Linear Algebra II

Course No. 100 222

Spring 2007

Michael Stoll

With some additions by Ronald van Luijk, 2024

Contents

1. Review of Eigenvalues, Eigenvectors and Characteristic Polynomial

We will heavily use most of what was discussed in *Linear Algebra I*, in particular the following.

(1) Vector spaces
(2) Subspaces and sums of subspaces
(3) Complementary subspaces
(4) Linear maps, as well as their associated kernels and ranks
(5) Bases of vector spaces (all vector spaces have a basis by Zorn's Lemma)
(6) Dimension
(7) The isomorphism $\varphi_B\colon F^n \to V$ associated to a basis B for a vector space V of dimension n over a field F.
(8) Matrices, and elementary operations on them
(9) Matrices associated to linear maps
(10) Determinants
(11) Cramer's rule
(12) Dimension formula for sums of vector spaces
(13) Dimension formula for linear maps
(14) Eigenvalues, eigenvectors, and eigenspaces of endomorphisms
(15) Diagonalizability of endomorphisms

We finished Linear Algebra I discussing eigenvalues and eigenvectors of endomorphisms and square matrices, and the question when they are *diagonalizable*. For your convenience, we repeat here the most relevant definitions and results.

Let V be a finite-dimensional F-vector space, $\dim V = n$, and let $f : V \to V$ be an endomorphism. Then for $\lambda \in F$, the λ-*eigenspace* of f was defined to be

$$E_\lambda(f) = \{v \in V : f(v) = \lambda v\} = \ker(f - \lambda \operatorname{id}_V)\,.$$

The scalar λ is an *eigenvalue* of f if $E_\lambda(f) \neq \{0\}$, i.e., if there is $0 \neq v \in V$ such that $f(v) = \lambda v$. Such a vector v is called an *eigenvector* of f for the eigenvalue λ.

The eigenvalues are exactly the roots (in F) of the *characteristic polynomial* of f,

$$P_f(x) = \det(x \operatorname{id}_V - f)\,,$$

which is a monic polynomial of degree n with coefficients in F.

The *geometric multiplicity* of λ as an eigenvalue of f is defined to be the dimension of the λ-eigenspace, whereas the *algebraic multiplicity* of λ as an eigenvalue of f is defined to be its multiplicity as a root of the characteristic polynomial.

The endomorphism f is said to be *diagonalizable* if there exists a basis for V consisting of eigenvectors of f. The matrix representing f relative to this basis is then a diagonal matrix, with the various eigenvalues appearing on the diagonal.

Since $n \times n$ matrices can be identified with endomorphisms $F^n \to F^n$, all notions and results makes sense for square matrices, too. A matrix $A \in \operatorname{Mat}(n, F)$ is diagonalizable if and only if it is similar to a diagonal matrix, i.e., if there is an invertible matrix $P \in \operatorname{Mat}(n, F)$ such that $P^{-1}AP$ is diagonal.

It is an important fact that the geometric multiplicity of an eigenvalue cannot exceed its algebraic multiplicity. An endomorphism or square matrix is diagonalizable if and only if the sum of the geometric multiplicities of all eigenvalues equals the dimension of the space. This in turn is equivalent to the two conditions (a) the characteristic polynomial is a product of linear factors, and (b) for each eigenvalue, algebraic and geometric multiplicities agree. For example, both conditions are satisfied if P_f is the product of n *distinct* monic linear factors.

Exercises.

(1) Are the vectors $\begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix}$, $\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$, and $\begin{pmatrix} 4 \\ -1 \\ -4 \end{pmatrix}$ linearly independent?

(2) Are the vectors $\begin{pmatrix} 2 \\ -1 \\ -2 \end{pmatrix}$, $\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$, and $\begin{pmatrix} 4 \\ -1 \\ -5 \end{pmatrix}$ linearly independent?

(3) For which $x \in \mathbb{R}$ are the vectors $\begin{pmatrix} 1 \\ x \\ 0 \end{pmatrix}$, $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \\ x \end{pmatrix}$ linearly dependent?

(4) Compute $\det(M)$ for

$$M = \begin{pmatrix} -3 & -1 & 0 & -2 \\ 0 & -2 & 0 & 0 \\ 1 & 0 & -1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

(5) Give the kernel and the image of the map $\mathbb{R}^5 \to \mathbb{R}^3$ given by $x \mapsto Ax$ with

$$A = \begin{pmatrix} 1 & -1 & 1 & 2 & 1 \\ 2 & -1 & 4 & 3 & 3 \\ -1 & 0 & -3 & -1 & 1 \end{pmatrix}.$$

(6) For any square matrix M show that $\mathrm{rk}(M^2) \leq \mathrm{rk}(M)$.

(7) Compute the characteristic polynomial, the complex eigenvalues and the complex eigenspaces of the matrix $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ viewed as a matrix over $\mathbb{C}$.

(8) Find the eigenvalues and eigenspaces of the matrix $A = \begin{pmatrix} 11 & 9 \\ -12 & -10 \end{pmatrix}$. Is A diagonalizable?

(9) Same question for $A = \begin{pmatrix} 3 & 1 \\ -1 & 1 \end{pmatrix}$.

(10) Show that $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ is not diagonalizable.

(11) Consider the map $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $x \mapsto Ax$ where $A = \begin{pmatrix} 3 & 1 \\ -2 & 0 \end{pmatrix}$. Show that $\mathbb{R}^2$ has a basis consisting of eigenvectors of f, and give the matrix of f with respect to this basis. For any positive integer n, give a formula for the matrix representation of f^n, first with respect to the basis of eigenvectors, and then with respect to the standard basis.

(12) Suppose that M is a diagonalizable matrix. Show that $M^2 + M$ is diagonalizable.

(13) Is every 3×3 matrix whose characteristic polynomial is $X^3 - X$ diagonalizable? Is every 3×3 matrix whose characteristic polynomial is $X^3 - X^2$ diagonalizable?

(14) Let the map $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ be the reflection in the plane $x + 2y + z = 0$. What are the eigenvalues and eigenspaces of f?

(15) What is the characteristic polynomial of the rotation map $\mathbb{R}^3 \to \mathbb{R}^3$ which rotates space around the line through the origin and the point $(1, 2, 3)$) by 180 degrees? Same question if we rotate by 90 degrees?

2. Direct Sums of Subspaces

The proof of the Jordan Normal Form Theorem, which is one of our goals, uses the idea to split the vector space V into subspaces on which the endomorphism can be more easily described. In order to make this precise, we introduce the notion of direct sum of linear subspaces of V.

2.1. Definition. Suppose I is an index set and $U_i \subset V$ (for $i \in I$) are linear subspaces of a vector space V satisfying

$$U_j \cap \left(\sum_{i \in I \setminus \{j\}} U_i \right) = \{0\} \tag{1}$$

for all $j \in I$. Then we write $\bigoplus_{i \in I} U_i$ for the subspace $\sum_{i \in I} U_i$ of V, and we call this sum the *(internal) direct sum* of the subspaces U_i. Whenever we use this notation, the hypothesis (1) is implied. If $I = \{1, 2, \ldots, n\}$, then we also write $U_1 \oplus U_2 \oplus \cdots \oplus U_n$.

2.2. Lemma. *Let V be a vector space, and $U_i \subset V$ (for $i \in I$) linear subspaces. Then the following statements are equivalent.*

(1) *Every $v \in V$ can be written uniquely as $v = \sum_{i \in I} u_i$ with $u_i \in U_i$ for all $i \in I$ (and only finitely many $u_i \neq 0$).*
(2) *$\sum_{i \in I} U_i = V$, and for all $j \in I$, we have $U_j \cap \sum_{i \in I \setminus \{j\}} U_i = \{0\}$.*
(3) *If we have any basis B_i of U_i for each $i \in I$, then these bases B_i are pairwise disjoint, and the union $\bigcup_{i \in I} B_i$ forms a basis for V.*
(4) *There exists a basis B_i of U_i for each $i \in I$ such that these bases B_i are pairwise disjoint, and the union $\bigcup_{i \in I} B_i$ forms a basis for V.*

By statement (2) of this lemma, if these conditions are satisfied, then V is the direct sum of the subspaces U_i, that is, we have $V = \bigoplus_{i \in I} U_i$.

Proof. "(1) $\Rightarrow$ (2)": Since every $v \in V$ can be written as a sum of elements of the U_i, we have $V = \sum_{i \in I} U_i$. Now assume that $v \in U_j \cap \sum_{i \neq j} U_i$. This gives two representations of v as $v = u_j = \sum_{i \neq j} u_i$. Since there is only one way of writing v as a sum of u_i's, this is only possible when $v = 0$.
"(2) $\Rightarrow$ (3)": Since the elements of any basis are nonzero, and B_i is contained in U_i for all i, it follows from $U_j \cap \sum_{i \in I \setminus \{j\}} U_i = \{0\}$ that $B_i \cap B_j = \emptyset$ for all $i \neq j$. Let $B = \bigcup_{i \in I} B_i$. Since B_i generates U_i and $\sum_i U_i = V$, we find that B generates V. To show that B is linearly independent, consider a linear combination

$$\sum_{i \in I} \sum_{b \in B_i} \lambda_{i,b} b = 0\,.$$

For any fixed $j \in I$, we can write this as

$$U_j \ni u_j = \sum_{b \in B_j} \lambda_{j,b} b = -\sum_{i \neq j} \sum_{b \in B_i} \lambda_{i,b} b \in \sum_{i \neq j} U_i\,.$$

By (2), this implies that $u_j = 0$. Since B_j is a basis for U_j, this is only possible when $\lambda_{j,b} = 0$ for all $b \in B_j$. Since $j \in I$ was arbitrary, this shows that all coefficients vanish.

"(3) $\Rightarrow$ (4)": This follows by choosing any basis B_i for U_i (see Remark 2.3).

"(4) $\Rightarrow$ (1)": Take a basis B_i for U_i for each $i \in I$ as in (4). Write $v \in V$ as a linear combination of the basis elements in $\bigcup_i B_i$. Since B_i is a basis for U_i, we may write the part of the linear combination coming from B_i as u_i, which yields $v = \sum_i u_i$ with $u_i \in U_i$. To see that the u_i are unique, we note that the u_i can be written as linear combinations of elements in B_i; the sum $v = \sum_i u_i$ is then a linear combination of elements in $\bigcup_i B_i$, which has to be the same as the original linear combination, because $\bigcup_i B_i$ is a basis for V. It follows that indeed all the u_i are uniquely determined. □

2.3. Remark. The proof of the implication (3) $\Rightarrow$ (4) implicitly assumes the existence of a basis B_i for each U_i. The existence of a basis B_i for U_i is clear when U_i is finite-dimensional, but for infinite-dimensional vector spaces this is more subtle. Using Zorn's Lemma, which is equivalent to the Axiom of Choice of Set Theory, one can prove that all vector spaces do indeed have a basis. See Appendix E of *Linear Algebra I, 2020 edition (or later)*. We will use this more often.

2.4. Remark. If U_1 and U_2 are linear subspaces of the vector space V, then statement $V = U_1 \oplus U_2$ is equivalent to U_1 and U_2 being complementary subspaces.

2.5. Lemma. *Suppose V is a vector space with subspaces U and U' such that $V = U \oplus U'$. If $U_1, \ldots, U_r$ are subspaces of U with $U = U_1 \oplus \cdots \oplus U_r$ and $U_1', \ldots, U_s'$ are subspaces of U' with $U' = U_1' \oplus \cdots \oplus U_s'$, then we have*

$$V = U_1 \oplus \cdots \oplus U_r \oplus U_1' \oplus \cdots \oplus U_s'.$$

Proof. This follows most easily from part (1) of Lemma 2.2. □

The converse of this lemma is trivial in the sense that if we have

$$V = U_1 \oplus \cdots \oplus U_r \oplus U_1' \oplus \cdots \oplus U_s',$$

then apparently the $r + s$ subspaces $U_1, \ldots, U_r, U_1', \ldots, U_s'$ satisfy the hypothesis (1), which implies that also the r subspaces $U_1, \ldots, U_r$ satisfy this hypothesis, as well as the subspaces $U_1', \ldots, U_s'$; then also the two subspaces $U = U_1 \oplus \cdots \oplus U_r$ and $U' = U_1' \oplus \ldots \oplus U_s'$ together satisfy the hypothesis and we have $V = U \oplus U'$.

In other words, we may write

$$(U_1 \oplus \cdots \oplus U_r) \oplus (U_1' \oplus \cdots \oplus U_s') = U_1 \oplus \cdots \oplus U_r \oplus U_1' \oplus \cdots \oplus U_s'$$

in the sense that if all the implied conditions of the form (1) are satisfied for one side of the equality, then the same holds for the other side, and the (direct) sums are then equal. In particular, we have $U_1 \oplus (U_2 \oplus \cdots \oplus U_r) = U_1 \oplus \cdots \oplus U_r$.

The following lemma states that if two subspaces intersect each other trivially, then one can be extended to a complementary space of the other. Its proof also suggests how we can do the extension explicitly.

2.6. Lemma. *Let U and U' be subspaces of a finite-dimensional vector space V satisfying $U \cap U' = \{0\}$. Then there exists a subspace $W \subset V$ with $U' \subset W$ that is a complementary subspace of U in V.*

Proof. Let $(u_1, \ldots, u_r)$ be a basis for U and $(v_1, \ldots, v_s)$ a basis for U'. Then by Lemma 2.2 we have a basis $(u_1, \ldots, u_r, v_1, \ldots, v_s)$ for $U + U' = U \oplus U'$. By the Basis Extension Theorem of Linear Algebra 1, we may extend this to a basis $(u_1, \ldots, u_r, v_1, \ldots, v_s, w_1, \ldots, w_t)$ for V. We now let W be the subspace generated by $v_1, \ldots, v_s, w_1, \ldots, w_t$. Then $(v_1, \ldots, v_s, w_1, \ldots, w_t)$ is a basis for W and clearly W contains U'. By Lemma 2.2 we conclude that U and W are complementary spaces. □

Next, we discuss the relation between endomorphisms of V and endomorphisms between the U_i.

2.7. Lemma and Definition. *Let V be a vector space with linear subspaces U_i $(i \in I)$ such that $V = \bigoplus_{i \in I} U_i$. For each $i \in I$, let $f_i : U_i \to U_i$ be an endomorphism. Then there is a unique endomorphism $f : V \to V$ such that $f|_{U_i} = f_i$ for all $i \in I$.*

We call f the *direct sum* of the f_i and write $f = \bigoplus_{i \in I} f_i$.

Proof. Let $v \in V$. Then we have $v = \sum_i u_i$ as above, therefore the only way to define f is by $f(v) = \sum_i f_i(u_i)$. This proves uniqueness. Since the u_i in the representation of v above are unique, f is a well-defined map, and it is clear that f is linear, so f is an endomorphism of V. □

2.8. Remark. If in the situation of Definition 2.7, V is finite-dimensional and we choose a basis B of V that is the concatenation of bases B_i of the U_i, then the matrix representing f relative to B will be a block diagonal matrix, where the diagonal blocks are the matrices representing the f_i relative to the bases B_i of the U_i. In this finite-dimensional case the number of indices $i \in I$ for which U_i is nonzero is finite, and it follows that the characteristic polynomial P_f equals

$$P_f = \prod_{i \in I} P_{f_i}.$$

In particular, we have $\det f = \prod_{i \in I} \det f_i$, and $\operatorname{Tr} f = \sum_{i \in I} \operatorname{Tr} f_i$ for the determinant and the trace.

2.9. Remark. An endomorphism $f : V \to V$ is diagonalizable if and only if the vector space V is the direct sum of the eigenspaces of f.

2.10. Lemma. *Let V be a vector space with linear subspaces U_i $(i \in I)$ such that $V = \bigoplus_{i \in I} U_i$. Let $f : V \to V$ be an endomorphism. Then there are endomorphisms $f_i : U_i \to U_i$ for $i \in I$ such that $f = \bigoplus_{i \in I} f_i$ if and only if each U_i is* invariant under f (or *f-invariant*), i.e., $f(U_i) \subset U_i$.

Proof. If $f = \bigoplus_i f_i$, then $f_i = f|_{U_i}$, hence $f(U_i) = f|_{U_i}(U_i) = f_i(U_i) \subset U_i$. Conversely, suppose that $f(U_i) \subset U_i$. Then we can define $f_i : U_i \to U_i$ to be the restriction of f to U_i; it is then clear that f_i is an endomorphism of U_i and that f equals $\bigoplus_i f_i$, as the two coincide on all the subspaces U_i, which together generate V. □

2.11. Example. Consider the linear map $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ that sends (x,y,z) to (y,z,x). This describes rotation over $2\pi/3$ around the line $U_1 = L(a)$ with $a = (1,1,1)$. The line U_1 is point-wise fixed by f, so it is f-invariant. The orthogonal complement $U_2 = a^\perp$ is an f-invariant plane, so we have $\mathbb{R}^3 = U_1 \oplus U_2$ and $f = f_1 \oplus f_2$ with $f_i = f|_{U_i}$. The vector $v_1 = a$ gives a basis for the line U_1. The vectors $v_2 = (1,-1,0)$ and $v_3 = (-1,0,1)$ form a basis (v_2, v_3) for the plane U_2. Putting these two bases together, we obtain a basis $B = (v_1, v_2, v_3)$ for $\mathbb{R}^3$ and by the Remark 2.8, the associated matrix $[f]_B^B$ is a block diagonal matrix. Indeed, from $f(v_1) = v_1$ and $f(v_2) = v_3$ and $f(v_3) = -v_2 - v_3$ we find

$$[f]_B^B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}.$$

Recall that if V is a vector space over a field F and $f\colon V \to V$ is an endomorphism, then we write

$$f^n = \underbrace{f \circ f \circ \cdots \circ f}_{n}.$$

More generally, if $p = \sum_{i=0}^d a_i x^i \in F[x]$ is a polynomial, then we define $p(f) = \sum_{i=0}^d a_i f^i$. Note that for two polynomials $p, q \in F[x]$, we have $(p \cdot q)(f) = p(f) \circ q(f)$. We now come to a relation between splittings of f as a direct sum and polynomials that vanish on f, that is, polynomials p with $p(f) = 0$ (where 0 denotes the zero endomorphism). We will see later that this includes the characteristic and the minimal polynomial of f (see Theorem 3.1 and Lemma 3.4).

We call two polynomials $p_1(x)$ and $p_2(x)$ *coprime* if there are polynomials $a_1(x)$ and $a_2(x)$ such that $a_1(x)p_1(x) + a_2(x)p_2(x) = 1$.

2.12. Lemma. *Let V be a vector space and $f : V \to V$ an endomorphism. Let $p(x) = p_1(x)p_2(x)$ be a polynomial such that $p(f) = 0$ and such that $p_1(x)$ and $p_2(x)$ are coprime. Let $U_i = \ker\big(p_i(f)\big)$, for $i = 1, 2$. Then $V = U_1 \oplus U_2$ and the U_i are f-invariant. In particular, $f = f_1 \oplus f_2$, where $f_i = f|_{U_i}$. Moreover, we have $U_1 = \operatorname{im}\big(p_2(f)\big)$ and $U_2 = \operatorname{im}\big(p_1(f)\big)$.*

Proof. Set $K_1 = \operatorname{im}\big(p_2(f)\big)$ and $K_2 = \operatorname{im}\big(p_1(f)\big)$. We first show that $K_i \subset U_i$ for $i = 1, 2$. Let $v \in K_1 = \operatorname{im}\big(p_2(f)\big)$, so $v = \big(p_2(f)\big)(u)$ for some $u \in V$. Then

$$\big(p_1(f)\big)(v) = \big(p_1(f)\big)\Big(\big(p_2(f)\big)(u)\Big) = \big(p_1(f) \circ p_2(f)\big)(u) = \big(p(f)\big)(u) = 0,$$

so $K_1 = \operatorname{im}\big(p_2(f)\big) \subset \ker\big(p_1(f)\big) = U_1$. The statement for $i = 2$ follows by symmetry.

Now we show that $U_1 \cap U_2 = \{0\}$. So let $v \in U_1 \cap U_2$. Then $\big(p_1(f)\big)(v) = \big(p_2(f)\big)(v) = 0$. Let $a_1(x), a_2(x)$ be such that $a_1(x)p_1(x) + a_2(x)p_2(x) = 1$. Using

$$\mathrm{id}_V = 1(f) = \big(a_1(x)p_1(x) + a_2(x)p_2(x)\big)(f) = a_1(f) \circ p_1(f) + a_2(f) \circ p_2(f),$$

we see that

$$v = \big(a_1(f)\big)\Big(\big(p_1(f)\big)(v)\Big) + \big(a_2(f)\big)\Big(\big(p_2(f)\big)(v)\Big) = \big(a_1(f)\big)(0) + \big(a_2(f)\big)(0) = 0.$$

Next, we show that $K_1 + K_2 = V$. Using the same relation above, and the fact that $p_i(f)$ and $a_i(f)$ commute, we find for $v \in V$ arbitrary that

$$v = \big(p_1(f)\big)\Big(\big(a_1(f)\big)(v)\Big) + \big(p_2(f)\big)\Big(\big(a_2(f)\big)(v)\Big) \in \operatorname{im}\big(p_1(f)\big) + \operatorname{im}\big(p_2(f)\big).$$

These statements together imply that $K_i = U_i$ for $i = 1, 2$, and $V = U_1 \oplus U_2$. Indeed, let $v \in U_1$. We can write $v = v_1 + v_2$ with $v_i \in K_i$. Then $U_1 \ni v - v_1 = v_2 \in U_2$, but $U_1 \cap U_2 = \{0\}$, so $v = v_1 \in K_1$.

Finally, we have to show that U_1 and U_2 are f-invariant. So let (e.g.) $v \in U_1$. Since f commutes with $p_1(f)$, we have

$$\big(p_1(f)\big)\big(f(v)\big) = \big(p_1(f) \circ f\big)(v) = \big(f \circ p_1(f)\big)(v) = f\Big(\big(p_1(f)\big)(v)\Big) = f(0) = 0\,,$$

(since $v \in U_1 = \ker\big(p_1(f)\big)$), hence $f(v) \in U_1$ as well. □

2.13. Example. Consider the linear map $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ from Example 2.11. Because $f^3 = \mathrm{id}$, we find that the polynomial $p = x^3 - 1$ vanishes on f, that is, we have $p(f) = 0$. We can factor p as $p = p_1 p_2$ with $p_1 = x - 1$ and $p_2 = x^2 + x + 1$. The polynomials p_1 and p_2 are coprime, as we have

$$1 = -\tfrac{1}{3}(x+2) \cdot p_1 + \tfrac{1}{3} \cdot p_2;$$

it also follows from Lemma 2.15. We recover U_1 and U_2 from Example 2.11 as follows. The linear map $p_1(f) = f - \mathrm{id}$ sends (x, y, z) to $(y - x, z - y, x - z)$, so we find $\ker\big(p_1(f)\big) = L((1,1,1)) = U_1$. The linear map $p_2(f) = f \circ f + f + \mathrm{id}$ sends (x, y, z) to $(x + y + z, x + y + z, x + y + z)$, so we find $\ker\big(p_2(f)\big) = U_2$.

2.14. Proposition. *Let V be a vector space and $f : V \to V$ an endomorphism. Let $p(x) = p_1(x)p_2(x) \cdots p_k(x)$ be a polynomial such that $p(f) = 0$ and such that the factors $p_i(x)$ are coprime in pairs. Let $U_i = \ker\big(p_i(f)\big)$. Then $V = U_1 \oplus \cdots \oplus U_k$ and the U_i are f-invariant. In particular, $f = f_1 \oplus \cdots \oplus f_k$, where $f_i = f|_{U_i}$.*

Proof. We proceed by induction on k. The case $k = 1$ is trivial. So let $k \geq 2$, and denote $q(x) = p_2(x) \cdots p_k(x)$. Then I claim that $p_1(x)$ and $q(x)$ are coprime. To see this, note that by assumption, we can write, for $i = 2, \ldots, k$,

$$a_i(x)p_1(x) + b_i(x)p_i(x) = 1\,.$$

Multiplying these equations, we obtain

$$A(x)p_1(x) + b_2(x) \cdots b_k(x)q(x) = 1\,;$$

note that all the terms except $b_2(x) \cdots b_k(x)q(x)$ that we get when expanding the product of the left hand sides contains a factor $p_1(x)$.

We can then apply Lemma 2.12 to $p(x) = p_1(x)q(x)$ and find that $V = U_1 \oplus U'$ and $f = f_1 \oplus f'$ with $U_1 = \ker\big(p_1(f)\big)$, $f_1 = f|_{U_1}$, and $U' = \ker\big(q(f)\big)$, $f' = f|_{U'}$. In particular, $q(f') = 0$. By induction, we then know that $U' = U_2 \oplus \cdots \oplus U_k$ with $U_j = \ker\big(p_j(f')\big)$ and $f' = f_2 \oplus \cdots \oplus f_k$, where $f_j = f'|_{U_j}$, for $j = 2, \ldots, k$. Finally, $\ker\big(p_j(f')\big) = \ker\big(p_j(f)\big)$ (since the latter is contained in U') and $f_j = f'|_{U_j} = f|_{U_j}$, so that we obtain the desired conclusion from Lemma 2.5. □

The following little lemma about polynomials is convenient if we want to apply Lemma 2.12.

2.15. Lemma. *If $p(x)$ is a polynomial (over F) and $\lambda \in F$ such that $p(\lambda) \neq 0$, then $(x-\lambda)^m$ and $p(x)$ are coprime for all $m \geq 0$.*

PROOF. For $m = 0$, this is trivial. Next, we consider $m = 1$. Let

$$q(x) = \frac{p(x)}{p(\lambda)} - 1\,;$$

this is a polynomial such that $q(\lambda) = 0$. Therefore, we can write $q(x) = (x-\lambda)r(x)$ with some polynomial $r(x)$. This gives us

$$-r(x)(x-\lambda) + \frac{1}{p(\lambda)}\,p(x) = 1\,.$$

Now for general $m \geq 1$, taking the mth power on both sides, we obtain an equation

$$\big(-r(x)\big)^m (x-\lambda)^m + a(x)p(x) = 1\,.$$

□

Exercises.

(1) Let $\phi\colon \mathbb{R}^3 \to \mathbb{R}^3$ be a rotation around the line through the origin and the point $(1,1,-1)$ by 120 degrees. Decompose $\mathbb{R}^3$ as a direct sum of two subspaces that are each stable under ϕ.

(2) Consider the vector space $V = \mathbb{R}^3$ with the linear map $\phi\colon V \to V$ given by the matrix

$$\begin{pmatrix} -1 & 0 & 1 \\ -2 & -1 & 1 \\ -3 & -1 & 2 \end{pmatrix}.$$

Decompose $\mathbb{R}^3$ as a direct sum of two non-trivial subspaces that are each stable under ϕ. [Theorem 3.1 (Cayley-Hamilton) states that for a square matrix A with characteristic polynomial P_A, we have $P_A(A) = 0$. You can verify and then use this for this specific matrix.]

(3) Same question for

$$\begin{pmatrix} 0 & 1 & 1 \\ 5 & -4 & -3 \\ -6 & 6 & 5 \end{pmatrix}.$$

(4) Consider the vector space $V = \mathbb{R}^4$ with the linear map $\phi\colon V \to V$ that permutes the standard basis vectors in a cycle of length 4. Decompose $\mathbb{R}^4$ into a direct sum of 3 subspaces that are all stable under ϕ.

(5) An endomorphism f of a vector space V is said to be a *projection* if $f^2 = f$. Suppose f is such a projection.
   (a) Show that the image of f is equal to the kernel of $f - \mathrm{id}_V$, i.e., we have $\operatorname{im} f = E_1$ with $E_1 = \ker(f - \mathrm{id}_V)$. Note that if E_1 is nonzero, then 1 is an eigenvalue for f and E_1 is the corresponding eigenspace.
   (b) Show that V is the direct sum of the kernel E_0 of f and the space E_1.
   (c) Show that $f = f_0 \oplus f_1$ where f_0 is the zero-map on E_0 and f_1 is the identity map on E_1.

(6) An endomorphism f of a vector space V is said to be a *reflection* if f^2 is the identity on V. Suppose f is such a reflection. Show that V is the direct sum of two subspaces U and W for which $f = \mathrm{id}_U \oplus (-\mathrm{id}_W)$.

3. The Cayley-Hamilton Theorem and the Minimal Polynomial

Let $A \in \mathrm{Mat}(n, F)$. We know that $\mathrm{Mat}(n, F)$ is an F-vector space of dimension n^2. Therefore, the elements $I, A, A^2, \dots, A^{n^2}$ cannot be linearly independent (because their number exceeds the dimension). If we define $p(A)$ in the obvious way for p a polynomial with coefficients in F (as we already did in the previous chapter), then we can deduce that there is a (non-zero) polynomial p of degree at most n^2 such that $p(A) = 0$ (0 here is the zero matrix). In fact, much more is true.

Consider a diagonal matrix $D = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. (This notation is supposed to mean that λ_j is the (j, j) entry of D; the off-diagonal entries are zero, of course.) Its characteristic polynomial is

$$P_D(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n)\,.$$

Since the diagonal entries are roots of P_D, we also have $P_D(D) = 0$. More generally, consider a diagonalizable matrix A. Then there is an invertible matrix Q such that $D = Q^{-1}AQ$ is diagonal. Since (Exercise!) $p(Q^{-1}AQ) = Q^{-1}p(A)Q$ for p a polynomial, we find

$$0 = P_D(D) = Q^{-1}P_D(A)Q = Q^{-1}P_A(A)Q \quad\Longrightarrow\quad P_A(A) = 0\,.$$

(Recall that $P_A = P_D$ — similar matrices have the same characteristic polynomial.)

The following theorem states that this is true for *all* square matrices (or endomorphisms of finite-dimensional vector spaces).

3.1. Theorem (Cayley-Hamilton). *Let* $A \in \mathrm{Mat}(n, F)$. *Then* $P_A(A) = 0$.

Proof. Here is a simple, but **wrong** “proof”. By definition, $P_A(x) = \det(xI - A)$, so, plugging in A for x, we have $P_A(A) = \det(AI - A) = \det(A - A) = \det(0) = 0$. (Exercise: find the mistake!)

For the correct proof, we need to consider matrices whose entries are polynomials. Since polynomials satisfy the field axioms except for the existence of inverses, we can perform all operations that do not require divisions. This includes addition, multiplication and determinants; in particular, we can use the adjugate matrix.

Let $B = xI - A$, then $\det(B) = P_A(x)$. Let $\tilde{B}$ be the adjugate matrix; then we still have $\tilde{B}B = \det(B)I$. The entries of $\tilde{B}$ come from determinants of $(n-1)\times(n-1)$ submatrices of B, therefore they are polynomials of degree at most $n - 1$. We can then write

$$\tilde{B} = x^{n-1}B_{n-1} + x^{n-2}B_{n-2} + \dots + xB_1 + B_0\,,$$

and we have the equality (of matrices with polynomial entries)

$$(x^{n-1}B_{n-1} + x^{n-2}B_{n-2} + \dots + B_0)(xI - A) = P_A(x)I = (x^n + b_{n-1}x^{n-1} + \dots + b_0)I\,,$$

where we have set $P_A(x) = x^n + b_{n-1}x^{n-1} + \dots + b_0$. Expanding the left hand side and comparing coefficients of like powers of x, we find the relations

$$B_{n-1} = I, \quad B_{n-2} - B_{n-1}A = b_{n-1}I, \quad \dots, \quad B_0 - B_1A = b_1I, \quad -B_0A = b_0I\,.$$

We multiply these from the right by A^n, A^{n-1}, …, A, I, respectively, and add:

$$\begin{array}{lcll}
B_{n-1}A^n & & & = A^n \\
B_{n-2}A^{n-1} & - & B_{n-1}A^n & = b_{n-1}A^{n-1} \\
\vdots & & \vdots & \quad\vdots \\
B_0A & - & B_1A^2 & = b_1A \\
 & - & B_0A & = b_0I \\
\hline
 & 0 & & = P_A(A)
\end{array}$$

□

3.2. Remarks.

(1) The reason why we cannot simply plug in A for x in the identity
$$\tilde{B} \cdot (xI - A) = P_A(x)I$$
is that whereas x (as a scalar) commutes with the matrices occurring as coefficients of powers of x, it is not a priori clear that A does so, too.

(2) Another idea of proof (and maybe easier to grasp) is to say that a 'generic' matrix is diagonalizable (if we assume F to be algebraically closed...), hence the statement holds for 'most' matrices. Since it is just a bunch of polynomial relations between the matrix entries, it then must hold for all matrices. This can indeed be turned into a proof, but unfortunately, this requires rather advanced tools from algebra.

(3) Of course, the statement of the theorem remains true for endomorphisms. Let $f : V \to V$ be an endomorphism of the finite-dimensional F-vector space V, then $P_f(f) = 0$ (which is the zero endomorphism in this case). For evaluating the polynomial at f, we have to interpret f^n as the n-fold composition $f \circ f \circ \cdots \circ f$, and $f^0 = \mathrm{id}_V$.

Our next goal is to define the *minimal polynomial* of a matrix or endomorphism, as the monic polynomial of smallest degree that has the matrix or endomorphism as a "root". However, we need to know a few more facts about polynomials in order to see that this definition makes sense.

3.3. Lemma (Polynomial Division). *Let f and g be polynomials with coefficients in F, with g monic. Then there are unique polynomials q and r with coefficients in F such that $r = 0$ or $\deg(r) < \deg(g)$ and such that*
$$f = qg + r\,.$$

PROOF. We first prove existence, by induction on the degree of f. If $f = 0$ or $\deg(f) < \deg(g)$, then we take $q = 0$ and $r = f$. So we now assume that $m = \deg(f) \geq \deg(g) = n$, $f = a_m x^m + \cdots + a_0$. Let $f' = f - a_m x^{m-n} g$, then (since $g = x^n + \dots$) $\deg(f') < \deg(f)$. By the induction hypothesis, there are q' and r such that $\deg(r) < \deg(g)$ or $r = 0$ and such that $f' = q'g + r$. Then $f = (q' + a_m x^{m-n})g + r$. (This proof leads to the well-known algorithm for polynomial long division.)

As to uniqueness, suppose we have $f = qg + r = q'g + r'$, with r and r' both of degree less than $\deg(g)$ or zero. Then
$$(q - q')g = r' - r\,.$$
If $q \neq q'$, then the degree of the left hand side is at least $\deg(g)$, but the degree of the right hand side is smaller, hence this is not possible. So $q = q'$, and therefore $r = r'$, too. □

Taking $g = x - \alpha$, this provides a different proof for case $k = 1$ of Example 8.4 of *Linear Algebra I, 2015 edition (or later).*

3.4. Lemma and Definition. *Let $A \in \mathrm{Mat}(n, F)$. Among all monic polynomials p with coefficients in F satisfying $p(A) = 0$, there is a unique polynomial M_A of minimal degree. If p is any polynomial with coefficients in F satisfying $p(A) = 0$, then p is divisible by M_A (as a polynomial).*

This polynomial M_A is called the *minimal* (or *minimum*) *polynomial* of A. Similarly, we define the minimal polynomial M_f of an endomorphism f of a finite-dimensional vector space.

Proof. It is clear that monic polynomials p with coefficients in F satisfying $p(A) = 0$ exist (by the Cayley-Hamilton Theorem 3.1, we can take $p = P_A$). So there will be such a polynomial of minimal degree. Now assume p and p' were two such monic polynomials of (the same) minimal degree with $p(A) = p'(A) = 0$. Then we would have $(p - p')(A) = p(A) - p'(A) = 0$. If $p \neq p'$, then we can divide $p - p'$ by its leading coefficient, leading to a monic polynomial q of smaller degree than p and p' with $q(A) = 0$, contradicting the minimality of the degree.

Now let p be any polynomial such that $p(A) = 0$. By Lemma 3.3, there are polynomials q and r, with $\deg(r) < \deg(M_A)$ or $r = 0$, such that $p = qM_A + r$. Plugging in A, we find that

$$0 = p(A) = q(A)M_A(A) + r(A) = q(A) \cdot 0 + r(A) = r(A)\,.$$

If $r \neq 0$, then $\deg(r) < \deg(M_A)$, but the degree of M_A is the minimal possible degree for a polynomial that vanishes on A, so we have a contradiction. Therefore $r = 0$ and hence $p = qM_A$. □

3.5. Remark. In *Introductory Algebra,* you will learn that the set of polynomials as discussed in the lemma forms an *ideal* in the polynomial ring $F[x]$ of all polynomials with coefficients in F, and that this ring is a *principal ideal domain,* which means that every ideal consists of the multiples of some fixed polynomial. The proof is exactly the same as for the lemma.

3.6. Warning. A priori, the minimal polynomial M_A of a matrix A depends on the field F we consider it over. For example, if A is a real matrix, then its minimal polynomial has minimal degree among all *real* polynomials $p \in \mathbb{R}[x]$ with $p(A) = 0$; if we consider the same matrix A as a complex matrix, then one might wonder if there are *complex* polynomials $p \in \mathbb{C}[x]$ with smaller degree. Exercise 11 shows that this is not the case, at least not for the fields $\mathbb{R}$ and $\mathbb{C}$ as in this example. With some more algebra, one can show that in fact the minimal polynomial of A is independent of the field F in general, which is why it is not reflected in the notation M_A.

By Lemma 3.4, the minimal polynomial divides the characteristic polynomial. As a simple example, consider the identity matrix I_n. Its characteristic polynomial is $(x - 1)^n$, whereas its minimal polynomial is $x - 1$. In some sense, this is typical, as the following result shows.

3.7. Proposition. *Let $A \in \mathrm{Mat}(n, F)$ and $\lambda \in F$. If λ is a root of the characteristic polynomial of A, then it is also a root of the minimal polynomial of A. In other words, both polynomials have the same linear factors.*

PROOF. If $P_A(\lambda) = 0$, then λ is an eigenvalue of A, so there is $0 \neq v \in F^n$ such that $Av = \lambda v$. Setting $M_A(x) = a_m x^m + \cdots + a_0$, we find

$$0 = M_A(A)v = \sum_{j=0}^{m} a_j A^j v = \sum_{j=0}^{m} a_j \lambda^j v = M_A(\lambda)v \,.$$

(Note that the terms in this chain of equalities are vectors.) Since $v \neq 0$, this implies $M_A(\lambda) = 0$.

By Lemma 3.4, we know that each root of M_A is a root of P_A, and we have just shown the converse. So both polynomials have the same linear factors. □

3.8. Remark. If F is algebraically closed (i.e., every non-zero polynomial is a product of linear factors), this shows that P_A is a multiple of M_A, and M_A^k is a multiple of P_A when k is large enough. In fact, the latter statement is true for general fields F (and can be interpreted as saying that both polynomials have the same irreducible factors). For the proof, one replaces F by a larger field F' such that both polynomials split into linear factors over F'. That this can always be done is shown in *Introductory Algebra.* See Exercise 12 for the case $F = \mathbb{R}$.

One nice property of the minimal polynomial is that it provides another criterion for diagonalizability.

3.9. Proposition. *Let $A \in \operatorname{Mat}(n, F)$. Then A is diagonalizable if and only if its minimal polynomial M_A is a product of distinct monic linear factors.*

PROOF. First assume that A is diagonalizable. It is easy to see that similar matrices have the same minimal polynomial (Exercise 3), so we can as well assume that A is already diagonal. But for a diagonal matrix, the minimal polynomial is just the product of factors $x - \lambda$, where λ runs through the distinct diagonal entries. (It is the monic polynomial of smallest degree that has all diagonal entries as roots.)

Conversely, assume that $M_A(x) = (x - \lambda_1) \cdots (x - \lambda_m)$ with $\lambda_1, \dots, \lambda_m \in F$ distinct. The polynomials $q_i = x - \lambda_i$ (with $1 \leq i \leq m$) are pairwise coprime, so by Proposition 2.14 the eigenspaces

$$U_i = E_{\lambda_i}(A) = \ker(A - \lambda_i I) = \ker q_i(A)$$

satisfy $F^n = U_1 \oplus \cdots \oplus U_m$. It then follows from Remark 2.9 that A is diagonalisable. □

3.10. Example. Consider the matrix

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

Is it diagonalizable?

Its characteristic polynomial is clearly $P_A(x) = (x-1)^3$, so its minimal polynomial must be $(x-1)^m$ for some $m \leq 3$. Since $A - I \neq 0$, we find $m > 1$ (in fact, $m = 3$), hence A is not diagonalizable.

On the other hand, the matrix (for $F = \mathbb{R}$, say)

$$B = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}$$

has $M_B(x) = P_B(x) = (x-1)(x-4)(x-6)$; therefore, B is diagonalizable. Exercise: what happens for fields F of small characteristic?

3.11. Remark. Let $f\colon V \to V$ be an endomorphism of a finite-dimensional vector space V with basis B. Then f is diagonalizable if and only if the matrix $A = [f]^B_B$ is. Furthermore, the characteristic and minimal polynomial of f are the same as those of the matrix A. Therefore, Lemma 3.4 and Propositions 3.7 and 3.9 also hold for f instead of A. (See also part (3) of Remark 3.2.)

3.12. Corollary. *Let $f\colon V \to V$ be a diagonalizable endomorphism of a finite-dimensional vector space V. Let $U \subset V$ be an f-invariant subspace. Then the restriction $f|_U$ is also diagonalizable.*

Proof. By Proposition 3.9, the minimal polynomial M_f of f is the product of distinct linear factors. The endomorphism $M_f(f|_U)$ is the restriction to U of $M_f(f) = 0$, so the minimal polynomial of $f|_U$ divides M_f by Lemma 3.4, and is therefore also the product of distinct linear factors. Proposition 3.9 then implies that $f|_U$ is diagonalizable. □

Exercises.

(1) What is the remainder when one divides the polynomial $x^5 + x$ by $x^2 + 1$?
(2) Give the minimal polynomial and the characteristic polynomial of the matrices
$$\begin{pmatrix} 2 & -3 & 3 \\ 3 & -4 & 3 \\ 3 & -3 & 2 \end{pmatrix}, \qquad \begin{pmatrix} 0 & -1 & 3 \\ 1 & -2 & 3 \\ 3 & -3 & 2 \end{pmatrix}.$$
(3) Let $A, P \in \mathrm{Mat}(n, F)$ be square matrices, with P invertible. Show that the matrices A and PAP^{-1} have the same minimal polynomial.
(4) Suppose that a 2×2 matrix A has two distinct eigenvalues λ and μ. Show that the image of the matrix $A - \lambda I$ is the eigenspace with eigenvalue μ.
(5) Is the matrix $\begin{pmatrix} 0 & 0 & -3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ diagonalizable over $\mathbb{R}$? And over $\mathbb{C}$?
(6) If $f\colon \mathbb{R}^3 \to \mathbb{R}^3$ is the projection on a plane through the origin, what is the minimum polynomial of f? What is the minimum polynomial of reflection in a plane through the origin?
(7) Compute the characteristic polynomial of the matrix
$$A = \begin{pmatrix} 1 & -9 & 4 \\ 1 & -4 & 1 \\ 1 & -7 & 3 \end{pmatrix}.$$
Compute A^3 (use Cayley-Hamilton!)
(8) Let V be the 4 dimensional vector space of polynomial functions $\mathbb{R} \to \mathbb{R}$ of degree at most 3. Let $T\colon V \to V$ be the map that sends a polynomial p to its derivative $T(p) = p'$. Show that T is a linear map. Is T diagonalizable?
(9) For each $\alpha \in \mathbb{R}$, determine the characteristic and minimal polynomials of
$$A_\alpha = \begin{pmatrix} 1-\alpha & \alpha & 0 \\ 2-\alpha & \alpha-1 & \alpha \\ 0 & 0 & -1 \end{pmatrix}.$$
For which values of α is A_α diagonalizable?

(10) Let M be a square matrix satisfying $M^3 = M$. What can you say about the eigenvalues of M? Show that M is diagonalizable.
(11) Let A be a real square matrix. Show that its minimal polynomial as a real matrix is the same as its minimal polynomial as a complex matrix.
(12) Let A be a real square matrix. Suppose that $f \in \mathbb{R}[x]$ is a quadratic polynomial without real roots that divides the characteristic polynomial P_A of A. Show that f also divides the minimal polynomial M_A of A.

4. The Structure of Nilpotent Endomorphisms

4.1. Definition. A matrix $A \in \mathrm{Mat}(n, F)$ is said to be *nilpotent,* if $A^m = 0$ for some $m \geq 1$. Similarly, if V is a finite-dimensional vector space and $f : V \to V$ is an endomorphism, then f is said to be *nilpotent* if for some $m \geq 1$ we have

$$f^m = \underbrace{f \circ f \circ \cdots \circ f}_{m \text{ times}} = 0.$$

It follows that the minimal polynomial of A or f is of the form x^m, where m is the smallest number that has the property required in the definition.

4.2. Proposition. *A nilpotent matrix or endomorphism of a finite-dimensional vector space is diagonalizable if and only if it is zero.*

PROOF. The minimal polynomial is x^m. Proposition 3.9 then implies that the matrix or endomorphism is diagonalizable if and only if $m = 1$. But then the minimal polynomial is x, which means that the matrix or endomorphism is zero. □

Theorem 4.8 tells us more about the structure of nilpotent endomorphisms. It is the main ingredient to proving the existence of the Jordan Normal Form. We first state some lemmas that will be useful for the proof of Theorem 4.8.

4.3. Lemma. *Let V be a vector space and $f \colon V \to V$ an endomorphism. Suppose $m > 0$ is an integer such that $f^m = 0$. If for each $j \in \{1, 2, \ldots, m\}$ we have a complementary subspace X_j of $\ker f^{j-1}$ inside $\ker f^j$, then we have*

$$V = X_1 \oplus X_2 \oplus \cdots \oplus X_m.$$

PROOF. Note that we have $\ker f^m = V$ and $\ker f^0 = \{0\}$. For all $j \in \{1, \ldots, m\}$, we have $\ker f^j = \ker f^{j-1} \oplus X_j$, so we find

$$\begin{aligned} V &= \ker f^m = \ker f^{m-1} \oplus X_m = (\ker f^{m-2} \oplus X_{m-1}) \oplus X_m = \\ &= \ker f^{m-2} \oplus X_{m-1} \oplus X_m = \cdots = \ker f^0 \oplus X_1 \oplus X_2 \oplus \ldots \oplus X_m = \\ &= X_1 \oplus X_2 \oplus \ldots \oplus X_m. \end{aligned}$$

□

4.4. Lemma. *Let $f\colon V \to W$ be a linear map of vector spaces, and $X \subset V$ and $Y \subset W$ subspaces such that $X \cap f^{-1}(Y) = \{0\}$. Then f restricts to an injective map $X \hookrightarrow W$, and we have $f(X) \cap Y = \{0\}$.*

Proof. The kernel of the restriction $\tilde{f} = f|_X\colon X \to W$ satisfies

$$\ker \tilde{f} = X \cap \ker f \subset X \cap f^{-1}(Y) = \{0\},$$

so $\tilde{f}$ is injective. The last part of the statement follows from the fact that, more generally, the restriction $X \cap f^{-1}(Y) \to f(X) \cap Y$ of f is surjective (the verification of this fact is left as an exercise for the reader). □

4.5. Lemma. *Let V be a vector space and $f\colon V \to V$ an endomorphism. Let $j \geq 1$ be an integer. If X is a complementary space of $\ker f^j$ inside $\ker f^{j+1}$, then f restricts to an injective map $X \hookrightarrow \ker f^j$ and we have $f(X) \cap \ker f^{j-1} = \{0\}$.*

Proof. Note that for every $i \geq 0$, we have $f^{-1}(\ker f^i) = \ker f^{i+1}$. For $i = j$, this implies that f restricts to a linear map $f'\colon \ker f^{j+1} \to \ker f^j$. For $i = j-1$ and $Y = \ker f^{j-1}$, it implies $f^{-1}(Y) = \ker f^j$, so we get

$$X \cap f'^{-1}(Y) \subset X \cap f^{-1}(Y) = \{0\}.$$

Hence, the statement follows directly from Lemma 4.4, applied to f', X, and Y. □

4.6. Remark. In terms of quotient spaces, Lemma 4.4 can be phrased by saying that f induces an injective map $V/f^{-1}(Y) \to W/Y$, which follows from one of the isomorphism theorems (analogous to those from group theory), applied to the linear map $V \to W/Y$ with kernel $f^{-1}(Y)$. Similarly, Lemma 4.5 can be phrased by saying that f induces an injective map $\ker f^{j+1}/\ker f^j \hookrightarrow \ker f^j/\ker f^{j-1}$.

4.7. Remark. Lemmas 2.6 and 4.5 together show that, under the conditions of Lemma 4.5, we can extend $f(X)$ to a complementary space X' of $\ker f^{j-1}$ inside $\ker f^j$. Then f restricts to an injective map $X \hookrightarrow X'$, and we can apply Lemma 4.5 to X' (if $j > 1$). If moreover, $m > 0$ is an integer such that $f^m = 0$, then this allows us to recursively define a sequence $X_m, \ldots, X_2, X_1$ of subspaces as in Lemma 4.3 with the extra property that f restricts to an injective map $X_j \hookrightarrow X_{j-1}$ for $1 < j \leq m$. This is the main idea for the proof of Theorem 4.8, which also keeps track of bases for the subspaces.

4.8. Theorem. *Let V be an F-vector space, $\dim V = n$, and let $f : V \to V$ be a nilpotent endomorphism. Then V has a basis $(v_1, v_2, \ldots, v_n)$ such that $f(v_j)$ is either zero or v_{j+1}.*

Proof. Let m be a positive integer such that $f^m = 0$. In each of m steps, numbered $j = m, m-1, \ldots, 2, 1$, we will construct an integer t_j and vectors $w_{j1}, \ldots, w_{jt_j} \in \ker f^j$ such that the elements

$$\left(f^{k-j}(w_{kl})\right)_{\substack{j \leq k \leq m \\ 1 \leq l \leq t_k}} \tag{2}$$

form a basis for a complementary space X_j of $\ker f^{j-1}$ inside $\ker f^j$. For $j = m$, we take any basis $(w_{m1}, \ldots, w_{mt_m})$ for a complementary subspace X_m of $\ker f^{m-1}$ inside $\ker f^m = V$. Assume $1 \leq j < m$ and suppose that we have already constructed integers and vectors as above in all steps $m, m-1, \ldots, j+1$. Then the elements

$$(3) \qquad \left(f^{k-(j+1)}(w_{kl})\right)_{\substack{j+1\le k\le m\\ 1\le l\le t_k}}$$

form a basis for a complementary space X_{j+1} of $\ker f^j$ inside $\ker f^{j+1}$. The map f restricts to an injective map $X_{j+1} \to \ker f^j$ by Lemma 4.5. This implies that the images

$$(4) \qquad \left(f^{k-j}(w_{kl})\right)_{\substack{j+1\le k\le m\\ 1\le l\le t_k}}$$

of the elements in (3) form a basis for the subspace $f(X_{j+1}) \subset \ker f^j$ (for linear independence, see Lemma 7.13 of *Linear Algebra I, 2015 edition (or later)*), which satisfies $f(X_{j+1}) \cap \ker f^{j-1} = \{0\}$, again by Lemma 4.5. By Lemma 2.6 we can extend the basis (4) for $f(X_{j+1})$ to a basis for a complementary subspace X_j of $\ker f^{j-1}$ inside $\ker f^j$; we denote the added basis vectors by $w_{j1}, w_{j2}, \ldots, w_{jt_j}$. Adding these elements to (4) gives (2), with the new elements corresponding to $k = j$.

By Lemma 4.3, we have $V = X_1 \oplus X_2 \oplus \ldots \oplus X_m$, so the bases (2) for the X_j are disjoint and their union forms a basis for V (see Lemma 2.2). Writing $i = k - j$, this union consists of the elements

$$(5) \qquad \left(f^{i}(w_{kl})\right)_{\substack{1\le k\le m\\ 1\le l\le t_k\\ 0\le i<k}}.$$

Note that for any indices $1 \le k \le m$ and $1 \le l \le t_k$, we have $w_{kl} \in \ker f^k$, so $f(f^{k-1}(w_{kl})) = 0$. Hence, if we order the elements of (5) lexicographically by their index triples (k, l, i), then we obtain a basis as mentioned in the theorem. $\square$

4.9. Remark. If $(v_1, \ldots, v_n)$ is a basis as in Theorem 4.8, then the matrix $A = (a_{ij})$ representing f with respect to $(v_n, \ldots, v_2, v_1)$, has all entries zero except $a_{j,j+1} = 1$ if $f(v_{n-j}) = v_{n+1-j}$. Therefore A is a *block diagonal matrix*

$$A = \left(\begin{array}{c|c|c|c} B_1 & 0 & \cdots & 0 \\ \hline 0 & B_2 & \cdots & 0 \\ \hline \vdots & \vdots & \ddots & \vdots \\ \hline 0 & 0 & \cdots & B_k \end{array}\right)$$

where for each i there is an integer $m \ge 1$ such that the i-th block B_i is the $m \times m$ block

$$B(m) = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

with all zeroes except for ones just above the diagonal. Note that we reversed the order of the basis elements! Also note that $B(m)^m = 0$, and for each integer $1 \le s < m$, the matrix $B(m)^s$ is the $m \times m$ matrix with all zeroes, except for ones on the diagonal that is s positions above the main diagonal.

4.10. Corollary. *Every nilpotent matrix is similar to a matrix of the form just described.*

Proof. This is clear from our discussion. □

4.11. Corollary. *A matrix $A \in \mathrm{Mat}(n, F)$ is nilpotent if and only if its characteristic polynomial is $P_A(x) = x^n$.*

Proof. If $P_A(x) = x^n$, then $A^n = 0$ by the Cayley-Hamilton Theorem 3.1, hence A is nilpotent. Conversely, if A is nilpotent, then it is similar to a matrix of the form above, which visibly has characteristic polynomial x^n. □

4.12. Remark. The statement of Corollary 4.11 would also follow from the fact that $P_A(x)$ divides some power of $M_A(x) = x^m$, see Remark 3.8. However, we have proved this only in the case that $P_A(x)$ splits into linear factors (which we know is true, but only after the fact).

4.13. Example. Consider

$$A = \begin{pmatrix} 3 & 4 & -7 \\ 1 & 2 & -3 \\ 2 & 3 & -5 \end{pmatrix} \in \mathrm{Mat}(3, \mathbb{R})\,.$$

We find

$$A^2 = \begin{pmatrix} -1 & -1 & 2 \\ -1 & -1 & 2 \\ -1 & -1 & 2 \end{pmatrix}$$

and $A^3 = 0$, so A is nilpotent. Let us find a basis as given in Theorem 4.8. The first step in the process comes down to finding a complementary subspace of $\ker(A^2) = L\big((2,0,1)^\top, (-1,1,0)^\top\big)$ within $\ker A^3 = \mathbb{R}^3$. We can take $(1,0,0)^\top$, for example, as the basis for a complement. This will be w_{31} in the notation of the proof of Theorem 4.8. We then have $Aw_{31} = (3,1,2)^\top$ and $A^2 w_{31} = (-1,-1,-1)^\top$, and these three already form a basis $B = (A^2 w_{31}, Aw_{31}, w_{31})$. With

$$Q = [\mathrm{id}]^B_E = \begin{pmatrix} -1 & 3 & 1 \\ -1 & 1 & 0 \\ -1 & 2 & 0 \end{pmatrix}$$

we obtain

$$Q^{-1}AQ = [\mathrm{id}]^E_B \cdot [f_A]^E_E \cdot [\mathrm{id}]^B_E = [f_A]^B_B = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\,.$$

The following proposition tells us how many blocks of each size to expect.

4.14. Proposition. Let $f\colon V \to V$ be a nilpotent endomorphism of a finite-dimensional vector space V. Let $B = (v_n, \dots, v_1)$ be a basis for V such that its reverse is a basis as in Theorem 4.8. Let $A = [f]^B_B$ be the associated matrix. For every integer $j \ge 0$ we set $r_j = \dim \ker f^j$, and for every integer $j \ge 1$ we set $s_j = r_j - r_{j-1}$ and $t_j = s_j - s_{j+1}$. Then for every integer $j \ge 1$ there are exactly t_j blocks of the form $B(j)$ of size $j \times j$ along the diagonal of A.

PROOF. The matrix A is described in Remark 4.9. Let $m_1, m_2, \ldots, m_k \geq 0$ be integers such that the blocks along the diagonal of A are $B(m_1), \ldots, B(m_k)$. For each integer $j \geq 0$, the matrix A^j is a block matrix with blocks $B(m_1)^j, \ldots, B(m_k)^j$ along the diagonal. Therefore, the matrix A^j is in row echelon form, and for every i, the first $\min(m_i, j)$ columns corresponding to the i-th block $B(m_i)^j$ do not contain a pivot, while the other columns do contain pivots. Hence, the kernel of A^j has dimension

$$r_j = \sum_{i=1}^{k} \min(m_i, j)$$

and we find

$$s_j = r_j - r_{j-1} = \sum_{i=1}^{k} \big(\min(m_i, j) - \min(m_i, j-1) \big).$$

As for integers a, b the value $\min(a, b) - \min(a, b-1)$ equals 0 for $a < b$ and it equals 1 otherwise, we conclude that s_j equals the number of blocks of size at least j. Therefore, the number of blocks of size exactly j is $s_j - s_{j+1} = t_j$. □

4.15. Remark. The t_k from the proof of Theorem 4.8 are the same as the t_k from the proof of Proposition 4.14. Indeed, for fixed integers $1 \leq k \leq m$ and $1 \leq l \leq t_k$, with t_k as in the proof of Theorem 4.8, the k elements $f^i(w_{kl})$ with $0 \leq i < k$ in (5) form a basis for a subspace that corresponds to a block of size $k \times k$, so there are t_k such blocks. Moreover, with r_k and s_k as in Proposition 4.14, the proof of Theorem 4.8 shows

$$\dim X_k = \dim \ker f^k - \dim \ker f^{k-1} = r_k - r_{k-1} = s_k.$$

This also implies for t_k as defined in the proof of Theorem 4.8 that we have

$$t_k = \dim X_k - \dim f(X_{k+1}) = \dim X_k - \dim X_{k+1} = s_k - s_{k+1}.$$

While this seems to give another proof of Proposition 4.14, this argument a priori only holds for bases that are obtained as in the proof of Theorem 4.8. It is however not hard to show that every basis as mentioned in Theorem 4.8 can indeed be obtained through the construction in the proof of Theorem 4.8, so it does yield a second proof.

4.16. Example. In Example 4.13, we have $\operatorname{rk} A = 2$ and $\operatorname{rk} A^2 = 1$ and $A^3 = 0$, so we get the following table.

j	r_j	s_j	t_j
0	0		
1	1	1	0
2	2	1	0
3	3	1	1
4	3	0	0
5	3	0	

We conclude, as we have seen in the example above, that there is an invertible matrix Q such that $Q^{-1}AQ$ consists of one block $B(3)$.

4.17. Corollary. *Let $A, A' \in \mathrm{Mat}(n, F)$ be two nilpotent matrices. Then A and A' are similar if and only if for each integer $1 \le j < n$ we have $\dim\ker A^j = \dim\ker A'^j$.*

Proof. For every integer $j \ge 0$, and every square matrix M, set $r_j(M) = \dim\ker M^j$. For $j \ge 1$, also set $s_j(M) = r_j(M) - r_{j-1}(M)$ and $t_j(M) = s_j(M) - s_{j+1}(M)$. Of course, if A and A' are similar, then $r_j(A) = r_j(A')$ for each j. Conversely, suppose that for each integer $1 \le j < n$ we have $r_j(A) = r_j(A')$. By Cayley-Hamilton, we have $A^n = A'^n = 0$, so for $j \ge n$ we have $r_j(A) = r_j(A')$ as well, as both equal n. For $j = 0$ both equal 0, so we have $r_j(A) = r_j(A')$ for all $j \ge 0$. This implies that for all $j \ge 1$ we have $s_j(A) = s_j(A')$ and $t_j(A) = t_j(A')$, so by Proposition 4.14, both A and A' are similar to a block diagonal matrix with $t_j(A) = t_j(A')$ blocks of the form $B(j)$ along the diagonal for every $j \ge 1$. Any two such matrices are similar to each other; in fact they can be obtained from each other by a permutation of the basis. By transitivity of similarity, also A and A' are similar. □

4.18. Example. Consider the real matrix

$$A = \begin{pmatrix} -5 & 10 & -8 & 4 & 1 \\ -4 & 8 & -10 & 8 & 2 \\ -3 & 6 & -12 & 12 & 3 \\ -2 & 4 & -8 & 4 & 10 \\ -1 & 2 & -4 & 2 & 5 \end{pmatrix}$$

and the linear map $f = f_A : \mathbb{R}^5 \to \mathbb{R}^5$ associated to it. We compute

$$A^2 = \begin{pmatrix} 0 & 0 & 0 & -18 & 36 \\ 0 & 0 & 0 & -36 & 72 \\ 0 & 0 & 0 & -54 & 108 \\ 0 & 0 & 0 & -36 & 72 \\ 0 & 0 & 0 & -18 & 36 \end{pmatrix} \qquad \text{and} \qquad A^3 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

so for $m = 3$ we have $A^m = 0$. The kernel $\ker A$ is generated by

$$x = (-3, 0, 3, 2, 1) \qquad \text{and} \qquad x' = (2, 1, 0, 0, 0).$$

(We urge the reader to verify this, either by bringing A into row echelon form by elementary row operations, or by verifying that A has rank 3, concluding that $\ker A$ has dimension 2, and checking that x and x' are linearly independent elements contained in $\ker A$.) The kernel $\ker A^2$ is generated by

$$e_1 = (1, 0, 0, 0, 0), \quad e_2 = (0, 1, 0, 0, 0), \quad e_3 = (0, 0, 1, 0, 0), \quad \text{and} \quad y = (0, 0, 0, 2, 1).$$

Clearly, we have $\ker A^3 = \mathbb{R}^5$. In terms of Proposition 4.14, with $r_j = \dim\ker A^j$, we find $r_0 = 0$ and $r_1 = 2$ and $r_2 = 4$ and $r_n = 5$ for $n \ge 3$; this yields $s_1 = 2$ and $s_2 = 2$ and $s_3 = 1$ and $s_4 = 0$. Finally, we obtain $t_1 = 0$ and $t_2 = 1$ and $t_3 = 1$, so we already find that the standard nilpotent form consists of one block of size 2 and one block of size 3.

To find an appropriate basis, we start with step $j = m = 3$ (as in the proof of Theorem 4.8) by picking a complementary space X_3 of $\ker A^2$ inside $\ker A^3 = \mathbb{R}^5$. Since $\dim\ker A^3 - \dim\ker A^2 = 5 - 4 = 1$, it suffices to pick any element of $\mathbb{R}^5$ that is not contained in $\ker A^2$. We choose $w_{31} = e_5 = (0, 0, 0, 0, 1)$, which gives $Aw_{31} = (1, 2, 3, 10, 5)$ and $A^2 w_{31} = 36(1, 2, 3, 2, 1)$ and $A^3 w_{31} = 0$. We take $X_3 = \langle w_{31} \rangle$. In the next step ($j = 2$), we extend $f(X_3) \subset \ker A^2$ to a complementary space X_2 of $\ker A$ inside $\ker A^2$. In order to do this, we follow the proof of Lemma 2.6: take a basis for $\ker A$ and for $f(X_3)$ and put the elements of these two bases as columns in a matrix; we also take generators for $\ker A^2$ and add these as columns to the matrix. We obtain

$$\left(\begin{array}{cc|c|cccc} -3 & 2 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 & 0 \\ 3 & 0 & 3 & 0 & 0 & 1 & 0 \\ 2 & 0 & 10 & 0 & 0 & 0 & 2 \\ 1 & 0 & 5 & 0 & 0 & 0 & 1 \end{array}\right).$$

A row echelon form for this matrix is

$$\left(\begin{array}{cc|c|cccc} 1 & 0 & 5 & 0 & 0 & 0 & 1 \\ 0 & 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 12 & 0 & 0 & -1 & 3 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right),$$

which has pivots in the first three columns as expected. Of the last four columns, only the first contains a pivot, so in order to extend $f(X_3)$ to a complementary space X_2 as mentioned, it suffices to add the first generator for $\ker A^2$, so we take $w_{21} = (1,0,0,0,0)$, which gives $Aw_{21} = -(5,4,3,2,1)$. The last step ($j = 1$), namely finding a complementary space X_1 for $\ker A^0 = \{0\}$ inside $\ker A$ that contains $f(X_2)$, turns out to be trivial. Indeed, $f(X_2)$ is generated by $A^2 w_{31}$ and Aw_{21}, so $\dim f(X_2) = 2 = \dim\ker A$, so we have $X_1 = f(X_2)$ and we do not need to extend.

Hence, we obtain a basis $B = (A^2w_{31}, Aw_{31}, w_{31}, Aw_{21}, w_{21})$ (note the order of the elements). If we denote the standard basis for $\mathbb{R}^5$ by E, the basis transformation matrix

$$P = [\mathrm{id}]_E^B = \begin{pmatrix} 36 & 1 & 0 & -5 & 1 \\ 72 & 2 & 0 & -4 & 0 \\ 108 & 3 & 0 & -3 & 0 \\ 72 & 10 & 0 & -2 & 0 \\ 36 & 5 & 1 & -1 & 0 \end{pmatrix}$$

satisfies

$$P^{-1}AP = [f_A]_B^B = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

4.19. Example. From small examples one does not always get a good idea of the general case, so we now do a bigger example. If the reader wishes to verify the calculations, we recommend using a computer.

Let M be the 11×11 real matrix

$$M = \begin{pmatrix} 14 & 15 & 0 & 8 & -40 & 32 & -2 & -72 & -8 & 0 & -20 \\ -29 & -34 & -7 & -16 & 55 & -64 & 14 & 137 & 16 & 0 & 31 \\ 6 & 10 & 2 & 4 & -18 & 15 & -2 & -33 & -5 & 0 & -10 \\ -3 & -2 & 2 & -1 & -10 & 0 & -2 & 3 & 0 & 1 & -6 \\ -6 & -7 & 0 & -4 & 24 & -15 & -1 & 34 & 4 & 0 & 12 \\ 14 & 7 & -4 & 6 & -28 & 24 & 5 & -56 & -4 & 0 & -12 \\ -3 & -4 & -1 & -2 & 9 & -8 & 2 & 17 & 2 & 0 & 5 \\ 10 & 7 & -2 & 5 & -26 & 20 & 2 & -46 & -4 & 0 & -12 \\ -67 & -77 & -14 & -38 & 130 & -148 & 30 & 319 & 36 & 1 & 72 \\ -53 & -54 & -2 & -28 & 102 & -108 & 10 & 241 & 26 & 1 & 52 \\ 12 & 15 & 2 & 8 & -42 & 30 & -1 & -66 & -8 & 0 & -22 \end{pmatrix}.$$

One checks that $M^4 = 0$, so M is nilpotent.

Moreover, one checks that M, M^2, and M^3 have rank $7, 4$, and 1, respectively. This gives the following table.

j	r_j	s_j	t_j
0	0		
1	4	4	1
2	7	3	0
3	10	3	2
4	11	1	1
5	11	0	0
6	11	0	

We conclude that there is an invertible matrix Q such that $Q^{-1}MQ$ is a block matrix consisting of one block $B(1)$, two blocks $B(3)$, and one block $B(4)$ along its diagonal.

To find such a matrix Q, we will construct a basis $(v_1, v_2, \ldots, v_{11})$ as in Theorem 4.8 following the proof of that theorem. We note that $M^m = 0$ for $m = 4$, so we start with $j = m = 4$. We want to pick a basis for a complementary space X_4 of $\ker M^3$ inside $\ker M^4 = \mathbb{R}^{11}$; given that we have $\dim \ker M^3 = 10$, we find $\dim X_4 = 1$, so it suffices to find one vector $w_{41} \in R^{11}$ that is not contained in $\ker M^3$. The 3-rd, 7-th, and 10-th column of M^3 are the only zero columns, so the standard basis vector e_i is not contained in $\ker M^3$ for $i \notin \{3, 7, 10\}$. Because the fourth column of M^3 contains relatively small numbers, we choose $w_{41} = e_4$. This gives

$$w_{41} = \begin{pmatrix} 0\\0\\0\\1\\0\\0\\0\\0\\0\\0\\0 \end{pmatrix}, \qquad Mw_{41} = \begin{pmatrix} 8\\-16\\4\\-1\\-4\\6\\-2\\5\\-38\\-28\\8 \end{pmatrix}, \qquad M^2w_{41} = \begin{pmatrix} 4\\-7\\3\\0\\-2\\0\\-1\\1\\-15\\-11\\4 \end{pmatrix}, \qquad M^3w_{41} = \begin{pmatrix} 1\\-2\\0\\-2\\0\\2\\0\\1\\-5\\-6\\0 \end{pmatrix}.$$

These vectors correspond to a block of the form $B(4)$. To check consistency, one could verify that indeed the last vector is contained in the kernel of M.

We continue with $j = 3$. We want to pick a basis for some complementary space X_3 of $\ker M^2$ inside $\ker M^3$ that contains $M^{4-j}w_{41} = Mw_{41}$ (this is indeed the only vector of the four that we already found that is contained in $\ker M^3$ but not in $\ker M^2$). We do this following the proof of Lemma 2.6. One computes that the kernel $\ker M^2$ is generated by the columns of the matrix

$$K_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 2 & 3 & 0 & 0 & 0 & 0 & 4 \\ 4 & 5 & -1 & 0 & -1 & 0 & 7 \\ -32 & -41 & 9 & 1 & 9 & 4 & -61 \\ -7 & -7 & 1 & 1 & 1 & 0 & -11 \\ -1 & -1 & 0 & 0 & -2 & 0 & -1 \end{pmatrix}.$$

Moreover, the kernel $\ker M^3$ is generated by the columns of the matrix

$$K_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & -2 & 2 & 0 & -5 & -1 & 0 \end{pmatrix}$$

Lemma 2.6 tells us that in order to extend Mw_{41} to a basis for a complementary space of $\ker M^2$ inside $\ker M^3$, we take the columns of K_2 together with one column Mw_{41}, and extend this to a basis for $\ker M^3$ by adding some of the columns of K_3. We do this by taking the extended matrix

$$(K_2|Mw_{41}|K_3)$$

and using elementary row operations to bring this into (reduced) row echelon form. This yields

$$\left(\begin{array}{ccccccc|c|cccccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -18 & 7 & -18 & -4 & 24 & -15 & 2 & -7 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 36 & -12 & 36 & 8 & -50 & 28 & -4 & 12 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & -7 & 4 & -8 & 0 & 10 & -12 & 0 & -4 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 2 & 3 & 1 & 3 & 2 & -5 & -2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 3 & 10 & -2 & 11 & 4 & -15 & 2 & -2 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 3 & -12 & 6 & -12 & 1 & 15 & -18 & 0 & -6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -16 & 6 & -16 & -2 & 22 & -17 & 1 & -6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 2 & 3 & 0 & 3 & 2 & -5 & -2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 4 & 2 & 1 & 2 & 4 & -4 & -9 & -2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5 & 2 & 2 & 2 & 4 & -5 & -10 & -2 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Since the first two columns of the right part of this matrix are the ones that contain a pivot, we see that we may add the corresponding first two columns of

K_3 to Mw_{41} to obtain a complementary space of $\ker M^2$ inside $\ker M^3$. The first two columns of K_3 are $w_{31} = e_1 + e_{11}$ and $w_{32} = e_2 + e_{11}$, so we find

$$X_3 = L\big(Mw_{41}, w_{31}, w_{32}\big).$$

Note as a consistency check that indeed we have $\dim X_3 + \dim\ker M^2 = \dim\ker M^3$, that is, $3 + 7 = 10$. For $1 \leq l \leq 2$, the vectors $w_{3l}, Mw_{3l}, M^2w_{3l}$ span a subspace that corresponds to a block of the form $B(3)$.

We proceed with $j = 2$. We want to pick a basis for some complementary space X_2 of $\ker M$ inside $\ker M^2$ that contains $M^{4-j}w_{41} = M^2w_{41}$ and $M^{3-j}w_{31} = Mw_{31}$ and $M^{3-j}w_{32} = Mw_{32}$ (these are indeed the only vectors of the ten that we found so far that are contained in $\ker M^2$ but not in $\ker M$). From $\dim X_2 = \dim\ker M^2 - \dim\ker M = 7 - 4 = 3$, we find that the linearly independent vectors M^2w_{41} and Mw_{31} and Mw_{32} already span X_2. This corresponds to the fact that there are no blocks of the form $B(2)$, as we had already seen.

Finally, for $j = 1$, we want to pick a basis for some complementary space X_1 of $\ker M^0 = \ker I_{11} = \{0\}$ inside $\ker M$ that contains $M^{4-j}w_{41} = M^3w_{41}$ and $M^{3-j}w_{31} = M^2w_{31}$ and $M^{3-j}w_{32} = M^2w_{32}$ (these are indeed the vectors among those that we found so far that are contained in $\ker M$ but not in $\ker M^0 = \{0\}$). We do this by writing down an extended matrix with M^3w_{41} and M^2w_{31} and M^2w_{32} as columns on the left, and four generators for $\ker M$ on the right, say

$$\left(\begin{array}{ccc|cccc}
1 & -2 & -1 & 1 & 0 & 0 & 0 \\
-2 & 4 & 2 & 2 & 4 & 0 & 0 \\
0 & -1 & 1 & 0 & 0 & 1 & 0 \\
-2 & 4 & 2 & -4 & -2 & 0 & 0 \\
0 & 2 & 3 & 0 & 0 & 0 & 1 \\
2 & -6 & -10 & 0 & -2 & -2 & -2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & -3 & -5 & 0 & -1 & -1 & -1 \\
-5 & 11 & 9 & 4 & 9 & 1 & 1 \\
-6 & 13 & 10 & -3 & 3 & 1 & 1 \\
0 & -4 & -6 & -1 & -1 & 0 & -2
\end{array}\right).$$

Note that here we have no columns coming from a basis for $\ker M^0 = \{0\}$. The associated reduced row echelon form is

$$\left(\begin{array}{ccc|cccc}
1 & 0 & 2 & 0 & -1 & 0 & 1 \\
0 & 1 & 4 & 0 & 0 & 1 & 1 \\
0 & 0 & 5 & 0 & 0 & 2 & 1 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right).$$

Since the first column on the right is the only column on the right with a pivot, we add only the first of the four chosen generators for $\ker M$, so

$$w_{11} = \begin{pmatrix} 1 \\ 2 \\ 0 \\ -4 \\ 0 \\ 0 \\ 0 \\ 0 \\ 4 \\ -3 \\ -1 \end{pmatrix}.$$

We conclude that

$$(w_{11}, w_{31}, Mw_{31}, M^2w_{31}, w_{32}, Mw_{32}, M^2w_{32}, w_{41}, Mw_{41}, M^2w_{41}, M^3w_{41})$$

is a basis as in Theorem 4.8. Putting the eleven vectors in reverse order, we obtain the basis B. If we let E denote the standard basis, and we set

$$Q = [\mathrm{id}]_E^B = \begin{pmatrix} 1 & 4 & 8 & 0 & -1 & -5 & 0 & -2 & -6 & 1 & 1 \\ -2 & -7 & -16 & 0 & 2 & -3 & 1 & 4 & 2 & 0 & 2 \\ 0 & 3 & 4 & 0 & 1 & 0 & 0 & -1 & -4 & 0 & 0 \\ -2 & 0 & -1 & 1 & 2 & -8 & 0 & 4 & -9 & 0 & -4 \\ 0 & -2 & -4 & 0 & 3 & 5 & 0 & 2 & 6 & 0 & 0 \\ 2 & 0 & 6 & 0 & -10 & -5 & 0 & -6 & 2 & 0 & 0 \\ 0 & -1 & -2 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 0 \\ 1 & 1 & 5 & 0 & -5 & -5 & 0 & -3 & -2 & 0 & 0 \\ -5 & -15 & -38 & 0 & 9 & -5 & 0 & 11 & 5 & 0 & 4 \\ -6 & -11 & -28 & 0 & 10 & -2 & 0 & 13 & -1 & 0 & -3 \\ 0 & 4 & 8 & 0 & -6 & -7 & 1 & -4 & -10 & 1 & -1 \end{pmatrix},$$

then we find

$$Q^{-1}MQ = [\mathrm{id}]_B^E[f_M]_E^E[\mathrm{id}]_E^B = [f_M]_B^B = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Exercises.

(1) Let A be a nilpotent $n \times n$ matrix. Show that $\mathrm{id}_n + A$ is invertible.
(2) Let A be a nilpotent $n \times n$ matrix. Show that $A^n = 0$.
(3) Let N be a 9×9 matrix for which $N^3 = 0$. Suppose that N^2 has rank 3. Prove that N has rank 6.
(4) Let N be a 12×12 matrix for which $N^4 = 0$.
  (a) Show that the kernel of N^2 contains the image of N^2.
  (b) Show that the rank of N is at most 9.

(c) Show that the rank of N is equal to 9 if the kernel of N^2 is equal to the image of N^2.

(5) Let A be a square matrix over any field. Suppose that $r > 0$ is an integer for which $\dim\ker A^r = \dim\ker A^{r+1}$. Show that for every integer $s > r$ we have $\ker A^r = \ker A^s$.

(6) For which $x \in \mathbb{R}$ is the following matrix nilpotent?

$$\begin{pmatrix} 2x & x & -1 \\ -4 & -1 & -3 \\ 5 & 2 & 3 \end{pmatrix}$$

(7) For each of the matrices

$$\begin{pmatrix} 4 & -4 & 12 \\ 1 & -1 & 3 \\ -1 & 1 & -3 \end{pmatrix} \qquad \begin{pmatrix} 2 & 0 & 8 \\ 0 & 1 & 1 \\ -1 & 1 & -3 \end{pmatrix}$$

give a basis for $\mathbb{R}^3$ for which the matrix sends each basis vector either to 0 or to the next basis vector in the basis.

(8) Do the same for the matrix

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ -5 & -2 & 2 & -1 \\ -3 & 0 & 2 & -1 \\ -5 & -2 & 2 & -1 \end{pmatrix}$$

5. The Jordan Normal Form Theorem

In this section, we will formulate and prove the Jordan Normal Form Theorem, which will tell us that any matrix whose characteristic polynomial is a product of linear factors is similar to a matrix of a very special near-diagonal form.

Just like true diagonal forms are related to eigenspaces, the Jordan normal form is related to so-called generalised eigenspaces.

5.1. Definition. Let V be a vector space over a field F, and $f\colon V \to V$ an endomorphism. Let $\lambda \in F$ be an element. The *generalised λ-eigenspace* of f is

$$\tilde{E}_\lambda(f) = \{\, v \in V \;:\; (f - \lambda\,\mathrm{id}_V)^l(v) = 0 \text{ for some } l \geq 1 \,\} = \bigcup_{l \geq 1} \ker\,(f - \lambda\,\mathrm{id}_V)^l\,.$$

We leave it to the reader to check that the generalised λ-eigenspace is indeed a subspace of V (Exercise 1). Clearly, it contains the λ-eigenspace

$$E_\lambda(f) = \ker(f - \lambda\,\mathrm{id}_V).$$

Moreover, if the generalised λ-eigenspace $\tilde{E}_\lambda(f)$ contains a nonzero element v, then for some integer $l \geq 1$ we have $(f - \lambda\,\mathrm{id}_V)^l(v) = 0$; for the smallest such integer l we set $w = (f - \lambda\,\mathrm{id}_V)^{l-1}(v)$ and find $w \neq 0$ and $(f - \lambda\,\mathrm{id}_V)(w) = 0$, so λ is an eigenvalue of f and w is an eigenvector for the eigenvalue λ. The nonzero elements of $\tilde{E}_\lambda(f)$ are called *generalised eigenvectors* for the eigenvalue λ.

The following theorem shows that if V is finite-dimensional, then there exists an integer m such that $\tilde{E}_\lambda(f) = \ker(f - \lambda\,\mathrm{id}_V)^m$ (cf. Exercise 2).

5.2. Theorem. *Let V be a finite-dimensional vector space over a field F, and let $f\colon V \to V$ be an endomorphism. Let $p \in F[x]$ be a polynomial with coefficients in F satisfying $p(f) = 0$. Let $\lambda \in F$ be any element, and factorise p as $p(x) = (x - \lambda)^m q(x)$ with $q(\lambda) \neq 0$. Set*

$$U = \ker(f - \lambda\,\mathrm{id}_V)^m \qquad \textit{and} \qquad U' = \ker q(f).$$

Then we have $\tilde{E}_\lambda(f) = U$ and $V = U \oplus U'$ and $f = f|_U \oplus f|_{U'}$. Moreover, if p is equal to the minimal polynomial M_f or the characteristic polynomial P_f of f, then the characteristic polynomial of $f|_U$ is a multiple of $(x-\lambda)^m$, and $\dim \tilde{E}_\lambda(f) \geq m$.

Proof. By Lemma 2.15, we know that $(x - \lambda)^m$ and q are coprime. The fact that $V = U \oplus U'$ and $f = f|_U \oplus f|_{U'}$ then follows directly from Lemma 2.12.

Since $q\big(f|_{U'}\big) = 0$, the minimal polynomial of $f|_{U'}$ divides q, so it does not have λ has a root. Then by Proposition 3.7, the characteristic polynomial $P_{f|_{U'}}$ of $f|_{U'}$ does not have a factor $x - \lambda$, so λ is not an eigenvalue of $f|_{U'}$. We conclude that $f - \lambda\,\mathrm{id}_V$ restricts to an automorphism of U', and hence so does every power $(f - \lambda\,\mathrm{id}_V)^l$. This implies that the rank of $(f - \lambda\,\mathrm{id}_V)^l$ is at least $\dim U'$ for every l, and therefore

$$\dim\ker(f - \lambda\,\mathrm{id}_V)^l \leq \dim V - \dim U' = \dim U.$$

For $l \geq m$, the containment $\ker(f - \lambda\,\mathrm{id}_V)^l \supset \ker(f - \lambda\,\mathrm{id}_V)^m = U$ is therefore an equality, which implies $\tilde{E}_\lambda(f) = U$.

Now suppose that p is equal to M_f or P_f. By definition of the minimal polynomial, we have $M_f(f) = 0$, and by the Cayley-Hamilton Theorem 3.1, we know that $P_f(f) = 0$, so in both cases all the above arguments apply. Since M_f divides P_f, in both cases the polynomial $(x - \lambda)^m$ divides the characteristic polynomial P_f, which by Remark 2.8 we know to be equal to $P_{f|_U} \cdot P_{f|_{U'}}$. We have seen that $x - \lambda$ does not divide $P_{f|_{U'}}$, so $(x-\lambda)^m$ is a divisor of $P_{f|_U}$. This implies $\dim U \geq m$. □

5.3. Remark. In fact, in the notation of Theorem 5.2, if we have $p = P_f$, then one can prove that $P_{f|_U} = (x - \lambda)^m$ and $\dim \tilde{E}_\lambda(f) = m$, as we will now sketch. As we have seen in Remark 3.8, it is a fact that the characteristic polynomial of an endomorphism divides some power of the minimal polynomial, though we have only proved that in the case that the characteristic polynomial splits into linear factors (see Proposition 3.7), and the case that $F = \mathbb{R}$ (see Exercise 12 of Chapter 3).
Since we know that the minimal polynomial of $f|_U$ divides $(x-\lambda)^m$, it would follow from this not-in-full-generality-proven fact that the characteristic polynomial $P_{f|_U}$ is a power of $x - \lambda$. As it is a multiple of $(x - \lambda)^m$ by Theorem 5.2 and it divides P_f, we would indeed find $P_{f|_U} = (x-\lambda)^m$ and hence $P_{f|_{U'}} = q(x)$ and $\dim U = m$.

The following theorem also concludes the equality $\dim \tilde{E}_\lambda(f) = m$ in the restricted case that the characteristic polynomial P_f splits into linear factors, and the multiplicity of λ as a root of P_f is m, just as it was above.

5.4. Theorem. *Let V be a finite-dimensional vector space over a field F, and let $f\colon V \to V$ be an endomorphism. Let $p \in F[x]$ be a polynomial with coefficients in F satisfying $p(f) = 0$ that splits into linear factors:*

$$p(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}\,,$$

where the $\lambda_i \in F$ *are distinct. Then for the* generalised λ_i-eigenspaces $U_i = \tilde{E}_{\lambda_i}(f)$ *of* f *we have* $U_i = \ker(f - \lambda_i \operatorname{id}_V)^{m_i}$ *and* $V = U_1 \oplus \cdots \oplus U_k$ *and* $f = f|_{U_1} \oplus \cdots \oplus f|_{U_k}$. *Moreover, if* $p = P_f$*, then* $\dim \tilde{E}_{\lambda_i}(f) = m_i$.

Proof. Write $p(x) = p_1(x) \cdots p_k(x)$ with $p_i(x) = (x - \lambda_i)^{m_i}$. By Theorem 5.2 we have $\tilde{E}_{\lambda_i}(f) = U_i = \ker p_i(f) = \ker(f - \lambda_i \operatorname{id}_V)^{m_i}$. By Lemma 2.15, we know that the $p_i(x)$ are coprime in pairs. The facts that V and f are direct sums as stated then follow from Proposition 2.14. Now suppose $p = P_f$. By the Cayley-Hamilton Theorem 3.1, we know that $P_f(f) = 0$, so all the arguments above apply. From Theorem 5.2 we then also find $\dim U_i = \dim \tilde{E}_{\lambda_i}(f) \geq m_i$ for all i, so we obtain

$$\dim V = \sum_i \dim U_i \geq \sum_i m_i = \deg P_f = \dim V,$$

which implies that all inequalities $\dim U_i \geq m_i$ are actually equalities. □

5.5. Theorem (Jordan Normal Form). *Let* V *be a finite-dimensional vector space, and let* $f : V \to V$ *be an endomorphism whose characteristic polynomial splits into linear factors:*

$$P_f(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k},$$

where the λ_i *are distinct. Then there is a basis for* V *such that the matrix representing* f *with respect to that basis is a block diagonal matrix with blocks of the form*

$$B(\lambda, m) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ 0 & 0 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{pmatrix} \in \operatorname{Mat}(m, F)$$

where $\lambda \in \{\lambda_1, \ldots, \lambda_k\}$.

Proof. We keep the notations of Theorem 5.4. We know that on U_i, we have $(f - \lambda_i \operatorname{id})^{m_i} = 0$, so $f|_{U_i} = \lambda_i \operatorname{id}_{U_i} + g_i$, where $g_i^{m_i} = 0$, i.e., g_i is nilpotent. By Theorem 4.8, there is a basis for U_i such that g_i is represented by a block diagonal matrix B_i with blocks of the form $B(0, m)$ (such that the sum of the m's is m_i). Therefore, $f|_{U_i}$ is represented by $B_i + \lambda_i I_{\dim U_i}$, which is a block diagonal matrix composed of blocks $B(\lambda_i, m)$ (with the same m's as before). The basis for V that is the concatenation of the various bases for the U_i then does what we want, compare Remark 2.8. □

We say that a matrix is in *Jordan normal form* if it is a diagonal block matrix with all blocks along the diagonal of the form $B(\lambda, m)$ for some $\lambda \in F$ and some integer $m \geq 0$.

5.6. Remark. Let V, f, and $\lambda_1, \ldots, \lambda_k \in F$ be as in Theorem 5.5. Let B be a basis as is claimed to exist, and let $A = [f]_B^B$ be the matrix associated to f with respect to B. Take any element $\lambda \in F$. For every integer $j \geq 0$ we set $r_j(\lambda) = \dim \ker(f - \lambda \operatorname{id}_V)^j$, and for every integer $j \geq 1$ we set $s_j(\lambda) = r_j(\lambda) - r_{j-1}(\lambda)$ and $t_j(\lambda) = s_j(\lambda) - s_{j+1}(\lambda)$. Then for every integer $j \geq 1$ there are exactly $t_j(\lambda)$ blocks of the form $B(\lambda, j)$ along the diagonal of A.

Indeed, for λ not a root of the characteristic polynomial P_f, we get $r_j(\lambda) = s_j(\lambda) = t_j(\lambda) = 0$ for all j, and no blocks of the form $B(\lambda, j)$ for any j. If $\lambda = \lambda_i$ for some i, then in terms of the notation of the proof of Theorem 5.5, we can apply Proposition 4.14 to the nilpotent endomorphisn $g_i = f|_{U_i} - \lambda \operatorname{id}_{U_i}$, which satisfies $g_i^{m_i} = 0$. Note that for every integer $j \geq 0$ the kernel $\ker(f - \lambda_i \operatorname{id}_V)^j$ is contained in $\ker(f - \lambda_i \operatorname{id}_V)^{m_i} = U_i$ by Theorem 5.4. Hence this kernel equals $\ker g_i^j$, and we find $r_j(\lambda_i) = \dim \ker g_i^j$. Proposition 4.14 then states that there are $t_j(\lambda_i)$ blocks of the form $B(0, j)$ in a diagonal block matrix for g_i, and these blocks correspond to blocks in A of the form $B(\lambda_i, j)$.

5.7. Corollary. Let $A, A' \in \operatorname{Mat}(n, F)$ be two square matrices such that the characteristic polynomial of A splits into linear factors, that is,

$$P_A(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k} .$$

Then A and A' are similar if and only if for each index $1 \leq i \leq k$ and each integer $1 \leq j \leq m_i$ we have $\dim \ker(A - \lambda_i I)^j = \dim \ker(A' - \lambda_i I)^j$.

Proof. If A and A' are similar, then the claimed equality of dimensions holds. For the converse, assume that for every index $1 \leq i \leq k$ and for each integer $1 \leq j \leq m_i$ we have $\dim \ker(A - \lambda_i I)^j = \dim \ker(A' - \lambda_i I)^j$. Then in particular, this holds for $j = m_i$. Since $\ker(A - \lambda_i I)^{m_i}$ is the generalised eigenspace associated to λ_i for A, we find that for each i, the dimension of the generalised eigenspace associated to λ_i is at least as large for A' as for A. Since the sum of the dimensions of all generalised eigenspaces for A and for A' are both equal to n, we find that equality holds for each i, and furthermore, A' has no other eigenvalues. It follows that the characteristic polynomials of A and A' are the same. From Remark 5.6 we conclude that A and A' are both similar to a block diagonal matrices B and B', respectively, where B and B' have the same blocks along the diagonal. For details, compare to the proof of Corollary 4.17. Then B and B' are similar, as they can be obtained from each other by a permutation of the basis. So by transitivity of similarity, also A and A' are similar. □

Here is a less precise, but for many applications sufficient version of Theorem 5.5.

5.8. Corollary. *Let V be a finite-dimensional vector space, and let $f : V \to V$ be an endomorphism whose characteristic polynomial splits into linear factors, as above. Then we can write $f = d + n$, with endomorphisms d and n of V, such that d is diagonalizable, n is nilpotent, and d and n commute: $d \circ n = n \circ d$.*

Proof. We just take d to be the endomorphism corresponding to the 'diagonal part' of the matrix given in Theorem 5.5 and n to be that corresponding to the 'nilpotent part' (obtained by setting all diagonal entries equal to zero). Since the two parts commute within each 'Jordan block,' the two endomorphisms commute. □

5.9. Example. Let us compute the Jordan Normal Form and a suitable basis for the endomorphism $f : \mathbb{R}^3 \to \mathbb{R}^3$ given by the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -4 & 0 & 3 \end{pmatrix} .$$

We first compute the characteristic polynomial:

$$P_f(x) = \begin{vmatrix} x & -1 & 0 \\ 0 & x & -1 \\ 4 & 0 & x-3 \end{vmatrix} = x^2(x-3)+4 = x^3-3x^2+4 = (x-2)^2(x+1)\,.$$

We see that it splits into linear factors, which is good. We now have to find the generalised eigenspaces. The eigenvalue -1 has algebraic multiplicity 1, so its generalised eigenspace has dimension 1. It is therefore equal to the eigenspace

$$E_{-1}(f) = \ker \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ -4 & 0 & 4 \end{pmatrix} = L\big((1,-1,1)^\top\big)\,,$$

so for a basis we can choose $v = (1,-1,1)^\top$. The other eigenspace is

$$E_2(f) = \ker \begin{pmatrix} -2 & 1 & 0 \\ 0 & -2 & 1 \\ -4 & 0 & 1 \end{pmatrix} = L\big((1,2,4)^\top\big)\,.$$

This space has only dimension 1, so f is not diagonalizable, and we have to look at the generalised eigenspace:

$$\ker\big((f-2\,\mathrm{id})^2\big) = \ker \begin{pmatrix} 4 & -4 & 1 \\ -4 & 4 & -1 \\ 4 & -4 & 1 \end{pmatrix} = L\big((1,1,0)^\top, (1,0,-4)^\top\big)\,.$$

To construct a basis for this generalised eigenspace, we follow the proof of Theorem 4.8, applied to the nilpotent endomorphism that is $f - 2\,\mathrm{id}$ restricted to its generalised eigenspace. We start with a basis for a complementary space of $\ker(f-2\,\mathrm{id})$ inside $\ker(f-2\,\mathrm{id})^2$. Such a complementary space has dimension $\dim\ker(f-2\,\mathrm{id})^2 - \dim\ker(f-2\,\mathrm{id}) = 2-1 = 1$, so we can take any element in $\ker(f-2\,\mathrm{id})^2$ that is not contained in $\ker(f-2\,\mathrm{id})$, say $w_{21} = (1,1,0)^\top$. As basis for this generalised eigenspace, we then obtain $(w_{21}, (f-2\,\mathrm{id})(w_{21}))$. Reversing the order, and adding the basis (v) for the generalised eigenspace for $\lambda = -1$, we get a basis

$$B = ((f-2\,\mathrm{id})(w_{21}), w_{21}, v) = \big((-1,-2,-4)^\top, (1,1,0)^\top, (1,-1,1)^\top\big)\,,$$

for $\mathbb{R}^3$. With

$$P = [\mathrm{id}]^B_E = \begin{pmatrix} -1 & 1 & 1 \\ -2 & 1 & -1 \\ -4 & 0 & 1 \end{pmatrix}$$

we obtain

$$[f_A]^B_B = [\mathrm{id}]^E_B \cdot [f_A]^E_E \cdot [\mathrm{id}]^B_E = P^{-1}AP = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix}\,.$$

As mentioned in Example 4.19, from small examples one does not always get an idea of the general case, so at the end of this chapter, we will do some bigger examples.

5.10. Application. One important application of the Jordan Normal Form Theorem is to the explicit solution of systems of linear first-order differential equations with constant coefficients. Such a system can be written

$$\frac{d}{dt}y(t) = A \cdot y(t)\,,$$

where y is a vector-valued function and A is a matrix. One can then show (Exercise) that there is a unique solution with $y(0) = y_0$ for any specified initial value y_0, and it is given by

$$y(t) = \exp(tA) \cdot y_0$$

with the matrix exponential

$$\exp(tA) = \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n\,.$$

If A is in Jordan Normal Form, the exponential can be easily determined. In general, A can be transformed into Jordan Normal Form, the exponential can be evaluated for the transformed matrix, then we can transform it back — note that

$$\exp(tP^{-1}AP) = P^{-1}\exp(tA)P\,.$$

5.11. Remark. Writing an endomorphism $f\colon V \to V$ as $f = n + d$ with d diagonalizable and n nilpotent and $d \circ n = n \circ d$ is very useful for computing powers of f, as for every positive integer k, the relation $d \circ n = n \circ d$ implies

$$f^k = \sum_{i=0}^{k} \binom{k}{i} d^{k-i} n^i,$$

and if $n^m = 0$ for some integer m, then all terms with $i \geq m$ vanish.

5.12. Remark. What can we do when the characteristic polynomial does not split into linear factors (which is possible when the field F is not algebraically closed)? In this case, we have to use a weaker notion than that of diagonalizability. Define the endomorphism $f : V \to V$ to be *semi-simple* if every f-invariant subspace $U \subset V$ has an f-invariant complementary subspace in V. One can show (exercise) that if the characteristic polynomial of f splits into linear factors, then f is semi-simple if and only if it is diagonalizable. The general version of the Jordan Normal Form Theorem then is as follows.

Let V be a finite-dimensional vector space, $f : V \to V$ an endomorphism. Then $f = s + n$ with endomorphisms s and n of V such that s is semi-simple, n is nilpotent, and $s \circ n = n \circ s$.

Unfortunately, we do not have the means and time to prove this result here.

However, we can state the result we get over $F = \mathbb{R}$.

5.13. Theorem (Real Jordan Normal Form). *Let V be a finite-dimensional real vector space, $f : V \to V$ an endomorphism. Then there is a basis for V such that the matrix representing f with respect to this basis is a block diagonal matrix* *with blocks of the form $B(\lambda, m)$ and of the form (with $\mu > 0$)*

$$B'(\lambda,\mu,m) = \begin{pmatrix} \lambda & -\mu & 1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \mu & \lambda & 0 & 1 & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & -\mu & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & \mu & \lambda & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda & -\mu & 1 & 0 \\ 0 & 0 & 0 & 0 & \cdots & \mu & \lambda & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \lambda & -\mu \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & \mu & \lambda \end{pmatrix} \in \mathrm{Mat}(\mathbb{R}, 2m)\,.$$

Blocks $B(\lambda, m)$ occur for eigenvalues λ of f; blocks $B'(\lambda,\mu,m)$ occur if $P_f(x)$ is divisible by $x^2 - 2\lambda x + \lambda^2 + \mu^2$.

Warning: the *Real Jordan Normal Form* in this theorem is *not* a Jordan Normal Form, unless the characteristic polynomial of f splits into linear factors, in which case only blocks of the form $B(\lambda, m)$ occur.

Proof. Here is a sketch that gives the main ideas. First choose any basis $B = (x_1, \ldots, x_n)$ for V, so that $\varphi_B \colon \mathbb{R}^n \to V$ given by $(\lambda_1, \ldots, \lambda_n) \mapsto \sum_i \lambda_i x_i$ is an isomorphism. Identifying V with $\mathbb{R}^n$ through this isomorphism reduces the problem to the case $V = \mathbb{R}^n$, which is naturally contained in $\mathbb{C}^n$, and $f \colon \mathbb{R}^n \to \mathbb{R}^n$ being given by a real $n \times n$ matrix A.

Over $\mathbb{C}$, the characteristic polynomial $P_f = P_A$ will split into linear factors. Some of them will be of the form $x - \lambda$ with $\lambda \in \mathbb{R}$, the others will be of the form $x - (\lambda + \mu i)$ with $\lambda, \mu \in \mathbb{R}$ and $\mu \neq 0$. These latter ones occur in pairs

$$(x - (\lambda + \mu i))(x - (\lambda - \mu i)) = x^2 - 2\lambda x + \lambda^2 + \mu^2\,.$$

If $v_1, \ldots, v_m \in \mathbb{C}^n$ is a basis for the generalised eigenspace (over $\mathbb{C}$) for the eigenvalue $\lambda + \mu i$, then $\bar{v}_1, \ldots, \bar{v}_m$ is a basis for the generalised eigenspace for the eigenvalue $\lambda - \mu i$, where $\bar{v}$ denotes the vector obtained from $v \in \mathbb{C}^n$ by replacing each coordinate with its complex conjugate. If we now consider

$$(v_1 + \bar{v}_1), i(v_1 - \bar{v}_1), \ldots, (v_m + \bar{v}_m), i(v_m - \bar{v}_m)\,,$$

then these vectors are in $\mathbb{R}^n$ and form a basis for the sum of the two generalised eigenspaces. If $(v_1, \ldots, v_m)$ gives rise to a Jordan block $B(\lambda + \mu i, m)$, then we have

$$\begin{aligned} f(v_i + \bar{v}_i) &= f(v_i) + f(\bar{v}_i) = f(v_i) + \overline{f(v_i)} \\ &= (\lambda + \mu i)v_i + v'_{i-1} + (\lambda - \mu i)\bar{v}_i + \overline{v'_{i-1}} \\ &= \lambda(v_i + \bar{v}_i) + \mu i(v_i - \bar{v}_i) + v'_{i-1} + \overline{v'_{i-1}}, \\ f(i(v_i - \bar{v}_i)) &= i f(v_i) - i f(\bar{v}_i) = i \cdot f(v_i) - i \cdot \overline{f(v_i)} \\ &= i(\lambda + \mu i)v_i + i v'_{i-1} - i(\lambda - \mu i)\bar{v}_i - i\overline{v'_{i-1}} \\ &= \lambda i(v_i - \bar{v}_i) - \mu(v_i + \bar{v}_i) + i(v'_{i-1} - \overline{v'_{i-1}}), \end{aligned}$$

for $v'_{i-1} = 0$ if $i = 1$ and $v'_{i-1} = v_{i-1}$ if $i > 1$, so the new basis gives rise to a block of the form $B'(\lambda, \mu, m)$. $\square$

5.14. Theorem. *Let V be a finite-dimensional vector space, $f_1, \dots, f_k : V \to V$ diagonalizable endomorphisms that commute in pairs. Then $f_1, \dots, f_k$ are* simultaneously diagonalizable, *i.e., there is a basis for V consisting of vectors that are eigenvectors for all the f_j at the same time. In particular, any linear combination of the f_j is again diagonalizable.*

Proof. First note that if f and g are commuting endomorphisms and v is a λ-eigenvector of f, then $g(v)$ is again a λ-eigenvector of f (or zero):

$$f\bigl(g(v)\bigr) = g\bigl(f(v)\bigr) = g(\lambda v) = \lambda g(v)\,.$$

We now proceed by induction on k. For $k = 1$, there is nothing to prove. So assume $k \ge 2$. We can write $V = U_1 \oplus \cdots \oplus U_l$, where the U_i are the nontrivial eigenspaces of f_k. By the observation just made, we have splittings, for $j = 1, \dots, k-1$,

$$f_j = f_j^{(1)} \oplus \cdots \oplus f_j^{(l)} \qquad \text{with } f_j^{(i)} : U_i \to U_i.$$

By Corollary 3.12, the restrictions $f_j^{(i)} \colon U_i \to U_i$ are diagonalizable, so by the induction hypothesis, $f_1^{(i)}, \dots, f_{k-1}^{(i)}$ are simultaneously diagonalizable on U_i, for each i. Since U_i consists of eigenvectors of f_k, any basis for U_i that consists of eigenvectors for all the f_j with $j < k$, will also consist of eigenvectors for all the f_j with $j \le k$, that is, including $j = k$. To get a suitable basis for V, we take the concatenation of the bases of the various U_i. □

To finish this section, here is a uniqueness statement related to Corollary 5.8.

5.15. Theorem. *The diagonalizable and nilpotent parts of f in Corollary 5.8 are uniquely determined.*

Proof. Let $f = d + n = d' + n'$, where d and n are constructed as in the Jordan Normal Form Theorem 5.5, so with d diagonalizable and n nilpotent and $d \circ n = n \circ d$, and where d' is diagonalizable, n' is nilpotent, and $d' \circ n' = n' \circ d'$. Then d' and n' commute with f (as $d' \circ f = d' \circ d' + d' \circ n' = d' \circ d' + n' \circ d' = f \circ d'$, same for n'). Now let g be any endomorphism commuting with f, and consider $v \in U_j = \ker((f - \lambda_j \operatorname{id})^{m_j})$. Then

$$(f - \lambda_j \operatorname{id})^{m_j}\bigl(g(v)\bigr) = g\bigl((f - \lambda_j \operatorname{id})^{m_j}(v)\bigr) = g(0) = 0\,,$$

so $g(v) \in U_j$, i.e., U_j is g-invariant. So $g = g_1 \oplus \cdots \oplus g_k$ splits as a direct sum of endomorphisms of the generalised eigenspaces U_j of f. Since on U_j, we have $f|_{U_j} = \lambda_j \operatorname{id} + n|_{U_j}$ and g commutes with f, we find that g_j commutes with $n|_{U_j}$ for all j, hence g commutes with n (and also with d).

Applying this to d' and n', we see that d and d' commute, and that n and n' commute. We can write

$$d - d' = n' - n\,;$$

then the right hand side is nilpotent (for this we need that n and n' commute!). By Theorem 5.14, the left hand side is diagonalizable, so from Proposition 4.2 we conclude $d - d' = n' - n = 0$, that is, $d' = d$ and $n' = n$. □

As promised, we will now give some bigger examples of matrices that we will put in Jordan normal form.

5.16. Example. Consider the matrix

$$A = \begin{pmatrix} 2 & 3 & 3 & 3 & 3 \\ 0 & -1 & 0 & -1 & -1 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

We want an invertible matrix Q and a matrix J in Jordan normal form such that $A = QJQ^{-1}$. The characteristic polynomial of A is $(x-2)(x+1)^4$, so the eigenvalues are 2 and -1. The dimensions of the generalised eigenspaces equal the algebraic multiplicities, so they equal 1 and 4, respectively.

The dimension of the eigenspace associated to an eigenvalue is at least 1, so for the eigenvalue $\lambda = 2$ the associated eigenspace $\ker(A - 2I)$ is the whole generalised eigenspace, as both have dimension 1. The element e_1 is contained in the eigenspace, so e_1 generates this subspace.

For the eigenvalue $\lambda = -1$, we follow the proof of Theorem 4.8 (as $A + I$ is nilpotent on the generalised eigenspace for $\lambda = -1$). We have

$$A+I = \begin{pmatrix} 3 & 3 & 3 & 3 & 3 \\ 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad (A+I)^2 = \begin{pmatrix} 9 & 9 & 9 & 9 & 9 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$(A+I)^3 = \begin{pmatrix} 27 & 27 & 27 & 27 & 27 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Because $(A+I)^3$ has rank 1 we have $\dim\ker(A+I)^3 = 5-1 = 4$. As the generalised eigenspace has dimension 4, the subspace $U = \ker(A+I)^3$ is the whole generalised eigenspace. For each $n = 1, 2, 3$, the kernel $\ker(A+I)^n$ is easy to determine, since $(A+I)^n$ is almost in row echelon form. We find

$$\begin{aligned} \ker(A+I) &= L\big((-1,1,0,0,0),(-1,0,1,0,0)\big), \\ \ker(A+I)^2 &= L\big((-1,1,0,0,0),(-1,0,1,0,0),(-1,0,0,1,0)\big), \\ \ker(A+I)^3 &= L\big((-1,1,0,0,0),(-1,0,1,0,0),(-1,0,0,1,0),(-1,0,0,0,1)\big). \end{aligned}$$

For the dimension $r_n(-1) = \dim\ker(A+I)^n$ we have $r_1(-1) = 2$ and $r_2(-1) = 3$ and $r_3(-1) = 4$. We get $s_1(-1) = 2$ and $s_2(-1) = 1$ and $s_3(-1) = 1$. We also get $t_1(-1) = 1$ and $t_2(-1) = 0$ and $t_3(-1) = 1$, so there are two Jordan blocks, one of size 1×1 and one of size 3×3.

For the largest block, we choose a complementary subspace of $\ker(A+I)^2$ inside $\ker(A+I)^3$; this complementary space has dimension $s_3 = r_3 - r_2 = 1$, so it suffices to pick one vector: a vector in $\ker(A+I)^3 \setminus \ker(A+I)^2$, so for example $w_{31} = (-1,0,0,0,1)$. The other two vectors associated to the 3×3 block are $(A+I)w_{31} = (0,-1,0,1,0)$ and $(A+I)^2 w_{31} = (0,-1,1,0,0)$.

Any complementary subspace for $\ker(A+I)$ inside $\ker(A+I)^2$ has dimension $s_2 = r_2 - r_1 = 1$ as well, so $(A+I)w_{31}$ already generates such a complementary space. A complementary subspace for $\ker(A+I)^0 = \{0\}$ inside $\ker(A+I)$ is equal to $\ker(A+I)$, which has dimension 2; we already have a vector, namely $(A+I)^2 w_{31} = (0,-1,1,0,0)$, so in order to generate $\ker(A+I)$, it suffices to add a vector from $\ker(A+I)$ that is not a multiple of $(A+I)^2 w_{31}$. For example, we may choose $w_{11} = (-1,1,0,0,0)$. This vector corresponds to the 1×1 block.

The vectors $e_1, (A+I)^2 w_{31}, (A+I)w_{31}, w_{31}, w_{11}$ form a basis B. If we put the vectors in this order in a matrix, then we get

$$Q = [\mathrm{id}]^B_E = \begin{pmatrix} 1 & 0 & 0 & -1 & -1 \\ 0 & -1 & -1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

where E is the standard basis. The associated Jordan normal form is then

$$J = [f_A]^B_B = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Indeed, one verifies $QJQ^{-1} = [\mathrm{id}]^B_E \cdot [f_A]^B_B \cdot [\mathrm{id}]^E_B = [f_A]^E_E = A$.

5.17. Example. We consider the real matrix

$$M = \begin{pmatrix} -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 0 & -1 & 3 & -3 & 3 & -3 & 3 & -3 & 3 & -3 \\ 0 & 0 & 2 & 0 & 1 & -1 & 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 2 & 1 & -1 & 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 2 & 0 & 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 & -1 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix},$$

which has characteristic polynomial $(x+1)^2(x-2)^8$. Therefore, we have to deal with the two generalised eigenspaces

$$U_1 = \ker(M+I)^2 \qquad \text{and} \qquad U_2 = \ker(M-2I)^8$$

of dimensions 2 and 8, respectively. Indeed, by Theorem 5.4, we have $\mathbb{R}^{10} = U_1 \oplus U_2$. Let $e_1, \ldots, e_{10} \in \mathbb{R}^{10}$ denote the standard basis vectors.

We start with the larger case, namely U_2. By definition of U_2, the restriction of f_{M-2I} to U_2 is nilpotent, as f^8_{M-2I} restricts to 0 on U_2. By finding a row echelon form for $(M-2I)^n$ for $1 \le n \le 3$, we find $r_1(2) = \dim\ker(M-2I) = 4$ and $r_2(2) = \dim\ker(M-2I)^2 = 7$ and $r_3(2) = \dim\ker(M-2I)^3 = 8$. For $n > 3$ we have

$$8 = \dim\ker(M-2I)^3 \le \dim\ker(M-2I)^n \le \dim U_2 = 8,$$

so we conclude $\ker(M-2I)^3 = U_2$ and $r_n(2) = \dim\ker(M-2I)^n = 8$ for $n \ge 3$. This yields the following table for $s_n(2) = r_n(2) - r_{n-1}(2)$ and $t_n(2) = s_n(2) -$

$s_{n+1}(2)$.

n	$r_n(2)$	$s_n(2)$	$t_n(2)$
0	0		
1	4	4	1
2	7	3	2
3	8	1	1
4	8	0	0
5	8	0	0

We conclude that in any Jordan Normal Form for M, there is one Jordan block for eigenvalue 2 of size 1, there are two of size 2, and there is one of size 3.

As mentioned before, the restriction of f_{M-2I} to U_2 is nilpotent by definition of U_2. In fact, we have $(f_{M-2I}|_{U_2})^3 = 0$. To find a suitable basis for U_2, we follow the proof of Theorem 4.8, applied to this nilpotent endomorphism of U_2. We consider the filtration

$$\{0\} \subset \ker(M-2I) \subset \ker(M-2I)^2 \subset \ker(M-2I)^3 = U_2$$

and we will choose integers $t_1, t_2, t_3 \geq 0$ (which should turn out to be the values $t_j(2)$ from the table above) and elements $w_{jl} \in \ker(M-2I)^j$ with $1 \leq j \leq 3$ and $1 \leq l \leq t_j$ such that for each index $1 \leq j \leq 3$ the sequence

$$\big((M-2I)^{k-j}(w_{kl})\big)_{\substack{j\leq k\leq 3\\ 1\leq l\leq t_k}}$$

is a basis for a complementary subspace X_j of $\ker(M-2I)^{j-1}$ inside $\ker(M-2I)^j$.

We had already brought $(M-2I)^n$ into row echelon form before and we can use that to find explicit bases for $\ker(M-2I)^n$ for $1 \leq n \leq 3$. We find

$$\begin{aligned}\ker(M-2I) &= \langle x_1, x_2, x_3, x_4\rangle,\\ \ker(M-2I)^2 &= \langle y_1, y_2, y_3, y_4, y_5, y_6, y_7\rangle,\\ \ker(M-2I)^3 &= \langle z_1, z_2, z_3, z_4, z_5, z_6, z_7, z_8\rangle,\end{aligned}$$

with

$$\begin{aligned}x_1 &= (0,1,0,-1,0,0,0,0,0,0),\\ x_2 &= (0,0,1,1,0,0,0,0,0,0),\\ x_3 &= (0,0,0,0,1,1,0,0,0,0),\\ x_4 &= (0,0,0,0,0,0,1,1,0,0),\end{aligned}
\quad
\begin{aligned}y_1 &= (0,1,0,0,0,0,0,0,1,0),\\ y_2 &= (0,0,1,0,0,0,0,0,-1,0),\\ y_3 &= (0,0,0,1,0,0,0,0,1,0),\\ y_4 &= (0,0,0,0,1,0,0,0,-1,0),\\ y_5 &= (0,0,0,0,0,1,0,0,1,0),\\ y_6 &= (0,0,0,0,0,0,1,0,-1,0),\\ y_7 &= (0,0,0,0,0,0,0,1,1,0),\end{aligned}
\quad
\begin{aligned}z_1 &= (0,1,0,0,0,0,0,0,0,-1),\\ z_2 &= (0,0,1,0,0,0,0,0,0,1),\\ z_3 &= (0,0,0,1,0,0,0,0,0,-1),\\ z_4 &= (0,0,0,0,1,0,0,0,0,1),\\ z_5 &= (0,0,0,0,0,1,0,0,0,-1),\\ z_6 &= (0,0,0,0,0,0,1,0,0,1),\\ z_7 &= (0,0,0,0,0,0,0,1,0,-1),\\ z_8 &= (0,0,0,0,0,0,0,0,1,1).\end{aligned}$$

In the first step, corresponding to $j = m$ in the notation of the proof of Theorem 4.8, we want a complementary subspace X_3 of $\ker(M-2I)^2$ inside the subspace $\ker(M-2I)^3 = U_2$. One way to do this is to put the basis elements $y_1, \ldots, y_7$ for $\ker(M-2I)^2$ as columns in a matrix, and add the generators $z_1, \ldots, z_8$ for

$\ker(M-2I)^3$ as more columns to the right:

$$\left(\begin{array}{ccccccc|cccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & -1 & 1 & -1 & 1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1
\end{array}\right).$$

The reduced row echelon form for this matrix is

$$\left(\begin{array}{ccccccc|cccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right).$$

Of the added columns to the right, only the first has a pivot. This implies that the first of the added generators, namely z_1, generates a complementary space of $\ker(M-2I)^2$ inside $\ker(M-2I)^3$. [Of course, we could have seen this without any computation. From the last coordinate, we see that no z_i is contained in $\ker(M-2I)^2$, as the last coordinate of all the y_i is 0; since $\ker(M-2I)^2$ has codimension 1 inside $\ker(M-2I)^3$ (meaning the difference of their dimensions is 1), any element in $\ker(M-2I)^3$ that is not contained in $\ker(M-2I)^2$ generates a complementary space of $\ker(M-2I)^2$ inside $\ker(M-2I)^3$.] So, we take $t_3 = 1$ and $w_{31} = z_1$ and $X_3 = \langle w_{31} \rangle$.

The second step corresponds to $j = 2$. We want to extend $(M-2I)(X_3)$, that is, the image of X_3 under multiplication by $M-2I$, to a complementary subspace X_2 of $\ker(M-2I)$ inside $\ker(M-2I)^2$. We follow the proof of Lemma 2.6. First, note that $(M-2I)(X_2)$ has basis $(M-2I)w_{31} = (0,0,1,1,1,1,1,0,-1,0)$. We put the basis elements $x_1, \ldots, x_4$ for $\ker(M-2I)$ as columns in a matrix, we add $(M-2I)w_{31}$ as a column to the right, and we finally add the generators $y_1, \ldots, y_7$ for $\ker(M-2I)^2$ as columns to the far right:

$$\left(\begin{array}{cccc|c|ccccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
-1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right).$$

The reduced row echelon form for this matrix is

$$\left(\begin{array}{cccc|c|ccccccc}
1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & -1 \\
0 & 0 & 0 & 0 & 0 & 1 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right).$$

So of the last seven columns, the first and the fourth contain a pivot. This means that if we add y_1 and y_4 to $(M-2I)w_{31}$, then we obtain a basis for a complementary space X_2 of $\ker(M-2I)$ inside $\ker(M-2I)^2$. Hence, we set $t_2 = 2$ and $w_{21} = y_1$ and $w_{22} = y_4$ and we denote the space $\langle (M-2I)w_{31}, w_{21}, w_{22}\rangle$ by X_2.

In the step corresponding to $j = 1$, we extend $(M-2I)(X_2)$ to a complementary space X_1 of $\ker(M-2I)^0$ inside $\ker(M-2I)$. Since we have $(M-2I)^0 = I$, we find $\ker(M-2I)^0 = \{0\}$, so $X_1 = \ker(M-2I)$. Note that $(M-2I)(X_2)$ is generated by

$$\begin{aligned}
(M-2I)^2 w_{31} &= (0,0,0,0,0,0,-1,-1,0,0), \\
(M-2I)w_{21} &= (0,0,1,1,1,1,1,1,0,0), \\
(M-2I)w_{22} &= (0,0,0,0,-1,-1,-1,-1,0,0).
\end{aligned}$$

We put these as columns in a matrix and add columns for the generators $x_1, \ldots, x_4$ for $\ker(M-2I)$.

$$\left(\begin{array}{ccc|cccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & -1 & 1 & 0 & 0 \\
0 & 1 & -1 & 0 & 0 & 1 & 0 \\
0 & 1 & -1 & 0 & 0 & 1 & 0 \\
-1 & 1 & -1 & 0 & 0 & 0 & 1 \\
-1 & 1 & -1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right).$$

The reduced row echelon form for this matrix is

$$\left(\begin{array}{ccc|cccc}
1 & 0 & 0 & 0 & 0 & 1 & -1 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & -1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right).$$

Since only the first of the right-most four columns has a pivot, it suffices to add x_1 to the elements we already had in order to get a basis for $\ker(M-2I)$. In other words, we set $t_1 = 1$ and $w_{11} = x_1$ and let X_1 be the subspace generated by

$$\big(\ (M-2I)^2 w_{31},\ (M-2I)w_{21},\ (M-2I)w_{22}, w_{11}\ \big).$$

We now reorder the elements of the bases for X_1, X_2, X_3 to get a basis

$$C = \big((M-2I)^2 w_{31}, (M-2I)w_{31}, w_{31}, (M-2I)w_{22}, w_{22}, (M-2I)w_{21}, w_{21}, w_{11}\big)$$

for the generalised eigenspace $X_1 \oplus X_2 \oplus X_3 = U_2$. Note that indeed the integers t_j coincide with the integers $t_j(2)$ in the table above.

We continue with the generalised eigenspace U_1. By definition of U_1, the restriction of f_{M+I} to U_1 is nilpotent, as f_{M+I}^2 restricts to 0 on U_1. It is easy to verify that $\ker(M+I)$ is generated by e_1, while $\ker(M+I)^2$ is generated by e_1 and e_2. We proceed exactly the same as for U_2, but everything is much easier in this case. The vector e_2 generates a complementary space of $\ker(M+I)$ inside $\ker(M+I)^2$, so we set $v_{21} = e_2$. Its image under $M+I$ is $(M+I)v_{21} = e_1$, which, as we said, generates $\ker(M+I)$. Together, v_{21} and $(M+I)v_{21} = e_1$ form a basis D for the generalised eigenspace U_1.

The bases C and D together yield the basis

$$B = \big((M-2I)^2 w_{31}, (M-2I)w_{31}, w_{31}, (M-2I)w_{22}, w_{22}, (M-2I)w_{21}, w_{21}, w_{11}, (M+I)v_{21}, v_{21}\big)$$

for $U_1 \oplus U_2 = \mathbb{R}^{10}$. If we let E denote the standard basis for $\mathbb{R}^{10}$, then the matrix $P = [\mathrm{id}]^B_E$ has the elements of B as columns, that is,

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

We now already know that $[f_M]^B_B = [\mathrm{id}]^E_B [f_M]^E_E [\mathrm{id}]^B_E = P^{-1}MP$ is a matrix in Jordan Normal Form, with Jordan blocks $B(2,3), B(2,2), B(2,2), B(2,1)$ and $B(-1,2)$ in this order along the diagonal (for this notation, see Theorem 5.5). Indeed, a simple but tedious calculation shows

$$P^{-1}MP = \begin{pmatrix} 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{pmatrix}.$$

Exercises.

(1) Let V be a vector space over a field F, and $f\colon V \to V$ an endomorphism, and $\lambda \in F$ an element. Show that the generalised λ-eigenspace $\tilde{E}_\lambda(f)$ is a subspace.

(2) Give an example of an endomorphism f of a vector space V over a field F, and an element $\lambda \in F$, such that the generalised λ-eigenspace $\tilde{E}_\lambda(f)$ is not equal to $\ker(f - \lambda \operatorname{id}_V)^l$ for any $l \geq 0$.

(3) Let $J \in \operatorname{Mat}(n, F)$ be a matrix in Jordan Normal Form. Factorise the minimal polynomial M_J of J as

$$M_J = (x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}$$

with $\lambda_i \neq \lambda_j$ for $i \neq j$. Show that for each i, the multiplicity m_i is equal to the largest ℓ for which there is a block $B(\lambda_i, \ell)$ that occurs in J.

(4) In each of the following cases, give an example of a real 4×4-matrix A with the given properties, or explain why such a matrix does not exist. Here I denotes the 4×4 identity matrix.
   (a) $A^2 = 0$ and A has rank 1;
   (b) $A^2 = 0$ and A has rank 2;
   (c) $A^2 = 0$ and A has rank 3;
   (d) A has rank 2, and $A - I$ has rank 1;
   (e) A has rank 2, and $A - I$ has rank 2;
   (f) A has rank 2, and $A - I$ has rank 3.

(5) Let V be a finite-dimensional vector space over any field F. Let f be an endomorphism of V, and let $\lambda \in F$ be any scalar. Suppose $r > 0$ is an integer satisfying $\operatorname{rk}(f - \lambda \operatorname{id}_V)^r = \operatorname{rk}(f - \lambda \operatorname{id}_V)^{r+1}$. Show that for all $s > r$ we have $\operatorname{im}(f - \lambda \operatorname{id}_V)^r = \operatorname{im}(f - \lambda \operatorname{id}_V)^s$.

(6) For the following matrices A, B give their Jordan normal forms, and decide if they are similar.

$$A = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 2 & 0 \\ 1 & 1 & 2 & -1 \\ 0 & 0 & 2 & 2 \end{pmatrix} \qquad B = \begin{pmatrix} 2 & 0 & 0 & -2 \\ 1 & 2 & 1 & 0 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

(7) Give the Jordan normal form of the matrix

$$\begin{pmatrix} 2 & 2 & 0 & -1 \\ 0 & 0 & 0 & 1 \\ 1 & 5 & 2 & -2 \\ 0 & -4 & 0 & 4 \end{pmatrix}$$

(8) Give the Jordan normal form of the matrix

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

(9) Let A be the 3×3 matrix

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{pmatrix}.$$

Compute A^{100}.

(10) Consider the matrix $A = \begin{pmatrix} 1 & 4 \\ -1 & 5 \end{pmatrix}$.
   (a) Give the eigenvalues and eigenspaces of A.

(b) Give a diagonalizable matrix D and a nilpotent matrix N for which $D + N = A$ and $DN = ND$.
(c) Give a formula for A^n when $n = 1, 2, 3, \ldots$

(11) For the matrix
$$A = \begin{pmatrix} 2 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$
give a diagonalizable matix D and a nilpotent matrix N so that $A = D+N$ and $ND = DN$.

(12) For $A = \begin{pmatrix} 2 & 1 & -1 \\ 0 & 4 & -2 \\ 0 & 2 & 0 \end{pmatrix}$ compute the matrix e^A.

(13) Let $\phi\colon \mathbb{R}^3 \to \mathbb{R}^3$ be the linear map given by $\phi(x) = Ax$ where A is the matrix
$$\begin{pmatrix} 3 & 1 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$
We proved in class that generalised eigenspaces for ϕ are ϕ-invariant. What are these spaces in this case? Give all other ϕ-invariant subspaces of $\mathbb{R}^3$.

(14) Compute the characteristic polynomial of the matrix
$$A = \begin{pmatrix} 1 & -2 & 2 & -2 \\ 1 & -1 & 2 & 0 \\ 0 & 0 & -1 & 2 \\ 0 & 0 & -1 & 1 \end{pmatrix}$$
Does A have a Jordan normal form as 4×4 matrix over $\mathbb{R}$? What is the Jordan normal form of A as a 4×4 matrix over $\mathbb{C}$?

(15) Let n be a positive integer, and A a real $n \times n$ matrix. The table below shows the dimension of several subspaces of $\mathbb{R}^n$.

U	$\dim U$
$\ker A$	3
$\ker A^2$	3
$\operatorname{im} A^3$	8
$\ker(A - 2I_n)$	3
$\ker(A - 2I_n)^2$	6
$\ker(A - 2I_n)^3$	7
$\ker(A - 3I_n)^2$	1

(a) Prove that we have $\ker A = \ker A^2 = \ker A^3$.
(b) Prove that $n = 11$.
(c) Prove that A has a Jordan normal form over $\mathbb{R}$ (this does not refer to the real Jordan Normal Form as in Theorem 5.13).
(d) Show that the Jordan normal form for A is the determined uniquely, up to the order of the Jordan blocks, by the information in the table. Give a Jordan normal form for A.

(16) Suppose that for a 20×20 matrix A the rank of A^i for $i = 0, 1, \ldots 9$ is given by the sequence $20, 15, 11, 7, 5, 3, 1, 0, 0, 0$. What sizes are the Jordan-blocks in the Jordan normal form of A?

(17) Let V be a complex vector space of dimension at most 3, and f an endomorphism of V. Show that f is determined, up to similarity, by its characteristic and minimal polynomial together.

6. The Dual Vector Space

6.1. Definition. Let V be an F-vector space. A *linear form* or *linear functional* on V is a linear map $\phi : V \to F$.

The *dual vector space* of V is $V^* = \operatorname{Hom}(V, F)$, the vector space of all linear forms on V.

Recall how the vector space structure on $V^* = \operatorname{Hom}(V, F)$ is defined: for $\phi, \psi \in V^*$ and $\lambda, \mu \in F$, we have, for $v \in V$,

$$(\lambda\phi + \mu\psi)(v) = \lambda\phi(v) + \mu\psi(v)\,.$$

6.2. Example. Consider the standard example $V = F^n$. Then the *coordinate maps*

$$p_j : (x_1, \dots, x_n) \longmapsto x_j$$

are linear forms on V.

The following result is important.

6.3. Proposition and Definition. *Let V be a finite-dimensional vector space with basis $(v_1, \dots, v_n)$. Then V^* has a unique basis $(v_1^*, \dots, v_n^*)$ such that*

$$v_i^*(v_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

This basis $(v_1^*, \dots, v_n^*)$ of V^* is called the dual basis of $(v_1, \dots, v_n)$ or the basis dual to $(v_1, \dots, v_n)$.

Proof. Since linear maps are uniquely determined by their images on a basis, there certainly exist unique linear forms $v_i^* \in V^*$ with $v_i^*(v_j) = \delta_{ij}$. We have to show that they form a basis of V^*. First, it is easy to see that they are linearly independent, by applying a linear combination to the basis vectors v_j:

$$0 = (\lambda_1 v_1^* + \dots + \lambda_n v_n^*)(v_j) = \lambda_1\delta_{1j} + \dots + \lambda_n\delta_{nj} = \lambda_j\,.$$

It remains to show that the v_i^* generate V^*. So let $\phi \in V^*$. Then

$$\phi = \phi(v_1)v_1^* + \dots + \phi(v_n)v_n^*\,,$$

since both sides take the same values on the basis $v_1, \dots, v_n$. □

It is important to keep in mind that the dual basis vectors depend on *all* of $v_1, \dots, v_n$ — the notation v_j^* is *not* intended to imply that v_j^* depends only on v_j!

Note that for $\phi \in V^*$, we have

$$\phi = \sum_{j=1}^{n} \phi(v_j)v_j^*\,,$$

and for $v \in V$, we have

$$v = \sum_{i=1}^{n} v_i^*(v)v_i$$

(write $v = \lambda_1 v_1 + \dots + \lambda_n v_n$, then $v_i^*(v) = \lambda_i$).

6.4. Example. Consider $V = F^n$, with the canonical basis $E = (e_1, \dots, e_n)$. Then the dual basis is $P = (p_1, \dots, p_n)$ consisting of the coordinate maps from Example 6.2.

6.5. Corollary. *If V is finite-dimensional, then* $\dim V^* = \dim V$.

Proof. Clear from Prop. 6.3. □

6.6. Remark. The statement in Corollary 6.5 is actually an equivalence, if we define dimension to be the cardinality of a basis. If V has infinite dimension, then the dimension of V^* is "even more infinite". This is related to the following fact. Let B be a basis for V. Then the power set of B, i.e., the set of all subsets of B, has larger cardinality than B. To each subset S of B, we can associate an element $\psi_S \in V^*$ such that $\psi_S(b) = 1$ for $b \in S$ and $\psi_S(b) = 0$ for $b \in B \setminus S$. Now there are certainly linear relations between the ψ_S, but one can show that, if B is infinite, no subset of $\{\psi_S : S \subset B\}$ whose cardinality is that of B can generate all the ψ_S. Therefore any basis for V^* must be of strictly larger cardinality than B.

Note that again, we are implicitly assuming that every vector space has a basis (cf. Remark 2.3). Also, we are using the fact that for any basis $B = (v_i)_{i \in I}$ of V and any collection $C = (w_i)_{i \in I}$ of elements in a vector space W, there is a linear map $\varphi \colon V \to W$ that sends v_i to w_i for each $i \in I$. Indeed, this follows from the fact that the map $\varphi_B \colon F^{(I)} \to V$ that sends $(\lambda_i)_{i \in I}$ to $\sum_i \lambda_i v_i$ is an isomorphism, so the map $\varphi \colon V \to W$ is $\varphi_C \circ \varphi_B^{-1}$. See Exercises 3.1.9, 4.4.7 of *Linear Algebra I, 2018 edition*, also to recall that $F^{(I)}$ denotes the vector space of all functions from $I \to F$ that are zero for all but finitely many elements of I.

6.7. Example. If $V = L(\sin, \cos)$ (a linear subspace of the real vector space of real-valued functions on $\mathbb{R}$), then the basis dual to $\sin, \cos$ is given by the functionals $f \mapsto f(\pi/2), f \mapsto f(0)$.

6.8. Theorem. *Let V be a vector space and $V^{**} = (V^*)^*$ its* bidual. *Then the map $\alpha_V : V \to V^{**}$ that sends $v \in V$ to the linear map $\alpha_V(v) \colon V^* \to F$ given by $V^* \ni \phi \mapsto \phi(v)$ is an injective homomorphism; moreover, α_V is an isomorphism when V is finite-dimensional.*

Proof. We sometimes denote the evaluation map $\alpha_V(v) \colon V^* \to F$ by ev_v, though this notation may also be used for any other evaluation map (cf. Example 6.10). Then $\alpha_V(v)$ is a linear form on V^* by the definition of the linear structure on V^*. Also, α_V is itself linear:

$$\begin{aligned} \alpha_V(\lambda v + \lambda' v')(\phi) &= \phi(\lambda v + \lambda' v') = \lambda \phi(v) + \lambda' \phi(v') \\ &= \lambda \alpha_V(v)(\phi) + \lambda' \alpha_V(v')(\phi) = (\lambda \alpha_V(v) + \lambda' \alpha_V(v'))(\phi) . \end{aligned}$$

In order to prove that α_V is injective, it suffices to show that its kernel is trivial. So let $0 \neq v \in V$. Using Zorn's Lemma from Set Theory (cf. Remark 2.3 and see Appendix E of *Linear Algebra I, 2020 edition, or later*), we can choose a basis for V containing v. Then there is a linear form ϕ on V such that $\phi(v) = 1$ (and $\phi(w) = 0$ on all the other basis elements, say). But this means $\alpha_V(v)(\phi) = 1$, so $\alpha_V(v) \neq 0$ and $v \notin \ker \alpha_V$.

Finally, if V is finite-dimensional, then by Corollary 6.5, we have $\dim V^{**} = \dim V^* = \dim V$, so α_V must be surjective as well (use $\dim \operatorname{im}(\alpha_V) = \dim V - \dim \ker(\alpha_V) = \dim V^{**}$.) □

6.9. Corollary. *Let V be a finite-dimensional vector space, and let $(\phi_1, \dots, \phi_n)$ be a basis for V^*. Then there is a unique basis $(v_1, \dots, v_n)$ of V with $\phi_i(v_j) = \delta_{ij}$.*

PROOF. By Prop. 6.3, there is a unique dual basis $(\phi_1^*, \dots, \phi_n^*)$ of $V^{**} = (V^*)^*$. Since α_V is an isomorphism, there are unique $v_1, \dots, v_n$ in V such that $\alpha_V(v_j) = \phi_j^*$. They form a basis for V, and

$$\phi_i(v_j) = \mathrm{ev}_{v_j}(\phi_i) = \alpha_V(v_j)(\phi_i) = \phi_j^*(\phi_i) = \delta_{ij} \,.$$

□

In other words, $(\phi_1, \dots, \phi_n)$ is the basis for V^* dual to $(v_1, \dots, v_n)$.

6.10. Example. Let V be the vector space of polynomials of degree less than n; then $\dim V = n$. For any $\alpha \in F$, the evaluation map

$$\mathrm{ev}_\alpha : V \ni p \mapsto p(\alpha) \in F$$

is a linear form on V. Now pick $\alpha_1, \dots, \alpha_n \in F$ distinct. Then $\mathrm{ev}_{\alpha_1}, \dots, \mathrm{ev}_{\alpha_n} \in V^*$ are linearly independent, hence form a basis. (This comes from the fact that the *Vandermonde matrix* $(\alpha_i^j)_{1\le i\le n, 0\le j\le n-1}$ has determinant $\prod_{i<j}(\alpha_j - \alpha_i) \neq 0$.) What is the basis for V dual to that? What we need are polynomials $p_1, \dots, p_n$ of degree less than n such that $p_i(\alpha_j) = \delta_{ij}$. So $p_i(x)$ has to be a multiple of $\prod_{j\neq i}(x - \alpha_j)$. We then obtain

$$p_i(x) = \prod_{j\neq i} \frac{x - \alpha_j}{\alpha_i - \alpha_j} \,,$$

these are exactly the *Lagrange interpolation polynomials.*

We then find that the unique polynomial of degree less than n that takes the value β_j on α_j, for all j, is given by

$$p(x) = \sum_{j=1}^{n} \beta_j p_j(x) = \sum_{j=1}^{n} \beta_j \prod_{i\neq j} \frac{x - \alpha_i}{\alpha_j - \alpha_i} \,.$$

So far, we know how to 'dualize' vector spaces (and bases). Now we will see how we can also 'dualize' linear maps.

6.11. Definition. Let V and W be F-vector spaces, $f : V \to W$ a linear map. Then the *transpose* or *dual* linear map of f is defined as

$$f^\top : W^* \longrightarrow V^* \,, \quad \psi \longmapsto f^\top(\psi) = \psi \circ f \,.$$

A diagram clarifies perhaps what is happening here.

$$V \xrightarrow{\ f\ } W \xrightarrow{\ \psi\ } F$$

The composition $\psi \circ f$ is a linear map from V to F, and is therefore an element of V^*. It is easy to see that $f^\top$ is again linear: for $\psi_1, \psi_2 \in W^*$ and $\lambda_1, \lambda_2 \in F$, we have

$$f^\top(\lambda_1\psi_1 + \lambda_2\psi_2) = (\lambda_1\psi_1 + \lambda_2\psi_2) \circ f = \lambda_1\psi_1 \circ f + \lambda_2\psi_2 \circ f = \lambda_1 f^\top(\psi_1) + \lambda_2 f^\top(\psi_2) \,.$$

Also note that for linear maps $f_1, f_2 : V \to W$ and scalars λ_1, λ_2, we have

$$(\lambda_1 f_1 + \lambda_2 f_2)^\top = \lambda_1 f_1^\top + \lambda_2 f_2^\top \,,$$

and for linear maps $f_1 : V_1 \to V_2$, $f_2 : V_2 \to V_3$, we obtain $(f_2 \circ f_1)^\top = f_1^\top \circ f_2^\top$ — note the reversal.

Another simple observation is that $\mathrm{id}_V^\top = \mathrm{id}_{V^*}$.

6.12. Proposition. *Let $f : V \to W$ be an isomorphism. Then $f^\top : W^* \to V^*$ is also an isomorphism, and $(f^\top)^{-1} = (f^{-1})^\top$.*

Proof. We have $f \circ f^{-1} = \mathrm{id}_W$ and $f^{-1} \circ f = \mathrm{id}_V$. This implies that

$$(f^{-1})^\top \circ f^\top = \mathrm{id}_{W^*} \quad \text{and} \quad f^\top \circ (f^{-1})^\top = \mathrm{id}_{V^*}\ .$$

The claim follows. □

We denote the standard scalar product (dot product) on F^n by $\langle \_, \_ \rangle$. While working with general vector spaces, it is often advisable to avoid choosing a basis, as there usually is no natural choice. However, the vector space F^n comes with a standard basis $E = (e_1, e_2, \ldots, e_n)$, and its dual $(F^n)^*$ with the associated dual basis $P = (p_1, \ldots, p_n)$ of coordinate maps (see Example 6.2). We denote the associated map $\varphi_P \colon F^n \to (F^n)^*$ by φ_n; it sends e_i to the linear form $p_i = \langle e_i, \_ \rangle$, which sends $x \in F^n$ to $\langle e_i, x \rangle$. We conclude that, in general, φ_n sends $a \in F^n$ to the linear form $\langle a, \_ \rangle$. Indeed, φ_n and the map $F^n \to (F^n)^*$ given by $a \mapsto \langle a, \_ \rangle$ coincide on a basis, so they are the same.

6.13. Lemma. *Let V be a finite-dimensional F-vector space with basis B of dimension n, and let B^* be the corresponding dual basis for the dual space V^*. Let $\varphi_B \colon F^n \to V$ and $\varphi_{B^*} \colon F^n \to V^*$ be the usual linear maps sending the i-th standard basis vector to the i-th vector in B and B^*, respectively. Then the composition $\varphi_B^\top \circ \varphi_{B^*} \colon F^n \to (F^n)^*$ is φ_n.*

Proof. It suffices to check that the two maps are the same on the standard basis vectors $e_i \in F^n$. Write $B = (v_1, \ldots, v_n)$ and $B^* = (v_1^*, \ldots, v_n^*)$. Then for each index $1 \le i \le n$, we have $\varphi_{B^*}(e_i) = v_i^*$, and therefore $(\varphi_B^\top \circ \varphi_{B^*})(e_i) = \varphi_B^\top(v_i^*) = v_i^* \circ \varphi_B$. For each index $1 \le j \le n$ we have $(v_i^* \circ \varphi_B)(e_j) = v_i^*(v_j) = \delta_{ij} = p_i(e_j)$, which implies that $v_i^* \circ \varphi_B = p_i = \varphi_n(e_i)$. The statement follows. □

The reason for calling $f^\top$ the "transpose" of f becomes clear through the following result.

6.14. Lemma. *Let m, n be nonnegative integers, and $A \in \mathrm{Mat}(m \times n, F)$ a matrix. Let $f_A \colon F^n \to F^m$ and $f_{A^\top} \colon F^m \to F^n$ be the linear maps associated to A and its transpose $A^\top$, respectively. Then we have $f_{A^\top} = \varphi_n^{-1} \circ f_A^\top \circ \varphi_m$ and the diagram*

$$\begin{array}{ccc} (F^m)^* & \xrightarrow{\ f_A^\top\ } & (F^n)^* \\ {\scriptstyle \varphi_m}\uparrow & & \uparrow{\scriptstyle \varphi_n} \\ F^m & \xrightarrow[\ f_{A^\top}\ ]{} & F^n \end{array}$$

commutes.

Proof. Both statements are equivalent to the equality $\varphi_n \circ f_{A^\top} = f_A^\top \circ \varphi_m$, which we now verify. For each $a \in F^m$ and $x \in F^n$, we have, if we identify them with an $m \times 1$ and an $n \times 1$ matrix, respectively,

$$\big((\varphi_n \circ f_{A^\top})(a)\big)(x) = \big(\varphi_n(A^\top a)\big)(x) = \langle A^\top a, x \rangle = (A^\top a)^\top x = a^\top A x,$$

and

$$\big((f_A^\top \circ \varphi_m)(a)\big)(x) = \big(f_A^\top(\langle \_, a \rangle)\big)(x) = \big(\langle a, \_ \rangle \circ f_A\big)(x) = \langle a, Ax \rangle = a^\top A x.$$

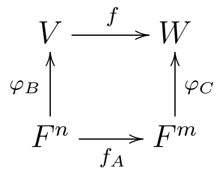
These are equal for all $x \in F^n$, so we conclude $(\varphi_n \circ f_{A^\top})(a) = (f_A^\top \circ \varphi_m)(a)$ for all $a \in F^m$, which implies $\varphi_n \circ f_{A^\top} = f_A^\top \circ \varphi_m$. $\square$

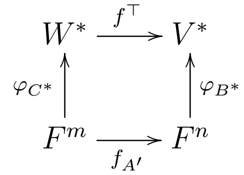
The following proposition is a generalisation of the previous lemma.

6.15. Proposition. *Let V and W be finite-dimensional vector spaces, with bases $B = (v_1, \dots, v_n)$ and $C = (w_1, \dots, w_m)$, respectively. Let $B^* = (v_1^*, \dots, v_n^*)$ and $C^* = (w_1^*, \dots, w_m^*)$ be the corresponding dual bases of V^* and W^*, respectively. Let $f : V \to W$ be a linear map, represented by the matrix A with respect to the given bases of V and W. Then the matrix representing $f^\top$ with respect to the dual bases is $A^\top$, that is*

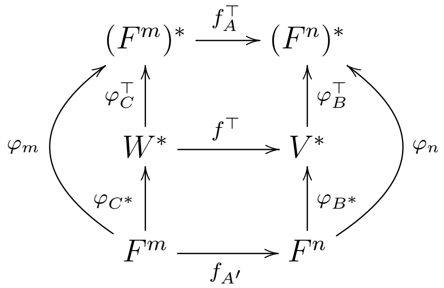
$$[f^\top]_{B^*}^{C^*} = ([f]_C^B)^\top.$$

Proof. We have the following two commutative diagrams





with $A = [f]_C^B$ and $A' = [f^\top]_{B^*}^{C^*}$. The dual of the first diagram can be combined with the second to obtain the following commutative diagram



where the two curved compositions are φ_m and φ_n by Lemma 6.13. We conclude from Lemma 6.14 that $f_{A'} = \varphi_n^{-1} \circ f_A^\top \circ \varphi_m = f_{A^\top}$, so $A' = A^\top$. $\square$

Alternative proof. Let $A = (a_{ij})_{1 \le i \le m, 1 \le j \le n}$; then

$$f(v_j) = \sum_{i=1}^{m} a_{ij} w_i \,.$$

We then have

$$\big(f^\top(w_i^*)\big)(v_j) = (w_i^* \circ f)(v_j) = w_i^*\big(f(v_j)\big) = w_i^*\Big(\sum_{k=1}^{m} a_{kj} w_k\Big) = a_{ij} \,.$$

Since we always have, for $\phi \in V^*$, that $\phi = \sum_{j=1}^{n} \phi(v_j) v_j^*$, this implies that

$$f^\top(w_i^*) = \sum_{j=1}^{n} a_{ij} v_j^* \,.$$

Therefore the columns of the matrix representing $f^\top$ with respect to the dual bases are exactly the rows of A. $\square$

Note that for every invertible matrix P we have $(P^{-1})^\top = (P^\top)^{-1}$; we will denote this matrix by $P^{-\top}$.

6.16. Corollary. *Let V be a finite-dimensional vector space, and let $B = (v_1, \dots, v_n)$ and $C = (w_1, \dots, w_n)$ be two bases of V. Let $B^* = (v_1^*, \dots, v_n^*)$ and $C^* = (w_1^*, \dots, w_n^*)$ be the corresponding dual bases. Then we have*

$$[\mathrm{id}_{V^*}]^{B^*}_{C^*} = \big([\mathrm{id}_V]^B_C\big)^{-\top}.$$

Proof. Using $\mathrm{id}_V^\top = \mathrm{id}_{V^*}$, we find from Proposition 6.15 that $[\mathrm{id}_{V^*}]^{C^*}_{B^*} = \big([\mathrm{id}_V]^B_C\big)^\top$. The statement now follows from the fact that the matrices $[\mathrm{id}_{V^*}]^{C^*}_{B^*}$ and $[\mathrm{id}_{V^*}]^{B^*}_{C^*}$ are each other's inverses. □

This corollary is reflected in the matrices we use to change bases. If $f\colon V \to V$ is an endomorphism and we set $A = [f]^B_B$ and $A' = [f]^C_C$, then for the matrix $P = [\mathrm{id}_V]^B_C$ we have $A' = PAP^{-1}$. The matrices $A^\top = [f^\top]^{B^*}_{B^*}$ and $A'^\top = [f^\top]^{C^*}_{C^*}$ are then related by $A'^\top = (PAP^{-1})^\top = P^{-\top}A^\top P^\top$.

As is to be expected, we have a compatibility between $f^{\top\top}$ and the canonical map α_V.

6.17. Proposition. *Let V and W be vector spaces, $f : V \to W$ a linear map. Then the following diagram commutes.*

$$\begin{array}{ccc} V & \xrightarrow{\;f\;} & W \\ {\scriptstyle\alpha_V}\downarrow & & \downarrow{\scriptstyle\alpha_W} \\ V^{**} & \xrightarrow{\;f^{\top\top}\;} & W^{**} \end{array}$$

Proof. We have to show that $f^{\top\top} \circ \alpha_V = \alpha_W \circ f$. So let $v \in V$ and $\psi \in W^*$. Then

$$\begin{aligned} f^{\top\top}\big(\alpha_V(v)\big)(\psi) &= (\alpha_V(v) \circ f^\top)(\psi) = \alpha_V(v)\big(f^\top(\psi)\big) \\ &= \alpha_V(v)(\psi \circ f) = (\psi \circ f)(v) \\ &= \psi\big(f(v)\big) = \alpha_W\big(f(v)\big)(\psi)\,. \end{aligned}$$

□

6.18. Proposition. *Let V and W be finite-dimensional vector spaces. Then*

$$\mathrm{Hom}(V, W) \ni f \longmapsto f^\top \in \mathrm{Hom}(W^*, V^*)$$

is an isomorphism.

Proof. By the observations made in Definition 6.11, the map is linear. We claim that the map

$$\mathrm{Hom}(W^*, V^*) \ni g \longmapsto \alpha_W^{-1} \circ g^\top \circ \alpha_V \in \mathrm{Hom}(V, W)\,,$$

is the inverse of the given map. Because $\mathrm{Hom}(V, W)$ and $\mathrm{Hom}(W^*, V^*)$ have the same finite dimension, it suffices to verify that the composition of the two maps in only one of the two orders is the identity. Indeed, by Proposition 6.17 we have

$$\alpha_W^{-1} \circ (f^\top)^\top \circ \alpha_V = f.$$

□

The following lemma states that every linear form on a subspace U of a vector space V can be extended to a linear form on V. Note that if $j\colon U \to V$ is an inclusion map, then $j^\top\colon V^* \to U^*$ is the restriction map that sends $\varphi \in V^*$ to $\varphi|_U$.

6.19. Lemma. *Let V be a vector space and $U \subset V$ a subspace. Let $j\colon U \hookrightarrow V$ denote the inclusion map. Then $j^\top\colon V^* \to U^*$ is surjective.*

Proof. Let $U' \subset V$ be a complementary space of U (using Zorn's Lemma if V is infinite-dimensional), and $\pi\colon V \to U$ the projection onto U along U'. That is, for $v = u + u'$ with $u \in U$ and $u' \in U'$, we have $\pi(v) = u$. Then we have $\pi \circ j = \mathrm{id}_U$, so $j^\top \circ \pi^\top = (\pi \circ j)^\top = \mathrm{id}_{U^*}$, which implies that $j^\top$ is surjective. □

6.20. Proposition. *Let $f\colon U \to V$ and $g\colon V \to W$ be two linear maps of vector spaces.*

(1) *If we have* $\mathrm{im}\, f \subset \ker g$, *then we have* $\mathrm{im}\, g^\top \subset \ker f^\top$.
(2) *If we have* $\ker g \subset \mathrm{im}\, f$, *then we have* $\ker f^\top \subset \mathrm{im}\, g^\top$.
(3) *If we have* $\mathrm{im}\, f = \ker g$, *then we have* $\mathrm{im}\, g^\top = \ker f^\top$.

Proof. (1) Suppose $\mathrm{im}\, f \subset \ker g$. Then the composition $g \circ f$ is the zero map. Hence so is the dual of this composition, which is the composition $f^\top \circ g^\top$ of the duals. This implies $\mathrm{im}\, g^\top \subset \ker f^\top$.

(2) Write g as the composition $g = j \circ \tilde{g}$ with $\tilde{g}\colon V \to \mathrm{im}\, g$ and $j\colon\ \mathrm{im}\, g \to W$ the inclusion map. Then we have $\ker g = \ker \tilde{g}$. From Lemma 6.19 we find that $j^\top$ is surjective, so from $g^\top = \tilde{g}^\top \circ j^\top$ we obtain $\mathrm{im}\, g^\top = \mathrm{im}\, \tilde{g}^\top$. Hence it suffices to prove the statement with $\tilde{g}$ instead of g, so without loss of generality, we may and will assume g is surjective.

Suppose $\ker g \subset \mathrm{im}\, f$. Take any $\varphi \in \ker f^\top$, so $f^\top(\varphi) = 0$, that is, for all $u \in U$ we have $\varphi(f(u)) = 0$. For each $w \in W$, there is a $v \in V$ with $g(v) = w$, since g is surjective; for $v, v' \in V$ with $g(v) = g(v') = w$, we have $v - v' \in \ker g \subset \mathrm{im}\, f$, so there is a $u \in U$ with $f(u) = v - v'$, and therefore $\varphi(v) = \varphi(v - v') + \varphi(v') = \varphi(f(u)) + \varphi(v') = \varphi(v')$. Hence, there is a well-defined map $\psi\colon W \to F$ with $\psi(g(v)) = \varphi(v)$ for all $v \in V$. To verify that ψ is linear, note that if $w = g(v)$ and $w' = g(v')$, then we have $w + w' = g(v + v')$, so

$$\psi(w + w') = \varphi(v + v') = \varphi(v) + \varphi(v') = \psi(w) + \psi(w');$$

The fact that ψ respects scalar multiplication follows similarly. We conclude that $\psi \in W^*$, and $\varphi = g^\top(\psi) \in \mathrm{im}\, g^\top$, so $\ker f^\top \subset \mathrm{im}\, g^\top$.

(3) This follows from the previous statements.

□

6.21. Definition. A sequence

$$V_0 \xrightarrow{f_1} V_1 \xrightarrow{f_2} \dots \xrightarrow{f_n} V_n$$

of composable linear maps is called *exact* if for all indices $1 \le i < n$ we have $\mathrm{im}\, f_i = \ker f_{i+1}$.

Proposition 6.20 states that if $U \to V \to W$ is an exact sequence, then the induced sequence $W^* \to V^* \to U^*$ is exact as well. Note that a linear map $f\colon V \to W$ is injective if and only if the sequence $0 \to V \xrightarrow{f} W$ is exact, while f is surjective if and only if the sequence $V \xrightarrow{f} W \to 0$ is exact.

6.22. Corollary. *Let $f\colon V \to W$ be a linear map of vector spaces. If f is injective, then $f^\top$ is surjective. If f is surjective, then $f^\top$ is injective.*

Proof. If f is injective, then the sequence $0 \to V \xrightarrow{f} W$ is exact. Then by Proposition 6.20 the sequence $W^* \xrightarrow{f^\top} V^* \to 0$ is exact, so $f^\top$ is surjective. As an alternative proof, we could have also written f as the composition $f = j \circ \tilde{f}$ of the isomorphism $\tilde{f}\colon V \to \operatorname{im} f$ induced by f, and the inclusion $j\colon \operatorname{im} f \to W$; then by Proposition 6.12 and Lemma 6.19, the map $f^\top = \tilde{f}^\top \circ j^\top$ is the composition of a surjection and an isomorphism, and thus surjective.

If f is surjective, then the sequence $V \xrightarrow{f} W \to 0$ is exact. Then by Proposition 6.20 the sequence $0 \to W^* \xrightarrow{f^\top} V^*$ is exact, so $f^\top$ is injective. □

6.23. Proposition. *Let $f\colon V \to W$ be a linear map of finite-dimensional vector spaces. Then we have* $\dim \operatorname{im} f = \dim \operatorname{im} f^\top$.

Proof. The map f is the composition of the surjection $\tilde{f}\colon V \to \operatorname{im} f$ induced by f and the inclusion $j\colon \operatorname{im} f \to W$. By Corollary 6.22, the dual map $f^\top$ is the composition of the surjective map $j^\top\colon W^* \to (\operatorname{im} f)^*$ and the injective map $\tilde{f}^\top\colon (\operatorname{im} f)^* \to V^*$. We conclude $\operatorname{im} f^\top = \operatorname{im} \tilde{f}^\top$ and hence

$$\dim \operatorname{im} f^\top = \dim \operatorname{im} \tilde{f}^\top = \dim(\operatorname{im} f)^* = \dim \operatorname{im} f.$$

□

Note that [**BR2**] claims (in Theorem 7.8) that we also have $\dim \ker(f^\top) = \dim \ker(f)$. However, this is **false** unless $\dim V = \dim W$!

6.24. Remark. The equality of dimensions $\dim \operatorname{im}(f^\top) = \dim \operatorname{im}(f)$ is, by Prop. 6.15, equivalent to the statement "row rank equals column rank" for matrices.

6.25. Definition. Let $A \in \operatorname{Mat}(m \times n, F)$ be a matrix. A *kernel matrix* of A is a matrix whose columns span the kernel of A.

If B is a kernel matrix of A, then we have $\operatorname{im} f_B = \ker f_A$. By Proposition 6.20, this implies $\operatorname{im} f_A^\top = \ker f_B^\top \subset (F^n)^*$. Applying φ_n^{-1}, we obtain the equality $\operatorname{im} f_{A^\top} = \ker f_{B^\top}$ by Lemma 6.14. This shows that $A^\top$ is a kernel matrix of $B^\top$.

Next, we study how subspaces relate to dualization.

6.26. Definition. Let V be a vector space and $S \subset V$ a subset. Then

$$S^\circ = \{\phi \in V^* : \phi(v) = 0 \text{ for all } v \in S\} \subset V^*$$

is called the *annihilator* of S.

The set S° is a linear subspace of V^*, since we can write

$$S^\circ = \bigcap_{v \in S} \ker\bigl(\alpha_V(v)\bigr).$$

Trivial examples are $\{0_V\}^\circ = V^*$ and $V^\circ = \{0_{V^*}\}$.

6.27. Remark. As we have seen before, if U is a subspace of a vector space V, and $j\colon U \to V$ is the inclusion map, then $j^\top\colon V^* \to U^*$ is the restriction map, which sends each linear form $\psi \in V^*$ to its restriction $\psi|_U$; we have

$$U^\circ = \ker j^\top.$$

6.28. Theorem. *Let V be a finite-dimensional vector space, $U \subset V$ a linear subspace. Then we have*

$$\dim U + \dim U^\circ = \dim V \quad \textit{and} \quad \alpha_V(U) = U^{\circ\circ}.$$

Proof. As in Remark 6.27, the dual of the inclusion $j\colon U \hookrightarrow V$ is a surjective map $V^* \to U^*$, of which the kernel is U°. Hence, we have $\dim U^\circ + \dim U^* = \dim V^*$, even if V were not finite-dimensional. Because V is finite-dimensional, we have $\dim V = \dim V^*$ and $\dim U = \dim U^*$, so the first equality follows. Applying it to U°, we obtain $\dim U = \dim U^{\circ\circ}$.

For the second equality, note that U° consists of all the linear forms on V that vanish on U. Hence, for every $u \in U$, the evaluation map $\mathrm{ev}_u\colon V^* \to F$ sending $\varphi \in V^*$ to $\varphi(u)$ sends all of U° to 0. This implies that the element $\alpha_V(u) = \mathrm{ev}_u \in V^{**}$ is contained in $U^{\circ\circ}$, so we have $\alpha_V(U) \subset U^{\circ\circ}$, even if V were not finite-dimensional. Because V is finite-dimensional, we have $\dim \alpha_V(U) = \dim U = \dim U^{\circ\circ}$, so the inclusion $\alpha_V(U) \subset U^{\circ\circ}$ is an equality. □

The theorem implies that we have $U^{\circ\circ} = U$ if we identify V and V^{**} via α_V.

6.29. Theorem. *Let $f\colon V \to W$ be a linear map of vector spaces. Then we have*

$$\big(\ker(f)\big)^\circ = \operatorname{im}(f^\top) \quad \textit{and} \quad \big(\operatorname{im}(f)\big)^\circ = \ker(f^\top).$$

Proof. Let $j\colon \ker f \to V$ be the inclusion map. Apply Proposition 6.20 to the exact sequence

$$\ker f \xrightarrow{j} V \xrightarrow{f} W$$

to get the exact sequence

$$W^* \xrightarrow{f^\top} V^* \xrightarrow{j^\top} (\ker f)^*,$$

which implies $\operatorname{im} f^\top = \ker j^\top = (\ker f)^\circ$, which proves the first equality. For the second equality, let $i\colon \operatorname{im} f \to W$ denote the inclusion map, and write f as the composition $f = i \circ \tilde{f}$ with $\tilde{f}\colon V \to \operatorname{im} f$ induced by f. Then $f^\top = \tilde{f}^\top \circ i^\top$, and since $\tilde{f}^\top$ is injective, we obtain $\ker f^\top = \ker i^\top = (\operatorname{im} f)^\circ$. □

6.30. Interpretation in Terms of Matrices. Let us consider the vector spaces $V = F^n$ and $W = F^m$ and a linear map $f : V \to W$. Then f is represented by a matrix A, and the image of f is the column space of A, i.e., the subspace of F^m spanned by the columns of A. We identify $V^* = (F^n)^*$ and $W^* = (F^m)^*$ with F^n and F^m via the dual bases consisting of the coordinate maps (see the text above Lemma 6.13). Then for $x \in W^*$, we have $x \in (\operatorname{im}(f))^\circ$ if and only if $x^\top y = \langle x, y\rangle = 0$ for all columns y of A, which is the case if and only if $x^\top A = 0$. This is equivalent to $A^\top x = 0$, which says that $x \in \ker(f^\top)$ — remember that $A^\top$ represents $f^\top : W^* \to V^*$.

Exercises.

(1) Define $\phi_i \colon \mathbb{R}^n \to \mathbb{R}$ by $\phi_i(x_1, \dots, x_n) = x_1 + x_2 + \cdots + x_i$ for $i = 1, 2, \dots n$. Show that $\Phi = (\phi_1, \dots, \phi_n)$ is a basis for $(\mathbb{R}^n)^*$, and compute a basis B for $\mathbb{R}^n$ of which Φ is the dual basis.

(2) Let V be an n-dimensional vector space, let $v_1, \dots, v_n \in V$ and let $\phi_1, \dots, \phi_n \in V^*$. Show that $\det((\phi_i(v_j))_{i,j})$ is non-zero if and only if $(v_1, \dots, v_n)$ is a basis for V and $(\phi_1, \dots, \phi_n)$ is a basis for V^*.

(3) Let V be the 3-dimensional vector space of polynomial functions $\mathbb{R} \to \mathbb{R}$ of degree at most 2. In each of the following cases, we define $\phi_i \in V^*$ for $i = 0, 1, 2$. In each case, indicate whether (ϕ_0, ϕ_1, ϕ_2) is a basis for V^*, and if so, give the dual basis for V.
   (a) $\phi_i(f) = f(i)$
   (b) $\phi_i(f) = f^{(i)}(0)$, i.e., the ith derivative of f evaluated at 0.
   (c) $\phi_i(f) = f^{(i)}(1)$
   (d) $\phi_i(f) = \int_{-1}^{i} f(x)dx$

(4) Let $V = \mathbb{R}[X]_2$ be the space of polynomials of degree at most two. Take $\alpha, \beta, \gamma \in V^*$ to be given by

$$\alpha(f) = f(0),$$
$$\beta(f) = \int_0^1 f(x)\, dx,$$
$$\gamma(f) = f'(0)$$

for all $f \in V$, and where f' denotes the derivative of f. Show that (α, β, γ) is a basis for V^*.

(5) For each positive integer n show that there are constants $a_1, a_2, \dots, a_n$ so that

$$\int_0^1 f(x)e^x dx = \sum_{i=1}^{n} a_i f(i)$$

for all polynomial functions $f \colon \mathbb{R} \to \mathbb{R}$ of degree less than n.

(6) Let A and B be matrices for which the product AB exists. Show that B is a kernel matrix of A if and only if we have $AB = 0$ and $\operatorname{rk} B = \dim \ker A$.

(7) Let Z be any set. For any set X, we call the set $Z^X = \operatorname{Map}(X, Z)$ the Z-dual of X; if the set Z is clear from the context (as it will be in this exercise), we will denote this Z-dual Z^X by $X^\times$, and we will write $X^{\times\times}$ for $(X^\times)^\times$.
   (a) Verify that any map $f \colon X \to Y$ of sets induces a map $f^\dagger \colon Y^\times \to X^\times$ on the associated Z-duals by sending $g \in Y^\times$ to the composition $g \circ f \in X^\times$.
   (b) Verify that for any set X there is a canonical map $\beta_X \colon X \to X^{\times\times}$ that sends $x \in X$ to the evaluation map $\operatorname{ev}_x \colon X^\times \to Z$, which sends $f \in X^\times$ to $f(x)$.
   (c) Show that for every set X we have $\beta_X^\dagger \circ \beta_{X^\times} = \operatorname{id}_{X^\times}$.
   (d) Show that for every vector space V we have $\alpha_V^\top \circ \alpha_{V^*} = \operatorname{id}_{V^*}$; if V is finite-dimensional, then $\alpha_V^\top = \alpha_{V^*}^{-1}$.
   (e) In the proof of Proposition 6.18, it was shown that the composition of two maps is the identity. Use the previous part of this exercise to show directly that the composition in the opposite order is the identity as well.

(8) Suppose we have a long exact sequence

$$0 \longrightarrow V_1 \longrightarrow V_2 \longrightarrow \cdots \longrightarrow V_n \longrightarrow 0$$

of vector spaces. Show that we have $\sum_{i=1}^{n}(-1)^i \dim V_i = 0$.
[Hint: first do the case $n = 3$].

(9) Suppose $f\colon U \to V$ and $g\colon V \to W$ are linear maps such that

$$U \xrightarrow{f} V \xrightarrow{g} W \longrightarrow 0$$

is an exact sequence. Suppose that $F_U\colon U \to U$ and $F_V\colon V \to V$ are endomorphisms such that $F_V \circ f = f \circ F_U$. Show that there exists an endomorphism $F_W\colon W \to W$ such that $F_W \circ g = g \circ F_V$. In other words, show that there exists an endomorphism F_W of W such that the following diagram commutes.

$$\begin{array}{ccccccc} U & \xrightarrow{f} & V & \xrightarrow{g} & W & \longrightarrow & 0 \\ \downarrow{\scriptstyle F_U} & & \downarrow{\scriptstyle F_V} & & \downarrow{\scriptstyle F_W} & & \\ U & \xrightarrow{f} & V & \xrightarrow{g} & W & \longrightarrow & 0 \end{array}$$

(10) Suppose V is a vector space and W is a subspace. Let $f\colon V \to V$ be a linear map so that $f(w) = w$ for $w \in W$. Show that $f^\top(\varphi) - \varphi \in W^\circ$ for all $\varphi \in V^*$.

Conversely, if you assume that $f^\top(\varphi) - \varphi \in W^\circ$ for all $\varphi \in V^*$, can you show that $f(w) = w$ for all $w \in W$?

(11) Let $f\colon U \to V$ and $g\colon V \to W$ be two linear maps of finite-dimensional vector spaces. Suppose that the dual sequence

$$W^* \xrightarrow{g^\top} V^* \xrightarrow{f^\top} U^*$$

is exact. Show that the sequence

$$U \xrightarrow{f} V \xrightarrow{g} W$$

is exact.

(12) * Let V be a finite-dimensional vector space and let $U \subset V$ and $W \subset V^*$ be subspaces. We identify V and V^{**} via α_V (so $W^\circ \subset V$). Show that

$$\dim(U^\circ \cap W) + \dim U = \dim(U \cap W^\circ) + \dim W\,.$$

(13) Let $\phi_1, \ldots, \phi_n \in (\mathbb{R}^n)^*$. Prove that the solution set C of the linear inequalities $\phi_1(x) \geq 0, \ldots, \phi_n(x) \geq 0$ has the following properties:

(a) $\alpha, \beta \in C \implies \alpha + \beta \in C$.

(b) $\alpha \in C,\ t \in \mathbb{R}_{\geq 0} \implies t\alpha \in C$.

(c) If $\phi_1, \ldots, \phi_n$ form a basis for $(\mathbb{R}^n)^*$, then

$$C = \left\{t_1\alpha_1 + \ldots + t_n\alpha_n : t_i \in \mathbb{R}_{\geq 0}, \forall i \in \{1, \ldots, n\}\right\},$$

where $\alpha_1, \ldots, \alpha_n$ is the basis for $\mathbb{R}^n$ dual to $\phi_1, \ldots, \phi_n$.

7. Norms on Real Vector Spaces

The following has some relevance for Analysis.

7.1. Definition. Let V be a real vector space. A *norm* on V is a map $V \to \mathbb{R}$, usually written $x \mapsto \|x\|$, such that

(i) $\|x\| \geq 0$ for all $x \in V$, and $\|x\| = 0$ if and only if $x = 0$;
(ii) $\|\lambda x\| = |\lambda|\|x\|$ for all $\lambda \in \mathbb{R}$, $x \in V$;
(iii) $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$ (triangle inequality).

7.2. Examples. If $V = \mathbb{R}^n$, then we have the following standard examples of norms.

(1) The maximum norm:
$$\|(x_1, \dots, x_n)\|_\infty = \max\{|x_1|, \dots, |x_n|\}\,.$$
(2) The Euclidean norm (see Section 9 below):
$$\|(x_1, \dots, x_n)\|_2 = \sqrt{x_1^2 + \dots + x_n^2}\,.$$
(3) The sum norm (or 1-norm):
$$\|(x_1, \dots, x_n)\|_1 = |x_1| + \dots + |x_n|\,.$$

7.3. Remark. A norm on a real vector space V induces a metric: we set
$$d(x, y) = \|x - y\|\,,$$
then the axioms of a metric (positivity, symmetry, triangle inequality) follow from the properties of a norm.

Recall that the usual Euclidean topology on $\mathbb{R}^n$ is induced by the Euclidean metric given by $d(x, y) = \|x - y\|_2$ for all $x, y \in \mathbb{R}^n$. With respect to this topology, we have the following result.

7.4. Lemma. *Every norm on $\mathbb{R}^n$ is continuous (as a map from $\mathbb{R}^n$ to $\mathbb{R}$).*

Proof. Note that the maximum norm on $\mathbb{R}^n$ is bounded from above by the Euclidean norm:
$$\max\{|x_j| : j \in \{1, \dots, n\}\} \leq \sqrt{x_1^2 + \dots + x_n^2}\,.$$
Let $\|\cdot\|$ be a norm, and set $C = \sum_{j=1}^n \|e_j\|$, where $e_1, \dots, e_n$ is the canonical basis for $\mathbb{R}^n$. Then for $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ we have
$$\begin{aligned}\|x\| &= \|(x_1, \dots, x_n)\| = \|x_1 e_1 + \dots + x_n e_n\| \leq \|x_1 e_1\| + \dots + \|x_n e_n\| \\ &= |x_1|\|e_1\| + \dots + |x_n|\|e_n\| \leq \max\{|x_1|, \dots, |x_n|\} \cdot C \leq \|x\|_2 \cdot C\,.\end{aligned}$$
From the triangle inequality, we then get
$$\big|\|x\| - \|y\|\big| \leq \|x - y\| \leq C \cdot \|x - y\|_2\,.$$
So for any $\varepsilon > 0$, if $\|x - y\|_2 < \varepsilon / C$, then $\big|\|x\| - \|y\|\big| < \varepsilon$. □

7.5. Definition. Let V be a real vector space, $x \mapsto \|x\|_1$ and $x \mapsto \|x\|_2$ two norms on V (any norms, not necessarily those of Example 7.2!). The two norms are said to be *equivalent*, if there are $C_1, C_2 > 0$ such that
$$C_1\|x\|_1 \leq \|x\|_2 \leq C_2\|x\|_1 \qquad \text{for all } x \in V.$$

7.6. Theorem. *On a finite-dimensional real vector space, all norms are equivalent.*

Proof. Without loss of generality, we can assume that our space is $\mathbb{R}^n$, and we can assume that one of the norms is the Euclidean norm $\|\cdot\|_2$ defined above. Let $S \subset \mathbb{R}^n$ be the unit sphere, i.e., $S = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$. We know from Analysis that S is compact (it is closed as the zero set of the continuous function $x \mapsto x_1^2 + \cdots + x_n^2 - 1$ and bounded). Let $\|\cdot\|$ be another norm on $\mathbb{R}^n$. Then $x \mapsto \|x\|$ is continuous by Lemma 7.4, hence it attains a maximum C_2 and a minimum C_1 on S. Then $C_2 \geq C_1 > 0$ (since $0 \notin S$). Now let $0 \neq x \in \mathbb{R}^n$, and let $e = \|x\|_2^{-1}x$; then $\|e\|_2 = 1$, so $e \in S$. This implies that $C_1 \leq \|e\| \leq C_2$, and therefore

$$C_1\|x\|_2 \leq \|x\|_2 \cdot \|e\| \leq C_2\|x\|_2 \,.$$

From $\|x\|_2 \cdot \|e\| = \big\|\|x\|_2 e\big\| = \|x\|$ we conclude $C_1\|x\|_2 \leq \|x\| \leq C_2\|x\|_2$. So every norm is equivalent to $\|\cdot\|_2$, which implies the claim, since equivalence of norms is an equivalence relation. □

7.7. Examples. If V is infinite-dimensional, then the statement of the theorem is no longer true. As a simple example, consider the space of finite sequences $(a_n)_{n\geq 0}$ (such that $a_n = 0$ for n sufficiently large). Then we can define norms $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_\infty$ as in Examples 7.2, but they are pairwise inequivalent now — consider the sequences $s_n = (1, \dots, 1, 0, 0, \dots)$ with n ones, then $\|s_n\|_1 = n$, $\|s_n\|_2 = \sqrt{n}$ and $\|s_n\|_\infty = 1$.

Here is a perhaps more natural example. Let V be the vector space $\mathcal{C}([0,1])$ of real-valued continuous functions on the unit interval. We can define norms

$$\|f\|_1 = \int_0^1 |f(x)|\,dx\,, \quad \|f\|_2 = \sqrt{\int_0^1 f(x)^2\,dx}\,, \quad \|f\|_\infty = \max\{|f(x)| : x \in [0,1]\}$$

in a similar way as in Examples 7.2, and again they are pairwise inequivalent. Taking $f(x) = x^n$, we have

$$\|f\|_1 = \frac{1}{n+1}\,, \quad \|f\|_2 = \frac{1}{\sqrt{2n+1}}\,, \quad \|f\|_\infty = 1\,.$$

Exercises.

Let V and W be normed vector spaces over $\mathbb{R}$. For a linear map $f\colon V \to W$ set

$$\|f\| = \sup_{x\in V,\ \|x\|=1} \|f(x)\|\,.$$

(1) Consider $V = \mathbb{R}^n$ with the standard inner product and the norm $\|\cdot\|_2$. Suppose that $f\colon V \to V$ is a diagonalizable map whose eigenspaces are orthogonal (i.e., V has an orthogonal basis consisting of eigenvectors of f). Show that $\|f\|$ as defined above is equal to the largest absolute value of an eigenvalue of f.

(2) (a) Show that $B(V,W) = \{f \in \mathrm{Hom}(V,W)\colon \|f\| < \infty\}$ is a subspace of $\mathrm{Hom}(V,W)$, and that $\|\cdot\|$ is a norm on $B(V,W)$.
   (b) Show that $B(V,W) = \mathrm{Hom}(V,W)$ if V is finite-dimensional.
   (c) Taking $V = W$ above, we obtain a norm on $B(V,V)$. Show that $\|f \circ g\| \leq \|f\| \cdot \|g\|$ for all $f, g \in B(V,V)$.

(3) Consider the rotation map $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ which rotates the plane by 45 degrees. For any norm on $\mathbb{R}^2$ the previous exercise defines a norm $\|f\|$ of f. Show that $\|f\| = 1$ when we take the standard Euclidean norm $\|\cdot\|_2$ on $\mathbb{R}^2$. What is $\|f\|$ when we take the maximum norm $\|\cdot\|_\infty$ on $\mathbb{R}^2$?

(4) Consider the vector space V of polynomial functions $[0,1] \to \mathbb{R}$ with the sup-norm: $\|f\| = \sup_{0\le x\le 1} |f(x)|$. Consider the functional $\phi \in V^*$ defined by $\phi(f) = f'(0)$. Show that $\phi \notin B(V, \mathbb{R})$. [Hint: consider the polynomials $(1-x)^n$ for $n = 1, 2, \ldots$.]

(5) What is the sine of the matrix $\begin{pmatrix} \pi & \pi \\ 0 & \pi \end{pmatrix}$?

8. Bilinear Forms

We have already seen multilinear maps when we were discussing the determinant in Linear Algebra I. Let us remind ourselves of the definition in the special case when we have two arguments.

8.1. Definition. Let V_1, V_2 and W be F-vector spaces. A map $\phi : V_1 \times V_2 \to W$ is *bilinear* if it is linear in both arguments, i.e.

$$\forall \lambda, \lambda' \in F, x, x' \in V_1, y \in V_2 : \phi(\lambda x + \lambda' x', y) = \lambda\phi(x, y) + \lambda'\phi(x', y) \quad \text{and}$$
$$\forall \lambda, \lambda' \in F, x \in V_1, y, y' \in V_2 : \phi(x, \lambda y + \lambda' y') = \lambda\phi(x, y) + \lambda'\phi(x, y') .$$

When $W = F$ is the field of scalars, ϕ is called a *bilinear form.*

If $V_1 = V_2 = V$ and $W = F$, then ϕ is a *bilinear form on* V. It is *symmetric* if $\phi(x, y) = \phi(y, x)$ for all $x, y \in V$, and *alternating* if $\phi(x, x) = 0$ for all $x \in V$. The latter property implies that ϕ is *skew-symmetric*, i.e. $\phi(x, y) = -\phi(y, x)$ for all $x, y \in V$. To see this, consider

$$0 = \phi(x+y, x+y) = \phi(x, x) + \phi(x, y) + \phi(y, x) + \phi(y, y) = \phi(x, y) + \phi(y, x) .$$

The converse holds if $\operatorname{char}(F) \neq 2$, since (taking $x = y$)

$$0 = \phi(x, x) + \phi(x, x) = 2\phi(x, x) .$$

We denote by $\operatorname{Bil}(V, W)$ the set of all bilinear forms $V \times W \to F$, and by $\operatorname{Bil}(V)$ the set of all bilinear forms on V. These sets are F-vector spaces in the usual way, by defining addition and scalar multiplication point-wise.

8.2. Examples. The standard 'dot product' on $\mathbb{R}^n$ is a symmetric bilinear form on $\mathbb{R}^n$.

The map that sends $\left(\binom{a}{b}, \binom{c}{d}\right) \in \mathbb{R}^2 \times \mathbb{R}^2$ to $\left|\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}\right| = ad - bc$ is an alternating bilinear form on $\mathbb{R}^2$.

The map $(A, B) \mapsto \operatorname{Tr}(A^\top B)$ is a symmetric bilinear form on $\operatorname{Mat}(m \times n, F)$.

If $K : [0,1]^2 \to \mathbb{R}$ is continuous, then the following defines a bilinear form on the space of continuous real-valued functions on $[0,1]$:

$$(f, g) \longmapsto \int_0^1\!\!\int_0^1 K(x, y) f(x) g(y)\, dx\, dy .$$

Evaluation defines a bilinear form on $V \times V^*$: $(v, \phi) \longmapsto \phi(v)$.

8.3. Definition. A bilinear form $\phi : V \times W \to F$ induces linear maps

$$\phi_L : V \longrightarrow W^*,\ v \mapsto \big(w \mapsto \phi(v,w)\big) \quad \text{and} \quad \phi_R : W \longrightarrow V^*,\ w \mapsto \big(v \mapsto \phi(v,w)\big)\,.$$

Indeed, by the definition of bilinear forms, the maps $w \mapsto \phi(v,w)$ (for any fixed $v \in V$) and $v \mapsto \phi(v,w)$ (for any fixed $w \in W$) are linear forms contained in W^* and V^*, respectively, so ϕ_L and ϕ_R are well-defined as maps. Then using the definition of bilinearity again, we see that ϕ_L and ϕ_R are themselves linear maps.

The subspace $\ker(\phi_L) \subset V$ is called the *left kernel* of ϕ; it is the set of all $v \in V$ such that $\phi(v,w) = 0$ for all $w \in W$. Similarly, the subspace $\ker(\phi_R) \subset W$ is called the *right kernel* of ϕ. The bilinear form ϕ is said to be *nondegenerate* if ϕ_L and ϕ_R are isomorphisms.

8.4. Remark. If $\phi : V \times W \to F$ is a nondegenerate bilinear form, then V and W have the same finite dimension (Exercise, cf. Remark 6.6).

8.5. Lemma. *Let $\phi : V \times W \to F$ be a bilinear form with V or W finite-dimensional. Then ϕ is nondegenerate if and only if both its left and right kernel are trivial.*

Proof. To prove this statement, first observe that the left and right kernels are certainly trivial when ϕ_L and ϕ_R are isomorphisms. For the converse statement, first suppose that W is finite-dimensional. Assume that the left and right kernels are trivial. Then ϕ_L is injective, and since W is finite-dimensional, we obtain $\dim V \le \dim W^* = \dim W$, so V is finite-dimensional as well. From ϕ_R being injective, we similarly get $\dim W \le \dim V$, so $\dim V = \dim W$ and ϕ_L and ϕ_R are isomorphisms. The case that V is finite-dimensional works analogously. □

8.6. Example. For the 'evaluation pairing' $\mathrm{ev}\colon V \times V^* \to F$, we find that the map $\mathrm{ev}_L\colon V \to V^{**}$ is α_V, and $\mathrm{ev}_R\colon V^* \to V^*$ is the identity. So this bilinar form ev is nondegenerate if and only if α_V is an isomorphism, which is the case if and only if V is finite-dimensional (see Remark 6.6).

8.7. Example. The standard scalar (dot) product ϕ on F^n given by $\phi(v,w) = \langle v, w\rangle$ is a nondegenerate symmetric bilinear form. In fact, here ϕ_L equals φ_n as defined in the paragraph above Lemma 6.13: it sends the standard basis vector e_j to the j-th coordinate map in $(F^n)^*$, so it maps a basis to a basis and is therefore an isomorphism.

8.8. Remarks.

(1) The bilinear form $\phi : V \times V \to F$ is symmetric if and only if $\phi_R = \phi_L$.

(2) Suppose V and W have the same finite dimension. If $\phi\colon V \times W \to F$ is a bilinear form, then ϕ is nondegenerate if and only if its left kernel is trivial (if and only if its right kernel is trivial).

Indeed, in this case, $\dim W^* = \dim V$, so if ϕ_L is injective, it is also surjective, hence an isomorphism. Proposition 6.12 gives that $\phi_L^\top$ is an isomorphism as well. By Theorem 6.8, the map α_W is also an isomorphism, so the identity $\phi_R = \phi_L^\top \circ \alpha_W$ (which we leave as an exercise for the reader) shows that ϕ_R is an isomorphism as well. If ϕ_R is injective, then we use the identity $\phi_L = \phi_R^\top \circ \alpha_V$ instead.

In fact, we can say a little bit more.

8.9. Proposition. *Let V and W be F-vector spaces. There is an isomorphism*

$$\beta_{V,W} : \mathrm{Bil}(V,W) \longrightarrow \mathrm{Hom}(V,W^*)\,, \quad \phi \longmapsto \phi_L$$

with inverse given by

$$f \longmapsto \big((v,w) \mapsto (f(v))(w)\big)\,.$$

PROOF. We leave the (by now standard) proof that the given maps are linear as an exercise. It remains to check that they are inverses of each other. Call the second map $\gamma_{V,W}$. So let $\phi : V \times W \to F$ be a bilinear form. Then $\gamma_{V,W}(\phi_L)$ sends (v,w) to $(\phi_L(v))(w) = \phi(v,w)$, so $\gamma_{V,W} \circ \beta_{V,W}$ is the identity. Conversely, let $f \in \mathrm{Hom}(V,W^*)$, and set $\phi = \gamma_{V,W}(f)$. Then for $v \in V$, the linear form $\phi_L(v)$ sends w to $(\phi_L(v))(w) = \phi(v,w) = (f(v))(w)$, so $\phi_L(v) = f(v)$ for all $v \in V$, hence $\phi_L = f$. This shows that $\beta_{V,W} \circ \gamma_{V,W}$ is also the identity map. □

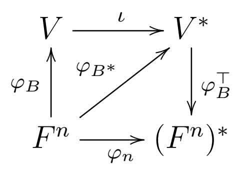
If $V = W$, we write $\beta_V : \mathrm{Bil}(V) \to \mathrm{Hom}(V,V^*)$ for this isomorphism.

8.10. Example. Let V now be finite-dimensional. We see that a nondegenerate bilinear form ϕ on V allows us to identify V with V^* via the isomorphism ϕ_L. Conversely, if we fix a basis $B = (v_1, \ldots, v_n)$, we also obtain an isomorphism $\iota : V \to V^*$ by sending v_j to v_j^*, where $B^* = (v_1^*, \ldots, v_n^*)$ is the dual basis for V^*. What is the bilinear form $\phi \colon V \times V \to F$ corresponding to this map? We have, for $v = \sum_{j=1}^n \lambda_j v_j$, $w = \sum_{j=1}^n \mu_j v_j$,

$$\begin{aligned}\phi(v,w) &= \big(\iota(v)\big)(w) = \Big(\iota\Big(\sum_{j=1}^n \lambda_j v_j\Big)\Big)\Big(\sum_{k=1}^n \mu_k v_k\Big) \\ &= \Big(\sum_{j=1}^n \lambda_j v_j^*\Big)\Big(\sum_{k=1}^n \mu_k v_k\Big) = \sum_{j,k=1}^n \lambda_i \mu_k\, v_j^*(v_k) = \sum_{j,k=1}^n \lambda_i \mu_k \delta_{jk} = \sum_{j=1}^n \lambda_j \mu_j\,.\end{aligned}$$

This is just the standard dot product if we identify V with F^n using the given basis; it is a symmetric bilinear form on V.

Alternatively, we note that $\varphi_{B^*} = \iota \circ \varphi_B$, so we obtain the following commutative diagram by Lemma 6.13.



Hence, indeed, if we identify V with F^n through φ_B (and likewise V^* with $(F^n)^*$ through $\varphi_B^\top$), then $\iota \colon V \to V^*$ corresponds to the map $\varphi_n \colon F^n \to (F^n)^*$, which sends $a \in F^n$ to the linear form $\langle \_, a\rangle$. As we have seen in Example 8.7, this map corresponds to the bilinear form that is the usual scalar (dot) product.

8.11. Proposition. *Let V, W be F-vector spaces, and let $\phi \colon V \times W \to F$ be a nondegenerate bilinear form. Then for every linear form $\psi \in W^*$ there is a unique $v \in V$ such that for every $w \in W$ we have $\psi(w) = \phi(v,w)$.*

PROOF. The condition that for every $w \in W$ we have $\psi(w) = \phi(v,w)$ is equivalent with the equality $\psi = \phi(v, \_)$, which means that $\psi = \phi_L(v)$. The claim now follows from the fact that $\phi_L \colon V \to W^*$ is an isomorphism. □

8.12. Example. Let V be the real vector space of polynomials of degree at most 2. Then

$$\phi\colon V \times V \to \mathbb{R}, \quad (p,q) \longmapsto \int_0^1 p(x)q(x)\,dx$$

is a bilinear form on V. It is nondegenerate since for $p \neq 0$, we have $\phi(p,p) > 0$. Evaluation at zero $p \mapsto p(0)$ defines a linear form on V, which by Proposition 8.11 must be representable in the form $p(0) = \phi(q,p)$ for some $q \in V$. To find q, we have to solve a linear system:

$$\begin{aligned}&\phi(a_0 + a_1x + a_2x^2, b_0 + b_1x + b_2x^2)\\ &\qquad = a_0b_0 + \tfrac{1}{2}(a_0b_1 + a_1b_0) + \tfrac{1}{3}(a_0b_2 + a_1b_1 + a_2b_0) + \tfrac{1}{4}(a_1b_2 + a_2b_1) + \tfrac{1}{5}a_2b_2\,,\end{aligned}$$

and we want to find a_0, a_1, a_2 such that this is always equal to b_0. This leads to

$$a_0 + \tfrac{1}{2}a_1 + \tfrac{1}{3}a_2 = 1\,, \quad \tfrac{1}{2}a_0 + \tfrac{1}{3}a_1 + \tfrac{1}{4}a_2 = 0\,, \quad \tfrac{1}{3}a_0 + \tfrac{1}{4}a_1 + \tfrac{1}{5}a_2 = 0$$

so $q(x) = 9 - 36x + 30x^2$, and

$$p(0) = \int_0^1 (9 - 36x + 30x^2)p(x)\,dx\,.$$

8.13. Representation by Matrices. Let $\phi : F^n \times F^m \to F$ be a bilinear form. Then we can represent ϕ by a matrix $A = (a_{ij}) \in \mathrm{Mat}(m \times n, F)$, with entries $a_{ij} = \phi(e_j, e_i)$. In terms of column vectors $x \in F^n$ and $y \in F^m$, we have

$$\phi(x,y) = y^\top Ax\,.$$

Similarly, if V and W are finite-dimensional F-vector spaces, and we fix bases $B = (v_1, \ldots, v_n)$ and $C = (w_1, \ldots, w_m)$ of V and W, respectively, then any bilinear form $\phi : V \times W \to F$ is given by a matrix relative to these bases, by identifying V and W with F^n and F^m in the usual way, that is, through the isomorphisms $\varphi_B\colon F^n \to V$ and $\varphi_C\colon F^m \to W$. If $A = (a_{ij})$ is the matrix as above, then $a_{ij} = \phi(v_j, w_i)$. If $v = x_1v_1 + \cdots + x_nv_n$ and $w = y_1w_1 + \cdots + y_mw_m$, then

$$\phi(v,w) = \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_jy_i\,.$$

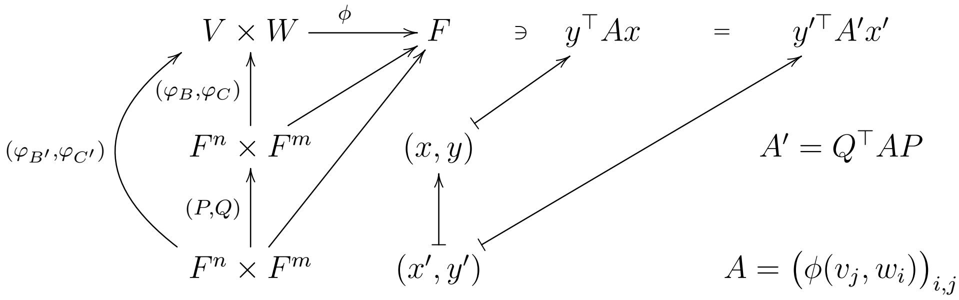
8.14. Proposition. *Let V and W be finite-dimensional F-vector spaces. Pick two bases $B = (v_1, \ldots, v_n)$ and $B' = (v'_1, \ldots, v'_n)$ of V and two bases $C = (w_1, \ldots, w_m)$ and $C' = (w'_1, \ldots, w'_m)$ of W. Let A be the matrix representing a bilinear form $\phi : V \times W \to F$ with respect to B and C, and let A' be the matrix representing ϕ with respect to B' and C'. Then for $P = [\mathrm{id}_V]^{B'}_B$ and $Q = [\mathrm{id}_W]^{C'}_C$ we have*

$$A' = Q^\top AP\,.$$

Proof. Let $x' \in F^n$ be the coefficients of $v \in V$ with respect to the new basis B'. Then $x = Px'$, where x represents v with respect to the old basis B. Similary for $y', y \in F^m$ representing $w \in W$ with respect to the two bases B' and B, respectively, we have $y = Qy'$. So

$$y'^\top A'x' = \phi(v,w) = y^\top Ax = y'^\top Q^\top APx'\,.$$

Given that this holds for all $x' \in F^n$ and all $y' \in F^m$, this implies the claim. $\square$



In particular, if V is an n-dimensional vector space V with basis B, and ϕ is a bilinear form on V, then ϕ is represented with respect to B by a square matrix $A \in \mathrm{Mat}(n, F)$. If we change the basis B to a basis B', then the new matrix will be $A' = P^\top AP$, with $P = [\mathrm{id}_V]_B^{B'} \in \mathrm{Mat}(n, F)$ invertible. Matrices A and A' for which there is an invertible matrix $P \in \mathrm{Mat}(n, F)$ such that $A' = P^\top AP$ are called *congruent.*

8.15. Remark. Let A be an $m \times n$ matrix over F. Then the associated bilinear form

$$F^n \times F^m \to F, \qquad (x, y) \mapsto y^\top Ax$$

can also be expressed using the standard dot products on F^m and F^n, both denoted by $\langle \_, \_ \rangle$, as we have

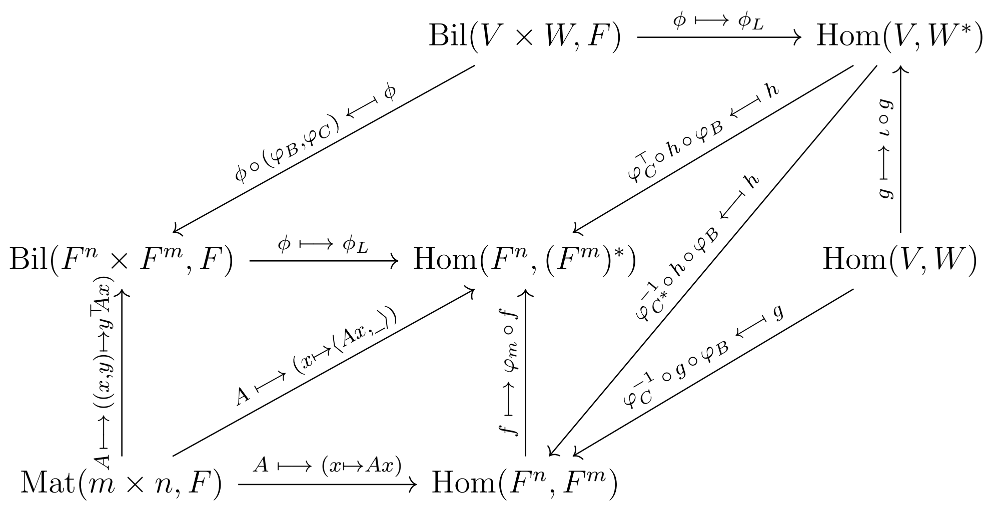
$$\langle y, Ax \rangle = y^\top Ax = (A^\top y)^\top x = \langle A^\top y, x \rangle.$$

8.16. Example. Let V be the real vector space of polynomials of degree less than n, and consider again the symmetric bilinear form

$$\phi(p, q) = \int_0^1 p(x)q(x)\, dx\,.$$

With respect to the standard basis $(1, x, \ldots, x^{n-1})$, it is represented by the "Hilbert matrix" $H_n = \left(\frac{1}{i+j-1}\right)_{1 \le i,j \le n}$.

For completeness, we summarize in one commutative diagram the ways to associate a matrix to linear maps and bilinear forms. Let V and W be finite-dimensional vector spaces, with bases B and C, respectively. Let C^* denote the dual basis for W^*. Also set $\iota = \varphi_{C^*} \circ \varphi_C^{-1} \colon W \to W^*$, which sends the i-th basis vector of C to the i-th basis vector of C^*. Recall that $\varphi_m = \varphi_C^\top \circ \varphi_{C^*} \colon F^m \to (F^m)^*$ sends $a \in F^m$ to $\langle a, \_ \rangle$. Then all maps in the following diagram are isomorphisms.



This diagram shows, for example, that if A is the matrix representing the bilinear form $\phi\colon V \times W \to F$ with respect to the bases B and C of V and W, respectively, then $A = [\phi_L]^B_{C^*}$ is also the matrix associated to the linear map $\phi_L\colon V \to W^*$ with respect to the bases B and C^*, since the map $\varphi_{C^*}^{-1} \circ \phi_L \circ \varphi_B$ is f_A.

8.17. Lemma. *Let $\phi\colon V \times W \to F$ be a bilinear form, and B and C bases of the finite-dimensional vector spaces V and W, respectively. Let A be the matrix that represents ϕ with respect to B and C. Then ϕ is nondegenerate if and only if A is invertible.*

PROOF. We have just seen that $A = [\phi_L]^B_{C^*}$, so the left kernel of ϕ corresponds to the kernel of A, which is trivial if and only if $\dim V = \operatorname{rk} A$. Similarly, the right kernel of ϕ is trivial if and only if $\dim W = \operatorname{rk} A$. The statement therefore follows from Lemma 8.5 and the fact that the equalities $\dim V = \dim W = \operatorname{rk} A$ are equivalent with A being invertible. □

8.18. Lemma. *Let ϕ be a bilinear form on the finite-dimensional vector space V, represented (with respect to some basis) by the matrix A. Then*

(1) *ϕ is symmetric if and only if $A^\top = A$;*
(2) *ϕ is skew-symmetric if and only if $A^\top + A = 0$;*
(3) *ϕ is alternating if and only if $A^\top + A = 0$ and all diagonal entries of A are zero.*

PROOF. Let $B = (v_1, \dots, v_n)$ be the basis for V. Since $a_{ij} = \phi(v_j, v_i)$, the implications "⇒" in the first three statements are clear. On the other hand, assume that $A^\top = \pm A$. Then

$$x^\top Ay = (x^\top Ay)^\top = y^\top A^\top x = \pm y^\top Ax\,,$$

which implies "⇐" in the first two statements. For the third statement, we compute $\phi(v, v)$ for $v = x_1 v_1 + \dots + x_n v_n$:

$$\phi(v,v) = \sum_{i,j=1}^n a_{ij} x_i x_j = \sum_{i=1}^n a_{ii} x_i^2 + \sum_{1 \le i < j \le n} (a_{ij} + a_{ji}) x_i x_j = 0\,,$$

since the assumption implies that both a_{ii} and $a_{ij} + a_{ji}$ vanish. □

8.19. Definition. Let $\phi\colon V \times W \to F$ be a bilinear form. For any subspace $U \subset W$ we set

$$U^\perp = \{v \in V : \phi(v,u) = 0 \text{ for all } u \in U\}\,.$$

For any subspace $U \subset V$ we set

$$U^\perp = \{w \in W : \phi(u,w) = 0 \text{ for all } u \in U\}\,.$$

In both cases we call $U^\perp$ the subspace *orthogonal* to U (with respect to ϕ).

8.20. Remark. Note that for a subspace $U \subset W$, the set $U^\perp$ is indeed a subspace, as it is the kernel of the composition of $\phi_L\colon V \to W^*$ with the restriction map $\mathrm{res}^W_U : W^* \to U^*$ that sends $\psi \in W^*$ to the restriction $\psi|_U$. Alternatively, $U^\perp$ is in this case the left kernel of the restricted bilinear form $V \times U \to F$.
Similarly, for a subspace $U \subset V$, the subspace $U^\perp$ is the kernel of the composition of $\phi_R\colon W \to V^*$ with the restriction map $\mathrm{res}^V_U\colon V^* \to U^*$. Alternatively, $U^\perp$ is in this case the right kernel of the restricted bilinear form $U \times W \to F$. Moreover, as the kernel of res^V_U is the annihilator U°, we also find $U^\perp = \phi_R^{-1}(U^\circ)$.

8.21. Remark. For a general bilinear form ϕ on V and a subspace $U \subset V$, the notation $U^\perp$ is ambiguous, as the left kernel of the restriction $V \times U \to F$ and the right kernel of the restriction $U \times V \to F$ need not coincide. If ϕ is symmetric, then they do coincide, and the space $U^\perp$ is well defined.

8.22. Example. Let V be a vector space over F, and consider the bilinear form $\mathrm{ev}\colon V \times V^* \to F$ of Example 8.6. Let $U \subset V$ be a subspace. Then the orthogonal subspace $U^\perp$ with respect to ev consists of all $f \in V^*$ that satisfy $f(u) = \mathrm{ev}(u,f) = 0$ for all $u \in U$. This means that the subspace $U^\perp = U^\circ$ is the annihilator of U. Note that this is a special case of Remark 8.20, as we have $\mathrm{ev}_R = \mathrm{id}_{V^*}$ (see Example 8.6).

8.23. Lemma. *Let $\phi\colon V \times W \to F$ be a nondegenerate bilinear form. Let U be a subspace of either V or W. Then we have* $\dim U + \dim U^\perp = \dim V = \dim W$. *Moreover, we have $(U^\perp)^\perp = U$.*

Proof. From Remark 8.4 we recall that V and W are finite-dimensional, and $\dim V = \dim W$. First suppose $U \subset W$. By Lemma 6.19, the restriction map $\mathrm{res}^W_U : W^* \to U^*$ is surjective. So is the map $\phi_L\colon V \to W^*$, and therefore so is the composition $V \to U^*$. The kernel of this composition is $U^\perp$, so we obtain $\dim V = \dim U^\perp + \dim U^* = \dim U^\perp + \dim U$. The case $U \subset V$ follows similarly by considering the composition of ϕ_R with the restriction map res^V_U, thus proving the identity $\dim U + \dim U^\perp = \dim V$ in all cases. Applying this identity to $U^\perp$ as well, we find $\dim(U^\perp)^\perp = \dim U$. For all $u \in U$ and all $w \in U^\perp$, we have $\phi(u,w) = 0$, so there is an inclusion $U \subset (U^\perp)^\perp$ of subspaces of the same finite dimension. Hence, this inclusion is an equality. □

We leave it to the reader to find an example of a bilinear form ϕ on a finite-dimensional vector space V that is degenerate and for which there is a subspace $U \subset V$ with $(U^\perp)^\perp \neq U$.

As with endomorphisms, we can also split bilinear forms into direct sums in some cases.

8.24. Definition. If $V = U \oplus U'$, ϕ is a bilinear form on V, ψ and ψ' are bilinear forms on U and U', respectively, and for all $u_1, u_2 \in U$, $u'_1, u'_2 \in U'$, we have

$$\phi(u_1 + u'_1, u_2 + u'_2) = \psi(u_1, u_2) + \psi'(u'_1, u'_2)\,,$$

then ϕ is the *orthogonal direct sum* of ψ and ψ'.

Given $V = U \oplus U'$ and ϕ, this is the case if and only if $\phi(u, u') = 0$ and $\phi(u', u) = 0$ for all $u \in U$, $u' \in U'$ (and then $\psi = \phi|_{U\times U}$, $\psi' = \phi|_{U'\times U'}$).

This can be generalised to an arbitrary number of summands.

If V is finite-dimensional and we represent ϕ by a matrix with respect to a basis that is compatible with the splitting, then the matrix will be block diagonal.

8.25. Proposition. *Let ϕ be a symmetric bilinear form on V, and let $U \subset V$ be a linear subspace such that $\phi|_{U\times U}$ is nondegenerate. Then $V = U \oplus U^\perp$, and ϕ splits accordingly as an orthogonal direct sum.*

When the restriction of ϕ to $U \times U$ is nondegenerate, we call $U^\perp$ the *orthogonal complement* of U.

Proof. We have to check a number of things. First, $U \cap U^\perp = \{0\}$ since $v \in U \cap U^\perp$ implies $\phi(v, u) = 0$ for all $u \in U$, but ϕ is nondegenerate on U, so v must be zero. Second, $U + U^\perp = V$: let $v \in V$, then $U \ni u \mapsto \phi(v, u)$ is a linear form on U, and since ϕ is nondegenerate on U, by Proposition 8.11 there must be $u' \in U$ such that $\phi(v, u) = \phi(u', u)$ for all $u \in U$. This means that $\phi(v - u', u) = 0$ for all $u \in U$, hence $v - u' \in U^\perp$, and we see that $v = u' + (v - u') \in U + U^\perp$ as desired. So we have $V = U \oplus U^\perp$. The last statement is clear, since by definition, ϕ is zero on $U \times U^\perp$. □

Theorem 8.27 gives the first and quite general classification result for symmetric bilinear forms on finite-dimensional vector spaces: they can always be diagonalized. We first state a useful lemma.

8.26. Lemma. *Assume that* $\mathrm{char}(F) \neq 2$, *let V be an F-vector space and ϕ a symmetric bilinear form on V. If $\phi \neq 0$, then there is $v \in V$ such that $\phi(v, v) \neq 0$.*

Proof. If $\phi \neq 0$, then there are $v, w \in V$ such that $\phi(v, w) \neq 0$. Note that we have

$$0 \neq 2\phi(v, w) = \phi(v, w) + \phi(w, v) = \phi(v + w, v + w) - \phi(v, v) - \phi(w, w)\,,$$

so at least one of $\phi(v, v)$, $\phi(w, w)$ and $\phi(v + w, v + w)$ must be nonzero. □

8.27. Theorem. *Assume that* $\mathrm{char}(F) \neq 2$, *let V be a finite-dimensional F-vector space and ϕ a symmetric bilinear form on V. Then there is a basis $(v_1, \ldots, v_n)$ of V such that ϕ is represented by a diagonal matrix with respect to this basis.*

Equivalently, every symmetric matrix $A \in \mathrm{Mat}(n, F)$ is congruent to a diagonal matrix.

Proof. If $\phi = 0$, there is nothing to prove. Otherwise, we proceed by induction on the dimension n. Since $\phi \neq 0$, by Lemma 8.26, there is $v_1 \in V$ such that $\phi(v_1, v_1) \neq 0$ (in particular, $n \geq 1$). Let $U = L(v_1)$, then ϕ is nondegenerate on U. By Prop. 8.25, we have an orthogonal splitting $V = L(v_1) \oplus U^\perp$. By induction ($\dim U^\perp = n - 1$), $U^\perp$ has a basis $(v_2, \ldots, v_n)$ such that $\phi|_{U^\perp \times U^\perp}$ is represented by a diagonal matrix. But then ϕ is also represented by a diagonal matrix with respect to the basis $(v_1, v_2, \ldots, v_n)$. $\square$

8.28. Remark. The entries of the diagonal matrix are not uniquely determined. For example, we can always scale the basis elements; this will multiply the entries by arbitrary nonzero squares in F. But this is not the only ambiguity. For example, we have

$$\begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}.$$

On the other hand, the *number* of nonzero entries is uniquely determined, since it is the rank of the matrix, which does not change when we multiply on the left or right by an invertible matrix.

8.29. Example. Let us see how we can find a diagonalizing basis in practice. Consider the bilinear form on F^3 (with $\text{char}(F) \neq 2$) given by the matrix

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

Following the proof above, we first have to find an element $v_1 \in F^3$ such that $v_1^\top A v_1 \neq 0$. Since the diagonal entries of A are zero, we cannot take one of the standard basis vectors. However, the proof of Lemma 8.26 tells us that (for example) $v_1 = (1, 1, 0)^\top$ will do. So we make a first change of basis to obtain

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

Now we have to find a basis for the orthogonal complement $L(v_1)^\perp$. This can be done by adding suitable multiples of v_1 to the other basis elements, in order to make the off-diagonal entries in the first row and column of the matrix zero. Here we have to add $-1/2$ times the first basis vector to the second, and add -1 times the first basis vector to the third. This gives

$$A'' = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} A' \begin{pmatrix} 1 & -\frac{1}{2} & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 \\ 0 & 0 & -2 \end{pmatrix}.$$

We are lucky: this matrix is already diagonal. (Otherwise, we would have to continue in the same way with the 2×2 matrix in the lower right.) The total change of basis is indicated by the product of the two change-of-basis matrices that we have used:

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\frac{1}{2} & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -\frac{1}{2} & -1 \\ 1 & \frac{1}{2} & -1 \\ 0 & 0 & 1 \end{pmatrix}$$

so the desired basis is $v_1 = (1, 1, 0)^\top$, $v_2 = (-\frac{1}{2}, \frac{1}{2}, 0)^\top$, $v_3 = (-1, -1, 1)^\top$.

8.30. Example. Consider the bilinear form ϕ on $\mathbb{R}^3$ given by $(x, y) \mapsto y^\top Ax$ with

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

First we switch the first two basis vectors to get a 1 in the top left. This yields

$$A' = P_1^\top A P_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \qquad \text{with} \quad P_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

From the new basis (e_2, e_1, e_3), in order to get generators for $e_2^\perp$, we subtract e_2 from the other two to get $(e_2, e_1 - e_2, e_3 - e_2)$. This corresponds to

$$A'' = P_2^\top A' P_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \qquad \text{with} \quad P_2 = \begin{pmatrix} 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The middle vector $e_1 - e_2$ is not orthogonal to itself, as the corresponding entry along the diagonal of A' is nonzero, so we keep it as second vector. In order to find generators for the orthogonal complement of the subspace spanned by e_2 and $e_1 - e_2$, we subtract this middle vector $e_1 - e_2$ from the last vector to obtain the basis $(e_2, e_1 - e_2, e_3 - e_1)$. This corresponds to

$$A''' = P_3^\top A'' P_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \text{with} \quad P_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}.$$

Setting

$$P = P_1 P_2 P_3 = \begin{pmatrix} 0 & 1 & -1 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

we find $P^\top AP = A'''$. Note that indeed the basis vectors $e_2, e_1 - e_2$, and $e_3 - e_1$, or better said, their coefficients with respect to the standard basis, are in the columns of P.

For algebraically closed fields like $\mathbb{C}$, we get a very nice result.

8.31. Theorem (Classification of Symmetric Bilinear Forms Over $\mathbb{C}$). *Let F be algebraically closed of characteristic different from 2, for example $F = \mathbb{C}$. Then every symmetric matrix $A \in \mathrm{Mat}(n, F)$ is congruent to a matrix*

$$\left(\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right),$$

and the rank $0 \leq r \leq n$ is uniquely determined.

Proof. By Theorem 8.27, A is congruent to a diagonal matrix, and we can assume that all zero diagonal entries come at the end. Let a_{jj} be a non-zero diagonal entry. Then we can scale the corresponding basis vector by $1/\sqrt{a_{jj}}$ (which exists in F, since F is algebraically closed); in the new matrix we get, this entry is then 1.

The uniqueness statement follows from the fact that $n - r$ is the dimension of the (left or right) kernel of the associated bilinear form. □

If $F = \mathbb{R}$, we have a similar statement. Let us first make a definition.

8.32. Definition. Let V be a real vector space, ϕ a symmetric bilinear form on V. Then ϕ is *positive definite* if

$$\phi(v,v) > 0 \qquad \text{for all } v \in V \setminus \{0\}.$$

8.33. Remark. A positive definite symmetric bilinear form ϕ on a finite-dimensional real vector space is nondegenerate: if $v \neq 0$, then $\phi(v,v) > 0$, so $\phi(v,v) \neq 0$. Hence v is not in the (left or right) kernel of ϕ. For example, this implies that the Hilbert matrix from Example 8.16 is invertible.

8.34. Theorem (Classification of Symmetric Bilinear Forms Over $\mathbb{R}$). *Every symmetric matrix $A \in \text{Mat}(n, \mathbb{R})$ is congruent to a unique matrix of the form*

$$\left(\begin{array}{c|c|c} I_r & 0 & 0 \\ \hline 0 & -I_s & 0 \\ \hline 0 & 0 & 0 \end{array}\right).$$

The number $r + s$ is the *rank* of A or of the corresponding bilinear form, the number $r - s$ is called the *signature* of A or of the corresponding bilinear form.

Proof. By Theorem 8.27, the matrix A is congruent to a diagonal matrix, so there is a basis $(v_1, \ldots, v_n)$ for $\mathbb{R}^n$ such that the bilinear form $\phi\colon (x,y) \mapsto y^\top A x$ is represented by a diagonal matrix D with respect to that basis. We can assume that the diagonal entries are ordered in such a way that we first have positive, then negative and then zero entries.

If d_{ii} is a non-zero diagonal entry of D, we scale the corresponding basis vector by $1/\sqrt{|d_{ii}|}$. Then the new diagonal matrix we get has positive entries 1 and negative entries -1, so it is of the form given in the statement.

The number $r + s$ is the rank of D, and hence of A, so it is uniquely determined. We claim that the number r is the maximal dimension of a subspace on which the bilinear form ϕ is positive definite. Indeed, if we let r' denote this maximal dimension, then we have $r \le r'$, as the bilinear form ϕ is positive definite on the subspace generated by $v_1, \ldots, v_r$. Moreover, if we have a subspace $U \subset \mathbb{R}^n$ on which ϕ is positive definite, then for the subspace $V \subset \mathbb{R}^n$ generated by $v_{r+1}, \ldots, v_n$ we have $U \cap V = \{0\}$, as any nonzero element $x \in U \cap V$ satisfies $\phi(x,x) > 0$ as well as $\phi(x) \le 0$; so we have $\dim U \le \dim \mathbb{R}^n - \dim V = n - (n - r) = r$, and we conclude $r' \le r$.

Therefore r and s only depend on the bilinear form, so they are uniquely determined. □

8.35. Example. Let V be again the real vector space of polynomials of degree ≤ 2. Consider the symmetric bilinear form on V given by

$$\phi(p,q) = \int_0^1 (2x-1)p(x)q(x)\,dx\,.$$

What are the rank and signature of ϕ?

We first find the matrix representing ϕ with respect to the standard basis $(1, x, x^2)$. Using $\int_0^1 (2x-1)x^n\,dx = \frac{2}{n+2} - \frac{1}{n+1} = \frac{n}{(n+1)(n+2)}$, we obtain

$$A = \begin{pmatrix} 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{3}{20} \\ \frac{1}{6} & \frac{3}{20} & \frac{2}{15} \end{pmatrix} = \frac{1}{60}\begin{pmatrix} 0 & 10 & 10 \\ 10 & 10 & 9 \\ 10 & 9 & 8 \end{pmatrix}.$$

The rank of this matrix is 2 (the kernel is generated by $10x^2 - 10x + 1$). We have that $\phi(x, x) = \frac{1}{6} > 0$ and $\phi(x-1, x-1) = \frac{1}{6} - 2\frac{1}{6} + 0 = -\frac{1}{6} < 0$, so r and s must both be at least 1. The only possibility is then $r = s = 1$, so the rank is 2 and the signature is 0. In fact, we have $\phi(x, x-1) = 0$, so

$$\sqrt{6}\,x\,, \quad \sqrt{6}(x-1)\,, \quad 10x^2 - 10x + 1$$

is a basis such that the matrix representing ϕ is

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

8.36. Theorem (Criterion for Positive Definiteness). *Let $A \in \mathrm{Mat}(n, \mathbb{R})$ be symmetric. Let A_j be the submatrix of A consisting of the upper left $j \times j$ block. Then (the bilinear form given by) A is positive definite if and only if $\det A_j > 0$ for all $1 \le j \le n$.*

Proof. First observe that if a matrix B represents a positive definite symmetric bilinear form, then $\det B > 0$: by Theorem 8.34, there is an invertible matrix P such that $P^\top BP$ is diagonal with entries 1, -1, or 0, and the bilinear form is positive definite if and only if all diagonal entries are 1, i.e., $P^\top BP = I$. But this implies $1 = \det(P^\top BP) = \det B(\det P)^2$, and since $(\det P)^2 > 0$, this implies $\det B > 0$.

Now if A is positive definite, then all A_j are positive definite, since they represent the restriction of the bilinear form to subspaces. So $\det A_j > 0$ for all j.

Conversely, assume that $\det A_j > 0$ for all j. We use induction on n. For $n = 1$ (or $n = 0$), the statement is clear. For $n \ge 2$, we apply the induction hypothesis to A_{n-1} and obtain that A_{n-1} is positive definite. Then there is an invertible matrix $P \in \mathrm{Mat}(n-1, \mathbb{R})$ such that

$$\left(\begin{array}{c|c} P^\top & 0 \\ \hline 0 & 1 \end{array}\right) A \left(\begin{array}{c|c} P & 0 \\ \hline 0 & 1 \end{array}\right) = \left(\begin{array}{c|c} I & b \\ \hline b^\top & \alpha \end{array}\right) =: B\,,$$

with some vector $b \in \mathbb{R}^{n-1}$ and $\alpha \in \mathbb{R}$. Setting

$$Q = \left(\begin{array}{c|c} I & -b \\ \hline 0 & 1 \end{array}\right),$$

we get

$$Q^\top BQ = \left(\begin{array}{c|c} I & 0 \\ \hline 0 & \beta \end{array}\right),$$

and so A is positive definite if and only if $\beta > 0$. But we have (note $\det Q = 1$)

$$\beta = \det(Q^\top BQ) = \det B = \det(P^\top)\det A \det P = (\det P)^2 \det A\,,$$

so $\beta > 0$, since $\det A = \det A_n > 0$, and A is positive definite. □

Exercises.

(1) Let V_1, V_2, U, W be vector spaces over a field F, and let $b\colon V_1 \times V_2 \to U$ be a bilinear map. Show that for each linear map $f : U \to W$ the composition $f \circ b$ is bilinear.
(2) Let V, W be vector spaces over a field F. If $b\colon V \times V \to W$ is both bilinear and linear, show that b is the zero map.

(3) Give an example of two vector spaces V, W over a field F and a bilinear map $b\colon V \times V \to W$ for which the image of b is not a subspace of W.

(4) Let V, W be two 2-dimensional subspaces of the standard $\mathbb{R}$-vector space $\mathbb{R}^3$. The restriction of the standard scalar product $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}$ to $\mathbb{R}^3 \times W$ is a bilinear map $b\colon \mathbb{R}^3 \times W \to \mathbb{R}$.
   (a) What is the left kernel of b? And the right kernel?
   (b) Let $b'\colon V \times W \to \mathbb{R}$ be the restriction of b to $V \times W$. Show that b' is degenerate if and only if the angle between V and W is 90°.

(5) Let $\phi\colon \mathbb{R}^4 \times \mathbb{R}^3 \to \mathbb{R}$ be the bilinear form given by $(x, y) \mapsto y^\top A x$ with

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \end{pmatrix}.$$

Let $f\colon \mathbb{R}^4 \to \mathbb{R}^4$ be the isomorphism given by

$$(x_1, x_2, x_3, x_4) \mapsto (x_1, x_1 + x_2, x_1 + x_2 + x_3, x_1 + x_2 + x_3 + x_4).$$

Let $g\colon \mathbb{R}^3 \to \mathbb{R}^3$ be the isomorphism given by

$$(x_1, x_2, x_3) \mapsto (x_1, x_1 + x_2, x_1 + x_2 + x_3).$$

Let $b\colon \mathbb{R}^4 \times \mathbb{R}^3 \to \mathbb{R}$ be the map given by $b(x, y) = \phi(f(x), g(y))$.
   (a) Determine the kernel of ϕ_L and ϕ_R.
   (b) Show that b is bilinear.
   (c) Give the matrix associated to b with respect to the standard bases for $\mathbb{R}^4$ and $\mathbb{R}^3$.

(6) Let V be a finite-dimensional vector space over F, and $\mathrm{ev}\colon V \times V^* \to F$ the bilinear form that sends (v, φ) to $\varphi(v)$. Let B be a basis for V, and B^* its dual basis for V^*. What is the matrix associated to ev with respect to the bases B and B^*?

(7) Let V be a vector space over $\mathbb{R}$, and let $b\colon V \times V \to \mathbb{R}$ be a symmetric bilinear map. Let the "quadratic form" associated to b be the map $q\colon V \to \mathbb{R}$ that sends $x \in V$ to $b(x, x)$. Show that b is uniquely determined by q.

(8) Let V be a vector space over $\mathbb{R}$, and let $b\colon V \times V \to \mathbb{R}$ be a bilinear map. Show that b can be uniquely written as a sum of a symmetric and a skew-symmetric bilinear form.

(9) Consider the real matrix

$$A = \begin{pmatrix} 1 & -2 & 4 \\ -2 & -2 & -2 \\ 4 & -2 & 1 \end{pmatrix}.$$

Find an invertible matrix P such that $P^\top A P$ is diagonal.

(10) Consider the real matrix

$$B = \begin{pmatrix} 1 & 2 & -4 \\ 2 & 3 & -3 \\ -4 & -3 & -9 \end{pmatrix}.$$

   (a) Show that $(-6, 5, 1)$ is an eigenvector of B.
   (b) Is B positive definite?
   (c) Find an invertible matrix Q such that $Q^\top B Q$ is diagonal.
   (d) What is the signature of B?

(11) Let V be the 3-dimensional vector space of polynomials of degree at most 2 with coefficients in $\mathbb{R}$. For $f, g \in V$ define the bilinear form $\phi\colon\ V \times V \to \mathbb{R}$ by

$$\phi(f,g) = \int_{-1}^{1} x f(x) g(x) dx.$$

(a) Is ϕ nondegenerate?
(b) Give a basis for V for which the matrix associated to ϕ is diagonal.
(c) Show that V has a 2-dimensional subspace U for which $U \subset U^{\perp}$.

(12) Let $e_1, \ldots, e_n$ be the standard basis for $V = \mathbb{R}^n$, and define a symmetric bilinear form ϕ on V by $\phi(e_i, e_j) = 2$ for all $i, j \in \{1, \ldots, n\}$. Give the signature of ϕ and a diagonalizing basis for ϕ.

(13) Suppose V is a vector space over $\mathbb{R}$ of finite dimension n with a nondegenerate bilinear form $\phi\colon\ V \times V \to \mathbb{R}$, and suppose that U is a subspace of V with $U \subset U^{\perp}$. Then show that the dimension of U is at most $n/2$.

(14) For $x \in \mathbb{R}$ consider the matrix

$$A_x = \begin{pmatrix} x & -1 \\ -1 & x \end{pmatrix}$$

(a) What is the signature of A_1 and A_{-1}?
(b) For which x is A_x positive definite?
(c) For which x is $\begin{pmatrix} x & -1 & 1 \\ -1 & x & 1 \\ 1 & 1 & 1 \end{pmatrix}$ positive definite?

(15) Let A be the matrix from exercise (14c). Find an invertible matrix P such that $P^{\top} A P$ is diagonal.

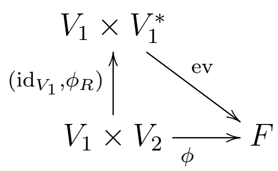
(16) Let V be a vector space over $\mathbb{R}$, let $b : V \times V \to \mathbb{R}$ be a skew-symmetric bilinear form, and let $x \in V$ be an element that is not in the left kernel of b.

(a) Show that there exist $y \in V$ such that $b(x, y) = 1$ and a linear subspace $U \subset V$ such that $V = \langle x, y\rangle \oplus U$ is an orthogonal direct sum with respect to b.
REMARK. The notation $\langle x, y\rangle$ denotes the subspace spanned by x and y, and of course has nothing to do with an inner product.
HINT. Take $U = \langle x, y\rangle^{\perp} = \{v \in V : b(x, v) = b(y, v) = 0\}$.
(b) Conclude that if $\dim V < \infty$, then then there exists a basis for V such that the matrix representing b with respect to this basis is a block diagonal matrix with blocks $B_1, \ldots, B_l$ of the form

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

and zero blocks $B_{l+1}, \ldots, B_k$.

(17) Let V_1, V_2 be vector spaces over F. Let $\phi\colon V_1 \times V_2 \to F$ be a bilinear form. Show that there is a commutative diagram



which shows that if ϕ is nondegenerate, and we use ϕ_R to identify V_2 with V_1^*, then ϕ corresponds to the evaluation pairing.

9. Inner Product Spaces

In many applications, we want to measure *distances* and *angles* in a real vector space. For this, we need an additional structure, a so-called *inner product*.

9.1. Definition. Let V be a real vector space. An *inner product* on V is a positive definite symmetric bilinear form on V. It is usually written in the form $(x, y) \mapsto \langle x, y\rangle \in \mathbb{R}$. Recall the defining properties:

(1) $\langle \lambda x + \lambda' x', y\rangle = \lambda\langle x, y\rangle + \lambda'\langle x', y\rangle$;
(2) $\langle y, x\rangle = \langle x, y\rangle$;
(3) $\langle x, x\rangle > 0$ for $x \neq 0$.

A real vector space together with an inner product on it is called a *real inner product space*.

Recall that an inner product on V induces an injective homomorphism $V \to V^*$, given by sending $x \in V$ to the linear form $y \mapsto \langle x, y\rangle$; this homomorphism is an isomorphism when V is finite-dimensional, in which case the inner product is nondegenerate.

Frequently, it is necessary to work with complex vector spaces. In order to have a similar structure there, we cannot use a bilinear form: if we want to have $\langle x, x\rangle$ to be real and positive, then we would get

$$\langle ix, ix\rangle = i^2\langle x, x\rangle = -\langle x, x\rangle ,$$

which would be negative. The solution to this problem is to consider *Hermitian* forms instead of symmetric bilinear forms. The difference is that they are *conjugate-linear* in the second argument.

9.2. Definition. Let V be a complex vector space. A *sesquilinear form* on V is a map $\phi : V \times V \to \mathbb{C}$ that is linear in the first and conjugate-linear in the second argument ("sesqui" means $1\frac{1}{2}$):

$$\phi(\lambda x + \lambda' x', y) = \lambda\phi(x, y) + \lambda'\phi(x', y) , \quad \phi(x, \lambda y + \lambda' y') = \bar{\lambda}\phi(x, y) + \bar{\lambda}'\phi(x, y') .$$

A *Hermitian form* on V is a sesquilinear form ϕ on V such that $\phi(y, x) = \overline{\phi(x, y)}$ for all $x, y \in V$. Note that this implies $\phi(x, x) \in \mathbb{R}$. The Hermitian form ϕ is *positive definite* if $\phi(x, x) > 0$ for all $x \in V \setminus \{0\}$. An *inner product* on the complex vector space V is a positive definite Hermitian form on V; in this context, the form is again usually written as $(x, y) \mapsto \langle x, y\rangle \in \mathbb{C}$.

Warning: this means that from now on, the notation $\langle x, y\rangle$ may refer to other pairings than the ordinary scalar (dot) product.

For an inner product on V, we have

(1) $\langle \lambda x + \lambda' x', y\rangle = \lambda\langle x, y\rangle + \lambda'\langle x', y\rangle$;
(2) $\langle y, x\rangle = \overline{\langle x, y\rangle}$;
(3) $\langle x, x\rangle > 0$ for $x \neq 0$.

A complex vector space together with an inner product on it is called a *complex inner product space* or *Hermitian inner product space*. A real or complex vector space with an inner product on it is an *inner product space*.

9.3. Definition. If V is a complex vector space, we denote by $\bar{V}$ the complex vector space with the same underlying set and addition as V, but with scalar multiplication modified by taking the complex conjugate: $\lambda \cdot v = \bar{\lambda} v$, where on the left, we have scalar multiplication on $\bar{V}$, and on the right, we have scalar multiplication on V. We call $\bar{V}$ the *complex conjugate* of V. If V is a real vector space, then we set $\bar{V} = V$.

9.4. Remark. Let V be a complex vector space. Note that any basis for V is also a basis for $\bar{V}$, so we have $\dim V = \dim \bar{V}$. Note that if $f\colon V \to W$ is a linear map, then it is also linear as a map from $\bar{V}$ to $\bar{W}$. If we denote this (same) map by $f'\colon \bar{V} \to \bar{W}$ to distinguish it from f, which has a different vector space structure on its domain and codomain, and B and C are finite bases for V and W, respectively, then we have $[f']^B_C = \overline{[f]^B_C}$.

We denote by $\bar{V}^* = (\bar{V})^*$ the dual of this complex conjugate space. If V is a complex inner product space, then the sesquilinear form $\phi\colon V \times V \to \mathbb{C}$ corresponds to a bilinear form $V \times \bar{V} \to \mathbb{C}$, and we again get homomorphisms

$$V \longrightarrow \bar{V}^*, \quad x \longmapsto (y \mapsto \langle x, y\rangle) = \langle x, \_\rangle$$

and

$$\bar{V} \longrightarrow V^*, \quad y \longmapsto (x \mapsto \langle x, y\rangle) = \langle \_, y\rangle .$$

These maps are injective because we have $\langle x, x\rangle \neq 0$ for $x \neq 0$. When V is finite-dimensional, this implies that they are isomorphisms, that is, the bilinear form $V \times \bar{V} \to \mathbb{C}$ is nondegenerate.

9.5. Remark. Note that the dual $\bar{V}^*$ of $\bar{V}$ is not the same as $\overline{V^*}$, which is the dual of V with the modified scalar multiplication. In fact, the map $\bar{V}^* \to \overline{V^*}$ that sends $\phi \in \bar{V}^*$ to the function $\bar{\phi}$ that sends $x \in V$ to $\overline{\phi(x)}$ is a homomorphism.

9.6. Examples. We have seen some examples of real inner product spaces already: the space $\mathbb{R}^n$ together with the usual scalar (dot) product is the standard example of a finite-dimensional real inner product space. An example of a different nature, important in analysis, is the space of continuous real-valued functions on an interval $[a, b]$, with the inner product

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx .$$

For complex inner product spaces, the finite-dimensional standard example is $\mathbb{C}^n$ with the standard (Hermitian) inner product

$$\langle (z_1, \dots, z_n), (w_1, \dots, w_n)\rangle = z_1\bar{w}_1 + \dots + z_n\bar{w}_n ,$$

so $\langle z, w\rangle = z \cdot \bar{w}$ in terms of the usual scalar (dot) product. Note that

$$\langle z, z\rangle = |z_1|^2 + \dots + |z_n|^2 \geq 0 .$$

The complex version of the function space example is the space of complex-valued continuous functions on $[a, b]$, with inner product

$$\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}\,dx .$$

9.7. Definition. Let V be an inner product space.

(1) For $x \in V$, we set $\|x\| = \sqrt{\langle x, x\rangle} \geq 0$. The vector x is a *unit vector* if $\|x\| = 1$.
(2) We say that $x, y \in V$ are *orthogonal*, $x \perp y$, if $\langle x, y\rangle = 0$.
(3) A subset $S \subset V$ is *orthogonal* if $x \perp y$ for all $x, y \in S$ such that $x \neq y$. The set S is an *orthonormal* set if in addition, $\|x\| = 1$ for all $x \in S$.
(4) A sequence $(v_1, \ldots, v_k)$ of elements in V is *orthogonal* if $v_i \perp v_j$ for all $1 \leq i < j \leq k$. The sequence is *orthonormal* if in addition, $\|v_i\| = 1$ for all $1 \leq i \leq k$.
(5) An *orthonormal basis* or *ONB* of V is a basis of V that is orthonormal.
(6) For any set $S \subset V$, we define $S^\perp$ as

$$S^\perp = \{v \in V \ : \ v \perp s \text{ for all } s \in S\}.$$

Note that being perpendicular is symmetric, that is, we have $x \perp y$ if and only if $y \perp x$. Also note that, as mentioned before, the inner product corresponds to a bilinear form $V \times \bar{V} \to F$ where F is $\mathbb{R}$ or $\mathbb{C}$. If $U \subset V$ is a subspace, then the definition of $U^\perp$ above coincides with the one given in Definition 8.19 with respect to this bilinear form (where we use that V and $\bar{V}$ are the same on the level of sets, and we may choose to view U as a subset of either V or $\bar{V}$). In particular, if V is finite-dimensional, and the inner product is therefore nondegenerate, then we find from Lemma 8.23 that $(U^\perp)^\perp = U$ and $\dim U + \dim U^\perp = \dim V$.

9.8. Proposition. Let V be a finite-dimensional inner product space, and $U \subset V$ a subspace. Then we have $V = U \oplus U^\perp$.

Proof. As mentioned just before the proposition, we have $\dim U + \dim U^\perp = \dim V$. Since inner products are positive definite, we have $U \cap U^\perp = \{0\}$, so the dimension theorem for subspaces gives

$$\dim U + \dim U^\perp = \dim(U + U^\perp) + \dim(U \cap U^\perp) = \dim(U + U^\perp).$$

We conclude $\dim(U + U^\perp) = \dim V$, so $U + U^\perp = V$. Because the intersection $U \cap U^\perp$ is trivial, we get $V = U \oplus U^\perp$. □

If we have $V = U \oplus U^\perp$, so in particular when V is finite-dimensional, then we call $U^\perp$ the *orthogonal complement* of U. If V is a real inner product space, then this coincides with Proposition 8.25 and the sentence below it; if V is a complex inner product space, then we can not apply Proposition 8.25 directly, as a complex inner product is not a bilinear form.

9.9. Proposition. *Let V be an inner product space.*

(1) *For $x \in V$ and a scalar λ, we have $\|\lambda x\| = |\lambda| \cdot \|x\|$.*
(2) **(Cauchy-Schwarz inequality)** *For $x, y \in V$, we have*

$$|\langle x, y\rangle| \leq \|x\| \cdot \|y\|,$$

with equality if and only if x and y are linearly dependent.
(3) **(Triangle inequality)** *For $x, y \in V$, we have $\|x + y\| \leq \|x\| + \|y\|$.*

Note that these properties imply that $\|\cdot\|$ is a norm on V in the sense of Section 7. In particular,

$$d(x,y) = \|x-y\|$$

defines a metric on V; we call $d(x,y)$ the *distance* between x and y. If $V=\mathbb{R}^n$ with the standard inner product, then this is just the usual Euclidean distance.

PROOF.

(1) We have

$$\|\lambda x\| = \sqrt{\langle \lambda x, \lambda x\rangle} = \sqrt{\lambda\bar\lambda\langle x,x\rangle} = \sqrt{|\lambda|^2\langle x,x\rangle} = |\lambda|\sqrt{\langle x,x\rangle} = |\lambda|\,\|x\|\,.$$

(2) This is clear when $y=0$, so assume $y\neq 0$. Consider

$$z = x - \frac{\langle x,y\rangle}{\|y\|^2}y\,;$$

then $\langle z,y\rangle = 0$ (in fact z is the projection of x on $y^\perp$). We find that

$$0 \le \langle z,z\rangle = \langle z,x\rangle = \langle x,x\rangle - \frac{\langle x,y\rangle}{\|y\|^2}\langle y,x\rangle = \|x\|^2 - \frac{|\langle x,y\rangle|^2}{\|y\|^2}\,,$$

which implies the inequality. If $x=\lambda y$, we have equality by the first part of the proposition. Conversely, if we have equality, we must have $z=0$, hence $x=\lambda y$ (with $\lambda = \langle x,y\rangle/\|y\|^2$).

(3) We have

$$\begin{aligned}\|x+y\|^2 &= \langle x+y,x+y\rangle = \langle x,x\rangle + \langle x,y\rangle + \langle y,x\rangle + \langle y,y\rangle \\ &= \|x\|^2 + 2\,\mathrm{Re}\langle x,y\rangle + \|y\|^2 \le \|x\|^2 + 2|\langle x,y\rangle| + \|y\|^2 \\ &\le \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 = (\|x\|+\|y\|)^2\,,\end{aligned}$$

using the Cauchy-Schwarz inequality.

□

Next we show that given any basis for a finite-dimensional inner product space, we can modify it in order to obtain an orthonormal basis. In particular, every finite-dimensional inner product space has orthonormal bases.

9.10. Theorem (Gram-Schmidt Orthonormalization Process). *Let V be an inner product space. Let $x_1,\dots,x_k\in V$ be linearly independent, and define*

$$\begin{aligned}y_1 &= x_1 \\ y_2 &= x_2 - \frac{\langle x_2,y_1\rangle}{\langle y_1,y_1\rangle}y_1 \\ y_3 &= x_3 - \frac{\langle x_3,y_1\rangle}{\langle y_1,y_1\rangle}y_1 - \frac{\langle x_3,y_2\rangle}{\langle y_2,y_2\rangle}y_2 \\ &\vdots \\ y_k &= x_k - \frac{\langle x_k,y_1\rangle}{\langle y_1,y_1\rangle}y_1 - \dots - \frac{\langle x_k,y_{k-1}\rangle}{\langle y_{k-1},y_{k-1}\rangle}y_{k-1}\,.\end{aligned}$$

Finally, set $z_i = y_i/\|y_i\|$ for $i=1,\dots,k$. Then $(z_1,\dots,z_k)$ is an orthonormal basis for $L(x_1,\dots,x_k)$.

PROOF. We first prove by induction on k that $(y_1, \dots, y_k)$ is an orthogonal basis for $L(x_1, \dots, x_k)$. The case $k = 1$ (or $k = 0$) is clear — $x_1 \neq 0$, so it is a basis for $L(x_1)$.

If $k \geq 2$, we know by the induction hypothesis that $y_1, \dots, y_{k-1}$ is an orthogonal basis for $L(x_1, \dots, x_{k-1})$. In particular, $y_1, \dots, y_{k-1}$ are nonzero, so y_k is well defined. Since $y_1, \dots, y_{k-1}$ are pairwise orthogonal, that is, $\langle y_i, y_j \rangle = 0$ for $i \neq j$, we find for $1 \leq j \leq k-1$ that

$$\langle y_k, y_j \rangle = \langle x_k, y_j \rangle - \sum_{i=1}^{k-1} \frac{\langle x_k, y_i \rangle}{\langle y_i, y_i \rangle} \cdot \langle y_i, y_j \rangle = \langle x_k, y_j \rangle - \langle x_k, y_j \rangle = 0.$$

Hence, in fact $y_1, \dots, y_k$ are pairwise orthogonal. By construction, we have an inclusion $L(y_1, \dots, y_k) \subset L(x_1, \dots, x_k)$. As it is also clear that x_k can be expressed in $y_1, \dots, y_k$, the opposite inclusion also holds. In particular, this implies that $L(y_1, \dots, y_k)$ has dimension k, so $(y_1, \dots, y_k)$ is linearly independent and hence an orthogonal basis for $L(x_1, \dots, x_k)$.

Since $y_1, \dots, y_k$ are linearly independent, they are nonzero, so we may indeed normalise and set $z_i = y_i/\|y_i\|$ for $i = 1, \dots, k$. After normalising, we have $\|z_i\| = 1$ and $\langle z_i, z_j \rangle = 0$ for $i \neq j$. Clearly, we have $L(z_1, \dots, z_k) = L(y_1, \dots, y_k) = L(x_1, \dots, x_k)$, so $(z_1, \dots, z_k)$ is an orthonormal basis for $L(x_1, \dots, x_k)$. □

9.11. Corollary. *Every finite-dimensional inner product space has an ONB.*

PROOF. Apply Theorem 9.10 to a basis for the space. □

9.12. Proposition. *Let V be an inner product space.*

1. *If $(v_1, v_2, \dots, v_k)$ is an orthogonal sequence of nonzero elements in V, then $v_1, \dots, v_k$ are linearly independent.*
2. *If $S \subset V$ is an orthogonal set of nonzero vectors, then S is linearly independent.*

PROOF.

1. Let $(v_1, v_2, \dots, v_k)$ be an orthogonal sequence of nonzero elements in V, and assume we have a linear combination
$$\sum_{i=1}^{k} \lambda_i v_i = 0\,.$$
Now we take the inner product with v_j for a fixed j:
$$0 = \Big\langle \sum_{i=1}^{k} \lambda_i v_i, v_j \Big\rangle = \sum_{i=1}^{k} \lambda_i \langle v_i, v_j \rangle = \lambda_j \langle v_j, v_j \rangle\,.$$
Since $v_j \neq 0$, we have $\langle v_j, v_j \rangle \neq 0$, therefore we must have $\lambda_j = 0$. Since this is true for every index $1 \leq j \leq k$, the linear combination is trivial.
2. By part (1), every finite subset of S is linearly independent, which makes the set S linearly independent by definition.

□

9.13. Proposition. *Suppose V is an n-dimensional inner product space. Then for every orthonormal sequence $(e_1, \dots, e_k)$ of elements in V, there are elements $e_{k+1}, \dots, e_n \in V$ such that $(e_1, \dots, e_n)$ is an ONB of V.*

Proof. By Proposition 9.12, the elements $e_1, \dots, e_k$ are linearly independent. Extend $e_1, \dots, e_k$ to a basis for V in some way and apply Theorem 9.10 to this basis. This will not change the first k basis elements, since they are already orthonormal. □

Orthonormal bases are rather nice, as we will see.

9.14. Theorem (Bessel's Inequality). *Let V be an inner product space, and let $(e_1, \dots, e_n)$ be an orthonormal sequence of elements in V. Then for all $x \in V$, we have the inequality*

$$\sum_{j=1}^{n} \big|\langle x, e_j\rangle\big|^2 \le \|x\|^2 .$$

Let $U = L(e_1, \dots, e_n)$ be the subspace spanned by $e_1, \dots, e_n$. Then for $x \in V$, the following statements are equivalent:

(1) *$x \in U$;*

(2) *$\sum_{j=1}^{n} \big|\langle x, e_j\rangle\big|^2 = \|x\|^2$;*

(3) *$x = \sum_{j=1}^{n} \langle x, e_j\rangle e_j$;*

(4) *for all $y \in V$, $\langle x, y\rangle = \sum_{j=1}^{n} \langle x, e_j\rangle \langle e_j, y\rangle$.*

In particular, statements (2) to (4) hold for all $x \in V$ when $(e_1, \dots, e_n)$ is an ONB of V.

When $(e_1, \dots, e_n)$ is an ONB, then (4) (and also (2)) is called *Parseval's Identity.* The relation in (3) is sometimes called the *Fourier expansion* of x relative to the given ONB.

Proof. Let $z = x - \sum_{j=1}^{n} \langle x, e_j\rangle e_j$. Then for any $1 \le k \le n$ we have

$$\langle z, e_k\rangle = \langle x, e_k\rangle - \sum_{j=1}^{n} \langle x, e_j\rangle \cdot \langle e_j, e_k\rangle = \langle x, e_k\rangle - \langle x, e_k\rangle = 0.$$

This implies $\langle z, z\rangle = \langle z, x\rangle$, so we find

$$0 \le \langle z, z\rangle = \langle z, x\rangle = \langle x, x\rangle - \sum_{j=1}^{n} \langle x, e_j\rangle \cdot \langle e_j, x\rangle = \|x\|^2 - \sum_{j=1}^{n} \big|\langle x, e_j\rangle\big|^2 .$$

This implies the inequality and also gives the implication (2) ⇒ (3), as equality in (2) implies $\langle z, z\rangle = 0$, so $z = 0$. The implication (3) ⇒ (4) is a simple calculation, and (4) ⇒ (2) follows by taking $y = x$. (3) ⇒ (1) is trivial. Finally, to show (1) ⇒ (3), let

$$x = \sum_{j=1}^{n} \lambda_j e_j .$$

Then

$$\langle x, e_k \rangle = \sum_{j=1}^{n} \lambda_j \langle e_j, e_k \rangle = \lambda_k \,,$$

which gives the relation in (3). □

Next, we want to discuss linear maps on inner product spaces.

9.15. Theorem. *Let V and W be two inner product spaces over the same field ($\mathbb{R}$ or $\mathbb{C}$), and let $f : V \to W$ be linear. Then there is at most one map $f^* : W \to V$ such that*

$$\langle f(v), w \rangle = \langle v, f^*(w) \rangle$$

for all $v \in V$, $w \in W$. If such a map exists, then it is linear. Moreover, if V is finite-dimensional, then such a map does exist.

Proof. Recall that we have an injective linear map $\bar{V} \to V^*$ that sends $x \in \bar{V}$ to $\langle \_, x \rangle$, and where we use $\bar{V} = V$ if the base field is $\mathbb{R}$. This injective map is an isomorphism if V is finite-dimensional. For $w \in W$ fixed, the map $V \ni v \mapsto \langle f(v), w \rangle$ is a linear form on V, so there is at most one element $x \in \bar{V}$ such that $\langle f(v), w \rangle = \langle v, x \rangle$ for all $v \in V$; if such an element exists, which is the case if V is finite-dimensional, then we set $f^*(w) = x$. Assume that $f^*(w)$ is defined for all $w \in W$. Now consider $w + w'$ for $w, w' \in W$. We find that $f^*(w + w')$ and $f^*(w) + f^*(w')$ both satisfy the relation, so by uniqueness, f^* is additive. Similary, considering λw for $w \in W$ and $\lambda \in \mathbb{R}$ or $\mathbb{C}$, we see that $f^*(\lambda w)$ and $\lambda f^*(w)$ must agree. Hence f^* is actually a linear map. □

Alternative proof. Let F be the field over which V and W are inner product spaces. Let $\phi\colon V \times \bar{V} \to F$ and $\psi\colon W \times \bar{W} \to F$ be the bilinear forms that correspond to the inner products on V and W, respectively. Then we have $\langle f(v), w \rangle = \langle v, f^*(w) \rangle$ for all $v \in V$ and all $w \in W$ if and only if we have $\phi_R \circ f^* = f^\top \circ \psi_R$, that is, the diagram

(6)
$$\begin{array}{ccc} W^* & \xrightarrow{f^\top} & V^* \\ {\scriptstyle \psi_R}\uparrow & & \uparrow{\scriptstyle \phi_R} \\ \bar{W} & \xrightarrow[f^*]{} & \bar{V} \end{array}$$

commutes. Note that ϕ_R is injective, so there is at most one such map f^*. Also because of injectivity, and the fact that the composition $f^\top \circ \psi_R$ is linear, the map f^* is linear if it exists. If V is finite-dimensional, then ϕ_R is an isomorphism, so there is such a map, as we can take $f^* = \phi_R^{-1} \circ f^\top \circ \psi_R$. □

9.16. Definition. Let V and W be inner product spaces over the same field.

(1) Let $f : V \to W$ be linear. If f^* exists with the property given in Theorem 9.15 (which is always the case when $\dim V < \infty$), then f^* is called the *adjoint* of f.
(2) If $f : V \to V$ has an adjoint f^*, and $f = f^*$, then f is *self-adjoint*.
(3) If $f : V \to V$ has an adjoint f^* and $f \circ f^* = f^* \circ f$, then f is *normal*.
(4) A linear map $f : V \to W$ is an *isometry* if it is an isomorphism and $\langle f(v), f(v') \rangle = \langle v, v' \rangle$ for all $v, v' \in V$.

9.17. Remark. Some books use an alternative definition for isometry. Indeed, Exercise 23 shows that an isomorphism of inner product spaces is an isometry if and only if it preserves lengths. Exercise 25 shows that we do not even need to require the map to be linear, if we assume it preserves all distances. Exercises 27 and 28 show that it also suffices to require angles to be preserved.

9.18. Examples. If $f\colon V \to V$ is self-adjoint or an isometry, then f is normal. For the second claim, note that every isometry $f\colon V \to W$, also between infinite-dimensional spaces, has an adjoint $f^* = f^{-1}$. (In fact, the converse is true as well: if an isomorphism $f\colon V \to W$ has an adjoint $f^* = f^{-1}$, then f is an isometry. The proof of Proposition 9.22 below includes a proof of this statement that does not rely on finite-dimensionality.)

9.19. Remark. While the property of the adjoint given in Theorem 9.15 may seem asymmetric, we also have

$$\langle w, f(v)\rangle = \overline{\langle f(v), w\rangle} = \overline{\langle v, f^*(w)\rangle} = \langle f^*(w), v\rangle$$

for all $v \in V$ and all $w \in W$, which is equivalent with $\phi_L \circ f^* = f^\top \circ \psi_L$.

9.20. Example. Consider the standard inner product on F^n and F^m (for $F = \mathbb{R}$ or $F = \mathbb{C}$). Let $A \in \mathrm{Mat}(m \times n, F)$ be a matrix and let $f = f_A\colon F^n \to F^m$ be the linear map given by multiplication by A. We denote the conjugate transpose $\bar{A}^\top$ by A^*. Then for every $v \in F^n$ and $w \in F^m$, we have

$$\langle f(v), w\rangle = \langle Av, w\rangle = (Av)^\top \cdot \overline{w} = v^\top \cdot A^\top \cdot \overline{w} = v^\top \cdot \overline{\bar{A}^\top w} = \langle v, A^* w\rangle$$

(where the dot denotes matrix multiplication), so the adjoint $f^*\colon F^m \to F^n$ of f is given by multiplication by the matrix A^*.

9.21. Proposition (Properties of the Adjoint). *Let V_1, V_2, V_3 be finite-dimensional inner product spaces over the same field, and let $f, g : V_1 \to V_2$, $h : V_2 \to V_3$ be linear. Then*

(1) $(f + g)^* = f^* + g^*$, $(\lambda f)^* = \bar{\lambda} f^*$;
(2) $(h \circ f)^* = f^* \circ h^*$;
(3) $(f^*)^* = f$.

Proof.

(1) We have for $v \in V_1$, $v' \in V_2$

$$\begin{aligned}\langle v, (f+g)^*(v')\rangle &= \langle (f+g)(v), v'\rangle = \langle f(v), v'\rangle + \langle g(v), v'\rangle \\ &= \langle v, f^*(v')\rangle + \langle v, g^*(v')\rangle = \langle v, (f^* + g^*)(v')\rangle\end{aligned}$$

and

$$\begin{aligned}\langle v, (\lambda f)^*(v')\rangle &= \langle (\lambda f)(v), v'\rangle = \langle \lambda f(v), v'\rangle = \lambda \langle f(v), v'\rangle \\ &= \lambda \langle v, f^*(v')\rangle = \langle v, \bar{\lambda} f^*(v')\rangle = \langle v, (\bar{\lambda} f^*)(v')\rangle\,.\end{aligned}$$

The claim follows from the uniqueness of the adjoint.

(2) We argue in a similar way. For $v \in V_1$, $v' \in V_3$,

$$\begin{aligned}\langle v, (h \circ f)^*(v')\rangle &= \langle (h \circ f)(v), v'\rangle = \langle h\big(f(v)\big), v'\rangle \\ &= \langle f(v), h^*(v')\rangle = \langle v, f^*\big(h^*(v')\big)\rangle = \langle v, (f^* \circ h^*)(v')\rangle\,.\end{aligned}$$

Again, the claim follows from the uniqueness of the adjoint.

(3) For all $v \in V_1$, $v' \in V_2$, we have

$$\langle v', f(v)\rangle = \overline{\langle f(v), v'\rangle} = \overline{\langle v, f^*(v')\rangle} = \langle f^*(v'), v\rangle = \langle v', (f^*)^*(v)\rangle \,,$$

which implies $\langle v', (f^*)^*(v) - f(v)\rangle = 0$. For $v' = (f^*)^*(v) - f(v)$, we find $\|v'\| = 0$, so $v' = 0$, and therefore $(f^*)^*(v) = f(v)$ for all v, so $f = (f^*)^*$.

□

Now we characterize isometries.

9.22. Proposition. *Let V and W be inner product spaces of the same finite dimension over the same field. Let $f : V \to W$ be linear. Then the following are equivalent.*

(1) *f is an isometry;*
(2) *f is an isomorphism and $f^{-1} = f^*$;*
(3) *$f \circ f^* = \mathrm{id}_W$;*
(4) *$f^* \circ f = \mathrm{id}_V$.*

PROOF. To show (1) ⇒ (2), we observe that for an isometry f and $v \in V$, $w \in W$, we have

$$\langle v, f^*(w)\rangle = \langle f(v), w\rangle = \langle f(v), f\bigl(f^{-1}(w)\bigr)\rangle = \langle v, f^{-1}(w)\rangle \,,$$

which implies $f^* = f^{-1}$. The implications (2) ⇒ (3) and (2) ⇒ (4) are clear. Now assume (say) that (4) holds (the argument for (3) is similar). Then f is injective, hence an isomorphism, and we get (2). Now assume (2), and let $v, v' \in V$. Then

$$\langle f(v), f(v')\rangle = \langle v, f^*\bigl(f(v')\bigr)\rangle = \langle v, v'\rangle \,,$$

so f is an isometry. □

9.23. Lemma. *Let V be a finite-dimensional inner product space over F with an orthonormal basis $B = (v_1, \dots, v_n)$. Consider the standard inner product on F^n. Then the isomorphism*

$$\varphi_B \colon F^n \to V, \quad (\lambda_1, \dots, \lambda_n) \mapsto \lambda_1 v_1 + \dots + \lambda_n v_n$$

is an isometry.

PROOF. We denote the standard inner product on F^n by $\langle \_, \_ \rangle$ as well. Note that if $v, v' \in V$ have coordinates $x = (x_1, \dots, x_n), x' = (x'_1, \dots, x'_n) \in F^n$ with respect to B (so that $\varphi_B(x) = v$ and $\varphi_B(x') = v'$), then we have $x_i = \langle v, v_i\rangle$ and $x'_i = \langle v', v_i\rangle$ by Theorem 9.14, which therefore also implies

$$\langle v, v'\rangle = x_1\overline{x'_1} + \dots + x_n\overline{x'_n} = \langle x, x'\rangle \,.$$

This shows that φ_B is indeed an isometry. □

9.24. Theorem. Let $f\colon V \to W$ be a linear map of finite-dimensional inner product spaces. Then we have

$$\operatorname{im}(f^*) = \bigl(\ker(f)\bigr)^\perp \qquad \text{and} \qquad \ker(f^*) = \bigl(\operatorname{im}(f)\bigr)^\perp .$$

Proof. Let F be the field over which V and W are inner product spaces. Let $\phi\colon V\times\bar{V} \to F$ and $\psi\colon W\times\bar{W} \to F$ be the bilinear forms that correspond to the inner products on V and W, respectively. Because V and W are finite-dimensional, the maps ϕ_R and ψ_R in the commutative diagram (6) are isomorphisms. Hence, they restrict to isomorphisms $\operatorname{im} f^* \to \operatorname{im} f^\top$ and $\ker f^* \to \ker f^\top$, respectively. By Remark 8.20, they also restrict to isomorphisms $(\ker f)^\perp \to (\ker f)^\circ$ and $(\operatorname{im} f)^\perp \to (\operatorname{im} f)^\circ$, respectively. Hence, the claimed identities follow after applying ϕ_R^{-1} and ψ_R^{-1} to the identities of Theorem 6.29, respectively. □

Alternative proof. We first show the inclusion $\operatorname{im}(f^*) \subset \bigl(\ker(f)\bigr)^\perp$. So let $z \in \operatorname{im}(f^*)$, say $z = f^*(y)$. Let $x \in \ker(f)$, then

$$\langle x, z\rangle = \langle x, f^*(y)\rangle = \langle f(x), y\rangle = \langle 0, y\rangle = 0\,,$$

so $z \in \bigl(\ker(f)\bigr)^\perp$. This inclusion implies

$$\dim\operatorname{im} f^* \le \dim(\ker f)^\perp = \dim V - \dim\ker f = \dim\operatorname{im} f. \tag{7}$$

The analogous inequality for f^* instead of f is

$$\dim\operatorname{im}(f^*)^* \le \dim\operatorname{im} f^*.$$

From the equality $(f^*)^* = f$ (see Proposition 9.21) we conclude

$$\dim\operatorname{im} f \le \dim\operatorname{im} f^*.$$

Combining this inequality with (7) shows that all inequalities are equalities, so $\operatorname{im}(f^*) = \bigl(\ker(f)\bigr)^\perp$. Applying this to f^* instead of f yields $\operatorname{im}(f) = \bigl(\ker(f^*)\bigr)^\perp$, which is equivalent to the second identity claimed in the theorem. □

Now we relate the notions of adjoint etc. to matrices representing the linear maps *with respect to orthonormal bases.*

9.25. Proposition. *Let V and W be two inner product spaces over the same field, let $B = (v_1, \dots, v_n)$ and $C = (w_1, \dots, w_m)$ be orthonormal bases of V and W, respectively, and let $f\colon V \to W$ be linear. If f is represented by the matrix A relative to the given bases, then the adjoint map f^* is represented by the conjugate transpose matrix $A^* = \bar{A}^\top$ with respect to the same bases, that is*

$$[f^*]^C_B = ([f]^B_C)^*.$$

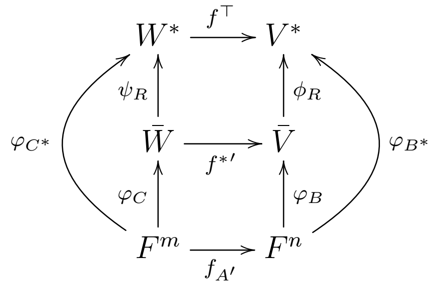
Note that when we have real inner product spaces, then $A^* = A^\top$ is simply the transpose.

Proof. Let $F = \mathbb{R}$ or $\mathbb{C}$ be the field of scalars. Let $\varphi_B\colon F^n \to V$ and $\varphi_C\colon F^m \to W$ be the usual maps associated to the bases B and C, respectively. By Lemma 9.23, these two maps are isometries, so we have $\varphi_B^* = \varphi_B^{-1}$ and $\varphi_C^* = \varphi_C^{-1}$. By definition, the map $\varphi_C^{-1} \circ f \circ \varphi_B\colon F^n \to F^m$ is given by multiplication by the matrix $A = [f]^B_C$. By Example 9.20, multiplication by the conjugate transpose A^* of A gives the adjoint of this map, which equals

$$\bigl(\varphi_C^{-1} \circ f \circ \varphi_B\bigr)^* = \varphi_B^* \circ f^* \circ (\varphi_C^{-1})^* = \varphi_B^{-1} \circ f^* \circ \varphi_C.$$

By definition, this map is also given by multiplication by $[f^*]^C_B$, so we conclude $[f^*]^C_B = A^* = ([f]^B_C)^*$. In other words, the matrix $\bar{A}^\top = A^*$ represents f^*. □

Alternative proof. To distinguish between the linear map $f^*\colon W \to V$ and the same map between the associated complex conjugate spaces, we write $f^{*\prime}\colon \bar{W} \to \bar{V}$ for the latter. Set $A' = [f^{*\prime}]^C_B$. Note that this means $f^{*\prime} \circ \varphi_C = \varphi_B \circ f_{A'}$, where $\varphi_B\colon F^n \to \bar{V}$ and $\varphi_C\colon F^m \to \bar{W}$ are the usual maps associated to the bases B for $\bar{V}$ and C for $\bar{W}$, respectively; this means that in terms of the scalar multiplication on V we have $\varphi_B\big((\lambda_1, \ldots, \lambda_n)\big) = \bar{\lambda}_1 v_1 + \cdots + \bar{\lambda}_n v_n$, and similarly for φ_C. Let B^* and C^* be the bases of V^* and W^* dual to B and C, respectively. Let $\phi\colon V \times \bar{V} \to F$ and $\psi\colon W \times \bar{W} \to F$ denote the bilinear forms associated to the inner products on V and W, respectively. Since $\phi_R\colon \bar{V} \to V^*$ and $\psi_R\colon \bar{W} \to W^*$ send orthonormal bases to their duals (exercise 9.5), we have $\varphi_{B^*} = \phi_R \circ \varphi_B$ and $\varphi_{C^*} = \psi_R \circ \varphi_C$. Then the commutative diagram (6) extends to the following commutative diagram.



We conclude $A' = [f^\top]^{C^*}_{B^*}$, so from Proposition 6.15 we find $A' = A^\top$. From Remark 9.4 we then conclude $[f^*]^C_B = \overline{[f^{*\prime}]^C_B} = \overline{A'} = \overline{A^\top} = A^*$. □

Warning. If the given bases are not orthonormal, then the statement is *wrong* in general.

9.26. Corollary. *Let V and W be two inner product spaces over the same field, let $B = (v_1, \ldots, v_n)$ and $C = (w_1, \ldots, w_m)$ be orthonormal bases of V and W, respectively, and let $f\colon V \to W$ be linear. Set $A = [f]^B_C$. We have the following.*

(1) *The map f is an isometry if and only if $A^* = A^{-1}$.*
(2) *Suppose $V = W$ and $B = C$. Then f is self-adjoint if and only if $A^* = A$.*
(3) *Suppose $V = W$ and $B = C$. Then f is normal if and only if $A^*A = AA^*$.*

Proof. Exercise. □

9.27. Definition. A matrix $A \in \mathrm{Mat}(n, \mathbb{R})$ is

(1) *symmetric* if $A^\top = A$;
(2) *normal* if $AA^\top = A^\top A$;
(3) *orthogonal* if $AA^\top = I_n$.

A matrix $A \in \mathrm{Mat}(n, \mathbb{C})$ is

(1) *Hermitian* if $A^* = A$;
(2) *normal* if $AA^* = A^*A$;
(3) *unitary* if $AA^* = I_n$.

These properties correspond to the properties "self-adjoint", "normal", "isometry" of the linear map given by A on the standard inner product space $\mathbb{R}^n$ or $\mathbb{C}^n$. Correspondingly, isometries of real inner product spaces are also called orthogonal maps, and isometries of complex inner product spaces are also called unitary maps.

9.28. Example. Lemma 9.23 was used to prove Proposition 9.25, and we can recover Lemma 9.23 from Proposition 9.25. Indeed, suppose V is an n-dimensional inner product space over F with $F = \mathbb{R}$ or $F = \mathbb{C}$, and let $B = (v_1, \ldots, v_n)$ be an orthonormal basis. Let E denote the standard (orthonormal) basis for F^n. Let $\varphi_B \colon F^n \to V$ be the map that sends $(\lambda_1, \ldots, \lambda_n)$ to $\sum_i \lambda_i v_i$. Then the associated matrix $A = [\varphi_B]_B^E$ is the identity, which is unitary, so φ_B is an isometry.

9.29. Example. Suppose V is an n-dimensional inner product space over F with $F = \mathbb{R}$ or $F = \mathbb{C}$, and let B and B' be two orthonormal bases for V. Then the base change matrix $P = [\mathrm{id}_V]_B^{B'}$ is unitary, because the identity map is an isometry.

Exercises.

(1) Let V be the vector space of continuous complex-valued functions defined on the interval $[0, 1]$, with the inner product $\langle f, g\rangle = \int_0^1 f(x)\overline{g(x)}\,dx$. Show that the set $\{x \mapsto e^{2\pi i k x} : k \in \mathbb{Z}\} \subset V$ is orthonormal. Is it a basis for V?

(2) Give an orthonormal basis for the 2-dimensional complex subspace V_3 of the standard inner product space $\mathbb{C}^3$ given by the equation $x_1 - ix_2 + ix_3 = 0$.

(3) For the real vector space V of polynomial functions $[-1, 1] \to \mathbb{R}$ with inner product given by

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)dx,$$

apply the Gram-Schmidt procedure to the elements $1, x, x^2, x^3$.

(4) For the real vector space V of continuous functions $[-\pi, \pi] \to \mathbb{R}$ with inner product given by

$$\langle f, g\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)dx$$

show that the functions

$$1/\sqrt{2}, \sin x, \cos x, \sin 2x, \cos 2x, \ldots$$

form an orthonormal set. [Note: for any function f the inner products with this list of functions is the sequence of Fourier coefficients of f.]

(5) Let F be $\mathbb{R}$ or $\mathbb{C}$, and let V be a finite-dimensional inner product space over F. Let $\phi \colon V \times \bar{V} \to F$ be the bilinear form corresponding to the inner product, and let $\phi_L \colon V \to \bar{V}^*$ and $\phi_R \colon \bar{V} \to V^*$ be the usual induced linear maps. Show that ϕ_L and ϕ_R send every orthonormal basis to its dual basis.

(6) Let A be an orthogonal $n \times n$ matrix with entries in $\mathbb{R}$. Show that $\det A = \pm 1$. If A is an orthogonal 2×2 matrix with entries in $\mathbb{R}$ and $\det A = 1$, show that A is a rotation matrix $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ for some $\theta \in \mathbb{R}$.

(7) For which values of $\alpha \in \mathbb{C}$ is the matrix $\begin{pmatrix} \alpha & \frac{1}{2} \\ \frac{1}{2} & \alpha \end{pmatrix}$ unitary?

(8) Show that the matrix of a normal transformation of a 2-dimensional real inner product space with respect to an orthonormal basis has one of the forms

$$\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} \alpha & \beta \\ \beta & \delta \end{pmatrix}.$$

(9) Let V be the vector space of infinitely differentiable functions $f\colon \mathbb{R} \to \mathbb{C}$ satisfying $f(x+2) = f(x)$ for all $x \in \mathbb{R}$. Consider the inner product on V given by $\langle p, q \rangle = \int_{-1}^{1} p(x)\overline{q(x)}dx$. Show that the operator $D : p \mapsto p''$ is self-adjoint.

(10) TRUE or FALSE? Give a proof or counterexample.
   (a) For any two real symmetric $n \times n$ matrices, the product is symmetric.
   (b) For any two real normal $n \times n$ matrices, the product is normal.
   (c) For any two real orthogonal $n \times n$ matrices, the product is orthogonal.

(11) Let n be a positive integer. Show that there exists an orthogonal anti-symmetric $n \times n$-matrix with real coefficients if and only if n is even.

(12) Let $V \subset \mathbb{R}^3$ be a plane through the origin. Let $\pi\colon \mathbb{R}^3 \to \mathbb{R}^3$ be the projection onto V.
   (a) Show that $\mathbb{R}^3$ has an orthonormal basis B of eigenvectors for π.
   (b) Show that for such a basis B, the associated matrix $[\pi]_B^B$ is diagonal.
   (c) Show that the matrix $[\pi]_E^E$, where E is the standard basis for $\mathbb{R}^3$, is symmetric.

(13) Consider $\mathbb{R}^n$ with the standard inner product, and let $V \subset \mathbb{R}^n$ be a subspace. Let A be the $n \times n$-matrix of orthogonal projection on V. Show that A is symmetric.

(14) Give an alternative proof of Proposition 9.21 that follows the ideas of the alternative proof of Theorem 9.15. (Hint: For (3), use Remark 9.19, the identity $\phi_L = \phi_R^\top \circ \alpha_V$ and its equivalent for W, and Proposition 6.17.)

(15) Let V be an inner product space and $U \subset V$ a finite-dimensional subspace. Let the inclusion map be denoted by $\iota\colon U \hookrightarrow V$. Show that we have $\ker \iota^* = U^\perp$.

(16) Suppose

$$U \xrightarrow{f} V \xrightarrow{g} W$$

is an exact sequence of linear maps between finite-dimensional inner product spaces. Show that there is an induced exact sequence

$$W \xrightarrow{g^*} V \xrightarrow{f^*} U\,.$$

(17) Check for all finite-dimensional inner product spaces in the results and exercises of this chapter whether the assumption of finite-dimensionality can be left out (possibly by replacing it by the assumption that certain adjoint maps exist). If so, give a proof of the stronger statement. If not, give a counterexample.

(18) Let V be a real inner product space, and $f\colon V \to V$ an endomorphism. Define the map

$$\phi\colon V \times V \to \mathbb{R}, \quad (x, y) \mapsto \langle f(x), y\rangle.$$

(a) Show that ϕ is a bilinear map.
(b) Show that if V is finite-dimensional, then every bilinear map is of this form.
(c) Show that ϕ is symmetric if and only if f is self-adjoint.

(19) Let V be a complex inner product space, and $f\colon V \to V$ an endomorphism. Define the map

$$\phi\colon V \times V \to \mathbb{C}, \quad (x, y) \mapsto \langle f(x), y\rangle.$$

(a) Show that ϕ is a sesquilinear map.
(b) Show that if V is finite-dimensional, then every sesquilinear map is of this form.
(c) Show that ϕ is a Hermitian form if and only if f is self-adjoint.

(20) Let V_1, V_2, W_1, and W_2 be vector spaces, and let $\phi\colon V_1 \times V_2 \to F$ and $\psi\colon W_1 \times W_2 \to F$ be two nondegenerate bilinear forms.

(a) Show that for every linear map $f\colon V_1 \to W_1$ there is a unique map $f^\dagger\colon W_2 \to V_2$ such that for all $x \in V_1$ and all $y \in W_2$ we have

$$\phi(x, f^\dagger(y)) = \psi(f(x), y).$$

(b) Show that we have

$$\operatorname{im} f^\dagger = (\ker f)^\perp \qquad \text{and} \qquad \ker f^\dagger = (\operatorname{im} f)^\perp.$$

(21) Show that an endomorphism f of an inner product space V is normal if and only if f has an adjoint f^* and for all $v, v' \in V$ we have

$$\langle f(v), f(v')\rangle = \langle f^*(v), f^*(v')\rangle.$$

(22) Let $f_1\colon V \to W_1$ and $f_2\colon V \to W_2$ be two linear maps of inner product spaces over the same field. Show that the following two conditions are equivalent.

(i) For all $v \in V$ we have $\|f_1(v)\| = \|f_2(v)\|$.
(ii) For all $v, v' \in V$ we have $\langle f_1(v), f_1(v')\rangle = \langle f_2(v), f_2(v')\rangle$.

(23) Let $f\colon V \to W$ be a linear map of inner product spaces over the same field.

(a) Show that f is an isometry if and only if f is an isomorphism and for all $v \in V$ we have $\|f(v)\| = \|v\|$.
(b) Suppose V and W have the same finite dimension. Show that f is an isometry if and only if for all $v \in V$ we have $\|f(v)\| = \|v\|$.

(24) Let $f_1\colon V \to W_1$ and $f_2\colon V \to W_2$ be two linear maps of inner product spaces over the same field. Suppose that the two equivalent conditions of Exercise 22 hold.

(a) Show that f_1 and f_2 have the same kernel.
(b) Show that there exists a unique isometry $g\colon \operatorname{im} f_1 \to \operatorname{im} f_2$ such that $f_2 = g \circ f_1$.

(25) Let $f_1\colon V \to W_1$ and $f_2\colon V \to W_2$ be any two maps of real inner product spaces that satisfy $f_1(0) = 0$ and $f_2(0) = 0$. Show that the following two conditions are equivalent.

(i) For all $v, v' \in V$ we have $\|f_1(v) - f_1(v')\| = \|f_2(v) - f_2(v')\|$.
(ii) For all $v, v' \in V$ we have $\langle f_1(v), f_1(v')\rangle = \langle f_2(v), f_2(v')\rangle$.

(26) Let $f\colon V \to W$ be any map of real inner product spaces of the same finite dimension that satisfies $f(0) = 0$. Show that f is an isometry if and only if for all $v, v' \in V$ we have $\|f(v) - f(v')\| = \|v - v'\|$.

(27) The Cauchy-Schwarz inequality allows us to define the angle between any two nonzero vectors x and y in the same real inner product space as the unique real number $\alpha \in [0, \pi]$ for which we have

$$\cos\alpha = \frac{\langle x, y\rangle}{\|x\| \cdot \|y\|}.$$

We denote this angle by $\angle(x, y)$. Suppose that V and W are real inner product spaces, and $f\colon V \to W$ is an isomorphism that preserves angles at 0, that is, for all $x, y \in V$ we have

$$\angle\big(f(x), f(y)\big) = \angle(x, y).$$

Show that f is the composition of an isometry with the multiplication by a scalar.

(28) Suppose that V and W are real inner product spaces of dimension at least 2, and $f\colon V \to W$ is a bijection that preserves general angles, that is, for all $x, y, z \in V$ we have

$$\angle\big(f(x) - f(z), f(y) - f(z)\big) = \angle(x - z, y - z).$$

Show that f is the composition of a translation, the multiplication by a scalar, and an isometry.

10. Orthogonal Diagonalization

In this section, we discuss the following question. Let V be an inner product space and $f : V \to V$ an endomorphism. When is it true that f has an *orthonormal* basis of eigenvectors (so can be orthogonally diagonalized or is *orthodiagonalizable* — nice word!)?

After a few general lemmas, we will first consider the case of complex inner product spaces, for which, as we will see, f has an orthonormal basis of eigenvectors if and only if f is normal.

10.1. Lemma. *Let V be a finite-dimensional inner product space and let $f : V \to V$ be an endomorphism. If f is orthodiagonalizable, then f is normal.*

Proof. If f is orthodiagonalizable, then there exists an orthonormal basis $(e_1, \dots, e_n)$ of V such that f is represented by a diagonal matrix D with respect to this basis. Now D is normal, hence so is f, by Corollary 9.26. □

The proof of the other direction is a little bit more involved. We begin with the following partial result.

10.2. Lemma. *Let V be an inner product space, and let $f : V \to V$ be normal.*

(1) *For all $v \in V$ we have $\|f^*(v)\| = \|f(v)\|$.*
(2) *If $f(v) = \lambda v$ for some $v \in V$, then $f^*(v) = \bar{\lambda} v$.*
(3) *If $f(v) = \lambda v$ and $f(w) = \mu w$ with $\lambda \neq \mu$, then $v \perp w$ (i.e., $\langle v, w\rangle = 0$).*

PROOF. For the first statement, note that

$$\begin{aligned}\|f^*(v)\|^2 &= \langle f^*(v), f^*(v)\rangle = \langle f\big(f^*(v)\big), v\rangle \\ &= \langle f^*\big(f(v)\big), v\rangle = \langle f(v), f(v)\rangle = \|f(v)\|^2 .\end{aligned}$$

For the second statement, note that

$$\begin{aligned}\langle f^*(v), f^*(v)\rangle &= \langle f(v), f(v)\rangle = |\lambda|^2\langle v, v\rangle \\ \langle \bar\lambda v, f^*(v)\rangle &= \bar\lambda\langle f(v), v\rangle = \bar\lambda\langle \lambda v, v\rangle = |\lambda|^2\langle v, v\rangle \\ \langle f^*(v), \bar\lambda v\rangle &= \lambda\langle v, f(v)\rangle = \lambda\langle v, \lambda v\rangle = |\lambda|^2\langle v, v\rangle \\ \langle \bar\lambda v, \bar\lambda v\rangle &= |\lambda|^2\langle v, v\rangle\end{aligned}$$

and so

$$\langle f^*(v) - \bar\lambda v, f^*(v) - \bar\lambda v\rangle = \langle f^*(v), f^*(v)\rangle - \langle \bar\lambda v, f^*(v)\rangle - \langle f^*(v), \bar\lambda v\rangle + \langle \bar\lambda v, \bar\lambda v\rangle = 0 .$$

This implies $f^*(v) - \bar\lambda v = 0$, so $f^*(v) = \bar\lambda v$.

For the last statement, we compute

$$\lambda\langle v, w\rangle = \langle f(v), w\rangle = \langle v, f^*(w)\rangle = \langle v, \bar\mu w\rangle = \mu\langle v, w\rangle .$$

Since $\lambda \neq \mu$ by assumption, we must have $\langle v, w\rangle = 0$. □

This result shows that the various eigenspaces are orthogonal in pairs, and we conclude that when f is a normal endomorphism of an inner product space, it is orthodiagonalizable if it is just diagonalizable. It remains to prove that this is the case.

10.3. Remark. Let V be an inner product space over F, and let $f : V \to V$ be normal. Let $\lambda \in F$ be an element. From Lemma 10.2(2), it follows that the eigenspace $E_\lambda(f)$ is contained in the eigenspace $E_{\bar\lambda}(f^*)$. Applying the same argument to f^*, and using $f^{**} = f$ (see Proposition 9.21), we also find the opposite inclusion, and we conclude $E_\lambda(f) = E_{\bar\lambda}(f^*)$. In particular, for $\lambda = 0$, we obtain $\ker f = \ker f^*$.

10.4. Lemma. *Let V be an inner product space over the field $F = \mathbb{R}$ or $\mathbb{C}$, let $f : V \to V$ be normal, and let $p \in F[X]$ be a polynomial. Then $p(f)$ is also normal.*

PROOF. Let $p(x) = a_m x^m + \cdots + a_0$. Then by Prop. 9.21,

$$p(f)^* = (a_m f^m + \cdots + a_1 f + a_0 \operatorname{id}_V)^* = \bar a_m (f^*)^m + \cdots + \bar a_1 f^* + \bar a_0 \operatorname{id}_V = \bar p(f^*) ,$$

where $\bar p$ is the polynomial whose coefficients are the complex conjugates of those of p. (If $F = \mathbb{R}$, then $p(f)^* = p(f^*)$.) Now $p(f)$ and $p(f)^* = \bar p(f^*)$ commute since f and f^* do, hence $p(f)$ is normal. □

10.5. Lemma. *Let V be a finite-dimensional inner product space, and let $f : V \to V$ be normal. Then $V = \ker(f) \oplus \mathrm{im}(f)$ is an orthogonal direct sum.*

PROOF. Let $v \in \ker(f)$ and $w \in \mathrm{im}(f)$. We have $f(v) = 0$, so $f^*(v) = 0$ by Lemma 10.2, and $w = f(u)$ for some $u \in V$. Then

$$\langle v, w\rangle = \langle v, f(u)\rangle = \langle f^*(v), u\rangle = \langle 0, u\rangle = 0\,,$$

so $v \perp w$. In particular, we have $\ker f \cap \mathrm{im}\, f = \{0\}$, because the inner product is positive definite. From $\dim \ker(f) + \dim \mathrm{im}(f) = \dim V$, we conclude

$$\dim(\ker(f) + \mathrm{im}(f)) = \dim \ker(f) + \dim \mathrm{im}(f) - \dim(\ker f \cap \mathrm{im}\, f) = \dim V,$$

so $\ker(f) + \mathrm{im}(f) = V$, which finishes the proof. □

ALTERNATIVE PROOF. Take $U = \mathrm{im}\, f$. From Proposition 9.8, we know that $V = U \oplus U^\perp$ is an orthogonal direct sum. From Theorem 9.24 and Remark 10.3, we find $U^\perp = (\mathrm{im}\, f)^\perp = \ker f^* = \ker f$. □

10.6. Lemma. *Let V be a finite-dimensional complex inner product space, and let $f : V \to V$ be normal. Then f is diagonalizable.*

PROOF. We will show that the minimal polynomial of f does not have multiple roots. So assume the contrary, namely that

$$M_f(x) = (x - \alpha)^2 g(x)$$

for some $\alpha \in \mathbb{C}$ and some polynomial g. We know that $f - \alpha\,\mathrm{id}_V$ is normal. Let $v \in V$ and consider $w = (f - \alpha\,\mathrm{id}_V)\big(g(f)(v)\big)$. Obviously $w \in \mathrm{im}(f - \alpha\,\mathrm{id}_V)$, but also $(f - \alpha\,\mathrm{id}_V)(w) = M_f(f)(v) = 0$, so $w \in \ker(f - \alpha\,\mathrm{id}_V)$. By the previous lemma, $w = 0$. Hence, f is already annihilated by the polynomial $(x - \alpha)g(x)$ of degree smaller than $M_f(x)$, a contradiction. □

ALTERNATIVE PROOF. We proceed by induction on $\dim V$. The base case $\dim V = 1$ (or $= 0$) is trivial. So assume $\dim V \geq 2$. Then f has at least one eigenvector v, say with eigenvalue λ. Let $U = \ker(f - \lambda\,\mathrm{id}_V) \neq 0$ be the eigenspace and $W = \mathrm{im}(f - \lambda\,\mathrm{id}_V)$. We know that $V = U \oplus W$ is an orthogonal direct sum by Lemma 10.5. Because f commutes with $f - \lambda\,\mathrm{id}_V$, we have that $f(U) \subset U$ and $f(W) \subset W$, so f is the direct sum of its restrictions to U and W. Then by uniqueness, f^* is also the direct sum of the adjoints of these restrictions, so normality of f implies normality of its restrictions. In particular, $f|_W : W \to W$ is again a normal map. By the induction hypothesis, $f|_W$ is diagonalizable. Since $f|_U = \lambda\,\mathrm{id}_U$ is trivially diagonalizable, f is diagonalizable. (The same proof would also prove directly that f is orthodiagonalizable.) □

So we have now proved the following statement, which is often referred to as the *Spectral Theorem* (though this may also refer to some other related theorems).

10.7. Theorem. *Let V be a finite-dimensional complex inner product space, and let $f : V \to V$ be a linear map. Then V has an orthonormal basis of eigenvectors for f if and only if f is normal.*

PROOF. Indeed, Lemma 10.1 states the "only if"-part. For the converse, assume f is normal. Then f is diagonalizable by Lemma 10.6, which means that the concatenation of any bases for the eigenspaces yields a basis for V. Lemma 10.2 shows that if we take the bases of the eigenspaces to be orthonormal, which we can do by applying Gram-Schmidt orthonormalization (Theorem 9.10) to any basis, then the concatenation is orthonormal as well, so f has an orthonormal basis of eigenvectors. □

This nice result leaves one question open: what is the situation for *real* inner product spaces? The key to this is the following observation.

10.8. Proposition. *Let V be a finite-dimensional complex inner product space, and let $f : V \to V$ be a linear map. Then f is normal with all eigenvalues real if and only if f is self-adjoint.*

Proof. We know that a self-adjoint map is normal. So assume now that f is normal. Then there is an ONB of eigenvectors, and with respect to this basis, f is represented by a diagonal matrix D, so we have $D^* = \bar{D}^\top = \bar{D}$. Obviously, we have that f is self-adjoint if and only if $D = D^*$, which reduces to $D = \bar{D}$, which happens if and only if all entries of D (i.e., the eigenvalues of f) are real. □

This implies the following.

10.9. Theorem. *Let V be a finite-dimensional real inner product space, and let $f : V \to V$ be linear. Then V has an orthonormal basis of eigenvectors for f if and only if f is self-adjoint.*

Proof. If f has an ONB of eigenvectors, then its matrix with respect to this basis is diagonal and so symmetric, hence f is self-adjoint.

For the converse, set $n = \dim V$, choose any orthonormal basis B for V and suppose that f is self-adjoint. Then the associated real matrix $A = [f]_B^B$ satisfies $A^* = A$ by Corollary 9.26. Hence, the associated map $f_A \colon \mathbb{C}^n \to \mathbb{C}^n$ is self-adjoint with respect to the standard Hermitian inner product (see Example 9.6). Therefore, the matrix A, viewed over $\mathbb{C}$, is normal and has all its eigenvalues (over $\mathbb{C}$) real by Proposition 10.8. The fact that A is normal over $\mathbb{C}$ implies that A is diagonalizable over $\mathbb{C}$ by Theorem 10.7. By Proposition 3.9 this means that the minimal polynomial $M_{A/\mathbb{C}}$ of A as a matrix over $\mathbb{C}$ is the product of distinct linear factors, which has the real eigenvalues as roots, and is therefore a polynomial with real coefficients satisfying $M_{A/\mathbb{C}}(A) = 0$. This implies that the minimal polynomial $M_{A/\mathbb{R}}$ of A as a matrix over $\mathbb{R}$ is a factor of $M_{A/\mathbb{C}}$, which means that $M_{A/\mathbb{R}}$ is also the product of distinct linear factors over $\mathbb{R}$. (In fact, we have $M_{A/\mathbb{R}} = M_{A/\mathbb{C}}$; if not, then some real factor p of $M_{A/\mathbb{C}}$ of smaller degree would satisfy $p(A) = 0$, and since p is also a polynomial over $\mathbb{C}$, this contradicts the minimality of $M_{A/\mathbb{C}}$ among complex polynomials that vanish on A.) Applying Proposition 3.9 again shows that A, and thus f, is also diagonalizable over $\mathbb{R}$. Lemma 10.2, (3) then shows that the eigenspaces are orthogonal in pairs. Hence, concatenating orthonormal bases for the different eigenspaces, obtainable with Gram-Schmidt orthonormalization (Theorem 9.10), yields an orthonormal basis of eigenvectors for V. □

In terms of matrices, this reads as follows.

10.10. Theorem. *Let A be a square matrix with real entries. Then A is orthogonally similar to a diagonal matrix (i.e., there is an orthogonal matrix P such that $P^{-1}AP$ is a diagonal matrix) if and only if A is symmetric. In this case, we can choose P to be orientation-preserving, i.e., to have* $\det P = 1$ *(and not* -1*).*

PROOF. The first statement follows from the previous theorem. To see that we can take P with $\det P = 1$, assume that we already have an orthogonal matrix Q such that $Q^{-1}AQ = D$ is diagonal, but with $\det Q = -1$. The diagonal matrix T with diagonal entries $(-1, 1, \ldots, 1)$ is orthogonal and $\det T = -1$, so $P = QT$ is also orthogonal, and $\det P = 1$. Furthermore,

$$P^{-1}AP = T^{-1}Q^{-1}AQT = TDT = D\,,$$

so P has the required properties. □

10.11. Remark. For an orthogonal matrix P, we have $P^{-1} = P^\top$, so we could have also written $P^\top AP$ in Theorem 10.10. If we want just any matrix P for which $P^\top AP$ is diagonal, and we do not need P to be orthogonal, then it is often easier to apply Theorem 8.27, especially when the eigenvalues of A are difficult to compute.

10.12. Remark. Note that we have two notions of *diagonalisation*, one for linear maps as in Linear Algebra I, and one for symmetric bilinear forms as in Theorem 8.27. Theorem 10.10 can be interpreted in both contexts. Indeed, suppose B and B' are two bases for a finite-dimensional vector space V over a field F, and set $P = [\mathrm{id}_V]^{B'}_B$. If $f\colon V \to V$ is an endomorphism of V, and $M = [f]^B_B$ and $M' = [f]^{B'}_{B'}$ are the two matrices associated to f with respect to the bases B and B', respectively, then we have $M' = P^{-1}MP$, so M' and M are similar.
If $\phi\colon V \times V \to F$ is a bilinear form on V, and A is the matrix that represents ϕ with respect to the basis B, while A' is the matrix that represents ϕ with respect to the basis B', then by Proposition 8.14 we have $A' = P^\top AP$, so A' and A are congruent.
To diagonalise the linear map f or the bilinear form ϕ, respectively, means to find a basis B' for which M' or A', respectively, is a diagonal matrix. If we already know an initial basis B, with the corresponding associated matrix M or A, respectively, then this goal is equivalent to finding an invertible matrix P, for which $P^{-1}MP$ or $P^\top AP$, respectively, is diagonal.
If V is a finite-dimensional real inner product space, and B and B' are to be orthonormal bases, then for $P = [\mathrm{id}_V]^{B'}_B$ we have $P^\top = P^{-1}$ by Example 9.29. Therefore, Theorem 10.10, which as a consequence of Theorem 10.9 was proved in the context of linear maps, can be reinterpreted in terms of bilinear forms: if ϕ is a symmetric bilinear form on V (not necessarily the one giving the inner product!), then there is an orthonormal basis B (with respect to the inner product) for V that diagonalises ϕ, that is, such that the matrix that represents ϕ with respect to B is diagonal. See Exercise 7.

Theorem 10.10 has a geometric interpretation. If A is a symmetric 2×2-matrix, then the equation

$$v^\top Av = 1 \tag{8}$$

in terms of $v = (x, y) \in \mathbb{R}^2$ defines a *conic section* in the plane. Our theorem implies that there is a *rotation* P such that $P^{-1}AP$ is diagonal. This means that in a suitably rotated coordinate system, our conic section has an equation of the form

$$a\,x^2 + b\,y^2 = 1\,,$$

where a and b are the eigenvalues of A. We can use their signs to classify the geometric shape of the conic section (ellipse, hyperbola, empty, degenerate).

The directions given by the eigenvectors of A are called the *principal axes* of the conic section (or of A), and the coordinate change given by P is called the *principal axes transformation*. Similar statements are true for higher-dimensional *quadrics* given by equation (8) when A is a larger symmetric matrix.

10.13. Example. Let us consider the conic section given by the equation

$$5\,x^2 + 4\,xy + 2\,y^2 = 1\,.$$

The matrix is

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}\,.$$

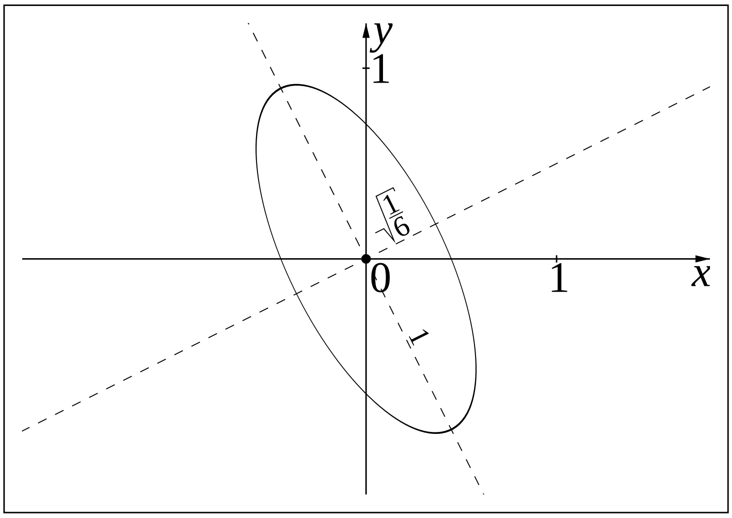
We have to find its eigenvalues and eigenvectors. The characteristic polynomial is $(X-5)(X-2)-4 = X^2 - 7\,X + 6 = (X-1)(X-6)$, so we have the two eigenvalues 1 and 6. This already tells us that we have an ellipse. To find the eigenvectors, we have to determine the kernels of $A - I$ and $A - 6I$. We get

$$A - I = \begin{pmatrix} 4 & 2 \\ 2 & 1 \end{pmatrix} \quad \text{and} \quad A - 6I = \begin{pmatrix} -1 & 2 \\ 2 & -4 \end{pmatrix}\,,$$

so the eigenvectors are multiples of $(1,-2)$ and of $(2,1)$. To get an orthonormal basis, we have to scale them appropriately; we also need to check whether we have to change the sign of one of them in order to get an orthogonal matrix with determinant 1. Here, we obtain

$$P = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \quad \text{and} \quad P^{-1}AP = \begin{pmatrix} 1 & 0 \\ 0 & 6 \end{pmatrix}\,.$$

To sketch the ellipse, note that the principal axes are in the directions of the eigenvectors and that the ellipse meets the first axis (in the direction of $(1,-2)$) at a distance of 1 from the origin and the second axis (in the direction of $(2,1)$) at a distance of $1/\sqrt{6}$ from the origin.



The ellipse $5\,x^2 + 4\,xy + 2\,y^2 = 1$.

10.14. Example. Consider the symmetric matrix

$$A = \begin{pmatrix} 5 & -2 & 4 \\ -2 & 8 & 2 \\ 4 & 2 & 5 \end{pmatrix}.$$

We will determine an orthogonal matrix Q and a diagonal matrix D such that $A = QDQ^\top$. The characteristic polynomial of A is the determinant of

$$tI - A = \begin{pmatrix} t-5 & 2 & -4 \\ 2 & t-8 & -2 \\ -4 & -2 & t-5 \end{pmatrix},$$

which is easily determined to be $P_A(t) = t(t-9)^2$, so we have eigenvalues 0 and 9. The eigenspace for eigenvalue $\lambda = 0$ is the kernel $\ker A$. From a row echelon form for A, which we will leave out here, we find that this kernel is generated by $(2, 1, -2)$. Normalising gives the unit vector $v_1 = \frac{1}{3}(2, 1, -2)$, which forms a basis for the eigenspace for $\lambda = 0$. The eigenspace for eigenvalue $\lambda = 9$ is the kernel of

$$A - 9I = \begin{pmatrix} -4 & -2 & 4 \\ -2 & -1 & 2 \\ 4 & 2 & -4 \end{pmatrix}.$$

A row echelon form for this matrix is

$$\begin{pmatrix} 2 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

from which we find that this eigenspace is generated by $w_1 = (1, 0, 1)$ and $w_2 = (1, -2, 0)$. Within this eigenspace we apply Gram-Schmidt orthonormalisation to find an orthonormal basis for the eigenspace. We find w_1 and

$$w_2 - \frac{\langle w_2, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 = w_2 - \tfrac{1}{2} w_1 = \tfrac{1}{2}(1, -4, -1).$$

After normalising this yields $v_2 = \frac{1}{\sqrt{2}}(1, 0, 1)$ and $v_3 = \frac{1}{3\sqrt{2}}(1, -4, -1)$.

Our new basis becomes $B = (v_1, v_2, v_3)$. By Lemma 10.2, the two eigenspaces are orthogonal to each other, so B is an orthonormal basis of eigenvectors. Hence, the matrix $Q = [\mathrm{id}]_E^B$ is orthogonal, that is, $Q^{-1} = Q^\top$. For the diagonal matrix

$$D = [f_A]_B^B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{pmatrix}$$

we find

$$A = [f_A]_E^E = [\mathrm{id}]_E^B \cdot [f_A]_B^B \cdot [\mathrm{id}]_B^E = QDQ^{-1} = QDQ^\top.$$

The matrix $Q = [\mathrm{id}]_E^B$ has the basis vectors of B as columns, so we have

$$Q = \begin{pmatrix} \frac{2}{3} & \frac{1}{\sqrt{2}} & \frac{1}{3\sqrt{2}} \\ \frac{1}{3} & 0 & -\frac{2}{3}\sqrt{2} \\ -\frac{2}{3} & \frac{1}{\sqrt{2}} & -\frac{1}{3\sqrt{2}} \end{pmatrix}.$$

Exercises.

(1) Suppose that A is a real symmetric 2×2 matrix of determinant 2 for which $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ is an eigenvector with eigenvalue -1.
  (a) What is the other eigenvalue of A?
  (b) What is the other eigenspace?
  (c) Determine A.

(2) Consider the quadratic form $q(x, y) = 11x^2 - 16xy - y^2$.
  (a) Find a real symmetric matrix A for which

$$q(x, y) = (x \ y) \cdot A \cdot \begin{pmatrix} x \\ y \end{pmatrix}.$$

  (b) Find real numbers a, b and an orthogonal map $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ so that $q(f(u, v)) = au^2 + bv^2$ for all $u, v \in \mathbb{R}$.
  (c) What values does $q(x, y)$ assume on the unit circle $x^2 + y^2 = 1$?

(3) What values does the quadratic form $q(x, y, z) = 2xy + 2xz + y^2 - 2yz + z^2$ assume when (x, y, z) ranges over the unit sphere $x^2 + y^2 + z^2 = 1$ in $\mathbb{R}^3$?

(4) Suppose that A is an anti-symmetric $n \times n$ matrix over the real numbers.
  (a) Show that every eigenvalue of A over the complex numbers lies in $i\mathbb{R}$.

(b) If n is odd, show that 0 is an eigenvalue of A.

(5) Let V be an inner product space and let $f : V \to V$ be an endomorphism. Suppose that V has an orthonormal basis of eigenvectors for f. Show that f has an adjoint and that f is normal (see Lemma 10.1).

(6) Let A be a symmetric matrix over $\mathbb{R}$. Show that its signature is equal to the number of positive eigenvalues minus the number of negative eigenvalues.

(7) Let V be a finite-dimensional real inner product space, and $\phi\colon V \times V \to \mathbb{R}$ a symmetric bilinear form on V. Use Exercise 9.18 and Theorem 10.9 to show that there exists an orthonormal basis B for V that diagonalises ϕ, that is, such that the matrix that represents ϕ with respect to B is diagonal.

11. External Direct Sums

Earlier in this course, we have discussed direct sums of linear subspaces of a vector space. In this section, we discuss a way to contruct a vector space out of a given family of vector spaces in such a way that the given spaces can be identified with linear subspaces of the new space, which becomes their direct sum.

11.1. Definition. Let F be a field, and let $(V_i)_{i \in I}$ be a family of F-vector spaces. The *(external) direct sum* of the spaces V_i is the vector space

$$V = \bigoplus_{i \in I} V_i = \Big\{ (v_i) \in \prod_{i \in I} V_i : v_i = 0 \text{ for all but finitely many } i \in I \Big\} .$$

Addition and scalar multiplication in V are defined component-wise.

If I is finite, say $I = \{1, 2, \dots, n\}$, then we also write

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_n ;$$

as a set, it is just the cartesian product $V_1 \times \cdots \times V_n$.

11.2. Proposition. *Let $(V_i)_{i \in I}$ be a family of F-vector spaces, and*

$$V = \bigoplus_{i \in I} V_i$$

their direct sum.

(1) *There are injective linear maps $\iota_j : V_j \to V$ given by*

$$\iota_j(v_j) = (0, \dots, 0, v_j, 0, \dots) \qquad \text{with } v_j \text{ in the } j\text{th position}$$

such that with $\tilde{V}_j = \iota_j(V_j)$, we have $V = \bigoplus_{j \in I} \tilde{V}_j$ as a direct sum of subspaces.

(2) *If B_j is a basis for V_j, then $B = \bigcup_{j \in I} \iota_j(B_j)$ is a basis for V.*

(3) *If W is another F-vector space, and $\phi_j : V_j \to W$ are linear maps, then there is a unique linear map $\phi : V \to W$ such that $\phi_j = \phi \circ \iota_j$ for all $j \in I$.*

PROOF.

(1) This is clear from the definitions, compare 2.2.

(2) This is again clear from 2.2.

(3) A linear map is uniquely determined by its values on a basis. Let B be a basis as in (2). The only way to get $\phi_j = \phi \circ \iota_j$ is to define $\phi(\iota_j(b)) = \phi_j(b)$ for all $b \in B_j$; this gives a unique linear map $\phi : V \to W$.

□

Statement (3) above is called the *universal property* of the direct sum. It is essentially the only thing we have to know about $\bigoplus_{i \in I} V_i$; the explicit construction is not really relevant (except to show that such an object exists).

12. The Tensor Product

As direct sums allow us to "add" vector spaces in a way (which corresponds to "adding" their bases by taking the disjoint union), the tensor product allows us to "multiply" vector spaces ("multiplying" their bases by taking a cartesian product). The main purpose of the tensor product is to "linearize" multilinear maps.

You may have heard of "tensors". They are used in physics (there is, for example, the "stress tensor" or the "moment of inertia tensor") and also in differential geometry (the "curvature tensor" or the "metric tensor"). Basically a tensor is an element of a tensor product (of vector spaces), like a vector is an element of a vector space. You have seen special cases of tensors already. To start with, a scalar (element of the base field F) or a vector or a linear form are trivial examples of tensors. More interesting examples are given by linear maps, endomorphisms, bilinear forms and multilinear maps in general.

The vector space of $m \times n$ matrices over F can be identified in a natural way with the tensor product $(F^n)^* \otimes F^m$. This identification corresponds to the interpretation of matrices as linear maps from F^n to F^m. The vector space of $m \times n$ matrices over F can also identified in a (different) natural way with $(F^m)^* \otimes (F^n)^*$; this corresponds to the interpretation of matrices as bilinear forms on $F^m \times F^n$.

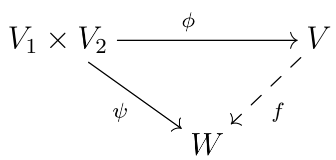
In these examples, we see that (for example), the set of all bilinear forms has the structure of a vector space. The tensor product generalizes this. Given two vector spaces V_1 and V_2, it produces a new vector space $V_1 \otimes V_2$ such that we have a natural identification

$$\mathrm{Bil}(V_1 \times V_2, W) \cong \mathrm{Hom}(V_1 \otimes V_2, W)$$

for all vector spaces W. Here $\mathrm{Bil}(V_1 \times V_2, W)$ denotes the vector space of bilinear maps from $V_1 \times V_2$ to W. The following definition states the property we want more precisely.

12.1. Definition. Let V_1 and V_2 be two vector spaces. A *tensor product* of V_1 and V_2 is a vector space V, together with a bilinear map $\phi : V_1 \times V_2 \to V$, satisfying the following "universal property":

For every vector space W and bilinear map $\psi : V_1 \times V_2 \to W$, there is a *unique* linear map $f : V \to W$ such that $\psi = f \circ \phi$.



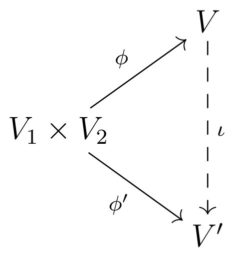
In other words, the canonical linear map

$$\mathrm{Hom}(V, W) \longrightarrow \mathrm{Bil}(V_1 \times V_2, W), \quad f \longmapsto f \circ \phi$$

is an isomorphism.

It is easy to see that there can be *at most* one tensor product in a very specific sense.

12.2. Lemma. *Any two tensor products (V, ϕ), (V', ϕ') are uniquely isomorphic in the following sense: There is a unique isomorphism $\iota : V \to V'$ such that $\phi' = \iota \circ \phi$.*



Proof. Since $\phi' : V_1 \times V_2 \to V'$ is a bilinear map, there is a unique linear map $\iota : V \to V'$ making the diagram above commute. For the same reason, there is a unique linear map $\iota' : V' \to V$ such that $\phi = \iota' \circ \phi'$. Now $\iota' \circ \iota : V \to V$ is a linear map satisfying $(\iota' \circ \iota) \circ \phi = \phi$, and id_V is another such map. But by the universal property, there is a unique such map, hence $\iota' \circ \iota = \mathrm{id}_V$. In the same way, we see that $\iota \circ \iota' = \mathrm{id}_{V'}$, therefore ι is an isomorphism. □

Because of this uniqueness, it is allowable to simply speak of "the" tensor product of V_1 and V_2 (provided it exists! — but see below). The tensor product is denoted $V_1 \otimes V_2$, and the bilinear map ϕ is written $(v_1, v_2) \mapsto v_1 \otimes v_2$.

It remains to show existence of the tensor product.

12.3. Proposition. *Let V_1 and V_2 be two vector spaces; choose bases B_1 of V_1 and B_2 of V_2. Let V be the vector space with basis $B = B_1 \times B_2$, and define a bilinear map $\phi : V_1 \times V_2 \to V$ via $\phi(b_1, b_2) = (b_1, b_2) \in B$ for $b_1 \in B_1$, $b_2 \in B_2$. Then (V, ϕ) is a tensor product of V_1 and V_2.*

Proof. Let $\psi : V_1 \times V_2 \to W$ be a bilinear map. We have to show that there is a unique linear map $f : V \to W$ such that $\psi = f \circ \phi$. Now if this relation is to be satisfied, we need to have $f((b_1, b_2)) = f(\phi(b_1, b_2)) = \psi(b_1, b_2)$. This fixes the values of f on the basis B, hence there can be at most one such linear map. It remains to show that the linear map thus defined satisfies $f(\phi(v_1, v_2)) = \psi(v_1, v_2)$ for all $v_1 \in V_1$, $v_2 \in V_2$. But this is clear since ψ and $f \circ \phi$ are two bilinear maps that agree on pairs of basis elements. □

12.4. Remark. This existence proof does not use that the bases are finite and so also works for infinite-dimensional vector spaces (given the fact that every vector space has a basis).

There is also a different construction that does not require the choice of bases. The price one has to pay is that one first needs to construct a gigantically huge space V (with *basis* $V_1 \times V_2$), which one then divides by another huge space (incorporating all relations needed to make the map $V_1 \times V_2 \to V$ bilinear) to end up with the relatively small space $V_1 \otimes V_2$. This is a kind of "brute force" approach, but it works.

Note that by the uniqueness lemma above, we always get "the same" tensor product, no matter which bases we choose.

12.5. Elements of $V_1 \otimes V_2$. What do the elements of $V_1 \otimes V_2$ look like? Some of them are values of the bilinear map $\phi : V_1 \times V_2 \to V_1 \otimes V_2$, so are of the form $v_1 \otimes v_2$. *But these are not all!* However, elements of this form span $V_1 \otimes V_2$, and since

$$\lambda(v_1 \otimes v_2) = (\lambda v_1) \otimes v_2 = v_1 \otimes (\lambda v_2)$$

(this comes from the bilinearity of ϕ), every element of $V_1 \otimes V_2$ can be written as a (finite) *sum* of elements of the form $v_1 \otimes v_2$.

The following result gives a more precise formulation that is sometimes useful.

12.6. Lemma. *Let V and W be two vector spaces, and let $w_1, \dots, w_n$ be a basis for W. Then every element of $V \otimes W$ can be written uniquely in the form*

$$\sum_{i=1}^{n} v_i \otimes w_i = v_1 \otimes w_1 + \dots + v_n \otimes w_n$$

with $v_1, \dots, v_n \in V$.

Proof. Let $x \in V \otimes W$; then by the discussion above, we can write

$$x = y_1 \otimes z_1 + \dots + y_m \otimes z_m$$

for some $y_1, \dots, y_m \in V$ and $z_1, \dots, z_m \in W$. Since $w_1, \dots, w_n$ is a basis for W, we can write

$$z_j = \alpha_{j1} w_1 + \dots + \alpha_{jn} w_n$$

with scalars α_{jk}. Using the bilinearity of the map $(y, z) \mapsto y \otimes z$, we find that

$$\begin{aligned} x &= y_1 \otimes (\alpha_{11} w_1 + \dots + \alpha_{1n} w_n) + \dots + y_m \otimes (\alpha_{m1} w_1 + \dots + \alpha_{mn} w_n) \\ &= (\alpha_{11} y_1 + \dots + \alpha_{m1} y_m) \otimes w_1 + \dots + (\alpha_{1n} y_1 + \dots + \alpha_{mn} y_m) \otimes w_n \,, \end{aligned}$$

which is of the required form.

For uniqueness, it suffices to show that

$$v_1 \otimes w_1 + \dots + v_n \otimes w_n = 0 \quad \Longrightarrow \quad v_1 = \dots = v_n = 0 \,.$$

Assume that $v_j \neq 0$. There is a bilinear form ψ on $V \times W$ such that $\psi(v_j, w_j) = 1$ and $\psi(v, w_i) = 0$ for all $v \in V$ and $i \neq j$. By the universal property of the tensor product, there is a linear form f on $V \otimes W$ such that $f(v \otimes w) = \psi(v, w)$. Applying f to both sides of the equation, we find that

$$0 = f(0) = f(v_1 \otimes w_1 + \dots + v_n \otimes w_n) = \psi(v_1, w_1) + \dots + \psi(v_n, w_n) = 1 \,,$$

a contradiction. □

In this context, one can think of $V \otimes W$ as being "the vector space W with scalars replaced by elements of V." This point of view will be useful when we want to enlarge the base field, e.g., in order to turn a real vector space into a complex vector space of the same dimension.

12.7. Basic Properties of the Tensor Product. Recall the axioms satisfied by a commutative "semiring" like the natural numbers:

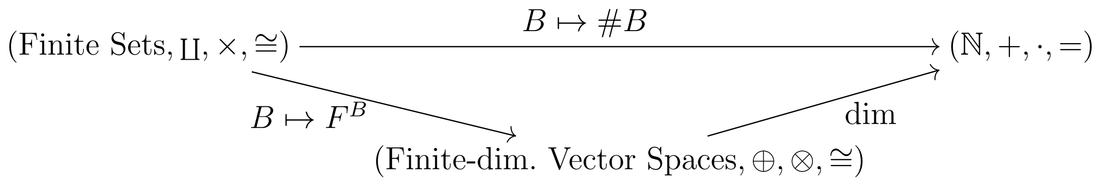
$$
\begin{aligned}
a + (b + c) &= (a + b) + c \\
a + b &= b + a \\
a + 0 &= a \\
a \cdot (b \cdot c) &= (a \cdot b) \cdot c \\
a \cdot b &= b \cdot a \\
a \cdot 1 &= a \\
a \cdot (b + c) &= a \cdot b + a \cdot c
\end{aligned}
$$

(The name "semi"ring refers to the fact that we do not require the existence of additive inverses.)

All of these properties have their analogues for vector spaces, replacing addition by direct sum, zero by the zero space, multiplication by tensor product, one by the one-dimensional space F, and equality by natural isomorphism:

$$
\begin{aligned}
U \oplus (V \oplus W) &\cong (U \oplus V) \oplus W \\
U \oplus V &\cong V \oplus U \\
U \oplus 0 &\cong U \\
U \otimes (V \otimes W) &\cong (U \otimes V) \otimes W \\
U \otimes V &\cong V \otimes U \\
U \otimes F &\cong U \\
U \otimes (V \oplus W) &\cong U \otimes V \ \oplus \ U \otimes W
\end{aligned}
$$

There is a kind of "commutative diagram":



Let us prove some of the properties listed above.

Proof. We show that $U \otimes V \cong V \otimes U$. We have to exhibit an isomorphism, or equivalently, linear maps going both ways that are inverses of each other. By the universal property, a linear map from $U \otimes V$ into any other vector space W is "the same" as a bilinear map from $U \times V$ into W. So we get a linear map $f : U \otimes V \to V \otimes U$ from the bilinear map $U \times V \to V \otimes U$ that sends (u, v) to $v \otimes u$. So we have $f(u \otimes v) = v \otimes u$. Similarly, there is a linear map $g : V \otimes U \to U \otimes V$ that satisfies $g(v \otimes u) = u \otimes v$. Since f and g are visibly inverses of each other, they are isomorphisms. $\square$

Before we go on to the next statement, let us make a note of the principle we have used.

12.8. Note. To give a linear map $f : U \otimes V \to W$, it is enough to specify $f(u \otimes v)$ for $u \in U$, $v \in V$. The map $U \times V \to W$, $(u, v) \mapsto f(u \otimes v)$ must be bilinear.

Proof. We now show that $U \otimes (V \otimes W) \cong (U \otimes V) \otimes W$. First fix $u \in U$. Then by the principle above, there is a linear map $f_u : V \otimes W \to (U \otimes V) \otimes W$ such that $f_u(v \otimes w) = (u \otimes v) \otimes w$. Now the map $U \times (V \otimes W) \to (U \otimes V) \otimes W$ that sends (u, x) to $f_u(x)$ is bilinear (check!), so we get a *linear* map $f : U \otimes (V \otimes W) \to (U \otimes V) \otimes W$ such that $f(u \otimes (v \otimes w)) = (u \otimes v) \otimes w$. Similarly, there is a linear map g in the other direction such that $g((u \otimes v) \otimes w) = u \otimes (v \otimes w)$. Since f and g are inverses of each other (this needs only be checked on elements of the form $u \otimes (v \otimes w)$ or $(u \otimes v) \otimes w$, since these span the spaces), they are isomorphisms. □

We leave the remaining two statements involving tensor products for the exercises.

Now let us look into the interplay of tensor products with linear maps.

12.9. Definition. Let $f : V \to W$ and $f' : V' \to W'$ be linear maps. Then $V \times V' \to W \otimes W'$, $(v, v') \mapsto f(v) \otimes f'(v')$ is bilinear and therefore corresponds to a linear map $V \otimes V' \to W \otimes W'$, which we denote by $f \otimes f'$. I.e., we have

$$(f \otimes f')(v \otimes v') = f(v) \otimes f'(v') .$$

12.10. Lemma. $\mathrm{id}_V \otimes \mathrm{id}_W = \mathrm{id}_{V \otimes W}$.

Proof. Obvious (check equality on elements $v \otimes w$). □

12.11. Lemma. *Let* $U \xrightarrow{f} V \xrightarrow{g} W$ *and* $U' \xrightarrow{f'} V' \xrightarrow{g'} W'$ *be linear maps. Then*

$$(g \otimes g') \circ (f \otimes f') = (g \circ f) \otimes (g' \circ f') .$$

Proof. Easy — check equality on $u \otimes u'$. □

12.12. Lemma. $\mathrm{Hom}(U, \mathrm{Hom}(V, W)) \cong \mathrm{Hom}(U \otimes V, W)$.

Proof. Let $f \in \mathrm{Hom}(U, \mathrm{Hom}(V, W))$ and define $\tilde{f}(u \otimes v) = \big(f(u)\big)(v)$ (note that $f(u) \in \mathrm{Hom}(V, W)$ is a linear map from V to W). Since $\big(f(u)\big)(v)$ is bilinear in u and v, this defines a linear map $\tilde{f} \in \mathrm{Hom}(U \otimes V, W)$. Conversely, given $\varphi \in \mathrm{Hom}(U \otimes V, W)$, define $\hat{\varphi}(u) \in \mathrm{Hom}(V, W)$ by $\big(\hat{\varphi}(u)\big)(v) = \varphi(u \otimes v)$. Then $\hat{\varphi}$ is a linear map from U to $\mathrm{Hom}(V, W)$, and the two linear(!) maps $f \mapsto \tilde{f}$ and $\varphi \mapsto \hat{\varphi}$ are inverses of each other. □

In the special case $W = F$, the statement of the lemma reads

$$\mathrm{Hom}(U, V^*) \cong \mathrm{Hom}(U \otimes V, F) = (U \otimes V)^* .$$

The following result is important, as it allows us to replace Hom spaces by tensor products (at least when the vector spaces involved are finite-dimensional).

12.13. Proposition. *Let V and W be two vector spaces. There is a natural linear map*

$$\phi : V^* \otimes W \longrightarrow \mathrm{Hom}(V,W)\,, \quad l \otimes w \longmapsto \big(v \mapsto l(v)w\big)\,,$$

which is an isomorphism when V or W is finite-dimensional.

Proof. We will give the proof here for the case that W is finite-dimensional, and leave the case "V finite-dimensional" for the exercises.

First we should check that ϕ is a well-defined linear map. By the general principle on maps from tensor products, we only need to check that $(l,w) \mapsto \big(v \mapsto l(v)w\big)$ is bilinear. Linearity in w is clear; linearity in l follows from the definition of the vector space structure on V^*:

$$(\alpha_1 l_1 + \alpha_2 l_2, w) \longmapsto \big(v \mapsto (\alpha_1 l_1 + \alpha_2 l_2)(v)\, w = \alpha_1 l_1(v)w + \alpha_2 l_2(v)w\big)$$

To show that ϕ is bijective when W is finite-dimensional, we choose a basis $w_1, \dots, w_n$ of W. Let $w_1^*, \dots, w_n^*$ be the basis for W^* dual to $w_1, \dots, w_n$. Define a map

$$\phi' : \mathrm{Hom}(V,W) \longrightarrow V^* \otimes W\,, \quad f \longmapsto \sum_{i=1}^n (w_i^* \circ f) \otimes w_i\,.$$

It is easy to see that ϕ' is linear. Let us check that ϕ and ϕ' are inverses. Recall that for all $w \in W$, we have

$$w = \sum_{i=1}^n w_i^*(w) w_i\,.$$

Now,

$$\begin{aligned}
\phi'\big(\phi(l \otimes w)\big) &= \sum_{i=1}^n \big(w_i^* \circ (v \mapsto l(v)w)\big) \otimes w_i \\
&= \sum_{i=1}^n (v \mapsto l(v) w_i^*(w)) \otimes w_i = \sum_{i=1}^n w_i^*(w) l \otimes w_i \\
&= l \otimes \sum_{i=1}^n w_i^*(w) w_i = l \otimes w\,.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\phi\big(\phi'(f)\big) &= \phi\Big(\sum_{i=1}^n (w_i^* \circ f) \otimes w_i\Big) = \sum_{i=1}^n \Big(v \mapsto w_i^*\big(f(v)\big) w_i\Big) \\
&= \Big(v \mapsto \sum_{i=1}^n w_i^*(f(v)) w_i\Big) = \big(v \mapsto f(v)\big) = f\,.
\end{aligned}$$

□

Now assume that $V = W$ is finite-dimensional. Then by the above,

$$\mathrm{Hom}(V,V) \cong V^* \otimes V$$

in a natural way. But $\mathrm{Hom}(V,V)$ contains a special element, namely id_V. What is the element of $V^* \otimes V$ that corresponds to it?

12.14. Remark. *Let $v_1, \dots, v_n$ be a basis for V, and let $v_1^*, \dots, v_n^*$ be the basis for V^* dual to it. Then, with ϕ the canonical map from above, we have*

$$\phi\Big(\sum_{i=1}^{n} v_i^* \otimes v_i\Big) = \mathrm{id}_V \ .$$

Proof. Apply ϕ' as defined above to id_V. □

On the other hand, there is a natural bilinear form on $V^* \times V$, given by evaluation: $(l, v) \mapsto l(v)$. This gives the following.

12.15. Lemma. *Let V be a finite-dimensional vector space. There is a linear form $T\colon V^* \otimes V \to F$ given by $T(l \otimes v) = l(v)$. It makes the following diagram commutative.*

$$\begin{array}{ccc} V^* \otimes V & \xrightarrow{\ \phi\ } & \mathrm{Hom}(V, V) \\ & {\scriptstyle T} \searrow \quad \swarrow {\scriptstyle \mathrm{Tr}} & \\ & F & \end{array}$$

Proof. That T is well-defined is clear by the usual principle. (The vector space structure on V^* is defined in order to make evaluation bilinear!) We have to check that the diagram commutes. Fix a basis $v_1, \dots, v_n$, with dual basis $v_1^*, \dots, v_n^*$, and let $f \in \mathrm{Hom}(V, V)$. Then $\phi^{-1}(f) = \sum_i (v_i^* \circ f) \otimes v_i$, hence $T(\phi^{-1}(f)) = \sum_i v_i^*(f(v_i))$. The terms in the sum are exactly the diagonal entries of the matrix A representing f with respect to $v_1, \dots, v_n$, so $T(\phi^{-1}(f)) = \mathrm{Tr}(A) = \mathrm{Tr}(f)$. □

The preceding operation is called "contraction". More generally, it leads to linear maps

$$U_1 \otimes \cdots \otimes U_m \otimes V^* \otimes V \otimes W_1 \otimes \cdots \otimes W_n \longrightarrow U_1 \otimes \cdots \otimes U_m \otimes W_1 \cdots \otimes W_n \, .$$

This in turn is used to define "inner multiplication"

$$(U_1 \otimes \cdots \otimes U_m \otimes V^*) \times (V \otimes W_1 \otimes \cdots \otimes W_n) \longrightarrow U_1 \otimes \cdots \otimes U_m \otimes W_1 \cdots \otimes W_n$$

(by first going to the tensor product). The roles of V and V^* can also be reversed. This is opposed to "outer multiplication", which is just the canonical bilinear map

$$(U_1 \otimes \cdots \otimes U_m) \times (W_1 \otimes \cdots \otimes W_n) \longrightarrow U_1 \otimes \cdots \otimes U_m \otimes W_1 \cdots \otimes W_n \, .$$

An important example of inner multiplication is composition of linear maps.

12.16. Lemma. *Let U, V, W be vector spaces. Then the following diagram commutes.*

$$\begin{array}{ccccc} (l \otimes v, l' \otimes w) & (U^* \otimes V) \times (V^* \otimes W) & \xrightarrow{\ \phi \times \phi\ } & \mathrm{Hom}(U, V) \times \mathrm{Hom}(V, W) & (f, g) \\ \big\downarrow & \big\downarrow & & \big\downarrow & \big\downarrow \\ l'(v)\, l \otimes w & U^* \otimes W & \xrightarrow{\ \phi\ } & \mathrm{Hom}(U, W) & g \circ f \end{array}$$

Proof. We have

$$\begin{aligned} \phi(l' \otimes w) \circ \phi(l \otimes v) &= \big(v' \mapsto l'(v')w\big) \circ \big(u \mapsto l(u)v\big) \\ &= \Big(u \mapsto l'\big(l(u)v\big)w = l'(v)l(u)w\Big) \\ &= \phi\big(l'(v)l \otimes w\big) \, . \end{aligned}$$

□

12.17. Remark. Identifying $\mathrm{Hom}(F^m, F^n)$ with the space $\mathrm{Mat}(n \times m, F)$ of $n \times m$-matrices over F, we see that matrix multiplication is a special case of inner multiplication of tensors.

12.18. Remark. Another example of inner multiplication is given by evaluation of linear maps: the following diagram commutes.

$$\begin{array}{ccccccc} (l \otimes w, v) & & (V^* \otimes W) \times V & \xrightarrow{\phi \times \mathrm{id}_V} & \mathrm{Hom}(V, W) \times V & & (f, v) \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ l(v)w & & W & =\!=\!= & W & & f(v) \end{array}$$

Complexification of Vector Spaces. Now let us turn to another use of the tensor product. There are situations when one has a real vector space, which one would like to turn into a complex vector space with "the same" basis. For example, suppose that $V_{\mathbb{R}}$ is a real vector space and $W_{\mathbb{C}}$ is a complex vector space (writing the field as a subscript to make it clear what scalars we are considering), then W can also be considered as a real vector space (just by restricting the scalar multiplication to $\mathbb{R} \subset \mathbb{C}$). We write $W_{\mathbb{R}}$ for this space. Note that $\dim_{\mathbb{R}} W_{\mathbb{R}} = 2 \dim_{\mathbb{C}} W_{\mathbb{C}}$ — if $b_1, \dots, b_n$ is a $\mathbb{C}$-basis for W, then $b_1, ib_1, \dots, b_n, ib_n$ is an $\mathbb{R}$-basis. Now we can consider an $\mathbb{R}$-linear map $f : V_{\mathbb{R}} \to W_{\mathbb{R}}$. Can we construct a $\mathbb{C}$-vector space $\tilde{V}_{\mathbb{C}}$ out of V in such a way that f extends to a $\mathbb{C}$-linear map $\tilde{f} : \tilde{V}_{\mathbb{C}} \to W_{\mathbb{C}}$? (Of course, for this to make sense, $V_{\mathbb{R}}$ has to sit in $\tilde{V}_{\mathbb{R}}$ as a subspace.)

It turns out that we can use the tensor product to do this.

12.19. Lemma and Definition. *Let V be a real vector space. The real vector space $\tilde{V} = \mathbb{C} \otimes_{\mathbb{R}} V$ can be given the structure of a complex vector space by defining scalar multiplication as follows.*

$$\lambda(\alpha \otimes v) = (\lambda\alpha) \otimes v$$

V is embedded into $\tilde{V}$ as a real subspace via $\iota : v \mapsto 1 \otimes v$.

This $\mathbb{C}$-vector space $\tilde{V}$ is called the *complexification* of V.

PROOF. We first have to check that the equation above leads to a well-defined $\mathbb{R}$-bilinear map $\mathbb{C} \times \tilde{V} \to \tilde{V}$. But this map is just

$$\mathbb{C} \times (\mathbb{C} \otimes_{\mathbb{R}} V) \longrightarrow \mathbb{C} \otimes_{\mathbb{R}} (\mathbb{C} \otimes_{\mathbb{R}} V) \cong (\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}) \otimes_{\mathbb{R}} V \stackrel{m \otimes \mathrm{id}_V}{\longrightarrow} \mathbb{C} \otimes_{\mathbb{R}} V\,,$$

where $m : \mathbb{C} \otimes_{\mathbb{R}} \mathbb{C} \to \mathbb{C}$ is induced from multiplication on $\mathbb{C}$ (which is certainly an $\mathbb{R}$-bilinear map). Since the map is in particular linear in the second argument, we also have the "distributive laws"

$$\lambda(x + y) = \lambda x + \lambda y\,, \qquad (\lambda + \mu)x = \lambda x + \mu x$$

for $\lambda, \mu \in \mathbb{C}$, $x, y \in \tilde{V}$. The "associative law"

$$\lambda(\mu x) = (\lambda\mu)x$$

(for $\lambda, \mu \in \mathbb{C}$, $x \in \tilde{V}$) then needs only to be checked for $x = \alpha \otimes v$, in which case we have

$$\lambda(\mu(\alpha \otimes v)) = \lambda((\mu\alpha) \otimes v) = (\lambda\mu\alpha) \otimes v = (\lambda\mu)(\alpha \otimes v)\,.$$

The last statement is clear. □

If we apply the representation of elements in a tensor product given in Lemma 12.6 to $\tilde{V}$, we obtain the following.

Suppose V has a basis $v_1, \ldots, v_n$. Then every element of $\tilde{V}$ can be written uniquely in the form

$$\alpha_1 \otimes v_1 + \cdots + \alpha_n \otimes v_n \qquad \text{for some } \alpha_1, \ldots, \alpha_n \in \mathbb{C}.$$

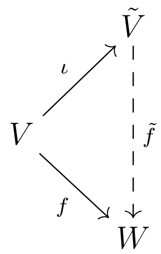
In this sense, we can consider $\tilde{V}$ to have “the same” basis as V, but we allow complex coordinates instead of real ones.

On the other hand, we can consider the basis $1, i$ of $\mathbb{C}$ as a real vector space, then we see that every element of $\tilde{V}$ can be written uniquely as

$$1 \otimes v + i \otimes v' = \iota(v) + i \cdot \iota(v') \qquad \text{for some } v, v' \in V.$$

In this sense, elements of $\tilde{V}$ have a real and an imaginary part, which live in V (identifying V with its image under ι in $\tilde{V}$).

12.20. Proposition. *Let V be a real vector space and W a complex vector space. Then for every $\mathbb{R}$-linear map $f : V_{\mathbb{R}} \to W_{\mathbb{R}}$, there is a unique $\mathbb{C}$-linear map $\tilde{f} : \tilde{V}_{\mathbb{C}} \to W_{\mathbb{C}}$ such that $\tilde{f} \circ \iota = f$ (where $\iota : V_{\mathbb{R}} \to \tilde{V}_{\mathbb{R}}$ is the map defined above).*



Proof. The map $\mathbb{C} \times V \to W$, $(\alpha, v) \mapsto \alpha f(v)$ is $\mathbb{R}$-bilinear. By the universal property of the tensor product $\tilde{V} = \mathbb{C} \otimes_{\mathbb{R}} V$, there is a unique $\mathbb{R}$-linear map $\tilde{f} : \tilde{V} \to W$ such that $\tilde{f}(\alpha \otimes v) = \alpha f(v)$. Then we have

$$\tilde{f}(\iota(v)) = \tilde{f}(1 \otimes v) = f(v)\,.$$

We have to check that $\tilde{f}$ is in fact $\mathbb{C}$-linear. It is certainly additive (being $\mathbb{R}$-linear), and for $\lambda \in \mathbb{C}$, $\alpha \otimes v \in \tilde{V}$,

$$\tilde{f}(\lambda(\alpha \otimes v)) = \tilde{f}((\lambda\alpha) \otimes v) = \lambda\alpha f(v) = \lambda \tilde{f}(\alpha \otimes v)\,.$$

Since any $\mathbb{C}$-linear map $\tilde{f}$ having the required property must be $\mathbb{R}$-linear and satisfy

$$\tilde{f}(\alpha \otimes v) = \tilde{f}(\alpha(1 \otimes v)) = \alpha \tilde{f}(1 \otimes v) = \alpha f(v)\,,$$

and since there is only one such map, $\tilde{f}$ is uniquely determined. □

12.21. Remark. The proposition can be stated in the form that

$$\mathrm{Hom}_{\mathbb{R}}(V, W) \xrightarrow{\cong} \mathrm{Hom}_{\mathbb{C}}(\tilde{V}, W)\,, \quad f \longmapsto \tilde{f}\,,$$

is an isomorphism. (The inverse is $F \mapsto F \circ \iota$.)

We also get that $\mathbb{R}$-linear maps between $\mathbb{R}$-vector spaces give rise to $\mathbb{C}$-linear maps between their complexifications.

12.22. Lemma. *Let $f : V \to W$ be an $\mathbb{R}$-linear map between two $\mathbb{R}$-vector spaces. Then $\mathrm{id}_{\mathbb{C}} \otimes f : \tilde{V} \to \tilde{W}$ is $\mathbb{C}$-linear, extends f, and is the only such map.*

PROOF. Consider the following diagram.

$$\begin{array}{ccc} V & \xrightarrow{\ f\ } & W \\ {\scriptstyle \iota_V}\downarrow & \overset{F}{\searrow} & \downarrow{\scriptstyle \iota_W} \\ \tilde{V} & \xrightarrow[\tilde{F}]{} & \tilde{W} \end{array}$$

Here, $F = \iota_W \circ f$ is an $\mathbb{R}$-linear map from V into the $\mathbb{C}$-vector space $\tilde{W}$, hence there is a unique $\mathbb{C}$-linear map $\tilde{F} : \tilde{V} \to \tilde{W}$ such that the diagram is commutative. We only have to verify that $\tilde{F} = \mathrm{id}_{\mathbb{C}} \otimes f$. But

$$(\mathrm{id}_{\mathbb{C}} \otimes f)(\alpha \otimes v) = \alpha \otimes f(v) = \alpha(1 \otimes f(v)) = \alpha(\iota_W \circ f)(v) = \alpha F(v) = \tilde{F}(\alpha \otimes v)\,.$$

□

13. Symmetric and Alternating Products

Note. The material in this section is not required for the final exam.

Now we want to generalize the tensor product construction (in a sense) in order to obtain similar results for symmetric and skew-symmetric (or alternating) bi- and multilinear maps.

13.1. Reminder. Let V and W be vector spaces. A bilinear map $f : V \times V \to W$ is called *symmetric* if $f(v, v') = f(v', v)$ for all $v, v' \in V$. f is called *alternating* if $f(v, v) = 0$ for all $v \in V$; this implies that f is *skew-symmetric*, i.e., $f(v, v') = -f(v', v)$ for all $v, v' \in V$. The converse is true if the field of scalars is not of characteristic 2.

Let us generalize these notions to multilinear maps.

13.2. Definition. Let V and W be vector spaces, and let $f : V^n \to W$ be a multilinear map.

(1) f is called *symmetric* if
$$f(v_{\sigma(1)}, v_{\sigma(2)}, \dots, v_{\sigma(n)}) = f(v_1, v_2, \dots, v_n)$$
for all $v_1, \dots, v_n \in V$ and all $\sigma \in S_n$.
The symmetric multilinear maps form a linear subspace of the space of all multilinear maps $V^n \to W$, denoted $\mathrm{Sym}(V^n, W)$.

(2) f is called *alternating* if
$$f(v_1, v_2, \dots, v_n) = 0$$
for all $v_1, \dots, v_n \in V$ such that $v_i = v_j$ for some $1 \le i < j \le n$.
The alternating multilinear maps form a linear subspace of the space of all multilinear maps $V^n \to W$, denoted $\mathrm{Alt}(V^n, W)$.

13.3. Remark. Since transpositions generate the symmetric group S_n, we have the following.

(1) f is symmetric if and only if it is a symmetric bilinear map in all pairs of variables, the other variables being fixed.
(2) f is alternating if and only if it is an alternating bilinear map in all pairs of variables, the other variables being fixed.
(3) Assume that the field of scalars has characteristic $\neq 2$. Then f is alternating if and only if
$$f(v_{\sigma(1)}, v_{\sigma(2)}, \dots, v_{\sigma(n)}) = \varepsilon(\sigma) f(v_1, v_2, \dots, v_n)$$
for all $v_1, \dots, v_n \in V$ and all $\sigma \in S_n$, where $\varepsilon(\sigma)$ is the sign of the permutation σ.

13.4. Example. We know from earlier that the determinant can be interpreted as an alternating multilinear map $V^n \to F$, where V is an n-dimensional vector space — consider the n vectors in V as the n columns in a matrix. Moreover, we had seen that up to scaling, the determinant is the only such map. This means that
$$\mathrm{Alt}(V^n, F) = F \det .$$

13.5. We have seen that we can express multilinear maps as elements of suitable tensor products: Assuming V and W to be finite-dimensional, a multilinear map $f : V^n \to W$ lives in
$$\mathrm{Hom}(V^{\otimes n}, W) \cong (V^*)^{\otimes n} \otimes W .$$
Fixing a basis $v_1, \dots, v_m$ of V and its dual basis $v_1^*, \dots, v_n^*$, any element of this tensor product can be written uniquely in the form
$$f = \sum_{i_1,\dots,i_n=1}^{m} v_{i_1}^* \otimes \cdots \otimes v_{i_n}^* \otimes w_{i_1,\dots,i_n}$$
with suitable $w_{i_1 \dots i_n} \in W$. How can we read off whether f is symmetric or alternating?

13.6. Definition. Let $x \in V^{\otimes n}$.

(1) x is called *symmetric* if $s_\sigma(x) = x$ for all $\sigma \in S_n$, where $s_\sigma : V^{\otimes n} \to V^{\otimes n}$ is the automorphism given by
$$s_\sigma(v_1 \otimes v_2 \otimes \cdots \otimes v_n) = v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \cdots \otimes v_{\sigma(n)} .$$
We will write $\mathrm{Sym}(V^{\otimes n})$ for the subspace of symmetric tensors.
(2) x is called *skew-symmetric* if $s_\sigma(x) = \varepsilon(\sigma)x$ for all $\sigma \in S_n$.
We will write $\mathrm{Alt}(V^{\otimes n})$ for the subspace of skew-symmetric tensors.

13.7. Proposition. *Let $f : V^n \to W$ be a multilinear map, identified with its image in $(V^*)^{\otimes n} \otimes W$. The following statements are equivalent.*

(1) *f is a symmetric multilinear map.*
(2) *$f \in (V^*)^{\otimes n} \otimes W$ lies in the subspace* $\mathrm{Sym}((V^*)^{\otimes n}) \otimes W$.
(3) *Fixing a basis as above in 13.5, in the representation of f as given there, we have*
$$w_{i_{\sigma(1)},\dots,i_{\sigma(n)}} = w_{i_1,\dots,i_n}$$
for all $\sigma \in S_n$.

Note that in the case $W = F$ and $n = 2$, the equivalence of (1) and (3) is just the well-known fact that symmetric matrices encode symmetric bilinear forms.

PROOF. Looking at (3), we have that $w_{i_1,\dots,i_n} = f(v_{i_1}, \dots, v_{i_n})$. So symmetry of f (statement (1)) certainly implies (3). Assuming (3), we see that f is a linear combination of terms of the form

$$\Big(\sum_{\sigma\in\mathfrak{S}_n} v^d_{i_\sigma(1)} \otimes \cdots \otimes v^d_{i_\sigma(n)}\Big) \otimes w$$

(with $w = w_{i_1,\dots,i_n}$), all of which are in the indicated subspace $\mathrm{Sym}((V^*)^{\otimes n}) \otimes W$ of $(V^*)^{\otimes n} \otimes W$, proving (2). Finally, assuming (2), we can assume $f = x \otimes w$ with $x \in \mathrm{Sym}((V^*)^{\otimes n})$ and $w \in W$. For $y \in V^{\otimes n}$ and $z \in (V^*)^{\otimes n} \cong (V^{\otimes n})^*$, we have $(s_\sigma(z))(s_\sigma(y)) = z(y)$. Since $s_\sigma(x) = x$, we get $x(s_\sigma(y)) = x(y)$ for all $\sigma \in S_n$, which implies that $f(s_\sigma(y)) = x(s_\sigma(y)) \otimes w = x(y) \otimes w = f(y)$. So f is symmetric. □

13.8. Proposition. *Let $f : V^n \to W$ be a multilinear map, identified with its image in $(V^*)^{\otimes n} \otimes W$. The following statements are equivalent.*

(1) *f is an alternating multilinear map.*
(2) *$f \in (V^*)^{\otimes n} \otimes W$ lies in the subspace $\mathrm{Alt}((V^*)^{\otimes n}) \otimes W$.*
(3) *Fixing a basis as above in 13.5, in the representation of f as given there, we have*

$$w_{i_{\sigma(1)},\dots,i_{\sigma(n)}} = \varepsilon(\sigma) w_{i_1,\dots,i_n}$$

for all $\sigma \in S_n$.

The proof is similar to the preceding one.

The equivalence of (2) and (3) in the propositions above, in the special case $W = F$ and replacing V^* by V, gives the following. (We assume that F is of characteristic zero, i.e., that $\mathbb{Q} \subset F$.)

13.9. Proposition. *Let V be an m-dimensional vector space with basis $v_1, \dots, v_m$.*

(1) *The elements*

$$\sum_{\sigma\in S_n} v_{i_{\sigma(1)}} \otimes \cdots \otimes v_{i_{\sigma(n)}}$$

for $1 \le i_1 \le i_2 \le \cdots \le i_n \le m$ form a basis for $\mathrm{Sym}(V^{\otimes n})$. In particular,

$$\dim \mathrm{Sym}(V^{\otimes n}) = \binom{m+n-1}{n}.$$

(2) *The elements*

$$\sum_{\sigma\in S_n} \varepsilon(\sigma) v_{i_{\sigma(1)}} \otimes \cdots \otimes v_{i_{\sigma(n)}}$$

for $1 \le i_1 < i_2 < \cdots < i_n \le m$ form a basis for $\mathrm{Alt}(V^{\otimes n})$. In particular,

$$\dim \mathrm{Alt}(V^{\otimes n}) = \binom{m}{n}.$$

PROOF. It is clear that the given elements span the spaces. They are linearly independent since no two of them involve the same basis elements of $V^{\otimes n}$. (In the alternating case, note that the element given above vanishes if two of the i_j are equal.) □

The upshot of this is that (taking $W = F$ for simplicity) we have identified

$$\operatorname{Sym}(V^n, F) = \operatorname{Sym}((V^*)^{\otimes n}) \subset (V^*)^{\otimes n} = (V^{\otimes n})^*$$

and

$$\operatorname{Alt}(V^n, F) = \operatorname{Alt}((V^*)^{\otimes n}) \subset (V^*)^{\otimes n} = (V^{\otimes n})^*$$

as subspaces of $(V^{\otimes n})^*$. But what we would like to have are spaces $\operatorname{Sym}^n(V)$ and $\operatorname{Alt}^n(V)$ such that we get identifications

$$\operatorname{Sym}(V^n, F) = \operatorname{Hom}(\operatorname{Sym}^n(V), F) = (\operatorname{Sym}^n(V))^*$$

and

$$\operatorname{Alt}(V^n, F) = \operatorname{Hom}(\operatorname{Alt}^n(V), F) = (\operatorname{Alt}^n(V))^* .$$

Now there is a general principle that says that subspaces are "dual" to quotient spaces: If W is a subspace of V, then W^* is a quotient space of V^* in a natural way, and if W is a quotient of V, then W^* is a subspace of V^* in a natural way. So in order to translate the subspace $\operatorname{Sym}(V^n, F)$ (or $\operatorname{Alt}(V^n, F)$) of the dual space of $V^{\otimes n}$ into the dual space of something, we should look for a suitable *quotient* of $V^{\otimes n}$!

13.10. Definition. Let V be a vector space, $n > 0$ an integer.

(1) Let $W \subset V^{\otimes n}$ be the subspace spanned by all elements of the form

$$v_1 \otimes v_2 \otimes \cdots \otimes v_n - v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \cdots \otimes v_{\sigma(n)}$$

for $v_1, v_2, \ldots, v_n \in V$ and $\sigma \in S_n$. Then the quotient space

$$\operatorname{Sym}^n(V) = S^n(V) = V^{\otimes n}/W$$

is called the *nth symmetric tensor power* of V. The image of $v_1 \otimes v_2 \otimes \cdots \otimes v_n$ in $S^n(V)$ is denoted $v_1 \cdot v_2 \cdots v_n$.

(2) Let $W \subset V^{\otimes n}$ be the subspace spanned by all elements of the form

$$v_1 \otimes v_2 \otimes \cdots \otimes v_n$$

for $v_1, v_2, \ldots, v_n \in V$ such that $v_i = v_j$ for some $1 \le i < j \le n$. Then the quotient space

$$\operatorname{Alt}^n(V) = \textstyle\bigwedge^n(V) = V^{\otimes n}/W$$

is called the *nth alternating tensor power* of V. The image of $v_1 \otimes v_2 \otimes \cdots \otimes v_n$ in $\bigwedge^n(V)$ is denoted $v_1 \wedge v_2 \wedge \cdots \wedge v_n$.

13.11. Theorem.

(1) *The map*

$$\varphi : V^n \longrightarrow S^n(V), \quad (v_1, v_2, \ldots, v_n) \longmapsto v_1 \cdot v_2 \cdots v_n$$

is multilinear and symmetric. For every multilinear and symmetric map $f : V^n \to U$, there is a unique linear map $g : S^n(V) \to U$ such that $f = g \circ \varphi$.

(2) *The map*

$$\psi : V^n \longrightarrow \textstyle\bigwedge^n(V), \quad (v_1, v_2, \ldots, v_n) \longmapsto v_1 \wedge v_2 \wedge \cdots \wedge v_n$$

is multilinear and alternating. For every multilinear and alternating map $f : V^n \to U$, there is a unique linear map $g : \bigwedge^n(V) \to U$ such that $f = g \circ \psi$.

These statements tell us that the spaces we have defined do what we want: We get identifications

$$\operatorname{Sym}(V^n, U) = \operatorname{Hom}(S^n(V), U) \quad \text{and} \quad \operatorname{Alt}(V^n, U) = \operatorname{Hom}(\textstyle\bigwedge^n(V), U)\,.$$

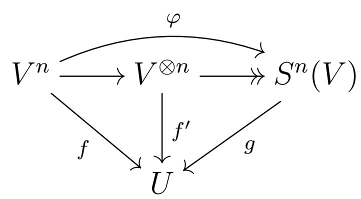
Proof. We prove the first part; the proof of the second part is analogous. First, it is clear that φ is multilinear: it is the composition of the multilinear map $(v_1, \ldots, v_n) \mapsto v_1 \otimes \cdots \otimes v_n$ and the linear projection map from $V^{\otimes n}$ to $S^n(V)$. We have to check that φ is symmetric. But

$$\varphi(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) - \varphi(v_1, \ldots, v_n) = v_{\sigma(1)} \cdots v_{\sigma(n)} - v_1 \cdots v_n = 0\,,$$

since it is the image in $S^n(V)$ of $v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(n)} - v_1 \otimes \cdots \otimes v_n \in W$. Now let $f : V^n \to U$ be multilinear and symmetric. Then there is a unique linear map $f' : V^{\otimes n} \to U$ corresponding to f, and by symmetry of f, we have

$$f'(v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(n)} - v_1 \otimes \cdots \otimes v_n) = 0\,.$$

So f' vanishes on all the elements of a spanning set of W. Hence it vanishes on W and therefore induces a unique linear map $g : S^n(V) = V^{\otimes n}/W \to U$.



□

The two spaces $\operatorname{Sym}(V^{\otimes n})$ and $S^n(V)$ (resp., $\operatorname{Alt}(V^{\otimes n})$ and $\bigwedge^n(V)$) are closely related. We assume that F is of characteristic zero.

13.12. Proposition.

(1) *The maps* $\operatorname{Sym}(V^{\otimes n}) \subset V^{\otimes n} \to S^n(V)$ *and*

$$S^n(V) \longrightarrow \operatorname{Sym}(V^{\otimes n})\,, \quad v_1 \cdot v_2 \cdots v_n \longmapsto \frac{1}{n!} \sum_{\sigma \in S_n} v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \cdots \otimes v_{\sigma(n)}$$

are inverse isomorphisms. In particular, if $b_1, \ldots, b_m$ *is a basis for* V*, then the elements*

$$b_{i_1} \cdot b_{i_2} \cdots b_{i_n} \qquad \text{with } 1 \le i_1 \le i_2 \le \cdots \le i_n \le m$$

form a basis for $S^n(V)$*, and* $\dim S^n(V) = \binom{m+n-1}{n}$.

(2) *The maps* $\operatorname{Alt}(V^{\otimes n}) \subset V^{\otimes n} \to \bigwedge^n(V)$ *and*

$$\textstyle\bigwedge^n(V) \longrightarrow \operatorname{Alt}(V^{\otimes n})\,, \quad v_1 \wedge v_2 \wedge \cdots \wedge v_n \longmapsto \displaystyle\frac{1}{n!} \sum_{\sigma \in S_n} \operatorname{sign}(\sigma) v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \cdots \otimes v_{\sigma(n)}$$

are inverse isomorphisms. In particular, if $b_1, \ldots, b_m$ *is a basis for* V*, then the elements*

$$b_{i_1} \wedge b_{i_2} \wedge \cdots \wedge b_{i_n} \qquad \text{with } 1 \le i_1 < i_2 < \cdots < i_n \le m$$

form a basis for $\bigwedge^n(V)$*, and* $\dim \bigwedge^n(V) = \binom{m}{n}$.

Proof. It is easy to check that the specified maps are well-defined linear maps and inverses of each other, so they are isomorphisms. The other statements then follow from the description in Prop. 13.9. □

Note that if $\dim V = n$, then we have

$$\bigwedge^n(V) = F\,(v_1 \wedge \cdots \wedge v_n)$$

for any basis $v_1, \ldots, v_n$ of V.

13.13. Corollary. *Let $v_1, \ldots, v_n \in V$. Then $v_1, \ldots, v_n$ are linearly independent if and only if $v_1 \wedge \cdots \wedge v_n \neq 0$.*

Proof. If $v_1, \ldots, v_n$ are linearly dependent, then we can express one of them, say v_n, as a linear combination of the others:

$$v_n = \lambda_1 v_1 + \cdots + \lambda_{n-1} v_{n-1}\,.$$

Then

$$\begin{aligned} v_1 \wedge \cdots \wedge v_{n-1} \wedge v_n &= v_1 \wedge \cdots \wedge v_{n-1} \wedge (\lambda_1 v_1 + \cdots + \lambda_{n-1} v_{n-1}) \\ &= \lambda_1 (v_1 \wedge \cdots \wedge v_{n-1} \wedge v_1) + \cdots + \lambda_{n-1} (v_1 \wedge \cdots \wedge v_{n-1} \wedge v_{n-1}) \\ &= 0 + \cdots + 0 = 0\,. \end{aligned}$$

On the other hand, when $v_1, \ldots, v_n$ are linearly independent, they form part of a basis $v_1, \ldots, v_n, \ldots, v_m$, and by Prop. 13.12, $v_1 \wedge \cdots \wedge v_n$ is a basis element of $\bigwedge^n(V)$, hence nonzero. □

13.14. Lemma and Definition. *Let $f : V \to W$ be linear. Then f induces linear maps $S^n(f) : S^n(V) \to S^n(W)$ and $\bigwedge^n(f) : \bigwedge^n(V) \to \bigwedge^n(W)$ satisfying*

$$S^n(f)(v_1 \cdots v_n) = f(v_1) \cdots f(v_n)\,, \qquad \bigwedge^n(f)(v_1 \wedge \cdots \wedge v_n) = f(v_1) \wedge \cdots \wedge f(v_n)\,.$$

Proof. The map $V^n \to S^n(W)$, $(v_1, \ldots, v_n) \mapsto f(v_1) \cdots f(v_n)$, is a symmetric multilinear map and therefore determines a unique linear map $S^n(f) : S^n(V) \to S^n(W)$ with the given property. Similarly for $\bigwedge^n(f)$. □

13.15. Proposition. *Let $f : V \to V$ be a linear map, with V an n-dimensional vector space. Then $\bigwedge^n(f) : \bigwedge^n(V) \to \bigwedge^n(V)$ is multiplication by $\det(f)$.*

Proof. Since $\bigwedge^n(V)$ is a one-dimensional vector space, $\bigwedge^n(f)$ must be multiplication by a scalar. We pick a basis $v_1, \ldots, v_n$ of V and represent f by a matrix A with respect to this basis. The scalar in question is the element $\delta \in F$ such that

$$f(v_1) \wedge f(v_2) \wedge \cdots \wedge f(v_n) = \delta\,(v_1 \wedge v_2 \wedge \cdots \wedge v_n)\,.$$

The vectors $f(v_1)$, …, $f(v_n)$ correspond to the columns of the matrix A, and δ is an alternating multilinear form on them. Hence δ must be $\det(A)$, up to a scalar factor. Taking f to be id_V, we see that the scalar factor is 1. □

13.16. Corollary. *Let V be a finite-dimensional vector space, $f, g : V \to V$ two endomorphisms. Then $\det(g \circ f) = \det(g)\det(f)$.*

Proof. Let $n = \dim V$. We have $\bigwedge^n(g \circ f) = \bigwedge^n g \circ \bigwedge^n f$, and the map on the left is multiplication by $\det(g \circ f)$, whereas the map on the right is multiplication by $\det(g)\det(f)$. □

We see that, similarly to the trace $\mathrm{Hom}(V, V) \cong V^* \otimes V \to F$, our constructions give us a natural (coordinate-free) definition of the determinant of an endomorphism.

Bibliography

[BR1] T.S. Blyth and E.F. Robertson: *Basic Linear Algebra.* Springer Undergraduate Mathematics Series, 2002.

[BR2] T.S. Blyth and E.F. Robertson: *Further Linear Algebra.* Springer Undergraduate Mathematics Series, 2002.

[J] K. Jänich: *Linear Algebra.* Springer Undergraduate Texts in Mathematics, 1994

[KM] A. Kostrykin and Y. Manin: *Linear Algebra and Geometry.* Gordan and Breach, 1988.

Index of notation

Index