Algebra 2

Chenglong Yu

February 5, 2026

Contents

1 Introduction

The course roughly covers the following three parts. You may refer to course website of Math 371 Spring 2020 at UPenn in my personal page for related materials.

Part I: Bilinar forms. Symmetric forms, Hermitian forms, and skew-symmetric forms. Orthogonality. Spectral Theorem. Conics and Quadrics. Key examples of classical groups, and their basic properties. Lie algebra (for such groups).

Part II: Group representations. Irreducible representations and unitary representations. Characters. Schur's Lemma. Modules over principal ideal domains. Free modules. Group rings. Noetherian rings. Structure of Abelian groups. Maschke's theorem. Constructions of representations, et cetera.

Part III: Field extensions, algebraic extensions and algebraic closures, splitting fields, separable and inseparable extensions, Galois extensions, Galois correspondences, cyclotomic extensions, solvability by radicals, et cetera.

2 Bilinear Forms

2.1 Basic Definitions

Definition 2.1 (Bilinear Form)**.** *Let V be a vector space over a field $\mathbb{K}$. A map*

$$\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{K}$$

is called a ***bilinear form*** *if it is linear in both components. That is:*

1. $\langle a\boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w} \rangle = a\langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$
2. $\langle \boldsymbol{u}, a\boldsymbol{v} + \boldsymbol{w} \rangle = a\langle \boldsymbol{u}, \boldsymbol{v} \rangle + \langle \boldsymbol{u}, \boldsymbol{w} \rangle$

for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$ and $a \in \mathbb{K}$.

Definition 2.2 (Symmetric and Skew-Symmetric Forms)**.** *A bilinear form $\langle \cdot, \cdot \rangle$ is called:*

1. ***Symmetric*** *if $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{w}, \boldsymbol{v} \rangle$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$.*
2. ***Skew-symmetric*** *(or alternating) if $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = -\langle \boldsymbol{w}, \boldsymbol{v} \rangle$ for all $\boldsymbol{v}, \boldsymbol{w} \in V$.*

There are following examples of bilinear forms.

Example 2.1 (Euclidean Space)**.** *Let $V = \mathbb{R}^n$. The* ***standard inner product*** *defined by*

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{st} = \sum_{i=1}^{n} x_i y_i = \boldsymbol{x}^T \boldsymbol{y}$$

is a symmetric bilinear form. It allows us to define the length of vectors and the angle between non-zero vectors.

Example 2.2 (Minkowski Space)**.** *Let $V = \mathbb{R}^{n+1}$. The* ***Lorentz form*** *is defined by*

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = x_1 y_1 + \cdots + x_{n-1} y_{n-1} - x_{n+1} y_{n+1}.$$

This is a symmetric bilinear form used in special relativity.

Example 2.3 (Matrix Space)**.** *Let $V = M_{m \times n}(\mathbb{R})$. Define*

$$\langle A, B \rangle = \operatorname{tr}(A^T B).$$

This is a symmetric bilinear form on the space of matrices.

For infinite-dimensional spaces, we have the following example.

Example 2.4. *Let V be the space of continuous real-valued functions on $[0, 1]$. Define*

$$\langle f, g \rangle = \int_0^1 f(x) g(x)\, dx.$$

This is a symmetric bilinear form on V.

2.2 Gram Matrices and Congruency

Next we consider finite-dimensional vector spaces V over a field $\mathbb{K}$ with $\dim V = n < \infty$. Let $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be a basis of V.

Definition 2.3 (Gram Matrix)**.** *The* ***Gram matrix*** *of a bilinear form $\langle \cdot, \cdot \rangle$ with respect to the basis $\mathcal{B}$ is the matrix $G_{\langle \cdot, \cdot \rangle, \mathcal{B}} \in M_n(\mathbb{K})$ defined by:*

$$(G_{\langle \cdot, \cdot \rangle, \mathcal{B}})_{ij} = \langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle.$$

By expansion of bilinearity, we have the following important property.

Proposition 2.1 (Matrix Representation of Bilinear Forms)**.** *If $\boldsymbol{u} = \sum x_i \boldsymbol{v}_i$ and $\boldsymbol{w} = \sum y_j \boldsymbol{v}_j$ are vectors in V with coordinate vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{K}^n$, then the value of the bilinear form can be computed via matrix multiplication:*

$$\langle \boldsymbol{u}, \boldsymbol{w} \rangle = \boldsymbol{x}^T G_{\langle \cdot, \cdot \rangle, \mathcal{B}} \boldsymbol{y}. \tag{1}$$

In fact the formula (1) can be used to define bilinear forms from arbitrary matrices A.

So if we consider the space of bilinear forms on V, it has a natural structure of $\mathbb{K}$-vector space stucture and it is isomorphic to the space of $n \times n$ matrices over $\mathbb{K}$. To summarize, we have the following proposition.

Proposition 2.2 (Matrix representation of bilinear forms)**.** *Let $\mathbf{Bil}(V)$ denote the space of bilinear forms on V. Then $\mathbf{Bil}(V)$ is a vector space over $\mathbb{K}$, and the map*

$$\begin{array}{rcl} \mathbf{Bil}(V) & \to & M_n(\mathbb{K}) \\ \langle \cdot, \cdot \rangle & \mapsto & G_{\langle \cdot, \cdot \rangle, \mathcal{B}} \end{array}$$

is a vector space isomorphism between the space of bilinear forms on V and the space of $n \times n$ matrices over $\mathbb{K}$.

The symmetric and skew-symmetric bilinear forms correspond to symmetric and skew-symmetric matrices, respectively.

Proposition 2.3. *Let $\langle \cdot, \cdot \rangle$ be a bilinear form on V and $A = G_{\langle \cdot, \cdot \rangle, \mathcal{B}}$ is the Gram matrix of $\langle \cdot, \cdot \rangle$ with respect to the basis $\mathcal{B}$. Then:*

1. *$\langle \cdot, \cdot \rangle$ is symmetric if and only if A is a symmetric matrix, i.e. $A = A^T$.*
2. *$\langle \cdot, \cdot \rangle$ is skew-symmetric if and only if A is a skew-symmetric matrix, i.e. $A = -A^T$.*

The dependence of Gram matrices on the choice of basis is described as follows.

Proposition 2.4 (Change of Basis)**.** *Let $\mathcal{B} = \{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ and $\mathcal{B}' = \{\boldsymbol{w}_1, \cdots, \boldsymbol{w}_n\}$ be two bases of V. Let P be the transition matrix from $\mathcal{B}$ to $\mathcal{B}'$ (i.e., $\boldsymbol{w}_j = \sum_i P_{ij} \boldsymbol{v}_i$). Then the Gram matrices are related by:*

$$G_{\langle \cdot, \cdot \rangle, \mathcal{B}'} = P^T G_{\langle \cdot, \cdot \rangle, \mathcal{B}} P.$$

Proof. By definition:

$$\begin{aligned} (G_{\langle \cdot, \cdot \rangle, \mathcal{B}'})_{jk} = \langle \boldsymbol{w}_j, \boldsymbol{w}_k \rangle &= \langle \sum_i P_{ij} \boldsymbol{v}_i, \sum_l P_{lk} \boldsymbol{v}_l \rangle \\ &= \sum_i \sum_l P_{ij} \langle \boldsymbol{v}_i, \boldsymbol{v}_l \rangle P_{lk} \\ &= \sum_i \sum_l (P^T)_{ji} (G_{\langle \cdot, \cdot \rangle, \mathcal{B}})_{il} P_{lk} \\ &= (P^T G_{\langle \cdot, \cdot \rangle, \mathcal{B}} P)_{jk}. \end{aligned}$$

□

Definition 2.4 (Congruency)**.** *Two square matrices A and B are called* ***congruent*** *if there exists an invertible matrix P such that $B = P^T AP$. It is straightforward to verify that congruency is an equivalence relation on the set of square matrices.*

Since all the invertible $n \times n$ matrices can appear as the change of basis matrix for an n-dimensional vector space, so two matrices are congruent if and only if they represent the same bilinear form under two bases.

Remark 2.1. *Coordinate change of bilinear forms corresponds to matrix congruency, whereas linear operators change coordinates via similarity ($P^{-1}AP$).*

Definition 2.5 (Isometry)**.** *Let $(V_1, \langle\cdot,\cdot\rangle_1)$ and $(V_2, \langle\cdot,\cdot\rangle_2)$ be vector spaces equipped with bilinear forms. A linear map $f : V_1 \to V_2$ is called an* ***isometry*** *if*

$$\langle f(\boldsymbol{u}), f(\boldsymbol{v})\rangle_2 = \langle \boldsymbol{u}, \boldsymbol{v}\rangle_1$$

for all $\boldsymbol{u}, \boldsymbol{v} \in V_1$.

Theorem 2.1. *Two finite-dimensional $\mathbb{K}$-vector spaces with bilinear forms are isometric if and only if their Gram matrices (under any chosen bases) are congruent.*

3 Symmetric Forms

Throughout this subsection, assume $\langle\cdot,\cdot\rangle$ is a **symmetric** bilinear form on a vector space V over a field $\mathbb{K}$ (where $\mathrm{char}(\mathbb{K}) \neq 2$, i.e. 2 is invertible in $\mathbb{K}$).

3.1 Diagonalization of Gram Matrices

The assumption on the characteristic is necessary for the following polarization identity.

Proposition 3.1 (Polarization Identity)**.** *The bilinear form is completely determined by its quadratic form $q(\boldsymbol{v}) = \langle \boldsymbol{v}, \boldsymbol{v}\rangle$. Specifically:*

$$\langle \boldsymbol{v}, \boldsymbol{w}\rangle = \frac{1}{2}\left(\langle \boldsymbol{v}+\boldsymbol{w}, \boldsymbol{v}+\boldsymbol{w}\rangle - \langle \boldsymbol{v}, \boldsymbol{v}\rangle - \langle \boldsymbol{w}, \boldsymbol{w}\rangle\right).$$

This implies that if $\langle\cdot,\cdot\rangle$ is not identically zero, there must exist some vector $\boldsymbol{v}$ such that $\langle \boldsymbol{v}, \boldsymbol{v}\rangle \neq 0$.

Theorem 3.1 (Diagonalization / Orthogonal Basis)**.** *There exists a basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ of V such that the Gram matrix of $\langle\cdot,\cdot\rangle$ is diagonal. That is, $\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle = 0$ for $i \neq j$.*

Proof. We proceed by induction on $n = \dim V$.

1. **Base case:** If $\langle\cdot,\cdot\rangle \equiv 0$, any basis works. If $n = 1$, any basis works.
2. **Inductive step:** Assume the statement holds for dimensions $< n$. If $\langle\cdot,\cdot\rangle \equiv 0$, we are done. Otherwise, by the polarization identity, there exists $\boldsymbol{v}_1 \in V$ such that $\langle \boldsymbol{v}_1, \boldsymbol{v}_1\rangle \neq 0$ (such a vector is called non-isotropic).

   Define $W = \{\boldsymbol{w} \in V \mid \langle \boldsymbol{v}_1, \boldsymbol{w}\rangle = 0\}$. This is the orthogonal complement of the line spanned by $\boldsymbol{v}_1$. Consider the map $\phi : V \to \mathbb{K}$ given by $\boldsymbol{w} \mapsto \langle \boldsymbol{v}_1, \boldsymbol{w}\rangle$. Since $\langle \boldsymbol{v}_1, \boldsymbol{v}_1\rangle \neq 0$, the map is non-zero, hence surjective. Thus $\dim W = \dim(\ker \phi) = n - 1$.

We claim $V = \operatorname{span}(\boldsymbol{v}_1) \oplus W$. For any $\boldsymbol{v} \in V$, let

$$\boldsymbol{w} = \boldsymbol{v} - \frac{\langle \boldsymbol{v}, \boldsymbol{v}_1 \rangle}{\langle \boldsymbol{v}_1, \boldsymbol{v}_1 \rangle} \boldsymbol{v}_1.$$

Then a direct check shows $\langle \boldsymbol{w}, \boldsymbol{v}_1 \rangle = 0$, so $\boldsymbol{w} \in W$. Thus $\boldsymbol{v} \in \operatorname{span}(\boldsymbol{v}_1) + W$. The intersection is zero because $\boldsymbol{v}_1$ is not in W.

By the induction hypothesis, W admits an orthogonal basis $\{\boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$. Then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is an orthogonal basis for V.

□

Under this orthogonal basis, the Gram matrix is diagonal:

$$G = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix},$$

where $d_i = \langle \boldsymbol{v}_i, \boldsymbol{v}_i \rangle$.

Remark 3.1. *The proof of the diagonalization can also be obtained from the theory of nondegeneracy criterion on subspaces (see Excercise 4.3).*

3.2 Sylvester's Law of Inertia

When $\mathbb{K} = \mathbb{R}$, we can scale the basis vectors to normalize the diagonal entries coefficients to be 1, -1, or 0. More precisely, we choose the basis vectors as follows:

- If $d_i > 0$, replace $\boldsymbol{v}_i$ by $\frac{1}{\sqrt{d_i}} \boldsymbol{v}_i$.
- If $d_i < 0$, replace $\boldsymbol{v}_i$ by $\frac{1}{\sqrt{-d_i}} \boldsymbol{v}_i$.
- If $d_i = 0$, leave $\boldsymbol{v}_i$ unchanged.

After this scaling, the Gram matrix becomes diagonal with entries in $\{1, -1, 0\}$. In fact, these numbers are determined by the bilinear form itself, independent of the choice of basis.

Theorem 3.2 (Sylvester's Law of Inertia)**.** *Let $\langle \cdot, \cdot \rangle$ be a real symmetric bilinear form on V. There exists a basis under which the Gram matrix is diagonal with entries in $\{1, -1, 0\}$. Usually, the basis is ordered such that:*

$$G = \begin{pmatrix} I_p & 0 & 0 \\ 0 & -I_q & 0 \\ 0 & 0 & 0_{n-p-q} \end{pmatrix}.$$

Furthermore, the integers p (index of positivity) and q (index of negativity) are invariants depending only on $\langle \cdot, \cdot \rangle$, not on the choice of basis.

The triple $(p, q, n - p - q)$ is called the **signature** of the form.

Proof of Uniqueness. The coordinate transformation allows us to write any symmetric matrix A as congruent to a diagonal matrix with diagonal entries $d_i \in \{1, -1, 0\}$. Suppose we have two such decompositions yielding indices (p, q) and (p', q'). Consider the subspaces corresponding to the basis vectors:

- Basis $\mathcal{B} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ gives p positive, q negative terms. Let $V^+ = \operatorname{span}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p)$. Then $\dim V^+ = p$.
- Basis $\mathcal{B}' = \{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ gives p' positive, q' negative terms. Let $W^{\leq 0} = \operatorname{span}(\boldsymbol{w}_{p'+1}, \ldots, \boldsymbol{w}_n)$. Then $\dim W^{\leq 0} = n - p'$.

If $p > p'$, then $\dim V^+ + \dim W^{\leq 0} > n$. So the dimension of the intersection

$$\dim V^+ \cap W^{\leq 0} = \dim V^+ \dim W^{\leq 0} - \dim(V^+ + W^{\leq 0}) \geq \dim V^+ \dim W^{\leq 0} - \dim V > 0$$

Then there is an nonzero vector $\boldsymbol{x} \in V^+ \cap W^{\leq 0}$. Write $\boldsymbol{x} = \sum_{i=1}^{p} a_i v_i = \sum_{j=p'+1}^{n} b_j w_j$. Then $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \sum a_i^2 > 0$ (from V^+) and $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = -\sum_{j=p'+1}^{p'+q'} -b_j^2 \leq 0$ (from $W^{\leq 0}$). This is a contradiction. Thus $p \leq p'$. By symmetry, $p' \leq p$, so $p = p'$. A similar argument shows $q = q'$. □

Corollary 3.1. *Two real symmetric matrices A and B of order n are congruent if and only if they have the same positive index of inertia and negative index of inertia.*

Remark 3.2. *Even though the signature $(p, q, n - p - q)$ is unique, the specific orthogonal basis achieving this signature is not unique. The subspaces V^+ contributing the positive part (or the subspaces for the negative part) are not unique. But the subspace corresponding to the zero part is unique. Try to define this subspace intrinsically. It is called the* ***radical*** *of the form.*

3.3 Positive Definite Forms

In the proof of Sylvester's law of inertia, we have constructed subspaces where the quadratic form is shows positive and negative properties.

Definition 3.1 (Definiteness)**.** *Let V be a real vector space and $\langle \cdot, \cdot \rangle$ be a symmetric bilinear form.*

1. *$\langle \cdot, \cdot \rangle$ is* ***positive definite*** *(denoted $\langle \cdot, \cdot \rangle > 0$) if $\langle \boldsymbol{v}, \boldsymbol{v} \rangle > 0$ for all $\boldsymbol{v} \neq \boldsymbol{0}$.*
2. *$\langle \cdot, \cdot \rangle$ is* ***negative definite*** *(denoted $\langle \cdot, \cdot \rangle < 0$) if $\langle \boldsymbol{v}, \boldsymbol{v} \rangle < 0$ for all $\boldsymbol{v} \neq \boldsymbol{0}$.*
3. *$\langle \cdot, \cdot \rangle$ is* ***positive semi-definite*** *(denoted $\langle \cdot, \cdot \rangle \geq 0$) if $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \geq 0$ for all $\boldsymbol{v} \in V$.*
4. *$\langle \cdot, \cdot \rangle$ is* ***negative semi-definite*** *(denoted $\langle \cdot, \cdot \rangle \leq 0$) if $\langle \boldsymbol{v}, \boldsymbol{v} \rangle \leq 0$ for all $\boldsymbol{v} \in V$.*

Proposition 3.2. *Let V be a finite-dimensional real vector space with a symmetric form $\langle \cdot, \cdot \rangle$. The form $\langle \cdot, \cdot \rangle$ is positive definite if and only if its index of positivity p equals $\dim(V)$.*

The proof for the uniqueness of signature also shows the following **intrinsic** characterization.

Proposition 3.3 (Characterization of Signature)**.** *The positive index of inertia p of a symmetric form $\langle \cdot, \cdot \rangle$ on V can be characterized by:*

$$p = \max\{\dim W \mid W \subseteq V \text{ is a subspace where } \langle \cdot, \cdot \rangle|_W \text{ is positive definite}\}.$$

Similarly, the negative index q is the maximal dimension of a subspace where $\langle \cdot, \cdot \rangle$ is negative definite.

When the symmetric form is positive definite, we also call the corresponding Gram matrix a **positive definite matrix**. More properties of positive definite matrices are summarized in the exercises.

3.4 Euclidean Spaces

Definition 3.2 (Euclidean Space)**.** *A real vector space V equipped with a positive definite symmetric bilinear form $\langle \cdot, \cdot \rangle$ is called a* ***Euclidean space*** *or an* ***inner product space****.*

See Excercise 4.3 for the definition of nondegeneracy and that if $(V, \langle \cdot, \cdot \rangle)$ is an inner product space, then it is non-degenerate and all its subspaces are also non-degenerate.

Proposition 3.4. *Every Euclidean space of dimension n is isometric to the standard Euclidean space $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{st})$.*

Example 3.1 (Polynomial Space)**.** *Let $V = P_n(\mathbb{R}) = \{f(x) \in \mathbb{R}[x] \mid \deg f \leq n\}$ (sometimes denoted $\mathbb{R}[x]_{\leq n}$). Define the bilinear form:*

$$\langle f, g \rangle = \int_0^1 f(x)g(x)\,dx.$$

Since $\int_0^1 (f(x))^2\,dx > 0$ for any non-zero polynomial, $\langle \cdot, \cdot \rangle$ is positive definite. Consider the standard basis $\{1, x, x^2, \dots\}$. The Gram matrix G under this basis has entries:

$$G_{ij} = \langle x^{i-1}, x^{j-1} \rangle = \int_0^1 x^{i+j-2}\,dx = \frac{1}{i+j-1}.$$

This matrix is known as the ***Hilbert matrix****.*

3.5 Gram-Schmidt process and QR Decomposition

While Sylvester's theorem guarantees an orthogonal basis, in Euclidean spaces we can construct an **orthonormal** basis algorithmically from any given basis.

Definition 3.3 (Orthonormal Basis)**.** *A basis $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$ of a Euclidean space $(V, \langle \cdot, \cdot \rangle)$ is* ***orthonormal*** *if*

$$\langle \boldsymbol{w}_i, \boldsymbol{w}_j \rangle = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}.$$

In terms of matrices, an orthonormal basis corresponds to orthogonal matrices.

Definition 3.4 (Orthogonal Matrix)**.** *A square real matrix Q is called an* ***orthogonal matrix*** *if its column vectors form an orthonormal basis of $\mathbb{R}^n$ under the standard inner product. Equivalently, Q is orthogonal if and only if $Q^TQ = I$ or $QQ^T = I$, or all the row vectors of Q form an orthonormal basis of $\mathbb{R}^n$.*

Theorem 3.3 (Gram-Schmidt Process)**.** *Let $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ be an arbitrary basis of a Euclidean space V. One can construct an orthonormal basis $\{\boldsymbol{w}_1, \dots, \boldsymbol{w}_n\}$ such that*

$$\operatorname{span}(\boldsymbol{w}_1, \dots, \boldsymbol{w}_k) = \operatorname{span}(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$$

for all $k = 1, \dots, n$.

Algorithm. The construction proceeds inductively:

1. Set $\tilde{\boldsymbol{w}}_1 = \boldsymbol{v}_1$. Since basis vector are not zero, the inner product $\langle \tilde{\boldsymbol{w}}_1, \tilde{\boldsymbol{w}}_1 \rangle > 0$. Normalize it to obtain $\boldsymbol{w}_1$:
$$\boldsymbol{w}_1 = \frac{\tilde{\boldsymbol{w}}_1}{\sqrt{\langle \tilde{\boldsymbol{w}}_1, \tilde{\boldsymbol{w}}_1 \rangle}}.$$

2. Set $\tilde{\boldsymbol{w}}_2 = \boldsymbol{v}_2 - \langle \boldsymbol{v}_2, \boldsymbol{w}_1 \rangle \boldsymbol{w}_1$. This vector is satisfies $\langle \tilde{\boldsymbol{w}}_2, \boldsymbol{w}_1 \rangle = 0$. From the construction, we also know that
$$\operatorname{span}(\tilde{\boldsymbol{w}}_2, \boldsymbol{w}_1) = \operatorname{span}(\boldsymbol{v}_1, \boldsymbol{v}_2).$$
So $\tilde{\boldsymbol{w}}_2 \neq 0$. Normalize it:
$$\boldsymbol{w}_2 = \frac{\tilde{\boldsymbol{w}}_2}{\sqrt{\langle \tilde{\boldsymbol{w}}_2, \tilde{\boldsymbol{w}}_2 \rangle}}.$$

3. In general, for step k, define
$$\tilde{\boldsymbol{w}}_k = \boldsymbol{v}_k - \sum_{j=1}^{k-1} \langle \boldsymbol{v}_k, \boldsymbol{w}_j \rangle \boldsymbol{w}_j$$
Then
$$\langle \tilde{\boldsymbol{w}}_k, \boldsymbol{w}_i \rangle = 0 \text{ for } i < k,$$
Inductively we have
$$\operatorname{span}(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_{k-1}), \tilde{\boldsymbol{w}}_k = \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1}), \tilde{\boldsymbol{w}}_k = \operatorname{span}(\boldsymbol{v}_1, \cdots, \boldsymbol{v}_{k-1}, \boldsymbol{v}_k)$$
so $\tilde{\boldsymbol{w}}_k \neq 0$. Then normalize:
$$\boldsymbol{w}_k = \frac{\tilde{\boldsymbol{w}}_k}{\sqrt{\langle \tilde{\boldsymbol{w}}_k, \tilde{\boldsymbol{w}}_k \rangle}}.$$

□

To summarize, the Gram-Schmidt process constructs an orthonormal basis $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ from any given basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ such that the transition matrix from $\{\boldsymbol{w}_i\}$ to $\{\boldsymbol{v}_i\}$ is upper triangular with positive diagonal entries $\frac{1}{\sqrt{\langle \tilde{\boldsymbol{w}}_i, \tilde{\boldsymbol{w}}_i \rangle}}$. Upper triangularity follows from the fact that the subspaces spanned by the first k basis vectors are preserved. So the change of basis matrix has the form:
$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ 0 & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_{nn} \end{pmatrix},$$
where $p_{ii} = \frac{1}{\sqrt{\langle \tilde{\boldsymbol{w}}_i, \tilde{\boldsymbol{w}}_i \rangle}} > 0$. The transition matrix from the orthonormal basis $\{\boldsymbol{v}_i\}$ back to the original basis $\{\boldsymbol{w}_i\}$ is then given by P^{-1}, which is also upper triangular with positive diagonal entries.
$$(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n) = (\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_n) P^{-1}$$

In terms of matrices, this is called the **QR decomposition** of a matrix.

Definition 3.5 (QR Decomposition)**.** *If V is the standard Euclidean space $\mathbb{R}^n$, then any basis $(\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n)$ gather together to form an invertible matrix A. The orthonormal basis vectors $Q = (\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_n)$ form an orthogonal matrix. The decomposition above can be rewritten as*
$$A = QR,$$
*where $R = P^{-1}$ is an upper triangular matrix with positive diagonal entries. This is called the **QR decomposition** of the matrix A.*

The uniquess of the QR decomposition is stated in Problem 4.2

4 Exercises

4.1 Useful practices

Please submit solutions to the following problems in this section. Some problems help you to review the material we have learned, and some problems introduce useful concepts and theorems not covered in class.

Excercise 4.1. *Practice the Gram-Schmidt process and the QR decomposition. You can choose either one of the following two problems to solve.*

1. *Let $V = P_{\leq 2}(\mathbb{R})$ be the vector space of real polynomials with degree at most 2. Define an inner product on V by*
$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$$
Given basis $1, x, x^2$ for V, use the Gram-Schmidt process to find an orthonormal basis for V.

2. *Calculate the QR decomposition for the matrix*
$$A = \begin{pmatrix} 3 & 2 & 100 \\ 4 & 0 & 0 \\ 0 & 0 & -5 \end{pmatrix}.$$

Excercise 4.2. *Prove the uniqueness of the QR decomposition: if A is an $n \times n$ invertible real matrix then there exists a unique $n \times n$ orthogonal matrix Q and a unique $n \times n$ upper triangular matrix R with positive diagonal entries such that $A = QR$. (You only need to prove the uniqueness part; the existence part is given by the Gram-Schmidt process.)*

Excercise 4.3. *Let V be a finite-dimensional vector space over field F. Define a symmetric form on V to be a bilinear form $\langle\cdot,\cdot\rangle : V \times V \to F$ similar as the real case. We call the symmetric form **non-degenerate** if for any $\boldsymbol{v} \in V$, $\langle \boldsymbol{v}, \boldsymbol{w}\rangle = 0$ for all $\boldsymbol{w} \in V$ implies $\boldsymbol{v} = 0$.*

1. *Show that the symmetric form $\langle\cdot,\cdot\rangle$ is non-degenerate if and only if for some (and hence any) basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ of V, the Gram matrix $A = (\langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle)_{1\leq i,j\leq n}$ is invertible.*

2. *Assume V has a nondegenerate symmetric form $\langle\cdot,\cdot\rangle$. Let W be a subspace of V. Define the orthogonal complement of W to be*
$$W^\perp = \{\boldsymbol{v} \in V : \langle \boldsymbol{v}, \boldsymbol{w}\rangle = 0 \text{ for all } \boldsymbol{w} \in W\}.$$
Prove that $W \bigoplus W^\perp = V$ if and only if the restriction of $\langle\cdot,\cdot\rangle$ on W is non-degenerate.

3. *When $\mathbb{R}$ is the base field, show that a positive definite symmetric form is non-degenerate.*

4. *For an inner product space V over $\mathbb{R}$, show that for any subspace W of V, $W \bigoplus W^\perp = V$.*

5. *In four-dimensional Minkowski space with the Lorentz form, find an one-dimensional subspace W such that W and $W^\perp$ do not form a direct sum of the whole space.*

Excercise 4.4. *In this problem, you will prove the **Cauchy-Schwarz inequality for Euclidean spaces** in an inner product space V. The **norm of an inner product space** is defined by $\|\boldsymbol{u}\| = \sqrt{\langle \boldsymbol{u}, \boldsymbol{u}\rangle}$. The Cauchy-Schwarz inequality states that for any $\boldsymbol{u}, \boldsymbol{v} \in V$,*
$$|\langle \boldsymbol{u}, \boldsymbol{v}\rangle| \leq \|\boldsymbol{u}\| \cdot \|\boldsymbol{v}\|.$$

Moreover, equality holds if and only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent. You can choose any one of the following two methods to prove it.

1. *Assume $\boldsymbol{v} \neq 0$. Consider the quadratic function of λ*

$$f(\lambda) = \langle \boldsymbol{u} - \lambda \boldsymbol{v}, \boldsymbol{u} - \lambda \boldsymbol{v} \rangle$$

   Show that this function is non-negative and deduce the Cauchy-Schwarz inequality from this.Show that equality holds if and only if $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly dependent.

2. *There is another method to reduce the Cauchy-Schwarz inequality to two dimensional case. Assume $\boldsymbol{u}$ and $\boldsymbol{v}$ are linearly independent (otherwise the inequality is trivial). Let $W = \operatorname{span}\{\boldsymbol{u}, \boldsymbol{v}\}$. Show that the Cauchy-Schwarz inequality holds in V if it holds in W. Then prove the Cauchy-Schwarz inequality in two-dimensional inner product space by directly considering the standard inner product on $\mathbb{R}^2$.*

Excercise 4.5. *We call a symmetric matrix A **positive definite** if it is the Gram matrix of any positive definite symmetric form.*

1. *Prove that a symmetric matrix A is positive definite if and only if there exists an invertible matrix P such that $A = P^T P$.*

2. *Show that if A is positive definite, then its determinant is positive.*

3. *Prove that a two by two symmetric matrix is positive definite if and only if it has positive trace and positive determinant.*

Excercise 4.6. *In the following, you will prove the **criterion for positive definiteness by principal minors**. A **principal minor** of a matrix A is the determinant of a square submatrix obtained by deleting certain rows and the corresponding columns. A **leading principal minor** is a principal minor obtained by deleting the last $n - k$ rows and columns for some k. In the following, show that a symmetric matrix A is positive definite if and only if all its leading principal minors are positive.*

1. *Show that if A is positive definite, then all its principal minors are positive. (Hint: consider the restriction of the corresponding symmetric form on the subspace spanned by the first k basis vectors.)*

2. *Use induction to show that if all leading principal minors of A are positive, then matrix A is positive definite. (Hint: use problem 4.3 and induction.)*

Excercise 4.7. *Let A be a real symmetric matrix where the diagonal elements are all 2, the elements on the two sub-diagonals are all -1, and all other elements are 0. Prove that A is positive definite.*

Excercise 4.8. *Artin chapter 8 1.1. Show that a bilinear form $\langle , \rangle$ on a real vector space V is a sum of a symmetric form and a skew-symmetric form. (skew-symmetric means alternating)*

Excercise 4.9. *Let g be a bilinear form on a real vector space V. Prove that if g satisfies $g(\boldsymbol{x}, \boldsymbol{y}) = 0$ if and only if $g(\boldsymbol{y}, \boldsymbol{x}) = 0$, then g is either symmetric or alternating.*

4.2 Optional problems

If you would like to try some additional problems, you can find them here and you do not need to submit them.

Excercise 4.10. *Prove that the* ***Hilbert matrix*** *of order* n*,*

$$H_n = (\frac{1}{i+j-1})_{n\times n}$$

is a positive definite matrix. (Hint: Use the symmetric form in Problem 4.1 (1).)

Excercise 4.11. *Prove the* ***reversed Cauchy-Schwarz inequality in Minkowski space****:*
For all $\boldsymbol{v} = (v_0, v_1, \dots, v_n), \boldsymbol{w} = (w_0, w_1, \dots, w_n) \in \mathbb{R}^{n+1}$ *satisfying*

$$v_0^2 - v_1^2 - \cdots - v_n^2 > 0 \text{ and } w_0^2 - w_1^2 - \cdots - w_n^2 > 0,$$

prove the following inequality

$$(v_0^2 - v_1^2 - \cdots - v_n^2)(w_0^2 - w_1^2 - \cdots - w_n^2) \leq (v_0w_0 - v_1w_1 - \cdots - v_nw_n)^2$$

and determine the necessary and sufficient condition for equality to hold.

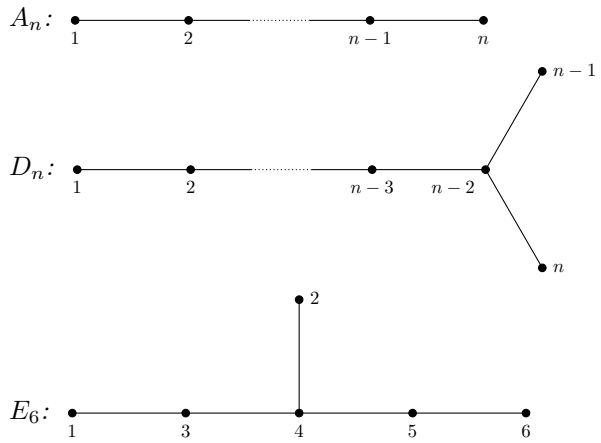
In terms of the Lorentz form $\langle \boldsymbol{v}, \boldsymbol{w}\rangle = v_0w_0 - v_1w_1 - \cdots - v_nw_n$*, the inequality can be rewritten as*

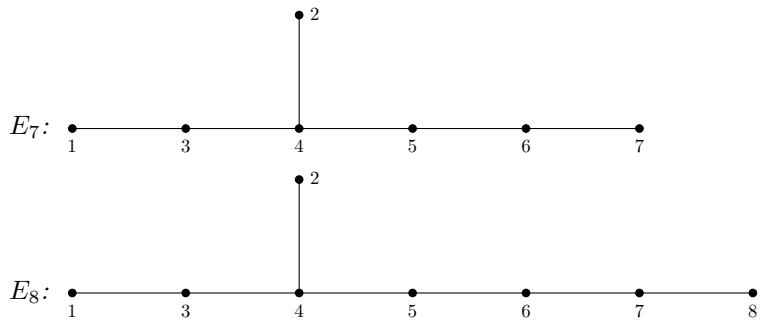
$$\langle \boldsymbol{v}, \boldsymbol{v}\rangle\langle \boldsymbol{w}, \boldsymbol{w}\rangle \leq \langle \boldsymbol{v}, \boldsymbol{w}\rangle^2.$$

when $\langle \boldsymbol{v}, \boldsymbol{v}\rangle > 0$ *and* $\langle \boldsymbol{w}, \boldsymbol{w}\rangle > 0$ *(in physics* $\boldsymbol{v}$ *and* $\boldsymbol{w}$ *are called time-like vectors and this implies two time-like vectors have positive product under the Lorentz form).*

Hint: Use similar method as in Problem 4.4 (2).

Excercise 4.12 (Challenge)**.** *In this problem, you will prove the* ***Cartan matrices*** *associated to* ***ADE Dynkin diagrams*** *are positive definite. For a graph* Γ *with vertices* $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$*, consider the* $n \times n$ *real symmetric matrix defined by* $A_\Gamma = (a_{ij})_{n\times n}$*, where* $a_{ij} = 2$ *when* $i = j$*,* $a_{ij} = -1$ *when* $i \neq j$ *and* $\boldsymbol{v}_i, \boldsymbol{v}_j$ *are adjacent (connected by an edge), and* $a_{ij} = 0$ *otherwise. Prove that for the following graphs* Γ*,* A_Γ *is positive definite:*





In fact, these graphs are exactly those connected and whose corresponding matrices are positive definite.

5 Geometry of Euclidean spaces: distance and projection

Cauchy–Schwartz inequality allows us to define angles and lengths in Euclidean spaces, which leads to the study of geometry in these spaces.

Definition 5.1. *Let $(V, \langle\cdot,\cdot\rangle)$ be a Euclidean space.*

1. *For any $\boldsymbol{v} \in V$, the **norm** (or length) of $\boldsymbol{v}$ is defined as*

$$\|\boldsymbol{v}\| = \sqrt{\langle \boldsymbol{v}, \boldsymbol{v}\rangle}.$$

2. *For any non-zero vectors $\boldsymbol{u}, \boldsymbol{v} \in V$, the **angle** $\theta \in [0, \pi]$ between $\boldsymbol{u}$ and $\boldsymbol{v}$ is defined by*

$$\cos\theta = \frac{\langle \boldsymbol{u}, \boldsymbol{v}\rangle}{\|\boldsymbol{u}\|\|\boldsymbol{v}\|}.$$

Note that the angle is well-defined due to the Cauchy-Schwartz inequality. When the angle is $\frac{\pi}{2}$ or equivalently when $\langle \boldsymbol{u}, \boldsymbol{v}\rangle = 0$, we say that the two vectors are **orthogonal**. The norm function satisfies the following properties:

1. $\|\boldsymbol{v}\| \geq 0$ for all $\boldsymbol{v} \in V$, and $\|\boldsymbol{v}\| = 0$ if and only if $\boldsymbol{v} = 0$.
2. For any scalar $c \in \mathbb{R}$ and vector $\boldsymbol{v} \in V$, $\|c\boldsymbol{v}\| = |c|\|\boldsymbol{v}\|$
3. (**Triangle inequality**) For any $\boldsymbol{u}, \boldsymbol{v} \in V$, $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$.
4. (**Pithagorean theorem**) For any $\boldsymbol{u}, \boldsymbol{v} \in V$, if $\boldsymbol{u} \perp \boldsymbol{v}$, then $\|\boldsymbol{u} + \boldsymbol{v}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$.

By triangle inequality, we can define the distance on V by $d(\boldsymbol{u}, \boldsymbol{v}) = \|\boldsymbol{u} - \boldsymbol{v}\|$. The distance function satisfies the following properties:

1. $d(\boldsymbol{u}, \boldsymbol{v}) \geq 0$ for all $\boldsymbol{u}, \boldsymbol{v} \in V$, and $d(\boldsymbol{u}, \boldsymbol{v}) = 0$ if and only if $\boldsymbol{u} = \boldsymbol{v}$.
2. $d(\boldsymbol{u}, \boldsymbol{v}) = d(\boldsymbol{v}, \boldsymbol{u})$ for all $\boldsymbol{u}, \boldsymbol{v} \in V$.
3. (**Triangle inequality**) For any $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in V$, $d(\boldsymbol{u}, \boldsymbol{w}) \leq d(\boldsymbol{u}, \boldsymbol{v}) + d(\boldsymbol{v}, \boldsymbol{w})$.

In toplogy, a distance function satisfying the above three properties is called a **metric**, and such a space V with metric d is called a **metric space**. The distance function can induce distance between subsets of V as follows:

Definition 5.2. *Let (V, d) be a metric space. For any two subsets A, B of V, the **distance** between A and B is defined as*

$$d(A, B) = \inf\{d(\boldsymbol{a}, \boldsymbol{b}) \mid \boldsymbol{a} \in A, \boldsymbol{b} \in B\}.$$

We will consider natural subsets in inner product spaces, for example, in space of functions with inner products defined by integrals, this is useful when we want to approximate a function by polynomials or trigonometric functions. Or in space of data points or matrices, the distance function helps us to construct clustering or compressing algorithms. The most natural form of subsets are subspaces.

Proposition 5.1. *Let $(V, \langle\cdot,\cdot\rangle)$ be a Euclidean space, and W be a subspace of V. For any $\boldsymbol{v} \in V$, there exists a unique vector $\boldsymbol{w}_0 \in W$ such that*

$$d(\boldsymbol{v}, W) = d(\boldsymbol{v}, \boldsymbol{w}_0).$$

Moreover, the vector $\boldsymbol{w}_0$ satisfies that $\boldsymbol{v} - \boldsymbol{w}_0 \in W^\perp$.

Proof. By the Gram-Schmidt process, we can find an orthonormal basis $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k\}$ of W and extend it to an orthonormal basis $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_k, \boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_n\}$ of V. Then any vector $\boldsymbol{v} \in V$ can be written as

$$\boldsymbol{v} = \sum_{i=1}^{n} \langle \boldsymbol{v}, \boldsymbol{u}_i \rangle \boldsymbol{u}_i.$$

Define

$$\boldsymbol{w}_0 = \sum_{i=1}^{k} \langle \boldsymbol{v}, \boldsymbol{u}_i \rangle \boldsymbol{u}_i \in W. \tag{2}$$

Then for any $\boldsymbol{w} \in W$, we have

$$\begin{aligned} d(\boldsymbol{v}, \boldsymbol{w})^2 &= \|\boldsymbol{v} - \boldsymbol{w}\|^2 \\ &= \|\boldsymbol{v} - \boldsymbol{w}_0 + \boldsymbol{w}_0 - \boldsymbol{w}\|^2 \\ &= \|\boldsymbol{v} - \boldsymbol{w}_0\|^2 + \|\boldsymbol{w}_0 - \boldsymbol{w}\|^2 \quad (\text{since } \boldsymbol{v} - \boldsymbol{w}_0 \perp \boldsymbol{w}_0 - \boldsymbol{w}) \\ &\geq \|\boldsymbol{v} - \boldsymbol{w}_0\|^2. \end{aligned}$$

Thus, $d(\boldsymbol{v}, W) = d(\boldsymbol{v}, \boldsymbol{w}_0)$. The equality holds if and only if $\boldsymbol{w} = \boldsymbol{w}_0$. □

In Problem 4.3 of Exercise 4, we have shown that for any subspace W of an inner product space V, $V = W \oplus W^\perp$. Thus, any vector $\boldsymbol{v} \in V$ can be uniquely written as $\boldsymbol{v} = \boldsymbol{w} + \boldsymbol{w}^\perp$ with $\boldsymbol{w} \in W$ and $\boldsymbol{w}^\perp \in W^\perp$. The formula (2) also gives an effective way to obtain such a decomposition via orthonormal basis of W.

Definition 5.3. *The map $\mathrm{Proj}_W : \boldsymbol{v} \to \boldsymbol{w}_0$ is called a projection map onto W and it is a linear transformation satisfying $\mathrm{Proj}_W \circ \mathrm{Proj}_W = \mathrm{Proj}_W$.*

Later a projection map will be an example of symmetric or self-adjoint operator.

Remark 5.1. *Notice that the distance function is invariant under translations, i.e., for any $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{a} \in V$, $d(\boldsymbol{u} + \boldsymbol{a}, \boldsymbol{v} + \boldsymbol{a}) = d(\boldsymbol{u}, \boldsymbol{v})$. Thus, Proposition 5.1 can also be used to discribe the distance between a point and an affine subspace (a translation of a subspace).*

6 Orthogonal Matrices

In QR decomposition, we have already seen orthogonal matrices. An invertible $n \times n$ real matrix A can be viewed as a basis of $\mathbb{R}^n$ by taking all the column vectors. The column vectors of an orthogonal matrix form an orthonormal basis of Euclidean space $(\mathbb{R}^n, \langle\cdot,\cdot\rangle_{\text{st}})$. Another interpretation of matrices is that they represent linear transformations

$$T_A \colon \mathbb{R}^n \to \mathbb{R}^n, \quad \boldsymbol{v} \mapsto A\boldsymbol{v}.$$

Under this interpretation, orthogonal matrices represent linear transformations that preserve the inner product, i.e.,

Proposition 6.1. *Any $A \in M_n(\mathbb{R})$ is orthogonal if and only if*

$$\langle T_A(\boldsymbol{u}), T_A(\boldsymbol{v})\rangle_{st} = \langle \boldsymbol{u}, \boldsymbol{v}\rangle_{st}$$

for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$.

Proof. For any $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, we have

$$\langle T_A(\boldsymbol{u}), T_A(\boldsymbol{v})\rangle_{\text{st}} = (A\boldsymbol{u})^T(A\boldsymbol{v}) = \boldsymbol{u}^T A^T A\boldsymbol{v}.$$

If A is orthogonal, then $A^TA = I$, so $\langle T_A(\boldsymbol{u}), T_A(\boldsymbol{v})\rangle_{\text{st}} = \boldsymbol{u}^T\boldsymbol{v} = \langle \boldsymbol{u}, \boldsymbol{v}\rangle_{\text{st}}$. Conversely, if $\langle T_A(\boldsymbol{u}), T_A(\boldsymbol{v})\rangle_{\text{st}} = \langle \boldsymbol{u}, \boldsymbol{v}\rangle_{\text{st}}$ for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$, then $\boldsymbol{u}^T A^T A\boldsymbol{v} = \boldsymbol{u}^T\boldsymbol{v}$ for all $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$. When $\boldsymbol{u} = e_i$ and $\boldsymbol{v} = e_j$, this implies that the ij-th entry of A^TA is δ_{ij}, so $A^TA = I$ and A is orthogonal. □

Such linear transformations are called **isometries** under the standard inner product by Definition 2.5. More generally,

Proposition 6.2. *For an n-dimensional Euclidean space $(V, \langle\cdot,\cdot\rangle)$, the isometries on V are exactly those linear transformations whose matrix representations under any orthonormal basis are orthogonal matrices.*

The proof is similar as Proposition 6.2.

All the isometries of $V, \langle\cdot,\cdot\rangle$ form a group under composition. So we have a subgroup of $\mathrm{GL}(n, \mathbb{R})$ consisting of orthogonal matrices.

Definition 6.1. *An **orthogonal group** of order n is defined as*

$$\mathrm{O}(n, \mathbb{R}) \coloneqq \{A \in M_n(\mathbb{R}) \mid A^TA = I\}$$

For Euclidean space $(V, \langle\cdot,\cdot\rangle)$, we also define the orthogonal group as the group of isometries on the space

$$\mathrm{O}(V) \coloneqq \{T \in \mathrm{GL}(V) \mid \langle T(\boldsymbol{u}), T(\boldsymbol{v})\rangle = \langle \boldsymbol{u}, \boldsymbol{v}\rangle \text{ for all } \boldsymbol{u}, \boldsymbol{v} \in V\}$$

The map determinant is a group homomorphism

$$\det \colon \mathrm{GL}(n, \mathbb{R}) \to \mathbb{R}^\times$$

For an orthogonal matrix A, its determinant is equal to ± 1, and both ± 1 can be achieved. Thus, we introduce the following normal subgroup of $\mathrm{O}(n, \mathbb{R})$ by the kernel of the determinant map.

Definition 6.2. *A **special orthogonal group** of order n is defined as*

$$\mathrm{SO}(n,\mathbb{R}) \coloneqq \{A \in \mathrm{O}(n,\mathbb{R}) \mid \det(A) = 1\}$$

The special orthogonal group represents all orientation-preserving isometries of Euclidean space.

Example 6.1. *First by direct computation of orthogonal basis in $\mathbb{R}^2$, we know that any orthogonal matrix in $\mathrm{O}(2,\mathbb{R})$ has the form*

$$\mathrm{O}(2,\mathbb{R}) = \{\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \mid \theta \in \mathbb{R}\} \bigsqcup \{\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \mid \theta \in \mathbb{R}\}$$

Computing the determinants, we see that

$$\mathrm{SO}(2,\mathbb{R}) = \{\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \mid \theta \in \mathbb{R}\}$$

represents all rotations by angle θ in $\mathbb{R}^2$.

The other coset represents all reflections in $\mathbb{R}^2$. One important observation is that the composition of two reflections is a rotation, i.e., for any two reflection matrices $P_1, P_2 \in \mathrm{O}(2,\mathbb{R}) \setminus \mathrm{SO}(2,\mathbb{R})$, we have $P_1P_2 \in \mathrm{SO}(2,\mathbb{R})$. If the two reflectoin axes form an angle $\frac{\theta}{2}$, then P_1P_2 is the rotation by angle θ.

The concept of reflections can be generalized to higher dimensions.

Definition 6.3 (reflections)**.** *A linear transformation $T\colon V \to V$ on a Euclidean space $(V, \langle\cdot,\cdot\rangle)$ is called a **reflection** if there exists a one-dimensional subspace $L \subset V$ such that*

(1) $T(\boldsymbol{v}) = \boldsymbol{v}$ for all $\boldsymbol{v} \in L$;

(2) $T(\boldsymbol{w}) = -\boldsymbol{w}$ for all $\boldsymbol{w} \in L^{\perp}$.

Conversely, given any nonzero vector $\boldsymbol{u} \in V$, we can define a reflection T with respect to the line spanned by $\boldsymbol{u}$ by

$$T(\boldsymbol{v}) = \boldsymbol{v} - 2\frac{\langle \boldsymbol{v}, \boldsymbol{u}\rangle}{\langle \boldsymbol{u}, \boldsymbol{u}\rangle}\boldsymbol{u}.$$

*This vector $\boldsymbol{u}$ is called a **normal vector** of the reflection. Such a reflection is also be denoted by $s_{\boldsymbol{u}}$.*

From the coset decomposition of $\mathrm{O}(n,\mathbb{R})$ by $\mathrm{SO}(n,\mathbb{R})$, we know that

$$\mathrm{O}(n,\mathbb{R}) = \mathrm{SO}(n,\mathbb{R}) \ \cup \ P \cdot \mathrm{SO}(n,\mathbb{R}),$$

where $P = \begin{pmatrix} -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$ is a reflection matrix with determinant -1.

6.1 Spectral theorem

The classification of similar classes of matrices is equivalently the orbit decomposition of the conjugation action of $\mathrm{GL}(n,\mathbb{R})$ on $M_n(\mathbb{R})$. The congruent classes of matrices are the orbit decomposition of the congruence action of $P \in \mathrm{GL}(n,\mathbb{R})$ on $A \in M_n(\mathbb{R})$ by P^TAP. If we restrict the action to $\mathrm{O}(n,\mathbb{R})$, we have the following definitions and theorem.

Definition 6.4.

(1) *For $A, B \in M_n(\mathbb{R})$, we say A, B are **orthogonally similar** if there exists $Q \in \mathrm{O}(n,\mathbb{R})$ such that $A = Q^TBQ$ $(= Q^{-1}BQ)$.*

(2) *If A is orthogonally similar to a diagonal matrix, we say A is **orthogonally diagonalizable** (over $\mathbb{R}$).*

We mainly deal with three special kinds of matrices:

1. symmetric matrices, i.e., $A^T = A$;
2. skew-symmetric matrices, i.e., $A^T = -A$;
3. orthogonal matrices, i.e., $A^TA = I$.

We study their orbits under orthogonal group actions. Among them, symmetric matrices have the best diagonalization property.

Theorem 6.1 (Spectral Theorem for Real Symmetric Matrices)**.** *Real symmetric matrices are orthogonally diagonalizable.*

Proof. Let $A \in M_n(\mathbb{R})$ with $A = A^T$. We proceed by induction on n. First, we show that A has a real eigenvalue. Let $\lambda \in \mathbb{C}$ be an eigenvalue of A, and let $A\boldsymbol{v} = \lambda\boldsymbol{v}$ for some $\boldsymbol{v} \in \mathbb{C}^n \setminus \{0\}$. Then

$$\overline{\boldsymbol{v}}^TA\boldsymbol{v} = \lambda\overline{\boldsymbol{v}}^T\boldsymbol{v},$$

where $\overline{\boldsymbol{v}}^T\boldsymbol{v} > 0$. Since

$$(\overline{\overline{\boldsymbol{v}}^TA\boldsymbol{v}})^T = \overline{\boldsymbol{v}}^TA\boldsymbol{v},$$

it follows that $\lambda \in \mathbb{R}$. Since $A - \lambda\mathrm{I}$ is a singular real matrix, the solution $\boldsymbol{v}$ to $(A - \lambda\mathrm{I})\boldsymbol{v} = 0$ can be chosen in $\mathbb{R}^n$.

Assuming $(\boldsymbol{v},\boldsymbol{v}) = 1$, extend $\boldsymbol{v} = \boldsymbol{v}_1$ to an orthonormal basis of $\mathbb{R}^n$, denoted by

$$Q_1 = (\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n)$$

Then

$$\begin{aligned} AQ_1 &= (\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n) \\ &= (\lambda\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_n) \\ &= (\boldsymbol{v}_1 \cdots \boldsymbol{v}_n)\begin{pmatrix}\lambda & * \\ 0 & *\end{pmatrix} \end{aligned}$$

Thus,

$$A = Q_1\begin{pmatrix}\lambda & * \\ * & *\end{pmatrix}Q_1^{-1} = Q_1\begin{pmatrix}\lambda & * \\ * & *\end{pmatrix}Q_1^T$$

Since A is symmetric, the above decomposition actually takes the form

$$A = Q_1\begin{pmatrix}\lambda & 0 \\ 0 & *\end{pmatrix}Q_1^{-1}.$$

The result follows by induction on n. □

In the proof, we can also obtain a useful proposition of invariant subspace for real linear transformations. This is essentially because irreducible polynomials over $\mathbb{R}$ have degree at most 2.

Proposition 6.3 (Invariant subspace of real linear transformations)**.** *Let $T\colon V \to V$ be a linear transformation on an $n \geq 1$-dimensional real vector space V. Then there exists a one-dimensional or two-dimensional T-invariant subspace of V.*

Proof. Let $A \in M_n(\mathbb{R})$. Then A has a complex eigenvalue $\lambda = a + b\sqrt{-1} \in \mathbb{C}$ with eigenvector $\boldsymbol{v} \in \mathbb{C}^n \setminus \{0\}$. Write $\boldsymbol{v} = \boldsymbol{u} + \sqrt{-1}\boldsymbol{w}$ with $\boldsymbol{u}, \boldsymbol{w} \in \mathbb{R}^n$. Then

$$A(\boldsymbol{u} + \sqrt{-1}\boldsymbol{w}) = (a + b\sqrt{-1})(\boldsymbol{u} + \sqrt{-1}\boldsymbol{w}).$$

Equating real and imaginary parts, we have

$$\boldsymbol{u} = a\boldsymbol{u} - b\boldsymbol{w}, \quad \boldsymbol{w} = b\boldsymbol{u} + a\boldsymbol{w}.$$

So the subspace spanned by $\boldsymbol{u}$ and $\boldsymbol{w}$ is a A-invariant subspace in $\mathbb{R}^n$. For general V, choose a basis of V and identify V with $\mathbb{R}^n$. The result follows. □

The notion of symmetric matrices can be generalized to self-adjoint operators on inner product spaces and we obtain a more intrinsic proof for the spectral theorem.

Definition 6.5. *Let V be a vector space equipped with a symmetric or skew-symmetric bilinear form g. A linear map $T\colon V \to V$ is called* ***self-adjoint*** *if*

$$g(T(\boldsymbol{u}), \boldsymbol{v}) = g(\boldsymbol{u}, T(\boldsymbol{v}))$$

holds for all $\boldsymbol{u}, \boldsymbol{v} \in V$.

Proposition 6.4. *Let V be a vector space with a symmetric or skew-symmetric bilinear form g. Suppose T is a self-adjoint linear map. If $W \subset V$ is an T-invariant subspace, then $W^\perp$ is also an T-invariant subspace.*

Proof. For any $\boldsymbol{u} \in W^\perp$ and $\boldsymbol{w} \in W$, we have

$$g(T(\boldsymbol{u}), \boldsymbol{w}) = g(\boldsymbol{u}, T(\boldsymbol{w})) \in g(\boldsymbol{u}, W) = 0.$$

Thus $T(\boldsymbol{u}) \in W^\perp$. □

Now assume (V, g) is an inner product space, and $T\colon V \to V$ is a self-adjoint linear map. Let $\mathcal{B} = \{e_1, \ldots, e_n\}$ be an orthonormal basis of V, and let A be the matrix representation of T with respect to $\mathcal{B}$, i.e.,

$$T(e_1, \ldots, e_n) = (e_1, \ldots, e_n)A$$

From $g(T(x), y) = g(x, T(y))$, we have:

$$\begin{pmatrix} T(e_1) \\ \vdots \\ T(e_n) \end{pmatrix} \cdot_g (e_1, \ldots, e_n) = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} \cdot_g (T(e_1), \cdots, T(e_n))$$

That is,

$$A^T \begin{pmatrix} T(e_1) \\ \vdots \\ T(e_n) \end{pmatrix} \cdot_g (e_1, \cdots, e_n) = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} \cdot_g (e_1, \cdots, e_n)A$$

Since $\mathcal{B}$ is an orthonormal basis,

$$\begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix} \cdot_g (e_1, \cdots, e_n) = G_{g,\mathcal{B}} = \mathrm{I}_n,$$

Thus $A = A^T$. So we have the following proposition.

Proposition 6.5 (Self-adjoint operators and symmetric matrices)**.** *Let (V, g) be an inner product space and $T\colon V \to V$ be a linear map. Then T is self-adjoint if and only if the matrix representation of T with respect to any orthonormal basis is symmetric.*

Theorem 6.2 (Spectral Theorem for Self-adjoint Operators)**.** *Let (V, g) be a finite-dimensional inner product space over $\mathbb{R}$. Then any self-adjoint operator $T\colon V \to V$ is orthogonally diagonalizable, i.e., there exists an orthonormal basis of V consisting of eigenvectors of T.*

Proof. Proposition 6.3 shows that T has a one-dimensional or two-dimensional invariant subspace $W_1 \subset V$. If W_1 is two dimensional, under orthonormal basis of W_1, the matrix representation of $T|_{W_1}$ is symmetric

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

The characteristic polynomial is

$$p(\lambda) = \lambda^2 - (a + c)\lambda + (ac - b^2).$$

The discriminant is

$$(a + c)^2 - 4(ac - b^2) = (a - c)^2 + 4b^2 \geq 0.$$

Thus $T|_{W_1}$ has a real eigenvalue, and hence W_1 has a one-dimensional invariant subspace. Therefore we can assume $\dim_{\mathbb{R}} W_1 = 1$. The orthogonal complement $W_1^{\perp}$ is also invariant by Proposition 6.4. Using induction on $\dim_{\mathbb{R}} V$, we can find an orthonormal basis of V consisting of eigenvectors of T. □

Next, we look at the conjugacy classes of orthogonal group.

Theorem 6.3 (Conjugacy classes of orthogonal matrices)**.** *Suppose $A \in \mathrm{O}(n, \mathbb{R})$. Then A is orthogonally similar to*

$$\mathit{diag}\{\begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix}, \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix}, \ldots, 1, \ldots, -1, \ldots\},$$

Or equivalently, for any orthogonal transformation $T\colon V \to V$ on an n-dimensional Euclidean space $(V, \langle\cdot,\cdot\rangle)$, there exists an orthonormal basis of V such that the matrix representation of T with respect to this basis is block diagonal with blocks of the form in $\mathrm{SO}(2, \mathbb{R})$, 1 or -1.

Proof. We use the same method as in the proof of Theorem 6.2. By Proposition 6.3, there exists a one-dimensional or two-dimensional A-invariant subspace $W_1 \subset \mathbb{R}^n$. If $\dim_{\mathbb{R}} W_1 = 1$, then W_1 is spanned by an eigenvector of A with eigenvalue 1 or -1 since A is orthogonal. If $\dim_{\mathbb{R}} W_1 = 2$, then under an orthonormal basis of W_1, the matrix representation of $A|_{W_1}$ is in $\mathrm{O}(2, \mathbb{R})$. When the matrix is in $\mathrm{SO}(2, \mathbb{R})$, it is of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

When the matrix is not in $\mathrm{SO}(2,\mathbb{R})$, it is of the form

$$\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix}.$$

The characteristic polynomial is

$$p(\lambda) = \lambda^2 - 1.$$

So the eigenvalues are 1 and -1. Thus, in any case, W_1 has a one-dimensional invariant subspace. The orthogonal complement $W_1^{\perp}$ is also invariant by Proposition 6.4. Using induction on n, we can find an orthonormal basis of $\mathbb{R}^n$ such that the matrix representation of A with respect to this basis is block diagonal with blocks of the form in $\mathrm{SO}(2,\mathbb{R})$, 1 or -1. □

Example 6.2. *For $A \in \mathrm{SO}(3,\mathbb{R})$, it is orthogonally similar to*

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Let

$$Q^T A Q = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and $Q = (v_1, v_2, v_3)$, then

$$A(v_1, v_2, v_3) = (v_1, v_2, v_3) \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

So the action of A on $\mathbb{R}^3$ is a rotation in the plane $\mathrm{span}_{\mathbb{R}}\{v_1, v_2\}$.

Thus we have the following corollary in geometry.

Corollary 6.1. *The composition of two rotations of three-dimensional Euclidean space along two intersecting lines is a rotation.*

If we also generalize the concept of rotation to higher dimensions as orthogonal transformations whose matrix representations under some orthonormal basis has one block in $\mathrm{SO}(2,\mathbb{R})$ and all others are 1 blocks, then we have the following corollary.

Corollary 6.2. *The composition of two reflections of an n-dimensional Euclidean space along two hyperplanes is a rotation.*

7 Singular Value Decomposition and Low Rank Approximation

In previous sections, we gave the structure theorem for self-adjoint operators on a given linear space, which is also known as the spectral theorem for self-adjoint operators. If $T\colon V \to W$ is a linear map between two different spaces, how do we find the canonical form of T?

The theory of equivalence canonical forms for matrices tells us that, given a matrix $A \in M_{m\times n}(\mathbb{R})$, there exist $P \in \mathrm{GL}(n,\mathbb{R})$, $Q \in \mathrm{GL}(m,\mathbb{R})$ such that:

$$Q^{-1} A P = \begin{pmatrix} \mathrm{I}_r & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{pmatrix}$$

where $r = \operatorname{rank} A$. Now we attach an inner product structure to the linear space, so we hope that P, Q can be chosen as orthogonal matrices.

Theorem 7.1 (Singular Value Decomposition)**.** *Let $A \in M_{m\times n}(\mathbb{R})$. There exist $P \in \mathrm{O}(n)$, $Q \in \mathrm{O}(m)$ such that:*

$$A = QDP^T,$$

where

$$D = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}_{m\times n}, \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0.$$

*The σ_i are called the **singular values** of A.*

Remark 7.1. *Suppose $A = QDP^T$ is the singular value decomposition of A. Let $Q = (w_1, \ldots, w_m), P = (v_1, \ldots, v_n)$, then*

$$A = \sum_{i=1}^{r} \sigma_i w_i v_i^T,$$

This is a commonly used form of the singular value decomposition.

Proof. We first prove the uniqueness of the singular values: If $A = QDP^T$, then

$$A^T A = P\mathrm{diag}\{\sigma_1^2, \ldots, \sigma_n^2\}P^T,$$

so $\sigma_1^2, \ldots, \sigma_n^2$ are the eigenvalues of $A^T A$, which are uniquely determined by A after sorting.

Next, we prove the existence of the singular value decomposition: Since $A^T A$ is symmetric, there exists $P \in \mathrm{O}(n)$ such that

$$P^{-1}(A^T A)P = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix},$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. Let $P = (v_1, \ldots, v_n)$, then $A^T A v_i = \lambda_i v_i$.

We show that $Av_1, \ldots, Av_n$ are mutually orthogonal:

$$\begin{aligned} \langle Av_i, Av_j\rangle_{\mathbb{R}^m} &= (Av_i)^T(Av_j) \\ &= v_i^T(A^T A v_j) \\ &= \lambda_j v_i^T v_j \\ &= \lambda_j \langle v_i, v_j\rangle_{\mathbb{R}^n}. \end{aligned}$$

Thus, it is 0 when $i \neq j$, and λ_i when $i = j$.

Assume $\lambda_1 \geq \cdots \geq \lambda_r > 0$, $\lambda_{r+1} = \cdots = \lambda_n = 0$. Let

$$w_i = \frac{Av_i}{||Av_i||_{\mathbb{R}^m}} = \frac{Av_i}{\sqrt{\lambda_i}}, \quad i = 1, \ldots, r$$

Then $w_1, \ldots, w_r$ are orthonormal. If we denote $\sigma_i := \sqrt{\lambda_i}$, then

$$(Av_1, \ldots, Av_n) = (w_1, \ldots, w_n) \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}_{m \times n}$$

That is,

$$AP = QD, \quad P \in \mathrm{O}(n),\ Q \in \mathrm{O}(m).$$

□

7.1 Low-Rank Approximation and Application: Image Compression

Given a matrix $A = (a_{ij})_{m \times n}$, where a_{ij} represents the grayscale of the pixel at (i, j), the amount of data needed to record this is mn, which is very large when m, n are large. Using SVD, we can write it as

$$A = \sum_{i=1}^{\min(m,n)} \sigma_i w_i v_i^T$$

Take $k \ll \min(m, n)$, then we have the approximation

$$A_k = \sum_{i=1}^{k} \sigma_i w_i v_i^T,$$

The storage data amount is now $k(m + n + 1)$.

Intuitively, A_k should be "very close" to A. Next, we describe this strictly. Define the **Frobenius inner product** on the matrix space $M_{m \times n}(\mathbb{R})$:

$$\langle A, B \rangle_F = \mathrm{tr}(A^T B).$$

Then the distance between matrices A, B is

$$\sqrt{\langle A - B, A - B \rangle}_F = \sqrt{\sum_{i,j} (a_{ij} - b_{ij})^2}.$$

Under Frobenius inner product, we have the following important theorem about the **low-rank approximation** using SVD.

Theorem 7.2 (Eckart-Young, Schmidt)**.** *A_k is the closest matrix to A among matrices of rank $\leq k$, i.e.,*

$$||A - A_k||_F = \min_{\mathrm{rank}\, B \leq k} ||A - B||_F.$$

Corollary 7.1. *Let $A = QDP^T$ and $D = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ be the singular value decomposition of matrix A, then*

$$\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2},$$

and

$$||A - A_k||_F = \sqrt{\sigma_{k+1}^2 + \cdots + \sigma_n^2}.$$

Before proving Theorem 7.2, we first present some important properties used in the proof.

Lemma 7.1. *Given matrices $A, B \in M_{m\times n}(\mathbb{R})$, for any $P \in \mathrm{O}(n), Q \in \mathrm{O}(m)$, we have*

$$\langle QAP^T, QBP^T\rangle_F = \langle A, B\rangle_F.$$

Lemma 7.2. *Given matrix $A \in M_{m\times n}(\mathbb{R})$, then*

$$\sigma_1(A) = \max_{0\neq v\in\mathbb{R}^n} \frac{|Av|_{\mathbb{R}^m}}{|v|_{\mathbb{R}^n}}.$$

Proof. Consider the symmetric matrix $M = A^TA$. According to the Min-Max principle for eigenvalues of symmetric matrices, we have

$$\lambda_1(A^TA) = \max_{0\neq v\in\mathbb{R}^n} \frac{\langle v, A^TAv\rangle_{\mathbb{R}^n}}{\langle v, v\rangle_{\mathbb{R}^n}} = \max_{0\neq v\in\mathbb{R}^n} \frac{\langle Av, Av\rangle_{\mathbb{R}^m}}{\langle v, v\rangle_{\mathbb{R}^n}},$$

Thus,

$$\sigma_1(A) = \sqrt{\lambda_1(A^TA)} = \max_{0\neq v\in\mathbb{R}^n} \frac{|Av|_{\mathbb{R}^m}}{|v|_{\mathbb{R}^n}}.$$

□

Lemma 7.3.

(1) *For any matrix $A \in M_{m\times n}(\mathbb{R})$, we have*

$$\sigma_\ell(A - A_k) = \sigma_{k+\ell}(A)$$

(2) *For any matrix $A, B \in M_{m\times n}(\mathbb{R})$, where $\operatorname{rank} B \le k$, then*

$$\sigma_\ell(A - B) \ge \sigma_{k+\ell}(A)$$

Proof. (1): Assume the singular value decomposition of A is denoted as

$$A = \sum_{i=1}^{\min\{m,n\}} \sigma_i w_i v_i^T,$$

then

$$A - A_k = \sum_{i=k+1}^{\min\{m,n\}} \sigma_i w_i v_i^T,$$

so $\sigma_\ell(A - A_k) = \sigma_{k+\ell}(A)$.

(2): Assume the singular value decomposition of A is denoted as

$$A = \sum_{i=1}^{\min\{m,n\}} \sigma_i w_i v_i^T.$$

We first prove the case $\ell = 1$: Let $W = \operatorname{span}_{\mathbb{R}}\{v_1, \ldots, v_{k+1}\}$. Since $\operatorname{rank} B \le k$, we have $\dim\ker B \ge n - k$, thus $\ker B \cap W \neq \varnothing$. Take $0 \neq v \in \ker B \cap W$, assume without loss of generality that $v = \sum_{i=1}^{k+1} a_i v_i$, then by Lemma 7.2,

$$\sigma_1(A - B) \ge \frac{|(A - B)v|_{\mathbb{R}^m}}{|v|_{\mathbb{R}^n}} = \frac{|Av|_{\mathbb{R}^m}}{|v|_{\mathbb{R}^n}} = \frac{\sqrt{\sum_{i=1}^{k+1} \sigma_i^2 a_i^2}}{\sqrt{\sum_{i=1}^{k+1} a_i^2}} \ge \sigma_{k+1}(A).$$

Now we prove the general case: By (1), we have

$$\begin{aligned}\sigma_\ell(A-B) &= \sigma_1\left((A-B)-(A-B)_{\ell-1}\right)\\ &= \sigma_1\left(A-(B+(A-B)_{\ell-1})\right)\end{aligned}$$

Since $\operatorname{rank} B \le k, \operatorname{rank}(A-B)_{\ell-1} \le \ell-1$, we have

$$\operatorname{rank}(B+(A-B)_{\ell-1}) \le k+\ell-1,$$

By the previous case,

$$\sigma_1\left(A-(B+(A-B)_{\ell-1})\right) \ge \sigma_{k+\ell}(A).$$

□

Now we give the proof of Theorem 7.2:

Proof.

$$\begin{aligned}\|A-B\|_F^2 &= \sum_{i=1}^{n}\sigma_i^2(A-B)\\ &\ge \sum_{i=1}^{n}\sigma_{i+k}^2(A)\\ &= \sum_{i=1}^{n}\sigma^2(A-A_k)^2\\ &= \|A-A_k\|_F.\end{aligned}$$

□

7.2 Application: Low-dimensional Fitting (PCA)

Problem 7.1. *Given m experiments, where each experiment yields an n-dimensional datum, assume $m \gg n$, then we obtain the following matrix:*

$$A = \begin{pmatrix}\alpha_1^T\\ \alpha_2^T\\ \vdots\\ \alpha_m^T\end{pmatrix}.$$

How can we determine if $\alpha_1, \dots, \alpha_m$ lie near some lower-dimensional linear subspace?

Proposition 7.1. *Given the singular value decomposition of a matrix $A \in M_{m\times n}(\mathbb{R})$ with $m > n$:*

$$A = \sum_{i=1}^{n}\sigma_i w_i v_i^T.$$

Let $W_k = \operatorname{span}_{\mathbb{R}}\{v_1, \dots, v_k\}$. Then W_k minimizes the following value:

$$\sum_{i=1}^{m}(\operatorname{dist}(\alpha_i, W))^2,$$

where W is a k-dimensional subspace of $\mathbb{R}^n$.

Proof. For any k-dimensional subspace W, take an orthonormal basis $\{u_1, \ldots, u_k\}$ of W. Then

$$\begin{aligned}(\mathrm{dist}(\alpha_i, W))^2 &= \|\alpha_i - \mathrm{Proj}_W \alpha_i\|^2 \\ &= \|\alpha_i - \sum_{j=1}^{k} \langle \alpha_i, u_j \rangle u_j\|^2 \\ &= \|\alpha_i^T - ((\alpha_i, u_1), \ldots, (\alpha_i, u_k)) \begin{pmatrix} u_1^T \\ \vdots \\ u_k^T \end{pmatrix} \|_2^2.\end{aligned}$$

Thus,

$$\sum_{i=1}^{m} (\mathrm{dist}(\alpha_i, W))^2 = \|A - \begin{pmatrix} \langle \alpha_1, u_1 \rangle & \ldots & \langle \alpha_1, u_k \rangle \\ \vdots & & \vdots \\ \langle \alpha_m, u_1 \rangle & \ldots & \langle \alpha_m, u_k \rangle \end{pmatrix} \begin{pmatrix} u_1^T \\ \vdots \\ u_k^T \end{pmatrix} \|_F^2.$$

On the other hand, since $\alpha = \sum_{i=1}^{n} \langle \alpha, v_i \rangle v_i$, we can write

$$\alpha^T = (\langle \alpha, v_1 \rangle, \ldots, \langle \alpha, v_n \rangle) P,$$

where $P = (v_1, \ldots, v_n)$. Thus

$$A = QDP^T,$$

where the i-th row of QD is $(\langle \alpha_i, v_1 \rangle, \ldots, \langle \alpha_i, v_n \rangle)$. Therefore

$$QD \begin{pmatrix} v_1^T \\ \vdots \\ v_k^T \\ 0 \\ \vdots \\ 0 \end{pmatrix} = A_k.$$

According to Theorem 7.2, W_k achieves the minimum of $\sum_{i=1}^{m} (\mathrm{dist}(\alpha_i, W))^2$. □

Proposition 7.2. *Let $\mu = \frac{1}{m} \sum_{i=1}^{m} (\alpha_i^T)$, and $\overline{\mu} = (1, \ldots, 1)^T \cdot \mu$. Perform singular value decomposition on $B = A - \overline{\mu}$ to obtain $B = QDP^T$, where $P = (v_1, \ldots, v_n)$. Then*

$$\mu + \mathrm{span}_{\mathbb{R}}(v_1, \ldots, v_k)$$

is the k-dimensional affine plane that minimizes the sum of squared distances to $\alpha_1, \ldots, \alpha_m$.

7.3 Application: Least Squares Method

To predict y from n-dimensional data $(a_1, \ldots, a_n)$,

$$y = x_1 a_1 + \cdots + x_n a_n.$$

After multiple experiments, we obtain a system of equations

$$\begin{pmatrix} \alpha_1^T \\ \alpha_2^T \\ \vdots \\ \alpha_m^T \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}.$$

When $m \gg n$, this system of equations most likely has no solution. How do we find the closest solution? That is, minimizing the sum of squared errors between the model predicted values and the m experiment results:

$$\sum_{i=1}^{m} (\alpha_i^T \cdot x - b)^2$$

Let W be the column space of A. Then the distance between b and W is

$$|b - \mathrm{Proj}_W b|.$$

Since Ax ranges over W, there exists x such that $Ax = \mathrm{Proj}_W b$. However, such x is not unique; they differ by elements in $\ker A$. Usually, we require minimizing the length of x to give a unique solution, which is called the optimal least squares solution.

Next, we introduce how to use singular value decomposition to provide the least squares solution. Given a matrix $A \in M_{m\times n}(\mathbb{R})$ and its singular value decomposition

$$A = QDP^T,$$

where

$$D = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}$$

Let

$$A^\dagger = PD^\dagger Q^T,$$

where

$$D^\dagger \begin{pmatrix} \sigma_1^{-1} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \sigma_2^{-1} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r^{-1} & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix}$$

Then we have the following theorem:

Theorem 7.3. *For $Ax = b$, $\widehat{x} = A^\dagger b$ is the* ***optimal least squares solution.***

Proof. First we prove that $A\widehat{x} - b \perp \mathrm{im}\, A$. That is, we verify

$$A^T(A(A^\dagger b) - b) = 0.$$

By direct calculation,

$$\begin{aligned} PD^TQ^T(QDP^TPD^\dagger Q^T - \mathrm{I}_m)b &= (PD^TDD^\dagger Q^T - PD^TQ^T)b \\ &= (PD^TQ^T - PD^TQ^T)b \\ &= 0. \end{aligned}$$

Next, we verify that $\ker A \perp A^\dagger b$. Since $\ker A = \ker(QDP^T) = \ker D$, and $A^\dagger b = PD^\dagger Q^T b$, we need to verify

$$D^\dagger Q^T b \perp \ker D,$$

which is evident from the expressions of D and $D^\dagger$. □

Remark 7.2. *If one wants to use*

$$y = x_1 a_1 + \cdots + x_n a_n + c$$

to fit the data, consider

$$\begin{pmatrix} \alpha_1^T, 1 \\ \alpha_2^T, 1 \\ \vdots \\ \alpha_m^T, 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ c \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Let

$$\widetilde{A} = \begin{pmatrix} \alpha_1^T, 1 \\ \alpha_2^T, 1 \\ \vdots \\ \alpha_m^T, 1 \end{pmatrix},$$

and consider $\widetilde{A}^\dagger$.

8 Excercises

8.1 Useful Exercises

Excercise 8.1. *Prove that any skew-symmetric matrix $A \in M_n(\mathbb{R})$ can be orthogonally similar to a block diagonal matrix with blocks of the form*

$$\begin{pmatrix} 0 & -\lambda \\ \lambda & 0 \end{pmatrix}$$

and possibly a 0 block if n is odd. Use this to show that any skew-symmetric matrix over $\mathbb{R}$ is congruent to a block diagonal matrix with blocks of the form

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

and possibly some 0 block.

Excercise 8.2. *Let V be the linear space consisting of all skew-symmetric real matrices of order n.*

1. *For any $A \in V$, prove that $I + A$ is invertible.*
2. *For any $A \in V$, define $f(A) = (I - A)(I + A)^{-1}$. Prove that $f(A)$ is an orthogonal matrix.*
3. *Give a characterization of the image of $f \colon V \to \mathrm{O}(n)$ in terms of eigenvalues, that is, which matrices can be written in the form $(I - A)(I + A)^{-1}$ for some $A \in V$.*

Excercise 8.3. *Let A be 2×2 real symmetric matrix*

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}.$$

Write down an orthogonal matrix Q which diagonalizes A in terms of a, b, c.

Excercise 8.4. *Consider the groups* $\mathrm{O}(2)$*, its subgroup* $\mathrm{SO}(2)$ *and group* $\mathrm{SO}(3)$*. Determine whether the following statements are correct. If correct, prove it; if incorrect, provide a counterexample:*

1. *Two elements in the group* $\mathrm{O}(2)$ *are conjugate if and only if they have the same trace.*
2. *Two elements in the group* $\mathrm{SO}(2)$ *are conjugate in the group* $\mathrm{SO}(2)$ *if and only if they have the same trace.*
3. *Two elements in the group* $\mathrm{SO}(2)$ *are conjugate in the group* $\mathrm{O}(2)$ *if and only if they have the same trace.*
4. *Two elements in the group* $\mathrm{SO}(3)$ *are conjugate if and only if they have the same trace.*

Excercise 8.5 (Cartan–Dieudonné theorem)**.** *Prove that any orthogonal transformation of Euclidean space $(V, \langle \cdot, \cdot \rangle)$ can be expressed as a composition of at most* $\dim V$ *reflections.*

(The nontrivial part of the original theorem is to show this also holds for any non-degenerate symmetric bilinear form over a field of characteristic not equal to 2.)

Excercise 8.6 (Courant–Fischer–Weyl Min-Max Principle)**.** *You may choose to prove either part (1) or part (2).*

1. *Let $(E, \langle \cdot, \cdot \rangle)$ be an n-dimensional real inner product space. Suppose T is a self-adjoint transformation on E with real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Prove that the eigenvalues of T can be characterized by the following min-max method:*

   $$\lambda_k = \min \{ \max \{ \langle T(x), x \rangle : x \perp W_k, |x| = 1 \} : W_k \subset E \text{ is a subspace}, \dim W_k = k - 1 \}$$

   Here, for a fixed $(k-1)$-dimensional subspace W_k, we first compute the maximum value

   $$\max \{ \langle T(x), x \rangle : x \perp W_k, |x| = 1 \}.$$

   Then we vary W_k over all $(k-1)$-dimensional subspaces and take the minimum of these maximum values.

2. *Alternatively, you may prove the following special case: Let A be an $n \times n$ real symmetric matrix and v be an arbitrary n-dimensional real column vector, where $|v|$ denotes the vector length under the standard inner product. Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be all eigenvalues of A. Prove that:*

   $$|Av| \leq \max\{|\lambda_1|, |\lambda_2|, \cdots, |\lambda_n|\}|v|.$$

8.2 Optional problems

You do not need to hand in these problems, but you are encouraged to discuss and try them.

Excercise 8.7 (Outer automorphisms of $\mathrm{SO}(n, \mathbb{R})$)**.** *An automorphism of a group G is called an* inner automorphism *if it is of the form $g \mapsto hgh^{-1}$ for some fixed $h \in G$. An automorphism which is not inner is called an* outer automorphism*. Consider the automorphism of* $\mathrm{SO}(n, \mathbb{R})$ *defined by $A \mapsto PAP^{-1}$ where $P \in \mathrm{O}(n, \mathbb{R})$ with* $\det P = -1$*. Is this an inner automorphism or an outer automorphism? Prove your answer. (The answer may depend on n.)*

Excercise 8.8 (Challenge Problem)**.** *You will obtain a standard form for Lorentz transformations on $\mathbb{R}^4$. Let e_i $(i = 1, \dots 4)$ be the standard basis for $\mathbb{R}^4$. Consider the symmetric bilinear on $\mathbb{R}^4$ defined by*

$$\langle x, y\rangle = x_1y_1 + x_2y_2 + x_3y_3 - x_4y_4.$$

A basis f_i $(i = 1, \dots 4)$ of $\mathbb{R}^4$ is called orthonormal if

$$\langle f_1, f_1\rangle = \langle f_2, f_2\rangle = \langle f_3, f_3\rangle = 1, \quad \langle f_4, f_4\rangle = -1, \quad \langle f_i, f_j\rangle = 0 \text{ if } i \neq j.$$

Suppose T is a Lorentz transformation on $\mathbb{R}^4$, that is, T is a linear transformation such that

$$\langle Tx, Ty\rangle = \langle x, y\rangle$$

for all $x, y \in \mathbb{R}^4$. Prove that there exists an orthonormal basis of $\mathbb{R}^4$ such that the matrix of T is block diagonal with blocks of the following types:

1. *A block of order 1 with entry ± 1.*
2. *A block of order 2 of the form*
$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$
3. *a block of order 2 of the form*
$$\pm\begin{pmatrix} \cosh\theta & \sinh\theta \\ \sinh\theta & \cosh\theta \end{pmatrix} \quad \text{or} \quad \pm\begin{pmatrix} \cosh\theta & \sinh\theta \\ -\sinh\theta & -\cosh\theta \end{pmatrix}.$$
4. *A block A of order 3 with eigenvalue $\lambda = \pm 1$ so that $(A - \lambda I)^3 = 0$ but $(A - \lambda I)^2 \neq 0$.*

Excercise 8.9. *If the Lorentz transformation T in Problem 8.8 is replaced by a transformation satisfying*

$$\langle Tx, y\rangle = -\langle x, Ty\rangle$$

can you obtain a similar result? State the result and prove it.

Excercise 8.10 (Cauchy Interlacing Theorem)**.** *Let A be an $n \times n$ real symmetric matrix, and let B be an $m \times m$ principal submatrix of A, where $m < n$. If the eigenvalues of A are $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$, and the eigenvalues of B are $\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_m$, then for all $1 \leqslant i \leqslant m$, we have*

$$\lambda_i \geqslant \mu_i \geqslant \lambda_{i+n-m}.$$

(Hint: Use the Courant-Fischer-Weyl min-max principle from Problem 8.6.)

Excercise 8.11 (Sylvester's Criterion)**.** *Use the Cauchy interlacing theorem to prove Sylvester's criterion: A symmetric matrix is positive definite if and only if all its leading principal minors are positive.*

9 Hermitian Forms and Unitary Matrices

Definition 9.1. *Let V be a finite-dimensional vector space over $\mathbb{C}$. A map $h\colon V \times V \to \mathbb{C}$ is called a* Hermitian form *, if it satisfies:*

(1) $h(\lambda_1\boldsymbol{x}_1 + \lambda_2\boldsymbol{x}_2, \boldsymbol{y}) = \overline{\lambda_1}h(\boldsymbol{x}_1, \boldsymbol{y}) + \overline{\lambda_2}h(\boldsymbol{x}_2, \boldsymbol{y})$;

(2) $h(\boldsymbol{x},\boldsymbol{y}) = \overline{h(\boldsymbol{y},\boldsymbol{x})}$.

Remark 9.1. *Condition (2) guarantees that $h(\boldsymbol{x},\boldsymbol{x})$ is real for all $\boldsymbol{x}$.*

Example 9.1. *The map*

$$h\colon \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$$
$$(\boldsymbol{x},\boldsymbol{y}) \mapsto \sum_{i=1}^{n} \overline{x}_i y_i$$

is called the standard Hermitian form on $\mathbb{C}^n$.

Problem 9.1. *Similar to symmetric bilinear forms and quadratic forms on real vector spaces, how can we recover h from $h(x,x)$?*

Lemma 9.1. *For a Hermitian form $h\colon V \times V \to \mathbb{C}$, we have*

$$\mathrm{Re}h(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{2}(h(\boldsymbol{x}+\boldsymbol{y},\boldsymbol{x}+\boldsymbol{y}) - h(\boldsymbol{x},\boldsymbol{x}) - h(\boldsymbol{y},\boldsymbol{y}))$$
$$\mathrm{Im}h(\boldsymbol{x},\boldsymbol{y}) = \frac{1}{2}(h(\sqrt{-1}\boldsymbol{x}+\boldsymbol{y},\sqrt{-1}\boldsymbol{x}+\boldsymbol{y}) - h(\boldsymbol{x},\boldsymbol{x}) - h(\boldsymbol{y},\boldsymbol{y})).$$

Moreover, Reh *is symmetric, and* Imh *is skew-symmetric (alternating).*

Given a Hermitian form $h\colon V\times V \to \mathbb{C}$, and a basis $\alpha_1,\cdots,\alpha_n$, the Gram matrix $(h(\alpha_i,\alpha_j)) = H$ satisfies $H = \overline{H}^T$. Such a matrix H is called a *Hermitian matrix*. For convenience, we denote $H^* = \overline{H}^T$ hereafter. For a Hermitian matrix H, it is also true that ReH is symmetric and ImH is skew-symmetric.

Definition 9.2. *For $A, B \in M_n(\mathbb{C})$, if there exists $P \in \mathrm{GL}_n(\mathbb{C})$ such that $PAP^* = B$, then A and B are said to be* congruent *(or Hermitian congruent).*

Now given a Hermitian form $h\colon V \times V \to \mathbb{C}$, and two bases $(\alpha_1,\cdots,\alpha_n), (\beta_1,\cdots,\beta_n)$ satisfying $\{\beta_1,\cdots,\beta_n\} = (\alpha_1,\cdots,\alpha_n)P$, then

$$\begin{aligned}(h(\beta_i,\beta_j))^* &= \begin{pmatrix}\beta_1\\ \vdots\\ \beta_n\end{pmatrix}\cdot\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix}\\ &= \overline{P}^T\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_n\end{pmatrix}\cdot(\alpha_1,\cdots,\alpha_n)P\\ &= \overline{P}^T\left(h(\alpha_i,\alpha_j)\right)P\end{aligned}$$

So the Gram matrices of the same Hermitian form under different bases are congruent.

Theorem 9.1. *Any Hermitian matrix is congruent to a diagonal matrix, and the diagonal entries must be real numbers.*

Definition 9.3. *A Hermitian form h on a vector space V over $\mathbb{C}$ is called:*

- Positive definite *($h > 0$), if $h(v,v) > 0$ for all $v \neq 0$;*
- Positive semi-definite *($h \geq 0$), if $h(v,v) \geq 0$ for all v;*

- Negative definite *($h < 0$), if $h(v, v) < 0$ for all $v \neq 0$;*
- Negative semi-definite *($h \leq 0$), if $h(v, v) \leq 0$ for all v.*

Theorem 9.2. *For a Hermitian matrix H, the following are equivalent:*

(1) *H is positive definite;*

(2) *H is congruent to the identity matrix;*

(3) *$H = PP^*$, where $P \in \mathrm{GL}_n(\mathbb{C})$;*

(4) *All leading principal minors of H are positive;*

(5) *All principal minors of H are positive.*

Theorem 9.3. *For a Hermitian matrix H, the following are equivalent:*

(1) *H is positive semi-definite;*

(2) *H is congruent to*
$$\begin{pmatrix} \mathrm{I}_r & \mathrm{O} \\ \mathrm{O} & \mathrm{O} \end{pmatrix};$$

(3) *$H = PP^*$, where $P \in M_n(\mathbb{C})$;*

(4) *All principal minors of H are nonnegative.*

Definition 9.4. *Given a finite-dimensional $\mathbb{C}$-vector space V and a Hermitian form $h \colon V \times V \to \mathbb{C}$, if $h > 0$, then (V, h) is called a Hermitian inner product space, or a* unitary space *.*

For convenience, we denote $\langle x, y\rangle := h(x, y)$.

(1) Length (Norm): For $x \in V$, $|x| := \sqrt{\langle x, x\rangle}$.

(2) Angle: For $x, y \in V \backslash \{0\}$, there exists $\theta(x, y) \in [0, \frac{\pi}{2}]$ such that
$$\cos(\theta(x, y)) = \frac{|\langle x, y\rangle|}{|x||y|}.$$

Lemma 9.2 (Cauchy-Schwarz Inequality)**.**
$$|\langle x, y\rangle| \leq |\langle x, x\rangle||\langle y, y\rangle|.$$

Proof. Let $\langle x, y\rangle = re^{\sqrt{-1}\theta}$. Consider $|te^{\sqrt{-1}\theta}x + y|^2$, where $t \in \mathbb{R}$. Then
$$t^2|x|^2 + 2rt + |y|^2 \geq 0, \quad \forall t \in \mathbb{R},$$
Thus the discriminant is ≤ 0, and equality holds if and only if x, y are linearly dependent over $\mathbb{C}$. □

Lemma 9.3 (Triangle Inequality)**.** $\forall x, y \in V$, $|x + y| \leq |x| + |y|$.

Proof. By definition,

$$|x+y|^2 = |x|^2 + \langle x, y\rangle + \langle y, x\rangle + |y|^2.$$

By Cauchy-Schwarz inequality,

$$|\langle x, y\rangle| \le |x||y|, \quad |\langle y, x\rangle| \le |x||y|,$$

Thus

$$|x+y|^2 \le |x|^2 + 2|x||y| + |y|^2 = (|x| + |y|)^2.$$

□

Lemma 9.4 (Parallelogram Identity)**.** *$\forall x, y \in V$, we have*

$$|x+y|^2 + |x-y|^2 = 2(|x|^2 + |y|^2).$$

Definition 9.5. *Given a unitary space (V, h), a basis $\{\alpha_1, \ldots, \alpha_n\}$ is called an **orthonormal basis** if $\langle \alpha_i, \alpha_j\rangle = \delta_{ij}$.*

Remark 9.2. *Similar to the inner product space case, starting from any basis, one can obtain an orthonormal basis via the Gram-Schmidt process.*

Theorem 9.4. *Given a unitary space (V, h) and two orthonormal bases $\{\alpha_1, \ldots, \alpha_n\}$ and $\{\beta_1, \ldots, \beta_n\}$. Assume $(\alpha_1, \ldots, \alpha_n) = (\beta_1, \ldots, \beta_n)P$, then*

$$PP^* = \mathrm{I}_n\,.$$

Definition 9.6. *A matrix $P \in M_n(\mathbb{C})$ is called a* unitary matrix *if $PP^* = \mathrm{I}_n$.*

Lemma 9.5. *If $P \in M_n(\mathbb{C})$ is a unitary matrix, then $\overline{P}, P^{-1}, P^*$ are all unitary matrices.*

Definition 9.7. *Two complex matrices $A, B \in M_n(\mathbb{C})$ are called* unitarily similar *if there exists a unitary matrix $U \in M_n(\mathbb{C})$ such that $A = U^*BU$.*

Proposition 9.1. *Any complex square matrix is unitarily similar to an upper triangular matrix.*

Proof. By induction. Let $A \in M_n(\mathbb{C})$, induct on n.

Let λ be an eigenvalue of A, $v \in \mathbb{C}^n$ be a λ-eigenvector. Assume without loss of generality $\|v\| = 1$. Extend v to an orthonormal basis of $\mathbb{C}^n$, say $v_1, \cdots, v_n$. Let $U = (v_1 \cdots v_n)$ be a unitary matrix. Then

$$AU = (Av_1 \cdots Av_n) = (v_1 \cdots v_n)\begin{pmatrix} \lambda & * \\ 0 & * \end{pmatrix}$$

That is $U^*AU = \begin{pmatrix} \lambda & * \\ 0 & * \end{pmatrix}$.

The result follows by induction. □

Definition 9.8. *A square matrix $N \in M_n(\mathbb{C})$ is called **normal** if $NN^* = N^*N$.*

Lemma 9.6. *If a normal matrix N is unitarily similar to $\begin{pmatrix} N_1 & N_2 \\ 0 & N_3 \end{pmatrix}$, where N_1, N_3 are square matrices, then it must be that $N_2 = 0$, and both N_1 and N_3 are normal.*

Proof. Suppose $\begin{pmatrix} N_1 & N_2 \\ 0 & N_3 \end{pmatrix} = U^*NU$ for a unitary matrix U. Then U^*NU commutes with $(U^*NU)^* = U^*N^*U$, implying U^*NU is normal. Thus

$$\begin{pmatrix} N_1 & N_2 \\ 0 & N_3 \end{pmatrix} \begin{pmatrix} N_1^* & 0 \\ N_2^* & N_3^* \end{pmatrix} = \begin{pmatrix} N_1^* & 0 \\ N_2^* & N_3^* \end{pmatrix} \begin{pmatrix} N_1 & N_2 \\ 0 & N_3 \end{pmatrix}$$

Comparing top-left blocks, $N_1N_1^* + N_2N_2^* = N_1^*N_1 \Rightarrow \operatorname{tr}(N_2N_2^*) = 0 \Rightarrow N_2 = 0$. □

Theorem 9.5 (Spectral Theorem for Normal Matrices)**.** *A normal matrix is unitarily similar to a diagonal matrix.*

Proof. Let $A \in M_n(\mathbb{C})$ be normal. Then A is unitarily similar to an upper triangular matrix

$$\begin{pmatrix} * & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & * \end{pmatrix}$$

By Lemma 9.6, it must be a diagonal matrix. □

Example 9.2. *Real symmetric, real skew-symmetric, orthogonal, Hermitian, unitary, and skew-Hermitian matrices are all normal matrices.*

Corollary 9.1.

1. *Any Hermitian matrix (including real symmetric matrices) is unitarily similar to* $\operatorname{diag}(\lambda_1, \cdots, \lambda_n)$*, where* λ_i *are real numbers.*
2. *Any unitary matrix (including orthogonal matrices) is unitarily similar to* $\operatorname{diag}(\lambda_1, \cdots, \lambda_n)$*, where* $\lambda_i = e^{\sqrt{-1}\theta_i}$*,* $\theta_i \in \mathbb{R}$*.*
3. *Any skew-Hermitian matrix (including skew-symmetric matrices) is unitarily similar to* $\operatorname{diag}(\lambda_1, \cdots, \lambda_n)$*, where* $\lambda_i = \sqrt{-1}g_i$*,* $g_i \in \mathbb{R}$*.*

Proof.

1. Suppose a Hermitian matrix is unitarily similar to $\operatorname{diag}(\lambda_1, \cdots, \lambda_n)$. Since the diagonal matrix is also Hermitian $\Rightarrow$ λ_i are real.
2. Suppose a unitary matrix is unitarily similar to a diagonal matrix $\operatorname{diag}(\lambda_1, \cdots, \lambda_n)$. Similarity preserves unitarity $\Rightarrow$ the modulus of each λ_i is 1.
3. Similar argument.

□

Definition 9.9. *Let* $(V, \langle \cdot, \cdot \rangle)$ *be an* n*-dimensional unitary space (i.e., the Gram matrix of* $\langle \cdot, \cdot \rangle$ *is a positive definite Hermitian matrix). A linear transformation* $\varphi \colon V \to V$ *is called a* ***normal transformation****,* ***Hermitian transformation****,* ***unitary transformation****, or* ***skew-Hermitian transformation****, if the matrix representation of* φ *under an orthonormal basis of* V *is a matrix of the corresponding type.*

Restating the Spectral Theorem from the viewpoint of linear transformations:

Lemma 9.7. *Let φ be a normal transformation on a unitary space V. If W is an invariant subspace of φ, then the orthogonal complement $W^\perp$ is also an invariant subspace.*

Theorem 9.6 (Spectral Theorem, Transformation Form)**.** *Let V be an n-dimensional unitary space, and $\varphi\colon V \to V$ be a normal transformation. Then there exists an orthonormal basis $v_1, \ldots, v_n$ of V such that each v_i is an eigenvector of φ. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues of φ, and let W_i be the eigenspace of φ corresponding to λ_i. Then*

$$\varphi = \lambda_1\pi_1 + \cdots + \lambda_k\pi_k,$$

*where $\pi_i\colon V \to W_i$ is the orthogonal projection. This is called the **spectral decomposition** of φ.*

Theorem 9.7. *Given an n-dimensional unitary space V and a linear transformation $\varphi\colon V \to V$, there exists a unique linear transformation $\varphi^*\colon V \to V$ such that for any $\alpha, \beta \in V$,*

$$\langle \varphi^*(\alpha), \beta\rangle = \langle \alpha, \varphi(\beta)\rangle.$$

*We call φ^* the **adjoint** of φ. Let A and B be the matrices of φ and φ^* under an orthonormal basis of V, respectively, then $B = A^* = \overline{A}^T$.*

Proof. Assume $\langle \varphi^*(\alpha), \beta\rangle = \langle \alpha, \varphi(\beta)\rangle$, $\forall \alpha, \beta$. Take an orthonormal basis $v_1, \ldots, v_n$. Then:

$$\begin{pmatrix} \varphi^*(v_1) \\ \vdots \\ \varphi^*(v_n) \end{pmatrix} = B \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \quad (\varphi(v_1), \cdots, \varphi(v_n)) = (v_1, \cdots, v_n)A$$

The matrix of inner products on the left is $B^T \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \cdot (v_1, \cdots, v_n) = B^T$, and the matrix of inner products on the right is $\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \cdot (v_1, \cdots, v_n)\overline{A} = \overline{A}$. Thus $B^T = \overline{A} \Rightarrow B = A^*$.

Conversely, to construct φ^*, one only needs to pick an orthonormal basis. Let A be the matrix of φ, then we use A^* to obtain φ^*. □

Theorem 9.8. *Let V_1, V_2 be n-dimensional unitary spaces, and $\gamma : V_1 \to V_2$ be a linear map. The following are equivalent:*

1. *γ preserves the inner product;*
2. *γ preserves the length of vectors (isometry);*
3. *$\gamma : V_1 \to V_2$ is a linear isomorphism and preserves the inner product;*
4. *γ maps any orthonormal basis of V_1 to an orthonormal basis of V_2;*
5. *γ maps some orthonormal basis to an orthonormal basis;*
6. *The matrix representation of γ with respect to orthonormal bases of V_1 and V_2 is a unitary matrix.*

Example 9.3. *When $n = 1$, a unitary matrix is just a complex number with modulus 1, i.e., $e^{\sqrt{-1}\theta}$ for some $\theta \in \mathbb{R}$. Thus, a unitary transformation on a 1-dimensional unitary space is just a rotation by angle θ. Or in other words, the group of unitary transformations on a 1-dimensional unitary space is isomorphic to the real special orthogonal group* $\mathrm{SO}(2)$.

When $n = 2$, we will discuss in detail the group $\mathrm{SU}(2) = \{A \in M_2(\mathbb{C}) : AA^* = \mathrm{I}_2, \det A = 1\}$ in the next lecture.

10 Application of spectral theorem: conics and quadrics

Definition 10.1. *A* conic *in $\mathbb{R}^2$ is the locus of points (x, y) satisfying a quadratic equation*

$$Ax_1^2 + Bx_1x_2 + Cx_2^2 + Dx_1 + Ex_2 + F = 0,$$

where $A, B, C, D, E, F \in \mathbb{R}$ and not all of A, B, C are zero.

The conic is called *nondegenerate* if it is not two lines, one line, a point, or empty set. If we allow rigid motion transformations of $\mathbb{R}^2$, $\boldsymbol{x} \mapsto P\boldsymbol{x} + \boldsymbol{b}$ where $P \in \mathrm{O}(2, \mathbb{R})$ and $\boldsymbol{b} \in \mathbb{R}^2$, then we can simplify the equation of a conic. Using the results on symmetric bilinear forms and quadratic forms, we can show the classification theorem for conics:

Theorem 10.1. *Any conic with nondegenerate quadratic part can be transformed by rigid motion into one of the following standard forms:*

1. *Ellipse: $\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 1$, where $a, b > 0$;*
2. *Hyperbola: $\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 1$, where $a, b > 0$;*
3. *Parabola: $x_1^2 = 2px_2$, where $p \neq 0$.*

Proof. The quadratic part $Ax_1^2 + Bx_1x_2 + Cx_2^2$ defines a symmetric bilinear form on $\mathbb{R}^2$ by Gram matrix

$$\begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix}.$$

Then there exists an orthogonal matrix $P \in \mathrm{O}(2, \mathbb{R})$ such that

$$P^T \begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix} P = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix},$$

where λ_1, λ_2 are the eigenvalues of the above matrix. Thus, under the change of variables $\boldsymbol{x} = P\boldsymbol{y}$, the conic equation becomes

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 + D'y_1 + E'y_2 + F = 0,$$

for some $D', E' \in \mathbb{R}$. Now we consider the following cases:

1. If λ_1, λ_2 have the same sign, then we can complete the square to obtain the standard form of an ellipse.
2. If λ_1, λ_2 have opposite signs, then we can complete the square to obtain the standard form of a hyperbola.

3. If one of λ_1, λ_2 is zero, then we can complete the square to obtain the standard form of a parabola.

□

Definition 10.2. *A* quadric *in* $\mathbb{R}^n$ *is the locus of points* $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ *satisfying a quadratic equation*

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j + \sum_{i=1}^{n} b_i x_i + c = 0,$$

where $a_{ij}, b_i, c \in \mathbb{R}$ *and not all of* a_{ij} *are zero.*

We call a quadric *nondegenerate* if it is not a union of hyperplanes, or a cone over lower dimensional quadric, or empty set. Similar to the conic case, using the results on symmetric bilinear forms and quadratic forms, we can show the classification theorem for three-dimensional quadrics:

Theorem 10.2. *Any quadric in* $\mathbb{R}^3$ *with nondegenerate quadratic part can be transformed by rigid motion into one of the following standard forms:*

1. *Ellipsoid:* $\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = 1$, *where* $a, b, c > 0$;
2. *Hyperboloid of one sheet:* $\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} - \frac{x_3^2}{c^2} = 1$, *where* $a, b, c > 0$;
3. *Hyperboloid of two sheets:* $-\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = 1$, *where* $a, b, c > 0$;
4. *Elliptic paraboloid:* $\frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} = 2px_3$, *where* $a, b, p > 0$;
5. *Hyperbolic paraboloid:* $\frac{x_1^2}{a^2} - \frac{x_2^2}{b^2} = 2px_3$, *where* $a, b, p > 0$.

11 Skew-symmetric Bilinear Forms and Symplectic Matrices

The classification of skew-symmetric bilinear forms on finite-dimensional vector spaces is similar to that of symmetric bilinear forms. We use one unified approach to summarize the idea of the proofs here.

Definition 11.1 (Nondegenerate Forms)**.** *If* $B\colon V \times V \to F$ *is a bilinear (or Hermitian) form on a finite-dimensional vector space* V *over a field* F*, then* B *is called* nondegenerate *if for any nonzero* $v \in V$*, there exists some* $u \in V$ *such that* $B(v, u) \neq 0$ *(or equivalently,* $B(u, v) \neq 0$*). Or equivalently, the Gram matrix of* B *under any basis of* V *is invertible.*

We will always assume B is symmetric or skew-symmetric bilinear form on V over a field F with $\mathrm{char}(F) \neq 2$, or Hermitian form on V over $\mathbb{C}$. Then the next definition of orthogonal complement has no ambiguity. Let W be a subspace of V. The *orthogonal complement* of W with respect to B is defined as

$$W^{\perp} = \{v \in V : B(v, w) = 0, \forall w \in W\}.$$

The definition of radical $V^{\perp}$ is also similar.

Definition 11.2 (Radical)**.** *The* radical *of B is defined as*

$$\mathrm{Rad}(B) = \{v \in V : B(v,u) = 0, \forall u \in V\} = V^{\perp}.$$

Then have an orthogonal decomposition theorem.

Proposition 11.1. *The vector space is decomposed into the orthogonal direct sum*

$$V = \mathrm{Rad}(B) \oplus W,$$

where W is a subspace of V such that the restriction of B on W is nondegenerate.

The next step is to classify nondegenerate forms and we have the following important theorem generalizing the theorem in Excercise 4.3.

Theorem 11.1. *Let B be a nondegenerate symmetric, or skew-symmetric (or Hermitian) form on a finite-dimensional vector space V over a field F (or over $\mathbb{C}$). Fix a subspace W of V. Then the following are equivalent:*

(1) *The restriction of B on W is nondegenerate;*

(2) *$V = W \oplus W^{\perp}$;*

(3) *The restriction of B on $W^{\perp}$ is nondegenerate.*

The proof is similar to that of Excercise 4.3.

Then we can classify skew-symmetric bilinear forms as follows.

Theorem 11.2. *Let B be a skew-symmetric bilinear form on a finite-dimensional vector space V over a field F whose characteristic is not 2. Then there exists a basis of V such that the Gram matrix of B under this basis is*

$$\begin{pmatrix} \mathrm{O}_r & \mathrm{I}_r & \mathrm{O} \\ -\mathrm{I}_r & \mathrm{O}_r & \mathrm{O} \\ \mathrm{O} & \mathrm{O} & \mathrm{O} \end{pmatrix},$$

where $2r = \mathrm{rank}(B)$. Or equivalently, there exists a basis $\{\alpha_1, \cdots, \alpha_r, \beta_1, \cdots, \beta_r, \gamma_1, \cdots, \gamma_{n-2r}\}$ of V such that

$$B(\alpha_i, \beta_j) = \delta_{ij}, \quad B(\beta_i, \alpha_j) = -\delta_{ij}, \quad B(\alpha_i, \alpha_j) = B(\beta_i, \beta_j) = B(\gamma_i, \cdot) = 0.$$

Proof. We reduce to nondegenerate case by the orthogonal decomposition by the radical. Then we prove by induction on the dimension of V. The base case $\dim V = 0$ is trivial. For $\dim V > 0$, we can find $u, v \in V$ such that $B(u,v) \neq 0$. Rescale u such that $B(u,v) = 1$. Let $W = \mathrm{span}\{u, v\}$. Then the restriction of B on W is nondegenerate. By the previous theorem, we have the orthogonal decomposition $V = W \oplus W^{\perp}$. By induction hypothesis, we can find a basis of $W^{\perp}$ such that the Gram matrix of the restriction of B on $W^{\perp}$ has the desired form. Combining with $\{u, v\}$ gives the desired basis of V. □

Corollary 11.1. *The rank of any skew-symmetric matrix is even.*

Corollary 11.2. *Any two nondegenerate skew-symmetric bilinear forms on a $2n$-dimensional vector space over a field F are equivalent.*

Definition 11.3 (Symplectic basis)**.** *A basis $\{\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_n\}$ of a $2n$-dimensional vector space V over a field F with a nondegenerate skew-symmetric bilinear form B is called a* symplectic basis *if*

$$B(\alpha_i, \beta_j) = \delta_{ij}, \quad B(\beta_i, \alpha_j) = -\delta_{ij}, \quad B(\alpha_i, \alpha_j) = B(\beta_i, \beta_j) = 0.$$

Definition 11.4 (Symplectic Matrix)**.** *A matrix $A \in M_{2n}(F)$ is called a* symplectic matrix *if it satisfies*

$$A^T J A = J,$$

where

$$J = \begin{pmatrix} \mathrm{O}_n & \mathrm{I}_n \\ -\mathrm{I}_n & \mathrm{O}_n \end{pmatrix}.$$

Proposition 11.2. *The set of all symplectic matrices in $M_{2n}(F)$ forms a group under matrix multiplication, denoted by* $\mathrm{Sp}(2n, F)$*.*

One nontrivial fact about symplectic matrices is that their determinants are always 1.

Theorem 11.3. *If $A \in \mathrm{Sp}(2n, F)$, then* $\det(A) = 1$*.*

Proof. Taking determinants on both sides of the equation $A^T J A = J$, we have

$$(\det A)^2 \det J = \det J.$$

Since $\det J = 1$, we have $(\det A)^2 = 1$. Thus $\det A = \pm 1$. To show $\det A = 1$, we consider the symplectic form on $\wedge^{2n} F^{2n}$ induced by J. Then we have

$$(\det A)\omega = \omega,$$

where ω is a nonzero element in $\wedge^{2n} F^{2n}$. Thus $\det A = 1$. □

Another nontrivial factor about symplectic matrices is that they are closed under taking transpose.

Theorem 11.4. *If $A \in \mathrm{Sp}(2n, F)$, then $A^T \in \mathrm{Sp}(2n, F)$.*

Proof. Taking transpose on both sides of the equation $A^T J A = J$, we have

$$A^T J^T A = J^T.$$

Since $J^T = -J$, we have

$$A^T J A = -J.$$

Multiplying both sides by -1, we have

$$A^T J A = J.$$

Thus $A^T \in \mathrm{Sp}(2n, F)$. □

12 Excercises

12.1 Useful Exercises

Excercise 12.1. *Suppose A is an invertible real square matrix with singular values $\sigma_1, \cdots, \sigma_n$. Find the singular values of A^{-1}.*

Excercise 12.2 (Polar Decomposition)**.** *In this problem, you will prove the polar decomposition of an invertible complex matrix using the singular value decomposition.*

1. *Try to state without proof the singular value decomposition theorem for complex matrices similar to the one stated in the class for real matrices.*

2. *Use it to prove the following: Let A be an $n \times n$ invertible complex matrix. Prove that there exist unitary matrices $U \in \mathrm{U}(n)$ and P is a positive definite hermitian matrix such that $A = UP$. This is called the polar decomposition of A.*

3. *Show that the polar decomposition is unique. (This shows that $\mathrm{GL}(n, \mathbb{C})$ is homeomorphic to $\mathrm{U}(n) \times \mathcal{H}_n^+$ where $\mathcal{H}_n^+$ is the set of all positive definite hermitian matrices, a convex cone in the vector space of all hermitian matrices.)*

Excercise 12.3. *Artin Chapter 8, Excercise 6.20*
Prove the circulant, the matrix below, is normal.

$$C = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_n \\ c_n & c_0 & c_1 & \cdots & c_{n-1} \\ c_{n-1} & c_n & c_0 & \cdots & c_{n-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_1 & c_2 & c_3 & \cdots & c_0 \end{bmatrix}$$

(How to diagonalize it? Hint: write C as a polynomial in the shift matrix. see Artin Chapter 8, Excercise 6.19)

Excercise 12.4. *Prove that for any square submatrix of a unitary matrix, the modulus of any of its complex eigenvalues does not exceed* 1.

Excercise 12.5. 1. *If A and B are normal matrices, is AB necessarily normal? What if we additionally assume $AB = BA$?*

2. *Determine if the matrix $A = \begin{pmatrix} \sqrt{-1} & -\sqrt{-1} \\ -\sqrt{-1} & \sqrt{-1} \end{pmatrix}$ is normal, Hermitian, or unitary.*

Excercise 12.6 (Characterization of Normal Matrices in terms of Singular Values)**.** *Let A be an $n \times n$ complex matrix. Let $\sigma_i(A)$ denote the i-th singular value of A (such that $\sigma_1(A) \geq \cdots \geq \sigma_n(A) \geq 0$), and let $\lambda_1, \cdots, \lambda_n$ be the complex eigenvalues of A (counted with algebraic multiplicity). Prove that $\sum_{i=1}^{n} \sigma_i(A)^2 \geq \sum_{i=1}^{n} |\lambda_i|^2$, with equality if and only if A is a normal matrix.*

(Hint: Recall that any complex matrix can be conjugate to upper triangular matrix by an invertible matrix, by choosing eigenvectors and induction. Prove that actually a unitary matrix can do this job, then compare the Frobenius norms of both sides.)

Excercise 12.7. *Lang Algebra Chapter XV, 1.*
Here we choose σ to be complex conjugation.

1. *Let E be a finite dimensional space over the complex numbers, and let*

$$h : E \times E \to \mathbb{C}$$

be a hermitian form. Write

$$h(x, y) = g(x, y) + if(x, y)$$

where g, f are real valued. Show that g, f are $\mathbb{R}$-bilinear, g is symmetric, f is alternating.

2. *Let E be finite dimensional over $\mathbf{C}$. Let $g : E \times E \to \mathbf{C}$ be $\mathbb{R}$-bilinear. Assume that for all $x \in E$, the map $y \mapsto g(x, y)$ is $\mathbb{C}$-linear, and that the $\mathbb{R}$-bilinear form*

$$f(x, y) = g(x, y) - g(y, x)$$

is real-valued on $E \times E$. Show that there exists a hermitian form h on E and a symmetric $\mathbb{C}$-bilinear form ψ on E such that $2ig = h + \psi$. Show that h and ψ are uniquely determined.

Excercise 12.8. *Let $X = A + \sqrt{-1}B$ be a complex square matrix, where $A, B \in M_n(\mathbb{R})$. Prove that X is a unitary matrix if and only if*

$$\begin{pmatrix} A & -B \\ B & A \end{pmatrix}$$

is an orthogonal matrix.

12.2 Optional problems

Excercise 12.9. *Prove the additive inequality of singular values. Let A, B be two $m \times n$ real matrices. Prove that*

$$\sigma_{k+l-1}(A + B) \leq \sigma_k(A) + \sigma_l(B)$$

for $k + l - 1 \leq \min(m, n)$.

Excercise 12.10. *Let (V, ω) be a symplectic space if ω is a non-degenerate skew-symmetric form on the F-vector space V. An F-linear transformation T is called a symplectic transformation if $\omega(T(v), T(w)) = \omega(v, w)$. A basis $(\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_n)$ is called a symplectic basis if $\omega(\alpha_i, \beta_j) = \delta_{ij}$ and $\omega(\alpha_i, \alpha_j) = \omega(\beta_i, \beta_j) = 0$. When $F = \mathbb{C}$, prove that for any symplectic transformation T of a symplectic space, there exists a symplectic basis such that the matrix of T under this basis has the form*

$$\begin{bmatrix} B_n & 0_n \\ 0_n & B_n^T \end{bmatrix}$$

where B_n is an n-dimensional Jordan normal form.

Excercise 12.11. *Prove the following inequality between singular values and eigenvalues. Let A be an $n \times n$ complex matrix with eigenvalues $\lambda_1, \cdots, \lambda_n$ (counted with algebraic multiplicity) and singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$. Then for any $k = 1, \cdots, n$,*

$$|\lambda_1 \lambda_2 \cdots \lambda_k| \leq \sigma_1 \sigma_2 \cdots \sigma_k.$$

Excercise 12.12 (Challenge Problem)**.** *Let A be a real square matrix of order n, and let the eigenvalues of $A^T A$ be $\lambda_1^2, \cdots, \lambda_n^2$, where $0 \leq \lambda_i \leq 1$ for $i = 1, 2, \cdots, n$. Prove that:*

$$\det(I_n - A) \geq (1 - \lambda_1)(1 - \lambda_2) \cdots (1 - \lambda_n)$$

13 Orthogonal representation of $SU(2)$

We will use the spectral theorem for unitary matrices to study the special unitary group of degree 2 over $\mathbb{C}$:

$$\mathrm{SU}(2) = \{A \in M_2(\mathbb{C}) : AA^* = \mathrm{I}_2, \det A = 1\}.$$

Then we can study the topology of SU(2) via its explicit form.

Proposition 13.1. *Any matrix $A \in \mathrm{SU}(2)$ can be written as*

$$A = \begin{pmatrix} \alpha & -\overline{\beta} \\ \beta & \overline{\alpha} \end{pmatrix},$$

where $\alpha, \beta \in \mathbb{C}$ satisfy $|\alpha|^2 + |\beta|^2 = 1$.

Proof. First it is straightforward to verify that any matrix of the above form is in SU(2). Conversely, let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SU}(2)$. Then the conditions $AA^* = \mathrm{I}_2$ and $\det A = 1$ imply that $A^{-1} = A^*$. Thus

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} \overline{a} & \overline{c} \\ \overline{b} & \overline{d} \end{pmatrix}.$$

Comparing the entries gives us the desired form.

□

So the topology of SU(2) is the same as that of the unit sphere S^3 in $\mathbb{R}^4$.

The conjugacy classes of SU(2) can be described as follows:

Proposition 13.2. *The conjugacy classes of* SU(2) *are determined by the trace. More precisely, two matrices $A, B \in \mathrm{SU}(2)$ are conjugate if and only if $\mathrm{tr}(A) = \mathrm{tr}(B)$.*

Proof. First the eigenvalues of any matrix $A \in \mathrm{SU}(2)$ are of the form $e^{\sqrt{-1}\theta}, e^{-\sqrt{-1}\theta}$ for some $\theta \in \mathbb{R}$. Thus the trace of A is $2\cos\theta$. If two matrices $A, B \in \mathrm{SU}(2)$ have the same trace, then they have the same eigenvalues. By the spectral theorem for unitary matrices, they are conjugate by a matrix $P \in U(2)$. And we can alway rescale P such that $\det P = 1$, so that $P \in \mathrm{SU}(2)$. If two matrices $A, B \in \mathrm{SU}(2)$ are conjugate, then they have the same eigenvalues, so they have the same trace. □

The trace zero matrices in SU(2) are of the form

$$\begin{pmatrix} x\sqrt{-1} & y + \sqrt{-1}z \\ -y + \sqrt{-1}z & -x\sqrt{-1} \end{pmatrix}, \text{ and } x^2 + y^2 + z^2 = 1.$$

Next we will show that there is a natural homomorphism from SU(2) to the special orthogonal group $\mathrm{SO}(3, \mathbb{R})$. Consider the Lie algebra of SU(2): Let

$$\mathfrak{su}(2) = \{X \in M_2(\mathbb{C}) : X^* = -X, \mathrm{tr}(X) = 0\}.$$

Then $\mathfrak{su}(2)$ is a 3-dimensional real vector space with basis

$$\sqrt{-1}\sigma_1 = \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}, \quad \sqrt{-1}\sigma_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \sqrt{-1}\sigma_3 = \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix},$$

where $\sigma_1, \sigma_2, \sigma_3$ are the Pauli matrices. Define an inner product on $\mathfrak{su}(2)$ by

$$\langle X, Y \rangle = -\frac{1}{2}\mathrm{tr}(XY).$$

Then $\{i\sigma_1, i\sigma_2, i\sigma_3\}$ is an orthonormal basis of $\mathfrak{su}(2)$ with respect to this inner product.

Proposition 13.3. *For any $A \in \mathrm{SU}(2)$, the map $\varphi_A : \mathfrak{su}(2) \to \mathfrak{su}(2)$ defined by*

$$\varphi_A(X) = AXA^{-1}$$

is an orthogonal transformation on the inner product space $\mathfrak{su}(2)$.

Proof. First we need to verify that $\varphi_A(X) \in \mathfrak{su}(2)$ for any $X \in \mathfrak{su}(2)$. Indeed,

$$(AXA^{-1})^* = (A^{-1})^* X^* A^* = -AXA^{-1},$$

and

$$\operatorname{tr}(AXA^{-1}) = \operatorname{tr}(X) = 0.$$

Next we verify that φ_A preserves the inner product. Indeed,

$$\begin{aligned}\langle \varphi_A(X), \varphi_A(Y) \rangle &= -\frac{1}{2}\operatorname{tr}(AXA^{-1}AYA^{-1}) \\ &= -\frac{1}{2}\operatorname{tr}(XY) \\ &= \langle X, Y \rangle.\end{aligned}$$

□

The main theorem we want to show is the following:

Theorem 13.1. *The map* $\Phi : \mathrm{SU}(2) \to \mathrm{SO}(3, \mathbb{R})$ *induced by*

$$\Phi(A) \text{“} =\text{”} \varphi_A$$

is a surjective group homomorphism with kernel $\{\pm \mathrm{I}_2\}$.

It is easier to use quaternions to prove this theorem. We introduce the the following matrices:

$$\mathbf{1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{i} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \mathbf{j} = \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}, \quad \mathbf{k} = \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix}.$$

Then any matrix in $\mathfrak{su}(2)$ can be written as

$$X = x\mathbf{i} + y\mathbf{j} + z\mathbf{k},$$

for some $x, y, z \in \mathbb{R}$. Thus we can identify $\mathfrak{su}(2)$ with $\mathbb{R}^3$ via the map

$$(x, y, z) \mapsto x\mathbf{i} + y\mathbf{j} + z\mathbf{k}.$$

Under this identification, the inner product on $\mathfrak{su}(2)$ corresponds to the standard inner product on $\mathbb{R}^3$. The multiplication rules of $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -\mathbf{1}, \quad \mathbf{ij} = \mathbf{k}, \quad \mathbf{jk} = \mathbf{i}, \quad \mathbf{ki} = \mathbf{j},$$

The group SU(2) is isomorphic to the group of unit quaternions

$$\{a\mathbf{1} + b\mathbf{i} + c\mathbf{j} + d\mathbf{k} : a, b, c, d \in \mathbb{R}, a^2 + b^2 + c^2 + d^2 = 1\}.$$

Proof of Theorem 13.1. First we check that the image of Φ is in $\mathrm{SO}(3, \mathbb{R})$. Since φ_A is an orthogonal transformation on $\mathfrak{su}(2)$, we only need to check that $\det(\varphi_A) = 1$. Note that SU(2) is connected and the determinant is a continuous map to $\{\pm 1\}$, so the determinant must be constantly 1.

In order to prove surjectivity, we first verify the group operation of SU(2) on the unit sphere of $\mathfrak{su}(2)$ is transitive. This is because of the conjugacy class description of SU(2): any two trace zero matrices in SU(2) are conjugate.

Next we show that for any fixed axis, we have a surjective homomorphism from the stabilizer subgroup to $\mathrm{SO}(2,\mathbb{R})$. Without loss of generality, we consider the axis $\mathbf{i}$. Then we can consider the subgroup

$$H = \{A_\theta \in \mathrm{SU}(2) : A_\theta = \cos(\theta)\mathbf{1} + \sin(\theta)\mathbf{i}\}.$$

Then $A_\theta^{-1} = \cos(\theta)\mathbf{1} - \sin(\theta)\mathbf{i}$. So φ_{A_θ} fixes $\mathbf{i}$:

$$\begin{aligned}\varphi_{A_\theta}(\mathbf{i}) &= A_\theta \mathbf{i} A_\theta^{-1} \\ &= (\cos(\theta)\mathbf{1} + \sin(\theta)\mathbf{i})\mathbf{i}(\cos(\theta)\mathbf{1} - \sin(\theta)\mathbf{i}) \\ &= \mathbf{i}.\end{aligned}$$

Next we compute the action of φ_{A_θ} on the orthogonal complement of $\mathbf{i}$:

$$\begin{aligned}\varphi_{A_\theta}(\mathbf{j}) &= A_\theta \mathbf{j} A_\theta^{-1} \\ &= (\cos(\theta)\mathbf{1} + \sin(\theta)\mathbf{i})\mathbf{j}(\cos(\theta)\mathbf{1} - \sin(\theta)\mathbf{i}) \\ &= \cos(2\theta)\mathbf{j} + \sin(2\theta)\mathbf{k},\end{aligned}$$

and

$$\begin{aligned}\varphi_{A_\theta}(\mathbf{k}) &= A_\theta \mathbf{k} A_\theta^{-1} \\ &= (\cos(\theta)\mathbf{1} + \sin(\theta)\mathbf{i})\mathbf{k}(\cos(\theta)\mathbf{1} - \sin(\theta)\mathbf{i}) \\ &= -\sin(2\theta)\mathbf{j} + \cos(2\theta)\mathbf{k}.\end{aligned}$$

Thus the map φ_{A_θ} restricted on the orthogonal complement of $\mathbf{i}$ is a rotation by angle 2θ. So we have a surjective homomorphism from H to $\mathrm{SO}(2,\mathbb{R})$.

Next we use a standard trick in group operations. Let $\mathbf{v}$ be any unit vector in $\mathfrak{su}(2)$. Then we can find some $P \in \mathrm{SU}(2)$ such that $\varphi_P(\mathbf{i}) = \mathbf{v}$. Then the group PHP^{-1} is in the stabilizer subgroup of $\mathbf{v}$. And we have a surjective homomorphism from PHP^{-1} to $\mathrm{SO}(2,\mathbb{R})$ by sending $PHP^{-1} \ni A \mapsto \varphi_P \circ \varphi_A \circ \varphi_{P^{-1}} \in \mathrm{SO}(2,\mathbb{R})$.

Finally any element in $\mathrm{SO}(3,\mathbb{R})$ fixed some $\mathbf{v}$ in the unit sphere of $\mathfrak{su}(2)$, we conclude that $\Phi : \mathrm{SU}(2) \to \mathrm{SO}(3,\mathbb{R})$ is surjective.

The kernel of Φ is $\{\pm \mathrm{I}_2\}$ since these are the only two matrices act trivially on $\mathfrak{su}(2)$. Or you can check it from the explicit centers from the quaternions.

□

One important consequence of Theorem 13.1 is that $\mathrm{SU}(2)$ is simply connected, and the special orthogonal group $\mathrm{SO}(3,\mathbb{R})$ has fundamental group $\mathbb{Z}/2\mathbb{Z}$. This can be reflected by the famous experiment of Dirac's belt, or the plate trick.

14 Examples of Lie groups and Lie algebras

Lie groups and Lie algebras are important tools to study continuous symmetries. For example, the orthogonal group $\mathrm{O}(3,\mathbb{R})$ is a Lie group which describes the continuous symmetries of the 3-dimensional space preserving the standard inner product. Continuous symmetry of a physic system usually corresponds to conservation laws. This is the famous Noether's theorem. If we consider all the symmetries of the 3-dimensional space preserving the distance, then we get a bigger Lie group containing both $\mathrm{O}(3,\mathbb{R})$ and translation group $\mathbb{R}^3$ as subgroups. The total dimension of the symmetries is 6 and they correpond to three angular momenta and three linear momenta in physics. If we combine the symmetry in time axis, then we also have another

dimension corresponding to energy. The definition and general properties involves differential geometry, and we will only give some examples here instead of proving the full theory for general cases.

A *Lie group* is a group G which is also a smooth manifold such that the group operations (multiplication and inversion) are smooth maps. A *Lie algebra* is a vector space $\mathfrak{g}$ over a field $\mathbb{R}$ equipped with a bilinear map $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ called the Lie bracket satisfying the following properties:

(1) $[x, x] = 0$ for any $x \in \mathfrak{g}$ (anticommutativity);

(2) $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0$ for any $x, y, z \in \mathfrak{g}$ (Jacobi identity).

The definition of Lie algebra is motivated by the commutator operation on square matrices. If we consider the general linear group $\mathrm{GL}(n, \mathbb{R})$, then its Lie algebra is the general linear Lie algebra $\mathfrak{gl}(n, \mathbb{R}) = M_n(\mathbb{R})$ with the Lie bracket defined by $[A, B] = AB - BA$ for any $A, B \in M_n(\mathbb{R})$. These two objects are related by the **exponential map**:

$$\exp : \mathfrak{gl}(n, \mathbb{R}) \to \mathrm{GL}(n, \mathbb{R}), \quad \exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!}.$$

The exponential map is a local diffeomorphism around the identity element of $\mathrm{GL}(n, \mathbb{R})$ and the zero element of $\mathfrak{gl}(n, \mathbb{R})$. This is because the following differential equation:

$$\frac{d}{dt} \exp(tA) = A \exp(tA) = \exp(tA)A,$$

with initial condition $\exp(0) = \mathrm{I}_n$ has a unique solution. So the differential of exp at 0 is the identity map:

$$d(\exp)_0(A) = A,$$

for any $A \in \mathfrak{gl}(n, \mathbb{R})$.

The exponential map also changes the addition in the Lie algebra to the multiplication in the Lie group. More precisely, we have the **Baker-Campbell-Hausdorff formula**:

$$\exp(tA) \exp(tB) = \exp\left(tA + tB + \frac{t^2}{2}[A, B] + \frac{t^3}{12}([A, [A, B]] + [B, [B, A]]) + \cdots\right),$$

for any $A, B \in \mathfrak{gl}(n, \mathbb{R})$.

When $A = B$, we have

$$\exp(tA) \exp(sA) = \exp((t + s)A).$$

So the bracket operation in the Lie algebra measures the noncommutativity of the multiplication in the Lie group.

The equality

$$\exp(tA) \exp(sA) = \exp((t + s)A)$$

can also be viewed as a group homomorphism from the additive group $\mathbb{R}$ to the Lie group $\mathrm{GL}(n, \mathbb{R})$. So the exponential map gives us a **one-parameter subgroup** of the Lie group $\mathrm{GL}(n, \mathbb{R})$ generated by $A \in \mathfrak{gl}(n, \mathbb{R})$.

Consider the subgroup called **special linear group** $\mathrm{SL}(n, \mathbb{R})$ is defined by

$$\mathrm{SL}(n, \mathbb{R}) = \{A \in \mathrm{GL}(n, \mathbb{R}) : \det A = 1\}.$$

Its Lie algebra is the **special linear Lie algebra**

$$\mathfrak{sl}(n,\mathbb{R}) = \{A \in M_n(\mathbb{R}) : \mathrm{tr}(A) = 0\}.$$

This can be read from the exponential map as follows:

$$\det(\exp(A)) = \exp(\mathrm{tr}(A)).$$

So $\exp(A) \in \mathrm{SL}(n,\mathbb{R})$ if and only if $\mathrm{tr}(A) = 0$.

Similarly, the **orthogonal group** $\mathrm{O}(n,\mathbb{R})$ is defined by

$$\mathrm{O}(n,\mathbb{R}) = \{A \in \mathrm{GL}(n,\mathbb{R}) : A^T A = \mathrm{I}_n\}.$$

Its Lie algebra is the **orthogonal Lie algebra**

$$\mathfrak{o}(n,\mathbb{R}) = \{A \in M_n(\mathbb{R}) : A^T = -A\}.$$

This can be read from the exponential map as follows:

$$(\exp(tA))^T \exp(tA) = \exp(tA^T)\exp(tA) = \exp(t(A^T + A)) + O(t^2).$$

So $\exp(A) \in \mathrm{O}(n,\mathbb{R})$ if and only if $A^T = -A$.

We can also embedd the **complex linear group** $\mathrm{GL}(n,\mathbb{C})$ into $\mathrm{GL}(2n,\mathbb{R})$ via the map

$$A + B\sqrt{-1} \mapsto \begin{pmatrix} A & -B \\ B & A \end{pmatrix},$$

for any $A, B \in M_n(\mathbb{R})$. Then the Lie algebra of $\mathrm{GL}(n,\mathbb{C})$ can be identified with

$$\mathfrak{gl}(n,\mathbb{C}) = \{A + B\sqrt{-1} : A, B \in M_n(\mathbb{R})\}$$

with the bracket operation defined by

$$[X, Y] = XY - YX,$$

for any $X, Y \in \mathfrak{gl}(n,\mathbb{C})$.

The **unitary group** $\mathrm{U}(n)$ is defined by

$$\mathrm{U}(n) = \{A \in \mathrm{GL}(n,\mathbb{C}) : AA^* = \mathrm{I}_n\}.$$

Its Lie algebra is the **unitary Lie algebra**

$$\mathfrak{u}(n) = \{A \in M_n(\mathbb{C}) : A^* = -A\}.$$

The **special unitary group** $\mathrm{SU}(n)$ is defined by

$$\mathrm{SU}(n) = \{A \in \mathrm{GL}(n,\mathbb{C}) : AA^* = \mathrm{I}_n, \det A = 1\}.$$

Its Lie algebra is the **special unitary Lie algebra**

$$\mathfrak{su}(n) = \{A \in M_n(\mathbb{C}) : A^* = -A, \mathrm{tr}(A) = 0\}.$$

A challenge problem is to compute the Lie algebra of the **symplectic group** $\mathrm{Sp}(2n,\mathbb{R})$ defined by

$$\mathrm{Sp}(2n,\mathbb{R}) = \{A \in \mathrm{GL}(2n,\mathbb{R}) : A^T J A = J\},$$

where

$$J = \begin{pmatrix} \mathrm{O}_n & \mathrm{I}_n \\ -\mathrm{I}_n & \mathrm{O}_n \end{pmatrix}.$$

Its Lie algebra is

$$\mathfrak{sp}(2n, \mathbb{R}) = \{A \in M_{2n}(\mathbb{R}) : A^T J + JA = 0\}.$$

A very nontrivial and powerful result in Lie group theory is the following, which we will not prove here.

Theorem 14.1. *For any Lie group G, its closed subgroup H is also a Lie group.*

In other words, if you have any continuous equation defining a subgroup of $\mathrm{GL}(n, \mathbb{R})$, then the solution set is also a Lie group.

The study of Lie group is usually reduced to the study of Lie algebra via the exponential map. The following theorem shows that any finite dimensional Lie algebra can be integrated to a Lie group.

Theorem 14.2 (Lie's Third Theorem)**.** *For any finite dimensional Lie algebra $\mathfrak{g}$ over $\mathbb{R}$, there exists a Lie group G such that the Lie algebra of G is isomorphic to $\mathfrak{g}$.*

The dimension of the Lie group is the same as the dimension of its Lie algebra as a vector space, or the number of parameters needed to describe the Lie group locally around the identity element.

Recall the spetral theorems for orthogonal and unitary matrices, we can now relate them by Lie groups and Lie algebras. For example, the spectral theorem for orthogonal matrices have 2×2 rotation blocks corresponding to complex eigenvalues and also 1×1 blocks corresponding to real eigenvalues ± 1. The corresponding blocks in the Lie algebra (or skew-symmetric matrices) are 2×2 skew-symmetric matrices and also 1×1 zero blocks. The exponential map sends the 2×2 skew-symmetric blocks to the 2×2 rotation blocks, and sends the 1×1 zero blocks to the 1×1 identity blocks. Under the exponential map, each 2×2 skew-symmetric block generates a circle subgroup which is isomorphic to $\mathrm{SO}(2, \mathbb{R})$, and this is a one-parameter subgroup of $\mathrm{O}(n, \mathbb{R})$.

Example 14.1. *Two Lie groups may share the same Lie algebra. For example, the Lie algebras of $\mathrm{SO}(3, \mathbb{R})$ and $\mathrm{SU}(2)$ are both isomorphic to $\mathfrak{so}(3, \mathbb{R})$. We can find this using explicit bases:*

$$\mathfrak{so}(3, \mathbb{R}) = \left\{ \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} : a, b, c \in \mathbb{R} \right\},$$

$$\mathfrak{su}(2) = \left\{ a \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix} + b \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix} : a, b, c \in \mathbb{R} \right\}.$$

The isomorphism can be given by sending

$$\begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix} \mapsto \frac{1}{2\sqrt{-1}} \left(a \begin{pmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{pmatrix} + b \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix} \right).$$

15 Excercies

15.1 Mandatory part

Excercise 15.1. *Is it true that the conjugacy classes of unitary group $\mathrm{U}(2)$ are determined by the trace and determinant? Prove your answer.*

Excercise 15.2.

Definition 15.1 (Orthogonal Group of Signature (p,q))**.** *Let p,q be non-negative integers. The orthogonal group of signature (p,q), denoted by $\mathrm{O}_{p,q}$, is defined as the group of all linear transformations on $\mathbb{R}^{p+q}$ that preserve the bilinear form*

$$\langle x,y\rangle = x_1y_1 + x_2y_2 + \cdots + x_py_p - x_{p+1}y_{p+1} - \cdots - x_{p+q}y_{p+q},$$

for all $x,y \in \mathbb{R}^{p+q}$.

Let W be the space of real trace-zero 2×2 matrices $W = \{A \in M_{2\times 2}(\mathbb{R}) | trace(A) = 0\}$. W has a basis $\mathbf{B} = (w_1, w_2, w_3)$, where

$$w_1 = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix}, w_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, w_3 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

1. *Show that the symmetric bilinear form defined by $\langle A, A'\rangle = \mathrm{trace}\,(AA')$ has signature $(2,1)$. (Hint: use basis $\mathbf{B}$)*
2. *Prove that $P \star A = PAP^{-1}$ defines a linear group operation of $\mathrm{SL}(2,\mathbb{R})$ on the space W.*
3. *Use this operation to define a group homomorphism $\varphi : \mathrm{SL}(2,\mathbb{R}) \to \mathrm{O}_{2,1}$.*
4. *Prove the kernel of this homomorphism is $\{\pm I\}$.*

This construction actually shows that $\mathrm{SL}(2,\mathbb{R})$ is a double cover of $\mathrm{SO}^+_{2,1}$, the connected component of $\mathrm{O}_{2,1}$ containing the identity matrix. Or equivalently, $\mathrm{PSL}(2,\mathbb{R}) = \mathrm{SL}(2,\mathbb{R})/\{\pm I\}$ is isomorphic to the spin group $\mathrm{SO}^+(2,1)$. It is an interesting question to show that the orthogonal group $\mathrm{O}_{2,1}$ has four connected components and identify the geometry of each component.

A similar excercise is to relate $\mathrm{SL}(2,\mathbb{C})$ to the orthogonal group $\mathrm{O}(3,1)$. see Artin Algebra Chapter 9, 4.8

Excercise 15.3. *Let A be the set of all $n\times n$ upper triangular matrices with real entries and all diagonal entries equal to 1. Find the Lie algebra of A and compute its dimension.*

Excercise 15.4. *Show that the intersection of symplectic group $\mathrm{Sp}(2n,\mathbb{R})$ and orthogonal group $\mathrm{O}(2n,\mathbb{R})$ in $\mathrm{GL}(2n,\mathbb{R})$ is isomorphic to the unitary group $\mathrm{U}(n)$. Here we use the embedding of these groups into $\mathrm{GL}(2n,\mathbb{R})$ as Section Examples of Lie groups and Lie algebras.*

Excercise 15.5. *Prove the Jacobi identity of Lie algebra $\mathfrak{gl}(n,\mathbb{R}) = M_n(\mathbb{R})$ using the properties of the matrix commutator. Here*

$$[A,B] = AB - BA.$$

15.2 Optional excercises

Excercise 15.6. *Prove that the Lie algebra of the symplectic group $\mathrm{Sp}(2n,\mathbb{R})$ is*

$$\mathfrak{sp}(2n,\mathbb{R}) = \{A \in M_{2n}(\mathbb{R}) : A^TJ + JA = 0\},$$

where

$$J = \begin{pmatrix} \mathrm{O}_n & \mathrm{I}_n \\ -\mathrm{I}_n & \mathrm{O}_n \end{pmatrix}.$$

Excercise 15.7. *Prove the second and the third order term in the Baker-Campbell-Hausdorff formula when X,Y are elements of $M_n(\mathbb{R})$. That is, prove*

$$\exp(tX)\exp(tY) = \exp\left(t(X+Y) + \frac{1}{2}t^2[X,Y] + \frac{1}{12}t^3([X,[X,Y]] + [Y,[Y,X]]) + \cdots\right)$$

You may use the following expansion of logarithm:

$$\log(I + A) = A - \frac{1}{2}A^2 + \frac{1}{3}A^3 - \cdots$$

16 Group representations: basic concepts

Group representation theory studies the ways in which a group can act on vector spaces via linear transformations. It is a powerful way to study symmetry and structure in mathematics. It is much more than a tool in abstract group theory; it provides deep insights into their nature. Actually, groups are usually appearing as their representations. Representations themselves are crucial and central objects in mathematics. This section introduces the basic definitions and concepts in group representation theory.

16.1 Linear operation, Matrix Representations and Conjugacy

Definition 16.1. *The **general linear group** over a field F is defined as:*

$$\mathrm{GL}(n, F) = \{A \in M_n(F) \mid \det A \neq 0\}.$$

Definition 16.2. *Let V be a finite-dimensional vector space over F. The **general linear group of** V is:*

$$\mathrm{GL}(V) = \{f \colon V \to V \mid f \text{ is an invertible linear transformation}\}.$$

Let $\dim V = n$ and fix a basis $B = (e_1, \ldots, e_n)$ of V. For any linear transformation $f \colon V \to V$, we have the matrix representation $R_B(f)$ such that

$$f(e_1, \ldots, e_n) = (e_1, \ldots, e_n) \cdot R_B(f).$$

If $B' = (v_1, \ldots, v_n)$ is another basis with change-of-basis matrix P (i.e., $(v_1, \ldots, v_n) = (e_1, \ldots, e_n) \cdot P$), then

$$R_{B'}(f) = P^{-1} R_B(f) P.$$

Definition 16.3. *A **representation** of a group G on a vector space V is a group homomorphism*

$$\rho \colon G \to \mathrm{GL}(V).$$

Equivalently, it is a group action $G \times V \to V$, denoted by $(g, v) \mapsto g \cdot v$, satisfying:

$$g \cdot (v + w) = g \cdot v + g \cdot w,$$
$$g \cdot (\lambda v) = \lambda (g \cdot v),$$

*for all $g \in G$, $v, w \in V$, and $\lambda \in F$. Such an action is called a **linear group action**.*

Given a representation $\rho : G \to \mathrm{GL}(V)$ and a basis B of V, we obtain a **matrix representation** or a group homomorphism:

$$R \colon G \to \mathrm{GL}(n, F), \quad g \mapsto R_B(\rho(g)).$$

Two matrix representations R and R' are said to be **conjugate** if there exists $P \in \mathrm{GL}(n, F)$ such that

$$P R(g) P^{-1} = R'(g) \quad \forall g \in G.$$

Definition 16.4. *Two representations $\rho : G \to \mathrm{GL}(V)$ and $\rho' : G \to \mathrm{GL}(V')$ are **isomorphic** if there exists an isomorphism $f : V \to V'$ such that*

$$f(g \cdot v) = g \cdot f(v) \quad \forall g \in G, v \in V.$$

From the discussion about change of basis, we see that

Proposition 16.1. *Two representations are isomorphic if and only if their corresponding matrix representations are conjugate.*

To summarize, we have three equivalent viewpoints of group representations:

- A group homomorphism $\rho\colon G \to \mathrm{GL}(V)$;
- A linear group action of G on V;
- A matrix representation $R\colon G \to \mathrm{GL}(n, F)$ for some n.

The isomorphism classes of finite-dimensional representations correspond to conjugacy classes of matrix representations.

Definition 16.5 (Faithful Representation)**.** *A representation $\rho : G \to \mathrm{GL}(V)$ is called **faithful** if it is injective as a group homomorphism, i.e., if $\ker \rho = \{e\}$. In terms of group actions, this means that if $g \cdot v = v$ for all $v \in V$, then $g = e$.*

16.2 Unitary Representations

Assume $F = \mathbb{C}$. Let V be a complex inner product space, with postive definite Hermitian form $\langle \cdot, \cdot \rangle$. A representation $\rho\colon G \to \mathrm{GL}(V)$ is called a **unitary representation** if $\mathrm{Im}\, \rho \subseteq U(V)$, where $U(V)$ is the unitary group.

Theorem 16.1. *Let G be a finite group and $\rho\colon G \to \mathrm{GL}(V)$ a representation. Then there exists a positive definite Hermitian form $\langle \cdot, \cdot \rangle$ on V that is G-invariant, i.e.,*

$$\langle g \cdot v, g \cdot w \rangle = \langle v, w \rangle \quad \forall g \in G, v, w \in V.$$

Hence, ρ is unitary with respect to this form.

Proof. Start with any positive definite Hermitian form $\langle \cdot, \cdot \rangle_0$ on V. Define a new form by averaging over the group:

$$\langle v, w \rangle_G = \frac{1}{|G|} \sum_{g \in G} \langle g \cdot v, g \cdot w \rangle_0.$$

One checks that $\langle \cdot, \cdot \rangle_G$ is still a positive definite Hermitian form and is G-invariant. □

This is a fundamental result in representation theory of finite groups, allowing us to always work with unitary representations when the field is $\mathbb{C}$. The key idea is to average an arbitrary inner product over the group to obtain a G-invariant one. This technique is known as **averaging** and will be used frequently in representation theory. For example, if you work with real representations, you can also try to prove that every real representation of a finite group admits an invariant real inner product, or equivalently, any matrix representation is conjugate to one with image in the orthogonal group $\mathrm{O}(n)$.

16.3 Invariant Subspaces and Orthogonal Decomposition

Definition 16.6. *Let $\rho\colon G \to \mathrm{GL}(V)$ be a representation. A subspace $W \subseteq V$ is called G-**invariant** if $g \cdot w \in W$ for all $g \in G$ and $w \in W$. This also means that the restriction of ρ to W defines a **subrepresentation**.*

Theorem 16.2. *Let G be a finite group and $\rho\colon G \to \mathrm{GL}(V)$ a unitary representation. For any G-invariant subspace $W \subseteq V$, the orthogonal complement $W^{\perp}$ (with respect to a G-invariant inner product) is also G-invariant, and we have the decomposition*

$$V = W \oplus W^{\perp}.$$

Proof. Choose a G-invariant inner product $\langle \cdot, \cdot \rangle$ on V. For $w \in W$ and $v \in W^{\perp}$, we have

$$\langle w, g \cdot v \rangle = \langle g^{-1} \cdot w, v \rangle = 0,$$

since $g^{-1} \cdot w \in W$. Hence $g \cdot v \in W^{\perp}$, so $W^{\perp}$ is G-invariant. □

16.4 Irreducible Representations

Definition 16.7. *A representation $\rho\colon G \to \mathrm{GL}(V)$ is called **irreducible** if V has no nontrivial G-invariant subspaces (i.e., the only G-invariant subspaces are $\{0\}$ and V itself).*

Example 16.1. *Let $F = \mathbb{F}_p$ (the finite field with p elements), $G = \mathbb{F}_p$ (additive group), and $V = \mathbb{F}_p^2$. Define the representation $R : G \to \mathrm{GL}(2, \mathbb{F}_p)$ by*

$$t \mapsto \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}.$$

Then the subspace $W = \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \mid x \in \mathbb{F}_p \right\}$ is G-invariant. However, there is no complementary G-invariant subspace W' such that $V = W \oplus W'$. This shows that the theorem on orthogonal decomposition requires an inner product (and hence the field to be $\mathbb{C}$ and the representation to be unitary).

16.5 Semisimplicity

A fundamental result in the representation theory of finite groups over $\mathbb{C}$ is that every representation can be decomposed into irreducible pieces. This property is known as complete reducibility or semisimplicity.

Theorem 16.3 (Maschke's Theorem)**.** *Let G be a finite group and V a finite-dimensional representation of G over $\mathbb{C}$. Then V is isomorphic to a direct sum of irreducible representations. That is,*

$$V \cong V_1^{\oplus m_1} \oplus V_2^{\oplus m_2} \oplus \cdots \oplus V_k^{\oplus m_k},$$

where the V_i are pairwise non-isomorphic irreducible representations of G.

Proof. We proceed by induction on $\dim V$. If V is irreducible, we are done. Otherwise, let $W \subset V$ be a nontrivial G-invariant subspace. By Theorem 3.2, we can find a G-invariant inner product on V, and then $V = W \oplus W^{\perp}$ with $W^{\perp}$ also G-invariant. By induction, both W and $W^{\perp}$ decompose into irreducibles, hence so does V. □

This theorem shows that irreducible representations are the basic building blocks of all representations of finite groups over $\mathbb{C}$. The study of these irreducible representations—their classification, dimensions, and characters—is the central goal of representation theory.

17 Examples of Group Representations

Now that we have established the basic framework, let us examine some concrete examples of group representations to illustrate these concepts in action.

Example 17.1 (Trivial Representation)**.** *The simplest example is the* ***trivial representation****:*

$$\rho\colon G \to \mathrm{GL}(1,\mathbb{C}) = \mathbb{C}^\times, \quad g \mapsto 1.$$

Here every group element acts as the identity transformation.

Example 17.2 (Dihedral Group)**.** *The dihedral group D_n has a natural representation as orthogonal transformations of the plane:*

$$D_n \to O(2) \subset \mathrm{GL}(2,\mathbb{R}) \subset \mathrm{GL}(2,\mathbb{C}).$$

This representation is generated by a rotation a and a reflection b satisfying the relations $a^n = 1$, $b^2 = 1$, and $bab^{-1} = a^{-1}$.

Definition 17.1 (Group Algebra)**.** *Given a group G and a field F, the* ***group algebra*** *$F[G]$ is defined as:*

$$F[G] = \left\{ \sum_{g \in G} a_g g \mid a_g \in F \right\}.$$

The group G acts on $F[G]$ by left multiplication:

$$h \cdot \left(\sum_{g \in G} a_g g \right) = \sum_{g \in G} a_g (hg).$$

This is called the ***regular representation*** *of G.*

Example 17.3 (Permutation Representation)**.** *Let G act on a finite set $X = \{x_1, \ldots, x_n\}$. Consider the vector space $F^X = \{\sum_{x \in X} a_x x \mid a_x \in F\}$. Then G acts on F^X by permuting the basis vectors:*

$$g \cdot \left(\sum_{x \in X} a_x x \right) = \sum_{x \in X} a_x (g \cdot x).$$

This yields a representation $G \to \mathrm{GL}(n, F)$.

For example, take $G = S_n$ acting on the standard basis $\{e_1, \ldots, e_n\}$ of $\mathbb{C}^n$. Then the subspace $W = \mathrm{span}\{e_1 + \cdots + e_n\}$ is G-invariant. With respect to the standard Hermitian form, its orthogonal complement

$$W^\perp = \left\{ \sum_{i=1}^n a_i e_i \mid \sum_{i=1}^n a_i = 0 \right\}$$

is also G-invariant, giving a decomposition $\mathbb{C}^n = W \oplus W^\perp$ as S_n-representations.

18 Constructions of Representations

Having seen several examples, we now turn to systematic methods for constructing new representations from existing ones. These constructions are fundamental for building more complex representations from simpler building blocks.

18.1 Direct Sums and Quotients

Definition 18.1 (Direct Sum)**.** *Let V and W be representations of G. Their **direct sum** $V \oplus W$ is the representation on the direct sum of vector spaces, with G acting componentwise:*

$$g \cdot (v, w) = (g \cdot v, g \cdot w).$$

In matrix form, if V and W have matrix representations R_V and R_W, then the representation on $V \oplus W$ is given by block diagonal matrices:

$$g \mapsto \begin{bmatrix} R_V(g) & 0 \\ 0 & R_W(g) \end{bmatrix}.$$

Definition 18.2 (Quotient Representation)**.** *Let $W \subseteq V$ be a G-invariant subspace. The **quotient representation** on V/W is defined by*

$$g \cdot (v + W) = (g \cdot v) + W.$$

In matrix form, after choosing a basis adapted to W, the representation on V takes the form:

$$g \mapsto \begin{bmatrix} R_W(g) & * \\ 0 & R_{V/W}(g) \end{bmatrix},$$

where R_W and $R_{V/W}$ are the matrix representations on W and V/W respectively.

18.2 Dual Representation

Another important construction is the dual representation, which corresponds to the contragredient action.

Definition 18.3 (Dual Representation)**.** *Let V be a representation of G. The **dual representation** $V^* = \mathrm{Hom}(V, \mathbb{C})$ is defined by*

$$(g \cdot f)(v) = f(g^{-1} \cdot v) \quad \forall f \in V^*, v \in V, g \in G.$$

In matrix terms, if we choose a basis $B = (v_1, \ldots, v_n)$ of V and the dual basis $B^* = (f_1, \ldots, f_n)$ of V^* (satisfying $f_i(v_j) = \delta_{ij}$), then the matrix representation for V^* is given by:

$$R_{V^*}(g) = (R_V(g)^{-1})^\top.$$

19 Excercies

19.1 Mandatory part

Excercise 19.1. *Let V be an irreducible representation of a finite group G over $\mathbb{C}$. Assume V is not the trivial representation. Prove that for any $v \in V$, we have $\sum_{g \in G} g \cdot v = 0$. (For cyclic groups, this is known to be an identity of roots of unity.)*

Excercise 19.2. *(Artin Algebra Chapter 10, 3.5) Let x be a generator of a cyclic group G of order p. Sending*

$$x \mapsto \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

defines a matrix representation $G \to \mathrm{GL}_2(\mathbb{F}_p)$. Prove that this representation is not the direct sum of irreducible representations.

Excercise 19.3. *(Artin Algebra Chapter 10, 3.4) Let $\langle \cdot, \cdot \rangle$ be a nondegenerate skew-symmetric form on a vector space V, and let ρ be a representation of a finite group G on V. Prove that the averaging process produces a G-invariant skew-symmetric form on V, and show by example that the form obtained in this way need not be nondegenerate.*

Excercise 19.4. *(Artin algebra Chapter 10, 2.2) Consider the standard two-dimensional complex representation of the dihedral group D_n. For which n is this an irreducible complex representation?*

Here the standard representation is given by the action of D_n as the group of symmetries of a regular n-gon in the plane, or equivalently, the representation defined by the matrices

$$r \mapsto \begin{pmatrix} \cos(2\pi/n) & -\sin(2\pi/n) \\ \sin(2\pi/n) & \cos(2\pi/n) \end{pmatrix}, \quad s \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

where r is a rotation by $2\pi/n$ and s is a reflection.

Excercise 19.5. *(Artin Algebra Chapter 10, 3.1) Let G be a cyclic group of order 3. The matrix*

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

has order 3, so it defines a matrix representation of G on $\mathbb{C}^2$. Use the averaging process to produce a G-invariant form from the standard Hermitian product $\langle X, Y \rangle = X^* Y$ on $\mathbb{C}^2$.

Excercise 19.6. *(Artin Algebra Chapter 10, 3.2) Let $\rho \colon G \to \mathrm{GL}(V)$ be a representation of a finite group G on a real vector space V. Prove the following:*

1. *There exists a G-invariant, positive definite symmetric form $\langle \cdot, \cdot \rangle$ on V.*
2. *ρ is a direct sum of irreducible representations.*
3. *Every finite subgroup of $\mathrm{GL}_n(\mathbb{R})$ is conjugate to a subgroup of $\mathrm{O}(n)$.*

Excercise 19.7. *Show that the dual representation of an irreducible representation is also irreducible.*

19.2 Optional excercises

Excercise 19.8. *In this part, we prove the semisimplicity theorem (Maschke's theorem) for any representation of group G and field F such that the characteristic of F does not divide the order of G, using the averaging technique.*

Let V be a representation of G over F, and let W be a G-invariant subspace of V. Prove that there exists a complementary G-invariant subspace W' of V such that $V = W \oplus W'$. (Hint: start with any complementary subspace and then use averaging to construct a projection onto W whose kernel is W'.)

Excercise 19.9. *Show that the dual representation is isomorphic to the original representation for any finite group representation if and only if there exists a nondegenerate G-invariant bilinear form on the representation space.*

20 Hom, tensor product and Schur's lemma

20.1 Hom space

Given two representations, we can form the space of linear maps between them, which itself carries a natural representation.

Definition 20.1 (Hom Representation)**.** *Let V and W be representations of G. The space* $\mathrm{Hom}(V, W)$ *of linear maps from V to W becomes a representation of G via:*

$$(g \cdot T)(v) = g \cdot T(g^{-1} \cdot v) \quad \forall T \in \mathrm{Hom}(V, W), v \in V, g \in G.$$

20.2 Tensor Product

Tensor products provide a way to combine representations, which is particularly important in physics and the study of product groups.

Definition 20.2 (Tensor Product of Representations)**.** *Let V and W be representations of G. Their* ***tensor product*** *$V \otimes W$ is the representation on the tensor product of vector spaces, with G acting diagonally:*

$$g \cdot (v \otimes w) = (g \cdot v) \otimes (g \cdot w).$$

If V has basis $\{v_1, \ldots, v_n\}$ and W has basis $\{w_1, \ldots, w_m\}$, then $\{v_i \otimes w_j\}$ is a basis of $V \otimes W$, and any element can be written as $\sum a_{ij} v_i \otimes w_j$. The action of G in coordinates is given by the Kronecker product of the matrices for V and W.

Lemma 20.1. *There is a natural isomorphism of vector spaces:*

$$\Gamma : V^* \otimes W \to \mathrm{Hom}(V, W),$$

defined by $\Gamma(f \otimes w)(v) = f(v)w$. Moreover, if V and W are representations of G, then Γ is an isomorphism of representations when $\mathrm{Hom}(V, W)$ *is equipped with the action*

$$(g \cdot T)(v) = g \cdot T(g^{-1} \cdot v).$$

Proof. For surjectivity, let $B = (v_1, \ldots, v_n)$ be a basis of V and $C = (w_1, \ldots, w_m)$ a basis of W. Given $T \in \mathrm{Hom}(V, W)$ with matrix (a_{ij}) relative to B and C, let $(f_1, \ldots, f_n)$ be the dual basis of B. Then

$$\Gamma\left(\sum_{i,j} a_{ij} f_j \otimes w_i\right)(v_k) = \sum_i a_{ik} w_i = T(v_k).$$

Thus Γ is surjective, and by dimension count it is an isomorphism. □

20.3 G-Homomorphisms

To compare different representations and understand their structure, we need the notion of maps that respect the group action.

Definition 20.3 (G-Homomorphism)**.** *Let V and W be representations of G. A linear map $f : V \to W$ is called a* ***G-homomorphism*** *(or* ***intertwining operator****) if*

$$f(g \cdot v) = g \cdot f(v) \quad \forall g \in G, v \in V.$$

We denote by $\mathrm{Hom}_G(V, W)$ *the space of all such maps.*

Proposition 20.1. *Let $f : V \to W$ be a G-homomorphism. Then* $\ker f$ *is a G-invariant subspace of V, and $\Im f$ is a G-invariant subspace of W. Moreover, the induced map $\bar{f} : V/\ker f \to \Im f$ is an isomorphism of representations.*

20.4 Schur's Lemma

One of the most powerful tools in representation theory is Schur's Lemma, which describes the structure of G-homomorphisms between irreducible representations.

Theorem 20.1 (Schur's Lemma)**.** *Let V and W be irreducible representations of G over $\mathbb{C}$. Then:*

1. *If V and W are not isomorphic, then $\mathrm{Hom}_G(V, W) = 0$.*
2. *If $V = W$, then $\mathrm{Hom}_G(V, V) = \mathbb{C} \cdot \mathrm{I}_V$.*

Proof. Let $T \in \mathrm{Hom}_G(V, W)$. Since V is irreducible and $\ker T$ is G-invariant, either $\ker T = V$ or $\ker T = 0$. Similarly, $\Im T$ is G-invariant in W, so either $\Im T = 0$ or $\Im T = W$.

If $T \neq 0$, then $\ker T = 0$ and $\Im T = W$, so T is an isomorphism. This proves (1).

For (2), let $T \in \mathrm{Hom}_G(V, V)$. Since $\mathbb{C}$ is algebraically closed, T has an eigenvalue λ. Then $T - \lambda \mathrm{I}_V \in \mathrm{Hom}_G(V, V)$ is not invertible, hence must be zero by the argument above. Thus $T = \lambda \mathrm{I}_V$. □

Conclusion

In this chapter we have introduced the fundamental concepts of group representation theory. Starting from basic definitions, we explored various constructions of representations and proved key results such as Schur's Lemma and Maschke's Theorem. These tools provide the foundation for the deeper study of group representations, including character theory and the classification of irreducible representations, which we will explore in subsequent chapters.

Index