Linear Algebra — Lecture 1: Vector Spaces

1 Basic Definitions and Properties

1.1 Vector spaces over a general field

Definition 1.1. *Let $(E, +)$ be a commutative group and let $(K, +, \times)$ be a field. We say that E is a* vector space over K *if there exists a map*

$$K \times E \longrightarrow E, \quad (\alpha, x) \mapsto \alpha \cdot x,$$

called scalar multiplication, *satisfying the following axioms:*

(i) *For all $x \in E$, $1_K \cdot x = x$.*

(ii) *For all $\alpha, \beta \in K$ and all $x \in E$,*

$$(\alpha + \beta) \cdot x = \alpha \cdot x + \beta \cdot x.$$

(iii) *For all $\alpha, \beta \in K$ and all $x \in E$,*

$$(\alpha\beta) \cdot x = \alpha \cdot (\beta \cdot x).$$

(iv) *For all $\alpha \in K$ and all $x, y \in E$,*

$$\alpha \cdot (x + y) = \alpha \cdot x + \alpha \cdot y.$$

In this case, the elements of E are called vectors *and are usually denoted by $x, y, z, \ldots$, while the elements of K are called* scalars *and are usually denoted by $\alpha, \beta, \gamma, \lambda, \ldots$.*

1.2 Consequences of the definition

Proposition 1.1. *Let E be a vector space over K. Then:*

(a) *For all $x \in E$, $0_K \cdot x = 0_E$.*

(b) *For all $\lambda \in K$, $\lambda \cdot 0_E = 0_E$.*

(c) *For all $\lambda \in K$ and all $x \in E$,*

$$\lambda \cdot x = 0_E \iff \lambda = 0_K \text{ or } x = 0_E.$$

Proof. (a) By axiom (ii),

$$0_K \cdot x = (0_K + 0_K) \cdot x = 0_K \cdot x + 0_K \cdot x,$$

hence $0_K \cdot x = 0_E$.

(b) By axiom (iv),

$$\lambda \cdot 0_E = \lambda \cdot (0_E + 0_E) = \lambda \cdot 0_E + \lambda \cdot 0_E,$$

hence $\lambda \cdot 0_E = 0_E$.

(c) ($\Rightarrow$) Suppose that $\lambda \cdot x = 0_E$ and $\lambda \neq 0_K$. Then

$$\lambda^{-1} \cdot (\lambda \cdot x) = \lambda^{-1} \cdot 0_E = 0_E.$$

By axiom (iii),

$$(\lambda^{-1}\lambda) \cdot x = 1_K \cdot x = x,$$

so $x = 0_E$.

($\Leftarrow$) The converse is immediate: if $\lambda = 0_K$ then $\lambda \cdot x = 0_E$ by (a), and if $x = 0_E$ then $\lambda \cdot x = 0_E$ by (b).

$\square$

2 In-class exercise

Exercise 2.1. *Let $\mathbb{R}_+^* = (0, +\infty)$ be the set of strictly positive real numbers.*

Define an internal operation $\oplus$ on $\mathbb{R}_+^*$ by

$$x \oplus y := xy, \quad \forall x, y \in \mathbb{R}_+^*,$$

and define an external operation (scalar multiplication) by

$$\lambda \cdot x := x^\lambda, \quad \forall \lambda \in \mathbb{R},\ \forall x \in \mathbb{R}_+^*.$$

Show that $(\mathbb{R}_+^*, \oplus, \cdot)$ is a vector space over $\mathbb{R}$.

Linear Algebra — Lecture 2: Examples of Vector spaces and Subspaces

1 Examples of Vector Spaces

1.1 Fundamental examples

1. Let K be a field. For any integer $n \geq 1$, the set K^n is a vector space over K with scalar multiplication defined componentwise by

$$\lambda \cdot (x_1, \ldots, x_n) := (\lambda x_1, \ldots, \lambda x_n).$$

2. Let K be a field and let $K[X]$ denote the set of polynomials with coefficients in K. Then $K[X]$ is a vector space over K with scalar multiplication defined by

$$\lambda \cdot \sum_{i=0}^{m} a_i X^i := \sum_{i=0}^{m} (\lambda a_i) X^i.$$

3. Let $E_1, \ldots, E_n$ be vector spaces over the same field K. Then the Cartesian product

$$E_1 \times \cdots \times E_n$$

is a vector space over K, with scalar multiplication defined by

$$\lambda \cdot (x_1, \ldots, x_n) := (\lambda x_1, \ldots, \lambda x_n).$$

4. Let E be a vector space over K and let A be a nonempty set. The set E^A of all functions from A to E is a vector space over K, with operations defined by

$$(\lambda f)(a) := \lambda f(a), \qquad (f+g)(a) := f(a) + g(a), \quad \forall a \in A.$$

Exercise 1.1. *Check that all the examples above are indeed vector spaces !*

2 Subspaces

Definition 2.1. *Let V be a vector space over a field $\mathbb{F}$. A subset $U \subset V$ is called a* subspace *if:*

(i) $0 \in U$,

(ii) *if* $u_1, u_2 \in U$, *then* $u_1 + u_2 \in U$,

(iii) *if* $u \in U$ *and* $\lambda \in \mathbb{F}$, *then* $\lambda u \in U$.

2.1 In-class exercises

Exercise 2.1. *Show that the only subspaces of $\mathbb{R}$ (viewed as a vector space over $\mathbb{R}$) are $\{0\}$ and $\mathbb{R}$.*

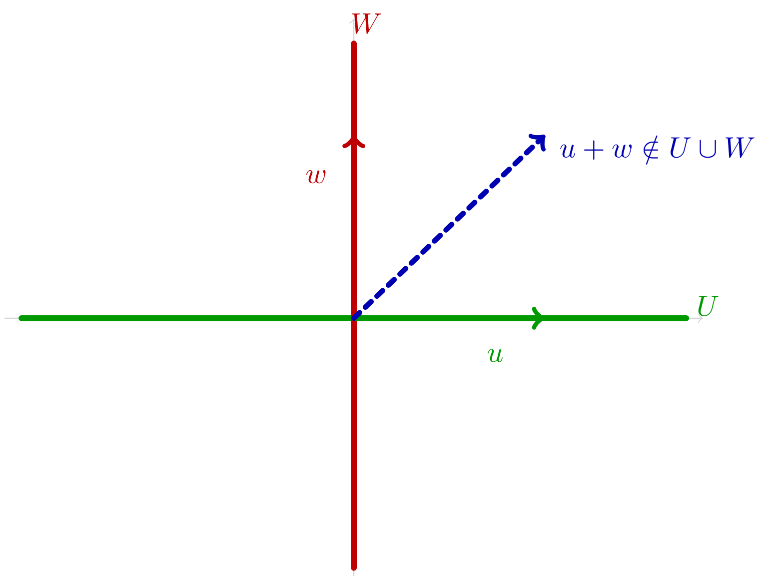
Exercise 2.2. *Let $\mathcal{U}$ be a family of subspaces of a vector space V. Show that*

$$W := \bigcap_{U \in \mathcal{U}} U$$

is a subspace of V.

Remark. **Warning.** The union of two subspaces is *not* a subspace in general. Indeed, if U and W are subspaces of V, and if $u \in U$ and $w \in W$, then in general

$$u + w \notin U \cup W.$$



Proposition 2.1. *Let E be a vector space over a field K, and let F and G be two subspaces of E. Then $F \cup G$ is a subspace of E if and only if*

$$F \subset G \quad \textit{or} \quad G \subset F.$$

Proof. ($\Rightarrow$) Suppose that $F \cup G$ is a subspace of E. We show that either $F \subset G$ or $G \subset F$.

Assume by contradiction that $F \not\subseteq G$ and $G \not\subseteq F$. Then there exist elements

$$x \in F \setminus G \quad \text{and} \quad y \in G \setminus F.$$

Since $F \cup G$ is a subspace, it is closed under addition, so

$$x + y \in F \cup G.$$

Hence either $x + y \in F$ or $x + y \in G$.

If $x + y \in F$, then

$$y = (x + y) - x \in F,$$

since F is a subspace. This contradicts $y \notin F$.

If $x + y \in G$, then

$$x = (x + y) - y \in G,$$

since G is a subspace. This contradicts $x \notin G$.

In both cases we reach a contradiction. Therefore,

$$F \subset G \quad \text{or} \quad G \subset F.$$

($\Leftarrow$) If $F \subset G$, then $F \cup G = G$, which is a subspace. Similarly, if $G \subset F$, then $F \cup G = F$, which is a subspace. $\square$

Linear Algebra — Lecture 3: Sums and Direct Sums of Subspaces

1 Examples of Subspaces

We start by recalling a few standard examples of subspaces.

Example 1.1 (Linear equation)**.** *Let $a_1, \ldots, a_n \in K$ and define*

$$S := \{(x_1, \ldots, x_n) \in K^n : a_1x_1 + \cdots + a_nx_n = 0\}.$$

Then S is a subspace of K^n.

Proof. Clearly $0 \in S$. If $x, y \in S$, then

$$a_1(x_1 + y_1) + \cdots + a_n(x_n + y_n) = 0,$$

so $x + y \in S$. If $\lambda \in K$ and $x \in S$, then

$$a_1(\lambda x_1) + \cdots + a_n(\lambda x_n) = \lambda(a_1x_1 + \cdots + a_nx_n) = 0,$$

so $\lambda x \in S$. □

Example 1.2 (System of linear equations)**.** *Let $(a_{ij})_{1\leq i\leq m,\, 1\leq j\leq n} \in K^{m\times n}$ and define*

$$S := \left\{ (x_1, \ldots, x_n) \in K^n : \begin{array}{c} a_{11}x_1 + \cdots + a_{1n}x_n = 0 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = 0 \end{array} \right\}.$$

Then S is a subspace of K^n.

Proof. For each $i = 1, \ldots, m$, define

$$S_i := \{x \in K^n : a_{i1}x_1 + \cdots + a_{in}x_n = 0\}.$$

Each S_i is a subspace of K^n, and $S = \bigcap_{i=1}^m S_i$. Since intersections of subspaces are subspaces, S is a subspace of K^n. □

Definition 1.1 (Degree of a polynomial)**.** *Let $p \in K[X]$ be a polynomial. If $p \neq 0$ and*

$$p(x) = a_0 + a_1x + \cdots + a_mx^m \quad \textit{with } a_m \neq 0,$$

we define $\deg(p) = m$. By convention, the zero polynomial is assigned degree $-\infty$.

Example 1.3 (Polynomials of bounded degree)**.** *Let $V = K[X]$ and fix $n \in \mathbb{N}$. Define*

$$K_n[X] := \{f \in K[X] : \deg(f) \leq n\}.$$

Then $K_n[X]$ is a subspace of $K[X]$.

Proof. First, $0 \in K_n[X]$ since $\deg(0) = -\infty \leq n$.

Let $f, g \in K_n[X]$. Then

$$\deg(f+g) \leq \max\{\deg(f), \deg(g)\} \leq n,$$

so $f + g \in K_n[X]$.

If $\lambda \in K$, then $\deg(\lambda f) = \deg(f) \leq n$ when $\lambda \neq 0$, and $\lambda f = 0$ when $\lambda = 0$, which also belongs to $K_n[X]$. Hence $K_n[X]$ is a subspace of $K[X]$. ☐

Remark. The set

$$\{f \in K[X] : f = 0 \text{ or } \deg(f) \geq n\}$$

is not a subspace of $K[X]$. Indeed, X^n and $-X^n + 1$ both belong to this set, but their sum equals 1, which does not.

2 Sum of Subspaces

Definition 2.1 (Sum of subspaces). *Let E be a vector space and let $F_1, \dots, F_p$ be subspaces of E with $p \geq 2$. We define*

$$F_1 + \cdots + F_p := \{x_1 + \cdots + x_p : \ x_i \in F_i\}.$$

Proposition 2.1. *The set $F_1 + \cdots + F_p$ is a subspace of E.*

Proof. First, $0 = 0 + \cdots + 0$ with $0 \in F_i$ for all i, hence $0 \in F_1 + \cdots + F_p$.

Let $x, y \in F_1 + \cdots + F_p$. Then there exist $x_i \in F_i$ and $y_i \in F_i$ such that

$$x = x_1 + \cdots + x_p, \qquad y = y_1 + \cdots + y_p.$$

Hence

$$x + y = (x_1 + y_1) + \cdots + (x_p + y_p).$$

Because each F_i is a subspace, $x_i + y_i \in F_i$, so $x + y \in F_1 + \cdots + F_p$.

Finally, let $\lambda \in K$ and $x = x_1 + \cdots + x_p$ with $x_i \in F_i$. Then

$$\lambda x = (\lambda x_1) + \cdots + (\lambda x_p),$$

and since each F_i is a subspace we have $\lambda x_i \in F_i$, hence $\lambda x \in F_1 + \cdots + F_p$. Therefore $F_1 + \cdots + F_p$ is a subspace of E. ☐

Proposition 2.2 (Smallest subspace containing the F_i). *Let V be a vector space and let $V_1, \dots, V_m$ be subspaces of V. Then $V_1 + \cdots + V_m$ is the smallest subspace of V containing $V_1, \dots, V_m$.*

Proof. Clearly each $V_k \subset V_1 + \cdots + V_m$ (take all other terms equal to 0). Now let U be a subspace of V such that $V_k \subset U$ for all k. If $v \in V_1 + \cdots + V_m$, then $v = v_1 + \cdots + v_m$ with $v_k \in V_k \subset U$. Since U is a subspace, $v \in U$. Thus $V_1 + \cdots + V_m \subset U$, proving minimality. ☐

3 Direct Sum (case of two subspaces)

Definition 3.1 (Direct sum of two subspaces)**.** *Let U, W be subspaces of a vector space V. We say that $U + W$ is a* direct sum *if every $v \in U + W$ can be written in a unique way as*

$$v = u + w, \qquad u \in U,\ w \in W.$$

In that case, we write $U \oplus W$.

Proposition 3.1 (Direct sum of two subspaces)**.** *Let U and W be subspaces of V. Then*

$$U + W \text{ is a direct sum} \quad \Longleftrightarrow \quad U \cap W = \{0\}.$$

Proof. ($\Rightarrow$) Assume $U + W$ is a direct sum and let $v \in U \cap W$. Then

$$0 = v + (-v),$$

where $v \in U$ and $-v \in W$ (since $v \in W$). But also $0 = 0 + 0$ with $0 \in U$ and $0 \in W$. By uniqueness of the representation of 0 as a sum of one element in U and one in W, we get $v = 0$. Hence $U \cap W = \{0\}$.

($\Leftarrow$) Assume $U \cap W = \{0\}$. To prove uniqueness, suppose $u + w = u' + w'$ with $u, u' \in U$ and $w, w' \in W$. Then

$$u - u' = w' - w.$$

The left-hand side belongs to U and the right-hand side belongs to W, hence $u - u' \in U \cap W = \{0\}$. So $u = u'$ and therefore also $w = w'$. Thus the decomposition is unique, so $U + W$ is a direct sum. $\square$

Linear Algebra — Lecture 4: Linear combinations and span

1 Direct sum of several subspaces

Definition 1.1 (Direct sum of p subspaces)**.** *Let E be a vector space and let $F_1, \dots, F_p$ be subspaces of E. We say that the sum $F_1 + \cdots + F_p$ is* direct *if every $x \in F_1 + \cdots + F_p$ admits a unique decomposition*

$$x = x_1 + \cdots + x_p, \qquad x_i \in F_i.$$

In that case, we write $F_1 \oplus \cdots \oplus F_p$.

Lemma 1.1. *The sum $F_1 + \cdots + F_p$ is direct if and only if, for all $(x_1, \dots, x_p) \in F_1 \times \cdots \times F_p$,*

$$x_1 + \cdots + x_p = 0 \quad \Longrightarrow \quad x_1 = \cdots = x_p = 0.$$

Proof. ($\Rightarrow$) If the sum is direct, then 0 has a unique decomposition. Since $0 = 0 + \cdots + 0$, any decomposition $0 = x_1 + \cdots + x_p$ forces $x_i = 0$ for all i.

($\Leftarrow$) Suppose the implication holds and assume $x = x_1 + \cdots + x_p = y_1 + \cdots + y_p$ with $x_i, y_i \in F_i$. Subtracting,

$$(x_1 - y_1) + \cdots + (x_p - y_p) = 0,$$

and each $x_i - y_i \in F_i$. By the implication, $x_i - y_i = 0$ for all i, hence the decomposition is unique. □

Exercise 1.1 (In-class exercise). *Let $p \geq 3$ and let $F_1, \dots, F_p$ be subspaces of a vector space E.*

(1) *Assume that $F_1 \cap F_2 = \{0\}$, $F_1 \cap F_3 = \{0\}$ and $F_2 \cap F_3 = \{0\}$. Are F_1, F_2, F_3 necessarily in direct sum?*

(2) *Prove the equivalence:*

$$F_1 + \cdots + F_p \text{ is direct} \quad \Longleftrightarrow \quad \forall k \in \{1, \dots, p-1\},\ \Big(\sum_{i=1}^{k} F_i\Big) \cap F_{k+1} = \{0\}.$$

2 Linear Combinations and Span

Let V be a vector space over a field $\mathbb{F}$. In this section, we introduce the notions of linear combinations and span, which will play a central role in our study of finite-dimensional vector spaces.

Definition 2.1 (Linear combination)**.** *Let $S \subset V$ be a subset. A vector $x \in V$ is called a* linear combination *of elements of S if there exist a nonnegative integer n, vectors $x_1, \dots, x_n \in S$, and scalars $a_1, \dots, a_n \in \mathbb{F}$ such that*

$$x = a_1 x_1 + \cdots + a_n x_n.$$

Definition 2.2 (Span)**.** *Let $S \subset V$. The* span *of S, denoted by* $\operatorname{span}(S)$*, is the set of all linear combinations of elements of S. When $S = \{x_1, \ldots, x_n\}$, we also write* $\operatorname{span}(x_1, \ldots, x_n)$*.*

Proposition 2.1. *Let V be a vector space and let $S, T \subset V$ with $S \subset T$. Then*

$$\operatorname{span}(S) \subset \operatorname{span}(T).$$

Proof. Every linear combination of elements of S is also a linear combination of elements of T. □

Proposition 2.2. *Let V be a vector space and let $S \subset V$. Then* $\operatorname{span}(S)$ *is a subspace of V that contains S.*

Proof. The zero vector belongs to $\operatorname{span}(S)$ by convention. Closure under addition and scalar multiplication follows directly from the definition of linear combination. Finally, if $x \in S$, then $x = 1 \cdot x$, hence $x \in \operatorname{span}(S)$. □

Proposition 2.3. *Let V be a vector space, let U be a subspace of V, and let $S \subset V$. Then*

$$\operatorname{span}(S) \subset U \quad \Longleftrightarrow \quad S \subset U.$$

Proof. If $\operatorname{span}(S) \subset U$, then in particular $S \subset \operatorname{span}(S)$, hence $S \subset U$.

Conversely, assume that $S \subset U$. Let $x \in \operatorname{span}(S)$. Then x is a linear combination of elements of S, so there exist $x_1, \ldots, x_n \in S$ and $a_1, \ldots, a_n \in \mathbb{F}$ such that

$$x = a_1 x_1 + \cdots + a_n x_n.$$

Since $x_i \in S \subset U$ and U is a subspace, each $a_i x_i$ belongs to U, and therefore $x \in U$. Thus $\operatorname{span}(S) \subset U$. □

Linear Algebra — Lecture 4: Linear combinations and span (2)

1 Properties of Span

Proposition 1.1. *Let V be a vector space, let U be a subspace of V, and let $S \subset V$. Then*

$$\operatorname{span}(S) \subset U \iff S \subset U.$$

Proof. If $\operatorname{span}(S) \subset U$, then in particular $S \subset \operatorname{span}(S)$, hence $S \subset U$.

Conversely, assume that $S \subset U$. Let $x \in \operatorname{span}(S)$. Then x is a linear combination of elements of S, so there exist $x_1, \ldots, x_n \in S$ and $a_1, \ldots, a_n \in \mathbb{F}$ such that

$$x = a_1 x_1 + \cdots + a_n x_n.$$

Since $x_i \in S \subset U$ and U is a subspace, each $a_i x_i$ belongs to U, and therefore $x \in U$. Thus $\operatorname{span}(S) \subset U$. □

Proposition 1.2. *Let V be a vector space and let $S \subset V$. Then* $\operatorname{span}(S)$ *is the smallest subspace of V containing S. Moreover,*

$$\operatorname{span}(S) = \bigcap_{\substack{U \leq V \\ S \subset U}} U,$$

where the intersection is taken over all subspaces U of V that contain S.

Proof. Let

$$W := \bigcap_{\substack{U \leq V \\ S \subset U}} U.$$

Since W is an intersection of subspaces of V, it is itself a subspace. Moreover, every subspace appearing in the intersection contains S, hence $S \subset W$.

We claim that W is the smallest subspace of V containing S. Indeed, if U is any subspace of V such that $S \subset U$, then U appears in the intersection defining W, and therefore $W \subset U$.

In particular, since $\operatorname{span}(S)$ is a subspace of V containing S, we have $W \subset \operatorname{span}(S)$. On the other hand, because W is a subspace containing S, the previous proposition implies that $\operatorname{span}(S) \subset W$. Thus $\operatorname{span}(S) = W$. □

2 Spanning Lists and Finite-Dimensional Vector Spaces

In the previous lecture, we introduced linear combinations and the notion of span. We now specialize to the case of *lists* of vectors and study when such lists generate an entire vector space. This will naturally lead us to the concept of finite-dimensional vector spaces.

Definition 2.1 (Spanning list). *Let V be a vector space and let $(v_1, \dots, v_m)$ be a list of vectors in V. If*

$$\operatorname{span}(v_1, \dots, v_m) = V,$$

we say that the list $(v_1, \dots, v_m)$ spans *V.*

Example 2.1 (A list that spans $\mathbb{F}^n$). *Let n be a positive integer. For each $k \in \{1, \dots, n\}$, let e_k denote the vector in $\mathbb{F}^n$ whose k-th entry is 1 and whose other entries are 0. Then the list*

$$(e_1, \dots, e_n)$$

spans $\mathbb{F}^n$. Indeed, if $x = (x_1, \dots, x_n) \in \mathbb{F}^n$, then

$$(x_1, \dots, x_n) = x_1 e_1 + \cdots + x_n e_n,$$

showing that every vector in $\mathbb{F}^n$ belongs to $\operatorname{span}(e_1, \dots, e_n)$.

Example 2.2 (Polynomial spaces). *Let $\mathbb{F}[X]$ denote the vector space of polynomials in one variable with coefficients in $\mathbb{F}$. Consider the set*

$$S = \{1, X, X^2, X^3, \dots\}.$$

By definition of a polynomial, every element of $\mathbb{F}[X]$ is a finite linear combination of elements of S. Hence

$$\operatorname{span}(S) = \mathbb{F}[X].$$

Example 2.3 (Function spaces). *Let X be a nonempty set and let $V = \mathbb{F}^X$ denote the vector space of all functions from X to $\mathbb{F}$, with pointwise addition and scalar multiplication.*

For each $x \in X$, define the function $\delta_x : X \to \mathbb{F}$ by

$$\delta_x(y) = \begin{cases} 1, & \text{if } y = x, \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$S = \{\delta_x : x \in X\}.$$

We claim that

$$\operatorname{span}(S) = \mathbb{F}^X \quad \text{if and only if} \quad X \text{ is finite.}$$

First assume that X is finite. Write $X = \{x_1, \dots, x_n\}$. Let $f \in \mathbb{F}^X$ be an arbitrary function. We claim that

$$f = f(x_1)\delta_{x_1} + \cdots + f(x_n)\delta_{x_n}.$$

Indeed, let $y \in X$. Then $y = x_i$ for a unique $i \in \{1, \dots, n\}$, and evaluating the right-hand side at y gives

$$\big(f(x_1)\delta_{x_1} + \cdots + f(x_n)\delta_{x_n}\big)(y) = f(x_i)\delta_{x_i}(y) = f(x_i) = f(y).$$

Thus the equality holds for every $y \in X$, and therefore $f \in \operatorname{span}(S)$. Since f was arbitrary, we conclude that $\operatorname{span}(S) = \mathbb{F}^X$.

Now assume that X is infinite. Define a function $f : X \to \mathbb{F}$ by

$$f(x) = 1 \quad \textit{for all } x \in X.$$

Suppose, toward a contradiction, that $f \in \operatorname{span}(S)$. Then there exist distinct elements $x_1, \dots, x_m \in X$ and scalars $a_1, \dots, a_m \in \mathbb{F}$ such that

$$f = a_1\delta_{x_1} + \cdots + a_m\delta_{x_m}.$$

Because X is infinite, there exists $x \in X$ such that $x \notin \{x_1, \dots, x_m\}$. Evaluating both sides at this x, we obtain

$$1 = f(x) = (a_1\delta_{x_1} + \cdots + a_m\delta_{x_m})(x) = 0,$$

since $\delta_{x_i}(x) = 0$ for all i. This is a contradiction.

Therefore $f \notin \operatorname{span}(S)$, and hence $\operatorname{span}(S) \neq \mathbb{F}^X$ when X is infinite.

We conclude that $\operatorname{span}(S) = \mathbb{F}^X$ if and only if X is finite.

Definition 2.2 (Finite-dimensional vector space)**.** *A vector space V is called* finite-dimensional *if there exists a finite list of vectors that spans V.*

Remark. The examples above show that $\mathbb{F}^n$ is finite-dimensional for every positive integer n, whereas $\mathbb{F}[X]$ and $\mathbb{F}^X$ (for infinite X) are not finite-dimensional.

Exercise 2.1 (In-class exercise). *Let E be a vector space over a field $\mathbb{K}$ and let $(u_1, \dots, u_{n+1})$ be a list of $n+1$ vectors in E. Assume that $(u_1, \dots, u_{n+1})$ spans E and that*

$$u_{n+1} \in \operatorname{span}(u_1, \dots, u_n).$$

Show that $(u_1, \dots, u_n)$ also spans E.

Exercise 2.2 (In-class exercise)**.** *Let F, G, and H be three subspaces of a vector space E. Is it necessarily true that*

$$H \cap (F + G) = (H \cap F) + (H \cap G)\,?$$

Linear Algebra — Lecture 4: Linear Independence

1 Linear Independence

Suppose $v_1, \dots, v_m \in V$ and $v \in \operatorname{span}(v_1, \dots, v_m)$. By the definition of span, there exist scalars $a_1, \dots, a_m \in \mathbb{F}$ such that

$$v = a_1 v_1 + \cdots + a_m v_m.$$

We now consider the question of whether the choice of scalars in the equation above is unique. Suppose $c_1, \dots, c_m \in \mathbb{F}$ is another set of scalars such that

$$v = c_1 v_1 + \cdots + c_m v_m.$$

Subtracting the last two equations, we obtain

$$0 = (a_1 - c_1)v_1 + \cdots + (a_m - c_m)v_m.$$

Thus, we have written the zero vector as a linear combination of $v_1, \dots, v_m$. If the only way to do this is by using zero for all the scalars in the linear combination, then each $a_k - c_k = 0$, which implies $a_k = c_k$ for all k. In this case, the representation of v as a linear combination of $v_1, \dots, v_m$ is unique.

This situation is so important that we give it a special name: *linear independence.* We now make this precise.

Definition 1.1 (Linearly independent)**.** *An ordered list $v_1, \dots, v_m$ of vectors in V is called* linearly independent *if the only choice of scalars $a_1, \dots, a_m \in \mathbb{F}$ such that*

$$a_1 v_1 + \cdots + a_m v_m = 0$$

is

$$a_1 = \cdots = a_m = 0.$$

The empty list $()$ is also declared to be linearly independent.

Remark. Throughout this section, by a *list* of vectors we always mean an *ordered list.* In particular, the order in which vectors appear in a list matters.

The reasoning above shows that $v_1, \dots, v_m$ is linearly independent if and only if each vector in $\operatorname{span}(v_1, \dots, v_m)$ has a unique representation as a linear combination of $v_1, \dots, v_m$.

Examples

Example 1.1 (Standard basis vectors)**.** *Let $n \geq 1$, and let $e_1, \ldots, e_n$ denote the standard basis vectors of $\mathbb{F}^n$, where e_k has a 1 in the k-th coordinate and 0 elsewhere. To see that the list $e_1, \ldots, e_n$ is linearly independent in $\mathbb{F}^n$, suppose $a_1, \ldots, a_n \in \mathbb{F}$ satisfy*

$$a_1 e_1 + \cdots + a_n e_n = 0.$$

Then

$$(a_1, \ldots, a_n) = (0, \ldots, 0),$$

which implies

$$a_1 = \cdots = a_n = 0.$$

Hence, the vectors $e_1, \ldots, e_n$ are linearly independent.

Remark. A list of length one in a vector space is linearly independent if and only if the vector in the list is not 0.

Remark. A list of length two in a vector space is linearly independent if and only if neither of the two vectors in the list is a scalar multiple of the other.

Definition 1.2 (Linearly dependent)**.** *An ordered list of vectors in V is called* linearly dependent *if it is not linearly independent. Equivalently, a list $v_1, \ldots, v_m$ of vectors in V is linearly dependent if there exist scalars $a_1, \ldots, a_m \in \mathbb{F}$, not all zero, such that*

$$a_1 v_1 + \cdots + a_m v_m = 0.$$

Theorem 1.1 (Linear Dependence Lemma)**.** *Suppose $v_1, \ldots, v_m$ is a linearly dependent list of vectors in V. Then there exists an index $k \in \{1, \ldots, m\}$ such that*

$$v_k \in \operatorname{span}(v_1, \ldots, v_{k-1}).$$

Moreover, if such an index k is chosen and the kth vector is removed from the list $v_1, \ldots, v_m$, then the span of the remaining list is the same as the span of the original list.

Proof. Because the list $v_1, \ldots, v_m$ is linearly dependent, there exist scalars $a_1, \ldots, a_m \in \mathbb{F}$, not all zero, such that

$$a_1 v_1 + \cdots + a_m v_m = 0.$$

Let k be the largest index in $\{1, \ldots, m\}$ such that $a_k \neq 0$. Solving for v_k, we obtain

$$v_k = -\frac{a_1}{a_k} v_1 - \cdots - \frac{a_{k-1}}{a_k} v_{k-1}.$$

This shows that $v_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$.

Now suppose that k is any index such that $v_k \in \operatorname{span}(v_1, \ldots, v_{k-1})$. Then there exist scalars $b_1, \ldots, b_{k-1} \in \mathbb{F}$ such that

$$v_k = b_1 v_1 + \cdots + b_{k-1} v_{k-1}.$$

Let $u \in \operatorname{span}(v_1, \ldots, v_m)$. Then there exist scalars $c_1, \ldots, c_m \in \mathbb{F}$ such that

$$u = c_1 v_1 + \cdots + c_m v_m.$$

Substituting the expression for v_k above into this equation shows that u can be written as a linear combination of the vectors obtained by removing v_k from the list $v_1, \ldots, v_m$. Hence, removing v_k does not change the span of the list. $\square$

Theorem 1.2 (Length of a linearly independent list)**.** *Let V be a finite-dimensional vector space. Then the length of every linearly independent list of vectors in V is less than or equal to the length of every spanning list of vectors in V.*

Can you remove 5 objects from a box that initially contained only 3 objects?

Proof. Suppose that $u_1, \ldots, u_m$ is a linearly independent list in V, and suppose that $w_1, \ldots, w_n$ is a list of vectors that spans V. We must show that $m \leq n$.

We proceed inductively, performing a process with m steps. At each step, we add one of the u's to a spanning list and remove one of the w's, while preserving the span.

Step 1. Let B be the list $w_1, \ldots, w_n$, which spans V. Adjoining u_1 at the beginning of this list produces the list

$$u_1, w_1, \ldots, w_n,$$

which is linearly dependent, since $u_1 \in \operatorname{span}(w_1, \ldots, w_n)$.

By the linear dependence lemma, one of the vectors in this list is a linear combination of the preceding ones. Because the list $u_1, \ldots, u_m$ is linearly independent, we have $u_1 \neq 0$. Hence u_1 cannot be in the span of the vectors preceding it (the span of the empty list is $\{0\}$). Therefore, the linear dependence lemma implies that we can remove one of the w's so that the resulting list B (still of length n), consisting of u_1 and the remaining w's, spans V.

Step k, for $k = 2, \ldots, m$. Assume that after step $k-1$, the list B (of length n) spans V and contains $u_1, \ldots, u_{k-1}$. Because B spans V, we have $u_k \in \operatorname{span}(B)$. Thus, adjoining u_k to B produces a linearly dependent list of length $n+1$.

By the linear dependence lemma, one of the vectors in this list is a linear combination of the preceding ones. Because $u_1, \ldots, u_k$ is linearly independent, this vector cannot be one of the u's. Therefore, it must be one of the remaining w's. Removing such a w yields a new list B of length n that still spans V and now contains $u_1, \ldots, u_k$.

After step m, all vectors $u_1, \ldots, u_m$ have been added. At each step, one u is added and one w is removed. Hence there must have been at least as many w's as u's to begin with. Therefore, $m \leq n$. □

Example 1.2. *No list of length 4 is linearly independent in $\mathbb{R}^3$.*

Indeed, the list

$$(1, 0, 0),\ (0, 1, 0),\ (0, 0, 1)$$

has length 3 and spans $\mathbb{R}^3$. By the previous theorem, the length of any linearly independent list in $\mathbb{R}^3$ is less than or equal to the length of any spanning list. Hence, no list of length greater than 3 can be linearly independent in $\mathbb{R}^3$.

Linear Algebra — Lecture 7: Bases (1)

1 In-class exercises

Exercise 1.1 (A family of trigonometric functions)**.** *Define functions* $f_1, f_2, f_3, f_4 : [0, 2\pi] \to \mathbb{R}$ *by*

$$f_1(x) = \cos x, \qquad f_2(x) = x \cos x, \qquad f_3(x) = \sin x, \qquad f_4(x) = x \sin x.$$

Show that (f_1, f_2, f_3, f_4) *is a linearly independent list in the real vector space* $C([0, 2\pi], \mathbb{R})$.

Exercise 1.2 (Exponential functions)**.** *Let* $n \in \mathbb{N}$. *For each* $k \in \{0, 1, \dots, n\}$, *define* $f_k : \mathbb{R} \to \mathbb{R}$ *by*

$$f_k(x) = e^{kx}.$$

Show that the family $(f_k)_{0 \le k \le n}$ *is linearly independent in the vector space* $\mathcal{F}(\mathbb{R}, \mathbb{R})$.

2 Bases

Previously, we studied linearly independent lists and spanning lists. We now bring these two notions together by introducing the concept of a basis.

Definition 2.1 (Basis)**.** *Let* V *be a vector space. A* basis *of* V *is a list of vectors in* V *that is linearly independent and spans* V.

Example 2.1 (The standard basis of $\mathbb{F}^n$)**.** *Let* n *be a positive integer and consider the vectors defined in previous lectures*

$$e_1, e_2, \dots, e_n \in \mathbb{F}^n.$$

Let

$$\mathcal{B} := \{e_1, e_2, \dots, e_n\}.$$

We have already seen that $\mathcal{B}$ *spans* $\mathbb{F}^n$. *Moreover, the list* $(e_1, \dots, e_n)$ *is linearly independent. Therefore,* $\mathcal{B}$ *is a basis of* $\mathbb{F}^n$. *This basis is called the* standard basis *of* $\mathbb{F}^n$.

Example 2.2 (The delta functions on a finite set)**.** *Let* X *be a* ***finite*** *nonempty set, and let* $\mathbb{F}^X$ *denote the vector space of all functions from* X *to* $\mathbb{F}$. *For each* $x \in X$, *define the function* $\delta_x : X \to \mathbb{F}$ *by*

$$\delta_x(y) := \begin{cases} 1, & \text{if } y = x, \\ 0, & \text{otherwise.} \end{cases}$$

We have seen in previous lectures that the set

$$\mathcal{B} := \{\delta_x : x \in X\}$$

spans $\mathbb{F}^X$.

We now show that this set is linearly independent. Let $x_1, \ldots, x_n$ *be distinct elements of* X, *and let* $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ *be scalars such that*

$$\alpha_1 \delta_{x_1} + \cdots + \alpha_n \delta_{x_n} = 0 \quad \textit{in } \mathbb{F}^X.$$

Evaluating this equality at x_i *gives*

$$0 = (\alpha_1 \delta_{x_1} + \cdots + \alpha_n \delta_{x_n})(x_i) = \alpha_i,$$

since $\delta_{x_j}(x_i) = 0$ *for* $j \neq i$ *and* $\delta_{x_i}(x_i) = 1$. *Hence* $\alpha_i = 0$ *for each* i, *and the family* $(\delta_x)_{x \in X}$ *is linearly independent. Therefore,* $\mathcal{B}$ *is a basis of* $\mathbb{F}^X$.

Theorem 2.1 (Criterion for a basis)**.** *Let* V *be a vector space over a field* $\mathbb{F}$. *A list* $(v_1, \ldots, v_n)$ *of vectors in* V *is a basis of* V *if and only if every vector* $v \in V$ *can be written uniquely in the form*

$$v = a_1 v_1 + \cdots + a_n v_n,$$

where $a_1, \ldots, a_n \in \mathbb{F}$.

Proof. First assume that $(v_1, \ldots, v_n)$ is a basis of V. Let $v \in V$. Because $(v_1, \ldots, v_n)$ spans V, there exist scalars $a_1, \ldots, a_n \in \mathbb{F}$ such that

$$v = a_1 v_1 + \cdots + a_n v_n.$$

To prove uniqueness, suppose that

$$v = c_1 v_1 + \cdots + c_n v_n$$

for some scalars $c_1, \ldots, c_n \in \mathbb{F}$. Subtracting the two expressions gives

$$0 = (a_1 - c_1)v_1 + \cdots + (a_n - c_n)v_n.$$

Because $(v_1, \ldots, v_n)$ is linearly independent, we conclude that

$$a_1 = c_1, \ \ldots, \ a_n = c_n.$$

Hence the representation is unique.

Conversely, assume that every vector $v \in V$ can be written uniquely in the form

$$v = a_1 v_1 + \cdots + a_n v_n.$$

Taking arbitrary $v \in V$ shows that $(v_1, \ldots, v_n)$ spans V. To prove linear independence, suppose that

$$0 = a_1 v_1 + \cdots + a_n v_n.$$

By uniqueness of the representation (applied to $v = 0$), we must have $a_1 = \cdots = a_n = 0$. Thus $(v_1, \ldots, v_n)$ is linearly independent and hence a basis of V. $\square$

Linear Algebra — Lecture 8: Bases (2)

1 In-class exercises

Exercise 1.1. *Let* (v_1, v_2, v_3, v_4) *be a basis of a vector space* V*. Show that*

$$(v_1 + v_2,\ v_2 + v_3,\ v_3 + v_4,\ v_4)$$

is also a basis of V*.*

Solution. Set

$$w_1 = v_1 + v_2, \qquad w_2 = v_2 + v_3, \qquad w_3 = v_3 + v_4, \qquad w_4 = v_4.$$

Step 1: the family generates V.

We express each v_i as a linear combination of w_1, w_2, w_3, w_4. Starting from the last vector, we obtain

$$\begin{aligned} v_4 &= w_4, \\ v_3 &= w_3 - w_4, \\ v_2 &= w_2 - w_3 + w_4, \\ v_1 &= w_1 - w_2 + w_3 - w_4. \end{aligned}$$

Thus each v_i belongs to $\operatorname{span}(w_1, w_2, w_3, w_4)$. Since (v_1, v_2, v_3, v_4) generates V, the family (w_1, w_2, w_3, w_4) also generates V.

Step 2: the family is linearly independent.

Suppose that

$$a_1 w_1 + a_2 w_2 + a_3 w_3 + a_4 w_4 = 0$$

for some scalars $a_1, a_2, a_3, a_4 \in \mathbb{R}$. Substituting the definitions of the w_i and regrouping terms, we obtain

$$a_1 v_1 + (a_1 + a_2) v_2 + (a_2 + a_3) v_3 + (a_3 + a_4) v_4 = 0.$$

Since (v_1, v_2, v_3, v_4) is a linearly independent family, the only linear combination of these vectors that equals the zero vector is the trivial one. Therefore, each coefficient must vanish:

$$a_1 = 0, \qquad a_1 + a_2 = 0, \qquad a_2 + a_3 = 0, \qquad a_3 + a_4 = 0.$$

It follows that $a_1 = a_2 = a_3 = a_4 = 0$, and hence (w_1, w_2, w_3, w_4) is linearly independent.

2 Bases

Before proceeding, we take a more conceptual point of view on the notion of a basis. beyond the finite-dimensional vector spaces case.

Definition 2.1 (Linearly independent sets)**.** *From now on, during this lecture, we will work with subsets rather than lists of vectors.*

Let V be a vector space over a field $\mathbb{K}$, and let $L \subset V$ be a nonempty subset.

(i) *If L is finite, say $L = \{x_1, \ldots, x_n\}$, we say that L is* linearly independent *if*

$$\alpha_1 x_1 + \cdots + \alpha_n x_n = 0 \quad \Longrightarrow \quad \alpha_1 = \cdots = \alpha_n = 0.$$

(ii) *If L is infinite, we say that L is* linearly independent *if every finite subset of L is linearly independent in the sense of* (i)*.*

(iii) *A subset of V that is not linearly independent is called* linearly dependent*.*

Lemma 2.1 (Adding one vector to a linearly independent set)**.** *Let V be a vector space over a field $\mathbb{K}$, let $L \subset V$ be a linearly independent set, and let $x \in V$. Then*

$$x \notin \operatorname{span}(L) \quad \Longleftrightarrow \quad L \cup \{x\} \textit{ is linearly independent.}$$

Proof. We prove the contrapositive statement:

$$x \in \operatorname{span}(L) \quad \Longleftrightarrow \quad L \cup \{x\} \text{ is linearly dependent.}$$

($\Rightarrow$) If $x \in \operatorname{span}(L)$, then there exist $x_1, \ldots, x_n \in L$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{K}$ not all zero, such that

$$x = \sum_{i=1}^{n} \alpha_i x_i.$$

Rewriting, we obtain a nontrivial linear relation

$$x - \sum_{i=1}^{n} \alpha_i x_i = 0$$

among vectors in $L \cup \{x\}$, so this set is linearly dependent.

($\Leftarrow$) Conversely, suppose that $L \cup \{x\}$ is linearly dependent. By definition, there exist vectors $x_1, \ldots, x_n \in L \cup \{x\}$ and scalars $\alpha_1, \ldots, \alpha_n \in \mathbb{K}$, not all zero, such that

$$\sum_{i=1}^{n} \alpha_i x_i = 0.$$

Since L is linearly independent, at least one of the x_i must be equal to x. Choose an index i_0 such that $x_{i_0} = x$. Necessarily, $\alpha_{i_0} \neq 0$, otherwise the relation would involve only vectors from L, contradicting independence of L. Solving for x, we obtain

$$x = -\frac{1}{\alpha_{i_0}} \sum_{\substack{i=1 \\ i \neq i_0}}^{n} \alpha_i x_i \in \operatorname{span}(L).$$

Thus $x \in \operatorname{span}(L)$, completing the proof. □

Theorem 2.1 (Characterizations of a basis in arbitrary vector spaces)**.** *Let V be a vector space over a field $\mathbb{K}$, and let $B \subset V$. The following statements are equivalent:*

(i) *B is a basis of V, that is, B is linearly independent and* $\mathrm{span}(B) = V$.

(ii) *B is a* minimal generating set *of V, meaning that* $\mathrm{span}(B) = V$ *and for every subset $A \subsetneq B$ one has* $\mathrm{span}(A) \neq V$.

(iii) *B is a* maximal linearly independent set *in V, meaning that B is linearly independent and for every subset $L \subset V$ such that $B \subsetneq L$, the set L is linearly dependent.*

The proof is postponed to the next lecture.

Linear Algebra — Lecture 9: Bases (3) and Introduction to Dimension

1 Bases

Lemma 1.1 (Adding one vector to a linearly independent set). *Let V be a vector space over a field $\mathbb{K}$, let $L \subset V$ be a linearly independent set, and let $x \in V$. Then*

$$x \notin \operatorname{span}(L) \quad \Longleftrightarrow \quad L \cup \{x\} \textit{ is linearly independent.}$$

Theorem 1.1 (Characterizations of a basis in arbitrary vector spaces). *Let V be a vector space over a field $\mathbb{K}$, and let $B \subset V$. The following statements are equivalent:*

(i) *B is a basis of V, that is, B is linearly independent and* $\operatorname{span}(B) = V$.

(ii) *B is a* minimal generating set *of V, meaning that* $\operatorname{span}(B) = V$ *and for every subset $A \subsetneq B$ one has* $\operatorname{span}(A) \neq V$.

(iii) *B is a* maximal linearly independent set *in V, meaning that B is linearly independent and for every subset $L \subset V$ such that $B \subsetneq L$, the set L is linearly dependent.*

Proof. We prove (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) $\Rightarrow$ (i).

(i)$\Rightarrow$(ii). Assume B is a basis of V. In particular, $\operatorname{span}(B) = V$. Let $A \subseteq B$ be a generating set of V, i.e. $\operatorname{span}(A) = V$. We claim that $A = B$.

Indeed, suppose for contradiction that $A \neq B$. Then there exists $x \in B \setminus A$. Since A generates V, we can write

$$x = \sum_{i=1}^{n} \alpha_i x_i \quad \text{for some } x_1, \dots, x_n \in A \text{ and } \alpha_1, \dots, \alpha_n \in \mathbb{K}.$$

Rearranging gives a nontrivial linear relation among vectors in B:

$$x - \sum_{i=1}^{n} \alpha_i x_i = 0,$$

where $x \in B$ and each $x_i \in A \subseteq B$, and the coefficient of x is $1 \neq 0$. This contradicts the fact that B is linearly independent. Hence $A = B$, so B is minimal among generating sets.

(ii)$\Rightarrow$(iii). Assume that B is a minimal generating set of V.

First, B is linearly independent. Suppose, by contradiction, that B is linearly dependent. Then there exist distinct vectors $x_1, \dots, x_n \in B$ and scalars $\alpha_1, \dots, \alpha_n \in \mathbb{K}$, not all zero, such that

$$\sum_{i=1}^{n} \alpha_i x_i = 0.$$

Choose an index i_0 such that $\alpha_{i_0} \neq 0$. Then

$$x_{i_0} \in \operatorname{span}\big(B \setminus \{x_{i_0}\}\big),$$

and hence

$$\operatorname{span}(B) = \operatorname{span}\big(B \setminus \{x_{i_0}\}\big),$$

which contradicts the minimality of B as a generating set. Therefore B is linearly independent.

Next, B is maximal linearly independent. Let $L \subset V$ be a linearly independent set such that $B \subseteq L$. We claim that $L = B$. Suppose, by contradiction, that $B \subsetneq L$, and choose $x \in L \setminus B$.

Since B generates V, we have $x \in \operatorname{span}(B)$. By the previous lemma, this implies that $B \cup \{x\}$ is linearly dependent. As $B \cup \{x\} \subseteq L$, it follows that L is linearly dependent, contradicting the assumption that L is linearly independent. Hence no such x exists, and therefore $L = B$.

Thus B is a maximal linearly independent subset of V.

(iii)$\Rightarrow$(i). Assume that B is a maximal linearly independent subset of V. We show that B generates V.

Suppose, by contradiction, that $\operatorname{span}(B) \neq V$. Then there exists $x \in V$ such that $x \notin \operatorname{span}(B)$. Since B is linearly independent, the previous lemma implies that $B \cup \{x\}$ is linearly independent.

This contradicts the maximality of B, since $B \subsetneq B \cup \{x\}$. Therefore $\operatorname{span}(B) = V$, and together with linear independence, we conclude that B is a basis of V.

We now specialize to finite-dimensional vector spaces. In this context, we can prove the existence of bases and relate them to dimension.

Theorem 1.2 (Every spanning list contains a basis)**.** *Every spanning list in a vector space can be reduced to a basis of the vector space.*

Proof. Suppose $v_1, \ldots, v_n$ spans V. We want to remove some of the vectors from $v_1, \ldots, v_n$ so that the remaining vectors form a basis of V. We do this through the multistep process described below.

Start with B equal to the list $v_1, \ldots, v_n$.

Step 1. If $v_1 = 0$, then delete v_1 from B. If $v_1 \neq 0$, then leave B unchanged.

Step k. If v_k is in $\operatorname{span}(v_1, \ldots, v_{k-1})$, then delete v_k from the list B. If v_k is not in $\operatorname{span}(v_1, \ldots, v_{k-1})$, then leave B unchanged.

Stop the process after step n, getting a list B. This list B spans V because our original list spanned V and we have discarded only vectors that were already in the span of the previous vectors. The process ensures that no vector in B is in the span of the previous ones. Thus B is linearly independent, by the linear dependence lemma. Hence B is a basis of V. $\square$

Theorem 1.3 (Basis of a finite-dimensional vector space)**.** *Every finite-dimensional vector space has a basis.*

Proof. By definition, a finite-dimensional vector space has a spanning list. The previous result tells us that each spanning list can be reduced to a basis. $\square$

Exercise 1.1 (Every linearly independent list extends to a basis)**.** *Every linearly independent list of vectors in a finite-dimensional vector space can be extended to a basis of the vector space.*

Theorem 1.4 (Every subspace of V is part of a direct sum equal to V)**.** *Suppose V is finite-dimensional and U is a subspace of V. Then there exists a subspace W of V such that*

$$V = U \oplus W.$$

Proof. Since V is finite-dimensional, the subspace U is also finite-dimensional. Thus U has a basis $u_1, \dots, u_m$. This list is linearly independent in V, so it can be extended to a basis

$$u_1, \dots, u_m, w_1, \dots, w_n$$

of V. Let

$$W = \operatorname{span}(w_1, \dots, w_n).$$

We show that $V = U \oplus W$ by proving that

$$V = U + W \quad \text{and} \quad U \cap W = \{0\}.$$

First, let $v \in V$. Because the vectors $u_1, \dots, u_m, w_1, \dots, w_n$ form a basis of V, there exist scalars $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbb{F}$ such that

$$v = a_1u_1 + \cdots + a_mu_m + b_1w_1 + \cdots + b_nw_n.$$

Set

$$u = a_1u_1 + \cdots + a_mu_m \in U, \qquad w = b_1w_1 + \cdots + b_nw_n \in W.$$

Then $v = u + w$, showing that $v \in U + W$. Hence $V = U + W$.

Now suppose $v \in U \cap W$. Then there exist scalars $a_1, \dots, a_m, b_1, \dots, b_n \in \mathbb{F}$ such that

$$v = a_1u_1 + \cdots + a_mu_m = b_1w_1 + \cdots + b_nw_n.$$

Subtracting, we obtain

$$a_1u_1 + \cdots + a_mu_m - b_1w_1 - \cdots - b_nw_n = 0.$$

Because the vectors $u_1, \dots, u_m, w_1, \dots, w_n$ are linearly independent, all coefficients must be zero. Thus $v = 0$, and therefore $U \cap W = \{0\}$. □

Introduction to Dimension

Although we have been discussing finite-dimensional vector spaces, we have not yet defined the dimension of such an object. How should dimension be defined?

A reasonable definition should force the dimension of $\mathbb{F}^n$ to equal n. Notice that the standard basis

$$(1, 0, \dots, 0),\ (0, 1, 0, \dots, 0),\ \dots,\ (0, \dots, 0, 1)$$

of $\mathbb{F}^n$ has length n. Thus we are tempted to define the dimension as the length of a basis.

However, a finite-dimensional vector space in general has many different bases, and our attempted definition makes sense only if all bases in a given vector space have the same length. Fortunately, that turns out to be the case, as we now show.

Theorem 1.5 (Basis length does not depend on the basis)**.** *Any two bases of a finite-dimensional vector space have the same length.*

Proof. Suppose V is finite-dimensional, and let B_1 and B_2 be two bases of V. Because B_1 is linearly independent and B_2 spans V, the length of B_1 is at most the length of B_2. Interchanging the roles of B_1 and B_2, we also see that the length of B_2 is at most the length of B_1. Thus the two bases have the same length. □

Now that we know that any two bases of a finite-dimensional vector space have the same length, we can formally define the dimension of such spaces.

Definition 1.1 (Dimension)**.** *The dimension of a finite-dimensional vector space is the length of any basis of the vector space.*

The dimension of a finite-dimensional vector space V is denoted by $\dim V$*.*

Theorem 1.6 (Dimension of a subspace)**.** *If V is finite-dimensional and U is a subspace of V, then*

$$\dim U \leq \dim V.$$

Proof. Suppose V is finite-dimensional and U is a subspace of V. Think of a basis of U as a linearly independent list of vectors in V, and think of a basis of V as a spanning list of V. Because the length of any linearly independent list in V is at most the length of any spanning list of V, it follows that

$$\dim U \leq \dim V.$$

□

Linear Algebra — Lecture 10: Dimension

1 Dimension

Although we have been discussing finite-dimensional vector spaces, we have not yet defined the dimension of such an object. How should dimension be defined?

A reasonable definition should force the dimension of $\mathbb{F}^n$ to equal n. Notice that the standard basis

$$(1, 0, \ldots, 0),\ (0, 1, 0, \ldots, 0),\ \ldots,\ (0, \ldots, 0, 1)$$

of $\mathbb{F}^n$ has length n. Thus we are tempted to define the dimension as the length of a basis.

However, a finite-dimensional vector space in general has many different bases, and our attempted definition makes sense only if all bases in a given vector space have the same length. Fortunately, that turns out to be the case, as we now show.

Theorem 1.1 (Basis length does not depend on the basis). *Any two bases of a finite-dimensional vector space have the same length.*

Proof. Suppose V is finite-dimensional, and let B_1 and B_2 be two bases of V. Because B_1 is linearly independent and B_2 spans V, the length of B_1 is at most the length of B_2. Interchanging the roles of B_1 and B_2, we also see that the length of B_2 is at most the length of B_1. Thus the two bases have the same length. □

Now that we know that any two bases of a finite-dimensional vector space have the same length, we can formally define the dimension of such spaces.

Definition 1.1 (Dimension). *The dimension of a finite-dimensional vector space is the length of any basis of the vector space. The dimension of a finite-dimensional vector space V is denoted by* $\dim V$.

Theorem 1.2 (Dimension of a subspace). *If V is finite-dimensional and U is a subspace of V, then*

$$\dim U \leq \dim V.$$

Proof. Suppose V is finite-dimensional and U is a subspace of V. Think of a basis of U as a linearly independent list of vectors in V, and think of a basis of V as a spanning list of V. Because the length of any linearly independent list in V is at most the length of any spanning list of V, it follows that

$$\dim U \leq \dim V.$$

□

Theorem 1.3 (Linearly independent list of the right length is a basis). *Suppose V is finite-dimensional. Then every linearly independent list of vectors in V of length* $\dim V$ *is a basis of V.*

Proof. Suppose $\dim V = n$ and $v_1, \ldots, v_n$ is linearly independent in V. The list $v_1, \ldots, v_n$ can be extended to a basis of V . However, every basis of V has length n, so in this case the extension is the trivial one, meaning that no elements are adjoined to $v_1, \ldots, v_n$. Thus $v_1, \ldots, v_n$ is a basis of V, as desired. $\square$

The next result is a useful consequence of the previous result.

Theorem 1.4 (Subspace of full dimension equals the whole space)**.** *Suppose that V is finite-dimensional and U is a subspace of V such that*

$$\dim U = \dim V.$$

Then $U = V$.

Proof. Let $u_1, \ldots, u_n$ be a basis of U. Thus $n = \dim U$, and by hypothesis we also have $n = \dim V$. Thus $u_1, \ldots, u_n$ is a linearly independent list of vectors in V (because it is a basis of U) of length $\dim V$. From previous Theorem, we see that $u_1, \ldots, u_n$ is a basis of V. In particular, every vector in V is a linear combination of $u_1, \ldots, u_n$. Thus $U = V$. $\square$

Remark (The dimension depends on the field). The real vector space $\mathbb{R}^2$ has dimension 2, whereas the complex vector space $\mathbb{C}$ has dimension 1. As sets, $\mathbb{R}^2$ can be identified with $\mathbb{C}$ (and addition is the same on both spaces, as is scalar multiplication by real numbers). Thus, when we talk about the dimension of a vector space, the role played by the choice of the field $\mathbb{F}$ cannot be neglected.

In class-exercise Let $n \geq 1$. What is $\dim_{\mathbb{R}}(\mathbb{C}^n)$? What is $\dim_{\mathbb{C}}(\mathbb{C}^n)$?

Theorem 1.5 (Spanning list of the right length is a basis)**.** *Suppose V is finite-dimensional. Then every spanning list of vectors in V of length* $\dim V$ *is a basis of V.*

Proof. Suppose $\dim V = n$ and $v_1, \ldots, v_n$ spans V. The list $v_1, \ldots, v_n$ can be reduced to a basis of V (by Theorem 2.30). However, every basis of V has length n, so in this case the reduction is the trivial one, meaning that no elements are deleted from $v_1, \ldots, v_n$. Thus $v_1, \ldots, v_n$ is a basis of V, as desired. $\square$

In-class exercise. Let E be a real vector space of finite dimension n, and let U, V, W be subspaces of E.

(a) Assume $\dim U + \dim V > n$. Show that $U \cap V \neq \{0\}$.

(b) Assume $\dim U + \dim V + \dim W > 2n$. What can you say about $U \cap V \cap W$?

Theorem 1.6 (Dimension of a sum)**.** *If V_1 and V_2 are subspaces of a finite-dimensional vector space, then*

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2).$$

Solution. (a) Since $U + V \subseteq E$, we have $\dim(U + V) \leq n$. Using the dimension formula,

$$\dim(U + V) = \dim U + \dim V - \dim(U \cap V) \leq n.$$

Hence

$$\dim(U \cap V) \geq \dim U + \dim V - n > 0,$$

so $U \cap V \neq \{0\}$.

(b) Set $W' := U \cap V$. By the same formula,

$$\dim W' = \dim(U \cap V) = \dim U + \dim V - \dim(U + V) \geq \dim U + \dim V - n.$$

Therefore,

$$\dim W' + \dim W \geq (\dim U + \dim V - n) + \dim W = \dim U + \dim V + \dim W - n > n.$$

Applying part (a) to the subspaces W' and W (inside E) gives

$$W' \cap W \neq \{0\}.$$

But $W' \cap W = (U \cap V) \cap W = U \cap V \cap W$, so

$$U \cap V \cap W \neq \{0\}.$$

□

Proof. Let $v_1, \dots, v_m$ be a basis of $V_1 \cap V_2$; thus $\dim(V_1 \cap V_2) = m$. Because $v_1, \dots, v_m$ is a basis of $V_1 \cap V_2$, it is linearly independent in V_1. Hence this list can be extended to a basis

$$v_1, \dots, v_m, u_1, \dots, u_j$$

of V_1 . Thus $\dim V_1 = m + j$. Also extend $v_1, \dots, v_m$ to a basis

$$v_1, \dots, v_m, w_1, \dots, w_k$$

of V_2; thus $\dim V_2 = m + k$.

We will show that

$$v_1, \dots, v_m, u_1, \dots, u_j, w_1, \dots, w_k \tag{1}$$

is a basis of $V_1 + V_2$. This will complete the proof, because then we will have

$$\dim(V_1 + V_2) = m + j + k = (m + j) + (m + k) - m = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2).$$

The list (1) is contained in $V_1 \cup V_2$ and thus is contained in $V_1 + V_2$. The span of this list contains V_1 and contains V_2 and hence is equal to $V_1 + V_2$. Thus to show that (1) is a basis of $V_1 + V_2$ we only need to show that it is linearly independent.

To prove that (1) is linearly independent, suppose

$$a_1 v_1 + \cdots + a_m v_m + b_1 u_1 + \cdots + b_j u_j + c_1 w_1 + \cdots + c_k w_k = 0,$$

where all the a's, b's, and c's are scalars. We need to prove that all the a's, b's, and c's equal 0. The equation above can be rewritten as

$$c_1 w_1 + \cdots + c_k w_k = -a_1 v_1 - \cdots - a_m v_m - b_1 u_1 - \cdots - b_j u_j. \tag{2}$$

This shows that $c_1 w_1 + \cdots + c_k w_k \in V_1$. All the w's are in V_2, so this implies that $c_1 w_1 + \cdots + c_k w_k \in V_1 \cap V_2$. Because $v_1, \dots, v_m$ is a basis of $V_1 \cap V_2$, we have

$$c_1 w_1 + \cdots + c_k w_k = d_1 v_1 + \cdots + d_m v_m$$

for some scalars $d_1, \dots, d_m$. But $v_1, \dots, v_m, w_1, \dots, w_k$ is linearly independent, so the last equation implies that all the c's (and d's) equal 0. Thus (2) becomes the equation

$$a_1 v_1 + \cdots + a_m v_m + b_1 u_1 + \cdots + b_j u_j = 0.$$

Because the list $v_1, \dots, v_m, u_1, \dots, u_j$ is linearly independent, this equation implies that all the a's and b's are 0, completing the proof. □

Linear Algebra — Lecture 11: Dimension (Conclusion) and Linear Maps (I)

Theorem 0.1 (Dimension of a sum)**.** *If V_1 and V_2 are subspaces of a finite-dimensional vector space V, then*

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2).$$

Proof. Let $v_1, \dots, v_m$ be a basis of $V_1 \cap V_2$; thus $\dim(V_1 \cap V_2) = m$. Because $v_1, \dots, v_m$ is a basis of $V_1 \cap V_2$, it is linearly independent in V_1. Hence this list can be extended to a basis

$$v_1, \dots, v_m, u_1, \dots, u_j$$

of V_1 . Thus $\dim V_1 = m + j$. Also extend $v_1, \dots, v_m$ to a basis

$$v_1, \dots, v_m, w_1, \dots, w_k$$

of V_2; thus $\dim V_2 = m + k$.

We will show that

$$v_1, \dots, v_m, u_1, \dots, u_j, w_1, \dots, w_k \tag{1}$$

is a basis of $V_1 + V_2$. This will complete the proof, because then we will have

$$\dim(V_1 + V_2) = m + j + k = (m + j) + (m + k) - m = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2).$$

The list (1) is contained in $V_1 \cup V_2$ and thus is contained in $V_1 + V_2$. The span of this list contains V_1 and contains V_2 and hence is equal to $V_1 + V_2$. Thus to show that (1) is a basis of $V_1 + V_2$ we only need to show that it is linearly independent.

To prove that (1) is linearly independent, suppose

$$a_1v_1 + \cdots + a_mv_m + b_1u_1 + \cdots + b_ju_j + c_1w_1 + \cdots + c_kw_k = 0,$$

where all the a's, b's, and c's are scalars. We need to prove that all the a's, b's, and c's equal 0. The equation above can be rewritten as

$$c_1w_1 + \cdots + c_kw_k = -a_1v_1 - \cdots - a_mv_m - b_1u_1 - \cdots - b_ju_j. \tag{2}$$

This shows that $c_1w_1 + \cdots + c_kw_k \in V_1$. All the w's are in V_2, so this implies that $c_1w_1 + \cdots + c_kw_k \in V_1 \cap V_2$. Because $v_1, \dots, v_m$ is a basis of $V_1 \cap V_2$, we have

$$c_1w_1 + \cdots + c_kw_k = d_1v_1 + \cdots + d_mv_m$$

for some scalars $d_1, \dots, d_m$. But $v_1, \dots, v_m, w_1, \dots, w_k$ is linearly independent, so the last equation implies that all the c's (and d's) equal 0. Thus (2) becomes the equation

$$a_1v_1 + \cdots + a_mv_m + b_1u_1 + \cdots + b_ju_j = 0.$$

Because the list $v_1, \dots, v_m, u_1, \dots, u_j$ is linearly independent, this equation implies that all the a's and b's are 0, completing the proof. □

1 Linear functions

2.1.1

Let V and W be vector spaces over a field F. We say that a function $f : V \to W$ is *linear*, or a *vector space homomorphism*, if

$$f(x+y) = f(x) + f(y), \qquad f(\lambda x) = \lambda f(x),$$

for every $x, y \in V$ and every $\lambda \in F$. If moreover $W = V$, we say that f is an *endomorphism* of V.

2.1.2 Examples

(a) If V is a vector space, then the identity function

$$\mathrm{id}_V : x \mapsto x$$

is linear. More generally, if U is a subspace of V, the inclusion $U \hookrightarrow V$ is linear.

(b) If V is a vector space, the constant function

$$x \mapsto 0$$

is linear.

(c) Let X be a set and let F^X be the vector space of all functions $X \to F$. If $x \in X$, the evaluation map

$$f \mapsto f(x)$$

is linear.

(d) Let $C(\mathbb{R})$ be the vector space of continuous functions $\mathbb{R} \to \mathbb{R}$ and $C^1(\mathbb{R})$ the space of continuously differentiable functions. The map

$$f \in C^1(\mathbb{R}) \mapsto f' \in C(\mathbb{R})$$

is linear.

(e) Let V be finite-dimensional with ordered basis $\mathcal{B} = (x_1, \ldots, x_n)$. For $x \in V$, we denote by

$$[x]_{\mathcal{B}} = (c_1, \ldots, c_n) \in F^n$$

the coordinate vector of x with respect to $\mathcal{B}$, i.e., the unique scalars $c_1, \ldots, c_n \in F$ such that

$$x = c_1 x_1 + \cdots + c_n x_n.$$

The coordinate map

$$x \in V \mapsto [x]_{\mathcal{B}} \in F^n$$

is linear.

Remark. Let $f : V \to W$ be a linear map. Then

$$f(0) = 0.$$

Indeed, by additivity,

$$f(0) = f(0+0) = f(0) + f(0).$$

Adding the additive inverse of $f(0)$ to both sides yields $f(0) = 0$.

Proposition 1.1. *Let V be a finite-dimensional vector space of dimension n and let $\mathcal{B} = \{x_1, \dots, x_n\}$ be a basis of V. If W is a vector space and $y_1, \dots, y_n \in W$ are n elements of W, then there exists a unique linear map $f : V \to W$ such that*

$$f(x_i) = y_i \quad \text{for each } i \in [n].$$

Proof. Let W be a vector space and let $y_1, \dots, y_n \in W$. We construct a function $f : V \to W$ as follows: if $x \in V$, then there exist uniquely determined scalars $a_1, \dots, a_n \in F$ such that

$$x = a_1 x_1 + \cdots + a_n x_n,$$

and we define

$$f(x) = a_1 y_1 + \cdots + a_n y_n.$$

Let us check that the function obtained in this way is linear and satisfies the condition in the statement.

- Suppose $x, x' \in V$. Let $a_1, \dots, a_n, b_1, \dots, b_n \in F$ be scalars such that

$$x = a_1 x_1 + \cdots + a_n x_n \quad \text{and} \quad x' = b_1 x_1 + \cdots + b_n x_n.$$

By the definition of f, we have

$$f(x) = a_1 y_1 + \cdots + a_n y_n, \quad f(x') = b_1 y_1 + \cdots + b_n y_n,$$

and therefore

$$f(x) + f(x') = (a_1 y_1 + \cdots + a_n y_n) + (b_1 y_1 + \cdots + b_n y_n) = (a_1 + b_1) y_1 + \cdots + (a_n + b_n) y_n. \tag{1}$$

On the other hand,

$$x + x' = (a_1 x_1 + \cdots + a_n x_n) + (b_1 x_1 + \cdots + b_n x_n) = (a_1 + b_1) x_1 + \cdots + (a_n + b_n) x_n,$$

so the definition of f gives

$$f(x + x') = (a_1 + b_1) y_1 + \cdots + (a_n + b_n) y_n. \tag{2}$$

Since the right-hand sides of (1) and (2) are equal, we obtain

$$f(x + x') = f(x) + f(x').$$

- Now let $\lambda \in F$ and $x \in V$. Let $a_1, \dots, a_n \in F$ be such that

$$x = a_1 x_1 + \cdots + a_n x_n,$$

so that

$$f(x) = a_1 y_1 + \cdots + a_n y_n.$$

Since

$$\lambda x = \lambda a_1 x_1 + \cdots + \lambda a_n x_n,$$

we obtain

$$f(\lambda x) = \lambda a_1 y_1 + \cdots + \lambda a_n y_n = \lambda(a_1 y_1 + \cdots + a_n y_n) = \lambda f(x).$$

- Finally, if $i \in [n]$, then

$$x_i = 0x_1 + \cdots + 0x_{i-1} + 1x_i + 0x_{i+1} + \cdots + 0x_n,$$

so the definition of f gives

$$f(x_i) = 0y_1 + \cdots + 0y_{i-1} + 1y_i + 0y_{i+1} + \cdots + 0y_n = y_i.$$

This proves existence. We now prove uniqueness.

Suppose $g, g' : V \to W$ are linear maps such that $g(x_i) = y_i$ and $g'(x_i) = y_i$ for each $i \in [n]$. If $x \in V$, there exist scalars $a_1, \ldots, a_n \in F$ such that

$$x = a_1x_1 + \cdots + a_nx_n.$$

Since g is linear,

$$g(x) = g(a_1x_1 + \cdots + a_nx_n) = a_1g(x_1) + \cdots + a_ng(x_n) = a_1y_1 + \cdots + a_ny_n.$$

Similarly,

$$g'(x) = a_1y_1 + \cdots + a_ny_n.$$

Hence $g(x) = g'(x)$ for all $x \in V$, and therefore $g = g'$. □

2 Image and preimage

2.2.1

Let V and W be vector spaces and let $f : V \to W$ be a linear map. If U is a subspace of V, the *image of* U *under* f is the subset

$$f(U) = \{\, f(x) : x \in U \,\}$$

of W.

On the other hand, if U is a subspace of W, the *preimage of* U *under* f is the subset

$$f^{-1}(U) = \{\, x \in V : f(x) \in U \,\}$$

of V.

In particular, we call *the image of* f and write $\mathrm{Im}(f)$ the set $f(V)$, and we call *the kernel of* f and write $\mathrm{Ker}(f)$ the preimage $f^{-1}(0)$ of the zero subspace of W.

2.2.2 Proposition

Let V and W be vector spaces and let $f : V \to W$ be a linear map.

(i) If U is a subspace of W, then the preimage $f^{-1}(U)$ is a subspace of V. In particular, the kernel $\mathrm{Ker}(f)$ of f is a subspace of V.

(ii) If U is a subspace of V, then the image $f(U)$ is a subspace of W. In particular, the image $\mathrm{Im}(f)$ of f is a subspace of W.

Proof. (i) Let U be a subspace of W. We verify that $f^{-1}(U)$ is a subspace of V.

- Since $f(0) = 0 \in U$, we have $0 \in f^{-1}(U)$.
- If $x, y \in f^{-1}(U)$, then $f(x) \in U$ and $f(y) \in U$. Since U is a subspace of W,

  $$f(x + y) = f(x) + f(y) \in U,$$

  so $x + y \in f^{-1}(U)$.
- If $\lambda \in F$ and $x \in f^{-1}(U)$, then

  $$f(\lambda x) = \lambda f(x) \in U,$$

  so $\lambda x \in f^{-1}(U)$.

(ii) Now let U be a subspace of V.

- Since $0 \in U$, we have $0 = f(0) \in f(U)$.
- If $x, y \in f(U)$, then there exist $v, w \in U$ such that $x = f(v)$ and $y = f(w)$. Then

  $$x + y = f(v) + f(w) = f(v + w) \in f(U),$$

  since $v + w \in U$.
- If $\lambda \in F$ and $x \in f(U)$, then there exists $v \in U$ such that $x = f(v)$. Then

  $$\lambda x = \lambda f(v) = f(\lambda v) \in f(U),$$

  since $\lambda v \in U$.

This shows that $f(U)$ is a subspace of W. $\square$

Linear Algebra — Lecture 12: Linear Maps (II)

1 Image and preimage

2.2.1

Let V and W be vector spaces and let $f : V \to W$ be a linear map. If U is a subspace of V, the *image of* U *under* f is the subset

$$f(U) = \{\, f(x) : x \in U \,\}$$

of W.

On the other hand, if U is a subspace of W, the *preimage of* U *under* f is the subset

$$f^{-1}(U) = \{\, x \in V : f(x) \in U \,\}$$

of V.

In particular, we call *the image of* f and write $\mathrm{Im}(f)$ the set $f(V)$, and we call *the kernel of* f and write $\mathrm{Ker}(f)$ the preimage $f^{-1}(0)$ of the zero subspace of W.

2.2.2 Proposition

Let V and W be vector spaces and let $f : V \to W$ be a linear map.

(i) If U is a subspace of W, then the preimage $f^{-1}(U)$ is a subspace of V. In particular, the kernel $\mathrm{Ker}(f)$ of f is a subspace of V.

(ii) If U is a subspace of V, then the image $f(U)$ is a subspace of W. In particular, the image $\mathrm{Im}(f)$ of f is a subspace of W.

Proof. (i) Let U be a subspace of W. We verify that $f^{-1}(U)$ is a subspace of V.

- Since $f(0) = 0 \in U$, we have $0 \in f^{-1}(U)$.
- If $x, y \in f^{-1}(U)$, then $f(x) \in U$ and $f(y) \in U$. Since U is a subspace of W,

  $$f(x + y) = f(x) + f(y) \in U,$$

  so $x + y \in f^{-1}(U)$.
- If $\lambda \in F$ and $x \in f^{-1}(U)$, then

  $$f(\lambda x) = \lambda f(x) \in U,$$

  so $\lambda x \in f^{-1}(U)$.

(ii) Now let U be a subspace of V.

- Since $0 \in U$, we have $0 = f(0) \in f(U)$.
- If $x, y \in f(U)$, then there exist $v, w \in U$ such that $x = f(v)$ and $y = f(w)$. Then

$$x + y = f(v) + f(w) = f(v + w) \in f(U),$$

  since $v + w \in U$.
- If $\lambda \in F$ and $x \in f(U)$, then there exists $v \in U$ such that $x = f(v)$. Then

$$\lambda x = \lambda f(v) = f(\lambda v) \in f(U),$$

  since $\lambda v \in U$.

This shows that $f(U)$ is a subspace of W. $\square$

In-class exercise 1 (Even and odd decomposition)**.** *Let $\mathcal{F}(\mathbb{R}, \mathbb{R})$ be the vector space of all functions $f : \mathbb{R} \to \mathbb{R}$. Define*

$$\mathcal{P} = \{f : \mathbb{R} \to \mathbb{R} : \ \forall x \in \mathbb{R}, \ f(x) = f(-x)\} \quad \text{and} \quad \mathcal{I} = \{f : \mathbb{R} \to \mathbb{R} : \ \forall x \in \mathbb{R}, \ f(x) = -f(-x)\}.$$

Show that $\mathcal{P}$ and $\mathcal{I}$ are complementary subspaces of $\mathcal{F}(\mathbb{R}, \mathbb{R})$, i.e.

$$\mathcal{F}(\mathbb{R}, \mathbb{R}) = \mathcal{P} \oplus \mathcal{I}.$$

Solution. Step 1: $\mathcal{P}$ and $\mathcal{I}$ are subspaces. *If $f, g \in \mathcal{P}$ and $\alpha, \beta \in \mathbb{R}$, then for all x,*

$$(\alpha f + \beta g)(-x) = \alpha f(-x) + \beta g(-x) = \alpha f(x) + \beta g(x) = (\alpha f + \beta g)(x),$$

so $\alpha f + \beta g \in \mathcal{P}$. Similarly, if $f, g \in \mathcal{I}$, then

$$(\alpha f + \beta g)(-x) = \alpha f(-x) + \beta g(-x) = -\alpha f(x) - \beta g(x) = -(\alpha f + \beta g)(x),$$

so $\alpha f + \beta g \in \mathcal{I}$.

Step 2: the intersection is trivial. *If $f \in \mathcal{P} \cap \mathcal{I}$, then for all x we have $f(x) = f(-x)$ and also $f(x) = -f(-x)$. Hence $f(x) = -f(x)$, so $f(x) = 0$ for all x, i.e. $f = 0$. Therefore $\mathcal{P} \cap \mathcal{I} = \{0\}$.*

Step 3: every function splits as even + odd. *Let $f \in \mathcal{F}(\mathbb{R}, \mathbb{R})$ and define*

$$g(x) = \frac{1}{2}\big(f(x) + f(-x)\big), \qquad h(x) = \frac{1}{2}\big(f(x) - f(-x)\big).$$

Then for all x,

$$g(-x) = \frac{1}{2}\big(f(-x) + f(x)\big) = g(x),$$

so $g \in \mathcal{P}$, and

$$h(-x) = \frac{1}{2}\big(f(-x) - f(x)\big) = -h(x),$$

so $h \in \mathcal{I}$. Finally,

$$g(x) + h(x) = \frac{1}{2}\big(f(x) + f(-x)\big) + \frac{1}{2}\big(f(x) - f(-x)\big) = f(x),$$

so $f = g + h \in \mathcal{P} + \mathcal{I}$.

Since $\mathcal{P} \cap \mathcal{I} = \{0\}$ and $\mathcal{P} + \mathcal{I} = \mathcal{F}(\mathbb{R}, \mathbb{R})$, we conclude $\mathcal{F}(\mathbb{R}, \mathbb{R}) = \mathcal{P} \oplus \mathcal{I}$.

In-class exercise 2 (Decomposing with a non-kernel vector)**.** *Suppose* $\varphi \in \mathcal{L}(V, \mathbb{F})$ *and* $\varphi \neq 0$. *Let* $u \in V$ *be such that* $u \notin \ker(\varphi)$ *(equivalently,* $\varphi(u) \neq 0$*). Prove that*

$$V = \ker(\varphi) \;\oplus\; \{au : \; a \in \mathbb{F}\}.$$

Solution. *Let* $U = \{au : \; a \in \mathbb{F}\} = (u)$.

Step 1: $\ker(\varphi) \cap U = \{0\}$. *If* $v \in \ker(\varphi) \cap U$*, then* $v = au$ *for some* $a \in \mathbb{F}$ *and also* $\varphi(v) = 0$. *Thus*

$$0 = \varphi(v) = \varphi(au) = a\,\varphi(u).$$

Since $\varphi(u) \neq 0$*, we must have* $a = 0$*, hence* $v = 0$.

Step 2: $V = \ker(\varphi) + U$. *Take any* $x \in V$ *and set*

$$a = \frac{\varphi(x)}{\varphi(u)} \in \mathbb{F}, \qquad y = x - au.$$

Then

$$\varphi(y) = \varphi(x - au) = \varphi(x) - a\,\varphi(u) = \varphi(x) - \frac{\varphi(x)}{\varphi(u)}\varphi(u) = 0,$$

so $y \in \ker(\varphi)$. *Hence* $x = y + au \in \ker(\varphi) + U$.

Since we have both $V = \ker(\varphi) + U$ *and* $\ker(\varphi) \cap U = \{0\}$*, it follows that*

$$V = \ker(\varphi) \oplus U = \ker(\varphi) \oplus \{au : \; a \in \mathbb{F}\}.$$

Definition 1.1 (Injective)**.** *A function* $T : V \to W$ *is called* injective *if*

$$Tu = Tv \text{ implies } u = v.$$

Proposition 1.1 (Injectivity $\Leftrightarrow$ kernel equals $\{0\}$)**.** *Let* $T \in \mathcal{L}(V, W)$. *Then* T *is injective if and only if*

$$\ker T = \{0\}.$$

Proof. First suppose T is injective. We want to prove that $\ker T = \{0\}$. We already know that $\{0\} \subseteq \ker T$. To prove the inclusion in the other direction, suppose $v \in \ker T$. Then

$$T(v) = 0 = T(0).$$

Because T is injective, this implies $v = 0$. Thus $\ker T = \{0\}$.

To prove the implication in the other direction, now suppose $\ker T = \{0\}$. We want to prove that T is injective. Suppose $u, v \in V$ and $Tu = Tv$. Then

$$0 = Tu - Tv = T(u - v).$$

Thus $u - v \in \ker T = \{0\}$. Hence $u - v = 0$, which implies $u = v$. Therefore T is injective. $\square$

Definition 1.2 (Surjective)**.** *A function* $T : V \to W$ *is called* surjective *if its range equals* W.

Theorem 1.1 (Fundamental theorem of linear maps)**.** *Suppose* V *is finite-dimensional and* $T \in \mathcal{L}(V, W)$. *Then* $\operatorname{Im}(T)$ *is finite-dimensional and*

$$\dim V = \dim \ker T + \dim \operatorname{Im}(T).$$

Proof. Let $u_1, \ldots, u_m$ be a basis of $\ker T$; thus $\dim \ker T = m$. The linearly independent list $u_1, \ldots, u_m$ can be extended to a basis

$$u_1, \ldots, u_m, v_1, \ldots, v_n$$

of V. Thus $\dim V = m + n$.

To complete the proof, we show that $\mathrm{Im}(T)$ is finite-dimensional and $\dim \mathrm{Im}(T) = n$ by proving that

$$Tv_1, \ldots, Tv_n$$

is a basis of $\mathrm{Im}(T)$.

Let $v \in V$. Because $u_1, \ldots, u_m, v_1, \ldots, v_n$ spans V, we can write

$$v = a_1u_1 + \cdots + a_mu_m + b_1v_1 + \cdots + b_nv_n,$$

where the a's and b's are in $\mathbb{F}$. Applying T to both sides gives

$$Tv = b_1Tv_1 + \cdots + b_nTv_n,$$

because each $u_k \in \ker T$. This shows that $Tv_1, \ldots, Tv_n$ spans $\mathrm{Im}(T)$. In particular, $\mathrm{Im}(T)$ is finite-dimensional.

To show linear independence, suppose $c_1, \ldots, c_n \in \mathbb{F}$ and

$$c_1Tv_1 + \cdots + c_nTv_n = 0.$$

Then

$$T(c_1v_1 + \cdots + c_nv_n) = 0,$$

so

$$c_1v_1 + \cdots + c_nv_n \in \ker T.$$

Because $u_1, \ldots, u_m$ spans $\ker T$, we can write

$$c_1v_1 + \cdots + c_nv_n = d_1u_1 + \cdots + d_mu_m,$$

where the d's are in $\mathbb{F}$. Since

$$u_1, \ldots, u_m, v_1, \ldots, v_n$$

is linearly independent, all coefficients must be zero. Thus $c_1 = \cdots = c_n = 0$, so $Tv_1, \ldots, Tv_n$ is linearly independent.

Therefore $Tv_1, \ldots, Tv_n$ is a basis of $\mathrm{Im}(T)$, and $\dim \mathrm{Im}(T) = n$. Since $\dim V = m + n$ and $\dim \ker T = m$, we conclude that

$$\dim V = \dim \ker T + \dim \mathrm{Im}(T).$$

□

Linear Algebra — Lecture 13: Linear Maps (III)

Corollary 0.1 (Linear map to a lower-dimensional space is not injective)**.** *Suppose V and W are finite-dimensional vector spaces such that*

$$\dim V > \dim W.$$

Then no linear map from V to W is injective.

Proof. Let $T \in \mathcal{L}(V, W)$. By the fundamental theorem of linear maps,

$$\dim \ker T = \dim V - \dim \operatorname{Im}(T).$$

Because $\operatorname{Im}(T) \subseteq W$, we have

$$\dim \operatorname{Im}(T) \leq \dim W.$$

Thus

$$\dim \ker T \geq \dim V - \dim W > 0.$$

Hence $\ker T$ contains nonzero vectors, so T is not injective. □

Example 0.1. *Define a linear map $T : \mathbb{F}^4 \to \mathbb{F}^3$ by*

$$T(z_1, z_2, z_3, z_4) = \left(\sqrt{7}\, z_1 + \pi z_2 + z_4,\ 9z_1 + 3z_2 + 2z_3,\ z_2 + 6z_3 + 7z_4\right).$$

Because $\dim \mathbb{F}^4 > \dim \mathbb{F}^3$, the previous corollary implies that T is not injective.

Corollary 0.2 (Linear map to a higher-dimensional space is not surjective)**.** *Suppose V and W are finite-dimensional vector spaces such that*

$$\dim V < \dim W.$$

Then no linear map from V to W is surjective.

Proof. Let $T \in \mathcal{L}(V, W)$. By the fundamental theorem of linear maps,

$$\dim \operatorname{Im}(T) = \dim V - \dim \ker T.$$

Hence

$$\dim \operatorname{Im}(T) \leq \dim V < \dim W.$$

Therefore $\operatorname{Im}(T) \neq W$, so T is not surjective. □

Corollary 0.3 (Injective, surjective, bijective in equal dimension)**.** *Let E and F be finite-dimensional vector spaces such that*

$$\dim(E) = \dim(F),$$

and let $f : E \to F$ be linear. The following are equivalent:

1. f is injective,

2. f is surjective,

3. f is bijective.

Proof. (i) $\Rightarrow$ (ii). Suppose f is injective. Then

$$\ker(f) = \{0\}, \quad \text{so} \quad \dim(\ker(f)) = 0.$$

By the fundamental theorem of linear maps,

$$\dim(E) = \dim(\ker(f)) + \dim(\mathrm{Im}(f)).$$

Hence

$$\dim(E) = \dim(\mathrm{Im}(f)).$$

Because $\mathrm{Im}(f) \subseteq F$ and $\dim(E) = \dim(F)$, we obtain

$$\mathrm{Im}(f) = F,$$

so f is surjective.

(ii) $\Rightarrow$ (iii). Suppose f is surjective. Then

$$\mathrm{Im}(f) = F, \quad \text{so} \quad \dim(\mathrm{Im}(f)) = \dim(F) = \dim(E).$$

Again using rank–nullity,

$$\dim(\ker(f)) = \dim(E) - \dim(\mathrm{Im}(f)) = 0.$$

Thus $\ker(f) = \{0\}$, so f is injective. Therefore f is bijective.

(iii) $\Rightarrow$ (i). This is immediate. □

Remark. The equivalences in the corollary above are no longer true in infinite-dimensional vector spaces.

Example 0.2 (Derivative map)**.** *Let K be a field and consider the vector space $K[X]$ of polynomials. Define*

$$f : K[X] \to K[X], \qquad f(P) = P',$$

where P' is the derivative of P.

Then f is linear and surjective but not injective. Indeed,

$$\mathrm{Im}(f) = K[X], \qquad \ker(f) = K.$$

Example 0.3 (Multiplication by X)**.** *Define*

$$g : K[X] \to K[X], \qquad g(P) = XP.$$

Then g is linear and injective but not surjective. Moreover,

$$\ker(g) = \{0\}, \qquad \mathrm{Im}(g) = \{P \in K[X] : P(0) = 0\}.$$

1 Linear systems and linear maps

We now connect systems of linear equations with linear maps.

Homogeneous systems

Fix positive integers m and n, and let $A_{j,k} \in \mathbb{F}$. Consider the homogeneous system of m equations in n variables

$$\begin{aligned} A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,n}x_n &= 0, \\ A_{2,1}x_1 + A_{2,2}x_2 + \cdots + A_{2,n}x_n &= 0, \\ &\vdots \\ A_{m,1}x_1 + A_{m,2}x_2 + \cdots + A_{m,n}x_n &= 0. \end{aligned}$$

Clearly $x_1 = \cdots = x_n = 0$ is a solution. The question is whether nonzero solutions exist.

Define a linear map

$$T : \mathbb{F}^n \to \mathbb{F}^m$$

by

$$T(x_1, \ldots, x_n) = \big(A_{1,1}x_1 + \cdots + A_{1,n}x_n,\ A_{2,1}x_1 + \cdots + A_{2,n}x_n,\ \ldots,\ A_{m,1}x_1 + \cdots + A_{m,n}x_n\big).$$

Then solving the homogeneous system is the same as solving

$$T(x_1, \ldots, x_n) = 0.$$

Thus the set of solutions equals $\ker(T)$.

Theorem 1.1 (Homogeneous system with more variables than equations)**.** *A homogeneous system of m linear equations in n variables, with $n > m$, has a nonzero solution.*

Proof. Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be the linear map associated to the system. If $n > m$, the previous corollary implies that T is not injective. Hence

$$\ker(T) \neq \{0\},$$

so the homogeneous system has a nonzero solution. □

Non-homogeneous systems

Now consider constants $c_1, \ldots, c_m \in \mathbb{F}$ and the system

$$\begin{aligned} A_{1,1}x_1 + A_{1,2}x_2 + \cdots + A_{1,n}x_n &= c_1, \\ A_{2,1}x_1 + A_{2,2}x_2 + \cdots + A_{2,n}x_n &= c_2, \\ &\vdots \\ A_{m,1}x_1 + A_{m,2}x_2 + \cdots + A_{m,n}x_n &= c_m. \end{aligned}$$

Using the same linear map $T : \mathbb{F}^n \to \mathbb{F}^m$ defined by

$$T(x_1, \ldots, x_n) = \big(A_{1,1}x_1 + \cdots + A_{1,n}x_n,\ A_{2,1}x_1 + \cdots + A_{2,n}x_n,\ \ldots,\ A_{m,1}x_1 + \cdots + A_{m,n}x_n\big),$$

the system can be written as

$$T(x_1, \ldots, x_n) = (c_1, \ldots, c_m).$$

Thus solving the system is equivalent to asking whether

$$(c_1, \ldots, c_m) \in \mathrm{Im}(T).$$

Theorem 1.2 (More equations than variables). *If a system of linear equations has more equations than variables, then for some choice of constants the system has no solution.*

Proof. Let $T : \mathbb{F}^n \to \mathbb{F}^m$ be defined as above. If $n < m$, then T is not surjective. Hence

$$\mathrm{Im}(T) \neq \mathbb{F}^m.$$

Therefore there exists $c \in \mathbb{F}^m$ such that $T(x) = c$ has no solution. $\square$

Linear Algebra — Lecture 14: Linear maps and matrices

1 Representing a Linear Map by a Matrix

We know that if $v_1, \dots, v_n$ is a basis of V and $T : V \to W$ is linear, then the values $Tv_1, \dots, Tv_n$ determine the values of T on arbitrary vectors in V. Matrices provide an efficient way of recording the values of the Tv_k's in terms of a basis of W.

1.1 Matrices

Definition 1.1. *Let $m, n \geq 0$ be integers. An $m \times n$ matrix A over a field $\mathbb{F}$ is a rectangular array of elements of $\mathbb{F}$ with m rows and n columns:*

$$A = \begin{pmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & & \vdots \\ A_{m,1} & \cdots & A_{m,n} \end{pmatrix}.$$

The notation $A_{j,k}$ denotes the entry in row j, column k of A.

1.2 Matrix of a Linear Map

Definition 1.2. *Let $T \in \mathcal{L}(V, W)$. Let $(v_1, \dots, v_n)$ be a basis of V and $(w_1, \dots, w_m)$ a basis of W.*

The matrix of T *with respect to these bases is the $m \times n$ matrix $\mathcal{M}(T)$ whose entries $A_{j,k}$ are defined by*

$$Tv_k = A_{1,k}w_1 + \cdots + A_{m,k}w_m, \qquad k = 1, \dots, n.$$

Remark. The k-th column of $\mathcal{M}(T)$ consists of the coordinates of $T(v_k)$ in the basis $(w_1, \dots, w_m)$.

$$\mathcal{M}(T) = \begin{array}{c} \\ w_1 \\ \vdots \\ w_m \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & \cdots & v_k & \cdots & v_n \end{array} \\ \left[\begin{array}{ccccc} & & A_{1,k} & & \\ & & \vdots & & \\ & & A_{m,k} & & \end{array}\right] \end{array}$$

Remark. If T is a linear map from an n-dimensional vector space to an m-dimensional vector space, then $\mathcal{M}(T)$ is an $m \times n$ matrix.

1.3 Example

Let

$$\mathcal{P}_3(\mathbb{R}) = \{a_0 + a_1x + a_2x^2 + a_3x^3 : a_i \in \mathbb{R}\}$$

be the vector space of real polynomials of degree at most 3, and

$$\mathcal{P}_2(\mathbb{R}) = \{b_0 + b_1x + b_2x^2 : b_i \in \mathbb{R}\}$$

the space of polynomials of degree at most 2.

Their standard bases are

$$(1, x, x^2, x^3) \quad \text{and} \quad (1, x, x^2),$$

respectively.

Let $D : \mathcal{P}_3(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ be the differentiation map defined by

$$Dp = p'.$$

We compute:

$$D(1) = 0, \quad D(x) = 1, \quad D(x^2) = 2x, \quad D(x^3) = 3x^2.$$

Thus the matrix of D is

$$\mathcal{M}(D) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

2 Addition and Scalar Multiplication of Matrices

2.1 Matrix Addition

Definition 2.1. *The sum of two matrices of the same size is the matrix obtained by adding corresponding entries.*

If

$$A = (A_{j,k}) \quad \textit{and} \quad C = (C_{j,k})$$

are $m \times n$ matrices, then

$$(A + C)_{j,k} = A_{j,k} + C_{j,k}.$$

Proposition 2.1. *Let $S, T \in \mathcal{L}(V, W)$. Then*

$$\mathcal{M}(S + T) = \mathcal{M}(S) + \mathcal{M}(T).$$

Proof. Let $(v_1, \ldots, v_n)$ be a basis of V and $(w_1, \ldots, w_m)$ a basis of W.

Suppose

$$S(v_k) = \sum_{j=1}^{m} A_{j,k}w_j, \qquad T(v_k) = \sum_{j=1}^{m} B_{j,k}w_j.$$

Then

$$(S + T)(v_k) = S(v_k) + T(v_k) = \sum_{j=1}^{m} (A_{j,k} + B_{j,k})w_j.$$

Thus the k-th column of $\mathcal{M}(S + T)$ is obtained by adding the k-th columns of $\mathcal{M}(S)$ and $\mathcal{M}(T)$, which proves the result. $\square$

2.2 Scalar Multiplication

Definition 2.2. *The product of a scalar λ and a matrix $A = (A_{j,k})$ is the matrix obtained by multiplying each entry by λ:*

$$(\lambda A)_{j,k} = \lambda A_{j,k}.$$

$$2\begin{pmatrix} 3 & 1 \\ -1 & 5 \end{pmatrix} + \begin{pmatrix} 4 & 2 \\ 1 & 6 \end{pmatrix} = \begin{pmatrix} 6 & 2 \\ -2 & 10 \end{pmatrix} + \begin{pmatrix} 4 & 2 \\ 1 & 6 \end{pmatrix} = \begin{pmatrix} 10 & 4 \\ -1 & 16 \end{pmatrix}.$$

Proposition 2.2. *Let $\lambda \in \mathbb{F}$ and $T \in \mathcal{L}(V, W)$. Then*

$$\mathcal{M}(\lambda T) = \lambda\,\mathcal{M}(T).$$

Proof. Suppose

$$T(v_k) = \sum_{j=1}^{m} A_{j,k} w_j.$$

Then

$$(\lambda T)(v_k) = \lambda T(v_k) = \sum_{j=1}^{m} (\lambda A_{j,k}) w_j.$$

Thus the k-th column of $\mathcal{M}(\lambda T)$ is λ times the k-th column of $\mathcal{M}(T)$, which proves the result. □

2.3 The Vector Space of Matrices

Definition 2.3. *Let m, n be positive integers. We denote by*

$$M_{m,n}(\mathbb{F})$$

the set of all $m \times n$ matrices with entries in $\mathbb{F}$.

Proposition 2.3. *$M_{m,n}(\mathbb{F})$ is a vector space of dimension mn.*

Proof. Matrix addition and scalar multiplication are defined entrywise, so the vector space axioms follow directly from the corresponding properties in $\mathbb{F}$.

The zero vector is the $m \times n$ matrix whose entries are all 0.

For $1 \leq j \leq m$ and $1 \leq k \leq n$, define $E_{j,k} \in M_{m,n}(\mathbb{F})$ by

$$(E_{j,k})_{r,s} = \begin{cases} 1 & \text{if } (r,s) = (j,k), \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to check that the family

$$\{E_{j,k} : 1 \leq j \leq m,\ 1 \leq k \leq n\}$$

forms a basis of $M_{m,n}(\mathbb{F})$. Hence the dimension is mn. □

3 Matrix Multiplication

Suppose $T : U \to V$ and $S : V \to W$ are linear maps. The composition $S \circ T : U \to W$ is again a linear map.

A natural question is:

What is the relation between the matrices of S, T, and ST?

We will define matrix multiplication so that

$$\mathcal{M}(ST) = \mathcal{M}(S)\mathcal{M}(T).$$

3.1 Definition of Matrix Multiplication

Definition 3.1. *Let A be an $m \times n$ matrix and B an $n \times p$ matrix. The product AB is the $m \times p$ matrix defined by*

$$(AB)_{j,k} = \sum_{r=1}^{n} A_{j,r}B_{r,k}.$$

Remark. The entry in row j, column k of AB is obtained by multiplying row j of A with column k of B, and then summing the results.

Remark. Matrix multiplication is not commutative in general:

$$AB \neq BA$$

even when both products are defined.

3.2 Matrices and Composition

Proposition 3.1. *Let $T : U \to V$ and $S : V \to W$ be linear maps. Then*

$$\mathcal{M}(S \circ T) = \mathcal{M}(S)\mathcal{M}(T).$$

Proof. Let $(u_1, \ldots, u_p)$ be a basis of U, $(v_1, \ldots, v_n)$ a basis of V, and $(w_1, \ldots, w_m)$ a basis of W.

Suppose

$$T(u_k) = \sum_{r=1}^{n} B_{r,k}v_r, \qquad S(v_r) = \sum_{j=1}^{m} A_{j,r}w_j.$$

Then

$$(S \circ T)(u_k) = S(T(u_k)) = S\left(\sum_{r=1}^{n} B_{r,k}v_r\right) = \sum_{r=1}^{n} B_{r,k}S(v_r) = \sum_{r=1}^{n} B_{r,k}\sum_{j=1}^{m} A_{j,r}w_j.$$

Reordering the sums gives

$$(S \circ T)(u_k) = \sum_{j=1}^{m}\left(\sum_{r=1}^{n} A_{j,r}B_{r,k}\right) w_j.$$

Thus the (j, k)-entry of $\mathcal{M}(S \circ T)$ equals

$$\sum_{r=1}^{n} A_{j,r}B_{r,k},$$

which is exactly the (j, k)-entry of $\mathcal{M}(S)\mathcal{M}(T)$. $\square$

3.3 In-class exercise

Exercise 3.1 (In-class exercise)**.** *Let $M_n(\mathbb{R})$ be the vector space of all $n \times n$ real matrices.*

Determine the center *of $M_n(\mathbb{R})$, that is, the set of matrices*

$$A \in M_n(\mathbb{R})$$

such that, for every

$$M \in M_n(\mathbb{R}),$$

we have

$$AM = MA.$$

Lecture: Rank of a Matrix (I)

4. The rank of a matrix

4.1 Column rank

Let $m, n \in \mathbb{N}$ and $A \in M_{m,n}(\mathbb{K})$ be a matrix with m rows and n columns over the field $\mathbb{K}$.

Definition 1 (Column rank)**.** *The* ***column rank*** *of A is the dimension of the subspace of $\mathbb{K}^m$ generated by the n columns of A. We denote it by $\mathrm{rg}(A)$.*

4.2 Characterization via linear maps

Proposition 1. *Let $m, n \in \mathbb{N}$. The column rank of a matrix $A \in M_{m,n}(\mathbb{K})$ is equal to the dimension of the image*

$$\mathrm{Im}(f)$$

of the linear map

$$f : \mathbb{K}^n \to \mathbb{K}^m, \qquad x \mapsto Ax.$$

Proof. Let $(e_1, \ldots, e_n)$ be the standard ordered basis of $\mathbb{K}^n$. The image of f is generated by the set

$$\{f(e_1), \ldots, f(e_n)\}.$$

For each $i \in [n]$, the vector

$$f(e_i) = Ae_i$$

is precisely the i-th column of the matrix A. This proves the proposition. $\square$

4.3 Basic properties

Proposition 2. *Let $m, n \in \mathbb{N}$, let $A \in M_{m,n}(\mathbb{K})$, and let*

$$f : x \in \mathbb{K}^n \mapsto Ax \in \mathbb{K}^m.$$

1. $0 \leq \mathrm{rg}(A) \leq \min\{m, n\}$, *and the column rank of A is zero if and only if A is the zero matrix.*
2. *The map f is injective if and only if $\mathrm{rg}(A) = n$.*
3. *The map f is surjective if and only if $\mathrm{rg}(A) = m$.*

Proof. Since $\mathrm{Im}(f)$ is a subspace of $\mathbb{K}^m$, whose dimension is m, we have

$$\dim(\mathrm{Im}(f)) \leq m.$$

On the other hand, by the rank–nullity theorem,

$$\dim(\mathrm{Im}(f)) \leq \dim \mathbb{K}^n = n.$$

These two inequalities prove (1).

Again from rank–nullity,

$$n = \dim \ker(f) + \mathrm{rg}(A).$$

Thus $\ker(f) = \{0\}$ if and only if $\mathrm{rg}(A) = n$, which proves (2).

Finally, since $\mathrm{Im}(f) \subseteq \mathbb{K}^m$, the map f is surjective if and only if

$$\dim(\mathrm{Im}(f)) = m,$$

which proves (3). $\square$

4.4 Rank of a product

Proposition 3. *Let m, n, p be three nonnegative integers. If $A \in M_{m,n}(\mathbb{K})$ and $B \in M_{n,p}(\mathbb{K})$, then*

$$\mathrm{rg}_c(AB) \leq \min\{\mathrm{rg}_c(A), \mathrm{rg}_c(B)\}.$$

Proof. Let $A \in M_{m,n}(\mathbb{K})$ and $B \in M_{n,p}(\mathbb{K})$, and let

$$f_A : \mathbb{K}^n \to \mathbb{K}^m, \qquad x \mapsto Ax,$$

$$f_B : \mathbb{K}^p \to \mathbb{K}^n, \qquad x \mapsto Bx$$

be the associated linear maps.

Since the composition $f_A \circ f_B$ is the map

$$x \in \mathbb{K}^p \longmapsto ABx \in \mathbb{K}^m,$$

we have

$$\mathrm{rg}_c(AB) = \dim \mathrm{Im}(f_{AB}) = \dim \mathrm{Im}(f_A \circ f_B).$$

Because

$$\mathrm{Im}(f_A \circ f_B) \subseteq \mathrm{Im}(f_A),$$

it follows that

$$\mathrm{rg}_c(AB) = \dim \mathrm{Im}(f_A \circ f_B) \leq \dim \mathrm{Im}(f_A) = \mathrm{rg}_c(A).$$

On the other hand,

$$\mathrm{Im}(f_A \circ f_B) = f_A\big(\mathrm{Im}(f_B)\big),$$

hence

$$\mathrm{rg}_c(AB) = \dim \mathrm{Im}(f_A \circ f_B) \leq \dim f_A\big(\mathrm{Im}(f_B)\big) \leq \dim \mathrm{Im}(f_B) = \mathrm{rg}_c(B).$$

These two inequalities imply that

$$\mathrm{rg}_c(AB) \leq \min\{\mathrm{rg}_c(A), \mathrm{rg}_c(B)\}.$$

□

4.5 Corollary

Let $l, m, n, p \in \mathbb{N}$ and let $A \in M_{m,n}(\mathbb{K})$.

1. If $B \in M_{n,p}(\mathbb{K})$ with $\mathrm{rg}(B) = n$, then

$$\mathrm{rg}(AB) = \mathrm{rg}(A).$$

2. If $C \in M_{l,m}(\mathbb{K})$ with $\mathrm{rg}(C) = m$, then

$$\mathrm{rg}(CA) = \mathrm{rg}(A).$$

Proof. Since $\mathrm{rg}(B) \leq n$, we have

$$\mathrm{rg}(AB) \leq \min\{\mathrm{rg}(A), \mathrm{rg}(B)\} \leq \min\{\mathrm{rg}(A), n\} = \mathrm{rg}(A).$$

The reverse inequality is immediate, hence equality. The second part is proved similarly. □

4.6 Sylvester inequality

Proposition 4 (Sylvester inequality)**.** *Let $m, n, p \in \mathbb{N}$. If $A \in M_{m,n}(\mathbb{K})$ and $B \in M_{n,p}(\mathbb{K})$, then*

$$\mathrm{rg}(A) + \mathrm{rg}(B) \le \mathrm{rg}(AB) + n.$$

Proof. Consider the linear maps

$$f_A : x \in \mathbb{K}^n \mapsto Ax \in \mathbb{K}^m, \qquad f_B : x \in \mathbb{K}^p \mapsto Bx \in \mathbb{K}^n.$$

From rank–nullity,

$$n = \dim\ker(f_A) + \mathrm{rg}(A),$$
$$p = \dim\ker(f_B) + \mathrm{rg}(B),$$
$$p = \dim\ker(f_A \circ f_B) + \mathrm{rg}(AB).$$

Hence

$$\mathrm{rg}(AB) + n = \mathrm{rg}(A) + \mathrm{rg}(B) + [\dim\ker(f_B) + \dim\ker(f_A) - \dim\ker(f_A \circ f_B)].$$

It suffices to show that the bracketed term is non-negative.
If $x \in \ker(f_A \circ f_B)$, then $f_B(x) \in \ker(f_A)$. Define

$$h : x \in \ker(f_A \circ f_B) \mapsto f_B(x) \in \ker(f_A).$$

Then

$$\mathrm{Im}(h) \subseteq \ker(f_A), \quad \ker(h) \subseteq \ker(f_B).$$

Applying rank–nullity to h gives

$$\dim\ker(f_A \circ f_B) = \dim\ker(h) + \dim\mathrm{Im}(h) \le \dim\ker(f_B) + \dim\ker(f_A).$$

This proves the inequality. □

4.7 Subadditivity of rank

Proposition 5. *Let m, n be nonnegative integers. If $A, B \in M_{m,n}(\mathbb{K})$, then*

$$\mathrm{rg}(A + B) \le \mathrm{rg}(A) + \mathrm{rg}(B).$$

Proof. Let A and B be two matrices in $M_{m,n}(\mathbb{K})$ and consider the corresponding linear maps

$$f_A : \mathbb{K}^n \to \mathbb{K}^m, \quad x \mapsto Ax,$$
$$f_B : \mathbb{K}^n \to \mathbb{K}^m, \quad x \mapsto Bx.$$

If $x \in \mathbb{K}^n$, then

$$(f_A + f_B)(x) = f_A(x) + f_B(x) \in \mathrm{Im}(f_A) + \mathrm{Im}(f_B).$$

Thus,

$$\mathrm{Im}(f_A + f_B) \subseteq \mathrm{Im}(f_A) + \mathrm{Im}(f_B).$$

Using the formula for the dimension of a sum of subspaces, we obtain

$$\begin{aligned}\mathrm{rg}(A + B) = \dim\mathrm{Im}(f_A + f_B) &\le \dim\big(\mathrm{Im}(f_A) + \mathrm{Im}(f_B)\big)\\ &= \dim\mathrm{Im}(f_A) + \dim\mathrm{Im}(f_B) - \dim\big(\mathrm{Im}(f_A) \cap \mathrm{Im}(f_B)\big)\\ &\le \dim\mathrm{Im}(f_A) + \dim\mathrm{Im}(f_B) = \mathrm{rg}(A) + \mathrm{rg}(B).\end{aligned}$$

This proves the proposition. □

Exercise.

Let A and B be two 3×3 matrices over $\mathbb{K}$ such that

$$AB = 0_3.$$

Show that at least one of the matrices has rank less than or equal to 1.

Solution 1 (using kernels and images).

Let u and v be the linear maps on $\mathbb{K}^3$ canonically associated with A and B.

Since $AB = 0$, we have

$$u \circ v = 0.$$

Therefore,

$$\mathrm{Im}(v) \subseteq \ker(u).$$

Hence,

$$\mathrm{rank}(v) = 3 - \dim\ker(v) \leq \dim\ker(u).$$

Thus,

$$\dim\ker(u) + \dim\ker(v) \geq 3.$$

This implies that either

$$\dim\ker(u) \geq 2 \quad \text{or} \quad \dim\ker(v) \geq 2.$$

Therefore,

$$\mathrm{rank}(A) = \mathrm{rank}(u) \leq 1 \quad \text{or} \quad \mathrm{rank}(B) = \mathrm{rank}(v) \leq 1.$$

Solution 2 (using Sylvester's inequality).

Since $AB = 0_3$, we have

$$\mathrm{rank}(AB) = 0.$$

By Sylvester's inequality,

$$\mathrm{rank}(A) + \mathrm{rank}(B) \leq \mathrm{rank}(AB) + 3.$$

Thus,

$$\mathrm{rank}(A) + \mathrm{rank}(B) \leq 3.$$

If both matrices had rank at least 2, we would have

$$\mathrm{rank}(A) + \mathrm{rank}(B) \geq 4,$$

which is impossible.

Therefore,

$$\mathrm{rank}(A) \leq 1 \quad \text{or} \quad \mathrm{rank}(B) \leq 1.$$

Lecture 16: Endomorphisms – Automorphisms and Projectors

1. Endomorphisms and Automorphisms

Definition 1. *Let E be a K-vector space.*

1. *A linear map $u : E \to E$ is called an* ***endomorphism*** *of E.*

2. *A bijective endomorphism is called an* ***automorphism*** *of E.*

Remark 1. *Assume that E is finite dimensional and u is an endomorphism of E. By the Rank–Nullity Theorem,*

$$\dim(E) = \dim(\ker u) + \dim(\operatorname{Im} u).$$

However, in general we do not necessarily have

$$E = \ker(u) \oplus \operatorname{Im}(u).$$

We now characterize when this decomposition holds.

2. Characterization of the Direct Sum Decomposition

Theorem 1. *Let E be a finite-dimensional K-vector space and $u : E \to E$ an endomorphism. The following statements are equivalent:*

1. $E = \ker(u) \oplus \operatorname{Im}(u)$.

2. $\operatorname{Im}(u) = \operatorname{Im}(u^2)$.

3. $\ker(u) = \ker(u^2)$.

4. $\ker(u) \cap \operatorname{Im}(u) = \{0\}$.

Proof. (1) $\Rightarrow$ (2). We always have $\operatorname{Im}(u^2) \subset \operatorname{Im}(u)$. Let $y \in \operatorname{Im}(u)$, so $y = u(x)$. Write $x = x_1 + x_2$ with $x_1 \in \ker(u)$ and $x_2 \in \operatorname{Im}(u)$. Then $y = u(x_2)$. Since $x_2 = u(z)$ for some z,

$$y = u(u(z)) = u^2(z).$$

Thus $\operatorname{Im}(u) \subset \operatorname{Im}(u^2)$.

(2) $\Rightarrow$ (3). We always have $\ker(u) \subset \ker(u^2)$. By rank-nullity and equality of image dimensions,

$$\dim(\ker u) = \dim(\ker u^2),$$

hence equality of kernels.

(3) $\Rightarrow$ (4). If $y \in \ker(u) \cap \operatorname{Im}(u)$, then $y = u(x)$ and $u(y) = 0$. Thus $u^2(x) = 0$, so $x \in \ker(u^2) = \ker(u)$. Hence $y = u(x) = 0$.

(4) $\Rightarrow$ (1). Since dimensions add correctly and the intersection is trivial,

$$E = \ker(u) \oplus \operatorname{Im}(u).$$

□

3. Projectors

Definition 2. *Let E be a K-vector space. A linear transformation $u : E \to E$ is called a* ***projector*** *if*

$$u^2 = u.$$

Proposition 1. *Let $u : E \to E$ be linear. Then u is a projector if and only if*

$$u(x) = x \quad \text{for every } x \in \operatorname{Im}(u).$$

Proof. If $u^2 = u$ and $x = u(v)$, then

$$u(x) = u(u(v)) = u(v) = x.$$

Conversely, if $u(x) = x$ for all $x \in \operatorname{Im}(u)$, then for any v,

$$u(u(v)) = u(v),$$

so $u^2 = u$. $\square$

Proposition 2. *If u is a projector, then*

$$E = \ker(u) \oplus \operatorname{Im}(u).$$

Proof. If $x \in \ker(u) \cap \operatorname{Im}(u)$, then $u(x) = x$ and $u(x) = 0$, hence $x = 0$.

For any $x \in E$,

$$x = (x - u(x)) + u(x),$$

with $x - u(x) \in \ker(u)$ and $u(x) \in \operatorname{Im}(u)$. $\square$

4. Projectors and Direct Sums

Proposition 3. *Let E be a K-vector space and suppose*

$$E = S \oplus T.$$

Then there exists a unique projector $u : E \to E$ such that

$$\ker(u) = S, \qquad \operatorname{Im}(u) = T.$$

Proof. Every $x \in E$ has a unique decomposition $x = s + t$ with $s \in S$, $t \in T$. Define

$$u(x) = t.$$

This is well-defined because the decomposition is unique. Linearity follows directly from the direct sum structure. Moreover,

$$u^2(x) = u(t) = t = u(x),$$

so u is a projector. It is immediate that $\operatorname{Im}(u) = T$ and $\ker(u) = S$. Uniqueness follows from uniqueness of the decomposition. $\square$

5. Matrix Form of a Projector

Suppose

$$E = S \oplus T$$

and choose a basis adapted to this decomposition: first a basis of S, then a basis of T.

In this basis, the matrix of the projector u with $\ker(u) = S$ and $\mathrm{Im}(u) = T$ is

$$\begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix}.$$

Indeed, u sends every vector in S to 0 and acts as the identity on T.

From this matrix form we immediately see that

$$u^2 = u,$$

and therefore every projector is diagonalizable, with eigenvalues only 0 and 1.

Lecture 17: Change of bases and Similarity

Invertible Matrices

Definition 1. *A square matrix $A \in M_n(K)$ is called **invertible** if there exists a matrix $B \in M_n(K)$ such that*

$$AB = BA = I_n.$$

*The matrix B is called the **inverse** of A and is denoted by A^{-1}.*

Proposition 1. *If A is invertible, then its inverse is unique.*

Proof. Suppose B and C both satisfy

$$AB = BA = I \quad \text{and} \quad AC = CA = I.$$

Then

$$B = BI = B(AC) = (BA)C = IC = C.$$

□

Basic Properties of the Inverse

Proposition 2. *Let A and C be invertible matrices of the same size.*

1. $(A^{-1})^{-1} = A$.
2. *The product AC is invertible and*

$$(AC)^{-1} = C^{-1}A^{-1}.$$

Proof: Exercise.

Change of bases

Theorem 1. *Let $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$.*

Let $(u_1, \ldots, u_m)$ be a basis of U, $(v_1, \ldots, v_n)$ a basis of V, and $(w_1, \ldots, w_p)$ a basis of W.

Then

$$\mathcal{M}(ST, (u_1, \ldots, u_m), (w_1, \ldots, w_p)) = \mathcal{M}(S, (v_1, \ldots, v_n), (w_1, \ldots, w_p))\mathcal{M}(T, (u_1, \ldots, u_m), (v_1, \ldots, v_n)).$$

Matrix of the Identity with Respect to Two Bases

Proposition 3. *Let $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ be two bases of V.*

Then the matrices

$$\mathcal{M}(I, (u_1, \ldots, u_n), (v_1, \ldots, v_n)) \quad \text{and} \quad \mathcal{M}(I, (v_1, \ldots, v_n), (u_1, \ldots, u_n))$$

are invertible and are inverses of each other.

Proof. We apply the product formula for matrices of compositions to the identity operator.

Recall that for linear maps S, T,

$$\mathcal{M}(ST, (u), (w)) = \mathcal{M}(S, (v), (w))\mathcal{M}(T, (u), (v)).$$

Now take both maps equal to the identity:

$$S = I \quad \text{and} \quad T = I.$$

Also take $w_k = u_k$ (so that the final basis is (u)).

Since $I \circ I = I$, the product formula gives

$$\mathcal{M}(I, (v), (u))\mathcal{M}(I, (u), (v)) = \mathcal{M}(I, (u), (u)).$$

But the matrix of the identity operator with respect to the same basis is the identity matrix. Hence

$$\mathcal{M}(I, (v), (u))\mathcal{M}(I, (u), (v)) = I_n.$$

Now interchange the roles of the two bases (u) and (v). Repeating the same argument gives

$$\mathcal{M}(I, (u), (v))\mathcal{M}(I, (v), (u)) = I_n.$$

Therefore each matrix is the inverse of the other, and in particular both are invertible.

□

Change-of-Basis Formula

Theorem 2. *Let $T \in \mathcal{L}(V)$ and let $(u_1, \dots, u_n)$ and $(v_1, \dots, v_n)$ be two bases of V.*

Define

$$A = \mathcal{M}(T, (u_1, \dots, u_n)), \qquad B = \mathcal{M}(T, (v_1, \dots, v_n)),$$

and

$$C = \mathcal{M}(I, (u_1, \dots, u_n), (v_1, \dots, v_n)).$$

Then

$$A = C^{-1}BC.$$

Proof. Using the composition formula for matrices, first write

$$T = I \circ T.$$

Then

$$\mathcal{M}(T, (u), (u)) = \mathcal{M}(I, (v), (u))\mathcal{M}(T, (u), (v)).$$

Since $\mathcal{M}(T, (u), (u)) = A$ and $\mathcal{M}(I, (v), (u)) = C^{-1}$, we obtain

$$A = C^{-1}\mathcal{M}(T, (u), (v)).$$

Now write

$$T = T \circ I.$$

Again by the composition formula,

$$\mathcal{M}(T,(u),(v)) = \mathcal{M}(T,(v),(v))\mathcal{M}(I,(u),(v)).$$

That is,

$$\mathcal{M}(T,(u),(v)) = BC.$$

Substituting gives

$$A = C^{-1}BC.$$

□

Similarity is an Equivalence Relation

Definition 2. *Let $A, B \in M_n(K)$.*

*We say that A and B are **similar** if there exists an invertible matrix $P \in M_n(K)$ such that*

$$B = P^{-1}AP.$$

Proposition 4. *The relation $\sim$ is an equivalence relation on $M_n(K)$.*

Proof. **Reflexivity.**

For any A, take $P = I_n$. Then

$$A = I^{-1}AI,$$

so $A \sim A$.

Symmetry.

If $A \sim B$, then

$$B = P^{-1}AP$$

for some invertible P.

Multiplying on the left by P and on the right by P^{-1} gives

$$A = PBP^{-1}.$$

Since P^{-1} is invertible, this shows

$$A = (P^{-1})^{-1}B(P^{-1}),$$

so $B \sim A$.

Transitivity.

Suppose $A \sim B$ and $B \sim C$.

Then there exist invertible matrices P and Q such that

$$B = P^{-1}AP, \qquad C = Q^{-1}BQ.$$

Substituting,

$$C = Q^{-1}(P^{-1}AP)Q = (PQ)^{-1}A(PQ).$$

Since PQ is invertible, we conclude $A \sim C$.

□

Remark 1. *The equivalence class of a matrix A under this relation is called the **similarity class** of A. Matrices in the same similarity class represent the same linear operator in different bases.*

Exercise

Let V be a finite-dimensional vector space and let $v_1, \dots, v_n$ be a basis of V.

Suppose $T \in \mathcal{L}(V)$ is invertible. Show that

$$\mathcal{M}(T^{-1}) = \big(\mathcal{M}(T)\big)^{-1},$$

where both matrices are taken with respect to the basis $v_1, \dots, v_n$.

In-Class Exercise

Let $M \in M_4(\mathbb{R})$ satisfy

$$M^2 + I_4 = 0.$$

Show that M is similar to the matrix

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

Let $f \in \mathcal{L}(\mathbb{R}^4)$ be the linear operator canonically associated with the matrix M. Since $M^2 + I_4 = 0$, we have

$$f^2 = -\mathrm{Id}.$$

Analysis.
We look for a basis (e_1, e_2, e_3, e_4) such that

$$f(e_1) = e_2, \quad f(e_2) = -e_1, \quad f(e_3) = e_4, \quad f(e_4) = -e_3.$$

Knowing e_1 and e_3 will be enough to define

$$e_2 = f(e_1), \qquad e_4 = f(e_3),$$

and the required relations will follow from $f^2 = -\mathrm{Id}$.

Construction.
Choose $e_1 \neq 0$ and define

$$e_2 = f(e_1).$$

Since $f^2 = -\mathrm{Id}$,

$$f(e_2) = f(f(e_1)) = -e_1.$$

Because $\dim \mathbb{R}^4 = 4$ and $\dim \operatorname{span}(e_1, e_2) = 2$, there exists

$$e_3 \notin \operatorname{span}(e_1, e_2).$$

Define

$$e_4 = f(e_3).$$

Again, using $f^2 = -\mathrm{Id}$,

$$f(e_4) = f(f(e_3)) = -e_3.$$

Linear independence.
Suppose
$$\lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3 + \lambda_4 e_4 = 0.$$
Since $e_2 = f(e_1)$ and $e_4 = f(e_3)$, this becomes
$$\lambda_1 e_1 + \lambda_2 f(e_1) + \lambda_3 e_3 + \lambda_4 f(e_3) = 0. \tag{1}$$
Applying f to (1) gives
$$\lambda_1 f(e_1) - \lambda_2 e_1 + \lambda_3 f(e_3) - \lambda_4 e_3 = 0. \tag{2}$$
Now compute
$$\lambda_3(1) - \lambda_4(2).$$
This yields
$$(\lambda_3\lambda_1 + \lambda_2\lambda_4)e_1 + (\lambda_3\lambda_2 - \lambda_4\lambda_1)f(e_1) + (\lambda_3^2 + \lambda_4^2)e_3 = 0.$$
Since $e_3 \notin \operatorname{span}(e_1, e_2)$, the coefficient of e_3 must vanish:
$$\lambda_3^2 + \lambda_4^2 = 0.$$
Over $\mathbb{R}$, this implies
$$\lambda_3 = \lambda_4 = 0.$$
Returning to (1) and (2), we obtain
$$\lambda_1 e_1 + \lambda_2 f(e_1) = 0,$$
$$\lambda_1 f(e_1) - \lambda_2 e_1 = 0.$$
As before, combining these gives
$$\lambda_1^2 + \lambda_2^2 = 0,$$
hence
$$\lambda_1 = \lambda_2 = 0.$$
Therefore
$$(e_1, e_2, e_3, e_4)$$
is a linearly independent family, hence a basis of $\mathbb{R}^4$.

In this basis, the matrix of f is
$$\begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$
Therefore M is similar to the desired matrix.

Lecture 18: Trace of an endomorphism

The Trace of a Matrix

Using the correspondence between matrices and linear maps studied previously, we can construct an important invariant of endomorphisms of a finite-dimensional vector space: its **trace**.

We begin by recalling the definition of the trace of a square matrix and its most basic properties.

Definition 1. *Let n be a nonnegative integer. The* ***trace*** *of a matrix $A = (a_{i,j})$ of size $n \times n$ with entries in K is the scalar*

$$\operatorname{tr}(A) := \sum_{i=1}^{n} a_{i,i}.$$

The **trace function** is the map

$$\operatorname{tr} : M_n(K) \longrightarrow K, \qquad A \longmapsto \operatorname{tr}(A).$$

Proposition 1. *(i) For every nonnegative integer n, the function*

$$\operatorname{tr} : M_n(K) \longrightarrow K$$

is linear. If $n > 0$, it is also surjective.

(ii) Let n and m be nonnegative integers. If $A \in M_{m,n}(K)$ and $B \in M_{n,m}(K)$, then

$$\operatorname{tr}(AB) = \operatorname{tr}(BA).$$

Remark 1. *In general, the matrices AB and BA whose traces we compare in part (ii) have different sizes.*

Proof. **(i)** Let n be a nonnegative integer. Let $A = (a_{i,j})$ and $B = (b_{i,j})$ be matrices in $M_n(K)$, and let $\alpha, \beta \in K$.

Then

$$\alpha A + \beta B = (\alpha a_{i,j} + \beta b_{i,j}),$$

so

$$\operatorname{tr}(\alpha A + \beta B) = \sum_{i=1}^{n} (\alpha a_{i,i} + \beta b_{i,i}) = \alpha \sum_{i=1}^{n} a_{i,i} + \beta \sum_{i=1}^{n} b_{i,i}.$$

Thus

$$\operatorname{tr}(\alpha A + \beta B) = \alpha \operatorname{tr}(A) + \beta \operatorname{tr}(B),$$

which shows that tr is linear.

Now suppose $n > 0$. To prove surjectivity, let $c \in K$ be arbitrary and consider the matrix $A = cE_{11}$, where E_{11} has a 1 in the $(1,1)$-position and zeros elsewhere. Then

$$\operatorname{tr}(A) = \operatorname{tr}(cE_{11}) = c \operatorname{tr}(E_{11}) = c.$$

Since every $c \in K$ is the trace of some matrix, the trace map is surjective.

(ii) Let $A = (a_{i,k}) \in M_{m,n}(K)$ and $B = (b_{k,j}) \in M_{n,m}(K)$.

If we write $AB = (c_{i,j})$, then

$$c_{i,j} = \sum_{k=1}^{n} a_{i,k} b_{k,j}.$$

Hence

$$\operatorname{tr}(AB) = \sum_{i=1}^{m} c_{i,i} = \sum_{i=1}^{m} \sum_{k=1}^{n} a_{i,k} b_{k,i}.$$

Similarly, if $BA = (d_{i,j})$, then

$$d_{i,j} = \sum_{k=1}^{m} b_{i,k} a_{k,j},$$

and therefore

$$\operatorname{tr}(BA) = \sum_{i=1}^{n} d_{i,i} = \sum_{i=1}^{n} \sum_{k=1}^{m} b_{i,k} a_{k,i}.$$

These two double sums coincide (by renaming indices), so

$$\operatorname{tr}(AB) = \operatorname{tr}(BA).$$

□

Proposition 2. *Let n be a nonnegative integer. If $A, C \in M_n(K)$ and C is invertible, then*

$$\operatorname{tr}(C^{-1}AC) = \operatorname{tr}(A).$$

Proof. Let $A, C \in M_n(K)$ and suppose C is invertible. By the second part of previous Proposition we have

$$\operatorname{tr}(C^{-1}AC) = \operatorname{tr}(CC^{-1}A) = \operatorname{tr}(IA) = \operatorname{tr}(A).$$

□

According to the second part of the previous proposition, whenever n and m are nonnegative integers and $A \in M_{m,n}(K)$, $B \in M_{n,m}(K)$, we have

$$\operatorname{tr}(AB) = \operatorname{tr}(BA).$$

For this reason, we say that the trace is cyclically invariant.

This immediately implies the following more general result.

Corollary 1. *Let $n_0, n_1, \ldots, n_k$ be nonnegative integers such that $n_0 = n_k$. Suppose*

$$A_1 \in M_{n_0,n_1}(K), \quad A_2 \in M_{n_1,n_2}(K), \quad \ldots, \quad A_k \in M_{n_{k-1},n_k}(K),$$

so that the products

$$A_1 A_2 \cdots A_k \quad \textit{and} \quad A_2 \cdots A_k A_1$$

are well-defined and are square matrices. Then

$$\operatorname{tr}(A_1 A_2 \cdots A_k) = \operatorname{tr}(A_2 \cdots A_k A_1).$$

It is important to note, however, that despite this property, the trace of a product of matrices does depend on the order of the factors in general: it is easy to find matrices $A, B, C \in M_2(K)$ such that

$$\operatorname{tr}(ABC) \neq \operatorname{tr}(ACB).$$

Theorem 1. *Let $T \in \mathcal{L}(V)$ be an endomorphism of a finite-dimensional vector space V. If $(u_1, \dots, u_n)$ and $(v_1, \dots, v_n)$ are two bases of V, then*

$$\operatorname{tr}\big(\mathcal{M}(T, (u_1, \dots, u_n))\big) = \operatorname{tr}\big(\mathcal{M}(T, (v_1, \dots, v_n))\big).$$

Proof. Let

$$A = \mathcal{M}(T, (u_1, \dots, u_n)) \quad \text{and} \quad B = \mathcal{M}(T, (v_1, \dots, v_n)).$$

By the change-of-basis formula, there exists an invertible matrix C such that

$$A = C^{-1}BC.$$

Since the trace is invariant under conjugation, we obtain

$$\operatorname{tr}(A) = \operatorname{tr}(B).$$

$\square$

Definition 2. *Let $T \in \mathcal{L}(V)$ be an endomorphism of a finite-dimensional vector space V.*

*The **trace of** T, denoted $\operatorname{tr}(T)$, is defined by*

$$\operatorname{tr}(T) = \operatorname{tr}\big(\mathcal{M}(T, (v_1, \dots, v_n))\big),$$

where $(v_1, \dots, v_n)$ is any basis of V.

Remark 2. *The previous theorem shows that this definition makes sense, since the value does not depend on the choice of basis.*

Homework Exercise

Let V be a finite-dimensional vector space and let $P \in \mathcal{L}(V)$ be a projector.

Prove that

$$\operatorname{tr}(P) = \dim(\operatorname{range} P).$$

Remark 3. *If two matrices do not have the same trace or do not have the same rank, then they cannot be similar.*

(a) The matrices

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix}$$

are not similar, since their traces are different.

(b) The matrices

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

have the same trace (zero), but they have different ranks: the first has rank 1 and the second has rank 2. Therefore they are not similar.

Remark 4. *It is important to note that two matrices can have the same trace and the same rank without being similar.*

(a) The matrices
$$A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}, \qquad H = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$
are not similar. Indeed, $H = 2I_2$, so for every invertible matrix P,
$$P^{-1}HP = H,$$
and therefore it is impossible to have $P^{-1}HP = A$.

Question: Do there exist matrices $S, T \in M_n(K)$ such that
$$ST - TS = I_n\,?$$

Proposition 3. *Let n be a nonnegative integer. The kernel of the trace map*
$$\mathrm{tr} : M_n(K) \to K$$
is the subspace generated by the set of matrices of the form
$$AB - BA,$$
with $A, B \in M_n(K)$.

Proof. Challenging exercise. □

Proposition 4. *Let n be a nonnegative integer. If $t : M_n(K) \to K$ is a linear map such that*
$$t(AB) = t(BA) \quad \textit{for all } A, B \in M_n(K),$$
then there exists a scalar $\lambda \in K$ such that
$$t = \lambda\,\mathrm{tr}\,.$$

Proof. Let $t : M_n(K) \to K$ satisfy the hypothesis of the statement, and assume $t \neq 0$ (otherwise the conclusion is immediate). If $A, B \in M_n(K)$, then
$$t(AB - BA) = t(AB) - t(BA) = 0,$$
so $AB - BA \in \ker(t)$.

By the previous Proposition , this implies
$$\ker(\mathrm{tr}) \subset \ker(t).$$

Since t is not the zero map and its codomain has dimension 1, it is surjective. Hence, by the Rank–Nullity Theorem,
$$\dim\ker(t) = \dim M_n(K) - 1 = n^2 - 1 = \dim\ker(\mathrm{tr}).$$

Therefore,
$$\ker(t) = \ker(\mathrm{tr}).$$

Now let $A \in M_n(K)$. Then

$$\mathrm{tr}(A - \mathrm{tr}(A)\, E_{11}) = \mathrm{tr}(A) - \mathrm{tr}(A)\,\mathrm{tr}(E_{11}) = 0,$$

since $\mathrm{tr}(E_{11}) = 1$. Thus

$$A - \mathrm{tr}(A)\, E_{11} \in \ker(\mathrm{tr}) = \ker(t).$$

Hence

$$0 = t\big(A - \mathrm{tr}(A)\, E_{11}\big) = t(A) - \mathrm{tr}(A)\, t(E_{11}),$$

so

$$t(A) = t(E_{11})\ \mathrm{tr}(A).$$

Setting $\lambda = t(E_{11})$ gives

$$t = \lambda\ \mathrm{tr}\,.$$

$\square$

Determinants: Permutations and Alternating Multilinear Forms

1. The Symmetric Group

Definition 1. *A* ***transposition*** *in* S_n *is a permutation that exchanges two elements* $i, j \in \{1, \dots, n\}$ *and leaves all others fixed.*

It is denoted by $\tau = (i\, j)$ *with* $i \neq j$*, and satisfies:*

$$\tau(i) = j, \quad \tau(j) = i, \quad \tau(x) = x \text{ for } x \neq i, j.$$

Definition 2. *A* ***cycle of length*** k *is a permutation*

$$\sigma = (x_1\, x_2\, \dots\, x_k)$$

where $x_1, \dots, x_k$ *are distinct elements of* $\{1, \dots, n\}$*, defined by*

$$\sigma(x_1) = x_2,\ \dots,\ \sigma(x_{k-1}) = x_k,\ \sigma(x_k) = x_1,$$

and fixing all other elements.

Example 1. S_2 *has order* 2*:*

$$S_2 = \{\mathrm{Id}, (1\,2)\}.$$

It is commutative.

S_3 *has order* 6*: it consists of the identity, three transpositions*

$$(1\,2),\ (1\,3),\ (2\,3),$$

and two 3-cycles

$$(1\,2\,3),\ (1\,3\,2).$$

It is not commutative.

Theorem 1. *Every permutation* $\sigma \in S_n$ *can be written as a product of transpositions.*

$$\sigma = \tau_1 \tau_2 \cdots \tau_k.$$

2. Inversions and Signature

Definition 3. *Let* $\sigma \in S_n$*. A pair* (i, j) *with* $i < j$ *is called an* ***inversion*** *if*

$$\sigma(i) > \sigma(j).$$

We denote by $I(\sigma)$ *the number of inversions.*

Definition 4. *The* ***signature*** *of* σ *is*

$$\varepsilon(\sigma) = (-1)^{I(\sigma)}.$$

It can also be defined by

$$\varepsilon(\sigma) = \prod_{1 \le i < j \le n} \frac{\sigma(j) - \sigma(i)}{j - i}.$$

A permutation is called

- even if $\varepsilon(\sigma) = +1$,
- odd if $\varepsilon(\sigma) = -1$.

Example 2. *(a) Let*

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 3 & 4 & 5 & 1 & 7 & 6 & 8 \end{pmatrix}.$$

We compute:

$$\sigma(1) = 2, \quad \sigma(2) = 3, \quad \sigma(3) = 4, \quad \sigma(4) = 5, \quad \sigma(5) = 1,$$

which gives the first cycle

$$\gamma_1 = (1\,2\,3\,4\,5).$$

Next,

$$\sigma(6) = 7, \quad \sigma(7) = 6,$$

which gives the second cycle

$$\gamma_2 = (6\,7).$$

Thus,

$$\sigma = \gamma_1\gamma_2 = \gamma_2\gamma_1,$$

and

$$\varepsilon(\sigma) = \varepsilon(\gamma_1)\varepsilon(\gamma_2) = (-1)^{5-1} \times (-1) = -1.$$

(b) Let

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 1 & 5 & 4 \end{pmatrix} = (1\,2\,3)(4\,5).$$

Therefore,

$$\varepsilon(\sigma) = \varepsilon((1\,2\,3))\varepsilon((4\,5)) = (-1)^{3-1} \times (-1) = -1.$$

Proposition 1 (Basic properties of the signature)**.** *The signature satisfies:*

1. $\varepsilon(\mathrm{Id}) = 1$.
2. *For all $\sigma, \rho \in S_n$,*
$$\varepsilon(\sigma\rho) = \varepsilon(\sigma)\,\varepsilon(\rho).$$
3. *For all $\sigma \in S_n$,*
$$\varepsilon(\sigma^{-1}) = \varepsilon(\sigma).$$

Proposition 2 (Signature of a cycle)**.** *If $c \in S_n$ is a cycle of length k, then*

$$\varepsilon(c) = (-1)^{k-1}.$$

In particular, a transposition $\tau = (i\,j)$ has signature -1, hence it is an odd permutation.

Proof. Let $c = (x_1\,x_2\,\ldots\,x_k)$. Then

$$c = (x_1\,x_k)(x_1\,x_{k-1})\cdots(x_1\,x_2),$$

which is a product of $k-1$ transpositions. Since each transposition has signature -1, we obtain

$$\varepsilon(c) = (-1)^{k-1}.$$

□

Theorem 2 (Disjoint cycle decomposition)**.** *Every permutation $\sigma \in S_n$ can be written as a product of disjoint cycles. Moreover, this decomposition is unique up to the order of the cycles.*

Example 3. *Let*

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 3 & 4 & 5 & 1 & 7 & 6 & 8 \end{pmatrix}.$$

We compute the cycles.
Starting with 1:

$$1 \mapsto 2 \mapsto 3 \mapsto 4 \mapsto 5 \mapsto 1,$$

so

$$\gamma_1 = (1\,2\,3\,4\,5).$$

Next,

$$6 \mapsto 7, \quad 7 \mapsto 6,$$

so

$$\gamma_2 = (6\,7).$$

Finally, $\sigma(8) = 8$, so 8 is fixed.
Thus,

$$\sigma = \gamma_1\gamma_2 = \gamma_2\gamma_1.$$

Using the previous proposition,

$$\varepsilon(\gamma_1) = (-1)^{5-1} = 1, \qquad \varepsilon(\gamma_2) = (-1)^{2-1} = -1.$$

Therefore,

$$\varepsilon(\sigma) = \varepsilon(\gamma_1)\varepsilon(\gamma_2) = 1 \cdot (-1) = -1.$$

3. Multilinear Maps

Let E and F be vector spaces over a field $\mathbb{K}$.

Definition 5. *A p-**linear map** from E to F is a map*

$$f : E^p = E \times \cdots \times E \longrightarrow F$$

such that for every family $(v_1, v_2, \ldots, v_p) \in E^p$ and for every $i \in \{1, \ldots, p\}$, the map

$$E \longrightarrow F$$

$$x \longmapsto f(v_1, \ldots, v_{i-1}, x, v_{i+1}, \ldots, v_p)$$

is linear.

*A p-**linear form** on E is a p-linear map from E to $\mathbb{K}$.*

Example 4. *1. The map $\mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ given by*

$$((x_1, x_2), (y_1, y_2)) \mapsto x_1y_1 + x_2y_2$$

is bilinear.

2. The map

$$(u,v) \mapsto \int_0^1 u(t)v(t)\,dt$$

is bilinear on $C^0([0,1])$.

If $(e_1,\dots,e_n)$ is a basis of E, and

$$u = \sum a_i e_i, \qquad v = \sum b_j e_j,$$

then

$$f(u,v) = \sum_{i,j} a_i b_j f(e_i,e_j).$$

Alternating p-linear maps

We now study an important class of p-linear maps.

Definition 6. *A p-linear map $f : E^p \to F$ is said to be* ***alternating*** *if for every $(u_1,\dots,u_p) \in E^p$ and for all $i \neq j$,*

$$u_i = u_j \quad \Longrightarrow \quad f(u_1,\dots,u_p) = 0.$$

Proposition 3. *A p-linear map f is alternating if and only if it is* ***antisymmetric****, that is, for all $(u_1,\dots,u_p) \in E^p$ and for all $i < j$,*

$$f(u_1,\dots,u_i,\dots,u_j,\dots,u_p) = -\,f(u_1,\dots,u_j,\dots,u_i,\dots,u_p).$$

Proof. Suppose first that f is alternating. Let $i < j$ and fix $u_1,\dots,u_p$. Define the map

$$g : E^2 \to F$$

by

$$g(x,y) = f(u_1,\dots,x,\dots,y,\dots,u_p),$$

where x is placed in the i-th position and y in the j-th position.

Then g is bilinear and satisfies

$$\forall x \in E, \quad g(x,x) = 0.$$

We claim that

$$\forall (x,y) \in E^2, \quad g(x,y) = -g(y,x).$$

Indeed, by bilinearity,

$$g(x+y,x+y) = g(x,x) + g(x,y) + g(y,x) + g(y,y).$$

Since $g(x,x) = g(y,y) = 0$ and $g(x+y,x+y) = 0$, we obtain

$$g(x,y) + g(y,x) = 0,$$

which proves the claim.

Therefore,

$$f(u_1,\dots,u_i,\dots,u_j,\dots,u_p) = -\,f(u_1,\dots,u_j,\dots,u_i,\dots,u_p).$$

Conversely, suppose f is antisymmetric and let $i \neq j$ with $u_i = u_j$. Exchanging u_i and u_j changes the sign of f, but the value itself does not change since the two vectors are equal. Hence

$$f(u_1, \ldots, u_p) = -f(u_1, \ldots, u_p),$$

so

$$f(u_1, \ldots, u_p) = 0.$$

□

Examples and Consequences of Alternating Maps

Example 5. *Let E be a vector space of dimension 2 with basis (e_1, e_2).*

- *The bilinear form*
$$(x_1e_1 + x_2e_2,\ y_1e_1 + y_2e_2) \longmapsto x_1y_2 - x_2y_1$$
is alternating.
- *The bilinear form*
$$(x_1e_1 + x_2e_2,\ y_1e_1 + y_2e_2) \longmapsto x_1y_1 + x_2y_2$$
is not alternating.

Proposition 4. *Let f be an alternating p-linear map from E to F. If $(u_1, u_2, \ldots, u_p)$ is a linearly dependent family in E, then*

$$f(u_1, u_2, \ldots, u_p) = 0.$$

Proof. Since the family $(u_1, u_2, \ldots, u_p)$ is linearly dependent, there exists an index i such that

$$u_i = \sum_{k \neq i} \lambda_k u_k.$$

By multilinearity of f in the i-th variable,

$$f(u_1, \ldots, u_p) = f(u_1, \ldots, u_{i-1}, \sum_{k \neq i} \lambda_k u_k, u_{i+1}, \ldots, u_p) = \sum_{k \neq i} \lambda_k f(u_1, \ldots, u_{i-1}, u_k, u_{i+1}, \ldots, u_p).$$

For each $k \neq i$, the family

$$(u_1, \ldots, u_{i-1}, u_k, u_{i+1}, \ldots, u_p)$$

contains two equal vectors, namely u_k in positions k and i. Since f is alternating, each term in the sum is zero. Therefore,

$$f(u_1, u_2, \ldots, u_p) = 0.$$

□

Let f be an alternating p-linear map from E to F and let $(u_1, u_2, \ldots, u_p) \in E^p$. The value $f(u_1, u_2, \ldots, u_p)$ is unchanged if one adds to one of the vectors u_j a linear combination of the others.

Proof. Let x be a linear combination of the family

$$(u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_p).$$

By the previous proposition,

$$f(u_1, \dots, u_{j-1}, x, u_{j+1}, \dots, u_p) = 0.$$

By multilinearity of f in the j-th variable,

$$\begin{aligned} f(u_1, \dots, u_{j-1}, u_j + x, u_{j+1}, \dots, u_p) &= f(u_1, \dots, u_p) + f(u_1, \dots, u_{j-1}, x, u_{j+1}, \dots, u_p) \\ &= f(u_1, \dots, u_p). \end{aligned}$$

□

Examples and Consequences of Alternating Maps

Example 6. *Let E be a vector space of dimension 2 with basis (e_1, e_2).*

- *The bilinear form*
$$(x_1e_1 + x_2e_2,\ y_1e_1 + y_2e_2) \longmapsto x_1y_2 - x_2y_1$$
is alternating.

- *The bilinear form*
$$(x_1e_1 + x_2e_2,\ y_1e_1 + y_2e_2) \longmapsto x_1y_1 + x_2y_2$$
is not alternating.

Proposition 5. *Let f be an alternating p-linear map from E to F. If $(u_1, u_2, \dots, u_p)$ is a linearly dependent family in E, then*

$$f(u_1, u_2, \dots, u_p) = 0.$$

Proof. Since the family $(u_1, u_2, \dots, u_p)$ is linearly dependent, there exists an index i such that

$$u_i = \sum_{k \neq i} \lambda_k u_k.$$

By multilinearity of f in the i-th variable,

$$f(u_1, \dots, u_p) = f(u_1, \dots, u_{i-1}, \sum_{k \neq i} \lambda_k u_k, u_{i+1}, \dots, u_p) = \sum_{k \neq i} \lambda_k f(u_1, \dots, u_{i-1}, u_k, u_{i+1}, \dots, u_p).$$

For each $k \neq i$, the family

$$(u_1, \dots, u_{i-1}, u_k, u_{i+1}, \dots, u_p)$$

contains two equal vectors, namely u_k in positions k and i. Since f is alternating, each term in the sum is zero. Therefore,

$$f(u_1, u_2, \dots, u_p) = 0.$$

□

Let f be an alternating p-linear map from E to F and let $(u_1, u_2, \dots, u_p) \in E^p$. The value $f(u_1, u_2, \dots, u_p)$ is unchanged if one adds to one of the vectors u_j a linear combination of the others.

Proof. Let x be a linear combination of the family

$$(u_1, \dots, u_{j-1}, u_{j+1}, \dots, u_p).$$

By the previous proposition,

$$f(u_1, \dots, u_{j-1}, x, u_{j+1}, \dots, u_p) = 0.$$

By multilinearity of f in the j-th variable,

$$\begin{aligned} f(u_1, \dots, u_{j-1}, u_j + x, u_{j+1}, \dots, u_p) &= f(u_1, \dots, u_p) + f(u_1, \dots, u_{j-1}, x, u_{j+1}, \dots, u_p) \\ &= f(u_1, \dots, u_p). \end{aligned}$$

□

Proposition 6. *Let f be an alternating n-linear map from E to F and let $\sigma \in S_n$ be a permutation. Then for every $(u_1, u_2, \dots, u_n) \in E^n$,*

$$f(u_{\sigma(1)}, u_{\sigma(2)}, \dots, u_{\sigma(n)}) = \varepsilon(\sigma)\, f(u_1, u_2, \dots, u_n).$$

Proof. By the previous theorem, every permutation σ can be written as a product of transpositions:

$$\sigma = \tau_1 \tau_2 \cdots \tau_k.$$

Applying the antisymmetry property successively, we obtain

$$\begin{aligned} f(u_{\sigma(1)}, u_{\sigma(2)}, \dots, u_{\sigma(n)}) &= \varepsilon(\tau_1)\, f(u_{\tau_2\cdots\tau_k(1)}, u_{\tau_2\cdots\tau_k(2)}, \dots, u_{\tau_2\cdots\tau_k(n)}) \\ &= \varepsilon(\tau_1)\varepsilon(\tau_2)\, f(u_{\tau_3\cdots\tau_k(1)}, u_{\tau_3\cdots\tau_k(2)}, \dots, u_{\tau_3\cdots\tau_k(n)}) \\ &\ \ \vdots \\ &= \varepsilon(\tau_1)\varepsilon(\tau_2)\cdots\varepsilon(\tau_k)\, f(u_1, u_2, \dots, u_n). \end{aligned}$$

Since the signature of a product is the product of the signatures,

$$\varepsilon(\tau_1)\varepsilon(\tau_2)\cdots\varepsilon(\tau_k) = \varepsilon(\tau_1\tau_2\cdots\tau_k) = \varepsilon(\sigma),$$

which proves the result. □

Determinants: Permutations and Alternating Multilinear Forms

0.1 Expression of an alternating n-linear map in dimension n

In this subsection E is a K-vector space of dimension n. Previously, we have seen that if f is an alternating n-linear map, then for every transposition $\tau \in S_n$ and for every $u \in E^n$ we have

$$f(u_{\tau(1)}, u_{\tau(2)}, \ldots, u_{\tau(n)}) = -f(u_1, u_2, \ldots, u_n),$$

or equivalently

$$f(u_{\tau(1)}, u_{\tau(2)}, \ldots, u_{\tau(n)}) = \varepsilon(\tau) f(u_1, u_2, \ldots, u_n), \tag{2}$$

where ε denotes the signature of S_n.

Proposition 1. *Let f be an alternating n-linear map from E to F and let σ be a permutation in S_n. Then for every $(u_1, u_2, \ldots, u_n) \in E^n$,*

$$f(u_{\sigma(1)}, u_{\sigma(2)}, \ldots, u_{\sigma(n)}) = \varepsilon(\sigma) f(u_1, u_2, \ldots, u_n).$$

Proof. By Theorem seen in previous lecture, σ can be written as a product of transpositions

$$\sigma = \tau_1 \tau_2 \cdots \tau_k.$$

Using equality (2) successively we obtain

$$f(u_{\sigma(1)}, u_{\sigma(2)}, \ldots, u_{\sigma(n)}) = \varepsilon(\tau_1) f(u_{\tau_2 \cdots \tau_k(1)}, u_{\tau_2 \cdots \tau_k(2)}, \ldots, u_{\tau_2 \cdots \tau_k(n)})$$

$$= \varepsilon(\tau_1)\varepsilon(\tau_2) f(u_{\tau_3 \cdots \tau_k(1)}, u_{\tau_3 \cdots \tau_k(2)}, \ldots, u_{\tau_3 \cdots \tau_k(n)})$$

$$\cdots$$

$$= \varepsilon(\tau_1)\varepsilon(\tau_2) \cdots \varepsilon(\tau_k) f(u_1, u_2, \ldots, u_n).$$

Since the signature of a product is the product of the signatures, hence

$$\varepsilon(\tau_1)\varepsilon(\tau_2) \cdots \varepsilon(\tau_k) = \varepsilon(\tau_1 \circ \tau_2 \circ \cdots \circ \tau_k) = \varepsilon(\sigma).$$

This proves the result. □

Proposition 2. *Let φ be an alternating n-linear map on E and let $B = (e_1, e_2, \ldots, e_n)$ be a basis of E. If $u_1, u_2, \ldots, u_n$ are vectors such that*

$$\forall j \in \{1, \ldots, n\}, \qquad u_j = \sum_{i=1}^{n} a_{i,j} e_i,$$

then

$$\varphi(u_1, u_2, \ldots, u_n) = \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}\, \varphi(e_1, e_2, \ldots, e_n). \tag{3}$$

Proof. One can check (exercise) that we have

$$\varphi(u_1, u_2, \ldots, u_n) = \sum_{1 \le i_1, \ldots, i_n \le n} a_{i_1,1} a_{i_2,2} \cdots a_{i_n,n} \varphi(e_{i_1}, e_{i_2}, \ldots, e_{i_n}).$$

We can rewrite the sum as

$$\sum_{\sigma \in \{1,\ldots,n\}^{\{1,\ldots,n\}}} a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n} \varphi(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}).$$

When σ is not a permutation (i.e. not a bijection), the family

$$(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)})$$

contains at least two equal vectors. Since φ is alternating, the image is therefore zero. Hence the sum reduces to permutations:

$$\varphi(u_1, u_2, \ldots, u_n) = \sum_{\sigma \in S_n} a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n} \varphi(e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}).$$

Using previous Proposition we obtain

$$\varphi(u_1, u_2, \ldots, u_n) = \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n} \varphi(e_1, e_2, \ldots, e_n).$$

□

Proposition 3. *The set of alternating p-linear maps from E to F is nonempty and stable under linear combinations. Hence it is a vector subspace of the vector space $F(E^p, F)$. In particular, the set of alternating p-linear forms on E is a vector space, denoted $\Lambda^p(E)$.*

Proof. Exercise □

0.2 Alternating n-linear forms on a vector space of dimension n

In this subsection we also assume that E is a K-vector space of dimension n. We first characterize alternating n-linear forms in dimensions 2 and 3 in order to understand the general case.

Case of dimension 2. Let E be a K-vector space of dimension 2 with basis (e_1, e_2) and let φ be an alternating bilinear form on E. Let u and v be two vectors in E written as

$$u = a_1 e_1 + a_2 e_2, \qquad v = b_1 e_1 + b_2 e_2.$$

By bilinearity of φ we obtain

$$\varphi(u, v) = \varphi(a_1 e_1 + a_2 e_2, b_1 e_1 + b_2 e_2)$$

$$= a_1 b_1 \varphi(e_1, e_1) + a_1 b_2 \varphi(e_1, e_2) + a_2 b_1 \varphi(e_2, e_1) + a_2 b_2 \varphi(e_2, e_2).$$

Since φ is alternating we have

$$\varphi(e_1, e_1) = \varphi(e_2, e_2) = 0, \qquad \varphi(e_2, e_1) = -\varphi(e_1, e_2),$$

hence

$$\varphi(u,v) = (a_1b_2 - a_2b_1)\varphi(e_1,e_2).$$

Therefore φ is proportional to the alternating bilinear form

$$\varphi_0 : E^2 \to K, \qquad (u,v) \mapsto a_1b_2 - a_2b_1. \tag{5}$$

Moreover, if we impose $\varphi(e_1,e_2) = 1$, then $\varphi = \varphi_0$.

Case of dimension 3. Let E be a K-vector space of dimension 3 with basis (e_1,e_2,e_3) and let φ be an alternating trilinear form on E. Let u_1,u_2,u_3 be three vectors in E written as

$$\forall j \in \{1,2,3\}, \qquad u_j = \sum_{i=1}^{3} a_{i,j}e_i.$$

By trilinearity we obtain

$$\varphi(u_1,u_2,u_3) = \sum_{k=1}^{3}\sum_{l=1}^{3}\sum_{m=1}^{3} a_{k,1}a_{l,2}a_{m,3}\varphi(e_k,e_l,e_m).$$

Since φ is alternating, $\varphi(e_k,e_l,e_m) = 0$ whenever k,l,m are not all distinct. Moreover,

$$\varphi(e_1,e_3,e_2) = \varphi(e_3,e_2,e_1) = \varphi(e_2,e_1,e_3) = -\varphi(e_1,e_2,e_3),$$

$$\varphi(e_2,e_3,e_1) = -\varphi(e_3,e_2,e_1) = \varphi(e_1,e_2,e_3),$$

$$\varphi(e_3,e_1,e_2) = -\varphi(e_1,e_3,e_2) = \varphi(e_1,e_2,e_3).$$

Hence we obtain

$$\varphi(u_1,u_2,u_3) = (+a_{1,1}a_{2,2}a_{3,3} - a_{1,1}a_{3,2}a_{2,3} + a_{2,1}a_{3,2}a_{1,3}$$

$$-a_{2,1}a_{1,2}a_{3,3} + a_{3,1}a_{1,2}a_{2,3} - a_{3,1}a_{2,2}a_{1,3})\varphi(e_1,e_2,e_3).$$

Therefore φ is proportional to the alternating trilinear form

$$\varphi_0 : E^3 \to K,$$

$$(u_1,u_2,u_3) \mapsto +a_{1,1}a_{2,2}a_{3,3} - a_{1,1}a_{3,2}a_{2,3} + a_{2,1}a_{3,2}a_{1,3}$$

$$-a_{2,1}a_{1,2}a_{3,3} + a_{3,1}a_{1,2}a_{2,3} - a_{3,1}a_{2,2}a_{1,3}. \tag{6}$$

If we impose $\varphi(e_1,e_2,e_3) = 1$, then $\varphi = \varphi_0$.

Theorem 1. *Let $B = (e_1,e_2,\dots,e_n)$ be a basis of E.*

(a) There exists a unique alternating n-linear form φ_0 on E such that

$$\varphi_0(e_1,e_2,\dots,e_n) = 1.$$

(b) Every alternating n-linear form on E is proportional to φ_0.

Proof. Let $u_j = \sum_{i=1}^{n} a_{i,j} e_i$ for every $j \in \{1, \dots, n\}$.

Define $\varphi_0 : E^n \to K$ by

$$\varphi_0(u_1, u_2, \dots, u_n) = \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}. \tag{7}$$

Formula (3) shows that every alternating n-linear form φ is proportional to φ_0, and more precisely

$$\varphi = \varphi(e_1, e_2, \dots, e_n)\varphi_0.$$

In particular, if $\varphi(e_1, e_2, \dots, e_n) = 1$, then $\varphi = \varphi_0$, which proves uniqueness. It remains to show that φ_0 is an alternating n-linear form satisfying $\varphi_0(e_1, e_2, \dots, e_n) = 1$.

The expression of φ_0 clearly shows that it is n-linear.

Let us prove that if $u_i = u_j$ for $i \neq j$, then

$$\varphi_0(u_1, u_2, \dots, u_n) = 0.$$

Let $\tau = (ij)$ be the transposition exchanging i and j, and let $A_n \subset S_n$ be the set of even permutations. The set of odd permutations is

$$A_n \tau = \{\sigma\tau \mid \sigma \in A_n\}.$$

Hence we can write

$$\begin{aligned}\varphi_0(u_1, u_2, \dots, u_n) &= \sum_{\sigma \in S_n} \varepsilon(\sigma) \prod_{k=1}^{n} a_{\sigma(k),k} \\ &= \sum_{\sigma \in A_n} \prod_{k=1}^{n} a_{\sigma(k),k} - \sum_{\sigma \in A_n} \prod_{k=1}^{n} a_{\sigma\tau(k),k}.\end{aligned}$$

Now we have

$$a_{\sigma\tau(i),i} = a_{\sigma(j),i} = a_{\sigma(j),j}, \qquad a_{\sigma\tau(j),j} = a_{\sigma(i),j} = a_{\sigma(i),i},$$

since $u_i = u_j$.

Moreover,

$$a_{\sigma\tau(k),k} = a_{\sigma(k),k} \quad \text{for } k \notin \{i, j\}.$$

Hence

$$\prod_{k=1}^{n} a_{\sigma\tau(k),k} = \prod_{k=1}^{n} a_{\sigma(k),k},$$

and therefore

$$\varphi_0(u_1, u_2, \dots, u_n) = 0.$$

Finally, let us verify that

$$\varphi_0(e_1, e_2, \dots, e_n) = 1.$$

By definition of φ_0,

$$\varphi_0(e_1, e_2, \dots, e_n) = \sum_{\sigma \in S_n} \varepsilon(\sigma) \delta_{\sigma(1),1} \delta_{\sigma(2),2} \cdots \delta_{\sigma(n),n},$$

where $(\delta_{i,j})_{1 \le i,j \le n}$ is the matrix of the components of $e_1, e_2, \dots, e_n$ in the basis $(e_1, e_2, \dots, e_n)$, that is, the identity matrix I_n.

In the above sum:

- if σ is not the identity, there exists i such that $\sigma(i) \neq i$, hence $\delta_{\sigma(i),i} = 0$, and therefore the product vanishes;
- if $\sigma = \mathrm{Id}$, then
$$\varepsilon(\sigma)\delta_{\sigma(1),1}\delta_{\sigma(2),2}\cdots\delta_{\sigma(n),n} = 1.$$

Hence
$$\varphi_0(e_1, e_2, \ldots, e_n) = 1.$$

□

Corollary 1. *Let E be a K-vector space of dimension n. The space $\Lambda^n(E)$ of alternating n-linear forms on E is a vector space of dimension 1, generated by φ_0.*

Definition of determinants

We now introduce the notion of determinant using the alternating n-linear form constructed in the previous section.

Definition 1. *Let E be a K-vector space of dimension n and let $B = (e_1, e_2, \dots, e_n)$ be a basis of E. Let φ_0 be the unique alternating n-linear form such that*

$$\varphi_0(e_1, e_2, \dots, e_n) = 1.$$

For vectors $u_1, u_2, \dots, u_n \in E$, the scalar

$$\det_B(u_1, u_2, \dots, u_n) = \varphi_0(u_1, u_2, \dots, u_n)$$

*is called the **determinant of the family of vectors** $(u_1, u_2, \dots, u_n)$ with respect to the basis B.*

The map

$$(u_1, u_2, \dots, u_n) \mapsto \det_B(u_1, u_2, \dots, u_n)$$

is called the determinant associated with the basis B.

Since $\Lambda^n(E)$ is a vector space of dimension 1, every alternating n-linear form on E is proportional to $\det_B$.

In particular, if B' is another basis of E, then there exists a scalar λ such that

$$\det_{B'} = \lambda \det_B .$$

Evaluating at B gives

$$\lambda = \det_{B'}(B),$$

and therefore for every $(u_1, \dots, u_n) \in E^n$ we have

$$\det_{B'}(u_1, \dots, u_n) = \det_{B'}(B) \det_B(u_1, \dots, u_n).$$

0.1 Determinant of an endomorphism

Proposition 1. *Let $f : E \to E$ be a linear map. There exists a unique scalar λ such that for every basis B and every $(u_1, \dots, u_n) \in E^n$,*

$$\det_B(f(u_1), f(u_2), \dots, f(u_n)) = \lambda \det_B(u_1, u_2, \dots, u_n).$$

*This scalar is called the **determinant of** f and is denoted*

$$\det(f).$$

Moreover, for any basis B,

$$\det(f) = \det_B(f(B)).$$

Proof. **Uniqueness.** Suppose that a scalar λ satisfies

$$\det_B(f(u_1), f(u_2), \ldots, f(u_n)) = \lambda \det_B(u_1, u_2, \ldots, u_n)$$

for every $(u_1, u_2, \ldots, u_n) \in E^n$.

Taking in particular $(u_1, u_2, \ldots, u_n) = B$, we obtain

$$\lambda = \det_B(f(B)).$$

Hence λ is uniquely determined.

Existence. Fix a basis B_0 of E. The map

$$(u_1, u_2, \ldots, u_n) \longmapsto \det_{B_0}(f(u_1), f(u_2), \ldots, f(u_n))$$

from E^n to K is clearly an alternating n-linear form. Therefore it is proportional to $\det_{B_0}$. Hence there exists a scalar λ such that for every $(u_1, u_2, \ldots, u_n) \in E^n$,

$$\det_{B_0}(f(u_1), f(u_2), \ldots, f(u_n)) = \lambda \det_{B_0}(u_1, u_2, \ldots, u_n). \tag{8}$$

Now let B be any basis of E. Since $\det_B$ is an alternating n-linear form, it is proportional to $\det_{B_0}$. Thus there exists a scalar α such that

$$\det_B = \alpha \det_{B_0}.$$

Multiplying relation (8) by α, we obtain, for every $(u_1, u_2, \ldots, u_n) \in E^n$,

$$\alpha \det_{B_0}(f(u_1), f(u_2), \ldots, f(u_n)) = \lambda \, \alpha \det_{B_0}(u_1, u_2, \ldots, u_n).$$

Since $\det_B = \alpha \det_{B_0}$, this becomes

$$\det_B(f(u_1), f(u_2), \ldots, f(u_n)) = \lambda \det_B(u_1, u_2, \ldots, u_n).$$

This proves the result. $\square$

In-class Exercise

Let $(e_1, \ldots, e_n)$ be a basis of a vector space E and let

$$a = \sum_{i=1}^{n} \alpha_i e_i.$$

Compute

$$\det(e_1 + b_1 a,\ e_2 + b_2 a,\ \ldots,\ e_n + b_n a),$$

where the determinant is taken with respect to the basis $(e_1, \ldots, e_n)$.

0.2 Determinant of a square matrix

We now consider square matrices in $M_n(K)$.

Definition 2. *Let A be a square matrix of order n. The determinant of A is defined as the determinant of its column vectors in the canonical basis of K^n. It is denoted* $\det(A)$ *or* $\det A$.

If $A = (a_{i,j})_{1\leq i,j\leq n}$, then the determinant of A is written

$$\det A = \begin{vmatrix} a_{1,1} & \cdots & a_{1,j} & \cdots & a_{1,n} \\ \vdots & & \vdots & & \vdots \\ a_{i,1} & \cdots & a_{i,j} & \cdots & a_{i,n} \\ \vdots & & \vdots & & \vdots \\ a_{n,1} & \cdots & a_{n,j} & \cdots & a_{n,n} \end{vmatrix}.$$

Corollary 1. *The determinant of a matrix is an alternating n-linear form of its column vectors.*

Example 1. *1. For a 2×2 matrix,*

$$\begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} = a_{1,1}a_{2,2} - a_{2,1}a_{1,2}.$$

2. For a 3×3 matrix,

$$\begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} = a_{1,1}a_{2,2}a_{3,3} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{3,2}$$

$$-a_{1,1}a_{2,3}a_{3,2} - a_{1,2}a_{2,1}a_{3,3} - a_{1,3}a_{2,2}a_{3,1}.$$

Proposition 2. *Let E be a K-vector space of dimension n. If $(u_1,\ldots,u_n)$ is a family of vectors in E whose matrix of coordinates in a basis $B = (e_i)_{1\leq i\leq n}$ is A, then*

$$\det A = \det_B(u_1,\ldots,u_n).$$

Proof. This follows directly from the previous definition. □

Example 2. *If D is a diagonal matrix with diagonal entries $\lambda_1,\ldots,\lambda_n$, then*

$$\det D = \prod_{i=1}^{n} \lambda_i.$$

Indeed, D is the matrix of the family $(\lambda_1 e_1, \lambda_2 e_2, \ldots, \lambda_n e_n)$ in the basis $B = (e_1,\ldots,e_n)$. Hence

$$\det D = \det_B(\lambda_1 e_1, \lambda_2 e_2, \ldots, \lambda_n e_n) = \left(\prod_{i=1}^{n} \lambda_i\right) \det_B(e_1,\ldots,e_n) = \prod_{i=1}^{n} \lambda_i.$$

Proposition 3. *Let f be an endomorphism of a K-vector space E of dimension n and let $B = (e_i)_{1\leq i\leq n}$ be a basis of E. If A is the matrix of f with respect to the basis B, then*

$$\det f = \det A.$$

Proof. Both determinants equal the determinant of the family $(f(e_1), f(e_2), \dots, f(e_n))$ in the basis B. $\square$

Proposition 4. *If $A = (a_{i,j})_{1 \le i,j \le n}$ is a square matrix of order n, then*

$$\det A = \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}.$$

Proof. Let $u_1, \dots, u_n$ be the column vectors of A expressed in a basis B. By the previous proposition,

$$\det A = \det_B(u_1, \dots, u_n).$$

The map

$$(u_1, \dots, u_n) \mapsto \det_B(u_1, \dots, u_n)$$

is alternating n-linear and satisfies $\det_B(B) = 1$. Hence by the uniqueness of the alternating n-linear form, it coincides with the form φ_0 defined previously. $\square$

Let us verify the formula when $n = 2$ and $n = 3$.

For $n = 2$, the symmetric group is

$$S_2 = \{\mathrm{Id}, \tau_{1,2}\},$$

where $\tau_{1,2} = (1\,2)$.

If

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix},$$

then

$$\det A = \varepsilon(\mathrm{Id}) a_{1,1} a_{2,2} + \varepsilon(\tau_{1,2}) a_{2,1} a_{1,2} = a_{1,1} a_{2,2} - a_{2,1} a_{1,2}.$$

For $n = 3$, the symmetric group is

$$S_3 = \{\mathrm{Id}, \tau_{1,2}, \tau_{1,3}, \tau_{2,3}, c_1, c_2\},$$

where

$$\tau_{1,2} = (1\,2), \quad \tau_{1,3} = (1\,3), \quad \tau_{2,3} = (2\,3), \quad c_1 = (1\,2\,3), \quad c_2 = (1\,3\,2).$$

If

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{pmatrix},$$

then

$$\det A = a_{1,1} a_{2,2} a_{3,3} + a_{1,2} a_{2,3} a_{3,1} + a_{1,3} a_{2,1} a_{3,2}$$
$$- a_{1,2} a_{2,1} a_{3,3} - a_{1,3} a_{2,2} a_{3,1} - a_{1,1} a_{2,3} a_{3,2}.$$

Proposition 5. *Let $A = (a_{i,j}) \in M_n(K)$ be an upper-triangular matrix with diagonal entries $\lambda_1, \dots, \lambda_n$. Then*

$$\det A = \lambda_1 \lambda_2 \cdots \lambda_n.$$

Proof. Recall the Leibniz formula for the determinant:

$$\det A = \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}.$$

Since A is upper-triangular, we have

$$a_{i,j} = 0 \quad \text{whenever } i > j.$$

Let $\sigma \in S_n$ be a permutation different from the identity $(1, 2, \ldots, n)$. Then there exists an index k such that

$$\sigma(k) > k.$$

The factor appearing in the corresponding term of the sum is

$$a_{\sigma(k),k},$$

which lies strictly below the diagonal. Hence

$$a_{\sigma(k),k} = 0.$$

Therefore every term in the Leibniz sum corresponding to a permutation $\sigma \neq (1, \ldots, n)$ vanishes.

The only permutation that gives a nonzero contribution is the identity permutation. The corresponding term is

$$a_{1,1} a_{2,2} \cdots a_{n,n}.$$

Since $a_{k,k} = \lambda_k$ for each $k = 1, \ldots, n$, we obtain

$$\det A = \lambda_1 \lambda_2 \cdots \lambda_n.$$

$\square$

Properties of the determinant

Throughout this section, E is a K-vector space of dimension n.

Theorem 1. *Let B be a basis of E. If $(u_1, u_2, \ldots, u_n)$ is a family of n vectors in E, then the following properties are equivalent:*

1. *$(u_1, u_2, \ldots, u_n)$ is a basis of E;*

2. *$\det_B(u_1, u_2, \ldots, u_n) \neq 0$.*

Proof. (1) $\Rightarrow$ (2). Assume that $B' = (u_1, u_2, \ldots, u_n)$ is a basis of E. Then $\det_{B'}$ is an alternating n-linear form on E, so since the space of alternating n-linear forms is one-dimensional, it is proportional to $\det_B$. Hence there exists $\lambda \in K$ such that

$$\det_{B'}(v_1, v_2, \ldots, v_n) = \lambda \det_B(v_1, v_2, \ldots, v_n)$$

for all $(v_1, v_2, \ldots, v_n) \in E^n$.

Applying this to $(u_1, u_2, \ldots, u_n)$, we obtain

$$\det_{B'}(u_1, u_2, \ldots, u_n) = \lambda \det_B(u_1, u_2, \ldots, u_n).$$

Since $(u_1, u_2, \ldots, u_n) = B'$, we have

$$\det_{B'}(u_1, u_2, \ldots, u_n) = 1.$$

Therefore,

$$\det_B(u_1, u_2, \ldots, u_n) \neq 0.$$

(2) $\Rightarrow$ (1). We prove the contrapositive. If $(u_1, u_2, \ldots, u_n)$ is not a basis of E, then the family is linearly dependent. Since $\det_B$ is alternating, it vanishes on every linearly dependent family. Hence

$$\det_B(u_1, u_2, \ldots, u_n) = 0.$$

This proves the equivalence. $\square$

Proposition 1. *Let f and g be endomorphisms of E, and let $\lambda \in K$. Then:*

1. *$\det(\lambda f) = \lambda^n \det(f)$;*

2. *$\det(f \circ g) = \det(f)\det(g)$.*

Proof. Let $B = (e_1, e_2, \ldots, e_n)$ be a basis of E. Then

$$\det(\lambda f) = \det_B(\lambda f(e_1), \lambda f(e_2), \ldots, \lambda f(e_n)) = \lambda^n \det_B(f(e_1), f(e_2), \ldots, f(e_n)),$$

so

$$\det(\lambda f) = \lambda^n \det(f).$$

Also,

$$\det(f \circ g) = \det_B(f(g(e_1)), f(g(e_2)), \ldots, f(g(e_n))).$$

By the defining property of $\det(f)$,

$$\det_B(f(g(e_1)), \ldots, f(g(e_n))) = \det(f) \det_B(g(e_1), \ldots, g(e_n)).$$

Hence

$$\det(f \circ g) = \det(f) \det(g).$$

$\square$

Proposition 2. *Let $A, B \in M_n(K)$ and let $\lambda \in K$. Then:*

1. $\det(\lambda A) = \lambda^n \det(A)$;
2. $\det(AB) = \det(A) \det(B)$.

Proof. Let f be the endomorphism of K^n canonically associated with A. The matrix of λf in the canonical basis is λA. Therefore,

$$\det(\lambda A) = \det(\lambda f) = \lambda^n \det(f) = \lambda^n \det(A).$$

Now let g be the endomorphism of K^n canonically associated with B. The matrix of $f \circ g$ in the canonical basis is AB. Thus,

$$\det(AB) = \det(f \circ g) = \det(f) \det(g) = \det(A) \det(B).$$

$\square$

Proposition 3. 1. *An endomorphism f of E is bijective if and only if $\det(f) \neq 0$. In that case,*

$$\det(f^{-1}) = \frac{1}{\det(f)}.$$

2. *A matrix $A \in M_n(K)$ is invertible if and only if $\det(A) \neq 0$. In that case,*

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

Proof. 1. Let $B = (e_1, e_2, \ldots, e_n)$ be a basis of E. Then

$$\det(f) = \det_B(f(e_1), f(e_2), \ldots, f(e_n)).$$

By the theorem above, $\det(f) \neq 0$ if and only if $(f(e_1), f(e_2), \ldots, f(e_n))$ is a basis of E, that is, if and only if f is bijective.

If f is invertible, then

$$f \circ f^{-1} = \mathrm{Id}_E.$$

Taking determinants and using multiplicativity, we obtain

$$\det(f) \det(f^{-1}) = \det(\mathrm{Id}_E) = 1.$$

Hence

$$\det(f^{-1}) = \frac{1}{\det(f)}.$$

2. Let f be the endomorphism of K^n canonically associated with A. Then

$$\det(A) = \det(f),$$

and A is invertible if and only if f is bijective. By part (1), this is equivalent to $\det(A) \neq 0$. If A is invertible, then

$$\det(A)\det(A^{-1}) = \det(I_n) = 1,$$

so

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

□

Corollary 1. *1. Let f and g be endomorphisms of E, with g invertible. Then*

$$\det(g \circ f \circ g^{-1}) = \det(f).$$

2. Let $A, B \in M_n(K)$ with B invertible. Then

$$\det(BAB^{-1}) = \det(A).$$

Proof. Using multiplicativity and the formula for the determinant of the inverse, we get

$$\det(g \circ f \circ g^{-1}) = \det(g)\det(f)\det(g^{-1}) = \det(g)\det(f)\frac{1}{\det(g)} = \det(f).$$

The matrix version is proved in the same way:

$$\det(BAB^{-1}) = \det(B)\det(A)\det(B^{-1}) = \det(A).$$

□

Proposition 4. *For every matrix $A \in M_n(K)$,*

$$\det(A) = \det(A^\top).$$

Proof. Consider the two maps on $(K^n)^n$ defined by

$$(C_1, C_2, \ldots, C_n) \longmapsto \det(C_1, C_2, \ldots, C_n)$$

and

$$(C_1, C_2, \ldots, C_n) \longmapsto \det \begin{pmatrix} C_1^\top \\ C_2^\top \\ \vdots \\ C_n^\top \end{pmatrix}.$$

Both are alternating n-linear forms on K^n. Since the space of alternating n-linear forms is one-dimensional, they are proportional. Thus there exists $\lambda \in K$ such that for every $A \in M_n(K)$,

$$\det(A) = \lambda \det(A^\top).$$

Taking $A = I_n$, we obtain

$$1 = \lambda \cdot 1,$$

hence $\lambda = 1$. Therefore,

$$\det(A) = \det(A^\top).$$

□

Corollary 2. *The determinant of a matrix is an alternating n-linear form of its rows.*

1 Let's compute...

Proposition 5. *Let $A \in M_n(K)$. Then the determinant has the following properties with respect to the rows of A:*

1. *If one multiplies a row by a scalar λ, then the determinant is multiplied by λ.*
2. *If one adds to a row a scalar multiple of another row, then the determinant does not change.*
3. *If two rows are exchanged, then the determinant changes sign.*

Proof. Since the determinant is an alternating n-linear form of the rows, properties (1) and (3) follow immediately.

Indeed, property (1) is just linearity in one row, and property (3) follows from the alternating property.

It remains to prove (2). Let $L_1, \dots, L_n$ be the rows of the matrix, and suppose that we replace the row L_i by $L_i + \lambda L_j$, where $i \neq j$.

By multilinearity of the determinant with respect to the i-th row, we have

$$\det(L_1, \dots, L_{i-1}, L_i + \lambda L_j, L_{i+1}, \dots, L_n)$$

$$= \det(L_1, \dots, L_{i-1}, L_i, L_{i+1}, \dots, L_n) + \lambda \det(L_1, \dots, L_{i-1}, L_j, L_{i+1}, \dots, L_n).$$

In the second determinant, the i-th row is L_j, while the j-th row is still L_j. Thus two rows are equal, so this determinant is equal to 0.

Therefore,

$$\det(L_1, \dots, L_{i-1}, L_i + \lambda L_j, L_{i+1}, \dots, L_n) = \det(L_1, \dots, L_n).$$

This proves (2). □

Example 1. *Let $a, b, c \in K$, and consider the determinant*

$$\Delta = \begin{vmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{vmatrix}.$$

We compute Δ using row operations.

First, replace

$$L_2 \leftarrow L_2 - L_1, \qquad L_3 \leftarrow L_3 - L_1.$$

By the previous proposition, these operations do not change the determinant. Hence

$$\Delta = \begin{vmatrix} 1 & a & a^2 \\ 0 & b-a & b^2-a^2 \\ 0 & c-a & c^2-a^2 \end{vmatrix}.$$

Now we factor $b-a$ from the second row and $c-a$ from the third row. Since

$$b^2 - a^2 = (b-a)(b+a), \qquad c^2 - a^2 = (c-a)(c+a),$$

we obtain

$$\Delta = (b-a)(c-a) \begin{vmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 1 & c+a \end{vmatrix}.$$

Next, replace

$$L_3 \leftarrow L_3 - L_2.$$

Again, this operation does not change the determinant, so

$$\Delta = (b-a)(c-a)\begin{vmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 0 & c-b \end{vmatrix}.$$

The matrix is now upper triangular. Therefore, its determinant is the product of its diagonal entries:

$$\begin{vmatrix} 1 & a & a^2 \\ 0 & 1 & b+a \\ 0 & 0 & c-b \end{vmatrix} = 1 \cdot 1 \cdot (c-b) = c-b.$$

Hence

$$\Delta = (b-a)(c-a)(c-b).$$

In-class Exercise

Let $A = [a_{i,j}] \in M_4(\mathbb{Q})$ be a matrix such that each entry is either -2 or 3. Show that

$$\det(A)$$

is an integer multiple of 125.

Cofactor expansion and Vandermonde determinants

Proposition 1. *Let*

$$A = \begin{pmatrix} & & & 0 \\ & A' & & \vdots \\ & & & 0 \\ * & \cdots & * & a_{n,n} \end{pmatrix} \in M_n(K),$$

where $A' \in M_{n-1}(K)$. Then

$$\det(A) = a_{n,n} \det(A').$$

Proof. By the permutation formula for the determinant,

$$\det(A) = \sum_{\sigma \in S_n} \varepsilon(\sigma)\, a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n),n}.$$

Since all the entries in the last column are zero except possibly the last one, any term with

$$\sigma(n) \neq n$$

vanishes. Therefore

$$\det(A) = \sum_{\substack{\sigma \in S_n \\ \sigma(n)=n}} \varepsilon(\sigma)\, a_{\sigma(1),1} a_{\sigma(2),2} \cdots a_{\sigma(n-1),n-1} a_{n,n}.$$

Now, if $\sigma \in S_n$ satisfies $\sigma(n) = n$, then its restriction to

$$\{1, \ldots, n-1\}$$

is a permutation $s \in S_{n-1}$. Conversely, every permutation $s \in S_{n-1}$ extends uniquely to a permutation $\sigma \in S_n$ such that $\sigma(n) = n$. Moreover,

$$\varepsilon(\sigma) = \varepsilon(s).$$

Hence

$$\det(A) = a_{n,n} \sum_{s \in S_{n-1}} \varepsilon(s)\, a_{s(1),1} a_{s(2),2} \cdots a_{s(n-1),n-1}.$$

The sum on the right is exactly $\det(A')$. Thus

$$\det(A) = a_{n,n} \det(A').$$

□

Minors and cofactors

Definition 1. *Let $A = (a_{i,j}) \in M_n(K)$ and let $i, j \in \{1, \ldots, n\}$.*

1. *The* minor *of the entry $a_{i,j}$ is the determinant*

$$\Delta_{i,j}$$

of the matrix obtained from A by deleting the i-th row and the j-th column.

2. The cofactor *of the entry $a_{i,j}$ is the scalar*

$$(-1)^{i+j}\Delta_{i,j}.$$

Theorem 1 (Cofactor expansion along a column)**.** *Let $A = (a_{i,j}) \in M_n(K)$. Then for every $j \in \{1, \dots, n\}$,*

$$\det(A) = \sum_{i=1}^{n} a_{i,j}(-1)^{i+j}\Delta_{i,j}.$$

Proof. Let $C_1, \dots, C_n$ be the columns of A, and let

$$B = (e_1, \dots, e_n)$$

be the canonical basis of K^n.

Fix $j \in \{1, \dots, n\}$. Since

$$C_j = \sum_{i=1}^{n} a_{i,j} e_i,$$

multilinearity of the determinant with respect to the j-th column gives

$$\begin{aligned}
\det(A) &= \det_B(C_1, \dots, C_{j-1}, C_j, C_{j+1}, \dots, C_n) \\
&= \det_B\left(C_1, \dots, C_{j-1}, \sum_{i=1}^{n} a_{i,j} e_i, C_{j+1}, \dots, C_n\right) \\
&= \sum_{i=1}^{n} a_{i,j} \det_B(C_1, \dots, C_{j-1}, e_i, C_{j+1}, \dots, C_n).
\end{aligned}$$

For each $i \in \{1, \dots, n\}$, set

$$D_{i,j} := \det_B(C_1, \dots, C_{j-1}, e_i, C_{j+1}, \dots, C_n).$$

Then

$$\det(A) = \sum_{i=1}^{n} a_{i,j} D_{i,j}.$$

We now compute $D_{i,j}$. Since the j-th column is e_i, we have

$$D_{i,j} = \begin{vmatrix}
a_{1,1} & a_{1,2} & \cdots & a_{1,j-1} & 0 & a_{1,j+1} & \cdots & a_{1,n} \\
a_{2,1} & a_{2,2} & \cdots & a_{2,j-1} & 0 & a_{2,j+1} & \cdots & a_{2,n} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
a_{i-1,1} & a_{i-1,2} & \cdots & a_{i-1,j-1} & 0 & a_{i-1,j+1} & \cdots & a_{i-1,n} \\
a_{i,1} & a_{i,2} & \cdots & a_{i,j-1} & 1 & a_{i,j+1} & \cdots & a_{i,n} \\
a_{i+1,1} & a_{i+1,2} & \cdots & a_{i+1,j-1} & 0 & a_{i+1,j+1} & \cdots & a_{i+1,n} \\
\vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\
a_{n,1} & a_{n,2} & \cdots & a_{n,j-1} & 0 & a_{n,j+1} & \cdots & a_{n,n}
\end{vmatrix}.$$

Now move the j-th column to the last position by performing $n - j$ column swaps, and then move the i-th row to the last position by performing $n - i$ row swaps. Each swap changes the sign of the determinant, so the total factor is

$$(-1)^{n-j}(-1)^{n-i} = (-1)^{i+j}.$$

Thus

$$D_{i,j} = (-1)^{i+j} \begin{vmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1,n} & 0 \\ a_{2,1} & a_{2,2} & \cdots & a_{2,j-1} & a_{2,j+1} & \cdots & a_{2,n} & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ a_{i-1,1} & a_{i-1,2} & \cdots & a_{i-1,j-1} & a_{i-1,j+1} & \cdots & a_{i-1,n} & 0 \\ a_{i+1,1} & a_{i+1,2} & \cdots & a_{i+1,j-1} & a_{i+1,j+1} & \cdots & a_{i+1,n} & 0 \\ \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,j-1} & a_{n,j+1} & \cdots & a_{n,n} & 0 \\ a_{i,1} & a_{i,2} & \cdots & a_{i,j-1} & a_{i,j+1} & \cdots & a_{i,n} & 1 \end{vmatrix}.$$

Expanding along the last column, we obtain

$$D_{i,j} = (-1)^{i+j}\Delta_{i,j}.$$

Therefore

$$\det(A) = \sum_{i=1}^{n} a_{i,j}(-1)^{i+j}\Delta_{i,j}.$$

□

Expansion along a row

Theorem 2 (Cofactor expansion along a row)**.** *Let $A = (a_{i,j}) \in M_n(K)$. Then for every $i \in \{1, \ldots, n\}$,*

$$\det(A) = \sum_{j=1}^{n} a_{i,j}(-1)^{i+j}\Delta_{i,j}.$$

Proof. Apply the previous theorem to the transpose $A^\top$. Since

$$\det(A^\top) = \det(A),$$

the formula for expansion along a row follows from the formula for expansion along a column. □

Vandermonde determinant

Definition 2. *Let $\alpha_1, \ldots, \alpha_n \in K$. The determinant*

$$V(\alpha_1, \ldots, \alpha_n) := \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1^2 & \alpha_2^2 & \cdots & \alpha_n^2 \\ \vdots & \vdots & & \vdots \\ \alpha_1^{n-1} & \alpha_2^{n-1} & \cdots & \alpha_n^{n-1} \end{vmatrix}$$

is called the Vandermonde determinant*.*

Theorem 3. *For every $n \geq 2$ and every $\alpha_1, \ldots, \alpha_n \in K$,*

$$V(\alpha_1, \ldots, \alpha_n) = \prod_{1 \leq i < j \leq n} (\alpha_j - \alpha_i).$$

Proof. We argue by induction on n.

For $n = 2$,

$$V(\alpha_1, \alpha_2) = \begin{vmatrix} 1 & 1 \\ \alpha_1 & \alpha_2 \end{vmatrix} = \alpha_2 - \alpha_1,$$

so the formula holds.

Assume now that the formula is true for size $n-1$, and let us prove it for size n.

Starting from the Vandermonde matrix, for each $i = n, n-1, \dots, 2$, replace the row L_i by

$$L_i - \alpha_1 L_{i-1}.$$

These operations do not change the determinant. We obtain

$$V(\alpha_1, \dots, \alpha_n) = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ 0 & \alpha_2 - \alpha_1 & \cdots & \alpha_n - \alpha_1 \\ 0 & \alpha_2(\alpha_2 - \alpha_1) & \cdots & \alpha_n(\alpha_n - \alpha_1) \\ \vdots & \vdots & & \vdots \\ 0 & \alpha_2^{n-2}(\alpha_2 - \alpha_1) & \cdots & \alpha_n^{n-2}(\alpha_n - \alpha_1) \end{vmatrix}.$$

Now expand along the first column. Since the only nonzero entry in that column is the first one, equal to 1, we get

$$V(\alpha_1, \dots, \alpha_n) = \begin{vmatrix} \alpha_2 - \alpha_1 & \alpha_3 - \alpha_1 & \cdots & \alpha_n - \alpha_1 \\ \alpha_2(\alpha_2 - \alpha_1) & \alpha_3(\alpha_3 - \alpha_1) & \cdots & \alpha_n(\alpha_n - \alpha_1) \\ \vdots & \vdots & & \vdots \\ \alpha_2^{n-2}(\alpha_2 - \alpha_1) & \alpha_3^{n-2}(\alpha_3 - \alpha_1) & \cdots & \alpha_n^{n-2}(\alpha_n - \alpha_1) \end{vmatrix}.$$

Factor $\alpha_k - \alpha_1$ from the $(k-1)$-st column, for each $k = 2, \dots, n$. This gives

$$V(\alpha_1, \dots, \alpha_n) = \left(\prod_{k=2}^{n} (\alpha_k - \alpha_1) \right) \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \alpha_2 & \alpha_3 & \cdots & \alpha_n \\ \vdots & \vdots & & \vdots \\ \alpha_2^{n-2} & \alpha_3^{n-2} & \cdots & \alpha_n^{n-2} \end{vmatrix}.$$

The remaining determinant is exactly

$$V(\alpha_2, \dots, \alpha_n).$$

Hence

$$V(\alpha_1, \dots, \alpha_n) = \left(\prod_{k=2}^{n} (\alpha_k - \alpha_1) \right) V(\alpha_2, \dots, \alpha_n).$$

By the induction hypothesis,

$$V(\alpha_2, \dots, \alpha_n) = \prod_{2 \le i < j \le n} (\alpha_j - \alpha_i).$$

Therefore

$$V(\alpha_1, \dots, \alpha_n) = \left(\prod_{k=2}^{n} (\alpha_k - \alpha_1) \right) \left(\prod_{2 \le i < j \le n} (\alpha_j - \alpha_i) \right).$$

This is precisely

$$\prod_{1 \le i < j \le n} (\alpha_j - \alpha_i).$$

The proof is complete. $\square$

Corollary 1. *For $\alpha_1, \ldots, \alpha_n \in K$, one has*

$$V(\alpha_1, \ldots, \alpha_n) \neq 0 \quad \Longleftrightarrow \quad \alpha_1, \ldots, \alpha_n \text{ are pairwise distinct.}$$

In-class Exercise

Let $z_0, \ldots, z_n$ be pairwise distinct complex numbers. Show that the family

$$\big((X - z_0)^n, \ldots, (X - z_n)^n\big)$$

is a basis of $\mathbb{C}_n[X]$, the vector space of complex polynomials of degree at most n.

Solution

The space $\mathbb{C}_n[X]$ has dimension $n + 1$, and the family

$$\big((X - z_0)^n, \ldots, (X - z_n)^n\big)$$

contains exactly $n + 1$ polynomials. Therefore, it is enough to prove that this family is linearly independent.

Consider the standard basis

$$(1, X, \ldots, X^n)$$

of $\mathbb{C}_n[X]$. For each $i \in \{0, \ldots, n\}$, the binomial formula gives

$$(X - z_i)^n = \sum_{j=0}^{n} \binom{n}{j} X^j (-z_i)^{n-j}.$$

Thus, the coordinate column of $(X - z_i)^n$ in the basis $(1, X, \ldots, X^n)$ is

$$\begin{pmatrix} (-z_i)^n \\ \binom{n}{1}(-z_i)^{n-1} \\ \vdots \\ \binom{n}{n-1}(-z_i) \\ 1 \end{pmatrix}.$$

Hence the determinant of the family in the basis $(1, X, \ldots, X^n)$ is

$$\det \begin{pmatrix} (-z_0)^n & (-z_1)^n & \cdots & (-z_n)^n \\ \binom{n}{1}(-z_0)^{n-1} & \binom{n}{1}(-z_1)^{n-1} & \cdots & \binom{n}{1}(-z_n)^{n-1} \\ \vdots & \vdots & & \vdots \\ \binom{n}{n-1}(-z_0) & \binom{n}{n-1}(-z_1) & \cdots & \binom{n}{n}(-z_n) \\ 1 & 1 & \cdots & 1 \end{pmatrix}.$$

Factoring out the binomial coefficients from the rows, we obtain

$$\binom{n}{1}\binom{n}{2}\cdots\binom{n}{n-1} \det \begin{pmatrix} (-z_0)^n & (-z_1)^n & \cdots & (-z_n)^n \\ (-z_0)^{n-1} & (-z_1)^{n-1} & \cdots & (-z_n)^{n-1} \\ \vdots & \vdots & & \vdots \\ (-z_0) & (-z_1) & \cdots & (-z_n) \\ 1 & 1 & \cdots & 1 \end{pmatrix}.$$

Reversing the order of the rows, this is a Vandermonde determinant in the numbers

$$-z_0, \dots, -z_n.$$

Since $z_0, \dots, z_n$ are pairwise distinct, the numbers $-z_0, \dots, -z_n$ are also pairwise distinct, so this Vandermonde determinant is nonzero.

Therefore, the family

$$\big((X - z_0)^n, \dots, (X - z_n)^n\big)$$

is linearly independent. Since it has $n + 1$ vectors in an $(n + 1)$-dimensional space, it is a basis of $\mathbb{C}_n[X]$.

Invariant Subspaces and Eigenvalues

Invariant subspaces

Suppose $f \in \mathcal{L}(V)$. If $m \geq 2$ and
$$V = V_1 \oplus \cdots \oplus V_m,$$
where each V_k is a nonzero subspace of V, then in order to understand the behavior of f, it is natural to try to understand the behavior of each restriction
$$f|_{V_k}.$$
Here $f|_{V_k}$ denotes the restriction of f to the smaller domain V_k. Since V_k is a smaller vector space than V, studying $f|_{V_k}$ should be easier than studying f.

However, if we want to use tools that are useful in the study of endomorphisms, then we face a problem: the map $f|_{V_k}$ may fail to map V_k into itself. In other words, $f|_{V_k}$ may fail to be an endomorphism of V_k. This leads us to consider subspaces of V that are mapped into themselves by f.

Definition 1. *Let $f \in \mathcal{L}(V)$. A subspace $U \subseteq V$ is said to be* invariant under f *if*
$$f(u) \in U \qquad \text{for every } u \in U.$$

Equivalently, a subspace $U \subseteq V$ is invariant under f if and only if the restriction
$$f|_U : U \to U$$
is an endomorphism of U.

Example 1. *Let $f \in \mathcal{L}(\mathcal{P}(\mathbb{R}))$ be defined by*
$$f(p) = p'.$$
Then $\mathcal{P}_4(\mathbb{R})$ is invariant under f, because if $p \in \mathcal{P}_4(\mathbb{R})$, then p' also has degree at most 4.

Example 2. *Let $f \in \mathcal{L}(V)$. Then the following subspaces of V are invariant under f:*
$$\{0\}, \qquad V, \qquad \ker(f), \qquad \operatorname{Im}(f).$$

Indeed:

- *$\{0\}$ is invariant under f because $f(0) = 0 \in \{0\}$.*
- *V is invariant under f because $f(v) \in V$ for every $v \in V$.*
- *$\ker(f)$ is invariant under f because if $u \in \ker(f)$, then $f(u) = 0$, hence $f(u) \in \ker(f)$.*
- *$\operatorname{Im}(f)$ is invariant under f because if $u \in \operatorname{Im}(f)$, then $u = f(v)$ for some $v \in V$, and therefore*
$$f(u) = f(f(v)) \in \operatorname{Im}(f).$$

A natural question is whether an endomorphism $f \in \mathcal{L}(V)$ must admit invariant subspaces other than $\{0\}$ and V. The previous example shows that $\ker(f)$ and $\operatorname{Im}(f)$ are invariant under f, but these do not always provide nontrivial invariant subspaces: it may happen that

$$\ker(f) = \{0\} \qquad \text{and} \qquad \operatorname{Im}(f) = V,$$

for instance when f is invertible.

We now turn to the simplest possible nontrivial invariant subspaces, namely those of dimension 1.

Take $v \in V$ with $v \neq 0$, and let

$$U = \{\lambda v : \lambda \in \mathbb{F}\} = \operatorname{span}(v).$$

Then U is a one-dimensional subspace of V, and every one-dimensional subspace of V is of this form for a suitable choice of v.

If U is invariant under f, then $f(v) \in U$, hence there exists $\lambda \in \mathbb{F}$ such that

$$f(v) = \lambda v.$$

Conversely, if

$$f(v) = \lambda v$$

for some $\lambda \in \mathbb{F}$, then $\operatorname{span}(v)$ is invariant under f.

This equation is important enough to deserve special terminology.

Eigenvalues

Definition 2. *Let $f \in \mathcal{L}(V)$. A scalar $\lambda \in \mathbb{F}$ is called an* eigenvalue *of f if there exists $v \in V$ such that*

$$v \neq 0 \qquad \textit{and} \qquad f(v) = \lambda v.$$

The condition $v \neq 0$ is essential, because every scalar $\lambda \in \mathbb{F}$ satisfies

$$f(0) = \lambda 0.$$

Thus V has a one-dimensional invariant subspace under f if and only if f has an eigenvalue.

The following result gives several equivalent ways of recognizing an eigenvalue in the finite-dimensional setting.

Proposition 1. *Suppose V is finite-dimensional, let $f \in \mathcal{L}(V)$, and let $\lambda \in \mathbb{F}$. Then the following are equivalent:*

1. *λ is an eigenvalue of f.*

2. *$f - \lambda \operatorname{Id}_V$ is not injective.*

3. *$f - \lambda \operatorname{Id}_V$ is not surjective.*

4. *$f - \lambda \operatorname{Id}_V$ is not invertible.*

Proof. By definition, λ is an eigenvalue of f if and only if there exists $v \in V$, $v \neq 0$, such that

$$f(v) = \lambda v.$$

This is equivalent to

$$(f - \lambda \operatorname{Id}_V)(v) = 0$$

for some nonzero vector $v \in V$. Thus λ is an eigenvalue of f if and only if

$$\ker(f - \lambda \operatorname{Id}_V) \neq \{0\},$$

which means exactly that $f - \lambda \operatorname{Id}_V$ is not injective.

Because V is finite-dimensional and $f - \lambda \operatorname{Id}_V$ is an endomorphism of V, injectivity, surjectivity, and invertibility are equivalent. Therefore (2), (3), and (4) are equivalent as well. $\square$

Definition 3. *Let $f \in \mathcal{L}(V)$, and let $\lambda \in \mathbb{F}$ be an eigenvalue of f. A vector $v \in V$ is called an* eigenvector *of f corresponding to λ if*

$$v \neq 0 \qquad \textit{and} \qquad f(v) = \lambda v.$$

In other words, a nonzero vector $v \in V$ is an eigenvector of an endomorphism $f \in \mathcal{L}(V)$ if and only if $f(v)$ is a scalar multiple of v. Moreover, because

$$f(v) = \lambda v \qquad \Longleftrightarrow \qquad (f - \lambda \operatorname{Id}_V)(v) = 0,$$

a vector $v \in V$ with $v \neq 0$ is an eigenvector of f corresponding to λ if and only if

$$v \in \ker(f - \lambda \operatorname{Id}_V).$$

Example 3. *Define $f \in \mathcal{L}(\mathbb{F}^2)$ by*

$$f(w, z) = (-z, w).$$

1. *First consider the case $\mathbb{F} = \mathbb{R}$. Then f is the counterclockwise rotation by 90° about the origin in $\mathbb{R}^2$. An endomorphism has an eigenvalue if and only if there exists a nonzero vector that is sent to a scalar multiple of itself. But a rotation by 90° sends every nonzero vector to a perpendicular vector, hence never to a scalar multiple of itself. Therefore, if $\mathbb{F} = \mathbb{R}$, then f has no eigenvalues, and consequently no eigenvectors.*

2. *Now consider the case $\mathbb{F} = \mathbb{C}$. To find the eigenvalues of f, we look for scalars $\lambda \in \mathbb{C}$ such that*

$$f(w, z) = \lambda(w, z)$$

has a solution other than $(w, z) = (0, 0)$.

Since

$$f(w, z) = (-z, w),$$

the equation $f(w, z) = \lambda(w, z)$ is equivalent to the system

$$-z = \lambda w, \qquad w = \lambda z.$$

Substituting the second equation into the first gives

$$-z = \lambda^2 z.$$

Now $z \neq 0$, because if $z = 0$, then the second equation implies $w = 0$, contradicting the fact that $(w, z) \neq (0, 0)$. Therefore we may divide by z, obtaining

$$-1 = \lambda^2.$$

Hence

$$\lambda = i \qquad \text{or} \qquad \lambda = -i.$$

Thus the eigenvalues of f are i and $-i$.

If $\lambda = i$, then from

$$w = \lambda z = iz$$

we get

$$z = -iw.$$

Hence the eigenvectors corresponding to i are exactly the vectors of the form

$$(w, -iw), \qquad w \in \mathbb{C},\ w \neq 0.$$

If $\lambda = -i$, then from

$$w = \lambda z = -iz$$

we get

$$z = iw.$$

Hence the eigenvectors corresponding to $-i$ are exactly the vectors of the form

$$(w, iw), \qquad w \in \mathbb{C},\ w \neq 0.$$

Theorem 1. *Let $f \in \mathcal{L}(V)$. Then every list of eigenvectors of f corresponding to distinct eigenvalues of f is linearly independent.*

Proof. Suppose the statement is false. Then there exists a smallest positive integer m such that there exist eigenvectors

$$v_1, \ldots, v_m$$

of f, corresponding to distinct eigenvalues

$$\lambda_1, \ldots, \lambda_m,$$

and such that the list $v_1, \ldots, v_m$ is linearly dependent.

Because m is minimal, there exist scalars

$$a_1, \ldots, a_m \in \mathbb{F},$$

all nonzero, such that

$$a_1 v_1 + \cdots + a_m v_m = 0.$$

Apply the endomorphism $f - \lambda_m \operatorname{Id}_V$ to both sides. Using the fact that

$$f(v_j) = \lambda_j v_j \qquad \text{for each } j = 1, \ldots, m,$$

we obtain

$$a_1(\lambda_1 - \lambda_m)v_1 + \cdots + a_{m-1}(\lambda_{m-1} - \lambda_m)v_{m-1} = 0.$$

Since the eigenvalues $\lambda_1, \dots, \lambda_m$ are distinct, we have

$$\lambda_j - \lambda_m \neq 0 \qquad \text{for } j = 1, \dots, m-1.$$

Because each a_j is nonzero, all the coefficients in the last relation are nonzero. Hence

$$v_1, \dots, v_{m-1}$$

is a linearly dependent list of eigenvectors corresponding to distinct eigenvalues, contradicting the minimality of m.

This contradiction proves the theorem. □

Corollary 1. *Suppose V is finite-dimensional. Then every endomorphism of V has at most* $\dim V$ *distinct eigenvalues.*

Proof. Let $f \in \mathcal{L}(V)$, and suppose that

$$\lambda_1, \dots, \lambda_m$$

are distinct eigenvalues of f. For each $j \in \{1, \dots, m\}$, choose an eigenvector v_j corresponding to λ_j. By the previous theorem, the list

$$v_1, \dots, v_m$$

is linearly independent. Therefore

$$m \leq \dim V.$$

This proves the result. □

Definition 4. *Let $f \in \mathcal{L}(V)$, and let*

$$p(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_m X^m$$

be a polynomial in $\mathbb{F}[X]$. We define the endomorphism $p(f) \in \mathcal{L}(V)$ by

$$p(f) = a_0 \operatorname{Id}_V + a_1 f + a_2 f^2 + \cdots + a_m f^m.$$

We have already seen that $\ker(f)$ and $\operatorname{Im}(f)$ are invariant under f. The same remains true for the kernel and the image of any polynomial in f.

Proposition 2. *Let $f \in \mathcal{L}(V)$, and let $p \in \mathbb{F}[X]$. Then* $\ker(p(f))$ *and* $\operatorname{Im}(p(f))$ *are invariant under f.*

Proof. First let us prove that $\ker(p(f))$ is invariant under f.

Let $u \in \ker(p(f))$. Then

$$p(f)(u) = 0.$$

Because $p(f)$ is a polynomial in f, it commutes with f. Hence

$$p(f)(f(u)) = f(p(f)(u)) = f(0) = 0.$$

Therefore

$$f(u) \in \ker(p(f)).$$

This proves that $\ker(p(f))$ is invariant under f.

Now let us prove that $\operatorname{Im}(p(f))$ is invariant under f.

Let $u \in \mathrm{Im}(p(f))$. Then there exists $v \in V$ such that

$$u = p(f)(v).$$

Applying f, we obtain

$$f(u) = f(p(f)(v)) = p(f)(f(v)),$$

again because f and $p(f)$ commute. Hence

$$f(u) \in \mathrm{Im}(p(f)).$$

Therefore $\mathrm{Im}(p(f))$ is invariant under f. $\square$

Eigenvalues and the Minimal Polynomial

Existence of eigenvalues on complex vector spaces

Theorem 1. *Let V be a finite-dimensional nonzero complex vector space, and let $T \in \mathcal{L}(V)$. Then T has an eigenvalue.*

Proof. Let $n = \dim V > 0$, and choose $v \in V$ with $v \neq 0$. Consider the list of vectors

$$v,\ Tv,\ T^2v,\ \dots,\ T^nv.$$

This is a list of $n+1$ vectors in a vector space of dimension n, so it is linearly dependent. Therefore there exist scalars $a_0, \dots, a_n \in \mathbb{C}$, not all equal to 0, such that

$$a_0v + a_1Tv + \cdots + a_nT^nv = 0.$$

Equivalently, if we define the polynomial

$$p(z) = a_0 + a_1z + \cdots + a_nz^n,$$

then

$$p(T)v = 0.$$

Among all nonconstant polynomials $r \in \mathcal{P}(\mathbb{C})$ such that

$$r(T)v = 0,$$

choose one of smallest degree, and denote it by p. Since p is a nonconstant complex polynomial, the fundamental theorem of algebra implies that p has a root $\lambda \in \mathbb{C}$. Hence there exists a polynomial $q \in \mathcal{P}(\mathbb{C})$ such that

$$p(z) = (z - \lambda)q(z)$$

for every $z \in \mathbb{C}$.

Evaluating this identity at the operator T, we obtain

$$p(T) = (T - \lambda I)q(T).$$

Applying both sides to v, we get

$$0 = p(T)v = (T - \lambda I)(q(T)v).$$

We now claim that $q(T)v \neq 0$. Indeed, if $q(T)v = 0$, then q would also be a nonconstant polynomial satisfying

$$q(T)v = 0,$$

and since

$$\deg q = \deg p - 1 < \deg p,$$

this would contradict the minimality of the degree of p.

Therefore $q(T)v$ is a nonzero vector and satisfies

$$(T - \lambda I)(q(T)v) = 0,$$

that is,

$$T(q(T)v) = \lambda\, q(T)v.$$

Hence λ is an eigenvalue of T, with corresponding eigenvector $q(T)v$.

This proves that every operator on a finite-dimensional nonzero complex vector space has an eigenvalue. $\square$

Existence, uniqueness, and degree of the minimal polynomial

We now explain carefully the proof of the following result.

Theorem 2. *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then there exists a unique monic polynomial $p \in \mathcal{P}(\mathbf{F})$ of smallest degree such that*

$$p(T) = 0.$$

Furthermore,

$$\deg p \leq \dim V.$$

Proof of existence

We proceed by induction on $\dim V$.

If $\dim V = 0$, then $V = \{0\}$, and the identity operator I is the zero operator on V. Hence the constant polynomial

$$p(z) = 1$$

satisfies

$$p(T) = I = 0.$$

Thus the result holds in this case.

Now assume $\dim V > 0$, and assume inductively that the desired statement is true for all operators on all vector spaces of smaller dimension.

Choose $u \in V$ with $u \neq 0$. Consider the list

$$u,\ Tu,\ T^2u,\ \ldots,\ T^{\dim V}u.$$

This list has length $1 + \dim V$, so it is linearly dependent. By the linear dependence lemma, there exists a smallest positive integer $m \leq \dim V$ such that $T^m u$ is a linear combination of

$$u,\ Tu,\ \ldots,\ T^{m-1}u.$$

Thus there exist scalars $c_0, c_1, \ldots, c_{m-1} \in \mathbf{F}$ such that

$$c_0 u + c_1 Tu + \cdots + c_{m-1} T^{m-1} u + T^m u = 0. \tag{1}$$

Define the polynomial

$$q(z) = c_0 + c_1 z + \cdots + c_{m-1} z^{m-1} + z^m.$$

Then q is monic and, by (1),

$$q(T)u = 0.$$

Now let k be any nonnegative integer. Because $q(T)$ is a polynomial in T, it commutes with T. Therefore

$$q(T)(T^k u) = T^k(q(T)u) = T^k(0) = 0.$$

In particular,

$$u,\ Tu,\ \ldots,\ T^{m-1}u \in \ker q(T).$$

By the minimality of m, the vectors

$$u,\ Tu,\ \ldots,\ T^{m-1}u$$

are linearly independent. Hence $\ker q(T)$ contains m linearly independent vectors, so

$$\dim \ker q(T) \geq m.$$

By rank-nullity,

$$\dim \operatorname{range} q(T) = \dim V - \dim \ker q(T) \leq \dim V - m.$$

Next we observe that $\operatorname{range} q(T)$ is invariant under T. Indeed, if $y \in \operatorname{range} q(T)$, then $y = q(T)x$ for some $x \in V$, and thus

$$Ty = Tq(T)x = q(T)Tx \in \operatorname{range} q(T),$$

again because T commutes with $q(T)$.

Therefore the restriction

$$T|_{\operatorname{range} q(T)}$$

is a well-defined linear operator on the vector space $\operatorname{range} q(T)$, whose dimension is at most $\dim V - m$, hence strictly smaller than $\dim V$.

By the induction hypothesis, there exists a monic polynomial $s \in \mathcal{P}(\mathbf{F})$ such that

$$s\big(T|_{\operatorname{range} q(T)}\big) = 0$$

and

$$\deg s \leq \dim V - m.$$

We claim that the product sq annihilates T. Let $v \in V$. Then

$$(sq)(T)(v) = s(T)(q(T)v).$$

But $q(T)v \in \operatorname{range} q(T)$, and s annihilates the restriction of T to that subspace. Hence

$$s(T)(q(T)v) = 0.$$

Since this holds for every $v \in V$, we conclude that

$$(sq)(T) = 0.$$

Moreover, sq is monic because both s and q are monic, and

$$\deg(sq) = \deg s + \deg q \leq (\dim V - m) + m = \dim V.$$

Thus we have produced a monic polynomial of degree at most $\dim V$ that annihilates T. This proves the existence part.

Proof of uniqueness

Let $p \in \mathcal{P}(\mathbf{F})$ be a monic polynomial of smallest degree such that

$$p(T) = 0.$$

Suppose $r \in \mathcal{P}(\mathbf{F})$ is another monic polynomial of the same degree such that

$$r(T) = 0.$$

Then

$$(p - r)(T) = p(T) - r(T) = 0.$$

Because p and r are monic and have the same degree, their leading terms cancel. Therefore

$$\deg(p - r) < \deg p.$$

If $p - r \neq 0$, then we may divide $p - r$ by its leading coefficient to obtain a monic polynomial h such that

$$\deg h < \deg p$$

and still

$$h(T) = 0.$$

Indeed, if c is the leading coefficient of $p - r$, then

$$h = \frac{1}{c}(p - r),$$

so

$$h(T) = \frac{1}{c}(p - r)(T) = \frac{1}{c} \cdot 0 = 0.$$

But this contradicts the minimality of the degree of p.

Hence $p - r = 0$, and so

$$p = r.$$

This proves uniqueness.

Eigenvalues and the Minimal Polynomial (II)

Existence of eigenvalues on complex vector spaces

Theorem 1. *Let V be a finite-dimensional nonzero complex vector space, and let $T \in \mathcal{L}(V)$. Then T has an eigenvalue.*

Proof of uniqueness

Let $p \in \mathcal{P}(\mathbf{F})$ be a monic polynomial of smallest degree such that

$$p(T) = 0.$$

Suppose $r \in \mathcal{P}(\mathbf{F})$ is another monic polynomial of the same degree such that

$$r(T) = 0.$$

Then

$$(p - r)(T) = p(T) - r(T) = 0.$$

Because p and r are monic and have the same degree, their leading terms cancel. Therefore

$$\deg(p - r) < \deg p.$$

If $p - r \neq 0$, then we may divide $p - r$ by its leading coefficient to obtain a monic polynomial h such that

$$\deg h < \deg p$$

and still

$$h(T) = 0.$$

Indeed, if c is the leading coefficient of $p - r$, then

$$h = \frac{1}{c}(p - r),$$

so

$$h(T) = \frac{1}{c}(p - r)(T) = \frac{1}{c} \cdot 0 = 0.$$

But this contradicts the minimality of the degree of p.

Hence $p - r = 0$, and so

$$p = r.$$

This proves uniqueness.

The minimal polynomial

The previous result justifies the following definition.

Definition 1 (Minimal polynomial)**.** *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then the **minimal polynomial** of T is the unique monic polynomial $p \in \mathcal{P}(\mathbf{F})$ of smallest degree such that*

$$p(T) = 0.$$

We now consider an example, which will also serve as motivation for a homework exercise.

Example 1 (Minimal polynomial of an operator on $\mathbf{F}^5$)**.** *Suppose $T \in \mathcal{L}(\mathbf{F}^5)$ and*

$$\mathcal{M}(T) = \begin{pmatrix} 0 & 0 & 0 & 0 & -3 \\ 1 & 0 & 0 & 0 & 6 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

with respect to the standard basis e_1, e_2, e_3, e_4, e_5.

Taking $v = e_1$, we compute

$$Te_1 = e_2, \quad T^2e_1 = Te_2 = e_3, \quad T^3e_1 = Te_3 = e_4,$$

$$T^4e_1 = Te_4 = e_5, \quad T^5e_1 = Te_5 = -3e_1 + 6e_2.$$

Thus

$$T^5e_1 = -3e_1 + 6Te_1,$$

which can be rewritten as

$$(3 - 6T + T^5)e_1 = 0.$$

Define the polynomial

$$p(z) = 3 - 6z + z^5.$$

Then the previous computation shows that

$$p(T)e_1 = 0.$$

We now observe that the list

$$e_1,\, Te_1,\, T^2e_1,\, T^3e_1,\, T^4e_1$$

coincides with

$$e_1,\, e_2,\, e_3,\, e_4,\, e_5,$$

which is a basis of $\mathbf{F}^5$. In particular, this list is linearly independent and spans the whole space.

Because $p(T)$ is a polynomial in T, it commutes with T. Hence for every nonnegative integer k we have

$$p(T)(T^ke_1) = T^k(p(T)e_1) = T^k(0) = 0.$$

Therefore $p(T)$ annihilates each vector in the list

$$e_1,\, Te_1,\, T^2e_1,\, T^3e_1,\, T^4e_1,$$

and since this list spans $\mathbf{F}^5$, we conclude that

$$p(T) = 0 \quad \text{on all of } \mathbf{F}^5.$$

It remains to show that p has the smallest possible degree. Suppose, for contradiction, that there exists a nonzero polynomial q with $\deg q < 5$ such that

$$q(T) = 0.$$

Then in particular

$$q(T)e_1 = 0.$$

Writing

$$q(z) = a_0 + a_1 z + \cdots + a_4 z^4,$$

we obtain

$$q(T)e_1 = a_0 e_1 + a_1 T e_1 + \cdots + a_4 T^4 e_1.$$

Using the computations above, this becomes

$$a_0 e_1 + a_1 e_2 + a_2 e_3 + a_3 e_4 + a_4 e_5 = 0.$$

Because e_1, e_2, e_3, e_4, e_5 *are linearly independent, it follows that*

$$a_0 = a_1 = a_2 = a_3 = a_4 = 0,$$

which contradicts the assumption that q *is nonzero.*

Hence no nontrivial polynomial of degree less than 5 *annihilates* T*. Therefore, the minimal polynomial of* T *is*

$$p(z) = 3 - 6z + z^5.$$

Proposition 1 (Eigenvalues are the zeros of the minimal polynomial)**.** *Suppose* V *is finite-dimensional and* $T \in \mathcal{L}(V)$*.*

(a) The zeros of the minimal polynomial of T *are precisely the eigenvalues of* T*.*

(b) If V *is a complex vector space, then the minimal polynomial of* T *has the form*

$$p(z) = (z - \lambda_1) \cdots (z - \lambda_m),$$

where $\lambda_1, \ldots, \lambda_m$ *is a list of all eigenvalues of* T*, possibly with repetitions.*

Proof. Let p be the minimal polynomial of T.

(a) First direction: zeros of p are eigenvalues.

Suppose $\lambda \in \mathbf{F}$ is a zero of p. Then p can be written as

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

where $q \in \mathcal{P}(\mathbf{F})$ is monic and

$$\deg q = \deg p - 1.$$

Evaluating at T, we obtain

$$p(T) = (T - \lambda I)q(T).$$

Because $p(T) = 0$, it follows that

$$(T - \lambda I)(q(T)v) = 0 \quad \text{for all } v \in V.$$

We now claim that there exists $v \in V$ such that $q(T)v \neq 0$. Indeed, if $q(T)v = 0$ for all $v \in V$, then $q(T) = 0$ as an operator. But this would contradict the minimality of p, because q is monic and

$$\deg q < \deg p.$$

Thus we can choose $v \in V$ such that $q(T)v \neq 0$. Set

$$w = q(T)v.$$

Then $w \neq 0$ and

$$(T - \lambda I)w = 0,$$

which means

$$Tw = \lambda w.$$

Hence λ is an eigenvalue of T.

Second direction: eigenvalues are zeros of p.

Now suppose $\lambda \in \mathbf{F}$ is an eigenvalue of T. Then there exists $v \in V$, $v \neq 0$, such that

$$Tv = \lambda v.$$

Applying T repeatedly, we obtain by induction that

$$T^k v = \lambda^k v \quad \text{for every nonnegative integer } k.$$

Let

$$p(z) = a_0 + a_1 z + \cdots + a_n z^n.$$

Then

$$p(T)v = a_0 v + a_1 Tv + \cdots + a_n T^n v = a_0 v + a_1 \lambda v + \cdots + a_n \lambda^n v = p(\lambda)v.$$

Because $p(T) = 0$, we have

$$p(T)v = 0.$$

Hence

$$p(\lambda)v = 0.$$

Since $v \neq 0$, it follows that

$$p(\lambda) = 0.$$

Thus λ is a zero of p.

This completes the proof of (a).

(b) If V is a complex vector space, then by the fundamental theorem of algebra, the polynomial p factors completely into linear terms:

$$p(z) = (z - \lambda_1) \cdots (z - \lambda_m).$$

By part (a), each λ_i is an eigenvalue of T, and every eigenvalue of T appears among the λ_i (possibly with repetition). $\square$

Example 2 (An operator whose eigenvalues cannot be found exactly)**.** *Let $T \in \mathcal{L}(\mathbb{C}^5)$ be defined by its action on the standard basis:*

$$T(e_1) = e_2, \quad T(e_2) = e_3, \quad T(e_3) = e_4, \quad T(e_4) = e_5, \quad T(e_5) = -3e_1 + 6e_2.$$

Let

$$z = (z_1, z_2, z_3, z_4, z_5) = z_1 e_1 + z_2 e_2 + z_3 e_3 + z_4 e_4 + z_5 e_5.$$

Then, by linearity,

$$T(z) = z_1T(e_1) + z_2T(e_2) + z_3T(e_3) + z_4T(e_4) + z_5T(e_5).$$

Substituting,

$$= z_1e_2 + z_2e_3 + z_3e_4 + z_4e_5 + z_5(-3e_1 + 6e_2).$$

Grouping terms, we obtain

$$T(z_1, z_2, z_3, z_4, z_5) = (-3z_5,\ z_1 + 6z_5,\ z_2,\ z_3,\ z_4).$$

As shown in the previous example, the minimal polynomial of T is

$$p(z) = 3 - 6z + z^5.$$

By Proposition above, the eigenvalues of T are exactly the zeros of p.

Proposition 2. *Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and $q \in \mathcal{P}(\mathbf{F})$. Then $q(T) = 0$ if and only if q is a polynomial multiple of the minimal polynomial of T.*

Proof. Let p be the minimal polynomial of T.

First suppose that $q(T) = 0$. By the division algorithm, there exist polynomials $s, r \in \mathcal{P}(\mathbf{F})$ such that

$$q = ps + r,$$

with $\deg r < \deg p$.

Evaluating at T, we obtain

$$0 = q(T) = p(T)s(T) + r(T) = r(T),$$

because $p(T) = 0$.

Thus $r(T) = 0$. If $r \neq 0$, let c be the leading coefficient of r and define

$$h = \frac{1}{c}r.$$

Then h is a monic polynomial, $\deg h = \deg r < \deg p$, and

$$h(T) = \frac{1}{c}r(T) = 0.$$

This contradicts the minimality of p, since p is the monic polynomial of smallest degree that annihilates T. Therefore $r = 0$, and hence

$$q = ps.$$

Conversely, suppose that q is a multiple of p, say $q = ps$ for some polynomial s. Then

$$q(T) = p(T)s(T) = 0,$$

because $p(T) = 0$.

This completes the proof. $\square$

Proposition 3. *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then T is not invertible if and only if the constant term of the minimal polynomial of T is 0.*

Proof. Let p be the minimal polynomial of T. Then

$$T \text{ is not invertible} \iff 0 \text{ is an eigenvalue of } T \iff 0 \text{ is a zero of } p \iff p(0) = 0.$$

The last condition means that the constant term of p is 0. $\square$

Proposition 4. *Suppose V is finite-dimensional, $T \in \mathcal{L}(V)$, and U is a subspace of V that is invariant under T. Then the minimal polynomial of T is a polynomial multiple of the minimal polynomial of $T|_U$.*

Proof. Let p be the minimal polynomial of T. Then $p(T) = 0$ on V. In particular,

$$p(T)u = 0 \quad \text{for all } u \in U.$$

Because U is invariant under T, it follows that

$$p(T|_U) = 0.$$

Let m be the minimal polynomial of $T|_U$. By definition, m is the monic polynomial of smallest degree such that $m(T|_U) = 0$. Since p is another polynomial that annihilates $T|_U$, it follows that m divides p. $\square$

Upper-triangular matrices and eigenvalues

A square matrix is called *upper triangular* if all entries below the diagonal are equal to 0. Equivalently, an $n \times n$ matrix is upper triangular if it has the form

$$\begin{pmatrix} \lambda_1 & * & * & \cdots & * \\ 0 & \lambda_2 & * & \cdots & * \\ 0 & 0 & \lambda_3 & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where the stars denote entries that may or may not be zero.

Remark. *If $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V, then the k-th column of $\mathcal{M}(T)$ is obtained by writing Tv_k as a linear combination of the basis vectors $v_1, \ldots, v_n$. Thus saying that $\mathcal{M}(T)$ is upper triangular means exactly that, for each k, the vector Tv_k involves only the first k basis vectors.*

Proposition 1. *Suppose $T \in \mathcal{L}(V)$ and $v_1, \ldots, v_n$ is a basis of V. Then the following conditions are equivalent:*

(a) The matrix of T with respect to $v_1, \ldots, v_n$ is upper triangular.

(b) $\operatorname{span}(v_1, \ldots, v_k)$ is invariant under T for each $k = 1, \ldots, n$.

(c) $Tv_k \in \operatorname{span}(v_1, \ldots, v_k)$ for each $k = 1, \ldots, n$.

Remark. *This proposition gives the geometric meaning of an upper-triangular matrix. Indeed, the matrix of T is upper triangular in the basis $v_1, \ldots, v_n$ if and only if we have a chain of invariant subspaces*

$$\operatorname{span}(v_1) \subseteq \operatorname{span}(v_1, v_2) \subseteq \cdots \subseteq \operatorname{span}(v_1, \ldots, v_n) = V.$$

In other words, upper-triangularity is equivalent to saying that the partial spans of the chosen basis are all preserved by T.

Example 1. *Define $T \in \mathcal{L}(\mathbf{F}^3)$ by*

$$T(x, y, z) = (2x + y,\ 5y + 3z,\ 8z).$$

Let e_1, e_2, e_3 denote the standard basis of $\mathbf{F}^3$. Then

$$Te_1 = 2e_1, \qquad Te_2 = e_1 + 5e_2, \qquad Te_3 = 3e_2 + 8e_3.$$

Thus

$$Te_1 \in \operatorname{span}(e_1), \qquad Te_2 \in \operatorname{span}(e_1, e_2), \qquad Te_3 \in \operatorname{span}(e_1, e_2, e_3).$$

Therefore the matrix of T with respect to the standard basis is upper triangular. Indeed,

$$\mathcal{M}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}.$$

This also shows that

$$\operatorname{span}(e_1) \subseteq \operatorname{span}(e_1, e_2) \subseteq \operatorname{span}(e_1, e_2, e_3)$$

is a chain of invariant subspaces.

Example 2 (A counterexample: rotation on $\mathbf{R}^2$)**.** *Define $T \in \mathcal{L}(\mathbf{R}^2)$ by*

$$T(x, y) = (-y, x).$$

This is rotation by $90°$. Let $e_1 = (1, 0)$ and $e_2 = (0, 1)$ be the standard basis of $\mathbf{R}^2$. Then

$$Te_1 = e_2.$$

But

$$e_2 \notin \operatorname{span}(e_1).$$

Therefore condition (c) above fails already for $k = 1$, so the matrix of T with respect to the standard basis is not upper triangular.

Indeed,

$$\mathcal{M}(T) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

which is not upper triangular.

This example shows concretely that upper-triangularity depends on the choice of basis: in the standard basis, the first basis vector is sent outside the span of the first basis vector.

We record two basic facts about operators that admit an upper-triangular matrix. The proofs below are written in a slightly more explicit way than in the text, with the goal of making the mechanism transparent for a first lecture.

Proposition 2. *Suppose $T \in \mathcal{L}(V)$ and V has a basis with respect to which T has an upper-triangular matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$. Then*

$$(T - \lambda_1 I) \cdots (T - \lambda_n I) = 0.$$

Proof. Let $v_1, \ldots, v_n$ be a basis of V with respect to which T has an upper-triangular matrix and diagonal entries $\lambda_1, \ldots, \lambda_n$. Thus, for each $k = 1, \ldots, n$,

$$Tv_k = \lambda_k v_k + \text{a linear combination of } v_1, \ldots, v_{k-1}.$$

Equivalently,

$$(T - \lambda_k I)v_k \in \operatorname{span}(v_1, \ldots, v_{k-1}).$$

We now show, by induction on k, that

$$(T - \lambda_1 I) \cdots (T - \lambda_k I)v_k = 0 \qquad \text{for each } k = 1, \ldots, n.$$

For $k = 1$, we have $Tv_1 = \lambda_1 v_1$, so

$$(T - \lambda_1 I)v_1 = 0.$$

Now suppose $k \geq 2$. As observed above,

$$(T - \lambda_k I)v_k \in \operatorname{span}(v_1, \ldots, v_{k-1}).$$

By the induction hypothesis, the operator

$$(T - \lambda_1 I) \cdots (T - \lambda_{k-1} I)$$

kills each of $v_1, \dots, v_{k-1}$, and hence kills every vector in $\operatorname{span}(v_1, \dots, v_{k-1})$. Therefore,

$$(T - \lambda_1 I) \cdots (T - \lambda_{k-1} I)(T - \lambda_k I)v_k = 0.$$

This proves the induction step.

Hence

$$(T - \lambda_1 I) \cdots (T - \lambda_n I)v_k = 0 \qquad \text{for every } k = 1, \dots, n.$$

Because $v_1, \dots, v_n$ is a basis of V, the operator

$$(T - \lambda_1 I) \cdots (T - \lambda_n I)$$

is 0 on a basis, and therefore is the 0 operator on all of V. □

Remark. *The key point is that each factor $(T - \lambda_k I)$ lowers one step in the flag*

$$\operatorname{span}(v_1) \subseteq \operatorname{span}(v_1, v_2) \subseteq \cdots \subseteq \operatorname{span}(v_1, \dots, v_n) = V.$$

Indeed,

$$(T - \lambda_k I)v_k \in \operatorname{span}(v_1, \dots, v_{k-1}),$$

so repeated application of the factors pushes basis vectors downward until they eventually become 0.

Example 3. *Define $T \in \mathcal{L}(\mathbf{F}^3)$ by*

$$T(x, y, z) = (2x + y,\ 5y + 3z,\ 8z).$$

With respect to the standard basis e_1, e_2, e_3 of $\mathbf{F}^3$, we have

$$\mathcal{M}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}.$$

Thus the proposition tells us that

$$(T - 2I)(T - 5I)(T - 8I) = 0.$$

Let us see this directly on the standard basis:

$$Te_1 = 2e_1, \qquad Te_2 = e_1 + 5e_2, \qquad Te_3 = 3e_2 + 8e_3.$$

Hence

$$(T - 2I)e_1 = 0,$$

$$(T - 5I)e_2 = e_1, \quad \text{so} \quad (T - 2I)(T - 5I)e_2 = (T - 2I)e_1 = 0,$$

and

$$(T - 8I)e_3 = 3e_2, \quad \text{so} \quad (T - 5I)(T - 8I)e_3 = 3e_1, \quad \text{and hence} \quad (T - 2I)(T - 5I)(T - 8I)e_3 = 0.$$

Thus the product kills each basis vector e_1, e_2, e_3, and so

$$(T - 2I)(T - 5I)(T - 8I) = 0$$

on all of $\mathbf{F}^3$.

Proposition 3. *Suppose $T \in \mathcal{L}(V)$ has an upper-triangular matrix with respect to some basis of V. Then the eigenvalues of T are precisely the entries on the diagonal of that upper-triangular matrix.*

Proof. Let $v_1, \dots, v_n$ be a basis of V with respect to which

$$\mathcal{M}(T) = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_2 & \cdots & * \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

We first show that each λ_k is an eigenvalue of T.

For $k = 1$, we have

$$Tv_1 = \lambda_1 v_1,$$

so λ_1 is an eigenvalue.

Now fix $k \in \{2, \dots, n\}$. Because the matrix is upper triangular,

$$Tv_k = \lambda_k v_k + \text{a linear combination of } v_1, \dots, v_{k-1},$$

and hence

$$(T - \lambda_k I)v_k \in \operatorname{span}(v_1, \dots, v_{k-1}).$$

Moreover, for each $j < k$,

$$(T - \lambda_k I)v_j \in \operatorname{span}(v_1, \dots, v_j) \subseteq \operatorname{span}(v_1, \dots, v_{k-1}).$$

Therefore $T - \lambda_k I$ maps $\operatorname{span}(v_1, \dots, v_k)$ into $\operatorname{span}(v_1, \dots, v_{k-1})$.

Now

$$\dim \operatorname{span}(v_1, \dots, v_k) = k \quad \text{and} \quad \dim \operatorname{span}(v_1, \dots, v_{k-1}) = k - 1.$$

Thus the restriction of $T - \lambda_k I$ to $\operatorname{span}(v_1, \dots, v_k)$ cannot be injective. Hence there exists a nonzero vector

$$v \in \operatorname{span}(v_1, \dots, v_k)$$

such that

$$(T - \lambda_k I)v = 0.$$

Therefore

$$Tv = \lambda_k v,$$

so λ_k is an eigenvalue of T.

We have proved that every diagonal entry $\lambda_1, \dots, \lambda_n$ is an eigenvalue of T.

Next we show that there are no other eigenvalues. Define

$$q(z) = (z - \lambda_1) \cdots (z - \lambda_n).$$

By the previous proposition,

$$q(T) = 0.$$

Hence the minimal polynomial of T divides q. Therefore every zero of the minimal polynomial of T is a zero of q. But the zeros of the minimal polynomial of T are exactly the eigenvalues of T. Consequently every eigenvalue of T is a zero of q, and thus must belong to the list

$$\lambda_1, \dots, \lambda_n.$$

So the eigenvalues of T are precisely the diagonal entries of the upper-triangular matrix. $\square$

Example 4. *For the operator*

$$T(x, y, z) = (2x + y,\ 5y + 3z,\ 8z),$$

the matrix in the standard basis is

$$\mathcal{M}(T) = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 5 & 3 \\ 0 & 0 & 8 \end{pmatrix}.$$

Hence the eigenvalues of T are exactly

$$2,\ 5,\ 8.$$

Notice that e_1 is an eigenvector with eigenvalue 2, because

$$Te_1 = 2e_1.$$

However, e_2 is not an eigenvector, since

$$Te_2 = e_1 + 5e_2.$$

Still, 5 is an eigenvalue because $T - 5I$ maps $\operatorname{span}(e_1, e_2)$ into $\operatorname{span}(e_1)$, so this restriction cannot be injective. Thus there exists some nonzero vector in $\operatorname{span}(e_1, e_2)$ that is an eigenvector with eigenvalue 5. Similarly, 8 is an eigenvalue because $T - 8I$ maps $\operatorname{span}(e_1, e_2, e_3)$ into $\operatorname{span}(e_1, e_2)$.

We now prove a fundamental characterization of when an operator admits an upper-triangular matrix.

Proposition 4. *Suppose V is finite-dimensional and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some basis of V if and only if the minimal polynomial of T has the form*

$$m_T(z) = (z - \lambda_1) \cdots (z - \lambda_m)$$

for some $\lambda_1, \ldots, \lambda_m \in \mathbf{F}$.

Corollary. *Suppose V is a finite-dimensional complex vector space and $T \in \mathcal{L}(V)$. Then T has an upper-triangular matrix with respect to some basis of V.*

Proof. The minimal polynomial of T is a polynomial with coefficients in $\mathbf{C}$. By the Fundamental Theorem of Algebra, every nonconstant polynomial over $\mathbf{C}$ factors completely into linear factors. Hence the minimal polynomial of T has the form

$$m_T(z) = (z - \lambda_1) \cdots (z - \lambda_m)$$

for some $\lambda_1, \ldots, \lambda_m \in \mathbf{C}$. The result now follows immediately from the proposition above. □

Remark. *This corollary is false over $\mathbf{R}$. The reason is that a real polynomial need not factor completely into linear factors over $\mathbf{R}$.*

Example 5 (A real operator that is not triangularizable over $\mathbf{R}$)**.** *Define $T \in \mathcal{L}(\mathbf{R}^2)$ by*

$$T(x, y) = (-y, x).$$

With respect to the standard basis of $\mathbf{R}^2$, the matrix of T is

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

which represents rotation by $90°$.

A direct computation shows that

$$T^2 = -I,$$

and hence

$$T^2 + I = 0.$$

Thus the polynomial

$$z^2 + 1$$

annihilates T. In fact, the minimal polynomial of T is

$$m_T(z) = z^2 + 1.$$

Over $\mathbf{R}$, this polynomial does not factor into linear factors, because it has no real roots.

Therefore, by the proposition above, T does not *have an upper-triangular matrix with respect to any basis of $\mathbf{R}^2$.*

However, over $\mathbf{C}$ we have

$$z^2 + 1 = (z - i)(z + i),$$

so if we regard the same operator as acting on $\mathbf{C}^2$, then it is *triangularizable over $\mathbf{C}$.*

Characteristic polynomial

Throughout this section, E denotes a finite-dimensional vector space over $\mathbf{F}$, with

$$\dim E = n.$$

We would like to associate to a square matrix, or to an endomorphism, a polynomial whose roots are exactly the eigenvalues.

If A is diagonal, or more generally upper triangular, with diagonal entries

$$\lambda_1, \dots, \lambda_n,$$

then the eigenvalues of A are precisely $\lambda_1, \dots, \lambda_n$. In that case it is natural to consider the polynomial

$$(\lambda_1 - X) \cdots (\lambda_n - X).$$

But this polynomial is exactly

$$\det(A - XI_n),$$

because $A - XI_n$ is again upper triangular, with diagonal entries

$$\lambda_1 - X, \dots, \lambda_n - X.$$

More generally, if λ is an eigenvalue of a square matrix A, then there exists a nonzero column vector v such that

$$Av = \lambda v.$$

Equivalently,

$$(A - \lambda I_n)v = 0.$$

Because $v \neq 0$, the matrix $A - \lambda I_n$ is not invertible, and hence

$$\det(A - \lambda I_n) = 0.$$

This shows that every eigenvalue of A is a root of the function

$$\lambda \longmapsto \det(A - \lambda I_n).$$

Definition 1. *Let $A \in M_n(\mathbf{F})$. The* characteristic polynomial *of A, denoted by $\chi_A(X)$, is defined by*

$$\chi_A(X) := \det(A - XI_n).$$

Remark. *Some authors define the characteristic polynomial as* $\det(XI_n - A)$ *instead. With that convention the characteristic polynomial is monic. The two definitions differ only by the factor $(-1)^n$, because*

$$\det(A - XI_n) = (-1)^n \det(XI_n - A).$$

The convention

$$\chi_A(X) = \det(A - XI_n)$$

has the advantage that

$$\chi_A(0) = \det(A).$$

Example 1. *Let*

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbf{F}).$$

Then

$$\chi_A(X) = \det \begin{pmatrix} a-X & b \\ c & d-X \end{pmatrix} = (a-X)(d-X) - bc.$$

Therefore

$$\chi_A(X) = X^2 - (a+d)X + (ad - bc)$$

and hence

$$\chi_A(X) = X^2 - \mathrm{tr}(A)X + \det(A).$$

Example 2. *Consider*

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

Because A is diagonal, its characteristic polynomial is

$$\chi_A(X) = \det(A - XI_n) = (2-X)^2(3-X).$$

Now let

$$p(X) = a_0 + a_1X + \cdots + a_mX^m$$

be any polynomial. Then

$$p(A) = a_0I + a_1A + \cdots + a_mA^m.$$

Since A is diagonal, every power A^k is diagonal, namely

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

Hence

$$p(A) = \begin{pmatrix} p(2) & 0 & 0 \\ 0 & p(2) & 0 \\ 0 & 0 & p(3) \end{pmatrix}.$$

Therefore

$$p(A) = 0 \quad \Longleftrightarrow \quad p(2) = 0 \text{ and } p(3) = 0.$$

So a polynomial annihilates A if and only if it has both 2 and 3 as roots. It follows that the monic polynomial of smallest degree annihilating A is

$$m_A(X) = (X-2)(X-3).$$

Thus

$$\chi_A(X) = -(X-2)^2(X-3), \qquad m_A(X) = (X-2)(X-3).$$

So the characteristic polynomial and the minimal polynomial need not be equal.

In this example they have the same roots, namely 2 and 3, but not with the same multiplicities.

Proposition 1. *For every $A \in M_n(\mathbf{F})$, the polynomial $\chi_A(X)$ has degree n. More precisely,*

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

In particular, the constant term of χ_A is $\det(A)$*.*

Proof. Write $A = (a_{i,j})$. Then

$$A - XI_n = (a_{i,j} - X\delta_{i,j})_{1 \leq i,j \leq n},$$

where $\delta_{i,j}$ denotes the Kronecker delta.

Using the permutation formula for the determinant, we obtain

$$\chi_A(X) = \det(A - XI_n) = \sum_{\sigma \in S_n} \varepsilon(\sigma) \prod_{j=1}^{n} (a_{\sigma(j),j} - X\delta_{\sigma(j),j}).$$

So $\chi_A(X)$ is a polynomial in X of degree at most n.

Now suppose $\sigma \neq \mathrm{id}$. Then there exist at least two indices j such that $\sigma(j) \neq j$. For those indices we have

$$\delta_{\sigma(j),j} = 0,$$

so in the corresponding product

$$\prod_{j=1}^{n} (a_{\sigma(j),j} - X\delta_{\sigma(j),j})$$

at least two factors do not involve X. Hence every such term has degree at most $n-2$.

Therefore the terms of degree n and $n-1$ can only come from the identity permutation. For $\sigma = \mathrm{id}$, the corresponding product is

$$(a_{1,1} - X) \cdots (a_{n,n} - X).$$

Its term of degree n is

$$(-X)^n = (-1)^n X^n,$$

and its term of degree $n-1$ is

$$(-1)^{n-1}(a_{1,1} + \cdots + a_{n,n}) X^{n-1} = (-1)^{n-1} \operatorname{tr}(A) X^{n-1}.$$

Finally, evaluating at $X = 0$ gives

$$\chi_A(0) = \det(A),$$

so the constant term is $\det(A)$. □

Example 3 (Rotation matrix)**.** *For*

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in M_2(\mathbf{R}),$$

we have

$$\chi_{R_\theta}(X) = \det \begin{pmatrix} \cos\theta - X & -\sin\theta \\ \sin\theta & \cos\theta - X \end{pmatrix}.$$

A direct computation gives

$$\chi_{R_\theta}(X) = X^2 - 2\cos(\theta) X + 1.$$

Proposition 2. *Similar matrices have the same characteristic polynomial. In other words, if*

$$A \in M_n(\mathbf{F}), \qquad P \in \mathrm{GL}_n(\mathbf{F}),$$

then

$$\chi_{PAP^{-1}} = \chi_A.$$

Consequently, A and PAP^{-1} have the same eigenvalues.

Proof. We compute

$$\chi_{PAP^{-1}}(X) = \det(PAP^{-1} - XI_n).$$

Because $XI_n = P(XI_n)P^{-1}$, this becomes

$$\det\big(P(A - XI_n)P^{-1}\big).$$

Using multiplicativity of the determinant,

$$\det\big(P(A - XI_n)P^{-1}\big) = \det(P)\det(A - XI_n)\det(P^{-1}) = \det(A - XI_n).$$

Therefore

$$\chi_{PAP^{-1}}(X) = \chi_A(X).$$

$\square$

Definition 2. *Let $f \in \mathcal{L}(E)$. The* characteristic polynomial *of f, denoted by $\chi_f(X)$, is defined as the characteristic polynomial of any matrix of f in a basis of E.*

Remark. *This definition is well-defined because two matrices representing the same endomorphism in two different bases are similar, and similar matrices have the same characteristic polynomial.*

Theorem 1. *Let $f \in \mathcal{L}(E)$ and let $\lambda \in \mathbf{F}$. Then*

$$\lambda \textit{ is an eigenvalue of } f \quad \Longleftrightarrow \quad \chi_f(\lambda) = 0.$$

Proof. By definition, λ is an eigenvalue of f if and only if

$$f - \lambda\,\mathrm{Id}_E$$

is not injective, equivalently, not invertible. Since E is finite-dimensional, this is equivalent to

$$\det(f - \lambda\,\mathrm{Id}_E) = 0.$$

By definition of the characteristic polynomial,

$$\det(f - \lambda\,\mathrm{Id}_E) = \chi_f(\lambda).$$

Therefore

$$\lambda \text{ is an eigenvalue of } f \quad \Longleftrightarrow \quad \chi_f(\lambda) = 0.$$

$\square$

Corollary. *An endomorphism of an n-dimensional vector space has at most n distinct eigenvalues.*

Proof. The eigenvalues of f are exactly the roots of the polynomial χ_f, and χ_f has degree n. $\square$

Corollary. *If E is a complex vector space of positive finite dimension and $f \in \mathcal{L}(E)$, then f has at least one eigenvalue.*

Proof. The characteristic polynomial χ_f is a nonconstant polynomial with complex coefficients, so by the Fundamental Theorem of Algebra it has a root in $\mathbf{C}$. By the previous theorem, that root is an eigenvalue of f. $\square$

Corollary. *If E is a real vector space of odd finite dimension and $f \in \mathcal{L}(E)$, then f has at least one real eigenvalue.*

Proof. The characteristic polynomial χ_f has odd degree, so it has a real root. By the theorem above, that real root is an eigenvalue of f. $\square$

Remark. *Suppose the characteristic polynomial of f splits over $\mathbf{F}$, and write*

$$\chi_f(X) = (-1)^n \prod_{\lambda \in \mathrm{Sp}(f)} (X - \lambda)^{m(\lambda)},$$

where the eigenvalues are counted with algebraic multiplicity.

Then

$$\mathrm{tr}(f) = \sum_{\lambda \in \mathrm{Sp}(f)} m(\lambda)\lambda, \qquad \det(f) = \prod_{\lambda \in \mathrm{Sp}(f)} \lambda^{m(\lambda)}.$$

Indeed, we already know that

$$\chi_f(X) = (-1)^n X^n + (-1)^{n-1}\,\mathrm{tr}(f) X^{n-1} + \cdots + \det(f).$$

On the other hand, if we expand the factorization

$$\chi_f(X) = (-1)^n \prod_{\lambda \in \mathrm{Sp}(f)} (X - \lambda)^{m(\lambda)},$$

then the coefficient of X^{n-1} is

$$(-1)^{n-1} \sum_{\lambda \in \mathrm{Sp}(f)} m(\lambda)\lambda,$$

because each eigenvalue λ appears exactly $m(\lambda)$ times among the roots of χ_f. Comparing the coefficient of X^{n-1} in the two expressions for χ_f, we obtain

$$\mathrm{tr}(f) = \sum_{\lambda \in \mathrm{Sp}(f)} m(\lambda)\lambda.$$

Similarly, the constant term of χ_f is $\det(f)$. But in the factorized expression, the constant term is obtained by multiplying all the roots together, counting multiplicities. Hence

$$\det(f) = \prod_{\lambda \in \mathrm{Sp}(f)} \lambda^{m(\lambda)}.$$

This does not contradict the usual definition of trace. The trace is always the sum of the diagonal entries of a matrix of f. The point is that, when the characteristic polynomial splits, the same number can also be recovered from the eigenvalues. Thus the trace may be computed in two different ways:

$$\mathrm{tr}(f) = \textit{sum of the diagonal entries}$$

and, when χ_f splits,

$$\operatorname{tr}(f) = \textit{sum of the eigenvalues counted with algebraic multiplicity.}$$

Likewise, the determinant can be computed either as the determinant of a matrix of f, or, when χ_f splits, as the product of the eigenvalues counted with algebraic multiplicity.

Proposition 3. *Let $f \in \mathcal{L}(E)$ and let F be a subspace of E which is invariant under f. Then the characteristic polynomial of the restriction $f_{|F}$ divides the characteristic polynomial of f.*

Proof. Let $\dim F = p$, and choose a basis

$$(e_1, \ldots, e_p)$$

of F, completed to a basis

$$(e_1, \ldots, e_p, e_{p+1}, \ldots, e_n)$$

of E.

Because F is invariant under f, the matrix of f in this basis has the block form

$$A = \begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix},$$

where A_1 is the matrix of $f_{|F}$ in the basis $(e_1, \ldots, e_p)$.

Then

$$A - XI_n = \begin{pmatrix} A_1 - XI_p & A_2 \\ 0 & A_3 - XI_{n-p} \end{pmatrix}.$$

Since this matrix is block upper triangular,

$$\chi_f(X) = \det(A - XI_n) = \det(A_1 - XI_p)\det(A_3 - XI_{n-p}).$$

But

$$\det(A_1 - XI_p) = \chi_{f_{|F}}(X).$$

Hence $\chi_{f_{|F}}$ divides χ_f. $\square$

Proposition 4. *Suppose*

$$E = E_1 \oplus \cdots \oplus E_r,$$

and suppose $f \in \mathcal{L}(E)$ leaves each E_i invariant. Let f_i denote the restriction of f to E_i. Then

$$\chi_f(X) = \chi_{f_1}(X) \cdots \chi_{f_r}(X).$$

Proof. Choose a basis of each E_i, and combine these bases into a basis of E adapted to the direct sum decomposition. In that basis, the matrix of f is block diagonal:

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r \end{pmatrix},$$

where A_i is the matrix of f_i.

Then

$$A - XI_n = \begin{pmatrix} A_1 - XI & 0 & \cdots & 0 \\ 0 & A_2 - XI & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_r - XI \end{pmatrix},$$

so

$$\chi_f(X) = \det(A - XI_n) = \prod_{i=1}^{r} \det(A_i - XI) = \prod_{i=1}^{r} \chi_{f_i}(X).$$

□

Example 4 (In-class exercise: companion matrix of a monic polynomial)**.** *Let*

$$P(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_{n-1} X^{n-1} + X^n$$

be a monic polynomial. Define f on a vector space with basis

$$(e_0, e_1, \ldots, e_{n-1})$$

by

$$f(e_i) = e_{i+1} \qquad (0 \le i \le n-2),$$

and

$$f(e_{n-1}) = -(a_0 e_0 + a_1 e_1 + \cdots + a_{n-1} e_{n-1}).$$

The matrix of f in this basis is called the companion matrix *of P and is denoted by C_P.*

Example 5 (In-class exercise: companion matrix of a monic polynomial)**.** *Let*

$$P(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_{n-1} X^{n-1} + X^n$$

be a monic polynomial. Define f on a vector space with basis

$$(e_0, e_1, \ldots, e_{n-1})$$

by

$$f(e_i) = e_{i+1} \qquad (0 \le i \le n-2),$$

and

$$f(e_{n-1}) = -(a_0 e_0 + a_1 e_1 + \cdots + a_{n-1} e_{n-1}).$$

The matrix of f in this basis is called the companion matrix *of P and is denoted by C_P.*

Exercise.

(1) Write down the matrix C_P explicitly.

(2) Prove that, with our convention

$$\chi_A(X) = \det(A - XI_n),$$

the characteristic polynomial of C_P is

$$\chi_{C_P}(X) = (-1)^n P(X).$$

(3) Prove that the minimal polynomial of C_P is

$$\mu_{C_P}(X) = P(X).$$

Diagonalizable Endomorphisms

1 Diagonalizable endomorphisms

Definition 1.1. *Let E be a finite-dimensional vector space over $\mathbf{K}$, and let $f \in \mathcal{L}(E)$.*

We say that f is diagonalizable *if there exists a basis $\mathcal{B}$ of E such that the matrix $\mathrm{Mat}_{\mathcal{B}}(f)$ is diagonal. Likewise, a matrix $A \in M_n(\mathbf{K})$ is said to be* diagonalizable *if there exists a diagonal matrix $D \in M_n(\mathbf{K})$ such that A is similar to D. In other words,*

$$\exists P \in GL_n(\mathbf{K}),\ \exists D \in M_n(\mathbf{K}) \text{ diagonal, such that } A = PDP^{-1}.$$

Theorem 1.2. *Let $f \in \mathcal{L}(E)$. The following assertions are equivalent:*

(1) f is diagonalizable;

(2) there exists a basis of E consisting of eigenvectors of f;

(3)
$$E = \sum_{\lambda \in \mathrm{Sp}(f)} E_\lambda(f);$$

(4)
$$\dim E = \sum_{\lambda \in \mathrm{Sp}(f)} \dim E_\lambda(f).$$

Proof. First observe that, since eigenspaces corresponding to distinct eigenvalues are in direct sum, assertion (3) is equivalent to

$$E = \bigoplus_{\lambda \in \mathrm{Sp}(f)} E_\lambda(f).$$

(1) $\Rightarrow$ (2). Assume that f is diagonalizable. Then there exists a basis

$$\mathcal{B} = (e_1, \dots, e_n)$$

of E such that

$$\mathrm{Mat}_{\mathcal{B}}(f) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

for some scalars $\lambda_1, \dots, \lambda_n \in \mathbf{K}$.

By definition of the matrix of a linear map in a basis, we have

$$f(e_i) = \lambda_i e_i, \qquad i = 1, \dots, n.$$

Since each $e_i \neq 0$, every e_i is an eigenvector of f. Thus $\mathcal{B}$ is a basis of E consisting of eigenvectors of f.

(2) $\Rightarrow$ (3). Assume that there exists a basis

$$\mathcal{B} = (e_1, \dots, e_n)$$

of E consisting of eigenvectors of f. For each i, there exists $\lambda_i \in \mathbf{K}$ such that

$$f(e_i) = \lambda_i e_i.$$

Hence

$$e_i \in E_{\lambda_i}(f) = \mathrm{Ker}(f - \lambda_i \mathrm{Id}).$$

Let $x \in E$. Since $\mathcal{B}$ is a basis, one can write

$$x = a_1 e_1 + \cdots + a_n e_n.$$

Each vector e_i belongs to some eigenspace of f, therefore x belongs to the sum of all eigenspaces. Thus

$$E \subseteq \sum_{\lambda \in \mathrm{Sp}(f)} E_\lambda(f).$$

The reverse inclusion is obvious, since each $E_\lambda(f)$ is a subspace of E. Therefore

$$E = \sum_{\lambda \in \mathrm{Sp}(f)} E_\lambda(f).$$

(3) $\Rightarrow$ (4). Assume that

$$E = \sum_{\lambda \in \mathrm{Sp}(f)} E_\lambda(f).$$

Since this sum is direct, we have

$$E = \bigoplus_{\lambda \in \mathrm{Sp}(f)} E_\lambda(f).$$

Hence

$$\dim(E) = \sum_{\lambda \in \mathrm{Sp}(f)} \dim(E_\lambda(f)).$$

(4) $\Rightarrow$ (1). Assume that

$$\dim(E) = \sum_{\lambda \in \mathrm{Sp}(f)} \dim(E_\lambda(f)).$$

Let

$$\mathrm{Sp}(f) = \{\mu_1, \dots, \mu_k\}$$

be the set of distinct eigenvalues of f. For each $j \in \{1, \dots, k\}$, let

$$d_j = \dim(E_{\mu_j}(f)),$$

and choose a basis

$$\mathcal{B}_j = (v_{j,1}, \dots, v_{j,d_j})$$

of the eigenspace $E_{\mu_j}(f)$.

Now consider the family

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

Since the eigenspaces $E_{\mu_1}(f), \dots, E_{\mu_k}(f)$ are in direct sum, and since each $\mathcal{B}_j$ is linearly independent in $E_{\mu_j}(f)$, it follows that $\mathcal{B}$ is linearly independent in E.

Moreover,

$$\mathrm{Card}(\mathcal{B}) = \sum_{j=1}^{k} \mathrm{Card}(\mathcal{B}_j) = \sum_{j=1}^{k} d_j = \sum_{j=1}^{k} \dim(E_{\mu_j}(f)) = \dim(E).$$

Hence $\mathcal{B}$ is a basis of E.

Now, for every $j \in \{1, \ldots, k\}$ and every $i \in \{1, \ldots, d_j\}$, since $v_{j,i} \in E_{\mu_j}(f)$, we have

$$f(v_{j,i}) = \mu_j v_{j,i}.$$

Therefore, in the basis $\mathcal{B}$, the matrix of f is diagonal, more precisely

$$\mathrm{Mat}_{\mathcal{B}}(f) = \begin{pmatrix} \mu_1 I_{d_1} & 0 & \cdots & 0 \\ 0 & \mu_2 I_{d_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_k I_{d_k} \end{pmatrix}.$$

Thus f is diagonalizable.

□

Theorem 1.3 (Necessary and sufficient condition for diagonalizability)**.** *Let E be a finite-dimensional vector space over $\mathbf{K}$, with $\dim(E) = n$, and let $f \in \mathcal{L}(E)$. Then f is diagonalizable if and only if:*

(a) the characteristic polynomial χ_f splits over $\mathbf{K}$;

(b) for every eigenvalue $\lambda \in \mathrm{Sp}(f)$,
$$\dim(E_\lambda(f)) = m(\lambda),$$
where $m(\lambda)$ denotes the algebraic multiplicity of λ as a root of χ_f.

Proof. For each eigenvalue $\lambda \in \mathrm{Sp}(f)$, set

$$d(\lambda) = \dim(E_\lambda(f)),$$

and let $m(\lambda)$ denote the multiplicity of λ in χ_f. Thus, when χ_f splits,

$$\chi_f(X) = \prod_{\lambda \in \mathrm{Sp}(f)} (\lambda - X)^{m(\lambda)}.$$

Suppose first that f is diagonalizable. Then, by the characterization of diagonalizable endomorphisms, we have

$$n = \sum_{\lambda \in \mathrm{Sp}(f)} d(\lambda). \tag{1}$$

Moreover, since f is diagonalizable, there exists a basis $\mathcal{B}$ of E such that

$$\mathrm{Mat}_{\mathcal{B}}(f) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

for some scalars $\lambda_1, \dots, \lambda_n \in \mathbf{K}$. Hence

$$\chi_f(X) = \det(\mathrm{Mat}_{\mathcal{B}}(f) - XI_n) = \prod_{i=1}^{n} (\lambda_i - X).$$

Therefore χ_f splits over $\mathbf{K}$. Grouping equal factors together, we may write

$$\chi_f(X) = \prod_{\lambda \in \mathrm{Sp}(f)} (\lambda - X)^{m(\lambda)},$$

and therefore

$$\sum_{\lambda \in \mathrm{Sp}(f)} m(\lambda) = n. \tag{2}$$

Now, for every eigenvalue λ, one always has

$$1 \leq d(\lambda) \leq m(\lambda).$$

Combining this with (1) and (2), we obtain

$$\sum_{\lambda \in \mathrm{Sp}(f)} d(\lambda) = \sum_{\lambda \in \mathrm{Sp}(f)} m(\lambda) = n.$$

Since each $d(\lambda) \leq m(\lambda)$, the equality of the sums forces

$$d(\lambda) = m(\lambda) \qquad \text{for every } \lambda \in \mathrm{Sp}(f).$$

Thus both (a) and (b) hold.

Conversely, assume that χ_f splits over $\mathbf{K}$ and that

$$d(\lambda) = m(\lambda) \qquad \text{for every } \lambda \in \mathrm{Sp}(f).$$

Since χ_f splits and has degree n, the sum of the algebraic multiplicities of its roots is

$$\sum_{\lambda \in \mathrm{Sp}(f)} m(\lambda) = n.$$

Using the assumption $d(\lambda) = m(\lambda)$, we get

$$\sum_{\lambda \in \mathrm{Sp}(f)} d(\lambda) = n.$$

That is,

$$\sum_{\lambda \in \mathrm{Sp}(f)} \dim(E_\lambda(f)) = \dim(E).$$

By the previous characterization of diagonalizable endomorphisms, this implies that f is diagonalizable. $\square$

Corollary 1.4 (A sufficient condition for diagonalizability)**.** *Let E be a finite-dimensional vector space over $\mathbf{K}$, with $\dim(E) = n$, and let $f \in \mathcal{L}(E)$. If f has n pairwise distinct eigenvalues, then f is diagonalizable.*

Proof. Let $\lambda_1, \dots, \lambda_n$ be n pairwise distinct eigenvalues of f. Since the characteristic polynomial χ_f has degree n, and since it already has n distinct roots in $\mathbf{K}$, it follows that

$$\chi_f(X) = \prod_{i=1}^{n} (\lambda_i - X).$$

In particular, χ_f splits over $\mathbf{K}$, and each eigenvalue λ_i has algebraic multiplicity

$$m(\lambda_i) = 1.$$

Now, for every eigenvalue λ_i, we always have

$$1 \leq \dim(E_{\lambda_i}(f)) \leq m(\lambda_i).$$

Since $m(\lambda_i) = 1$, this becomes

$$1 \leq \dim(E_{\lambda_i}(f)) \leq 1,$$

and therefore

$$\dim(E_{\lambda_i}(f)) = 1 \qquad \text{for every } i = 1, \dots, n.$$

Thus, for every eigenvalue, the geometric multiplicity is equal to the algebraic multiplicity. By the previous theorem, f is diagonalizable. $\square$

Corollary 1.5. *Let $A \in M_n(\mathbf{K})$. If A has n pairwise distinct eigenvalues, then A is diagonalizable.*

Example 1.6. *Let $n \in \mathbb{N}$, and consider the vector space $\mathbb{R}_n[X]$ of real polynomials of degree at most n. Define*

$$\varphi : \mathbb{R}_n[X] \to \mathbb{R}_n[X], \qquad \varphi(P) = P - (X+1)P'.$$

(1) Show that φ is a well-defined endomorphism of $\mathbb{R}_n[X]$.

(2) Determine the eigenvalues of φ.

(3) Deduce that φ is diagonalizable.

Diagonalizable Endomorphisms and Minimal Polynomial

Exercise 0.1 (In-class exercise: Fibonacci and diagonalization)**.** The Fibonacci sequence $F_0, F_1, F_2, \ldots$ is defined by

$$F_0 = 0, \qquad F_1 = 1, \qquad \text{and} \qquad F_n = F_{n-2} + F_{n-1} \quad \text{for } n \geq 2.$$

Define $T \in \mathcal{L}(\mathbb{R}^2)$ by

$$T(x, y) = (y, x + y).$$

(a) Show that

$$T^n(0, 1) = (F_n, F_{n+1})$$

for each nonnegative integer n.

(b) Find the eigenvalues of T.

(c) Find a basis of $\mathbb{R}^2$ consisting of eigenvectors of T.

(d) Use the solution to (c) to compute $T^n(0, 1)$. Conclude that

$$F_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1 + \sqrt{5}}{2} \right)^n - \left(\frac{1 - \sqrt{5}}{2} \right)^n \right]$$

for each nonnegative integer n.

0.1 Minimal polynomial and diagonalization

We now give a useful characterization of diagonalizable operators in terms of annihilating polynomials and the minimal polynomial.

Theorem 0.2. *Let E be a finite-dimensional vector space over a field $\mathbb{K}$, and let*

$$u \in \mathcal{L}(E).$$

Then the following assertions are equivalent:

1. *u is diagonalizable.*
2. *u admits an annihilating polynomial that splits into distinct linear factors.*
3. *The minimal polynomial μ_u of u splits into distinct linear factors.*

Notation and tools. We denote by

$$\mathrm{Sp}(u)$$

the set of eigenvalues of u. For each $\lambda \in \mathrm{Sp}(u)$, we denote by

$$E(\lambda) = \ker(u - \lambda \, \mathrm{Id}_E)$$

the eigenspace associated to λ.

Recall that if

$$\mathrm{Sp}(u) = \{\lambda_1, \ldots, \lambda_p\},$$

then u is diagonalizable if and only if

$$E = \bigoplus_{i=1}^{p} E(\lambda_i).$$

We shall also use the following lemma.

[Kernel decomposition lemma] Let $P_1, \dots, P_r \in \mathbb{K}[X]$ be pairwise coprime polynomials, and let

$$P = P_1 \cdots P_r.$$

Then

$$\ker(P(u)) = \bigoplus_{i=1}^{r} \ker(P_i(u)).$$

Proof. We prove

$$(1) \implies (2) \implies (3) \implies (1).$$

Step 1: $(1) \implies (2)$**.** Assume that u is diagonalizable. Write

$$\mathrm{Sp}(u) = \{\lambda_1, \dots, \lambda_p\}.$$

Since u is diagonalizable, we have

$$E = \bigoplus_{i=1}^{p} E(\lambda_i).$$

For each $i = 1, \dots, p$, define

$$P_i(X) = X - \lambda_i,$$

and set

$$P(X) = \prod_{i=1}^{p} P_i(X) = \prod_{i=1}^{p} (X - \lambda_i).$$

Then P splits into distinct linear factors.

Now let $x_i \in E(\lambda_i)$. Since x_i is an eigenvector corresponding to λ_i, we have

$$u(x_i) = \lambda_i x_i,$$

and therefore

$$P_i(u)(x_i) = u(x_i) - \lambda_i x_i = 0.$$

Hence

$$P(u)(x_i) = \Big(\prod_{j \neq i} P_j(u)\Big) \circ P_i(u)(x_i) = 0.$$

Now let $x \in E$. Since

$$E = \bigoplus_{i=1}^{p} E(\lambda_i),$$

there exist unique vectors

$$x_1 \in E(\lambda_1), \dots, x_p \in E(\lambda_p)$$

such that

$$x = x_1 + \cdots + x_p.$$

By linearity,

$$P(u)(x) = P(u)(x_1) + \cdots + P(u)(x_p) = 0.$$

Thus

$$P(u) = 0.$$

So P is an annihilating polynomial for u, and it splits into distinct linear factors.

Step 2: (2) $\Longrightarrow$ (3)**.** Assume that u admits an annihilating polynomial P that splits into distinct linear factors. Let μ_u be the minimal polynomial of u. Since the minimal polynomial divides every annihilating polynomial of u, we have

$$\mu_u \mid P.$$

Since P is a product of distinct linear factors, the same is true for μ_u. Therefore μ_u also splits into distinct linear factors.

Step 3: (3) $\Longrightarrow$ (1)**.** Assume that

$$\mu_u(X) = \prod_{i=1}^{r}(X - \lambda_i),$$

where the λ_i are pairwise distinct.

Apply the kernel decomposition lemma with

$$P_i(X) = X - \lambda_i.$$

Since the polynomials $P_1, \ldots, P_r$ are pairwise coprime, we obtain

$$\ker(\mu_u(u)) = \bigoplus_{i=1}^{r} \ker(P_i(u)) = \bigoplus_{i=1}^{r} \ker(u - \lambda_i \operatorname{Id}_E).$$

That is,

$$\ker(\mu_u(u)) = \bigoplus_{i=1}^{r} E(\lambda_i).$$

But by definition of the minimal polynomial,

$$\mu_u(u) = 0,$$

so

$$\ker(\mu_u(u)) = E.$$

Therefore,

$$E = \bigoplus_{i=1}^{r} E(\lambda_i).$$

Hence u is diagonalizable. $\square$

Remark 0.3. This theorem gives a very practical criterion for diagonalizability: an operator is diagonalizable if and only if its minimal polynomial has no repeated root.

Kernel decomposition lemma and Cayley-Hamilton theorem

Lemma 0.1 (Kernel decomposition lemma)**.** *Let E be a vector space, let $u \in \mathcal{L}(E)$, and let $P_1, \dots, P_r$ be polynomials which are pairwise coprime. Set*

$$P = P_1 \cdots P_r.$$

Then

$$\ker(P(u)) = \bigoplus_{i=1}^{r} \ker(P_i(u)).$$

We will assume for the proof the following results:

Theorem 0.1 (Bézout's theorem for polynomials)**.** *If P and Q are two coprime polynomials, then there exist polynomials U and V such that*

$$UP + VQ = 1.$$

Proposition 0.2 (Product of polynomials and polynomial endomorphisms)**.** If P and Q are two polynomials and $u \in \mathcal{L}(E)$, then

$$P(u) \circ Q(u) = Q(u) \circ P(u) = (PQ)(u).$$

Proposition 0.3 (Direct sum of subspaces)**.** If

$$F = F_1 \oplus F_2 \qquad \text{and} \qquad F_2 = G_1 \oplus G_2 \oplus \cdots \oplus G_r,$$

then

$$F = F_1 \oplus G_1 \oplus G_2 \oplus \cdots \oplus G_r.$$

Proof. We proceed by induction on r.

Fix a vector space E and an endomorphism $u \in \mathcal{L}(E)$. Let $\mathcal{P}(r)$ be the statement:

If $P_1, \dots, P_r$ are pairwise coprime polynomials and

$$P = P_1 \cdots P_r,$$

then

$$\ker(P(u)) = \bigoplus_{i=1}^{r} \ker(P_i(u)).$$

Base case: $r = 2$**.** Let P_1 and P_2 be coprime polynomials, and set

$$P = P_1 P_2.$$

We must prove that

$$\ker(P(u)) = \ker(P_1(u)) \oplus \ker(P_2(u)).$$

We will show:

1. $\ker(P_1(u)) + \ker(P_2(u)) \subset \ker(P(u))$;

2. $\ker(P(u)) \subset \ker(P_1(u)) + \ker(P_2(u))$;

3. $\ker(P_1(u))$ and $\ker(P_2(u))$ are in direct sum.

1. We prove that

$$\ker(P_1(u)) + \ker(P_2(u)) \subset \ker(P(u)).$$

Take

$$x \in \ker(P_1(u)) + \ker(P_2(u)).$$

Then $x = x_1 + x_2$ with

$$P_1(u)(x_1) = 0 \qquad \text{and} \qquad P_2(u)(x_2) = 0.$$

Let us compute $P(u)(x)$. We have

$$P(u)(x) = P(u)(x_1) + P(u)(x_2),$$

hence

$$P(u)(x) = (P_1P_2)(u)(x_1) + (P_1P_2)(u)(x_2).$$

Using the fact that

$$(P_1P_2)(u) = P_1(u) \circ P_2(u) = P_2(u) \circ P_1(u),$$

we get

$$P(u)(x) = P_2(u)\big(P_1(u)(x_1)\big) + P_1(u)\big(P_2(u)(x_2)\big) = 0.$$

Therefore

$$x \in \ker(P(u)).$$

2. We prove that

$$\ker(P(u)) \subset \ker(P_1(u)) + \ker(P_2(u)).$$

Since P_1 and P_2 are coprime, Bézout's theorem gives polynomials U and V such that

$$UP_1 + VP_2 = 1.$$

This equality in the polynomial algebra gives, after evaluation at u,

$$\mathrm{Id}_E = (UP_1)(u) + (VP_2)(u).$$

Now let $x \in \ker(P(u))$. Then

$$x = (UP_1)(u)(x) + (VP_2)(u)(x).$$

Set

$$y = (UP_1)(u)(x), \qquad z = (VP_2)(u)(x).$$

Then $x = y + z$.

We claim that

$$y \in \ker(P_2(u)) \qquad \text{and} \qquad z \in \ker(P_1(u)).$$

Indeed,

$$P_2(u)(y) = P_2(u)\big((UP_1)(u)(x)\big) = (P_2UP_1)(u)(x).$$

Since polynomial endomorphisms commute,

$$(P_2UP_1)(u) = U(u) \circ (P_1P_2)(u) = U(u) \circ P(u),$$

so

$$P_2(u)(y) = U(u)(P(u)(x)) = U(u)(0) = 0.$$

Thus $y \in \ker(P_2(u))$.

Similarly,

$$z \in \ker(P_1(u)).$$

Hence

$$x \in \ker(P_1(u)) + \ker(P_2(u)).$$

3. We prove that

$$\ker(P_1(u)) \cap \ker(P_2(u)) = \{0\}.$$

Take

$$x \in \ker(P_1(u)) \cap \ker(P_2(u)).$$

Using again Bézout's identity evaluated at u,

$$\mathrm{Id}_E = (UP_1)(u) + (VP_2)(u),$$

and applying this to x, we obtain

$$x = (UP_1)(u)(x) + (VP_2)(u)(x) = U(u)(P_1(u)(x)) + V(u)(P_2(u)(x)) = 0 + 0 = 0.$$

Therefore

$$\ker(P_1(u)) \cap \ker(P_2(u)) = \{0\},$$

so the sum is direct.

Combining steps (1), (2), and (3), we conclude that

$$\ker(P(u)) = \ker(P_1(u)) \oplus \ker(P_2(u)).$$

Induction step. Let $r \geq 2$, and assume that $\mathcal{P}(r)$ holds. We prove $\mathcal{P}(r+1)$.

Let $P_1, \ldots, P_{r+1}$ be pairwise coprime polynomials, and define

$$P = P_1 \cdots P_{r+1}, \qquad Q = P_2 \cdots P_{r+1},$$

so that

$$P = P_1 Q.$$

Since P_1 is coprime with each P_i for $i \geq 2$, it is coprime with Q. By the case $r = 2$, we have

$$\ker(P(u)) = \ker(P_1(u)) \oplus \ker(Q(u)).$$

By the induction hypothesis applied to $Q = P_2 \cdots P_{r+1}$,

$$\ker(Q(u)) = \bigoplus_{i=2}^{r+1} \ker(P_i(u)).$$

Therefore,

$$\ker(P(u)) = \ker(P_1(u)) \oplus \bigoplus_{i=2}^{r+1} \ker(P_i(u)) = \bigoplus_{i=1}^{r+1} \ker(P_i(u)).$$

This proves $\mathcal{P}(r+1)$, and hence the result for all $r \geq 2$. $\square$

The Cayley–Hamilton theorem

Theorem 0.4 (Cayley–Hamilton)**.** *Let E be a finite-dimensional vector space over a field $\mathbb{K}$, and let $f \in \mathcal{L}(E)$. Then the characteristic polynomial χ_f is an annihilating polynomial of f, that is,*

$$\chi_f(f) = 0_{\mathcal{L}(E)}.$$

Proof. We first show that for every vector $x \in E$,

$$\chi_f(f)(x) = 0_E.$$

This will imply that $\chi_f(f) = 0_{\mathcal{L}(E)}$.

If $x = 0_E$, the conclusion is immediate. Assume now that $x \neq 0_E$.

Consider the set

$$I = \left\{p \in \mathbb{N} \ : \ \big(x, f(x), \dots, f^p(x)\big) \text{ is linearly independent}\right\}.$$

Since $x \neq 0$, the set I is nonempty. Moreover, because $\dim E = n$, any family of more than n vectors in E is linearly dependent, so I is bounded above by $n-1$. Hence I has a greatest element, say s.

Define the subspace

$$E(x) = \operatorname{Span}\big(x, f(x), \dots, f^s(x)\big).$$

Then

$$B = \big(x, f(x), \dots, f^s(x)\big)$$

is a basis of $E(x)$. Since s is the greatest element of I, the family

$$\big(x, f(x), \dots, f^s(x), f^{s+1}(x)\big)$$

is linearly dependent. Therefore, there exist scalars $a_0, \dots, a_s \in \mathbb{K}$ such that

$$f^{s+1}(x) = \sum_{k=0}^{s} a_k f^k(x).$$

It follows that

$$f(E(x)) = \operatorname{Span}\big(f(x), f^2(x), \dots, f^{s+1}(x)\big) \subset E(x),$$

so the subspace $E(x)$ is stable under f.

Let

$$g = f_{|E(x)} \in \mathcal{L}(E(x)).$$

Since $E(x)$ is f-stable, the characteristic polynomial χ_g divides χ_f. Thus there exists a polynomial Q such that

$$\chi_f = Q\chi_g.$$

Hence

$$\chi_f(f) = Q(f) \circ \chi_g(f).$$

Therefore, it is enough to prove that

$$\chi_g(f)(x) = 0_E.$$

We now compute the matrix of g in the basis

$$B = \big(x, f(x), \dots, f^s(x)\big).$$

We have

$$g(x) = f(x), \qquad g(f(x)) = f^2(x), \qquad \dots, \qquad g(f^{s-1}(x)) = f^s(x),$$

and

$$g(f^s(x)) = f^{s+1}(x) = \sum_{k=0}^{s} a_k f^k(x).$$

Therefore, the matrix of g in the basis B is

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_0 \\ 1 & 0 & \cdots & 0 & a_1 \\ 0 & 1 & \cdots & 0 & a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & a_s \end{pmatrix}.$$

Thus

$$\chi_g(X) = \det(A - XI).$$

So

$$\chi_g(X) = \begin{vmatrix} -X & 0 & \cdots & 0 & a_0 \\ 1 & -X & \cdots & 0 & a_1 \\ 0 & 1 & \ddots & 0 & a_2 \\ \vdots & \vdots & \ddots & -X & \vdots \\ 0 & 0 & \cdots & 1 & a_s - X \end{vmatrix}.$$

Now replace the first row L_1 by

$$L_1 + XL_2 + X^2L_3 + \cdots + X^sL_{s+1}.$$

Then we obtain

$$\chi_g(X) = \begin{vmatrix} 0 & 0 & \cdots & 0 & R(X) \\ 1 & -X & \cdots & 0 & a_1 \\ 0 & 1 & \ddots & 0 & a_2 \\ \vdots & \vdots & \ddots & -X & \vdots \\ 0 & 0 & \cdots & 1 & a_s - X \end{vmatrix},$$

where

$$R(X) = -X^{s+1} + \sum_{k=0}^{s} a_k X^k.$$

Expanding along the first row gives

$$\chi_g(X) = (-1)^s R(X).$$

Therefore,

$$\chi_g(f)(x) = (-1)^s \left(\sum_{k=0}^{s} a_k f^k(x) - f^{s+1}(x) \right) = 0_E,$$

by the relation defining the coefficients $a_0, \dots, a_s$.

Hence

$$\chi_f(f)(x) = Q(f)\big(\chi_g(f)(x)\big) = Q(f)(0_E) = 0_E.$$

Since this holds for every $x \in E$, we conclude that

$$\chi_f(f) = 0_{\mathcal{L}(E)}.$$

$\square$

Corollary 0.5. *Let $f \in \mathcal{L}(E)$.*

1. *The minimal polynomial μ_f divides the characteristic polynomial χ_f.*
2. *The degree of the minimal polynomial is less than or equal to* $\dim E$.
3. *The minimal polynomial and the characteristic polynomial have the same roots, possibly with different multiplicities.*

Linear Algebra – Inner Products and Norms

1. Bilinear forms

Let E be a vector space over $\mathbb{R}$.

Definition 1. *A* bilinear form *on E is a map*

$$b : E \times E \to \mathbb{R}$$

such that:

- *for each fixed $y \in E$, the map $x \mapsto b(x, y)$ is linear;*
- *for each fixed $x \in E$, the map $y \mapsto b(x, y)$ is linear.*

Definition 2. *Let b be a bilinear form on E.*

- *We say that b is* symmetric *if*

$$b(x, y) = b(y, x) \qquad \text{for all } x, y \in E.$$

- *We say that b is* positive *if*

$$b(x, x) \geq 0 \qquad \text{for all } x \in E.$$

- *We say that b is* positive definite *if*

$$b(x, x) \geq 0 \quad \text{for all } x \in E,$$

  and

$$b(x, x) = 0 \iff x = 0.$$

2. Inner products

Definition 3. *Let E be a vector space over $\mathbb{R}$. An* inner product *on E is a map*

$$\langle \cdot, \cdot \rangle : E \times E \to \mathbb{R}$$

such that:

1. *for each fixed $y \in E$, the map $x \mapsto \langle x, y \rangle$ is linear;*
2. $$\langle x, y \rangle = \langle y, x \rangle \qquad \text{for all } x, y \in E;$$

3.

$$\langle x, x \rangle > 0 \qquad \text{for all } x \neq 0.$$

A vector space equipped with an inner product is called an inner product space.

Remark 1. *Over $\mathbb{R}$, an inner product is exactly a symmetric positive definite bilinear form.*

Remark 2. *If E is a complex vector space, then the definition is modified: one requires*

$$\langle x, y \rangle = \overline{\langle y, x \rangle},$$

together with linearity in one variable and conjugate-linearity in the other. In this lecture, we work over $\mathbb{R}$.

3. Examples of inner products

Example 1 (The standard inner product on $\mathbb{R}^n$)**.** *For $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$, define*

$$\langle x, y \rangle = x_1 y_1 + \cdots + x_n y_n.$$

This is an inner product on $\mathbb{R}^n$.

Example 2 (Weighted inner product on $\mathbb{R}^n$)**.** *Let $a_1, \ldots, a_n > 0$. Define*

$$\langle x, y \rangle = a_1 x_1 y_1 + \cdots + a_n x_n y_n.$$

Then this is also an inner product on $\mathbb{R}^n$.

Example 3 (The space $\ell^2(\mathbb{R})$)**.** *Let*

$$\ell^2(\mathbb{R}) = \left\{ x = (x_n)_{n \geq 1} : \sum_{n=1}^{\infty} x_n^2 < \infty \right\}.$$

For $x = (x_n)$ and $y = (y_n)$ in $\ell^2(\mathbb{R})$, define

$$\langle x, y \rangle = \sum_{n=1}^{\infty} x_n y_n.$$

This is an inner product on $\ell^2(\mathbb{R})$.

Remark 3. *The series $\sum_{n=1}^{\infty} x_n y_n$ is well-defined for $x, y \in \ell^2(\mathbb{R})$ by the Cauchy–Schwarz inequality for series.*

Example 4 (Continuous functions on the interval $[0, 1]$)**.** *Let $C([0, 1])$ be the vector space of continuous real-valued functions on $[0, 1]$. Define*

$$\langle f, g \rangle = \int_0^1 f(t) g(t) \, dt.$$

Then this is an inner product on $C([0, 1])$.

Proof. We only check positive definiteness. If

$$\langle f, f\rangle = \int_0^1 f(t)^2\, dt = 0,$$

then $f(t)^2 = 0$ for all $t \in [0,1]$, since f^2 is continuous and nonnegative. Hence $f = 0$. □

Example 5 (Inner product on matrices)**.** *Let $M_{n,p}(\mathbb{R})$ be the vector space of all $n \times p$ real matrices. Define*

$$\langle A, B\rangle = \mathrm{tr}(A^T B).$$

Then

$$\langle A, B\rangle = \sum_{i=1}^{n}\sum_{j=1}^{p} a_{ij} b_{ij},$$

so this is an inner product on $M_{n,p}(\mathbb{R})$.

Proof. We check the three properties:

- Linearity follows from the linearity of the trace.
- Symmetry follows from
$$\mathrm{tr}(A^T B) = \mathrm{tr}((A^T B)^T) = \mathrm{tr}(B^T A).$$
- Also,
$$\langle A, A\rangle = \mathrm{tr}(A^T A) = \sum_{i,j} a_{ij}^2 \geq 0,$$
and this is equal to 0 if and only if $A = 0$.

□

4. The Cauchy–Schwarz inequality

Theorem 1 (Cauchy–Schwarz)**.** *Let $(E, \langle\cdot,\cdot\rangle)$ be an inner product space. Then for all $x, y \in E$,*

$$|\langle x, y\rangle| \leq \sqrt{\langle x, x\rangle}\,\sqrt{\langle y, y\rangle}.$$

Equivalently,

$$|\langle x, y\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle.$$

Proof. If $y = 0$, then $\langle x, y\rangle = 0$, so the conclusion is immediate.

Assume now that $y \neq 0$. For every $\lambda \in \mathbb{R}$,

$$0 \leq \langle x - \lambda y, x - \lambda y\rangle.$$

Expanding,

$$0 \leq \langle x, x\rangle - 2\lambda\langle x, y\rangle + \lambda^2\langle y, y\rangle.$$

Choose

$$\lambda = \frac{\langle x, y\rangle}{\langle y, y\rangle}.$$

Substituting this value gives

$$0 \le \langle x, x\rangle - \frac{\langle x, y\rangle^2}{\langle y, y\rangle}.$$

Therefore,

$$\langle x, y\rangle^2 \le \langle x, x\rangle\langle y, y\rangle.$$

Since the left-hand side is nonnegative, this is equivalent to

$$|\langle x, y\rangle| \le \sqrt{\langle x, x\rangle}\,\sqrt{\langle y, y\rangle}.$$

□

Remark 4. *Equality holds in Cauchy–Schwarz if and only if x and y are linearly dependent.*

5. Norms

Definition 4. *Let E be a vector space over $\mathbb{R}$. A* norm *on E is a map*

$$\|\cdot\| : E \to [0, \infty)$$

such that for all $x, y \in E$ and all $\lambda \in \mathbb{R}$:

1. $$\|\lambda x\| = |\lambda|\,\|x\|;$$

2. $$\|x + y\| \le \|x\| + \|y\|;$$

3. $$\|x\| = 0 \iff x = 0.$$

A vector space equipped with a norm is called a normed vector space*.*

6. The norm induced by an inner product

Proposition 1. *Let $(E, \langle\cdot,\cdot\rangle)$ be an inner product space. Define*

$$\|x\| = \sqrt{\langle x, x\rangle} \qquad \text{for } x \in E.$$

Then $\|\cdot\|$ is a norm on E.

Proof. We check the three axioms.

(1) Homogeneity. For $\lambda \in \mathbb{R}$,

$$\|\lambda x\|^2 = \langle \lambda x, \lambda x\rangle = \lambda^2\langle x, x\rangle = |\lambda|^2\|x\|^2.$$

Since both sides are nonnegative,

$$\|\lambda x\| = |\lambda|\,\|x\|.$$

(2) Positive definiteness. By the definition of inner product,

$$\langle x, x\rangle \ge 0$$

for all x, and

$$\langle x, x\rangle = 0 \iff x = 0.$$

Hence

$$\|x\| = \sqrt{\langle x, x\rangle} \geq 0$$

and

$$\|x\| = 0 \iff x = 0.$$

(3) Triangle inequality. We compute

$$\|x+y\|^2 = \langle x+y, x+y\rangle = \langle x, x\rangle + 2\langle x, y\rangle + \langle y, y\rangle.$$

By Cauchy–Schwarz,

$$\langle x, y\rangle \leq |\langle x, y\rangle| \leq \|x\|\,\|y\|.$$

Therefore,

$$\|x+y\|^2 \leq \|x\|^2 + 2\|x\|\,\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2.$$

Taking square roots, we obtain

$$\|x+y\| \leq \|x\| + \|y\|.$$

$\square$

7. Final examples

Example 6. *On $\mathbb{R}^n$ with the standard inner product,*

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2},$$

which is the usual Euclidean norm.

Example 7. *On $\ell^2(\mathbb{R})$, the norm induced by the inner product is*

$$\|x\|_2 = \left(\sum_{n=1}^{\infty} x_n^2\right)^{1/2}.$$

Example 8. *On $C([0,1])$ with*

$$\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt,$$

the induced norm is

$$\|f\|_2 = \left(\int_0^1 |f(t)|^2\,dt\right)^{1/2}.$$

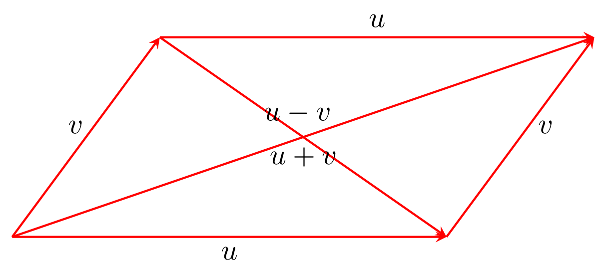
Introduction to Orthogonality

The next result is called the *parallelogram identity* because of its geometric interpretation: in every parallelogram, the sum of the squares of the lengths of the diagonals equals the sum of the squares of the lengths of the four sides.

If the sides of the parallelogram are represented by the vectors u and v, then its diagonals are $u+v$ and $u-v$. Therefore the geometric statement becomes

$$\|u+v\|^2 + \|u-v\|^2 = 2\big(\|u\|^2 + \|v\|^2\big).$$

The proof in an inner product space is more straightforward than the usual Euclidean geometry proof.



The diagonals of this parallelogram are $u+v$ and $u-v$.

Theorem 1 (Parallelogram identity)**.** *Suppose $u, v \in V$. Then*

$$\|u+v\|^2 + \|u-v\|^2 = 2\big(\|u\|^2 + \|v\|^2\big).$$

Proof. We have

$$\|u+v\|^2 + \|u-v\|^2 = \langle u+v, u+v\rangle + \langle u-v, u-v\rangle.$$

Expanding both inner products, we obtain

$$\langle u+v, u+v\rangle = \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \langle v, u\rangle,$$

and

$$\langle u-v, u-v\rangle = \|u\|^2 + \|v\|^2 - \langle u, v\rangle - \langle v, u\rangle.$$

Adding these two equalities, the cross terms cancel, so

$$\|u+v\|^2 + \|u-v\|^2 = 2\big(\|u\|^2 + \|v\|^2\big),$$

as desired. □

Inclass Exercise 1. ***Part A.*** *Let E be a real vector space equipped with a norm $\|\cdot\|$ satisfying the parallelogram identity recalled above. Let $\varphi : E \times E \to \mathbb{R}$ be defined by*

$$\varphi(x, y) = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right).$$

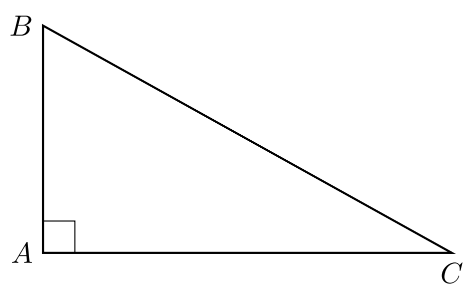
We want to show that φ is an inner product on E from which $\|\cdot\|$ arises. To do this, prove the following identities, where $x, y, z \in E$ and $\lambda \in \mathbb{R}$:

1. $\varphi(x, x) = \|x\|^2$, $\varphi(x, y) = \varphi(y, x)$, *and* $\varphi(x, 0) = 0$.
2. $2\varphi(x, z) + 2\varphi(y, z) = \varphi(x+y, 2z)$ *and* $2\varphi(x, z) = \varphi(x, 2z)$.
3. $\varphi(x+y, z) = \varphi(x, z) + \varphi(y, z)$.
4. $\varphi(\lambda x, z) = \lambda\varphi(x, z)$. *(One should begin with $\lambda \in \mathbb{N}$, then $\mathbb{Z}$, $\mathbb{Q}$, and $\mathbb{R}$.)*

Deduce that φ is an inner product.

In Euclidean geometry, the Pythagorean theorem says that if a triangle has a right angle, then the square of the length of the hypotenuse is the sum of the squares of the lengths of the two legs.

This suggests that, in an inner product space, perpendicularity should correspond to the vanishing of the cross term in the expansion of a squared norm.



If the angle at A is a right angle, then by the Pythagorean theorem,

$$\|BC\|^2 = \|AB\|^2 + \|AC\|^2.$$

On the other hand,

$$\overrightarrow{BC} = \overrightarrow{AC} - \overrightarrow{AB},$$

so

$$\|BC\|^2 = \|\overrightarrow{AC} - \overrightarrow{AB}\|^2 = \|AC\|^2 - 2\langle AC, AB\rangle + \|AB\|^2.$$

Comparing with the Pythagorean identity, we obtain

$$\langle AC, AB\rangle = 0.$$

This motivates the following definition.

Definition 1 (Orthogonality). *Let V be an inner product space. Two vectors $u, v \in V$ are said to be* orthogonal *if*

$$\langle u, v\rangle = 0.$$

In that case, we write

$$u \perp v.$$

Proposition 1 (Pythagorean identity)**.** *Let V be an inner product space, and let $u, v \in V$.*

If $u \perp v$, then

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

If V is a real inner product space, then the converse also holds:

$$u \perp v \quad \Longleftrightarrow \quad \|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

Proof. We compute

$$\|u + v\|^2 = \langle u + v, u + v \rangle = \|u\|^2 + \langle u, v \rangle + \langle v, u \rangle + \|v\|^2.$$

If $u \perp v$, then $\langle u, v \rangle = 0$. Hence also $\langle v, u \rangle = 0$, and therefore

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

Now suppose that V is real and that

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

From the expansion above, we get

$$\|u\|^2 + 2\langle u, v \rangle + \|v\|^2 = \|u\|^2 + \|v\|^2,$$

so

$$2\langle u, v \rangle = 0.$$

Thus $\langle u, v \rangle = 0$, which means $u \perp v$. $\square$

Definition 2 (Orthogonal and orthonormal lists)**.** *A list of vectors*

$$e_1, \ldots, e_m$$

in V is called orthogonal *if*

$$\langle e_j, e_k \rangle = 0 \qquad \text{whenever } j \neq k.$$

The list is called orthonormal *if it is orthogonal and each vector has norm* 1*. Equivalently,*

$$\langle e_j, e_k \rangle = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$

Remark 1. *Orthogonal and especially orthonormal lists are very easy to work with. The next results explain why.*

Proposition 2 (Norm of an orthonormal linear combination)**.** *Suppose $e_1, \ldots, e_m$ is an orthonormal list of vectors in a real inner product space V. Then for all $a_1, \ldots, a_m \in \mathbb{R}$,*

$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = a_1^2 + \cdots + a_m^2.$$

Proof. Let
$$u = a_1 e_1 + \cdots + a_m e_m.$$
Then
$$\|u\|^2 = \langle u, u\rangle = \left\langle \sum_{j=1}^{m} a_j e_j, \sum_{k=1}^{m} a_k e_k \right\rangle.$$
Expanding, we obtain
$$\|u\|^2 = \sum_{j=1}^{m} \sum_{k=1}^{m} a_j a_k \langle e_j, e_k\rangle.$$
Since the list is orthonormal, we have
$$\langle e_j, e_k\rangle = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$
Therefore all the off-diagonal terms vanish, and only the diagonal terms remain:
$$\|u\|^2 = \sum_{j=1}^{m} a_j^2.$$
Hence
$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = a_1^2 + \cdots + a_m^2.$$
□

Corollary 1 (Orthonormal lists are linearly independent)**.** *Every orthonormal list of vectors is linearly independent.*

Proof. Suppose $e_1, \ldots, e_m$ is an orthonormal list and that
$$a_1 e_1 + \cdots + a_m e_m = 0.$$
Taking norms and using the previous proposition, we obtain
$$0 = \|a_1 e_1 + \cdots + a_m e_m\|^2 = |a_1|^2 + \cdots + |a_m|^2.$$
A sum of nonnegative numbers can be zero only if each term is zero. Thus
$$a_1 = \cdots = a_m = 0.$$
Therefore $e_1, \ldots, e_m$ is linearly independent. □

Orthogonality and Gram–Schmidt procedure

Theorem 1 (Bessel's inequality). *Suppose $e_1, \dots, e_m$ is an orthonormal list of vectors in V. If $v \in V$, then*

$$|\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2 \le \|v\|^2.$$

Proof. Let

$$u = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m$$

and

$$w = v - u.$$

Then

$$v = u + w.$$

We claim that $w \perp e_k$ for every $k \in \{1, \dots, m\}$. Indeed, for each k,

$$\langle w, e_k\rangle = \left\langle v - \sum_{j=1}^{m} \langle v, e_j\rangle e_j,\ e_k \right\rangle = \langle v, e_k\rangle - \sum_{j=1}^{m} \langle v, e_j\rangle \langle e_j, e_k\rangle.$$

Because $e_1, \dots, e_m$ is an orthonormal list, we have

$$\langle e_j, e_k\rangle = \begin{cases} 1, & j = k, \\ 0, & j \ne k. \end{cases}$$

Hence

$$\langle w, e_k\rangle = \langle v, e_k\rangle - \langle v, e_k\rangle = 0.$$

Thus w is orthogonal to each e_k, and therefore $w \perp u$.

Since $v = u + w$ and $u \perp w$, the Pythagorean theorem gives

$$\|v\|^2 = \|u\|^2 + \|w\|^2 \ge \|u\|^2.$$

Finally, since u is a linear combination of the orthonormal list $e_1, \dots, e_m$, Proposition 6.24 implies that

$$\|u\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2.$$

Therefore,

$$|\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_m\rangle|^2 \le \|v\|^2.$$

□

Definition 1 (Orthonormal basis). *An* orthonormal basis *of V is an orthonormal list of vectors in V that is also a basis of V.*

Proposition 1. *Suppose V is finite-dimensional. Then every orthonormal list of vectors in V of length* $\dim V$ *is an orthonormal basis of V.*

Proof. Every orthonormal list of vectors in V is linearly independent. Hence any orthonormal list of length $\dim V$ is a basis of V. □

Proposition 2 (Writing a vector using an orthonormal basis)**.** *Suppose $e_1, \ldots, e_n$ is an orthonormal basis of V, and let $u, v \in V$. Then:*

1.
$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n.$$

2.
$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2.$$

3.
$$\langle u, v \rangle = \langle u, e_1 \rangle \langle v, e_1 \rangle + \cdots + \langle u, e_n \rangle \langle v, e_n \rangle.$$

Proof. Because $e_1, \ldots, e_n$ is a basis of V, there exist scalars $a_1, \ldots, a_n$ such that

$$v = a_1 e_1 + \cdots + a_n e_n.$$

Taking the inner product of both sides with e_k, we get

$$\langle v, e_k \rangle = a_1 \langle e_1, e_k \rangle + \cdots + a_n \langle e_n, e_k \rangle = a_k,$$

because $e_1, \ldots, e_n$ is orthonormal. Thus

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n,$$

which proves (1).

Now (2) follows immediately from (1) and the formula for the norm of a linear combination of an orthonormal list:

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2.$$

Finally, taking the inner product of u with both sides of (1), we obtain

$$\langle u, v \rangle = \langle u,\ \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n \rangle.$$

By linearity,

$$\langle u, v \rangle = \langle v, e_1 \rangle \langle u, e_1 \rangle + \cdots + \langle v, e_n \rangle \langle u, e_n \rangle.$$

Reordering the factors gives

$$\langle u, v \rangle = \langle u, e_1 \rangle \langle v, e_1 \rangle + \cdots + \langle u, e_n \rangle \langle v, e_n \rangle.$$

This proves (3). □

Theorem 2 (Gram–Schmidt procedure)**.** *Suppose $v_1, \ldots, v_m$ is a linearly independent list of vectors in V. Let $f_1 = v_1$. For $k = 2, \ldots, m$, define f_k inductively by*

$$f_k = v_k - \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2} f_1 - \cdots - \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2} f_{k-1}.$$

For each $k = 1, \ldots, m$, let

$$e_k = \frac{f_k}{\|f_k\|}.$$

Then $e_1, \ldots, e_m$ is an orthonormal list of vectors in V such that

$$\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(e_1, \ldots, e_k)$$

for each $k = 1, \ldots, m$.

Example 1 (An orthonormal basis of $P_2(\mathbb{R})$)**.** *We make $P_2(\mathbb{R})$ into an inner product space by defining*

$$\langle p, q\rangle = \int_{-1}^{1} p(t)q(t)\,dt$$

for all $p, q \in P_2(\mathbb{R})$.

We know that $1, x, x^2$ is a basis of $P_2(\mathbb{R})$, but it is not an orthonormal basis. Indeed,

$$\langle 1, x^2\rangle = \int_{-1}^{1} t^2\,dt = \frac{2}{3} \neq 0,$$

so the vectors are not all orthogonal. Also,

$$\|1\|^2 = \int_{-1}^{1} 1\,dt = 2 \neq 1.$$

We now apply the Gram–Schmidt procedure to

$$v_1 = 1, \qquad v_2 = x, \qquad v_3 = x^2.$$

First,

$$f_1 = v_1 = 1.$$

Hence

$$\|f_1\|^2 = \int_{-1}^{1} 1\,dt = 2.$$

Next,

$$f_2 = v_2 - \frac{\langle v_2, f_1\rangle}{\|f_1\|^2} f_1 = x - \frac{\langle x, 1\rangle}{\|f_1\|^2}.$$

Because

$$\langle x, 1\rangle = \int_{-1}^{1} t\,dt = 0,$$

we get

$$f_2 = x.$$

Thus

$$\|f_2\|^2 = \int_{-1}^{1} t^2\,dt = \frac{2}{3}.$$

Now,

$$f_3 = v_3 - \frac{\langle v_3, f_1\rangle}{\|f_1\|^2} f_1 - \frac{\langle v_3, f_2\rangle}{\|f_2\|^2} f_2.$$

Substituting $v_3 = x^2$, $f_1 = 1$, and $f_2 = x$, we obtain

$$f_3 = x^2 - \frac{1}{2}\langle x^2, 1\rangle - \frac{3}{2}\langle x^2, x\rangle x.$$

Now

$$\langle x^2, 1\rangle = \int_{-1}^{1} t^2\,dt = \frac{2}{3}, \qquad \langle x^2, x\rangle = \int_{-1}^{1} t^3\,dt = 0,$$

so

$$f_3 = x^2 - \frac{1}{3}.$$

Therefore,

$$\|f_3\|^2 = \int_{-1}^{1} \left(t^2 - \frac{1}{3}\right)^2 dt = \int_{-1}^{1} \left(t^4 - \frac{2}{3}t^2 + \frac{1}{9}\right) dt = \frac{8}{45}.$$

Normalizing f_1, f_2, f_3, *we obtain*

$$e_1 = \frac{f_1}{\|f_1\|} = \frac{1}{\sqrt{2}}, \qquad e_2 = \frac{f_2}{\|f_2\|} = \sqrt{\frac{3}{2}}\,x,$$

and

$$e_3 = \frac{f_3}{\|f_3\|} = \sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right).$$

Hence

$$\left\{\frac{1}{\sqrt{2}}, \sqrt{\frac{3}{2}}\,x, \sqrt{\frac{45}{8}}\left(x^2 - \frac{1}{3}\right)\right\}$$

is an orthonormal basis of $P_2(\mathbb{R})$.

Proof. We prove the result by induction on k.

For $k = 1$, we have

$$f_1 = v_1 \qquad \text{and} \qquad e_1 = \frac{f_1}{\|f_1\|} = \frac{v_1}{\|v_1\|}.$$

Thus $\|e_1\| = 1$. Also, e_1 is a nonzero scalar multiple of v_1, so

$$\operatorname{span}(v_1) = \operatorname{span}(e_1).$$

Hence the conclusion holds for $k = 1$.

Now fix k with $2 \le k \le m$, and assume inductively that

$$e_1, \ldots, e_{k-1}$$

is an orthonormal list and that

$$\operatorname{span}(v_1, \ldots, v_{k-1}) = \operatorname{span}(e_1, \ldots, e_{k-1}). \tag{*}$$

We first show that $f_k \neq 0$. Because $v_1, \ldots, v_m$ is linearly independent, we have

$$v_k \notin \operatorname{span}(v_1, \ldots, v_{k-1}).$$

By the inductive hypothesis,

$$\operatorname{span}(v_1, \ldots, v_{k-1}) = \operatorname{span}(e_1, \ldots, e_{k-1}).$$

Since each e_j is a nonzero scalar multiple of f_j, it follows that

$$\operatorname{span}(e_1, \ldots, e_{k-1}) = \operatorname{span}(f_1, \ldots, f_{k-1}).$$

Therefore

$$v_k \notin \operatorname{span}(f_1, \ldots, f_{k-1}).$$

On the other hand, by definition,

$$f_k = v_k - \frac{\langle v_k, f_1\rangle}{\|f_1\|^2} f_1 - \cdots - \frac{\langle v_k, f_{k-1}\rangle}{\|f_{k-1}\|^2} f_{k-1}.$$

If $f_k = 0$, then v_k would be a linear combination of $f_1, \dots, f_{k-1}$, a contradiction. Hence

$$f_k \neq 0.$$

Thus $e_k = f_k / \|f_k\|$ is well defined, and clearly

$$\|e_k\| = 1.$$

Next we show that e_k is orthogonal to each of $e_1, \dots, e_{k-1}$. Let $j \in \{1, \dots, k-1\}$. Then

$$\langle e_k, e_j \rangle = \frac{1}{\|f_k\| \|f_j\|} \langle f_k, f_j \rangle.$$

So it suffices to prove that $\langle f_k, f_j \rangle = 0$. Using the definition of f_k, we get

$$\langle f_k, f_j \rangle = \left\langle v_k - \sum_{i=1}^{k-1} \frac{\langle v_k, f_i \rangle}{\|f_i\|^2} f_i,\ f_j \right\rangle.$$

By linearity,

$$\langle f_k, f_j \rangle = \langle v_k, f_j \rangle - \sum_{i=1}^{k-1} \frac{\langle v_k, f_i \rangle}{\|f_i\|^2} \langle f_i, f_j \rangle.$$

Because $e_1, \dots, e_{k-1}$ is orthonormal, the vectors $f_1, \dots, f_{k-1}$ are pairwise orthogonal. Hence

$$\langle f_i, f_j \rangle = 0 \quad \text{for } i \neq j,$$

and

$$\langle f_j, f_j \rangle = \|f_j\|^2.$$

Therefore all terms in the sum vanish except the one with $i = j$, and so

$$\langle f_k, f_j \rangle = \langle v_k, f_j \rangle - \frac{\langle v_k, f_j \rangle}{\|f_j\|^2} \|f_j\|^2 = \langle v_k, f_j \rangle - \langle v_k, f_j \rangle = 0.$$

Thus

$$\langle e_k, e_j \rangle = 0 \qquad \text{for all } j = 1, \dots, k-1.$$

Since $\|e_k\| = 1$, it follows that

$$e_1, \dots, e_k$$

is an orthonormal list.

It remains to prove that

$$\operatorname{span}(v_1, \dots, v_k) = \operatorname{span}(e_1, \dots, e_k).$$

From the definition of f_k, we can rewrite

$$v_k = f_k + \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2} f_1 + \cdots + \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2} f_{k-1}.$$

Thus

$$v_k \in \operatorname{span}(f_1, \dots, f_k) = \operatorname{span}(e_1, \dots, e_k).$$

Also, by the inductive hypothesis,

$$v_1, \ldots, v_{k-1} \in \operatorname{span}(e_1, \ldots, e_{k-1}) \subseteq \operatorname{span}(e_1, \ldots, e_k).$$

Hence

$$\operatorname{span}(v_1, \ldots, v_k) \subseteq \operatorname{span}(e_1, \ldots, e_k).$$

Now both lists

$$v_1, \ldots, v_k \qquad \text{and} \qquad e_1, \ldots, e_k$$

are linearly independent: the first by hypothesis, and the second because every orthonormal list is linearly independent. Therefore both subspaces above have dimension k. Since one is contained in the other and both have the same dimension, they must be equal. Hence

$$\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(e_1, \ldots, e_k).$$

This completes the induction, and therefore the proof. □

Theorem 3 (Existence of orthonormal bases)**.** *Every finite-dimensional inner product space has an orthonormal basis.*

Proof. Suppose V is finite-dimensional. Choose a basis of V. Applying the Gram–Schmidt procedure to this basis, we obtain an orthonormal list of vectors in V of length $\dim V$. By the result that every orthonormal list of length $\dim V$ is an orthonormal basis, this list is an orthonormal basis of V. □

Orthogonal Complements

Definition 1 (Orthogonal complement)**.** *Let $U \subseteq V$. The* orthogonal complement *of U, denoted by $U^\perp$, is defined by*

$$U^\perp = \{v \in V : \langle u, v\rangle = 0 \text{ for every } u \in U\}.$$

Remark 1. *The orthogonal complement $U^\perp$ depends on both V and U. Usually the ambient inner product space V is clear from the context, so it is omitted from the notation.*

Proposition 1 (Properties of orthogonal complements)**.** *Let $U, G, H \subseteq V$. Then:*

1. *If $U \subseteq V$, then $U^\perp$ is a subspace of V.*
2. *$\{0\}^\perp = V$.*
3. *$V^\perp = \{0\}$.*
4. *If $U \subseteq V$, then*
$$U \cap U^\perp \subseteq \{0\}.$$
5. *If $G \subseteq H$, then*
$$H^\perp \subseteq G^\perp.$$

Proof. 1. First, $0 \in U^\perp$ because $\langle u, 0\rangle = 0$ for every $u \in U$.

Now let $v, w \in U^\perp$. For every $u \in U$,

$$\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle = 0 + 0 = 0.$$

Hence $v + w \in U^\perp$.

If $\lambda \in \mathbb{R}$ and $v \in U^\perp$, then for every $u \in U$,

$$\langle u, \lambda v\rangle = \lambda\langle u, v\rangle = \lambda \cdot 0 = 0.$$

Hence $\lambda v \in U^\perp$. Therefore $U^\perp$ is a subspace of V.

2. Let $v \in V$. Since $\langle 0, v\rangle = 0$, we have $v \in \{0\}^\perp$. Thus $\{0\}^\perp = V$.

3. If $v \in V^\perp$, then in particular $\langle v, v\rangle = 0$, which implies $v = 0$. Hence $V^\perp = \{0\}$.

4. If $u \in U \cap U^\perp$, then $u \in U$ and $u \in U^\perp$. Therefore

$$\langle u, u\rangle = 0,$$

so $u = 0$. Thus $U \cap U^\perp \subseteq \{0\}$.

5. Suppose $G \subseteq H$, and let $v \in H^{\perp}$. Then $\langle h, v\rangle = 0$ for every $h \in H$. Since every element of G belongs to H, it follows that $\langle g, v\rangle = 0$ for every $g \in G$. Hence $v \in G^{\perp}$, so

$$H^{\perp} \subseteq G^{\perp}.$$

□

Proposition 2 (Direct sum decomposition)**.** *Suppose U is a finite-dimensional subspace of V. Then*

$$V = U \oplus U^{\perp}.$$

Proof. We first show that

$$V = U + U^{\perp}.$$

Let $v \in V$, and let $e_1, \ldots, e_m$ be an orthonormal basis of U. Define

$$u = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m$$

and

$$w = v - u.$$

Because each $e_j \in U$, we have $u \in U$. Also, for each $k = 1, \ldots, m$,

$$\langle w, e_k\rangle = \left\langle v - \sum_{j=1}^{m} \langle v, e_j\rangle e_j,\ e_k \right\rangle = \langle v, e_k\rangle - \sum_{j=1}^{m} \langle v, e_j\rangle \langle e_j, e_k\rangle = \langle v, e_k\rangle - \langle v, e_k\rangle = 0.$$

Thus w is orthogonal to every vector in $\operatorname{span}(e_1, \ldots, e_m) = U$, so $w \in U^{\perp}$.

Therefore $v = u + w$ with $u \in U$ and $w \in U^{\perp}$, proving that $V = U + U^{\perp}$.

From the previous proposition, $U \cap U^{\perp} = \{0\}$. Hence

$$V = U \oplus U^{\perp}.$$

□

Corollary 1 (Dimension of the orthogonal complement)**.** *Suppose V is finite-dimensional and U is a subspace of V. Then*

$$\dim U^{\perp} = \dim V - \dim U.$$

Proof. This follows immediately from

$$V = U \oplus U^{\perp}.$$

□

Proposition 3 (Orthogonal complement of the orthogonal complement)**.** *Suppose U is a finite-dimensional subspace of V. Then*

$$U = (U^{\perp})^{\perp}.$$

Proof. We first show that

$$U \subseteq (U^{\perp})^{\perp}.$$

If $u \in U$ and $w \in U^{\perp}$, then by definition $\langle u, w\rangle = 0$. Hence $u \in (U^{\perp})^{\perp}$.

Now let $v \in (U^{\perp})^{\perp}$. By the direct sum decomposition,

$$v = u + w \qquad \text{with } u \in U,\ w \in U^{\perp}.$$

Then

$$w = v - u.$$

Since $v \in (U^\perp)^\perp$ and $u \in U \subseteq (U^\perp)^\perp$, it follows that $w \in (U^\perp)^\perp$. Therefore

$$w \in U^\perp \cap (U^\perp)^\perp.$$

By the property $W \cap W^\perp \subseteq \{0\}$, we get $w = 0$. Thus $v = u \in U$, showing that

$$(U^\perp)^\perp \subseteq U.$$

Hence

$$U = (U^\perp)^\perp.$$

□

Proposition 4. *Suppose U is a finite-dimensional subspace of V. Then*

$$U^\perp = \{0\} \quad \iff \quad U = V.$$

Proof. If $U^\perp = \{0\}$, then by the previous proposition,

$$U = (U^\perp)^\perp = \{0\}^\perp = V.$$

Conversely, if $U = V$, then

$$U^\perp = V^\perp = \{0\}.$$

□

Example 1 (In-class exercise)**.** *Let $V = C[-1,1]$, the vector space of continuous real-valued functions on $[-1,1]$, with inner product*

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx.$$

Let

$$U = \{f \in C[-1,1] : f(0) = 0\}.$$

Show that:

1. $U^\perp = \{0\}$.
2. *This shows that the statements*

$$V = U \oplus U^\perp \qquad \text{and} \qquad U = (U^\perp)^\perp$$

may fail without the hypothesis that U is finite-dimensional.

Solution. We show first that $U^\perp = \{0\}$.

Let $g \in U^\perp$. We want to prove that $g = 0$. Suppose, for contradiction, that $g \neq 0$. Then there exists $x_0 \in [-1,1]$ such that $g(x_0) \neq 0$. Replacing g by $-g$ if necessary, we may assume

$$g(x_0) > 0.$$

By continuity of g, there exists an interval $I \subseteq [-1,1]$ containing x_0 such that

$$g(x) > 0 \qquad \text{for all } x \in I.$$

We now construct $f \in U$ such that

$$\int_{-1}^{1} f(x)g(x)\,dx > 0,$$

which will contradict the assumption that $g \in U^{\perp}$.

If $x_0 \neq 0$, we may choose I so that $0 \notin I$. Let f be any continuous function such that

$$f \geq 0, \qquad f \not\equiv 0, \qquad \operatorname{supp}(f) \subseteq I.$$

Since $0 \notin I$, we have $f(0) = 0$, so $f \in U$. Moreover,

$$\int_{-1}^{1} f(x)g(x)\,dx = \int_I f(x)g(x)\,dx > 0,$$

because $f \geq 0$, $f \not\equiv 0$, and $g > 0$ on I. This is a contradiction.

If $x_0 = 0$, then $g(0) > 0$. By continuity, there exists $\delta > 0$ such that

$$g(x) > 0 \qquad \text{for all } x \in [-\delta, \delta].$$

Define

$$f(x) = \begin{cases} x^2(\delta - |x|), & |x| \leq \delta, \\ 0, & |x| > \delta. \end{cases}$$

Then f is continuous, $f \geq 0$, $f \not\equiv 0$, and $f(0) = 0$. Hence $f \in U$. Again,

$$\int_{-1}^{1} f(x)g(x)\,dx > 0,$$

a contradiction.

Therefore $g = 0$, and so

$$U^{\perp} = \{0\}.$$

Now $U \neq V$, because for example the constant function 1 belongs to V but not to U. Hence

$$U \oplus U^{\perp} = U \oplus \{0\} = U \neq V,$$

so the identity $V = U \oplus U^{\perp}$ fails.

Also,

$$(U^{\perp})^{\perp} = \{0\}^{\perp} = V \neq U,$$

so the identity $U = (U^{\perp})^{\perp}$ also fails.

This shows that both statements can fail when U is not finite-dimensional. $\square$

Definition 2 (Orthogonal projection)**.** *Suppose U is a finite-dimensional subspace of V. The* orthogonal projection *of V onto U is the operator*

$$P_U \in \mathcal{L}(V)$$

defined as follows: for each $v \in V$, write

$$v = u + w,$$

where $u \in U$ and $w \in U^{\perp}$. Then define

$$P_U v = u.$$

The direct sum decomposition

$$V = U \oplus U^{\perp}$$

shows that each $v \in V$ can be uniquely written in the form

$$v = u + w$$

with $u \in U$ and $w \in U^{\perp}$. Thus $P_U v$ is well defined.

Orthogonal Projection and the Riesz Representation Theorem

Definition 1 (Orthogonal projection)**.** *Let U be a finite-dimensional subspace of an inner product space V. For each $v \in V$, write*

$$v = u + w, \qquad \text{with } u \in U,\ w \in U^{\perp}.$$

The orthogonal projection *of v onto U is defined by*

$$P_U v = u.$$

Proposition 1 (Minimizing distance to a subspace)**.** *Suppose U is a finite-dimensional subspace of V, $v \in V$, and $u \in U$. Then*

$$\|v - P_U v\| \leq \|v - u\|.$$

Furthermore, the inequality above is an equality if and only if $u = P_U v$.

Proof. We have

$$\begin{aligned} \|v - P_U v\|^2 &\leq \|v - P_U v\|^2 + \|P_U v - u\|^2 \\ &= \|(v - P_U v) + (P_U v - u)\|^2 \\ &= \|v - u\|^2, \end{aligned}$$

where the first line above holds because

$$0 \leq \|P_U v - u\|^2,$$

the second line above comes from the Pythagorean theorem [which applies because

$$v - P_U v \in U^{\perp} \quad \text{and} \quad P_U v - u \in U],$$

and the third line above holds by simple algebra. Taking square roots gives the desired inequality.

The inequality proved above is an equality if and only if

$$\|v - P_U v\|^2 = \|v - P_U v\|^2 + \|P_U v - u\|^2,$$

which happens if and only if

$$\|P_U v - u\| = 0,$$

which happens if and only if $u = P_U v$. □

Example 1 (In-class exercise)**.** *Let*

$$E = \mathbb{R}_4[X]$$

with inner product

$$\langle P, Q \rangle = \frac{1}{2} \int_{-1}^{1} P(x) Q(x)\, dx.$$

1. *Find an orthonormal basis of* $\operatorname{span}(1, X, X^2)$ *using the Gram–Schmidt process.*
2. *Show that the orthogonal projection of* X^4 *onto* $\operatorname{span}(1, X, X^2)$ *is of the form*

$$aX^2 + b$$

   for some real numbers a, b *(which you do not need to compute).*
3. *Deduce that*

$$\int_{-1}^{1} (x^4 - \alpha x^2 - \beta x - \gamma)^2 \, dx$$

   is minimized for a unique choice of α, β, γ, *and express that choice in terms of* a *and* b.

Solution. Set

$$U = \operatorname{span}(1, X, X^2).$$

For the first vector, note that

$$\|1\|^2 = \langle 1, 1 \rangle = \frac{1}{2} \int_{-1}^{1} 1 \, dx = 1.$$

Hence we may take

$$e_1 = 1.$$

Next,

$$\langle X, e_1 \rangle = \frac{1}{2} \int_{-1}^{1} x \, dx = 0,$$

so X is already orthogonal to e_1. Also,

$$\|X\|^2 = \frac{1}{2} \int_{-1}^{1} x^2 \, dx = \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}.$$

Thus

$$e_2 = \frac{X}{\|X\|} = \sqrt{3}\, X.$$

Now we orthogonalize X^2. First,

$$\langle X^2, e_1 \rangle = \frac{1}{2} \int_{-1}^{1} x^2 \, dx = \frac{1}{3},$$

and

$$\langle X^2, e_2 \rangle = \left\langle X^2, \sqrt{3}\, X \right\rangle = \frac{\sqrt{3}}{2} \int_{-1}^{1} x^3 \, dx = 0.$$

Therefore

$$X^2 - \langle X^2, e_1 \rangle e_1 - \langle X^2, e_2 \rangle e_2 = X^2 - \frac{1}{3}.$$

Its norm is

$$\left\| X^2 - \frac{1}{3} \right\|^2 = \frac{1}{2} \int_{-1}^{1} \left(x^2 - \frac{1}{3} \right)^2 dx.$$

Expanding,

$$\left(x^2 - \frac{1}{3} \right)^2 = x^4 - \frac{2}{3} x^2 + \frac{1}{9},$$

so

$$\left\| X^2 - \frac{1}{3} \right\|^2 = \frac{1}{2} \int_{-1}^{1} x^4\, dx - \frac{1}{3} \int_{-1}^{1} x^2\, dx + \frac{1}{18} \int_{-1}^{1} 1\, dx.$$

Using

$$\int_{-1}^{1} x^4\, dx = \frac{2}{5}, \qquad \int_{-1}^{1} x^2\, dx = \frac{2}{3}, \qquad \int_{-1}^{1} 1\, dx = 2,$$

we get

$$\left\| X^2 - \frac{1}{3} \right\|^2 = \frac{1}{5} - \frac{2}{9} + \frac{1}{9} = \frac{4}{45}.$$

Hence

$$\left\| X^2 - \frac{1}{3} \right\| = \frac{2}{\sqrt{45}},$$

and therefore

$$e_3 = \frac{X^2 - \frac{1}{3}}{\|X^2 - \frac{1}{3}\|} = \frac{\sqrt{45}}{2} \left(X^2 - \frac{1}{3} \right).$$

So an orthonormal basis of $\operatorname{span}(1, X, X^2)$ is

$$\boxed{e_1 = 1, \qquad e_2 = \sqrt{3}\, X, \qquad e_3 = \frac{\sqrt{45}}{2} \left(X^2 - \frac{1}{3} \right).}$$

For part (2), the orthogonal projection of X^4 onto U is

$$P_U(X^4) = \langle X^4, e_1 \rangle e_1 + \langle X^4, e_2 \rangle e_2 + \langle X^4, e_3 \rangle e_3.$$

Now

$$\langle X^4, e_2 \rangle = \left\langle X^4, \sqrt{3}\, X \right\rangle = \frac{\sqrt{3}}{2} \int_{-1}^{1} x^5\, dx = 0$$

because x^5 is odd. Thus

$$P_U(X^4) = \langle X^4, e_1 \rangle e_1 + \langle X^4, e_3 \rangle e_3.$$

Since $e_1 = 1$ and e_3 is of the form $c\left(X^2 - \frac{1}{3}\right)$, it follows that

$$P_U(X^4)$$

is a linear combination of 1 and $X^2 - \frac{1}{3}$, hence of the form

$$\boxed{P_U(X^4) = aX^2 + b}$$

for some real numbers a, b.

For part (3), note that

$$\int_{-1}^{1} (x^4 - \alpha x^2 - \beta x - \gamma)^2\, dx = 2\, \|X^4 - (\alpha X^2 + \beta X + \gamma)\|^2.$$

Thus minimizing the integral is equivalent to minimizing

$$\|X^4 - Q\|^2 \qquad \text{over } Q \in U.$$

By the characterization of orthogonal projection, the unique minimizer is

$$Q = P_U(X^4).$$

Since

$$P_U(X^4) = aX^2 + b,$$

the unique minimizing choice is

$$\boxed{\alpha = a, \qquad \beta = 0, \qquad \gamma = b.}$$

□

Theorem 1 (Riesz representation theorem, real case)**.** *Suppose V is a finite-dimensional real inner product space and $\varphi : V \to \mathbb{R}$ is a linear functional. Then there exists a unique vector $v \in V$ such that*

$$\varphi(u) = \langle u, v \rangle \qquad \textit{for every } u \in V.$$

Proof. Let $e_1, \ldots, e_n$ be an orthonormal basis of V. Then every $u \in V$ can be written as

$$u = \langle u, e_1 \rangle e_1 + \cdots + \langle u, e_n \rangle e_n.$$

Applying φ and using linearity, we get

$$\varphi(u) = \varphi\big(\langle u, e_1 \rangle e_1 + \cdots + \langle u, e_n \rangle e_n\big)$$

$$= \langle u, e_1 \rangle \varphi(e_1) + \cdots + \langle u, e_n \rangle \varphi(e_n).$$

Now define

$$v = \varphi(e_1) e_1 + \cdots + \varphi(e_n) e_n.$$

Then

$$\langle u, v \rangle = \langle u, \varphi(e_1) e_1 + \cdots + \varphi(e_n) e_n \rangle$$
$$= \langle u, e_1 \rangle \varphi(e_1) + \cdots + \langle u, e_n \rangle \varphi(e_n) = \varphi(u).$$

Thus there exists $v \in V$ such that

$$\varphi(u) = \langle u, v \rangle \qquad \text{for every } u \in V.$$

To prove uniqueness, suppose $v_1, v_2 \in V$ satisfy

$$\varphi(u) = \langle u, v_1 \rangle = \langle u, v_2 \rangle \qquad \text{for every } u \in V.$$

Then

$$0 = \langle u, v_1 \rangle - \langle u, v_2 \rangle = \langle u, v_1 - v_2 \rangle \qquad \text{for every } u \in V.$$

Taking $u = v_1 - v_2$, we obtain

$$0 = \langle v_1 - v_2, v_1 - v_2 \rangle = \|v_1 - v_2\|^2.$$

Hence $v_1 - v_2 = 0$, so $v_1 = v_2$. Therefore the vector v is unique. □

Linear Algebra
Adjoints and Conjugate Transposes

Until now, we have mostly worked with real inner product spaces. Today we will also briefly discuss complex inner product spaces, because the adjoint is most naturally connected with the conjugate transpose of a matrix.

Throughout this lecture, $\mathbb{F}$ denotes either $\mathbb{R}$ or $\mathbb{C}$.

1. Inner products over $\mathbb{C}$

In a real inner product space, the inner product satisfies

$$\langle x, y\rangle = \langle y, x\rangle.$$

Over $\mathbb{C}$, this symmetry must be modified. The correct condition is

$$\langle x, y\rangle = \overline{\langle y, x\rangle}.$$

Definition 1 (Complex inner product)**.** *Let V be a vector space over $\mathbb{C}$. An inner product on V is a function*

$$\langle\cdot,\cdot\rangle : V \times V \to \mathbb{C}$$

such that:

1. *for every fixed $y \in V$, the map*
$$x \mapsto \langle x, y\rangle$$
is linear;
2. *for all $x, y \in V$,*
$$\langle x, y\rangle = \overline{\langle y, x\rangle};$$
3. *for every nonzero $x \in V$,*
$$\langle x, x\rangle > 0.$$

The first condition says that the inner product is linear in the first slot. Thus

$$\langle \lambda x_1 + \mu x_2, y\rangle = \lambda\langle x_1, y\rangle + \mu\langle x_2, y\rangle.$$

The second condition implies that the inner product is conjugate-linear in the second slot:

$$\langle x, \lambda y\rangle = \overline{\lambda}\langle x, y\rangle.$$

Indeed,

$$\langle x, \lambda y\rangle = \overline{\langle \lambda y, x\rangle} = \overline{\lambda\langle y, x\rangle} = \overline{\lambda}\,\overline{\langle y, x\rangle} = \overline{\lambda}\langle x, y\rangle.$$

Example 1. *The standard inner product on $\mathbb{C}^n$ is*

$$\langle x, y\rangle = x_1\overline{y_1} + \cdots + x_n\overline{y_n}.$$

If

$$x = (x_1, \ldots, x_n), \qquad y = (y_1, \ldots, y_n),$$

then

$$\langle x, y\rangle = \sum_{j=1}^{n} x_j\overline{y_j}.$$

Remark 1. *If all scalars are real, then complex conjugation does nothing. Thus this definition reduces to the usual definition of a real inner product.*

Definition 2 (Adjoint)**.** *Let V, W be finite-dimensional inner product spaces over $\mathbb{F}$, and let*

$$T \in \mathcal{L}(V, W).$$

The adjoint of T is the function

$$T^* : W \to V$$

such that

$$\langle Tv, w\rangle_W = \langle v, T^*w\rangle_V$$

for every $v \in V$ and every $w \in W$.

Notice carefully the direction:

$$T : V \to W, \qquad T^* : W \to V.$$

The adjoint goes in the opposite direction.

We now prove that the adjoint exists and is unique in finite dimensions.

Theorem 1. *Let V, W be finite-dimensional inner product spaces, and let*

$$T \in \mathcal{L}(V, W).$$

Then there exists a unique linear map

$$T^* \in \mathcal{L}(W, V)$$

such that

$$\langle Tv, w\rangle_W = \langle v, T^*w\rangle_V$$

for every $v \in V$ and every $w \in W$.

Proof. Fix $w \in W$. Define

$$\varphi_w : V \to \mathbb{F}$$

by

$$\varphi_w(v) = \langle Tv, w\rangle_W.$$

This is a linear functional on V, because T is linear and the inner product is linear in the first slot.

By the Riesz representation theorem, there exists a unique vector $u \in V$ such that

$$\varphi_w(v) = \langle v, u\rangle_V$$

for every $v \in V$. In other words, there exists a unique vector $u \in V$ such that

$$\langle Tv, w\rangle_W = \langle v, u\rangle_V$$

for every $v \in V$.

We define

$$T^*w = u.$$

Thus

$$\langle Tv, w\rangle_W = \langle v, T^*w\rangle_V$$

for every $v \in V$.

This defines a function

$$T^* : W \to V.$$

We now prove that T^* is linear. Let $w_1, w_2 \in W$. For every $v \in V$,

$$\begin{aligned}
\langle v, T^*(w_1 + w_2)\rangle &= \langle Tv, w_1 + w_2\rangle \\
&= \langle Tv, w_1\rangle + \langle Tv, w_2\rangle \\
&= \langle v, T^* w_1\rangle + \langle v, T^* w_2\rangle \\
&= \langle v, T^* w_1 + T^* w_2\rangle.
\end{aligned}$$

Therefore

$$\langle v, T^*(w_1 + w_2) - T^* w_1 - T^* w_2\rangle = 0$$

for all $v \in V$. Taking

$$v = T^*(w_1 + w_2) - T^* w_1 - T^* w_2,$$

we get

$$T^*(w_1 + w_2) = T^* w_1 + T^* w_2.$$

Now let $\lambda \in \mathbb{F}$ and $w \in W$. For every $v \in V$,

$$\begin{aligned}
\langle v, T^*(\lambda w)\rangle &= \langle Tv, \lambda w\rangle \\
&= \overline{\lambda}\langle Tv, w\rangle \\
&= \overline{\lambda}\langle v, T^* w\rangle \\
&= \langle v, \lambda T^* w\rangle.
\end{aligned}$$

Therefore

$$T^*(\lambda w) = \lambda T^* w.$$

Thus T^* is linear.

Finally, uniqueness follows from the uniqueness in the Riesz representation theorem. More explicitly, suppose $S : W \to V$ also satisfies

$$\langle Tv, w\rangle = \langle v, Sw\rangle$$

for every $v \in V$ and every $w \in W$. Then

$$\langle v, T^* w - Sw\rangle = 0$$

for every $v \in V$. Taking $v = T^* w - Sw$, we get

$$T^* w = Sw.$$

Since this holds for every $w \in W$, we conclude that

$$T^* = S.$$

□

5. A concrete example

Example 2. *Define* $T : \mathbb{R}^3 \to \mathbb{R}^2$ *by*

$$T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1).$$

We compute T^*.

Let

$$x = (x_1, x_2, x_3) \in \mathbb{R}^3, \qquad y = (y_1, y_2) \in \mathbb{R}^2.$$

Then

$$\begin{aligned} \langle Tx, y\rangle &= \langle (x_2 + 3x_3, 2x_1), (y_1, y_2)\rangle \\ &= (x_2 + 3x_3)y_1 + 2x_1y_2 \\ &= 2x_1y_2 + x_2y_1 + 3x_3y_1 \\ &= \langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1)\rangle. \end{aligned}$$

Therefore

$$T^*(y_1, y_2) = (2y_2, y_1, 3y_1).$$

Notice that

$$T : \mathbb{R}^3 \to \mathbb{R}^2, \qquad T^* : \mathbb{R}^2 \to \mathbb{R}^3.$$

The matrix of T in the standard bases is

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

The transpose is

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

This matrix represents precisely the map

$$(y_1, y_2) \mapsto (2y_2, y_1, 3y_1).$$

Thus, in this example,

$$\mathcal{M}(T^*) = \mathcal{M}(T)^t.$$

6. Matrix of the adjoint

We now prove the general result.

Theorem 2. *Let* $T \in \mathcal{L}(V, W)$. *Suppose*

$$e_1, \ldots, e_n$$

is an orthonormal basis of V, *and*

$$f_1, \ldots, f_m$$

is an orthonormal basis of W. *Then*

$$\mathcal{M}(T^*) = (\mathcal{M}(T))^*,$$

where $(\mathcal{M}(T))^*$ *denotes the conjugate transpose, that is,*

$$\big((\mathcal{M}(T))^*\big)_{jk} = \overline{(\mathcal{M}(T))_{kj}}.$$

In the real case, this becomes

$$\mathcal{M}(T^*) = (\mathcal{M}(T))^t.$$

Proof. Write

$$A = \mathcal{M}(T), \qquad B = \mathcal{M}(T^*).$$

First, we describe the entries of A. The k-th column of A is obtained by writing Te_k as a linear combination of the basis vectors

$$f_1, \dots, f_m.$$

Since this basis is orthonormal,

$$Te_k = \langle Te_k, f_1\rangle f_1 + \cdots + \langle Te_k, f_m\rangle f_m.$$

Thus the entry in row j, column k of A is

$$A_{jk} = \langle Te_k, f_j\rangle.$$

Now we describe the entries of B. Since $T^* : W \to V$, the k-th column of B is obtained by writing $T^* f_k$ in the orthonormal basis

$$e_1, \dots, e_n.$$

Thus the entry in row j, column k of B is

$$B_{jk} = \langle T^* f_k, e_j\rangle.$$

Using conjugate symmetry and the definition of the adjoint,

$$\begin{aligned} B_{jk} &= \langle T^* f_k, e_j\rangle \\ &= \overline{\langle e_j, T^* f_k\rangle} \\ &= \overline{\langle Te_j, f_k\rangle} \\ &= \overline{A_{kj}}. \end{aligned}$$

Therefore

$$B_{jk} = \overline{A_{kj}}.$$

This means exactly that

$$B = A^*.$$

Hence

$$\mathcal{M}(T^*) = (\mathcal{M}(T))^*.$$

□

Remark 2. *The orthonormality assumption is important. If the bases are not orthonormal, the matrix of the adjoint is not necessarily the transpose or conjugate transpose of the matrix.*

7. Example with a complex matrix

Example 3. *Let $T : \mathbb{C}^3 \to \mathbb{C}^2$ be the linear map whose matrix in the standard bases is*

$$A = \begin{pmatrix} 2 & 3+4i & 7 \\ 6 & 5 & 8i \end{pmatrix}.$$

Since the standard bases are orthonormal, the matrix of $T^* : \mathbb{C}^2 \to \mathbb{C}^3$ is the conjugate transpose of A:

$$A^* = \begin{pmatrix} 2 & 6 \\ 3-4i & 5 \\ 7 & -8i \end{pmatrix}.$$

Thus

$$\mathcal{M}(T^*) = \begin{pmatrix} 2 & 6 \\ 3-4i & 5 \\ 7 & -8i \end{pmatrix}.$$