Notes on linear algebra

Danny Nygård Hansen

30th May 2023

1 • Introduction

These notes cover aspects of linear algebra that I have not found satisfactory expositions of elsewhere. We generally restrict ourselves to the finite-dimensional case, unless results can be generalised without significant effort. For instance, in the context of inner product spaces there is of course no loss in generality by restricting to the real or the complex numbers, and the elementary theory of Hilbert space adjoints is not simplified substantially by the assumption of finite dimension, so we make no such assumption. On the other hand, we only prove the spectral theorem for normal operators on finite-dimensional spaces.

2 • Linear equations and matrices

2.1. *Linear equations*

Throughout we let $\mathbb{F}$ denote an arbitrary field and R a commutative ring. Let m and n be positive integers. A *linear equation in n unknowns* is an equation on the form

$$l\colon a_1x_1 + \cdots + a_nx_n = b,$$

where $a_1,\ldots,a_n,b \in \mathbb{F}$. A *solution* to l is an element $v = (v_1,\ldots,v_n) \in \mathbb{F}^n$ such that

$$a_1v_1 + \cdots + a_nv_n = b.$$

A *system of linear equations in n unknowns* is a tuple $L = (l_1,\ldots,l_m)$, where each l_i is a linear equation in n unknowns. An element $v \in \mathbb{F}^n$ is a *solution* to L if it is a solution to each linear equation $l_1,\ldots,l_m$.

Let L and L' be systems of linear equations in n unknowns. We say that L and L' are *solution equivalent* if they have the same solutions. Furthermore, we say that they are *combination equivalent* if each equation in L' is a linear combination of the equations in L, and vice versa. Clearly, if L and L' are combination equivalent they are also solution equivalent, but the converse does not hold.

2.2. Matrices

It is well-known that a system of linear equations is equivalent to a matrix equation on the form $Ax = b$, where $A \in \mathrm{Mat}_{m,n}(\mathbb{F})$, $x \in \mathbb{F}^n$ and $b \in \mathbb{F}^m$. Recall the *elementary row operations* on A:

(1) multiplication of one row of A by a nonzero scalar,

(2) addition to one row of A a scalar multiple of another (different) row, and

(3) interchange of two rows of A.

If e is an elementary row operation, we write $e(A)$ for the matrix obtained when applying e to A. Clearly each elementary row operation e has an 'inverse', i.e. an elementary row operation e' such that $e'(e(A)) = e(e'(A)) = A$. Two matrices $A, B \in \mathrm{Mat}_{m,n}(\mathbb{F})$ are called *row-equivalent* if A is obtained by applying a finite sequence of elementary row operations to B (and vice versa, though this need not be assumed since each elementary row operation has an inverse).

Clearly, if $A, B \in \mathrm{Mat}_{m,n}(\mathbb{F})$ are row-equivalent, then the systems of equations $Ax = 0$ and $Bx = 0$ are combination equivalent, hence have the same solutions.

Definition 2.1

A matrix $H \in \mathrm{Mat}_{m,n}(\mathbb{F})$ is called *row-reduced* if

(i) the first nonzero entry of each nonzero row in H is 1, and

(ii) each column of H containing the leading nonzero entry of some row has all its other entries equal 0.

If H is row-reduced, it is called a *row-reduced echelon matrix* if it also has the following properties:

(iii) Every row of H only containing zeroes occur below every row which has a nonzero entry, and

(iv) if rows $1, \dots, r$ are the nonzero rows of H, and if the leading nonzero entry of row i occurs in column k_i, then $k_1 < \cdots < k_r$.

An *elementary matrix* is a matrix obtained by applying a single elementary row operation to the identity matrix I. It is easy to show that if e is an elementary row operation and $E = e(I) \in \mathrm{Mat}_m(\mathbb{F})$, then $e(A) = EA$ for $A \in \mathrm{Mat}_{m,n}(\mathbb{F})$. If $B \in \mathrm{Mat}_{m,n}(\mathbb{F})$, then A and B are row-equivalent if and only if $A = PB$, where $P \in \mathrm{Mat}_m(\mathbb{F})$ is a product of elementary matrices.

Proposition 2.2

Every matrix in $\mathrm{Mat}_{m,n}(\mathbb{F})$ *is row-equivalent to a unique row-reduced echelon matrix.*

Proof. The usual Gauss–Jordan elimination algorithm proves existence. If $H, K \in \mathrm{Mat}_{m,n}(R)$ are row-equivalent row-reduced echelon matrices, we claim that $H = K$. We prove this by induction in n. If $n = 1$ then this is obvious, so assume that $n > 1$. Let H_1 and K_1 be the matrices obtained by deleting the nth column in H and K respectively. Then H_1 and K_1 are also row-equivalent[1] and row-reduced echelon matrices, so by induction $H_1 = K_1$. Thus if H and K differ, they must differ in the nth column.

Let H_2 be the matrix obtained by deleting columns in H, only keeping those columns containing pivots, as well as keeping the nth column. Define K_2 similarly. Thus we have deleted the same columns in H and K, so H_2 and K_2 are also row-equivalent. Say that the number of columns in H_2 and K_2 is $r + 1$, and write the matrices on the form

$$H_2 = \begin{pmatrix} I_r & h \\ 0 & h' \end{pmatrix} \quad \text{and} \quad K_2 = \begin{pmatrix} I_r & k \\ 0 & k' \end{pmatrix},$$

where $h, k \in \mathbb{F}^r$ and $h', k' \in \mathbb{F}^{m-r}$ are column vectors. Since H_2 and K_2 are row-equivalent, the systems $H_2x = 0$ and $K_2x = 0$ are solution equivalent. If $h' = 0$, then $H_2x = 0$ has the solution $(-h, 1)$. But this is also a solution to $K_2x = 0$, so $h = k$ and $k' = 0$. If $h' \neq 0$, then $H_2x = 0$ only has the trivial solution. But then $K_2x = 0$ also only has the trivial solution, and hence $k' \neq 0$. But that must be because both H_2 and K_2 has a pivot in the rightmost column, so also in this case $H_2 = K_2$. □

2.3. Invertible matrices

Notice that elementary matrices are invertible, since elementary row operations are invertible.

[1] It should be obvious that deleting columns preserves row-equivalence, but we give a more precise argument: If $P \in \mathrm{Mat}_m(\mathbb{F})$ is a product of elementary matrices and $a_1, \ldots, a_n \in \mathbb{F}^m$ are the columns in A, then the columns in PA are $Pa_1, \ldots, Pa_m$. Thus elementary row operations are applied to each column independently of the other columns.

LEMMA 2.3

If $A \in \mathrm{Mat}_n(\mathbb{F})$, then the following are equivalent:

(i) *A is invertible,*

(ii) *A is row-equivalent to I_n,*

(iii) *A is a product of elementary matrices, and*

(iv) *the system $Ax = 0$ has only the trivial solution $x = 0$.*

PROOF. *(i) ⇔ (ii)*: Let $H \in \mathrm{Mat}_n(\mathbb{F})$ be a row-reduced echelon matrix that is row-equivalent to A. Then $H = PA$, where $P \in \mathrm{Mat}_n(\mathbb{F})$ is a product of elementary matrices. Then $A = P^{-1}H$, so A is invertible if and only if H is. But the only invertible row-reduced echelon matrix is the identity matrix, so (i) and (ii) are equivalent.

(ii) ⇒ (iii): As above, there exists a product P of elementary matrices such that $I_n = PA$, so $A = P^{-1}$.

(iii) ⇒ (i): This is obvious since elementary matrices are invertible.

(ii) ⇔ (iv): If A and I_n are row-equivalent, then the systems $Ax = 0$ and $I_n x = 0$ have the same solutions. Conversely, assume that $Ax = 0$ only has the trivial solution. If $H \in \mathrm{Mat}_{m,n}(\mathbb{F})$ is a row-reduced echelon matrix that is row-equivalent to A, then $Hx = 0$ has no nontrivial solution. Thus if r is the number of nonzero rows in H, then $r \geq n$. But then $r = n$, so H must be the identity matrix. □

PROPOSITION 2.4

Let $A \in \mathrm{Mat}_n(\mathbb{F})$. Then the following are equivalent:

(i) *A is invertible,*

(ii) *A has a left inverse, and*

(iii) *A has a right inverse.*

PROOF. If A has a left inverse, then $Ax = 0$ has no nontrivial solution, so A is invertible. If A has a right inverse $B \in \mathrm{Mat}_n(\mathbb{F})$, i.e. $AB = I$, then B has a left inverse and is thus invertible. But then A is the inverse of B and hence is itself invertible. □

3 • Bases and coordinates

3.1. Bases

Let V be a vector space. A *Hamel basis* for V is a linearly independent set $\mathcal{V} \subseteq V$ that spans V, i.e. for every $v \in V$ there exist unique (up to ordering) $v_1, \dots, v_n \in \mathcal{V}$ and $\alpha_1, \dots, \alpha_n \in \mathbb{F}$ such that $v = \sum_{i=1}^n \alpha_i v_i$. In other words, a Hamel basis is a maximal linearly independent subset of V.

If V is an inner product space, a subset $\mathcal{O}$ of V is said to be *orthogonal* if $v \neq w$ implies $v \perp w$ for $v, w \in \mathcal{O}$. Furthermore, if every element of $\mathcal{O}$ is a unit vector, then $\mathcal{O}$ is called *orthonormal*. If $\mathcal{V}$ is a Hamel basis for V that is also an orthogonal/orthonormal set, then $\mathcal{V}$ is called an *orthogonal/orthonormal Hamel basis*. If every vector in an orthogonal set $\mathcal{O}$ is nonzero, then $\mathcal{O}$ gives rise to an orthonormal set by dividing each vector by its norm. This clearly preserves the span of every subset of $\mathcal{O}$; in particular, if $\mathcal{O}$ is an orthogonal Hamel basis then this modification yields an orthonormal Hamel basis.

There is also another notion of basis in an inner product space: A maximal orthonormal subset of V is called a *Hilbert basis*. An orthonormal Hamel basis is thus a Hilbert basis, but not vice-versa. For instance, the 'standard basis' of $\mathbb{R}^{\mathbb{N}}$ consisting of sequences $e_n = (0, \dots, 0, 1, 0, \dots)$ with a 1 in the nth place and zeros elsewhere is a Hilbert basis for the space l^2, but it is not a Hamel basis for l^2 since its linear span is the *coproduct* $\mathbb{R}^{\oplus\mathbb{N}}$, i.e. the subspace of l^2 of sequences with finitely many nonzero elements.

Zorn's lemma can be used to show that every vector space has a Hamel basis, and that every inner product space has a Hilbert basis (every inner product space of course also has a Hamel basis). However, not every inner product space has an *orthonormal* Hamel basis. For instance, let $\mathcal{H}$ be an infinite-dimensional Hilbert space, let $\mathcal{O}$ be an infinite orthonormal subset of $\mathcal{H}$, and let $(e_n)_{n\in\mathbb{N}}$ be a sequence of distinct elements from $\mathcal{O}$. Then the sum of the series $\sum_{n=1}^n e_n/n$ lies in $\mathcal{H}$ by completeness, but this cannot be expressed as a finite linear combination of elements in $\mathcal{O}$, since the terms in the sum become arbitrarily small.

The argument above in particular shows that the (Hamel) dimension of an infinite-dimensional Hilbert space $\mathcal{H}$ is uncountable. For if $\mathcal{I}$ is any countable, linearly independent collection of elements from $\mathcal{H}$, then the Gram–Schmidt process yields an orthonormal collection $\mathcal{O} \subseteq \mathcal{H}$ with $\operatorname{span}\mathcal{O} = \operatorname{span}\mathcal{I}$. But the above shows that $\mathcal{O}$ cannot span $\mathcal{H}$, so neither can $\mathcal{I}$.

Let V be a vector space. A *series* of subspaces U_i of V is a finite or infinite decreasing sequence

$$V = U_0 \supsetneq U_1 \supsetneq U_2 \supsetneq \cdots.$$

If the sequence is finite, then the *length* of the series is the number of strict inclusions. If the sequence is infinite, then we say that the length of the series is ∞. The maximal length of a series of subspaces of V is denoted $l(V)$.

In the proposition below, we write $\dim V = \infty$ if the dimension of V is infinite.

PROPOSITION 3.1

Let V be a vector space. Then $\dim V = l(V)$.

PROOF. First assume that V is finite-dimensional, and let $\mathcal{V} = (v_1,\dots,v_n)$ be a basis for V. Then there is a series

$$V = \operatorname{span}(v_1,\dots,v_n) \supsetneq \operatorname{span}(v_1,\dots,v_{n-1}) \supsetneq \cdots \supsetneq \operatorname{span}(v_1) \supsetneq 0$$

of subspaces of V, so $\dim V \le l(V)$. Conversely, let

$$V = U_0 \supsetneq U_1 \supsetneq U_2 \supsetneq \cdots$$

be a series of subspaces of V. If the series ends with 0, remove it. Hence all subspaces in the series are nontrivial. Then choose for each i an element $v_i \in U_i \setminus U_{i+1}$, and collect them in a set $\mathcal{I}$. It is clear that $\mathcal{I}$ is linearly independent, hence finite. Thus the series is also finite with length $|\mathcal{I}| - 1$. Adding back 0 to the series we obtain a series that is at least as long as the original sequence, and that is of length $|\mathcal{I}| \le \dim V$. Since the sequence was arbitrary, $l(V) \le \dim V$.

Next assume that V is infinite-dimensional. Then V contains a sequence $(v_i)_{i\in\mathbb{N}}$ that is linearly independent, so the series

$$V \supseteq \operatorname{span}\{v_i \mid i \in \mathbb{N}\} \supsetneq \operatorname{span}\{v_i \mid i \ge 2\} \supsetneq \operatorname{span}\{v_i \mid i \ge 3\} \supsetneq \cdots$$

is infinite, and $l(V) = \infty$. Conversely, assume that V has an infinite series. As above we construct a linearly independent set $\mathcal{I}$ whose size equals the length of the sequence. Thus V contains an infinite linearly independent set, so $\dim V = \infty$. □

3.2. *Coordinate maps and matrices*

For $A \in \operatorname{Mat}_{m,n}(\mathbb{F})$ we define the map $M_A \colon \mathbb{F}^n \to \mathbb{F}^m$ by $M_A v = Av$.

PROPOSITION 3.2

Let $(e_1,\dots,e_n)$ be the standard basis for $\mathbb{F}^n$. The map

$$\begin{aligned} \mathcal{M} \colon \mathcal{L}(\mathbb{F}^n,\mathbb{F}^m) &\to \operatorname{Mat}_{m,n}(\mathbb{F}), \\ T &\mapsto \big(Te_1 \mid \cdots \mid Te_n\big), \end{aligned}$$

is a linear isomorphism with inverse $A \mapsto M_A$. The matrix $\mathcal{M}(T)$ is called the standard matrix representation *of T. If $T \colon \mathbb{F}^n \to \mathbb{F}^m$ and $S \colon \mathbb{F}^m \to \mathbb{F}^l$ are linear maps, then*

(i) $Tv = \mathcal{M}(T)v$ *for all* $v \in \mathbb{F}^n$.

(ii) $\mathcal{M}(\mathrm{id}_{\mathbb{F}^n}) = I$.

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

(iv) T *is invertible if and only if* $\mathcal{M}(T)$ *is invertible, in which case* $\mathcal{M}(T^{-1}) = \mathcal{M}(T)^{-1}$.

Proof. The map $A \mapsto M_A$ is clearly linear, so to prove the first point it suffices to show that this is the inverse of $\mathcal{M}$. Let $T \in \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$. Then

$$M_{\mathcal{M}(T)}\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix} = \mathcal{M}(T)\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix} = \begin{pmatrix}Te_1 \mid \cdots \mid Te_n\end{pmatrix}\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix} = \sum_{i=1}^n \alpha_i Te_i = T\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix}$$

for $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$. Conversely, for $A \in \mathrm{Mat}_{m,n}(\mathbb{F})$ we have

$$\mathcal{M}(M_A) = \begin{pmatrix}M_A e_1 \mid \cdots \mid M_A e_n\end{pmatrix} = \begin{pmatrix}Ae_1 \mid \cdots \mid Ae_n\end{pmatrix} = A,$$

since Ae_i is the ith column of A. We prove the remaining claims:

Proof of (i): Simply notice that $Tv = M_{\mathcal{M}(T)}v = \mathcal{M}(T)v$.

Proof of (ii): This is obvious from the definition of $\mathcal{M}$.

Proof of (iii): Let $v \in \mathbb{F}^n$ and notice that

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

by (i). Since this holds for all v, the claim follows.

Proof of (iv): This follows easily from (ii) and (iii). □

Let V be a finite-dimensional $\mathbb{F}$-vector space. If $\mathcal{V} = (v_1, \ldots, v_n)$ is an ordered basis for V, then for every $v \in V$ there are unique $\alpha_1, \ldots, \alpha_n \in \mathbb{F}$ such that $v = \sum_{i=1}^n \alpha_i v_i$. Hence the map $\varphi_{\mathcal{V}} \colon V \to \mathbb{F}^n$ given by $\varphi_{\mathcal{V}}(v) = (\alpha_1, \ldots, \alpha_n)$ is well-defined. Furthermore, it is clearly linear, and since $\mathcal{V}$ is a basis it is also bijective, hence a linear isomorphism. The map $\varphi_{\mathcal{V}}$ is called the *coordinate map* with respect to $\mathcal{V}$, and the vector $[v]_{\mathcal{V}} = \varphi_{\mathcal{V}}(v)$ is called the *coordinate vector* of v with respect to $\mathcal{V}$.

Now let $\mathcal{W}$ be another ordered basis for V. The composition $\varphi_{\mathcal{W},\mathcal{V}} = \varphi_{\mathcal{W}} \circ \varphi_{\mathcal{V}}^{-1}$ is called the *change of basis operator* from $\mathcal{V}$ to $\mathcal{W}$, and this makes the diagram

$$\begin{tikzcd} & \mathbb{F}^n \arrow[dd, "\varphi_{\mathcal{W},\mathcal{V}}"] \\ V \arrow[ur, "\varphi_{\mathcal{V}}"] \arrow[dr, "\varphi_{\mathcal{W}}"'] & \\ & \mathbb{F}^n \end{tikzcd} \tag{3.1}$$

commute. Its standard matrix is denoted ${}_{\mathcal{W}}[\square]_{\mathcal{V}}$. This has the expected properties:

Proposition 3.3

Let $\mathcal{V}$, $\mathcal{W}$ and $\mathcal{U}$ be ordered bases for a finite-dimensional $\mathbb{F}$-vector space V. Then

(i) $[v]_{\mathcal{W}} = \varphi_{\mathcal{W},\mathcal{V}}([v]_{\mathcal{V}})$ *for all* $v \in V$. *In particular,* $[v]_{\mathcal{W}} = {}_{\mathcal{W}}[\square]_{\mathcal{V}} \cdot [v]_{\mathcal{V}}$.

(ii) $\varphi_{\mathcal{V},\mathcal{V}}$ *is the identity map. In particular,* ${}_{\mathcal{V}}[\square]_{\mathcal{V}}$ *is the identity matrix.*

(iii) $\varphi_{\mathcal{U},\mathcal{W}} \circ \varphi_{\mathcal{W},\mathcal{V}} = \varphi_{\mathcal{U},\mathcal{V}}$. *In particular,* ${}_{\mathcal{U}}[\square]_{\mathcal{W}} \cdot {}_{\mathcal{W}}[\square]_{\mathcal{V}} = {}_{\mathcal{U}}[\square]_{\mathcal{V}}$.

(iv) $\varphi_{\mathcal{W},\mathcal{V}}$ *(resp.* ${}_{\mathcal{W}}[\square]_{\mathcal{V}}$*) is invertible with inverse* $\varphi_{\mathcal{V},\mathcal{W}}$ *(resp.* ${}_{\mathcal{V}}[\square]_{\mathcal{W}}$*).*

Proof. All claims about change of basis matrices follow by Proposition 3.2 from the corresponding claims about change of basis operators.

The claim (i) follows by commutativity of the diagram (3.1), i.e.

$$\varphi_{\mathcal{W},\mathcal{V}}([v]_{\mathcal{V}}) = (\varphi_{\mathcal{W}} \circ \varphi_{\mathcal{V}}^{-1}) \circ \varphi_{\mathcal{V}}(v) = \varphi_{\mathcal{W}}(v) = [v]_{\mathcal{W}}.$$

Claim (ii) is an immediate consequence of the definition of $\varphi_{\mathcal{V},\mathcal{V}}$. The remaining claims are proved similarly to (i). $\square$

Next consider a linear map $T\colon V \to W$. If $\mathcal{V} \in V^n$ and $\mathcal{W} \in W^m$ are bases for V and W respectively, then the diagram

$$\begin{tikzcd}
V \arrow[r, "\varphi_{\mathcal{V}}"] \arrow[d, "T"'] & \mathbb{F}^n \arrow[d, "\varphi_{\mathcal{W}} \circ T \circ \varphi_{\mathcal{V}}^{-1}"] \\
W \arrow[r, "\varphi_{\mathcal{W}}"'] & \mathbb{F}^n
\end{tikzcd}$$

commutes. The map $\varphi_{\mathcal{W}} \circ T \circ \varphi_{\mathcal{V}}^{-1}$ is the *basis representation of* T with respect to the bases $\mathcal{V}$ and $\mathcal{W}$. This is a linear map $\mathbb{F}^n \to \mathbb{F}^m$, so it has a standard matrix which we denote ${}_{\mathcal{W}}[T]_{\mathcal{V}}$. This is called the *matrix representation* of T with respect to the bases $\mathcal{V}$ and $\mathcal{W}$.

Proposition 3.4

Let V and W be finite-dimensional $\mathbb{F}$-vector spaces with ordered bases $\mathcal{V} = (v_1, \dots, v_n) \in V^n$ and $\mathcal{W} \in W^m$, respectively. The map

$$\begin{aligned} {}_{\mathcal{W}}[\,\cdot\,]_{\mathcal{V}}\colon \mathcal{L}(V,W) &\to \mathrm{Mat}_{m,n}(\mathbb{F}), \\ T &\mapsto {}_{\mathcal{W}}[T]_{\mathcal{V}}, \end{aligned}$$

is a linear isomorphism. Let $T\colon V \to W$ and $S\colon W \to U$ be linear maps, and let $\mathcal{U} \in U^l$ be an ordered basis for U. Then

(i) ${}_{\mathcal{W}}[T]_{\mathcal{V}} = \big([Tv_1]_{\mathcal{W}} \mid \cdots \mid [Tv_n]_{\mathcal{W}}\big)$.

(ii) $[Tv]_{\mathcal{W}} = {}_{\mathcal{W}}[T]_{\mathcal{V}} \cdot [v]_{\mathcal{V}}$ *for all* $v \in V$.

(iii) *If* $\mathcal{V}'$ *is another basis for* V*, then* ${}_{\mathcal{V}'}[\mathrm{id}_V]_{\mathcal{V}} = {}_{\mathcal{V}'}[\square]_{\mathcal{V}}$.

(iv) ${}_{\mathcal{U}}[S \circ T]_{\mathcal{V}} = {}_{\mathcal{U}}[S]_{\mathcal{W}} \cdot {}_{\mathcal{W}}[T]_{\mathcal{V}}$.

(v) T *is invertible if and only if* ${}_{\mathcal{W}}[T]_{\mathcal{V}}$ *is invertible, in which case* ${}_{\mathcal{V}}[T^{-1}]_{\mathcal{W}} = {}_{\mathcal{W}}[T]_{\mathcal{V}}^{-1}$.

PROOF. For the first claim, notice that the map $T \mapsto \varphi_{\mathcal{W}} \circ T \circ \varphi_{\mathcal{V}}^{-1}$ is a linear isomorphism, since pre- and postcomposition with linear isomorphisms are themselves linear isomorphisms. Composing this map with $\mathcal{M}$ yields ${}_{\mathcal{W}}[\cdot]_{\mathcal{V}}$, so this is a linear isomorphism by Proposition 3.2.

Proof of (i): If $(e_1, \dots, e_n)$ is the standard basis for $\mathbb{F}^n$, then the definition of the standard matrix representation yields that the ith column of ${}_{\mathcal{W}}[T]_{\mathcal{V}}$ is given by

$${}_{\mathcal{W}}[T]_{\mathcal{V}} \cdot e_i = \mathcal{M}(\varphi_{\mathcal{W}} \circ T \circ \varphi_{\mathcal{V}}^{-1}) \cdot e_i = (\varphi_{\mathcal{W}} \circ T \circ \varphi_{\mathcal{V}}^{-1}) e_i = \varphi_{\mathcal{W}}(Tv_i) = [Tv_i]_{\mathcal{W}},$$

as claimed.

Proof of (ii): Notice that

$$\begin{aligned} [Tv]_{\mathcal{W}} &= (\varphi_{\mathcal{W}} \circ T)(v) \\ &= (\varphi_{\mathcal{W}} \circ T \circ \varphi_{\mathcal{V}}^{-1}) \circ \varphi_{\mathcal{V}}(v) \\ &= (\varphi_{\mathcal{W}} \circ T \circ \varphi_{\mathcal{V}}^{-1})([v]_{\mathcal{V}}) \\ &= {}_{\mathcal{W}}[T]_{\mathcal{V}} \cdot [v]_{\mathcal{V}}. \end{aligned}$$

where the last equality follows from Proposition 3.2(i).

Proof of (iii): This is obvious from the definitions of ${}_{\mathcal{V}'}[\mathrm{id}_V]_{\mathcal{V}}$ and ${}_{\mathcal{V}'}[\square]_{\mathcal{V}}$.

Proof of (iv): Notice that

$$\varphi_{\mathcal{U}} \circ (S \circ T) \circ \varphi_{\mathcal{V}}^{-1} = (\varphi_{\mathcal{U}} \circ S \circ \varphi_{\mathcal{W}}^{-1}) \circ (\varphi_{\mathcal{W}} \circ T \circ \varphi_{\mathcal{V}}^{-1})$$

The claim then follows from Proposition 3.2(iii).

Proof of (v): This is an immediate consequence of either (iv) or of Proposition 3.2(iv). □

PROPOSITION 3.5

Let $\mathcal{V} = (v_1, \dots, v_n)$ *be an ordered basis for an* $\mathbb{F}$*-vector space* V*, and let* $T \colon V \to V$ *be a linear isomorphism. Let* $\mathcal{W} = (w_1, \dots, w_n)$ *where* $w_i = Tv_i$*. Then* $\mathcal{W}$ *is an* *ordered basis for V and*

$$\varphi_{\mathcal{W},\mathcal{V}} = \varphi_{\mathcal{V}} \circ T^{-1} \circ \varphi_{\mathcal{V}}^{-1}, \quad \text{or} \quad {}_{\mathcal{W}}[\square]_{\mathcal{V}} = {}_{\mathcal{V}}[T^{-1}]_{\mathcal{V}}.$$

In particular, if $V = \mathbb{F}^n$ and $\mathcal{V}$ is the standard basis $\mathcal{E}$, then

$$\varphi_{\mathcal{W},\mathcal{E}} = T^{-1}, \quad \text{or} \quad {}_{\mathcal{W}}[\square]_{\mathcal{E}} = \mathcal{M}(T^{-1}).$$

We think of this result as follows: If we change basis by applying an invertible linear transformation T, we obtain the coordinate vectors corresponding to the transformed basis by applying T^{-1} (in the old basis). This says that if we perform a *passive transformation*, i.e. a change of basis while keeping vectors themselves fixed, the coordinates change by the inverse of said transformation.

PROOF. Let $v \in V$ and write $v = \sum_{i=1}^n \alpha_i v_i$. Then

$$Tv = \sum_{i=1}^n \alpha_i T v_i = \sum_{i=1}^n \alpha_i w_i = \varphi_{\mathcal{W}}^{-1} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \varphi_{\mathcal{W}}^{-1} \circ \varphi_{\mathcal{V}}(v),$$

implying that

$$\varphi_{\mathcal{W},\mathcal{V}} = \varphi_{\mathcal{W}} \circ \varphi_{\mathcal{V}}^{-1} = (T \circ \varphi_{\mathcal{V}}^{-1})^{-1} \circ \varphi_{\mathcal{V}}^{-1} = \varphi_{\mathcal{V}} \circ T^{-1} \circ \varphi_{\mathcal{V}}^{-1}$$

as claimed. ☐

[TODO] Recall that two matrices $A, B \in \mathrm{Mat}_n(\mathbb{F})$ are *similar* if there exists an invertible matrix $P \in \mathrm{Mat}_n(\mathbb{F})$ such that $A = PBP^{-1}$.

4 • Determinants

4.1. Existence of determinants

If $M_1, \dots, M_n, N$ are modules over a commutative ring R, a map

$$\varphi\colon M_1 \times \cdots \times M_n \to N$$

is called *n-linear* if, for all i, the maps $m_i \mapsto \varphi(m_1, \dots, m_n)$ are linear for all choices of $m_j \in M_j$ where $j \neq i$. Since there is a natural isomorphism $\mathrm{Mat}_{m,n}(R) \cong (R^n)^m$, a map $\varphi\colon \mathrm{Mat}_{m,n}(R) \to N$ that is linear in each row is also called n-linear.

In the case $M_1 = \cdots = M_n$, we call φ *alternating* if $\varphi(m_1, \dots, m_n) = 0$ whenever $m_i = m_j$ for some $i \neq j$. Furthermore, φ is called *skew-symmetric* if

$$\begin{aligned}\varphi(m_1, \dots, m_{i-1}, m_i, m_{i+1}, \dots, m_{j-1}, m_j, m_{j+1}, \dots, m_n)& \\ = -\varphi(m_1, \dots, m_{i-1}, m_j, m_{i+1}, \dots, m_{j-1}, m_i, m_{j+1}, \dots, m_n)&\end{aligned}$$

for all $i < j$.

LEMMA 4.1

Let M and N be R-modules, and let $\varphi\colon M^n \to N$ be an n-linear map.

(i) *If φ is alternating, then φ is skew-symmetric. If* $\operatorname{char} R \neq 2$ *then the converse also holds.*

(ii) *If $\varphi(m_1,\dots,m_n) = 0$ whenever $m_i = m_{i+1}$ for some $i = 1,\dots,n-1$, then φ is alternating.*

We shall not use the converse direction of Lemma 4.1(i) but we include it for completeness.

PROOF. *Proof of (i)*: Consider $m_1,\dots,m_n \in M$, and let $1 \leq i < j \leq n$. Define a map $\psi\colon M \times M \to N$ by

$$\psi(a,b) = \varphi(m_1,\dots,m_{i-1},a,m_{i+1},\dots,m_{j-1},b,m_{j+1},\dots,m_n),$$

and notice that it suffices to show that $\psi(m_i,m_j) = -\psi(m_j,m_i)$. But ψ is 2-linear and alternating, so for $a,b \in M$ we have

$$\psi(a+b,a+b) = \psi(a,a) + \psi(a,b) + \psi(b,a) + \psi(b,b) = \psi(a,b) + \psi(b,a).$$

Thus $\psi(m_i,m_j) = -\psi(m_j,m_i)$, so φ is skew-symmetric as claimed.

Conversely, if $\operatorname{char} R \neq 2$ and ψ is skew-symmetric, then since $\psi(a,b) = -\psi(b,a)$, letting $a = b$ we have $2\psi(a,a) = 0$, so $\psi(a,a) = 0$.

Proof of (ii): The argument above shows that, in particular, if $A,B \in M^n$, and B is obtained from A by interchanging two adjacent elements, then $\varphi(B) = -\varphi(A)$. Assuming now that B is obtained from A by interchanging the ith and jth elements in A, with $i < j$, we claim that we may obtain B by successively interchanging adjacent elements of A. Writing $A = (m_1,\dots,m_n)$, we first perform $j-i$ such interchanges and arrive that the tuple

$$(m_1,\dots,m_{i-1},m_{i+1},\dots,m_{j-1},m_j,m_i,m_{j+1},\dots,m_n),$$

moving m_i to the right $j-i$ places. Next we perform another $j-i-1$ interchanges, moving m_j to the left until we reach

$$B = (m_1,\dots,m_{i-1},m_j,m_{i+1},\dots,m_{j-1},m_i,m_{j+1},\dots,m_n).$$

Since each interchange results in a sign change, we have

$$\varphi(B) = (-1)^{2(j-i)-1}\varphi(A) = -\varphi(A).$$

If $m_i = m_j$ for $i < j$, then we claim that $\varphi(A) = 0$. For let B be obtained from A by interchanging m_{i+1} and m_j. Then $\varphi(B) = 0$, so $\varphi(A) = -\varphi(B) = 0$ by the above argument, and hence φ is alternating as claimed. ∎

DEFINITION 4.2: *Determinant functions*

If n be a positive integer, a *determinant function* is a map $\varphi\colon \mathrm{Mat}_n(R) \to R$ that is n-linear, alternating, and which satisfies $\varphi(I_n) = 1$.

If $A \in \mathrm{Mat}_n(R)$ with $n > 1$ and $1 \leq i, j \leq n$, denote by $M(A)_{i,j}$ the matrix in $\mathrm{Mat}_{n-1}(R)$ obtained by removing the the ith row and the jth column of A. This is called the (i,j)*-th minor* of A. If $\varphi\colon \mathrm{Mat}_{n-1}(R) \to R$ is an $(n-1)$-linear function and $A \in \mathrm{Mat}_n(R)$, then we write $\varphi_{i,j}(A) = \varphi(M(A)_{i,j})$. Then $\varphi_{i,j}\colon \mathrm{Mat}_n(R) \to R$ is clearly linear in all rows except row i, and is independent of row i.

THEOREM 4.3: *Construction of determinants*

Let $n > 1$, and let $\varphi\colon \mathrm{Mat}_{n-1}(R) \to R$ be alternating and $(n-1)$-linear. For $j = 1, \dots, n$ define a map $\psi_j\colon \mathrm{Mat}_n(R) \to R$ by

$$\psi_j(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \varphi_{i,j}(A),$$

for $A = (a_{ij}) \in \mathrm{Mat}_n(R)$. Then ψ_j is alternating and n-linear. If φ is a determinant function, then so is ψ_j.

PROOF. Let $A = (a_{ij}) \in \mathrm{Mat}_n(R)$. Then $A \mapsto a_{ij}$ is independent of all rows except row i, and $\varphi_{i,j}$ is linear in all rows except row i. Thus $A \mapsto a_{ij}\varphi_{i,j}(A)$ is linear in all rows except row i. Conversely, $A \mapsto a_{ij}$ is linear in row i, and $\varphi_{i,j}$ is independent of row i, so $A \mapsto a_{ij}\varphi_{i,j}(A)$ is also linear in row i. Since ψ_j is a linear combination of n-linear maps, is it itself n-linear.

Now assume that A has two equal adjacent rows, say $a_k, a_{k+1} \in R^n$. If $i \neq k$ and $i \neq k+1$, then $M(A)_{i,j}$ has two equal rows, so $\varphi_{i,j}(A) = 0$. Thus

$$\psi_j(A) = (-1)^{k+j} a_{kj} \varphi_{k,j}(A) + (-1)^{k+1+j} a_{(k+1)j} \varphi_{k+1,j}(A).$$

Since $a_k = a_{k+1}$ we also have $a_{kj} = a_{(k+1)j}$ and $M(A)_{k,j} = M(A)_{k+1,j}$. Thus $\psi_j(A) = 0$, so Lemma 4.1(ii) implies that ψ_j is alternating.

Finally suppose that φ is a determinant function. Then $M(I_n)_{j,j} = I_{n-1}$ and we have

$$\psi_j(I_n) = (-1)^{j+j} \varphi_{j,j}(I_n) = \varphi(I_{n-1}) = 1,$$

so ψ_j is also a determinant function. □

COROLLARY 4.4: *Existence of determinants*

For every positive integer n, there exists a determinant function $\mathrm{Mat}_n(R) \to R$.

PROOF. The identity map on $\mathrm{Mat}_1(R) \cong R$ is a determinant function for $n = 1$, and Theorem 4.3 allows us to recursively construct a determinant for each $n > 1$. □

4.2. *Uniqueness of determinants*

THEOREM 4.5: *Uniqueness of determinants*

Let n be a positive integer. There is precisely one determinant function on $\mathrm{Mat}_n(R)$, namely the function $\det\colon \mathrm{Mat}_n(R) \to R$ given by

$$\det A = \sum_{\sigma \in S_n} (\operatorname{sgn}\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)}$$

for $A = (a_{ij}) \in \mathrm{Mat}_n(R)$. If $\varphi\colon \mathrm{Mat}_n(R) \to R$ is any alternating n-linear function, then

$$\varphi(A) = (\det A)\varphi(I_n).$$

We use the notation det for the unique determinant on $\mathrm{Mat}_n(R)$ for all n.

PROOF. Let $e_1, \dots, e_n$ denote the rows of I_n, and denote the rows of a matrix $A = (a_{ij}) \in \mathrm{Mat}_n(R)$ by $a_1, \dots, a_n$. Then $a_i = \sum_{j=1}^n a_{ij} e_j$, so

$$\varphi(A) = \sum_{k_1,\dots,k_n} a_{1k_1} \cdots a_{nk_n} \varphi(e_{k_1}, \dots, e_{k_n}),$$

where the sum is taken over all $k_i = 1, \dots, n$. Since φ is alternating we have $\varphi(e_{k_1}, \dots, e_{k_n}) = 0$ if two of the indices $k_1, \dots, k_n$ are equal. Thus it suffices to sum over those sequences $(k_1, \dots, k_n)$ that are permutations of $(1, \dots, n)$, and so

$$\varphi(A) = \sum_{\sigma \in S_n} a_{1\sigma(1)} \cdots a_{n\sigma(n)} \varphi(e_{\sigma(1)}, \dots, e_{\sigma(n)}).$$

Next notice that, since φ is also skew-symmetric by Lemma 4.1(i), we have $\varphi(e_{\sigma(1)}, \dots, e_{\sigma(n)}) = (-1)^m \varphi(e_1, \dots, e_n)$, where m is the number of transpositions of $(1, \dots, n)$ it takes to obtain the permutation $(\sigma(1), \dots, \sigma(n))$. But then $(-1)^m$ is just the sign of σ, so

$$\varphi(A) = \sum_{\sigma \in S_n} (\operatorname{sgn}\sigma) a_{1\sigma(1)} \cdots a_{n\sigma(n)} \varphi(I_n).$$

Finally, if φ is a determinant function, then $\varphi(I_n) = 1$, so we must have $\varphi = \det$. The rest of the theorem follows directly from this. □

4.3. Properties of determinants

Theorem 4.6

Let $A, B \in \mathrm{Mat}_n(R)$. Then

$$\det AB = (\det A)(\det B).$$

In particular, $\det\colon \mathrm{GL}_n(R) \to R^*$ *is a group homomorphism.*

Proof. The map $\varphi\colon \mathrm{Mat}_n(R) \to R$ given by $\varphi(A) = \det AB$ is clearly n-linear and alternating. Hence $\varphi(A) = (\det A)\varphi(I)$, and $\varphi(I) = \det B$.

Furthermore, if A is invertible, then $1 = \det I = (\det A)(\det A^{-1})$. Thus $\det A \in R^*$, so det is a group homomorphism as claimed. □

Corollary 4.7

If $A, B \in \mathrm{Mat}_n(\mathbb{F})$ are similar matrices, then $\det A = \det B$.

Proof. Let $P \in \mathrm{Mat}_n(\mathbb{F})$ be such that $A = PBP^{-1}$. Theorem 4.6 then implies that

$$\det A = (\det P)(\det B)(\det P^{-1}) = (\det B)(\det PP^{-1}) = \det B. \qquad \square$$

Corollary 4.7 allows us to define the determinant of a general linear operator $T\colon V \to V$ on a finite-dimensional $\mathbb{F}$-vector space. If $\mathcal{V}$ and $\mathcal{W}$ are bases for V, then the matrix representations ${}_{\mathcal{V}}[T]_{\mathcal{V}}$ and ${}_{\mathcal{W}}[T]_{\mathcal{W}}$ are similar. This allows us to define the determinant $\det T$ of T as the matrix representation ${}_{\mathcal{V}}[T]_{\mathcal{V}}$ for any basis $\mathcal{V}$.

Proposition 4.8

Let $A_{11}, \ldots, A_{nn}$ be square matrices with entries in R and consider the block matrix

$$M = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ 0 & A_{22} & \cdots & A_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & A_{nn} \end{pmatrix},$$

where the remaining A_{ij} are matrices of appropriate dimensions. Then $\det M = \prod_{i=1}^n \det A_{ii}$.

Proof. By induction it suffices to consider the case where M has the block form

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

where $A \in \mathrm{Mat}_r(R)$, $B \in \mathrm{Mat}_s(R)$ and $C \in \mathrm{Mat}_{r,s}(R)$ for appropriate integers r, s. Notice that if we define the matrices

$$M_1 = \begin{pmatrix} I_r & 0 \\ 0 & B \end{pmatrix} \quad \text{and} \quad M_2 = \begin{pmatrix} A & C \\ 0 & I_s \end{pmatrix},$$

then $M = M_1 M_2$. But using Theorem 4.3 we easily see that $\det M_1 = \det B$ and $\det M_2 = \det A$, so it follows that

$$\det M = (\det M_1)(\det M_2) = (\det A)(\det B)$$

as desired. □

PROPOSITION 4.9

Let $A \in \mathrm{Mat}_n(R)$. *Then* $\det A = \det A^\top$.

PROOF. Writing $A = (a_{ij})$, first notice that

$$\det A^\top = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma^{-1}) a_{\sigma(1)1} \cdots a_{\sigma(n)n},$$

since $\operatorname{sgn} \sigma = \operatorname{sgn} \sigma^{-1}$. Next notice that, if $j = \sigma(i)$, then $a_{\sigma(i)i} = a_{j\sigma^{-1}(j)}$. Since R is commutative, it follows that

$$\det A^\top = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma^{-1}) a_{1\sigma^{-1}(1)} \cdots a_{n\sigma^{-1}(n)},$$

and since $\sigma \mapsto \sigma^{-1}$ is a bijection on S_n, it follows that $\det A^\top = \det A$ as desired. □

Let $A \in \mathrm{Mat}_n(R)$. For $1 \leq i, j \leq n$, the (i,j)*-th cofactor* of A is the number $A_{i,j} = (-1)^{i+j} \det M(A)_{i,j}$, where we recall that $M(A)_{i,j}$ is the (i,j)-th minor of A. The *cofactor matrix* of A is the matrix $\operatorname{cof} A \in \mathrm{Mat}_n(R)$ whose (i,j)-th entry is the cofactor $A_{i,j}$. Note that

$$(A^\top)_{i,j} = (-1)^{i+j} \det M(A^\top)_{i,j} = (-1)^{j+i} \det M(A)_{j,i} = A_{j,i},$$

so $\operatorname{cof} A^\top = (\operatorname{cof} A)^\top$. Of greater importance than the cofactor matrix is the *adjoint matrix* of A, written $\operatorname{adj} A$, which is just the transpose of $\operatorname{cof} A$. That is, the (i,j)-th entry of $\operatorname{adj} A$ is the cofactor $A_{j,i}$. Similar to the cofactor matrix we have

$$\operatorname{adj} A^\top = (\operatorname{cof} A^\top)^\top = \operatorname{cof} A = (\operatorname{adj} A)^\top.$$

We have the following:

PROPOSITION 4.10

Let $A \in \mathrm{Mat}_n(R)$. *Then*

$$(\operatorname{adj} A)A = (\det A)I = A(\operatorname{adj} A).$$

PROOF. Writing $A = (a_{ij})$ and fixing some $j \in \{1, \dots, n\}$, Theorem 4.3 implies that

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det M(A)_{i,j} = \sum_{i=1}^{n} a_{ij} A_{i,j},$$

which is just the (j, j)-th entry in the product $(\operatorname{adj} A)A$.

Next we claim that if $k \neq j$, then $\sum_{i=1}^{n} a_{ik} A_{i,j} = 0$. Let $B = (b_{ij}) \in \mathrm{Mat}_n(R)$ be the matrix obtained from A by replacing the jth column of A by its kth column. Then B has two equal columns, so $\det B = 0$. Also, $b_{ij} = a_{ik}$ and $M(B)_{i,j} = M(A)_{i,j}$, so it follows that

$$\begin{aligned} 0 = \det B &= \sum_{i=1}^{n} (-1)^{i+j} b_{ij} \det M(B)_{i,j} \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ik} \det M(A)_{i,j} = \sum_{i=1}^{n} a_{ik} A_{i,j}. \end{aligned}$$

That is, the (j, k)-th entry of the product $(\operatorname{adj} A)A$ is zero, so the off-diagonal entries of $(\operatorname{adj} A)A$ are zero. In total we thus have $(\operatorname{adj} A)A = (\det A)I$.

Finally we prove the equality $A(\operatorname{adj} A) = (\det A)I$, Applying the first equality to $A^\top$ yields

$$(\operatorname{adj} A^\top)A^\top = (\det A^\top)I = (\det A)I,$$

and transposing we get

$$A(\operatorname{adj} A) = A(\operatorname{adj} A^\top)^\top = (\det A)I$$

as desired. □

COROLLARY 4.11

Let $A \in \mathrm{Mat}_n(R)$. *The following are equivalent:*

(i) *A is a (two-sided) unit in* $\mathrm{Mat}_n(R)$.

(ii) *A is a left- or right-unit in* $\mathrm{Mat}_n(R)$.

(iii) $\det A$ *is a unit in* R.

PROOF. If A is e.g. a left-unit, then Theorem 4.6 implies that

$$1 = \det I_n = (\det A)(\det A^{-1}),$$

so $\det A$ is a unit in R. Conversely, if $\det A$ is a unit then Proposition 4.10 implies that $(\det A)^{-1}(\operatorname{adj} A)$ is a two-sided inverse of A. □

Notice that this gives us a second proof of the fact that a matrix is invertible just when it has either a left- or right-inverse. In fact, we see that this holds for matrices with entries in any commutative ring.

4.4. Determinants and eigenvalues

Let V be a vector space of dimension $n < \infty$. If $T \in \mathcal{L}(V)$, then recall that an *eigenvalue* of T is an element $\lambda \in \mathbb{F}$ such that there is a nonzero vector $v \in V$ with $Tv = \lambda v$. The set of eigenvalues of T is called the *spectrum* of T and is denoted $\operatorname{Spec} T$. Clearly $\lambda \in \operatorname{Spec} T$ if and only if $\lambda I - T$ is not injective, i.e. if $\det(\lambda I - T) = 0$. This motivates the definition of the *characteristic polynomial* $p_T(t) \in \mathbb{F}[t]$ of T, given by $p_T(t) = \det(tI - T)$. The eigenvalues of T are then precisely the roots of $p_T(t)$.

PROPOSITION 4.12

Let $T \in \mathcal{L}(V)$.

(i) $p_T(t)$ *is a monic polynomial of degree* n.

(ii) *The constant term of* $p_T(t)$ *equals* $(-1)^n \det T$.

(iii) *The coefficient of* t^{n-1} *in* $p_T(t)$ *equals* $-\operatorname{tr} T$.

Assume further that $p_T(t)$ *splits over* $\mathbb{F}$. *Then:*

(iv) T *has an eigenvalue.*

(v) $\det T$ *is the product of the eigenvalues of* T.

(vi) $\operatorname{tr} T$ *is the sum of the eigenvalues of* T.

The condition that $p_T(t)$ splits over $\mathbb{F}$ means that $p_T(t)$ decomposes into a product of linear factors on the form $t - a \in \mathbb{F}[t]$ (up to multiplication by a constant). This is in particular the case if $\mathbb{F}$ is algebraically closed.

PROOF. *(i)*: Let $A = (a_{ij}) \in \operatorname{Mat}_n(\mathbb{F})$ be a matrix representation of T. The (i,j)-th entry of $tI - A$ is then $t\delta_{ij} - a_{ij}$, so

$$\det(tI - T) = \sum_{\sigma \in S_n} (\operatorname{sgn} \sigma)(t\delta_{1\sigma(1)} - a_{1\sigma(1)}) \cdots (t\delta_{n\sigma(n)} - a_{n\sigma(n)}) \tag{4.1}$$

by Theorem 4.5. Thus $p_T(t)$ is a polynomial in t. Furthermore, the only entries in $tI - A$ containing t are the diagonal entries, and the largest number of such entries occurring in a single term of (4.1) is n, so $\deg p_T(t) \leq n$. But notice that there is only one term in which t appears n times, namely the term corresponding to the identity permutation in S_n, giving the product of the diagonal entries in $tI - A$. This term equals

$$(t - a_{11})(t - a_{22}) \cdots (t - a_{nn}), \tag{4.2}$$

and multiplying out we see that the only resulting term containing t^n is t^n itself. Hence $p_T(t)$ is monic and of degree n. Thus we may write $p_T(t) = \sum_{i=0}^{n} c_i t^i$ for appropriate $c_0, \ldots, c_n \in \mathbb{F}$.

(ii): Simply notice that

$$(-1)^n \det T = \det(-T) = p_T(0) = c_0$$

by n-linearity of det and the definition of $p_T(t)$.

(iii): The only way for one of the terms in (4.1) to contain the factor t^{n-1} is for at least $n-1$ of the b_{ij} to be a diagonal element. But in choosing $n-1$ elements along the diagonal we are forced to also choose the final diagonal element, since otherwise σ would not be a permutation. Hence the factor t^n can only appear in the product (4.2). It is then clear that

$$c_{n-1} = -(a_{11} + \cdots + a_{nn}) = -\operatorname{tr} T$$

as claimed.

(iv): Now assume that $p_T(t)$ splits over $\mathbb{F}$. Then some linear factor $t - \lambda \in \mathbb{F}[t]$ divides $p_T(t)$, which implies that $\lambda \in \mathbb{F}$ is an eigenvalue of T.

(v): Since $p_T(t)$ is monic we have

$$p_T(t) = (t - \lambda_1)(t - \lambda_2) \cdots (t - \lambda_n)$$

for appropriate $\lambda_1, \ldots, \lambda_n \in \mathbb{F}$. These are then the (not necessarily distinct) eigenvalues of T. Thus $p_T(0) = (-1)^n \lambda_1 \cdots \lambda_n$, and the claim follows from (ii).

(vi): We similarly find that $c_{n-1} = -(\lambda_1 + \cdots + \lambda_n)$, so the final claim follows from (iii). □

4.5. Proofs without determinants

Existence of eigenvalues

Assume that $\mathbb{F}$ is algebraically closed, and consider $T \in \mathcal{L}(V)$. For $d \in \mathbb{N}$, let $\mathbb{F}[t]_d$ denote the vector space of polynomials in $\mathbb{F}[t]$ with degree strictly less than d, such that $\dim \mathbb{F}[t]_d = d$. Consider the map $\mathrm{ev}_T : \mathbb{F}[t]_{n^2+1} \to \mathcal{L}(V)$ given by $\mathrm{ev}_T(p) = p(T)$. This cannot be injective, so there is some nonzero $p(t) \in \mathbb{F}[t]_{n^2+1}$ such that $p(T) = 0$. Note that $p(t)$ cannot be constant.

Since $\mathbb{F}$ is algebraically closed, there exist $c, \lambda_1, \ldots, \lambda_m \in \mathbb{F}$ such that $p(t) = c\prod_{i=1}^m (t - \lambda_i)$. But then

$$0 = p(T) = c\prod_{i=1}^m (T - \lambda_i I),$$

so at least one $T - \lambda_i I$ is not injective. Hence λ_i is an eigenvalue of T.

Trace is sum of eigenvalues

Corollary 4.13

Let $\mathbb{F}$ be algebraically closed, and let $T \in \mathcal{L}(V)$. Then the sum of the eigenvalues of T is $\operatorname{tr} T$.

Proof. Let $A \in \mathrm{Mat}_n(\mathbb{F})$ be an upper triangular matrix [TODO reference to later, perhaps move things around.] for T. The diagonal elements of A are the eigenvalues, and the trace of T is just the sum of these elements. □

4.6. Cross products

Definition 4.14: *Cross products*

Let $v = (\alpha_1, \alpha_2, \alpha_3)$ and $w = (\beta_1, \beta_2, \beta_3)$ be vectors in $\mathbb{R}^3$. The *cross product* of v and w is the vector

$$v \times w = \begin{pmatrix} \alpha_2\beta_3 - \alpha_3\beta_2 \\ \alpha_3\beta_1 - \alpha_1\beta_3 \\ \alpha_1\beta_2 - \alpha_2\beta_1 \end{pmatrix}.$$

Denote the standard basis on $\mathbb{R}^3$ by $\mathcal{E} = (e_1, e_2, e_3)$. We easily see that $e_i \times e_j = e_k$ when (i, j, k) is a cyclic permutation of $(1, 2, 3)$.

Lemma 4.15

Let $v, w, u \in \mathbb{R}^3$. Then

$$\langle u, v \times w\rangle = \det(u, v, w).$$

Proof. By multilinearity of the inner product and of determinants, it suffices to prove the lemma when u is a basis vector. But it is clear that

$$\langle e_i, v \times w\rangle = \det(e_i, v, w),$$

as desired. □

The product $\langle u, v\times w\rangle$ is called the *(scalar) triple product* of u, v and w, and is denoted $[u,v,w]$. We call it the *scalar* triple product to distinguish it from the *vector* triple product $u\times(v\times w)$, whose properties we will examine in Corollary 4.18. The scalar triple product has some very nice properties summarised in the following proposition:

PROPOSITION 4.16

Let $u,v,w\in\mathbb{R}^3$.

(i) *The cross product map $(v,w)\mapsto v\times w$ is bilinear.*

(ii) $v\times w=-w\times v$.

(iii) *The triple product $[u,v,w]$ is invariant under cyclic permutations, i.e.*

$$[u,v,w]=[v,w,u]=[w,u,v]$$

and invariant under interchange of inner product and cross product, i.e.

$$\langle u, v\times w\rangle=[u,v,w]=\langle u\times v, w\rangle.$$

(iv) *$v\times w=0$ if and only if v and w are linearly dependent.*

(v) *$v\times w$ is orthogonal to both v and w.*

PROOF. The first three claims follow from Lemma 4.15 since the determinant is multilinear and alternating (hence skew-symmetric).

For the fourth claim, if v and w are linearly dependent then $\det(u,v,w)=0$ for all $u\in\mathbb{R}^3$, so $v\times w=0$. Conversely, if v and w are linearly independent, then extending to a basis (u,v,w) for $\mathbb{R}^3$ we have $\det(u,v,w)\neq 0$, implying that $v\times w\neq 0$.

To prove the final claim, notice that

$$\langle v, v\times w\rangle=\det(v,v,w)=0,$$

and similarly for w. □

PROPOSITION 4.17

Let $a,b,v,w\in\mathbb{R}^3$. Then

$$\langle a\times b, v\times w\rangle=\det\begin{pmatrix}\langle a,v\rangle & \langle b,v\rangle\\ \langle a,w\rangle & \langle b,w\rangle\end{pmatrix}.$$

In particular,

$$\|v \times w\|^2 = \det\begin{pmatrix} \|v\|^2 & \langle v, w\rangle \\ \langle v, w\rangle & \|w\|^2 \end{pmatrix}.$$

The latter identity is just Lagrange's identity in three dimensions. If θ is the angle between v and w, then $\langle v, w\rangle = \|v\|\|w\|\cos\theta$, so

$$\|v \times w\|^2 = \|v\|^2\|w\|^2 - \langle v, w\rangle^2 = \|v\|^2\|w\|^2(1 - \cos^2\theta) = \|v\|^2\|w\|^2\sin^2\theta.$$

Hence $\|v \times w\| = \|v\|\|w\||\sin\theta|$, which is the area of the parallelogram spanned by v and w. If $u \in \mathbb{R}^3$ is another vector and φ is the angle between u and the normal of the plane spanned by v and w (e.g. $v \times w$), then

$$\big|[u, v, w]\big| = |\langle u, v \times w\rangle| = \|u\|\|v \times w\||\cos\varphi| = \|u\|\|v\|\|w\||\sin\theta\cos\varphi|.$$

But this is the volume of the parallelepiped spanned by u, v and w. This gives a geometric interpretation (or 'proof') of the invariance of the scalar triple product.

Proof. By linearity it suffices to prove the identity when the four vectors are basis vectors. If $a = b$ or $v = w$ then both sides are zero, so we may assume that $a = e_i$, $b = e_j$, $v = e_k$ and $v = e_l$ with $i \neq j$ and $k \neq l$. By potentially swapping a and b and/or v and w we may assume that $e_i \times e_j = e_p$ and $e_k \times e_l = e_q$ for some $p, q \in \{1, 2, 3\}$.

If $p = q$ then $i = k$ and $j = l$, so both sides equal 1. If instead $p \neq q$, then the two cross products on the left-hand side are orthogonal, so the inner product is zero. Furthermore, either k or l equals p, so one of the rows in the right-hand side matrix is zero, and hence the determinant is zero. □

Corollary 4.18

Let $u, v, w \in \mathbb{R}^3$. Then

$$u \times (v \times w) = v\langle u, w\rangle - w\langle u, v\rangle. \tag{4.3}$$

In particular, the cross product satisfies the Jacobi identity

$$u \times (v \times w) + v \times (w \times u) + w \times (u \times v) = 0. \tag{4.4}$$

The identity (4.3) is sometimes called the 'bac-cab rule', a name that would have been self-explanatory had we used the names a, b and c instead of u, v and w. Note that to conform to this rule we need to write the vectors before the scalars.

PROOF. For $x \in \mathbb{R}^3$ we have

$$\begin{aligned}\langle x, u \times (v \times w)\rangle &= [x, u, v \times w] \\ &= \langle x \times u, v \times w\rangle \\ &= \det\begin{pmatrix} \langle x, v\rangle & \langle u, v\rangle \\ \langle x, w\rangle & \langle u, w\rangle \end{pmatrix} \\ &= \langle x, v\rangle\langle u, w\rangle - \langle u, v\rangle\langle x, w\rangle \\ &= \big\langle x, v\langle u, w\rangle - w\langle u, v\rangle\big\rangle.\end{aligned}$$

The claim then follows since x was arbitrary. □

LEMMA 4.19

Let $A \in \mathrm{Mat}_d(\mathbb{R})$. Every neighbourhood of A contains an invertible matrix different from A. In particular, there exists a sequence $(A_n)_{n\in\mathbb{N}}$ of invertible matrices different from A such that $A_n \to A$ for $n \to \infty$.

Since $\mathrm{Mat}_d(\mathbb{R})$ is a finite-dimensional vector space, it has a unique vector space topology. More concretely, all norms on $\mathrm{Mat}_d(\mathbb{R})$ are Lipschitz equivalent, so we may choose whatever norm we wish. We choose the Euclidean norm, identifying $\mathrm{Mat}_d(\mathbb{R})$ with $\mathbb{R}^{d^2}$.

PROOF. Let $t \in \mathbb{R} \setminus \{0\}$. Then $A - tI$ is invertible if and only if $\det(A - tI) = 0$, but $\det(A - tI)$ is a polynomial in t, so it has finitely many roots. Hence the nonzero roots of $\det(A - tI)$ are bounded away from zero, so since $A - tI \to A$ as $t \to 0$, the claim follows. □

PROPOSITION 4.20: *Transformation of cross products*

Let $u, v, w \in \mathbb{R}^3$, and let $A \in \mathrm{Mat}_3(\mathbb{R})$. Then we have the following:

(i) $[Au, Av, Aw] = (\det A)[u, v, w]$.

(ii) $Av \times Aw = (\operatorname{cof} A)(v \times w) = (\operatorname{adj} A)^\top (v \times w)$.

(iii) *If A is orthogonal, then* $A(v \times w) = (\det A)(Av \times Aw)$.

This gives a geometric interpretation of the determinant. If $[u, v, w]$ is the signed volume of the parallelepiped spanned by u, v and w, and $[Au, Av, Aw]$ is the signed volume of the parallelepiped spanned by Au, Av and Aw, then $\det A$ is the factor by which this volume increasing when applying A to each of u, v and w. In particular, this explains why the determinant of A is zero if and only if A is singular: This means that A sends a basis of $\mathbb{R}^3$ to a linearly dependent set, and the parallelepiped spanned by such a set has zero volume.

PROOF. *Proof of (i)*: Simply notice that

$$[Au, Av, Aw] = \det(Au, Av, Aw) = (\det A)\det(u, v, w) = (\det A)\langle u, v \times w\rangle,$$

where the second equality follows since $\det(Au, Av, Aw)$ is also the determinant of the matrix

$$\big(Au \mid Av \mid Aw\big) = A\big(u \mid v \mid w\big),$$

and the determinant is multiplicative.

Proof of (ii): First assume that A is invertible. Then replacing u with $A^{-1}u$ in (i) we obtain

$$\begin{aligned}\langle u, Av \times Aw\rangle &= (\det A)\langle A^{-1}u, v \times w\rangle \\ &= (\det A)\langle u, (A^{-1})^{\top}(v \times w)\rangle \\ &= \langle u, (\operatorname{cof} A)(v \times w)\rangle,\end{aligned}$$

where the last equality follows from Proposition 4.10. Hence we obtain the desired identity when A is invertible. Finally notice that both the maps $A \mapsto \operatorname{cof} A$ and $A \mapsto Av \times Aw$ are continuous. Hence the claim for general A follows from Lemma 4.19.

Proof of (iii): Notice that $A^{-1} = A^{\top}$, so this follows immediately from (ii). □

If A is a proper rotation, i.e. if A is orthogonal and $\det A = 1$, then Proposition 4.20(iii) implies that $A(v \times w) = Av \times Av$. This allows us to define a cross product on any three-dimensional inner product space, when this is equipped with an orientation.

First, if $\mathcal{V}$ and $\mathcal{W}$ are ordered bases for any finite-dimensional real vector space V, then we say that $\mathcal{V}$ and $\mathcal{W}$ have the *same orientation* if the change of basis operator $\varphi_{\mathcal{W},\mathcal{V}}$ has positive determinant. It follows that orientation partitions the set of ordered bases for V into two *orientation classes*, each called an *orientation* of V. If V is equipped with an orientation $\mathcal{O}$, then we call this class the *positive orientation* of V, and the other class the *negative orientation* of V. An ordered basis for V is called *positive* if it lies in $\mathcal{O}$ and *negative* if it does not.

Returning to the case where V is three-dimensional and equipped with an orientation, let $\mathcal{V}$ and $\mathcal{W}$ be positive ordered orthonormal bases for V. For vectors $v, w \in V$ we can then consider the cross products of their coordinate vectors, i.e.

$$[v]_{\mathcal{V}} \times [w]_{\mathcal{V}} \quad \text{and} \quad [v]_{\mathcal{W}} \times [w]_{\mathcal{W}}.$$

Since ${}_{\mathcal{W}}[\square]_{\mathcal{V}}$ is orthogonal with determinant 1, we have

$${}_{\mathcal{W}}[\square]_{\mathcal{V}}([v]_{\mathcal{V}} \times [w]_{\mathcal{V}}) = {}_{\mathcal{W}}[\square]_{\mathcal{V}} \cdot [v]_{\mathcal{V}} \times {}_{\mathcal{W}}[\square]_{\mathcal{V}} \cdot [w]_{\mathcal{V}} = [v]_{\mathcal{W}} \times [w]_{\mathcal{W}}.$$

Hence we have

$$\varphi_{\mathcal{V}}^{-1}([v]_{\mathcal{V}} \times [w]_{\mathcal{V}}) = \varphi_{\mathcal{W}}^{-1}([v]_{\mathcal{W}} \times [w]_{\mathcal{W}}),$$

so we may define the cross product of v and w as $v \times w = \varphi_{\mathcal{V}}^{-1}([v]_{\mathcal{V}} \times [w]_{\mathcal{V}})$ where $\mathcal{V}$ is any positive ordered orthonormal basis for V. Notice that this means that $[v \times w]_{\mathcal{V}} = [v]_{\mathcal{V}} \times [w]_{\mathcal{V}}$.

This allows us to generalise most of the above results to general vector spaces. For instance, using that the coordinate map $\varphi_{\mathcal{V}}$ is an isometry, the scalar triple product of $u, v, w \in V$ is given by

$$[u, v, w] = \langle u, v \times w \rangle = \langle [u]_{\mathcal{V}}, [v \times w]_{\mathcal{V}} \rangle = \langle [u]_{\mathcal{V}}, [v]_{\mathcal{V}} \times [w]_{\mathcal{V}} \rangle = \Big[[u]_{\mathcal{V}}, [v]_{\mathcal{V}}, [w]_{\mathcal{V}}\Big],$$

and hence it has all the properties of the scalar triple product on $\mathbb{R}^3$, such as invariance under cyclic permutations. Notice also that it is indeed a *scalar* quantity, in the sense that it is invariant under a change of basis. Similarly, the 'bac-cab rule' (4.3) becomes

$$\begin{aligned}
[u \times (v \times w)]_{\mathcal{V}} &= [u]_{\mathcal{V}} \times [v \times w]_{\mathcal{V}} \\
&= [u]_{\mathcal{V}} \times ([v]_{\mathcal{V}} \times [w]_{\mathcal{V}}) \\
&= [v]_{\mathcal{V}} \langle [u]_{\mathcal{V}}, [w]_{\mathcal{V}} \rangle - [w]_{\mathcal{V}} \langle [u]_{\mathcal{V}}, [v]_{\mathcal{V}} \rangle \\
&= [v]_{\mathcal{V}} \langle u, w \rangle - [w]_{\mathcal{V}} \langle u, v \rangle \\
&= [v \langle u, w \rangle - w \langle u, v \rangle]_{\mathcal{V}}.
\end{aligned}$$

Hence $u \times (v \times w) = v\langle u, w \rangle - w \langle u, v \rangle$ since $\varphi_{\mathcal{V}}$ is an isomorphism. In particular, the cross product on V also satisfies the Jacobi identity (4.4), so V becomes a Lie algebra whose Lie bracket is given by the cross product, i.e. $[v, w] = v \times w$.

5 • Complexification

If W is a complex vector space, then we may restrict the scalar multiplication $\mathbb{C} \times W \to W$ to a map $\mathbb{R} \times W \to W$. When we equip W with this restricted scalar multiplication instead of the original one, we call the resulting space the *real version of* W and denote it by $W_{\mathbb{R}}$.

Conversely, if V is a real vector space then we define the *complexification of* V as the vector space $V^{\mathbb{C}}$ whose underlying set is $V \times V$, and which is equipped with componentwise addition and the complex scalar multiplication

$$(\alpha + \mathrm{i}\beta)(v, u) = (\alpha v - \beta u, \alpha u + \beta v),$$

for $\alpha, \beta \in \mathbb{R}$ and $v, u \in V$. Notice that the map $v \mapsto (v, 0)$ is injective (and real linear), that $(v, 0) + (w, 0) = (v + w, 0)$, and that $\alpha(v, 0) = (\alpha v, 0)$ for $\alpha \in \mathbb{R}$, so $V^{\mathbb{C}}$ contains an isomorphic copy of V, and we may identify elements $v \in V$ with elements $(v, 0) \in V^{\mathbb{C}}$. Furthermore, notice that $(v, u) = (v, 0) + \mathrm{i}(u, 0)$, so by the above identification we may write $(v, u) = v + \mathrm{i}\,u$.

We briefly study the relationship between a real vector space and its complexification.

PROPOSITION 5.1

If $\mathcal{B}$ is a basis for V, then $\mathcal{B}^{\mathbb{C}} = \{b + \mathrm{i}\,0 \mid b \in \mathcal{B}\}$ is a basis for $V^{\mathbb{C}}$. In particular, $\dim_{\mathbb{R}} V = \dim_{\mathbb{C}} V^{\mathbb{C}}$.

PROOF. Let $v + \mathrm{i}\,u \in V^{\mathbb{C}}$. Then there are real numbers α_b and β_b (finitely many nonzero) such that $v = \sum_{b \in \mathcal{B}} \alpha_b b$ and $u = \sum_{b \in \mathcal{B}} \beta_b b$. But then

$$v + \mathrm{i}\,u = \sum_{b \in \mathcal{B}} \alpha_b b + \mathrm{i} \sum_{b \in \mathcal{B}} \beta_b b = \sum_{b \in \mathcal{B}} (\alpha_b + \mathrm{i}\,\beta_b) b = \sum_{b \in \mathcal{B}} (\alpha_b + \mathrm{i}\,\beta_b)(b + \mathrm{i}\,0),$$

so $\mathcal{B}^{\mathbb{C}}$ spans $V^{\mathbb{C}}$. Furthermore, if $v + \mathrm{i}\,u = 0$, then the previous computation shows that $\sum_{b \in \mathcal{B}} \alpha_b b = 0 = \sum_{b \in \mathcal{B}} \beta_b b$. Linear independence of $\mathcal{B}$ then implies that $\alpha_b = \beta_b = 0$ for all $b \in \mathcal{B}$. $\square$

EXAMPLE 5.2. Notice that $(\mathbb{R}^n)^{\mathbb{C}} \cong \mathbb{C}^n$. The above proposition then implies that the standard basis for $\mathbb{R}^n$ gives rise to a basis for $\mathbb{C}^n$, and we notice that this is precisely the standard basis. ⌟

We now show how to extend linear maps defined between real vector spaces to the complexifications of those spaces. If $T \colon V \to W$ is a linear map between real vector spaces, then we define the complexification of T by

$$\begin{aligned} T^{\mathbb{C}} \colon V^{\mathbb{C}} &\to W^{\mathbb{C}}, \\ v + \mathrm{i}\,u &\mapsto Tv + \mathrm{i}\,Tu. \end{aligned}$$

That is, $T^{\mathbb{C}}$ is just the product map $T \times T$. This is easily seen to be complex-linear.

PROPOSITION 5.3

Let V be a real vector space, and let $T \in \mathcal{L}(V)$. If $\lambda \in \mathbb{R}$ is an eigenvalue of the complexification $T^{\mathbb{C}}$ of T, then λ is also an eigenvalue of T. Furthermore, if $v + \mathrm{i}\,u \in E_{T^{\mathbb{C}}}(\lambda)$ then $v, u \in E_T(\lambda)$.

Note that this does not mean that v and u are eigenvectors of T since they might be zero. But if $v + \mathrm{i}\,u$ is an eigenvector of $T^{\mathbb{C}}$, then at least one of v and u is nonzero and hence an eigenvector of T.

PROOF. Let $v + \mathrm{i}\, u \in V^{\mathbb{C}}$ be an eigenvector of $T^{\mathbb{C}}$ corresponding to λ. Then

$$Tv + \mathrm{i}\, Tu = T^{\mathbb{C}}(v + \mathrm{i}\, u) = \lambda(v + \mathrm{i}\, u) = \lambda v + \mathrm{i}\, \lambda u.$$

It follows that $Tv = \lambda v$ and $Tu = \lambda u$ as desired. □

If V is finite-dimensional and $\mathcal{V}$ is an ordered basis for V, then $\mathcal{V}^{\mathbb{C}}$ carries the obvious ordering. Since V and $V^{\mathbb{C}}$ have the same dimension, the following result is not surprising:

PROPOSITION 5.4

Let V and W be a finite-dimensional real vector spaces, and consider $T\colon V \to W$. If $\mathcal{V} = (v_1, \dots, v_n)$ and $\mathcal{W}$ are ordered bases of V and W respectively, then

$${}_{\mathcal{W}^{\mathbb{C}}}[T^{\mathbb{C}}]_{\mathcal{V}^{\mathbb{C}}} = {}_{\mathcal{W}}[T]_{\mathcal{V}}.$$

PROOF. By Proposition 3.4(i), the ith column of ${}_{\mathcal{W}^{\mathbb{C}}}[T^{\mathbb{C}}]_{\mathcal{V}^{\mathbb{C}}}$ is given by

$$[T^{\mathbb{C}}(v_i + \mathrm{i}\, 0)]_{\mathcal{W}^{\mathbb{C}}} = [Tv_i + \mathrm{i}\, 0]_{\mathcal{W}^{\mathbb{C}}} = [Tv_i]_{\mathcal{W}},$$

that is, the ith column of ${}_{\mathcal{W}}[T]_{\mathcal{V}}$, which proves the claim. □

Finally, if V is a real normed space, then we define a norm on $V^{\mathbb{C}}$ by the equation

$$\|v + \mathrm{i}\, u\|^2 = \|v\|^2 + \|u\|^2. \tag{5.1}$$

Furthermore, if V is an inner product space, then we define an inner product on $V^{\mathbb{C}}$ by

$$\langle v + \mathrm{i}\, u, x + \mathrm{i}\, y\rangle = \langle v, x\rangle + \langle u, y\rangle + \mathrm{i}(\langle u, x\rangle - \langle v, y\rangle). \tag{5.2}$$

The norm induced by this inner product agrees with the norm defined by (5.1). Notice that the identity (5.2) holds in any *complex* inner product space, where the notation $v + \mathrm{i}\, u$ instead means the sum of v and the scalar product of i and u (in justifying this claim, the reader will recall that the inner product on a complex space is *sesqui*linear).

6 • Operator adjoints

Let $\mathcal{C}$ be a locally small category, and let $f\colon A \to B$ be an arrow in $\mathcal{C}$. For every object C, precomposition with f then induces an arrow

$$\begin{aligned} \mathrm{Hom}_{\mathcal{C}}(f, C)\colon \mathrm{Hom}_{\mathcal{C}}(B, C) &\to \mathrm{Hom}_{\mathcal{C}}(A, C), \\ g &\mapsto g \circ f. \end{aligned}$$

This gives rise to a contravariant functor $\mathrm{Hom}_{\mathcal{C}}(-, C)\colon \mathcal{C} \to \mathbf{Set}$. Specialising to the case where $\mathcal{C}$ is the category $\mathbb{F}$-**Vect** and where C is the field $\mathbb{F}$ (considered as a vector space), we obtain the functor $(-)^*$ sending a vector space V to its (algebraic) dual V^*, and a linear map T to its pullback. Since we will use the notation T^* for the Hilbert space adjoint, we instead write $T^\dagger$ for the pullback of T, following Folland (2007). We also call this the *operator adjoint* of T:

Definition 6.1: *Operator adjoints*

Let V and W be $\mathbb{F}$-vector spaces, and let $T\colon V \to W$ be a linear map. The *(operator) adjoint* of T is the pullback

$$\begin{aligned} T^\dagger\colon W^* &\to V^*, \\ \varphi &\mapsto \varphi \circ T. \end{aligned}$$

This already satisfies $\mathrm{id}_V^* = \mathrm{id}_{V^*}$ and $(ST)^\dagger = T^\dagger S^\dagger$ by functoriality, so that in particular $(T^{-1})^\dagger = (T^\dagger)^{-1}$ when T is invertible. Furthermore, it is easy to show that the map $T \mapsto T^\dagger$ is linear. It is also injective, since if $Tv \neq Sv$ then there is a $\varphi \in W^*$ such that $\varphi(Tv) \neq \varphi(Sv)$. If V and W are finite-dimensional, it is therefore a linear isomorphism.

Proposition 6.2

Let $T \in \mathcal{L}(V, W)$.

(i) $\ker T^\dagger = (\operatorname{im} T)^0$.

(ii) $\operatorname{im} T^\dagger = (\ker T)^0$.

Proof. Roman (2008, Theorem 3.19). □

6.1. *Finite-dimensional spaces*

Corollary 6.3

If $T \in \mathcal{L}(V, W)$ with V and W finite-dimensional, then $\operatorname{rank} T^\dagger = \operatorname{rank} T$.

Proof. Recall that the dimension of $(\ker T)^0$ equals the codimension of $\ker T$, which is just $\dim V - \dim\ker T$ when V is finite-dimensional (cf. Roman 2008, Theorem 3.15). We then have

$$\operatorname{rank} T^\dagger = \dim\operatorname{im} T^\dagger = \dim(\ker T)^0 = \dim V - \dim\ker T = \dim\operatorname{im} T = \operatorname{rank} T$$

by Proposition 6.2, as desired. □

Note that if $\mathcal{V} = (v_1, \dots, v_n)$ is an ordered basis for V, $\mathcal{V}^*$ the corresponding dual basis, and $\mathcal{V}^{**}$ the double dual basis, then for $\varphi = \varphi_1 v_1^* + \cdots + \varphi_n v_n^*$ we have

$$v_i^{**}(\varphi) = \varphi_i = \varphi(v_i),$$

since both $v_i^*(v_j) = \delta_{ij}$ and $v_i^{**}(v_j^*) = \delta_{ij}$, by definition of the dual basis.

PROPOSITION 6.4

If $T \in \mathcal{L}(V, W)$ is a linear map between finite-dimensional vector spaces, and $\mathcal{V}$ and $\mathcal{W}$ are ordered bases for V and W respectively, then

$$_{\mathcal{V}^*}[T^\dagger]_{\mathcal{W}^*} = \big(_{\mathcal{W}}[T]_{\mathcal{V}}\big)^\top.$$

PROOF. Write $\mathcal{V} = (v_1, \dots, v_n)$ and $\mathcal{W} = (w_1, \dots, w_m)$. Then

$$\big(_{\mathcal{W}}[T]_{\mathcal{V}}\big)_{ij} = \big([Tv_j]_{\mathcal{W}}\big)_i = w_i^*(Tv_j),$$

and

$$\big(_{\mathcal{V}^*}[T^\dagger]_{\mathcal{W}^*}\big)_{ij} = \big([T^\dagger w_j^*]_{\mathcal{V}^*}\big)_i = v_i^{**}(T^\dagger w_j^*) = T^\dagger w_j^*(v_i) = w_j^*(Tv_i).$$

These expressions are the same, but with i and j switched. □

COROLLARY 6.5

The row rank and the column rank of a matrix $A \in \mathrm{Mat}_{m,n}(\mathbb{F})$ are equal.

PROOF. The matrix representation of the multiplication operator M_A with respect to the standard bases on $\mathbb{F}^n$ and $\mathbb{F}^m$ is just A itself, and Proposition 6.4 then implies that the matrix representation of $(M_A)^\dagger$ with respect to the dual bases is $A^\top$. But the rank of an operator equals the rank of any matrix representation of that operator, so Corollary 6.3 implies that A and $A^\top$ have the same (column) rank. Finally, the column rank of $A^\top$ is the row rank of A, proving the claim. □

6.2. Topological vector spaces

If V and W are topological vector spaces and V^* and W^* denote their respective *continuous* dual spaces, then we may also consider the adjoint $T^\dagger$ of a *continuous* linear map $T: V \to W$. It then turns out that $T^\dagger$ is also continuous:

PROPOSITION 6.6

Let $T: V \to W$ be a continuous linear map, and let $T^\dagger: W^* \to V^*$ be its adjoint.

(i) *$T^\dagger$ is continuous with respect to the weak*-topologies on W^* and V^*.*

(ii) *If V and W are normed $\mathbb{K}$-vector spaces, then $T^\dagger$ is continuous with respect to the operator norms on W^* and V^*, and $\|T^\dagger\| = \|T\|$.*

PROOF. Let $(\varphi_i)_{i\in I}$ be a net in W^* that converges to some $\varphi \in W^*$. That is, $\varphi_i(w) \to \varphi(w)$ for all $w \in W$, so in particular $\varphi_i(Tv) \to \varphi(Tv)$ for all $v \in V$. But then $T^\dagger\varphi_i = \varphi_i \circ T$ converges to $T^\dagger\varphi = \varphi \circ T$, so $T^\dagger$ is continuous as claimed.

If V and W are normed, then

$$\|T^\dagger\varphi\| = \|\varphi \circ T\| \leq \|\varphi\|\|T\|$$

for all $\varphi \in W^*$, implying that $T^\dagger$ is bounded with $\|T^\dagger\| \leq \|T\|$. If $T \neq 0$, then let $v \in V$ with $\|v\| = 1$ such that $Tv \neq 0$. The Hahn–Banach theorem then furnishes a $\varphi \in W^*$ with $\|\varphi\| = 1$ and $\varphi(Tv) = \|Tv\|$ (cf. Folland 2007, Theorem 5.8(b)). It follows that

$$\|T^\dagger\| \geq \|T^\dagger\varphi\| \geq |(T^\dagger\varphi)v| = |\varphi(Tv)| = \|Tv\|.$$

This inequality then holds for all $v \in V$ with $\|v\| = 1$, implying that $\|T^\dagger\| \geq \|T\|$. □

TODO: $T \mapsto T^\dagger$ not necessarily injective now?

7 • Triangularisation

Recall that a matrix $A = (a_{ij}) \in \mathrm{Mat}_n(R)$ is called *upper triangular* if $a_{ij} = 0$ whenever $i > j$. If V is an n-dimensional $\mathbb{F}$-vector space and $\mathcal{V}$ is an ordered basis for V, then we say that the operator $T \in \mathcal{L}(V)$ is *upper triangular with respect to* $\mathcal{V}$ if the matrix representation ${}_{\mathcal{V}}[T]_{\mathcal{V}}$ is upper triangular.

A subspace U of a vector space V is said to be *invariant under* $T \in \mathcal{L}(T)$ if $T(U) \subseteq U$.

PROPOSITION 7.1

Let V be an $\mathbb{F}$-vector space with $n = \dim V < \infty$, and let $\mathcal{V} = (v_1, \dots, v_n)$ be an ordered basis for V. An operator $T \in \mathcal{L}(V)$ is upper triangular with respect to $\mathcal{V}$ if and only if $\mathrm{span}(v_1, \dots, v_i)$ is invariant under T for all $i \in \{1, \dots, n\}$.

PROOF. This is obvious. □

LEMMA 7.2

Let V be an $\mathbb{F}$-vector space, and let $T \in \mathcal{L}(V)$ be an isomorphism. If U is a finite-dimensional subspace of V that is invariant under T, then U is also invariant *under T^{-1}.*

PROOF. Since U is finite-dimensional and $T|_U : U \to U$ is injective, applying the rank–nullity theorem implies that $T|_U$ is also surjective. Hence if $u \in U$, then there exists a $v \in U$ such that $Tv = u$. It follows that

$$T^{-1}u = T^{-1}Tv = v \in U,$$

so U is invariant under T^{-1}. □

PROPOSITION 7.3

Let V be a finite-dimensional $\mathbb{F}$-vector space, and let $\mathcal{V}$ be an ordered basis for V. If $T \in \mathcal{L}(V)$ is an isomorphism that is upper triangular with respect to $\mathcal{V}$, then T^{-1} is also upper triangular with respect to $\mathcal{V}$.

In particular, the subset of $\mathrm{GL}_n(\mathbb{F})$ consisting of upper triangular matrices is a subgroup.

PROOF. This is an obvious consequence of the above two results. □

LEMMA 7.4

Let $A \in \mathrm{Mat}_n(\mathbb{F})$ be upper triangular. Then A is invertible if and only if all its diagonal elements are nonzero.

PROOF. Denote the diagonal elements of A by $\lambda_1, \dots, \lambda_n$, and let $(e_1, \dots, e_n)$ be the standard basis of $\mathbb{F}^n$. First assume that the diagonal elements are nonzero. Then notice that $e_1 \in R(A)$, and that

$$Ae_i = a_1 e_1 + \cdots + a_{i-1} e_{i-1} + \lambda_i e_i$$

for appropriate $a_1, \dots, a_{i-1} \in \mathbb{F}$. By induction we then have $e_i \in R(A)$. Since $(e_1, \dots, e_n)$ is a basis, this implies that $R(A) = \mathbb{F}^n$.

Conversely, assume that some diagonal element λ_i is zero. Then

$$A \operatorname{span}(e_1, \dots, e_i) \subseteq \operatorname{span}(e_1, \dots, e_{i-1}),$$

so the null-space of A is nontrivial, and hence A is singular. □

LEMMA 7.5

Let $A \in \mathrm{Mat}_n(\mathbb{F})$ be upper triangular. Then the eigenvalues of A are its diagonal elements.

PROOF. Let $\lambda \in \mathbb{F}$, and denote the diagonal elements of A by $\lambda_1, \dots, \lambda_n$. By Lemma 7.4, the matrix $\lambda I - A$ is singular if and only if $\lambda - \lambda_i = 0$ for some i, and hence $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A. □

Proposition 7.6

Let $\mathbb{F}$ be algebraically closed, and let V be a finite-dimensional $\mathbb{F}$-vector space. If $T \in \mathcal{L}(V)$, then V has an ordered basis with respect to which T is upper triangular.

Proof. This is obvious if $\dim V = 1$, so assume that $n = \dim V > 1$, and assume that the claim is true for $\mathbb{F}$-vector spaces of dimension $n-1$. Since $\mathbb{F}$ is algebraically closed, T has an eigenvector $v_1 \in V$. Then $U = \operatorname{span}(v_1)$ is invariant under T, so we may define a linear operator[2] $\tilde{T} \in \mathcal{L}(V/U)$ by $\tilde{T}(v+U) = Tv + U$. Since $\dim V/U = n-1$, by induction there is a basis $v_2 + U, \dots, v_n + U$ of V/U with respect to which the matrix of $\tilde{T}$ is upper triangular. It is easy to show that the collection $v_1, \dots, v_n$ is linearly independent, hence a basis for V.

Now notice that

$$Tv_i + U = \tilde{T}(v_i + U) \in \operatorname{span}(v_2 + U, \dots, v_i + U)$$

for $i \in \{2, \dots, n\}$. That is, there exist $a_2, \dots, a_i \in \mathbb{F}$ such that

$$Tv_i + U = (a_2 v_2 + \cdots + a_i v_i) + U.$$

But then $Tv_i \in \operatorname{span}(v_1, \dots, v_i)$ for all $i \in \{2, \dots, n\}$, and since U is invariant under T this also holds for $i = 1$. Hence T is upper triangular with respect to the basis $v_1, \dots, v_n$ of V. ☐

Theorem 7.7: *Schur's Theorem*

Let V be a finite-dimensional complex inner product space. If $T \in \mathcal{L}(V)$, then V has an ordered orthonormal basis with respect to which T is upper triangular.

Proof. By Proposition 7.6 V has an ordered basis $\mathcal{V} = (v_1, \dots, v_n)$ with respect to which ${}_{\mathcal{V}}[T]_{\mathcal{V}}$ is upper triangular. Now apply the Gram–Schmidt procedure to $\mathcal{V}$ and obtain an orthonormal basis $\mathcal{U} = (u_1, \dots, u_n)$ for V such that

$$\operatorname{span}(u_1, \dots, u_i) = \operatorname{span}(v_1, \dots, v_i)$$

for all $i \in \{1, \dots, n\}$. Then Proposition 7.1 shows that ${}_{\mathcal{U}}[T]_{\mathcal{U}}$ is also upper triangular, proving the claim. ☐

8 • Orthonormal diagonalisation

[2] The operator $\tilde{T}$ may arise as follows: Let $\pi \colon V \to V/U$ be the quotient map. Then $U \subseteq \ker(\pi \circ T)$ since U is invariant under T, so $\pi \circ T$ descends to a linear map $\tilde{T} \colon V/U \to V/U$.

8.1. Hilbert space adjoints

If V is a Hilbert space, for $v \in V$ let φ_v denote the element in the continuous dual V^* given by $\varphi_v(u) = \langle u, v\rangle$. Further, let $\Phi_V \colon V \to V^*$ denote the (conjugate-)linear isomorphism $v \mapsto \varphi_v$. Then

$$|\varphi_v(u)| = |\langle u, v\rangle| \leq \|u\| \|v\|,$$

implying that $\|\varphi_v\| \leq \|v\|$. Furthermore,

$$|\varphi_v(v)| = \|v\|^2,$$

so in fact $\|\varphi_v\| = \|v\|$. In other words, Φ_V is an isometry and in particular continuous.

DEFINITION 8.1: *Hilbert space adjoints*

Let V and W be Hilbert spaces, and let $T \in \mathcal{B}(V,W)$. The *(Hilbert space) adjoint* of T is the operator $T^* \colon W \to V$ given by

$$T^* = \Phi_V^{-1} \circ T^\dagger \circ \Phi_W.$$

Properties of the operator adjoint $T^\dagger$ are often inherited by the Hilbert space adjoint: By Proposition 6.6(ii) $T^\dagger$ is continuous, so T^* is also continuous. And Φ_V and Φ_W are *conjugate*-linear, so T^* is linear. Furthermore, if $S \in \mathcal{B}(W,U)$ then

$$\begin{aligned}
(ST)^* &= \Phi_V^{-1} \circ (ST)^\dagger \circ \Phi_U \\
&= \Phi_V^{-1} \circ T^\dagger \circ S^\dagger \circ \Phi_U \\
&= (\Phi_V^{-1} \circ T^\dagger \circ \Phi_W) \circ (\Phi_W^{-1} \circ S^\dagger \circ \Phi_U) \\
&= T^* S^*.
\end{aligned}$$

Finally, notice that

$$\|T^*\| \leq \|\Phi_V^{-1}\| \|T^\dagger\| \|\Phi_W\| = \|T^\dagger\|,$$

since Φ_V and Φ_W are isometric isomorphisms. The opposite inequality follows similarly, so in total $\|T^*\| = \|T^\dagger\| = \|T\|$ by Proposition 6.6(ii).

PROPOSITION 8.2

Let V, W be Hilbert spaces, and let $T \in \mathcal{B}(V,W)$. For all $w \in W$ we have $T^\dagger \varphi_w = \varphi_{T^* w}$. In particular, T^* is the unique linear operator $W \to V$ with the property that

$$\langle v, T^* w\rangle_V = \langle Tv, w\rangle_W,$$

for all $v \in V$ and $w \in W$. Furthermore, $T^{**} = T$, i.e. the map $T \mapsto T^*$ is an involution.

The adjoint T^* is equivalently characterised by the identity

$$\langle T^* w, v\rangle_V = \langle w, Tv\rangle_W,$$

by complex conjugation. Note that either of these two identities is often taken as the definition of T^*, and existence is proved without appealing to the operator adjoint. In this case, many of the above properties are proved using the uniqueness part of this proposition.

PROOF. First notice that T^* indeed has this property. For $w \in W$ we have

$$\varphi_{T^* w} = \Phi_V(T^* w) = (T^\dagger \circ \Phi_W)(w) = T^\dagger \varphi_w,$$

so for $v \in V$ it thus follows that

$$\langle v, T^* w\rangle_V = \varphi_{T^* w}(v) = T^\dagger \varphi_w(v) = \varphi_w(Tv) = \langle Tv, w\rangle_W,$$

as desired. Furthermore, if $S \colon W \to V$ is another such operator, then $\langle v, Sw\rangle_V = \langle v, T^* w\rangle_V$ for all v and w, so $S = T^*$. The final claim that $T^{**} = T$ follows by uniqueness. $\square$

An operator $U \colon V \to W$ is an *isometry* if

$$\langle Uv, Uu\rangle_W = \langle v, u\rangle_V$$

for all $v, u \in V$. Clearly U is injective. If U is also surjective (e.g. if $\dim V = \dim W < \infty$), then it is called *unitary*. Notice that if U is an isometry, then

$$\langle U^* U v, u\rangle_V = \langle Uv, Uu\rangle_W = \langle v, u\rangle_V,$$

implying that $U^* U = \mathrm{id}_V$, and the converse clearly also holds. If U is also surjective, then it is an isomorphism and so also $UU^* = \mathrm{id}_W$ (an operator with this property is called a *coisometry*). In this case $U^* = U^{-1}$.

In the case $W = V$ we say that T is *normal* if $TT^* = T^* T$, and that T is *self-adjoint* if $T^* = T$. Clearly both self-adjoint and unitary operators (with $V = W$) are normal.

8.2. Properties of adjoints

PROPOSITION 8.3

Let V and W be real Hilbert spaces, and let $T \in \mathcal{B}(V,W)$. Then we have

$$(T^{\mathbb{C}})^* = (T^*)^{\mathbb{C}},$$

i.e., the adjoint of the complexification of T is the complexification of the adjoint of T. In particular

(i) *T is normal if and only if $T^{\mathbb{C}}$ is normal, and*

(ii) *T is self-adjoint if and only if $T^{\mathbb{C}}$ is self-adjoint.*

PROOF. For $v,u,x,y \in V$ we have

$$\begin{aligned}\langle (T^*)^{\mathbb{C}}(x+\mathrm{i}y), v+\mathrm{i}u\rangle &= \langle T^*x + \mathrm{i}T^*y, v+\mathrm{i}u\rangle \\ &= \langle T^*x, v\rangle + \langle T^*y, u\rangle + \mathrm{i}(\langle T^*y, u\rangle - \langle T^*x, v\rangle) \\ &= \langle x, Tv\rangle + \langle y, Tu\rangle + \mathrm{i}(\langle y, Tu\rangle - \langle x, Tv\rangle) \\ &= \langle x+\mathrm{i}y, Tv+\mathrm{i}Tu\rangle \\ &= \langle x+\mathrm{i}y, T^{\mathbb{C}}(v+\mathrm{i}u)\rangle.\end{aligned}$$

Uniqueness of adjoints thus yields the claim.

Assume that T is normal. Then

$$T^{\mathbb{C}}(T^{\mathbb{C}})^* = T^{\mathbb{C}}(T^*)^{\mathbb{C}} = (TT^*)^{\mathbb{C}} = (T^*T)^{\mathbb{C}} = (T^*)^{\mathbb{C}}T^{\mathbb{C}} = (T^{\mathbb{C}})^*T^{\mathbb{C}},$$

so $T^{\mathbb{C}}$ is normal. The converse follows similarly. If T is self-adjoint, then

$$(T^{\mathbb{C}})^* = (T^*)^{\mathbb{C}} = T^{\mathbb{C}},$$

and similarly if $T^{\mathbb{C}}$ is self-adjoint. □

PROPOSITION 8.4

Let V be a finite-dimensional inner product space, and let $T \in \mathcal{L}(V)$ and $\lambda \in \mathbb{K}$. Then $\lambda\,\mathrm{id}_V - T$ is invertible if and only if $\overline{\lambda}\,\mathrm{id}_V - T^*$ is invertible. In other words, λ is an eigenvalue of T if and only if $\overline{\lambda}$ is an eigenvalue of T^*.

PROOF. Since the map $T \mapsto T^*$ is an involution it suffices to prove one implication, so assume that $\lambda\,\mathrm{id}_V - T$ is invertible. Then there exists an $S \in \mathcal{L}(V)$ such that

$$S(\lambda\,\mathrm{id}_V - T) = (\lambda\,\mathrm{id}_V - T)S = \mathrm{id}_V,$$

and taking adjoints we find that

$$(\overline{\lambda}\,\mathrm{id}_V - T^*)S^* = S^*(\overline{\lambda}\,\mathrm{id}_V - T^*) = \mathrm{id}_V.$$

That is, $\overline{\lambda}\,\mathrm{id}_V - T^*$ is invertible as claimed. □

REMARK 8.5. Note that this does *not* say that $v \in V$ is an eigenvector of T^* if it is an eigenvector of T. A counterexample is given by the matrix

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

which has the eigenvector $(1,0)$ with eigenvalue 1. However, while 1 is also an eigenvalue of the transpose $A^\top$ (with eigenvector $(1,1)$), $(1,0)$ is not an eigenvector of $A^\top$.

While this does not hold in general, recall that in Proposition 8.6(ii) we saw that it holds for *normal* operators. ⌟

PROPOSITION 8.6

Let $T \in \mathcal{B}(V)$ be a normal operator.

(i) $\|Tv\| = \|T^*v\|$ *for all* $v \in V$.

(ii) *If $\lambda \in \mathbb{K}$ is an eigenvalue of T, then $\bar{\lambda}$ is an eigenvalue of T^* with the same eigenvectors. In other words,* $E_T(\lambda) = E_{T^*}(\bar{\lambda})$.

(iii) *If $\mu \in \mathbb{K}$ is another eigenvalue of T distinct from λ, then $E_T(\lambda)$ and $E_T(\mu)$ are orthogonal.*

(iv) *If T is self-adjoint, then it has an eigenvalue and all its eigenvalues are real.*

(v) *If T is unitary, then all its eigenvalues lie on the unit circle* $S^1 \subseteq \mathbb{C}$.

In Corollary 8.12 we will prove the converses of (iv) and (v) under the assumption that V is finite-dimensional and that T is normal, using the spectral theorem (cf. Theorem 8.11). We will use (iv) in the proof of the spectral theorem, and we have proved (v) already to make explicit that it does not depend on the spectral theorem, and that the proof does not require that V is finite-dimensional.

PROOF. *Proof of (i):* Notice that

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle T^*Tv, v\rangle = \langle TT^*v, v\rangle = \langle T^*v, T^*v\rangle = \|T^*v\|^2.$$

Proof of (ii): If T is normal then so is $\lambda\,\mathrm{id}_V - T$, so (i) implies that

$$\|(\lambda\,\mathrm{id}_V - T)v\| = \|(\bar{\lambda}\,\mathrm{id}_V - T^*)v\|,$$

so $v \in V$ is an eigenvector for T with eigenvalue λ if and only if v is an eigenvector for T^* with eigenvalue $\bar{\lambda}$.

Proof of (iii): Let $v \in E_T(\lambda)$ and $u \in E_T(\mu)$. Since w is also an eigenvector for T^* with eigenvalue $\overline{\mu}$, we have

$$\lambda\langle v,u\rangle = \langle Tv,u\rangle = \langle v,T^*u\rangle = \mu\langle v,u\rangle.$$

Since $\lambda \neq \mu$ we must have $\langle v,u\rangle = 0$ as claimed.

Proof of (iv): If T is self-adjoint and $v \in V$ is an eigenvector for T with $\lambda \in \mathbb{K}$, then

$$\lambda\langle v,v\rangle = \langle Tv,v\rangle = \langle v,Tv\rangle = \overline{\lambda}\langle v,v\rangle,$$

and since $v \neq 0$ we must have $\lambda = \overline{\lambda}$. Hence λ is real.

If $\mathbb{K} = \mathbb{C}$ then V has a complex eigenvalue, which is real by the above argument. Assume instead that $\mathbb{K} = \mathbb{R}$ and consider the complexification $T^{\mathbb{C}}$ of T. This is self-adjoint by Proposition 8.3, so it has a real eigenvalue by the above. But then Proposition 5.3 implies that this also is an eigenvalue of T.

Proof of (v): Let $\lambda \in \mathbb{K}$ be an eigenvalue of T with eigenvector v. Then

$$\langle v,v\rangle = \langle Tv,Tv\rangle = \langle \lambda v,\lambda v\rangle = \lambda\overline{\lambda}\langle v,v\rangle = |\lambda|^2\langle v,v\rangle,$$

so $|\lambda| = 1$. □

Let $T\colon V \to V$ be an operator on an $\mathbb{F}$-vector space V, and let U be a subspace of V that is invariant under T (i.e., $T(U) \subseteq U$). If W is a complement of V, i.e. $V = U \oplus W$, then W is not necessarily invariant under T. However, we have the following:

Lemma 8.7

Let V be a Hilbert space and let $T \in \mathcal{B}(V)$. If a subspace U of V is invariant under T, then $U^{\perp}$ is invariant under T^*.

Proof. Let $v \in U^{\perp}$. For $u \in U$ we have $Tu \in U$, so

$$\langle T^*v,u\rangle = \langle v,Tu\rangle = 0.$$

Since this holds for all $u \in U$, it follows that $T^*v \in U^{\perp}$ as desired. □

8.3. Adjoints and coordinates

Lemma 8.8

Let V and W be finite-dimensional inner product spaces, and let $\mathcal{V}$ and $\mathcal{W}$ be ordered orthonormal bases for V and W.

(i) *The coordinate map $\varphi_{\mathcal{V}}$ is unitary, i.e.*

$$\langle [v]_{\mathcal{V}},[u]_{\mathcal{V}}\rangle = \langle v,u\rangle \tag{8.1}$$

for all $v, u \in V$.

Let further $T : V \to W$ be a linear map, and let $A \in \mathrm{Mat}_{m,n}(\mathbb{K})$.

(ii) $(M_A)^* = M_{A^*}$. *In particular, if $V = \mathbb{K}^n$ and $W = \mathbb{K}^m$ then $\mathcal{M}(T^*) = \mathcal{M}(T)^*$.*

(iii) $({}_{\mathcal{W}}[T]_{\mathcal{V}})^* = {}_{\mathcal{V}}[T^*]_{\mathcal{W}}$.

Proof. *(i)*: By bi- or sesquilinearity of the inner product it suffices to prove (8.1) for a basis for V. And writing $\mathcal{V} = (v_1, \dots, v_n)$ we find that

$$\langle [v_i]_{\mathcal{V}}, [v_j]_{\mathcal{V}} \rangle = \langle e_i, e_j \rangle = \delta_{ij} = \langle v_i, v_j \rangle$$

for $1 \le i, j \le n$.

(ii): Notice that

$$\langle M_{A^*} w, v \rangle = \langle A^* w, v \rangle = v^*(A^* w) = (Av)^* w = \langle w, Av \rangle = \langle w, M_A v \rangle$$

for all $v \in \mathbb{K}^n$ and $w \in \mathbb{K}^m$. By uniqueness of the adjoint operator, it follows that $(M_A)^* = M_{A^*}$. Furthermore, we have

$$M_{\mathcal{M}(T^*)} = T^* = (M_{\mathcal{M}(T)})^* = M_{\mathcal{M}(T)^*}.$$

It follows that $\mathcal{M}(T^*) = \mathcal{M}(T)^*$.

(iii): Notice that

$$(\varphi_{\mathcal{W}} \circ T \circ \varphi_{\mathcal{V}}^{-1})^* = (\varphi_{\mathcal{W}} \circ T \circ \varphi_{\mathcal{V}}^*)^* = \varphi_{\mathcal{V}} \circ T^* \circ \varphi_{\mathcal{W}}^* = \varphi_{\mathcal{V}} \circ T^* \circ \varphi_{\mathcal{W}}^{-1},$$

and taking standard matrix representations, it follows from (ii) that $({}_{\mathcal{W}}[T]_{\mathcal{V}})^* = {}_{\mathcal{V}}[T^*]_{\mathcal{W}}$. □

Lemma 8.9

Let V be a finite-dimensional vector space, let $T \in \mathcal{L}(V)$, and let $\mathcal{V}$ be an ordered basis for V. Then $v \in V$ is an eigenvector for T if and only if $[v]_{\mathcal{V}}$ is an eigenvector for ${}_{\mathcal{V}}[T]_{\mathcal{V}}$ with the same eigenvalue.

Proof. Let $\lambda \in \mathbb{F}$ be the eigenvalue of v. Then

$${}_{\mathcal{V}}[T]_{\mathcal{V}} \cdot [v]_{\mathcal{V}} = [Tv]_{\mathcal{V}} = [\lambda v]_{\mathcal{V}} = \lambda [v]_{\mathcal{V}}.$$

For the converse, a similar calculation shows that $[Tv]_{\mathcal{V}} = [\lambda v]_{\mathcal{V}}$. Since $\varphi_{\mathcal{V}}$ is an isomorphism, it follows that $Tv = \lambda v$ as desired. □

8.4. The spectral theorem

LEMMA 8.10

V be a finite-dimensional inner product space over $\mathbb{K}$, and consider $T \in \mathcal{L}(V)$. If either

(i) $\mathbb{K} = \mathbb{R}$ *and T is self-adjoint, or*

(ii) $\mathbb{K} = \mathbb{C}$ *and T is normal,*

then T is orthogonally diagonalisable.

PROOF. Assume that either $\mathbb{K} = \mathbb{R}$ and T is self-adjoint, or that $\mathbb{K} = \mathbb{C}$ and T is normal. We prove by induction in $n = \dim V$ that T is orthogonally diagonalisable. If $n = 1$ then this follows since T has an eigenvalue, so assume that the claim is proved for operators on spaces of dimension strictly less than n.

Let $\lambda \in \operatorname{Spec} T$, and consider the corresponding eigenspace $E_T(\lambda)$. If $d := \dim E_T(\lambda) = n$, then any orthonormal basis of $E_T(\lambda)$ will suffice. Assume therefore that $0 < d < n$.

The space $E_T(\lambda) = E_{T^*}(\overline{\lambda})$ is clearly invariant under both T and T^*. It follows from Lemma 8.7 that $E_T(\lambda)^\perp$ is also invariant under both T and T^*. We furthermore have $\dim E_T(\lambda)^\perp = n - d$ and $0 < n - d < n$. Let $T_\parallel \in \mathcal{L}(E_T(\lambda))$ and $T_\perp \in \mathcal{L}(E_T(\lambda)^\perp)$ denote the restrictions of T to $E_T(\lambda)$ and $E_T(\lambda)^\perp$ respectively. Both $T_\parallel$ and $T_\perp$ are also self-adjoint or normal, depending on the hypothesis, so the induction hypothesis furnishes orthonormal bases $\mathcal{U}$ and $\mathcal{W}$ for $E_T(\lambda)$ and $E_T(\lambda)^\perp$ consisting of eigenvectors of T. But then $\mathcal{V} = \mathcal{U} \cup \mathcal{W}$ is an orthonormal basis for V as desired. □

THEOREM 8.11: *The spectral theorem*

Let V be a finite-dimensional inner product space over $\mathbb{K}$, and let $T \in \mathcal{L}(V)$. Then the following are equivalent:

(i) $\mathbb{K} = \mathbb{R}$ *and T is self-adjoint, or* $\mathbb{K} = \mathbb{C}$ *and T is normal.*

(ii) *T is orthogonally diagonalisable.*

(iii) *T has the orthogonal spectral resolution*

$$T = \sum_{\lambda \in \operatorname{Spec} T} \lambda P_\lambda,$$

where P_λ is the orthogonal projection onto the eigenspace $E_T(\lambda)$. In particular,

V is an orthogonal direct sum of the eigenspaces of T, i.e.

$$V = \bigodot_{\lambda \in \operatorname{Spec} T} E_T(\lambda).$$

(iv) *T is unitarily (when $\mathbb{K} = \mathbb{C}$) or orthogonally (when $\mathbb{K} = \mathbb{R}$) equivalent to a multiplication operator $M_A \in \mathcal{L}(\mathbb{K}^n)$ where A is a diagonal matrix, and the diagonal of A contains the eigenvalues of T with multiplicity. If $\mathcal{V}$ is an ordered orthonormal basis for V consisting of eigenvectors for T, then we may choose $A = {}_{\mathcal{V}}[T]_{\mathcal{V}}$ and*

$$T = \varphi_{\mathcal{V}}^{-1} \circ M_A \circ \varphi_{\mathcal{V}},$$

with $\varphi_{\mathcal{V}}$ unitary.

Note that the first part of property (iii) means that

$$\mathrm{id}_V = \sum_{\lambda \in \operatorname{Spec} T} P_\lambda$$

is a resolution of the identity, i.e. that $P_\lambda P_\mu = 0$ for $\lambda \neq \mu$, and that this is composed of orthogonal projections.

Proof. *(i) $\Rightarrow$ (ii)*: This is just Lemma 8.10.

(i) & (ii) $\Rightarrow$ (iii): The first claim says that distinct eigenspaces are orthogonal, which is just a restatement of Proposition 8.6(iii). To prove the second, let $\mathcal{V} = (v_1, \ldots, v_n)$ be an orthonormal basis for V consisting of eigenvectors for T, and let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues. Then for any $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$ we have $P_{\lambda_i} v = \alpha_i v_i$, so

$$\left(\sum_{\lambda \in \operatorname{Spec} T} P_\lambda \right) v = \sum_{\lambda \in \operatorname{Spec} T} P_\lambda v = \sum_{i=1}^n \alpha_i v_i = v.$$

For the third claim, notice that

$$\left(\sum_{\lambda \in \operatorname{Spec} T} \lambda P_\lambda \right) v = \sum_{\lambda \in \operatorname{Spec} T} \lambda P_\lambda v = \sum_{i=1}^n \lambda_i \alpha_i v_i = \sum_{i=1}^n \alpha_i T v_i = Tv.$$

The final claim follows from the first two.

(iii) $\Rightarrow$ (ii): This follows from the decomposition of V into an orthogonal sum of eigenspaces, by constructing an orthonormal basis for each eigenspace.

(ii) $\Rightarrow$ (iv): Let $\mathcal{V} = (v_1, \ldots, v_n)$ be an ordered orthonormal basis for $\mathcal{V}$ consisting of eigenvectors for T with corresponding eigenvalues $\lambda_1, \ldots, \lambda_n$, and

consider the matrix representation ${}_{\mathcal{V}}[T]_{\mathcal{V}}$. If $(e_1,\dots,e_n)$ is the standard basis on $\mathbb{K}^n$, then Lemma 8.9 implies that the vectors $[v_i]_{\mathcal{V}} = e_i$ are eigenvectors for ${}_{\mathcal{V}}[T]_{\mathcal{V}}$. Hence ${}_{\mathcal{V}}[T]_{\mathcal{V}}$ is diagonal, so the basis representation $\varphi_{\mathcal{V}} \circ T \circ \varphi_{\mathcal{V}}^{-1}$ is multiplication by a diagonal matrix. Next notice that

$$T = \varphi_{\mathcal{V}}^{-1} \circ (\varphi_{\mathcal{V}} \circ T \circ \varphi_{\mathcal{V}}^{-1}) \circ \varphi_{\mathcal{V}},$$

so it suffices to show that $\varphi_{\mathcal{V}}$ is unitary (orthogonal). But this follows by Lemma 8.8.

(iv) ⇒ (i): First assume that $\mathbb{K} = \mathbb{C}$. Since $\varphi_{\mathcal{V}}$ is unitary we have $\varphi_{\mathcal{V}}^{-1} = \varphi_{\mathcal{V}}^*$, so

$$T^* = (\varphi_{\mathcal{V}}^* \circ M_A \circ \varphi_{\mathcal{V}})^* = \varphi_{\mathcal{V}}^* \circ M_A^* \circ \varphi_{\mathcal{V}} = \varphi_{\mathcal{V}}^{-1} \circ M_{A^*} \circ \varphi_{\mathcal{V}}.$$

Since A is diagonal, T clearly commutes with T^*, hence is normal.

If instead $\mathbb{K} = \mathbb{R}$, the same argument shows that $T^* = \varphi_{\mathcal{V}}^{-1} \circ M_{A^\top} \circ \varphi_{\mathcal{V}}$, but since A is diagonal this is just T, so T is self-adjoint. □

COROLLARY 8.12

Let $T \in \mathcal{L}(V)$ be a normal operator on a complex vector space V.

(i) *T is self-adjoint if and only if* $\operatorname{Spec} T \subseteq \mathbb{R}$.

(ii) *T is unitary if and only if* $\operatorname{Spec} T \subseteq S^1$.

Note that this does not hold on a real vector space, since then a normal operator is not necessarily diagonalisable.

PROOF. *Proof of (i)*: The 'only if' part follows from Proposition 8.6(iv), so assume that $\operatorname{Spec} T \subseteq \mathbb{R}$ and notice that

$$T^* = \Bigg(\sum_{\lambda \in \operatorname{Spec} T} \lambda P_\lambda\Bigg)^* = \sum_{\lambda \in \operatorname{Spec} T} \bar{\lambda} P_\lambda^* = \sum_{\lambda \in \operatorname{Spec} T} \lambda P_\lambda,$$

since each $\lambda \in \mathbb{R}$, and each P_λ is an orthogonal projection, hence self-adjoint.

Alternatively, choose a diagonal matrix $A \in \operatorname{Mat}_n(\mathbb{K})$ in accordance with Theorem 8.11(iv). Since the diagonal of A contains the eigenvalues of T, we have $A^* = A$, and so it follows that $T^* = T$.

Proof of (ii): Similarly, the 'only if' part is just Proposition 8.6(v). Assume that $\operatorname{Spec} T \subseteq S^1$ and notice that

$$T^* = \sum_{\lambda \in \operatorname{Spec} T} \bar{\lambda} P_\lambda.$$

Since the projections P_λ are pairwise orthogonal, we have

$$T^*T = \sum_{\lambda\in\operatorname{Spec} T} \bar{\lambda}\lambda P_\lambda = \sum_{\lambda\in\operatorname{Spec} T} |\lambda|^2 P_\lambda = \sum_{\lambda\in\operatorname{Spec} T} P_\lambda = \mathrm{id}_V,$$

so U is unitary.

Alternatively, let A be as above. Then all diagonal elements in A are nonzero, so A is invertible, and we clearly have $A^*A = I_n$. Hence also $T^*T = \mathrm{id}_V$, so T is unitary. □

9 • Projections

Let V be an $\mathbb{F}$-vector space. A linear operator $P\colon V \to V$ is called a *projection* if it is idempotent, i.e. if $P^2 = P$.

Proposition 9.1

A linear map $P\colon V \to V$ is a projection if and only if there exist subspaces U and W of V such that $V = U \oplus W$ and $P|_U = \iota_U$. In this case $U = \operatorname{im} P$ and $W = \ker P$.

We say that P is the projection onto U along W.

Proof. Assume that P is a projection, and let $v \in \operatorname{im} P$. Then $v = Pu$ for some $u \in V$, and

$$Pv = P^2u = Pu = v.$$

If also $v \in \ker P$, then $v = 0$. Furthermore, for any $v \in V$ we have $v = Pv + (v - Pv) \in \operatorname{im} P \oplus \ker P$, so $\operatorname{im} P$ and $\ker P$ are indeed complements in V.

The converse is obvious, and so is the characterisation of U and W. □

Now let V be a real finite-dimensional[3] inner product space. A projection $P\colon V \to V$ is *orthogonal* if $\operatorname{im} P$ and $\ker P$ are orthogonal subspaces of V.

Proposition 9.2

A projection $P\colon V \to V$ is orthogonal if and only if P is self-adjoint.

Proof. Say that P is a projection onto U along W. Assume that P is orthogonal and let $v, w \in V$. Since then $Pv \in U$ and $v - Pv \in W$, and similarly for w, we get

$$\langle v - Pv, Pw\rangle = 0 = \langle Pv, w - Pw\rangle.$$

[3] Since projection operators are clearly bounded, the discussion below readily generalises to infinite-dimensional inner product spaces.

This implies that

$$\langle v, Pw\rangle = \langle Pv, Pw\rangle = \langle Pv, w\rangle = \langle v, P^* w\rangle,$$

which shows that $P = P^*$.

Conversely assume that P is self-adjoint. For $u \in U$ and $w \in W$ we then have

$$\langle u, w\rangle = \langle Pu, w\rangle = \langle u, Pw\rangle = \langle u, 0\rangle = 0,$$

so U and W are orthogonal. □

PROPOSITION 9.3

Let $T\colon V \to W$ be an injective linear operator between real inner product spaces V and W, and let P be the orthogonal projection onto $\operatorname{im} T$. Then $P = T(T^*T)^{-1}T^*$.

PROOF. First note that T^*T is indeed injective (hence invertible) since T is. This follows from the identity $\ker T^* = (\operatorname{im} T)^\perp$.

Next notice that the rank of P is $\dim \operatorname{im} T$. But T^* is surjective since T is injective, so the rank of $T(T^*T)^{-1}T^*$ is also $\dim \operatorname{im} T$. It thus suffices to show that P and $T(T^*T)^{-1}T^*$ agree on $\operatorname{im} T$, and writing $w = Tv$ we have

$$T(T^*T)^{-1}T^*w = T(T^*T)^{-1}(T^*T)v = Tv = w,$$

as desired. □

References

Axler, Sheldon (2015). *Linear Algebra Done Right*. 3rd ed. Springer. 340 pp. ISBN: 978-3-319-11079-0. DOI: 10.1007/978-3-319-11080-6.

Folland, Gerald B. (2007). *Real Analysis: Modern Techniques and Their Applications*. 2nd ed. Wiley. 386 pp. ISBN: 0-471-31716-0.

Hoffman, Kenneth and Ray Kunze (1971). *Linear Algebra*. 2nd ed. Prentice-Hall. 407 pp.

Roman, Steven (2008). *Advanced Linear Algebra*. 3rd ed. Springer. 522 pp. ISBN: 978-0-387-72828-5.

Thomsen, Jesper Funch (2017). *Lineær algebra*. Department of Mathematics, Aarhus University. 251 pp.