

LINEAR ALGEBRA II

DRAFT VERSION

SERGEY MOZGOVOY

CONTENTS

1. Vector spaces and linear maps	2
1.1. Recollection from LA I	2
1.2. Matrix associated with a linear map	3
1.3. Base change	4
1.4. Rank and nullity	5
1.5. Sums and direct sums	7
1.6. Intersection of subspaces	9
1.7. Quotient space	10
2. Linear operators	12
2.1. Linear operators	12
2.2. Invariant subspaces and eigenvectors	13
2.3. Diagonalizable operators	15
2.4. Polynomials of matrices and linear operators	17
2.5. Jordan matrices	19
2.6. Generalized eigenvectors	20
2.7. Nilpotent operators	22
2.8. Examples	24
2.9. Applications	25
3. Inner products	27
3.1. Dual spaces	27
3.2. Bilinear forms	30
3.3. Euclidean vector spaces	32
3.4. Unitary vector spaces	35
3.5. Orthogonal complement	36
3.6. Self-adjoint operators	37
3.7. Orthogonal operators and matrices	39
3.8. Quadratic forms	40
3.9. Positive definiteness	42
3.10. Applications	43

1. VECTOR SPACES AND LINEAR MAPS

1.1. Recollection from LA I.

1.1.1. *Vector spaces.* We will consider (finite-dimensional) vector spaces over $K = \mathbb{R}$ or $K = \mathbb{C}$ (or over any other field K). Elements of K are called scalars. Recall that a vector space V is a set equipped with operations

- (1) vector addition $V \times V \rightarrow V$, $(u, v) \mapsto u + v$,
- (2) scalar multiplication $K \times V \rightarrow V$, $(\lambda, v) \mapsto \lambda v$,

that satisfy certain axioms. A subset $0 \in U \subset V$ is called a **subspace** (vector subspace, linear subspace) if, under the operations of V , it is a vector space. Equivalently, if

- (1) $u + v \in U$ for all $u, v \in U$.
- (2) $\lambda u \in U$ for all $\lambda \in K$, $u \in U$.

Given a vector space V and a set of vectors $S = \{v_1, \dots, v_r\}$ in V , we define a **span** (linear span, vector space **generated** by S)

$$\text{span}(S) = \text{span}(v_1, \dots, v_r) = \left\{ \sum_i \lambda_i v_i \mid \lambda_i \in K \right\} \subset V.$$

It is the minimal subspace of V that contains vectors $v_1, \dots, v_r$. We can similarly define $\text{span}(S)$ for an arbitrary set $S \subset V$.

1.1.2. *Linear maps.* A map $\phi: U \rightarrow V$ between two vector spaces is called a **linear map** if

- (1) $\phi(u + v) = \phi(u) + \phi(v)$ for all $u, v \in U$.
- (2) $\phi(cu) = c\phi(u)$ for any scalar $c \in K$ and any vector $u \in U$.

Given linear maps $\phi: U \rightarrow V$ and $\psi: U \rightarrow V$, we define a new linear map

$$\phi + \psi: U \rightarrow V, \quad (\phi + \psi)(u) = \phi(u) + \psi(u), \quad u \in U.$$

For any scalar $c \in K$, define a linear map

$$c\phi: U \rightarrow V, \quad (c\phi)(u) = c \cdot \phi(u), \quad u \in U.$$

Remark 1.1. A bijective linear map $\phi: U \rightarrow V$ is called an **isomorphism**. In this case the vector spaces U, V are said to be **isomorphic** and we write $U \simeq V$. A surjective linear map is called an **epimorphism**. An injective linear map is called a **monomorphism**.

1.1.3. *Linear map associated with a matrix.* Given an $m \times n$ matrix

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

we define the linear map associated with it

$$A = L_A: K^n \rightarrow K^m, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_j a_{1j}x_j \\ \vdots \\ \sum_j a_{mj}x_j \end{pmatrix}.$$

Let $\mathbf{e} = (e_1, \dots, e_n)$ be the standard basis of K^n and $\mathbf{f} = (f_1, \dots, f_m)$ be the standard basis of K^m . Then

$$L_A(e_j) = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix} = \sum_i a_{ij} f_i. \quad (1)$$

1.2. Matrix associated with a linear map. Let U be a vector space with a basis $\mathbf{e} = (e_1, \dots, e_n)$, V be a vector space with a basis $\mathbf{f} = (f_1, \dots, f_m)$ and $\phi: U \rightarrow V$ be a linear map. Using formula (1) as a motivation, we define the $m \times n$ matrix associated with ϕ

$$[\phi]_{\mathbf{e}, \mathbf{f}} = A = (a_{ij}), \quad \phi(e_j) = \sum_i a_{ij} f_i.$$

This means that the j -th column of the matrix A is given by the coordinates of the vector $\phi(e_j)$ with respect to the basis $\mathbf{f}$. Note that the matrix A depends on the choice of the bases $\mathbf{e}$ and $\mathbf{f}$.

Given a basis $\mathbf{e} = (e_1, \dots, e_n)$ of U and a vector $u \in U$, define the **coordinate vector** of u

$$[u]_{\mathbf{e}} = (x_1, \dots, x_n)^T \in K^n, \quad u = x_1 e_1 + \dots + x_n e_n.$$

Lemma 1.2. *Let $\phi: U \rightarrow V$ be a linear map, $u \in U$ and $\mathbf{e}, \mathbf{f}$ be bases of U, V respectively. Then*

$$[\phi(u)]_{\mathbf{f}} = [\phi]_{\mathbf{e}, \mathbf{f}} \cdot [u]_{\mathbf{e}}.$$

Proof. Let $A = [\phi]_{\mathbf{e}, \mathbf{f}}$ and $x = [u]_{\mathbf{e}} \in K^n$. Then

$$\phi(u) = \phi\left(\sum_j x_j e_j\right) = \sum_j x_j \phi(e_j) = \sum_j x_j \sum_i a_{ij} f_i = \sum_i \left(\sum_j a_{ij} x_j\right) f_i = \sum_i (Ax)_i f_i.$$

This implies $[\phi(u)]_{\mathbf{f}} = Ax$. □

The following result is the statement that composition of linear maps corresponds to the multiplication of matrices.

Proposition 1.3 (Composition). *Let $\phi: U \rightarrow V$ and $\psi: V \rightarrow W$ be linear maps and let $\mathbf{e}, \mathbf{f}, \mathbf{f}'$ be bases of U, V, W respectively. Then*

$$[\psi \circ \phi]_{\mathbf{e}, \mathbf{f}'} = [\psi]_{\mathbf{f}, \mathbf{f}'} \cdot [\phi]_{\mathbf{e}, \mathbf{f}},$$

where $\psi \circ \phi: U \rightarrow W$ is the composition map, defined by $(\psi \circ \phi)(u) = \psi(\phi(u))$, for $u \in U$.

Proof. Let $A = [\phi]_{\mathbf{e}, \mathbf{f}}$, $B = [\psi]_{\mathbf{f}, \mathbf{f}'}$ and $C = [\psi \circ \phi]_{\mathbf{e}, \mathbf{f}'}$. Then

$$\sum_i c_{ij} f'_i = \psi \phi(e_j) = \psi \left(\sum_k a_{kj} f_k \right) = \sum_k a_{kj} \sum_i b_{ik} f'_i = \sum_i \left(\sum_k b_{ik} a_{kj} \right) f'_i.$$

Taking the coefficient of the basis vector f'_i , we obtain

$$c_{ij} = \sum_k b_{ik} a_{kj} = (BA)_{ij}.$$

Therefore $C = BA$. □

1.3. Base change. Let V be a vector space and $\mathbf{e}, \mathbf{e}'$ be two bases of V . Define the **transition matrix**

$$M_{\mathbf{e}, \mathbf{e}'} = M_{\mathbf{e} \rightarrow \mathbf{e}'} = (m_{ij}) = [\text{id}]_{\mathbf{e}, \mathbf{e}'},$$

where $\text{id}: V \rightarrow V, x \mapsto x$, is the identity map. Explicitly, this means

$$e_j = \text{id}(e_j) = \sum_i m_{ij} e'_i,$$

hence the j -th column of $M_{\mathbf{e} \rightarrow \mathbf{e}'}$ is given by the coordinates of e_j with respect to the basis $\mathbf{e}'$.

Lemma 1.4. *Given two bases $\mathbf{e}, \mathbf{e}'$ of a vector space V and a vector $v \in V$, we have*

$$[v]_{\mathbf{e}'} = M_{\mathbf{e} \rightarrow \mathbf{e}'} \cdot [v]_{\mathbf{e}}.$$

Proof. We obtain from Lemma 1.2 $[v]_{\mathbf{e}'} = [\text{id}]_{\mathbf{e}, \mathbf{e}'} \cdot [v]_{\mathbf{e}} = M_{\mathbf{e} \rightarrow \mathbf{e}'} \cdot [v]_{\mathbf{e}}$. □

Proposition 1.5. *We have*

- (1) $M_{\mathbf{e} \rightarrow \mathbf{e}} = I$ (the identity matrix).
- (2) $M_{\mathbf{e}' \rightarrow \mathbf{e}} = M_{\mathbf{e} \rightarrow \mathbf{e}'}^{-1}$ (the inverse matrix).

Proof. (1) The entries of $M_{\mathbf{e} \rightarrow \mathbf{e}} = (m_{ij})$ satisfy $e_j = \sum_i m_{ij} e_i$, hence $m_{ij} = \delta_{ij}$.

(2) We have (by applying Proposition 1.3)

$$M_{\mathbf{e}' \rightarrow \mathbf{e}} \cdot M_{\mathbf{e} \rightarrow \mathbf{e}'} = [\text{id}]_{\mathbf{e}', \mathbf{e}} \cdot [\text{id}]_{\mathbf{e}, \mathbf{e}'} = [\text{id} \circ \text{id}]_{\mathbf{e}, \mathbf{e}} = [\text{id}]_{\mathbf{e}, \mathbf{e}} = I,$$

hence the statement. □

Transition matrices allow us to express matrices of linear maps with respect to different bases.

Proposition 1.6. *Let $\phi: U \rightarrow V$ be a linear map and let $\mathbf{e}, \mathbf{e}'$ be two bases of U and $\mathbf{f}, \mathbf{f}'$ be two bases of V . Then*

$$[\phi]_{\mathbf{e}', \mathbf{f}'} = M_{\mathbf{f} \rightarrow \mathbf{f}'} \cdot [\phi]_{\mathbf{e}, \mathbf{f}} \cdot M_{\mathbf{e}' \rightarrow \mathbf{e}}.$$

Proof. Applying Proposition 1.3, we obtain

$$M_{\mathbf{f} \rightarrow \mathbf{f}'} \cdot [\phi]_{\mathbf{e}, \mathbf{f}} \cdot M_{\mathbf{e}' \rightarrow \mathbf{e}} = [\text{id}]_{\mathbf{f}, \mathbf{f}'} \cdot [\phi]_{\mathbf{e}, \mathbf{f}} \cdot \text{id}_{\mathbf{e}', \mathbf{e}} = [\text{id} \circ \phi \circ \text{id}]_{\mathbf{e}', \mathbf{f}'} = [\phi]_{\mathbf{e}', \mathbf{f}'}.$$

□

1.3.1. Elementary operations and base change. Let $\phi: U \rightarrow V$ be a linear map and

$$A = (a_{ij}) = [\phi]_{\mathbf{e}, \mathbf{f}}, \quad \phi(e_j) = \sum_i a_{ij} f_i$$

be its matrix with respect to a basis $\mathbf{e}$ of U and a basis $\mathbf{f}$ of V . Consider a new basis

$$\mathbf{f}' = (f_1 - \lambda f_2, f_2, f_3, \dots), \quad \lambda \in K,$$

of V and the corresponding matrix $B = (b_{ij}) = [\phi]_{\mathbf{e}, \mathbf{f}'}$. Then

$$\phi(e_j) = \sum_i b_{ij} f'_i = b_{1j}(f_1 - \lambda f_2) + b_{2j}f_2 + \dots = b_{1j}f_1 + (b_{2j} - \lambda b_{1j})f_2 + \dots$$

This implies that for all j

$$b_{1j} = a_{1j} \quad \forall j \neq 2, \quad b_{2j} = a_{2j} + \lambda a_{1j}.$$

Therefore the matrix B is obtained from A by an elementary row operation, where the first row is multiplied by λ and added to the second row. Similarly, every elementary row operation (as well as their composition) corresponds to a base change on the vector space V . In the same way, every elementary column operation corresponds to a base change on the vector space U .

In particular, performing Gauss elimination and reducing the matrix A to a row echelon form corresponds to a choice of a different basis of V .

1.4. Rank and nullity.

Definition 1.7. Let $\phi: U \rightarrow V$ be a linear map. Define

- (1) The **kernel** (or the **null space**) of ϕ to be $\text{Ker } \phi = \{u \in U \mid \phi(u) = 0\} \subset U$.
- (2) The **image** of ϕ to be $\text{Im } \phi = \{\phi(u) \mid u \in U\} \subset V$.

Lemma 1.8. Given a linear map $\phi: U \rightarrow V$, we have

- (1) $\text{Ker } \phi \subset U$ is a subspace of U .
- (2) $\text{Im } \phi \subset V$ is a subspace of V .

Proof. (1) If $u, v \in \text{Ker } \phi$, then $\phi(u) = \phi(v) = 0 \implies \phi(u + v) = \phi(u) + \phi(v) = 0 \implies u + v \in \text{Ker } \phi$. For any scalar $\lambda \in K$ and $u \in \text{Ker } \phi$, we have $\phi(\lambda u) = \lambda\phi(u) = 0$, hence $\lambda u \in \text{Ker } \phi$. These properties imply that $\text{Ker } \phi \subset U$ is a subspace. $\square$

Lemma 1.9. Let $\phi: U \rightarrow V$ be a linear map. Then ϕ is injective $\iff \text{Ker } \phi = 0$.

Proof. If ϕ is injective and $u \in \text{Ker } \phi$, then $\phi(u) = \phi(0) = 0$, hence $u = 0$. This implies that $\text{Ker } \phi = 0$. Conversely, let $\text{Ker } \phi = 0$ and $u, v \in U$ be such that $\phi(u) = \phi(v)$. Then $\phi(u - v) = \phi(u) - \phi(v) = 0$, hence $u - v \in \text{Ker } \phi$ and $u - v = 0$. This implies that $u = v$, hence ϕ is injective. $\square$

Because of the above results we can consider dimensions of vector spaces $\text{Ker } \phi$ and $\text{Im } \phi$.

Definition 1.10. Let $\phi: U \rightarrow V$ be a linear map. Define

- (1) The **rank** of ϕ to be $\text{rk}(\phi) = \dim(\text{Im } \phi)$.
- (2) The **nullity** of ϕ to be $\text{null}(\phi) = \dim(\text{Ker } \phi)$.

Remark 1.11.

- (1) Given an $m \times n$ matrix A , consider the corresponding linear map

$$A = L_A: K^n \rightarrow K^m, \quad x \mapsto A \cdot x$$

and define

$$\text{rk } A = \text{rk } L_A, \quad \text{null } A = \text{null } L_A.$$

- (2) The kernel $\text{Ker } L_A$ is equal to the set of solutions of the system of linear equations

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = 0 \\ \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = 0 \end{cases}$$

corresponding to the matrix A . This set is also called the **null space** of A . We conclude that $\text{null } A$ is equal to the dimension of the null space of A .

- (3) The image $\text{Im } L_A$ is equal to the **column space** of A , generated by the columns of A . Therefore $\text{rk } A$ is equal to the dimension of the column space of A , also called the **column rank** of A . It is equal to the maximal number of linearly independent columns of A . Note that the rank of the linear map L_A is independent of a choice of basis.
- (4) Similarly, define the **row space** of A to be the space generated by the rows of A and define the **row rank** of A to be the dimension of this space. It is equal to the maximal number of linearly independent rows of A . Applying transposition, we can transform rows into columns. Therefore the row rank of A is equal to the column rank $\text{rk } A^T$ of the transposed matrix A^T . We will see that the column and the row ranks of a matrix are equal.

Example 1.12. Consider the matrix $A = \begin{pmatrix} 1 & 2 & 1 & 1 & 1 \\ 2 & 5 & 0 & 1 & 0 \end{pmatrix}$. Applying elementary row operations, we obtain

$$\begin{pmatrix} 1 & 2 & 1 & 1 & 1 \\ 2 & 5 & 0 & 1 & 0 \end{pmatrix} \xrightarrow{r_2 - 2r_1} \begin{pmatrix} 1 & 2 & 1 & 1 & 1 \\ 0 & 1 & -2 & -1 & -2 \end{pmatrix} \xrightarrow{r_1 - 2r_2} \begin{pmatrix} 1 & 0 & 5 & 3 & 5 \\ 0 & 1 & -2 & -1 & -2 \end{pmatrix}.$$

This means that we have free variables x_3, x_4, x_5 and pivot variables

$$x_2 = 2x_3 + x_4 + 2x_5, \quad x_1 = -5x_3 - 3x_4 - 5x_5.$$

The null space $\text{Ker } A$ consists of the vectors

$$\begin{pmatrix} -5x_3 - 3x_4 - 5x_5 \\ 2x_3 + x_4 + 2x_5 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = x_3 \begin{pmatrix} -5 \\ 2 \\ 1 \\ 0 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -3 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_5 \begin{pmatrix} -5 \\ 2 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

hence $\text{Ker } A$ is the span of the above three vectors and $\text{null } A = \dim(\text{Ker } A) = 3$. On the other hand, the linear map $A: K^5 \rightarrow K^2$ is surjective, hence $\text{Im } A = K^2$ and $\text{rk } A = 2$. Note that $\text{rk } A + \text{null } A = 5$, the dimension of the domain of A . We will see that this is true in general.

Theorem 1.13 (Rank-nullity theorem). *Let $\phi: U \rightarrow V$ be a linear map. Then*

$$\text{rk}(\phi) + \text{null}(\phi) = \dim U.$$

Proof. Choosing bases $\mathbf{e} = (e_1, \dots, e_n)$, $\mathbf{f} = (f_1, \dots, f_m)$ of U and V respectively, we can represent ϕ as an $m \times n$ matrix $A = [\phi]_{\mathbf{e}, \mathbf{f}}$ so that $\text{rk } A = \text{rk } \phi$ and $\text{null } A = \text{null } \phi$. We know that elementary row and column operations on A correspond to base changes in V and U . These base changes don't affect $\text{rk } \phi$ and $\text{null } \phi$, hence new matrices will have the same rank and nullity as A . We can transform the matrix A to the form

$$B = \begin{pmatrix} I_k & 0_{k, n-k} \\ 0_{m-k, k} & 0_{m-k, n-k} \end{pmatrix}.$$

This matrix has k pivots and $n - k$ free variables, hence $\text{rk } \phi = \text{rk } B = k$, $\text{null } \phi = \text{null } B = n - k$ and $\text{rk } \phi + \text{null } \phi = k + (n - k) = n = \dim U$. $\square$

Lemma 1.14. *For any linear map $\phi: U \rightarrow V$, we have $\text{rk } \phi \leq \min \{\dim U, \dim V\}$.*

Proof. We have $\text{rk } \phi \leq \dim V$ as $\text{Im } \phi \subset V$. On the other hand $\text{rk } \phi + \text{null } \phi = \dim U$, hence $\text{rk } \phi \leq \dim U$. $\square$

Theorem 1.15. *The column and the row ranks of a matrix are equal.*

Proof. Let A be an $m \times n$ matrix. The column rank of A is equal to $\text{rk } A$ and the row rank of A is equal to $\text{rk } A^\top$, the rank of the transposed matrix. We have seen in the proof of the rank-nullity theorem that the rank of A is unchanged under elementary row and column operations. The same applies to $\text{rk } A^\top$. We can transform the matrix A to the form $B = \begin{pmatrix} I_k & 0_{k, n-k} \\ 0_{m-k, k} & 0_{m-k, n-k} \end{pmatrix}$. Then $\text{rk } A = \text{rk } B = k$, $\text{rk } A^\top = \text{rk } B^\top = k$. Therefore $\text{rk } A = \text{rk } A^\top$. $\square$

Lemma 1.16. *Let $\phi: U \rightarrow V$ be an injective (or surjective) linear map and $\dim U = \dim V$. Then ϕ is an isomorphism.*

Proof. If ϕ is injective, then $\text{Ker } \phi = 0$ and $\text{null } \phi = 0$. Therefore by the rank-nullity theorem $\text{rk } \phi = \dim U = \dim V$, hence $\text{Im } \phi \subset V$ has the same dimension as V . This implies that $\text{Im } \phi = V$, hence ϕ is surjective. Therefore ϕ is an isomorphism. $\square$

1.5. Sums and direct sums.

Definition 1.17. Let V be a vector space and $V_1, \dots, V_r$ be its subspaces. The **sum** $\sum_i V_i = V_1 + \dots + V_r \subset V$ is defined to be the set of all vectors of the form $v_1 + \dots + v_r$, where $v_i \in V$ for $1 \leq i \leq r$. It is said to be a **direct sum** if $0 + \dots + 0$ is the only way to write $0 \in V$ as a sum $v_1 + \dots + v_r$ with $v_i \in V_i$. In this case it is denoted by $\bigoplus_i V_i = V_1 \oplus \dots \oplus V_r$.

Lemma 1.18. Let $V_1, \dots, V_r \subset V$ be subspaces. Then $V_1 + \dots + V_r \subset V$ is a subspace.

Example 1.19. Let $v_1, \dots, v_r \in V$ and let $V_i = \text{span}(v_i) = Kv_i = \{\lambda v_i \mid \lambda \in K\}$. Then

$$V_1 + \dots + V_r = \{\lambda_1 v_1 + \dots + \lambda_r v_r \mid \lambda_i \in K\} = \text{span}(v_1, \dots, v_r).$$

Example 1.20. Let $V = K^{m+n}$, $V_1 = \text{span}(e_1, \dots, e_m)$, $V_2 = \text{span}(e_{m+1}, \dots, e_{m+n})$. Then $V = V_1 \oplus V_2$.

Remark 1.21. Let $V_1, V_2 \subset V$ be subspaces of a vector space. Then $V_1 + V_2$ is a direct sum $\iff V_1 \cap V_2 = 0$. Indeed, if $V_1 + V_2$ is not a direct sum, then we have $v_1 + v_2 = 0$ for some $0 \neq v_i \in V_i$. Therefore $v_1 = -v_2$ is contained in $V_1 \cap V_2$, hence $V_1 \cap V_2 \neq 0$. The converse is similar. On the other hand, for $r \geq 3$, there exist subspaces $V_1, \dots, V_r \subset V$ such that $V_i \cap V_j = 0$ for $i \neq j$, while $V_1 + \dots + V_r$ is not a direct sum. For example, consider $V = K^2$, $V_1 = Ke_1$, $V_2 = Ke_2$, $V_3 = K(e_1 + e_2)$.

Lemma 1.22. Let $V_1, \dots, V_r \subset V$ be subspaces. Then the following are equivalent

- (1) $V_1 + \dots + V_r$ is a direct sum.
- (2) Every vector in $V_1 + \dots + V_r$ can be uniquely written in the form $v_1 + \dots + v_r$ with $v_i \in V_i$.

Proof. (1) $\implies$ (2). Assume that we have two representations $v = \sum_i v_i = \sum_i w_i$, where $v_i, w_i \in V_i$. Then $\sum_i (v_i - w_i) = 0$, where $v_i - w_i \in V_i$. By our assumption we conclude that $v_i - w_i = 0$, hence $v_i = w_i$ for all i .

(2) $\implies$ (1). Is clear. □

Theorem 1.23. Let $V_1, V_2 \subset V$ be subspaces. Then

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2).$$

Proof. Consider a basis $(e_1, \dots, e_k)$ of $V_1 \cap V_2$ and extend it to a basis $(e_1, \dots, e_k, f_1, \dots, f_l)$ of V_1 and a basis $(e_1, \dots, e_k, f'_1, \dots, f'_m)$ of V_2 . We claim that

$$(e_1, \dots, e_k, f_1, \dots, f_l, f'_1, \dots, f'_m)$$

is a basis of $V_1 + V_2$. It generates $V_1 + V_2$ as it contains a basis of V_1 and a basis of V_2 . Assume that we have a linear dependence

$$\sum a_i e_i + \sum b_i f_i + \sum c_i f'_i = 0, \quad a_i, b_i, c_i \in K.$$

Then $v = \sum a_i e_i + \sum b_i f_i = -\sum c_i f'_i$ is contained in $V_1 \cap V_2$, hence is a linear combination of $e_1, \dots, e_k$. This linear combination together with expression $v = -\sum c_i f'_i$ give two representations of v with respect to the basis $(e_1, \dots, e_k, f'_1, \dots, f'_m)$ of V_2 . We conclude that $c_i = 0$ for all i . This implies that $\sum a_i e_i + \sum b_i f_i = 0$, hence $a_i = 0$ and $b_i = 0$. We conclude that the above collection of vectors is indeed a basis of $V_1 + V_2$. Therefore

$$\dim(V_1 + V_2) = k + l + m, \quad \dim V_1 = k + l, \quad \dim V_2 = k + m, \quad \dim(V_1 \cap V_2) = k$$

and we conclude the statement of the theorem. □

Corollary 1.24. We have $\dim(V_1 \oplus V_2) = \dim V_1 + \dim V_2$.

Remark 1.25. Given vector spaces $V_1, \dots, V_r$, we define their **(external) direct sum** $V = \bigoplus_i V_i = V_1 \oplus \dots \oplus V_r$ to be the set $\{(v_1, \dots, v_r) \mid v_i \in V_i\}$ equipped with a vector space structure

$$(v_1, \dots, v_r) + (w_1, \dots, w_r) = (v_1 + w_1, \dots, v_r + w_r), \quad \lambda(v_1, \dots, v_r) = (\lambda v_1, \dots, \lambda v_r).$$

Every vector space V_i can be considered as a subspace of V , where $v_i \in V_i$ is identified with $(0, \dots, 0, \overset{i}{v_i}, 0, \dots, 0)$. Then the previous (internal) direct sum $\bigoplus_i V_i$ is equal to V , hence we can identify both direct sums.

Remark 1.26 (Direct sum of matrices). Let $\phi_i: V_i \rightarrow V_i$ be linear operators for $1 \leq i \leq r$ and let $V = \bigoplus_{i=1}^r V_i$. Then we can define a linear operator

$$\phi = \bigoplus_{i=1}^r \phi_i: V \rightarrow V, \quad (v_1, \dots, v_r) \mapsto (\phi_1 v_1, \dots, \phi_r v_r).$$

Let $\mathbf{e}^{(i)}$ be a basis of V_i and $A_i = [\phi_i]_{\mathbf{e}^{(i)}}$ be the matrix associated with ϕ_i . Then the matrix of ϕ with respect to the basis $\mathbf{e} = (\mathbf{e}^{(1)}, \dots, \mathbf{e}^{(r)})$ of V is equal to

$$[\phi]_{\mathbf{e}} = \bigoplus_{i=1}^r A_i = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A_r \end{pmatrix},$$

called the **direct sum** of the matrices $A_1, \dots, A_r$.

1.6. Intersection of subspaces. Let $V_1, V_2 \subset V = K^n$ be subspaces and let $(e_1, \dots, e_k)$ and $(f_1, \dots, f_l)$ be their respective bases. Then every vector in the intersection $V_1 \cap V_2$ is of the form

$$a_1 e_1 + \dots + a_k e_k = -(b_1 f_1 + \dots + b_l f_l), \quad a_i, b_j \in K.$$

Considering $e_i, f_j \in K^n$ as column vectors, the above equation can be interpreted as a system of linear equations corresponding to an $n \times (k+l)$ matrix A (with $k+l$ variables a_i, b_j). The space of solutions has dimension (by the rank-nullity theorem)

$$\text{null } A = k + l - \text{rk } A = \dim V_1 + \dim V_2 - \dim(V_1 + V_2) = \dim(V_1 \cap V_2).$$

This implies that in order to determine a basis of $V_1 \cap V_2$ we should find a basis of $\text{Ker } A$ and, for each vector $(a_1, \dots, a_k, b_1, \dots, b_l)$ in this basis, consider the vector $a_1 e_1 + \dots + a_k e_k$ in $V_1 \cap V_2$.

Example 1.27. Consider $V = K^3$ and its subspaces

$$V_1 = \text{span} \left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right), \quad V_2 = \text{span} \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right).$$

Note that the first two vectors of V_1 form its basis. By the previous discussion we need to find the null space of

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \xrightarrow{r_2 - r_1} \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \xrightarrow{\substack{r_3 - r_2 \\ r_1 - r_3 + r_2}} \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

We choose a free variable x_4 and obtain $x_3 = x_4$, $x_2 = -x_4$, $x_1 = -x_4$. Therefore the kernel is generated by the vector $(1, 1, -1, -1)^T$. The corresponding basis vector of $V_1 \cap V_2$ is

$$1 \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}.$$

1.7. Quotient space. Let V be a vector space and $U \subset V$ be a subspace. Define a binary relation on V (it corresponds to a subset of $V \times V$)

$$v \sim w \iff v - w \in U.$$

It is an **equivalence relation** as it satisfies the axioms

- (1) $v \sim v$ (reflexivity).
- (2) $v \sim w \iff w \sim v$ (symmetry).
- (3) $u \sim v, v \sim w \implies u \sim w$ (transitivity).

For every $v \in V$, consider its **equivalence class** (or coset)

$$[v] = \{w \in V \mid w \sim v\} = v + U = \{v + u \mid u \in U\}.$$

The set of all equivalence classes is denoted by V/U and is called the **quotient space**. It is equipped with a vector space structure with addition and scalar multiplication given by

$$[u] + [v] = [u + v], \quad \lambda[v] = [\lambda v], \quad u, v \in V, \lambda \in K.$$

There is a surjective linear map

$$\pi: V \rightarrow V/U, \quad v \mapsto [v],$$

called the **quotient map**.

Theorem 1.28. *Given a subspace $U \subset V$, we have $\dim V/U = \dim V - \dim U$.*

Proof. Consider the quotient map $\pi: V \rightarrow V/U$. Then $\text{Ker } \pi = U$ and $\text{Im } \pi = V/U$. Therefore by the rank-nullity theorem

$$\dim V/U + \dim U = \text{rk } \pi + \text{null } \pi = \dim V$$

and the statement follows. $\square$

Remark 1.29. Given subspaces $V_1, V_2 \subset V$, consider the composition

$$\phi: V_1 \rightarrow V_1 + V_2 \rightarrow (V_1 + V_2)/V_2.$$

This map is surjective and $\text{Ker } \phi = V_1 \cap V_2$. Therefore by the rank-nullity theorem

$$\dim V_1 = \text{rk } \phi + \text{null } \phi = \dim(V_1 + V_2)/V_2 + \dim V_1 \cap V_2 = \dim(V_1 + V_2) - \dim V_2 + \dim V_1 \cap V_2.$$

We obtain $\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim V_1 \cap V_2$ which is Theorem 1.23.

Theorem 1.30. *Let $V = U \oplus W$. Then the composition*

$$\phi: W \rightarrow V \xrightarrow{\pi} V/U, \quad w \mapsto [w] = w + U,$$

is an isomorphism.

Proof. If $w \in W$ and $[w] = 0$, then $w \in U$, hence $w \in W \cap U = 0$ and $w = 0$. This implies that $\text{Ker } \phi = 0$ and ϕ is injective. Every vector $v \in V$ can be written in the form $v = u + w$ for some $u \in U$ and $w \in W$. Then $v - w = u \in U$, hence $[v] = [w] = \phi(w)$ and ϕ is surjective. $\square$

Let us discuss a way to find a basis of V/U . A collection of vectors $(e_1, \dots, e_k)$ in V is called a basis of V relative to U if these vectors, together with a basis of U , give a basis of V . Equivalently, this means that $(e_1, \dots, e_k)$ are linearly independent and

$$V = \text{span}(e_1, \dots, e_k) \oplus U.$$

Lemma 1.31. *Let $U \subset V$ be a subspace and $(e_1, \dots, e_k)$ be a basis of V relative to U . Then $([e_1], \dots, [e_k])$ is a basis of V/U .*

First proof. By the previous result, there is an isomorphism $W = \text{span}(e_1, \dots, e_k) \simeq V/U$. Under this isomorphism, the basis $(e_1, \dots, e_k)$ of W is mapped to the basis $([e_1], \dots, [e_k])$ of V/U . $\square$

Second proof. Let $(f_1, \dots, f_l)$ be a basis of U so that $(e_1, \dots, e_k, f_1, \dots, f_l)$ is a basis of V . For every $v \in V$, we can write $v = \sum_i a_i e_i + \sum_j b_j f_j$. Then $u = \sum b_j f_j \in U$ and $[v] = [v - u] = [\sum_i a_i e_i] = \sum_i a_i [e_i]$, hence these vectors span V/U . If these vectors are linearly dependent, then we have $0 = \sum_i a_i [e_i] = [\sum_i a_i e_i]$, hence $\sum_i a_i e_i \in U$ and we can write $\sum_i a_i e_i = \sum_j b_j f_j$ for some $b_j \in K$. As all these vectors form a basis of V , we conclude that $a_i = 0$. $\square$

Remark 1.32. Using the above notation, the vector space V/U has a basis $([e_1], \dots, [e_k])$, hence $\dim V/U = k$. On the other hand $\dim V = k + \dim U$, hence $\dim V/U = \dim V - \dim U$, giving an alternative proof of Theorem 1.28.

Let $V = K^n$ and $U = \text{span}(f_1, \dots, f_l) \subset V$ (these vectors can be linearly dependent). In order to find a basis of V relative to U , we write down a matrix consisting of the column vectors

$$(f_1, \dots, f_l, e_1, \dots, e_n)$$

where $(e_1, \dots, e_n)$ is the standard basis of V (or any other basis). Then we perform column operations to bring the first l columns to the column echelon form and then bring the last n columns to the column echelon form. The non-zero columns among the last n columns give a relative basis.

Example 1.33. Let $V = K^3$ and let $U = \text{span} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 2 \\ 1 & 0 & 1 \end{pmatrix}$. Then U is generated by the first two vectors and we can remove the third vector and consider

$$\left(\begin{array}{cc|cc} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{array} \right) \xrightarrow{c_3 - c_1} \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & -1 & 1 & 0 \\ 1 & 0 & -1 & 0 & 1 \end{array} \right) \rightarrow \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 \end{array} \right) \rightarrow \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{array} \right)$$

Therefore there is a relative basis consisting of the vector $(0 \ 0 \ 1)^\top$.

2. LINEAR OPERATORS

2.1. Linear operators. A **linear operator** (or a linear transformation or an endomorphism) is a linear map $\phi: V \rightarrow V$. Let $\mathbf{e} = (e_1, \dots, e_n)$ be a basis of V . Then we define the matrix associated with ϕ relative to basis $\mathbf{e}$

$$[\phi]_{\mathbf{e}} = [\phi]_{\mathbf{e}, \mathbf{e}} = (a_{ij}), \quad \phi(e_j) = \sum_i a_{ij} e_i.$$

Lemma 2.1. Let $\phi: V \rightarrow V$ be a linear operator and $\mathbf{e}, \mathbf{f}$ be two bases of V . Then

$$[\phi]_{\mathbf{f}} = M_{\mathbf{f} \rightarrow \mathbf{e}}^{-1} \cdot [\phi]_{\mathbf{e}} \cdot M_{\mathbf{f} \rightarrow \mathbf{e}},$$

where $M_{\mathbf{f} \rightarrow \mathbf{e}}$ is the transition matrix.

Proof. We have $[\phi]_{\mathbf{f}} = [\phi]_{\mathbf{f}, \mathbf{f}} = M_{\mathbf{e} \rightarrow \mathbf{f}} \cdot [\phi]_{\mathbf{e}, \mathbf{e}} \cdot M_{\mathbf{f} \rightarrow \mathbf{e}} = M_{\mathbf{f} \rightarrow \mathbf{e}}^{-1} \cdot [\phi]_{\mathbf{e}} \cdot M_{\mathbf{f} \rightarrow \mathbf{e}}$. $\square$

Remark 2.2. Two square matrices A, B are called **similar** or **conjugate** if there exists an invertible matrix M such that $B = M^{-1}AM$. Note that in this case $A = N^{-1}BN$, where $N = M^{-1}$. The previous result implies that the matrices $[\phi]_{\mathbf{f}}$ and $[\phi]_{\mathbf{e}}$ are conjugate.

Example 2.3. Consider the vector space $V = K^2$, the standard basis $\mathbf{e}$ of V and the basis $\mathbf{f} = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right)$. Then

$$M = M_{\mathbf{f} \rightarrow \mathbf{e}} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad M^{-1} = M_{\mathbf{e} \rightarrow \mathbf{f}} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}$$

Consider a linear map $A: V \rightarrow V$ given by the matrix $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Then

$$Af_1 = \begin{pmatrix} 1 \\ 3 \end{pmatrix} = -2f_1 + 3f_2, \quad Af_2 = \begin{pmatrix} 3 \\ 7 \end{pmatrix} = -4f_1 + 7f_2.$$

Therefore the matrix of this linear map with respect to the basis $\mathbf{f}$ is $B = \begin{pmatrix} -2 & -4 \\ 3 & 7 \end{pmatrix}$. We can check that

$$M^{-1}AM = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -2 & -4 \\ 3 & 7 \end{pmatrix} = B.$$

Remark 2.4. We used the fact that if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible, then $A^{-1} = \frac{1}{\det A} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.

2.2. Invariant subspaces and eigenvectors. Recall that a linear operator is a linear map $\phi: V \rightarrow V$.

Definition 2.5. Let $\phi: V \rightarrow V$ be a linear operator.

- (1) A subspace $U \subset V$ is called an **invariant subspace** of ϕ (or ϕ -invariant) if $\phi(U) \subset U$. This means that $\phi(u) \in U$, for all $u \in U$.
- (2) A nonzero vector $v \in V$ is called an **eigenvector** of ϕ if $\phi(v) = \lambda v$ for some $\lambda \in K$. The scalar λ is called an **eigenvalue** of ϕ .

Remark 2.6. Given an $n \times n$ matrix A , we consider the corresponding linear map $A: K^n \rightarrow K^n$ and similarly define invariant subspaces, eigenvectors and eigenvalues of A .

Example 2.7. Let $v \in V$ be an eigenvector of ϕ . Then $U = \text{span}(v) = Kv$ is an invariant subspace. Indeed, assume that $\phi(v) = \lambda v$. Then, for every $cv \in Kv$ with $c \in K$, we have $\phi(cv) = c\phi(v) = (c\lambda)v \in Kv$, hence $\phi(U) \subset U$. Conversely, assume that $U \subset V$ is a one-dimensional invariant subspace. Let $U = Kv$, for some $v \in V$. Then $\phi(v) \in U = Kv$, hence $\phi(v) = \lambda v$, for some $\lambda \in K$. This implies that v is an eigenvector of ϕ .

Lemma 2.8. Let A, B be commuting square matrices ($AB = BA$) of size n . Then $\text{Ker } B \subset K^n$ is an A -invariant subspace.

Proof. If $v \in \text{Ker } B$, then $BA(v) = AB(v) = 0$, hence $A(v) \in \text{Ker } B$. □

Theorem 2.9. Let A be an $n \times n$ matrix. Then $\lambda \in K$ is an eigenvalue of $A \iff$ it is a root of the polynomial

$$\chi_A(t) = \det(tI - A),$$

called the **characteristic polynomial** of the matrix A .

Proof. Assume that $Av = \lambda v$ for some $0 \neq v \in K^n$. Then $(\lambda I - A)v = 0$, hence $v \in \text{Ker}(\lambda I - A)$. This implies that $\lambda I - A$ is not invertible (is singular), hence $\det(\lambda I - A) = 0$. Therefore $\chi_A(\lambda) = 0$. Conversely, if $\det(\lambda I - A) = 0$, then $\lambda I - A$ is singular, hence there exists $0 \neq v \in \text{Ker}(\lambda I - A)$. This implies that $Av = \lambda v$ and λ is an eigenvalue of A . □

Example 2.10. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then

$$\begin{aligned} \chi_A(t) &= \begin{vmatrix} t-a & -b \\ -c & t-d \end{vmatrix} = (t-a)(t-d) - bc = t^2 - (a+d)t + (ad-bc) \\ &= t^2 - \text{tr}(A)t + \det(A). \end{aligned}$$

Example 2.11. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. Then $\chi_A(t) = t^2 - 2t + 1 = (t-1)^2$. This polynomial has a unique root $\lambda = 1$. This is the only eigenvalue of A . If $x \in K^2$ is the corresponding eigenvector, then $Ax = \lambda x$, hence $(A - \lambda I)x = 0$. We have

$$A - \lambda I = A - I = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad (A - \lambda I)x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ 0 \end{pmatrix}.$$

Therefore we require $x_2 = 0$, hence $x \in K \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. All nonzero vectors of this form are eigenvectors of A .

Example 2.12. Given $\theta \in \mathbb{R}$, consider the linear map (rotation by angle θ anti-clockwise)

$$e^{i\theta}: \mathbb{C} \rightarrow \mathbb{C}, \quad z \mapsto e^{i\theta}z = (\cos \theta + i \sin \theta)z.$$

The matrix of this linear map with respect to the basis $(1, i)$ of $\mathbb{C}$ over $\mathbb{R}$ is given by

$$A = [e^{i\theta}] = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

Note that $\det A = \cos^2 \theta + \sin^2 \theta = 1$ and the characteristic polynomial is equal to $\chi_A(t) = t^2 - 2(\cos \theta)t + 1$. In particular, for $\theta = \frac{\pi}{2}$, we have $e^{i\theta} = i$, $A = [e^{i\theta}] = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $\chi_A(t) = t^2 + 1$. This polynomial does not have real roots (elements $t \in \mathbb{R}$ such that $\chi_A(t) = 0$). Therefore A does not have eigenvalues or eigenvectors over $\mathbb{R}$.

Lemma 2.13. *A square matrix (linear operator) over $K = \mathbb{C}$ has an eigenvector.*

Proof. Every (non-constant) polynomial over $\mathbb{C}$ has a root and is a product of linear factors. In particular, the characteristic polynomial $\chi_A(t)$ has a root $\lambda \in \mathbb{C}$. By Theorem 2.9, λ is an eigenvalue of A , hence A has an eigenvector. $\square$

Lemma 2.14. *If A and B are conjugate matrices, then $\chi_A(t) = \chi_B(t)$.*

Proof. Assume that $B = MAM^{-1}$ for an invertible matrix M . Then $\det(tI - B) = \det(tI - MAM^{-1}) = \det(M(tI - A)M^{-1}) = \det(M^{-1}M(tI - A)) = \det(tI - A)$. Therefore $\chi_A(t) = \chi_B(t)$. $\square$

Corollary 2.15. *Let $\phi: V \rightarrow V$ be a linear operator, $\mathbf{e} = (e_1, \dots, e_n)$ be a basis of V and $A = [\phi]_{\mathbf{e}}$ be the corresponding matrix of ϕ . Then*

$$\chi_{\phi}(t) = \chi_A(t) = \det(tI - A)$$

*is independent of the basis $\mathbf{e}$. It is called the **characteristic polynomial** of ϕ .*

Proof. Let $\mathbf{f}$ be another basis of V , $A = [\phi]_{\mathbf{e}}$ and $B = [\phi]_{\mathbf{f}}$. Then $B = M_{\mathbf{e} \rightarrow \mathbf{f}} \cdot A \cdot M_{\mathbf{f} \rightarrow \mathbf{e}} = MAM^{-1}$, where $M = M_{\mathbf{e} \rightarrow \mathbf{f}}$ is the transition matrix. Now we apply the previous result. $\square$

Remark 2.16. The above results imply that in order to find eigenvectors and eigenvalues of a matrix A , we should compute the characteristic polynomial $\chi_A(t)$, find the roots of $\chi_A(t)$ and, for each root λ , find the kernel $\text{Ker}(\lambda I - A) = \text{Ker}(A - \lambda I)$.

Remark 2.17. Let $\phi: V \rightarrow V$ be a linear operator and $U \subset V$ be an invariant subspace. Let $(e_1, \dots, e_k)$ be a basis of U and $(f_1, \dots, f_l)$ be its complement to a basis of V . Then the matrix of ϕ with respect to the basis $(e_1, \dots, e_k, f_1, \dots, f_l)$ is of the form $\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$, where A is a $k \times k$ matrix corresponding to ϕ acting on U and C is an $l \times l$ matrix corresponding to ϕ acting on V/U . This implies that $\chi_{\phi}(t) = \det(tI - A) \cdot \det(tI - C)$.

2.3. Diagonalizable operators.

Definition 2.18. A linear operator $\phi: V \rightarrow V$ is called **diagonalizable** if one of the following equivalent conditions is satisfied

- (1) V is a direct sum of one-dimensional invariant subspaces.
- (2) There exists a basis of V in which the matrix of ϕ is diagonal.

Proof of equivalence. (1) $\implies$ (2). Assume that $V = \bigoplus_{i=1}^n V_i$, where $\dim V_i = 1$ and V_i are invariant. If $0 \neq e_i \in V_i$, then $\phi(e_i) \in V_i = Ke_i$, hence $\phi(e_i) = \lambda_i e_i$ for some $\lambda_i \in K$. This implies that the matrix of ϕ with respect to the basis $(e_1, \dots, e_n)$ is the diagonal matrix $\text{diag}(\lambda_1, \dots, \lambda_n)$.
(2) $\implies$ (1). Let $(e_1, \dots, e_n)$ be a basis of V in which the matrix of ϕ is a diagonal matrix $\text{diag}(\lambda_1, \dots, \lambda_n)$. Then $V = \bigoplus_i V_i$, where $V_i = Ke_i$. We have $\phi(e_i) \in Ke_i$, hence $\phi(V_i) = \phi(Ke_i) \subset Ke_i = V_i$. This means that subspaces V_i are invariant. $\square$

Remark 2.19. Note that the second condition means that there exists a basis of V consisting of eigenvectors of ϕ .

Remark 2.20. We say that an $n \times n$ matrix A is diagonalizable if the corresponding linear operator $L_A: K^n \rightarrow K^n$ is.

Lemma 2.21. A matrix A is diagonalizable $\iff$ it is conjugate to a diagonal matrix.

Proof. Assume that $L_A: K^n \rightarrow K^n$ is diagonalizable. Then there exists a basis $\mathbf{f} = (f_1, \dots, f_n)$ such that $B = [L_A]_{\mathbf{f}}$ is diagonal. Note that $A = [L_A]_{\mathbf{e}}$, where $\mathbf{e} = (e_1, \dots, e_n)$ is the standard basis. We have

$$[L_A]_{\mathbf{f}} = M_{\mathbf{e} \rightarrow \mathbf{f}} \cdot [L_A]_{\mathbf{e}} \cdot M_{\mathbf{f} \rightarrow \mathbf{e}} = M_{\mathbf{f} \rightarrow \mathbf{e}}^{-1} \cdot [L_A]_{\mathbf{e}} \cdot M_{\mathbf{f} \rightarrow \mathbf{e}}.$$

This implies $B = M^{-1}AM$, where $M = M_{\mathbf{f} \rightarrow \mathbf{e}} = (f_1 | \dots | f_n)$ is the transition matrix.

Conversely, assume that $B = M^{-1}AM$ is diagonal, for some invertible matrix M . Let us define a basis $\mathbf{f} = (f_1, \dots, f_n)$ consisting of the columns of the matrix M , that is, $M = (f_1 | \dots | f_n)$. Then $M_{\mathbf{f} \rightarrow \mathbf{e}} = M$, hence

$$[L_A]_{\mathbf{f}} = M_{\mathbf{f} \rightarrow \mathbf{e}}^{-1} \cdot [L_A]_{\mathbf{e}} \cdot M_{\mathbf{f} \rightarrow \mathbf{e}} = M^{-1}AM$$

is diagonal. This implies that A is diagonalizable. $\square$

Example 2.22. Let $A = \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix}$. Then

$$\chi_A(t) = \det(tI - A) = \begin{vmatrix} t & 2 \\ -1 & t-3 \end{vmatrix} = t(t-3) + 2 = (t-1)(t-2).$$

This implies that A has eigenvalues $\lambda_1 = 1$, $\lambda_2 = 2$. The corresponding eigenvectors are vectors in the null spaces of the matrices

$$A - I = \begin{pmatrix} -1 & -2 \\ 1 & 2 \end{pmatrix}, \quad A - 2I = \begin{pmatrix} -2 & -2 \\ 1 & 1 \end{pmatrix}.$$

We can take $f_1 = \begin{pmatrix} 2 \\ -1 \end{pmatrix} \in \text{Ker}(A - I)$ and $f_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \in \text{Ker}(A - 2I)$. Then $M = M_{\mathbf{f} \rightarrow \mathbf{e}} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}$. We can see that $M^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$ and

$$M^{-1}AM = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 0 & -2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

is diagonal.

Example 2.23. Consider the matrix $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$, for some $\lambda \in K$. Then $\chi_A(t) = (t - \lambda)^2$, hence A has only one eigenvalue λ . The corresponding eigenvectors are contained in $\text{Ker}(A - \lambda I) = \text{Ker} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = K \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. This implies that there is only one eigenvector (up to a scalar), hence A is not diagonalizable.

Lemma 2.24. Let $\lambda_1, \dots, \lambda_r$ be distinct eigenvalues of A . Then the sum $\sum_{i=1}^r \text{Ker}(A - \lambda_i I)$ is direct.

Proof. Let $v_i \in V_i = \text{Ker}(A - \lambda_i I)$ be such that $\sum_i v_i = 0$ and some v_i are nonzero (there are at least two of them). We can assume that the number of non-zero summands is minimal. We have $Av_i = \lambda_i v_i$ for all i . If $v_k \neq 0$, then

$$(A - \lambda_k I) \sum_i v_i = \sum_i (\lambda_i - \lambda_k) v_i = 0$$

has one less non-zero summand as $(\lambda_k - \lambda_k)v_k = 0$. This is a contradiction. $\square$

Corollary 2.25. Let A be an $n \times n$ matrix having n distinct eigenvalues. Then A is diagonalizable.

Proof. Let $\lambda_1, \dots, \lambda_n$ be distinct eigenvalues and $v_1, \dots, v_n$ be distinct eigenvectors. Then $Kv_i \subset \text{Ker}(A - \lambda_i I)$, hence $\bigoplus_{i=1}^n Kv_i \subset K^n$ is a direct sum by the previous result. These spaces have the same dimension, hence $K^n = \bigoplus_{i=1}^n Kv_i$ and $(v_1, \dots, v_n)$ is a basis of V . $\square$

Remark 2.26. The previous lemma actually implies that if $V = \sum_i \text{Ker}(A - \lambda_i I)$, then A is diagonalizable.

2.4. Polynomials of matrices and linear operators.

2.4.1. *Polynomials.* We define a **polynomial** over K to be an expression of the form

$$f(t) = \sum_{i=0}^n f_i t^i = f_n t^n + \cdots + f_1 t + f_0, \quad f_i \in K.$$

Let $K[t]$ denote the set of all polynomials over K . Given two polynomials $f, g \in K[t]$ and $\lambda \in K$, we can define the sum $f + g \in K[t]$, the product $fg \in K[t]$ and the scalar product $\lambda f \in K[t]$ in an obvious way.

We define the **degree** $\deg f$ of a nonzero polynomial f to be the maximal $n \geq 0$ such that $f_n \neq 0$. For the zero polynomial, we define $\deg 0 = -\infty$. We say that a polynomial f is **monic** if $f_n = 1$, where $n = \deg f$.

Theorem 2.27 (Euclidean division). *Given nonzero polynomials $f, g \in K[t]$, there exist polynomials $q \in K[t]$ (called the **quotient**) and $r \in K[t]$ (called the **remainder**) such that*

$$f = qg + r, \quad \deg r < \deg g.$$

The pair (q, r) is uniquely determined.

Remark 2.28. Note that if $f = qg$, then $r = 0$, hence $\deg r = -\infty < \deg g$ automatically.

Proof. Let $d = \deg g$ and assume for simplicity that $g_d = 1$. Let $q \in K[t]$ be such that $r = f - qg$ is of minimal possible degree. If $e = \deg r \geq d$, consider $\bar{q} + r_e t^{e-d}$. Then

$$\bar{r} = f - \bar{q}g = r - r_e t^{e-d}g = \sum_{i=0}^{e-1} r_i t^i - r_e t^{e-d} \sum_{i=0}^{d-1} g_i t^i$$

has degree $< e$, contradicting to the choice of r . Therefore $\deg r < d = \deg g$. Given a different pair of polynomials $(\bar{q}, \bar{r})$ with $f = \bar{q}g + \bar{r}$ and $\deg \bar{r} < \deg g$. We obtain $r - \bar{r} = (\bar{q} - q)g$, hence $\deg(r - \bar{r}) \geq \deg g$, a contradiction. $\square$

2.4.2. *Evaluation of polynomials.* Given a polynomial $f(t) = \sum_{i=0}^d f_i t^i$ and a linear operator $\phi: V \rightarrow V$, we define a new linear operator

$$f(\phi) = \sum_{i=0}^d f_i \phi^i = f_d \phi^d + \cdots + f_1 \phi + f_0 \text{id},$$

where $\phi^i = \phi \circ \cdots \circ \phi$ denotes the composition of i maps equal ϕ . Similarly, for any square matrix A , we define a new square matrix

$$f(A) = \sum_{i=0}^d f_i A^i = f_d A^d + \cdots + f_1 A + f_0 I.$$

We say that the polynomial $f(t)$ **annihilates** the linear operator ϕ if $f(\phi) = 0$ (similarly for matrices).

Remark 2.29. Given a block upper-triangular matrix $T = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$, we have

$$f(T) = \begin{pmatrix} f(A) & * \\ 0 & f(C) \end{pmatrix}.$$

Let $M_{m,n}(K)$ denote the set (vector space) of all $m \times n$ matrices with entries in K .

Lemma 2.30. *We have $f(MAM^{-1}) = Mf(A)M^{-1}$, for $A, M \in M_{n,n}(K)$ with invertible M .*

Proof. For any $k \geq 0$, we have

$$(MAM^{-1})^k = (MAM^{-1}) \dots (MAM^{-1}) = MA \dots AM^{-1} = MA^k M^{-1}.$$

Therefore $f(MAM^{-1}) = \sum_k f_k(MAM^{-1})^k = M(\sum_k f_k A^k) M^{-1} = Mf(A)M^{-1}$. $\square$

Theorem 2.31. *For any $A \in M_{n,n}(K)$, there exists a nonzero polynomial $f \in K[t]$ that annihilates A .*

Proof. Note that the set of matrices $M_{n,n}(K)$ is a vector space (with the usual addition and scalar multiplication of matrices) of dimension n^2 . This implies that the matrices (considered as vectors in $M_{n,n}(K)$)

$$I, A, A^2, \dots, A^N$$

are linearly dependent, for $N \geq n^2$. Therefore there exist $f_0, \dots, f_N \in K$, not all of which are zero, such that $\sum_{i=0}^N f_i A^i = 0$. Then $f(A) = 0$, for $f(t) = \sum_{i=0}^N f_i t^i$. $\square$

Definition 2.32. Define the **minimal polynomial** of $A \in M_{n,n}(K)$ to be the monic polynomial $p \in K[t]$ of minimal degree such that $p(A) = 0$.

Lemma 2.33. *Let $p \in K[t]$ be the minimal polynomial of $A \in M_{n,n}(K)$ and $f \in K[t]$ be such that $f(A) = 0$. Then p is a factor of f (meaning that $f = pg$ for some $g \in K[t]$).*

Proof. We can divide with a remainder $f = qp + r$, where $\deg r < \deg p$. Then $f(A) = p(A) = 0$ implies $r(A) = 0$. By the minimality of $\deg p$, we conclude that $r = 0$. $\square$

Theorem 2.34 (Cayley-Hamilton Theorem). *Let $\phi: V \rightarrow V$ be a linear operator and $\chi_\phi(t)$ be its characteristic polynomial. Then $\chi_\phi(\phi) = 0$.*

Remark 2.35. Similarly, if A is an $n \times n$ matrix, then $\chi_A(A) = 0$.

Proof. Let $f(t) = \chi_\phi(t)$ be the characteristic polynomial of ϕ and let $\lambda \in K$ be its root. Let e_1 be an eigenvector corresponding to the eigenvalue λ . We extend it to a basis $(e_1, \dots, e_n)$ of V .

Then the matrix of ϕ relative to this basis is of the form $A = \begin{pmatrix} \lambda & * \\ 0 & C \end{pmatrix}$. We have

$$f(t) = \det(tI - A) = (t - \lambda)g(t), \quad g(t) = \det(tI - C) = \chi_C(t).$$

We can assume by induction that $g(C) = \chi_C(C) = 0$. Then

$$g(A) = \begin{pmatrix} g(\lambda) & * \\ 0 & g(C) \end{pmatrix} = \begin{pmatrix} g(\lambda) & * \\ 0 & 0 \end{pmatrix}.$$

This implies that $g(A)(V) \subset Ke_1$. Therefore

$$f(A)(V) = (A - \lambda I)g(A)(V) \subset (A - \lambda I)(Ke_1) = 0$$

as $Ae_1 = \lambda e_1$. We conclude that $f(A) = 0$. $\square$

Remark 2.36. The above theorem implies that the minimal polynomial of a matrix A is a factor of $\chi_A(t)$. These polynomials can be different. For example, consider a diagonal matrix $A = \text{diag}(\lambda, \dots, \lambda)$ of dimension n . Then the minimal polynomial of A is $p(t) = t - \lambda$, while the characteristic polynomial of A is $\chi_A(t) = (t - \lambda)^n$. One can show that generally $\chi_A(t)$ is a factor of some power of the minimal polynomial.

2.5. Jordan matrices.

Definition 2.37. (1) An $n \times n$ matrix of the form

$$J_{n,\lambda} = J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & \dots & 0 & \lambda \end{pmatrix}$$

is called a **Jordan block** having dimension n and eigenvalue λ .

(2) A block diagonal matrix

$$J_{n_1}(\lambda_1) \oplus \dots \oplus J_{n_r}(\lambda_r)$$

consisting of diagonal blocks $J_{n_i}(\lambda_i)$ is called a **Jordan matrix**.

(3) Define a **Jordan basis** for a linear operator $\phi: V \rightarrow V$ to be a basis of V in which the matrix of ϕ is a Jordan matrix. This matrix is called the **Jordan normal form** of ϕ .

Remark 2.38. Let A be a square $n \times n$ matrix and let $V = K^n$ be equipped with the standard basis $\mathbf{e}$. If $\mathbf{f}$ is a Jordan basis for $L_A: V \rightarrow V$, then

$$J = [L_A]_{\mathbf{f}} = M_{\mathbf{f} \rightarrow \mathbf{e}}^{-1} \cdot A \cdot M_{\mathbf{f} \rightarrow \mathbf{e}}$$

is a Jordan matrix, conjugate to A . We call it the Jordan normal form of A .

Example 2.39. Let $A = J_n(\lambda)$ and let $(e_1, \dots, e_n)$ be the standard basis of K^n . Then $Ae_i = \lambda e_i + e_{i-1}$ for $2 \leq i \leq n$. Therefore $e_{i-1} = (A - \lambda I)e_i = Be_i$, where $B = A - \lambda I$. This implies that $e_i = B^{n-i}e_n$ and the basis $(e_1, \dots, e_n)$ can be written in the form

$$B^{n-1}e_n, \dots, Be_n, e_n.$$

Theorem 2.40 (Jordan normal form). *Let V be a vector space over $\mathbb{C}$ and $\phi: V \rightarrow V$ be a linear operator. Then there exists a Jordan basis for ϕ . Equivalently, for any square matrix A over $\mathbb{C}$, there exists an invertible matrix M such that $J = M^{-1}AM$ is a Jordan matrix. The matrix J is unique (up to a permutation of Jordan blocks).*

We will prove this theorem in several steps occupying the next two sections.

Remark 2.41. Consider the matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ over $\mathbb{R}$. Then $\chi_A(t) = t^2 + 1$, hence A does not have eigenvalues or eigenvectors (over $\mathbb{R}$). This implies that A is not conjugate to a Jordan matrix (over $\mathbb{R}$) as a Jordan matrix always has eigenvalues.

Remark 2.42. The minimal polynomial of $J = J_n(\lambda)$ is $p(t) = (t - \lambda)^n$. Indeed, $p(t)$ is the characteristic polynomial of J , hence $p(J) = 0$. The minimal polynomial should be a factor of $p(t)$, hence it is of the form $(t - \lambda)^k$ for some $1 \leq k \leq n$. One can check that $(J - \lambda I)^k \neq 0$ for $1 \leq k < n$. Consider an arbitrary Jordan matrix $J = \bigoplus_{\lambda \in K, n \geq 1} J_n(\lambda)^{\oplus m_{\lambda,n}}$. For every $\lambda \in K$, the minimal polynomial of $\bigoplus_{n \geq 1} J_n(\lambda)^{\oplus m_{\lambda,n}}$ is $(t - \lambda)^{n_\lambda}$, where $n_\lambda = \max \{n \geq 1 \mid m_{\lambda,n} \neq 0\}$ is the maximal size of a Jordan block with eigenvalue λ appearing in J . The minimal polynomial of J is $\prod_{\lambda \in K} (t - \lambda)^{n_\lambda}$.

2.6. Generalized eigenvectors. Let $A: V \rightarrow V$ be a linear operator (or a square matrix if a basis of V is chosen). We will often write $A - \lambda$ instead of $A - \lambda I$.

Definition 2.43. A vector $v \in V$ is called a **generalized eigenvector** of A corresponding to $\lambda \in K$, if $(A - \lambda)^k v = 0$ for some $k \geq 1$. The set $V(\lambda) \subset V$ consisting of all generalized eigenvectors of A corresponding to λ is called a **generalized eigenspace** of A .

Remark 2.44. For any $\lambda \in K$, we have a chain of subspaces

$$\text{Ker}(A - \lambda) \subset \text{Ker}(A - \lambda)^2 \subset \dots \subset V.$$

This chain stabilizes, meaning that $\text{Ker}(A - \lambda)^m = \text{Ker}(A - \lambda)^{m+1} = \dots$ for some $m \geq 1$. Then $V(\lambda) = \text{Ker}(A - \lambda)^m$.

Lemma 2.45. *The generalized eigenspace $V(\lambda) \subset V$ of $A: V \rightarrow V$ is an A -invariant subspace and $V(\lambda) \neq 0 \iff \lambda$ is an eigenvalue of A .*

Proof. (1) We have seen that $V(\lambda) = \text{Ker}(A - \lambda)^m$ for some $m \geq 1$. This implies that $V(\lambda)$ is a subspace. If $v \in V(\lambda)$, then $(A - \lambda)^m(Av) = A(A - \lambda)^m v = 0$, hence $Av \in \text{Ker}(A - \lambda)^m = V(\lambda)$. This implies that $V(\lambda)$ is A -invariant.

(2) If $V(\lambda) \neq 0$, let $0 \neq v \in V(\lambda)$ and let $(A - \lambda)^k v = 0$ with the minimal $k \geq 1$. Then $w = (A - \lambda)^{k-1} v \neq 0$ and $(A - \lambda)w = 0$. Therefore λ is an eigenvalue. Conversely, if $\lambda \in K$ is an eigenvalue, then $\text{Ker}(A - \lambda) \neq 0$, hence $V(\lambda) \neq 0$. $\square$

Lemma 2.46. *Assume that $f, g \in K[t]$ are coprime (don't have non-constant common factors). Then there exist $u, v \in K[t]$ such that $uf + vg = 1$.*

Proof. Let $I = \{uf + vg \mid u, v \in K[t]\}$ and let $p \in I$ be a non-zero polynomial of minimal degree. Given $h \in I$, we can divide it with a remainder $h = qp + r$, where $\deg r < \deg p$. As $h, p \in I$, we have $r = h - qp \in I$, hence $r = 0$ by the minimality of $\deg p$. This implies $h = qp$. As $f, g \in I$, we obtain that $f = q_1 p$ and $g = q_2 p$, for some $q_1, q_2 \in K[t]$. Therefore f, g have a common factor p . By our assumption this is a constant polynomial and we can assume $p = 1$. By construction, it is of the form $p = 1 = uf + vg$ for some polynomials u, v . $\square$

Theorem 2.47. *Let $A: V \rightarrow V$ be a linear operator (or a square matrix) annihilated by a polynomial $f(t) = \prod_{i=1}^r (t - \lambda_i)^{m_i}$ with distinct roots λ_i . Then V has a direct sum decomposition*

$$V = V_1 \oplus \dots \oplus V_r, \quad V_i = \text{Ker}(A - \lambda_i)^{m_i}.$$

Proof. It is enough to show that if $f = gh$, where g and h are coprime, then we have a direct sum decomposition

$$V = U \oplus W, \quad U = \text{Ker } g(A), \quad W = \text{Ker } h(A).$$

Namely, we can consider $g = (t - \lambda_1)^{m_1}$, $h = \prod_{i=2}^r (t - \lambda_i)^{m_i}$ and proceed by induction on $\dim V$. The subspace $W \subset V$ is A -invariant: if $v \in W = \text{Ker } h(A)$, then $h(A)(Av) = A(h(A)v) = 0$, hence $Av \in \text{Ker } h(A) = W$. The restriction $A|_W$ satisfies $h(A|_W) = h(A)|_W = 0$, hence by induction $W = \bigoplus_{i=2}^r \text{Ker}(A - \lambda_i I)^{m_i}$ and then

$$V = U \oplus W = \bigoplus_{i=1}^r \text{Ker}(A - \lambda_i I)^{m_i}.$$

By Lemma 2.46, there exist polynomials $p, q \in K[t]$ such that $pg + qh = 1$. Therefore, for any $v \in V$,

$$(q(A)h(A) + p(A)g(A))(v) = Iv = v.$$

Consider the summands $u = q(A)h(A)(v)$ and $w = p(A)g(A)(v)$. We have

$$g(A)(u) = g(A)q(A)h(A)(v) = q(A)f(A)(v) = 0,$$

hence $u \in \text{Ker } g(A) = U$ and similarly $w \in \text{Ker } h(A) = W$. As $v = u + w$, we obtain $V = U + W$. If $v \in U \cap W$, then $g(A)v = h(A)v = 0$, hence $v = (pg + qh)(A)v = 0$. This implies $U \cap W = 0$, hence $V = U \oplus W$. $\square$

Remark 2.48. In particular, we can apply the above theorem to the characteristic polynomial $\chi_A(t)$ of a linear operator (or a square matrix) A .

Corollary 2.49. Let $A: V \rightarrow V$ be annihilated by a polynomial $f(t) = \prod_{i=1}^r (t - \lambda_i)^{m_i}$ with distinct roots λ_i . Then $V(\lambda_i) = \text{Ker}(A - \lambda_i)^{m_i}$ and $V(\lambda) = 0$ for $\lambda \neq \lambda_i$.

Proof. We will just show that $V(\lambda_i) = \text{Ker}(A - \lambda_i)^{m_i}$ (if $\lambda \neq \lambda_i$ for all i , we can consider $\lambda_0 = \lambda$, $m_0 = 0$ and $f(t) = \prod_{i=0}^r (t - \lambda_i)^{m_i}$). We have $\text{Ker}(A - \lambda_i)^{m_i} \subset V(\lambda_i)$. If $v \in V(\lambda_i)$, then $v \in \text{Ker}(A - \lambda_i)^m$ for some $m \geq m_i$. Let $g(t) = (t - \lambda_i)^m \prod_{j \neq i} (t - \lambda_j)^{m_j}$. Then $g(A) = f(A) = 0$ and by Theorem 2.47 we have

$$V = \text{Ker}(A - \lambda_i)^m \oplus \bigoplus_{j \neq i} \text{Ker}(A - \lambda_j)^{m_j} = \text{Ker}(A - \lambda_i)^{m_i} \oplus \bigoplus_{j \neq i} \text{Ker}(A - \lambda_j)^{m_j}.$$

As $m \geq m_i$, we have $\text{Ker}(A - \lambda_i)^{m_i} \subset \text{Ker}(A - \lambda_i)^m$ and the above equality implies that $\text{Ker}(A - \lambda_i)^{m_i} = \text{Ker}(A - \lambda_i)^m$. Therefore $v \in \text{Ker}(A - \lambda_i)^{m_i}$ and $V(\lambda_i) \subset \text{Ker}(A - \lambda_i)^{m_i}$. $\square$

Remark 2.50. Because of the decomposition $V = \bigoplus_i \text{Ker}(A - \lambda_i)^{m_i} = \bigoplus_i V(\lambda_i)$, to prove Theorem 2.40 about the Jordan normal form it is enough to show that the restriction of A to $V(\lambda)$ has a Jordan basis. Let us assume that $V = V(\lambda)$ (hence λ is the unique eigenvalue of A). Then $(A - \lambda)^m = 0$ for some $m \geq 1$. A Jordan basis of A is also a Jordan basis of $B = A - \lambda I$ which satisfies $B^m = 0$. Such operator (or matrix) is called **nilpotent**. Next we will construct a Jordan basis of a nilpotent operator.

Example 2.51. Consider the matrix

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

Its characteristic polynomial is $\chi_A(t) = t^2(t - 1)$ and the eigenvalues are $\lambda = 0$ and $\lambda = 1$. Generalized eigenspaces of A are $V(0) = \text{Ker } A^2$ and $V(1) = \text{Ker}(A - I)$. We obtain

$$A^2 = \begin{pmatrix} 2 & -1 & 0 \\ 2 & -1 & 0 \\ 2 & -1 & 0 \end{pmatrix}, \quad A - I = \begin{pmatrix} 0 & -1 & 1 \\ 0 & -2 & 2 \\ 1 & -1 & 0 \end{pmatrix},$$

$$V(0) = \text{Ker } A^2 = \text{span} \left(\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right), \quad V(1) = \text{Ker}(A - I) = \text{span} \left(\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right).$$

Note that $\text{Ker } A = \text{span} \left(\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \right)$. For a Jordan basis of $A|_{V(0)}$, consider a vector in $\text{Ker } A^2 \setminus \text{Ker } A$, for example $v = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Then $Av = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$ and $A^2v = 0$. The matrix of $A|_{V(0)}$ with respect to the basis (Av, v) is equal to $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = J_2(0)$ which is a Jordan block. For a Jordan basis of $A|_{V(1)}$, consider a generator of $V(1)$, for example $w = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. We found a Jordan basis of A

$$\mathbf{f} = (Av, v, w) = \left(\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right).$$

The corresponding Jordan normal form of A is

$$J = J_2(0) \oplus J_1(1) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that $J = [A]_{\mathbf{f}} = M^{-1}AM$, where $M = M_{\mathbf{f} \rightarrow \mathbf{e}} = (Av|v|w)$.

2.7. Nilpotent operators.

Definition 2.52.

- (1) A linear operator $A: V \rightarrow V$ is called **nilpotent** if $A^r = 0$ for some $r \geq 1$. The minimal r with this property is called the **index** of A .
- (2) Given a vector $0 \neq v \in V$, define its **index** (relative to A) to be the minimal $r \geq 1$ such that $A^r v = 0$.

If A has index r , then there exists $v \in V$ having index r .

Lemma 2.53. *Let $v \in V$ have index r relative to A . Then the vectors*

$$v, Av, \dots, A^{r-1}v$$

are linearly independent.

Proof. We can assume that $V = \text{span}(v, Av, \dots, A^{r-1}v)$ (this space is A -invariant) so that $A^r = 0$ on V . Assume that $\sum_{i=0}^{r-1} a_i A^i v = 0$ with some $a_i \neq 0$ and let $0 \leq j \leq r-1$ be the minimal index with $a_j \neq 0$. Dividing by $-a_j$, we can assume that $a_j = -1$. Then

$$A^j v = \sum_{k \geq 0} a_{j+k+1} A^{j+k+1} v = A^{j+1} \sum_{k \geq 0} a_{j+k+1} A^k v = A^{j+1} w.$$

Therefore $A^{r-1} v = A^{r-j-1} A^j v = A^{r-j-1} A^{j+1} w = A^r w = 0$. This contradicts to the assumption that v has index r . $\square$

Definition 2.54. Let $A: V \rightarrow V$ be a linear operator. A vector $v \in V$ is called **cyclic** relative to A (or A -cyclic) if the vectors $v, Av, A^2 v, \dots$ generate V . The vector space V is called **cyclic** relative to A if there exists a cyclic vector $v \in V$ relative to A .

Remark 2.55. Given a vector $v \in V$, consider the subspace $U = \text{span}(v, Av, A^2 v, \dots) \subset V$. It is an A -invariant, cyclic subspace of V , called the cyclic subspace generated by v . Let $r \geq 1$ be the maximal number such that the vectors $v, Av, \dots, A^{r-1} v$ are linearly independent. Then we can write $A^r v = \sum_{i=0}^{r-1} c_i A^i v$ for some $c_i \in K$ and the above vectors form a basis of U .

Remark 2.56. Let $v \in V$ be a cyclic vector of index r . By Lemma 2.53, the vectors

$$A^{r-1} v, \dots, Av, v$$

form a basis of V . The matrix of A with respect to this basis has the form

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

which is the Jordan block $J_{r,0}$.

Theorem 2.57. *Let $A: V \rightarrow V$ be a nilpotent linear operator. Then A has a Jordan basis.*

Proof. We need to show that there exists a decomposition $V = \bigoplus_i V_i$, where V_i are A -invariant and cyclic relative to A . We have seen that $A|_{V_i}$ is represented by a Jordan block in an appropriate basis.

Let $W = A(V) \subset V$. By induction, there exists a decomposition $W = \bigoplus_{i=1}^m W_i$, where W_i are A -invariant and have cyclic vectors $w_i \in W_i$ of index r_i . We have $w_i = Av_i$ for some $v_i \in V$. Then $V_i = \text{span}(v_i, Av_i, \dots, A^{r_i} v_i)$ are A -invariant cyclic spaces.

We claim that $V' = \sum_i V_i$ is a direct sum. Assume that there are polynomials $f_i(t)$ of degree $\leq r_i$ such that $\sum_i f_i(A)v_i = 0$. Then $\sum_i f_i(A)w_i = 0$ and we conclude that $f_i(A)w_i = 0$ for all i . We have $A^{r_i}w_i = 0$, hence $f_i(t) = c_i t^{r_i}$ for some $c_i \in K$ (otherwise the vectors $w_i, \dots, A^{r_i-1}w_i$ are linearly dependent). Therefore $0 = \sum_i f_i(A)v_i = \sum_i c_i A^{r_i-1}w_i$ and we conclude that $c_i = 0$. This implies $f_i(t) = 0$, hence $V' = \sum_i V_i$ is a direct sum as required.

Next, we claim that $V' + \text{Ker } A = V$. For any $v \in V$, we have $Av \in W = AV'$ (as $W_i = AV_i$), hence $Av = Av'$ for some $v' \in V'$. Then $v - v' \in \text{Ker } A$ and our claim follows. This implies that we can choose vectors $u_1, \dots, u_k$ in $\text{Ker } A$ that form a basis of V relative to V' . Then

$$V = V' \oplus \text{span}(u_1, \dots, u_k) = \bigoplus_{i=1}^m V_i \oplus \bigoplus_{i=1}^k Ku_i$$

is the required decomposition. $\square$

Proof of uniqueness. It is enough to prove uniqueness of the Jordan normal form for nilpotent matrices. If $J = J_n(0)$, then $\text{null } J = \dim \text{Ker } J = 1$, $\text{null } J^2 = 2$ and generally $\text{null } J^k = \min\{k, n\}$. If A is conjugate to $J = \bigoplus_{n \geq 1} J_n(0)^{\oplus m_n}$, then

$$\begin{aligned} \text{null } J^k &= \sum_{n \geq 1} m_n \min\{k, n\} = \sum_{n=1}^{k-1} nm_n + k \sum_{n \geq k} m_n, \\ \text{null } J^k - \text{null } J^{k-1} &= \sum_{n \geq k} m_n, \\ m_k &= (\text{null } J^k - \text{null } J^{k-1}) - (\text{null } J^{k+1} - \text{null } J^k). \end{aligned}$$

Therefore the multiplicities m_n are uniquely determined by $\text{null } J^k = \text{null } A^k$ for $k \geq 0$. $\square$

Remark 2.58. To find a Jordan basis of a nilpotent operator A , we construct cyclic vectors of cyclic summands (defined in the theorem) as follows. Let r be the index of A and let $V_k = \text{Ker } A^k$ for $k \geq 0$. We obtain a chain of subspaces

$$0 = V_0 \subset V_1 \subset \dots \subset V_r = V.$$

- (1) Let $W_r = 0$ and let $(e_1^{(r)}, \dots, e_{m_r}^{(r)})$ be a basis of $V = V_r$ relative to $V_{r-1} + W_r = V_{r-1}$.
- (2) Assuming that $W_{k+1} \subset V_{k+1}$ and $(e_i^{(k+1)} \in V_{k+1})_i$ are already constructed, define

$$W_k = A \left(W_{k+1} \oplus \text{span} \left(e_i^{(k+1)} : i \geq 1 \right) \right) \subset AV_{k+1} \subset V_k$$

and let $(e_1^{(k)}, \dots, e_{m_k}^{(k)})$ be a basis of V_k relative to $V_{k-1} + W_k$.

Then the vectors $e_i^{(k)}$ are the required cyclic vectors and the Jordan basis of A consists of vectors

$$A^{k-1}e_i^{(k)}, \dots, Ae_i^{(k)}, e_i^{(k)}, \quad 1 \leq k \leq r, 1 \leq i \leq m_k.$$

2.8. Examples.

Example 2.59. Consider the matrix $A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$. Then $\chi_A(t) = t^2 - 2t + 1 = (t - 1)^2$ and the only eigenvalue of A is $\lambda = 1$. The matrix $B = A - I = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$ is nilpotent ($B^2 = 0$) and we have

$$\text{Ker } B \subset \text{Ker } B^2 = K^2,$$

where $\text{Ker } B = \text{span} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. For a Jordan basis of B (and A) we choose a basis vector of $\text{Ker } B^2 = K^2$ relative to $\text{Ker } B$, for example $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Then $Bv = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and the Jordan basis is

$$\mathbf{f} = (Bv, v) = \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right).$$

The corresponding Jordan normal form of A is $J = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = J_2(1)$. We have

$$J = [A]_{\mathbf{f}} = M^{-1}AM, \quad M = M_{\mathbf{f} \rightarrow \mathbf{e}} = (Bv|v) = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

Example 2.60. Consider the matrix

$$A = \begin{pmatrix} -3 & -2 & -1 \\ 5 & 3 & 2 \\ 1 & 1 & 0 \end{pmatrix}.$$

We can check directly that $A^3 = 0$, hence A is nilpotent. Alternatively, we can verify that $\chi_A(t) = t^3$, hence $A^3 = \chi_A(A) = 0$. We have

$$\text{Ker } A \subset \text{Ker } A^2 \subset \text{Ker } A^3 = K^3,$$

where

$$A^2 = \begin{pmatrix} -2 & -1 & -1 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{pmatrix}, \quad \text{Ker } A^2 = \text{span} \left(\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right).$$

Choose a basis vector of $\text{Ker } A^3 = K^3$ relative to $\text{Ker } A^2$, for example, $v = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Then

$$Av = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} \in \text{Ker } A^2, \quad A^2v = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix} \in \text{Ker } A$$

The Jordan basis of A is $\mathbf{f} = (A^2v, Av, v)$ and the corresponding Jordan matrix is

$$J = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = J_3(0).$$

We have $J = M^{-1}AM$, where $M = M_{\mathbf{f} \rightarrow \mathbf{e}} = (A^2v|Av|v)$.

2.9. Applications. We can write a Jordan block in the form

$$J = J_n(\lambda) = \lambda I + J_n(0).$$

Therefore

$$J^k = (\lambda I + J_n(0))^k = \sum_{i=0}^k \binom{k}{i} \lambda^{k-i} J_n(0)^i$$

$$= \begin{pmatrix} \lambda^k & k\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \binom{k}{3}\lambda^{k-3} & \dots\dots\dots \\ 0 & \lambda^k & k\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \dots\dots\dots \\ 0 & 0 & \lambda^k & k\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \dots\dots \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ 0 & \dots\dots\dots & \dots\dots\dots & \lambda^k & k\lambda^{k-1} \\ 0 & \dots\dots\dots & \dots\dots\dots & 0 & \lambda^k \end{pmatrix},$$

where

$$\binom{k}{i} = \frac{k!}{i!(k-i)!} = \frac{k(k-1)\dots(k-i+1)}{1\dots i}$$

are the binomial coefficients and $k! = 1 \dots k$ (with $0! = 1$). Using the polynomial $f(t) = t^k$, we can write

$$\lambda^k = f(\lambda), \quad k\lambda^{k-1} = f'(\lambda), \quad \binom{k}{i}\lambda^{k-i} = \frac{f^{(i)}(\lambda)}{i!},$$

where $f^{(i)}(t)$ is the i -th derivative of $f(t)$. We obtain for an arbitrary polynomial $f(t)$

$$f(J) = \begin{pmatrix} f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} & \frac{f'''(\lambda)}{3!} & \dots\dots\dots \\ 0 & f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} & \dots\dots\dots \\ 0 & 0 & f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2!} & \dots\dots \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ 0 & \dots\dots\dots & \dots\dots\dots & f(\lambda) & f'(\lambda) \\ 0 & \dots\dots\dots & \dots\dots\dots & 0 & f(\lambda) \end{pmatrix}.$$

A similar formula can be written for an arbitrary Jordan matrix J . Assume now that A is a square $n \times n$ matrix with a Jordan basis $\mathbf{v}$ and a Jordan matrix J . Then $J = [A]_{\mathbf{v}} = M^{-1}AM$, where $M = M_{\mathbf{v} \rightarrow \mathbf{e}}$. Therefore $A = MJM^{-1}$ and

$$f(A) = f(MJM^{-1}) = Mf(J)M^{-1}.$$

Using the above formula for $f(J)$, we can compute $f(A)$.

Example 2.61. Consider a sequence $x_{n+2} = x_n + x_{n+1}$ with $x_0 = 0$, $x_1 = 1$ (called a Fibonacci sequence). Setting $v_n = \begin{pmatrix} x_n \\ x_{n+1} \end{pmatrix}$, we obtain

$$v_{n+1} = \begin{pmatrix} x_{n+1} \\ x_n + x_{n+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_n \\ x_{n+1} \end{pmatrix} = Av_n, \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Therefore $v_n = A^n v_0$. We have $\chi_A(t) = t^2 - t - 1$ with roots $\lambda_{1,2} = \frac{1}{2}(1 \pm \sqrt{5})$. Then $A - \lambda_1 I = \begin{pmatrix} -\lambda_1 & 1 \\ 1 & * has a null space generated by $\begin{pmatrix} 1 \\ \lambda_1 \end{pmatrix}$ and $A - \lambda_2 I$ has a null space$

generated by $\begin{pmatrix} 1 \\ \lambda_2 \end{pmatrix}$. We obtain a Jordan basis

$$f_1 = \begin{pmatrix} 1 \\ \lambda_1 \end{pmatrix}, \quad f_2 = \begin{pmatrix} 1 \\ \lambda_2 \end{pmatrix}$$

and the corresponding (diagonal) Jordan matrix $J = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$. The transition matrix is $M = M_{f \rightarrow e} = \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix}$ with $\det M = \lambda_2 - \lambda_1 = -\sqrt{5}$ and $M^{-1} = \frac{1}{-\sqrt{5}} \begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix}$. Then $A = MJM^{-1}$ and

$$\begin{aligned} A^n &= MJ^nM^{-1} = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix} \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} \begin{pmatrix} -\lambda_2 & 1 \\ \lambda_1 & -1 \end{pmatrix} \\ &= \frac{1}{\sqrt{5}} \begin{pmatrix} \lambda_1^n & \lambda_2^n \\ \lambda_1^{n+1} & \lambda_2^{n+1} \end{pmatrix} \begin{pmatrix} -\lambda_2 & 1 \\ \lambda_1 & -1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} * & \lambda_1^n - \lambda_2^n \\ * & * \end{pmatrix}. \end{aligned}$$

Therefore

$$\begin{aligned} v_n &= \begin{pmatrix} x_n \\ x_{n+1} \end{pmatrix} = A^n v_0 = A^n \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}} \begin{pmatrix} \lambda_1^n - \lambda_2^n \\ * \end{pmatrix}, \\ x_n &= \frac{1}{\sqrt{5}} (\lambda_1^n - \lambda_2^n) = \frac{(1 + \sqrt{5})^n - (1 - \sqrt{5})^n}{2^n \sqrt{5}}. \end{aligned}$$

Example 2.62. Let $A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$ and let us compute

$$e^A = \sum_{n \geq 0} \frac{A^n}{n!}.$$

The Jordan normal form of A is (see Example 2.59)

$$J = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = M^{-1}AM, \quad M = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad M^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix}.$$

Considering the function $f(t) = e^t$, we obtain $f'(t) = e^t$, hence

$$e^J = \begin{pmatrix} f(1) & f'(1) \\ 0 & f(1) \end{pmatrix} = \begin{pmatrix} e & e \\ 0 & e \end{pmatrix}$$

$$e^A = e^{MJM^{-1}} = Me^J M^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} e & e \\ 0 & e \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} e & 2e \\ e & e \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 2e & -e \\ e & 0 \end{pmatrix}$$

Example 2.63. Let us compute A^{100} for the above matrix A . If $f(t) = t^{100}$, then $f(1) = 1$ and $f'(1) = 100$. Therefore

$$A^{100} = MJ^{100}M^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 100 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 101 \\ 1 & 100 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 101 & -100 \\ 100 & -99 \end{pmatrix}$$

3. INNER PRODUCTS

3.1. Dual spaces.

Definition 3.1. Let V be a vector space over K .

- (1) Define a **linear functional** on V to be a linear map $f: V \rightarrow K$.
- (2) Define the **dual space** of V to be the set of all linear functionals on V , denoted by V^* .

The dual space V^* has a vector space structure with addition and scalar multiplication defined pointwise. Namely, for any $f, g \in V^*$ and $c \in K$, we define

$$f + g \in V^*, \quad (f + g)(v) = f(v) + g(v), \quad v \in V,$$

$$cf \in V^*, \quad (cf)(v) = c \cdot f(v), \quad v \in V.$$

Define a **pairing**

$$\langle -, - \rangle : V^* \times V \rightarrow K, \quad \langle f, v \rangle = f(v), \quad f \in V^*, v \in V.$$

It is **bilinear**, meaning that

$$\langle c_1 f_1 + c_2 f_2, v \rangle = c_1 \langle f_1, v \rangle + c_2 \langle f_2, v \rangle, \quad \langle f, c_1 v_1 + c_2 v_2 \rangle = c_1 \langle f, v_1 \rangle + c_2 \langle f, v_2 \rangle,$$

for $v, v_1, v_2 \in V$, $f, f_1, f_2 \in V^*$, $c_1, c_2 \in K$. Similarly, we define a bilinear pairing

$$\langle -, - \rangle : V \times V^* \rightarrow K, \quad \langle v, f \rangle = f(v), \quad f \in V^*, v \in V.$$

Example 3.2. One can identify vectors of $V = K^n$ with column vectors (or $n \times 1$ matrices) and vectors of V^* with row vectors (or $1 \times n$ matrices). Then, given $v \in V$ and $f \in V^*$, we can interpret $f(v) \in K$ as a matrix product $f \cdot v$ which is a 1×1 matrix. For example, $f = \begin{pmatrix} 1 & 2 & 1 \end{pmatrix}$ defines a linear functional

$$f: K^3 \rightarrow K, \quad x \mapsto f \cdot x = \begin{pmatrix} 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 + 2x_2 + x_3.$$

Lemma 3.3. Let V be a vector space with a basis $(e_1, \dots, e_n)$. For any $y \in K^n$, there exists a unique linear functional $f \in V^*$ such that $f(e_i) = y_i$ for all i .

Proof. Every vector $v \in V$ can be written uniquely in the form $v = \sum_i x_i e_i$, where $x_i \in K$. For any $f \in V^*$, we have

$$f(v) = \sum_i x_i f(e_i),$$

hence the values $f(e_i) = y_i$ determine f uniquely. Conversely, given $y \in K^n$, we define $f: V \rightarrow K$ by the formula

$$f(v) = \sum_i x_i y_i, \quad v = \sum_i x_i e_i \in V.$$

Then f is a linear functional and $f(e_i) = y_i$. □

Theorem 3.4. Let V be a vector space with a basis $\mathbf{e} = (e_1, \dots, e_n)$. For every $1 \leq j \leq n$, define $e_j^* \in V^*$ by the formula

$$e_j^*(e_i) = \delta_{ij}, \quad 1 \leq i \leq n.$$

Then $(e_1^*, \dots, e_n^*)$ is a basis of V^* , called the **dual basis** for the basis $\mathbf{e}$.

Proof. (1) The linear functionals $e_j^* \in V^*$ are uniquely determined by the previous lemma.

(2) Given $f \in V^*$, let $y_i = \langle f, e_i \rangle$ for all i . Then $g = \sum_j y_j e_j^*$, satisfies

$$\langle g, e_i \rangle = \sum_j y_j \langle e_j^*, e_i \rangle = \sum_j y_j \delta_{ij} = y_i = \langle f, e_i \rangle.$$

This implies that for any $v = \sum_i x_i e_i$, we have

$$\langle g, v \rangle = \sum_i x_i \langle g, e_i \rangle = \sum_i x_i \langle f, e_i \rangle = \langle f, v \rangle,$$

hence $f = g = \sum_j y_j e_j^*$ and V^* is generated by $e_1^*, \dots, e_n^*$.

(3) If $\sum_j y_j e_j^* = 0$, then

$$0 = \left\langle \sum_j y_j e_j^*, e_i \right\rangle = \sum_j y_j \langle e_j^*, e_i \rangle = \sum_j y_j \delta_{ij} = y_i$$

for all i . This implies that $e_1^*, \dots, e_n^*$ are linearly independent, hence they form a basis of V^* . $\square$

Remark 3.5. For any $v \in V$, $f \in V^*$, we have

$$v = \sum_i \langle e_i^*, v \rangle e_i, \quad f = \sum_i \langle f, e_i \rangle e_i^*.$$

Example 3.6. Let $V = K^2$ be equipped with a basis $e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. The dual basis e_1^*, e_2^* consists of row vectors such that $e_i^* \cdot e_j = \delta_{ij}$. Let us define matrices $A = (e_1 | e_2)$ and $B = \begin{pmatrix} e_1^* \\ e_2^* \end{pmatrix}$ (with rows equal the row vectors e_i^*). Then we obtain $B \cdot A = I$, hence $B = A^{-1}$. In our example we have $A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$, hence $B = A^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$. We conclude that

$$e_1^* = \begin{pmatrix} 1 & 0 \end{pmatrix}, \quad e_2^* = \begin{pmatrix} -1 & 1 \end{pmatrix}.$$

Remark 3.7. The above theorem implies that the vector spaces V and V^* have the same dimension. Therefore they are isomorphic to each other. However, there is no canonical isomorphism between them (independent of a basis choice).

Lemma 3.8. Let $\phi: U \rightarrow V$ be a linear map. Then there exists a unique linear map $\phi^*: V^* \rightarrow U^*$ satisfying the condition

$$\langle \phi^*(f), u \rangle = \langle f, \phi(u) \rangle, \quad \forall f \in V^*, u \in U.$$

The map ϕ^* is called the **transpose** (or **dual**) of ϕ .

Proof. Uniqueness follows immediately from the above formula. For any $f \in V^*$, we define $\phi^*(f): U \rightarrow K$ by the formula $\phi^*(f)(u) = \langle f, \phi(u) \rangle$, for $u \in U$. We verify that $\phi^*(f)$ is a linear functional, hence $\phi^*(f) \in U^*$. Finally, we verify that the map $V^* \rightarrow U^*$, $f \mapsto \phi^*(f)$ is linear. $\square$

Theorem 3.9. Let $\phi: U \rightarrow V$ be a linear map, $\mathbf{e}, \mathbf{f}$ be the bases of U and V respectively, and $A = [\phi]_{\mathbf{e}, \mathbf{f}}$ be the corresponding matrix of ϕ . Then the matrix of $\phi^*: V^* \rightarrow U^*$ with respect to the dual bases $\mathbf{f}^*, \mathbf{e}^*$ is equal to the transpose matrix A^T .

Proof. Let B be the matrix of $\phi^*: V^* \rightarrow U^*$ with respect to the dual bases. We have $\phi(e_j) = \sum_k a_{kj} f_k$ and $\phi^*(f_i^*) = \sum_k b_{ki} e_k^*$. Therefore

$$\langle \phi^*(f_i^*), e_j \rangle = \sum_k b_{ki} \langle e_k^*, e_j \rangle = b_{ji} = \langle f_i^*, \phi(e_j) \rangle = \sum_k a_{kj} \langle f_i^*, f_k \rangle = a_{ij}.$$

This implies $b_{ji} = a_{ij}$, hence $B = A^T$. $\square$

Let $V^{**} = (V^*)^*$ be the dual space of V^* (the **double dual** of V). We mentioned earlier that V and V^* are not canonically isomorphic. But it turns out that V and V^{**} are (if V is finite-dimensional).

Theorem 3.10. *Let V be a (finite-dimensional) vector space. Then there is a (canonical) linear map*

$$J: V \rightarrow V^{**}, \quad J(v)(f) = f(v), \quad v \in V, f \in V^*,$$

which is an isomorphism. We say that V is a reflexive space.

Proof. One can verify that $J(v): V^* \rightarrow K$ is linear, hence $J(v) \in V^{**}$. One can also verify that the map $V \rightarrow V^{**}$, $v \mapsto J(v)$ is linear. We know that $\dim V^{**} = \dim V^* = \dim V$. Therefore to show that J is an isomorphism, it is enough to show J is injective or that $\text{Ker } J = 0$. If $v = \sum_i x_i e_i \in \text{Ker } J$, then $f(v) = J(v)(f) = 0$ for all $f \in V^*$. Therefore $0 = \langle e_j^*, v \rangle = \sum_i x_i \langle e_j^*, e_i \rangle = x_j$ for all j . This implies that $v = 0$, hence $\text{Ker } J = 0$. $\square$

3.2. Bilinear forms.

Definition 3.11. Let V be a vector space over K . A **bilinear form** on V is a map

$$\sigma: V \times V \rightarrow K$$

such that, for all $v_1, v_2, w \in V$, $c_1, c_2 \in K$, we have

- (1) $\sigma(c_1v_1 + c_2v_2, w) = c_1\sigma(v_1, w) + c_2\sigma(v_2, w)$ (linearity in the first argument).
- (2) $\sigma(w, c_1v_1 + c_2v_2) = c_1\sigma(w, v_1) + c_2\sigma(w, v_2)$ (linearity in the second argument).

It is called **symmetric** if $\sigma(v, w) = \sigma(w, v)$, for all $v, w \in V$.

Remark 3.12. Given vector spaces V, W , one can similarly define a bilinear form

$$\sigma: V \times W \rightarrow K.$$

Example 3.13. The dot product on K^n

$$\sigma(x, y) = \sum_{i=1}^n x_i y_i = x^\top \cdot y, \quad x, y \in K^n,$$

is a symmetric bilinear form. More generally, for any $n \times n$ matrix A , define a bilinear form on K^n by

$$\sigma(x, y) = \sum_{i,j} a_{ij} x_i y_j = x^\top A y.$$

Note that for the standard basis of K^n we have

$$\sigma(e_i, e_j) = e_i^\top A e_j = e_i^\top \begin{pmatrix} a_{1j} \\ \vdots \\ a_{nj} \end{pmatrix} = a_{ij}.$$

Definition 3.14. Let σ be a bilinear form on V and $\mathbf{e}$ be a basis of V . Define the matrix $A = [\sigma]_{\mathbf{e}}$ of σ with respect to the basis $\mathbf{e}$ to be the matrix with entries $a_{ij} = \sigma(e_i, e_j)$.

Remark 3.15. If σ is a symmetric bilinear form, then its matrix A is **symmetric**, meaning that $a_{ij} = a_{ji}$ (or $A^\top = A$). Indeed, $a_{ij} = \sigma(e_i, e_j) = \sigma(e_j, e_i) = a_{ji}$.

Lemma 3.16. Let $A = [\sigma]_{\mathbf{e}}$ be the matrix of a bilinear form σ on V with respect to a basis $\mathbf{e}$. Then for all vectors $u, v \in V$, we have

$$\sigma(u, v) = [u]_{\mathbf{e}}^\top \cdot A \cdot [v]_{\mathbf{e}},$$

where $[v]_{\mathbf{e}}$ is the coordinate vector of v with respect to the basis $\mathbf{e}$.

Proof. Let $x = [u]_{\mathbf{e}} \in K^n$ and $y = [v]_{\mathbf{e}} \in K^n$. Then $u = \sum_i x_i e_i$, $v = \sum_i y_i e_i$ and

$$\sigma(u, v) = \sum_{i,j} x_i y_j \sigma(e_i, e_j) = \sum_{i,j} x_i a_{ij} y_j = \sum_i x_i (Ay)_i = x^\top Ay.$$

□

Theorem 3.17. Let $\sigma: V \times V \rightarrow K$ be a bilinear form, $\mathbf{e}, \mathbf{f}$ be two bases of V and $A = [\sigma]_{\mathbf{e}}$, $B = [\sigma]_{\mathbf{f}}$ be the corresponding matrices of σ . Then

$$B = M^\top A M,$$

where $M = M_{\mathbf{f} \rightarrow \mathbf{e}}$ is the transition matrix.

Proof. We know that $[v]_{\mathbf{e}} = M \cdot [v]_{\mathbf{f}}$ (see Lemma 1.4). Therefore

$$\sigma(u, v) = [u]_{\mathbf{e}}^\top A [v]_{\mathbf{e}} = (M[u]_{\mathbf{f}})^\top A (M[v]_{\mathbf{f}}) = [u]_{\mathbf{f}}^\top \cdot (M^\top A M) \cdot [v]_{\mathbf{f}}$$

On the other hand $\sigma(u, v) = [u]_{\mathbf{f}}^\top B [v]_{\mathbf{f}}$ for all $u, v \in V$. We conclude that $B = M^\top A M$. □

Definition 3.18. A bilinear form $\sigma: V \times V \rightarrow K$ is called **non-degenerate** (in the first argument) if $\sigma(v, w) = 0$ for all $w \in V$ implies $v = 0$. Similarly, we define non-degenerate forms in the second argument.

Theorem 3.19. Let $\sigma: V \times V \rightarrow K$ be a bilinear form and A be its matrix with respect to some basis. Then following are equivalent

- (1) σ is non-degenerate in the first argument.
- (2) σ is non-degenerate in the second argument.
- (3) A is invertible.

We call σ **non-degenerate** under any of these conditions.

Proof. It is enough to show that (2) and (3) are equivalent. Let $A = [\sigma]_{\mathbf{e}}$ be the matrix of σ with respect to a basis $\mathbf{e}$. Define a linear map

$$\phi = \sigma': V \rightarrow V^*, \quad v \mapsto \sigma(-, v) \in V^*,$$

where $\sigma(-, v): V \rightarrow K, u \mapsto \sigma(u, v)$. The matrix B of the map ϕ with respect to the basis $\mathbf{e}$ of V and the dual basis $\mathbf{e}^*$ of V^* satisfies $\sigma(-, e_j) = \phi(e_j) = \sum_k b_{kj} e_k^*$. Therefore

$$a_{ij} = \sigma(e_i, e_j) = \sum_k b_{kj} \langle e_i, e_k^* \rangle = b_{ij}$$

and $A = B$. Assume that σ is non-degenerate (in the second argument). If $v \in \text{Ker } \phi$, then $\phi(v) = \sigma(-, v) = 0$, hence $\sigma(u, v) = 0$ for all $u \in V$. Therefore $v = 0$. This implies that $\text{Ker } \phi = 0$ and ϕ is injective. As $\dim V^* = \dim V$, we conclude that ϕ is an isomorphism. Therefore $A = B$ is invertible. The converse statement goes through the same lines. $\square$

Corollary 3.20. Let $\sigma: V \times V \rightarrow K$ be a non-degenerate bilinear form. Then, for every linear functional $f: V \rightarrow K$, there exists a unique $v \in V$ such that $f = \sigma(-, v)$, meaning that $f(u) = \sigma(u, v)$ for all $u \in V$.

Proof. We have seen that the map $\sigma': V \rightarrow V^*, v \mapsto \sigma(-, v)$ is an isomorphism. Therefore, for any $f \in V^*$, there exists a unique $v \in V$ such that $f = \sigma'(v) = \sigma(-, v)$. $\square$

3.3. Euclidean vector spaces. In this section we will denote our bilinear form by $(-, -)$.

Definition 3.21. A bilinear form $(-, -)$ on a real vector space V is called **positive definite** if

$$(v, v) > 0 \quad \forall v \in V \setminus \{0\}.$$

A real $n \times n$ matrix A is called **positive definite** if $x^T A x > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$.

Definition 3.22. An **inner product** on a real vector space V is a map

$$(-, -): V \times V \rightarrow \mathbb{R}$$

satisfying (for all $u, v, u_1, u_2 \in V, c_1, c_2 \in \mathbb{R}$)

- (1) $(u, v) = (v, u)$ (symmetry).
- (2) $(c_1 u_1 + c_2 u_2, v) = c_1(u_1, v) + c_2(u_2, v)$ (linearity in the first argument).
- (3) $(v, v) > 0$ for all $v \in V \setminus \{0\}$ (positive definiteness).

Equivalently, it is a bilinear, symmetric, positive definite form on V . A real vector space equipped with an inner product is called a **Euclidean vector space**.

Example 3.23. Consider the dot product on $\mathbb{R}^n$

$$(x, y) = \sum_{i=1}^n x_i y_i, \quad x, y \in \mathbb{R}^n$$

It is a symmetric, bilinear form. Moreover, $(x, x) = \sum_i x_i^2 > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$. Therefore this form is positive definite, hence it is an inner product on $\mathbb{R}^n$.

Remark 3.24. Note that an inner product is non-degenerate as $(v, v) \neq 0$ for all $v \in V \setminus \{0\}$.

Definition 3.25. Given a Euclidean vector space V , we define the **length** (or **norm**) of a vector $v \in V$ to be

$$\|v\| = \sqrt{(v, v)}.$$

A vector of length 1 is called a **unit vector**.

Definition 3.26. Let V be a Euclidean vector space.

- (1) A collection of vectors $(v_1, \dots, v_n)$ in V is called **orthogonal** if $(v_i, v_j) = 0$ for $i \neq j$.
- (2) A collection of vectors $(v_1, \dots, v_n)$ in V is called **orthonormal** if it is orthogonal and $\|v_i\| = 1$. Equivalently, $(v_i, v_j) = \delta_{ij}$ for all i, j .
- (3) In the same way we define **orthogonal** and **orthonormal bases** of V .

Example 3.27. Let $V = \mathbb{R}^n$ be equipped with the dot product. Then the standard basis $(e_1, \dots, e_n)$ is an orthonormal basis as $(e_i, e_j) = \delta_{ij}$ for all i, j . But there are other orthogonal and orthonormal bases in V . For example, let $V = \mathbb{R}^2$ and let $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Then $(v_1, v_2) = 0$, hence these vectors form an orthogonal basis.

This basis is not orthonormal as $\|v_1\| = \|v_2\| = \sqrt{2}$. To get an orthonormal basis, we consider

$$f_1 = \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad f_2 = \frac{v_2}{\|v_2\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Then $\|f_i\| = \frac{\|v_i\|}{\|v_i\|} = 1$ and $(f_i, f_j) = \delta_{ij}$ for all i, j .

Lemma 3.28. Let $\mathbf{e} = (e_1, \dots, e_n)$ be an orthogonal basis of V . Then, for any vector $v \in V$, we have

$$v = \sum_i x_i e_i, \quad x_i = \frac{(v, e_i)}{(e_i, e_i)}.$$

If $\mathbf{e}$ is an orthonormal basis, then $v = \sum_i (v, e_i) e_i$.

Proof. Let $v = \sum_j x_j e_j$ for some $x \in \mathbb{R}^n$. Then $(v, e_i) = \sum_j x_j (e_j, e_i) = x_i (e_i, e_i)$, hence $x_i = \frac{(v, e_i)}{(e_i, e_i)}$. If $\mathbf{e}$ is orthonormal, then $(e_i, e_i) = \|e_i\|^2 = 1$, hence $x_i = (v, e_i)$. $\square$

Theorem 3.29 (Gram-Schmidt orthogonalization process). *Any Euclidean vector space V has an orthonormal basis. Given a basis $(v_1, \dots, v_n)$ of V , we define $w_1 = \frac{v_1}{\|v_1\|}$ and then inductively*

$$w_{k+1} = \frac{u}{\|u\|}, \quad u = v_{k+1} - \sum_{i=1}^k (v_{k+1}, w_i) w_i.$$

Then $(w_1, \dots, w_n)$ is an orthonormal basis such that $w_k \in \text{span}(v_1, \dots, v_k)$, for all k .

Proof. Let $(v_1, \dots, v_n)$ be any basis of V . We will construct an orthonormal basis $(w_1, \dots, w_n)$ of V having the property that every w_k is a linear combination of $v_1, \dots, v_k$. First we set $w_1 = \frac{v_1}{\|v_1\|}$ so that $\|w_1\| = 1$. Assuming that vectors $w_1, \dots, w_k$ are already constructed, let us consider a vector of the form

$$u = v_{k+1} - \sum_{i=1}^k c_i w_i, \quad c_i \in K.$$

Then

$$(u, w_i) = (v_{k+1}, w_i) - c_i (w_i, w_i) = (v_{k+1}, w_i) - c_i, \quad 1 \leq i \leq k.$$

If we set $c_i = (v_{k+1}, w_i)$, then $(u, w_i) = 0$ for $1 \leq i \leq k$. The vector u is nonzero as otherwise $v_{k+1} = \sum_{i=1}^k c_i w_i \in \text{span}(v_1, \dots, v_k)$ contradicting to the assumption that $(v_1, \dots, v_n)$ are linearly independent. We define $w_{k+1} = \frac{u}{\|u\|}$ so that $\|w_{k+1}\| = 1$ and

$$(w_{k+1}, w_i) = \frac{1}{\|u\|} (u, w_i) = 0, \quad 1 \leq i \leq k.$$

This implies that vectors $w_1, \dots, w_{k+1}$ are orthonormal. Assuming that $\text{span}(w_1, \dots, w_k) = \text{span}(v_1, \dots, v_k)$ (by induction), we obtain that

$$w_{k+1} \in \text{span}(v_{k+1}, w_1, \dots, w_k) = \text{span}(v_1, \dots, v_{k+1}), \quad v_{k+1} \in \text{span}(w_1, \dots, w_{k+1}).$$

Therefore $\text{span}(w_1, \dots, w_{k+1}) = \text{span}(v_1, \dots, v_{k+1})$. Continuing this procedure we obtain an orthonormal basis $(w_1, \dots, w_n)$ of V . $\square$

Remark 3.30. Alternatively, we can define first an orthogonal basis $(w_1, \dots, w_n)$ by the formulas $w_1 = v_1$ and

$$w_{k+1} = v_{k+1} - \sum_{i=1}^k \frac{(v_{k+1}, w_i)}{(w_i, w_i)} w_i.$$

Then we get an orthonormal basis by considering vectors $\frac{w_i}{\|w_i\|}$. In this way we avoid taking square roots in the process of orthogonalization.

Example 3.31. Let $V = \mathbb{R}^3$ (we will use row vectors) and let $v_1 = (1, 1, 0)$, $v_2 = (1, 1, 1)$, $v_3 = (1, 0, 1)$. We define

$$w_1 = v_1 = (1, 1, 0), \quad w_2 = v_2 - \frac{(v_2, w_1)}{(w_1, w_1)} w_1 = (0, 0, 1),$$

$$w_3 = v_3 - \frac{(v_3, w_1)}{(w_1, w_1)} w_1 - \frac{(v_3, w_2)}{(w_2, w_2)} w_2 = \frac{1}{2}(1, -1, 0).$$

One can check that w_1, w_2, w_3 are indeed orthogonal. The corresponding orthonormal basis is $\frac{1}{\sqrt{2}}(1, 1, 0)$, $(0, 0, 1)$, $\frac{1}{\sqrt{2}}(1, -1, 0)$.

Theorem 3.32 (Bessel's inequality). *Let $e_1, \dots, e_n$ be an orthonormal sequence of vectors in V (meaning that $(e_i, e_j) = \delta_{ij}$). For any $v \in V$, we have*

$$\sum_i (v, e_i)^2 \leq \|v\|^2.$$

Proof. Let $a_i = (v, e_i)$. Then

$$\begin{aligned} 0 &\leq \left(v - \sum_i a_i e_i, v - \sum_i a_i e_i \right) = (v, v) - 2 \sum_i a_i (v, e_i) + \sum_{i,j} a_i a_j (e_i, e_j) \\ &= (v, v) - 2 \sum_i a_i^2 + \sum_i a_i^2 = (v, v) - \sum_i a_i^2. \end{aligned}$$

Therefore $\sum_i a_i^2 \leq (v, v) = \|v\|^2$. □

Theorem 3.33 (Cauchy-Schwarz inequality). *We have*

$$(v, w) \leq \|v\| \cdot \|w\|.$$

Proof. We can assume that $w \neq 0$. Then the vector $e = w/\|w\|$ is a unit vector. By the previous result $(v, w/\|w\|)^2 = (v, e)^2 \leq \|v\|^2$, hence $(v, w)^2 \leq \|v\|^2 \cdot \|w\|^2$. □

Remark 3.34. We define the **angle** $\theta \in [0, \pi]$ between two nonzero vectors $v, w \in V$ by the formula

$$\cos \theta = \frac{(v, w)}{\|v\| \cdot \|w\|}.$$

Note that the right hand side is contained in $[-1, 1]$ by the Cauchy-Schwarz inequality.

3.4. Unitary vector spaces. Recall, that given a complex number $z = a + bi \in \mathbb{C}$ (with $a, b \in \mathbb{R}$), we define the **complex conjugate** $\bar{z} = a - bi$.

Definition 3.35. An **inner product** on a complex vector space V is a map

$$(-, -): V \times V \rightarrow \mathbb{C}$$

satisfying (for all $u, v, u_1, u_2 \in V, c_1, c_2 \in \mathbb{C}$)

- (1) $(u, v) = \overline{(v, u)}$ (conjugate symmetry).
- (2) $(c_1 u_1 + c_2 u_2, v) = c_1(u_1, v) + c_2(u_2, v)$ (linearity in the first argument).
- (3) $(v, v) > 0$ for all $v \in V \setminus \{0\}$ (positive definiteness).

A complex vector space equipped with an inner product is called a **unitary vector space**.

Remark 3.36. Our earlier definition of an inner product on a real vector space can be formulated using the above three axioms, where the conjugate sign is ignored.

Example 3.37. Define a pairing on $V = \mathbb{C}^n$ by the formula

$$(x, y) = \sum_{i=1}^n x_i \bar{y}_i, \quad x, y \in \mathbb{C}^n.$$

Note that $(y, x) = \sum_i y_i \bar{x}_i = \sum_i \overline{x_i \bar{y}_i} = \overline{(x, y)}$, hence the first axiom is satisfied. The pairing is linear in the first variable, hence the second axiom is satisfied. Finally, $(x, x) = \sum_i x_i \bar{x}_i = \sum_i |x_i|^2 > 0$ if $x \in \mathbb{C}^n \setminus \{0\}$. Therefore the third axiom is satisfied. This implies that the above pairing is an inner product.

Remark 3.38. A map satisfying the first two axioms is called a **Hermitian form**. Note that

$$(v, c_1 u_1 + c_2 u_2) = \overline{(c_1 u_1 + c_2 u_2, v)} = \bar{c}_1 \overline{(u_1, v)} + \bar{c}_2 \overline{(u_2, v)} = \bar{c}_1 (v, u_1) + \bar{c}_2 (v, u_2).$$

Therefore such map satisfies

- (1) $(c_1 u_1 + c_2 u_2, v) = c_1(u_1, v) + c_2(u_2, v)$.
- (2) $(v, c_1 u_1 + c_2 u_2) = \bar{c}_1(v, u_1) + \bar{c}_2(v, u_2)$.

A map satisfying these two properties is called a **sesquilinear form**. An inner product is a sesquilinear form that is conjugate symmetric and positive definite.

Remark 3.39. Given a vector space V over $\mathbb{C}$, define the **complex conjugate space** $\bar{V}$ to be the set V with the same vector addition and with the new scalar multiplication

$$\lambda \circ v = \bar{\lambda} v, \quad \lambda \in \mathbb{C}, v \in \bar{V}.$$

Then a sesquilinear form $\sigma: V \times V \rightarrow \mathbb{C}$ corresponds to a bilinear form $\sigma: V \times \bar{V} \rightarrow \mathbb{C}$ as

$$\sigma(u, \lambda \circ v) = \sigma(u, \bar{\lambda} v) = \lambda \sigma(u, v).$$

Remark 3.40. Given a unitary vector space, one can define orthogonal and orthonormal bases, apply Gram-Schmidt orthogonalization process, prove Bessel and Cauchy-Schwartz inequalities in the same way as for Euclidean spaces.

Theorem 3.41 (Bessel's inequality). *Let $e_1, \dots, e_n$ be an orthonormal sequence of vectors in a unitary vector space V . Then, for any $v \in V$, we have*

$$\sum_i |(v, e_i)|^2 \leq \|v\|^2.$$

Theorem 3.42 (Cauchy-Schwarz inequality). *For any vectors $v, w \in V$, we have*

$$|(v, w)| \leq \|v\| \cdot \|w\|.$$

3.5. Orthogonal complement. Let V be a Euclidean vector space (or a unitary vector space).

Definition 3.43. Given a subspace $U \subset V$, define its **orthogonal complement**

$$U^\perp = \{v \in V \mid (v, u) = 0 \ \forall u \in U\}.$$

Lemma 3.44. For any subspace $U \subset V$, the orthogonal complement $U^\perp \subset V$ is a subspace.

Proof. If $v, v' \in U^\perp$, then $(v + v', u) = (v, u) + (v', u) = 0$ for all $u \in U$. Therefore $v + v' \in U^\perp$. If $c \in K$, then $(cv, u) = c(v, u) = 0$ for all $u \in U$. Therefore $cv \in U^\perp$. $\square$

Theorem 3.45. For any subspace $U \subset V$, we have $V = U \oplus U^\perp$.

Proof. If $v \in U \cap U^\perp$, then $(v, v) = 0$, hence $v = 0$. This implies that $U \cap U^\perp = \{0\}$. Let $(u_1, \dots, u_k)$ be an orthonormal basis of U . For any $v \in V$, consider $u = \sum_i (v, u_i)u_i \in U$ and $w = v - u = v - \sum_i (v, u_i)u_i$. Then

$$(w, u_j) = (v, u_j) - \sum_i (v, u_i)(u_i, u_j) = (v, u_j) - (v, u_j) = 0, \quad \forall 1 \leq j \leq k.$$

This implies that $(w, u') = 0$ for all $u' \in U$, hence $w \in U^\perp$. Therefore $v = u + w \in U + U^\perp$, hence $V = U + U^\perp$. We conclude that $V = U \oplus U^\perp$. $\square$

Remark 3.46. Given a symmetric non-degenerate bilinear form σ on a vector space V and a subspace $U \subset V$, we can similarly define $U^\perp = \{v \in V \mid \sigma(v, u) = 0 \ \forall u \in U\}$ and prove that $V = U \oplus U^\perp$.

Remark 3.47. Let $V = \mathbb{R}^n$ be equipped with the dot product and let $U = \text{span}(u_1, \dots, u_k) \subset V$ be a subspace. Then $x \in U^\perp$ if and only if $u_i^\top \cdot x = 0$ for all i . Let A be a $k \times n$ matrix with the rows $u_i^\top$:

$$A = (u_1 | \dots | u_k)^\top.$$

Then we can write the above condition as $A \cdot x = 0$. Therefore

$$U^\perp = \text{Ker } A.$$

Example 3.48. Let $V = \mathbb{R}^3$ be equipped with the dot product and let

$$U = \text{span}\left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}\right) \subset V.$$

To find $U^\perp$, we consider $A = (u_1 | u_2)^\top = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix}$. Then $U^\perp = \text{Ker } A$. Consider

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix} \xrightarrow{r_2 - r_1} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \xrightarrow{r_1 - r_2} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Fixing the free variable $x_3 = 1$, we get $x_1 = 1$, $x_2 = -1$. Therefore $U^\perp = \text{Ker } A = \text{span}\left(\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}\right)$.

3.6. Self-adjoint operators. Let V be a Euclidean or a unitary vector space.

Lemma 3.49. *Given a linear operator $A: V \rightarrow V$, there exists a unique linear operator $A^*: V \rightarrow V$ satisfying*

$$(Au, v) = (u, A^*v) \quad \forall u, v \in V.$$

*It is called the **adjoint operator** of A .*

Proof. For any $v \in V$, consider the linear functional $f: V \rightarrow K, u \mapsto (Au, v)$. By Corollary 3.20, there exists a unique vector $A^*v \in V$ such that $f = (-, A^*v)$. Then $f(u) = (Au, v) = (u, A^*v)$ for all $u \in V$. This defines a map $A^*: V \rightarrow V$ satisfying the required relation and one can check that A^* is linear. If $B: V \rightarrow V$ is another linear operator satisfying the required relation, then $(Au, v) = (u, A^*v) = (u, Bv)$, hence $(u, A^*v - Bv) = 0$ for all $u \in V$. Therefore $A^*v - Bv = 0$ and we conclude that $A^* = B$. This proves uniqueness. $\square$

Remark 3.50. Note that $(A^*v, u) = \overline{(u, A^*v)} = \overline{(Au, v)} = (v, Au)$. Therefore $(A^*)^* = A$.

Definition 3.51. A linear operator $A: V \rightarrow V$ is called **self-adjoint** if $A^* = A$. Equivalently

$$(Au, v) = (u, Av) \quad \forall u, v \in V.$$

Example 3.52. Let $V = \mathbb{R}^n$ be equipped with the dot product. Given an $n \times n$ matrix A , consider the corresponding linear map $A: \mathbb{R}^n \rightarrow \mathbb{R}^n$ and its adjoint operator A^* . Then

$$(A^*e_i, e_j) = \sum_k (a_{ki}^* e_k, e_j) = a_{ji}^* = (e_i, Ae_j) = \sum_k (e_i, a_{kj} e_k) = \sum_k a_{kj} (e_i, e_k) = a_{ij}.$$

This implies that the matrix of the adjoint operator A^* is the transpose matrix $A^\top$. The linear operator A is self-adjoint if and only if the corresponding matrix is symmetric ($A = A^\top$).

Example 3.53. Let $V = \mathbb{C}^n$ be equipped with the inner product $(x, y) = \sum_i x_i \bar{y}_i$. Given an $n \times n$ matrix A , consider the corresponding linear map $A: \mathbb{C}^n \rightarrow \mathbb{C}^n$ and its adjoint operator A^* . Then

$$(A^*e_i, e_j) = \sum_k (a_{ki}^* e_k, e_j) = a_{ji}^* = (e_i, Ae_j) = \sum_k (e_i, a_{kj} e_k) = \sum_k \bar{a}_{kj} (e_i, e_k) = \bar{a}_{ij}.$$

This implies that the matrix of the adjoint operator A^* is equal to $(\bar{A})^\top = \bar{A}^\top$, called the **conjugate transpose** of A . The linear operator A is self-adjoint if and only if the corresponding matrix satisfies $A = \bar{A}^\top$. Such matrices are called **Hermitian** or self-adjoint.

Lemma 3.54. *Let V be a unitary vector space and $A: V \rightarrow V$ be a self-adjoint operator. Then*

- (1) *All eigenvalues of A are real.*
- (2) *If $u, v \in V$ are eigenvectors of A corresponding to two distinct eigenvalues, then u, v are orthogonal, meaning that $(u, v) = 0$.*

Proof. (1) Assume that $Av = \lambda v$ for some $\lambda \in \mathbb{C}$ and $0 \neq v \in V$. Then

$$\lambda(v, v) = (\lambda v, v) = (Av, v) = (v, Av) = (v, \lambda v) = \bar{\lambda}(v, v).$$

Therefore $\bar{\lambda} = \lambda$ and $\lambda \in \mathbb{R}$.

(2) Let $Au = \lambda u$ and $Av = \mu v$ for some $\lambda \neq \mu$. We have seen that $\lambda, \mu \in \mathbb{R}$. Therefore

$$(Au, v) = (\lambda u, v) = \lambda(u, v) = (u, Av) = (u, \mu v) = \mu(u, v).$$

This implies that $(\lambda - \mu)(u, v) = 0$, hence $(u, v) = 0$. $\square$

Example 3.55. Let $A = \begin{pmatrix} 3 & -1 \\ -1 & 3 \end{pmatrix}$. Then $\chi_A(t) = t^2 - 6t + 8 = (t - 2)(t - 4)$. The eigenvalues are $\lambda = 2$ and $\lambda = 4$. For $\lambda = 2$, we consider $A - 2I = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ and find the eigenvector $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. For $\lambda = 4$, we consider $A - 4I = \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}$ and find the eigenvector $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. We observe that $(v_1, v_2) = 1 - 1 = 0$, hence v_1, v_2 are orthogonal.

Corollary 3.56. Let A be a real symmetric matrix (or a Hermitian matrix). Then

- (1) All eigenvalues of A are real.
- (2) If u, v are eigenvectors of A corresponding to two distinct eigenvalues, then u, v are orthogonal, meaning that $(u, v) = 0$.

Proof. The matrix A defines a linear operator $\mathbb{C}^n \rightarrow \mathbb{C}^n$ which is self-adjoint. Now we apply the previous result. $\square$

Theorem 3.57 (Spectral theorem). Let V be a unitary vector space and $A: V \rightarrow V$ be a self-adjoint operator. Then there exists an orthonormal basis of V consisting of eigenvectors of A (the matrix of A with respect to this basis is diagonal).

Proof. Let $\lambda \in \mathbb{R}$ be an eigenvalue of A . Then $U = \text{Ker}(A - \lambda)$ is A -invariant. The subspace $U^\perp \subset V$ is also A -invariant: for every $v \in U^\perp$ and $u \in U$, we have $(Av, u) = (v, Au) = 0$ as $Au \in U$. Therefore $Av \in U^\perp$. We proved earlier that $V = U \oplus U^\perp$. By induction, the restriction $A|_{U^\perp}$ is diagonalizable in an orthonormal basis of $U^\perp$. We extend this basis by an orthonormal basis of U . Then A is diagonal in this basis. $\square$

Remark 3.58. Let V be a unitary vector space. A linear operator $A: V \rightarrow V$ is called **normal** if $A^*A = AA^*$. For example, self-adjoint operators ($A^* = A$) and unitary operators ($A^*A = I$, see below) are normal. One can prove a spectral theorem for normal operators (existence of an orthonormal basis of V consisting of eigenvectors of A) in the same way as Theorem 3.57.

Remark 3.59. Let $\mathbf{e}, \mathbf{f}$ be orthonormal bases and $M = M_{\mathbf{f} \rightarrow \mathbf{e}}$. Then $f_j = \sum_k m_{kj} e_k$. Therefore

$$\delta_{ij} = (f_i, f_j) = \sum_{k,l} m_{ki} m_{lj} (e_k, e_l) = \sum_k m_{ki} m_{kj} = (M^\top M)_{ij}$$

and $M^\top M = I$. Such matrices M are called **orthogonal**. Note that $M_{\mathbf{e} \rightarrow \mathbf{f}} = M^{-1} = M^\top$.

Theorem 3.60 (Spectral theorem). Let A be a real symmetric matrix. Then there exists an orthogonal matrix M such that $M^{-1}AM = M^\top AM$ is diagonal.

Proof. The matrix A defines a linear operator $A: \mathbb{C}^n \rightarrow \mathbb{C}^n$ which is self-adjoint. By the previous theorem, there exists an orthonormal basis $\mathbf{f}$ of $\mathbb{C}^n$ such that $[A]_{\mathbf{f}}$ is diagonal. The vectors f_i are eigenvectors of A corresponding to real eigenvalues λ_i . Therefore $f_i \in \text{Ker}(A - \lambda_i)$ and we can choose them to be real vectors. Let $\mathbf{e}$ be the standard basis of $\mathbb{C}^n$ (which is orthonormal). Then $[A]_{\mathbf{f}} = M^{-1}AM$, where $M = M_{\mathbf{f} \rightarrow \mathbf{e}}$ is an orthogonal matrix by the previous remark. $\square$

Example 3.61. In the previous example we had an orthogonal basis $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. We make it orthonormal by considering $f_1 = v_1 / \|v_1\| = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $f_2 = v_2 / \|v_2\| = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Then $M = (f_1 | f_2) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ is an orthogonal matrix: $M^\top M = I$. One can check that $M^\top AM = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}$.

3.7. Orthogonal operators and matrices. Let V be a Euclidean vector space.

Definition 3.62. A linear operator $A: V \rightarrow V$ is called **orthogonal** if

$$(Au, Av) = (u, v) \quad \forall u, v \in V.$$

Remark 3.63. In the case of a unitary vector space V , a linear operator $A: V \rightarrow V$ satisfying the above condition is called a **unitary operator**.

Theorem 3.64. *Given a linear operator $A: V \rightarrow V$, the following are equivalent*

- (1) A is orthogonal.
- (2) $\|Av\| = \|v\|$ for all $v \in V$.
- (3) $A^*A = I$.

Proof. (1) $\implies$ (2). We have $\|Av\|^2 = (Av, Av) = (v, v) = \|v\|^2$.

(2) $\implies$ (1). If $\|Av\| = \|v\|$, then $(Av, Av) = (v, v)$. This implies

$$2(u, v) = (u+v, u+v) - (u, u) - (v, v) = (A(u+v), A(u+v)) - (Au, Au) - (Av, Av) = 2(Au, Av).$$

(1) $\implies$ (3). We have $(u, v) = (Au, Av) = (u, A^*Av)$ for all u, v . Therefore $A^*Av = v$, hence $A^*A = I$.

(3) $\implies$ (1). We have $(Au, Av) = (u, A^*Av) = (u, Iv) = (u, v)$. □

Remark 3.65. Let $V = \mathbb{R}^n$ be equipped with the dot product. An $n \times n$ matrix A defines a linear operator $L_A: \mathbb{R}^n \rightarrow \mathbb{R}^n$. Then the adjoint operator L_A^* corresponds to the transpose A^T . Therefore L_A is an orthogonal operator if and only if $A^T A = A A^T = I$. We called such matrices **orthogonal** in Remark 3.59.

Remark 3.66. Let $A = (f_1 | \dots | f_n)$. Then the entries of $A^T A$ have the form $f_i^T f_j$. Therefore A is orthogonal if and only if $(f_i, f_j) = f_i^T f_j = \delta_{ij}$. This means that $(f_1, \dots, f_n)$ is an orthonormal basis. A similar condition can be formulated for the rows of A .

Lemma 3.67. *If A is an orthogonal matrix, then $\det A = \pm 1$.*

Proof. We have $A^T A = I$, hence $1 = \det I = \det(A^T A) = \det(A^T) \cdot \det(A) = \det(A)^2$. Therefore $\det(A) = \pm 1$. □

Example 3.68. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be an orthogonal matrix with $\det A = 1$. Then $a^2 + c^2 = b^2 + d^2 = 1$ and $ab + cd = 0$. The last condition implies that $(b, d) = \lambda(-c, a)$ for some $\lambda \in \mathbb{R}$. Then $\det A = ad - bc = \lambda(a^2 + c^2) = \lambda$, hence $\lambda = 1$. Therefore $(b, d) = (-c, a)$ and we have just one condition $a^2 + c^2 = 1$. There exists $\theta \in [0, 2\pi)$ such that $a = \cos \theta$, $c = \sin \theta$. Then $A = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}$ (compare to Example 2.12). Every operator of this form is orthogonal.

Note that for general θ , the matrix A can not be diagonalized (over real numbers). On the other hand, we can consider A as a unitary operator (matrix) and then we can diagonalize A over complex numbers according to Remark 3.58.

3.8. Quadratic forms.

Definition 3.69. Let V be a real vector space. A map $q: V \rightarrow \mathbb{R}$ is called a **quadratic form** if there exists a basis $(e_1, \dots, e_n)$ of V and scalars $a_{ij} \in \mathbb{R}$ such that

$$q(x_1 e_1 + \dots + x_n e_n) = \sum_{i,j} a_{ij} x_i x_j.$$

Remark 3.70. If the above condition holds for one basis of V , then it holds for any basis. Indeed, let $\mathbf{e}, \mathbf{f}$ be two bases of V and let A be a matrix such that $q(v) = \sum_{i,j} a_{ij} x_i x_j = x^\top A x$, for $v = \sum_i x_i e_i$ (or $x = [v]_{\mathbf{e}}$). We have $x = My$, where $y = [v]_{\mathbf{f}}$ and $M = M_{\mathbf{f} \rightarrow \mathbf{e}}$ is the transition matrix. Therefore

$$q(v) = x^\top A x = (My)^\top A (My) = y^\top (M^\top A M) y = y^\top B y, \quad B = M^\top A M.$$

This means that $q(y_1 f_1 + \dots + y_n f_n) = \sum_{i,j} b_{ij} y_i y_j$.

Example 3.71. Let $\sigma: V \times V \rightarrow \mathbb{R}$ be a bilinear form. Then $q(v) = \sigma(v, v)$ is a quadratic form. Indeed, given a basis $\mathbf{e} = (e_1, \dots, e_n)$, consider the matrix $A = [\sigma]_{\mathbf{e}}$ with entries $a_{ij} = \sigma(e_i, e_j)$. Then

$$\sigma(v, w) = \sum_{i,j} a_{ij} x_i y_j, \quad v = \sum_i x_i e_i, \quad w = \sum_i y_i e_i.$$

In particular, $q(v) = \sigma(v, v) = \sum_{i,j} a_{ij} x_i x_j$, for $v = \sum_i x_i e_i$. If σ is symmetric, then we can reconstruct it from q by the formula

$$\sigma(v, w) = \frac{1}{2} (q(v + w) - q(v) - q(w)).$$

This gives a 1-1 correspondence between quadratic forms and symmetric bilinear forms.

Remark 3.72. In what follows we will consider quadratic forms on $\mathbb{R}^n$ of the form

$$q(x) = x^\top A x = (A x, x),$$

where A is a symmetric matrix and $(x, y) = \sum_i x_i y_i$ is the dot product. Note that there is a 1-1 correspondence between quadratic forms, symmetric bilinear forms, symmetric matrices and self-adjoint operators. Therefore we can apply our earlier results about symmetric matrices (and self-adjoint operators) to the study of quadratic forms.

Example 3.73. Consider the quadratic form on $\mathbb{R}^4$ (Minkowski space)

$$q(x_1, x_2, x_3, t) = x_1^2 + x_2^2 + x_3^2 - t^2.$$

We will see later that there is no change of variables (of the form $x = My$ for some invertible matrix M) such that the above quadratic form takes the form $\sum_i y_i^2$.

Theorem 3.74 (Canonical form). *Let $q(x) = x^\top A x$, for a real symmetric $n \times n$ matrix A .*

- (1) *There exists an orthogonal change of variables $x = My$ (meaning that M is an orthogonal matrix) such that $q(x) = \sum_{i,j} a_{ij} x_i x_j = \sum_i \lambda_i y_i^2$, where λ_i are (real) eigenvalues of A .*
- (2) *There exists a change of variables $x = Mz$ (meaning that M is invertible) such that $q(x) = \sum_{i,j} a_{ij} x_i x_j = \sum_i \varepsilon_i z_i^2$, where $\varepsilon_i \in \{1, -1, 0\}$.*

Proof. (1) By the spectral theorem, there exists an orthogonal matrix M such that $M^\top A M = \text{diag}(\lambda_1, \dots, \lambda_n)$, where λ_i are the eigenvalues of A . With $x = My$, we obtain $q(x) = (My)^\top A (My) = y^\top (M^\top A M) y = \sum_i \lambda_i y_i^2$.

(2) Using the same notation as before, let us define new variables $z_i = \begin{cases} \sqrt{|\lambda_i|} \cdot y_i, & \lambda_i \neq 0, \\ y_i, & \lambda_i = 0. \end{cases}$

Then $\lambda_i y_i^2 = \varepsilon_i z_i^2$, where

$$\varepsilon_i = \operatorname{sgn}(\lambda_i) = \begin{cases} 1, & \lambda_i > 0, \\ 0, & \lambda_i = 0, \\ -1, & \lambda_i < 0. \end{cases}$$

Therefore $q(x) = \sum_i \lambda_i y_i^2 = \sum_i \varepsilon_i z_i^2$. $\square$

Example 3.75. Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and $q(x) = x^T A x = 2x_1 x_2$. Then there are eigenvalues $\lambda_1 = 1$, $\lambda_2 = -1$ and orthonormal eigenvectors $f_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $f_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Consider the transition matrix $M = (f_1 | f_2) = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ and new coordinates $x = My$, hence $x_1 = \frac{1}{\sqrt{2}}(y_1 + y_2)$, $x_2 = \frac{1}{\sqrt{2}}(y_1 - y_2)$. Then the quadratic form in new coordinates is

$$2x_1 x_2 = \frac{2}{2}(y_1 + y_2)(y_1 - y_2) = y_1^2 - y_2^2.$$

Theorem 3.76 (Sylvester's law of inertia). *Let A be a real symmetric $n \times n$ matrix and*

$$q(x) = x^T A x = \sum_i \varepsilon_i y_i^2,$$

*for some change of variables $x = My$ and $\varepsilon_i \in \{1, -1, 0\}$. Then the numbers n_+, n_-, n_0 of positive, negative, zero values among ε_i are independent of the choice of variables. They are equal respectively to the number of positive, negative, zero eigenvalues of A . The triple (n_+, n_-, n_0) is called the **inertia** (or the **signature**) of the quadratic form q (or the matrix A).*

Proof. We have to show that given invertible matrices M_1, M_2 such that

$$M_1^T A M_1 = D_1 = \operatorname{diag}(\lambda_1, \dots, \lambda_n), \quad M_2^T A M_2 = D_2 = \operatorname{diag}(\mu_1, \dots, \mu_n),$$

the numbers of positive, negative, zero values among λ_i and μ_i are the same. We will do this for positive values. Assume that

- (1) $\lambda_i > 0$ for $1 \leq i \leq k$ and $\lambda_i \leq 0$ for $k < i \leq n$,
- (2) $\mu_i > 0$ for $1 \leq i \leq l$ and $\mu_i \leq 0$ for $l < i \leq n$

and assume that $l > k$ without loss of generality. We have

$$D_2 = M_2^T ((M_1^T)^{-1} D_1 M_1^{-1}) M_2 = M^T D_1 M, \quad M = M_1^{-1} M_2.$$

Let $(e_1, \dots, e_n)$ be the standard basis and let $f_i = M e_i$. Then

$$e_i^T D_1 e_i = \lambda_i \leq 0, \quad k < i \leq n$$

and $e_i^T D_1 e_j = 0$ for $i \neq j$, hence $x^T D_1 x \leq 0$ for $x \in U = \operatorname{span}(e_{k+1}, \dots, e_n)$. On the other hand

$$f_i^T D_1 f_i = (M e_i)^T D_1 (M e_i) = e_i^T D_2 e_i = \mu_i > 0, \quad 1 \leq i \leq l.$$

and similarly $f_i^T D_1 f_j = 0$ for $i \neq j$. Therefore $x^T D_1 x > 0$ for $0 \neq x \in V = \operatorname{span}(f_1, \dots, f_l)$. This implies that $U \cap V = 0$. But $\dim U + \dim V = (n - k) + l > n$, hence $U \cap V \neq 0$. This is a contradiction. $\square$

Remark 3.77. If A in the above theorem is an invertible matrix (meaning that the corresponding bilinear form is non-degenerate), then $n_0 = 0$ and $n_+ + n_- = n$. This implies that the triple (n_+, n_-, n_0) is uniquely determined by the value $n_+ - n_-$, usually called the **signature** of the quadratic form (or the matrix A). For example, the quadratic form $q(x) = x_1^2 + x_2^2 + x_3^2 - x_4^2$ has signature 2.

3.9. Positive definiteness. Recall that a real symmetric $n \times n$ matrix A is called **positive definite** if $x^T Ax > 0$ for all $x \in \mathbb{R}^n \setminus \{0\}$. Given such a matrix, the corresponding bilinear form $\sigma(x, y) = x^T Ay$ is an inner product on $\mathbb{R}^n$. Therefore it is important to give a characterization of such matrices.

Definition 3.78. Let A be an $n \times n$ matrix. We define a **minor of order k** of A to be the determinant of a $k \times k$ submatrix of A (obtained by removing $n - k$ rows and $n - k$ columns from A). If the submatrix is the square upper-left submatrix of A , then the corresponding minor is called a **corner (principal) minor** of A .

Example 3.79. Let $A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \end{pmatrix}$. Then the corner minors of A are 1 and $1 \cdot 3 - 2 \cdot 2 = -1$.

Theorem 3.80. Let A be a real symmetric $n \times n$ matrix. Then the following are equivalent

- (1) A is positive definite.
- (2) All eigenvalues of A are positive.
- (3) The corner minors of A are positive (Sylvester's criterion).

Proof. (1) $\implies$ (2) and (2) $\implies$ (1). We know that there exists an orthogonal matrix M such that $M^T AM = \text{diag}(\lambda_1, \dots, \lambda_n)$, where λ_i are eigenvalues of A . Every vector $x \in \mathbb{R}^n$ can be written as $x = My$ for some $y \in \mathbb{R}^n$. Then

$$x^T Ax = (My)^T AMy = y^T (M^T AM)y = \sum_i \lambda_i y_i^2.$$

If A is positive definite, we consider $y = e_i$, $x = Ay$ and obtain $\lambda_i = x^T Ax > 0$. If all $\lambda_i > 0$, then we obtain $\sum_i \lambda_i y_i^2 > 0$ for all $y \neq 0$. Therefore A is positive definite.

(1) $\implies$ (3). Let A_k be the left top $k \times k$ submatrix of A . Considering $x \in \mathbb{R}^n$ with $x_{k+1} = \dots = x_n = 0$, we obtain $x^T Ax = \sum_{1 \leq i, j \leq k} a_{ij} x_i x_j$, hence A_k is positive definite. This implies that all eigenvalues of A_k are positive, hence $\det A_k > 0$.

(3) $\implies$ (1). We use induction on the size of A . Let A be an $(n+1) \times (n+1)$ matrix. Then we can write it in the form $A = \begin{pmatrix} A_n & v \\ v^T & \lambda \end{pmatrix}$, where $v \in K^n$ is a column vector of dimension n and $\lambda \in K$. Consider a matrix $M = \begin{pmatrix} I_n & x \\ 0 & 1 \end{pmatrix}$, where $x \in K^n$ is to be determined. Then

$$M^T AM = \begin{pmatrix} I_n & 0 \\ x^T & 1 \end{pmatrix} \begin{pmatrix} A_n & v \\ v^T & \lambda \end{pmatrix} \begin{pmatrix} I_n & x \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} A_n & v \\ * & * \end{pmatrix} \begin{pmatrix} I_n & x \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} A_n & A_n x + v \\ * & * \end{pmatrix}$$

We have $\det A_n > 0$, hence A_n is invertible and we can find $x \in K^n$ such that $A_n x + v = 0$. The matrix $M^T AM$ is symmetric as $(M^T AM)^T = M^T A^T M = M^T AM$. Therefore $M^T AM = \begin{pmatrix} A_n & 0 \\ 0 & \mu \end{pmatrix}$ for some $\mu \in K$. Note that

$$\det(M^T AM) = \det A_n \cdot \mu = (\det M)^2 \cdot \det A = \det A.$$

Therefore $\mu > 0$. By induction A_n is positive definite and has positive eigenvalues, hence $M^T AM$ has positive eigenvalues and is positive definite. For every $0 \neq x \in K^n$, we can write $x = My$, for some $0 \neq y \in K^n$, and then $x^T Ax = y^T (M^T AM)y > 0$. This implies that A is positive definite. $\square$

Example 3.81. Consider the matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. Its corner minors are 2 and $2 \cdot 1 - 1 \cdot 1 = 1$. Both are positive, hence A is positive definite. We can also prove this directly:

$$x^T Ax = 2x_1^2 + 2x_1x_2 + x_2^2 = x_1^2 + (x_1 + x_2)^2 > 0$$

for $x \in \mathbb{R}^2 \setminus \{0\}$.

3.10. Applications.

3.10.1. *Minimum of quadratic functions.* A quadratic function $f(x) = ax^2 + bx + c$ with $a > 0$ attains its minimum at the extremal point satisfying $2ax + b = 0$, hence $x = -\frac{1}{2}a^{-1}b$. One can formulate a similar result in higher dimensions.

Theorem 3.82. *Consider a quadratic function*

$$f: \mathbb{R}^n \rightarrow \mathbb{R}, \quad x \mapsto x^\top Ax + b^\top x + c,$$

where A is a symmetric positive definite $n \times n$ matrix and $b, c \in \mathbb{R}^n$. Then f attains its unique minimum at $x = -\frac{1}{2}A^{-1}b$.

Proof. If $x_0 = -\frac{1}{2}A^{-1}b$, then $b = -2Ax_0$ and $f(x_0) = x_0^\top Ax_0 - 2x_0^\top Ax_0 + c = -x_0^\top Ax_0 + c$. Therefore

$$f(x) - f(x_0) = (x^\top Ax - 2x^\top Ax_0 + c) - (-x_0^\top Ax_0 + c) = (x - x_0)^\top A(x - x_0) > 0$$

if $x \neq x_0$. Therefore $f(x) > f(x_0)$ for all $x \neq x_0$. $\square$

3.10.2. *Min/Max of quadratic functions over spheres.*

Theorem 3.83. *Consider the function $f: \mathbb{R}^n \rightarrow \mathbb{R}$, $x \mapsto x^\top Ax$, where A is a symmetric $n \times n$ matrix. Then the minimal/maximal values of $f(x)$ over the sphere $\|x\| = 1$ are the minimal/maximal eigenvalues of A . These values are attained at the (unit) eigenvectors.*

Proof. There exists an orthogonal matrix M such that $M^\top AM = \text{diag}(\lambda_1, \dots, \lambda_n)$. If $x = My$, then $(x, x) = x^\top x = (My)^\top (My) = y^\top (M^\top M)y = y^\top y = (y, y)$. Therefore $\|x\| = \|y\|$. On the other hand $f(x) = x^\top Ax = \sum_i \lambda_i y_i^2$. Assuming that $\lambda_1 \leq \dots \leq \lambda_n$, we have

$$\sum_i \lambda_i y_i^2 \leq \lambda_n \sum_i y_i^2 = \lambda_n \|y\|^2 = \lambda_n$$

if $\|y\| = 1$. The equality holds for $y = e_n$. Note that $M = M_{\mathbf{f} \rightarrow \mathbf{e}} = (f_1 | \dots | f_n)$, where f_i are the (orthonormal) eigenvectors of A . Therefore the maximal value is attained at $x = My = Me_n = f_n$ which is an eigenvector. The statement about the minimal value is similar. $\square$

Example 3.84. Let us find the min/max values of $f(x_1, x_2) = x_1^2 + 4x_1x_2 + 4x_2^2$ over $x_1^2 + x_2^2 = 1$. We have $f(x) = x^\top Ax$, where $A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$. Then $\chi_A(t) = t^2 - 5t$ and the eigenvalues are $\lambda_1 = 0$, $\lambda_2 = 5$. The eigenvectors are $v_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 2 \\ -1 \end{pmatrix}$, $v_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$. These are the vectors where respectively the minimum 0 and the maximum 5 are attained. For example $f(v_1) = v_1^\top Av = 0$

3.10.3. *Second derivative test.* Let $f(x_1, \dots, x_n)$ be a differentiable function in n variables.

(1) A point $p \in \mathbb{R}^n$ is called **critical** if $\frac{\partial f}{\partial x_i}(p) = 0$ for all i .

(2) Given a critical point p , define the **Hessian matrix** $A = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(p) \right)_{i,j}$.

We can approximate $f(x)$ at $x = p + h$ with $\|h\| \ll 1$ as (Taylor's expansion)

$$f(x) \sim f(p) + \sum_i \frac{\partial f}{\partial x_i}(p)h_i + \frac{1}{2} \sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(p)h_i h_j = \frac{1}{2}h^\top Ah + f(p).$$

We have seen that if A is positive definite, then the quadratic function on the right attains its minimum at $h = -A^{-1}b = 0$ (as $b = 0$ in our setting). This corresponds to $x = p + h = p$. We obtain, for a critical point p , that

- (1) If the Hessian matrix A at p is positive definite, then p is a local minimum.
- (2) Similarly, if the Hessian matrix A at p is negative definite (meaning that $-A$ is positive definite or that A has negative eigenvalues), then p is a local maximum.
- (3) If the Hessian matrix A has positive and negative eigenvalues, then p is a saddle point.