$. This leads to the following definition.

Definition 14.8 Let $\mathcal{V}$ be a finite dimensional K-vector space, and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ have the eigenvalue $\lambda \in K$. If the characteristic polynomial of f has the form

$$P_f = (t - \lambda)^d \cdot g$$

for some $g \in K[t]$ with $g(\lambda) \neq 0$, then d is called the *algebraic multiplicity* of the eigenvalue λ of f. It is denoted by $a(\lambda, f)$.

If $\lambda_1, \ldots, \lambda_k$ are the *pairwise distinct* eigenvalues of f with corresponding algebraic multiplicities $a(\lambda_1, f), \ldots, a(\lambda_k, f)$, and if $\dim(\mathcal{V}) = n$, then

$$a(\lambda_1, f) + \ldots + a(\lambda_k, f) \leq n,$$

since $\deg(P_f) = \dim(\mathcal{V}) = n$.

Example 14.9 The endomorphism $f : \mathbb{R}^{4,1} \to \mathbb{R}^{4,1}$, $v \mapsto Fv$ with

$$F = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \in \mathbb{R}^{4,4},$$

has the characteristic polynomial $P_f = (t-1)^2(t^2+1)$. The only real root of P_f is 1, and $a(\lambda_1, f) = 2 < 4 = \dim(\mathbb{R}^{4,1})$.

Lemma 14.10 *If $\mathcal{V}$ is a finite dimensional K-vector space and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$, then*

$$g(\lambda, f) \leq a(\lambda, f)$$

for every eigenvalue λ of f.

Proof Let $\lambda \in K$ be an eigenvalue of f with geometric multiplicity $m = g(\lambda, f)$. Then there exist m linear independent eigenvectors $v_1, \ldots, v_m \in \mathcal{V}$ of f corresponding to the eigenvalue λ. If $m = \dim(\mathcal{V})$, then these m eigenvectors form a basis B of $\mathcal{V}$. If $m < \dim(\mathcal{V}) = n$, then we can extend the m eigenvectors to a basis $B = \{v_1, \ldots, v_m, v_{m+1}, \ldots, v_n\}$ of $\mathcal{V}$.

We have $f(v_j) = \lambda v_j$ for $j = 1, \ldots, m$ and, therefore,

$$[f]_{B,B} = \begin{bmatrix} \lambda I_m & Z_1 \\ 0 & Z_2 \end{bmatrix}$$

for two matrices $Z_1 \in K^{m,n-m}$ and $Z_2 \in K^{n-m,n-m}$. Using (1) in Lemma 7.10 we obtain

$$P_f = \det(t I_n - [f]_{B,B}) = (t - \lambda)^m \cdot \det(t I_{n-m} - Z_2),$$

which implies $a(\lambda, f) \geq m = g(\lambda, f)$. $\qquad\square$

In the following we will try to find a basis of $\mathcal{V}$, so that the eigenvalues of a given endomorphism f can be read off easily from its matrix representation. The easiest forms of matrices in this sense are diagonal and triangular matrices, since their eigenvalues are just their diagonal entries.

14.2 Diagonalizability

In this section we will analyze when for a given endomorphism has a diagonal matrix representation. We formally define this property as follows.

Definition 14.11 Let $\mathcal{V}$ be a finite dimensional K-vector space. An endomorphism $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ is called *diagonalizable*, if there exists a basis B of $\mathcal{V}$, such that $[f]_{B,B}$ is a diagonal matrix.

Accordingly, a matrix $A \in K^{n,n}$ is diagonalizable when there exists a matrix $S \in GL_n(K)$ with $A = SDS^{-1}$ for a diagonal matrix $D \in K^{n,n}$.

In order to analyze the diagonalizablility, we begin with a sufficient condition for the linear independence of eigenvectors. This condition also holds when $\mathcal{V}$ is infinite dimensional.

Lemma 14.12 *Let $\mathcal{V}$ be a K-vector space and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$. If $\lambda_1, \ldots, \lambda_k \in K$, $k \geq 2$, are pairwise distinct eigenvalues of f with corresponding eigenvectors $v_1, \ldots, v_k \in \mathcal{V}$, then $v_1, \ldots, v_k$ are linearly independent.*

Proof We prove the assertion by induction on k. Let $k = 2$ and let v_1, v_2 be eigen-vectors of f corresponding to the eigenvalues $\lambda_1 \neq \lambda_2$. Let $\mu_1, \mu_2 \in K$ with $\mu_1 v_1 + \mu_2 v_2 = 0$. Applying f on both sides of this equation as well as multiplying the equation with λ_2 yields the two equations

$$\mu_1 \lambda_1 v_1 + \mu_2 \lambda_2 v_2 = 0,$$
$$\mu_1 \lambda_2 v_1 + \mu_2 \lambda_2 v_2 = 0.$$

Subtracting the second equation from the first, we get $\mu_1(\lambda_1 - \lambda_2)v_1 = 0$. Since $\lambda_1 \neq \lambda_2$ and $v_1 \neq 0$, we have $\mu_1 = 0$. Then from $\mu_1 v_1 + \mu_2 v_2 = 0$ we also obtain $\mu_2 = 0$, since $v_2 \neq 0$. Thus, v_1 and v_2 are linearly independent.

The proof of the inductive step is analogous. We assume that the assertion holds for some $k \geq 2$. Let $\lambda_1, \ldots, \lambda_{k+1}$ be pairwise distinct eigenvalues of f with corre-sponding eigenvectors $v_1, \ldots, v_{k+1}$, and let $\mu_1, \ldots, \mu_{k+1} \in K$ satisfy

$$\mu_1 v_1 + \ldots + \mu_k v_k + \mu_{k+1} v_{k+1} = 0.$$

Applying f to this equation yields

$$\mu_1 \lambda_1 v_1 + \ldots + \mu_k \lambda_k v_k + \mu_{k+1} \lambda_{k+1} v_{k+1} = 0,$$

while a multiplication with λ_{k+1} gives

$$\mu_1 \lambda_{k+1} v_1 + \ldots + \mu_k \lambda_{k+1} v_k + \mu_{k+1} \lambda_{k+1} v_{k+1} = 0.$$

Subtracting this equation from the previous one we get

$$\mu_1(\lambda_1 - \lambda_{k+1})v_1 + \ldots + \mu_k(\lambda_k - \lambda_{k+1})v_k = 0.$$

Since $\lambda_1, \ldots, \lambda_{k+1}$ are pairwise distinct and $v_1, \ldots, v_k$ are linearly independent by the induction hypothesis, we obtain $\mu_1 = \cdots = \mu_k = 0$. But then $\mu_{k+1} v_{k+1} = 0$ implies that also $\mu_{k+1} = 0$, so that $v_1, \ldots, v_{k+1}$ are linearly independent. $\square$

Using this result we next show that the sum of eigenspaces corresponding to pairwise distinct eigenvalues is direct (cp. Theorem 9.31).

Lemma 14.13 *Let $\mathcal{V}$ be a K-vector space and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$. If $\lambda_1, \ldots, \lambda_k \in K$, $k \geq 2$, are pairwise distinct eigenvalues of f, then the corresponding eigenspaces satisfy*

$$\mathcal{V}_f(\lambda_i) \cap \sum_{\substack{j=1 \\ j \neq i}}^{k} \mathcal{V}_f(\lambda_j) = \{0\}$$

for all $i = 1, \ldots, k$.

Proof Let i be fixed and let

$$v \in \mathcal{V}_f(\lambda_i) \cap \sum_{\substack{j=1 \\ j \neq i}}^{k} \mathcal{V}_f(\lambda_j).$$

In particular, we have $v = \sum_{j \neq i} v_j$ for some $v_j \in \mathcal{V}_f(\lambda_j)$, $j \neq i$. Then $-v + \sum_{j \neq i} v_j = 0$, and the linear independence of eigenvectors corresponding to pairwise distinct eigenvalues (cp. Lemma 14.12) implies $v = 0$.   □

The following theorem gives necessary and sufficient conditions for the diagonalizability of an endomorphism on a finite dimensional vector space.

Theorem 14.14 *If $\mathcal{V}$ is a finite dimensional K-vector space and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$, then the following statements are equivalent:*

(1) f is diagonalizable.
(2) There exists a basis of $\mathcal{V}$ consisting of eigenvectors of f.
(3) The characteristic polynomial P_f decomposes into $n = \dim(\mathcal{V})$ linear factors over K, i.e.,

$$P_f = (t - \lambda_1) \cdot \ldots \cdot (t - \lambda_n)$$

with the eigenvalues $\lambda_1, \ldots, \lambda_n \in K$ of f, and for every eigenvalue λ_j we have $g(\lambda_j, f) = a(\lambda_j, f)$.

Proof

(1) ⇔ (2): If $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ is diagonalizable, then there exists a basis $B = \{v_1, \ldots, v_n\}$ of $\mathcal{V}$ and scalars $\lambda_1, \ldots, \lambda_n \in K$ with

$$[f]_{B,B} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}, \tag{14.1}$$

and hence $f(v_j) = \lambda_j v_j$, $j = 1, \ldots, n$. The scalars $\lambda_1, \ldots, \lambda_n$ are thus eigenvalues of f, and the corresponding eigenvectors are $v_1, \ldots, v_n$.
If, on the other hand, there exists a basis $B = \{v_1, \ldots, v_n\}$ of $\mathcal{V}$ consisting of eigenvectors of f, then $f(v_j) = \lambda_j v_j$, $j = 1, \ldots, n$, for scalars $\lambda_1, \ldots, \lambda_n \in K$ (the corresponding eigenvalues), and hence $[f]_{B,B}$ has the form (14.1).

(2) ⇒ (3): Let $B = \{v_1, \ldots, v_n\}$ be a basis of $\mathcal{V}$ consisting of eigenvectors of f, and let $\lambda_1, \ldots, \lambda_n \in K$ be the corresponding eigenvalues. Then $[f]_{B,B}$ has the form (14.1) and hence

$$P_f = (t - \lambda_1) \cdot \ldots \cdot (t - \lambda_n),$$

so that P_f decomposes into linear factors over K.

We still have to show that $g(\lambda_j, f) = a(\lambda_j, f)$ for every eigenvalue λ_j. The eigenvalue λ_j has the algebraic multiplicity $m_j := a(\lambda_j, f)$ if and only if λ_j occurs m_j times on the diagonal of the (diagonal) matrix $[f]_{B,B}$. This holds if and only if exactly m_j vectors of the basis B are eigenvectors of f corresponding to the eigenvalue λ_j. Each of these m_j linearly independent vectors is a element of the eigenspace $\mathcal{V}_f(\lambda_j)$ and, hence,

$$\dim(\mathcal{V}_f(\lambda_j)) = g(\lambda_j, f) \geq m_j = a(\lambda_j, f).$$

From Lemma 14.10 we know that $g(\lambda_j, f) \leq a(\lambda_j, f)$, and thus $g(\lambda_j, f) = a(\lambda_j, f)$.

(3) $\Rightarrow$ (2): Let $\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_k$ be the pairwise distinct eigenvalues of f with corresponding geometric and algebraic multiplicities $g(\widetilde{\lambda}_j, f)$ and $a(\widetilde{\lambda}_j, f)$, $j = 1, \ldots, k$, respectively. Since P_f decomposes into linear factors, we have

$$\sum_{j=1}^{k} a(\widetilde{\lambda}_j, f) = n = \dim(\mathcal{V}).$$

Now $g(\widetilde{\lambda}_j, f) = a(\widetilde{\lambda}_j, f)$, $j = 1, \ldots, k$, implies that

$$\sum_{j=1}^{k} g(\widetilde{\lambda}_j, f) = n = \dim(\mathcal{V}).$$

By Lemma 14.13 we obtain (cp. also Theorem 9.31)

$$\mathcal{V}_f(\widetilde{\lambda}_1) \oplus \ldots \oplus \mathcal{V}_f(\widetilde{\lambda}_k) = \mathcal{V}.$$

If we select bases of the respective eigenspaces $\mathcal{V}_f(\widetilde{\lambda}_j)$, $j = 1, \ldots, k$, then we get a basis of $\mathcal{V}$ that consists of eigenvectors of f.

$\square$

Theorem 14.14 and Lemma 14.12 imply an important sufficient condition for diagonalizability.

Corollary 14.15 *If $\mathcal{V}$ is an n-dimensional K-vector space and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ has n pairwise distinct eigenvalues, then f is diagonalizable.*

The condition of having $n = \dim(\mathcal{V})$ pairwise distinct eigenvalues is, however, not necessary for the diagonalizability of an endomorphism. A simple counterexample is the identity $\mathrm{Id}_\mathcal{V}$, which has the n-fold eigenvalue 1, while $[\mathrm{Id}_\mathcal{V}]_{B,B} = I_n$ holds for every basis B of $\mathcal{V}$. On the other hand, there exist endomorphisms with multiple eigenvalues that are not diagonalizable.

Example 14.16 The endomorphism

$$f : \mathbb{R}^{2,1} \to \mathbb{R}^{2,1}, \quad v \mapsto Fv \quad \text{with} \quad F = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix},$$

has the characteristic polynomial $(t - 1)^2$ and thus only has the eigenvalue 1. We have $\ker(\mathcal{V}_f(1)) = \text{span}\{[1, 0]^T\}$ and thus $g(1, f) = 1 < a(1, f) = 2$. By Theorem 14.14, f is not diagonalizable.

14.3 Triangulation and Schur's Theorem

If the property $g(\lambda_j, f) = a(\lambda_j, f)$ does not hold for every eigenvalue λ_j of f, then f is not diagonalizable. However, as long as the characteristic polynomial P_f decomposes into linear factors, we can find a special basis B such that $[f]_{B,B}$ is a triangular matrix.

Theorem 14.17 *If $\mathcal{V}$ is a finite dimensional K-vector space and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$, then the following statements are equivalent:*

(1) The characteristic polynomial P_f decomposes into linear factors over K.
(2) There exists a basis B of $\mathcal{V}$ such that $[f]_{B,B}$ is upper triangular, i.e., f can be triangulated.

Proof

(2) $\Rightarrow$ (1): If $n = \dim(\mathcal{V})$ and $[f]_{B,B} = [r_{ij}] \in K^{n,n}$ is upper triangular, then $P_f = (t - r_{11}) \cdot \ldots \cdot (t - r_{nn})$.

(1) $\Rightarrow$ (2): We show the assertion by induction on $n = \dim(\mathcal{V})$. The case $n = 1$ is trivial, since then $[f]_{B,B} \in K^{1,1}$.

Suppose that the assertion holds for an $n \geq 1$, and let $\dim(\mathcal{V}) = n + 1$. By assumption,

$$P_f = (t - \lambda_1) \cdot \ldots \cdot (t - \lambda_{n+1}),$$

where $\lambda_1, \ldots, \lambda_{n+1} \in K$ are the eigenvalues of f. Let $v_1 \in \mathcal{V}$ be an eigenvector corresponding to the eigenvalue λ_1. We extend this vector to a basis $B = \{v_1, w_2, \ldots, w_{n+1}\}$ of $\mathcal{V}$. With $B_{\mathcal{W}} := \{w_2, \ldots, w_{n+1}\}$ and $\mathcal{W} := \text{span } B_{\mathcal{W}}$ we have $\mathcal{V} = \text{span}\{v_1\} \oplus \mathcal{W}$ and

$$[f]_{B,B} = \begin{bmatrix} \lambda_1 & a_{12} & \cdots & a_{1,n+1} \\ 0 & a_{22} & \cdots & a_{2,n+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n+1,2} & \cdots & a_{n+1,n+1} \end{bmatrix}.$$

We define $h \in \mathcal{L}(\mathcal{W}, \text{span}\{v_1\})$ and $g \in \mathcal{L}(\mathcal{W}, \mathcal{W})$ by

$$h(w_j) := a_{1j}v_1 \quad \text{and} \quad g(w_j) := \sum_{k=2}^{n+1} a_{kj}w_k, \quad j = 2, \ldots, n+1.$$

Then $f(w) = h(w) + g(w)$ for all $w \in \mathcal{W}$, and

$$[f]_{B,B} = \begin{bmatrix} \lambda_1 & [h]_{B_{\mathcal{W}}, \{v_1\}} \\ 0 & [g]_{B_{\mathcal{W}}, B_{\mathcal{W}}} \end{bmatrix}.$$

Consequently,

$$(t - \lambda_1)P_g = P_f = (t - \lambda_1) \cdot \ldots \cdot (t - \lambda_{n+1}),$$

and hence $P_g = (t - \lambda_2) \cdot \ldots \cdot (t - \lambda_{n+1})$. Now $\dim(\mathcal{W}) = n$ and the characteristic polynomial of $g \in \mathcal{L}(\mathcal{W}, \mathcal{W})$ decomposes into linear factors. By the induction hypothesis there exists a basis $\widehat{B}_{\mathcal{W}} = \{\widehat{w}_2, \ldots, \widehat{w}_{n+1}\}$ of $\mathcal{W}$ such that $[g]_{\widehat{B}_{\mathcal{W}}, \widehat{B}_{\mathcal{W}}}$ upper triangular. Thus, for the basis $B_1 := \{v_1, \widehat{w}_2, \ldots, \widehat{w}_{n+1}\}$ the matrix $[f]_{B_1, B_1}$ is upper triangular.   □

A "matrix version" of this theorem reads as follows: The characteristic polynomial P_A of $A \in K^{n,n}$ decomposes into linear factors over K if and only if A can be triangulated, i.e., there exists a matrix $S \in GL_n(K)$ with $A = SRS^{-1}$ for an upper triangular matrix $R \in K^{n,n}$.

Corollary 14.18 *Let $\mathcal{V}$ be a finite dimensional Euclidian or unitary vector space and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$. If P_f decomposes over $\mathbb{R}$ (in the Euclidian case case) or $\mathbb{C}$ (in the unitary case) into linear factors, then there exists an orthonormal basis B of $\mathcal{V}$, such that $[f]_{B,B}$ is upper triangular.*

Proof If P_f decomposes into linear factors, then by Theorem 14.17 there exists a basis B_1 of $\mathcal{V}$, such that $[f]_{B_1, B_1}$ is upper triangular. Applying the Gram-Schmidt method to the basis B_1, we obtain an orthonormal basis B_2 of $\mathcal{V}$, such that $[\text{Id}_{\mathcal{V}}]_{B_1, B_2}$ is upper triangular (cp. Theorem 12.11). Then

$$[f]_{B_2, B_2} = [\text{Id}_{\mathcal{V}}]_{B_1, B_2}[f]_{B_1, B_1}[\text{Id}_{\mathcal{V}}]_{B_2, B_1} = [\text{Id}_{\mathcal{V}}]_{B_2, B_1}^{-1}[f]_{B_1, B_1}[\text{Id}_{\mathcal{V}}]_{B_2, B_1}.$$

The invertible upper triangular matrices form a group with respect to the matrix multiplication (cp. Theorem 4.13). Thus, all matrices in the product on the right hand side are upper triangular, and hence $[f]_{B_2, B_2}$ is upper triangular.   □

Example 14.19 Consider the Euclidian vector space $\mathbb{R}[t]_{\leq 1}$ with the scalar product $\langle p, q \rangle = \int_0^1 p(t)q(t)\,dt$, and the endomorphism

$$f : \mathbb{R}[t]_{\leq 1} \to \mathbb{R}[t]_{\leq 1}, \quad \alpha_1 t + \alpha_0 \mapsto 2\alpha_1 t + \alpha_0.$$

We have $f(1) = 1$ and $f(t) = 2t$, i.e., the polynomials 1 and t are eigenvectors of f corresponding to the (distinct) eigenvalues 1 and 2. Thus, $\widehat{B} = \{1, t\}$ is a basis of $\mathbb{R}[t]_{\leq 1}$, and $[f]_{\widehat{B}, \widehat{B}}$ is a diagonal matrix. Note that $\widehat{B}$ is not an orthonormal basis, since in particular $\langle 1, t \rangle \neq 0$.

Since P_f decomposes into linear factors, Corollary 14.18 guarantees the existence of an orthonormal basis B for which $[f]_{B, B}$ is upper triangular. In the proof of the implication $(1) \Rightarrow (2)$ of Theorem 14.17 one chooses any eigenvector of f, and then proceeds inductively in order to obtain the triangulation of f. In this example, let us use $q_1 = 1$ as the first vector. This vector is an eigenvector of f with norm 1 corresponding to the eigenvalue 1. If $q_2 \in \mathbb{R}[t]_{\leq 1}$ is a vector with norm 1 and $\langle q_1, q_2 \rangle = 0$, then $B = \{q_1, q_2\}$ is an orthonormal basis for which $[f]_{B, B}$ is an upper triangular matrix. We construct the vector q_2 by orthogonalizing t against q_1 using the Gram-Schmidt method:

$$\widehat{q_2} = t - \langle t, q_1 \rangle q_1 = t - \frac{1}{2},$$
$$\|\widehat{q_2}\| = \left\langle t - \frac{1}{2}, \ t - \frac{1}{2} \right\rangle^{1/2} = \frac{1}{\sqrt{12}},$$
$$q_2 = \|\widehat{q_2}\|^{-1} \widehat{q_2} = \sqrt{12}t - \sqrt{3}.$$

This leads to the triangulation

$$[f]_{B, B} = \begin{bmatrix} 1 & \sqrt{3} \\ 0 & 2 \end{bmatrix} \in \mathbb{R}^{2,2}.$$

We could also choose $q_1 = \sqrt{3}t$, which is an eigenvector of f with norm 1 corresponding to the eigenvalue 2. Orthogonalizing the vector 1 against q_1 leads to the second basis vector $q_2 = -3t + 2$. With the corresponding basis B_1 we obtain the triangulation

$$[f]_{B_1, B_1} = \begin{bmatrix} 2 & -\sqrt{3} \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{2,2}.$$

This example shows that in the triangulation of f the elements above the diagonal can be different for different orthonormal bases. Only the diagonal elements are (except for their order) uniquely determined, since they are the eigenvalues of f. A more detailed statement about the uniqueness is given in Lemma 14.22.

In the next chapter we will prove the Fundamental Theorem of Algebra, which states that every non-constant polynomial over $\mathbb{C}$ decomposes into linear factors. This result has the following corollary, which is known as *Schur's theorem*.[1]

---

[1] Issai Schur (1875–1941).

Corollary 14.20 *If $\mathcal{V}$ is a finite dimensional unitary vector space, then every endo-morphism on $\mathcal{V}$ can be* unitarily triangulated, *i.e., for each $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ there exists an orthonormal basis B of $\mathcal{V}$, such that $[f]_{B,B}$ is upper triangular. The matrix $[f]_{B,B}$ is called a* Schur form *of f.*

If $\mathcal{V}$ is the unitary vector space $\mathbb{C}^{n,1}$ with the standard scalar product, then we obtain the following "matrix version" of Corollary 14.20.

Corollary 14.21 *If $A \in \mathbb{C}^{n,n}$, then there exists a unitary matrix $Q \in \mathbb{C}^{n,n}$ with $A = QRQ^H$ for an upper triangular matrix $R \in \mathbb{C}^{n,n}$. The matrix R is called a* Schur form *of A.*

The following result shows that a Schur form of a matrix $A \in \mathbb{C}^{n,n}$ with n pairwise distinct eigenvalues is "almost unique".

Lemma 14.22 *Let $A \in \mathbb{C}^{n,n}$ have n pairwise distinct eigenvalues, and let $R_1, R_2 \in \mathbb{C}^{n,n}$ be two Schur forms of A. If the diagonals of R_1 and R_2 are equal then $R_1 = UR_2U^H$ for a unitary diagonal matrix U.*

Proof Exercise.   □

A survey of the results on unitary similarity of matrices can be found in the article [Sha91].

MATLAB-Minute.
Consider for $n \geq 2$ the matrix

$$
A = \begin{bmatrix} 1 & 2 & 3 & \cdots & n \\ 1 & 3 & 4 & \cdots & n+1 \\ 1 & 4 & 5 & \cdots & n+2 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & n+1 & n+2 & \dots & 2n-1 \end{bmatrix} \in \mathbb{C}^{n,n}.
$$

Compute a Schur form of A using the command [U,R] = schur(A) for $n = 2, 3, 4, \dots 10$. What are the eigenvalues of A? Formulate a conjecture about the rank of A for general n. Can you prove your conjecture?

Exercises

(In the following exercises K is an arbitrary field.)

14.1. Let $\mathcal{V}$ be a vector space and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ have the eigenvalue λ. Show that $\operatorname{im}(\lambda \operatorname{Id}_{\mathcal{V}} - f)$ is an f-invariant subspace.

14.2. Let $\mathcal{V}$ be a finite dimensional vector space and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ be bijective. Show that f and f^{-1} have the same invariant subspaces.

14.3. Let V be an n-dimensional K-vector space, let $f \in \mathcal{L}(V, V)$, and let $\mathcal{U}$ be an m-dimensional f-invariant subspace of V. Show that a basis B of V exists such that

$$[f]_{B,B} = \begin{bmatrix} A_1 & A_2 \\ 0 & A_3 \end{bmatrix}$$

for some matrices $A_1 \in K^{m,m}$, $A_2 \in K^{m,n-m}$ and $A_3 \in K^{n-m,n-m}$.

14.4. Let $K \in \{\mathbb{R}, \mathbb{C}\}$ and $f : K^{4,1} \to K^{4,1}$, $v \mapsto Fv$ with

$$F = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}.$$

Compute P_f and determine for $K = \mathbb{R}$ and $K = \mathbb{C}$ the eigenvalues of f with their algebraic and geometric multiplicities, as well as the associated eigenspaces.

14.5. Consider the vector space $\mathbb{R}[t]_{\leq n}$ with the standard basis $\{1, t, \ldots, t^n\}$ and the endomorphism

$$f : \mathbb{R}[t]_{\leq n} \to \mathbb{R}[t]_{\leq n}, \quad \sum_{i=0}^{n} \alpha_i t^i \mapsto \sum_{i=2}^{n} i(i-1)\alpha_i t^{i-2} = \frac{d^2}{dt^2} p.$$

Compute P_f, the eigenvalues of f with their algebraic and geometric multiplicities, and examine whether f is diagonalizable or not. What changes if one considers as map the kth derivative (for $k = 3, 4, \ldots, n$)?

14.6. Examine whether the following matrices

$$A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \in \mathbb{Q}^{2,2}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 2 & 0 \\ -1 & 1 & 1 \end{bmatrix} \in \mathbb{Q}^{3,3}, \quad C = \begin{bmatrix} 3 & 1 & 0 & -2 \\ 0 & 2 & 0 & 0 \\ 2 & 2 & 2 & -4 \\ 0 & 0 & 0 & 2 \end{bmatrix} \in \mathbb{Q}^{4,4}$$

are diagonalizable.

14.7. Is the set of all diagonalizable and invertible matrices a subgroup of $GL_n(K)$?

14.8. Let $n \in \mathbb{N}_0$. Consider the $\mathbb{R}$-vector space $\mathbb{R}[t]_{\leq n}$ and the map

$$f : \mathbb{R}[t]_{\leq n} \to \mathbb{R}[t]_{\leq n}, \quad p(t) \mapsto p(t+1) - p(t).$$

Show that f is linear. For which n is f diagonalizable?

14.9. Let V be an $\mathbb{R}$-vector space with the basis $\{v_1, \ldots, v_n\}$. Examine whether the following endomorphisms are diagonalizable or not:

(a) $f(v_j) = v_j + v_{j+1}$, $j = 1, \ldots, n-1$, and $f(v_n) = v_n$,

(b) $f(v_j) = jv_j + v_{j+1}$, $j = 1, \ldots, n-1$, and $f(v_n) = nv_n$.

14.10. Let $\mathcal{V}$ be a finite dimensional Euclidian vector space and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ with $f + f^{ad} = 0 \in \mathcal{L}(\mathcal{V}, \mathcal{V})$. Show that $f \neq 0$ if and only if f is not diagonalizable.

14.11. Let $\mathcal{V}$ be a $\mathbb{C}$-vector space and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ with $f^2 = -\mathrm{Id}_{\mathcal{V}}$. Determine all possible eigenvalues of f.

14.12. Let $\mathcal{V}$ be a finite dimensional vector space and $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$. Show that $P_f(f) = 0 \in \mathcal{L}(\mathcal{V}, \mathcal{V})$.

14.13. Let $\mathcal{V}$ be a finite dimensional K-vector space, let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ and

$$p = (t - \mu_1) \cdot \ldots \cdot (t - \mu_m) \in K[t]_{\leq m}.$$

Show that $p(f)$ is bijective if and only if $\mu_1, \ldots, \mu_m$ are not eigenvalues of f.

14.14. Determine conditions for the entries of the matrices

$$A = \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} \in \mathbb{R}^{2,2},$$

such that A is diagonalizable or can be triangulated.

14.15. Determine an endomorphism on $\mathbb{R}[t]_{\leq 3}$ that is not diagonalizable and that cannot be triangulated.

14.16. Let $\mathcal{V}$ be a vector space with $\dim(\mathcal{V}) = n$. Show that $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ can be triangulated if and only if there exist subspaces $\mathcal{V}_0, \mathcal{V}_1, \ldots, \mathcal{V}_n$ of $\mathcal{V}$ with

(a) $\mathcal{V}_j \subset \mathcal{V}_{j+1}$ for $j = 0, 1, \ldots, n-1$,
(b) $\dim(\mathcal{V}_j) = j$ for $j = 0, 1, \ldots, n$, and
(c) $\mathcal{V}_j$ is f-invariant for $j = 0, 1, \ldots, n$.

14.17. Prove Lemma 14.22.