Linear Algebra II

Course No. 100 222

Spring 2007

Michael Stoll

Contents

1. Review of Eigenvalues, Eigenvectors and Characteristic Polynomial 2
2. The Cayley-Hamilton Theorem and the Minimal Polynomial 2
3. The Structure of Nilpotent Endomorphisms 7
4. Direct Sums of Subspaces 9
5. The Jordan Normal Form Theorem 11
6. The Dual Vector Space 16
7. Norms on Real Vector Spaces 22
8. Bilinear Forms 24
9. Inner Product Spaces 32
10. Orthogonal Diagonalization 39
11. External Direct Sums 43
12. The Tensor Product 44
13. Symmetric and Alternating Products 53

References 59

1. Review of Eigenvalues, Eigenvectors and Characteristic Polynomial

Recall the topics we finished *Linear Algebra I* with. We were discussing eigenvalues and eigenvectors of endomorphisms and square matrices, and the question when they are *diagonalizable*. For your convenience, I will 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

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

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 of 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.

2. The Cayley-Hamilton Theorem and the Minimal Polynomial

Let $A \in \operatorname{Mat}(n, F)$. We know that $\operatorname{Mat}(n, F)$ 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, 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 $D = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. (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).

2.1. **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

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

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

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

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

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

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

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

□

2.2. **Remarks.**

(1) The reason why we cannot simply plug in A for x in $\tilde{B}(xI - A) = P_A(x)I$ 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. We will discuss this in more detail in the *Introductory Algebra* course, where polynomial rings will be studied in some detail.

(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.

2.3. **Lemma (Polynomial Division).** *Let f and g be polynomials, with g monic. Then there are unique polynomials q and r 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 $\deg(f) < \deg(g)$, 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 $g = x^n + \ldots$) $\deg(f') < \deg(f)$. 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 Theorem 17.13 of *Linear Algebra I*.

2.4. **Lemma and Definition.** *Let $A \in \mathrm{Mat}(n, F)$. There is a unique monic polynomial $M_A(x)$ of minimal degree such that $M_A(A) = 0$. If $p(x)$ is any polynomial 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 $p(A) = 0$ exist (by the Cayley-Hamilton Theorem 2.1, we can take $p = P_A$). So there will be one such 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 2.3, there are polynomials q and r, $\deg(r) < \deg(M_A)$ 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 $p = qM_A$. □

2.5. **Remark.** In *Introductory Algebra,* you will learn that the set of polynomials as discussed in the lemma forms an *ideal* and that the polynomial 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.

By Lemma 2.4, 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.

2.6. **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 2.4, 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. □

2.7. **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.*

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

2.8. **Proposition.** *Let $A \in \mathrm{Mat}(n, F)$. 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), 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. We write $M_A(x) = (x - \lambda_j)p_j(x)$ for polynomials p_j. I claim that

$$E_{\lambda_j}(A) = \mathrm{im}\bigl(p_j(A)\bigr)\,.$$

To show that $\operatorname{im}\bigl(p_j(A)\bigr) \subset E_{\lambda_j}(A)$, consider

$$0 = M_A(A)v = (A - \lambda_j I)p_j(A)v\,.$$

If $w \in \operatorname{im}\bigl(p_j(A)\bigr)$, then $w = p_j(A)v$ for some $v \in F^n$, hence

$$Aw - \lambda_j w = (A - \lambda_j I)w = 0\,.$$

To show that $E_{\lambda_j}(A) \subset \operatorname{im}\bigl(p_j(A)\bigr)$, let $v \in E_{\lambda_j}(A)$, so that $Av = \lambda_j v$. Then

$$p_j(A)v = p_j(\lambda_j)v\,,$$

and $p_j(\lambda_j) \neq 0$, so

$$v = p_j(A)\bigl(p_j(\lambda_j)^{-1}v\bigr) \in \operatorname{im}\bigl(p_j(A)\bigr)\,.$$

Finally, I claim that $\sum_{j=1}^m \operatorname{im}\bigl(p_j(A)\bigr) = F^n$. To see this, note that the polynomial

$$p(x) = \sum_{j=1}^m p_j(\lambda_j)^{-1}p_j(x)$$

has degree less than m and assumes the value 1 at the m distinct elements $\lambda_1, \ldots,$ λ_m of F. There is only one such polynomial, and this is $p(x) = 1$. So for $v \in F^n$, we have

$$v = Iv = p(A)v = \sum_{j=1}^m p_j(\lambda_j)^{-1}p_j(A)v \in \sum_{j=1}^m \operatorname{im}\bigl(p_j(A)\bigr)\,.$$

Both claims together imply that

$$\sum_{j=1}^m \dim E_{\lambda_j}(A) \geq n\,,$$

which in turn (by Cor. 17.10 of *Linear Algebra I*) implies that A is diagonalizable. □

2.9. **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 $A - I \neq 0$, $m > 1$ (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 $M_B(x) = P_B(x) = (x-1)(x-4)(x-6)$; B therefore is diagonalizable.

Exercise: what happens for fields F of small characteristic?

3. The Structure of Nilpotent Endomorphisms

3.1. **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 $f^m = \underbrace{f \circ f \circ \cdots \circ f}_{m \text{ times}} = 0$ for some $m \geq 1$.

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.

3.2. **Corollary.** *A nilpotent matrix or endomorphism is diagonalizable if and only if it is zero.*

Proof. The minimal polynomial is x^m. Prop. 2.6 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. □

The following result tells us more about the structure of nilpotent endomorphisms.

3.3. **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 $M_f(x) = x^m$. We will do induction on m. When $m = 0$ (then $V = \{0\}$) or $m = 1$, the claim is trivial, so we assume $m \geq 2$. Then $\{0\} \subsetneq \ker(f) \subsetneq V$, and we can consider the quotient space $W = V/\ker(f)$. Then f induces an endomorphism $\tilde{f} : W \to W$ that satisfies $\tilde{f}^{m-1} = 0$. (Note that $f^{m-1}(V) \subset \ker(f)$.) So by induction, there is a basis $w_1, \ldots, w_k$ of W such that $\tilde{f}(w_j) = 0$ or w_{j+1}. We can then write the basis in the form

$$(w_1, \ldots, w_k) = \bigl(w'_1, \tilde{f}(w'_1), \ldots, \tilde{f}^{e_1}(w'_1), w'_2, \tilde{f}(w'_2), \ldots, \tilde{f}^{e_2}(w'_2), \ldots, w'_\ell, \ldots, \tilde{f}^{e_\ell}(w'_\ell)\bigr)$$

where $\tilde{f}^{e_j+1}(w'_j) = 0$ for $j = 1, \ldots, \ell$. We now lift w'_j to an element $v'_j \in V$ (i.e., we pick $v'_j \in V$ that maps to w'_j under the canonical epimorphism $V \to W$). Then

$$v'_1, f(v'_1), \ldots, f^{e_1}(v'_1), v'_2, f(v'_2), \ldots, f^{e_2}(v'_2), \ldots, v'_\ell, f(v'_\ell), \ldots, f^{e_\ell}(v'_\ell)$$

are linearly independent in V (since their images in W are linearly independent; note that the image of $f^i(v'_j)$ is $\tilde{f}^i(w'_j)$). We must have $f^{e_j+1}(v'_j) \in \ker(f)$ (since its image in W is zero). Note that their linear hull L is a complementary subspace of $\ker(f)$. I claim that the extended sequence

$$v'_1, f(v'_1), \ldots, f^{e_1+1}(v'_1), v'_2, f(v'_2), \ldots, f^{e_2+1}(v'_2), \ldots, v'_\ell, f(v'_\ell), \ldots, f^{e_\ell+1}(v'_\ell)$$

is still linearly independent. So assume we have a linear combination that vanishes. Then we get an equality of two vectors, one a linear combination of what we had before, the other a linear combination of the new vectors. So the first vector is in L, the second in $\ker(f)$, but $L \cap \ker(f) = \{0\}$, so both vectors have to vanish. Since the vectors we had previously are linearly independent, all their coefficients must vanish. It remains to show that the new vectors are linearly independent as well. So assume that we have

$$\lambda_1 f^{e_1+1}(v'_1) + \lambda_2 f^{e_2+1}(v'_2) + \cdots + \lambda_\ell f^{e_\ell+1}(v'_\ell) = 0\,.$$

This implies that

$$f\bigl(f^{e_1}(v'_1) + \lambda_2 f^{e_2}(v'_2) + \cdots + \lambda_\ell f^{e_\ell}(v'_\ell)\bigr) = 0\,,$$

so $f^{e_1}(v'_1) + \lambda_2 f^{e_2}(v'_2) + \cdots + \lambda_\ell f^{e_\ell}(v'_\ell) \in \ker(f)$. But this vector is also in L, hence it must be zero, and since the vectors involved in this linear combination are linearly independent (they are part of a basis of L), all the λ_j must vanish.

Finally, pick a basis $v'_{\ell+1}, \ldots, v'_{n-k}$ of a complementary subspace of the linear hull of $f^{e_1+1}(v'_1), \ldots, f^{e_\ell+1}(v'_\ell)$ in $\ker(f)$. Then

$$(v_1, \ldots, v_n) = \bigl(v'_1, f(v'_1), \ldots, f^{e_1+1}(v'_1), \ldots, v'_\ell, f(v'_\ell), \ldots, f^{e_\ell+1}(v'_\ell), v'_{\ell+1}, \ldots, v'_{n-k}\bigr)$$

is a basis of V with the required properties. □

3.4. **Remark.** The matrix $A = (a_{ij})$ representing f with respect to $v_n, \ldots, v_2, v_1$, where $v_1, \ldots, v_n$ is a basis as in Thm. 3.3 above, 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*

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

with blocks of the form

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

3.5. **Corollary.** *A nilpotent matrix is similar to a matrix of the form just described.*

Proof. This is clear from our discussion. □

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

Proof. If $P_A(x) = x^n$, then $A^n = 0$ by the Cayley-Hamilton Theorem 2.1, 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. □

3.7. **Remark.** The statement of Cor. 3.6 would also follow from the fact that $P_A(x)$ divides some power of $M_A(x) = x^m$, see Remark 2.7. 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).

3.8. **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 Thm. 3.3. The first step in the process comes down to finding a complementary subspace of

$\ker(A^2) = L\big((2,0,1)^\top, (0,2,1)^\top\big)$. We can take $(1,0,0)^\top$, for example, as the basis of a complement. This will be v_1' in the notation of the proof above. We then have $Av_1' = (3,1,2)^\top$ and $A^2 v_1' = (-1,-1,-1)^\top$, and these three form a basis. Reversing the order, we get

$$\begin{pmatrix} -1 & 3 & 1 \\ -1 & 1 & 0 \\ -1 & 2 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 3 & 4 & -7 \\ 1 & 2 & -3 \\ 2 & 3 & -5 \end{pmatrix} \begin{pmatrix} -1 & 3 & 1 \\ -1 & 1 & 0 \\ -1 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

4. Direct Sums of Subspaces

The proof of the Jordan Normal Form Theorem, which is our next goal, 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.

4.1. **Lemma and Definition.** *Let V be a vector space, $U_i \subset V$ (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 B_i is a basis of U_i, for $i \in I$, then the B_i are pairwise disjoint, and $\bigcup_{i\in I} B_i$ is a basis of V.*

If these conditions are satisfied, we say that V *is the direct sum of the subspaces U_i* and write $V = \bigoplus_{i\in I} U_i$. If $I = \{1, 2, \ldots, n\}$, we also write $V = U_1 \oplus U_2 \oplus \cdots \oplus U_n$.

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)": Let $B = \bigcap_{i\in I} B_i$, considered as a 'multiset' (i.e., such that elements can occur several times). Since B_i generates U_i and $\sum_i U_i = V$, 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 of 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) $\Rightarrow$ (1)": Write $v \in V$ as a linear combination of the basis elements in $\bigcup_i B_i$. Since B_i is a basis of U_i, this shows that $v = \sum_i u_i$ with $u_i \in U_i$. Since there is only one way of writing v as such a linear combination, all the u_i are uniquely determined. □

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

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

4.3. **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. □

4.4. **Remark.** If in the situation just described V is finite-dimensional and we choose a basis of V that is the union of bases of the U_i, then the matrix representing f relative to that basis will be a block diagonal matrix, where the diagonal blocks are the matrices representing the f_i relative to the bases of the U_i.

4.5. **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 endomorphims $f_i : U_i \to U_i$ 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 $f = \bigoplus_i f_i$. □

We now come to a relation between splittings of f as a direct sum and the characteristic or minimal polynomial of f.

4.6. **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, *i.e., there are polynomials $a_1(x)$ and $a_2(x)$ such that $a_1(x)p_1(x) + a_2(x)p_2(x) = 1$. 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 $f_i = f|_{U_i}$.*

Proof. We first show that

$$\operatorname{im}\big(p_2(f)\big) \subset U_1 = \ker\big(p_1(f)\big) \quad \text{and} \quad \operatorname{im}\big(p_1(f)\big) \subset U_2 = \ker\big(p_2(f)\big)\,.$$

Let $v \in \operatorname{im}\big(p_2(f)\big)$, so $v = \big(p_2(f)\big)(u)$ for some $u \in U$. Then

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

so $\operatorname{im}\big(p_2(f)\big) \subset \ker\big(p_1(f)\big)$; the other statement is proved in the same way.

Now we show that $U_1 \cap U_2 = \{0\}$. So let $v \in U_1 \cap U_2$. Then $\big(p_1(f)\big)(v) = \big(p_2(f)\big)(v) = 0$. 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 $\operatorname{im}\big(p_1(f)\big) + \operatorname{im}\big(p_2(f)\big) = V$. Using the same relation above, 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 $\operatorname{im}\big(p_1(f)\big) = U_2$, $\operatorname{im}\big(p_2(f)\big) = U_1$ and $V = U_1 \oplus U_2$: Let $v \in U_1$. We can write $v = v_1 + v_2$ with $v_i \in \operatorname{im}\big(p_i(f)\big)$. Then $U_1 \ni v - v_1 = v_2 \in U_2$, but $U_1 \cap U_2 = \{0\}$, so $v = v_1 \in \operatorname{im}\big(p_1(f)\big)$.

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

4.7. **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 4.6 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 $f'_{U_j} = f|_{U_j}$, so that we obtain the desired conclusion. □

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.

But first we need a little lemma about polynomials.

5.1. **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 $m \geq 1$.*

Proof. First, 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, taking the mth power on both sides, we obtain an equation

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

□

Now we can feed this into Prop. 4.7.

5.2. **Theorem.** *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 $V = U_1 \oplus \cdots \oplus U_k$, where $U_j = \ker(f - \lambda\,\mathrm{id}_V)^{m_j}$ is the generalized λ_j-eigenspace *of f.*

Proof. Write $P_f(x) = p_1(x)\cdots p_k(x)$ with $p_j(x) = (x-\lambda_j)^{m_j}$. By Lemma 5.1, we know that the $p_j(x)$ are coprime in pairs. By the Cayley-Hamilton Theorem 2.1, we know that $P_f(f) = 0$. The result then follows from Prop. 4.7. □

5.3. **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 of V such that the matrix representing f with respect to that basis is a block diagonal matrix with blocks of the form

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

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

Proof. We keep the notations of Thm. 5.2. We know that on U_j, $(f - \lambda_j\,\mathrm{id})^{m_j} = 0$, so $f|_{U_j} = \lambda_j\,\mathrm{id}_{U_j} + g_j$, where $g_j^{m_j} = 0$, i.e., g_j is nilpotent. By Thm. 3.3, there is a basis of U_j such that g_j is represented by a block diagonal matrix B_j with blocks of the form $B(0, m)$ (such that the sum of the m's is m_j). Therefore, $f|_{U_j}$ is represented by $B_j + \lambda_j I_{\dim U_j}$, which is a block diagonal matrix composed of blocks $B(\lambda_j, m)$ (with the same m's as before). The basis of V that is given by the union of the various bases of the U_j then does what we want, compare Remark 4.4. □

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

5.4. **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 Thm. 5.3 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. □

5.5. **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 generalized eigenspaces. We have a simple eigenvalue -1; the eigenspace is

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

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 generalized eigenspace:

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

So we can take as our basis

$$(f - 2\,\mathrm{id}_V)\big((1,1,0)^\top\big) = (-1,-2,-4)^\top, (1,1,0)^\top, (1,-1,1)^\top,$$

and we find

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

5.6. **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 exponental 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

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

5.7. **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 $F = \mathbb{R}$.

5.8. **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 of 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 $x^2 - 2\lambda x + \lambda^2 + \mu^2$.

Proof. Here is a sketch that gives the main ideas. Over $\mathbb{C}$, the characteristic polynomial 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 $v_1, \ldots, v_m$ is a basis of the generalized eigenspace (over $\mathbb{C}$) for the eigenvalue $\lambda + \mu i$, then $\bar{v}_1, \ldots, \bar{v}_m$ is a basis of the generalized 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

$$\tfrac{1}{2}(v_1 + \bar{v}_1), \tfrac{1}{2i}(v_1 - \bar{v}_1), \ldots, \tfrac{1}{2}(v_m + \bar{v}_m), \tfrac{1}{2i}(v_m - \bar{v}_m) ,$$

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

5.9. **Theorem.** *Let V be a finite-dimensional vector space, $f_1, \ldots, f_k : V \to V$ diagonalizable endomorphisms that commute in pairs. Then $f_1, \ldots, f_k$ are* simultaneously diagonalizable*, i.e., there is a basis of 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 $k \geq 2$. 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 $j = 1, \ldots, k - 1$,

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

By the induction hypothesis, $f_1^{(i)}, \ldots, f_{k-1}^{(i)}$ are simultaneously diagonalizable on U_i, for each i. Since U_i consists of eigenvectors of f_k, any basis of U_i that consists of eigenvectors of all the f_j, $j < k$, will also consist of eigenvectors for all the f_j, $j \leq k$. To get a suitable basis of V, we take the union of the bases of the various U_i. □

To finish this section, here is a uniqueness statement related to Cor. 5.4.

5.10. **Theorem.** *The diagonalizable and nilpotent parts of f in Cor. 5.4 are uniquely determined.*

Proof. Let $f = d + n = d' + n'$, where d and n are constructed as in the Jordan Normal Form Theorem 5.3 and $d \circ n = n \circ d$, $d' \circ n' = n' \circ d'$. Then d' and n' commute with f ($d' \circ f = d' \circ d' + d' \circ n' = d' \circ d' + n' \circ d' = f \circ d'$, 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

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

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 generalized 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 Thm. 5.9, the left hand side is diagonalizable, so we can assume it is represented by a diagonal matrix. But the only nilpotent diagonal matrix is the zero matrix, therefore $d - d' = n' - n = 0$, i.e., $d' = d$ and $n' = n$. □

6. The Dual Vector Space

6.1. **Defintion.** 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 $V^* = \mathrm{Hom}(V, F)$, the vector space of all linear forms on V.

Recall how the vector space structure on $V^* = \mathrm{Hom}(V, F)$ 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:

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

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

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

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 $v^* \in V^*$, we have

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

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 $e_1, \dots, e_n$. Then the dual basis is $p_1, \dots, p_n$.

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 Cor. 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 of 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 $b_S^* \in V^*$ such that $b_S^*(b) = 1$ for $b \in S$ and $b_S^*(b) = 0$ for $b \in B \setminus S$. Now there are certainly linear relations between the b_S^*, but one can show that no subset of $\{b_S^* : S \subset B\}$ whose cardinality is that of B can generate all the b_S^*. Therefore any basis of V^* must be of stricltly larger cardinality than B.

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. Then there is a canonical injective homomorphism $\alpha_V : V \to V^{**}$ of V into its* bidual $V^{**} = (V^*)^*$. *It is an isomorphism when V is finite-dimensional.*

Proof. In order to construct α_V, we have to associate to each $v \in V$ a linear form on V^*. More or less the only thing we can do with a linear form is to evaluate it on elements of V. Therefore we set

$$\alpha_V(v) = (V^* \ni \phi \mapsto \phi(v) \in F) .$$

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 $0 \neq v \in V$ such that $\alpha_V(v) = 0$. We can choose a basis of 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 $\alpha_V(v)$ cannot be zero, and we get the desired contradiction.

Finally, if V is finite-dimensional, then by Cor. 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 $v_1^*, \ldots, v_n^*$ be a basis of V^*. Then there is a unique basis $v_1, \ldots, v_n$ of V such that $v_i^*(v_j) = \delta_{ij}$.*

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

$$v_i^*(v_j) = \alpha_V(v_j)(v_i^*) = v_j^{**}(v_i^*) = \delta_{ij}\,.$$

□

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 $\alpha_1, \ldots, \alpha_n \in F$ distinct. Then $\mathrm{ev}_{\alpha_1}, \ldots, \mathrm{ev}_{\alpha_n} \in V^*$ 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 of V dual to that? What we need are polynomials $p_1, \ldots, p_n$ 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

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

A diagram clarifies perhaps what is happening here.

$$V \xrightarrow{f} W \xrightarrow{w^*} F$$

The composition $w^* \circ f$ 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 $w_1^*, w_2^* \in W^*$ and $\lambda_1, \lambda_2 \in F$, we have

$$f^\top(\lambda_1 w_1^* + \lambda_2 w_2^*) = (\lambda_1 w_1^* + \lambda_2 w_2^*) \circ f = \lambda_1 w_1^* \circ f + \lambda_2 w_2^* \circ f = \lambda_1 f^\top(w_1^*) + \lambda_2 f^\top(w_2^*)\,.$$

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. □

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

6.13. **Proposition.** *Let V and W be finite-dimensional vector spaces, with bases $v_1, \dots, v_n$ and $w_1, \dots, w_m$, respectively. Let $v_1^*, \dots, v_n^*$ and $w_1^*, \dots, w_m^*$ 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 $A^\top$.*

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 $v^* \in V^*$, that $v^* = \sum_{j=1}^{n} v^*(v_j) v_j^*$, 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. □

6.14. **Corollary.** *Let V be a finite-dimensional vector space, and let $v_1, \dots, v_n$ and $w_1, \dots, w_n$ be two bases of V. Let $v_1^*, \dots, v_n^*$ and $w_1^*, \dots, w_n^*$ be the corresponding dual bases. If P is the basis change matrix associated to changing the basis of V from $v_1, \dots, v_n$ to $w_1, \dots, w_n$, then the basis change matrix associated to changing the basis of V^* from $v_1^*, \dots, v_n^*$ to $w_1^*, \dots, w_n^*$ is $(P^\top)^{-1} = (P^{-1})^\top =: P^{-\top}$.*

Proof. By definition (see Linear Algebra I, Def. 14.3), P is the matrix representing the identity map id_V with respect to the bases $w_1, \dots, w_n$ (on the domain side) and $v_1, \dots, v_n$ (on the target side). By Prop. 6.13, the matrix representing $\mathrm{id}_{V^*} = \mathrm{id}_V^\top$ with respect to the basis $v_1^*, \dots, v_n^*$ (domain) and $w_1^*, \dots, w_n^*$ (target) is $P^\top$. So this is the basis change matrix associated to changing the basis from $w_1^*, \dots, w_n^*$ to $v_1^*, \dots, v_n^*$; hence we have to take the inverse in order to get the matrix we want. □

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

6.15. **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 $w^* \in W^*$. Then

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

□

6.16. **Proposition.** *Let V be a vector space. Then we have $\alpha_V^\top \circ \alpha_{V^*} = \mathrm{id}_{V^*}$. If V is finite-dimensional, then $\alpha_V^\top = \alpha_{V^*}^{-1}$.*

Proof. Let $v^* \in V^*$, $v \in V$. Then

$$\alpha_V^\top\big(\alpha_{V^*}(v^*)\big)(v) = \big(\alpha_{V^*}(v^*) \circ \alpha_V\big)(v) = \alpha_{V^*}(v^*)\big(\alpha_V(v)\big) = \big(\alpha_V(v)\big)(v^*) = v^*(v)\,,$$

so $\alpha_V^\top\big(\alpha_{V^*}(v^*)\big) = v^*$, and $\alpha_V^\top \circ \alpha_{V^*} = \mathrm{id}_{V^*}$.

If $\dim V < \infty$, then $\dim V^* = \dim V < \infty$, and α_{V^*} is an isomorphism; the relation we have shown then implies that $\alpha_V^\top = \alpha_{V^*}^{-1}$. □

6.17. **Corollary.** *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 Def. 6.11, the map is linear. We have another map

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

and by Prop. 6.15 and Prop. 6.16, the two maps are inverses of each other:

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

and

$$(\alpha_W^{-1} \circ \phi^\top \circ \alpha_V)^\top = \alpha_V^\top \circ \phi^{\top\top} \circ (\alpha_W^\top)^{-1} = \alpha_{V^*}^{-1} \circ \phi^{\top\top} \circ \alpha_{W^*} = \phi\,.$$

□

Next, we study how subspaces relate to dualization.

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

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

is called the *annihilator* of S.

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

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

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

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

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

So we have $W^{\circ\circ} = W$ if we identify V and V^{**} via α_V.

Proof. Let $v_1, \dots, v_n$ be a basis of V such that $v_1, \dots, v_m$ is a basis of W (where $m = \dim W \le \dim V = n$). Let $v_1^*, \dots, v_n^*$ be the corresponding dual basis of V^*. Then for $v^* = \lambda_1 v_1^* + \dots + \lambda_n v_n^* \in V^*$,

$$\begin{aligned} v^* \in W^\circ &\iff \forall w \in W : v^*(w) = 0 \\ &\iff \forall j \in \{1, \dots, m\} : v^*(v_j) = 0 \\ &\iff \forall j \in \{1, \dots, m\} : \lambda_j = 0, \end{aligned}$$

so $v_{m+1}^*, \dots, v_n^*$ is a basis of W°. The dimension formula follows.

In a similar way, we find that $\alpha_V(v_1), \dots, \alpha_V(v_m)$ is a basis of $W^{\circ\circ}$, therefore $\alpha_V(W) = W^{\circ\circ}$.

Alternatively, we trivially have $\alpha_V(W) \subset W^{\circ\circ}$ (check the definitions!), and since both spaces have the same dimension by the first claim, they must be equal. □

Finally, we discuss how annihilators and transposes of linear maps are related.

6.20. **Theorem.** *Let V and W be vector spaces, $f : V \to W$ a linear map. Then*

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

If V and W are finite-dimensional, we have the equality

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

Proof. Let $w^* \in W^*$. Then we have

$$\begin{aligned} w^* \in \bigl(\operatorname{im}(f)\bigr)^\circ &\iff w^*\bigl(f(v)\bigr) = 0 \text{ for all } v \in V \\ &\iff \bigl(f^\top(w^*)\bigr)(v) = 0 \text{ for all } v \in V \\ &\iff f^\top(w^*) = 0 \iff w^* \in \ker(f^\top). \end{aligned}$$

Similarly, for $v^* \in V^*$, we have

$$\begin{aligned} v^* \in \operatorname{im}(f^\top) &\iff \exists w^* \in W^* : v^* = w^* \circ f \\ &\implies \forall v \in \ker(f) : v^*(v) = 0 \\ &\iff v^* \in \bigl(\ker(f)\bigr)^\circ. \end{aligned}$$

For the other direction, write $W = f(V) \oplus U$. Assume that $v^* \in V^*$ satisfies $v^*|_{\ker(f)} = 0$. Define w^* on $f(V)$ by $w^*(f(v)) = v^*(v)$; this is well-defined since v^* takes the same value on all preimages under f of a given element of W. We can extend w^* to all of W by setting $w^*|_U = 0$. Then $v^* = w^* \circ f$.

The dimension formula now follows:

$$\dim \operatorname{im}(f^\top) = \dim\bigl(\ker(f)\bigr)^\circ = \dim V - \dim \ker(f) = \dim \operatorname{im}(f).$$

□

6.21. **Remark.** The equality of dimensions $\dim \operatorname{im}(f^\top) = \dim \operatorname{im}(f)$ is, by Prop. 6.13, equivalent to the statement "row rank equals column rank" for matrices.

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

6.22. **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. Then for $x \in W^*$, we have $x \in (\operatorname{im}(f))^\circ$ if and only if $x^\top y = x \cdot y = 0$ 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 $f^\top : W^* \to V^*$.

If we define a *kernel matrix* of A to be a matrix whose columns span the kernel of A, the this says the following. Let $A \in \operatorname{Mat}(m \times n, F)$ and $B \in \operatorname{Mat}(m \times k, F)$ be matrices. Then the image of B (i.e., the column space of B) is the annihilator of the image of A if and only if B is a kernel matrix of $A^\top$. The condition for B to be a kernel matrix of $A^\top$ means

$$A^\top B = 0 \qquad \text{and} \qquad A^\top x = 0 \Longrightarrow \exists y : x = By\,.$$

Since the relation "U is the annihilator of U'" is symmetric (if we identify a space with its bidual), we find that B is a kernel matrix of $A^\top$ if and only if A is a kernel matrix of $B^\top$.

The statement $(\ker(f))^\circ = \operatorname{im}(f^\top)$ translates into "if B is a kernel matrix of A, then $A^\top$ is a kernel matrix of $B^\top$", which follows from the equivalence just stated.

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.

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

Proof. We use the maximum metric on $\mathbb{R}^n$:

$$d(x, y) = \max\{|x_j - y_j| : j \in \{1, \dots, n\}\}.$$

Let $\|\cdot\|$ be a norm, and let $C = \sum_{j=1}^n \|e_j\|$, where $e_1, \dots, e_n$ is the canonical basis of $\mathbb{R}^n$. Then we have

$$\begin{aligned} \|(x_1, \dots, x_n)\| &= \|x_1 e_1 + \dots + x_n e_n\| \le \|x_1 e_1\| + \dots + \|x_n e_n\| \\ &= |x_1|\|e_1\| + \dots + |x_n|\|e_n\| \le \max\{|x_1|, \dots, |x_n|\}C. \end{aligned}$$

From the triangle inequality, we then get

$$\big|\|y\| - \|x\|\big| \le \|y - x\| \le C d(y, x).$$

So if $d(y, x) < \varepsilon/C$, then $\big|\|y\| - \|x\|\big| < \varepsilon$. □

7.5. **Definition.** Let V be a real vector space, $x \mapsto \|x\|_1$ and $x \mapsto \|x\|_2$ two norms on V. The two norms are said to be *equivalent*, if there are $C_1, C_2 > 0$ such that

$$C_1\|x\|_1 \le \|x\|_2 \le C_2\|x\|_1 \qquad \text{for all } x \in V.$$

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 $x \mapsto x_1^2 + \dots + x_n^2 - 1$ 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 $C_2 \ge C_1 > 0$ (since $0 \notin S$). Now let $0 \ne x \in V$, and let $e = \|x\|_2^{-1} x$; then $\|e\|_2 = 1$, so $e \in S$. This implies that $C_1 \le \|e\| \le C_2$, and therefore

$$C_1\|x\|_2 \le \|x\|_2 \cdot \|e\| = \big\|\|x\|_2 e\big\| = \|x\| \le C_2\|x\|_2.$$

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 $(a_n)_{n \ge 0}$ (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

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

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 on $V \times W$, 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. **Lemma and Definition.** *A bilinear form $\phi : V \times W \to F$ induces linear maps*

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

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 *non-degenerate* if ϕ_L and ϕ_R are isomorphisms. In this case V and W have the same finite dimension (Exercise).

If V and W are finite-dimensional, then ϕ is non-degenerate if and only if both its left and right kernels are trivial.

Proof. First, 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, so ϕ_L and ϕ_R are well-defined as maps into W^* and V^*, respectively. Then using the definition of bilinearity again, we see that ϕ_L and ϕ_R are themselves linear maps.

To prove the last statement, first observe that the left and right kernels are certainly trivial when ϕ_L and ϕ_R are isomorphisms. For the converse statement, assume that the left and right kernels are trivial, so ϕ_L and ϕ_R are both injective. Now note that $\phi_R^\top \circ \alpha_V = \phi_L$ (Exercise). Since V is finite-dimensional, α_V is an isomorphism, hence $\phi_R^\top$ is injective, since ϕ_L is. This implies that ϕ_R is surjective (by Thm. 6.20), so ϕ_R (and in the same way, ϕ_L) is an isomorphism. □

8.4. **Example.** For the 'evaluation pairing' $\mathrm{ev} : V \times V^* \to F$, we find that the map $\mathrm{ev}_L : V \to V^{**}$ is α_V, and $\mathrm{ev}_R : V^* \to V^*$ is the identity. So this bilinar form ev is non-degenerate if and only if α_V is an isomorphism, which is the case if and only if V is finite-dimensional.

8.5. **Example.** The standard dot product on F^n is a non-degenerate symmetric bilinear form. In fact, here ϕ_L sends the standard basis vector e_j to the jth coordinate map in $(F^n)^*$, so it maps a basis to a basis and is therefore an isomorphism.

8.6. **Remarks.**

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

(2) If ϕ is a bilinear form on the finite-dimensional vector space V, then ϕ is non-degenerate if and only if its left kernel is trivial (if and only if its right kernel is trivial).

Indeed, in this case, $\dim V^* = \dim V$, so if ϕ_L is injective, it is also surjective, hence an isomorphism. But then $\phi_R = \phi_L^\top \circ \alpha_V$ is an isomorphism as well.

In fact, we can say a little bit more.

8.7. **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 let $\phi = \gamma_{V,W}(f)$. Then for $v \in V$, $\phi_(v)$ 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.

8.8. **Example.** Let V now be finite-dimensional. We see that a non-degenerate bilinear form ϕ on V allows us to identify V with V^* via the isomorphism ϕ_L. On the other hand, if we fix a basis $v_1, \dots, v_n$, we also obtain an isomorphism $\iota : V \to V^*$ by sending v_j to v_j^*, where $v_1^*, \dots, v_n^*$ is the dual basis of V^*. What is the bilinear form corresponding to this map? We have, for $v = \sum_{j=1}^n \lambda_j v_j$, $w = \sum_{j=1}^n \mu_j v_j$,

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

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.

8.9. **Corollary.** *Let V be a finite-dimensional vector space, and let ϕ be a non-degenerate bilinear form on V. Then every linear form $\psi \in V^*$ is represented as $\psi(w) = \phi(v, w)$ for a unique $v \in V$.*

Proof. The equality $\psi = \phi(v, \cdot)$ means that $\psi = \phi_L(v)$. The claim now follows from the fact that ϕ_L is an isomorphism. □

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

$$\phi : (p, q) \longmapsto \int_0^1 p(x)q(x)\,dx$$

is a bilinear form on V. It is non-degenerate 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 Cor. 8.9 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:

$$\begin{aligned}&\phi(a_0 + a_1 x + a_2 x^2, b_0 + b_1 x + b_2 x^2) \\ &\qquad = a_0 b_0 + \frac{1}{2}(a_0 b_1 + a_1 b_0) + \frac{1}{3}(a_0 b_2 + a_1 b_1 + a_2 b_0) + \frac{1}{4}(a_1 b_2 + a_2 b_2) + \frac{1}{5} a_2 b_2 \,,\end{aligned}$$

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

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

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

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

8.11. **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

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

Similarly, if V and W are finite-dimensional F-vector spaces, and we fix bases $v_1, \dots, v_n$ of V and $w_1, \dots, w_m$ of W, 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. If $A = (a_{ij})$ is the matrix as above, then $a_{ij} = \phi(v_j, w_i)$. If $v = x_1 v_1 + \dots + x_n v_n$ and $w = y_1 w_1 + \dots + y_m w_m$, then

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

8.12. **Proposition.** *Let V and W be finite-dimensional F-vector spaces. Pick two bases $v_1, \dots, v_n$ and $v'_1, \dots, v'_n$ of V and two bases $w_1, \dots, w_m$ and $w'_1, \dots, w'_m$ of W. Let A be the matrix representing the bilinear form $\phi : V \times W \to F$ with respect to $v_1, \dots, v_n$ and $w_1, \dots, w_m$, and let A' be the matrix representing ϕ with respect to $v'_1, \dots, v'_n$ and $w'_1, \dots, w'_m$. If P is the basis change matrix associated to changing the basis of V from $v_1, \dots, v_n$ to $v'_1, \dots, v'_n$, and Q is the basis change matrix associated to changing the basis of W from $w_1, \dots, w_m$ to $w'_1, \dots, w'_m$, then we have*

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

Proof. Let $x' \in F^n$ be the coefficients of $v \in V$ w.r.t. the new basis $v'_1, \dots, v'_n$. Then $x = Px'$, where x represents v w.r.t. the old basis $v_1, \dots, v_n$. Similary for $y', y \in F^m$ representing $w \in W$ w.r.t. the two bases, we have $y = Qy'$. So

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

which implies the claim. □

In particular, if ϕ is a bilinear form on the n-dimensional vector space V, then ϕ is represented (w.r.t. any given basis) by a square matrix $A \in \mathrm{Mat}(n, F)$. If we change the basis, then the new matrix will be $B = P^\top A P$, with $P \in \mathrm{Mat}(n, F)$ invertible. Matrices A and B such that there is an invertible matrix $P \in \mathrm{Mat}(n, F)$ such that $B = P^\top A P$ are called *congruent*.

8.13. **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 $1, x, \dots, x^{n-1}$, it is represented by the "Hilbert matrix" $H_n = \left(\frac{1}{i+j-1}\right)_{1 \le i,j \le n}$.

8.14. **Lemma.** *Let ϕ be a bilinear form on the finite-dimensional vector space V, represented (w.r.t. 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.*
(4) *ϕ is non-degenerate if and only if $\det A \neq 0$.*

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

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

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

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

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

The last statement follows by observing that the left kernel of ϕ corresponds to the kernel of A, which is trivial if and only if $\det A \neq 0$, i.e., A is invertible. □

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

8.15. **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 $u_1, u_2 \in U$, $u'_1, u'_2 \in U'$, 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 generalized to an arbitrary number of summands.

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

8.16. **Definition.** Let ϕ be a symmetric bilinear form on V, and let $U \subset V$ be a linear subspace. Then

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

is the subspace *orthogonal* to U (with respect to ϕ).

8.17. **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 non-degenerate. Then $V = U \oplus U^\perp$, and ϕ splits accordingly as an orthogonal direct sum.*

In this case, we also 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 non-degenerate 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 non-degenerate on U, by Cor. 8.9 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 $U \times U^\perp$. □

Here is a first and quite general classification result for symmetric bilinear forms: they can always be diagonalized.

8.18. **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. □

8.19. **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 8.18, 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 non-degenerate on U. By Prop. 8.17, 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 $v_1, v_2, \ldots, v_n$. □

8.20. **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.

8.21. **Example.** Let us see how we can find a diagonalizing basis in practice. Consider the bilinear form on F^3 (with $\mathrm{char}(F) \neq 2$) 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, Lemma 8.18 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 of 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

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

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 P's 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 $v_1 = (1,1,0)^\top$, $v_2 = (-\frac{1}{2}, \frac{1}{2}, 0)^\top$, $v_3 = (-1,-1,1)^\top$.

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

8.22. **Theorem (Classification of Symmetric Bilinear Forms Over $\mathbb{C}$).** *Let F be algebraically closed, 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 $0 \le r \le n$ is uniquely determined.

Proof. By Thm. 8.19, 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.

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

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

8.24. **Remark.** A positive definite symmetric bilinear form is non-degenerate: if $v \neq 0$, then $\phi(v, v) > 0$, so $\neq 0$, hence v is not in the (left or right) kernel of v. For example, this implies that the Hilbert matrix from Example 8.13 is invertible.

8.25. **Theorem (Classification of Symmetric Bilinear Forms Over $\mathbb{R}$).** *Every symmetric matrix $A \in \mathrm{Mat}(n, \mathbb{R})$ 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 Thm. 8.19, A is congruent to a diagonal matrix, and 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 a_{ii} is a non-zero diagonal entry, we scale the corresponding basis vector by $1/\sqrt{|a_{ii}|}$. 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 the form as before, and the number r is the maximal dimension of a subspace on which the bilinear form is positive definite, therefore r and s only depend on the bilinear form, hence are uniquely determined. □

8.26. **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

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

8.27. **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 $1 \leq j \leq n$.*

Proof. First observe that if a matrix B represents a positive definite symmetric bilinear form, then $\det B > 0$: by Thm. 8.25, 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 $n \geq 2$, 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. □

9. Inner Product Spaces

In many applications, we want to measure *distances* an *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.

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 $x \in V \setminus \{0\}$. A positive definite Hermitian form on the complex vector space V is also called an *inner product* on V; in this context, the form is usually again written as $(x, y) \mapsto \langle x, y\rangle \in \mathbb{C}$. 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*. A real or complex vector space with an inner product on it is an *inner product space*.

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 denote by $\bar{V}^* = \overline{V^*}$ the dual of this space. If V is a complex inner product space, we get again injective homomorphisms

$$V \longrightarrow \bar{V}^*\,, \quad x \longmapsto (y \mapsto \langle x, y\rangle)$$

and

$$\bar{V} \longrightarrow V^*\,, \quad x \longmapsto (y \mapsto \langle y, x\rangle)\,,$$

which are isomorphisms when V is finite-dimensional.

9.3. **Examples.** We have seen some examples of real inner product spaces already: the space $\mathbb{R}^n$ together with the usual 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

$$\langle f, g\rangle = \int_a^b f(x)g(x)\,dx\,.$$

As for complex inner product spaces, the finite-dimensional standard example is $\mathbb{C}^n$ with the 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 inner 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

$$\langle f, g\rangle = \int_a^b f(x)\overline{g(x)}\,dx\,.$$

9.4. **Definition.** Let V be an inner product space.

(1) For $x \in V$, we set $\|x\| = \sqrt{\langle x, x\rangle} \geq 0$. 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 $x \neq y$. S is an *orthonormal* set if in addition, $\|x\| = 1$ for all $x \in S$.
(4) $v_1, \ldots, v_n \in V$ is an *orthonormal basis* or *ONB* of V if the vectors form a basis that is an orthonormal set.

9.5. **Proposition.** *Let V be an inner product space.*

(1) *For $x \in V$ and a scalar λ, we have $\|\lambda x\| = |\lambda|\,\|x\|$.*
(2) **(Cauchy-Schwarz inequality)** *For $x, y \in V$, we have $|\langle x, y\rangle| \leq \|x\|\,\|y\|$, 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 $\langle z, y\rangle = 0$. We find that

$$0 \leq \langle z, z\rangle = \langle z, x\rangle = \langle x, x\rangle - \frac{\langle x, y\rangle}{\|y\|^2}\langle y, x\rangle = \|x\|^2 - \frac{|\langle x, y\rangle|^2}{\|y\|^2}\,,$$

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

$$\begin{aligned}\|x + y\|^2 &= \langle x + y, x + y\rangle = \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \\ &= \|x\|^2 + 2\,\mathrm{Re}\langle x, y\rangle + \|y\|^2 \leq \|x\|^2 + 2|\langle x, y\rangle| + \|y\|^2 \\ &\leq \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 = \left(\|x\| + \|y\|\right)^2\,,\end{aligned}$$

using the Cauchy-Schwarz inequality.

□

Next we show that given any basis of 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.

9.6. **Theorem (Gram-Schmidt Orthonormalization Process).** *Let V be an inner product space. For $0 \neq x \in V$ write $x^\bullet = x/\|x\|$. Let $x_1, \ldots, x_k \in V$ be linearly independent, and define*

$$\begin{aligned} y_1 &= x_1^\bullet \\ y_2 &= (x_2 - \langle x_2, y_1\rangle y_1)^\bullet \\ y_3 &= (x_3 - \langle x_3, y_1\rangle y_1 - \langle x_3, y_2\rangle y_2)^\bullet \\ &\vdots \\ y_k &= (x_k - \langle x_k, y_1\rangle y_1 - \cdots - \langle x_k, y_{k-1}\rangle y_{k-1})^\bullet \,. \end{aligned}$$

Then $y_1, \ldots, y_k$ is an orthonormal basis of $L(x_1, \ldots, x_k)$.

Proof. By induction on k. The case $k = 1$ (or $k = 0$) is clear — $x_1 \neq 0$, hence $y_1 = x_1^\bullet$ is defined, and we have $\|y_1\| = 1$.

If $k \geq 2$, we know by the induction hypothesis that $y_1, \ldots, y_{k-1}$ is an ONB of $L(x_1, \ldots, x_{k-1})$. Let

$$z = x_k - \langle x_k, y_1\rangle y_1 - \cdots - \langle x_k, y_{k-1}\rangle y_{k-1} \,.$$

Since $x_1, \ldots, x_k$ are linearly independent, $z \neq 0$. Also, for $1 \leq j \leq k-1$, we have that

$$\langle z, y_j\rangle = \langle x_k, y_j\rangle - \sum_{i=1}^{k-1} \langle x_k, y_i\rangle\langle y_j, y_i\rangle = \langle x_k, y_j\rangle - \langle x_k, y_j\rangle = 0$$

since $y_1, \ldots, y_{k-1}$ are orthonormal. With $y_k = z^\bullet$, we then also have $\langle y_k, y_j\rangle = 0$ for $j < k$, and $\|y_k\| = 1$. The remaining orthonormality relations hold by the induction hypothesis. That $L(y_1, \ldots, y_k) = L(x_1, \ldots, x_k)$ is clear from the construction. □

9.7. **Corollary.** *Every finite-dimensional inner product space has an ONB.*

Proof. Apply Thm.9.6 to a basis of the space. □

9.8. **Corollary.** *If V is an n-dimensional inner product space, and $\{e_1, \ldots, e_k\} \subset V$ is an orthonormal set, then there are $e_{k+1}, \ldots, e_n \in V$ such that $e_1, \ldots, e_n$ is an ONB of V.*

Proof. Extend $e_1, \ldots, e_k$ to a basis of V in some way and apply Thm. 9.6 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.

9.9. **Proposition.** *Let V be an inner product space and $S \subset V$ an orthogonal set of non-zero vectors. Then S is linearly independent.*

Proof. Let $T \subset S$ be finite, and assume we have a linear combination

$$\sum_{s \in T} \lambda_s s = 0 \,.$$

Now we take the inner product with a fixed $t \in T$:

$$0 = \langle \sum_{s \in T} \lambda_s s, t\rangle = \sum_{s \in T} \lambda_s \langle s, t\rangle = \lambda_t \langle t, t\rangle \,.$$

Since $t \neq 0$, we have $\langle t, t\rangle \neq 0$, therefore we must have $\lambda_t = 0$. Since this works for any $t \in T$, the linear combination must have been trivial. □

9.10. **Theorem (Bessel's Inequality).** *Let V be an inner product space, and let $\{e_1, \dots, e_n\} \subset V$ be an orthonormal set. Then for all $x \in V$, we have the inequality*

$$\sum_{j=1}^{n} \left|\langle x, e_j\rangle\right|^2 \leq \|x\|^2 .$$

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) *$\sum_{j=1}^{n} \left|\langle x, e_j\rangle\right|^2 = \|x\|^2$;*

(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

$$0 \leq \langle z, z\rangle = \langle x, x\rangle - \sum_{j=1}^{n} \langle x, e_j\rangle\langle e_j, x\rangle = \|x\|^2 - \sum_{j=1}^{n} \left|\langle x, e_j\rangle\right|^2 .$$

This implies the inequality and also gives the implication (2) ⇒ (3). 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.

9.11. **Theorem.** *Let V and W be two inner product spaces over the same field ($\mathbb{R}$ or $\mathbb{C}$), with V finite-dimensional, and let $f : V \to W$ be linear. Then there is a unique linear map $f^* : W \to V$ such that*

$$\langle f(v), w \rangle = \langle v, f^*(w) \rangle$$

for all $v \in V$, $w \in W$.

Proof. For $w \in W$ fixed, the map $V \ni v \mapsto \langle f(v), w \rangle$ is a linear form on V. Since V is finite-dimensional, there is a unique $f^*(w) \in V$ satisfying the desired relation for all $v \in V$. Now consider $w + w'$. We find that $f^*(w + w')$ and $f^*(w) + f^*(w')$ both satisfy the relation, so by uniqueness, f^* is additive. Similary, considering λw, we see that $f^*(\lambda w)$ and $\lambda f^*(w)$ must agree. Hence f^* is actually a linear map. □

9.12. **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 Thm. 9.11 (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) An isomorphism $f : V \to W$ is an *isometry* if $\langle f(v), f(w) \rangle = \langle v, w \rangle$ for all $v, w \in V$.

9.13. **Examples.** If $f : V \to V$ is self-adjoint or an isometry, then f is normal. For the second claim, note that an automorphism f is an isometry if and only if $f^* = f^{-1}$. (See also below.)

9.14. **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 $v \in V_1$, $v' \in V_2$,
$$\langle v', f(v)\rangle = \overline{\langle f(v), v'\rangle} = \overline{\langle v, f^*(v')\rangle} = \langle f^*(v'), v\rangle = \langle v', (f^*)^*(v)\rangle \,,$$
and the claim follows.

□

Now we characterize isometries.

9.15. **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 $v \in V$, $w \in W$,
$$\langle v, f^*(w)\rangle = \langle f(v), w\rangle = \langle f(v), f\bigl(f^{-1}(w)\bigr)\rangle = \langle v, f^{-1}(w)\rangle \,,$$
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. □

9.16. **Theorem.** *Let V be a finite-dimensional inner product space and $f : V \to V$ a linear map. Then we have*
$$\mathrm{im}(f^*) = \bigl(\ker(f)\bigr)^\perp \qquad \textit{and} \qquad \ker(f^*) = \bigl(\mathrm{im}(f)\bigr)^\perp .$$

Proof. We first show the inclusions $\mathrm{im}(f^*) \subset \bigl(\ker(f)\bigr)^\perp$ and $\ker(f^*) \subset \bigl(\mathrm{im}(f)\bigr)^\perp$. So let $z \in im(f^*)$, 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 $z \in \bigl(\ker(f)\bigr)^\perp$. If $y \in \ker(f^*)$ and $z = f(x) \in \mathrm{im}(f)$, then
$$\langle z, y\rangle = \langle f(x), y\rangle = \langle x, f^*(y)\rangle = \langle x, 0\rangle = 0 \,,$$
so $y \in \bigl(\mathrm{im}(f)\bigr)^\perp$. Now we have
$$\begin{aligned} \dim \mathrm{im}(f) &= \dim V - \dim\bigl(\mathrm{im}(f)\bigr)^\perp \\ &\le \dim V - \dim \ker(f^*) \\ &= \dim \mathrm{im}(f^*) \\ &\le \dim\bigl(\ker(f)\bigr)^\perp \\ &= \dim V - \dim \ker(f) \\ &= \dim \mathrm{im}(f) \,. \end{aligned}$$
So we must have equality throughout, which implies the result. □

Now we relate the notions of adjoint etc. to matrices representing the linear maps *with respect to orthonormal bases.*

9.17. **Proposition.** *Let V and W be two inner product spaces over the same field, let $v_1, \dots, v_n$ and $w_1, \dots, w_m$ be orthonormal bases of V and W, respectively, and let $f : V \to W$ 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.*

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. First note that if $v, v' \in V$ have coordinates $x, x' \in F^n$ with respect to the ONB $v_1, \dots, v_n$, then

$$\langle v, v' \rangle = x \cdot \bar{x}' = x^\top \bar{x}' ;$$

this follows from Thm. 9.10. A similar statement holds for the inner product of W. So for $x \in F^n$, $y \in F^m$ representing $v \in V$ and $w \in W$, we have

$$\langle v, f^*(w) \rangle = \langle f(v), w \rangle = (Ax)^\top \bar{y} = x^\top A^\top \bar{y} = x^\top (\overline{\bar{A}^\top y}) .$$

This shows that $\bar{A}^\top = A^*$ represents f^*. □

Warning. If the given bases are not orthonormal, then the statement is *wrong* in general.

9.18. **Corollary.** *In the situation above, if $V = W$, then we have the following:*

(1) *f is self-adjoint if and only if $A^* = A$;*
(2) *f is normal if and only if $A^*A = AA^*$;*
(3) *f is an isometry if and only if $AA^* = I$.*

Proof. This is clear. □

9.19. **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 $\mathbb{C}^n$.

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?

Let us first consider the case of complex inner product spaces.

Our goal in the following is to show that f has an orthonormal basis of eigenvectors (so can be orthogonally diagonalized or is *orthodiagonalizable* — nice word!) if and only if f is normal.

10.1. **Lemma.** *Let V be an inner product space and $f : V \to V$ an endomorphism. If f is orthodiagonalizable, then f is normal.*

Proof. If f is orthodiagonalizable, then there is 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 Cor. 9.18. □

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) $\|f^*(v)\| = \|f(v)\|$.
(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

$$\begin{aligned}\|f^*(v)\|^2 &= \langle f^*(v), f^*(v)\rangle = \langle f\bigl(f^*(v)\bigr), v\rangle \\ &= \langle f^*\bigl(f(v)\bigr), v\rangle = \langle f(v), f(v)\rangle = \|f(v)\|^2 .\end{aligned}$$

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

$$\langle f^*(v) - \bar{\lambda} v, f^*(v) - \bar{\lambda} v\rangle = \langle f^*(v), f^*(v)\rangle - \langle \bar{\lambda} v, f^*(v)\rangle - \langle f^*(v), \bar{\lambda} v\rangle + \langle \bar{\lambda} v, \bar{\lambda} v\rangle = 0 .$$

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 a complex inner product space, it is orthodiagonalizable if it is just diagonalizable. It remains to prove that this is the case.

10.3. **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 $p(x) = a_m x^m + \dots + a_0$. Then by Prop. 9.14,

$$p(f)^* = (a_m f^{\circ m} + \dots + a_1 f + a_0 \,\mathrm{id}_V)^* = \bar{a}_m (f^*)^{\circ m} + \dots + \bar{a}_1 f^* + \bar{a}_0 \,\mathrm{id}_V = \bar{p}(f^*) ,$$

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. □

10.4. **Lemma.** *Let V be a finite-dimensional inner product space, and let $f : V \to V$ be normal. Then $V = \ker(f) \oplus \operatorname{im}(f)$ is an orthogonal direct sum.*

Proof. Let $v \in \ker(f)$ and $w \in \operatorname{im}(f)$. We show that $v \perp w$. The claim follows, since $\dim \ker(f) + \dim \operatorname{im}(f) = \dim V$.

We have $f(v) = 0$, so $f^*(v) = 0$ by Lemma 10.2, and $w = f(u)$. Then

$$\langle v, w\rangle = \langle v, f(u)\rangle = \langle f^*(v), u\rangle = \langle 0, u\rangle = 0 .$$

□

10.5. **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 give two proofs. In the first proof, we 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 $f - \alpha \operatorname{id}_V$ is normal. Let $v \in V$ and consider $w = (f - \alpha \operatorname{id}_V)\big(g(f)(v)\big)$. Obviously $w \in \operatorname{im}(f - \alpha \operatorname{id}_V)$, but also $(f - \alpha \operatorname{id}_V)(w) = M_f(f)(v) = 0$, so $w \in \ker(f - \alpha \operatorname{id}_V)$. 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.

The second proof proceeds by induction on $\dim V$. The base case $\dim V = 1$ (or $= 0$) is trivial. So assume $\dim V \geq 2$. f has at least one eigenvector v, say with eigenvalue λ. Let $U = \ker(f - \lambda \operatorname{id}_V) \neq 0$ be the eigenspace and $W = \operatorname{im}(f - \lambda \operatorname{id}_V)$. We know that $V = U \oplus W$ is an orthogonal direct sum, and we have that $f(W) \subset W$, so $f|_W : W \to W$ is again a normal map. By induction, $f|_W$ is diagonalizable. Since $f|_U = \lambda \operatorname{id}_U$ is trivially diagonalizable, f is diagonalizable. (The same proof would also prove directly that f is orthodiagonalizable.) □

So we have now proved the following.

10.6. **Theorem.** *Let V be a finite-dimensional complex inner product space, and let $f : V \to V$ be a linear map. Then f has an orthonormal basis of eigenvectors if and only if f is normal.*

This nice result leaves one question open: what is the situation for *real* inner product spaces? The key to this is the following observation.

10.7. **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. Obviously, we have that D is self-adjoint if and only if $\bar{D}^\top = D$ if and only if $\bar{D} = D$ if and only if all entries of D (i.e., the eigenvalues of f) are real. □

This implies the following.

10.8. **Theorem.** *Let V be a finite-dimensioanl real inner product space, and let $f : V \to V$ be linear. Then f has an orthonormal basis of eigenvectors 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.

On the other hand, if f is self-adjoint, we know that it is diagonalizable over $\mathbb{C}$ with all eigenvalues real, so it is diagonalizable (over $\mathbb{R}$). Lemma 10.2, (3) then shows that the eigenspaces are orthogonal in pairs. □

In terms of matrices, this reads as follows.

10.9. **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: $PP^\top = I$, 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 diagional 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. □

This statement has a geometric interpretation. If A is a symmetric 2×2-matrix, then the equation

$$\mathbf{x}^\top A\mathbf{x} = 1 \tag{1}$$

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 (1) when A is a larger symmetric matrix.

10.10. **Example.** Let us consider the conic section given by the equation

$$5\,x^2 + 4\,xy + 2\,y^2 = 1\,.$$

The matrix is

$$A = \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}\,.$$

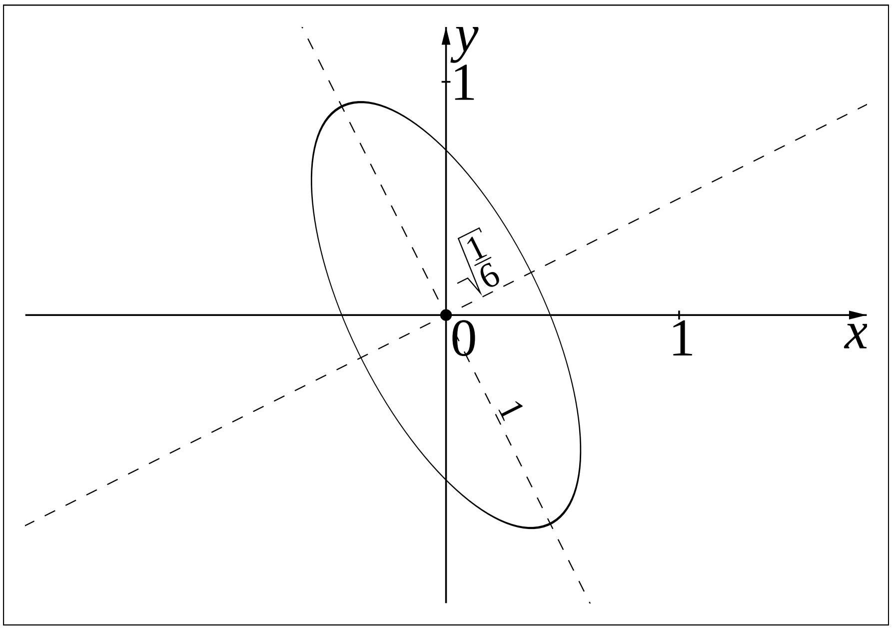
We have to find its eigenvalues and eigenvectors. The characteristic polynomial is $(X - 5)(X - 2) - 4 = X^2 - 7\,X + 6 = (X - 1)(X - 6)$, so we have the two eigenvalues 1 and 6. This already tells us that we have an ellipse. To find the eigenvectors, we have to determine the kernels of $A - I$ and $A - 6I$. 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 $(1\ {-2})^\top$ and of $(2\ 1)^\top$. To get an orthonormal basis, we have to scale them appropriately; we also need to check whether we have to change the sign on 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 $(1, -2)^\top$) at a distance of 1 from the origin and the second axis (in the direction of $(2, 1)^\top$) at a distance of $1/\sqrt{6}$ from the origin.



The ellipse $5\,x^2 + 4\,xy + 2\,y^2 = 1$.

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 $V = \bigoplus_{i \in I}$ 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 bais of V_j, then $B = \bigcup_{j \in I} \iota_j(B_j)$ is a basis of 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 4.1.

(2) This is again clear from 4.1.

(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.

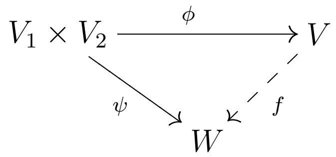
The vector space of $m \times n$ matrices over F can be identified in a natural way with the tensor product $(F^n)^* \otimes F^m$. This identification corresponds to the interpretation of matrices as linear maps from F^n to F^m. The vector space of $m \times n$ matrices over F can also identified in 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 $\psi = f \circ \phi$.



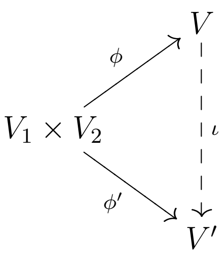
In other words, the canonical linear map

$$\mathrm{Hom}(V, W) \longrightarrow \mathrm{Bil}(V_1 \times V_2, W)\,, \quad f \longmapsto f \circ \phi$$

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. □

Beacuse 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 of W. Then every element of $V \otimes W$ can be written uniquely in the form*

$$\sum_{i=1}^{n} v_i \otimes w_i = v_1 \otimes w_1 + \cdots + v_n \otimes w_n$$

with $v_1, \dots, v_n \in V$.

Proof. Let $x \in V \otimes W$; then by the discussion above, we can write

$$x = y_1 \otimes z_1 + \cdots + y_m \otimes z_m$$

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 of W, we can write

$$z_j = \alpha_{j1} w_1 + \cdots + \alpha_{jn} w_n$$

with scalars α_{jk}. Using the bilinearity of the map $(y, z) \mapsto y \otimes z$, we find that

$$\begin{aligned} x &= y_1 \otimes (\alpha_{11} w_1 + \cdots + \alpha_{1n} w_n) + \cdots + y_m \otimes (\alpha_{m1} w_1 + \cdots + \alpha_{mn} w_n) \\ &= (\alpha_{11} y_1 + \cdots + \alpha_{m1} y_m) \otimes w_1 + \cdots + (\alpha_{1n} y_1 + \cdots + \alpha_{mn} y_m) \otimes w_n , \end{aligned}$$

which is of the required form.

For uniqueness, it suffices to show that

$$v_1 \otimes w_1 + \cdots + v_n \otimes w_n = 0 \quad \Longrightarrow \quad v_1 = \cdots = v_n = 0\,.$$

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

$$0 = f(0) = f(v_1 \otimes w_1 + \cdots + v_n \otimes w_n) = \psi(v_1, w_1) + \cdots + \psi(v_n, w_n) = 1\,,$$

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":

$$\begin{array}{ccc}
(\text{Finite Sets}, \amalg, \times, \cong) & \xrightarrow{\quad B \mapsto \#B \quad} & (\mathbb{N}, +, \cdot, =) \\
& {\scriptstyle B \mapsto F^B} \searrow \qquad \nearrow {\scriptstyle \dim} & \\
& (\text{Finite-dim. Vector Spaces}, \oplus, \otimes, \cong) &
\end{array}$$

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. □

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 of 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,

$$\phi\big(\phi'(f)\big) = \phi\Big(\sum_{i=1}^{n}(w_i^* \circ f) \otimes w_i\Big) = \sum_{i=1}^{n}\Big(v \mapsto w_i^*\big(f(v)\big)w_i\Big)$$
$$= \Big(v \mapsto \sum_{i=1}^{n} w_i^*(f(v))w_i\Big) = \big(v \mapsto f(v)\big) = f\,.$$

□

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?

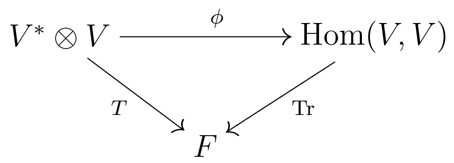
12.14. **Remark.** *Let $v_1, \ldots, v_n$ be a basis of V, and let $v_1^*, \ldots, v_n^*$ be the basis of V^* dual to it. Then, with ϕ the canonical map from above, we have*

$$\phi\Big(\sum_{i=1}^{n} v_i^* \otimes v_i\Big) = \mathrm{id}_V\ .$$

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 $T : V^* \otimes V \to F$ given by $T(l \otimes v) = l(v)$. It makes the following diagram commutative.*



Proof. That T is well-defined is clear by the usual princpiple. (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 $v_1, \ldots, v_n$, with dual basis $v_1^*, \ldots, v_n^*$, 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 $v_1, \ldots, v_n$, 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.*

$$\begin{array}{ccccc} (l\otimes v, l'\otimes w) & (U^*\otimes V)\times(V^*\otimes W) & \xrightarrow{\phi\times\phi} & \mathrm{Hom}(U,V)\times\mathrm{Hom}(V,W) & (f,g) \\ \downarrow & \downarrow & & \downarrow & \downarrow \\ l'(v)\, l\otimes w & U^*\otimes W & \xrightarrow{\phi} & \mathrm{Hom}(U,W) & g\circ f \end{array}$$

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.

$$\begin{array}{ccccc} (l\otimes w, v) & (V^*\otimes W)\times V & \xrightarrow{\phi\times\mathrm{id}_V} & \mathrm{Hom}(V,W)\times V & (f,v) \\ \downarrow & \downarrow & & \downarrow & \downarrow \\ l(v)w & W & = & W & f(v) \end{array}$$

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 of 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

$$\mathbb{C} \times (\mathbb{C} \otimes_{\mathbb{R}} V) \longrightarrow \mathbb{C} \otimes_{\mathbb{R}} (\mathbb{C} \otimes_{\mathbb{R}} V) \cong (\mathbb{C} \otimes_{\mathbb{R}} \mathbb{C}) \otimes_{\mathbb{R}} V \xrightarrow{m \otimes \mathrm{id}_V} \mathbb{C} \otimes_{\mathbb{R}} V\,,$$

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 $v_1, \dots, v_n$. Then every element of $\tilde{V}$ can be written uniquely in the form

$$\alpha_1 \otimes v_1 + \cdots + \alpha_n \otimes v_n \qquad \text{for some } \alpha_1, \dots, \alpha_n \in \mathbb{C}.$$

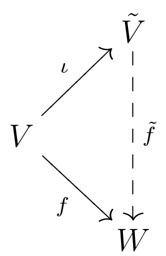
In this sense, we can consider $\tilde{V}$ to have "the same" basis as V, but we allow complex coordinates instead of real ones.

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

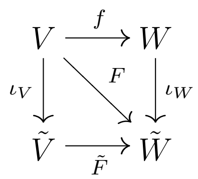
$$\mathrm{Hom}_{\mathbb{R}}(V, W) \xrightarrow{\cong} \mathrm{Hom}_{\mathbb{C}}(\tilde{V}, W)\,, \quad f \longmapsto \tilde{f}\,,$$

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.



Here, $F = \iota_W \circ f$ is an $\mathbb{R}$-linar 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 imples 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

$$f = \sum_{i_1, \dots, i_n = 1}^{m} v_{i_1}^* \otimes \dots \otimes v_{i_n}^* \otimes w_{i_1, \dots, i_n}$$

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, indentified 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, indentified 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 of $\operatorname{Sym}(V^{\otimes n})$. In particular,
$$\dim \operatorname{Sym}(V^{\otimes n}) = \binom{m+n-1}{n}.$$

(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 of $\operatorname{Alt}(V^{\otimes n})$. In particular,
$$\dim \operatorname{Alt}(V^{\otimes n}) = \binom{m}{n}.$$

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 $v_1, v_2, \dots, v_n \in V$ 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
$$\mathrm{Alt}^n(V) = \bigwedge^n(V) = V^{\otimes n}/W$$
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 $v_1 \wedge v_2 \wedge \cdots \wedge v_n$.

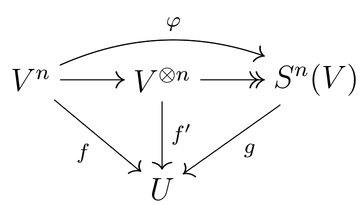
13.11. **Theorem.**

(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*
$$\psi : V^n \longrightarrow \bigwedge^n(V)\,, \quad (v_1, v_2, \ldots, v_n) \longmapsto v_1 \wedge v_2 \wedge \cdots \wedge v_n$$
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
$$\mathrm{Sym}(V^n, U) = \mathrm{Hom}(S^n(V), U) \quad \text{and} \quad \mathrm{Alt}(V^n, U) = \mathrm{Hom}(\textstyle\bigwedge^n(V), U)\,.$$

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 $\mathrm{Sym}(V^{\otimes n})$ and $S^n(V)$ (resp., $\mathrm{Alt}(V^{\otimes n})$ and $\bigwedge^n(V)$) are closely related. We assume that F is of characteristic zero.

13.12. **Proposition.**

(1) *The maps* $\mathrm{Sym}(V^{\otimes n}) \subset V^{\otimes n} \to S^n(V)$ *and*

$$S^n(V) \longrightarrow \mathrm{Sym}(V^{\otimes n}), \quad v_1 \cdot v_2 \cdots v_n \longmapsto \frac{1}{n!} \sum_{\sigma \in S_n} v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \cdots \otimes v_{\sigma(n)}$$

are inverse isomorphisms. In particular, if $b_1, \ldots, b_m$ *is a basis of* 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 of $S^n(V)$*, and* $\dim S^n(V) = \binom{m+n-1}{n}$.

(2) *The maps* $\mathrm{Alt}(V^{\otimes n}) \subset V^{\otimes n} \to \bigwedge^n(V)$ *and*

$$\bigwedge^n(V) \longrightarrow \mathrm{Alt}(V^{\otimes n}), \quad v_1 \wedge v_2 \wedge \cdots \wedge v_n \longmapsto \frac{1}{n!} \sum_{\sigma \in S_n} \mathrm{sign}(\sigma) v_{\sigma(1)} \otimes v_{\sigma(2)} \otimes \cdots \otimes v_{\sigma(n)}$$

are inverse isomorphisms. In particular, if $b_1, \ldots, b_m$ *is a basis of* 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 of $\bigwedge^n(V)$*, 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 $V^n \to S^n(W)$, $(v_1, \dots, v_n) \mapsto f(v_1) \cdots f(v_n)$, 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 $v_1, \dots, v_n$ 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* $\det(g \circ f) = \deg(g)\det(f)$.

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.

References

[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.