Appendix B

Linear Algebra

B.1 Vector spaces

Definition B.1. A *field* $\mathbb{F}$ is a set together with two binary operations, usually called "addition" and "multiplication" and denoted by "+" and "$*$", such that $\forall a, b, c \in \mathbb{F}$, the following axioms hold,

- commutativity: $a + b = b + a$, $ab = ba$;
- associativity: $a + (b + c) = (a + b) + c$, $a(bc) = (ab)c$;
- identity: $a + 0 = a$, $a1 = a$;
- invertibility: $a + (-a) = 0$, $aa^{-1} = 1$ $(a \neq 0)$;
- distributivity: $a(b + c) = ab + ac$.

Definition B.2. A *vector space* or *linear space* over a field $\mathbb{F}$ is a set $\mathcal{V}$ together with two binary operations "+" and "$\times$" respectively called vector addition and scalar multiplication that satisfy the following axioms:

(VSA-1) commutativity
$\forall \mathbf{u}, \mathbf{v} \in \mathcal{V}$, $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$;

(VSA-2) associativity
$\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$, $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$;

(VSA-3) compatibility
$\forall \mathbf{u} \in \mathcal{V}$, $\forall a, b \in \mathbb{F}$, $(ab)\mathbf{u} = a(b\mathbf{u})$;

(VSA-4) additive identity
$\exists \mathbf{0} \in \mathcal{V}$, $\forall \mathbf{u} \in \mathcal{V}$, s.t. $\mathbf{u} + \mathbf{0} = \mathbf{u}$;

(VSA-5) additive inverse
$\forall \mathbf{u} \in \mathcal{V}$, $\exists \mathbf{v} \in \mathcal{V}$, s.t. $\mathbf{u} + \mathbf{v} = \mathbf{0}$;

(VSA-6) multiplicative identity
$\exists 1 \in \mathbb{F}$, s.t. $\forall \mathbf{u} \in \mathcal{V}$, $1\mathbf{u} = \mathbf{u}$;

(VSA-7) distributive laws

$$\forall \mathbf{u}, \mathbf{v} \in \mathcal{V},\ \forall a, b \in \mathbb{F},\ \begin{cases} (a + b)\mathbf{u} = a\mathbf{u} + b\mathbf{u}, \\ a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}. \end{cases}$$

The elements of $\mathcal{V}$ are called *vectors* and the elements of $\mathbb{F}$ are called *scalars*.

Definition B.3. A *real vector space* or a *complex vector space* is a vector space with $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, respectively.

Exercise B.4. Show that a complex vector space always induces another real vector space.

Example B.5. The simplest vector space is $\{\mathbf{0}\}$. Another simple example of a vector space over a field $\mathbb{F}$ is $\mathbb{F}$ itself, equipped with its standard addition and multiplication.

B.1.1 Subspaces

Definition B.6. A subset $\mathcal{U}$ of $\mathcal{V}$ is called a *subspace* of $\mathcal{V}$ if $\mathcal{U}$ is also a vector space when equipped with the same addition and scalar multiplication on V.

Definition B.7. Suppose $\mathcal{U}_1, \ldots, \mathcal{U}_m$ are subsets of $\mathcal{V}$. The *sum* of $\mathcal{U}_1, \ldots, \mathcal{U}_m$ is the set of all possible sums of elements of $\mathcal{U}_1, \ldots, \mathcal{U}_m$:

$$\mathcal{U}_1 + \ldots + \mathcal{U}_m := \left\{ \sum_{j=1}^{m} \mathbf{u}_j : \mathbf{u}_j \in \mathcal{U}_j \right\}. \tag{B.1}$$

Example B.8. For $\mathcal{U} = \{(x, x, y, y) \in \mathbb{F}^4 : x, y \in \mathbb{F}\}$ and $\mathcal{W} = \{(x, x, x, y) \in \mathbb{F}^4 : x, y \in \mathbb{F}\}$, we have

$$\mathcal{U} + \mathcal{W} = \{(x, x, z, y) \in \mathbb{F}^4 : x, y, z \in \mathbb{F}\}.$$

Lemma B.9. Suppose $\mathcal{U}_1, \ldots, \mathcal{U}_m$ are subspaces of $\mathcal{V}$. Then $\mathcal{U}_1 + \ldots + \mathcal{U}_m$ is the smallest subspace of $\mathcal{V}$ that contains $\mathcal{U}_1, \ldots, \mathcal{U}_m$.

Definition B.10. Suppose $\mathcal{U}_1, \ldots, \mathcal{U}_m$ are subspaces of $\mathcal{V}$. The sum $\mathcal{U}_1 + \ldots + \mathcal{U}_m$ is called a *direct sum* if each element in $\mathcal{U}_1 + \ldots + \mathcal{U}_m$ can be written in only one way as a sum $\sum_{j=1}^{m} \mathbf{u}_j$ with $\mathbf{u}_j \in \mathcal{U}_j$ for each $j = 1, \ldots, m$. In this case we write the direct sum as $\mathcal{U}_1 \oplus \ldots \oplus \mathcal{U}_m$.

Exercise B.11. Show that $\mathcal{U}_1 + \mathcal{U}_2 + \mathcal{U}_3$ is not a direct sum:

$$\begin{aligned} \mathcal{U}_1 &= \{(x, y, 0) \in \mathbb{F}^3 : x, y \in \mathbb{F}\}, \\ \mathcal{U}_2 &= \{(0, 0, z) \in \mathbb{F}^3 : z \in \mathbb{F}\}, \\ \mathcal{U}_3 &= \{(0, y, y) \in \mathbb{F}^3 : y \in \mathbb{F}\}. \end{aligned}$$

Lemma B.12. Suppose $\mathcal{U}_1, \ldots, \mathcal{U}_m$ are subspaces of $\mathcal{V}$. Then $\mathcal{U}_1 + \ldots + \mathcal{U}_m$ is a direct sum if and only if the only way to write $\mathbf{0}$ as a sum $\sum_{j=1}^{m} \mathbf{u}_j$, where $\mathbf{u}_j \in \mathcal{U}_j$ for each $j = 1, \ldots, m$, is by taking each $\mathbf{u}_j$ equal to $\mathbf{0}$.

Theorem B.13. Suppose $\mathcal{U}$ and $\mathcal{W}$ are subspaces of $\mathcal{V}$. Then $\mathcal{U} + \mathcal{W}$ is a direct sum if and only if $\mathcal{U} \cap \mathcal{W} = \{\mathbf{0}\}$.

B.1.2 Span and linear independence

Definition B.14. A *list of length* n or n*-tuple* is an ordered collection of n elements (which might be numbers, other lists, or more abstract entities) separated by commas and surrounded by parentheses: $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

Definition B.15. A vector space composed of all the n-tuples of a field $\mathbb{F}$ is known as a *coordinate space*, denoted by $\mathbb{F}^n$ ($n \in \mathbb{N}^+$).

Example B.16. The properties of forces or velocities in the real world can be captured by a coordinate space $\mathbb{R}^2$ or $\mathbb{R}^3$.

Example B.17. The set of continuous real-valued functions on the interval $[a, b]$ forms a real vector space.

Notation 25. For a set $\mathcal{S}$, define a vector space

$$\mathbb{F}^{\mathcal{S}} := \{f : \mathcal{S} \to \mathbb{F}\}.$$

$\mathbb{F}^n$ is a special case of $\mathbb{F}^{\mathcal{S}}$ because n can be regarded as the set $\{1, 2, \ldots, n\}$ and each element in $\mathbb{F}^n$ can be considered as a function $\{1, 2, \ldots, n\} \mapsto \mathbb{F}$.

Definition B.18. A *linear combination* of a list of vectors $\{\mathbf{v}_i\}$ is a vector of the form $\sum_i a_i \mathbf{v}_i$ where $a_i \in \mathbb{F}$.

Example B.19. $(17, -4, 2)$ is a linear combination of $(2, 1, -3), (1, -2, 4)$ because

$$(17, -4, 2) = 6(2, 1, -3) + 5(1, -2, 4).$$

Example B.20. $(17, -4, 5)$ is not a linear combination of $(2, 1, -3), (1, -2, 4)$ because there do not exist numbers a_1, a_2 such that

$$(17, -4, 5) = a_1(2, 1, -3) + a_2(1, -2, 4).$$

Solving from the first two equations yields $a_1 = 6$, $a_2 = 5$, but $5 \neq -3 \times 6 + 4 \times 5$.

Definition B.21. The *span* of a list of vectors $(\mathbf{v}_i)$ is the set of all linear combinations of $(\mathbf{v}_i)$,

$$\mathrm{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m) = \left\{ \sum_{i=1}^{m} a_i \mathbf{v}_i : \ a_i \in \mathbb{F} \right\}. \qquad \text{(B.2)}$$

In particular, the span of the empty set is $\{\mathbf{0}\}$. We say that $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$ *spans* $\mathcal{V}$ if $\mathcal{V} = \mathrm{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$.

Example B.22.

$$\begin{aligned}(17, -4, 2) &\in \mathrm{span}((2, 1, -3), (1, -2, 4))\\ (17, -4, 5) &\notin \mathrm{span}((2, 1, -3), (1, -2, 4))\end{aligned}$$

Definition B.23. A vector space $\mathcal{V}$ is called *finite dimensional* if some list of vectors span $\mathcal{V}$; otherwise it is *infinite dimensional*.

Example B.24. Let $\mathbb{P}_m(\mathbb{F})$ denote the set of all polynomials with coefficients in $\mathbb{F}$ and degree at most m,

$$\mathbb{P}_m(\mathbb{F}) = \left\{ p : \mathbb{F} \to \mathbb{F};\ p(z) = \sum_{i=0}^{m} a_i z^i, a_i \in \mathbb{F} \right\}. \qquad \text{(B.3)}$$

Then $\mathbb{P}_m(\mathbb{F})$ is a finite-dimensional vector space for each non-negative integer m. The set of all polynomials with coefficients in $\mathbb{F}$, denoted by $\mathbb{P}(\mathbb{F}) := \mathbb{P}_{+\infty}(\mathbb{F})$, is infinite-dimensional. Both are subspaces of $\mathbb{F}^{\mathbb{F}}$ for $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

Definition B.25. A list of vectors $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$ in $\mathcal{V}$ is called *linearly independent* iff

$$a_1\mathbf{v}_1 + \ldots + a_m\mathbf{v}_m = \mathbf{0} \ \Rightarrow\ a_1 = \cdots = a_m = 0. \qquad \text{(B.4)}$$

Otherwise the list of vectors is called *linearly dependent*.

Example B.26. The empty list is declared to be linearly independent. A list of one vector $(\mathbf{v})$ is linearly independent iff $\mathbf{v} \neq \mathbf{0}$. A list of two vectors is linearly independent iff neither vector is a scalar multiple of the other.

Example B.27. The list $(1, z, \ldots, z^m)$ is linearly independent in $\mathbb{P}_m(\mathbb{F})$ for each $m \in \mathbb{N}$.

Example B.28. $(2, 3, 1)$, $(1, -1, 2)$, and $(7, 3, 8)$ is linearly dependent in $\mathbb{R}^3$ because

$$2(2, 3, 1) + 3(1, -1, 2) + (-1)(7, 3, 8) = (0, 0, 0).$$

Example B.29. Every list of vectors containing the $\mathbf{0}$ vector is linearly dependent.

Lemma B.30 (Linear dependence lemma)**.** Suppose $V = (\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m)$ is a linearly dependent list in $\mathcal{V}$. Then there exists $j \in \{1, 2, \ldots, m\}$ such that

- $\mathbf{v}_j \in \mathrm{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{j-1})$;
- if the jth term is removed from V, the span of the remaining list equals $\mathrm{span}(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$.

Lemma B.31. In a finite-dimensional vector space, the length of every linearly independent list of vectors is less than or equal to the length of every spanning list of vectors.

B.1.3 Bases

Definition B.32. A *basis* of a vector space $\mathcal{V}$ is a list of vectors in $\mathcal{V}$ that is linearly independent and spans $\mathcal{V}$.

Definition B.33. The *standard basis* of $\mathbb{F}^n$ is the list of vectors

$$(1, 0, \cdots, 0)^T,\ (0, 1, 0, \cdots, 0)^T,\ \ldots,\ (0, \cdots, 0, 1)^T. \qquad \text{(B.5)}$$

Example B.34. $(z^0, z^1, \ldots, z^m)$ is a basis of $\mathbb{P}_m(\mathbb{F})$ in (B.3).

Lemma B.35. A list of vectors $(\mathbf{v}_1, \ldots, \mathbf{v}_n)$ is a basis of $\mathcal{V}$ iff every vector $\mathbf{u} \in \mathcal{V}$ can be written uniquely as

$$\mathbf{u} = \sum_{i=1}^{n} a_i \mathbf{v}_i, \qquad \text{(B.6)}$$

where $a_i \in \mathbb{F}$.

Lemma B.36. Every spanning list in a vector space $\mathcal{V}$ can be reduced to a basis of $\mathcal{V}$.

Lemma B.37. Every linearly independent list of vectors in a finite-dimensional vector space can be extended to a basis of that vector space.

B.1.4 Dimension

Lemma B.38. Any two bases of a finite-dimensional vector space have the same length.

Proof. Suppose B_1 and B_2 are two bases of V. Then B_1 is linearly independent in V and B_2 spans V. By Lemma B.31, the length of B_1 is no greater than B_2. The proof is completed by switching the roles of B_1 and B_2. □

Definition B.39. The *dimension* of a finite-dimensional vector space $\mathcal{V}$, denoted $\dim \mathcal{V}$, is the length of any basis of the vector space.

Lemma B.40. If $\mathcal{V}$ is finite-dimensional, then every spanning list of vectors in $\mathcal{V}$ with length $\dim \mathcal{V}$ is a basis of $\mathcal{V}$.

Lemma B.41. If $\mathcal{V}$ is finite-dimensional, then every linearly independent list of vectors in $\mathcal{V}$ with length $\dim \mathcal{V}$ is a basis of $\mathcal{V}$.

B.2 Linear maps

Definition B.42. A *linear map* or *linear transformation* between two vector spaces $\mathcal{V}$ and $\mathcal{W}$ is a function $T : \mathcal{V} \to \mathcal{W}$ that satisfies

(LNM-1) additivity
$\forall \mathbf{u}, \mathbf{v} \in \mathcal{V}$, $T(\mathbf{u} + \mathbf{v}) = T\mathbf{u} + T\mathbf{v}$;

(LNM-2) homogeneity
$\forall a \in \mathbb{F}$, $\forall \mathbf{v} \in \mathcal{V}$, $T(a\mathbf{v}) = a(T\mathbf{v})$,

where $\mathbb{F}$ is a scalar field. In particular, a linear map $T : \mathcal{V} \to \mathcal{W}$ is called a *(linear) operator* if $\mathcal{W} = \mathcal{V}$.

Notation 26. The set of all linear maps from $\mathcal{V}$ to $\mathcal{W}$ is denoted by $\mathcal{L}(\mathcal{V}, \mathcal{W})$. The set of all linear operators from $\mathcal{V}$ to itself is denoted by $\mathcal{L}(\mathcal{V})$.

Example B.43. The differentiation operator on $\mathbb{R}[x]$ is a linear map $T \in \mathcal{L}(\mathbb{R}[x], \mathbb{R}[x])$

Example B.44. $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m) \simeq \mathbb{F}^{m\times n}$ is a vector space with the zero map $\mathbf{0}$ as the additive identity.

Lemma B.45. The set $\mathcal{L}(\mathcal{V}, \mathcal{W})$, equipped with scalar multiplication $(aT)\mathbf{v} = a(T\mathbf{v})$ and vector addition $(S + T)\mathbf{v} = S\mathbf{v} + T\mathbf{v}$, is a vector space.

Proof. The scalar field $\mathbb{F}$ of $\mathcal{L}(\mathcal{V}, \mathcal{W})$ is the same as that of $\mathcal{V}$ and $\mathcal{W}$. So multiplicative identity is still 1, the same as that of $\mathbb{F}$. However, the additive identity is the zero map $\mathbf{0} \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. □

Definition B.46. The *identity map*, denoted I, is the function on a vector space that assigns to each element the same element:

$$I\mathbf{v} = \mathbf{v}. \tag{B.7}$$

Definition B.47. A *complex linear functional* is a linear map $T : \mathcal{V} \to \mathbb{C}$ with $\mathbb{C}$ being the underlying field of $\mathcal{V}$. A *real linear functional* is a map $T : \mathcal{V} \to \mathbb{R}$ such that (LNM-1) and (LNM-2) in Definition B.42 hold for $\mathbb{F} = \mathbb{R}$.

Lemma B.48. Let V be a complex vector space and f a complex linear functional on V. Then the real part $\mathrm{Re}\, f(x) = u(x)$ is related to f by

$$\forall x \in V, \quad f(x) = u(x) - \mathbf{i}u(\mathbf{i}x). \tag{B.8}$$

Proof. Any $\alpha, \beta \in \mathbb{R}$ and $z = \alpha + \mathbf{i}\beta \in \mathbb{C}$ satisfy

$$z = \mathrm{Re}\, z - \mathbf{i}\mathrm{Re}\,(\mathbf{i}z).$$

Set $z = f(x)$ and we have

$$\begin{aligned} f(x) &= \mathrm{Re}\, f(x) - \mathbf{i}\mathrm{Re}\,(\mathbf{i}f(x)) \\ &= u(x) - \mathbf{i}\mathrm{Re}\,(f(\mathbf{i}x)) = u(x) - \mathbf{i}u(\mathbf{i}x). \end{aligned}$$ □

Lemma B.49. Let V be a complex vector space and $u : V \to \mathbb{R}$ a real linear functional on V. Then the function $f : V \to \mathbb{C}$ defined by (B.8) is a complex linear functional.

Proof. The additivity (LNM-1) of f follows from the additivity of u and (B.8). For any $c \in \mathbb{R}$, we have $f(cx) = cf(x)$ from (B.8). The rest follows from the additivity of f and

$$\begin{aligned} f(\mathbf{i}x) &= u(\mathbf{i}x) - \mathbf{i}u(\mathbf{i}^2x) \\ &= u(\mathbf{i}x) + \mathbf{i}u(x) = \mathbf{i}(u(x) - \mathbf{i}u(\mathbf{i}x)) = \mathbf{i}f(x). \end{aligned}$$ □

B.2.1 Null spaces and ranges

Definition B.50. The *null space* of a linear map $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ is the subset of $\mathcal{V}$ consisting of those vectors that T maps to the additive identity $\mathbf{0}$:

$$\mathrm{null}\, T = \{\mathbf{v} \in \mathcal{V} : T\mathbf{v} = \mathbf{0}\}. \tag{B.9}$$

Example B.51. The null space of the differentiation map in Example B.43 is $\mathbb{R}$.

Theorem B.52. A linear map $T \in \mathcal{L}(V, W)$ is injective if and only if $\mathrm{null}\, T = \{\mathbf{0}\}$.

Definition B.53. The *range* of a linear map $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ is the subset of $\mathcal{W}$ consisting of those vectors that are of the form $T\mathbf{v}$ for some $\mathbf{v} \in \mathcal{V}$:

$$\mathrm{range}\ T = \{T\mathbf{v} : \mathbf{v} \in \mathcal{V}\}. \tag{B.10}$$

Example B.54. The range of $A \in \mathbb{C}^{m\times n}$ is the span of its column vectors.

Theorem B.55. The range of a linear map $T \in \mathcal{L}(V, W)$ is a subspace of W.

Theorem B.56 (The counting theorem or the fundamental theorem of linear maps)**.** If $\mathcal{V}$ is a finite-dimensional vector space and $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, then range T is a finite-dimensional subspace of $\mathcal{W}$ and

$$\dim \mathcal{V} = \dim \mathrm{null}\ T + \dim \mathrm{range}\ T. \tag{B.11}$$

Theorem B.57. For an operator $T \in \mathcal{L}(\mathcal{V})$ on a finite-dimensional vector space $\mathcal{V}$, the following are equivalent:

(a) T is invertible;

(b) T is injective;

(c) T is surjective.

B.2.2 The matrix of a linear map

Definition B.58. The *matrix of a linear map* $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ *with respect to the bases* $(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n)$ of $\mathcal{V}$ and $(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m)$ of $\mathcal{W}$, denoted by

$$M_T := M(T, (\mathbf{v}_1, \ldots, \mathbf{v}_n), (\mathbf{w}_1, \ldots, \mathbf{w}_m)), \tag{B.12}$$

is the $m \times n$ matrix $A(T)$ whose entries $a_{i,j} \in \mathbb{F}$ satisfy the linear system

$$\forall j = 1, 2, \ldots, n, \qquad T\mathbf{v}_j = \sum_{i=1}^{m} a_{i,j}\mathbf{w}_i. \tag{B.13}$$

Corollary B.59. The matrix M_T in (B.12) of a linear map $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ satisfies

$$T[\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n] = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m]M_T. \tag{B.14}$$

Proof. This follows directly from (B.13). □

B.2.3 Duality

Dual vector spaces

Definition B.60. The *dual space* of a vector space V is the vector space of all linear functionals on V,

$$V' = \mathcal{L}(V, \mathbb{F}). \tag{B.15}$$

Definition B.61. For a basis $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of a vector space V, its *dual basis* is the list $\varphi_1, \ldots, \varphi_n$ where each $\varphi_j \in V'$ is

$$\varphi_j(\mathbf{v}_k) = \begin{cases} 1 & \text{if } k = j, \\ 0 & \text{if } k \neq j. \end{cases} \tag{B.16}$$

Exercise B.62. Show that the dual basis is a basis of the dual space.

Lemma B.63. A finite-dimensional vector space V satisfies

$$\dim V' = \dim V. \tag{B.17}$$

Proof. This follows from Definition B.60 and the identity $\dim \mathcal{L}(V, W) = \dim(V)\dim(W)$. □

Definition B.64. The *double dual space* of a vector space V, denoted by V'', is the dual space of V'.

Lemma B.65. The function $\Lambda : V \to V''$ defined as

$$\forall v \in V, \forall \varphi \in V', \qquad (\Lambda v)(\varphi) = \varphi(v) \tag{B.18}$$

is a linear bijection.

Proof. It is easily verified that Λ is a linear map. The rest follows from Definitions B.60, B.64, and Lemma B.63. □

Dual linear maps

Definition B.66. The *dual map* of a linear map $T : V \to W$ is the linear map $T' : W' \to V'$ defined as

$$\forall \varphi \in W', \qquad T'(\varphi) = \varphi \circ T. \tag{B.19}$$

Exercise B.67. Denote by D the linear map of differentiation $Dp = p'$ on the vector space $\mathcal{P}(\mathbb{R})$ of polynomials with real coefficients. Under the dual map of D, what is the image of the linear functional $\varphi(p) = \int_0^1 p$ on $\mathcal{P}(\mathbb{R})$?

Theorem B.68. The matrix of T' is the transpose of the matrix of T.

Proof. Let $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$, $(\varphi_1, \cdots, \varphi_n)$, $(\mathbf{w}_1, \cdots, \mathbf{w}_m)$, $(\psi_1, \cdots, \psi_m)$, be bases of V, V', W, W', respectively. Denote by A and C the matrices of $T : V \to W$ and $T' : W' \to V'$, respectively. We have

$$\psi_j \circ T = T'(\psi_j) = \sum_{r=1}^{n} c_{r,j}\varphi_r.$$

By Corollary B.59, applying this equation to $\mathbf{v}_k$ yields

$$(\psi_j \circ T)(\mathbf{v}_k) = \sum_{r=1}^{n} c_{r,j}\varphi_r(\mathbf{v}_k) = c_{k,j}.$$

On the other hand, we have

$$\begin{aligned}(\psi_j \circ T)(\mathbf{v}_k) &= \psi_j(T\mathbf{v}_k) = \psi_j\left(\sum_{r=1}^{m} a_{r,k}w_r\right) \\ &= \sum_{r=1}^{m} a_{r,k}\psi_j(w_r) = a_{j,k}.\end{aligned}$$

□

Definition B.69. The *double dual map* of a linear map $T : V \to W$ is the linear map $T'' : V'' \to W''$ defined as $T'' = (T')'$.

Theorem B.70. For $T \in \mathcal{L}(V)$ and Λ in (B.18), we have

$$T'' \circ \Lambda = \Lambda \circ T. \tag{B.20}$$

Proof. Definition B.69 and equation (B.18) yields

$$\begin{aligned}&\forall v \in V, \ \forall \varphi \in V', \\ (T'' \circ \Lambda)v\varphi &= ((T')'\Lambda v)\varphi = (\Lambda v \circ T')\varphi = \Lambda v(T'\varphi) \\ &= (T'\varphi)(v) = \varphi(Tv) = \Lambda(Tv)(\varphi) \\ &= (\Lambda \circ T)v\varphi,\end{aligned}$$

where the third step is natural since T' send V' to V'. □

Corollary B.71. For $T \in \mathcal{L}(V)$ where V is finite-dimensional, the double dual map is

$$T'' = \Lambda \circ T \circ \Lambda^{-1}. \tag{B.21}$$

Proof. This follows directly from Theorem B.70 and Lemma B.65. □

The null space and range of the dual of a linear map

Definition B.72. For $U \subset V$, the *annihilator* of U, denoted U^0, is defined by

$$U^0 := \{\varphi \in V' : \forall \mathbf{u} \in U, \ \varphi(\mathbf{u}) = 0\}. \tag{B.22}$$

Exercise B.73. Let $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4, \mathbf{e}_5$ denote the standard basis of $V = \mathbb{R}^5$, and $\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5$ its dual basis of V'. Suppose

$$U = \operatorname{span}(\mathbf{e}_1, \mathbf{e}_2) = \{(x_1, x_2, 0, 0, 0) \in \mathbb{R}^5 : x_1, x_2 \in \mathbb{R}\}.$$

Show that $U^0 = \operatorname{span}(\varphi_3, \varphi_4, \varphi_5)$.

Exercise B.74. Let $i : U \hookrightarrow V$ be an inclusion. Show that $\operatorname{null} i' = U^0$.

Lemma B.75. Suppose $U \subset V$. Then U^0 is a subspace of V'.

Exercise B.76. Suppose V is finite-dimensional. Prove every linear map on a subspace of V can be extended to a linear map on V.

Lemma B.77. Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U + \dim U^0 = \dim V. \tag{B.23}$$

Proof. Apply Theorem B.56 to the dual of an inclusion $i' : V' \to U'$ and we have

$$\begin{aligned} &\dim \operatorname{range} i' + \dim \operatorname{null} i' = \dim V' \\ \Rightarrow &\dim \operatorname{range} i' + \dim U^0 = \dim V, \end{aligned}$$

where the second line follows from Exercise B.74 and Lemma B.63. For any $\varphi \in U'$, Exercise B.76 states that $\varphi \in U'$ can be extended to $\psi \in V'$ such that $i'(\psi) = \varphi$. Hence i' is surjective and we have $U' = \operatorname{range} i'$. The proof is then completed by Lemma B.63. $\square$

Lemma B.78. Any linear map $T \in \mathcal{L}(V, W)$ satisfies

$$\operatorname{null} T' = (\operatorname{range} T)^0. \tag{B.24}$$

Proof. Definitions B.50, B.53, B.66, and B.72 yield

$$\begin{aligned} \varphi \in \operatorname{null} T' &\Leftrightarrow 0 = T'(\varphi) = \varphi \circ T \\ &\Leftrightarrow \forall v \in V, \varphi(Tv) = 0 \\ &\Leftrightarrow \varphi(\operatorname{range} T) = 0 \\ &\Leftrightarrow \varphi \in (\operatorname{range} T)^0. \end{aligned}$$ $\square$

Lemma B.79. For finite-dimensional vector spaces V and W, any linear map $T \in \mathcal{L}(V, W)$ satisfies

$$\dim \operatorname{null} T' = \dim \operatorname{null} T + \dim W - \dim V. \tag{B.25}$$

Proof. Lemma B.78 and Theorem B.56 yield

$$\begin{aligned} \dim \operatorname{null} T' &= \dim(\operatorname{range} T)^0 = \dim W - \dim(\operatorname{range} T) \\ &= \dim W - \dim V + \dim(\operatorname{null} T) \\ &= \dim \operatorname{null} T + \dim W - \dim V. \end{aligned}$$ $\square$

Corollary B.80. For finite-dimensional vector spaces V and W, any linear map $T \in \mathcal{L}(V, W)$ is surjective if and only if T' is injective.

Proof. T is surjective $\Leftrightarrow$ $W = \operatorname{range} T$ $\Leftrightarrow$ $(\operatorname{range} T)^0 = \{0\}$ $\Leftrightarrow$ $\operatorname{null} T' = \{0\}$ $\Leftrightarrow$ T' is injective. The second step follows from Lemma B.77 applied to W:

$$\dim W = \dim(\operatorname{range} T) + \dim(\operatorname{range} T)^0.$$ $\square$

Lemma B.81. For finite-dimensional vector spaces V and W, any linear map $T \in \mathcal{L}(V, W)$ satisfies

$$\dim \operatorname{range} T' = \dim \operatorname{range} T. \tag{B.26}$$

Proof. Theorem B.56, Lemma B.78, and Lemma B.77 yield

$$\begin{aligned} \dim \operatorname{range} T' &= \dim W - \dim \operatorname{null} T' \\ &= \dim W - \dim(\operatorname{range} T)^0 \\ &= \dim(\operatorname{range} T). \end{aligned}$$ $\square$

Lemma B.82. For finite-dimensional vector spaces V and W, any linear map $T \in \mathcal{L}(V, W)$ satisfies

$$\operatorname{range} T' = (\operatorname{null} T)^0. \tag{B.27}$$

Proof. Theorem B.56, Lemma B.78, and Lemma B.77 yield

$$\begin{aligned} \varphi \in \operatorname{range} T' &\Rightarrow \exists \psi \in W' \text{ s.t } T'(\psi) = \varphi \\ &\Rightarrow \forall v \in \operatorname{null} T, \varphi(v) = \psi(Tv) = 0 \\ &\Rightarrow \varphi \in (\operatorname{null} T)^0. \end{aligned}$$

The proof is completed by

$$\begin{aligned} \dim \operatorname{range} T' &= \dim(\operatorname{range} T) \\ &= \dim V - \dim \operatorname{null} T \\ &= \dim(\operatorname{null} T)^0. \end{aligned}$$ $\square$

Corollary B.83. For finite-dimensional vector spaces V and W, any linear map $T \in \mathcal{L}(V, W)$ is injective if and only if T' is surjective.

Proof. T is injective $\Leftrightarrow$ $\operatorname{null} T = \{0\}$ $\Leftrightarrow$ $(\operatorname{null} T)^0 = V'$ $\Leftrightarrow$ $\operatorname{range} T' = V'$ $\Leftrightarrow$ T' is surjective. The second step follows from Lemmas B.77 and B.63, and the third step follows from Lemma B.82. $\square$

Matrix ranks

Definition B.84. For a matrix $A \in \mathbb{F}^{m \times n} : \mathbb{F}^n \to \mathbb{F}^m$, its *column space* (or range or image) consists of all linear combinations of its columns, its *row space* (or coimage) is the column space of A^T, its *null space* (or kernel) is the null space of A as a linear operator, and the *left null space* (or cokernel) is the null space of A^T.

Definition B.85. The *column rank* and *row rank* of a matrix $A \in \mathbb{F}^{m \times n}$ is the dimension of its column space and row space, respectively.

Lemma B.86. Let A_T denote the matrix of a linear operator $T \in \mathcal{L}(V, W)$. Then the column rank of A_T is the dimension of $\operatorname{range} T$.

Proof. For $\mathbf{u} = \sum_i c_i \mathbf{v}_i$, Corollary B.59 yields

$$T\mathbf{u} = \sum_i c_i T\mathbf{v}_i = T[\mathbf{v}_1, \ldots, \mathbf{v}_n]\mathbf{c} = [\mathbf{w}_1, \ldots, \mathbf{w}_m]A_T\mathbf{c}.$$

Hence we have

$$\{T\mathbf{u} : \mathbf{c} \in \mathbb{F}^n\} = \{[\mathbf{w}_1, \ldots, \mathbf{w}_m]A_T\mathbf{c} : \mathbf{c} \in \mathbb{F}^n\}.$$

The LHS is range$T$ while $\{A_T\mathbf{c} : \mathbf{c} \in \mathbb{F}^n\}$ is the column space of A_T. Since $(\mathbf{w}_1, \ldots, \mathbf{w}_m)$ is a basis, by Definition B.85 the column rank of the matrix $[\mathbf{w}_1, \ldots, \mathbf{w}_m]$ is m. Taking dim to both sides of the above equation yields the conclusion. Note that the RHS is a subspace of $\mathbb{F}^m$ (why?) and the dimension of it does not depend on the special choice of its basis, hence we can choose $(\mathbf{w}_1, \ldots, \mathbf{w}_m)$ to be the standard basis and then $[\mathbf{w}_1, \ldots, \mathbf{w}_m]$ is simply the identity matrix. □

Theorem B.87. For any $A \in \mathbb{F}^{m\times n}$, its row rank equals its column rank.

Proof. Define a linear map $T : \mathbb{F}^n \to \mathbb{F}^m$ as $T\mathbf{x} = A\mathbf{x}$. Clearly, A is the matrix of T for the standard bases of $\mathbb{F}^n$ and $\mathbb{F}^m$. Then we have,

$$\begin{aligned}
\text{column rank of } A &= \dim \operatorname{range} T \\
&= \dim \operatorname{range} T' \\
&= \text{column rank of the matrix of } T' \\
&= \text{column rank of } A^T \\
&= \text{row rank of } A,
\end{aligned}$$

where the first step follows from Lemma B.86, the second from Lemma B.81, the third from Lemma B.86, the fourth from Theorem B.68, and the last from the definition of matrix transpose and matrix products. □

Definition B.88. The *rank* of a matrix is its column rank.

Theorem B.89 (Fundamental theorem of linear algebra)**.** For a matrix $A \in \mathbb{F}^{m\times n} : \mathbb{F}^n \to \mathbb{F}^m$, its column space and row space both have dimension $r \le \min(m, n)$; its null space and left null space have dimensions $n - r$ and $m - r$, respectively. In addition, we have

$$\mathbb{F}^m = \operatorname{range} A \oplus \operatorname{null} A^T, \tag{B.28a}$$

$$\mathbb{F}^n = \operatorname{range} A^T \oplus \operatorname{null} A, \tag{B.28b}$$

where range$A \perp$ nullA^T and range$A^T \perp$ null$A$.

Proof. The first sentence is a rephrase of Theorem B.87 and follows from Theorem B.56. For the second sentence, we only prove (B.28b). $\mathbf{x} \in$ null$A$ implies $\mathbf{x} \in \mathbb{F}^n$ and $A\mathbf{x} = \mathbf{0}$. The latter expands to

$$\begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \cdots \\ \mathbf{a}_m^T \end{bmatrix} \mathbf{x} = \begin{bmatrix} 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix},$$

which implies that $\forall j = 1, 2, \ldots, m$, $\mathbf{a}_j \perp \mathbf{x}$. Hence $\mathbf{x}$ is orthogonal to each basis vector of rangeA^T. The rest of the proof follows from Lemma B.81, Theorem B.68, Theorem B.56. □

B.3 Eigenvalues, eigenvectors, and invariant subspaces

B.3.1 Invariant subspaces

Definition B.90. Under a linear operator $T \in \mathcal{L}(\mathcal{V})$, a subspace $\mathcal{U}$ of $\mathcal{V}$ is *invariant* if $\mathbf{u} \in \mathcal{U}$ implies $T\mathbf{u} \in \mathcal{U}$.

Example B.91. Under $T \in \mathcal{L}(\mathcal{V})$, each of the following subspaces of $\mathcal{V}$ is invariant: $\{\mathbf{0}\}$, $\mathcal{V}$, null T, and range T.

Definition B.92. A number $\lambda \in \mathbb{F}$ is called an *eigenvalue of an operator* $T \in \mathcal{L}(\mathcal{V})$ if there exists $\mathbf{v} \in \mathcal{V}$ such that $T\mathbf{v} = \lambda\mathbf{v}$ and $\mathbf{v} \neq \mathbf{0}$. Then the vector $\mathbf{v}$ is called an *eigenvector* of T corresponding to λ.

Lemma B.93. Suppose V is finite-dimensional. $\lambda \in V$ is an eigenvalue of $T \in \mathcal{L}(V)$ if and only if $T - \lambda I$ is not injective.

Proof. This follows directly from Definition B.92. □

Example B.94. For each eigenvector $\mathbf{v}$ of $T \in \mathcal{L}(\mathcal{V})$, the subspace span$(\mathbf{v})$ is a one-dimensional invariant subspace of $\mathcal{V}$.

Lemma B.95. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of $T \in \mathcal{L}(\mathcal{V})$ with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_m$. Then $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is linearly independent.

Lemma B.96. Suppose $\mathcal{V}$ is finite-dimensional. Then each operator on $\mathcal{V}$ has at most $\dim \mathcal{V}$ distinct eigenvalues.

Definition B.97. Suppose $T \in \mathcal{L}(V)$ and U is an invariant subspace of V under T. The *restriction operator* $T|_U \in \mathcal{L}(U)$ is defined by

$$\forall \mathbf{u} \in U, \quad T|_U(\mathbf{u}) = T\mathbf{u}. \tag{B.29}$$

B.3.2 Existence of eigenvalues

Notation 27. Suppose $T \in \mathcal{L}(\mathcal{V})$ and $p \in \mathbb{P}(\mathbb{F})$ is a polynomial given by

$$p(z) = a_0 + a_1 z + \cdots + a_m z^m$$

for $z \in \mathbb{F}$. Then $p(T)$ is the operator given by

$$p(T) = a_0 I + a_1 T + \cdots + a_m T^m,$$

where $I = T^0$ is the identity operator.

Example B.98. Suppose $D \in \mathcal{L}(\mathbb{P}(\mathbb{R}))$ is the differentiation operator defined by $Dq = q'$ and p is the polynomial defined by $p(x) = 7 - 3x + 5x^2$. Then we have

$$p(D) = 7 - 3D + 5D^2, \qquad (p(D))q = 7q - 3q' + 5q''.$$

Definition B.99. The *product polynomial* of two polynomials $p, q \in \mathbb{P}(\mathbb{F})$ is the polynomial defined by

$$\forall z \in \mathbb{F}, \qquad (pq)(z) := p(z)q(z). \tag{B.30}$$

Lemma B.100. Any $T \in \mathcal{L}(\mathcal{V})$ and $p, q \in \mathbb{P}(\mathbb{F})$ satisfy

$$(pq)(T) = p(T)q(T) = q(T)p(T). \tag{B.31}$$

Theorem B.101 (Existence of eigenvalues)**.** Every operator $T \in \mathcal{L}(V)$ on a finite-dimensional, nonzero, complex vector space V has an eigenvalue.

Proof. Write $n := \dim V$. For a nonzero $\mathbf{v} \in V$, the $n+1$ vectors $(\mathbf{v}, T\mathbf{v}, T\mathbf{v}, \cdots, T^n\mathbf{v})$ must be linear dependent, i.e.,

$$\mathbf{0} = a_0\mathbf{v} + a_1 T\mathbf{v} + \cdots + a_n T^n \mathbf{v} = (a_0 + a_1 T + \cdots + a_n T^n)\mathbf{v}$$

implies that there exists $j \in [1, n]$ such that $a_j \neq 0$. By the fundamental theorem of algebra, the polynomial $\sum_{i=0}^{n} a_i T^i$ has m roots, say, $\lambda_1, \ldots, \lambda_m$, and thus for some $c \in \mathbb{C} \setminus \{0\}$, we have

$$\mathbf{0} = c(T - \lambda_1 I)(T - \lambda_2 I) \cdots (T - \lambda_m I)\mathbf{v}.$$

Hence $T - \lambda_j I$ is not injective for some λ_j and the proof is completed by Lemma B.93. □

B.3.3 Upper-triangular matrices

Definition B.102. The *matrix of a linear operator* $T \in \mathcal{L}(\mathcal{V})$ is the matrix of the linear map $T \in \mathcal{L}(\mathcal{V}, \mathcal{V})$, c.f. Definition B.58.

Theorem B.103. Suppose $T \in \mathcal{L}(\mathcal{V})$ and $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of $\mathcal{V}$. Then the following are equivalent:

(a) the matrix of T with respect to $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is upper triangular;

(b) $T\mathbf{v}_j \in \operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_j)$ for each $j = 1, \ldots, n$;

(c) $\operatorname{span}(\mathbf{v}_1, \ldots, \mathbf{v}_j)$ is invariant under T for each $j = 1, \ldots, n$.

Theorem B.104. Every linear operator $T \in \mathcal{L}(\mathcal{V})$ on a finite-dimensional complex vector space $\mathcal{V}$ has an upper-triangular matrix with respect to some basis of $\mathcal{V}$.

Theorem B.105. Suppose $T \in \mathcal{L}(\mathcal{V})$ has an upper-triangular matrix with respect to some basis of $\mathcal{V}$. Then T is invertible if and only if all the entries on the diagonal of that upper-triangular matrix are nonzero.

Theorem B.106. Suppose $T \in \mathcal{L}(\mathcal{V})$ has an upper-triangular matrix with respect to some basis of $\mathcal{V}$. Then the eigenvalues of T are precisely the entries on the diagonal of that upper-triangular matrix.

B.3.4 Eigenspaces and diagonal matrices

Definition B.107. A *diagonal entry* of a matrix is an entry of the matrix of which the row index equals the column index. The *diagonal* of a matrix consists of all diagonal entries of the matrix. A *diagonal matrix* is a square matrix that is zero everywhere except possibly along the diagonal.

Definition B.108. The *eigenspace* of $T \in \mathcal{L}(\mathcal{V})$ corresponding to $\lambda \in \mathbb{F}$ is

$$E(\lambda, T) := \operatorname{null}(T - \lambda I). \tag{B.32}$$

Lemma B.109. Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of $T \in \mathcal{L}(\mathcal{V})$ on a finite-dimensional space $\mathcal{V}$. Then

$$E(\lambda_1, T) + \cdots + E(\lambda_m, T)$$

is a direct sum and

$$\dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T) \leq \dim \mathcal{V}. \tag{B.33}$$

Definition B.110. An operator $T \in \mathcal{L}(\mathcal{V})$ is *diagonalizable* if it has a diagonal matrix with respect to some basis of $\mathcal{V}$.

Theorem B.111 (Conditions of diagonalizability)**.** Suppose $\lambda_1, \ldots, \lambda_m$ are distinct eigenvalues of $T \in \mathcal{L}(\mathcal{V})$ on a finite-dimensional space $\mathcal{V}$. Then the following are equivalent:

(a) T is diagonalizable;

(b) $\mathcal{V}$ has a basis consisting of eigenvectors of T;

(c) there exist one-dimensional subspaces $U_1, \ldots, U_n$ of $\mathcal{V}$, each invariant under T, such that $\mathcal{V} = U_1 \oplus \cdots \oplus U_n$;

(d) $\mathcal{V} = E(\lambda_1, T) \oplus \cdots \oplus E(\lambda_m, T)$;

(e) $\dim \mathcal{V} = \dim E(\lambda_1, T) + \cdots + \dim E(\lambda_m, T)$.

Corollary B.112. An operator $T \in \mathcal{L}(\mathcal{V})$ is diagonalizable if T has $\dim \mathcal{V}$ distinct eigenvalues.

B.4 Operators on complex vector spaces

B.4.1 Generalized eigenvectors

Lemma B.113. For a linear operator $T \in \mathcal{L}(V)$, we have

$$\{\mathbf{0}\} = \operatorname{null} T^0 \subseteq \operatorname{null} T^1 \subseteq \cdots \subseteq \operatorname{null} T^k \subseteq \operatorname{null} T^{k+1} \subseteq \cdots. \tag{B.34}$$

Proof. Suppose $\mathbf{v} \in \operatorname{null} T^k$ for $k \in \mathbb{N}$. Thus $T^k\mathbf{v} = 0$. Then $T^{k+1}\mathbf{v} = TT^k\mathbf{v} = T\mathbf{0} = \mathbf{0}$ and therefore $\mathbf{v} \in \operatorname{null} T^{k+1}$. □

Lemma B.114. Suppose a linear operator $T \in \mathcal{L}(V)$ satisfies $\operatorname{null} T^m = \operatorname{null} T^{m+1}$. Then we have

$$\operatorname{null} T^m = \operatorname{null} T^{m+1} = \operatorname{null} T^{m+2} = \cdots. \tag{B.35}$$

Proof. By Lemma B.113, it suffices to show

$$\forall k \in \mathbb{N}^+, \quad \operatorname{null} T^{m+k+1} \subseteq \operatorname{null} T^{m+k},$$

which indeed holds because

$$\begin{aligned}
\mathbf{v} \in \operatorname{null} T^{m+k+1} \Rightarrow\ & T^{m+k+1}\mathbf{v} = \mathbf{0} \Rightarrow T^{m+1}(T^k\mathbf{v}) = \mathbf{0} \\
\Rightarrow\ & T^k\mathbf{v} \in \operatorname{null} T^{m+1} = \operatorname{null} T^m \\
\Rightarrow\ & T^m T^k \mathbf{v} = \mathbf{0} \Rightarrow T^{m+k}\mathbf{v} = \mathbf{0} \\
\Rightarrow\ & \mathbf{v} \in \operatorname{null} T^{m+k}. \quad \square
\end{aligned}$$

Lemma B.115. A linear operator $T \in \mathcal{L}(V)$ satisfies

$$\operatorname{null} T^n = \operatorname{null} T^{n+1} = \operatorname{null} T^{n+2} = \cdots, \tag{B.36}$$

where $n = \dim V$.

Proof. By Lemma B.114, it suffices to show

$$\operatorname{null} T^n = \operatorname{null} T^{n+1}.$$

Suppose this is not true. Then Lemmas B.113 and B.114 yield

$$\{\mathbf{0}\} = \operatorname{null} T^0 \subset \operatorname{null} T^1 \subset \cdots \subset \operatorname{null} T^n \subset \operatorname{null} T^{n+1},$$

where the symbol"⊂" means strict inclusion, c.f. Definition A.2. At each strict inclusion in the above chain, the dimension of the space increases by at least 1, and thus $\dim \operatorname{null} T^{n+1} > n$. But the dimension of any subspace of V cannot exceed that of V. □

Theorem B.116. A linear operator $T \in \mathcal{L}(V)$ satisfies

$$V = \operatorname{null} T^n \oplus \operatorname{range} T^n \tag{B.37}$$

where $n = \dim V$.

Proof. We first show $\operatorname{null} T^n \cap \operatorname{range} T^n = \{\mathbf{0}\}$. Indeed, if $\mathbf{v} \in \operatorname{null} T^n \cap \operatorname{range} T^n$, then $T^n\mathbf{v} = \mathbf{0}$ and there exists $\mathbf{u}$ such that $\mathbf{v} = T^n\mathbf{u}$. Hence $T^{2n}\mathbf{u} = \mathbf{0}$. Lemma B.115 further implies $T^n\mathbf{u} = \mathbf{0}$ and thus $\mathbf{v} = \mathbf{0}$.

By Theorem B.13, $\operatorname{null} T^n + \operatorname{range} T^n$ is a direct sum. Then (B.37) follows from

$$\begin{aligned}\dim(\operatorname{null} T^n \oplus \operatorname{range} T^n) &= \dim \operatorname{null} T^n + \dim \operatorname{range} T^n \\ &= \dim V,\end{aligned}$$

where the second step follows from the fundamental theorem of linear maps (Theorem B.56). □

Example B.117. For the operator $T \in \mathcal{L}(\mathbb{C}^3)$ given by

$$T(z_1, z_2, z_3) = (4z_2, 0, 5z_3), \tag{B.38}$$

$\operatorname{null} T + \operatorname{range} T$ is not a direct sum of $\mathbb{C}^3$ because

$$\begin{aligned}&\operatorname{null} T = \{(z_1, 0, 0) : z_1 \in \mathbb{C}\},\\ &\operatorname{range} T = \{(z_1, 0, z_3) : z_1, z_3 \in \mathbb{C}\}.\end{aligned}$$

In contrast, $T^3(z_1, z_2, z_3) = (0, 0, 125z_3)$ and thus

$$\begin{aligned}&\operatorname{null} T^3 = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbb{C}\},\\ &\operatorname{range} T^3 = \{(0, 0, z_3) : z_3 \in \mathbb{C}\},\\ &\operatorname{null} T^3 \oplus \operatorname{range} T^3 = \mathbb{C}^3.\end{aligned}$$

Definition B.118. A *generalized eigenvector of a linear operator* $T \in \mathcal{L}(V)$ corresponding to the eigenvalue λ of T is a nonzero vector $\mathbf{v} \in V$ satisfying

$$\exists j \in \mathbb{N}^+ \text{ s.t. } (T - \lambda I)^j\mathbf{v} = \mathbf{0}. \tag{B.39}$$

Definition B.119. The *generalized eigenspace of a linear operator* $T \in \mathcal{L}(V)$ corresponding to the eigenvalue λ of T, denoted $G(\lambda, T)$, is the set of all generalized eigenvectors of T corresponding to λ along with the zero vector.

Lemma B.120. A generalized eigenspace $G(\lambda, T)$ satisfies

$$\forall T \in \mathcal{L}(V), \forall \lambda \in \mathbb{F},\ G(\lambda, T) = \operatorname{null}(T - \lambda I)^{\dim V}. \tag{B.40}$$

Proof. Suppose $\mathbf{v} \in \operatorname{null}(T - \lambda I)^{\dim V}$. Then Definitions B.118 and B.119 imply $\mathbf{v} \in G(\lambda, T)$. Conversely, $\mathbf{v} \in G(\lambda, T)$ implies that $(T - \lambda I)^j\mathbf{v} = \mathbf{0}$ for some $j \in \mathbb{N}^+$. Then we have $\mathbf{v} \in \operatorname{null}(T - \lambda I)^{\dim V}$ from Lemmas B.113 and B.115. □

Definition B.121. The *multiplicity or algebraic multiplicity of an eigenvalue* λ of an operator T is the dimension of the corresponding generalized eigenspace,

$$m_a(\lambda) := \dim G(\lambda, T) = \dim \operatorname{null}(T - \lambda I)^{\dim V} \tag{B.41}$$

while the *geometric multiplicity of an eigenvalue* λ of T is the dimension of the corresponding eigenspace,

$$m_g(\lambda) := \dim E(\lambda, T) = \dim \operatorname{null}(T - \lambda I). \tag{B.42}$$

The *index of an eigenvalue* λ of T is the smallest integer k for which

$$\operatorname{null}\,(T - \lambda I)^k = \operatorname{null}\,(T - \lambda I)^{k+1}. \tag{B.43}$$

Corollary B.122. Geometric multiplicity and algebraic multiplicity satisfy

$$1 \le m_g(\lambda) \le m_a(\lambda). \tag{B.44}$$

Proof. The case $j = 1$ in Definition B.118 implies that every eigenvector of T is a generalized eigenvector of T. Hence

$$\forall T \in \mathcal{L}(V),\ \forall \lambda \in \mathbb{F},\ E(\lambda, T) \subseteq G(\lambda, T).$$

Then Definition B.121 completes the proof. □

Definition B.123. An eigenvalue λ of A is *defective* iff

$$m_g(\lambda) < m_a(\lambda). \tag{B.45}$$

A is *defective* iff A has one or more defective eigenvalues.

Example B.124. Eigenvalues of the operator $T \in \mathcal{L}(\mathbb{C}^3)$ in (B.38) are 0 and 5, with the corresponding eigenspaces as

$$\begin{aligned}E(0, T) &= \{(z_1, 0, 0) : z_1 \in \mathbb{C}\},\\ E(5, T) &= \{(0, 0, z_3) : z_3 \in \mathbb{C}\}.\end{aligned}$$

The generalized eigenspaces are deduced as follows,

$$\begin{aligned}&T^3(z_1, z_2, z_3) = (0, 0, 125z_3),\\ &G(0, T) = \{(z_1, z_2, 0) : z_1, z_2 \in \mathbb{C}\},\\ &(T - 5I)^3(z_1, z_2, z_3) = (-125z_1 + 300z_2, -125z_2, 0),\\ &G(5, T) = E(5, T) = \{(0, 0, z_3) : z_3 \in \mathbb{C}\}.\end{aligned}$$

Since $m_g(5) = 1 = m_a(5)$ and $m_g(0) = 1 < m_a(0) = 2$, the eigenvalue 5 is not defective but the eigenvalue 0 is. Hence the operator T is defective.

Lemma B.125. Generalized eigenvectors of distinct eigenvalues of an operator $T \in \mathcal{L}(V)$ are linearly independent.

Proof. Let $(\lambda_i, \mathbf{v}_i)$ be a generalized eigenpair of T. Define

$$\mathbf{w} := (T - \lambda_1 I)^k \mathbf{v}_1,$$

where $k \in \mathbb{N}$ is the largest integer such that $\mathbf{w} \neq \mathbf{0}$. By Definition B.118, $(T - \lambda_1 I)\mathbf{w} = \mathbf{0}$ and thus $\mathbf{w}$ is an eigenvector of T. Consequently, we have

$$(*) : \quad \forall j \in \mathbb{N}^+, \ \forall \lambda \in \mathbb{C}, \quad (T - \lambda I)^j \mathbf{w} = (\lambda_1 - \lambda)^j \mathbf{w}.$$

Write $n := \dim V$ and define a polynomial of T by

$$p(T) = (T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n;$$

By Lemma B.100, the factors of $p(T)$ commute.

Suppose $\mathbf{0} = \sum_{i=1}^m a_i \mathbf{v}_i$ where each $a_i \in \mathbb{C}$. Then the application of $p(T)$ to this equation yields $a_1 = 0$ because of the distinctness of the eigenvalues and

$$\begin{aligned} \mathbf{0} &= a_1 (T - \lambda_1 I)^k (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n \mathbf{v}_1 \\ &= a_1 (T - \lambda_2 I)^n \cdots (T - \lambda_m I)^n \mathbf{w} \\ &= a_1 (\lambda_1 - \lambda_2)^n \cdots (\lambda_1 - \lambda_m)^n \mathbf{w}, \end{aligned}$$

where the first equality follows from Lemma B.120, the second from the definition of $\mathbf{w}$, and the third from $(*)$.

Similarly, the above arguments applied to the other generalized eigenpairs yield $a_2 = \cdots = a_m = 0$, which implies the linear independence of generalized eigenvectors. □

B.4.2 Nilpotent operators

Definition B.126. An operator $N \in \mathcal{L}(V)$ is *nilpotent* iff $N^k = \mathbf{0}$ for some $k \in \mathbb{N}^+$.

Example B.127. The differentiation operator on the vector space of polynomials of degree at most m is nilpotent. The operator $N \in \mathcal{L}(\mathbb{F}^3)$ given by $N(x, y, z) = (y, z, 0)$ is nilpotent because $N^3 = \mathbf{0}$.

Lemma B.128. A nilpotent operator $N \in \mathcal{L}(V)$ satisfies

$$N^{\dim V} = \mathbf{0}. \tag{B.46}$$

Proof. Definitions B.126 and B.119 imply $G(0, N) = V$. The rest follows from Lemma B.120. □

Lemma B.129. Any nilpotent operator $N \in \mathcal{L}(V)$ has a strictly upper triangular matrix M, i.e., $\forall i \geq j$, $M_{i,j} = 0$.

Proof. Write $n := \dim V$. Choose a basis of $\operatorname{null} N$, extend it to a basis of $\operatorname{null} N^2$, …, $\operatorname{null} N^n$, and we have a matrix U whose columns are the vectors of this basis:

$$U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n].$$

Corollary B.59 states that

$$N[\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n] = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n] M.$$

The fact of $\mathbf{u}_1 \in \operatorname{null} N$ dictates that all entries in the first column of M must be zero. Let $\mathbf{u}_j$ be the first basis vector in $\operatorname{null} N^2$ but not in $\operatorname{null} N$. Then $N\mathbf{u}_j \in \operatorname{null} N$, i.e., $N\mathbf{u}_j$ must be a linear combination of $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{j-1}$. Hence the first $j-1$ entries in the jth column of M could be nonzero while all other entries must be zero. Proceeding in this fashion completes the proof. □

Lemma B.130. If an operator $N \in \mathcal{L}(V)$ is nilpotent, then $I + N$ has a square root, i.e., there exists an operator $M \in \mathcal{L}(V)$ such that $M^2 = I + N$.

Proof. By Definition B.126, $N^m = \mathbf{0}$ for some $m \in \mathbb{N}^+$. For operators of the form

$$M = I + \sum_{i=1}^{m-1} a_i N^i,$$

$M^2 = I + N$ yields $\left(I + \sum_{i=1}^{m-1} a_i N^i\right)^2 = I + N$, which is equivalent to

$$2a_1 = 1, \ 2a_2 + a_1^2 = 0, \ 2a_3 + 2a_1 a_2 = 0, \ \ldots,$$

where $a_1 = \frac{1}{2}$ and each a_j can be uniquely determined from $a_1, a_2, \ldots, a_{j-1}$. □

B.4.3 Operator decomposition

Lemma B.131. Suppose $p(T)$ is a polynomial of an operator $T \in \mathcal{L}(V)$. Then both $\operatorname{null} p(T)$ and $\operatorname{range} p(T)$ are invariant under T.

Proof. $\mathbf{v} \in \operatorname{null} p(T)$ implies $p(T)\mathbf{v} = \mathbf{0}$ and

$$p(T)(T\mathbf{v}) = T(p(T)\mathbf{v}) = T\mathbf{0} = \mathbf{0}.$$

Hence $T\mathbf{v} \in \operatorname{null} p(T)$, and, by Definition B.90, $\operatorname{null} p(T)$ is invariant under T. Similarly,

$$\mathbf{v} \in \operatorname{range} p(T) \ \Rightarrow \ \exists \mathbf{u} \in V \text{ s.t. } p(T)\mathbf{u} = \mathbf{v}$$

and thus

$$T\mathbf{v} = Tp(T)\mathbf{u} = p(T)(T\mathbf{u}).$$

Since $T\mathbf{u} \in V$, we have $T\mathbf{v} \in \operatorname{range} p(T)$ and, by Definition B.90, $\operatorname{range} p(T)$ is invariant under T. □

Theorem B.132 (Decomposing operators on complex vector spaces)**.** Suppose V is a complex vector space, $T \in \mathcal{L}(V)$, and $\lambda_1, \lambda_2, \ldots, \lambda_m$ are all the distinct eigenvalues of T. Then

(DOC-1) $V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T)$;

(DOC-2) each $G(\lambda_j, T)$ is invariant under T;

(DOC-3) each $(T - \lambda_j I)|_{G(\lambda_j, T)}$ is nilpotent.

Proof. Write $n := \dim V$. Lemma B.120 states that for each $j = 1, 2, \ldots, m$ we have $G(\lambda_j, T) = \operatorname{null}(T - \lambda_j I)^n$. Then (DOC-3) follows from Definition B.126 and (DOC-2) follows from Lemma B.131 by choosing $p(z) = (z - \lambda_j)^n$.

(DOC-1) can be proven by an induction on n. The induction basis of $n = 1$ clearly holds. As the induction hypothesis, (DOC-1) holds on all vector spaces of which the dimension is smaller than n.

By Theorem B.101, T has an eigenvalue λ_1. The application of Lemma B.116 to $T - \lambda_1 I$ yields

$$(*) : \quad V = G(\lambda_1, T) \oplus U,$$

where $U = \operatorname{range}(T - \lambda_1 I)^n$. If U is empty, (DOC-1) holds trivially; otherwise Lemma B.131 implies that U is invariant under T, furnishing the restriction operator $T|_U \in \mathcal{L}(U)$ in (B.29). By $(*)$ and Theorem B.101, $T|_U$ has distinct eigenvalues $\lambda_2, \ldots, \lambda_m$, each of which is different from λ_1. Since $\dim G(\lambda_1, T) \geq 1$, we have $\dim U < n$ and thus we can apply the induction hypothesis to U to obtain

$$(**): \quad U = G(\lambda_2, T|_U) \oplus \cdots \oplus G(\lambda_m, T|_U).$$

Combining $(**)$ with $(*)$ completes the proof if we can show

$$\forall j = 2, \ldots, m, \quad G(\lambda_j, T|_U) = G(\lambda_j, T).$$

Indeed, apply $\cap G(\lambda_j, T)$ to both sides of $(*)$ and we have

$$\begin{aligned} G(\lambda_j, T) &= (G(\lambda_1, T) \cap G(\lambda_j, T)) \oplus (U \cap G(\lambda_j, T)) \\ &= G(\lambda_j, T|_U), \end{aligned}$$

where the second step follows from Lemma B.125 and the identity $G(\lambda_j, T|_U) = G(\lambda_j, T) \cap U$. □

Example B.133. For the operator $T \in \mathcal{L}(\mathbb{C}^3)$ in (B.38), T does not have enough eigenvectors to span $\mathbb{C}^3$. Example B.124 shows that

$$\begin{aligned} E(0, T) \oplus E(5, T) \subset \mathbb{C}^3, \\ G(0, T) \oplus G(5, T) = \mathbb{C}^3. \end{aligned}$$

Corollary B.134. Suppose V is a complex vector space and $T \in \mathcal{L}(V)$. Then there is a basis of V that consists of generalized eigenvectors of T.

Proof. This follows from (DOC-1) in Theorem B.132. □

Definition B.135. A *block diagonal matrix* A is a square matrix of the form

$$\begin{pmatrix} A_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & A_m \end{pmatrix}, \tag{B.47}$$

where $A_1, \ldots, A_m$ are square matrices along the diagonal and all other entries of A is 0.

Theorem B.136. Suppose V is a complex vector space, $T \in \mathcal{L}(V)$, and $\lambda_1, \lambda_2, \ldots, \lambda_m$ are all the distinct eigenvalues of T with multiplicities $d_1, d_2, \ldots, d_m$. Then there is a basis of V with respect to which T has a block diagonal form (B.47) where each A_j is a d_j-by-d_j upper triangular matrix of the form

$$\begin{bmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{bmatrix}. \tag{B.48}$$

Proof. By Theorem B.132, each $(T - \lambda_j I)|_{G(\lambda_j,T)}$ is nilpotent. By Lemma B.129, we can choose a basis of $G(\lambda_j, T)$ such that the matrix of $(T - \lambda_j I)|_{G(\lambda_j,T)}$ is strictly upper triangular. Then the form of (B.48) follows from

$$T|_{G(\lambda_j,T)} = (T - \lambda_j I)|_{G(\lambda_j,T)} + \lambda_j I|_{G(\lambda_j,T)}.$$

The rest follows from (DOC-1) in Theorem B.132. □

Example B.137. For $T \in \mathcal{L}(\mathbb{F}^3)$ given by

$$T(x, y, z) = (6x + 3y + 4z, 6y + 2z, 7z), \tag{B.49}$$

the matrix of T with respect to the standard basis is

$$T_S = \begin{pmatrix} 6 & 3 & 4 \\ 0 & 6 & 2 \\ 0 & 0 & 7 \end{pmatrix}.$$

It is readily verified that the eigenvalues of T are 6 and 7, with the corresponding generalized eigenspaces as

$$\begin{aligned} G(6, T) &= \operatorname{span}\{(1, 0, 0), (0, 1, 0)\}; \\ G(7, T) &= \operatorname{span}\{(10, 2, 1)\}. \end{aligned}$$

The matrix of T with respect to the basis

$$\{\mathbf{v}_1 = (1, 0, 0), \mathbf{v}_2 = (0, 1, 0), \mathbf{v}_3 = (10, 2, 1)\}$$

is of the block diagonal form:

$$T_B = \begin{pmatrix} \begin{bmatrix} 6 & 3 \\ 0 & 6 \end{bmatrix} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \begin{matrix} 0 & 0 \end{matrix} & [7] \end{pmatrix}.$$

More precisely, by Corollary B.59, we have

$$T_S[\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3] T_B.$$

Corollary B.138. Any operator $T \in \mathcal{L}(V)$ on a complex vector space V can be decomposed as $T = \Lambda + N$ where Λ is diagonalizable, N is nilpotent, and $\Lambda N = N\Lambda$.

Proof. By Theorem B.136, T has a block diagonal form where the diagonal matrices are $T_1, T_2, \cdots, T_p$ and each T_j can be decomposed as $T_j = N_j + \Lambda_j$ with $N_j = T_j - \lambda_j I_j$ and $\Lambda_j = \lambda_j I_j$. Clearly N_j is nilpotent and Λ_j is diagonalizable. Also, any matrix commutes with the identity matrix, hence we have $N_j \lambda_j I_j = \lambda_j I_j N_j$. The rest follows from the block diagonal form of T. □

Example B.139. If p is a polynomial of degree k, then $p(a + x)$ can be expressed as a Taylor series,

$$p(a + x) = \sum_{i=0}^{k} \frac{p^{(i)}(a)}{i!} x^i.$$

The formula is an algebraic identity and can be generalized to an operator $T = \Lambda + N$ with the help of $\Lambda N = N\Lambda$,

$$p(\Lambda + N) = \sum_{i=0}^{k} \frac{p^{(i)}(\Lambda)}{i!} N^i = \sum_{i=0}^{m} \frac{p^{(i)}(\Lambda)}{i!} N^i,$$

where the nilpotent operator N satisfies $N^{m+1} = \mathbf{0}$.

The same approach works if p is not a polynomial, but an infinite power series with its radius of convergence as ∞. In particular, if $p(x) = e^x$, we have

$$e^T = \sum_{i=0}^{+\infty} \frac{e^\Lambda}{i!} N^i = e^\Lambda \sum_{i=0}^{m} \frac{1}{i!} N^i, \tag{B.50}$$

which can be adopted as a definition of matrix exponentials.

Theorem B.140. Any invertible operator $T \in \mathcal{L}(V)$ on a complex vector space V has a square root.

Proof. Let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be all distinct eigenvalues of T. By (DOC-3) in Theorem B.132, for each $j = 1, 2, \ldots, m$ there exists a niplotent operator $N_j \in \mathcal{L}(G(\lambda_j, T))$ such that $T|_{G(\lambda_j,T)} = \lambda_j I + N_j$. Since T is invertible, $\lambda_j \neq 0$ and

$$\forall j = 1, 2, \ldots, m, \quad T|_{G(\lambda_j,T)} = \lambda_j \left(I + \frac{N_j}{\lambda_j} \right).$$

By Lemma B.130 and the condition of $\mathbb{F} = \mathbb{C}$, there exists an operator $R_j \in \mathcal{L}(G(\lambda_j, T))$ such that $R_j^2 = T|_{G(\lambda_j,T)}$.

By (DOC-1) in Theorem B.132, any vector $\mathbf{v} \in V$ can be uniquely expressed in the form $\mathbf{v} = \sum_{i=1}^m \mathbf{u}_i$ where $\mathbf{u}_i \in G(\lambda_i, T)$. Then it is straightforward to verify that the operator $R \in \mathcal{L}(V)$ given below satisfies $R^2 = T$:

$$R(\mathbf{v}) = \sum_{j=1}^m R_j \mathbf{u}_j. \qquad \square$$

B.4.4 Jordan basis

Definition B.141. A *Jordan block* of order k has the form

$$J(\lambda, k) = \lambda I_k + S_k, \tag{B.51}$$

where

$$(S_k)_{i,j} = \begin{cases} 1, & i = j - 1, \\ 0, & \text{otherwise.} \end{cases}$$

Example B.142. The Jordan blocks of orders 1, 2, and 3 are

$$J(\lambda, 1) = \lambda, \quad J(\lambda, 2) = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \quad J(\lambda, 3) = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}.$$

Example B.143. The nilpotent operator

$$N(x, y, z) = (0, x, y)$$

can be exploited to construct a basis of $\mathbb{F}^3$: $(N^2\mathbf{v}, N\mathbf{v}, \mathbf{v})$ where $\mathbf{v} = (1, 0, 0)$. With respect to this basis, the matrix of N is the Jordan block $J(0, 3)$.

Example B.144. The nilpotent operator

$$N(z_1, z_2, z_3, z_4, z_5, z_6) = (0, z_1, z_2, 0, z_4, 0)$$

can be exploited to construct a basis of $\mathbb{F}^6$:

$$(N^2\mathbf{v}_1, N\mathbf{v}_1, \mathbf{v}_1, N\mathbf{v}_2, \mathbf{v}_2, \mathbf{v}_3)$$

where

$$\begin{aligned} \mathbf{v}_1 &= (1, 0, 0, 0, 0, 0), \\ \mathbf{v}_2 &= (0, 0, 0, 1, 0, 0), \\ \mathbf{v}_3 &= (0, 0, 0, 0, 0, 1). \end{aligned}$$

With respect to this basis, the matrix of N is the block diagonal matrix

$$\begin{pmatrix} J(0,3) & & \\ & J(0,2) & \\ & & J(0,1) \end{pmatrix}.$$

Lemma B.145. For a nilpotent operator $N \in \mathcal{L}(V)$, there exist $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in V$ and $m_1, m_2, \ldots, m_n \in \mathbb{N}$ such that

(a) $N^{m_1}\mathbf{v}_1, \cdots, N\mathbf{v}_1, \mathbf{v}_1, \cdots, N^{m_n}\mathbf{v}_n, \cdots, N\mathbf{v}_n, \mathbf{v}_n$ form a basis of V;

(b) $N^{m_1+1}\mathbf{v}_1 = \cdots = N^{m_n+1}\mathbf{v}_n = \mathbf{0}$.

Proof. If range N is empty, Theorem B.56 implies that null $N = V$. Then $N = \mathbf{0}$ and both (a) and (b) hold trivially. Hereafter we prove this lemma by an induction on $\dim V$. The induction basis for $\dim V = 1$ clearly holds. Hereafter we assume that $\dim V > 1$ and (a) and (b) hold on all vector spaces of smaller dimensions.

Because N is nilpotent, N is not injective. By Theorem B.57, N is not surjective either. The range of N, a subspace of V (c.f. Theorem B.55), has a smaller dimension than V. Thus we can apply our induction hypothesis to range N to obtain that, for $N|_{\text{range}\,N} \in \mathcal{L}(\text{range}\,N)$, there exist vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \text{range}\,N$ and $m_1, m_2, \ldots, m_n \in \mathbb{N}$ such that

$$(*): \quad N^{m_1}\mathbf{v}_1, \ldots, N\mathbf{v}_1, \mathbf{v}_1, \cdots, N^{m_n}\mathbf{v}_n, \ldots, N\mathbf{v}_n, \mathbf{v}_n$$

is a basis of range N and (b) holds for this basis. Then

$$\forall j = 1, \ldots, n, \quad \exists \mathbf{u}_j \in V \text{ s.t. } \mathbf{v}_j = N\mathbf{u}_j.$$

Next, we claim that the following is an independent list of vectors in V:

$$(\square): \quad N^{m_1+1}\mathbf{u}_1, \ldots, N\mathbf{u}_1, \mathbf{u}_1, \cdots, N^{m_n+1}\mathbf{u}_n, \ldots, N\mathbf{u}_n, \mathbf{u}_n.$$

Indeed, suppose a linear combination of $(\square)$ equals $\mathbf{0}$. Then apply N to it and we get a linear combination of $(*)$ equal to $\mathbf{0}$ and thus all the coefficients in the original linear combination must be 0 except for those of the vectors

$$N^{m_1+1}\mathbf{u}_1 = N^{m_1}\mathbf{v}_1, \ \ldots, \ N^{m_n+1}\mathbf{u}_n = N^{m_n}\mathbf{v}_n.$$

Again, the linear independence of the list $(*)$ dictates that all coefficients of the above vectors must be 0.

By Lemma B.37, we can extend $(\square)$ to a basis of V,

$$(\triangle): \quad N^{m_1+1}\mathbf{u}_1, \ldots, \mathbf{u}_1, \cdots, N^{m_n+1}\mathbf{u}_n, \ldots, \mathbf{u}_n, \mathbf{w}_1, \ldots, \mathbf{w}_p.$$

Since $N\mathbf{w}_j \in \text{range}\,N$, each $N\mathbf{w}_j$ is in the span of $(*)$, thus there exists $\mathbf{x}_j$ in the span of $(\square)$ such that $N\mathbf{w}_j = N\mathbf{x}_j$. Define $\mathbf{u}_{n+j} = \mathbf{w}_j - \mathbf{x}_j$ and we have $N\mathbf{u}_{n+j} = \mathbf{0}$. Therefore the following list of vectors satisfies (a) and (b):

$$N^{m_1+1}\mathbf{u}_1, \ldots, \mathbf{u}_1, \cdots, N^{m_n+1}\mathbf{u}_n, \ldots, \mathbf{u}_n, \mathbf{u}_{n+1}, \ldots, \mathbf{u}_{n+p}.$$

This completes the induction and the proof. $\square$

Definition B.146. A *Jordan basis for a linear operator* $T \in \mathcal{L}(V)$ is a basis of V with respect to which the matrix of T is a block diagonal matrix of the form

$$J = \begin{pmatrix} J(\lambda_1, k_1) & & & \\ & J(\lambda_2, k_2) & & \\ & & \ddots & \\ & & & J(\lambda_p, k_p) \end{pmatrix}, \tag{B.52}$$

where the λ_j's might not be distinct.

Theorem B.147 (Jordan)**.** Any operator $T \in \mathcal{L}(V)$ on a complex vector space has a Jordan basis.

Proof. If T is a nilpotent operator N, we consider the basis given in Lemma B.145(a). For each j, N annihilates $N^{m_j}\mathbf{v}_j$ in the list $N^{m_j}\mathbf{v}_j, \cdots, N\mathbf{v}_j, \mathbf{v}_j$ and sends another vector to its previous. Hence N has a block diagonal matrix where each matrix on the diagonal is the Jordan block $J(0, k_j)$. Thus the statement holds for nilpotent operators.

Otherwise let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be all distinct eigenvalues of T. By Theorem B.132, we have the decomposition

$$V = G(\lambda_1, T) \oplus \cdots \oplus G(\lambda_m, T),$$

with each $N_j := (T - \lambda_j I)|_{G(\lambda_j, T)}$ being nilpotent. By the previous paragraph, each N_j has a Jordan basis. The combination of all these Jordan bases is a Jordan basis for T. □

B.5 Inner product spaces

B.5.1 Inner products

Definition B.148. Denote by $\mathbb{F}$ the underlying field of a vector space $\mathcal{V}$. The *inner product* $\langle \mathbf{u}, \mathbf{v} \rangle$ on $\mathcal{V}$ is a function $\mathcal{V} \times \mathcal{V} \to \mathbb{F}$ that satisfies

(IP-1) real positivity: $\forall \mathbf{v} \in \mathcal{V}$, $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$;

(IP-2) definiteness: $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ iff $\mathbf{v} = \mathbf{0}$;

(IP-3) additivity in the first slot: $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$, $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$;

(IP-4) homogeneity in the first slot: $\forall a \in \mathbb{F}$, $\forall \mathbf{v}, \mathbf{w} \in \mathcal{V}$, $\langle a\mathbf{v}, \mathbf{w} \rangle = a \langle \mathbf{v}, \mathbf{w} \rangle$;

(IP-5) conjugate symmetry: $\forall \mathbf{v}, \mathbf{w} \in \mathcal{V}$, $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$.

An *inner product space* is a vector space $\mathcal{V}$ equipped with an inner product on $\mathcal{V}$.

Corollary B.149. An inner product has additivity in the second slot, i.e. $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$.

Corollary B.150. An inner product has conjugate homogeneity in the second slot, i.e.

$$\forall a \in \mathbb{F},\ \forall \mathbf{v}, \mathbf{w} \in \mathcal{V}, \qquad \langle \mathbf{v}, a\mathbf{w} \rangle = \bar{a} \langle \mathbf{v}, \mathbf{w} \rangle. \tag{B.53}$$

Exercise B.151. Prove Corollaries B.149 and B.150 from Definition B.148.

Definition B.152. The *Euclidean inner product* on $\mathbb{F}^n$ is

$$\langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{n} v_i \overline{w_i}. \tag{B.54}$$

B.5.2 Norms induced from inner products

Definition B.153. Let $\mathbb{F}$ be the underlying field of an inner product space $\mathcal{V}$. The *norm induced by an inner product* on $\mathcal{V}$ is a function $\mathcal{V} \to \mathbb{F}$:

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}. \tag{B.55}$$

Definition B.154. For $p \in [1, \infty)$, the *Euclidean ℓ_p norm* of a vector $\mathbf{v} \in \mathbb{F}^n$ is

$$\|\mathbf{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{\frac{1}{p}} \tag{B.56}$$

and the *Euclidean ℓ_∞ norm* is

$$\|\mathbf{v}\|_\infty = \max_i |v_i|. \tag{B.57}$$

Theorem B.155 (Equivalence of norms)**.** Any two norms $\|\cdot\|_N$ and $\|\cdot\|_M$ on a finite dimensional vector space $\mathcal{V} = \mathbb{C}^n$ satisfy

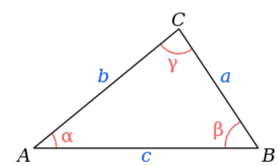
$$\exists c_1, c_2 \in \mathbb{R}^+, \text{ s.t. } \forall \mathbf{x} \in \mathcal{V},\ c_1\|\mathbf{x}\|_M \leq \|\mathbf{x}\|_N \leq c_2\|\mathbf{x}\|_M. \tag{B.58}$$

Definition B.156. The angle between two vectors $\mathbf{v}, \mathbf{w}$ in an inner product space with $\mathbb{F} = \mathbb{R}$ is the number $\theta \in [0, \pi]$,

$$\theta = \arccos \frac{\langle \mathbf{v}, \mathbf{w} \rangle}{\|\mathbf{v}\|\|\mathbf{w}\|}. \tag{B.59}$$

Theorem B.157 (The law of cosines)**.** Any triangle satisfies

$$c^2 = a^2 + b^2 - 2ab\cos\gamma. \tag{B.60}$$



Proof. The dot product of AB to $AB = CB - CA$ yields

$$c^2 = \langle AB, CB \rangle - \langle AB, CA \rangle.$$

The dot products of CB and CA to $AB = CB - CA$ yield

$$\begin{aligned} \langle CB, AB \rangle &= a^2 - \langle CB, CA \rangle; \\ -\langle CA, AB \rangle &= -\langle CA, CB \rangle + b^2. \end{aligned}$$

The proof is completed by adding up all three equations and applying (B.59). □

Theorem B.158 (The law of cosines: abstract version)**.** Any induced norm on a real vector space satisfies

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\langle \mathbf{u}, \mathbf{v} \rangle. \tag{B.61}$$

Proof. Definitions B.153 and B.148 and $\mathbb{F} = \mathbb{R}$ yield

$$\begin{aligned} \|\mathbf{u} - \mathbf{v}\|^2 &= \langle \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle - \langle \mathbf{u}, \mathbf{v} \rangle - \langle \mathbf{v}, \mathbf{u} \rangle \\ &= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\langle \mathbf{u}, \mathbf{v} \rangle. \end{aligned}$$ □

B.5.3 Norms and induced inner-products

Definition B.159. A function $\|\cdot\| : \mathcal{V} \to \mathbb{F}$ is a *norm* for a vector space $\mathcal{V}$ iff it satisfies

(NRM-1) real positivity: $\forall \mathbf{v} \in \mathcal{V}$, $\|\mathbf{v}\| \geq 0$;

(NRM-2) point separation: $\|\mathbf{v}\| = 0 \Rightarrow \mathbf{v} = \mathbf{0}$.

(NRM-3) absolute homogeneity: $\forall a \in \mathbb{F}, \forall \mathbf{v} \in \mathcal{V}$, $\|a\mathbf{v}\| = |a|\|\mathbf{v}\|$;

(NRM-4) triangle inequality: $\forall \mathbf{u}, \mathbf{v} \in \mathcal{V}$, $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$.

The function $\|\cdot\| : \mathcal{V} \to \mathbb{F}$ is called a *semi-norm* iff it satifies (NRM-1,3,4). A *normed vector space* (or simply a *normed space*) is a vector space $\mathcal{V}$ equipped with a norm on $\mathcal{V}$.

Exercise B.160. Explain how (NRM-1,2,3,4) relate to the geometric meaning of the norm of vectors in $\mathbb{R}^3$.

Lemma B.161. The norm induced by an inner product is a norm as in Definition B.159.

Proof. The induced norm as in (B.55) satisfies (NRM-1,2) trivially. For (NRM-3),

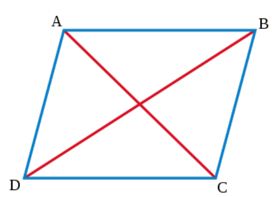
$$\|a\mathbf{v}\|^2 = \langle a\mathbf{v}, a\mathbf{v}\rangle = a\,\langle \mathbf{v}, a\mathbf{v}\rangle = a\bar{a}\,\langle \mathbf{v}, \mathbf{v}\rangle = |a|^2\|\mathbf{v}\|^2.$$

To prove (NRM-4), we have

$$\begin{aligned}\|\mathbf{u}+\mathbf{v}\|^2 &= \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle \\ &= \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle + \langle \mathbf{u}, \mathbf{v}\rangle + \overline{\langle \mathbf{u}, \mathbf{v}\rangle} \\ &\leq \langle \mathbf{u}, \mathbf{u}\rangle + \langle \mathbf{v}, \mathbf{v}\rangle + 2|\,\langle \mathbf{u}, \mathbf{v}\rangle\,| \\ &\leq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\|\mathbf{u}\|\|\mathbf{v}\| \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2,\end{aligned}$$

where the second step follows from (IP-5) and the fourth step from Cauchy-Schwarz inequality. □

Theorem B.162 (The parallelogram law)**.** The sum of squares of the lengths of the four sides of a parallelogram equals the sum of squares of the two diagonals.



More precisely, we have in the above plot

$$(AB)^2+(BC)^2+(CD)^2+(DA)^2 = (AC)^2+(BD)^2. \quad \text{(B.62)}$$

Proof. Apply the law of cosines to the two diagonals, add the two equations, and we obtain (B.62). □

Theorem B.163 (The parallelogram law: abstract version)**.** Any induced norm (B.55) satisfies

$$2\|\mathbf{u}\|^2 + 2\|\mathbf{v}\|^2 = \|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2. \quad \text{(B.63)}$$

Proof. Replace $\mathbf{v}$ in (B.61) with $-\mathbf{v}$ and we have

$$\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\,\langle \mathbf{u}, \mathbf{v}\rangle\,.$$

(B.63) follows from adding the above equation to (B.61). □

Exercise B.164. In the case of Euclidean ℓ_p norms, show that the parallelogram law (B.63) holds if and only if $p = 2$.

Theorem B.165. The induced norm (B.55) holds for some inner product $\langle\cdot,\cdot\rangle$ if and only if the parallelogram law (B.63) holds for every pair of $\mathbf{u}, \mathbf{v} \in \mathcal{V}$.

Exercise B.166. Prove Theorem B.165.

Example B.167. By Theorem B.165 and Exercise B.164, the ℓ^1 and ℓ^∞ spaces do not have a corresponding inner product for the Euclidean ℓ_1 and ℓ_∞ norms.

B.5.4 Orthonormal bases

Definition B.168. Two vectors $\mathbf{u}, \mathbf{v}$ are called *orthogonal* if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$, i.e., their inner product is the additive identity of the underlying field.

Example B.169. An inner product on the vector space of continuous real-valued functions on the interval $[-1, 1]$ is

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

f and g are said to be orthogonal if the integral is zero.

Theorem B.170 (Pythagorean)**.** If $\mathbf{u}, \mathbf{v}$ are orthogonal, then $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

Proof. This follows from (B.61) and Definition B.168. □

Theorem B.171 (Cauchy-Schwarz inequality)**.**

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|, \quad \text{(B.64)}$$

where the equality holds iff one of $\mathbf{u}, \mathbf{v}$ is a scalar multiple of the other.

Proof. For any complex number λ, (IP-1) implies

$$\begin{aligned}&\langle \mathbf{u}+\lambda\mathbf{v}, \mathbf{u}+\lambda\mathbf{v}\rangle \geq 0 \\ \Rightarrow &\langle \mathbf{u}, \mathbf{u}\rangle + \lambda\,\langle \mathbf{v}, \mathbf{u}\rangle + \bar{\lambda}\,\langle \mathbf{u}, \mathbf{v}\rangle + \lambda\bar{\lambda}\,\langle \mathbf{v}, \mathbf{v}\rangle \geq 0.\end{aligned}$$

If $\mathbf{v} = \mathbf{0}$, (B.64) clearly holds. Otherwise (B.64) follows from substituting $\lambda = -\frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\langle \mathbf{v}, \mathbf{v}\rangle}$ into the above equation. □

Exercise B.172. To explain the choice of λ in the proof of Theorem B.171, what is the geometric meaning of (B.64) in the plane? When will the equality hold?

Example B.173. If $x_i, y_i \in \mathbb{R}$, then for any $n \in \mathbb{N}^+$

$$\left|\sum_{i=1}^n x_i y_i\right|^2 \leq \sum_{j=1}^n x_j^2 \sum_{k=1}^n y_k^2.$$

Example B.174. If $f, g : [a, b] \to \mathbb{R}$ are continuous, then

$$\left|\int_a^b f(x)g(x)\mathrm{d}x\right|^2 \leq \left(\int_a^b f^2(x)\mathrm{d}x\right)\left(\int_a^b g^2(x)\mathrm{d}x\right)$$

Definition B.175. A list of vectors $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m)$ is called *orthonormal* if the vectors in it are pairwise orthogonal and each vector has norm 1, i.e.

$$\begin{cases} \forall i = 1, 2, \dots, m, & \|\mathbf{e}_i\| = 1; \\ \forall i \neq j, & \langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0. \end{cases} \tag{B.65}$$

Definition B.176. An *orthonormal basis* of an inner-product space $\mathcal{V}$ is an orthonormal list of vectors in $\mathcal{V}$ that is also a basis of $\mathcal{V}$.

Theorem B.177. If $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$ is an orthonormal basis of $\mathcal{V}$, then

$$\forall \mathbf{v} \in \mathcal{V}, \quad \mathbf{v} = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{e}_i \rangle \mathbf{e}_i, \tag{B.66a}$$

$$\|\mathbf{v}\|^2 = \sum_{i=1}^{n} |\langle \mathbf{v}, \mathbf{e}_i \rangle|^2. \tag{B.66b}$$

Lemma B.178. Every finite-dimensional inner-product space has an orthonormal basis.

Theorem B.179 (Schur)**.** Every linear operator $T \in \mathcal{L}(\mathcal{V})$ on a finite-dimensional complex vector space $\mathcal{V}$ has an upper-triangular matrix with respect to some orthonormal basis of $\mathcal{V}$.

Proof. This follows from Theorem B.104, Lemma B.178 and the Gram-Schmidt process. □

Definition B.180. A *linear functional* on $\mathcal{V}$ is a linear map from $\mathcal{V}$ to $\mathbb{F}$, or, it is an element of $\mathcal{L}(\mathcal{V}, \mathbb{F})$.

Theorem B.181 (Riesz representation theorem)**.** If $\mathcal{V}$ is a finite-dimensional vector space, then

$$\forall \varphi \in \mathcal{V}', \ \exists! \mathbf{u} \in \mathcal{V} \text{ s.t. } \forall \mathbf{v} \in \mathcal{V}, \quad \varphi(\mathbf{v}) = \langle \mathbf{v}, \mathbf{u} \rangle. \tag{B.67}$$

Proof. Let $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$ be an orthonormal basis of $\mathcal{V}$.

$$\begin{aligned} \varphi(\mathbf{v}) &= \varphi\left(\sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{e}_i \rangle \mathbf{e}_i\right) = \sum_{i=1}^{n} \langle \mathbf{v}, \mathbf{e}_i \rangle \varphi(\mathbf{e}_i) \\ &= \sum_{i=1}^{n} \left\langle \mathbf{v}, \overline{\varphi(\mathbf{e}_i)} \mathbf{e}_i \right\rangle = \left\langle \mathbf{v}, \sum_{i=1}^{n} \overline{\varphi(\mathbf{e}_i)} \mathbf{e}_i \right\rangle, \end{aligned}$$

where the last two steps follow from Corollaries B.149 and B.150.

As for the uniqueness, suppose that $\exists \mathbf{u}_1, \mathbf{u}_2 \in \mathcal{V}$ s.t. $\varphi(\mathbf{v}) = \langle \mathbf{v}, \mathbf{u}_1 \rangle = \langle \mathbf{v}, \mathbf{u}_2 \rangle$. Then for each $\mathbf{v} \in \mathcal{V}$,

$$0 = \langle \mathbf{v}, \mathbf{u}_1 \rangle - \langle \mathbf{v}, \mathbf{u}_2 \rangle = \langle \mathbf{v}, \mathbf{u}_1 - \mathbf{u}_2 \rangle.$$

Taking $\mathbf{v} = \mathbf{u}_1 - \mathbf{u}_2$ shows that $\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{0}$. □

B.6 Operators on inner-product spaces

B.6.1 Adjoint and self-adjoint operators

Definition B.182. The *adjoint of a linear map* $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$ between inner-product spaces is a function $T^* : \mathcal{W} \to \mathcal{V}$ that satisfies

$$\forall \mathbf{v} \in \mathcal{V}, \ \forall \mathbf{w} \in \mathcal{W}, \qquad \langle T\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, T^*\mathbf{w} \rangle. \tag{B.68}$$

Example B.183. Define a linear operator $T : \mathbb{R}^3 \to \mathbb{R}^2$,

$$T(x_1, x_2, x_3) = (x_2 + 3x_3, 2x_1).$$

Then $T^*(y_1, y_2) = (2y_2, y_1, 3y_1)$ because

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

Lemma B.184. If $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$, then $T^* \in \mathcal{L}(\mathcal{W}, \mathcal{V})$.

Proof. Use Definition B.42. □

Theorem B.185. The adjoint of a linear map has the following properties.

(ADJ-1) additivity:
$\forall S, T \in \mathcal{L}(\mathcal{V}, \mathcal{W}), \qquad (S + T)^* = S^* + T^*;$

(ADJ-2) conjugate homogeneity:
$\forall T \in \mathcal{L}(\mathcal{V}, \mathcal{W}), \forall a \in \mathbb{F}, \qquad (aT)^* = \bar{a}T^*;$

(ADJ-3) adjoint of adjoint:
$\forall T \in \mathcal{L}(\mathcal{V}, \mathcal{W}), \ (T^*)^* = T;$

(ADJ-4) identity: $I^* = I$;

(ADJ-5) products: let $\mathcal{U}$ be an inner-product space,
$\forall T \in \mathcal{L}(\mathcal{V}, \mathcal{W}), \forall S \in \mathcal{L}(\mathcal{W}, \mathcal{U}), \ (ST)^* = T^* S^*.$

Proof. Use Definitions B.182 and B.148. □

Lemma B.186. $T \in \mathcal{L}(V, W)$ and T^* satisfy

(a) $\operatorname{null} T^* = (\operatorname{range} T)^\perp$;

(b) $\operatorname{range} T^* = (\operatorname{null} T)^\perp$;

(c) $\operatorname{null} T = (\operatorname{range} T^*)^\perp$;

(d) $\operatorname{range} T = (\operatorname{null} T^*)^\perp$.

Definition B.187. The *conjugate transpose*, or *Hermitian transpose*, or *Hermitian conjugate*, or *adjoint matrix*, of a matrix $A \in \mathbb{C}^{m \times n}$ is the matrix $A^* \in \mathbb{C}^{n \times m}$ defined by

$$(A^*)_{ij} = \overline{a_{ji}}, \tag{B.69}$$

where $\overline{a_{ji}}$ denotes the complex conjugate of the entry a_{ji}.

Exercise B.188. Show that the conjugate transpose is an adjoint operator in $\mathcal{L}(\mathcal{V}, \mathcal{W})$ with $\mathcal{V} = \mathbb{C}^n$ and $\mathcal{W} = \mathbb{C}^m$.

Definition B.189. A matrix $U \in \mathbb{C}^{n \times n}$ is *unitary* iff $U^* U = I$. A matrix $U \in \mathbb{R}^{n \times n}$ is *orthogonal* iff $U^T U = I$.

Theorem B.190. A matrix $U \in \mathbb{C}^{n \times n}$ is unitary if and only if its columns form an orthonormal basis for $\mathbb{C}^n$.

Proof. This follows from considering the (i, j)th element of $U^* U$ and applying $U^* U = I$ in Definition B.189. □

Corollary B.191. A unitary matrix U preserves norms and inner products. More precisely, we have

$$\forall \mathbf{v}, \mathbf{w} \in \mathbb{C}^n, \qquad \langle U\mathbf{v}, U\mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle.$$

Proof. This follows from Definitions B.182 and B.189. □

Theorem B.192. Every unitary matrix $U \in \mathbb{C}^{2\times 2}$ with $\det U = 1$ is of the form

$$U = \begin{bmatrix} a & b \\ -\bar{b} & \bar{a} \end{bmatrix}, \tag{B.70}$$

where $|a|^2 + |b|^2 = 1$.

Proof. Let

$$U = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then Theorem B.190 and the condition $\det U = 1$ yield

$$\begin{aligned} a\bar{b} + c\bar{d} &= 0, \\ ad - cb &= 1. \end{aligned}$$

In other words, the linear system

$$\begin{bmatrix} \bar{b} & \bar{d} \\ d & -b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

has solution $x = a, y = c$. Furthermore, Theorem B.190 and the form of U yield $|b|^2 + |d|^2 = 1$. Hence the solution $x = a, y = c$ is unique and we have $a = \bar{d}$ and $c = -\bar{b}$, which completes the proof. □

Theorem B.193. Let $T \in \mathcal{L}(\mathcal{V}, \mathcal{W})$. Suppose $e_1, \ldots, e_n$ is an orthonormal basis of $\mathcal{V}$ and $f_1, \ldots, f_m$ is an orthonormal basis of $\mathcal{W}$. Then

$$M(T^*, (f_1, \ldots, f_m), (e_1, \ldots, e_n))$$

is the conjugate transpose of

$$M(T, (e_1, \ldots, e_n), (f_1, \ldots, f_m)).$$

Proof. By Corollary B.59, we have

$$T[e_1, \ldots, e_n] = [f_1, \ldots, f_m]M_T.$$

The orthonormality of the two bases and Definition B.152 further imply

$$M_T = \begin{bmatrix} \langle Te_1, f_1\rangle & \langle Te_2, f_1\rangle & \cdots & \langle Te_n, f_1\rangle \\ \langle Te_1, f_2\rangle & \langle Te_2, f_2\rangle & \cdots & \langle Te_n, f_2\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle Te_1, f_m\rangle & \langle Te_2, f_m\rangle & \cdots & \langle Te_n, f_m\rangle \end{bmatrix}.$$

The proof is completed by repeating the above derivation for T^* and then applying Definitions B.148 and B.182. □

Lemma B.194. Suppose $\mathcal{V}$ is a complex inner product space and $T \in \mathcal{L}(\mathcal{V})$. If

$$\forall \mathbf{v} \in \mathcal{V}, \qquad \langle T\mathbf{v}, \mathbf{v}\rangle = 0, \tag{B.71}$$

then $T = \mathbf{0}$.

Proof. By Definition B.148 and (B.71), we have, $\forall \mathbf{u}, \mathbf{w} \in \mathcal{V}$,

$$\begin{aligned} \langle T\mathbf{u}, \mathbf{w}\rangle =& \frac{\langle T(\mathbf{u}+\mathbf{w}), \mathbf{u}+\mathbf{w}\rangle - \langle T(\mathbf{u}-\mathbf{w}), \mathbf{u}-\mathbf{w}\rangle}{4} \\ &+ \mathbf{i}\frac{\langle T(\mathbf{u}+\mathbf{i}\mathbf{w}), \mathbf{u}+\mathbf{i}\mathbf{w}\rangle - \langle T(\mathbf{u}-\mathbf{i}\mathbf{w}), \mathbf{u}-\mathbf{i}\mathbf{w}\rangle}{4} \\ =& 0. \end{aligned}$$

Setting $\mathbf{w} = T\mathbf{u}$ completes the proof. □

Definition B.195. An operator $T \in \mathcal{L}(\mathcal{V})$ is *self-adjoint* iff $T = T^*$, i.e.

$$\forall \mathbf{v}, \mathbf{w} \in \mathcal{V}, \qquad \langle T\mathbf{v}, \mathbf{w}\rangle = \langle \mathbf{v}, T\mathbf{w}\rangle. \tag{B.72}$$

Lemma B.196. Every eigenvalue of a self-adjoint operator T is real.

Proof. Let $(\lambda, \mathbf{u})$ be an eigenpair of T. We have

$$\lambda\|\mathbf{u}\|^2 = \langle \lambda\mathbf{u}, \mathbf{u}\rangle = \langle T\mathbf{u}, \mathbf{u}\rangle = \langle \mathbf{u}, T\mathbf{u}\rangle = \langle \mathbf{u}, \lambda\mathbf{u}\rangle = \bar{\lambda}\|\mathbf{u}\|^2,$$

where the third step follows from Definition B.195. Then $\mathbf{u} \neq \mathbf{0}$ implies $\lambda = \bar{\lambda}$. □

Theorem B.197. Suppose $\mathcal{V}$ is a complex inner product space and $T \in \mathcal{L}(\mathcal{V})$. Then T is self-adjoint if and only if

$$\forall \mathbf{v} \in \mathcal{V}, \quad \langle T\mathbf{v}, \mathbf{v}\rangle \in \mathbb{R}. \tag{B.73}$$

Proof. By Definitions B.148, B.182, and B.195, we have

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

Then Lemma B.194 completes the proof. □

Lemma B.198. Suppose $\mathcal{V}$ is a real inner product space and $T \in \mathcal{L}(\mathcal{V})$. If T is self-adjoint and satisfies

$$\forall \mathbf{v} \in \mathcal{V}, \qquad \langle T\mathbf{v}, \mathbf{v}\rangle = 0, \tag{B.74}$$

then $T = \mathbf{0}$.

Proof. By the self-adjointness and the underlying field being real, we have

$$\langle T\mathbf{w}, \mathbf{u}\rangle = \langle \mathbf{w}, T\mathbf{u}\rangle = \langle T\mathbf{u}, \mathbf{w}\rangle,$$

which, together with Definition B.148, implies

$$\langle T\mathbf{u}, \mathbf{w}\rangle = \frac{\langle T(\mathbf{u}+\mathbf{w}), \mathbf{u}+\mathbf{w}\rangle - \langle T(\mathbf{u}-\mathbf{w}), \mathbf{u}-\mathbf{w}\rangle}{4}.$$

Setting $\mathbf{w} = T\mathbf{u}$ completes the proof. □

B.6.2 Normal operators

Definition B.199. An operator $T \in \mathcal{L}(\mathcal{V})$ is *normal* iff $TT^* = T^*T$.

Corollary B.200. Every self-adjoint operator is normal.

Lemma B.201. $T \in \mathcal{L}(\mathcal{V})$ is normal if and only if

$$\forall \mathbf{v} \in \mathcal{V}, \qquad \|T\mathbf{v}\| = \|T^*\mathbf{v}\|. \tag{B.75}$$

Proof. By Lemma B.198 and Definition B.182, we have

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

The positivity of a norm completes the proof. □

Lemma B.202. $T \in \mathcal{L}(\mathcal{V})$ is normal if and only if each eigenvector of T is also an eigenvector of T^*.

Proof. If T is normal, so is $T - \lambda I$. By Lemma B.201, an eigenpair $(\lambda, \mathbf{u})$ of T satisfies

$$0 = \|(T - \lambda I)\mathbf{u}\| = \|(T - \lambda I)^*\mathbf{u}\| = \|(T^* - \bar{\lambda} I)\mathbf{u}\|,$$

and thus $\mathbf{u}$ is also an eigenvector of T^*.

Conversely, suppose each eigenvector $\mathbf{u}$ of T is also an eigenvector of T^*. Then the above equation implies that the corresponding eigenvalue of T^* is the conjugate of that of T. It suffices to prove that these eigenvectors form a basis of V, because then we have

$$TU = U\Lambda, \quad T^*U = U\Lambda^*,$$

where U is the matrix of these eigenvectors. Thus

$$TT^*U = TU\Lambda^* = U\Lambda\Lambda^* = U\Lambda^*\Lambda = T^*U\Lambda = T^*TU$$

and we have $TT^* = T^*T$ because U is nonsingular. By Theorem B.111, it suffices to show that T is diagonalizable. By Theorem B.147, we only need to show that for any eigenpair $(\lambda, \mathbf{u})$ of T,

$$(A - \lambda I)^2\mathbf{u} = \mathbf{0} \ \Rightarrow\ (A - \lambda I)\mathbf{u} = \mathbf{0},$$

because this condition will annihilate all Jordan blocks of size greater than 1. Define $\mathbf{v} = (A - \lambda I)\mathbf{u}$ and we have

$$\begin{aligned} \langle A\mathbf{u}, \mathbf{v}\rangle &= \langle \mathbf{u}, A^*\mathbf{v}\rangle = \left\langle \mathbf{u}, \bar{\lambda}\mathbf{v}\right\rangle = \langle \lambda\mathbf{u}, \mathbf{v}\rangle, \\ \langle A\mathbf{u}, \mathbf{v}\rangle &= \langle \mathbf{v} + \lambda\mathbf{u}, \mathbf{v}\rangle = \|\mathbf{v}\|^2 + \langle \lambda\mathbf{u}, \mathbf{v}\rangle, \end{aligned}$$

which imply $\mathbf{v} = 0$, i.e., $(A - \lambda I)\mathbf{u} = \mathbf{0}$. □

Theorem B.203. For a linear operator $T \in \mathcal{L}(\mathcal{V})$ on a two-dimensional real inner product space $\mathcal{V}$, the following are equivalent:

(a) T is normal but not self-adjoint.

(b) The matrix of T with respect to every orthonormal basis of $\mathcal{V}$ has the form

$$M(T) = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \tag{B.76}$$

where $b \neq 0$.

Proof. (b) ⇒ (a) trivially holds, so we only prove (a) ⇒ (b). Let (e_1, e_2) be an orthonormal basis of $\mathcal{V}$ and set

$$M(T, (e_1, e_2)) = \begin{bmatrix} a & c \\ b & d \end{bmatrix}.$$

By Definition B.199, we have

$$\begin{aligned} \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} &= \begin{bmatrix} a^2 + c^2 & ab + cd \\ ab + cd & b^2 + d^2 \end{bmatrix} \\ = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} &= \begin{bmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{bmatrix}. \end{aligned}$$

$b^2 = c^2$ and the condition of T being not self-adjoint further yields $c = -b \neq 0$, which, together with $ab + cd = ac + bd$, yields $a = d$. □

B.6.3 The spectral theorems

Theorem B.204 (Complex spectral)**.** For a linear operator $T \in \mathcal{L}(\mathcal{V})$ with $\mathbb{F} = \mathbb{C}$, the following are equivalent:

(a) T is normal;

(b) $\mathcal{V}$ has an orthonormal basis consisting of eigenvectors of T;

(c) T has a diagonal matrix with respect to some orthonormal basis of $\mathcal{V}$.

Corollary B.205. A normal operator T whose eigenvalues are real is self-adjoint.

Proof. By Theorem B.204, we write $T = V\Lambda V^{-1}$ and thus $T^* = V\Lambda^* V^{-1}$. Then all entries in Λ being real implies $T = T^*$. □

Theorem B.206 (Real spectral)**.** For a linear operator $T \in \mathcal{L}(\mathcal{V})$ with $\mathbb{F} = \mathbb{R}$, the following are equivalent:

(a) T is self-adjoint;

(b) $\mathcal{V}$ has an orthonormal basis consisting of eigenvectors of T;

(c) T has a diagonal matrix with respect to some orthonormal basis of $\mathcal{V}$.

B.6.4 Isometries

Definition B.207. An operator $S \in \mathcal{L}(\mathcal{V})$ is called a (linear) *isometry* iff

$$\forall \mathbf{v} \in \mathcal{V}, \qquad \|S\mathbf{v}\| = \|\mathbf{v}\|. \tag{B.77}$$

Theorem B.208. An operator $S \in \mathcal{L}(\mathcal{V})$ on a real inner product space is an isometry if and only if there exists an orthonormal basis of $\mathcal{V}$ with respect to which $\mathcal{S}$ has a block diagonal matrix such that each block on the diagonal is a 1-by-1 matrix containing 1 or -1, or, is a 2-by-2 matrix of the form

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{B.78}$$

where $\theta \in (0, \pi)$.

Corollary B.209. For an operator $\mathcal{S} \in \mathcal{L}(\mathcal{V})$ on a two-dimensional real inner product space, the following are equivalent:

(a) $\mathcal{S}$ is an isometry;

(b) $\mathcal{S}$ is either an identity or a reflection or a rotation.

B.6.5 Singular value decomposition

Definition B.210. An operator $T \in \mathcal{L}(V)$ is *positive semi-definite* iff

$$\forall \mathbf{v} \in V, \quad \langle T\mathbf{v}, \mathbf{v}\rangle \geq 0 \tag{B.79}$$

and is *positive definite* iff

$$\forall \mathbf{v} \in V \setminus \{\mathbf{0}\}, \quad \langle T\mathbf{v}, \mathbf{v}\rangle > 0. \tag{B.80}$$

An operator is *positive* if it is self-adjoint and positive semi-definite.

Corollary B.211. For any linear operator $f \in \mathcal{L}(\mathcal{V})$, both $f^* \circ f$ and $f \circ f^*$ are positive.

Proof. By Definition B.195, $f^* \circ f$ is self-adjoint since

$$\langle (f^* \circ f)\mathbf{u}, \mathbf{v}\rangle = \langle f\mathbf{u}, f\mathbf{v}\rangle = \langle \mathbf{u}, (f^* \circ f)\mathbf{v}\rangle .$$

Suppose $(\lambda, \mathbf{u})$ is an eigenpair of $(f^* \circ f)$. Then we have

$$\begin{aligned} &\lambda \langle \mathbf{u}, \mathbf{u}\rangle = \langle (f^* \circ f)\mathbf{u}, \mathbf{u}\rangle = \langle f\mathbf{u}, f\mathbf{u}\rangle \\ \Rightarrow &\lambda = \frac{\langle f\mathbf{u}, f\mathbf{u}\rangle}{\langle \mathbf{u}, \mathbf{u}\rangle} \geq 0. \end{aligned}$$

Similar arguments apply to $f \circ f^*$. □

Definition B.212. The *singular values* of a linear map $f : \mathbb{C}^n \to \mathbb{C}^m$ are the non-negative square roots of the eigenvalues of $f^* \circ f$, usually sorted in non-increasing order as

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_n = 0,$$

where r is the rank of f.

Theorem B.213. For any matrix $A \in \mathbb{C}^{m\times n}$ with rank r, there exist orthonormal bases $\mathbb{C}^n = \operatorname{span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ and $\mathbb{C}^m = \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m\}$ such that

$$\forall j = 1, 2, \ldots, r, \quad \begin{cases} A\mathbf{u}_j = \sigma_j \mathbf{v}_j; \\ A^*\mathbf{v}_j = \sigma_j \mathbf{u}_j, \end{cases} \tag{B.81a}$$

$$\begin{cases} \forall j = r+1, r+2, \ldots, n, \quad A\mathbf{u}_j = \mathbf{0}; \\ \forall j = r+1, r+2, \ldots, m, \quad A^*\mathbf{v}_j = \mathbf{0}, \end{cases} \tag{B.81b}$$

where σ_j's are the singular values of A in Definition B.212.

Proof. The matrix $A^*A \in \mathbb{C}^{n\times n}$ is self-adjoint and thus normal. By Theorem B.204, $\mathbb{C}^n$ has an orthonormal basis that are also eigenvectors of A^*A; choose them to be $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$. By Definition B.212 we have

$$A^*A\mathbf{u}_j = \sigma_j^2 \mathbf{u}_j.$$

Then we choose

$$\forall j = 1, 2, \ldots, r, \quad \mathbf{v}_j := \frac{1}{\sigma_j} A\mathbf{u}_j,$$

which implies

$$\langle \mathbf{v}_i, \mathbf{v}_j\rangle = \frac{1}{\sigma_i\sigma_j} \langle A\mathbf{u}_i, A\mathbf{u}_j\rangle = \frac{1}{\sigma_i\sigma_j} \langle \mathbf{u}_i, A^*A\mathbf{u}_j\rangle = \frac{\sigma_j}{\sigma_i} \langle \mathbf{u}_i, \mathbf{u}_j\rangle .$$

Then the orthonormality of $\mathbf{u}_j$'s implies that the first r $\mathbf{v}_j$'s are orthonormal and if $r < m$ we can extend them by the Gram-Schmidt process to arrive at an orthonormal basis of $\mathbb{C}^m$. Therefore (B.81a) holds.

The first line of (B.81b) follows from $\text{null}A = \text{null}(A^*A)$, which is implied by

$$\|A\mathbf{u}\|^2 = \langle A\mathbf{u}, A\mathbf{u}\rangle = \langle \mathbf{u}, A^*A\mathbf{u}\rangle .$$

The rank of A^* is r. If $r = m$, the second line of (B.81b) holds vacuously. Otherwise $r < m$. The fundamental theorem of linear algebra (Theorem B.89) implies that A^* has rank $m - r$, which completes the proof. □

Definition B.214. The *singular value decomposition* (SVD) of a rectangular matrix $A \in \mathbb{C}^{m\times n}$ is the factorization $A = V\Sigma U^*$ where Σ is a diagonal matrix with its diagonal entries as the singular values in Definition B.212 and V and U are unitary matrices whose columns are respectively the vectors $\mathbf{v}_j$'s and $\mathbf{u}_j$'s specified in Theorem B.213, which are also called the *left singular vectors* and the *right singular vectors* of A, respectively. We also refer to the sequence of triples $(\sigma_j, \mathbf{u}_j, \mathbf{v}_j)$ as the *singular system* of A.

Definition B.215. Two matrices $A, B \in \mathbb{R}^{n\times n}$ are called *similar* iff there exists an invertible matrix P such that $B = P^{-1}AP$. The map $A \mapsto P^{-1}AP$ is called a *similarity transformation* or *conjugation of the matrix A*.

B.7 Trace and determinant

Definition B.216. The *trace of a matrix A*, denoted by Trace A, is the sum of the diagonal entries of A.

Lemma B.217. The trace of a matrix is the sum of its eigenvalues, each of which is repeated according to its multiplicity.

Definition B.218. A *permutation of a set A* is a bijective function $\sigma : A \to A$.

Definition B.219. Let σ be a permutation of $A = \{1, 2, \ldots, n\}$ and let s denote the number of pairs of integers (j, k) with $1 \leq j < k \leq n$ such that j appears after k in the list $(m_1, \ldots, m_n)$ given by $m_i = \sigma(i)$. The *sign of the permutation* σ is 1 if s is even and -1 if s is odd.

Definition B.220. The *signed volume of a parallelotope* spanned by n vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ is a function $\delta : \mathbb{R}^{n\times n} \to \mathbb{R}$ that satisfies

(SVP-1) $\delta(I) = 1$;

(SVP-2) $\delta(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = 0$ if $\mathbf{v}_i = \mathbf{v}_j$ for some $i \neq j$;

(SVP-3) δ is linear, i.e., $\forall j = 1, \ldots, n$, $\forall c \in \mathbb{R}$,

$$\begin{aligned} &\delta(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{v} + c\mathbf{w}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n) \\ = &\,\delta(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{v}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n) \\ &+ c\delta(\mathbf{v}_1, \ldots, \mathbf{v}_{j-1}, \mathbf{w}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n). \end{aligned} \tag{B.82}$$

Exercise B.221. Give a geometric proof that the signed volume of the parallelogram determined by the two vectors $\mathbf{v}_1 = (a,b)^T$ and $\mathbf{v}_2 = (c,d)^T$ is

$$\delta(\mathbf{v}_1, \mathbf{v}_2) = ad - bc = \left\langle \mathbf{v}_1^{\perp}, \mathbf{v}_2 \right\rangle. \tag{B.83}$$

Lemma B.222. Adding a multiple of one vector to another does not change the signed volume.

Proof. This follows directly from (SVP-2,3). □

Lemma B.223. If the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly dependent, then $\delta(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n) = 0$.

Proof. WLOG, we assume $\mathbf{v}_1 = \sum_{i=2}^n c_i \mathbf{v}_i$. Then the result follows from (SVP-2,3). □

Lemma B.224. The signed volume δ is alternating, i.e.,

$$\delta(\mathbf{v}_1, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n) = -\delta(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_i, \ldots, \mathbf{v}_n). \tag{B.84}$$

Exercise B.225. Prove Lemma B.224 using (SVP-2,3).

Lemma B.226. Let M_σ denote the matrix of a permutation $\sigma : E \to E$ where E is the set of standard basis vectors in (B.5). Then we have $\delta(M_\sigma) = \mathrm{sgn}(\sigma)$.

Proof. There is a one-to-one correspondence between the vectors in the matrix

$$M_\sigma = [e_{\sigma(1)}, e_{\sigma(2)}, \ldots, e_{\sigma(n)}]$$

and the scalars in the one-line notation

$$(\sigma(1)\ \sigma(2)\ \ldots\ \sigma(n)).$$

A sequence of transpositions taking σ to the identity map also takes M_σ to the identity matrix. By Lemma B.224, each transposition yields a multiplication factor -1. Definition B.219 and (SVP-1) give $\delta(M_\sigma) = \mathrm{sgn}(\sigma)\delta(I) = \mathrm{sgn}(\sigma)$. □

Definition B.227 (Leibniz formula of determinants)**.** The *determinant of a square matrix* $A \in \mathbb{R}^{n\times n}$ is

$$\det A = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^n a_{\sigma(i),i}, \tag{B.85}$$

where the sum is over the symmetric group S_n of all permutations and $a_{\sigma(i),i}$ is the element of A at the $\sigma(i)$th row and the ith column.

Lemma B.228. The determinant of a matrix is the product of its eigenvalues, each of which is repeated according to its multiplicity.

Theorem B.229. The signed volume function satisfying (SVP-1,2,3) in Definition B.220 is unique and is the same as the determinant in (B.85).

Proof. Let the parallelotope be spanned by the column vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. We have

$$\begin{aligned}
&\delta \begin{bmatrix} v_{11} & v_{12} & \ldots & v_{1n} \\ v_{21} & v_{22} & \ldots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \ldots & v_{nn} \end{bmatrix} \\
&= \sum_{i_1=1}^n v_{i_1 1} \delta \begin{bmatrix} | & v_{12} & \ldots & v_{1n} \\ e_{i_1} & v_{22} & \ldots & v_{2n} \\ | & \vdots & \ddots & \vdots \\ | & v_{n2} & \ldots & v_{nn} \end{bmatrix} \\
&= \sum_{i_1,i_2=1}^n v_{i_1 1} v_{i_2 2} \delta \begin{bmatrix} | & | & v_{13} & \ldots & v_{1n} \\ e_{i_1} & e_{i_2} & v_{23} & \ldots & v_{2n} \\ | & | & \vdots & \ddots & \vdots \\ | & | & v_{n2} & \ldots & v_{nn} \end{bmatrix} \\
&= \cdots \\
&= \sum_{i_1,i_2,\ldots,i_n=1}^n v_{i_1 1} v_{i_2 2} \cdots v_{i_n n} \delta \begin{bmatrix} | & | & \ldots & | \\ e_{i_1} & e_{i_2} & \ldots & e_{i_n} \\ | & | & \ldots & | \end{bmatrix} \\
&= \sum_{\sigma \in S_n} v_{\sigma(1),1} v_{\sigma(2),2} \cdots v_{\sigma(n),n} \delta \begin{bmatrix} | & | & \ldots & | \\ e_{\sigma(1)} & e_{\sigma(2)} & \ldots & e_{\sigma(n)} \\ | & | & \ldots & | \end{bmatrix} \\
&= \sum_{\sigma \in S_n} v_{\sigma(1),1} v_{\sigma(2),2} \cdots v_{\sigma(n),n} \mathrm{sgn}(\sigma) \\
&= \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_{i=1}^n v_{\sigma(i),i},
\end{aligned}$$

where the first four steps follow from (SVP-3), the sixth step from Lemma B.226, and the fifth step from (SVP-2). In other words, the signed volume $\delta(\cdot)$ is zero for any $i_j = i_k$ and hence the only nonzero terms are those of which $(i_1, i_2, \ldots, i_n)$ is a permutation of $(1, 2, \ldots, n)$. □

Exercise B.230. Use the formula in (B.85) to show that $\det A = \det A^T$.

Definition B.231. The i,j *cofactor* of $A \in \mathbb{R}^{n\times n}$ is

$$C_{ij} = (-1)^{i+j} M_{ij}, \tag{B.86}$$

where M_{ij} is the i,j *minor of a matrix* A, i.e. the determinant of the $(n-1)\times(n-1)$ matrix that results from deleting the i-th row and the j-th column of A.

Theorem B.232 (Laplace formula of determinants)**.** Given fixed indices $i, j \in 1, 2, \ldots, n$, the determinant of an n-by-n matrix $A = [a_{ij}]$ is given by

$$\det A = \sum_{j'=1}^n a_{ij'} C_{ij'} = \sum_{i'=1}^n a_{i'j} C_{i'j}. \tag{B.87}$$

Exercise B.233. Prove Theorem B.232 by induction.