Section 6.3 Orthogonal Projections

ELOs:

- Uniquely write a vector in $\mathbb{R}^n$ as a linear combination of vectors in W and $W^\perp$.
- Apply the Best Approximation Theorem to find the unique $\hat{\mathbf{y}} \in W \subset \mathbb{R}^n$ for the vector $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W and $\| \mathbf{y} - \hat{\mathbf{y}} \|$ is minimized.

Motivation: Find the closest vector $\hat{\mathbf{y}} \in W \subset \mathbb{R}^n$ to $\mathbf{y} \in \mathbb{R}^n$.

Last section: we did this in $\mathbb{R}^2$...now do in $\mathbb{R}^n$.

Example: Suppose $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis for $\mathbb{R}^3$. Let $W = \text{span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Write $\mathbf{y} \in \mathbb{R}^3$ as the sum of a vector $\hat{\mathbf{y}} \in W$ and a vector $\mathbf{z} \in W^\perp$.

This might be handy.

$$\boxed{c_j = \frac{\vec{w} \cdot \vec{u}_j}{\vec{u}_j \cdot \vec{u}_j}}$$

Theorem 8 *(The Orthogonal Decomposition Theorem)*
Let W be a subspace of $\mathbb{R}^n$. Then, each $\mathbf{y} \in \mathbb{R}^n$ can be uniquely represented in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z}$$

where $\hat{\mathbf{y}} \in W$ and $\mathbf{z} \in W^\perp$. That is, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is any orthogonal basis of W then

$$\text{proj}_W \mathbf{y} = \hat{\mathbf{y}} = \left(\frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right)\mathbf{u}_1 + \cdots + \left(\frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\right)\mathbf{u}_p$$

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

$\text{proj}_W \mathbf{y} = \hat{\mathbf{y}}$ is called the orthogonal projection of $\mathbf{y}$ onto W.

Existence

Proof: By Theorem 5, we know for $\hat{y} \in W$, we have

$$\hat{y} = \left(\frac{\vec{y}\cdot\vec{u}_1}{\|\vec{u}_1\|^2}\right)\vec{u}_1 + \cdots + \left(\frac{\vec{y}\cdot\vec{u}_p}{\|\vec{u}_p\|^2}\right)\vec{u}_p .$$

Let $\vec{z} = \vec{y} - \hat{y}$. Since $\{\vec{u}_1, \dots, \vec{u}_p\}$ is orthonormal set. $\vec{u}_j \cdot \vec{u}_j = 1 = \|\vec{u}_j\|^2$.

$\vec{u}_j \cdot \vec{u}_i = 0 \ \forall i \neq j, \ i,j \in \{1, \dots, p\}$ and

$$\Rightarrow \vec{z}\cdot\vec{u}_1 = (\vec{y}-\hat{y})\cdot\vec{u}_1 = \vec{y}\cdot\vec{u}_1 - \left[\left(\frac{\vec{y}\cdot\vec{u}_1}{\cancel{\|\vec{u}_1\|^2}}\right)\cancel{\vec{u}_1\cdot\vec{u}_1} + 0 + \cdots + 0\right] = \vec{y}\cdot\vec{u}_1 - \vec{y}\cdot\vec{u}_1 = \vec{0}.$$

likewise $\vec{z}\cdot\vec{u}_i = \vec{0} \quad \forall i = 1, \dots, p. \Rightarrow z \in W^\perp$.

Uniqueness

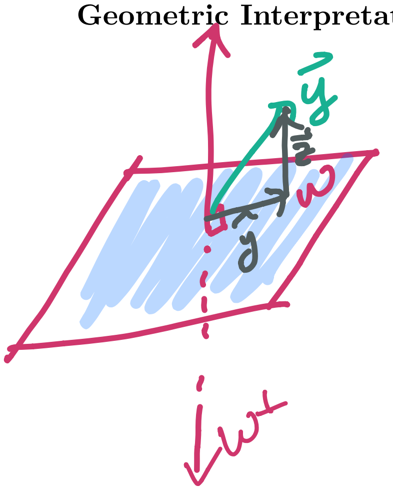
Suppose $\exists \ \vec{q}$ s.t. $\vec{y} = \vec{q} + \vec{w}$ w/ $\vec{q} \in W, \vec{w} \in W^\perp$.

$\Rightarrow \vec{z} + \hat{y} = \vec{q} + \vec{w} \iff \underbrace{\hat{y} - \vec{q}}_{\in W} = \underbrace{\vec{w} - \vec{z}}_{\in W^\perp}$ but the only vector that's both in W and $W^\perp$ is $\vec{0}$.

$\Rightarrow \vec{y} = \hat{y} + \vec{z}$ is unique. //

Example: Let $\mathbf{u}_1 = \begin{bmatrix}3\\0\\1\end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix}0\\1\\0\end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix}0\\3\\10\end{bmatrix}$. 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.

Geometric Interpretation of the Orthogonal Projection



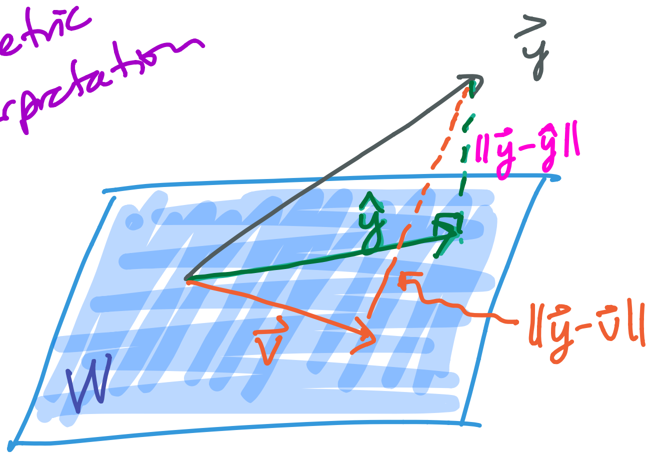
Qn: what if $y \in W$, what is $\text{proj}_W \vec{y}$?

Note: The orthogonal projection of $\mathbf{y}$ is the sum of its projections onto one-dimensional subspaces that are mutually orthogonal.

Properties of Orthogonal Projections

> **Theorem 9** *(The Best Approximation Theorem) Let W be a subspace of $\mathbb{R}^n$. Let $\mathbf{y}$ be any vector in $\mathbb{R}^n$. Let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then, $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$. That is*
>
> $$\| \mathbf{y} - \hat{\mathbf{y}} \| < \| \mathbf{y} - \mathbf{v} \|$$
>
> *for all $\mathbf{v} \neq \hat{\mathbf{y}} \in W$.*

geometric interpretation



$\|\vec{y}-\hat{y}\| < \|\vec{y}-\vec{v}\|$
$\forall \vec{v} \in W \ (\vec{v} \neq \hat{y})$

Example: Find the closest point to $\mathbf{y}$ in W=span$\{\mathbf{u}_1, \mathbf{u}_2\}$ where $\mathbf{y} = \begin{bmatrix} 2 \\ 4 \\ 0 \\ -2 \end{bmatrix}$, $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, and $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$.

Find the distance from $\mathbf{y}$ to W.

Note: "The" distance from a pt $\vec{y} \in \mathbb{R}^n$ to a subspace $W \subset \mathbb{R}^n$ is the distance from $\vec{y}$ to to the nearest point in W, i.e. $proj_W \vec{y}$.

Theorem 10 *If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then*

$$proj_W \mathbf{y} = \hat{\mathbf{y}} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p$$

If $U = \underbrace{[\mathbf{u}_1 \ \mathbf{u}_2 \ \ldots \ \mathbf{u}_p]}_{\text{matrix w/ } \vec{u}_i \text{ as columns}}$, *then*

$$\boxed{proj_W \mathbf{y} = UU^T\mathbf{y}}$$

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

And voila we're back to doing matrix multiplication!

Proof

read through this carefully & pay attention to summations

$$\hat{y} = (\vec{y} \cdot \vec{u}_1)\vec{u}_1 + \cdots + (\vec{y} \cdot \vec{u}_p)\vec{u}_p = (\vec{u}_1 \cdot \vec{y})\vec{u}_1 + \cdots + (\vec{u}_p \cdot \vec{y})\vec{u}_p$$

$$= (\vec{u}_1^T \vec{y})\vec{u}_1 + \cdots + (\vec{u}_p^T \vec{y})\vec{u}_p$$

$$= \vec{u}_1^T \vec{y} \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1n} \end{bmatrix} + \cdots + \vec{u}_p^T \vec{y} \begin{bmatrix} u_{p1} \\ u_{p2} \\ \vdots \\ u_{pn} \end{bmatrix} = \begin{bmatrix} \vec{u}_1^T \vec{y}\, u_{11} + \cdots + \vec{u}_p^T y\, u_{p1} \\ \vdots \\ \vec{u}_1^T \vec{y}\, u_{1n} + \cdots + \vec{u}_p^T \vec{y}\, u_{pn} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{p} (\vec{u}_i^T \vec{y})\, u_{i1} \\ \vdots \\ \sum_{i=1}^{p} (\vec{u}_i^T \vec{y})\, u_{in} \end{bmatrix}$$

$$UU^T = \begin{bmatrix} u_{11} & u_{21} & \cdots & u_{p1} \\ u_{12} & u_{22} & \cdots & u_{p2} \\ \vdots & \vdots & & \vdots \\ u_{1n} & u_{2n} & \cdots & u_{pn} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ u_{21} & u_{22} & \cdots & u_{2n} \\ u_{p1} & u_{p2} & \cdots & u_{pn} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{p} u_{i1}^2 & \sum_{i=1}^{p} u_{i1}u_{i2} & \cdots & \sum_{i=1}^{p} u_{i1}u_{in} \\ \vdots & & & \\ \sum_{i=1}^{p} u_{i1}u_{in} & \sum_{i=1}^{p} u_{i2}u_{in} & \cdots & \sum_{i=1}^{p} u_{in}^2 \end{bmatrix}$$

$$\Rightarrow UU^T\vec{y} = \begin{bmatrix} \sum_{i=1}^{p} u_{i1}^2 & \sum_{i=1}^{p} u_{i1}u_{i2} & \cdots & \sum_{i=1}^{p} u_{i1}u_{in} \\ \vdots & & & \\ \sum_{i=1}^{p} u_{i1}u_{in} & \sum_{i=1}^{p} u_{i2}u_{in} & \cdots & \sum_{i=1}^{p} u_{in}^2 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$= \begin{bmatrix} y_1 \sum_{i=1}^{p} u_{i1}^2 + y_2 \sum_{i=1}^{p} u_{i1}u_{i2} + \cdots + y_n \sum_{i=1}^{p} u_{i1}u_{in} \\ y_1 \sum_{i=1}^{p} u_{i1}u_{i2} + y_2 \sum_{i=1}^{p} u_{i2}^2 + y_3 \sum_{i=1}^{p} u_{i2}u_{i3} + \cdots + y_n \sum_{i=1}^{p} u_{i2}u_{in} \\ \vdots \\ y_1 \sum_{i=1}^{p} u_{i1}u_{in} + y_2 \sum_{i=1}^{p} u_{i2}u_{in} + \cdots + y_{n-1} \sum_{i=1}^{p} u_{i,n-1}u_{in} + y_n \sum_{i=1}^{p} u_{in}^2 \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=1}^{n} \sum_{i=1}^{p} u_{i1}u_{ij}y_j \\ \vdots \\ \sum_{j=1}^{n} \sum_{i=1}^{p} u_{in}u_{ij}y_j \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{p} u_{i1} \sum_{j=1}^{n} u_{ij}y_j \\ \vdots \\ \sum_{i=1}^{p} u_{in} \sum_{j=1}^{n} u_{ij}y_j \end{bmatrix}$$

now notice $\vec{u}_k^T\vec{y} = \begin{bmatrix} u_{k1} & \cdots & u_{kn} \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = u_{k1}y_1 + u_{k2}y_2 + \cdots + u_{kn}y_n$

$\forall k = 1, \ldots, p$ $\qquad = \sum_{j=1}^{n} u_{kj}y_j$

$$\Rightarrow UU^T\vec{y} = \begin{bmatrix} \sum_{i=1}^{p} u_{i1}(\vec{u}_i^T\vec{y}) \\ \vdots \\ \sum_{i=1}^{p} u_{in}(\vec{u}_i^T\vec{y}) \end{bmatrix} = \text{proj}_W\vec{y}. \qquad /\!/$$

How do we FIND an orthonormal basis? Yay, we find out!

Section 6.4 The Gram-Schmidt Process

ELOs:

- Construct an orthogonal (and orthonormal) basis for a subspace W using the Gram-Schmidt Process.
- Use the Gram-Schmidt process to find a QR factorization for a matrix A.

Goal: ~~Form~~ create an orthogonal basis for a subspace W.

Example: Suppose $W = \text{span}\{\mathbf{x}_1, \mathbf{x}_2\}$ where $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$.

Find an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2\}$ for W.

① let's start w/ $\vec{x}_1$ & then find a vector $\perp$ to $\vec{x}_1$.

i.e let $\vec{v}_1 = \vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ let $W_1 = \text{span}\{\vec{v}_1\}$

you finish

② and let $\vec{v}_2 = \vec{x}_2 - \text{proj}_{W_1} \vec{x}_2 = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} - \dfrac{\vec{x}_2 \cdot \vec{v}_1}{\|\vec{v}_1\|^2} \vec{v}_1$

③ check that $\vec{v}_1 \perp \vec{v}_2$.

Example: Suppose $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is a basis for a subspace W of $\mathbb{R}^4$. Describe an orthogonal basis for W.

① $\vec{v}_1 = \vec{x}_1$

$W_1 = \text{span}\{\vec{v}_1\}$

② $\vec{v}_2 = \vec{x}_2 - \text{proj}_{W_1} \vec{x}_2 =$

$W_2 = \text{span}\{\vec{v}_1, \vec{v}_2\}$

then ③ $\vec{v}_3 = \vec{x}_3 - \text{proj}_{W_2} \vec{x}_3$

It turns out the process we were doing always works!

Theorem 11 *The Gram-Schmidt Process*
Given a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ *for a nonzero subspace* W *of* $\mathbb{R}^n$. *Define*

$$\mathbf{v}_1 = \vec{x}_1 \qquad \text{let } W_1 = \operatorname{span}\{\vec{v}_1\}$$

$$\mathbf{v}_2 = \vec{x}_2 - \operatorname{proj}_{W_1}\vec{x}_2 \qquad W_2 = \operatorname{span}\{\vec{v}_1, \vec{v}_2\}$$

$$\mathbf{v}_3 = \vec{x}_3 - \operatorname{proj}_{W_2}\vec{x}_3 \qquad \vdots$$

$$\vdots$$

$$\mathbf{v}_p = \vec{x}_p - \operatorname{proj}_{W_{p-1}}\vec{x}_p \qquad W_{p-1} = \operatorname{span}\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{p-1}\}$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ *is an orthogonal basis for* W. *Moreover,*

$$\mathit{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \mathit{span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \text{ for } 1 \le k \le p$$

Note: We can always scale vectors to make them length 1 so, we can construct an **orthonormal basis** by adding one more step to the Gram-Schmidt process and normalizing each vector in our orthogonal basis. An **orthonormal basis** is an orthogonal basis of unit vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$.

Example: Suppose $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ where $\mathbf{x}_1 = \begin{bmatrix}1\\2\\3\\0\end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix}1\\2\\0\\0\end{bmatrix}$, $\mathbf{x}_3 = \begin{bmatrix}1\\0\\0\\1\end{bmatrix}$ is a basis for $W \subset \mathbb{R}^4$.
Describe an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for W.

QR Factorizations

we can interpret G.S. process as factoring the matrix whose columns are our original basis.

> **Theorem 12** *(The QR Factorization)*
> *If A is an $m \times n$ matrix with linearly independent columns, then*
> $$A = QR$$
> *where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A and R is an $n \times n$ upper triangular invertible matrix with positive entries on the diagonal.*

notice A does not need to be square

Note: Since Q has orthonormal columns, $Q^TQ = I$. Therefore, $Q^TA = Q^T(QR) = (Q^TQ)R = IR = R$.

Example: Find a QR factorization of $A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 3 \end{bmatrix}$.

Step 1 let $\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ $\vec{x}_2 = \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix}$ and $W = \text{span}\{\vec{x}_1, \vec{x}_2\}$.

Use Gram-Schmidt process to find orthogonal vectors $\vec{v}_1$ & $\vec{v}_2$.

Step 2: normalize those vectors $\vec{u}_1 = \dfrac{\vec{v}_1}{\|\vec{v}_1\|}$, $\vec{u}_2 = \dfrac{\vec{v}_2}{\|\vec{v}_2\|}$.

Step 3: $Q = [\vec{u}_1 \ \ \vec{u}_2]$

Step 4: Find R using $Q^TA = R$.

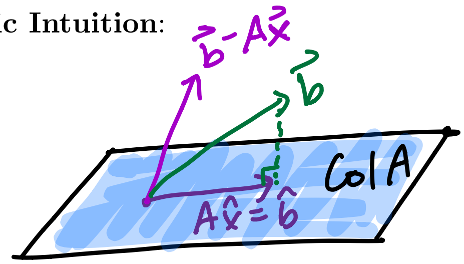
Section 6.5 Least-Squares Problems

ELOs:

- Define the least-squares problem.
- Find the least-squares solutions of a matrix using the normal equations and using QR factorization.
- Identify when a least-squares solution is unique.
- Derive a formula for the least-squares error.

Goal: If $A\mathbf{x} = \mathbf{b}$ has no solution, we want to find $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}}$ is as close as possible to $\mathbf{b}$.

Geometric Intuition:



Try to solve $A\vec{x} = \vec{b} \iff A\vec{x} - \vec{b} = \vec{0}$

but there's no solution because $\vec{b} \notin \text{Col}(A)$

$\Rightarrow$ Solve $A\hat{x} = \hat{b}$ where $\hat{b}$ is projection of $\vec{b}$ onto Col(A).

i.e. find $\hat{x}$ s.t. $A\hat{x} = \hat{b}$.

notice it says "a" not "the".

Definition 1 *If A is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$, a least-squares solution of $A\mathbf{x} = \mathbf{b}$ is a vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that*

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

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

we're minimizing distance between vectors

Note: For any $\mathbf{x} \in \mathbb{R}^n$, $A\mathbf{x} \in$ Col A. We want to find $\hat{\mathbf{x}}$ such that $A\hat{\mathbf{x}}$ is the closest point in Col A to $\mathbf{b}$.

get best approx solution $\hat{x}$ by solving

$$A\hat{x} = \text{proj}_{\text{Col}A}\vec{b}$$

Remember:

$m \times n$ matrix U has orthonormal columns iff $U^TU = I$.

and ① $\|U\vec{x}\| = \|\vec{x}\|$

② $(U\vec{x}) \cdot (U\vec{y}) = \vec{x} \cdot \vec{y}$

③ $(U\vec{x}) \cdot (U\vec{y}) = 0$ iff $\vec{x} \cdot \vec{y} = 0$.

Question: What if $\mathbf{b} \in$ Col A?

Theorem 13 *The set of least-squares solutions of* $A\mathbf{x} = \mathbf{b}$ *is the set of all solutions of the normal equations* $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.

Proof

Suppose $\hat{x}$ satisfies $A\hat{x} = \hat{b}$. Then $\vec{b} - \hat{b}$ is orthogonal to Col A (since $\hat{b} \in \text{Col} A$).

$\Rightarrow \vec{b} - \hat{b}$ orthog to each column of A.

$\Rightarrow \vec{b} - \hat{b}$ orthog to each row of A^T.

$\Rightarrow A^T(\vec{b} - \hat{b}) = \vec{0} \iff A^T(\vec{b} - A\hat{x}) = \vec{0}$

$\iff \boxed{A^T\vec{b} = A^T A\hat{x}}$. i.e. a least squares solution $\hat{x}$ satisfies this eqn.

//

Example: Find a least-squares solution to the inconsistent system $A\mathbf{x} = \mathbf{b}$ where $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \\ 2 & 2 \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

$A^T = \begin{bmatrix} 2 & 0 & 2 \\ 0 & 1 & 2 \end{bmatrix} \Rightarrow A^T\vec{b} =$

$A^T\vec{b} = A^T A\hat{x} \iff$

Observe: When A^TA is invertible, we still have $A^T\vec{b} = A^TA\hat{x}$ as least squares solution.

If $(A^TA)^{-1}$ exists, then

$$A^T\vec{b} = (A^TA)\hat{x}$$

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

$$\Leftrightarrow \boxed{\hat{x} = (A^TA)^{-1}A^T\vec{b}}$$

Note: $(A^TA)^{-1} \neq A^{-1}(A^T)^{-1}$ since A is not necessarily square.

Note: In some cases, there may be more than one possible least squares solution. For example, the least squares solution is not unique when the normal equations have free variables.

Theorem 14 *Let A be an $m \times n$ matrix. The following statements are logically equivalent:*

- $A\mathbf{x} = \mathbf{b}$ *has a unique least-squares solution* $\forall\ \mathbf{b} \in \mathbb{R}^m$.
- *The columns of A are linearly independent.*
- A^TA *is invertible.*

When these statements are true, the least-squares solution is

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

Definition 2 *The least-squares error is the distance between* $\mathbf{b}$ *and the vector* $A\hat{\mathbf{x}}$.

$$\|\ \mathbf{b} - A\hat{\mathbf{x}}\ \|$$

Exercise: Determine the least-squares error for the least-squares solution of $A\mathbf{x} = \mathbf{b}$ from the Example.

(on last page)

Theorem 15 *Given an $m \times n$ matrix A with linearly independent columns. Let $A = QR$ be the QR factorization of A. Then, for each $\mathbf{b} \in \mathbb{R}^m$, $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution given by*

$$\boxed{\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}.}$$

Proof: $A\hat{x} = A(R^{-1}Q^T\vec{b}) = QR(R^{-1}Q^T)\vec{b}$

$= QIQ^T\vec{b} = QQ^T\vec{b} = \text{proj}_{\text{Col}A}\vec{b} = \hat{b}$

since Q has orthogonal columns

$\Rightarrow$ proposed $\hat{x}$ is indeed solution to least squares problem.

Example: Find the unique least-squares solution of

$$\begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$$

where

Step 1: Find QR factorization. (done)

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

Step 2: $Q^T\vec{b} =$

Step 3: Solve $R\hat{x} = Q^T\vec{b}$ (i.e. $\hat{x} = R^{-1}Q^T\vec{b}$)