Note 10

Linear maps between vector spaces

Given two vector spaces V_1 and V_2, both over the same field $\mathbb{F}$, a linear map is a function from V_1 to V_2 that is compatible with scalar multiplication and vector addition. More precisely, we have the following:

Definition 10.1

Let V_1 and V_2 be vector spaces over a field $\mathbb{F}$. Then a *linear map* from V_1 to V_2 is a function $L : V_1 \to V_2$ such that:

1. $L(\mathbf{u} + \mathbf{v}) = L(\mathbf{u}) + L(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in V_1$,
2. $L(c \cdot \mathbf{u}) = c \cdot L(\mathbf{u})$ for all $c \in \mathbb{F}$ and $\mathbf{u} \in V_1$.

A linear map is also called a *linear transformation*. Note that in the formula $L(\mathbf{u} + \mathbf{v}) = L(\mathbf{u}) + L(\mathbf{v})$, the $+$ in $\mathbf{u} + \mathbf{v}$ denotes vector addition in V_1, while the $+$ in $L(\mathbf{u}) + L(\mathbf{v})$ denotes vector addition in V_2. Similarly, in the formula $L(c \cdot \mathbf{u}) = c \cdot L(\mathbf{u})$, the $\cdot$ in $c \cdot \mathbf{u}$ denotes the scalar multiplication in V_1, while in $c \cdot L(\mathbf{u})$, it denotes the scalar multiplication in V_2.

While in the previous chapter, we studied one vector space at the time, linear maps connect different vector spaces with each other. Linear maps respect the vector space structure: choosing the scalar c equal to 0 and using equation (9-1), one obtains for example

$$L(\mathbf{0}) = \mathbf{0}, \tag{10-1}$$

where the $\mathbf{0}$ on the left-hand side of the equation denotes the zero vector in V_1 and the one on the right denotes the zero vector in V_2. Similarly, choosing $c = -1$ and using equation (9-2), one obtains that

$$L(-\mathbf{u}) = -L(\mathbf{u}). \tag{10-2}$$

Of course, there are many possible functions between two vector spaces and in general not many will be linear. Let us consider some examples.

Example 10.2

Consider the following function from $\mathbb{R}$ to $\mathbb{R}$. Which ones are linear maps?

1. $f : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto x^2$,
2. $g : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto 2x + 1$,
3. $h : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto 2x$.

Answer:

1. $f : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto x^2$. This is not a linear map. We have for example $f(1+1) = f(2) = 4$, but if f would have been a linear map, we should have had $f(1+1) = f(1) + f(1) = 1 + 1 = 2$.
2. $g : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto 2x + 1$. This is not a linear map either, even though the graph of this function is a line. We have $g(0) = 1$, but if g would have been a linear map, we should have had $g(0) = 0$ by equation 10-1.
3. $h : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto 2x$. This is a linear map. For all $x, y \in \mathbb{R}$ we have $h(x + y) = 2(x + y) = 2x + 2y = h(x) + h(y)$ and for all $c \in \mathbb{R}$ and $x \in \mathbb{R}$, we have $c \cdot h(x) = c2x = 2cx = h(c \cdot x)$.

More general, linear maps from $\mathbb{R}$ to $\mathbb{R}$ are precisely those functions whose graph is a straight line passing through the origin. In other words, they are functions $L : \mathbb{R} \to \mathbb{R}$ such that $x \mapsto a \cdot x$ for some constant $a \in \mathbb{R}$. The reason is that if $L : \mathbb{R} \to \mathbb{R}$ is a linear map, then for all $x \in \mathbb{R}$, we have $L(x) = L(x \cdot 1) = x \cdot L(1)$. In the last equality, we used property 2 from Definition 10.1. Setting $a = L(1)$, we indeed obtain that $L(x) = a \cdot x$ for all $x \in \mathbb{R}$.

We will see that there is a strong connection between linear maps and matrices. For this reason, we start with studying linear maps coming from matrices and afterwards return to studying linear maps in a more general setting.

10.1 Linear maps using matrices

Let us start by defining a large class of linear maps.

Definition 10.3

Let $\mathbb{F}$ be a field. Given a matrix $\mathbf{A} \in \mathbb{F}^{m\times n}$, define the function $L_\mathbf{A} : \mathbb{F}^n \to \mathbb{F}^m$ by defining $L_\mathbf{A}(\mathbf{v}) = \mathbf{A} \cdot \mathbf{v}$ for all $\mathbf{v} \in \mathbb{F}^n$.

It turns out that all functions $L_\mathbf{A} : \mathbb{F}^n \to \mathbb{F}^m$ defined above are linear.

Lemma 10.4

The function $L_\mathbf{A} : \mathbb{F}^n \to \mathbb{F}^m$ in Definition 10.3 is a linear map.

Proof. We need to check the two conditions from Definition 10.1. First of all

$$\begin{aligned} L_\mathbf{A}(\mathbf{u}+\mathbf{v}) &= \mathbf{A}\cdot(\mathbf{u}+\mathbf{v}) \\ &= \mathbf{A}\cdot\mathbf{u}+\mathbf{A}\cdot\mathbf{v} \\ &= L_\mathbf{A}(\mathbf{u})+L_\mathbf{A}(\mathbf{v}), \text{ for all } \mathbf{u},\mathbf{v}\in\mathbb{F}^n. \end{aligned}$$

Secondly:

$$L_\mathbf{A}(c\cdot\mathbf{u}) = \mathbf{A}\cdot(c\cdot\mathbf{u}) = c\cdot(\mathbf{A}\cdot\mathbf{u}) = c\cdot L_\mathbf{A}(\mathbf{u}) \text{ for all } c\in\mathbb{F} \text{ and } \mathbf{u}\in\mathbb{F}^n.$$

□

In Example 10.2, we saw that the function $h : \mathbb{R} \to \mathbb{R}$, $x \mapsto 2x$ was a linear map. It is actually a very special case of Definition 10.3: if we choose $n = m = 1$, $\mathbb{F} = \mathbb{R}$ and $\mathbf{A} = [2]$ in Definition 10.3, we find the function h. Instead of $\mathbf{A} = [2]$, we could also just have written $\mathbf{A} = 2$. Indeed, when writing down a 1×1 matrix, it is quite common to leave the brackets [ ] out.

Example 10.5

Let $\mathbb{F} = \mathbb{R}$ and choose

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \in \mathbb{R}^{2\times 4}.$$

Then the corresponding linear map $L_{\mathbf{A}} : \mathbb{R}^4 \to \mathbb{R}^2$ works as follows:

$$L_{\mathbf{A}}\left(\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \begin{bmatrix} v_1 + v_2 + v_3 + v_4 \\ v_1 + 2v_2 + 3v_3 + 4v_4 \end{bmatrix}.$$

So for example

$$L_{\mathbf{A}}\left(\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad L_{\mathbf{A}}\left(\begin{bmatrix} 0 \\ 2 \\ -1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 5 \end{bmatrix} \quad \text{and} \quad L_{\mathbf{A}}\left(\begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

As Example 10.5, it is possible that a linear map $L_{\mathbf{A}}$ maps a vector to the zero vector. The set of such vectors has a special name:

Definition 10.6

Let $\mathbb{F}$ be a field. Given a matrix $\mathbf{A} \in \mathbb{F}^{m\times n}$, the of the matrix $\mathbf{A}$, denoted by $\ker \mathbf{A}$, is the following set of vectors:

$$\ker \mathbf{A} = \{\mathbf{v} \in \mathbb{F}^n \mid \mathbf{A} \cdot \mathbf{v} = \mathbf{0}\}.$$

Note that one can equivalently define the kernel of a matrix $\mathbf{A}$ to be all vectors from $\mathbb{F}^n$ that are mapped to the zero vector by the linear map $L_{\mathbf{A}}$. We can also think of the vectors in the kernel as precisely those vectors that are solutions to the homogeneous system of linear equations with coefficient matrix $\mathbf{A}$.

|||| Remark 10.7

A remark about terminology is in place here. Some authors prefer to use the words *null space*, *right kernel* or *right null space* for what we have called the kernel of a matrix. The reason for adding the word "right" is that we have multiplied the matrix with a column vector from the right. One could also have considered the set of row vectors $\mathbf{u} \in \mathbb{F}^{1\times m}$ such that $\mathbf{u} \cdot \mathbf{A} = \mathbf{0}$. This set is called the *left kernel* of $\mathbf{A}$ or sometimes also the *left null space*.

One of the reasons that we introduced the notion of kernel of a matrix, is that it actually is a subspace. Let us show this in the following lemma.

|||| Lemma 10.8

Let $\mathbb{F}$ be a field and $\mathbf{A} \in \mathbb{F}^{m\times n}$ a matrix. Then the kernel of $\mathbf{A}$ is a subspace of $\mathbb{F}^n$.

Proof. First of all, note that $\mathbf{0} \in \ker \mathbf{A}$ so that $\ker \mathbf{A}$ is not the empty set. This means that if we set $W = \ker \mathbf{A}$, then the requirement that W is not empty in Lemma 9.33 is met.

Let $\mathbf{u}, \mathbf{v} \in \ker \mathbf{A}$ and $c \in \mathbb{F}$. Then

$$\mathbf{A} \cdot (\mathbf{u} + c \cdot \mathbf{v}) = \mathbf{A} \cdot \mathbf{u} + \mathbf{A} \cdot (c \cdot \mathbf{v}) = \mathbf{A} \cdot \mathbf{u} + c \cdot (\mathbf{A} \cdot \mathbf{v}) = \mathbf{0} + c \cdot \mathbf{0} = \mathbf{0}. \tag{10-3}$$

Here we used that $\mathbf{A} \cdot \mathbf{u} = \mathbf{0}$ and $\mathbf{A} \cdot \mathbf{v} = \mathbf{0}$, since $\mathbf{u}, \mathbf{v} \in \ker \mathbf{A}$. Equation (10-3) implies that $\mathbf{u} + c \cdot \mathbf{v} \in \ker \mathbf{A}$. Then Lemma 9.33 implies that $\ker \mathbf{A}$ is a subspace of $\mathbb{F}^n$. □

The dimension of $\ker \mathbf{A}$ is called the *nullity* of the matrix $\mathbf{A}$. It is denoted by null$\mathbf{A}$.

|||| Example 10.9

Let $\mathbb{F} = \mathbb{R}$ and consider the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix}.$$

Compute a basis for $\ker \mathbf{A}$ and compute the nullity of $\mathbf{A}$.

Answer: The kernel of $\mathbf{A}$ consists of all vectors $\mathbf{v} = (v_1, v_2, v_3, v_4) \in \mathbb{R}^4$ such that $\mathbf{A} \cdot \mathbf{v} = \mathbf{0}$. We have for example seen in Example 10.5 that the vector $(1, -1, -1, 1)$ is mapped to $(0, 0)$ by the linear map $L_{\mathbf{A}}$. Therefore $(1, -1, -1, 1) \in \ker \mathbf{A}$.

We can think of the vectors in the kernel as precisely those vectors that are solutions to the homogeneous system of two linear equations with coefficient matrix $\mathbf{A}$. To describe all these solutions, we follow the same procedure as explained in Example 6.28 and Theorem 6.29. Hence, we first bring the matrix $\mathbf{A}$ in reduced row echelon form:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \underset{R_2 \leftarrow R_2 - R_1}{\longrightarrow} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 \end{bmatrix} \underset{R_1 \leftarrow R_1 - R_2}{\longrightarrow} \begin{bmatrix} 1 & 0 & -1 & -2 \\ 0 & 1 & 2 & 3 \end{bmatrix}.$$

Now we can see that $\mathbf{v} = (v_1, v_2, v_3, v_4) \in \ker \mathbf{A}$ if and only if $v_1 - v_3 - 2v_4 = 0$ and $v_2 + 2v_3 + 3v_4 = 0$. Similarly as in Example 6.28 (or directly using Theorem 6.29), we see that

$$\ker \mathbf{A} = \left\{ t_1 \cdot \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} + t_2 \cdot \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix} \mid t_1, t_2 \in \mathbb{R} \right\}.$$

Hence the vectors

$$\begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix}$$

span $\ker \mathbf{A}$. In fact, Corollary 9.39 tells us that these two vectors form a basis of $\ker \mathbf{A}$. Hence a basis for $\ker \mathbf{A}$ is given by

$$\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -3 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

The nullity of the matrix $\mathbf{A}$ is by definition the dimension of the subspace $\ker \mathbf{A}$. Since we have just computed a basis of $\ker \mathbf{A}$ and this basis consists of two vectors, we conclude that the nullity of $\mathbf{A}$ is two. In other words: $\text{null}\mathbf{A} = 2$.

We have already observed that we can think of the vectors in the kernel as precisely those vectors that are solutions to the homogeneous system of linear equations with coefficient matrix $\mathbf{A}$. Using Corollary 9.39, we obtain the following result, which often is called the *rank-nullity theorem for matrices*.

|||| Theorem 10.10

Let $\mathbb{F}$ be a field and $\mathbf{A} \in \mathbb{F}^{m \times n}$ a matrix. Then

$$\rho(\mathbf{A}) + \text{null}(\mathbf{A}) = n,$$

where $\rho(\mathbf{A})$ denotes the rank of the matrix $\mathbf{A}$ and $\text{null}(\mathbf{A})$ its nullity.

Proof. Using Corollary 9.39, we see that the kernel of $\mathbf{A}$ has a basis containing precisely $n - \rho(\mathbf{A})$ many vectors. Hence $\text{null}(\mathbf{A}) = \dim \ker(\mathbf{A}) = n - \rho(\mathbf{A})$. This implies that $\rho(\mathbf{A}) + \text{null}(\mathbf{A}) = n$. □

We have seen in Lemma 10.8, that the kernel of a matrix $\mathbf{A} \in \mathbb{F}^{m \times n}$ is a linear subspace of $\mathbb{F}^n$. In other words: $\ker \mathbf{A}$ is a linear subspace of the domain of the linear map $L_{\mathbf{A}} : \mathbb{F}^n \to \mathbb{F}^m$. To a matrix $\mathbf{A} \in \mathbb{F}^{m \times n}$ one can also associate a linear subspace of $\mathbb{F}^m$, the codomain of the linear map $L_{\mathbf{A}} : \mathbb{F}^n \to \mathbb{F}^m$. We do this in the following definition.

|||| Definition 10.11

Let $\mathbb{F}$ be a field. Given a matrix $\mathbf{A} \in \mathbb{F}^{m \times n}$, the *column space* of the matrix $\mathbf{A}$, denoted by $\text{colsp}\mathbf{A}$, is the subspace of $\mathbb{F}^m$ spanned by the columns of $\mathbf{A}$. The dimension of the column space of a matrix $\mathbf{A}$ is called the *column rank* of $\mathbf{A}$.

|||| Lemma 10.12

Let $\mathbb{F}$ be a field and $\mathbf{A} \in \mathbb{F}^{m \times n}$ a matrix. Then $\text{colsp}\mathbf{A}$, the column space of the matrix $\mathbf{A}$ is precisely the image of the linear map $L_{\mathbf{A}} : \mathbb{F}^n \to \mathbb{F}^m$.

Proof. An element from the column space of a matrix $\mathbf{A}$ is by definition a linear combination of the columns of $\mathbf{A}$. On the other hand, an element of the image of the linear map $L_{\mathbf{A}} : \mathbb{F}^n \to \mathbb{F}^m$ is of the form $\mathbf{A} \cdot \mathbf{v}$ for some vector $\mathbf{v} = (v_1, \ldots, v_n) \in \mathbb{F}^n$. Using Definition 7.10, we can rewrite this as the linear combination of the columns of $\mathbf{A}$ as follows:

$$\mathbf{A} \cdot \mathbf{v} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = v_1 \cdot \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + \cdots v_n \cdot \begin{bmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{bmatrix}.$$

Hence the image of the linear map $L_{\mathbf{A}}$ consists precisely of all linear combinations of the columns of $\mathbf{A}$. But this is precisely the column space of the matrix $\mathbf{A}$. □

|||| Remark 10.13

Because of Lemma 10.12, the column space of a matrix $\mathbf{A}$ is sometimes also called the *range* or the *image* of $\mathbf{A}$.

We have previously introduced the rank of a matrix in Definition 6.22. The rank $\rho(\mathbf{A})$ of a matrix $\mathbf{A}$ as defined in Definition 6.22 is sometimes more properly called the *row rank* of the matrix $\mathbf{A}$, since one can show that the dimension of the vector space spanned by the rows of $\mathbf{A}$ is equal to $\rho(\mathbf{A})$. It turns out however, that for any matrix, its row rank and column rank are the same. Therefore, we will from now on simply call the column rank of a matrix $\mathbf{A}$, the rank of the matrix and denote it by $\rho(\mathbf{A})$, using the same notation as in Definition 6.22.

It is not obvious from Definitions 6.22 and 10.11 that row rank and column rank of a matrix are always the same. A reader willing to accept this can skip the remainder of this section, but for the interested reader, we give a short proof of why row rank and column rank are always the same.

|||| Theorem 10.14

Let $\mathbb{F}$ be a field and $\mathbf{A} \in \mathbb{F}^{m \times n}$ a matrix. Then the row rank and the column rank of the matrix $\mathbf{A}$ are the same.

Proof. The row rank $\rho(\mathbf{A})$ of a matrix $\mathbf{A}$ is by definition equal to the number of pivots in the reduced row echelon form of $\mathbf{A}$. On the other hand, Theorem 9.38 implies that the number of vectors in a basis of the column space of $\mathbf{A}$ is also equal to this number of pivots. Hence the dimension of the column space of $\mathbf{A}$ is also equal to $\rho(\mathbf{A})$. □

10.2 Linear maps between general vector spaces

In the previous section, we have focused on linear maps coming from matrices, but Definition 10.1 allows for much more general linear maps. It turns out that the notions of kernel and image also make sense in the general setting. Let us first consider some more examples.

|||| Example 10.15

Let $\mathbb{F} = \mathbb{C}$ and consider the complex vector space $\mathbb{C}[Z]$ (see Example 9.6). Recall that $\mathbb{C}[Z]$ denotes the set of all polynomials with coefficients in $\mathbb{C}$. Now consider the map $D : \mathbb{C}[Z] \to \mathbb{C}[Z]$ defined by $D(a_0 + a_1 Z + a_2 Z^2 + \cdots + a_n Z^n) = a_1 + 2a_2 Z + \cdots + n a_n Z^{n-1}$. In words, the map D sends a polynomial $p(Z)$ to its derivative $p(Z)'$. One can show that D is a linear map.

|||| Example 10.16

Let $V_1 = \mathbb{F}^{n\times n}$ and $V_2 = \mathbb{F}$. Given a square matrix $\mathbf{A} \in \mathbb{F}^{n\times n}$, the *trace*, denoted by $\mathrm{Tr}(\mathbf{A})$, is defined as the sum of the elements on its diagonal. In other words:

$$\mathrm{Tr}\left(\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}\right) = a_{11} + \cdots + a_{nn}.$$

Question: Is the map $\mathrm{Tr} : \mathbb{F}^{n\times n} \to \mathbb{F}$, defined by $\mathbf{A} \mapsto \mathrm{Tr}(\mathbf{A})$ a linear map?

Answer: To find out whether or not the trace map $\mathrm{Tr} : \mathbb{F}^{n\times n} \to \mathbb{F}$ as defined above, is linear, we check if all conditions in Definition 10.1 are satisfied. First of all, using the notation from Definition 10.1, we have $V_1 = \mathbb{F}^{n\times n}$ and $V_2 = \mathbb{F}$. We should first check that these are vector spaces over a field $\mathbb{F}$. Both are indeed vector spaces over $\mathbb{F}$: For V_1, see Example 9.5 with $m = n$ and for V_2, see Example 9.2 with $n = 1$.

Now we need to check if Tr satisfies the two conditions from Definition 10.1. Let us choose arbitrary $c \in \mathbb{F}$ and $\mathbf{u}, \mathbf{v} \in \mathbb{F}^{n\times n}$. Hence we can write

$$\mathbf{u} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \quad \text{and} \quad \mathbf{v} = \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \dots & b_{nn} \end{bmatrix}.$$

Then

$$\mathbf{u}+\mathbf{v} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} + \begin{bmatrix} b_{11} & \dots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \dots & b_{nn} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & \dots & a_{1n}+b_{1n} \\ \vdots & & \vdots \\ a_{n1}+b_{n1} & \dots & a_{nn}+b_{nn} \end{bmatrix}$$

and

$$c \cdot \mathbf{u} = c \cdot \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} = \begin{bmatrix} c \cdot a_{11} & \dots & c \cdot a_{1n} \\ \vdots & & \vdots \\ c \cdot a_{n1} & \dots & c \cdot a_{nn} \end{bmatrix}.$$

Hence

$$\mathrm{Tr}(\mathbf{u}+\mathbf{v}) = a_{11}+b_{11}+\cdots a_{nn}+b_{nn} = a_{11}+\cdots a_{nn}+b_{11}+\cdots b_{nn} = \mathrm{Tr}(\mathbf{u})+\mathrm{Tr}(\mathbf{v})$$

and

$$\mathrm{Tr}(c \cdot \mathbf{u}) = c \cdot a_{11} + \cdots c \cdot a_{nn} = c \cdot (a_{11}+\cdots a_{nn}) = c \cdot \mathrm{Tr}(\mathbf{u}).$$

We can conclude that $\mathrm{Tr} : \mathbb{F}^{n \times n} \to \mathbb{F}$, defined by $\mathbf{A} \mapsto \mathrm{Tr}(\mathbf{A})$ is a linear map.

|||| Example 10.17

Let $\mathbb{F} = \mathbb{R}$ and consider the map $m_5 : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $m_5(v_1, v_2) = (5v_1, 5v_2)$. In other words, the effect of map m_5 on a vector is that it multiplies a vector with the scalar 5. Visually, this means that the direction of a vector is not changed, but its length becomes five times longer. One can show that this is a linear map of real vector spaces.

More generally, one can show that if $\mathbb{F}$ is a field and $c \in \mathbb{F}$ is a scalar, then the map $m_c : \mathbb{F}^n \to \mathbb{F}^n$ defined by $m_c(\mathbf{u}) = c \cdot \mathbf{u}$ is a linear map of vector spaces.

|||| Example 10.18

For $\alpha \in \mathbb{R}$, consider the map $R_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $R_\alpha(v_1, v_2) = (\cos(\alpha) \cdot v_1 - \sin(\alpha) \cdot v_2, \sin(\alpha) \cdot v_1 + \cos(\alpha) \cdot v_2)$. Geometrically, the effect of R_α on $(v_1, v_2) \in \mathbb{R}^2$ is a rotation over an angle α against the clock, where the rotation has center in $(0,0)$. For example, if $\alpha = \pi/2$, then $R_{\pi/2}(v_1, v_2) = (-v_2, v_1)$. One can show that R_α is a linear map.

|||| Example 10.19

We choose $\mathbb{F} = \mathbb{C}$. Let V_1 be the set of polynomials in $\mathbb{C}[Z]$ of degree at most three and similarly let V_2 be the set of polynomials in $\mathbb{C}[Z]$ of degree at most four. Both V_1 and V_2 are vector spaces over $\mathbb{C}$. A possible basis for V_1 is given by the set $\{1, Z, Z^2, Z^3\}$, while a basis for V_2 is $\{1, Z, Z^2, Z^3, Z^4\}$. Hence $\dim V_1 = 4$ and $\dim V_2 = 5$. Now define the map $L : V_1 \to V_2$ by $p(Z) \mapsto (i + 2Z) \cdot p(Z)$. Note that indeed for any $p(Z) \in V_1$, we have $(i + 2Z) \cdot p(Z) \in V_2$ using equation (4-1). One can show that L is a linear map. Indeed, if $p_1(Z), p_2(Z) \in V_1$ and $c \in \mathbb{C}$ one has

$$\begin{aligned} L(p_1(Z) + p_2(Z)) &= (i + 2Z) \cdot (p_1(Z) + p_2(Z)) \\ &= (i + 2Z) \cdot p_1(Z) + (i + 2Z) \cdot p_2(Z) \\ &= L(p_1(Z)) + L(p_2(Z)) \end{aligned}$$

and

$$L(c \cdot p_1(Z)) = (i + 2Z) \cdot c \cdot p_1(Z) = c \cdot (i + 2Z) \cdot p_1(Z) = c \cdot L(p_1(Z)).$$

|||| Example 10.20

As a final example of a linear map, we consider the map $\mathrm{ev} : \mathbb{C}[Z] \to \mathbb{C}^2$ defined by $p(Z) \mapsto (p(0), p(1))$. So for example $L(Z^2 + Z + 1) = (0^2 + 0 + 1, 1^2 + 1 + 1) = (1, 3)$. One can show that ev is a linear map.

We finish this section with some general properties of linear maps. First we consider the composition of two linear maps, see Section 2.2 for the definition of the composite of two functions.

|||| Theorem 10.21

Let $\mathbb{F}$ be a field and V_1, V_2, V_3 vector spaces over $\mathbb{F}$. Further, suppose that $L_1 : V_1 \to V_2$ and $L_2 : V_2 \to V_3$ are linear maps. Then the composition $L_2 \circ L_1 : V_1 \to V_3$ is also a linear map.

Proof. Let us choose arbitrary $\mathbf{u}, \mathbf{v} \in V_1$ and $c \in \mathbb{F}$. Then using linearity of L_1 and L_2 as well as the definition of the composition of two functions, we obtain that

$$\begin{aligned} (L_2 \circ L_1)(\mathbf{u} + \mathbf{v}) &= L_2(L_1(\mathbf{u} + \mathbf{v})) \\ &= L_2(L_1(\mathbf{u}) + L_1(\mathbf{v})) \\ &= L_2(L_1(\mathbf{u})) + L_2(L_1(\mathbf{v})) \\ &= (L_2 \circ L_1)(\mathbf{u}) + (L_2 \circ L_1)(\mathbf{v}). \end{aligned}$$

and

$$(L_2 \circ L_1)(c \cdot \mathbf{u}) = L_2(L_1(c \cdot \mathbf{u})) = L_2(c \cdot L_1(\mathbf{u})) = c \cdot L_2(L_1(\mathbf{u})) = c \cdot (L_2 \circ L_1)(\mathbf{u}).$$

Hence by Definition 10.1, the map $L_2 \circ L_1 : V_1 \to V_3$ is a linear map. □

Since any function $f : A \to B$ has an image, namely the set $\text{image}(f) = \{f(a) \mid a \in A\}$, see Section 2.2, a linear map $L : V_1 \to V_2$ has an image as well. In view of Lemma 10.12, this generalizes the idea of the notion of the column space of a matrix to the setting of general linear maps. One can show that the image of a linear map $L : V_1 \to V_2$ is a subspace of V_2. The notion of a kernel can also directly be generalized.

|||| Definition 10.22

Let $\mathbb{F}$ be a field and V_1 and V_2 vector spaces over $\mathbb{F}$. Given a linear map $L : V_1 \to V_2$, the *kernel* of the map L is:

$$\ker L = \{\mathbf{v} \in V_1 \mid L(\mathbf{v}) = \mathbf{0}\}.$$

Similarly as in the case of the kernel of a matrix, one can show that the kernel of a linear map $L : V_1 \to V_2$ is a subspace of V_1.

|||| Example 10.23

Let us revisit Example 10.15. We considered the linear map $D : \mathbb{C}[Z] \to \mathbb{C}[Z]$, sending a polynomial $p(Z)$ to its derivative. The only polynomials whose derivative is 0 are constant polynomials, that is to say polynomials of the form $p(Z) = a_0$. Hence $\ker D = \{a_0 \mid a_0 \in \mathbb{C}\} = \mathbb{C}$. Note that $\{1\}$ is a basis of $\ker D$, so that we can conclude that $\dim \ker D = 1$.

With a view to a later application of the theory to differential equations, we consider another example involving derivatives.

|||| Example 10.24

Let C_∞ be the vector space of all infinitely differentiable functions from $\mathbb{R}$ to $\mathbb{R}$, see Example 9.35. Let us consider the map $L : C_\infty \to C_\infty$ where $f \mapsto f' - f$. As usual f' denotes the derivative of the function f. Since $f \in C_\infty$, also f' is infinitely often differentiable, so that $f' \in C_\infty$. One can show that L is a linear map. Using Definition 10.22, we see that $\ker L = \{f \in C_\infty \mid f' - f = 0\}$. In other words: the kernel of L consists of those functions $f \in C_\infty$ such that the derivative of f is the same as f itself. In yet other words: the kernel of L consists exactly of all solutions in C_∞ of the differential equation $f' = f$. An example of a function satisfying this differential equation is the exponential function $\exp : \mathbb{R} \to \mathbb{R}$ defined by $x \mapsto e^x$. Also all scalar multiples $f = c \cdot \exp$ with $c \in \mathbb{R}$, are solutions to the differential equation $f' = f$. It is in fact possible to show that there are no more solutions in C_∞. Hence $\ker L$ turns out to be a one-dimensional subspace of C_∞ with basis $\{\exp\}$.

|||| Remark 10.25

The exponential function was also discussed in Example 2.22, but there its codomain was defined to be $\mathbb{R}_{\geq 0}$. Strictly speaking, the exponential function from Example 2.22 is therefore not the same function as the exponential function we used in this example. However, since both functions map any $x \in \mathbb{R}$ to exactly the same value, namely e^x, it is a bit overkill to use different notations for these functions. For this reason we have denoted both functions with exp.

|||| Example 10.26

As a final example of the kernel of a linear map, we consider the map ev from Example 10.20. The map $\mathrm{ev} : \mathbb{C}[Z] \to \mathbb{C}^2$ was defined by $p(Z) \mapsto (p(0), p(1))$. Hence we have

$$\ker \mathrm{ev} = \{p(Z) \in \mathbb{C}[Z] \mid (p(0), p(1)) = (0, 0)\} = \{p(Z) \in \mathbb{C}[Z] \mid p(0) = 0 \wedge p(1) = 0\}.$$

It is possible to describe the kernel of ev more specifically. Let us start by describing the set of polynomials $p(Z)$ satisfying $p(0) = 0$, that is to say, such that 0 is a root of $p(Z)$. Using Lemma 4.20, we conclude that

$$\{p(Z) \in \mathbb{C}[Z] \mid p(0) = 0\} = \{Z \cdot q(Z) \mid q(Z) \in \mathbb{C}[Z]\}.$$

Now if both $p(0) = 0$ and $p(1) = 0$, then we see that $p(Z) = Z \cdot q(Z)$ for some $q(Z) \in \mathbb{C}[Z]$, and $p(1) = 0$. But this is equivalent with saying that $p(Z) = Z \cdot q(Z)$ for some $q(Z) \in \mathbb{C}[Z]$ and $q(1) = 0$. Using Lemma 4.20 again, but now for $q(Z)$ and the root 1, we see that $q(Z) = (Z - 1) \cdot s(Z)$ for some $s(Z) \in \mathbb{C}[Z]$. Hence we obtain $p(Z) \in \ker \mathrm{ev}$ if and only if

$p(Z) = Z \cdot (Z-1) \cdot s(X)$ for some $s(Z) \in \mathbb{C}[Z]$. We conclude that

$$\ker \mathrm{ev} = \{p(Z) \in \mathbb{C}[Z] \mid p(0) = 0 \wedge p(1) = 0\} = \{Z \cdot (Z-1) \cdot s(Z) \mid s(Z) \in \mathbb{C}[Z]\}.$$

10.3 Linear maps between finite dimensional vector spaces

Let us assume that we are given a finite dimensional vector space V over a field $\mathbb{F}$, say $\dim V = n$. In such a setting, we can choose an ordered basis of V, say $\beta = (\mathbf{v}_1, \dots, \mathbf{v}_n)$, where $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent vectors in V. As we have seen in Definition 9.17, for each $\mathbf{v} \in V$, we can produce a unique coordinate vector $[\mathbf{v}]_\beta \in \mathbb{F}^n$. This means that we can define a function $\phi_\beta : V \to \mathbb{F}^n$ by $\mathbf{v} \mapsto [\mathbf{v}]_\beta$. Now combining Lemma 9.18 and Definition 10.1, we can immediately conclude that the function ϕ_β is a linear map. Given a vector $(c_1, \dots, c_n) \in \mathbb{F}^n$, it is simple to write down a vector of V having $(c_1, \dots, c_n)$ as its coordinate vector (with respect to β). Indeed, the vector $\mathbf{v} = c_1 \cdot \mathbf{v}_1 + \cdots + c_n \cdot \mathbf{v}_n$ is that vector and it is the only vector with coordinates $(c_1, \dots, c_n)$ according to Lemma 9.16! What we in fact have found is the inverse function of ϕ_β. Let us put these statements in a lemma and give a complete proof.

Lemma 10.27

Let $\mathbb{F}$ be a field, V a vector space over $\mathbb{F}$ of dimension n, and $\beta = (\mathbf{v}_1, \dots, \mathbf{v}_n)$, an ordered basis of V. Then the function $\phi_\beta : V \to \mathbb{F}^n$ defined by $\mathbf{v} \mapsto [\mathbf{v}]_\beta$ is a linear map. Moreover, the function $\psi_\beta : \mathbb{F}^n \to V$ defined by $(c_1, \dots, c_n) \mapsto c_1 \cdot \mathbf{v}_1 + \cdots + c_n \cdot \mathbf{v}_n$ is the inverse of ϕ_β and also a linear map.

Proof. We have already shown in the discussion before this lemma that $\phi_\beta : V \to \mathbb{F}^n$ is a linear map of vector spaces over $\mathbb{F}$. Now let us denote by $\psi_\beta : \mathbb{F}^n \to V$ the map defined by $(c_1, \dots, c_n) \mapsto c_1 \cdot \mathbf{v}_1 + \cdots + c_n \cdot \mathbf{v}_n$. We first show that ψ_β is the inverse function ϕ_β^{-1}. In order to check this, we need to show that $\psi_\beta \circ \phi_\beta(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in V$ as well as that $\phi_\beta \circ \psi_\beta(c_1, \dots, c_n) = (c_1, \dots, c_n)$ for all $(c_1, \dots, c_n) \in \mathbb{F}^n$. We have

$$(\psi_\beta \circ \phi_\beta)(\mathbf{v}) = \psi_\beta([\mathbf{v}]_\beta) = \mathbf{v}$$

and

$$(\phi_\beta \circ \psi_\beta)(c_1, \dots, c_n) = \phi_\beta(c_1 \cdot \mathbf{v}_1 + \cdots + c_n \cdot \mathbf{v}_n) = (c_1, \dots, c_n).$$

It is left to the reader to check that ψ_β is a linear map. $\square$

The reason the linear maps ϕ_β and ψ_β are so useful, is that they can be used to describe a general linear map more explicitly. More to the point, suppose that we are given a linear map $L : V_1 \to V_2$ as in Definition 10.1, but that we know that both V_1 and V_2 are finite dimensional vector spaces, say that $\dim V_1 = m$ and $\dim V_2 = n$. This means that we can choose an ordered basis of V_1, say $\beta = (\mathbf{v}_1, \ldots, \mathbf{v}_m)$, where $\mathbf{v}_1, \ldots, \mathbf{v}_m$ are linearly independent vectors in V_1. Similarly, we can choose an ordered basis of V_2, say $\gamma = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$, where $\mathbf{w}_1, \ldots, \mathbf{w}_n \in V_2$ are linearly independent vectors in V_2. Then instead of studying the abstract linear map $L : V_1 \to V_2$, we will study the function $\phi_\gamma \circ L \circ \psi_\beta : \mathbb{F}^m \to \mathbb{F}^n$. The effect is that the abstract vector spaces V_1 and V_2 have been replaced by the more down to earth vector spaces $\mathbb{F}^m$ and $\mathbb{F}^n$. Using Theorem 10.21 in combination with Lemma 10.27, we can also conclude that the function $\phi_\gamma \circ L \circ \psi_\beta$ actually is a linear map of vector spaces over $\mathbb{F}$, since it is the composite of linear maps.

We have in Section 10.1 seen that any matrix $\mathbf{A} \in \mathbb{F}^{m\times n}$ gives rise to a linear map $L_\mathbf{A} : \mathbb{F}^m \to \mathbb{F}^n$, by defining $\mathbf{v} \mapsto \mathbf{A} \cdot \mathbf{v}$. In fact, any linear map from $\mathbb{F}^m$ to $\mathbb{F}^n$ is of this form. Let us show this now:

|||| Lemma 10.28

Let $\mathbb{F}$ be a field and $\tilde{L} : \mathbb{F}^m \to \mathbb{F}^n$ a linear map. Then there exists exactly one matrix $\mathbf{A} \in \mathbb{F}^{m\times n}$ such that $\tilde{L} = L_\mathbf{A}$. Moreover, if we denote by $\mathbf{e}_1, \ldots, \mathbf{e}_m$ the standard basis of $\mathbb{F}^m$, then $\mathbf{A}$ is the matrix whose columns consist of $\tilde{L}(\mathbf{e}_1), \ldots, \tilde{L}(\mathbf{e}_m)$.

Proof. If $\mathbf{v} = (c_1, \ldots, c_m) \in \mathbb{F}^m$, then $\mathbf{v} = c_1 \cdot \mathbf{e}_1 + \cdots + c_m \cdot \mathbf{e}_m$, since the i-th standard basis vector of $\mathbb{F}^m$ has a one in coordinate i and zeroes otherwise. Since $\tilde{L}$ is a linear map, we have $\tilde{L}(\mathbf{v}) = \tilde{L}(c_1 \cdot \mathbf{e}_1 + \cdots + c_m \cdot \mathbf{e}_m) = c_1 \cdot \tilde{L}(\mathbf{e}_1) + \cdots + c_m \cdot \tilde{L}(\mathbf{e}_m)$. Hence the matrix $\mathbf{A}$ with columns $\tilde{L}(\mathbf{e}_1), \ldots, \tilde{L}(\mathbf{e}_m)$ satisfies that $\mathbf{A} \cdot \mathbf{v} = c_1 \cdot \tilde{L}(\mathbf{e}_1) + \cdots + c_m \cdot \tilde{L}(\mathbf{e}_m) = \tilde{L}(\mathbf{v})$. This shows that $\tilde{L} = L_\mathbf{A}$.

What is left to show is that the matrix $\mathbf{A}$ is unique. Suppose that there exist another matrix $\mathbf{B} \in \mathbb{F}^{m\times n}$ such that $\tilde{L} = L_\mathbf{B}$. We want to show that $\mathbf{A} = \mathbf{B}$. If $\mathbf{A} \neq \mathbf{B}$, one can find a column, say column i, where the matrices $\mathbf{A}$ and $\mathbf{B}$ are distinct. Note that the ith column of $\mathbf{A}$ equals $\tilde{L}(\mathbf{e}_i)$ by construction of the matrix $\mathbf{A}$. On the other hand, $\tilde{L}(\mathbf{e}_i) = L_\mathbf{B}(\mathbf{e}_i) = \mathbf{B} \cdot \mathbf{e}_i$, which is precisely the ith column of $\mathbf{B}$. Apparently, the ith columns of $\mathbf{A}$ and $\mathbf{B}$ are both equal to $L(\mathbf{e}_i)$ and not distinct after all. This contradiction show that the assumption $\mathbf{A} \neq \mathbf{B}$ cannot be valid and therefore that $\mathbf{A} = \mathbf{B}$. $\square$

Given a linear map $L : V_1 \to V_2$ we will apply this lemma to the associated linear map $\tilde{L} = \phi_\gamma \circ L \circ \psi_\beta : \mathbb{F}^m \to \mathbb{F}^n$. Let us before continuing with the general theory, first consider an example.

|||| Example 10.29

We revisit Example 10.19. In that example V_1 was the vector space consisting of polynomials in $\mathbb{C}[Z]$ of degree at most three and V_2 the vector space of polynomials in $\mathbb{C}[Z]$ of degree at most four. Hence as ordered basis for V_1, we can choose $\beta = (1, Z, Z^2, Z^3)$, while a possible ordered basis for V_2 is given by $\gamma = (1, Z, Z^2, Z^3, Z^4)$. The linear map $L : V_1 \to V_2$ described in Example 10.19 mapped a polynomial $p(Z)$ to $(i + 2Z) \cdot p(Z)$.

Let us start by explaining what the linear map $\phi_\gamma : V_2 \to \mathbb{F}^5$ is in this case. An element in V_2 is a polynomial of degree at most four. Hence $\mathbf{v} \in V_2$ is a polynomial of the form $a_0 + a_1 Z + \cdots + a_4 Z^4$ with $a_0, a_1, \ldots, a_4 \in \mathbb{C}$, which is already written as a linear combination of the vectors in the ordered basis $(1, Z, \ldots, Z^4)$. Hence $\phi_\gamma(a_0 + a_1 Z + \cdots + a_4 Z^4) = (a_0, a_1, \ldots, a_4)$, or in vector notation:

$$\phi_\gamma(a_0 + a_1 Z + \cdots + a_4 Z^4) = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_4 \end{bmatrix}.$$

Similarly,

$$\psi_\beta\left(\begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}\right) = b_0 + b_1 Z + b_2 Z^2 + b_3 Z^3.$$

We can describe the linear map L by figuring out what happens with the vectors in the chosen ordered basis β when L is applied. It is convenient to express the outcome as a linear combination of the vectors in the chosen ordered basis γ. We obtain:

$$L(1) = (i + 2Z) \cdot 1 = i + 2Z, \quad L(Z) = (i + 2Z) \cdot Z = iZ + 2Z^2,$$

$$L(Z^2) = (i + 2Z) \cdot Z^2 = iZ^2 + 2Z^3, \quad L(Z^3) = (i + 2Z) \cdot Z^3 = iZ^3 + 2Z^4.$$

Now let us compute the matrix $\mathbf{A}$ described in Lemma 10.28. We need to compute $\tilde{L}(\mathbf{e}_i)$ for $i = 1, \ldots, 4$, where $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4)$ is the standard ordered basis of $\mathbb{F}^4$ and $\tilde{L} = \phi_\gamma \circ L \circ \psi_\beta$:

$\mathbb{F}^4 \to \mathbb{F}^5$. Then we find:

$$\begin{aligned} \tilde{L}(\mathbf{e}_1) = (\phi_\gamma \circ L \circ \psi_\beta)(\mathbf{e}_1) &= (\phi_\gamma \circ L \circ \psi_\beta)\left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}\right) \\ &= (\phi_\gamma \circ L)(1) \\ &= \phi_\gamma(i + 2Z) \\ &= \begin{bmatrix} i \\ 2 \\ 0 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

and similarly

$$\tilde{L}(\mathbf{e}_2) = \begin{bmatrix} 0 \\ i \\ 2 \\ 0 \\ 0 \end{bmatrix}, \quad \tilde{L}(\mathbf{e}_3) = \begin{bmatrix} 0 \\ 0 \\ i \\ 2 \\ 0 \end{bmatrix}, \quad \text{and} \quad \tilde{L}(\mathbf{e}_4) = \begin{bmatrix} 0 \\ 0 \\ 0 \\ i \\ 2 \end{bmatrix}.$$

Using Lemma 10.28, we see that $\tilde{L} = \phi_\gamma \circ L \circ \psi_\beta = L_{\mathbf{A}}$, where

$$\mathbf{A} = \begin{bmatrix} i & 0 & 0 & 0 \\ 2 & i & 0 & 0 \\ 0 & 2 & i & 0 \\ 0 & 0 & 2 & i \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

|||| Definition 10.30

Let $\mathbb{F}$ be a field and $L : V_1 \to V_2$ a linear map between two finite dimensional vector spaces, say $\dim V_1 = m$ and $\dim V_2 = n$. Let β be an ordered basis of V_1 and γ one of V_2. Then we denote with ${}_\gamma[L]_\beta \in \mathbb{F}^{m\times n}$ the matrix described in Lemma 10.28 when applied to the linear map $\tilde{L} = \phi_\gamma \circ L \circ \psi_\beta : \mathbb{F}^m \to \mathbb{F}^n$. We say that the matrix ${}_\gamma[L]_\beta$ is the *matrix representation* of L with respect to the ordered bases β and γ. One also calls ${}_\gamma[L]_\beta$ the *mapping matrix* of L with respect to the ordered bases β and γ.

To avoid unnecessary computations, let us describe the mapping matrix ${}_\gamma[L]_\beta$ more directly:

|||| Lemma 10.31

Let $\mathbb{F}$ be a field and $L : V_1 \to V_2$ a linear map between two finite dimensional vector spaces, say $\dim V_1 = m$ and $\dim V_2 = n$. Let $\beta = (\mathbf{v}_1, \dots, \mathbf{v}_n)$ be an ordered basis of V_1 and γ one of V_2. Then the mapping matrix of L with respect to the ordered bases β and γ has $[L(\mathbf{v}_1)]_\gamma, \dots, [L(\mathbf{v}_n)]_\gamma$ as columns. That is to say:

$${}_\gamma[L]_\beta = [[L(\mathbf{v}_1)]_\gamma \cdots [L(\mathbf{v}_n)]_\gamma].$$

Proof. Combining Definition 10.30 and Lemma 10.28, we see that ${}_\gamma[L]_\beta$ has columns $\tilde{L}(\mathbf{e}_1), \dots, \tilde{L}(\mathbf{e}_n)$, where $\mathbf{e}_1, \dots, \mathbf{e}_n$ is the standard basis for $\mathbb{F}^n$ and $\tilde{L} = \phi_\gamma \circ L \circ \psi_\beta$. Now note that for all i between 1 and n, we have $\psi_\beta(\mathbf{e}_i) = \mathbf{v}_i$ using the definition of ψ_β given in Lemma 10.27. Further, $\phi_\gamma(\mathbf{w}) = [\mathbf{w}]_\gamma$ for all $\mathbf{w} \in V_2$ by definition of the map ϕ_β. Hence we see that for all i between 1 and n, we have

$$\tilde{L}(\mathbf{e}_i) = (\phi_\gamma \circ L \circ \psi_\beta)(\mathbf{e}_i) = (\phi_\gamma \circ L)(\mathbf{v}_i) = \phi_\gamma(L(\mathbf{v}_i)) = [L(\mathbf{v}_i)]_\gamma.$$

□

In Example 10.29, we already computed the matrix representation of a linear map (the matrix denoted by $\mathbf{A}$ in the example). Let us consider a few more examples.

|||| Example 10.32

This example is a continuation of Example 10.18. There, for $\alpha \in \mathbb{R}$, we considered the linear map $R_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $R_\alpha(v_1, v_2) = (\cos(\alpha) \cdot v_1 - \sin(\alpha) \cdot v_2, \sin(\alpha) \cdot v_1 + \cos(\alpha) \cdot v_2)$. Choosing the standard ordered basis $\beta = \gamma = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ for $\mathbb{R}^2$ both in case of the domain and the codomain of the linear map R_α, we obtain that

$${}_\gamma[R_\alpha]_\beta = \begin{bmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{bmatrix}. \tag{10-4}$$

The point of representing a linear map L with the matrix ${}_\gamma[L]_\beta$, is that the structure of the original linear map is "encoded" in this matrix. The following theorem makes this more precise.

Theorem 10.33

Let $\mathbb{F}$ be a field and V_1, V_2 and V_3 three finite dimensional vector spaces over $\mathbb{F}$. Further, let β, γ and δ be ordered bases of respectively V_1, V_2 and V_3. Then one has

1. $[L(\mathbf{v})]_\gamma = {}_\gamma[L]_\beta \cdot [\mathbf{v}]_\beta$ for any linear map $L : V_1 \to V_2$ and any $\mathbf{v} \in V_1$.
2. ${}_\delta[M \circ L]_\beta = {}_\delta[M]_\gamma \cdot {}_\gamma[L]_\beta$ for any linear maps $L : V_1 \to V_2$ and $M : V_2 \to V_3$.

Proof. We first prove the first item. Let us write $\mathbf{A} = {}_\gamma[L]_\beta$ for convenience. We have seen that $\phi_\gamma \circ L \circ \psi_\beta = L_\mathbf{A}$, using the notation from Lemma 10.27. Hence $\phi_\gamma \circ L = L_\mathbf{A} \circ (\psi_\beta)^{-1} = L_\mathbf{A} \circ \phi_\beta$. But then for any $\mathbf{v} \in V_1$, we obtain that $(\phi_\gamma \circ L)(\mathbf{v}) = (L_\mathbf{A} \circ \phi_\beta)(\mathbf{v})$. Simplifying the left-hand and right-hand side, we find that

$$(\phi_\gamma \circ L)(\mathbf{v}) = \phi_\gamma(L(\mathbf{v})) = [L(\mathbf{v})]_\gamma$$

and

$$(L_\mathbf{A} \circ \phi_\beta)(\mathbf{v}) = L_\mathbf{A}(\phi_\beta(\mathbf{v})) = L_\mathbf{A}([\mathbf{v}]_\beta) = {}_\gamma[L]_\beta \cdot [\mathbf{v}]_\beta.$$

Hence $[L(\mathbf{v})]_\gamma = {}_\gamma[L]_\beta \cdot [\mathbf{v}]_\beta$, which is what we needed to show.

The proof of the second item is somewhat similar. We write $\mathbf{A} = {}_\gamma[L]_\beta$ and $\mathbf{B} = {}_\delta[M]_\gamma$ for convenience. We have $L_\mathbf{A} = \phi_\gamma \circ L \circ \psi_\beta$ and $L_\mathbf{B} = \phi_\delta \circ M \circ \psi_\gamma$, which implies that $L_\mathbf{B} \circ L_\mathbf{A} = \phi_\delta \circ M \circ \psi_\gamma \circ \phi_\gamma \circ L \circ \psi_\beta$. Now using that ψ_γ and ϕ_γ are each other's inverses, see Lemma 10.27, we obtain that $L_\mathbf{B} \circ L_\mathbf{A} = \phi_\delta \circ M \circ L \circ \psi_\beta$. Since on the one hand $L_\mathbf{B} \circ L_\mathbf{A} = L_{\mathbf{B} \cdot \mathbf{A}}$ and on the other hand $\phi_\delta \circ M \circ L \circ \psi_\beta = L_\mathbf{C}$ with $\mathbf{C} = {}_\delta[M \circ L]_\beta$, this implies that ${}_\delta[M \circ L]_\beta = \mathbf{B} \cdot \mathbf{A} = {}_\delta[M]_\gamma \cdot {}_\gamma[L]_\beta$. This is what we wanted to show. □

The first item in this theorem simply tells us that the matrix ${}_\gamma[L]_\beta$ contains all information we need to know to describe the linear map L: computing $L(\mathbf{v})$ and then computing the coordinate vector of the outcome with respect to the ordered basis γ of V_2, is exactly the same as multiplying the matrix ${}_\gamma[L]_\beta$ with the coordinate vector of $\mathbf{v}$ with respect to the ordered basis β of V_1. The second item says that composition of linear maps behaves nice with respect to matrix representations. Let us look at an example of this.

|||| Example 10.34

We continue with Example 10.32. We have seen that if we choose β and γ to be the standard basis of $\mathbb{R}^2$, then ${}_\gamma[R_\alpha]_\beta$ is as in equation (10-4). Recall that the map $R_\alpha : \mathbb{R}^2 \to \mathbb{R}^2$ itself, geometrically can be described as a rotation over an angle α against the clock with midpoint in the origin. In particular $R_{\pi/2}$ corresponds with a rotation over $\pi/2$ radians (90 degrees). This means that $R_{\pi/2} \circ R_{\pi/2} = R_\pi$, a rotation over π radians (180 degrees). In particular, this means that $R_\pi(v_1, v_2) = (-v_1, -v_2)$. Let us check the second item in Theorem 10.33 for $V_1 = V_2 = V_3 = \mathbb{R}^2$, $\beta = \gamma = \delta = (\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix})$ and $L = M = R_{\pi/2}$. Then on the one hand we have

$$ {}_\delta[R_{\pi/2}]_\gamma = {}_\gamma[R_{\pi/2}]_\beta = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} $$

and therefore

$$ {}_\delta[R_{\pi/2}]_\gamma \cdot {}_\gamma[R_{\pi/2}]_\beta = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}. $$

On the other hand, using equation (10-4) for $\alpha = \pi$, we see that

$$ {}_\delta[R_{\pi/2} \circ R_{\pi/2}]_\beta = {}_\delta[R_\pi]_\beta = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}. $$

We conclude that indeed ${}_\delta[R_{\pi/2} \circ R_{\pi/2}]_\beta = {}_\delta[R_{\pi/2}]_\gamma \cdot {}_\gamma[R_{\pi/2}]_\beta$, just as it should be.

If one would do the same computation for $M = R_{\alpha_1}$ and $L = R_{\alpha_2}$ and use that $R_{\alpha_1} \circ R_{\alpha_2} = R_{\alpha_1+\alpha_2}$, one obtains that

$$ \begin{bmatrix} \cos(\alpha_1) & -\sin(\alpha_1) \\ \sin(\alpha_1) & \cos(\alpha_1) \end{bmatrix} \cdot \begin{bmatrix} \cos(\alpha_2) & -\sin(\alpha_2) \\ \sin(\alpha_2) & \cos(\alpha_2) \end{bmatrix} = \begin{bmatrix} \cos(\alpha_1+\alpha_2) & -\sin(\alpha_1+\alpha_2) \\ \sin(\alpha_1+\alpha_2) & \cos(\alpha_1+\alpha_2) \end{bmatrix}. $$

This identity actually implies the addition formulas for the cosine and the sine that we used in the proof of Lemma 3.18:

$$ \cos(\alpha_1 + \alpha_2) = \cos(\alpha_1)\cos(\alpha_2) - \sin(\alpha_1)\sin(\alpha_2) $$

and

$$ \sin(\alpha_1 + \alpha_2) = \sin(\alpha_1)\cos(\alpha_2) + \cos(\alpha_1)\sin(\alpha_2). $$

|||| Example 10.35

Let $\mathbb{F} = \mathbb{R}$ and $V_1 = \mathbb{R}^2$, $V_2 = \mathbb{R}^2$ and let

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}.$$

Denote by $L_{\mathbf{A}} : \mathbb{R}^2 \to \mathbb{R}^2$ the linear map defined by $\mathbf{v} \mapsto \mathbf{A} \cdot \mathbf{v}$. We have for example

$$L_{\mathbf{A}}\left(\begin{bmatrix} -1 \\ 2 \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

and

$$L_{\mathbf{A}}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

Question:

1. Choosing the standard ordered bases $\beta = \gamma = (\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix})$ for V_1 and V_2, compute ${}_{\gamma}[L_{\mathbf{A}}]_{\beta}$.
2. Choosing the ordered bases $\beta = \gamma = (\begin{bmatrix} -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix})$ for V_1 and V_2, compute ${}_{\gamma}[L_{\mathbf{A}}]_{\beta}$.

Answer:

1. Since γ is chosen to be the standard basis and

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = v_1 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + v_2 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

we see that $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_{\gamma} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ for all $v_1, v_2 \in \mathbb{R}$. Using Lemma 10.31, we see that

$${}_{\gamma}[L_{\mathbf{A}}]_{\beta} = [[L_{\mathbf{A}}(\mathbf{e}_1)]_{\gamma}\ [L_{\mathbf{A}}(\mathbf{e}_2)]_{\gamma}] = [L_{\mathbf{A}}(\mathbf{e}_1)\ L_{\mathbf{A}}(\mathbf{e}_2)] = [\mathbf{A} \cdot \mathbf{e}_1\ \mathbf{A} \cdot \mathbf{e}_2] = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix} = \mathbf{A}.$$

One can in fact see in a similar way that for any field $\mathbb{F}$ and any matrix $\mathbf{A} \in \mathbb{F}^{m \times n}$, one has ${}_{\gamma}[L_{\mathbf{A}}]_{\beta} = \mathbf{A}$ if β and γ are the standard ordered bases of $\mathbb{F}^m$ and $\mathbb{F}^n$.

2. Now we choose the ordered bases $\beta = \gamma = (\begin{bmatrix} -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix})$ for V_1 and V_2. Using Lemma 10.31, we see that

$$\begin{aligned} {}_\gamma[L_{\mathbf{A}}]_\beta &= \left[[L_{\mathbf{A}}\left(\begin{bmatrix} -1 \\ 2 \end{bmatrix}\right)]_\gamma \; [L_{\mathbf{A}}(\begin{bmatrix} 1 \\ 1 \end{bmatrix})]_\gamma \right] = \left[[\mathbf{A} \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix}]_\gamma \; [\mathbf{A} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}]_\gamma \right] \\ &= \left[\begin{bmatrix} 1 \\ -2 \end{bmatrix}_\gamma \begin{bmatrix} 2 \\ 2 \end{bmatrix}_\gamma \right]. \end{aligned}$$

Now in order to compute $[\mathbf{w}]_\gamma$ for $\mathbf{w} \in \mathbb{R}^2$, one needs in general to solve a linear system of equations. More precisely, let us write $\mathbf{w} = (w_1, w_2)$, then we want to find $c_1, c_2 \in \mathbb{R}^2$ such that

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} = c_1 \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} + c_2 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Therefore we need to solve the system of linear equations in the indeterminates c_1 and c_2 given by:

$$\begin{bmatrix} -1 & 1 \\ 2 & 1 \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}.$$

This can in principle be done using the theory of Chapter 6 or by multiplying the system on both sides of the equality sign with the matrix

$$\begin{bmatrix} -1 & 1 \\ 2 & 1 \end{bmatrix}^{-1} = \frac{1}{3} \begin{bmatrix} -1 & 1 \\ 2 & 1 \end{bmatrix}.$$

However, in this case we are lucky, since we can see directly that

$$\begin{bmatrix} 1 \\ -2 \end{bmatrix} = (-1) \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} \quad \text{and hence} \quad \begin{bmatrix} 1 \\ -2 \end{bmatrix}_\gamma = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 2 \\ 2 \end{bmatrix} = 2 \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \text{implying} \quad \begin{bmatrix} 2 \\ 2 \end{bmatrix}_\gamma = \begin{bmatrix} 0 \\ 2 \end{bmatrix}.$$

We conclude that

$${}_\gamma[L_{\mathbf{A}}]_\beta = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix}.$$

The result is a surprisingly nice looking matrix: it is a diagonal matrix (see Definition 8.5).

As a last item in this section, we consider matrices of the form ${}_\gamma[L]_\beta$ in case L is the identity map from a vector space V to itself: $\mathrm{id}_V : V \to V, \mathbf{v} \mapsto \mathbf{v}$. Here β and γ are two, possibly distinct, ordered bases of V. From the first part of Theorem 10.33, we see that

$$_\gamma[\mathrm{id}_V]_\beta \cdot [\mathbf{v}]_\beta = [\mathbf{v}]_\gamma \quad \text{for all } \mathbf{v} \in V. \tag{10-5}$$

In words equation (10-5) states that if one multiplies the matrix $_\gamma[\mathrm{id}_V]_\beta$ with the β-coordinate vector of a vector $\mathbf{v}$ in V, the outcome is the γ-coordinate vector of $\mathbf{v}$. For this reason, the matrix $_\gamma[\mathrm{id}_V]_\beta$ is called a *change of coordinates matrix* also known as a *change of basis matrix* .

|||| Example 10.36

Let $V = \{p(Z) \in \mathbb{C}[Z] \mid \deg p(Z) \leq 3\}$. Then $\beta = (1, Z, Z^2, Z^3)$ and $\gamma = (Z^3, Z^2, Z, 1)$ are two ordered bases of V. Question: Compute the corresponding change of coordinates matrix $_\gamma[\mathrm{id}_V]_\beta$.

Answer: Using Lemma 10.31, what we need to do is to compute $[1]_\gamma, [Z]_\gamma, [Z^2]_\gamma$ and $[Z^3]_\gamma$. Since the only difference between β and γ is the order of the basis vectors this is not so hard to do. For example

$$[1]_\gamma = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

since 1 is the fourth basis vector of γ. Proceeding similarly for the other basis vectors, one obtains the desired change of coordinates matrix:

$$_\gamma[\mathrm{id}_V]_\beta = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

We finish this section with a few facts on change of coordinates matrices that will come in handy later.

Lemma 10.37

Let $\mathbb{F}$ be a field, V a vector space over $\mathbb{F}$ of finite dimension n and β, γ and δ ordered bases of V. Then

1. ${}_\delta[\mathrm{id}_V]_\gamma \cdot {}_\gamma[\mathrm{id}_V]_\beta = {}_\delta[\mathrm{id}_V]_\beta$,
2. ${}_\beta[\mathrm{id}_V]_\beta = \mathbf{I}_n$, where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix, and
3. $({}_\gamma[\mathrm{id}_V]_\beta)^{-1} = {}_\beta[\mathrm{id}_V]_\gamma$.

Proof. The first item follows directly from the second item in Theorem 10.33. The second item is clear, since if the ordered basis β is not changed, the coordinates of a vector with respect to β do not change either. For the third item, note that according to the first and second part of the theorem, we have ${}_\gamma[\mathrm{id}_V]_\beta \cdot {}_\beta[\mathrm{id}_V]_\gamma = {}_\gamma[\mathrm{id}_V]_\gamma = \mathbf{I}_n$ and similarly ${}_\beta[\mathrm{id}_V]_\gamma \cdot {}_\gamma[\mathrm{id}_V]_\beta = {}_\beta[\mathrm{id}_V]_\gamma = \mathbf{I}_n$. Hence $({}_\gamma[\mathrm{id}_V]_\beta)^{-1} = {}_\beta[\mathrm{id}_V]_\gamma$. □

10.4 Usages of the matrix representation of a linear map

Now that we have the ability to represent linear maps between finite dimensional vector spaces with a matrix, we will use this to describe in more detail how to compute the kernel and image of a linear map. We start with a more general description of solutions to equations involving a linear map.

Theorem 10.38

Let $\mathbb{F}$ be a field and $L : V_1 \to V_2$ a linear map between vector spaces over $\mathbb{F}$. Further, let a vector $\mathbf{w} \in V_2$ be given and denote by $S = \{\mathbf{v} \in V_1 \mid L(\mathbf{v}) = \mathbf{w}\}$. Then exactly one of the following two possibilities occurs:

1. $S = \emptyset$. This is the case if and only if $\mathbf{w} \notin \mathrm{image} L$.
2. $S = \{\mathbf{v}_p + \mathbf{v} \mid \mathbf{v} \in \ker L\}$, where $\mathbf{v}_p \in V_1$ is a vector such that $L(\mathbf{v}_p) = \mathbf{w}$.

Proof. If $S = \emptyset$, then the equation $L(\mathbf{v}) = \mathbf{w}$ has no solutions. This is equivalent to the statement that no vector $\mathbf{v} \in V_1$ is mapped to $\mathbf{w}$. This in turn is the same as saying that $\mathbf{w}$ is not in the image of L.

If $S \neq \emptyset$, we may conclude that there exists a vector $\mathbf{v}_p \in V_1$ such that $L(\mathbf{v}_p) = \mathbf{w}$. If $\tilde{\mathbf{v}}$ is some vector, such that $L(\tilde{\mathbf{v}}) = \mathbf{w}$, then using linearity of L, we see that $L(\tilde{\mathbf{v}} - \mathbf{v}_p) = \mathbf{w} - \mathbf{w} = \mathbf{0}$. Hence $\tilde{\mathbf{v}} - \mathbf{v}_p \in \ker L$. Since $\tilde{\mathbf{v}} = \mathbf{v}_p + (\tilde{\mathbf{v}} - \mathbf{v}_p)$ and, as we already have seen $\tilde{\mathbf{v}} - \mathbf{v}_p \in \ker L$, this shows that $S \subseteq \{\mathbf{v}_p + \mathbf{v} \mid \mathbf{v} \in \ker L\}$. Conversely, if a vector is of the form $\mathbf{v}_p + \mathbf{v}$ for some $\mathbf{v} \in \ker L$, then $L(\mathbf{v}_p + \mathbf{v}) = L(\mathbf{v}_p) + L(\mathbf{v}) = \mathbf{w} + \mathbf{0} = \mathbf{w}$. This shows that $\{\mathbf{v}_p + \mathbf{v} \mid \mathbf{v} \in \ker L\} \subseteq S$. Combining both inclusions, we may conclude that $S = \{\mathbf{v}_p + \mathbf{v} \mid \mathbf{v} \in \ker L\}$. □

Hence the structure of the solution set of an equation of the form $L(\mathbf{v}) = \mathbf{w}$ is completely determined. The vector $\mathbf{v}_p$, if it exists, is called a *particular solution*. Notice how similar this is to Theorem 6.10. This is not a coincidence. After all, the solution set to a system of linear equations with augmented matrix $[\mathbf{A}|\mathbf{b}]$ is exactly the same as the solution set to the equation $L_{\mathbf{A}}(\mathbf{v}) = \mathbf{b}$. Moreover, $\ker L_{\mathbf{A}}$ is exactly the same as the solution set to the homogeneous system of linear equations with coefficient matrix $\mathbf{A}$. Hence, Theorem 6.10 is really just a special case of Theorem 10.38.

In case both V_1 and V_2 are finite dimensional vector spaces, we can computationally solve an equation of the form $L(\mathbf{v}) = \mathbf{w}$ by solving a suitable system of linear equations. We make this more precise in the following theorem.

Theorem 10.39

Let $\mathbb{F}$ be a field and $L : V_1 \to V_2$ a linear map between finite dimensional vector spaces over $\mathbb{F}$. Let $\beta = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ be an ordered basis of V_1 and $\gamma = (\mathbf{w}_1, \ldots, \mathbf{w}_m)$ be an ordered basis of V_2. Then

$$\{\mathbf{v} \in V_1 \mid L(\mathbf{v}) = \mathbf{w}\} = \{c_1 \cdot \mathbf{v}_1 + \cdots c_n \cdot \mathbf{v}_n \mid \mathbf{c} = (c_1, \ldots, c_n) \text{ satisfies } {}_\gamma[L]_\beta \cdot \mathbf{c} = [\mathbf{w}]_\gamma\}.$$

In particular

$$\ker L = \{c_1 \cdot \mathbf{v}_1 + \cdots c_n \cdot \mathbf{v}_n \mid (c_1, \ldots, c_n) \in \ker {}_\gamma[L]_\beta\}.$$

Proof. Applying Lemma 10.27 to the vector space V_1 and the given ordered basis β, we see that the linear maps $\phi_\beta : V_1 \to \mathbb{F}^n$ defined by $\mathbf{v} \mapsto [\mathbf{v}]_\beta$ and $\psi_\beta : \mathbb{F}^n \to V_1$ defined by $(c_1, \ldots, c_n) \mapsto c_1 \cdot \mathbf{v}_1 + \cdots + c_n \cdot \mathbf{v}_n$ are inverses of each other.

Assume that $L(\mathbf{v}) = \mathbf{w}$, then $[L(\mathbf{v})]_\gamma = [\mathbf{w}]_\gamma$, which using the first item in Theorem 10.33 implies that ${}_\gamma[L]_\beta[\mathbf{v}]_\beta = [\mathbf{w}]_\gamma$. If we write $(c_1, \dots, c_n) = [\mathbf{v}]_\beta$, then $\mathbf{v} = c_1 \cdot \mathbf{v}_1 + \cdots + c_n \cdot \mathbf{v}_n$. This shows that

$$\{\mathbf{v} \in V_1 \mid L(\mathbf{v}) = \mathbf{w}\} \subseteq \{c_1 \cdot \mathbf{v}_1 + \cdots c_n \cdot \mathbf{v}_n \mid \mathbf{c} = (c_1, \dots, c_n) \text{ satisfies } {}_\gamma[L]_\beta \cdot \mathbf{c} = [\mathbf{w}]_\gamma\}.$$

Conversely, assume that $\mathbf{c} = (c_1, \dots, c_n) \in \mathbb{F}^n$ satisfying ${}_\gamma[L]_\beta \cdot \mathbf{c} = [\mathbf{w}]_\gamma$ is given. The vector $\mathbf{v} = c_1 \cdot \mathbf{v}_1 + \cdots c_n \cdot \mathbf{v}_n$ has the property that $[\mathbf{v}]_\beta = \mathbf{c}$. Therefore ${}_\gamma[L]_\beta \cdot [\mathbf{v}]_\beta = [\mathbf{w}]_\gamma$. Using Theorem 10.33 again, we see that $[L(\mathbf{v})]_\gamma = [\mathbf{w}]_\gamma$. But then $L(\mathbf{v}) = \mathbf{w}$. This shows that

$$\{\mathbf{v} \in V_1 \mid L(\mathbf{v}) = \mathbf{w}\} \supseteq \{c_1 \cdot \mathbf{v}_1 + \cdots c_n \cdot \mathbf{v}_n \mid \mathbf{c} = (c_1, \dots, c_n) \text{ satisfies } {}_\gamma[L]_\beta \cdot \mathbf{c} = [\mathbf{w}]_\gamma\}.$$

Combining the above two inclusions, we see that the first part of the theorem follows.

Choosing $\mathbf{w} = \mathbf{0}$, the statement on $\ker L$ follows. □

The point of this theorem is that in order to compute all solutions to the equation $L(\mathbf{v}) = \mathbf{w}$, it is enough to compute all solutions to the equation ${}_\gamma[L]_\beta \cdot \mathbf{c} = [\mathbf{w}]_\gamma$. The latter equation is a system of linear equation with augmented matrix $[{}_\gamma[L]_\beta | [\mathbf{w}]_\gamma]$, which we can solve using the techniques from Chapter 6. The fact that the kernel of a linear map can be computed using the matrix representation of that map, has a nice consequence known as the *rank-nullity theorem for linear maps*.

|||| Corollary 10.40

Let $\mathbb{F}$ be a field and $L : V_1 \to V_2$ a linear map between finite dimensional vector spaces over $\mathbb{F}$. Then

$$\dim(\ker L) + \dim(\text{image} L) = \dim V_1.$$

Proof. If $\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ is basis of $\ker L$, then $\{[\mathbf{v}_1]_\beta, \dots, [\mathbf{v}_d]_\beta\}$ is a basis of $\ker {}_\gamma[L]_\beta$ using Theorem 9.19. Hence $\dim \ker L = \dim \ker {}_\gamma[L]_\beta$. Moreover, $\dim \text{image} L = \dim \text{image}_\gamma[L]_\beta$ using Corollary 9.40. Then the result follows from the rank-nullity theorem for matrices (see Theorem 10.10). □

|||| Example 10.41

This example is a variation of Example 10.26. In that example, we consider the map ev : $\mathbb{C}[Z] \to \mathbb{C}^2$ defined by $p(Z) \mapsto (p(0), p(1))$ and computed its kernel. Let $V_1 \subseteq \mathbb{C}[Z]$ be the subspace of $\mathbb{C}[Z]$ consisting of all polynomials of degree at most three. Then $\beta = (1, Z, Z^2, Z^3)$ is an ordered basis of V_1. For $\mathbb{C}^2$, we choose the standard ordered basis $(\mathbf{e}_1, \mathbf{e}_2)$. Now let us consider the linear map $L : V_1 \to \mathbb{C}^2$ defined by $L(p(Z)) = (p(0), p(1))$. In other words: we restrict the domain of ev to V_1, but otherwise do not change anything.

Questions: What is the kernel of the linear map L described above? What are all solutions to the equation $L(p(Z)) = (5, 8)$?

Answer:

We can compute ker L in several ways, but let use Theorem 10.39. To compute ker L, we first compute the kernel of ${}_\gamma[L]_\beta$. We have $L(1) = (1,1) = 1 \cdot \mathbf{e}_1 + 1 \cdot \mathbf{e}_2$, $L(Z) = (0,1) = 1 \cdot \mathbf{e}_2$, $L(Z^2) = (0,1) = 1 \cdot \mathbf{e}_2$, and $L(Z^3) = (0,1) = 1 \cdot \mathbf{e}_2$. Hence

$${}_\gamma[L]_\beta = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

Computing the reduced row echelon form of this matrix in this case just amounts to subtracting the first row from the second row. One finds the matrix:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{bmatrix}.$$

Hence using Theorem 6.29 and Corollary 9.39, we find that a basis of $\ker {}_\gamma[L]_\beta$ is given by the set

$$\left\{ \begin{bmatrix} 0 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

and hence a basis of ker L is given by the set $\{-Z + Z^2, -Z + Z^3.\}$ Hence

$$\ker L = \{t_1 \cdot (-Z + Z^2) + t_2 \cdot (-Z + Z^3) \mid t_1, t_2 \in \mathbb{C}\}.$$

To solve the final question about the solutions to the equation $L(p(Z)) = (5, 8)$, we use Theorem 10.38. All we still need to do is to compute a particular solution. We could in principle again transform the equation into a system of linear equations. Doing this would give rise to a system of inhomogeneous linear equations with augmented matrix

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 5 \\ 1 & 1 & 1 & 1 & 8 \end{array}\right],$$

which has reduced row echelon form

$$\left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 5 \\ 0 & 1 & 1 & 1 & 3 \end{array}\right].$$

A particular solution (c_1, c_2, c_3, c_4) should satisfy $c_1 = 5$ and $c_2 + c_3 + c_4 = 3$. Therefore $(5, 3, 0, 0)$ is a particular solution, which corresponds to the polynomial $f(Z) = 5 + 3Z$. Using Theorem 10.38, we conclude that all solutions to the equation $L(p(Z)) = (5, 8)$ form the set

$$\{5 + 3Z + t_1 \cdot (-Z + Z^2) + t_2 \cdot (-Z + Z^3) \mid t_1, t_2 \in \mathbb{C}\}.$$

Just as an aside: another way to compute $\ker L$ is to use that we already have computed the kernel of $\mathrm{ev} : \mathbb{C}[Z] \to \mathbb{C}^2$ in Example 10.26. Then

$$\begin{array}{rcl} \ker L & = & \ker \mathrm{ev} \cap V_1 \\ & = & \{Z \cdot (Z-1) \cdot r(Z) \mid s(Z) \in \mathbb{C}[Z]\} \cap V_1 \\ & = & \{Z \cdot (Z-1) \cdot r(Z) \mid s(Z) \in \mathbb{C}[Z], \deg s(Z) \leq 1\}. \end{array}$$

Here we used that $Z \cdot (Z-1) \cdot s(Z) \in V_1$ precisely if $\deg(Z \cdot (Z-1) \cdot s(Z)) \leq 3$. Since $\deg(Z \cdot (Z-1) \cdot s(Z)) = 1 + 1 + \deg s(Z)$, we see that $Z \cdot (Z-1) \cdot s(Z) \in V_1$ precisely if $\deg s(Z) \leq 1$. It is left to the reader to check that this computation of $\ker L$ gives exactly the same result as before.