Notes on linear algebra

Danny Nygård Hansen

13th 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

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

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 $A \in \mathrm{Mat}_{m,n}(\mathbb{F})$, $x \in \mathbb{F}^n$ 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 1.1

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 $1, \ldots, r$ 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 1.2

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 $H_2 x = 0$ and $K_2 x = 0$ are solution equivalent. If $h' = 0$, then $H_2 x = 0$ has the solution $(-h, 1)$. But this is also a solution to $K_2 x = 0$, so $h = k$ and $k' = 0$. If $h' \neq 0$, then $H_2 x = 0$ only has the trivial solution. But then $K_2 x = 0$ 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 $H_2 = K_2$. □

1.3. *Invertible matrices*

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

LEMMA 1.3

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.

---

[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 $a_1, \dots, a_n \in \mathbb{F}^m$ are the columns in A, then the columns in PA are $Pa_1, \dots, Pa_m$. 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 $H \in \mathrm{Mat}_{m,n}(\mathbb{F})$ 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 1.4

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. □

2 • Coordinates

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

Proposition 2.1

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

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

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 $\alpha_1, \dots, \alpha_n \in \mathbb{F}$. 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 $\mathcal{V} = (v_1, \dots, v_n)$ 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 $\varphi_{\mathcal{V}} \colon V \to \mathbb{F}^n$ 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

$$\begin{tikzcd} & \mathbb{F}^n \arrow[dd, "\varphi_{\mathcal{W},\mathcal{V}}"] \\ V \arrow[ur, "\varphi_{\mathcal{V}}"] \arrow[dr, "\varphi_{\mathcal{W}}"'] & \\ & \mathbb{F}^n \end{tikzcd} \tag{2.1}$$

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

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

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 $\mathcal{W}$. We show below that 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 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

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

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

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

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

(iv) *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 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 ${}_{\mathcal{V}'}[\mathrm{id}_V]_{\mathcal{V}}$ 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). □

PROPOSITION 2.4

Let $\mathcal{V} = (v_1, \dots, v_n)$ be an ordered basis for an $\mathbb{F}$-vector space V, and let $T \colon V \to V$ 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 $A = PBP^{-1}$.

3 • Determinants

3.1. 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

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

for all $i < j$.

LEMMA 3.1

Let M and N be R-modules, and let $\varphi \colon M^n \to N$ 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 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 $\psi \colon M \times M \to N$ 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 3.2: *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 3.3: *Construction of determinants*

Let $n > 1$, and let $\varphi\colon \mathrm{Mat}_{n-1}(R) \to R$ be alternating and $(n-1)$-linear. For $j = 1,\ldots,n$ 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 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. □

Corollary 3.4: *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 3.3 allows us to recursively construct a determinant for each $n > 1$. □

3.2. Uniqueness of determinants

Theorem 3.5: *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 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

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

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. □

3.3. Properties of determinants

THEOREM 3.6

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 3.7

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 3.6 then implies that

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

□

Corollary 3.7 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 3.8

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 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. □

Proposition 3.9

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. $\square$

Let $A \in \mathrm{Mat}_n(R)$. For $1 \le i, j \le n$, 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 *adjoint matrix* of A is the matrix $\operatorname{adj} A \in \mathrm{Mat}_n(R)$ whose (i,j)-th entry is the cofactor $A_{j,i}$. 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 $\operatorname{adj} A^\top = (\operatorname{adj} A)^\top$. We have the following:

Proposition 3.10

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 $j \in \{1, \ldots, n\}$, 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 $(\operatorname{adj} A)A$.

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 3.11

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 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 $\operatorname{Spec} T$. 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 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* $-\operatorname{tr} T$.

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

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

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

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

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

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

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

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 $\mathrm{ev}_T \colon \mathbb{F}[t]_{n^2+1} \to \mathcal{L}(V)$ 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 $c,\lambda_1,\dots,\lambda_m \in \mathbb{F}$ 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 $\operatorname{tr} T$.

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 • Triangularisation and diagonalisation

4.1. 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 4.1

Let V be an $\mathbb{F}$-vector space with $n = \dim V < \infty$, and let $\mathcal{V} = (v_1, \ldots, v_n)$ 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 $\mathrm{span}(v_1, \ldots, v_i)$ *is invariant under T for all $i \in \{1, \ldots, n\}$.*

Proof. This is obvious. □

Lemma 4.2

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 $T|_U \colon U \to U$ 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 4.3

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 4.4

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

$$Ae_i = a_1e_1 + \cdots + a_{i-1}e_{i-1} + \lambda_i e_i$$

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

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

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

LEMMA 4.5

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 4.4, 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 4.6

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 $U = \mathrm{span}(v_1)$ 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

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

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

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

[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 $\tilde{T}\colon V/U \to V/U$.

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 4.7: *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 4.6 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 4.1 shows that ${}_{\mathcal{U}}[T]_{\mathcal{U}}$ is also upper triangular, proving the claim. □

4.2. *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

$$\langle T^*w, v\rangle_V = \langle w, Tv\rangle_W \tag{4.1}$$

for all $v \in V$ and $w \in W$. There first of all exists such $L^*w \in V$ since, if $(v_1,\dots,v_n)$ is an orthonormal basis for V, then

$$\begin{aligned}\langle w, Tv\rangle_W &= \Big\langle w, \sum_{i=1}^n \langle v, v_i\rangle_V Tv_i \Big\rangle_W \\ &= \sum_{i=1}^n \langle v_i, v\rangle_V \langle w, Tv_i\rangle_W \\ &= \Big\langle \sum_{i=1}^n \langle w, Tv_i\rangle_W v_i, v \Big\rangle_V.\end{aligned}$$

Hence we may choose $T^*w = \sum_{i=1}^n \langle w, Tv_i\rangle_W v_i$. Furthermore, this vector is unique since (4.1) implies that

$$T^*w = \sum_{i=1}^n \langle T^*w, v_i\rangle_V = \sum_{i=1}^n \langle w, Tv_i\rangle_W.$$

In particular, T^*w does not depend on a choice of basis for V. Notice that taking complex conjugates in (4.1) we find that $T^{**} = T$.

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 a self-adjoint operator is normal.

PROPOSITION 4.8

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

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

(ii) *If* $\lambda \in \mathbb{K}$ *is an eigenvalue of* T, *then* $\overline{\lambda}$ *is an eigenvalue of* T^* *with the same eigenvectors.*

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

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

PROOF. *(i)*: Notice that

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

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

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

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

(iii): Since w is an eigenvector for T^* with eigenvalue $\overline{\mu}$ we have

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

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

(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. If instead $\mathbb{K} = \mathbb{R}$, then let $\mathcal{V}$ be an orthonormal basis for V and consider the matrix representation $A = {}_{\mathcal{V}}[T]_{\mathcal{V}}$. This is a real symmetric matrix, in particular a matrix with *complex* entries, i.e. an operator $\mathbb{C}^n \to \mathbb{C}^n$. Hence it has a complex eigenvalue λ, which is real by the above. This means that $\lambda I - A$ is singular when considered as an operator $\mathbb{C}^n \to \mathbb{C}^n$. But then it is clearly singular as an operator $\mathbb{R}^n \to \mathbb{R}^n$, so λ is an eigenvalue of T. ◻

Let $T : V \to V$ 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 V, i.e. $V = U \oplus W$, then W is not necessarily invariant under T. However, we have the following:

LEMMA 4.9

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. □

THEOREM 4.10: *The spectral theorem*

Let $\mathbb{F}$ be either the real or the complex numbers, let V be a finite-dimensional inner product space over $\mathbb{F}$, and consider $T \in \mathcal{L}(V)$. Then T is orthogonally diagonalisable if and only if

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

(ii) $\mathbb{F} = \mathbb{C}$ *and T is normal.*

PROOF. Assume that either $\mathbb{F} = \mathbb{R}$ and T is self-adjoint, or that $\mathbb{F} = \mathbb{C}$ 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 $0 < d < n$.

Clearly $E_T(\lambda)$ is invariant under T, and we claim that it is also invariant under T^*. If T is self-adjoint this is obvious, and if T is normal then for all $w \in E_T(\lambda)$,

$$TT^* w = T^* T w = T^*(\lambda w) = \lambda T^* w,$$

so we also have $T^*(w) \in E_T(\lambda)$. It follows from Lemma 4.9 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 these operators 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.

Conversely, assume that $\mathcal{V}$ is an orthonormal basis of V consisting of eigenvectors for T, and let $A \in \mathrm{Mat}_n(\mathbb{F})$ be the matrix of T with respect to $\mathcal{V}$. Then A is diagonal with the eigenvalues of T on its diagonal. If $\mathbb{F} = \mathbb{R}$ then the eigenvalues of T are real, so A is a real symmetric matrix, and hence T

is self-adjoint. If instead $\mathbb{F} = \mathbb{C}$, then since A is diagonal we have $A^*A = AA^*$, which implies that T is normal. □

5 • Complex numbers

It is well-known that a complex number $z = a + \mathrm{i}b$ has a representation as a matrix

$$A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

and that the subring of $\mathrm{Mat}_2(\mathbb{R})$ consisting of such matrices is isomorphic to $\mathbb{C}$. Letting $r = |z| = \sqrt{\det A}$ we obtain a matrix $Q = A/r \in \mathrm{SO}(2)$. Let us call the pair (r, Q) the *geometric representation* of z.

Let $\mathbb{C}^*$ denote the group of nonzero complex numbers under multiplication. We define an action of $\mathbb{C}^*$ on $\mathbb{R}^2$ as follows: If $v \in \mathbb{R}^2$ then, in the notation above, we let $zv = rQv$; that is, z acts on v by applying the rotation matrix Q and scaling by r.

Alternatively, given $v = (x, y) \in \mathbb{R}^2$ form the complex number $w = x + \mathrm{i}y$ with corresponding matrix

$$B = \begin{pmatrix} x & -y \\ y & x \end{pmatrix}.$$

Then zw has the corresponding matrix rQB, the first column of which is $zv = rQv$. Thus the action of $\mathbb{C}^*$ on $\mathbb{R}^2$ is also obtained by considering a vector in $\mathbb{R}^2$ as a complex number and performing complex multiplication.

Lemma 5.1

The action of $\mathbb{C}^*$ on $\mathbb{R}^2$ preserves angles.

Proof. Let $z \in \mathbb{C}^*$ have have the geometric representation (r, Q), and let $v, u \in \mathbb{R}^2$. Then notice that

$$\langle zv, zu \rangle = r^2 \langle Qv, Qu \rangle = r^2 \langle v, u \rangle,$$

since Q is orthogonal. In particular we have $\|zv\| = r\|v\|$. If $\theta \in [0, \pi]$ is the angle between zv and zu, then the Cauchy–Schwarz inequality implies that

$$\cos\theta = \frac{\langle zv, zu \rangle}{\|zv\|\,\|zu\|} = \frac{r^2\langle v, u \rangle}{r^2\|v\|\,\|u\|} = \frac{\langle v, u \rangle}{\|v\|\,\|u\|},$$

which is just the cosine of the angle between v and u. This proves the lemma. □

Now let $U \subseteq \mathbb{C}$ be a nonempty open set, and let $f: U \to \mathbb{C}$ be a holomorphic function that does not attain the value zero.[3] Considering U and $\mathbb{C}$ as real two-dimensional manifolds, let $T_p f: T_p U \to T_{f(p)}\mathbb{C}$ be the tangent map of f at $p \in U$. The Jacobian matrix of f at p is then simply the matrix corresponding to the complex number $f'(p)$, so if $v \in T_p U$, then the vector $T_p f(v) \in T_{f(p)}\mathbb{C} \cong \mathbb{R}^2$ is just the action of $f'(p)$ on v. The lemma then implies that, for $v, u \in T_p U$,

$$\langle T_p f(v), T_p f(u)\rangle = \langle f'(p)v, f'(p)u\rangle = |f'(p)|^2 \langle v, u\rangle.$$

Since f is holomorphic it is smooth as a function on $\mathbb{R}^2$, the map $p \mapsto |f'(p)|^2$ is also smooth and nonzero everywhere, and so f is conformal.

6 • Gray codes

[This doesn't belong here, I just needed a LaTeX editor to write the proof.]

If a and b are binary strings of the same length, we denote the bitwise exclusive disjunction of a and b by $a \oplus b$. We denote the concatenation of a with b either by $a \circ b$ or ab. Also, if b is a binary string, denote by $b^{\gg}$ the right logical shift of b, i.e. the string obtained by removing the rightmost bit of b and appending a 0 on the left of the result.

Let $n \in \mathbb{N}$. For a number $k \in \mathbb{N}$ with $k < 2^n$ we denote the n-bit binary representation of k by $\mathrm{bin}_n(k)$. Furthermore, we denote the n-bit Gray code for k by $\mathrm{gr}_n(k)$. By definition, $\mathrm{gr}_0(0) = \lambda$ and

$$\mathrm{gr}_{n+1}(k) = \begin{cases} 0 \circ \mathrm{gr}_n(k), & k < 2^n, \\ 1 \circ \mathrm{gr}_n(2^{n+1} - 1 - k), & k \geq 2^n. \end{cases}$$

for all $n \in \mathbb{N}$ and $(n+1)$-bit numbers k. We claim the following:

Proposition 6.1

Let $n \in \mathbb{N}$, and let $k \in \mathbb{N}$ be an n-bit number. Writing $\mathrm{bin}_n(k) = b_{n-1} \cdots b_0$ we have $\mathrm{gr}_n(k) = a_{n-1} \cdots a_0$, where $a_{n-1} = b_{n-1}$ and

$$a_i = b_{i+1} \oplus b_i \tag{6.1}$$

for $i \in \{0, \ldots, n-2\}$. That is,

$$\mathrm{gr}_n(k) = b_{n-1}(b_{n-1} \oplus b_{n-2}) \cdots (b_1 \oplus b_0).$$

[3] If f is not identically zero, then $f^{-1}(\mathbb{C}^*)$ is a nonempty open subset of $\mathbb{C}$, so this is a very natural assumption.

Conversely we have

$$b_i = a_i \oplus \cdots \oplus a_{n-1}.$$

The formula (6.1) also holds in the case $i = n-1$ if we let $b_n = 0$, i.e. we prepend zeros if necessary.

PROOF. If $n = 0$, then the claim is obvious since there are no 0-bit numbers. Now let k be an $(n+1)$-bit number, so that $k < 2^{n+1}$, and write $\mathrm{bin}_{n+1}(k) = b_n \cdots b_0$. If $k < 2^n$, then $b_n = 0$ and $\mathrm{gr}_{n+1}(k) = 0 \circ \mathrm{gr}_n(k)$. By induction we have

$$\begin{aligned}\mathrm{gr}_n(k) &= b_{n-1}(b_{n-1} \oplus b_{n-2}) \cdots (b_1 \oplus b_0) \\ &= (b_n \oplus b_{n-1})(b_{n-1} \oplus b_{n-2}) \cdots (b_1 \oplus b_0),\end{aligned}$$

so it follows that

$$\mathrm{gr}_{n+1}(k) = b_n \circ \mathrm{gr}_n(k) = b_n(b_n \oplus b_{n-1})(b_{n-1} \oplus b_{n-2}) \cdots (b_1 \oplus b_0)$$

as claimed. If instead $k \geq 2^n$, then $b_n = 1$. Writing $k = 2^n + r$ with $0 \leq r < 2^n$ we have $\mathrm{bin}_n(r) = b_{n-1} \cdots b_0$. Now notice that $\mathrm{bin}_n(2^n - 1 - r) = \bar{b}_{n-1} \cdots \bar{b}_0$ since

$$(\bar{b}_{n-1} \cdots \bar{b}_0)_2 + r + 1 = (\bar{b}_{n-1} \cdots \bar{b}_0)_2 + (b_{n-1} \cdots b_0)_2 + 1 = 2^n.$$

By induction we have

$$\begin{aligned}\mathrm{gr}_n(2^n - 1 - r) &= \bar{b}_{n-1}(\bar{b}_{n-1} \oplus \bar{b}_{n-2}) \cdots (\bar{b}_1 \oplus \bar{b}_0) \\ &= (b_n \oplus b_{n-1})(b_{n-1} \oplus b_{n-2}) \cdots (b_1 \oplus b_0)\end{aligned}$$

since $b_n = 1$, so it follows that

$$\begin{aligned}\mathrm{gr}_{n+1}(k) &= b_n \circ \mathrm{gr}_n(2^n - 1 - r) \\ &= b_n(b_n \oplus b_{n-1})(b_{n-1} \oplus b_{n-2}) \cdots (b_1 \oplus b_0)\end{aligned}$$

as desired.

For the final claim, simply notice that

$$\begin{aligned}a_i \oplus \cdots \oplus a_{n-1} &= (b_i \oplus b_{i+1}) \oplus (b_{i+1} \oplus b_{i+2}) \oplus \cdots \oplus (b_{n-2} \oplus b_{n-1}) \oplus b_{n-1} \\ &= b_i \oplus (b_{i+1} \oplus b_{i+1}) \oplus (b_{i+2} \oplus \cdots \oplus b_{n-2}) \oplus (b_{n-1} \oplus b_{n-1}) \\ &= b_i.\end{aligned}$$

Alternatively we may notice that (6.1) defines a linear system of equations with coefficients in $\mathbb{Z}/2\mathbb{Z}$ and invert this. □

Corollary 6.2

For $n \in \mathbb{N}$ and any n-bit number k, we have

$$\mathrm{gr}_n(k) = \mathrm{bin}_n(k) \oplus \mathrm{bin}_n(k)^{\gg}.$$

Proof. Writing $\mathrm{bin}_n(k) = b_{n-1} \cdots b_0$, the proposition implies that

$$\begin{aligned}\mathrm{gr}_n(k) &= b_{n-1}(b_{n-1} \oplus b_{n-2}) \cdots (b_1 \oplus b_0) \\ &= (0 \oplus b_{n-1})(b_{n-1} \oplus b_{n-2}) \cdots (b_1 \oplus b_0).\end{aligned}$$

But $\mathrm{bin}_n(k)^{\gg} = 0b_{n-1} \cdots b_1$, so the claim follows. □

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.

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

Knuth, Donald E. (2011). *The Art of Programming, Volume 4A: Combinatorial Algorithms, Part 1*. 1st ed. Addison-Wesley. 883 pp. ISBN: 978-0-201-03804-0.

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