Chapter 6 : Orthogonality

ENSIA, 2024

Contents

The Inner Product

Definition 1

Let $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$ and $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ be two column vectors in $\mathbb{R}^n$. The real number

$$(u^T)v = u_1 v_1 + \cdots + u_n v_n$$

is called inner product of u and v, and it is written as $u \cdot v$.

Remark 1

The inner product is also referred to as a dot product.

Example

Example 1

Let $u = \begin{pmatrix} 2 \\ 1 \\ -1 \end{pmatrix}$ and $v = \begin{pmatrix} -3 \\ 5 \\ -2 \end{pmatrix}$. Then

$$u \cdot v = 2(-3) + 1 \times 5 + (-1)(-2) = 1.$$

Inner Product Properties

Theorem 1

Let u, v and w be vectors in $\mathbb{R}^n$. Then we have the following properties

a) $u \cdot v = v \cdot u$

b) $(u + v) \cdot w = u \cdot w + v \cdot w$

c) $(cu) \cdot v = c(u \cdot v) = u \cdot (cv)$ for all $c \in \mathbb{R}$

d) $u \cdot u \geq 0$ and $u \cdot u = 0$ if and only if $u = 0$

Corollary 1

For all $u_1, \cdots, u_p, w \in \mathbb{R}^n$ and for all $c_1, \cdots, c_p \in \mathbb{R}$, we have

$$\left(c_1 u_1 + \cdots + c_p u_p\right) \cdot w = c_1 (u_1 \cdot w) + \cdots + c_p (u_p \cdot w)$$

Norm and Unit Vector

Definition 2

Let $v \in \mathbb{R}^n$. The norm of v is the real number defined by

$$||v|| = \sqrt{{v_1}^2 + \cdots + {v_n}^2}.$$

A vector whose norm is 1 is called a unit vector.

Remark 2

The norm is also referred to as a length.

Distance in $\mathbb{R}^{\boldsymbol{n}}$

Definition 3

For $u, v \in \mathbb{R}^n$, the distance between u and v is given by

$$dist(u, v) = \|u - v\|.$$

Example 2

We compute the distance between the vectors $u = (2,5)$ and $v = (3,2)$.

We have $u - v = \begin{pmatrix} 2 \\ 5 \end{pmatrix} - \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$, then

$$dist(u, v) = \sqrt{(-1)^2 + 3^2} = \sqrt{10}.$$

Orthogonal Vectors and Pythagorean Theorem

Definition 4

Two vectors u and v in $\mathbb{R}^n$ are orthogonal if $u \cdot v = 0$.

Theorem 2 (The Pythagorean Theorem)

Two vectors u and v are orthogonal if and only if

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2.$$

Orthogonal Complements

Definition 5

Let W be a subspace of $\mathbb{R}^n$ and let $z \in \mathbb{R}^n$. We say that z is orthogonal to W if z is orthogonal to every vector of W. The set of all vectors z that are orthogonal to W is called the orthogonal complement of W and is denoted by $W^\perp$ (and read as "W perpendicular").

Theorem 3

1) Let A be a finite spanning set of a subspace W of $\mathbb{R}^n$. Then a vector x is in $W^\perp$ if and only if x is orthogonal to every vector in A.

2) $W^\perp$ is a subspace of $\mathbb{R}^n$.

Null Space, Column Space and Row Space

Definition 6

The null space of an $m \times n$ matrix A, written as $\mathrm{Nul}\ A$, is the set of all solutions of the homogeneous equation $Ax = 0$. In set notation,

$$\mathrm{Nul}\ A = \{x \in \mathbb{R}^n : Ax = 0\}.$$

The column space (resp. row space) of an $m \times n$ matrix A, written as $\mathrm{Col}\ A$ (resp. $\mathrm{Row}\ A$) is the set of all linear combinations of the columns (resp. rows) of A.

Remark 3

We have obviously

$$\mathrm{Row}\ A = \mathrm{Col} A^T$$

Null Space and Column Space

Proposition 1

The null space (resp. the column space) of a matrix A is a subspace of $\mathbb{R}^n$(resp. $\mathbb{R}^m$).

Theorem 4

Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:

$$(\mathrm{Row}\ A)^{\perp} = \mathrm{Nul}\ A \text{ and } (\mathrm{Col}\ A)^{\perp} = \mathrm{Nul} A^T.$$

Orthogonal Sets

Definition 7

A set of vectors $\{u_1, \cdots, u_p\}$ in $\mathbb{R}^n$ is said to be an orthogonal set if each pair of distinct vectors from the set is orthogonal, that is, if $u_i \cdot u_j = 0$ whenever $i \neq j$.

Theorem 5

If $S = \{u_1, \cdots, u_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

Orthogonal Basis

Definition 8

An orthogonal basis for a subspace W of $\mathbb{R}^n$ is a basis for W that is also an orthogonal set.

Theorem 6

Let $\{u_1, \cdots, u_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$, and let $y \in W$. If

$$y = c_1 u_1 + \cdots + c_p u_p$$

then we have

$$c_j = \frac{y \cdot u_j}{u_j \cdot u_j} \text{ for all } j \in \{1, \cdots, p\}.$$

Example

Example 3

Let $u_1 = (3, 1, 1)$, $u_2 = (-1, 2, 1)$ and $u_3 = (-1/2, -2, 7/2)$. Then $S = \{u_1, u_2, u_3\}$ is an orthogonal basis for $\mathbb{R}^3$. Using Theorem 5, we can express the vector $y = (6, 1, -8)$ as a linear combination of the vectors in S. We have

$$y = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2 + \frac{y \cdot u_3}{u_3 \cdot u_3} u_3$$

$$= \frac{11}{11} u_1 + \frac{-12}{6} u_2 + \frac{-33}{33/2} u_3$$
$$= u_1 - 2u_2 - 2u_3.$$

Orthogonal Projection onto a Vector

Given a nonzero vector u in $\mathbb{R}^n$, consider the problem of decomposing a vector y in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of u and the other orthogonal to u. We wish to write

$$y = \hat{y} + z \qquad (1)$$

where $\hat{y} = \alpha u$ for some scalar α and z is some vector orthogonal to u.

Proposition 2

The decomposition (1) is satisfied with z orthogonal to u if and only if

$$\alpha = \frac{y \cdot u}{u \cdot u}.$$

Definition 9

The vector $\hat{y}$ is called the orthogonal projection of y onto u, and the vector z is called the component of y orthogonal to u.

Example

Example 4

Let $y = (7, 6)$ and $u = (4, 2)$. Then

$$\alpha = \frac{y \cdot u}{u \cdot u} = \frac{40}{20} = 2.$$

So the orthogonal projection of y onto u is

$$\hat{y} = 2u = (8, 4)$$

and the component of y orthogonal to u is

$$z = y - \hat{y} = (7, 6) - (8, 4) = (-1, 2).$$

Therefore, the decomposition of y is given by

$$y = \hat{y} + z = (8, 4) + (-1, 2).$$

Orthonormal Basis

Definition 10

A set of vectors $\{u_1, \cdots, u_p\}$ in $\mathbb{R}^n$ is said to be an orthonormal set if it is an orthogonal set of unit vectors. An orthonormal basis of a subspace W is an orthonormal set which spans W.

Example 5

The standard basis $(e_1, \cdots, e_n)$ of $\mathbb{R}^n$ is an orthonormal basis.

Matrices with Orthonormal Columns

Theorem 7

A $m \times n$ matrix U has orthonormal columns if and only if

$$U^T U = I$$

Theorem 8

Let U be an $m \times n$ matrix with orthonormal columns and let $x, y \in \mathbb{R}^n$. Then

a) $\|Ux\| = \|x\|$

b) $(Ux) \cdot (Uy) = x \cdot y$

c) $(Ux) \cdot (Uy) = 0$ if and only if $x \cdot y = 0$

Orthogonal Matrices

Definition 11

An orthogonal matrix is a square invertible matrix U verifying

$$U^{-1} = U^T$$

The Orthogonal Decomposition Theorem

Theorem 9 (The orthogonal Decomposition Theorem)

Let W be a subspace of $\mathbb{R}^n$. Then each $y \in \mathbb{R}^n$ can be written uniquely in the form

$$y = \hat{y} + z \qquad (1)$$

where $\hat{y} \in W$ and $z \in W^{\perp}$.

In fact, if $\{u_1, \cdots, u_p\}$ is any orthogonal basis of W, then

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \cdots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p \qquad (2)$$

and

$$z = y - \hat{y}.$$

Orthogonal Projection onto a Subspace

Definition 12

The vector $\hat{y}$ in (1) of Theorem 9 is called the orthogonal projection of y onto W, and it is denoted by

$$\mathrm{proj}_W y.$$

Remark 4

The uniqueness of the decomposition (1) shows that the orthogonal projection $\mathrm{proj}_W y$ depends only on W and not on a particular choice of a basis of W.

Example

Example 6

Let $u_1 = (2, 5, -1), u_2 = (-2, 1, 1), y = (1, 2, 3)$ and set $W = \langle u_1, u_2 \rangle$. The set $\{u_1, u_2\}$ is orthogonal and the orthogonal projection of y onto W is given by

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2 = \frac{1}{5}(-2, 10, 1).$$

So we can write

$$y = (1, 2, 3) = \frac{1}{5}(-2, 10, 1) + \frac{1}{5}(7, 0, 14) = \hat{y} + z,$$

where $\hat{y} \in W$ and $z \in W^{\perp}$.

Best Approximation Theorem

Theorem 10 (The Best Approximation Theorem)

Let W be a subspace of $\mathbb{R}^n$, let $y \in \mathbb{R}^n$ and let $\hat{y}$ be the orthogonal projection of y onto W. Then $\hat{y}$ is the closest point in W to y, that is

$$||y - \hat{y}|| < ||y - v|| \qquad (3)$$

for all $v \in W, v \neq \hat{y}$.

Definition 13

The vector $\hat{y}$ in (3) is called the best approximation to y by elements of W.

Remark 5

The distance from y to v, given by $||y - v||$, can be regarded as the error of using v in place of y. Theorem 10 says that this error is minimized for $v = \hat{y}$.

Proof

Proof of Theorem 10

Let $v \in W, v \neq \hat{y}$. Then $\hat{y} - v \in W$. By the ODT, $y - \hat{y} \in W^{\perp}$. In particular $y - \hat{y}$ is orthogonal to $\hat{y} - v$.

Since $y - v = (y - \hat{y}) + (\hat{y} - v)$, the Pythagorean Theorem gives :

$$\|y - v\|^2 = \|y - \hat{y}\|^2 + \|\hat{y} - v\|^2$$

$\|\hat{y} - v\|^2 > 0$ because $\hat{y} - v \neq 0$, so

$$\|y - v\|^2 > \|y - \hat{y}\|^2$$

And finally

$$\|y - \hat{y}\| < \|y - v\| \,.$$

Example

Example 7

Let $u_1 = (2, 5, -1), u_2 = (-2, 1, 1), y = (1, 2, 3)$ and $W = \langle u_1, u_2 \rangle$, as in Example 6, then

$$\hat{y} = \frac{1}{5}(-2, 10, 1)$$

is the closest point in W to y.

Orthogonal Projection with Orthonormal Basis

Theorem 11

Let $\{u_1, \cdots, u_p\}$ be an orthonormal basis for a subspace W of $\mathbb{R}^n$, and let $y \in \mathbb{R}^n$. Then

$$\text{proj}_W y = (y \cdot u_1)u_1 + \cdots + (y \cdot u_p)u_p \qquad (4).$$

If $U = (u_1 u_2 \cdots u_p)$, then

$$\text{proj}_W y = UU^T y.$$

Gram-Schmidt Process

Theorem 12 (The Gram-Schmidt Process)

Let $\{x_1, \cdots, x_p\}$ be a basis for a nonzero subspace W of $\mathbb{R}^n$, and define

$$v_1 = x_1$$

$$v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1$$

$$v_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2$$

$$\vdots$$

$$v_p = x_p - \frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_p \cdot v_2}{v_2 \cdot v_2} v_2 - \cdots - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1}.$$

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

$$\langle v_1, \cdots, v_k \rangle = \langle x_1, \cdots, x_k \rangle \text{ for } 1 \leq k \leq p.$$

Proof

For $1 \leq k \leq p$, let $W_k = \langle x_1, x_2, \cdots, x_k \rangle$. Set $v_1 = x_1$, then $\langle v_1 \rangle = \langle x_1 \rangle$.

Suppose, for some $k < p$, we have constructed $v_1, \cdots, v_k$ so that $\{v_1, \cdots, v_k\}$ is an orthogonal basis for W_k. Define

$$v_{k+1} = x_{k+1} - \mathrm{proj}_{W_k} x_{k+1}$$

By the ODT, v_{k+1} is orthogonal to W_k.

Note that $v_{k+1} \in W_{k+1}$.

Furthermore, $v_{k+1} \neq 0$ since $x_{k+1} \notin W_k$.

Proof

Hence $\{v_1, \cdots, v_{k+1}\}$ is an orthogonal set of nonzero vectors in the $(k+1) -$ dimensional space W_{k+1}. Thus this set is an orthogonal basis for W_{k+1}. Therefore $W_{k+1} = \langle v_1, \cdots, v_{k+1} \rangle$.

The process stops when $k + 1 = p$.

Example

Example 8

Let $W = \langle x_1, x_2 \rangle$ where $x_1 = (3, 6, 0)$ and $x_2 = (1, 2, 2)$. Following the Gram-Schmidt Process, we will construct an orthogonal basis for W. Set

$v_1 = x_1$ and

$$v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 = x_2 - \frac{x_2 \cdot x_1}{x_1 \cdot x_1} x_1 = (1, 2, 2) - \frac{15}{45}(3, 6, 0) = (0, 0, 2).$$

Then $\{v_1 = (3, 6, 0), v_2 = (0, 0, 2)\}$ is an orthogonal basis for W.

If we want to construct an orthonormal basis for W, it suffices to normalize the vectors v_1 and v_2, that is we take $v_1' = \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{5}}(1, 2, 0)$ and $v_2' = \frac{v_2}{\|v_2\|} = (0, 0, 1)$.

Now $\left\{v_1{}' = \frac{1}{\sqrt{5}}(1, 2, 0), v_2{}' = (0, 0, 1)\right\}$ is an orthonormal basis for W.

Example

Example 9

Consider the subspace of $\mathbb{R}^4$ given by

$$W = \langle x_1, x_2, x_3 \rangle,$$

where $x_1 = (1,1,1,1)$, $x_2 = (0,1,1,1)$ and $x_3 = (0,0,1,1)$. We will construct an orthonormal basis for W.

QR Factorization

Theorem 13

If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for $\mathrm{Col}\ A$ and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

To prove this theorem, we will use the following lemma.

Lemma 1

Suppose $A = QR$, where Q is $m \times n$ and R is $n \times n$. If the columns of A are linearly independent, then R is invertible.

Proof of Lemma

Proof of Lemma

Let $x \in \mathbb{R}^n$, such that $Rx = 0$, then $Ax = 0$.

Set $A = (u_1, u_2, \cdots, u_n)$. We have

$$Ax = 0 \Leftrightarrow x_1 u_1 + x_2 u_2 + \cdots + x_n u_n = 0$$

With $x = (x_1, \cdots, x_n)$.

Since the columns of A are LI, it follows that $x_1 = \cdots = x_n = 0$,

$x = 0$. This means that R is invertible.

Proof of Theorem 13

Proof of the Theorem 13

Set $A = (x_1, \cdots, x_n)$.

The columns of A form a basis of $col(A)$.

By the ODT, there exists an orthonormal basis $\{u_1, u_2, \cdots, u_n\}$ of $col(A)$, and we have

$$\langle u_1, u_2, \cdots, u_k \rangle = \langle x_1, x_2, \cdots, x_k \rangle \text{ for all } 1 \le k \le n. \quad (1)$$

Set $Q = (u_1, u_2, \cdots, u_n)$.

From (1), we have that for all k, $1 \le k \le n$, there exist

$r_{1k}, r_{2k}, \cdots, r_{kk} \in \mathbb{R}$ such that $x_k = r_{1k}u_1 + r_{2k}u_2 + \cdots + r_{kk}u_k$.

Proof of Theorem 13

WLOG we can suppose $r_{kk} \geq 0$, since otherwise we can replace $r_{kk} < 0$ by $-r_{kk}$ and u_k by $-u_k$.

Set $r_k = \begin{pmatrix} r_{1k} \\ r_{2k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^n$ and $r = (r_1, r_2, \cdots, r_n)$.

We have $x_k = Qr_k$ for all $k \in \{1,2,\cdots,n\}$.

Proof

$$A = (x_1, \cdots, x_n) = (Qr_1, Qr_2, \cdots, Qr_n) = QR$$

Since A has LI columns, so R is invertible. Since R is upper triangular, then $detR = \prod_{i=1}^{n} r_{ii} \neq 0$. Since $r_{ii} \geq 0, \forall i$, then $r_{ii} > 0, \forall i$.

Example

Example 10

Find a QR factorization of

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Least-Squares Problems

Definition 14

If A is $m \times n$ and b is in $\mathbb{R}^m$, a least-squares solution of $Ax = b$ is an $\hat{x}$ in $\mathbb{R}^n$ such that

$$||b - A\hat{x}|| \leq ||b - Ax||$$

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

Theorem 14

The set of least-squares solutions of $Ax = b$ coincides with the nonempty set of solutions of the normal equations $A^T Ax = A^T b$

Least-Squares Problems

Theorem 15

Let A be an $m \times n$ matrix. The following statements are logically equivalent :

a. The equation $Ax = b$ has a unique least-squares solution for each b in $\mathbb{R}^m$.

b. The columns of A are linearly independent.

c. The matrix ${}^tAA$ is invertible.

When these statements are true, the least-squares solution $\hat{x}$ is given by

$$\hat{x} = (A^T)^{-1}A^T b$$