Lecture Notes for

Linear Algebra

Notes typed by
Sayan Das

Fourth Semester 2025
Jadavpur University

Contents

Syllabus

Dual spaces. — Dual spaces, dual basis, double dual, transpose of a linear transformation and its matrix in the dual basis, annihilators. Eigen spaces of a linear operator, the minimal polynomial for a linear operator, diagonalisability, invariant subspaces and Cayley-Hamilton theorem.

Inner product spaces and orthogonality. — Inner product spaces and norms, Gram-Schmidt orthogonalisation process, orthogonal complements, best approximation, Bessel's inequality, the adjoint of a linear operator. Normal, unitary, self-adjoint and positive operators. Orthogonal projections and Spectral theorem.

Determinant and generalised inverses. — Characterisation of determinant as multi-linear function, Generalized inverses of rectangular matrices, Moore-Penrose (MP) inverse, Singular value decomposition (SVD) of a matrix. reduced SVD, nearest low rank matrix, applications of SVD.

Chapter 1.

1

Dual spaces

1.1. Dual spaces

1.1.1. Linear functionals. — Vector spaces arise naturally from the set of solutions to a system of linear equations (hence why they're also called *linear spaces*). The *dual space* of a vector space arises from the set of *linear functionals* over a vector space. Linear functionals are the foundation of the subject of *functional analysis* and find applications in the theory of distributions as well as quantum mechanics. Let V, W be two $\mathbb{F}$-vector spaces. Consider the set of all $T : V \to W$ such that T is a linear transformation. Denote this set by $\mathcal{L}(V, W)$, i.e.,

$$\mathcal{L}(V, W) := \left\{ T : V \to W \,\middle|\, T \text{ is linear} \right\}.$$

As linear transformations are homomorphisms between vector spaces, $\mathcal{L}(V, W)$ is oft denoted $\mathrm{Hom}(V, W)$. This $\mathcal{L}(V, W)$ forms an $\mathbb{F}$-vector space, where for all $f, g \in \mathcal{L}(V, W),\ \lambda \in \mathbb{F}, v \in V$

$$\begin{aligned}(f + g)(v) &= f(v) + g(v),\\ (\lambda f)(v) &= \lambda f(v).\end{aligned}$$

If V and W are finite-dimensional with $\dim(V) = m,\ \dim(W) = n$, then $V \cong \mathbb{F}^m,\ W \cong \mathbb{F}^n$ and $V \times W \cong V \oplus W$.[1] Thus $V \times W \cong \mathbb{F}^{m+n}$. In particular, we have $\mathcal{L}(V, W) \cong \mathbb{F}^{m \times n}$, thus $\dim(\mathcal{L}(V, W)) = mn$. If $W = \mathbb{F}$, we call the elements of $\mathcal{L}(V, \mathbb{F})$ *linear functionals.* More concretely, a linear functional on an $\mathbb{F}$-vector space V is a function $f : V \to \mathbb{F}$ such that $f(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 f(x_1) + \lambda_2 f(x_2)$ for all $x_1, x_2 \in V, \lambda_1, \lambda_2 \in \mathbb{F}$.

1.1.2. Example. — Let $\mathcal{P} = \mathbb{C}[t]$ be the $\mathbb{C}$-vector space of all complex polynomials in t, and when we specify a subscript n (so $\mathcal{P}_n$) with degree at most n. Define $f(x(t)) = x(0)$ (the constant term) of every polynomial $x(t) \in \mathcal{P}$. This is a linear functional $f : \mathcal{P} \to \mathbb{C}$. More generally, for any n scalars $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ and real $t_1, \ldots, t_n$ the function

$$f(x(t)) = \lambda_1 x(t_1) + \cdots + \lambda_n x(t_n)$$

is a linear functional $f \in \mathcal{L}(\mathcal{P}, \mathbb{C})$.

Another linear functional, in a sense a limiting case of the above, is obtained as follows. Let (a, b) be an open interval on the real t-axis and let $g(t) : (a, b) \to \mathbb{C}$ be integrable; for a complex-valued function such as g we simply have $\int g = \int \Re(g) + i \int \Im(g)$. Then

$$f(x(t)) = \int_a^b g(t)x(t)dt$$

defines a linear functional $f \in \mathcal{L}(\mathcal{P}, \mathbb{C})$.

[1] Recall that $U = V \oplus W$ if $U = V + W$ and $V \cap W = \{0\}$. Or equivalently, if $U = V + W$ and the representation of every $w \in W$ as $u + v, u \in U, v \in V$ is unique.

1.1.3. Example. — Let $C[0, 2\pi]$ be the vector space of all continuous $f : [0, 2\pi] \to \mathbb{R}$. Then for any $g(t) \in C[0, 2\pi]$ we have that the *nth Fourier coefficient* $h(x(t))$ of $x(t) \in C[0, 2\pi]$,

$$h(x(t)) = \frac{1}{2\pi} \int_0^{2\pi} x(t)g(t)dt$$

is a linear functional on $C[0, 2\pi]$. So $h \in \mathcal{L}(C[0, 2\pi], \mathbb{R})$.

In general we have linear functionals $\eta \in \mathcal{L}(C[a, b], \mathbb{R})$ defined by (for fixed $g_0 \in C[a, b]$),

$$\eta : f \mapsto \int_a^b f(t)g_0(t)dt.$$

1.1.4. Example. — Let $C^n(U)$ be the subspace of $\mathbb{R}^U$ for a subset $U \subseteq \mathbb{R}$ consisting of all n-times differentiable $f : U \to \mathbb{R}$. Then the differential operator $D : C^n(\mathbb{R}) \to \mathbb{R}$ sending $f \mapsto f'$ is a linear functional $D \in \mathcal{L}(C^n(U), \mathbb{R})$.

Let $C^\infty(U)$ denote the vector space of all infinitely differentiable functions $f \in \mathbb{R}^U$ where U is a compact subset of $\mathbb{R}$. Then a *distribution* on U is a continuous linear functional $f \in \mathcal{L}(C^\infty(U), \mathbb{R})$.

1.1.5. Example. — Let $\mathrm{tr} : \mathbb{F}^{n\times n} \to \mathbb{F}$ be the *trace* of a square matrix of order n, defined as $tr\left((a_{ij})_{\substack{1\leqslant i\leqslant n\\ 1\leqslant j\leqslant n}}\right) = \sum_{i=1}^{n} a_{ii}$. Then $\mathrm{tr}(\lambda A + \mu B) = \lambda\,\mathrm{tr}(A) + \mu\,\mathrm{tr}(B)$, so it is a linear functional on $\mathbb{F}^{n\times n}$, i.e., $\mathrm{tr} \in \mathcal{L}(\mathbb{F}^{n\times n}, \mathbb{F})$.

1.1.6. Example. — Given a field $\mathbb{F}$, $\mathbb{F}^n$ is an n-dimensional $\mathbb{F}$-vector space. The *inner product* $v \cdot w$ of $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{F}^n$ is the scalar $v \cdot w = v^T w = \sum_{i=1}^n v_i w_i \in \mathbb{F}$.

Hence any $v \in \mathbb{F}^n$ defines a linear functional in $\mathcal{L}(\mathbb{F}^n, \mathbb{F})$ by $w \mapsto v \cdot w$. The *exterior product* is defined as $v \wedge w = vw^T \in \mathbb{F}^{n\times n}$. In the Dirac notation of quantum mechanics, inner products are called *bra-ket products* while exterior products are called *ket-bra products*.

1.1.7. Example. — Let V be a finite dimensional $\mathbb{F}$-vector space with $\dim(V) = n$ and $\beta = \{x_1, \ldots, x_n\}$ be an ordered basis for V. Let the coordinate vector of x wrt β be

$$[x]_\beta = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

Then for each $1 \leqslant i \leqslant n$, we define the *ith projection map* $f_i(x) := a_i$, which yields the linear functionals $f_i \in \mathcal{L}(V, \mathbb{F})$. In particular, the linear functionals $f_i(x_j) = \delta_{ij}$ yield exactly the *Kronecker delta*

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

1.1.8. Definition (Linear functional). — A **linear functional** on an $\mathbb{F}$-vector space V is a linear transformation from V to its field of scalars, i.e., $f : V \to \mathbb{F}$. Equivalently, $f \in \mathcal{L}(V, \mathbb{F})$.

1.1.9. Definition (Dual space). — For an $\mathbb{F}$-vector space V, the **dual space** of V is the vector space of all linear functionals on V, i.e., $V' = \mathcal{L}(V, \mathbb{F})$.

1.1.10. Theorem. — *Let V be a finite-dimensional $\mathbb{F}$-vector space. Then* $\dim(V) = \dim(V')$.

Proof. $\dim(V') = \dim(\mathcal{L}(V, \mathbb{F})) = \dim_{\mathbb{F}}(V) \dim_{\mathbb{F}}(\mathbb{F}) = \dim(V)$. □

1.1.11. Definition (Dual basis). — Let V be an $\mathbb{F}$-vector space and $B = \{v_1, v_2, \ldots, v_n\}$ be a basis of V. Then the **dual basis** of B is the set $B' = \{\phi_1, \phi_2, \ldots, \phi_n\}$ of elements of V', where each ϕ_i is a linear functional on V s.t.

$$\phi_i(v_j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i. \end{cases}$$

1.1.12. Example. — The dual basis of the standard basis $\{e_1, \ldots, e_n\}$ of $\mathbb{F}^n$ is $\{\phi_1, \ldots, \phi_n\}$ where

$$\phi_i(e_j) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i. \end{cases}$$

1.1.13. Example. — The dual basis of the standard basis $\{(0, 1), (1, 0)\}$ of $\mathbb{R}^2$ is $\{\phi_1, \phi_2\}$ where

$$\begin{aligned} \phi_1(1, 0) &= 1, & \phi_2(1, 0) &= 0 \\ \phi_1(0, 1) &= 0, & \phi_2(0, 1) &= 1 \end{aligned}$$

and as a linear functional is of the form $\phi(x, y) = ax + by$, we get $\phi_1(x, y) = x, \phi_2(x, y) = y$.

1.1.14. Question. — Find the dual basis of:

1. $\{(-1, 2), (0, 1)\}$ of $\mathbb{R}^2$. *Ans.* $\{\phi_1, \phi_2\}$ *given by* $\phi_1(x, y) = -x,\ \phi_2(x, y) = 2x + y$.
2. $\{(2, 1), (3, 1)\}$ of $\mathbb{R}^2$. *Ans.* $\{\phi_1, \phi_2\}$ *given by* $\phi_1(x, y) = -x + 3y,\ \phi_2(x, y) = x - 2y$.
3. $\{(1, 0, -1), (-1, 1, 0), (0, 1, 1)\}$ of $\mathbb{R}^3$. *Ans.* $\{\phi_1, \phi_2, \phi_3\}$ *given by*

$$\phi_1(x, y, z) = \frac{1}{2}(x + y - z),\ \phi_2(x, y, z) = \frac{1}{2}(-x + y - z),\ \phi_3(x, y, z) = \frac{1}{2}(x + y + z).$$

The next result shows that the dual basis of a basis of V consists of the linear functionals on V that yield the coefficients for expressing a vector in V as a linear combination of the basis vectors.

1.1.15. Theorem. — *Suppose $\{v_1, \ldots, v_n\}$ is a basis of V and $\{\phi_1, \ldots, \phi_n\}$ is the dual basis. Then*

$$v = \phi_1(v)v_1 + \cdots + \phi_n(v)v_n \quad \forall v \in V. \tag{1.1.1}$$

Proof. Since $v \in V$ and $\{v_1, \ldots, v_n\}$ is a basis of V, there exist $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ s.t.

$$v = \lambda_1 v_1 + \cdots + \lambda_n v_n.$$

Then for $i = 1, \ldots, n$ we have $\phi_i(v) = \lambda_i$. □

The next result shows that the dual basis is indeed a basis of the dual space.

1.1.16. Theorem. — *Let V be a finite-dimensional $\mathbb{F}$-vector space. Then the dual space of a basis of V is a basis of V'.*

Proof. Suppose $\{v_1, \ldots, v_n\}$ is a basis of V and $\{\phi_1, \ldots, \phi_n\}$ is the dual basis. Then to show linear independence, suppose there exist $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$ s.t.

$$\lambda_1\phi_1 + \cdots + \lambda_n\phi_n = 0.$$

But for each $k = 1, \ldots, n$ we have

$$(\lambda_1\phi_1 + \cdots + \lambda_n\phi_n)(v_k) = \lambda_k.$$

Thus, $\lambda_1 = \cdots = \lambda_n = 0$. So $\{\phi_1, \ldots, \phi_n\}$ is a linearly independent set in V' with $n = \dim(V')$ elements. Hence, $\{\phi_1, \ldots, \phi_n\}$ is a basis of V'. □

1.1.17. Question. — Let $A, B \in O(n)$, [2] s.t. $\det(A) + \det(B) = 0$.
Show that $A + B \notin \mathrm{GL}_n(\mathbb{R})$. [3]

Solution. As A and B are real orthogonal and $\det(A) = -\det(B)$, we have

$$\det(A)\det(B) = -1.$$

Hence, $\det(A + B)$

$$= \det\left(A(B^T + A^T)B\right) = -\det(B^T + A^T) = -\det\left((B + A)^T\right) = -\det(B + A)$$

and the assertion follows. □

1.1.18. Question. — Let $A \in SO(2)$. [4] Show that there exists θ s.t. $A = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$.

Solution. In order for a matrix to be in $SO(2)$, it has to be (a) orthogonal, and (b) have determinant 1.

For (a) we need in particular that each row is a unit vector. Every unit vector has the form $(\cos\theta, \sin\theta)$ for some θ, so our matrix necessarily has the form

$$\begin{pmatrix} \cos\theta & \sin\theta \\ \cos\phi & \sin\phi \end{pmatrix}$$

In addition the rows must be perpendicular to each other. Our two unit vectors are perpendicular if and only if $\phi = \theta \pm \frac{\pi}{2}$, up to irrelevant multiples of 2π. By some basic trigonometric identities, this means that the possibilites are now

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}$$

The first of these always has determinant 1, so it is in $SO(2)$. The second has determinant -1, so it is not in $SO(2)$. □

[2] orthogonal group of order n i.e. orthogonal matrices of order n
[3] general linear group of order n i.e. invertible matrices of order n
[4] special orthogonal group of order 2 i.e. orthogonal matrices of order 2 and determinant 1

1.1.19. Question. — Let $A, B \in O(n)$ where n is odd s.t. $\det(A) = \det(B)$. Then show that $A - B \notin \mathrm{GL}_n(\mathbb{R})$.

1.1.20. Question. — Let $A \in \mathrm{GL}_n(\mathbb{R})$ s.t. sum of each row (resp. column) of A is r (resp. c). Show that the sum of each row (resp. column) of A^{-1} is r^{-1} (resp. c^{-1}).

1.1.21. Definition (Dual map or transpose map). — Let V_1, V_2 be two $\mathbb{F}$-vector spaces and $T \in \mathcal{L}(V_1, V_2)$. Then the **dual of T** is a linear map $T' \in \mathcal{L}(V_2', V_1')$ defined by

$$T'(\phi) = \phi \circ T \quad \forall \phi \in V_2'. \tag{1.1.2}$$

In particular, $T'(\phi) \in V_1'$.

1.1.22. Proposition. — *Dual map is a linear map.*

Proof.

$$\begin{aligned} T'(\lambda\phi_1 + \phi_2) &= (\lambda\phi_1 + \phi_2) \circ T \\ &= (\lambda\phi_1 \circ T) + (\phi_2 \circ T) \\ &= \lambda(\phi_1 \circ T) + (\phi_2 \circ T) \\ &= \lambda T'(\phi_1) + T'(\phi_2). \end{aligned}$$

□

1.1.23. Theorem (Properties of dual map). — *Let $T \in \mathcal{L}(V_1, V_2)$. Then,*

1. $(S + T)' = S' + T' \quad \forall S \in \mathcal{L}(V_1, V_2)$.
2. $(\lambda T)' = \lambda T' \quad \forall \lambda \in \mathbb{F}$.
3. $(ST)' = T'S' \quad \forall S \in \mathcal{L}(V_1, V_2)$.

Proof. 1.

$$\begin{aligned} (S+T)'(\phi) &= \phi \circ (S + T) \\ &= (\phi \circ S) + (\phi \circ T) \\ &= S'(\phi) + T'(\phi). \end{aligned}$$

2.

$$\begin{aligned} (\lambda T)'(\phi) &= \phi \circ (\lambda T) \\ &= \lambda(\phi \circ T) \\ &= \lambda T'(\phi). \end{aligned}$$

3.

$$\begin{aligned} (ST)'(\phi) &= \phi \circ (ST) \\ &= (\phi \circ S) \circ T \\ &= S'(\phi) \circ T \\ &= T'(S'(\phi)) \\ &= (T'S')(\phi). \end{aligned}$$

□

1.1.24. Definition (Annihilator). — Let V be an $\mathbb{F}$-vector space and W be a nonempty subset of V. Then the **annihilator** of W, denoted by W°, is defined by

$$W^\circ = \{\phi \in V' : \phi(w) = 0 \quad \forall w \in W\}. \tag{1.1.3}$$

1.1.25. Proposition. — *W° is a subspace of the dual space V'.*

Proof. Let $\phi_1, \phi_2 \in V'$ and $\lambda \in \mathbb{F}$.

$$\begin{aligned} \phi_1, \phi_2 \in W^\circ &\implies \lambda\phi_1(w) = \phi_2(w) = 0 \quad \forall w \in W \\ &\implies (\lambda\phi_1 + \phi_2)(w) = \lambda\phi_1(w) + \phi_2(w) = 0 \\ &\implies \lambda\phi_1 + \phi_2 \in W^\circ. \end{aligned}$$

□

1.1.26. Example. — In $\mathbb{R}^5$, with standard basis $\{e_1, e_2, e_3, e_4, e_5\}$, let the corresponding dual basis be $\{\phi_1, \phi_2, \phi_3, \phi_4, \phi_5\}$ of $(\mathbb{R}^5)'$.

Let $W = \text{span}(e_1, e_2) = \{(x_1, x_2, 0, 0, 0) : x_1, x_2 \in \mathbb{R}\}$. Then $W^\circ = \text{span}(\{\phi_3, \phi_4, \phi_5\})$.

$\dim(W) = 2, \dim(W^\circ) = 3$ so $\dim(W^\circ) = \dim(\mathbb{R}^5) - \dim(W)$. Every ϕ_i $(i = 1, \ldots, 5)$ is such that

$$\phi_i(x_1, x_2, x_3, x_4, x_5) = x_i.$$

Let $\phi \in \text{span}(\{\phi_3, \phi_4, \phi_5\})$ then there exist $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ s.t.

$$\phi = \lambda_1\phi_3 + \lambda_2\phi_4 + \lambda_3\phi_5.$$

If $\tilde{v} = (x_1, x_2, 0, 0, 0) \in W$ then

$$\begin{aligned} &\phi(\tilde{v}) \\ &= \lambda_1\phi_3(\tilde{v}) + \lambda_2\phi_4(\tilde{v}) + \lambda_3\phi_5(\tilde{v}) \\ &= 0 + 0 + 0 = 0 \in W^\circ. \end{aligned}$$

Thus, $\text{span}(\{\phi_3, \phi_4, \phi_5\}) \subseteq W^\circ$.

Let $\phi \in W^\circ$, then

$$\phi = \lambda_1\phi_3 + \lambda_2\phi_4 + \lambda_3\phi_5.$$

As $e_1 \in W$ and $\phi \in W^\circ$,

$$\begin{aligned} \phi(e_1) &= (\lambda_1\phi_3 + \lambda_2\phi_4 + \lambda_3\phi_5)(e_1) \\ &= \lambda_1\phi_3(e_1) + \lambda_2\phi_4(e_1) + \lambda_3\phi_5(e_1) \\ &= \lambda_1 \in \mathbb{R}. \end{aligned}$$

Thus, $W^\circ = \text{span}(\{\phi_3, \phi_4, \phi_5\})$.

1.1.27. Theorem. — *Let V be a finite-dimensional $\mathbb{F}$-vector space and W a subspace of V. Then*

$$\dim(W^\circ) = \dim(V) - \dim(W). \tag{1.1.4}$$

Proof. Let $i : W \to V$ be the inclusion map defined by $i(w) = w \quad \forall w \in W$.

Then i is a linear map and the dual map $i' : V' \to W'$ is a linear map too. By rank-nullity theorem, [5]

$$\dim(\ker i') + \dim(\operatorname{im} i') = \dim(V').$$

Now, $\ker i' = W^\circ$ so $\dim(\ker i') = \dim(W^\circ)$. Also $\dim(V') = \dim(V)$. Then,

$$\dim(W^\circ) + \dim(\operatorname{im} i') = \dim(V).$$

If $\phi \in W'$ then ϕ can be extended to a linear functional ψ on V. Now, $i'(\psi) = \psi \circ i = \phi$, which implies that $\phi \in \operatorname{im} i'$ so $W' \subseteq \operatorname{im} i'$. Also, $\operatorname{im} i' \subseteq W'$, so $\operatorname{im} i' = W'$.

Using $\dim(W) = \dim(W')$, we get $\dim(W^\circ) = \dim(V) - \dim(W)$. □

1.1.28. Theorem. — *Let V be a finite-dimensional $\mathbb{F}$-vector space and W a subspace of V. Then*

1. $W^\circ = \{0\} \iff W = V$.
2. $W^\circ = V' \iff W = \{0\}$.

Proof. Using Theorem 1.1.27,

$$\begin{aligned} &W^\circ = \{0\} \\ \iff &\dim(W^\circ) = 0 \\ \iff &\dim(V) = \dim(W) \\ \iff &V = W. \end{aligned}$$

Similarly,

$$\begin{aligned} &W^\circ = V' \\ \iff &\dim(W^\circ) = \dim(V') = \dim(V) \\ \iff &\dim(W) = \dim(V) - \dim(W^\circ) = 0 \\ \iff &W = \{0\}. \end{aligned}$$

□

[5] $\operatorname{rank}(T) = \dim(\operatorname{im} T)$, $\operatorname{nullity}(T) = \dim(\ker T)$.

1.1.29. Theorem. — *Let V_1, V_2 be two $\mathbb{F}$-vector spaces and $T \in \mathcal{L}(V_1, V_2)$. Then*

1. $\ker T' = (\operatorname{im} T)^\circ$.
2. $\dim(\ker T') = \dim(\ker T) + \dim(V_2) - \dim(V_1)$.

Proof. 1. By definition, $T' : \underset{\phi \mapsto \phi\circ T}{V_2' \to V_1'}$ so

$$\begin{aligned}\phi \in \ker T' &\implies T'(\phi) = 0\\ &\implies \phi \circ T = 0\\ &\implies (\phi \circ T)(v) = 0 \quad \forall v \in V_1\\ &\implies \phi \in (\operatorname{im} T)^\circ.\end{aligned}$$

So $\ker T' \subseteq (\operatorname{im} T)^\circ$. Now,

$$\begin{aligned}&\phi \in (\operatorname{im} T)^\circ\\ &\implies (\phi \circ T)(v) = 0 \quad \forall v \in V_1\\ &\implies \phi \circ T = 0\\ &\implies T'(\phi) = 0 \implies \phi \in \ker T'.\end{aligned}$$

Thus, $\ker T' = (\operatorname{im} T)^\circ$.

2.

$$\begin{aligned}\dim(\ker T') &= \dim((\operatorname{im} T)^\circ) = \dim(V_2) - \dim(\operatorname{im} T)\\ &= \dim(V_2) - (\dim(V_1) - \dim(\ker T)) \quad \text{(by rank-nullity theorem)}\\ &= \dim(\ker T) + \dim(V_2) - \dim(V_1).\end{aligned}$$

□

1.1.30. Theorem. — *Let V_1, V_2 be two $\mathbb{F}$-vector spaces and $T \in \mathcal{L}(V_1, V_2)$. Then T is surjective $\iff$ T' is injective.*

Proof. T is surjective so $\operatorname{im} T = V_2$. Now,

$$\begin{aligned}\operatorname{im} T = V_2 &\iff (\operatorname{im} T)^\circ = \{0\}\\ &\iff \ker T' = \{0\}\end{aligned}$$

which is equivalent to saying that T' is injective. □

1.1.31. Theorem. — *Let V_1, V_2 be two $\mathbb{F}$-vector spaces and $T \in \mathcal{L}(V_1, V_2)$. Then*

1. $\dim(\operatorname{im} T') = \dim(\operatorname{im} T)$.

2. $\operatorname{im} T' = (\ker T)^\circ$.

Proof.

$$\begin{aligned}\dim(\operatorname{im} T') &= \dim(V_2') - \dim(\ker T') \\ &= \dim(V_2') - \dim((\operatorname{im} T)^\circ) \\ &= \dim(V_2') - \dim(V_2) + \dim(\operatorname{im} T).\end{aligned}$$

Hence, $\dim(\operatorname{im} T') = \dim(\operatorname{im} T)$. For the second part, let $\phi \in \operatorname{im} T'$.

Then $\exists \psi \in V_2'$ s.t. $\phi = T'(\psi)$.

Now, if $v \in \ker T$ then

$$\begin{aligned}\phi(v) = 0 &\implies T'(\psi(v)) = 0 \\ &\implies (\psi \circ T)(v) = 0 \\ &\implies \psi(T(v)) = \psi(0) = 0.\end{aligned}$$

So $\phi \in (\ker T)^\circ$. Now,

$$\dim(\operatorname{im} T') = \dim(\operatorname{im} T) = \dim(V_1) - \dim(\ker T) = \dim((\ker T)^\circ).$$

□

Alternative proof of T injective $\iff$ T' surjective.

$$\begin{aligned}T \text{ is injective } &\iff \ker T = \{0\} \\ &\iff (\ker T)^\circ = V_1' \\ &\iff \operatorname{im} T' = V_1'.\end{aligned}$$

□

1.1.32. Definition (Matrix representation of dual maps). — Let V_1 be a finite-dimensional $\mathbb{F}$-vector space with an ordered basis $\beta_1 = \{v_1, \ldots, v_n\}$ along with its dual ordered basis $\beta_1' = \{\phi_1, \ldots, \phi_n\}$ of V_1'. Let V_2 be another finite-dimensional $\mathbb{F}$-vector space with an ordered basis $\beta_2 = \{u_1, \ldots, u_m\}$ along with its dual ordered basis $\beta_2' = \{\psi_1, \ldots, \psi_m\}$ of V_2'.

Let $T : \underset{\beta_1}{V_1} \to \underset{\beta_2}{V_2}$ have matrix $m(T)$ while $T' : \underset{\beta_2'}{V_2'} \to \underset{\beta_1'}{V_1'}$ has $m(T')$. Then

$$m(T') = (m(T))^T.$$

1.1.33. Remark. — This is why dual map is also called transpose map. Also, matrix multiplication is defined the way it is so that

$$m(ST) = m(S)m(T) \text{ holds.}$$

1.1.34. Theorem. — *Let V_1, V_2 be two $\mathbb{F}$-vector spaces and $T \in \mathcal{L}(V_1, V_2)$. Then*

$$m(T') = (m(T))^T. \tag{1.1.5}$$

1.1.35. Theorem. — *Let $A \in \mathbb{F}^{m\times n}$. Then column rank of A = row rank of A.*

Proof. Let $A = m(T)$ where $T \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$. Using dual map then,

$$\begin{aligned} &\text{column rank of } A \\ &= \text{column rank of } m(T) \\ &= \text{column rank of } m(T') \\ &= \text{column rank of } (m(T))^T \\ &= \text{row rank of } m(T). \end{aligned}$$

□

1.1.36. Remark. — Rank is invariant under row reduction since RREF is unique.

1.1.37. Definition (Double dual space)**.** — Let V be an $\mathbb{F}$-vector space and V' be the dual space of V. The dual space V'' of V' is the **double dual space** of V.

1.1.38. Theorem. — $V \cong V''$.

Proof. Define $T : V \to V''$. Then $T_v(\phi) = \phi(v) \quad \forall v \in V$. □

This is independent of choice of basis. If V is finite-dimensional, then $\dim(V) = \dim(V') - \dim(V'')$.

1.2. Invariant subspaces

1.2.1. Definition (T-invariant subspace)**.** — Let V be an $\mathbb{F}$-vector space and $T : V \to V$ be a linear operator. Let W be an $\mathbb{F}$-subspace of V. Then W is called ***T*-invariant** or **invariant under *T*** if $T(W) \subseteq W$.

1.2.2. Remark. — This is analogous to the notion of a characteristic subgroup H of G which is a subgroup s.t. $f(H) \subseteq H$ for every automorphism $f : G \to G$.

Inner automorphism is defined: $f_g(x) = gxg^{-1} \quad \forall x \in G$.

Normal subgroups are invariant under inner automorphisms, but characteristic subgroups are invariant under any automorphisms. Hence, every characteristic subgroup is normal but the converse is not true.

***T*-invariant subspaces.** — Thus, for a T-invariant subspace W, $T(w) \in W$ for all $w \in W$. The restriction of T to W is a linear operator over W.

Trivially, $\{0\}, V$ are always invariant subspaces of V.

Further for any linear operator $T \in \mathcal{L}(V)$,

$$\ker T = \{x \in V : T(v) = 0\}, \quad \operatorname{im} T = \{T(x) : x \in V\}$$

are invariant under T.

Let $x \in \ker T$. Then

$$T(x) = 0 \implies T(T(x)) = T(0) = 0 \implies T(x) \in \ker T.$$

So, $T(\ker T) \subseteq \ker T$. Similarly, $T(\operatorname{im} T) \subseteq \operatorname{im} T$.

If T is a nonsingular (invertible) linear operator, then $\ker T = \{0\}$ and $\operatorname{im} T = V$.

1.2.3. Question. — Let $V = \mathbb{R}^3$ over $\mathbb{R}$ and define $f_1, f_2, f_3 \in V'$ as follows:

$$f_1(x,y,z) = x - 2y, f_2(x,y,z) = x + y + z, f_3(x,y,z) = y - 3z.$$

Show that $\{f_1, f_2, f_3\}$ is a basis for V' and then find a basis of V for which it is the dual basis.

Solution.

$$\begin{aligned}
&\lambda_1 f_1 + \lambda_2 f_2 + \lambda_3 f_3 = 0 \\
&\implies \lambda_1(x-2y) + \lambda_2(x+y+z) + \lambda_3(y-3z) = 0 \\
&\implies (\lambda_1+\lambda_2)x + (\lambda_2+\lambda_3-2\lambda_1)y + (\lambda_2 - 3\lambda_3)z = 0 \\
&\implies -\lambda_2 = \lambda_1,\ 3\lambda_2 + \lambda_3 = 0,\ \lambda_2 = 3\lambda_3
\end{aligned}$$

solving which yields $\lambda_1 = \lambda_2 = \lambda_3 = 0$.

Let $B = \{v_1, v_2, v_3\}$ be a basis of V for which $\{f_1, f_2, f_3\}$ is a dual basis. Then, using Definition 1.1.11, we get that

$$B = \left\{ \left(\frac{2}{5}, \frac{-3}{10}, \frac{-1}{10}\right), \left(\frac{3}{5}, \frac{3}{10}, \frac{1}{10}\right), \left(\frac{1}{5}, \frac{1}{10}, \frac{-3}{10}\right) \right\}.$$

□

1.2.4. Question. — Let V be an $\mathbb{F}$-vector space, $T \in \mathcal{L}(V)$ and W a subspace of V.

Prove that W is T-invariant in $V \iff W^\circ$ is T' in V'.

Solution. Let W be T-invariant subspace of V. Then $T(W) \subseteq W$. So $T(w) \in W \quad \forall w \in W$. Now, $\forall w \in W, f \in W^\circ$,

$$\begin{aligned}
(T'(f))(w) &= (f \circ T)(w) \\
&= f(T(w)) \\
&= 0 \because f \in W^\circ, T(w) \in W.
\end{aligned}$$

So $T'(f) \in W^\circ$, i.e. $T'(W^\circ) \subseteq W^\circ$.

Conversely, suppose that $T'(W^\circ) \subseteq W^\circ$. Then $T'(f) \in W^\circ \quad \forall f \in W^\circ$.

If $T(w) \notin W$ for some $w \in W$ then $\exists f \in W^\circ : (f \circ T)(w) \neq 0 \implies T'(f) \notin W^\circ$. Absurdity. Hence, $T(w) \in W \quad \forall w \in W$. □

1.2.5. Question. — Let $V = \{a + bx : a, b \in \mathbb{R}\}$. (real polynomials with deg $\leqslant 1$.)

Find a basis $\{v_1, v_2\}$ of V which is dual to the basis $\{\phi_1, \phi_2\}$ of V' defined by

$$\phi_1(f(x)) = \int_0^1 f(x)dx, \quad \phi_2(f(x)) = \int_0^2 f(x)dx.$$

Solution. Let $v_1 = a + bx, v_2 = c + dx$.

Then $\phi_1(v_1) = 1, \phi_1(v_2) = 0$ while $\phi_2(v_1) = 0, \phi_2(v_2) = 1$.

Integrating and solving the the resulting system in each case, we get

$$v_1 = 2 - 2x,\ v_2 = \frac{-1}{2} + x.$$

□

1.2.6. Question. — Let V be an $\mathbb{F}$-vector space and W_1, W_2 be two subspaces of V. Prove that

1. $(W_1 + W_2)^\circ = W_1^\circ \cap W_2^\circ$.
2. $(W_1 \cap W_2)^\circ = W_1^\circ + W_2^\circ$.

1.2.7. Question. — Let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be a linear functional defined by

$$\phi(x, y) = x - 2y \quad \forall x, y \in \mathbb{R}.$$

Find the dual map for each of the following operators on $\mathbb{R}^2$,

1. $T(x, y) = (x, 0)$.
2. $T(x, y) = (y, x + y)$.
3. $T(x, y) = (2x - 3y, 5x + 2y)$.

Solution. $(T'(\phi))(x, y) = (\phi \circ T)(x, y) = \phi(T(x, y))$

1. $= \phi(x, 0) = x$.
2. $= \phi(y, x + y) = y - 2x - 2y = -2x - y$.
3. $= \phi(2x - 3y, 5x + 2y) = 2x - 3y - 10x - 4y = -8x - 7y$.

□

1.3. Diagonalisation of a linear operator

Eigenvalues and eigenvectors. —

For real square matrices $A \in \mathbb{R}^{n\times n}$ we had $\lambda \in \mathbb{R}$ is an eigenvalue of A if $Ax = \lambda x$ for some $x \neq 0$.	Let V be a finite-dimensional $\mathbb{F}$-vector space then $U = \{\lambda v : \lambda \in \mathbb{F}\} = \text{span}(v)$ is a subspace of V of dimension 1.

If U is invariant under a linear operator $T : V \to V$ then $T(u) \in U \quad \forall u \in U$. Thus, $T(u) = \lambda u$ for some scalar $\lambda \in \mathbb{F}$.

Then $\text{span}(u)$ is a 1-dimensional T-invariant subspace of U.

1.3.1. Definition (Eigenvalue of a linear operator)**.** — Let T be a linear operator on V and U be T-invariant in V. If $T(u) = \lambda u$ holds for some $u \in U : u \neq 0$ then λ is an **eigenvalue** of T.

1.3.2. Example. — Let $T \in \mathcal{L}(\mathbb{R}^3)$ s.t. $T(x, y, z) = (7x + 3z, 3x + 6y + 9z, -6y)$.
Then $T(3, 1, -1) = (18, 6, -6) = 6(3, 1, -1)$. So $\lambda = 6$ is an eigenvalue of T.

1.3.3. Theorem. — *Let V be a finite-dimensional $\mathbb{F}$-vector space. Then the following are equivalent for $T \in \mathcal{L}(V)$,*

1. *λ is an eigenvalue of T.*
2. *$T - \lambda I$ is not injective.*[6]
3. *$T - \lambda I$ is not surjective.*
4. *$T - \lambda I$ is not bijective.*
5. $\det(T - \lambda I) = 0$. *In other words, $T - \lambda I$ is singular.*

Proof. (1.) $\iff$ (2.):

$$\begin{aligned} &\exists v \neq 0 \text{ s.t. } T(v) = \lambda v \\ \iff & (T - \lambda I)(v) = 0 \\ \iff & \ker(T - \lambda I) \neq \{0\} \\ \iff & T \text{ is not injective.} \end{aligned}$$

(2.) $\iff$ (3.) $\iff$ (4.): Follows from the fact that
$T - \lambda I$ is injective $\iff$ $T - \lambda I$ is surjective $\iff$ $T - \lambda I$ is bijective.

(5.) $\iff$ (1.):

$$\begin{aligned} &\det(T - \lambda I) = 0 \\ \iff & (T - \lambda I)(v) = 0 \\ \iff & T(v) = \lambda v \end{aligned}$$

for some $v \neq 0$. □

[6] I is the identity operator $I : V \to V$ s.t. $x \mapsto x$.

1.3.4. Theorem. — *Let V be a finite-dimensional $\mathbb{F}$-vector space and $T \in \mathcal{L}(V)$. Let $v \neq 0$ be an eigenvector of T corresponding to an eigenvalue λ. Let $\lambda' \in \mathbb{F}$. Then $\lambda' v$ is also an eigenvector of T corresponding to the same eigenvalue λ.*

Proof.

$$\begin{aligned} T(\lambda' v) &= \lambda' T(v) \\ &= \lambda'(\lambda v) \\ &= (\lambda' \lambda) v \\ &= (\lambda \lambda') v \\ &= \lambda(\lambda' v). \end{aligned}$$

□

1.3.5. Example. — Let $T \in \mathcal{L}(\mathbb{F}^2)$ s.t. $T(x, y) = (-y, x)$.

For $\mathbb{F} = \mathbb{R}$, there is no $\lambda \in \mathbb{R} : T(v) = \lambda v$ so no real eigenvalues.

But for $\mathbb{F} = \mathbb{C}$, T has eigenvalues $\lambda = \pm i$. This is a reflection of the fact that $\mathbb{C}$ is *algebraically closed*.

Eigenvalues of a linear operator T always exist iff $\mathbb{F}$ is an algebraically closed field. The eigenvalues of T satisfy some polynomial

$$f(x) = a_0 + a_1 x + \cdots + a_n x^n, \text{ or}$$

$$f(T) = a_0 + a_1 T + \cdots + a_n T^n,$$

where $T^n = \underbrace{T \circ T \circ \cdots \circ T}_{n \text{ times}}$.

This polynomial is in fact the characteristic polynomial of T.

1.3.6. Definition (Characteristic polynomial)**.** — Let V be an $\mathbb{F}$-vector space and $T \in \mathcal{L}(V)$, then the polynomial

$$c_T(x) = \det(T - xI)$$

is called the **characteristic polynomial** of T.

1.3.7. Question. — Show that 0 is an eigenvalue of $T \in \mathcal{L}(V)$ iff T is singular.

Solution. $\det(T - (0)I) = 0 \iff \det(T) = 0.$ □

1.3.8. Question. — If λ is an eigenvalue of a nonsingular operator T, then show that λ^{-1} is an eigenvalue of T^{-1}.

Solution. There exists $v \neq 0$ s.t.

$$\begin{aligned} & T(v) = \lambda v \\ \implies & (T^{-1}T)(v) = T^{-1}(\lambda v) \\ & \qquad = \lambda T^{-1}(v) \\ & \implies v = \lambda T^{-1}(v) \\ & \implies \lambda^{-1} v = T^{-1}(v). \end{aligned}$$

□

Minimal polynomial. —

1.3.9. Theorem (Existence and uniqueness)**.** — *Let V be a finite-dimensional $\mathbb{F}$-vector space with* $\dim(V) = n$ *and $T \in \mathcal{L}(V)$. Then there exists a unique monic polynomial $m_T(x) \in \mathbb{F}[x]$ of smallest degree s.t. $m_T(T) = 0$. Furthermore,* $\deg(m_T(x)) \leqslant n$.

Proof. Consider $I[x] = \{f(x) \in \mathbb{F}[x] : f(T) = 0\}$. Then $I[x]$ is an ideal of $\mathbb{F}[x]$. We know that, as $\mathbb{F}$ is a field, $\mathbb{F}[x]$ must be PID. Thus, $I[x] = \langle m_1(x) \rangle$, i.e.,

$$f(x) = m_1(x)q(x), \quad q(x) \in \mathbb{F}[x].$$

Let, $m_1(x) = a_0 + a_1x + a_2x^2 + \cdots + a_kx^k \in \mathbb{F}[x]$ be the least degree polynomial in $I[x]$ s.t. $m_1(T) = 0$. Set $m_T(x) = a_k^{-1}m_1(x)$. Then $m_T(x)$ is a monic polynomial s.t.

$$m_T(T) = a_k^{-1}m_1(T) = 0.$$

Thus, $I[x] = \langle m_T(x) \rangle$. Let $m'_T(x)$ be another polynomial s.t. $m'_T(T) = 0$. Then

$$m_T(x) \mid m'_T(x),$$

so $m'_T(x) = cm_T(x)$ for some $c \in \mathbb{F}$. If $m'_T(x)$ is monic, then $c = 1$ so $m'_T(x) = m_T(x)$. ☐

1.3.10. Definition (Minimal polynomial)**.** — Let V be a finite-dimensional $\mathbb{F}$-vector space with $T \in \mathcal{L}(V)$. Then the **minimal polynomial** of T is the unique monic polynomial $m_T(x) \in \mathbb{F}[x]$ of smallest degree s.t. $m_T(T) = 0$.

1.3.11. Theorem. — *Let $m_T(x)$ be the minimal polynomial of $T \in \mathcal{L}(V)$. Then for any polynomial $f(x) \in \mathbb{F}[x] : f(T) = 0$,*

$$m_T(x) \mid f(x).$$

In particular, $m_T(x) \mid c_T(x)$.

So, $\deg(m_T(x)) \leqslant \deg(c_T(x))$.

Proof. By division algorithm, $\exists q(x), r(x) \in \mathbb{F}[x]$ s.t.

$$f(x) = q(x)m_T(x) + r(x).$$

Thus, $0 = f(T) = q(T)m_T(T) + r(T) = 0 + r(T)$. Then, we must have $r(x) = 0$ for all x as $\deg(m_T(x)) \leqslant \deg(r(x))$. Thus $f(x) = q(x)m_T(x)$ which means $m_T(x) \mid f(x)$.

By Cayley-Hamilton theorem, $c_T(T) = 0$. So $m_T(x) \mid c_T(x)$ as well. ☐

1.3.12. Theorem. — *Let V be a finite-dimensional $\mathbb{F}$-vector space with $T \in \mathcal{L}(V)$, $m_T(x)$ minimal polynomial of T.*

A scalar λ is an eigenvalue of T iff $m_T(\lambda) = 0$. Hence, the characteristicpolynomial and minimal polynomial have the same zeros.

Proof. $m_T(x) \mid c_T(x) \implies c_T(x) = q(x)m_T(x) \implies c_T(\lambda) = q(\lambda)m_T(\lambda)$. If λ is a zero of $m_T(x)$ then $c_T(\lambda) = q(\lambda)(0) = 0$ so λ is a zero of $c_T(x)$, i.e., an eigenvalue of T.

Conversely, let λ be an eigenvalue of T. Then $c_T(\lambda) = 0$. Let $x \neq 0$ be an eigenvector corresponding to λ,

$$0 = 0(x) = m_T(T)(x) = m_T(\lambda)x \implies m_T(\lambda) = 0.$$

So $m_T(x)$ and $c_T(x)$ have the same zeros. ☐

1.3.13. Theorem. — *Let V be a finite-dimensional $\mathbb{F}$-vector space with $T \in \mathcal{L}(V)$, $m_T(x)$ minimal polynomial of T, $c_T(x)$ characteristic polynomial of T. Suppose that $c_T(x)$ factors as*

$$c_T(x) = (x - \lambda_1)^{n_1}(x - \lambda_2)^{n_2} \cdots (x - \lambda_k)^{n_k} \tag{1.3.1}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct eigenvalues of T. Then there exist integers $m_1, m_2, \ldots, m_k$ s.t. $1 \leqslant m_i \leqslant n_j$ for $j = 1, 2, \ldots, k$ and

$$m_T(x) = (x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}. \tag{1.3.2}$$

Equivalent conditions for invertibility. — T is invertible $\iff$ 0 is not an eigenvalue of T $\iff$ 0 is not a root of $m_T(x)$ $\iff$ the constant term of $m_T(x)$ is nonzero.

Relationship between characteristic and minimal polynomial. — Let V be a finite-dimensional $\mathbb{F}$-vector space with $T \in \mathcal{L}(V)$, then $m_T(x), c_T(x)$ always exist and

$$m_T(x) \mid c_T(x).$$

In particular,

$$c_T(x) = m_T(x) \iff \left\{v, T(v), T^2(v), \ldots, T^{n-1}(v)\right\} \text{ is a basis of } V.$$

Let $c_T(x) = a_0 + a_1 x + a_2 x^2 + \cdots + x^n = m_T(x)$.

1.3.14. Definition (Companion matrix). — The **companion matrix** of the monic polynomial

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1}x^{n-1} + x^n$$

is defined as

$$C(p) = \begin{pmatrix} 0 & 0 & \cdots & 0 & \cdots & -a_0 \\ 1 & 0 & \cdots & 0 & \cdots & -a_1 \\ 0 & 1 & \cdots & 0 & \cdots & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & -a_{n-1} \end{pmatrix}. \tag{1.3.3}$$

1.3.15. Definition (Non-derogatory matrix). — $A \in \mathbb{F}^{n\times n}$ is **non-derogatory** if $c_A(x) = m_A(x)$.

1.3.16. Question. — Find the minimal polynomial of the operator $T \in \mathcal{L}(\mathbb{R}^2)$ s.t.

$$T(x, y) = (x, 0).$$

Solution. As $T(1,0) = (1,0)$, $T(0,1) = (0,0)$, the matrix wrt standard basis is

$$A = m(T) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

Then $c_A(x) = x(x-1) = m_T(x)$, i.e. $m_T(x) = x(x-1)$.

This is because $A(A - I) = 0$ so $m_T(x)$ has to divide $x(x-1)$ but $A \neq 0$ and $A - I \neq 0$ so $m_T(x) = x(x-1)$. □

1.3.17. Question. — Find the minimal polynomial of the operator $T \in \mathcal{L}(\mathbb{R}^2)$ s.t.

$$T(x, y) = \left(x + 4y, \frac{1}{2}x - y\right).$$

Solution. We get $c_T(x) = x^2 - 3$ from

$$m(T) = \begin{pmatrix} 1 & 4 \\ \frac{1}{2} & -1 \end{pmatrix}.$$

So $T^2 - 3I = 0$ needs to be satisfied by the minimal polynomial. But $T \neq \pm\sqrt{3}I$, so $m_T(x) = x^2 - 3$. □

Two matrices may have the same characteristic polynomial but different minimal polynomials.

1.3.18. Question. — Find the characteristic and minimal polynomial of

$$A = \begin{pmatrix} 4 & -2 & 2 \\ 6 & -3 & 4 \\ 3 & -2 & 3 \end{pmatrix} \text{ and } B = \begin{pmatrix} 3 & -2 & 2 \\ 4 & -4 & 6 \\ 2 & -3 & 5 \end{pmatrix}.$$

Solution. $c_A(x) = \det(A - xI) = (x-1)^2(x-2) = \det(B - xI) = c_B(x)$.
But $m_A(x) = (x-1)(x-2) \neq (x-1)^2(x-2) = m_B(x)$. □

1.3.19. Proposition. — $c_A(x) = c_{A^T}(x)$, $m_A(x) = m_{A^T}(x)$.

Proof. $m_A(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{n-1}x^{n-1} + x^n$ with $m_A(A) = 0$. But

$$m_A(A^T) = m_A(A)^T = 0^T = 0$$

as $p(A^T) = p(A)^T$ for any polynomial $p(x)$.
So $m_A(x)$ is the minimal polynomial of A and A^T. Consequently, $c_A(x) = c_{A^T}(x)$. □

1.3.20. Question. — Find a matrix A having minimal polynomial $x^3 - 8x^2 + 5x + 7$. Is A invertible ? Justify.

Solution. $A = \begin{pmatrix} 0 & 0 & -7 \\ 1 & 0 & -5 \\ 0 & 1 & 8 \end{pmatrix}$. Then A is invertible as the constant term of the minimal polynomial is $7 \neq 0$. □

1.3.21. Definition (Block diagonal matrix)**.** — A **block diagonal matrix** is of the form

$$M = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

where A and B are square matrices.
In this case, $m_M(x) = \operatorname{lcm}(m_A(x), m_B(x))$.

1.3.22. Question. — Find the minimal polynomial of

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

Solution. $M = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ where

$$A = \begin{pmatrix} 2 & 5 \\ 0 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 4 & 2 & 0 \\ 3 & 5 & 0 \\ 0 & 0 & 7 \end{pmatrix}.$$

$c_A(x) = (x-2)^2,\ c_B(x) = (x-2)(x-7)^2$. Now,

$$m_A(x) = (x-2)^2,\ m_B(x) = (x-2)(x-7).$$

So $m_M(x) = \operatorname{lcm}\big((x-2)^2,\ (x-2)(x-7)\big) = (x-2)^2(x-7)$. □

Eigenspace, algebraic multiplicity and geometric multiplicity. — Let V be a finite-dimensional $\mathbb{F}$-vector space and $T \in \mathcal{L}(V)$ with eigenvalue $\lambda \in \mathbb{F}$. Then

$$W_\lambda = \{v \in V : T(v) = \lambda v\} = \ker(T - \lambda I)$$

is a subspace of V, called the *eigenspace* of λ in V.

Furthermore, λ is an eigenvalue of $T \iff \ker(T-\lambda I) \neq \{0\} \iff W_\lambda \neq \{0\}$.

W_λ is a T-invariant subspace of V,

$$T(W_\lambda) \subseteq W_\lambda.$$

1.3.23. Definition (Algebraic and geometric multiplicity)**.** — The **algebraic multiplicity** of an eigenvalue λ of T is the greatest integer k s.t. $(x-\lambda)^k$ is a factor of the characteristic polynomial of T.

The **geometric multiplicity** of λ is the dimension of the eigenspace W_λ, $\dim(W_\lambda)$.

In general, given $T \in \mathcal{L}(V)$ with

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

$$m_T(x) = (x-\lambda_1)^{s_1}(x-\lambda_2)^{s_2}\cdots(x-\lambda_k)^{s_k}$$

we have $\dim(W_{\lambda_j}) \leqslant r_j$, i.e., geometric multiplicity $\leqslant$ algebraic multiplicity. If $\dim(W_j) = r_j$, then λ_j is a **regular eigenvalue** of T.

Similar matrices and operators. — $A, B \in \mathbb{F}^{n\times n}$ are similar if there exists $P \in \mathrm{GL}_n(\mathbb{F})$: $A = P^{-1}BP$. We can generalise this to linear operators.

1.3.24. Definition (Similar operators)**.** — Let V be a finite-dimensional $\mathbb{F}$-vector space with two linear operators $T_1, T_2 \in \mathcal{L}(V)$. Then T_1 is **similar** to T_2 if there exists an invertible linear operator $T_3 \in \mathcal{L}(V)$ s.t. $T_1 = T_3^{-1}T_2T_3$.

So, $T_1 \sim T_2 \implies T_2 = (T_3^{-1})^{-1}T_1T_3^{-1} = T_4^{-1}T_1T_4$.

1.3.25. Proposition. — $T_1 \sim T_2 \implies T_1^n \sim T_2^n \quad \forall n \in \mathbb{N}$.

Proof.

$$\begin{aligned} T_1^n &= \left(T_3^{-1}T_2T_3\right)^n \\ &= \left(T_3^{-1}T_2T_3\right)\cdots\left(T_3^{-1}T_2T_3\right) \\ &= T_3^{-1}\left(T_2\cdots T_2\right)T_3 \\ &= T_3^{-1}T_2^nT_3. \end{aligned}$$

□

1.3.26. Proposition. — *$T_1 \sim T_2$ and T_1 non-singular $\implies$ T_2 non-singular and $T_1^{-1} \sim T_2^{-1}$.*

Proof. $T_1 = T_3^{-1}T_2T_3 \implies T_1^{-1} = (T_3^{-1}T_2T_3)^{-1} = T_3^{-1}T_2^{-1}(T_3^{-1})^{-1} = T_3^{-1}T_2^{-1}T_3$. □

1.3.27. Proposition. — *T_2 non-singular $\implies$ $T_1T_2 \sim T_2T_1$.*

Proof. $T_2^{-1}(T_2T_1)T_2 = (T_2^{-1}T_2)T_1T_2 = T_1T_2$. □

1.3.28. Theorem. — *$T_1 \sim T_2 \implies c_{T_1}(x) = c_{T_2}(x)$. In other words, similar operators have the same eigenvalues.*

Proof. $T_1 = T_3^{-1}T_2T_3$. So

$$\begin{aligned} \det(T_1 - xI) &= \det(T_3^{-1}T_2T_3 - xT_3^{-1}T_3) \\ &= \det(T_3^{-1}(T_2 - xI)T_3) \\ &= \det(T_3^{-1})\det(T_2 - xI)\det(T_3) \\ &= \det(T_2 - xI). \end{aligned}$$

Thus, $c_{T_1}(x) = c_{T_2}(x)$. □

Two operators that have the same eigenvalues must have the same characteristic polynomials.

1.3.29. Example. — 1. Let $T_1, T_2 \in \mathcal{L}(\mathbb{R}^2)$ be defined by

$$T_1(x, y) = (3x + 6y, 3y), \; T_2(x, y) = (3x, 3y).$$

Then $c_{T_1}(x) = (x - 3)^2 = c_{T_2}(x)$.

2. Let $T_1, T_2 \in \mathcal{L}(\mathbb{R}^2)$ be defined by

$$T_1(x,y) = (x,y),\ T_2(x,y) = (x+y,y).$$

Then $c_{T_1}(x) = (x-1)^2 = c_{T_2}(x)$.

1.3.30. Definition (Diagonalisable operator). — Let V be a finite-dimensional $\mathbb{F}$-vector space. An operator $T \in \mathcal{L}(V)$ is **diagonalisable** over $\mathbb{F}$ if there exists an ordered basis B of V wrt which the matrix of T, $[m(T)]_B$, is diagonalisable.

In other words, $[m(T)]_B$ is a diagonal matrix, called the *diagonal form* of T.

1.3.31. Remark. — $A \in \mathbb{F}^{n\times n}$ is diagonalisable over $\mathbb{F}$ if $\exists P \in \mathrm{GL}_n(\mathbb{F}) : P^{-1}AP$ is a diagonal matrix.

$$P^{-1}[m(T)]_B P = [m(T)]_{B'}.$$

Suppose that $B = \{v_1, \ldots, v_n\}$ is a basis of V s.t.

$$[m(T)]_B = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

then

$$\begin{aligned} T(v_1) &= \lambda_1 v_1 + 0\cdot v_2 + \cdots + 0 \cdot v_n = \lambda_1 v_1 \\ &\vdots \\ T(v_n) &= 0\cdot v_1 + 0 \cdot v_2 + \cdots + \lambda_n v_n = \lambda_n v_n \end{aligned}$$

thus $\boxed{T_j(v_j) = \lambda_j v_j.}$

1.3.32. Question. — Check if $T \in \mathcal{L}(\mathbb{R}^2)$ is diagonalisable, where

$$T(x,y) = (41x + 7y, -20x + 74y).$$

Solution. $[m(T)]_B = \begin{pmatrix} 41 & 7 \\ -20 & 74 \end{pmatrix}$ wrt standard basis.

Let $B' = \{(1,4),(7,5)\}$. Then we get

$$[m(T)]_{B'} = \begin{pmatrix} 69 & 0 \\ 0 & 46 \end{pmatrix}.$$

Taking $P = \begin{pmatrix} 1 & 7 \\ 4 & 5 \end{pmatrix}$ we get $P^{-1}[m(T)]_B P = [m(T)]_{B'}$ so T is diagonalisable. □

Necessary and sufficient conditions for diagonalisability. — Let V be a finite-dimensional $\mathbb{F}$-vector space with $\dim(V) = n$.

1.3.33. Theorem. — *$T \in \mathcal{L}(V)$ is diagonalisable iff every eigenvalue λ of T is regular.*

1.3.34. Remark. — $\dim(W_\lambda) = \dim(\ker(T - \lambda I)) = n - \mathrm{rank}(T - \lambda I)$ by rank-nullity theorem.

1.3.35. Theorem. — *$T \in \mathcal{L}(V)$ is diagonalisable iff $m_T(x)$ is a product of distinct linear factors, i.e.,*

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

1.3.36. Theorem. — *Let $T \in \mathcal{L}(V)$ have k distinct eigenvalues $\lambda_1, \ldots, \lambda_k$. Then the following are equivalent:*

1. *T is diagonalisable.*
2. *$V = W_{\lambda_1} \oplus \cdots \oplus W_{\lambda_k}$, where $W_{\lambda_j} = \ker(T - \lambda_j I)$ is the eigenspace of λ_j.*
3. *The characteristic polynomial of T splits[7] over $\mathbb{F}$ and each eigenvalue is regular.*
4. *The minimal polynomial of T is $m_T(x) = (x - \lambda_1) \cdots (x - \lambda_k)$.*

1.3.37. Theorem (Enough eigenvalues $\implies$ diagonalisable)**.** — *If T has $n = \dim(V)$ distinct eigenvalues, then $T \in \mathcal{L}(V)$ is diagonalisable.*

The above condition is sufficient but not necessary - for example on $\mathbb{R}^3$ take the operator $T(x, y, z) = (6x, 6y, 7z)$. Then it has only two eigenvalues 6 and 7 and $\dim(\mathbb{R}^3) = 3$ but T is diagonalisable.

1.3.38. Example. — Let $T \in \mathcal{L}(\mathbb{R}^2)$ defined by

$$T(x, y) = (x + y, 2y).$$

Then $c_T(x) = (x - 1)(x - 2)$ and taking B as standard ordered basis,

$$[m(T)]_B = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$$

is diagonalisable.

1.3.39. Question. — Check whether the following are diagonalisable:

1. $T \in \mathcal{L}(\mathbb{R}^2)$ s.t. $T(x, y) = (x, y)$. *Ans. Diagonalisable.*
2. $T \in \mathcal{L}(\mathbb{R}^3)$ s.t. $T(x, y, z) = (y + z, z + x, x + y)$. *Ans. Diagonalisable.*
3. $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. *Ans. Not diagonalisable.*
4. $A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$. *Ans. Not diagonalisable.*
5. $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}$. *Ans. Diagonalisable.*
6. $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$. *Ans. Diagonalisable over $\mathbb{C}$, but not over $\mathbb{R}$.*

[7] For example, $(x^2 + 1)$ splits over $\mathbb{R}$ and $\mathbb{Q}(\sqrt{2})$ as $(x^2 + 1) = (x - \sqrt{2})(x + \sqrt{2})$.

1.3.40. Definition (Idempotent and nilpotent operators)**.** — $T \in \mathcal{L}(V)$ is

1. **idempotent** iff $T^2 = T$.
2. **nilpotent** iff $T^n = 0$ for some $n \in \mathbb{N}$.

1.3.41. Proposition. — *Idempotent operators are diagonalisable but nonzero nilpotent operators are not diagonalisable.*

Proof. For idempotent operator, $T^2 - T = 0$ so

$$m_T(x) = x,\ (x-1) \text{ or } x(x-1)$$

hence diagonalisable.

For nonzero nilpotent operator, $T^n = 0$ so $m_T(x) = x^m$, $m \leqslant n$ which is not a product of distinct linear factors. Hence not diagonalisable. □

1.3.42. Definition (Triangularisable)**.** — Let V be a finite-dimensional $\mathbb{F}$-vector space and $T \in \mathcal{L}(V)$, then T is **triangularisable** over $\mathbb{F}$ if there exists an ordered basis B of V wrt which the matrix of T is upper or lower triangular.

1.3.43. Remark. — $A \in \mathbb{F}^{n \times n}$ is triangularisable over $\mathbb{F}$ if $\exists P \in \mathrm{GL}_n(\mathbb{F}) : P^{-1}AP$ is upper triangular.

1.3.44. Example. — $T \in \mathcal{L}(\mathbb{R}^3)$ s.t.

$$T(x, y, z) = (2x + y, 5y + 3z, 8z)$$

is triangularisable wrt the standard ordered basis.

1.3.45. Theorem. — *The following are equivalent for $T \in \mathcal{L}(V)$:*

1. *T is triangularisable.*
2. *Every nonzero T-invariant subspace of V contains an eigenvector of T.*
3. *$c_T(x)$ splits over $\mathbb{F}$.*

1.3.46. Theorem. — *$T \in \mathcal{L}(V)$ is triangularisable iff $m_T(x)$ is a product of linear polynomials over $\mathbb{F}$ that are not necessarily distinct.*

1.3.47. Example. — $T \in \mathcal{L}(\mathbb{R}^3)$ s.t.

$$T(x, y, z) = (2x - y + 4z, y, 4x + z)$$

is triangularisable.

1.3.48. Question. — Is $T \in \mathcal{L}(\mathbb{R}^4)$ defined by

$$T(x, y, z, w) = (x - y, -2y + 3z - w, 4x - 5z, x + y - w)$$

is triangularisable ?

1.3.49. Definition (Simultaneously diagonalisable and triangularisable)**.** — Let V be a finite-dimensional $\mathbb{F}$-vector space and $T_1, T_2 \in \mathcal{L}(V)$ be diagonalisable (resp. triangularisable). Then T_1 and T_2 are **simultaneosuly diagonalisable** (resp. **simultaneosuly triangularisable**) iff $T_1T_2 = T_2T_1$.

1.3.50. Question. — Let $T \in \mathcal{L}(\mathbb{R}^3)$ s.t.

$$T(x, y, z) = (7x - y - 2z, -x + 7y + 2z, -2x + 2y + 10z).$$

Is T diagonalisable over $\mathbb{R}$? *Ans. Diagonalisable.*

1.3.51. Question. — Is $T \in \mathcal{L}(\mathbb{R}^3)$ defined by

$$T(x, y, z) = (-2x - y + z, 2x + y - 3z, -z)$$

diagonalisable or triangularisable over $\mathbb{R}$? *Ans. Triangularisable but not diagonalisable.*

1.3.52. Question. — Are the following diagonalisable or triangularisable over $\mathbb{R}$?

1. $T \in \mathcal{L}(\mathbb{R}^3)$ s.t.
$$T(x, y, z) = (x + z, 2y + z, -x + 3z).$$
Ans. Not diagonalisable but triangularisable.

2. $T \in \mathcal{L}(\mathbb{R}^3)$ s.t.
$$T(x, y, z) = (-z, x + z, y + z).$$
Ans. Not diagonalisable but triangularisable.

1.3.53. Question. — Let $A \in \mathbb{R}^{3\times 3}$. Prove that if $A \not\sim B$ where B is triangular over $\mathbb{R}$, then $A \sim D$ where D is diagonal over $\mathbb{C}$. Hence conclude that

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

is not triangularisable over $\mathbb{R}$ but diagonalisable over $\mathbb{C}$.

Solution. $c_T(x) = ax^3 + vx^2 + cx + d \in \mathbb{R}[x]$, which is of odd degree so at least one real root exists. Assume

$$c_T(x) = (x - \lambda_1)g(x),$$

then as A is not similar to any triangular matrix on $\mathbb{R}$, $g(x)$ does not split over $\mathbb{R}$ so it has no real roots. Thus

$$c_T(x) = (x - \lambda_1)(x - \lambda)(x - \overline{\lambda}) = m_T(x)$$

which is diagonalisable over $\mathbb{C}$.

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

has $c_A(x) = (x - 1)(x^2 + 1) = (x - 1)(x - i)(x + i)$ which splits over $\mathbb{C}$ but not $\mathbb{R}$ as $(x^2 + 1)$ cannot split over $\mathbb{R}$. Hence A is not triangularisable over $\mathbb{R}$ but diagonalisable over $\mathbb{C}$. □

1.3.54. Question. — Let $A = \begin{pmatrix} 0 & 1 & 0 \\ 2 & -2 & 2 \\ 2 & -3 & 2 \end{pmatrix}$. Is A similar to a triangular matrix over $\mathbb{R}$? Justify. *Ans. Triangularisable.*

1.3.55. Question. — Are A and B simultaneosuly diagonalisable over $\mathbb{R}$ if

$$A = \begin{pmatrix} 3 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 3 & -2 & -2 \\ 1 & 0 & -2 \\ 3 & -3 & -1 \end{pmatrix}?$$

Justify. *Ans. Simultaneously diagonalisable.*

1.3.56. Question. — Is every invertible matrix diagonalisable ? Is every diagonalisable matrix invertible ?

Solution. Counterexample to first claim is

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

Counterexample to second claim is

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

□

1.3.57. Question. — Are A and B simultaneosuly diagonalisable over $\mathbb{R}$ if

$$A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 2 \\ -2 & 1 \end{pmatrix}?$$

Justify.

1.3.58. Question. — Given an example of

1. two diagonalisable matrices A, B s.t. $A + B$ is not diagonalisable.
2. two triangularisable matrices A, B s.t. $A + B$ is not triangularisable.
3. two diagonalisable matrices A, B s.t. AB is not diagonalisable.
4. two triangularisable matrices A, B s.t. AB is not triangularisable.

Solution. 1.

$$A = \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix} \text{ and } B = \begin{pmatrix} -1 & 0 \\ 0 & 2 \end{pmatrix}.$$

2.

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} -1 & 0 \\ -1 & -1 \end{pmatrix}.$$

3.

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

4.

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

□

Chapter 2.

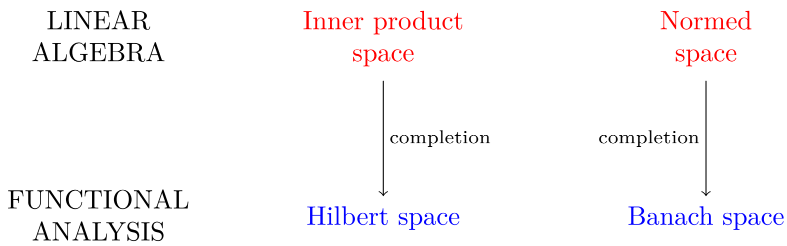
Inner product spaces and orthogonality

2

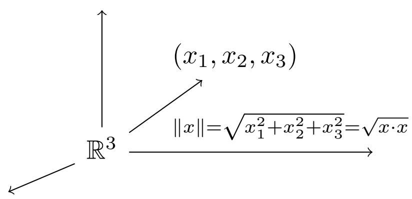
2.1. Inner product and norm

2.1.1. Overview. — *Inner products* and *norms* are functions that let us measure distances or equip our vector space with some distance function (metric) or topology. They are generalisations of the Euclidean inner product (dot product) and Euclidean norm (distance), and it only makes sense to consider vector spaces equipped with such functions over $\mathbb{R}$ and $\mathbb{C}$ due to issues with ordering. Hence we will assume $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$ unless otherwise stated.

It is nicer to do analysis on such a space if every Cauchy sequence in it is convergent to some limit in that space - *completeness*. A complete inner product space is called a *Hilbert space* whereas a complete normed space is called a *Banach space*. Hilbert and Banach spaces form the core of the subject of functional analysis and the mathematics of quantum theory.



2.1.2. Euclidean norm and inner product. — On the real line $\mathbb{R}$ the distance of x from 0 is measured simply by the absolute value $|x|$. In the $\mathbb{R}^2$ plane we measure distance from origin by $\|x\| = \|(x_1, x_2)\| = \sqrt{x_1^2 + x_2^2} = \sqrt{x \cdot x}$, where $x \cdot y = x_1 y_1 + x_2 y_2$ is the inner (dot) product. Generalisation to $\mathbb{R}^3$ and the Euclidean n-space $\mathbb{R}^n$ is straightforward.



2.1.3. Definition (Euclidean inner product and norm)**.** — For any two vectors $x, y \in \mathbb{R}^n$, the **dot product** or **Euclidean inner product** of $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ is defined by $x \cdot y = x_1 y_1 + \cdots + x_n y_n \in \mathbb{R}$. The **Euclidean norm** of x is $\|x\| = \sqrt{x \cdot x} \in \mathbb{R}$.

2.1.4. Remark. — The Euclidean inner product is scalar-valued, hence it's often called a *scalar product*. Also note that

$$x \cdot x = x_1^2 + \cdots + x_n^2 = \|x\|^2.$$

We want to generalise the Euclidean norm and inner product to any V over $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Observe that the Euclidean inner product satisfies the following properties:

1. $x \cdot x \geqslant 0 \; \forall x \in \mathbb{R}^n$ with equality iff $x = 0 \in \mathbb{R}^n$.
2. $\forall y \in \mathbb{R}^n$ we have that $x \mapsto x \cdot y$ is a linear map from $\mathbb{R}^n$ to $\mathbb{R}$ (in fact a linear functional).
3. $x \cdot y = y \cdot x \; \forall x, y \in \mathbb{R}^n$.

This is only for $\mathbb{R}^n$ as an $\mathbb{R}$-vector space. What about $\mathbb{C}^n$ as a $\mathbb{C}$-vector space ? We use the property of complex conjugation:

$$\forall z \in \mathbb{C} : z\overline{z} = (\Re(z) + i\Im(z))(\Re(z) - i\Im(z)) = \Re(z)^2 + \Im(z)^2 = |z|^2.$$

2.1.5. Definition. — Let $z = (z_1, \cdots, z_n), w = (w_1, \cdots, w_n) \in \mathbb{C}^n$. Then the *complex inner product* and *norm* are, respectively,

$$z \cdot w = z_1\overline{w}_1 + \cdots + z_n\overline{w}_n, \quad \|z\| = \sqrt{z \cdot z} = \sqrt{z_1\overline{z}_1 + \cdots + z_n\overline{z}_n}.$$

2.1.6. Remark. — The complex inner product differs from the Euclidean in that symmetry (Remark 2.1.4, property 3.) no longer holds; instead, we have *conjugate symmetry* $z \cdot w = \overline{w \cdot z}$.

From this point onwards, unless otherwise stated, $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

2.1.7. Definition (Inner product)**.** — Let V be an $\mathbb{F}$-vector space. An **inner product** on V is a map

$$\langle \cdot, \cdot \rangle : \underset{(x,y)}{V \times V} \underset{\longmapsto \langle x,y \rangle}{\longrightarrow \mathbb{F}}$$

satisfying the following properties:

1. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle \quad \forall x, y, z \in V$ (additivity in first component).
2. $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle \quad \forall \lambda \in \mathbb{F}, x, y \in V$ (homogeneity or linearity in first component).
3. $\langle x, y \rangle = \overline{\langle y, x \rangle} \quad \forall x, y \in V$ (conjugate symmetry).
4. $\langle x, x \rangle > 0 \quad \forall x \in V$ (positivity).
   In particular, $\langle x, x \rangle = 0 \iff x = 0 \in V$ (definiteness).

We then say $\langle \cdot, \cdot \rangle$ is an inner product defined on V, making it into an **inner product space**.

2.1.8. Remark. — Inner product is needed to define angles and distances in a vector space. $\mathbb{F} = \mathbb{R} \implies$ real inner product space, $\mathbb{F} = \mathbb{C} \implies$ complex inner product space. Physicists often define linearity in the second component instead, whereas most mathematicians define inner products with linearity in the first component.

As $\overline{a} = a \; \forall a \in \mathbb{R}$, when $\mathbb{F} = \mathbb{R}$ we have $\langle x, y \rangle = \langle y, x \rangle$.

If $\langle \cdot, \cdot \rangle$ is an inner product on V, then $\varepsilon \langle \cdot, \cdot \rangle$ is also an inner product on V for every $\varepsilon > 0$, $\varepsilon \in \mathbb{R}$.

If $\mathbb{F} = \mathbb{R}$ then $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ is a bilinear map (linear in both components):

$$\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$$
$$\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle .$$

If $\mathbb{F} = \mathbb{C}$ then $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$, but $\langle x, \lambda y \rangle = \overline{\lambda} \langle x, y \rangle$.

2.1.9. Example. — Let $V = \mathbb{F}^n$ ($\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$) and $x, y \in \mathbb{F}^n : x = (x_1, \cdots, x_n), y = (y_1, \cdots, y_n)$. Then

$$\langle x, y \rangle = x_1 \overline{y}_1 + \cdots + x_n \overline{y}_n \tag{2.1.1}$$

is the *usual*, *standard* or *Euclidean inner product* on $\mathbb{F}^n$. For $\lambda_1, \ldots, \lambda_n > 0$, $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$,

$$\langle x, y \rangle = \lambda_1 x_1 \overline{y}_1 + \cdots + \lambda_n x_n \overline{y}_n \tag{2.1.2}$$

is also an inner product on $\mathbb{F}^n$.

2.1.10. Example. — Let $V = C[a, b]$ be the $\mathbb{R}$-vector space of all continuous functions $f \in \mathbb{R}^{[a,b]}$ $(-\infty < a < b < \infty)$. Then an inner product on V is given by

$$\langle f, g \rangle = \int_a^b f(x) g(x) dx. \tag{2.1.3}$$

In particular, if $V = C[-1, 1]$ then

$$\langle f, g \rangle = \int_{-1}^1 f(x) g(x) dx \tag{2.1.4}$$

and if $V = C[0, 1]$ then

$$\langle f, g \rangle = \int_0^1 f(x) g(x) dx. \tag{2.1.5}$$

If $f \in \mathbb{C}^{[a,b]}$ instead, then

$$\langle f, g \rangle = \int_a^b f(x) \overline{g(x)} dx. \tag{2.1.6}$$

2.1.11. Example. — Let $V = \mathbb{F}^{n \times n}$ (the space of square matrices of order n over $\mathbb{F}$). Let $A = (a_{ij})_{n \times n} \in \mathbb{F}^{n \times n}$. Then the *adjoint* of A is defined as the conjugate transpose of A,

$$A^* = \overline{(a_{ij})}_{n \times n}^T = \overline{A}^T = \overline{A^T} = \overline{(a_{ji})}_{n \times n}. \tag{2.1.7}$$

If

$$A = \begin{pmatrix} 1 + 2i & 2 + 3i \\ 3 + 4i & 4 + 5i \end{pmatrix},$$

then

$$\overline{A} = \begin{pmatrix} 1 - 2i & 2 - 3i \\ 3 - 4i & 4 - 5i \end{pmatrix}$$

so that

$$A^* = \begin{pmatrix} 1 - 2i & 3 - 4i \\ 2 - 3i & 4 - 5i \end{pmatrix}.$$

We can define an inner product on $\mathbb{F}^{n \times n}$ by

$$\langle A, B \rangle = \operatorname{tr}(B^* A). \tag{2.1.8}$$

2.1.12. Example. — Let $V = \mathcal{P}(\mathbb{R})$. Then

$$\langle f, g \rangle_1 = f(0)g(0) + \int_{-1}^{1} f'(x)g'(x)dx \tag{2.1.9}$$

and

$$\langle f, g \rangle_2 = \int_{0}^{\infty} f(x)g(x)e^{-x}dx \tag{2.1.10}$$

are two possible inner products on $\mathcal{P}(\mathbb{R})$.

2.1.13. Question. — Does there exist a non-Euclidean inner product on $\mathbb{R}^2$?

2.1.14. Example. — For any $x, y \in \mathbb{R}^2$, we define

$$\langle x, y \rangle = ax_1y_1 + c(x_1y_2 + x_2y_1) + bx_2y_2 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T \begin{pmatrix} a & c \\ c & b \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x^T Ay \tag{2.1.11}$$

where $a, b, c \in \mathbb{R}$ are arbitrary. Then $\langle x, y \rangle = x^T Ay$ is an inner product on $\mathbb{R}^2$ iff $a > 0$ and $ab - c^2 > 0$, i.e. $\det(A) > 0$. For example, $\langle x, y \rangle = x_1y_1 + 2(x_1y_2 + x_2y_1) + 5x_2y_2$.

2.1.15. Theorem (Basic properties of inner product). — *Let V be an $\mathbb{F}$-inner product space. Then the following hold:*

1. *For all $y \in V$ (fixed), there is a linear functional $V \to \mathbb{F}$ defined by $x \mapsto \langle x, y \rangle$.*
2. $\langle x, 0 \rangle = \langle 0, x \rangle = 0 \quad \forall x \in V.$
3. $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle \quad \forall x, y, z \in V.$
4. $\langle x, \lambda y \rangle = \overline{\lambda} \langle x, y \rangle \quad \forall \lambda \in \mathbb{F}, x, y \in V.$

Proof.

1. Let $\lambda \in \mathbb{F}, a, b \in V$. Then

$$\lambda a + b \mapsto \langle \lambda a + b, y \rangle = \langle \lambda a, y \rangle + \langle b, y \rangle = \lambda \langle a, y \rangle + \langle b, y \rangle .$$

2. Observe that

$$\langle 0, x \rangle = \langle x, x \rangle + \langle -x, x \rangle = \langle x, x \rangle - \langle x, x \rangle = 0.$$

Hence, $\langle x, 0 \rangle = \overline{\langle 0, x \rangle} = \overline{0} = 0$.

3. $\langle x, y + z \rangle = \overline{\langle y + z, x \rangle} = \overline{\langle y, x \rangle + \langle z, x \rangle} = \overline{\langle y, x \rangle} + \overline{\langle z, x \rangle} = \langle x, y \rangle + \langle x, z \rangle .$
4. $\langle x, \lambda y \rangle = \overline{\langle \lambda y, x \rangle} = \overline{\lambda} \overline{\langle y, x \rangle} = \overline{\lambda} \langle x, y \rangle .$

□

2.1.16. Definition (Norm). — Let V be an $\mathbb{F}$-inner product space. For any $x \in V$ the **norm** of x is defined by

$$\|x\| = \sqrt{\langle x, x \rangle} = (\langle x, x \rangle)^{\frac{1}{2}} . \tag{2.1.12}$$

In other words, $\|x\|^2 = \langle x, x \rangle$.

2.1.17. Example. — Let $V = \mathbb{F}^n$. Then

$$\forall x \in \mathbb{F}^n : \|x\| = \sqrt{|x_1|^2 + \cdots + |x_n|^2}.$$

2.1.18. Example. — Let $V = C[-1,1]$. Then

$$\forall f \in C[-1,1] : \|f\| = \sqrt{\langle f, f\rangle} = \sqrt{\int_{-1}^{1} f^2(x)dx}.$$

2.1.19. Theorem (Basic properties of norm)**.** — *Let V be an $\mathbb{F}$-inner product space, then for any $x \in V, \lambda \in \mathbb{F}$ we have the following:*

1. $\|x\| = 0 \iff x = 0$.
2. $\|\lambda x\| = |\lambda|\|x\|$.

Proof. 1. $\|x\| = 0 \iff \sqrt{\langle x, x\rangle} = 0 \iff x = 0$.

2. $\|\lambda x\| = \sqrt{\langle \lambda x, \lambda x\rangle} = \sqrt{\lambda\overline{\lambda}}\sqrt{\langle x, x\rangle} = \sqrt{|\lambda|^2}\sqrt{\langle x, x\rangle} = |\lambda|\|x\|$.

□

2.1.20. Theorem. — *Let V be an $\mathbb{F}$-inner product space. Then for any $x, y \in V$ we have the following inequalities:*

1. $|\langle x, y\rangle| \leqslant \|x\|\|y\|$. *(Cauchy-Schwarz inequality)*
   Equality holds $\iff$ x and y are linearly dependent, i.e. $x = \lambda y$ for some $\lambda \in \mathbb{F}$.
2. $\|x + y\| \leqslant \|x\| + \|y\|$. *(Triangle inequality)*
3. $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$. *(Parallelogram law)*

Proof. 1. If $y = 0$ then this is trivial. So let $y \neq 0$. Then $\langle y, y\rangle \neq 0$. Now for any $\lambda \in \mathbb{F}$,

$$\begin{aligned}
0 &\leqslant \|x - \lambda y\|^2 \\
&= \langle x - \lambda y, x - \lambda y\rangle \\
&= \langle x, x - \lambda y\rangle - \lambda \langle y, x - \lambda y\rangle \\
&= \langle x, x\rangle - \overline{\lambda}\langle x, y\rangle - \lambda \langle y, x\rangle + \lambda\overline{\lambda}\langle y, y\rangle .
\end{aligned}$$

Setting $\lambda = \dfrac{\langle x, y\rangle}{\langle y, y\rangle}$, we get

$$0 \leqslant \|x\|^2 - \frac{\langle x, y\rangle}{\|y\|^2}. \tag{2.1.13}$$

Equality holds in (2.1.13) iff $x = \lambda y$ for some $\lambda \in \mathbb{F}$.

2.

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

3.

$$\begin{aligned}\|x+y\|^2+\|x-y\|^2 &= \langle x+y, x+y\rangle + \langle x-y, x-y\rangle\\ &= 2\langle x,x\rangle + 2\langle y,y\rangle\\ &= 2\|x\|^2+2\|y\|^2.\end{aligned}$$

□

2.1.21. Example. — If $V=\mathbb{R}^2$ then $\langle x,y\rangle = \|x\|\|y\|\cos(\theta)$ and

$$\begin{aligned}|\langle x,y\rangle| &= \|x\|\|y\|\cos(\theta)\\ &\leqslant \|x\|\|y\|.\end{aligned}$$

2.1.22. Theorem. — *Every inner product space is a metric space.*

Proof. Let V be an inner product space over $\mathbb{F}$. Then $d: V\times V\to\mathbb{R}$ defined by

$$d(x,y)=\|x-y\| \quad \forall x,y\in V$$

defines a metric space (V,d). □

2.1.23. Remark. — Every finite dimensional inner product space is a complete metric space, hence a complete inner product space. Such a complete inner product space is called a Hilbert space.

2.2. Orthogonality and orthonormality

2.2.1. Definition (Orthogonal)**.** — Let V be an $\mathbb{F}$-inner product space. Then two vectors $x,y\in V,\ x\neq y$ are said to be **orthogonal** if $\langle x,y\rangle=0$.

A nonempty subset $S\subseteq V$ is orthogonal if $\langle x,y\rangle = 0 \quad \forall x,y\in S,\ x\neq y$.

2.2.2. Definition (Orthonormal)**.** — Let V be an $\mathbb{F}$-inner product space. A nonempty subset $S\subseteq V$ is **orthonormal** if S is orthogonal and $\|x\|=1\ \forall x\in S$.

2.2.3. Remark. — If $\{x_1,\ldots,x_n\}$ is an orthonormal set of vectors in V, then

$$\langle x_j,x_k\rangle = \begin{cases}1 & \text{if } j=k\\ 0 & \text{if } j\neq k\end{cases} \quad \because \langle x_k,x_k\rangle = \|x_k\|^2. \tag{2.2.1}$$

2.2.4. Example. — Let $V=\mathbb{F}^n$, then the standard basis $\{e_1,\ldots,e_n\}$ is a natural example of an orthonormal set.

2.2.5. Example. —

$$\left\{\left(\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}},\frac{1}{\sqrt{3}}\right),\left(-\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}},0\right),\left(\frac{1}{\sqrt{6}},\frac{1}{\sqrt{6}},-\frac{2}{\sqrt{6}}\right)\right\}$$

is an orthonormal set of vectors in $\mathbb{R}^3$.

2.2.6. Example. — Consider $\mathbb{F}^{n\times n} = \mathbb{R}^{n\times n}$ or $\mathbb{C}^{n\times n}$. Take the elementary matrices $\{E_{ij} : i, j = 1, \ldots, n\}$, where

$$E_{ij} = \begin{pmatrix} 0 & & & & \\ & \ddots & & \text{\Large 0} & \\ & & 1 & & \\ & \text{\Large 0} & & \ddots & \\ & & & & 0 \end{pmatrix}$$

where the ith row, jth column is 1 and every other entry is 0. With $\langle A, B\rangle = \operatorname{tr}(B^*A)$, we get that $\{E_{ij} : i, j = 1, \ldots, n\}$ is an orthonormal set over $\mathbb{F}^{n\times n}$.

2.2.7. Example. — Let $V = \mathbb{R}^2$. Let $x = (x_1, x_2),\ y = (y_1, y_2) \in \mathbb{R}^2$. Then

$$\langle x, y\rangle = 0 \iff x_1y_1 + x_2y_2 = 0.$$

If we take a straight line from origin through x then this straight line is perpendicular to the straight line passing through y.

2.2.8. Theorem. — *In an $\mathbb{F}$-inner product space V, orthogonal subsets*[1] *of nonzero elements of V are linearly independent.*

Proof. Let $S \subseteq V$ be an orthogonal subset of nonzero vectors in V. Let $x_1, \ldots, x_n \in S$ such that

$$\lambda_1x_1 + \cdots + \lambda_nx_n = 0 \tag{2.2.2}$$

for some $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$. Then for $1 \leqslant i \leqslant n$,

$$\begin{aligned} \lambda_i\|x_i\|^2 &= \lambda_i \langle x_i, x_i\rangle \\ &= \lambda_1 \langle x_1, x_i\rangle + \cdots + \lambda_i \langle x_i, x_i\rangle + \cdots + \lambda_n \langle x_n, x_i\rangle \\ &= \langle \lambda_1x_1 + \cdots + \lambda_nx_n, x_i\rangle \\ &= \langle 0, x_i\rangle = 0. \end{aligned}$$

As $\|x_i\| = 1$, we have $\lambda_i = 0$ for $1 \leqslant i \leqslant n$. Hence S is linearly independent. □

2.2.9. Theorem (Bessel's inequality)**.** — *Let V be an $\mathbb{F}$-inner product space. Let $\{e_1, \ldots, e_n\}$ be an orthonormal subset of V. Then for all $x \in V$, we have*

$$|\langle x, e_1\rangle|^2 + \cdots + |\langle x, e_n\rangle|^2 \leqslant \|x\|^2. \tag{2.2.3}$$

Proof. Let $y = x - (\langle x, e_1\rangle e_1 + \cdots + \langle x, e_n\rangle e_n)$. Then

$$\|y\|^2 = \langle y, y\rangle = \|x\|^2 - \left(|\langle x, e_1\rangle|^2 + \cdots + |\langle x, e_n\rangle|^2\right) \geqslant 0. \tag{2.2.4}$$

Hence, $|\langle x, e_1\rangle|^2 + \cdots + |\langle x, e_n\rangle|^2 \leqslant \|x\|^2$. □

2.2.10. Definition (Orthonormal basis)**.** — Let V be an $\mathbb{F}$-inner product space. A subset B of V is called an **orthonormal basis** of V if B is a basis for V and an orthonormal set.

[1] If V is infinite dimensional then we consider all possible finite subsets of S, in that case all finite subsets of S are linearly independent.

2.2.11. Example. — The standard basis $\{e_1, \ldots, e_n\}$ of $\mathbb{F}^n$ over $\mathbb{F}$ is an orthonormal basis.

2.2.12. Example. — The standard basis $\{E_{ij} : i, j = 1, \ldots, n\}$ of $\mathbb{F}^{n\times n}$ over $\mathbb{F}$ is an orthonormal basis.

2.2.13. Example. — $\left\{\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right), \left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}\right), \left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right), \left(-\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right)\right\}$ is an orthonormal basis of $\mathbb{F}^4$ over $\mathbb{F}$.

2.2.14. Remark. — $\underset{\text{space}}{V} \longleftrightarrow \underset{basis}{B}$ For a general basis $B = \{x_1, \ldots, x_n\}$, verifying linear independence is not easy. Showing $\lambda_k = 0$ for all λ_k just from $\sum_{k=1}^n \lambda_k x_k = 0$ is nontrivial in general.

For an orthonormal basis, the implication $\sum_{k=1}^n \lambda_k x_k = v \implies \langle v, x_k\rangle = \lambda_k$ is immediate. So $\sum_{k=1}^n \lambda_k x_k = 0 \implies \lambda_k = \langle 0, x_k\rangle = 0$. This is the advantage of orthonormal bases over usual bases.

2.2.15. Theorem. — *Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of an $\mathbb{F}$-inner product space V. If $x, y \in V$ then*

1. $x = \langle x, e_1\rangle e_1 + \cdots + \langle x_n, e_n\rangle e_n$.
2. $\|x\|^2 = |\langle x, e_1\rangle|^2 + \cdots + |\langle x, e_n\rangle|^2$. *(Bessel's identity)*
3. $\langle x, y\rangle = \langle x, e_1\rangle \langle e_1, y\rangle + \cdots + \langle x, e_n\rangle \langle e_n, y\rangle$ *(Parseval's identity)*

   $= \langle x, e_1\rangle \overline{\langle y, e_1\rangle} + \cdots + \langle x, e_n\rangle \overline{\langle y, e_n\rangle}$.

Proof. 1. Let $x \in V : x = \lambda_1 e_1 + \cdots + \lambda_n e_n \implies \langle x, e_i\rangle = \lambda_i$.

So, $x = \langle x, e_1\rangle e_1 + \cdots + \langle x_n, e_n\rangle e_n$.

2. For any $x \in V$,

$$\begin{aligned}\|x\|^2 = \langle x, x\rangle &= \langle x, \langle x, e_1\rangle e_1 + \cdots + \langle x_n, e_n\rangle e_n\rangle \\ &= \langle x, \langle x, e_1\rangle e_1\rangle + \cdots + \langle x, \langle x, e_n\rangle e_n\rangle \\ &= \overline{\langle x, e_1\rangle} \langle x, e_1\rangle + \cdots + \overline{\langle x, e_n\rangle} \langle x, e_n\rangle \\ &= |\langle x, e_1\rangle|^2 + \cdots + |\langle x, e_n\rangle|^2 .\end{aligned}$$

3. Let $x, y \in V$. Then

$$\begin{aligned}\langle x, y\rangle &= \langle x, \langle y, e_1\rangle e_1 + \cdots + \langle y, e_n\rangle e_n\rangle \\ &= \langle x, \langle y, e_1\rangle e_1\rangle + \cdots + \langle x, \langle y, e_n\rangle e_n\rangle \\ &= \overline{\langle y, e_1\rangle} \langle x, e_1\rangle + \cdots + \overline{\langle y, e_n\rangle} \langle x, e_n\rangle \\ &= \langle x, e_1\rangle \overline{\langle y, e_1\rangle} + \cdots + \langle x, e_n\rangle \overline{\langle y, e_n\rangle} \\ &= \langle x, e_1\rangle \langle e_1, y\rangle + \cdots + \langle x, e_n\rangle \langle e_n, y\rangle .\end{aligned}$$

□

2.2.16. Theorem. — *Every finite dimensional inner product space has an orthonormal basis.*

Proof. Construction by Gram-Schmidt orthonormalisation (Theorem 2.2.17). □

2.2.17. Theorem (Gram-Schmidt orthonormalisation process)**.** — *Let V be an $\mathbb{F}$-inner product space and $\{v_1, \ldots, v_m\}$ be a linearly independent set of vectors in V. Let $f_1 = v_1$. For $2 \leqslant k \leqslant m$, define f_k inductively by*

$$f_k = v_k - \frac{\langle v_k, f_1 \rangle}{\|f_1\|^2} f_1 - \cdots - \frac{\langle v_k, f_{k-1} \rangle}{\|f_{k-1}\|^2} f_{k-1}. \tag{2.2.5}$$

For $1 \leqslant k \leqslant m$, let $e_k = \dfrac{f_k}{\|f_k\|}$. Then $\{e_1, \ldots, e_m\}$ is an orthonormal set of vectors in V s.t.

$$\operatorname{span}(v_1, \ldots, v_k) = \operatorname{span}(e_1, \ldots, e_k) \quad \textit{for } 1 \leqslant k \leqslant m.$$

2.2.18. Example. — Consider the basis $B = \{(0,1,1), (1,0,1), (1,1,0)\}$ of $\mathbb{R}^3$.
Let $f_1 = (0,1,1) \implies \|f_1\| = \sqrt{2}$. Now,

$$\begin{aligned} f_2 &= (1,0,1) - \frac{\langle (1,0,1), (0,1,1) \rangle}{2}(0,1,1) \\ &= (1,0,1) - \left(0, \frac{1}{2}, \frac{1}{2}\right) \\ &= \left(1, -\frac{1}{2}, \frac{1}{2}\right) \implies \|f_2\| = \sqrt{\frac{3}{2}}. \\ f_3 &= (1,1,0) - \frac{\langle (1,1,0), (0,1,1) \rangle}{2}(0,1,1) - \frac{\left\langle (1,1,0), \left(1, -\frac{1}{2}, \frac{1}{2}\right) \right\rangle}{\frac{3}{2}} \left(1, -\frac{1}{2}, \frac{1}{2}\right) \\ &= (1,1,0) - \left(0, \frac{1}{2}, \frac{1}{2}\right) - \left(\frac{1}{3}, -\frac{1}{6}, \frac{1}{6}\right) \\ &= \left(\frac{2}{3}, \frac{2}{3}, -\frac{2}{3}\right) \implies \|f_3\| = \frac{2}{\sqrt{3}}. \end{aligned}$$

Then we have

$$e_1 = \frac{(0,1,1)}{\sqrt{2}}, e_2 = \frac{\left(1, -\frac{1}{2}, \frac{1}{2}\right)}{\sqrt{\frac{3}{2}}}, e_3 = \frac{\left(\frac{2}{3}, \frac{2}{3}, -\frac{2}{3}\right)}{\frac{2}{\sqrt{3}}}.$$

After simplifications, $e_1 = \frac{1}{\sqrt{2}}(0,1,1), e_2 = \frac{1}{\sqrt{6}}(2,-1,1), e_3 = \frac{1}{\sqrt{3}}(1,1,-1)$.

Hence, $\left\{ \dfrac{1}{\sqrt{2}}(0,1,1), \dfrac{1}{\sqrt{6}}(2,-1,1), \dfrac{1}{\sqrt{3}}(1,1,-1) \right\}$ is an orthonormal basis of $\mathbb{R}^3$.

2.2.19. Example. — Use Gram-Schmidt orthonormalisation process to construct an orthonormal basis from the following bases of $\mathbb{R}^3$:

1. $\{(1,1,1), (0,1,1), (0,0,1)\}$ *Ans.* $\left\{ \frac{1}{\sqrt{3}}(1,1,1), \frac{1}{\sqrt{6}}(-2,1,1), \frac{1}{\sqrt{2}}(0,-1,1) \right\}$.
2. $\{(2,1,2), (4,1,0), (3,1,-1)\}$ *Ans.* $\left\{ \frac{1}{3}(2,1,2), \frac{1}{\sqrt{2}}(1,0,-1), \frac{1}{3\sqrt{2}}(-1,4,1) \right\}$.

2.2.20. Example. — Take the polynomial inner product space

$$V = \mathcal{P}_2(\mathbb{R}) = \{f \in \mathbb{R}[x] : \deg f \leqslant 2\}.$$

For all $f, g \in \mathcal{P}_2(\mathbb{R})$,

$$\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)dx.$$

Then $\{1, x, x^2\}$ is a basis of V. Taking $f_1 = 1$, we then have $\|f_1\| = \sqrt{\int_{-1}^{1} dx} = \sqrt{2}$, so

$$\begin{aligned}
f_2 &= x - \frac{\langle x, f_1\rangle}{\|f_1\|^2} f_1 = x - \frac{1}{2}\int_{-1}^{1} x dx = x,\\
\|f_2\| &= \sqrt{\int_{-1}^{1} x^2 dx} = \sqrt{\frac{2}{3}}.\\
f_3 &= x^2 - \frac{\langle x^2, f_1\rangle}{\|f_1\|^2} f_1 - \frac{\langle x^2, f_2\rangle}{\|f_2\|^2} f_2\\
&= x^2 - \frac{1}{2}\int_{-1}^{1} x^2 dx - \frac{3}{2}x\int_{-1}^{1} x^3 dx\\
&= x^2 - \frac{1}{3},\\
\|f_3\| &= \sqrt{\int_{-1}^{1}\left(x^5 - \frac{2}{3}x^2 + \frac{1}{9}\right)dx} = \sqrt{\frac{8}{45}}.
\end{aligned}$$

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 V.

2.2.21. Example. — Let $V = \mathcal{P}(\mathbb{R})$. This is an infinite dimensional vector space. A basis for V is given by $\{1, x, x^2, \dots, x^n, \dots\}$, and we want to find an orthononormal basis of the form

$$\{\alpha_0 P_0(x), \alpha_2 P_2(x), \dots, \alpha_n P_n(x), \dots\},$$

where $P_n(1) = 1$ and $\alpha_n = \|P_n(x)\|^{-1} = \left(\int_{-1}^{1}(P_n(x))^2 dx\right)^{-\frac{1}{2}}$ for all n. Let $P_n(x) = \frac{f_n(x)}{f_n(1)}$.

$$\begin{aligned}
\text{Then } f_0 = 1 &\implies \|f_0\|^2 = 2, \quad P_0 = 1,\\
f_1 &= x - \frac{1}{2}\int_{-1}^{1} x dx = x \implies \|f_1\|^2 = \frac{2}{3}, \quad P_1 = x,\\
f_2 &= x^2 - \frac{1}{2}\int_{-1}^{1} x^2 dx - \frac{3}{2}x\int_{-1}^{1} x^3 dx\\
&= x^2 - \frac{1}{3} \implies \|f_2\|^2 = \frac{8}{45},\\
P_2(x) &= \frac{f_2(x)}{f_2(1)} = \frac{3}{2}\left(x^2 - \frac{1}{3}\right) = \frac{1}{2}(3x^2 - 1).
\end{aligned}$$

We get $\alpha_n = \|P_n(x)\|^{-1} = \left(\int_{-1}^{1}(P_n(x))^2 dx\right)^{-\frac{1}{2}} = \left(\sqrt{\frac{2}{2n+1}}\right)^{-1}$, and the polynomials $P_0(x) = 1,\ P_1(x) = x,\ P_2(x) = \frac{1}{2}(3x^2 - 1), \dots$ are exactly the *Legendre polynomials.*

2.2.22. Theorem (Extension theorem)**.** — *Let V be an $\mathbb{F}$-inner product space of dimension n. If $\{x_1, \ldots, x_k\}$ is an orthonormal subset of V then there exist $x_{k+1}, \ldots, x_n \in V$ s.t. $\{x_1, \ldots, x_n\}$ is an orthonormal basis of V.*

2.3. Inner product space isomorphisms

2.3.1. Definition. — Let V and W be two $\mathbb{F}$-inner product spaces. Then a mapping $f : V \to W$ is an **inner product space isomorphism** if it is a vector space isomorphism that preserves inner products, i.e.

$$\langle f(x), f(y) \rangle = \langle x, y \rangle \quad \forall x, y \in V.$$

2.3.2. Theorem. — *Let V and W be two $\mathbb{F}$-inner product spaces. Let $\{e_1, \ldots, e_n\}$ be an orthonormal basis of V. Then $f : V \to W$ is an inner product space isomorphism iff $\{f(e_1), \ldots, f(e_n)\}$ is an orthonormal basis of W.*

2.3.3. Theorem. — *Let V be a finite dimensional $\mathbb{F}$-inner product space and $T : V \to V$ be a linear operator. Then T has an upper triangular matrix with respect to some orthonormal basis of V iff the minimal polynomial of T is of the form*

$$m_T(x) = (x - \lambda_1) \cdots (x - \lambda_m), \quad \lambda_1, \ldots, \lambda_m \in \mathbb{F}.$$

2.3.4. Theorem (Schur's theorem)**.** — *Every linear operator on a finite-dimensional complex inner product space has an upper-triangular matrix with respect to some orthonormal basis.*

2.3.5. Theorem (Riesz representation theorem)**.** — *Let V be a finite-dimensional $\mathbb{F}$-inner product space and $\phi : V \to \mathbb{F}$ be a linear functional. Then there exists a unique vector $v \in V$ s.t.*

$$\phi(u) = \langle u, v \rangle \quad \forall u \in V.$$

2.3.6. Example. — Let $V = C[-1, 1]$, which is an infinite dimensional real inner product space. Then

$$\langle f, g \rangle = \int_{-1}^{1} f(x) g(x) dx.$$

We define $\phi : C[-1, 1] \to \mathbb{R}$ by $\phi(f) = f(0)$. Then there is no $g \in C[-1, 1] : \phi(f) = \langle f, g \rangle$. Hence, the Riesz representation theorem (Theorem 2.3.5) may fail on an infinite-dimensional inner product space.

2.4. Orthogonal complement

Non-uniqueness of complementary subspaces. — In an $\mathbb{F}$-vector space V, every subspace has a complement. A complement always exists, but is not necessarily unique. Consider the real plane $V = \mathbb{R}^2$. Any line W passing through the origin in $\mathbb{R}^2$ forms a subspace. Any line V passing through the origin and not parallel to W then is a complement of W.

However, we can define a specific kind of complementary subspace in an inner product space (or, more generally, in any space equipped with a bilinear form) that both exists and is unique.

2.4.1. Definition (Orthogonal complement). — The **orthogonal complement** $W^\perp$ of a subset W of an F-inner product space V is defined by

$$W^\perp = \{v \in V : \langle u, v\rangle = 0 \quad \forall u \in W\}. \tag{2.4.1}$$

2.4.2. Example. — In $V = \mathbb{R}^2$ we have the following three subspaces:
$X = \{(x, 0) : x \in \mathbb{R}\}$, $Y = \{(0, y) : y \in \mathbb{R}\}$ and $D = \{(x, x) : x \in \mathbb{R}\}$. Then

$$\mathbb{R}^2 = X \oplus Y = X \oplus D = Y \oplus D$$

so all three are complementary subspaces. But the orthogonal complement of any one subspace is unique: for example, $\mathbb{R}^2 = X \oplus X^\perp, \quad X^\perp = Y$.

2.4.3. Example. — In $V = \mathbb{R}^3$ consider the subset $W = \{(2, 3, 5)\}$. Then its orthogonal complement is the following plane passing through origin,

$$W^\perp = \{(x, y, z) \in \mathbb{R}^3 : 2x + 3y + 5z = 0\}.$$

The orthogonal complement of $U = W^\perp$ is then $U^\perp = \{(2t, 3t, 5t) : t \in \mathbb{R}\}$.

2.4.4. Example. — In $V = \mathbb{R}^5$ consider $W = \{(a, b, 0, 0, 0) : a, b \in \mathbb{R}\}$.
Then $W^\perp = \{(0, 0, x, y, z) : x, y, z \in \mathbb{R}\}$.

2.4.5. Example. — $C[-1, 1] = W \oplus W^\perp$, where

$$W = \{f(x) : f(-x) = f(x)\}, \quad W^\perp = \{f(x) : f(-x) = -f(x)\}.$$

2.4.6. Example. — $\mathbb{R}^{n\times n} = S \oplus S^\perp$, where

$$S = \{A \in \mathbb{R}^{n\times n} : A = A^T\}, \quad S^\perp = \{A \in \mathbb{R}^{n\times n} : A = -A^T\}.$$

2.4.7. Theorem (Properties of orthogonal complement). — *Let V be an $\mathbb{F}$-inner product space.*

1. *If $W \subseteq V$ then $W^\perp$ is a subspace of V.*
2. $\{0\}^\perp = V, \quad V^\perp = \{0\}$.
3. *If $W \subseteq V$ then $W \cap W^\perp \subseteq \{0\}$.*
4. *If $\dim(V) < \infty$ and W_1, W_2 are subspaces of V, then*
   a) $W_1 \subseteq W_2 \implies W_2^\perp \subseteq W_1^\perp$.
   b) $(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp$.
   c) $(W_1 \cap W_2)^\perp = W_1^\perp + W_2^\perp$.

Proof. 1. $w \in W \implies \langle w, 0\rangle = 0 \implies 0 \in W^\perp$. Also,

$$\begin{aligned} x, y \in W^\perp, w \in W, \lambda \in \mathbb{F} \implies & \langle w, x + \lambda y\rangle \\ &= \langle w, x\rangle + \overline{\lambda}\,\langle x, y\rangle \\ &= 0 + 0 = 0 \implies x + \lambda y \in W^\perp. \end{aligned}$$

2. Let $v \in V \implies \langle 0, v\rangle = 0 \implies v \in \{0\}^\perp$. So, $V \subseteq \{0\}^\perp$, but $\{0\}^\perp \subseteq V$ as $\{0\} \subseteq V$. Hence, $\{0\}^\perp = V$.

   Similarly, let $v \in V^\perp \implies \langle v, v\rangle = 0 \implies v = 0$. Hence, $V^\perp = \{0\}$.

3. Let $w \in W \cap W^\perp \implies \langle w, w\rangle = 0 \implies w = 0 \implies w \in \{0\} \implies W \cap W^\perp \subseteq \{0\}$.

4. a) Let $w \in W_2^\perp \implies \langle u, w\rangle = 0 \quad \forall u \in W_2$. As $W_1 \subseteq W_2$, this implies that $\langle u, w\rangle = 0 \quad \forall u \in W_1 \implies w \in W_1^\perp$. Hence $W_2^\perp \subseteq W_1$.

   b) As $W_1^\perp$ and $W_2^\perp$ are subspaces of V, $0 \in W_1^\perp \cap W_2^\perp$.

   Let $v \in (W_1 + W_2)^\perp \implies \langle w_1 + w_2, v\rangle = 0 \quad \forall w_1 \in W_1, w_2 \in W_2$.

   Then, setting $w_1 = 0$ we get $\langle w_2, v\rangle = 0 \implies v \in W_2^\perp$ and setting $w_2 = 0$ we get $\langle w_1, v\rangle = 0 \implies v \in W_1 \perp$. So $(W_1 + W_2)^\perp \subseteq W_1^\perp \cap W_2^\perp$.

   Let $v \in W_1^\perp \cap W_2^\perp \implies \langle w_1, v\rangle = \langle w_2, v\rangle = 0 \quad \forall w_1 \in W_1, w_2 \in W_2$. Thus,

$$\langle w_1 + w_2, v\rangle = \langle w_1, v\rangle + \langle w_2, v\rangle = 0 + 0 = 0 \quad \forall w_1 \in W_1, w_2 \in W_2.$$

   Hence, $v \in (W_1 + W_2)^\perp$. So $(W_1 + W_2)^\perp = W_1^\perp \cap W_2^\perp$.

   c) In a finite dimensional space $(W^\perp)^\perp = W$. (cf. Theorem 2.4.10) So,

$$\begin{aligned}&(W_1 \cap W_2)^\perp \\ &= \big((W_1^\perp)^\perp \cap (W_2^\perp)^\perp\big)^\perp \\ &= \big((W_1^\perp + W_2^\perp)^\perp\big)^\perp \\ &= W_1^\perp + W_2^\perp.\end{aligned}$$

□

2.4.8. Theorem. — *Let V be an $\mathbb{F}$-inner product space and W be a finite dimensional subspace of V. Then $V = W \oplus W^\perp$.*

Proof. Let $v \in V$. As W is a finite dimensional subspace of V, there exists an orthonormal basis, say, $\{e_1, \dots, e_m\}$ of W. Also, $\langle v - \langle v, e_1\rangle e_1 - \cdots - \langle v, e_m\rangle e_m, e_k\rangle = \langle v, e_k\rangle - \langle v, e_k\rangle = 0$ for $k = 1, \dots, m$.

Then

$$\begin{aligned}v = {}&\langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m \in W \\ &+ v - \langle v, e_1\rangle e_1 - \cdots - \langle v, e_m\rangle e_m \in W^\perp.\end{aligned}$$

Thus $V = W + W^\perp$. From Theorem 2.4.7 (3.), we know that $W \cap W^\perp \subseteq \{0\}$. But as W and $W^\perp$ are subspaces of V, $\{0\} \subseteq W \cap W^\perp$.

Hence, $V = W + W^\perp, \quad W \cap W^\perp = \{0\} \implies V = W \oplus W^\perp$. □

2.4.9. Example. — Let $W = \{f \in C[-1, 1] : f(0) = 0\}$. Then $C[-1, 1] \neq W \oplus W^\perp$ as W is infinite-dimensional.

Indeed, here $W^\perp = \{0\}$.

2.4.10. Theorem. — *Let V be an $\mathbb{F}$-inner product space. Let V be finite dimensional and W be a subspace of V. Then*

$$\dim(W^{\perp}) = \dim(V) - \dim(W), \quad (W^{\perp})^{\perp} = W.$$

Proof. As $V = W \oplus W^{\perp}$, we have

$$\dim V = \dim W + \dim W^{\perp} \implies \dim W^{\perp} = \dim V - \dim W.$$

Further, $\dim (W^{\perp})^{\perp} = \dim V - \dim W^{\perp} = \dim W$. □

Alternative proof. Let $x \in W$, then $\langle x, y\rangle = 0 \quad \forall y \in W^{\perp} \implies x \in (W^{\perp})^{\perp}$. So $W \subseteq (W^{\perp})^{\perp}$.

Let $v \in (W^{\perp})^{\perp}$, then $v = u + w, \quad u \in W, w \in W^{\perp}$. We have $v - u = w \in W^{\perp}$.

Now, $u \in W \subseteq (W^{\perp})^{\perp}$ and $v \in (W^{\perp})^{\perp}$ so $v - u \in (W^{\perp})^{\perp}$. So $v - u \in W^{\perp} \cap (W^{\perp})^{\perp}$. Then by Theorem 2.4.7 (3.), we have $v - u = 0 \implies v = u \in W$. So $(W^{\perp})^{\perp} \subseteq W$.

Hence, $W = (W^{\perp})^{\perp}$. □

2.4.11. Corollary. — *If W is a finite dimensional subspace of V, then*

$$W^{\perp} = \{0\} \iff W = V.$$

Proof. $W = (W^{\perp})^{\perp} = \{0\}^{\perp} = V$. □

Projection operator. — Let V be an $\mathbb{F}$-vector space that is the direct sum of two subspaces W_1 and W_2,

$$V = W_1 \oplus W_2.$$

Then every vector $v \in V$ can be uniquely written as a sum $v = w_1 + w_2$, where $w_1 \in W_1, w_2 \in W_2$. We think of w_1 as the *projection* of v on W_1 along W_2.

2.4.12. Definition (Projection operator)**.** — The **projection operator** on W_1 along W_2 is the linear operator P on V defined by

$$P(v) = w_1 \quad \forall v = w_1 + w_2 \in V = W_1 \oplus W_2. \tag{2.4.2}$$

2.4.13. Theorem. — *A linear operator $P : V \to V$ is a projection iff P is an idempotent operator on V, i.e. $P^2 = P$.*

Proof. Let $P : V \to V$ be a projection operator on a subspace W and let $v \in V$. Then $P^2(v) = P(v)$ as $P(v) \in W$ and $P(w) = w \quad \forall w \in W$.

Now let $P : V \to V$ be a linear operator such that $P^2 = P$. Let $v \in V$, then

$$v = \underbrace{P(v)}_{\in \mathrm{im}(P)} + \underbrace{(v - P(v))}_{\in \ker(P)}.$$

Thus, $V = \mathrm{im}(P) \oplus \ker(P)$. By definition, $P(V) = \mathrm{im}(P)$. Also, for any $w \in \mathrm{im}(P)$ we have $P(w) = P(P(v))$ for some $v \in V$.

But $P^2 = P$, so $w = P(w) = P(v)$. Hence, P is a projection operator on $\mathrm{im}(P)$. □

2.4.14. Remark. — If $V = W_1 \oplus W_2$ and $P : V \to V$ is the projection on W_1 along W_2, then $\mathrm{im}(P) = W_1$ and $\ker(P) = W_2$.

2.4.15. Corollary. — *If V is a finite dimensional $\mathbb{F}$-vector space and P_1, P_2 are two projection operators on V, then*

$$\operatorname{im}(P_1) = \operatorname{im}(P_2) \iff P_1P_2 = P_2 \quad \& \quad P_2P_1 = P_1.$$

2.4.16. Question. — Is it possible to have operators $P_1, P_2 : V \to V$ with $\ker(P_1) = \ker(P_2)$ and $\operatorname{im}(P_1) \neq \operatorname{im}(P_2)$? Justify. Is the converse true ?

2.4.17. Question. — Is it possible to have projection operators P_1, P_2 s.t. $P_1P_2 = 0$ but $P_2P_1 \neq 0$?

2.4.18. Question. — Let U be the subspace of $\mathbb{R}^3$ defined by $U = \{(x, x, 0) : x \in \mathbb{R}\}$. Find a subspace W of $\mathbb{R}^3$ s.t. $\mathbb{R}^3 = U \oplus W$. Is W unique ? Justify.

Find a projection $P_1 : \mathbb{R}^3 \to \mathbb{R}^3$ s.t. $\operatorname{im}(P_1) = \{0\}$ and $\ker(P_1) = W$. Find also a projection $P_2 : \mathbb{R}^3 \to \mathbb{R}^3$ s.t. $\operatorname{im}(P_2) = W$ and $\ker(P_2) = W$.

2.4.19. Question. — Let V be an $\mathbb{F}$-vector space and $P : V \to V$ a projection operator. If $f(x) \in \mathbb{F}[x]$ then show that $f(P) = aI + bP$ for some $a, b \in \mathbb{F}$. What are a, b in terms of coefficients of $f(x)$?

2.4.20. Question. — Let $V = W_1 \oplus W_2$. Show that a linear operator $P : V \to V$ is a projection operator iff $I - P$ is a projection operator.

Further, show that if P is the projection on W_1 along W_2 then $I - P$ is the projection on W_2 along W_1.

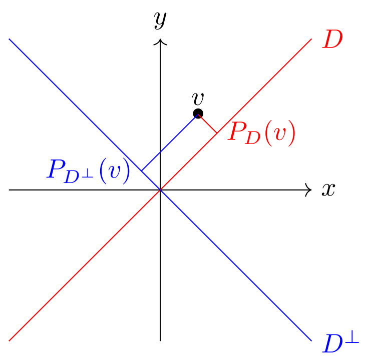
2.4.21. Question. — If T is a linear operator on V s.t. $T^2(I - T) = T(I - T)^2$ then show that T is a projection operator on V.

2.4.22. Question. — Let P_1, P_2 be two projection operators on V and $P_1P_2 = P_2P_1$. Then show that $P_1 + P_2 - P_1P_2$ is a projection on V.

If $\operatorname{char}(\mathbb{F}) \neq 2$ then show that $P_1 + P_2$ is a projection operator on V iff $P_1P_2 = P_2P_1 = 0$.

Orthogonal projection. — We know $V = W \oplus W^\perp$ for any finite dimensional subspace W of V. For example,

$$\begin{aligned}
&\mathbb{R}^2 = X \oplus Y = D \oplus D^\perp,\\
&X = \{(x, 0) : x \in \mathbb{R}\}, \quad X^\perp = Y = \{(0, y) : y \in \mathbb{R}\},\\
&D = \{(x, x) : x \in \mathbb{R}\}, \quad D^\perp = \{(-x, x) : x \in \mathbb{R}\}.
\end{aligned}$$



2.4.23. Definition (Orthogonal projection). — Let W be a finite dimensional subspace of V. The **orthogonal projection** of V onto W is the linear operator $P_W : V \to V$ s.t.

$$P_W(v) = u, \quad \forall v = u + w \in V = W \oplus W^\perp.$$

2.4.24. Theorem (Properties of orthogonal projection). — *Suppose W is a finite dimensional subspace of V. Then*

1. $P_W \in \mathcal{L}(V)$ *(linear operator on V).*
2. $P_W(u) = u \quad \forall u \in W$.
3. $P_W(w) = 0 \quad \forall w \in W^\perp$.
4. $\operatorname{im}(P_W) = W, \quad \ker(P_W) = W^\perp$.
5. $v - P_W(v) \in W^\perp \quad \forall v \in V$.
6. $P_W^2 = P_W$.
7. $\|P_W(v)\| \leqslant \|v\| \quad \forall v \in V$.
8. *If $\{e_1, \ldots, e_m\}$ is an orthonormal basis of W and $v \in V$ then*

$$P_W(v) = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_m\rangle e_m$$

Minimisation problem. — The following problem often arises, the remarkable simplicity of the solution to which has led to many important applications of inner product spaces outside of pure maths:

Given a subspace W of V and a point $v \in V$, find a point $u \in U$ such that $\|v - u\|$ is as small as possible. The next result shows that $u = P_W(v)$ is the unique solution of this minimization problem.

2.4.25. Theorem. — *Let V be an $\mathbb{F}$-inner product space. Let W be a finite dimensional subspace of V and $v \in V$. Then*

$$\|v - P_W(v)\| \leqslant \|v - w\| \quad \forall w \in W \tag{2.4.3}$$

with equality iff $P_W(v) = w$.

Proof.

$$\begin{aligned}
&\|v - P_W(v)\|^2 \\
&\leqslant \|v - P_W(v)\|^2 + \|P_W(v) - w\|^2 \\
&= \|v - P_W(v) + P_W(v) - w\|^2 \quad \text{(Pythagoras theorem)} \\
&= \|v - w\|^2.
\end{aligned}$$

Equality holds iff $\|P_W(v) - w\| = 0 \iff w = P_W(v)$. □

2.5. Self-adjoint linear operators

2.5.1. Definition (Adjoint)**.** — Let V and W be two finite dimensional $\mathbb{F}$-inner product spaces. Let $T : V \to W$ be a linear transformation. The **adjoint** of T is the function $T^* : W \to V$ s.t.

$$\langle T(v), w\rangle = \langle v, T^*(w)\rangle \quad \forall v \in V, w \in W. \tag{2.5.1}$$

2.5.2. Remark. — To see why this makes sense, suppose $T \in \mathcal{L}(V, W)$ and fix $w \in W$. Consider the linear functional

$$v \mapsto \langle T(v), w\rangle$$

on V, then this linear functional depends on T and w. By Riesz representation theorem (Theorem 2.3.5) there exists a unique vector in V s.t. this linear functional is given by taking the inner product with it. We call this unique vector $T^*(w)$, in other words $T^*(w)$ is the unique vector in V s.t.

$$\langle T(v), w\rangle = \langle v, T^*(w)\rangle \quad \forall v \in V.$$

In the equation above, the inner product on the left takes place in W and the inner product on the right takes place in V, but we use the same notation for both inner products.

2.5.3. Proposition (Adjoint of a linear map is a linear map)**.** —

$$T \in \mathcal{L}(V, W) \implies T^* \in \mathcal{L}(W, V).$$

Proof.

$$\begin{aligned}\langle T(v), w_1 + w_2\rangle &= \langle T(v), w_1\rangle + \langle T(v), w_2\rangle \\ &= \langle v, T^*(w_1)\rangle + \langle v, T^*(w_2)\rangle \\ &= \langle v, T^*(w_1) + T^*(w_2)\rangle .\end{aligned}$$

So $T^*(w_1 + w_2) = T^*(w_1) + T^*(w_2)$. Further,

$$\begin{aligned}\langle T(v), \lambda w\rangle &= \overline{\lambda} \langle T(v), w\rangle \\ &= \overline{\lambda} \langle v, T^*(w)\rangle \\ &= \langle v, \lambda T^*(w)\rangle .\end{aligned}$$

So $T^*(\lambda w) = \lambda T^*(w)$. □

2.5.4. Example. — Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be a linear map defined by

$$T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1) \quad \forall (x_1, x_2, x_3) \in \mathbb{R}^3.$$

Let $(y_1, y_2) \in \mathbb{R}^2$, then

$$\begin{aligned}\langle T(x_1, x_2, x_3), (y_1, y_2)\rangle &= \langle (x_2 + 3x_3, 2x_1), (y_1, y_2)\rangle \\ &= x_2y_1 + 3x_3y_1 + 2x_1y_2 \\ &= \langle (x_1, x_2, x_3), (2y_2, y_1, 3y_1)\rangle .\end{aligned}$$

So we have $T^* : \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$T^*(y_1, y_2) = (2y_2, y_1, 3y_1) \quad \forall (y_1, y_2) \in \mathbb{R}^2.$$

2.5.5. Theorem (Properties of adjoint of a linear map). — *Let $T \in \mathcal{L}(V, W)$, then*

1. $(S+T)^* = S^* + T^* \quad \forall S \in \mathcal{L}(V, W)$.
2. $(\lambda T)^* = \overline{\lambda} T^* \quad \forall \lambda \in \mathbb{F}$.
3. $(T^*)^* = T$.
4. $(ST)^* = T^* S^* \quad \forall S \in \mathcal{L}(W, U)$, *where U is a finite dimensional $\mathbb{F}$-inner product space.*
5. $I^* = I$, *where I is the identity operator on V.*
6. *T invertible $\implies T^*$ invertible and $(T^*)^{-1} = (T^{-1})^*$.*

Proof. Let $v \in V$ and $w \in W$.

1. If $S \in \mathcal{L}(V, W)$, then
$$\begin{aligned} \langle (S+T)(v), w \rangle &= \langle S(v), w \rangle + \langle T(v), w \rangle \\ &= \langle v, S*(w) \rangle + \langle v, T*(w) \rangle \\ &= \langle v, (S^* + T^*)(w) \rangle . \end{aligned}$$
2. If $\lambda \in \mathbb{F}$, then
$$\langle (\lambda T)(v), w \rangle = \lambda \langle T(v), w \rangle = \lambda \langle v, T^*(w) \rangle = \left\langle v, \overline{\lambda} T^*(w) \right\rangle .$$
3. We have
$$\langle T^*(w), v \rangle = \overline{\langle v, T^*(w) \rangle} = \overline{\langle T(v), w \rangle} = \langle w, T(v) \rangle$$
so $(T^*)^*(v) = T(v)$.
4. Let $S \in \mathcal{L}(W, U)$ and $u \in U$. Then
$$\langle (S \circ T)(v), u \rangle = \langle S(T(v)), u \rangle = \langle T(v), S*(u) \rangle = \langle v, T^*(S^*(u)) \rangle .$$
5. Suppose $u \in V$ then $\langle Iu, v \rangle = \langle u, v \rangle$, so $I^* = I$.
6. Suppose T is invertible. Then $T^{-1}T = I$. Taking adjoints on both sides,
$$\begin{aligned} \left(T^{-1}T\right)^* &= I^* \\ \implies T^*(T^{-1})^* &= I \end{aligned}$$

   Similarly, we get $(T^{-1})^* T^* = I$ from $TT^{-1} = I$. Thus, $(T^*)^{-1} = (T^*)^{-1}$.

□

2.5.6. Remark. — Similarly we have adjoints for square matrices $A, B \in \mathbb{F}^{n \times n}$.

1. $(A+B)^* = A^* + B^*$.
2. $(\lambda A)^* = \overline{\lambda} A^*$.
3. $(A^*)^* = A$.
4. $(AB)^* = B^* A^*$.
5. $I_n^* = I$.
6. $(A^*)^{-1} = (A^{-1})^* \quad \forall A \in \mathrm{GL}_n(\mathbb{F})$.

2.5.7. Theorem. — *Let $T \in \mathcal{L}(V, W)$, then*

1. $\ker(T^*) = (\operatorname{im}(T))^{\perp}$.
2. $\operatorname{im}(T^*) = (\ker(T))^{\perp}$.
3. $\ker(T) = (\operatorname{im}(T^*))^{\perp}$.
4. $\operatorname{im}(T) = (\ker(T^*))^{\perp}$.

Proof. Let $w \in W$. Then we begin by proving (1.)

$$\begin{aligned} w \in \ker(T^*) &\iff T^*(w) = 0 \\ &\iff \langle v, T^*(w) \rangle = 0 \quad \forall v \in V \\ &\iff \langle T(v), w \rangle = 0 \quad \forall v \in V \\ &\iff w \in (\operatorname{im}(T))^{\perp}. \end{aligned}$$

Thus, $\ker(T^*) = (\operatorname{im}(T))^{\perp}$. Taking the orthogonal complement on both sides yields (4.), using Theorem 2.4.10. As $(T^*)^* = T$, we get (3.) from (1.) by replacing T with T^*. Finally, we get (2.) from (4.) by replacing T with T^*. □

2.5.8. Theorem. — *If $T \in \mathcal{L}(V, W)$ then the following are equivalent*

1. *T is an inner product space isomorphism.*
2. *T is a vector space isomorphism and $T^{-1} = T^*$.*
3. $TT^* = \mathrm{id}_W$.
4. $T^*T = \mathrm{id}_V$.

Proof. Let T be an inner product space isomorphism, then

$$\begin{aligned} \langle T(x), y \rangle &= \left\langle T(x), T(T^{-1}(y)) \right\rangle \\ &= \left\langle x, T^{-1}(y) \right\rangle = \langle x, T^*(y) \rangle. \end{aligned}$$

So (1.) $\implies$ (2.), and (2.) trivially implies (3.) and (4.).

Now, (4.) $\implies$ (1.) as $T^*T = \mathrm{id}_V \implies T$ is injective $\implies T$ is bijective.

Also, (2.) $\implies$ (1.) as $\langle T(x), T(y) \rangle = \langle x, T^*(T(y)) \rangle = \langle x, y \rangle$. □

2.5.9. Definition (conjugate transpose)**.** — Let $A \in \mathbb{F}^{m \times n}$. Then the **conjugate transpose** of A is the $n \times m$ matrix $A^* \in \mathbb{F}^{n \times m}$ obtained by taking the complex conjugate of each entry in the transpose of A,

$$(a^*)_{ij} = (\overline{a})_{ji}.$$

When $\mathbb{F} = \mathbb{R}$, $A^* = A^T$. When $\mathbb{F} = \mathbb{C}$, $A^* = (\overline{A})^T$.

2.5.10. Example. — Let $A = \begin{pmatrix} 2 & 3+4i & 7 \\ 6 & 5 & 8i \end{pmatrix}$. Then $A^* = \begin{pmatrix} 2 & 6 \\ 3-4i & 5 \\ 7 & -8i \end{pmatrix}$.

2.5.11. Example. —

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}, \quad A^* = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}.$$

2.5.12. Theorem. — *Let V and W be two finite dimensional $\mathbb{F}$-inner product spaces. Suppose $T \in \mathcal{L}(V, W)$, $\{e_1, \dots, e_n\}$ is an ordered orthonormal basis of V and $\{f_1, \dots, f_m\}$ is an ordered orthonormal basis of W. If $m(T)$ is the matrix representation of T w.r.t. these bases, then the adjoint map $T* \in \mathcal{L}(W, V)$ is represented by $(m(T))^*$ w.r.t. the same bases. In other words,*

$$m(T^*) = (m(T))^*.$$

2.5.13. Definition (Self-adjoint)**.** — Let V be a finite dimensional $\mathbb{F}$-inner product space and $T \in \mathcal{L}(V)$ be a linear operator. Then T is **self-adjoint** if $T = T^*$.

In other words, T is self-adjoint if

$$\langle T(v), w \rangle = \langle v, T(w) \rangle \quad \forall v, w \in V.$$

In matrix representation, $m(T) = (m(T))^*$.

2.5.14. Example. — Let $\lambda \in \mathbb{F}$ and $T : \mathbb{F}^2 \to \mathbb{F}^2$ be a linear operator defined by

$$T(x, y) = (2x + \lambda y, 3x + 7y) \quad \forall (x, y) \in \mathbb{F}^2.$$

Then T is self-adjoint iff

$$m(T) = \begin{pmatrix} 2 & \lambda \\ 3 & 7 \end{pmatrix} = (m(T))^*.$$

But the adjoint is

$$m(T)^* = \begin{pmatrix} 2 & 3 \\ \overline{\lambda} & 7 \end{pmatrix},$$

so T is self-adjoint iff $\lambda = 3$.

In particular, a matrix $A \in \mathbb{F}^{n \times n}$ is self-adjoint iff $A = A^*$.

We know that the eigenvalues of a real symmetric matrix are real. This can be generalised.

2.5.15. Theorem. — *Every eigenvalue of a self-adjoint linear operator on a complex inner product space is real.*

Proof. Let V be a complex inner product space and λ an eigenvalue of T. As T is self-adjoint,

$$\begin{aligned} \lambda \|v\|^2 &= \lambda \langle v, v \rangle \\ &= \langle \lambda v, v \rangle \\ &= \langle T(v), v \rangle \\ &= \langle v, T(v) \rangle \\ &= \overline{\lambda} \langle v, v \rangle \\ &= \overline{\lambda} \|v\|^2. \\ \implies \lambda \|v\|^2 &= \overline{\lambda} \|v\|^2 \\ \implies \lambda &= \overline{\lambda}. \end{aligned}$$

Hence, λ is real. □

2.5.16. Theorem. — *Let V be a complex inner product space and $T \in \mathcal{L}(V)$ be a linear operator. Then $T(v)$ is orthogonal to v for all $v \in V$ iff $T = 0$, i.e.*

$$\langle T(v), v\rangle = 0 \iff T = 0.$$

Proof. If $T = 0$, then $\langle T(v), v\rangle = \langle 0, v\rangle = 0$ holds trivially.

Suppose $\langle T(v), v\rangle = 0$, and let $x, y \in V$. Then,

$$\begin{aligned}\langle T(x), y\rangle &= \frac{\langle T(x+y), x+y\rangle - \langle T(x-y), x-y\rangle}{4} \\ &+ \frac{\langle T(x+iy), x+iy\rangle - \langle T(x-iy), x-iy\rangle}{4} \\ &\implies \langle T(x), y\rangle = 0 \quad \forall x, y \in V.\end{aligned}$$

So, in particular, setting $y = T(x)$, we get $\langle T(x), T(x)\rangle = 0 \implies T(x) = 0 \quad \forall x \in V.$ □

2.5.17. Remark. — The above result doesn't hold in $\mathbb{F} = \mathbb{R}$. Consider for instance $T \in \mathcal{L}(\mathbb{R}^2)$ defined by

$$T(x, y) = (-y, x) \quad \forall (x, y) \in \mathbb{R}^2.$$

Then $\langle T(x, y), (x, y)\rangle = 0 \forall (x, y) \in \mathbb{R}^2$ but $T \neq 0$.

2.5.18. Theorem. — *Let V be a complex inner product space and $T \in \mathcal{L}(V)$ be a linear operator. Then T is self-adjoint iff $\langle T(v), v\rangle \in \mathbb{R}$.*

Proof. Let $v \in V$ and T^* be the adjoint of T. Then,

$$\begin{aligned}\langle T^*(v), v\rangle &= \langle v, T^*(v)\rangle \\ &= \langle v, T(v)\rangle .\end{aligned}$$

T is self-adjoint, so

$$\begin{aligned}&T - T^* = 0 \\ \iff &\langle (T - T^*)(v), v\rangle = 0 \\ \iff &\langle T(v), v\rangle - \langle T^*(v), v\rangle = 0 \\ \iff &\langle T(v), v\rangle - \langle v, T(v)\rangle = 0 \\ \iff &\langle T(v), v\rangle \in \mathbb{R}.\end{aligned}$$

□

2.5.19. Remark. — Again, the above result doesn't hold in $\mathbb{F} = \mathbb{R}$. Consider for instance $T \in \mathcal{L}(\mathbb{R}^2)$ defined by

$$T(x, y) = (2x - 3y, 3x + 2y) \quad \forall (x, y) \in \mathbb{R}^2.$$

Then $m(T) = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}$ which is not real symmetric, so not self-adjoint even though $\langle T(x, y), (x, y)\rangle = \langle (2x - 3y, 3x + 2y), (x, y)\rangle \in \mathbb{R}$.

2.5.20. Theorem. — *Let V be an $\mathbb{F}$-inner product space and $T \in \mathcal{L}(V)$ be self-adjoint. Then*

$$\langle T(v), v\rangle = 0 \quad \forall v \in V \iff T = 0.$$

Proof. If $T = 0$ then it is trivial.

Suppose $\langle T(v), v\rangle = 0$. Then, as T is self-adjoint,

$$\langle T(x), y\rangle = \frac{\langle T(x+y), x+y\rangle - \langle T(x-y), x-y\rangle}{4} = 0.$$

In particular, set $y = T(x)$, then $\langle T(x), T(x)\rangle = 0 \implies T = 0$. □

2.6. Normal linear operators

Let V be a finite-dimensional $\mathbb{F}$-inner product space. Let $T \in \mathcal{L}(V)$ be a linear operator.

2.6.1. Definition (Normal operator)**.** — T is called a **normal operator** if T commutes with its adjoint T^*, i.e. $TT^* = T^*T$.

Every self-adjoint operator is (trivially) a normal operator. The converse need not be true.

2.6.2. Example. — 1. $T \in \mathcal{L}(\mathbb{F}^2)$ defined by $T(x, y) = (2x - 3y, 3x + 2y)$.

$m(T) = \begin{pmatrix} 2 & -3 \\ 3 & 2 \end{pmatrix}$ wrt standard ordered basis of $\mathbb{F}^2$.

Its adjoint is $m(T)^* = \begin{pmatrix} 2 & 3 \\ -3 & 2 \end{pmatrix} \neq m(T)$, so it is not self-adjoint, but

$$m(T)m(T)^* = m(T)^*m(T) = \begin{pmatrix} 13 & 0 \\ 0 & 13 \end{pmatrix} \text{ so it is normal.}$$

2. Let $A \in \mathbb{C}^{2\times 2} : A = \begin{pmatrix} 2 & i \\ 1 & 2 \end{pmatrix}$. Then again A is not self-adjoint as

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

but it is normal as

$$AA^* = A^*A = \begin{pmatrix} 5 & 2+2i \\ 2-2i & 5 \end{pmatrix}.$$

2.6.3. Question. — Show that each of the matrices

$$A = \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

is normal but neither $A + B$ nor AB is normal.

2.6.4. Lemma. — *Let $T \in \mathcal{L}(V)$, then T is normal iff*

$$\|T^*(v)\| = \|T(v)\| \quad \forall v \in V.$$

Proof.

$$\begin{aligned}
&TT^* = T^*T \\
&\implies \|TT^* - T^*T\| = \|0\| = 0 \\
&\implies \langle (TT^* - T^*T)(v), v\rangle = 0 \quad \forall v \in V \\
&\implies \langle (TT^*)(v), v\rangle = \langle (T^*T)(v), v\rangle \\
&\implies \langle T^*(v), T^*(v)\rangle = \langle T(v), T(v)\rangle \\
&\implies \|T(v)\|^2 = \|T^*(v)\|^2.
\end{aligned}$$

Hence, $\|T^*(v)\| = \|T(v)\| \quad \forall v \in V$. □

2.6.5. Theorem. — *Let $P \in \mathcal{L}(V)$ be a projection operator*[2]*, then P is a normal operator iff P is a self-adjoint operator.*

Proof. Let P be normal, then $PP^* = P^*P$. Thus, $\|P^*(v)\| = \|P(v)\| \quad \forall v \in V$.
So, $P(v) = 0 \iff P^*(v) = 0 \quad \forall v \in V$. Let $y = v - P(v)$.

$$\begin{aligned}
\implies P(y) &= P(v - P(v)) \\
&= P(v) - P^2(v) \\
&= 0 \quad \because P^2 = P
\end{aligned}$$

Then

$$\begin{aligned}
0 &= P^*(y) \\
&= P^*(v - P(v)) \\
&= P^*(v) - (P^*P)(v) \\
P^*(v) &= (P^*)(v).
\end{aligned}$$

So $P^* = P^*P$. Then $P = (P^*)^* = (P^*P)^* = P^*(P^*)^* = P^*P = P^*$. □

2.6.6. Theorem. — *Let $T \in \mathcal{L}(V)$ be normal, then*

1. $\ker(T) = \ker(T^*)$.
2. $\operatorname{im}(T) = \operatorname{im}(T^*)$.
3. $V = \ker(T) \oplus \operatorname{im}(T)$.
4. *$T - \lambda I$ is normal for all $\lambda \in \mathbb{F}$.*
5. *For all $\lambda \in \mathbb{F}, v \in V$,*
$$T(v) = \lambda v \iff T^*(v) = \overline{\lambda} v.$$

Proof. For (1.) let $v \in \ker(T)$,

$$\begin{aligned}
&\iff T(v) = 0 \\
&\iff \|T(v)\| = 0 \\
&\iff \|T^*(v)\| = 0 \\
&\iff T^*(v) = 0 \\
&\iff v \in \ker(T^*).
\end{aligned}$$

[2] i.e. an idempotent operator

For (2.), $\operatorname{im}(T) = (\ker(T^*))^\perp = (\ker(T))^\perp = \operatorname{im}(T^*)$.
For (3.),

$$\begin{aligned} V &= \ker(T) \oplus (\ker(T))^\perp \\ &= \ker(T) \oplus \operatorname{im}(T^*) \\ &= \ker(T) \oplus \operatorname{im}(T). \end{aligned}$$

For (4.), we need to show $(T - \lambda I)(T - \lambda I)^* = (T - \lambda I)^*(T - \lambda I)$.

$$\begin{aligned} &(T - \lambda I)(T - \lambda I)^* \\ &= (T - \lambda I)(T^* - \overline{\lambda} I) \\ &= (TT^* - \lambda T^* - \overline{\lambda} T^* - |\lambda|^2 I) \\ &= (T^* - \overline{\lambda} I)(T - \lambda I) \\ &= (T - \lambda I)^*(T - \lambda I). \end{aligned}$$

Finally, (5.) is a direct consequence of (1.),

$$\begin{aligned} T(v) = \lambda v &\iff (T - \lambda I)v = 0 \\ &\iff v \in \ker(T - \lambda I) \\ &\iff v \in \ker(T - \lambda I)^* \\ &\iff v \in \ker(T^* - \overline{\lambda} I) \\ &\iff (T^* - \overline{\lambda} I)(v) = 0, \end{aligned}$$

and hence $T^*(v) = \overline{\lambda} v$. □

2.6.7. Theorem. — *Let $T \in \mathcal{L}(V)$ be normal. Then the eigenvectors of T corresponding to distinct eigenvalues are orthogonal.*

2.6.8. Theorem. — *Let V be a finite-dimensional complex inner product space. Then $T \in \mathcal{L}(V)$ is normal iff there exist commuting self-adjoint operators[3] A, B such that $T = A + iB$.*

2.6.9. Theorem. — *Let V be a finite-dimensional $\mathbb{F}$-inner product space. Then $T \in \mathcal{L}(V)$ is normal iff there exist commuting operators A, B s.t. A is self-adjoint, B is skew[4] and $T = A + B$.*

Proof. Let

$$A = \frac{1}{2}(T + T^*), \quad B = \frac{1}{2}(T - T^*).$$

Then $T = A + B$. Further,

$$\begin{aligned} T^*T - TT^* &= (A + B)^*(A + B) - (A + B)(A + B)^* \\ &= (A^* + B^*)(A + B) - (A + B)(A^* + B^*) \\ &= (A - B)(A + B) - (A + B)(A - B) \\ &= 2(AB - BA) = 0. \end{aligned}$$

Hence, T is normal. □

[3] this means $A = A*, B = B^*, AB = BA$
[4] $B^* = -B$

2.7. Spectral theorem

Invertible quadratic expressions. — Suppose $b, c \in \mathbb{R} : b^2 < 4c$. Let $x \in \mathbb{R}$, then by completing the square

$$x^2 + bx + c = \left(x + \frac{b}{2}\right)^2 + \left(c - \frac{b^2}{4}\right) > 0.$$

In particular, $x^2 + bx + c$ is an invertible real number, a convoluted way of saying that it is nonzero. Replacing the real number x with a self-adjoint linear operator yields the next result.

2.7.1. Theorem. — *Let V be a finite-dimensional $\mathbb{F}$-inner product space. Let $T \in \mathcal{L}(V)$ be self-adjoint. Let $b, c \in \mathbb{R} : b^2 < 4c$. Then $T^2 + bT + cI$ is an invertible linear operator.*

Proof. Let $v \in V \setminus \{0\}$. Then

$$\begin{aligned}
\left\langle (T^2 + bT + cI)(v), v \right\rangle &= \left\langle T^2(v), v \right\rangle + b \left\langle T(v), v \right\rangle + c \left\langle v, v \right\rangle \\
&= \|T(v)\|^2 + b \left\langle T(v), v \right\rangle + c\|v\|^2 \\
\text{(by Cauchy-Schwarz, Theorem 2.1.20)} \quad &\geqslant \|T(v)\|^2 - |b|\|T(v)\|\|v\| + c\|v\|^2 \\
&= \left(\|T(v)\| - \frac{|b|\|v\|}{2}\right)^2 + \left(c - \frac{b^2}{4}\right)\|v\|^2 \\
&> 0.
\end{aligned}$$

As $(T^2 + bT + cI)(v) \neq 0 \quad \forall v \neq 0$, $\ker(T^2 + bT + cI) = \{0\}$ which means $T^2 + bT + cI$ is injective, so $T^2 + bT + cI$ is bijective as it is linear. Hence, $T^2 + bT + cI$ is invertible. ☐

2.7.2. Theorem. — *Let $T \in \mathcal{L}(V)$ be self-adjoint. Then the minimal polynomial $m_T(x)$ of T is given by*

$$m_T(x) = (x - \lambda_1)(x - \lambda_2) \cdots (x - \lambda_n) \tag{2.7.1}$$

for some $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$.

2.7.3. Theorem (Real Spectral Theorem)**.** — *Let V be a finite-dimensional real inner product space and $T \in \mathcal{L}(V)$. Then the following are equivalent:*

1. *T is self-adjoint.*
2. *T has a diagonal matrix wrt some orthonormal basis of V.*
3. *V has an orthonormal basis consisting of eigenvectors of T.*

2.7.4. Theorem (Complex Spectral Theorem)**.** — *Let V be a finite-dimensional complex inner product space and $T \in \mathcal{L}(V)$. Then the following are equivalent:*

1. *T is normal.*
2. *T has a diagonal matrix wrt some orthonormal basis of V.*
3. *V has an orthonormal basis consisting of eigenvectors of T.*

2.8. Positive linear operators

2.8.1. Definition (Positive and positive-definite)**.** — Let V be a finite-dimensional $\mathbb{F}$-inner product space and $T \in \mathcal{L}(V)$.
T is **positive** if T is self-adjoint and $\langle T(v), v\rangle \geqslant 0 \quad \forall v \in V$.
T is **positive-definite** if T is self-adjoint and $\langle T(v), v\rangle > 0 \quad \forall v \in V$.

2.8.2. Example. —

1. $T \in \mathcal{L}(\mathbb{F}^2)$ given by $T(x, y) = (2x - y, -x + y)$ with
$$m(T) = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$
Then
$$\langle T(w, z), (w, z)\rangle = 2|w|^2 - 2\Re(w\overline{z}) + |z|^2 = |w - z|^2 + |w|^2 \geqslant 0 \forall (w, z) \in \mathbb{F}^2$$
so T is a positive operator.
2. If W is a subspace of V then the orthogonal projection P_W is a positive operator.
3. If $T \in \mathcal{L}(V)$ is self-adjoint and $b, c \in \mathbb{R} : b^2 < 4c$, then $T^2 + bT + cI$ is a positive operator.
4. Consider $T \in \mathcal{L}(\mathbb{R}^2)$ defined by
$$T(x, y) = \left(x \cos\theta + y \sin\theta, \quad -x \sin\theta + y \cos\theta\right).$$
For which values of θ is T positive ?
In this case we want for all $v \in \mathbb{R}^2$,
$$\langle T(v), v\rangle = \|v\|\|Av\| \cos\theta \geqslant 0$$
so $\theta \in [-\pi/2, \pi/2]$.

2.8.3. Definition (Square root)**.** — An operator S is called a **square root** of an operator T if $S^2 = T$.

2.8.4. Example. — $T \in \mathcal{L}(\mathbb{F}^3)$ given by $T(x, y, z) = (z, 0, 0)$ has a square root $S \in \mathcal{L}(\mathbb{F}^3)$ defined by
$$S(x, y, z) = (y, z, 0).$$
This is because $S^2 = T$.

2.8.5. Theorem (Characterisation of positive-definite operators)**.** — *Let $T \in \mathcal{L}(V)$. Then the following are equivalent:*

1. *T is positive-definite.*
2. *T is self-adjoint and all eigenvalues of T are strictly positive.*
3. *Wrt some orthonormal basis of V, the matrix of T is a diagonal matrix with only (strictly) positive numbers on the (main) diagonal.*
4. *T has an invertible positve square root.*

5. *T has an invertible self-adjoint square root.*

6. *$T = S^*S$ for some invertible $S \in \mathcal{L}(V)$.*

Linear maps that preserve norms are sufficiently important to deserve a name.

2.8.6. Definition (Unitary operator or isometry)**.** — $T \in \mathcal{L}(V)$ is **unitary** if $\|T(v)\| = \|v\|$.

2.8.7. Remark. — Although unitary operator and isometry are the same for operators on finite-dimensional spaces, a unitary operator maps a space to itself whereas an isometry may map into an altogether different space.

2.8.8. Example. —

$$m(T) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta. \end{pmatrix}$$

2.8.9. Theorem (Characterisation of unitary operators)**.** — *Let $T \in \mathcal{L}(V)$ and $\{e_1, \ldots, e_n\}$ be an orthonormal basis of V. Then the following are equivalent:*

1. *T is unitary.*

2. *$T^*T = TT^* = I$.*

3. *T is invertible with $T^{-1} = T^*$.*

4. *$\langle T(x), T(y)\rangle = \langle x, y\rangle \quad \forall x, y \in V$.*

5. *$\{T(e_1), \ldots, T(e_n)\}$ is an orthonormal basis of V.*

6. *The rows of $m(T)$ wrt the basis $\{e_1, \ldots, e_n\}$ form an orthonormal basis of $\mathbb{F}^n$ wrt the Euclidean inner product.*

7. *T^* is unitary.*

Analogies between complex numbers and linear operators. —

$\mathbb{C}$	$\mathcal{L}(V)$
$z = \overline{z}$	$T = T^*$
$z\overline{z} = \overline{z}z = 1$	$TT^* = T^*T = I$

2.8.10. Theorem. — *Let V be a finite-dimensional $\mathbb{F}$-inner product space and let $T \in \mathcal{L}(V)$ be unitary. If λ is an eigenvalue of T, then $|\lambda| = 1$.*

Proof. $\|T(v)\| = \|v\|$ and $\exists\, v \neq 0 : T(v) = \lambda v$,

$$\begin{aligned} &\implies \|T(v)\| = |\lambda|\|v\| \\ &\implies \|v\| = |\lambda|\|v\| \\ &\implies 1 = |\lambda| \quad \because v \neq 0. \end{aligned}$$

So $|\lambda| = 1$. □

Chapter 3.

3

Determinant and generalised inverses

3.1. Determinant as a multilinear function

Bilinear and multilinear forms. — The inner product is an example of what we call a *bilinear form* - it is linear in 2 input parameters. Similarly, we can have *k-linear* or *multilinear forms* which are linear in k input parameters. The determinant is one such example of a multilinear form.

3.1.1. Definition (Determinant). — Let $\mathbb{K}$ be a field with the standard basis $\{e_1, \dots, e_n\}$. Then the **determinant** is a function $\det : \underbrace{\mathbb{K}^n \times \cdots \times \mathbb{K}^n}_{n \text{ times}} \to \mathbb{K}$ that is uniquely characterised by the following 3 properties:

1. For $\lambda \in \mathbb{K}, \quad x, y \in \mathbb{K}^n$, (multilinear, or n-linear)
$$\begin{aligned} &\det(v_1, \dots, v_{i-1}, x + \lambda y, v_{i+1}, \dots, v_n) \\ =&\det(v_1, \dots, v_{i-1}, x, v_{i+1}, \dots, v_n) + \lambda \det(v_1, \dots, v_{i-1}, y, v_{i+1}, \dots, v_n). \end{aligned}$$
2. $\det(v_1, \dots, v_n) = 0$ if $v_i = v_j$ for some $1 \leqslant i < j \leqslant n$. (antisymmetric, or alternating)
3. $\det(e_1, \dots, e_n) = 1$. (identity)

3.1.2. Remark. — The above can be summarised in one sentence: the order n determinant is the unique alternating n-form $\det \in \bigwedge^n (\mathbb{K}^n)^*$ for which $\det(e_1, \dots, e_n) = 1$.

Also observe that from the alternating property we have
$$\begin{aligned} &\det(v_1, \dots, v_i, \dots, v_j, \dots, v_n) + \det(v_1, \dots, v_j, \dots, v_i, \dots, v_n) \\ &= \det(2v_1, \dots, v_i + v_j, \dots, v_j + v_i, \dots, 2v_n) \\ &= 0 \end{aligned}$$
so $\det(v_1, \dots, v_i, \dots, v_j, \dots, v_n) = -\det(v_1, \dots, v_j, \dots, v_i, \dots, v_n)$.

This is why det is *antisymmetric*.

3.2. Singular value decomposition

3.2.1. Definition/Proposition (Singular value decomposition). — A **singular value decomposition** (SVD) of a matrix $A \in \mathbb{K}^{m \times n}$ of rank r is a factorisation
$$A = ULV^*$$
where $L = \operatorname{diag}(l_1, \dots, l_r),\ l_r > 0$ and $U^*U = I_r = V^*V$.

3.2.2. Corollary. — *Let $A \in \mathbb{K}^{m\times n}$ with $r = \operatorname{rank}(A)$. Then A contains at least one $r \times r$ nonsingular matrix B such that*

$$A = \begin{pmatrix} B & C \\ D & E \end{pmatrix}. \tag{3.2.1}$$

3.3. Generalised inverses

Motivation. — Given two square matrices A, B with $AB = BA = I$ and $\det(A) \neq 0$ the inverse of A is $B = A^{-1}$. A square matrix is invertible iff it is non-singular. Note here that A and B are *square* matrices i.e. $A, B \in \mathbb{K}^{n\times n}$.

Is it possible to generalise the notion of an inverse to any $A \in \mathbb{K}^{m\times n}$? The inverse B of a square matrix A satisfies $ABA = A$. So if $A \in \mathbb{K}^{m\times n}$ then we want some $B \in \mathbb{K}^{n\times m}$ such that $ABA = A$.

3.3.1. Definition (Generalised inverse)**.** — Let $A \in \mathbb{K}^{m\times n}$. If there exists a matrix $B \in \mathbb{K}^{n\times m}$ such that $ABA = A$ then B is called a **generalised inverse** (or g-inverse) of A.

3.3.2. Theorem. — *Let $A \in \mathbb{K}^{m\times n}$ and G be a g-inverse of A. Then AG and GA are idempotent and*

$$\operatorname{rank}(AG) = \operatorname{rank}(GA) = \operatorname{rank}(A). \tag{3.3.1}$$

Proof. $AGA = A$ so $(AG)^2 = (AG)(AG) = (AGA)G = AG$. Hence AG is idempotent. Similarly, $(GA)^2 = (GA)(GA) = G(AGA) = GA$. Hence GA is idempotent.

Now, $\boxed{\operatorname{rank}(AB) \leqslant \min(\operatorname{rank}(A), \operatorname{rank}(B))}$. Thus,

$$\operatorname{rank}(AG) \leqslant \operatorname{rank}(A) = \operatorname{rank}((AG)A) \leqslant \operatorname{rank}(AG).$$

Similarly,

$$\operatorname{rank}(GA) \leqslant \operatorname{rank}(A) = \operatorname{rank}(A(GA)) \leqslant \operatorname{rank}(GA).$$

Hence $\operatorname{rank}(AG) = \operatorname{rank}(GA) = \operatorname{rank}(A)$. □

3.3.3. Remark. — Every matrix over a field has a g-inverse. In general, the g-inverse of a matrix is not unique. In fact, there are infinitely many g-inverses of a matrix $A \in \mathbb{K}^{m\times n}$. But if $m = n$ and $\det(A) \neq 0$ then A has the unique g-inverse A^{-1} which is just the inverse of A.

3.3.4. Theorem. — *If G_1, G_2 are two g-inverses of $A \in \mathbb{K}^{m\times n}$ then for any $\lambda \in \mathbb{K}$*

$$\lambda G_1 + (1-\lambda)G_2$$

is a g-inverse of A.

Proof. We are given $AG_1A = A = AG_2A$. Now for any $\lambda \in \mathbb{K}$ we have

$$\begin{aligned} &A\left(\lambda G_1 + (1-\lambda)G_2\right)A \\ &= \lambda AG_1A + (1-\lambda)AG_2A \\ &= \lambda A + (1-\lambda)A \\ &= A. \end{aligned}$$

Hence $A\left(\lambda G_1 + (1-\lambda)G_2\right)A = A$ so $\lambda G_1 + (1-\lambda)G_2$ is a g-inverse of A. □

Finding a g-inverse of a matrix. — Given a matrix $A \in \mathbb{K}^{m\times n}$ we want to find a g-inverse of A.

First compute $r = \operatorname{rank}(A)$. We then get an $r \times r$ non-singular matrix B such that $A = \begin{pmatrix} B & C \\ D & E \end{pmatrix}$ by SVD of A (using (3.2.1)). Then

$$G = \begin{pmatrix} B^{-1} & 0 \\ 0 & 0 \end{pmatrix} \tag{3.3.2}$$

is a g-inverse of A.

3.3.5. Example. — 1. Let $A = \begin{pmatrix} 1 & 7 & 1 & 4 \\ 1 & 2 & 0 & 1 \\ 0 & 5 & 1 & 3 \end{pmatrix}$. We have $\operatorname{rank}(A) = 2$, so $B = \begin{pmatrix} 1 & 7 \\ 1 & 2 \end{pmatrix}$, then $B^{-1} = \frac{1}{5}\begin{pmatrix} -2 & 7 \\ 1 & -1 \end{pmatrix}$. So a g-inverse of A is $G = \begin{pmatrix} -2/5 & 7/5 & 0 \\ 1/5 & -1/5 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

2. Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$. As $\det(A) = 0$, inverse will not exist, but $\operatorname{rank}(A) = 2$, so we have $B = \begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}$. Then $B^{-1} = -\frac{1}{3}\begin{pmatrix} 5 & -2 \\ -4 & 1 \end{pmatrix}$, so a g-inverse of A is $G = \begin{pmatrix} -5/3 & 2/3 & 0 \\ 4/3 & -1/3 & 0 \\ 0 & 0 & 0 \end{pmatrix}$.

3.3.6. Question. — Does there exist a g-inverse of a singular matrix ?

The answer is yes; consider $A = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}$. Then $\det(A) = 0$.

A g-inverse of this A is then given by $G = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.

3.3.7. Definition (Moore-Penrose pseudoinverse)**.** — If A^+ is a g-inverse of $A \in \mathbb{K}^{m\times n}$ such that

1. $A^+AA^+ = A^+$
2. $(AA^+)^* = AA^+$
3. $(A^+A)^* = A^+A$

then A^+ is called the **Moore-Penrose pseudoinverse** of A.

3.3.8. Theorem. — *Let $A \in \mathbb{K}^{n\times n}$. Then $\det(A) \neq 0 \iff A^{-1}$ is the unique g-inverse and Moore-Penrose pseudoinverse of A.*

Proof. We have trivially $AA^{-1}A = A$ so A^{-1} is a g-inverse. If G is a g-inverse of A then

$$AGA = A \implies A^{-1}(AGA)A^{-1} = A^{-1}(A)A^{-1} \implies G = A^{-1}.$$

Hence, A^{-1} is the unique g-inverse of A. □

CHAPTER 4.

Exam review

4

4.1. Internal assessment

4.1.1. Question. — Define a linear functional. Give 3 distinct examples of linear functionals on $\mathbb{R}^{[0,1]}$.

Let V be an $\mathbb{F}$-vector space and V' its dual space. Then show that the dual map of the identity operator on V is the identity operator on V'.

4.1.2. Question. — Let V be a finite dimensional $\mathbb{F}$-vector space. Let $T : V \to V$ be a linear operator.

(a) Prove that T has at most $1 + \dim(\operatorname{im}(T))$ distinct eigenvalues.

(b) Let the minimal polynomial of T be $2x^5 - 3x^4 + 5x^3 - 6x^2 + 7x - 6$. Is T invertible ? Justify.

(c) Suppose $T^4 = T$. Is T diagonalisable ? Justify.

Solutions. —

1. Let V be a vector space over a field $\mathbb{F}$. A linear functional ϕ is a linear transformation $\phi : V \to \mathbb{F}$, i.e. ϕ maps V into $\mathbb{F}$. $\mathbb{R}^{[0,1]}$ is the real vector space of all functions $f : [0,1] \to \mathbb{R}$. Three distinct examples of linear functionals on $\mathbb{R}^{[0,1]}$ are

   a) $\phi(f(x)) = f(0)$,

   b) $\phi(f(x)) = f(1/2)$ and

   c) $\phi(f(x)) = f(1)$

   for all $x \in [0,1],\ f \in \mathbb{R}^{[0,1]}$.

   Let $\operatorname{id} : V \to V$ be the identity operator on V, i.e. $\operatorname{id}(x) = x \quad \forall x \in V$. Let $\operatorname{id}' : V' \to V'$ be the dual mapping of the identity operator id. Then by definition

   $$\operatorname{id}'(\phi(x)) = (\phi \circ \operatorname{id})(x) = \phi(\operatorname{id}(x)) = \phi(x) \quad \forall \phi \in V', x \in V.$$

   Hence, id' is the identity operator on V'.

2. (a) Let $\lambda_1, \ldots, \lambda_m$ be the distinct nonzero eigenvalues of T. Let $v_1, \ldots, v_m$ be the corresponding eigenvectors. Then $\{v_1, \ldots, v_m\}$ is linearly independent.

   Now if v is an eigenvector corresponding to a nonzero eigenvalue λ then

   $$T(v) = \lambda v \implies T(v/\lambda) = v \implies v \in \operatorname{im}(T).$$

   Hence, $\{v_1, \ldots, v_m\} \subseteq \operatorname{im}(T)$. Also the eigenvalue may not be nonzero. Thus, $m \leqslant 1 + \dim(\operatorname{im}(T))$.

(b) A linear operator T is invertible iff its minimal polynomial $m_T(x)$ has nonzero constant term, i.e. $m_T(0) \neq 0$. We have $m_T(0) = -6 \neq 0$ so T is invertible.

(c) We assume $\mathbb{F}$ is algebraically closed (say $\mathbb{F} = \mathbb{C}$). Then

$$T^4 = T \implies T^4 - T = T(T^3 - I) = 0.$$

So T satisfies

$$f(x) = x(x^3 - 1) = x(x-1)(x-\omega)(x-\omega^2) = 0,$$

where ω is a primitive cube root of unity. Then $m_T(x) \mid f(x)$.

So either $m_T(x) = x,\ x(x-1),\ x(x-1)(x-\omega)$ or $x(x-1)(x-\omega)(x-\omega^2)$. In every case, the minimal polynomial $m_T(x)$ splits as a product of linear terms. Hence, T is diagonalisable.

Bibliography

[Axl24] Sheldon Axler. *Linear Algebra Done Right*. Springer, 2024. URL: https://linear.axler.net/LADR4e.pdf.