

Inner Product Spaces

Math 5250 — Week 5

Inner Products over $\mathbb{R}$

Inner Products

Definition. Let V be a real vector space. An *inner product* on V is a map $J : V \times V \rightarrow \mathbb{R}$ such that:

- a. (symmetry) $J(u, v) = J(v, u)$ for all $u, v \in V$;
- b. (linearity) $J(au_1 + bu_2, v) = aJ(u_1, v) + bJ(u_2, v)$. By symmetry, J is also linear in the second variable, so J is *bilinear*;
- c. (positive definiteness) $J(v, v) > 0$ for all nonzero $v \in V$.

We often write $J(u, v) = \langle u, v \rangle$.

A real vector space together with an inner product is a (real) *inner product space*.

Example: The Dot Product

Example. The Euclidean dot product on $\mathbb{R}^n$:

$$J((a_1, \dots, a_n), (b_1, \dots, b_n)) = \sum_{i=1}^n a_i b_i = (a_1, \dots, a_n) \cdot (b_1, \dots, b_n).$$

Inner product spaces make sense in infinite dimensions too, but we focus on the finite-dimensional case.

Geometry in an Inner Product Space

An inner product lets us measure the size of a vector and the distance between vectors:

$$\|v\| = \sqrt{\langle v, v \rangle}, \quad d(u, v) = \sqrt{\langle u - v, u - v \rangle} = \|u - v\|.$$

This distance is a *metric*:

- a. $d(u, v) \geq 0$, and $d(u, v) = 0$ if and only if $u = v$;
- b. $d(u, w) \leq d(u, v) + d(v, w)$ (the triangle inequality).

Example. For the standard dot product, this is the Pythagorean distance

$$d(u, v) = \sqrt{\sum_{i=1}^n (u_i - v_i)^2},$$

and $\{v : \|v\| = r\}$ is a sphere of radius r .

The Matrix of an Inner Product

Let $E = \{e_1, \dots, e_n\}$ be a basis for V . Define the $n \times n$ matrix $[J]^E$ by

$$[J]_{ij}^E = \langle e_i, e_j \rangle.$$

$$\begin{bmatrix} e_1 & e_2 \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

Then

$$\left\langle \sum a_i e_i, \sum b_j e_j \right\rangle = \begin{bmatrix} a_1 & \dots & a_n \end{bmatrix} [J]^E \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix},$$

$a_1 x_{11} + a_2 x_{21}$
 $a_1 x_{12} + a_2 x_{22}$
 $a_1 b_1 x_{11} + a_2 b_1 x_{21}$
 $+ b_2 a_1 x_{12}$
 $+ b_2 a_2 x_{22}$

that is,

$$\langle a, b \rangle = [a]_E^\top [J]^E [b]_E.$$

Definition. If X is $m \times n$, its *transpose* $X^\top$ is the $n \times m$ matrix with $X_{ij}^\top = X_{ji}$.

Remark. $(XY)^\top = Y^\top X^\top$.

$$X = (x_{ij}) \quad n \times m$$

$$X^\top = (x_{ji}) \quad m \times n$$

Properties of $[J]^E$

If $X = [J]^E$ comes from an inner product, then:

- a. X is symmetric; $(X^T = X)$
- b. X has rank n ;
- c. X is positive definite: $[a]_E^T X [a]_E > 0$ when $[a]_E \neq 0$;
- d. if T is another basis and $[a]_T = [1]_E^T [a]_E$, then

$$[J]^E = ([1]_E^T)^T [J]^T [1]_E^T.$$

$$[a]_E^T [J]^E [b]_E = [a]_E^T ([1]_E^T)^T [J]^T [1]_E^T [b]_E$$

$$= ([a]_E^T [1]_E^T)^T [J]^T [1]_E^T [b]_E$$

$$= ([a]_T^T)^T [J]^T [b]_T$$

$$[J]_{ij} = \langle e_i, e_j \rangle$$

$$[J]_{ji} = \langle e_j, e_i \rangle$$

$$\langle a, b \rangle = [a]^T X [b]$$

Inner Products from Matrices

Proposition. If X is a real symmetric positive definite $n \times n$ matrix, then

$$J(a, b) = a^\top X b$$

is an inner product on $\mathbb{R}^n$, and $X = [J]^E$ for the standard basis E .

Proposition. If J is an inner product on V and E is a basis, then $X = [J]^E$ is symmetric and positive definite, and

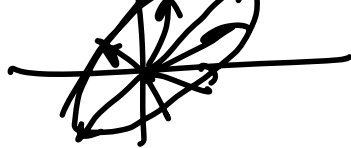
$$J(a, b) = [a]_E^\top [J]^E [b]_E.$$

In these coordinates, the surfaces $\{a : J(a, a) = r\}$ are ellipsoids.

$$X = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$$

symmetric

$$[a_1 \ a_2] \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} =$$



$$2a^2 + 2ab + 3b^2 = R^2$$

ellipse? hyperbola?

$$[a_1 \ a_2] \begin{bmatrix} 2a_1 + 3a_2 \\ a_1 + 3a_2 \end{bmatrix} >$$

$$2a_1^2 + a_1a_2 + a_1a_2 + 3a_2^2$$

$$2a_1^2 + 2a_1a_2 + 3a_2^2 > 0$$

$$2a_1^2 + 10a_1a_2 + 3a_2^2$$

Orthogonality

Orthogonality

Let V be an n -dimensional inner product space with inner product J .

Definition. Vectors $a, b \in V$ are *orthogonal* (relative to J) if $J(a, b) = 0$.

Example. For the standard Euclidean inner product on $\mathbb{R}^n$, orthogonal means perpendicular.

The Gram-Schmidt Process

V inner product space

$W \subseteq V$ a subspace
 $e_1, \dots, e_m$ a basis for W

Proposition. (Gram-Schmidt) Let $E = \{e_1, \dots, e_m\}$ be a linearly independent subset of V spanning W . Then there are vectors $U = \{u_1, \dots, u_m\}$ such that:

1. U is an *orthonormal* set:
 - a. $J(u_i, u_j) = 0$ if $i \neq j$;
 - b. $\|u_i\| = 1$ for $i = 1, \dots, m$;
2. for $j = 1, \dots, m$, the span W_j of $U_j = \{u_1, \dots, u_j\}$ equals the span of $E_j = \{e_1, \dots, e_j\}$.

Consequences of Gram-Schmidt

Part (2) says the change of basis is triangular:

$$u_1 = a_{11}e_1$$

$$u_2 = a_{12}e_1 + a_{22}e_2$$

$$\vdots$$

$$u_j = a_{1j}e_1 + \cdots + a_{jj}e_j$$

If E spans V , then $[J]^U$ is the identity matrix: in the U basis, J is the standard dot product,

$$J(a, b) = [a]_U \cdot [b]_U = \sum_{i=1}^n a_i b_i, \quad a = \sum a_i u_i, \quad b = \sum b_i u_i.$$



Proof of Gram-Schmidt

We proceed by induction. Let

$$u_1 = \frac{e_1}{\|e_1\|}.$$

This handles the case where W is one-dimensional.

Suppose $u_1, \dots, u_j$ are orthonormal, their span equals the span of E_j , and the change of basis is triangular. Let

$$v_{j+1} = e_{j+1} - \sum_{i=1}^j J(e_{j+1}, u_i) u_i.$$

Proof of Gram-Schmidt (continued)

For $k = 1, \dots, j$:

$$J(v_{j+1}, u_k) = J(e_{j+1}, u_k) - \sum_{i=1}^j J(e_{j+1}, u_i) J(u_i, u_k).$$

Since $J(u_i, u_k) = 0$ unless $i = k$, where it is 1, this is zero.

Also, the span of $\{u_1, \dots, u_j, v_{j+1}\}$ equals the span of $\{e_1, \dots, e_{j+1}\}$.

Let

$$u_{j+1} = \frac{v_{j+1}}{\|v_{j+1}\|}.$$

This constructs U as desired, and the result follows by induction. $\square$

Expansion in an Orthonormal Basis

Proposition. If $u_1, \dots, u_n$ is an orthonormal basis for V , then for every $v \in V$,

$$v = \sum_{i=1}^n J(v, u_i) u_i.$$

$$\left\langle v - \sum_{i=1}^n J(v, u_i) u_i, u_j \right\rangle = 0$$

$$\left\langle v - \sum_{i=1}^n J(v, u_i) u_i, u_j \right\rangle = 0$$

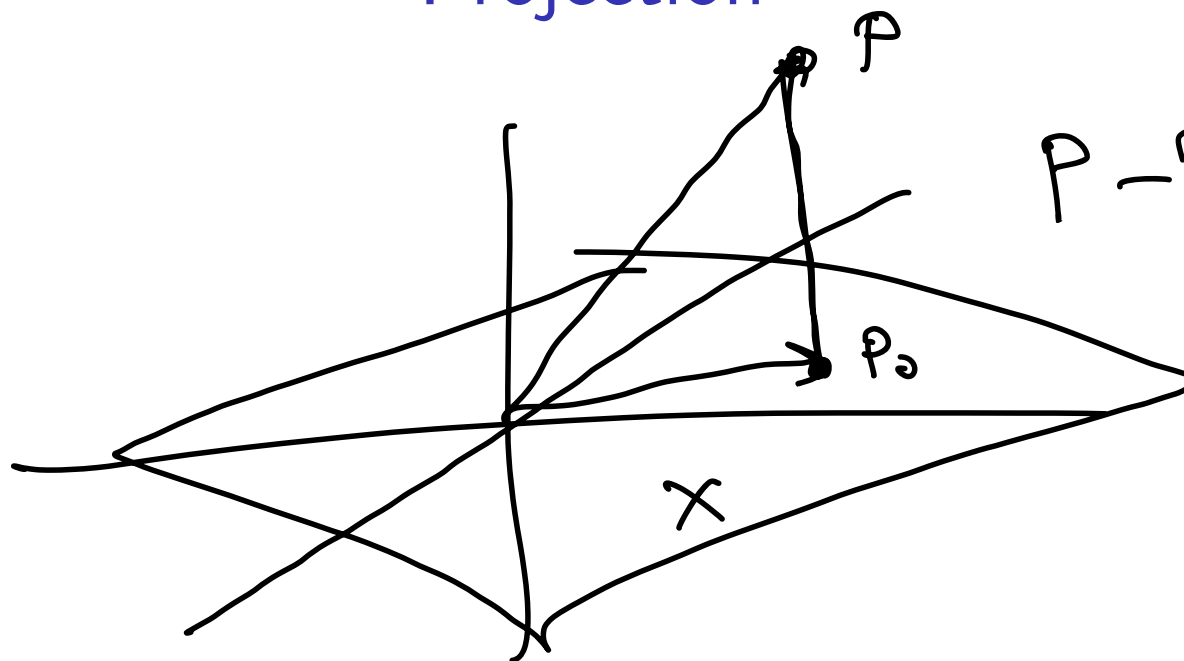
for all $j = 1, \dots, n$.

If $\langle v, u_i \rangle = 0$ for all i then $v = 0$.

$[v]_{u \cdot I} \cdot [i] = 0$ for all u_i
coordinates of $[v]$

in $\mathcal{O}$ basis are given by $v=0$.

Projection



$$P - P_0 \perp X$$

Orthogonal Complements

Let $W \subset V$ be a subspace.

Definition. The *orthogonal complement* of W is

$$W^\perp = \{v \in V : J(v, w) = 0 \text{ for all } w \in W\}.$$

Proposition.

- a. $W^\perp$ is a subspace of V .
- b. The map $W \oplus W^\perp \rightarrow V$, $(x, y) \mapsto x + y$, is an isomorphism.
In particular, $\dim W + \dim W^\perp = n$.
- c. $(W^\perp)^\perp = W$.

$W^\perp$ subspace

$W \oplus W^\perp \longrightarrow V$ is an isomorphism

$$v \in W \cap W^\perp \Rightarrow J(v, v) = 0 \Rightarrow v = 0$$

$$v \in V$$

Take a basis $w_1, \dots, w_n$ for W
 $u_1, \dots, u_n$ orthonormal basis for $W^\perp$.

$$v = v - \sum_{\substack{i \\ W^\perp}} \langle v, u_i \rangle u_i + \underbrace{\sum_{\substack{i \\ W}} \langle v, w_i \rangle w_i}_{\in W}$$

$$\cancel{W}^\perp \perp = W \quad \begin{matrix} k = \dim W \\ n = \dim V \end{matrix}$$

Choose a basis $w_1, \dots, w_k$ for W
and complete it to a basis $w_1, \dots, w_k, w_{k+1}, \dots, w_n$

By G-S $\underbrace{u_1, \dots, u_k}_{\text{basis for } W}, u_{k+1}, \dots, u_n$

$u_{k+1}, \dots, u_n$ basis for $W^\perp$.

Orthogonal Projection

Proposition. Inverting $W \oplus W^\perp \rightarrow V$ gives a map $P_W : V \rightarrow V$ with image W and kernel $W^\perp$, the *orthogonal projection* onto W . Similarly, $P_{W^\perp}$ has image $W^\perp$ and kernel W . Moreover:

- ▶ $P_W P_W = P_W$ and $P_{W^\perp} P_{W^\perp} = P_{W^\perp}$ (they are *idempotent*);
- ▶ $P_W + P_{W^\perp}$ is the identity.

In an orthonormal basis $u_1, \dots, u_n$ with $u_1, \dots, u_m$ spanning W :

$$[P_W] = \begin{pmatrix} I_{m \times m} & 0 \\ 0 & 0 \end{pmatrix}, \quad [P_{W^\perp}] = \begin{pmatrix} 0 & 0 \\ 0 & I_{(n-m) \times (n-m)} \end{pmatrix}.$$

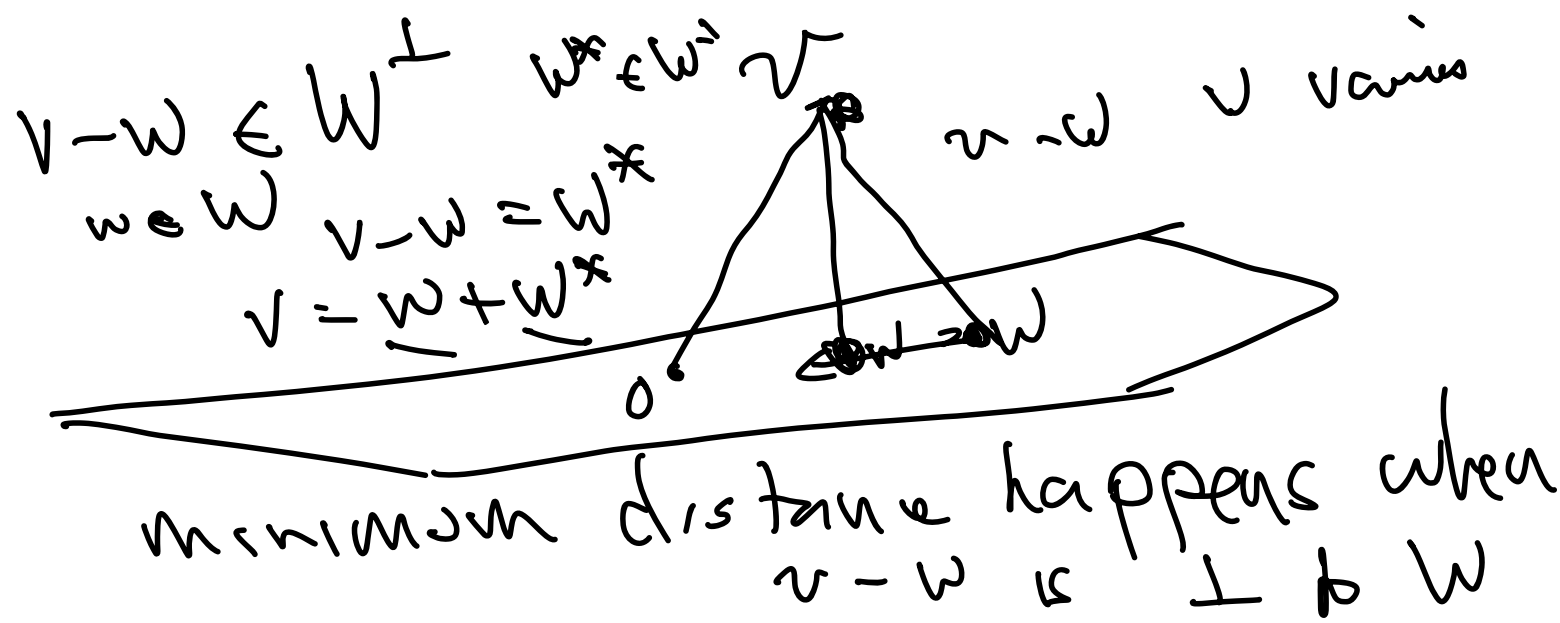
Closest Points

Distance to a Subspace

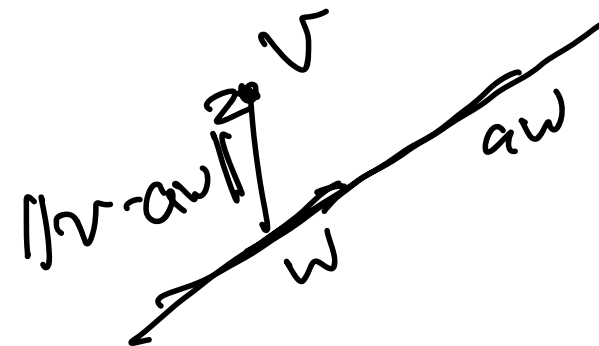
The *distance* from $v \in V$ to W is the minimum of $\|v - w\|$ as w ranges over W .

Proposition. The distance from v to W is

$$\|v - P_W(v)\| = \|P_{W^\perp}(v)\|.$$



The One-Dimensional Case



If W is spanned by w and $v \notin W$, then

$$\|v - aw\|^2 = J(v, v) - 2aJ(v, w) + a^2J(w, w)$$

is minimized at $a = \frac{J(v, w)}{J(w, w)}$. The vector from the closest point to v is

$$v_0 = v - \frac{J(v, w)}{J(w, w)}w = P_{W^\perp}(v),$$

and

$$\|v_0\|^2 = J(v, v) - \frac{J(v, w)^2}{J(w, w)} > 0.$$

The Cauchy-Schwarz Inequality

Proposition. (Cauchy-Schwarz) For v, w in a real inner product space,

$$J(v, w)^2 \leq J(v, v)J(w, w),$$

or equivalently

$$|J(v, w)| \leq \sqrt{J(v, v)J(w, w)},$$

with equality if and only if v is a multiple of w .

Application: Ordinary Least Squares

The Least Squares Problem

Given measurements $(x_{i1}, \dots, x_{ik})_{i=1}^N$ and $(y_i)_{i=1}^N$, we suspect a linear relation

$$y = \sum_{j=1}^k m_j x_j + b$$

up to random error.

The *least squares* choice of m_j and b minimizes

$$E(m, b) = \sum_{i=1}^N \left(y_i - \sum_{j=1}^k m_j x_{ij} - b \right)^2.$$

Least Squares as Projection

Let $Y \in \mathbb{R}^N$ have entries y_i , let X be the $N \times (k + 1)$ matrix with i^{th} row $(x_{i1}, \dots, x_{ik}, 1)$, and let $M = (m_1, \dots, m_k, b) \in \mathbb{R}^{k+1}$. Then

$$E(m, b) = \|Y - XM\|^2.$$

As M varies, XM ranges over the column space of X , so E is minimized when $Y - XM$ is orthogonal to every column of X :

$$X^\top(Y - XM) = 0.$$

The Normal Equations

Let $D = X^\top X$, a $(k + 1) \times (k + 1)$ matrix. If X has rank $k + 1$, then D is invertible and

$$M = D^{-1}X^\top Y.$$

The closest point in the column space of X is

$$XM = XD^{-1}X^\top Y,$$

the projection of Y onto the column space of X .

Projection Operator in General

Choose a basis E for V and write J for $[J]^E$. Let X be the matrix with columns $[w_1]_E, \dots, [w_k]_E$, where $w_1, \dots, w_k$ is a basis for W .

For $Y \in V$, we want $Y - XM$ orthogonal to the columns of X :

$$X^\top J(Y - XM) = 0 \quad \implies \quad M = (X^\top JX)^{-1} X^\top JY.$$

So the matrix of P_W is

$$[P_W]_E = X(X^\top JX)^{-1} X^\top J,$$

and one checks directly that $[P_W][P_W] = [P_W]$.

Complex (Hermitian) Inner Products

Hermitian Inner Products

Definition. Let V be a finite-dimensional complex vector space. A *Hermitian* inner product on V is a function $J : V \times V \rightarrow \mathbb{C}$ that:

1. is linear in the first variable and conjugate linear in the second:

$$J(au+bv, w) = aJ(u, w) + bJ(v, w), \quad J(u, av+bw) = \bar{a}J(u, v) + \bar{b}J(u, w)$$

2. is conjugate symmetric: $J(u, v) = \overline{J(v, u)}$ (so $J(u, u)$ is real);
3. is positive definite: $J(u, u) \geq 0$, with $J(u, u) = 0$ only if $u = 0$.

As in the real case, $\|v\| = \sqrt{J(v, v)}$.

The Matrix of a Hermitian Inner Product

Let $E = \{v_1, \dots, v_n\}$ be a basis and set $[J]_{ij}^E = J(v_i, v_j)$. Since

$$J\left(\sum a_i v_i, \sum b_j v_j\right) = \sum a_i \overline{b_j} J(v_i, v_j),$$

we have

$$J(u, v) = [u]_E^\top [J]^E \overline{[v]_E}.$$

The matrix $[J]^E$ is *Hermitian*: its transpose is its conjugate,

$$([J]^E)^\top = \overline{[J]^E}.$$

Gram-Schmidt for Hermitian Inner Products

If $\{v_1, \dots, v_n\}$ is a basis for a Hermitian inner product space, let

$$u_1 = \frac{v_1}{\|v_1\|},$$

and, assuming $u_1, \dots, u_k$ are orthonormal,

$$u'_{k+1} = v_{k+1} - \sum_{i=1}^k J(v_{k+1}, u_i) u_i, \quad u_{k+1} = \frac{u'_{k+1}}{\|u'_{k+1}\|}.$$

This gives an orthonormal basis $u_1, \dots, u_n$ for V in which the span of $u_1, \dots, u_j$ equals the span of $v_1, \dots, v_j$.