Algebra 2

Chenglong Yu

January 14, 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 Problem 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 Problem 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 $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n\}$ be an arbitrary basis of a Euclidean space V. One can construct an orthonormal basis $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n\}$ such that*

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

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

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 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.

Problem 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*

$$A = \begin{pmatrix} 3 & 2 & 100 \\ 4 & 0 & 0 \\ 0 & 0 & -5 \end{pmatrix}.$$

Problem 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.)*

Problem 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 $A = (\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle)_{1 \le i,j \le n}$ 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.*

Problem 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 $\boldsymbol{u}, \boldsymbol{v} \in V$,*

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \le \|\boldsymbol{u}\| \cdot \|\boldsymbol{v}\|.$$

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 $\mathbb{R}^2$.*

Problem 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.*

Problem 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 3 and induction.)*

Problem 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.*

Problem 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)*

Problem 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.

Problem 10. *Prove that the **Hilbert matrix** of order n,*

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

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

Problem 11. *Prove the **reversed Cauchy-Schwarz inequality in Minkowski space**:*
For all $\boldsymbol{v} = (v_0, v_1, \ldots, v_n), \boldsymbol{w} = (w_0, w_1, \ldots, w_n) \in \mathbb{R}^{n+1}$ 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

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

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

In terms of the Lorentz form $\langle \boldsymbol{v}, \boldsymbol{w} \rangle = v_0 w_0 - v_1 w_1 - \cdots - v_n w_n$, the inequality can be rewritten as

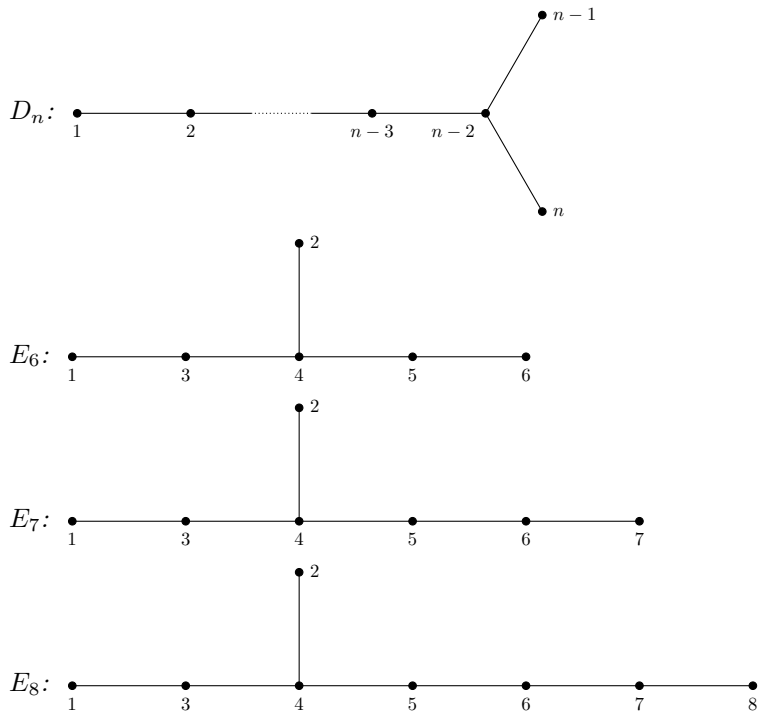
$$\langle \boldsymbol{v}, \boldsymbol{v} \rangle \langle \boldsymbol{w}, \boldsymbol{w} \rangle \leq \langle \boldsymbol{v}, \boldsymbol{w} \rangle^2.$$

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 (2).

Problem 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:*

A_n:



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 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* $\mathrm{Proj}_W \colon \boldsymbol{v} \to \boldsymbol{w}_0$ *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 $(\mathbb{R}^n, \langle \cdot, \cdot \rangle_{\mathrm{st}})$. 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

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

If A is orthogonal, then $A^T A = I$, so $\langle T_A(\boldsymbol{u}), T_A(\boldsymbol{v}) \rangle_{\mathrm{st}} = \boldsymbol{u}^T \boldsymbol{v} = \langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\mathrm{st}}$. Conversely, if $\langle T_A(\boldsymbol{u}), T_A(\boldsymbol{v}) \rangle_{\mathrm{st}} = \langle \boldsymbol{u}, \boldsymbol{v} \rangle_{\mathrm{st}}$ 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 $A^T A$ is δ_{ij}, so $A^T A = I$ 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*

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

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 $P_1 P_2 \in \mathrm{SO}(2,\mathbb{R})$. If the two reflectoin axes form an angle $\frac{\theta}{2}$, then $P_1 P_2$ 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 $P^T AP$. 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 $A = Q^T BQ$ $(= Q^{-1}BQ)$.*

(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., $A^T A = I$.

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

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

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

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

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 $\mathcal{B} = \{e_1, \dots, e_n\}$ be an orthonormal basis of V, and let A be the matrix representation of T with respect to $\mathcal{B}$, i.e.,

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

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

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

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*

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

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:

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

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

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

*. 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

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

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

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

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

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

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

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

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

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

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

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

Thus,

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

□

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 $\mathrm{rank}\, B \leq k$, then*

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

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 $\operatorname{rank} B \leq k$, we have $\dim \ker B \geq n - k$, 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,

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

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

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

By the previous case,

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

□

Now we give the proof of Theorem 7.2:

Proof.

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

□

7.2 Application: Low-dimensional Fitting

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 $\alpha_1, \ldots, \alpha_m$ 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 $W_k = \mathrm{span}_{\mathbb{R}}\{v_1, \ldots, v_k\}$. Then W_k minimizes the following value:

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

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

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*

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

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 $(a_1,\ldots,a_n)$,

$$y = x_1a_1 + \cdots + x_na_n.$$

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

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

Since Ax ranges over W, there exists x such that $Ax = \text{Proj}_W b$. 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 $A\widehat{x} - b \perp \operatorname{im} A$. 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*

$$y = x_1a_1 + \cdots + x_na_n + c$$

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

Problem 13. *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.*

Problem 14 (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*.)*

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) $h(\lambda_1 x_1 + \lambda_2 x_2, y) = \lambda_1 h(x_1, y) + \lambda_2 h(x_2, y)$;

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

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

Example 9.1. *The map*

$$\begin{aligned} h\colon \mathbb{C}^n \times \mathbb{C}^n &\to \mathbb{C} \\ (x, y) &\mapsto \sum_{i=1}^{n} x_i \overline{y}_i \end{aligned}$$

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*

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

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

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

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,

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

Thus

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

□

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 $\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. Since the diagonal matrix is also Hermitian $\Rightarrow \lambda_i$ are real.
2. Suppose a unitary matrix is unitarily similar to a diagonal matrix $\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$. 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.*

Index