

Notes on linear algebra

Danny Nygård Hansen

19th November 2022

1 • Linear equations and matrices

1.1. 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

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

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

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

A *system of linear equations in n unknowns* is a tuple $L = (l_1, \dots, l_m)$, 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, \dots, 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.

1.2. Matrices

It is well-known that a system of linear equations is equivalent to a matrix equation on the form $Ax = b$, where $A \in \text{Mat}_{m,n}(\mathbb{F})$, $x \in \mathbb{F}^n$ 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 $A, B \in \text{Mat}_{m,n}(\mathbb{F})$ 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 $A, B \in \text{Mat}_{m,n}(\mathbb{F})$ are row-equivalent, then the systems of equations $Ax = 0$ and $Bx = 0$ are combination equivalent, hence have the same solutions.

DEFINITION 1.1

A matrix $H \in \text{Mat}_{m,n}(\mathbb{F})$ 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 $k_1 < \dots < k_r$.

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 $E = e(I) \in \text{Mat}_m(\mathbb{F})$, then $e(A) = EA$ for $A \in \text{Mat}_{m,n}(\mathbb{F})$. If $B \in \text{Mat}_{m,n}(\mathbb{F})$, then A and B are row-equivalent if and only if $A = PB$, where $P \in \text{Mat}_m(\mathbb{F})$ is a product of elementary matrices.

PROPOSITION 1.2

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

PROOF. The usual Gauss–Jordan elimination algorithm proves existence. If $H, K \in \text{Mat}_{m,n}(\mathbb{F})$ 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 n th

column in H and K respectively. Then H_1 and K_1 are also row-equivalent¹ and row-reduced echelon matrices, so by induction $H_1 = K_1$. Thus if H and K differ, they must differ in the n th column.

Let H_2 be the matrix obtained by deleting columns in H , only keeping those columns containing pivots, as well as keeping the n th 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 $H_2 = K_2$. \square

1.3. Invertible matrices

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

LEMMA 1.3

If $A \in \text{Mat}_n(\mathbb{F})$, 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) \Leftrightarrow (ii): Let $H \in \text{Mat}_n(\mathbb{F})$ be a row-reduced echelon matrix that is row-equivalent to A . Then $H = PA$, where $P \in \text{Mat}_n(\mathbb{F})$ 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.

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

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

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

(ii) \Leftrightarrow (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 $H \in \text{Mat}_{m,n}(\mathbb{F})$ 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. \square

PROPOSITION 1.4

Let $A \in \text{Mat}_n(\mathbb{F})$. 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 $B \in \text{Mat}_n(\mathbb{F})$, 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. \square

2 • Coordinates

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

PROPOSITION 2.1

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

$$\begin{aligned} \mathcal{M}: \mathcal{L}(\mathbb{F}^n, \mathbb{F}^m) &\rightarrow \text{Mat}_{m,n}(\mathbb{F}), \\ T &\mapsto (Te_1 \mid \cdots \mid Te_n), \end{aligned}$$

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

- (i) $Tv = \mathcal{M}(T)v$ for all $v \in \mathbb{F}^n$.
- (ii) $\mathcal{M}(\text{id}_{\mathbb{F}^n}) = I$.

(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

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

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

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

since Ae_i is the i th 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). \square

Let V be a finite-dimensional \mathbb{F} -vector space. If $\mathcal{V} = (v_1, \dots, v_n)$ is an ordered basis for V , then for every $v \in V$ there are unique $\alpha_1, \dots, \alpha_n \in \mathbb{F}$ such that $v = \sum_{i=1}^n \alpha_i v_i$. Hence the map $\varphi_{\mathcal{V}}: V \rightarrow \mathbb{F}^n$ given by $\varphi_{\mathcal{V}}(v) = (\alpha_1, \dots, \alpha_n)$ 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

$$\begin{array}{ccc} & & \mathbb{F}^n \\ & \nearrow \varphi_{\mathcal{V}} & \downarrow \varphi_{\mathcal{W}, \mathcal{V}} \\ V & & \mathbb{F}^n \\ & \searrow \varphi_{\mathcal{W}} & \end{array} \quad (2.1)$$

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

PROPOSITION 2.2

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 2.1](#) from the corresponding claims about change of basis operators.

The claim (i) follows by commutativity of the diagram (2.1), 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 $T: V \rightarrow W$. If $\mathcal{V} \in V^n$ and $\mathcal{W} \in W^m$ are bases for V and W respectively, then the diagram

$$\begin{array}{ccc} V & \xrightarrow{\varphi_{\mathcal{V}}} & \mathbb{F}^n \\ T \downarrow & & \downarrow \varphi_{\mathcal{W}} \circ T \circ \varphi_{\mathcal{V}}^{-1} \\ W & \xrightarrow{\varphi_{\mathcal{W}}} & \mathbb{F}^m \end{array}$$

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 $\mathbb{F}^n \rightarrow \mathbb{F}^m$, 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 2.3

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

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

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

- (i) $[Tv]_{\mathcal{W}} = {}_{\mathcal{W}}[T]_{\mathcal{V}} \cdot [v]_{\mathcal{V}}$ for all $v \in V$.
- (ii) If \mathcal{V}' is another basis for V , then ${}_{\mathcal{V}'}[\text{id}_V]_{\mathcal{V}} = {}_{\mathcal{V}'}[\square]_{\mathcal{V}}$.

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

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

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 2.1](#).

Proof of (i): 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 2.1\(i\)](#).

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

Proof of (iii): 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 2.1\(iii\)](#).

Proof of (iv): This is an immediate consequence of either [\(iii\)](#) or of [Proposition 2.1\(iv\)](#). \square

PROPOSITION 2.4

Let $\mathcal{V} = (v_1, \dots, v_n)$ be an ordered basis for an \mathbb{F} -vector space V , and let $T: V \rightarrow V$ 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

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

implying that

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

as claimed. \square

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

3 • Determinants

3.1. Existence of determinants

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

$$\varphi: M_1 \times \dots \times M_n \rightarrow N$$

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 $\text{Mat}_{m,n}(R) \cong (R^n)^m$, a map $\varphi: \text{Mat}_{m,n}(R) \rightarrow N$ that is linear in each row is also called *n-linear*.

In the case $M_1 = \dots = M_n$, 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 3.1

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

- (i) If φ is alternating, then φ is skew-symmetric. If $\text{char } R \neq 2$ 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 3.1\(i\)](#) 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 $\psi: M \times M \rightarrow N$ 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 $\text{char } R \neq 2$ 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 i th and j th 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. \square

DEFINITION 3.2: Determinant functions

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

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

THEOREM 3.3: Construction of determinants

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

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

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

PROOF. Let $A = (a_{ij}) \in \text{Mat}_n(R)$. 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 3.1(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. \square

COROLLARY 3.4: Existence of determinants

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

PROOF. The identity map on $\text{Mat}_1(R) \cong R$ is a determinant function for $n = 1$, and **Theorem 3.3** allows us to recursively construct a determinant for each $n > 1$. \square

3.2. Uniqueness of determinants

THEOREM 3.5: Uniqueness of determinants

Let n be a positive integer. There is precisely one determinant function on $\text{Mat}_n(R)$,

namely the function $\det: \text{Mat}_n(R) \rightarrow R$ given by

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

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

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

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

PROOF. Let e_1, \dots, e_n denote the rows of I_n , and denote the rows of a matrix $A = (a_{ij}) \in \text{Mat}_n(R)$ 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 3.1\(i\)](#), 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

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

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. \square

3.3. Properties of determinants

THEOREM 3.6

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

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

In particular, $\det: \text{GL}_n(R) \rightarrow R^*$ is a group homomorphism.

PROOF. The map $\varphi: \text{Mat}_n(R) \rightarrow R$ 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. \square

COROLLARY 3.7

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

PROOF. Let $P \in \text{Mat}_n(\mathbb{F})$ be such that $A = PBP^{-1}$. [Theorem 3.6](#) then implies that

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

[Corollary 3.7](#) allows us to define the determinant of a general linear operator $T: V \rightarrow V$ 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 3.8

Let A_{11}, \dots, 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 $A \in \text{Mat}_r(R)$, $B \in \text{Mat}_s(R)$ and $C \in \text{Mat}_{r,s}(R)$ 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 3.3](#) 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. \square

PROPOSITION 3.9

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

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

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

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

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

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

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

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

We have the following:

PROPOSITION 3.10

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

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

PROOF. Writing $A = (a_{ij})$ and fixing some $j \in \{1, \dots, n\}$, Theorem 3.3 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 $(\text{adj } A)A$.

Next we claim that if $k \neq j$, then $\sum_{i=1}^n a_{ik} A_{i,j} = 0$. Let $B = (b_{ij}) \in \text{Mat}_n(R)$ be the matrix obtained from A by replacing the j th column of A by its k th

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 3.11

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

- (i) A is a (two-sided) unit in $\operatorname{Mat}_n(R)$.
- (ii) A is a left- or right-unit in $\operatorname{Mat}_n(R)$.
- (iii) $\det A$ is a unit in R .

PROOF. If A is e.g. a left-unit, then [Theorem 3.6](#) 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 3.10](#) 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.

3.4. 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 $\text{Spec } T$. Clearly $\lambda \in \text{Spec } T$ 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 3.12

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 $-\text{tr } T$.

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

$$\det(tI - T) = \sum_{\sigma \in S_n} (\text{sgn } \sigma) (t\delta_{1\sigma(1)} - a_{1\sigma(1)}) \cdots (t\delta_{n\sigma(n)} - a_{n\sigma(n)}) \quad (3.1)$$

by [Theorem 3.5](#). 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 (3.1) 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

$$(t - a_{11})(t - a_{22}) \cdots (t - a_{nn}), \quad (3.2)$$

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 $c_0, \dots, c_n \in \mathbb{F}$.

(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 (3.1) 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 (3.2). It is then clear that

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

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 $\lambda_1, \dots, \lambda_n \in \mathbb{F}$. 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 $c_{n-1} = -(\lambda_1 + \dots + \lambda_n)$, so the final claim follows from (iii). \square

3.5. 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 $\operatorname{ev}_T: \mathbb{F}[t]_{n^2+1} \rightarrow \mathcal{L}(V)$ given by $\operatorname{ev}_T(p) = p(T)$. 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 $c, \lambda_1, \dots, \lambda_m \in \mathbb{F}$ 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 3.13

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

PROOF. Let $A \in \text{Mat}_n(\mathbb{F})$ 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. \square

3.6. Cross products

DEFINITION 3.14: 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 3.15

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. \square

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 3.18](#). The scalar triple product has some very nice properties summarised in the following proposition:

PROPOSITION 3.16

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 3.15](#) 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

$$[v, v, w] = \det(v, v, w) = 0,$$

and similarly for w . □

PROPOSITION 3.17

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

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

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 $w = e_l$ 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. \square

COROLLARY 3.18

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

$$u \times (v \times w) = v\langle u, w \rangle - w\langle u, v \rangle. \quad (3.3)$$

In particular, the cross product satisfies the Jacobi identity

$$u \times (v \times w) + v \times (w \times u) + w \times (u \times v) = 0. \quad (3.4)$$

The identity (3.3) 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

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

The claim then follows since x was arbitrary. \square

LEMMA 3.19

Let $A \in \text{Mat}_d(\mathbb{R})$. 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 $A_n \rightarrow A$ for $n \rightarrow \infty$.

Since $\text{Mat}_d(\mathbb{R})$ is a finite-dimensional vector space, it has a unique vector space topology. More concretely, all norms on $\text{Mat}_d(\mathbb{R})$ are Lipschitz equivalent, so we may choose whatever norm we wish. We choose the Euclidean norm, identifying $\text{Mat}_d(\mathbb{R})$ 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 $A - tI \rightarrow A$ as $t \rightarrow 0$, the claim follows. \square

PROPOSITION 3.20: Transformation of cross products

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

- (i) $[Au, Av, Aw] = (\det A)[u, v, w]$.
- (ii) $Av \times Aw = (\text{cof } A)(v \times w) = (\text{adj } A)^\top (v \times w)$.
- (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 increases 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

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

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

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

where the last equality follows from [Proposition 3.10](#). Hence we obtain the desired identity when A is invertible. Finally notice that both the maps $A \mapsto \text{cof } A$ and $A \mapsto Av \times Aw$ are continuous. Hence the claim for general A follows from [Lemma 3.19](#).

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



If A is a proper rotation, i.e. if A is orthogonal and $\det A = 1$, then [Proposition 3.20\(iii\)](#) implies that $A(v \times w) = Av \times Aw$. 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

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

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’ (3.3) becomes

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

Hence $u \times (v \times w) = v \langle u, w \rangle - w \langle u, v \rangle$ since φ_V is an isomorphism. In particular, the cross product on V also satisfies the Jacobi identity (3.4), so V becomes a Lie algebra whose Lie bracket is given by the cross product, i.e. $[v, w] = v \times w$.

4 • Complexification

If W is a complex vector space, then we may restrict the scalar multiplication $\mathbb{C} \times W \rightarrow W$ to a map $\mathbb{R} \times W \rightarrow W$. 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

$$(a + ib)(v, u) = (av - bu, au + bv),$$

for $a, b \in \mathbb{R}$ and $v, u \in V$. We denote the vector (v, u) by $v + iu$.

If $T: V \rightarrow W$ is a linear map between real vector spaces, then we define the complexification of T by

$$\begin{aligned}
 T^{\mathbb{C}}: V^{\mathbb{C}} &\rightarrow W^{\mathbb{C}}, \\
 v + iu &\mapsto Tv + iTu.
 \end{aligned}$$

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

If V is a real inner product space, then we define an inner product by

$$\langle v + iu, x + iy \rangle = \langle v, x \rangle + \langle u, y \rangle + i(\langle u, x \rangle - \langle v, y \rangle).$$

Notice that this identity holds in any *complex* inner product space, where the notation $v + iu$ 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 *sesquilinear*).

5 • Operator adjoints

DEFINITION 5.1: Operator adjoints

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

$$\begin{aligned} T^*: W^* &\rightarrow V^*, \\ \varphi &\mapsto \varphi \circ T. \end{aligned}$$

Note that this is just the action of the dual functor on maps in the category of \mathbb{F} -vector spaces. Hence it already satisfies $\text{id}_V^* = \text{id}_{V^*}$ and $(ST)^* = T^*S^*$, so that in particular $(T^{-1})^* = (T^*)^{-1}$ when T is invertible. Furthermore, it is easy to show that the map $T \mapsto T^*$ 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 5.2

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

- (i) $\ker T^* = (\text{im } T)^0$.
- (ii) $\text{im } T^* = (\ker T)^0$.

PROOF. Roman (2008, Theorem 3.19). □

COROLLARY 5.3

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

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

$$\text{rank } T^* = \dim \text{im } T^* = \dim (\ker T)^0 = \dim V - \dim \ker T = \dim \text{im } T = \text{rank } T,$$

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 $\varphi = \varphi_1 v_1^* + \dots + \varphi_n v_n^*$ 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 5.4

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

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

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

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

and

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

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

COROLLARY 5.5

The row rank and the column rank of a matrix $A \in \text{Mat}_{m,n}(\mathbb{F})$ 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 5.4](#) then implies that the matrix representation of $(M_A)^*$ 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 5.3](#) 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. \square

If V is a finite-dimensional inner product space, for $v \in V$ let φ_v denote the element in V^* given by $\varphi_v(w) = \langle w, v \rangle$. Further, let $\Phi_V: V \rightarrow V^*$ denote the (conjugate-)linear isomorphism $v \mapsto \varphi_v$.

THEOREM 5.6

Let V and W be finite-dimensional inner product spaces, and let $T \in \mathcal{L}(V, W)$. Denoting the Hilbert space adjoint of T by $T^{\dagger}: W \rightarrow V$ we have

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

i.e. the diagram

$$\begin{array}{ccc} V & \xrightleftharpoons[T^{\dagger}]{} & W \\ \Phi_V \downarrow & & \downarrow \Phi_W \\ V^* & \xleftarrow{T^*} & W^* \end{array}$$

commutes. [TODO also commutes when T is there?]

PROOF. Simply notice that, for $v \in V$ and $\varphi \in W^*$, we have

$$T^* \varphi(v) = \varphi(Tv) = \langle Tv, \Phi_W^{-1}(\varphi) \rangle = \langle v, T^\dagger \Phi_W^{-1}(\varphi) \rangle = \Phi_V(T^\dagger \Phi_W^{-1}(\varphi))(v),$$

which implies the claim. \square

6 • Triangularisation and diagonalisation

6.1. Triangularisation

Recall that a matrix $A = (a_{ij}) \in \text{Mat}_n(R)$ 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 $T \in \mathcal{L}(V)$* if $T(U) \subseteq U$.

PROPOSITION 6.1

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 $\text{span}(v_1, \dots, v_i)$ is invariant under T for all $i \in \{1, \dots, n\}$.

PROOF. This is obvious. \square

LEMMA 6.2

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 $T|_U: U \rightarrow U$ 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} . \square

PROPOSITION 6.3

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 $\text{GL}_n(\mathbb{F})$ consisting of upper triangular matrices is a subgroup.

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

LEMMA 6.4

Let $A \in \text{Mat}_n(\mathbb{F})$ 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

$$Ae_i = a_{i1}e_1 + \dots + a_{i-1}e_{i-1} + \lambda_i e_i$$

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

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

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

LEMMA 6.5

Let $A \in \text{Mat}_n(\mathbb{F})$ 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 6.4, 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 . \square

PROPOSITION 6.6

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 $U = \text{span}(v_1)$ is invariant under T , so we may define a linear operator² $\tilde{T} \in \mathcal{L}(V/U)$ by

² The operator \tilde{T} may arise as follows: Let $\pi: V \rightarrow V/U$ 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 $\tilde{T}: V/U \rightarrow V/U$.

$\tilde{T}(v + U) = Tv + U$. 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

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

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

$$Tv_i + U = (a_2v_2 + \dots + a_iv_i) + U.$$

But then $Tv_i \in \text{span}(v_1, \dots, v_i)$ 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 . \square

THEOREM 6.7: 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 6.6 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

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

for all $i \in \{1, \dots, n\}$. Then Proposition 6.1 shows that ${}_{\mathcal{U}}[T]_{\mathcal{U}}$ is also upper triangular, proving the claim. \square

6.2. Orthonormal diagonalisation

Let V and W be finite-dimensional inner product spaces, and let $T \in \mathcal{L}(V, W)$. Recall that the *adjoint* of T is the operator $T^* \in \mathcal{L}(W, V)$ with the property that

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

or by complex conjugation equivalently

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

for all $v \in V$ and $w \in W$. An operator with this property is unique if it exists, since if $S \in \mathcal{L}(W, V)$ is another such operator, then $\langle v, Sw \rangle_V = \langle v, T^*w \rangle_V$ for all v and w , so $S = T^*$.

For existence, for $w \in W$ define $\psi_w \in V^*$ by $\psi_w(v) = \langle Tv, w \rangle$, and let $\Psi_W: W \rightarrow V^*$ be the map $\Psi_W(w) = \psi_w$. Then define $T^* = \Phi_V^{-1} \circ \Psi_W$. Both Φ_V and Ψ_W are (conjugate-)linear, so T^* is linear. Furthermore we have

$$\langle v, T^*w \rangle_V = \langle v, \Phi_V^{-1} \circ \Psi_W(w) \rangle_V = \psi_w(v) = \langle Tv, w \rangle_W$$

as required.



An operator $U: V \rightarrow W$ 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 (i.e. if $\dim V = \dim W < \infty$), then it is called *unitary*. Notice that if U is an isometry, then

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

implying that $U^*U = \text{id}_V$, and the converse clearly also holds. If U is also surjective, then it is an isomorphism and so also $UU^* = \text{id}_W$ (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.

LEMMA 6.8

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.

$$\langle [v]_{\mathcal{V}}, [u]_{\mathcal{V}} \rangle = \langle v, u \rangle \quad (6.1)$$

for all $v, u \in V$.

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

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

PROOF. (i): By bi- or sesquilinearity of the inner product it suffices to prove (6.1) 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 $1 \leq i, j \leq n$.

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

PROPOSITION 6.9

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

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

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

and taking adjoints we find that

$$(\bar{\lambda} \text{id}_V - T^*)S^* = S^*(\bar{\lambda} \text{id}_V - T^*) = \text{id}_V.$$

That is, $\bar{\lambda} \text{id}_V - T^*$ is invertible as claimed. \square

REMARK 6.10. 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^T (with eigenvector $(1, 1)$), $(1, 0)$ is not an eigenvector of A^T .

While this does not hold in general, in [Proposition 6.14\(ii\)](#) we will see that it holds for *normal* operators. \lrcorner

PROPOSITION 6.11

Let V and W be real inner product spaces, and let $T \in \mathcal{L}(V, W)$. 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

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

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. □

LEMMA 6.12

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. □

LEMMA 6.13

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 .

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

$$Tv + iTu = T^{\mathbb{C}}(v + iu) = \lambda(v + iu) = \lambda v + i\lambda u.$$

It follows that $Tv = \lambda v$ as desired. \square

PROPOSITION 6.14

Let $T \in \mathcal{L}(V)$ 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 6.18](#) we will prove the converses of (iv) and (v) under the assumption that T is normal, using the spectral theorem (cf. [Theorem 6.17](#)). 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.

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 $\lambda \text{id}_V - T$, so (i) implies that

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

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 $\bar{\mu}$, 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

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

and since $v \neq 0$ we must have $\lambda = \bar{\lambda}$. 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 6.11](#), so it has a real eigenvalue by the above. But then [Lemma 6.13](#) 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

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

so $|\lambda| = 1$. □

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

LEMMA 6.15

Let $T \in \mathcal{L}(V)$ be an operator on a finite-dimensional inner product space 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. □

LEMMA 6.16

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 $\lambda \in \text{Spec } T$, 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 $E_T(\lambda) = E_{T^*}(\bar{\lambda})$ is clearly invariant under both T and T^* . It follows from Lemma 6.15 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. \square

THEOREM 6.17: 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

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

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.

$$V = \bigoplus_{\lambda \in \text{Spec } T} E_T(\lambda).$$

- (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

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

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 6.16.

(i) & (ii) \Rightarrow (iii): The first claim says that distinct eigenspaces are orthogonal, which is just a restatement of Proposition 6.14(iii). To prove the second, let $\mathcal{V} = (v_1, \dots, v_n)$ be an orthonormal basis for V consisting of eigenvectors for T , and let $\lambda_1, \dots, \lambda_n$ be the corresponding eigenvalues. Then for any $v = \alpha_1 v_1 + \dots + \alpha_n v_n$ we have $P_{\lambda_i} v = \alpha_i v_i$, so

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

For the third claim, notice that

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

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 $\mathcal{V} = (v_1, \dots, v_n)$ be an ordered orthonormal basis for \mathcal{V} consisting of eigenvectors for T with corresponding eigenvalues $\lambda_1, \dots, \lambda_n$, 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 6.12 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 6.8.

(iv) \Rightarrow (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 $T^* = \varphi_{\mathcal{V}}^{-1} \circ M_{A^T} \circ \varphi_{\mathcal{V}}$, but since A is diagonal this is just T , so T is self-adjoint. \square

COROLLARY 6.18

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

- (i) T is self-adjoint if and only if $\text{Spec } T \subseteq \mathbb{R}$.
- (ii) T is unitary if and only if $\text{Spec } T \subseteq S^1$.

PROOF. *Proof of (i):* The ‘only if’ part follows from Proposition 6.14(iv), so assume that $\text{Spec } T \subseteq \mathbb{R}$ and notice that

$$T^* = \left(\sum_{\lambda \in \text{Spec } T} \lambda P_\lambda \right)^* = \sum_{\lambda \in \text{Spec } T} \bar{\lambda} P_\lambda^* = \sum_{\lambda \in \text{Spec } T} \lambda P_\lambda,$$

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

Alternatively, choose a diagonal matrix $A \in \text{Mat}_n(\mathbb{K})$ in accordance with Theorem 6.17(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 6.14(v). Assume that $\text{Spec } T \subseteq S^1$ and notice that

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

Since the projections P_λ are pairwise orthogonal, we have

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

so T 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 $T^*T = \text{id}_V$, so T is unitary. \square

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](https://doi.org/10.1007/978-3-319-11080-6).
- 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). *Linear algebra*. Department of Mathematics, Aarhus University. 251 pp.