Inner product on a vector space

- The inner product on $\mathbb{R}^n$ generalizes to an arbitrary vector space.

An **inner product** on a vector space V is a function that, to each pair of vectors $\mathbf{u}$ and $\mathbf{v}$ in V, associates a real number $\langle \mathbf{u}, \mathbf{v} \rangle$ and satisfies the following axioms, for all $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ in V and all scalars c:

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$

A vector space with an inner product is called an **inner product space**.

- Example: Let $a_1, \cdots, a_n$ be positive real numbers, we can define a "weighted" inner product on $\mathbb{R}^n$ as $\langle \mathbf{u}, \mathbf{v} \rangle = a_1 u_1 v_1 + \cdots a_n u_n v_n$.
- Example

EXAMPLE 7 For f, g in $C[a, b]$, set

$$\langle f, g \rangle = \int_a^b f(t)g(t)\,dt$$

Show that (5) defines an inner product on $C[a, b]$.

Proof: The properties 1), 2), 3) of the inner product is clear. For the last property, notice that

$\langle f, f \rangle = \int_a^b f(t)^2\,dt \geq 0$, and it is zero iff $f = 0$. $\boxtimes$

Lengths, distances and orthogonality

- Let V be an inner product space. For $\mathbf{v} \in V$, we define its length to be $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$, and the distance between $\mathbf{u}, \mathbf{v} \in V$ to be $\|\mathbf{u} - \mathbf{v}\|$.
- Two vectors $\mathbf{u}, \mathbf{v} \in V$ are said to be orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.
- Pythagoras theorem: Two vectors $\mathbf{u}, \mathbf{v} \in V$ are orthogonal if and only if $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u} + \mathbf{v}\|^2$.
- Let W be a sub vector space of V, we can define its orthogonal complement $W^\perp$ as $\{\mathbf{v} \in V \mid \langle \mathbf{v}, \mathbf{w} \rangle = 0, \forall\, \mathbf{w} \in W\}$.
- Theorem: Let W be a sub vector space of V, then $W^\perp$ remains a sub vector space of V, and we have $W \cap W^\perp = \{0\}, W + W^\perp = V$.

- So any vector $\mathbf{v} \in V$ can be written uniquely as $\mathbf{v} = \mathrm{Proj}_W(\mathbf{v}) + \mathrm{Proj}_{W^\perp}(\mathbf{v})$, with $\mathrm{Proj}_W(\mathbf{v}) \in W, \mathrm{Proj}_{W^\perp}(\mathbf{v}) \in W^\perp$.
- The projection formula holds also in the new more general setting. In particular, for a line $L = \mathrm{Span}\{\mathbf{u}\}$, we have $\mathrm{Proj}_L(\mathbf{v}) = \dfrac{\langle \mathbf{v}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle}\mathbf{u}$.
- Theorem

The Cauchy–Schwarz Inequality

For all $\mathbf{u}, \mathbf{v}$ in V,

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \le \|\mathbf{u}\|\,\|\mathbf{v}\| \tag{4}$$

Proof. By Pythagoras' theorem

$$\|v\|^2 = \|\mathrm{Proj}_W(v)\|^2 + \|v - \mathrm{Proj}_W(v)\|^2$$

$$\geq \|\mathrm{Proj}_W(v)\|^2.$$

$$\Rightarrow \|v\| \geq \|\mathrm{Proj}_W(v)\|$$

In particular, let $W = \mathrm{Span}\{u\}$,

then $$\|v\| \geq \left\| \frac{\langle v, u\rangle}{\langle u, u\rangle} u \right\|$$

$$= \frac{|\langle v, u\rangle|}{\|u\|^2} \|u\|$$

$$\Rightarrow \|u\| \cdot \|v\| \geq \|\langle u, v\rangle\|. \quad ☒$$

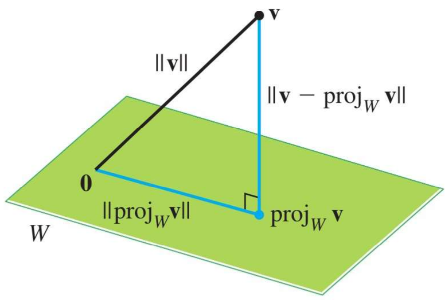


FIGURE 2
The hypotenuse is the longest side.

- Theorem

The Triangle Inequality
For all $\mathbf{u}, \mathbf{v}$ in V,
$$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

Proof: $$\|u+v\|^2 = \langle u+v, u+v\rangle$$

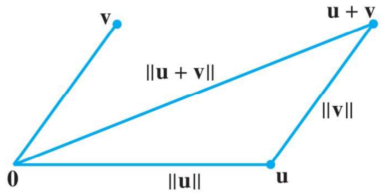
$$= \langle u, u\rangle + \langle v, v\rangle + 2\langle u, v\rangle$$

$$= \|u\|^2 + \|v\|^2 + 2\langle u, v\rangle$$

Cauchy - Schwartz $$\leq \|u\|^2 + \|v\|^2 + 2\|u\| \cdot \|v\|$$

$$= (\|u\| + \|v\|)^2$$

$$\Rightarrow \|u+v\| \leq \|u\| + \|v\|. \quad ☒$$



Quadratic forms and symmetric matrices

- The Euclidean norm square $\|\mathbf{v}\|^2 = v_1^2 + \cdots + v_n^2$ is a particular quadratic form (二次型) on $\mathbb{R}^n$.
- In general, a quadratic form on $\mathbb{R}^n$ is a function

$$Q(\mathbf{x}) = c_1 x_1^2 + \cdots + c_n x_n^2 + \sum_{i \neq j = 1}^{n} c_{ij} x_i x_j \,.$$

- The terms $c_{ij} x_i x_j$ are called cross-product terms (交叉项).

- Notice that it can be written in matrix form: Let $A = (a_{ij})$ be the $n \times n$ symmetric matrix such that $a_{ii} = c_i$ and $a_{ij} = a_{ji} = c_{ij}/2$, then $Q(\mathbf{x}) = \mathbf{x}^t A \mathbf{x}$. The matrix A is called the matrix of the quadratic form.
- Similar to the relationship between Euclidean norm and the inner product, we can associate the function $B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^t A \mathbf{y}$. It is clear that $Q(\mathbf{x}) = B(\mathbf{x}, \mathbf{x})$.
- As A is symmetric, the function $B(\mathbf{x}, \mathbf{y}) = \mathbf{x}^t A \mathbf{y}$ satisfies the properties:

1. $B(\mathbf{x}, \mathbf{y}) = B(\mathbf{y}, \mathbf{x})$,
2. $B(\mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}) = B(\mathbf{x}_1, \mathbf{y}) + B(\mathbf{x}_2, \mathbf{y})$,
3. $B(c\mathbf{x}, \mathbf{y}) = cB(\mathbf{x}, \mathbf{y})$.

- For this reason, B is called a symmetric bilinear form (对称双线性形式).

- Consider the change of variable $\mathbf{x} = P\mathbf{y}$, the quadratic form will be changed to $Q(\mathbf{y}) = \mathbf{y}^t(P^tAP)\mathbf{y}$.
- Question: Is it possible to find a change of variable such that the matrix of the quadratic form is nice? For example, diagonal?
- In case that P is orthogonal, i.e. $P^{-1} = P^t$, the transformation becomes $P^{-1}AP$ and is related to the diagonalization of the matrix A.
- Definition: The symmetric matrix A is said to be orthogonally diagonalizable if there exists an orthogonal matrix P such that $P^{-1}AP$ is diagonal.
- It is clear that A is orthogonally diagonalizabel if and only the eigenvectors of A forms an orthonormal basis of $\mathbb{R}^n$.

Diagonalization of symmetric matrices

- Theorem

The Spectral Theorem for Symmetric Matrices

An $n \times n$ symmetric matrix A has the following properties:

a. A has n real eigenvalues, counting multiplicities.

b. The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic equation.

c. The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal.

d. A is orthogonally diagonalizable.

- Geometrically, this means that we can find an orthonormal basis of $\mathbb{R}^n$ such that A acts as dilation with respect to this coordinate system.

- Lemma: Let A be an $n \times n$ symmetric matrix, then all the complex roots of the characteristic equation $\det(A - \lambda I) = 0$ are in fact real.

Proof: Let $\det(A - \lambda I) = \prod_{i=1}^{r} (\lambda_i - \lambda)^{m_i}$ with $\lambda_i \in \mathbb{C}$.

Consider A as a linear transformation $A : \mathbb{C}^n \to \mathbb{C}^n$, then for each λ_i, we have an eigenvector

$$A v_i = \lambda_i v_i .$$

Over $\mathbb{C}^n$, we can define an inner product similar to the Euclidean inner product: For $z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \in \mathbb{C}^n$, $w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{C}^n$,

complex conjugate

let $\langle z, w \rangle = \bar{z}_1 w_1 + \cdots + \bar{z}_n w_n = \bar{z}^t \cdot w$.

It is clear that the inner product has the properties

1) $\langle z, w \rangle = \overline{\langle w, z \rangle}$
2) $\langle z, c_1 w_1 + c_2 w_2 \rangle = c_1 \langle z, w_1 \rangle + c_2 \langle z, w_2 \rangle$, $\forall c_1, c_2 \in \mathbb{C}$,
3) $\langle z, z \rangle$ is real ≥ 0 and $\langle z, z \rangle = 0$ iff $z = 0$.

Now consider the inner product $\langle v_i, A v_i \rangle = \bar{v}_i^t A v_i$.

A real symmetric $= (A v_i)^t v_i$

$= \langle A v_i, v_i \rangle$

On the one hand,

(*) $\langle v_i, A v_i \rangle = \langle v_i, \lambda_i v_i \rangle = \lambda_i \underbrace{\langle v_i, v_i \rangle}_{>0}$.

On the other hand,

(**) $\langle v_i, A v_i \rangle = \langle A v_i, v_i \rangle = \langle \lambda_i v_i, v_i \rangle = \bar{\lambda}_i \langle v_i, v_i \rangle$

Compare (*), (**), get $\lambda_i = \bar{\lambda}_i$, so $\lambda_i \in \mathbb{R}$ as claim. $\boxed{\text{S}}$

- Lemma: Let A be an $n \times n$ symmetric matrix, let $\mathbf{v}$ be an eigenvector of A, then A preserves the orthogonal complement of $\mathrm{Span}\{\mathbf{v}\}$.

Proof Suppose that $Av = \lambda v$ for some $\lambda \in \mathbb{R}$.

For $w \in \mathrm{Span}\{v\}^{\perp}$, we have

$$\langle v, Aw \rangle = v^t A w = (Av)^t w = \langle Av, w \rangle$$
$$= \langle \lambda v, w \rangle = 0$$

So $Aw \in \mathrm{Span}\{v\}^{\perp}$, and hence A preserves $\mathrm{Span}\{v\}^{\perp}$.

$\boxtimes$

- Proof of the theorem:

a) follows from Lemma 1.

For the remaining assertions, we'll prove that A is orthogonally diagonalizable, i.e. there exists an orthogonal matrix P such that $A = P \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} P^t$. $(P^{-1} = P^t)$.

This implies the assertion b), c), d). Indeed, let $P = [v_1 \cdots v_n]$, then $\{v_1, \cdots, v_n\}$ form an orthonormal basis, and $A = P \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} P^t = P \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} P^{-1}$

is equivalent to $AP = P\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}$, i.e.

$$A[v_1 \cdots v_n] = [v_1 \cdots v_n]\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}$$

$$\Leftrightarrow \quad A v_i = \lambda_i v_i, \quad i = 1, \cdots, n.$$

We prove the assertion that A is orthogonally diagonalisable by induction. The case $n=1$ is evident. Suppose that the assertion is proved for $n-1$. Then for an $n \times n$ matrix A, it is known from lemma 1 that

$$\det(A - \lambda I) = \prod_{i=1}^{r} (\lambda_i - \lambda)^{m_i}, \quad \text{with } \lambda_i \in \mathbb{R}.$$

Let v_1 be an eigenvector with eigenvalue λ_1,

$$(*) \qquad A v_1 = \lambda v_1 .$$

Without loss of generality, we can assume that $\| v_1 \| = 1$.

By lemma 2, it is know that A preserves $\mathrm{Span}\{v_1\}^{\perp}$.

Let $\{v_2, \cdots, v_n\}$ be an orthonormal basis of $\mathrm{Span}\{\vec{v}_1\}$.

then $\{v_1, \cdots, v_n\}$ is an orthonormal basis of $\mathbb{R}^n$.

Notice that $(*)$ and the fact that A preserves $\mathrm{Span}\{v_1\}^{\perp}$ implies:

$$A\underbrace{[v_1\ v_2 \cdots v_n]}_{P_1 \text{ orthogonal.}} = [v_1\ v_2 \cdots v_n]\left(\begin{array}{c|ccc}\lambda_1 & 0 & \cdots & 0 \\ \hline 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{array}\right)$$

$$\Rightarrow \quad A = P_1 \begin{pmatrix} \lambda_1 & \\ & \boxed{A'} \end{pmatrix} P_1^t .$$

↑ induction hypothesis

$$A' = P_2 \begin{pmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} P_2^t .$$

$$\Rightarrow \quad A = P_1 \begin{pmatrix} \lambda_1 & \\ & P_2 \begin{pmatrix} \lambda_2 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} P_2^t \end{pmatrix} P_1^t$$

$$= \underbrace{P_1 \left(\begin{array}{c|c} 1 & \\ \hline & P_2 \end{array}\right)}_{P} \left(\begin{array}{c|ccc} \lambda_1 & & & \\ \hline & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{array}\right) \underbrace{\left(\begin{array}{c|c} 1 & \\ \hline & P_2^t \end{array}\right) P_1^t}_{P^t}$$

P product of orthogonal matrices remain orthogonal.

☒

EXAMPLE 3 Orthogonally diagonalize the matrix $A = \begin{bmatrix} 3 & -2 & 4 \\ -2 & 6 & 2 \\ 4 & 2 & 3 \end{bmatrix}$, whose characteristic equation is

$$0 = -\lambda^3 + 12\lambda^2 - 21\lambda - 98 = -(\lambda - 7)^2(\lambda + 2)$$

Applications to quadratic forms

- Recall that a quadratic form can be written as $Q(\mathbf{x}) = \mathbf{x}^t A\mathbf{x}$ for a symmetric matrix A, and that a change of variable $\mathbf{x} = P\mathbf{y}$ by an orthogonal matrix P will change Q to $Q(\mathbf{y}) = \mathbf{y}^t(P^t AP)\mathbf{y}$.
- Combined with the spectral theorem of symmetric matrices, we get the theorem

The Principal Axes Theorem

Let A be an $n \times n$ symmetric matrix. Then there is an orthogonal change of variable, $\mathbf{x} = P\mathbf{y}$, that transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into a quadratic form $\mathbf{y}^T D\mathbf{y}$ with no cross-product term.

- The column vectors of P are the eigenvectors of A, they are called the principle axes (主轴) of the quadratic form Q.

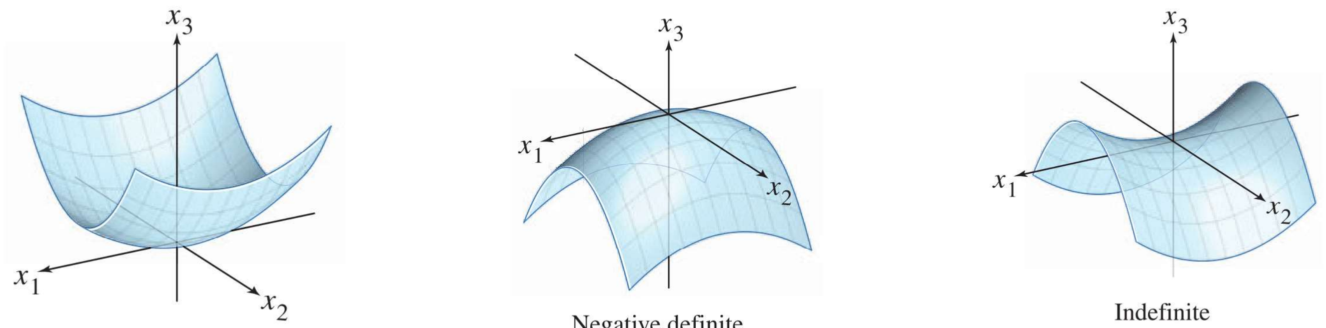
- In other words, with respect to the coordinate system given by the principal axes, the matrix for the quadratic form will be diagonal, i.e. it will be of the form $Q(\mathbf{y}) = \lambda_1 y_1^2 + \cdots + \lambda_n y_n^2$.
- Definition

A quadratic form Q is:

a. **positive definite** if $Q(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$,

b. **negative definite** if $Q(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{0}$,

c. **indefinite** if $Q(\mathbf{x})$ assumes both positive and negative values.



Positive definite   Negative definite   Indefinite

- Theorem

Quadratic Forms and Eigenvalues

Let A be an $n \times n$ symmetric matrix. Then a quadratic form $\mathbf{x}^T A \mathbf{x}$ is:

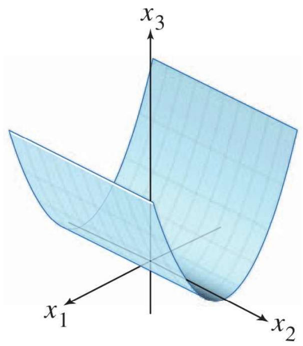
a. positive definite if and only if the eigenvalues of A are all positive,

b. negative definite if and only if the eigenvalues of A are all negative, or

c. indefinite if and only if A has both positive and negative eigenvalues.

Proof: The change of variable $\vec{y} = P\vec{x}$ with P orthogonal doesn't change whether Q is positive definite, negative definite or indefinite. By the principle axe theorem, we can assume that $Q = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2$, $\lambda_1, \ldots, \lambda_n$ eigenvalues of A. The assertion is clear for such Q.

- A quadratic form is called degenerate, if the associated matrix has 0 as one of its eigenvalues. With respect to the principal axis, it takes the form $Q(\mathbf{y}) = \lambda_1 y_1^2 + \cdots + \lambda_r y_r^2$, with $\lambda_1, \cdots, \lambda_r \neq 0$ and $r < n$.



(b) $z = 3x_1^2$