Chapter 10
Linear Maps

In this chapter we study maps between vector spaces that are compatible with the two vector space operations, addition and scalar multiplication. These maps are called linear maps or homomorphisms. We first investigate their most important properties and then show that in the case of finite dimensional vector spaces every linear map can be represented by a matrix, when bases in the respective spaces have been chosen. If the bases are chosen in a clever way, then we can read off important properties of a linear map from its matrix representation. This central idea will arise frequently in later chapters.

10.1 Basic Definitions and Properties of Linear Maps

We start our investigations with the definition of linear maps between vector spaces.

Definition 10.1 Let $\mathcal{V}$ and $\mathcal{W}$ be K-vector spaces. A map $f : \mathcal{V} \to \mathcal{W}$ is called *linear*, when

(1) $f(\lambda v) = \lambda f(v)$, and
(2) $f(v + w) = f(v) + f(w)$,

hold for all $v, w \in \mathcal{V}$ and $\lambda \in K$. The set of all these maps is denoted by $\mathcal{L}(\mathcal{V}, \mathcal{W})$.

A linear map $f : \mathcal{V} \to \mathcal{W}$ is also called a *linear transformation* or (vector space) *homomorphism*. A bijective linear map is called an *isomorphism*. If there exists an isomorphism between $\mathcal{V}$ and $\mathcal{W}$, then the spaces $\mathcal{V}$ and $\mathcal{W}$ are called *isomorphic*, which we denote by

$$\mathcal{V} \cong \mathcal{W}.$$

A map $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ is called an *endomorphism*, and a bijective endomorphism is called an *automorphism*.

J. Liesen and V. Mehrmann, *Linear Algebra*, Springer Undergraduate Mathematics Series, DOI 10.1007/978-3-319-24346-7_10

It is an easy exercise to show that the conditions (1) and (2) in Definition 10.1 hold if and only if

$$f(\lambda v + \mu w) = \lambda f(v) + \mu f(w)$$

holds for all $\lambda, \mu \in K$ and $v, w \in \mathcal{V}$.

Example 10.2

(1) Every matrix $A \in K^{n,m}$ defines a map

$$A : K^{m,1} \to K^{n,1}, \quad x \mapsto Ax.$$

This map is linear, since

$$A(\lambda x) = \lambda Ax \quad \text{for all } x \in K^{m,1} \text{ and } \lambda \in K,$$
$$A(x + y) = Ax + Ay \quad \text{for all } x, y \in K^{m,1}$$

(cp. Lemmas 4.3 and 4.4).

(2) The map trace : $K^{n,n} \to K$, $A = [a_{ij}] \mapsto \text{trace}(A) := \sum_{i=1}^{n} a_{ii}$, is linear (cp. Exercise 8.8).

(3) The map

$$f : \mathbb{Q}[t]_{\leq 3} \to \mathbb{Q}[t]_{\leq 2}, \quad \alpha_3 t^3 + \alpha_2 t^2 + \alpha_1 t + \alpha_0 \;\mapsto\; 2\alpha_2 t^2 + 3\alpha_1 t + 4\alpha_0,$$

is linear. (Show this as an exercise). The map

$$g : \mathbb{Q}[t]_{\leq 3} \to \mathbb{Q}[t]_{\leq 2}, \quad \alpha_3 t^3 + \alpha_2 t^2 + \alpha_1 t + \alpha_0 \;\mapsto\; \alpha_2 t^2 + \alpha_1 t + \alpha_0^2,$$

is not linear. For example, if $p_1 = t + 2$ and $p_2 = t + 1$, then $g(p_1 + p_2) = 2t + 9 \neq 2t + 5 = g(p_1) + g(p_2)$.

The set of linear maps between vector spaces forms a vector space itself.

Lemma 10.3 *Let $\mathcal{V}$ and $\mathcal{W}$ be K-vector spaces. For $f, g \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ and $\lambda \in K$ define $f + g$ and $\lambda \cdot f$ by*

$$(f + g)(v) := f(v) + g(v),$$
$$(\lambda \cdot f)(v) := \lambda f(v),$$

for all $v \in \mathcal{V}$. Then $(\mathcal{L}(\mathcal{V}, \mathcal{W}), +, \cdot)$ is a K-vector space.

Proof Cp. Exercise 9.4. □

The next result deals with the existence and uniqueness of linear maps.

Theorem 10.4 *Let $\mathcal{V}$ and $\mathcal{W}$ be K-vector spaces, let $\{v_1, \ldots, v_m\}$ be a basis of $\mathcal{V}$, and let $w_1, \ldots, w_m \in \mathcal{W}$. Then there exists a unique linear map $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ with $f(v_i) = w_i$ for $i = 1, \ldots, m$.*

Proof For every $v \in \mathcal{V}$ there exist (unique) coordinates $\lambda_1^{(v)}, \ldots, \lambda_m^{(v)}$ with $v = \sum_{i=1}^m \lambda_i^{(v)} v_i$ (cp. Lemma 9.22). We define the map $f : \mathcal{V} \to \mathcal{W}$ by

$$f(v) := \sum_{i=1}^m \lambda_i^{(v)} w_i \quad \text{for all } v \in \mathcal{V}.$$

By definition, $f(v_i) = w_i$ for $i = 1, \ldots, m$.

We next show that f is linear. For every $\lambda \in K$ we have $\lambda v = \sum_{i=1}^m (\lambda \lambda_i^{(v)}) v_i$, and hence

$$f(\lambda v) = \sum_{i=1}^m (\lambda \lambda_i^{(v)}) w_i = \lambda \sum_{i=1}^m \lambda_i^{(v)} w_i = \lambda f(v).$$

If $u = \sum_{i=1}^m \lambda_i^{(u)} v_i \in \mathcal{V}$, then $v + u = \sum_{i=1}^m (\lambda_i^{(v)} + \lambda_i^{(u)}) v_i$, and hence

$$f(v+u) = \sum_{i=1}^m (\lambda_i^{(v)} + \lambda_i^{(u)}) w_i = \sum_{i=1}^m \lambda_i^{(v)} w_i + \sum_{i=1}^m \lambda_i^{(u)} w_i = f(v) + f(u).$$

Thus, $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$.

Suppose that $g \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ also satisfies $g(v_i) = w_i$ for $i = 1, \ldots, m$. Then for every $v = \sum_{i=1}^m \lambda_i^{(v)} v_i$ we have

$$f(v) = f\Big(\sum_{i=1}^m \lambda_i^{(v)} v_i\Big) = \sum_{i=1}^m \lambda_i^{(v)} f(v_i) = \sum_{i=1}^m \lambda_i^{(v)} w_i = \sum_{i=1}^m \lambda_i^{(v)} g(v_i) = g\Big(\sum_{i=1}^m \lambda_i^{(v)} v_i\Big) = g(v),$$

and hence $f = g$, so that f is indeed uniquely determined. □

Theorem 10.4 shows that the map $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ is uniquely determined by the images of f at the given basis vectors of $\mathcal{V}$. Note that the image vectors $w_1, \ldots, w_m \in \mathcal{W}$ may be linearly dependent, and that $\mathcal{W}$ may be infinite dimensional.

In Definition 2.12 we have introduced the image and pre-image of a map. We next recall these definitions for completeness and introduce the kernel of a linear map.

Definition 10.5 If $\mathcal{V}$ and $\mathcal{W}$ are K-vector spaces and $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, then the *kernel* and the *image* of f are defined by

$$\ker(f) := \{v \in \mathcal{V} \mid f(v) = 0\}, \quad \mathrm{im}(f) := \{f(v) \mid v \in \mathcal{V}\}.$$

For $w \in \mathcal{W}$ the *pre-image* of w in the space $\mathcal{V}$ is defined by

$$f^{-1}(w) := f^{-1}(\{w\}) = \{v \in \mathcal{V} \mid f(v) = w\}.$$

The kernel of a linear map is sometimes called the *null space* (or *nullspace*) of the map, and some authors use the notation $\mathrm{null}(f)$ instead of $\ker(f)$.

Note that the pre-image $f^{-1}(w)$ is a set, and that f^{-1} here does *not* mean the inverse map of f (cp. Definition 2.12). In particular, we have $f^{-1}(0) = \ker(f)$, and if $w \notin \operatorname{im}(f)$, then $f^{-1}(w) = \emptyset$,

Example 10.6 For $A \in K^{n,m}$ and the corresponding map $A \in \mathcal{L}(K^{m,1}, K^{n,1})$ from (1) in Example 10.2 we have

$$\ker(A) = \{x \in K^{m,1} \mid Ax = 0\} \quad \text{and} \quad \operatorname{im}(A) = \{Ax \mid x \in K^{m,1}\}.$$

Note that $\ker(A) = \mathscr{L}(A, 0)$ (cp. Definition 6.1). Let $a_j \in K^{n,1}$ denote the jth column of A, $j = 1, \ldots, m$. For $x = [x_1, \ldots, x_m]^T \in K^{m,1}$ we then can write

$$Ax = \sum_{j=1}^{m} x_j a_j.$$

Clearly, $0 \in \ker(A)$. Moreover, we see from the representation of Ax that $\ker(A) = \{0\}$ if and only if the columns of A are linearly independent. The set $\operatorname{im}(A)$ is given by the linear combinations of the columns of A, i.e., $\operatorname{im}(A) = \operatorname{span}\{a_1, \ldots, a_m\}$.

Lemma 10.7 *If $\mathcal{V}$ and $\mathcal{W}$ are K-vector spaces, then for every $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ the following assertions hold:*

(1) $f(0) = 0$ *and* $f(-v) = -f(v)$ *for all* $v \in \mathcal{V}$.
(2) *If f is an isomorphism, then* $f^{-1} \in \mathcal{L}(\mathcal{W}, \mathcal{V})$.
(3) $\ker(f)$ *is a subspace of* $\mathcal{V}$ *and* $\operatorname{im}(f)$ *is a subspace of* $\mathcal{W}$.
(4) f *is surjective if and only if* $\operatorname{im}(f) = \mathcal{W}$.
(5) f *is injective if and only if* $\ker(f) = \{0\}$.
(6) *If f is injective and if $v_1, \ldots, v_m \in \mathcal{V}$ are linearly independent, then $f(v_1), \ldots, f(v_m) \in \mathcal{W}$ are linearly independent.*
(7) *If $v_1, \ldots, v_m \in \mathcal{V}$ are linearly dependent, then $f(v_1), \ldots, f(v_m) \in \mathcal{W}$ are linearly dependent, or, equivalently, if $f(v_1), \ldots, f(v_m) \in \mathcal{W}$ are linearly independent, then $v_1, \ldots, v_m \in \mathcal{V}$ are linearly independent.*
(8) *If $w \in \operatorname{im}(f)$ and if $u \in f^{-1}(w)$ is arbitrary, then*

$$f^{-1}(w) = u + \ker(f) := \{u + v \mid v \in \ker(f)\}.$$

Proof

(1) We have $f(0_{\mathcal{V}}) = f(0_K \cdot 0_{\mathcal{V}}) = 0_K \cdot f(0_{\mathcal{V}}) = 0_{\mathcal{V}}$ as well as $f(v) + f(-v) = f(v + (-v)) = f(0) = 0$ for all $v \in \mathcal{V}$.
(2) The existence of the inverse map $f^{-1} : \mathcal{W} \to \mathcal{V}$ is guaranteed by Theorem 2.20, so we just have to show that f^{-1} is linear. If $w_1, w_2 \in \mathcal{W}$, then there exist uniquely determined $v_1, v_2 \in \mathcal{V}$ with $w_1 = f(v_1)$ and $w_2 = f(v_2)$. Hence,

$$\begin{aligned} f^{-1}(w_1 + w_2) &= f^{-1}(f(v_1) + f(v_2)) = f^{-1}(f(v_1 + v_2)) = v_1 + v_2 \\ &= f^{-1}(w_1) + f^{-1}(w_2). \end{aligned}$$

Moreover, for every $\lambda \in K$ we have

$$f^{-1}(\lambda w_1) = f^{-1}(\lambda f(v_1)) = f^{-1}(f(\lambda v_1)) = \lambda v_1 = \lambda f^{-1}(w_1).$$

(3) and (4) are obvious from the corresponding definitions.

(5) Let f be injective and $v \in \ker(f)$, i.e., $f(v) = 0$. From (1) we know that $f(0) = 0$. Since $f(v) = f(0)$, the injectivity of f yields $v = 0$. Suppose now that $\ker(f) = \{0\}$ and let $u, v \in \mathcal{V}$ with $f(u) = f(v)$. Then $f(u - v) = 0$, i.e., $u - v \in \ker(f)$, which implies $u - v = 0$, i.e., $u = v$.

(6) Let $\sum_{i=1}^{m} \lambda_i f(v_i) = 0$. The linearity of f yields

$$f\Big(\sum_{i=1}^{m} \lambda_i v_i\Big) = 0, \quad \text{i.e.,} \quad \sum_{i=1}^{m} \lambda_i v_i \in \ker(f).$$

Since f is injective, we have $\sum_{i=1}^{m} \lambda_i v_i = 0$ by (5), and hence $\lambda_1 = \cdots = \lambda_m = 0$ due to the linear independence of $v_1, \ldots, v_m$. Thus, $f(v_1), \ldots, f(v_m)$ are linearly independent.

(7) If $v_1, \ldots, v_m$ are linearly dependent, then $\sum_{i=1}^{m} \lambda_i v_i = 0$ for some $\lambda_1, \ldots, \lambda_m \in K$ that are not all equal to zero. Applying f on both sides and using the linearity yields $\sum_{i=1}^{m} \lambda_i f(v_i) = 0$, hence $f(v_1), \ldots, f(v_m)$ are linearly dependent.

(8) Let $w \in \operatorname{im}(f)$ and $u \in f^{-1}(w)$.
If $v \in f^{-1}(w)$, then $f(v) = f(u)$, and thus $f(v - u) = 0$, i.e., $v - u \in \ker(f)$ or $v \in u + \ker(f)$. This shows that $f^{-1}(w) \subseteq u + \ker(f)$.
If, on the other hand, $v \in u + \ker(f)$, then $f(v) = f(u) = w$, i.e., $v \in f^{-1}(w)$. This shows that $u + \ker(f) \subseteq f^{-1}(w)$. □

Example 10.8 Consider a matrix $A \in K^{n,m}$ and the corresponding map $A \in \mathcal{L}(K^{m,1}, K^{n,1})$ from (1) in Example 10.2. For a given $b \in K^{n,1}$ we have $A^{-1}(b) = \mathscr{L}(A, b)$. If $b \notin \operatorname{im}(A)$, then $\mathscr{L}(A, b) = \emptyset$ (case (1) in Corollary 6.6). Now suppose that $b \in \operatorname{im}(A)$ and let $\widehat{x} \in \mathscr{L}(A, b)$ be arbitrary. Then (8) in Lemma 10.7 yields

$$\mathscr{L}(A, b) = \widehat{x} + \ker(A),$$

which is the assertion of Lemma 6.2. If $\ker(A) = \{0\}$, i.e., the columns of A are linearly independent, then $|\mathscr{L}(A, b)| = 1$ (case (2) in Corollary 6.6). If $\ker(A) \neq \{0\}$, i.e., the columns of A are linearly dependent, then $|\mathscr{L}(A, b)| > 1$ (case (3) in Corollary 6.6). If $\{w_1, \ldots, w_\ell\}$ is a basis of $\ker(A)$, then

$$\mathscr{L}(A, b) = \Big\{\widehat{x} + \sum_{i=1}^{\ell} \lambda_i w_i \,\Big|\, \lambda_1, \ldots, \lambda_\ell \in K\Big\}.$$

Thus, the solutions of $Ax = b$ depend of $\ell \leq m$ parameters.

The following result, which gives an important dimension formula for linear maps, is also known as the *rank-nullity theorem*: The dimension of the image of f is equal to the rank of a matrix associated with f (cp. Theorem 10.22 below), and the dimension of the kernel (or null space) of f is sometimes called the *nullity*[1] of f.

Theorem 10.9 *Let $\mathcal{V}$ and $\mathcal{W}$ be K-vector spaces and let $\mathcal{V}$ be finite dimensional. Then for every $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ we have the dimension formula*

$$\dim(\mathcal{V}) = \dim(\mathrm{im}(f)) + \dim(\ker(f)).$$

Proof Let $v_1, \ldots, v_n \in \mathcal{V}$. If $f(v_1), \ldots, f(v_n) \in \mathcal{W}$ are linearly independent, then by (7) in Lemma 10.7 also $v_1, \ldots, v_n$ are linearly independent, and thus $\dim(\mathrm{im}(f)) \leq \dim(\mathcal{V})$. Since $\ker(f) \subseteq \mathcal{V}$, we have $\dim(\ker(f)) \leq \dim(\mathcal{V})$, so that $\mathrm{im}(f)$ and $\ker(f)$ are both finite dimensional.

Let $\{w_1, \ldots, w_r\}$ and $\{v_1, \ldots, v_k\}$ be bases of $\mathrm{im}(f)$ and $\ker(f)$, respectively, and let $u_1 \in f^{-1}(w_1), \ldots, u_r \in f^{-1}(w_r)$. We will show that $\{u_1, \ldots, u_r, v_1, \ldots, v_k\}$ is a basis of $\mathcal{V}$, which then implies the assertion.

If $v \in \mathcal{V}$, then by Lemma 9.22 there exist (unique) coordinates $\mu_1, \ldots, \mu_r \in K$ with $f(v) = \sum_{i=1}^{r} \mu_i w_i$. Let $\widetilde{v} := \sum_{i=1}^{r} \mu_i u_i$, then $f(\widetilde{v}) = f(v)$, and hence $v - \widetilde{v} \in \ker(f)$, which gives $v - \widetilde{v} = \sum_{i=1}^{k} \lambda_i v_i$ for some (unique) coordinates $\lambda_1, \ldots, \lambda_k \in K$. Therefore,

$$v = \widetilde{v} + \sum_{i=1}^{k} \lambda_i v_i = \sum_{i=1}^{r} \mu_i u_i + \sum_{i=1}^{k} \lambda_i v_i,$$

and thus $v \in \mathrm{span}\{u_1, \ldots, u_r, v_1, \ldots, v_k\}$. Since $\{u_1, \ldots, u_r, v_1, \ldots, v_k\} \subset \mathcal{V}$, we have

$$\mathcal{V} = \mathrm{span}\{u_1, \ldots, u_r, v_1, \ldots, v_k\},$$

and it remains to show that $u_1, \ldots, u_r, v_1, \ldots, v_k$ are linearly independent. If

$$\sum_{i=1}^{r} \alpha_i u_i + \sum_{i=1}^{k} \beta_i v_i = 0,$$

then

$$0 = f(0) = f\left(\sum_{i=1}^{r} \alpha_i u_i + \sum_{i=1}^{k} \beta_i v_i\right) = \sum_{i=1}^{r} \alpha_i f(u_i) = \sum_{i=1}^{r} \alpha_i w_i$$

and thus $\alpha_1 = \cdots = \alpha_r = 0$, because $w_1, \ldots, w_r$ are linearly independent. Finally, the linear independence of $v_1, \ldots, v_k$ implies that $\beta_1 = \cdots = \beta_k = 0$. □

[1]This term was introduced in 1884 by James Joseph Sylvester (1814–1897).

Example 10.10

(1) For the linear map

$$f : \mathbb{Q}^{3,1} \to \mathbb{Q}^{2,1}, \quad \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \mapsto \begin{bmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} = \begin{bmatrix} \alpha_1 + \alpha_3 \\ \alpha_1 + \alpha_3 \end{bmatrix},$$

we have

$$\text{im}(f) = \left\{ \begin{bmatrix} \alpha \\ \alpha \end{bmatrix} \,\middle|\, \alpha \in \mathbb{Q} \right\}, \quad \ker(f) = \left\{ \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ -\alpha_1 \end{bmatrix} \,\middle|\, \alpha_1, \alpha_2 \in \mathbb{Q} \right\}.$$

Hence $\dim(\text{im}(f)) = 1$ and $\dim(\ker(f)) = 2$, so that indeed $\dim(\text{im}(f)) + \dim(\ker(f)) = \dim(\mathbb{Q}^{3,1})$.

(2) If $A \in K^{n,m}$ and $A \in \mathcal{L}(K^{m,1}, K^{n,1})$ are as in (1) in Example 10.2, then

$$m = \dim(K^{m,1}) = \dim(\ker(A)) + \dim(\text{im}(A)).$$

Thus, $\dim(\text{im}(A)) = m$ if and only if $\dim(\ker(A)) = 0$. This holds if and only if $\ker(A) = \{0\}$, i.e., if and only if the columns of A are linearly independent (cp. Example 10.6). If, on the other hand, $\dim(\text{im}(A)) < m$, then $\dim(\ker(A)) = m - \dim(\text{im}(A)) > 0$, and thus $\ker(A) \neq \{0\}$. In this case the columns of A are linearly dependent, since there exists an $x \in K^{m,1} \setminus \{0\}$ with $Ax = 0$.

Corollary 10.11 *If $\mathcal{V}$ and $\mathcal{W}$ are K-vector spaces with $\dim(\mathcal{V}) = \dim(\mathcal{W}) \in \mathbb{N}$ and if $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, then the following statements are equivalent:*

(1) f is injective.
(2) f is surjective.
(3) f is bijective.

Proof If (3) holds, then (1) and (2) hold by definition. We now show that (3) is implied by (1) as well as by (2).

If f is injective, then $\ker(f) = \{0\}$ (cp. (5) in Lemma 10.7) and the dimension formula of Theorem 10.9 yields $\dim(\mathcal{W}) = \dim(\mathcal{V}) = \dim(\text{im}(f))$. Thus, $\text{im}(f) = \mathcal{W}$ (cp. Lemma 9.27), so that f is also surjective.

If f is surjective, i.e., $\text{im}(f) = \mathcal{W}$, then the dimension formula and $\dim(\mathcal{W}) = \dim(\mathcal{V})$ yield

$$\dim(\ker(f)) = \dim(\mathcal{V}) - \dim(\text{im}(f)) = \dim(\mathcal{W}) - \dim(\text{im}(f)) = 0.$$

Thus, $\ker(f) = \{0\}$, so that f is also injective. □

Using Theorem 10.9 we can also characterize when two finite dimensional vector spaces are isomorphic.

Corollary 10.12 *Two finite dimensional K-vector spaces $\mathcal{V}$ and $\mathcal{W}$ are isomorphic if and only if* $\dim(\mathcal{V}) = \dim(\mathcal{W})$.

Proof If $\mathcal{V} \cong \mathcal{W}$, then there exists a bijective map $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. By (4) and (5) in Lemma 10.7 we have $\mathrm{im}(f) = \mathcal{W}$ and $\ker(f) = \{0\}$, and the dimension formula of Theorem 10.9 yields

$$\dim(\mathcal{V}) = \dim(\mathrm{im}(f)) + \dim(\ker(f)) = \dim(\mathcal{W}) + \dim(\{0\}) = \dim(\mathcal{W}).$$

Let now $\dim(\mathcal{V}) = \dim(\mathcal{W})$. We need to show that there exists a bijective $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. Let $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_n\}$ be bases of $\mathcal{V}$ and $\mathcal{W}$. By Theorem 10.4 there exists a unique $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ with $f(v_i) = w_i$, $i = 1, \ldots, n$. If $v = \lambda_1 v_1 + \ldots + \lambda_n v_n \in \ker(f)$, then

$$\begin{aligned} 0 = f(v) \;&=\; f(\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_1 f(v_1) + \ldots + \lambda_n f(v_n) \\ &= \lambda_1 w_1 + \ldots + \lambda_n w_n. \end{aligned}$$

Since $w_1, \ldots, w_n$ are linearly independent, we have $\lambda_1 = \cdots = \lambda_n = 0$, hence $v = 0$ and $\ker(f) = \{0\}$. Thus, f is injective. Moreover, the dimension formula yields $\dim(\mathcal{V}) = \dim(\mathrm{im}(f)) = \dim(\mathcal{W})$ and, therefore, $\mathrm{im}(f) = \mathcal{W}$ (cp. Lemma 9.27), so that f is also surjective. □

Example 10.13

(1) The vector spaces $K^{n,m}$ and $K^{m,n}$ both have the dimension $n \cdot m$ and are therefore isomorphic. An isomorphism is given by the linear map $A \mapsto A^T$.
(2) The $\mathbb{R}$-vector spaces $\mathbb{R}^{1,2}$ and $\mathbb{C} = \{x + \mathbf{i}y \mid x, y \in \mathbb{R}\}$ both have the dimension 2 and are therefore isomorphic. An isomorphism is given by the linear map $[x, y] \mapsto x + \mathbf{i}y$.
(3) The vector spaces $\mathbb{Q}[t]_{\leq 2}$ and $\mathbb{Q}^{1,3}$ both have dimension 3 and are therefore isomorphic. An isomorphism is given by the linear map $\alpha_2 t^2 + \alpha_1 t + \alpha_0 \mapsto [\alpha_2, \alpha_1, \alpha_0]$.

Although Mathematics is a formal and exact science, where smallest details matter, one sometimes uses an "abuse of notation" in order to simplify the presentation. We have used this for example in the inductive existence proof of the echelon form in Theorem 5.2. There we kept, for simplicity, the indices of the larger matrix $A^{(1)}$ in the smaller matrix $A^{(2)} = [a_{ij}^{(2)}]$. The matrix $A^{(2)}$ had, of course, an entry in position $(1, 1)$, but this entry was denoted by $a_{22}^{(2)}$ rather than $a_{11}^{(2)}$. Keeping the indices in the induction made the argument much less technical, while the proof itself remained formally correct.

An abuse of notation should always be justified and should not be confused with a "misuse" of notation. In the field of Linear Algebra a justification is often given by an isomorphism that identifies vector spaces with each other. For example, the constant polynomials over a field K, i.e., polynomials of the form αt^0 with $\alpha \in K$, are often written simply as α, i.e., as elements of the field itself. This is justified since

$K[t]_{\leq 0}$ and K are isomorphic K-vector spaces (of dimension 1). We already used this identification above. Similarly, we have identified the vector space $\mathcal{V}$ with $\mathcal{V}^1$ and written just v instead of (v) in Sect. 9.3. Another common example in the literature is the notation K^n that in our text denotes the set of n-tuples with elements from K, but which is often used for the (matrix) sets of the "column vectors" $K^{n,1}$ or the "row vectors" $K^{1,n}$. The actual meaning then should be clear from the context. An attentive reader can significantly benefit from the simplifications due to such abuses of notation.

10.2 Linear Maps and Matrices

Let $\mathcal{V}$ and $\mathcal{W}$ be finite dimensional K-vector spaces with bases $\{v_1, \ldots, v_m\}$ and $\{w_1, \ldots, w_n\}$, respectively, and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. By Lemma 9.22, for every $f(v_j) \in \mathcal{W}$, $j = 1, \ldots, m$, there exist (unique) coordinates $a_{ij} \in K$, $i = 1, \ldots, n$, with

$$f(v_j) = a_{1j} w_1 + \ldots + a_{nj} w_n.$$

We define $A := [a_{ij}] \in K^{n,m}$ and write, similarly to (9.3), the m equations for the vectors $f(v_j)$ as

$$(f(v_1), \ldots, f(v_m)) = (w_1, \ldots, w_n)A. \tag{10.1}$$

The matrix A is determined uniquely by f and the given bases of $\mathcal{V}$ and $\mathcal{W}$.

If $v = \lambda_1 v_1 + \ldots + \lambda_m v_m \in \mathcal{V}$, then

$$\begin{aligned} f(v) = f(\lambda_1 v_1 + \ldots + \lambda_m v_m) \;&=\; \lambda_1 f(v_1) + \ldots + \lambda_m f(v_m) \\ &= (f(v_1), \ldots, f(v_m)) \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix} \\ &= ((w_1, \ldots, w_n)\, A) \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix} \\ &= (w_1, \ldots, w_n) \left(A \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix} \right). \end{aligned}$$

The coordinates of $f(v)$ with respect to the given basis of $\mathcal{W}$ are therefore given by

$$A \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{bmatrix}.$$

Thus, we can compute the coordinates of $f(v)$ simply by multiplying the coordinates of v with A. This motivates the following definition.

Definition 10.14 The uniquely determined matrix in (10.1) is called the *matrix representation* of $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ with respect to the bases $B_1 = \{v_1, \ldots, v_m\}$ of $\mathcal{V}$ and $B_2 = \{w_1, \ldots, w_n\}$ of $\mathcal{W}$. We denote this matrix by $[f]_{B_1,B_2}$.

The construction of the matrix representation and Definition 10.14 can be consistently extended to the case that (at least) one of the K-vector spaces has dimension zero. If, for instance, $m = \dim(\mathcal{V}) \in \mathbb{N}$ and $\mathcal{W} = \{0\}$, then $f(v_j) = 0$ for every basis vector v_j of $\mathcal{V}$. Thus, every vector $f(v_j)$ is an empty linear combination of vector of the basis $\emptyset$ of $\mathcal{W}$. The matrix representation of f then is an empty matrix of size $0 \times m$. If also $\mathcal{V} = \{0\}$, then the matrix representation of f is an empty matrix of size 0×0.

There are many different notations for the matrix representation of linear maps in the literature. The notation should reflect that the matrix depends on the linear map f and the given bases B_1 and B_2. Examples of alternative notations are $[f]_{B_2}^{B_1}$ and $M(f)_{B_1,B_2}$ (where "M" means "matrix").

An important special case is obtained for $\mathcal{V} = \mathcal{W}$, hence in particular $m = n$, and $f = \mathrm{Id}_{\mathcal{V}}$, the identity on $\mathcal{V}$. We then obtain

$$(v_1, \ldots, v_n) = (w_1, \ldots, w_n)[\mathrm{Id}_{\mathcal{V}}]_{B_1,B_2}, \tag{10.2}$$

so that $[\mathrm{Id}_{\mathcal{V}}]_{B_1,B_2}$ is exactly the matrix P in (9.4), i.e., the coordinate transformation matrix in Theorem 9.25. On the other hand,

$$(w_1, \ldots, w_n) = (v_1, \ldots, v_n)\,[\mathrm{Id}_{\mathcal{V}}]_{B_2,B_1},$$

and thus

$$\left([\mathrm{Id}_{\mathcal{V}}]_{B_1,B_2}\right)^{-1} = [\mathrm{Id}_{\mathcal{V}}]_{B_2,B_1}.$$

Example 10.15

(1) Consider the vector space $\mathbb{Q}[t]_{\le 1}$ with the bases $B_1 = \{1, t\}$ and $B_2 = \{t + 1, t - 1\}$. Then the linear map

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

has the matrix representations

$$[f]_{B_1,B_1} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \quad [f]_{B_1,B_2} = \begin{bmatrix} \frac{1}{2} & 1 \\ -\frac{1}{2} & 1 \end{bmatrix}, \quad [f]_{B_2,B_2} = \begin{bmatrix} \frac{3}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{3}{2} \end{bmatrix}.$$

(2) For the vector space $K[t]_{\le n}$ with the basis $B = \{t^0, t^1, \ldots, t^n\}$ and the linear map

$$f : K[t]_{\leq n} \to K[t]_{\leq n},$$
$$\alpha_n t^n + \alpha_{n-1} t^{n-1} + \ldots + \alpha_1 t + \alpha_0 \mapsto \alpha_0 t^n + \alpha_1 t^{n-1} + \ldots + \alpha_{n-1} t + \alpha_n,$$

we have $f(t^j) = t^{n-j}$ for $j = 0, 1, \ldots, n$, so that

$$[f]_{B,B} = \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix} \in K^{n+1,n+1}.$$

Thus, $[f]_{B,B}$ is a permutation matrix.

Theorem 10.16 *Let $\mathcal{V}$ and $\mathcal{W}$ be finite dimensional K-vector spaces with bases $B_1 = \{v_1, \ldots, v_m\}$ and $B_2 = \{w_1, \ldots, w_n\}$, respectively. Then the map*

$$\mathcal{L}(\mathcal{V}, \mathcal{W}) \to K^{n,m}, \quad f \mapsto [f]_{B_1,B_2},$$

is an isomorphism. Hence $\mathcal{L}(\mathcal{V}, \mathcal{W}) \cong K^{n,m}$ and $\dim(\mathcal{L}(\mathcal{V}, \mathcal{W})) = \dim(K^{n,m}) = n \cdot m$.

Proof In this proof we denote the map $f \mapsto [f]_{B_1,B_2}$ by mat, i.e., $\mathrm{mat}(f) = [f]_{B_1,B_2}$. We first show that this map is linear. Let $f, g \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, $\mathrm{mat}(f) = [f_{ij}]$ and $\mathrm{mat}(g) = [g_{ij}]$. For $j = 1, \ldots, m$ we have

$$(f+g)(v_j) = f(v_j) + g(v_j) = \sum_{i=1}^{n} f_{ij} w_i + \sum_{i=1}^{n} g_{ij} w_i = \sum_{i=1}^{n} (f_{ij} + g_{ij}) w_i,$$

and thus $\mathrm{mat}(f+g) = [f_{ij} + g_{ij}] = [f_{ij}] + [g_{ij}] = \mathrm{mat}(f) + \mathrm{mat}(g)$. For $\lambda \in K$ and $j = 1, \ldots, m$ we have

$$(\lambda f)(v_j) = \lambda f(v_j) = \lambda \sum_{i=1}^{n} f_{ij} w_i = \sum_{i=1}^{n} (\lambda f_{ij}) w_i,$$

and thus $\mathrm{mat}(\lambda f) = [\lambda f_{ij}] = \lambda [f_{ij}] = \lambda\, \mathrm{mat}(f)$.

It remains to show that mat is bijective. If $f \in \ker(\mathrm{mat})$, i.e., $\mathrm{mat}(f) = 0 \in K^{n,m}$, then $f(v_j) = 0$ for $j = 1, \ldots, m$. Thus, $f(v) = 0$ for all $v \in \mathcal{V}$, so that $f = 0$ (the zero map) and mat is injective (cp. (5) in Lemma 10.7). If, on the other hand, $A = [a_{ij}] \in K^{n,m}$ is arbitrary, we define the linear map $f : \mathcal{V} \to \mathcal{W}$ via $f(v_j) := \sum_{i=1}^{n} a_{ij} w_i$, $j = 1, \ldots, m$ (cp. the proof of Theorem 10.4). Then $\mathrm{mat}(f) = A$ and hence mat is also surjective (cp. (4) in Lemma 10.7).

Corollary 10.12 now shows that $\dim(\mathcal{L}(\mathcal{V}, \mathcal{W})) = \dim(K^{n,m}) = n \cdot m$ (cp. also Example 9.20). □

Theorem 10.16 shows, in particular, that $f, g \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ satisfy $f = g$ if and only if $[f]_{B_1,B_2} = [g]_{B_1,B_2}$ holds for given bases B_1 of $\mathcal{V}$ and B_2 of $\mathcal{W}$. Thus, we can prove the equality of linear maps via the equality of their matrix representations.

We now consider the map from the elements of a finite dimensional vector space to their coordinates with respect to a given basis.

Lemma 10.17 *If $B = \{v_1, \dots, v_n\}$ is a basis of a K-vector space $\mathcal{V}$, then the map*

$$\Phi_B : \mathcal{V} \to K^{n,1}, \quad v = \lambda_1 v_1 + \ldots + \lambda_n v_n \mapsto \Phi_B(v) := \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix},$$

is an isomorphism, called the coordinate map *of $\mathcal{V}$ with respect to the basis B.*

Proof The linearity of Φ_B is clear. Moreover, we obviously have $\Phi_B(\mathcal{V}) = K^{n,1}$, i.e., Φ_B is surjective. If $v \in \ker(\Phi_B)$, i.e., $\lambda_1 = \cdots = \lambda_n = 0$, then $v = 0$, so that $\ker(\Phi_B) = \{0\}$ and Φ_B is also injective (cp. (5) in Lemma 10.7). □

Example 10.18 In the vector space $K[t]_{\le n}$ with the basis $B = \{t^0, t^1, \dots, t^n\}$ we have

$$\Phi_B(\alpha_n t^n + \alpha_{n-1} t^{n-1} + \ldots + \alpha_1 t + \alpha_0) = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix} \in K^{n+1}.$$

On the other hand, the basis $\widetilde{B} = \{t^n, t^{n-1}, \dots, t^0\}$ yields

$$\Phi_{\widetilde{B}}(\alpha_n t^n + \alpha_{n-1} t^{n-1} + \ldots + \alpha_1 t + \alpha_0) = \begin{bmatrix} \alpha_n \\ \alpha_{n-1} \\ \vdots \\ \alpha_0 \end{bmatrix} \in K^{n+1}.$$

If B_1 and B_2 are bases of the finite dimensional vector spaces $\mathcal{V}$ and $\mathcal{W}$, respectively, then we can illustrate the meaning and the construction of the matrix representation $[f]_{B_1,B_2}$ of $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ in the following *commutative diagram*:

$$\begin{array}{ccc} \mathcal{V} & \xrightarrow{\quad f \quad} & \mathcal{W} \\ {\scriptstyle \Phi_{B_1}} \downarrow & & \downarrow {\scriptstyle \Phi_{B_2}} \\ K^{m,1} & \xrightarrow{[f]_{B_1,B_2}} & K^{n,1} \end{array}$$

We see that different compositions of maps yield the same result. In particular, we have

$$f = \Phi_{B_2}^{-1} \circ [f]_{B_1,B_2} \circ \Phi_{B_1}, \tag{10.3}$$

where the matrix $[f]_{B_1,B_2} \in K^{n,m}$ is interpreted as a linear map from $K^{m,1}$ to $K^{n,1}$, and we use that the coordinate map Φ_{B_2} is bijective and hence invertible. In the same way we obtain

$$\Phi_{B_2} \circ f \;=\; [f]_{B_1,B_2} \circ \Phi_{B_1},$$

i.e.,

$$\Phi_{B_2}(f(v)) = [f]_{B_1,B_2}\Phi_{B_1}(v) \quad \text{for all } v \in \mathcal{V}. \tag{10.4}$$

In words, the coordinates of $f(v)$ with respect to the basis B_2 of $\mathcal{W}$ are given by the product of $[f]_{B_1,B_2}$ and the coordinates of v with respect to the basis B_1 of $\mathcal{V}$.

We next show that the consecutive application of linear maps corresponds to the multiplication of their matrix representations.

Theorem 10.19 *Let $\mathcal{V}$, $\mathcal{W}$ and $\mathcal{X}$ be K-vector spaces. If $f \in \mathcal{L}(\mathcal{V},\mathcal{W})$ and $g \in \mathcal{L}(\mathcal{W},\mathcal{X})$, then $g \circ f \in \mathcal{L}(\mathcal{V},\mathcal{X})$. Moreover, if $\mathcal{V}$, $\mathcal{W}$ and $\mathcal{X}$ are finite dimensional with respective bases B_1, B_2 and B_3, then*

$$[g \circ f]_{B_1,B_3} \;=\; [g]_{B_2,B_3}\,[f]_{B_1,B_2}.$$

Proof Let $h := g \circ f$. We show first that $h \in \mathcal{L}(\mathcal{V},\mathcal{X})$. For $u, v \in \mathcal{V}$ and $\lambda, \mu \in K$ we have

$$\begin{aligned} h(\lambda u + \mu v) &= g(f(\lambda u + \mu v)) = g(\lambda f(u) + \mu f(v)) \\ &= \lambda g(f(u)) + \mu g(f(v)) = \lambda h(u) + \mu h(v). \end{aligned}$$

Now let $B_1 = \{v_1, \ldots, v_m\}$, $B_2 = \{w_1, \ldots, w_n\}$ and $B_3 = \{x_1, \ldots, x_s\}$. If $[f]_{B_1,B_2} = [f_{ij}]$ and $[g]_{B_2,B_3} = [g_{ij}]$, then for $j = 1, \ldots, m$ we have

$$\begin{aligned} h(v_j) = g(f(v_j)) = g\left(\sum_{k=1}^{n} f_{kj} w_k\right) &= \sum_{k=1}^{n} f_{kj}\, g(w_k) = \sum_{k=1}^{n} f_{kj} \sum_{i=1}^{s} g_{ik} x_i \\ &= \sum_{i=1}^{s}\left(\sum_{k=1}^{n} f_{kj} g_{ik}\right) x_i = \sum_{i=1}^{s} \underbrace{\left(\sum_{k=1}^{n} g_{ik} f_{kj}\right)}_{=:\, h_{ij}} x_i. \end{aligned}$$

Thus, $[h]_{B_1,B_3} = [h_{ij}] = [g_{ij}]\,[f_{ij}] = [g]_{B_2,B_3}\,[f]_{B_1,B_2}$. □

Using this theorem we can study how a change of the bases affects the matrix representation of a linear map.

Corollary 10.20 *Let $\mathcal{V}$ and $\mathcal{W}$ be finite dimensional K-vector spaces with bases B_1, $\widetilde{B}_1$ of $\mathcal{V}$ and B_2, $\widetilde{B}_2$ of $\mathcal{W}$. If $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, then*

$$[f]_{B_1,B_2} = [\mathrm{Id}_{\mathcal{W}}]_{\widetilde{B}_2,B_2}[f]_{\widetilde{B}_1,\widetilde{B}_2}[\mathrm{Id}_{\mathcal{V}}]_{B_1,\widetilde{B}_1}. \tag{10.5}$$

In particular, the matrices $[f]_{B_1,B_2}$ and $[f]_{\widetilde{B}_1,\widetilde{B}_2}$ are equivalent.

Proof Applying Theorem 10.19 twice to the identity $f = \mathrm{Id}_{\mathcal{W}} \circ f \circ \mathrm{Id}_{\mathcal{V}}$ yields

$$\begin{aligned}[f]_{B_1,B_2} &= [(\mathrm{Id}_{\mathcal{W}} \circ f) \circ \mathrm{Id}_{\mathcal{V}}]_{B_1,B_2}\\ &= [\mathrm{Id}_{\mathcal{W}} \circ f]_{\widetilde{B}_1,B_2}\,[\mathrm{Id}_{\mathcal{V}}]_{B_1,\widetilde{B}_1}\\ &= [\mathrm{Id}_{\mathcal{W}}]_{\widetilde{B}_2,B_2}\,[f]_{\widetilde{B}_1,\widetilde{B}_2}\,[\mathrm{Id}_{\mathcal{V}}]_{B_1,\widetilde{B}_1}.\end{aligned}$$

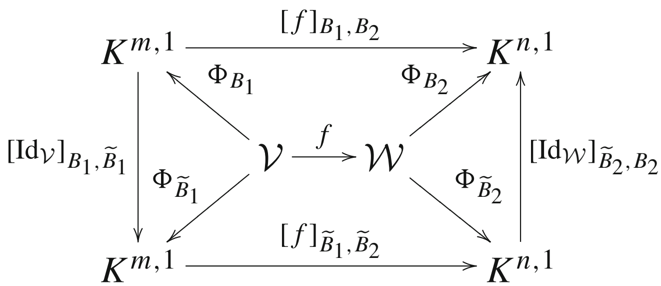
The matrices $[f]_{B_1,B_2}$ and $[f]_{\widetilde{B}_1,\widetilde{B}_2}$ are equivalent, since both $[\mathrm{Id}_{\mathcal{W}}]_{\widetilde{B}_2,B_2}$ and $[\mathrm{Id}_{\mathcal{V}}]_{B_1,\widetilde{B}_1}$ are invertible. □

If $\mathcal{V} = \mathcal{W}$, $B_1 = B_2$, and $\widetilde{B}_1 = \widetilde{B}_2$, then (10.5) becomes

$$[f]_{B_1,B_1} = [\mathrm{Id}_{\mathcal{V}}]_{\widetilde{B}_1,B_1}[f]_{\widetilde{B}_1,\widetilde{B}_1}[\mathrm{Id}_{\mathcal{V}}]_{B_1,\widetilde{B}_1} = ([\mathrm{Id}_{\mathcal{V}}]_{B_1,\widetilde{B}_1})^{-1}[f]_{\widetilde{B}_1,\widetilde{B}_1}[\mathrm{Id}_{\mathcal{V}}]_{B_1,\widetilde{B}_1}.$$

Thus, the matrix representations $[f]_{B_1,B_1}$ and $[f]_{\widetilde{B}_1,\widetilde{B}_1}$ of the endomorphism $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ are *similar* (cp. Definition 8.11).

The following commutative diagram illustrates Corollary 10.20:



(10.6)

Analogously to (10.3) we have

$$f = \Phi_{B_2}^{-1} \circ [f]_{B_1,B_2} \circ \Phi_{B_1} = \Phi_{\widetilde{B}_2}^{-1} \circ [f]_{\widetilde{B}_1,\widetilde{B}_2} \circ \Phi_{\widetilde{B}_1}.$$

Example 10.21 For the following bases of the vector space $\mathbb{Q}^{2,2}$,

$$B_1 = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\},$$

$$B_2 = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \right\},$$

we have the coordinate transformation matrices

$$[\mathrm{Id}_{\mathcal{V}}]_{B_1,B_2} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

and

$$[\mathrm{Id}_{\mathcal{V}}]_{B_2,B_1} = ([\mathrm{Id}_{\mathcal{V}}]_{B_1,B_2})^{-1} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

The coordinate maps are

$$\Phi_{B_1}\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{bmatrix}, \quad \Phi_{B_2}\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = \begin{bmatrix} a_{22} \\ a_{11} - a_{12} - a_{22} \\ a_{12} \\ a_{21} \end{bmatrix},$$

and one can easily verify that

$$\Phi_{B_2}\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = ([\mathrm{Id}_{\mathcal{V}}]_{B_1,B_2} \circ \Phi_{B_1})\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right).$$

Theorem 10.22 *Let $\mathcal{V}$ and $\mathcal{W}$ be K-vector spaces with* $\dim(\mathcal{V}) = m$ *and* $\dim(\mathcal{W}) = n$*, respectively. Then there exist bases B_1 of $\mathcal{V}$ and B_2 of $\mathcal{W}$ such that*

$$[f]_{B_1,B_2} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} \in K^{n,m},$$

where $0 \le r = \dim(\mathrm{im}(f)) \le \min\{n, m\}$*. Furthermore,* $r = \mathrm{rank}(F)$*, where F is the matrix representation of f with respect to arbitrary bases of $\mathcal{V}$ and $\mathcal{W}$, and we define* $\mathrm{rank}(f) := \mathrm{rank}(F) = \dim(\mathrm{im}(f))$.

Proof Let $\widetilde{B}_1 = \{\widetilde{v}_1, \ldots, \widetilde{v}_m\}$ and $\widetilde{B}_2 = \{\widetilde{w}_1, \ldots, \widetilde{w}_n\}$ be two arbitrary bases of $\mathcal{V}$ and $\mathcal{W}$, respectively. Let $r := \mathrm{rank}([f]_{\widetilde{B}_1,\widetilde{B}_2})$. Then by Theorem 5.11 there exist invertible matrices $Q \in K^{n,n}$ and $Z \in K^{m,m}$ with

$$Q\,[f]_{\widetilde{B}_1,\widetilde{B}_2}\,Z = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}, \tag{10.7}$$

where $r = \operatorname{rank}([f]_{\widetilde{B}_1,\widetilde{B}_2}) \leq \min\{n,m\}$. Let us introduce two new bases $B_1 = \{v_1, \ldots, v_m\}$ and $B_2 = \{w_1, \ldots, w_n\}$ of $\mathcal{V}$ and $\mathcal{W}$ via

$$(v_1, \ldots, v_m) := (\widetilde{v}_1, \ldots, \widetilde{v}_m)Z,$$
$$(w_1, \ldots, w_n) := (\widetilde{w}_1, \ldots, \widetilde{w}_n)Q^{-1}, \quad \text{hence} \quad (\widetilde{w}_1, \ldots, \widetilde{w}_n) = (w_1, \ldots, w_n)Q.$$

Then, by construction,

$$Z = [\mathrm{Id}_{\mathcal{V}}]_{B_1,\widetilde{B}_1}, \quad Q = [\mathrm{Id}_{\mathcal{W}}]_{\widetilde{B}_2,B_2}.$$

From (10.7) and Corollary 10.20 we obtain

$$\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} = [\mathrm{Id}_{\mathcal{W}}]_{\widetilde{B}_2,B_2} \, [f]_{\widetilde{B}_1,\widetilde{B}_2} \, [\mathrm{Id}_{\mathcal{V}}]_{B_1,\widetilde{B}_1} = [f]_{B_1,B_2}\,.$$

We thus have found bases B_1 and B_2 that yield the desired matrix representation of f. Every other choice of bases leads, by Corollary 10.20, to an equivalent matrix which therefore also has rank r. It remains to show that $r = \dim(\operatorname{im}(f))$.

The structure of the matrix $[f]_{B_1,B_2}$ shows that

$$f(v_j) = \begin{cases} w_j, & 1 \leq j \leq r, \\ 0, & r+1 \leq j \leq m. \end{cases}$$

Therefore, $v_{r+1}, \ldots, v_m \in \ker(f)$, which implies that $\dim(\ker(f)) \geq m - r$. On the other hand, $w_1, \ldots, w_j \in \operatorname{im}(f)$ and thus $\dim(\operatorname{im}(f)) \geq r$. Theorem 10.9 yields

$$\dim(\mathcal{V}) = m = \dim(\operatorname{im}(f)) + \dim(\ker(f)),$$

and hence $\dim(\ker(f)) = m - r$ and $\dim(\operatorname{im}(f)) = r$. □

Example 10.23 For $A \in K^{n,m}$ and the corresponding map $A \in \mathcal{L}(K^{m,1}, K^{n,1})$ from (1) in Examples 10.2 and 10.6, we have $\operatorname{im}(A) = \operatorname{span}\{a_1, \ldots, a_m\}$. Thus, $\operatorname{rank}(A)$ is equal to the number of linearly independent columns of A. Since $\operatorname{rank}(A) = \operatorname{rank}(A^T)$ (cp. (4) in Theorem 5.11), this number is equal to the number of linearly independent rows of A.

Theorem 10.22 is a first example of a general strategy that we will use several times in the following chapters:

By choosing appropriate bases, the matrix representation should reveal a desired information about a linear map in an efficient way.

In Theorem 10.22 this information is the rank of the linear map f, i.e., the dimension of its image.

The dimension formula for linear maps can be generalized to the composition of maps as follows.

Theorem 10.24 *If $\mathcal{V}$, $\mathcal{W}$ and $\mathcal{X}$ are finite dimensional K-vector spaces, $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ and $g \in \mathcal{L}(\mathcal{W}, \mathcal{X})$, then*

$$\dim(\operatorname{im}(g \circ f)) = \dim(\operatorname{im}(f)) - \dim(\operatorname{im}(f) \cap \ker(g)).$$

Proof Let $\widetilde{g} := g|_{\operatorname{im}(f)}$ be the restriction of g to the image of f, i.e., the map

$$\widetilde{g} \in \mathcal{L}(\operatorname{im}(f), \mathcal{X}), \quad v \mapsto g(v).$$

Applying Theorem 10.9 to $\widetilde{g}$ yields

$$\dim(\operatorname{im}(f)) = \dim(\operatorname{im}(\widetilde{g})) + \dim(\ker(\widetilde{g})).$$

Now

$$\operatorname{im}(\widetilde{g}) = \{g(v) \in \mathcal{X} \mid v \in \operatorname{im}(f)\} = \operatorname{im}(g \circ f)$$

and

$$\ker(\widetilde{g}) = \{v \in \operatorname{im}(f) \mid \widetilde{g}(v) = 0\} = \operatorname{im}(f) \cap \ker(g),$$

imply the assertion. □

Note that Theorem 10.22 with $\mathcal{V} = \mathcal{W}$, $f = \operatorname{Id}_{\mathcal{V}}$, and $g \in \mathcal{L}(\mathcal{V}, \mathcal{X})$ gives $\dim(\operatorname{im}(g)) = \dim(\mathcal{V}) - \dim(\ker(g)$, which is equivalent to Theorem 10.9.

If we interpret matrices $A \in K^{n,m}$ and $B \in K^{s,n}$ as linear maps, then Theorem 10.24 implies the equation

$$\operatorname{rank}(BA) = \operatorname{rank}(A) - \dim(\operatorname{im}(A) \cap \ker(B)).$$

For the special case $K = \mathbb{R}$ and $B = A^T$ we have the following result.

Corollary 10.25 *If $A \in \mathbb{R}^{n,m}$, then* $\operatorname{rank}(A^T A) = \operatorname{rank}(A)$.

Proof Let $w = [\omega_1, \ldots, \omega_n]^T \in \operatorname{im}(A) \cap \ker(A^T)$. Then $w = Ay$ for a vector $y \in \mathbb{R}^{m,1}$. Multiplying this equation from the left by A^T, and using that $w \in \ker(A^T)$, we obtain $0 = A^T w = A^T Ay$, which implies

$$0 = y^T A^T Ay = w^T w = \sum_{j=1}^{n} \omega_j^2.$$

Since this holds only for $w = 0$, we have $\operatorname{im}(A) \cap \ker(A^T) = \{0\}$. □

Exercises

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

10.1 Consider the linear map on $\mathbb{R}^{3,1}$ given by the matrix $A = \begin{bmatrix} 2 & 0 & 1 \\ 2 & 1 & 0 \\ 4 & 1 & 1 \end{bmatrix} \in \mathbb{R}^{3,3}$.
Determine $\ker(A)$, $\dim(\ker(A))$ and $\dim(\mathrm{im}(A))$.

10.2 Construct a map $f \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ such that for linearly independent vectors $v_1, \ldots, v_r \in \mathcal{V}$ the images $f(v_1), \ldots, f(v_r) \in \mathcal{W}$ are linearly dependent.

10.3 The map

$$f : \mathbb{R}[t]_{\leq n} \to \mathbb{R}[t]_{\leq n-1},$$
$$\alpha_n t^n + \alpha_{n-1} t^{n-1} + \ldots + \alpha_1 t + \alpha_0 \mapsto n\alpha_n t^{n-1} + (n-1)\alpha_{n-1} t^{n-2} + \ldots + 2\alpha_2 t + \alpha_1,$$

is called the *derivative* of the polynomial $p \in \mathbb{R}[t]_{\leq n}$ with respect to the variable t. Show that f is linear and determine $\ker(f)$ and $\mathrm{im}(f)$.

10.4 For the bases $B_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ of $\mathbb{R}^{3,1}$ and $B_2 = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$ of $\mathbb{R}^{2,1}$, let $f \in \mathcal{L}(\mathbb{R}^{3,1}, \mathbb{R}^{2,1})$ have the matrix representation $[f]_{B_1,B_2} = \begin{bmatrix} 0 & 2 & 3 \\ 1 & -2 & 0 \end{bmatrix}$.

(a) Determine $[f]_{\widetilde{B}_1,\widetilde{B}_2}$ for the bases $\widetilde{B}_1 = \left\{ \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} \right\}$ of $\mathbb{R}^{3,1}$ and $\widetilde{B}_2 = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ of $\mathbb{R}^{2,1}$.

(b) Determine the coordinates of $f([4,\ 1,\ 3]^T)$ with respect to the basis $\widetilde{B}_2$.

10.5 Construct a map $f \in \mathcal{L}(K[t], K[t])$ with the following properties:

(1) $f(pq) = (f(p))q + p(f(q))$ for all $p, q \in K[t]$.
(2) $f(t) = 1$.

Is this map uniquely determined by these properties or are there further maps with the same properties?

10.6 Let $\alpha \in K$ and $A \in K^{n,n}$. Show that the maps

$$K[t] \to K, \quad p \mapsto p(\alpha), \qquad \text{and} \qquad K[t] \to K^{m,m}, \quad p \mapsto p(A),$$

are linear and justify the name *evaluation homomorphism* for this map.

10.7 Let $S \in GL_n(K)$. Show that the map $f : K^{n,n} \to K^{n,n}$, $A \mapsto S^{-1}AS$ is an isomorphism.

10.8 Let K be a field with $1 + 1 \neq 0$ and let $A \in K^{n,n}$. Consider the map

$$f : K^{n,1} \to K, \quad x \mapsto x^T A x.$$

Is f a linear map? Show that $f = 0$ if and only if $A + A^T = 0$.

10.9 Let $\mathcal{V}$ be a $\mathbb{Q}$-vector space with the basis $B_1 = \{v_1, \ldots, v_n\}$ and let $f \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ be defined by

$$f(v_j) = \begin{cases} v_j + v_{j+1}, & j = 1, \ldots, n-1, \\ v_1 + v_n, & j = n. \end{cases}$$

(a) Determine $[f]_{B_1,B_1}$.

(b) Let $B_2 = \{w_1, \ldots, w_n\}$ with $w_j = j v_{n+1-j}$, $j = 1, \ldots, n$. Show that B_2 is a basis of $\mathcal{V}$. Determine the coordinate transformation matrices $[\mathrm{Id}_{\mathcal{V}}]_{B_1,B_2}$ and $[\mathrm{Id}_{\mathcal{V}}]_{B_2,B_1}$, as well as the matrix representations $[f]_{B_1,B_2}$ and $[f]_{B_2,B_2}$.

10.10 Can you extend Theorem 10.19 consistently to the case $\mathcal{W} = \{0\}$? What are the properties of the matrices $[g \circ f]_{B_1,B_3}$, $[g]_{B_2,B_3}$ and $[f]_{B_1,B_2}$?

10.11 Consider the map

$$\begin{aligned} f :\ & \mathbb{R}[t]_{\leq n} \to \mathbb{R}[t]_{\leq n+1}, \\ & \alpha_n t^n + \alpha_{n-1} t^{n-1} + \ldots + \alpha_1 t + \alpha_0 \mapsto \frac{1}{n+1} \alpha_n t^{n+1} \\ & \qquad + \frac{1}{n} \alpha_{n-1} t^n + \ldots + \frac{1}{2} \alpha_1 t^2 + \alpha_0 t. \end{aligned}$$

(a) Show that f is linear. Determine $\ker(f)$ and $\mathrm{im}(f)$.

(b) Choose bases B_1, B_2 in the two vector spaces and verify that for your choice $\mathrm{rank}([f]_{B_1,B_2}) = \dim(\mathrm{im}(f))$ holds.

10.12 Let $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, $n \geq 2$, be pairwise distinct numbers and let n polynomials in $\mathbb{R}[t]$ be defined by

$$p_j = \prod_{\substack{k=1 \\ k \neq j}}^{n} \left(\frac{1}{\alpha_j - \alpha_k} (t - \alpha_k) \right), \quad j = 1, \ldots, n.$$

(a) Show that the set $B = \{p_1, \ldots, p_n\}$ is a basis of $\mathbb{R}[t]_{\leq n-1}$. (This basis is called the *Lagrange basis*[2] of $\mathbb{R}[t]_{\leq n-1}$.)

(b) Show that the corresponding coordinate map is given by

[2] Joseph-Louis de Lagrange (1736–1813).

$$\Phi_B : \mathbb{R}[t]_{\leq n-1} \to \mathbb{R}^{n,1}, \quad p \mapsto \begin{bmatrix} p(\alpha_1) \\ \vdots \\ p(\alpha_n) \end{bmatrix}.$$

(*Hint:* You can use Exercise 7.8 (b).)

10.13 Verify different paths in the commutative diagram (10.6) for the vector spaces and bases of Example 10.21 and linear map $f : \mathbb{Q}^{2,2} \to \mathbb{Q}^{2,2}$, $A \mapsto FA$ with $F = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$.