Note 11

The eigenvalue problem and diagonalization

Let us take a look at Example 10.35 again. In it, we considered the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

and the linear map $L_{\mathbf{A}} : \mathbb{R}^2 \to \mathbb{R}^2$ associated to it. We further saw that if we chose the same ordered basis $\beta = (\begin{bmatrix} -1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix})$ for both the domain and the codomain of $L_{\mathbf{A}}$, then the resulting mapping matrix ${}_\beta[L_{\mathbf{A}}]_\beta$ of $L_{\mathbf{A}}$ was particularly nice:

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

In this chapter, we investigate to which extent this can be done for an arbitrary square matrix.

11.1 Eigenvalues and eigenvectors

We start out by studying linear maps $L : V \to V$. The difference with our previous studies of linear maps is that we now assume that the domain of L is the same as the codomain of L, namely the vector space V.

|||| Definition 11.1

Let $\mathbb{F}$ be a field, V a vector space over $\mathbb{F}$ and $L : V \to V$ a linear map. Let $\mathbf{v} \in V$ be a nonzero vector and $\lambda \in \mathbb{F}$ a scalar such that

$$L(\mathbf{v}) = \lambda \cdot \mathbf{v}.$$

Then the vector $\mathbf{v}$ is called an *eigenvector* of the linear map L with *eigenvalue* λ.

Note that by definition an eigenvector is always a nonzero vector. The reason for this is to avoid uninteresting solutions to the equation $L(\mathbf{v}) = \lambda \cdot \mathbf{v}$. Indeed, if one chooses $\mathbf{v} = \mathbf{0}$ and any $\lambda \in \mathbb{F}$, then it will hold that $L(\mathbf{v}) = \lambda \cdot \mathbf{v}$, since $L(\mathbf{0}) = \mathbf{0}$ and $\lambda \cdot \mathbf{0} = \mathbf{0}$. Further note that an eigenvalue always is an element from the field $\mathbb{F}$ over which V is a vector space. Intuitively, what an eigenvector of a linear operator L is, is a vector that is scaled when L operates on it. Indeed, we can think of $\lambda \cdot \mathbf{v}$ as a scaling of the vector $\mathbf{v}$ by a factor λ. For matrices one can also talk about eigenvectors and eigenvalues:

|||| Definition 11.2

Let $\mathbb{F}$ be a field, n a positive integer and $\mathbf{A} \in \mathbb{F}^{n \times n}$ a matrix. Let $\mathbf{v} \in \mathbb{F}^n$ be a nonzero vector and $\lambda \in \mathbb{F}$ a scalar such that

$$\mathbf{A} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}.$$

Then the vector $\mathbf{v}$ is called an *eigenvector* of the matrix $\mathbf{A}$ with *eigenvalue* λ.

Note that this definition assumed that the matrix $\mathbf{A}$ is a square matrix. As we have seen, a matrix $\mathbf{A} \in \mathbb{F}^{m \times n}$ gives rise to a linear map $L_{\mathbf{A}} : \mathbb{F}^n \to \mathbb{F}^m$. If $m = n$, we therefore see that a square matrix $\mathbf{A} \in \mathbb{F}^{n \times n}$ gives rise to a linear map $L_{\mathbf{A}} : \mathbb{F}^n \to \mathbb{F}^n$. Note that $\mathbf{v} \in \mathbb{F}^n$ is an eigenvector of a square matrix $\mathbf{A} \in \mathbb{F}^{n \times n}$ if and only if $\mathbf{v} \in \mathbb{F}^n$ is an eigenvector of the linear map $L_{\mathbf{A}} : \mathbb{F}^n \to \mathbb{F}^n$. In that sense Definition 11.2 is just a special case of Definition 11.1. Also for matrices it holds that if the field is specified to be $\mathbb{F}$, then its eigenvalues are by definition elements of that field $\mathbb{F}$.

Let us consider some examples.

Example 11.3

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

$$\mathbf{A} = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix} \in \mathbb{R}^{2\times 2}.$$

Question: Determine all possible eigenvalues of the matrix $\mathbf{A}$ as well as a corresponding eigenvector.

Answer: Assume that $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2 \setminus \{(0,0)\}$ is an eigenvector with eigenvalue λ. The equation $\mathbf{A} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}$ is equivalent to the two equations $-v_1 = \lambda v_1$ and $2v_2 = \lambda v_2$. These two equations can be rewritten as $(-1-\lambda)v_1 = 0$ and $(2-\lambda)v_2 = 0$, which in turn can be written in matrix form as follows:

$$\begin{bmatrix} -1-\lambda & 0 \\ 0 & 2-\lambda \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{11-1}$$

Now we distinguish three cases.

In the first case we assume that $-1-\lambda \neq 0$ and $2-\lambda \neq 0$. In other words: we assume that $\lambda \neq -1$ and $\lambda \neq 2$. In this case the diagonal elements of the matrix occurring in equation (11-1) are both nonzero. Hence the only solution to equation (11-1) is $(v_1, v_2) = (0,0)$. However, eigenvectors are by definition not equal to the zero vector, so we conclude that in this case there are no eigenvectors with eigenvalue λ.

In case two, we assume that $\lambda = -1$. In this case equation (11-1) has solutions of the form $(v_1, 0)$, where $v_1 \in \mathbb{R}$ can be chosen freely. Hence $\lambda = -1$ is an eigenvalue of the given matrix $\mathbf{A}$. As eigenvector we can choose any vector of the form $\begin{bmatrix} v_1 \\ 0 \end{bmatrix}$ as long as $v_1 \neq 0$. For example $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is an eigenvector of the given matrix $\mathbf{A}$ with eigenvalue -1.

Finally, as the third and final case, we assume that $\lambda = 2$. In this case equation (11-1) has solutions of the form $(0, v_2)$, where $v_2 \in \mathbb{R}$ can be chosen freely. Hence $\lambda = 2$ is an eigenvalue of the given matrix $\mathbf{A}$. As eigenvector we can choose any vector of the form $\begin{bmatrix} 0 \\ v_2 \end{bmatrix}$ as long as $v_2 \neq 0$. For example $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is an eigenvector of the given matrix $\mathbf{A}$ with eigenvalue 2.

Also in case V is an infinite dimensional vector space, the definition of eigenvectors and eigenvalues makes sense. We consider an example of this type.

|||| Example 11.4

Let us consider the linear map $D : \mathbb{C}[Z] \to \mathbb{C}[Z]$ defined in Example 10.15. In particular, we are working over the field $\mathbb{C}$, since in Example 10.15 we considered $\mathbb{C}[Z]$ as a complex vector space. The map D was defined by sending a polynomial to its derivative.

Question: What are the possible eigenvalues of D? Also for each possible eigenvalue, find a corresponding eigenvector.

Answer: We are looking for nonzero polynomials $p(Z)$ in $\mathbb{C}[Z]$ and scalars $\lambda \in \mathbb{C}$ such that $D(p(Z)) = \lambda \cdot p(Z)$. Let $p(Z) = a_0 + a_1 Z + a_2 Z^2 + \cdots + a_n Z^n$ be a nonzero polynomial. Since $D(a_0 + a_1 Z + a_2 Z^2 + \cdots + a_n Z^n) = a_1 + 2a_2 Z + \cdots + n a_n Z^{n-1}$, the degree of the polynomial $D(p(Z))$ will typically be one less than the degree of the polynomial $p(Z)$ itself. The only exception is if $p(Z) = a_0$, in which case $D(p(Z)) = 0$. Hence $D(p(Z)) = \lambda \cdot p(Z)$ can only hold for constant polynomials. If $p(Z)$ is a constant polynomial, then $p(Z) = a_0$ and $D(a_0) = 0 = 0 \cdot a_0$. This shows that 0 is the only eigenvalue that the linear map D has. Any polynomial $p(Z) = a_0$ with $a_0 \in \mathbb{C} \setminus \{0\}$ is an eigenvector of D with eigenvalue 0. The reason that the zero polynomial is not an eigenvector of D is that by definition eigenvectors must be nonzero. As this example shows, eigenvalues themselves can be zero.

The previous two examples, may suggest that a linear map always has at least one eigenvector, but this is not the case. Let us consider such an example.

|||| Example 11.5

The rotation map $R_{\pi/2} : \mathbb{R}^2 \to \mathbb{R}^2$ from Example 10.18 was defined by $R_{\pi/2}(v_1, v_2) = (-v_2, v_1)$.

Question: Does the linear map $R_{\pi/2} : \mathbb{R}^2 \to \mathbb{R}^2$ have any eigenvectors?

Answer: Let us first give an intuitive answer and after that one using the definitions more directly. What the linear map $R_{\pi/2}$ does geometrically, is to take a vector as input and return as output the vector rotated over $\pi/2$ radians against the clock. If a nonzero vector would be an eigenvector, that would mean that rotation over $\pi/2$ radians would output a scaling of the input vector. This is intuitively not possible, so what we expect is that the linear map $R_{\pi/2}$ has no eigenvectors at all.

Let us now proceed to prove this using the definitions. If the map $R_{\pi/2} : \mathbb{R}^2 \to \mathbb{R}^2$ does have eigenvectors, there exists $(v_1, v_2) \in \mathbb{R}^2 \setminus \{(0,0)\}$ and $\lambda \in \mathbb{R}$ such that $R_{\pi/2}(v_1, v_2) =$

$\lambda \cdot (v_1, v_2)$. Equivalently $(-v_2, v_1) = (\lambda \cdot v_1, \lambda \cdot v_2)$, which in turn can be rewritten as the two equations $-\lambda v_1 - v_2 = 0$ and $v_1 - \lambda v_2 = 0$. Formulated in matrix form, we would get the matrix equation

$$\begin{bmatrix} -\lambda & -1 \\ 1 & -\lambda \end{bmatrix} \cdot \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Adding λ times the second row to the first row, we obtain the equation $-(1+\lambda^2)v_2 = 0$. But this implies that $v_2 = 0$, since $\lambda^2 + 1$ is not zero for any $\lambda \in \mathbb{R}$. Then using the second row, we also see that $v_1 = 0$. We conclude that $(v_1, v_2) = (0, 0)$, but eigenvectors were not allowed to be the zero vector. Hence the linear map $R_{\pi/2} : \mathbb{R}^2 \to \mathbb{R}^2$ has no eigenvectors.

The procedure of determining the possible eigenvectors and eigenvalues in the previous examples was quite ad hoc. Fortunately, there is a procedure that always works in case V has a finite dimension. We will explain this procedure now, starting with eigenvalues of a square matrix.

|||| Theorem 11.6

Let $\mathbb{F}$ be a field and $\mathbf{A} \in \mathbb{F}^{n \times n}$ a square matrix. Then $\lambda \in \mathbb{F}$ is an eigenvalue of $\mathbf{A}$ if and only if $\det(\mathbf{A} - \lambda \cdot \mathbf{I}_n) = 0$, where $\mathbf{I}_n$ denotes the $n \times n$ identity matrix.

Proof. If $\lambda \in \mathbb{F}$ is an eigenvalue of the matrix $\mathbf{A}$, then there exists a nonzero vector $\mathbf{v} \in \mathbb{F}^n$ such that $\mathbf{A} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}$. Since $\lambda \cdot \mathbf{v} = \lambda \cdot (\mathbf{I}_n \cdot \mathbf{v}) = (\lambda \cdot \mathbf{I}_n) \cdot \mathbf{v}$, we see that the equation $\mathbf{A} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}$ can be rewritten as $\mathbf{A} \cdot \mathbf{v} = (\lambda \cdot \mathbf{I}_n)\mathbf{v}$, which in turn can be rewritten as $(\mathbf{A} - \lambda \cdot \mathbf{I}_n) \cdot \mathbf{v} = \mathbf{0}$. This shows that the homogeneous system of linear equations with coefficient matrix $\mathbf{A} - \lambda \cdot \mathbf{I}_n$ has a nonzero solution. Using Corollary 8.26 for the square matrix $\mathbf{A} - \lambda \cdot \mathbf{I}_n$, we conclude that $\det(\mathbf{A} - \lambda \cdot \mathbf{I}_n) = 0$.

Conversely, if $\det(\mathbf{A} - \lambda \cdot \mathbf{I}_n) = 0$, Corollary 8.26 implies that the homogeneous system of linear equations with coefficient matrix $\mathbf{A} - \lambda \cdot \mathbf{I}_n$ has a nonzero solution. Any such nonzero solution $\mathbf{v} \in \mathbb{F}^n$ then satisfies $(\mathbf{A} - \lambda \cdot \mathbf{I}_n) \cdot \mathbf{v} = \mathbf{0}$. This can be rewritten as $\mathbf{A} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}$. Hence $\mathbf{v}$ is an eigenvalue of $\mathbf{A}$ with eigenvalue λ. □

For a given square matrix $\mathbf{A} \in \mathbb{F}^{n \times n}$, the expression $\det(\mathbf{A} - Z \cdot \mathbf{I}_n)$ is a polynomial in $\mathbb{F}[Z]$ of degree n. This polynomial is called the *characteristic polynomial* of $\mathbf{A}$. We will denote it by $p_{\mathbf{A}}(Z)$. The roots of this polynomial in the field $\mathbb{F}$ are exactly all the eigenvalues of the matrix $\mathbf{A}$.

|||| Example 11.7

Theorem 11.6 makes it possible to describe all possible eigenvalues of a square matrix. For example, for the matrix

$$\mathbf{A} = \begin{bmatrix} -1 & 0 \\ 0 & 2 \end{bmatrix} \in \mathbb{R}^{2\times 2}$$

that we considered in Example 11.3, we have

$$p_{\mathbf{A}}(Z) = \det(\mathbf{A} - Z \cdot \mathbf{I}_2) = \det\left(\begin{bmatrix} -1-Z & 0 \\ 0 & 2-Z \end{bmatrix}\right) = (-1-Z)\cdot(2-Z) = (Z+1)\cdot(Z-2).$$

Therefore the roots of the characteristic polynomial $p_{\mathbf{A}}(Z)$ are precisely -1 and 2. This means that the eigenvalues of the matrix $\mathbf{A}$ are -1 and 2. Looking back at Example 11.3 we can make the answer given there a bit shorter, since the first case we considered is no longer needed. Indeed, in the first case, we considered all λ such that $\lambda \neq -1$ and $\lambda \neq 2$, but now we know already that there are no eigenvectors with such an eigenvalue λ.

|||| Example 11.8

As a second example, let us consider the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in \mathbb{R}^{2\times 2}.$$

We have seen in Example 10.34 that his matrix represents the linear map $R_{\pi/2} : \mathbb{R}^2 \to \mathbb{R}^2$ if the standard ordered basis is chosen for $\mathbb{R}^2$. In this case

$$p_{\mathbf{A}}(Z) = \det(\mathbf{A} - Z \cdot \mathbf{I}_2) = \det\left(\begin{bmatrix} -Z & -1 \\ 1 & -Z \end{bmatrix}\right) = Z^2 + 1.$$

Since we are working over the real numbers $\mathbb{R}$ and the polynomial $Z^2 + 1$ has no roots in $\mathbb{R}$, we conclude that the matrix $\mathbf{A}$, when studied over $\mathbb{R}$, has no eigenvalues and hence no eigenvectors.

Now that we know how to find eigenvalues of a square matrix, it is natural to ask how to find eigenvectors. We will return to that question in the next section. In the remainder of this section, we will explain how to find eigenvalues of an arbitrary linear map $L : V \to V$ in case V is a finite dimensional vector space.

Theorem 11.9

Let $\mathbb{F}$ be a field, V a vector space over $\mathbb{F}$ of dimension n and $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ an ordered basis of V. Then $\lambda \in \mathbb{F}$ is an eigenvalue of a linear map $L : V \to V$ if and only if $\det({}_\beta[L]_\beta - \lambda \cdot \mathbf{I}_n) = 0$.

Proof. If $\lambda \in \mathbb{F}$ is an eigenvalue of the linear map $L : V \to V$, then there exists a nonzero vector $\mathbf{v} \in V$ such that $L(\mathbf{v}) = \lambda \cdot \mathbf{v}$. Hence by the first item in Theorem 10.33, we have ${}_\beta[L]_\beta \cdot [\mathbf{v}]_\beta = [L(\mathbf{v})]_\beta = [\lambda \cdot \mathbf{v}]_\beta$. Applying Lemma 9.17, we have $[\lambda \cdot \mathbf{v}]_\beta = \lambda \cdot [\mathbf{v}]_\beta$. Combining these two equations, we obtain that ${}_\beta[L]_\beta \cdot [(\mathbf{v})]_\beta = \lambda \cdot [\mathbf{v}]_\beta$. Hence $[\mathbf{v}]_\beta$ is an eigenvector of the matrix ${}_\beta[L]_\beta$ with eigenvalue λ.

Conversely, suppose that $\det({}_\beta[L]_\beta - \lambda \cdot \mathbf{I}_n) = 0$ for some $\lambda \in \mathbb{F}$. Then by Theorem 11.6, λ is an eigenvalue of the matrix ${}_\beta[L]_\beta$. Hence there exists a nonzero vector $\mathbf{c} = (c_1, \ldots, c_n) \in \mathbb{F}^n$ that is an eigenvector of the matrix ${}_\beta[L]_\beta$ with eigenvalue λ. Now define $\mathbf{v} = c_1 \cdot \mathbf{v}_1 + \cdots + c_n \cdot \mathbf{v}_n \in V$. Then $\mathbf{c} = [\mathbf{v}]_\beta$. Hence we have ${}_\beta[L]_\beta \cdot [(\mathbf{v})]_\beta = \lambda \cdot [\mathbf{v}]_\beta$, which implies $[L(\mathbf{v})]_\beta = \lambda \cdot [\mathbf{v}]_\beta = [\lambda \cdot \mathbf{v}]_\beta$. This implies that $L(\mathbf{v}) = \lambda \cdot \mathbf{v}$. Hence λ is an eigenvalue of the linear map $L : V \to V$. $\square$

This theorem shows that if V is a finite dimensional vector space, we can reduce the calculation of eigenvalues of a linear map $L : V \to V$ directly to the calculation of the eigenvalues of the square matrix ${}_\beta[L]_\beta$ representing the linear map. Here it does not matter at all, which ordered basis of V one chooses. For future use, let us nonetheless investigate the effect of choosing another ordered basis on the matrix representing L. Here the change of coordinate matrices introduced in equation (10-5) will play an important role.

Lemma 11.10

Let $\mathbb{F}$ be a field, V a vector space over $\mathbb{F}$ of dimension n, and $L : V \to V$ a linear map. Further let β and γ be two ordered bases of V and denote by $\mathrm{id}_V : V \to V$ the identity map $\mathbf{v} \mapsto \mathbf{v}$. Then

$$ {}_\gamma[L]_\gamma = ({}_\beta[\mathrm{id}_V]_\gamma)^{-1} \cdot {}_\beta[L]_\beta \cdot {}_\beta[\mathrm{id}_V]_\gamma. $$

Proof. We know from Lemma 10.37 that $({}_\beta[\mathrm{id}_V]_\gamma)^{-1} = {}_\gamma[\mathrm{id}_V]_\beta$. Hence

$$\begin{aligned}({}_\beta[\mathrm{id}_V]_\gamma)^{-1} \cdot {}_\beta[L]_\beta \cdot {}_\beta[\mathrm{id}_V]_\gamma &= {}_\gamma[\mathrm{id}_V]_\beta \cdot {}_\beta[L]_\beta \cdot {}_\beta[\mathrm{id}_V]_\gamma \\ &= {}_\gamma[\mathrm{id}_V]_\beta \cdot {}_\beta[L \circ \mathrm{id}_V]_\gamma \\ &= {}_\gamma[\mathrm{id}_V \circ L \circ \mathrm{id}_V]_\gamma \\ &= {}_\gamma[L]_\gamma.\end{aligned}$$

In the second and third equality, we used the first item of Theorem 10.33. ☐

Two square matrices $\mathbf{A} \in \mathbb{F}^{n\times n}$ and $\mathbf{B} \in \mathbb{F}^{n\times n}$ are called *similar* if there exists an invertible matrix $\mathbf{Q} \in \mathbb{F}^{n\times n}$ such that $\mathbf{A} = \mathbf{Q}^{-1} \cdot \mathbf{B} \cdot \mathbf{Q}$. Hence Lemma 11.10 can be rephrased in words as follows: the effect of choosing a different ordered basis of V is that the matrix representing L is replaced by a similar matrix. It turns out that this lemma also explains why it does not matter which ordered basis one chooses when computing the eigenvalues of a linear map. In fact, we have the following:

|||| Theorem 11.11

Let $\mathbb{F}$ be a field, V a vector space over $\mathbb{F}$ of dimension n, and $L : V \to V$ a linear map. Further let β and γ be two ordered bases of V. Then the characteristic polynomials of ${}_\beta[L]_\beta$ and ${}_\gamma[L]_\gamma$ are identical.

Proof. For convenience, let us write $\mathbf{Q} = {}_\beta[\mathrm{id}_V]_\gamma$. Using Lemma 11.10, we see that:

$$\begin{aligned}p_{{}_\gamma[L]_\gamma}(Z) &= \det({}_\gamma[L]_\gamma - Z \cdot \mathbf{I}_n) \\ &= \det(\mathbf{Q}^{-1} \cdot {}_\beta[L]_\beta \cdot \mathbf{Q} - Z \cdot \mathbf{I}_n) \\ &= \det(\mathbf{Q}^{-1} \cdot {}_\beta[L]_\beta \cdot \mathbf{Q} - Z \cdot \mathbf{Q}^{-1} \cdot \mathbf{Q}) \\ &= \det(\mathbf{Q}^{-1} \cdot {}_\beta[L]_\beta \cdot \mathbf{Q} - Z \cdot \mathbf{Q}^{-1} \cdot \mathbf{I}_n \cdot \mathbf{Q}) \\ &= \det(\mathbf{Q}^{-1} \cdot ({}_\beta[L]_\beta - Z \cdot \mathbf{I}_n) \cdot \mathbf{Q}).\end{aligned}$$

At this point Theorem 8.23 comes in handy. Using this theorem, we can namely continue as follows:

$$\begin{aligned}
p_{\gamma[L]_\gamma}(Z) &= \det(\mathbf{Q}^{-1} \cdot ({}_\beta[L]_\beta - Z \cdot \mathbf{I}_n) \cdot \mathbf{Q}) \\
&= \det(\mathbf{Q}^{-1}) \cdot \det({}_\beta[L]_\beta - Z \cdot \mathbf{I}_n) \cdot \det(\mathbf{Q}) \\
&= \det(\mathbf{Q}^{-1}) \cdot \det(\mathbf{Q}) \cdot \det({}_\beta[L]_\beta - Z \cdot \mathbf{I}_n) \\
&= \det(\mathbf{Q})^{-1} \cdot \det(\mathbf{Q}) \cdot \det({}_\beta[L]_\beta - Z \cdot \mathbf{I}_n) \\
&= 1 \cdot \det({}_\beta[L]_\beta - Z \cdot \mathbf{I}_n) \\
&= \det({}_\beta[L]_\beta - Z \cdot \mathbf{I}_n) \\
&= p_{\beta[L]_\beta}(Z).
\end{aligned}$$

This is exactly what we wanted to show. □

|||| Corollary 11.12

With the same notation as before, we have $\det {}_\beta[L]_\beta = \det {}_\gamma[L]_\gamma$.

Proof. This follows by putting $Z = 0$ in the characteristic polynomials $p_{\beta[L]_\beta}(Z)$ and $p_{\gamma[L]_\gamma}(Z)$. □

We can now define the characteristic polynomial of a linear map $L : V \to V$ as long as V is a finite dimensional vector space.

|||| Definition 11.13

Let $\mathbb{F}$ be a field, V a vector space over $\mathbb{F}$ of finite dimension n, and $L : V \to V$ a linear map. Then the *characteristic polynomial* is defined to be that polynomial $p_L(Z) = \det({}_\beta[L]_\beta - Z \cdot \mathbf{I}_n) \in \mathbb{F}[Z]$, where β is some ordered basis of V.

The reason this definition makes sense, is that by Theorem 11.11, the choice of the ordered basis β does not matter: a different choice will not change the corresponding characteristic polynomial. In a similar way, based on Corollary 11.12, one can define the determinant of such a linear map: $\det L = \det {}_\beta[L]_\beta$.

|||| Example 11.14

As an example, we will consider a linear map, similar to the linear map $D : \mathbb{C}[Z] \to \mathbb{C}[Z]$ from Example 10.15. However, since $\mathbb{C}[Z]$ is an infinitely dimensional vector space, we will modify the domain and codomain of the map a bit. More precisely, let V be the complex vector space of polynomials of degree at most three. Then we can define $\tilde{D} : V \to V$ as $p(Z) \mapsto p(Z)'$.

Question: What is the characteristic polynomial $p_{\tilde{D}}(\lambda)$ of the linear map $\tilde{D}$?

Answer: Let us choose the ordered basis $\beta = (1, Z, Z^2, Z^3)$ of V. Since $\tilde{D}(1) = 0, \tilde{D}(Z) = 1, \tilde{D}(Z^2) = 2Z$ and $\tilde{D}(Z^3) = 3Z^2$, we see that

$$_\beta[\tilde{D}]_\beta = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and therefore} \quad _\beta[\tilde{D}]_\beta - Z \cdot \mathbf{I}_4 = \begin{bmatrix} -Z & 1 & 0 & 0 \\ 0 & -Z & 2 & 0 \\ 0 & 0 & -Z & 3 \\ 0 & 0 & 0 & -Z \end{bmatrix}.$$

We see that $_\beta[\tilde{D}]_\beta - Z \cdot \mathbf{I}_4$ is an upper triangular matrix (see Definition 8.7). This means that its determinant is simply the product of the elements on its diagonal, see Theorem 8.8. Therefore the characteristic polynomial of $\tilde{D}$ is $p_{\tilde{D}}(Z) = (-Z)^4 = Z^4$.

11.2 Eigenspaces

So far, we have focused mainly on how to find the eigenvalues of a matrix and a linear map. In this section, we will focus on finding all possible eigenvectors for a given eigenvalue.

|||| Theorem 11.15

Let $\mathbb{F}$ be a field, V a vector space over $\mathbb{F}$ of finite dimension n, and $L : V \to V$ a linear map. Suppose that $\lambda \in \mathbb{F}$ is an eigenvalue of L. Then the set

$$E_\lambda = \{\mathbf{v} \in V \mid L(\mathbf{v}) = \lambda \cdot \mathbf{v}\}$$

is a subspace of V.

Proof. Let $\mathbf{u}, \mathbf{v} \in E_\lambda$ and $c \in \mathbb{F}$. According to Lemma 9.32, we can conclude that E_λ is a subspace of V, if we can show that $\mathbf{u} + c \cdot \mathbf{v} \in E_\lambda$. Now note that

$$L(\mathbf{u} + c \cdot \mathbf{v}) = L(\mathbf{u}) + c \cdot L(\mathbf{v}) = \lambda \cdot \mathbf{u} + c \cdot \lambda \cdot \mathbf{v} = \lambda \cdot (\mathbf{u} + c \cdot \mathbf{v}).$$

Hence indeed $\mathbf{u} + c \cdot \mathbf{v} \in E_\lambda$, which is what we needed to show. □

For square matrices, this theorem has a direct consequence.

|||| Corollary 11.16

Let $\mathbb{F}$ be a field and $\mathbf{A} \in \mathbb{F}^{n \times n}$ a square matrix. Suppose that $\lambda \in \mathbb{F}$ is an eigenvalue of $\mathbf{A}$. Then the set $E_\lambda = \{\mathbf{v} \in V \mid \mathbf{A} \cdot \mathbf{v} = \lambda \cdot \mathbf{v}\}$ is a subspace of $\mathbb{F}^n$.

Proof. This follows from Theorem 11.15 by applying it to the linear map $L_\mathbf{A} : \mathbb{F}^n \to \mathbb{F}^n$, $\mathbf{v} \mapsto \mathbf{A} \cdot \mathbf{v}$. □

For a given linear map $L : V \to V$ for a finite dimensional vector space V and an eigenvalue λ of L, the subspace E_λ is called the *eigenspace corresponding to the eigenvalue* λ of the linear map L. Similarly, for a given square matrix $\mathbf{A} \in \mathbb{F}^{n \times n}$, the subspace E_λ is called the eigenspace corresponding to the eigenvalue λ of the matrix $\mathbf{A}$.

Now that we know that the set of all eigenvectors of a given eigenvalue λ together with the zero vector, forms a subspace E_λ, we can describe all possible eigenvectors for a given eigenvalue by giving a basis of this subspace E_λ. Fortunately, this turns out to be yet another application of the theory of systems of linear equations. First of all, we have:

|||| Lemma 11.17

Let $L : V \to V$ be a linear map of vector spaces over a field $\mathbb{F}$ and assume that $\dim V = n$. Suppose that $\lambda \in \mathbb{F}$ is an eigenvalue of L. Then $E_\lambda = \ker(L - \lambda \cdot \mathrm{id}_V)$. Similarly, if $\mathbf{A} \in \mathbb{F}^{n \times n}$ is a matrix and $\lambda \in \mathbb{F}$ is an eigenvalue of $\mathbf{A}$, then $E_\lambda = \ker(\mathbf{A} - \lambda \cdot \mathbf{I}_n)$.

Proof. By definition, we have $\mathbf{v} \in E_\lambda$ if and only if $L(\mathbf{v}) = \lambda \cdot \mathbf{v}$. Note that $L(\mathbf{v}) = \lambda \cdot \mathbf{v}$ if and only if $(L - \lambda \cdot \mathrm{id}_n)(\mathbf{v}) = \mathbf{0}$, which in turn is equivalent to saying that $\mathbf{v} \in \ker(L - \lambda \cdot \mathbf{I}_n)$. The second part of the lemma involving the matrix $\mathbf{A}$ can be proved similarly. $\square$

As we have observed before, computing vectors in the kernel of some matrix $\mathbf{B}$, is exactly the same as finding solutions to the homogeneous system of linear equations with coefficient matrix $\mathbf{B}$. Moreover, we already know how to compute a basis for the solution space of a homogeneous system of linear equations using Corollary 9.38 and Theorem 6.29. Hence, we do not need to develop new tools when computing a basis for the eigenspace E_λ of a matrix. Also when dealing with the similar problem for linear maps, we do not need any new tools: Theorem 10.39 implies that we can compute the kernel of a linear map $L : V \to V$ by computing the kernel of a matrix ${}_\beta[L]_\beta$ representing the linear map, where β is an ordered basis of V. This settles the computation of eigenvectors completely. Let us illustrate this in two examples.

|||| Example 11.18

First, let us consider the matrix

$$\mathbf{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \in \mathbb{C}^{2\times 2}.$$

We have encountered this matrix before in Example 11.8, but there is one important difference: in this example we work over the complex numbers $\mathbb{C}$. This was indicated by introducing the matrix as an element in $\mathbb{C}^{2\times 2}$, rather than as element of $\mathbb{R}^{2\times 2}$. First of all, we have, just as in Example 11.8, that

$$p_{\mathbf{A}}(Z) = \det(\mathbf{A} - Z \cdot \mathbf{I}_2) = \det\left(\begin{bmatrix} -Z & -1 \\ 1 & -Z \end{bmatrix}\right) = Z^2 + 1.$$

Since we are working over the field $\mathbb{C}$, the polynomial $Z^2 + 1$ has two roots namely i and $-i$.

Question: Find a basis for the eigenspace E_i.

Answer: We know from Lemma 11.17 that $E_i = \ker(\mathbf{A} - i \cdot \mathbf{I}_2)$. We have

$$\mathbf{A} - i \cdot \mathbf{I}_2 = \begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix}.$$

To compute the kernel of this matrix, we bring it in reduced row echelon form:

$$\begin{bmatrix} -i & -1 \\ 1 & -i \end{bmatrix} \underset{R_2 \leftarrow R_2 - i \cdot R_1}{\longrightarrow} \begin{bmatrix} -i & -1 \\ 0 & 0 \end{bmatrix} \underset{R_1 \leftarrow i \cdot R_1}{\longrightarrow} \begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix}.$$

This means that $\mathbf{v} = (v_1, v_2) \in \ker(\mathbf{A} - i \cdot \mathbf{I}_2)$ if and only if $v_1 = i \cdot v_2$. Hence:

$$E_i = \ker(\mathbf{A} - i \cdot \mathbf{I}_2) = \left\{ c \cdot \begin{bmatrix} i \\ 1 \end{bmatrix} \mid c \in \mathbb{C} \right\}.$$

A basis of E_i is therefore given by

$$\left\{ \begin{bmatrix} i \\ 1 \end{bmatrix} \right\}.$$

This completely answers the question. In a similar way, one can show that a basis of E_{-i} is given by

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

|||| Example 11.19

Let us revisit the linear map $\tilde{D} : V \to V$ introduced in Example 11.14. In that example V was the complex vector space of polynomials of degree at most three and $\tilde{D} : V \to V$ was defined by $p(Z) \mapsto p(Z)'$. We have already seen in Example 11.14 that $p_{\tilde{D}}(Z) = Z^4$. Hence $\tilde{D}$ has only one eigenvalue, namely 0.

Question: Compute a basis for the eigenspace E_0.

Answer: We know by Lemma 11.17 that $E_0 = \ker(\tilde{D} - 0 \cdot \mathrm{id}_V) = \ker \tilde{D}$. In order to compute a basis of $\ker \tilde{D}$, we first compute the kernel of a matrix ${}_\beta[\tilde{D}]_\beta \in \mathbb{C}^{4\times 4}$ representing $\tilde{D}$. Let us choose the ordered basis $\beta = (1, Z, Z^2, Z^3)$ of V. We have already seen in Example 11.14 that in that case:

$${}_\beta[\tilde{D}]_\beta = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

This matrix is already in echelon form and we can directly see that $(v_1, v_2, v_3, v_4) \in \ker {}_\beta[\tilde{D}]_\beta$ if and only if $v_2 = 0$ and $v_3 = 0$ and $v_4 = 0$. Therefore,

$$\ker {}_\beta[\tilde{D}]_\beta = \left\{ c \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \mid c \in \mathbb{C} \right\}.$$

We see that a basis for $\ker{}_\beta[\tilde{D}]_\beta$ is given by

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

The basis vector $(1,0,0,0)$ corresponds to the polynomial $1 \cdot 1 + 0 \cdot Z + 0 \cdot Z^2 + 0 \cdot Z^3 = 1$. Hence using Theorem 10.39, we see that

$$E_0 = \ker \tilde{D} = \{c \cdot 1 \mid c \in \mathbb{C}\} = \mathbb{C}$$

and that a basis of E_0 is given by $\{1\}$.

Let us finish this section with a theoretical consideration about eigenvectors that will become very important later on. We start with a definition.

|||| Definition 11.20

Let $\mathbb{F}$ be a field, V a finite dimensional vector space and $L : V \to V$ be a linear map. Suppose that $\lambda \in \mathbb{F}$ is an eigenvalue of L. Then we define the *algebraic multiplicity* $\mathrm{am}(\lambda)$ of the eigenvalue λ to be the multiplicity of λ as root in the characteristic polynomial $p_L(Z)$ of L. Further, we define the *geometric multiplicity* $\mathrm{gm}(\lambda)$ of the eigenvalue λ to be the dimension of E_λ.
Similarly for an eigenvalue $\lambda \in \mathbb{F}$ of a square matrix $\mathbf{A} \in \mathbb{F}^{n \times n}$, we define $\mathrm{am}(\lambda)$ to be the multiplicity of λ as root in the characteristic polynomial $p_{\mathbf{A}}(Z)$ of $\mathbf{A}$ and $\mathrm{gm}(\lambda) = \dim E_\lambda$.

|||| Example 11.21

In Example 11.18, the eigenvalue i is a root of multiplicity 1 in the characteristic polynomial $p_{\mathbf{A}}(Z) = Z^2 + 1$. Hence $\mathrm{am}(i) = 1$. In that example, we also saw that E_i is a vector space of dimension one. Hence $\mathrm{gm}(i) = 1$.

|||| Example 11.22

In Example 11.19, the eigenvalue 0 is a root of multiplicity 4 in the characteristic polynomial $p_{\tilde{D}}(Z) = Z^4$. Hence $\mathrm{am}(0) = 4$. In that example, we also saw that E_0 is a vector space of dimension one. Hence $\mathrm{gm}(0) = 1$ in this case.

As the last example shows, the algebraic and the geometric multiplicity of an eigenvalue need not be the same. We do have the following theorem stating that $1 \leq \mathrm{gm}(\lambda) \leq \mathrm{am}(\lambda)$. A reader willing to accept this statement can continue to the next section.

|||| Theorem 11.23

Let $\mathbb{F}$ be a field and $\lambda \in \mathbb{F}$ an eigenvalue of a linear map $L : V \to V$, with $\dim V = n < \infty$, or an eigenvalue of a square matrix $\mathbf{A} \in \mathbb{F}^{n \times n}$. Then $1 \leq \mathrm{gm}(\lambda) \leq \mathrm{am}(\lambda) \leq n$.

Proof. First of all, if λ is an eigenvalue, there by definition exists at least one eigenvector. Hence $\mathrm{gm}(\lambda) = \dim E_\lambda \geq 1$.

Now suppose that λ is an eigenvalue and let us write $s = \mathrm{gm}(\lambda)$ for convenience. We will prove the theorem in case λ is an eigenvalue of a linear map $L : V \to V$ only, since the case of a matrix $\mathbf{A}$ follows by considering the linear map $L_{\mathbf{A}} : \mathbb{F}^n \to \mathbb{F}^n$. Since $\dim E_\lambda = \mathrm{gm}(\lambda) = s$, any basis of E_λ contains precisely s vectors. Let us choose such a basis, say $\{\mathbf{v}_1, \ldots, \mathbf{v}_s\}$. Now choose vectors $\mathbf{v}_{s+1}, \ldots, \mathbf{v}_n \in V$ such that $\beta = \mathbf{v}_1, \ldots, \mathbf{v}_s, \mathbf{v}_{s+1}, \ldots, \mathbf{v}_n$ is an ordered basis of V. Since $L(\mathbf{v}_i) = \lambda \cdot \mathbf{v}_i$ for all i between 1 and s, we have

$$_\beta[L]_\beta = \begin{bmatrix} \lambda \cdot \mathbf{I}_s & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix},$$

for some matrices $\mathbf{B} \in \mathbb{F}^{s \times (n-s)}$ and $\mathbf{D} \in \mathbb{F}^{(n-s) \times (n-s)}$ and where $\mathbf{0}$ denotes the $(n-s) \times s$ matrix all of whose coefficients are zero. Then

$$_\beta[L]_\beta - Z \cdot \mathbf{I}_n = \begin{bmatrix} (\lambda - Z) \cdot \mathbf{I}_s & \mathbf{B} \\ \mathbf{0} & \mathbf{D} - Z \cdot \mathbf{I}_{n-s} \end{bmatrix}$$

and hence

$$\begin{aligned} p_L(Z) &= \det({}_\beta[L]_\beta - Z \cdot \mathbf{I}_n) \\ &= \det\left(\begin{bmatrix} (\lambda - Z) \cdot \mathbf{I}_s & \mathbf{B} \\ \mathbf{0} & \mathbf{D} - Z \cdot \mathbf{I}_{n-s} \end{bmatrix}\right) \\ &= (\lambda - Z)^s \cdot \det(\mathbf{D} - Z \cdot \mathbf{I}_{n-s}). \end{aligned}$$

In the last equality, we used induction on s and developed the determinant in the first column to prove the induction basis as well as to perform the induction step. Now it is clear that the multiplicity of λ in $p_L(Z)$ is at least s. In other words: $\mathrm{am}(\lambda) \geq s = \mathrm{gm}(\lambda)$, which is exactly what we wanted to show. The final inequality $\mathrm{am}(\lambda) \leq n$ follows, since $\mathrm{am}(\lambda)$ is the multiplicity of the root λ in the polynomial $p_L(Z)$ and $\deg p_L(Z) = n$. $\square$

11.3 Diagonalization

In this section, we describe when a linear map can be represented by a particularly nice matrix: a diagonal matrix. In other words: we will describe when a linear map has a diagonal mapping matrix. To achieve this, we need to be able to choose a particularly nice ordered basis. Therefore we start with a lemma.

|||| Lemma 11.24

Let $\mathbb{F}$ be a field, V a finite dimensional vector space over $\mathbb{F}$ and $L : V \to V$ a linear map. Further, suppose that $\lambda_1, \ldots, \lambda_r \in \mathbb{F}$ are distinct eigenvalues of L and write $d_i = \mathrm{gm}(\lambda_i)$ for $i = 1, \ldots, r$. If $(\mathbf{v}_1^{(i)}, \ldots, \mathbf{v}_{d_i}^{(i)})$ for $i = 1, \ldots, r$ are ordered bases of E_{λ_i}, then the vectors

$$\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_{d_1}^{(1)}, \ldots, \mathbf{v}_1^{(r)}, \ldots, \mathbf{v}_{d_r}^{(r)}$$

are linearly independent.

Proof. We will prove the lemma using induction on r.

If $r = 1$, there is nothing to prove, since we assume that $(\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_{d_1}^{(1)})$ is an ordered basis of E_{λ_1}. Then the vectors $\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_{d_1}^{(1)}$ are certainly linearly independent. Now let $r > 1$ and assume as induction hypothesis that the lemma is correct if there are $r - 1$

distinct eigenvalues. Suppose that

$$\sum_{i=1}^{r} \sum_{j=1}^{d_i} \alpha_{i,j} \cdot \mathbf{v}_j^{(i)} = \mathbf{0}, \tag{11-2}$$

for certain $\alpha_{i,j} \in \mathbb{F}$. We need to show that $\alpha_{i,j} = 0$ for all $i = 1, \ldots, r$ and $j = 1, \ldots, d_i$. Applying the linear map L to this equation and using that $L(\mathbf{v}_j^{(i)}) = \lambda_i \cdot \mathbf{v}_j^{(i)}$, we see that $\sum_{i=1}^{r} \sum_{j=1}^{d_i} \alpha_{i,j} \cdot \lambda_i \cdot \mathbf{v}_j^{(i)} = \mathbf{0}$, which can be rewritten as

$$\sum_{i=1}^{r} \lambda_i \cdot \sum_{j=1}^{d_i} \alpha_{i,j} \cdot \mathbf{v}_j^{(i)} = \mathbf{0}. \tag{11-3}$$

Multiplying equation (11-2) with λ_r and subtracting equation (11-3) from the result, the term corresponding to $i = r$ cancels, while the result still equals $\mathbf{0}$. In other words:

$$\lambda_r \cdot \sum_{i=1}^{r-1} \sum_{j=1}^{d_i} \alpha_{i,j} \cdot \mathbf{v}_j^{(i)} - \sum_{i=1}^{r-1} \lambda_i \cdot \sum_{j=1}^{d_i} \alpha_{i,j} \cdot \mathbf{v}_j^{(i)} = \lambda_r \cdot \sum_{i=1}^{r} \sum_{j=1}^{d_i} \alpha_{i,j} \cdot \mathbf{v}_j^{(i)} - \sum_{i=1}^{r} \lambda_i \cdot \sum_{j=1}^{d_i} \alpha_{i,j} \cdot \mathbf{v}_j^{(i)} = \mathbf{0}.$$

Combining the first two sums into one, we obtain:

$$\sum_{i=1}^{r-1} \sum_{j=1}^{d_i} (\lambda_r - \lambda_i) \cdot \alpha_{i,j} \cdot \mathbf{v}_j^{(i)} = \sum_{i=1}^{r-1} (\lambda_r - \lambda_i) \cdot \sum_{j=1}^{d_i} \alpha_{i,j} \cdot \mathbf{v}_j^{(i)} = \mathbf{0}.$$

Now we can apply the induction hypothesis and conclude that $(\lambda_r - \lambda_i) \cdot \alpha_{i,j} = 0$ for $i = 1, \ldots, r-1$ and $j = 1, \ldots, d_i$. Since all eigenvalues were assumed to be distinct, we see that $\lambda_r - \lambda_i \neq 0$ for all i between 1 and $r-1$. Hence $\alpha_{i,j} = 0$ for $i = 1, \ldots, r-1$ and $j = 1, \ldots, d_i$. Substituting this in equation (11-2), we obtain that $\sum_{j=1}^{d_r} \alpha_{r,j} \cdot \mathbf{v}_j^{(r)} = \mathbf{0}$, but then we may also conclude that $\alpha_{r,j} = 0$ for $j = 1, \ldots d_r$, since we assumed that $(\mathbf{v}_1^{(r)}, \ldots, \mathbf{v}_{d_r}^{(r)})$ is an ordered basis of E_{λ_r}. This completes the induction step. Hence by the induction principle, we may conclude that the lemma holds for all r. $\square$

As we have seen before, in order to be able to represent a linear map $L : V \to V$ by a matrix ${}_\beta[L]_\beta$, we need to choose an ordered basis β of V. The vectors in Lemma 11.24 are linearly independent, which is a good start, but may not span the entire space V. The next lemma clarifies when the eigenvectors span V.

|||| **Lemma 11.25**

Let $\mathbb{F}$ be a field, V a vector space over $\mathbb{F}$ of dimension n, and $L : V \to V$ a linear map. Then the following two items are equivalent:

1. The eigenvectors of L span V.
2. The characteristic polynomial of L is of the form

$$p_L(Z) = (-1)^n \cdot (Z - \lambda_1)^{m_1} \cdots (Z - \lambda_r)^{m_r}$$

for certain $\lambda_1, \ldots, \lambda_r \in \mathbb{F}$ and positive integers $m_1, \ldots, m_r$. Moreover, for each eigenvalue λ_i its algebraic and geometric multiplicity is the same: $\mathrm{am}(\lambda_i) = \mathrm{gm}(\lambda_i)$ for $i = 1, \ldots, r$.

Proof. To show that the two items are logically equivalent, we first show 1. $\Rightarrow$ 2 and afterwards 2. $\Rightarrow$ 1.

1. $\Rightarrow$ 2.: assume that the eigenvectors of L span V. Then we can find a basis S of V consisting of eigenvectors only. Let $\lambda_1, \ldots, \lambda_r \in \mathbb{F}$ be the eigenvalues of L and order the eigenvectors in S such that the eigenvectors with eigenvalue λ_1 come first, then those with eigenvalue λ_2, and so on, ending with the eigenvectors in S with eigenvalue λ_r. We then have constructed an ordered basis

$$\beta = (\mathbf{v}_1^{(1)}, \ldots, \mathbf{v}_{n_1}^{(1)}, \ldots, \mathbf{v}_1^{(r)}, \ldots, \mathbf{v}_{n_r}^{(r)}),$$

where for $i = 1, \ldots, r$, the vectors $\mathbf{v}_1^{(i)}, \ldots, \mathbf{v}_{n_i}^{(i)}$ are the eigenvectors in S with eigenvalue λ_i.

Now on the one hand, we have $n_1 + n_2 + \cdots + n_r = n$, since the number of vectors in the ordered basis β is the same as the dimension of V. On the other hand, for all i, we have $n_i \leq \mathrm{gm}(\lambda_i)$, since $\mathbf{v}_1^{(i)}, \ldots, \mathbf{v}_{n_i}^{(i)}$ are linearly independent vectors in E_{λ_i} and $\dim E_{\lambda_i} = \mathrm{gm}(\lambda_i)$. Therefore, we have:

$$\begin{aligned} n &= n_1 + \cdots + n_r \\ &\leq \mathrm{gm}(\lambda_1) + \cdots + \mathrm{gm}(\lambda_r) \\ &\leq \mathrm{am}(\lambda_1) + \cdots + \mathrm{am}(\lambda_r) \\ &\leq \deg p_L(Z) \\ &= n. \end{aligned}$$

Since we both started and ended with n, all inequalities have to be equalities. This shows that $\mathrm{gm}(\lambda_i) = \mathrm{am}(\lambda_i)$ for all $i = 1, \ldots, r$ and that $p_L(Z)$ is of the form as stated in item 2.

2. $\Rightarrow$ 1.: Now assume that $p_L(Z) = (-1)^n \cdot (Z - \lambda_1)^{m_1} \cdots (Z - \lambda_r)^{m_r}$ for certain distinct $\lambda_1, \ldots, \lambda_r \in \mathbb{F}$ and positive integers $m_1, \ldots, m_r$ and that $\mathrm{am}(\lambda_i) = \mathrm{gm}(\lambda_i)$ for all $i = 1, \ldots, r$. Note that by definition, we have $m_i = \mathrm{am}(\lambda_i)$, which in turn implies that $m_i = \mathrm{gm}(\lambda_i)$, since we assume that $\mathrm{am}(\lambda_i) = \mathrm{gm}(\lambda_i)$ for all i. We conclude that $n = \deg p_L(Z) = \mathrm{gm}(\lambda_1) + \cdots + \mathrm{gm}(\lambda_r)$. On the other hand, by Lemma 11.24, we can find precisely $\mathrm{gm}(\lambda_1) + \cdots + \mathrm{gm}(\lambda_r)$ linearly independent eigenvectors of L. Combining these statements, we can conclude that we can find an ordered basis of V consisting of eigenvectors. In particular, the eigenvectors span V, which is what we wanted to show. □

Now we are ready to show the main result of this section.

|||| Definition 11.26

Let a linear map $L : V \to V$ be given, where V is a finite dimensional vector space over a field $\mathbb{F}$. Then one says that L can be diagonalized, if there exists an ordered basis β of V such that the corresponding mapping matrix ${}_\beta[L]_\beta$ is a diagonal matrix. Likewise, if $\mathbf{A} \in \mathbb{F}^{n \times n}$ is a square matrix, then one says that $\mathbf{A}$ can be diagonalized, if $\mathbf{A}$ is similar to a diagonal matrix.

|||| Theorem 11.27

Let V a finite dimensional vector space over a field $\mathbb{F}$. A linear map $L : V \to V$ can be diagonalized if and only if the characteristic polynomial of L is of the form $p_L(Z) = (-1)^n \cdot (Z - \lambda_1)^{m_1} \cdots (Z - \lambda_r)^{m_r}$ for certain $\lambda_1, \ldots, \lambda_r \in \mathbb{F}$, and $\mathrm{am}(\lambda_i) = \mathrm{gm}(\lambda_i)$ for each eigenvalue λ_i.

Proof. Using Lemma 11.25, it is enough to show that L can be diagonalized if and only if its eigenvectors span V.

Therefore, first assume that L can be diagonalized. Then there exists an ordered basis

β of V such that ${}_\beta[L]_\beta$ is a diagonal matrix. But this implies that each vector in β is an eigenvector. Hence V can be spanned by eigenvectors.

Conversely, assume that V can be spanned by eigenvectors. Then there exists an ordered basis $\beta = \mathbf{v}_1, \ldots, \mathbf{v}_n$ of V, containing eigenvectors only. The corresponding matrix ${}_\beta[L]_\beta$ is a diagonal matrix, with the eigenvalues of $\mathbf{v}_1, \ldots, \mathbf{v}_n$ on its diagonal. $\square$

|||| Corollary 11.28

A matrix $\mathbf{A} \in \mathbb{F}^{n\times n}$ can be diagonalized if and only if the characteristic polynomial of $\mathbf{A}$ is of the form $p_\mathbf{A}(Z) = (-1)^n \cdot (Z-\lambda_1)^{m_1} \cdots (Z-\lambda_r)^{m_r}$ for certain $\lambda_1, \ldots, \lambda_r \in \mathbb{F}$, and $\mathrm{am}(\lambda_i) = \mathrm{gm}(\lambda_i)$ for each eigenvalue λ_i.

Proof. This follows from Theorem 11.27 by applying it to the linear map $L_\mathbf{A} : \mathbb{F}^n \to \mathbb{F}^n$. $\square$

|||| Corollary 11.29

Let V be a finite dimensional complex vector space. A linear map $L : V \to V$, can be diagonalized if and only if $\mathrm{am}(\lambda_i) = \mathrm{gm}(\lambda_i)$ for each eigenvalue λ_i of L. Similarly, a complex matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is similar to a diagonal matrix if and only if $\mathrm{am}(\lambda_i) = \mathrm{gm}(\lambda_i)$ for each eigenvalue λ_i of $\mathbf{A}$.

Proof. If the field we work over is $\mathbb{C}$, it follows from Theorem 4.23 that the characteristic polynomial $p_L(Z)$ can be written as a product of its leading term and terms of the form $Z-\lambda$. Hence this condition in Theorem 11.27 is always satisfied if $\mathbb{F} = \mathbb{C}$ and can therefore be removed. Theorem 11.27 then implies what we want. The proof of the corollary in the case of a complex matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ is similar. $\square$

|||| Example 11.30

Consider the linear map $\tilde{D} : V \to V$ introduced in Example 11.14. In that example V was the complex vector space of polynomials of degree at most three and $\tilde{D} : V \to V$ was defined by $p(Z) \mapsto p(Z)'$.

Question: Can the linear map $\tilde{D}$ be diagonalized?

Answer: From Example 11.14, we see that $p_{\tilde{D}}(Z) = Z^4$. Using the notation of Theorem 11.27, we see that $r = 1$ and $\lambda_1 = 0$. Moreover, $\mathrm{am}(0) = 4$, since 0 is a root with multiplicity four of $p_{\tilde{D}}(Z)$. In Example 11.19, we have seen that E_0 is a one dimensional vector space with basis $\{1\}$. Hence $\mathrm{gm}(0) = \dim E_0 = 1$. Since $\mathrm{gm}(0) < \mathrm{am}(0)$ Theorem 11.27 implies that the linear map $\tilde{D}$ cannot be diagonalized.

|||| Example 11.31

Consider the matrix

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

that we also studied in Example 11.18 and that also occurred in Example 11.8.

Question 1: Can the matrix $\mathbf{A}$ be diagonalized when working over the real numbers $\mathbb{R}$? If yes, compute a matrix $\mathbf{Q} \in \mathbb{R}^{2\times 2}$ such that $\mathbf{Q}^{-1} \cdot \mathbf{A} \cdot \mathbf{Q}$ is a diagonal matrix.

Question 2: Can the matrix $\mathbf{A}$ be diagonalized when working over the complex numbers $\mathbb{C}$? If yes, compute a matrix $\mathbf{Q} \in \mathbb{C}^{2\times 2}$ such that $\mathbf{Q}^{-1} \cdot \mathbf{A} \cdot \mathbf{Q}$ is a diagonal matrix.

Answer to Question 1: We have computed in Example 11.8, that $p_{\mathbf{A}}(Z) = Z^2 + 1$. Since $Z^2 + 1$ has no real roots, it cannot be written in the form as required in Corollary 11.28. Therefore, the matrix $\mathbf{A}$ is not diagonalizable over $\mathbb{R}$.

Answer to Question 2: The characteristic polynomial $p_{\mathbf{A}}(Z) = Z^2 + 1$ has two complex roots, namely i and $-i$. Furthermore, we have $Z^2 + 1 = (Z - i) \cdot (Z + i)$. Hence $\mathrm{am}(i) = 1$ and $\mathrm{am}(-i) = 1$. Since by Theorem 11.23, we know that $1 \leq \mathrm{gm}(\lambda) \leq \mathrm{am}(\lambda)$ for any eigenvalue λ, we conclude that $\mathrm{gm}(i) = \mathrm{am}(i) = 1$ and $\mathrm{gm}(-i) = \mathrm{am}(-i) = 1$. Hence all in Corollary 11.28 are satisfied. We conclude that the given matrix $\mathbf{A}$ is diagonalizable over the complex numbers.

Now we explicitly compute an invertible matrix $\mathbf{Q} \in \mathbb{C}^{2\times 2}$ such that $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$ is a diagonal matrix. Let us denote by ϵ the standard basis of $\mathbb{C}^2$. Then ${}_{\epsilon}[L_{\mathbf{A}}]_{\epsilon} = \mathbf{A}$. To diagonalize $\mathbf{A}$, we simply diagonalize the corresponding linear map $L_{\mathbf{A}}$. In order to do that, we need to find an ordered basis of $\mathbb{C}^2$ consisting of eigenvectors. In Example 11.18, we saw that:

$$E_i \quad \text{has basis} \quad \left\{ \begin{bmatrix} i \\ 1 \end{bmatrix} \right\} \qquad \text{and} \qquad E_{-i} \quad \text{has basis} \quad \left\{ \begin{bmatrix} -i \\ 1 \end{bmatrix} \right\}.$$

Hence $\beta = \left(\begin{bmatrix} i \\ 1 \end{bmatrix}, \begin{bmatrix} -i \\ 1 \end{bmatrix} \right)$ is an ordered basis of $\mathbb{C}^2$ consisting of eigenvectors only. Using this ordered basis, we find that the mapping matrix ${}_{\beta}[L_{\mathbf{A}}]_{\beta}$ is a diagonal matrix with the eigenvalues of the vectors in the ordered matrix β on its diagonal. Hence

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

To find the matrix $\mathbf{Q}$ such that $\mathbf{Q}^{-1} \cdot \mathbf{A} \cdot \mathbf{Q}$, now observe that

$${}_{\beta}[L_{\mathbf{A}}]_{\beta} = {}_{\beta}[\mathrm{id}_{\mathbb{C}^2}]_{\epsilon} \cdot {}_{\epsilon}[L_{\mathbf{A}}]_{\epsilon} \cdot {}_{\epsilon}[\mathrm{id}_{\mathbb{C}^2}]_{\beta} = {}_{\epsilon}[\mathrm{id}_{\mathbb{C}^2}]_{\beta}^{-1} \cdot \mathbf{A} \cdot {}_{\epsilon}[\mathrm{id}_{\mathbb{C}^2}]_{\beta}.$$

Hence we can simply choose $\mathbf{Q} = {}_{\epsilon}[\mathrm{id}_{\mathbb{C}^2}]_{\beta}$, the change of coordinate matrix from β-coordinates to ϵ-coordinates. This matrix just contains the eigenvectors in β as columns. Hence

$$\mathbf{Q} = {}_{\epsilon}[\mathrm{id}_{\mathbb{C}^2}]_{\beta} = \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}$$

is the matrix we were looking for. Concretely, we have

$$\begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}^{-1} \cdot \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

11.4 Fibonacci numbers revisited

In Example 5.2, more precisely in equation (5-3), we gave an example of a recursively defined sequence of numbers $F_1, F_2, F_3, \ldots$ called the Fibonacci numbers:

$$F_n = \begin{cases} 1 & \text{if } n = 1, \\ 1 & \text{if } n = 2, \\ F_{n-1} + F_{n-2} & \text{if } n \geq 3. \end{cases} \tag{11-4}$$

This recursion can also be expressed using matrices. Indeed, directly from equation (11-4), one sees that

$$\begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} F_{n-1} \\ F_{n-2} \end{bmatrix} \quad \text{for all } n \geq 3.$$

This matrix form makes it possible to find a closed form for the Fibonacci numbers. First of all, we have the following:

|||| Lemma 11.32

For all $n \geq 2$ it holds that:

$$\begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{n-2} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Proof. This can be shown using induction on n with base case $n = 2$. The details are left to the reader. □

To find a closed formula for F_n, it is enough to find a closed formula for powers of the matrix

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

We will diagonalize $\mathbf{P}$ to do this. The point is that if a matrix can be diagonalized, it is possible to find a closed formula for its powers:

|||| Lemma 11.33

Let $\mathbb{F}$ be a field and $\mathbf{A} \in \mathbb{F}^{n \times n}$ a square matrix. Let $\mathbf{Q} \in \mathbb{F}^{n \times n}$ be an invertible matrix such that $\mathbf{Q} \cdot \mathbf{A} \cdot \mathbf{Q}^{-1}$ is a diagonal matrix $\mathbf{D}$ with the elements $d_1, \ldots, d_n$ on its diagonal. Then

$$\mathbf{A}^n = \mathbf{Q}^{-1} \cdot \mathbf{D}^n \cdot \mathbf{Q}.$$

Moreover, $\mathbf{D}^n$ is a diagonal matrix with the elements $d_1^n, \ldots, d_n^n$ on its diagonal.

Proof. With induction on n one can show that $(\mathbf{Q}\cdot\mathbf{A}\cdot\mathbf{Q}^{-1})^n = \mathbf{Q}\cdot\mathbf{A}^n\cdot\mathbf{Q}^{-1}$ for all $n \geq 1$. Since $\mathbf{Q}\cdot\mathbf{A}\cdot\mathbf{Q}^{-1} = \mathbf{D}$, the result then follows. Also showing that $\mathbf{D}^n$ is a diagonal matrix with the elements $d_1^n, \ldots, d_n^n$ on its diagonal, can readily be shown by induction on n. $\square$

The point of this lemma is that it makes the computation of powers of a matrix relatively easy if the matrix is diagonalizable. Now let us return to the matrix $\mathbf{P}$. The characteristic polynomial of $\mathbf{P}$ is

$$p_{\mathbf{P}}(Z) = \det\left(\begin{bmatrix} 1-Z & 1 \\ 1 & -Z \end{bmatrix}\right) = Z^2 - Z - 1.$$

Hence the eigenvalues of $\mathbf{P}$ are $\lambda_1 = \frac{1+\sqrt{5}}{2}$ and $\lambda_2 = \frac{1-\sqrt{5}}{2}$. This already means that the matrix $\mathbf{P}$ is diagonalizable. To find the desired change of coordinate matrix, we need to calculate a basis of the eigenspaces. To calculate a basis of the eigenspace E_{λ_1}, note that

$$\mathbf{P} - \lambda_1\cdot\mathbf{I}_2 = \begin{bmatrix} \frac{1-\sqrt{5}}{2} & 1 \\ 1 & \frac{-1-\sqrt{5}}{2} \end{bmatrix} \underset{R_2 \leftarrow R_2 - \lambda_2\cdot R_1}{\longrightarrow} \begin{bmatrix} \frac{1-\sqrt{5}}{2} & 1 \\ 0 & 0 \end{bmatrix}$$

Hence we see that a basis of E_{λ_1} is given by

$$\left\{\begin{bmatrix} -1 \\ \frac{1-\sqrt{5}}{2} \end{bmatrix}\right\}.$$

Similarly, one can show that a basis of E_{λ_2} is given by

$$\left\{\begin{bmatrix} -1 \\ \frac{1+\sqrt{5}}{2} \end{bmatrix}\right\}.$$

Hence

$$\begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ \frac{1-\sqrt{5}}{2} & \frac{1+\sqrt{5}}{2} \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} -1 & -1 \\ \frac{1-\sqrt{5}}{2} & \frac{1+\sqrt{5}}{2} \end{bmatrix},$$

which implies that

$$\mathbf{P} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ \frac{1-\sqrt{5}}{2} & \frac{1+\sqrt{5}}{2} \end{bmatrix} \cdot \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} \cdot \begin{bmatrix} -1 & -1 \\ \frac{1-\sqrt{5}}{2} & \frac{1+\sqrt{5}}{2} \end{bmatrix}^{-1}$$

Now applying Lemma 11.33 in order to compute powers of $\mathbf{P}$ and Lemma 11.32, we see that

$$\begin{aligned} \begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}^{n-2} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -1 & -1 \\ \frac{1-\sqrt{5}}{2} & \frac{1+\sqrt{5}}{2} \end{bmatrix} \cdot \begin{bmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^{n-2} & 0 \\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^{n-2} \end{bmatrix} \cdot \begin{bmatrix} -1 & -1 \\ \frac{1-\sqrt{5}}{2} & \frac{1+\sqrt{5}}{2} \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \end{aligned}$$

After working out all the matrix products on the right-hand side, one obtains that

$$\begin{bmatrix} F_n \\ F_{n-1} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{5}} \cdot \left(\frac{1+\sqrt{5}}{2}\right)^n - \frac{1}{\sqrt{5}} \cdot \left(\frac{1-\sqrt{5}}{2}\right)^n \\ \frac{1}{\sqrt{5}} \cdot \left(\frac{1+\sqrt{5}}{2}\right)^{n-1} - \frac{1}{\sqrt{5}} \cdot \left(\frac{1-\sqrt{5}}{2}\right)^{n-1} \end{bmatrix},$$

which explains where equation (5-4) came from.

In this section we focused on the Fibonacci numbers, but very similar techniques can be used to find closed formulas for other recursively defined sequences of numbers, but we will not pursue this further here.

11.5 Extra: What if diagonalization is not possible?

This section is not required reading and can be skipped. It is meant as extra material for a student who has the time and motivation for it.

As we have seen in the previous section, diagonalization of a linear map $L : V \to V$ is not always possible. In this section, we discuss the well-known *Jordan normal form*. The key for diagonalization was to study the eigenspace $E_\lambda = \ker(L - \lambda \cdot \mathrm{id}_V)$ for a given eigenvalue λ. We defined $\mathrm{gm}(\lambda) = \dim E_\lambda$ and have seen that in order to be able to diagonalize a matrix or linear map, it was important that the condition $\mathrm{gm}(\lambda) = \mathrm{am}(\lambda)$ is met. It turns out that if $\mathrm{gm}(\lambda) < \mathrm{am}(\lambda)$, one needs to study the kernels of powers of the linear map $L - \lambda \cdot \mathrm{id}_V$. Here the i-th power f^i of a function $f : V \to V$ should be understood as the i-fold composite of f with itself (so $f^1 = f$, $f^2 = f \circ f$, etcetera).

Lemma 11.34

Let $\mathbb{F}$ be a field, V an n-dimensional vector space over $\mathbb{F}$ and $L : V \to V$ a linear map. Further assume that $\lambda \in \mathbb{F}$ is an eigenvalue of L. Then

$$\ker(L - \lambda \cdot \mathrm{id}_V) \subseteq \ker((L - \lambda \cdot \mathrm{id}_V)^2) \subseteq \ker((L - \lambda \cdot \mathrm{id}_V)^3) \subseteq \dots$$

Moreover,

1. if $\ker((L - \lambda \cdot \mathrm{id}_V)^i) = \ker((L - \lambda \cdot \mathrm{id}_V)^{i+1})$ for some positive integer i, then $\ker((L - \lambda \cdot \mathrm{id}_V)^i) = \ker((L - \lambda \cdot \mathrm{id}_V)^m)$ for all $m \geq i$, and
2. $\ker((L - \lambda \cdot \mathrm{id}_V)^n) = \ker((L - \lambda \cdot \mathrm{id}_V)^{n+1})$.

Proof. It is clear that for all i, the kernel of $(L-\lambda\cdot\mathrm{id}_V)^{i+1}$ is a subspace of the kernel of $(L-\lambda\cdot\mathrm{id}_V)^i$. If equality holds for some i, then by the rank-nullity theorem for linear maps, see Corollary 10.40, we also obtain that the images of the linear maps $(L-\lambda\cdot\mathrm{id}_V)^{i+1}$ and $(L-\lambda\cdot\mathrm{id}_V)^i$ are the same. But then also

$$\begin{aligned}\mathrm{im}(L-\lambda\cdot\mathrm{id}_V)^{i+2} &= (L-\lambda\cdot\mathrm{id}_V)^{i+2}(V)\\ &= (L-\lambda\cdot\mathrm{id}_V)((L-\lambda\cdot\mathrm{id}_V)^{i+1}(V))\\ &= (L-\lambda\cdot\mathrm{id}_V)((L-\lambda\cdot\mathrm{id}_V)^{i}(V))\\ &= (L-\lambda\cdot\mathrm{id}_V)^{i+1}(V)\\ &= \mathrm{im}(L-\lambda\cdot\mathrm{id}_V)^{i+1}\\ &= \mathrm{im}(L-\lambda\cdot\mathrm{id}_V)^{i}.\end{aligned}$$

But then, again using the rank-nullity theorem for linear maps, we see that the kernel of $(L-\lambda\cdot\mathrm{id}_V)^{i+2}$ is equal to the kernel of $(L-\lambda\cdot\mathrm{id}_V)^i$. Using induction on m, one can similarly show that for any $m\geq i$ the kernel of $(L-\lambda\cdot\mathrm{id}_V)^m$ is equal to the kernel of $(L-\lambda\cdot\mathrm{id}_V)^i$.

Now consider the sequence of subspaces:

$$\ker(L-\lambda\cdot\mathrm{id}_V)\subseteq\ker((L-\lambda\cdot\mathrm{id}_V)^2)\subseteq\ker((L-\lambda\cdot\mathrm{id}_V)^3)\subseteq\dots$$

By the previous, we know that if equality holds at some point in the sequence, then equalities will hold from then on. Therefore there exists $e\geq 1$ such that

$$\ker(L-\lambda\cdot\mathrm{id}_V)\subsetneq\cdots\subsetneq\ker((L-\lambda\cdot\mathrm{id}_V)^e)=\ker((L-\lambda\cdot\mathrm{id}_V)^{e+1})=\dots$$

For every strict inclusion, the dimension of the subspace increases by at least one. Since $\dim V=n$ and $\dim\ker(L-\lambda\cdot\mathrm{id}_V)=\dim E_\lambda\geq 1$, this can occur at most n times. Hence $e\leq n$. In particular $\ker((L-\lambda\cdot\mathrm{id}_V)^n)=\ker((L-\lambda\cdot\mathrm{id}_V)^{n+1})$. □

|||| Theorem 11.35

Let $\mathbb{F}$ be a field, V an n-dimensional vector space over $\mathbb{F}$ and $L:V\to V$ a linear map. Further assume that $\lambda\in\mathbb{F}$ is an eigenvalue of L. Further, write $U=\mathrm{im}(L-\lambda\cdot\mathrm{id}_V)^n$ and $W=\ker(L-\lambda\cdot\mathrm{id}_V)^n$. Then

1. $L(U)\subseteq U$ and $L(W)\subseteq W$.
2. $\dim U+\dim W=\dim V$ and $U\cap W=\{\mathbf{0}\}$.
3. Any vector in V can be written as the sum of a vector in U and a vector in W.

Proof. We have seen in the third item of Lemma 11.34 that $W = \ker(L - \lambda \cdot \mathrm{id}_V)^n = \ker(L - \lambda \cdot \mathrm{id}_V)^{n+1}$. Now choose $\mathbf{w} \in W$. Then also $(L - \lambda \cdot \mathrm{id}_V)(\mathbf{w}) \in W$, since $(L-\lambda \cdot \mathrm{id}_V)^n((L-\lambda \cdot \mathrm{id}_V)(\mathbf{w})) = (L-\lambda \cdot \mathrm{id}_V)((L-\lambda \cdot \mathrm{id}_V)^n(\mathbf{w})) = (L-\lambda \cdot \mathrm{id}_V)(\mathbf{0}) = \mathbf{0}$. Hence $L(\mathbf{w}) - \lambda \cdot \mathbf{w} \in W$, which implies that $L(\mathbf{w}) \in W$. We may conclude that $L(W) \subseteq W$. Similarly, if $\mathbf{u} \in U$, then $(L - \lambda \cdot \mathrm{id}_V)(\mathbf{u}) \in U$, since if $\mathbf{u} = (L - \lambda \cdot \mathrm{id}_V)^n(\mathbf{v})$ for some $\mathbf{v} \in V$, then $(L - \lambda \cdot \mathrm{id}_V)(\mathbf{u}) = (L - \lambda \cdot \mathrm{id}_V)((L - \lambda \cdot \mathrm{id}_V)^n(\mathbf{v})) = (L - \lambda \cdot \mathrm{id}_V)^n((L - \lambda \cdot \mathrm{id}_V)(\mathbf{v})) \in U$. Hence $L(\mathbf{u}) - \lambda \cdot \mathbf{u} \in U$, which implies that $L(\mathbf{u}) \in U$. We may conclude that $L(U) \subseteq U$.

The rank-nullity theorem for linear maps applied to the linear map $(L - \lambda \cdot \mathrm{id}_V)^n : V \to V$ immediately implies that $\dim U + \dim W = \dim V$. Now, we prove that $U \cap W = \{\mathbf{0}\}$. Let $\mathbf{u} \in U \cap W$. We wish to show that $\mathbf{u} = \mathbf{0}$. First of all, since $\mathbf{u} \in U$, there exists $\mathbf{v} \in V$ such that $\mathbf{u} = (L - \lambda \cdot \mathrm{id}_V)^n(\mathbf{v})$. Second, since $\mathbf{u} \in W$, we have $(L - \lambda \cdot \mathrm{id}_V)^n(\mathbf{u}) = \mathbf{0}$. Combining these two, we see that

$$(L - \lambda \cdot \mathrm{id}_V)^{2n}(\mathbf{v}) = (L - \lambda \cdot \mathrm{id}_V)^n(\mathbf{u}) = \mathbf{0}.$$

In other words, $\mathbf{v} \in \ker(L - \lambda \cdot \mathrm{id}_V)^{2n}$. However, Lemma 11.34 implies that $\ker(L - \lambda \cdot \mathrm{id}_V)^{2n} = \ker(L - \lambda \cdot \mathrm{id}_V)^n$ and hence $\mathbf{v} \in \ker(L - \lambda \cdot \mathrm{id}_V)^n$. But then $\mathbf{u} = (L - \lambda \cdot \mathrm{id}_V)^n(\mathbf{v}) = \mathbf{0}$, which is what we wanted to show.

Given an ordered basis $\beta_U = (\mathbf{u}_1, \ldots, \mathbf{u}_r)$ of U and an ordered basis $\beta_W = (\mathbf{w}_1, \ldots, \mathbf{w}_s)$ of W, joining the two together yields an ordered basis $\beta = (\beta_U, \beta_W)$ of V. Indeed, the fact that $U \cap W = \{\mathbf{0}\}$ can be used to show that the vectors in β are linearly independent, while the identity $\dim U + \dim W = \dim V$ implies that β contains exactly n vectors. Now given an arbitrarily chosen $\mathbf{v} \in V$, we can write $\mathbf{v}$ in exactly one way as a linear combination of the $\mathbf{u}_i$ and the $\mathbf{w}_j$, say $\mathbf{v} = \sum_i \alpha_i \cdot \mathbf{u}_i + \sum_j \beta_j \cdot \mathbf{w}_j$. Now observing that $\sum_i \alpha_i \cdot \mathbf{u}_i \in U$ and $\sum_j \beta_j \cdot \mathbf{w}_j \in W$, the last item in the theorem follows. □

|||| Corollary 11.36

Using the same notation as in Theorem 11.35, write $p_L(Z) = (\lambda - Z)^{\mathrm{am}(\lambda)} \cdot q(Z)$ for a suitably chosen $q(Z) \in \mathbb{F}[Z]$. Denote by $L|_W : W \to W$, respectively $L|_U : U \to U$, the linear maps obtained by restricting the domain and codomain of L to U, respectively W. Then $p_L(Z) = p_{L|_U}(Z) \cdot p_{L|_W}(Z)$ and λ is not a root of $p_{L|_U}(Z)$.

Proof. Given an ordered basis $\beta_U = (\mathbf{u}_1, \ldots, \mathbf{u}_r)$ of U and an ordered basis $\beta_W = (\mathbf{w}_1, \ldots, \mathbf{w}_s)$ of W, we have already seen in the proof of Theorem 11.35, that $\beta = (\beta_U, \beta_W)$

is an ordered basis of V. Since we know that $L(U) \subseteq U$ and $L(W) \subseteq W$, the matrix ${}_\beta[L]_\beta \in \mathbb{F}^{n\times n}$ will have the form

$$ {}_\beta[L]_\beta = \begin{bmatrix} {}_{\beta_U}[L]_{\beta_U} & \mathbf{0} \\ \mathbf{0} & {}_{\beta_W}[L]_{\beta_W} \end{bmatrix}. \tag{11-5}$$

This implies that $p_L(Z) = p_{L|_U}(Z) \cdot p_{L|_W}(Z)$. Now observe that λ cannot be a root of $p_{L|_U}(Z)$. Indeed, if this would be the case, then there would exist a nonzero $\mathbf{u} \in U$ such that $L|_U(\mathbf{u}) = \lambda \cdot \mathbf{u}$. Since by definition of the linear map $L|_U$, we have $L|_U(\mathbf{u}) = L(\mathbf{u})$, this would imply that $(L - \lambda \cdot \mathrm{id}_V)(\mathbf{u}) = \mathbf{0}$. But then $\mathbf{u} \in \ker(L - \lambda \cdot \mathrm{id}_V)$, implying that $\mathbf{u} \in \ker(L - \lambda \cdot \mathrm{id}_V)^n = W$. Since we have seen that $U \cap W = \{\mathbf{0}\}$, we would obtain that $\mathbf{u} = \mathbf{0}$, contrary to our assumption. □

Now let us return to what we are trying to achieve: to find a matrix representing L that is as simple as possible. Equation (11-5) is an important step on the way. Indeed, we have reduced the problem into two simpler ones: finding a simple matrix representing $L|_U$ and one representing $L|_W$. Moreover, λ is not a root of the characteristic polynomial of $L|_U$, so to deal with the eigenvalue λ of L, we only have to continue with the study of the linear map $L|_W$. Let us first get an intuitive idea of what may be going on. If $\mathrm{am}(\lambda) = \mathrm{gm}(\lambda)$, we can find an ordered basis β_W of W consisting of eigenvectors for L only, all having λ as eigenvalue. Then

$$ {}_{\beta_W}[L]_{\beta_W} = [\lambda \cdot \mathbf{I}_s]\,, $$

where $s = \dim W$. What we did in the previous section is essentially to repeat this procedure for another eigenvalue and split the matrix representing $L|_U$ further up in smaller blocks. As long as the algebraic and geometric multiplicity of the eigenvalues is always the same, we end up with diagonalizing the entire matrix.

So what happens if $\mathrm{am}(\lambda) > \mathrm{gm}(\lambda)$? We can still find an ordered basis β_W of W containing the eigenvectors with eigenvalue λ, but we also need some more vectors in β_W that are not eigenvectors. To put this in a different way: if $\mathrm{am}(\lambda) = \mathrm{gm}(\lambda)$, then $W = E_\lambda$, so that $\ker(L - \lambda \cdot \mathrm{id}_V) = \ker((L - \lambda \cdot \mathrm{id}_V)^n)$. However, if $\mathrm{am}(\lambda) > \mathrm{gm}(\lambda)$, then W contains E_λ, but is not equal to it. Then apparently $\ker(L - \lambda \cdot \mathrm{id}_V) \subsetneq \ker((L - \lambda \cdot \mathrm{id}_V)^n)$. This implies in particular that $\ker(L - \lambda \cdot \mathrm{id}_V) \subsetneq \ker((L - \lambda \cdot \mathrm{id}_V)^2)$ using Lemma 11.34. If we choose a vector $\mathbf{w} \in \ker((L - \lambda \cdot \mathrm{id}_V)^2) \setminus \ker(L - \lambda \cdot \mathrm{id}_V)$, it has the nice property that $L(\mathbf{w}) - \lambda \cdot \mathbf{w} = (L - \lambda \cdot \mathrm{id}_V)(\mathbf{w}) \in \ker(L - \lambda \cdot \mathrm{id}_V) = E_\lambda$. Let us define $\mathbf{v} = L(\mathbf{w}) - \lambda \cdot \mathbf{w}$. Further, we see that: $L(\mathbf{v}) = \lambda \cdot \mathbf{v}$, since $\mathbf{v} \in E_\lambda$, and $L(\mathbf{w}) = \lambda \cdot \mathbf{w} + \mathbf{v}$, by the way we defined $\mathbf{v}$. So if we only consider the effect of L on the two-dimensional subspace of V spanned by $\mathbf{v}$ and $\mathbf{w}$, which has ordered basis $\mathbf{v}, \mathbf{w}$, we can represent L by the matrix

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

This gives a first idea of what to expect more in general for a matrix representing $L|_W$. In particular, it motivates the following:

|||| Definition 11.37

Let $\mathbb{F}$ be a field and $\lambda \in \mathbb{F}$. A *Jordan block* of size e is a matrix $\mathbf{J}_e(\lambda) \in \mathbb{F}^{e\times e}$ of the form

$$\mathbf{J}_e(\lambda) = \begin{bmatrix} \lambda & 1 & & \mathbf{0} \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ \mathbf{0} & & & \lambda \end{bmatrix}.$$

|||| Lemma 11.38

Let $L : V \to V$ be a linear map, $\dim V = n$ and λ an eigenvalue of L. Further, let $W = \ker((L - \lambda \cdot \mathrm{id}_V)^n)$. Then there exists an ordered basis of W such that $L|_W : W \to W$, the restriction of L to W has a mapping matrix $\mathbf{D}(\lambda)$ of the form

$$\mathbf{D}(\lambda) = \begin{bmatrix} \mathbf{J}_{e_1}(\lambda) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{J}_{e_s}(\lambda) \end{bmatrix}$$

for a certain positive integer s and certain positive integers $e_1, \ldots, e_s$.

Proof. It will be convenient to write $W_i = \ker((L - \lambda \cdot \mathrm{id}_V)^i)$ and $r_i = \dim W_i$. Note that $W_1 = E_\lambda$. Now let e be the largest exponent such that $W_{e-1} \subsetneq W_e$. Then $W_e = W$ and $r_1 < r_2 < \cdots < r_e = \dim W$. Let $\beta = (\mathbf{w}_1, \ldots, \mathbf{w}_{r_e})$ be an ordered basis of W with the additional property that $(\mathbf{w}_1, \ldots, \mathbf{w}_{r_i})$ is an ordered basis of W_i for all i.

We now gradually construct another ordered basis, say γ, of W. First of all, we add to γ the vectors $\mathbf{w}_i$ for any i between $r_{e-1} + 1$ and r_e. By construction, for any i between $r_{e-1} + 1$ and r_e the vector $\mathbf{w}_i$ lies in W_e, but not in W_{e-1}. Now, consider the vectors $\mathbf{w}_{i,j} = (L - \lambda \cdot \mathrm{id}_V)^j(\mathbf{w}_i)$, for $j = 0, \ldots, e-1$. Note that the vector $\mathbf{w}_{i,j}$ lies in W_{e-j}, but not in W_{e-j-1}. Also note that $\mathbf{w}_i = \mathbf{w}_{i,0}$ for any i between $r_{e-1} + 1$ and r_e.

First we claim that vectors $\mathbf{w}_{i,e-1}$ are linearly independent. Indeed if $\sum_{i=r_{e-1}+1}^{r_e} \alpha_i \cdot$

$\mathbf{w}_{i,e-1} = 0$, then $\sum_{i=r_{e-1}+1}^{r_e} \alpha_i \cdot \mathbf{w}_i \in W_{e-1}$ and hence in the span of the vectors $\mathbf{w}_1, \ldots, \mathbf{w}_{r_{e-1}}$. But since the vectors $\mathbf{w}_1, \ldots, \mathbf{w}_{r_e}$ are linearly independent, this implies that α_i for all $i = r_{e-1}+1, \ldots, r_e$. Next we claim that the vectors $\mathbf{w}_{i,j}$, with $i = r_{e-1}+1, \ldots, r_e$ and $j = 0, \ldots, e-1$ are linearly independent. If $\sum_i \sum_{j=0}^{e-1} \alpha_{i,j} \cdot \mathbf{w}_{i,j} = 0$, then applying $(L - \lambda \cdot \mathrm{id}_V)^{e-1}$ yields the equation $\sum_i \alpha_{i,0} \cdot \mathbf{w}_{i,e-1} = 0$. Hence $\alpha_{i,0} = 0$ for all i. Now applying lower and lower powers of $(L - \lambda \cdot \mathrm{id}_V)$ to the equation $\sum_i \sum_{j=0}^{e-1} \alpha_{i,j} \cdot \mathbf{w}_{i,j} = 0$, one obtains inductively that $\alpha_{i,j} = 0$ for all i and j.

Therefore, it makes sense to include all the vectors $\mathbf{w}_{i,j}$ in γ. More precisely, we now set $\gamma = (\mathbf{w}_{r_{e-1}+1,e-1}, \ldots, \mathbf{w}_{r_{e-1}+1,0}, \ldots, \mathbf{w}_{r_e,e-1}, \ldots, \mathbf{w}_{r_e,0})$. We have $(L - \lambda \cdot \mathrm{id}_V)(\mathbf{w}_{i,j}) = \mathbf{w}_{i,j+1}$ for $j = 0, \ldots, e-2$ and $(L - \lambda \cdot \mathrm{id}_V)(\mathbf{w}_{i,e-1}) = \mathbf{0}$. This implies that

$$L(\mathbf{w}_{i,j}) = \lambda \cdot \mathbf{w}_{i,j} + \mathbf{w}_{i,j+1} \quad \text{for } j = 0, \ldots, e-2 \text{ and} \quad L(\mathbf{w}_{i,e-1}) = \lambda \cdot \mathbf{w}_{i,e-1}.$$

We have now in fact shown that the restriction of L to the subspace spanned by the $\mathbf{w}_{i,j}$ can be represented by a block diagonal matrix with $r_e - r_{e-1}$ many matrices $\mathbf{J}_e(\lambda)$ on its diagonal.

For any j between 0 and $e-1$, the vectors $\mathbf{w}_{i,j}$, with i varying from $r_{e-1}+1$ to r_e, span a subspace of W_{e-j}. If for all j, this subspace is equal to W_{e-j}, then γ is an ordered basis of W, giving rise to a matrix representing W in Jordan normal form, as we have seen. Otherwise, let $\tilde{j}$ be the smallest value of j such that this subspace is not all of W_{e-j} and define $\tilde{e} = e - \tilde{j}$. Then let $\tilde{\mathbf{w}}_1, \ldots, \tilde{\mathbf{w}}_d \in W_{\tilde{e}}$ be vectors such that

$$(\mathbf{w}_1, \ldots, \mathbf{w}_{r_{\tilde{e}-1}}, \tilde{\mathbf{w}}_1, \ldots, \tilde{\mathbf{w}}_d, \mathbf{w}_{r_{e-1}+1,\tilde{j}}, \ldots, \mathbf{w}_{r_e,\tilde{j}})$$

is an ordered basis of $W_{\tilde{e}}$. Now we proceed similarly as in the start, defining vectors $\tilde{\mathbf{w}}_{i,j} = (L - \lambda \cdot \mathrm{id}_V)^j(\tilde{\mathbf{w}}_i)$, which we add to γ and which will give rise to Jordan blocks of size $\tilde{e}$ in the matrix representing L.

Continuing in this way, we end up with an ordered basis of γ giving rise to a matrix in Jordan normal form that represents the restriction of L to W. □

Theorem 11.39

Let $\mathbb{F}$ be a field, V a finite dimensional vector space, and $L : V \to V$. Suppose that there exist distinct $\lambda_1, \ldots, \lambda_r \in \mathbb{F}$ and positive integers $m_1, \ldots, m_r$ such that $p_L(Z) = (-1)^n \cdot (Z - \lambda_1)^{m_1} \cdots (Z - \lambda_r)^{m_r}$. Then the linear map can be represented by a matrix of the form

$$\begin{bmatrix} \mathbf{D}(\lambda_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{D}(\lambda_r) \end{bmatrix},$$

where each matrix $\mathbf{D}(\lambda_i) \in \mathbb{F}^{m_i \times m_i}$ is of the form as in Lemma 11.38.

Proof. We prove the theorem with induction of the number of eigenvalues. We use the same notation for U and W as in Theorem 11.35. If $r = 1$, $W = V$ and hence $L|_W = L$. Hence the result follows from Lemma 11.38. Now let $r > 1$. Let $\lambda = \lambda_r \in \mathbb{F}$ be an eigenvalue of L. From Lemma 11.38, we conclude that we can choose an ordered basis for W such that $L|_W$ is represented by a block diagonal matrix with Jordan blocks $\mathbf{J}_{e_i}(\lambda_r)$ on its diagonal. Further, from Corollary 11.36, λ_r is not an eigenvalue of $L|_U$, while the characteristic polynomial of $L|_U$ is a divisor of $P_L(Z)$. Hence the induction hypothesis applies. $\square$

The matrix given in Theorem 11.39 is said to be the *Jordan normal form* of the matrix $\mathbf{A}$.

Corollary 11.40

Let $\mathbf{A} \in \mathbb{C}^{n \times n}$ be a complex matrix. Then $\mathbf{A}$ is similar to a matrix in Jordan normal form.

Proof. If we work over the complex numbers, it follows from Theorem 4.23 that $p_{\mathbf{A}}(Z)$ can be written as a product of its leading term, which is $(-1)^n$, and terms of the form $Z - \lambda$. Hence Theorem 11.39 applies. $\square$