Lecture 1: dot product, length and angle

Shengkui Ye

January 30, 2023

1 Orthogonal basis and projections

For two vectors $x = (x_1, x_2, ..., x_n)^T, y = (y_1, y_2, ..., y_n)^T \in \mathbb{R}^n$, we already know that the dot product $x \circ y = x_1y_1 + x_2y_2 + \cdots + x_ny_n = x^Ty$. The length of the vector x is $\| x \| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$. The angle between nonzero vectors x, y is $\measuredangle(x, y) = \arccos \frac{x \circ y}{\|x\|\|y\|}$.

Definition 1 *An orthogonal basis S of $\mathbb{R}^n$ is a basis such that any two distinct elements $u, v \in S$ are orthogonal.*

Lemma 2 *Let $S = \{v_1, v_2, ..., v_n\}$ be an orthogonal basis of $\mathbb{R}^n$. Any element $x \in \mathbb{R}^n$ is a linear combination*

$$x = a_1v_1 + a_2v_2 + \cdots + a_nv_n$$

with $a_i = \frac{x \circ v_i}{\|v_i\|^2}$ for each i.

Proof. Note that $x \circ v_i = a_iv_i \circ v_i$. ■

Lemma 3 *Let H be a subspace of $\mathbb{R}^n$. Any element $x \in \mathbb{R}^n$ is writen uniquely as $x = x_1 + x_2$ with $x_1 \in H$ and $x_2 \in H^\perp$. The x_1 is called the projection of x onto H, denoted by $proj_H(x)$.*

Proof. For the existence, let $x_1 \in H$ be a vector such that $\|x - x_1\| = \inf_{y \in H} \|x - y\|$. Choose $x_2 = x - x_1$. By properties of triangles, we know that $x - x_1$ is orthogonal to x_1. The existence can also be proved by assuming that H has an orthogonal basis (saying $\{v_1, ..., v_k\}$) which can be extended to be an orthogonal basis of $\mathbb{R}^n$. Then $x_1 = \sum_{i=1}^k a_iv_i = \sum_{i=1}^k \frac{x \circ v_i}{\|v_i\|^2} v_i$ by the previous lemma.

If $x = x_1' + x_2'$ with $x_1' \in H$ and $x_2' \in H^\perp$, then $x_1 - x_1' = x_2' - x_2 \in H \cap H^\perp = \{0\}$. ■

If H is spanned by a nonzero vector u, then $x_1 = ku$ for some k. Then $x \circ u = (x_1 + x_2) \circ u = ku \circ u$ and thus $k = \frac{x \circ u}{u \circ u}$.

Example 4 *Let $x = [7, 6]^T$ and $u = [4, 2]^T$. Find $proj_u(x)$.*

A set $\{v_1, v_2, \cdots, v_k\}$ is orthonormal if $v_i \circ v_i = 1$ and $v_i \circ v_j = 0$ for any $i \neq j$.

Example 5 *Show that* $\{(1/\sqrt{2}, 1/\sqrt{2})^T, (1/\sqrt{2}, -1/\sqrt{2})^T\}$ *is orthonormal.*

Lemma 6 *An* $m \times n$ *matrix* U *has orthonormal columns if and only if* $U^T U = I$. *An orthogonal matrix* A *is a square invertible matrix* A *such that* $A^T A = I$.

Lemma 7 *Let* U *be an* $m \times n$ *matrix with orthonormal columns, and let* $x, y \in \mathbb{R}^n$. *Then*
a) $\|Ux\| = \|x\|$;
b) $Ux \circ Uy = x \circ y$;
c) $Ux \circ Uy = 0$ *if and only* $x \circ y = 0$.

Lemma 8 *Let* $f : \mathbb{R}^n \to \mathbb{R}^n$ *be a linear map. Suppose that* f *is distance-preserving, i.e.* $\|f(x) - f(y)\| = \|x - y\|$ *for any* $x, y \in \mathbb{R}^n$. *Then* f *is angle-preserving, i.e.* $\measuredangle(f(x), f(y)) = \measuredangle(x, y)$ *for any* $x, y \in \mathbb{R}^n$;

Proof. Suppose that f is distance-preserving. We have $\|f(x)\| = \|x\|$ and $\|f(x + y)\| = \|x + y\|$ for any $x, y \in \mathbb{R}^n$. But this implies that

$$\begin{aligned} \|f(x+y)\|^2 &= \|f(x) + f(y)\|^2 = (f(x) + f(y)) \circ (f(x) + f(y)) \\ &= f(x) \circ f(x) + f(y) \circ f(y) + 2f(x) \circ f(y) \\ &= \|x + y\|^2 \\ &= x \circ x + y \circ y + 2x \circ y \end{aligned}$$

and $f(x) \circ f(y) = x \circ y$. Note that $\cos \measuredangle(x, y) = \frac{x \circ y}{\|x\|\|y\|} = \frac{f(x) \circ f(y)}{\|f(x)\|\|f(y)\|} = \cos \measuredangle(f(x), f(y))$, which gives $\measuredangle(f(x), f(y)) = \measuredangle(x, y)$. ■

Lemma 9 *A linear map* $f : \mathbb{R}^n \to \mathbb{R}^n$ *is distance-preserving if and only if the standard representation matrix* A_f *of* f *is orthogonal.*

Proof. Let $x \in \mathbb{R}^n$ be arbitrary vector. If A is orthogonal, we have $\|f(x)\|^2 = \|Ax\|^2 = Ax \circ Ax = (Ax)^T Ax = x^T (A^T A)x = x \circ x = \|x\|^2$. Therefore, f is distance-preserving.

Suppose that f is distance-preserving. The proof of the previous lemma shows that $f(x) \circ f(y) = x \circ y$ for any $x, y \in \mathbb{R}^n$. Choose $x, y \in \{e_1, e_2, ..., e_n\}$, the standard basis, to get that $f(e_i) \circ f(e_i) = e_i^T A^T A e_j = e_i \circ e_j$, which is the (i, j)-th entry of $A^T A$. Therefore, $A^T A = I_n$. ■

Example 10 *Show that* $\begin{bmatrix} \frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{bmatrix}$ *is orthogonal.*

Theorem 11 *(Orthogonal decomposition theorem) Let* W *be a subspace of* $\mathbb{R}^n$. *Then each element* x *in* $\mathbb{R}^n$ *is a sum* $\hat{x} + z$ *with* $\hat{x} \in W$ *and* $z \in W^\perp$. *In fact, if* $\{u_1, u_2, \cdots, u_p\}$ *is any orthogonal basis of* W, *then*

$$\hat{x} = \frac{x \circ u_1}{u_1 \circ u_1} u_1 + \frac{x \circ u_2}{u_2 \circ u_2} u_2 + \cdots + \frac{x \circ u_p}{u_p \circ u_p} u_p$$

and $z = x - \hat{x}$.

Proof. Let $\{u_1, u_2, \cdots, u_p\}$ be an orthogonal basis of W. Extend this set to be a basis $\{u_1, u_2, \cdots, u_n\}$ of $\mathbb{R}^n$. The proof is finished. ∎

Corollary 12 *If $\{u_1, u_2, \cdots, u_p\}$ is any* orthonormal *basis of W, then the projection of $x \in \mathbb{R}^n$ onto W is*

$$\hat{x} = UU^T x$$

where $U = [u_1, u_2, \cdots, u_p]$.

Example 13 *Let* $u_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, u_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$ *and* $y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. *Show that $\{u_1, u_2\}$ is an orthogonal basis for $W = \mathrm{Span}\{u_1, u_2\}$. Write y as a sum of a vector in W and a vector in the orthogonal complement of W.*

2 The Gram-Schmidt process

Example 14 *Let* $u_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, u_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$ *and $W = \mathrm{Span}\{u_1, u_2\}$. Find an orthogonal basis of W.*

Proof. Take $\{u_1, u_2 - \frac{u_2 \circ u_1}{u_1 \circ u_1} u_1\}$. ∎

Theorem 15 *(Gram-Schmidt process) Given a basis $\{x_1, x_2, \cdots, x_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define*

$$\begin{aligned} v_1 &= x_1, \\ v_2 &= x_2 - \frac{x_2 \circ v_1}{v_1 \circ v_1} v_1, \cdots, \\ v_p &= x_p - \frac{x_p \circ v_1}{v_1 \circ v_1} v_1 - \frac{x_p \circ v_2}{v_2 \circ v_2} v_2 - \cdots - \frac{x_p \circ v_{p-1}}{v_{p-1} \circ v_{p-1}} v_{p-1}. \end{aligned}$$

Then $\{v_1, v_2, \cdots, v_p\}$ is an orthogonal basis for W. Moreover,

$$\mathrm{Span}\{v_1, \cdots, v_k\} = \mathrm{Span}\{x_1, \cdots, x_k\}, \textit{for each } k \leq p.$$

Proof. Inductively, we assume that $\mathrm{Span}\{v_1, \cdots, v_{k-1}\} = \mathrm{Span}\{x_1, \cdots, x_{k-1}\}$. Since $x_k = v_k + z_{k-1}$ for a vector $z_{k-1} \in \mathrm{Span}\{v_1, \cdots, v_{k-1}\}$, we see that $x_k \in \mathrm{Span}\{v_1, \cdots, v_{k-1}, v_k\}$ and similarly $v_k \in \mathrm{Span}\{x_1, \cdots, x_{k-1}, x_k\}$. ∎

Example 16 *Let* $x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$. *Find an orthonormal basis of* $\mathrm{Span}\{x_1, x_2, x_3\}$.

Corollary 17 *(QR factorization) Let A be an invertible matrix. Then $A = QR$ for an orthogonal matrix Q and an upper triangular matrix R.*

Proof. Let $A = [x_1, x_2, \cdots, x_n]$. The Gram-Schmidt process produces a matrix $P = [v_1, v_2, \cdots, v_n]$. Note that $A = PS$ for a strictly upper triangular matrix S. Let $D = \text{diag}(v_1 \circ v_1, v_2 \circ v_2, \cdots, v_n \circ v_n)$ be the diagonal matrix, and $Q = [\frac{v_1}{v_1 \circ v_1}, \frac{v_2}{v_2 \circ v_2}, \cdots, \frac{v_n}{v_n \circ v_n}]$. Therefore, we have $A = (PD^{-1})DS = QR$, with $R = DS$. Note that $QQ^T = I_n$. ■

3 Least-square problem

For a matrix $A_{m\times n}$ and $b \in \mathbb{R}^n$, we know that $Ax = b$ may not have a solution. An element $x_0 \in \mathbb{R}^n$ is called a least-square problem for $Ax = b$ if

$$\|b - Ax_0\| \leq \|b - Ax\|$$

for any $x \in \mathbb{R}^n$.

Theorem 18 *A vector x_0 is a least-square solution of $Ax = b$ if and only if $A^T Ax_0 = A^T b$.*

Proof. Denote by Ax_0 the projection of b onto the column space $Col(A)$. By the orthogonal decomposition theorem, $b - Ax_0 \in Col(A)^\perp = \text{Nul}A^T$. Therefore, $A^T(b - Ax_0) = 0$ and $A^T Ax_0 = A^T b$. Conversely, when $A^T Ax_0 = A^T b$, we have $A^T(b - Ax_0) = 0$ and $b - Ax_0 = \text{Nul}A^T = \text{Col}(A)^\perp$. This implies that $\|b - Ax_0\| \leq \|b - Ax\|$ for any $x \in \mathbb{R}^n$. ■

Corollary 19 *$Ax = b$ has a unique least-square solution if and only if the columns of A are linearly independent.*

Proof. By the previous theorem, it's enough to prove that $A^T A$ is invertible if and only if the columns of A are linearly independent. When $A^T A$ is invertible, $n = rank(A^T A) \leq rank(A)$. Therefore, the columns of A are linearly independent. Conversely, when the columns of A are linearly independent, $Ax = 0$ has only the trivial solution $x = 0$. If $A^T Ax = 0$, then $0 = x^T A^T Ax = (Ax)^T(Ax)$, which implies that $Ax = 0$ and thus $x = 0$. Therefore, $A^T A$ is invertible. ■

Example 20 *Find a least-square solution for $Ax = b$, where*

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, b = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

Lecture 2: inner product, length and angle

Shengkui Ye

February 13, 2023

1 Inner product: definitions

The following are generalizations of the dot product.

Definition 1 *An inner product* $\langle , \rangle$ *on a real vector space* V *is a function* $\langle , \rangle : V \times V \to \mathbb{R}$ *such that*

1. $\langle u, v\rangle = \langle v, u\rangle$ *for any* $v, u \in V$;
2. $\langle v, a_1u_1 + a_2u_2\rangle = a_1\langle v, u_1\rangle + a_2\langle v, u_2\rangle$ *for any* $u_1, u_2, v \in V$ *and any* $a_1, a_2 \in \mathbb{R}$;
3. $\langle u, u\rangle \geq 0$ *and* $\langle u, u\rangle = 0$ *if and only if* $u = 0$.

If the function $\langle , \rangle$ satisfies only condition 1) and 2), we call $\langle , \rangle$ a symmetric bilinear form.

Remark 2 *Sometimes, the inner product is defined on complex vector spaces by replacing* $\mathbb{R}$ *with* $\mathbb{C}$ *and the condition 1) is* $\langle u, v\rangle = \overline{\langle v, u\rangle}$, *the complex conjugation.*

Example 3 $\langle u, v\rangle = u \circ v$ *is an inner product on* $V = \mathbb{R}^n$.

Example 4 *A matrix* $A_{n\times n}$ *is called symmetric if* $A^T = A$. *The function* $\langle x, y\rangle = x^T Ay$ *is a symmetric bilinear form. If* A *is diagonal with positive diagonal entries, then* $\langle , \rangle$ *is an inner product on* $\mathbb{R}^n$.

Example 5 *Let* $V = M_{m\times n}(\mathbb{R})$ *(the vector space of all* $m \times n$ *real matrices). The function* $\langle x, y\rangle = Trace(x^T y)$ *is an inner product on* V.

Example 6 *Let* $C[a, b]$ *be the set of all continuous functions on the closed interval* $[a, b]$. *Then* $\langle f, g\rangle = \int_a^b fg\mathrm{d}x$ *is an inner product.*

We denote $\|x\| = \sqrt{\langle x, x\rangle} \geq 0$ as the length of $x \in V$. For two vectors $x, y \in V$, the distance $d(x, y) = \|x - y\|$. Two vectors x, y are orthogonal if $\langle x, y\rangle = 0$.

Lemma 7 *(Cauchy-Schwarz inequality) For any $x, y \in V$, we have $\langle x, y\rangle \leq \|x\|\|y\|$. Furthermore, we have $\|x+y\| \leq \|x\| + \|y\|$.*

Proof. For any real number t, we have $0 \leq \langle tx+y, tx+y\rangle = t^2\langle x, x\rangle + 2t\langle x, y\rangle + \langle y, y\rangle$. Therefore, $4\langle x, y\rangle^2 - 4\langle x, x\rangle\langle y, y\rangle \leq 0$ and thus $\langle x, y\rangle \leq \|x\|\|y\|$.

Note that

$$\begin{aligned} \|x+y\|^2 &= \langle x+y, x+y\rangle = \langle x, x\rangle + \langle y, y\rangle + 2\langle x, y\rangle \\ &\leq \langle x, x\rangle + \langle y, y\rangle + 2\|x\|\|y\| = (\|x\| + \|y\|)^2. \end{aligned}$$

■

When $x, y \in \mathbb{R}^n$, the law of cosine gives that

$$\begin{aligned} \|x-y\|^2 &= \|x\|^2 + \|y\|^2 - 2\|x\|\|y\|\cos\phi, \\ \langle x, y\rangle &= \|x\|\|y\|\cos\phi, \end{aligned}$$

where ϕ is the angle between vector x and y. In general inner-product space, if $\langle x, y\rangle = \|x\|\|y\|\cos\phi$, we still view $\phi \in [0, \pi)$ as an angle between x and y. In particular, when $\langle x, y\rangle = 0$, we call x, y are orthogonal. Using these general concepts, we can still talk about orthogonal, orthnormal basis and do Gram-Schmidt orthogonalization process.

Lemma 8 *Let $\{v_1, v_2, ..., v_n\}$ be an orthonormal basis of an inner product space V. We have*

$$\|\sum_{i=1}^n a_i v_i\| = \sum_{i=1}^n |a_i|^2.$$

Definition 9 *Let V be a subspace of an inner product space $(W, \langle,\rangle)$ (i.e. W is a real vector space together with an inner product $\langle,\rangle$). The orthogonal complement $V^\perp = \{x \in W \mid \langle x, y\rangle = 0\}$.*

Lemma 10 *Let A be an $m \times n$ matrix.*

1) For any $x \in \mathbb{R}^m, y \in \mathbb{R}^n$, we have $x \circ Ay = A^T x \circ y$.

2) Then $(\mathrm{Col}A)^\perp = \mathrm{Nul}A^T$, $(\mathrm{Row}A)^\perp = \mathrm{Nul}A$.

Proof. Note that $x \circ Ay = x^T Ay = (A^T x)^T y = (A^T x) \circ y$. 2) follows 1): for any $x \in \mathrm{Nul}A^T$ we have $A^T x = 0$ and thus $x \circ Ay = 0$ for any y, which proves that $x \in (\mathrm{Col}A)^\perp$. On the other hand, for any $x \in (\mathrm{Col}A)^\perp$ we have $x \circ Ay = 0$ for any y. But $A^T x \circ y = 0$ for any y, which implies that $A^T x = 0$ by choosing y in a basis. ■

Lemma 11 *1) The orthgonal complement $V^\perp$ is a vector subspace of W.*

2) $W = V \bigoplus V^\perp$, the direct sum.

3) $(V^\perp)^\perp = V$.

Proof. 1) For any $x, y \in V^\perp$, we have $\langle ax+by, v\rangle = a\langle x, v\rangle + b\langle y, z\rangle = 0$ for any $v \in V$ and arbitrary $a, b \in \mathbb{R}$. This shows that $ax + by \in V^\perp$.

2) Choose a basis B for V and extend this set to be a basis C of W. Apply the Gram-Schmidt orthogonalization process to get an orthogonal basis S of W. Each element $x \in W$ is a linear combination

$$x = \sum_{s \in S} a_s s = \sum_{s \in S \cap V} a_s s + \sum_{s \in S \setminus V} a_s s \in V + V^{\perp}.$$

It is enough to show that $V \cap V^{\perp} = \{0\}$. Actually, any $x \in V \cap V^{\perp}$ has $\langle x, x \rangle = 0$ implying $x = 0$.

3) Since any $v \in V$ is orthogonal to any $x \in V^{\perp}$, we have $V \subset (V^{\perp})^{\perp}$. If there is $x \in (V^{\perp})^{\perp} \setminus V$, we have

$$x = \sum_{s \in S} a_s s = \sum_{s \in S \cap V} a_s s + \sum_{s \in S \setminus V} a_s s,$$

with $\sum_{s \in S \setminus V} a_s s \neq 0$, where S is an orthogonal basis as in 1). However, $\langle x, \sum_{s \in S \setminus V} a_s s \rangle = \langle \sum_{s \in S \setminus V} a_s s, \sum_{s \in S \setminus V} a_s s \rangle > 0$, a contradiction to the fact that x is orthogonal to $V^{\perp}$. ■

2 Inner products and matrices

For a complex matrix $A_{n \times m}$, its conjugate transpose is the $m \times n$ matrix $A^* = (\bar{a}_{ji})$, where $\bar{a}_{ji} = a - b\mathbf{i}$ (complex onjugate) if $a_{ji} = a + b\mathbf{i}, a, b \in \mathbb{R}$. A square complex matrix A is called Hermitian (or self-adjoint) if $A = A^*$. Note that real Hermitian matrix is symmetric.

Lemma 12 *Let $(V, \langle , \rangle)$ be an inner product space of dimension n. There is a Hermitian matrix $A_{n \times n}$ such that $\langle x, y \rangle = x^T A y$ (or $\langle x, y \rangle = x^* A y$ when the ground field is $\mathbb{C}$) for any $x, y \in V$.*

Proof. Choose a basis $\{e_1, e_2, ..., e_n\}$. Let $A = (\langle e_i, e_j \rangle)_{1 \leq i,j \leq n}$. For any $x = \sum x_i e_i, y = \sum y_i e_i$, we have $\langle x, y \rangle = \sum x_i y_j \langle e_i, e_j \rangle = x^T A y$ (or $\langle x, y \rangle = \sum \bar{x}_i y_j \langle e_i, e_j \rangle = x^* A y$). By the definition of inner products, we have $\langle e_i, e_j \rangle = \overline{\langle e_j, e_i \rangle}$. ■

In the above lemma, we actually assume that x is the same as its coordinate vector with respect to the basis. We call the matrix A the representation matrix of the inner product with respect to the basis $\{e_1, e_2, ..., e_n\}$.

Lemma 13 *Let $(V = \mathbb{F}^n, \langle , \rangle)$ be an inner product space for $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, with a representation matrix A (with respect to the standard basis). A set $\{v_1, v_2, ..., v_n\}$ is an orthonormal basis if and only $[v_1, v_2, ..., v_n]^* A [v_1, v_2, ..., v_n] = A$.*

When the inner product on $\mathbb{C}^n$ is the standard one (i.e. $A = I_n$), we have that a set $\{v_1, v_2, ..., v_n\}$ is an orthonormal basis if and only if $[v_1, v_2, ..., v_n]^* [v_1, v_2, ..., v_n] = I_n$. We call a square complex matrix B unitary if $B^* B = I_n$. Note that a real unitary matrix is orthogonal.

Theorem 14 *Let $(V, \langle , \rangle)$ be a complex inner product space of dimension n. For any complex $n \times n$ matrix A, there is an orthogonal basis $\{v_1, v_2, ..., v_n\}$ of V such that the representation matrix of A is upper triangular.*

Proof. By the Jordan canonical theorem, there is an invertible matrix P and an upper triangular matrix U such that $A = PUP^{-1}$. Apply the Gram-Schmidt orthogonalization to get a QR-decomposition $P = QR$. Then $A = QRUR^{-1}Q^{-1}$. Note that RUR^{-1} is upper triangular and the columns of Q are orthogonal. ■

Corollary 15 *(Schur's theorem) For any matrix $A_{n\times n}$, there is a unitary matrix P such that $PAP^{-1} = PAP^*$ is upper triangular. In other words, any square complex matrix is conjugate to an upper triangular matrix by a unitary matrix.*

Proof. Consider the standard inner product $\langle x, y\rangle = x^*y$ on $\mathbb{C}^n$ and apply the previous theorem. ■

Lemma 16 *(Riesz representation theorem) Let $(V, \langle , \rangle)$ be an inner product space over the field $F = \mathbb{R}$ or $\mathbb{C}$. Suppose that $f : V \to F$ is a linear map (usually called linear functional). There exists a unique $y \in V$ such that $f(x) = \langle x, y\rangle$ for any $x \in V$.*

Proof. Existence. If $f = 0$, we just choose $y = 0$. Otherwise, the complement $(\ker f)^{\perp}$ is of dimension one. Choose v to be a unit vector of $(\ker f)^{\perp}$ and let $y = f(v)v$. For any $x \in V$, we have $x = x_1 + a_1v$ for some $x_1 \in \ker f$ and $a_1 \in F$. Therefore, $f(x) = a_1f(v) = \langle x, f(v)v\rangle$.

Uniqueness. If there are two vectors y_1, y_2 both satisfying $\langle x, y_1\rangle = \langle x, y_2\rangle$ for any $x \in V$. Then $\langle x, y_1 - y_2\rangle = 0$, imply $y_1 = y_2$ by choosing $x = y_1 - y_2$. ■

Corollary 17 *Let $f : V \to W$ be a linear map between two inner product spaces $(V, \langle , \rangle_V), (W, \langle , \rangle_W)$. There exits a unique linear map $f^* : W \to V$ such that*

$$\langle f(x), y\rangle_W = \langle x, f^*(y)\rangle_V$$

for any $x \in V, y \in W$. The function f^* is called the adjoint of f.

Proof. Existence. Fix any $y \in W$, the function $\langle f(-), y\rangle_W : V \to F$ is a linear functional. The Riesz representation theorem implies that there is a unique element $z \in V$ satisfying $\langle f(-), y\rangle_W = \langle x, z\rangle_V$. Define $f^*(y) = z$.

Uniqueness. If there is another g^* satisfying $\langle f(x), y\rangle_W = \langle x, f^*(y)\rangle_V = \langle x, g^*(y)\rangle_V$ for any x, y, we must have $\langle x, f^*(y) - g^*(y)\rangle_V = 0$ which implies that $f^*(y) = g^*(y)$. ■

Example 18 *Consider the dot product on $\mathbb{R}^n$ and the standard inner product on $\mathbb{C}^n$. We have the adjoint of a real matrix A is its transpose A^T and the adjoint of a complex matrix A is its conjugate transpose. (hint: in this case, $\langle Ax, y\rangle = x^*A^*y = \langle x, A^*y\rangle$.)*

The following is general version of Lemma 10.

Lemma 19 *Let $f : V \to W$ be a linear map between two inner product spaces. We have the following:*
1) $(f^*)^* = f$.
2) $(\ker f)^\perp = \operatorname{Im} f^*$.

3 Inner product spaces and isometries

Definition 20 *A linear map $f : V \to W$ between inner product spaces is distance-preserving (or isometric) if $\| f(x) \| = \| x \|$ for any $x \in V$.*

Lemma 21 *A linear map $f : V \to W$ is distance-preserving if and only if $\langle f(x), f(y) \rangle = \langle x, y \rangle$ for any $x, y \in V$.*

Lemma 22 *Let V be a vector space together with an inner product defined by $\langle x, y \rangle = x^* P y$ for a matrix P. A linear map $f : V \to V$ is isometric if and only if $A^* P A = P$, where A is the standard representation matrix of f.*

Proof. Note that $\langle f(x), f(y) \rangle = \langle x, y \rangle$ if and only $(Ax)^* P A y = x^* A^* P A y = x^* P y$. Choose $x, y \in \{e_1, e_2, ..., e_n\}$, the standard basis. ■

Corollary 23 *A complex matrix $A_{n\times n}$ preserves the standard distance on $\mathbb{C}^n$ if and only if $A^* A = I_n$, i.e. A is unitary.*

4 Inner products and norms

Definition 24 *A normed vector space is a vector space V (over $\mathbb{R}$ or $\mathbb{C}$) together with a function (called a norm): $\| \, \| : V \to \mathbb{R}$ sastisfying*
1) Homogeneity: $\| \alpha v \| = |\alpha| \| v \|$ for any for all vectors v and all scalars α;
2) Triangle inequality: $\| x + y \| \le \| x \| + \| y \|$ for any $x, y \in V$;
3) positivity: $\| x \| \ge 0$ for any vector x, and $\| x \| = 0$ if and only if $x = 0$.

It is obvious that an inner product $\langle , \rangle$ gives a norm $\| x \| = \sqrt{\langle x, x \rangle}$. But not every norm is from an inner product.

Example 25 *Let $V = \mathbb{R}^n$ or $\mathbb{C}^n$. For any $x \in V$, define $\| x \|_p = (|x_1|^p + |x_2|^p + \cdots + |x_n|^p)^{1/p}$ for $1 \le p < \infty$ and $\| x \|_\infty = \max\{|x_i| : i = 1, 2, ..., n\}$.*

Example 26 *Let $V = C[0, 1]$ be the vector space of continuous functions on the closed interval $[0, 1]$. Define $\| f \|_p = (\int_0^1 |f|^p dx)^{1/p}$ for $1 \le p < \infty$.*

Theorem 27 *A norm in a normed space is obtained from some inner product if and only if it satisfies the Parallelogram Identity*

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

for any $x, y \in V$.

Proof. When the norm comes from an inner product, $\|x+y\|^2 + \|x-y\|^2 = \langle x+y, x+y\rangle + \langle x-y, x-y\rangle = 2(\|x\|^2 + \|y\|^2)$.

Conversely, for real vector spaces we define $\langle x,y\rangle = \frac{1}{4}(\|x+y\|^2 - \|x-y\|^2)$ (called Polarization identities). We check the conditions of an inner product. It is obvious that $\langle x,y\rangle = \langle y,x\rangle$, and $\langle x,x\rangle \geq 0, \langle x,x\rangle = 0$ if and only if $x=0$. It is enough to prove that $\langle x,y\rangle$ is bilinear. By the parallelogram law we have

$$2\|x+z\|^2 + 2\|y\|^2 = \|x+y+z\|^2 + \|x-y+z\|^2.$$

Therefore,

$$\begin{aligned}
\|x+y+z\|^2 &= 2\|x+z\|^2 + 2\|y\|^2 - \|x-y+z\|^2 \\
&= 2\|y+z\|^2 + 2\|x\|^2 - \|y-x+z\|^2 \\
\|x+y+z\|^2 &= \|x\|^2 + \|y\|^2 + \|x+z\|^2 + \|y+z\|^2 - \frac{1}{2}\|x-y+z\|^2 - \frac{1}{2}\|y-x+z\|^2. \\
\|x+y-z\|^2 &= \|x\|^2 + \|y\|^2 + \|x-z\|^2 + \|y-z\|^2 - \frac{1}{2}\|x-y-z\|^2 - \frac{1}{2}\|y-x-z\|^2.
\end{aligned}$$

$$\begin{aligned}
\langle x+y, z\rangle &= \frac{1}{4}\left(\|x+y+z\|^2 - \|x+y-z\|^2\right) \\
&= \frac{1}{4}\left(\|x+z\|^2 - \|x-z\|^2\right) + \frac{1}{4}\left(\|y+z\|^2 - \|y-z\|^2\right) \\
&= \langle x,z\rangle + \langle y,z\rangle
\end{aligned}$$

Inductively, we have $\langle nx,z\rangle = n\langle x,z\rangle$ for each integer n. Similarly, we have $\langle x,z\rangle = \langle n\frac{1}{n}x,z\rangle = n\langle\frac{1}{n}x,z\rangle$ and $\langle\frac{1}{n}x,z\rangle = \frac{1}{n}\langle x,z\rangle$ for each nonzero n. This actually means for any rational number $q=\frac{m}{n}$ we have $\langle qx,z\rangle = \langle\frac{m}{n}x,z\rangle = q\langle x,z\rangle$. Note that $t \to \frac{1}{t}\langle tx,z\rangle \in \mathbb{R}$ is continuous on $\mathbb{R}\backslash\{0\}$. Since every real number is a limit of a rational sequence, we have that $\langle rx,z\rangle = r\langle x,z\rangle$ for every rational number r.

For complex vector spaces, define

$$\begin{aligned}
\langle x,y\rangle &= \frac{1}{4}(\|x+y\|^2 + \mathbf{i}\|\mathbf{i}x+y\|^2 - \|-x+y\|^2 - \mathbf{i}\|-\mathbf{i}x+y\|^2) \\
&= \frac{1}{4}\sum_{k=0}^{3}\mathbf{i}^k\|\mathbf{i}^k x+y\|^2 .
\end{aligned}$$

It's obvious that $\langle x,y\rangle = \overline{\langle y,x\rangle}$. A similar argument proves the bilinear property of the real and imaginary parts. ∎

Corollary 28 *For $p \neq 1$, the norm $\|-\|_p$ does not come from an inner product since the Parallelogram identity does not hold. Let e_1, e_2 be elements of the standard basis. We have*

$$2(\|e_1\|_p^2 + \|e_2\|_p^2) = 2 \neq \|e_1+e_2\|_p^2 + \|e_1-e_2\|_p^2 = 2^{2/p} + 2^{2/p}.$$

Lecture 3: Symmetric matrices and quadratic forms

Shengkui Ye

March 20, 2023

1 Symmetric matrices

A square matrix A is symmetric if $A = A^T$. For example, $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

Lemma 1.1 *If A is symmetric, then any two eigenvectors corresponding to distinct eigenvalues are orthogonal. In other words, if $Ax_1 = \lambda_1 x_1$ and $Ax_2 = \lambda_2 x_2$ with $\lambda_1 \neq \lambda_2$ then $x_1 \circ x_2 = 0$.*

Proof. Note that $x_2^T \lambda_1 x_1 = x_2^T (Ax_1) = x_2^T A^T x_1 = (Ax_2)^T x_1 = \lambda_2 x_2^T x_1$, which implies $x_2^T x_1 = 0$ since $\lambda_1 \neq \lambda_2$. ■

An $n \times n$ matrix A is said to be orthogonal diagonalizable if there is an orthogonal matrix P (i.e. $P^{-1} = P^T$) such that $P^{-1}AP$ is diagonal.

Lemma 1.2 *An $n \times n$ symmetric (real) matrix A has n real eigenvalues, counting multiplicities. For each eigenvalue λ, there is a real eigenvector x corresponding to it.*

Proof. Suppose that $Ax = \lambda x$ for a complex value λ and a complex vector x. Let x^* be the complex conjugate transpose. Then $x^*Ax = x^*\lambda x = \lambda \|x\|^2$, but $x^*Ax = (Ax)^*x = (\lambda x)^*x = \lambda^* x^* x$. This implies $\lambda = \lambda^*$ and thus λ is real. The Fundamental Theorem of Algebra proves that A has n eigenvalues and thus the symmetric matrix A has n real eigenvalues. Since $A - \lambda I$ has determinant zero, $(A - \lambda I)x = 0$ has a nonzero solution in $\mathbb{R}^n$. ■

Theorem 1.3 *(Spectral theorem) An $n \times n$ matrix A is orthogonal diagonalizable if and only if A is symmetric.*

Proof. If there exists orthogonal matrix P and diagonal matrix D such that $P^{-1}AP = D$, then $A = PDP^{-1} = PDP^T$ is symmetric.

The othe direction can be proved by induction. When $n = 1$, there is nothing to prove. Suppose the statement is true for $n-1$. Let λ be a real eigenvalue of A, with a unit real eigenvector vector x (the existence follows the previous lemma). Extend x to be a basis B of $\mathbb{R}^n$ and apply the Gram-Schmidt process to get an orthonormal basis $B = \{x_1 = x, x_2, x_3, \cdots, x_n\}$. Let $P_1 = [x_1, x_2, ..., x_n]$ and $C = P_1^{-1}AP_1$. Note that the first column of C is $[\lambda, 0, 0, \cdots, 0]^T$. Moreover C is symmetric, since P_1 is orthogonal. Therefore, the first row of C is $[\lambda, 0, 0, \cdots, 0]$. Write

$$C = \begin{bmatrix} \lambda & 0 \\ 0 & C_1 \end{bmatrix}$$

for a symmetric matrix C_1. The induction step implies that there exists orthogonal matrix P_2 such that $P_2^{-1}C_1P_2$ is diagonal. Therefore, we take $P = P_1 \begin{bmatrix} 1 & 0 \\ 0 & P_2 \end{bmatrix}$ such that $P^{-1}AP$ is diagonal. ■

Example 1.4 *Let* $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$. *Find the orthogonal diagonalization if exits.*

When A is symmetric, there is an orthogonal matrix P such that $P^{-1}AP = D$, a diagonal matrix. Suppose that $P = [u_1, u_2, \cdots, u_n]$. Then $AP = PD$ and thus $[Au_1, Au_2, \cdots, Au_n] = [d_1u_1, d_2u_2, \cdots, d_nu_n]$ where d_i is the i-th diagonal entry of D. Since $Au_i = d_iu_i$ for each i, we know that d_i is an eigenvalue and u_i is the corresponding eigenvector. Moreover, $A = PDP^{-1} = PDP^T = [d_1u_1, d_2u_2, \cdots, d_nu_n][u_1, u_2, \cdots, u_n]^T = d_1u_1u_1^T + d_2u_2u_2^T + \cdots + d_nu_nu_n^T$. This sum is called the spectral decomposition of A.

2 Applications: Quadratic forms

Definition 2.1 *A quadratic form Q is function defined on $\mathbb{R}^n$ such that $Q(x) = x^TAx$ for a symmetric matrix A. In other words, $Q(x)$ is a degree-two homogenous polynomial.*

Example 2.2 $Q(x) = 3x_1^2 + 4x_2^2 = [x_1, x_2] \begin{bmatrix} 3 & \\ & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ *is a quadratic form.*

Example 2.3 *Write $Q(x) = x_1x_2 + x_2^2$ as the form x^TAx for some symmetric matrix A.*

Example 2.4 *Let $Q(x) = x^TAx$ be a quadratic form. For an invertible matrix P, let $y = P^{-1}x$. Then $x = Py$ and $Q(x) = y^TP^TAPy$ is another quadratic form of y, which is called a change of variable.*

For a general degree-two homogenous polynomial $Q(x) = \sum_{i,j=1}^n a_{ij}x_ix_j$, is there a canonical form after change of variables? If there is such one, how to reduce $Q(x)$ to the canonical form?

Lemma 2.5 *Any quadratic form $Q(x) = x^TAx$ could be transformed to the diagonal form. In other words, there exists an orthogonal matrix P such that $x = Py$ and*

$$Q(x) = y^T(P^TAP)y = a_1y_1^2 + a_2y_2^2 + \cdots + a_ny_n^2$$

for some real numbers $a_1, a_2, \cdots, a_n$.

Proof. It is enough to note that $P^T AP$ could be diagonal for some orthogonal matrix P. ■

Example 2.6 *Let $Q(x) = x_1^2 - 8x_1x_2 - 5x_2^2$. Reduce $Q(x)$ to be the canonical form by change of variables.*

Definition 2.7 *A quadratic form $Q(x) = x^T Ax$ (or the coefficient matrix A) is*

a) positive definite if $Q(x) > 0$ for any $x \neq 0$;
b) negative definite if $Q(x) < 0$ for any $x \neq 0$;
c) indefinite if $Q(x)$ assumes both positive and negative values.
d) positive semi-definite if $Q(x) \geq 0$ for any x.

Example 2.8 *Suppose that $Q(x) = x^T Ax$ for a symmetric matrix A. If all eigenvalues of A are positive, then Q is positive definite. Similarly, if all the eigenvalues are negative, then Q is negative definite.*

Corollary 2.9 *A symmetric matrix A is positive semi-definite (resp. definite) if and only if $A = R^T R$ for a (resp. invertible) matrix R.*

Proof. For any x, we have $x^T Ax = x^T R^T Rx = \langle Rx, Rx\rangle \geq 0$. When R is invertible, $\langle Rx, Rx\rangle = 0$ if and only $x = 0$. ■

Lemma 2.10 *Let $A_{n\times n}$ be a positive definite matrix. Define $\langle x, y\rangle := x^T Ay$. Then $\langle x, y\rangle$ is an inner product on $\mathbb{R}^n$.*

Proof. It's easy to check that $\langle x, y\rangle$ is symmetric (as A is symmetric) and bilinear. When A is positive definite, $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0$ if and only $x = 0$. ■

3 Applications: Quadratic curves

In high school, we already studied three kinds of curves: ellipse, hyperbola, parabola. These curves are defined by two-variable degree-two polynomials. It turns out that these are the only three cases (in a genuine sense).

Definition 3.1 *A quadratic curves is a plane curve in $\mathbb{R}^2$ defined by a degree-two two-variable polynomial*

$$ax^2 + bxy + cy^2 + dx + ey + f = 0, \tag{1}$$

where $a, b, c, d, e, f \in \mathbb{R}$.

Theorem 3.2 *Any quadratic curve is one of the following:*

1) ellipse; 2) hyperbola; 3) parabola; 4) intersecting lines; 5) parallel lines, or 6) a single point.

Proof. Write

$$\begin{aligned} & ax^2 + bxy + cy^2 + dx + ey + f \\ = \quad & (x, y) \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} d & \\ & e \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + f. \end{aligned}$$

Since $\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}$ is symmetric, there is an orthogonal matrix P such that

$$\begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix} = P^T \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} P$$

for some real numbers d_1, d_2. Change the variables by letting $\begin{bmatrix} x' \\ y' \end{bmatrix} = P \begin{bmatrix} x \\ y \end{bmatrix}$. The equation (1) becames

$$d_1 x'^2 + d_2 x'^2 + d'x' + e'y' + f' = 0. \tag{2}$$

Since the polynomial is still of degree 2, we may assume that $d_1 \neq 0$. If $d_2 \neq 0$, the previous equation (2) can be written as

$$d_1(x' + a_1)^2 + d_2(y' + a_2)^2 + f'' = 0$$

for some real coefficients. Change the variables again by letting $x' + a_1 = x'', y' + a_2 = y''$. We have

$$d_1 x''^2 + d_2 y''^2 = g \tag{3}$$

for some real numbers d_1, d_2, g. After exchanging x'', y'' and the sign of d_1, we can assume that $d_1 > 0$.

Case 1) $d_2 > 0$. If $g > 0$, the equation (3) gives an ellipse. If $g = 0$, the equation (3) gives a point. If $g < 0$, the equation (**) does not have real solutions (or imaginary ellipse).

Case 2) $d_2 < 0$. If $g \neq 0$, the equation (3) gives a hyperbola. If $g = 0$, the equation (3) gives intersecting of two lines.

Case 3) $d_2 = 0$. The equation (2) can be written as

$$d_1(x' + a_1)^2 + e'y' + f'' = 0. \tag{4}$$

If $e' \neq 0$, we have $d_1 x''^2 + e'y'' = 0$, for some $x'' = x' + a', y'' = x'' + b''$, which gives a parabola. Suppose that $e' = 0$. If $f'' < 0$, the equation (4) gives a pair of parrell lines. If $f'' > 0$, the equation (4) has no real solutions (or a imaginary circle). If $f'' = 0$, the equation (4) actually is a single point.

■

Remark 3.3 *Ellipse, hyperbola and parabola are called non-degenerate quadratic curves, while the intersecting curves, parallel lines, and a single point are called degenerated quadratic curves.*

Example 3.4 *Determine the type of the quadratic curve* $x^2+xy+y^2+x+1 = 0$.

4 Applications: extreme values and singular values

Theorem 4.1 *Let A be a symmetric matrix with an orthogonal diagonalization $A = PDP^{-1}$, with the diagonal entries of D arranged as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, and P is an orthogonal matrix. Then*

$$\lambda_1 = \max_{\|x\|=1} x^T Ax, \lambda_n = \min_{\|x\|=1} x^T Ax,$$

with the extreme values are achived when x are the corresponding eigenvectors.

Proof. Let $y = (y_1, \cdots, y_n)^T = P^T x$. When $\|x\| = 1$, we have $\|y\| = 1$. Note that

$$\begin{aligned} x^T Ax &= x^T PDP^{-1}x^T = (P^T x)^T DP^T x = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \\ &\leq \lambda_1(y_1^2 + y_2^2 + \cdots + y_n^2) = \lambda_1. \end{aligned}$$

The maximum is achived when $y = (1, 0, \cdots, 0)^T$ and $x = Py$, an eigenvalue corresponding to λ_1. Similarly, $x^T Ax \geq \lambda_n(y_1^2 + y_2^2 + \cdots + y_n^2)$, with the mimumn is achived when $y = (0, \cdots, 0, 1)^T$ and $x = Py$, an eigenvalue corresponding to λ_n. ■

Definition 4.2 *Let $A_{m\times n}$ be a matrix. A singular value σ_i of A is the square root of an eigenvalue λ_i of $A^T A$, i.e. $\sigma_i = \sqrt{\lambda_i(A^T A)}$.*

Note that $A^T A$ is symmetric and positive semi-definite. There is an orthogonal diagonalization $A^T A = PDP^{-1}$. Let P_i be a column of P, ie. an eigenvector. Then $P_i^T A^T AP_i = \lambda_i P_i^T P_i$, which implies that $\|AP_i\| = \sigma_i$. View A as a linear map $\mathbb{R}^n \to \mathbb{R}^m$, with $\{P_1, ..., P_n\}$ an orthonormal basis of $\mathbb{R}^n$. The singular value σ_i is the length $\|AP_i\|$.

Lemma 4.3 *Suppose that the eigenvalues of $A^T A$ are $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_k > \lambda_{k+1} = \lambda_{k+2} = \cdots = 0$, with corresponding eigenvectors $v_1, v_2, ..., v_n$. Then $\{Av_1, Av_2, \cdots, Av_k\}$ is an orthogonal basis of $Col(A)$.*

Proof. Note that $v_1, v_2, ..., v_n$ form an orthogonal basis of R^n. This means $Col(A)$ is spanned by $\{Av_1, Av_2, ..., Av_n\}$. But $Av_{k+1} = 0 = Av_{k+2} = \cdots = Av_n$. Moreover, $Av_i \circ Av_j = v_i^T A^T Av_j = 0, Av_i \circ Av_i = \lambda_i \|v_i\|^2$ for any $i \neq j \leq k$. Therefore, $\{Av_1, Av_2, \cdots, Av_k\}$ is an orthogonal basis. ■

Theorem 4.4 *(singular value decomposition) Let $A_{m\times n}$ be a matrix of rank r. There exist a diagonal matrix $D_{r\times r}$ (with diagonal entries the singular values of A) and orthogonal matrices $U_{m\times m}, V_{n\times n}$ such that*

$$A = U \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}_{m\times n} V^T.$$

Proof. As in the previous lemma, let λ_i and v_i be the eigenvalues and eigenvectors (with $\|v_i\| = 1$) of $A^T A$. Let $u_i = \frac{Av_i}{\|Av_i\|}, i \leq r$. Extend $\{u_1, u_2, ..., u_r\}$ to be an orthonormal basis $\{u_1, u_2, ..., u_r, u_{r+1}, ..., u_m\}$ of $\mathbb{R}^m$. Take $U = [u_1, u_2, \cdots, u_m]$ and $V = [v_1, v_2, \cdots, v_n]$. It can be directly checked that

$$\begin{aligned} A[v_1, v_2, \cdots, v_n] &= [Av_1, Av_2, \cdots, Av_n] \\ &= [\sigma_1 u_1, \sigma_2 u_2, \cdots, \sigma_r u_r, 0, \cdots, 0] \\ &= U \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

The result is proved by noting that $V^{-1} = V^T$. ■

Example 4.5 *Find the singular value decomposition (SVD) of* $A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

Example 4.6 *Let* $A = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \end{bmatrix}$, *viewed as a linear map* $\mathbb{R}^3 \to \mathbb{R}^2$. *Find a unit vector* $v \in \mathbb{R}^3$ *such that* $\|Av\|$ *is the maximum.*

Lecture 4: Symmetric matrices and quadratic forms

Shengkui Ye

April 17, 2023

1 Self-adjoint operators

Recall that a self-adjoint operator is a linear map $f : V \to V$ on an inner product space satisfying $f = f^*$, i.e. $\langle f(x), y\rangle = \langle x, f(y)\rangle$ for any $x, y \in V$. This is a generalization of a symmetric matrix. Many properties on symmetric matrices are still true for self-adjoint operators.

Lemma 1.1 *Let $f = f^*$ be a self-adjoint operator in an inner product space V. We have the following:*

1) all eigenvalues of f are real;

2) eigenvectors from distinct eigenvalues are orthogonal;

Proof. Suppose that the inner product is represented by A and B is the standard matrix of f. We have $B^*A = AB$. Suppose that $Bx = \lambda x$ for a complex value λ and a complex vector x. Then $x^*ABx = x^*A\lambda x = \lambda\|x\|^2$, but $x^*ABx = (Bx)^*Ax = (\lambda x)^*Ax = \lambda^* x^*Ax$. This implies $\lambda = \lambda^*$ and thus λ is real.

Suppose that $Bx_1 = \lambda_1 x_1, Bx_2 = \lambda_2 x_2$ for $\lambda_1 \neq \lambda_2$ and eigenvectors x_1, x_2. We have $\langle x_2, Bx_1\rangle = \langle x_2, \lambda_1 x_1\rangle = \langle Bx_2, x_1\rangle = \lambda_2\langle x_2, x_1\rangle$, which implies $x_2^T x_1 = 0$ since $\lambda_1 \neq \lambda_2$. ■

Theorem 1.2 *(spectral theorem)Let $f = f^*$ be a self-adjoint operator on a finite-dimensional inner product space V over a field $F = \mathbb{R}$ or $\mathbb{C}$. There exists an orthonormal basis on which the representation matrix of f is a real diagonal matrix. In particular, for any Hermitian (or self-adjoint) matrix A, there exist a unitary matrix U and a real diagonal matrix D such that $A = UDU^*$.*

Proof. Consider the characteristic polynomial of f. Over the complex numbers $\mathbb{C}$, there is an eigenvalue λ, which is actually real since f is self-adjoint. Choose a unit eigenvector v_1, i.e. $f(v_1) = \lambda_1 v_1$. The orthogonal complement $(Fv_1)^\perp$ is invariant under the transformation by f ($\forall x \in (Fv_1)^\perp$, we have $\langle v_1, fx\rangle = \langle f^*v_1, x\rangle = \langle fv_1, x\rangle = \lambda_1\langle v_1, x\rangle = 0$). We repeat the argument to choose another eigenvector $v_2 \in (Fv_1)^\perp$. After finitely many steps, we get an orthogonal basis $\{v_1, ..., v_n\}$ on which the representation matrix of f is real diagonal.

The complex case can be proved as following. Schur's theorem implies that there is an orthogonal basis on which the representation matrix of f is an upper triangular matrix, i.e. $f = URU^{-1}$ for an upper triangular matrix (here we denote f as its standard representation matrix). Suppose that the inner product is represented by a matrix A. Note that a self-adjoint upper triangular matrix must be diagonal with real entries. Actually, we have $(fx)^*Ay = x^*Afy$ and $f^*A = Af, (U^{-1})^*R^*U^*A = AURU^{-1}, R^*U^*AU = U^*AUR$, (noting that $U^*AU = I_n$), implying $R^* = R$ and R must be diagonal. ■

Recall that a square real matrix A is orthogonal diagonalizable if and only if A is symmetric. Can we have a similar result for unitary matrices? We already know that a self-adjoint matrix is diagonalizable by a unitary matrix. It turns out that the converse is not true.

Definition 1.3 *A linear map (or matrix) $N : V \to V$ on an inner product space V is normal, if $NN^* = N^*N$.*

Example 1.4 *A self-adjoint matrix is normal. An orthgonal (or unitary) matrix is normal. A unitary diagonalizable matrix is normal. Unitary conjugates of a normal matrix is normal.*

Lemma 1.5 *A linear map (or matrix) $N : V \to V$ is normal if and only if*

$$\| Nx \| = \| N^*x \|, \text{ for any } x \in V.$$

Proof. If N is normal, we have $\| Nx \|^2 = \langle Nx, Nx\rangle = \langle x, N^*Nx\rangle = \langle x, NN^*x\rangle = \langle N^*x, N^*x\rangle = \| N^*x \|^2$ for any x.

Conversely, the Polarization Identities imply for any $x, y \in V$ that

$$\begin{aligned} \langle N^*Nx, y\rangle &= \langle Nx, Ny\rangle = \frac{1}{4}\sum_{k=0}^{3} \mathbf{i}^k \| Nx + \mathbf{i}^k Ny \| \\ &= \frac{1}{4}\sum_{k=0}^{3} \mathbf{i}^k \| N(x + \mathbf{i}^k y) \| \\ &= \frac{1}{4}\sum_{k=0}^{3} \mathbf{i}^k \| N^*(x + \mathbf{i}^k y) \| \\ &= \langle N^*x, N^*y\rangle = \langle NN^*x, y\rangle \end{aligned}$$

and thus $N^*N = NN^*$. ■

Theorem 1.6 *Any normal linear map in a* complex *vector space has an orthonormal basis consisting of eigenvectors. In particular, a complex matrix is unitary diagonalizable if and only if it is normal.*

Proof. Schur's theorem implies that there is an orthogonal basis on which the representation matrix of f is an upper triangular matrix A. It is enough to prove that an upper triangular normal matrix must be diagonal. Suppose that

$$A = \begin{bmatrix} a_{11} & * \\ 0 & A' \end{bmatrix}.$$

Since $AA^* = A^*A$, the $(1,1)$-th entries are $\bar{a}_{11}a_{11} = a_{11}\bar{a}_{11} + a_{12}\bar{a}_{12} + \cdots + a_{1n}\bar{a}_{1n}$. This gives that $a_{12} = a_{13} = ... = a_{1n} = 0$. Repeat this argument to prove that A is diagonal.

We already know that a unitary diagonalizable matrix is normal. The converse is proved by choosing the standard inner product on $\mathbb{C}^n$. ■

2 Polar and singular decomposition

Definition 2.1 *A self-adjoint linear map $f : V \to V$ on an inner product space V is called positive definite if*

$$\langle fx, x\rangle > 0, \forall x \neq 0.$$

Similarly, f is called positive semi-definite if

$$\langle fx, x\rangle \geq 0, \forall x \in V.$$

Example 2.2 *For any complex matrix $B_{m\times n}$, the product B^*B is positive semi-definite, since $\langle B^*Bx, x\rangle = \langle Bx, Bx\rangle \geq 0$ for any $x \in \mathbb{C}^n$.*

Theorem 2.3 *For a self-adjoint linear map f, we have the following.*

1) f is positive definite if and only if the eigenvalues of f are positive.

2) f is positive semi-definite if and only if the eigenvalues of f are non-negative.

Proof. By Lemma 1.2, there is an orthonormal basis on which the representation matrix of f is diagonal. A diagonal matrix is positive definite if and only if the diagonal entries are positive. ■

Remark 2.4 *It is interesting to note that the positive definiteness of a self-adjoint linear map f depends only on its eigenvalues, independent of the basis and the inner product.*

Corollary 2.5 *Let A be a positive semidefinite operator. There exists a unique positive semi-definite operator B such that $A = B^2$. We denote $B = A^{\frac{1}{2}} = \sqrt{A}$.*

Proof. Existence. There is a basis S on which A is diagonal with positive diagonal entries $\lambda_1 \geq \lambda_2 \geq ...\lambda_n \geq 0$. Define B as the the linear map whose representation matrix on the basis is $\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq ...\sqrt{\lambda_n} \geq 0$.

Uniqueness. Suppose that $A = C^2$ for a self-adjoint positive semi-definite matrix C. Choose an orthogonal basis S' on which C is diagonal with diagonal entries $\mu_1 \geq \mu_2 \geq ...\mu_n \geq 0$. Then A has eigenvalues $\mu_1^2 \geq \mu_2^2 \geq ...\mu_n^2 \geq 0$. Moreover, $Ax = \lambda x$ if and only if $Cx = \sqrt{\lambda}x$. Therefore, $Bx = \sqrt{\lambda}x$ for any eigenvector x of A. This implies $B = C$. ■

Lemma 2.6 *For any linear map $A : V \to V$ on an inner product space V. We have*

$$\|\sqrt{A^*A}x\| = \|Ax\|, \forall x \in V.$$

Proof. $\|\sqrt{A^*A}x\|^2 = \langle\sqrt{A^*A}x, \sqrt{A^*A}x\rangle = \langle x, A^*Ax\rangle = \langle Ax, Ax\rangle = \|Ax\|^2$. ■

Theorem 2.7 *(Polar decomposition) For any linear map $A : V \to V$ on an inner product space V. There is an unitary operator U such that*

$$A = U\sqrt{A^*A}.$$

Proof. By the previous lemma, we have $\ker A = \ker\sqrt{A^*A} = \operatorname{Im}(\sqrt{A^*A})^{*\perp} = \operatorname{Im}(\sqrt{A^*A})^{\perp}$ since $\sqrt{A^*A}$ is self-adjoint. We will define U explicitly by specifying its image on $\operatorname{Im}(\sqrt{A^*A})\bigoplus\ker A = V$. For any $x \in \operatorname{Im}(\sqrt{A^*A})$, choose $y \in V$ such that $\sqrt{A^*A}y = x$. Define $U_1 : \operatorname{Im}(\sqrt{A^*A}) \to \operatorname{Im} A$ by $Ux = Ay$. If another y' has $\sqrt{A^*A}y' = x$, we have $\sqrt{A^*A}(y - y') = 0$ and $y - y' \in \ker A$. This checks that U is well-defined on $\operatorname{Im}(\sqrt{A^*A})$. Note that $\operatorname{Im} A = (\ker A^*)^{\perp}$. Since the subspace $\ker A$ is isomorphic to $\ker A^*$ (by the rank theorem), we can choose an isometry $U_2 : \ker A \to \ker A^* = (\operatorname{Im} A)^{\perp}$. It can be directly checked that $U = U_1 \oplus U_2$ is unitary and $A = U\sqrt{A^*A}$. ■

The following is a general singular value decomposition.

Theorem 2.8 *For any linear map $A : V_1 \to V_2$ between inner product spaces V_1, V_2. There exists orthonormal base $\{v_1, v_2, ..., v_m\}$ for V_1 and $\{w_1, w_2, ..., w_n\}$ for V_2, such that the representation matrix A is diagonal with diagonal entries the singular values of A. In other words,*

$$A = [w_1, ..., w_n]D[v_1, v_2, ..., w_m].$$

3 Matrix norms

Let $A_{n\times m} : \mathbb{C}^m \to \mathbb{C}^n$ be a complex matrix.

Definition 3.1 *The real number* $\sup\{\|Ax\| : \|x\| \le 1\}$ *is called the operator norm of A and denoted as $\|A\|$.*

Theorem 3.2 *Let $M_{n\times m}(\mathbb{C})$ be the vector space of all $n \times m$ matrices. We have the following.*

Lemma 3.3 *1) $(M_{n\times m}(\mathbb{C}), \|-\|)$ is a normed space;*

2) $\|Ax\| \le \|A\|\|x\|$ for any $x \in \mathbb{C}^m$;

3) $\|AB\| \le \|A\|\|B\|$ if AB can be defined;

3) $\|A\| = s_1 \le \|A\|_2 = trace(A^*A) = \sum s_i^2$, where s_i's are the singular values.

Proof. 1) The conditions for a normed space can be checked directly. 2) It's obvious that $\|A0\| = 0$. For nonzero x, we have $\|Ax\| = \|A\frac{x}{\|x\|}\|x\|\| = \|A\frac{x}{\|x\|}\|\|x\| \le \|A\|\|x\|$. 3) Note that $\sup\{\|Ax\| : \|x\| \le 1\} = \|Ax_0\|$ for some $x_0 \in \{x : \|x\| \le 1\}$ (a continuous function can achieve its supremum on a compact set). Suppose that $\|AB\| = \|ABx_0\|$. By 2), we have $\|ABx_0\| \le \|A\|\|Bx_0\| \le \|A\|\|B\|$. 4) follows the theorem of singular value decomposition. ■

4 Canonical forms of orthogonal matrices

Theorem 4.1 *Let A be an $n \times n$ orthogonal matrix.*

1) If $\det A = 1$, then A is orthogonal conjugate to

$$\begin{bmatrix} R_{\phi_1} & & & \\ & \ddots & & \\ & & R_{\phi_k} & \\ & & & I_{n-2k} \end{bmatrix}$$

where $R_{\phi_i} = \begin{bmatrix} \cos\phi_i & -\sin\phi_i \\ \sin\phi_i & \cos\phi_i \end{bmatrix}$ is the rotation matrix of angle ϕ_i.

2) If $\det A = -1$, then A is orthgonal conjugate to

$$\begin{bmatrix} R_{\phi_1} & & & & \\ & \ddots & & & \\ & & R_{\phi_k} & & \\ & & & I_l & \\ & & & & -1 \end{bmatrix}.$$

Proof. View A as a complex matrix. If $Ax = \lambda x$ for a unit vector x, we have $\|Ax\| = \|\lambda x\|$ implying $|\lambda| = 1$. Note that $A\bar{x} = \bar{\lambda}\bar{x}$. If $\lambda \neq \pm 1$, write $\lambda = \cos\phi + i\sin\phi$ and $x = x_1 + ix_2$ for real vectors x_1, x_2. It can directly check that

$$A[x_1, x_2] = [x_1, x_2] \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}.$$

Note that $\lambda \neq \bar{\lambda}$, which implies $x \perp \bar{x}$ and thus $x_1^T x_1 = x_2^T x_2$, $x_1 \perp x_2$. Moreover, the complement $Span_{\mathbb{R}}\{x_1, x_2\}^{\perp}$ is invariant under A. If $\lambda = \pm 1$, we can choose a real eigenvector x and consider the complement $Span_{\mathbb{R}}\{x_1, x_2\}^{\perp}$. Note that the number of -1 must be even when $\det A = 1$, while the number is odd when $\det A = -1$. An inductive argument finishes the proof after reordering the elements in the basis. ■

Lecture 5 : Symmetric matrices and quadratic forms

Shengkui Ye

May 3, 2023

1 Symmetric bilinear forms

Definition 1.1 *A symmetric bilinear form on a real vector space V over a field F is a function $\langle , \rangle : V \times V \to F$ such that*

1. $\langle u, v\rangle = \langle v, u\rangle$ *for any* $v, u \in V$;
2. $\langle v, a_1u_1 + a_2u_2\rangle = a_1\langle v, u_1\rangle + a_2\langle v, u_2\rangle$ *for any* $u_1, u_2, v \in V$ *and any* $a_1, a_2 \in F$;

Example 1.2 *Let $V = \mathbb{R}^3$. The function*

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

is a symmetric bilinear form.

Lemma 1.3 *A symmetric bilinear form $\langle , \rangle : V \times V \to F$ can always be represented by $\langle x, y\rangle = x^TAy$ for some symmetric matrix A.*

For a quadratic form $q(x) = x^TAx = \sum_{1\leq i,j\leq n} \frac{a_{ij}}{2} x_ix_j$, we already know that for an orthogonal matrix P the new form $q(Px) = x^TP^TAPx$ is a sum of squares. But for a change of variable $y = Sx$ (for an invertible matrix S), we may still have $q(Sx) = x^TS^TASx$ a sum of squares. In this section, we will study some invariants of $q(x)$ which depend only on A, not on S.

Definition 1.4 *Two square real matrices A, B are congruent if there is an invertible matrix S such that $B = SAS^T$. Similarly, we call two square complex matrices A, B congruent if there is an invertible matrix S such that $B = SAS^*$.*

Definition 1.5 *For a Hermitian matrix A (i.e. $A^* = A$), let n_+, n_-, n_0 be the number of positive, negative, zero eigenvalues, respectively. We can the triple (n_+, n_-, n_0) the signature of A.*

Theorem 1.6 *(Sylvester's law of inertia) Two Hermitian matrices A, B are congruent if and only if they have the same signature (i.e. they have the same number of of positive, negative, zero eigenvalues.)*

Proof. Since A, B are Hermitian, there exist unitary matrices Q_1, Q_2 such that $Q_1 A Q_1^* = D_1, Q_2 B Q_2^* = D_2$ are both real diagonal matrices. After permutation of diagonal elements and changing the absolute values, we see that D_1, D_2 are congruent, which implies that A, B are congruent.

Suppose that $B = SAS^*$ for an invertible matrix S. Since A is Hermitian, there is a unitary matrix U such that $A = UDU^*$ for a real diagonal matrix D. Then $B = SUDU^*S^*$. We claim that $n_+(B) = \max\{\dim V : V < F^n$ is a subspace on which B is positive definite$\}$. Actually, $B = VD'V^*$ for a unitary matrix V and a real diagonal matrix D'. Let V be the subspace spanned by the eigenvectors corresponding to the positive eigenvalues of D' (and B). We see that B is positive definite o V. If W is a subspace on which B is positive definite with the maximal $\dim W$, we know that the orthogonal complement $W^\perp$ is B-invariant (for any $x \in W, y \in W^\perp$, we have $\langle x, By\rangle = \langle B^*x, y\rangle = 0$). Since B has positive eigenvalues on W, this shows $\dim W \leq n_+$. Note that $n_+(B) = n_+(D) = n_+(A)$. Similarly, we have $n_-(B) = n_-(A), n_0(B) = n_0(A)$. ■

Corollary 1.7 *The maximal dimension of a positive definite subspace for quadratic form $q(x) = x^T Ax$ is n_+.*

2 Dual space

The following is a generalization of orthogonal complement.

Definition 2.1 *Let V be a vector over a field F. Its dual space is $V^* = \{f \mid f : V \to F$ is linear$\}$.*

Exercise 2.2 *Check that V^* is a vector space over F.*

Example 2.3 *Let $V = C[0,1]$ be the vector space of continuous functions. The integration $\int_0^1$ is a linear functional, i.e. a linear map from V to $\mathbb{R}$.*

Lemma 2.4 *Let V be a vector space. We have $V \cong (V^*)^*$, i.e. the dual of the dual of V is isomorphic to V.*

Definition 2.5 *A bilinear form $x^T Ay$ is non-degenerated if A is invertible.*

Lemma 2.6 *Let $\langle,\rangle : V \times V \to F$ be a symmetric bilinear form. The following are equivalent.*

1) $\langle,\rangle$ is non-degenerated.

2) The map $V \to V^*$,

$$x \longmapsto \langle -, x\rangle$$

is isomorphic of vector spacs. Here $\langle -, x\rangle$ is a linear function $y \mapsto \langle y, x\rangle$ for any $y \in V$.