Algebra B

Lecture notes

Mariano Suárez-Álvarez

Version: November 12, 2024

Contents

Capítulo 1 The structure of linear maps

§1.1. Eigenvalues, eigenvectors and eigenspaces

Linear maps

{subsect:eig:maps}

We fix in this section a field $\Bbbk$, a vector space V over $\Bbbk$ and a linear map $f : V \to V$.

A scalar $\lambda \in \Bbbk$ is an ***eigenvalue*** for f if there exists a non-zero vector v in V such that $f(v) = \lambda v$, and any such vector is an ***eigenvector*** for f corresponding to that eigenvalue. We have to keep in mind that the eigenvalues of f are elements of the field $\Bbbk$: this is very important at the moment of deciding if an linear map has eigenvalues and finding them. For example, the reader can easily verify that the $\mathbb{R}$-linear map $\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right) \in \mathbb{R}^2 \mapsto \left(\begin{smallmatrix}-y\\x\end{smallmatrix}\right) \in \mathbb{R}^2$ has no eigenvalues, while the $\mathbb{C}$-linear map $\left(\begin{smallmatrix}x\\y\end{smallmatrix}\right) \in \mathbb{C}^2 \mapsto \left(\begin{smallmatrix}-y\\x\end{smallmatrix}\right) \in \mathbb{C}^2$ has two, namely i and $-i$, even though the two linear maps «have the same formulas».

For each scalar $\lambda \in \Bbbk$ we consider in V the subset

$$E_\lambda(f) := \{v \in V : f(v) = \lambda v\}.$$

{prop:eig:lif}

Proposition 1.1.1. *Let $f : V \to V$ be a linear map, and let λ be an element of $\Bbbk$.*

(i) $E_\lambda(f)$ is a subspace of V, and it coincides with the kernel of the linear map $\lambda \cdot \mathrm{id}_V - f : V \to V$.

(ii) The subspace $E_\lambda(f)$ is non-zero if and only if the scalar λ is an eigenvalue for f.

When $E_\lambda(f) \neq 0$, so that λ is an eigevalue for f, we call $E_\lambda(f)$ the ***eigenspace*** of f corresponding to the eigenvalue λ. In that case, the non-zero elements of $E_\lambda(f)$ are precisely the eigenvectors of f corresponding to λ.

Proof. (i) Let K be the kernel of the map $\lambda \cdot \mathrm{id}_V - f : V \to V$. We will show that $E_\lambda(f) = K$, and then it will follow immediately that $E_\lambda(f)$ is a subspace of V, since the kernel of any linear map is a subspace of its domain. Now a vector $v \in V$ belongs to $E_\lambda(f)$ exactly when $f(v) = \lambda v = \lambda \cdot \mathrm{id}_V(v)$, and this occurs exactly when $(\lambda \cdot \mathrm{id}_V - f)(v) = 0$, that is, when $v \in K$. This proves what we wanted.

(ii) If $E_\lambda(f)$ is not the zero subspace, then there is a non-zero vector $v \in V$ such that $v \in E_\lambda(f)$, so that $f(v) = \lambda v$: this vector v is then an eigenvector for f with eigenvalue λ and, in particular, λ is an eigenvalue of f. Conversely, if λ is an eigenvalue of f then there exists a non-zero vector $v \in V$ such that $f(v) = \lambda v$, and that vector is a non-zero element of $E_\lambda(f)$, so that $E_\lambda(f)$ is a non-zero subspace of V. □

An immediate consequence of the proposition is the following criterion:

{coro:eig:lif}

Corollary 1.1.2. *Let V be a vector space over a field $\Bbbk$, let $f : V \to V$ be a linear map, and let λ be an element of $\Bbbk$. The following two statements are equivalent:*

(a) *The scalar λ is an eigenvalue of f.*

(b) *The linear map $\lambda \cdot \mathrm{id}_V - f : V \to V$ is* not *injective.*

If the vector space V is finite-dimensional, *then these two statements are also equivalent to the following third one:*

(c) *The linear map $\lambda \cdot \mathrm{id}_V - f : V \to V$ is* not *bijective.*

Proof. Proposition 1.1.1 tells us that λ is an eigenvalue for f exactly when the subspace $E_\lambda(f)$ is non-zero, and we also know that the linear map $\lambda \cdot \mathrm{id}_V - f : V \to V$ is injective exactly when its kernel is zero: since $E_\lambda(f) = \ker(\lambda \cdot \mathrm{id}_V - f)$, this implies immediately that $(a) \Leftrightarrow (b)$, and proves the first part of the proposition.

Let us now suppose that the vector space V is finite-dimensional. If the map $\lambda \cdot \mathrm{id}_V - f : V \to V$ is not injective then it is obviously not bijective, and this gives us the implication $(b) \Rightarrow (c)$. On the other hand, as the vector space V is finite-dimensional whenever the linear map $\lambda \cdot \mathrm{id}_V - f : V \to V$ is not bijective it is also not injective, and this gives us the remaining implication $(c) \Rightarrow (b)$. □

Observation 1.1.3. When proving the implication $(c) \Rightarrow (b)$ of the corollary we used the following result:

> *an endomorphism $f : V \to V$ of a finite-dimensional vector space V is injective if and only if it is bijective.*

If we remove the hypothesis of finite-dimensionality then the result is no longer true. For example, the $\mathbb{R}$-linear map

$$f : p \in \mathbb{R}[X] \mapsto Xp \in \mathbb{R}[X]$$

is injective but not surjective, so it is not bijective — we leave the verification of this as an exercise for the reader. As a consequence of this fact, if we remove the hypothesis of finite-dimensionality from Corollary 1.1.2 then the statement (c) is no longer equivalent to the other two. It is because of this that in all what follows we will mostly restrict our attention to endomorphisms of finite-dimensional vector spaces. There is a similar theory for infinite-dimensional vector spaces but it is more complicated.

Matrices

{subsect:eig:mats}

Let now $A \in \mathrm{M}_n(\Bbbk)$ be a matrix of size $n \times n$ with entries in the field $\Bbbk$. We say that a scalar $\lambda \in \Bbbk$ is an ***eigenvalue*** for A if there is a non-zero vector $v \in \Bbbk^n$ such that $Av = \lambda v$, and any such vector is an ***eigenvector*** for A corresponding to that eigenvalue. Just as for linear maps, for any $\lambda \in \Bbbk$ we put

$$E_\lambda(A) \coloneqq \{v \in \Bbbk^n : Av = \lambda v\}$$

and we have the following:

{prop:eig:mats}

Proposition 1.1.4. *Let $A \in \mathrm{M}_n(\Bbbk)$ and let $\lambda \in \Bbbk$. The subset $E_\lambda(A)$ is a subspace of V, and the following statements are equivalent:*

(a) The scalar λ is an eigenvalue for A.
(b) The subspace $E_\lambda(A)$ is non-zero.
(c) The determinant of the matrix $\lambda \cdot I_n - A$ is zero.
(d) The rank of the matrix $\lambda \cdot I_n - A$ is less than n.

When the subspace $E_\lambda(A)$ is non-zero we call it the ***eigenspace*** of A corresponding to the eigenvalue λ. Its elements are the eigenvectors of A with eigenvalue λ and the zero vector.

Proof. If x and y are two elements of $E_\lambda(A)$, so that $Ax = \lambda x$ and $Ay = \lambda y$, and α and β are two scalars in $\Bbbk$, then we have that

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay = \alpha\lambda x + \beta\lambda y = \lambda(\alpha x + \beta y),$$

and this tells us that $\alpha x + \beta y \in E_\lambda(A)$. It follows from this that $E_\lambda(A)$ is a subspace of $\Bbbk^n$.

The scalar $\lambda \in \Bbbk$ is an eigenvalue for A exactly when there is a non-zero vector $v \in \Bbbk^n$ such that $Av = \lambda v$, that is, when there is a non-zero element in $E_\lambda(A)$: this shows that $(a) \Leftrightarrow (b)$ holds.

On the other hand, a vector $v \in \Bbbk^n$ belongs to $E_\lambda(A)$ if and only if $Av = \lambda v$ or, equivalently, if and only if $(\lambda \cdot I_n - A)v = 0$. This means that the subspace $E_\lambda(A)$ is non-zero exactly when there are non-zero solutions to the equation $(\lambda \cdot I_n - A)v = 0$, and this happens exactly when the determinant of the matrix $\lambda \cdot I_n - A$ is zero, which in turn happens exactly when the rank of that matrix is less than n. This shows that the statements (b), (c) and (d) are equivalent. □

The connection between the two cases

In all that we will do in what follows we will always have to consider two different but similar situations: the case in which are working with an endomorphism $f : V \to V$ of a finite-dimensional vector space over a field $\Bbbk$, and the case in which we are working with a matrix $A \in \mathrm{M}_n(\Bbbk)$ with entries in $\Bbbk$. The two cases are closely related, as we know. Let us recall how.

Suppose first that we have a matrix $A \in \mathrm{M}_n(\Bbbk)$. We can then consider the vector space $V = \Bbbk^n$ and the linear map

$$f_A : x \in V \mapsto Ax \in V.$$

The eigenvalues, eigenvectors and eigenspaces of the matrix A that we defined in Section 1.1.2 are exactly the same as the eigenvalues, eigenvectors and eigenspaces of the linear map f_A that we defined in Section 1.1.1. Indeed,

- a scalar λ is an eigenvalue for A exactly when there is a non-zero vector v in $\Bbbk^n$ such that $Av = \lambda v$, and this happens precisely when there is a non-zero vector v in $\Bbbk^n$ such that $f_A(v) = \lambda v$, that is, when λ is an eigenvalue for f_A;
- if λ is a scalar, then a non-zero vector v of $\Bbbk^n$ is an eigenvector for A with eigenvalue λ if and only if $Av = \lambda v$, and clear this occurs exactly when $f_A(v) = \lambda v$, that is, when v is an eigenvector for f_A with eigenvalue λ; and finally
- for each λ in $\Bbbk$ we have $E_\lambda(f_A) = E_\lambda(f_A)$.

This is very direct — going in the other direction is slightly more elaborated.

Let us suppose now that we have an endomorphism $f : V \to V$ of a finite dimensional vector space V, let n be the dimension of V, and let $\mathscr{B} = (v_1, v_2, \dots, v_n)$ an ordered basis for V. Let moreover $A \coloneqq [f]_{\mathscr{B}}$ be the matrix of f with respect to $\mathscr{B}$, so that $A = (a_{i,j})$ is the unique element of $\mathrm{M}_n(\Bbbk)$ such that

$$f(v_i) = a_{1,i}v_1 + a_{2,i}v_2 + \cdots + a_{n,i}v_n \qquad \text{for each } i \in \{1, 2, \dots, n\}. \tag{1.1}$$ {eq:fvi}

In other words, the entries that appear in ith column of A are the coefficients of the vector $f(v_i)$ in terms of the basis $\mathscr{B}$, in the order given by that ordered basis.

If v is an element of V, then there exist scalars $c_1, \dots, c_n \in \Bbbk$, all uniquely determined by v, such that $v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$, and as usual we write

$$[v]_{\mathscr{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

for the element of $\Bbbk^n$ whose entries are the coordinates of v with respect to the ordered basis $\mathscr{B}$. In this way we obtain a function

$$v \in V \mapsto [v]_{\mathscr{B}} \in \Bbbk^n$$ {eq:az:a}

that we know to be an isomorphism of vector spaces. In particular, this function is linear, so that

$$[\alpha v + \beta w]_{\mathscr{B}} = \alpha[v]_{\mathscr{B}} + \beta[w]_{\mathscr{B}} \tag{1.2}$$ {eq:az:c1}

whenever $v, w \in V$ and $\alpha, \beta \in \Bbbk$, and injective, so that

$$[v]_{\mathscr{B}} = 0 \implies v = 0 \tag{1.3}$$ {eq:az:c2}

for all vectors v in V.

Lemma 1.1.5. *Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space V, and let n be the dimension of V, let $\mathscr{B}$ be an ordered basis for V, and let $A \coloneqq [f]_{\mathscr{B}} \in \mathrm{M}_n(\Bbbk)$ be the matrix of f with respect to $\mathscr{B}$. Finally, let λ be a scalar in $\Bbbk$.*

(*i*) *If v is an element of $E_\lambda(f)$, then $[v]_{\mathscr{B}}$ is an element of $E_\lambda(A)$. The function*

$$v \in E_\lambda(f) \mapsto [v]_{\mathscr{B}} \in E_\lambda(f)$$

is an isomorphism of vector spaces.

(*ii*) *The scalar λ is an eigenvalue of the linear map f if and only if it is an eigenvalue of the matrix A.*

(*iii*) *A vector v in V is an eigenvector for the linear map f with eigenvalue λ if and only if the vector of its coordinates $[v]_{\mathscr{B}}$ is an eigenvector for the matrix A with eigenvalue λ.*

Proof. Suppose that v is an element of $E_\lambda(f)$. Since $\mathscr{B}$ is an ordered basis for V, there are scalars $c_1, c_2, \dots, c_n \in \Bbbk$ such that $v = c_1v_1 + c_2v_2 + \dots + c_nv_n$ and, since $\mathscr{B}$ is linearly independent and $v \neq 0$, not all of those n scalars are zero. Since $v \in E_\lambda(f)$, we have that $f(v) = \lambda v$. The right hand side of this equality is

$$\lambda c_1 v_1 + \lambda c_2 v_2 + \dots + \lambda c_n v_n, \tag{1.4}$$ {eq:az:1}

while the left hand side is

$$f(v) = f(c_1v_1 + c_2v_2 + \dots + c_nv_n) = c_1 f(v_1) + c_2 f(v_2) + \dots + c_n f(v_n),$$

which according to (1.1) is

$$\begin{aligned}
&= c_1(a_{1,1}v_1 + a_{2,1}v_2 + \dots + a_{n,1}v_n) + c_2(a_{1,2}v_1 + a_{2,2}v_2 + \dots + a_{n,2}v_n) \\
&\qquad + \dots + c_n(a_{1,n}v_1 + a_{2,n}v_2 + \dots + a_{n,n}v_n) \\
&= (c_1a_{1,1} + c_2a_{1,2} + \dots + c_na_{1,n})v_1 + (c_1a_{2,1} + c_2a_{2,2} + \dots + c_na_{2,n})v_2 \\
&\qquad + \dots + (c_1a_{n,1} + c_2a_{n,2} + \dots + c_na_{n,n})v_n.
\end{aligned}$$

Since this last expression is equal to (1.4) and $\mathscr{B}$ is linearly independent, we see that

$$\begin{aligned} c_1 a_{1,1} + c_2 a_{1,2} + \cdots + c_n a_{1,n} &= \lambda c_1, \\ c_1 a_{2,1} + c_2 a_{2,2} + \cdots + c_n a_{2,n} &= \lambda c_2, \\ &\vdots \quad \vdots \\ c_1 a_{n,1} + c_2 a_{n,2} + \cdots + c_n a_{n,n} &= \lambda c_n. \end{aligned}$$

{eq:az:2}

We can rewrite these n equalities in terms of the matrix $A := [f]_{\mathscr{B}} = (a_{i,j})$ and the vector

$$c := [v]_{\mathscr{B}} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \in \Bbbk^n$$

in the form $Ac = \lambda c$ and, in particular, we see that $[v]_{\mathscr{B}}$ is an element of $E_\lambda(A)$. This proves the first statement of (i), and shows that we indeed have a function

$$F : v \in E_\lambda(f) \mapsto [v]_{\mathscr{B}} \in E_\lambda(A).$$

Let us now show that this function is an isomorphism of vector spaces.

- If v and w are two elements of $E_\lambda(f)$ and α and β are two scalars in $\Bbbk$, then

  $$F(\alpha v + \beta w) = [\alpha v + \beta w]_{\mathscr{B}} = \alpha[v]_{\mathscr{B}} + \beta[w]_{\mathscr{B}} = \alpha F(v) + \beta F(w),$$

  because (1.2) holds. This tells us that the function F is linear.
- If $v \in E_\lambda(f)$ is such that $F(v) = [v]_{\mathscr{B}} = 0$, then we know from (1.3) that $v = 0$, and the linear function F is therefore injective.
- Finally, we need to check that the function F is surjective. Let us suppose that

  $$c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$$

  is an element of $E_\lambda(A)$, so that $Ac = \lambda c$. This last equality means that the following n equalities hold:

  $$\begin{aligned} a_{1,1}c_1 + a_{1,2}c_2 + \cdots + a_{1,n}c_n &= \lambda c_1, \\ a_{2,1}c_1 + a_{2,2}c_2 + \cdots + a_{2,n}c_n &= \lambda c_2, \\ &\vdots \quad \vdots \\ a_{n,1}c_1 + a_{n,2}c_2 + \cdots + a_{n,n}c_n &= \lambda c_n. \end{aligned}$$

Let us now consider the element $v \coloneqq c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$ of V. We can compute that

$$\begin{aligned} f(v) &= f(c_1v_1 + c_2v_2 + \cdots + c_nv_n) = c_1 f(v_1) + c_2 f(v_2) + \cdots + c_n f(v_n) \\ &= c_1(a_{1,1}v_1 + a_{2,1}v_2 + \cdots + a_{n,1}v_n) + c_2(a_{1,2}v_1 + a_{2,2}v_2 + \cdots + a_{n,2}v_n) \\ &\qquad + \cdots + c_n(a_{1,n}v_1 + a_{2,n}v_2 + \cdots + a_{n,n}v_n) \\ &= (c_1a_{1,1} + c_2a_{1,2} + \cdots + c_na_{1,n})v_1 + (c_1a_{2,1} + c_2a_{2,2} + \cdots + c_na_{2,n})v_2 \\ &\qquad + \cdots + (c_1a_{n,1} + c_2a_{n,2} + \cdots + c_na_{n,n})v_n \\ &= \lambda c_1 v_2 + \lambda c_2 v_2 + \cdots + \lambda c_n v_n \\ &= \lambda v, \end{aligned}$$

so that $v \in E_\lambda(f)$. Since of course $[v]_{\mathscr{B}} = c$, we see that $c = F(v)$. This shows that the function F is surjective.

With this we have completed the proof of part (i) of the lemma.

A scalar $\lambda \in \mathbb{k}$ is an eigenvalue for the linear map f if and only if the subspace $E_\lambda(f)$ of V is non-zero, and that subspace is isomorphic to $E_\lambda(A)$, so it is non-zero exactly when λ is an eigenvalue for the matrix A. This proves part (ii) of the lemma.

On the other hand, if a vector v in V is an eigenvector for the linear map f with eigenvalue λ, then it is non-zero and belongs to $E_\lambda(f)$, and we know now that implies that $[v]_{\mathscr{B}}$ is non-zero and belongs to $E_\lambda(A)$, so that it is an eigenvector for the matrix A with eigenvalue λ. Conversely, if $v \in V$ is such that $[v]_{\mathscr{B}}$ is an eigenvector for the matrix A with eigenvalue λ, then $[v]_{\mathscr{B}}$ is an element of $E_\lambda(A)$ and according to (i) there is a vector $w \in E_\lambda(f)$ such that $[w]_{\mathscr{B}} = [v]_{\mathscr{B}}$. This implies that $v = w$, so that v is a non-zero element of $E_\lambda(f)$ and, therefore, an eigenvector for f with eigenvalue λ. This proves the third part of the lemma. □

The observations made above allow us to «translate» any problem regarding the eigenvalues and eigenvectors of a linear map into one regarding the eigenvalues and eigenvalues of a matrix, and *vice versa*.

§1.2. Characteristic polynomials

Let n be a positive integer and let A be a matrix in $\mathrm{M}_n(\mathbb{k})$. The ***characteristic polynomial*** of A is the polynomial

$$\chi_A(X) \coloneqq \det(X \cdot I_n - A) \in \mathbb{k}[X].$$

The entries of the matrix $X \cdot I_n - A$ are polynomials — those that appear along its diagonal are the polynomials of degree 1 of the form $X - a_{1,1}, X - a_{2,2}, \dots, X - a_{n,n}$, while those that appear in the other entries are in fact constant polynomials — so the determinant of $X \cdot I_n - A$ is itself a polynomial. We will compute below its degree and some of its coefficients.

Example 1.2.1. If $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \in \mathrm{M}_2(\mathbb{Q})$, then

$$X \cdot I_n - A = \begin{pmatrix} X-1 & -2 \\ -3 & X-4 \end{pmatrix},$$

and therefore the characteristic polynomial of A is

$$\chi_A(X) = \begin{vmatrix} X-1 & -2 \\ -3 & X-4 \end{vmatrix} = (X-1)(X-4) - 6 = X^2 - 5X - 2 \in \mathbb{Q}[X].$$

{ex:chi:tri}

Example 1.2.2. Suppose that $A = (a_{i,j}) \in \mathrm{M}_n(\Bbbk)$ is an upper triangular matrix, so that for all choices of i and j in $\{1, 2, \dots, n\}$ we have that

$$i < j \implies a_{i,j} = 0.$$

In that case, the matrix $X \cdot I_n - A$ is also upper triangular, and therefore its determinant is easy to compute: it is simply the product of the diagonal entries of that matrix. We thus have that

$$\chi_A(X) = \begin{pmatrix} X - a_{1,1} & -a_{1,2} & -a_{1,3} & \cdots & -a_{1,n} \\ 0 & X - a_{2,2} & -a_{2,3} & \cdots & -a_{2,n} \\ 0 & 0 & X - a_{3,3} & \cdots & -a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & X - a_{n,n} \end{pmatrix} = (X - a_{1,1})(X - a_{2,2}) \cdots (X - a_{n,n}).$$

Of course, we can make a similar calculation for lower triangular matrices.

We are interested in the characteristic polynomial of a matrix because its roots are precisely the eigenvalues of the latter. In order to show this fundamental fact this we will use a simple result about polynomials and determinants that we will not prove.

Suppose that $P = (p_{i,j})$ is an $n \times n$ matrix whose entries are polynomials in $\Bbbk[X]$ and that λ is an element of $\Bbbk$. We can then do two things:

- First, we can construct the matrix $(p_{i,j}(\lambda)) \in \mathrm{M}_n(\Bbbk)$ whose entries are the values of the polynomials that appear in the original matrix P evaluated at λ, and then compute its determinant:

$$\det\begin{pmatrix} p_{1,1}(\lambda) & \cdots & p_{1,n}(\lambda) \\ \vdots & \ddots & \vdots \\ p_{n,1}(\lambda) & \cdots & p_{n,n}(\lambda) \end{pmatrix}$$

This is an element of $\Bbbk$.

- On the other hand, we can instead first compute the determinant $\det P$ of the matrix P, which is polynomial in $\Bbbk[X]$, and *then* evaluate the result at λ, obtained again an element of $\Bbbk$ that we can write $(\det P)(\lambda)$.

The result we need is that these two scalars are the same.

Example 1.2.3. Let us consider the matrix $P = \begin{pmatrix} X-1 & X+2 \\ X^2-3 & 7 \end{pmatrix}$ with entries in $\mathbb{Q}[X]$ and the scalar $\lambda = 2 \in \mathbb{Q}$. If we evaluate the entries of P at 2, we obtained the matrix $\begin{pmatrix} 1 & 4 \\ 1 & 7 \end{pmatrix}$, whose determinant is 3. On the other hand, the determinant of P is

$$\begin{vmatrix} X-1 & X+2 \\ X^2-3 & 7 \end{vmatrix} = (X-1)7 - (X+2)(X^2-3) = -X^3 - 2X^2 + 10X - 1$$

and the value of this element of $\mathbb{Q}[X]$ at 2 is also 3, as it should be.

Using this fact we can easily prove what we want:

Proposition 1.2.4. *Let n be a positive integer and let A be a matrix in $\mathrm{M}_n(\Bbbk)$. A scalar $\lambda \in \Bbbk$ is an eigenvalue of A if and only if $\chi_A(\lambda) = 0$.* {prop:chi:mats:eig}

Proof. Let λ be an element of $\Bbbk$. Using the result mentioned above, we see that

the determinant of the matrix $\lambda \cdot I_n - A$ coincides with the value of the polynomial $\chi_A(X) = \det(X \cdot I_n - A)$ at λ. (1.5) {eq:eigdet}

We know from Proposition 1.1.4 that λ is an eigenvalue of A if and only if the determinant of $\lambda \cdot I_n - A$ is zero, and (1.5) tells us that this happens exactly when $\chi_A(\lambda) = 0$: this is precisely what the proposition claims. □

In cases where we can compute the characteristic polynomial of a matrix, this proposition gives us an efficient way to check whether a given scalar is an eigenvalue of the matrix or not.

Example 1.2.5. Let us consider the matrix $A = \begin{pmatrix} -1 & -1 \\ 2 & -2 \end{pmatrix}$ in $\mathrm{M}_2(\mathbb{C})$. Its characteristic polynomial is

$$\chi_A(X) = \det\begin{pmatrix} X+1 & 1 \\ -2 & X+2 \end{pmatrix} = (X+1)(X+2)+2 = X^2+3X+4 \in \mathbb{C}[X],$$

and the roots of this polynomial are the numbers $(-3+i\sqrt{7})/2$ and $(-3-i\sqrt{7})/2$. The proposition tells us that these two numbers are precisely the eigenvalues of A.

If we instead view the matrix A as an element of $\mathrm{M}_2(\mathbb{Q})$ then the characteristic polynomial is still $\chi_A(X) = X^2+3X+4$, now an element of $\mathbb{Q}[X]$, but is irreducible, so it has no roots in $\mathbb{Q}$. This tells us that A has no eigenvalues as a matrix over $\mathbb{Q}$.

Example 1.2.6. Let n be a positive integer and let A be a upper triangular matrix in $\mathrm{M}_n(\mathbb{k})$,

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ 0 & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ 0 & 0 & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{n,n} \end{pmatrix}.$$

As we noted in Example 1.2.2, the characteristic of polynomial of A is then

$$\chi_A(X) = (X-a_{1,1})(X-a_{2,2})\cdots(X-a_{n,n}) \in \mathbb{k}[X]$$

and therefore the roots of χ_A and the eigenvalues of A are the scalars $a_{1,1}$, $a_{2,2}$, …, $a_{n,n}$ that appear along the diagonal of the matrix.

In Proposition 1.2.8 below we will partially describe the characteristic polynomial of a matrix. To prove it we will use the following simple observation about polynomials:

Lemma 1.2.7. *Let n be a positive integer. If a_1, a_2, …, a_n are elements of $\mathbb{k}$, then the product $(X-a_1)(X-a_2)\dots(X-a_n)$ is a monic polynomial of degree n in which the coefficient of X^{n-1} is $-(a_1+a_2+\cdots+a_n)$ and whose constant coefficient is $(-1)^n a_1 a_2 \cdots a_n$.*

Proof. We will prove the lemma by induction with respect to the integer n. When n is 1, the claim is immediate, so there is nothing to do in that case. Let then k be a positive integer, let us suppose that the claim of the lemma is true when n is k, and let a_1, a_2, …, a_{k+1} be $k+1$ elements of $\mathbb{k}$. The hypothesis implies that the product $(X-a_1)(X-a_2)\cdots(X-a_k)$ is a monic polynomial of degree k in which the coefficient of X^{k-1} is $-(a_1+a_2+\cdots+a_k)$ and the constant coefficient is

$(-1)^k a_1 a_2 \cdots a_k$. In other words, there are elements $b_1, b_2, \dots, b_{k-2}$ in $\Bbbk$ such that

$$\begin{aligned}&(X-a_1)(X-a_2)\cdots(X-a_k)\\ &\qquad\qquad = X^k - (a_1+a_2+\cdots+a_k)X^{k-1} + b_{k-2}X^{k-2} + \cdots + b_1 X + (-1)^k a_1 a_2 \cdots a_k.\end{aligned}$$

We therefore have that

$$\begin{aligned}&(X-a_1)(X-a_2)\cdots(X-a_k)(X-a_{k+1})\\ &\quad = \Big(X^k - (a_1+a_2+\cdots+a_k)X^{k-1} + b_{k-2}X^{k-2} + \cdots + b_1 X + (-1)^k a_1 a_2 \cdots a_k\Big)(X-a_{k+1})\end{aligned}$$

and distributing we see immediately that this is

$$= X^{k+1} - (a_1+a_2+\cdots+a_k+a_{k+1})X^{k-1} + \cdots + (-1)^{k+1} a_1 a_2 \cdots a_k a_{k+1}.$$

This tells us that the claim of the lemma is also true when n is $k+1$ and, by induction, that it is actually true for all positive integers n. □

Using this lemma and Leibniz's formula for the determinant of a matrix we can easily compute the degree of the characteristic polynomial of a matrix and three of its coefficients.

Proposition 1.2.8. *Let n be a positive integer and let A be a matrix in $\mathrm{M}_n(\Bbbk)$. The characteristic polynomial χ_A of A is monic and has degree exactly n. Moreover, the coefficient of X^{n-1} in χ_A is equal to $-\operatorname{tr} A$, and the constant coefficient is $(-1)^n \det A$, so that* {prop:chi:mats:deg

$$\chi_A(X) = X^n - \operatorname{tr} A \cdot X + \cdots + (-1)^n \det A.$$

Proof. Let $B = (b_{i,j})$ be the matrix $X \cdot I_n - A$, so that the characteristic polynomial χ_A of A is $\det B$. According to Leibniz's formula, we have

$$\det B = \sum_{\sigma} \operatorname{sgn}(\sigma) \cdot b_{1,\sigma(1)} b_{2,\sigma(2)} \dots b_{n,\sigma(n)}. \tag{1.6}$$ {eq:lei:1}

In this sum there is one term for each permutation σ of the set $[\![n]\!]$, and for each such permutation σ we have written $\operatorname{sgn}(\sigma)$ for the *sign* of σ. Let us look in some detail at the terms of the sum.

- If σ is the identity permutation id, so that $\sigma(i) = i$ for all $i \in [\![n]\!]$, then we have $\operatorname{sgn}(\sigma) = 1$ and the term corresponding to σ in the sum above is

  $$\operatorname{sgn}(\sigma) \cdot b_{1,\sigma(1)} b_{2,\sigma(2)} \dots b_{n,\sigma(n)} = b_{1,1} b_{2,2} \cdots b_{n,n} = (X - a_{1,1})(X - a_{2,2}) \dots (X - a_{2,2}).$$

  According to the lemma, this is a monic polynomial of degree n in which the coefficient of X^{n-1} is $-(a_{1,1} + a_{2,2} + \cdots + a_{n,n}) = -\operatorname{tr} A$.

- Let now σ be a permutation of $[\![n]\!]$ that is not the identity permutation. There is then an element k in $[\![n]\!]$ such that $\sigma(k) \neq k$. As σ is an injective function $[\![n]\!] \to [\![n]\!]$, we also have that $\sigma(\sigma(k)) \neq \sigma(k)$: we thus see that the set $\{i \in [\![n]\!] : \sigma(i) = i\}$ has at most $n-2$ elements, so that in the product

  $$b_{1,\sigma(1)} b_{2,\sigma(2)} \dots b_{n,\sigma(n)}$$

  at most $n-2$ of the factors are diagonal entries of the matrix B and therefore are polynomials of degree 1, while the remaining factors are all elements of $\Bbbk$. It follows immediately from this that this product, and thus the term corresponding to the permutation σ in the sum (1.6), have both degree at most equal to $n-2$.

Of course, we have that

$$\det B = b_{1,1} b_{2,2} \dots b_{n,n} + \sum_{\sigma \neq \mathrm{id}} \operatorname{sgn}(\sigma) \cdot b_{1,\sigma(1)} b_{2,\sigma(2)} \dots b_{n,\sigma(n)}.$$

Our first observation above tells us that the first summand here is a monic polynomial of degree n in which the coefficient of X^{n-1} is $-\operatorname{tr} A$, and our second observation implies that the second summand is a polynomial of degree at most $n-2$. It follows from all this that $\det B$ itself is a monic polynomial of degree n in which the coefficient of X^{n-1} is $-\operatorname{tr} A$, and thus that the claim of the proposition holds. □

This proposition has the following very important corollary:

{coro:chi:mats:n}

Corollary 1.2.9. *Let n be a positive integer. A matrix in $\mathrm{M}_n(\Bbbk)$ has at most n eigenvalues in $\Bbbk$.*

Proof. Indeed, if A is a matrix in $\mathrm{M}_n(\Bbbk)$, then the characteristic polynomial of χ_A has degree n, and the eigenvalues of A are precisely the roots of χ_A in $\Bbbk$, so there are at most n of them. □

In the following example we will compute the characteristic polynomials of a family of matrices that will be very important later.

{ex:companion}

Example 1.2.10. Let n be a positive integer, and let $p = a_0 + a_1X + \cdots + a_{n-1}X^{n-1} + X^n$ be a monic polynomial of degree n. The ***companion matrix*** of the polynomial p is the $n \times n$ matrix

$$C(p) \coloneqq \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & -a_{n-3} \\ 0 & 0 & 0 & \cdots & 1 & 0 & -a_{n-2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -a_{n-1} \end{pmatrix}$$

For example, when n is 1, 2 or 3, the matrices $C(p)$ are, respectively,

$$\begin{pmatrix} -a_0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & -a_0 \\ 1 & -a_1 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 & -a_0 \\ 1 & 0 & -a_1 \\ 0 & 1 & -a_2 \end{pmatrix}.$$

We want to show that

> *the characteristic polynomial $\chi_{C(p)}$ of the companion matrix of p is precisely p.*

To do this we need to compute the determinant of the matrix

$$X \cdot I_n - C(p) = \begin{pmatrix} X & 0 & 0 & \cdots & 0 & 0 & a_0 \\ -1 & X & 0 & \cdots & 0 & 0 & a_1 \\ 0 & -1 & X & \cdots & 0 & 0 & a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & X & 0 & a_{n-3} \\ 0 & 0 & 0 & \cdots & -1 & X & a_{n-2} \\ 0 & 0 & 0 & \cdots & 0 & -1 & a_{n-1} + X \end{pmatrix}$$

If we add to the first row of this matrix its second row multiplied by X, its third row multiplied by X^2, and so on up to its nth row multiplied by X^{n-1} we obtain the matrix

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & p(X) \\ -1 & X & 0 & \cdots & 0 & 0 & a_1 \\ 0 & -1 & X & \cdots & 0 & 0 & a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & X & 0 & a_{n-3} \\ 0 & 0 & 0 & \cdots & -1 & X & a_{n-2} \\ 0 & 0 & 0 & \cdots & 0 & -1 & a_{n-1} + X \end{pmatrix}$$

Now, using Laplace's formula to expand this determinant along its first row we see that that determinant is

$$(-1)^{n+1}p(X)\begin{vmatrix} -1 & X & 0 & \cdots & 0 & 0 \\ 0 & -1 & X & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & X & 0 \\ 0 & 0 & 0 & \cdots & -1 & X \\ 0 & 0 & 0 & \cdots & 0 & -1 \end{vmatrix} = (-1)^{n+1}p(X)(-1)^{n-1} = p(X),$$

because the last matrix is upper triangular. This proves what we wanted.

{ex:chi:but}

Exercise 1.2.11. Let A be a matrix in $\mathrm{M}_n(\Bbbk)$ that is block upper triangular, that is, such that there is an integer m with $0 < m < n$ and matrices $A_{1,1} \in \mathrm{M}_m(\Bbbk)$, $A_{2,2} \in \mathrm{M}_{n-m}(\Bbbk)$ and $A_{1,2} \in \mathrm{M}_{m,n-m}(\Bbbk)$ such that

$$A = \begin{pmatrix} A_{1,1} & A_{1,2} \\ 0 & A_{2,2} \end{pmatrix}.$$

Prove that $\chi_A(X) = \chi_{A_{1,1}}(X) \cdot \chi_{A_{2,2}}(X)$. Extend this results for general block upper triangular matrices of the form

$$\begin{pmatrix} A_{1,1} & A_{1,2} & A_{1,3} & \cdots & A_{1,n-1} & A_{1,n} \\ 0 & A_{2,2} & A_{2,3} & \cdots & A_{2,n-1} & A_{2,n} \\ 0 & 0 & A_{3,3} & \cdots & A_{3,n-1} & A_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & A_{n-1,n-1} & A_{n-1,n} \\ 0 & 0 & 0 & \cdots & 0 & A_{n,n} \end{pmatrix}$$

for some choice of positive integers $m_1, \ldots, m_n$ and matrices $A_{i,j} \in \mathrm{M}_{m_i,m_j}(\Bbbk)$ for each $i, j \in \{1, 2, \ldots, n\}$.

We have defined the characteristic polynomial of matrices, and we will next define characteristic polynomials for linear maps. We start with a simple observation:

Lemma 1.2.12. *Let n be a positive integer. If A and B are two matrices in $\mathrm{M}_n(\Bbbk)$ that are similar, then the characteristic polynomials of A and of B are equal.*

In view of Proposition 1.2.8, this lemma implies in particular that two similar matrices have the same determinant and the same trace — something that we already know, of course.

Proof. Indeed, if A and B are matrices in $\mathrm{M}_n(\Bbbk)$ that are similar, so that there exists an invertible matrix C in $\mathrm{M}_n(\Bbbk)$ such that $A = CBC^{-1}$, then characteristic polynomial of A is

$$\begin{aligned}\chi_A(X) &= \det(X \cdot I_n - A) = \det(X \cdot CC^{-1} - CBC^{-1}) = \det C(X \cdot I_n - B)C^{-1} \\ &= \det C \cdot \det(X \cdot I_n - B) \cdot \det C^{-1} = \det(X \cdot I_n - B) = \chi_B(X).\end{aligned}$$ □

Suppose now that V is a non-zero finite-dimensional vector space over a field $\Bbbk$, let n be its dimension, and let $\mathscr{B} = (v_1, 2, \dots, v_n)$ be an ordered basis for V. We can then construct the matrix $[f]_{\mathscr{B}}$ of the linear map f with respect to the ordered basis $\mathscr{B}$, and define the ***characteristic polynomial*** χ_f of f to be the characteristic polynomial of the matrix $[f]_{\mathscr{B}}$,

$$\chi_f \coloneqq \chi_{[f]_{\mathscr{B}}}.$$

Of course, for this to make sense we need to verify that the polynomial $\chi_{[f]_{\mathscr{B}}}$ depends only on the linear map f and not on the choice of the ordered basis $\mathscr{B}$. To do that, let us suppose that $\mathscr{B}' = (v'_1, \dots, v'_n)$ is another ordered basis for V. We know then that there exists an invertible matrix $C = (c_{i,j})$ in $\mathrm{M}_n(\Bbbk)$, the ***change of basis matrix***, such that $v'_i = \sum_{j=1}^{n} c_{j,i} v_j$ for each $i \in \{1, 2, \dots, n\}$, and that moreover $[f]_{\mathscr{B}} \cdot C = C \cdot [f]_{\mathscr{B}'}$. It follows from this, of course, that $[f]_{\mathscr{B}} = C \cdot [f]_{\mathscr{B}'} \cdot C^{-1}$ and, according to the lemma, that the two matrices $[f]_{\mathscr{B}}$ and $[f]_{\mathscr{B}'}$ have the same characteristic polynomial. This proves what we needed.

The following proposition combines the results of Proposition 1.2.4, Proposition 1.2.8, and Corollary 1.2.9 and reinterprets them in terms of linear maps.

{prop:chi:f}

Proposition 1.2.13. *Let V be a finite-dimensional vector space, let n be the dimension of V, and let $f : V \to V$ be a linear map.*

(*i*) *A scalar λ is an eigenvalue for f if and only if it is a root of the characteristic polynomial χ_f.*

(*ii*) *The characteristic polynomial χ_f is monic and of degree exactly n. The coefficient of X^{n-1} in χ_f is $-\operatorname{tr} f$ and the constant term is $(-1)^n \det f$.*

(*iii*) *The map f has at most n eigenvalues.* □

Exercise 1.2.14. Prove the proposition.

Observation 1.2.15. We have defined the characteristic polynomial of an endomorphism $f : V \to V$ only when the vector space V is *finite-dimensional*, and our definition only makes sense in that situation. Indeed, if V is infinite-dimensional then any «matrix» for f with respect to a basis of V will necessarily be an infinite matrix, and we cannot compute determinants of such a thing.

§1.3. The linear independence of eigenvectors

The purpose of this section is to establish a fundamental property of eigenvectors. To do that we need the following simple lemma.

{lemma:pfv}

Lemma 1.3.1. *Let $f : V \to V$ be an endomorphism of a vector space, and let $v \in V$ be an eigenvector of f with eigenvalue $\lambda \in \Bbbk$. If $p \in \Bbbk[X]$ is a polynomial, then $p(f)(v) = p(\lambda) \cdot v$.*

Proof. As $f(v) = \lambda v$, we have that $f^2(v) = f(f(v)) = f(\lambda v) = \lambda^2 v$ and, more generally, that $f^i(v) = \lambda^i v$ for all $i \in \mathbb{N}_0$. Let now $p \in \Bbbk[X]$ be a polynomial, so that there is a non-negative integer d and scalars $a_0, a_1, a_2, \dots, a_d \in \Bbbk$ such that $p(X) = a_0 + a_1 X + a_d f^2 + \cdots + a_d X^d$. We can then compute that

$$\begin{aligned} p(f)(v) &= (a_0 \mathrm{id}_V + a_1 f + a_2 f^2(v) + \cdots + a_d f^d)(v) \\ &= a_0 v + a_1 f(v) + a_2 f^2(v) + \cdots + a_d f^d(v) \\ &= a_0 v + a_1 \lambda v + a_2 \lambda^2 v + \cdots + a_d \lambda^d v \\ &= (a_0 + a_1 \lambda + a_2 \lambda^2 + \cdots + a_d \lambda^d) v \\ &= p(\lambda) \cdot v, \end{aligned}$$

and this is precisely what the lemma claims. □

Using this lemma we can prove that eigenvectors with different eigenvalues are always linearly independent.

{prop:eig:li}

Proposition 1.3.2. *Let $f : V \to V$ be an endomorphism of a linear map, let $v_1, v_2 \dots, v_k$ be eigenvectors of f, and let $\lambda_1, \lambda_2, \dots, \lambda_k \in \Bbbk$ be the corresponding eigenvalues. If the scalars $\lambda_1, \lambda_2, \dots, \lambda_k$ are pairwise different, then the vectors $v_1, v_2, \dots, v_k$ are linearly independent.*

Proof. Let us suppose the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$ are pairwise different, and that we have scalars $a_1, a_2, \dots, a_k \in \Bbbk$ such that

$$0 = a_1 v_1 + a_2 v_2 + \cdots + a_k v_k. \qquad \text{\{eq:lip\}}$$

Let i be an element of $\{1, 2, \dots, k\}$, and let us consider the polynomial

$$p_i(X) = (X - \lambda_1) \cdots (X - \lambda_{i-1})(X - \lambda_{i+1}) \cdots (X - \lambda_k) \in \Bbbk[X]$$

with is the product of all the differences $X - \lambda_j$ with $1 \leq j \leq k$ and $j \neq i$. If $l \in \{1, 2, \dots, k\}$, then the lemma implies at once that

$$p_i(f)(v_l) = p_i(\lambda_l) \cdot v_l = \begin{cases} 0 & \text{if } l \neq i; \\ p_i(\lambda_i) \cdot v_i & \text{if } l = i. \end{cases}$$

It follows from this that

$$0 = p_i(f)(a_1v_1 + a_2v_2 + \cdots + a_kv_k) = a_1p_i(f)(v_1) + a_2p_i(f)(v_2) + \cdots + a_kp_i(f)(v_k)$$
$$= a_ip_i(\lambda_i)\cdot v_i.$$

Since $v_i \neq 0$ and $p_i(\lambda_i) \neq 0$, this allows us to conclude that $a_i = 0$ and, since this is so for all choices of i in $\{1, 2, \ldots, k\}$, it proves that the vectors $v_1, v_2, \ldots, v_k$ are linearly independent. □

Eigenvectors are by definition non-zero vectors, and we used this in the proof of the proposition. We can deduce from the proposition a slightly different statement which is often useful:

{coro:eig:lix}

Corollary 1.3.3. *Let $f : V \to V$ be an endomorphism of a linear map, and let $\lambda_1, \lambda_2, \ldots, \lambda_k \in \Bbbk$ be pairwise different eigenvalues of f. If $v_1 \in E_{\lambda_1}(f)$, $v_2 \in E_{\lambda_2}(f)$, …, $v_k \in E_{\lambda_k}(f)$ are such that $v_1 + v_2 + \cdots + v_k = 0$, then $v_1 = v_2 = \cdots = v_k = 0$.*

Proof. Let $v_1 \in E_{\lambda_1}(f), v_2 \in E_{\lambda_2}(f), \ldots, v_k \in E_{\lambda_k}(f)$ be such that $v_1 + v_2 + \cdots + v_k = 0$, let us suppose that the set $I := \{i \in \{1, 2, \ldots, k\} : v_i \neq 0\}$ is not empty, and let $i_1, \ldots, i_r$ be its elements listed in increasing order. We then have that $v_{i_1}, v_{i_2}, \ldots, v_{i_r}$ are eigenvectors of f with corresponding eigenvalues $\lambda_{i_1}, \lambda_{i_2}, \ldots, \lambda_{i_r}$, and that $v_{i_1} + v_{i_2} + \cdots + v_{i_r} = 0$: this is absurd, since the proposition tells us that the vectors $v_{i_1}, v_{i_2}, \ldots, v_{i_r}$ are linearly independent. We thus see that the set I is empty, and the corollary is therefore true. □

Using Proposition 1.3.2 we can give a different proof of the last part of Proposition 1.2.13:

Corollary 1.3.4. *An endomorphism of a finite-dimensional vector space V has at most* $\dim V$ *eigenvalues.*

Proof. Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space V, and let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be pairwise different eigenvalues of f. There exist then eigenvectors $v_1, v_2, \ldots, v_k$ of f with those eigenvalues, and the proposition tells us that these r vectors are linearly independent. It follows from this that their number k is at most $\dim V$, and this proves the corollary. □

A more technical consequence of the linear independence of eigenvectors that will be extremely useful below is the following one:

{coro:concat-bases

Corollary 1.3.5. *Let $f : V \to V$ be an endomorphism of a linear map, let $\lambda_1, \ldots, \lambda_k \in \Bbbk$ be pairwise different eigenvalues of f, let $E_{\lambda_1}(f), \ldots, E_{\lambda_k}(f)$ be the corresponding eigenspaces, and let $d_1, \ldots, d_k$ be the dimensions of these subspaces. If for each $i \in \{1, 2, \ldots, k\}$ the sequence $\mathscr{B}_i = (v_{i,1}, \ldots, v_{i,d_i})$*

is an ordered basis for $E_{\lambda_i}(f)$, then the sequence

$$\mathscr{B} = (\underbrace{v_{1,1}, v_{1,2}, \dots, v_{1,d_1}}_{\mathscr{B}_1}, \underbrace{v_{2,1}, v_{2,2}, \dots, v_{2,d_2}}_{\mathscr{B}_2}, \dots, \dots, \underbrace{v_{k,1}, v_{k,2}, \dots, v_{k,d_k}}_{\mathscr{B}_k})$$

formed by concatenating the ordered bases $\mathscr{B}_1$, $\mathscr{B}_2$, …, $\mathscr{B}_k$ in order is an ordered basis for the subspace

$$E_{\lambda_1}(f) + E_{\lambda_2}(f) + \dots + E_{\lambda_k}(f)$$

of V and, in particular, the dimension of this sum is $d_1 + d_2 + \dots + d_k$.

Proof. Let us start by verifying that the vectors that appear in the sequence $\mathscr{B}$ are linearly independent. Let us suppose that we have scalars $a_{1,1}$, $a_{1,2}$, …, a_{1,d_1}, $a_{2,1}$, $a_{2,2}$, …, a_{2,d_2}, …, $a_{k,1}$, $a_{k,2}$, …, a_{k,d_k} in $\Bbbk$ such that

$$\begin{multline} a_{1,1}v_{1,1} + a_{1,2}v_{1,2} + \dots + a_{1,d_1}v_{1,d_1} + a_{2,1}v_{2,1} + a_{2,2}v_{2,2} + \dots + a_{2,d_2}v_{2,d_2} \\ + \dots + \dots + a_{k,1}v_{k,1} + a_{k,2}v_{k,2} + \dots + a_{k,d_k}v_{k,d_k} = 0. \qquad (1.7) \end{multline}$$

{eq:rrx}

For each $i \in \{1, 2, \dots, k\}$ we consider the vector $w_i \coloneqq a_{i,1}v_{i,1} + a_{i,2}v_{i,2} + \dots + a_{i,d_i}v_{i,d_i}$, which belongs to $E_{\lambda_i}(f)$. The equality above then tells us that $w_1 + w_2 + \dots + w_k = 0$, and then Corollary 1.3.3 allows us to conclude that in fact the k vectors $w_1, \dots, w_k$ are all zero. If now i is an element of $\{1, 2, \dots, k\}$, then we have that $a_{i,1}v_{i,1} + a_{i,2}v_{i,2} + \dots + a_{i,d_i}v_{i,d_i} = 0$ and, since $\mathscr{B}_i = (v_{i,1}, v_{i,2}, \dots, v_{i,d_i})$ is an ordered basis for $E_{\lambda_i}(f)$, we have that all the scalars $a_{i,1}$, $a_{i,2}$, …, a_{i,d_i} are zero. All the coefficients in the linear relation (1.7) are zero, and this proves the linear independence of the sequence of vectors $\mathscr{B}$.

To complete the proof we need to show that the sequence $\mathscr{B}$ generates the subspace $U \coloneqq E_{\lambda_1}(f) + E_{\lambda_2}(f) + \dots + E_{\lambda_k}(f)$. First of all, each element of the sequence $\mathscr{B}$ belongs to one of the basis $\mathscr{B}_1, \mathscr{B}_2, \dots, \mathscr{B}_k$ and therefore to one of the subspaces $E_{\lambda_1}(f), E_{\lambda_2}(f), \dots, E_{\lambda_k}(f)$, so to the subspace U. This implies, of course, that $\langle \mathscr{B} \rangle \subseteq U$.

Let now u be an element of U, so that there are $u_1 \in E_{\lambda_1}(f)$, $u_2 \in E_{\lambda_2}(f), \dots, u_k \in E_{\lambda_k}(f)$ such that $u = u_1 + u_2 + \dots + u_k$. Since $\mathscr{B}_1, \mathscr{B}_2, \dots, \mathscr{B}_k$ are ordered bases of $E_{\lambda_1}(f), E_{\lambda_2}(f), \dots, E_{\lambda_k}(f)$, we have that $u_i \in \langle \mathscr{B}_i \rangle \subseteq \langle \mathscr{B} \rangle$ for each $i \in \{1, 2, \dots, k\}$, and therefore that $u = u_1 + u_2 + \dots + u_k \in \langle \mathscr{B} \rangle$. This tells us that $U \subseteq \langle \mathscr{B} \rangle$ and, putting everything together, that $U = \langle \mathscr{B} \rangle$, as we wanted. □

§1.4. Diagonalizability

We say that an endomorphism $f : V \to V$ of a vector space is ***diagonalizable*** if there is a basis $\mathscr{B}$ of V whose elements are eigenvectors for f. This condition means, in a sense, that there are «enough» eigenvectors of V, and the following result makes this clear:

Proposition 1.4.1. *An endomorphism $f : V \to V$ of a vector space V is diagonalizable if and only if every element of V is a linear combination of eigenvectors of f.* {prop:diag:2}

To prove this result we will use the fact that every subset of a vector space that spans it contains a basis. This is true for all vector spaces, even infinite-dimensional ones — although it is possible that the reader has only seen it proved under the additional hypothesis of finite-dimensionality. For most of our purposes, the extra generality can be ignored.

Proof. Let $f : V \to V$ be an endomorphism of a vector space V, and let us suppose first that f is diagonalizable, so that there is a basis $\mathscr{B}$ of V whose elements are eigenvectors of f. If v is an arbitrary element of V, then there exists elements $v_1, v_2, \dots, v_k$ of $\mathscr{B}$ and scalars $a_1, a_2, \dots, a_k$ in $\Bbbk$ such that $v = a_1v_1 + a_2v_2 + \cdots + a_kv_k$, and therefore v is a linear combination of eigenvectors of f. This shows that the condition given by the proposition is necessary for the endomorphism f to be diagonalizable.

It is also sufficient. Indeed, if it is satisfied, then V is the span of the set of eigenvectors of f, so this set contains a basis of V and therefore the map f is diagonalizable. □

The choice of name for this notion is explained by the following lemma.

Lemma 1.4.2. *Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space V and let $\mathscr{B}$ be an ordered basis for V. The elements of $\mathscr{B}$ are eigenvectors for f if and only if the matrix $[f]_{\mathscr{B}}$ is diagonal.*

In other words, the endomorphism f is diagonalizable if we can pick an ordered basis of V with respect to which the matrix of f is diagonal.

Proof. Let the ordered basis $\mathscr{B}$ be $(v_1, v_2, \dots, v_n)$. If the elements of $\mathscr{B}$ are eigenvectors for f, then there exist scalars $\lambda_1, \lambda_2, \dots, \lambda_n$ in $\Bbbk$ such that $f(v_i) = \lambda_i v_i$ for each $i \in \{1, 2, \dots, n\}$, and in that case the matrix of f with respect to $\mathscr{B}$ is clearly

$$[f]_{\mathscr{B}} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix},$$

which is a diagonal matrix. This proves that the condition in the lemma is necessary.

Suppose now that the matrix $[f]_{\mathscr{B}}$ is $(a_{i,j}) \in \mathrm{M}_n(\Bbbk)$ is diagonal: this means precisely that for each $i \in \{1, 2, \dots, n\}$ we have $f(v_i) = a_{i,i}v_i$, so that v_i, being non-zero, is an eigenvector for f of eigenvalue $a_{i,i}$. This shows that the condition in the lemma is also sufficient. □

The general idea is that a linear map is diagonalizable exactly it possesses «enough» eigenvectors to generate its domain. The following characterization of diagonalizability is a concrete version of this statement.

{prop:diag:1}

Proposition 1.4.3. *Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space V, let $\lambda_1, \lambda_2, \dots, \lambda_k \in \Bbbk$ be the eigenvalues of f listed without repetitions, and let $E_{\lambda_1}(f), E_{\lambda_2}(f), \dots, E_{\lambda_k}(f)$ be the corresponding eigenspaces. The following statements are equivalent:*

(*a*) *The endomorphism f is diagonalizable.*

(*b*) $V = E_{\lambda_1}(f) + E_{\lambda_2}(f) + \dots + E_{\lambda_k}(f)$.

(*c*) $\dim V = \dim E_{\lambda_1}(f) + \dim E_{\lambda_2}(f) + \dots + \dim E_{\lambda_k}(f)$.

Proof. Let us write $W := E_{\lambda_1}(f) + E_{\lambda_2}(f) + \dots + E_{\lambda_k}(f)$.

(*a*) ⇒ (*b*) Let us suppose that the endomorphism f is diagonalizable, so that there is a ordered basis $\mathscr{B} = (v_1, v_2, \dots, v_n)$ of V whose elements are eigenvalues for f. If $i \in \{1, 2, \dots, n\}$, then the vector v_i is an eigenvector for f: since the scalars $\lambda_1, \lambda_2, \dots, \lambda_k$ are the eigenvalues of f, there exists an index $j \in \{1, 2, \dots, k\}$ such that $f(v_i) = \lambda_j v_i$, and therefore $v_i \in E_{\lambda_j}(f) \subseteq W$. We see that the subspace W contains all the elements of the basis $\mathscr{B}$: as $\mathscr{B}$ spans V, this implies that, in fact, V is contained in W, so that $V = W$.

(*b*) ⇒ (*c*) Let us now suppose that $V = W$, and for each $i \in \{1, 2, \dots, k\}$ let $d_i := \dim E_{\lambda_i}(f)$ and pick an ordered basis $\mathscr{B}_i = (v_{i,1}, v_{i,2}, \dots, v_{i,d_i})$ for $E_{\lambda_i}(f)$. The sequence

$$\mathscr{B} = (\underbrace{v_{1,1}, v_{1,2}, \dots, v_{1,d_1}}_{\mathscr{B}_1}, \underbrace{v_{2,1}, v_{2,2}, \dots, v_{2,d_2}}_{\mathscr{B}_2}, \dots\dots\dots, \underbrace{v_{k,1}, v_{k,2}, \dots, v_{k,d_k}}_{\mathscr{B}_k})$$

formed by concatenating the ordered bases $\mathscr{B}_1, \mathscr{B}_2, \dots, \mathscr{B}_k$ in order is, according to Corollary 1.3.5, a basis for W, and, in view of our hypothesis, of V: the dimension of V is then equal to the length of $\mathscr{B}$, which is, of course, equal to

$$d_1 + d_2 + \dots + d_k = \dim E_{\lambda_1}(f) + \dim E_{\lambda_2}(f) + \dots + \dim E_{\lambda_k}(f).$$

(*c*) ⇒ (*a*) Let us now suppose that $\dim V = \dim E_{\lambda_1}(f) + \dim E_{\lambda_2}(f) + \dots + \dim E_{\lambda_k}(f)$. Just as before, for each index $i \in \{1, 2, \dots, k\}$ we let $d_i := \dim E_{\lambda_i}(f)$, pick an ordered basis $\mathscr{B}_i = (v_{i,1}, v_{i,2}, \dots, v_{i,d_i})$ for $E_{\lambda_i}(f)$, and construct the sequence

$$\mathscr{B} = (v_{1,1}, v_{1,2}, \dots, v_{1,d_1}, v_{2,1}, v_{2,2}, \dots, v_{2,d_2}, \dots\dots\dots, v_{k,1}, v_{k,2}, \dots, v_{k,d_k}).$$

We know from Corollary 1.3.5 that $\mathscr{B}$ is linearly independent. On the other hand, its length is $d_1 + d_2 + \cdots + d_k = \dim V$, so it is in fact a basis for V: as its elements are eigenvectors for f, this shows that the endomorphism f is diagonalizable. □

We say that a matrix $A \in \mathrm{M}_n(\Bbbk)$ is ***diagonalizable*** if there is an invertible matrix $C \in \mathrm{M}_n(\Bbbk)$ such that the product CAC^{-1} is a diagonal matrix. This notion is connected with the one of diagonalizability for linear maps in the usual way:

Lemma 1.4.4. *Let n be a positive integer. A matrix $A \in \mathrm{M}_n(\Bbbk)$ is diagonalizable if and only if the corresponding linear map $f_A : x \in \Bbbk^n \mapsto Ax \in \Bbbk^n$ is diagonalizable.*

Proof. Let A be a matrix in $\mathrm{M}_n(\Bbbk)$, and let us suppose that the linear map $f_A : x \in \Bbbk^n \mapsto Ax \in \Bbbk^n$ is diagonalizable, so that there is an ordered basis $\mathscr{B} = (v_1, v_2, \dots, v_n)$ of $\Bbbk^n$ whose elements are eigenvectors of f_A, and let $\lambda_1, \lambda_2, \dots, \lambda_n \in \Bbbk$ be the corresponding eigenvalues. Let P be the matrix in $\mathrm{M}_n(\Bbbk)$ whose columns are the vectors $v_1, \dots, v_n$, and let D be the diagonal matrix there whose diagonal entries are, in order, the scalars $\lambda_1, \lambda_2, \dots, \lambda_n$.

The matrix P is an invertible matrix because its columns are linearly independent. On the other hand, the matrix AP has as columns the vectors $Av_1, Av_2, \dots, Av_n$, which coincide with the vectors $f_A(v_1), f_A(v_2), \dots, f_A(v_n)$, and these are precisely the vectors $\lambda_1 v_1, \lambda_2 v_2, \dots, \lambda_n v_n$: a direct calculation shows that there are the columns of the matrix PD, so that $AP = PD$. If we let $C \coloneqq P^{-1}$, this tells us that CAC^{-1} is D, a diagonal matrix, so that the matrix A is diagonalizable.

Let us now suppose, to prove the converse implication, that the matrix A is diagonalizable, so that there is an invertible matrix C such that the matrix $D \coloneqq CAC^{-1}$ is diagonal. Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the diagonal entries of the matrix D in order, and let $v_1, v_2, \dots, v_n$ be the columns of the matrix C^{-1}. Clearly the sequence $\mathscr{B} \coloneqq (v_1, v_2, \dots, v_n)$ is a basis for $\Bbbk^n$. The equality $CAC^{-1} = D$ implies that $f_A(v_) = Av_i = \lambda_i v_i$ for each $i \in \{1, 2, \dots, n\}$, so that the vectors $v_1, v_2, \dots, v_n$ are eigenvectors. We thus see that the linear map f_A is diagonalizable, as there is an ordered basis for V whose elements are eigenvectors of f_A. □

As a consequence of this, we have a version of Proposition 1.4.3 for matrices:

Proposition 1.4.5. *Let A be a matrix in $\mathrm{M}_n(\Bbbk)$, let let $\lambda_1, \lambda_2, \dots, \lambda_k \in \Bbbk$ be the eigenvalues of A listed without repetitions, and let $E_{\lambda_1}(A), E_{\lambda_2}(A), \dots, E_{\lambda_k}(A)$ be the corresponding eigenspaces. The following statements are equivalent:*

(*a*) *The matrix A is diagonalizable.*

(*b*) $V = E_{\lambda_1}(A) + E_{\lambda_2}(A) + \cdots + E_{\lambda_k}(A)$.

(*c*) $\dim V = \dim E_{\lambda_1}(A) + \dim E_{\lambda_2}(A) + \cdots + \dim E_{\lambda_k}(A)$. □

Exercise 1.4.6. Provide the details of the proof of this proposition.

§1.5. Minimal polynomials

The following result is fundamental in all that follows.

{prop:min:1}

Proposition 1.5.1. *Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space V. There exists a unique monic polynomial $\mu \in \Bbbk[X]$ such that*

- *$\mu(f) = 0$, and*
- *every polynomial $p \in \Bbbk[X]$ such that $p(f) = 0$ is divisible by μ.*

We call the polynomial μ described by this proposition the ***minimal polynomial*** of f and often write μ_f for it — or simply μ, if there is no risk for confusion. As we will see later, it encodes useful information about the endomorphism f. In the proof of this proposition we will see that the degree of μ_f is at most $(\dim V)^2$, but below we will find a much better bound.

Proof. Let $n \coloneqq \dim V$. We know that $\hom(V, V)$ is a vector space of dimension n^2, so its $n^2 + 1$ elements

$$\mathrm{id}_V,\ f,\ f^2,\ \dots,\ f^{n^2-1},\ f^{n^2}$$

cannot be linearly independent. There exist then n^2 scalars $a_0, a_1, \dots, a_{n^2} \in \Bbbk$ not all simultaneously zero and such that

$$a_0\mathrm{id}_V + a_1 f + a_2 f^2 + \cdots + a_{n^2} f^{n^2} = 0$$

in $\hom(V, V)$. If we consider now the polynomial

$$p(X) \coloneqq a_0\mathrm{id}_V + a_1 X + a_2 X^2 + \cdots + a_{n^2} X^{n^2} \in \Bbbk[X],$$

which is not the zero polynomial, then we have that $p(f) = 0$. This tells us that that the subset

$$I \coloneqq \{h \in \Bbbk[X] : h(f) = 0\}$$

of $\Bbbk[X]$ contains non-zero elements, and that we may therefore consider the integer

$$d \coloneqq \min\{\deg h : h \in I \smallsetminus 0\},$$

as the set whose minimum we are taking is a non-empty subset of $\mathbb{N}_0$. This number is at most equal to n^2, since we constructed above an element of $I \smallsetminus 0$ whose degree is at most n^2.

Let μ_0 be an element of $I \smallsetminus 0$ whose degree is exactly d and let a be its principal coefficient of μ_0. Of course, $a \neq 0$, so we can consider the polynomial $\mu \coloneqq a^{-1} \cdot \mu$: this is clearly monic of degree d, and has $\mu(f) = a^{-1} \cdot \mu_0(f) = 0$, so it is a monic element of $I \smallsetminus 0$ of degree d.

We have, as we observed, that μ is monic and has $\mu(f) = 0$. Let p be any element of $\Bbbk[X]$ such that $p(f) = 0$. Since μ is not the zero polynomial, we can find by division two polynomials q and r in $\Bbbk[X]$ with $p = q \cdot \mu + r$ such that either $r = 0$ or $\deg r < \deg \mu$.

Suppose for a moment that $r \neq 0$. Since $r = p - q \cdot \mu$ we have that

$$r(f) = p(f) - q(f) \circ \mu(f) = 0,$$

and this tells us that r belongs to $I \smallsetminus 0$, so that

$$\deg \mu > \deg r \geq \min\{\deg h : h \in I \smallsetminus 0\} = d = \deg \mu.$$

This is absurd. We must therefore have that r is zero, and therefore that $p = q \cdot \mu$, that is, that μ divides p. This shows that μ has the two properties listed in the statement of the proposition.

To complete the proof of the proposition we have to check that μ is the unique monic polynomial with those two properties. To do that, let us suppose that τ is another element of $\Bbbk[X]$ that is monic, has $\tau(f) = 0$, and divides every polynomial p of $\Bbbk[X]$ that vanishes on f.

As μ divides every polynomial that vanishes on f and $\tau(f) = 0$, there exists a polynomial u such that $\tau = a \cdot \mu$. Similarly, since τ divides every polynomial that vanishes on f and $\mu(f) = 0$, there exists a polynomial v such that $\mu = v \cdot \tau$. We have that $\mu = uv \cdot \mu$: since μ is not the zero polynomial, this implies that $uv = 1$, and thus that the polynomials u and v are in fact non-zero scalars. Moreover, since $\mu = u \cdot \tau$ and μ and τ are both monic polynomials, we must have that $u = 1$. In particular, we see with this that $\mu = \tau$, and this proves what we want. □

The way in which we found the minimal polynomial in the proof of this proposition does not lend itself to practical calculation. The following proposition gives a much more calculational approach.

{prop:min:2}

Proposition 1.5.2. *Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space V. There is a smallest non-negative integer $d \in \mathbb{N}_0$ such that the linear maps*

$$\mathsf{id}_V,\ f,\ f^2,\ \dots,\ f^d$$

are linearly dependent in the vector space $\hom(V,V)$*, and there is a unique choice of scalars $a_0, a_1, \dots, a_{d-1}$ such that*

$$a_0\mathsf{id}_V + a_1 f + a_2 f^2 + \cdots + a_{d-1}f^{d-1} + f^d = 0.$$

The polynomial

$$a_0 + a_1 X + a_2 X^2 + \cdots + a_{d-1}X^{d-1} + X^d \in \Bbbk[X]$$

is the minimal polynomial of f.

Proof. Let $\mu_f \in \Bbbk[X]$ be the minimal polynomial of f, let d be its degree, and let $a_0, a_1, \dots, a_{d-1}$ be the scalars such that $\mu_f(X) = a_0 + a_1X + \cdots + a_{d-1}X^{d-1} + X^d$.

Let E be the set of all non-negative integers e such that the linear maps $\mathsf{id}_V, f, f^2, \dots, f^e$ are linearly dependent. This set is not empty: since $0 = \mu_f(f) = a_0\mathsf{id}_V + a_1f + \cdots + a_{d-1}f^{d-1} + f^d$, the number d belongs to E. Since E is a non-empty subset of $\mathbb{N}_0$, we may therefore consider its minimum $\epsilon \coloneqq \min E$. We noted above that $d \in I$, so that $\epsilon \leq d$.

Since ϵ belongs to E, the linear maps $\mathsf{id}_V, f, \dots, f^\epsilon$ are linearly independent, and there exist scalars $b_0, b_1, \dots, b_\epsilon \in \Bbbk$ that are not all zero and have $b_0\mathsf{id}_V + b_1f + \cdots + b_\epsilon f^\epsilon = 0$. We must have that $b_\epsilon \neq 0$: if that were not the case, we would have that the scalars $b_0, b_1, \dots, b_{\epsilon-1}$ are not all zero and have $b_0\mathsf{id}_V + b_1f + \cdots + b_{\epsilon-1}f^{\epsilon-1} = 0$, so that $\epsilon - 1 \in I$: this is absurd, since ϵ is the minimal element of I. Let $c_i \coloneqq b_i/b_\epsilon$ for each $i \in \{1, 2, \dots, \epsilon\}$ and let p be the polynomial $c_0 + c_1X + \cdots + c_{\epsilon-1}X^{\epsilon-1} + X^\epsilon$, then we have that

$$p(f) = c_0 + c_1f + \cdots + c_{\epsilon-1}f^{\epsilon-1} + f^\epsilon = b_\epsilon^{-1}(b_0\mathsf{id}_V + b_1f + \cdots + b_\epsilon f^\epsilon) = 0.$$

It follows from this that the minimal polynomial μ_f divides p. In particular, since p is not the zero polynomial, this tells us that $\epsilon = \deg p \geq \deg \mu_f = d$. As also $\epsilon \leq d$, we in fact have that $d = e$, and then, since μ_f divides p and both polynomials are monic, we have that $\mu_f = p$. We claim that the scalars $c_0, c_1, \dots, c_{\epsilon-1}$ are the only ones for which $c_0 + c_1f + \cdots + c_{\epsilon-1}f^{\epsilon-1} + f^\epsilon = 0$. Indeed, if $c'_0, c'_1, \dots, c'_{\epsilon-1}$ is another sequence of elements of $\Bbbk$ such that $c'_0 + c'_1f + \cdots + c'_{\epsilon-1}f^{\epsilon-1} + f^\epsilon = 0$, then we have that

$$(c_0 - c'_0) + (c_1 - c'_1)f + \cdots + (c_{\epsilon-1} - c'_{\epsilon-1})f^{\epsilon-1} = 0,$$

so that the polynomial $q(X) = (c_0 - c'_0) + (c_1 - c'_1)X + \cdots + (c_{\epsilon-1} - c'_{\epsilon-1})X^{\epsilon-1}$ vanishes on f and is therefore divisible by μ_f: as its degree is strictly smaller that the degree of μ_f, we can deduce from this that it is actually the zero polynomial and thus that $c_i = c'_i$ for all $i \in \{1, 2, \dots, \epsilon - 1\}$. The proposition follows from these observations. □

As usual, there are versions of these results for matrices. We collect them in the following proposition.

{prop:min:mats}

Proposition 1.5.3. *Let n be a positive integer and let A be a matrix in $\mathrm{M}_n(\Bbbk)$. There exists a unique monic polynomial $\mu_A \in \Bbbk[X]$ such that*

- *$\mu_A(A) = 0$, and*
- *every polynomial $p \in \Bbbk[X]$ such that $p(A) = 0$ is divisible by μ.*

The degree of μ_A is the smallest non-negative integer $d \in \mathbb{N}_0$ such that the $d + 1$ matrices

$$I_n,\ A,\ A^2,\ \dots,\ A^d$$

are linearly independent, there is exactly one choice of scalars a_0, a_1, …, a_{d-1} in $\Bbbk$ such that

$$a_0 I_n + a_1 A + a_2 A^2 + \cdots + a_{d-1} A^{d-1} + A^d = 0,$$

and those scalars are such that $\mu_A(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_{d-1} X^{d-1} + X^d$. □

Of course, we call the polynomial μ_A the ***minimal polynomial*** of the matrix A.

Exercise 1.5.4. Deduce this proposition from Propositions 1.5.1 and 1.5.2.

In practice, we use Proposition 1.5.2 and its analogue for matrices when trying to compute the minimal polynomial of endomorphisms and matrices. The following is a typical example of how that calculation is carried out.

Example 1.5.5. Let us consider the matrix

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

The list of matrices of length 1

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

is linearly independent, the list of matrices of length 2

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \ A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}$$

is also linearly independent, and the list of length 3

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \ A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}, \ A^2 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 2 & 1 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 3 & 0 & 1 & 1 \end{pmatrix}$$

is also linearly independent. On the other hand, the list of length 4

$$I = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}, A^2 = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 2 & 1 & 1 & 0 \\ 0 & 0 & 2 & 0 \\ 3 & 0 & 1 & 1 \end{pmatrix}, A^3 = \begin{pmatrix} 2 & 0 & 2 & 0 \\ 4 & 1 & 1 & 0 \\ 2 & 0 & -2 & 0 \\ 5 & 0 & 3 & 1 \end{pmatrix}$$

is not linearly independent, since

$$2I - 2A - A^2 + A^3 = 0.$$

According to the proposition, then, the minimal polynomial of A is

$$\mu_A(X) = 2 - 2X - X^2 + X^3.$$

The idea we used in the proof of Proposition 1.5.1 to find the minimal polynomial of a linear map is extremely useful. The following exercise gives its general form.

{ex:mgen}

Exercise 1.5.6. Let I be a subspace of the vector space $\Bbbk[X]$ which is an ***ideal***, that is, such that

whenever $p \in I$ and $q \in \Bbbk[X]$ we have that $pq \in I$.

Prove that if I is not the zero subspace if $\Bbbk[X]$ there is in I a unique monic element m such that $I = \{pm : p \in \Bbbk[X]\}$. We call m the ***monic generator*** of the ideal I.

If $f : V \to V$ is an endomorphism of a finite-dimensional vector space, then the subspace $\{p \in \Bbbk[X] : p(f) = 0\}$ of $\Bbbk[X]$ is a non-zero ideal and its monic generator is precisely the minimal polynomial of f.

Observation 1.5.7. Proposition 1.5.1 states that an endomorphism of a finite-dimensional vector space has a well-determined minimal polynomial with the properties described there. The hypothesis of finite-dimensionality is important for that. In general, if $f : V \to V$ is an endomorphism of an arbitrary vector space, we can consider the subspace

$$I \coloneqq \{p \in \Bbbk[X] : p(f) = 0\}$$

of $\Bbbk[X]$, as in the proof of that proposition, and if I is not the zero subspace we call the unique monic generator of I, in the sense of Exercise 1.5.6, the ***minimal polynomial*** of f. The thing is, without some hypothesis on V or on f it may well be the case that I is actually the zero subspace of $\Bbbk[X]$. In that case, of course, there is certainly no monic element in I, and the only sensible candidate for a minimal polynomial is the zero polynomial.

It is easy to construct examples of this situation. For example, the maps $L, R, S : \mathbb{R}[X] \to \mathbb{R}[X]$

such that

$$L(p) = Xp, \qquad R(p) = p', \qquad S(p) = p(2X)$$

for all $p \in \mathbb{R}[X]$ have vanishing minimal polynomial. On the other hand, and endomorphism of an infinite-dimensional vector space may have a non-zero minimal polynomial: a simple example of this is the map

$$T : p \in \mathbb{R}[X] \mapsto p(-X) \in \mathbb{R}[X],$$

whose minimal polynomial is $X^2 - X$.

§1.6. The Cayley–Hamilton theorem

The minimal polynomial and the characteristic polynomial of an endomorphism are closely related. We will explore that relation in this section. We start with the fact that they have the same roots.

{prop:chimu:roots}

Proposition 1.6.1. *Let $f : V \to V$ be a endomorphism of a finite-dimensional vector space. The characteristic polynomial χ of f and the minimal polynomial μ of f have the same roots in $\Bbbk$.*

Proof. Let λ be a root of the minimal μ of f, so that there is a polynomial $q \in \Bbbk[X]$ such that $\mu(X) = (X - \lambda)q(X)$. As $\mu \neq 0$, we have that $q \neq 0$ and that $\deg q < \deg \mu$. In particular, μ does not divide q and we know that the linear map $q(f) : V \to V$ is not zero, so that there is a vector v in V such that $w := q(f)(v) \neq 0$. Now $\mu(f) = 0$, and therefore

$$0 = \mu(f)(v) = \big((f - \lambda \cdot \mathsf{id}_V) \circ q(f)\big)(v) = (f - \lambda \cdot \mathsf{id}_V)(w),$$

so $f(w) = \lambda w$. As w is not zero, this tells us that λ is an eigenvalue of f and, in particular, that it is a root of the characteristic polynomial χ of f.

Let now λ be a root of the characteristic polynomial χ. It is then an eigenvalue of f, so there exists a non-zero vector v in V such that $f(v) = \lambda v$. Now Lemma 1.3.1 tells us that

$$0 = \mu(f)(v) = \mu(\lambda) \cdot v,$$

and this implies that $\mu(\lambda) = 0$, since $v \neq 0$. This proves the proposition. □

Next we pass on to the Cayley–Hamilton theorem, one of the fundamental results of linear algebra.

Theorem 1.6.2 (Cayley–Hamilton). *Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space, and let χ_f and μ_f be its characteristic and minimal polynomials, respectively. We have that $\chi_f(f) = 0$, and that μ_f divides χ_f.* {thm:CH}

Proof. We want to show that $\chi_f(f) = 0$. Since $\chi_f(f)$ is a linear map $V \to V$, to do this we have to check that $\chi_f(f)(v) = 0$ for all non-zero vectors v of V. Let us do that.

Let us put $n \coloneqq \dim V$, and let v be a non-zero vector of V. Let us consider the set E of all non-negative integers $e \in \mathbb{N}$ such that the vectors d vectors

$$v,\ f(v),\ f^2(v),\ \dots, f^{e-1}(v)$$

are linearly independent. This set is non-empty because it contains 1, since we are supposing that v is non-zero. On the other hand, the set E is finite: if $e > n$ then no e vectors of V are linearly independent. We may therefore consider the number $d \coloneqq \max E$.

We clearly have that $1 \le d \le n$, that the d vectors

$$v,\ f(v),\ f^2(v),\ \dots, f^{d-1}(v) \tag{1.8}$$ {eq:vs}

are linearly independent, and that the $d+1$ vectors

$$v,\ f(v),\ f^2(v),\ \dots, f^{d}(v)$$

are linearly dependent. It follows easily from this that we can find scalars $a_0, a_1, \dots, a_{d-1} \in \Bbbk$ such that

$$a_0 v + a_1 f(v) + a_2 f^2(v) + \cdots + a_{d-1} f^{d-1}(v) + f^d(v) = 0. \tag{1.9}$$ {eq:vs:2}

If we let p be the polynomial $a_0 + a_1 X + a_2 X^2 + \cdots + a_{d-1} X^{d-1} + X^d \in \Bbbk[X]$, then what we have is that $p(f)(v) = 0$.

As the d vectors listed in (1.8) are linearly independent, we can find other vectors $w_1, \dots, w_{n-d}$ in V such that

$$\mathscr{B} = (v, f(v), \dots, f^{d-1}(v), w_1, \dots, w_{n-d})$$

is an ordered basis for V. Since we have that

$$f\big(f^i(v)\big) = f^{i+1}(v) \qquad \text{for each } i \in \{0, \dots, d-1\}$$

and

$$f\big(f^{d-1}(v)\big) = -a_0 v - a_1 f(v) - a_2 f^2(v) - \cdots - a_{d-1} f^{d-1}(v)$$

because of (1.9), the matrix of f with respect to the ordered basis $\mathscr{B}$ is of the form

$$\left(\begin{array}{cccccc|ccc}
0 & 0 & 0 & \cdots & 0 & -a_0 & * & \cdots & * \\
1 & 0 & 0 & \cdots & 0 & -a_1 & * & \cdots & * \\
0 & 1 & 0 & \cdots & 0 & -a_2 & * & \cdots & * \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0 & -a_{n-2} & * & \cdots & * \\
0 & 0 & 0 & \cdots & 1 & -a_{n-1} & * & \cdots & * \\
\hline
0 & 0 & 0 & \cdots & 0 & 0 & * & \cdots & * \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & * & \ddots & * \\
0 & 0 & 0 & \cdots & 0 & 0 & * & \cdots & *
\end{array}\right).$$

In other terms, there exist matrices $A \in \mathrm{M}_{n-d}(\Bbbk)$ and $B \in \mathrm{M}_{d,n-d}(\Bbbk)$ such that we have a block decomposition

$$[f]_{\mathscr{B}} = \begin{pmatrix} C(p) & B \\ 0 & A \end{pmatrix},$$

with $C(p)$ the companion matrix of the polynomial p. It then follows from the results in Exercise 1.2.11 and Example 1.2.10 that the characteristic polynomial χ_f, which is the same as the characteristic polynomial of the matrix $[f]_{\mathscr{B}}$, is

$$\chi_f(X) = \chi_{C(p)}(X) \cdot \chi_A(X) = p(X) \cdot \chi_A(X).$$

In particular, we can now compute that

$$\chi_f(f)(v) = \big(\chi_A(f) \circ p(f)\big)(v) = \chi_A(f)\big(p(f)(v)\big) = 0,$$

since $p(f)(v) = 0$. As we noted at the beginning of the proof, the first claim of the theorem, that $\chi_f(f) = 0$, follows from this. Finally, since the minimal polynomial μ_f divides every polynomial that vanishes on f, the second claim of the theorem follows from the first one. □

An immediate consequence of this theorem is a bound for the degree of the minimal polynomial of a linear map:

Corollary 1.6.3. *If $f : V \to V$ is an endomorphism of a finite-dimensional vector space V and μ_f is its minimal polynomial, then* $\deg \mu_f \leq \dim V$.

Proof. Indeed, in that situation the theorem tells us that μ_f divides the characteristic polynomial χ_f of f, and we know the degree of the latter is $\dim V$. □

This bound is in fact tight: there exist endomorphism $f : V \to V$ of finite-dimensional vector spaces whose minimal and characteristic polynomials coincide, so that the degree of their minimal polynomial is equal to $\dim V$ — we say such endomorphisms are ***non-derogatory***.

Example 1.6.4. Let n be a positive integer, let $p = a_0 + a_1X + \cdots + a_{n-1}X^{n-1} + X^n$ be a monic polynomial in $\Bbbk[X]$ of degree n, and let $f : \Bbbk^n \to \Bbbk^n$ be the linear map whose matrix with respect to the standard basis of $\Bbbk^n$ is $C(p)$, the companion matrix of the polynomial p,

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & -a_{n-3} \\ 0 & 0 & 0 & \cdots & 1 & 0 & -a_{n-2} \\ 0 & 0 & 0 & \cdots & 0 & 1 & -a_{n-1} \end{pmatrix}.$$

Let $\mu = c_0 + c_1X + \cdots + c_{d-1}X^{d-1} + X^d$ be the minimal polynomial of f, and suppose for a moment that its degree d is strictly smaller than n. If $(e_1, \dots, e_n)$ is the standard ordered basis of $\Bbbk^n$, then $f(e_i) = e_{i+1}$ for each $i \in \{1, 2, \dots, n-1\}$, and therefore $f^i(e_1) = e_{1+i}$ whenever $0 \le i < n-2$. It follows from this that

$$0 = \mu(f)(e_1) = c_0e_1 + c_1f(e_1) + \cdots + c_{d-1}f^{d-1}(e_1) + f^d(e_1) = c_0e_1 + c_1e_2 + \cdots + c_{d-1}e_d + e_{d+1}.$$

This is of course impossible, and the contradiction arose from our hypothesis that $d < n$. It follows then that the degree of μ is n, the degree of the characteristic polynomial χ of f, and since μ divides χ and both polynomials are monic, we have that $\mu = \chi$.

The minimal polynomial of an endomorphism $f : V \to V$ of a finite-dimensional vector space V has the property that it divides all polynomials that vanish on f. The characteristic polynomials of f has a closely related property: to prove this, we start by considering the case of matrices. To obtain this, in turn, we will need the following simple observation.

Lemma 1.6.5. *Let n be a positive integer, and let A and B be two square matrices of size n with entries in $\Bbbk[X]$. If $AB = BA$, then for all positive integers i we have that* {lemma:aibi}

$$A^i - B^i = (A - B)\sum_{j=0}^{i-1} A^{i-1-j}B^j$$

Proof. Let us suppose that the two matrices A and B are such that $AB = BA$, and let i be a positive integer. We have that

$$(A-B)\sum_{j=0}^{i-1}A^{i-1-j}B^j=\sum_{j=0}^{i-1}A^{i-j}B^j-\sum_{j=0}^{i-1}BA^{i-1-j}B^j.$$

Since $AB=BA$ we have that $BA^k=A^kB$ for all $k\in\mathbb{N}_0$, and therefore this is

$$\begin{aligned}&=\sum_{j=0}^{i-1}A^{i-j}B^j-\sum_{j=0}^{i-1}A^{i-1-j}B^{j+1}\\&=\sum_{j=0}^{i-1}A^{i-j}B^j-\sum_{j=1}^{i}A^{i-j}B^j\\&=A^i-B^i.\end{aligned}$$

This proves the lemma. □

This lemma and a simple calculation with matrices proves what we want:

Proposition 1.6.6. *Let n be a positive integer, and let A be an element of $\mathrm{M}_n(\Bbbk)$. If g is an element of $\Bbbk[X]$ such that $g(A)=0$, then the characteristic polynomial χ_A of A divides g^n.*

Proof. Let g be an element of $\Bbbk[X]$ such that $g(f)=0$. If g is the zero polynomial then it is obvious that χ_A divides g^n, so we may suppose that that is not the case. There are then a non-negative integer d and scalars $g_0, g_1, \dots, g_d$ in $\Bbbk$ such that $g=\sum_{i=0}^d g_iX^i$ and $g_d\neq 0$.

Let us consider the matrix $X\cdot I_n$, which is a square matrix of size n whose entries are elements of $\Bbbk[X]$. We can evaluate the polynomial g at $X\cdot I_n$: as this matrix is diagonal, we have that $g(X\cdot I_n)=g(X)\cdot I_n$, and therefore it is clear that

$$\det g(X\cdot I_n)=g(X)^n. \qquad \text{{eq:detgxin}}$$

On the other hand, we have that, since $g(A)=0$,

$$g(X\cdot I_n)=g(X\cdot I_n)-g(A)=\sum_{i=0}^d g_i\cdot((X\cdot I_n)^i-A^i)=\sum_{i=1}^d g_i\cdot((X\cdot I_n)^i-A^i),$$

because the term corresponding to $i=0$ in the first sum vanishes. Using Lemma 1.6.5 and the fact that the matrices $X\cdot I_n$ and A commute, we see from this that

$$g(X\cdot I_n)=\sum_{i=1}^d g_i\cdot(X\cdot I_n-A)\cdot\sum_{j=0}^{i-1}X^{i-1-j}A^j=(X\cdot I_n-A)\cdot C,$$

with C the matrix $\sum_{i=1}^d g_i\cdot\sum_{j=0}^{i-1}X^{i-1-j}A^j$. In particular, we have that

$$g(X)^n=\det(X\cdot I_n)=\det(X\cdot I_n-A)\cdot\det C,$$

and we can conclude that the characteristic polynomial χ_A divides g^n: this is what the proposition claims. □

We can now deduce as a corollary the correspoding fact about linear maps.

{coro:mugn}

Corollary 1.6.7. *Let V be a non-zero finite-dimensional vector space, let n be the dimension of V, and let $f : V \to V$ be an endomorphism of V. If g is any polynomial in $\Bbbk[X]$ such that $g(f) = 0$, then the characteristic polynomial χ_f of f divides g^n.*

Proof. Let g be an element of $\Bbbk[X]$ such that $g(f) = 0$. Let $\mathscr{B}$ be an ordered basis for the vector space V, let n be the dimension of V, and let $A \coloneqq [f]_{\mathscr{B}} \in \mathrm{M}_n(\Bbbk)$ be the matrix of f with respect to $\mathscr{B}$. As $g(f) = 0$, we have that $g(A) = 0$, and the proposition we have just proved tells us that the characteristic polynomial χ_A of A, which coincides with the characteristic polynomial χ_f of f, divides g^n. This proves the corollary. □

The most important application of this result is the following generalization of Proposition 1.6.1:

Proposition 1.6.8. *Let $f : V \to V$ be an endomorphism of a non-zero finite dimensional vector space V. The characteristic polynomial χ_f of f and the minimal polynomial μ_f of f have the same irreducible factors in $\Bbbk[X]$.*

Proof. Let n be the dimension of V. The Cayley–Hamilton Theorem 1.6.2 tells us that μ_f divides χ_f, and Corollary 1.6.7 that χ_f divides μ_f^n.

- If p is an irreducible factor of μ_f, then $p \mid \mu_f \mid \chi_f$, and thus p divides χ_f and is, therefore, one of its irreducible factors.
- If p is an irreducible factor of χ_f, then we have that $p \mid \chi_f \mid \mu_f^n$ and, since p is irreducible, that in fact p divides μ_f.

These two observations prove the proposition. □

Corollary 1.6.9. *Let $f : V \to V$ be an endomorphism of a non-zero finite dimensional vector space V. If the minimal polynomial μ_f of f is irreducible, then* $\deg \mu_f$ *divides* $\dim V$ *and there is a positive integer k such that the characteristic polynomial of f is $\chi_f = \mu_f^k$.*

Proof. Let us suppose that the minimal polynomial μ_f of f is irreducible. Since μ_f and the characteristic polynomial χ_f have the same irreducible factors, this tells us that μ_f is the only irreducible factor of χ_f, and thus, since both polynomials are monic, that there is a positive integer k such that $\chi_f = \mu_f^k$. In particular, this implies that $\dim V = \deg \chi_f = k \cdot \deg \mu_f$, so that the degree of μ_f divides $\dim V$. □

Example 1.6.10. Let V be a non-zero finite dimensional real vector space. A ***complex structure*** on V is a linear map $f : V \to V$ such that $f^2 = -\mathrm{id}_V$. Clearly, such a linear map is a zero of the polynomial $p \coloneqq X^2 + 1 \in \mathbb{R}[X]$ and thus its minimal polynomial μ_f divides p. As V is a non-zero space, the polynomial μ_f is not constant: as it is a non-constant factor of p and p is irreducible in $\mathbb{R}[X]$, we see that in fact $\mu_f = X^2 + 1$. In particular, the minimal polynomial μ_f is irreducible: since the degree of μ_f is 2, the corollary allows us to conclude that $\dim V$ is an even positive integer. In this way we obtain the following important observation:

> *the dimension of a finite-dimensional real vector space that admits a complex structure is even.*

§1.7. Invariant subspaces and restrictions

Let $f : V \to V$ be an endomorphism of a vector space V. A subspace W of V is ***f-invariant*** if for all vectors w of W we have that also $f(w) \in W$. The following lemma exhibits simple examples of such subspaces.

{lemma:finv:1}

Lemma 1.7.1. *Let $f : V \to V$ be an endomorphism of a vector space V.*

(*i*) *The subspaces 0 and V of V are f-invariant of V.*

(*ii*) *The kernel $\ker f$ and the image $\operatorname{img} f$ of f are f-invariant subspaces of V.*

(*iii*) *If $\lambda \in \Bbbk$ is a scalar, then the subspace $E_\lambda(f) = \{v \in V : f(v) = \lambda v\}$ is an f-invariant subspace of V.*

(*iv*) *If $p \in \Bbbk[X]$ is a polynomial, then the subspace $\ker p(f)$ is f-invariant.*

We call 0 and V the ***trivial***f-invariant subspaces, and all other invariant subspaces ***non-trivial***. Notice that if in (*iv*) we take $p(X) = X - \lambda$, then the subspace $\ker p(f)$ is precisely the same as $E_\lambda(f)$, so this fourth part of the lemma generalizes the third one.

Proof. That the subspaces 0 and V are f-invariant is obvious. If v is in $\ker f$, then $f(v) = 0 \in \ker f$, because $\ker f$ is a subspace of V, and this shows that $\ker f$ is an f-invariant subspace. On the other hand, if v is in $\operatorname{img} f$, then $f(v)$ is also in $\operatorname{img} f$, simply because it is the image of v, and this tells us that $\operatorname{img} f$ is also an f-invariant subspace. This proves the first two claims of the lemma.

Let $\lambda \in \Bbbk$ be a scalar. If v is an element of $E_\lambda(f)$, so that $f(v) = \lambda v$, then $f(v) \in E_\lambda(f)$ because $E_\lambda(f)$ is a subspace of V and contains v: this shows that $E_\lambda(f)$ is an f-invariant subspace of V, and thus that the third claim of the lemma holds.

Finally, let $p = a_0 + a_1 X + \cdots + a_d X^d \in \Bbbk[X]$ be a polynomial, and suppose that $v \in V$ is an element of $\ker p(f)$, so that $0 = p(f)(v) = a_0 v + a_1 f(v) + \cdots + a_d f^d(v)$. We then have that

$$p(f)(f(v)) = a_0 f(v) + a_1 f^2(v) + \cdots + a_d f^{d+1}(v) = f\big(a_0 v + a_1 f(v) + \cdots + a_d f^d(v)\big) = f(0) = 0,$$

so that $f(v) \in \ker p(f)$. This shows that $\ker p(f)$ is an p-invariant subspace. □

Example 1.7.2. Let us consider the $\mathbb{Q}$-linear map

$$f : (x, y) \in \mathbb{Q}^2 \mapsto (2x, 3y) \in \mathbb{Q}^2.$$

We know that 0 and $\mathbb{Q}^2$ are f-invariant subspaces. Also, since 2 and 3 are the eigenvalues of f, the eigenspaces $E_2(f) = \langle e_1 \rangle$ and $E_2(f) = \langle e_2 \rangle$ are also f-invariant subspaces. We claim that these four are the only f-invariant subspaces of V.

Indeed, suppose that W is an f-invariant subspace of $\mathbb{Q}^2$ that is different from 0 and from $\mathbb{Q}^2$. The dimension of W is therefore equal to 1, and if we let $v = (a, b)$ be any non-zero vector in W we have that $W = \langle (a, b) \rangle$. As W is f-invariant and (a, b) belongs to W, we must have that $(2a, 3b) = f(a, b)$ also belongs to W, and this implies that there exists a scalar $\lambda \in \mathbb{Q}$ such that $(2a, 3b) = \lambda(a, b)$. Since (a, b) is not the zero vector in $\mathbb{Q}^2$, this tells us that λ is an eigenvalue of f, so it is equal to 2 or to 3, and that (a, b) is an eigenvector of f corresponding to that eigenvalue, so that it is a non-zero multiple of e_1 or a non-zero multiple of e_2. It follows from this that W is equal to $E_2(f)$ or to $E_3(f)$. This proves what we want.

Lemma 1.7.1 tells us that the zero subspace is always invariant. The next one characterizes the one-dimensional invariant subspaces, and shows that, in some sense, the notion of invariant subspaces generalizes that of eigenvectors.

Lemma 1.7.3. *Let $f : V \to V$ be an endomorphism of a vector space V. A one-dimensional subspace W of V is f-invariant if and only if there is an eigenvector v of f such that $W = \langle v \rangle$.*

Proof. Suppose first that W is a one-dimensional f-invariant subspace of V, and let v be any non-zero element of W. We have that $W = \langle v \rangle$, because $\dim W = 1$, and, since W is invariant, that $f(v) \in W = \langle v \rangle$, so that there is a scalar $\lambda \in \Bbbk$ such that $f(v) = \lambda v$: we thus see that W is the span of an eigenvector of f. The condition given by the lemma is therefore necessary.

Suppose now, to check that that condition is also sufficient, that there is an eigenvector v of f such that $W = \langle v \rangle$, and let $\lambda \in \Bbbk$ be the eigenvalue corresponding to v, so that $f(v) = \lambda v$. If w is an arbitrary element of W, then there is a scalar $a \in \Bbbk$ such that $w = av$, and then $f(w) = f(av) = af(v) = a\lambda v \in W$. This shows that W is f-invariant. □

Using this lemma as a starting point we can easily describe the linear maps that have the maximum possible number of invariant subspaces.

Example 1.7.4. Let us suppose that $f : V \to V$ is an endomorphism of a vector space such that *every* subspace of V is f-invariant. If v is a non-zero vector, then the one-dimensional subspace $\langle w \rangle$ is thus invariant and the lemma implies immediately that w is an eigenvector for f. We see with this that *all* non-zero elements of V are eigenvectors of f.

For each non-zero vector v in V there is then a scalar λ_v such that $f(v) = \lambda_v v$. We claim that, in fact, the scalar λ_v is independent of v. Indeed, let v and w be two non-zero elements of V. To check that $\lambda_v = \lambda_w$ we consider two cases:

- If v and w are linearly dependent, then there is a non-zero scalar a in $\Bbbk$ such that $v = aw$, and then we have that $\lambda_v v = f(v) = f(aw) = af(w) = a\lambda_w w = \lambda_w v$: as $v \neq 0$, this implies that $\lambda_v = \lambda_w$ in this case.
- If instead v and w are linearly independent, then in particular $v + w \neq 0$, so that $f(v + w) = \lambda_{v+w}(v + w)$, and therefore

$$\lambda_{v+w}v + \lambda_{v+w}w = \lambda_{v+w}(v + w) = f(v + w) = f(v) + f(w) = \lambda_v v + \lambda_w w.$$

  The linear independence of v and w then tells us that $\lambda_v = \lambda_{v+w} = \lambda_w$.

The conclusion of this is that there exists a scalar λ such that $f(v) = \lambda v$ for all non-zero elements v in V. We of course also have that $f(0) = \lambda 0$, so in fact we see that $f = \lambda \cdot \mathsf{id}_V$. We have showed part of the following statement:

all subspaces of V are f-invariant if and only if f is a scalar multiple of id_V.

What remains to verify in order to establish this statement is that if $\lambda \in \Bbbk$ is a scalar and $f : V \to V$ is the map $\lambda \cdot \mathsf{id}_V$, then all subspaces of V are f-invariant. This is immediate.

Example 1.7.5. For each $m \in \mathbb{N}_0$ we write $\mathbb{R}[X]_{\leq m}$ for the real vector space of all polynomials of degree at most m. We fix $n \in \mathbb{N}_0$, put $V := \mathbb{R}[X]_{\leq n}$, and the linear map

$$f : p \in V \mapsto p' \in V.$$

Suppose that W is a non-zero f-invariant subspace of V. Since W is not the zero subspace, it contains non-zero polynomials; on the other hand, all non-zero elements of W have degree at most n. This implies that we we can consider the number $d := \max\{\deg p : p \in W \setminus 0\}$, as the set whose maximum we are taking is a bounded and non-empty subset of $\mathbb{N}_0$. Since d is the maximum degree of the non-zero elements of W, it is clear that $W \subseteq \mathbb{R}[X]_{\leq d}$, and we want to show that in fact we have that $W = \mathbb{R}[X]_{\leq d}$. To do this it is enough to check that W has dimension at least $d + 1$, since we know that $\dim \mathbb{R}[X]_{\leq d} = d + 1$.

There is some polynomial q in W of degree exactly d. Since W is f-invariant, we have that the $d+1$ polynomials q, $f(q) = q'$, $f^2(q) = q''$, …, $f^d(q) = q^{(d)}$ all belong to W. They are all non-zero and their degrees are exactly d, $d-1$, $d-2$, …, 0. It follows from this that those $d+1$ polynomials are linearly independent. Indeed, suppose that they are not, so that there are are scalars a_0, a_1, …, a_d in $\Bbbk$, not all zero, such that

$$a_0 q + a_1 q' + a_2 q'' + \cdots + a_d q^{(d)} = 0.$$

As not all of those scalars are zero, it makes sense to consider the smallest index $i \in \{0, \dots, d\}$ such that $a_i \neq 0$: we then have that

$$-a_i q^{(i)} = \underbrace{a_{i+1} q^{(i+1)} + \cdots + a_d q^{(d)}}_{\text{degree at most } d-i-1}.$$

This is absurd, since $a_u q^{(i)}$ is a polynomial of degree exactly $d-i$, strictly greater than $d-i-1$.

It follows from this that W contains the $d+1$ polynomial q, q', …, $q^{(d)}$ that are linearly independent and therefore that $\dim W \geq d+1$, as we wanted to show. We have proved the following statement

every f-invariant subspace of V is of the form $\mathbb{R}[X]_{\leq d}$ for some $d \in \{0, \dots, n\}$.

On the other hand, it is easy to check that each of the $d+1$ subspaces $\mathbb{R}[X]_{\leq 0}$, $\mathbb{R}[X]_{\leq 1}$, …, $\mathbb{R}[X]_{\leq d}$ is f-invariant, so putting everything together we can conclude that those $d+1$ subspaces of $\mathbb{R}[X]_{\leq n}$ are all the f-invariant subspaces.

Example 1.7.6. Let us next consider the endomorphism $f : (x, y) \in \mathbb{Q}^2 \mapsto (-y, x) \in \mathbb{Q}^2$. If W is a non-zero f-invariant subspace of $\mathbb{Q}^2$, then we can pick a non-zero vector $v = (a, b)$ in W, and the f-invariance of W implies that also $w = (-b, a) = f(a, b)$ belongs to W. As $\det\left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right) = a^2 + b^2 \neq 0$, the two vectors v and w are linearly independent and therefore the subspace W has dimension at least 2. Of course, this implies that in fact $W = \mathbb{Q}^2$. We see with this that there are exactly two f-invariant subspaces in $\mathbb{Q}^2$, the zero subspace and $\mathbb{Q}^2$ itself.

This example provides us with an endomorphism $f : V \to V$ of a 2-dimensional vector space that only admits *trivial* f-invariant subspaces. We will see later in Example 1.7.14 that we can construct examples with this property of arbitrary finite dimension, but to do this we will need a better way to verify that there are no non-trivial invariant subspaces.

When we have a subspace given as the span of some set of vectors, the following result provides a convenient criterion to check its invariance.

Lemma 1.7.7. *Let $f : V \to V$ be an endomorphism of a vector space V, and let $v_1, v_2, \dots, v_n \in V$ be elements of V. The span $W \coloneqq \langle v_1, v_2, \dots, v_n \rangle$ is f-invariant if and only if for each $i \in \{1, 2, \dots, n\}$ we have that $f(v_i) \in W$.* {lemma:finv:gens}

Proof. If W is f-invariant, then we have that $f(v_1), f(v_2), \dots, f(v_n) \in W$, as $v_1, v_2, \dots, v_n \in W$.

To prove the converse, let us suppose that $f(v_1), f(v_2) \dots, f(v_n) \in W$, and let w be an arbitrary element of W. As W is the span of $\{v_1, v_2, \dots, v_n\}$, there are scalars $a_1, a_2, \dots, a_n \in \Bbbk$ such that $w = a_1v_1 + a_2v_2 + \cdots + a_nv_n$, and then

$$f(w) = f(a_1v_1 + a_2v_2 + \cdots + a_nv_n) = a_1f(v_1) + a_2f(v_2) + \cdots + a_nf(v_n) \in W,$$

since the vectors $f(v_1)$, $f(v_2)$, …, $f(v_n)$ are all in W and W is a subspace, so that W is an f-invariant subspace. This proves the lemma. □

The result of the following exercise tells us that invariant subspaces interact nicely with the standard operations of subspaces.

Exercise 1.7.8. Let $f : V \to V$ be an endomorphism of a vector space V. {ex:finv:oplus}

(1) Show that if W_1 and W_2 are f-invariant subspaces of V, then so are $W_1 + W_2$ and $W_1 \cap W_2$.

(2) Show that if W is an f-invariant subspace of V, then for all polynomials $p \in \Bbbk[X]$ the subspace W is also $p(f)$-invariant.

Let $f : V \to V$ be an endomorphism of a vector space W and let W be an f-invariant subspace of V. We have that $f(w) \in W$ for each vector w in W, and this implies that we can consider the function

$$f_W : w \in W \mapsto f(w) \in W.$$

It is immediate to check that this is in fact a linear function. We call it the ***restriction*** of f to the f-invariant subspace W. Even though the original map f and the restriction f_W «do the same» to elements of W we will insist in distinguishing them, because their domains and codomains are different. In this situation we of course have that $f_W(w) = f(w)$ for all $w \in W$. More generally, we have the following simple result:

Lemma 1.7.9. *Let $f : V \to V$ be an endomorphism of a vector space V and let W be an f-invariant subspace of V. If $p \in \Bbbk[X]$ and $w \in W$, then* {lemma:res:pol}

$$p(f_W)(w) = p(f)(w).$$

Proof. We claim that

$$f_W^i(u) = f^i(u) \text{ *for all* } u \in W \text{ *and all* } i \in \mathbb{N}_0.$$

When $i = 0$, this is clear: for all $u \in W$ we have that $f_W^0(u) = \mathsf{id}_W(u) = u = \mathsf{id}_V(u) = f^0(u)$. On the other hand, if $i \in \mathbb{N}_0$ is such that $f_W^i(u) = f^i(u)$ for all $u \in W$, then also for all $u \in W$ we have that

$$\begin{aligned} f_W^{i+1}(u) &= f_W(f_W^i(u)) = f_W(f^i(u)) && \text{because of the hypothesis} \\ &= f(f^i(u)) = f^{i+1}(u) && \text{because } f^i(u) \in W. \end{aligned}$$

Our claim therefore follows by induction.

Now suppose that $p = a_0 + a_1X + a_2X^2 + \cdots + a_dX^d$ is an element of $\Bbbk[X]$. If w is a vector in W, we then have that

$$\begin{aligned} p(f_W)(w) &= (a_0\mathsf{id}_V + a_1f_W + a_2f_W^2 + \cdots + a_df_W^d)(w) \\ &= a_0w + a_1f_W(w) + a_2f_W^2(w) + \cdots + a_df_W^d(w) \\ &= a_0w + a_1f(w) + a_2f^2(w) + \cdots + a_df^d(w) \\ &= (a_0\mathsf{id}_V + a_1f + a_2f^2 + \cdots + a_df^d)(w) \\ &= p(f)(w). \end{aligned}$$

This is what the lemma asserts. □

The result of the following exercise describes what happens when we restrict a restriction.

{exer:restrict-iterate

Exercise 1.7.10. Let $f : V \to V$ be an endomorphism of a vector space V. If W is an f-invariant subspace of V, and U is an f_W-invariant subspace of W, then U is an f-invariant subspace of V and we have that $(f_W)_U = f_U$.

There is a useful and instructive way of expressing the invariance of subspaces in terms of basis and matrices that we next explain. Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space and let us suppose that W is an f-invariant subspace of V. We put $n \coloneqq \dim V$ and $m \coloneqq \dim W$. We let $\mathscr{B}_W = (v_1, v_2, \ldots, v_m)$ be an ordered basis of W, write $f_W : W \to W$ for the restriction of f to W, and $(a_{i,j})$ for the matrix $[f_W]_{\mathscr{B}_W} \in \mathrm{M}_m(\Bbbk)$ of that restriction f_W with respect to the ordered basis $\mathscr{B}_W$. As usual, this means that

$$f(v_i) = a_{1,i}v_1 + \cdots + a_{m,i}v_m$$

for each $i \in \{1, 2, \ldots, m\}$. As we know, we can complete $\mathscr{B}_W$ to a basis of the whole space V: there exist vectors $v_{m+1}, \ldots, v_n$ in V such that $\mathscr{B} = (v_1, v_2, \ldots, v_m, v_{m+1}, \ldots, v_n)$ is an ordered basis of V. If i is an element of $\{1, 2, \ldots, m\}$, then the ith vector of the ordered basis $\mathscr{B}$ is v_i, and this is an element of W: since W is f-invariant, we know that the image $f(v_i)$ is also an element of W, and it can therefore we written as a linear combination of the elements of the basis $\mathscr{B}_W$ of W with which we started: in fact, we have that

$$f(v_i) = a_{1,i}v_1 + \cdots + a_{m,i}v_i + 0v_{m+1} + \cdots + 0v_n,$$

and this tells us that the ith column of the matrix $[f]_{\mathscr{B}}$ of f with respect to the basis $\mathscr{B}$ has its last $n-m$ entries all equal to zero. This is true for all the first m columns of that matrix, so that matrix has in fact the following block upper triangular decomposition:

$$[f]_{\mathscr{B}} = \left(\begin{array}{ccc:ccc} a_{1,1} & \cdots & a_{1,m} & * & \cdots & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & \cdots & a_{m,m} & * & \cdots & * \\ \hdashline 0 & \cdots & 0 & * & \cdots & * \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & * & \cdots & * \end{array}\right)$$

in which the block at the top leftmost position is precisely the matrix $[f_W]_{\mathscr{B}_W}$.

This argument shows that whenever a linear map admits a non-trivial invariant subspace we can find bases with respect to which the matrix of the linear map is block upper triangular. Moreover, this observation allows us to prove the first claim of the following proposition that establishes the relation between the characteristic polynomial of a linear map and that of a its restriction to an invariant subspace.

{prop:res:chi}

Proposition 1.7.11. *Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space V. If W is an f-invariant subspace of V and $f_W : W \to W$ is the restriction of f to W, then $\chi_{f_W} \mid \chi_f$.*

Proof. Let W be an f-invariant subspace of V and let $f_W : W \to W$ be the restriction of the map f to W. Set $n := \dim V$, $m := \dim W$, and let $\mathscr{B}_W = (v_1, v_2, \dots, v_m)$ be an ordered basis of W and let $v_{m+1}, \dots, v_n$ be vectors of V such that $\mathscr{B} = (v_1, v_2, \dots, v_m, v_{m+1}, \dots, v_n)$ is an ordered basis for V. As we observed above, the matrix of f with respect to $\mathscr{B}$ is block upper triangular, of the form

$$[f]_{\mathscr{B}} = \begin{pmatrix} [f_W]_{\mathscr{B}_W} & C \\ 0 & B \end{pmatrix},$$

for some matrices $B \in \mathrm{M}_{n-m}(\Bbbk)$ and $C \in \mathrm{M}_{m,n}(\Bbbk)$. If follows from this, of course, that the matrix $X \cdot I_n - [f]_{\mathscr{B}}$ is also block upper triangular,

$$X \cdot I_n - [f]_{\mathscr{B}} = \begin{pmatrix} X \cdot I_m - [f_W]_{\mathscr{B}_W} & -C \\ 0 & X \cdot I_{n-m} - B \end{pmatrix},$$

and therefore that

$$\begin{aligned} \chi_f(X) &= \det\bigl(X \cdot I_n - [f]_{\mathscr{B}}\bigr) \\ &= \det\bigl(X \cdot I_m - [f_W]_{\mathscr{B}_W}\bigr) \cdot \det\bigl(X \cdot I_{n-m} - B\bigr) \\ &= \chi_{f_W}(X) \cdot \det\bigl(X \cdot I_{n-m} - B\bigr). \end{aligned}$$

This tells us that the characteristic polynomial χ_{f_W} of the restriction f_W divides the characteristic polynomial χ_f of f, and proves the proposition. □

The following is a simple and useful application of the proposition we have just proved.

Corollary 1.7.12. *Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space V. If λ is an element of $\Bbbk$ and n is the dimension of the subspace $E_\lambda(f)$, then the polynomial $(X - \lambda)^m$ divides the characteristic polynomial χ_f of f.*

Proof. Let λ be an element of $\Bbbk$, let n be the dimension of the subspace $E_\lambda(f)$, and let $\mathscr{B}$ be an ordered basis for $E_\lambda(f)$. The matrix $[f_{E_\lambda(f)}]_{\mathscr{B}}$ of the restriction of f to $E_\lambda(f)$ with respect to $\mathscr{B}$ is $\lambda \cdot I_n$, so the characteristic polynomial of that restriction is

$$\chi_{f_{E_\lambda(f)}} = (X - \lambda)^n.$$

As the subspace $E_\lambda(f)$ is f-invariant, the proposition tells us then that $(X - \lambda)^n$ divides χ_f, and this is what the corollary claims. □

The same relation holds for minimal polynomials, in fact, as we show next. The argument to do this is rather different, though.

{prop:res:mu}

Proposition 1.7.13. *Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space V. If W is an f-invariant subspace of V and $f_W : W \to W$ is the restriction of f to W, then $\mu_{f_W} \mid \mu_f$.*

Proof. Let W be an f-invariant subspace of V and let $f_W : W \to W$ be the restriction of f to W. According to Lemma 1.7.9, for all $w \in W$ we have that $\mu_f(f_W)(w) = \mu_f(f)(w) = 0$, so that in fact $\mu_f(f_W) = 0$. Because of the characteristic property of the minimal polynomial μ_{f_W} of the restriction f_W, then, we have that μ_{f_W} divides μ_f, as the proposition claims. □

We can use these results about divisibility to produce examples of non-derogatory linear maps.

{ex:res:none}

Example 1.7.14. Let n be a positive integer. The polynomial $p(X) = X^n + 2 \in \mathbb{Q}[X]$ is monic and irreducible. Let $C \coloneqq C(p) \in \mathrm{M}_n(\mathbb{Q})$ be the companion matrix of p, and let us consider the linear map $f : x \in \mathbb{Q}^n \mapsto Cx \in \mathbb{Q}^n$. The matrix of f with respect to the standard ordered basis of $\mathbb{Q}^n$ is precisely C, so the characteristic polynomial χ_f of f is the characteristic polynomial of $C(p)$, which we computed in Example 1.2.10 to be p.

Now suppose that W is a non-zero f-invariant subspace of $\mathbb{Q}^n$. According to Proposition 1.7.11, the characteristic polynomial χ_{f_W} of the restriction f_W divides χ_f. Since p is irreducible and $\deg \chi_{f_W} = \dim W > 0$, this implies that in fact $\chi_{f_W} = p$, so that $\dim W = \deg \chi_{f_W} = \deg p = n$. We thus see that $W = \mathbb{Q}^n$. This shows that there are no non-trivial f-invariant subspaces in $\mathbb{Q}^n$.

§1.8. Criteria for triangularizability and diagonalizability

In this section we will establish criteria for the triangularizability and diagonalizability of linear maps and matrices in terms of their minimal polynomials.

We say that a non-zero polynomial p in $\mathbb{k}[x]$ ***splits completely over*** $\mathbb{k}$ if it has a factorization of the form

$$a(X-\lambda_1)^{r_1}(X-\lambda_2)^{r_2}\cdots(X-\lambda_k)^{r_k},$$

with $a \in \mathbb{k}$, $k \in \mathbb{N}_0$, $r_1, r_2, \dots, r_k \in \mathbb{N}$, and $\lambda_1, \lambda_2, \dots, \lambda_k \in \mathbb{k}$ pairwise different. In that case, of course, the scalars $\lambda_1, \dots, \lambda_k$ are precisely the roots of p, listed without repetitions, and the numbers $r_1, \dots, r_k$ are their respective multiplicities. If the multiplicity of one of those roots is 1, we say that root is ***simple***, and if all the roots of p are simple we say that p is ***without multiplicities*** or ***separable***. We will use the following lemma a few times.

Lemma 1.8.1. *Let p and q be two polynomials in $\mathbb{k}[X]$ and suppose that p divides q.*

(i) If q splits completely over $\mathbb{k}$, then so does p.

(ii) If additionally q is without multiplicities, then so is p. □

Exercise 1.8.2. Prove the lemma.

Let us recall that, more generally, *any* non-zero element p of $\mathbb{k}[x]$ has a factorization of the form

$$aq_1^{r_1}q_2^{r_2}\cdots q_k^{r_k} \tag{1.10}$$ {eq:fact}

with $a \in \mathbb{k}$, $r_1, r_2, \dots, r_k \in \mathbb{N}$, and $q_1, q_2, \dots, q_k$ pairwise different monic polynomials that are irreducible in $\mathbb{k}[x]$, and that moreover that factorization is uniquely determined by p, except for the ordering of the factors. As in the previous case, we say that p is ***without multiplicities*** is ***without multiplicities*** or ***separable*** if each of the numbers $r_1, r_2, \dots, r_k$ is equal to 1. The polynomial p splits completely over $\mathbb{k}$ exactly when each of the irreducible factors $q_1, q_2, \dots, q_k$ that appear in its factorization (1.10) has degree 1.

Observation 1.8.3. The polynomial $X^2 + 1$ can be viewed both as an element of $\mathbb{C}[X]$ and as an element of $\mathbb{R}[X]$: it splits completely over $\mathbb{C}$, as it can be factored as $(X - i)(X + i)$ in $\mathbb{C}[X]$, but it does not split completely over $\mathbb{R}$. We thus see that the field plays an important role in deciding whether a polynomial splits completely or not.

On the other hand, it can be shown that a polynomial p in $\Bbbk[X]$ is separable if and only if it is coprime with its derivative or, equivalently, if $\gcd(p, p') = 1$. It follows from this that if we have two fields $\Bbbk$ and $\mathbb{K}$, and one is a subfield of the other, so that $\Bbbk \subseteq \mathbb{K}$, then a polynomial with coefficients in $\Bbbk$ is separable in $\Bbbk[X]$ if and only if it is separable in $\mathbb{K}[X]$.

The condition that $\Bbbk$ is a subfield of $\mathbb{K}$ here is important. For example, the polynomial $X^2 + 1$ can be viewed as an element of $\mathbb{R}[X]$, and there is is separable as it is in fact irreducible, and as an element of $\mathbb{F}_2[X]$, where $\mathbb{F}_2$ denotes a field with two elements, and in this last case it is not separable, as $X^2 + 1$ splits as $(X + 1)^2$ when working over $\mathbb{F}_2$.

The key step to do what we want is the following technical lemma:

{lemma:tech}

Lemma 1.8.4. *Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space V whose minimal polynomial μ splits completely over $\Bbbk$. If W is a* proper *f-invariant subspace of V, then there exists a vector v in $V \smallsetminus W$ and an eigenvalue λ of f such that $(f - \lambda \cdot \mathrm{id}_V)(v) \in W$.*

Proof. As μ splits completely over $\Bbbk$, it has a factorization of the form

$$(X - \lambda_1)^{r_1}(X - \lambda_2)^{r_2} \cdots (X - \lambda_k)^{r_k},$$

with $k \in \mathbb{N}$ and $\lambda_1, \lambda_2, \dots, \lambda_k \in \Bbbk$ pairwise different. Let W be a proper f-invariant subspace of V, let u be a vector in $V \smallsetminus W$, and let us consider the set

$$I \coloneqq \{p \in \Bbbk[X] : p(f)(u) \in W\}.$$

Since $\mu(f) = 0$, we have that $\mu(f)(u) = 0 \in W$, so that $\mu \in I$: as $\mu \neq 0$, this tells us that the set $I \smallsetminus 0$ is not empty, so that neither is the set $\{\deg p : p \in I \smallsetminus 0\}$. This is then a non-empty subset of $\mathbb{N}_0$, and we may therefore consider its minimum element, which we will write d, and an element q of $I \smallsetminus 0$ such that $\deg q = d$.

The polynomial q is not constant: if it were, it would be equal to a non-zero scalar α, and then we would have that $W \ni q(f)(u) = (\alpha \cdot \mathrm{id}_V)(u) = \alpha u$, which is absurd, as $u \notin W$ and $\alpha \neq 0$.

Also, since q is not zero, we can divide the minimal polynomial μ by q: there exist polynomials s and r in $\Bbbk[X]$ such that $\mu = s \cdot q + r$ and either $r = 0$ or $r \neq 0$ and $\deg r < \deg q$. Let us suppose for a moment that the second alternative holds: as $r = \mu - s \cdot q$, we have that $r(f) = \mu(f) - s(f) \circ q(f)$ and

$$r(f)(u) = \mu(f)(u) - s(f)\big(q(f)(u)\big) = 0 \in W.$$

This is impossible, since $r \in I \setminus 0$ and $\deg r$ is strictly smaller than the degree of μ, which is d, the minimal degree of an element of $I \setminus 0$. This contradiction implies that we must have that $r = 0$, so that

$$s \cdot q = \mu = (X - \lambda_1)^{r_1}(X - \lambda_2)^{r_2} \cdots (X - \lambda_k)^{r_k}.$$

In particular, this tells us that the polynomial q, which is not constant, divides the product $(X - \lambda_1)^{r_1}(X - \lambda_2)^{r_2} \cdots (X - \lambda_k)^{r_k}$, and we can conclude that there is an index $i \in \{1, 2, \dots, k\}$ such that $q(\lambda_i) = 0$. Of course, this implies that there is a polynomial $b \in \Bbbk[X]$ such that $q(X) = (X - \lambda_i)b(X)$. Clearly b is not the zero polynomial, because q is not the zero polynomial, and the degree of b is $\deg q - 1 = d - 1$. In view of the way we chose the number d we therefore have that $b \notin I$, that is, that $b(f)(u) \notin W$.

If we now put $v \coloneqq b(f)(u)$, then we have that $v \notin W$ and, as $q(f) = (f - \lambda_i \cdot \mathrm{id}_V) \circ b(f)$, that

$$(f - \lambda_i)(v) = (f - \lambda_i \cdot \mathrm{id}_V)\big(b(f)(u)\big) = q(f)(u) \in W.$$

This proves the lemma, since the scalar λ_i, being a root of the minimal polynomial of f, is an eigenvalue of f. □

We will also need the following easy criterion for linear independence, whose proof we will leave to the reader.

{lemma:td:li}

Lemma 1.8.5. *Let V be a vector space, let $v_1, \dots, v_m$ be linearly independent vectors in V, and let v be an element of V. The following two statements are equivalent:*

(*a*) *The vectors $v_1, \dots, v_m, v$ are linearly dependent.*

(*b*) *The vector v belongs to $\langle v_1, v_2, \dots, v_m \rangle$.* □

Exercise 1.8.6. Prove this lemma.

The idea that we used in the proof of Lemma 1.8.4 can be generalized a bit, as in the following exercise.

Exercise 1.8.7. Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space, let W be an f-invariant subspace and let v be a vector in V. Show that if v is not in W, then the set $I \coloneqq \{p \in \Bbbk[X] : p(f)(v) \in W\}$ is a non-zero ideal of $\Bbbk[X]$, so that there is a unique monic polynomial c in I such that $I = \{pc : p \in \Bbbk[X]\}$. We call this polynomial c the ***f-conductor*** of v into W.

Triangularizability

Our first application of Lemma 1.8.4 is a criterion to decide if a map is triangularizable. Of course, we say that an endomorphism $f : V \to V$ of a vector space is ***triangularizable*** if there is an ordered basis $\mathscr{B}$ of V such that the matrix $[f]_{\mathscr{B}}$ of f with respect to $\mathscr{B}$ is upper triangular.

{prop:tri}

Proposition 1.8.8. *Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space V. If the minimal polynomial μ of f splits completely over $\Bbbk$, then f is triangularizable.*

Proof. Let us suppose that the minimal polynomial μ splits completely over $\Bbbk$. Since it is monic, there are then a positive integer $k \in \mathbb{N}$, pairwise different scalars $\lambda_1, \dots, \lambda_k \in \Bbbk$ pairwise different, and positive integers $r_1, \dots, r_k \in \mathbb{N}$ such that

$$(X - \lambda_1)^{r_1}(X - \lambda_2)^{r_2} \cdots (X - \lambda_k)^{r_k}.$$

If V is the zero space, then the claim of the proposition is obvious, so we may suppose that V is non-zero. Let $\mathscr{W}$ be the set of all f-invariant subspaces W of V such that the restriction $f_W : W \to W$ is triangularizable. The zero subspace belongs to $\mathscr{W}$, so that $\mathscr{W}$ is not the empty set. As all elements of $\mathscr{W}$ have dimension at most equal to $\dim V$, we can therefore consider the number

$$d \coloneqq \max\{\dim W : W \in \mathscr{W}\}$$

and choose an element U in $\mathscr{W}$ such that $\dim U = d$. As U is in $\mathscr{W}$, it is f-invariant and the restriction $f_U : U \to U$ is triangularizable, so that there exists an ordered basis $\mathscr{B}_U = (u_1, u_2, \dots, u_n)$ of U such that the matrix $[f_U]_{\mathscr{B}_U}$ is upper triangular.

Let us suppose for a moment that U is a proper subspace of V. In that case, Lemma 1.8.4 tells us that there is a vector v in $V \setminus U$ and an eigenvalue λ of f such that $(f - \lambda \cdot \mathrm{id}_V)(v) \in U$. Since v is not in U and $(u_1, u_2, \dots, u_n)$ is an ordered basis for U, we know from Lemma 1.8.5 that $\mathscr{B} \coloneqq (u_1, u_2, \dots, u_n, v)$ is a linearly independent sequence of elements of V. Let $U' \coloneqq \langle u_1, u_2, \dots, u_n, v\rangle$ be the span of $\mathscr{B}$, so that $\mathscr{B}$ is in fact a basis for U'. Now we can make the following observations.

- As $U \subseteq U'$, we have that $f(u_i) \in U \subseteq U'$ for all $i \in [\![n]\!]$. On the other hand, if we put $u \coloneqq (f - \lambda \cdot \mathrm{id}_V)(v)$, then $f(v) = u + \lambda v \in U'$ because $v \in U'$ and $u \in U \subseteq U'$. It follows then from Lemma 1.7.7 that U' is an f-invariant subspace of V.
- The vector u is in U, so there are scalars $a_1, a_2, \dots, a_n$ in $\Bbbk$ such that $u = a_1u_1 + a_2u_2 + \cdots + a_nu_n$. $f(v) = \lambda v + u$, with $u \in U$, the matrix of f_U with respect to the basis $\mathscr{B}$ of U' is the upper triangular block matrix

$$[f_{U'}]_{\mathscr{B}} = \begin{pmatrix} [f_U]_{\mathscr{B}_U} & * \\ 0 & \lambda \end{pmatrix}$$

in which $*$ denotes the column vector with components $(a_1, a_2, \dots, a_n)$. As the matrix $[f_U]_{\mathscr{B}_U}$ is upper triangular, we see that the matrix $[f_{U'}]_{\mathscr{B}}$ is also upper triangular, so that the restriction $f_{U'}$ is triangularizable.

With this we see that U' is an element of the set $\mathscr{W}$, but this is absurd, since

$$\dim U' = \dim U + 1 > \dim U = d = \max\{\dim W : W \in \mathscr{W}\}.$$

This contradiction arose from our supposition that the subspace U is a proper subspace of V, and we can therefore conclude that U is equal to V. As then V belongs to the set $\mathscr{W}$, this tells us that the linear map f is triangularizable, which is what we wanted to show. □

The condition in the proposition is in fact also necessary for triangularizability:

Corollary 1.8.9. *An endomorphism $f : V \to V$ of a finite-dimensional vector space V is triangularizable if and only if its minimal polynomial splits completely over $\Bbbk$.*

Proof. The proposition states precisely that the condition given in the corollary is sufficient for the triangularizability of f, so we need only check that it is also necessary.

In order to do that, let us suppose that the linear map f is diagonalizable, so that there is an ordered basis $\mathscr{B}$ such that the matrix $[f]_{\mathscr{B}}$ is upper triangular, and let n be the dimension of V. If $a_{1,1}$, $a_{2,2}$, …, $a_{n,n}$ are the entries of $[f]_{\mathscr{B}}$ that appear along its diagonal, we know that the characteristic polynomial of f is $(X - a_{1,1})(X - a_{2,2})\cdots(X - a_{n,n})$. In particular, the minimal polynomial of f is a divisor of that product, and is thus splits completely over $\Bbbk$. This proves what we wanted. □

We say that the field $\Bbbk$ is ***algebraically closed*** if every non-constant polynomial in $\Bbbk[X]$ has a root in $\Bbbk$, and when that is the case, in fact, every polynomial in $\Bbbk[X]$ factors as a product of polynomials of degree 1 and therefore splits completely over $\Bbbk$. For example, the field $\mathbb{C}$ of complex numbers is algebraically closed — this is precisely the content of the celebrated *Fundamental Theorem of Algebra*. There are many algebraically closed fields, even though the reader, at this stage, probably knows only of $\mathbb{C}$.

For algebraically closed fields the hypothesis that appears in Proposition 1.8.8 holds automatically, so we have the following result:

Corollary 1.8.10. *Suppose that the field $\Bbbk$ is algebraically closed. Every endomorphism $f : V \to V$ of a finite-dimensional vector space is triangularizable.* □

Exercise 1.8.11. Let V be a finite-dimensional vector space, let n be the dimension of f, and let $f : V \to V$ be a linear map. Prove that there exists an ordered basis $\mathscr{B}$ of V such that the matrix $[f]_{\mathscr{B}}$ is *strictly* upper triangular if and only if the characteristic polynomial of f is X^n.

Exercise 1.8.12. Let V be a finite-dimensional vector space, and let $f : V \to V$ be a linear map. Show that f is triangularizable if and only if there is an ordered basis $\mathscr{B}$ of V such that the matrix $[f]_{\mathscr{B}}$ of f with respect to $\mathscr{B}$ is *lower* triangular.

Diagonalizability

Next we deal with diagonalizability of linear maps. A diagonalizable map is, of course, triangularizable, so what we have already done tells us that its minimal polinomial has to split completely, but this is not enough.

{prop:diag}

Proposition 1.8.13. *Let $f : V \to V$ be an endomorphism of a finite-dimensional vector space V. If the minimal polynomial μ of f splits completely over $\Bbbk$ and is without multiplicities, then f is diagonalizable.*

Proof. Let us suppose that there are a positive integer k and pairwise different scalars $\lambda_1, \lambda_2, \dots, \lambda_k$ in $\Bbbk$ such that the minimal polynomial of f is

$$\mu = (X - \lambda_1)(X - \lambda_2)\cdots(X - \lambda_k).$$

For each $i \in \{1, 2, \dots, k\}$ let $E_{\lambda_i}(f)$ be the eigenspace of f corresponding to λ_i, and let us consider the subspace

$$W \coloneqq E_{\lambda_1}(f) + E_{\lambda_2}(f) + \cdots + E_{\lambda_k}(f)$$

of V. We know that the subspaces $E_{\lambda_1}(f), E_{\lambda_2}(f), \dots, E_{\lambda_k}(f)$ are f-invariant from Lemma 1.7.1, and the result of Exercise 1.7.8 implies that their sum W is also f-invariant.

Let us suppose that W is a proper subspace of V. Lemma 1.8.4 then tells us that there is a vector $v \in V$ and an eigenvalue λ of f such that $v \notin W$ and $(f - \lambda \cdot \mathrm{id}_V)(v) \in W$. Let us put

$$w \coloneqq f(v) - \lambda v,$$

which is an element of W. Since λ is an eigenvalue of f, it is a root of the characteristic polynomial of f and, according to Proposition 1.6.1, it is therefore also a root of the minimal polynomial of f. It follows from this that there is a polynomial $p \in \Bbbk[X]$ such that

$$\mu(X) = (X - \lambda)p(X). \tag{1.11}$$ {eq:mux}

This implies that

$$0 = \mu(f)(v) = \big((f - \lambda \cdot \mathrm{id}_V) \circ p(f)\big)(v) = f(p(f)(v)) - \lambda p(f)(v),$$

so that $p(f)(v)$ belongs to the eigenspace $E_\lambda(f)$ and, therefore, that

$$p(f)(v) \in W.$$

{eq:pfv}

On the other hand, the polynomial $p(X) - p(\lambda)$ also has λ as a root, so there is another polynomial $q \in \Bbbk[X]$ such that

$$p(X) - p(\lambda) = q(X)(X - \lambda),$$

and using this we see that

$$\begin{aligned} p(f)(v) - p(\lambda) \cdot v &= \big(p(f) - p(\lambda) \cdot \mathrm{id}_V\big)(v) = \big(q(f) \circ (f - \lambda \cdot \mathrm{id}_V)\big)(v) \\ &= q(f)\big(f(v) - \lambda v\big) = q(f)(w). \end{aligned}$$

Rearranging this equality, we find that

$$p(\lambda) \cdot v = p(f)(v) + q(f)(w)$$

and — since $p(f)(v) \in W$ and also $q(f)(w)$, because $w \in W$ and W is f-invariant — we conclude that $p(\lambda) \cdot v \in W$. As $v \notin W$, this allows us to conclude that $p(\lambda) = 0$. Going back to the equality (1.11) we see that λ is a root of μ of multiplicity at least 2: this is absurd, since by hypothesis all the roots of μ are simple.

This contradiction arouse from our hypothesis that the subspace W is a proper subspace of V. We must therefore have that $W = V$, and this implies, according to Proposition 1.4.3, that the linear map f is diagonalizable. □

An easy application of the proposition is the following corollary.

Corollary 1.8.14. *Let $f : V \to V$ be a diagonalizable endomorphism of a finite-dimensional vector space V. If W is an f-invariant subspace of V, then the restriction $f_W : W \to W$ is also diagonalizable.*

Proof. Let W be an f-invariant subspace of V. Since f is diagonalizable, the minimal polynomial μ_f of f splits completely over $\Bbbk$ and has simple roots. The minimal polynomial of the restriction μ_{f_W} divides μ_f, according to Proposition 1.7.13, and then it also splits completely over $\Bbbk$ and has simple roots. Proposition 1.8.13 allows us then to conclude that the restriction f_W is also diagonalizable. □

Another extremely useful application of the proposition to linear maps satisfying a rather special condition is:

Corollary 1.8.15. *If $f : V \to V$ is an endomorphism of a finite-dimensional complex vector space and there exists a positive integer n such that $f^n = \mathsf{id}_V$, then f is diagonalizable.*

In this situation we say that the endomorphism f has ***finite order***.

Proof. Let $f : V \to V$ be an endomorphism of a finite-dimensional complex vector space and suppose that there is a positive integer n such that $f^n = \mathsf{id}_V$. This means that the polynomial $p(X) \coloneqq X^n - 1 \in \mathbb{C}[X]$ has $p(f) = 0$ and is therefore divisible by the minimal polynomial μ_f of f. That polynomial p can be factored as

$$X^n - 1 = \prod_{k=0}^{n-1}(X - e^{2\pi ik/n}),$$

so it splits completely over $\mathbb{C}$ and has simple roots. As μ_f divides it, the same is true of μ_f and Proposition 1.8.13 tells us that f is diagonalizable. □

The sufficient condition given by Proposition 1.8.13 for diagonalizability is in fact also necessary, so we have a complete characterization of diagonalizability in terms of minimal polynomials:

Corollary 1.8.16. *An endomorphism of a finite-dimensional vector space is diagonalizable if and only if its minimal polynomial splits completely over $\Bbbk$ and is without multiplicities.*

Proof. The proposition tells us that an endomorphism of a finite-dimensional vector space is diagonalizable if its minimal polynomial splits completely without multiplicities. We therefore only have to prove that the necessity of this condition.

Let us suppose that $f : V \to V$ is an endomorphism of a finite-dimensional vector space that is diagonalizable, so that there is an ordered basis $\mathscr{B} = (v_1, v_2, \dots, v_n)$ for V whose elements are eigenvectors of f. Let $\lambda_1, \dots, \lambda_k$ be the eigenvalues of f listed without repetitions, and let us consider the polynomial

$$p(X) = (X - \lambda_1)(X - \lambda_2)\cdots(X - \lambda_k) \in \Bbbk[X].$$

We claim that $p(f) = 0$, and to verify this claim is is enough that we check that $p(f)(v_i) = 0$ for all $i \in \{1, 2, \dots, n\}$, since $\mathscr{B}$ is a basis for V. Let $i \in \{1, 2, \dots, n\}$. As v_i is an eigenvector for f and since $\lambda_1, \dots, \lambda_k$ are the eigenvalues of f, there exists some index j in $\{1, 2, \dots, k\}$ such that $f(v_i) = \lambda_j v_i$. We therefore have that $p(f)(v_i) = p(\lambda_j) \cdot v_i = 0$, as we wanted.

Now, since $p(f) = 0$, the characteristic property of the minimal polynomial μ of f implies at once that μ divides f: it follows from this, since the polynomial p completely splits over $\Bbbk$ and is without multiplicities, that μ has the same two properties. This proves the corollary. □

§1.9. Direct sum decompositions

Let V be a vector space. We say that a finite family of subspaces $W_1, W_2, \dots, W_k$ is ***independent*** if the following condition holds:

if $w_1 \in W_1, w_2 \in W_2, \dots, w_k \in W_k$ *are such that* $w_1+w_2+\cdots+w_k = 0$*, then* $w_1 = w_2 = \cdots = w_k = 0$.

This notion of independence of subspaces generalizes the usual one of linear independence of vectors, in the following precise way:

{lemma:indep:one}

Lemma 1.9.1. *Let V be a vector space, and let $v_1, v_2, \dots, v_k$ be non-zero vectors in V. The subspaces $\langle v_1\rangle, \langle v_2\rangle, \dots, \langle v_k\rangle$ are independent if and only if the vectors $v_1, v_2, \dots, v_k$ are linearly independent.*

Proof. Let us suppose first that the subspaces $\langle v_1\rangle, \langle v_2\rangle, \dots, \langle v_k\rangle$ are independent, and that $a_1, a_2, \dots, a_k \in \Bbbk$ are scalars such that $a_1v_1 + a_2v_2 + \cdots + a_kv_k = 0$. If for each $i \in [\![k]\!]$ we put $w_i \coloneqq a_iv_i$, then we have $w_i \in \langle v_i\rangle$ for each $i \in [\![k]\!]$ and $w_1+w_2+\cdots+w_k = 0$: the hypothesis implies then that $w_1 = w_2 = \cdots = w_k = 0$. As the vectors $v_1, v_2, \dots, v_k$ are all non-zero, this implies in turn that $a_1 = a_2 = \cdots = a_k = 0$. This proves that the condition given by the lemma for the subspaces $\langle v_1\rangle, \langle v_2\rangle, \dots, \langle v_k\rangle$ to be independent is necessary.

Let us now show that it is also sufficient. Let us suppose that the vectors $v_1, v_2, \dots, v_k$ are linearly independent, and let $w_1, w_2, \dots, w_k$ be elements of the subspaces $\langle v_1\rangle, \langle v_2\rangle, \dots, \langle v_k\rangle$ such that $w_1 + w_2 + \cdots + w_k = 0$. There are scalars $a_1, a_2, \dots, a_k$ in $\Bbbk$ such that $w_1 = a_1v_1$, $w_2 = a_2v_2, \dots,$ $w_k = a_kv_k$, and the hypothesis is then that $a_1v_1+a_2v_2+\cdots+a_kv_k = 0$: since the vectors $v_1, v_2, \dots, v_k$ are linearly independent, we see that $a_1 = a_2 = \cdots = a_k = 0$ and thus that $w_1 = w_2 = \cdots = w_k = 0$. This shows that the subspaces $\langle v_1\rangle, \langle v_2\rangle, \dots, \langle v_k\rangle$ are independent, as we want. □

Lemma 1.9.1 is a special case of part of the following characterization of independence:

{lemma:indep:man

Proposition 1.9.2. *Let $W_1, W_2, \dots, W_k$ be non-zero subspaces of a vector space V. The following two statements are equivalent:*

(a) The subspaces $W_1, W_2, \dots, W_r$ are independent.

(b) Whenever $w_1, w_2, \dots, w_r$ are non-zero elements of $W_1, W_2, \dots, W_r$, respectively, we have that $w_1, w_2, \dots, w_r$ are linearly independent.

Proof. Let us suppose first that the subspaces $W_1, W_2, \dots, W_k$ are independent, and let $w_1, w_2, \dots, w_r$ be non-zero elements of $W_1, W_2, \dots, W_k$, respectively, and let $a_1, a_2, \dots, a_k \in \Bbbk$ be scalars such that $a_1w_1 + a_2w_2 + \cdots + a_kw_k = 0$. The vectors $a_1w_1, a_2w_2, \dots, a_kw_k$ belong to $W_1, W_2, \dots, W_k$, respectively, and their sum is 0, so the hypothesis on the subspaces $W_1, W_2, \dots, W_k$ implies that $a_1w_1 = a_2w_2 = \cdots = a_kw_k = 0$. As the vectors $w_1, w_2, \dots, w_k$ are all non-zero, this allows us to concldue that in fact $a_1 = a_2 = \dots = a_k = 0$. We thus see that the vectors $w_1, w_2, \dots, w_k$ are linearly independent, and this proves the implication $(a) \Rightarrow (b)$.

In order to prove the converse implication, let us suppose the statement (b) holds, and let w_1, $w_2, \dots, w_k$ be elements of $W_1, W_2, \dots, W_k$, respectively, such that $w_1 + w_2 + \dots + w_k = 0$. For each $i \in [\![k]\!]$,

- either $w_i \neq 0$, and then we put $v_i \coloneqq w_i$ and $a_i \coloneqq 1$,
- or $w_i = 9$, and in this case we chose an arbitrary non-zero element v_i in W_i and put $a_i \coloneqq= 0$.

In this way we have that $v_1, v_2, \dots, v_k$ are non-zero elements of $W_1, W_2, \dots, W_k$, respectively, so that they are linearly independent, and that $a_i v_i = w_i$ for all $i \in [\![k]\!]$, so that $a_1 v_1 + a_2 v_2 + \dots + a_k v_k = 0$. We therefore have that $a_1 = a_2 = \dots = a_k = 0$, so that $w_1 = w_2 = \dots = w_k = 0$. This shows that the subspaces $W_1, W_2, \dots, W_k$ are independent. □

Let W_1 and W_2 be two subspaces of a vector space V, and let us prove that

$$\text{the subspaces } W_1 \text{ and } W_2 \text{ are independent if and only if } W_1 \cap W_2 = 0. \tag{1.12}$$

{eq:indep:2}

First, let us suppose that W_1 and W_2 are independent and consider a vector v in $W_1 \cap W_2$. As $v + (-v) = 0$, $v \in W_1$ and $-v \in W_2$, then hypothesis implies at once that $v = 0$: this shows that $W_1 \cap W_2 = 0$ and, therefore, that the condition in (1.12) is necessary. To prove the converse, let us suppose that $W_1 \cap W_2 = 0$ and that $w_1 \in W_1$ and $w_2 \in W_2$ are vectors such that $w_1 + w_2 = 0$. In that case we have that $W_1 \ni w_1 = -w_2 \in W_2$, so that $w_1 \in W_1 \cap W_2 = 0$: this tells us that $w_1 = 0$ and, of course, that then also $w_2 = -w_1 = 0$, so that W_1 and W_2 are independent subspaces.

Our claim (1.12) shows that the independence of two subspaces has a very simple restatement in terms of their intersection. A similar result is false for families of subspaces with more than two elements, as the following example shows.

Example 1.9.3. Let $V = \mathbb{Q}^2$, let $k \in \mathbb{N}$, and for each $i \in \{1, 2, \dots, k\}$ let $W_i \coloneqq \langle e_1 + i e_2 \rangle$. We have that $W_1 \cap W_2 \cap \dots \cap W_k = 0$ and, in fact, that $W_i \cap W_j = 0$ whenever i and j are two different elements of $\{1, 2, \dots, k\}$, but the family of k subspaces $W_1, \dots, W_k$ is not independent unless $k \leq 2$. This follows immediately from Lemma 1.9.1.

The equivalence of the two statements listed in the following proposition provides a way to *fix* this, giving a criterion for independence in terms of intersections and sums.

Proposition 1.9.4. *Let V be a vector space, and let $W_1, W_2, \dots, W_k$ be subspaces of V. The following statements are equivalent:*

(a) *The subspaces $W_1, W_2, \dots, W_k$ are independent.*

(b) *For each $j \in \{2, \dots, k\}$ we have that $(W_1 + W_2 + \dots + W_{j-1}) \cap W_j = 0$.*

Proof. Let us suppose that the statement (a) is true, so that the subspaces $W_1, W_2, \dots, W_k$ are independent, and let j be an element of $\{2, \dots, k\}$. Let v be an element of $(W_1 + W_2 + \dots + W_{j-1}) \cap W_j$. Since v is in $W_1 + W_2 + \dots + W_{j-1}$, there exist $w_1 \in W_1$, $w_2 \in W_2$, …, $w_{j-1} \in W_{j-1}$ such that $v = w_1 + w_2 + \dots + w_{j-1}$. Let us put $w_j \coloneqq -v$, and $w_i = 0$ for each $i \in \{j+1, \dots, j\}$. As $w_1 + w_2 + \dots + w_k = 0$, the independence of the subspaces $W_1, W_2, \dots, W_k$ implies that $w_1 = w_2 = \dots = w_k = 0$ and, in particular, that $v = -w_j = 0$. This shows that the only element of $(W_1 + W_2 + \dots + W_{j-1}) \cap W_j$ is the zero vector, so that the statement (b) holds.

Let us now suppose that the statement (b) holds, let $w_1 \in W_1$, $w_2 \in W_2$, …, $w_k \in W_k$ be vectors such that $w_1 + w_2 + \dots + w_k = 0$, and, to reach a contradiction, let us suppose that the set $I \coloneqq \{i \in [\![k]\!] : w_i \neq 0\}$ is not empty. In that case we may consider the number $j \coloneqq \max I$.

The definition of j implies that $w_{i+1} = w_{i+2} = \dots = w_k = 0$, so that $w_1 + w_2 + \dots + w_j = 0$ and therefore that $-w_j = w_1 + w_2 + \dots + w_{j-1}$: this implies that w_j is an element of the intersection $(W_1 + W_2 + \dots + W_{j-1}) \cap W_j$, and this is absurd for that intersection is zero while w_j is not. This contradiction arose from out supposition that the set I is not empty, so we see that we must have $w_1 = w_2 = \dots = w_k = 0$. This shows that the subspaces $W_1, W_2, \dots, W_k$ are independent, so that the statement (a) holds. The proof of the proposition is therefore complete. □

There is a third condition equivalent to independence that is often useful — we leave the proof of this as an exercise for the reader — and that, in fact, is used in many textbooks as the definition of independence.

Exercise 1.9.5. A finite family of subspaces $W_1, W_2, \dots, W_k$ of a vector space V is independent if and only if for each $i \in \{1, 2, \dots, k\}$ we have that $W_i \cap (W_1 + \dots + W_{i-1} + W_{i+1} + \dots + W_k) = 0$.

We can also give conditions for independence in terms of bases:

Proposition 1.9.6. *Let V be a vector space, let $W_1, W_2, \dots, W_k$ be finite-dimensional subspaces of V, and let $d_1, d_2, \dots, d_k$ be their dimensions. The following statements are equivalent:*

(a) *The subspaces $W_1, W_2, \dots, W_k$ are independent.*

(b) *Whenever $(v_{1,1}, v_{1,2}, \dots, v_{1,d_1})$, $(v_{2,1}, v_{2,2}, \dots, v_{2,d_2})$, …, $(v_{k,1}, v_{k,2}, \dots, v_{k,d_k})$ are ordered bases for the subspaces $W_1, W_2, \dots, W_k$ the sequence*

$$(v_{1,1}, v_{1,2}, \dots, v_{1,d_1}, v_{2,1}, v_{2,2}, \dots, v_{2,d_2}, \dots, \dots, v_{k,1}, v_{k,2}, \dots, v_{k,d_k})$$

is an ordered basis for $W_1 + W_2 + \dots + W_k$.

(c) *There exist ordered bases $(v_{1,1}, v_{1,2}, \dots, v_{1,d_1})$, $(v_{2,1}, v_{2,2}, \dots, v_{2,d_2})$, …, $(v_{k,1}, v_{k,2}, \dots, v_{k,d_k})$ of the subspaces $W_1, W_2, \dots, W_k$ such that the sequence*

$$(v_{1,1}, v_{1,2}, \dots, v_{1,d_1}, v_{2,1}, v_{2,2}, \dots, v_{2,d_2}, \dots, \dots, v_{k,1}, v_{k,2}, \dots, v_{k,d_k})$$

is an ordered basis for $W_1 + W_2 + \dots + W_k$.

When these conditions hold, then we have that

$$\dim(W_1 + W_2 + \cdots + W_k) = \dim W_1 + \dim W_2 + \cdots + \dim W_k.$$

This proposition has an easy yet long and notationally annoying proof.

Proof. Let us write W for the subspace $W_1 + W_2 + \cdots + W_k$.

$(a) \Rightarrow (b)$ Let us suppose that the statement (a) holds, and that $\mathscr{B}_1 = (v_{1,1}, v_{1,2}, \dots, v_{1,d_1})$, $\mathscr{B}_2 = (v_{2,1}, v_{2,2}, \dots, v_{2,d_2})$, …, $\mathscr{B}_k = (v_{k,1}, v_{k,2}, \dots, v_{k,d_k})$ are ordered bases for the subspaces W_1, W_2, …, W_k, and let us consider the sequence

$$\mathscr{B} \coloneqq (v_{1,1}, v_{1,2}, \dots, v_{1,d_1}, v_{2,1}, v_{2,2}, \dots, v_{2,d_2}, \dots, \dots, v_{k,1}, v_{k,2}, \dots, v_{k,d_k}).$$

We have to show that $\mathscr{B}$ is an ordered basis for W.

Let w be an element of W, so that there are vectors $w_1, w_2, \dots, w_k$ in W_1, W_2, …, W_k such that $w = w_1 + w_2 + \cdots + w_k$. If $i \in [\![k]\!]$, then the span of the sequence $\mathscr{B}_i$ is W_i, so there exist scalars $a_{i,1}$, $a_{i,2}$, …, a_{i,d_i} in $\Bbbk$ such that $w_i = a_{i,1}v_{i,1} + a_{i,2}v_{i,2} + \cdots + a_{i,d_i}v_{i,d_i}$. We therefore have that

$$\begin{aligned} w &= w_1 + w_2 + \cdots + w_k \\ &= a_{1,1}v_{1,1} + a_{1,2}v_{1,2} + \cdots + a_{1,d_1}v_{1,d_1} + a_{2,1}v_{2,1} + a_{2,2}v_{2,2} + \cdots + a_{2,d_2}v_{2,d_2} \\ &\qquad + \cdots + a_{k,1}v_{k,1} + a_{k,k}v_{k,k} + \cdots + a_{k,d_k}v_{k,d_k} \end{aligned}$$

and we see that w is in the span of the sequence $\mathscr{B}$.

Next, let us suppose that

$$a_{1,1},\ a_{1,2},\ \dots,\ a_{1,d_1}, a_{2,1},\ a_{2,2},\ \dots,\ a_{2,d_2}, \dots, a_{i,1},\ a_{i,2},\ \dots,\ a_{i,d_i} \tag{1.13}$$

{eq:as}

are scalars in $\Bbbk$ such that

$$\begin{aligned} 0 &= a_{1,1}v_{1,1} + a_{1,2}v_{1,2} + \cdots + a_{1,d_1}v_{1,d_1} + a_{2,1}v_{2,1} + a_{2,2}v_{2,2} + \cdots + a_{2,d_2}v_{2,d_2} \\ &\qquad + \cdots + a_{k,1}v_{k,1} + a_{k,k}v_{k,k} + \cdots + a_{k,d_k}v_{k,d_k}. \end{aligned}$$

We can then consider the vectors

$$\begin{aligned} w_1 &\coloneqq a_{1,1}v_{1,1} + a_{1,2}v_{1,2} + \cdots + a_{1,d_1}v_{1,d_1}, \\ w_2 &\coloneqq a_{2,1}v_{2,1} + a_{2,2}v_{2,2} + \cdots + a_{2,d_2}v_{2,d_2}, \\ \vdots\ &\ \ \vdots \\ w_k &\coloneqq a_{k,1}v_{k,1} + a_{k,2}v_{k,2} + \cdots + a_{k,d_k}v_{k,d_k}, \end{aligned}$$

which belong to the subspaces W_1, W_2, …, W_k, respectively. We have that $w_1 + w_2 + \cdots + w_k = 0$, so the hypothesis (a) implies that $w_1 = w_2 = \cdots = w_k = 0$. In particular, for each $i \in [\![k]\!]$ we have that $a_{i,1}v_{i,1} + a_{i,2}v_{i,2} + \cdots + a_{i,d_1}v_{i,d_i} = 0$ and therefore, since the sequence $\mathscr{B}_i$ is linearly independent, we have that $a_{i,1} = a_{i,2} = \cdots = a_{i,d_i} = 0$. This tells us that each the the scalars listed in (1.13) is zero and thus that the sequence $\mathscr{B}$ is linearly independent.

Putting everything together, we can conclude that the sequence $\mathscr{B}$ is a basis for W, so that the statement (b) holds.

$(b) \Rightarrow (c)$ Let us suppose that the statement (b) holds, and let us choose ordered bases $\mathscr{B}_1 = (v_{1,1}, v_{1,2}, \dots, v_{1,d_1})$, $\mathscr{B}_2 = (v_{2,1}, v_{2,2}, \dots, v_{2,d_2})$, …, $\mathscr{B}_k = (v_{k,1}, v_{k,2}, \dots, v_{k,d_k})$ for the subspaces W_1, W_2, …, W_k arbitrarily — this is, of course, possible. The hypothesis then tells us that the sequence

$$\mathscr{B} \coloneqq (v_{1,1}, v_{1,2}, \dots, v_{1,d_1}, v_{2,1}, v_{2,2}, \dots, v_{2,d_2}, \dots, \dots, v_{k,1}, v_{k,2}, \dots, v_{k,d_k})$$

is an ordered basis for the subspace W, and we see that the statement (c) holds.

$(c) \Rightarrow (a)$ Finally, let us suppose that there exist ordered bases $\mathscr{B}_1 = (v_{1,1}, v_{1,2}, \dots, v_{1,d_1})$, $\mathscr{B}_2 = (v_{2,1}, v_{2,2}, \dots, v_{2,d_2})$, …, $\mathscr{B}_k = (v_{k,1}, v_{k,2}, \dots, v_{k,d_k})$ for the subspaces W_1, W_2, …, W_k such that the sequence

$$\mathscr{B} \coloneqq (v_{1,1}, v_{1,2}, \dots, v_{1,d_1}, v_{2,1}, v_{2,2}, \dots, v_{2,d_2}, \dots, \dots, v_{k,1}, v_{k,2}, \dots, v_{k,d_k}) \tag{1.14}$$ {eq:bsbs}

is an ordered basis for the subspace W and prove that the subspaces W_1, W_2, …, W_k are independent. To do that, let us suppose that w_1, w_2, …, w_k are element of those subspaces such that $w_1 + w_2 + \cdots + w_k = 0$. For each $i \in [\![k]\!]$ the sequence $\mathscr{B}_i$ is an ordered basis for W_i, so there exist scalars $a_{i,1}$, $a_{i,2}$, …, a_{i,d_i} in $\Bbbk$ such that $w_i = a_{i,1}v_{i,1} + a_{i,2}v_{i,2} + \cdots + a_{i,d_i}v_{i,d_i}$, and therefore we have that

$$\begin{aligned} 0 &= w_1 + w_2 + \cdots + w_k \\ &= a_{1,1}v_{1,1} + a_{1,2}v_{1,2} + \cdots + a_{1,d_1}v_{1,d_1} + a_{2,1}v_{2,1} + a_{2,2}v_{2,2} + \cdots + a_{2,d_2}v_{2,d_2} \\ &\qquad + \cdots + a_{k,1}v_{k,1} + a_{k,k}v_{k,k} + \cdots + a_{k,d_k}v_{k,d_k}. \end{aligned}$$

Since the sequence $\mathscr{B}$ of (1.14) is linearly independent, we see that all the scalars $a_{1,1}$, $a_{1,2}$, …, a_{1,d_1}, $a_{2,1}$, $a_{2,2}$, …, a_{2,d_2}, ……, $a_{k,1}$, $a_{k,2}$, …, a_{k,d_k} are equal to 0, and therefore, of course, all the vectors w_1, w_2, …, w_k are zero. This proves that the subspaces W_1, W_2, …, W_k are independent, and thus that the statement (a) holds. □

If V is a vector space and W, W_1, …, W_k are subspaces of V, we write

$$W = W_1 \oplus W_2 \oplus \cdots \oplus W_k \tag{1.15}$$ {eq:decom}

to mean that

- W is equal to the sum $W_1 + W_2 + \cdots + W_k$, and
- the subspaces W_1, W_2, …, W_k are independent.

In this situation we say that W is the ***direct sum*** of the subspaces $W_1, W_2, \dots, W_k$, and that (1.15) is a ***direct sum decomposition*** of W.

The key reason for which this notion is interesting is that direct sum decompositions play a role similar in linear algebra as partitions of sets in set theory or combinatorics. The following lemma makes explicit one of the ways we often use direct sums:

{lemma:oplus:desc

Lemma 1.9.7. *Let V is a vector space, let W, W_1, …, W_k be subspaces of V, and suppose that $W = W_1 \oplus W_2 \oplus \cdots \oplus W_k$. If v is an element of V, then there is a unique choice of $w_1 \in W_1$, $w_2 \in W_2, \dots, w_k \in W_k$ such that $v = w_1 + w_2 + \cdots + w_k$.*

Proof. Let v be an element of W. As $W = W_1 + W_2 + \cdots + W_k$, there are vectors $w_1 \in W_1$, $w_2 \in W_2$, …, $w_k \in W_k$ such that $w = w_1 + w_2 + \cdots + w_k$. This proves the existence claim of the lemma.

Let us now suppose that the vectors $w_1' \in W_1$, $w_2' \in W_2$, …, $w_k' \in W_k$ are also such that $w = w_1 + w_2 + \cdots + w_k$. In that case we have that $w_1 - w_1' \in W_1$, $w_2 - w_2' \in W_2, \dots, w_k - w_k' \in W_k$, and that

$$(w_1 - w_1') + (w_2 - w_2') + \cdots + (w_k - w_k') = 0,$$

and since the hypothesis implies that the subspaces W_1, W_2, …, W_k are independent we can conclude that $w_1 - w_1' = w_2 - w_2' = \cdots = w_k - w_k' = 0$, so that $w_i = w_i'$ for all $i \in [\![k]\!]$. This proves the uniqueness claim of the lemma. □

Direct sum decompositions, in some sense, generalize bases, as the result of the following exercise suggests.

Exercise 1.9.8. Let V be a vector space, and let $v_1, v_2, \dots, v_n$ be non-zero elements of V. We have a direct sum decomposition $V = \langle v_1 \rangle \oplus \langle v_2 \rangle \oplus \cdots \oplus \langle v_n \rangle$ if and only if the sequence $(v_1, v_2, \dots, v_n)$ is an ordered basis for V.

The following is a well-known example of a direct sum decomposition.

{ex:mats:as}

Example 1.9.9. Let n be a positive integer. Let us write S_+ and S_- for the subspaces of $\mathrm{M}_n(\mathbb{R})$ of the symmetric and anti-symmetric matrices, respectively, so that

$$S_+ = \{A \in \mathrm{M}_n(\mathbb{R}) : A^t = A\}, \qquad S_- = \{A \in \mathrm{M}_n(\mathbb{R}) : A^t = -A\}.$$

If A is an element of $S_+ \cap S_-$, then we have that $A^t = A$ and that $A^t = -A$, so that in fact $A = -A$ and thus that $A = 0$: this shows that $S_+ \cap S_- = 0$. As we observed above, this implies that S_+ and S_- are independent subspaces of $\mathrm{M}_n(\mathbb{R})$. On the other hand, if A is an arbitrary element of $\mathrm{M}_n(\mathbb{R})$

then we have that

$$A = \underbrace{\left(\frac{A + A^t}{2}\right)}_{\in S_+} + \underbrace{\left(\frac{A - A^t}{2}\right)}_{\in S_-} \in S_+ + S_-,$$

and we see with thus that $\mathrm{M}_n(\mathbb{R}) = S_+ + S_-$. Putting everything together, we can conclude that we have a direct sum decomposition

$$\mathrm{M}_n(\mathbb{R}) = S_+ \oplus S_-$$

of the space of all real $n \times n$ matrices. The exact same statement, with the exact same proof, is true if we replace the field $\mathbb{R}$ by any other field in which 2 is non-zero.

We leave the details of a very similar example to the responsability of the reader.

{exer:pols:eo}

Exercise 1.9.10. Let $\mathbb{R}[X]$ be the vector space of all real polynomials, and let $\mathscr{E}$ and $\mathscr{O}$ be the subspaces of all even and odd polynomials, respectively, so that

$$\mathscr{E} = \{p \in \mathbb{R}[X] : p(-X) = p(X)\}, \qquad \mathscr{O} = \{p \in \mathbb{R}[X] : p(-X) = -p(X)\}.$$

Prove that $\mathbb{R}[X] = \mathscr{E} \oplus \mathscr{O}$.

For us, the most relevant example of a direct sum decomposition is the one described in the following proposition.

{prop:oplus:eig}

Proposition 1.9.11. *Let $f : V \to V$ be a diagonalizable endomorphism of a vector space V that has finitely many eigenvalues. If $\lambda_1, \lambda_2, \dots, \lambda_k$ are the eigenvalues of f listed without repetitions, then*

$$V = E_{\lambda_1}(f) \oplus E_{\lambda_2}(f) \oplus \cdots \oplus E_{\lambda_k}(f).$$

Proof. Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be the eigenvalues of f listed without repetitions. If $w_1, w_2, \dots, w_k$ are non-zero elements of $E_{\lambda_1}(f), E_{\lambda_2}(f), \dots, E_{\lambda_k}(f)$, respectively, then they are in fact eigenvectors with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$, and Proposition 1.3.2 tells us that they are linearly independent. It follows from this and Proposition 1.9.2 that the eigenspaces $E_{\lambda_1}(f), E_{\lambda_2}(f), \dots, E_{\lambda_k}(f)$ are independent subspaces of V.

Let now v be an arbitrary element of V. As f is diagonalizable, we know from Proposition 1.4.1 that v is equal to a linear combination of eigenvectors of f, and this tells us that it is in the span of the union $E_{\lambda_1}(f) \cup E_{\lambda_2}(f) \cup \cdots \cup E_{\lambda_k}(f)$, which is precisely the sum $E_{\lambda_1}(f) + E_{\lambda_2}(f) + \cdots + E_{\lambda_k}(f)$. This tells us that this last sum coincides with V, and therefore, with what we proved above, that in fact $V = E_{\lambda_1}(f) \oplus E_{\lambda_2}(f) \oplus \cdots \oplus E_{\lambda_k}(f)$, as we want. □

This proposition, in fact, has as special cases the direct sum decompositions that we described in Example 1.9.9 and in Exercise 1.9.10. Let us see how.

Example 1.9.12. Let n be a positive integer, and let $T : \mathrm{M}_n(\mathbb{R}) \to \mathrm{M}_n(\mathbb{R})$ be the linear map such that $T(A) = A^t$ for each matrix $A \in \mathrm{M}_n(\mathbb{R})$. We clearly have that $T^2 = \mathrm{id}_{\mathrm{M}_n(\mathbb{R})}$, so the polynomial $X^2 - 1 = (X-1)(X+1)$ vanishes on T and is therefore divisible by the minimal polynomial μ_T of T. As $(X-1)(X+1)$ splits completely over $\mathbb{R}$ without multiplicities, so does μ_T, and this implies, as we know, that T is a diagonalizable map. Since $\mathrm{M}_n(\mathbb{R})$ is finite-dimensional, T has finitely many eigenvalues, and the proposition thus applies to T. The eigenspace $E_1(T)$ of T corresponding to the eigenvalue 1 is the space of all matrices A such that $T(A) = A$, that is, such that $A^t = A$, and that is the space S_+ of all symmetric matrices. Similarly, the eigenspace $E_{-1}(T)$ corresponding to the eigenvalue -1 is the space S_- of all anti-symmetric matrices. It follows from this that the direct sum decomposition $\mathrm{M}_n(\mathbb{R}) = E_1(T) \oplus E_{-1}(T)$ given by Proposition 1.9.11 coincides with the direct sum decomposition $\mathrm{M}_n(\mathbb{R}) = S_+ \oplus S_-$ that we described in Example 1.9.9.

Exercise 1.9.13. Let $S : \mathbb{R}[X] \to \mathbb{R}[X]$ be the linear map such that $S(p) = p(-X)$ for all $p \in \mathbb{R}[X]$. Show that S has exactly two eigenvalues, 1 and -1, that it is diagonalizable, and that the eigenspaces $E_1(S)$ and $E_{-1}(S)$ coincide with the subspaces $\mathscr{E}$ and $\mathscr{O}$ of even and odd polynomials of Exercise 1.9.10. Conclude that the direct sum decomposition $\mathbb{R}[X] = \mathscr{E} \oplus \mathscr{O}$ of that exercise is a special case of the decompositions given by Proposition 1.9.11.

Observation 1.9.14. In Proposition 1.9.11 we have the hypothesis that the linear map $f : V \to V$ have finitely many eigenvalues. This hypothesis holds automatically if the vector space V is finite-dimensional, as we know, but in general need not be true. For example, the linear map $Q : \mathbb{R}[X] \mapsto \mathbb{R}[X]$ such that $Q(p) = p(2X)$ for each polynomial $p \in \mathbb{R}[X]$ is diagonalizable and its eigenvalues are precisely the integers of the form 2^i with $i \in \mathbb{N}_0$. For such a linear map a result exactly like Proposition 1.9.11 holds, but with an infinite direct sum. Since we have not defined this, we omit the details.

{lemma:oplus-asso

Lemma 1.9.15. *Let W be a finite-dimensional vector space, and let U and V be subspaces of W such that $W = U \oplus V$. If $U_1, U_2, \dots, U_r$ and $V_1, V_2, \dots, V_s$ are subspaces of U and of V, respectively, such that $U = U_1 \oplus U_2 \oplus \cdots \oplus U_r$ and $V = V_1 \oplus V_2 \oplus \cdots \oplus V_r$, then*

$$V = U_1 \oplus U_2 \oplus \cdots \oplus U_r \oplus V_1 \oplus V_2 \oplus \cdots \oplus V_s.$$

Proof. □

§1.10. Indecomposable endomorphisms

We say that an endomorphism $f : V \to V$ is ***decomposable*** if there exist non-zero f-invariant subspaces W_1 and W_2 of V such that $V = W_1 \oplus W_2$, and that it is ***indecomposable*** if $V \neq 0$ and f is not decomposable. These two notions have straightforward interpretations in terms of matrices.

Indeed, if $f : V \to V$ is a decomposable endomorphism of a finite-dimensional vector space V and W_1 and W_2 are two non-zero f-invariant subspaces of V such that $V = W_1 \oplus W_2$, and $\mathscr{B}_1 = (v_1, v_2, \dots, v_r)$ and $\mathscr{B}_2 = (w_1, \dots, w_s)$ are ordered bases of W_1 and of W_2, then

$$\mathscr{B} = (v_1, v_2, \dots, v_r, w_1, \dots, w_s)$$

is an ordered basis of V with respect to which the matrix of f is of the form

$$[f]_{\mathscr{B}} = \begin{pmatrix} [f_{W_1}]_{\mathscr{B}_1} & 0 \\ 0 & [f_{W_2}]_{\mathscr{B}_2} \end{pmatrix},$$

with $f_{W_1} : W_1 \to W_1$ and $f_{W_2} : W_2 \to W_2$ the restrictions of f to the subspaces W_1 and W_2, as usual. This tells us that when the linear map f is decomposable we can find a basis of V with respect to which the matrix of f is a block diagonal matrix in a non-trivial way, and it is easy to check that this is in fact also a sufficient condition for decomposability. In other words, we have that

> *a linear map $f : V \to V$ is decomposable if and only if there is an ordered basis of V with respect to which the matrix of f is a block diagonal matrix in a non-trivial way.*

Of course, it follows from this that an endomorphism $f : V \to V$ of a non-zero finite-dimensional vector space V is indecomposable if and only if there is no way to choose a basis of V with respect to which its matrix is block diagonal.

Observation 1.10.1. According to our definitions, an endomorphism $f : 0 \to 0$ of a zero vector space is neither decomposable nor indecomposable. On the other hand, an endomorphism $f : V \to V$ of a non-zero vector space is either decomposable or indecomposable, and not both.

The following lemma gives a convenient criterion to check the indecomposability of linear maps:

{lemma:indec:suf}

Lemma 1.10.2. *Let V be a vector space, and let $f : V \to V$ be a linear map whose characteristic polynomial splits completely over the field $\Bbbk$. If f has exactly one eigenvalue λ and the corresponding eigenspace $E_\lambda(f)$ is one-dimensional, then f is an indecomposable endomorphism.*

Proof. Let us suppose that the linear map f has exactly one eigenvalue λ and that it is not indecomposable. Since $V \neq 0$, this means that there exist non-zero f-invariant subspaces W_1 and W_2 such that $V = W_1 \oplus W_2$, and we can consider the restrictions $f_{W_1} : W_1 \to W_1$ and $f_{W_2} : W_2 \to W_2$

of f to those two invariant subspaces. The characteristic polynomials of f_{W_1} and of f_{W_2} divide the characteristic polynomial of f, so they split completely over $\Bbbk$ because χ_f does: since they are non-constant because the subspaces W_1 and W_2 are non-zero, this tells us that they have roots in $\Bbbk$, and the hypothesis on f implies that, in fact, the scalar λ is a root of both. This means that λ is an eigenvalue of f_{W_1} and of f_{W_2}, so there exist non-zero vectors $w_1 \in W_1$ and $w_2 \in W_2$ such that $f(w_1) = \lambda w_1$ and $f(w_2) = \lambda w_2$. Since $V = W_1 \oplus W_2$, the vectors w_1 and w_2 are linearly independent. As they clearly belong to the eigenspace $E_\lambda(f)$, it follows from this that $\dim E_\lambda(f) \geq 2$. This proves the lemma. □

Using the lemma we can exhibit indecomposable endomorphism with ease. Let n be a positive integer, let λ be a scalar in $\Bbbk$, and let $f : \Bbbk^n \to \Bbbk^n$ be the linear map whose matrix with respect to the standard ordered basis $(e_1, e_2, \dots, e_n)$ of $\Bbbk^n$ is

$$J_n(\lambda) \coloneqq \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix},$$

so that for each $i \in \{1, \dots, n\}$ we have that

$$f(e_i) = \begin{cases} \lambda e_1 & \text{if } i = 1; \\ \lambda e_i + e_{i-1} & \text{if } 1 < i \leq n. \end{cases}$$

The matrix above is upper triangular, so it is immediate that the characteristic polynomial of f is $\chi_f(X) = (X - \lambda)^n$: this polynomial splits completely over $\Bbbk$ and has λ as its only root. Moreover, the linear map $f - \lambda \mathrm{id}_V$ clearly has rank $n - 1$, so its kernel has dimension 1, and this tells us that the eigenspace $E_\lambda(f)$ has dimension 1. We are thus in the hypotheses of the lemma, and we can conclude that the linear map f is indecomposable. We remark that the minimal polynomial of this linear map is also $(X - \lambda)^n$.

The main objective of this section is to show that, in fact, *all* indecomposable endomorphisms are of this form. To get to that result, we will need a few preliminary lemmas. We start with the easy observation that «shifting» a linear map preserves indecomposability.

{lemma:indec:shift}

Lemma 1.10.3. *If $f : V \to V$ is an indecomposable linear map and $\lambda \in \Bbbk$ is a scalar, then the linear map $f - \lambda \cdot \mathrm{id}_V : V \to V$ is also indecomposable.*

Proof. Let $f : V \to V$ be a linear map, let $\lambda \in \Bbbk$ be a scalar, and suppose that V is not zero and that the linear map $f - \lambda \cdot \mathrm{id}_V : V \to V$ is not indecomposable, so that there exist non-zero subspaces

W_1 and W_2 that are $(f - \lambda \cdot \mathrm{id}_V)$-invariant and such that $V = W_1 \oplus W_2$. If v is an element of W_1, then the $(f - \lambda \cdot \mathrm{id}_V)$-invariance of W_1 tells us that $f(v) - \lambda v \in W_1$, and that therefore $f(v)$ is an element of W_1: we see with this that W_1 is f-invariant. Of course, the same argument shows that W_2 is f-invariant and, since $V = W_1 \oplus W_2$ and both subspaces are non-zero, we see that the linear map f is not indecomposable. The lemma follows from this. □

The second lemma that we need is a well-known property of coprime polynomials, a special case of the so-called ***Bézout's identity***.

{lemma:bezout}

Lemma 1.10.4. *Let p and q be two polynomials in $\Bbbk[X]$. If p and q are coprime, then there exists polynomials u and v in $\Bbbk[X]$ such that $up + vq = 1$.*

Proof. Let us suppose that the polynomials p and q are coprime and consider the subset $I \coloneqq \{up + vq : u, v \in \Bbbk[X]\}$ of $\Bbbk[X]$, which is in fact a subspace. Since p and q are coprime, at least one of the two is non-zero: as it belongs to I, we see that there are non-zero elements in I. We may therefore pick a polynomial m in I that is monic and whose degree is $d \coloneqq \min\{\deg h : h \in I \setminus 0\}$. Since m belongs to I, there are polynomials $u_0, v_0 \in \Bbbk[X]$ such that $m = u_0 p + v_0 q$.

There are polynomials $a, r \in \Bbbk[X]$ such that $p = am + r$ and either $r = 0$ or $\deg r < d$. In fact, as

$$r = p - am = p - au_0 p - av_0 q = (1 - au_0)p + (-av_0)q \in I,$$

the choice of d implies that we must have that $r = 0$, for otherwise we would have in I a non-zero polynomial of degree strictly smaller than that of m: this means that m divides p. Of course, the same argument can be used to show that m also divides q, so that m is a common divisor of p and q: since p and q are coprime, we thus see that $m = 1$. This proves the lemma, since it tells us that $u_0 p + v_0 q = 1$. □

Third, we need an easy fact about invariant subspaces:

{ex:finv:ker}

Exercise 1.10.5. Let $f : V \to V$ be a linear map. Show that if $p \in \Bbbk[X]$ is a polynomial, then the subspace $\ker p(f)$ of V is f-invariant.

In order to describe indecomposable morphisms we start by showing that their minimal polynomials are of a very simple form.

{lemma:indec:min}

Lemma 1.10.6. *If $f : V \to V$ is an indecomposable endomorphism of a finite-dimensional vector space whose characteristic polynomials splits completely over $\Bbbk$, then the minimal polynomial of f is of the form $(X - \lambda)^\ell$ for some scalar λ in $\Bbbk$ and some positive integer ℓ.*

Proof. Let $f : V \to V$ be an indecomposable endomorphism of a finite-dimensional vector space, let us suppose that the characteristic polynomial of f splits completely over $\Bbbk$, and let $\mu \in \Bbbk[X]$ be its minimal polynomial. Since V is not the zero vector space, the degree of μ is positive and, as the characteristic polynomial of f splits completely over $\Bbbk$, there is a $\lambda \in \Bbbk$ such that $\mu(\lambda) = 0$. There are then $\ell \in \mathbb{N}$ and $p \in \Bbbk[X]$ such that $\mu(X) = (X - \lambda)^\ell p(X)$ and $p(\lambda) \neq 0$, and, according to Lemma 1.10.4, there are polynomials $r, s \in \Bbbk[X]$ such that $1 = s \cdot p + r \cdot (X - \lambda)^\ell$. Evaluating both sides of this equality at f we find that

$$\mathsf{id}_V = s(f) \circ p(f) + r(f) \circ (f - \lambda \cdot \mathsf{id}_V)^\ell. \tag{1.16}$$

{eq:fgcd}

If we consider the linear maps $\pi_1 \coloneqq s(f) \circ p(f) : V \to V$ and $\pi_2 \coloneqq r(f) \circ (f - \lambda \cdot \mathsf{id}_V)^\ell : V \to V$, we can rewrite the equality (1.16) in the form

$$\mathsf{id}_V = \pi_1 + \pi_2. \tag{1.17}$$

{eq:pi:1}

Let us consider now the subspaces $W_1 \coloneqq \ker(f - \lambda \cdot \mathsf{id}_V)^\ell$ and $W_2 \coloneqq \ker p(f)$ of V, and show that $V = W_1 \oplus W_2$:

- Let w be a vector in the intersection $W_1 \cap W_2$. Since $w \in W_1$, we have that $(f - \lambda \cdot \mathsf{id}_V)^\ell(w) = 0$ and therefore that $\pi_2(w) = 0$; since $w \in W_2$, we have that $p(f)(w) = 0$ and therefore that $\pi_1(w) = 0$. According to (1.17), then, $w = \mathsf{id}_V(w) = \pi_1(w) + \pi_2(w) = 0$. This tells us that the intersection $W_1 \cap W_2$ is zero and therefore that the subspaces W_1 and W_2 are independent.
- Let now w be an arbitrary vector in V. We have that

  $$(f - \lambda \cdot \mathsf{id}_V)^\ell(\pi_1(w)) = \big((f - \lambda \cdot \mathsf{id}_V)^\ell \circ s(f) \circ p(f)\big)(w) = \big(\mu(f) \circ s(f)\big)(w) = 0,$$

  so that $\pi_1(w) \in W_1$, and that

  $$p(f)(\pi_2(w)) = \big(p(f) \circ r(f) \circ (f - \lambda \cdot \mathsf{id}_V)^\ell\big)(w) = \big(\mu(f) \circ r(f)\big)(w) = 0,$$

  so that $\pi_2(w) \in W_2$. In view of (1.17), this implies that $w = \pi_1(w) + \pi_2(w) \in W_1 + W_2$, and we can conclude that $V = W_1 + W_2$.

According to Exercise 1.10.5, the subspaces W_1 and W_2 are f-invariant: since we have that $V = W_1 \oplus W_2$ and the map f is indecomposable, one of W_1 or W_2 has to be the zero subspace. As λ is an eigenvalue of f, there is a non-zero vector $w \in V$ such that $f(w) = \lambda w$, and clearly we have that $(f - \lambda \cdot \mathsf{id}_V)^\ell(w) = 0$, so that $w \in W_1$. It follows from this that $W_2 = \ker p(f) = 0$, and we see that the linear map $p(f) : V \to V$ is bijective. This and the fact that $0 = \mu(f) = p(f) \circ (f - \lambda \cdot \mathsf{id}_V)^\ell$ imply together that $(f - \lambda \cdot \mathsf{id}_V)^\ell = 0$, so that the minimal polynomial μ, which is of the form $(X - \lambda)^\ell p(X)$, divides $(X - \lambda)^\ell$. Of course, we can conclude from this that $p = 1$, so that $\mu = (X - \lambda)^\ell$, as the lemma claims. □

Using the information about the minimal polynomial that this lemma gives, we can now completely describe indecomposable endomorphisms:

Proposition 1.10.7. *Let $f : V \to V$ be an indecomposable endomorphism of a finite-dimensional vector space V such that the characteristic polynomial of f splits completely over $\Bbbk$. There exist an ordered basis $\mathscr{B}$ of V and a scalar $\lambda \in \Bbbk$ such that the matrix of f with respect to $\mathscr{B}$ is the matrix* {prop:block}

$$J_n(\lambda) := \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix},$$

with $n = \dim V$.

Proof. Let n be the dimension of V. According to Lemma 1.10.6, there is a scalar $\lambda \in \Bbbk$ and a positive integer ℓ such that the minimal polynomial of f is $\mu_f(X) = (X - \lambda)^\ell$.

Step 1. Let us consider the linear map $h := f - \lambda \cdot \mathsf{id}_V : V \to V$. We know from Lemma 1.10.3 that h is also an indecomposable endomorphism of V, and it is clear that its minimal polynomial is $\mu_h(X) = X^\ell$. This implies that h^ℓ is the zero map and that $h^{\ell-1}$ is not the zero map, so that there exists a vector $x \in V$ such that $h^{\ell-1}(x) \neq 0$. We claim that the ℓ vectors

$$x,\ h(x),\ h^2(x),\ \dots,\ h^{\ell-1}(x) \tag{1.18}$$ {eq:hix}

are linearly independent. Indeed, let us suppose that we have scalars $a_0, a_1, \dots, a_{\ell-1} \in \Bbbk$, not all zero, such that

$$a_0 x + a_1 h(x) + a_2 h^2(x) + \cdots + a_{\ell-1} h^{\ell-1}(x) = 0.$$

We can then consider the number $k := \min\{i \in \{0, \dots, \ell-1\} : a_i \neq 0\}$: of course, we have that $a_k h^k(x) + a_{k+1} h^{k+1}(x) + \cdots + a_{\ell-1} h^{\ell-1}(x) = 0$ and therefore that

$$0 = h^{\ell-k-1}\big(a_k h^k(x) + a_{k+1} h^{k+1}(x) + \cdots + a_{\ell-1} h^{\ell-1}(x)\big) = a_k h^{\ell-1}(x).$$

Since $h^{\ell-1}(x) \neq 0$ and $a_k \neq 0$, this is absurd. This proves our claim.

Step 2. Let U be the subspace of V generated by the ℓ vectors listed in (1.18), which has dimension ℓ. As we know, we can find $n - \ell$ vectors $v_1, v_2, \dots, v_{n-\ell}$ such that the sequence

$$(x,\ h(x),\ h^2(x),\ \dots,\ h^{\ell-1}(x), v_1,\ v_2,\ \dots,\ v_{n-\ell})$$

is an ordered basis for V. There is a unique linear map $\Phi : V \to \Bbbk$ such that

$$\Phi(h^i(x)) = \begin{cases} 0 & \text{if } 0 \le i < \ell - 1; \\ 1 & \text{if } i = \ell - 1; \end{cases} \qquad \text{for each } i \in \{0, \dots, \ell - 1\}$$

and

$$\Phi(v_i) = 0 \qquad\qquad \text{for each } i \in \{1, 2, \dots, n - \ell\}.$$

Using all this we can define a linear function $\pi : V \to V$ putting, for each $v \in V$,

$$\pi(v) \coloneqq \sum_{i=0}^{\ell-1} \Phi(h^i(v)) \cdot h^{\ell-1-i}(x).$$

Step 3. Let $W \coloneqq \ker \pi$ be the kernel of the linear map π. We want to show now that U and W are h-invariant subspaces of V.

- Since U is generated by the vectors x, $h(x)$, $h^2(x)$, …, $h^{\ell-1}(x)$ and the image under h of each of them is either one of them or zero, it is clear that U is h-invariant.
- Suppose that w is an element of W, so that

  $$0 = \pi(w) = \sum_{i=0}^{\ell-1} \Phi(h^i(w)) \cdot h^{\ell-1-i}(x).$$

  We can then compute that

  $$\begin{aligned}
  \pi(h(w)) &= \sum_{i=0}^{\ell-1} \Phi(h^{i+1}(w)) \cdot h^{\ell-1-i}(x) \\
  &= \sum_{j=1}^{\ell} \Phi(h^j(w)) \cdot h^{\ell-j}(x) && \text{changing the index of summation} \\
  &= \sum_{j=0}^{\ell-1} \Phi(h^j(w)) \cdot h^{\ell-j}(x) && \text{because } h^\ell(x) = 0 \text{ and } h^\ell(w) = 0 \\
  &= h\left(\sum_{j=0}^{\ell-1} \Phi(h^j(w)) \cdot h^{\ell-1-j}(x) \right) \\
  &= h(\pi(w)) = 0,
  \end{aligned}$$

  so that $h(w) \in W$. This shows that W is also an h-invariant subspace.

Step 4. We will show next that $V = U \oplus W$.

- We claim that $\pi(u) = u$ for all $u \in U$ and to check this, since U is generated by the ℓ vectors listed in (1.18), it is enough that we show that $\pi(h^k(x)) = h^k(x)$ for each $k \in \{0, \dots, \ell - 1\}$.

  Let the k be an element of $\{0, \dots, \ell - 1\}$. We have that

  $$\pi(h^k(x)) = \sum_{i=0}^{\ell-1} \underbrace{\Phi(h^{i+k}(x))}_{} \cdot h^{\ell-1-i}(x).$$

  If an element i of $\{0, \dots, \ell - 1\}$ is such that $i + k \geq \ell$, then $h^{i+k}(x) = 0$ since $h^\ell = 0$, and if instead $i + k < \ell - 1$ then $\Phi(h^{i+k}(x)) = 0$: it follows from this and the formula above that $\pi(h^k(x)) = \Phi(h^{\ell-1}(x)) \cdot h^k(x) = h^k(x)$, as we want.

- Let v be a vector in V. The value $\pi(v)$ is a linear combination of the ℓ vectors listed in (1.18), so $\pi(v)$ belongs to the subspace U, and what we have shown implies then that $\pi(\pi(v)) = \pi(v)$. We therefore have that $\pi(v - \pi(v)) = 0$, so that $v - \pi(v) \in W$, and thus that $v = \pi(v) + (v - \pi(v)) \in U + W$. We see with this that $V = U + W$.
- On the other hand, if v is an element of $U \cap W$, then on one hand we have that $\pi(v) = 0$, because $v \in W$, and on the other that $\pi(v) = v$, because $v \in U$: putting the two things together we see that $v = 0$. This shows that the subspaces U and W are independent, so that in fact $V = U \oplus W$.

Step 5. We have that $V = U \oplus W$ and that U and W are h-invariant: as $h : V \to V$ is an indecomposable endomorphism and $U \neq 0$ because $0 \neq x \in U$, we must have that $W = 0$. This tells us that $V = U$, so that the sequence of ℓ linearly independent vectors listed in (1.18) generate V. It follows from this, of course, that the sequence

$$\mathscr{B} = (h^{\ell-1}(x), h^{\ell-2}(x), \dots, h(x), x)$$

is an ordered basis of V, and the matrix of the linear map h with respect to this ordered basis is

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

This implies that the matrix of the linear map $f = h + \lambda$ is the matrix $J_n(\lambda)$ described in the statement of the proposition. This proves what we wanted. □

As a corollary, we can show that the condition for indecomposability given in Lemma 1.10.2 is also necessary:

Corollary 1.10.8. *An endomorphism $f : V \to V$ of a finite-dimensional vector space whose characteristic polynomial splits completely over $\Bbbk$ is indecomposable if and only if it has exactly one eigenvalue λ whose corresponding eigenspace $E_\lambda(f)$ is one-dimensional.*

Proof. Lemma 1.10.2 tells us that the condition is suficient. Suppose, on the other hand, that $f : V \to V$ is an indecomposable endomorphism of a finite-dimensional vector space whose characteristic polynomial splits completely over $\Bbbk$: according to Proposition 1.10.7 there are a scalar $\lambda \in \Bbbk$ and an ordered basis $\mathscr{B} = (v_1, v_2, \dots, v_n)$ of V such that the matrix $[f]_{\mathscr{B}}$ is the matrix $J_n(\lambda)$ with $n = \dim V$. The characteristic polynomial of f is then clearly $(X - \lambda)^n$, so that λ is the unique eigenvalue of f, and the corresponding eigenspace is $E_\lambda(f) = \langle v_1 \rangle$, which is one-dimensional. □

§1.11. The Jordan canonical form of an endomorphism

In the previous section we have carried out all the hard work needed to obtain the Jordan canonical form of an endomorphism. In this one we will put the pieces together.

{lemma:jcf:desc}

Lemma 1.11.1. *Let $f : V \to V$ be an endomorphism of a non-zero finite-dimensional vector space. There exist non-zero f-invariant subspaces $W_1, W_2, \dots, W_r$ of V such that*

- *$V = W_1 \oplus W_2 \oplus \cdots \oplus W_r$ and*
- *for each index $i \in \{1, 2, \dots, r\}$ the restriction $f_{W_i} : W_i \to W_i$ of f to W_i is an indecomposable endomorphism of W_i.*

Proof. Let us suppose that the lemma is not true, so that there exists an endomorphism $f : V \to V$ of a non-zero finite dimensional vector space V such that

there is no collection of non-zero f-invariant subspaces $W_1, W_2, \dots, W_r$ of V such that $V = W_1 \oplus W_2 \oplus \cdots \oplus W_r$ and each of the restrictions $f_{W_1}, f_{W_2}, \dots, f_{W_r}$ is indecomposable. (1.19) {eq:cihip}

We can moreover suppose that the endomorphism f is such that the number $\dim V$ is as small as possible.

The endomorphism $f : V \to V$ is not indecomposable: if it were, we could take $r = 1$, $W_1 = V$ and then have that $V = W_1$ and that the restriction f_{W_1}, which is simply f, is indecomposable, contradicting (1.19). As V is a non-zero space, we thus see that there exist two non-zero f-invariant subspaces P and Q of V such that $V = P \oplus Q$.

As $\dim V = \dim P + \dim Q$ and the subspaces P and Q are non-zero, we have $\dim P < \dim V$ and $\dim Q < \dim V$. In view of the way we chose the endomorphism f, this implies that the claim of the lemma is true for the restrictions f_P and f_Q, that is, there exist

- non-zero f_P-invariant subspaces $P_1, P_2, \dots, P_r$ of P such that $P = P_1 \oplus P_2 \oplus \cdots \oplus P_r$ and each of the restrictions $(f_P)_{P_1}, (f_P)_{P_2}, \dots, (f_P)_{P_r}$ is indecomposable, and
- non-zero f_Q-invariant subspaces $Q_1, Q_2, \dots, Q_s$ of Q such that $Q = Q_1 \oplus Q_2 \oplus \cdots \oplus Q_s$ and each of the restrictions $(f_Q)_{Q_1}, (f_Q)_{Q_2}, \dots, (f_Q)_{Q_s}$ is indecomposable.

As $V = P \oplus Q$, $P = P_1 \oplus P_2 \oplus \cdots \oplus P_r$ and $Q = Q_1 \oplus Q_2 \oplus \cdots \oplus Q_s$, we know from Lemma 1.9.15 that

$$V = P_1 \oplus P_2 \oplus \cdots \oplus P_r \oplus Q_1 \oplus Q_2 \oplus \cdots \oplus Q_s.$$

On the other hand, we know from Exercise 1.7.10 that the subspaces $P_1, P_2, \dots, P_r$ are f-invariant, because they are f_P-invariant, and that the restrictions $f_{P_1}, f_{P_2}, \dots, f_{P_r}$ are indecomposable, because they coincide with the restrictions $(f_P)_{P_1}, (f_P)_{P_2}, \dots, (f_P)_{P_r}$. Similarly, we know that the subspaces $Q_1, Q_2, \dots, Q_s$ is f-invariant, because they are f_Q-invariant, and that the restrictions $f_{Q_1}, f_{Q_2}, \dots, f_{Q_s}$ are indecomposable, because they coincide with the restrictions $(f_Q)_{Q_1}, (f_Q)_{Q_2}, \dots, (f_Q)_{Q_s}$. This contradicts (1.19), and this contradiction proves the lemma. □

This lemma almost immediately gives us the existence of the Jordan canonical form of an endomorphism. As in the previous section, for each positive integer n and each scalar $\lambda \in \Bbbk$ we consider the matrix

$$J_n(\lambda) \coloneqq \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix} \in \mathrm{M}_n(\Bbbk),$$

which we call a ***Jordan block*** of size n and eigenvalue λ.

{thm:jordan:f:exist}

Theorem 1.11.2. *Let $f : V \to V$ be an endomorphism of a non-zero finite-dimensional vector space whose characteristic polynomial splits completely over $\Bbbk$. There exist*

- *positive integers r and $n_1, n_2, \dots, n_r$ such that $n_1 + n_2 + \cdots + n_r = \dim V$,*
- *scalars $\lambda_1, \lambda_2, \dots, \lambda_r$, and*
- *an ordered basis $\mathscr{B}$ of V*

such that the matrix of f with respect to $\mathscr{B}$ is the block diagonal matrix

$$[f]_{\mathscr{B}} = \begin{pmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{n_r}(\lambda_r) \end{pmatrix},$$

and the scalars $\lambda_1, \lambda_2, \dots, \lambda_r$ are the eigenvalues of f, possibly listed with repetitions.

A matrix of this form — a block diagonal matrix whose diagonal blocks are Jordan blocks of various sizes and eigenvalues — is called a ***matrix in Jordan form***.

Proof. According to Lemma 1.11.1, there exists a positive integer r and non-zero f-invariant subspaces $W_1, W_2, \dots, W_r$ of V with $V = W_1 \oplus W_2 \oplus \cdots \oplus W_r$ and such that each of the restrictions $f_{W_1} : W_1 \to W_1, \dots, f_{W_r} : W_r \to W_r$ is an indecomposable endomorphism. The characteristic polynomials of these r restrictions divide the characteristic polynomial of f, so they split completely over the field $\Bbbk$, and then, according to Proposition 1.10.7, if we put $n_i \coloneqq \dim W_i$ for each $i \in \{1, 2, \dots, r\}$, there exist scalars $\lambda_1, \dots, \lambda_r \in \Bbbk$ and ordered bases $\mathscr{B}_1 = (v_{1,1}, v_{1,2}, \dots, v_{1,n_1})$, $\mathscr{B}_2 = (v_{2,1}, v_{2,2}, \dots, v_{2,n_2}), \dots, \mathscr{B}_r = (v_{r,1}, v_{r,2}, \dots, v_{r,n_r})$ of the subspaces $W_1, W_2, \dots, W_r$, respectively, such that $[f_{W_i}]_{\mathscr{B}_i} = J_{n_i}(\lambda_i)$ for each $i \in \{1, 2, \dots, r\}$. Since $V = W_1 \oplus W_2 \oplus \dots \oplus W_r$, we know that the sequence

$$\mathscr{B} = (v_{1,1}, v_{1,2}, \dots, v_{1,n_1}, v_{2,1}, v_{2,2}, \dots, v_{2,n_2}, \dots, \dots, v_{r,1}, v_{r,2}, \dots, v_{r,n_r})$$

is an ordered basis for V, and it is easy to see that the matrix of f with respect to it is

$$[f]_{\mathscr{B}} = \begin{pmatrix} [f_{W_1}]_{\mathscr{B}_1} & & & \\ & [f_{W_2}]_{\mathscr{B}_2} & & \\ & & \ddots & \\ & & & [f_{W_r}]_{\mathscr{B}_r} \end{pmatrix} = \begin{pmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{n_r}(\lambda_r) \end{pmatrix}.$$

This proves the first claim of theorem, and the second one follows immediately from it. □

The existence theorem we have proved has an accompanying uniqueness result. The key to obtaining it is a very simple observation about the ranks of powers of Jordan blocks. In general, for each matrix A in $\mathrm{M}_n(\Bbbk)$, each $k \in \mathbb{N}_0$ and each scalar $\lambda \in \Bbbk$ we set

$$\begin{aligned} \rho_k(A, \lambda) &\coloneqq \operatorname{rank}(A - \lambda \cdot I_n)^k, \\ \Delta_k(A, \lambda) &\coloneqq \rho_{k-1}(A, \mu) - 2\rho_k(A, \mu) + \rho_{k+1}(A, \mu). \end{aligned}$$

Similarly, if $f : V \to V$ is an endomorphism of a finite-dimensional vector space, then for each $\lambda \in \Bbbk$ and each $k \in \mathbb{N}_0$ we put

$$\begin{aligned} \rho_k(f, \lambda) &\coloneqq \operatorname{rank}(f - \lambda \cdot \mathsf{id}_V)^k, \\ \Delta_k(f, \lambda) &\coloneqq \rho_{k-1}(f, \mu) - 2\rho_k(f, \mu) + \rho_{k+1}(f, \mu). \end{aligned}$$

As usual, these two definitions — one for matrices and one for endomorphisms — are closely related:

{ex:ranks:sim}

Exercise 1.11.3. Let V be finite-dimensional vector space, and let $\mathscr{B}$ be an ordered basis for V. Prove that if $f : V \to V$ is an endomorphism of V, then for all $k \in \mathbb{N}_0$ and all $\lambda \in \Bbbk$ we have that

$$\rho_k(f, \lambda) = \rho_k([f]_{\mathscr{B}}, \lambda), \qquad \Delta_k(f, \lambda) = \Delta_k([f]_{\mathscr{B}}, \lambda),$$

These numbers are additive in the following sense:

{ex:ranks}

Exercise 1.11.4. Let n and $n_1, n_2, \dots, n_r$ be positive integers such that $n = n_1 + n_2 + \dots + n_r$. Show that if A is a matrix in $\mathrm{M}_n(\Bbbk)$ that is a block diagonal matrix of the form

$$A = \begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_r \end{pmatrix},$$

with $A_1 \in \mathrm{M}_{n_1}(\Bbbk)$, $A_2 \in \mathrm{M}_{n_2}(\Bbbk)$, …, $A_r \in \mathrm{M}_{n_r}(\Bbbk)$, then

$$\rho_k(A,\lambda) = \rho_k(A_1,\lambda) + \rho_k(A_2,\lambda) + \cdots + \rho_k(A_r,\lambda)$$

and

$$\Delta_k(A,\lambda) = \Delta_k(A_1,\lambda) + \Delta_k(A_2,\lambda) + \cdots + \Delta_k(A_r,\lambda),$$

for all $\lambda \in \Bbbk$ and all $k \in \mathbb{N}_0$.

The third important observation that we need to make about these ranks is that they are invariant under similarly:

Exercise 1.11.5. {exer:delta-sim}

(1) Let V be a finite-dimensional vector space, and let $f, g : V \to V$ be an endomorphism of V. If f and g are similar, so that there exists a bijective endomorphism $h : V \to V$ such that $f = h \circ g \circ h^{-1}$, then

$$\rho_k(f,\lambda) = \rho_k(g,\lambda), \qquad \Delta_k(f,\lambda) = \Delta_k(g,\lambda)$$

for all $k \in \mathbb{N}_0$ and all $\lambda \in \Bbbk$.

(2) Let n be a positive integer. If A and B are two matrices in $\mathrm{M}_n(\Bbbk)$ that are similar, so that there exists an invertible matrix C in $\mathrm{M}_n(\Bbbk)$ such that $A = CBC^{-1}$, then

$$\rho_k(A,\lambda) = \rho_k(B,\lambda), \qquad \Delta_k(A,\lambda) = \Delta_k(B,\lambda)$$

for all $k \in \mathbb{N}_0$ and all $\lambda \in \Bbbk$.

It is easy to compute these numbers when the matrix A is a Jordan block. We start with the ranks:

Lemma 1.11.6. *Let n be a positive integer, let $\lambda \in \Bbbk$ be a scalar, and let A be the matrix $J_n(\lambda)$.* {lemma:ranks:block

(*i*) *For each $k \in \mathbb{N}_0$ and each $\mu \in \Bbbk$ we have that*

$$\rho_k(A,\mu) = \begin{cases} n & \text{if } \mu \neq \lambda; \\ n-k & \text{if } \mu = \lambda \text{ and } 0 \leq k \leq n; \\ 0 & \text{if } \mu = \lambda \text{ and } k > n. \end{cases}$$

(*ii*) *For each $k \in \mathbb{N}$ and each $\mu \in \Bbbk$ we have that*

$$\Delta_k(A,\mu) = \begin{cases} 1 & \textit{if } \mu = \lambda \textit{ and } k = n; \\ 0 & \textit{in any other case.} \end{cases}$$

Proof. (*i*) Let $\mu \in \Bbbk$ be a scalar. The matrix $A - \mu \cdot I_n$ is

$$\begin{pmatrix} \lambda-\mu & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda-\mu & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda-\mu & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda-\mu & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda-\mu \end{pmatrix}$$

If $\mu \neq \lambda$, then this matrix and all its powers are invertible, so that their rank is n. If instead $\mu = \lambda$, then the matrix is

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix},$$

and a direct calculation shows that the ranks of its powers are as described in the statement lemma.

(*ii*) We can compute the expression that appears in the statement in the various cases.

- If $\mu \neq \lambda$, then $\rho_k(A,\mu) = n$ for all $k \in \mathbb{N}_0$, so that clearly

  $$\Delta_k(A,\mu) = \rho_{k-1}(A,\mu) - 2\rho_k(A,\mu) + \rho_{k+1}(A,\mu) = 0$$

  for all $k \in \mathbb{N}$.
- If $0 \leq k < n$, then $\rho_{k-1}(A,\lambda) = n-(k-1)$, $\rho_k(A,\lambda) = n-k$ and $\rho_{k+1}(A,\lambda) = n-(k+1)$, so

  $$\begin{aligned} \Delta_k(A,\mu) &= \rho_{k-1}(A,\lambda) - 2\rho_k(A,\lambda) + \rho_{k+1}(A,\lambda) \\ &= (n-(k-1)) - 2(n-k) + (n-(k+1)) = 0. \end{aligned}$$

- We have $\rho_{n-1}(A,\lambda) = 1$, $\rho_n(A,\lambda) = 0$ and $\rho_{n+1}(A,\lambda) = 0$, so

  $$\Delta_n(A,\mu) = \rho_{n-1}(A,\lambda) - 2\rho_n(A,\lambda) + \rho_{n+1}(A,\lambda) = 1 - 2\cdot 0 + 0 = 1.$$

- Finally, if $k > n$, then $\rho_{k-1}(A,\lambda) = 0$, $\rho_k(A,\lambda) = 0$ and $\rho_{k+1}(A,\lambda) = 0$, so

  $$\Delta_k(A,\mu) = \rho_{k-1}(A,\lambda) - 2\rho_k(A,\lambda) + \rho_{k+1}(A,\lambda) = 0.$$

These observations taken together prove the lemma. □

This lemma, together with the results of the exercises that precede it, leads us very easily to the uniqueness of the Jordan form of an endomorphism.

Theorem 1.11.7. *Let $f : V \to V$ be an endomorphism of a non-zero finite-dimensional vector space and let us suppose there are positive integers r and $n_1, n_2, \dots, n_r$ such that $n_1 + n_2 + \cdots + n_r = \dim V$, scalars $\lambda_1, \lambda_2, \dots, \lambda_r$, and an ordered basis $\mathscr{B}$ of V such that the matrix of f with respect to $\mathscr{B}$ is the block diagonal matrix* {thm:jordan:f:uniqu

$$[f]_{\mathscr{B}} = \begin{pmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{n_r}(\lambda_r) \end{pmatrix}. \tag{1.20}$$ {eq:fbjx}

The number of Jordan blocks in this matrix with eigenvalue μ and size m is $\Delta_k(f,\mu)$.

Proof. Let A be the matrix that appears in (1.20). Using the result of Exercises 1.11.3 and 1.11.4 we see that for each $\mu \in \Bbbk$ and each $k \in \Bbbk$ we have

$$\Delta_k(f,\mu) = \Delta_k(A,\mu) = \Delta_k(J_{n_1}(\lambda_1),\mu) + \Delta_k(J_{n_2}(\lambda_2),\mu) + \cdots + \Delta_k(J_{n_r}(\lambda_r),\mu).$$

It follows from Lemma 1.11.6 that for each $i \in \{1,2,\dots,r\}$ the ith term appearing in this sum is equal to 1 if $\lambda_i = \mu$ and $n_i = k$, and to 0 in any other case. This tells us that $\Delta_k(f,\mu)$ is precisely the number of Jordan blocks of size k and eigenvalue μ that appear in the matrix (1.20). □

We view this theorem as a uniqueness result for the Jordan form of an endomorphism. Indeed, let us suppose that $f : V \to V$ is an endomorphism of a finite-dimensional vector space such that

- there are positive integers r and $n_1, n_2, \dots, n_r$ such that $n_1 + n_2 + \cdots + n_r = \dim V$, and scalars $\lambda_1, \lambda_2, \dots, \lambda_r$ and an ordered basis $\mathscr{B}$ of V such that the matrix of f with respect to $\mathscr{B}$ is the block diagonal matrix

$$[f]_{\mathscr{B}} = \begin{pmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{n_r}(\lambda_r) \end{pmatrix}, \tag{1.21}$$ {eq:fbj:xb1}

- there are positive integers s and $m_1, m_2, \dots, m_s$ such that $m_1 + m_2 + \cdots + m_s = \dim V$, and scalars $\mu_1, \mu_2, \dots, \mu_r$ and an ordered basis $\mathscr{B}'$ of V such that the matrix of f with respect to $\mathscr{B}'$ is the block diagonal matrix of the form

$$[f]_{\mathscr{B}'} = \begin{pmatrix} J_{n_1}(\mu_1) & & & \\ & J_{n_2}(\mu_2) & & \\ & & \ddots & \\ & & & J_{n_s}(\mu_s) \end{pmatrix}. \tag{1.22}$$ {eq:fbj:xb2}

It then follows from the theorem that for all $k \in \mathbb{N}$ and all $\mu \in \mathbb{k}$ the matrices (1.21) and (1.22) have the same number of Jordan blocks of size k and eigenvalue μ, since that number is $\Delta_k(f,\mu)$. Clearly this implies that the total number of Jordan blocks in the two matrices coincide, so that $r = s$, and that the two lists of pairs

$$(n_1,\lambda_1),(n_2,\lambda_2),\dots,(n_r,\lambda_r)$$

and

$$(m_1,\mu_1),(m_2,\mu_2),\dots,(m_s,\mu_s)$$

have the same elements — taking into account repetitions — but possibly in different order. We therefore say that the two matrices (1.21) and (1.22) coincide *up to permutation of the blocks.*

We have been working throughout with endomorphisms, but as usual all our results have analogues about matrices. The following proposition states the end result:

{thm:jordan:mats}

Theorem 1.11.8. *Let n be a positive integer and let A be an element in $\mathrm{M}_n(\mathbb{k})$ whose characteristic polynomial χ_A splits completely over $\mathbb{k}$. There exist*

- *an invertible matrix C in $\mathrm{M}_n(\mathbb{k})$,*
- *positive integers r and n_1, n_2, …, n_r such that $n_1 + n_2 + \cdots + n_r = n$, and*
- *scalars λ_1, λ_2, …, λ_r in $\mathbb{k}$*

such that

$$CAC^{-1} = \begin{pmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{n_r}(\lambda_r) \end{pmatrix}.$$

For each positive integer k and each scalar $\mu \in \mathbb{k}$ the number of Jordan blocks appearing in this matrix of size k and eigenvalue μ is exactly $\Delta_k(A,\mu)$.

Proof. The characteristic polynomial of the linear map $f : x \in \mathbb{k}^n \mapsto Ax \in \mathbb{k}^n$ coincides with χ_A, so it splits completely over $\mathbb{k}$. We can therefore apply Theorems 1.11.2 and 1.11.7 to f. According to the first of these two theorems there exists an ordered basis $\mathscr{B} = (v_1, v_2, \dots, v_n)$ of $\mathbb{k}^n$, positive integers r and n_1, n_2, …, n_r such that $n_1 + n_2 + \cdots + n_r = n$, and scalars λ_1, λ_2, …, $\lambda_r \in \mathbb{k}$ such that the matrix of f with respect to $\mathscr{B}$ is

$$[f]_{\mathscr{B}} = \begin{pmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{n_r}(\lambda_r) \end{pmatrix}. \tag{1.23}$$

{eq:j1}

Let $\mathscr{S}$ be the standard ordered basis of $\Bbbk^n$. We know that the matrix of f with respect to $\mathscr{S}$ is

$$[f]_{\mathscr{S}} = A. \tag{1.24}$$

{eq:j2}

On the other hand, if $C \in \mathrm{M}_n(\Bbbk)$ is the change-of-basis matrix from $\mathscr{S}$ to $\mathscr{B}$, which is of course invertible, then

$$C \cdot [f]_{\mathscr{S}} \cdot C^{-1} = [f]_{\mathscr{B}}. \tag{1.25}$$

{eq:j3}

Putting together the three equalities (1.23), (1.24) and (1.25) we see that the first claim of the theorem holds. The second claim, in turn, follows immediately from Theorem 1.11.7 and the fact that $\Delta_k(f,\mu) = \Delta_k(A,\mu)$ for all $k \in \mathbb{N}_0$ and all $\mu \in \Bbbk$. □

Theorems 1.11.2, 1.11.7 and 1.11.8 give enough information to actually compute in practice the Jordan canonical forms of endomorphisms and matrices. This often requires a lot of calculation, though.

Example 1.11.9. Let us consider the following element of $\mathrm{M}_{12}(\mathbb{Q})$,

$$A = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

This is a strictly upper triangular matrix, so its characteristic polynomial is easy to compute: we have $\chi_A(X) = X^{12}$, which is a polynomial that splits completely over $\mathbb{Q}$, and therefore its only eigenvalue is 0. A calculation shows that the ranks of the powers of A are

k	0	1	2	3	4	5	6	$\cdots$
$\rho_k(A,0)$	12	7	4	1	0	0	0	$\cdots$

and therefore we can compute the following table:

k	1	2	3	4	5	$\cdots$
$\Delta_k(A,0)$	2	0	2	1	0	$\cdots$

There are therefore five Jordan blocks in the Jordan canonical form of A, all with eigenvalue 0, and their sizes are 1, 1, 3, 3, and 4. The Jordan canonical form of A is thus the matrix

$$\begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}$$

Example 1.11.10. Let us now consider the matrix

$$A = \begin{pmatrix}
-7 & -4 & -5 & 1 & -7 & 3 & 4 & 4 \\
7 & 5 & 3 & 3 & 3 & 2 & -4 & -4 \\
-1 & -1 & 2 & -1 & 1 & -1 & 2 & 1 \\
7 & 3 & 3 & 4 & 4 & 1 & -3 & -4 \\
4 & 2 & 2 & -1 & 6 & -3 & -2 & -2 \\
-5 & -3 & -3 & 0 & -2 & 1 & 3 & 2 \\
-2 & -2 & -2 & 3 & -2 & 3 & 4 & 0 \\
0 & 1 & 0 & 0 & -1 & 1 & -1 & 2
\end{pmatrix}.$$

Using a computer we find that its characteristic polynomial is

$$\begin{aligned}
\chi_A(X) &= x^8 - 17x^7 + 118x^6 - 424x^5 + 800x^4 - 592x^3 - 416x^2 + 1024x - 512 \\
&= (x+1)(x-4)^2(x-2)^5,
\end{aligned}$$

so that its eigenvalues are −1, 4 and 2. We can also compute the following table of ranks

k	0	1	2	3	4	5	6	7	8	$\cdots$
$\rho_k(A,-1)$	8	7	7	7	7	7	7	7	7	$\cdots$
$\rho_k(A,4)$	8	7	6	6	6	6	6	6	6	$\cdots$
$\rho_k(A,2)$	8	6	4	3	3	3	3	3	3	$\cdots$

and therefore we have the following table

k	1	2	3	4	5	$\cdots$
$\Delta_k(A,-1)$	1	0	0	0	0	$\cdots$
$\Delta_k(A,4)$	0	1	0	0	0	$\cdots$
$\Delta_k(A,2)$	0	1	1	0	0	$\cdots$

The Jordan canonical form of the matrix A therefore has four blocks in total: 1 of size 1 with eigenvalue −1, 1 of size 2 with eigenvalue 4, and 2 of sizes 2 and 3 with eigenvalue 2. That Jordan canonical form is thus

$$\begin{pmatrix}
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 4 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 2
\end{pmatrix}.$$

Essentially all the information about an endomorphism or a matrix can be read off its Jordan canonical form. The following proposition describes how to do this for some of the invariants we have studied in these notes.

{prop:jordan:apps}

Proposition 1.11.11. *Let $f : V \to V$ be an endomorphism of a non-zero finite-dimensional vector space whose characteristic polynomial splits completely over $\Bbbk$ and let λ be an eigenvalue of f.*

(*i*) *The sum of the sizes of the Jordan blocks with eigenvalue λ that appear in the Jordan canonical form of f is equal to the multiplicity of λ as a root of the characteristic polynomial of f.*

(*ii*) *The number of Jordan blocks with eigenvalue λ that appear in the Jordan canonical form of f is equal to the dimension of the eigenspace $E_\lambda(f)$.*

(*iii*) *The size of the largest Jordan block with eigenvalue λ that appears in the Jordan canonical form of f is equal to the multiplicity of λ as a root of the minimal polynomial of f.*

Proof. □

It follows immediately from the third part of this proposition that if the minimal polynomial of the linear map f is without multiplicities then f is diagonalizable, as we already know from Proposition 1.8.13: indeed, the proposition tells us that in that case all the Jordan blocks that appear in the Jordan canonical form of f are of size 1, so that this matrix is in fact diagonal.

The information given by Proposition 1.11.11 can often be used to simplify the process of finding the Jordan canonical form of endomorphisms and matrices. Let us give some examples of this.

Example 1.11.12. Let $f : V \to V$ be an endomorphism of a complex vector space of dimension 5 whose characteristic polynomial is $(X-5)^3(X+7)^2$ and whose minimal polynomial is $(X-5)^2(X+7)^2$, and let J be the Jordan canonical form of f.

- The sum of the sizes of the Jordan blocks of J with eigenvalue 5 is 3, and the maximum size of those blocks is 2: it follows that those blocks are 2, of sizes 1 and 2.
- On the other hand, the sum of the sizes of the blocks of eigenvalue −7 is 2, and the maximum size of those blocks is also 2: this tells us that there is exactly one block of eigenvalue −7, of size 2.

We can therefore conclude that the Jordan canonical form of the map f is the matrix

$$\left(\begin{array}{c|cc|ccc}
5 & 0 & 0 & 0 & 0 & 0 \\
\hline
0 & 5 & 1 & 0 & 0 & 0 \\
0 & 0 & 5 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & -7 & 1 & 0 \\
0 & 0 & 0 & 0 & -7 & 1 \\
0 & 0 & 0 & 0 & 0 & -7
\end{array}\right).$$

In this example we were able to determined the Jordan form of the map f using only information about its characteristic and minimal polynomials. That is not possible in general, as our next example shows.

Example 1.11.13. If $f : \mathbb{Q}^6 \to \mathbb{Q}^6$ is a $\mathbb{Q}$-linear map with characteristic and minimal polynomials X^6 and X^3, respectively, both of which split completely over $\mathbb{Q}$, then the only eigenvalue of f is 0, and the maximal size of a Jordan block of f with that eigenvalue is 3. This information is not enough to find the Jordan canonical form of f: indeed, there are three possible canonical forms compatible with this information, namely the matrices

$$\left(\begin{array}{c|c|c|ccc} 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right), \qquad \left(\begin{array}{c|cc|ccc} 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

and

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Notice that if we knew the dimension of the kernel of f, which is the eigenspace $E_0(f)$, then we could determine the canonical form of f, for this number coincides with the number of Jordan blocks in that canonical form.

Example 1.11.14. Let n be a positive integer and let $f : \mathbb{C}^n \to \mathbb{C}^n$ be a linear map whose only eigenvalue is 0. This tells us that the only root of the characteristic polynomial χ_f is 0, and since the field $\mathbb{C}$ is algebraically closed this implies that in fact $\chi_f(X) = X^n$. We let r be the number of Jordan blocks in the canonical form of f, and write $n_1, n_2, \ldots, n_r$ for their sizes. Clearly, we can suppose that the indices the blocks are chosen so that $n_1 \geq n_2 \geq \cdots \geq n_r$. Since 0 is the only eigenvalue of f, the sequence $(n_1, n_2, \ldots, n_r)$ completely determines the Jordan canonical form of f. That sequence is called the ***Jordan type*** of f.

A decreasing sequence $(n_1, \ldots, n_r)$ of positive integers with $n = n_1 + n_2 + \cdots + n_r$ is called a ***partition*** of the number n, and the integers $n_1, n_2, \ldots, n_r$ are called the ***parts*** of the partition. Table 1.1 on page 76 lists the partitions of the first few positive integers. The sequences that appear in it thus describe the possible shapes of the Jordan canonical form of the map f. From this table we see immediately, for example, that

- if $n = 7$, $\mu_f = X^3$, and the kernel of f has dimension 4, then the Jordan canonical form of f has four blocks of sizes 3, 2, 1 and 1.
- if $n = 8$, $\mu_f = X^4$ and the kernel of f^2 has dimension 5, then the Jordan canonical form of f has three blocks of sizes 4, 3 and 1.

n	partitions of n
1	(1)
2	(2), $(1,1)$
3	(3), $(2,1)$, $(1,1,1)$
4	(4), $(3,1)$, $(2,2)$, $(2,1,1)$, $(1,1,1,1)$
5	(5), $(4,1)$, $(3,2)$, $(3,1,1)$, $(2,2,1)$, $(2,1,1,1)$, $(1,1,1,1,1)$
6	(6), $(5,1)$, $(4,2)$, $(4,1,1)$, $(3,3)$, $(3,2,1)$, $(3,1,1,1)$, $(2,2,2)$, $(2,2,1,1)$, $(2,1,1,1,1)$, $(1,1,1,1,1,1)$
7	(7), $(6,1)$, $(5,2)$, $(5,1,1)$, $(4,3)$, $(4,2,1)$, $(4,1,1,1)$, $(3,3,1)$, $(3,2,2)$, $(3,2,1,1)$, $(3,1,1,1,1)$, $(2,2,2,1)$, $(2,2,1,1,1)$, $(2,1,1,1,1,1)$, $(1,1,1,1,1,1,1)$
8	(8), $(7,1)$, $(6,2)$, $(6,1,1)$, $(5,3)$, $(5,2,1)$, $(5,1,1,1)$, $(4,4)$, $(4,3,1)$, $(4,2,2)$, $(4,2,1,1)$, $(4,1,1,1,1)$, $(3,3,2)$, $(3,3,1,1)$, $(3,2,2,1)$, $(3,2,1,1,1)$, $(3,1,1,1,1,1)$, $(2,2,2,2)$, $(2,2,2,1,1)$, $(2,2,1,1,1,1)$, $(2,1,1,1,1,1,1)$, $(1,1,1,1,1,1,1,1)$

Table 1.1. The partitions of the first few integers.

{table:partitions}

The number of partitions of an integer n is usually written $p(n)$ and called the ***partition number***. There is an immense amount of work devoted to understanding these partition numbers. We can tabulate the first few:

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$p(n)$	1	2	3	5	7	11	15	22	30	42	56	77	101	135	176	231

These numbers grow very fast: for example,

$$\begin{aligned} p(100) &= 190\,569\,292 \sim 2\cdot 10^{8}, \\ p(1\,000) &= 24\,061\,467\,864\,032\,622\,473\,692\,149\,727\,991 \sim 2\cdot 10^{31}, \\ p(10\,000) &= 36\,167\,251\,325\,636\,293\,988\,820\,471\,890\,953\,695\,495\,016\,030\,339\,315\,650\,422 \\ &\qquad 081\,868\,605\,887\,952\,568\,754\,066\,420\,592\,310\,556\,052\,906\,916\,435\,144 \sim 3\cdot 10^{106}, \end{aligned}$$

and it can be proved, in fact, that

$$p(n) \sim \frac{1}{4n\sqrt{3}} \exp\left(\pi\sqrt{\frac{2n}{3}}\right) \qquad \text{as } n \text{ tends to infinity.}$$

In this way we find information about the possible Jordan forms of matrices of size n whose characteristic polynomial is X^n. We can also inquire about those matrices of size n whose characteristic polynomial is X^n and whose minimal polynomial is X^m for some $m \in \{1,\dots,n\}$. The

$p(n,m)$		m														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
n	1	1														
	2	1	2													
	3	1	2	3												
	4	1	3	4	5											
	5	1	3	5	6	7										
	6	1	4	7	9	10	11									
	7	1	4	8	11	13	14	15								
	8	1	5	10	15	18	20	21	22							
	9	1	5	12	18	23	26	28	29	30						
	10	1	6	14	23	30	35	38	40	41	42					
	11	1	6	16	27	37	44	49	52	54	55	56				
	12	1	7	19	34	47	58	65	70	73	75	76	77			
	13	1	7	21	39	57	71	82	89	94	97	99	100	101		
	14	1	8	24	47	70	90	105	116	123	128	131	133	134	135	

Table 1.2. The number $p(n,m)$ of partitions of the number n with parts not larger than m. {table:pnm}

Jordan form of such a matrix only has Jordan blocks of eigenvalue 0 and, as we have seen above, the maximum size of those blocks is m. The type of such a matrix is thus a partition of n of the form $(n_1, n_2, \dots, n_r)$ in which $n_r = m$, and clearly then $(n_2, \dots, n_r)$ is a partition of $n - m$ whose parts are not larger than m. It follows from these observations if n and m are two positive integers such that $1 \leq m \leq n$, then

there are as many matrices in Jordan form with characteristic polynomial X^n and minimal polynomial X^m as there are partitions of $n - m$ with parts not larger than m. (1.26) {eq:resp}

We usually write $p(n,k)$ for the number of partitions of n with parts not larger than m, so the number of matrices in Jordan form described in (1.26) is $p(n, n-m)$. Table 1.2 on page 1.2 gives these numbers for small values of n and m.

This and a lot more of information about this beautiful subject can be found in George Andrews's classical book [And98].

§1.12. Similarity

Let us recall that two endomorphisms $f, g : V \to V$ of a vector space V are ***similar*** if there exists a bijective endomorphism $h : V \to V$ such that $f = h \circ g \circ h^{-1}$ and that in that case we write $f \sim f$. Analogously, two matrices A and B in $\mathrm{M}_n(\Bbbk)$ are ***similar*** if there exists an invertible matrix C in $\mathrm{M}_n(\Bbbk)$ such that $A = CBC^{-1}$, and in that case we write $A \sim B$.

Exercise 1.12.1.

(1) Let V be a vector space. Prove that similarity is an equivalence relation on the set $\mathrm{End}(V)$ of all endomorphisms of V.

(2) Let n be a positive number. Prove that similarity is an equivalence relation on the set $\mathrm{M}_n(\Bbbk)$ of all matrices of size n with entries in $\Bbbk$.

(3) Let n be a positive number, and let A and B be two matrices in $\mathrm{M}_n(\Bbbk)$. Prove that the matrices A and B are similar if and only if the associated linear maps $f_A : x \in \Bbbk^n \mapsto Ax \in \Bbbk^n$ and $f_B : x \in \Bbbk^n \mapsto Bx \in \Bbbk^n$ are similar.

The following proposition gives the basic characterization of similarity for endomorphisms.

Proposition 1.12.2. *Let V be a finite-dimensional vector space, and let $f, g : V \to V$ be two endomorphisms of V. The following statements are equivalent:*

(a) The endomorphisms f and g are similar.

(b) For every ordered basis $\mathscr{B}$ of V the matrices $[f]_{\mathscr{B}}$ and $[g]_{\mathscr{B}}$ are similar.

(c) For every ordered basis $\mathscr{B}$ of V there exists another ordered basis $\mathscr{B}'$ of V such that $[f]_{\mathscr{B}} = [g]_{\mathscr{B}'}$.

Proof. Let n be the dimension of V.

$(a) \Rightarrow (b)$ Let us suppose that the endomorphisms f and g are similar, so that there exists a bijective endomorphism $h : V \to V$ such that $f = h \circ g \circ h^{-1}$, and let $\mathscr{B}$ be an ordered basis for V. The matrix $C \coloneqq [h]_{\mathscr{B}}$ is then invertible, its inverse is $[h^{-1}]_{\mathscr{B}}$, and we have that

$$[f]_{\mathscr{B}} = [h \circ g \circ h^{-1}]_{\mathscr{B}} = [h]_{\mathscr{B}} \cdot [g]_{\mathscr{B}} \cdot [h^{-1}]_{\mathscr{B}} = C \cdot [g]_{\mathscr{B}} \cdot C^{-1},$$

so the matrices $[f]_{\mathscr{B}}$ and $[g]_{\mathscr{B}}$ are similar.

$(b) \Rightarrow (c)$ Let us suppose that the statement (b) holds, and let $\mathscr{B} = (v_1, v_2, \dots, v_n)$ be any ordered basis for V. According to the hypothesis, there exists an invertible matrix $C = (c_{i,j})$ in $\mathrm{M}_n(\Bbbk)$ such that $[f]_{\mathscr{B}} = C \cdot [g]_{\mathscr{B}} \cdot C^{-1}$. If for every $i \in [\![n]\!]$ we put $w_i \coloneqq c_{1,i}v_1 + c_{2,i}v_2 + \cdots + c_{n,i}v_n$, then the sequence $\mathscr{B}' \coloneqq (w_1, w_2, \dots, w_n)$ is an ordered basis for V because the matrix C is invertible, and C and C^{-1} are, in fact, the change of basis matrices $C(\mathscr{B}', \mathscr{B})$ and $C(\mathscr{B}, \mathscr{B}')$. It follows from this that

$$\begin{aligned}[g]_{\mathscr{B}'} &= C(\mathscr{B},\mathscr{B}')\cdot[g]_{\mathscr{B}}\cdot C(\mathscr{B}',\mathscr{B})\\ &= C(\mathscr{B},\mathscr{B}')\cdot C\cdot[f]_{\mathscr{B}}\cdot C^{-1}\cdot C(\mathscr{B}',\mathscr{B})\\ &= [f]_{\mathscr{B}}.\end{aligned}$$

We thus see that the statement (c) also holds.

$(c) \Rightarrow (a)$ Let us finally suppose that the statement (c) holds, let us consider any ordered basis $\mathscr{B} = (v_1, v_2, \dots, v_n)$ for V, and let $\mathscr{B}' = (w_1, w_2, \dots, w_n)$ be an ordered basis for V such that $[f]_{\mathscr{B}} = [g]_{\mathscr{B}'}$, whose existence is guaranteed by the hypothesis. Since $\mathscr{B}'$ is a basis for V, there exists exactly one linear map $h : V \to V$ such that $h(w_i) = v_i$ for all $i \in [\![n]\!]$, and since the image of h clairly contains $\mathscr{B}$ we see that the map h is surjective and this bijective. If the matrix $[f]_{\mathscr{B}}$ is the matrix $(a_{i,j}) \in \mathrm{M}_n(\Bbbk)$, then we have that $f(v_i) = \sum_{j=1}^{n} a_{j,i}v_i$ and $g(w_i) = \sum_{j=1}^{n} a_{j,i}w_i$ for all $i \in [\![n]\!]$, and therefore that

$$h(g(w_i)) = h\left(\sum_{j=1}^{n} a_{j,i}w_i\right) = \sum_{j=1}^{n} a_{j,i}h(w_i) = \sum_{j=1}^{n} a_{j,i}v_i = f(v_i) = f(h(w_i))$$

for all $i \in [\![n]\!]$. As $\mathscr{B}'$ is a basis for V, this implies that, in fact, $h \circ g = f \circ h$ or, equivalently, that $f = h \circ g \circ h^{-1}$, so that the endomorphisms f and g are similar. □

An important special case that we need to understand well is that of similarity of Jordan matrices:

Proposition 1.12.3. *Let r, s, n_1, n_2, …, n_r and m_1, m_2, …, m_s be positive integers such that*

$$n_1 + n_2 + \cdots + n_r = m_1 + m_2 + \cdots + m_s,$$

and let λ_1, λ_2, …, λ_r and μ_1, μ_2, …, μ_s be elements of $\Bbbk$. The Jordan matrices

$$J := \begin{pmatrix} J_{n_1}(\lambda_1) & & & \\ & J_{n_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{n_r}(\lambda_r) \end{pmatrix} \quad \text{and} \quad K := \begin{pmatrix} J_{m_1}(\mu_1) & & & \\ & J_{m_2}(\mu_2) & & \\ & & \ddots & \\ & & & J_{m_s}(\mu_s) \end{pmatrix}$$

are similar if and only if $r = s$ and there exists a bijective function $\pi : [\![r]\!] \to [\![s]\!]$ such that $n_i = m_{\pi(i)}$ and $\lambda_i = \mu_{\pi(i)}$ for all $i \in [\![r]\!]$.

Proof. Let us suppose first that $r = s$ and that there is a bijective function $\pi : [\![r]\!] \to [\![s]\!]$ such that

$n_i = m_{\pi(i)}$ and $\lambda_i = \mu_{\pi(i)}$ for all $i \in [\![r]\!]$, and let us consider the block matrix

$$C \coloneqq \begin{pmatrix} C_{1,1} & \cdots & C_{1,r} \\ \vdots & \ddots & \vdots \\ C_{r,1} & \cdots & C_{r,r} \end{pmatrix}$$

that for each choice of i and j in $[\![r]\!]$ has

$$C_{i,j} = \begin{cases} I \in \mathrm{M}_{n_j}(\Bbbk) & \text{if } \pi(i) = j; \\ 0 \in \mathrm{M}_{n_i,n_j}(\Bbbk) & \text{if } \pi(i) \neq j. \end{cases}$$

Each column and each row of the matrix C has exactly one non-zero entry, so C is an invertible matrix. In fact, by computing directly we can easily veryfy that its inverse matrix is the block matrix

$$C^{-1} \coloneqq \begin{pmatrix} D_{1,1} & \cdots & D_{1,r} \\ \vdots & \ddots & \vdots \\ D_{r,1} & \cdots & D_{r,r} \end{pmatrix}$$

with

$$D_{i,j} = \begin{cases} I \in \mathrm{M}_{n_j}(\Bbbk) & \text{if } \pi^{-1}(i) = j; \\ 0 \in \mathrm{M}_{n_i,n_j}(\Bbbk) & \text{if } \pi^{-1}(i) \neq j \end{cases}$$

for each choice of i and j in $[\![r]\!]$.

□

Using the results of the previous section we can provide a much better criterion for similarity. Doing this is, in fact, the original motivation that lead Camille Jordan to develop the theory.

Proposition 1.12.4. *Let V be a finite-dimensional vector space, and let $f, g : V \to V$ be two endomorphisms of V whose characteristic polynomials split completely over $\Bbbk$. The following statements are equivalent:*

(*a*) *The endomorphisms f and g are similar.*

(*b*) *The Jordan forms of f and of g are the same up to the ordering of the blocks.*

(*c*) *For all $k \in \mathbb{N}$ and all $\lambda \in \Bbbk$ we have that $\Delta_k(f, \mu) = \Delta_k(g, \mu)$.*

Notice that since we are supposing that the characteristic polynomials of f and g split completely over $\Bbbk$ the two maps do have Jordan forms, so the statement of the proposition makes sense.

Proof. (a) ⇒ (c) If the endomorphisms f and g are similar, then we know from Exercise 1.11.5 that for all $k \in \mathbb{N}$ and all $\lambda \in \Bbbk$ we have that $\Delta_k(f,\mu) = \Delta_k(g,\mu)$.

(c) ⇒ (b) Let $\mathscr{B}$ and $\mathscr{B}'$ be ordered bases of V such that the matrices $J \coloneqq [f]_{\mathscr{B}}$ and $K \coloneqq [g]_{\mathscr{B}'}$ are Jordan matrices. Exercise 1.11.3 tells us that $\Delta_k(f,\mu) = \Delta_k(J,\mu)$ and $\Delta_k(g,\mu) = \Delta_k(K,\mu)$ for all $k \in \mathbb{N}$ and all $\lambda \in \Bbbk$. If the statement (c) holds, then all this implies that $\Delta_k(J,\mu) = \Delta_k(K,\mu)$ for all $k \in \mathbb{N}$ and all $\mu \in \Bbbk$, so that the matrices J and K have the same Jordan blocks, so that the statement (b) also holds.

(c) ⇒ (a) Let us suppose that the statement (b) holds. There are ordered bases $\mathscr{B}$ and $\mathscr{B}'$ for V such that the matrices $J \coloneqq [f]_{\mathscr{B}}$ and $K \coloneqq [g]_{\mathscr{B}'}$ are in Jordan form and, up to a permutation of the blocks, equal. There is then an ordered basis $\mathscr{B}''$ that is a rearrangement of the basis $\mathscr{B}'$ for which the matrices $[f]_{\mathscr{B}}$ and $[g]_{\mathscr{B}''}$ are in Jordan form and *equal*: this implies, as we know, that the endomorphisms f and g are equal. □

Capítulo 2
Inner product spaces

§2.1. Inner products

In this chapter we will write $\mathbb{k}$ to refer either to the field $\mathbb{R}$ of real numbers or to the field $\mathbb{C}$ of complex numbers. In both cases, if λ is an element of $\mathbb{k}$ we will write $\overline{\lambda}$ for the complex conjugate of λ. Of course, if $\mathbb{k}$ is $\mathbb{R}$ then we have that $\overline{\lambda} = \lambda$ for all $\lambda \in \mathbb{k}$.

If V is a vector space over the field $\mathbb{k}$, then an ***inner product*** on V is a function

$$\langle -, - \rangle : V \times V \to \mathbb{k}$$

that for all $x, x', y \in V$ and all $\lambda \in \mathbb{k}$ has

(IP$_1$) $\langle x + x', y \rangle = \langle x, y \rangle + \langle x', y \rangle$, {ax:pi:1}

(IP$_2$) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$, {ax:pi:2}

(IP$_3$) $\langle x, y \rangle = \overline{\langle y, x \rangle}$, {ax:pi:3}

(IP$_4$) $\langle x, x \rangle > 0$ if $x \neq 0$. {ax:pi:4}

Let us remark that for all $x \in V$ the third condition implies that $\langle x, x \rangle = \overline{\langle x, x \rangle}$ and, therefore, that the scalar $\langle x, x \rangle$ is a real number: in particular, this tells us that the fourth condition in this definition makes sense. On the other hand, conditions (IP$_1$) and (IP$_2$) imply that an inner product is a linear function of its first argument, that is, that for each $v \in V$ the map

$$u \in V \mapsto \langle u, v \rangle \in \mathbb{k}$$

is linear. If $\mathbb{k} = \mathbb{R}$, then condition (IP$_3$) in turn implies immediately that $\langle -, - \rangle$ is also a linear function of its second argument. If instead $\mathbb{k} = \mathbb{C}$, then an inner product is a ***semilinear*** function of its second argument: this means that for all $u, v, v' \in V$ and all $\lambda \in \Bbbk$ we have that

$$\langle u, v + v' \rangle = \langle u, v \rangle + \langle u, v' \rangle, \qquad \langle u, \lambda v \rangle = \overline{\lambda} \langle u, v \rangle.$$

In any case, it follows from the linearity with respect to the first variable that

$$\langle 0, 0 \rangle = \langle 0 \cdot 0, 0 \rangle = 0 \cdot \langle 0, 0 \rangle = 0$$

and this, together with **(IP$_4$)**, implies that in fact for each $x \in V$ we have

$$\langle x, x \rangle = 0 \iff x = 0.$$

An ***inner product space*** is an ordered pair $(V, \langle -, - \rangle)$ in which V is a vector space over $\Bbbk$ and $\langle -, - \rangle$ is an inner product defined on V. Except in special situations we will write simply V instead of the pair $(V, \langle -, - \rangle)$ and we will say that V itself is an inner product space, leaving the notation for the inner product implicit.

The following examples present important families of inner product spaces.

Example 2.1.1. Let n be a positive integer, and let us consider the real vector space $\mathbb{R}^n$ and on it the function

$$\langle -, - \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$$

that on each pair of vectors $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ of $\mathbb{R}^n$ takes the value

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

This is an inner product on $\mathbb{R}^n$, and we call it the ***standard inner product*** on that vector space. Let us verify in detail that the conditions of the definition are indeed satisfied:

- If $x = (x_1, x_2, \dots, x_n)$, $x' = (x'_1, x'_2, \dots, x'_n)$ and $y = (y_1, y_2, \dots, y'_n)$ are elements of $\mathbb{R}^n$, then of course we have that $x + y = (x_1 + x'_1, x_2 + x'_2, \dots, x_n + x'_n)$ and therefore

$$\begin{aligned} \langle x + x', y \rangle &= (x_1 + x'_1) y_1 + (x_2 + x'_2) y_2 + \cdots + (x_n + x'_n) y_n \\ &= (x_1 y_1 + x_2 y_2 + \cdots + x_n y_n) + (x'_1 y_1 + x'_2 y_2 + \cdots + x'_n y_n) \\ &= \langle x, y \rangle + \langle x', y \rangle. \end{aligned}$$

- If $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y'_n)$ are elements of $\mathbb{R}^n$, and λ one of $\mathbb{R}$, then that $\lambda x = (\lambda x_1, \lambda x_2, \dots, \lambda x_n)$ and therefore

$$\begin{aligned} \langle \lambda x, y \rangle &= (\lambda x_1) y_1 + (\lambda x_2) y_2 + \cdots + (\lambda x_n) y_n \\ &= \lambda (x_1 y_1 + x_2 y_2 + \cdots + x_n y_n) \\ &= \lambda \langle x, y \rangle. \end{aligned}$$

- If $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y'_n)$ are elements of $\mathbb{R}^n$, then

  $$\begin{aligned} \langle x, y\rangle &= x_1y_1 + x_2y_2 + \cdots + x_ny_n \\ &= y_1x_1 + y_2x_2 + \cdots + y_nx_n \\ &= \langle y, x\rangle. \end{aligned}$$

- Finally, if $x = (x_1, x_2, \dots, x_n)$ is a non-zero element of $\mathbb{R}^n$, then we have that

  $$\langle x, x\rangle = x_1x_1 + x_2x_2 + \cdots + x_nx_n = x_1^2 + x_2^2 + \cdots + x_n^2 > 0,$$

  because all the terms in the sum $x_1^2 + x_2^2 + \cdots + x_n^2$ are non-negative and, since $x \neq 0$, at least one of them is strictly positive.

Example 2.1.2. Let n be a positive integer, and let us now consider the complex vector space $\mathbb{C}^n$ and on it the function

$$\langle -, -\rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$$

that on each pair of vectors $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ of $\mathbb{C}^n$ takes the value

$$\langle x, y\rangle = x_1\overline{y_1} + x_2\overline{y_2} + \cdots + x_n\overline{y_n}.$$

This is an inner product on $\mathbb{C}^n$, and we call it, as in the previous example, the ***standard inner product*** on that vector space. Let us check that the conditions of the definition are indeed satisfied:

- If $x = (x_1, x_2, \dots, x_n)$, $x' = (x'_1, x'_2, \dots, x'_n)$ and $y = (y_1, y_2, \dots, y'_n)$ are elements of $\mathbb{C}^n$, then $x + y = (x_1 + x'_1, x_2 + x'_2, \dots, x_n + x'_n)$ and thus

  $$\begin{aligned} \langle x + x', y\rangle &= (x_1 + x'_1)\overline{y_1} + (x_2 + x'_2)\overline{y_2} + \cdots + (x_n + x'_n)\overline{y_n} \\ &= (x_1\overline{y_1} + x_2\overline{y_2} + \cdots + x_n\overline{y_n}) + (x'_1\overline{y_1} + x'_2\overline{y_2} + \cdots + x'_n\overline{y_n}) \\ &= \langle x, y\rangle + \langle x', y\rangle. \end{aligned}$$

- If $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y'_n)$ are elements of $\mathbb{C}^n$, and λ is a complex number, then $\lambda x = (\lambda x_1, \lambda x_2, \dots, \lambda x_n)$ and therefore

  $$\begin{aligned} \langle \lambda x, y\rangle &= (\lambda x_1)\overline{y_1} + (\lambda x_2)\overline{y_2} + \cdots + (\lambda x_n)\overline{y_n} \\ &= \lambda(x_1\overline{y_1} + x_2\overline{y_2} + \cdots + x_n\overline{y_n}) \\ &= \lambda\langle x, y\rangle. \end{aligned}$$

- If $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y'_n)$ are elements of $\mathbb{C}^n$, then

  $$\begin{aligned} \langle x, y\rangle &= x_1\overline{y_1} + x_2\overline{y_2} + \cdots + x_n\overline{y_n} \\ &= \overline{y_1\overline{x_1} + y_2\overline{x_2} + \cdots + y_n\overline{x_n}} \\ &= \overline{\langle y, x\rangle}. \end{aligned}$$

- Finally, if $x = (x_1, x_2, \dots, x_n)$ is a non-zero element of $\mathbb{R}^n$, then we have that

  $$\langle x, x\rangle = x_1\overline{x_1} + x_2\overline{x_2} + \cdots + x_n\overline{x_n} = |x_1| + |x_2| + \cdots + |x_n| > 0,$$

  because all the terms in the sum $|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2$ are non-negative and, since $x \neq 0$, at least one of them is strictly positive.

Example 2.1.3. Let us now fix a non-negative integer n and consider the real vector space $\mathbb{R}[X]_{\leq n}$ of all real polynomials of degree at most n. If p and q are two elements of $\mathbb{R}[X]_{\leq n}$, then we can consider p and q as functions $\mathbb{R} \to \mathbb{R}$ as usual, and as such they are both continuous: in particular, it makes sense to consider the integral $\int_{-1}^{1} p(x)q(x)\,\mathrm{d}x$. It follows from this that there is a function

$$\langle -, -\rangle : \mathbb{R}[X]_{\leq n} \times \mathbb{R}[X]_{\leq n} \to \mathbb{R}$$

that for each choice of p and q in $\mathbb{R}[X]_{\leq n}$ has

$$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)\,\mathrm{d}x.$$

This is an inner product on the real vector space $\mathbb{R}[X]_{\leq n}$.

- If p, q and r are three elements of $\mathbb{R}[X]_{\leq n}$, then

  $$\begin{aligned}\langle p + q, r\rangle &= \int_{-1}^{1} (p(x) + q(x))r(x)\,\mathrm{d}x \\ &= \int_{-1}^{1} p(x)r(x)\,\mathrm{d}x + \int_{-1}^{1} q(x)r(x)\,\mathrm{d}x \\ &= \langle p, r\rangle + \langle q, r\rangle.\end{aligned}$$

- If p and q are elements of $\mathbb{R}[X]_{\leq n}$ and λ is a real number, then

  $$\langle \lambda p, q\rangle = \int_{-1}^{1} \lambda p(x)q(x)\,\mathrm{d}x = \lambda \int_{-1}^{1} p(x)q(x)\,\mathrm{d}x = \lambda\langle p, q\rangle.$$

- If p and q are elements of $\mathbb{R}[X]_{\leq n}$, then

  $$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)\,\mathrm{d}x = \int_{-1}^{1} q(x)x(x)\,\mathrm{d}x = \langle q, p\rangle.$$

- Finally, let p be a non-zero element of $\mathbb{R}[X]_{\leq n}$. There exists a number x_0 in the interval $(-1, 1)$ such that $p(x_0) \neq 0$: if that were not the case, then all elements of that interval would be roots of p, and a polynomial with an infinite number of roots is identically zero. Of course $p(x_0)^2 > 0$ and, since the function $x \in \mathbb{R} \mapsto p(x)^2 \in \mathbb{R}$ is continuous, this implies that there is a positive number ϵ such that $(x_0 - \epsilon, x_0 + \epsilon) \subseteq (-1, 1)$ and $p(x)^2 > \frac{1}{2}p(x_0)^2$ for all $x \in (x_0 - \epsilon, x_0 + \epsilon)$.

We have that

$$\begin{aligned}\langle p, p\rangle &= \int_{-1}^{1} p(x)^2 \,\mathrm{d}x \\ &\geq \int_{x_0-\epsilon}^{x_0+\epsilon} p(x)^2 \,\mathrm{d}x \qquad \text{because } p^2 \text{ is a non-negative function} \\ &\geq \int_{x_0-\epsilon}^{x_0+\epsilon} \tfrac{1}{2} p(x_0)^2 \,\mathrm{d}x \\ &= \epsilon p(x_0)^2 > 0.\end{aligned}$$

We will now present a final example of an inner product space that will be useful in what follows to construct certain counterexamples.

{exam:FoX}

Example 2.1.4. Let X be a set, and let us consider the set $\mathscr{F}(X)$ of all functions $X \to \Bbbk$, with its usual structure of a vector space over $\Bbbk$. Let us recall that whenever $f : X \to \Bbbk$ and $g : X \to \Bbbk$ are two elements of $\mathscr{F}(X)$ and a and b are two elements of $\Bbbk$ the function $af + bg : \mathbb{N}_0 \to \Bbbk$ takes on each element x of X the value

$$(af + bg)(x) = a \cdot f(x) + b \cdot g(x).$$

Similarly, if $f : X \to \Bbbk$ and $g : X \to \Bbbk$ are two elements of $\mathscr{F}(X)$ we can define a new function, the product $f \cdot g : X \to \Bbbk$, putting, for every $x \in X$,

$$(f \cdot g)(x) \coloneqq f(x) \cdot g(x).$$

The ***support*** of an element $f : X \to \Bbbk$ of $\mathscr{F}(X)$ is the set

$$\sigma(f) \coloneqq \{x \in X : f(x) \neq 0\}.$$

It is easy to check that

> *if $f : X \to \Bbbk$ and $g : X \to \Bbbk$ are two elements of $\mathscr{F}(X)$ and a and b are two scalars in $\Bbbk$, then the sets $\sigma(af + bg)$ and $\sigma(f \cdot g)$ are both contained in $\sigma(f) \cup \sigma(g)$,*

and it follows at once from this that if two elements $f : X \to \Bbbk$ and $g : X \to \Bbbk$ of $\mathscr{F}(X)$ are such that the sets $\sigma(f)$ and $\sigma(g)$ are finite, then the set $\sigma(af + bg)$ is also finite. This implies immediately that the set

$$\mathscr{F}_0(X) \coloneqq \{f \in \mathscr{F}(X) : \text{the set } \sigma(f) \text{ is finite}\}$$

is a subspace of $\mathscr{F}(X)$. We call it the ***space of functions of finite support*** on X.

If f and g are two elements of the subspace $\mathscr{F}_0(X)$, then the function $\overline{g} : x \in X \mapsto \overline{g(x)} \in \Bbbk$ has the same support as g, so it also belongs to $\mathscr{F}_0(X)$, and therefore the support $\sigma(f \cdot \overline{g})$ of the product $f \cdot \overline{g}$ is a finite set contained in $\sigma(f) \cup \sigma(g)$: we may therefore consider the scalar

$$\langle f, g \rangle \coloneqq \sum_{x \in \sigma(f) \cup \sigma(g)} f(x)\overline{g(x)}.$$

In this way we obtain a function

$$\langle -, - \rangle : \mathscr{F}_0(X) \times \mathscr{F}_0(X) \to \Bbbk.$$

This is, in fact, an inner product on the vector space $\mathscr{F}_0(X)$. We will leave the details of the verification of this claim to the reader.

Exercise 2.1.5. Prove that the function $\langle -, - \rangle : \mathscr{F}_0(X) \times \mathscr{F}_0(X) \to \Bbbk$ constructed in the previous example is an inner product on the vector space $\mathscr{F}_0(X)$.

A very basic application of inner products is that they allow us to recognize the zero vector and to compare vectors:

{prop:zero}

Proposition 2.1.6. *Let V be an inner product space, and let x and x' be two elements of V.*

(*i*) $\langle x, y \rangle = 0$ *for all* $y \in V$ *if and only if* $x = 0$.

(*ii*) $\langle x, y \rangle = \langle x', y \rangle$ *for all* $y \in V$ *if and only if* $x = x'$.

The interest of these two statements is that they allow us to compare vectors by comparing numbers: for example, the second part says that two vectors x and x' are equal exactly when for all vectors y the two numbers $\langle x, y \rangle$ and $\langle x', y \rangle$ are equal.

Proof. If $x = 0$, then, as we observed above, $\langle x, x \rangle = 0$. On the other hand, if $\langle x, y \rangle = 0$ for all $y \in V$, then in particular we have that $\langle x, x \rangle = 0$, and therefore, according to the condition (**IP**$_4$), we must have $x = 0$. This proves the first part of the proposition. On the other hand, since $\langle x, y \rangle - \langle x', y \rangle = \langle x - x', y \rangle$ for all $y \in V$, the second part follows immediately from the first one. □

We can restrict an inner product defined on a vector space to any subspace, and in this way we can construct examples of inner product spaces:

Proposition 2.1.7. *Let V be an inner product space and let W be a subspace of V. The function*

$$\langle -, - \rangle_W : (w, w') \in W \times W \mapsto \langle w, w' \rangle \in \Bbbk$$

that is obtained by restricting the inner product $\langle -, - \rangle : V \times V \to \Bbbk$ of V to the subset $W \times W$ of $V \times V$ is an inner product on W.

From now on we will always consider a subspace of an inner product space to be an inner product space itself with respect to this restricted inner product.

Proof. We have to show that the function $\langle -,-\rangle_W$ defined in the statement of the proposition satisfies the four properties ($\mathbf{IP}_1$)–($\mathbf{IP}_4$) of the definition, and this is immediate because the function $\langle -,-\rangle$ does. □

Let us finish this section by describing *all* inner products on the vector space $\mathbb{R}^2$.

{exam:r2}

Example 2.1.8. Let us suppose that $\langle -,-\rangle : \mathbb{R}^2\times\mathbb{R}^2 \to \mathbb{R}$ is an inner product on the vector space $\mathbb{R}^2$, and let (e_1, e_2) be the standard ordered basis for $\mathbb{R}^2$. We can consider the four numbers

$$a \coloneqq \langle e_1, e_1\rangle, \qquad b \coloneqq \langle e_1, e_2\rangle, \qquad c \coloneqq \langle e_2, e_1\rangle, \qquad d \coloneqq \langle e_2, e_2\rangle.$$

Condition ($\mathbf{IP}_4$) tells us that $a > 0$ and $d > 0$, and condition ($\mathbf{IP}_3$) that $b = c$, and it follows from this that the real matrix

$$A \coloneqq \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is symmetric and has positive diagonal entries. On the other hand, for all $t \in \mathbb{R}$ we have that

$$0 \leq \langle te_1 + e_2, te_1 + e_2\rangle = \langle te_1, te_1\rangle + \langle e_2, te_1\rangle + \langle te_1, e_2\rangle + \langle e_2, te_2\rangle = at^2 + 2bt + c.$$

We thus see that the polynomial function

$$t \in \mathbb{R} \mapsto at^2 + 2bt + c \in \mathbb{R}$$

is everywhere non-negative, so that it has at most one real root and therefore its discriminant $\Delta = 4b^2 - 4ac$ is non-positive. The conclusion of this is that

> *if $\langle -,-\rangle : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is an inner product on $\mathbb{R}^2$ and we put $a \coloneqq \langle e_1, e_1\rangle$, $b \coloneqq \langle e_1, e_2\rangle$, $c \coloneqq \langle e_2, e_1\rangle$, and $d \coloneqq \langle e_2, e_2\rangle$, then the matrix $A \coloneqq \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is symmetric and its diagonal entries and determinant are positive.*

Let us now show that, conversely,

> *if $A \coloneqq \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is symmetric matrix whose diagonal entries and determinant are positive, then there is an inner product $\langle -,-\rangle : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is an inner product on $\mathbb{R}^2$ such that $a = \langle e_1, e_1\rangle$, $b = \langle e_1, e_2\rangle$, $c = \langle e_2, e_1\rangle$, and $d = \langle e_2, e_2\rangle$.*

To do this, let us suppose $A \coloneqq \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ is a matrix that is symmetric and has positive diagonal entries and determinant, and let us consider the function

$$\langle -,-\rangle : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$$

such that

$$\langle (x_1, x_2), (y_1, y_2) \rangle = ax_1y_1 + bx_2y_1 + cx_1y_2 + dx_2y_2$$

for all choices of two elements (x_1, x_2) and (y_1, y_2) of $\mathbb{R}^2$. It is immediately clear that $a = \langle e_1, e_1 \rangle$, $b = \langle e_1, e_2 \rangle$, $c = \langle e_2, e_1 \rangle$, and $d = \langle e_2, e_2 \rangle$, so to prove what we want we need only verify that $\langle -, - \rangle$ is an inner product on $\mathbb{R}^2$.

- If $x = (x_1, x_2)$, $x' = (x_1, x_2)$ and $y = (y_1, y_2)$ are three elements of $\mathbb{R}^2$, then

  $$\begin{aligned} \langle x + x', y \rangle &= \langle (x_1 + x_1', x_2 + x_2'), (y_1, y_2) \rangle \\ &= a(x_1 + x_1')y_1 + b(x_2 + x_2')y_1 + c(x_1 + x_1')y_2 + d(x_2 + x_2')y_2 \\ &= (ax_1y_1 + bx_2y_1 + cx_1y_2 + dx_2y_2) + (ax_1'y_1 + bx_2'y_1 + cx_1'y_2 + dx_2'y_2) \\ &= \langle x, y \rangle + \langle x', y \rangle. \end{aligned}$$

- If $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are two elements of $\mathbb{R}^2$ and λ is a rea number, then

  $$\begin{aligned} \langle \lambda x, y \rangle &= \langle (\lambda x_1, \lambda x_2), (y_1, y_2) \rangle \\ &= a(\lambda x_1)y_1 + b(\lambda x_2)y_1 + c(\lambda x_1)y_2 + d(\lambda x_2)y_2 \\ &= \lambda(ax_1y_1 + bx_2y_1 + cx_1y_2 + dx_2y_2) \\ &= \lambda \langle x, y \rangle. \end{aligned}$$

- If $x = (x_1, x_2)$ and $y = (y_1, y_2)$ are two elements of $\mathbb{R}^2$, then

  $$\begin{aligned} \langle (x_1, x_2), (y_1, y_2) \rangle &= ax_1y_1 + bx_2y_1 + cx_1y_2 + dx_2y_2 \\ &= ay_1x_1 + by_2x_1 + cy_1x_2 + dy_2x_2 \qquad \text{because } b = c \\ &= \langle (y_1, y_2), (x_1, x_2) \rangle. \end{aligned}$$

- If $x = (x_1, x_2)$ is a non-zero element of $\mathbb{R}^2$, then, since $b = c$, we have that

  $$\langle x, x \rangle = ax_1^2 + 2bx_1x_2 + dx_2^2.$$

  If $x_2 = 0$, then we must have $x_1 \neq 0$ and therefore $\langle x, x \rangle = ax_1^2 > 0$ since $a > 0$. If instead $x_2 \neq 0$, then putting $t \coloneqq x_1/x_2$ we have that

  $$\langle x, x \rangle = x_2^2(at^2 + 2bt + d)$$

  and again this is a positive: we have that $x_2^2 > 0$ and that $at^2 + 2bt + d > 0$, as the polynomial $aX^2 + 2bX + d$ has negative discriminant and positive constant term.

These observations provide a description of all the inner products on $\mathbb{R}^2$. One way to phrase the result is the following. Let us write $\mathscr{I}(\mathbb{R}^2)$ for the set of all inner products on $\mathbb{R}^2$, and $\mathscr{S}_2$ for the set of all symmetric matrices in $\mathrm{M}_2(\mathbb{R})$ with positive diagonal entries and determinant. What we have proved above implies immediately that we have a function

$$\Phi : \langle -, - \rangle \in \mathscr{I}(\mathbb{R}^2) \longmapsto \begin{pmatrix} \langle e_1, e_1 \rangle & \langle e_1, e_2 \rangle \\ \langle e_2, e_1 \rangle & \langle e_2, e_2 \rangle \end{pmatrix} \in \mathscr{S}_2,$$

and that it is surjective. In fact, it is a bijection — this is a consequence of the following exercise.

Exercise 2.1.9. Prove that the function Φ defined at the end of Example 2.1.8 is injective.

§2.2. Norms and metrics

When we have an inner product on a vector space V we are able to talk about the *size* of the elements of V and about the distance that separates two elements of V. In this section we will explain how this is done. In the next one, moreover, we will show how to measure angles in an inner product space, and with all this structure we will be able to do geometry in inner product spaces much as we do in the Euclidean plane.

Let V be a vector space over $\Bbbk$. A ***norm*** on V is a function $\|-\| : V \to \mathbb{R}_{\geq 0}$ that for each choice of $x, y \in V$ and $\lambda \in \Bbbk$ satisfies the following conditions:

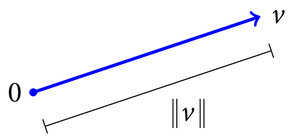
{ax:norm:1}

(N$_1$) $\|x\| = 0$ if and only if $x = 0$; {ax:norm:2}

(N$_2$) $\|\lambda x\| = |\lambda| \|x\|$; and {ax:norm:3}

(N$_3$) $\|x + y\| \leq \|x\| + \|y\|$.

We call (N$_3$) the ***triangular inequality***. Usually we view the norm $\|v\|$ of a vector v of V as a measure of its *size* or, more directly, as its *length*.



The interest of this notion for us is that on every inner product space we can define, in a canonical way, a norm:

{prop:norm}

Proposition 2.2.1. *Let V be an inner product space. The function*

$$\|-\| : x \in V \mapsto \langle x, x \rangle^{1/2} \in \mathbb{R}$$

is a norm on V and for each $x, y \in V$ we have that

$$|\langle x, y\rangle| \leq \|x\|\|y\|. \tag{2.1}$$

{eq:cs}

Moreover, the equality appearing here holds if and only if the set $\{x, y\}$ is linearly dependent.

From now on we will consider on every inner product space the the norm described in this proposition, which we call the norm associated to the inner product. The inequality (2.1) that appears in this proposition is the *Cauchy–Bunyakovsky–Schwartz inequality*, after Augustin-Louis Cauchy (1789–1857, France), Viktor Bunyakovsky (1804–1889, Russia) and Hermann Schwarz (1843–1921, Germany). The first of these three authors proved the inequality in the special case of the vector space $\Bbbk^n$ with its standard inner product, the second one for vector spaces of functions with an inner product given by an integral, and the third one proved essentially the general case considered here.

Proof. We have to verify that the three conditions of the definition of norms are satisfied. The first two are immediate, and in order to prove the third one we will first establish the inequality (2.1).

Let $x, y \in V$. If $y = 0$, then the inequality (2.1) is obvious, so we may suppose that $y \neq 0$. For each $\lambda \in \Bbbk$ we have that

$$0 \leq \|x - \lambda y\|^2 = \langle x - \lambda y, x - \lambda y\rangle = \langle x, x\rangle - \lambda\langle y, x\rangle - \bar{\lambda}\big(\langle x, y\rangle - \lambda\langle y, y\rangle\big).$$

In particular, if we take $\lambda = \langle x, y\rangle / \langle y, y\rangle$, the expression between parenthesis becomes zero and therefore we see that

$$0 \leq \langle x, x\rangle - \frac{\langle x, y\rangle\langle y, x\rangle}{\langle y, y\rangle} = \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y^2\|}.$$

This is equivalent to the inequality (2.1). Moreover, if the equality holds then clearly $\|x - \lambda y\|^2 = 0$, so that $x = \lambda y$ and x and y are linearly dependent.

We will now prove, using the Cauchy–Bunyakovsky–Schwartz inequality, that the third condition in the definition of norms is also satisfied. If $x, y \in V$, then

$$\begin{aligned}
\|x + y\|^2 &= \langle x + y, x + y\rangle \\
&= \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\
&= \|x\|^2 + 2\operatorname{Re}\langle x, y\rangle + \|y, y\| \\
&\leq \|x\|^2 + 2|\langle x, y\rangle| + \|y, y\| \\
&\leq \|x\|^2 + 2\|x\|\|y\| + \|y, y\| \\
&= \big(\|x\| + \|y\|\big)^2
\end{aligned}$$

and clearly this implies that the triangular inequality holds. □

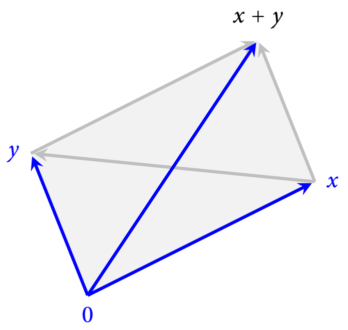


Figure 2.1. The Parallelogram Law tells us that the sum of the squares of the lengths of the diagonals of a parallelogram is equal to the sum of the squares of the lenths of its sides.

{fig:parallelogram}

We say that an element v of an inner product space V is a ***unit vector*** if $\|v\| = 1$. Clearly, a unit-vector is necessarily non-zero. On the other hand, if w is an arbitrary non-zero element of V, then the vector $v \coloneqq w/\|w\|$ is immediately seen to be a unit vector: we say that v is obtained from w by ***normalization***.

A norm that is constructed from an inner product satisfies certain geometric conditions of which the following one is the most important:

{prop:pi:para}

Proposition 2.2.2. *Let V be an inner product space and let $\|-\|$ be the norm on V associated to the inner product of V.*

(i) (Parallelogram law) *If $x, y \in V$, then*

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

(ii) *If $\mathbb{k} = \mathbb{R}$, then for each $x, y \in V$ we have that*

$$\langle x, y\rangle = \tfrac{1}{4}\|x+y\|^2 - \tfrac{1}{4}\|x-y\|^2.$$

If instead $\mathbb{k} = \mathbb{C}$, then for each $x, y \in V$ we have that

$$\langle x, y\rangle = \tfrac{1}{4}\|x+y\|^2 - \tfrac{1}{4}\|x-y\|^2 + \tfrac{i}{4}\|x+iy\|^2 - \tfrac{i}{4}\|x-iy\|^2.$$

The *Parallelogram Law* has a very direct geometrical interpretation which we have represented graphically in Figure 2.1. On the other hand, the second part of this proposition tells us that the inner product of an inner product space is completely determined by the associated norm.

Proof. To prove the first part we let x and y be any two vectors in V and compute:

$$\begin{aligned}\|x+y\|^2+\|x-y\|^2 &= \langle x+y,x+y\rangle+\langle x-y,x-y\rangle \\ &= \big(\langle x,x\rangle+\langle x,y\rangle+\langle y,x\rangle+\langle y,y\rangle\big)+\big(\langle x,x\rangle-\langle x,y\rangle-\langle y,x\rangle+\langle y,y\rangle\big) \\ &= 2\langle x,x\rangle+2\langle y,y\rangle \\ &= 2\|x\|^2+2\|y\|^2.\end{aligned}$$

If $\mathbb{k}=\mathbb{R}$, we can also compute that

$$\begin{aligned}\|x+y\|^2-\|x-y\|^2 &= \langle x+y,x+y\rangle-\|x-y,x-y\| \\ &= \big(\langle x,x\rangle+\langle x,y\rangle+\langle y,x\rangle+\langle y,y\rangle\big)-\big(\langle x,x\rangle-\langle x,y\rangle-\langle y,x\rangle+\langle y,y\rangle\big) \\ &= 4\langle x,y\rangle,\end{aligned}$$

and this proves the statement about real inner product spaces made in the second part of the proposition. The statement about complex inner product spaces can be proved in the same way: we leave that to the reader. □

Exercise 2.2.3. Complete the proof of Proposition 2.2.2.

The first part of Proposition 2.2.2 tells us that the Parallelogram Law is a necessary condition for a norm to be associated to an inner product. The following result que of Pascual Jordan (1902–1980, Germany) and John von Neumann (1903–1957, Hungary) published in [JVN35] states that it is also a sufficient condition.

Proposition 2.2.4. *Let V be a vector space, and let $\|-\|:V\to\mathbb{R}_{\geq 0}$ be a norm on V. There exists an inner product on V whose associated norm is $\|-\|$ if and only if $\|-\|$ satisfies the* Parallelogram Law, *that is, if for every choice of x and y in V we have that*

$$\|x+y\|^2+\|x-y\|^2=2\|x\|^2+2\|y\|^2.$$

We will only prove this result in the case of real inner product spaces — the argument for the complex case is similar but more involved. We will not make use if this result in what follows, so the reader may wish to skip this proof.

Proof. As we noted above, we only have to show that the condition is sufficient, and we will suppose that $\mathbb{k}=\mathbb{R}$. If the condition hold, then for all x, x' and y in V we have that

$$\|x+x'+y\|^2=\|(x+y)+x'\|^2=2\|x+y\|^2+2\|x'\|^2-\|x+y-x'\|^2$$

and that

$$\|x+x'-y\|^2=\|x+(x'-y)\|^2=2\|x\|^2+2\|x'-y\|^2-\|x-x'+y\|^2.$$

Substracting we see that

$$\|x+x'+y\|^2-\|x+x'-y\|^2=2\|x+y\|^2+2\|x'\|^2-2\|x\|^2-2\|x'-y\|^2$$

and interchanging the roles of x and x' we conclude that also

$$\|x+x'+y\|^2-\|x+x'-y\|^2=2\|x'+y\|^2+2\|x\|^2-2\|x'\|^2-2\|x-y\|^2.$$

From this two equalities we see that

$$\|x+x'+y\|^2-\|x+x'-y\|^2=\|x+y\|^2-\|x-y\|^2+\|x'+y\|^2-\|x'-y\|^2 \tag{2.2}$$

{eq:jvn:o}

Let $\langle -,-\rangle : V\times V\to\mathbb{R}$ be the function such that

$$\langle x,y\rangle=\tfrac{1}{4}\|x+y\|^2-\tfrac{1}{4}\|x-y\|^2$$

whenever x and y are in V, and let us check that it is an inner product on V.

- If x, x', $y\in V$, then $\langle x+x',y\rangle=\langle x,y\rangle+\langle x',y\rangle$, since this equality is, according to the definition of the function $\langle -,-\rangle$, equivalent to (2.2). This means, of course, that the condition **(IP$_1$)** is satisfied.
- If x and y are elements of V, then

  $$\langle x,y\rangle=\tfrac{1}{4}\|x+y\|^2-\tfrac{1}{4}\|x-y\|^2=\tfrac{1}{4}\|y+x\|^2-\tfrac{1}{4}\|y-x\|^2=\langle y,x\rangle$$

  and

  $$\langle x,x\rangle=\tfrac{1}{4}\|x+x\|^2-\tfrac{1}{4}\|x-x\|^2=\|x\|^2\geq 0, \tag{2.3}$$

  {eq:jvn:x}

  so the conditions **(IP$_3$)** and **(IP$_4$)** are satisfied.
- For each rational number $r\in\mathbb{Q}$ let $P(r)$ be the statement

  for all x, $y\in V$ we have $\langle rx,y\rangle=r\langle x,y\rangle$.

  We want to show that the statement $P(r)$ holds for all $r\in\mathbb{Q}$, and we will do so in several steps.

  - First, let us note that if r is a rational number and the statement $P(r)$ holds, then for each x, $y\in V$ we have that

    $$\langle (r+1)x,y\rangle=\langle rx+x,y\rangle=\langle rx,y\rangle+\langle x,y\rangle=r\langle x,y\rangle+\langle x,y\rangle=(r+1)\langle x,y\rangle,$$

    so that the statement $P(r+1)$ also holds. Since $P(1)$ evidently holds, this implies that, in fact, $P(r)$ holds for all $r\in\mathbb{N}$.

  - We claim that the statement $P(r)$ holds if r is a positive rational number. Indeed, if $r = \frac{p}{q}$ with $p, q \in \mathbb{N}$, then for each $x, y \in V$ we have that

    $$p\langle x, y\rangle = \langle px, y\rangle = \left\langle q\left(\tfrac{p}{q}x\right), y\right\rangle = q\left\langle \tfrac{p}{q}x, y\right\rangle,$$

    because $P(p)$ and $P(q)$ hold, so that $\left\langle \frac{p}{q}x, y\right\rangle = \frac{p}{q}\langle x, y\rangle$.
  - The statement $P(0)$ holds, since

    $$\langle 0x, y\rangle = \tfrac{1}{4}\|0x + y\|^2 - \tfrac{1}{4}\|0x - y\|^2 = \tfrac{1}{4}\|y\|^2 - \tfrac{1}{4}\|y\|^2 = 0 = 0\langle x, y\rangle.$$

    On the other hand, if the statement $P(r)$ holds for some $r \in \mathbb{Q}$ then the statement $P(-r)$ also holds: for all x and y in V we have in that case that

    $$0 = \langle 0, y\rangle = \langle rx - rx, y\rangle = \langle rx + (-r)x, y\rangle = \langle rx, y\rangle + \langle -rx, y\rangle,$$

    so that $\langle -rx, y\rangle = -\langle rx, y\rangle = -r\langle x, y\rangle$.

  Putting all this together we see at once that we can conclude that $P(r)$ holds for al $r \in \mathbb{Q}$, as we wanted.

- Let again x and y be two elements V and let us show now that

  $$|\langle x, y\rangle| \leq \|x\|\|y\|. \tag{2.4}$$

  {eq:jvn:cs}

  This is obvious if y is 0, so we may suppose that is not the case. In view of what we have already proved, we know that for every $r \in \mathbb{Q}$ we have that

  $$0 \leq \|x - ry\|^2 = \langle x - ry, x - ry\rangle = \langle x, x\rangle - 2r\langle x, y\rangle + r^2\langle y, y\rangle.$$

  This tells us that the polynomial

  $$\|y\|^2X^2 - 2\langle x, y\rangle X + \|x\|^2 \in \mathbb{R}[X]$$

  does not take negative values on $\mathbb{Q}$: since it is a continuous function, it follows from this that in fact it does not take negative values on any point in $\mathbb{R}$ and, in particular, its discriminant is non-negative, that is, we have that $\langle x, y\rangle^2 - \|x\|^2\|y\|^2 \leq 0$. The inequality (2.4) is a consequence of this.

- If $x, y \in V$, then the function $\zeta : t \in \mathbb{R} \mapsto \langle tx, y\rangle \in \mathbb{R}$ is continuous. To see this it is enough to note that, according to what we already know abut the function $\langle -, -\rangle$, we have that

  $$|\zeta(s) - \zeta(t)| = \big|\langle sx, y\rangle - \langle tx, y\rangle\big| = \big|\langle (s - t)x, y\rangle\big| \leq \|(s - t)x\|\|y\| \leq |s - t|\|x\|\|y\|.$$

- Finally, let x and y be two elements of V. According to the two last steps me made we know that the function $\zeta : t \in \mathbb{R} \mapsto \langle tx, y\rangle - t\langle x, y\rangle \in \mathbb{R}$ is continuous and that it vanishes on $\mathbb{Q}$: this implies, of course, that it is identically zero and thus that for all $t \in \mathbb{R}$ we have $\langle tx, y\rangle = t\langle x, y\rangle$. The function $\langle -, -\rangle$ therefore satisfies the condition (**IP**$_2$).

We can conclude from all this that the function $\langle -, - \rangle$ is an inner product on V, and according to the equality (2.3) the norm assocated to that inner product is precisely the norm $\|-\|$ with which we started. This proves the proposition. □

As the *Parallelogram Law* is a necessary condition on a norm for it to be associated to an inner product, we can use to to exhibit examples of norms which do not have that property.

{exam:notinner}

Example 2.2.5. The function

$$\|-\| : (x, y) \in \mathbb{R}^2 \mapsto |x| + |y| \in \mathbb{R}_{\geq 0}$$

is a norm on $\mathbb{R}^2$, as the reader can easily verify. If $e_1 = (1, 0)$ y $e_2 = (0, 1)$ are the vectors in the standard basis of $\mathbb{R}^2$, then

$$\|e_1 + e_2\|^2 + \|e_1 - e_2\|^2 = 8 \neq 4 = \|e_1\|^2 + 2\|e_2\|^2.$$

This shows that the norm $\|-\|$ does not satisfy the *Parallelogram Law* and, therefore, that there does not exists an inner product on $\mathbb{R}^2$ for which it is the associated norm,

Exercise 2.2.6. Prove that the function $\|-\|$ defined in the Example 2.2.5 is a norm on the vector space $\mathbb{R}^2$.

Norms allow us to measure the length of vectors, and thanks to that we can also measure the *distance* between two vectors. We end this section explaining how this works.

If V is a vector space, then a function $d : V \times V \to \mathbb{R}_{\geq 0}$ is a ***distance function*** on V if whenever x, y, and z are elements of V we have that

(**M**$_1$) $d(x, y) = d(y, x)$;

(**M**$_2$) $d(x, y) = 0$ if and only if $x = y$; and

(**M**$_3$) $d(x, z) \leq d(x, y) + d(y, z)$.

We say that that distance function is ***invariant*** if additionally

(**M**$_4$) $d(x, y) = d(x + z, y + z)$

and that it is ***homogenous*** if for each $\lambda \in \Bbbk$ we have that

(**M**$_5$) $d(\lambda x, \lambda y) = |\lambda|\, d(x, y)$.

{prop:dist}

Proposition 2.2.7. *Let V be a vector space, and let $\|-\| : V \to \mathbb{R}$ be a norm on V. The function*

$$d : (v, w) \in V \times V \mapsto \|v - w\| \in \mathbb{R}_{\geq 0}$$

is a homogenous and invariant distance function on V.

Proof. This follows immediately from the definitions. We leave the verification of the details of this to the reader as an exercise. □

If V is a vector space, then there are many distance functions $d : V \times V \to \mathbb{R}_{\geq 0}$ that do not arise from a norm as in the proposition above. Let us see a simple example of this.

{exam:notnorm}

Example 2.2.8. Let us consider the one-dimensional real vector space $\mathbb{R}$, and the function $d : \mathbb{R} \times \mathbb{R} \to \mathbb{R}_{\geq 0}$ such that

$$d(x, y) = \max\{|x - y|, 1\}$$

whenever x and y are in $\mathbb{R}$. It is easy to check that this is a distance function on $\mathbb{R}$. It does not arise from a norm because it is a bounded function and

> *if $d : V \times V \to \mathbb{R}_{\geq 0}$ is a distance function on a non-zero vector space that arises from a norm as in Proposition 2.2.7, then d is not bounded.*

We leave the verification of these facts as an exercise for the reader.

Exercise 2.2.9. Verify all the claims of Example 2.2.8.

§2.3. Orthogonality

Let V be an inner product space. Two vectors x and y are ***orthogonal*** if $\langle x, y \rangle = 0$, and in that case we write $x \perp y$. It follows immediately from the condition (IP$_3$) in the definition of inner products that $\perp$ is a symmetric relation on the set V.

Lemma 2.3.1. *Let V be an inner product space.*

(*i*) *Every vector in V is orthogonal to* 0.

(*ii*) *If $x \in V$ is such that $x \perp y$ for all $y \in V$, then* $x = 0$.

Proof. The first claim is immediate, and the second one follows at once from the first part of Proposition 2.1.6. □

We extend the usage of the relation $\perp$ to sets: if A and B are two subsets of an inner product space V, then we will write $A \perp B$ to indicate that every vector of A is orthogonal to every vector of B. Again, this is a symmetric relation on the set of all subsets of V.

The geometric idea behind the definition of orthogonality is very simple: two vectors are orthogonal if the angle they form is a straight angle. Of course, at this point this statement does not make sense because we have no definition of what the angle determined by two vectors is. In any case, we have the following result, whose first part should remind the reader of Pythagoras' Theorem on right triangles.

Proposition 2.3.2. *Let V be an inner product space, and let x and y be two vectors in V.* {prop:pi:pitagoras}

(*i*) *If $x \perp y$, then $\|x+y\|^2 = \|x\|^2 + \|y\|^2$.* {prop:pi:pitagoras:1

More generally, we have that

(*ii*) $\|x+y\|^2 = \|x\|^2 + \|y\|^2 + 2\operatorname{Re}\langle x, y\rangle$. {prop:pi:pitagoras:2

Proof. The first part of the proposition follows immediately from the second one, so it will be enough that we check the latter. This can be done by direct calculation: if x and y are two vectors in V, then the definition of the norm associated to the inner product of V implies that

$$\begin{aligned}\|x+y\|^2 &= \langle x+y, x+y\rangle \\ &= \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle \\ &= \|x\|^2 + \|y\|^2 + \langle x,y\rangle + \overline{\langle x,y\rangle} \\ &= \|x\|^2 + \|y\|^2 + 2\operatorname{Re}\langle x,y\rangle. \qquad \square\end{aligned}$$

With the second part of Proposition 2.3.2 as motivation, we can make the following definition: if V is an inner product space and x and y are two non-zero vectors in V, then the ***angle*** determined by x and y is the unique real number $\theta(x,y) \in [0,\pi]$ such that

$$\cos\theta(x,y) = \frac{\operatorname{Re}\langle x,y\rangle}{\|x\|\cdot\|y\|}. \tag{2.5}$$ {eq:pi:angulo}

This definition does make sense: from the Cauchy–Bunyakovsky–Schwartz inequality we know that

$$|\operatorname{Re}\langle x,y\rangle| \leq |\langle x,y\rangle| \leq \|x\|\|y\|,$$

so the quotient that appears on the right of the definition (2.5) is an element of the interval $[-1,1]$ and, therefore, is the cosine of exactly one number in the interval $[0,\pi]$.

Using this definition our intuition becomes a fact: two non-zero vectors in an inner product space are orthogonal if and only if the angle determined by them is $\pi/2$. Similarly, we can reinterpret the second part of Proposition 2.3.2 as a version of the *Law of cosines* from Euclidean geometry:

Proposition 2.3.3. *Let V be an inner product space. If x and y are two non-zero vectors in V and $\theta(x,y)$ is the angle they determine, then*

$$\|x-y\|^2 = \|x\|^2 + \|y^2\| - 2\cdot\|x\|\cdot\|y\|\cdot\cos\theta(x,y).$$

Proof. This equality can be obtained by replacing y by $-y$ in the equality of Proposition (*ii*) and using the definition of $\theta(x, y)$. □

If V is an inner product space and A is a subset of V, then we say that

- A is ***orthogonal*** if for each choice of two different elements x and y of A we have that $x \perp y$,

and that

- A is ***orthonormal*** if it is orthonormal and additionally $\|x\| = 1$ for each $x \in A$.

There is a useful relation between orthogonality, which is a geometric condition, and linear independence, which is an algebraic condition:

{prop:pi:ortogonal}
{prop:pi:ortogonal:
{prop:pi:orthogona

Proposition 2.3.4. *Let V be an inner product space and let A be a subset of V.*

(*i*) *If A is orthogonal and $0 \notin A$, then A is linearly independent.*

(*ii*) *If A is orthonormal, then it is linearly independent.*

Proof. Let us suppose that the set A is orthogonal and does not contain 0. Let n be a positive integer, let $x_1, x_2, \dots, x_n$ be pairwise different elements of A, and let $\lambda_1, \lambda_2, \dots, \lambda_n \in \Bbbk$ be scalars such that $\sum_{i=1}^{n} \lambda_i x_i = 0$. If $j \in \{1, \dots, n\}$, then

$$0 = \langle 0, x_j \rangle = \left\langle \sum_{i=1}^{n} \lambda_i x_i, x_j \right\rangle = \sum_{i=1}^{n} \lambda_i \langle x_i, x_j \rangle.$$

Since A is orthonormal, the only term in this sum which is possibly non-zero is the one in which the index i is equal to j, so we have that $\lambda_j \langle a_j, a_j \rangle = 0$. Now $0 \notin A$, so $a_j \neq 0$ and therefore $\langle a_j, a_j \rangle \neq 0$, and we can conclude that $\lambda_j = 0$. As this is true for each $j \in \{1, \dots, n\}$, we see that the set A is linearly independent. This proves part (*i*) of the proposition, and the second one follows immediately from it since an orthonormal set is orthogonal and does not contain 0. □

If A is an orthonormal set of an inner product space, then we know that every element in the span $\langle A \rangle$ of A can be written in a unique way as a linear combination of the elements of A. In fact, we can describe exactly what the coefficients that appear in that linear combination are: this is the content of the first part of the following result.

{prop:pi:ortogonal-
{prop:pi:ortogonal:
{prop:pi:ortogonal:

Proposition 2.3.5. *Let V be an inner product space and let A be an orthonormal subset of V.*

(*i*) *Let n be a positive integer, let $x_1, x_2, \dots, x_n$ be pairwise different elements of A, let $\lambda_1, \lambda_2, \dots, \lambda_n$ be scalars in $\Bbbk$, and put $x \coloneqq \sum_{i=1}^{n} \lambda_i x_i$. For every $j \in \{1, \dots, n\}$ we have that $\lambda_j = \langle x, x_j \rangle$.*

(*ii*) *For each $x \in \langle A \rangle$ the set $A_x = \{y \in A : \langle x, y \rangle \neq 0\}$ is finite and $x = \sum_{y \in A_x} \langle x, y \rangle y$.*

Proof. (*i*) If $j \in \{1, \dots, n\}$, then

$$\langle x, x_j \rangle = \left\langle \sum_{i=1}^{n} \lambda_i x_i, x_j \right\rangle = \sum_{i=1}^{n} \lambda_i \langle x_i, x_j \rangle.$$

As the set A is orthonormal, the only term in this sum which is possibly non-zero is the one in which the index i is equal to j, and it is equal to $\lambda_j \langle x_j, x_j \rangle = \lambda_j$. This proves what we want.

(*ii*) Let x be an element of $\langle A \rangle$, so that there exist a non-negative integer n, pairwise different elements $x_1, x_2, \dots, x_n$ of A, and non-zero scalars $\lambda_1, \lambda_2, \dots, \lambda_n$ in $\Bbbk$ such that $x = \sum_{i=1}^{n} \lambda_i x_i$. If $y \in A$, then $\langle x, y \rangle = \sum_{i=1}^{n} \lambda_i \langle x_i, y \rangle$: as A is orthonormal, all terms in this sum are zero if $y \notin \{x_1, x_2, \dots, x_n\}$, and this shows that the set A_x described in the statement of the proposition is contained on $\{x_1, x_2, \dots, x_n\}$ and, in particular, finite. Finally, the fact that $\lambda_i = \langle x, x_i \rangle$ for each $i \in \{1, \dots, n\}$ is now a consequence of (*i*). □

A very important special case of the previous results is that of orthonormal bases:

{corol:coefs}

Corollary 2.3.6. *Let V be an finite-dimensional inner product space, let n be the dimension of V, and let $\mathscr{B} = \{x_1, x_2, \dots, x_n\}$ be an orthonormal basis for V. For each $x \in V$ we have that*

$$x = \sum_{i=1}^{n} \langle x, x_i \rangle x_i.$$

Proof. This is just a special case of the first part of Proposition 2.3.5. □

This corollary tells us that it is very easy to write a vector as a linear combination of an orthonormal basis, as the needed coefficients can be directly computed as inner products. Of course, for this to be actually useful we need to have orthonormal bases at our disposal. The purpose of the following section is establishing the fact that finite-dimensional inner product spaces always have orthonormal bases.

§2.4. The Gram–Schmidt orthonormalization process

For the results of the previous section to be useful we have to be able to construct orthonormal bases for our finite-dimensional inner product spaces: the following proposition shows that it is very easy.

{prop:pi:gs}

Proposition 2.4.1. *Let V be an inner product space, let n be a positive integer, and let $(v_1, v_2, \dots, v_n)$ be a linearly independent sequence of elements of V of length n. We can define the vectors*

$$w_1 \coloneqq \frac{v_1}{\|v_1\|},$$

and recursively, for each i in $\{2, \dots, n\}$,

$$\tilde{w}_i \coloneqq v_i - \sum_{j=1}^{i-1} \langle v_i, w_j \rangle w_j, \qquad w_i \coloneqq \frac{\tilde{w}_i}{\|\tilde{w}_i\|},$$

and the sequence $(w_1, w_2, \dots, w_n)$ that we obtain in this way is orthonormal and has the same span in V as $(v_1, v_2, \dots, v_n)$.

We say that the sequence $(w_1, w_2, \dots, w_n)$ is obtained from the sequence $(v_1, v_2, \dots, v_n)$ by the *Gram–Schmidt orthonormalization algorithm* — the name remembers Jørgen Pedersen Gram (1850–1916, Denmark) and Erhard Schmidt (1876–1959, Germany). Let us remark that the claim made in the proposition that we can define the vectors w_1, w_2, …, w_n is really saying that each of the vectors v_1, $\tilde{w}_2$, $\tilde{w}_3$, …, $\tilde{w}_n$ is non-zero, as we can then divide them by their norms to find the vectors w_1, w_2, …, w_n.

Proof. We will prove the proposition by induction with respect to the positive integer n.

Suppose first that $n = 1$. That the sequence (v_1) is linearly independent means simply that the vector v_1 is not 0, and therefore its norm $\|v_1\|$ is not 0 and it makes sense to construct the vector $w_1 = v_1/\|v_1\|$. This is clearly a unit vector, and the span of the sequence (w_1) is obviously the same as the span of the origina sequence (v_1). This proves the proposition in this case.

Let us suppose next that $n > 1$. Since the sequence $(v_1, v_2, \dots, v_n)$ is linearly independent, so is the sequence $(v_1, v_2, \dots, v_{n-1})$ of length $n-1$. The inductive hypothesis then allows us to conclude that the sequence $(w_1, w_2, \dots, w_{n-1})$ is orthonormal and has the same span as $(v_1, v_2, \dots, v_{n-1})$. This implies that the vectors w_1, w_2, …, w_{n-1}, v_n are linearly independent, as they span the same subspace of V as the vectors v_1, v_2, …, v_{n-1}, v_n. In particular, the vector

$$\tilde{w}_n = v_n - \sum_{i=1}^{n-1} \langle v_n, w_i \rangle w_i$$

is non-zero and $\|\tilde{w}_n\| \neq 0$. We claim that $\tilde{w}_n$ is orthogonal to each of the vectors w_1, w_2, …, w_{n-1}. Indeed, if $k \in \{1, \dots, n-1\}$ we have that

$$\langle \tilde{w}_n, w_k \rangle = \left\langle v_n - \sum_{j=1}^{n-1} \langle v_n, w_j \rangle w_j, w_k \right\rangle = \langle v_n, w_k \rangle - \sum_{j=1}^{n-1} \langle v_n, w_j \rangle \langle w_j, w_k \rangle,$$

and the orthonormality of $(w_1, w_2, \dots, w_{n-1})$ implies that $\langle w_j, w_k \rangle$ is 0 if $j \neq k$ and 1 if $j = k$, so that this sum reduces to

$$\langle v_n, w_k \rangle - \langle v_n, w_k \rangle = 0.$$

Of course, it then follows from this that

$$\langle w_n, w_k \rangle = \left\langle \frac{\tilde{w}_n}{\|\tilde{w}_n\|}, w_k \right\rangle = \frac{\langle \tilde{w}_n, w_k \rangle}{\|\tilde{w}_n\|} = 0.$$

We thus see that the sequence $(w_1, w_2, \dots, w_n)$ is orthonormal. This completes the induction and proves the proposition. □

The main application of this proposition is the result we are looking for:

{coro:bon}

Corollary 2.4.2. *Every finite-dimensional inner product space has orthonormal bases.*

Proof. Let V be a finite-dimensional inner product space, let n be its dimension, and let $\mathscr{B} = (v_1, v_2, \dots, v_n)$ be an ordered basis for V. The proposition tells us that there is an orthonormal sequence

is, an ordered basis, so it is an orthonormal ordered basis for V. This proves the corollary. □

{ex:g-s:1}

Example 2.4.3. Let us consider the vector space $\mathbb{R}^3$ with its standard inner product and the ordered basis $\mathscr{B} \coloneqq (v_1, v_2, v_3)$ of $\mathbb{R}^3$, with

$$v_1 = (1,1,1), \qquad v_2 = (1,1,0), \qquad v_3 = (0,1,1).$$

We will carry out the Gram–Schmidt orthonormalization algorithm stating from $\mathscr{B}$. Since

$$\|v_1\|^2 = \langle v_1, v_2 \rangle = 3,$$

we have that

$$w_1 = \frac{v_1}{\|v_1\|} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right).$$

Next, we compute that

$$\langle v_1, w_1 \rangle = \frac{2}{\sqrt{3}},$$

$$\tilde{w}_2 = v_2 - \langle v_1, w_1 \rangle w_1 = (1,1,0) - \frac{2}{\sqrt{3}} \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right) = \left(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3} \right),$$

and

$$\|w_2\|^2 = \sqrt{\frac{2}{3}}$$

so that

$$w_2 = \frac{\tilde{w}_2}{\|w_2\|} = \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\sqrt{\frac{2}{3}}\right).$$

Finally,

$$\begin{aligned}\langle v_3, w_1\rangle &= \frac{2}{\sqrt{3}},\\ \langle v_3, w_2\rangle &= -\frac{1}{\sqrt{6}},\end{aligned}$$

so that

$$\begin{aligned}\tilde{w}_3 &= \left(-\frac{1}{2}, \frac{1}{2}, 0\right),\\ \|\tilde{w}_3\| &= \frac{1}{\sqrt{2}},\end{aligned}$$

and

$$w_3 = \frac{w_3}{\|\tilde{w}_3\|} = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right).$$

We thus see that the orthonormal ordered basis that we can construct from $\mathscr{B}$ applying the Gram–Schmidt orthonormalization procedure is that of the three vectors

$$\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \qquad \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\sqrt{\frac{2}{3}}\right), \qquad \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right).$$

This example shows that as we move along the Gram–Schmidt orthonormalization procedure the entries of the vectors with which we deal get messier and messier. This is due, of course, to the normalization which we apply each time to go from the vector $\tilde{w}_i$ to the vector w_i. Sometimes it is more convenient to leave all this normalizations to the end. The resulting procedure is described in the following exercise.

Exercise 2.4.4. Let V be an inner product space, let n be a positive integer, let $(v_1, v_2, \dots, v_n)$ be a linearly independent sequence of elements of V of length n, and define recursively the vectors

$$\begin{aligned}u_1 &\coloneqq v_1,\\ u_2 &\coloneqq v_2 - \frac{\langle v_2, u_1\rangle}{\langle u_1, u_1\rangle}u_1,\\ u_3 &\coloneqq v_3 - \frac{\langle v_3, u_1\rangle}{\langle u_1, u_1\rangle}u_1 - \frac{\langle v_3, u_2\rangle}{\langle u_2, u_2\rangle}u_2,\end{aligned}$$

and, more generally, for each $i \in \{2, \dots, n\}$,

$$u_i \coloneqq v_1 - \sum_{j=1}^{i-1} \frac{\langle v_i, u_j \rangle}{\langle u_j, u_j \rangle} u_j.$$

Prove that the sequence $(u_1, u_2, \dots, u_n)$ is orthogonal, does not contain 0 and spans the same subspace as $(v_1, v_2, \dots, v_n)$, so that the sequence $(u_1/\|u_1\|, u_2/\|u_2\|, \dots, u_n/\|u_n\|)$ is an orthonormal ordered basis for that subspace. In fact, this last orthonormal ordered basis coincides with the one that the Gram–Schmidt procedure constructs.

Example 2.4.5. Let us show how the algorithm described in the exercise works in the situation of Example 2.4.3, so that we work with the vector space $\mathbb{R}^3$ endowed with its standard inner product and the ordered basis $\mathscr{B} \coloneqq (v_1, v_2, v_3)$ with

$$v_1 = (1,1,1), \qquad v_2 = (1,1,0), \qquad v_3 = (0,1,1).$$

Now we have that $u_1 = v_1 = (1,1,1)$, that $\langle u_1, u_1 \rangle = 3$ and $\langle v_2, u_1 \rangle = 2$, so that

$$u_2 = v_2 - \frac{\langle v_2, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 = \left(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3} \right).$$

Finally, we have that $\langle u_2, u_2 \rangle = 2/3$, $\langle v_3, u_1 \rangle = 2$ and $\langle v_3, u_2 \rangle = -1/3$, so that

$$u_3 = v_3 - \frac{\langle v_3, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 - \frac{\langle v_3, u_2 \rangle}{\langle u_2, u_2 \rangle} u_2 = \left(-\frac{1}{2}, \frac{1}{2}, 0 \right).$$

We have therefore obtained an orthogonal ordered basis (u_1, u_2, u_3) with vectors

$$(1,1,1), \qquad \left(\frac{1}{3}, \frac{1}{3}, -\frac{2}{3} \right), \qquad \left(-\frac{1}{2}, \frac{1}{2}, 0 \right).$$

If we normalize them to obtain an ortho*normal* basis, we will obtain the same one that we found in Example 2.4.3.

Example 2.4.6. Let us consider the vector space $\mathbb{R}[X]_{\leq 4}$ of real polynomials of degree at most 4 endowed with the inner product $\langle -, - \rangle$ such that

$$\langle p, q \rangle = \int_{-1}^{1} p(x) q(x) \, \mathrm{d}x$$

for all p and q in $\mathbb{R}[X]_{\leq 4}$, and carry out the modified Gram–Schmidt orthonormalization procedure starting from the basis $\mathscr{B} = (1, X, X^2, X^3, X^4)$ of monic monomials ordered by degree. Let us write $v_i = X^i$ for each $i \in \{0, \dots, 4\}$.

- We put
  $$u_0 \coloneqq v_0 = 1,$$
  so
  $$\langle u_0, u_0 \rangle = \int_{-1}^{1} 1 \cdot 1 \,\mathrm{d}x = 2.$$
- We have
  $$\langle v_1, u_0 \rangle = \int_{-1}^{1} x \cdot 1 \,\mathrm{d}x = 0,$$
  so we set
  $$u_1 \coloneqq v_1 - \frac{\langle v_1, u_0 \rangle}{\langle u_0, u_0 \rangle} u_0 = v_1 = X.$$
  We then have that
  $$\langle u_1, u_1 \rangle = \int_{-1}^{1} x \cdot x \,\mathrm{d}x = \frac{2}{3}.$$
- As
  $$\langle v_2, u_0 \rangle = \int_{-1}^{1} x^2 \cdot 1 \,\mathrm{d}x = \frac{2}{3},$$
  $$\langle v_2, u_1 \rangle = \int_{-1}^{1} x^2 \cdot x \,\mathrm{d}x = 0$$
  we put
  $$u_2 \coloneqq v_2 - \frac{\langle v_2, u_0 \rangle}{\langle u_0, u_0 \rangle} u_0 - \frac{\langle v_2, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 = X^2 - \frac{1}{3},$$
  and thus
  $$\langle u_2, u_2 \rangle = \int_{-1}^{1} \left(x^2 - \frac{2}{3} \right)^2 \,\mathrm{d}x = \frac{8}{45}.$$
- Next, we have that
  $$\langle v_3, u_0 \rangle = \int_{-1}^{1} x^3 \cdot 1 \,\mathrm{d}x = 0,$$
  $$\langle v_3, u_1 \rangle = \int_{-1}^{1} x^3 \cdot x \,\mathrm{d}x = \frac{2}{5},$$
  $$\langle v_3, u_2 \rangle = \int_{-1}^{1} x^3 \cdot \left(x^2 - \frac{2}{3} \right) \,\mathrm{d}x = 0,$$
  so we set
  $$u_3 \coloneqq v_3 - \frac{\langle v_3, u_0 \rangle}{\langle u_0, u_0 \rangle} u_0 - \frac{\langle v_3, u_1 \rangle}{\langle u_1, u_1 \rangle} u_1 - \frac{\langle v_3, u_2 \rangle}{\langle u_2, u_2 \rangle} u_2 = x^3 - \frac{3}{5} x,$$
  and therefore
  $$\langle u_3, u_3 \rangle = \int_{-1}^{1} \left(x^3 - \frac{3}{5} x \right)^2 \,\mathrm{d}x = \frac{8}{175}.$$

- Finally, we compute that

$$\begin{aligned}\langle v_4, u_0\rangle &= \int_{-1}^{1} x^4 \cdot 1 \,\mathrm{d}x = \frac{2}{5},\\ \langle v_4, u_1\rangle &= \int_{-1}^{1} x^4 \cdot x \,\mathrm{d}x = 0,\\ \langle v_4, u_2\rangle &= \int_{-1}^{1} x^4 \cdot \left(x^2 - \frac{2}{3}\right) \mathrm{d}x = 0,\\ \langle v_4, u_4\rangle &= \int_{-1}^{1} x^4 \cdot \left(x^3 - \frac{3}{5}x\right) \mathrm{d}x = \frac{16}{105},\end{aligned}$$

so our last polynomial is

$$u_4 \coloneqq v_4 - \frac{\langle v_4, u_0\rangle}{\langle u_0, u_0\rangle} u_0 - \frac{\langle v_4, u_1\rangle}{\langle u_1, u_1\rangle} u_1 - \frac{\langle v_4, u_2\rangle}{\langle u_2, u_2\rangle} u_2 - \frac{\langle v_4, u_3\rangle}{\langle u_3, u_3\rangle} u_3 = X^4 - \frac{6}{7}X^2 + \frac{3}{35},$$

which has

$$\langle u_4, u_4\rangle = \int_{-1}^{1} \left(x^4 - \frac{6}{7}x^2 + \frac{3}{35}\right)^2 \mathrm{d}x = \frac{128}{11025}.$$

We thus obtain an orthogonal basis $(u_0, u_1, u_2, u_3, u_4)$ for our vector space, and normalizing its elements an orthonormal ordered basis $(q_0, q_1, q_2, q_3, q_4)$.

i	0	1	2	3	4
u_i	1	X	$X^2 - \frac{1}{3}$	$X^3 - \frac{3}{5}X$	$X^4 - \frac{6}{7}X^2 + \frac{3}{35}$
q_i	$\frac{1}{\sqrt{2}}$	$\sqrt{\frac{3}{2}}X$	$\frac{1}{2}\sqrt{\frac{5}{2}}\left(3X^2 - 1\right)$	$\frac{1}{2}\sqrt{\frac{7}{2}}\left(5X^3 - 3X\right)$	$\frac{3}{8\sqrt{2}}\left(35X^4 - 30X^2 + 3\right)$

The polynomials that we obtain in this way are scalar multiples of the so called ***Legendre polynomials*** — named after Adrien-Marie Legendre (France, 1752–1833), who first studied them in connection with his study of the Newtonian gravitation potential function — which have innumerable applications both in mathematics and in the natural and applied sciences.

§2.5. Orthogonal approximation

We saw earlier how an inner product on a vector space allows us to measure the distance between any two points in that vector space. We will now generalize this a bit: we now want to measure the distance from a vector to a set. Let V be an inner product space. If x is a vector in V and S is a non-empty subset of V, then the ***distance*** from x to S is the number

$$d(v, S) \coloneqq \inf\{d(x, y) : y \in S\}.$$

Let us remark that this makes sense: the set $\{d(x, y) : y \in S\}$ is a non-empty subset of $[0, +\infty)$, so it has a well-determined infimum.

The following proposition shows how to compute this distance in the special case in which the set S is a subspace of V.

{prop:vS}

Proposition 2.5.1. *Let V be an inner product space, let S be a finite-dimensional subspace of V, and let $x \in V$. If $\mathscr{B} = (x_1, x_2, \dots, x_n)$ is an ordered orthonormal basis for S and*

$$x_S \coloneqq \sum_{j=1}^{n} \langle x, x_j \rangle x_j,$$

then

(i) $x - x_S \perp S$,

(ii) $d(x, S) = d(x, x_S)$,

(iii) $d(x, x_S) < d(x, y)$ *for all* $y \in S \setminus \{x_S\}$, *and*

(iv) $d(x, S)^2 = \langle x, x \rangle - \sum_{i=1}^{n} |\langle x, x_i \rangle|^2.$

This proposition tells us that the vector x_S is an element of S whose distance to x is equal to $d(x, S)$ and that it is moreover the only point of S with that property. We call the vector x_S the ***orthogonal projection*** of x onto the subspace S.

Proof. Let $y \in S$. As $\mathscr{B}$ is an orthonormal basis for S, we have $y = \sum_{i=1}^{n} \langle y, x_i \rangle x_i$ and therefore

$$\begin{aligned}
\langle x - x_S, y \rangle &= \Big\langle x - \sum_{j=1}^{n} \langle x, x_j \rangle x_j, \sum_{i=1}^{n} \langle y, x_i \rangle x_i \Big\rangle \\
&= \sum_{i=1}^{n} \overline{\langle y, x_i \rangle} \langle x, x_i \rangle - \sum_{j=1}^{n} \sum_{i=1}^{n} \langle x, x_j \rangle \overline{\langle y, x_i \rangle} \langle x_j, x_i \rangle \\
&= \sum_{i=1}^{n} \overline{\langle y, x_i \rangle} \langle x, x_i \rangle - \sum_{j=1}^{n} \langle x, x_j \rangle \overline{\langle y, x_j \rangle} = 0.
\end{aligned}$$

This tells us that $x - x_S \perp S$.

As $x_S \in S$ it is clear that $d(x, S) \leq d(x, x_S)$. If, on the other hand, we have a vector y in S, then

$$\begin{aligned}
d(x, y)^2 &= \|x - y\|^2 \\
&= \|(x - x_S) + (x_S - y)\|^2
\end{aligned}$$

and, since $x - x_S \perp x_S - y$ because $x_S - y \in S$, this is

$$
\begin{aligned}
&= \|x - x_S\|^2 + \|x_S - y\|^2 \\
&\geq \|x - x_S\|^2,
\end{aligned}
$$

so that $d(x, y) \geq d(x, x_S)$. This tells us that $d(x, S) \geq d(x, x_S)$, and proves the equality (*ii*).

If $y \in S \smallsetminus \{x_S\}$, then $d(x_S, y) > 0$ and, just as before,

$$d(x, y)^2 = d(x, x_S)^2 + d(x_S, y)^2 > d(x, x_S)^2,$$

so that $d(x, y) > d(x, x_S)$. Finally,

$$
\begin{aligned}
d(x, S)^2 = d(x, x_S)^2 = \|x - x_S\|^2 &= \left\langle x - \sum_{i=1}^n \langle x, x_i\rangle x_i, x - \sum_{j=1}^n \langle x, x_j\rangle x_j \right\rangle \\
&= \langle x, x\rangle - \sum_{j=1}^n \big\langle x, \langle x, x_j\rangle x_j\big\rangle - \sum_{i=1}^n \big\langle \langle x, x_i\rangle x_i, x\big\rangle + \sum_{i=1}^n \sum_{j=1}^n \big\langle \langle x, x_i\rangle x_i, \langle x, x_j\rangle x_j\big\rangle \\
&= \langle x, x\rangle - \sum_{j=1}^n \overline{\langle x, x_j\rangle} \langle x, x_j\rangle - \sum_{i=1}^n \langle x, x_i\rangle \langle x_i, x\rangle + \sum_{i=1}^n \sum_{j=1}^n \langle x, x_i\rangle \overline{\langle x, x_j\rangle} \langle x_i, x_j\rangle.
\end{aligned}
$$

Since the ordered basis $(x_1, x_2, \dots, x_n)$ is orthonormal, this is the same as

$$\langle x, x\rangle - \sum_{j=1}^n \overline{\langle x, x_j\rangle} \langle x, x_j\rangle - \sum_{i=1}^n \langle x, x_i\rangle \langle x_i, x\rangle + \sum_{i=1}^n \langle x, x_i\rangle \overline{\langle x, x_i\rangle}.$$

The first and third sums that appear here cancel each other, and therefore this expression has the same value as

$$\langle x, x\rangle - \sum_{i=1}^n \langle x, x_i\rangle \langle x_i, x\rangle = \langle x, x\rangle - \sum_{i=1}^n \langle x, x_i\rangle \overline{\langle x, x_i\rangle} = \langle x, x\rangle - \sum_{i=1}^n |\langle x, x_i\rangle|^2.$$

The equality that appears in the last part of the property is therefore true. □

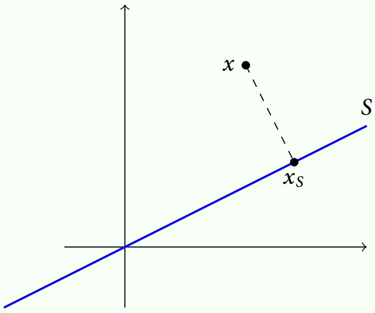
Let us consider two simple examples of this result.

Example 2.5.2. Let us consider the vector space $\mathbb{R}^2$ with its usual inner product, its 1-dimensional subspace S spanned by the vector $v = (2, 3)$, and the vector $x = (0, 3)$. Normalizing v we obtain the unit vector $w = (2/\sqrt{5}, 1/\sqrt{5})$, and the one-element sequence $\mathscr{B} = (w)$ is therefore an orthonormal basis for S. According to the proposition, the point

$$x_S \coloneqq \langle x, w\rangle w = \left(\frac{14}{5}, \frac{7}{5}\right)$$

is the point of S at the minimum distance from x, and that minimum distance is

$$d(x, S) = d(x, x_S) = \|x - x_S\| = \frac{4}{\sqrt{5}}.$$



Example 2.5.3. Let us now consider the vector space $\mathbb{R}[X]_{\leq 4}$ of all real polynomials of degree at most 4 endowed with the inner product $\langle -, - \rangle : \mathbb{R}[X]_{\leq 4} \times \mathbb{R}[X]_{\leq 4} \to \mathbb{R}$ that on each pair of polynomials p and q takes the value

$$\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)\,\mathrm{d}x.$$

Let us consider the subspace $S \coloneqq \{1, X, X^2\}$ of $\mathbb{R}[X]_{\leq 2}$, whose elements are the polynomials of degree at most 1. Of course, $\mathscr{B} = (1, X, X^2)$ is an ordered basis for S, and using the Gram–Schmidt orthonormalization procedure we can find immediately that $\mathscr{B}' = (p_0, p_1, p_2)$, with

$$p_0 = \frac{1}{\sqrt{2}}, \qquad p_1 = \sqrt{\frac{3}{2}}X, \qquad p_2 = \frac{1}{2}\sqrt{\frac{5}{2}}\left(3X^2 - 1\right),$$

is an orthonormal ordered basis for S.

Let $q = aX^4 + bX^3 + cX^2 + dX + e$ be an arbitrary element of $\mathbb{R}[X]_{\leq 4}$. We can compute that

$$\begin{aligned}
\langle q, p_0 \rangle &= \int_{-1}^{1} \left(ax^4 + bx^3 + cx^2 + dx + e\right) \cdot \frac{1}{\sqrt{2}}\,\mathrm{d}x = \sqrt{2}\left(\frac{a}{5} + \frac{c}{3} + e\right), \\
\langle q, p_1 \rangle &= \int_{-1}^{1} \left(ax^4 + bx^3 + cx^2 + dx + e\right) \cdot \sqrt{\frac{3}{2}}X\,\mathrm{d}x = \sqrt{\frac{2}{3}}\left(\frac{3}{5}b + d\right), \\
\langle q, p_2 \rangle &= \int_{-1}^{1} \left(ax^4 + bx^3 + cx^2 + dx + e\right) \cdot \frac{1}{2}\sqrt{\frac{5}{2}}\left(3X^2 - 1\right)\,\mathrm{d}x = \sqrt{\frac{2}{5}}\left(\frac{4}{7}a + \frac{2}{3}c\right).
\end{aligned}$$

It follows from this that the orthogonal projection of the polynomial q onto the subspace S is

$$q_S = \sqrt{2}\left(\frac{a}{5} + \frac{c}{3} + e\right) \cdot p_0 + \sqrt{\frac{2}{3}}\left(\frac{3}{5}b + d\right) \cdot p_1 + \sqrt{\frac{2}{5}}\left(\frac{4}{7}a + \frac{2}{3}c\right) \cdot p_2.$$

§2.6. Orthogonal complements

Let V be an inner product space. If S is an arbitrary subset of V, then the ***orthogonal complement*** of S is the subset

$$S^\perp \coloneqq \{x \in V : x \perp s \text{ for all } s \in S\}.$$

The following proposition describes the basic properties of this construction.

{prop:perp}

Proposition 2.6.1. *Let V be an inner product space.*

(*i*) *If S is a subset of V, then $S^\perp$ is a subspace of V.*

(*ii*) *We have that $0^\perp = V$ and $V^\perp = 0$.*

{prop-part:perp:rev

(*iii*) *If S and T are two subsets of V such that $S \subseteq T$, then $T^\perp \subseteq S^\perp$.*

(*iv*) *If S is a subset of V, then $\langle S \rangle^\perp = S^\perp$.*

{prop:perp:cap}

(*v*) *If S and T are two subspaces of V, then $(S+T)^\perp = S^\perp \cap T^\perp$.*

Proof. (*i*) Let x and y be two elements of $S^\perp$, and let a and b be two scalars in $\Bbbk$. For each $s \in S$ we have that

$$\langle ax + by, s\rangle = a\langle x, s\rangle + b\langle y, s\rangle = 0,$$

because both summands are zero. This tells us that $ax + by \in S^\perp$. As $S^\perp$ is not empty, since $0 \in S^\perp$, then this implies that $S^\perp$ is a subspace of V.

(*ii*) Every vector is orthogonal to 0, so that $V \subseteq 0^\perp$. The first equality in the statement is therefore immediate. On the other hand, if $x \in V^\perp$, then $\langle x, y\rangle = 0$ for all $y \in V$ and Proposition 2.1.6 tells us that $x = 0$. Since of course $0 \in V^\perp$, this proves the second equality in the statement.

(*iii*) Let S and T be two subsets of V such that $S \subseteq T$. If y is an element of $T^\perp$, then for each $s \in S$ we have that $y \perp s$, since $s \in T$, and therefore $x \in S^\perp$. We thus have that $T^\perp \subseteq S^\perp$, as we want.

(*iv*) Since $S \subseteq \langle S \rangle$, part (*iii*) of the proposition implies that $\langle S\rangle^\perp \subseteq S^\perp$. Let, on the other hand, x and y be elements of $S^\perp$ and of $\langle S \rangle$, respectively. There exist then a non-negative integer n, elements $s_1, s_2, \dots, s_n$ of S, and scalars $\lambda_1, \lambda_2, \dots, \lambda_n$ in $\Bbbk$ such that $y = \sum_{i=1}^n \lambda_i s_i$ and, as a consequence of this,

$$\langle x, y\rangle = \left\langle x, \sum_{i=1}^n \lambda_i s_i \right\rangle = \sum_{i=1}^n \overline{\lambda}_i \langle x, s_i\rangle = 0$$

because $x \in S^\perp$. This proves that $S^\perp \subseteq \langle S\rangle^\perp$.

(*v*) Let now S and T be two subspaces of V. As S and T are contained in $S+T$, part (*ii*) of the proposition tells us that $(S+T)^\perp \subseteq S^\perp$ and $(S+T)^\perp \subseteq T^\perp$, so that $(S+T)^\perp \subseteq S^\perp \cap T^\perp$. On the other hand, if $x \in S^\perp \cap T^\perp$ and $y \in S + T$, so that there exist $s \in S$ and $t \in T$ such that $y = s + t$, then

$$\langle x, y\rangle = \langle x, s+t\rangle = \langle x, s\rangle + \langle x, t\rangle = 0.$$

We see with this that $x \in (S+T)^\perp$ and, in conclusion, that $S^\perp \cap T^\perp \subseteq (S+T)^\perp$. □

When S is a subset of an inner product space, we write $S^{\perp\perp} \coloneqq (S^{\perp})^{\perp}$ and $S^{\perp\perp\perp} \coloneqq (S^{\perp\perp})^{\perp}$. It should be noticed that $S^{\perp\perp\perp}$ is the same set as $(S^{\perp})^{\perp\perp}$.

{prop:perp2}
{prop:perp2:2}

Proposition 2.6.2. *Let V be an inner product space. If S is a subset of V, then*

(*i*) $S \cap S^{\perp} \subseteq 0$,

(*ii*) $S \subseteq S^{\perp\perp}$, *and*

(*iii*) $S^{\perp} = S^{\perp\perp\perp}$.

Proof. (*i*) If $x \in S \cap S^{\perp}$, then $\langle x, x \rangle = 0$, since $x \in S$ and $x \in S^{\perp}$, and therefore $x = 0$.

(*ii*) If $s \in S$, then for all $t \in S^{\perp}$ we have that $s \perp t$. This tells us that $s \in (S^{\perp})^{\perp} = S^{\perp\perp}$.

(*iii*) We have just proved that $S \subseteq S^{\perp\perp}$, so part (*iii*) of Proposition 2.6.1 implies that $S^{\perp\perp\perp} = (S^{\perp\perp})^{\perp} \subseteq S^{\perp}$. On the other hand, part (*ii*) tells us that also $S^{\perp} \subseteq (S^{\perp})^{\perp\perp} = S^{\perp\perp\perp}$. □

In the situation of Proposition (*ii*) in general the inclusion is strict. The equality does hold, though, in a very important case:

{prop:perp2:fin}

Proposition 2.6.3. *Let V be an inner product space. If S is a finite-dimensional subspace of V, then $V = S \oplus S^{\perp}$ and $S = S^{\perp\perp}$.*

Proof. Let $x \in V$. According to Proposition 2.5.1, there is an $x_S \in S$ such that $x - x_S \perp S$, that is, such that $x - x_S \in S^{\perp}$. We then have that $x = x_S + (x - x_S) \in S + S^{\perp}$ and therefore that $V = S + S^{\perp}$. As we know that also $S \cap S^{\perp} = 0$, we can conclude that $V = S \oplus S^{\perp}$.

In order to prove that $S = S^{\perp\perp}$ it is enough, in view of Proposition (*ii*), to show that $S^{\perp\perp} \subseteq S$. Let then x be an element of $S^{\perp\perp}$. As $V = S \oplus S^{\perp}$, there exist $s \in S$ and $t \in S^{\perp}$ such that $x = s + t$. Since $x \in S^{\perp\perp}$, we have that

$$0 = \langle x, t \rangle = \langle s + t, t \rangle = \langle s, t \rangle + \langle t, t \rangle = \langle t, t \rangle,$$

and therefore $t = 0$: this implies that $x = s + t = s \in S$. □

An important corollary of this is obtained simply by taking dimensions:

Corollary 2.6.4. *Let V be a finite-dimensional inner product space. If $S \subseteq V$ is a subspace, then*

$$\dim V = \dim S + \dim S^{\perp}.$$

Proof. According to the proposition we have that $V = S \oplus S^{\perp}$, and taking dimensions we immediately obtain the equality in the corollary. □

Before continuining we want to exhibit an example of a subspace S of an inner product space V such that $S^{\perp\perp}$ is different from S, in order to show that the inclusion asserted by part (*ii*) of Proposition (*i*) can be strict. Of course, in view of Proposition 2.6.3, such a subspace S must necessary be infinite-dimensional.

Example 2.6.5. Let us consider the vector space $\mathscr{F}_0(\mathbb{N})$ of all functions $\mathbb{N}_0 \to \Bbbk$ with finite support endowed with the inner inner product that we described in Example 2.1.4. If $f : \mathbb{N}_0 \to \Bbbk$ is an element of $\mathscr{F}_0(\mathbb{N}_0)$, then the support $\sigma(f)$ is a finite subset of $\mathbb{N}_0$ and we can therefore consider the scalar {exam:perppperp}

$$\phi(f) \coloneqq \sum_{n\in\sigma(f)} f(n).$$

In this way we obtain a function $\phi : \mathscr{F}_0(\mathbb{N}_0) \to \Bbbk$. We remark that

if f is an element of $\mathscr{F}_0(\mathbb{N}_0)$ and T is a finite subset of $\mathbb{N}_0$ such that $\sigma(f) \subseteq T$, then $\phi(f) = \sum_{n\in T} f(n)$. (2.6) {eq:sums}

Indeed, in that situation we have that

$$\sum_{n\in T} f(n) = \sum_{n\in\sigma(f)} f(n) + \sum_{n\in T\smallsetminus\sigma(f)} f(n)$$

because T is the disjoint union of $\sigma(f)$ and $T \smallsetminus \sigma(f)$, and this is

$$= \sum_{n\in\sigma(f)} f(n) = \phi(f)$$

because clearly $f(n) = 0$ for all $n \in T \smallsetminus \sigma(f)$.

Let $f : \mathbb{N}_0 \to \Bbbk$ and $g : \mathbb{N}_0 \to \Bbbk$ be two elements of $\mathscr{F}_0(\mathbb{N}_0)$, and let a and b be two elements of $\Bbbk$. As we noted in Example 2.1.4, we have that $\sigma(af + bg) \subseteq \sigma(f) \cup \sigma(g)$, and therefore according to (2.6)

$$\begin{aligned}
\phi(af + gb) &= \sum_{n\in\sigma(f)\cup\sigma(g)} (af + bg)(n) \\
&= \sum_{n\in\sigma(f)\cup\sigma(g)} \big(af(n) + bg(n)\big) \\
&= a \sum_{n\in\sigma(f)\cup\sigma(g)} f(n) + b \sum_{n\in\sigma(f)\cup\sigma(g)} g(n)
\end{aligned}$$

and since $\sigma(f)$ and $\sigma(g)$ are of course contained in $\sigma(f) \cup \sigma(g)$ that same observation tells us that this is

$$= a \sum_{n\in\sigma(f)} f(n) + b \sum_{n\in\sigma(g)} g(n) = a\phi(f) + b\phi(g).$$

This tells us that the function $\phi : \mathscr{F}_0(\mathbb{N}_0) \to \Bbbk$ is linear and, in particular, that

$$S \coloneqq \ker \phi = \{f \in \mathscr{F}_0(\mathbb{N}_0) : \phi(f) = 0\}$$

is a subspace of $\mathscr{F}_0(\mathbb{N}_0)$. In fact, it is a *proper* subspace. for example, the function $u : \mathbb{N}_0 \to \Bbbk$ such that

$$u(n) = \begin{cases} 1 & \text{if } n = 0; \\ 0 & \text{otherwise} \end{cases}$$

has $\sigma(u) = \{0\}$, so that $u \in \mathscr{F}_0(\mathbb{N}_0)$, and

$$\phi(u) = \sum_{n \in \sigma(u)} u(n) = u(0) = 1,$$

so that $u \notin S$.

We claim that

$$S^\perp = 0. \tag{2.7}$$ {eq:sperp}

Assuming that, then we see that $S^{\perp\perp} = 0^\perp = \mathscr{F}_0(\mathbb{N}_0)$ is different from S, and that we have the example that we were looking for.

In order to prove (2.7) we fix an element $f : \mathbb{N}_0 \to \Bbbk$ of $S^\perp$ and prove that it is necessarily the zero function. Let k be a positive integer, and let $g : \mathbb{N}_0 \to \Bbbk$ be the function that for each $n \in \mathbb{N}_0$ has

$$g(n) = \begin{cases} 1 & \text{if } n = 0; \\ -1 & \text{if } n = k; \\ 0 & \text{in any other case.} \end{cases}$$

It is clear that $\sigma(g) = \{0, k\}$, a finite subset of $\mathbb{N}_0$, so that g belongs to $\mathscr{F}_0(\mathbb{N}_0)$, and moreover that

$$\phi(g) = \sum_{n \in \sigma(g)} g(n) = g(1) + g(k) = 0,$$

so that g is an element of the subspace S. As f belongs to $S^\perp$, we have that

$$0 = \langle f, g \rangle = \sum_{n \in \sigma(f) \cap \sigma(g)} f(n) \cdot \overline{g(n)} = f(0) \cdot \overline{g(0)} + f(k) \cdot \overline{g(k)} = f(0) - f(k),$$

and therefore $f(k) = f(0)$. We can therefore conclude from this that

$$f(k) = f(0) \text{ for all } k \in \mathbb{N}_0.$$

If the number $f(0)$ was different from zero, then this would tell us that $f(k) \neq 0$ for all $k \in \mathbb{N}_0$, so that $\sigma(f) = \mathbb{N}_0$, and this is absurd, as f belongs to $\mathscr{F}_0(\mathbb{N}_0)$. We thus see that we must have $f(0) = 0$ and therefore that in fact $f(k) = 0$ for all $k \in \mathbb{N}_0$, so that $f = 0$, as we wanted.

§2.7. Orthogonal projections

An endomorphism $p : V \to V$ of an vector space V is a ***projection*** if $p^2 = p$. The identity map $\mathrm{id}_V : V \to V$ and the zero map $0_V : V \to V$ are clearly projections: we call them the ***trivial projections***. The most basic two properties of projections are given by the following lemma.

Lemma 2.7.1. *Let V be a vector space. If $p : V \to V$ is a projection, then*

(i) a vector $x \in V$ is in the image of p if and only if $x = p(x)$, and {prop:pi:proy:1}

(ii) there is a decomposition $V = \operatorname{img} p \oplus \ker p$. {prop:pi:proy:2}

Proof. Let $p : V \to V$ be a projection.

(i) If x is in the image of p, so that there is a $y \in V$ such that $p(y) = x$, then $p(x) = p^2(x) = p(y) = x$. Conversely, it is clear that if $p(x) = x$ then x is in $\operatorname{img} p$.

(ii) If $x \in V$, then $p(x - p(x)) = p(x) - p^2(x) = 0$, so $x - p(x) \in \ker p$ and

$$x = p(x) + (x - p(x)) \in \operatorname{img} p + \ker p.$$

We see in this way that $V = \operatorname{img} p + \ker p$. This sum is direct: if $x \in \operatorname{img} p \cap \ker p$, then because $x \in \operatorname{img} p$ the first part tells us that $x = p(x)$ and, because $x \in \ker p$, this implies that $x = 0$. □

The second part of this lemma tells us that a projection $p : V \to V$ determines a direct sum decomposition $V = \operatorname{img} p \oplus \ker p$ of its domain V. Our next result can be viewed as a converse of that statement: it tells us that any direct sum decomposition $V = S \oplus T$ of a vector space V as a direct sum of two of its subspaces arises in that way from a projection.

Proposition 2.7.2. *Let V be a vector space and let S be a subspace of V. If T is a complement of S in V, then there exists exactly one projection $p : V \to V$ such that $\operatorname{img} p = S$ and $\ker p = T$.* {prop:proy:ik}

Proof. Let us suppose that T is a complement of S in V, so that $V = S \oplus T$. There is a function $p : V \to V$ such that whenever $x \in V$ and $s \in S$ and $t \in T$ are the unique vectors in S and T, respectively, such that $x = s + t$ has $p(x) = s$. Let us show that p is a projection that satifsied the conditions in the statement.

- Let x and y be vectors in V and let $a, b \in \Bbbk$ be scalars. If $s_1, s_2 \in S$ and $t_1, t_2 \in T$ are such that $x = s_1 + t_1$ and $y = s_2 + t_2$, then $p(x) = s_1$, $p(y) = s_2$ and, since

  $$ax + by = (as_1 + bs_2) + (at_1 + bt_2)$$

  with $as_1 + bs_2 \in S$ and $at_1 + bt_s \in T$,

  $$p(ax + by) = as_1 + bs_2 = ap(x) + bp(y).$$

This tells us that the function p is linear.

- If $x \in V$, $s \in S$ and $t \in T$ are such that $x = s + t$, then $p(x) = s \in S$: it follows from this that $\operatorname{img} p \subseteq S$. On the other hand, if $s \in S$, then it is clear that $p(s) = s$ and, therefore, that $S \subseteq \operatorname{img} p$. We see in this way that $\operatorname{img} p = S$ and that $p^2 = p$, so that p is a projection with image equal to S.
- Finally, if $x \in V$ is an element of $\ker p$ and $s \in S$ and $t \in T$ are such that $x = s + t$, then $0 = p(x) = s$: this tells us that $x = t \in T$ and, in consequence, that $\ker p \subseteq T$. Conversely, if $t \in T$ then the definition of p implies immediately that $p(t) = 0$: we thus have that $\ker p = T$.

Let us suppose now that $q : V \to V$ is another projector such that $\operatorname{img} q = S$ and $\ker q = T$. If $x \in V$ and $s \in S$ and $t \in T$ are such that $x = s + t$, then since $s \in \operatorname{img} q$, from Proposition (*i*) we know that $s = q(s)$; on the other hand, since $t \in \ker q$ we have that $q(t) = 0$. It follows from this that $q(x) = q(s) + q(t) = s = p(s)$. We thus see that $q = p$, and this proves the uniqueness claimed by the proposition. □

A consequence of this proposition is that we usually have many projections on a vector space. For example, we have the following result:

Corollary 2.7.3. *Every subspace of a finite-dimensional vector space V is the image of a projection $p : V \to V$.*

In fact, the hypothesis that the vector space V be finite-dimensional is not necessary for this to be true. We will not prove this here.

Proof. Let V be a finite-dimensional vector space and let S be a subspace of V. We know that there exists a subspace T of V such that $V = S \oplus T$, and Proposition 2.7.2 tells us that there is a projection $p : V \to V$ such that $\operatorname{img} p = S$ and $\ker p = T$. This proves the corollary. □

It should be noticed that, in fact, a subspace S of a vector space V is, in general, the image of *many* projections $p : V \to V$, for unless S is 0 or V there are many subspaces T of V such that $V = S \oplus T$.

{exam:pt}

Example 2.7.4. Let us consider the vector space $\mathbb{R}^2$ and its subspace $S = \langle(1,0)\rangle$. For every $t \in \mathbb{R}$ the subspace $T_t := \langle(t,1)\rangle$ is such that $\mathbb{R}^2 = S \oplus T_t$, and the linear map

$$p_t : (x,y) \in \mathbb{R}^2 \mapsto (x - ty, 0) \in \mathbb{R}^2$$

can be easily seen to be a projection such that $\operatorname{img} p_t = S$ and $\ker p_t = T_t$. Moreover, if $p : \mathbb{R}^2 \to \mathbb{R}^2$ is a projection such that $\operatorname{img} p = S$, then there exists exactly one scalar $t \in \Bbbk$ such that $p = p_t$.

Exercise 2.7.5. Verify all the claims made in Example 2.7.4

Let V be now an inner product space. We say that a projection $p : V \to V$ is ***orthogonal*** if $\operatorname{img} p \perp \ker p$.

{prop:op}

Proposition 2.7.6. *Let V be an inner product space. A projection $p : V \to V$ is orthogonal if and only if $(\operatorname{img} p)^\perp = \ker p$ and $(\ker p)^\perp = \operatorname{img} p$.*

Proof. Let $p : V \to V$ be a projection. If the condition of the proposition is satisfied, then $\ker p = (\operatorname{img} p)^\perp$ and, of course, we thus have that $\operatorname{img} p \perp \ker p$. That condition is therefore sufficient.

To show that it is also necessary, let us suppose that the projection p is orthogonal. We then have that $\operatorname{img} p \perp \ker p$ and thus that $\operatorname{img} p \subseteq \ker p^\perp$ and $\ker p \subseteq \operatorname{img} p^\perp$. We will show that in fact the two equalities hold. Let x be an element of $(\ker p)^\perp$. As p is a projection, we know that $V = \operatorname{img} p \oplus \ker p$, so that we can find $y \in \operatorname{img} p$ y $z \in \ker p$ be such that $x = y + z$. We then have that

$$0 = \langle x, z\rangle = \langle y + z, z\rangle = \langle y, z\rangle + \langle z, z\rangle,$$

and the first summand in this last expression is zero because $\operatorname{img} p \perp \ker p$. We see that $\langle z, z\rangle = 0$, so that $z = 0$ and $x = x + y = y \in \operatorname{img} p$. We can conclude from this that $(\ker p)^\perp \subseteq \operatorname{img} p$.

In exactly the same way we can show that $(\operatorname{img} p)^\perp \subseteq \ker p$. □

Exercise 2.7.7. Complete the proof of Proposition 2.7.6.

We proved above that every finite-dimensional subspace of a vector space is the image of a projection, and that in fact in most cases there are many projections that have it as image. If we restrict ourselves to orthogonal projections this changes:

{prop:proyector}

Proposition 2.7.8. *Let V be an inner product space. If S is a finite-dimensional subspace of V, then there exists exactly one orthogonal projection $p : V \to V$ such that $\operatorname{img} p = S$. Moreover, that projection p has $\ker p = S^\perp$ and, if $\mathscr{B} = (x_1, \dots, x_n)$, is an orthonormal order basis for S, then for every $x \in V$ we have that*

$$p(x) = \sum_{i=1}^{n} \langle x, x_i\rangle x_i.$$

Proof. Let S be a finite-dimensional subspace of V and let $\mathscr{B} = (x_1, \dots, x_n)$ be an ordered ortonormal basis for S. The function $p : V \to V$ such that $p(x) = \sum_{i=1}^{n} \langle x, x_i\rangle x_i$ for each $x \in V$ is clearly linear and, according to Proposition 2.5.1, has $x - p(x) \in S^\perp$ for each $x \in V$.

If $x \in \ker p$, then $x = x - p(x) \in S^\perp$ and, conversely, if $x \in S^\perp$, so that in particular, $\langle x, x_i \rangle = 0$ for each $i \in \{1, \dots, n\}$, then $p(x) = \sum_{i=1}^n \langle x, x_i \rangle x_i = 0$. We see with this that $\ker p = S^\perp$. From the definition of p it is clear that $\operatorname{img} p \subseteq S$. On the other hand, if $x \in S$, then Corollary 2.3.6 tells us that $x = \sum_{i=1}^n \langle x, x_i \rangle x_i$ because $\mathscr{B}$ is an ortonormal basis for S and, as a consequence of this, that $x = p(x) \in \operatorname{img} p$. It follows from this that $\operatorname{img} p = S$ and that $p^2 = p$, as we want. □

§2.8. Riesz's representation theorem

Let us recall that if V is a vector space, then the ***dual space*** of V is the vector space $V^* \coloneqq \hom(V, \Bbbk)$ of all linear functions $V \to \Bbbk$.

{lema:Phi}

Lemma 2.8.1. *Let V be an inner product space, and for each $y \in V$ let us consider the function*

$$\phi_y : x \in V \mapsto \langle x, y \rangle \in \Bbbk.$$

(*i*) *For each $y \in V$ the function ϕ_y is linear, so that it is an element of the dual space V^*.*

(*ii*) *The function $\Phi : y \in V \mapsto \phi_y \in V^*$ is injective and* ***semilinear****, that is, whenever y and y' are elements of V and a one of $\Bbbk$ we have that*

$$\Phi(y + y') = \Phi(y) + \Phi(y'), \qquad \Phi(ay) = \bar{a}\Phi(y).$$

Proof. (*i*) Let y be an element of V. If $x, x' \in V$ and $a, b \in \Bbbk$, then

$$\phi_y(ax + bx') = \langle ax + bx', y \rangle = a\langle x, y \rangle + b\langle x', y \rangle = a\phi_y(x) + b\phi_y(x'),$$

so the function ϕ_y is linear.

(*ii*) Let y and y' be elements of V. For each element y of V we have that

$$\phi_{y+y'}(x) = \langle x, y + y' \rangle = \langle x, y \rangle + \langle x, y' \rangle = \phi_y(x) + \phi_{y'}(x) = (\phi_y + \phi_{y'})(x),$$

so $\Phi(y + y') = \phi_{y+y'} = \phi_y + \phi_{y'} = \Phi(y) + \Phi(y')$. In a similar way, if $y \in V$ and $a \in \Bbbk$, then for each $x \in V$ we have that

$$\phi_{ay}(x) = \langle y, ay \rangle = \bar{a}\langle x, y \rangle = \bar{a}\phi_y(x) = (\bar{a}\phi_y)(x),$$

so that $\Phi(ay) = \phi_{ay} = \bar{a}\phi_y = \bar{a}\Phi(y)$. This proves that the function Φ is semilinear.

Finally, if y and y' are element of V such that $\Phi(y) = \Phi(y')$, then for all $x \in V$ we have that

$$\langle x, y - y' \rangle = \langle x, y \rangle - \langle x, y' \rangle = \Phi(y)(x) - \Phi(y')(x) = 0,$$

and then $y - y' = 0$, that is, $y = y'$. We can therefore conclude that the function Φ is injective. □

The following result is known as *Riesz's Representation Theorem*, because a version of it was proved by Frigyes Riesz (1880–1956, Hungary).

{teorema:riesz}

Theorem 2.8.2. *Let V be a finite-dimensional inner products space. For every element f of V^* there exists a unique vector $x_f \in V$ such that $\phi_{x_f} = x$, that is, such that for every $x \in V$ we have*

$$f(x) = \langle x, x_f \rangle. \tag{2.8}$$

{eq:riesz}

We cannot remove the hypothesis that the vector space V be finite-dimensional that appears here, since the resulting statement is false.

Proof. Let f be an element of V^*. Let n be the dimension of V, let $\mathscr{B} = (x_1, \dots, x_n)$ be an orthonormal ordered basis for V, and let us consider the vector

$$x_f := \sum_{i=1}^{n} \overline{f(x_i)} x_i.$$

If $x \in V$, then $x = \sum_{i=1}^{n} \langle x, x_i \rangle x_i$ and therefore

$$f(x) = f\left(\sum_{i=1}^{n} \langle x, x_i \rangle x_i\right) = \sum_{i=1}^{n} \langle x, x_i \rangle f(x_i) = \left\langle x, \sum_{i=1}^{n} \overline{f(x_i)} x_i \right\rangle = \langle x, x_f \rangle.$$

This tells us that the vector x_f satisfies the condition (2.8). On the other hand, if $x_f' \in V$ is another vector that satisfies that condition, then for all $x \in V$ we have that

$$\langle x, x_f - x_f' \rangle = \langle x, x_f \rangle - \langle x, x_f' \rangle = f(x) - f(x) = 0,$$

and this is only possible if $x_f = x_f'$. This proves the uniqueness claimed in the theorem. □

An immediate consequence of this result is the surjectivity of the function Φ of Lemma 2.8.1.

Corollary 2.8.3. *The function $\Phi : V \to V^*$ of Lemma 2.8.1 is a bijection.*

Proof. Indeed, Theorem 2.8.2 tells us that the function Φ is surjective, and we already know that it is injective. □

§2.9. Adjoint functions

Let V and W be two inner product spaces. We will sometimes write $\langle -, - \rangle_V$ and $\langle -, - \rangle_W$ for the inner products of V and of W, to be explicit about which inner product we have in mind, but most often we will wrote both of them in the form $\langle -, - \rangle$, leaving the reader to use the context to figure out to which inner product we are referring.

If $f : V \to W$ is a linear map, then we say that a linear function $g : W \to V$, in the opposite direction, is ***adjoint*** to f if for every $v \in V$ and every $w \in W$ we have that

$$\langle f(x), y \rangle_W = \langle x, g(y) \rangle_V.$$

Example 2.9.1. Let V be any inner product space, and let $\lambda \in \Bbbk$ be a scalar. The linear map $f : x \in V \mapsto \lambda x \in V$ has $g : y \in V \mapsto \overline{\lambda} y \in V$ as an adjoint. Indeed, whenever x and y are elements of V, we have that

$$\langle f(x), y \rangle = \langle \lambda x, y \rangle = \lambda \langle x, y \rangle = \langle x, \overline{\lambda} y \rangle = \langle x, g(y) \rangle,$$

and this means that g is an adjoint for f.

Example 2.9.2. Let us consider the vector space $\mathbb{R}^2$ endowed with its standard inner product, and the maps $f : (x, y) \in \mathbb{R}^2 \mapsto (0, x, y) \in \mathbb{R}^3$ and $g : (x, y, z) \in \mathbb{R}^3 \mapsto (y, z) \in \mathbb{R}^2$. If $v = (x, y) \in \mathbb{R}^2$ and $w = (x', y', z') \in \mathbb{R}^3$, then we have that

$$\langle f(v), w \rangle = \langle (0, x, y), (x', y', z') \rangle = xy' + yz' = \langle (x, y), (u', z') \rangle = \langle v, g(w) \rangle,$$

and this tells us that the map g is adjoint to f.

We will show below that in often encountered situations linear maps do have adjoints. For now, we simply note that a linear map can have at most one adjoint:

Lemma 2.9.3. *Let V and W be two inner product spaces. A linear map $f : V \to W$ has at most one adjoint.*

In view of this, whenever a linear map $f : V \to W$ has an adjoint map, it has exactly one, and we will write it in the form $f^* : W \to V$.

Proof. Let $f : V \to W$ be a linear map, and let us suppose that $g, g' : W \to V$ are two linear maps adjoint to f. If $y \in W$, then for all $x \in V$ we have that

$$\langle x, g(y) - g'(y) \rangle_V = \langle x, g(y) \rangle_V - \langle x, g'(y) \rangle_V = \langle f(x), y \rangle_W - \langle f(x), y \rangle_W = 0,$$

so that $g(y) = g'(y)$. We can conclude from this that $g = g'$. □

When dealing with adjoints it is useful to know that going from a map to its adjoint is a semilinear, involutive operation: this is precisely the content of the following result:

Proposition 2.9.4. *Let V and W be two inner product spaces.*

(*i*) *If $f, g : V \to W$ are two linear maps that have adjoints, then for each $a, b \in \Bbbk$ the linear map $af + bg : V \to W$ has an adjoint and, in fact, we have that $(af + bg)^* = \overline{a} f^* + \overline{b} g^*$.*

(*ii*) *If $f : V \to W$ is a linear map that has an adjoint, then the map $f^* : W \to V$ also has an adjoint and $(f^*)^* = f$.*

Proof. (*i*) If $x \in V$ and $y \in W$, then

$$\begin{aligned}\langle (af + bg)(x), y\rangle &= \langle af(x) + bf(x), y\rangle = a\langle f(x), y\rangle + b\langle g(x), y\rangle \\ &= a\langle x, f^*(y)\rangle + b\langle x, g^*(y)\rangle = \langle x, \overline{a} f^*(y) + \overline{b} g^*(y)\rangle \\ &= \langle x, (\overline{a} f^* + \overline{b} g^*)(y)\rangle.\end{aligned}$$

This tells us that the linear function $\overline{a} f^* + \overline{b} g^* : V \to W$ is adjoint to $af + gb$, so that the latter has an adjoint and $(af + bg)^* = \overline{a} f^* + \overline{b} g^*$.

(*ii*) If $x \in V$ and $y \in W$, then

$$\langle f^*(x), y\rangle = \overline{\langle y, f^*(x)\rangle} = \overline{\langle f(y), x\rangle} = \langle x, f(y)\rangle,$$

and with this we see that $f : V \to W$ is adjoint to $f^* : W \to V$, that is, that $(f^*)^* = f$. □

The operation of taking adjoints is also compatible with composition, in the following sense:

Proposition 2.9.5.

(*i*) *If V is an inner product space, then the identity function $\mathsf{id}_V : V \to V$ has an adjoint and, in fact, $(\mathsf{id}_V)^* = \mathsf{id}_V$.*

(*ii*) *If V, W and U are inner product spaces and $f : V \to W$ and $g : W \to U$ are linear maps that have adjoints, then the composition $g \circ f : V \to U$ has an adjoint and $(g \circ f)^* = f^* \circ g^*$.*

Proof. To prove the first part it is enough that we observe that whenever x and y are elements of V we have that

$$\langle \mathsf{id}_V(x), y\rangle = \langle x, y\rangle = \langle x, \mathsf{id}_V(y)\rangle,$$

so that id_V is adjoint to itself, that is, $(\mathsf{id}_V)^* = \mathsf{id}_V$. On the other hand, if $f : V \to W$ and $g : W \to U$ are linear functions that have adjoints, then for every $x \in V$ and every $y \in U$ we have that

$$\langle (g \circ f)(x), y\rangle = \langle g(f(x)), y\rangle = \langle f(x), g^*(y)\rangle = \langle x, f^*(g^*(y))\rangle = \langle x, (f^* \circ g^*)(y)\rangle.$$

This means that $g \circ f$ has $f^* \circ g^*$ as an adjoint, so that $(g \circ f)^* = f^* \circ g^*$. □

We will use now Riesz's Representation Theorem 2.8.2 to show that a linear map between inner product spaces with finite-dimensional domain has an adjoint.

{thm:adjoint}

Theorem 2.9.6. *Let V and W be inner product spaces. If V is finite-dimensional, then every linear map $f : V \to W$ has an adjoint.*

Proof. Let us suppose that V is finite-dimensional and consider a linear map $f : V \to W$. If $y \in W$, then then function

$$\psi_y : x \in V \mapsto \langle f(x), y\rangle_W \in \Bbbk$$

is linear, so Theorem 2.8.2 tells us that there exists exactly one vector $f^*(y)$ in V with the property that $\psi_y(x) = \langle x, f^*(y)\rangle_V$ for all $x \in V$. In this way we see that there is a function $f^* : W \to V$ such that for all $x \in W$ and every $y \in V$ we have

$$\langle f(x), y\rangle_V = \langle x, f^*(y)\rangle_W.$$

To show that this map f^* is an adjoint to f, and with that complete the proof of the theorem, we have to show that f^* is a linear function.

- Let y, $y' \in W$. Whenever $x \in V$ we have that

$$\begin{aligned}\langle x, f^*(y+y')\rangle_V &= \langle f(x), y+y'\rangle_W = \langle f(x), y\rangle_W + \langle f(x), y'\rangle_W \\ &= \langle x, f^*(y)\rangle_W + \langle x, f^*(y')\rangle_W = \langle x, f^*(y) + f^*(y)\rangle_W,\end{aligned}$$

  and this implies that $f^*(y+y') = f^*(y) + f^*(y')$.
- On the other hand, if $y \in W$ and $\lambda \in \Bbbk$, then for each $x \in W$ we have that

$$\begin{aligned}\langle x, f^*(\lambda y)\rangle_V &= \langle f(x), \lambda y\rangle_W = \langle \bar\lambda f(x), y\rangle_W = \langle f(\bar\lambda x), y\rangle_W = \langle \bar\lambda x, f^*(y)\rangle_V \\ &= \langle x, \lambda f^*(y)\rangle_V,\end{aligned}$$

  so that $f^*(\lambda w) = \lambda f^*(w)$.

We thus see that the map f^* is indeed linear, as we wanted. □

One way to describe a linear map is by giving its matrix with respect to a pair of bases of its domain and codomain. The following proposition does this for the adjoint of a map.

{prop:pi:adj:mat}

Proposition 2.9.7. *Let V and W be finite-dimensional inner product spaces, and let $\mathscr{B}$ and $\mathscr{B}'$ be* orthonormal *ordered bases for V and for W', respectively. If $f : V \to W$ is a linear map, then the* *matrix of the adjoint map $f^* : W \to V$ with respect to the bases $\mathscr{B}'$ and $\mathscr{B}$ is*

$$[f^*]^{\mathscr{B}'}_{\mathscr{B}} = \overline{([f]^{\mathscr{B}}_{\mathscr{B}'})^{\mathrm{t}}}.$$

Proof. Let us suppose that the bases for V and W are $\mathscr{B} = (x_1, \dots, x_m)$ and $\mathscr{B}' = (y_1, \dots, y_n)$, and that the matrix of f with respect to them is $[f]^{\mathscr{B}}_{\mathscr{B}'} = (a_{i,j}) \in \mathrm{M}_{n,m}(\Bbbk)$. If $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n\}$, then

$$\overline{\langle f^*(y_j), x_i \rangle} = \langle x_i, f^*(y_j) \rangle = \langle f(x_i), y_j \rangle = \left\langle \sum_{k=1}^{m} a_{k,i} y_k, y_j \right\rangle = \sum_{k=1}^{m} a_{k,i} \langle y_k, y_j \rangle = a_{j,i}$$

because the basis $\mathscr{B}'$ is orthonormal, and therefore $\langle f^*(y_j), x_i \rangle = \overline{a_{j,i}}$. According to Corollary 2.3.6, we can deduce from this that

$$f^*(y_j) = \sum_{i=1}^{m} \overline{a_{j,i}}\, x_i.$$

and thus that the matrix of f^* with respect to the bases $\mathscr{B}'$ and $\mathscr{B}$ is $(\overline{a_{j,i}})_{i,j}$, that is, the matrix obtained from $[f]^{\mathscr{B}}_{\mathscr{B}'}$ by first transposing and then conjugating, as the proposition claims. □

It is important to keep in mind that the equality claimed by the proposition is only valid, in general, if the bases $\mathscr{B}$ and $\mathscr{B}'$ are orthonormal.

Example 2.9.8. Let us consider the vector space $\Bbbk^2$ endowed with its standard inner product, the matrix $A = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right) \in \mathrm{M}_2(\Bbbk)$, and the linear map $f : x \in \Bbbk^2 \mapsto Ax \in \Bbbk^2$. If $\mathscr{B} = (e_1, e_2)$ is the standard ordered basis for $\Bbbk^2$, then $[f]^{\mathscr{B}}_{\mathscr{B}} = A$ and, since $\mathscr{B}$ is in fact an orthonormal basis for $\Bbbk^2$, that $[f^*]^{\mathscr{B}}_{\mathscr{B}} = A^{\mathrm{t}}$. If now put $e_2' := e_1 + e_2$, then $\mathscr{B}' = (e_1, e_2')$ is also an ordered basis for $\Bbbk^2$, but not an orthonormal one, and we have $[f]^{\mathscr{B}'}_{\mathscr{B}'} = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right)$ while $[f^*]^{\mathscr{B}'}_{\mathscr{B}'} = \left(\begin{smallmatrix} -1 & 0 \\ 1 & 0 \end{smallmatrix}\right)$.

Exercise 2.9.9. Let V and W be finite-dimensional inner product spaces. We proved in Theorem 2.9.6 that every linear map $V \to W$ has an adjoint and computed in Proposition 2.9.7 the matrix of that adjoint with respect to a pair of orthonormal bases. We can reverse the reasoning and use this idea to prove that adjoints exist.

Let $f : V \to W$ be a linear map and let $\mathscr{B}$ and $\mathscr{B}'$ be orthonormal bases for V and for W, respectively. Prove — without using Theorem 2.9.6 and Proposition 2.9.7 — that there is a unique linear map $g : W \to V$ such that $[g]^{\mathscr{B}}_{\mathscr{B}'} = \overline{([f]^{\mathscr{B}'}_{\mathscr{B}})^t}$ and that it is an adjoint to f.

A linear map $f : V \to W$ determines uniquely its adjoint $f^* : W \to V$ — when it has one, of course — and therefore we should be able to answer any question about f^* by looking at f. The following proposition deals with the description of the kernel and the image of f^*.

{prop:adj-im-ker}

Proposition 2.9.10. *Let V and W be two inner product spaces. If $f : V \to W$ is a linear map that has an adjoint, then*

(i) $\ker f^* = (\operatorname{img} f)^\perp$,

(ii) $\operatorname{img} f^* \subseteq (\ker f)^\perp$, *and*

(iii) $(\operatorname{img} f^*)^\perp \subseteq \ker f$.

If additionally V is finite-dimensional, then we also have that

(iv) $\operatorname{img} f^* = \ker f^\perp$.

Proof. (i) Let $x \in \ker f^*$. If $y \in \operatorname{img} f$, so that there exists a $z \in V$ such that $f(z) = y$, then $\langle y, x\rangle = \langle f(z), x\rangle = \langle z, f^*(x)\rangle = 0$. This tells us that $x \in (\operatorname{img} f)^\perp$. Conversely, if $x \in (\operatorname{img} f)^\perp$, then for all $y \in V$ we have that $\langle y, f^*(x)\rangle = \langle f(y), x\rangle = 0$, so that $f^*(x) = 0$, that is, $x \in \ker f^*$.

(ii) Let $x \in \operatorname{img} f^*$, so that there is a $y \in V$ such that $x = f^*(y)$. For each $z \in \ker f$ we have that $\langle z, x\rangle = \langle z, f^*(y)\rangle = \langle f(z), x\rangle = 0$, so $x \in (\ker f)^\perp$: this shows that $\operatorname{img} f^* \subseteq \ker f^\perp$.

(iii) Let $x \in (\operatorname{img} f^*)^\perp$. If $y \in V$, then $\langle f(x), y\rangle = \langle x, f^*(y)\rangle = 0$, so $f(x) = 0$. We can conclude from this that $(\operatorname{img} f^*)^\perp \subseteq \ker f^*$.

(iv) Let us suppose now that V is finite-dimensional. According to part (ii) of this proposition, we have that $\operatorname{img} f^* \subseteq (\ker f)^\perp$, and part (iii) of Proposition 2.6.1 then implies that also $\ker f = (\ker f)^{\perp\perp} \subseteq (\operatorname{img} f^*)^\perp$. Together with part (iii) of this proposition this lets us conclude that, in fact, $\ker f = (\operatorname{img} f^*)^\perp$. Similarly, according to part(iii), we have that $(\operatorname{img} f^*)^\perp \subseteq \ker f$, so part (iii) of Proposition 2.6.1 tells us that also $(\ker f)^\perp \subseteq (\operatorname{img} f^*)^{\perp\perp} = \operatorname{img} f^*$: this and part (ii) allow us to conclude that $(\ker f)^\perp = \operatorname{img} f^*$. □

In general, the equality of part (iv) of Proposition 2.9.10 does not hold.

Proposition 2.9.11. *Let V be a finite-dimensional inner product space, and let $f : V \to V$ be a linear map. If $\lambda \in \Bbbk$ is an eigenvalue for f, then $\overline{\lambda}$ is an eigenvalue for f^*.*

Proof. Let λ be an element of $\Bbbk$ such that $\overline{\lambda}$ is *not* an eigenvalue of f^*, so that the linear map $f^* - \overline{\lambda} \cdot \mathrm{id}_V : V \to V$ is injective and, therefore, bijective. There is then a linear map $g : V \to V$ such that $\mathrm{id}_V = (f^* - \overline{\lambda} \cdot \mathrm{id}_V) \circ g$, and thus also

$$\mathrm{id}_V = \mathrm{id}_V^* = g^* \circ (f^* - \overline{\lambda} \cdot \mathrm{id}_V)^* = g^* \circ (f - \lambda \cdot \mathrm{id}_V).$$

The map $f - \lambda \cdot \mathrm{id}_V$ is therefore injective, and this tells us that the number λ is not an eigenvalue of f. This proves the contraposition of the claim of the proposition. □

This proposition tells us that the eigenvalues of the adjoint of an endomorphism $f : V \to V$ of a finite-dimensional vector space are the conjugates of the eigenvalues of f, but does not say anything about the corresponding eigenvectors. In general, if λ is an eigenvector of f, then there is very little relation between the eigenvectors of f with eigenvalue λ and the eigenvectors of f^* with eigenvalue $\overline{\lambda}$.

§2.10. Self-adjoint linear maps

Let V be an inner product space. A linear map $f : V \to V$ that is its own adjoint is said to be ***self-adjoint***. In other words, we say that f is self-adjoint if for all x and y in V we have that

$$\langle f(x), y\rangle = \langle x, f(y)\rangle.$$

{prop:auto}

Proposition 2.10.1. *Let V be a finite-dimensional inner product space and let $\mathscr{B}$ be an orthonormal ordered basis for V. A linear map $f : V \to V$ is self-adjoint if and only if*

$$[f]^{\mathscr{B}}_{\mathscr{B}} = \overline{([f]^{\mathscr{B}}_{\mathscr{B}})^{\mathrm{t}}}.$$

In particular,

(i) if $\mathbb{k} = \mathbb{R}$, then f is self-adjoint if and only if the matrix $[f]^{\mathscr{B}}_{\mathscr{B}}$ is symmetric, and

(ii) if $\mathbb{k} = \mathbb{C}$, then f is self-adjoint if and only if the matrix $[f]^{\mathscr{B}}_{\mathscr{B}}$ is hermitian.

Proof. Let $f : V \to V$. According to Proposition 2.9.7, we have that $[f^*]^{\mathscr{B}}_{\mathscr{B}} = \overline{[f]^{\mathscr{B}}_{\mathscr{B}}}$, and it follows from this that $f = f^*$ exactly when $[f]^{\mathscr{B}}_{\mathscr{B}} = \overline{([f]^{\mathscr{B}}_{\mathscr{B}})^{\mathrm{t}}}$. □

A useful class of examples of self-adjoint maps is that of orthonormal projections, as the following result shows:

Proposition 2.10.2. *Let V be an inner product space. A projection $p : V \to V$ is orthogonal if and only if it is self-adjoint.*

Proof. Let $p : V \to V$ be a projection, and let us suppose at first that it is orthogonal. Let x, $y \in V$. As $V = \operatorname{img} p \oplus \ker p$, there are x', $y' \in \operatorname{img} p$ and x'', $y'' \in \ker p$ such that $x = x' + x''$ and $y = y' + y''$. Moreover, we have that $p(x) = x'$, $p(y) = y'$ and, according to the hypothesis, that $\operatorname{img} p \perp \ker p$, so $\langle x', y''\rangle = \langle x'', y'\rangle = 0$. Using all this we find that que

$$\begin{aligned}\langle p(x), y\rangle &= \langle x', y' + y''\rangle = \langle x', y'\rangle + \langle x', y''\rangle = \langle x', y'\rangle = \langle x', y'\rangle + \langle x'', y'\rangle \\ &= \langle x' + x'', y'\rangle = \langle x, p(y)\rangle,\end{aligned}$$

and therefore that p is its own adjoint.

Let us suppose next that the projection p is self-adjoint. If $x \in \operatorname{img} p$ and $y \in \ker p$, then the self-adjointness of p implies that $\langle x, y\rangle = \langle p(x), y\rangle = \langle x, p(y)\rangle = 0$. We can therefore conclude that $\operatorname{img} p \perp \ker p$, so that the projection p is an orthogonal one. □

We want to study the diagonalizability of self-adjoint maps. A first step is the following proposition that gives information about the eigenvalues and eigenvectors of such a map.

Proposition 2.10.3. *Let V be an inner product space, and let $f : V \to V$ be a self-adjoint linear map.* {prop:pi:auto}

(*i*) *Every eigenvalue of f is a real number.* {prop:pi:auto:real}

(*ii*) *If x and y are eigenvectors for f corresponding to different eigenvalues, then $x \perp y$.* {prop:pi:auto:perp}

Proof. (*i*) Let $\lambda \in \Bbbk$ be an eigenvalue of f, and let $x \in V$ un autovector for f with eigenvalue λ. We then have that

$$\lambda\langle x, x\rangle = \langle \lambda x, x\rangle = \langle f(x), x\rangle = \langle x, f(x)\rangle = \langle x, \lambda x\rangle = \overline{\lambda}\langle x, x\rangle$$

and, since $\langle x, x\rangle \neq 0$, this implies that $\lambda = \overline{\lambda}$, so that $\lambda \in \mathbb{R}$.

(*ii*) Let x, $y \in V$ be eigenvectors with eigenvalues λ, $\mu \in \Bbbk$, respectively, and suppose that $\lambda \neq \mu$. From part (*i*) we know that λ, $\mu \in \mathbb{R}$, and therefore

$$\lambda\langle x, y\rangle = \langle \lambda x, y\rangle = \langle f(x), y\rangle = \langle x, f(y)\rangle = \langle x, \mu y\rangle = \overline{\mu}\langle x, y\rangle = \mu\langle x, y\rangle.$$

It folows from this that $(\lambda - \mu)\langle x, y\rangle = 0$ and, since $\lambda \neq \mu$, that $\langle x, y\rangle = 0$. □

Apart from being real, the single most important property of the eigenvalues of a self-adjoint linear map is that they exist — at least when we are working with a finite dimensional vector space.

Proposition 2.10.4. *Let V be a finite-dimensional inner product space. A self-adjoint map $f : V \to V$ has at least one eigenvalue.*

Proof. If $\Bbbk = \mathbb{C}$ then we know that every linear map $V \to V$ has an eigenvalue, so in that case there is nothing to prove — and, in fact, in this case we do not need the hypothesis. We assume from now on that $\Bbbk = \mathbb{R}$.

Let $n \coloneqq \dim V$, let $\mathscr{B}$ be an orthonormal basis for V, and let $A = [f]_{\mathscr{B}} \in M_n(\mathbb{R})$. As f is self-adjoint, the matrix A is symmetric. Let us consider the endomorphism $g : x \in \mathbb{C}^n \mapsto Ax \in \mathbb{C}^n$ of the *complex* vector space $\mathbb{C}^n$, and let $\mathscr{B}'$ be the standard ordered basis of $\mathbb{C}^n$, which is orthonormal with respect to the standar inner product on that space. As $[g]_{\mathscr{B}'} = A = \overline{A^{t}} = \overline{[g]^{t}_{\mathscr{B}'}} = [g^*]_{\mathscr{B}'}$ because the matrix A is real and symmetric, the map g is self-adjoint and its eigenvalues are real. This tells us that all the roots of the characteristic polynomial $\chi_g \in \mathbb{C}[X]$ are real. Since the field $\mathbb{C}$ is algebraically closed, there is therefore some real number $\lambda \in \mathbb{R}$ such that $\chi_g(\lambda) = 0$.

Now, the characteristic polynomial χ_f of f coincides with the characteristic polynomial χ_A

of A, and this one coincides in turn with the characteristic polynomial χ_g of g: it follows from this that $\chi_f(\lambda) = \chi_g(\lambda) = 0$, so that λ is an eigenvalue for f. □

Putting together what we have done so far we can prove the most important property of the self-adjoint endomorphisms of finite-dimensional inner product spaces:

{prop:pi:auto:diag}

Proposition 2.10.5. *Let V be a finite-dimensional inner product space. If $f : V \to V$ is a self-adjoint endomorphism of V, then there exists an orthonormal ordered basis $\mathscr{B}$ for V whose elements are eigenvectors for f, so that the matrix $[f]_{\mathscr{B}}$ is diagonal.*

Proof. We will proceed by induction with respecto to the number $\dim V$. If $\dim V = 0$, then there is nothing to prove, and this establishes the base case. Let us then suppose that $n = \dim V$ is a positive integer.

Let $f : V \to V$ be a self-adjoint linear map. According to the previous proposition, there exists an eigenvalue $\lambda \in \Bbbk$, and therefore there exists a non-zero vector $x_1 \in V \smallsetminus 0$ such that $f(x_1) = \lambda x_1$. In fact, we can suppose that $\|x\| = 1$ — if that is not the case, we can simply replace x_1 by $x_1/\|x_1\|$.

Let now $W \coloneqq \langle x_1 \rangle^\perp$. If $x \in W$, then

$$\langle f(x), x_1 \rangle = \langle x, f(x_1) \rangle = \langle x, \lambda x_1 \rangle = \lambda \langle x, x_1 \rangle = 0,$$

so that $f(x) \in W$: this tells us that the subspace W is f-invariant, and that we can thus consider the restriction $f_W : W \to W$ of f to W. Let us view W as an inner product space with the inner product $\langle -, - \rangle_W$ obtained by restricting that of V. For all vector x and y in W we have that

$$\langle f_W(x), y \rangle_W = \langle f(x), y \rangle = \langle x, f(y) \rangle = \langle x, f_W(y) \rangle_W,$$

and the endomorphism f_W of W is therefore self-adjoint. As $\dim W = \dim V - 1$, we can inductively suppose that there is an orthonormal ordered basis $(x_2, \dots, x_n)$ of W whose elements are eigenvectors for f_W. It follows immediately from this that $\mathscr{B} = (x_1, \dots, x_n)$ is an orthonormal ordered basis for V whose elements are eigenvectors for f, and this proves the proposition. □

We can go a bit further and show that the property of the proposition actually characterizes self-adjoint linear maps:

Corollary 2.10.6. *Let V be a finite-dimensional inner product space. A linear function $f : V \to V$ is self-adjoint if and only if there exists an orthonormal ordered basis $\mathscr{B}$ such that the matrix $[f]^{\mathscr{B}}_{\mathscr{B}}$ is diagonal and real.*

Proof. The necessity of the condition is a consequence of Proposition 2.10.5 and part (i) of Proposition 2.10.3. Its sufficiency, on the other hand, follows immediately from Proposition 2.9.7. □

§2.11. Normal linear maps

Let V be an inner product space. A linear map $f : V \to V$ is ***normal*** if it has an adjoint and commutes with it, so that

$$f^* f = f f^*.$$

It is clear that a self-adjoint map is normal, but the converse is false.

{ejemplo:alpha}

Example 2.11.1. Let us consider the vector space $\mathbb{R}^2$ endowed with its standard inner product, and let $\mathscr{B}$ be the standard ordered bases, which is orthonormal. Let $\alpha \in \mathbb{R}$ and let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map such that

$$[f]_{\mathscr{B}} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}$$

As $\mathscr{B}$ is an orthonormal ordered basis, we know that

$$[f^*]_{\mathscr{B}} = \overline{([f]^{\mathscr{B}}_{\mathscr{B}})^{\mathrm{t}}} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix},$$

and a direct calculation shows that $[f^*]_{\mathscr{B}}[f]_{\mathscr{B}} = I_2 = [f]_{\mathscr{B}}[f^*]_{\mathscr{B}}$, so thatr manera que $f^* f = \mathsf{id}_V = f f^*$. We thus see that f is normal. It is clear, on the other hand, that $f^* = f$ if and only if $\sin\alpha = -\sin\alpha$, and this happens if and only if α is an integer multiple of π.

The class of normal maps generalizes that of self-adjoint maps, and one can expect it to have similar properties. The following result is what corresponds to Proposition 2.10.3:

{prop:adj-lambda}

Proposition 2.11.2. *Let V be an inner product space, let $f : V \to V$ be a normal linear map, and let $x \in V$ and $\lambda \in \Bbbk$. The following statements are equivalent:*

(a) x is an eigenvector for f with eigenvalue λ.

(b) x is an eigenvector for f^* with eigenvalue $\overline{\lambda}$.

Proof. Let $g = f - \lambda\mathsf{id}_V$. We know that g has an adjoint, and that in fact $g^* = f^* - \overline{\lambda}\mathsf{id}_V$. Since f es normal, we can compute that

$$\begin{aligned} gg^* &= (f - \lambda\mathsf{id}_V)(f^* - \overline{\lambda}\mathsf{id}_V) \\ &= ff^* - \lambda f^* - \overline{\lambda} f + \lambda\overline{\lambda} \\ &= f^* f - \overline{\lambda} f - \lambda f^* + \overline{\lambda}\lambda \\ &= (f^* - \overline{\lambda}\mathsf{id}_V)(f - \lambda\mathsf{id}_V) \\ &= g^* g, \end{aligned}$$

so that g is also normal. Using this we see that

$$\|g(x)\|^2 = \langle g(x), g(x)\rangle = \langle x, g^*(g(x))\rangle = \langle x, g(g^*(x))\rangle = \langle g^*(x), g^*(x)\rangle = \|g^*(x)\|^2,$$

and it follows from this that $g(v) = 0$ exactly when $g^*(v) = 0$. In view of the definition of g, this means that $f(v) = \lambda v$ exactly when $f^*(v) = \overline{\lambda}v$, and this is precisely what the proposition asserts. □

Corresponding to the diagonalizability of self-adjoint linear maps, we have here that normal maps are diagonalizable, but only in the complex case.

{teorema:normales

Theorem 2.11.3. *Let V be a finite-dimensional* complex *inner product space. If $f : V \to V$ is a normal linear map, then there exists an orthonormal ordered basis for V whose elements are eigenvectors for f.*

Proof. We proceed by induction with respect to the dimension of V, noting that when $\dim V = 0$ there is nothing to prove. Let then V be a non-zero finite-dimensional complex inner product space. As $\mathbb{C}$ is algebraically closed, there exist $\lambda \in \mathbb{C}$ and $x_1 \in V$ such that $\|x_1\| = 1$ and $f(x_1) = \lambda x_1$. According to Proposition 2.11.2, we also have that $f^*(x_1) = \overline{\lambda}x_1$.

Le $W := \langle x\rangle^\perp$. If $y \in W$, then $y \perp x$ and we have that

$$\langle f(y), x\rangle = \langle y, f^*(x)\rangle = \langle y, \overline{\lambda}x\rangle = \lambda\langle y, x\rangle = 0$$

and

$$\langle f^*(y), x\rangle = \langle y, f(x)\rangle = \langle y, \lambda x\rangle = \overline{\lambda}\langle y, x\rangle = 0.$$

We thus see that the subspace W is f- and f^*-invariant and that, in particular, we can consider the restrictions $f_W, (f^*)_W : W \to W$. If $\langle -, -\rangle_W$ is the inner product of the subspace W, then for all x and y in W we have that

$$\langle f_W(x), y\rangle_W = \langle f(x), y\rangle = \langle x, f^*(y)\rangle = \langle x, (f^*)_W(y)\rangle_W,$$

and this tells us that $f_W : W \to W$ has $(f^*)_W : W \to W$ as adjoint, so that $(f_W)^* = (f^*)_W$. Moreover, if $x \in W$, we have

$$f_W((f_W)^*(x)) = f_W((f^*)_W(x)) = f(f^*(x)) = f^*(f(x)) = (f^*)_W(f_W(x)),$$

and we thus have that $f_W(f_W)^* = (f_W)^* f_W$, so that f_W is normal.

The map $f_W : W \to W$ is therefore a normal endomorphism of the complex inner product space W. As $\dim W = \dim V - 1$, we can inductively suppose that there is an orthonormal ordered basis $(v_2, \dots, v_n)$ for W whose elements are eigenvectors for f_W. Of course, it follows from this that $(v_1, \dots, v_n)$ is an orthonormal ordered basis for V whose elements are eigenvectors for f. This completes the induction and proves the proposition. □

The theorem tells us that that a normal endomorphism is diagonalizable with respect to an orthonormal ordered basis, and this property in fact characterizes normality:

Corollary 2.11.4. *Let V be a finite-dimensional complex inner product space, and let $f : V \to V$ be a linear map. The following two statements are equivalent:*

(*a*) *The function f is normal.*

(*b*) *There exists an orthonormal ordered basis of V whose elements are eigenvectors for f.*

Proof. The implication $(a) \Rightarrow (b)$ is precisely the claim of Theorem 2.11.3. Let us prove the converse implication.

Let us suppose that $\mathscr{B}$ is an orthonormal ordered basis for V whose elements are eigenvectors for f, so that the matrix $[f]_{\mathscr{B}}$ is diagonal. As $[f^*]_{\mathscr{B}} = \overline{[f]^{t}_{\mathscr{B}}}$ is also diagonal, the matrices $[f]_{\mathscr{B}}$ and $[f^*]_{\mathscr{B}}$ commute and, therefore, the maps f and f^* commute: in other words, the map f is normal. □

Example 2.11.5. In Theorem 2.11.3 the hypothesis that the vector space be complex is imporant. Indeed, the linear function $f : \mathbb{R}^2 \to \mathbb{R}^2$ constructed in Example 2.11.1 is normal for all $\alpha \in \mathbb{R}$ but does not have any eigenvector if α is not an integer multiple of π.

§2.12. Orthogonal and unitary linear maps

If V is an inner product space, and $f : V \to V$ is an endomorphism of V such that

$$\langle f(x), f(y)\rangle = \langle x, y\rangle$$

for all $x, y \in V$, we say that f is ***unitary*** when the field $\Bbbk$ is $\mathbb{C}$ and that it is ***orthogonal*** when the field $\Bbbk$ is $\mathbb{R}$.

The unitary or orthogonal endomorphisms of an inner product space are those that preserve its inner product, and therefore they also preserve the norm associated to that inner product. The following result shows that this last condition is also sufficient for unitarity or orthogonality:

Proposition 2.12.1. *Let V be an inner product space, and let $f : V \to V$ be an endomorphism of V. The following statements are equivalent:*

(*a*) *The function f is unitary if $\Bbbk = \mathbb{C}$ or orthogonal if $\Bbbk = \mathbb{R}$.*

(*b*) *For every $x \in V$ we have that $\|f(x)\| = \|x\|$.*

Proof. $(a) \Rightarrow (b)$ If f is unitary or orthogonal, and $x \in V$, then

$$\|f(x)\| = \sqrt{\langle f(x), f(x)\rangle} = \sqrt{\langle x, x\rangle} = \|x\|.$$

$(b) \Rightarrow (a)$ Let us suppose that for every $x \in V$ we have that $\|f(x)\| = \|x\|$, and let x and y be two elements of V. If $\Bbbk = \mathbb{R}$, the Corollary 2.2.2 allows us to compute that

$$\langle f(x), f(y)\rangle = \tfrac{1}{4}\|f(x) + f(y)\|^2 - \tfrac{1}{4}\|f(x) - f(y)\|^2 = \tfrac{1}{4}\|f(x+y)\|^2 - \tfrac{1}{4}\|f(x-y)\|^2$$

and the hypothesis implies that this is

$$= \tfrac{1}{4}\|x + y\|^2 - \tfrac{1}{4}\|x - y\|^2 = \langle x, y\rangle,$$

so that f is orthogonal. Similarly, if $\Bbbk = \mathbb{C}$ that same corollary tells us that

$$\begin{aligned}
&\langle f(x), f(y)\rangle \\
&\quad = \tfrac{1}{4}\|f(x) + f(y)\|^2 - \tfrac{1}{4}\|f(x) - f(y)\|^2 \tfrac{i}{4}\|f(x) + if(y)\|^2 - \tfrac{i}{4}\|f(x) - if(y)\|^2 \\
&\quad = \tfrac{1}{4}\|f(x+y)\|^2 - \tfrac{1}{4}\|f(x-y)\|^2 + \tfrac{i}{4}\|f(x+iy)\|^2 - \tfrac{i}{4}\|f(x-iy)\|^2 \\
&\quad = \tfrac{1}{4}\|x+y\|^2 - \tfrac{1}{4}\|x-y\|^2 + \tfrac{i}{4}\|x+iy\|^2 - \tfrac{i}{4}\|x-iy\|^2 \\
&\quad = \langle x, y\rangle,
\end{aligned}$$

so that f is in this case unitary. □

We can characterize unitary and orthogonal maps in terms of their adjoints:

Proposition 2.12.2. *Let V be an inner product space, and let $f : V \to V$ be an endomorphism of V that admits an adjoint map $f^* : V \to V$. The following statements are equivalent:*

(a) The endomorphism f is unitary if $\Bbbk = \mathbb{C}$ or orthogonal if $\Bbbk = \mathbb{R}$.

(b) We have that $f^* \circ f = \mathsf{id}_V = f \circ f^*$, so that f^* is invertivle and has $f^{-1} = f^*$.

Proof. $(a) \Rightarrow (b)$ If f is unitary or orthogonal and $x \in V$, for each $y \in V$ we have that

$$\langle (f^* \circ f)(x), y\rangle = \langle f^*(f(x)), y\rangle = \langle f(x), f(y)\rangle = \langle x, y\rangle,$$

so that $(f^* \circ f)(x) = x$: this shows that $f^* \circ f = \mathsf{id}_V$. A similar argument shows that $f \circ f^* = \mathsf{id}_V$.

$(b) \Rightarrow (a)$ Let us suppose that $f^* \circ f = \mathsf{id}_V$ and $f \circ f^* = \mathsf{id}_V$. If$x$, $y \in V$, then

$$\langle f(x), f(y)\rangle = \langle f^*(f(x)), y\rangle = \langle x, y\rangle,$$

and this tells us that f is unitary or orthogonal, depending on the case. □

We can also characterize unitary or orthogonal maps in terms of what they do to orthonormal bases:

Proposition 2.12.3. *Let V be a finite-dimensional inner product space, and let $f : V \to V$ be a linear map. The following statements are equivalent:* {prop:f-orth}

- (a) *The endomorphism f is unitary if $\Bbbk = \mathbb{C}$ or orthogonal if $\Bbbk = \mathbb{R}$.*
- (b) *There exists an orthonormal ordered basis $(v_1, \dots, v_n)$ of V such that $(f(v_1), \dots, f(v_n))$ is also an orthonormal ordered basis for V.*
- (c) *For every orthonormal ordered basis $(v_1, \dots, v_n)$ of V, the sequence $(f(v_1), \dots, f(v_n))$ is also an orthonormal ordered basis for V.*

Proof. $(c) \Rightarrow (b)$ Let us suppose that the statement (c) holds. We know there exists an orthonormal ordered basis $(v_1, \dots, v_n)$ for V, and the hypothesis implies that $(f(v_1), \dots, f(v_n))$ is also an orthonormal ordered basis for V: this means that the statement (b) holds.

$(b) \Rightarrow (a)$ Let us suppose that there is an orthonormal ordered basis $(v_1, \dots, v_n)$ of V such that $(f(v_1), \dots, f(v_n))$ is also an orthonormal basis, so that $\langle f(v_i), f(v_j)\rangle = \delta_{i,j}$ for all i and j in $\{1, \dots, n\}$. Let x and y be two vectors in V, and let $a_1, \dots, a_n$ and $b_1, \dots, v_n$ be scalars in $\Bbbk$ such that $x = \sum_{i=1}^n a_i v_i$ and $y = \sum_{i=1}^n b_i v_i$. We then have that

$$\langle x), y)\rangle = \left\langle \sum_{i=1}^n a_i v_i, \sum_{j=1}^n b_j v_j \right\rangle = \sum_{i=1}^n \sum_{j=1}^n a_i \overline{b_j} \langle v_i, v_j\rangle = \sum_{i=1}^n a_i \overline{b_j}$$

and

$$\langle f(x), f(y)\rangle = \left\langle \sum_{i=1}^n a_i f(v_i), \sum_{j=1}^n b_j f(v_j) \right\rangle = \sum_{i=1}^n \sum_{j=1}^n a_i \overline{b_j} \langle f(v_i), f(v_j)\rangle = \sum_{i=1}^n a_i \overline{b_j},$$

so that in fact $\langle f(x), f(y)\rangle = \langle x, y\rangle$. This shows that f is unitary or orthogonal.

$(a) \Rightarrow (c)$ Let us suppose finally that f is unitary or orthogonal, and let $(v_1, \dots, v_n)$ be an orthonormal ordered basis for V. For all i and j in $\{1, \dots, n\}$ we then have that $\langle f(v_i), f(v_j)\rangle = \langle v_i, v_j\rangle = \delta_{i,j}$, and this tells us that the sequence $(v_1, \dots, v_n)$ is orthonormal: in particular, it is linearly independent and, since $n = \dim V$, a basis of V. □

§2.13. Orthogonal and unitary matrices

What we did in the previous sections with self-adjoint, normal, unitary and orthogonal linear maps has variants for matrices that are very useful. We deal with that here.

Let $n \in \mathbb{N}$ and let us consider the real vector space $\mathbb{R}^n$ and the complex vector space $\mathbb{C}^n$ endowed with their standard inner products.

- We say that a matrix $A \in \mathrm{M}_n(\mathbb{R})$ is ***orthogonal*** if whenever x and y are vectors in $\mathbb{R}^n$ we have that $\langle Ax, Ay\rangle = \langle x, y\rangle$. We write $O_n(\mathbb{R})$ for the set of all orthogonal matrices in $\mathrm{M}_n(\mathbb{R})$.
- Similarly, we say that a matrix $A \in \mathrm{M}_n(\mathbb{C})$ is ***unitary*** if for each x, $y \in \mathbb{C}^n$ we have that $\langle Ax, Ax\rangle = \langle x, y\rangle$. We write $\mathrm{U}_n(\mathbb{C})$ for the set of all unitary matrices in $\mathrm{M}_n(\mathbb{C})$.

The following proposition shows that the unitarity/orthogonality of matrices is closel related to the unitarity/orthogonality of linear maps, and gives two useful criteria to check those properties.

Proposition 2.13.1. *Let $n \in \mathbb{N}$ and let us consider the real vector space $\mathbb{R}^n$ and the complex vector space $\mathbb{C}^n$ endowed with their standard inner products.*

(i) *If $A \in \mathrm{M}_n(\mathbb{R})$, then the following statements are equivalent:*

(a) *The matrix A is orthogonal.*

(b) *The linear function $f_A : x \in \mathbb{R}^n \mapsto Ax \in \mathbb{R}^n$ is orthogonal.*

(c) *We have that $AA^{\mathrm{t}} = I_n = A^{\mathrm{t}}A$.*

(d) *The set of columns of A is an orthonormal basis for $\mathbb{R}^n$.*

(ii) *If $A \in \mathrm{M}_n(\mathbb{C})$, the the following statements are equivalent:*

(a) *The matrix A is unitary.*

(b) *The linear function $f_A : x \in \mathbb{C}^n \mapsto Ax \in \mathbb{C}^n$ is unitary.*

(c) *We have that $A\overline{A}^{\mathrm{t}} = I_n = \overline{A}^{\mathrm{t}}A$.*

(d) *The set of columns of A is an orthonormal basis for $\mathbb{C}^n$.*

Proof. We will only prove the first part of the proposition, as the proof of the second one is completely similar. Let us fix a matrix A in $\mathrm{M}_n(\mathbb{R})$.

$(a) \Leftrightarrow (b)$ For all x and y in $\mathbb{R}^n$ we of course have that $\langle f_A(x), f_A(y)\rangle = \langle Ax, Ay\rangle$, so if one of the two sides of this equality is equal to $\langle x, y\rangle$ so is the other one. This tells us that the matrix A is orthogonal if and only if the liner map f_A is an orthogonal.

$(b) \Rightarrow (c)$ Let us suppose that the function f_A is orthogonal, so that f_A and its adjoint f_A^* are mutually inverse. If $\mathscr{B}$ is the standard ordered basis of $\mathbb{R}^n$, then we have that $[f_A]_{\mathscr{B}} = A$ and $[f_A^*]_{\mathscr{B}} = [f_A]_{\mathscr{B}}^{\mathrm{t}} = A^t$, and we thus have that

$$AA^{\mathrm{t}} = [f_A]_{\mathscr{B}}[f_A^*]_{\mathscr{B}} = [f_A f_A^*]_{\mathscr{B}} = [\mathsf{id}_V]_{\mathscr{B}} = I_n$$

and

$$A^{\mathrm{t}}A = [f_A^*]_{\mathscr{B}} = [f_A]_{\mathscr{B}}[f_A^* f_A]_{\mathscr{B}} = [\mathsf{id}_V]_{\mathscr{B}} = I_n.$$

$(c) \Rightarrow (d)$ Let us suppose that $A^{\mathrm{t}}A = I_n$, and write $v_1, \ldots, v_n$ for the columns of A, which are elements of $\mathbb{R}^n$. For all i and j in $\{1, \ldots, n\}$ the (i, j)th entry of the product $A^{\mathrm{t}}A$ is $\langle v_i, v_j\rangle$, and this means that if $A^{\mathrm{t}}A = I_n$ then the sequence $(v_1, \ldots, v_n)$ is orthonormal and therefore an orthonormal ordered basis for $\mathbb{R}^n$.

$(d) \Rightarrow (b)$ Finally, let us suppose that the columns $v_1, \ldots, v_n$ are an orthonormal basis for $\mathbb{R}^n$. The standard ordered basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$ is orthonormal, and we have that $f_A(e_i) = v_i$ for each $i \in \{1, \dots, n\}$: in view of Proposition 2.12.3 we can conclude from this that the map f_A is orthogonal. □

Our results about diagonalization of self-adjoint linear maps have the following consequence for matrices:

Proposition 2.13.2. *Let $n \in \mathbb{N}$.*

(*i*) *If $A \in \mathrm{M}_n(\mathbb{R})$ is a symmetric matrix, then there exists an orthogonal matrix $C \in \mathrm{O}_n(\mathbb{R})$ such that $C^{\mathrm{t}}AC$ is diagonal.*

(*ii*) *If $A \in \mathrm{M}_n(\mathbb{C})$ is a hermitian matrix, then there exists a unitary matrix $C \in \mathrm{U}_n(\mathbb{C})$ such that C^*AC is diagonal.*

Proof. □

The following result — which is a direct consequence of the Gram–Schmidt orthonormalization process — tells us that an arbitrary matrix in $\mathrm{M}_n(\Bbbk)$ has a factorization as a product of an orthogonal or unitary matrix and an upper triangles matrix. This factorization is usually known as the ***QR-factorization*** and is of great interest in applications. For example, it is very often used when solving linear equations or least squares optimization problems, and it is the basis of the so called *QR-algorithm* to compute the eigenvalues of a real matrix.

Proposition 2.13.3. *Let $n \in \mathbb{N}$ and let $A \in \mathrm{M}_n(\Bbbk)$.*

(*i*) *There are a matrix Q that is orthogonal if $\Bbbk = \mathbb{R}$ and unitary if $\Bbbk = \mathbb{C}$, and an upper triangular matrix $R \in \mathrm{M}_n(\Bbbk)$ such that $A = QR$.*

(*ii*) *If A is invertible, then we can choose these matrices Q and R so that the entries along the diagonal of R are positive real numbers, and in that case both Q and R are uniquely determined by A.*

Proof. □

Bibliography

[And98] George E. Andrews, *The theory of partitions*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1998, Reprint of the 1976 original. MR 1634067

[JVN35] P. Jordan and J. Von Neumann, *On inner products in linear, metric spaces*, Ann. of Math. (2) **36** (1935), no. 3, 719–723. MR 1503247

Notations

Index