Linear Algebra II

Peter Philip*

Lecture Notes

Created for the Class of Spring Semester 2019 at LMU Munich

July 29, 2019

Contents

1 Affine Subspaces and Geometry 3
  1.1 Affine Subspaces ................................................. 3
  1.2 Affine Hull and Affine Independence .............................. 5
  1.3 Affine Bases .............................................................. 8
  1.4 Barycentric Coordinates and Convex Sets .......................... 10
  1.5 Affine Maps ............................................................... 12
  1.6 Affine Geometry ......................................................... 17

2 Duality 19
  2.1 Linear Forms and Dual Spaces ...................................... 19
  2.2 Annihilators .............................................................. 23
  2.3 Hyperplanes and Linear Systems .................................... 27
  2.4 Dual Maps ................................................................. 30

3 Symmetric Groups 33

4 Multilinear Maps and Determinants 40
  4.1 Multilinear Maps ....................................................... 41
  4.2 Alternating Multilinear Maps and Determinants .................... 43
  4.3 Determinants of Matrices and Linear Maps ........................ 51

5 Direct Sums and Projections 63

*E-Mail: philip@math.lmu.de
## CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6  Eigenvalues</td>
<td>67</td>
</tr>
<tr>
<td>7  Commutative Rings, Polynomials</td>
<td>80</td>
</tr>
<tr>
<td>8  Characteristic Polynomial, Minimal Polynomial</td>
<td>96</td>
</tr>
<tr>
<td>9  Jordan Normal Form</td>
<td>105</td>
</tr>
<tr>
<td>10 Vector Spaces with Scalar Products</td>
<td>121</td>
</tr>
<tr>
<td>10.1 Definition, Orthogonality</td>
<td>121</td>
</tr>
<tr>
<td>10.2 The Adjoint Map</td>
<td>128</td>
</tr>
<tr>
<td>10.3 Hermitian, Unitary, and Normal Maps</td>
<td>133</td>
</tr>
<tr>
<td>A  Multilinear Maps</td>
<td>140</td>
</tr>
<tr>
<td>References</td>
<td>140</td>
</tr>
</tbody>
</table>
1 Affine Subspaces and Geometry

1.1 Affine Subspaces

**Definition 1.1.** Let $V$ be a vector space over the field $F$. Then $M \subseteq V$ is called an affine subspace of $V$ if, and only if, there exists a vector $v \in V$ and a (vector) subspace $U \subseteq V$ such that $M = v + U$. We define $\dim M := \dim U$ to be the dimension of $M$ (this notion of dimension is well-defined by the following Lem. 1.2(a)).

Thus, the affine subspaces of a vector space $V$ are precisely the translations of vector subspaces $U$ of $V$, i.e. the cosets of subspaces $U$, i.e. the elements of quotient spaces $V/U$.

**Lemma 1.2.** Let $V$ be a vector space over the field $F$.

(a) If $M$ is an affine subspace of $V$, then the vector subspace corresponding to $M$ is unique, i.e. if $M = v_1 + U_1 = v_2 + U_2$ with $v_1, v_2 \in V$ and vector subspaces $U_1, U_2 \subseteq V$, then

$$U_1 = U_2 = \{u - v : u, v \in M\}. \quad (1.1)$$

(b) If $M = v + U$ is an affine subspace of $V$, then the vector $v$ in this representation is unique if, and only if, $U = \{0\}$.

**Proof.** (a): Let $M = v_1 + U_1$ with $v_1 \in V$ and vector subspace $U_1 \subseteq V$. Moreover, let $U := \{u - v : u, v \in M\}$. It suffices to show $U_1 = U$. Suppose, $u_1 \in U_1$. Since $v_1 \in M$ and $v_1 + u_1 \in M$, we have $u_1 = v_1 + u_1 - v_1 \in U$, showing $U_1 \subseteq U$. If $a \in U$, then there are $u_1, u_2 \in U_1$ such that $a = v_1 + u_1 - (v_1 + u_2) = u_1 - u_2 \in U_1$, showing $U \subseteq U_1$, as desired.

(b): If $U = \{0\}$, then $M = \{v\}$ and $v$ is unique. If $M = v + U$ with $0 \neq u \in U$, then $M = v + U = v + u + U$ with $v \neq v + u$. ■

**Definition 1.3.** In the situation of Def. 1.1, we call affine subspaces of dimension 0 points, of dimension 1 lines, and of dimension 2 planes – in $\mathbb{R}^2$ and $\mathbb{R}^3$, such objects are easily visualized and they then coincide with the points, lines, and planes with which one is already familiar.

Affine spaces and vector spaces share many structural properties. In consequence, one can develop a theory of affine spaces that is in many respects analogous to the theory of vector spaces, as will be illustrated by some of the notions and results presented in the following. We start by defining so-called affine combinations, which are, for affine spaces, what linear combinations are for vector spaces:
Definition 1.4. Let $V$ be a vector space over the field $F$ with $v_1, \ldots, v_n \in V$ and $\lambda_1, \ldots, \lambda_n \in F, n \in \mathbb{N}$. Then $\sum_{i=1}^n \lambda_i v_i$ is called an affine combination of $v_1, \ldots, v_n$ if, and only if, $\sum_{i=1}^n \lambda_i = 1$.

Theorem 1.5. Let $V$ be a vector space over the field $F, \emptyset \neq M \subseteq V$. Then $M$ is an affine subspace of $V$ if, and only, if $M$ is closed under affine combinations. More precisely, the following statements are equivalent:

(i) $M$ is an affine subspace of $V$.

(ii) If $n \in \mathbb{N}, v_1, \ldots, v_n \in M$, and $\lambda_1, \ldots, \lambda_n \in F$ with $\sum_{i=1}^n \lambda_i = 1$, then $\sum_{i=1}^n \lambda_i v_i \in M$.

If $\text{char } F \neq 2$, then (i) and (ii) are also equivalent to

(iii) If $v_1, v_2 \in M$ and $\lambda_1, \lambda_2 \in F$ with $\lambda_1 + \lambda_2 = 1$, then $\lambda_1 v_1 + \lambda_2 v_2 \in M$.

Proof. Exercise. ■

The following Th. 1.6 is the analogon of [Phi19, Th. 5.7] for affine spaces:

Theorem 1.6. Let $V$ be a vector space over the field $F$.

(a) Let $I \neq \emptyset$ be an index set and $(M_i)_{i \in I}$ a family of affine subspaces of $V$. Then the intersection $M := \bigcap_{i \in I} M_i$ is either empty or it is an affine subspace of $V$.

(b) In contrast to intersections, unions of affine subspaces are almost never affine subspaces. More precisely, if $M_1$ and $M_2$ are affine subspaces of $V$ and $\text{char } F \neq 2$ (i.e. $1 \neq -1$ in $F$), then

$$M_1 \cup M_2 \text{ is an affine subspace of } V \iff \left(M_1 \subseteq M_2 \lor M_2 \subseteq M_1\right) \quad (1.2)$$

(where "$\iff$" also holds for $\text{char } F = 2$, but cf. Ex. 1.7 below).

Proof. (a): Let $M \neq \emptyset$. We use the characterization of Th. 1.5(ii) to show $M$ is an affine subspace: If $n \in \mathbb{N}, v_1, \ldots, v_n \in M$, and $\lambda_1, \ldots, \lambda_n \in F$ with $\sum_{k=1}^n \lambda_k = 1$, then $v := \sum_{k=1}^n \lambda_k v_k \in M_i$ for each $i \in I$, implying $v \in M$. Thus, $M$ is an affine subspace of $V$.

(b): If $M_1 \subseteq M_2$, then $M_1 \cup M_2 = M_2$, which is an affine subspace of $V$. If $M_2 \subseteq M_1$, then $M_1 \cup M_2 = M_1$, which is an affine subspace of $V$. For the converse, we
now assume \( \text{char } F \neq 2 \), \( M_1 \not\subseteq M_2 \), and \( M_1 \cup M_2 \) is an affine subspace of \( V \). We have to show \( M_2 \subseteq M_1 \). Let \( m_1 \in M_1 \setminus M_2 \) and \( m_2 \in M_2 \). Since \( M_1 \cup M_2 \) is an affine subspace, \( m_2 + m_2 - m_1 \in M_1 \cup M_2 \) by Th. 1.5(ii). If \( m_2 + m_2 - m_1 \in M_2 \), then \( m_1 = m_2 + m_2 - (m_2 + m_2 - m_1) \in M_2 \), in contradiction to \( m_1 \notin M_2 \). Thus, \( m_2 + m_2 - m_1 \in M_1 \). Since \( \text{char } F \neq 0 \), we have \( 2 := 1 + 1 \neq 0 \) in \( F \), implying \( m_2 = \frac{1}{2}(m_2 + m_2 - m_1) + \frac{1}{2}m_1 \in M_1 \), i.e. \( M_2 \subseteq M_1 \).

Example 1.7. Consider \( F := \mathbb{Z}_2 = \{0, 1\} \) and the vector space \( V := F^2 \) over \( F \). Then \( M_1 := U_1 := \{(0, 0), (1, 0)\} = \{(1, 0)\} \) is a vector subspace and, in particular, an affine subspace of \( V \). The set \( M_2 := (0, 1) + U_1 = \{(0, 1), (1, 1)\} \) is also an affine subspace. Then \( M_1 \cup M_2 = V \) is an affine subspace, even though neither \( M_1 \subseteq M_2 \) nor \( M_2 \subseteq M_1 \).

1.2 Affine Hull and Affine Independence

Next, we will define the affine hull of a subset \( A \) of a vector space, which is the affine analogon to the linear notion of the span of \( A \) (which is sometimes also called the linear hull of \( A \)):

**Definition 1.8.** Let \( V \) be a vector space over the field \( F \), \( \emptyset \neq A \subseteq V \), and

\[
\mathcal{M} := \{ M \in \mathcal{P}(V) : A \subseteq M \land M \text{ is affine subspace of } V \},
\]

where we recall that \( \mathcal{P}(V) \) denotes the power set of \( V \). Then the set

\[
\text{aff } A := \bigcap_{M \in \mathcal{M}} M
\]

is called the affine hull of \( A \). We call \( A \) a generating set of \( \text{aff } A \).

The following Prop. 1.9 is the analogon of [Phi19, Prop. 5.9] for affine spaces:

**Proposition 1.9.** Let \( V \) be a vector space over the field \( F \) and \( \emptyset \neq A \subseteq V \).

(a) \( \text{aff } A \) is an affine subspace of \( V \), namely the smallest affine subspace of \( V \) containing \( A \).

(b) \( \text{aff } A \) is the set of all affine combinations of elements from \( A \), i.e.

\[
\text{aff } A = \left\{ \sum_{i=1}^{n} \lambda_i a_i : n \in \mathbb{N} \land \lambda_1, \ldots, \lambda_n \in F \land a_1, \ldots, a_n \in A \land \sum_{i=1}^{n} \lambda_i = 1 \right\}.
\]

(c) If \( A \subseteq B \subseteq V \), then \( \text{aff } A \subseteq \text{aff } B \).

(d) \( A = \text{aff } A \) if, and only if, \( A \) is an affine subspace of \( V \).
(e) \( \text{aff \ aff } A = \text{aff } A \).

Proof. (a): Since \( A \subseteq \text{aff } A \) implies \( \text{aff } A \neq \emptyset \), (a) is immediate from Th. 1.6(a).

(b): Let \( W \) denote the right-hand side of (1.3). If \( M \) is an affine subspace of \( V \) and \( A \subseteq M \), then \( W \subseteq M \), since \( M \) is closed under affine combinations, showing \( W \subseteq \text{aff } A \).

On the other hand, suppose \( n, n_1, \ldots, n_N \in \mathbb{N}, a_1^k, \ldots, a_{n_k}^k \in A \) for each \( k \in \{1, \ldots, N\} \), \( \lambda_1^k, \ldots, \lambda_{n_k}^k \in F \) for each \( k \in \{1, \ldots, N\} \), and \( \alpha_1, \ldots, \alpha_N \in F \) such that

\[ \forall k \in \{1, \ldots, N\} \quad \sum_{i=1}^{n_k} \lambda_i^k = \sum_{i=1}^N \alpha_i = 1. \]

Then

\[ \sum_{k=1}^N \alpha_k \sum_{i=1}^{n_k} \lambda_i^k a_i^k \in W, \]

since

\[ \sum_{k=1}^N \sum_{i=1}^{n_k} \alpha_k \lambda_i^k = \sum_{k=1}^N \alpha_k = 1, \]

showing \( W \) to be an affine subspace of \( V \) by Th. 1.5(ii). Thus, \( \text{aff } A \subseteq W \), completing the proof of \( \text{aff } A = W \).

(c) is immediate from (b).

(d): If \( A = \text{aff } A \), then \( A \) is an affine subspace by (a). For the converse, while it is clear that \( A \subseteq \text{aff } A \) always holds, if \( A \) is an affine subspace, then \( A \in \mathcal{M} \), where \( \mathcal{M} \) is as in Def. 1.8, implying \( \text{aff } A \subseteq A \).

(e) now follows by combining (d) with (a). \( \blacksquare \)

**Proposition 1.10.** Let \( V \) be a vector space over the field \( F \), \( A \subseteq V \), \( M = v + U \) with \( v \in A \) and \( U \) a vector subspace of \( V \). Then the following statements are equivalent:

(i) \( \text{aff } A = M \).

(ii) \( \langle -v + A \rangle = U \).

Proof. Exercise. \( \blacksquare \)

We will now define the notions of *affine dependence/independence*, which are, for affine spaces, what linear dependence/independence are for vector spaces:

**Definition 1.11.** Let \( V \) be a vector space over the field \( F \).

(a) A vector \( v \in V \) is called *affinely dependent* on a subset \( U \) of \( V \) (or on the vectors in \( U \)) if, and only if, there exists \( n \in \mathbb{N} \) and \( u_1, \ldots, u_n \in U \) such that \( v \) is an affine combination of \( u_1, \ldots, u_n \). Otherwise, \( v \) is called *affinely independent* of \( U \).
(b) A subset $U$ of $V$ is called **affinely independent** if, and only if, whenever $0 \in V$ is written as a linear combination of distinct elements of $U$ such that the coefficients have sum 0, then all coefficients must be $0 \in F$, i.e. if, and only if,

$$
\left( n \in \mathbb{N} \land W \subseteq U \land \#W = n \land \sum_{u \in W} \lambda_u u = 0 \land \forall_{u \in W} \lambda_u \in F \land \sum_{u \in W} \lambda_u = 0 \right) \Rightarrow \forall_{u \in W} \lambda_u = 0. \quad (1.4)
$$

Sets that are not affinely independent are called **affinely dependent**.

As a caveat, it is underlined that, in Def. 1.11(b) above, one does not consider affine combinations of the vectors $u \in U$, but special linear combinations (this is related to the fact that 0 is only an affine combination of vectors in $U$, if $\text{aff } U$ is a vector subspace of $V$).

**Remark 1.12.** It is immediate from Def. 1.11 that if $v \in V$ is linearly independent of $U \subseteq V$, then it is also affinely independent of $U$, and, if $U \subseteq V$ is linearly independent, then $U$ is also affinely independent. However, the converse is, in general, not true (cf. Ex. 1.13(b),(c) below).

**Example 1.13.** Let $V$ be a vector space over the field $F$.

(a) $\emptyset$ is affinely independent: Indeed, if $U = \emptyset$, then the left side of the implication in (1.4) is always false (since $W \subseteq U$ means $\#W = 0$), i.e. the implication is true.

(b) Every singleton set $\{v\}$, $v \in V$, is affinely independent, since $\lambda_1 = \sum_{i=1}^1 \lambda_i = 0$ means $\lambda_1 = 0$ (if $v = 0$, then $\{v\}$ is not linearly independent, cf. [Phi19, Ex. 5.13(b)]).

(c) Every set $\{v, w\}$ with two distinct vectors $v, w \in V$ is affinely independent (but not linearly independent for $w = \alpha v$ with some $\alpha \in F$): $0 = \lambda v - \lambda w = \lambda(v - w)$ implies $\lambda = 0$ or $v = w$.

There is a close relationship between affine independence and linear independence:

**Proposition 1.14.** Let $V$ be a vector space over the field $F$ and $U \subseteq V$. Then the following statements are equivalent:

(i) $U$ is affinely independent.

(ii) If $u_0 \in U$, then $U_0 := \{u - u_0 : u \in U \setminus \{u_0\}\}$ is linearly independent.
(iii) The set \( X := \{(u, 1) \in V \times F : u \in U \} \) is a linearly independent subset of the vector space \( V \times F \).

Proof. Exercise. ■

The following Prop. 1.15 is the analogon of [Phi19, Prop. 5.1 4(a)-(c)] for affine spaces:

**Proposition 1.15.** Let \( V \) be a vector space over the field \( F \) and \( U \subseteq V \).

(a) \( U \) is affinely dependent if, and only if, there exists \( u_0 \in U \) such that \( u_0 \) is affinely dependent on \( U \setminus \{u_0\} \).

(b) If \( U \) is affinely dependent and \( U \subseteq M \subseteq V \), then \( M \) is affinely dependent as well.

(c) If \( U \) is affinely independent und \( M \subseteq U \), then \( M \) is affinely independent as well.

Proof. (a): Suppose, \( U \) is affinely dependent. Then there exists \( W \subseteq U \), \( \#W = n \in \mathbb{N} \), such that \( \sum_{u \in W} \lambda_u u = 0 \) with \( \lambda_u \in F \), \( \sum_{u \in W} \lambda_u = 0 \), and there exists \( u_0 \in W \) with \( \lambda_{u_0} \neq 0 \). Then

\[
\begin{align*}
u_0 &= -\lambda_{u_0}^{-1} \sum_{u \in W \setminus \{u_0\}} \lambda_u u = \sum_{u \in W \setminus \{u_0\}} (-\lambda_{u_0}^{-1} \lambda_u) u, \\
&= \sum_{u \in W \setminus \{u_0\}} (-\lambda_{u_0}^{-1} \lambda_u) = (-\lambda_{u_0}^{-1}) \cdot (-\lambda_{u_0}) = 1,
\end{align*}
\]

showing \( u_0 \) to be affinely dependent on \( U \setminus \{u_0\} \). Conversely, if \( u_0 \in U \) is affinely dependent on \( U \setminus \{u_0\} \), then there exists \( n \in \mathbb{N} \), distinct \( u_1, \ldots, u_n \in U \setminus \{u_0\} \), and \( \lambda_1, \ldots, \lambda_n \in F \) with \( \sum_{i=1}^{n} \lambda_i = 1 \) such that

\[
u_0 = \sum_{i=1}^{n} \lambda_i u_i \Rightarrow -u_0 + \sum_{i=1}^{n} \lambda_i u_i = 0,
\]

showing \( U \) to be affinely dependent, since the coefficient of \( u_0 \) is \(-1 \neq 0\) and \(-1 + \sum_{i=1}^{n} \lambda_i = 0\).

(b) and (c) are now both immediate from (a). ■

### 1.3 Affine Bases

**Definition 1.16.** Let \( V \) be a vector space over the field \( F \), let \( M \subseteq V \) be an affine subspace, and \( B \subseteq V \). Then \( B \) is called an **affine basis** of \( M \) if, and only if, \( B \) is a generating set for \( M \) (i.e. \( M = \text{aff} B \)) that is also affinely independent.

There is a close relationship between affine bases and vector space bases:

**Proposition 1.17.** Let \( V \) be a vector space over the field \( F \), let \( M \subseteq V \) be an affine subspace, and let \( B \subseteq M \) with \( v \in B \). Then the following statements are equivalent:
(i) $B$ is an affine basis of $M$.

(ii) $B_0 := \{b - v : b \in B \setminus \{v\}\}$ is a vector space basis of the vector space $U := \{v_1 - v_2 : v_1, v_2 \in M\}$.

Proof. As a consequence of Lem. 1.2(a), we know $U$ to be a vector subspace of $V$ and $M = a + U$ for each $a \in M$. Moreover, $v \in B \subseteq M$ implies $B_0 \subseteq U$. According to Prop. 1.14, $B$ is affinely independent if, and only if, $B_0$ is linearly independent. According to Prop. 1.10, $\text{aff} B = M$ holds if, and only if, $\langle -v + B \rangle = U$, which, since $B_0 = (-v + B) \setminus \{0\}$, holds if, and only if, $\langle B_0 \rangle = U$. ■

The following Th. 1.18 is the analogon of [Phi19, Th. 5.17] for affine spaces:

**Theorem 1.18.** Let $V$ be a vector space over the field $F$, let $M \subseteq V$ be an affine subspace, and let $\emptyset \neq B \subseteq V$. Then the following statements (i) – (iii) are equivalent:

(i) $B$ is an affine basis of $M$.

(ii) $B$ is a maximal affinely independent subset of $M$, i.e. $B$ is affinely independent and each set $A \subseteq M$ with $B \subsetneq A$ is affinely dependent.

(iii) $B$ is a minimal generating set for $M$, i.e. $\text{aff} B = M$ and $\text{aff} A \subsetneq M$ for each $A \subsetneq B$.

Proof. Let $v \in B$, and let $B_0$ and $U$ be as in Prop. 1.17 above. Then, due to Prop. 1.14, $B$ is a maximal affinely independent subset of $M$ if, and only if, $B_0$ is a maximal linearly independent subset of $U$. Moreover, due to Prop. 1.10, $B$ is a minimal (affine) generating set for $M$ if, and only if, $B_0$ is a minimal (linear) generating set for $U$. Thus, the equivalences of Th. 1.18 follow by combining Prop. 1.17 with [Phi19, Th. 5.17]. ■

The following Th. 1.19 is the analogon of [Phi19, Th. 5.23] for affine spaces:

**Theorem 1.19.** Let $V$ be a vector space over the field $F$ and let $M \subseteq V$ be an affine subspace.

(a) If $S \subseteq M$ is affinely independent, then there exists an affine basis of $M$ that contains $S$.

(b) $M$ has an affine basis $B \subseteq M$.

(c) Affine bases of $M$ have a unique cardinality, i.e. if $B \subseteq M$ and $\hat{B} \subseteq M$ are both affine bases of $M$, then there exists a bijective map $\phi : B \rightarrow \hat{B}$.

(d) If $B$ is an affine basis of $M$ and $S \subseteq M$ is affinely independent, then there exists $C \subseteq B$ such that $\hat{B} := S \cup C$ is an affine basis of $M$. 

Proof. Let \( v \in V \) and let \( U \) be a vector subspace of \( V \) such that \( M = v + U \). Then \( v \in M \) and \( U = \{v_1 - v_2 : v_1, v_2 \in M\} \) according to Lem. 1.2(a).

(a): It suffices to consider the case \( S \neq \emptyset \). Thus, let \( v \in S \). According to Prop. 1.14(ii), \( S_0 := \{x - v : x \in S \setminus \{v\}\} \) is a linearly independent subset of \( U \). According to [Phi19, Th. 5.23(a)], \( U \) has a vector space basis \( S_0 \subseteq B_0 \subseteq U \). Then, by Prop. 1.17, \( S \subseteq (v + B_0) \cup \{v\} \) is an affine basis of \( M \).

(b) is immediate from (a).

(c): Let \( B \subseteq M \) and \( \tilde{B} \subseteq M \) be affine bases of \( M \). Moreover, let \( b \in B \) and \( \tilde{b} \in \tilde{B} \). Then, by Prop. 1.17, \( B_0 := \{x - b : x \in B \setminus \{b\}\} \) and \( \tilde{B}_0 := \{x - \tilde{b} : x \in \tilde{B} \setminus \{\tilde{b}\}\} \) are both vector space bases of \( U \). Thus, by [Phi19, Th. 5.23(c)], there exists a bijective map \( \psi : B_0 \rightarrow \tilde{B}_0 \). Then, clearly, the map

\[
\phi : B \rightarrow \tilde{B}, \quad \phi(x) := \begin{cases} \tilde{b} & \text{for } x = b, \\ \tilde{b} + \psi(x - b) & \text{for } x \neq b, \end{cases}
\]

is well-defined and bijective, thereby proving (c).

(d): If \( B \subseteq \text{aff } S \), then, according to Prop. 1.9(c),(e), \( M = \text{aff } B \subseteq \text{aff } S = \text{aff } S \), i.e. \( \text{aff } S = M \), as \( M \) is an affine subspace containing \( S \). Thus, \( S \) is itself an affine basis of \( M \) and the statement holds with \( C := \emptyset \). It remains to consider the case, where there exists \( b \in B \setminus S \) such that \( S \cup \{b\} \) is affinely independent. Then, by Prop. 1.17, \( B_0 := \{x - b : x \in B \setminus \{b\}\} \) is a vector space basis of \( U \) and, by Prop. 1.14(ii), \( S_0 := \{x - b : x \in S\} \) is a linearly independent subset of \( U \). Thus, by [Phi19, Th. 5.23(d)], there exists \( C_0 \subseteq B_0 \) such that \( \tilde{B}_0 := S_0 \cup C_0 \) is a vector space basis of \( U \) and, then, using Prop. 1.17 once again, \( (b + \tilde{B}_0) \cup \{b\} = S \cup C \) with \( C := (b + C_0) \cup \{b\} \subseteq B \) is an affine basis of \( M \). \( \qed \)

### 1.4 Barycentric Coordinates and Convex Sets

The following Th. 1.20 is the analogon of [Phi19, Th. 5.19] for affine spaces:

**Theorem 1.20.** Let \( V \) be a vector space over the field \( F \) and assume \( M \subseteq V \) is an affine subspace with affine basis \( B \) of \( M \). Then each vector \( v \in M \) has unique barycentric coordinates with respect to the affine basis \( B \), i.e., for each \( v \in M \), there exists a unique finite subset \( B_v \) of \( B \) and a unique map \( c : B_v \rightarrow F \setminus \{0\} \) such that

\[
v = \sum_{b \in B_v} c(b) b \quad \land \quad \sum_{b \in B_v} c(b) = 1. \quad (1.5)
\]

**Proof.** The existence of \( B_v \) and the map \( c \) follows from the fact that the affine basis \( B \) is an affine generating set, \( \text{aff } B = M \). For the uniqueness proof, consider finite sets \( B_v, \tilde{B}_v \subseteq B \) and maps \( c : B_v \rightarrow F \setminus \{0\}, \tilde{c} : \tilde{B}_v \rightarrow F \setminus \{0\} \) such that

\[
v = \sum_{b \in B_v} c(b) b = \sum_{b \in B_v} \tilde{c}(b) b \quad \land \quad \sum_{b \in B_v} c(b) = \sum_{b \in B_v} \tilde{c}(b) = 1.
\]
Extend both $c$ and $\tilde{c}$ to $A := B_v \cup \tilde{B}_v$ by letting $c(b) := 0$ for $b \in \tilde{B}_v \setminus B_v$ and $\tilde{c}(b) := 0$ for $b \in B_v \setminus \tilde{B}_v$. Then
\[
0 = \sum_{b \in A} \left( c(b) - \tilde{c}(b) \right) b,
\]
such that the affine independence of $A$ implies $c(b) = \tilde{c}(b)$ for each $b \in A$, which, in turn, implies $B_v = \tilde{B}_v$ and $c = \tilde{c}$.

Example 1.21. With respect to the affine basis $\{0,1\}$ of $\mathbb{R}$ over $\mathbb{R}$, the barycentric coordinates of $\frac{1}{3}$ are $\frac{2}{3}$ and $\frac{1}{3}$, whereas the barycentric coordinates of 5 are $-4$ and 5.

Remark 1.22. Let $V$ be a vector space over the field $F$ and assume $M \subseteq V$ is an affine subspace with affine basis $B$ of $M$.

(a) Caveat: In the literature, one also finds the notion of affine coordinates, however, this notion of affine coordinates is usually (but not always, so one has to use care) defined differently from the notion of barycentric coordinates as defined in Th. 1.20 above: For the affine coordinates, one designates one point $x_0 \in B$ to be the origin of $M$. Let $v \in M$ and let $c : B_v \rightarrow F \setminus \{0\}$ be the map yielding the barycentric coordinates according to Th. 1.20. We write $\{x_0\} \cup B_v = \{x_0, x_1, \ldots, x_n\}$ with distinct elements $x_1, \ldots, x_n \in M$ (if any) and we set $c(x_0) := 0$ in case $x_0 \notin B_v$. Then
\[
v = \sum_{i=0}^{n} c(x_i) x_i \quad \land \quad \sum_{i=0}^{n} c(x_i) = 1,
\]
which, since $1 - \sum_{i=1}^{n} c(x_i) = c(x_0)$, is equivalent to
\[
v = x_0 + \sum_{i=1}^{n} c(x_i) (x_i - x_0) \quad \land \quad \sum_{i=0}^{n} c(x_i) = 1.
\]
One calls the $c(x_1), \ldots, c(x_n)$, given by the map $c_n := c|_{B_v \setminus \{x_0\}}$, the affine coordinates of $v$ with respect to the affine coordinate system $\{x_0\} \cup (-x_0 + B)$ (for $v = x_0$, $c_n$ turns out to be the empty map).

(b) If $x_1, \ldots, x_n \in M$ are distinct points that are affinely independent and $n := n \cdot 1 \neq 0$ in $F$, then one sometimes calls
\[
v := \frac{1}{n} \sum_{i=1}^{n} x_i \in M
\]
the barycenter of $x_1, \ldots, x_n$.

Definition and Remark 1.23. Let $V$ be a vector space over $\mathbb{R}$ (we restrict ourselves to vector spaces over $\mathbb{R}$, since, for a scalar $\lambda$ we will need to know what it means for $\lambda$ to be positive, i.e. $\lambda > 0$ needs to be well-defined). Let $v_1, \ldots, v_n \in V$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$, $n \in \mathbb{N}$. Then we call the affine combination $\sum_{i=1}^{n} \lambda_i v_i$ of $v_1, \ldots, v_n$ a convex combination of $v_1, \ldots, v_n$ if, and only if, in addition to $\sum_{i=1}^{n} \lambda_i = 1$, one has $\lambda_i \geq 0$ for each $i \in \{1, \ldots, n\}$. Moreover, we call $C \subseteq V$ convex if, and only if, $C$...
is closed under convex combinations, i.e. if, and only if, \( n \in \mathbb{N}, v_1, \ldots, v_n \in C \), and \( \lambda_1, \ldots, \lambda_n \in \mathbb{R}^+ \) with \( \sum_{i=1}^{n} \lambda_i = 1 \), implies \( \sum_{i=1}^{n} \lambda_i v_i \in C \) (analogous to Th. 1.5, \( C \subseteq V \) is then convex if, and only if, each convex combination of merely two elements of \( C \) is again in \( C \)). Note that, in contrast to affine subspaces, we allow convex sets to be empty. Clearly, the convex subsets of \( \mathbb{R} \) are precisely the intervals (open, closed, half-open, bounded or unbounded). Convex subsets of \( \mathbb{R}^2 \) include triangles and disks.

Analogous to the proof of Th. 1.6(a), one can show that arbitrary intersections of convex sets are always convex, and, analogous to the definition of the affine hull in Def. 1.8, one defines the \textit{convex hull} \( \text{conv} A \) of a set \( A \subseteq V \) by letting

\[
C := \{ C \in \mathcal{P}(V) : A \subseteq C \land C \text{ is convex subset of } V \},
\]

\[
\text{conv} A := \bigcap_{C \in C} C.
\]

Then Prop. 1.9 and its proof still work completely analogously in the convex situation and one obtains \( \text{conv} A \) to be the smallest convex subset of \( V \) containing \( A \), where \( \text{conv} A \) consists precisely of all convex combinations of elements from \( A \); \( A = \text{conv} A \) holds if, and only if, \( A \) is convex; \( \text{conv conv} A = \text{conv} A \); and \( \text{conv} A \subseteq \text{conv} B \) for each \( A \subseteq B \subseteq V \). If \( n \in \mathbb{N}_0 \) and \( A = \{x_0, x_1, \ldots, x_n\} \subseteq V \) is an affinely independent set, consisting of the \( n+1 \) distinct points \( x_0, x_1, \ldots, x_n \), then \( \text{conv} A \) is called an \( n \)-dimensional \textit{simplex} (or simply an \( n \)-\textit{simplex}) with \textit{vertices} \( x_0, x_1, \ldots, x_n \); \( 0 \)-simplices are called \textit{points}, \( 1 \)-simplices \textit{line segments}, \( 2 \)-simplices \textit{triangles}, and \( 3 \)-simplices \textit{tetrahedra}. If \( \{e_1, \ldots, e_d\} \) denotes the standard basis of \( \mathbb{R}^d \), \( d \in \mathbb{N} \), then \( \text{conv} \{e_1, \ldots, e_{n+1}\}, 0 \leq n < d \), is called the \textit{standard} \( n \)-simplex in \( \mathbb{R}^d \).

### 1.5 Affine Maps

We first study a special type of affine map, namely so-called \textit{translations}.

**Definition 1.24.** Let \( V \) be a vector space over the field \( F \). If \( v \in V \), then the map

\[
T_v : V \rightarrow V, \quad T_v(x) := x + v,
\]

is called a \textit{translation}, namely, the translation by \( v \) or the translation with \textit{translation vector} \( v \). Let \( T(V) := \{T_v : v \in V\} \) denote the set of translations on \( V \).

**Proposition 1.25.** Let \( V \) be a vector space over the field \( F \).

(a) If \( v \in V \) and \( A, B \subseteq V \), then \( T_v(A + B) = v + A + B \). In particular, translations map affine subspaces of \( V \) into affine subspaces of \( V \).

(b) If \( v \in V \), then \( T_v \) is bijective with \( (T_v)^{-1} = T_{-v} \). In particular, \( T(V) \subseteq S_V \), where \( S_V \) denotes the symmetric group on \( V \) according to [Phi19, Ex. 4.9(b)].

(c) Nontrivial translations are not linear: More precisely, \( T_v \) with \( v \in V \) is linear if, and only if, \( v = 0 \) (i.e. \( T_v = \text{Id} \)).
1 AFFINE SUBSPACES AND GEOMETRY

(d) If \( v, w \in V \), then \( T_v \circ T_w = T_{v+w} = T_w \circ T_v \).

(e) \((T(V), \circ)\) is a commutative subgroup of \((S_V, \circ)\). Moreover, \((T(V), \circ) \cong (V, +)\), where

\[ I : (V, +) \rightarrow (T(V), \circ), \quad I(v) := T_v, \]

constitutes a group isomorphism.

Proof. Exercise. ■

We will now define affine maps, which are, for affine spaces, what linear maps are for vector spaces:

Definition 1.26. Let \( V \) and \( W \) be vector spaces over the field \( F \). A map \( A : V \rightarrow W \) is called affine if, and only if, there exists a linear map \( L \in \mathcal{L}(V, W) \) and \( w \in W \) such that

\[
\forall x \in V \quad A(x) = (T_w \circ L)(x) = w + L(x)
\]

(i.e. the affine maps are precisely the compositions of linear maps with translations). We denote the set of all affine maps from \( V \) into \( W \) by \( \mathcal{A}(V, W) \).

Proposition 1.27. Let \( V, W, X \) be vector spaces over the field \( F \).

(a) If \( L \in \mathcal{L}(V, W) \) and \( v \in V \), then \( L \circ T_v = T_{Lv} \circ L \in \mathcal{A}(V, W) \).

(b) If \( A \in \mathcal{A}(V, W) \), \( L \in \mathcal{L}(V, W) \), and \( w \in W \), then \( A = T_w \circ L \) if, and only if, \( T_{-w} \circ A = L \). In particular, \( A = T_w \circ L \) is injective (resp. surjective, resp. bijective) if, and only if, \( L \) is injective (resp. surjective, resp. bijective).

(c) If \( A : V \rightarrow W \) is an affine and bijective, then \( A^{-1} \) is also affine.

(d) If \( A : V \rightarrow W \) and \( B : W \rightarrow X \) are affine, then so is \( B \circ A \).

(e) Define \( GA(V) := \{ A \in \mathcal{A}(V, V) : \text{A bijective}\} \). Then \((GA(V), \circ)\) forms a subgroup of the symmetric group \((S_V, \circ)\) (then, clearly, \(GL(V)\) forms a subgroup of \(GA(V)\), cf. [Phi19, Cor. 6.23]).

Proof. (a): If \( L \in \mathcal{L}(V, W) \) and \( v, x \in V \), then

\[
(L \circ T_v)(x) = L(v + x) = Lv + Lx = (T_{Lv} \circ L)(x),
\]

proving \( L \circ T_v = T_{Lv} \circ L \).

(b) is due to the bijectivity of \( T_w \): One has, since \( T_{-w} \circ T_w = \text{Id} \),

\[
A = T_w \circ L \quad \Leftrightarrow \quad T_{-w} \circ A = T_{-w} \circ T_w \circ L = \text{Id} \circ L = L.
\]
Moreover, for each \( x, y \in V \) and \( z \in W \), one has
\[
Ax = w + Lx = w + Ly = Ay \iff Lx = Ly,
\]
\[
z + w = Ax = w + Lx \iff z = Lx,
\]
\[
z - w = Lx \iff z = w + Lx = Ax,
\]
proving \( A = T_w \circ L \) is injective (resp. surjective, resp. bijective) if, and only if, \( L \) is injective (resp. surjective, resp. bijective).

(c): If \( A = T_w \circ L \) with \( L \in \mathcal{L}(V,W) \) and \( w \in W \) is affine and bijective, then, by (b), \( L \) is bijective. Thus, \( A^{-1} = L^{-1} \circ (T_w)^{-1} = L^{-1} \circ T_{-w} \), which is affine by (a).

(d): If \( A = T_w \circ L, B = T_x \circ K \) with \( L \in \mathcal{L}(V,W), w \in W, K \in \mathcal{L}(W,X), x \in X \), then
\[
\forall \ a \in V \quad (B \circ A)(a) = B(w + La) = x + Kw + (K \circ L)(a) = (T_{Kw+x} \circ (K \circ L))(a),
\]
showing \( B \circ A \) to be affine.

(e) is an immediate consequence of (c) and (d). \( \blacksquare \)

**Proposition 1.28.** Let \( V \) and \( W \) be vector spaces over the field \( F \).

(a) Let \( v \in V, \ w \in W, \ L \in \mathcal{L}(V,W) \), and let \( U \) be a vector subspace of \( V \). Then
\[
(T_w \circ L)(v + U) = w + Lv + L(U)
\]
(in particular, each affine image of an affine subspace is an affine subspace). Moreover, if \( A := T_w \circ L \) and \( S \subseteq V \) such that \( M := v + U = \text{aff } S \), then \( A(M) = w + Lv + L(U) = \text{aff}(A(S)) \).

(b) Let \( y \in W, \ L \in \mathcal{L}(V,W) \), and let \( U \) be a vector subspace of \( W \). Then \( L^{-1}(U) \) is a vector subspace of \( V \) and
\[
\forall \ v \in L^{-1}(y) \quad L^{-1}(y + U) = v + L^{-1}(U)
\]
(in particular, each linear preimage of an affine subspace is either empty or an affine subspace).

(c) If \( M \subseteq W \) is an affine subspace of \( W \) and \( A \in \mathcal{A}(V,W) \), then \( A^{-1}(M) \) is either empty or an affine subspace of \( V \).

**Proof.** Exercise. \( \blacksquare \)

The following Prop. 1.29 is the analogon of [Phi19, Prop. 6.5(a),(b)] for affine spaces (but cf. Caveat 1.30 below):

**Proposition 1.29.** Let \( V \) and \( W \) be vector spaces over the field \( F \), and let \( A : V \rightarrow W \) be affine.
(a) If $A$ is injective, then, for each affinely independent subset $S$ of $V$, $A(S)$ is an affinely independent subset of $W$.

(b) $A$ is surjective if, and only if, for each subset $S$ of $V$ with $V = \text{aff } S$, one has $W = \text{aff } (A(S))$.

Proof. Let $w \in W$ and $L \in \mathcal{L}(V, W)$ be such that $A = T_w \circ L$.

(a): If $A$ is injective, $S \subseteq V$ is affinely independent, and $\lambda_1, \ldots, \lambda_n \in F$; $s_1, \ldots, s_n \in S$ distinct; $n \in \mathbb{N}$; such that $\sum_{i=1}^{n} \lambda_i = 0$ and

$$0 = \sum_{i=1}^{n} \lambda_i A(s_i) = \sum_{i=1}^{n} \lambda_i (w + Ls_i) = \left( \sum_{i=1}^{n} \lambda_i \right) w + L \left( \sum_{i=1}^{n} \lambda_i s_i \right),$$

then $\sum_{i=1}^{n} \lambda_i s_i = 0$ by [Phi19, Prop. 6.3(d)], implying $\lambda_1 = \cdots = \lambda_n = 0$ and, thus, showing that $A(S)$ is also affinely independent.

(b): If $A$ is not surjective, then $\text{aff } (A(V)) = A(V) \neq W$, since $A(V)$ is an affine subspace of $W$ by Prop. 1.28(a). Conversely, if $A$ is surjective, $S \subseteq V$, and $\text{aff } S = V$, then

$$W = A(V) = A(\text{aff } S) \overset{\text{Prop. 1.28(a)}}{=} \text{aff } (A(S)),$$

thereby establishing the case. ■

Caveat 1.30. Unlike in [Phi19, Prop. 6.5(a)], the converse of Prop. 1.29(a) is, in general, not true: If $\dim V \geq 1$ and $A \equiv w \in W$ is constant, then $A$ is affine, not injective, but it maps every nonempty affinely independent subset of $V$ (in fact, every nonempty subset of $V$) onto the affinely independent set $\{w\}$.

Corollary 1.31. Let $V$ and $W$ be vector spaces over the field $F$, and let $A : V \rightarrow W$ be affine and injective. If $M \subseteq V$ is an affine subspace and $B$ is an affine basis of $M$, then $A(B)$ is an affine basis of $A(M)$ (Caveat 1.30 above shows that the converse is, in general, not true).

Proof. Since $B$ is affinely independent, $A(B)$ is affinely independent by Prop. 1.29(a). On the other hand, $A(M) = \text{aff } (A(B))$ by Prop. 1.28(a). ■

The following Prop. 1.32 shows that affine subspaces are precisely the images of vector subspaces under translations and also precisely the sets of solutions to linear systems with nonempty sets of solutions:

Proposition 1.32. Let $V$ be a vector space over the field $F$ and $M \subseteq V$. Then the following statements are equivalent:

(i) $M$ is an affine subspace of $V$.

(ii) There exists $v \in V$ and a vector subspace $U \subseteq V$ such that $M = T_v(U)$. 
There exists a linear map \( L \in \mathcal{L}(V, V) \) and a vector \( b \in V \) such that \( \emptyset \neq M = L^{-1}\{b\} = \{x \in V : Lx = b\} \) (if \( V \) is finite-dimensional, then \( L^{-1}\{b\} = \mathcal{L}(L|b) \), where \( \mathcal{L}(L|b) \) denotes the set of solutions to the linear system \( Lx = b \) according to [Phi19, Th. 5.27(c)].

Proof. “(i)\(\Leftrightarrow\)(ii)”: By the definition of affine subspaces, (i) is equivalent to the existence of \( v \in V \) and a vector subspace \( U \subseteq V \) such that \( M = v + U = T_v(U) \), which is (ii).

“(iii)\(\Rightarrow\)(i)” holds, since the restricted translations \( L \in \mathcal{L}(V, V) \) and \( b \in V \) such that \( \emptyset \neq M = L^{-1}\{b\} \). Let \( x_0 \in M \). Then, by [Phi19, Th. 4.20(d)], \( M = x_0 + \ker L \), showing \( M \) to be an affine subspace.

“(i)\(\Rightarrow\)(iii)” holds, since the restricted translations \( L \in \mathcal{L}(V, V) \) and \( b \in V \) such that \( \emptyset \neq M = L^{-1}\{b\} \). Let \( b = Lx_0 \). Then \( M = L^{-1}\{b\} \). Indeed, if \( u \in U \), then \( L(u + v) = Lv + 0 = Lv = b \), showing \( M \subseteq L^{-1}\{b\} \). If \( L(u + v) = w = b = Lv \), then \( u + w = v + u + w - v = v + U = M \) (since \( L(u + w - v) = Lw - Lv = w - Lv \)).

The following Th. 1.33 is the analogon of [Phi19, Th. 6.9] for affine spaces:

**Theorem 1.33.** Let \( V \) and \( W \) be vector spaces over the field \( F \). Moreover, let \( M_V = v + U_V \subseteq V \) and \( M_W = w + U_W \subseteq W \) be affine subspaces of \( V \) and \( W \), respectively, where \( v \in V \), \( w \in W \), \( U_V \) is a vector subspace of \( V \) and \( U_W \) is a vector subspace of \( W \). Let \( B_V \) be an affine basis of \( M_V \) and let \( B_W \) be an affine basis of \( M_W \). Then the following statements are equivalent:

(i) There exists a linear isomorphism \( L : U_V \longrightarrow U_W \) such that \( M_W = (T_w \circ L \circ T_{-v})(M_V) \).

(ii) \( U_V \) and \( U_W \) are linearly isomorphic.

(iii) \( \dim M_V = \dim M_W \).

(iv) \#\(B_V\) = \#\(B_W\) (i.e. there exists a bijective map from \( B_V \) onto \( B_W \)).

Proof. “(i)\(\Rightarrow\)(ii)” is trivially true.

“(ii)\(\Rightarrow\)(i)” holds, since the restricted translations \( T_{-v} : M_V \longrightarrow U_V \) and \( T_w : U_W \longrightarrow M_W \) are, clearly, bijective.

“(ii)\(\Leftrightarrow\)(iii)” holds, since the restricted translations \( T_{-v} : M_V \longrightarrow U_V \) and \( T_w : U_W \longrightarrow M_W \) are, clearly, bijective.

“(iii)\(\Leftrightarrow\)(iv)” holds, since the restricted translations \( T_{-v} : M_V \longrightarrow U_V \) and \( T_w : U_W \longrightarrow M_W \) are, clearly, bijective. Thus, if \( \dim M_V = \dim M_W \), then there exists a bijective map...
$\phi : S_V \rightarrow S_W$, implying $(T_y \circ \phi \circ T_x) : B_V \setminus \{x\} \rightarrow B_W \setminus \{y\}$ to be bijective as well. Conversely, if $\psi : B_V \setminus \{x\} \rightarrow B_W \setminus \{y\}$ is bijective, so is $(T_{-y} \circ \psi \circ T_x) : S_V \rightarrow S_W$, implying $\dim M_V = \dim M_W$. ■

Analogous to [Phi19, Def. 6.17], we now consider, for vector spaces $V, W$ over the field $F$, $A(V, W)$ with pointwise addition and scalar multiplication, letting, for each $A, B \in A(V, W), \lambda \in F$,

$$(A + B) : V \rightarrow W, \quad (A + B)(x) := A(x) + B(x),$$

$$(\lambda \cdot A) : V \rightarrow W, \quad (\lambda \cdot A)(x) := \lambda \cdot A(x) \quad \text{for each } \lambda \in F.$$

The following Th. 1.34 corresponds to [Phi19, Th. 6.18] and [Phi19, Th. 6.21] for linear maps.

**Theorem 1.34.** Let $V$ and $W$ be vector spaces over the field $F$. Addition and scalar multiplication on $A(V, W)$, given by the pointwise definitions above, are well-defined in the sense that, if $A, B \in A(V, W)$ and $\lambda \in F$, then $A + B \in A(V, W)$ and $\lambda A \in A(V, W)$. Moreover, with these pointwise defined operations, $A(V, W)$ forms a vector space over $F$.

**Proof.** According to [Phi19, Ex. 5.2(c)], it only remains to show that $A(V, W)$ is a vector subspace of $F(V, W) = W^V$. To this end, let $A, B \in A(V, W)$ with $A = T_{w_1} \circ L_1$, $B = T_{w_2} \circ L_2$, where $w_1, w_2 \in W$, $L_1, L_2 \in \mathcal{L}(V, W)$, and let $\lambda \in F$. If $v \in V$, then

$$(A + B)(v) = w_1 + L_1 v + w_2 + L_2 v = w_1 + w_2 + (L_1 + L_2)v,$$

$$(\lambda A)(v) = \lambda w_1 + \lambda L_1 v,$$

proving $A + B = T_{w_1+w_2} \circ (L_1 + L_2) \in A(V, W)$ and $\lambda A = T_{\lambda w_1} \circ (\lambda L_1) \in A(V, W)$, as desired. ■

### 1.6 Affine Geometry

The subject of affine geometry is concerned with the relationships between affine subspaces, in particular, with the way they are contained in each other.

**Definition 1.35.** Let $V$ be a vector space over the field $F$ and let $M, N \subseteq V$ be affine subspaces.

(a) We define the incidence $M \mathcal{I} N$ by

$$M \mathcal{I} N \iff \left( M \subseteq N \lor N \subseteq M \right).$$

If $M \mathcal{I} N$ holds, then we call $M, N$ incident or $M$ incident with $N$ or $N$ incident with $M$.

(b) If $M = v + U_M$ and $N = w + U_N$ with $v, w \in V$ and $U_M, U_N$ vector subspaces of $V$, then we call $M, N$ parallel (denoted $M \parallel N$) if, and only if, $U_M \subseteq U_N$. 

Proposition 1.36. Let $V$ be a vector space over the field $F$ and let $M, N \subseteq V$ be affine subspaces.

(a) If $M \parallel N$, then $M \cap N = \emptyset$.

(b) If $n \in \mathbb{N}_0$ and $\mathcal{A}_n$ denotes the set of affine subspaces with dimension $n$ of $V$, then the parallelity relation of Def. 1.35(b) constitutes an equivalence relation on $\mathcal{A}_n$.

(c) If $\mathcal{A}$ denotes the set of all affine subspaces of $V$, then, for $\dim V \geq 2$, the parallelity relation of Def. 1.35(b) is not transitive (in particular, not an equivalence relation) on $\mathcal{A}$.

Proof. (a): Let $M = v + U_M$ and $N = w + U_N$ with $v, w \in V$ and $U_M, U_N$ vector subspaces of $V$. Without loss of generality, assume $U_M \subseteq U_N$. Assume there exists $x \in M \cap N$. Then, if $y \in M$, then $y - x \in U_M \subseteq U_N$, implying $y = x + (y - x) \in N$ and $M \subseteq N$.

(b): It is immediate from Def. 1.35 that $\parallel$ is reflexive and symmetric. It remains to show $\parallel$ is transitive on $\mathcal{A}_n$. Thus, suppose $M = v + U_M$, $N = w + U_N$, $P = z + U_P$ with $v, w, z \in V$ and $U_M, U_N, U_P$ vector subspaces of dimension $n$ of $V$. If $M \parallel N$, then $U_M \cap U_N$ and $\dim U_M = \dim U_N = n$ implies $U_M = U_N$ by [Phi19, Th. 5.27(d)]. In the same way, $N \parallel P$ implies $U_N = U_P$. But then $U_M = U_P$ and $M \parallel P$, proving transitivity of $\parallel$.

(c): Let $u, w \in V$ be linearly independent, $U := \langle \{u\} \rangle$, $W := \langle \{w\} \rangle$. Then $U \parallel V$ and $W \parallel V$, but $U \nparallel W$ (e.g., due to (a)). \hfill \blacksquare

Caveat 1.37. The statement of Prop. 1.36(b) becomes false if $n \in \mathbb{N}_0$ is replaced by an infinite cardinality: In an adaptation of the proof of Prop. 1.36(c), suppose $V$ is a vector space over the field $F$, where the distinct vectors $v_1, v_2, \ldots$ are linearly independent, and define $B := \{v_i : i \in \mathbb{N}\}$, $U := \langle B \setminus \{v_1\} \rangle$, $W := \langle B \setminus \{v_2\} \rangle$, $X := \langle B \rangle$. Then, clearly, $U \parallel X$ and $W \parallel X$, but $U \nparallel W$ (e.g., due to Prop. 1.36(a)).

Proposition 1.38. Let $V$ be a vector space over the field $F$.

(a) If $x, y \in V$ with $x \neq y$, then there exists a unique line $l \subseteq V$ (i.e. a unique affine subspace $l$ of $V$ with $\dim l = 1$) such that $x, y \in l$. Moreover, this affine subspace is given by

$$l = x + \langle \{x - y\} \rangle. \quad (1.7)$$

(b) If $x, y, z \in V$ and there does not exist a line $l \subseteq V$ such that $x, y, z \in l$, then there exists a unique plane $p \subseteq V$ (i.e. a unique affine subspace $p$ of $V$ with $\dim p = 2$) such that $x, y, z \in p$. Moreover, this affine subspace is given by

$$p = x + \langle \{y - x, z - x\} \rangle. \quad (1.8)$$

(c) If $v_1, \ldots, v_n \in V$, $n \in \mathbb{N}$, then $\text{aff}\{v_1, \ldots, v_n\} = v_1 + \langle \{v_2 - v_1, \ldots, v_n - v_1\} \rangle$. 

2 Duality

Proof. Exercise. ■

**Proposition 1.39.** Let \( V, W \) be vector spaces over the field \( F \) and let \( M, N \subseteq V \) be affine subspaces.

(a) If \( A \in \mathcal{A}(V, W) \), then \( M \parallel N \) implies \( A(M) \parallel A(N) \), and \( M \parallel N \) implies \( A(M) \parallel A(N) \).

(b) If \( v \in V \), then \( T_v(M) \parallel M \).

**Proof.** (a): Let \( A \in \mathcal{A}(V, W) \). Then \( M \parallel N \) implies \( A(M) \parallel A(N) \), since \( M \subseteq N \) implies \( A(M) \subseteq A(N) \) and \( A(M), A(N) \) are affine subspaces of \( W \) due to Prop. 1.28(a). Moreover, if \( M = v + U_M, N = w + U_N \) with \( v, w \in V \) and \( U_M, U_N \) vector subspaces of \( V \), \( A = T_x \circ L \) with \( x \in W \) and \( L \in \mathcal{L}(V, W) \), then \( A(M) = x + Lv + L(U_M) \) and \( A(N) = x + Lw + L(U_N) \), such that \( M \parallel N \) implies \( A(M) \parallel A(N) \), since \( U_M \subseteq U_N \) implies \( L(U_M) \subseteq L(U_N) \).

(b) is immediate from \( T_v(M) = v + w + U \) for \( M = w + U \) with \( w \in V \) and \( U \) a vector subspace of \( V \). ■

2 Duality

2.1 Linear Forms and Dual Spaces

If \( V \) is a vector space over the field \( F \), then maps from \( V \) into \( F \) are often of particular interest and importance. Such maps are sometimes called functionals or forms. Here, we will mostly be concerned with linear forms: Let us briefly review some examples of linear forms that we already encountered in [Phi19]:

**Example 2.1.** Let \( V \) be a vector space over the field \( F \).

(a) Let \( B \) be a basis of \( V \). If \( c_v : B_v \rightarrow F \setminus \{0\}, B_v \subseteq V \), are the corresponding coordinate maps (i.e. \( v = \sum_{b \in B_v} c_v(b) b \) for each \( v \in V \)), then, for each \( b \in B \), the projection onto the coordinate with respect to \( b \),

\[
\pi_b : V \rightarrow F, \quad \pi_b(v) := \begin{cases} 
c_v(b) & \text{for } b \in B_v, \\
0 & \text{for } b \notin B_v,
\end{cases}
\]

is a linear form (cf. [Phi19, Ex. 6.7(b)]).

(b) Let \( I \) be a nonempty set, \( V := \mathcal{F}(I, F) = F^I \) (i.e. the vector space of functions from \( I \) into \( F \)). Then, for each \( i \in I \), the projection onto the \( i \)th coordinate

\[
\pi_i : V \rightarrow F, \quad \pi_i(f) := f(i),
\]

is a linear form (cf. [Phi19, Ex. 6.7(c)]).
(c) Let $F := \mathbb{K}$, where, as in [Phi19], we write $\mathbb{K}$ if $\mathbb{K}$ may stand for $\mathbb{R}$ or $\mathbb{C}$. Let $V$ be the set of convergent sequences in $\mathbb{K}$. Then

$$A : V \rightarrow \mathbb{K}, \quad A(z_n)_{n \in \mathbb{N}} := \lim_{n \to \infty} z_n,$$

is a linear form (cf. [Phi19, Ex. 6.7(e)(i)]).

(d) Let $a, b \in \mathbb{R}$, $a \leq b$, $I := [a, b]$, and let $V := \mathcal{R}(I, \mathbb{K})$ be the set of all $\mathbb{K}$-valued Riemann integrable functions on $I$. Then

$$J : V \rightarrow \mathbb{K}, \quad J(f) := \int_I f,$$

is a linear form (cf. [Phi19, Ex. 6.7(e)(iii)]).

**Definition 2.2.** Let $V$ be a vector space over the field $F$.

(a) The functions from $V$ into $F$ (i.e. the elements of $\mathcal{F}(V, F) = F^V$) are called *functionals* or *forms* on $V$. In particular, the elements of $\mathcal{L}(V, F)$ are called *linear functionals* or *linear forms* on $V$.

(b) The set

$$V' := \mathcal{L}(V, F) \quad (2.1)$$

is called the (linear\(^2\)) *dual space* (or just the *dual*) of $V$ (in the literature, one often also finds the notation $V^*$ instead of $V'$). We already know from [Phi19, Th. 6.18] that $V'$ constitutes a vector space over $F$.

**Corollary 2.3.** Let $V$ be a vector space over the field $F$. Then each linear form $\alpha : V \rightarrow F$ is uniquely determined by its values on a basis of $V$. More precisely, if $B$ is a basis of $V$, $(\lambda_b)_{b \in B}$ is a family in $F$, and, for each $v \in V$, $c_v : B_v \rightarrow F \setminus \{0\}$, $B_v \subseteq V$, is the corresponding coordinate map (i.e. $v = \sum_{b \in B_v} c_v(b) b$ for each $v \in V$), then

$$\forall b \in B_v \quad \hat{\alpha}(b) = \lambda_b,$$

implies $\alpha = \hat{\alpha}$.

**Proof.** Corollary 2.3 constitutes a special case of [Phi19, Th. 6.6].

---

\(^2\)In Functional Analysis, where the vector space $V$ over $\mathbb{K}$ is endowed with the additional structure of a topology (e.g., $V$ might be the normed space $\mathbb{K}^n$), one defines the (topological) dual $V'_{\text{top}}$ of $V$ (there usually also just denoted as $V'$ or $V^*$) to consist of all linear functionals on $V$ that are also *continuous* with respect to the topology on $V$ (cf. [Phi17c, Ex. 3.1]). Depending on the topology on $V$, $V'_{\text{top}}$ can be much smaller than $V' - V'_{\text{top}}$ tends to be much more useful in an analysis context.
Corollary 2.4. Let $V$ be a vector space over the field $F$ and let $B$ be a basis of $V$. Using Cor. 2.3, define linear forms $\alpha_b \in V'$ by letting

$$\forall_{(b,a)\in B \times B} \alpha_b(a) := \delta_{ba} = \begin{cases} 1 & \text{for } a = b, \\ 0 & \text{for } a \neq b. \end{cases} \quad (2.3)$$

Define

$$B' := \{\alpha_b : b \in B\}. \quad (2.4)$$

(a) Then $B'$ is linearly independent.

(b) If $V$ is finite-dimensional, $\dim V = n \in \mathbb{N}$, then $B'$ constitutes a basis for $V'$ (in particular, $\dim V = \dim V'$). In this case, $B'$ is called the dual basis of $B$ (and $B$ the dual basis of $B'$).

(c) If $\dim V = \infty$, then $\langle B' \rangle \subsetneq V'$ and, in particular, $B'$ is not a basis of $V'$ (in fact, in this case, one has $\dim V' > \dim V$, see [Jac75, pp. 244-248]).

Proof. Cor. 2.4(a),(b),(c) constitute special cases of the corresponding cases of [Phi19, Th. 6.19].

Definition 2.5. If $V$ is a vector space over the field $F$ with $\dim V = n \in \mathbb{N}$ and $B := (b_1, \ldots, b_n)$ is an ordered basis of $V$, then we call $B' := (\alpha_1, \ldots, \alpha_n)$, where

$$\forall_{i \in \{1,\ldots,n\}} \left( \alpha_i \in V' \quad \land \quad \alpha_i(b_j) = \delta_{ij} \right), \quad (2.5)$$

the ordered dual basis of $B$ (and $B$ the ordered dual basis of $B'$) – according to Cor. 2.4(b), $B'$ is, indeed, an ordered basis of $V'$.

Example 2.6. Consider $V := \mathbb{R}^2$. If $b_1 := (1,0), b_2 := (1,1)$, then $B := (b_1, b_2)$ is an ordered basis of $V$. Then the ordered dual basis $B' = (\alpha_1, \alpha_2)$ of $V'$ consists of the maps $\alpha_1, \alpha_2 \in V'$ with $\alpha_1(b_1) = \alpha_2(b_2) = 1, \alpha_1(b_2) = \alpha_2(b_1) = 0$, i.e. with, for each $(v_1, v_2) \in V$,

$$\begin{align*}
\alpha_1(v_1, v_2) &= \alpha_1((v_1 - v_2)b_1 + v_2b_2) = v_1 - v_2, \\
\alpha_2(v_1, v_2) &= \alpha_2((v_1 - v_2)b_1 + v_2b_2) = v_2.
\end{align*}$$

Notation 2.7. Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, with ordered basis $B = (b_1, \ldots, b_n)$. Moreover, let $B' = (\alpha_1, \ldots, \alpha_n)$ be the corresponding ordered dual basis of $V'$. If one then denotes the coordinates of $v \in V$ with respect to $B$ as the column vector

$$v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix},$$

then one typically denotes the coordinates of $\gamma \in V'$ with respect to $B'$ as the row vector

$$\gamma = \begin{pmatrix} \gamma_1 & \cdots & \gamma_n \end{pmatrix}$$

(this has the advantage that one then can express $\gamma(v)$ as a matrix product, cf. Rem. 2.8(a) below).
Remark 2.8. We remain in the situation of Not. 2.7 above.

(a) We obtain

\[
\gamma(v) = \left( \sum_{k=1}^{n} \gamma_k \alpha_k \right) \left( \sum_{l=1}^{n} v_l b_l \right) = \sum_{l=1}^{n} \sum_{k=1}^{n} \gamma_l v_k \alpha_k = \sum_{l=1}^{n} \sum_{k=1}^{n} \gamma_l v_k \delta_{kl}
\]

\[
= \sum_{k=1}^{n} \gamma_k v_k = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.
\]

(b) Let \( \tilde{B}_V := (\tilde{v}_1, \ldots , \tilde{v}_n) \) be another ordered basis of \( V \) and \( (c_{ji}) \in GL_n(F) \) such that

\[
\forall i \in \{1, \ldots , n\} \quad \tilde{v}_i = \sum_{j=1}^{n} c_{ji} v_j.
\]

If \( \tilde{B}_V' := (\tilde{\alpha}_1, \ldots , \tilde{\alpha}_n) \) denotes the ordered dual basis corresponding to \( \tilde{B}_V \) and \( (d_{ji}) := (c_{ji})^{-1} \), then

\[
\forall i \in \{1, \ldots , n\} \quad \tilde{\alpha}_i = \sum_{j=1}^{n} d_{ji} \alpha_j = \sum_{j=1}^{n} d_{ij} \alpha_j,
\]

where \( (d_{ji}) \) denotes the transpose of \( (d_{ji}) \), i.e.

\[
(d_{ji}) \in GL_n(F) \quad \text{with} \quad \forall (j,i) \in \{1, \ldots , n\} \times \{1, \ldots , n\} \quad d_{ji}^t := d_{ij}.
\]

Indeed, for each \( k, l \in \{1, \ldots , n\} \), we obtain

\[
\left( \sum_{j=1}^{n} d_{kj} \alpha_j \right) (\tilde{v}_l) = \left( \sum_{j=1}^{n} d_{kj} \alpha_j \right) \left( \sum_{i=1}^{n} c_{il} v_i \right) = \sum_{j=1}^{n} \sum_{i=1}^{n} d_{kj} c_{il} \delta_{ji} = \sum_{j=1}^{n} d_{kj} c_{jl} = \delta_{kl}.
\]

Proposition 2.9. Let \( V \) be a vector space over the field \( F \). If \( U \) is a vector subspace of \( V \) and \( v \in V \setminus U \), then there exists \( \alpha \in V' \), satisfying

\[
\alpha(v) = 1 \quad \wedge \quad \forall u \in U \quad \alpha(u) = 0.
\]

Proof. Let \( B_U \) be a basis of \( U \). Then \( B_v := \{v\} \cup B_U \) is linearly independent and, according to [Phi19, Th. 5.23(a)], there exists a basis \( B \) of \( V \) such that \( B_v \subseteq B \). According to Cor. 2.3,

\[
\alpha : V \rightarrow F, \quad \forall b \in B \quad \alpha(b) := \begin{cases} 1 & \text{for } b = v, \\ 0 & \text{for } b \neq v, \end{cases}
\]

defines an element of \( V' \), which, clearly, satisfies the required conditions. \(\blacksquare\)
Definition 2.10. Let $V$ be a vector space over the field $F$.

(a) The map
\[ \langle \cdot, \cdot \rangle : V \times V' \to F, \quad \langle v, \alpha \rangle := \alpha(v), \] (2.6)
is called the dual pairing corresponding to $V$.

(b) The dual of $V'$ is called the bidual or the second dual of $V$. One writes $V'' := (V')'$.

(c) The map
\[ \Phi : V \to V'', \quad v \mapsto \Phi v, \quad (\Phi v)(\alpha) := \alpha(v) \] (2.7)
is called the canonical embedding of $V$ into $V''$ (cf. Th. 2.11 below).

Theorem 2.11. Let $V$ be a vector space over the field $F$.

(a) The canonical embedding $\Phi : V \to V''$ of (2.7) is a linear monomorphism (i.e. a linear isomorphism $\Phi : V \cong \text{Im} \Phi \subseteq V''$).

(b) If $\dim V = n \in \mathbb{N}$, then $\Phi$ is a linear isomorphism $\Phi : V \cong V''$ (in fact, the converse is also true, i.e., if $\Phi$ is an isomorphism, then $\dim V < \infty$, cf. the remark in Cor. 2.4(c)).

Proof. (a): Exercise.

(b): According to Cor. 2.4(b), $n = \dim V = \dim V' = \dim V''$. Thus, by [Phi19, Th. 6.10], the linear monomorphism $\Phi$ is also an epimorphism. ■

Corollary 2.12. Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$. If $B' = \{\alpha_1, \ldots, \alpha_n\}$ is a basis of $V'$, then there exists a basis $B$ of $V$ such that $B$ and $B'$ are dual.

Proof. According to Th. 2.11(b), the canonical embedding $\Phi : V \to V''$ of (2.7) constitutes a linear isomorphism. Let $B'' = \{f_1, \ldots, f_n\}$ be the basis of $V''$ that is dual to $B'$ and, for each $i \in \{1, \ldots, n\}$, $b_i := \Phi^{-1}(f_i)$. Then, as $\Phi$ is a linear isomorphism, $B := \{b_1, \ldots, b_n\}$ is a basis of $V$. Moreover, $B$ and $B'$ are dual:
\[ \forall_{i,j \in \{1, \ldots, n\}} \alpha_i(b_j) = (\Phi b_j)(\alpha_i) = f_j(\alpha_i) = \delta_{ij}, \]
where we used that $B'$ and $B''$ are dual. ■

2.2 Annihilators

Definition 2.13. Let $V$ be a vector space over the field $F$, $M \subseteq V$, $S \subseteq V'$. Moreover, let $\Phi : V \to V''$ denote the canonical embedding of (2.7). Then
\[ M^\perp := \left\{ \alpha \in V' : \forall_{v \in M} \alpha(v) = 0 \right\} = \begin{cases} V'' & \text{for } M = \emptyset, \\ \bigcap_{v \in M} \ker(\Phi v) & \text{for } M \neq \emptyset \end{cases} \]
is called the \textit{(forward) annihilator} of $M$ in $V'$,
\begin{equation}
S^\top := \left\{ v \in V : \forall \alpha \in S \, \alpha(v) = 0 \right\} = \left\{ \bigcap_{\alpha \in S} \ker \alpha \right\} \quad \text{for } S = \emptyset,
\end{equation}
is called the \textit{(backward) annihilator} of $S$ in $V$. In view of Rem. 2.15 and Ex. 2.16(b) below, one also calls $v \in V$ and $\alpha \in V'$ such that
\begin{equation}
\alpha(v) \overset{\text{(2.6)}}{=} \langle v, \alpha \rangle = 0
\end{equation}
perpendicular or orthogonal and, in consequence, sets $M^\perp$ and $S^\top$ are sometimes called $M$ perp and $S$ perp, respectively.

\textbf{Lemma 2.14.} Let $V$ be a vector space over the field $F$, $M \subseteq V$, $S \subseteq V'$. Then $M^\perp$ is a subspace of $V'$ and $S^\top$ is a subspace of $V$. Moreover,
\begin{equation}
M^\perp = \langle M \rangle^\perp, \quad S^\top = \langle S \rangle^\top.
\end{equation}

\textit{Proof.} Since $M^\perp$ and $S^\top$ are both intersections of kernels of linear maps, they are subspaces, since kernels are subspaces by [Phi19, Prop. 6.3(c)] and intersections of subspaces are subspaces by [Phi19, Th. 5.7(a)]. Moreover, it is immediate from Def. 2.13 that $M^\perp \supseteq \langle M \rangle^\perp$ and $S^\top \supseteq \langle S \rangle^\top$. On the other hand, consider $\alpha \in M^\perp$ and $v \in S^\top$. Let $\alpha_1, \ldots, \alpha_n \in F$, $n \in \mathbb{N}$. If $v_1, \ldots, v_n \in M$, then
\begin{equation}
\alpha \left( \sum_{i=1}^n \lambda_i v_i \right) = \sum_{i=1}^n \lambda_i \alpha(v_i) \overset{\text{Def. 2.13}}{=} 0,
\end{equation}
showing $\alpha \in \langle M \rangle^\perp$ and $M^\perp \subseteq \langle M \rangle^\perp$. Analogously, if $\alpha_1, \ldots, \alpha_n \in S$, then
\begin{equation}
\left( \sum_{i=1}^n \lambda_i \alpha_i \right)(v) = \sum_{i=1}^n \lambda_i \alpha_i(v) \overset{\text{Def. 2.13}}{=} 0,
\end{equation}
showing $v \in \langle S \rangle^\top$ and $S^\top \subseteq \langle S \rangle^\top$. \hfill \blacksquare

\textbf{Remark 2.15.} On real vector spaces $V$, one can study so-called \textit{scalar products} (also called \textit{inner products}), $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$, $(v, w) \mapsto \langle v, w \rangle \in \mathbb{R}$, which, as part of their definition, have the requirement of being \textit{bilinear forms}, i.e., for each $v \in V$, $\langle v, \cdot \rangle : V \rightarrow \mathbb{R}$ is a linear form and, for each $w \in V$, $\langle \cdot, w \rangle : V \rightarrow \mathbb{R}$ is a linear form (we will come back to vector spaces with scalar products again in a later section). One then calls vectors $v, w \in V$ \textit{perpendicular} or \textit{orthogonal} with respect to $\langle \cdot, \cdot \rangle$ if, and only if, $\langle v, w \rangle = 0$ so that the notions of Def. 2.13 can be seen as generalizing orthogonality with respect to scalar products (also cf. Ex. 2.16(b) below).

\textbf{Example 2.16. (a)} Let $V$ be a vector space over the field $F$ and let $U$ be a subspace of $V$ with $B_U$ being a basis of $U$. Then, according to [Phi19, Th. 5.23(a)], there exists a basis $B$ of $V$ such that $B_U \subseteq B$. Then Cor. 2.3 implies
\begin{equation}
\forall \alpha \in V' \left( \alpha \in U^\perp \iff \forall b \in B_U \left( \alpha(b) = 0 \right) \right).
\end{equation}
(b) Consider the real vector space $\mathbb{R}^2$ and let
\[ \langle \cdot, \cdot \rangle : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}, \quad \langle (v_1, v_2), (w_1, w_2) \rangle := v_1 w_1 + v_2 w_2, \]
denote the so-called Euclidean scalar product on $\mathbb{R}^2$. Then, clearly, for each $w = (w_1, w_2) \in \mathbb{R}^2$,
\[ \alpha_w : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad \alpha_w(v) := \langle v, w \rangle = v_1 w_1 + v_2 w_2, \]
defines a linear form on $\mathbb{R}^2$. Let $v := (1, 2)$. Then the span of $v$, i.e. $l_v := \{(\lambda, 2\lambda) : \lambda \in \mathbb{R}\}$, represents the line through $v$. Moreover, for each $w = (w_1, w_2) \in \mathbb{R}^2$,
\[ \alpha_w \in \{v\}^\perp = l_v^\perp \iff \alpha_w(v) = w_1 + 2w_2 = 0 \iff w_1 = -2w_2 \iff w \in l_v^\perp := \{(-2\lambda, \lambda) : \lambda \in \mathbb{R}\}. \]
Thus, $l_v^\perp$ is spanned by $(-2, 1)$ and we see that $l_v^\perp$ consists precisely of the linear forms $\alpha_w$ that are given by vectors $w$ that are perpendicular to $v$ in the Euclidean geometrical sense (i.e. in the sense usually taught in high school geometry).

The following notions defined for linear forms in connection with subspaces can sometimes be useful when studying annihilators:

**Definition 2.17.** Let $V$ be a vector space over the field $F$ and let $U$ be a subspace of $V$. Then
\[ R : V' \rightarrow U', \quad Rf := f\lfloor_U, \]
is called the restriction operator from $V$ to $U$;
\[ I : (V/U)' \rightarrow V', \quad (Ig)(v) := g(v + U), \]
is called the inflation operator from $V/U$ to $V$.

**Theorem 2.18.** Let $V$ be a vector space over the field $F$ and let $U$ be a subspace of $V$ with the restriction operator $R$ and the inflation operator $I$ defined as in Def. 2.17.

(a) $R : V' \rightarrow U'$ is a linear epimorphism with $\ker R = U^\perp$. Moreover,
\[ \dim U^\perp + \dim U' = \dim V' \quad (2.9) \]
and
\[ U' \cong V'/U^\perp. \quad (2.10) \]
(see [Phi19, Th. 6.8(a)] for the precise meaning of (2.9) in case at least one of the occurring cardinalities is infinite). If $\dim V = n \in \mathbb{N}$, then one also has
\[ n = \dim V = \dim U^\perp + \dim U. \quad (2.11) \]

(b) $I$ is a linear isomorphism $I : (V/U)' \cong U^\perp$. 
Proof. (a): Let \( \alpha, \beta \in V' \) and \( \lambda, \mu \in F \). Then, for each \( u \in U \),

\[
R(\lambda \alpha + \mu \beta)(u) = \lambda \alpha(u) + \mu \beta(u) = \lambda(R\alpha)(u) + \mu(R\beta)(u) = (\lambda(R\alpha) + \mu(R\beta))(u),
\]

showing \( R \) to be linear. Moreover, for each \( \alpha \in V' \), one has

\[
\alpha \in \ker R \iff \forall u \in U \alpha(u) = 0 \iff \alpha \in U^\perp,
\]

proving \( \ker R = U^\perp \). Let \( B_U \) be a basis of \( U \). Then, according to [Phi19, Th. 5.23(a)], there exists \( C \subseteq V \) such that \( B_U \cup C \) is a basis of \( V \). Consider \( \alpha \in U' \). Using Cor. 2.3, define \( \beta \in V' \) by setting

\[
\beta(b) := \begin{cases} 
\alpha(b) & \text{for } b \in B_U, \\
0 & \text{for } b \in C.
\end{cases}
\]

Then, clearly, \( R\beta = \alpha \), showing \( R \) to be surjective. Thus, we have

\[
\dim V' \overset{\text{[Phi19, Th. 6.8(a)]}}{=} \dim \ker R + \dim \operatorname{Im} R = \dim U^\perp + \dim U',
\]

thereby proving (2.9). Next, applying the isomorphism theorem of [Phi19, Th. 6.16(a)] yields

\[
U' = \operatorname{Im} R \cong V'/\ker R = V'/U^\perp,
\]

which is (2.10). Finally, if \( \dim V = n \in \mathbb{N} \), then

\[
n = \dim V \overset{\text{Cor. 2.4(b)}}{=} \dim V' \overset{(2.9)}{=} \dim U^\perp + \dim U' \overset{\text{Cor. 2.4(b)}}{=} \dim U^\perp + \dim U,
\]

proving (2.11).

(b): Exercise. \( \blacksquare \)

**Theorem 2.19.** Let \( V \) be a vector space over the field \( F \).

(a) If \( U \) is a subspace of \( V \), then \( (U^\perp)^\top = U \).

(b) If \( S \) is a subspace of \( V' \), \( S \subseteq (S^\top)^\perp \). If \( \dim V = n \in \mathbb{N} \), then one even has \( (S^\top)^\perp = S \).

(c) If \( U_1, U_2 \) are subspaces of \( V \), then

\[
(U_1 + U_2)^\perp = U_1^\perp \cap U_2^\perp, \quad (U_1 \cap U_2)^\perp = U_1^\perp + U_2^\perp.
\]

(d) If \( S_1, S_2 \) are subspaces of \( V' \), then

\[
(S_1 + S_2)^\top = S_1^\top \cap S_2^\top, \quad (S_1 \cap S_2)^\top \supseteq S_1^\top + S_2^\top.
\]

If \( \dim V = n \in \mathbb{N} \), then one also has

\[
(S_1 \cap S_2)^\top = S_1^\top + S_2^\top.
\]
Proof. (a): Exercise.
(b): According to Def. 2.13, we have
\[(S^T)^\perp := \{ \alpha \in V' : \forall v \in S^T \alpha(v) = 0 \},\]
showing \( S \subseteq (S^T)^\perp \). Now assume \( \dim V = n \in \mathbb{N} \) and suppose there exists \( \alpha \in (S^T)^\perp \setminus S \).
Then, according to Prop. 2.9, there exists \( f(a) \): Exercise.
Proof. (c): Exercise.
Proof. (d): Exercise.

2 DUALITY

2.3 Hyperplanes and Linear Systems

In the present section, we combine duality with the theory of affine spaces of Sec. 1 and with the theory of linear systems of [Phi19, Sec. 8].
Definition 2.20. Let $V$ be a vector space over the field $F$. If $\alpha \in V' \setminus \{0\}$ and $r \in F$, then the set
\[ H_{\alpha,r} := \alpha^{-1}\{r\} = \{v \in V : \alpha(v) = r\} \subseteq V \]
is called a hyperplane in $V$.

Notation 2.21. Let $V$ be a vector space over the field $F$, $v \in V$, and $\alpha \in V'$. We then write
\[ v^\perp := \{v\}^\perp, \quad \alpha^\top := \{\alpha\}^\top. \]

Theorem 2.22. Let $V$ be a vector space over the field $F$.

(a) Each hyperplane $H$ in $V$ is an affine subspace of $V$, where $\dim V = 1 + \dim H$, i.e. $\dim V = \dim H$ if $V$ is infinite-dimensional, and $\dim H = n - 1$ if $\dim V = n \in \mathbb{N}$. More precisely, if $0 \neq \alpha \in V'$ and $r \in F$, then
\[ \forall w \in V : \alpha(w) \neq 0 \quad H_{\alpha,r} = r \frac{w}{\alpha(w)} + \alpha^\top = r \frac{w}{\alpha(w)} + \ker \alpha. \quad (2.12) \]

(b) If $\dim V = n \in \mathbb{N}$ and $M$ is an affine subspace of $V$ with $\dim M = n - 1$, then $M$ is a hyperplane in $V$, i.e. there exist $0 \neq \alpha \in V'$ and $r \in F$ such that $M = H_{\alpha,r}$.

(c) Let $\alpha, \beta \in V' \setminus \{0\}$ and $r, s \in F$. Then
\[ H_{\alpha,r} = H_{\beta,s} \iff \exists \lambda \in F \quad \left( \beta = \lambda \alpha \land s = \lambda r \right). \]
Moreover, if $\alpha = \beta$, then $H_{\alpha,r}$ and $H_{\beta,s}$ are parallel.

Proof. (a): Let $0 \neq \alpha \in V'$ and $r \in F$ with $w \in V$ such that $\alpha(w) \neq 0$. Define $v := r \frac{w}{\alpha(w)}$. Then $\alpha(v) = r$, i.e. $v \in H_{\alpha,r} = \alpha^{-1}\{r\}$. Thus, by [Phi19, Th. 4.20(d)], we have
\[ H_{\alpha,r} = v + \ker \alpha = v + \{x \in V : \alpha(x) = 0\} = v + \alpha^\top, \]
proving (2.12). In particular, we have $\dim H_{\alpha,r} = \dim \ker \alpha$ and, by [Phi19, Th. 6.8(a)],
\[ \dim V = \dim \ker \alpha + \dim \text{Im} \alpha = \dim H_{\alpha,r} + \dim F = \dim H_{\alpha,r} + 1, \]
therby completing the proof of (a).

(b),(c): Exercise.

Proposition 2.23. Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$. If $M \subseteq V$ is an affine subspace of $V$ with $\dim M = m \in \mathbb{N}_0$, $m < n$, then $M$ is the intersection of $n - m$ hyperplanes in $V$.

Proof. If $\dim M = m$, then there exists a linear subspace $U$ of $V$ with $\dim U = m$ and $v \in V$ such that $M = v + U$. Then, according to (2.11), $\dim U^\perp = n - m$. Let $\{\alpha_1, \ldots, \alpha_{n-m}\}$ be a basis of $U^\perp$. Define
\[ \forall i \in \{1, \ldots, n-m\} \quad r_i := \alpha_i(v). \]
We claim $M = N := \bigcap_{i=1}^{n-m} H_{\alpha_i,r_i}$: Indeed, if $x = v + u$ with $u \in U$, then

$$\forall_{i \in \{1, \ldots, n-m\}} \alpha_i(x) = \alpha_i(v) + \alpha_i(u) \alpha_i \subseteq U \Rightarrow r_i + 0 = r_i,$$

showing $x \in N$ and $M \subseteq N$. Conversely, let $x \in N$. Then

$$\forall_{i \in \{1, \ldots, n-m\}} \alpha_i(x - v) = r_i - r_i = 0, \text{ i.e. } x - v \in \alpha_i^\top,$$

implying

$$x - v \in \langle \{\alpha_1, \ldots, \alpha_{n-m}\} \rangle^\top = (U^\top)^{\text{Th. 2.19(a)}} U,$$

showing $x \in v + U = M$ and $N \subseteq M$ as claimed.

**Example 2.24.** Let $F$ be a field. As in [Phi19, Sec. 8.1], consider the linear system

$$\forall_{j \in \{1, \ldots, m\}} \sum_{k=1}^{n} a_{jk} x_k = b_j, \tag{2.13}$$

where $m, n \in \mathbb{N}$; $b_1, \ldots, b_m \in F$ and the $a_{ji} \in F$, $j \in \{1, \ldots, m\}$, $i \in \{1, \ldots, n\}$.

We know that we can also write (2.13) in matrix form as $Ax = b$ with $A = (a_{ji}) \in \mathcal{M}(m, n, F)$, $m, n \in \mathbb{N}$, and $b = (b_1, \ldots, b_m) \in \mathcal{M}(m, 1, F) \cong F^m$. The set of solutions to (2.13) is

$$\mathcal{L}(A|b) = \{x \in F^n : Ax = b\}.$$

If we now define the linear forms

$$\forall_{j \in \{1, \ldots, m\}} \alpha_j : F^n \rightarrow F, \quad \alpha_j \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} := (a_{j1} \ldots a_{jn}) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \sum_{k=1}^{n} a_{jk} v_k,$$

then we can rewrite (2.13) as

$$\left(\forall_{j \in \{1, \ldots, m\}} x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in H_{\alpha_j,b_j}\right) \iff x \in \bigcap_{j=1}^{m} H_{\alpha_j,b_j}. \tag{2.14}$$

Thus, we have $\mathcal{L}(A|b) = \bigcap_{j=1}^{m} H_{\alpha_j,b_j}$ and we can view (2.14) as a geometric interpretation of (2.13), namely that the solution vectors $x$ are required to lie in the intersection of the $m$ hyperplanes $H_{\alpha_1,b_1}, \ldots, H_{\alpha_m,b_m}$. Even though we know from [Phi19, Th. 8.15] that the elementary row operations of [Phi19, Def. 8.13] do not change the set of solutions $\mathcal{L}(A|b)$, it might be instructive to reexamine this fact in terms of linear forms and hyperplanes: The elementary row operation of row switching merely corresponds to changing the order of the $H_{\alpha_j,b_j}$ in the intersection yielding $\mathcal{L}(A|b)$. The elementary row operation of row multiplication $r_j \mapsto \lambda r_j \ (0 \neq \lambda \in F)$ does not change $\mathcal{L}(A|b)$ due to $H_{\alpha_j,b_j} = H_{\lambda \alpha_j,\lambda b_j}$ according to Th. 2.22(c). The elementary row operation of row addition $r_j \mapsto r_j + \lambda r_i$
(λ ∈ F, i ≠ j) replaces $H_{α_j, b_j}$ by $H_{α_j + λα_i, b_j + λb_i}$. We verify, once again, what we already know form [Phi19, Th. 8.15], namely

$$L(A|b) = \bigcap_{k=1}^{m} H_{α_k, b_k} = M := \left( \bigcap_{k=1}^{m} H_{α_k, b_k} \right) \cap H_{α_j + λα_i, b_j + λb_i}:$$

If $x ∈ L(A|b)$, then $(α_j + λα_i)(x) = b_j + λb_i$, showing $x ∈ H_{α_j + λα_i, b_j + λb_i}$ and $x ∈ M$. Conversely, if $x ∈ M$, then $α_j(x) = (α_j + λα_i)(x) - λα_i(x) = b_j + λb_i - λb_i = b_j$, showing $x ∈ H_{α_j, b_j}$ and $x ∈ L(A|b)$.

### 2.4 Dual Maps

**Theorem 2.25.** Let $V, W$ be vector spaces over the field $F$. If $A ∈ L(V, W)$, then there exists a unique map $A' : W' → V'$ such that (using the notation of (2.6))

$$\forall \ v ∈ V, \ ∀ \ β ∈ W', \ (A'β)(v) = \langle v, A'β \rangle = \langle Av, β \rangle = β(Av). \quad (2.15)$$

Moreover, this map turns out to be linear, i.e. $A' ∈ L(W', V')$.

**Proof.** Clearly, given $A ∈ L(V, W)$, (2.15) uniquely defines a map $A' : W' → V'$ (for each $β ∈ W'$, (2.15) defines the map $(A'β) = β ∘ A ∈ L(V, F) = V'$). It merely remains to check that $A'$ is linear. To this end, let $β, β_1, β_2 ∈ W'$, $λ ∈ F$, and $v ∈ V$. Then

$$(A'(β_1 + β_2))(v) = (β_1 + β_2)(Av) = β_1(Av) + β_2(Av) = (A'β_1)(v) + (A'β_2)(v)$$

$$= (A'β_1 + A'β_2)(v),$$

$$(A'(λβ))(v) = (λβ)(Av) = λ(A'β)(v),$$

showing $A'(β_1 + β_2) = A'β_1 + A'β_2$, $A'(λβ) = λA'(β)$, and the linearity of $A'$.

**Definition 2.26.** Let $V, W$ be vector spaces over the field $F$, $A ∈ L(V, W)$. Then the map $A' ∈ L(W', V')$ given by Th. 2.26 is called the dual map corresponding to $A$ (or the transpose of $A$).

**Theorem 2.27.** Let $F$ be a field, $m, n ∈ \mathbb{N}$. Let $V, W$ be finite-dimensional vector spaces over $F$, $\dim V = n, \dim W = m$, where $B_V := (v_1, \ldots, v_n)$ is an ordered basis of $V$ and $B_W := (w_1, \ldots, w_m)$ is an ordered basis of $W$. Moreover, let $B_V' = (α_1, \ldots, α_n)$ and $B_W' = (β_1, \ldots, β_m)$ be the corresponding (ordered) dual bases of $V'$ and $W'$, respectively. If $A ∈ L(V, W)$ and $(a_{ji}) ∈ M(m, n, F)$ is the matrix corresponding to $A$ with respect to $B_V$ and $B_W$, then the transpose of $(a_{ji})$, i.e.

$$(a'_{ji})_{(j,i)∈\{1,...,n\}×\{1,...,m\}} ∈ M(n, m, F), \quad \forall \ (j,i)∈\{1,...,n\}×\{1,...,m\} \quad a'_{ji} := a_{ij}$$

is the matrix corresponding to $A'$ with respect to $B_W'$ and $B_V'$. 
Proof. If \((a_{ji})\) is the matrix corresponding to \(A\) with respect to \(B_V\) and \(B_W\), then (cf. [Phi19, Th. 7.10(b)])

\[
\forall \quad Av_i = \sum_{j=1}^{m} a_{ji}w_j,
\]

and we have to show

\[
\forall \quad A'\beta_j = \sum_{i=1}^{n} a_{ij}^\top \alpha_i = \sum_{i=1}^{n} a_{ji} \alpha_i.
\]

(2.17)

Indeed, one computes, for each \(j \in \{1, \ldots, m\}\) and for each \(k \in \{1, \ldots, n\}\),

\[
(A'\beta_j)v_k = \beta_j(Av_k) = \beta_j \left( \sum_{l=1}^{m} a_{lk}w_l \right) = \sum_{l=1}^{m} a_{lk}\beta_j(w_l) = \sum_{l=1}^{m} a_{lk}\delta_{jl} = a_{jk}
\]

\[
= \sum_{i=1}^{n} a_{ji}\delta_{ik} = \sum_{i=1}^{n} a_{ji}\alpha_i(v_k) = \left( \sum_{i=1}^{n} a_{ji} \alpha_i \right) v_k,
\]

thereby proving (2.17).

\[\blacksquare\]

Remark 2.28. As in Th. 2.27, let \(m, n \in \mathbb{N}\), let \(V, W\) be finite-dimensional vector spaces over the field \(F\), \(\dim V = n\), \(\dim W = m\), where \(B_V := (v_1, \ldots, v_n)\) is an ordered basis of \(V\) and \(B_W := (w_1, \ldots, w_m)\) is an ordered basis of \(W\) with corresponding (ordered) dual bases \(B'_V = (\alpha_1, \ldots, \alpha_n)\) of \(V'\) and \(B'_W = (\beta_1, \ldots, \beta_m)\) of \(W'\). Let \(A \in \mathcal{L}(V, W)\) with dual map \(A' \in \mathcal{L}(W', V')\).

(a) If \((a_{ji}) \in \mathcal{M}(m, n, F)\) is the matrix corresponding to \(A\) with respect to \(B_V\) and \(B_W\) and if one represents elements of the duals as column vectors, then, according to Th. 2.27, one obtains, for \(\gamma = \sum_{i=1}^{m} \gamma_i \beta_i \in W'\) and \(\epsilon := A'(\gamma) = \sum_{i=1}^{n} \epsilon_i \alpha_i \in V'\) with \(\gamma_1, \ldots, \gamma_m, \epsilon_1, \ldots, \epsilon_n \in F\),

\[
\begin{pmatrix}
\epsilon_1 \\
\vdots  \\
\epsilon_n
\end{pmatrix} = (a_{ji})^\top
\begin{pmatrix}
\gamma_1 \\
\vdots  \\
\gamma_m
\end{pmatrix}.
\]

However, if one adopts the convention of Not. 2.7 to represent elements of the duals as row vectors, then one applies transposes in the above equation to obtain

\[
\begin{pmatrix}
\epsilon_1 & \ldots & \epsilon_n
\end{pmatrix} = \begin{pmatrix}
\gamma_1 & \ldots & \gamma_m
\end{pmatrix} (a_{ji}),
\]

showing that this notation allows \(A\) and \(A'\) to be represented by the same matrix \((a_{ji})\).

(b) As in [Phi19, Th. 7.14] and Rem. 2.8(b) above, we now consider basis transitions

\[
\forall \quad v_i = \sum_{j=1}^{n} c_{ji}v_j, \quad i \in \{1, \ldots, n\}, \quad \forall \quad w_i = \sum_{j=1}^{m} f_{ji}w_j, \quad i \in \{1, \ldots, m\}.
\]
\( \tilde{B}_V := (\tilde{v}_1, \ldots, \tilde{v}_n), \tilde{B}_W := (\tilde{w}_1, \ldots, \tilde{w}_m) \). We then know from [Phi19, Th. 7.14] that the matrix representing \( A \) with respect to \( \tilde{B}_V \) and \( \tilde{B}_W \) is \( (f_{ji})^{-1}(a_{ji})(c_{ji}) \). Thus, according to Th. 2.26, the matrix representing \( A' \) with respect to the dual bases \( \tilde{B}'_V = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_n) \) and \( \tilde{B}'_W = (\tilde{\beta}_1, \ldots, \tilde{\beta}_m) \) is \( ((f_{ji})^{-1}(a_{ji})(c_{ji}))^\top = (c_{ji})^\top(a_{ji})^\top((f_{ji})^{-1})^\top \).

Of course, we can, alternatively, observe that, by Rem. 2.8(b), the basis transition from \( B_{W'} \) to \( \tilde{B}'_V \) is given by \( ((c_{ji})^{-1})^\top \) and the basis transition from \( B_{W'} \) to \( \tilde{B}'_W \) is given by \( ((f_{ji})^{-1})^\top \) and compute the matrix representing \( A' \) with respect to \( \tilde{B}'_V \) and \( \tilde{B}'_W \) via Th. 2.26 and [Phi19, Th. 7.14] to obtain \( (c_{ji})^\top(a_{ji})^\top((f_{ji})^{-1})^\top \), as before. If \( \gamma = \sum_{i=1}^m \gamma_i \tilde{\alpha}_i \in W' \) and \( \epsilon := A'(\gamma) = \sum_{i=1}^n \epsilon_i \tilde{\alpha}_i \in V' \) with \( \gamma_1, \ldots, \gamma_m, \epsilon_1, \ldots, \epsilon_n \in F \), then this yields

\[
(\epsilon_1 \ldots \epsilon_n) = (\gamma_1 \ldots \gamma_m)(f_{ji})^{-1}(a_{ji})(c_{ji}).
\]

(c) Comparing with [Phi19, Rem. 7.24], we observe that the dual map \( A' \in \mathcal{L}(W', V') \) is precisely the transpose map \( A^\top \) of the map \( A \) considered in [Phi19, Rem. 7.24]. Moreover, as a consequence of Th. 2.26, the rows of the matrix \( (a_{ji}) \), representing \( A \), span \( \text{Im} A' \) in the same way that the columns of \( (a_{ji}) \) span \( \text{Im} A \).

**Theorem 2.29.** Let \( V, W \) be vector spaces over the field \( F \).

(a) The duality map \( ' : \mathcal{L}(V, W) \longrightarrow \mathcal{L}(W', V'), \ A \mapsto A' \), is linear.

(b) If \( X \) is another vector space over \( F \), \( A \in \mathcal{L}(V, W) \) and \( B \in \mathcal{L}(W, X) \), then

\[
(BA)' = A'B'.
\]

**Proof.** (a): Let \( A, B \in \mathcal{L}(V, W) \), \( \lambda \in F \), \( \beta \in W' \), and \( v \in V \). Then we compute

\[
(A + B)'(\beta)(v) = \beta((A + B)(v)) = \beta(Av + Bv) = \beta(Av) + \beta(Bv) = (A'\beta)(v) + (B'\beta)(v) = (A' + B')(\beta)(v),
\]

\[
(\lambda A)'(\beta)(v) = \beta((\lambda A)(v)) = \lambda\beta(Av) = (\lambda A')(\beta)(v),
\]

showing \( (A + B)' = A' + B' \), \( (\lambda A)' = \lambda A' \), and the linearity of \( ' \).

(b): If \( \gamma \in X' \), then

\[
(B \circ A)'(\gamma) = \gamma \circ (B \circ A) = (\gamma \circ B) \circ A = A'(\gamma \circ B) = A'(B'\gamma) = (A' \circ B')(\gamma),
\]

showing \( (B \circ A)' = A' \circ B' \).  

**Theorem 2.30.** Let \( V, W \) be vector spaces over the field \( F \) and \( A \in \mathcal{L}(V, W) \).

(a) \( \ker A' = (\text{Im} A)^\perp \).

(b) \( \ker A = (\text{Im} A')^\perp \).

(c) \( A \) is an epimorphism if, and only if, \( A' \) is a monomorphism.
(d) If $A'$ is an epimorphism, then $A$ is a monomorphism. If $A$ is a monomorphism and $\dim V = n \in \mathbb{N}$, then $A'$ is an epimorphism.

(e) If $A'$ is an isomorphism, then $A$ is an isomorphism. If $A$ is an isomorphism and $\dim V = n \in \mathbb{N}$, then $A'$ is an isomorphism.

Proof. (a) is due to the equivalence

$$\beta \in \ker A' \iff \left( \forall v \in V, \beta(Av) = (A'\beta)(v) = 0 \right) \iff \beta \in (\Im A)^\perp.$$ 

(b): Exercise.

(c): If $A$ is an epimorphism, then $\Im A = W$, implying

$$\ker A' \overset{(a)}{=} (\Im A)^\perp = W^\perp = \{0\},$$

showing $A'$ to be a monomorphism. Conversely, if $A'$ is a monomorphism, then $\ker A' = \{0\}$, implying

$$\Im A \overset{\text{Th. 2.19(a)}}{=} ((\Im A)^\perp)^\perp \overset{(a)}{=} (\ker A')^\perp = \{0\}^\perp = W,$$

showing $A$ to be an epimorphism.

(d): Exercise.

(e) is now immediate from combining (c) and (d).

Theorem 2.31. Let $V, W$ be vector spaces over the field $F$ with canonical embeddings $\Phi_V : V \rightarrow V''$ and $\Phi_W : W \rightarrow W''$ according to Def. 2.10(c). Let $A \in \mathcal{L}(V, W)$ and $A'' := (A')' \in \mathcal{L}(V'', W'')$. Then we have

$$\Phi_W \circ A = A'' \circ \Phi_V.$$  \hspace{1cm} (2.18)

Proof. If $v \in V$ and $\beta \in W'$, then

$$((\Phi_W \circ A)(v))(\beta) = \beta(Av) = (A'\beta)(v) = (\Phi_V v)(A'\beta) = ((\Phi_V v) \circ A')(\beta) = (A''(\Phi_V v))(\beta) = ((A'' \circ \Phi_V)(v))(\beta)$$

proves (2.18).

3 Symmetric Groups

In preparation for the introduction of the notion of determinant (which we will find to be a useful tool to further study linear endomorphisms between finite-dimensional vector spaces), we revisit the symmetric group $S_n$ of [Phi19, Ex. 4.9(b)].
Definition 3.1. Let $k, n \in \mathbb{N}$, $k \leq n$. A permutation $\pi \in S_n$ is called a $k$-cycle if, and only if, there exist $k$ distinct numbers $i_1, \ldots, i_k \in \{1, \ldots, n\}$ such that

$$
\pi(i) = \begin{cases} 
  i_{j+1} & \text{if } i = i_j, j \in \{1, \ldots, k-1\}, \\
  i_1 & \text{if } i = i_k, \\
  i & \text{if } i \notin \{i_1, \ldots, i_k\}.
\end{cases}
$$

(3.1)

A 2-cycle is also known as a transposition.

Notation 3.2. Let $n \in \mathbb{N}$, $\pi \in S_n$.

(a) One writes

$$
\pi = \left( \begin{array}{c}
  i_1 \\
  \pi(i_1) \\
  \vdots \\
  \pi(i_n)
\end{array} \right).
$$

where $\{i_1, \ldots, i_n\} = \{1, \ldots, n\}$.

(b) If $\pi$ is a $k$-cycle as in (3.1), $k \in \mathbb{N}$, then one also writes

$$
\pi = (i_1 \ i_2 \ \ldots \ i_k).
$$

(3.2)

Example 3.3. (a) Consider $\pi \in S_5$,

$$
\pi = \begin{pmatrix}
  5 & 4 & 3 & 2 & 1 \\
  3 & 4 & 1 & 2 & 5
\end{pmatrix} = (3\ 1\ 5).
$$

Then

$$
\pi(1) = 5, \quad \pi(2) = 2, \quad \pi(3) = 1, \quad \pi(4) = 4, \quad \pi(5) = 3.
$$

(b) Letting $\pi \in S_5$ be as in (a), we have (recalling that the composition on $S_n$ is merely the usual composition of maps)

$$
\pi = \begin{pmatrix}
  5 & 4 & 3 & 2 & 1 \\
  4 & 1 & 2 & 5 & 3
\end{pmatrix} \begin{pmatrix}
  1 & 2 & 3 & 4 & 5 \\
  2 & 3 & 4 & 5 & 1
\end{pmatrix} = (3\ 5)(3\ 1).
$$

Lemma 3.4. Let $\alpha, k, n \in \mathbb{N}$, $\alpha \leq k \leq n$, and consider distinct numbers $i_1, \ldots, i_k \in \{1, \ldots, n\}$. Then, in $S_n$, the following statements hold true:

(a) One has

$$
(i_1 \ i_2 \ \ldots \ i_k) = (i_\alpha \ i_{\alpha+1} \ \ldots \ i_k \ i_1 \ \ldots \ i_{\alpha-1}).
$$

(b) Let $n \geq 2$. If $1 < \alpha < k$, then

$$
(i_1 \ i_2 \ \ldots \ i_k)(i_1 \ i_\alpha) = (i_1 \ i_{\alpha+1} \ \ldots \ i_k)(i_2 \ \ldots \ i_\alpha);
$$

and, moreover,

$$
(i_1 \ i_2 \ \ldots \ i_k)(i_1 \ i_k) = (i_2 \ \ldots \ i_k)(i_1).
$$
(c) Let \( n \geq 2 \). Given \( \beta, l \in \mathbb{N} \), \( \beta \leq l \leq n \), and distinct numbers \( j_1, \ldots, j_l \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\} \), one has
\[
(i_1 \ldots i_k)(j_1 \ldots j_l)(i_1 j_1) = (j_1 i_2 \ldots i_k i_1 j_2 \ldots j_l).
\]

Proof. Exercise. 

**Notation 3.5.** Let \( M, N \) be sets. Define
\[
S(M, N) := \{ (f : M \to N) : f \text{ bijective} \}.
\]

**Proposition 3.6.** Let \( M, N \) be sets with \( \#M = \#N = n \in \mathbb{N}_0 \), \( S := S(M, N) \) (cf. Not. 3.5). Then \( \#S = n! \); in particular \( \#S_M = n! \).

Proof. We conduct the proof via induction: If \( n = 0 \), then \( S \) contains precisely the empty map (i.e. the empty set) and \( \#S = 1 = 0! \) is true. If \( n = 1 \) and \( M = \{a\}, N = \{b\} \), then \( S \) contains precisely the map \( f : M \to N, f(a) = b \), and \( \#S = 1 = 1! \) is true. For the induction step, fix \( n \in \mathbb{N} \) and assume \( \#M = \#N = n + 1 \). Let \( a \in M \) and
\[
\mathcal{A} := \bigcup_{b \in N} S(M \setminus \{a\}, N \setminus \{b\}).
\]
Since the union in \( (3.3) \) is finite and disjoint, one has
\[
\#\mathcal{A} = \sum_{b \in N} \#S(M \setminus \{a\}, N \setminus \{b\}) \overset{\text{ind. hyp.}}{=} \sum_{b \in N} (n!) = (n + 1) \cdot n! = (n + 1)!. \tag{3.3}
\]
Thus, it suffices to show
\[
\phi : S \to \mathcal{A}, \quad \phi(f) : M \setminus \{a\} \to N \setminus \{f(a)\}, \quad \phi(f) := f|_{M \setminus \{a\}},
\]
is well-defined and bijective. If \( f : M \to N \) is bijective, then \( f|_{M \setminus \{a\}} : M \setminus \{a\} \to N \setminus \{f(a)\} \) is bijective as well, i.e. \( \phi \) is well-defined. Suppose \( f, g \in S \) with \( f \neq g \). If \( f(a) \neq g(a) \), then \( \phi(f) \neq \phi(g) \), as they have different ranges. If \( f(a) = g(a) \), then there exists \( x \in M \setminus \{a\} \) with \( f(x) \neq g(x) \), implying \( \phi(f)(x) = f(x) \neq g(x) = \phi(g)(x) \), i.e., once again, \( \phi(f) \neq \phi(g) \). Thus, \( \phi \) is injective. Now let \( h \in S(M \setminus \{a\}, N \setminus \{b\}) \) for some \( b \in N \). Letting
\[
f : M \to N, \quad f(x) := \begin{cases} b & \text{for } x = a, \\ h(x) & \text{for } x \neq a,
\end{cases}
\]
we have \( \phi(f) = h \), showing \( \phi \) to be surjective as well. 

**Theorem 3.7.** Let \( n \in \mathbb{N} \).

(a) Each permutation can be decomposed into finitely many disjoint cycles: For each \( \pi \in S_n \), there exists a decomposition of \( \{1, \ldots, n\} \) into disjoint sets \( A_1, \ldots, A_N \), \( N \in \mathbb{N} \), i.e.
\[
\{1, \ldots, n\} = \bigcup_{i=1}^{N} A_i \quad \text{and} \quad A_i \cap A_j = \emptyset \quad \text{for } i \neq j,
\]
\[
\text{where } A_i = \{ i_1, \ldots, i_k \} \quad \text{and} \quad i_1 < i_2 < \cdots < i_k.
\]
such that \( A_i \) consists of the distinct elements \( a_{i1}, \ldots, a_{i,N_i} \), and

\[
\pi = (a_{N_1} \ldots a_{N,N_N}) \cdots (a_{11} \ldots a_{1,N_1}).
\] (3.5)

The decomposition (3.5) is unique up to the order of the cycles.

(b) If \( n \geq 2 \), then every permutation \( \pi \in S_n \) is the composition of finitely many transpositions, where each transposition permutes two juxtaposed elements, i.e.

\[
\forall \pi \in S_n \exists \tau_1, \ldots, \tau_n \in T \pi = \tau_n \circ \cdots \circ \tau_1,
\] (3.6)

where \( T := \{(i, i+1) : i \in \{1, \ldots, n-1\}\} \).

Proof. (a): We prove the statement by induction on \( n \). For \( n = 1 \), there is nothing to prove. Let \( n > 1 \) and choose \( i \in \{1, \ldots, n\} \). We claim that

\[
\exists k \in \mathbb{N} \left( \pi^k(i) = i \land \forall l \in \{1, \ldots, k-1\} \pi^l(i) \neq i \right). 
\] (3.7)

Indeed, since \( \{1, \ldots, n\} \) is finite, there must be a smallest \( k \in \mathbb{N} \) such that \( \pi^k(i) \in A_1 := \{i, \pi(i), \ldots, \pi^{k-1}(i)\} \). Since \( \pi \) is bijective, it must be \( \pi^k(i) = i \) and \( (i, \pi(i), \ldots, \pi^{k-1}(i)) \) is a \( k \)-cycle. We are already done in case \( k = n \). If \( k < n \), then consider \( B := \{1, \ldots, n\} \setminus A_1 \). Then, again using the bijectivity of \( \pi \), \( \pi \mid_B \) is a permutation on \( B \) with \( 1 \leq \#B < n \). By induction, there are disjoint sets \( A_2, \ldots, A_N \) such that \( B = \bigcup_{j=2}^{N} A_j \), \( A_j \) consists of the distinct elements \( a_{j1}, \ldots, a_{j,N_j} \) and

\[
\pi \mid_B = (a_{N_1} \ldots a_{N,N_N}) \cdots (a_{21} \ldots a_{2,N_2}).
\]

Since \( \pi = (i, \pi(i), \ldots, \pi^{k-1}(i)) \circ \pi \mid_B \), this finishes the proof of (3.5). If there were another, different, decomposition of \( \pi \) into cycles, say, given by disjoint sets \( B_1, \ldots, B_M \), \( \{1, \ldots, n\} = \bigcup_{l=1}^{M} B_l \), \( M \in \mathbb{N} \), then there were \( A_i \neq B_j \) and \( k \in A_i \cap B_j \). But then \( k \) were in the cycle given by \( A_i \) and in the cycle given by \( B_j \), implying \( A_i = \{\pi^l(k) : l \in \mathbb{N}\} = B_j \), in contradiction to \( A_i \neq B_j \).

(b): We first show that every \( \pi \in S_n \) is a composition of finitely many transpositions (not necessarily transpositions from the set \( T \)): According to (a), it suffices to show that every cycle is a composition of finitely many transpositions. Since each 1-cycle is the identity, it is \( (i) = \text{Id} = (1 \ 2) (1 \ 2) \) for each \( i \in \{1, \ldots, n\} \). If \( (i_1 \ldots i_k) \) is a \( k \)-cycle, \( k \in \{2, \ldots, n\} \), then

\[
(i_1 \ldots i_k) = (i_1 i_2) (i_2 i_3) \cdots (i_{k-1} i_k).
\] (3.8)

Indeed,

\[
\forall i \in \{1, \ldots, n\} (i_1 i_2) (i_2 i_3) \cdots (i_{k-1} i_k)(i) = \begin{cases} i_1 & \text{for } i = i_k, \\ i_{l+1} & \text{for } i = i_l, \ l \in \{1, \ldots, k-1\}, \\ i & \text{for } i \notin \{i_1, \ldots, i_k\}, \end{cases}
\]
proving (3.8). To finish the proof of (b), we observe that every transposition is a composition of finitely many elements of $T$: If $i, j \in \{1, \ldots, n\}$, $i < j$, then
\[(i\ j) = (i\ i+1) \cdots (j-2\ j-1)(j-1\ j) \cdots (i+1\ i+2)(i\ i+1) : \quad (3.9)\]
Indeed,
\[
\forall_{k \in \{1, \ldots, n\}} (i\ i+1) \cdots (j-2\ j-1)(j-1\ j) \cdots (i+1\ i+2)(i\ i+1)(k) = \begin{cases} j & \text{for } k = i, \\ i & \text{for } k = j, \\ k & \text{for } i < k < j, \\ k & \text{for } k \not\in \{i, i+1, \ldots, j\}, \end{cases}
\]
proving (3.9). ■

Remark 3.8. Let $n \in \mathbb{N}$, $n \geq 2$. According to Th. 3.7(a), each $\pi \in S_n$ has a unique decomposition into $N \in \mathbb{N}$ cycles as in (3.5). If $\tau \in S_n$ is a transposition, then, as a consequence of Lem. 3.4(a),(b),(c), the corresponding cycle decomposition of $\pi \tau$ has precisely $N+1$ cycles (if Lem. 3.4(b) applies) or precisely $N-1$ cycles (if Lem. 3.4(c) applies).

Definition 3.9. Let $k \in \mathbb{Z}$. We call the integer $k$ even if, and only if, $k \equiv 0 \pmod{2}$ (cf. notation introduced in [Phi19, Ex. 4.27(a)]; we call $k$ odd if, and only if, $k \equiv 1 \pmod{2}$) (i.e. $k$ is even if, and only if, 2 is a divisor of $k$; $k$ is odd if, and only if, 2 is no divisor of $k$, cf. [Phi19, Def. D.2(a)]). The property of being even or odd is called the parity of the integer $k$.

Example 3.10. Let $\text{Id} \in S_3$ be the identity on $\{1, 2, 3\}$. Then
\[
\text{Id} = (1)(2)(3) = (1\ 2)(1\ 2) = (3\ 2)(3\ 2) = (1\ 2)(1\ 2)(3\ 2)(3\ 2), \\
(1\ 2\ 3) = (1\ 2)(2\ 3) = (2\ 3)(1\ 3) = (1\ 2)(1\ 3)(2\ 3)(1\ 2),
\]
illustrating that, for $n \geq 2$, one can write elements $\pi$ of $S_n$ as products of transpositions with varying numbers of factors. However, while the number $k \in \mathbb{N}_0$ of factors in such products representing $\pi$ is not unique, we will prove in the following Th. 3.11(a) that the parity of $k$ is uniquely determined by $\pi \in S_n$.

Theorem 3.11. Let $n \in \mathbb{N}$.

(a) Let $n \geq 2$, $\pi \in S_n$, $N \in \mathbb{N}$, $k \in \mathbb{N}_0$. If
\[
\pi = \prod_{i=1}^{k} h_i 
\]
with transpositions $h_1, \ldots, h_k \in S_n$ and $\pi$ is decomposed into $N$ cycles according to Th. 3.7(a), then
\[
k \equiv n - N \pmod{2}, \quad (3.11)
\]
i.e. the parity of $k$ is uniquely determined by $\pi$. 

(b) The map

\[ \text{sgn} : S_n \rightarrow \{-1, 1\}, \quad \text{sgn}(\pi) := \begin{cases} 1 & \text{for } n = 1, \\ (-1)^k (a) = (-1)^{n-N \text{ (mod 2)}} & \text{for } n \geq 2, \end{cases} \]

where, for \( n \geq 2 \), \( k = k(\pi) \) and \( N = N(\pi) \) are as in (a), constitutes a group epimorphism (here, \( \{-1, 1\} \) is considered as a multiplicative subgroup of \( \mathbb{R} \) (or \( \mathbb{Q} \)) — as we know all groups with two elements to be isomorphic, we have \( (\{-1, 1\}, \cdot) \cong (\mathbb{Z}_2, +) \)). One calls \( \text{sgn}(\pi) \) the sign, signature, or signum of the permutation \( \pi \).

Proof. (a): We conduct the proof via induction on \( k \): For \( k = 0 \), the product in (3.10) is empty, i.e.

\[ \pi = \text{Id} = (1) \cdots (n), \]

yielding \( N = n \), showing (3.11) to hold. If \( k = 1 \), then \( \pi = h_1 \) is a transposition and, thus, has \( N = n-1 \) cycles, showing (3.11) to hold once again. Now assume (3.11) for \( k \geq 1 \) by induction and consider \( \pi = \prod_{i=1}^{k+1} h_i = (\prod_{i=1}^{k} h_i) h_{k+1} \). Thus, \( \pi = \pi_k h_{k+1} \), where \( \pi_k := \prod_{i=1}^{k} h_i \). If \( \pi_k \) has \( N_k \) cycles, then, by induction,

\[ k \equiv n - N_k \text{ (mod 2)}. \quad (3.12) \]

Moreover, from Rem. 3.8 we know \( N = N_k + 1 \) or \( N = N_k - 1 \). In both cases, (3.12) implies \( k + 1 \equiv n - N \text{ (mod 2)} \), completing the induction.

(b): For \( n = 1 \), there is nothing to prove. For \( n \geq 2 \), we first note \( \text{sgn} \) to be well-defined, as the number of cycles \( N(\pi) \) is uniquely determined by \( \pi \in S_n \) (and each \( \pi \in S_n \) can be written as a product of transpositions by Th. 3.7(b)). Next, we note \( \text{sgn} \) to be surjective, since, for the identity, we can choose \( k = 0 \), i.e. \( \text{sgn}(\text{Id}) = (-1)^0 = 1 \), and, for each transposition \( \tau \in S_n \) (as \( n \geq 2 \), \( S_n \) contains at least the transposition \( \tau = (1 \ 2) \)), we can choose \( k = 1 \), i.e. \( \text{sgn}(\tau) = (-1)^1 = -1 \). To verify \( \text{sgn} \) to be a homomorphism, let \( \pi, \sigma \in S_n \). By Th. 3.7(b), there are transpositions \( \tau_1, \ldots, \tau_k, h_1, \ldots, h_{k+}\) \( \in S_n \) such that

\[ \pi = \prod_{i=1}^{k} \tau_i, \quad \sigma = \prod_{i=1}^{k} h_i \Rightarrow \pi \sigma = \left( \prod_{i=1}^{k} \tau_i \right) \left( \prod_{i=1}^{k} h_i \right), \]

implying

\[ \text{sgn}(\pi \sigma) = (-1)^{k_+ + k_{\pi}} = (-1)^{k_k} (-1)^{k_{\sigma}} = \text{sgn}(\pi) \text{ sgn}(\sigma), \]

thus, completing the proof that \( \text{sgn} \) constitutes a homomorphism. \hfill \blacksquare

Proposition 3.12. Let \( n \in \mathbb{N} \).

(a) One has

\[ \forall \pi \in S_n \quad \text{sgn}(\pi) = \prod_{1 \leq i < j \leq n} \frac{\pi(i) - \pi(j)}{i - j}. \quad (3.13) \]
(b) Let \( n \geq 2, \pi \in S_n \). As in Th. 3.7(a), we write \( \pi \) as a product of \( N \in \mathbb{N} \) cycles,

\[
\pi = (a_{N1} \ldots a_{N,N}) \cdots (a_{11} \ldots a_{1,N}) = \prod_{i=1}^{N} \gamma_i, \tag{3.14}
\]

where, for each \( i \in \{1, \ldots, N\} \), \( \gamma_i := (a_{i1} \ldots a_{i,N}) \) is a cycle of length \( N_i \in \mathbb{N} \).

Then

\[
\text{sgn}(\pi) = (-1)^{\sum_{i=1}^{N} (N_i - 1)}. \tag{3.15}
\]

Proof. (a): For each \( \pi \in S_n \), let \( \sigma(\pi) \) denote the value given by the right-hand side of (3.13). If \( n = 1 \), then \( S_n = \{\text{Id}\} \) and \( \sigma(\text{Id}) = 1 = \text{sgn}(\text{Id}) \), since the product in (3.13) is empty (and, thus, equal to 1). For \( n \geq 2 \), we first show

\[
\forall \pi_1, \pi_2 \in S_n \quad \sigma(\pi_1 \circ \pi_2) = \sigma(\pi_1)\sigma(\pi_2) : \tag{3.16}
\]

For each \( \pi_1, \pi_2 \in S_n \), one computes

\[
\sigma(\pi_1 \circ \pi_2) = \prod_{1 \leq i < j \leq n} \frac{\pi_1(\pi_2(i)) - \pi_1(\pi_2(j))}{i-j} = \prod_{1 \leq i < j \leq n} \left( \frac{\pi_1(\pi_2(i)) - \pi_1(\pi_2(j))}{\pi_2(i) - \pi_2(j)} \cdot \frac{\pi_2(i) - \pi_2(j)}{i-j} \right) \cdot \sigma(\pi_1)\sigma(\pi_2),
\]

thereby establishing the case. Next, if \( \tau \in S_n \) is a transposition, then there exist elements \( i, j \in \{1, \ldots, n\} \) such that \( i < j \) and \( \tau = (i \ j) \). Thus,

\[
\sigma(\tau) = \frac{\tau(i) - \tau(j)}{i-j} = \frac{j-i}{i-j} = -1
\]

holds for each transposition \( \tau \). In consequence, if \( \pi \in S_n \) is the composition of \( k \in \mathbb{N} \) transpositions, then

\[
\sigma(\pi) = (-1)^k \text{Th. 3.11(b)} = \text{sgn}(\pi),
\]

proving (a).

(b): Using (3.14) together with the homomorphism property of \( \text{sgn} \) given by Th. 3.11(b), if suffices to show that

\[
\text{sgn}(\gamma) = (-1)^{k-1} \tag{3.16}
\]

holds for each cycle \( \gamma := (i_1 \ldots i_k) \in S_n \), where \( i_1, \ldots, i_k \) are distinct elements of \( \{1, \ldots, n\} \), \( k \in \mathbb{N} \). According to (3.8), we have

\[
\gamma = (i_1 \ldots i_k) = (i_1, i_2)(i_2, i_3) \cdots (i_{k-1}, i_k),
\]

showing \( \gamma \) to be the product of \( k - 1 \) transpositions, thereby proving (3.16) and the proposition. \( \square \)
**Definition 3.13.** Let \( n \in \mathbb{N}, \ n \geq 2, \) and let \( sgn : S_n \to \{1, -1\} \) be the group homomorphism defined in Th. 3.11(b) above. We call \( \pi \in S_n \) **even** if, and only if, \( sgn(\pi) = 1; \) we call \( \pi \) **odd** if, and only if, \( sgn(\pi) = -1. \) The property of being even or odd is called the **parity** of the permutation \( \pi. \) Moreover, we call 
\[ A_n := \ker sgn = \{ \pi \in S_n : sgn(\pi) = 1 \} \]
the **alternating group** on \( \{1, \ldots, n\}. \)

**Proposition 3.14.** Let \( n \in \mathbb{N}, \ n \geq 2. \)

(a) \( A_n \) is a normal subgroup of \( S_n \) and one has
\[ S_n/A_n \cong \text{Im} \ sgn = \{ 1, -1 \} \cong \mathbb{Z}_2 \]
(3.17)
(where \( \{1, -1\} \) is considered with multiplication and \( \mathbb{Z}_2 = \{0, 1\} \) is considered with addition modulo 2).

(b) For each transposition \( \tau \in S_n, \) one has \( \tau \notin A_n, \) \( S_n = (A_n\tau) \cup A_n, \) where we recall that \( \cup \) denotes a disjoint union and \( A_n\tau \) denotes the coset \( \{ \pi\tau : \pi \in A_n \}. \) Moreover, 
\[ \#A_n = \#(A_n\tau) = (n!/2). \]

**Proof.** (a): As the kernel of a homomorphism, \( A_n \) is a normal subgroup by [Phi19, Ex. 4.24(a)] and, thus, (3.17) is immediate from the isomorphism theorem [Phi19, Th. 4.26(b)].

(b): Let \( \tau \in S_n \) be a transposition. Since \( sgn(\tau) = (-1)^1 = -1, \) \( \tau \notin A_n. \) Moreover, if \( \pi \in A_n, \) then \( sgn(\pi\tau) = sgn(\pi)sgn(\tau) = 1 \cdot (-1) = -1, \) showing \( (A_n\tau) \cap A_n = \emptyset. \) Let \( \pi \in S_n \setminus A_n. \) Then
\[ sgn(\pi) = -1 \Rightarrow sgn(\pi\tau) = 1 \Rightarrow \pi\tau \in A_n \Rightarrow \pi = \pi\tau \in A_n\tau, \]
showing \( S_n = (A_n\tau) \cup A_n. \) To prove \( \#A_n = \#(A_n\tau), \) we note the maps \( \phi : A_n \to A_n\tau, \phi(\pi) := \pi\tau, \phi^{-1} : A_n\tau \to A_n, \phi^{-1}(\pi) := \pi\tau, \) to be inverses of each other and, thus, bijective. Moreover, as we know \( \#S_n = n! \) from Prop. 3.6, we have \( n! = \#S_n = \#A_n + \#(A_n\tau) = 2 \cdot (\#A_n), \) thereby completing the proof. ■

# 4 Multilinear Maps and Determinants

Employing the preparations of the previous section, we are now in a position to formulate an ad hoc definition of the notion of determinant. However, it seems more instructive to first study some more general related notions that embed determinants into a context that is also of independent interest and importance. Determinants are actually rather versatile objects: We will see below that, given a finite-dimensional vector space \( V \) over a field \( F, \) they can be viewed as maps \( \det : \mathcal{L}(V, V) \to F, \) assigning numbers to linear endomorphisms. However, they can also be viewed as functions \( \det : \mathcal{M}(n, F) \to F, \)
assigning numbers to quadratic matrices, and also as polynomial functions \( \det : F^{n^2} \rightarrow F \) of degree \( n \) in \( n^2 \) variables. But the point of view, we will focus on first, is determinants as (alternating) multilinear forms (i.e. \( F \)-valued maps) \( \det : V^n \rightarrow F \).

We start with a general introduction to multilinear maps, which have many important applications, not only in Algebra, but also in Analysis, where they, e.g., occur in the form of higher order total derivatives of maps from \( \mathbb{R}^n \) to \( \mathbb{R}^m \) (cf. [Phi16b, Sec. 4.6]) and when studying the integration of so-called differential forms (see, e.g., [For17, §19] and [Kön04, Sec. 13.1]).

### 4.1 Multilinear Maps

**Definition 4.1.** Let \( V \) and \( W \) be vector spaces over the field \( F \), \( \alpha \in \mathbb{N} \). We call a map

\[
L : V^\alpha \rightarrow W
\]

**multilinear** (more precisely, \( \alpha \) times linear, **bilinear** for \( \alpha = 2 \)) if, and only if, it is linear in each component, i.e., for each \( x^1, \ldots, x^{i-1}, x^{i+1}, \ldots, x^{\alpha}, v, w \in V \), \( i \in \{1, \ldots, \alpha\} \) and each \( \lambda, \mu \in F \):

\[
L(x^1, \ldots, x^{i-1}, v^i \lambda + w^i \mu, x^{i+1}, \ldots, x^{\alpha}) = \lambda L(x^1, \ldots, x^{i-1}, v, x^{i+1}, \ldots, x^{\alpha}) + \mu L(x^1, \ldots, x^{i-1}, w, x^{i+1}, \ldots, x^{\alpha}),
\]

where we note that, here, and in the following, the superscripts merely denote upper indices and not exponentiation. We denote the set of all \( \alpha \) times linear maps from \( V^\alpha \) into \( W \) by \( \mathcal{L}^\alpha(V, W) \). We also set \( \mathcal{L}^0(V, W) := W \). In extension of Def. 2.2(a), we also call \( L \in \mathcal{L}^\alpha(V, F) \) a **multilinear form**, a **bilinear form** for \( \alpha = 2 \).

**Remark 4.2.** In the situation of Def. 4.1, each \( \mathcal{L}^\alpha(V, W) \), \( \alpha \in \mathbb{N}_0 \), constitutes a vector space over \( F \). It is a subspace of the vector space over \( F \) of all functions from \( V^\alpha \) into \( W \), since, clearly, if \( K, L : V^\alpha \rightarrow W \) are both \( \alpha \) times linear and \( \lambda, \mu \in F \), then \( \lambda K + \mu L \) is also \( \alpha \) times linear.

The following Th. 4.3 is in generalization of [Phi19, Th. 6.6].

**Theorem 4.3.** Let \( V \) and \( W \) be vector spaces over the field \( F \), \( \alpha \in \mathbb{N} \). Moreover, let \( B \) be a basis of \( V \). Then each \( \alpha \) times linear map \( L \in \mathcal{L}^\alpha(V, W) \) is uniquely determined by its values on \( B^\alpha \): More precisely, if \( (w_b)_{b \in B^\alpha} \) is a family in \( W \), and, for each \( v \in V \), \( B_v \) and \( c_v : B_v \rightarrow F \setminus \{0\} \) are as in [Phi19, Th. 5.19] (providing the coordinates of \( v \) with respect to \( B \) in the usual way), then the map

\[
L : V^\alpha \rightarrow W, \quad L(v^1, \ldots, v^\alpha) = L \left( \sum_{b^1 \in B_v} c_{v^1}(b^1) b^1, \ldots, \sum_{b^\alpha \in B_v} c_{v^\alpha}(b^\alpha) b^\alpha \right) := \sum_{(b^1, \ldots, b^\alpha) \in B_v \times \cdots \times B_v} c_{v^1}(b^1) \cdots c_{v^\alpha}(b^\alpha) w_{(b^1, \ldots, b^\alpha)},
\]

(4.3)
is $\alpha$ times linear, and $\tilde{L} \in \mathcal{L}^\alpha(V, W)$ with
\[ \forall b \in B^\alpha \quad \tilde{L}(b) = w_b, \] (4.4)
implies $L = \tilde{L}$.

Proof. Exercise: Apart from the more elaborate notation, everything works as in the proof of [Phi19, Th. 6.6].

The following Th. 4.4 is in generalization of [Phi19, Th. 6.19].

**Theorem 4.4.** Let $V$ and $W$ be vector spaces over the field $F$, let $B_V$ and $B_W$ be bases of $V$ and $W$, respectively, let $\alpha \in \mathbb{N}$. Given $b_1, \ldots, b_\alpha \in B_V$ and $b \in B_W$, and using Th. 4.3, define maps $L_{b_1, \ldots, b_\alpha, b} \in \mathcal{L}^\alpha(V, W)$ by letting
\[ L_{b_1, \ldots, b_\alpha, b}(\tilde{b}_1, \ldots, \tilde{b}_\alpha) := \begin{cases} b & \text{for } (\tilde{b}_1, \ldots, \tilde{b}_\alpha) = (b_1, \ldots, b_\alpha), \\ 0 & \text{otherwise}. \end{cases} \] (4.5)
Let $\mathcal{B} := \{ L_{b_1, \ldots, b_\alpha, b} : (b_1, \ldots, b_\alpha) \in (B_V)^\alpha, b \in B_W \}$.

(a) $\mathcal{B}$ is linearly independent.

(b) If $V$ is finite-dimensional, $\dim V = n \in \mathbb{N}$, $B_V = \{b_1, \ldots, b_\alpha\}$, then $\mathcal{B}$ constitutes a basis for $\mathcal{L}^\alpha(V, W)$. If, in addition, $\dim W = m \in \mathbb{N}$, $B_W = \{w_1, \ldots, w_m\}$, then we can write
\[ \dim \mathcal{L}^\alpha(V, W) = (\dim V)^\alpha \cdot \dim W = n^\alpha \cdot m. \] (4.6)

(c) If $\dim V = \infty$ and $\dim W \geq 1$, then $\langle \mathcal{B} \rangle \subsetneq \mathcal{L}^\alpha(V, W)$ and, in particular, $\mathcal{B}$ is not a basis of $\mathcal{L}^\alpha(V, W)$.

Proof. (a): We verify that the elements of $\mathcal{B}$ are linearly independent: Let $M, N \in \mathbb{N}$. Let $(b_1^1, \ldots, b_\alpha^1), \ldots, (b_1^N, \ldots, b_\alpha^N) \in (B_V)^\alpha$ be distinct and let $w^1, \ldots, w^M \in B_W$ be distinct as well. Assume $\lambda_{lk} \in F$ to be such that
\[ L := \sum_{l=1}^M \sum_{k=1}^N \lambda_{lk} L_{b_1^l, \ldots, b_\alpha^l, w^l} = 0. \]
Let $\tilde{k} \in \{1, \ldots, N\}$. Then
\[ 0 = L(b_1^{\tilde{k}}, \ldots, b_\alpha^{\tilde{k}}) = \sum_{l=1}^M \sum_{k=1}^N \lambda_{lk} L_{b_1^l, \ldots, b_\alpha^l, w^l}(b_1^{\tilde{k}}, \ldots, b_\alpha^{\tilde{k}}) = \sum_{l=1}^M \lambda_{l\tilde{k}} w^l \]
implies $\lambda_{l\tilde{k}} = \cdots = \lambda_{M\tilde{k}} = 0$ due to the linear independence of the $w^l \in B_W$. As this holds for each $\tilde{k} \in \{1, \ldots, N\}$, we have established the linear independence of $\mathcal{B}$.

(b): According to (a), it remains to show $\langle \mathcal{B} \rangle = \mathcal{L}^\alpha(V, W)$. Let $L \in \mathcal{L}^\alpha(V, W)$ and $(i_1, \ldots, i_\alpha) \in \{1, \ldots, n\}^\alpha$. Then there exists a finite set $B_{(i_1, \ldots, i_\alpha)} \subseteq B_W$ such
that \( L(b^{i_1}, \ldots, b^{i_\alpha}) = \sum_{w \in B(i_1, \ldots, i_\alpha)} \lambda_w w \) with \( \lambda_w \in F \). Now let \( w_1, \ldots, w_M, M \in \mathbb{N} \), be an enumeration of the finite set \( \bigcup_{t \in \{1, \ldots, n\}^\alpha} B_I \). Then there exist \( \lambda_j(i_1, \ldots, i_\alpha) \in F \), \( (j, (i_1, \ldots, i_\alpha)) \in \{1, \ldots, M\} \times \{1, \ldots, n\}^\alpha \), such that

\[
\forall (i_1, \ldots, i_\alpha) \in \{1, \ldots, n\}^\alpha \quad L(b^{i_1}, \ldots, b^{i_\alpha}) = \sum_{j=1}^M \lambda_j(i_1, \ldots, i_\alpha) w_j.
\]

Letting \( \tilde{L} := \sum_{j=1}^M \sum_{(i_1, \ldots, i_\alpha) \in \{1, \ldots, n\}^\alpha} \lambda_j(i_1, \ldots, i_\alpha) L(b^{i_1}, \ldots, b^{i_\alpha}, w_j) \), we claim \( \tilde{L} = L \). Indeed,

\[
\forall (j_1, \ldots, j_\alpha) \in \{1, \ldots, n\}^\alpha \quad \tilde{L}(b^{j_1}, \ldots, b^{j_\alpha}) = \sum_{j=1}^M \sum_{(i_1, \ldots, i_\alpha) \in \{1, \ldots, n\}^\alpha} \lambda_j(i_1, \ldots, i_\alpha) L(b^{j_1}, \ldots, b^{j_\alpha}, w_j)
\]

\[
= \sum_{j=1}^M \sum_{(i_1, \ldots, i_\alpha) \in \{1, \ldots, n\}^\alpha} \lambda_j(i_1, \ldots, i_\alpha) \delta(i_1, \ldots, i_\alpha, (j_1, \ldots, j_\alpha)) w_j
\]

\[
= \sum_{j=1}^M \lambda_j(j_1, \ldots, j_\alpha) w_j = L(b^{j_1}, \ldots, b^{j_\alpha}),
\]

proving \( \tilde{L} = L \) by Th. 4.3. Since \( L \in \langle B \rangle \), the proof of (b) is complete.

(c): As \( \dim W \geq 1 \), there exists \( w \in B_W \). If \( L \in \langle B \rangle \), then \( \{b \in (B_V)^\alpha : L(b) \neq 0\} \) is finite. Thus, if \( B_V \) is infinite, then the map \( L \in \mathcal{L}^\alpha(V, W) \) with \( L(b) := w \) for each \( b \in (B_V)^\alpha \) is not in \( \langle B \rangle \), proving (c).

\section*{4.2 Alternating Multilinear Maps and Determinants}

\textbf{Definition 4.5.} Let \( V \) and \( W \) be vector spaces over the field \( F \), \( \alpha \in \mathbb{N} \). Then \( A \in \mathcal{L}^\alpha(V, W) \) is called alternating if, and only if,

\[
\forall (v_1, \ldots, v_\alpha) \in V^\alpha \quad \exists_{i, j \in \{1, \ldots, \alpha\}, i \neq j} v^i = v^j \Rightarrow A(v^1, \ldots, v_\alpha) = 0. \tag{4.7}
\]

Moreover, define the sets

\[
\text{Alt}^0(V, W) := W, \quad \text{Alt}^\alpha(V, W) := \{A \in \mathcal{L}^\alpha(V, W) : A \text{ alternating}\}
\]

(note that this immediately yields \( \text{Alt}^1(V, W) = \mathcal{L}(V, W) \)).

\textbf{Remark 4.6.} Let \( V \) and \( W \) be vector spaces over the field \( F \), \( \alpha \in \mathbb{N} \). Then \( \text{Alt}^\alpha(V, W) \) is a vector subspace of \( \mathcal{L}^\alpha(V, W) \): Indeed, \( 0 \in \text{Alt}^\alpha(V, W) \) and, if \( \lambda \in F \) and \( A, B \in \mathcal{L}^\alpha(V, W) \) satisfy (4.7), then \( \lambda A \) and \( A + B \) satisfy (4.7) as well.

\textbf{Notation 4.7.} Let \( V, W \) be a sets and \( \alpha \in \mathbb{N} \). Then, for each \( f \in \mathcal{F}(V^\alpha, W) = W^{V^\alpha} \) and each permutation \( \pi \in S_\alpha \), define

\[
(\pi f) : V^\alpha \to W, \quad (\pi f)(v_1, \ldots, v_\alpha) := f(v_{\pi(1)}, \ldots, v_{\pi(\alpha)}).
\]
Lemma 4.8. Let \( V, W \) be a sets and \( \alpha \in \mathbb{N} \). Then
\[
\forall \pi_1, \pi_2 \in S_{\alpha}, \quad \forall f \in \mathcal{F}(V^{\alpha}, W) \quad (\pi_1 \pi_2) f = \pi_1 (\pi_2 f). \tag{4.8}
\]

Proof. For each \( (v_1, \ldots, v_\alpha) \in V^\alpha \), let \( (w_1, \ldots, w_\alpha) := (v_{\pi_1(1)}, \ldots, v_{\pi_1(\alpha)}) \in V^\alpha \). Then, for each \( i \in \{1, \ldots, \alpha\} \), we have \( w_i = v_{\pi_2(i)} = v_{\pi_1(\pi_2(i))} \). Thus, we compute
\[
(\pi_1 \pi_2)(v_1, \ldots, v_\alpha) = f(v_{\pi_1(1)}, \ldots, v_{\pi_1(\alpha)}) = f(w_{\pi_2(1)}, \ldots, w_{\pi_2(\alpha)}) = (\pi_2 f)(w_1, \ldots, w_\alpha)
= (\pi_1 \pi_2)(v_1, \ldots, v_\alpha) = (\pi_1 f)(v_1, \ldots, v_\alpha),
\]
thereby establishing (4.8).

Proposition 4.9. Let \( V \) and \( W \) be vector spaces over the field \( F \), \( \alpha \in \mathbb{N} \). Then, given \( A \in \mathcal{L}^\alpha(V, W) \), the following statements are equivalent for char \( F \neq 2 \), where \( \text{"(i) } \Rightarrow \text{(ii)"} \) also holds for char \( F = 2 \).

(i) \( A \) is alternating.

(ii) For each permutation \( \pi \in S_{\alpha} \), one has
\[
\forall (v^1, \ldots, v^\alpha) \in V^\alpha \quad A(v^{\pi(1)}, \ldots, v^{\pi(\alpha)}) = \text{sgn}(\pi) A(v^1, \ldots, v^\alpha). \tag{4.9}
\]

Proof. \( \text{"(i) } \Rightarrow \text{(ii)"} \): We first prove (4.9) for transpositions \( \pi = (i \ i + 1) \) with \( i \in \{1, \ldots, \alpha - 1\} \). Let \( v^k \in V \), \( k \in \{1, \ldots, \alpha\} \setminus \{i, i + 1\} \), and define
\[
B : V \times V \to W, \quad B(v, w) := A(v^1, \ldots, v^{i-1}, v, v^{i+2}, \ldots, v^\alpha).
\]
Then, as \( A \) is \( \alpha \) times linear and alternating, \( B \) is bilinear and alternating. Thus, for each \( v, w \in V \),
\[
0 = B(v + w, v + w) = B(v, v) + B(v, w) + B(w, v) + B(w, w) = B(v, w) + B(w, v),
\]
implying \( B(v, w) = -B(w, v) \), proving (4.9) for this case. For general \( \pi \in S_{\alpha} \), (4.9) now follows from Th. 3.7(b), Th. 3.11(b), and Lem. 4.8: Let \( T := \{(i \ i + 1) \in S_{\alpha} : i \in \{1, \ldots, \alpha - 1\}\} \). Then, given \( \pi \in S_{\alpha} \), Th. 3.7(b) implies the existence of \( \pi_1, \ldots, \pi_N \in T \), \( N \in \mathbb{N} \), such that \( \pi = \pi_1 \cdots \pi_N \). Thus, for each \( (v^1, \ldots, v^\alpha) \in V^\alpha \),
\[
A(v^{\pi(1)}, \ldots, v^{\pi(\alpha)}) = (\pi A)(v^1, \ldots, v^\alpha) \overset{\text{Lem. 4.8}}{=} (\pi_1(\ldots (\pi_N A)(v^1, \ldots, v^\alpha)) \overset{\text{Th. 3.11(b)}}{=} = (-1)^N A(v^1, \ldots, v^\alpha),
\]
proving (4.9).

\( \text{"(ii) } \Rightarrow \text{(i)"} \): Let char \( F \neq 2 \). Let \( (v^1, \ldots, v^\alpha) \in V^\alpha \) and suppose \( i, j \in \{1, \ldots, \alpha\} \) are such that \( i \neq j \) as well as \( v^i = v^j \). Then
\[
A(v^1, \ldots, v^\alpha) \overset{(ii)}{=} \text{sgn}(ij) A(v^1, \ldots, v^\alpha) = -A(v^1, \ldots, v^\alpha).
\]
Thus, \( 2A(v^1, \ldots, v^\alpha) = 0 \) and \( 2 \neq 0 \) implies \( A(v^1, \ldots, v^\alpha) = 0 \).
The following Ex. 4.10 shows that “(i) ⇐ (ii)” can not be expected to hold in Prop. 4.9 for char \( F = 2 \):

**Example 4.10.** Let \( F := \mathbb{Z}_2 = \{0, 1\} \), \( V := F \). Consider the bilinear map \( A : V^2 \rightarrow F \), \( A(\lambda, \mu) := \lambda \mu \). Then \( A \) is not alternating, since \( A(1, 1) = 1 \cdot 1 = 1 \neq 0 \). However, (4.9) does hold for \( A \) (due to \( 1 = -1 \in F \)): \( A(1, 1) = 1 = -1 = -A(1, 1) \) (and \( 0 = A(0, 0) = -A(0, 0) = A(0, 1) = -A(1, 0) \)).

**Proposition 4.11.** Let \( V \) and \( W \) be vector spaces over the field \( F \), \( \alpha \in \mathbb{N} \). Then, given \( A \in \mathcal{L}^\alpha(V, W) \), the following statements are equivalent:

(i) \( A \) is alternating.

(ii) The implication in (4.7) holds whenever \( j = i + 1 \) (i.e. whenever two juxtaposed arguments are identical).

(iii) If the family \( (v^1, \ldots, v^\alpha) \) in \( V \) is linearly dependent, then \( A(v^1, \ldots, v^\alpha) = 0 \).

*Proof.* Exercise.

**Definition 4.12.** Let \( F \) be a field and \( \alpha := F^\alpha \), \( \alpha \in \mathbb{N} \). Then \( \det \in \text{Alt}^\alpha(V, F) \) is called a determinant if, and only if,

\[
\det(e_1, \ldots, e_\alpha) = 1,
\]

where \( e_1, \ldots, e_\alpha \) denote the standard basis vectors of \( V \).

**Definition 4.13.** Let \( V \) be a vector space over the field \( F \), \( \alpha \in \mathbb{N} \). We define the map \( \Lambda^\alpha : (V')^\alpha \rightarrow \text{Alt}^\alpha(V, F) \), called outer product or wedge product of linear forms, as follows:

\[
\Lambda^\alpha(\omega_1, \ldots, \omega_\alpha)(v^1, \ldots, v^\alpha) := \sum_{\pi \in S_\alpha} \text{sgn}(\pi) \omega_{\pi(1)}(v^1) \cdots \omega_{\pi(\alpha)}(v^\alpha) \tag{4.10}
\]

(cf. Th. 4.15 below). Given linear forms \( \omega_1, \ldots, \omega_\alpha \in V' \), it is common to also use the notation

\[
\omega_1 \wedge \cdots \wedge \omega_\alpha := \Lambda^\alpha(\omega_1, \ldots, \omega_\alpha).
\]

**Lemma 4.14.** Let \( V \) be a vector space over the field \( F \), \( \alpha \in \mathbb{N} \). Moreover, let \( \Lambda^\alpha \) denote the wedge product of Def. 4.13. Then one has

\[
\forall \omega_1, \ldots, \omega_\alpha \in V' \quad \forall v_1, \ldots, v^\alpha \in V \quad \Lambda^\alpha(\omega_1, \ldots, \omega_\alpha)(v^1, \ldots, v^\alpha) := \sum_{\pi \in S_\alpha} \text{sgn}(\pi) \omega_1(v_{\pi(1)}) \cdots \omega_\alpha(v_{\pi(\alpha)}). \tag{4.11}
\]

*Proof.* Using the commutativity of multiplication in \( F \), the bijectivity of permutations, as well as \( \text{sgn}(\pi) = \text{sgn}(\pi^{-1}) \), we obtain

\[
\Lambda^\alpha(\omega_1, \ldots, \omega_\alpha)(v^1, \ldots, v^\alpha) = \sum_{\pi \in S_\alpha} \text{sgn}(\pi) \omega_{\pi(1)}(v^1) \cdots \omega_{\pi(\alpha)}(v^\alpha) = \sum_{\pi \in S_\alpha} \text{sgn}(\pi) \omega_1(v_{\pi^{-1}(1)}) \cdots \omega_\alpha(v_{\pi^{-1}(\alpha)}) = \sum_{\pi \in S_\alpha} \text{sgn}(\pi) \omega_1(v^\pi(1)) \cdots \omega_\alpha(v^\pi(\alpha)),
\]
Theorem 4.15. Let $V$ be a vector space over the field $F$, $\alpha \in \mathbb{N}$. Moreover, let $\Lambda^\alpha$ denote the wedge product of Def. 4.13.

(a) $\Lambda^\alpha$ is well-defined, i.e. it does, indeed, map $(V')^\alpha$ into $\text{Alt}^\alpha(V, F)$.

(b) $\Lambda^\alpha \in \text{Alt}^\alpha(V', \text{Alt}^\alpha(V, F))$.

Proof. (a): Let $\omega_1, \ldots, \omega_\alpha \in V'$ and $A := \Lambda^\alpha(\omega_1, \ldots, \omega_\alpha)$. First, we show $A$ to be $\alpha$ times linear: To this end, let $\lambda, \mu \in F$ and $x, v^1, \ldots, v^\alpha \in V$. Then

$$A(v^1, \ldots, v^{i-1}, \lambda v^i + \mu x, v^{i+1}, \ldots, v^\alpha)$$

proving $A \in \mathcal{L}^\alpha(V, F)$. It remains to show $A$ is alternating. To this end, let $i, j \in \{1, \ldots, \alpha\}$ with $i < j$. Then, according to Prop. 3.14(b). $\tau := (i, j) \notin A_\alpha$ and $S_\alpha = (A_\alpha) \cup A_\alpha$. If $k \in \{1, \ldots, \alpha\} \setminus \{i, j\}$, then, for each $\pi \in S_\alpha$, $(\pi \tau)(k) = \pi(k)$. Thus, if $(v^1, \ldots, v^\alpha) \in V^\alpha$ is such that $v^i = v^j$, then

$$A(v^1, \ldots, v^\alpha) \overset{(4.11)}{=} \sum_{\pi \in S_\alpha} \text{sgn}(\pi) \omega_1(v^{\pi(1)}) \cdots \omega_\alpha(v^{\pi(\alpha)})$$

$$= \sum_{\pi \in A_\alpha} \left( \text{sgn}(\pi) \omega_1(v^{\pi(1)}) \cdots \omega_\alpha(v^{\pi(\alpha)}) + \text{sgn}(\pi \tau) \omega_1(v^{\pi(1)}) \cdots \omega_\alpha(v^{\pi(\alpha)}) \right)$$

$$= \sum_{\pi \in A_\alpha} \left( \text{sgn}(\pi) \omega_1(v^{\pi(1)}) \cdots \omega_\alpha(v^{\pi(\alpha)}) - \text{sgn}(\pi) \omega_1(v^{\pi(1)}) \cdots \omega_\alpha(v^{\pi(\alpha)}) \right) = 0,$$

showing $A$ to be alternating.

(b) follows from (a): According to (a), $\Lambda^\alpha$ maps $(V')^\alpha$ into $\text{Alt}^\alpha(V', F)$. Thus, if $\Phi : V \to V''$ is the canonical embedding, then, for each $v^1, \ldots, v^\alpha \in V^\alpha$ and each $\omega_1, \ldots, \omega_\alpha \in V'$,

$$\Lambda^\alpha(\Phi v^1, \ldots, \Phi v^\alpha)(\omega_1, \ldots, \omega_\alpha) = \sum_{\pi \in S_\alpha} \text{sgn}(\pi) (\Phi v^{\pi(1)})(\omega_1) \cdots (\Phi v^{\pi(\alpha)})(\omega_\alpha)$$

$$= \sum_{\pi \in S_\alpha} \text{sgn}(\pi) \omega_1(v^{\pi(1)}) \cdots \omega_\alpha(v^{\pi(\alpha)})$$

$$= \Lambda^\alpha(\omega_1, \ldots, \omega_\alpha)(v^1, \ldots, v^\alpha),$$

thereby completing the proof. 

$\blacksquare$
Theorem 4.17. Let $v^1, \ldots, v^\alpha \in V$ such that, for each $j \in \{1, \ldots, \alpha\}$, $v^j = \sum_{i=1}^\alpha a_{ji} e_i$ with $(a_{ji}) \in M(\alpha, F)$, let
\[
\det(v^1, \ldots, v^\alpha) := \sum_{\pi \in S_\alpha} \operatorname{sgn}(\pi) a_{1\pi(1)} \cdots a_{\alpha\pi(\alpha)}.
\] (4.12)

Then $\det$ is a determinant according to Def. 4.12. Moreover,
\[
\det(v^1, \ldots, v^\alpha) = \sum_{\pi \in S_\alpha} \operatorname{sgn}(\pi) a_{\pi(1)1} \cdots a_{\pi(\alpha)\alpha}
\] (4.13)
also holds.

Proof. For each $i \in \{1, \ldots, \alpha\}$, let $\omega_i := \pi e_i : V \rightarrow F$ be the projection onto the coordinate with respect to $e_i$ according to Ex. 2.1(a). Then, if $v^j$ is as above, we have $\omega_i(v^j) = a_{ji}$. Thus,
\[
\Lambda^\alpha(\omega_1, \ldots, \omega_\alpha)(v^1, \ldots, v^\alpha) = \sum_{\pi \in S_\alpha} \operatorname{sgn}(\pi) \omega_{\pi(1)}(v^1) \cdots \omega_{\pi(\alpha)}(v^\alpha)
\] (4.11)
\[
= \det(v^1, \ldots, v^\alpha),
\]
showing $\det \in \operatorname{Alt}^\alpha(V, F)$ by Th. 4.15(a). Then we also have
\[
\det(v^1, \ldots, v^\alpha) = \sum_{\pi \in S_\alpha} \operatorname{sgn}(\pi) a_{\pi(1)1} \cdots a_{\pi(\alpha)\alpha},
\]
proving (4.13). Moreover, for $(v^1, \ldots, v^\alpha) = (e_1, \ldots, e_\alpha)$, $(a_{ji}) = (\delta_{ji})$ holds. Thus,
\[
\det(e_1, \ldots, e_\alpha) = \sum_{\pi \in S_\alpha} \operatorname{sgn}(\pi) \delta_{1\pi(1)} \cdots \delta_{\alpha\pi(\alpha)} = \operatorname{sgn}(\operatorname{Id}) = 1,
\]
which completes the proof. ■

According to Th. 4.17(b) below, the map defined by (4.12) is the only determinant in $\operatorname{Alt}^\alpha(F^\alpha, F)$.

Theorem 4.17. Let $V$ and $W$ be vector spaces over the field $F$, let $B_V$ and $B_W$ be bases of $V$ and $W$, respectively, let $\alpha \in \mathbb{N}$. Moreover, as in Cor. 2.4, let $B' := \{\omega_b : b \in B_V\}$, where
\[
\forall_{(b, a) \in B'_V \times B'_V} \left( \omega_b \in V', \quad \omega_b(a) := \delta_{ba} \right).
\] (4.14)

In addition, assume $<$ to be a strict total order on $B_V$ (for $\dim V = \infty$, the order on $B_V$ exists due to the axiom of choice, cf. [Phi19, Th. A.52(iv)]) and define
\[
(B_V)^\alpha_{\text{ord}} := \{(b^1, \ldots, b^\alpha) \in (B_V)^\alpha : b^1 < \cdots < b^\alpha\}.
\]

Given $(b^1, \ldots, b^\alpha) \in (B_V)^\alpha_{\text{ord}}$ and $y \in B_W$, and using Th. 4.15(a), define maps
\[
A_{b^1, \ldots, b^\alpha, y}^\alpha : \operatorname{Alt}^\alpha(V, W), \quad A_{b^1, \ldots, b^\alpha, y}^\alpha := \Lambda^\alpha(\omega_{b^1}, \ldots, \omega_{b^\alpha}) y.
\] (4.15)

Let $B := \{A_{b^1, \ldots, b^\alpha, y}^\alpha : y \in B_W, (b^1, \ldots, b^\alpha) \in (B_V)^\alpha_{\text{ord}}\}$. 

(a) \( B \) is linearly independent.

(b) If \( V \) is finite-dimensional, \( \dim V = n \in \mathbb{N}, B_V = \{ b^1 < \cdots < b^n \} \), then \( B \) constitutes a basis for \( \text{Alt}^\alpha(V, W) \) (note \( B = \emptyset, \dim \text{Alt}^\alpha(V, W) = 0 \) for \( \alpha > n \)). If, in addition, \( \dim W = m \in \mathbb{N} \), then we can write

\[
\dim \text{Alt}^\alpha(V, W) = \begin{cases} 
\binom{n}{\alpha} \cdot m & \text{for } \alpha \leq n, \\
0 & \text{for } \alpha > n.
\end{cases}
\]

(4.16)

In particular, \( \dim \text{Alt}^\alpha(F^\alpha, F) = 1 \), showing that the map \( \text{det} \in \text{Alt}^\alpha(F^\alpha, F) \) is uniquely determined by Def. 4.12 and given by (4.12).

(c) If \( V \) is infinite-dimensional and \( \dim W \geq 1 \), then \( B \) is not a basis for \( \text{Alt}^\alpha(V, W) \).

Proof. (a): We verify that the elements of \( B \) are linearly independent: Note that 

\[
\forall \, (b^1, \ldots, b^\alpha) \in (B_V)^\alpha \text{ord} \quad \forall \, \text{id} \neq \pi \in S_\alpha \quad (b^{\pi(1)}, \ldots, b^{\pi(\alpha)}) \notin (B_V)^\alpha \text{ord}.
\]

Thus, if 

\[
(b^1, \ldots, b^\alpha), (c^1, \ldots, c^\alpha) \in (B_V)^\alpha \text{ord},
\]

then

\[
\Lambda^\alpha(\omega_{b^1}, \ldots, \omega_{b^\alpha})(c^1, \ldots, c^\alpha) = \sum_{\pi \in S_\alpha} \text{sgn}(\pi) \omega_{b^{\pi(1)}}(c^1) \cdots \omega_{b^{\pi(\alpha)}}(c^\alpha)
\]

\[
\overset{(4.17),(4.14)}{=} \begin{cases} 
1 & \text{for } (b^1, \ldots, b^\alpha) = (c^1, \ldots, c^\alpha), \\
0 & \text{otherwise}.
\end{cases}
\]

(4.18)

Now let \( M, N \in \mathbb{N} \), let \( (b^1_1, \ldots, b^\alpha_1), \ldots, (b^1_N, \ldots, b^\alpha_N) \in (B_V)^\alpha \text{ord} \) be distinct, and let \( y_1, \ldots, y_M \in B_W \) be distinct as well. Assume \( \lambda_{lk} \in F \) to be such that

\[
A := \sum_{l=1}^M \sum_{k=1}^N \lambda_{lk} A_{b^1_k, \ldots, b^\alpha_k, y_l} = 0.
\]

Let \( \bar{k} \in \{1, \ldots, N\} \). Then

\[
0 = A(b^1_{\bar{k}}, \ldots, b^\alpha_{\bar{k}}) = \sum_{l=1}^M \sum_{k=1}^N \lambda_{lk} \Lambda^\alpha(\omega_{b^1_k}, \ldots, \omega_{b^\alpha_k})(b^1_k, \ldots, b^\alpha_k) y_l \overset{(4.18)}{=} \sum_{l=1}^M \lambda_{lk} y_l
\]

implies \( \lambda_{1\bar{k}} = \cdots = \lambda_{M\bar{k}} = 0 \) due to the linear independence of the \( y_l \in B_W \). As this holds for each \( \bar{k} \in \{1, \ldots, N\} \), we have established the linear independence of \( B \).

(b): Note that

\[
\#(B_V)^\alpha \overset{(4.17)}{=} \# \{ I \subseteq B_V : \# I = \alpha \} \overset{\text{[Phil6a, Prop. 5.18(b)]}}{=} \binom{n}{\alpha}.
\]
According to (a), it remains to show \( \langle B \rangle = \text{Alt}^\alpha(V, W) \). Let \( A \in \text{Alt}^\alpha(V, W) \) and \((b^1, \ldots, b^\alpha) \in (B_V)^{\alpha}\). Then there exists a finite set \( B_{(i_1, \ldots, i_\alpha)} \subseteq B_W \) such that
\[
A(b^1, \ldots, b^\alpha) = \sum_{y \in B_{(i_1, \ldots, i_\alpha)}} \lambda_y y \text{ with } \lambda_y \in F.
\]
Now let \( y_1, \ldots, y_M, M \in \mathbb{N}, \) be an enumeration of the finite set
\[
\bigcup_{(i_1, \ldots, i_\alpha) \in \{1, \ldots, n\}^\alpha: (b^1, \ldots, b^\alpha) \in (B_V)^{\alpha}} B_{(i_1, \ldots, i_\alpha)}.
\]
Then, for each \((j, (b^1, \ldots, b^\alpha)) \in \{1, \ldots, M\} \times (B_V)^{\alpha}\), there exists \( \lambda_{j,(i_1, \ldots, i_\alpha)} \in F \), such that
\[
A(b^1, \ldots, b^\alpha) = \sum_{j=1}^M \lambda_{j,(i_1, \ldots, i_\alpha)} y_j.
\]
Letting \( \tilde{A} := \sum_{j=1}^M \sum_{(b^1, \ldots, b^\alpha) \in (B_V)^{\alpha}} \lambda_{j,(i_1, \ldots, i_\alpha)} A_{b^1, \ldots, b^\alpha, y_j} \), we claim \( \tilde{A} = A \). Indeed, for each \((b^1, \ldots, b^\alpha) \in (B_V)^{\alpha}\),
\[
\tilde{A}(b^1, \ldots, b^\alpha) = \sum_{j=1}^M \sum_{(b^1, \ldots, b^\alpha) \in (B_V)^{\alpha}} \lambda_{j,(i_1, \ldots, i_\alpha)} A_{\alpha} \left( \omega_{b^1}, \ldots, \omega_{b^\alpha} \right) (b^1, \ldots, b^\alpha) y_j.
\]
\[
= \sum_{j=1}^M \lambda_{j,(j_1, \ldots, j_\alpha)} y_j = A(b^1, \ldots, b^\alpha),
\]
proving \( \tilde{A} = A \) by (4.9) and Th. 4.3. Since \( \tilde{A} \in \langle B \rangle \), the proof of (b) is complete.

(c): Exercise. \( \blacksquare \)

The following Th. 4.18 compiles some additional rules of importance for alternating multilinear maps:

**Theorem 4.18.** Let \( V \) and \( W \) be vector spaces over the field \( F \), \( \alpha \in \mathbb{N} \). The following rules hold for each \( A \in \text{Alt}^\alpha(V, W) \):

(a) The value of \( A \) remains unchanged if one argument is replaced by the sum of that argument and a linear combination of the other arguments, i.e., if \( \lambda_1, \ldots, \lambda_\alpha \in F \), \( v^1, \ldots, v^\alpha \in V \), and \( i \in \{1, \ldots, \alpha\} \), then
\[
A(v^1, \ldots, v^\alpha) = A \left( v^1, \ldots, v^{i-1}, v^i + \sum_{\substack{j=1 \atop j \neq i}}^\alpha \lambda_j v^j, v^{i+1}, \ldots, v^\alpha \right).
\]

(b) If \( v^1, \ldots, v^\alpha \in V \) and \( a_{ji} \in F \) are such that
\[
\forall_{j \in \{1, \ldots, \alpha\}} w^j = \sum_{i=1}^\alpha a_{ji} v^i,
\]
then
\[
A(v^1, \ldots, v^\alpha) = A \left( w^1, \ldots, w^\alpha \right).
\]
then

\[ A(w^1, \ldots, w^\alpha) = \left( \sum_{\pi \in S_\alpha} \text{sgn}(\pi) a_{1\pi(1)} \cdots a_{\alpha\pi(\alpha)} \right) A(v^1, \ldots, v^\alpha) \]

\[ = \det(x^1, \ldots, x^\alpha) A(v^1, \ldots, v^\alpha), \quad (4.19) \]

where

\[ \forall \ j \in \{1, \ldots, \alpha\} \quad x^j := \sum_{i=1}^{\alpha} a_{ji} e_i \in F^\alpha \]

and \( \{e_1, \ldots, e_\alpha\} \) is the standard basis of \( F^\alpha \).

(c) Suppose the family \((v^1, \ldots, v^\alpha)\) in \( V \) is linearly independent and such that there exist \( w^1, \ldots, w^\alpha \in \{\{v^1, \ldots, v^\alpha\}\} \) with \( A(w^1, \ldots, w^\alpha) \neq 0 \). Then \( A(v^1, \ldots, v^\alpha) \neq 0 \) as well.

**Proof.** (a): One computes, for each \( i,j \in \{1, \ldots, \alpha\} \) with \( i \neq j \) and \( \lambda_j \neq 0 \):

\[ A(v^1, \ldots, v^i + \lambda_j v^j, \ldots, v^\alpha) \]

\[ \overset{\lambda_j^{-1} A(v^1, \ldots, v^i, \ldots, \lambda_j v^j, \ldots, v^\alpha)}{= \alpha} \]

\[ \overset{\lambda_j^{-1} A(v^1, \ldots, v^i, \lambda_j v^j, \ldots, v^\alpha) + \lambda_j^{-1} A(v^1, \ldots, \lambda_j v^j, \ldots, v^\alpha)}{=} (4.7) \]

\[ \overset{\lambda_j^{-1} A(v^1, \ldots, v^i, \lambda_j v^j, \ldots, v^\alpha) + 0}{=} A(v^1, \ldots, v^\alpha). \]

The general case of (a) then follows via a simple induction.

(b): We calculate

\[ A(w^1, \ldots, w^\alpha) \overset{\lambda_j^{-1} A(w^1, \ldots, w^i, \ldots, \lambda_j w^j, \ldots, w^\alpha)}{=} \left( \sum_{\pi \in S_\alpha} \text{sgn}(\pi) a_{1\pi(1)} \cdots a_{\alpha\pi(\alpha)} \right) A(v^1, \ldots, v^\alpha), \quad (4.9) \]

thereby proving the first equality in (4.19). The second equality in (4.19) is now an immediate consequence of Cor. 4.16.

(c): Exercise.  ■
4.3 Determinants of Matrices and Linear Maps

In Def. 4.12, we defined a determinant as an alternating multilinear form on $F^n$. In the following, we will see that determinants can be particularly useful when they are considered to be maps on quadratic matrices or maps on linear endomorphisms on vector spaces of finite dimension. We begin by defining determinants on quadratic matrices:

**Definition 4.19.** Let $F$ be a field and $n \in \mathbb{N}$. Then the map

$$\det : M(n, F) \longrightarrow F, \quad \det(a_{ji}) := \sum_{\pi \in S_n} \text{sgn}(\pi) a_{1\pi(1)} \cdots a_{n\pi(n)},$$

(4.20)

is called the determinant on $M(n, F)$.

**Notation 4.20.** If $F$ is a field, $n \in \mathbb{N}$, and $\det : M(n, F) \longrightarrow F$ is the determinant, then, for $(a_{ji}) \in M(n, F)$, one commonly uses the notation

$$\begin{vmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{nn} \end{vmatrix} := \det(a_{ji}).$$

(4.21)

**Notation 4.21.** Let $F$ be a field and $n \in \mathbb{N}$. As in [Phi19, Rem. 7.4(b)], we denote the columns and rows of a matrix $A := (a_{ji}) \in M(n, F)$ as follows:

$$c_i := c^A_i := \begin{pmatrix} a_{1i} \\ \vdots \\ a_{ni} \end{pmatrix}, \quad r_i := r^A_i := (a_{i1} \ldots a_{in}).$$

**Remark 4.22.** Let $F$ be a field and $n \in \mathbb{N}$. If we consider the rows $r_1, \ldots, r_n$ of the matrix $(a_{ji}) \in M(n, F)$ as elements of $F^n$, then, by Cor. 4.16,

$$\det(a_{ji}) = \det(r_1, \ldots, r_n),$$

where the second det is the map $\det : (F^n)^n \longrightarrow F$ defined as in Cor. 4.16. As a further consequence of Cor. 4.16, in combination with Th. 4.17(b), we see that det of Def. 4.19 is the unique form on $M(n, F)$ that is multilinear and alternating in the rows of the matrix and that assigns the value 1 to the identity matrix.

**Example 4.23.** Let $F$ be a field. We evaluate (4.20) explicitly for $n = 1, 2, 3$:

(a) For $(a_{11}) \in M(1, F)$, we have

$$|a_{11}| = \det(a_{11}) = \sum_{\pi \in S_1} \text{sgn}(\pi) a_{1\pi(1)} = \text{sgn}(\text{Id}) a_{11} = a_{11}.$$
(b) For \((a_{ji}) \in M(2, F)\), we have
\[
\begin{vmatrix}
  a_{11} & a_{12} \\
  a_{21} & a_{22}
\end{vmatrix} = \det(a_{ji}) = \sum_{\pi \in S_2} \text{sgn}(\pi) a_{1\pi(1)} a_{2\pi(2)} \\
= \text{sgn}(\text{Id}) a_{11} a_{22} + \text{sgn}(12) a_{12} a_{21} = a_{11} a_{22} - a_{12} a_{21},
\]
i.e. the determinant is the product of the elements of the main diagonal minus the product of the elements of the other diagonal:
\[
\begin{array}{cc}
  + & - \\
  a_{11} & a_{12} \\
  - & a_{21} & a_{22}
\end{array}
\]

(c) For \((a_{ji}) \in M(3, F)\), we have
\[
\begin{vmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23} \\
  a_{31} & a_{32} & a_{33}
\end{vmatrix} = \det(a_{ji}) = \sum_{\pi \in S_3} \text{sgn}(\pi) a_{1\pi(1)} a_{2\pi(2)} a_{3\pi(3)} \\
= \text{sgn}(\text{Id}) a_{11} a_{22} a_{33} + \text{sgn}(12) a_{12} a_{21} a_{33} + \text{sgn}(13) a_{13} a_{22} a_{31} \\
+ \text{sgn}(23) a_{11} a_{23} a_{32} + \text{sgn}(123) a_{12} a_{23} a_{31} + \text{sgn}(132) a_{13} a_{21} a_{32} \\
= a_{11} a_{22} a_{33} + a_{12} a_{23} a_{31} + a_{13} a_{21} a_{32} - (a_{12} a_{21} a_{33} + a_{13} a_{22} a_{31} + a_{11} a_{23} a_{32}).
\]
To remember this formula, one can use the following tableau:
\[
\begin{array}{ccccc}
  + & + & + & - & - & - \\
  a_{11} & a_{12} & a_{13} & a_{11} & a_{12} \\
  a_{21} & a_{22} & a_{23} & a_{21} & a_{22} \\
  a_{31} & a_{32} & a_{33} & a_{31} & a_{32} \\
  - & - & - & + & + & +
\end{array}
\]
One writes the first two columns of the matrix once again on the right and then takes the product of the first three diagonals in the direction of the main diagonal with a positive sign and the first three diagonals in the other direction with a negative sign.

We can now translate some of the results of Sec. 4.2 into results on determinants of matrices:

**Corollary 4.24.** Let \(F\) be a field, \(n \in \mathbb{N}\), and \(A := (a_{ji}) \in M(n, F)\). Let \(r_1, \ldots, r_n\) denote the rows of \(A\) and let \(c_1, \ldots, c_n\) denote the columns of \(A\). Then the following rules hold for the determinant:
(a) $\det(\text{Id}_n) = 1$, where $\text{Id}_n$ denotes the identity matrix in $\mathcal{M}(n, F)$.

(b) $\det(A^t) = \det(A)$.

(c) $\det$ is multilinear with regard to matrix rows as well as multilinear with regard to matrix columns, i.e., for each $v \in \mathcal{M}(1, n, F)$, $w \in \mathcal{M}(n, 1, F)$, $i \in \{1, \ldots, n\}$, and $\lambda, \mu \in F$:

$$
\det \begin{pmatrix}
    r_1 \\
    \vdots \\
    r_{i-1} \\
    \lambda r_i + \mu v \\
    r_{i+1} \\
    \vdots \\
    r_n
\end{pmatrix}
= \lambda \det(A) + \mu \det \begin{pmatrix}
    r_1 \\
    \vdots \\
    r_{i-1} \\
    v \\
    r_{i+1} \\
    \vdots \\
    r_n
\end{pmatrix},
$$

$$
\det (c_1 \ldots c_{i-1} \lambda c_i + \mu w \ c_{i+1} \ldots c_n)
= \lambda \det(A) + \mu \det (c_1 \ldots c_{i-1} w \ c_{i+1} \ldots c_n).
$$

(d) If $\lambda \in F$, then $\det(\lambda A) = \lambda^n \det(A)$.

(e) For each permutation $\pi \in S_n$, one has

$$
\det \begin{pmatrix}
    r_{\pi(1)} \\
    \vdots \\
    r_{\pi(n)}
\end{pmatrix}
= \det (c_{\pi(1)} \ldots c_{\pi(n)}) = \text{sgn}(\pi) \det(A).
$$

In particular, switching rows $i$ and $j$ or columns $i$ and $j$, where $i, j \in \{1, \ldots, n\}$, $i \neq j$, changes the sign of the determinant, i.e.

$$
\det \begin{pmatrix}
    r_1 \\
    \vdots \\
    r_i \\
    \vdots \\
    r_j \\
    \vdots \\
    r_n
\end{pmatrix}
= -\det \begin{pmatrix}
    r_1 \\
    \vdots \\
    r_j \\
    \vdots \\
    r_i \\
    \vdots \\
    r_n
\end{pmatrix},
$$

$$
\det (c_1 \ldots c_i \ldots c_j \ldots c_n) = -\det (c_1 \ldots c_j \ldots c_i \ldots c_n).
$$

(f) The following statements are equivalent:

(i) $\det A = 0$

(ii) The rows of $A$ are linearly dependent.

(iii) The columns of $A$ are linearly dependent.
Multiplication Rule: If $B := (b_{ji}) \in \mathcal{M}(n, F)$, then $\det(AB) = \det(A) \det(B)$.

$\det(A) = 0$ if, and only if, $A$ is singular. If $A$ is invertible, then $\det(A^{-1}) = (\det(A))^{-1}$.

The value of the determinant remains the same if one row of a matrix is replaced by the sum of that row and a scalar multiple of another row. More generally, the determinant remains the same if one row of a matrix is replaced by the sum of that row and a linear combination of the other rows. The statement also remains true if the word “row” is replaced by “column”. Thus, if $\lambda_1, \ldots, \lambda_n \in F$ and $i \in \{1, \ldots, n\}$, then

$$\det(A) = \det(\begin{pmatrix} r_1 \\ \vdots \\ r_i \\ \vdots \\ r_n \end{pmatrix}) = \det \left( \begin{pmatrix} r_1 \\ \vdots \\ r_i + \sum_{j \neq i}^{n} \lambda_j r_j \\ \vdots \\ r_{i+1} \\ \vdots \\ r_n \end{pmatrix} \right),$$

$$\det(A) = \det(c_1, \ldots, c_n) = \det \left( \begin{pmatrix} c_1, \ldots, c_{i-1}, c_i + \sum_{j=1}^{n} \lambda_j c_j, c_{i+1}, \ldots, c_n \end{pmatrix} \right).$$

Proof. As already observed in Rem. 4.22, $\det : \mathcal{M}(n, F) \to F$ can be viewed as the unique map $\det \in \text{Alt}^n(F^n, F)$ with $\det(e_1, \ldots, e_n) = 1$, where one considers

$$A = \begin{pmatrix} r_1 \\ \vdots \\ r_n \end{pmatrix} \in (F^n)^n :$$

Since, for each $j \in \{1, \ldots, n\}$,

$$r_j = (a_{j1} \ldots a_{jn}) = \sum_{i=1}^{n} a_{ji} e_i,$$

$$c_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix} = \sum_{i=1}^{n} a_{ij} e_i,$$

we have

$$\det(A) \overset{(4.20)}{=} \sum_{\pi \in S_n} \text{sgn}(\pi) a_{1\pi(1)} \cdots a_{n\pi(n)} \overset{(4.12)}{=} \det(r_1, \ldots, r_n)$$

$$\overset{(4.13)}{=} \sum_{\pi \in S_n} \text{sgn}(\pi) a_{\pi(1)1} \cdots a_{\pi(n)n} \overset{(4.12)}{=} \det(c_1, \ldots, c_n). \quad (4.22)$$
(a): \[ \det(\text{Id}_n) = \det(e_1, \ldots, e_n) = 1. \]

(b) is immediate from (4.22).

(c) is due to \( \det \in \text{Alt}^n(F^n, F) \) and (4.22).

(d) is a consequence of (c).

(e) is due to \( \det \in \text{Alt}^n(F^n, F) \), (4.22), and (4.9).

(f): Again, we use \( \det \in \text{Alt}^n(F^n, F) \) and (4.22): If \( \det(A) = 0 \), then \( \det(\text{Id}_n) = 1 \) and Th. 4.18(c) imply (ii) and (iii). Conversely, if (ii) or (iii) holds, then \( \det(A) = 0 \) follows from Prop. 4.11(iii).

(g): Let \( C := (c_{ji}) := AB \). Using Not. 4.21, we have

\[ \forall j \in \{1, \ldots, n\} \quad \left( r_j^C = r_j^A B = \sum_{i=1}^n a_{ji} r_i^B, \quad r_j^A = \sum_{i=1}^n a_{ji} e_i \right) \]

and, thus,

\[ \det(AB) = \det(r_1^C, \ldots, r_n^C) = \det(r_1^A, \ldots, r_n^A) \det(r_1^B, \ldots, r_n^B) = \det(A) \det(B). \]

(h): As a consequence of [Phi19, Th. 7.17(a)], \( A \) is singular if, and only if, the columns of \( A \) are linearly dependent. Thus, \( \det(A) = 0 \) if, and only if, \( A \) is singular due to (f). Moreover, if \( A \) is invertible, then

\[ 1 = \det(\text{Id}_n) = \det(AA^{-1}) \quad \text{(g)} \quad \det(A) \det(A^{-1}). \]

(i) is due to \( \det \in \text{Alt}^n(F^n, F) \), (4.22), and Th. 4.18(a). \( \blacksquare \)

**Theorem 4.25 (Block Matrices).** Let \( F \) be a field. The determinant of so-called block matrices over \( F \), where one block is a zero matrix (all entries 0), can be computed as the product of the determinants of the corresponding blocks. More precisely, if \( n, m \in \mathbb{N} \), then

\[
\begin{vmatrix}
a_{11} & \ldots & a_{1n} \\
\vdots & \ddots & \vdots \\
a_{n1} & \ldots & a_{nm} \\
0 & \ldots & 0 & b_{11} & \ldots & b_{1m} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & b_{m1} & \ldots & b_{mm}
\end{vmatrix} = \det(a_{ji}) \det(b_{ji}). \tag{4.23}
\]

**Proof.** Suppose \((a_{ji}) \in \mathcal{M}(n + m, F)\), where

\[
a_{ji} = \begin{cases} 
b_{i-n,j-n} & \text{for } i, j > n, \\
0 & \text{if } j > n \text{ and } i \leq n.
\end{cases} \tag{4.24}
\]

Via obvious embeddings, we can consider \( S_n \subseteq S_{m+n} \) and \( T_m := S_{\{n+1, \ldots, n+m\}} \subseteq S_{m+n} \). If \( \pi \in S_{m+n} \), then there are precisely two possibilities: Either there exist \( \omega \in S_n \) and
4 MULTILINEAR MAPS AND DETERMINANTS

\( \sigma \in T_m \) such that \( \pi = \omega \sigma \), or there exists \( j \in \{n+1, \ldots, m+n\} \) such that \( i := \pi(j) \leq n \), in which case \( a_{j \pi(j)} = 0 \). Thus,

\[
\det(a_{ji}) = \sum_{\pi \in S_{n+m}} \text{sgn}(\pi) a_{1\pi(1)} \cdots a_{n+m,\pi(n+m)}
\]

\[
= \sum_{(\omega, \sigma) \in S_n \times T_m} \text{sgn}(\omega) \text{sgn}(\sigma) a_{1\omega(1)} \cdots a_{n,\omega(n)} a_{n+1,\sigma(n+1)} \cdots a_{n+m,\sigma(n+m)}
\]

\[
= \left( \sum_{\omega \in S_n} \text{sgn}(\omega) a_{1\omega(1)} \cdots a_{n,\omega(n)} \right) \left( \sum_{\sigma \in T_m} \text{sgn}(\sigma) a_{n+1,\sigma(n+1)} \cdots a_{n+m,\sigma(n+m)} \right)
\]

\[
= \left( \sum_{\pi \in S_n} \text{sgn}(\pi) a_{1\pi(1)} \cdots a_{n,\pi(n)} \right) \left( \sum_{\pi \in S_m} \text{sgn}(\pi) b_{1\pi(1)} \cdots b_{m,\pi(m)} \right)
\]

\[
= \det(a_{ji}) \mid_{\{1, \ldots, n\} \times \{1, \ldots, n\}} \det(b_{ji}),
\]

proving (4.23). \( \blacksquare \)

**Corollary 4.26.** Let \( F \) be a field and \( n \in \mathbb{N} \). If \( (a_{ji}) \in M(n, F) \) is upper triangular or lower triangular, then

\[
\det(a_{ji}) = \prod_{k=1}^{n} a_{kk}.
\]

**Proof.** For \( (a_{ji}) \) upper triangular, the statement follows from Th. 4.25 via an obvious induction on \( n \in \mathbb{N} \). If \( (a_{ji}) \) is lower triangular, then the transpose \( (a_{ji})^t \) is upper triangular and the statement the follows from Cor. 4.24(b). \( \blacksquare \)

**Definition 4.27.** Let \( F \) be a field, \( n \in \mathbb{N}, n \geq 2 \), \( A = (a_{ji}) \in M(n, F) \). For each \( j, i \in \{1, \ldots, n\} \), let \( M_{ji} \in M(n-1, F) \) be the \( (n-1) \times (n-1) \) submatrix of \( A \) obtained by deleting the \( j \)th row and the \( i \)th column of \( A \) – the \( M_{ji} \) are called the minor matrices of \( A \); define

\[
A_{ji} := (-1)^{i+j} \det(M_{ji}),
\]

where the \( A_{ji} \) are called cofactors of \( A \) and the \( \det(M_{ji}) \) are called the minors of \( A \). Let \( A := (A_{ji})^t \) denote the transpose of the matrix of cofactors of \( A \), called the adjugate matrix of \( A \).

**Lemma 4.28.** Let \( F \) be a field, \( n \in \mathbb{N}, n \geq 2 \), \( A = (a_{ji}) \in M(n, F) \). For each \( j, i \in \{1, \ldots, n\} \), let \( R(j, i) \in M(n, F) \) be the matrix obtained from \( A \) by replacing the \( j \)-th row with the standard (row) basis vector \( e_i \), and let \( C(j, i) \in M(n, F) \) be the matrix obtained from \( A \) by replacing the \( i \)-th column with the standard (column) basis vector \( e_j \),
\[ R(j, i) := \begin{pmatrix} a_{11} & \ldots & a_{1,i-1} & a_{1i} & a_{1,i+1} & \ldots & a_{1n} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{j-1,1} & \ldots & a_{j-1,i-1} & a_{j-1,i} & a_{j-1,i+1} & \ldots & a_{j-1,n} \\ 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \\ a_{j+1,1} & \ldots & a_{j+1,i-1} & a_{j+1,i} & a_{j+1,i+1} & \ldots & a_{j+1,n} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{n1} & \ldots & a_{n,i-1} & a_{ni} & a_{n,i+1} & \ldots & a_{nn} \end{pmatrix}, \]

\[ C(j, i) := \begin{pmatrix} a_{11} & \ldots & a_{1,i-1} & 0 & a_{1,i+1} & \ldots & a_{1n} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{j-1,1} & \ldots & a_{j-1,i-1} & 0 & a_{j-1,i+1} & \ldots & a_{j-1,n} \\ a_{ij} & \ldots & a_{ij-1} & 1 & a_{ij+1} & \ldots & a_{jn} \\ a_{j+1,1} & \ldots & a_{j+1,i-1} & 0 & a_{j+1,i+1} & \ldots & a_{j+1,n} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{n1} & \ldots & a_{n,i-1} & 0 & a_{ni+1} & \ldots & a_{nn} \end{pmatrix}. \]

Then, we have
\[ \forall_{j, i \in \{1, \ldots, n\}} A_{ji} = \det(R(j, i)) = \det(C(j, i)). \tag{4.26} \]

**Proof.** Let \( j, i \in \{1, \ldots, n\} \), and let \( M_{ji} \) denote the corresponding minor matrix of \( A \) according to Def. 4.27. Then
\[
\det(R(j, i)) \overset{\text{Cor. 4.24(c)}}{=} (-1)^{i+j} \begin{vmatrix} 1 & 0 \\ M_{ji} & * \end{vmatrix} = \det(C(j, i)) \overset{\text{Cor. 4.24(c)}}{=} (-1)^{i+j} \begin{vmatrix} 1 & * \\ 0 & M_{ji} \end{vmatrix} \]
\[
\overset{(4.23)}{=} (-1)^{i+j} \det(M_{ji}) \overset{(4.25)}{=} A_{ji},
\]
thereby proving (4.26).

**Theorem 4.29.** Let \( F \) be a field, \( n \in \mathbb{N}, n \geq 2, A = (a_{ji}) \in M(n, F) \). Moreover, let \( \tilde{A} := (A_{ji})^1 \) be the adjugate matrix of \( A \) according to Def. 4.27. Then the following holds:

(a) \( A\tilde{A} = \tilde{A}A = (\det A) \operatorname{Id}_n \).

(b) If \( \det A \neq 0 \), then \( \tilde{A} = (\det A)^{n-1} \).

(c) If \( \det A \neq 0 \), then \( A^{-1} = (\det A)^{-1} \tilde{A} \).

(d) Laplace Expansion by Rows: \( \det A = \sum_{i=1}^n a_{ji} A_{ji} \) (expansion with respect to the \( j \)th row).

(e) Laplace Expansion by Columns: \( \det A = \sum_{j=1}^n a_{ji} A_{ji} \) (expansion with respect to the \( i \)th column).
Proof. (a): Let $C := (c_{ji}) := AA$, $D := (d_{ji}) := \tilde{A}A$. Also let $R(j, i)$ and $C(j, i)$ be as in Lem. 4.28. Then, we compute, for each $j, i \in \{1, \ldots, n\}$,

\begin{equation}
\begin{aligned}
c_{ji} &= \sum_{k=1}^{n} a_{jk} A_{ki}^{t} \quad \text{(4.26)} \\
&= \sum_{k=1}^{n} a_{jk} \det\left(R(i, k)\right) \quad \text{Cor. 4.24(c)} \\
&= \det \begin{pmatrix}
r_{i}^{A} \\
\vdots \\
\end{pmatrix}
\begin{pmatrix}
r_{i-1}^{A} \\
\vdots \\
\end{pmatrix} \\
&= \det\left(\begin{pmatrix}
\det(A) & \text{for } i = j, \\
0 & \text{for } i \neq j,
\end{pmatrix}\right)
\end{aligned}
\end{equation}

for $i = j$, for $i \neq j$,

proving (a).

(b): We obtain

\[
\det(A) \det(\tilde{A}) = \det(\tilde{A}A) \quad \text{(a)} = \det\left(\left(\det(A) \text{ Id}_{n}\right)\right) \quad \text{Cor. 4.24(d)} = \left(\det(A)\right)^{n},
\]

which, for $\det(A) \neq 0$, implies $\det(\tilde{A}) = \left(\det(A)\right)^{n-1}$.

(c) is immediate from (a).

(d): From (a), we obtain

\[
\det(A) = \sum_{i=1}^{n} a_{ji} A_{ij}^{t} = \sum_{i=1}^{n} a_{ji} A_{ji},
\]

proving (d).

(e): From (a), we obtain

\[
\det(A) = \sum_{j=1}^{n} A_{ij}^{t} a_{ji} = \sum_{j=1}^{n} a_{ji} A_{ji},
\]

proving (e).

\[\blacksquare\]

Example 4.30. (a) We use Ex. 4.23(b) and Th. 4.29(d) to compute

\[
D_{1} := \begin{vmatrix} 
1 & 2 \\
3 & 4
\end{vmatrix}:
\]

From Ex. 4.23(b), we obtain

\[
D_{1} = 1 \cdot 4 - 2 \cdot 3 = -2,
\]
which we also obtain when expanding with respect to the first row according to Th. 4.29(d). Expanding with respect to the second row, we obtain
\[ D_1 = -3 \cdot 2 + 4 \cdot 1 = -2. \]

(b) We use Ex. 4.23(c) and Th. 4.29(e) to compute
\[ D_2 := \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} : \]

From Ex. 4.23(c), we obtain
\[ D_2 = 1 \cdot 5 \cdot 9 + 2 \cdot 6 \cdot 7 + 3 \cdot 4 \cdot 8 - 3 \cdot 5 \cdot 7 - 1 \cdot 6 \cdot 8 - 2 \cdot 4 \cdot 9 = 45 + 84 + 96 - 105 - 48 - 72 = 0. \]

Expanding with respect to the third column according to Th. 4.29(d), we obtain
\[ D_2 = 3 \cdot \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} - 6 \cdot \begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} + 9 \cdot \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} = 3 \cdot (-3) - 6 \cdot (-6) + 9 \cdot (-3) = -9 + 36 - 27 = 0. \]

**Theorem 4.31** (Cramer’s Rule). Let \( F \) be a field and \( A := (a_{ji}) \in M(n, F) \), \( n \in \mathbb{N} \), \( n \geq 2 \). If \( b_1, \ldots, b_n \in F \) and \( \det(A) \neq 0 \), then the linear system
\[ \forall \quad j \in \{1, \ldots, n\} \quad \sum_{k=1}^{n} a_{jk} x_k = b_j, \]
has a unique solution \( x \in F^n \), which is given by
\[ \forall \quad j \in \{1, \ldots, n\} \quad x_j = (\det A)^{-1} \sum_{k=1}^{n} A_{kj} b_k, \quad (4.27) \]
where \( A_{kj} \) denote the cofactors of \( A \) according to Def. 4.27.

**Proof.** In matrix form, the linear system reads \( Ax = b \), which, for \( \det(A) \neq 0 \), is equivalent to \( x = A^{-1}b \), where \( A^{-1} = (\det A)^{-1} \tilde{A} \) by Th. 4.29(c). Since \( \tilde{A} := (A_{ji})^t \), we have
\[ \forall \quad j \in \{1, \ldots, n\} \quad x_j = (\det A)^{-1} \sum_{k=1}^{n} A_{jk} b_k = (\det A)^{-1} \sum_{k=1}^{n} A_{kj} b_k, \]
proving (4.27). \( \blacksquare \)

**Definition 4.32.** Let \( n \in \mathbb{N} \). An element \( p = (p_1, \ldots, p_n) \in (\mathbb{N}_0)^n \) is called a multi-index; \( |p| := p_1 + \cdots + p_n \) is called the degree of the multi-index. Let \( R \) be a ring with unity. If \( x = (x_1, \ldots, x_n) \in R^n \) and \( p = (p_1, \ldots, p_n) \) is a multi-index, then we define
\[ x^p := x_1^{p_1} x_2^{p_2} \cdots x_n^{p_n}. \quad (4.28) \]

Each function from \( R^n \) into \( R \), \( x \mapsto x^p \), is called a monomial function (in \( n \) variables); the degree of \( p \) is called the degree of the monomial. A function \( P \) from \( R^n \) into \( R \) is
called a polynomial function (in $n$ variables) if, and only if, it is a linear combination of monomials, i.e. if, and only if $P$ has the form

$$P : \mathbb{R}^n \to \mathbb{R}, \quad P(x) = \sum_{|p| \leq k} a_p x^p, \quad k \in \mathbb{N}_0, \quad a_p \in \mathbb{R}. \quad (4.29)$$

The degree\(^3\) of $P$, denoted $\deg(P)$, is the largest number $d \leq k$ such that there is $p \in (\mathbb{N}_0)^n$ with $|p| = d$ and $a_p \neq 0$. If all $a_p = 0$, i.e. if $P \equiv 0$, then $P$ is the zero polynomial function and its degree is defined to be $-\infty$ (some authors use $-1$ instead). If $F$ is a field, then we also define a rational function as a quotient of two polynomials: If $P, Q : \mathbb{F}^n \to \mathbb{F}$ are polynomials, then

$$\left(\frac{P}{Q}\right) : \mathbb{F}^n \setminus Q^{-1}\{0\} \to \mathbb{F}, \quad \left(\frac{P}{Q}\right)(x) := \frac{P(x)}{Q(x)}; \quad (4.30)$$

is called a rational function (in $n$ variables).

**Remark 4.33.** Let $F$ be a field and $n \in \mathbb{N}$. Comparing Def. 4.19 with Def. 4.32, we observe that

$$\det : \mathbb{F}^{n^2} \to \mathbb{F}, \quad \det(a_{ji}) := \sum_{\pi \in S_n} \text{sgn}(\pi) a_{1\pi(1)} \cdots a_{n\pi(n)},$$

is a polynomial function of degree $n$ (and in $n^2$ variables). According to Th. 4.29(c) and (4.25), we have

$$\text{inv} : \text{GL}_n(F) \to \text{GL}_n(F), \quad \text{inv}(a_{ji}) := (a_{ji})^{-1} = \frac{((-1)^{i+j} \det(M_{ij}))}{\det(a_{ji})},$$

where the $M_{ji} \in \mathcal{M}(n-1, F)$ denote the minor matrices of $(a_{ji})$. Thus, for each $(k, l) \in \{1, \ldots, n\}^2$, the component function

$$\text{inv}_{kl} : \text{GL}_n(F) \to \mathbb{F}, \quad \text{inv}_{kl}(a_{ji}) = \frac{(-1)^{k+l} \det(M_{lk})}{\det(a_{ji})},$$

is a rational function $\text{inv}_{kl} = P_{kl}/\det$, where $P_{kl}$ is a polynomial of degree $n - 1$.

**Remark 4.34 (Computation of Determinants).** For $n \leq 3$, one may well use the formulas of Ex. 4.23 to compute determinants. However, the larger $n$ becomes, the less advisable is the use of formula (4.20) to compute $\det(a_{ji})$, as this requires $n \cdot n!$ multiplications (note that $n!$ grows faster for $n \to \infty$ than $n^k$ for each $k \in \mathbb{N}$). Sometimes Th. 4.29(d),(e) can help, if one can expand with respect to a row or column, where many entries are 0. However, for a generic large $n \times n$ matrix $A$, it is usually the best strategy\(^4\) to transform $A$ to triangular form (e.g. by using Gaussian elimination according to [Phi19, Alg. 8.17]) and then use Cor. 4.26 (the number of multiplications then

\(^3\)Caveat: In general, $\deg$, as defined here, is not a well-defined function of $P$, as it actually depends on the representation (4.29) of $P$, rather than on $P$ itself, cf. Rem. 7.14 below.

\(^4\)If $A$ has a special structure (e.g. many zeros) and/or the field $F$ has a special structure (e.g. $F \in \{\mathbb{R}, \mathbb{C}\}$), then there might well be more efficient methods to compute $\det(A)$, studied in the fields Numerical Analysis and Numerical Linear Algebra.
merely grows proportional to \( n^3 \): We know from Cor. 4.24(e),(i) that the operations of the Gaussian elimination algorithm do not change the determinant, except for sign changes, when switching rows. For the same reasons, one should use Gaussian elimination (or even more efficient algorithms adapted to special situations) rather than the neat-looking explicit formula of Cramer’s rule of Th. 4.31 to solve large linear systems and rather than Th. 4.29(c) to compute inverses of large matrices.

**Theorem 4.35 (Vandermonde Determinant).** Let \( F \) be a field and \( \lambda_0, \lambda_1, \ldots, \lambda_n \in F, \ n \in \mathbb{N} \). Moreover, let

\[
V := \begin{pmatrix}
1 & \lambda_0 & \ldots & \lambda_0^n \\
1 & \lambda_1 & \ldots & \lambda_1^n \\
\vdots & \vdots & & \vdots \\
1 & \lambda_n & \ldots & \lambda_n^n 
\end{pmatrix} \in M(n+1, F), \tag{4.31}
\]

which is known as the corresponding Vandermonde matrix. Then its determinant, the so-called Vandermonde determinant, is given by

\[
\det(V) = \prod_{k,l=0}^{n} (\lambda_k - \lambda_l). \tag{4.32}
\]

**Proof.** The proof can be conducted by induction with respect to \( n \): For \( n = 1 \), we have

\[
\det(V) = \begin{vmatrix}
1 & \lambda_0 \\
1 & \lambda_1
\end{vmatrix} = \lambda_1 - \lambda_0 = \prod_{k,l=0}^{1} (\lambda_k - \lambda_l),
\]

showing (4.32) holds for \( n = 1 \). Now let \( n > 1 \). Using Cor. 4.24(i), we add the \((-\lambda_0)\)-fold of the \( n \)th column to the \((n+1)\)st column, we obtain in the \((n+1)\)st column

\[
\begin{pmatrix}
0 \\
\lambda_1^n - \lambda_1^{n-1} \lambda_0 \\
\vdots \\
\lambda_n^n - \lambda_n^{n-1} \lambda_0
\end{pmatrix}.
\]

Next, one adds the \((-\lambda_0)\)-fold of the \((n+1)\)st column to the \( n \)th column, and, successively, the \((-\lambda_0)\)-fold of the \( m \)th column to the \((m+1)\)st column. One finishes, in the \( n \)th step, by adding the \((-\lambda_0)\)-fold of the first column to the second column, obtaining

\[
\det(V) = \begin{vmatrix}
1 & \lambda_0 & \ldots & \lambda_0^n \\
1 & \lambda_1 & \ldots & \lambda_1^n \\
\vdots & \vdots & & \vdots \\
1 & \lambda_n & \ldots & \lambda_n^n 
\end{vmatrix} = \begin{vmatrix}
1 & 0 & 0 & \ldots & 0 \\
1 & \lambda_1 - \lambda_0 & \lambda_1^2 - \lambda_1 \lambda_0 & \ldots & \lambda_1^n - \lambda_1^{n-1} \lambda_0 \\
\vdots & \vdots & \vdots & & \vdots \\
1 & \lambda_n - \lambda_0 & \lambda_n^2 - \lambda_n \lambda_0 & \ldots & \lambda_n^n - \lambda_n^{n-1} \lambda_0
\end{vmatrix}.
\]

Applying (4.23), then yields

\[
\det(V) = 1 \cdot \begin{vmatrix}
\lambda_1 - \lambda_0 & \lambda_1^2 - \lambda_1 \lambda_0 & \ldots & \lambda_1^n - \lambda_1^{n-1} \lambda_0 \\
\vdots & \vdots & & \vdots \\
\lambda_n - \lambda_0 & \lambda_n^2 - \lambda_n \lambda_0 & \ldots & \lambda_n^n - \lambda_n^{n-1} \lambda_0
\end{vmatrix}.
\]
We now use multilinearity to factor out, for each \( k \in \{1, \ldots, n\} \), \((\lambda_k - \lambda_0)\) from the \( k \)th row, arriving at

\[
\det(V) = \prod_{k=1}^{n} (\lambda_k - \lambda_0) \prod_{\substack{k,l=1 \atop k>l}}^{n} (\lambda_k - \lambda_l) = \prod_{\substack{k,l=0 \atop k>l}}^{n} (\lambda_k - \lambda_l),
\]

which is precisely the Vandermonde determinant of the \( n - 1 \) numbers \( \lambda_1, \ldots, \lambda_n \). Using the induction hypothesis, we obtain

\[
\det(V) = \prod_{k=1}^{n} (\lambda_k - \lambda_0) \prod_{\substack{k,l=1 \atop k>l}}^{n} (\lambda_k - \lambda_l)
\]

completing the induction proof of (4.32).

**Remark and Definition 4.36** (Determinant on Linear Endomorphisms). Let \( V \) be a finite-dimensional vector space over the field \( F \), \( n \in \mathbb{N} \), \( \dim V = n \), and \( A \in \mathcal{L}(V, V) \). Moreover, let \( B_1 = (v_1, \ldots, v_n) \) and \( B_2 = (w_1, \ldots, w_n) \) be ordered bases of \( V \). If \( (a_{ji}^{(1)}) \in \mathcal{M}(n, F) \) is the matrix corresponding to \( A \) with respect to \( B_1 \) and \( (a_{ji}^{(2)}) \in \mathcal{M}(n, F) \) is the matrix corresponding to \( A \) with respect to \( B_2 \), then we know from [Phi19, Th. 7.14] that there exists \( (c_{ji}) \in \text{GL}_n(F) \) such that

\[
(a_{ji}^{(2)}) = (c_{ji})^{-1}(a_{ji}^{(1)})(c_{ji})
\]

(namely \( (c_{ji}) \) such that, for each \( i \in \{1, \ldots, n\} \), \( w_i = \sum_{j=1}^{n} c_{ji} v_j \)). Thus,

\[
\det(a_{ji}^{(2)}) = (\det(c_{ji})^{-1}) \det(a_{ji}^{(1)}) \det(c_{ji}) = \det(a_{ji}^{(1)})
\]

and, in consequence,

\[
\det : \mathcal{L}(V, V) \rightarrow F, \quad \det(A) := \det(a_{ji}^{(1)})
\]

is well-defined.

**Corollary 4.37.** Let \( V \) be a finite-dimensional vector space over the field \( F \), \( n \in \mathbb{N} \), \( \dim V = n \).

(a) If \( A, B \in \mathcal{L}(V, V) \), then \( \det(AB) = \det(A) \det(B) \).

(b) If \( A \in \mathcal{L}(V, V) \), then \( \det(A) = 0 \) if, and only if, \( A \) is not bijective. If \( A \) is bijective, then \( \det(A^{-1}) = (\det(A))^{-1} \).

(c) If \( A \in \mathcal{L}(V, V) \) and \( \lambda \in F \), then \( \det(\lambda A) = \lambda^n \det(A) \).

**Proof.** Let \( B_V \) be an ordered basis of \( V \).

(a): Let \( (a_{ji}), (b_{ji}) \in \mathcal{M}(n, F) \) be the matrices corresponding to \( A, B \) with respect to \( B_V \). Then

\[
\det(AB) \overset{[\text{Phi19, Th. 7.10(a)}]}{=} \det ((a_{ji}) (b_{ji})) \overset{\text{Cor. 4.24(g)}}{=} \det(a_{ji}) \det(b_{ji}) = \det(A) \det(B).
\]
(b): Since \((a_{ji})\) (with \((a_{ji})\) as before) is singular if, and only if, \(A\) is not bijective, (b) is immediate from Cor. 4.24(h).

(c): If \((a_{ji})\) is as before, then
\[
\det(\lambda A) = \det(\lambda (a_{ji})) \overset{\text{Cor. 4.24(d)}}{=} \lambda^n \det(a_{ji}) = \lambda^n \det(A),
\]
thereby completing the proof. ■

5 Direct Sums and Projections

In [Phi19, Def. 5.10], we defined sums of arbitrary (finite or infinite) families of subspaces. In [Phi19, Def. 5.28], we defined the direct sum for two subspaces. We will now extend the notion of direct sum to arbitrary families of subspaces:

**Definition 5.1.** Let \(V\) be a vector space over the field \(F\), let \(I\) be an index set and let \((U_i)_{i \in I}\) be a family of subspaces of \(V\). We say that \(V\) is the direct sum of the family of subspaces \((U_i)_{i \in I}\) if, and only if, the following two conditions hold:

(i) \(V = \sum_{i \in I} U_i\).

(ii) For each finite \(J \subseteq I\) and each family \((u_j)_{j \in J}\) in \(V\) such that \(u_j \in U_j\) for each \(j \in J\), one has
\[
0 = \sum_{j \in J} u_j \implies \forall_{j \in J} u_j = 0.
\]

If \(V\) is the direct sum of the \(U_i\), then we write \(V = \bigoplus_{i \in I} U_i\).

**Proposition 5.2.** Let \(V\) be a vector space over the field \(F\), let \(I\) be an index set and let \((U_i)_{i \in I}\) be a family of subspaces of \(V\). Then the following statements are equivalent:

(i) \(V = \bigoplus_{i \in I} U_i\).

(ii) For each \(v \in V\), there exists a unique finite subset \(J_v\) of \(I\) and a unique map \(\sigma_v : J_v \to V \setminus \{0\}\), \(j \mapsto u_j(v) := \sigma_v(j)\), such that
\[
v = \sum_{j \in J_v} \sigma_v(j) = \sum_{j \in J_v} u_j(v) \quad \land \quad \forall_{j \in J_v} u_j(v) \in U_j.
\]

(iii) \(V = \sum_{i \in I} U_i\) and
\[
\forall_{j \in I} U_j \cap \sum_{i \in I \setminus \{j\}} U_i = \{0\}.
\]

(iv) \(V = \sum_{i \in I} U_i\) and, letting \(I' := \{i \in I : U_i \neq \{0\}\}\), each family \((u_i)_{i \in I'}\) in \(V\) with \(u_i \in U_i \setminus \{0\}\) for each \(i \in I'\) is linearly independent.
Proof. “(i) \Leftrightarrow (ii)”: According to the definition of \( V = \sum_{i \in I} U_i \), the existence of \( J_v \) and \( \sigma_v \) such that (5.1) holds is equivalent to \( V = \sum_{i \in I} U_i \). If (5.1) holds and \( I_v \subseteq I \) is finite such that \( v = \sum_{j \in I_v} \tau_v(j) \) with \( \tau_v : I_v \rightarrow V \setminus \{0\} \) and \( \tau_v(j) \in U_j \) for each \( j \in I_v \), then define \( \sigma_v(j) := 0 \) for each \( j \in I_v \setminus J_v \) and \( \tau_v(j) := 0 \) for each \( j \in J_v \setminus I_v \). Then

\[
0 = v - v = \sum_{j \in J_v \cup I_v} (\sigma_v(j) - \tau_v(j))
\]

and Def. 5.1(ii) implies \( \sigma_v(j) = \tau_v(j) \) for each \( j \in J_v \cup I_v \) as well as \( J_v = I_v \). Conversely, assume there exists \( J \subseteq I \) finite and \( u_j \in U_j \ (j \in J) \) such that

\[
0 = \sum_{j \in J} u_j \quad \exists \ u_{j_0} \in J \ u_{j_0} \neq 0.
\]

Then

\[
u_{j_0} = \sum_{j \in J \setminus \{j_0\}} (-u_j)
\]

shows (ii) does not hold (as \( v := u_{j_0} \) has two different representations).

“(i) \Rightarrow (iii)”: If (i) holds and \( v \in U_j \cap \sum_{i \in \{j\}^c} U_i \) for some \( j \in I \), then there exists a finite \( J \subseteq I \setminus \{j\} \) such that \( v = \sum_{i \in J} u_i \) with \( u_i \in U_i \) for each \( i \in J \). Since \( -v \in U_j \) and

\[
0 = -v + \sum_{i \in J} u_i
\]

Def. 5.1(ii) implies \( -v = 0 = v \), i.e. (iii).

“(iii) \Rightarrow (i)”: Let \( J \subseteq I \) be finite such that \( 0 = \sum_{i \in J} u_i \) with \( u_i \in U_i \) for each \( i \in J \). Then

\[
\forall j \in J \ u_j = - \sum_{i \in J \setminus \{j\}} u_j \in U_j \cap \sum_{i \in J \setminus \{j\}} U_i
\]

i.e. (iii) implies \( u_j = 0 \) and Def. 5.1(ii).

“(iv) \Rightarrow (i)”: If \( J \subseteq I' \) is finite and \( u_j \in U_j \setminus \{0\} \) for each \( j \in J \) with \( (u_j)_{j \in J} \) linearly independent, then \( \sum_{j \in J} u_j \neq 0 \), implying Def. 5.1(ii) via contraposition.

“(i) \Rightarrow (iv)”: Suppose \((u_i)_{i \in I'}\) is a family in \( V \) such that \( u_i \in U_i \setminus \{0\} \) for each \( i \in I' \). If \( J \subseteq I' \) is finite and \( 0 = \sum_{j \in J} \lambda_j u_j \) for some \( \lambda_j \in K \), then Def. 5.1(ii) implies \( \lambda_j u_j = 0 \) and, thus, \( \lambda_j = 0 \), for each \( j \in J \), yielding the linear independence of \((u_i)_{i \in I'}\). \( \blacksquare \)

**Proposition 5.3.** Let \( V \) be a vector space over the field \( F \).

(a) Let \( B \) be a basis of \( V \) with a decomposition \( B = \bigcup_{i \in I} B_i \). If, for each \( i \in I \), \( U_i := \langle B_i \rangle \), then \( V = \bigoplus_{i \in I} U_i \). In particular, \( V = \bigoplus_{b \in B \setminus \{b\}} \).

(b) If \((U_i)_{i \in I}\) is a family of subspaces of \( V \) such that \( B_i \) is a basis of \( U_i \) and \( V = \bigoplus_{i \in I} U_i \), then the \( B_i \) are pairwise disjoint and \( B := \bigcup_{i \in I} B_i \) forms a basis of \( V \).

**Proof.** Exercise.
Example 5.4. Consider the vector space $V := \mathbb{R}^2$ over $\mathbb{R}$ and let $U_1 := \langle \{(1, 0)\} \rangle$, $U_2 := \langle \{(0, 1)\} \rangle$, $U_3 := \langle \{(1, 1)\} \rangle$. Then $V = U_1 + U_2$ and $U_i \cap U_j = \{0\}$ for each $i, j \in \{1, 2, 3\}$ with $i \neq j$. In particular, the sum $V = U_1 + U_2 + U_3$ is not a direct sum, showing Prop. 5.2(iii) can, in general, not be replaced by the condition $U_i \cap U_j = \{0\}$ for each $i, j \in I$ with $i \neq j$.

Definition 5.5. Let $S$ be a set. Then $P : S \to S$ is called a projection if, and only if, $P^2 = P \circ P = P$.

Remark 5.6. For each set $S$, $\text{Id} : S \to S$ is a projection. Moreover, for each $x \in S$, the constant map $f_x : S \to S$, $f_x \equiv x$, is a projection. If $V := S$ is a vector space over a field $F$, then $f_x$ is linear, if and only if, $x = 0$. While this shows that not every projection on a vector space is linear, here, we are interested in linear projections. We will see in Th. 5.8 below that there is a close relationship between linear projections and direct sums.

Example 5.7. (a) Let $V$ be a vector space over the field $F$ and let $B$ be a basis of $V$. If $c_v : B_v \to F \setminus \{0\}$, $B_v \subseteq V$ finite, is the coordinate map for $v \in V$, then, clearly, for each $b \in B$,

$$P_b : V \to V, \quad P_b(v) := \begin{cases} c_v(b) b & \text{for } b \in B_v, \\ 0 & \text{otherwise,} \end{cases}$$

is a linear projection.

(b) Let $I$ be a nonempty set, can consider the vector space $V := F(I, F) = F^I$ over the field $F$. If $e_i : I \to F$, $e_i(j) := \delta_{ij}$,

then, clearly, for each $i \in I$,

$$P_i : V \to V, \quad P_i(f) := f(i) e_i,$$

is a linear projection.

(c) Let $n \in \mathbb{N}$ and let $V$ be the vector space over $\mathbb{R}$, consisting of all functions $f : \mathbb{R} \to \mathbb{R}$ such that the $n$-th derivative $f^{(n)}$ exists. Then, clearly,

$$P : V \to V, \quad P(f)(x) := \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k,$$

is a linear projection, where the image of $P$ consists of the subspace of all polynomial functions of degree at most $n$.

Theorem 5.8. Let $V$ be a vector space over the field $F$. 
(a) Let \((U_i)_{i \in I}\) be a family of subspaces of \(V\) such that \(V = \bigoplus_{i \in I} U_i\). According to Prop. 5.2(ii), for each \(v \in V\), there exists a unique finite subset \(J_v\) of \(I\) and a unique map \(\sigma_v : J_v \to V \setminus \{0\}\), \(j \mapsto u_j(v) := \sigma_v(j)\), such that
\[
v = \sum_{j \in J_v} \sigma_v(j) = \sum_{j \in J_v} u_j(v) \quad \land \quad \forall j \in J_v \quad u_j(v) \in U_j.
\]
Thus, we can define
\[
\forall j \in I \quad P_j : V \to V, \quad P_j(v) := \begin{cases} \sigma_v(j) & \text{for } j \in J_v, \\ 0 & \text{otherwise.} \end{cases}
\]
Then each \(P_j\) is a linear projection with \(\text{Im} P_j = U_j\) and \(\text{ker} P_j = \bigoplus_{i \in I \setminus \{j\}} U_i\). Moreover, if \(i \neq j\), then \(P_i P_j = 0\). Defining
\[
\left(\sum_{i \in I} P_i\right) : V \to V, \quad \left(\sum_{i \in I} P_i\right)(v) := \sum_{i \in J_v} P_i(v),
\]
we have \(\text{Id} = \sum_{i \in I} P_i\).

(b) Let \((P_i)_{i \in I}\) be a family of projections in \(\mathcal{L}(V, V)\) such that, for each \(v \in V\), the set \(J_v := \{i \in I : P_i(v) \neq 0\}\) is finite. If \(\text{Id} = \sum_{i \in I} P_i\) (where \(\sum_{i \in I} P_i\) is defined as in (5.4)) and \(P_i P_j = 0\) for each \(i, j \in \{1, \ldots, n\}\) with \(i \neq j\), then
\[
V = \bigoplus_{i \in I} \text{Im} P_i.
\]

(c) If \(P \in \mathcal{L}(V, V)\) is a projection, then \(V = \ker P \oplus \text{Im} P\), \(\text{Im} P = \ker(\text{Id} - P)\), \(\ker P = \text{Im}(\text{Id} - P)\).

Proof. (a): If \(v, w \in V\) and \(\lambda \in F\), then, extending \(\sigma_v\) and \(\sigma_w\) by 0 to \(J_v \cup J_w\), we have
\[
\forall j \in J_v \cup J_w \quad \left(\sigma_{v+w}(j) = \sigma_v(j) + \sigma_w(j) \in U_j \quad \land \quad \sigma_{\lambda v}(j) = \lambda \sigma_v(j) \in U_j\right),
\]
showing, for each \(i \in I\), the linearity of \(P_i\) as well as \(\text{Im} P_i = U_i\) and \(P_i^2 = P_i\). Clearly, if \(j \in I\), then \(\bigoplus_{i \in I \setminus \{j\}} U_i \subseteq \ker P_j\). On the other hand, if \(v \in \ker P_j\), then \(j \notin J_v\), showing \(v \in \bigoplus_{i \in I \setminus \{j\}} U_i\). If \(i \neq j\) and \(v \in V\), then \(P_j v \in U_j \subseteq \ker P_i\), showing \(P_i P_j = 0\). Finally, for each \(v \in V\), we compute
\[
\left(\sum_{i \in I} P_i\right)(v) = \sum_{i \in J_v} P_i(v) = \sum_{i \in J_v} \sigma_v(i) = v,
\]
thereby completing the proof of (a).

(b): For each \(v \in V\), we have
\[
v = \text{Id} v = \sum_{i \in J_v} P_i(v) \in \sum_{i \in J_v} \text{Im} P_i,
\]
proving \( V = \sum_{i \in J} \text{Im} \, P_i \). Now let \( J \subseteq I \) be finite such that \( 0 = \sum_{j \in J} u_j \) with \( u_j \in \text{Im} \, P_j \) for each \( j \in J \). Then there exist \( v_j \in V \) such that \( u_j = P_j v_j \). Thus, we obtain

\[
\forall \ i \in J \quad 0 = P_i(0) = P_i \left( \sum_{j \in J} u_j \right) = \sum_{j \in J} P_i P_j v_j = P_i \sum_{j \in J} P_j v_j = P_i v_i = u_i,
\]

showing Def. 5.1(ii) to hold and proving (b).

(c): We have

\[
(Id - P)^2 = (Id - P)(Id - P) = Id - P - P + P^2 = Id - P,
\]

showing \( Id - P \) to be a projection. On the other hand, \( P(Id - P) = (Id - P)P = P - P^2 = P - P = 0 \), i.e.

\[
V = \text{Im} \, P \oplus \text{Im}(Id - P) \tag{5.5}
\]

according to (b). We show \( \ker P = \text{Im}(Id - P) \) next: Let \( v \in V \) and \( x := (Id - P)v \in \text{Im}(Id - P) \). Then \( Px = P(v - P^2v) = Pv - Pv = 0 \), showing \( x \in \ker P \) and \( \text{Im}(Id - P) \subseteq \ker P \). Conversely, let \( x \in \ker P \). By (5.5), we write \( x = v_1 + v_2 \) with \( v_1 \in \text{Im} \, P \) and \( v_2 \in \text{Im}(Id - P) \subseteq \ker P \). Then \( v_1 = x - v_2 \in \ker P \) as well, i.e. \( v_1 \in \text{Im} \, P \cap \ker P \). Then there exists \( w \in V \) such that \( v_1 = Pw \) and we compute

\[
v_1 = Pw = P^2w = Pv_1 = 0,
\]

as \( v_1 \in \ker P \) as well. Thus, \( x = v_2 \in \text{Im}(Id - P) \), showing \( \ker P \subseteq \text{Im}(Id - P) \) as desired. From (5.5), we then also have \( V = \text{Im} \, P \oplus \ker P \). Since we have seen \( Id - P \) to be a projection as well, we also obtain \( \ker(Id - P) = \text{Im}(Id - (Id - P)) = \text{Im} \, P \). □

6 Eigenvalues

**Definition 6.1.** Let \( V \) be a vector space over the field \( F \) and \( A \in \mathcal{L}(V, V) \).

(a) We call \( \lambda \in F \) an *eigenvalue* of \( A \) if, and only if, there exists \( 0 \neq v \in V \) such that

\[
Av = \lambda v. \tag{6.1}
\]

Then each \( 0 \neq v \in V \) such that (6.1) holds is called an *eigenvector* of \( A \) for the eigenvalue \( \lambda \); the set

\[
E_A(\lambda) := \ker (\lambda \, \text{Id} - A) = \{ v \in V : Av = \lambda v \} \tag{6.2}
\]

is then called the *eigenspace* of \( A \) with respect to the eigenvalue \( \lambda \). The set

\[
\sigma(A) := \{ \lambda \in F : \lambda \text{ eigenvalue of } A \} \tag{6.3}
\]

is called the *spectrum* of \( A \).
(b) We call \( A \) diagonalizable if, and only if, there exists basis \( B \) of \( V \) such that each \( v \in B \) is an eigenvector of \( A \).

**Remark 6.2.** Let \( V \) be a finite-dimensional vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \), and assume \( A \in \mathcal{L}(V,V) \) to be diagonalizable. Then there exists a basis \( B = \{v_1, \ldots, v_n\} \) of \( V \), consisting of eigenvectors of \( A \), i.e. \( Av_i = \lambda_i v_i \), \( \lambda_i \in \sigma(A) \), for each \( i \in \{1, \ldots, n\} \). Thus, with respect to \( B \), \( A \) is represented by the diagonal matrix

\[
\text{diag}(\lambda_1, \ldots, \lambda_n) = \begin{pmatrix}
\lambda_1 \\
\vdots \\
\lambda_n
\end{pmatrix},
\]

which explains the term diagonalizable.

It will be a main goal of the present and the following sections to investigate under which conditions, given \( A \in \mathcal{L}(V,V) \) with \( \dim V < \infty \), \( V \) has a basis \( B \) such that, with respect to \( B \), \( A \) is represented by a diagonal matrix. While such a basis does not always exist, we will see that there always exist bases such that the representing matrix has a particularly simple structure, a so-called normal form.

**Theorem 6.3.** Let \( V \) be a vector space over the field \( F \) and \( A \in \mathcal{L}(V,V) \).

(a) \( \lambda \in F \) is an eigenvalue of \( A \) if, and only if, \( \ker(\lambda \text{Id} - A) \neq \{0\} \), i.e. if, and only if, \( \lambda \text{Id} - A \) is not injective.

(b) For each eigenvalue \( \lambda \) of \( A \), the eigenspace \( E_A(\lambda) \) constitutes a subspace of \( V \).

(c) Let \( (v_\lambda)_{\lambda \in \sigma(A)} \) be a family in \( V \) such that, for each \( \lambda \in \sigma(A) \), \( v_\lambda \) is an eigenvector for \( \lambda \). Then \( (v_\lambda)_{\lambda \in \sigma(A)} \) is linearly independent (in particular, \( \#\sigma(A) \leq \dim V \)).

(d) \( A \) is diagonalizable if, and only if, \( V \) is the direct sum of the eigenspaces of \( V \), i.e. if, and only if,

\[
V = \bigoplus_{\lambda \in \sigma(A)} E_A(\lambda). \tag{6.4}
\]

(e) Let \( A \) be diagonalizable and, for each \( \lambda \in \sigma(A) \), let \( P_\lambda : V \to V \) be the projection with

\[
\text{Im} \, P_\lambda = E_A(\lambda) \quad \text{and} \quad \ker P_\lambda = \bigoplus_{\mu \in \sigma(A) \setminus \{\lambda\}} E_A(\mu)
\]

given by (d) in combination with Th. 5.8(a). Then

\[
A = \sum_{\lambda \in \sigma(A)} \lambda P_\lambda \tag{6.5a}
\]

and

\[
\forall_{\lambda \in \sigma(A)} \quad AP_\lambda = P_\lambda A, \tag{6.5b}
\]

where (6.5a) is known as the spectral decomposition of \( A \).
Proof. (a) holds, as, for each \( \lambda \in F \) and each \( v \in V \),

\[
Av = \lambda v \iff \lambda v - Av = 0 \iff (\lambda \text{Id} - A)v = 0 \iff v \in \ker(\lambda \text{Id} - A).
\]

(b) holds, as \( E_A(\lambda) \) is the kernel of a linear map.

(c): Seeking a contradiction, assume \((v_\lambda)_{\lambda \in \sigma(A)}\) to be linearly dependent. Then there exists a minimal family of vectors \((v_\lambda_1, \ldots, v_\lambda_k)\), \(k \in \mathbb{N}\), such that \(\lambda_1, \ldots, \lambda_k \in \sigma(A)\) and there exist \(c_1, \ldots, c_k \in F \setminus \{0\}\) with

\[
0 = \sum_{i=1}^{k} c_i v_{\lambda_i}.
\]

We compute

\[
0 = 0 - 0 = A \left( \sum_{i=1}^{k} c_i v_{\lambda_i} \right) - \lambda_k \sum_{i=1}^{k} c_i v_{\lambda_i} = \sum_{i=1}^{k} c_i \lambda_i v_{\lambda_i} - \lambda_k \sum_{i=1}^{k} c_i v_{\lambda_i} = \sum_{i=1}^{k-1} c_i (\lambda_i - \lambda_k) v_{\lambda_i}.
\]

As we had chosen the family \((v_{\lambda_1}, \ldots, v_{\lambda_k})\) to be minimal, we obtain \(c_i (\lambda_i - \lambda_k) = 0\) for each \(i \in \{1, \ldots, k - 1\}\), which is a contradiction, since \(c_i \neq 0\) as well as \(\lambda_i \neq \lambda_k\).

(d): If \(A\) is diagonalizable, \(V\) has basis \(B\), consisting of eigenvectors of \(A\). Letting, for each \(\lambda \in \sigma(A)\), \(B_\lambda := \{b \in B : Ab = \lambda b\}\), we have that \(B_\lambda\) is a basis of \(E_A(\lambda)\). Since we have \(B = \bigcup_{\lambda \in \sigma(A)} B_\lambda\) (6.4) now follows from Prop. 5.3(a). Conversely, if (6.4) holds, then \(V\) has a basis of eigenvectors of \(A\) by means of Prop. 5.3(b).

(e): Exercise. \(\blacksquare\)

**Corollary 6.4.** Let \(V\) be a vector space over the field \(F\) and \(A \in \mathcal{L}(V, V)\). If \(\dim V = n \in \mathbb{N}\) and \(A\) has \(n\) distinct eigenvalues \(\lambda_1, \ldots, \lambda_n \in F\), then \(A\) is diagonalizable.

**Proof.** Due to Th. 6.3(b),(c), we must have

\[
V = \bigoplus_{i=1}^{n} E_A(\lambda_i),
\]

showing \(A\) to be diagonalizable by Th. 6.3(d). \(\blacksquare\)

The following examples illustrate the dependence of diagonalizability, and even of the mere existence of eigenvalues, on the structure of the field \(F\).

**Example 6.5.** (a) Let \(K \in \{\mathbb{R}, \mathbb{C}\}\) and let \(V\) be a vector space over \(K\) with \(\dim V = 2\) and ordered basis \(B := (v_1, v_2)\). Consider \(A \in \mathcal{L}(V, V)\) such that

\[
Av_1 = v_2, \quad Av_2 = -v_1.
\]
With respect to $B$, $A$ is then given by the matrix $M := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. In consequence,

$$M^2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix},$$

showing $A^2 = -\text{Id}$ as well. Suppose $\lambda \in \sigma(A)$ and $v \in E_A(\lambda)$. Then

$$-v = A^2v = \lambda^2 v \implies \lambda^2 = -1.$$  

Thus, for $K = \mathbb{R}$, $A$ has no eigenvalues, $\sigma(A) = \emptyset$. For $K = \mathbb{C}$, we obtain

$$A(v_1 + iv_2) = v_2 - iv_1 = -i(v_1 + iv_2),$$

$$A(v_1 - iv_2) = v_2 + iv_1 = i(v_1 - iv_2),$$

showing $A$ to be diagonalizable with $\sigma(A) = \{i, -i\}$ and $\{v_1 + iv_2, v_1 - iv_2\}$ being a basis of $V$ of eigenvectors of $A$.

(b) Over $\mathbb{R}$, consider the vector spaces

$$V_1 := \{(f : \mathbb{R} \to \mathbb{R}) : f \text{ polynomial function}\},$$

$$V_2 := \{\{(\exp_a : \mathbb{R} \to \mathbb{R}) : a \in \mathbb{R}\}\},$$

where

$$\forall a \in \mathbb{R} \quad \exp_a : \mathbb{R} \to \mathbb{R}, \quad \exp_a(x) := e^{ax}.$$  

For $i \in \{1, 2\}$, consider the linear map

$$D : V_i \to V_i, \quad D(f) := f'.$$

If $P \in V_1 \setminus \{0\}$, then $\deg(DP) < \deg(P)$. If $P \in V_1$ is constant, then $DP = 0 \cdot P = 0$. In consequence $0 \in \mathbb{R}$ is the only eigenvalue of $D : V_1 \to V_1$. On the other hand, for each $a \in \mathbb{R}$, $D(\exp_a) = a \exp_a$, showing $\sigma(D) = \mathbb{R}$ for $D : V_2 \to V_2$. In this case, $D$ is even diagonalizable, since $B := \{\exp_a : a \in \mathbb{R}\}$ is a basis of $V_2$ of eigenvectors of $D$.

(c) Let $V$ be a vector space over the field $F$ and assume $\text{char } F \neq 2$. Moreover, let $A \in \mathcal{L}(V, V)$ such that $A^2 = \text{Id}$ (for $\dim V = 2$, a nontrivial example is given by each $A$ represented by the matrix $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ with respect to some ordered basis of $V$). We claim $A$ to be diagonalizable with $\sigma(A) = \{-1, 1\}$ and

$$V = E_A(-1) \oplus E_A(1):$$

Indeed, if $x \in \text{Im}(A + \text{Id})$, then there exists $v \in V$ such that $x = (A + \text{Id})v$, implying

$$Ax = A(A + \text{Id})v = (A^2 + A)v = (\text{Id} + A)v = x,$$
showing \( \text{Im}(A + \text{Id}) \subseteq E_A(1) \). If \( x \in \text{Im}(\text{Id} - A) \), then there exists \( v \in V \) such that \( x = (\text{Id} - A)v \), implying

\[
Ax = A(\text{Id} - A)v = (A - A^2)v = (A - \text{Id})v = -x,
\]

showing \( \text{Im}(\text{Id} - A) \subseteq E_A(-1) \). Now let \( v \in V \) be arbitrary. We use \( 2 \neq 0 \) in \( F \) to obtain

\[
v = \frac{1}{2}(v + Av) + \frac{1}{2}(v - Av) \in \text{Im}(\text{Id} + A) + \text{Im}(\text{Id} - A) \subseteq E_A(1) + E_A(-1),
\]

showing \( V = E_A(1) + E_A(-1) \).

(d) Let \( V \) be a vector space over the field \( F \) with \( \dim V = 2 \) and ordered basis \( B := (v_1, v_2) \). Consider \( A \in \mathcal{L}(V, V) \) such that \( Av_1 = v_1 \) and \( Av_2 = v_1 + v_2 \).

With respect to \( B \), \( A \) is then given by the matrix \( \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \). Due to \( Av_1 = v_1 \), we have \( 1 \in \sigma(A) \). Let \( v \in V \). Then there exist \( c_1, c_2 \in F \) such that \( v = c_1v_1 + c_2v_2 \).

If \( \lambda \in \sigma(A) \) and \( 0 \neq v \in E_A(\lambda) \), then

\[
\lambda(c_1v_1 + c_2v_2) = \lambda v = Av = c_1v_1 + c_2v_1 + c_2v_2.
\]

As the coordinates with respect to the basis \( B \) are unique, we conclude \( \lambda c_1 = c_1 + c_2 \) and \( \lambda c_2 = c_2 \). If \( c_2 \neq 0 \), then the second equation yields \( \lambda = 1 \). If \( c_2 = 0 \), then \( c_1 \neq 0 \) and the first equation yields \( \lambda = 1 \). Altogether, we obtain \( \sigma(A) = \{1\} \) and, since \( A \neq \text{Id} \), \( A \) is not diagonalizable.

**Definition 6.6.** Let \( S \) be a set and \( A : S \rightarrow S \). Then \( U \subseteq S \) is called \( A \)-invariant if, and only if, \( A(U) \subseteq U \).

**Proposition 6.7.** Let \( V \) be a vector space over the field \( F \) and let \( U \subseteq V \) be a subspace. If \( A \in \mathcal{L}(V, V) \) is diagonalizable and \( U \) is \( A \)-invariant, then \( A|U \) is diagonalizable as well.

**Proof.** Let \( A \) be diagonalizable, let \( U \) be \( A \)-invariant, and set \( A_U := A|U \). Clearly,

\[
\forall \lambda \in \sigma(A) \quad U \cap E_A(\lambda) = \begin{cases} E_{Au}(\lambda) & \text{for } \lambda \in \sigma(A_U), \\ \{0\} & \text{for } \lambda \notin \sigma(A_U). \end{cases}
\]

As \( A \) is diagonalizable, from Th. 6.3(d), we know

\[
V = \bigoplus_{\lambda \in \sigma(A)} E_A(\lambda).
\]

(6.6)

It suffices to show

\[
U = W := \sum_{\lambda \in \sigma(A)} (U \cap E_A(\lambda)),
\]
since, then
\[ U = \bigoplus_{\lambda \in \sigma(A_U)} E_{A_U}(\lambda) \]
due to Th. 6.3(d) and Prop. 5.2(iv). Thus, seeking a contradiction, let \( u \in U \setminus W \) (note \( u \neq 0 \)). Then there exist distinct \( \lambda_1, \ldots, \lambda_n \in \sigma(A) \), \( n \in \mathbb{N} \), such that \( u = \sum_{i=1}^{n} v_i \) with \( v_i \in E_A(\lambda_i) \setminus \{0\} \) for each \( i \in \{1, \ldots, n\} \), where we may choose \( u \in U \setminus W \) such that \( n \in \mathbb{N} \) is minimal. Since \( U \) is \( A \)-invariant, we know
\[ Au = \sum_{i=1}^{n} Av_i = \sum_{i=1}^{n} \lambda_i v_i \in U. \]
As \( \lambda_n u \in U \) as well, we conclude
\[ Au - \lambda_n u = \sum_{i=1}^{n-1} (\lambda_i - \lambda_n) v_i \in U \]
as well. Since \( u \in U \setminus W \) was chosen such that \( n \) is minimal, we must have \( Au - \lambda_n u \in U \cap W \). Thus, there exists a finite set \( \sigma_u \subseteq \sigma(A) \) such that
\[ Au - \lambda_n u = \sum_{i=1}^{n-1} (\lambda_i - \lambda_n) v_i = \sum_{\lambda \in \sigma_u} \lambda v_{\lambda} \land \forall \lambda \in \sigma_u \ 0 \neq \lambda \in U \cap E_A(\lambda). \quad (6.7) \]
Since the sum in (6.6) is direct, (6.7) and Prop. 5.2(ii) imply \( \sigma_u = \{\lambda_1, \ldots, \lambda_{n-1}\} \) and
\[ \forall i \in \{1, \ldots, n-1\} \quad \left( w_{\lambda_i} = (\lambda_i - \lambda_n) v_i \land \lambda_i \neq \lambda_n \Rightarrow v_i \in U \cap E_A(\lambda_i) \subseteq W \right). \]
On the other hand, this then implies
\[ v_n = u - \sum_{i=1}^{n-1} v_i \in U \quad \forall v_{\lambda_n} \in E_A(\lambda_n) \Rightarrow v_n \in W, \]
yielding the contradiction \( u = v_n + \sum_{i=1}^{n-1} v_i \in W \). Thus, the assumption that there exists \( u \in U \setminus W \) was false, proving \( U = W \) as desired. \( \blacksquare \)

We will now use Prop. 6.7 to prove a result regarding the \textit{simultaneous} diagonalizability of linear endomorphisms:

**Theorem 6.8.** Let \( V \) be a vector space over the field \( F \) and let \( A_1, \ldots, A_n \in \mathcal{L}(V, V) \), \( n \in \mathbb{N} \), be diagonalizable linear endomorphisms. Then the \( A_1, \ldots, A_n \) are simultaneously diagonalizable (i.e. there exists a basis \( B \) of \( V \), consisting of eigenvectors of \( A_i \) for each \( i \in \{1, \ldots, n\} \)) if, and only if,
\[ \forall i,j \in \{1, \ldots, n\} \quad A_i A_j = A_j A_i. \quad (6.8) \]
6 EIGENVALUES

Proof. Suppose $B$ is a basis of $V$ such that

$$
\forall \ i \in \{1, \ldots, n\} \quad \forall \ b \in B \quad \exists \ \lambda_{i,b} \in \sigma(A_i) \quad A_i b = \lambda_{i,b} b.
$$

(6.9)

Then

$$
\forall \ i, j \in \{1, \ldots, n\} \quad A_i A_j b = \lambda_{i,b} \lambda_{j,b} b = A_j A_i b,
$$

proving (6.8). Conversely, assume (6.8) to hold. We prove (6.9) via induction on $n \in \mathbb{N}$. For technical reasons, we actually prove (6.9) via induction on $n \in \mathbb{N}$ in the following, clearly, equivalent form: There exists a family $(V_k)_{k \in K}$ of subspaces of $V$ such that

(i) $V = \bigoplus_{k \in K} V_k$.

(ii) For each $k \in K$, $V_k$ has a basis $B_k$, consisting of eigenvectors of $A_i$ for each $i \in \{1, \ldots, n\}$.

(iii) For each $k \in K$ and each $i \in \{1, \ldots, n\}$, $V_k$ is contained in some eigenspace of $A_i$, i.e.

$$
\forall \ k \in K \quad \forall \ i \in \{1, \ldots, n\} \quad \exists \ \lambda_{i,k} \in \sigma(A_i) \quad \left( V_k \subseteq E_{A_i}(\lambda_{i,k}), \quad \forall \ v \in V_k \ A_i v = \lambda_{i,k} v \right).
$$

(iv) For each $k, l \in K$ with $k \neq l$, there exists $i \in \{1, \ldots, n\}$ such that $V_k$ and $V_l$ are not contained in the same eigenspace of $A_i$, i.e.

$$
\forall \ k, l \in K \quad \left( k \neq l \implies \exists \ i \in \{1, \ldots, n\} \quad \lambda_{i,k} \neq \lambda_{i,l} \right).
$$

For $n = 1$, we can simply use $K := \sigma(A_1)$ and, for each $\lambda \in K$, $V_\lambda := E_{A_1}(\lambda)$. Thus, consider $n > 1$. By induction, assume (i) – (iv) to hold with $n$ replaced by $n - 1$. It suffices to show that the spaces $V_k$, $k \in K$, are all $A_n$-invariant, i.e. $A_n(V_k) \subseteq V_k$: Then, according to Prop. 6.7, $A_{n_k} := A_n | V_k$ is diagonalizable, i.e.

$$
V_k = \bigoplus_{\lambda \in \sigma(A_{n_k})} E_{A_{n_k}}(\lambda).
$$

Now each $V_{k,\lambda} := E_{A_{n_k}}(\lambda)$ has a basis $B_{k,\lambda}$, consisting of eigenvectors of $A_{n_k}$. Since, for each $i \in \{1, \ldots, n - 1\}$, $B_{k,\lambda} \subseteq V_k \subseteq E_{A_i}(\lambda_{i,k})$, $B_{k,\lambda}$ consists of eigenvectors of $A_i$ for each $i \in \{1, \ldots, n\}$. Letting $K_n := \{(k, \lambda) : k \in K, \lambda \in \sigma(A_{n_k})\}$, (i) – (iv) then hold with $K$ replaced by $K_n$. Thus, it remains to show $A_n(V_k) \subseteq V_k$ for each $k \in K$: Fix $v \in V_k$, $k \in K$. We have

$$
\forall \ j \in \{1, \ldots, n - 1\} \quad A_j(A_n v) \stackrel{(6.8)}{=} A_n(A_j v) = A_n(\lambda_{jk} v).
$$

Moreover, there exists a finite set $K_v \subseteq K$ such that

$$
A_n v = \sum_{l \in K_v} v_l + \sum_{l \in K_v} \forall \ v_l \in V_l \setminus \{0\}.
$$
Then
\[ \forall j \in \{1, \ldots, n-1\} \quad \lambda_{jk} \sum_{l \in K_v} v_l = \lambda_{jk} (A_n v) = A_n (\lambda_{jk} v) = A_j (A_n v) = \sum_{l \in K_v} A_j v_l = \sum_{l \in K_v} \lambda_{jl} v_l. \]

As the sum in (i) is direct, Prop. 5.2(ii) implies \( \lambda_{jk} v_l = \lambda_{jl} v_l \) for each \( l \in K_v \). For each \( l \in K_v \), we have \( v_l \neq 0 \), implying \( \lambda_{jk} = \lambda_{jl} \) for each \( j \in \{1, \ldots, n-1\} \). Thus, by (iv), \( k = l \), i.e. \( K_v = \{k\} \) and \( A_n v = v_k \in V_k \) as desired. \( \blacksquare \)

In general, computing eigenvalues is a difficult task, and we will say more about this issue below. The following results can sometimes help, where Th. 6.9(a) is most useful for \( \dim V \) small:

**Theorem 6.9.** Let \( V \) be a vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \). Let \( A \in \mathcal{L}(V,V) \).

(a) \( \lambda \in F \) is an eigenvalue of \( A \) if, and only if,
\[ \det(\lambda \text{Id} - A) = 0. \]

(b) If there exists a basis \( B \) of \( V \) such that the matrix \( (a_{ji}) \in \mathcal{M}(n, F) \) of \( A \) with respect to \( B \) is upper or lower triangular, then the diagonal elements \( a_{ii} \) are precisely the eigenvalues of \( A \), i.e. \( \sigma(A) = \{a_{ii} : i \in \{1, \ldots, n\}\} \).

**Proof.** (a): According to Th. 6.3(a), \( \lambda \in \sigma(A) \) is equivalent to \( \lambda \text{Id} - A \) not being injective, which (as \( V \) is finite-dimensional) is equivalent to \( \det(\lambda \text{Id} - A) = 0 \) by Cor. 4.37(b).

(b): For each \( \lambda \in F \), we have
\[ \det(\lambda \text{Id} - A) = \det(\lambda \text{Id}_n - (a_{ji})) = \prod_{i=1}^{n} (\lambda - a_{ii}). \]

Thus, by (a), \( \sigma(A) = \{a_{ii} : i \in \{1, \ldots, n\}\} \). \( \blacksquare \)

**Example 6.10.** Consider the vector space \( V := \mathbb{R}^2 \) over \( \mathbb{R} \) and, with respect to the standard basis, let \( A \in \mathcal{L}(V,V) \) be given by the matrix \( M := \begin{pmatrix} 3 & -2 \\ 1 & 0 \end{pmatrix} \). Then, for each \( \lambda \in \mathbb{R} \),
\[ \det(\lambda \text{Id} - A) = \begin{vmatrix} \lambda - 3 & 2 \\ -1 & \lambda \end{vmatrix} = (\lambda - 3) \cdot \lambda + 2 = \lambda^2 - 3\lambda + 2 = (\lambda - 1)(\lambda - 2), \]
i.e. \( \sigma(A) = \{1, 2\} \) by Th. 6.9(a). Since
\[ M \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \Rightarrow v_1 = v_2, \quad M \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 2v_1 \\ 2v_2 \end{pmatrix} \Rightarrow v_1 = 2v_2, \]
\( B := \{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \} \) is a basis of eigenvectors of \( A \).
Theorem 6.9(a) gives rise to the following definition:

**Definition 6.11.** If $V$ is a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, and $A \in \mathcal{L}(V, V)$, then the function 

$$p_A : F \to F, \quad p_A(t) := \det(t \text{Id} - A),$$

is called the *characteristic polynomial function* of $A$ (some authors call $p_A$ the characteristic polynomial of $A$ – here, we distinguish $p_A$ from a related, but different object, which we will call the characteristic polynomial of $A$ and which we will introduce in a later section). As it turns out, $p_A$ is, indeed, a polynomial function, cf. Rem. 6.12(a) below.

**Remark 6.12.** Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, and $A \in \mathcal{L}(V, V)$.

(a) Let $p_A$ be the characteristic polynomial function of $A$ as defined in Def. 6.11. If $B$ is a basis of $V$, the matrix $(a_{ji}) \in \mathcal{M}(n, F)$ represents $A$ with respect to $B$, and, for each $t \in F$, $(c_{ji}(t)) := (t \text{Id}_n - (a_{ji}))$, then 

$$\forall t \in F \quad p_A(t) = \det(t \text{Id}_n - (a_{ji})) = \prod_{i=1}^{n} (t - a_{ii}) + \sum_{\pi \in S_n \setminus \{\text{Id}\}} \text{sgn}(\pi) \prod_{i=1}^{n} c_{\pi(i)}(t).$$

Clearly, the degree of the first summand is $n$ and, for each $\pi \in S_n \setminus \{\text{Id}\}$, the degree of the corresponding summand is at most $n - 2$, showing $p_A$ to be a polynomial function of degree $n$ (as a caveat, consult the footnote to the degree definition in Def. 4.32 and Rem. 7.14 below). Moreover, we see $p_A$ to be a *monic* polynomial, i.e. the coefficient of $t^n$ is 1.

(b) According to Th. 6.9(a), $\sigma(A)$ is precisely the set of zeros of $p_A$.

(c) Some authors prefer to define the characteristic polynomial function of $A$ as the polynomial function 

$$q_A : F \to F, \quad q_A(t) := \det(A - t \text{Id}) = (-1)^n p_A(t).$$

While $q_A$ still has the property that $\sigma(A)$ is precisely the set of zeros of $q_A$, $q_A$ is monic only for $n$ even. On the other hand, $q_A$ has the advantage that $q_A(0) = \det(A)$.

(d) According to (b), the task for finding the eigenvalues of $A$ is the same as the task of finding the zeros of the polynomial function $p_A$. So one might hope that only particularly simple polynomial functions can occur as characteristic polynomial functions. However, this is not the case: Indeed, *every* monic polynomial function of degree $n$ occurs as a characteristic polynomial function: Let $a_1, \ldots, a_n \in F$, and 

$$P : F \to F, \quad P(t) := t^n + \sum_{i=1}^{n} a_i t^{n-i}.$$
We define the companion matrix of $P$ to be

$$M(P) := \begin{pmatrix} -a_1 & -a_2 & -a_3 & \cdots & -a_{n-1} & -a_n \\ 1 & 0 & 1 & \cdots & 0 \\ & & & \ddots & \ddots & 0 \\ & & & & & 1 \end{pmatrix}$$

and claim $p_A = P$, if $A \in \mathcal{L}(F^n, F^n)$ is the linear map, represented by $M(P)$ with respect to the standard basis of $F^n$: Indeed, using Laplace expansion with respect to the first row, we obtain, for each $t \in F$,

$$p_A(t) = \begin{vmatrix} t + a_1 & a_2 & a_3 & \cdots & a_{n-1} & a_n \\ -1 & t & -1 & \cdots & 0 & 1 \\ & & & \ddots & \ddots & 1 \\ & & & & & t \\ & & & & & -1 \end{vmatrix}$$

$$= (-1)^{n+1}a_n(-1)^{n-1} + (-1)^na_{n-1}(-1)^{n-2}t + \cdots + (-1)^2a_2(-1)^1t^{n-2} + (-1)^2(t+a_1)t^{n-1}$$

$$= \sum_{i=0}^{n-2}(-1)^{2(n-i)}a_{n-i}t^i + (-1)^2(t+a_1)t^{n-1}$$

$$= t^n + \sum_{i=0}^{n-1}a_{n-i}t^i = t^n + \sum_{i=1}^{n}a_i t^{n-i} = P(t).$$

(e) In Ex. 6.5(a), we saw that the considered linear endomorphism $A$ had eigenvalues for $F = \mathbb{C}$, but no eigenvalues for $F = \mathbb{R}$, which we can now relate to the fact that $p_A(t) = t^2 + 1$ has no zeros over $\mathbb{R}$, but $p_A(t) = (t-i)(t+i)$ with zeros $\pm i$ over $\mathbb{C}$. One calls a field $F$ algebraically closed if, and only if, every polynomial $P$ with coefficients in $F$ and degree $n \in \mathbb{N}$, can be written in the form $P(t) = c \prod_{i=1}^{n}(t-\lambda_i)$, where $c \in F$ and $\lambda_1, \ldots, \lambda_n \in F$ are precisely the zeros of $P$ ($\mathbb{R}$ is not algebraically closed, but $\mathbb{C}$ is algebraically closed due to the fundamental theorem of algebra, cf. [Phi16a, Th. 8.32]). It is an important result of Algebra that every field $F$ is contained in an algebraically closed field (see, e.g., [Bos13, Cor. 3.4.7], [Lan05, Cor. V.2.6]).

---

5One calls a field an $F$ algebraic closure of $F$ if, and only if, $F$ is algebraically closed and $F$ is minimal in the sense that each $\lambda \in F$ is a zero of some polynomial with coefficients in $F$ (one then says that $F$ is an algebraic extension of $F$). The mentioned references [Bos13, Cor. 3.4.7], [Lan05, Cor. V.2.6] actually show that every field is contained in an algebraic closure. If $F$ is not algebraically closed, then the algebraic closure of $F$ is not unique, but at least all algebraic closures of $F$ are isomorphic.
Given that eigenvalues are precisely the zeros of the characteristic polynomial function, and given that, according to (d), every polynomial function of degree \( n \) can occur as the characteristic polynomial of a matrix, it is not surprising that computing eigenvalues is, in general, a difficult task, even if \( F \) is algebraically closed. It is another result of Algebra that, for a generic polynomial of degree at least 5, it is not possible to obtain its zeros using so-called radicals (which are, roughly, zeros of polynomials of the form \( t^k - \lambda \), \( k \in \mathbb{N}, \lambda \in F \), see, e.g., [Bos13, Def. 6.1.1] for a precise definition) in finitely many steps (cf., e.g., [Bos13, Cor. 6.1.7]). In practice, one often has to make use of approximative numerical methods (see, e.g., [Phi17b, Sec. I]). Having said that, let us note that the problem of computing eigenvalues is, indeed, typically easier than the general problem of computing zeros of polynomials. This is due to the fact that the difficulty of computing the zeros of a polynomial depends tremendously on the form in which the polynomial is given: It is typically hard if the polynomial is expanded into the form \( p(t) = \sum_{i=0}^{n} a_i t^i \), but it is easy (trivial, in fact), if the polynomial is given in a factored form \( p(t) = c \prod_{i=1}^{n} (t - \lambda_i) \).

If the characteristic polynomial function is given implicitly by a matrix, one is, in general, somewhere between the two extremes. In particular, for a large matrix, it usually makes no sense to compute the characteristic polynomial function in its expanded form (this is an expensive task in itself and, in the process, one even loses the additional structure given by the matrix). It makes much more sense to use methods tailored to the computation of eigenvalues, and, if available, one should make use of additional structure a matrix might have.

**Remark 6.13.** Let \( V \) be a vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \), and \( A \in \mathcal{L}(V, V) \). Moreover, let \( \lambda \in \sigma(A) \). Clearly, one has

\[
\{0\} \subseteq \ker(A - \lambda \mathrm{Id}) \subseteq \ker(A - \lambda \mathrm{Id})^2 \subseteq \ldots
\]

and the inclusion can be strict for at most \( n \) times. Let

\[
r(\lambda) := \min \{k \in \mathbb{N} : \ker(A - \lambda \mathrm{Id})^k = \ker(A - \lambda \mathrm{Id})^{k+1}\}.
\]

Then

\[
\forall k \in \mathbb{N} \quad \ker(A - \lambda \mathrm{Id})^{r(\lambda)} = \ker(A - \lambda \mathrm{Id})^{r(\lambda)+k};
\]

indeed, otherwise, let \( k_0 := \min\{k \in \mathbb{N} : \ker(A - \lambda \mathrm{Id})^{r(\lambda)} \subsetneq \ker(A - \lambda \mathrm{Id})^{r(\lambda)+k}\} \). Then there exists \( v \in V \) such that \( (A-\lambda \mathrm{Id})^{r(\lambda)+k_0}v = 0 \), but \( (A-\lambda \mathrm{Id})^{r(\lambda)+k_0-1}v \neq 0 \). However, that means \( w := (A-\lambda \mathrm{Id})^{k_0-1}v \in \ker(A - \lambda \mathrm{Id})^{r(\lambda)+1} \), but \( w \notin \ker(A - \lambda \mathrm{Id})^{r(\lambda)} \), in contradiction to the definition of \( r(\lambda) \).

**Definition 6.14.** Let \( V \) be a vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \), and \( A \in \mathcal{L}(V, V) \). Moreover, let \( \lambda \in \sigma(A) \). The number

\[
m_a(\lambda) := \dim \ker(A - \lambda \mathrm{Id})^{r(\lambda)} \in \{1, \ldots, n\},
\]

where \( r(\lambda) \) is given by (6.10), is called the **algebraic multiplicity** of the eigenvalue \( \lambda \), whereas

\[
m_g(\lambda) := \dim \ker(A - \lambda \mathrm{Id}) \in \{1, \ldots, n\}
\]
is called its \textit{geometric multiplicity}. We call \( \lambda \) \textit{simple} if, and only if, \( m_g(\lambda) = m_a(\lambda) = 1 \); we call \( \lambda \) \textit{semisimple} if, and only if, \( m_g(\lambda) = m_a(\lambda) \). For each \( k \in \{1, \ldots, r(\lambda)\} \), the space

\[
E_A^k(\lambda) := \ker(A - \lambda \text{Id})^k
\]

is called the \textit{generalized eigenspace} of rank \( k \) of \( A \), corresponding to the eigenvalue \( \lambda \); each \( v \in E_A^k(\lambda) \setminus E_A^{k-1}(\lambda) \), \( k \geq 2 \), is called a \textit{generalized eigenvector} of rank \( k \) with corresponding to the eigenvalue \( \lambda \) (an eigenvector \( v \in E_A(\lambda) \) is sometimes called a \textit{generalized eigenvector} of rank \( 1 \)).

\begin{proposition}
Let \( V \) be a vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \), and \( A \in \mathcal{L}(V,V) \).

(a) If \( \lambda \in \sigma(A) \) and \( r(\lambda) \) is given by (6.10), then

\[
1 \leq r(\lambda) \leq n.
\]

If \( k \in \mathbb{N} \) is such that \( 1 \leq k < k + 1 \leq r(\lambda) \), then

\[
1 \leq m_g(\lambda) = \dim E_A^1(\lambda) \leq \dim E_A^k(\lambda) < \dim E_A^{k+1}(\lambda) \leq \dim E_A^{r(\lambda)}(\lambda) = m_a(\lambda) \leq n,
\]

which implies, in particular,

\[
\begin{align*}
    r(\lambda) &\leq m_a(\lambda), \\
    1 &\leq m_g(\lambda) \leq m_a(\lambda) \leq n, \\
    0 &\leq m_a(\lambda) - m_g(\lambda) \leq n - 1.
\end{align*}
\]

(b) If \( \lambda \in \sigma(A) \), then, for each \( k \in \{1, \ldots, r(\lambda)\} \), the generalized eigenspace \( E_A^k(\lambda) \) is \( A \)-invariant, i.e.

\[
A(E_A^k(\lambda)) \subseteq E_A^k(\lambda).
\]

(c) If \( A \) is diagonalizable, then \( m_g(\lambda) = m_a(\lambda) \) holds for each \( \lambda \in \sigma(A) \) (but cf. Ex. 6.16 below).

\end{proposition}

\textit{Proof.} (a): Both (6.14) and (6.15) are immediate from Rem. 6.13 together with the definitions of \( r(\lambda), m_g(\lambda) \) and \( m_a(\lambda) \). Then (6.16a) follows from (6.15), since \( \dim E_A^{k+1}(\lambda) - \dim E_A^k(\lambda) \geq 1 \); (6.16b) is immediate from (6.15); (6.16c) is immediate from (6.16b).

(b): Due to \( A(A - \lambda \text{Id}) = (A - \lambda \text{Id})A \), one has

\[
\forall k \in \mathbb{N}_0 \quad A(\ker(A - \lambda \text{Id})^k) \subseteq \ker(A - \lambda \text{Id})^k:
\]

Indeed, if \( v \in \ker(A - \lambda \text{Id})^k \), then

\[
(A - \lambda \text{Id})^k(Av) = A(A - \lambda \text{Id})^k v = 0.
\]

(c): Exercise. □
Example 6.16. Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, $n \geq 2$. Let $\lambda \in F$. We show that there always exists a map $A \in \mathcal{L}(V, V)$ such that $\lambda \in \sigma(A)$ and such that the difference between $m_a(\lambda)$ and $m_g(\lambda)$ maximal, namely

$$m_a(\lambda) - m_g(\lambda) = n - 1 :$$

Let $B = (v_1, \ldots, v_n)$ be an ordered basis of $V$, and let $A \in \mathcal{L}(V, V)$ be such that

$$Av_1 = \lambda v_1 \quad \land \quad \forall i \in \{2, \ldots, n\} \quad Av_i = \lambda v_i + v_{i-1}.$$  

Then, with respect to $B$, $A$ is represented by the matrix

$$M := \begin{pmatrix} \lambda & 1 \\ & \lambda & 1 \\ & & \ddots & \ddots \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}.$$  

We use an induction over $k \{1, \ldots, n\}$ to show

$$\forall_{k \{1, \ldots, n\}} \quad \left( \forall_{1 \leq i \leq k} \ (A - \lambda \text{Id})^k v_i = 0 \quad \land \quad \forall_{k < i \leq n} \ (A - \lambda \text{Id})^k v_i = v_{i-k} \right): \quad (6.17)$$

For $k = 1$, we have

$$(A - \lambda \text{Id})v_1 = \lambda v_1 - \lambda v_1 = 0, \quad \forall_{1 < i \leq n} \ (A - \lambda \text{Id})v_i = \lambda v_i + v_{i-1} - \lambda v_i = v_{i-1},$$

as needed. For $k > 1$, we have

$$\forall_{1 \leq i < k} \ (A - \lambda \text{Id})^k v_i \overset{\text{ind.hyp.}}{=} 0, \quad (A - \lambda \text{Id})^k v_k \overset{\text{ind.hyp.}}{=} (A - \lambda \text{Id}) v_1 = 0,$$

$$\forall_{k < i \leq n} \ (A - \lambda \text{Id})^k v_i \overset{\text{ind.hyp.}}{=} (A - \lambda \text{Id}) v_{i-k+1} = \lambda v_{i-k+1} + v_{i-k} - \lambda v_{i-k+1} = v_{i-k},$$

completing the induction. In particular, since $\dim V = n$, (6.17) yields, for each $1 < k \leq n$, $\dim E_A^k(\lambda) - \dim E_A^{k-1}(\lambda) = 1$ and $\dim E_A^k(\lambda) = k$, implying $m_g(\lambda) = 1$ and $m_a(\lambda) = n$, as claimed.

Definition 6.17. Let $n \in \mathbb{N}$. Consider the vector space $V := F^n$ over the field $F$. All of the notions we introduced in this section for linear endomorphisms $A \in \mathcal{L}(V, V)$ (e.g. eigenvalue, eigenvector, eigenspace, multiplicity of an eigenvalue, diagonalizability, etc.), one also defines for quadratic matrices $M \in \mathcal{M}(n, F)$: The notions are then meant with respect to the linear map $A_M$ that $M$ represents with respect to the standard basis of $F^n$.

Example 6.18. Let $F$ be field and $n \in \mathbb{N}$. If $M \in \mathcal{M}(n, F)$ is diagonalizable, then, according to Def. 6.17, there exists a regular matrix $T \in \text{GL}_n(F)$ and a diagonal matrix $D \in \mathcal{M}(n, F)$ such that $D = T^{-1}MT$. A simple induction then shows

$$\forall_{k \in \mathbb{N}_0} \quad M^k = TD^kT^{-1}.$$
Clearly, if one knows $T$ and $T^{-1}$, this can tremendously simplify the computation of $M^k$, especially if $k$ is large and $M$ is fully populated. However, computing $T$ and $T^{-1}$ can also be difficult, and it depends on the situation if it is a good option to pursue this route.

## 7 Commutative Rings, Polynomials

We have already seen in the previous section that the eigenvalues of a quadratic matrix are precisely the zeros of its characteristic polynomial functions. In order to further study the relation between certain polynomials and the structure of a matrix (and the structure of corresponding linear maps), we will need to investigate some of the general theory of polynomials. We will take this opportunity to also learn more about the general theory of commutative rings, which is of algebraic interest beyond our current interest in matrix-related polynomials.

**Definition 7.1.** Let $R$ be a commutative ring with unity. We call

\[ R[X] := R^N_0 := \{(f : \mathbb{N}_0 \to R) : \#f^{-1}(R \setminus \{0\}) < \infty\} \tag{7.1} \]

the set of *polynomials* over $R$ (i.e. a polynomial over $R$ is a sequence $(a_i)_{i \in \mathbb{N}_0}$ in $R$ such that all, but finitely many, of the entries $a_i$ are 0, cf. [Phi19, Ex. 5.16(c)]). We then have the pointwise-defined addition and scalar multiplication on $R[X]$, which it inherits from $R^N_0$:

\[ \forall f, g \in R[X], \quad (f + g)(i) := f(i) + g(i), \]

\[ \forall f \in R[X], \quad \forall \lambda \in R, \quad (\lambda \cdot f)(i) := \lambda f(i), \tag{7.2} \]

where we know from [Phi19, Ex. 5.16(c)] that, with these compositions, $R[X]$ forms a vector space over $R$, provided $R$ is a field and, then, $B = \{e_i : i \in \mathbb{N}_0\}$, where

\[ \forall i \in \mathbb{N}_0, \quad e_i : \mathbb{N}_0 \to R, \quad e_i(j) := \delta_{ij}, \]

provides the standard basis of the vector space $R[X]$. In the current context, we will now write $X^i := e_i$ and we will call these polynomials *monomials*. Furthermore, we define a multiplication on $R[X]$ by letting

\[ ((a_i)_{i \in \mathbb{N}_0}, (b_i)_{i \in \mathbb{N}_0}) \mapsto (c_i)_{i \in \mathbb{N}_0} := (a_i)_{i \in \mathbb{N}_0} \cdot (b_i)_{i \in \mathbb{N}_0}, \]

\[ c_i := \sum_{k+l=i} a_k b_l := \sum_{(k, l) \in (\mathbb{N}_0)^2: k+l=i} a_k b_{i-k} \tag{7.3} \]

If $f := (a_i)_{i \in \mathbb{N}_0} \in R[X]$, then we call the $a_i \in R$ the *coefficients* of $f$, and we define the *degree* of $f$ by

\[ \deg f := \begin{cases} -\infty & \text{for } f \equiv 0, \\ \max\{i \in \mathbb{N}_0 : a_i \neq 0\} & \text{for } f \neq 0 \end{cases} \tag{7.4} \]
(defining \( \deg(0) = -\infty \) instead of \( \deg(0) = -1 \) has the advantage that formulas (7.5a), (7.5b) below then also hold for the zero polynomial). If \( \deg f = n \in \mathbb{N}_0 \) and \( a_n = 1 \), then the polynomial \( f \) is called \textit{monic}.

Remark 7.2. In the situation of Def. 7.1, using the notation \( X^i = e_i \), we can write addition, scalar multiplication, and multiplication in the following, perhaps, more familiar-looking forms: If \( \lambda \in R \), \( f = \sum_{i=0}^n f_i X^i \), \( g = \sum_{i=0}^n g_i X^i \), \( n \in \mathbb{N}_0 \), \( f_0, \ldots, f_n, g_0, \ldots, g_n \in R \), then

\[
\begin{align*}
  f + g &= \sum_{i=0}^n (f_i + g_i) X^i, \\
  \lambda f &= \sum_{i=0}^n (\lambda f_i) X^i, \\
  fg &= \sum_{i=0}^{n^2} \left( \sum_{k+l=i} f_k g_l \right) X^i.
\end{align*}
\]

Recall from [Phi19, Def. and Rem. 4.40] that an element \( x \) in a ring with unity \( R \) is called \textit{invertible} if, and only if, there exists \( x \in R \) such that \( xx = 1 \), and that \((R^*, \cdot)\) denotes the group of invertible elements of \( R \).

Lemma 7.3. Let \( R \) be a ring with unity and \( x \in R^* \).

(a) \( x \) is not a zero divisor.

(b) If \( S \) is also a ring with unity and \( \phi : R \rightarrow S \) is a unital ring homomorphism, then \( \phi(x) \in S^* \).

Proof. (a): Let \( x \in R^* \) and \( \overline{x} \in R \) such that \( x\overline{x} = \overline{x}x = 1 \). If \( y \in R \) such that \( xy = 0 \), then \( y = 1 \cdot y = \overline{x}xy = 0 \); if \( y \in R \) such that \( yx = 0 \), then \( y = y \cdot 1 = yx\overline{x} = 0 \). In consequence, \( x \) is not a zero divisor.

(b): Let \( \overline{x} \in R \) such that \( x\overline{x} = \overline{x}x = 1 \). Then

\[
1 = \phi(1) = \phi(x\overline{x}) = \phi(x)\phi(\overline{x}) = \phi(x)\phi(\overline{x})\phi(x),
\]

proving \( \phi(x) \in S^* \).

Theorem 7.4. Let \( R \) be a commutative ring with unity.

(a) If \( f, g \in R[X] \) with \( f = (f_i)_{i \in \mathbb{N}_0}, g = (g_i)_{i \in \mathbb{N}_0} \), then

\[
\begin{align*}
  \deg(f + g) &= \begin{cases} 
    -\infty \leq \max\{\deg f, \deg g\} & \text{if } f = -g, \\
    \max\{i \in \mathbb{N}_0 : f_i \neq -g_i\} \leq \max\{\deg f, \deg g\} & \text{otherwise},
  \end{cases} \\
  \deg(fg) &\leq \deg f + \deg g.
\end{align*}
\]
If the highest nonzero coefficient of \( f \) or of \( g \) is not a zero divisor (e.g. if this coefficient is an invertible element, cf. Lem. 7.3), then one even has
\[
\deg(fg) = \deg f + \deg g.
\] (7.5c)

(b) \( (R[X], +, \cdot) \) forms a commutative ring with unity, where 1 = \( X^0 \) is the neutral element of multiplication.

Proof. (a) If \( f \equiv 0 \), then \( f + g = g \) and \( fg \equiv 0 \), i.e. the degree formulas hold if \( f \equiv 0 \) or \( g \equiv 0 \). It is also immediate from (7.2) that (7.5a) holds in the remaining case. If \( \deg f = n \in \mathbb{N}_0 \), \( \deg g = m \in \mathbb{N}_0 \), then, for each \( i \in \mathbb{N}_0 \) with \( i > m + n \), we have, for \( k, l \in \mathbb{N}_0 \) with \( k + l = i \) that \( k > m \) or \( l > n \), showing \( (fg)_i = \sum_{k+l=i} f_k g_l = 0 \), proving (7.5b). If \( f_n \) is not a zero divisor, then \( (fg)_{m+n} = f_n g_m \neq 0 \), proving (7.5c).

(b): We already know from [Phi19, Ex. 4.9(e)] that \( (R[X], +) \) forms a commutative group. To verify associativity of multiplication, let \( a, b, c, d, f, g, h \in R[X] \),
\[
\tag{7.5d}
\begin{align*}
  a &:= (a_i)_{i \in \mathbb{N}_0},
  b &:= (b_i)_{i \in \mathbb{N}_0},
  c &:= (c_i)_{i \in \mathbb{N}_0},
  d &:= (d_i)_{i \in \mathbb{N}_0},
  f &:= (f_i)_{i \in \mathbb{N}_0},
  g &:= (g_i)_{i \in \mathbb{N}_0},
  h &:= (h_i)_{i \in \mathbb{N}_0},
\end{align*}
\]
such that \( d := ab \), \( f := bc \), \( g := (ab)c \), \( h := a(bc) \). Then, for each \( i \in \mathbb{N}_0 \),
\[
g_i = \sum_{k+l=i} d_k c_l = \sum_{k+l=i} \sum_{m+n=k} a_m b_n c_l = \sum_{m+n+l=i} a_m b_n c_l
= \sum_{m+k=i} \sum_{n+l=k} a_m b_n c_l = \sum_{m+k=i} a_m f_k = h_i,
\]
proving \( g = h \), as desired. To verify distributivity, let \( a, b, c, d, f, g \in R[X] \) be as before, but this time such that \( d := ab \), \( f := ac \), and \( g := a(b + c) \). Then, for each \( i \in \mathbb{N}_0 \),
\[
g_i = \sum_{k+l=i} a_k (b_l + c_l) = \sum_{k+l=i} a_k b_l + \sum_{k+l=i} a_k c_l = d_i + f_i,
\]
proving \( g = d + f \), as desired. To verify commutativity of multiplication, let \( a, b, c, d \in R[X] \) be as before, but this time such that \( c := ab \), \( d := ba \). Then, for each \( i \in \mathbb{N}_0 \),
\[
c_i = \sum_{k+l=i} a_k b_l = \sum_{k+l=i} b_l a_k = d_i,
\]
proving \( c = d \), as desired. Finally, if \( b := X^0 \), then \( b_0 = 1 \) and \( b_i = 0 \) for \( i > 0 \), yielding, for \( c := ab \) and each \( i \in \mathbb{N}_0 \),
\[
c_i = \sum_{k+l=i} a_k b_l = \sum_{k=0} a_k b_0 = a_i,
\]
showing \( X^0 \) to be neutral and completing the proof. □
Definition 7.5. Let $R, R'$ be rings (with unity). We call $R'$ a ring extension of $R$ if, and only if, there exists a (unital) ring monomorphism $\iota : R \rightarrow R'$ (if $R'$ is a ring extension of $R$, then one might even identify the elements of $R$ and $\iota(R)$ and consider $R$ to be a subset of $R'$). If $R, R'$ are fields, then one also calls $R'$ a field extension of $R$.

Example 7.6. (a) If $R$ is a commutative ring with unity, then $R[X]$ is a ring extension of $R$ via the unital ring monomorphism

$$\iota : R \rightarrow R[X], \quad \iota(r) := rX^0 :$$

Indeed, $\iota$ is unital, since $\iota(1) = X^0$; $\iota$ is a ring homomorphism, since, for each $r, s \in R$, $\iota(r + s) = (r + s)X^0 = rX^0 + sX^0 = \iota(r) + \iota(s)$ and $\iota(rs) = rsX^0 = rX^0 \cdot sX^0 = \iota(r) \iota(s)$; $\iota$ is injective, since, for $r \neq 0$, $\iota(r) = rX^0 \neq 0$.

(b) If $R$ is a ring (with unity) and $n \in \mathbb{N}$, then the matrix ring $\mathcal{M}(n, R)$ (cf. [Phi19, Ex. 7.7(c)]) is a ring extension of $R$ via the (unital) ring monomorphism

$$\iota : R \rightarrow \mathcal{M}(n, R), \quad \iota(r) := \text{diag}(r, \ldots, r) :$$

Indeed, $\iota$ is a ring homomorphism, since, for each $r, s \in R$,

$$\iota(r + s) = \text{diag}(r + s, \ldots, r + s) = \text{diag}(r, \ldots, r) + \text{diag}(s, \ldots, s) = \iota(r) + \iota(s),$$

$$\iota(rs) = \text{diag}(r, \ldots, r) \text{diag}(s, \ldots, s) \overset{\text{Phi19, (7.28)}}{=} \text{diag}(r) \text{diag}(s, \ldots, s),$$

$\iota$ is injective, since, for $r \neq 0$, $\iota(r) = \text{diag}(r, \ldots, r) \neq 0$. If $R$ is a ring with unity, then $\iota$ is unital, since $\iota(1) = \text{Id}_n$.

Proposition 7.7. Let $R$ be a commutative ring with unity without nonzero zero divisors. Then $R[X]$ has no nonzero zero divisors and $(R[X])^* = R^*$.

Proof. Exercise. ■

Example 7.8. Let $R := \mathbb{Z}_4 = \mathbb{Z}/(4\mathbb{Z})$. Then (due to $4 = 0$ in $\mathbb{Z}_4$)

$$(2X^1 + 1X^0)(2X^1 + 1X^0) = 0X^2 + 0X^1 + 1X^0 = X^0,$$

showing $2X^1 + 1X^0 \in (R[X])^* \setminus R^*$, i.e. $(R[X])^* \neq R^*$ can occur if $R$ has nonzero zero divisors. This also provides an example, where the degree formula (7.5c) does not hold.

Definition and Remark 7.9. Let $R$ be a commutative ring with unity and let $R'$ be a ring extension of $R$, where $\iota : R \rightarrow R'$ is a unital ring monomorphism. For each $x \in R'$, the map

$$\epsilon_x : R[X] \rightarrow R', \quad f \mapsto \epsilon_x(f) = \epsilon_x \left( \sum_{i=0}^{\deg f} f_i X^i \right) := \sum_{i=0}^{\deg f} f_i x^i, \quad (7.6)$$

is called the substitution homomorphism or evaluation homomorphism corresponding to $x$ (a typical example, where one wants to use a proper ring extension rather than
$R$, is the substitution of matrices from $\mathcal{M}(n, R)$, $n \in \mathbb{N}$, for $X$): Indeed, if $x \in R'$, then $\epsilon_x$ is unital, since $\epsilon_x(X^0) = x^0 = 1$; $\epsilon_x$ is a ring homomorphism, since, for each $f = \sum_{i=0}^{\deg f} f_i X^i \in R[X]$ and $g = \sum_{i=0}^{\deg g} g_i X^i \in R[X]$, one has

\[
\epsilon_x(f + g) = \sum_{i=0}^{\deg(f+g)} (f_i + g_i) x^i = \sum_{i=0}^{\deg f} f_i x^i + \sum_{i=0}^{\deg g} g_i x^i = \epsilon_x(f) + \epsilon_x(g),
\]

\[
\epsilon_x(f g) = \sum_{i=0}^{\deg(f g)} \left( \sum_{k+l=i} f_k g_l \right) x^i = \sum_{k=0}^{\deg f} \sum_{l=0}^{\deg g} f_k g_l x^{k+l} = \left( \sum_{k=0}^{\deg f} f_k x^k \right) \left( \sum_{l=0}^{\deg g} g_l x^l \right) = \epsilon_x(f) \epsilon_x(g).
\]

Moreover, $\epsilon_x$ is linear, since, for each $\lambda \in R$,

\[
\epsilon_x(\lambda f) = \sum_{i=0}^{\deg f} (\lambda f_i) x^i = \lambda \sum_{i=0}^{\deg f} f_i x^i = \lambda \epsilon_x(f).
\]

We call $x \in R'$ a zero or a root of $f \in R[X]$ if, and only if, $\epsilon_x(f) = 0$.

**Theorem 7.10** (Polynomial Interpolation). Let $F$ be a field and $n \in \mathbb{N}_0$. Given $n + 1$ pairs $(x_0, y_0), (x_1, y_1), \ldots, (x_n, y_n) \in F^2$, such that $x_i \neq x_j$ for $i \neq j$, there exists a unique interpolating polynomial $f \in F[X]$ with $\deg f \leq n$, satisfying

\[
\forall i \in \{0, 1, \ldots, n\} \quad \epsilon_{x_i}(f) = y_i. \tag{7.7}
\]

Moreover, one can identify $f$ as the Lagrange interpolating polynomial, which is given by the explicit formula

\[
f := \sum_{j=0}^{n} y_j L_j, \quad \text{where} \quad L_j := \prod_{i=0}^{n} \frac{X^1 - x_i X^0}{x_j - x_i} \tag{7.8}
\]

(the $L_j$ are called Lagrange basis polynomials).

**Proof.** Since $f = f_0 X^0 + f_1 X^1 + \cdots + f_n X^n$, (7.7) is equivalent to the linear system

\[
\begin{align*}
    f_0 + f_1 x_0 + \cdots + f_n x_0^n &= y_0 \\
    f_0 + f_1 x_1 + \cdots + f_n x_1^n &= y_1 \\
    & \vdots \\
    f_0 + f_1 x_n + \cdots + f_n x_n^n &= y_n,
\end{align*}
\]

for the $n + 1$ unknowns $f_0, \ldots, f_n \in F$. This linear system has a unique solution if, and only if, the determinant

\[
D := \begin{vmatrix}
1 & x_0 & x_0^2 & \ldots & x_0^n \\
1 & x_1 & x_1^2 & \ldots & x_1^n \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_n & x_n^2 & \ldots & x_n^n
\end{vmatrix} \tag{7.10}
\]

is non-zero.
does not vanish. Comparing with (4.31), we see \( D \) to be a Vandermonde determinant and (4.32) yields
\[
D = \prod_{i,j=0 \atop i>j}^n (x_i - x_j).
\]
In particular, \( D \neq 0 \), as we assumed \( x_i \neq x_j \) for \( i \neq j \), proving both existence and uniqueness of \( f \) (bearing in mind that the \( X^i \) form a basis of \( F[X] \)). It remains to identify \( f \) with the Lagrange interpolating polynomial. To this end, set \( g := \sum_{j=0}^n y_j L_j \).

Since, clearly, \( \deg g \leq n \), it merely remains to show \( g \) satisfies (7.7). Since the \( x_i \) are distinct, we obtain
\[
\forall k \in \{0, \ldots, n\} \quad \epsilon_x (g) = \sum_{j=0}^n y_j \epsilon_x (L_j) = \sum_{j=0}^n y_j \delta_{jk} = y_k,
\]
thereby establishing the case.

\[\square\]

**Remark 7.11.** Let \( R \) be a commutative ring with unity. In Def. 4.32, we defined polynomial functions (of several variables). The following Th. 7.12 illuminates the relation between the polynomials of Def. 7.1 and the polynomial functions (of one variable) of Def. 4.32. Let \( \text{Pol}(R) \) denote the set of polynomial functions from \( R \) into \( R \). Clearly, \( \text{Pol}(R) \) is a subring with unity of \( R \) (cf. [Phi19, Ex. 4.41(a)]). If \( R \) is a field, then \( \text{Pol}(R) \) also is a vector subspace of \( R \).

**Theorem 7.12.** Let \( R \) be a commutative ring with unity and consider the map
\[
\phi : R[X] \rightarrow \text{Pol}(R), \quad f \mapsto \phi(f), \quad \phi(f)(x) := \epsilon_x (f).
\]

(a) \( \phi \) is a unital ring epimorphism. If \( R \) is a field, then \( \phi \) is also a linear epimorphism.

(b) If \( R \) is finite, then \( \phi \) is not a monomorphism.

(c) If \( F := R \) is an infinite field, then \( \phi \) is an isomorphism.

**Proof.** (a): If \( f = X^0 \), then \( \phi(f) \equiv 1 \). We also know from Def. and Rem. 7.9 that, for each \( x \in R \), \( \epsilon_x \) is a linear ring homomorphism. Thus, if \( f, g \in R[X] \) and \( \lambda \in R \), then, for each \( x \in R \),
\[
\phi(f + g)(x) = \epsilon_x (f + g) = \epsilon_x (f) + \epsilon_x (g) = (\phi(f) + \phi(g))(x),
\]
\[
\phi(fg)(x) = \epsilon_x (fg) = \epsilon_x (f) \epsilon_x (g) = (\phi(f) \phi(g))(x),
\]
\[
\phi(\lambda f)(x) = \epsilon_x (\lambda f) = \lambda \epsilon_x (f) = (\lambda \phi(f))(x).
\]
Moreover, \( \phi \) is an epimorphism since, if \( P \in \text{Pol}(R) \) with \( P(x) = \sum_{i=0}^n f_i x^i \), where \( f_0, \ldots, f_n \in R \), \( n \in \mathbb{N}_0 \), then \( P = \phi(f) \) with \( f = \sum_{i=0}^n f_i X^i \).

(b): If \( R \) is finite, then \( R^R \) and \( \text{Pol}(R) \subseteq R^R \) are finite, whereas \( R[X] = R^R_{\text{fin}} \) is infinite (also cf. 7.13 below).
(c): If $F$ is infinite, then, for each $n \in \mathbb{N}_0$, there exist distinct points $x_0, \ldots, x_n \in F$. Then Th. 7.10 implies that $f \equiv 0$ is the unique element of $F[X]$ with $\deg f \leq n$ such that $\epsilon_x(f) = 0$ for each $i \in \{0, \ldots, n\}$, implying $f \equiv 0$ to be the unique element of $F[X]$ such that $\phi(f) \equiv 0$. Thus, $\ker \phi = \{0\}$ and $\phi$ is a monomorphism.

**Example 7.13.** If $R$ is a finite commutative ring with unity, then $f := \prod_{\lambda \in R} (X^1 - \lambda) \in R[X] \setminus \{0\}$, but, using $\phi$ of (7.11), $\phi(f) \equiv 0$. For a concrete example, consider the field with two elements, $R := \mathbb{Z}_2 = \{0, 1\}$. Then, $0 \neq f := X^2 + X^1 = X^1(X^1 + 1) \in R[X]$, but, for each $x \in R$, $\phi(f)(x) = x(x + 1) = 0$.

**Remark 7.14.** (a) If $R$ is a finite commutative ring with unity, then Th. 7.12(b) shows that the representation $P(x) = \sum_{i=0}^n a_i x^i$ of $P \in \text{Pol}(R)$ with $a_i \in R$, $n \in \mathbb{N}_0$ is not unique (e.g., over $R := \mathbb{Z}_2$, $P(x) = 1$ and $x^2 + x + 1$ are two different representations of $P \equiv 1$). Thus, while $\deg f$ is always well-defined for $f \in R[X]$ by Def. 7.1, if $R$ is finite, then deg, as given by Def. 4.32, is not a well-defined function on $\text{Pol}(R)$, as it depends on the representation of $P \in \text{Pol}(R)$, rather than on $P$ itself.

(b) If $F$ is a field, then the proof of Th. 7.10 shows that, if $P \in \text{Pol}(F)$ can be written as $P(x) = \sum_{i=0}^n a_i x^i$ with $a_i \in F$, $n \in \mathbb{N}_0$, then the representation is unique if, and only if, $F$ has at least $n + 1$ elements. In particular, the monomial functions $x \mapsto x^i$, $i \in \{0, \ldots, n\}$, are linearly independent if, and only if, $F$ has at least $n + 1$ elements.

Next, we will prove a remainder theorem for polynomials, which can be seen as an analog of the remainder theorem for integers (cf. [Phi19, Th. D.1]):

**Theorem 7.15** (Remainder Theorem). Let $R$ be a commutative ring with unity. Let $g = \sum_{i=0}^d g_i X^i \in R[X]$, $\deg g = d$, where $g_d \in R^*$. Then, for each $f \in R[X]$, there exist unique polynomials $q, r \in R[X]$ such that

\[ f = qg + r \quad \wedge \quad \deg r < d. \tag{7.12} \]

**Proof.** Uniqueness: Suppose $f = qg + r = q'g + r'$ with $q, q', r, r' \in R[X]$ and $\deg r, \deg r' < d$. Then

\[ 0 = (q - q')g + (r - r') \quad \text{via (7.5a), (7.5c)} \quad \Rightarrow \quad \deg(q - q') + \deg g = \deg(r - r'). \]

However, since $\deg(r - r') < d$ and $\deg g = d$, this can only hold for $q = q'$, which, in turn, implies $r = r'$ as well.

Existence: We prove the existence of $q, r \in R[X]$, satisfying (7.12), via induction on $n := \deg f \in \mathbb{N}_0 \cup \{-\infty\}$: If $\deg f < d$, then set $q := 0$ and $r := f$ (this, in particular, takes care of the base case $f \equiv 0$). Now suppose $f = \sum_{i=0}^n f_i X^i$ with $f_n \neq 0$, $n \geq d$. Define

\[ h := f - f_n g_d^{-1} X^{n-d} g. \]
Then the degree of the subtracted polynomial is \( n \), where the coefficient of \( X^n \) is \( f_n g^{-1} d = f_n \), implying \( \deg h < n \). Thus, by the induction hypothesis, there exist \( q_h, r_h \in R[X] \) such that \( h = q_h g + r_h \) and \( \deg r_h < d \). Then

\[
f = h + f_n g^{-1} X^{n-d} g = q_h g + r_h + f_n g^{-1} X^{n-d} g = (q_h + f_n g^{-1} X^{n-d}) g + r_h.
\]

Hence, letting \( q := q_h + f_n g^{-1} X^{n-d} \) and \( r := r_h \) completes the proof. \( \square \)

**Definition 7.16.** Let \( R \) be a commutative ring with unity and \( 1 \neq 0 \).

(a) We call \( R \) **integral** or an **integral domain** if, and only if, \( R \) does not have any nonzero zero divisors.

(b) We call \( R \) **Euclidean** if, and only if, \( R \) is an integral domain and there exists a map \( \deg : R \setminus \{0\} \rightarrow \mathbb{N}_0 \) such that, for each \( f, g \in R \) with \( g \neq 0 \), there exist \( q, r \in R \), satisfying

\[
f = qg + r \quad \land \quad \left( \deg r < \deg g \lor r = 0 \right). \quad (7.13)
\]

The map \( \deg \) is then called a **degree map** or a **Euclidean map** of \( R \).

**Example 7.17.** (a) Every field \( F \) is a Euclidean ring, where we can choose \( \deg : F^* = F \setminus \{0\} \rightarrow \mathbb{N}_0, \deg f = 0 \), as the degree map: If \( g \in F^* \), then, given \( f \in F \), we choose \( q := fg^{-1} \) and \( r := 0 \). Then, clearly, (7.13) holds.

(b) \( \mathbb{Z} \) is a Euclidean ring, where we can choose \( \deg : \mathbb{Z} \setminus \{0\} \rightarrow \mathbb{N}_0, \deg(k) := |k| \), as the degree map: According to [Phi19, Th. D.1], for each \( f, g \in \mathbb{N} \), there exist \( q, r \in \mathbb{N}_0 \) such that \( f = qg + r \) and \( 0 \leq r < g \), also implying \( -f = -qg - r \) with \( |-r| < g \), \( f = -q(-g) + r \), and \( -f = q(-g) - r \), showing that (7.13) can always be satisfied (for \( f = 0 \), merely set \( q := r := 0 \)).

(c) If \( F \) is a field, then \( F[X] \) is a Euclidean ring, where we can choose the degree map as in Def. 7.1: If \( F \) is a field, then, for each \( 0 \neq g \in F[X] \), we can apply Th. 7.15 to obtain (7.13).

**Definition 7.18.** Let \( R \) be a commutative ring with unity.

(a) \( a \subseteq R \) is called an **ideal** in \( R \) if, and only if, the following two conditions hold:

(i) \( (a, +) \) is a subgroup of \( (R, +) \).

(ii) For each \( x \in R \) and each \( a \in a \), one has \( ax \in a \) (which, as \( 1 \in R \), is equivalent to \( aR = a \)).

(b) An ideal \( a \subseteq R \) is called principal if, and only if, there exists \( a \in R \) such that \( a = (a) := aR \).

(c) \( R \) is called principal if, and only if, every ideal in \( R \) is principal.

(d) \( R \) is called a **principal ideal domain** if, and only if, \( R \) is both principal and integral.
Remark 7.19. Let $R$ be a commutative ring with unity and let $a \subseteq R$ be an ideal. Since $(a, +)$ is a subgroup of $(R, +)$ and $a, b \in a$ implies $ab \in a$, $a$ is always a subring of $R$. However, as $1 \in a$ implies $a = R$, $(0) \neq a$ is a subring with unity if, and only if, $a = R$.

Proposition 7.20. Let $R$ be a commutative ring with unity.

(a) $\{0\} = (0)$ and $(1) = R$ are principal ideals of $R$.

(b) If $S$ is a ring and $\phi : R \rightarrow S$ is a ring homomorphism, then $\ker \phi = \phi^{-1}\{0\}$ is an ideal in $R$.

(c) If $a$ and $b$ are ideals in $R$, then $a + b$ is an ideal in $R$.

(d) If $(a_i)_{i \in I}$ is a family of ideals in $R$, $I \neq \emptyset$, then $a := \bigcap_{i \in I} a_i$ is an ideal in $R$ as well.

(e) If $I \neq \emptyset$ is an index set, totally ordered by $\leq$ and $(a_i)_{i \in I}$ is an increasing family of ideals in $R$ (i.e., for each $i, j \in I$ with $i \leq j$, one has $a_i \subseteq a_j$), then $a := \bigcup_{i \in I} a_i$ is an ideal in $R$ as well.

Proof. Exercise.

Theorem 7.21. If $R$ is a Euclidean ring, then $R$ is a principal ideal domain.

Proof. Let $R$ be a Euclidean ring with degree map $\deg : R \setminus \{0\} \rightarrow \mathbb{N}_0$. Moreover, let $a \subseteq R$ be an ideal, $a \neq (0)$. Let $a \in a \setminus \{0\}$ be such that $\deg(a) = \min\{\deg(x) : 0 \neq x \in a\}$.

Then $a = (a)$: Indeed, let $f \in a$. According to (7.13), $f = qa + r$ with $q, r \in R$ and $\deg(r) < \deg(a)$ or $r = 0$. Then $r = f - qa \in a$ and the choice of $a$ implies $r = 0$ and $f = qa \in (a)$, showing $a \subseteq (a)$. As $(a) \subseteq a$ also holds (since $a$ is an ideal), we have $a = (a)$, as desired.

Example 7.22. (a) If $F$ is a field, then $(0)$ and $F = (1)$ are the only ideals in $F$ (in particular, each field is a principal ideal domain): Indeed, if $a$ is an ideal in $F$, $0 \neq a \in a$, and $x \in F$, then $x = xa^{-1}a \in a$.

(b) $\mathbb{Z}$ and $F[X]$ (where $F$ is a field) are principal ideal domains according to Th. 7.21, since we know from Ex. 7.17(b),(c) that both rings are Euclidean rings.

(c) According to Rem. 7.19, a proper subring with unity $S$ of the commutative ring with unity $R$ can never be an ideal in $R$ (and then the unital ring monomorphism $\iota : S \rightarrow R$, $\iota(x) := x$, shows that the subring $S = \text{Im} \phi$ does not need to be an ideal). For example, $\mathbb{Z}$ is a subring of $\mathbb{Q}$, but not an ideal in $\mathbb{Q}$; $\mathbb{Q}$ is a subring of $\mathbb{R}$, but not an ideal in $\mathbb{R}$. 
(d) The ring $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ is principal: $(2) = \{0, 2\}$ and, if $a$ is an ideal in $\mathbb{Z}_4$ with $3 \in a$, then $3 + 3 = 2 \in a$ and $2 + 3 = 1 \in a$, showing $a = \mathbb{Z}_4$. However, since $2 \cdot 2 = 0$, $\mathbb{Z}_4$ is not a principal ideal domain.

(e) The set $A := 2\mathbb{Z} \cup 3\mathbb{Z} \subseteq \mathbb{Z}$ satisfies Def. 7.18(a)(ii): If $k, l \in \mathbb{Z}$, then $kl \cdot 2 \in 2\mathbb{Z} \subseteq A$ and $kl \cdot 3 \in 3\mathbb{Z} \subseteq A$, but $A$ is not an ideal in $\mathbb{Z}$, since $2 + 3 \notin A$. This example also shows that unions of ideals need not be ideals.

(f) The ring $\mathbb{Z}[X]$ is not principal: Let

$$a := \{(f_i)_{i \in \mathbb{N}_0} \in \mathbb{Z}[X] : f_0 \text{ is even}\}.$$

Then, clearly, $a$ is an ideal in $\mathbb{Z}[X]$. Moreover, $2X^0 \in a$ and $X^1 \in \mathbb{Z}[X]$. However, if $f \in a$ such that $2 = 2X^0 \in (f)$, then $f \in \{-2, 2\}$, showing $X^1 \notin (f)$. Thus, the ideal $a$ is not principal.

We now want to show that the analogue of the fundamental theorem of arithmetic [Phi19, Th. D.6] holds in every Euclidean ring (in particular, in $F[X]$, if $F$ is a field) and even in every principal ideal domain. We begin with some preparations:

**Definition 7.23.** Let $R$ be an integral domain.

(a) We call $x, y \in R$ associated if, and only if, there exists $a \in R^*$ such that $x = ay$.

(b) Let $x, y \in R$. We define $x$ to be a divisor of $y$ (and also say that $x$ divides $y$, denoted $x \mid y$) if, and only if, there exists $c \in R$ such that $y = cx$. If $x$ is no divisor of $y$, then we write $x \not\mid y$.

(c) Let $\emptyset \neq M \subseteq R$. We call $d \in R$ a greatest common divisor of the elements of $M$ if, and only if,

$$\forall x \in M \ d \mid x \land \left( \forall r \in R \left( \forall x \in M \ r \mid x \right) \Rightarrow r \mid d \right). \quad (7.14)$$

(d) $0 \neq p \in R \setminus R^*$ is called irreducible if, and only if,

$$\forall x, y \in R \left( p = xy \Rightarrow (x \in R^* \lor y \in R^*) \right). \quad (7.15)$$

Otherwise, $p$ is called reducible.

(e) $0 \neq p \in R \setminus R^*$ is called prime if, and only if,

$$\forall x, y \in R \left( p \mid xy \Rightarrow (p \mid x \lor p \mid y) \right). \quad (7.16)$$

Before looking at some examples, we prove two propositions:
Proposition 7.24. Let $R$ be an integral domain.

(a) Let $\emptyset \neq M \subseteq R$. If $r, d \in R$ are both greatest common divisors of the elements of $M$, then $r, d$ are associated.

(b) If $0 \neq p \in R \setminus R^*$ is prime, then $p$ is irreducible.

(c) If $a_1, \ldots, a_n \in R \setminus \{0\}$, $n \in \mathbb{N}$, and $d \in R$ is such that we have the equality of ideals

$$
(a_1) + \cdots + (a_n) = (d),
$$

(7.17)

then $d$ is a greatest common divisor of $a_1, \ldots, a_n$.

Proof. (a): If $r, d \in R$ are both greatest common divisors of the elements of $M$, then $r \mid d$ and $d \mid r$, i.e. there exist $a, c \in R$ such that $d = cr$ and $r = ad$, implying $d = cad$ and $d(1 - ca) = 0$. Thus, since $R$ does not have any nonzero zero divisors, $1 = ca$ and $c, a \in R^*$, i.e. $r, d$ are associated.

(b): Let $0 \neq p \in R \setminus R^*$ be prime and assume $p = xy = 1 \cdot xy$. Then $p \mid xy$ and, as $p$ is prime, $p \mid x$ or $p \mid y$. By possibly renaming $x, y$, we may assume $p \mid x$, i.e. there exists $c \in R$ with $x = cp$, implying $p = xy = cpy$ and $(1 - cy)p = 0$. Thus, since $R$ does not have any nonzero zero divisors, $1 = cy$ and $c, y \in R^*$, i.e. $p$ is irreducible.

(c): As $a_1, \ldots, a_n \in (d)$, there exist $c_1, \ldots, c_n \in R$ such that $a_i = c_id$ for each $i \in \{1, \ldots, n\}$, showing $d \mid a_i$ for each $i \in \{1, \ldots, n\}$. Now suppose $r \in R$ is such that $r \mid a_i$ for each $i \in \{1, \ldots, n\}$, i.e. there exist $x_1, \ldots, x_n \in R$ with $a_i = x_ir$ for each $i \in \{1, \ldots, n\}$. On the other hand $d \in (a_1) + \cdots + (a_n)$ implies the existence of $s_1, \ldots, s_n \in R$ with $d = \sum_{i=1}^{n} s_ia_i$. Then

$$
d = \sum_{i=1}^{n} s_ia_i = \sum_{i=1}^{n} s_ix_ir = \left(\sum_{i=1}^{n} s_ix_i\right)r,
$$

showing $r \mid d$ and proving (c). \(\blacksquare\)

Proposition 7.25. Let $R$ be a principal ideal domain.

(a) Bézout’s Lemma, cf. [Phil19, Th. D.4]: If $a_1, \ldots, a_n \in R \setminus \{0\}$, $n \in \mathbb{N}$, and $d \in R$ is a greatest common divisor of $a_1, \ldots, a_n$, then (7.17) holds. In particular,

$$
\exists \quad x_1a_1 + \cdots + x_na_n = d,
$$

(7.18)

which is known as Bézout’s identity (usually for $n = 2$). An important special case is that, if $1$ is a greatest common divisor of $a_1, \ldots, a_n$, then

$$
\exists \quad x_1a_1 + \cdots + x_na_n = 1.
$$

(7.19)

(b) Let $0 \neq p \in R \setminus R^*$. Then $p$ is prime if, and only if, $p$ is irreducible.
Proof. (a): Let \( d \) be a greatest common divisor of \( a_1, \ldots, a_n \). Since \( R \) is a principal ideal domain and using Prop. 7.24(c), (7.17) must hold with some greatest common divisor \( d_1 \) of \( a_1, \ldots, a_n \). Then, by Prop. 7.24(a), there exists \( r \in R^* \) such that \( d = r d_1 \), implying \( (d) = (d_1) \), proving (a).

(b): Due to Prop. 7.24(b), it only remains to prove that \( p \) is prime if it is irreducible. Thus, assume \( p \) to be irreducible and let \( x, y \in R \) such that \( p \mid xy \), i.e. \( xy = ap \) with \( a \in R \). If \( p \nmid x \), then 1 is a greatest common divisor \( p, x \) and, according to (7.19), there exist \( r, s \in R \) such that \( rp + sx = 1 \). Then
\[
y = y \cdot 1 = y(rp + sx) = yrp + sx = (yr + sa)p,
\]
showing \( p \mid y \) and \( p \) prime.

Example 7.26. Let \( R \) be an integral domain.

(a) For each \( x \in R \), due to \( x = 1 \cdot x \), \( 1 \mid x \) and \( x \mid x \).

(b) If \( F := R \) is a field, then \( F \setminus F^* = \{0\} \), i.e. \( F \) has neither irreducible elements nor prime elements.

(c) If \( R = \mathbb{Z} \), then \( R^* = \{-1, 1\} \), i.e. \( p \in \mathbb{Z} \) is irreducible if, and only if, \( |p| \) is a prime number in \( \mathbb{N} \) (and, since \( \mathbb{Z} \) is a principal ideal domain by Ex. 7.22(b), \( p \in \mathbb{Z} \) is irreducible if, and only if, it is prime).

(d) For each \( \lambda \in R \), \( X - \lambda = X^1 - \lambda X^0 \in R[X] \) is irreducible due to (7.5c). For \( R = \mathbb{R} \), \( X^2 + 1 \) is irreducible: Otherwise, there exist \( \lambda, \mu \in \mathbb{R} \) and \( X^2 + 1 = (X + \lambda)(X + \mu) \), yielding the contradiction \( 0 = \epsilon_{-\lambda}(X + \lambda)(X + \mu)) = \epsilon_{-\lambda}(X^2 + 1) = \lambda^2 + 1 \).

(e) Suppose
\[
R := \mathbb{Q} + X^1 \mathbb{R}[X] = \{(f_i)_{i \in \mathbb{N}_0} \in \mathbb{R}[X] : f_0 \in \mathbb{Q}\}.
\]
Clearly, \( R \) is a subring of \( \mathbb{R}[X] \). Then \( X = X^1 \in R \) is irreducible, but \( X \) is not prime, since \( X \mid 2X^2 = (\sqrt{2}X)(\sqrt{2}X) \), but \( X \nmid \sqrt{2}X \), since \( \sqrt{2} \notin \mathbb{Q} \). Then, as a consequence of Prop. 7.25(b), \( R \) can not be a principal ideal domain. Indeed, the ideal \( a := (X) + (\sqrt{2}X) \) is not principal in \( R \): Clearly, \( X \) and \( \sqrt{2}X \) are not common multiples of any noninvertible \( f \in R \).

Lemma 7.27. Let \( R \) be a principal ideal domain. Let \( I \neq \emptyset \) be an index set totally ordered by \( \leq \), let \( (a_i)_{i \in I} \) be an increasing family of ideals in \( R \). According to Prop. 7.20(e), we can form the ideal \( a := \bigcup_{i \in I} a_i \). Then there exists \( i_0 \in I \) such that \( a = a_{i_0} \).

Proof. As \( R \) is principal, there exists \( a \in R \) such that \( (a) = a \). Since \( a \in a \), there exists \( i_0 \in I \) such that \( a \in a_{i_0} \), implying \( (a) \subseteq a_{i_0} \subseteq a = (a) \) and establishing the case.

Theorem 7.28 (Existence of Prime Factorization). Let \( R \) be a principal ideal domain. If \( 0 \neq a \in R \setminus R^* \), then there exist prime elements \( p_1, \ldots, p_n \in R, n \in \mathbb{N} \), such that
\[
a = p_1 \cdots p_n.
\]
Proof. Let $S$ be the set of all ideals $(a)$ in $R$ that are generated by elements $0 \neq a \in R \setminus R^*$ that do not have a prime factorization as in (7.20). We need to prove $S = \emptyset$. Seeking a contradiction, assume $S \neq \emptyset$ and note that set inclusion $\subseteq$ provides a partial order on $S$. If $C \neq \emptyset$ is a totally ordered subset of $S$, then, by Prop. 7.20(e), $a := \bigcup_{c \in C} c$ is an ideal in $R$ and, by Lem. 7.27, there exists $c \in C$ such that $a = c$, showing $a \in S$ to provide an upper bound for $C$. Thus, Zorn’s lemma [Phi19, Th. 5.22] applies, yielding a maximal element $m \in S$ (i.e. maximal in $S$ with respect to $\subseteq$). Then there exists $a \in R \setminus R^*$ such that $m = (a)$ and $a$ does not have a prime factorization. In particular, $a$ is not prime, i.e. $a$ must be reducible by Prop. 7.25(b). Thus, there exist $a_1, a_2 \in R \setminus (R^* \cup \{0\})$ such that $a = a_1a_2$. Then $(a) \subseteq (a_1)$ and $(a) \subseteq (a_2)$: Indeed, if $a_1 = ra = ra_1a_2$ with $r \in R$, then $0 = a_1(1-r)a_2$ would imply $a_2 \in R^*$ and analogously for $(a) = (a_2)$. Due to the maximality of $m = (a)$ in $S$, we conclude $(a_1), (a_2) \notin S$. Thus, $a_1, a_2$ both must have prime factorizations, yielding the desired contradiction that $a = a_1a_2$ must have a prime factorization as well.

Remark 7.29. In particular, we obtain from Th. 7.28 that each $k \in \mathbb{Z} \setminus \{-1, 0, 1\}$ and each $f \in F[X]$ with $F$ being a field and $\deg f \geq 1$ has a prime factorization. However, for $R = \mathbb{Z}$ and $R = F[X]$, we can prove the existence of a prime factorization for each $0 \neq a \in R \setminus R^*$ in a simpler way and without making use of Zorn’s lemma: Let $\deg : R \setminus \{0\} \to \mathbb{N}_0$ be the degree map as in Ex. 7.17(b),(c): We conduct the proof via induction on $\deg(a) \in \mathbb{N}$: If $a$ itself is prime, then there is nothing to prove, and this, in particular, takes care of the base case of the induction. If $a$ is not prime, then it is reducible, i.e. $a = a_1a_2$ with $a_1, a_2 \in R \setminus (R^* \cup \{0\})$. In particular, $1 \leq \deg a_1, \deg a_2 < \deg a$. Thus, by induction $a_1, a_2$ both have prime factorizations, implying $a$ to have a prime factorization as well.

Theorem 7.30 (Uniqueness of Prime Factorization). Let $R$ be an integral domain and $x \in R$. Suppose

$$x = p_1 \cdots p_n = a r_1 \cdots r_m, \quad (7.21)$$

where $a \in R^*$, $p_1, \ldots, p_n \in R$, $n \in \mathbb{N}$, are prime and $r_1, \ldots, r_m \in R$, $m \in \mathbb{N}$, are irreducible. Then $m = n$ and there exists a permutation $\pi \in S_n$ such that, for each $i \in \{1, \ldots, n\}$, $r_i$ and $p_{\pi(i)}$ are associated (cf. Def. 7.23(a)).

Proof. The proof is conducted via induction on $n \in \mathbb{N}$. Since $p_1$ is prime and $p_1 | r_1 \cdots r_m$, there exists $i \in \{1, \ldots, m\}$ such that $p_1 | r_i$. Since $r_i$ is irreducible, we must have $r_i = a_1 p_1$ with $a_1 \in R^*$. For $n = 1$ (the induction base case), this yields $m = 1$ and $p_1, r_1$ associated, as desired. For $n > 1$, we have

$$p_2 \cdots p_n = a a_1 r_1 \cdots r_{i-1} r_{i+1} \cdots r_m$$

and we employ the induction hypothesis to obtain $n - 1 = m - 1$ (i.e. $n = m$) and a bijective map $\sigma : \{1, \ldots, n\} \setminus \{i\} \to \{2, \ldots, n\}$ such that, for each $j \in \{2, \ldots, n\}$, $r_j$ and $p_{\sigma(j)}$ are associated. Then

$$\pi : \{1, \ldots, n\} \to \{1, \ldots, n\}, \quad \pi(k) := \begin{cases} 1 & \text{for } k = i, \\ \sigma(k) & \text{for } k \neq i, \end{cases}$$
defines a permutation \( \pi \in S_n \) such that for each \( j \in \{1, \ldots, n\} \), \( r_j \) and \( p_{\pi(j)} \) are associated.

**Corollary 7.31.** If \( R \) is a principal ideal domain (e.g. \( R = \mathbb{Z} \) or \( R = F[X] \), where \( F \) is a field), then each \( 0 \neq a \in R \setminus R^* \) admits a factorization into prime elements, which is unique up to the order of the primes and up to association (rings \( R \) with this property are called factorial, factorial integral domains are called unique factorization domains).

**Proof.** One merely combines Th. 7.28 with Th. 7.30.

**Definition 7.32.** Let \( F \) be a field. We call \( F \) algebraically closed if, and only if, for each \( f \in F[X] \) with \( \deg f \geq 1 \), there exists \( \lambda \in F \) such that \( \epsilon_\lambda(f) = 0 \), i.e. such that \( \lambda \) is a zero of \( f \), as defined in Def. and Rem. 7.9 (cf. Rem. 6.12(e) and Th. 7.35).

**Notation 7.33.** Let \( R \) be a commutative ring with unity. One commonly uses the simplified notation \( X := X^1 \in R[X] \) and \( s := sX^0 \in R[X] \) for each \( s \in R \).

**Proposition 7.34.** Let \( R \) be a commutative ring with unity. For each \( f \in R[X] \) with \( \deg f = n \in \mathbb{N} \) and each \( s \in R \), there exists \( q \in R[X] \) with \( \deg q = n - 1 \) such that

\[
\tag{7.22a}
f = \epsilon_s(f) + (X - s)q = \epsilon_s(f)X^0 + (X^1 - sX^0)q,
\]

where \( \epsilon_s \) is the substitution homomorphism according to Def. and Rem. 7.9. In particular, if \( s \) is a zero of \( f \), then

\[
\tag{7.22b}
f = (X - s)q.
\]

**Proof.** According to Th. 7.15, there exist \( q, r \in R[X] \) such that \( f = q(X - s) + r \) with \( \deg r < \deg(X - s) = 1 \). Thus, \( r \in R \) and \( \deg q = n - 1 \) by (7.5c) (which holds, as \( X - s \) is monic). Applying \( \epsilon_s \) to \( f = q(X - s) + r \) then yields \( \epsilon_s(f) = \epsilon_s(q)(s - s) + r = r \), proving (7.22).

**Theorem 7.35.** Let \( F \) be a field. Then the following statements are equivalent:

(i) \( F \) is algebraically closed.

(ii) For each \( f \in F[X] \) with \( \deg f = n \in \mathbb{N} \), there exists \( c \in F \) and \( \lambda_1, \ldots, \lambda_n \in F \) (not necessarily distinct), such that

\[
\tag{7.23}
f = c \prod_{i=1}^n (X^1 - \lambda_i).
\]

(iii) \( f \in F[X] \) is irreducible if, and only if, \( \deg f = 1 \).

**Proof.** “(i) \( \iff \) (ii)”: If \( F \) is algebraically closed, then (ii) follows from Prop. 7.34 by combining (7.22b) with a straightforward induction. That (ii) implies (i) is immediate.

“(i) \( \iff \) (iii)” : We already noted in Ex. 7.26(d) that each \( X - \lambda \) with \( \lambda \in F \) is irreducible, i.e. each \( sX - \lambda \) with \( s \in F \setminus \{0\} \) is irreducible as well. If \( F \) is algebraically closed and
deg $f > 1$, then (7.22b) shows $f$ to be reducible. Conversely, if (iii) holds, then an induction over $n = \deg f \in \mathbb{N}$ shows each $f \in F[X]$ with $\deg f \in \mathbb{N}$ to have a zero: Indeed, $f = aX + b$ with $a, b \in R, a \neq 0$, has $-ba^{-1}$ as a zero, and, if $\deg f > 1$, then $f$ is reducible, i.e. there exist $g, h \in F[X]$ with $1 \leq \deg g, \deg h < \deg f$ such that $f = gh$. By induction, $g$ and $h$ must have a zero, i.e. $f$ must have a zero as well. ■

**Corollary 7.36.** Let $F$ be a field. If $f \in F[X]$ with $\deg f = n \in \mathbb{N}_0$, then $f$ has at most $n$ zeros. Moreover, there exists $k \in \{0, \ldots, n\}$ and $q \in F[X]$ with $\deg q = n - k$ and such that

$$f = q \prod_{j=1}^{k} (X - \lambda_j), \quad (7.24a)$$

where $q$ does not have any zeros in $F$ and $N := \{\lambda_1, \ldots, \lambda_k\} = \{\lambda \in F : \epsilon_\lambda(f) = 0\}$ is the set of zeros of $f$ ($N = \emptyset$ and $f = q$ is possible, and it can also occur that all $\lambda_j$ in (7.24a) are identical). We can rewrite (7.24a) as

$$f = q \prod_{j=1}^{l} (X - \mu_j)^{m_j}, \quad (7.24b)$$

where $\mu_1, \ldots, \mu_l \in F, l \in \{0, \ldots, k\}$, are the distinct zeros of $f$, and $m_j \in \mathbb{N}$ with $\sum_{j=1}^{l} m_j = k$. Then $m_j$ is called the multiplicity of the zero $\mu_j$ of $f$.

If $F$ is algebraically closed, then $k = n$ and $q \in F$.

**Proof.** If $f \in F[X]$ with $\deg f \leq n \in \mathbb{N}_0$ has $n + 1$ distinct zeros, then the uniqueness part of Th. 7.10 yields $f \equiv 0$ and $\deg f = -\infty < n$. The representation (7.24a) follows from (7.22b) combined with a straightforward induction. The algebraically closed case is immediate from Th. 7.35(ii).

**Example 7.37.** (a) Since $\mathbb{C}$ is algebraically closed by [Phi16a, Th. 8.32], for each $f \in \mathbb{C}[X]$ with $n := \deg f \in \mathbb{N}$, there exist numbers $c, \lambda_1, \ldots, \lambda_n \in \mathbb{C}$ such that

$$f = c \prod_{j=1}^{n} (X - \lambda_j) \quad (7.25)$$

(the $\lambda_1, \ldots, \lambda_n$ are precisely all the zeros of $f$, some or all of which might be identical).

(b) For each $f \in \mathbb{R}[X]$ with $n := \deg f \in \mathbb{N}$, there exist numbers $n_1, n_2 \in \mathbb{N}_0$ and $c, \xi_1, \ldots, \xi_{n_1}, \alpha_1, \ldots, \alpha_{n_2}, \beta_1, \ldots, \beta_{n_2} \in \mathbb{R}$ such that $n = n_1 + 2n_2$ and

$$n = n_1 + 2n_2, \quad (7.26a)$$

and

$$f = c \prod_{j=1}^{n_1} (X - \xi_j) \prod_{j=1}^{n_2} (X^2 + \alpha_jX + \beta_j) : (7.26b)$$
Indeed, if \( f \) has only real coefficients, then we can take complex conjugates to obtain, for each \( \lambda \in \mathbb{C} \),

\[
\epsilon_\lambda(f) = 0 \Rightarrow \overline{\epsilon_\lambda(f)} = \epsilon_{\overline{\lambda}}(f) = 0,
\]

showing that the nonreal zeros of \( f \) (if any) must occur in conjugate pairs. Moreover,

\[
(X - \lambda)(X - \overline{\lambda}) = X^2 - (\lambda + \overline{\lambda})X + \lambda\overline{\lambda} = X^2 - 2X \operatorname{Re} \lambda + |\lambda|^2,
\]

showing that (7.25) implies (7.26). This also shows that, if \( f \in \mathbb{R}[X] \) is irreducible, then \( \deg f \in \{1, 2\} \).

In [Phi19, Ex. 4.38], we saw how to obtain the field of rational numbers \( \mathbb{Q} \) from the ring of integers \( \mathbb{Z} \). The same construction actually still works if \( \mathbb{Z} \) is replaced by an arbitrary integral domain \( R \), resulting in the so-called field of fractions of \( R \) (in the following section, we will use the field of fractions of \( F[X] \) in the definition of the characteristic polynomial of \( A \in \mathcal{L}(V,V) \), where \( V \) is a vector space over \( F \)). This gives rise to the following Th. 7.38.

**Theorem 7.38.** Let \( R \) be an integral domain. One defines the field of fractions\(^6\) \( F \) of \( R \) as the quotient set \( F := R / \sim \) with respect to the following equivalence relation \( \sim \) on \( R \times (R \setminus \{0\}) \), where the relation \( \sim \) on \( R \times (R \setminus \{0\}) \) is defined by

\[
(a, b) \sim (c, d) :\Leftrightarrow ad = bc,
\]

where, as usual, we will write

\[
\frac{a}{b} := a/b := [(a, b)]
\]

for the equivalence class of \((a, b)\) with respect to \( \sim \). Addition on \( F \) is defined by

\[
+: F \times F \rightarrow F, \quad \left( \frac{a}{b}, \frac{c}{d} \right) \mapsto \frac{a}{b} + \frac{c}{d} := \frac{ad + bc}{bd}.
\]

Multiplication on \( F \) is defined by

\[
\cdot : F \times F \rightarrow F, \quad \left( \frac{a}{b}, \frac{c}{d} \right) \mapsto \frac{a}{b} \cdot \frac{c}{d} := \frac{ac}{bd}.
\]

Then \((F, +, \cdot)\) does, indeed, form a field, where \( 0/1 \) and \( 1/1 \) are the neutral elements with respect to addition and multiplication, respectively, \((-a/b)\) is the additive inverse to \( a/b \), whereas \( b/a \) is the multiplicative inverse to \( a/b \) with \( a \neq 0 \). The map

\[
\iota : R \rightarrow F, \quad \iota(k) := \frac{k}{1},
\]

is a unital ring monomorphism and it is customary to identify \( R \) with \( \iota(R) \), just writing \( k \) instead of \( \frac{k}{1} \).

\(^6\)Caveat: The field of fractions of \( R \) should not be confused with the quotient field or factor field of \( R \) with respect to a maximal ideal \( m \) in \( R \) — this is a different construction, leading to different objects (e.g., for \( R = \mathbb{Z} \) to the finite fields \( \mathbb{Z}_p \), where \( p \in \mathbb{N} \) is prime).
Proof. Exercise. ■

Example 7.39. (a) \( \mathbb{Q} \) is the field of fractions of \( \mathbb{Z} \).

(b) If \( R \) is an integral domain, then we know from Prop. 7.7 that \( R[X] \) is an integral domain as well. The field of fractions of \( R[X] \) is denoted by \( R(X) \) and is called the field of rational fractions over \( R \).

8 Characteristic Polynomial, Minimal Polynomial

We will now apply the theory of polynomials to further study linear endomorphisms on finite-dimensional vector spaces. The starting point is Th. 6.9(a), which states that the eigenvalues of \( A \in \mathcal{L}(V,V) \) are precisely the zeros of its characteristic polynomial function \( p_A \). In order to make the results of the previous section available, instead of associating a polynomial function with \( A \), we now need to associate an actual polynomial. The idea is to replace \( t \mapsto \det(t \text{Id} - A) \) with \( \det(X \text{Id}_n - M_A) \), where \( M_A \) is the matrix of \( A \) with respect to an ordered basis of \( V \). If \( V \) is a vector space over the field \( F \), then the entries of the matrix \( X \text{Id}_n - M_A \) are elements of the ring \( F[X] \). However, we defined determinants only for matrices with entries in fields. Thus, to make the following definition consistent with our definition of determinants, we consider the elements of \( X \text{Id}_n - M_A \) to be elements of \( F(X) \), the field of rational fractions over \( F \) (cf. Ex. 7.39(b)):

**Definition 8.1.** Let \( V \) be a vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \), and \( A \in \mathcal{L}(V,V) \). Moreover, let \( B \) be an ordered basis of \( A \) and let \( M_A \in \mathcal{M}(n, F) \) be the matrix of \( A \) with respect to \( B \). Since \( F(X) \) is a field extension of \( F \), we may consider \( M_A \) as an element of \( \mathcal{M}(n, F(X)) \). We define

\[
\chi_A := \det(X \text{Id}_n - M_A) \in F[X]
\]

to be the characteristic polynomial of \( A \).

**Proposition 8.2.** Let \( V \) be a vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \), and \( A \in \mathcal{L}(V,V) \).

(a) The characteristic polynomial \( \chi_A \) is well-defined by Def. 8.1, i.e. if \( B_1 \) are \( B_2 \) are ordered bases of \( V \) and \( M_1, M_2 \) are the matrices of \( A \) with respect to \( B_1, B_2 \), respectively, then

\[
\chi_1 := \det(X \text{Id}_n - M_1) = \chi_2 := \det(X \text{Id}_n - M_2).
\]

(b) If, for each \( t \in F \), \( \epsilon_t : F[X] \to F \) is the substitution homomorphism according to Def. and Rem. 7.9, then

\[
p_A : F \to F, \quad p_A(t) = \epsilon_t(\chi_A),
\]

is the relation between the characteristic polynomial \( \chi_A \) and the characteristic polynomial function \( p_A \) of \( A \).
(c) The spectrum $\sigma(A)$ is precisely the set of zeros of $\chi_A$.

Proof. (a): Let $T \in \text{GL}_n(F)$ be such that $M_2 = T^{-1}M_1T$. Then

$$\chi_2 = \det(X \text{ Id}_n - T^{-1}M_1T) = \det(T^{-1}(X \text{ Id}_n - M_1)T) = (\det T^{-1})\chi_1(\det T) = \chi_1,$$

proving (a).

(b) is immediate from comparing Def. 8.1 and Def. 6.11.

(c) follows by combining (b) with Th. 6.9(a).

\[\blacksquare\]

Remark 8.3. Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, and $A \in \mathcal{L}(V,V)$. The statements of Rem. 6.12 in regard to the characteristic polynomial function $p_A$ also apply correspondingly to $\chi_A$. In particular, if $B$ is an ordered basis of $V$, the matrix $(a_{ji}) \in \mathcal{M}(n,F)$ represents $A$ with respect to $B$, and we let $(c_{ji}) := (X \text{ Id}_n - (a_{ji}))$, then

$$\chi_A = \det (X \text{ Id}_n - (a_{ji})) = \prod_{i=1}^{n} (X - a_{ii}) + \sum_{\pi \in S_n \setminus \{\text{Id}\}} \text{sgn}(\pi) \prod_{i=1}^{n} c_{i\pi(i)} \tag{8.1}$$

shows $\chi_A$ to be monic with $\deg \chi_A = n$. The argument of Rem. 6.12(d) shows that every monic polynomial in $F[X]$ of degree $n$ occurs as a characteristic polynomial.

Theorem 8.4. Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, and $A \in \mathcal{L}(V,V)$. Then there exists an ordered basis $B$ of $V$ such that the matrix $M$ of $A$ with respect to $B$ is triangular if, and only if, there exist $\lambda_1, \ldots, \lambda_l \in F$, $l \in \mathbb{N}$, and $n_1, \ldots, n_l \in \mathbb{N}$ with

$$\sum_{i=1}^{l} n_i = n \quad \land \quad \chi_A = \prod_{i=1}^{l} (X - \lambda_i)^{n_i}. \tag{8.2}$$

In this case, $\sigma(A) = \{\lambda_1, \ldots, \lambda_l\}$,

$$\forall \quad i \in \{1, \ldots, l\} \quad n_i = m_\alpha(\lambda_i) \tag{8.3}$$

(i.e. the algebraic multiplicity of $\lambda_i$ is precisely the multiplicity of $\lambda_i$ as a zero of $\chi_A$), and one can choose $B$ such that $M$ has the upper diagonal form

$$M = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots \\ 0 & \cdots & \cdots & \lambda_l \end{pmatrix}, \tag{8.4}$$

where each $\lambda_i$ occurs precisely $n_i$ times on the diagonal.
Proof. If there exists a basis $B$ of $V$ such that the matrix $M = (m_{ji})$ of $A$ with respect to $B$ is triangular, then

$$
\chi_A \overset{\text{Def. 8.1}}{=} \det(X \text{ Id}_n - M) \overset{\text{Cor. 4.26}}{=} \prod_{i=1}^{n}(X - m_{ii}).
$$

Combining factors, where the $m_{ii}$ are equal, yields (8.2). For the converse, we assume (8.2) and prove the existence of the basis $B$ such that $M$ has the from of (8.4) via induction on $n$. For $n = 1$, there is nothing to prove. Thus, let $n > 1$. Then $\lambda_1$ must be an eigenvector of $A$ with some eigenvector $0 \neq v_1 \in V$. Then, if $B_1 := (v_1, \ldots, v_n)$ is an ordered basis of $V$, the matrix $M_1$ of $A$ with respect to $B_1$ has the block form

$$
M_1 = \begin{pmatrix}
\lambda_1 & * \\
0 & N
\end{pmatrix}, \quad N \in M(n-1, F).
$$

According to Th. 4.25, we obtain

$$
\prod_{i=1}^{l}(X - \lambda_i)^{n_i} = \chi_A = (X - \lambda_1) \chi_N \Rightarrow \chi_N = (X - \lambda_1)^{n_1-1} \prod_{i=2}^{l}(X - \lambda_i)^{n_i}.
$$

Let $U := \langle \{v_1\} \rangle$ and $W := V/\langle \{v_1\} \rangle$. Then $\dim W = n - 1$ and, by [Phi9, Cor. 6.13(a)], $B_W := (v_2 + U, \ldots, v_n + U)$ is an ordered basis of $W$. Let $A_1 \in \mathcal{L}(W, W)$ be such that, with respect to $B_W$, $A_1$ has the matrix $N$. Then, by induction hypothesis, there exists an ordered basis $C_W = (w_2 + U, \ldots, w_n + U)$ of $W$ $(w_2, \ldots, w_n \in V)$, such that, with respect to $C_W$, the matrix $N_1 \in M(n-1, F)$ of $A_1$ has the form (8.4), except that $\lambda_1$ occurs precisely $n_1 - 1$ times on the diagonal. That $N_1$ is the matrix of $A_1$ means, for $N_1 = (\nu_{ji})_{(j,i) \in \{2, \ldots, n\}^2}$ that

$$
\forall_{i \in \{2, \ldots, n\}} A_1(w_i + U) = \sum_{j=2}^{n} \nu_{ji}(w_j + U).
$$

Then, by [Phi9, Cor. 6.13(b)], $B := (v_1, w_2, \ldots, w_n)$ is an ordered basis of $V$ and, with respect to $B$, the matrix $M$ of $A$ has the form (8.4): According to [Phi9, Th. 7.14], there exists an $(n - 1) \times (n - 1)$ transition matrix $T_1 = (t_{ji})_{(j,i) \in \{2, \ldots, n\}^2}$ such that $N_1 = T_1^{-1}NT_1$ and

$$
\forall_{i \in \{2, \ldots, n\}} w_i = \sum_{j=2}^{n} t_{ji}v_j, \quad w_i + U = \sum_{j=2}^{n} t_{ji}(v_j + U),
$$

implying

$$
M = \begin{pmatrix}
1 & 0 & 0 \\
0 & T_1^{-1} & * \\
0 & 0 & N
\end{pmatrix}
\begin{pmatrix}
\lambda_1 & * \\
0 & N
\end{pmatrix}
\begin{pmatrix}
1 & 0 \\
0 & T_1
\end{pmatrix}
= \begin{pmatrix}
\lambda_1 & * \\
0 & N_1
\end{pmatrix}.
$$

It remains to verify (8.3). Letting $w_1 := v_1$, we have $B = (w_1, \ldots, w_n)$. For each
$k \in \{1, \ldots, n_1\}$ and the standard column basis vector $e_k$, we obtain

$$(M - \lambda_1 \text{Id}_n) e_k = \left( (m_{ji}) - \lambda_1 (\delta_{ji}) \right) e_k = \begin{pmatrix} m_{1k} \\ \vdots \\ m_{k-1,k} \\ \lambda - \lambda_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Rightarrow (A - \lambda_1 \text{Id}) w_k = \sum_{\alpha=1}^{k-1} m_{\alpha k} w_\alpha,$$

showing $(A - \lambda_1 \text{Id}) w_k \in \langle \{w_1, \ldots, w_{k-1}\} \rangle$ and $\{w_1, \ldots, w_{n_1}\} \subseteq \ker(A - \lambda_1 \text{Id})^{n_1}$. On the other hand, for each $k \in \{n_1 + 1, \ldots, n\}$, we obtain

$$(M - \lambda_1 \text{Id}_n) e_k = \begin{pmatrix} m_{1k} \\ \vdots \\ m_{k-1,k} \\ \lambda - \lambda_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \Rightarrow (A - \lambda_1 \text{Id}) w_k = (\lambda - \lambda_1) w_k + \sum_{\alpha=1}^{k-1} m_{\alpha k} w_\alpha,$$

where $\lambda \in \{\lambda_2, \ldots, \lambda_l\}$, showing $w_k \notin \ker(A - \lambda_1 \text{Id})^{N}$ for each $N \in \mathbb{N}$. Thus,

$$n_1 = \dim \ker(A - \lambda_1 \text{Id})^{n_1} = \dim \ker(A - \lambda_1 \text{Id})^{r(\lambda_1)} = m_\alpha(\lambda_1),$$

where $r(\lambda_1)$ is as defined in Rem. 6.13. Now note that $\lambda_1$ was chosen arbitrarily is the above argument. The same argument shows the existence of a basis $B'$ such that $\lambda_i$, $i \in \{1, \ldots, l\}$, appears in the upper left block of $M$. In particular, we obtain $n_i = m_\alpha(\lambda_i)$ for each $i \in \{1, \ldots, l\}$.

**Corollary 8.5.** Let $V$ be a vector space over the algebraically closed field $F$, $\dim V = n \in \mathbb{N}$, and $A \in \mathcal{L}(V, V)$. Then there exists an ordered basis $B$ of $V$ such that the matrix $M$ of $A$ with respect to $B$ is triangular.

**Proof.** This is immediate from combining Th. 8.4 with Th. 7.35(ii).}

**Theorem 8.6.** Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, and $A \in \mathcal{L}(V, V)$. There exists a unique monic polynomial $0 \neq \mu_A \in F[X]$ (called the minimal polynomial of $A$), satisfying the following two conditions:

(i) $\epsilon_A(\mu_A) = 0$, where $\epsilon_A : F[X] \rightarrow \mathcal{L}(V, V)$ is the substitution homomorphism according to Def. and Rem. 7.9 (noting $\mathcal{L}(V, V)$ to be a ring extension of $F$ via the unital ring monomorphism $\iota : F \rightarrow \mathcal{L}(V, V)$, $\iota(a) := a \text{Id}$).

(ii) For each $f \in F[X]$ such that $\epsilon_A(f) = 0$, $\mu_A$ is a divisor of $f$, i.e. $\mu_A | f$.  

8 CHARACTERISTIC POLYNOMIAL, MINIMAL POLYNOMIAL

Proof. Let \( \mathfrak{a} := \{ f \in F[X] : \epsilon_A(f) = 0 \} \). Clearly, \( \mathfrak{a} \) is an ideal in \( F[X] \), and, thus, as \( F[X] \) is a principal ideal domain, there exists \( g \in F[X] \) such that \( \mathfrak{a} = (g) \). Clearly, \( g \) satisfies both (i) and (ii). We need to show \( g \neq 0 \), i.e. \( \mathfrak{a} \neq \{0\} \). To this end, note that, since \( \dim \mathcal{L}(V,V) = n^2 \), the \( n^2 + 1 \) maps \( \text{Id}, A, A^2, \ldots, A^{n^2} \in \mathcal{L}(V,V) \) must be linearly dependent, i.e. there exist \( c_0, \ldots, c_{n^2} \in F \), not all 0, such that

\[
0 = \sum_{i=0}^{n^2} c_i A^i,
\]

showing \( 0 \neq f := \sum_{i=0}^{n^2} c_i X^i \in \mathfrak{a} \). If \( h \in F[X] \) also satisfies (i) and (ii), then \( h \mid g \) and \( g \mid h \), implying \( g, h \) to be associated. In consequence, \( \mu_A \) is the unique monic such element of \( F[X] \). \( \blacksquare \)

Remark 8.7. Let \( F \) be a field.

(a) We extend Def. 6.17 to the characteristic polynomial and to the minimal polynomial: Let \( n \in \mathbb{N} \). Consider the vector space \( V := F^n \) over the field \( F \). If \( M \in \mathcal{M}(n, F) \), then \( \chi_M \) and \( \mu_M \) denote the characteristic polynomial and the minimal polynomial of the linear map \( A_M \) that \( M \) represents with respect to the standard basis of \( F^n \).

(b) Let \( V \) be a vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \). As \( \mathcal{M}(n, F) \) is a ring extension of \( F \), we can plug \( M \in \mathcal{M}(n, F) \) into elements of \( F[X] \). Moreover, if \( f \in F[X], A \in \mathcal{L}(V,V) \) and \( M \in \mathcal{M}(n, F) \) represents \( A \) with respect to a basis \( B \) of \( V \), then, due to [Phi19, Th. 7.10(a)], \( \epsilon_M(f) \) represents \( \epsilon_A(f) \) with respect to \( B \).

Example 8.8. Let \( F \) be a field. Consider

\[
M := \begin{pmatrix}
0 & 0 & 1 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix} \in \mathcal{M}(3, F)
\]

We claim that the minimal polynomial is \( \mu_M = X^2 \): Indeed, \( M^2 = 0 \) implies \( \epsilon_M(X^2) = 0 \), and, if \( f = \sum_{i=0}^{n} f_i X^i \in F[X], n \in \mathbb{N} \), then \( \epsilon_M(f) = f_0 \text{Id}_3 + f_1 M \). Thus, if \( \epsilon_M(f) = 0 \), then \( f_0 = f_1 = 0 \), implying \( X^2 \mid f \) and showing \( \mu_M = X^2 \).

Theorem 8.9 (Cayley-Hamilton). Let \( V \) be a vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \), and \( A \in \mathcal{L}(V,V) \). If \( \chi_A \) and \( \mu_A \) denote the characteristic and the minimal polynomial of \( A \), respectively, then the following statements hold true:

(a) \( \chi_A(A) := \epsilon_A(\chi_A) = 0 \).

(b) \( \mu_A \mid \chi_A \) and, in particular, \( \deg \mu_A \leq \deg \chi_A = n \).

(c) \( \lambda \in \sigma(A) \) if, and only if, \( \mu_A(\lambda) := \epsilon_A(\mu_A) = 0 \), i.e. the eigenvalues of \( A \) are precisely the zeros of the minimal polynomial \( \mu_A \).

(d) If \#\sigma(A) = n \ (i.e. if \( A \) has \( n \) distinct eigenvalues), then \( \chi_A = \mu_A \).
Proof. (a): Let $B$ be an ordered basis of $V$ and let $(m_{ji}) := M_A$ be the matrix of $A$ with respect to $B$. Moreover, let $N$ be the adjugate matrix of $X \text{Id}_n - M_A$, i.e., up to factors of $\pm 1$, $N$ contains the determinants of the $(n-1) \times (n-1)$ submatrices of $X \text{Id}_n - M_A$. According to Th. 4.29(a), we then have

$$\chi_A \text{Id}_n = \det(X \text{Id}_n - M_A) \text{Id}_n = N (X \text{Id}_n - M_A).$$

(8.5)

Since $X \text{Id}_n - M_A$ contains only entries of degree at most 1 ($\deg(X - m_{ii}) = 1$, all other entries having degree 0 or degree $-\infty$), each entry $n_{ji}$ of $N$ has degree at most $n - 1$, i.e.

$$\forall (j,i) \in \{1, \ldots, n\}^2 \quad \exists b_{0,j,i}, \ldots, b_{n-1,j,i} \in F \quad n_{ji} = \sum_{k=0}^{n-1} b_{k,j,i} X^k.$$ 

If, for each $k \in \{0, \ldots, n - 1\}$, we let $B_k := (b_{k,j,i}) \in \mathcal{M}(n, F)$, then $N = \sum_{k=0}^{n-1} B_k X^k$. Plugging this into (8.5) yields

$$\chi_A \text{Id}_n = (B_0 + B_1 X + \cdots + B_{n-1} X^{n-1}) (X \text{Id}_n - M_A)$$

$$= -B_0 M_A + (B_0 - B_1 M_A) X + (B_1 - B_2 M_A) X^2$$

$$+ \cdots + (B_{n-2} - B_{n-1} M_A) X^{n-1} + B_{n-1} X^n.$$ 

(8.6)

Writing $\chi_A = X^n + \sum_{i=0}^{n-1} a_i X^i$ with $a_0, \ldots, a_{n-1} \in F$, the coefficients in front of each $X^i$ in (8.6) must agree: Indeed, in each entry of the respective matrix, we have an element of $F[X]$ and, in each entry, the coefficients of $X^i$ must agree (due to the linear independence of the $X^i$) — hence, the matrix coefficients of $X^i$ in (8.6) must agree as well. This yields

$$a_0 \text{Id}_n = -B_0 M_A,$$

$$a_1 \text{Id}_n = B_0 - B_1 M_A,$$

$$\vdots$$

$$a_{n-1} \text{Id}_n = B_{n-2} - B_{n-1} M_A,$$

$$\text{Id}_n = B_{n-1}.$$ 

Thus, $\epsilon_{M_A}(\chi_A)$ turns out to be the telescoping sum

$$\epsilon_{M_A}(\chi_A) = (M_A)^n + \sum_{i=0}^{n-1} a_i (M_A)^i = \sum_{i=0}^{n-1} a_i \text{Id}_n (M_A)^i + \text{Id}_n (M_A)^n$$

$$= -B_0 M_A + \sum_{i=1}^{n-1} (B_{i-1} - B_i M_A) (M_A)^i + B_{n-1} (M_A)^n = 0.$$ 

As $\phi : \mathcal{L}(V, V) \rightarrow \mathcal{M}(n, F)$, $\phi(A) := M_A$, is a ring isomorphism by [Phi19, Th. 7.10(a)], we also obtain $\epsilon_A(\chi_A) = \phi^{-1}(\epsilon_{M_A}(\chi_A)) = 0$, thereby proving (a).

(b) is an immediate consequence of (a) in combination with Th. 8.6(ii).
(c): Suppose \( \lambda \in \sigma(A) \) and let \( 0 \neq v \in V \) be a corresponding eigenvector. Also let \( m_0, \ldots, m_l \in F \) be such that \( \mu_A = \sum_{i=0}^{l} m_i X^i, \ l \in \mathbb{N} \). Then we compute

\[
0 = (\epsilon_A(\mu_A)) v = \left( \sum_{i=0}^{l} m_i A^i \right) v = \sum_{i=0}^{l} m_i (A^i v) = \sum_{i=0}^{l} m_i (\lambda^i v)
\]

showing \( \epsilon_\lambda(\mu_A) = 0 \). Conversely, if \( \lambda \in F \) is such that \( \epsilon_\lambda(\mu_A) = 0 \), then (b) implies \( \epsilon_\lambda(\chi_A) = 0 \), i.e. \( \lambda \in \sigma(A) \) by Th. 8.2(c).

(d): If \( \#\sigma(A) = n \), then \( \mu_A \) has \( n \) distinct zeros by (c), implying \( \deg \mu_A = n \). Since \( \mu_A \) is monic and \( \mu_A | \chi_A \) by (b), we have \( \mu_A = \chi_A \) as claimed. \( \blacksquare \)

Example 8.10. Let \( F \) be a field. If \( M := \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}(2, F) \), then \( \chi_M = X^2 \). Since \( \mu_M | \chi_M \) and \( \epsilon_M(X) = M \neq 0 \), we must have \( \mu_M = \chi_M \). On the other hand, if \( N := \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{M}(2, F) \), then \( \chi_N = X^2 \) and \( \mu_N = X \). Since \( N \) is diagonalizable, \( M \) is not diagonalizable (cf. Ex. 6.5(d) and Ex. 6.16), but \( \chi_M = \chi_N \); we can, in general, not decide diagonalizability merely by looking at the characteristic polynomial. However, we will in Th. 8.14 below that the \textit{minimal} polynomial does allow one to decide diagonalizability.

Caveat 8.11. One has to use care when substituting matrices and endomorphisms into polynomials: For example, one must be aware that in the expression \( X \text{Id}_n - M \), the polynomial \( X \) is a \textit{scalar}. Thus, when substituting a matrix \( B \in \mathcal{M}(n, F) \) for \( X \), one must \textit{not} use matrix multiplication between \( B \) and \( \text{Id}_n \): For example, for \( n = 2 \), \( B := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \), and \( M := \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \), we obtain \( B^2 = B \) and

\[
\epsilon_B(\det(X \text{Id}_n - M)) = \epsilon_B(X^2) = B^2 \neq 0 = \det B = \det(B - M).
\]

The following result further clarifies the relation between \( \chi_A \) and \( \mu_A \):

Proposition 8.12. Let \( V \) be a vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \), and \( A \in \mathcal{L}(V, V) \).

(a) One has \( \chi_A | (\mu_A)^n \) and, in particular, each irreducible factor of \( \chi_A \) must be an irreducible factor of \( \mu_A \).

(b) There exists an ordered basis \( B \) of \( V \) such that the matrix \( M \) of \( A \) with respect to \( B \) is triangular if, and only if, there exist \( \lambda_1, \ldots, \lambda_l \in F, l \in \mathbb{N} \), and \( n_1, \ldots, n_l \in \mathbb{N} \) with

\[
\mu_A = \prod_{i=1}^{l} (X - \lambda_i)^{n_i}.
\]
Proof. (a): Let $M \in \mathcal{M}(n,F)$ be a matrix representing $A$ with respect to some ordered basis of $V$. Let $G$ be an algebraically closed field with $F \subseteq G$ (cf. Rem. 6.12(e)). We can consider $M$ as an element of $M \in \mathcal{M}(n,G)$ and, then, $\sigma(M)$ is precisely the set of zeros of both $\chi_M = \chi_A$ and $\mu_A$ in $G$. As $G$ is algebraically closed, for each $\lambda \in \sigma(M)$, there exists $m_\lambda, n_\lambda \in \mathbb{N}$ such that $m_\lambda \leq n_\lambda \leq n$ and
\[
\mu_A = \prod_{\lambda \in \sigma(M)} (X - \lambda)^{m_\lambda} | \chi_A = \prod_{\lambda \in \sigma(M)} (X - \lambda)^{n_\lambda}.
\]
Letting $q := \prod_{\lambda \in \sigma(M)} (X - \lambda)^{n_\lambda - m_\lambda}$, we have $q \in G[X]$ as well as $q = (\mu_A)^n (\chi_A)^{-1}$, i.e. $q \chi_A = (\mu_A)^n$, proving $\chi_A | (\mu_A)^n$ (in both $G[X]$ and $F[X]$, since $q = (\mu_A)^n (\chi_A)^{-1} \in F(X) \cap G[X] = F[X]$).

(b) follows by combining (a) with Th. 8.4. ■

**Theorem 8.13.** Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, and $A \in \mathcal{L}(V,V)$. Suppose the minimal polynomial $\mu_A$ can be written in the form $\mu_A = g_1 \cdots g_l$, $l \in \mathbb{N}$, where $g_1, \ldots, g_l \in F[X]$ are such that, whenever $i \neq j$, then 1 is a greatest common divisor of $g_i$ and $g_j$. Then
\[
V = \bigoplus_{i=1}^{l} \ker g_i(A) = \bigoplus_{i=1}^{l} \ker \epsilon_A(g_i).
\]

Proof. Define
\[
\forall_{i \in \{1, \ldots, l\}} \quad h_i := \prod_{\kappa=1, \kappa \neq i}^{l} g_k.
\]
Then 1 is a greatest common divisor of $h_1, \ldots, h_l$: Indeed, as 1 is a greatest common divisor of $g_i$ and $g_j$ for $i \neq j$, the sets of prime factors of the $g_i$ must all be disjoint. Thus, if $f \in F[X]$ is a divisor of $h_i$, $i \in \{1, \ldots, l\}$, then $f$ does not share a prime factor with $g_i$. If $f | h_i$ holds for each $i \in \{1, \ldots, l\}$, then $f$ does not share a prime factor with any $g_i$, implying $f \in F \setminus \{0\}$, i.e. 1 is a greatest common divisor of $h_1, \ldots, h_l$. In consequence, (7.19) implies
\[
\exists_{f_1, \ldots, f_l \in F[X]} \quad 1 = \sum_{i=1}^{l} f_i h_i
\]
and
\[
\forall_{v \in V} \quad v = \text{Id} v = \epsilon_A(1)v = \sum_{i=1}^{l} \epsilon_A(f_i) \epsilon_A(h_i)v.
\]
(8.7)
We verify that, for each $i \in \{1, \ldots, l\}$, $\epsilon_A(f_i) \epsilon_A(h_i)v \in \ker \epsilon_A(g_i)$: Indeed, since $g_i h_i = \mu_A$, one has
\[
\epsilon_A(g_i) \epsilon_A(f_i) \epsilon_A(h_i)v = \epsilon_A(f_i) \epsilon_A(\mu_A)v = 0.
\]
Thus, (8.7) proves $\sum_{i=1}^{l} \ker \epsilon_A(g_i)$. According to Prop. 5.2(iii), it remains to show

$$\forall \ i \in \{1, \ldots, l\} \ U := \ker \epsilon_A(g_i) \cap \sum_{j \in \{1, \ldots, l\} \setminus \{i\}} \ker \epsilon_A(g_j) = \{0\}.$$ 

To this end, fix $i \in \{1, \ldots, l\}$ and note $\epsilon_A(g_i)(U) = \{0\} = \epsilon_A(h_i)(U)$. On the other hand, 1 is a greatest common divisor of $g_i$, $h_i$, i.e. (7.19) provides $r_i, s_i \in F[X]$ such that $1 = r_i g_i + s_i h_i$, yielding

$$\{0\} = (\epsilon_A(r_i) \epsilon_A(g_i) + \epsilon_A(s_i) \epsilon_A(h_i))(U) = \epsilon_A(1)(U) = \text{Id}(U) = U,$$

thereby completing the proof. □

**Theorem 8.14.** Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, and $A \in \mathcal{L}(V, V)$. Then $A$ is diagonalizable if, and only if, there exist distinct $\lambda_1, \ldots, \lambda_l \in F$, $l \in \mathbb{N}$, such that $\mu_A = \prod_{i=1}^{l} (X - \lambda_i)$.

**Proof.** Suppose $A$ is diagonalizable and let $B$ be a basis of $V$, consisting of eigenvectors of $A$. Define $g := \prod_{\lambda \in \sigma(A)} (X - \lambda)$. For each $b \in B$, there exists $\lambda \in \sigma(A)$ such that $Ab = \lambda b$. Thus,

$$\epsilon_A(g)(b) = \prod_{\lambda \in \sigma(A)} (A - \lambda) b = 0.$$

According to Th. 8.9(b), we have $\mu_A | g$. Since, by Th. 8.9(c), each $\lambda \in \sigma(A)$ is a zero of $\mu_A$, we have $\deg \mu_A = \deg g$. As both $\mu_A$ and $g$ are monic, this means $\mu_A = g$. Conversely, suppose $\mu_A = \prod_{i=1}^{l} (X - \lambda_i)$ with distinct $\lambda_1, \ldots, \lambda_l \in F$, $l \in \mathbb{N}$. Then, by Th. 8.13,

$$V = \bigoplus_{i=1}^{l} \ker \epsilon_A(X - \lambda_i) = \bigoplus_{i=1}^{l} \ker (A - \lambda_i \text{Id}) = \bigoplus_{i=1}^{l} E_A(\lambda_i),$$

proving $A$ to be diagonalizable by Th. 6.3(d). □

**Example 8.15.** (a) Let $V$ be vector space over $\mathbb{C}$, $\dim V = n \in \mathbb{N}$. If $A \in \mathcal{L}(V, V)$ is such that there exists $m \in \mathbb{N}$ with $A^m = \text{Id}$, then $A^m - \text{Id} = 0$ and $\mu_A | (X^m - 1) = \prod_{k=1}^{m} (X - \zeta_k)$, $\zeta_k := e^{k2\pi i/m}$. As the roots of unity $\zeta_k$ are all distinct (cf. [Phi16a, Cor. 8.31]), $A$ is diagonalizable by Th. 8.14.

(b) Let $V$ be vector space over the field $F$, $\dim V = n \in \mathbb{N}$, and let $P \in \mathcal{L}(V, V)$ be a projection, i.e. $P^2 = P$. Then $P^2 - P = 0$ and $\mu_P | (X^2 - X) = X(X - 1)$. Thus, we obtain the three cases

$$\mu_P = \begin{cases} X & \text{for } P = 0, \\ X - 1 & \text{for } P = \text{Id}, \\ X(X - 1) & \text{otherwise.} \end{cases}$$
(c) Let \( V \) be vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \), and let \( A \in \mathcal{L}(V, V) \) be a so-called involution i.e. \( A^2 = \text{Id} \). Then \( A^2 - \text{Id} = 0 \) and \( \mu_A \mid (X^2 - 1) = (X + 1)(X - 1) \). Thus, we obtain the three cases

\[
\mu_A = \begin{cases} 
X - 1 & \text{for } A = \text{Id}, \\
X + 1 & \text{for } A = -\text{Id}, \\
(X + 1)(X - 1) & \text{otherwise}.
\end{cases}
\]

If \( A \neq \pm \text{Id} \), then, according to Th. 8.14, \( A \) is diagonalizable if, and only if, \( 1 \neq -1 \), i.e. if, and only if, \( \text{char} F \neq 2 \). Even though \( A \neq \pm \text{Id} \) is not diagonalizable for \( \text{char} F = 2 \), there still exists an ordered basis \( B \) of \( V \) such that the matrix \( M \) of \( A \) with respect to \( B \) is triangular (all diagonal elements being 1), due to Prop. 8.12(b).

**Proposition 8.16.** Let \( V \) be a vector space over the real numbers \( \mathbb{R} \), \( \dim V = n \in \mathbb{N} \), and \( A \in \mathcal{L}(V, V) \).

(a) There exists a vector subspace \( U \) of \( V \) such that \( \dim U \in \{1, 2\} \) and \( U \) is \( A \)-invariant (i.e. \( A(U) \subseteq U \)).

(b) There exists an ordered basis \( B \) of \( V \) and matrices \( M_1, \ldots, M_l \), \( l \in \mathbb{N} \), such that each \( M_i \) is either \( 1 \times 1 \) or \( 2 \times 2 \) over \( \mathbb{R} \), and such that the matrix \( M \) of \( A \) with respect to \( B \) has the block triangular form

\[
M = \begin{pmatrix} 
M_1 & * & * \\
& \ddots & * \\
& & M_l
\end{pmatrix}.
\]

*Proof.* Exercise. \( \blacksquare \)

\section{Jordan Normal Form}

If \( V \) is a vector space over the algebraically closed field \( F \), \( \dim V = n \in \mathbb{N} \), \( A \in \mathcal{L}(V, V) \), then one can always find an ordered basis \( B \) of \( V \) such that the corresponding matrix \( M \) of \( A \) is in so-called Jordan normal form, which is an especially simple (upper) triangular form, where the eigenvalues are found on the diagonal, the value 1 can, possibly, occur directly above the diagonal, and all other entries are 0. However, we will also see below, that, if \( F \) is not algebraically closed, then one can still obtain a normal form for \( M \), albeit, in general, it is more complicated than the Jordan normal form.

**Definition 9.1.** Let \( V \) be a vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \), \( A \in \mathcal{L}(V, V) \).

(a) A vector subspace \( U \) of \( V \) is called \( A \)-cyclic if, and only if, \( U \) is \( A \)-invariant (i.e. \( A(U) \subseteq U \)) and

\[
\exists \ v \in V \quad U = \langle \{A^i v : i \in \mathbb{N}_0\} \rangle.
\]
Proposition 9.2. Let \( V \) be a finite-dimensional vector space over the field \( F \), \( r \in \mathbb{N} \), \( A \in \mathcal{L}(V,V) \). Suppose \( V \) is \( A \)-cyclic,

\[
\{A^i v : i \in \mathbb{N}_0\}, \quad v \in V,
\]

and

\[
\exists \begin{pmatrix} a_0, \ldots, a_r \in F \end{pmatrix} \quad \mu_A = \sum_{i=0}^{r} a_i X^i, \quad \text{deg} \mu_A = r.
\]

Then \( \chi_A = \mu_A \), \( \dim V = r \), \( B := (v, Av, \ldots, A^{r-1}v) \) is an ordered basis of \( V \), and, with respect to \( B \), \( A \) has the matrix

\[
M = \begin{pmatrix}
0 & 0 & \cdots & \cdots & -a_0 \\
1 & 0 & \cdots & \cdots & -a_1 \\
0 & 1 & \cdots & \cdots & -a_2 \\
\vdots & \vdots & \ddots & \cdots & \vdots \\
\cdots & \cdots & \cdots & 0 & -a_{r-2} \\
& & & 1 & -a_{r-1}
\end{pmatrix} \in \mathcal{M}(r,F). \tag{9.1}
\]

Proof. To show that \( v, Av, \ldots, A^{r-1}v \) are linearly independent, let \( \lambda_0, \ldots, \lambda_{r-1} \in F \) such that

\[
0 = \sum_{i=0}^{r-1} \lambda_i A^i v.
\tag{9.2}
\]

Define \( g := \sum_{i=0}^{r-1} \lambda_i X^i \in F[X] \). We need to show \( g = 0 \). Indeed, we have

\[
\forall_{i \in \mathbb{N}_0} \quad \epsilon_A(g) A^i v = A^i \epsilon_A(g) v \overset{(9.2)}{=} 0,
\]

which, as the \( A^i \) generate \( V \), implies \( \epsilon_A(g) = 0 \). Since \( \deg g < \deg \mu_A \), this yields \( g = 0 \) and the linear independence of \( v, Av, \ldots, A^{r-1}v \). We show \( \langle B \rangle = V \) next: Seeking a contradiction, assume \( A^m \notin \langle B \rangle \), where we choose \( m \in \mathbb{N} \) to be minimal. Then \( m \geq r \).

Then there exist \( \lambda_0, \ldots, \lambda_{r-1} \in F \) such that \( A^{m-1} v = \sum_{i=0}^{r-1} \lambda_i A^i v \). Then \( m > r \) yields the contradiction

\[
A^m v = \sum_{i=1}^{r} \lambda_i A^i v \in \langle B \rangle.
\]

However, \( m = r \) also yields a contradiction due to

\[
0 = \epsilon_A(\mu_A)v = A^r v + \sum_{i=0}^{r-1} a_i A^i v \quad \Rightarrow \quad A^r v \in \langle B \rangle.
\]

Thus, \( B \) is an ordered basis of \( V \). Then \( \deg \chi_A = r \), implying \( \chi_A = \mu_A \). Finally, if \( e_k \) is the \( k \)th standard column basis vector of \( F^r \), \( k \in \{1, \ldots, r\} \), then \( Me_k = e_{k+1} \) for \( k \in \{1, \ldots, r-1\} \) and \( Me_r = -\sum_{i=0}^{r-1} a_i e_i \), showing \( M \) to be the matrix of \( A \) with respect to \( B \), since \( A(A^{r-1}v) = A^r v = -\sum_{i=0}^{r-1} a_i A^i v \).
Proposition 9.3. Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, $A \in \mathcal{L}(V, V)$. Moreover, let $U$ be an $A$-invariant subspace of $V$, $1 \leq l := \dim U < n$, and define the linear maps

$$A_U : U \rightarrow U,$$
$$A_{V/U} : V/U \rightarrow V/U,$$
$$A_U : A|_U,$$
$$A_{V/U}(v + U) := Av + U.$$

Then the following holds:

(a) Let $B := (v_1, \ldots, v_n)$ be an ordered basis of $V$ such that $B_U := (v_1, \ldots, v_i)$ is an ordered basis of $U$. Then the matrix $M$ of $A$ with respect to $B$ has the block form

$$M = (m_{ji}) = \left( \begin{array}{cc} M_U & * \\ 0 & M_{V/U} \end{array} \right), \quad M_U \in \mathcal{M}(l, F), \quad M_{V/U} \in \mathcal{M}(n - l, F),$$

where $M_U$ is the matrix of $A_U$ with respect to $B_U$ and $M_{V/U}$ is the matrix of $A_{V/U}$ with respect to the ordered basis $B_{V/U} := (v_{i+1} + U, \ldots, v_n + U)$ of $V/U$.

(b) $\chi_A = \chi_A \chi_{A_U}$.

(c) $\mu_{A_U} | \mu_A$ and $\mu_{A_{V/U}} | \mu_A$.

Proof. $A_U$ is well-defined, since $U$ is $A$-invariant, and linear as $A$ is linear. Moreover, $A_{V/U}$ is well-defined, since, for each $v, w \in V$ with $v, w \in U$, $Av + U = Av + A(w - v) + U = Aw + U$; $A_{V/U}$ is linear, since, for each $v, w \in V$ and $\lambda \in F$,

$$A_{V/U}((v + w) + U) = A(v + w) + U = Av + U + Aw + U$$
$$= A_{V/U}(v + U) + A_{V/U}(w + U),$$
$$A_{V/U}(\lambda v + U) = A(\lambda v) + U = \lambda(Av + U) = \lambda A_{V/U}(v + U).$$

(a): Since $U$ is $A$-invariant, we have

$$\forall i \in \{1, \ldots, l\} \quad \exists m_{i, 1}, \ldots, m_{i, l} \in F \quad Av_i = \sum_{j=1}^{l} m_{ji} v_j,$$

showing $M$ to have the claimed form with $M_U$ being the matrix of $A_U$ with respect to $B_U$. Moreover, by [Phi19, Cor. 6.13(a)], $B_{V/U}$ is, indeed, an ordered basis of $V/U$ and

$$\forall i \in \{l+1, \ldots, n\} \quad A_{V/U}(v_i + U) = Av_i + U = \sum_{j=1}^{n} m_{ji} v_j + U = \sum_{j=l+1}^{n} m_{ji} v_j + U,$$

proving $M_{V/U}$ to be the matrix of $A_{V/U}$ with respect to $B_{V/U}$.

(b): We compute

$$\chi_A = \det(X \operatorname{Id}_n - M) \quad \text{Th.}4.25 \quad \det(X \operatorname{Id}_l - M_U) \det(X \operatorname{Id}_{n-l} - M_{V/U}) \quad \text{Th.}4.25 \quad \chi_{A_U} \chi_{A_{V/U}}.$$

(c): Since $\epsilon_A(\mu_A)v = 0$ for each $v \in V$, $\epsilon_A(\mu_A)v = 0$ for each $v \in U$, proving $\epsilon_{A_U}(\mu_A) = 0$ and $\mu_{A_U} | \mu_A$. Similarly, $\epsilon_{A_{V/U}}(\mu_A)(v + U) = \epsilon_A(\mu_A)v + U = 0$ for each $v \in V$, proving $\epsilon_{A_{V/U}}(\mu_A) = 0$ and $\mu_{A_{V/U}} | \mu_A$. \hfill $\blacksquare$
Comparing Prop. 9.3(b),(c) above, one might wonder if the analogon of Prop. 9.3(b) also holds for the minimal polynomials. The following Ex. 9.4 shows that, in general, it does not:

**Example 9.4.** Let $F$ be field and $V := F^2$. Then, for $A := \text{Id} \in \mathcal{L}(V,V)$, $\mu_A = X - 1$. If $U$ is an arbitrary 1-dimensional subspace of $V$, then, using the notation of Prop. 9.3, $\mu_{A_U} = X - 1 = \mu_{A_{V/U}}$, i.e. $\mu_{A_U} \mu_{A_{V/U}} = (X - 1)^2 \neq \mu_A$.

**Lemma 9.5.** Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, $A \in \mathcal{L}(V,V)$. Suppose $V$ is $A$-cyclic.

(a) If $\mu_A = gh$ with $g, h \in F[X]$, then we have:

(i) $\dim \ker \epsilon_A(h) = \deg h$.

(ii) $\ker \epsilon_A(h) = \text{Im} \epsilon_A(g)$.

(b) If $\lambda \in \sigma(A)$, then $\dim E_A(\lambda) = 1$, i.e. every eigenspace of $A$ has dimension 1.

**Proof.** (a): To prove (i), let $U := \text{Im} h(A) = \epsilon_A(h)(V)$ and define $A_U, A_{V/U}$ as in Prop. 9.3. As $V$ is $A$-cyclic (say, generated by $v \in V$), $U$ is $A_U$-cyclic (generated by $\epsilon_A(h)v$) and $V/U$ is $A_{V/U}$-cyclic (generated by $v + U$). Thus, Prop. 9.2 yields

$$\chi_A = \mu_A, \quad \chi_{A_U} = \mu_{A_U}, \quad \chi_{A_{V/U}} = \mu_{A_{V/U}},$$

implying

$$gh = \mu_A = \chi_A \overset{\text{Prop. 9.3(b)}}{=} \chi_{A_U} \chi_{A_{V/U}} = \mu_{A_U} \mu_{A_{V/U}}. \quad (9.3)$$

If $v \in V$, then $\epsilon_A(g)\epsilon_A(h)v = \epsilon_A(\mu_A)v = 0$, showing $\epsilon_{A_U}(g) = 0$ and $\mu_{A_U} \mid g$, $\deg \mu_{A_U} \leq \deg g$. Similarly, $\epsilon_{A_{V/U}}(h)(v + U) = \epsilon_A(h)v + U = 0$ (since $\epsilon_A(h)v \in U$), proving $\epsilon_{A_{V/U}}(h) = 0$ and $\mu_{A_{V/U}} \mid h$, $\deg \mu_{A_{V/U}} \leq \deg h$. Since we also have $\deg g + \deg h = \deg \mu_{A_U} + \deg \mu_{A_{V/U}}$ by (9.3), we must have $\deg g = \deg \mu_{A_U}$ and $\deg h = \deg \mu_{A_{V/U}}$. Thus,

$$\dim U = \deg \chi_{A_U} = \deg \mu_{A_U} = \deg g = \deg \mu_A - \deg h.$$

According to the isomorphism theorem [Phi19, Th. 6.16(a)], we know

$$V/\ker \epsilon_A(h) \cong \text{Im} \epsilon_A(h) = U,$$

implying

$$\deg h = \deg \mu_A - \dim U = \dim V - \dim U = \dim V - \dim \left(V/\ker \epsilon_A(h)\right) = \dim V - \left(\dim V - \dim \ker \epsilon_A(h)\right) = \dim \ker \epsilon_A(h),$$

thereby proving (i). We proceed to prove (ii): Since $\epsilon_A(h)(\text{Im} \epsilon_A(g)) = \{0\}$, we have $\text{Im} \epsilon_A(g) \subseteq \ker \epsilon_A(h)$. To prove equality, we note that (i) must also hold with $g$ instead of $h$ and compute

$$\dim \text{Im} \epsilon_A(g) = \dim V - \dim \ker \epsilon_A(g) \overset{(i)}{=} \deg \chi_A - \deg g = \deg \mu_A - \deg g = \deg h \overset{(i)}{=} \dim \ker \epsilon_A(h),$$

where $\text{Im} \epsilon_A(g)$ is the image of $g$ under the isomorphism $\epsilon_A$. This completes the proof of (ii).
then showing $\mu$. Since $v$ and $\mu$ to be $A$ such that $\mu = (X - \lambda)g$. Hence,

$$ \dim E_A(\lambda) = \dim \ker(A - \lambda \text{Id}) \overset{(a)(i)}{=} \deg(X - \lambda) = 1, $$

proving (b).

**Lemma 9.6.** Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, $A \in \mathcal{L}(V, V)$. Suppose $\mu_A = g^r$, where $g \in F[X]$ is irreducible and $r \in \mathbb{N}$. If $U$ is an $A$-cyclic subspace of $V$ such that $U$ has maximal dimension, then $\mu_{A|_U} = \mu_A$ and there exists an $A$-invariant subspace $W$ such that $V = U \oplus W$.

**Proof.** Letting $A_U := A |_U$, we show $\mu_{A_U} = \mu_A$ first: According to Prop. 9.3(c), we have $\mu_{A_U} |_{\mu_A}$, i.e. there exists $1 \leq r_1 \leq r$ such that $\mu_{A_U} = g^{r_1}$. According to Prop. 9.2, $\chi_{A_U} = \mu_{A_U}$, implying $\dim U = \deg \mu_{A_U} = r_1 \deg g$. Let $0 \neq v \in V$ and define $U_1 := \{(A^i v : i \in \mathbb{N}_0)\}$. Then $U_1$ is an $A$-cyclic subspace of $V$ and the maximality of $V$ implies $\mu_{A|_U} = g^{r_1}$, implying $\epsilon_A(g^{r_1})(U_1) = \{0\}$. As $0 \neq v \in V$ was arbitrary, this shows $\epsilon_A(g^{r_1}) = 0$, implying $\mu_A |_{g^{r_1}} = \mu_{A_U}$, showing $\mu_{A_U} = \mu_A$ (and $r_1 = r$) as claimed.

The proof of the existence of $W$ is now conducted via induction on $n \in \mathbb{N}$. For $n = 1$, we have $U = V$ and there is nothing to prove ($W = \{0\}$). Thus, let $n > 1$. If $U = V$, then, as for $n = 1$, we can merely set $W := \{0\}$. Thus, $U \neq V$. First, consider the case that $V/U$ is $A_{V/U}$- reducible, where $A_{V/U}$ is as in Prop. 9.3. Then there exist $A$-invariant subspaces $\emptyset \subsetneq V_1, V_2 \subseteq V$ such that $V/U = (V_1/U) \oplus (V_2/U)$, i.e. $V_1 + V_2 + U = V$ and $V_1 \cap V_2 \subseteq U$ (since $v \in V_1 \cap V_2$ implies $v + U \in (V_1/U) \cap (V_2/U) = U$). Replacing $V_1, V_2$ by $V_1 + U, V_2 + U$, respectively, we may also assume $V_1 \cap V_2 = U$. As $U \subseteq V_1, V_2$ and $\dim V_1, \dim V_2 < n$, we can use the induction hypothesis to obtain $A$-invariant subspaces $W_1, W_2$ of $V_1, V_2$, respectively, such that $V_1 = U \oplus W_1, V_2 = U \oplus W_2$. Let $W := W_1 \oplus W_2$ (the sum is direct, as $w \in W_1 \cap W_2$ implies $w \in V_1 \cap V_2 = U$, i.e. $w = 0$). If $v \in V$, then $V = V_1 + V_2$ implies the existence of $w_1, w_2 \in U$ and $w_1 \in W_1, w_2 \in W_2$ such that $v = u_1 + u_1 + w_2 + w_2$, showing $V = U + W$. Using $V/U = (V_1/U) \oplus (V_2/U)$ as well as $W_1 \cong V_1/U, W_2 \cong V_2/U$ by [Phi19, Th. 6.16(b)], we compute

$$ \dim V - \dim U + \dim (U \cap W) = \dim W = \dim W_1 + \dim W_2 $$

$$ = \dim (V_1/U) + \dim (V_2/U) $$

$$ = \dim (V/U) = \dim V - \dim U, $$

showing $U \cap W = \{0\}$ and $V = U \oplus W$. It remains to consider the case that $V/U$ is $A_{V/U}$-irreducible. As $\dim V/U < n$, the induction hypothesis applies to $V/U$, implying $V/U$ to be $A_{V/U}$-cyclic, i.e. there exists $v \in V$ with $V/U = \{(A_{V/U})^i(v + U) : i \in \mathbb{N}_0\}$. Since $\mu := \mu_{A_{V/U}} |_{\mu_A}$ by Prop. 9.3(c), we have $\mu = g^s$ with $1 \leq s \leq r$. Then

$$ \epsilon_A(g^{r-s}) \epsilon_A(g^s) \mu v = \epsilon_A(\mu_A) v = 0, $$

showing

$$ \epsilon_A(g^s) v = \epsilon_A(\mu) v \in U \cap \ker \epsilon_A(g^{r-s}). \quad (9.4) $$
Since $\mu|\mu_A = \mu_{A^r}$, we can apply Lem. 9.5(a)(ii) to obtain
\[ U \cap \ker \epsilon_A(g^{r-*)} = \ker \epsilon_{A^r} (g^{r-*)} = \text{Im} \epsilon_{A^r} (g^r), \]
which, when combined with (9.4), yields some $u_v \in U$ with
\[ \epsilon_A(g^r) u_v = \epsilon_A(g^r) v. \quad (9.5) \]
This, finally, allows us to define
\[ w_0 := v - u_v, \quad W := \langle \{ A^i w_0 : i \in \mathbb{N}_0 \} \rangle. \]
Clearly, $W$ is $A$-invariant. If $x \in V$, then $x \in x + U \cap V/U$ and, since $\dim(V/U) < n$, there exist $u \in U$ and $\lambda_1, \ldots, \lambda_n \in F$ such that
\[ x = u + \sum_{i=0}^{n} \lambda_i A^i v = u + \sum_{i=0}^{n} \lambda_i A^i (w_0 + u_v) = u + \sum_{i=0}^{n} \lambda_i A^i u_v + \sum_{i=0}^{n} \lambda_i A^i w_0, \]
proving $V = U + W$. Since
\[ \epsilon_A(g^r) w_0 = \epsilon_A(g^r) v - \epsilon_A(g^r) u_v \overset{(9.5)}{=} 0, \]
we have $\dim W = \deg \mu_{A^r} \leq \deg(g^r) = \deg \mu = \dim(V/U)$, implying (as $V = U + W$) $\dim W = \dim(V/U)$ and $V = U \oplus W$. \hspace{1cm} \blacksquare

**Theorem 9.7.** Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, $A \in \mathcal{L}(V, V)$.

(a) If $V$ is $A$-irreducible, then $V$ is $A$-cyclic.

(b) Suppose $A = g^r$, where $g \in F[X]$ is irreducible and $r \in \mathbb{N}$. Then $V$ is $A$-irreducible, if, and only if, $V$ is $A$-cyclic.

(c) There exist subspaces $U_1, \ldots, U_l$ of $V$, $l \in \mathbb{N}$, such that
\[ V = \bigoplus_{i=1}^{l} U_i \]
and each $U_i$ is both $A$-irreducible and $A$-cyclic.

**Proof.** (a): We must have $\mu_A = g^r$ with $r \in \mathbb{N}$ and $g \in F[X]$ irreducible, since, otherwise $V$ is $A$-reducible by Th. 8.13. In consequence, $V$ is $A$-cyclic by Lem. 9.6.

(b): According to (a), it only remains to show that $V$ being $A$-cyclic implies $V$ to be $A$-irreducible. Thus, let $V$ be $A$-cyclic and $V = V_1 \oplus V_2$ with $A$-invariant subspaces $V_1, V_2 \subseteq V$. Then, by Prop. 9.3(c), there exist $1 \leq r_1, r_2 \leq r$ such that $\mu_{A^{r_1}} = g^{r_1}$ and $\mu_{A^{r_2}} = g^{r_2}$, where we choose $V_1$ such that $r_1 \leq r_2 \leq r_1$. Then $\epsilon_A(\mu_{A^{r_1}})(V_1) = \epsilon_A(\mu_{A^{r_2}})(V_2) = \{0\}$, showing $\mu_A = \mu_{A^{r_1}}$. As $\chi_A = \mu_A$ by Prop. 9.2, we have $\dim V_1 \geq \deg \mu_{A^{r_1}} = \deg \chi_A = \dim V$, showing $V_1 = V$ as desired.
(c): The proof is conducted via induction on \( n \in \mathbb{N} \). If \( V \) is \( A \)-irreducible, then, by (a), \( V \) is also \( A \)-cyclic and the statement holds (in particular, this yields the base case \( n = 1 \)). If \( V \) is \( A \)-reducible, then there exist \( A \)-invariant subspaces \( V_1, V_2 \) of \( V \) such that \( V = V_1 \oplus V_2 \) with \( \dim V_1, \dim V_2 < n \). Thus, by induction hypothesis, both \( V_1 \) and \( V_2 \) can be written as a direct sum of subspaces that are both \( A \)-irreducible and \( A \)-cyclic, proving the same for \( V \). 

We now have all preparations in place to prove the existence of normal forms having matrices with block diagonal form, where the blocks all look like the matrix of (9.1). However, before we state and prove the corresponding theorem, we still provide a proposition that will help to address the uniqueness of such normal forms:

**Proposition 9.8.** Let \( V \) be a vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \), \( A \in \mathcal{L}(V,V) \). Moreover, suppose we have a decomposition

\[
V = \bigoplus_{i=1}^l U_i, \quad l \in \mathbb{N},
\]

(9.6)

where the \( U_1, \ldots, U_l \) all are \( A \)-invariant and \( A \)-irreducible (and, thus, \( A \)-cyclic by Th. 9.7(a)) subspaces of \( V \).

(a) If \( \mu_A = g_1^{r_1} \cdots g_m^{r_m} \) is the prime factorization of \( \mu_A \) (i.e. each \( g_i \in F[X] \) is irreducible, \( r_1, \ldots, r_m \in \mathbb{N} \), \( m \in \mathbb{N} \)), then

\[
\ker \epsilon_A(g_i^{r_i}) \quad (9.7)
\]

and the decomposition of (9.6) is a refinement of (9.7) in the sense that each \( U_i \) is contained in some \( \ker \epsilon_A(g_i^{r_i}) \).

(b) If \( \mu_A = g^r \) with \( g \in F[X] \) irreducible, \( r \in \mathbb{N} \), then, for each \( i \in \{1, \ldots, l\} \), we have \( \mu_{AU_i} = g^{r_i} \), \( \dim U_i = r_i \deg g \), with \( 1 \leq r_i \leq r \). If

\[
\forall_{k \in \mathbb{N}_0} l_k := \# \{ i \in \{1, \ldots, l\} : \mu_{AU_i} = g^k \}
\]

(i.e. \( l_k \) is the number of summands \( U_i \) with \( \mu_{AU_i} = g^k \)), then

\[
l = \sum_{k=1}^r l_k, \quad \dim V = (\deg g) \sum_{k=1}^r k l_k, \quad (9.8)
\]

and

\[
\forall_{s \in \{0, \ldots, r\}} \dim \text{Im} \epsilon_A(g^s) = (\deg g) \sum_{k=s}^r l_k (k - s). \quad (9.9)
\]

In consequence, the numbers \( l_k \) are uniquely determined by \( A \).
Proof. (a): That (9.7) is an immediate consequence of Th. 8.13. Moreover, if \(U\) is an \(A\)-invariant and \(A\)-irreducible subspace of \(V\), then \(\mu_{A_U} | \mu_A\) and Th. 8.13 imply \(\mu_{A_U} = g^s_j\) for some \(j \in \{1, \ldots, m\}\) and \(1 \leq s \leq r_j\). In consequence, \(U \subseteq \ker \, \epsilon_A(g^r_j)\) as claimed.

(b): The proof of (a) already showed that, for each \(i \in \{1, \ldots, l\}\), \(\mu_{A_{U_i}} = g^{r_i}\), with \(1 \leq r_i \leq r\), and then \(\dim U_i = r_i \deg g\) by Prop. 9.2. From the definitions of \(r\) and \(l_k\), it is immediate that \(l = \sum_{k=1}^r l_k\). As the sum in (9.6) is direct, we obtain

\[
\dim V = \sum_{i=1}^l \dim U_i = \sum_{k=1}^r l_k k \deg g = (\deg g) \sum_{k=1}^r k l_k.
\]

To prove (9.9), fix \(s \in \{0, \ldots, r\}\) and set \(h := g^s\). For each \(v \in V\), there exist \(u_1, \ldots, u_l \in V\) such that \(u_i \in U_i\) and \(v = \sum_{i=1}^l u_i\). Then

\[
\epsilon_A(h) v = \sum_{i=1}^l \epsilon_A(h) u_i,
\]

implying

\[
\ker \epsilon_A(h) = \bigoplus_{i=1}^l \ker \epsilon_{A_{U_i}}(h), \tag{9.10}
\]

due to the fact that \(\epsilon_A(h) v = 0\) if, and only if, \(\epsilon_A(h) u_i = 0\) for each \(i \in \{1, \ldots, l\}\). In the case, where \(\mu_{A_{U_i}} = g^{r_i}\) with \(r_i \leq s\), we have \(\ker \epsilon_{A_{U_i}}(h) = U_i\), i.e.

\[
\dim \ker \epsilon_{A_{U_i}}(h) = \dim U_i = r_i \deg g, \tag{9.11a}
\]

while, in the case, where \(\mu_{A_{U_i}} = g^{r_i}\) with \(s < r_i\), we can apply Lem. 9.5(a)(i) to obtain

\[
\dim \ker \epsilon_{A_{U_i}}(h) = \deg(g^s) = s \deg g. \tag{9.11b}
\]

Putting everything together yields

\[
\dim \text{Im} \, \epsilon_A(g^s) = \dim \text{Im} \, \epsilon_A(h) = \dim V - \dim \ker \epsilon_A(h)
\]

\[
\overset{(9.8), (9.10)}{=} (\deg g) \sum_{k=1}^r k l_k - \sum_{i=1}^l \dim \ker \epsilon_{A_{U_i}}(h) \overset{(9.11)}{=} (\deg g) \sum_{k=s}^r l_k (k - s),
\]

proving (9.9). To see that the \(l_k\) are uniquely determined by \(A\), observe \(l_k = 0\) for \(k > r\) and \(k = 0\), and, for \(1 \leq k \leq r\), (9.9) implies the recursion

\[
l_r = (\deg g)^{-1} \dim \text{Im} \, \epsilon_A(g^{r-1}), \tag{9.12a}
\]

\[
\forall s \in \{1, \ldots, r-1\} \quad l_s = (\deg g)^{-1} \dim \text{Im} \, \epsilon_A(g^{s-1}) - \sum_{k=s+1}^r l_k (k - (s - 1)), \tag{9.12b}
\]

thereby completing the proof.
Theorem 9.9 (General Normal Form). Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, $A \in \mathcal{L}(V, V)$.

(a) There exist subspaces $U_1, \ldots, U_l$ of $V$, $l \in \mathbb{N}$, such that each $U_i$ is $A$-cyclic and $A$-irreducible, satisfying

$$V = \bigoplus_{i=1}^l U_i.$$ 

Moreover, for each $i \in \{1, \ldots, l\}$, there exists $v_i \in U_i$ such that

$$B_i := \{v_i, Av_i, \ldots, A^{r_i-1}v_i\}, \quad r_i := \dim U_i,$$

is a basis of $U_i$. Then $\mu_{A|_{U_i}} = \sum_{k=0}^{r_i} a_k^{(i)} X^k$ with $a_0^{(i)}, \ldots, a_{r_i}^{(i)} \in F$, $A_{U_i} := A|_{U_i}$, and, with respect to the ordered basis

$$B := (v_1, A^{r_1-1}v_1, \ldots, v_l, A^{r_l-1}v_l),$$

$A$ has the block diagonal matrix

$$M := \begin{pmatrix}
M_1 & 0 & \cdots & 0 \\
0 & M_2 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & M_l
\end{pmatrix},$$

each block having the form of (9.1), namely

$$\forall i \in \{1, \ldots, l\} \quad M_i = \begin{pmatrix}
0 & 0 & \cdots & \cdots & -a_0^{(i)} \\
1 & 0 & \cdots & \cdots & -a_1^{(i)} \\
0 & 1 & \ddots & \vdots & \vdots \\
\vdots & \ddots & \ddots & 0 & -a_{r_i-2}^{(i)} \\
\cdots & \cdots & \cdots & 1 & -a_{r_i-1}^{(i)}
\end{pmatrix}.$$

(b) If

$$V = \bigoplus_{i=1}^m W_i$$

is another decomposition of $V$ into $A$-invariant and $A$-irreducible subspaces $W_1, \ldots, W_m$ of $V$, $m \in \mathbb{N}$, then $m = l$ and there exist a permutation $\pi \in S_l$ and $T \in \text{GL}(V)$ such that

$$TA = AT \quad \land \quad \forall i \in \{1, \ldots, l\} \quad T(U_i) = W_{\pi(i)}. \quad (9.13)$$
Proof. (a): The existence of the claimed decomposition was already shown in Th. 9.7(c) and the remaining statements are then provided by Prop. 9.2, where the \( A \)-invariance of the \( U_i \) yields the block diagonal structure of \( M \).

(b): We divide the proof into three steps:

Step 1: Assume \( U \) and \( W \) to be \( A \)-cyclic subspaces of \( V \) such that \( 1 \leq s := \dim U = \dim W \leq n \) and let \( u \in U, w \in W \) be such that \( B_U := \{ u, \ldots, A^{s-1}u \}, \ B_W := \{ w, \ldots, A^{s-1}w \} \) are bases of \( U, W \), respectively. Define \( S \in \mathcal{L}(U, W) \) by letting

\[
\forall i \in \{0, \ldots, s-1\} \quad S(A^i u) := A^i w
\]

(then \( S \) is invertible, as it maps the basis \( B_U \) onto the basis \( B_W \)). We verify \( SA = AS \): Indeed,

\[
\forall i \in \{0, \ldots, s-2\} \quad SA(A^i u) = S(A^{i+1} u) = A^{i+1} w = A(A^i w) = AS(A^i u)
\]

and, letting \( a_0, \ldots, a_s \in F \) be such that \( \mu_{A_U} = \mu_{A_W} = \sum_{i=0}^{s} a_i X^i \) (cf. Prop. 9.2),

\[
SA(A^{s-1} u) = S(A^s u) = S \left( - \sum_{i=0}^{s-1} a_i A^i u \right) = - \sum_{i=0}^{s-1} a_i A^i w
\]

\[
= A^s w = A(A^{s-1} w) = AS(A^{s-1} u).
\]

Step 2: Assume \( \mu_A = g^s \), where \( g \in F[X] \) is irreducible and \( s \in \mathbb{N} \). Letting

\[
\forall k \in \mathbb{N}_0 \quad I(U, k) := \{ i \in \{1, \ldots, l \} : \mu_{A_{U_i}} = g^k \},
\]

\[
I(W, k) := \{ i \in \{1, \ldots, m \} : \mu_{A_{W_i}} = g^k \},
\]

the uniqueness of the numbers \( l_k \) in Prop. 9.8(b) shows

\[
\forall k \in \mathbb{N}_0 \quad l_k = \# I(U, k) = \# I(W, k)
\]

and, in particular, \( m = l \), and the existence of \( \pi \in S_l \) such that, for each \( k \in \mathbb{N} \) and each \( i \in I(U, k) \), one has \( \pi(i) \in I(W, k) \). Thus, by Step 1, for each \( i \in \{1, \ldots, l\} \), there exists an invertible \( S_i \in \mathcal{L}(U_i, W_{\pi(i)}) \) such that \( S_i A = A S_i \). Define \( T \in GL(V) \) by letting, for each \( v \in V \) such that \( v = \sum_{i=1}^{l} u_i \) with \( u_i \in U_i \), \( Tv := \sum_{i=1}^{l} S_i u_i \). Then, clearly, \( T \) satisfies (9.13).

Step 3: We now consider the general situation of (b). Let \( \mu_A = g_1^{r_1} \cdots g_s^{r_s} \) be the prime factorization of \( \mu_A \) (i.e. each \( g_i \in F[X] \) is irreducible, \( r_1, \ldots, r_s \in \mathbb{N}, s \in \mathbb{N} \)). According to Prop. 9.8(a), there exist sets \( I_1, \ldots, I_s \subseteq \{1, \ldots, l\} \) and \( J_1, \ldots, J_s \subseteq \{1, \ldots, m\} \) such that

\[
\forall k \in \{1, \ldots, s\} \quad \ker \epsilon_A(g_k^{l_k}) = \bigoplus_{i \in I_k} U_i = \bigoplus_{i \in J_k} W_i.
\]

Then, by Step 2, we have \( \# I_k = \# J_k \) for each \( k \in \{1, \ldots, s\} \), implying

\[
l = \sum_{k=1}^{s} \# I_k = \sum_{k=1}^{s} \# J_k = m
\]
and the existence of a permutation \( \pi \in S_l \) such that \( \pi(I_k) = J_k \) for each \( k \in \{1, \ldots, s\} \).

Again using Step 2, we can now, in addition, choose \( \pi \in S_l \) such that

\[
\forall \ i \in \{1, \ldots, l\} \quad \exists \ T_i \in \mathcal{L}(U_i, W_m) \quad \left( T_i \text{ invertible} \land T_i A = A T_i \right).
\]

Define \( T \in \text{GL}(V) \) by letting, for each \( v \in V \) such that \( v = \sum_{i=1}^l u_i \) with \( u_i \in U_i \),

\[
Tv := \sum_{i=1}^l T_i u_i.
\]

Then, clearly, \( T \) satisfies (9.13).

The following Prop. 9.10 can sometimes be helpful in actually finding the \( U_i \) of Th. 9.9(a):

**Proposition 9.10.** Let \( V \) be a vector space over the field \( F \), \( \dim V = n \in \mathbb{N} \), \( A \in \mathcal{L}(V, V) \), and let \( \mu_A = g_1^r_1 \cdots g_m^r_m \) be the prime factorization of \( \mu_A \) (i.e. each \( g_i \in F[X] \) is irreducible, \( r_1, \ldots, r_m \in \mathbb{N}, m \in \mathbb{N} \)). For each \( i \in \{1, \ldots, m\} \), let \( V_i := \ker A(g_i^{r_i}) \).

(a) For each \( i \in \{1, \ldots, m\} \), we have \( \mu_{A_{V_i}} = g_i^{r_i} \) (recall \( V = \bigoplus_{i=1}^m V_i \)).

(b) For each \( i \in \{1, \ldots, m\} \), in the decomposition \( V = \bigoplus_{k=1}^l U_k \) of Th. 9.9(a), there exists at least one \( U_k \) with \( U_k \subseteq V_i \) and \( \dim U_k = \deg(g_i^{r_i}) \).

(c) As in (b), let \( V = \bigoplus_{k=1}^l U_k \) be the decomposition of Th. 9.9(a). Then, for each \( i \in \{1, \ldots, m\} \) and each \( k \in \{1, \ldots, l\} \) such that \( U_k \subseteq V_i \), one has \( \dim U_k \leq \deg(g_i^{r_i}) \).

**Proof.** (a): Let \( i \in \{1, \ldots, m\} \). According to Prop. 9.3(c), we have \( \mu_{A_{V_i}} = g_i^s \) with \( 1 \leq s \leq r_i \). For each \( v \in V_i \), there are \( v_1, \ldots, v_m \in V \) such that \( v = \sum_{i=1}^m v_i \) and \( v_i \in V_i \), implying

\[
\epsilon_A(g_1^{r_1} \cdots g_i^{r_i-1} g_i^s g_{i+1}^{r_{i+1}} \cdots g_m^{r_m}) v = 0
\]

and \( r_i \leq s \), i.e. \( r_i = s \).

(b): Let \( i \in \{1, \ldots, m\} \). According to (a), we have \( \mu_{A_{V_i}} = g_i^{r_i} \). Using the uniqueness of the decomposition \( V = \bigoplus_{k=1}^l U_k \) is the sense of Th. 9.9(b) together with Lem. 9.6, there exists \( U_k \) with \( U_k \subseteq V_i \) such that \( U_k \) is an \( A \)-cyclic subspace of \( V_i \) of maximal dimension. Then Lem. 9.6 also yields \( \mu_{A_{U_k}} = \mu_{A_{V_i}} = g_i^{r_i} \), which, as \( U_k \) is \( A \)-cyclic, implies \( \dim U_k = \deg(g_i^{r_i}) \).

(c): Let \( i \in \{1, \ldots, m\} \). Again, we know \( \mu_{A_{V_i}} = g_i^{r_i} \) by (a). If \( k \in \{1, \ldots, l\} \) is such that \( U_k \subseteq V_i \), then Prop. 9.3(c) yields \( \mu_{A_{U_k}}|_{\mu_{A_{V_i}}} = g_i^{r_i} \), showing \( \dim U_k \leq \deg(g_i^{r_i}) \).

**Remark 9.11.** In general, in the situation of Prop. 9.10, the knowledge of \( \mu_A \) and \( \chi_A \) does not suffice to uniquely determine the normal form of Th. 9.9(a). In general, for each \( g_i \) and each \( s \in \{1, \ldots, r_i\} \), one needs to determine \( \dim \text{Im} \epsilon_A(g_i^s) \), which then determine the numbers \( l_k \) of (9.9), i.e. the number of subspaces \( U_j \) with \( \mu_{A_{U_j}} = g_i^k \). This then determines the matrix \( M \) of Th. 9.9(a) (up to the order of the diagonal blocks), since one obtains precisely \( l_k \) many blocks of size \( k \deg g_i \) and the entries of these blocks are given by the coefficients of \( g_i^k \).
Example 9.12. (a) Let $F$ be a field and $V := F^6$. Assume $A \in \mathcal{L}(V, V)$ has

$$\chi_A = (X - 2)^2(X - 3)^4, \quad \mu_A = (X - 2)^2(X - 3)^3.$$ 

We want to determine the decomposition $V = \bigoplus_{i=1}^I U_i$ of Th. 9.9(a) and the matrix $M$ with respect to the corresponding basis of $V$ given in Th. 9.9(a): We know from Prop. 9.10(b) that we can choose $U_1 \subseteq \ker(A - 2 \text{Id})^2$ with $\dim U_1 = 2$ and $U_2 \subseteq \ker(A - 3 \text{Id})^3$ with $\dim U_2 = 3$. As $\dim V = 6$, this then yields $V = U_1 \oplus U_2 \oplus U_3$ with $\dim U_3 = 1$. We also know $\sigma(A) = \{2, 3\}$ and, according to Th. 8.4, the algebraic multiplicities are $m_a(2) = 2$, $m_a(3) = 4$. Moreover,

$$4 = m_a(3) = \dim \ker(A - 3 \text{Id})^4 = \dim \ker \epsilon_A((X - 3)^4),$$

implying $U_3 \subseteq \ker(A - 3 \text{Id})^4$. As $(X - 2)^2 = X^2 - 4X + 4$ and $(X - 3)^3 = X^3 - 9X^2 + 27X - 27$, $M$ has the form

$$M = \begin{pmatrix}
0 & -4 & 0 & 0 & 27 \\
1 & 4 & 1 & 0 & -27 \\
& & 1 & 9 & \\
& & & 3
\end{pmatrix}.$$ 

(b) Let $F := \mathbb{Z}_2 = \{0, 1\}$ and $V := F^3$. Assume, with respect to some basis of $V$, $A \in \mathcal{L}(V, V)$ has the matrix

$$N := \begin{pmatrix}
1 & 1 & 1 \\
1 & 0 & 1 \\
1 & 0 & 0
\end{pmatrix}.$$ 

We compute (using $1 = -1$ and $0 = 2$ in $F$)

$$\chi_A = \det(X \text{Id}_3 - N) = \begin{vmatrix}
X - 1 & -1 & -1 \\
-1 & X & -1 \\
-1 & 0 & X
\end{vmatrix} = X^3 - X^2 - 2X - 1 = X^3 + X^2 + 1.$$ 

Since $\chi_A = X(X^2 + X) + 1$ and $\chi_A = (X + 1)X^2 + 1$, $\chi_A$ is irreducible. Thus $\chi_A = \mu_A, V$ is $A$-irreducible and $A$-cyclic and the matrix of $A$ with respect to the basis of Th. 9.9(a) is (again making use of $-1 = 1$ in $F$)

$$M = \begin{pmatrix}
0 & 0 & 1 \\
1 & 0 & 0 \\
0 & 1 & 1
\end{pmatrix}.$$
For fields that are algebraically closed, we can improve the normal form of Th. 9.9 to the so-called Jordan normal form:

**Theorem 9.13** (Jordan Normal Form). Let $V$ be a vector space over the field $F$, $\dim V = n \in \mathbb{N}$, $A \in \mathcal{L}(V,V)$. Assume there exist distinct $\lambda_1, \ldots, \lambda_m \in K$ such that

$$\mu_A = \prod_{i=1}^{m} (X - \lambda_i)^{r_i}, \quad \sigma(A) = \{\lambda_1, \ldots, \lambda_m\}, \quad m, r_1, \ldots, r_m \in \mathbb{N} \quad (9.14)$$

(if $F$ is algebraically closed, then (9.14) always holds).

(a) There exist subspaces $U_1, \ldots, U_l$ of $V$, $l \in \mathbb{N}$, such that each $U_i$ is $A$-cyclic and $A$-irreducible, satisfying

$$V = \bigoplus_{k=1}^{l} U_k.$$ 

Moreover, for each $k \in \{1, \ldots, l\}$, there exists $v_k \in U_k$ and $i = i(k) \in \{1, \ldots, m\}$ such that

$$U_k \subseteq \ker(A - \lambda_i \text{Id})^{r_i}$$

and

$$J_k := \{v_k, (A - \lambda_i \text{Id})v_k, \ldots, (A - \lambda_i \text{Id})^{s_k-1}v_k\},$$

$$s_k := \dim U_k \leq r_i \leq \dim \ker(A - \lambda_i \text{Id})^{r_i},$$

is a basis of $U_k$ (note that, in general, the same $i$ will correspond to many distinct subspaces $U_k$). Then, with respect to the ordered basis

$$J := ((A - \lambda_{i(1)} \text{Id})^{s_1-1}v_1, \ldots, v_1, \ldots, (A - \lambda_{i(l)} \text{Id})^{s_l-1}v_l, \ldots, v_l),$$

$A$ has a matrix in Jordan normal form, i.e. the block diagonal matrix

$$N := \begin{pmatrix} N_1 & 0 & \ldots & 0 \\ 0 & N_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & N_l \end{pmatrix},$$

each block (called a Jordan block) having the form

$$N_k = (\lambda_{i(k)}) \in \mathcal{M}(1, F) \quad \text{for } s_k = 1,$$

$$N_k = \begin{pmatrix} \lambda_{i(k)} & 1 & 0 & \ldots & 0 \\ \lambda_{i(k)} & \lambda_{i(k)} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & \lambda_{i(k)} & 1 & \lambda_{i(k)} \end{pmatrix} \in \mathcal{M}(s_k, F) \quad \text{for } s_k > 1.$$
(b) In the situation of (a) and recalling from Def. 6.14 that, for each \( \lambda \in \sigma(A) \), \( r(\lambda) \) is such that \( m_\lambda(\lambda) = \dim \ker(A - \lambda \text{Id})^{r(\lambda)} \) and, for each \( s \in \{1, \ldots, r(\lambda)\} \),

\[
E_A^s(\lambda) = \ker(A - \lambda \text{Id})^s
\]

is called the corresponding generalized eigenspace of rank \( s \) of \( A \), each \( v \in E_A^s(\lambda) \setminus E_A^{s-1}(\lambda) \), \( s \geq 2 \), is called a generalized eigenvector of rank \( s \), we obtain

\[
\forall i \in \{1, \ldots, m\} \quad r_i = r(\lambda_i),
\]

\[
\forall k \in \{1, \ldots, l\} \quad U_k \subseteq E_A^{r(k)}(\lambda_{i(k)}) \subseteq E_A^{r_i(k)}(\lambda_{i(k)}).
\]

Moreover, for each \( k \in \{1, \ldots, l\} \), \( v_k \) is a generalized eigenvector of rank \( s_k \) and the basis \( J_k \) consists of generalized eigenvectors, containing precisely one generalized eigenvector of rank \( s \) for each \( s \in \{1, \ldots, s_k\} \). Define

\[
\forall i \in \{1, \ldots, m\} \quad \forall s \in \mathbb{N} \quad l(i, s) := \#\{k \in \{1, \ldots, l\} : U_k \subseteq \ker(A - \lambda_i \text{Id})^r, \dim U_k = s\}
\]

(\text{thus,} \( l(i, s) \) \text{is the number of Jordan blocks of size} \( s \) \text{corresponding to the eigenvalue} \( \lambda_i \) \text{- apart from the slightly different notation used here, the} \( l(i, s) \) \text{are precisely the numbers called} \( l_k \) \text{in Prop. 9.8(b)}.). Then, for each \( i \in \{1, \ldots, m\} \),

\[
\forall i \in \{1, \ldots, m\} \quad \forall s \in \mathbb{N} \quad l(i, s) = \dim \ker(A - \lambda_i \text{Id})^s - \dim \ker(A - \lambda_i \text{Id})^{s-1} - \sum_{j=s+1}^{r_i} l(i, j) (j - (s - 1)).
\]

Thus, in general, one needs to determine, for each \( i \in \{1, \ldots, m\} \) and each \( s \in \{1, \ldots, r_i\} \), \( \dim \ker(A - \lambda_i \text{Id})^s \) to know the precise structure of \( N \).

(c) For the sake of completeness and convenience, we restate Th. 9.9(b): If

\[
V = \bigoplus_{i=1}^L W_i
\]

is another decomposition of \( V \) into \( A \)-invariant and \( A \)-irreducible subspaces \( W_1, \ldots, W_L \) of \( V \), \( L \in \mathbb{N} \), then \( L = l \) and there exist a permutation \( \pi \in S_l \) and \( T \in \text{GL}(V) \) such that

\[
TA = AT \quad \forall i \in \{1, \ldots, l\} \quad T(U_i) = W_{\pi(i)}.
\]

Proof. (a),(b): The \( A \)-cyclic and \( A \)-irreducible subspaces \( U_1, \ldots, U_l \) are given by Th. 9.9(a). As in Th. 9.9(a), for each \( k \in \{1, \ldots, l\} \), let \( v_k \in U_k \) be such that

\[
B_k := \{v_k, Av_k, \ldots, A^{s_k-1}v_k\}, \quad s_k := \dim U_k.
\]
Consider the matrices $J$ that Proposition 9.10(a),(c) yields $U$ with respect to the ordered basis $(w_1,\ldots,w_l)$. Finally, (9.15), i.e. the formulas for the $V_i$ and (9.16b) were inferred from (9.9), using that, in the current situation, $g = X - \lambda_i$, $V_i := \ker(A - \lambda_i X)^r$, and (with $V_i := \ker(A - \lambda_i X)^r$,)
\[
\forall i \in \{1,\ldots,m\} \quad \dim \ker(A - \lambda_i X)^r = \dim \ker(A - \lambda_i X)^s.
\]
(c) was already proved as it is merely a restatement of Th. 9.9(b). 

Example 9.14. (a) In Ex. 9.12(a), we considered $V := F^6$ ($F$ some field) and $A \in \mathcal{L}(V,V)$ with
\[
\chi_A = (X - 2)^2(3 - 3)^4, \quad \mu_A = (X - 2)^2(3 - 3)^3.
\]
We obtained $V = \bigoplus_{i=1}^{3} U_i$ with $U_1 \subseteq \ker(A - 2 X)^2$ with $\dim U_1 = 2$ and $U_2, U_3 \subseteq \ker(A - 3 X)^3$ with $\dim U_2 = 3$, $\dim U_3 = 1$. Thus, the corresponding matrix in Jordan normal form is
\[
N = \begin{pmatrix}
2 & 1 \\
3 & 1 \\
3 & 3
\end{pmatrix}.
\]

(b) Consider the matrices
\[
N_1 := \begin{pmatrix}
2 & 1 & 2 \\
2 & 1 & 2 \\
2 & 1 & 2
\end{pmatrix}, \quad N_2 := \begin{pmatrix}
2 & 1 & 2 \\
2 & 1 & 2 \\
2 & 1 & 2
\end{pmatrix}
\]
over some field \( F \). Both matrices are in Jordan normal form with
\[
\chi_{N_1} = \chi_{N_2} = (X - 2)^8, \quad \mu_{N_1} = \mu_{N_2} = (X - 2)^3.
\]
Both have the same total number of Jordan blocks, namely 4, which corresponds to
\[
\dim \ker(N_1 - 2 \text{Id}) = \dim \ker(N_2 - 2 \text{Id}) = 4.
\]
The differences appear in the generalized eigenspaces of higher order: \( N_1 \) has two linearly independent generalized eigenvectors of rank 2, whereas \( N_2 \) has three linearly independent generalized eigenvectors of rank 2, yielding
\[
\dim \ker(N_1 - 2 \text{Id})^2 = \dim \ker(N_1 - 2 \text{Id})^2 = 2, \quad \text{i.e.} \quad \dim \ker(N_1 - 2 \text{Id})^2 = 6,
\]
\[
\dim \ker(N_2 - 2 \text{Id})^2 = \dim \ker(N_2 - 2 \text{Id})^2 = 3, \quad \text{i.e.} \quad \dim \ker(N_2 - 2 \text{Id})^2 = 7.
\]
Next, \( N_1 \) has two linearly independent generalized eigenvectors of rank 3, whereas \( N_2 \) has one linearly independent generalized eigenvector of rank 3, yielding
\[
\dim \ker(N_1 - 2 \text{Id})^3 = \dim \ker(N_1 - 2 \text{Id})^3 = 2, \quad \text{i.e.} \quad \dim \ker(N_1 - 2 \text{Id})^3 = 8,
\]
\[
\dim \ker(N_2 - 2 \text{Id})^3 = \dim \ker(N_2 - 2 \text{Id})^3 = 1, \quad \text{i.e.} \quad \dim \ker(N_2 - 2 \text{Id})^3 = 8.
\]
From (9.15a), we obtain (with \( i = 1 \))
\[
l_{N_1}(1, 3) = 2, \quad l_{N_2}(1, 3) = 1,
\]
corresponding to \( N_1 \) having two blocks of size 3 and \( N_2 \) having one block of size 3. From (9.15b), we obtain (with \( i = 1 \))
\[
l_{N_1}(1, 2) = 8 - 4 - 2(3 - 1) = 0, \quad l_{N_2}(1, 2) = 8 - 4 - 1(3 - 1) = 2,
\]
corresponding to \( N_1 \) having no blocks of size 2 and \( N_2 \) having two blocks of size 2.
To check consistency, we use (9.15b) again to obtain
\[
l_{N_1}(1, 1) = 8 - 0 - 0(2 - 0) - 2(3 - 0) = 2, \quad l_{N_2}(1, 1) = 8 - 0 - 2(2 - 0) - 1(3 - 0) = 1,
\]
corresponding to \( N_1 \) having two blocks of size 1 and \( N_2 \) having one block of size 1.

**Remark 9.15.** In the situation of Th. 9.13, we saw that, in order to find a matrix \( N \) in Jordan normal form for \( A \), according to Th. 9.13(b), in general, for each \( i \in \{1, \ldots, m\} \) and each \( s \in \{1, \ldots, r_i\} \), one has to know \( \dim \ker(A - \lambda_i \text{Id})^s \). On the other hand, given a matrix \( M \) of \( A \), one might also want to find the transition matrix \( T \in \text{GL}_n(F) \) such that \( M = TNT^{-1} \). As it turns out, if one has already determined generalized eigenvectors forming bases of \( \ker(A - \lambda_i \text{Id})^s \), one may use these same vectors for the columns of \( T \): Indeed, if \( M = TNT^{-1} \) and \( t_1, \ldots, t_n \) denote the columns of \( T \), then, if \( j \in \{1, \ldots, n\} \) corresponds to a column of \( N \) with \( \lambda_i \) being the only nonzero entry, then
\[
Mt_j = TNT^{-1}t_j = TNe_j = T\lambda_i e_j = \lambda_i t_j,
\]
showing \( t_j \) to be a corresponding eigenvector. If \( j \in \{1, \ldots, n\} \) corresponds to a column of \( N \) having nonzero entries \( \lambda_i \) and 1 (above the \( \lambda_i \)), then
\[
(M - \lambda_i \text{Id}_n)t_j = (TNT^{-1} - \lambda_i \text{Id}_n)t_j = TNe_j - \lambda_i t_j = T(e_{j-1} + \lambda_i e_j) - \lambda_i t_j = t_{j-1} + \lambda_i t_j - \lambda_i t_j = t_{j-1}
\]
showing \( t_j \) to be a generalized eigenvector of rank \( \geq 2 \) corresponding to the Jordan block containing the index \( j \).
10 Vector Spaces with Scalar Products

In this section, the field $F$ will always be $\mathbb{R}$ or $\mathbb{C}$. As before, we write $K$ if $K$ may stand for $\mathbb{R}$ or $\mathbb{C}$.

10.1 Definition, Orthogonality

**Definition 10.1.** Let $X$ be a vector space over $K$. A function $\langle \cdot, \cdot \rangle : X \times X \rightarrow K$ is called an *inner product* or a *scalar product* on $X$ if, and only if, the following three conditions are satisfied:

(i) $\langle x, x \rangle \in \mathbb{R}^+$ for each $0 \neq x \in X$ (i.e. an inner product is *positive definite*).

(ii) $\langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle$ for each $x, y, z \in X$ and each $\lambda, \mu \in K$ (i.e. an inner product is $K$-linear in its first argument).

(iii) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ for each $x, y \in X$ (i.e. an inner product is *conjugate-symmetric*, even symmetric for $K = \mathbb{R}$).

**Lemma 10.2.** For each inner product $\langle \cdot, \cdot \rangle$ on a vector space $X$ over $K$, the following formulas are valid:

(a) $\langle x, \lambda y + \mu z \rangle = \overline{\lambda} \langle y, x \rangle + \overline{\mu} \langle z, x \rangle$ for each $x, y, z \in X$ and each $\lambda, \mu \in K$, i.e. $\langle \cdot, \cdot \rangle$ is conjugate-linear in its second argument, even linear for $K = \mathbb{R}$. Together with Def. 10.1(ii), this means that $\langle \cdot, \cdot \rangle$ is a sesquilinear form, even a bilinear form for $K = \mathbb{R}$.

(b) $\langle 0, x \rangle = \overline{\langle x, 0 \rangle} = 0$ for each $x \in X$.

**Proof.** (a): One computes, for each $x, y, z \in X$ and each $\lambda, \mu \in K$,

$$\langle x, \lambda y + \mu z \rangle \overset{\text{Def. 10.1(iii)}}{=} \overline{\lambda} \langle y, x \rangle + \overline{\mu} \langle z, x \rangle \overset{\text{Def. 10.1(ii)}}{=} \overline{\lambda} \langle y, x \rangle + \overline{\mu} \langle z, x \rangle \overset{\text{Def. 10.1(iii)}}{=} \overline{\lambda} \langle y, x \rangle + \overline{\mu} \langle z, x \rangle.$$

(b): One computes, for each $x \in X$,

$$\langle x, 0 \rangle \overset{\text{Def. 10.1(ii)}}{=} \langle 0, x \rangle = \langle 0 x, x \rangle \overset{\text{Def. 10.1(ii)}}{=} 0 \langle x, x \rangle = 0,$$

thereby completing the proof of the lemma.

**Remark 10.3.** If $X$ is a vector space over $K$ with an inner product $\langle \cdot, \cdot \rangle$, then the map $\| \cdot \| : X \rightarrow \mathbb{R}_0^+$, $\| x \| := \sqrt{\langle x, x \rangle}$, defines a norm on $X$ (cf. [Phi16b, Prop. 1.65]). One calls this the norm *induced* by the inner product.
Definition 10.4. Let $X$ be a vector space over $\mathbb{K}$. If $\langle \cdot, \cdot \rangle$ is an inner product on $X$, then $(X, \langle \cdot, \cdot \rangle)$ is called an inner product space or a pre-Hilbert space. An inner product space is called a Hilbert space if, and only if, $(X, \| \cdot \|)$ is a Banach space, where $\| \cdot \|$ is the induced norm, i.e. $\|x\| := \sqrt{\langle x, x \rangle}$. Frequently, the inner product on $X$ is understood and $X$ itself is referred to as an inner product space or Hilbert space.

Example 10.5. (a) On the space $\mathbb{K}^n$, $n \in \mathbb{N}$, we define an inner product by letting, for each $z = (z_1, \ldots, z_n) \in \mathbb{K}^n$, $w = (w_1, \ldots, w_n) \in \mathbb{K}^n$:

$$ z \cdot w := \sum_{j=1}^n z_j \overline{w_j} \quad (10.1) $$

(called the standard inner product on $\mathbb{K}^n$, also the Euclidean inner product for $\mathbb{K} = \mathbb{R}$). Let us verify that (10.1), indeed, defines an inner product in the sense of Def. 10.1: If $z \neq 0$, then there is $j_0 \in \{1, \ldots, n\}$ such that $z_{j_0} \neq 0$. Thus, $z \cdot z = \sum_{j=1}^n |z_j|^2 \geq |z_{j_0}|^2 > 0$, i.e. Def. 10.1(i) is satisfied. Next, let $z, w, u \in \mathbb{K}^n$ and $\lambda, \mu \in \mathbb{K}$. One computes

$$ (\lambda z + \mu w) \cdot u = \sum_{j=1}^n (\lambda z_j + \mu w_j) \overline{u_j} = \sum_{j=1}^n \lambda z_j \overline{u_j} + \sum_{j=1}^n \mu w_j \overline{u_j} = \lambda (z \cdot u) + \mu (w \cdot u), $$

i.e. Def. 10.1(ii) is satisfied. For Def. 10.1(iii), merely note that

$$ z \cdot w = \sum_{j=1}^n z_j \overline{w_j} = \sum_{j=1}^n w_j \overline{z_j} = \overline{w} \cdot z. $$

Hence, we have shown that (10.1) defines an inner product according to Def. 10.1. Due to [Phi16b, Prop. 1.59(b)], the induced norm is complete, i.e. $\mathbb{K}^n$ with the inner product of (10.1) is a Hilbert space.

(b) Let $a, b \in \mathbb{R}$, $a < b$. We define an inner product on the space $X := C[a,b]$ of continuous $\mathbb{K}$-valued functions on $[a,b]$ by letting

$$ \langle \cdot, \cdot \rangle : X \times X \longrightarrow \mathbb{K}, \quad \langle f, g \rangle := \int_a^b \overline{f} \, g : \quad (10.2) $$

We verify that (10.2), indeed, defines an inner product: If $f \in X$, $f \neq 0$, then there exists $t \in [a,b]$ such that $|f(t)| = \alpha \in \mathbb{R}^+$. As $f$ is continuous, there exists $\delta > 0$ such that

$$ \forall \, s \in [a,b] \cap [t-\delta, t+\delta] \quad |f(s)| > \frac{\alpha}{2}, $$

implying

$$ \langle f, f \rangle = \int_a^b |f|^2 \geq \int_{[a,b] \cap [t-\delta, t+\delta]} |f|^2 > 0. $$
Next, let $f, g, h \in X$ and $\lambda, \mu \in \mathbb{K}$. One computes

$$\langle \lambda f + \mu g, h \rangle = \int_{\Omega} (\lambda f + \mu g) \overline{h} \, d\mu = \lambda \int_{\Omega} f \overline{h} \, d\mu + \mu \int_{\Omega} g \overline{h} \, d\mu = \lambda \langle f, h \rangle + \mu \lambda \langle g, h \rangle.$$  

Moreover,

$$\forall f, g \in X \quad \langle f, g \rangle = \int_{\Omega} f \overline{g} \, d\mu = \int_{\Omega} \overline{f} g \, d\mu = \langle g, f \rangle,$$

showing $\langle \cdot, \cdot \rangle$ to be an inner product on $X$. As it turns out, $(X, \| \cdot \|)$ is an example of an inner product space that is not a Hilbert space: With respect to the norm induced by $\langle \cdot, \cdot \rangle$ (usually called the 2-norm, denoted $\| \cdot \|_2$), $X$ is dense in $L^2[a,b]$ (the space of square-integrable functions with respect to Lebesgue (or Borel) measure on $[a,b]$, cf., e.g. [Phi17a, Th. 2.49(a)]), which is the completion of $X$ with respect to $\| \cdot \|_2$.

**Definition 10.6.** Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{K}$.

(a) $x, y \in X$ are called orthogonal or perpendicular (denoted $x \perp y$) if, and only if, $\langle x, y \rangle = 0$.

(b) Let $E \subseteq X$. Define the perpendicular space $E^\perp$ to $E$ (called $E_{\perp}$) by

$$E^\perp := \left\{ y \in X : \forall x \in E \quad \langle x, y \rangle = 0 \right\}.$$  

(10.3)

Caveat: As is common, we use the same symbol to denote the perpendicular space that we used to denote the forward annihilator in Def. 2.13, even though these objects are not the same: The perpendicular space is a subset of $X$, whereas the forward annihilator is a subset of $X'$. In the following, when dealing with inner product spaces, $E^\perp$ will always mean the perpendicular space.

(c) If $X = V_1 \oplus V_2$ with subspaces $V_1, V_2$ of $X$, then we call $X$ the orthogonal sum of $V_1, V_2$ if, and only if, $v_1 \perp v_2$ for each $v_1 \in V_1, v_2 \in V_2$. In this case, we also write $X = V_1 \perp V_2$.

(d) Let $S \subseteq X$. Then $S$ is an orthogonal system if, and only if, $x \perp y$ for each $x, y \in S$ with $x \neq y$. A unit vector is $x \in X$ such that $\|x\| = 1$ (with respect to the induced norm on $X$). Then $S$ is called an orthonormal system if, and only if, $S$ is an orthogonal system consisting entirely of unit vectors. Finally, $S$ is called an orthonormal basis if, and only if, it is a maximal orthonormal system in the sense that, if $S \subseteq T \subseteq X$ and $T$ is an orthonormal system, then $S = T$ (caveat: if $X$ is an infinite-dimensional Hilbert space, then an orthonormal basis of $X$ is not(!) a vector space basis of $X$).

**Lemma 10.7.** Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{K}$, $E \subseteq X$.

(a) $E \cap E^\perp \subseteq \{0\}$. 

(b) $E^\perp$ is a subspace of $X$.

(c) $X^\perp = \{0\}$ and $\{0\}^\perp = X$.

Proof. (a): If $x \in E \cap E^\perp$, then $\langle x, x \rangle = 0$, implying $x = 0$.

(b): We have $0 \in E^\perp$ and

$$\forall \lambda, \mu \in \mathbb{K} \forall y_1, y_2 \in E^\perp \forall x \in E \langle x, \lambda y_1 + \mu y_2 \rangle = \langle x, \lambda y_1 \rangle + \langle x, \mu y_2 \rangle = 0,$$

showing $\lambda y_1 + \mu y_2 \in E^\perp$, i.e. $E^\perp$ is a subspace of $X$.

(c): Is $x \in X^\perp$, then $0 = \langle x, x \rangle$, implying $x = 0$. On the other hand, $\langle x, 0 \rangle = 0$ for each $x \in X$ by Lem. 10.2(b).

Proposition 10.8. Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{K}$ and let $S \subseteq X$ be an orthogonal system.

(a) $S \setminus \{0\}$ is linearly independent.

(b) Pythagoras’ Theorem: If $s_1, \ldots, s_n \in S$ are distinct, $n \in \mathbb{N}$, then

$$\left\| \sum_{i=1}^{n} s_i \right\|^2 = \sum_{i=1}^{n} \| s_i \|^2,$$

where $\| \cdot \|$ is the induced norm on $X$.

Proof. (a): Suppose $n \in \mathbb{N}$ and $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ together with $s_1, \ldots, s_n \in S \setminus \{0\}$ distinct are such that $\sum_{i=1}^{n} \lambda_i s_i = 0$. Then, as $\langle s_k, s_j \rangle = 0$ for each $k \neq j$, we obtain

$$\forall j \in \{1, \ldots, n\} \quad 0 = \langle 0, s_j \rangle = \left\langle \sum_{i=1}^{n} \lambda_i s_i, s_j \right\rangle = \sum_{i=1}^{n} \lambda_i \langle s_i, s_j \rangle = \lambda_j \langle s_j, s_j \rangle,$$

which yields $\lambda_j = 0$ by Def. 10.1(i). Thus, we have shown that $\lambda_j = 0$ for each $j \in \{1, \ldots, n\}$, which establishes the linear independence of $S \setminus \{0\}$.

(b): We compute

$$\left\| \sum_{i=1}^{n} s_i \right\|^2 = \left\langle \sum_{i=1}^{n} s_i, \sum_{j=1}^{n} s_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \langle s_i, s_j \rangle = \sum_{i=1}^{n} \langle s_i, s_i \rangle = \sum_{i=1}^{n} \| s_i \|^2,$$

thereby establishing the case.

To obtain orthogonal systems and orthonormal systems in inner product spaces, one can apply the algorithm provided by the following Th. 10.9:
\[ v_0 := x_0, \quad v_n := x_n - \sum_{k=0, v_k \neq 0}^{n-1} \frac{\langle x_n, v_k \rangle}{\|v_k\|^2} v_k \quad (10.4) \]

for each \( n \in \mathbb{N} \), additionally assuming that \( n \) is less than or equal to the max index of the sequence \( x_0, x_1, \ldots \) if the sequence is finite. Then the set \( \{v_0, v_1, \ldots\} \) constitutes an orthogonal system. Of course, by omitting the \( v_k = 0 \) and by dividing each \( v_k \neq 0 \) by its norm, one can also obtain an orthonormal system (nonempty if at least one \( v_k \neq 0 \).

Moreover, \( v_n = 0 \) if, and only if, \( x_n \in \text{span}\{x_0, \ldots, x_{n-1}\} \). In particular, if the \( x_0, x_1, \ldots \) are all linearly independent, then so are the \( v_0, v_1, \ldots \).

**Proof.** We show by induction on \( n \in \mathbb{N}_0 \), that, for each \( 0 \leq m < n \), \( v_n \perp v_m \). For \( n = 0 \), there is nothing to show. Thus, let \( n > 0 \) and \( 0 \leq m < n \). By induction, \( \langle v_k, v_m \rangle = 0 \) for each \( 0 \leq k, m < n \) such that \( k \neq m \). For \( v_i = 0 \), \( \langle v_i, v_m \rangle = 0 \) is clear. Otherwise,

\[ \langle v_n, v_m \rangle = \left\langle x_n - \sum_{k=0, v_k \neq 0}^{n-1} \frac{\langle x_n, v_k \rangle}{\|v_k\|^2} v_k, v_m \right\rangle = \langle x_n, v_m \rangle - \frac{\langle x_n, v_m \rangle}{\|v_m\|^2} \langle v_m, v_m \rangle = 0, \]

thereby establishing the case. So we know that \( v_0, v_1, \ldots \) constitutes an orthogonal system. Next, by induction, for each \( n \), we obtain \( v_n \in \text{span}\{x_0, \ldots, x_n\} \) directly from (10.4). Thus, \( v_n = 0 \) implies \( x_n = \sum_{k=0, v_k \neq 0}^{n-1} \frac{\langle x_n, v_k \rangle}{\|v_k\|^2} v_k \in \text{span}\{x_0, \ldots, x_{n-1}\} \). Conversely, if \( x_n \in \text{span}\{x_0, \ldots, x_{n-1}\} \), then

\[
\dim \text{span}\{v_0, \ldots, v_{n-1}, v_n\} = \dim \text{span}\{x_0, \ldots, x_{n-1}, x_n\} = \dim \text{span}\{x_0, \ldots, x_{n-1}\} = \dim \text{span}\{v_0, \ldots, v_{n-1}\},
\]

which, due to Prop. 10.8(a), implies \( v_n = 0 \). Finally, if all \( x_0, x_1, \ldots \) are linearly independent, then all \( v_k \neq 0 \), \( k = 0, 1, \ldots \), such that the \( v_0, v_1, \ldots \) are linearly independent by Prop. 10.8(a). \( \blacksquare \)

**Example 10.10.** In the space \( C[-1, 1] \) with the inner product according to Ex. 10.5(b), consider

\[
\forall \quad x_i : [-1, 1] \rightarrow \mathbb{K}, \quad x_i(x) := x^i.
\]

We check that the first four orthogonal polynomials resulting from applying (10.4) to \( x_0, x_1, \ldots \) are given by

\[
v_0(x) = 1, \quad v_1(x) = x, \quad v_2(x) = x^2 - \frac{1}{3}, \quad v_3(x) = x^3 - \frac{3}{5} x.
\]

One has \( v_0 = x_0 \equiv 1 \) and, then, obtains successively from (10.4):

\[
v_1(x) = x_1(x) - \frac{\langle x_1, v_0 \rangle}{\|v_0\|^2} v_0(x) = x - \frac{\langle x_1, v_0 \rangle}{\|v_0\|^2} = x - \frac{\int_{-1}^{1} x \, dx}{\int_{-1}^{1} \, dx} = x.
\]
\( v_2(x) = x_2(x) - \frac{\langle x_2, v_0 \rangle}{\|v_0\|^2} v_0(x) - \frac{\langle x_2, v_1 \rangle}{\|v_1\|^2} v_1(x) = x^2 - \frac{\int_{-1}^{1} x^2 \, dx}{2} - \frac{\int_{-1}^{1} x^3 \, dx}{2} x \)

\( = x^2 - \frac{1}{3} \cdot \frac{x}{3} \).

\( v_3(x) = x_3(x) - \frac{\langle x_3, v_0 \rangle}{\|v_0\|^2} v_0(x) - \frac{\langle x_3, v_1 \rangle}{\|v_1\|^2} v_1(x) - \frac{\langle x_3, v_2 \rangle}{\|v_2\|^2} v_2(x) \)

\( = x^3 - \frac{\int_{-1}^{1} x^3 \, dx}{2} - \int_{-1}^{1} x^4 \, dx - \frac{\int_{-1}^{1} x^3 (x^2 - \frac{1}{3}) \, dx}{\int_{-1}^{1} (x^2 - \frac{1}{3})^2 \, dx} \left( x^2 - \frac{1}{3} \right) \)

\( = x^3 - \frac{2}{7} x = x^3 - \frac{3}{5} x. \)

**Definition 10.11.** Let \((X, \langle \cdot, \cdot \rangle)\) and \((Y, \langle \cdot, \cdot \rangle)\) be inner product space over \(\mathbb{K}\). If \(A \in \mathcal{L}(X, Y)\) is a linear isomorphism, then we call \(A\) isometric (and \(X\) and \(Y\) isometrically isomorphic) if, and only if

\[ \forall x_1, x_2 \in X \quad \langle Ax_1, Ax_2 \rangle = \langle x_1, x_2 \rangle \]

(due to the fact that \(A\) then preserves the metric induced by the induced norm).

**Theorem 10.12.** Let \((X, \langle \cdot, \cdot \rangle)\) be a finite-dimensional inner product space over \(\mathbb{K}\).

(a) An orthonormal system \(S \subseteq X\) is an orthonormal basis of \(X\) if, and only if, \(S\) is a vector space basis of \(X\).

(b) \(X\) has an orthonormal basis.

(c) If \((Y, \langle \cdot, \cdot \rangle)\) is an inner product space over \(\mathbb{K}\), then \(X\) and \(Y\) are isometrically isomorphic if, and only if, \(\dim X = \dim Y\).

(d) If \(U\) is a subspace of \(X\), then the following holds:

(i) \(X = U \perp U^\perp\).

(ii) \(\dim U^\perp = \dim X - \dim U\).

(iii) \((U^\perp)^\perp = U\).

**Proof.** (a): Let \(S\) be an orthonormal system. If \(S\) is an orthonormal basis, then, by Prop. 10.8(a) and Th. 10.9, \(S\) is a maximal linearly independent subset of \(X\), showing \(S\) to be a vector space basis of \(X\). Conversely, if \(S\) is a vector space basis of \(X\), then, again using Prop. 10.8(a), \(S\) must be a maximal orthonormal system, i.e. an orthonormal basis of \(X\).

(b): According to Th. 10.9, applying Gram-Schmidt orthogonalization to a vector space basis of \(X\) yields an orthonormal basis of \(X\).

(c): We already know that the existence of a linear isomorphism between \(X\) and \(Y\) implies \(\dim X = \dim Y\). Conversely, assume \(n := \dim X = \dim Y \in \mathbb{N}\), and let \(B_X = \)
then the map

\[ \psi : X \rightarrow X', \quad \psi(y) := \alpha_y, \]

where

\[ \alpha_y : X \rightarrow \mathbb{K}, \quad \alpha_y(a) := (a, y), \]

is bijective and conjugate-linear (in particular, each \( \alpha \in X' \) can be represented by \( y \in X \), and, if \( \mathbb{K} = \mathbb{R} \), then \( \psi \) is a linear isomorphism).

**Proof.** According to Def. 10.1(ii), for each \( y \in X \), \( \alpha_y \) is linear, i.e. \( \psi \) is well-defined. Moreover,

\[
\forall \lambda, \mu \in \mathbb{K} \quad \forall y_1, y_2 \in X \quad \forall a \in X \\
\psi(\lambda y_1 + \mu y_2)(a) = \langle a, \lambda y_1 + \mu y_2 \rangle = \overline{\lambda} \langle a, y_1 \rangle + \overline{\mu} \langle a, y_2 \rangle = (\overline{\lambda} \psi(y_1) + \overline{\mu} \psi(y_2))(a),
\]

\[
\psi(y + z)(a) = \langle a, y + z \rangle = \langle a, y \rangle + \langle a, z \rangle = \psi(y)(a) + \psi(z)(a).
\]
showing $\psi$ to be conjugate-linear. It remains to show $\psi$ is bijective. Let $\{x_1, \ldots, x_n\}$ be a basis of $X$, $\dim X = n$. It suffices to show that $B := \{\alpha_{x_1}, \ldots, \alpha_{x_n}\}$ is a basis of $X'$. As $\dim X = \dim X'$, it even suffices to show that $B$ is linearly independent. To this end, let $\lambda_1, \ldots, \lambda_n \in \mathbb{K}$ be such that $0 = \sum_{i=1}^n \lambda_i \alpha_{x_i}$. Then, for each $v \in X$,

$$0 = \left( \sum_{i=1}^n \lambda_i \alpha_{x_i} \right) v = \sum_{i=1}^n \lambda_i \alpha_{x_i} v = \sum_{i=1}^n \lambda_i \langle v, x_i \rangle = \left( v, \sum_{i=1}^n \lambda_i x_i \right),$$

showing $\sum_{i=1}^n \lambda_i x_i \in X^\perp = \{0\}$. As the $x_i$ are linearly independent, this yields $\lambda_1 = \cdots = \lambda_n = 0$ and the desired linear independence of $B$. \hfill \blacksquare

**Remark 10.14.** The above Th. 10.13 is a finite-dimensional version of the Riesz representation theorem for Hilbert spaces, which states that Th. 10.13 remains true if $X$ is an infinite-dimensional Hilbert space and $X'$ is replaced by the topological dual $X'_\text{top}$ of $X$, consisting of all linear forms on $X$ that are also continuous with respect to the induced norm on $X$, cf. [Phi17c, Ex. 3.1] (recall that, if $X$ is finite-dimensional, then all linear forms on $X$ are automatically continuous, cf. [Phi16b, Ex. 2.16]).

### 10.2 The Adjoint Map

**Definition 10.15.** Let $(X_1, \langle \cdot, \cdot \rangle)$, $(X_2, \langle \cdot, \cdot \rangle)$ be a finite-dimensional inner product space over $\mathbb{K}$. Moreover, let $A \in \mathcal{L}(X_1, X_2)$, let $A' \in \mathcal{L}(X_2', X_1')$ be the dual map according to Def. 2.26, and let $\psi_1 : X_1 \rightarrow X_1'$, $\psi_2 : X_2 \rightarrow X_2'$ be the maps given by the Th. 10.13. Then the map

$$A^* : X_2 \rightarrow X_1, \quad A^* := \psi_1^{-1} \circ A' \circ \psi_2,$$

(10.7)

is called the adjoint map of $A$.

**Theorem 10.16.** Let $(X_1, \langle \cdot, \cdot \rangle)$, $(X_2, \langle \cdot, \cdot \rangle)$ be finite-dimensional inner product spaces over $\mathbb{K}$. Let $A \in \mathcal{L}(X_1, X_2)$.

(a) One has $A^* \in \mathcal{L}(X_2, X_1)$, and $A^*$ is the unique map $X_2 \rightarrow X_1$ such that

$$\forall x \in X_1 \quad \forall y \in X_2 \quad \langle Ax, y \rangle = \langle x, A^* y \rangle.$$  

(10.8)

(b) One has $A^{**} = A$.

(c) One has that $A \mapsto A^*$ is a conjugate-linear bijection of $\mathcal{L}(X_1, X_2)$ onto $\mathcal{L}(X_2, X_1)$.

(d) $(\text{Id}_{X_1})^* = \text{Id}_{X_1}$.

(e) If $(X_3, \langle \cdot, \cdot \rangle)$ is another finite-dimensional inner product space over $\mathbb{K}$ and $B \in \mathcal{L}(X_2, X_3)$, then

$$\left( B \circ A \right)^* = A^* \circ B^*.$$  

(f) $\ker(A^*) = (\text{Im} A)^\perp$ and $\text{Im}(A^*) = (\ker A)^\perp$. 


(g) $A$ is a monomorphism if, and only if, $A^*$ is an epimorphism.

(h) $A$ is a an epimorphism if, and only if, $A^*$ is a monomorphism.

(i) $A^{-1} \in \mathcal{L}(X_2, X_1)$ exists if, and only if, $(A^*)^{-1} \in \mathcal{L}(X_1, X_2)$ exists, and, in that case,

$$(A^*)^{-1} = (A^{-1})^*.$$

Proof. (a): Let $A \in \mathcal{L}(X_1, X_2)$. Then $A^* \in \mathcal{L}(X_2, X_1)$, since $A'$ is linear, and $\psi_1^{-1}$ and $\psi_2$ are both conjugate-linear. Moreover, we know $A'$ is the unique map on $X'_2$ such that

$$\forall x \in X_1 \forall y \in X_2, \quad A'(\beta)(x) = \beta(A(x)).$$

Thus, for each $x \in X_1$ and each $y \in X_2$,

$$\langle x, A^* y \rangle = \langle x, (\psi_1^{-1} \circ A' \circ \psi_2)(y) \rangle = \psi_1((\psi_1^{-1} \circ A' \circ \psi_2)(y))(x)$$

$$= A'(\psi_2(y))(x) = \psi_2(y)(Ax) = \langle Ax, y \rangle,$$

proving (10.8). For each $y \in X_2$ and each $x \in X_1$, we have $(Ax, y) = (\psi_2(y))(Ax) = ((\psi_2(y) \circ A)(x)$. Then Th. 10.13 and (10.8) imply $A^*(y) = \psi_1^{-1}((\psi_2(y) \circ A)$, showing $A^*$ to be uniquely determined by (10.8).

(b): According to (a), $A^{**}$ is the unique map $X_1 \rightarrow X_2$ such that

$$\forall x \in X_1 \forall y \in X_2, \quad \langle A^* y, x \rangle = \langle y, A^{**} x \rangle.$$

Comparing with (10.8) yields $A = A^{**}$.

(c): If $A, B \in \mathcal{L}(X_1, X_2)$ and $\lambda \in \mathbb{K}$, then, for each $y \in X_2$,

$$(A + B)^*(y) = (\psi_1^{-1} \circ (A + B)' \circ \psi_2)(y) = (\psi_1^{-1} \circ (A' + B') \circ \psi_2)(y)$$

$$= (\psi_1^{-1} \circ A' \circ \psi_2)(y) + (\psi_1^{-1} \circ B' \circ \psi_2)(y) = (A^* + B^*)(y)$$

and

$$(\lambda A)^*(y) = (\psi_1^{-1} \circ (\lambda A)' \circ \psi_2)(y) = \overline{\lambda}(\psi_1^{-1} \circ A' \circ \psi_2)(y) = (\overline{\lambda}A^*)(y),$$

showing $A \mapsto A^*$ to be conjugate-linear. Moreover, $A \mapsto A^*$ is bijective due to (b).

(d): One has $(\text{Id}_{X_1})^* = \psi_1^{-1} \circ (\text{Id}_{X_1})' \circ \psi_1 = \psi_1^{-1} \circ \text{Id}_{X'_1} \circ \psi_1 = \text{Id}_{X_1}$.

(e): Let $\psi_3 : X_3 \rightarrow X'_3$ be given by Th. 10.13. Then

$$A^* \circ B^* = \psi_1^{-1} \circ A' \circ \psi_2 \circ \psi_3 \circ B' \circ \psi_3 = \psi_1^{-1} \circ (B \circ A)' \circ \psi_3 = (B \circ A)^*.$$

(f): We have

$$y \in \ker(A^*) \iff \forall x \in X_1 \langle x, A^* y \rangle = 0 \iff \forall x \in X_1 \langle Ax, y \rangle = 0 \iff y \in (\text{Im} A)^\perp.$$

Applying the first part with $A$ replaced by $A^*$ yields

$$\ker A = \ker A^{**} = (\text{Im}(A^*))^\perp.$$
and, thus, using Th. 10.12(d)(iii),

\[ \text{Im}(A^*) = (\text{Im}(A^*))^\perp = (\ker A)^\perp. \]

(g): According to (f), we have

\[ \ker A = \{0\} \iff \text{Im}(A^*) = (\ker A)^\perp = \{0\}^\perp = X_1. \]

(h): According to (f), we have

\[ \ker(A^*) = \{0\} \iff \text{Im} A = (\ker(A^*))^\perp = \{0\}^\perp = X_2. \]

(i): One has

\[ A^{-1} \in \mathcal{L}(X_2, X_1) \text{ exists} \iff (A^{-1})' = (A')^{-1} \in \mathcal{L}(X'_1, X'_2) \text{ exists} \]

\[ \iff (A^*)^{-1} = (\psi_1^{-1} \circ A' \circ \psi_2)^{-1} \in \mathcal{L}(X_1, X_2) \text{ exists.} \]

Moreover, if \( A^{-1} \in \mathcal{L}(X_2, X_1) \) exists, then

\[ (A^*)^{-1} = \psi_2^{-1} \circ (A^{-1})' \circ \psi_1 = (A^{-1})^*, \]

completing the proof.

For extensions of Def. 10.15 and Th. 10.16 to infinite-dimensional Hilbert spaces, see [Phi17c, Def. 4.34], [Phi17c, Cor. 4.35].

**Definition 10.17.** Let \( m, n \in \mathbb{N} \) and let \( M := (m_{kl}) \in \mathcal{M}(m, n, \mathbb{K}) \) be an \( m \times n \) matrix over \( \mathbb{K} \). We call

\[ \overline{M} := (\overline{m_{kl}}) \in \mathcal{M}(m, n, \mathbb{K}) \]

the **complex conjugate** matrix of \( M \) and

\[ M^* := (\overline{M})^t = (\overline{M^t}) \in \mathcal{M}(n, m, \mathbb{K}) \]

the **adjoint** matrix of \( M \) (thus, for \( \mathbb{K} = \mathbb{R} \), the adjoint matrix is the same as the transpose matrix).

**Theorem 10.18.** Let \((X, \langle \cdot, \cdot \rangle), (Y, \langle \cdot, \cdot \rangle)\) be finite-dimensional inner product spaces over \( \mathbb{K} \). Let \( A \in \mathcal{L}(X, Y) \).

(a) Let \( B_X := (x_1, \ldots, x_n) \) and \( B_Y := (y_1, \ldots, y_m) \) be ordered orthonormal bases of \( X \) and \( Y \), respectively. If \( M := (m_{kl}) \in \mathcal{M}(m, n, \mathbb{K}) \) is the matrix of \( A \) with respect to \( B_X \) and \( B_Y \), then the adjoint matrix \( M^* \) represents the adjoint map \( A^* \in \mathcal{L}(Y, X) \) with respect to \( B_Y \) and \( B_X \).

(b) Let \( X = Y, A \in \mathcal{L}(X, X) \). Then \( \det(A^*) = \overline{\det A} \). If \( \chi_A = \sum_{k=0}^n a_k X^k \) is the characteristic polynomial of \( A \), then \( \chi_{A^*} = \sum_{k=0}^n \overline{a_k} X^k \) and

\[ \sigma(A^*) = \{ \overline{\lambda} : \lambda \in \sigma(A) \}. \]
Proof. (a): Suppose
\[
\forall \, l \in \{1, \ldots, n\} \quad Ax_l = \sum_{k=1}^{m} m_{kl} y_k \quad \land \quad \forall \, k \in \{1, \ldots, m\} \quad A^* y_k = \sum_{l=1}^{n} n_{kl} x_l
\]
with \( n_{kl} \in \mathbb{K} \). Then, for each \((k, l) \in \{1, \ldots, m\} \times \{1, \ldots, n\}\),
\[
m_{kl} = \langle Ax_l, y_k \rangle = \langle x_l, A^* y_k \rangle = \overline{n_{kl}},
\]
showing \( M^* \) to be the matrix of \( A^* \) with respect to \( B_Y \) and \( B_X \).

(b): For \( M \) as in (a), it is
\[
\det(A^*) = \det(M^*) = \det \left( (M)^t \right) = \det(M) = \det(A),
\]
where \(^*\) holds, as the forming of complex conjugates commutes with the forming of sums and products of complex numbers. Next, using this fact again together with the linearity of forming the transpose of a matrix, we compute
\[
\chi_{A^*} = \det(X \text{ Id}_n - M^*) = \det \left( X \text{ Id}_n - (M)^t \right) = \det \left( (X \text{ Id}_n - M)^t \right) = \det \left( X \text{ Id}_n - M \right) = \sum_{k=0}^{n} a_k X^k.
\]
Thus,
\[
\lambda \in \sigma(A^*) \iff \epsilon_\lambda(\chi_{A^*}) = \sum_{k=0}^{n} a_k \lambda^k = 0 \iff \epsilon_{\overline{\lambda}}(\chi_{A}) = \sum_{k=0}^{n} a_k (\overline{\lambda})^k = 0 \iff \overline{\lambda} \in \sigma(A),
\]
thereby proving (b). \( \blacksquare \)

**Definition 10.19.** Let \((X, \langle \cdot, \cdot \rangle)\) be an inner product space over \( \mathbb{K} \) and let \( U \) be a subspace of \( X \) such that \( X = U \perp U^\perp \). Then the linear projection \( P_U : X \to U \), \( P_U(u + v) := u \) for \( u + v \in X \) with \( u \in U \) and \( v \in U^\perp \) is called the **orthogonal projection** from \( X \) onto \( U \).

**Theorem 10.20.** Let \((X, \langle \cdot, \cdot \rangle)\) be an inner product space over \( \mathbb{K} \), let \( U \) be a subspace of \( X \) such that \( X = U \perp U^\perp \), and let \( P_U : X \to U \) be the orthogonal projection onto \( U \). Moreover, let \( \| \cdot \| \) denote the induced norm on \( X \).

(a) One has
\[
\forall \, x \in X \quad \forall \, u \in U \quad \left( u \neq P_U(x) \Rightarrow \|P_U(x) - x\| < \|u - x\| \right),
\]
i.e. \( P_U(x) \) is the strict minimum of the function \( u \mapsto \|u - x\| \) from \( U \) into \( \mathbb{R}_0^+ \) and \( P_U(x) \) can be viewed as the best approximation to \( x \) in \( U \).
(b) If \( B_U := \{u_1, \ldots, u_n\} \) is an orthonormal basis of \( U \), \( \dim U = n \in \mathbb{N} \), then

\[
\forall x \in X \quad P_U(x) = \sum_{i=1}^{n} \langle x, u_i \rangle u_i.
\]

Proof. (a): Let \( x \in X \) and \( u \in U \) with \( u \neq P_U(x) \). Then \( P_U(x) - u \in U \) and \( x - P_U(x) \in \ker P_U = U^\perp \).

Thus,

\[
\|u - x\|^2 - \|P_U(x) - x\|^2 - \|u - P_U(x)\|^2 > \|P_U(x) - x\|^2,
\]

thereby proving (a).

(b): Let \( x \in X \). As \( B_U \) is a basis of \( U \), there exist \( \lambda_1, \ldots, \lambda_n \in \mathbb{K} \) such that

\[
P_U(x) = \sum_{i=1}^{n} \lambda_i u_i.
\]

Recalling \( x - P_U(x) \in \ker P_U = U^\perp \), we obtain

\[
\forall i \in \{1, \ldots, n\} \quad \langle x, u_i \rangle - \langle P_U(x), u_i \rangle = \langle x - P_U(x), u_i \rangle = 0
\]

and, thus,

\[
\forall i \in \{1, \ldots, n\} \quad \langle x, u_i \rangle = \langle P_U(x), u_i \rangle = \lambda_i,
\]

thereby proving (b). \( \blacksquare \)

**Proposition 10.21.** Let \( (X, \langle \cdot, \cdot \rangle) \) be an inner product space over \( \mathbb{K} \) with induced norm \( \| \cdot \| \). Let \( P \in \mathcal{L}(X, X) \) be a projection. Then \( P \) is an orthogonal projection onto \( U := \text{Im} P \) (i.e. \( X = \text{Im} P \perp \ker P \)) if, and only if \( P = P^* \).

Proof. First, assume \( X = \text{Im} P \perp \ker P \). Then, for each \( x = x_x + x_y \) and \( y = y_y + y_y \) with \( x_x, y_y \in \text{Im} P \) and \( x_v, y_v \in \ker P \), we have

\[
\langle Px, y \rangle = \langle x_x, y_y \rangle + \langle x_v, y_v \rangle = \langle x_x, y_y \rangle + 0 = \langle x_x, y_y \rangle + \langle y_v, x_y \rangle = \langle x, Py \rangle,
\]

proving \( P = P^* \). Conversely, assume \( P = P^* \). Then, for each \( x \in X \),

\[
\|Px\|^2 = \langle Px, Px \rangle = \langle P^2 x, x \rangle \leq \|Px\| \|x\|,
\]

where (*) holds due to the Cauchy-Schwarz inequality [Phi16b, (1.41)]. In consequence,

\[
\forall x \in X \quad \|Px\| \leq \|x\|.
\]

Now let \( u \in \text{Im} P \) and \( 0 \neq v \in \ker P \). Define

\[
y := u - \frac{\langle u, v \rangle}{\|v\|^2} v.
\]

Then \( \langle y, v \rangle = 0 \), implying

\[
\|y\|^2 \geq \|Py\|^2 = \|u\|^2 - \frac{\langle u, v \rangle}{\|v\|^2} v \quad \text{Prop. 10.8(b)} = \left\| u - \frac{\langle u, v \rangle}{\|v\|^2} v \right\|^2 + \left\| \frac{\langle u, v \rangle}{\|v\|^2} v \right\|^2 = \|y\|^2 + \frac{\langle u, v \rangle^2}{\|v\|^2}.
\]

As this yields \( \langle u, v \rangle = 0 \), we have \( X = \text{Im} P \perp \ker P \), as desired. \( \blacksquare \)
Example 10.22. Let \( n \in \mathbb{N}_0 \) and define
\[
T_n := \left\{ (f : [-\pi, \pi] \to \mathbb{C}) : f(t) = \sum_{k=-n}^{n} \gamma_k e^{ikt} \wedge \gamma_{-n}, \ldots, \gamma_n \in \mathbb{C} \right\}
\]
(due to Euler’s formula, relating the exponential function to sine and cosine, the elements of \( T_n \) are known as trigonometric polynomials). Clearly, \( U := T_n \) is a subspace of the space \( X := C[-\pi, \pi] \) of Ex. 10.5(b). As it turns out, we even have \( X = U \perp U \perp \) (we will not prove this here, but it follows from [Phi17c, Th. 4.20(e)], since the finite-dimensional subspace \( U \) is automatically a closed subspace, cf. [Phi17c, Th. 1.16(b)]). Thus, one has an orthogonal projection \( P_U \) from \( X \) onto \( U \), yielding the best approximation of a continuous function by a trigonometric polynomial. Moreover, if one has an orthonormal basis of \( U \), then one can use Th. 10.20(b) to compute \( P_U(x) \) for each function \( x \in X \).

We verify that an orthonormal basis is given by the functions
\[
u_k : [-\pi, \pi] \to \mathbb{C}, \quad \nu_k(t) := \frac{e^{ikt}}{\sqrt{2\pi}} \quad (k \in \{-n, \ldots, n\}) :\]

One computes, for each \( k, l \in \{-n, \ldots, n\} \),
\[
\int_{-\pi}^{\pi} \nu_k \overline{\nu_l} = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(k-l)t} \, dt = \begin{cases}
1 & \text{for } k = l, \\
\frac{1}{2\pi} [e^{i(k-l)t}]_{-\pi}^{\pi} = 0 & \text{for } k \neq l.
\end{cases}
\]

Thus, for each \( x \in X \), the orthogonal projection is
\[
P_U(x) = \frac{1}{\sqrt{2\pi}} \sum_{k=-n}^{n} \nu_k \int_{-\pi}^{\pi} x(t) e^{-ikt} \, dt.
\]

For example, let \( x(t) = t \) and \( n = 1 \). Then, since
\[
\int_{-\pi}^{\pi} e^{ikt} \, dt = \left[ \frac{te^{ikt}}{i} \right]_{-\pi}^{\pi} = \int_{-\pi}^{\pi} \frac{e^{ikt}}{i} \, dt = 2\pi i - 0 = 2\pi i,
\]
\[
\int_{-\pi}^{\pi} t \, dt = 0,
\]
\[
\int_{-\pi}^{\pi} te^{-ikt} \, dt = -2\pi i,
\]

\( P_U(x) = 2\pi i + 0 - 2\pi i = 0 \).

10.3 Hermitian, Unitary, and Normal Maps

Definition 10.23. Let \((X, \langle \cdot, \cdot \rangle)\) be a finite-dimensional inner product space over \( K \) and \( A \in \mathcal{L}(X, X) \). Moreover, let \( M \in \mathcal{M}(n, K) \), \( n \in \mathbb{N} \).
(a) We call $A$ normal if, and only if, $AA^* = A^*A$; likewise, we call $M$ normal if, and only if, $MM^* = M^*M$. We define

$$\text{Nor}(X) := \{ A \in \mathcal{L}(X, X) : A \text{ is normal} \},$$
$$\text{Nor}_n(K) := \{ M \in \mathcal{M}(n, K) : M \text{ is normal} \}.$$  

(b) We call $A$ Hermitian or self-adjoint if, and only if, $A = A^*$; likewise, we call $M$ Hermitian or self-adjoint if, and only if, $M = M^*$ (thus, for $K = \mathbb{R}$, Hermitian is the same as symmetric). We define

$$\text{Herm}(X) := \{ A \in \mathcal{L}(X, X) : A \text{ is Hermitian} \},$$
$$\text{Herm}_n(K) := \{ M \in \mathcal{M}(n, K) : M \text{ is Hermitian} \},$$

and, for $K = \mathbb{R}$,

$$\text{Sym}(X) := \{ A \in \mathcal{L}(X, X) : A \text{ is symmetric} \},$$
$$\text{Sym}_n(\mathbb{R}) := \{ M \in \mathcal{M}(n, \mathbb{R}) : M \text{ is symmetric} \}.$$  

(c) We call $A \in \text{GL}(X)$ unitary if, and only if, $A^{-1} = A^*$; likewise, we call $M \in \text{GL}_n(K)$ unitary if, and only if, $M^{-1} = M^*$. If $K = \mathbb{R}$, then we also call unitary maps and matrices orthogonal. We define

$$\text{U}(X) := \{ A \in \mathcal{L}(X, X) : A \text{ is unitary} \},$$
$$\text{U}_n(K) := \{ M \in \mathcal{M}(n, K) : M \text{ is unitary} \},$$

and, for $K = \mathbb{R}$,

$$\text{O}(X) := \{ A \in \mathcal{L}(X, X) : A \text{ is orthogonal} \},$$
$$\text{O}_n(\mathbb{R}) := \{ M \in \mathcal{M}(n, \mathbb{R}) : M \text{ is orthogonal} \}.$$  

Proposition 10.24. Let $(X, \langle \cdot , \cdot \rangle)$ be a finite-dimensional inner product space over $K$, $n \in \mathbb{N}$. We have $\text{Herm}(X) \subseteq \text{Nor}(X)$, $\text{Herm}_n(K) \subseteq \text{Nor}_n(K)$, $\text{U}(X) \subseteq \text{Nor}(X)$, $\text{U}_n(K) \subseteq \text{Nor}_n(K)$.

Proof. That Hermitian implies normal is immediate. If $A \in \text{U}(X)$, then $AA^* = A^*A = \text{Id}$, i.e. $A \in \text{Nor}(X)$. \hfill \blacksquare

Proposition 10.25. Let $(X, \langle \cdot , \cdot \rangle)$ be a finite-dimensional inner product space over $K$, $n \in \mathbb{N}$.

(a) $\text{U}(X)$ is a subgroup of $\text{GL}(X)$; $\text{U}_n(K)$ is a subgroup of $\text{GL}_n(K)$.

(b) Let $A, B \in \text{Herm}(X)$. Then $A + B \in \text{Herm}(X)$. If $A \in \text{GL}(X)$, then $A^{-1} \in \text{Herm}(X)$. If $AB = BA$, then $AB \in \text{Herm}(X)$. The analogous results also hold for Hermitian matrices.
Proof. (a): If $A, B \in U(X)$, then $(AB)^{-1} = B^{-1}A^{-1} = B^*A^* = (AB)^*$, showing $AB \in U(X)$. Also $A = (A^{-1})^{-1} = (A^*)^*$, showing $A^{-1} \in U(X)$ and establishing the case.

(b): If $A, B \in \text{Herm}(X)$, then $(A + B)^* = A^* + B^* = A + B$, showing $A + B \in \text{Herm}(X)$. If $A \in \text{Herm}(X) \cap \text{GL}(X)$, then $(A^{-1})^* = (A^*)^{-1} = A^{-1}$, proving $A^{-1} \in \text{Herm}(X) \cap \text{GL}(X)$. If $AB = BA$, then $(AB)^* = B^*A^* = BA = AB$, showing $AB \in \text{Herm}(X)$.

In general, for normal maps and normal matrices, neither the sum nor the product is normal (however, one can show that, if $A, B$ are normal with $AB = BA$, then $A + B$ and $AB$ are normal – this makes use of Fuglede’s theorem and is not quite as easy as one might think).

**Proposition 10.26.** Let $(X, \langle \cdot, \cdot \rangle)$ be an inner product space over $\mathbb{K}$, $\dim X = n \in \mathbb{N}$, with ordered orthonormal basis $B := (x_1, \ldots, x_n)$. Let $\| \cdot \|$ denote the induced norm. Moreover, let $U \in \mathcal{L}(X, X)$, and let $M = (m_{kl}) \in \mathcal{M}(n, \mathbb{K})$, $M^* = (m_{kl}^*) \in \mathcal{M}(n, \mathbb{K})$ be the respective matrices of $U$ and $U^*$ with respect to $B$. Then the following statements are equivalent:

(i) $U$ and $M$ are unitary.

(ii) $U^*$ and $M^*$ are unitary.

(iii) The columns of $M$ form an orthonormal basis of $\mathbb{K}^n$ with respect to the standard inner product on $\mathbb{K}^n$.

(iv) The rows of $M$ form an orthonormal basis of $\mathbb{K}^n$ with respect to the standard inner product on $\mathbb{K}^n$.

(v) $M^t$ is unitary.

(vi) $\overline{M}$ is unitary.

(vii) $\langle Ux, Uy \rangle = \langle x, y \rangle$ holds for each $x, y \in X$.

(viii) $\|Ux\| = \|x\|$ for each $x \in X$.

Proof. “(i)$\iff$(ii)”: $U^{-1} = U^*$ is equivalent to $\text{Id} = UU^*$, which is equivalent to $(U^*)^{-1} = U = (U^*)^*$.

“(i)$\iff$(iii)”: $M^{-1} = M^*$ implies

$$
\begin{pmatrix}
  m_{1k} \\
  \vdots \\
  m_{nk}
\end{pmatrix} \cdot \begin{pmatrix}
  m_{1j} \\
  \vdots \\
  m_{nj}
\end{pmatrix} = \sum_{l=1}^n m_{lk} \overline{m}_{lj} = \sum_{l=1}^n m_{jl}^* m_{lk} = \begin{cases}
  0 & \text{for } k \neq j, \\
  1 & \text{for } k = j,
\end{cases}
$$

(10.9)

showing that the columns of $M$ form an orthonormal basis of $\mathbb{K}^n$ with respect to the standard inner product on $\mathbb{K}^n$. Conversely, if the columns of $M$ form an orthonormal
basis of $\mathbb{K}^n$ with respect to the standard inner product, then they satisfy (10.9), which implies $M^*M = \text{Id}$.

“(i)⇔(iv)”: $M^{-1} = M^*$ implies

\[
\begin{pmatrix}
m_{k1} \\
\vdots \\
m_{kn}
\end{pmatrix} \cdot \begin{pmatrix}
m_{j1} \\
\vdots \\
m_{jn}
\end{pmatrix} = \sum_{l=1}^{n} m_{kl} m_{jl} = \sum_{l=1}^{n} m_{kl} m_{lj}^* = \begin{cases} 
0 & \text{for } k \neq j, \\
1 & \text{for } k = j,
\end{cases}
\]

(10.10)

showing that the rows of $M$ form an orthonormal basis of $\mathbb{K}^n$ with respect to the standard inner product on $\mathbb{K}^n$. Conversely, if the rows of $M$ form an orthonormal basis of $\mathbb{K}^n$ with respect to the standard inner product, then they satisfy (10.10), which implies $M^*M = \text{Id}$.

“(i)⇔(v)”: Since the rows of $M$ are the columns of $M^t$, the equivalence of (i) and (v) is immediate from (iii) and (iv).

“(i)⇔(vi)”: Since $M = (M^*)^t$, the equivalence of (i) and (vi) is immediate from (ii) and (v).

“(i)⇒(vii)”: For each $x,y \in X$:

\[\langle Ux, Uy \rangle = \langle U^*Ux, y \rangle = \langle \text{Id} x, y \rangle = \langle x, y \rangle.\]

“(vii)⇒(i)”: Since $\forall x,y \in X$ $
\langle \text{Id} x, y \rangle = \langle x, y \rangle = \langle Ux, Uy \rangle = \langle x, U^*Uy \rangle$
implies $U^*U = \text{Id}$, we obtain $U$ to be unitary.

“(vii)⇒(viii)”: For each $x \in X$: $\|Ux\|^2 = \langle Ux, Ux \rangle = \langle x, x \rangle = \|x\|^2$.

“(viii)⇒(vii)”: Let $x, y \in X$. Then (viii) implies

\[
\langle Ux, Ux \rangle + \langle Ux, Uy \rangle + \langle Uy, Ux \rangle + \langle Uy, Uy \rangle = \langle U(x + y), U(x + y) \rangle
\]

\[
= \langle x + y, x + y \rangle = \langle x, x \rangle + \langle x, y \rangle + \langle y, x \rangle + \langle y, y \rangle
\]

and, thus,

\[\langle Ux, Uy \rangle + \langle Uy, Ux \rangle = \langle x, y \rangle + \langle y, x \rangle.\]

Similarly,

\[
\langle Ux, Ux \rangle - i\langle Ux, Uy \rangle + i\langle Uy, Ux \rangle + \langle Uy, Uy \rangle = \langle U(x + iy), U(x + iy) \rangle
\]

\[
= \langle x + iy, x + iy \rangle = \langle x, x \rangle - i\langle x, y \rangle + i\langle y, x \rangle + \langle y, y \rangle
\]

and, thus,

\[\langle Ux, Uy \rangle - \langle Uy, Ux \rangle = \langle x, y \rangle - \langle y, x \rangle.\]

Adding both results and dividing by 2 then yields (vii).

In preparation for results on the diagonalizability of normal, Hermitian, and unitary maps, we prove the following proposition:
Proposition 10.27. Let \((X, \langle \cdot, \cdot \rangle)\) be an inner product space over \(\mathbb{K}\), \(\dim X = n \in \mathbb{N}\), and \(A \in \text{Nor}(X)\).

(a) If \(0 \neq x \in X\) is an eigenvector to the eigenvalue \(\lambda \in \mathbb{K}\) of \(A\), then \(x\) is also an eigenvector to the eigenvalue \(\overline{\lambda}\) of \(A^*\).

(b) If \(U\) is an \(A\)-invariant subspace of \(X\), then \(A(U^\perp) \subseteq U^\perp\) and \(A^*(U) \subseteq U\).

Proof. (a): It suffices to show \(\langle A^*x - \overline{\lambda}x, A^*x - \overline{\lambda}x \rangle = 0\). To this end, we use \(Ax = \lambda x\) to compute

\[
\langle A^*x - \overline{\lambda}x, A^*x - \overline{\lambda}x \rangle = \langle A^*x, A^*x \rangle - \overline{\lambda} \langle x, A^*x \rangle - \lambda \langle A^*x, x \rangle + \lambda \overline{\lambda} \langle x, x \rangle
\]

thereby establishing the case.

(b): Let \(\dim U = m \in \mathbb{N}\) and let \(B_U := \{u_1, \ldots, u_m\}\) be an orthonormal basis of \(U\). As \(U\) is \(A\)-invariant, there exist \(a_{kl} \in \mathbb{K}\) such that

\[
\forall \ l \in \{1, \ldots, m\} \quad Au_l = \sum_{k=1}^{m} a_{kl} u_k.
\]

Define

\[
\forall \ l \in \{1, \ldots, m\} \quad x_l := A^* u_l - \sum_{k=1}^{m} \overline{a}_{lk} u_k.
\]

To show the \(A^*\)-invariance of \(U\), it suffices to show that, for each \(l \in \{1, \ldots, m\}\), \(x_l = 0\), i.e. \(\langle x_l, x_l \rangle = 0\). To this end, for each \(l \in \{1, \ldots, m\}\), we compute

\[
\langle x_l, x_l \rangle = \langle A^* u_l, A^* u_l \rangle - \sum_{k=1}^{m} a_{lk} \langle u_l, u_k \rangle - \sum_{k=1}^{m} \overline{a}_{lk} \langle u_l, u_k \rangle + \sum_{j=1}^{m} \sum_{k=1}^{m} a_{lj} \overline{a}_{lk} \langle u_k, u_j \rangle
\]

\[
= \langle A^* Au_l, u_l \rangle - \sum_{k=1}^{m} a_{lk} \langle u_l, Au_k \rangle - \sum_{k=1}^{m} \overline{a}_{lk} \langle Au_l, u_l \rangle - \sum_{k=1}^{m} \sum_{j=1}^{m} a_{jk} \overline{a}_{lk} \langle u_j, u_l \rangle + \sum_{k=1}^{m} |a_{lk}|^2
\]

\[
= \sum_{k=1}^{m} |a_{kl}|^2 - \sum_{k=1}^{m} \sum_{j=1}^{m} \overline{a}_{jk} a_{lk} \langle u_l, u_j \rangle - \sum_{k=1}^{m} \sum_{j=1}^{m} a_{jk} \overline{a}_{lk} \langle u_j, u_l \rangle + \sum_{k=1}^{m} |a_{lk}|^2
\]

\[
= \sum_{k=1}^{m} |a_{kl}|^2 - \sum_{k=1}^{m} |a_{lk}|^2 - \sum_{k=1}^{m} |a_{lk}|^2 + \sum_{k=1}^{m} |a_{lk}|^2 = \sum_{k=1}^{m} |a_{kl}|^2 - \sum_{k=1}^{m} |a_{lk}|^2,
\]

implying

\[
\sum_{l=1}^{m} \langle x_l, x_l \rangle = 0.
\]
As \( \langle x_1, x_l \rangle \geq 0 \) for each \( l \in \{1, \ldots, m\} \), this implies the desired \( \langle x_l, x_l \rangle = 0 \) for each \( l \in \{1, \ldots, m\} \), thereby proving \( A^*(U) \subseteq U \). We will now make use of this result to also show \( A(U^\perp) \subseteq U^\perp \): Let \( u \in U \) and \( x \in U^\perp \). Then

\[
\langle u, Ax \rangle = \langle A^* u, x \rangle = 0,
\]

proving \( Ax \in U^\perp \).

**Theorem 10.28.** Let \( (X, \langle \cdot, \cdot \rangle) \) be an inner product space over \( \mathbb{C} \), \( \dim X = n \in \mathbb{N} \), and \( A \in \text{Nor}(X) \). Then there exists an orthonormal basis \( B \) of \( X \), consisting of eigenvectors of \( A \). In particular, \( A \) is diagonalizable. Moreover, if \( M \in \text{Nor}_n(\mathbb{C}) \), then there exists a unitary matrix \( U \in U_n(\mathbb{C}) \) such that \( D = U^{-1} MU \) is a diagonal matrix.

**Proof.** We prove the existence of the orthonormal basis of eigenvectors via induction on \( n \in \mathbb{N} \). For \( n = 1 \), there is nothing to prove. Thus, let \( n > 1 \). As \( \mathbb{C} \) is algebraically closed, there exists \( \lambda \in \sigma(A) \). Let \( 0 \neq v \in X \) be a corresponding eigenvector and \( U := \text{span}\{v\} \). According Prop. 10.27(b), both \( U \) and \( U^\perp \) are \( A \)-invariant. Moreover, \( A|_{U^\perp} \) is normal, since, if \( A \) and \( A^* \) commute on \( X \), they also commute on the subspace \( U^\perp \). Thus, by induction hypothesis, \( U^\perp \) has an orthonormal basis \( B' \), consisting of eigenvectors of \( A \). Thus, \( X \) also has an orthonormal basis, consisting of eigenvectors of \( A \). Now let \( M \in \text{Nor}_n(\mathbb{C}) \). We consider \( M \) as a normal map on \( \mathbb{C}^n \) with the standard inner product \( \langle \cdot, \cdot \rangle \), where the standard basis of \( \mathbb{C}^n \) constitutes an orthonormal basis. Then we know there exists an ordered orthonormal basis \( B := (x_1, \ldots, x_n) \) of \( \mathbb{C}^n \) such that, with respect to \( B \), \( M \) has the diagonal matrix \( D \). Thus, there exists \( U = (u_{kl}) \in \text{GL}_n(\mathbb{C}) \) such that \( D = U^{-1} MU \) and

\[
\forall l \in \{1, \ldots, n\} \quad x_l = \sum_{k=1}^{n} u_{kl} e_k.
\]

Then

\[
\forall l, j \in \{1, \ldots, n\} \quad \sum_{k=1}^{n} u_{kl} \overline{u}_{kj} = \sum_{k=1}^{n} \sum_{m=1}^{n} u_{kl} \overline{u}_{mj} \langle e_k, e_m \rangle = \langle x_l, x_j \rangle = \delta_{lj},
\]

showing the columns of \( U \) to be orthonormal and \( U \) to be unitary.

**Example 10.29.** Consider \( \mathbb{R}^2 \) with the standard inner product. We already know from Ex. 6.5(a) that \( M := \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \) has no real eigenvalues (the characteristic polynomial is \( \chi_M = X^2 + 1 \)). On the other hand, \( M \) is unitary (in particular, normal), showing, one can not expect Th. 10.28 to hold with \( \mathbb{C} \) replaced by \( \mathbb{R} \).

**Theorem 10.30.** Let \( (X, \langle \cdot, \cdot \rangle) \) be an inner product space over \( \mathbb{K} \), \( \dim X = n \in \mathbb{N} \), and \( A \in \text{Herm}(X) \). Then \( \sigma(A) \subseteq \mathbb{R} \) and there exists an orthonormal basis \( B \) of \( X \), consisting of eigenvectors of \( A \). In particular, \( A \) is diagonalizable. Moreover, if \( M \in \text{Nor}_n(\mathbb{C}) \), then there exists a unitary matrix \( U \in U_n(\mathbb{C}) \) such that \( D = U^{-1} MU \) is a diagonal matrix. Also, in particular, for \( \mathbb{K} = \mathbb{R} \), each \( A \in \text{Sym}(X) \) is diagonalizable, and, if \( M \in \text{Sym}_n(\mathbb{R}) \), then there exists an orthogonal matrix \( U \in O_n(\mathbb{R}) \) such that \( D = U^{-1} MU \) is a diagonal matrix.
Proof. Let \( \lambda \in \sigma(A) \) and let \( 0 \neq x \in X \) be a corresponding eigenvector. Then, according to Prop. 10.27(a), \( \lambda x = Ax = A^*x = \overline{\lambda}x \), showing \( \lambda = \overline{\lambda} \) and \( \lambda \in \mathbb{R} \). As \( A \in \text{Herm}(X) \) implies \( A \) to be normal, the case \( \mathbb{K} = \mathbb{C} \) is now immediate from Th. 10.28. It remains to consider \( \mathbb{K} = \mathbb{R} \). First, consider \( n = 2 \). If \( B_0 := (x_1, x_2) \) is an ordered orthonormal basis of \( X \), then the matrix \( M \in \mathcal{M}(2, \mathbb{R}) \) of \( A \) with respect to \( B_0 \) must have the form

\[
M = \begin{pmatrix} a & b \\ b & c \end{pmatrix}
\]

with \( a, b, c \in \mathbb{R} \). Thus, the characteristic polynomial is

\[
\chi_A = (X - a)(X - c) - b^2 = X^2 - (a + c)C + ac - b^2
\]

with the zeros

\[
\lambda = \frac{a + c}{2} \pm \sqrt{\frac{(a + c)^2}{4} - ac + b^2} = \frac{a + c}{2} \pm \sqrt{\frac{(a - c)^2 + 4b^2}{4}} \in \mathbb{R},
\]

showing \( A \) to be diagonalizable in this case. The rest of the proof is now conducted analogous to the proof of Th. 10.28: We prove the existence of the orthonormal basis of eigenvectors via induction on \( n \in \mathbb{N} \). For \( n = 1 \), there is nothing to prove. Thus, let \( n > 1 \). According to Prop. 8.16(a), there exists an \( A \)-invariant subspace \( W \) of \( X \) such that \( \dim W \in \{1, 2\} \). If \( \dim W = 1 \), then \( A \) has an eigenvalue \( \lambda \) and a corresponding eigenvector \( 0 \neq v \in X \). If \( \dim W = 2 \), then \( A \mid W \) is, clearly, also Hermitian, and the above-considered case \( n = 2 \) yields that \( A \mid W \) (and, thus, \( A \)) has an eigenvalue \( \lambda \) and a corresponding eigenvector \( 0 \neq v \in X \). Let \( U := \text{span}\{v\} \). According Prop. 10.27(b), both \( U \) and \( U^\perp \) are \( A \)-invariant. Moreover, \( A \mid U^\perp \) is also Hermitian and, by induction hypothesis, \( U^\perp \) has an orthonormal basis \( B' \), consisting of eigenvectors of \( A \). Thus, \( X \) also has an orthonormal basis, consisting of eigenvectors of \( A \). Now let \( M \in \text{Sym}_n(\mathbb{R}) \). We consider \( M \) as a symmetric map on \( \mathbb{R}^n \) with the standard inner product \( \langle \cdot, \cdot \rangle \), where the standard basis of \( \mathbb{R}^n \) constitutes an orthonormal basis. Then we know there exists an ordered orthonormal basis \( B := (x_1, \ldots, x_n) \) of \( \mathbb{R}^n \) such that, with respect to \( B \), \( M \) has the diagonal matrix \( D \). Thus, there exists \( U = (u_{kl}) \in \text{GL}_n(\mathbb{R}) \) such that \( D = U^{-1}MU \) and

\[
\forall l \in \{1, \ldots, n\} \quad x_l = \sum_{k=1}^{n} u_{kl} e_k.
\]

Then

\[
\forall l, j \in \{1, \ldots, n\} \quad \sum_{k=1}^{n} u_{kl} \overline{u}_{kj} = \sum_{k=1}^{n} \sum_{m=1}^{n} u_{kl} u_{mj} \langle e_k, e_m \rangle = \langle x_l, x_j \rangle = \delta_{lj},
\]

showing the columns of \( U \) to be orthonormal and \( U \) to be orthogonal. \( \blacksquare \)
A Multilinear Maps

**Theorem A.1.** Let $V$ and $W$ be vector spaces over the field $F$, $\alpha \in \mathbb{N}$. Then, as vector spaces over $F$, $\mathcal{L}(V, \mathcal{L}^{\alpha-1}(V, W))$ and $\mathcal{L}^\alpha(V, W)$ are isomorphic via the isomorphism

$$
\Phi : \mathcal{L}(V, \mathcal{L}^{\alpha-1}(V, W)) \longrightarrow \mathcal{L}^\alpha(V, W),
\Phi(L)(x^1, \ldots, x^\alpha) := L(x^1)(x^2, \ldots, x^\alpha).
$$

*(A.1)*

**Proof.** Since $L$ is linear and $L(x^1)$ is $(\alpha - 1)$ times linear, $\Phi(L)$ is, indeed, an element of $\mathcal{L}^\alpha(V, W)$, showing that $\Phi$ is well-defined by (A.1). Next, we verify $\Phi$ to be linear: If $\lambda \in F$ and $K, L \in \mathcal{L}(V, \mathcal{L}^{\alpha-1}(V, W))$, then

$$
\Phi(\lambda L)(x^1, \ldots, x^\alpha) = (\lambda L)(x^1)(x^2, \ldots, x^\alpha) = \lambda(L(x^1)(x^2, \ldots, x^\alpha)) = \lambda\Phi(L)(x^1, \ldots, x^\alpha)
$$

and

$$
\Phi(K + L)(x^1, \ldots, x^\alpha) = (K + L)(x^1)(x^2, \ldots, x^\alpha) = (K(x^1) + L(x^1))(x^2, \ldots, x^\alpha)
= K(x^1)(x^2, \ldots, x^\alpha) + L(x^1)(x^2, \ldots, x^\alpha)
= \Phi(K)(x^1, \ldots, x^\alpha) + \Phi(L)(x^1, \ldots, x^\alpha)
= (\Phi(K) + \Phi(L))(x^1, \ldots, x^\alpha),
$$

proving $\Phi$ to be linear. Now we show $\Phi$ to be injective. To this end, we show that, if $L \neq 0$, then $\Phi(L) \neq 0$. If $L \neq 0$, then there exist $x^1, \ldots, x^\alpha \in V$ such that $L(x^1)(x^2, \ldots, x^\alpha) \neq 0$, showing that $\Phi(L) \neq 0$ as needed. To verify $\Phi$ is also surjective, let $K \in \mathcal{L}^\alpha(V, W)$. Define $L : V \longrightarrow \mathcal{L}^{\alpha-1}(V, W)$ by letting

$$
L(x^1)(x^2, \ldots, x^\alpha) := K(x^1, \ldots, x^\alpha).
$$

*(A.2)*

Then, clearly, for each $x^1 \in V$, $L(x^1) \in \mathcal{L}^{\alpha-1}(V, W)$. Moreover, $L$ is linear, i.e. $L \in \mathcal{L}(V, \mathcal{L}^{\alpha-1}(V, W))$. Comparing (A.2) with (A.1) shows $\Phi(L) = K$, i.e. $\Phi$ is surjective, completing the proof. ■

**References**


REFERENCES


