6. Orthogonality and Least-Squares

6.1 Inner product, length, and orthogonality

❁ Definition

The inner product of two vector $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\boldsymbol{R}^n$ is written $\boldsymbol{u} \cdot \boldsymbol{v}$.

If $\boldsymbol{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$, $\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$, then $\boldsymbol{u} \cdot \boldsymbol{v} = u_1 v_1 + u_2 v_2 + \ldots + u_n v_n$.

❁ Theorem 1

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

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

(c) $(c\boldsymbol{u}) \cdot \boldsymbol{v} = c(\boldsymbol{u} \cdot \boldsymbol{v}) = \boldsymbol{u} \cdot (c\boldsymbol{v})$

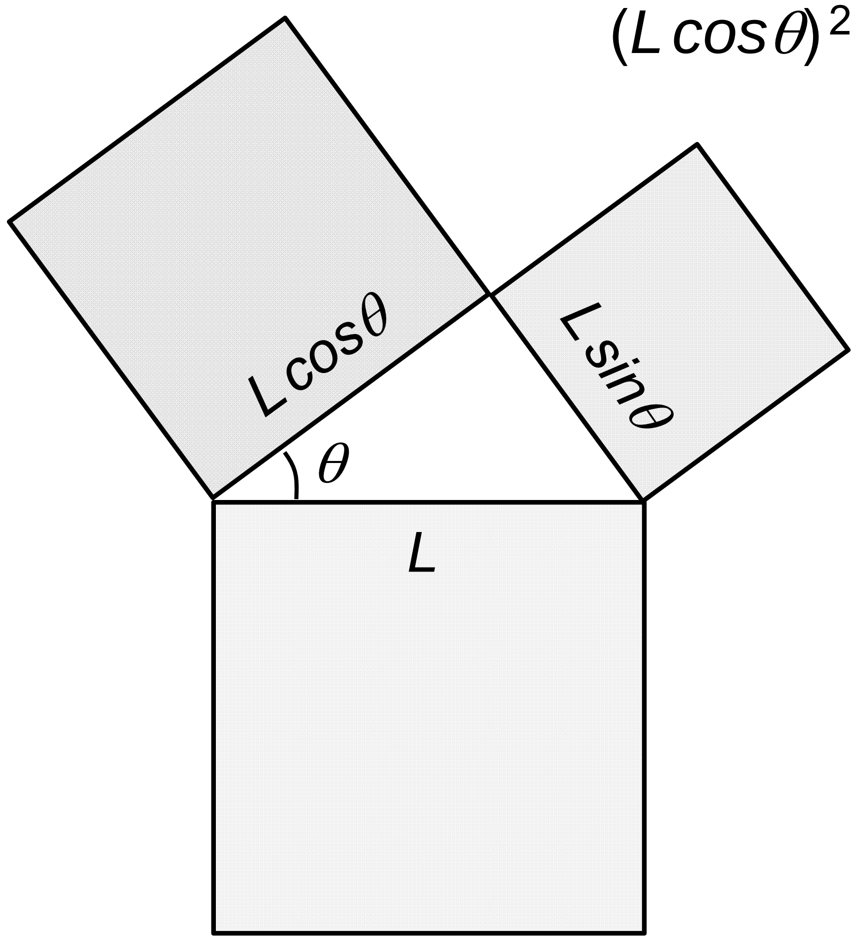
(d) $\boldsymbol{u} \cdot \boldsymbol{u} \geq 0$, and $\boldsymbol{u} \cdot \boldsymbol{u} = 0 \Leftrightarrow \boldsymbol{u} = \boldsymbol{0}$.

- Definition
  The length (or norm) of $\boldsymbol{v}$ is the non-negative scalar $\|\boldsymbol{v}\|$ defined by $\|v\| = \sqrt{\boldsymbol{v}\cdot\boldsymbol{v}} = \sqrt{v_1^2+\cdots+v_n^2}$ and $\|\boldsymbol{v}\|^2 = \boldsymbol{v}\cdot\boldsymbol{v}$.
- Note: $\| c\boldsymbol{v} \| = |c|\, \|\boldsymbol{v}\|$.
- Definition
  The distance between $\boldsymbol{u}$ and $\boldsymbol{v}$, written as dist($\boldsymbol{u}$, $\boldsymbol{v}$), is the length of the vector $\boldsymbol{u} - \boldsymbol{v}$. That is, dist($\boldsymbol{u}$, $\boldsymbol{v}$) = $\| \boldsymbol{u} - \boldsymbol{v} \|$.

Orthogonal vectors

- Definition
  Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\boldsymbol{R}^n$ are orthogonal if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$.
- Theorem 2 (The Pythagorean Theorem, 畢氏定理)
  Two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal if and only if
  $\| \boldsymbol{u} + \boldsymbol{v} \|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$.

$$(L\cos\theta)^2 + (L\sin\theta)^2 = L^2$$

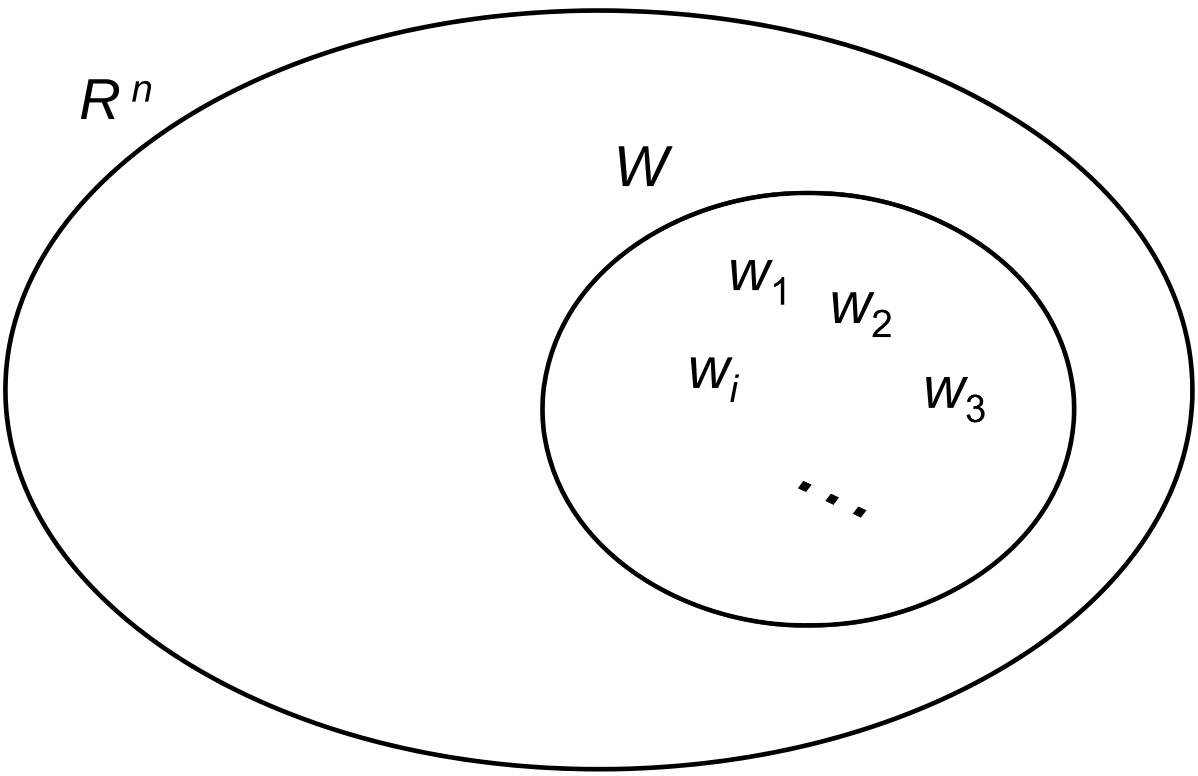


Orthogonal complements

- Definition
  The set of all vectors ***u*** that are orthogonal to every vector ***w*** in ***W***, then we say that the set is the orthogonal complement of ***W*** , and denote by $\boldsymbol{W}^{\perp}$. (next page)

- Note
  1. A vector ***x*** is in $\boldsymbol{W}^{\perp}$ if and only if ***x*** is orthogonal to every vector in a set that spans ***W***.
  2. $\boldsymbol{W}^{\perp}$ is a subspace of $\boldsymbol{R}^n$ if ***W*** is a subspace of $\boldsymbol{R}^n$.
  3. Zero vector is orthogonal to any vector.

- Theorem 3
  Let ***A*** be an $m \times n$ matrix. Then
  (*i*) $(\text{Row } \boldsymbol{A})^{\perp} = \text{Nul } \boldsymbol{A}$.
  (*ii*) $(\text{Col } \boldsymbol{A})^{\perp} = \text{Nul } \boldsymbol{A}^T$. ($\{\boldsymbol{x} \mid \boldsymbol{A}^T\boldsymbol{x} = \boldsymbol{0}\} \equiv \{\boldsymbol{x} \mid \boldsymbol{x}^T\boldsymbol{A} = \boldsymbol{0}\} \equiv$ left null space of ***A***)



$\boldsymbol{W}^{\perp} = \{ x \mid x \in R^n,\ x \perp w_i\text{'s}\}$

- $W^{\perp}$ is a subspace of R^n for any subset W of R^n.

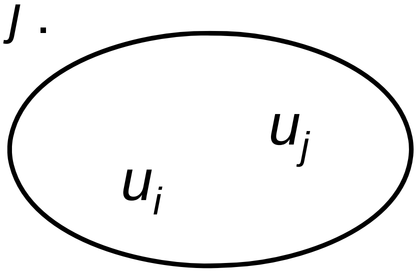
  If w is an arbitrary vector in W, then
  1. $0 \in W^{\perp}$
  2. If $u \in W^{\perp}$,
     then $cu \cdot w = c\,(u \cdot w) = 0$; thus $cu \in W^{\perp}$
  3. If u and $v \in W^{\perp}$,
     then $(u+v) \cdot w = u \cdot w + v \cdot w = 0$;
     thus $(u+v) \in W^{\perp}$

- If W is a subset of R^n, then $(W^{\perp})^{\perp} = W$?

- Exercises of Section 6.1.

6.2 Orthogonal sets

- A set of vectors $\{u_1, \ldots, u_n\}$ in R^n is said to be an orthogonal set if $u_i \cdot u_j = 0$ for all $i \neq j$.



- Theorem 4
  If $S = \{u_1, \ldots, u_p\}$ is an orthogonal set of nonzero vectors in R^n, then S is linearly independent and hence is a basis for the subspace spanned by S.

- An orthogonal basis for a subspace W of R^n is a basis for W that is also an orthogonal set.

- Theorem 5

  Let $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ be an orthogonal basis for a subspace $\boldsymbol{W}$ of $\boldsymbol{R}^n$. Then each $\boldsymbol{y}$ in $\boldsymbol{W}$ has a unique representation as a linear combination of $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$. In fact, if

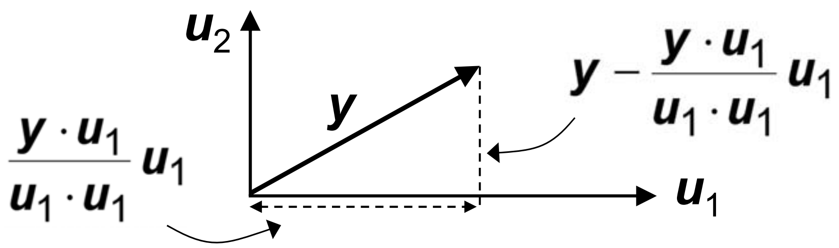
  $$\boldsymbol{y} = c_1\boldsymbol{u}_1 + \ldots + c_p\boldsymbol{u}_p$$

  then $c_j = \dfrac{\boldsymbol{y}\cdot\boldsymbol{u}_j}{\boldsymbol{u}_j\cdot\boldsymbol{u}_j}$, $j = 1, 2, \ldots, p$.

- Note 投影分量

  (1) $\boldsymbol{y}_j = c_j\,\boldsymbol{u}_j = \dfrac{\boldsymbol{y}\cdot\boldsymbol{u}_j}{\boldsymbol{u}_j\cdot\boldsymbol{u}_j}\boldsymbol{u}_j$ is the orthogonal projection of $\boldsymbol{y}$ onto $\boldsymbol{u}_j$.

  (2) $\boldsymbol{y} - \boldsymbol{y}_j = \boldsymbol{y} - \dfrac{\boldsymbol{y}\cdot\boldsymbol{u}_j}{\boldsymbol{u}_j\cdot\boldsymbol{u}_j}\boldsymbol{u}_j$ is component of $\boldsymbol{y}$ orthogonal to $\boldsymbol{u}_j$.



- A set $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p\}$ is an orthonormal set if it is an orthogonal set of unit vectors.

- Normalize the length of vector $\boldsymbol{u} = [u_1 \ldots u_n]^T$ to $\boldsymbol{u}'$

  $$\boldsymbol{u}' = \frac{1}{\sqrt{u_1^2 + \ldots + u_n^2}}[u_1 \ \cdots \ u_n]^T.$$

- Theorem 6

  An $m\times n$ matrix $\boldsymbol{U}$ has orthonormal columns if and only if $\boldsymbol{U}^T\boldsymbol{U} = \mathrm{I}$.

  Proof.

  $$\boldsymbol{U}^T\boldsymbol{U} = \begin{bmatrix}\boldsymbol{u}_1^T \\ \vdots \\ \boldsymbol{u}_n^T\end{bmatrix}[\boldsymbol{u}_1 \cdots \ \boldsymbol{u}_n] = \begin{bmatrix}\boldsymbol{u}_1^T\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_1^T\boldsymbol{u}_n \\ \vdots & \cdots & \vdots \\ \boldsymbol{u}_n^T\boldsymbol{u}_1 & \cdots & \boldsymbol{u}_n^T\boldsymbol{u}_n\end{bmatrix}.$$

  內積

- Theorem 7

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

(*a*) $\|\boldsymbol{Ux}\| = \|\boldsymbol{x}\|$ (preserving length)

(*b*) $(\boldsymbol{Ux}) \cdot (\boldsymbol{Uy}) = \boldsymbol{x} \cdot \boldsymbol{y}$ (preserving orthogonality)

(*c*) $(\boldsymbol{Ux}) \cdot (\boldsymbol{Uy}) = 0$ if and only if $\boldsymbol{x} \cdot \boldsymbol{y} = 0$

Proof.

(*a*) $$\boldsymbol{U}\boldsymbol{x} = [\boldsymbol{u}_1 \ \boldsymbol{u}_2 \ \dots \ \boldsymbol{u}_n]\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\boldsymbol{u}_1 + x_2\boldsymbol{u}_2 + \dots + x_n\boldsymbol{u}_n$$

$$\begin{aligned}
\|\boldsymbol{U}\boldsymbol{x}\| &= \sqrt{(\boldsymbol{U}\boldsymbol{x})\cdot(\boldsymbol{U}\boldsymbol{x})} \\
&= \sqrt{(x_1\boldsymbol{u}_1 + .. + x_n\boldsymbol{u}_n)\cdot(x_1\boldsymbol{u}_1 + .. + x_n\boldsymbol{u}_n)} \\
&= \sqrt{x_1\boldsymbol{u}_1\cdot(x_1\boldsymbol{u}_1 + .. + x_n\boldsymbol{u}_n) + .. + x_n\boldsymbol{u}_n\cdot(x_1\boldsymbol{u}_1 + .. + x_n\boldsymbol{u}_n)} \\
&= \sqrt{x_1\boldsymbol{u}_1\cdot(x_1\boldsymbol{u}_1 + .. + x_n\boldsymbol{u}_n) + .. + x_n\boldsymbol{u}_n\cdot(x_1\boldsymbol{u}_1 + .. + x_n\boldsymbol{u}_n)} \\
&= \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} = \sqrt{\boldsymbol{x}\cdot\boldsymbol{x}} = \|\boldsymbol{x}\|.
\end{aligned}$$

(*b*) $(\boldsymbol{Ux}) \cdot (\boldsymbol{Uy})$

$= (x_1\boldsymbol{u}_1 + \dots + x_n\boldsymbol{u}_n) \cdot (y_1\boldsymbol{u}_1 + \dots + y_n\boldsymbol{u}_n)$

$= x_1 y_1 + x_2 y_2 + \dots + x_n y_n$

- Exercises of Section 6.2.

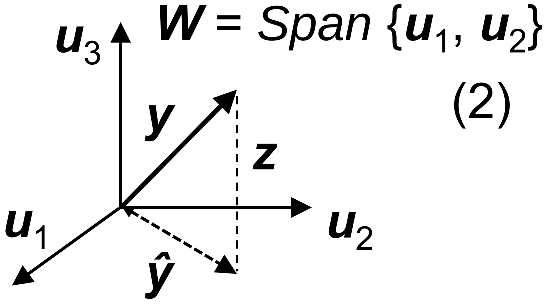
6.3 Orthogonal projections

- Purpose
  To find the projection of a vector on a subspace.

Theorem 8 (The Orthogonal Decomposition Theorem)
Let $\boldsymbol{W}$ be a subspace of $\boldsymbol{R}^n$. Then each $\boldsymbol{y}$ in $\boldsymbol{R}^n$ can be written uniquely in the form $\boldsymbol{y} = \hat{\boldsymbol{y}} + \boldsymbol{z}$, (1)
where $\hat{\boldsymbol{y}}$ is in $\boldsymbol{W}$ and $\boldsymbol{z}$ is in $\boldsymbol{W}^{\perp}$. In fact, if $\{ \boldsymbol{u}_1, \ldots, \boldsymbol{u}_p \}$ is any orthogonal basis of $\boldsymbol{W}$, then

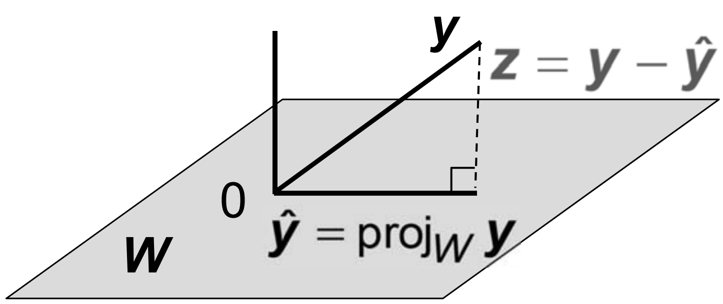
$$\hat{\boldsymbol{y}} = \frac{\boldsymbol{y} \cdot \boldsymbol{u}_1}{\boldsymbol{u}_1 \cdot \boldsymbol{u}_1}\boldsymbol{u}_1 + \ldots + \frac{\boldsymbol{y} \cdot \boldsymbol{u}_p}{\boldsymbol{u}_p \cdot \boldsymbol{u}_p}\boldsymbol{u}_p \qquad (2)$$

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



- The vector $\hat{\boldsymbol{y}}$ in Eq.(1) is called the orthogonal projection of $\boldsymbol{y}$ onto $\boldsymbol{W}$ and is often written as $proj_w\ \boldsymbol{y}$.

- For example,
  The orthogonal projection of $\boldsymbol{y}$ onto $\boldsymbol{W}$.



- Ex.2.

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

Solution.

The orthogonal projection of ***y*** onto ***W*** is

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

$$= \frac{9}{30}\begin{bmatrix}2\\5\\-1\end{bmatrix} + \frac{3}{6}\begin{bmatrix}-2\\1\\1\end{bmatrix} = \frac{9}{30}\begin{bmatrix}2\\5\\-1\end{bmatrix} + \frac{15}{30}\begin{bmatrix}-2\\1\\1\end{bmatrix} = \frac{1}{30}\begin{bmatrix}-12\\60\\6\end{bmatrix} = \begin{bmatrix}-2/5\\2\\1/5\end{bmatrix}$$

So

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

The desired decomposition of ***y*** is

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

Properties of orthogonal projections

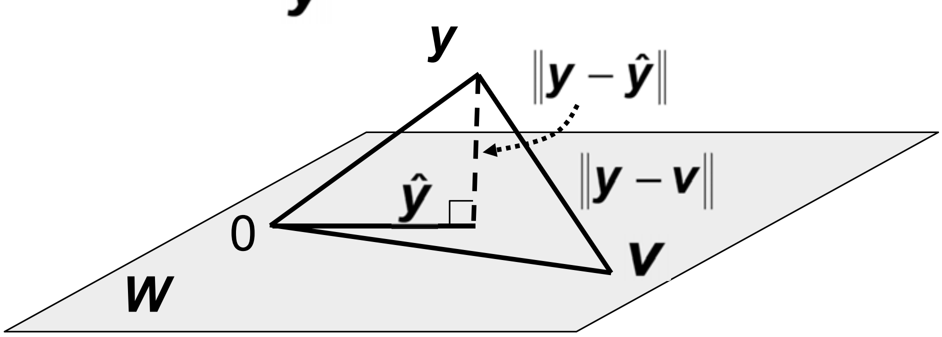
- If $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots \boldsymbol{u}_p\}$ is an orthogonal basis for ***W*** and if ***y*** is in $\boldsymbol{W} = \text{Span}\{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_p\}$, then $\text{proj}_{\boldsymbol{W}}\,\boldsymbol{y} = \boldsymbol{y}$.

Theorem 9 (The Best approximation theorem)

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

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

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



- Ex.4. The distance from a point $\boldsymbol{y}$ in $\boldsymbol{R}^n$ to a subspace $\boldsymbol{W}$ is defined as the distance from $\boldsymbol{y}$ to the nearest point in $\boldsymbol{W}$. Find the distance from y to

$\boldsymbol{W}$ = Span $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$, where $\boldsymbol{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}$, $\boldsymbol{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}$, $\boldsymbol{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$.

Solution.

By the best approximation theorem, the distance from $\boldsymbol{y}$ to $\boldsymbol{W}$ is $\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|$, where $\hat{\boldsymbol{y}} = \text{proj}_{\boldsymbol{W}}\, \boldsymbol{y}$. Since $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ is orthogonal basis for $\boldsymbol{W}$,

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

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

$\|\boldsymbol{y} - \hat{\boldsymbol{y}}\|^2 = 3^2 + 6^2 = 45.$ The distance from $\boldsymbol{y}$ to $\boldsymbol{W}$ is $\sqrt{45} = 3\sqrt{5}$.

Theorem 10

If $\{\boldsymbol{u}_1, .., \boldsymbol{u}_p\}$ is an orthonormal basis for a subspace W of $\boldsymbol{R}^n$, then

$$\text{proj}_W\, \boldsymbol{y} = (\boldsymbol{y} \cdot \boldsymbol{u}_1)\, \boldsymbol{u}_1 + (\boldsymbol{y} \cdot \boldsymbol{u}_2)\, \boldsymbol{u}_2 + \ldots + (\boldsymbol{y} \cdot \boldsymbol{u}_p)\, \boldsymbol{u}_p \qquad (4)$$

If $\boldsymbol{U} = [\boldsymbol{u}_1\ \boldsymbol{u}_2 \ldots \boldsymbol{u}_p]$, then

$$\text{proj}_W\, \boldsymbol{y} = \boldsymbol{U}\,\boldsymbol{U}^T \boldsymbol{y} \text{ for all } \boldsymbol{y} \text{ in } \boldsymbol{R}^n. \qquad (5)$$

Proof.

$$\text{proj}_W\, \boldsymbol{y} = (\boldsymbol{y} \cdot \boldsymbol{u}_1)\, \boldsymbol{u}_1 + (\boldsymbol{y} \cdot \boldsymbol{u}_2)\, \boldsymbol{u}_2 + \ldots + (\boldsymbol{y} \cdot \boldsymbol{u}_p)\, \boldsymbol{u}_p = \boldsymbol{U}\begin{bmatrix} \boldsymbol{y}\cdot\boldsymbol{u}_1 \\ \boldsymbol{y}\cdot\boldsymbol{u}_2 \\ \vdots \\ \boldsymbol{y}\cdot\boldsymbol{u}_p \end{bmatrix}$$

Since $\boldsymbol{y}\cdot\boldsymbol{u}_i = \boldsymbol{u}_i\cdot\boldsymbol{y} = \boldsymbol{u}_i^T \boldsymbol{y}$

$$\text{proj}_W\, \boldsymbol{y} = \boldsymbol{U}\begin{bmatrix} \boldsymbol{u}_1^T\boldsymbol{y} \\ \boldsymbol{u}_2^T\boldsymbol{y} \\ \vdots \\ \boldsymbol{u}_p^T\boldsymbol{y} \end{bmatrix} = \boldsymbol{U}\,(\boldsymbol{U}^T \boldsymbol{y}) = \boldsymbol{U}\,\boldsymbol{U}^T \boldsymbol{y}$$

- Exercises of Section 6.3.

6.4 The Gram - Schmidt process

- Purpose

  The Gram - Schmidt process is a simple algorithm for producing an orthogonal or orthonormal basis for nonzero subspace of $\boldsymbol{R}^n$.

- Theorem 11 (The Gram - Schmidt process)

  Let a basis { $\boldsymbol{x}_1$, … , $\boldsymbol{x}_p$ } for a subspace $\boldsymbol{W}$ of $\boldsymbol{R}^n$, define

$$\boldsymbol{v}_1 = \boldsymbol{x}_1,$$

$$\boldsymbol{v}_2 = \boldsymbol{x}_2 - \frac{\boldsymbol{x}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1,$$

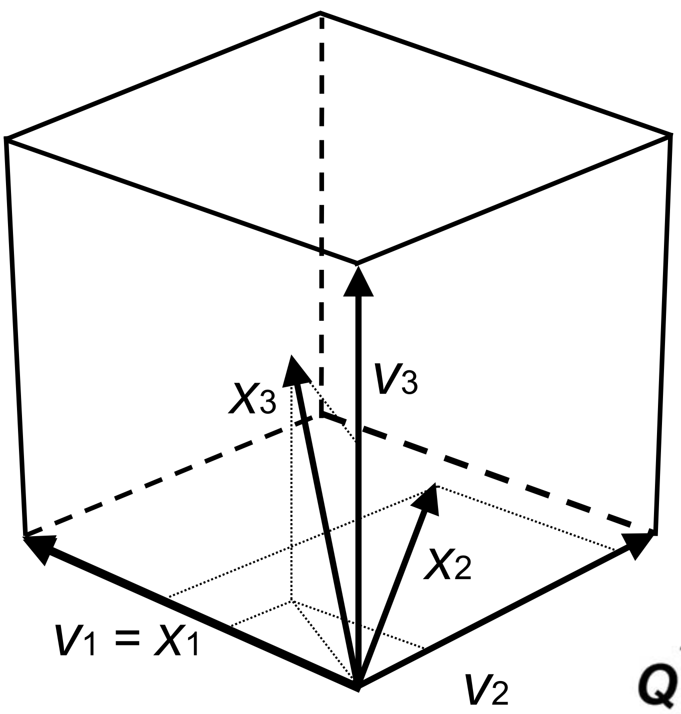
$$\boldsymbol{v}_3 = \boldsymbol{x}_3 - \frac{\boldsymbol{x}_3 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \frac{\boldsymbol{x}_3 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2, \quad ..$$

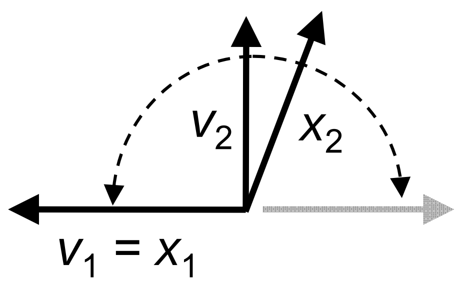
$$\boldsymbol{v}_p = \boldsymbol{x}_p - \frac{\boldsymbol{x}_p \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 - \;\; .... \;\; - \frac{\boldsymbol{x}_p \cdot \boldsymbol{v}_{p-1}}{\boldsymbol{v}_{p-1} \cdot \boldsymbol{v}_{p-1}}\boldsymbol{v}_{p-1}.$$

Non-zero linear independent set
⇓
Orthogonal set

  Then { $\boldsymbol{v}_1$, … , $\boldsymbol{v}_p$ } is an orthogonal basis for $\boldsymbol{W}$. In addition *Span* { $\boldsymbol{v}_1$, … , $\boldsymbol{v}_k$ } = *Span* { $\boldsymbol{x}_1$, … , $\boldsymbol{x}_k$ } for $1 \le k \le p$.

- The geometric meaning of the Gram - Schmidt process (A three-dimensional case)



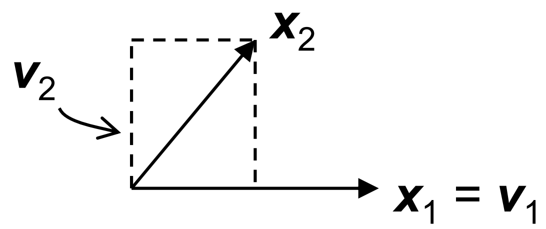


$$\boldsymbol{Q}^T\boldsymbol{A} = \begin{bmatrix} - & \boldsymbol{v}_1 & - \\ - & \boldsymbol{v}_2 & - \\ - & \boldsymbol{v}_3 & - \end{bmatrix} \begin{bmatrix} | & | & | \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 \\ | & | & | \end{bmatrix}$$

Proof.

$$\boldsymbol{x}_2 = \frac{\boldsymbol{x}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 + \frac{\boldsymbol{x}_2 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2$$

$$\Rightarrow \frac{\boldsymbol{x}_2 \cdot \boldsymbol{v}_2}{\boldsymbol{v}_2 \cdot \boldsymbol{v}_2}\boldsymbol{v}_2 = \boldsymbol{x}_2 - \frac{\boldsymbol{x}_2 \cdot \boldsymbol{v}_1}{\boldsymbol{v}_1 \cdot \boldsymbol{v}_1}\boldsymbol{v}_1 .$$



Orthonormal bases

- An orthonormal basis is constructed easily from an orthogonal basis { $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p$ }; simply normalize all $\boldsymbol{v}_k$ by $\boldsymbol{v}_k / \|\boldsymbol{v}_k\| \rightarrow \boldsymbol{v}_k$.

QR factorization of matrices

- This factorization is widely used in computer algorithm for various computations, such as solving equations and finding eigenvalues. (Exercise 23 of Section 5.2.)

- Theorem 12 **(The *QR* factorization)**
  If $\boldsymbol{A}$ is an $m \times n$ matrix with linearly independent columns, then $\boldsymbol{A}$ can be factored as $\boldsymbol{A} = \boldsymbol{QR}$, where $\boldsymbol{Q}$ is an $m \times n$ matrix whose columns form an orthonormal basis for Col $\boldsymbol{A}$ and $\boldsymbol{R}$ is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

- Ex.4.

  Factor $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ into $\boldsymbol{QR}$.

  Answer.

  (*a*) Find $\boldsymbol{Q}$. $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \right\} \xrightarrow{\text{orthogonalize}} \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix} \right\}$

$$\xrightarrow{\text{orthonormalize}} \left\{ \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} -3/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \\ 1/\sqrt{12} \end{bmatrix}, \begin{bmatrix} 0 \\ -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix} \right\} \Rightarrow \boldsymbol{Q} = \begin{bmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{bmatrix}.$$

(b) To find $\boldsymbol{R}$.

Since $\boldsymbol{A} = \boldsymbol{QR}$ and $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \mathrm{I}$ (Theorem 6 of page 10)

$\Rightarrow \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q}\,\boldsymbol{R} \Rightarrow \boldsymbol{Q}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{R}$.

$$\Rightarrow \boldsymbol{R} = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} \\ 0 & -2/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 3/2 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 2/\sqrt{6} \end{bmatrix}.$$

❁ Exercises of Section 6.4.

❁ Problem

$$\boldsymbol{A} = \boldsymbol{QR} \text{ and } \boldsymbol{Q}^T\boldsymbol{Q} = \mathrm{I}$$
$$\Downarrow$$
$$\boldsymbol{Q}^T\boldsymbol{A} = \boldsymbol{Q}^T\boldsymbol{QR} \Rightarrow \boldsymbol{Q}^T\boldsymbol{A} = \boldsymbol{R}$$
$$\Downarrow$$
$$\boldsymbol{QQ}^T\boldsymbol{A} = \boldsymbol{QR} = \boldsymbol{A}$$
$$\Downarrow$$
$$\boldsymbol{QQ}^T = \boldsymbol{I} \ ?$$

No !

$$\boldsymbol{Q} = \begin{bmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{bmatrix} \Rightarrow \boldsymbol{QQ}^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

$$\Rightarrow \boldsymbol{QQ}^T\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = \boldsymbol{A}.$$

6.5 Least-squares problems

- For a linear system, $\boldsymbol{Ax} = \boldsymbol{b}$, a solution is demanded but none exists. The best one can do is to find an $\boldsymbol{x}$ that makes $\boldsymbol{Ax}$ as close as possible to $\boldsymbol{b}$. The general least-squares problem is just to find an $\boldsymbol{x}$ that makes $||\boldsymbol{b} - \boldsymbol{Ax}||$ as small as possible. The term least-squares arises from the fact that $||\boldsymbol{b} - \boldsymbol{Ax}||$ is the square root of a sum of squares.

- Definition

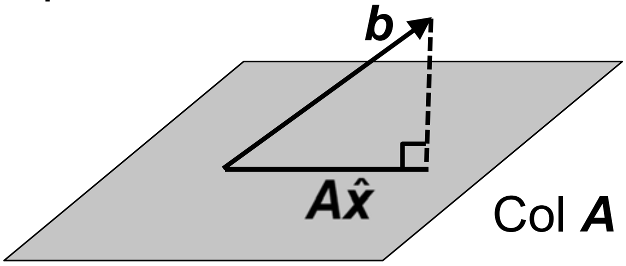
  If $\boldsymbol{A}$ is an $m\times n$ matrix and $\boldsymbol{b}$ is in $\boldsymbol{R}^m$, a least-squares solution of $\boldsymbol{Ax} = \boldsymbol{b}$ is an $\hat{\boldsymbol{x}}$ in $\boldsymbol{R}^n$ such that

  $$||\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}|| \leq ||\boldsymbol{b} - \boldsymbol{Ax}|| \text{ for all } \boldsymbol{x} \text{ in } \boldsymbol{R}^n.$$

Theorem 9 on page 16

- Note

  No matter that $\boldsymbol{x}$ is selected, the vector $\boldsymbol{Ax}$ will necessarily be in the column space Col $\boldsymbol{A}$. So we seek an $\boldsymbol{x}$ that makes $\boldsymbol{Ax}$ the closest point in Col $\boldsymbol{A}$ to $\boldsymbol{b}$.



If $\hat{\boldsymbol{b}}$ is the orthogonal projection of $\boldsymbol{b}$ on Col $\boldsymbol{A}$, then $\boldsymbol{A}\hat{\boldsymbol{x}} = \hat{\boldsymbol{b}}$. It means that $\boldsymbol{Ax} = \hat{\boldsymbol{b}}$ is consistent and there is a solution $\hat{\boldsymbol{x}}$ in $\boldsymbol{R}^n$. By the orthogonal decomposition principle, the projection $\hat{\boldsymbol{b}}$ has the property that $\boldsymbol{b} - \hat{\boldsymbol{b}}$ is orthogonal to Col $\boldsymbol{A}$, so $\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}$ is orthogonal to each column of $\boldsymbol{A}$. If $\boldsymbol{a}_j$ is any column of $\boldsymbol{A}$, then $\boldsymbol{a}_j$ is orthogonal to $(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}})$ and $\boldsymbol{a}_j^T(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}) = 0$. Since each $\boldsymbol{a}_j^T$ is a row of $\boldsymbol{A}^T$,
$\boldsymbol{A}^T(\boldsymbol{b} - \boldsymbol{A}\hat{\boldsymbol{x}}) = \boldsymbol{0} \Rightarrow \boldsymbol{A}^T\boldsymbol{b} - \boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{0} \Rightarrow \boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$.

- Theorem 13
  The set of least-squares solutions of $\boldsymbol{Ax} = \boldsymbol{b}$ coincides with the nonempty set of solutions of the normal equations $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$. If $\boldsymbol{A}^T\boldsymbol{A}$ is invertible, then $\hat{\boldsymbol{x}} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}$.

- Ex.1.
  Find a least-squares solution of the inconsistent system

$$\boldsymbol{Ax} = \boldsymbol{b} \text{ for } \boldsymbol{A} = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \text{ and } \boldsymbol{b} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}.$$

Solution. $$\hat{\boldsymbol{x}} = \left(\begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$
$$= \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

- In many calculations, $\boldsymbol{A}^T\boldsymbol{A}$ is invertible, but this is not always the case.

- Ex.2. Find a least-squares solution of $\boldsymbol{Ax} = \boldsymbol{b}$ for

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \text{ and } \boldsymbol{b} = \begin{bmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{bmatrix}.$$

where $\boldsymbol{A}^T\boldsymbol{A}$ is not invertible.

Solution. $$\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b} \Rightarrow \begin{bmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{bmatrix} \hat{\boldsymbol{x}} = \begin{bmatrix} 4 \\ -4 \\ 2 \\ 6 \end{bmatrix}$$

The augmented matrix $$\begin{bmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{bmatrix} \rightarrow \dots \rightarrow \begin{bmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \Rightarrow \begin{cases} x_1 + x_4 = 3 \\ x_2 - x_4 - 5 \\ x_3 - x_4 = -2 \end{cases}.$$

Solution $$\hat{\boldsymbol{x}} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 - x_4 \\ -5 + x_4 \\ -2 + x_4 \\ x_4 \end{bmatrix} = \begin{bmatrix} 3 \\ -5 \\ -2 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

- Theorem 14
  The matrix A^TA is invertible if and only if the columns of A are linearly independent. In this case, the equation $Ax = b$ has only one least-squares solution $\hat{x}$, and it is given by $\hat{x} = (A^TA)^{-1} A^Tb$.

- Note
  The distance from b to $A\hat{x}$ is called the least-squares error.

Alternative calculations of least-square solutions

- Meaning
  In some cases, the normal equations for a least-squares problem can be ill-conditioned; that is, small errors in the calculations of the entries of A^TA can sometimes cause relatively large errors in the solution $\hat{x}$. If the columns of A are linearly independent, the least-squares solution can often be computed more reliably through a ***QR*** factorization of A .

- ill-conditioned
  The condition number associated with the linear equation $Ax = b$ gives a bound on how inaccurate the solution x will be after approximate solution.

  Conditioning is a property of the matrix, not the algorithm or floating point accuracy of the computer used to solve the corresponding system.

  The rate at which the solution x will change with respect to a change in b. Thus, if the condition number is large, even a small error in b may cause a large error in x.

The condition number is the maximum ratio of the relative error in ***x*** divided by the relative error in ***b***,

$$\frac{\left\|\boldsymbol{A}^{-1}\mathbf{e}\right\| / \left\|\boldsymbol{A}^{-1}\boldsymbol{b}\right\|}{\left\|\mathbf{e}\right\| / \left\|\boldsymbol{b}\right\|} = \left\|\boldsymbol{A}\right\| \left\|\boldsymbol{A}^{-1}\right\|$$

, where $\|\boldsymbol{x}\|$ is 2 norm.

- For example,

$$\begin{bmatrix} 1 & 1 \\ 1 & 1.001 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 1 & 1.001 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2 \\ 2.001 \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

where condition number is

$$\left\| \begin{bmatrix} 1 & 1 \\ 1 & 1.001 \end{bmatrix} \right\| \left\| \begin{bmatrix} 1001 & -1000 \\ -1000 & 1000 \end{bmatrix} \right\| = \sqrt{4.002001}\sqrt{4002001} = 4002$$

- The larger condition number is, the more ill conditioned the coefficient matrix is.

- Theorem 15
  Given an $m \times n$ matrix $\boldsymbol{A}$ with linearly independent columns, let $\boldsymbol{A} = \boldsymbol{QR}$ be a $\boldsymbol{QR}$ factorization of $\boldsymbol{A}$. Then for each $\boldsymbol{b}$ in $\boldsymbol{R}^m$, the equation $\boldsymbol{Ax} = \boldsymbol{b}$ has a unique least-squares solution,
  $\hat{\boldsymbol{x}} = \boldsymbol{R}^{-1}\boldsymbol{Q}^T\boldsymbol{b}$.

  Proof.

$$\begin{aligned} \hat{\boldsymbol{x}} &= (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b} \quad \text{since } \boldsymbol{A} = \boldsymbol{QR} \\ &= (\boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{QR})^{-1}\boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{b} \\ &= (\boldsymbol{R}^T\boldsymbol{I}\boldsymbol{R})^{-1}\boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{b} \\ &= (\boldsymbol{R}^T\boldsymbol{R})^{-1}\boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{b} \\ &= \boldsymbol{R}^{-1}(\boldsymbol{R}^T)^{-1}\boldsymbol{R}^T\boldsymbol{Q}^T\boldsymbol{b} \\ &= \boldsymbol{R}^{-1}\boldsymbol{Q}^T\boldsymbol{b}. \end{aligned}$$

- Note

Since $\boldsymbol{R}$ is an upper triangular matrix, $\hat{\boldsymbol{x}}$ should be calculated from the equation

$$\boldsymbol{R}\hat{\boldsymbol{x}} = \boldsymbol{Q}^T\boldsymbol{b}.$$

It is much faster to solve this equation by back-substitution or row operations than to compute $\boldsymbol{R}^{-1}$ and use the preceding equation.

- Ex.5.

Find the least-squares solution of $\boldsymbol{Ax} = \boldsymbol{b}$ for

$$\boldsymbol{A} = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix} \text{ and } \boldsymbol{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}.$$

Solution.

$$\boldsymbol{A} = \boldsymbol{QR} = \begin{bmatrix} 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & -1/2 \\ 1/2 & -1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}.$$

$$\boldsymbol{R}\hat{\boldsymbol{x}} = \boldsymbol{Q}^T\boldsymbol{b} = \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \\ 4 \end{bmatrix}$$

$$\Rightarrow \hat{\boldsymbol{x}} = \begin{bmatrix} 10 \\ -6 \\ 2 \end{bmatrix}.$$

- Exercises for Section 6.5.