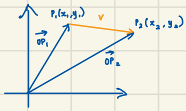
Vectors Whose Initial Point not @ Origin

$$v = \overrightarrow{P_1P_2}$$
$$= \overrightarrow{OP_2} - \overrightarrow{OP_1}$$



Finding the Components of a Vector

$$\overrightarrow{P_1P_2} = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$$

initial point : $(x_1\ y_1\ z_1)$

terminal point : $(x_2\ y_2\ z_2)$

SET	TUPLE
>> unordered collection of distinct objects	>> ordered collection of objects
$\{1,2,2,3\} = \{1,2,3\}$	$(1,2,2,3) \neq (1,2,3)$
$\{1,2,3\} = \{3,2,1\}$	$(1,2,3) \neq (3,2,1)$

Length / Magnitude / Norm of Vector

$$\|v\| = \sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}$$

OR

$$\|v\|^2 = v \cdot v = v_1^2 + v_2^2 + \ldots + v_n^2$$

>> Norm is never negative

>> Either : 0 (zero vector only) or positive

Unit Vector, u

"normalise the vector"

"vector of norm 1"

$$u = \frac{v}{\|v\|}$$

Distance between u and v

$$\text{dist}(u, v) = \|u - v\|$$
$$= \sqrt{(u_1 - v_1)^2 + \ldots + (u_n - v_n)^2}$$

Theorem 3.2.1

(a) $\|v\| \geq 0$

(b) $\|v\| = 0 \iff v = 0$

(c) $\|kv\| = |k|\,\|v\|$

Theorem 3.2.2

(a) $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$

(b) $\vec{u} \cdot (\vec{v} + \vec{w}) = \vec{u} \cdot \vec{v} + \vec{u} \cdot \vec{w}$

(c) $k(\vec{u} \cdot \vec{v}) = (k\vec{u}) \cdot \vec{v}$

(d) $\vec{v} \cdot \vec{v} \geq 0$ and $\vec{v} \cdot \vec{v} = 0$ if and only if $\vec{v} = \vec{0}$

Theorem 3.2.3

(a) $\vec{0} \cdot \vec{v} = \vec{v} \cdot \vec{0} = 0$

(b) $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{u} \cdot \vec{v}$

(c) $\vec{u} \cdot (\vec{v} - \vec{w}) = \vec{u} \cdot \vec{v} - \vec{u} \cdot \vec{w}$

(d) $(\vec{u} - \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} - \vec{v} \cdot \vec{w}$

(e) $k(\vec{u} \cdot \vec{v}) = \vec{u} \cdot (k\vec{v})$

Sample Questions

>> $v = (1, -2, 2, 0)$. Find unit vector.

$$\|v\| = \sqrt{(1)^2 + (-2)^2 + (2)^2 + 0}$$
$$= 3$$

$$u = \frac{v}{\|v\|} = \frac{1}{3}(1, -2, 2, 0) = \left(\frac{1}{3}, -\frac{2}{3}, \frac{2}{3}, 0\right)$$

Check $\|u\| = 1$:

$$\|u\| = \sqrt{\left(\tfrac{1}{3}\right)^2 + \left(-\tfrac{2}{3}\right)^2 + \left(\tfrac{2}{3}\right)^2 + 0} = 1$$

>> Compute distance between u & v.

$u = (7, 1)$, $v = (3, 2)$

$$u - v = (4, -1)$$

$$\|u - v\| = \sqrt{4^2 + (-1)^2} = \sqrt{17}$$

Dot product

θ between u & v satisfies $0 < \theta < \pi$

$$u \cdot v = \|u\| \|v\| \cos\theta$$ ← anti-clockwise measurement

$\Rightarrow$ θ is acute if $u \cdot v > 0$

$\Rightarrow$ θ is obtuse if $u \cdot v < 0$

$\theta = \frac{\pi}{2}$ if $u \cdot v = 0$

When coordinates are given,

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \cdot \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$$

if u & v are vectors in R^n

$$= (x_1 x_2) + (y_1 y_2) + (z_1 z_2)$$

Length of vector in terms of dot product : $\|v\| = \sqrt{v \cdot v}$

$u = v$

Orthogonality Definition

2 non-zero vectors $\vec{u}$ and $\vec{v}$ in R^n is orthogonal (perpendicular) if $u \cdot v = 0$;
$\vec{0}$ in R^n is orthogonal to every vector in R^n

Linear Combination of Standard unit vectors

$$v = (v_1, v_2, v_3)$$
$$= v_1(1,0,0) + v_2(0,1,0) + v_3(0,0,1)$$
$$= v_1 \vec{i} + v_2 \vec{j} + v_3 \vec{k}$$

Equation of straight line

$$a(x - x_0) + b(y - y_0) = 0$$

Equation of plane

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

Dot product properties

if u, v, w are vectors in R^n

$$u \cdot v = u^T v = v^T u$$
$$Au \cdot v = u \cdot A^T v$$
$$u \cdot Av = A^T u \cdot v$$

A^T : rows of A becomes columns of A

Theorem 3.2.4

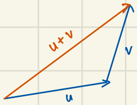
Cauchy - Schwarz Inequality

$$|u \cdot v| \le \|u\| \|v\|$$

OR

$$|u_1 v_1 + \ldots + u_n v_n| \le \sqrt{u_1^2 + \ldots + u_n^2} \sqrt{v_1^2 + \ldots + v_n^2}$$

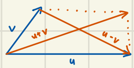
Theorem 3.2.5



(a) $\|u + v\| \le \|u\| + \|v\|$

(b) $d(u,v) \le d(u,w) + d(w,v)$

Theorem 3.2.6



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

" sum of squares of 2 diagonals = sum of squares of 4 sides "

Theorem 3.2.7

$$u \cdot v = \frac{1}{4}\|u + v\|^2 - \frac{1}{4}\|u - v\|^2$$
$$\|u + v\|^2 = (u + v) \cdot (u + v)$$
$$= \|u\|^2 + 2(u \cdot v) + \|v\|^2$$
$$\|u - v\|^2 = (u - v) \cdot (u - v)$$
$$= \|u\|^2 - 2(u \cdot v) + \|v\|^2$$

Theorem 3.3.1

Line equation :

$ax + by = c$ has normal (a, b)

Plane equation :

$ax + by + cz = d$ has normal (a, b, c)

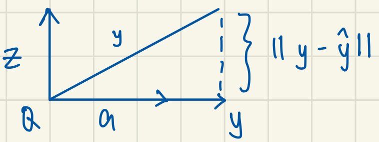
Orthogonal Projections (Theorem 3.3.2)

y as the sum of 2 orthogonal vectors.

$y = \hat{y} + (y - \hat{y}) = y + z$

$\hat{y}$: scalar multiple of a

$y - \hat{y}$: orthogonal to a



$\hat{y} = \text{proj}_a y = \frac{y \cdot a}{a \cdot a} a$ (aka projection)

(vector component y along a)

$z = y - \hat{y} = y - \frac{y \cdot a}{a \cdot a} a$

(vector component of u orthogonal to a)

($a \cdot a$ to normalize vector y)

$$w_1 = \text{proj}_{\mathbf{a}}\mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} \quad \textit{(vector component of } \mathbf{u} \textit{ along } \mathbf{a}\textit{)} \quad = ||u||cos\theta \tag{10}$$

$$w_2 = \mathbf{u} - \text{proj}_{\mathbf{a}}\mathbf{u} = \mathbf{u} - \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} \quad \textit{(vector component of } \mathbf{u} \textit{ orthogonal to } \mathbf{a}\textit{)} \tag{11}$$

EXAMPLE 5 Vector Component of u Along a

Let $\mathbf{u} = (2, -1, 3)$ and $\mathbf{a} = (4, -1, 2)$. Find the vector component of $\mathbf{u}$ along $\mathbf{a}$ and the vector component of $\mathbf{u}$ orthogonal to $\mathbf{a}$.

Solution

$$\mathbf{u} \cdot \mathbf{a} = (2)(4) + (-1)(-1) + (3)(2) = 15$$
$$\|\mathbf{a}\|^2 = 4^2 + (-1)^2 + 2^2 = 21$$

Thus the vector component of $\mathbf{u}$ along $\mathbf{a}$ is

$$\text{proj}_{\mathbf{a}}\mathbf{u} = \frac{\mathbf{u} \cdot \mathbf{a}}{\|\mathbf{a}\|^2}\mathbf{a} = \tfrac{15}{21}(4, -1, 2) = \left(\tfrac{20}{7}, -\tfrac{5}{7}, \tfrac{10}{7}\right)$$

and the vector component of $\mathbf{u}$ orthogonal to $\mathbf{a}$ is

$$\mathbf{u} - \text{proj}_{\mathbf{a}}\mathbf{u} = (2, -1, 3) - \left(\tfrac{20}{7}, -\tfrac{5}{7}, \tfrac{10}{7}\right) = \left(-\tfrac{6}{7}, -\tfrac{2}{7}, \tfrac{11}{7}\right)$$

As a check, you may wish to verify that the vectors $\mathbf{u} - \text{proj}_{\mathbf{a}}\mathbf{u}$ and $\mathbf{a}$ are perpendicular by showing that their dot product is zero. ◀

Norm of vector component

$\hat{y} = \frac{|y \cdot a|}{\|a\|}$

if there's an angle θ,

$\hat{y} = \|y\| \, |\cos\theta|$

Theorem 3.3.3

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

where u & v are orthogonal vectors in R^n

EXAMPLE 6 Theorem of Pythagoras in R^4

We showed in Example 1 that the vectors

$$\mathbf{u} = (-2, 3, 1, 4) \quad \text{and} \quad \mathbf{v} = (1, 2, 0, -1)$$

are orthogonal. Verify the Theorem of Pythagoras for these vectors.

Solution We leave it for you to confirm that

$$\mathbf{u} + \mathbf{v} = (-1, 5, 1, 3)$$
$$\|\mathbf{u} + \mathbf{v}\|^2 = 36$$
$$\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = 30 + 6$$

Thus, $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ ◀

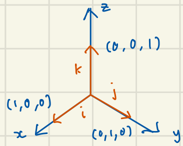
Orthogonal Sets

A set of vectors $\{u_1, \dots u_p\}$ in $\mathbb{R}^n$ is a orthogonal if each pair of distinct vectors is orthogonal, if $u_i \cdot u_j = 0$ whenever $i \neq j$

⇒ show a set is orthogonal by doing dot product of the vectors on each other

Standard basis

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



Orthogonal Basis

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

Linear Combination of orthogonal basis

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

Theorem 4

Note. $p \leq n$

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

⇒ if there a scalar $c_1 \dots c_p$, they must be 0

⇒ relationship is trivial solution 0

PROOF If $\mathbf{0} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ for some scalars $c_1, \dots, c_p$, then

$$\begin{aligned} 0 = \mathbf{0}\cdot\mathbf{u}_1 &= (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)\cdot\mathbf{u}_1 \\ &= (c_1\mathbf{u}_1)\cdot\mathbf{u}_1 + (c_2\mathbf{u}_2)\cdot\mathbf{u}_1 + \cdots + (c_p\mathbf{u}_p)\cdot\mathbf{u}_1 \\ &= c_1(\mathbf{u}_1\cdot\mathbf{u}_1) + c_2(\mathbf{u}_2\cdot\mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p\cdot\mathbf{u}_1) \\ &= c_1(\mathbf{u}_1\cdot\mathbf{u}_1) \end{aligned}$$

because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \dots, \mathbf{u}_p$. Since $\mathbf{u}_1$ is nonzero, $\mathbf{u}_1\cdot\mathbf{u}_1$ is not zero and so $c_1 = 0$. Similarly, $c_2, \dots, c_p$ must be zero. Thus S is linearly independent. ■

Theorem 5

Let $\{u_1 \dots u_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each y in W, the weights in the linear combination

$$y = c_1 \vec{u}_1 + \dots + c_p \vec{u}_p$$

are given by

$$c_j = \frac{y \cdot u_j}{u_j \cdot u_j} \quad (j = 1 \dots p)$$

⇒ division by $u_j \cdot u_j$ only required when not unit vector.

EXAMPLE 2 The set $S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ in Example 1 is an orthogonal basis for $\mathbb{R}^3$. Express the vector $\mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$ as a linear combination of the vectors in S.

SOLUTION Compute

$$\mathbf{y}\cdot\mathbf{u}_1 = 11, \qquad \mathbf{y}\cdot\mathbf{u}_2 = -12, \qquad \mathbf{y}\cdot\mathbf{u}_3 = -33$$
$$\mathbf{u}_1\cdot\mathbf{u}_1 = 11, \qquad \mathbf{u}_2\cdot\mathbf{u}_2 = 6, \qquad \mathbf{u}_3\cdot\mathbf{u}_3 = 33/2$$

By Theorem 5,

$$\begin{aligned} \mathbf{y} &= \frac{\mathbf{y}\cdot\mathbf{u}_1}{\mathbf{u}_1\cdot\mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}\cdot\mathbf{u}_2}{\mathbf{u}_2\cdot\mathbf{u}_2}\mathbf{u}_2 + \frac{\mathbf{y}\cdot\mathbf{u}_3}{\mathbf{u}_3\cdot\mathbf{u}_3}\mathbf{u}_3 \\ &= \frac{11}{11}\mathbf{u}_1 + \frac{-12}{6}\mathbf{u}_2 + \frac{-33}{33/2}\mathbf{u}_3 \\ &= \mathbf{u}_1 - 2\mathbf{u}_2 - 2\mathbf{u}_3 \end{aligned}$$

■

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} -1/2 \\ -2 \\ 7/2 \end{bmatrix}$$

Orthonormal Sets

EXAMPLE 5 Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis of $\mathbb{R}^3$, where

$$\mathbf{v}_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$$

SOLUTION Compute

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = -3/\sqrt{66} + 2/\sqrt{66} + 1/\sqrt{66} = 0$$
$$\mathbf{v}_1 \cdot \mathbf{v}_3 = -3/\sqrt{726} - 4/\sqrt{726} + 7/\sqrt{726} = 0$$
$$\mathbf{v}_2 \cdot \mathbf{v}_3 = 1/\sqrt{396} - 8/\sqrt{396} + 7/\sqrt{396} = 0$$

Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set. Also,

$$\mathbf{v}_1 \cdot \mathbf{v}_1 = 9/11 + 1/11 + 1/11 = 1$$
$$\mathbf{v}_2 \cdot \mathbf{v}_2 = 1/6 + 4/6 + 1/6 = 1$$
$$\mathbf{v}_3 \cdot \mathbf{v}_3 = 1/66 + 16/66 + 49/66 = 1$$

which shows that $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are unit vectors. Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal set. Since the set is linearly independent, its three vectors form a basis for $\mathbb{R}^3$. See Fig. 6. ■

Square Matrixes

See Theorem 6 & 7.

⇒ orthogonal matrix is a square invertible matrix

⇒ SQUARE matrixes with orthonormal columns / rows is an orthogonal matrix

EXAMPLE 6 Let $U = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix}$. Notice that U has orthonormal columns and

$$U^T U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 2/3 & -2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Verify that $\|U\mathbf{x}\| = \|\mathbf{x}\|$.

SOLUTION

$$U\mathbf{x} = \begin{bmatrix} 1/\sqrt{2} & 2/3 \\ 1/\sqrt{2} & -2/3 \\ 0 & 1/3 \end{bmatrix} \begin{bmatrix} \sqrt{2} \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$$
$$\|U\mathbf{x}\| = \sqrt{9+1+1} = \sqrt{11}$$
$$\|\mathbf{x}\| = \sqrt{2+9} = \sqrt{11}$$ ■

Theorems 6 and 7 are particularly useful when applied to *square* matrices. An **orthogonal matrix** is a square invertible matrix U such that $U^{-1} = U^T$. By Theorem 6, such a matrix has orthonormal columns.[1] It is easy to see that any *square* matrix with orthonormal columns is an orthogonal matrix. Surprisingly, such a matrix must have orthonormal *rows*, too.

Theorem 6

An $m \times n$ matrix Q has orthonormal columns $\leftrightarrow$ $Q^T Q = I$

⇒ implies that $Q^{-1} = Q^T$

⇒ Q^T is also an orthogonal matrix

Theorem 7

(a) $\| U x \| = \| x \|$

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

(c) $(Ux) \cdot (Uy) = 0$

$\leftrightarrow$ $x \cdot y = 0$

(d) (a) & (c) say that linear mapping $\vec{x} \mapsto U\vec{x}$ preserves lengths and orthogonality

Theorem 8

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 $y = \hat{y} + z$

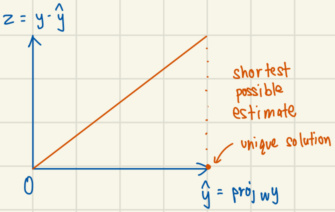
where $\hat{y}$ is in W and z is in $W^\perp$.

if $\{u_1 \dots u_p\}$ is any orthogonal basis of W,

then $$y = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \dots \frac{y \cdot u_p}{u_p \cdot u_p} u_p$$

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

Note : $W^\perp \Rightarrow$ set of all vectors orthogonal to the subspace W



EXAMPLE 2 Let $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Observe that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Write $\mathbf{y}$ as the sum of a vector in W and a vector orthogonal to W.

SOLUTION The orthogonal projection of $\mathbf{y}$ onto W is

sum of a vector in W

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2$$

$$= \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{15}{30}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

Also

Vertical Component

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix} \Rightarrow \text{orthognal to } W$$

Theorem 8 ensures that $\mathbf{y} - \hat{\mathbf{y}}$ is in $W^\perp$. To check the calculations, however, it is a good idea to verify that $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$ and hence to all of W. The desired decomposition of $\mathbf{y}$ is

$$\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

ORTHOGONAL TO EACH OTHER

3

Theorem 10

Note : $p < n$

If $\{u_1 \dots u_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W y = (y \cdot u_1)u_1 + \dots + (y \cdot u_p)u_p$$

If $U = [u_1 \; u_2 \dots u_p]$ then

$$\text{proj}_W y = UU^T y \text{ for all } y \text{ in } \mathbb{R}^n$$

$\Rightarrow$ requires ORTHONORMALITY

$\hookrightarrow$ if not use theorem 8

Theorem 9

Best Approximation Theorem

Let W be a subspace of $\mathbb{R}^n$, let y be any vector 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 in y.

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

for all v in W distinct from $\hat{y}$

$\Rightarrow$ UNIQUE value

* if y is in $W = \text{span}\{u_1 \dots u_p\}$, then $\text{proj}_W y = y$.

Conclusion : if y in the space of W, $\hat{y} = y$ using theorem 8.
else, best approximation according to theorem 9

EXAMPLE 3 If $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, and $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, as in Example 2, then the closest point in W to $\mathbf{y}$ is

orthogonal projection of y onto W

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2}\mathbf{u}_2 = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

closest to y

EXAMPLE 4 The distance from a point $\mathbf{y}$ in $\mathbb{R}^n$ to a subspace W is defined as the distance from $\mathbf{y}$ to the nearest point in W. Find the distance from $\mathbf{y}$ to $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, where

$$\mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

SOLUTION By the Best Approximation Theorem, the distance from $\mathbf{y}$ to W is $\|\mathbf{y} - \hat{\mathbf{y}}\|$, where $\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$. Since $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for W,

$$\hat{\mathbf{y}} = \frac{15}{30}\mathbf{u}_1 + \frac{-21}{6}\mathbf{u}_2 = \frac{1}{2}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{7}{2}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

$$\mathbf{y} - \hat{\mathbf{y}} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix}$$

$$\|\mathbf{y} - \hat{\mathbf{y}}\|^2 = 3^2 + 6^2 = 45$$

The distance from $\mathbf{y}$ to W is $\sqrt{45} = 3\sqrt{5}$.

Finding solution of x for $Ax = b$

$A = QR$

Q : $m \times n$ matrix with orthonormal columns (having same col. Space of A)

R : upper triangular matrix

solution

$$Ax = b$$
$$QRx = b$$
$$Q^T QRx = Q^T b \quad \text{hence}$$
$$Rx = Q^T b$$

Note: if b not in $C(A)$ [no soln], then the found x will only result in the orthogonal projection of b onto $C(A)$

Conversion of normal basis to orthornormal basis

Gram - Schmidt process

Given a basis $\{x_1 \dots x_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$v_1 = x_1$$
$$v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1$$
$$v_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2$$
$$\vdots$$

$\therefore \{v_1 \dots v_p\}$ is an orthogonal basis for W

$\text{Span}\{v_1 \dots v_k\} = \text{Span}\{x_1 \dots x_k\}$ for $1 \leq k \leq p$

linear combination of each other

Theorem 12

The QR Factorization

A : $m \times n$ matrix with linearly independent columns

$\Rightarrow$ can be factored as $A = QR$

if $m = n$, Q orthogonal ie. $Q^T Q = QQ^T = I$

orthonormal

Q : $m \times n$ matrix whose columns form an orthonormal basis for Col A

R : $n \times n$ upper triangular invertible matrix with positive entries on its diagonal

$\Rightarrow$ If $A \in \mathbb{R}^{m \times n}$ has full column rank, ie A has linearly independent columns.

MUST!!

$A = QR \Rightarrow$ Col vectors of A linearly independent!

EXAMPLE 1 Let $W = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$, where $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for W.

SOLUTION The subspace W is shown in Fig. 1, along with $\mathbf{x}_1$, $\mathbf{x}_2$, and the projection $\mathbf{p}$ of $\mathbf{x}_2$ onto $\mathbf{x}_1$. The component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$ is $\mathbf{x}_2 - \mathbf{p}$, which is in W because it is formed from $\mathbf{x}_2$ and a multiple of $\mathbf{x}_1$. Let $\mathbf{v}_1 = \mathbf{x}_1$ and

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{p} = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{x}_1}{\mathbf{x}_1 \cdot \mathbf{x}_1}\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{15}{45}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

Then $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal set of nonzero vectors in W. Since $\dim W = 2$, the set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is a basis for W. ■

EXAMPLE 2 Let $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$. Then $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is clearly linearly independent and thus is a basis for a subspace W of $\mathbb{R}^4$. Construct an orthogonal basis for W.

SOLUTION

Step 1. Let $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \text{Span}\{\mathbf{x}_1\} = \text{Span}\{\mathbf{v}_1\}$.

Step 2. Let $\mathbf{v}_2$ be the vector produced by subtracting from $\mathbf{x}_2$ its projection onto the subspace W_1. That is, let

$$\begin{aligned} \mathbf{v}_2 &= \mathbf{x}_2 - \text{proj}_{W_1} \mathbf{x}_2 \\ &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \qquad \text{Since } \mathbf{v}_1 = \mathbf{x}_1 \\ &= \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} \end{aligned}$$

As in Example 1, $\mathbf{v}_2$ is the component of $\mathbf{x}_2$ orthogonal to $\mathbf{x}_1$, and $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for the subspace W_2 spanned by $\mathbf{x}_1$ and $\mathbf{x}_2$.

Note: $v_2' = v_2 * 4$ to get rid of denominator in v_2

Step 3. Let $\mathbf{v}_3$ be the vector produced by subtracting from $\mathbf{x}_3$ its projection onto the subspace W_2. Use the orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2'\}$ to compute this projection onto W_2:

$$\text{proj}_{W_2} \mathbf{x}_3 = \underbrace{\frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1}_{\text{Projection of } \mathbf{x}_3 \text{ onto } \mathbf{v}_1} + \underbrace{\frac{\mathbf{x}_3 \cdot \mathbf{v}_2'}{\mathbf{v}_2' \cdot \mathbf{v}_2'}\mathbf{v}_2'}_{\text{Projection of } \mathbf{x}_3 \text{ onto } \mathbf{v}_2'} = \frac{2}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \frac{2}{12}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

Then $\mathbf{v}_3$ is the component of $\mathbf{x}_3$ orthogonal to W_2, namely,

$$\mathbf{v}_3 = \mathbf{x}_3 - \text{proj}_{W_2} \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

Gram–Schmidt Process for QR Factorisation- IMPT-need L I
⇒ linearly independent

Example [edit]

Consider the decomposition of

$$A = \begin{pmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{pmatrix}.$$

Recall that an orthonormal matrix Q has the property

$$Q^T Q = I.$$

Then, we can calculate Q by means of Gram–Schmidt as follows:

$$U = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \mathbf{u}_3) = \begin{pmatrix} 12 & -69 & -58/5 \\ 6 & 158 & 6/5 \\ -4 & 30 & -33 \end{pmatrix};$$

$$Q = \left(\frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} \quad \frac{\mathbf{u}_2}{\|\mathbf{u}_2\|} \quad \frac{\mathbf{u}_3}{\|\mathbf{u}_3\|} \right) = \begin{pmatrix} 6/7 & -69/175 & -58/175 \\ 3/7 & 158/175 & 6/175 \\ -2/7 & 6/35 & -33/35 \end{pmatrix}$$

not a yet, need to divide

because e

only if vectors inside column vectors is orthonormal, $A^TQ = I$

Thus, we have

$$Q^T A = Q^T Q\,R = R;$$

$$R = Q^T A = \begin{pmatrix} 14 & 21 & -14 \\ 0 & 175 & -70 \\ 0 & 0 & 35 \end{pmatrix}$$

Hence,

$$A = \begin{pmatrix} 12 & -51 & 4 \\ 6 & 167 & -68 \\ -4 & 24 & -41 \end{pmatrix} =$$

$$\begin{pmatrix} 6/7 & -69/175 & -58/175 \\ 3/7 & 158/175 & 6/175 \\ -2/7 & 6/35 & -33/35 \end{pmatrix} \times \begin{pmatrix} 14 & 21 & -14 \\ 0 & 175 & -70 \\ 0 & 0 & 35 \end{pmatrix}$$

Q: Orthogonal Matrix $\qquad$ R: Upper Triangular Matrix

Least Squares Solution for Inconsistent Eqns

When b not in $C(A)$;
b not a linear combinations of columns of A.
⇒ when rows > cols
⇒ linearly dependent (have free variables)
⇒ Rank (A) < Rank $(A|b)$

Solution:
Find x such that Ax is as close to b as possible
⇒ 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 $\mathbb{R}^n$

When $Ax = b$ is inconsistent, b does not lie in col space of A

Normal Equation

⇒ linked to Least Square Solution

$$Ax = b$$

↓ Multiply both sides by A^T

$$\underbrace{A^T A x = A^T b}_{\text{Normal equation}}$$

Theorem 6.4.1

Best Approximation Theorem
If W is a finite-dimensional subspace of an inner product space V, and if b is a vector in V, then $\text{proj}_W b$ is the best approximation to b from W in the sense that

$$\|b - \text{proj}_W b\| < \|b - w\|$$

for every vector w in W that is different from $\text{proj}_W b$.

Theorem 14

A : $m \times n$ matrix
(a) equation $Ax = b$ has a unique least-square solution for each b in $\mathbb{R}^m$
(b) cols of A are linearly independent
(c) matrix $A^T A$ is invertible
⇒ if (a), (b), (c) == True,
Least-square solution $\hat{x} = (A^T A)^{-1} A^T b$

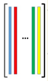
Consistency in a System of Equations

$Ax = b$

Consistent (solutions exist)

- b is in column space of A, i.e, b is formed by linear combinations of A's columns.
- Rank (A) = Rank $(A|b)$, i.e, rank of A is same as that of the augmented matrix.

Inconsistent (solutions don't exist)



$M \gg N$

- b is NOT in column space of A, i.e, b is NOT formed by linear combinations of A's columns.
- Occurs when $M \gg N$ (**over-determined**), i.e, there exist more equations than unknowns.
- The rows of A are dependent but, their corresponding b values are not consistent.
- Rank (A) < Rank $(A|b)$, i.e, rank of A is less than that of the augmented matrix.

Consistency in a System of Equations

A system of equations can be consistent or inconsistent. What does that mean?

A system of equations $Ax = b$ is consistent if there is a solution, and it is inconsistent if there is no solution. However, consistent system of equations does not mean a unique solution, that is, a consistent system of equation may have a unique solution or infinite solutions.

$Ax = b$

Consistent System — if rank(A) = rank$(A|b)$: rank of coefficient matrix = rank of augmented matrix

- Unique Solution — if rank$(A) = N$: rank of coefficient matrix = rank of augmented matrix = no. of unknowns
- Infinite Solutions — if rank$(A) < N$: rank of coefficient matrix = rank of augmented matrix < no. of unknowns

Inconsistent System — if rank(A) < rank$(A|b)$: rank of coefficient matrix < rank of augmented matrix

echelon form

NOTE: Rank (A) is the maximum number of independent rows or columns of A.

You can find number of independent row or columns by:
1. row reduction process
2. rank(A) in MATLAB

Ref: https://en.wikipedia.org/wiki/Augmented_matrix
Ref: https://www.mathsisfun.com/algebra/matrix-rank.html

Note: rank (A) > rank $(A|b)$ is never possible. Why?

Examples

rank(A) = rank$(A|b)$ = N

Consistent and Unique Solution

a) The system of equations

$$\begin{bmatrix}2 & 4\\1 & 3\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}6\\4\end{bmatrix}$$

is a consistent system of equations as it has a unique solution, that is,

$$\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}1\\1\end{bmatrix}.$$

```
>> A_b = [ 2 4 6; 1 3 4]

A_b =

     2     4     6
     1     3     4

>> rank(A_b)

ans =

     2
```

Inconsistent and No solutions Exist

c) The system of equations

$$\begin{bmatrix}2 & 4\\1 & 2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}6\\4\end{bmatrix}$$

```
>> A = [2 4; 1 2]

A =

     2     4
     1     2

>> rank(A)

ans =

     1
```

```
>> A_b = [ 2 4 6; 1 2 4]

A_b =

     2     4     6
     1     2     4

>> rank(A_b)

ans =

     2
```

rank(A) < rank$(A|b)$

Consistent and Having Infinite Solutions

b) The system of equations

$$\begin{bmatrix}2 & 4\\1 & 2\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}6\\3\end{bmatrix}$$

```
>> A_b = [ 2 4 6; 1 2 3]

A_b =

     2     4     6
     1     2     3

>> rank(A_b)

ans =

     1
```

is also a consistent system of equations but it has infinite solutions as given as follows.

Expanding the above set of equations,

$2x + 4y = 6$
$x + 2y = 3$

you can see that they are the same equation. Hence any combination of (x, y) that satisfies

$2x + 4y = 6$

is a solution. For example $(x, y) = (1,1)$ is a solution and other solutions include $(x, y) = (0.5,1.25)$, $(x, y) = (0,\ 1.5)$ and so on.

rank(A) = rank$(A|b)$ < N

no. of equations

Ref: Introduction to Matrix Algebra, Autar Kaw (pg 68)

Examples

b. The columns of A are linearly independent.
c. The matrix A^TA is invertible.

EXAMPLE 1 Find a least-squares solution of the inconsistent system $A\mathbf{x} = \mathbf{b}$ for

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

A & b are linearly independent ⇒ will result in unique soln

SOLUTION To use normal equations (3), compute:

$$A^TA = \begin{bmatrix}4 & 0 & 1\\0 & 2 & 1\end{bmatrix}\begin{bmatrix}4 & 0\\0 & 2\\1 & 1\end{bmatrix} = \begin{bmatrix}17 & 1\\1 & 5\end{bmatrix}$$

$$A^T\mathbf{b} = \begin{bmatrix}4 & 0 & 1\\0 & 2 & 1\end{bmatrix}\begin{bmatrix}2\\0\\11\end{bmatrix} = \begin{bmatrix}19\\11\end{bmatrix}$$

Then the equation $A^TA\mathbf{x} = A^T\mathbf{b}$ becomes

$$\underbrace{\begin{bmatrix}17 & 1\\1 & 5\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix} = \begin{bmatrix}19\\11\end{bmatrix}}_{\text{normal equation}}$$

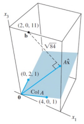


FIGURE 3

Row operations can be used to solve this system, but since A^TA is invertible and 2×2, it is probably faster to compute

$$(A^TA)^{-1} = \frac{1}{84}\begin{bmatrix}5 & -1\\-1 & 17\end{bmatrix}$$

and then to solve $A^TA\mathbf{x} = A^T\mathbf{b}$ as

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} = \frac{1}{84}\begin{bmatrix}5 & -1\\-1 & 17\end{bmatrix}\begin{bmatrix}19\\11\end{bmatrix} = \frac{1}{84}\begin{bmatrix}84\\168\end{bmatrix} = \begin{bmatrix}1\\2\end{bmatrix}$$

In many calculations, A^TA is invertible, but this is not always the case. The next

Lay, Linear Algebra and its Applications (4th Edition)

6.5 Least-Squares Problems 361

Examples

EXAMPLE 3 Given A and $\mathbf{b}$ as in Example 1, determine the least-squares error in the least-squares solution of $A\mathbf{x} = \mathbf{b}$.

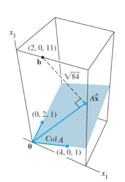


FIGURE 3

SOLUTION From Example 1,

$$\mathbf{b} = \begin{bmatrix}2\\0\\11\end{bmatrix} \quad \text{and} \quad A\hat{\mathbf{x}} = \begin{bmatrix}4 & 0\\0 & 2\\1 & 1\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix} = \begin{bmatrix}4\\4\\3\end{bmatrix}$$

Hence

$$\mathbf{b} - A\hat{\mathbf{x}} = \begin{bmatrix}2\\0\\11\end{bmatrix} - \begin{bmatrix}4\\4\\3\end{bmatrix} = \begin{bmatrix}-2\\-4\\8\end{bmatrix}$$

and

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| = \sqrt{(-2)^2 + (-4)^2 + 8^2} = \sqrt{84}$$

The least-squares error is $\sqrt{84}$. For any $\mathbf{x}$ in $\mathbb{R}^2$, the distance between $\mathbf{b}$ and the vector $A\mathbf{x}$ is at least $\sqrt{84}$. See Fig. 3. Note that the least-squares solution $\hat{\mathbf{x}}$ itself does not appear in the figure.

Examples

EXAMPLE 2 Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

not linearly independent

$$A = \begin{bmatrix}1 & 1 & 0 & 0\\1 & 1 & 0 & 0\\1 & 0 & 1 & 0\\1 & 0 & 1 & 0\\1 & 0 & 0 & 1\\1 & 0 & 0 & 1\end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix}-3\\-1\\0\\2\\5\\1\end{bmatrix}$$

Note the linear dependency in the rows and columns of A:

- Column 1 = Column 2 + Column 3 + Column 4
- Rows 1 & 2 are same, but their corresponding b values are different (inconsistent)
- Rows 3 & 4 are same, but their corresponding b values are different (inconsistent)
- Rows 5 & 6 are same, but their corresponding b values are different (inconsistent)

SOLUTION Compute

$$A^TA = \begin{bmatrix}1 & 1 & 1 & 1 & 1 & 1\\1 & 1 & 0 & 0 & 0 & 0\\0 & 0 & 1 & 1 & 0 & 0\\0 & 0 & 0 & 0 & 1 & 1\end{bmatrix}\begin{bmatrix}1 & 1 & 0 & 0\\1 & 1 & 0 & 0\\1 & 0 & 1 & 0\\1 & 0 & 1 & 0\\1 & 0 & 0 & 1\\1 & 0 & 0 & 1\end{bmatrix} = \underbrace{\begin{bmatrix}6 & 2 & 2 & 2\\2 & 2 & 0 & 0\\2 & 0 & 2 & 0\\2 & 0 & 0 & 2\end{bmatrix}}_{A^TA}$$

$$A^T\mathbf{b} = \begin{bmatrix}1 & 1 & 1 & 1 & 1 & 1\\1 & 1 & 0 & 0 & 0 & 0\\0 & 0 & 1 & 1 & 0 & 0\\0 & 0 & 0 & 0 & 1 & 1\end{bmatrix}\begin{bmatrix}-3\\-1\\0\\2\\5\\1\end{bmatrix} = \begin{bmatrix}4\\-4\\2\\6\end{bmatrix}$$

Note that A^TA is always a square matrix.

Lay, Linear Algebra and its Applications (4th Edition)

The augmented matrix for $A^TA\mathbf{x} = A^T\mathbf{b}$ is

Reduced to

$$\underbrace{\left[\begin{array}{cccc|c}6 & 2 & 2 & 2 & 4\\2 & 2 & 0 & 0 & -4\\2 & 0 & 2 & 0 & 2\\2 & 0 & 0 & 2 & 6\end{array}\right]}_{\text{augmented matrix}} \sim \begin{bmatrix}1 & 0 & 0 & 1 & 3\\0 & 1 & 0 & -1 & -5\\0 & 0 & 1 & -1 & -2\\0 & 0 & 0 & 0 & 0\end{bmatrix}$$

A^TA $\quad$ A^Tb

The general solution is $x_1 = 3 - x_4$, $x_2 = -5 + x_4$, $x_3 = -2 + x_4$, and x_4 is free. So the general least-squares solution of $A\mathbf{x} = \mathbf{b}$ has the form

$$\hat{\mathbf{x}} = \begin{bmatrix}3\\-5\\-2\\0\end{bmatrix} + x_4\begin{bmatrix}-1\\1\\1\\1\end{bmatrix}$$

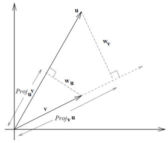
Note: Here, there are infinitely many solutions with the same least square error.
Note: Here, A^TA is not invertible (its determinant is 0).

A^TA may not be invertible if:
- some columns are linearly dependent (i.e. we have redundant features) **(as in this example)**
  - solution: remove the linear dependency
- too many features ($m < n$)
  - solution: delete some features, there are too many features for the amount of data we have

Ref: http://mlwiki.org/index.php/Normal_Equation

Ref: Andrew Ng discussing this phenomenon-
https://www.coursera.org/lecture/machine-learning/normal-equation-noninvertibility-zSiE6

Projection matrix for a space spanned by a single vector v



$$Proj_{\mathbf{v}}\mathbf{u} = \mathbf{v}\left(\frac{\mathbf{u}^T\mathbf{v}}{||\mathbf{v}||^2}\right),$$

Ref: 6.2.2_Orthogonal_Projections

$$\begin{aligned} proj_{\mathbf{v}}\mathbf{u} &= \frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v} \\ &= \mathbf{v}\frac{\mathbf{v}\cdot\mathbf{u}}{\mathbf{v}\cdot\mathbf{v}} \end{aligned}$$

Next, use the transpose definition of the inner product followed by the associative property of multiplication. Remember, when performing the dot product, a scalar multiplier may be placed anywhere you wish.

$$\begin{aligned} proj_{\mathbf{v}}\mathbf{u} &= \frac{1}{\mathbf{v}^T\mathbf{v}}\mathbf{v}(\mathbf{v}^T\mathbf{u}) \\ &= \frac{1}{\mathbf{v}^T\mathbf{v}}(\mathbf{v}\mathbf{v}^T)\mathbf{u} \\ &= \frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}}\mathbf{u} \end{aligned}$$

The expression $\mathbf{v}\mathbf{v}^T$ is called an **outer product** (the transpose operator is outside the product versus its inside position in the inner product). If we define $P = \frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}}$, then the projection formula becomes

$$proj_{\mathbf{v}}\mathbf{u} = P\mathbf{u}, \text{ where } P = \frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}}.$$

The matrix P is called the projection matrix. You can project any vector onto the vector $\mathbf{v}$ by multiplying by the matrix P.

Projection Matrix for col(A) and the Least Squares Solution

Consider solving the system of equations: $Ax = b$

- If b is not in the col(A), we find the least squares solution $\hat{x}$.

  $\hat{x}$ satisfies: $A^TA\hat{x} = A^Tb$

  Thus, $\hat{x} = (A^TA)^{-1}A^Tb$

Qn: **What is projection of b in the column space of A?**

Ans: $\hat{b} = proj_{Col(A)}b = A\hat{x}$

NOTE:

$$\hat{b} = proj_{Col(A)}b = A\hat{x}$$

$$\Rightarrow \hat{b} = proj_{Col(A)}b = A(A^TA)^{-1}A^T\, b$$

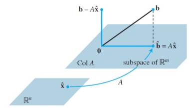


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

$P = A(A^TA)^{-1}A^T$ is the **projection** matrix

The **vector** b can be **projected** into **column space of** A by **multiplying** it with **projection** matrix P.

Projection matrix P maps the actual response values b with predicted values $\hat{b}$.

Ref: https://en.wikipedia.org/wiki/Projection_matrix

Properties of Projection Matrix

When P is multiplied by vector b, the resulting vector $\hat{b} = Pb = A\hat{x}$ is the least squares (nearest) solution in the column space of A.

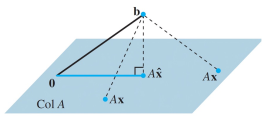


FIGURE 1 The vector $\mathbf{b}$ is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other $\mathbf{x}$.

Properties of Projection Matrix:

1. $P^T = P$ (symmetric)
2. $P^N = P$ [Idempotent Property]

4.3 *Projection matrices.* A matrix $P \in \mathbf{R}^{n\times n}$ is called a *projection matrix* if $P = P^T$ and $P^2 = P$.

(a) Show that if P is a projection matrix then so is $I - P$.

(b) Suppose that the columns of $U \in \mathbf{R}^{n\times k}$ are orthonormal. Show that UU^T is a projection matrix. (Later we will show that the converse is true: every projection matrix can be expressed as UU^T for some U with orthonormal columns.)

(c) Suppose $A \in \mathbf{R}^{n\times k}$ is full rank, with $k \leq n$. Show that $A(A^TA)^{-1}A^T$ is a projection matrix.

(d) If $S \subseteq \mathbf{R}^n$ and $x \in \mathbf{R}^n$, the point y in S closest to x is called the *projection of x on S*. Show that if P is a projection matrix, then $y = Px$ is the projection of x on $\mathcal{R}(P)$. (Which is why such matrices are called projection matrices ...)

Solution:

(a) To show that $I - P$ is a projection matrix we need to check two properties:

i. $I - P = (I-P)^T$ (symmetric)

ii. $(I-P)^2 = I - P$. (Idempotent)

The first one is easy: $(I-P)^T = I - P^T = I - P$ because $P = P^T$ (P is a projection matrix.) The show the second property we have

$$\begin{aligned} (I-P)^2 &= I - 2P + P^2 \\ &= I - 2P + P \qquad \text{(since } P = P^2\text{)} \\ &= I - P \end{aligned}$$

and we are done.

(b) Since the columns of U are orthonormal we have $U^TU = I$. Using this fact it is easy to prove that UU^T is a projection matrix, *i.e.*, $(UU^T)^T = UU^T$ and $(UU^T)^2 = UU^T$. Clearly, $(UU^T)^T = (U^T)^TU^T = UU^T$ and

$$\begin{aligned} (UU^T)^2 &= (UU^T)(UU^T) \\ &= U(U^TU)U^T \\ &= UU^T \qquad \text{(since } U^TU = I\text{).} \end{aligned}$$

(c) First note that $\left(A(A^TA)^{-1}A^T\right)^T = A(A^TA)^{-1}A^T$ because

$$\begin{aligned} \left(A(A^TA)^{-1}A^T\right)^T &= (A^T)^T\left((A^TA)^{-1}\right)^TA^T \\ &= A\left((A^TA)^T\right)^{-1}A^T \\ &= A(A^TA)^{-1}A^T. \end{aligned}$$

Also $\left(A(A^TA)^{-1}A^T\right)^2 = A(A^TA)^{-1}A^T$ because

$$\begin{aligned} \left(A(A^TA)^{-1}A^T\right)^2 &= \left(A(A^TA)^{-1}A^T\right)\left(A(A^TA)^{-1}A^T\right) \\ &= A\left((A^TA)^{-1}A^TA\right)(A^TA)^{-1}A^T \\ &= A(A^TA)^{-1}A^T \qquad \text{(since } (A^TA)^{-1}A^TA = I\text{).} \end{aligned}$$

(d) To show that Px is the projection of x on $\mathcal{R}(P)$ we verify that the "error" $x - Px$ is orthogonal to *any* vector in $\mathcal{R}(P)$. Since $\mathcal{R}(P)$ is nothing but the span of the columns of P we only need to show that $x - Px$ is orthogonal to the columns of P, or in other words, $P^T(x - Px) = 0$. But

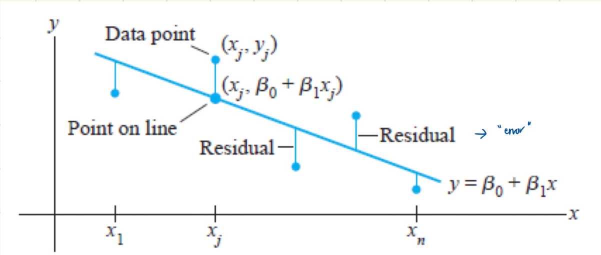
$$\begin{aligned} P^T(x - Px) &= P(x - Px) \qquad \text{(since } P = P^T\text{)} \\ &= Px - P^2x \\ &= 0 \qquad \text{(since } P^2 = P\text{)} \end{aligned}$$

and we are done.

key : Least square solution = solving normal equation

Least Squares Line

Simplest relation between two variables x and y is the linear equation $y = \beta_0 + \beta_1 x$.



Predicted y-value		Observed y-value
$\beta_0 + \beta_1 x_1$	=	y_1
$\beta_0 + \beta_1 x_2$	=	y_2
$\vdots$		$\vdots$
$\beta_0 + \beta_1 x_n$	=	y_n

$$X\beta = y, \quad \text{where } X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

x coordinates ↑   ↑ y coordinates

★ X must have only 2 cols

EXAMPLE 1 Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points $(2, 1)$, $(5, 2)$, $(7, 3)$, and $(8, 3)$.

SOLUTION Use the x-coordinates of the data to build the design matrix X in (1) and the y-coordinates to build the observation vector $\mathbf{y}$:

$$X = \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}$$

For the least-squares solution of $X\beta = \mathbf{y}$, obtain the normal equations (with the new notation):

$$X^T X \beta = X^T \mathbf{y}$$

That is, compute

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}$$

$$X^T \mathbf{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

The normal equations are

$$\begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 9 \\ 57 \end{bmatrix}$$

Hence

$$\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 4 & 22 \\ 22 & 142 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 142 & -22 \\ -22 & 4 \end{bmatrix} \begin{bmatrix} 9 \\ 57 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 24 \\ 30 \end{bmatrix} = \begin{bmatrix} 2/7 \\ 5/14 \end{bmatrix}$$

Thus the least-squares line has the equation

Best least-square line → (only one) $\quad y = \frac{2}{7} + \frac{5}{14} x$

$\beta_0 + \beta_1 x$

Thus the least-squares line has the equation

$$y = \frac{2}{7} + \frac{5}{14} x$$

See Figure 2.

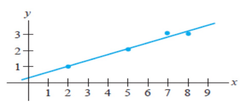


FIGURE 2 The least-squares line $y = \frac{2}{7} + \frac{5}{14} x$.