3. LINEAR MAPPINGS AND DIAGONALIZATION

In this chapter, we first consider functions that transform elements from one vector space into another one. These functions comply with the fundamental properties of linearity, already considered above, and are therefore called *linear mappings*. In many instances, we are able to formulate matrices to represent linear mappings. Afterwards, we consider a concept that has many applications in the sciences and engineering: the *eigenvector*. Such a vector is invariant under a linear mapping (except for a constant factor called *eigenvalue*) and for certain linear mappings (and associated matrices) it will be possible to find a basis of the vector space such that the associated matrix is diagonal, known as the *diagonalization* of the matrix.

3.1 Linear mappings and matrices

3.1.1 Introduction and basic properties

Example 3.1

Let us start with an example: Let $\mathbf{p} = \left[x_1, y_1, z_1\right]$ be a point in the vector space $\mathbb{R}^3$ (for simplicity, we use row vectors here). Then, there exists a unique projection to the vector space $\mathbb{R}^2$ spanned by x and y : such a point is simply $\mathbf{q} = \left[x_1, y_1\right]$ (in a three-dimensional Cartesian coordinate system, $\mathbf{q}$ lies in the plane $\left[x, y, 0\right]$). We regard $\left[x_1, y_1, z_1\right] \to \left[x_1, y_1\right]$ as a *mapping* T from $\mathbb{R}^3$ to $\mathbb{R}^2$ and write

$$T : \mathbb{R}^3 \to \mathbb{R}^2 \; T\left(\mathbf{p}\right) = \mathbf{q}.$$

This mapping conserves *linearity*:

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \\ z_1 + z_2 \end{bmatrix} \to \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

$$\alpha \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} \alpha x_2 \\ \alpha y_2 \\ \alpha z_2 \end{bmatrix} \to \begin{bmatrix} \alpha x_1 \\ \alpha y_1 \end{bmatrix} = \alpha \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \; \alpha \in \mathbb{R}$$

We see that *linear mappings* are functions defined on vector spaces that preserve linear combinations. Other names (mostly used synonymously, but strongly depending on the context) are linear transformations, linear operators, or linear maps.

Definition 3.1

Given two vector spaces V and W, we say that $T: V \to W$ is a linear mapping if $\forall \mathbf{u},\mathbf{v} \in V$ and $\forall \alpha \in \mathbb{R}$ it verifies:

(a) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$ (additivity).

(b) $T(\alpha \mathbf{u}) = \alpha T(\mathbf{u})$ (homogeneity).

These two properties can be combined in a single statement:

$$T(\alpha \mathbf{u} + \beta \mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v}) \ \forall \mathbf{u},\mathbf{v} \in V \text{ and } \forall \alpha, \beta \in \mathbb{R}$$

Example 3.2

The mapping $D: \mathbf{P} \to \mathbf{P}$ defined by $D(p(x)) = p'(x)$ is linear.

Proof. We show that the differentiation of polynomials is a linear mapping. Let f and g represent polynomials, written in the canonical basis of polynomials $\mathbf{P}$. Then, we use knowledge of calculus: $D(f+g) = D(f) + D(g)$ and $D(\alpha f) = \alpha D(f)$ and differentiation of polynomials (as actually of any differentiable function) is linear. Note that we need to have a well-defined vector space. ■

Example 3.3

The mapping $T: \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} y \\ x \end{bmatrix}$$

is linear.

Proof. For all vectors $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^2$ and all $\alpha \in \mathbb{R}$, we have:

$$T\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) = T\left(\begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}\right) = \begin{bmatrix} y_1 + y_2 \\ x_1 + x_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ x_1 \end{bmatrix} + \begin{bmatrix} y_2 \\ x_2 \end{bmatrix} = T\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}\right) + T\left(\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right)$$

and

$$T\left(\alpha \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}\right) = T\left(\begin{bmatrix} \alpha x_1 \\ \alpha y_1 \end{bmatrix}\right) = \begin{bmatrix} \alpha y_1 \\ \alpha x_1 \end{bmatrix} = \alpha \begin{bmatrix} y_1 \\ x_1 \end{bmatrix} = \alpha T\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}\right)$$

■

This simple example allows us to draw a link to the first chapter. The linear mapping of this example is interchanging x and y components of a vector. This can also be achieved by multiplying a matrix to the vector:

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} y \\ x \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

Hence, a linear mapping can be represented (for some relevant cases) by

$$T(\mathbf{u}) = A\mathbf{u},$$

where A is a suitable matrix. We will come back to the matrix associated to a linear mapping below.

Example 3.4

The mapping $T:\mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} y \\ x^2 \end{bmatrix}$$

is not linear.

Proof. It is sufficient to show that the additivity property is not fulfilled:

$$T\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) = T\left(\begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}\right) = \begin{bmatrix} y_1 + y_2 \\ (x_1 + x_2)^2 \end{bmatrix},$$

while

$$T\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}\right) + T\left(\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) = \begin{bmatrix} y_1 \\ x_1^2 \end{bmatrix} + \begin{bmatrix} y_2 \\ x_2^2 \end{bmatrix} = \begin{bmatrix} y_1 + y_2 \\ x_1^2 + x_2^2 \end{bmatrix}$$

■

Theorem 3.1

Let $T:V \to W$ be a linear mapping. Then:

(a) $T(0) = 0$.

(b) $T(-\mathbf{u}) = -T(\mathbf{u})$.

(c) $T(a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + \cdots + a_n\mathbf{u}_n) = a_1T(\mathbf{u}_1) + a_2T(\mathbf{u}_2) + \cdots + a_nT(\mathbf{u}_n)$.

Observe that in part (a), the nullvector on the left hand side is the nullvector from V, on the left hand side there is the nullvector from W. In particular, part (a) means that for a linear mapping, the nullvector maps to the nullvector. In part (c), the number of vectors n is not specified.

Example 3.5

The mapping $T:\mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+1 \\ y \end{bmatrix}$$

is not linear since

$$T\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

3.1.2 Kernel and image of a linear mapping

Definition 3.2

Let $T:V \to W$ be a linear mapping. We define the kernel and image of T, respectively, as follows:

$$\begin{aligned} \mathrm{Ker}T &= \left\{\mathbf{x} \in V : T(\mathbf{x}) = 0\right\}, \\ \mathrm{Im}T &= \left\{T(\mathbf{x}) \in W : \mathbf{x} \in V\right\}. \end{aligned}$$

Observe that $\mathrm{Ker}T$ is a subspace of V and $\mathrm{Im}T$ is a subspace of W. Once we establish the associated matrix for a linear mapping, we will see how the kernel is analogous to the nullspace and the image to the column space of the matrix.

Theorem 3.2

Let $T:V \to W$ be a linear mapping and let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be a system of generators of V. Then, $\{T(\mathbf{u}_1), T(\mathbf{u}_2), \ldots, T(\mathbf{u}_n)\}$ is a system of generators of $\mathrm{Im}T$.

Example 3.6

Find the kernel and image of the linear mapping $T:\mathbb{R}^3 \to \mathbb{R}^3$ defined by

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+z \\ y \\ x+2y+z \end{bmatrix}.$$

For the vector space of this example, we know a very convenient system of generators: the canonical basis of $\mathbb{R}^3$:

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

We apply the linear mapping to these vectors and obtain three vectors

$$\left\{T\left(\begin{bmatrix}1\\0\\0\end{bmatrix}\right),T\left(\begin{bmatrix}0\\1\\0\end{bmatrix}\right),T\left(\begin{bmatrix}0\\0\\1\end{bmatrix}\right)\right\}=\left\{\begin{bmatrix}1\\0\\1\end{bmatrix},\begin{bmatrix}0\\1\\2\end{bmatrix},\begin{bmatrix}1\\0\\1\end{bmatrix}\right\}.$$

Two of these vectors are linearly dependent, and hence we conclude

$$\operatorname{Im}T=\operatorname{Span}\left\{\begin{bmatrix}1\\0\\1\end{bmatrix},\begin{bmatrix}0\\1\\2\end{bmatrix}\right\}$$

The kernel is formed by the vectors that map to the nullvector.

$$\begin{aligned}x+z&=0,\\ y&=0,\\ x+2y+z&=0,\end{aligned}$$

From that set of equations we obtain

$$\begin{aligned}x&=-z,\\ y&=0,\end{aligned}$$

with z being arbitrary, and therefore

$$\operatorname{Ker}T=\operatorname{Span}\left\{\begin{bmatrix}1\\0\\-1\end{bmatrix}\right\}$$

Before we proceed, we have to review some general classification of functions:

Definition 3.3

Let $f : A \to B$ be a function. Then,

(a) f is injective if and only if $x \neq y$ implies $f(x) \neq f(y) \forall x, y \in A$.

(b) f is surjective if and only if $\forall b \in B, \exists a \in A$ such that $f(a) = b$.

(c) f is bijective if and only if f is injective and f is surjective.

With this classification it is now possible to confirm the following properties for injective and surjective (and hence, bijective) linear mappings:

Theorem 3.3

Let $T : V \to W$ be a linear mapping. Then:

(a) T is injective if and only if $\text{Ker}T = \{0\}$.

(b) T is surjective if and only if $\text{Im}T = W$.

(c) T is bijective if and only if $\text{Ker}T = \{0\}$ and $\text{Im}T = W$.

There are alternative characterizations for injective, surjective and bijective linear mappings, provided by the following theorem:

Theorem 3.4

Let $T : V \to W$ be a linear mapping. Then:

(a) T is injective if and only if for each independent set $\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$, the set $\{T(\mathbf{u}_1), T(\mathbf{u}_2), \dots, T(\mathbf{u}_n)\}$ is linearly independent.

(b) T is surjective if and only for each system of generators of $V, \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$, the set $\{T(\mathbf{u}_1), T(\mathbf{u}_2), \dots, T(\mathbf{u}_n)\}$ is a system of generators of W.

(c) T is bijective if and only if for each basis of $V, \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n\}$, the set $\{T(\mathbf{u}_1), T(\mathbf{u}_2), \dots, T(\mathbf{u}_n)\}$ is a basis for W.

Notice that n for the three cases is not necessarily the same number, as the number of independent vectors can be *smaller* than the dimension, and the number of generatoring vectors can be *larger* than the dimension, respectively.

Let us now move on to generate linear mappings from basic operations.

Definition 3.4

Let $f, g: V \to W$ be two linear mappings and $\lambda \in \mathbb{R}$. Then, we can define the following operations:

(a) $f + g: V \to W, (f+g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$

(b) $\lambda f: V \to W, (\lambda f)(\mathbf{x}) = \lambda f(\mathbf{x})$

Theorem 3.5

With the operations of linear mappings, the set of linear mappings between two vector spaces V and W is itself a vector space.

As for general functions, we can also define the composition of linear mappings and find that its result is also a linear mapping, by the next theorem.

Theorem 3.6

Let $f: V \to W$ and $g: W \to U$ be two linear mappings. Then, their composition $g \circ f: V \to U$, defined by $(g \circ f)(\mathbf{x}) = g(f(\mathbf{x}))$ is also a linear mapping.

We also would like to define an inverse linear mapping and, for that, we use the general result that if a function is bijective, it has a unique inverse function. In particular, if $f: V \to W$ is a bijective linear mapping, then there is a unique function $f^{-1}: W \to V$ such that:

$$f^{-1}(f(\mathbf{v})) = \mathbf{v} \text{ for all } \mathbf{v} \in V,$$
$$f(f^{-1}(\mathbf{w})) = \mathbf{w} \text{ for all } \mathbf{w} \in W.$$

It turns out that the inverse of a linear mapping is also a linear mapping.

Theorem 3.7

For each bijective linear mapping Let $f: V \to W$, its inverse mapping $f^{-1}: W \to V$ is also a linear mapping, i.e.,

$$f^{-1}(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha f^{-1}(\mathbf{x}) + \beta f^{-1}(\mathbf{y}) \ \forall \mathbf{x}, \mathbf{y} \in W \text{ and } \forall \alpha, \beta \in \mathbb{R}.$$

3.1.3 Associated matrices

Here, we will see that linear mappings between finite-dimensional vector spaces have a matrix representation. In fact, given a linear mapping $T: V \to W$, this matrix representation or ***associated matrix***, denoted by $M_{\mathcal{C}}^{\mathcal{B}}(T)$ (or just $M_{\mathcal{C}}^{\mathcal{B}}$ for the sake of simplicity) has the useful property that

$$[T(x)]_{\mathcal{C}} = M_{\mathcal{C}}^{\mathcal{B}}[x]_{\mathcal{B}}, \tag{3.1}$$

where $\mathcal{B}$ and $\mathcal{C}$ are bases for V and W, respectively. In other words, for each $\mathbf{x} \in V$, the matrix $M_{\mathcal{C}}^{\mathcal{B}}$ should transform the coordinates of $\mathbf{x}$ into the coordinates of its image under T. In this way, the effect that $M_{\mathcal{C}}^{\mathcal{B}}$ has on $[\mathbf{x}]_{\mathcal{B}}$ is analogous to the effect that T has on $\mathbf{x}$.

An important observation is that $M_{\mathcal{C}}^{\mathcal{B}}$ depends on the bases that are chosen for the vector spaces and, of course, the mapping itself. We now discuss how to construct the associated matrix of a linear mapping relative to the bases $\mathcal{B}$ and $\mathcal{C}$.

Construction of the matrix associated to a linear mapping

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis for V. Then, for each $\mathbf{x} \in V$, we can write

$$\mathbf{x} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_n\mathbf{b}_n,$$

which means that its vector of coordinates relative to $\mathcal{B}$ is

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

If we apply the linear mapping T to both sides of $\mathbf{x} = x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_n\mathbf{b}_n$, then we obtain

$$T(\mathbf{x}) = T(x_1\mathbf{b}_1 + x_2\mathbf{b}_2 + \cdots + x_n\mathbf{b}_n),$$

and, due to the linearity of T, we have

$$T(\mathbf{x}) = x_1T(\mathbf{b}_1) + x_2T(\mathbf{b}_2) + \cdots + x_nT(\mathbf{b}_n).$$

Since taking coordinates preserves linear combinations, we deduce that the coordinates of $T(\mathbf{x})$ relative to the basis $\mathcal{C}$ are given by

$$
\begin{aligned}
[T(\mathbf{x})]_{\mathcal{C}} &= \left[x_1 T(\mathbf{b}_1) + x_2 T(\mathbf{b}_2) + \cdots + x_n T(\mathbf{b}_n)\right]_{\mathcal{C}} \\
&= x_1\left[T(\mathbf{b}_1)\right]_{\mathcal{C}} + x_2\left[T(\mathbf{b}_2)\right]_{\mathcal{C}} + \cdots + x_2\left[T(\mathbf{b}_n)\right]_{\mathcal{C}}.
\end{aligned}
$$

or, in matrix notation,

$$
[T(x)]_{\mathcal{C}} = \left[\left[T(\mathbf{b}_1)\right]_{\mathcal{C}} \quad \left[T(\mathbf{b}_2)\right]_{\mathcal{C}} \quad \cdots \quad \left[T(\mathbf{b}_n)\right]_{\mathcal{C}}\right][x]_{\mathcal{B}}, \tag{3.2}
$$

where each $\left[T(\mathbf{b}_i)\right]_{\mathcal{C}}$ is an $m \times 1$ (column) vector.

Comparing (3.1) and (3.2), we see that the $m \times n$ matrix

$$
M_{\mathcal{C}}^{\mathcal{B}} = \left[\left[T(\mathbf{b}_1)\right]_{\mathcal{C}} \quad \left[T(\mathbf{b}_2)\right]_{\mathcal{C}} \quad \cdots \quad \left[T(\mathbf{b}_n)\right]_{\mathcal{C}}\right], \tag{3.3}
$$

is precisely the associated matrix for T that we were looking for.

Example 3.7

Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear mapping such that

$$
T\left(\begin{bmatrix}1\\0\end{bmatrix}\right) = \begin{bmatrix}1\\2\\3\end{bmatrix},\ T\left(\begin{bmatrix}0\\1\end{bmatrix}\right) = \begin{bmatrix}2\\1\\0\end{bmatrix},
$$

and let $\mathcal{E}_2$ and $\mathcal{E}_3$ be the canonical bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. Find $M_{\mathcal{E}_3}^{\mathcal{E}_2}(T) \equiv M_{\mathcal{E}_3}^{\mathcal{E}_2}$. Then, compute $T(\mathbf{x})$ for

$$
\mathbf{x} = \begin{bmatrix}1\\1\end{bmatrix}.
$$

According (3.3), the columns of $M_{\mathcal{E}_3}^{\mathcal{E}_2}$ are

$$
\left[T\left(\begin{bmatrix}1\\0\end{bmatrix}\right)\right]_{\mathcal{E}_3} = \begin{bmatrix}1\\2\\3\end{bmatrix},\ \text{and}\ \left[T\left(\begin{bmatrix}0\\1\end{bmatrix}\right)\right]_{\mathcal{E}_3} = \begin{bmatrix}2\\1\\0\end{bmatrix}.
$$

Therefore,

$$
M_{\mathcal{E}_3}^{\mathcal{E}_2} = \left[\left[T\left(\begin{bmatrix}1\\0\end{bmatrix}\right)\right]_{\mathcal{E}_3} \left[T\left(\begin{bmatrix}0\\1\end{bmatrix}\right)\right]_{\mathcal{E}_3}\right] = \begin{bmatrix}1 & 2\\2 & 1\\3 & 0\end{bmatrix}
$$

In this case, $T(\mathbf{x})$ can be computed directly (even if a definition for T is not explicitly given) using $M_{\mathcal{E}_3}^{\mathcal{E}_2}$, because $T(\mathbf{x}) = [T(\mathbf{x})]_{\mathcal{E}_3}$ and $\mathbf{x} = [\mathbf{x}]_{\mathcal{E}_2}$.

Therefore,

$$T(\mathbf{x})]_{\mathcal{E}_3} = M_{\mathcal{E}_3}^{\mathcal{E}_2}[\mathbf{x}]_{\mathcal{E}_2},$$

becomes

$$T\left(\begin{bmatrix}1\\1\end{bmatrix}\right) = \begin{bmatrix}1 & 2\\2 & 1\\3 & 0\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}3\\3\\3\end{bmatrix}.$$

Example 3.8

Let $D:\mathbf{P}_3 \to \mathbf{P}_3$ be the linear mapping defined as

$$D(p(x)) = p'(x).$$

Find $M_{\mathcal{S}_3}^{\mathcal{S}_3}(D) \equiv M_{\mathcal{S}_3}^{\mathcal{S}_3}$, where $\mathcal{S}_3 = \{1, x, x^2, x^3\}$ is the standard basis for $\mathbf{P}_3$.
Note that, in this example, D is a mapping from $\mathbf{P}_3$ to itself. Moreover, we are asked to use the same basis to represent the elements in $\mathbf{P}_3$ as well as their images under D. By (3.3), we have

$$M_{\mathcal{S}_3}^{\mathcal{S}_3} = \left[[D(1)]_{\mathcal{S}_3} \quad [D(x)]_{\mathcal{S}_3} \quad \left[D(x^2)\right]_{\mathcal{S}_3} \quad \left[D(x^3)\right]_{\mathcal{S}_3}\right] = \left[[0]_{\mathcal{S}_3} \quad [1]_{\mathcal{S}_3} \quad [2x]_{\mathcal{S}_3} \left[3x^2\right]_{\mathcal{S}_3}\right].$$

The explicit expression of $M_{\mathcal{S}_3}^{\mathcal{S}_3}$ is obtained by writing out the derivatives in the last equality as linear combinations of $\mathcal{S}_3$:

$$\begin{aligned}
0 &= 0\cdot 1 + 0\cdot x + 0\cdot x^2 + 0\cdot x^3,\\
1 &= 1\cdot 1 + 0\cdot x + 0\cdot x^2 + 0\cdot x^3,\\
2x &= 0\cdot 1 + 2\cdot x + 0\cdot x^2 + 0\cdot x^3,\\
3x^2 &= 0\cdot 1 + 0\cdot x + 3\cdot x^2 + 0\cdot x^3.
\end{aligned}$$

Therefore,

$$M_{\mathcal{S}_3}^{\mathcal{S}_3} = \begin{bmatrix}0 & 1 & 0 & 0\\0 & 0 & 2 & 0\\0 & 0 & 0 & 3\\0 & 0 & 0 & 0\end{bmatrix}.$$

Let us mention an interesting property of this linear mapping. Observe that the first column of $M_{\mathcal{S}_3}^{\mathcal{S}_3}$ consists entirely of zeros and, obviously, it is not a pivot column. Therefore, the equation

$$M_{\mathcal{S}_3}^{\mathcal{S}_3}\mathbf{x} = 0,$$

has a non trivial solution (the first column corresponds to a free variable). This means that there is at least one non zero polynomial $p(x)$ such that $D(p(x)) = 0$. In other words, $\mathrm{Ker}D \neq \{0\}$. It follows from Theorem 3.3 that D is not injective. Indeed, it is well known that for any constant $k, D(k) = 0$. Consequently, any polynomial of the form $f(x) = p(x) + k, k \in \mathbb{R}$, satisfies the equation

$$D(f(x)) = p'(x).$$

Example 3.9

Let $T : \mathrm{P}_3 \to \mathrm{P}_4$ be the linear mapping defined as

$$T(p(x)) = xp(x).$$

Let us compute $M_{\mathcal{S}_4}^{\mathcal{B}}(T) \equiv M_{\mathcal{S}_4}^{\mathcal{B}}$ where $\mathcal{B} = \{1, 1+x, x+x^2, x^2+x^3\}$ and $\mathcal{S}_4 = \{1, x, x^2, x^3, x^4\}$.

Since we are using the standard basis for $\mathbf{P}_4$, it is easy to write the explicit expression for $M_{\mathcal{S}_4}^{\mathcal{B}}$. By (3.3), we have

$$\begin{aligned} M_{\mathcal{S}_4}^{\mathcal{B}} &= \left[[T(1)]_{\mathcal{S}_4} \quad [T(1+x)]_{\mathcal{S}_4} \quad \left[T\left(x+x^2\right)\right]_{\mathcal{S}_4} \; \left[T\left(x^2+x^3\right)\right]_{\mathcal{S}_4} \right] \\ &= \left[[x]_{\mathcal{S}_4} \quad \left[x+x^2\right]_{\mathcal{S}_4} \quad \left[x^2+x^3\right]_{\mathcal{S}_4} \; \left[x^3+x^4\right]_{\mathcal{S}_4} \right] \\ &= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \end{aligned}$$

The matrices of injective, surjective, and bijective linear mappings

It is possible to characterize the injective and surjective character of a linear mapping $T : V \to W$ in terms of the pivots of its associated matrix independently of the bases we choose for the vector spaces involved.

In the discussion below, $\mathcal{B}=\{\mathbf{b}_1,\mathbf{b}_2,\ldots,\mathbf{b}_n\}$ will denote a basis for V and $\mathcal{C}$ will denote a basis for W.

Injectivity. By Theorem 3.4, we know that T is an injective linear mapping if and only if $\{T(\mathbf{b}_1),T(\mathbf{b}_2),\ldots,T(\mathbf{b}_n)\}$ is a linearly independent set of vectors. Moreover recall that their coordinate vectors relative to any basis constitute a linearly independent set of column vectors. Hence,

$$\left\{\left[T(\mathbf{b}_1)\right]_{\mathcal{C}},\left[T(\mathbf{b}_2)\right]_{\mathcal{C}},\ldots,\left[T(\mathbf{b}_n)\right]_{\mathcal{C}}\right\}$$

is a linearly independet set of vectors in $\mathbb{R}^m$ with $m=\dim W$. These vectors are the columns of the associated matrix $M_{\mathcal{C}}^{\mathcal{B}}$. Therefore, the injectivity of T passes down to its associated matrix $M_{\mathcal{C}}^{\mathcal{B}}$ as follows: T is injective if and only if all the columns of $M_{\mathcal{C}}^{\mathcal{B}}$ are linearly independent, or, equivalently, all the columns of $M_{\mathcal{C}}^{\mathcal{B}}$ are pivot columns. Consequently, the equation

$$[\mathbf{y}]_{\mathcal{C}}=M_{\mathcal{C}}^{\mathcal{B}}[\mathbf{x}]_{\mathcal{B}},\ \mathbf{y}\in \operatorname{Im}T,$$

has a unique solution. If we compute the rank of $M_{\mathcal{C}}^{\mathcal{B}}$ associated with an injective linear mapping T, then we find that

$$\operatorname{rank}M_{\mathcal{C}}^{\mathcal{B}}=\#\text{ pivots of }M_{\mathcal{C}}^{\mathcal{B}}=\#\text{ columns of }M_{\mathcal{C}}^{\mathcal{B}}=\dim V.$$

This has the following important consequence:

If there is an injective linear mapping $T:V\to W$, then

- $n=\dim V\le m$ because any linearly independent set in W has at most m elements.
- #columns of $M_{\mathcal{C}}^{\mathcal{B}}\le\#$ rows of $M_{\mathcal{C}}^{\mathcal{B}}$.

Surjectivity. Again, by Theorem 3.4, we know that T is surjective if and only if the set $\{T(\mathbf{b}_1),T(\mathbf{b}_2),\ldots,T(\mathbf{b}_n)\}$ is a generating system of W. Since taking coordinates preserve linear relations, this means that the set

$$\left\{\left[T(\mathbf{b}_1)\right]c,\left[T(\mathbf{b}_2)\right]c,\ldots,\left[T(\mathbf{b}_n)\right]c\right\}$$

is a generating system for $\mathbb{R}^m$ with $m=\dim W$. These vectors are the columns of the associated matrix $M_{\mathcal{C}}^{\mathcal{B}}$, and therefore surjectivity translates into the statement that all rows of $M_{\mathcal{C}}^{\mathcal{B}}$ have a pivot (in order to avoid degenerate rows), which is the same as confirming that the equation

$$[\mathrm{y}]_C=M_{\mathcal{C}}^{\mathcal{B}}\,[\mathrm{x}]_{B,}\quad \mathrm{y}\in\operatorname{Im}T$$

has a solution for all $\mathbf{b} \in W$. Therefore, computing the rank of $M_{\mathcal{C}}^{\mathcal{B}}$, we obtain

$$\operatorname{rank} M_{\mathcal{C}}^{\mathcal{B}} = \# \text{ pivots of } M_{\mathcal{C}}^{\mathcal{B}} = \# \text{ rows of } M_{\mathcal{C}}^{\mathcal{B}} = \dim W .$$

This has the following important consequence:

If there is a surjective linear mapping $T : V \to W$, then

- $n = \dim V \geqslant m$ because any generatig set for W has at least m elements.
- #columns of $M_{\mathcal{C}}^{\mathcal{B}} \geq$ #rows of $M_{\mathcal{C}}^{\mathcal{B}}$.

Bijectivity. From the previous discussion, we deduce that T is bijective (T is both injective and surjective) if and only if the set of vectors $\{T(\mathbf{b}_1), T(\mathbf{b}_2), \ldots, T(\mathbf{b}_n)\}$ forms a basis for W . Putting what we know about injectivity and surjectivity together

$$\operatorname{rank} M_{\mathcal{C}}^{\mathcal{B}} = \# \text{ pivots of } M_{\mathcal{C}}^{\mathcal{B}} = \# \text{ columns of } M_{\mathcal{C}}^{\mathcal{B}} = \# \text{ rows of } M_{\mathcal{C}}^{\mathcal{B}}$$

This has the following significant implication:

If there is a bijective linear mapping $T : V \to W$, then

- $\dim V = \dim W$.
- $M_{\mathcal{C}}^{\mathcal{B}}$ is an invertible square matrix.

Let us consider a few examples.

Example 3.10

Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear mapping such that

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

We already found the associated matrix with respect to the canonical bases:

$$M_{\mathcal{E}_3}^{\mathcal{E}_2} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 0 \end{bmatrix}.$$

Its row reduced echelon form is given by

$$M_{\mathcal{E}_3}^{\mathcal{E}_2} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

We see that all columns of $M_{\mathcal{E}_3}^{\mathcal{E}_2}$ are pivot columns and therefore T is injective. We can also see that $\dim\mathbb{R}^2 < \dim\mathbb{R}^3$, and hence T cannot be surjective (and, thus, not bijective).

Example 3.11

Let $D : \mathbf{P}_3 \to \mathbf{P}_3$ a linear mapping defined by $D\left(p\left(x\right)\right) = p'\left(x\right)$. We already know the associated matrix relative to the standar basis $\mathcal{S}_3 = \left\{1, x, x^2, x^3\right\}$:

$$M_{\mathcal{S}_3}^{\mathcal{S}_3} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This matrix has less pivot columns than columns (hence, D is not injective) and less pivot rows than rows (hence, D not surjective).

Example 3.12

Let $T : \mathbf{P}_3 \to \mathbf{P}_4$ a linear mapping defined by $T\left(p\left(x\right)\right) = xp\left(x\right)$. We already know the associated matrix relative to the basis $\mathcal{B} = \left\{1, 1 + x, x + x^2, x^2 + x^3\right\}$ for $\mathbf{P}_3$, and the standard basis $\mathcal{S}_4 = \left\{1, x, x^2, x^3, x^4\right\}$:

$$M_{\mathcal{S}_4}^{\mathcal{B}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

After computing its row reduced echelon form:

$$M_{\mathcal{S}_4}^{\mathcal{B}} \sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

we find that all its columns are pivot columns, which means that T is injective. However, $\dim \mathbf{P}_3 < \dim \mathbf{P}_4$, so T cannot be surjective.

3.1.4 Associated matrices and the change of basis

Whenever we compute a matrix associated with a linear mapping $T : V \to W$, we explicitly use a fixed basis for each vector space V and W. This means that the explicit expression of the associated matrix depends on the specific choice of these bases. Nevertheless, since the underlying mapping T is independent of the choice of bases, we should expect to find a relationship between associated matrices relative to two different choices of bases. In this section, we will investigate this in detail and describe how an associated matrix varies under a change of bases. Let us start with an illustrative example.

Example 3.13

Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be a linear mapping defined by

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y \\ y-z \end{bmatrix}$$

Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ and $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2\}$ be bases for $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively. We will consider two cases for illustration: when $\mathcal{B}$ and $\mathcal{C}$ are the canonical bases and when they are not, and give the associated matrices for both cases.

Case 1: $\mathcal{B}$ and $\mathcal{C}$ are the canonical bases $\mathcal{E}_3$ and $\mathcal{E}_2$, respectively. The presentation may appear unnecessarily detailed here, but it is illustrative to see the differences with the more complicated case 2 below. We have

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{b}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \ \mathbf{b}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

We calculate, by prescription, $T(\mathbf{b}_i)$, for $i = 1,2,3$:

$$T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

Actually, the resulting vectors are already written in terms the canonical basis $\mathcal{E}_2$, as one can readily confirm:

$$\left[\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right]_{\mathcal{E}_2} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \left[\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right]_{\mathcal{E}_2} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \left[\begin{bmatrix} 0 \\ -1 \end{bmatrix}\right]_{\mathcal{E}_2} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}.$$

Therefore, we have

$$M_{\mathcal{E}_2}^{\mathcal{E}_3} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix}.$$

Case 2: $\mathcal{B}$ and $\mathcal{C}$ are now given by

$$\mathbf{b}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},\ \mathbf{b}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix},\ \mathbf{b}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

$$\mathbf{c}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix},\ \mathbf{c}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

We calculate, by prescription, $T(\mathbf{b}_i)$, for $i = 1,2,3$:

$$T\left(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ -1 \end{bmatrix},\ T\left(\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 0 \end{bmatrix},\ T\left(\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 0 \end{bmatrix}.$$

Now, we need to compute $\left[T(\mathbf{b}_1)\right]_{\mathcal{C}}, \left[T(\mathbf{b}_2)\right]_{\mathcal{C}}$, and $\left[T(\mathbf{b}_3)\right]_{\mathcal{C}}$. For this, we must solve the three systems of equations whose augmented matrices are:

$$\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & T(\mathbf{b}_1) \end{bmatrix}, \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & T(\mathbf{b}_2) \end{bmatrix}, \begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & T(\mathbf{b}_3) \end{bmatrix}.$$

All three systems can be solved simultaneously by setting up the augmented matrix

$$\begin{bmatrix} \mathbf{c}_1 & \mathbf{c}_2 & T(\mathbf{b}_1) & T(\mathbf{b}_2) & T(\mathbf{b}_3) \end{bmatrix},$$

and reducing it to its row reduced echelon form:

$$\begin{bmatrix} 1 & -1 & 1 & 1 & 2 \\ 2 & 1 & -1 & 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & 1 & -1 & -\frac{2}{3} & -\frac{4}{3} \end{bmatrix}.$$

Therefore,

$$M_{\mathcal{C}}^{\mathcal{B}} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ -1 & -\frac{2}{3} & -\frac{4}{3} \end{bmatrix}.$$

As seen in the previous example, while a linear mapping $T : V \to W$ may be independent of the choice of bases for V and W, the explicit expression for its associated matrix will change under different choices of bases.

Our goal here is to start with the associated matrix $M_{\mathcal{C}}^{\mathcal{B}}$ relative to the old bases $\mathcal{B}$ (for V) and $\mathcal{C}$ (for W), and use it to produce the associated matrix relative to new bases $\hat{\mathcal{B}}$ (for V) and $\hat{\mathcal{C}}$ (for W). Let us start by recalling the fundamental property of the associated matrix relative to the old bases $\mathcal{B}$ and $\mathcal{C}$:

$$[T(\mathbf{v})]_{\mathcal{C}} = M_{\mathcal{C}}^{\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}, \text{ for all } \mathbf{v} \in V.$$

Moreover, we can use matrices of change of bases to write the coordinates of the vectors $\mathbf{v} \in V$ and $T(\mathbf{v}) \in W$ relative to the new bases:

$$[\mathbf{v}]_{\mathcal{B}} = P_{\mathcal{B}}^{\hat{\mathcal{B}}}[\mathbf{v}]_{\hat{\mathcal{B}}}, \text{ and } [T(\mathbf{v})]_{\mathcal{C}} = P_{\mathcal{C}}^{\hat{\mathcal{C}}}[T(\mathbf{v})]_{\hat{\mathcal{C}}}.$$

If we insert these two identities in the previous equation, then we obtain

$$P_{\mathcal{C}}^{\hat{\mathcal{C}}}[T(\mathbf{v})]_{\hat{\mathcal{C}}} = M_{\mathcal{C}}^{\mathcal{B}} P_{\mathcal{B}}^{\hat{\mathcal{B}}}[\mathbf{v}]_{\hat{\mathcal{B}}}.$$

Multiplying both sides of the equation on the left by $\left(P_{\mathcal{C}}^{\hat{\mathcal{C}}}\right)^{-1} = P_{\hat{\mathcal{C}}}^{\mathcal{C}}$, we get

$$[T(\mathbf{v})]_{\hat{\mathcal{C}}} = \left[P_{\hat{\mathcal{C}}}^{\mathcal{C}} M_{\mathcal{C}}^{\mathcal{B}} P_{\mathcal{B}}^{\hat{\mathcal{B}}}\right][\mathbf{v}]_{\hat{\mathcal{B}}}.$$

Comparing this equation with the fundamental property satisfied by $M_{\hat{\mathcal{C}}}^{\hat{\mathcal{B}}}$:

$$[T(\mathbf{v})]_{\hat{\mathcal{C}}} = M_{\hat{\mathcal{C}}}^{\hat{\mathcal{B}}}[\mathbf{v}]_{\hat{\mathcal{B}}} \text{ for all v } \in V,$$

we can finally conclude that

$$M_{\hat{\mathcal{C}}}^{\hat{\mathcal{B}}} = P_{\hat{\mathcal{C}}}^{\mathcal{C}} M_{\mathcal{C}}^{\mathcal{B}} P_{\mathcal{B}}^{\hat{\mathcal{B}}}.$$

Hence, we can obtain the associated matrix for T relative to the new bases $\hat{\mathcal{B}}$ and $\hat{\mathcal{C}}$ through a matrix product of three matrices. Note that, on one hand, $M_{\mathcal{C}}^{\mathcal{B}}$ and $M_{\hat{\mathcal{C}}}^{\hat{\mathcal{B}}}$ are both $m \times n$ matrices, where $n = \dim V$ and $m = \dim W$. On the other hand, the matrix $P_{\hat{\mathcal{C}}}^{\mathcal{C}}$ is $m \times m$ and the matrix $P_{\mathcal{B}}^{\hat{\mathcal{B}}}$ is $n \times n$.

We state the above important result as a theorem.

Theorem 3.8

Let $T : V \to W$ be a linear mapping and let $M_{\mathcal{C}}^{\mathcal{B}}$ be its associated matrix relative to the basis $\mathcal{B}$ for V and the basis $\mathcal{C}$ for W. If $\hat{\mathcal{B}}$ and $\hat{\mathcal{C}}$ are new bases for V and W, respectively, then the associate matrix for T relative to the new bases is given by

$$M_{\hat{\mathcal{C}}}^{\hat{\mathcal{B}}} = P_{\hat{\mathcal{C}}}^{\mathcal{C}} M_{\mathcal{C}}^{\mathcal{B}} P_{\mathcal{B}}^{\hat{\mathcal{B}}},$$

where $P_{\mathcal{B}}^{\hat{\mathcal{B}}}$ and $P_{\hat{\mathcal{C}}}^{\mathcal{C}}$ are matrices of change of basis.

Example 3.14

We review the example from above. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be a linear mapping such that

$$T\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \; T\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}.$$

Its associated matrix relative to the canonical bases is

$$M_{\mathcal{E}_3}^{\mathcal{E}_2} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 0 \end{bmatrix}.$$

We wish to determine $M_{\mathcal{C}}^{\mathcal{B}}$ relative to the new bases given by

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} \text{ and } \mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}.$$

By Theorem 3.8, we can obtain $M_{\mathcal{C}}^{\mathcal{B}}$ as follows:

$$M_{\mathcal{C}}^{\mathcal{B}} = P_{\mathcal{C}}^{\mathcal{E}_3} M_{\mathcal{E}_3}^{\mathcal{E}_2} P_{\mathcal{E}_2}^{\mathcal{B}}.$$

Writing the matrix of change of basis from $\mathcal{B}$ and $\mathcal{C}$ to the corresponding canonical bases is straightforward since their columns consist of the basis vectors themselves:

$$P_{\mathcal{E}_2}^{\mathcal{B}} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \; P_{\mathcal{E}_3}^{\mathcal{C}} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

However, we need to find $P_{\mathcal{C}}^{\mathcal{E}_3} = \left(P_{\mathcal{E}_3}^{\mathcal{C}}\right)^{-1}$:

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right] \sim \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 1 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{array}\right].$$

Therefore,

$$P_{\mathcal{C}}^{\mathcal{E}_3} = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

We can finally compute $M_{\mathcal{C}}^{\mathcal{B}}$:

$$M_{\mathcal{C}}^{\mathcal{B}} = P_{\mathcal{C}}^{\mathcal{E}_3} M_{\mathcal{E}_3}^{\mathcal{E}_2} P_{\mathcal{E}_2}^{\mathcal{B}} = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ -1 & 0 \\ 3 & 3 \end{bmatrix}.$$

Example 3.15

Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear mapping defined by

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y \\ y-z \end{bmatrix}.$$

Its associated matrix relative to the canonical bases is

$$M_{\mathcal{E}_2}^{\mathcal{E}_3} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix}.$$

If we wish to determine its associated matrix relative to the bases $\mathcal{B}$ and $\mathcal{C}$ given by

$$\mathcal{B} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\} \text{ and } \mathcal{C} = \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\},$$

then we need to compute the matrices of change of bases $P^{\mathcal{B}}_{\mathcal{E}_3}$ and $P^{\mathcal{E}_2}_{\mathcal{C}}$. We can write down directly the matrices

$$P^{\mathcal{C}}_{\mathcal{E}_2} = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix}, \; P^{\mathcal{B}}_{\mathcal{E}_3} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

We only need to compute $P^{\mathcal{E}_2}_{\mathcal{C}} = \left(P^{\mathcal{C}}_{\mathcal{E}_2} \right)^{-1}$:

$$P^{\mathcal{E}_2}_{\mathcal{C}} = \left(P^{\mathcal{C}}_{\mathcal{E}_2} \right)^{-1} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

By Theorem 3.8, we can obtain $M^{\mathcal{B}}_{\mathcal{C}}$ as follows:

$$M^{\mathcal{B}}_{\mathcal{C}} = P^{\mathcal{E}_2}_{\mathcal{C}} M^{\mathcal{E}_3}_{\mathcal{E}_2} P^{\mathcal{B}}_{\mathcal{E}_3} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ -1 & -\frac{2}{3} & -\frac{4}{3} \end{bmatrix}.$$

3.2 Eigenvalues, eigenvectors, and diagonalization

In this section, we consider a topic that has a wide range of applications in applied mathematics and by extension in the sciences and engineering: the eigenvectors and eigenvalues of a linear mapping. In the presence of some favourable conditions (namely, when the mapping is *diagonalizable*), they offer a way to understand the structure of a linear mapping that map a vector space to itself and, in turn, understand the structure of the involved vector space. Moreover, the study of diagonalizable linear mappings opens the avenue to more advanced methods (e.g., orthogonal or Fourier decomposition of vectors) that we discuss in the next chapter.

3.2.1 Introduction

For a finite-dimensional vector space V with $\dim V = n$, let $T : V \to V$ be a linear mapping, and let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n\}$ be a basis for V. Since maps V into itself then, for each $\mathbf{x}$ in V, the vector $T(\mathbf{x})$ can also be written in the same basis $\mathcal{B}$. The matrix associated with T relative to $\mathcal{B}$ is given by

$$M_{\mathcal{B}}^{\mathcal{B}} = \left[\left[T(\mathbf{b}_1) \right]_{\mathcal{B}} \quad \left[T(\mathbf{b}_2) \right]_{\mathcal{B}} \quad \cdots \quad \left[T(\mathbf{b}_n) \right]_{\mathcal{B}} \right].$$

For the sake of simplicity, the associated matrix $M_{\mathcal{B}}^{\mathcal{B}}$ can simply be denoted as $M^{\mathcal{B}}$. We must mention that any linear mapping that maps a vector space to itself is usually called a *linear operator*. So, from now on, $M^{\mathcal{B}}$ can be understood as the associated matrix of a linear operator relative to a basis $\mathcal{B}$.

In the previous section, we discussed the behavior of the associated matrices of a linear mapping $T : V \to W$ under a change of bases. In the particular case when T is a linear operator (that is, $W = V$), we can consider the change of basis from $\mathcal{B}$ to some other basis $\mathcal{C}$. In this case, by Theorem 3.8, we have

$$M^{\mathcal{C}} = P_{\mathcal{C}}^{\mathcal{B}} M^{\mathcal{B}} P_{\mathcal{B}}^{\mathcal{C}},$$

where $P_{\mathcal{B}}^{\mathcal{C}}$ is the matrix of change of basis from $\mathcal{C}$ to $\mathcal{B}$ and $P_{\mathcal{C}}^{\mathcal{B}} = \left(P_{\mathcal{B}}^{\mathcal{C}} \right)^{-1}$.

The fundamental goal is to find a basis $\mathcal{C}$ (when possible) such that $M^{\mathcal{C}}$ is a diagonal matrix

$$M^{\mathcal{C}} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix},$$

where the empty spaces are understood to be filled with 0 's. If such a basis $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$

exists, then the explicit expression

$$M^{\mathcal{C}} = \left[[T(\mathbf{c}_1)]_{\mathcal{C}} \quad [T(\mathbf{c}_2)]_{\mathcal{C}} \quad \cdots \quad [T(\mathbf{c}_n)]_{\mathcal{C}} \right] = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix},$$

shows that

$$T\left(\mathbf{c}_i\right) = \lambda_i \mathbf{c}_i, \quad \text{for } i = 1,2,\ldots,n.$$

This means that if the linear mapping is applied to a vector of this particular basis, then we obtain a multiple of the same basis vector! This is quite extraordinary since, in general, the application of a linear mapping to an input vector *does not* produce an output vector that is proportional to the input vector. When a non zero vector is mapped by T to a scalar multiple of itself, such a vector is called an *eigenvector*, which comes from German "eigen," meaning "self." The scalar constant that multiplies and eigenvector is called an *eigenvalue*. Each eigenvector is associated with an eigenvalue. We elaborate on this in the next section.

3.2.2 Eigenvalues and eigenvectors

Definition 3.5

Let $T : V \to V$ be a linear mapping. Any **non zero** vector $\mathbf{x} \in V$ such that

$$T\left(\mathbf{x}\right) = \lambda \mathbf{x},$$

for some scalar λ is called an eigenvector of T. The scalar λ is called the eigenvalue associated with $\mathbf{x}$. Eigenvalues are allowed to be zero.

Observation 3.1

The zero vector is excluded as eigenvector, because it would satisfy the equation for *any* value of λ. From the above introductory comments, we would expect that there are exactly $n = \dim V$ eigenvectors and eigenvalues. While sometimes this can be indeed the case, however, this topic will be treated with care and detail below.

Example 3.16

Let $T : \mathbb{C}^2 \to \mathbb{C}^2$ be the linear mapping defined by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -y \\ x \end{bmatrix}.$$

Recall that $\mathbb{C}^2 = \left\{ \begin{bmatrix} x & y \end{bmatrix}^\top : x, y \in \mathbb{C} \right\}$. Verify that $\lambda_1 = i$ and $\lambda_2 = -i$ are eigenvalues of T. If possible, find a basis for $\mathbb{C}^2$ such that its elements are eigenvectors of T and write its associated matrix relative to such basis.

We consider first $\lambda_1 = i$. If λ_i is an eigenvalue of T, then there is some vector $[xy]^\top \in \mathbb{C}^2$

such that

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = i \begin{bmatrix} x \\ y \end{bmatrix}.$$

However, by the definition of T, the same vector must satisfy

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -y \\ x \end{bmatrix}.$$

This implies that

$$\begin{aligned} ix &= -y, \\ iy &= x. \end{aligned}$$

Solving this system of equations, we obtain

$$\begin{bmatrix} x \\ y \end{bmatrix} = \alpha \begin{bmatrix} i \\ 1 \end{bmatrix}, \ \alpha \in \mathbb{R}.$$

This means that all **non zero** scalar multiples of $\begin{bmatrix} i \\ 1 \end{bmatrix}$ are eigenvectors of T associated with λ_1.

Now, we consider $\lambda_2 = -i$. Then, the following two equations must hold for some $\begin{bmatrix} x & y \end{bmatrix}^\top \in \mathbb{C}^2$:

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = -i \begin{bmatrix} x \\ y \end{bmatrix}, \text{ and } T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -y \\ x \end{bmatrix}.$$

This implies that

$$\begin{aligned} -ix &= -y, \\ -iy &= x. \end{aligned}$$

The solution of this system of equation is given by

$$\begin{bmatrix} x \\ y \end{bmatrix} = \beta \begin{bmatrix} -i \\ 1 \end{bmatrix}, \ \beta \in \mathbb{R}.$$

This means that all **non zero** scalar multiples of $\begin{bmatrix} -i \\ 1 \end{bmatrix}$ are eigenvectors of T associated with λ_2. We observe that the set

$$\mathcal{C} = \left\{ \begin{bmatrix} i \\ 1 \end{bmatrix}, \begin{bmatrix} -i \\ 1 \end{bmatrix} \right\},$$

formed by one eigenvector associated with λ_1 and one eigenvector associated with λ_2, is a linearly independent set. Moreover, since $\dim \mathbb{C}^2 = 2$, we conclude that $\mathcal{C}$ is a basis for $\mathbb{C}^2$.

It is not hard to verify that the associated matrix of T relative to $\mathcal{C}$ is

$$M^{\mathcal{C}} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

Note that the order that you choose for the elements of the basis is important. In fact, if we change the order of the elements of $\mathcal{C}$, then we obtain a new basis

$$\hat{\mathcal{C}} = \left\{ \begin{bmatrix} -i \\ 1 \end{bmatrix}, \begin{bmatrix} i \\ 1 \end{bmatrix} \right\},$$

and the associate matrix of T relative to $\hat{\mathcal{C}}$ is

$$M^{\hat{\mathcal{C}}} = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix} \neq M^{\mathcal{C}}.$$

Observation 3.2

The attentive reader would have noticed that in the previous example, the eigenvectors associated with each eigenvalues are not unique: any non zero scalar multiple of the set $\mathcal{C}$ is also an eigevector of T. This means that it is possible to normalize eigenvectors depending on the context, e.g., to obtain an eigenvector of unit length. It also implies that it is allowed to multiply an eigenvector by -1, i.e., changing the sign of all coordinate values at once, if convenient. In fact, we have the following general result.

Theorem 3.9

Let $T: V \to V$ be a linear operator. If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$ are eigenvectors associated with an eigenvalue λ, then any non zero vector $\mathbf{v} \in \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is also an eigenvector associated with λ.

Proof. Choose any non zero vector $\mathbf{v} \in \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$. Then, $\mathbf{v}$ can be written as a linear combination of the eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$. So, for some scalars $a_1, a_2, \ldots, a_r$, we can write

$$\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_r\mathbf{v}_r.$$

By the linearity of T and the fact that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$, are eigenvectors associated with the egienvalue λ, we have that

$$\begin{aligned} T(\mathbf{v}) &= T(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_r\mathbf{v}_r) \\ &= a_1T(\mathbf{v}_1) + a_2T(\mathbf{v}_2) + \cdots + a_rT(\mathbf{v}_r) \\ &= a_1\lambda\mathbf{v}_1 + a_2\lambda\mathbf{v}_2 + \cdots + a_r\lambda\mathbf{v}_r \\ &= \lambda(a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_r\mathbf{v}_r) \\ &= \lambda\mathrm{v}. \end{aligned}$$

Hence, $T(\mathbf{v}) = \lambda\mathbf{v}$, which proves that $\mathbf{v}$ is an eigenvector of T associated with λ. ■

The following important theorem will be needed later.

Theorem 3.10

Let $T : V \to V$ be a linear operator and let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$ be eigenvectors associated with *distinct* eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_m$, i.e., $\lambda_i \neq \lambda_j$ for $i \neq j$. Then, $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ is a linearly independent set.

Proof. We prove this theorem by contradiction. Suppose that $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\}$ is a linearly dependent set. Therefore, since $\mathbf{x}_1 \neq 0$, we know that one of the vectors in the set is a linear combination of the preceeding vectors (see Theorem 2.6). Let p be the least index such that $\mathbf{x}_{p+1}$ is a linear combination of the preceeding (linearly independent) vectors. Then there exist scalars $c_1, \cdots, c_p$ such that

$$c_1\mathbf{x}_1 + \cdots + c_p\mathbf{v}_0 = \mathbf{v}_{p+1}.$$

On one hand, applying T to both sides of this equation and using the fact that $T(\mathbf{x}_k) = \lambda_k\mathbf{x}_k$ for each k, we obtain

$$\lambda_1c_1\mathbf{x}_1 + \cdots + \lambda_pc_p\mathbf{x}_p = \lambda_{p+1}\mathbf{v}_{p+1}.$$

On the other hand, if we multiply both sides by λ_{p+1}, we get

$$\lambda_{p+1}c_1\mathbf{x}_1 + \cdots + \lambda_{p+1}c_p\mathbf{x}_p = \lambda_{p+1}\mathbf{v}_{p+1}.$$

Subtracting the previous two equations, we obtain

$$\left(\lambda_1 - \lambda_{p+1}\right)c_1\mathbf{x}_1 + \cdots + \left(\lambda_p - \lambda_{p+1}\right)c_p\mathbf{x}_p = 0.$$

Since $\left\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\right\}$ is linearly independent, the weights in the equation above are all zero. But none of the factors $\lambda_i - \lambda_{p+1}$ are zero, because the eigenvalues are distinct. Hence $c_i = 0$ for $i = 1, \ldots, p$, which means that $\mathbf{v}_{p+1} = 0$, which is impossible. Hence, $\left\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m\right\}$ cannot be linearly dependent and therefore it must be linearly independent. ■

We note that the reciprocal of the previous theorem is not true in general. That is, it is quite possible for a linear operator to have several linearly independent eigenvectors associated with one eigenvalue. What the previous theorem is saying is that an eigenvector of a linear operator can not be associated with two distinct eigenvalues.

The following definition is motivated by Theorem 3.9.

Definition 3.6

Let $T: V \to V$ be a linear operator. For each eigenvalue λ of T, we define the set

$$E_\lambda = \mathrm{Span}\left\{\mathbf{x} : T\left(\mathbf{x}\right) = \lambda\mathbf{x}\right\}.$$

It can easily be deduced from Theorem 3.9 that the set E_λ for each eigenvalue λ of T is a vector subspace of V. So we will call E_λ the ***eigenspace associated with*** λ. Observe that E_λ is composed of all the (non zero) linear combinations of the eigenvectors associated with λ, together with $0 \in V$ (otherwise, E_λ is not a vector subspace), which is not an eigenvector. By Theorem 3.10 we know that any eigenvector of T cannot belong to two different eigenspaces. Therefore we have that $E_{\lambda_i} \cap E_{\lambda_j} = \left\{0\right\}$ whenever $\lambda_i \neq \lambda_j$.

We are now paving the way for a central theorem for eigenvectors and eigenvalues. Let $T: V \to V$ be a linear operator. Our goal will be to characterize the eigenspaces of T. In order to do this, we search for solutions of the equation

$$T\left(\mathbf{x}\right) = \lambda\mathbf{x},$$

where λ is a scalar and $\mathbf{x}$ is a non zero vector in V. This equation can also be written as

$$\left(T - \lambda Id\right)\left(\mathbf{x}\right) = 0,$$

where Id denotes the identity operator on V defined by $Id(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in V$. This means that λ is an eigenvalue of T with associated eigenvector $\mathbf{x}$ if and only if the above equation has a non trivial solution. Consequently, we have the following characterization of the eigenspaces of T.

Theorem 3.11

Let $T : V \to V$ be a linear operator. Then, for each eigenvalue λ of T

$$E_{\lambda} = \operatorname{Ker}(T - \lambda Id).$$

Now, let us fix a basis $\mathcal{B}$ of V and let $M^{\mathcal{B}}$ be the associated matrix of T relative to $\mathcal{B}$. Then the associated matrix of the operator $T - \lambda Id$ relative to $\mathcal{B}$ is given by $M^{\mathcal{B}} - \lambda I_n$, where I_n is the identity matrix and $n = \dim V$. In this way, we can translante the above discussion about eigenspaces to associated matrices and coordinate vectors: the vector $\mathbf{x} \in V$ satisfies $(T - \lambda Id)(\mathbf{x}) = 0$ if and only if

$$\left(M^{\mathcal{B}} - \lambda I_n\right)[\mathbf{x}]_{\mathcal{B}} = 0 \in \mathbb{R}^n.$$

We immediately have the following result.

Theorem 3.12

Let $T : V \to V$ be a linear operator. Let $\mathcal{B}$ be a basis for V and let $M^{\mathcal{B}}$ be the associated matrix of T relative to $\mathcal{B}$. Then, for each eigenvalue λ of T:

(a) $\mathbf{x} \in E_{\lambda}$ if and only if $[\mathbf{x}]_{\mathcal{B}} \in \operatorname{Nul}\left(M^{\mathcal{B}} - \lambda I_n\right)$, with $n = \dim V$.

(b) $\dim E_{\lambda} = \dim \operatorname{Nul}\left(M^{\mathcal{B}} - \lambda I_n\right)$.

So far, we have described the eigenspaces of a linear operator assuming that we already know the eigenvalues. However, we still need to describe how to find the eigenvalues, so we make the following observation. The homogeneous equation $\left(M^{\mathcal{B}} - \lambda I_n\right)[\mathbf{x}]_{\mathcal{B}} = 0$ has a non trivial solution if and only if the value of λ is chosen so that the matrix $M^{\mathcal{B}} - \lambda I_n$ is not invertible. Equivalently, the values of λ must be choses so that

$$\det\left(M^{\mathcal{B}} - \lambda I_n\right) = 0.$$

It turns out that $\det\left(M^{\mathcal{B}} - \lambda I_n\right)$ is a polynomial of degree at most n and, therefore, if λ is an eigenvalue of T then it is a root of this polynomial. This means that a linear operator has at most n distinct eigenvalues.

This polynomial is important enough to deserve a special name.

Definition 3.7

Let $T:V \to V$ be a linear operator and let $M^{\mathcal{B}}$ be its associated matrix relative to a basis $\mathcal{B}$.

(a) The polynomial $p(\lambda) = \det\left(M^{\mathcal{B}} - \lambda I_n\right)$ of degree$n = \dim V$ is called the characteristic polynomial of T.

(b) $p(\lambda) = \det\left(M^{\mathcal{B}} - \lambda I_n\right) = 0$ is called the characteristic equation of T.

We must remark that important fact that the characteristic polynomial of a linear operator is independent of the basis that we choose to compute the associated matrix. To show this, suppose that P is a matrix of change of basis from $\mathcal{B}$ to some new basis, and let $M = P^{-1}M^{\mathcal{B}}P$ is the associated matrix relative to the new basis. Then,

$$\det\left(M - \lambda I_n\right) = \det\left[P^{-1}M^{\mathcal{B}}P - \lambda P^{-1}I_nP\right] = \det\left[P^{-1}\left(M^{\mathcal{B}} - \lambda I_n\right)P\right].$$

By the multiplicative property of determinants $\left(\det AB = (\det A)(\det B)\right)$, we have

$$\det\left(M - \lambda I_n\right) = \left(\det P^{-1}\right)\left[\det\left(M^{\mathcal{B}} - \lambda I_n\right)\right]\left(\det P\right) = \det\left(M^{\mathcal{B}} - \lambda I_n\right),$$

where we have used the fact that $\left(\det P^{-1}\right)\left(\det P\right) = 1$. This means that the characteristic polynomial (and its roots) does not change regardless of its matrix representation.

Theorem 3.13

Let $T:V \to V$ be a linear operator and $n = \dim V$. Then T has at most n linear independent eigenvectors.

Proof. Since the characteristic polynomial has degree n, then T has at most n distinct eigenvalues. It follows from Theorem 3.10 that T has at most n linearly independent eigenvectors. ■

Example 3.17

Find the eigenvalues and eigenvectors of $T:\mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T(\mathbf{x}) = A\mathbf{x} \text{ with } A = \begin{bmatrix} 2 & 3 \\ 3 & -6 \end{bmatrix}.$$

(Although it is not explicitly mentioned, the expressions for $\mathbf{x} \in \mathbb{R}^2$ and T are given in terms of the canonical basis.)

We may proceed as follows.

(1) First, we find the eigenvalues. Recall that the eigenvalues of T are the roots of its characteristic polynomial given by

$$p(\lambda) = \det(A - \lambda I_2) = \det\begin{bmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{bmatrix} = \lambda^2 + 4\lambda - 21 = (\lambda+7)(\lambda-3).$$

Therefore, the eigenvalues of T are

$$\lambda_1 = -7,\ \lambda_2 = 3.$$

(2) Now, we obtain the eigenvectors associated with each eigenvalue.

(a) Let us start with $\lambda_1 = -7$. The eigenvectors associated with this eigenvalue are vectors in the eigenspace $E_{\lambda_1} = \mathrm{Ker}(T + 7Id)$. Since we have (implicitly) fixed the canonical basis, we have that $\mathbf{x} = [\mathbf{x}]_{\mathcal{E}_2}$ and $M^{\mathcal{E}_2} = A$. Moreover, by Theorem 3.12 we have that $\mathbf{x} \in E_{\lambda_1}$ if and oly if $\mathbf{x} \in \mathrm{Nul}(A + 7I_2)$; that is,

$$(A + 7I_2)\mathbf{x} = 0.$$

Noticing that

$$A + 7I_2 = \begin{bmatrix} 9 & 3 \\ 3 & 1 \end{bmatrix} \sim \begin{bmatrix} 1 & \frac{1}{3} \\ 0 & 0 \end{bmatrix},$$

we deduce that

$$\mathrm{Nul}(A + 7I_2) = \mathrm{Span}\left\{\begin{bmatrix} -\frac{1}{3} \\ 1 \end{bmatrix}\right\}.$$

Therefore, $\dim E_{\lambda_1} = 1$ and every eigenvector associated with λ_1 is of the form

$$\mathbf{x} = \mu\begin{bmatrix} -\frac{1}{3} \\ 1 \end{bmatrix}\ \mu \neq 0,$$

(eigenvectors must be non zero vectors).

(b) Similarly, for $\lambda_1 = 3$, we have

$$\mathrm{Nul}(A - 3I_2) = \mathrm{Span}\left\{\begin{bmatrix} 3 \\ 1 \end{bmatrix}\right\}.$$

Therefore, $\dim E_{\lambda_2} = 1$ and every eigenvector associated with λ_2 is of the form

$$\mathbf{x} = \eta \begin{bmatrix} 3 \\ 1 \end{bmatrix} \quad \eta \neq 0.$$

We would like to remark that the two eigenvalues are distinct $(\lambda_1 \neq \lambda_2)$. Then, by Theorem 3.10, the set $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2\}$ with

$$\mathbf{v}_1 = \begin{bmatrix} -\frac{1}{3} \\ 1 \end{bmatrix}, \; \mathbf{v}_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix},$$

is linearly independent since $\mathbf{v}_1$ and $\mathbf{v}_2$ are eigenvectors associated with λ_1 and λ_2, respectively. Furthermore, $\dim \mathbb{R}^2 = 2$, so these two linearly independent eigenvectors form a basis of $\mathbb{R}^2$! It turns out that if we write the matrix representation of T relative to $\mathcal{B}$, we obtain

$$M^{\mathcal{B}} = \begin{bmatrix} -7 & 0 \\ 0 & 3 \end{bmatrix},$$

which, conveniently, is a diagonal matrix whose non zero element are precisely the eigenvalues of T.

Dxample 3.18

Here, we revisit Example 3.16 with the emphasis on showing the systematic procedure of calculating eigenvalues and eigenvectors starting from a linear operator.

Let $T : \mathbb{C}^2 \to \mathbb{C}^2$ be a linear operator defined by

$$T(\mathbf{x}) = A\mathbf{x}, \quad \text{with } A = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

We have implicitly fixed the canonical basis for $\mathbb{C}^2$ to represent the operator as a matrix. Find the eigenvalues and eigenvectors of T.

(1) First, let us find the eigenvalues of T. The characteristic polynomial is given by

$$p(\lambda) = \det(A - \lambda I_2) = \det \begin{bmatrix} -\lambda & 1 \\ -1 & -\lambda \end{bmatrix} = \lambda^2 + 1.$$

The eigenvalues of T are the roots of the characteristic polynomial. In this case, we have two distinct eigenvalues given by $\lambda_1 = i$ and $\lambda_2 = -i$. Notice that these eigenvalues are complex numbers.

(2) Next, we describe the eigenvectors associated with each of the eigenvalues.

(a) For $\lambda_1 = i, E_{\lambda_1} = \text{Ker}(T - iId)$. As in Example 3.17, we have fixed the canonical basis and, therefore, E_{λ_1} coincides with $\text{Nul}(A - iI_2)$. Hence, we compute

$$A - iI_2 = \begin{bmatrix} -i & 1 \\ -1 & -i \end{bmatrix} \sim \begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix},$$

and therefore $E_{\lambda_1} = \text{Span}\left\{\begin{bmatrix} -i \\ 1 \end{bmatrix}\right\}$.

(b) Similarly, for $\lambda_2 = -i$, we have $E_{\lambda_2} = \text{Nul}(A + iI_2)$. Since,

$$A + iI_2 = \begin{bmatrix} i & 1 \\ -1 & i \end{bmatrix} \sim \begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix},$$

we obtain $E_{\lambda_2} = \text{Span}\left\{\begin{bmatrix} i \\ 1 \end{bmatrix}\right\}$.

Now, consider the set

$$\mathcal{B} = \left\{\begin{bmatrix} -i \\ 1 \end{bmatrix}, \begin{bmatrix} i \\ 1 \end{bmatrix}\right\},$$

composed of one eigenvector associated with each eigenvalue. Since the eigenvalues are distinct, $\mathcal{B}$ is a linearly independent set in $\mathbb{C}^2$ with $\dim \mathbb{C}^2 = 2$. Therefore, $\mathcal{B}$ is a basis for $\mathbb{C}^2$. It is straightforward to verify that the matrix representation of T in this new basis is

$$M^{\mathcal{B}} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

3.2.3 Multiplicity of eigenvalues

So far, we have encountered examples of linear operators whose eigenvalues are all distinct and, consequently, it was possible to construct a basis for the involved vector space consisting entirely of eigenvectors of the linear operators. However, if we take into account the multiplicity of the linear factors of the characteristic polynomial, then it is not the case that the eigenvalues of every linear operator are all distinct as we show in the following example.

Example 3.19

Let $T:\mathbb{R}^3 \to \mathbb{R}^3$ be a linear operator defined by

$$T(\mathbf{x}) = A\mathbf{x} \;\text{ with }\; A = \begin{bmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{bmatrix}$$

As usual, we have fixed the canonical basis of $\mathbb{R}^3$ and, thus, A is the matrix associated with T relative to the canonical basis.

Let us compute the characteristic polynomial of T :

$$p(\lambda) = \det(A - \lambda I_3) = \det\begin{bmatrix} 4-\lambda & -1 & 6 \\ 2 & 1-\lambda & 6 \\ 2 & -1 & 8-\lambda \end{bmatrix} = -(\lambda-2)^2(\lambda-9).$$

Notice that $p(\lambda)$ has two distinct real roots $\lambda_1 = 2$ and $\lambda_2 = 9$. However, if we take into account the multiplicity of each linear factor, then $p(\lambda)$ has three roots: one **double** root $\lambda_1 = 2$ and a **single** root $\lambda_2 = 9$. Hence, T does not have three distinct eigenvalues and, unfortunately, this means that the existence of a basis for $\mathbb{R}^3$ consisting entirely of eigenvectors of T is not guaranteed. At this point, we know that we can choose two linearly independent eigenvectors: one from E_{λ_1} and another one from E_{λ_2} because $\lambda_1 \neq \lambda_2$. But it is not clear if a third appropriate eigenvector exists or, even worse, where to find it.

However, if we are lucky enough, one of the eigenspaces could have dimension 2 and, therefore, we could obtain the missing third eigenvector from such eigenspace. In fact, this is precisely what happens in this example. You can verify that

$$E_{\lambda_1} = \operatorname{Span}\left\{\begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}\right\} \;\text{ and }\; E_{\lambda_2} = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right\}.$$

Fortunately, $\dim E_{\lambda_1} = 2$ and, therefore, we can choose two linearly independent eigenvectors associated with λ_1; for instance

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

Moreover, the set obtained from B by adding **any** eigenvector associated with λ_2 will be linearly independent (because $\lambda_1 \neq \lambda_2$). In total, we have three linearly independent eigenvectors that can be used to construct a basis for $\mathbb{R}^3$. For example, This means that the set

$$\mathcal{B} = \left\{ \begin{bmatrix} 1/2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\},$$

is a basis for $\mathbb{R}^3$ consisting entirely of eigenvectors of T.

The following example shows that not every linear operator $T: V \to V$ admits enough linearly independent eigenvectors to constitute a basis for V.

Example 3.20

Fix the canonical basis of $\mathbb{R}^3$, and let $T: \mathbb{R}^3 \to \mathbb{R}^3$ be a linear operator defined by

$$T(\mathbf{x}) = A\mathbf{x} \text{ with } A = \begin{bmatrix} 2 & 2 & 3 \\ 0 & 2 & 2 \\ 0 & 0 & 2 \end{bmatrix}.$$

Let us examine the eigenvalues and eigenvectors of T.

(1) First, we find the eigenvalues. The characteristic polynomial of T is

$$p(\lambda) = \det(A - \lambda I_3) = \det \begin{bmatrix} 2-\lambda & 2 & 3 \\ 0 & 2-\lambda & 2 \\ 0 & 0 & 2-\lambda \end{bmatrix} = (2-\lambda)^3.$$

Therefore, $\lambda = 2$ is a triple root of $p(\lambda)$ which, in turn, is the only eigenvalue of T.

(2) The corresponding eigenspace for $\lambda = 2$ is $E_2 = \mathrm{Nul}(A - 2I_3)$. We have

$$E_2 = \mathrm{Nul}(A - 2I_3) = \mathrm{Nul} \begin{bmatrix} 0 & 2 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} = \mathrm{Nul} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \mathrm{Span} \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \right\}.$$

In this case, T only admits one linearly independent eigenvector since $\dim E_2 = 1$. Therefore, it is not possible to produce a basis for $\mathbb{R}^3$ consisting entirely of eigevenctors of T.

There are two quantities that play an important role in the general analysis of the eigenspaces of a linear operator: the multiplicity of each eigenvalue as a root of the characteristic polynomial and the dimension of the corresponding eigenspaces. These quantities deserve their own terminology.

Definition 3.8

Let $T: V \to V$ be a linear operator and let $p(x)$ be its characteristic polynomial. For each distinct eigenvalue λ of T, we have

$$p(x) = (x-\lambda)^{m_\lambda} q(\lambda),$$

where $q(x)$ is a non zero polynomial such that $q(\lambda) \neq 0$, and m_λ is a positive integer.

(a) We say that m_λ is the **algebraic multiplicity** of λ.

(b) We say that $\dim E_\lambda$ is the **geometric multiplicity** of λ.

The following theorem states the relationship between the algebraic and geometric multiplicities of the eigenvalues of a linear operator.

Theorem 3.14

Let $T: V \to V$ be a linear operator. Then, for each distinct eigenvalue λ,

$$1 \leq \dim E_\lambda \leq m_\lambda.$$

3.2.4 Diagonalization of linear operators

Now, we return to the original motivation of this chapter: given a linear operator $T: V \to V$, find a basis of V such that the matrix associated with T is a diagonal matrix. Let us introduce the following definition.

Definition 3.9

A linear operator $T: V \to V$ is said to be diagonalizable if there is a basis $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_n\}$ of V made up entirely of eigenvectors of T.

If a linear operator is diagonalizable, then its associated matrix relative to the basis of eigenvector $\mathcal{C}$ is of the form

$$M^{\mathcal{C}} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix},$$

where $T(\mathbf{c}_i) = \lambda_i \mathbf{c}_i$, for $i = 1,2,\dots,n$ (see the introductory comments of this section). As we saw before, not every linear operator is diagonalizable.

An immediate consequence of the above definition is the following theorem.

Theorem 3.15

Let $T : V \to V$ be a linear operator and $n = \dim V$. Then T is diagonalizable if and only if it admits n linear independent eigenvectors.

The previous theorem, together with Theorem 3.10, yields the following result.

Corollary 3.1

Let $T : V \to V$ be a linear operator and $n = \dim V$. If T has n distinct eigenvalues, then T is diagonalizable.

Observe that from Definition 3.8, we have that T has n distinct eigenvalues if and only if $m_\lambda = 1$; that is, each eigenvalue of T has algebraic multiplicity equal to 1 . Furthermore, by Theorem 3.14, this means that $\dim E_\lambda = 1$ for each eigenvalue λ .

In general, however, a linear operator may admit eigenvalues with algebraic multiplicity greater than 1. In this general setting, a linear operator is diagonalizable if and only if, for each eigenvalue λ with algebraic multiplicity m_λ , there are m_λ linear independent eigenvectors associated with λ . We formalize this fact in the following theorem.

Theorem 3.16

For a linear operator $T : V \to V$, let $\lambda_1, \lambda_2, \dots, \lambda_k$, be all the distinct eigenvalues of T . Then T is diagonalizable if and only if

$$\dim E_{\lambda_1} + \dim E_{\lambda_2} + \cdots + \dim E_{\lambda_k} = \dim V$$

Equivalently, T is diagonalizable if and only if

$$\dim E_{\lambda_j} = m_{\lambda_j}, \quad \text{for } j = 1,2,\dots,k.$$

Now, we turn our attention the the associated matrix of a diagonalizable linear operator.

Theorem 3.17

Let $T : V \to V$ be a diagonalizable linear operator, $n = \dim V$. Then there is a basis for $V, \mathcal{C} = \{\mathbf{c}_1, \ldots, \mathbf{c}_n\}$, formed by eigenvectors of T; that is, $T(\mathbf{c}_i) = \lambda_i \mathbf{c}_i$ for $i = 1, 2, \ldots, n$. Moreover, the matrix associated with T relative to $\mathcal{C}$ is a diagonal matrix given by

$$M^{\mathcal{C}} = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}.$$

Some remarks are in order. In Theorem 3.17, the associated matrix $M^{\mathcal{C}}$ is a diagonal matrix with the eigenvalues of T in its main diagonal. If $\mathcal{B}$ is a different basis for V, then

$$M^{\mathcal{B}} = P_{\mathcal{B}}^{\mathcal{C}} M^{\mathcal{C}} \left(P_{\mathcal{B}}^{\mathcal{C}}\right)^{-1}$$

where the change of basis matrix is given by

$$P_{\mathcal{B}}^{\mathcal{C}} = \left[[\mathbf{c}_1]_{\mathcal{B}} \quad \cdots \quad [\mathbf{c}_n]_{\mathcal{B}} \right].$$

Notice that the columns of this change of basis matrix are the coordinates of the eigenvectors of T relative to $\mathcal{B}$. Moreover, the eigenvalues of T appear in the main diagonal of $M^{\mathcal{C}}$ in the same order as the corresponding eigenvectors in the basis $\mathcal{C}$.

Most textbooks present the topic of diagonalization of linear operators as a method for factorizing matrices (referring to the associated matrix $M^{\mathcal{B}}$) into a product PDP^{-1}, where P is an invertible matrix and D is a diagonal matrix. However, these notes are intended to approach diagonalization from an operator point of view.

Finally, observe that we have total freedom in choosing the order of the eigenvectors in the basis $\mathcal{C}$. This means that the diagonal associated matrix for T is only uniquely determined up to the order in which the eigenvectors of T appear in $\mathcal{C}$.

Example 3.21

Fix the standard basis of $\mathbb{R}^3$. Let $T:\mathbb{R}^3 \to \mathbb{R}^3$ be a linear operator defined by $T(\mathbf{x}) = A\mathbf{x}$, with

$$A = \begin{bmatrix} 1 & 3 & 3 \\ -3 & -5 & -3 \\ 3 & 3 & 1 \end{bmatrix}.$$

If possible, diagonalize the linear operator T.

(1) First, we find the eigenvalues. The characteristic equation is

$$p(\lambda) = \det(A - \lambda I_3) = \det\begin{bmatrix} 1-\lambda & 3 & 3 \\ -3 & -5-\lambda & -3 \\ 3 & 3 & 1-\lambda \end{bmatrix} = -(\lambda - 1)(\lambda + 2)^2.$$

Therefore, the eigenvalues of T are $\lambda_1 = 1$ and $\lambda_2 = -2$ with algebraic multiplicities $m_{\lambda_1} = 1$ and $m_{\lambda_2} = 2$, respectively.

(2) Now we find the eigenvectors.

(a) For $\lambda_1 = 1$, we have $m_{\lambda_1} = 1$, so we expect know that $\dim E_{\lambda_1} = 1$. Indeed,

$$E_{\lambda_1} = \text{Nul}(A - I_3) = \text{Nul}\begin{bmatrix} 0 & 3 & 3 \\ -3 & -6 & -3 \\ 3 & 3 & 0 \end{bmatrix} = \text{Span}\left\{\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}\right\}.$$

(b) For $\lambda_2 = -2$ with $m_{\lambda_2} = 2$, we have $1 \le \dim E_{\lambda_2} \le 2$. Let us compute

$$E_{\lambda_2} = \text{Nul}(A + 2I_3) = \text{Nul}\begin{bmatrix} 3 & 3 & 3 \\ -3 & -3 & -3 \\ 3 & 3 & 3 \end{bmatrix} = \text{Span}\left\{\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}.$$

Therefore, $\dim E_{\lambda_2} = 2$.

Since $\dim E_{\lambda_1} + \dim E_{\lambda_2} = \dim \mathbb{R}^3$, by Theorem 3.16, T is diagonalizable.

(3) Let us compute a diagonal associated matrix for T. For this, we must construct a basis for $\mathbb{R}^3$ consisting entirely of eigenvectors. We are free to choose the order of the eigenvectors, but we need to be consistent. Let

$$\mathcal{C} = \left\{\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}\right\}$$

be the basis of eigenvectors. Then, $A = PDP^{-1}$ with

$$D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix},\ P = \begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Here, $D \equiv M^{\mathcal{C}}$ and $P \equiv P^{\mathcal{C}}_{\mathcal{E}_3}$ ($\mathcal{E}_3$ denotes the canonical basis).

Example 3.22

Fix the canonical basis of $\mathbb{R}^3$. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ defined by $T(\mathbf{x}) = A\mathbf{x}$, with

$$A = \begin{bmatrix} 2 & 4 & 3 \\ -4 & -6 & -3 \\ 3 & 3 & 1 \end{bmatrix}.$$

If possible, diagonalize the linear operator T.

(1) First, we find the eigenvalues. The characteristic equation is

$$p(\lambda) = \det(A - \lambda I_3) = \det \begin{bmatrix} 2-\lambda & 4 & 3 \\ -4 & -6-\lambda & -3 \\ 3 & 3 & 1-\lambda \end{bmatrix} = -(\lambda - 1)(\lambda + 2)^2.$$

Therefore, the eigenvalues of T are $\lambda_1 = 1$ and $\lambda_2 = -2$ with algebraic multiplicities $m_{\lambda_1} = 1$ and $m_{\lambda_2} = 2$, respectively. Note that the characteristic polynomial is identical to the one of the previous example, and therefore the eigenvalues are the same as well.

(2) Now we find the eigenvectors. Let us start examining the eigenvectors associated with $\lambda_2 = -2$ with $m_{\lambda_2} = 2$, since there resides the only possiblility for the matrix not being diagonalizable if $\dim E_{\lambda_2} \neq m_{\lambda_2}$. We compute

$$E_{\lambda_2} = \text{Nul}(A + 2I_3) = \text{Nul} \begin{bmatrix} 4 & 4 & 3 \\ -4 & -4 & -3 \\ 3 & 3 & 3 \end{bmatrix} = \text{Nul} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \text{Span}\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

Since $\dim E_{\lambda_2} \neq m_{\lambda_2}$, it is not possible to find a basis for $\mathbb{R}^3$ consisting entirely of eigenvectors of T and, therefore, T is not diagonalizable.