# Section 6.4-6.5 The Gram-Schmit Process, least-square problems, and applications to linear models

Gexin Yu
gyu@wm.edu

College of William and Mary

# Example

Ex. Let $x_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$, $x_2 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix}$, and $x_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$, $y = \begin{bmatrix}2\\-1\\3\\4\end{bmatrix}$, and $W = Span\{x_1, x_2, x_3\}$. Find the projection of $y$ onto $W$.

$$\text{Proj}_W \hat{y} = \frac{y \cdot x_1}{x_1 \cdot x_1} x_1 + \frac{y \cdot x_2}{x_2 \cdot x_2} x_2 + \frac{y \cdot x_3}{x_3 \cdot x_3} x_3$$

WRONG! not orthogonal set!

# Orthogonal basis

- Given a basis for a subspace $W$ of $\mathbf{R}^n$, how to find an orthogonal basis for $W$?
- Let $\{x_1, x_2, \ldots, x_p\}$ be a basis for $W$.
- The idea is as follows: let $v_1 = x_1$ and take $W_1 = Span\{v_1\}$, then project $x_2$ to $W_1$ and let $v_2$ be the component of $x_2$ orthogonal to $W_1$; then let $W_2 = Span\{v_1, v_2\}$, and project $x_3$ to $W_2$ and let $v_3$ be the component of $x_3$ orthogonal to $W_2$. Then $\{v_1, v_2, v_3\}$ is an orthogonal basis for $W$; and so on.
- This is so-called Gram-Schmidt Process.

# The Gram-Schmidt Process

- Given a basis $\{x_1, x_2, \ldots, x_p\}$ for a nonzero subspace $W$ of $\mathbf{R}^n$, define

$$
\begin{aligned}
v_1 &= x_1 \\
v_2 &= x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 \quad = x_2 - \mathrm{Proj}_{W_1} x_2 \\
v_3 &= x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2 \quad = x_3 - \mathrm{Proj}_{W_2} x_3 \\
&\cdots \\
v_p &= x_p - \frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_p \cdot v_2}{v_2 \cdot v_2} v_2 - \ldots - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1}
\end{aligned}
$$

Then $\{v_1, v_2, \ldots, v_p\}$ is an orthogonal basis for $W$.

# Example

Ex. Let $W = span\{x_1, x_2\}$, where $x_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $x_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Construct an orthogonal basis $\{v_1, v_2\}$ for $W$.

# Example

Ex. Let $W = span\{x_1, x_2\}$, where $x_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $x_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Construct an orthogonal basis $\{v_1, v_2\}$ for $W$.

- So let $v_1 = x_1$.

# Example

Ex. Let $W = span\{x_1, x_2\}$, where $x_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $x_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Construct an orthogonal basis $\{v_1, v_2\}$ for $W$.

- So let $v_1 = x_1$.
- let

$$v_2 = x_2 - p = x_2 - \frac{x_2 \cdot x_1}{x_1 \cdot x_1} x_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{15}{45} \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

# Example

Ex. Let $W = span\{x_1, x_2\}$, where $x_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}$ and $x_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$. Construct an orthogonal basis $\{v_1, v_2\}$ for $W$.

- So let $v_1 = x_1$.
- let

$$v_2 = x_2 - p = x_2 - \frac{x_2 \cdot x_1}{x_1 \cdot x_1} x_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{15}{45} \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

- Then $\{v_1, v_2\}$ is an orthogonal basis for $W$.

# Example

Ex. Let $x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, and $x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ and $W = Span\{x_1, x_2, x_3\}$.
Construct an orthogonal basis $\{v_1, v_2, v_3\}$ for $W$.

# Example

Ex. Let $x_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$, $x_2 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix}$, and $x_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$ and $W = Span\{x_1, x_2, x_3\}$.

Construct an orthogonal basis $\{v_1, v_2, v_3\}$ for $W$.

$v_1 = x_1$

- $v_2 = x_2 - proj_{W_1} x_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix} - \frac{3}{4}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \begin{bmatrix}-3/4\\1/4\\1/4\\1/4\end{bmatrix}$.

# Example

Ex. Let $x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, and $x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ and $W = Span\{x_1, x_2, x_3\}$.

Construct an orthogonal basis $\{v_1, v_2, v_3\}$ for $W$.

- $v_2 = x_2 - proj_{W_1} x_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$.
- $v_3 = x_3 - proj_{W_2} x_3 = x_3 - (\frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2) = \begin{bmatrix} -3/8 \\ -13/24 \\ 11/24 \\ 11/24 \end{bmatrix}$.

# Example

Ex. Let $x_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, and $x_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$ and $W = Span\{x_1, x_2, x_3\}$.

Construct an orthogonal basis $\{v_1, v_2, v_3\}$ for $W$.

- $v_2 = x_2 - proj_{W_1} x_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$.
- $v_3 = x_3 - proj_{W_2} x_3 = x_3 - (\frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2) = \begin{bmatrix} -3/8 \\ -13/24 \\ 11/24 \\ 11/24 \end{bmatrix}$.
- $\{v_1, v_2, v_3\}$ is an orthogonal basis for $W$.

Least-square problems

# Least-squares problems

- We know that some linear systems $Ax = b$ may not be consistent.

# Least-squares problems

- We know that some linear systems $Ax = b$ may not be consistent.
- Note that no matter what $x$ we select, the vector $Ax$ will necessarily be in the column space of $A$, $Col\ A$.

# Least-squares problems

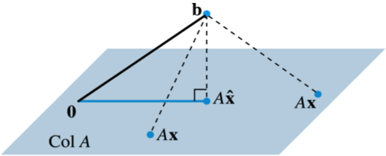
- We know that some linear systems $Ax = b$ may not be consistent.
- Note that no matter what $x$ we select, the vector $Ax$ will necessarily be in the column space of $A$, $Col\ A$.
- So we seek an $x$ in $Col\ A$ so that $Ax$ is the closest point to $b$.

# Least-squares problems

- We know that some linear systems $Ax = b$ may not be consistent.
- Note that no matter what $x$ we select, the vector $Ax$ will necessarily be in the column space of $A$, $Col\ A$.
- So we seek an $x$ in $Col\ A$ so that $Ax$ is the closest point to $b$.
- Definition: If $A$ is $m \times n$ and $b$ in $\mathbf{R}^n$, a least-squares solution of $Ax = b$ is an $\hat{x}$ in $\mathbf{R}^n$ such that $||b - A\hat{x}|| \leq ||b - Ax||$ for all $x \in \mathbf{R}^n$.

# Least-squares problems

- We know that some linear systems $Ax = b$ may not be consistent.
- Note that no matter what $x$ we select, the vector $Ax$ will necessarily be in the column space of $A$, *Col* $A$.
- So we seek an $x$ in *Col* $A$ so that $Ax$ is the closest point to $b$.
- Definition: If $A$ is $m \times n$ and $b$ in $\mathbf{R}^n$, a least-squares solution of $Ax = b$ is an $\hat{x}$ in $\mathbf{R}^n$ such that $||b - A\hat{x}|| \leq ||b - Ax||$ for all $x \in \mathbf{R}^n$.




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

$u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$

$\|u\| = \sqrt{u_1^2 + \cdots + u_n^2}$

# Solution to the general least-squares problem

- Given $A$ and $b$, apply the Best Approximation Theorem to the subspace $Col\ A$. Let $\hat{b} = proj_{Col\ A} b$.

# Solution to the general least-squares problem

- Given $A$ and $b$, apply the Best Approximation Theorem to the subspace $Col\ A$. Let $\hat{b} = proj_{Col\ A} b$.
- Because $\hat{b}$ is in the column space of $A$, the equation $Ax = \hat{b}$ is consistent, and there is an $\hat{x}$ in $\mathbf{R}^n$ such that $A\hat{x} = \hat{b}$.

# Solution to the general least-squares problem

- Given $A$ and $b$, apply the Best Approximation Theorem to the subspace $Col\ A$. Let $\hat{b} = proj_{Col\ A} b$.
- Because $\hat{b}$ is in the column space of $A$, the equation $Ax = \hat{b}$ is consistent, and there is an $\hat{x}$ in $\mathbf{R}^n$ such that $A\hat{x} = \hat{b}$.
- Since $\hat{b}$ is the closest point in $Col\ A$ to $b$, a vector $\hat{x}$ is a least-squares solution of $Ax = b$ if and only if $\hat{x}$ satisfies $A\hat{x} = \hat{b}$.

least-square solution to $Ax = b$

$\Updownarrow$

find $\hat{x}$ s.t. $A\hat{x} = Proj_{Col A} b$

# Solution to the general least-squares problem

- Given $A$ and $b$, apply the Best Approximation Theorem to the subspace $Col\ A$. Let $\hat{b} = proj_{Col\ A} b$.
- Because $\hat{b}$ is in the column space of $A$, the equation $Ax = \hat{b}$ is consistent, and there is an $\hat{x}$ in $\mathbf{R}^n$ such that $A\hat{x} = \hat{b}$.
- Since $\hat{b}$ is the closest point in $Col\ A$ to $b$, a vector $\hat{x}$ is a least-squares solution of $Ax = b$ if and only if $\hat{x}$ satisfies $A\hat{x} = \hat{b}$.
- Such an $\hat{x}$ in $\mathbf{R}^n$ is a list of weights that will build $\hat{b}$ out of the columns of $A$.

# Solution to the general least-squares problem

- Given $A$ and $b$, apply the Best Approximation Theorem to the subspace $Col\ A$. Let $\hat{b} = proj_{Col\ A} b$.
- Because $\hat{b}$ is in the column space of $A$, the equation $Ax = \hat{b}$ is consistent, and there is an $\hat{x}$ in $\mathbf{R}^n$ such that $A\hat{x} = \hat{b}$.
- Since $\hat{b}$ is the closest point in $Col\ A$ to $b$, a vector $\hat{x}$ is a least-squares solution of $Ax = b$ if and only if $\hat{x}$ satisfies $A\hat{x} = \hat{b}$.
- Such an $\hat{x}$ in $\mathbf{R}^n$ is a list of weights that will build $\hat{b}$ out of the columns of $A$.
- Suppose $\hat{x}$ satisfies $A\hat{x} = \hat{b}$.

# Solution to the general least-squares problem

- Given $A$ and $b$, apply the Best Approximation Theorem to the subspace $Col\ A$. Let $\hat{b} = proj_{Col\ A} b$.
- Because $\hat{b}$ is in the column space of $A$, the equation $Ax = \hat{b}$ is consistent, and there is an $\hat{x}$ in $\mathbf{R}^n$ such that $A\hat{x} = \hat{b}$.
- Since $\hat{b}$ is the closest point in $Col\ A$ to $b$, a vector $\hat{x}$ is a least-squares solution of $Ax = b$ if and only if $\hat{x}$ satisfies $A\hat{x} = \hat{b}$.
- Such an $\hat{x}$ in $\mathbf{R}^n$ is a list of weights that will build $\hat{b}$ out of the columns of $A$.
- Suppose $\hat{x}$ satisfies $A\hat{x} = \hat{b}$.
- By the Orthogonal Decomposition Theorem, the projection $\hat{b}$ has the property that $b - \hat{b}$ is orthogonal to $Col\ A$, so $b - A\hat{x}$ is orthogonal to each column of $A$.

- If $a_j$ is any column of $A$, then $a_j \cdot (b - A\hat{x}) = 0$, and $a_j^T(b - A\hat{x}) = 0$.

- If $a_j$ is any column of $A$, then $a_j \cdot (b - A\hat{x}) = 0$, and $a_j^T(b - A\hat{x}) = 0$.
- Since each $a_j^T$ is a row of $A^T$, we have $A^T(b - A\hat{x}) = 0$.

- If $a_j$ is any column of $A$, then $a_j \cdot (b - A\hat{x}) = 0$, and $a_j^T(b - A\hat{x}) = 0$.
- Since each $a_j^T$ is a row of $A^T$, we have $A^T(b - A\hat{x}) = 0$.
- Thus $A^T b - A^T A\hat{x} = 0$, and $A^T A\hat{x} = A^T b$.

- If $a_j$ is any column of $A$, then $a_j \cdot (b - A\hat{x}) = 0$, and $a_j^T(b - A\hat{x}) = 0$.
- Since each $a_j^T$ is a row of $A^T$, we have $A^T(b - A\hat{x}) = 0$.
- Thus $A^Tb - A^TA\hat{x} = 0$, and $A^TA\hat{x} = A^Tb$.
- These calculations show that each least-squares solution of $Ax = b$ satisfies the equation $A^TAx = A^Tb$.

- If $a_j$ is any column of $A$, then $a_j \cdot (b - A\hat{x}) = 0$, and $a_j^T(b - A\hat{x}) = 0$.
- Since each $a_j^T$ is a row of $A^T$, we have $A^T(b - A\hat{x}) = 0$.
- Thus $A^Tb - A^TA\hat{x} = 0$, and $A^TA\hat{x} = A^Tb$.
- These calculations show that each least-squares solution of $Ax = b$ satisfies the equation $A^TAx = A^Tb$.
- The matrix equation $A^TAx = A^Tb$ represents a system of equations called the normal equations for $Ax = b$.

- If $a_j$ is any column of $A$, then $a_j \cdot (b - A\hat{x}) = 0$, and $a_j^T(b - A\hat{x}) = 0$.

- Since each $a_j^T$ is a row of $A^T$, we have $A^T(b - A\hat{x}) = 0$.

- Thus $A^Tb - A^TA\hat{x} = 0$, and $A^TA\hat{x} = A^Tb$.

- These calculations show that each least-squares solution of $Ax = b$ satisfies the equation $A^TAx = A^Tb$.

- The matrix equation $A^TAx = A^Tb$ represents a system of equations called the normal equations for $Ax = b$.

- A solution to $A^TAx = A^Tb$ is often denoted by $\hat{x}$.

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

**THM** The set of least-squares solutions of $Ax = b$ coincides with the nonempty set of solutions of the normal equation $A^T Ax = A^T b$.

**PF.** The set of least-squares solutions is nonempty and each least-squares solution $\hat{x}$ satisfies the normal equations.

$C_1 \subseteq C_2$

$C_2 \subseteq C_1$

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

PF. The set of least-squares solutions is nonempty and each least-squares solution $\hat{x}$ satisfies the normal equations.

- Conversely, suppose $\hat{x}$ satisfies $A^T A\hat{x} = A^T b$.

**THM** The set of least-squares solutions of $Ax = b$ coincides with the nonempty set of solutions of the normal equation $A^T Ax = A^T b$.

**PF.** The set of least-squares solutions is nonempty and each least-squares solution $\hat{x}$ satisfies the normal equations.

- Conversely, suppose $\hat{x}$ satisfies $A^T A\hat{x} = A^T b$.
- Then $\hat{x}$ satisfies $A^T(b - A\hat{x}) = 0$, which shows that $b - A\hat{x}$ is orthogonal to the rows of $A^T$, and hence is orthogonal to the columns of $A$.

**THM** The set of least-squares solutions of $Ax = b$ coincides with the nonempty set of solutions of the normal equation $A^T Ax = A^T b$.

**PF.** The set of least-squares solutions is nonempty and each least-squares solution $\hat{x}$ satisfies the normal equations.

- Conversely, suppose $\hat{x}$ satisfies $A^T A\hat{x} = A^T b$.
- Then $\hat{x}$ satisfies $A^T(b - A\hat{x}) = 0$, which shows that $b - A\hat{x}$ is orthogonal to the rows of $A^T$, and hence is orthogonal to the columns of $A$.
- Since the columns of $A$ span $Col\ A$, the vector $b - A\hat{x}$ is orthogonal to all of $Col\ A$.

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

PF. The set of least-squares solutions is nonempty and each least-squares solution $\hat{x}$ satisfies the normal equations.

- Conversely, suppose $\hat{x}$ satisfies $A^T A\hat{x} = A^T b$.
- Then $\hat{x}$ satisfies $A^T(b - A\hat{x}) = 0$, which shows that $b - A\hat{x}$ is orthogonal to the rows of $A^T$, and hence is orthogonal to the columns of $A$.
- Since the columns of $A$ span *Col* $A$, the vector $b - A\hat{x}$ is orthogonal to all of *Col* $A$.
- Hence the equation $b = A\hat{x} + (b - A\hat{x})$ is a decomposition of $b$ into the sum of a vector in *Col* $A$ and a vector orthogonal to *Col* $A$.

THM The set of least-squares solutions of $Ax = b$ coincides with the nonempty set of solutions of the normal equation $A^TAx = A^Tb$.

PF. The set of least-squares solutions is nonempty and each least-squares solution $\hat{x}$ satisfies the normal equations.

- Conversely, suppose $\hat{x}$ satisfies $A^TA\hat{x} = A^Tb$.
- Then $\hat{x}$ satisfies $A^T(b - A\hat{x}) = 0$, which shows that $b - A\hat{x}$ is orthogonal to the rows of $A^T$, and hence is orthogonal to the columns of $A$.
- Since the columns of $A$ span $Col\ A$, the vector $b - A\hat{x}$ is orthogonal to all of $Col\ A$.
- Hence the equation $b = A\hat{x} + (b - A\hat{x})$ is a decomposition of $b$ into the sum of a vector in $Col\ A$ and a vector orthogonal to $Col\ A$.
- By the uniqueness of the orthogonal decomposition, $A\hat{x}$ must be the orthogonal projection of $b$ onto $Col\ A$.

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

PF. The set of least-squares solutions is nonempty and each least-squares solution $\hat{x}$ satisfies the normal equations.

- Conversely, suppose $\hat{x}$ satisfies $A^T A\hat{x} = A^T b$.
- Then $\hat{x}$ satisfies $A^T(b - A\hat{x}) = 0$, which shows that $b - A\hat{x}$ is orthogonal to the rows of $A^T$, and hence is orthogonal to the columns of $A$.
- Since the columns of $A$ span $Col\ A$, the vector $b - A\hat{x}$ is orthogonal to all of $Col\ A$.
- Hence the equation $b = A\hat{x} + (b - A\hat{x})$ is a decomposition of $b$ into the sum of a vector in $Col\ A$ and a vector orthogonal to $Col\ A$.
- By the uniqueness of the orthogonal decomposition, $A\hat{x}$ must be the orthogonal projection of $b$ onto $Col\ A$.
- That is, $A\hat{x} = \hat{b}$ and $\hat{x}$ is a least-squares solution.

# Example

Ex. Find a least-squares solution of the inconsistent system $Ax = b$ for

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

# Example

Ex. Find a least-squares solution of the inconsistent system $Ax = b$ for

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

$\Updownarrow$

$A^T A x = A^T b$

Sol. To use the normal equation, compute:

$$A^T A = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}, A^T b = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

# Example

Ex. Find a least-squares solution of the inconsistent system $Ax = b$ for

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

Sol. To use the normal equation, compute:

$$A^T A = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}, A^T b = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

- Then the equation $A^T Ax = A^T b$ becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

# Example

Ex. Find a least-squares solution of the inconsistent system $Ax = b$ for

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

Sol. To use the normal equation, compute:

$$A^T A = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}, A^T b = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

- Then the equation $A^T Ax = A^T b$ becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

- Solve it, we have $\hat{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.

- Note that the matrix equation $A^T Ax = A^T b$ may have infinite many solutions.

- Note that the matrix equation $A^T Ax = A^T b$ may have infinite many solutions.
- When will we have a unique least-squares solution to $Ax = b$?

- Note that the matrix equation $A^T Ax = A^T b$ may have infinite many solutions.

- When will we have a unique least-squares solution to $Ax = b$?

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

When these statements are true, the least-squares solution $\hat{x}$ is given by $\hat{x} = (A^T A)^{-1} A^T b$.

- Note that the matrix equation $A^T Ax = A^T b$ may have infinite many solutions.

- When will we have a unique least-squares solution to $Ax = b$?

THM Let $A$ be an $m \times n$ matrix. The following statements are logically equivalent:
  - The equation $Ax = b$ has a unique least-square solution for each $b$ in $\mathbf{R}^m$.

When these statements are true, the least-squares solution $\hat{x}$ is given by $\hat{x} = (A^T A)^{-1} A^T b$.

- Note that the matrix equation $A^T Ax = A^T b$ may have infinite many solutions.

- When will we have a unique least-squares solution to $Ax = b$?

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

  - The equation $Ax = b$ has a unique least-square solution for each $b$ in $\mathbf{R}^m$.
  - The columns of $A$ are linearly independent.

When these statements are true, the least-squares solution $\hat{x}$ is given by $\hat{x} = (A^T A)^{-1} A^T b$.

- Note that the matrix equation $A^T Ax = A^T b$ may have infinite many solutions.

- When will we have a unique least-squares solution to $Ax = b$?

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

  - The equation $Ax = b$ has a unique least-square solution for each $b$ in $\mathbf{R}^m$.
  - The columns of $A$ are linearly independent.
  - The matrix $A^T A$ is invertible.

When these statements are true, the least-squares solution $\hat{x}$ is given by $\hat{x} = (A^T A)^{-1} A^T b$.

$A^T A x = A^T b$

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

- Note that the matrix equation $A^T Ax = A^T b$ may have infinite many solutions.

- When will we have a unique least-squares solution to $Ax = b$?

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

  - The equation $Ax = b$ has a unique least-square solution for each $b$ in $\mathbf{R}^m$.
  - The columns of $A$ are linearly independent.
  - The matrix $A^T A$ is invertible.

  When these statements are true, the least-squares solution $\hat{x}$ is given by $\hat{x} = (A^T A)^{-1} A^T b$.

- The distance from $b$ to $A\hat{x}$, $||b - A\hat{x}||$, is called the least-squares error of this approximation.

# Alternative calculation

A =

- When the columns of $A[u_1 \ u_2 \ \ldots \ u_p]$ are orthogonal, we know exactly the orthogonal projection of $b$ on $Col\ A$:

$$\hat{b} = \frac{b \cdot u_1}{u_1 \cdot u_1} u_1 + \ldots + \frac{b \cdot u_p}{u_p \cdot u_p} u_p$$

# Alternative calculation

- When the columns of $A[u_1 \; u_2 \; \dots \; u_p]$ are orthogonal, we know exactly the orthogonal projection of $b$ on $Col\ A$:

$$u_1x_1+u_2x_2+\dots = Ax = \hat{b} = \frac{b \cdot u_1}{u_1 \cdot u_1}u_1 + \dots + \frac{b \cdot u_p}{u_p \cdot u_p}u_p$$

- Now to get the least-squares solution to $A\hat{x} = \hat{b}$, we just need to read:

$$x_1 = \frac{b \cdot u_1}{u_1 \cdot u_1}, x_2 = \frac{b \cdot u_2}{u_2 \cdot u_2}, \dots, x_p = \frac{b \cdot u_p}{u_p \cdot u_p}$$

# Alternative calculation

- When the columns of $A[u_1 \ u_2 \ \dots \ u_p]$ are orthogonal, we know exactly the orthogonal projection of $b$ on $Col\ A$:

$$\hat{b} = \frac{b \cdot u_1}{u_1 \cdot u_1} u_1 + \ldots + \frac{b \cdot u_p}{u_p \cdot u_p} u_p$$

- Now to get the least-squares solution to $A\hat{x} = \hat{b}$, we just need to read:

$$x_1 = \frac{b \cdot u_1}{u_1 \cdot u_1}, x_2 = \frac{b \cdot u_2}{u_2 \cdot u_2}, \ldots, x_p = \frac{b \cdot u_p}{u_p \cdot u_p}$$

- Such matrices often appear in linear regression problems.

# Example

Ex. Find a least-squares solution of $Ax = b$ for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, b = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

# Example

Ex. Find a least-squares solution of $Ax = b$ for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, b = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

Sol. We know that

$$\hat{b} = \frac{b \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{b \cdot u_2}{u_2 \cdot u_2} u_2 = 2u_1 + 1/2u_2 = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix}$$

# Example

Ex. Find a least-squares solution of $Ax = b$ for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, b = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

Sol. We know that

$$\hat{b} = \frac{b \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{b \cdot u_2}{u_2 \cdot u_2} u_2 = 2u_1 + 1/2 u_2 = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix}$$

- Now we solve $A\hat{x} = \hat{b}$: $\hat{x} = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}$.

# Example

Ex. Find a least-squares solution of $Ax = b$ for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, b = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

Sol. We know that

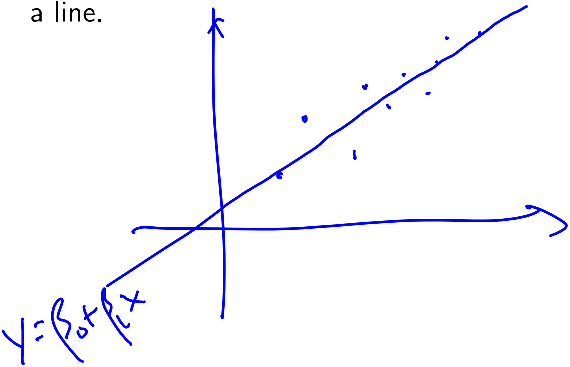
$$\hat{b} = \frac{b \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{b \cdot u_2}{u_2 \cdot u_2} u_2 = 2u_1 + 1/2 u_2 = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix}$$

- Now we solve $A\hat{x} = \hat{b}$: $\hat{x} = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}$.
- We can also get the least-squares error: $||b - A\hat{x}|| = \sqrt{\ }$.

Applications to linear models

# Applications to linear models

- The simplest relation between two variables $x$ and $y$ is the linear equation $y = \beta_0 + \beta_1 x$. Experimental data often produce points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ that, when graphed, seem to lie close to a line.



# Applications to linear models

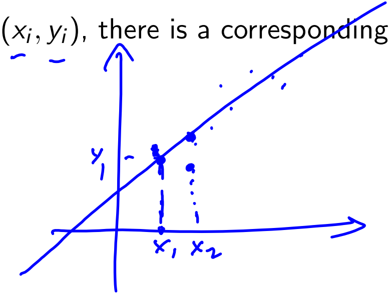
- The simplest relation between two variables $x$ and $y$ is the linear equation $y = \beta_0 + \beta_1 x$. Experimental data often produce points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ that, when graphed, seem to lie close to a line.
- We want to determine $\beta_0$ and $\beta_1$ that make the line as "close" to the points as possible.

# Applications to linear models

- The simplest relation between two variables $x$ and $y$ is the linear equation $y = \beta_0 + \beta_1 x$. Experimental data often produce points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ that, when graphed, seem to lie close to a line.
- We want to determine $\beta_0$ and $\beta_1$ that make the line as "close" to the points as possible.
- Suppose that $\beta_0$ and $\beta_1$ are fixed, and consider the line $y = \beta_0 + \beta_1 x$.

# Applications to linear models

- The simplest relation between two variables $x$ and $y$ is the linear equation $y = \beta_0 + \beta_1 x$. Experimental data often produce points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ that, when graphed, seem to lie close to a line.
- We want to determine $\beta_0$ and $\beta_1$ that make the line as "close" to the points as possible.
- Suppose that $\beta_0$ and $\beta_1$ are fixed, and consider the line $y = \beta_0 + \beta_1 x$.
- For each point $(x_i, y_i)$, there is a corresponding point $(x_i, \beta_0 + \beta_1 x_i)$ on the line.




# Applications to linear models

- The simplest relation between two variables $x$ and $y$ is the linear equation $y = \beta_0 + \beta_1 x$. Experimental data often produce points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ that, when graphed, seem to lie close to a line.
- We want to determine $\beta_0$ and $\beta_1$ that make the line as "close" to the points as possible.
- Suppose that $\beta_0$ and $\beta_1$ are fixed, and consider the line $y = \beta_0 + \beta_1 x$.
- For each point $(x_i, y_i)$, there is a corresponding point $(x_i, \beta_0 + \beta_1 x_i)$ on the line.
- We call $y_i$ the observed value of $y$ and $\beta_0 + \beta_1 x_i$ the predicted $y$-value. The difference of an observed $y$-value and the predicted $y$-value is called a residual.

# Applications to linear models

- The simplest relation between two variables $x$ and $y$ is the linear equation $y = \beta_0 + \beta_1 x$. Experimental data often produce points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ that, when graphed, seem to lie close to a line.
- We want to determine $\beta_0$ and $\beta_1$ that make the line as "close" to the points as possible.
- Suppose that $\beta_0$ and $\beta_1$ are fixed, and consider the line $y = \beta_0 + \beta_1 x$.
- For each point $(x_i, y_i)$, there is a corresponding point $(x_i, \beta_0 + \beta_1 x_i)$ on the line.
- We call $y_i$ the observed value of $y$ and $\beta_0 + \beta_1 x_i$ the predicted $y$-value. The difference of an observed $y$-value and the predicted $y$-value is called a residual.
- The least-squares line, or the line of regression of $y$ on $x$, is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares of the residuals.

- The $\beta_0$ and $\beta_1$ satisfy the following

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}$$

$$\beta_0 + \beta_1 x_1 = y_1$$
$$\beta_0 + \beta_1 x_2 = y_2$$
$$\vdots$$
$$\beta_0 + \beta_1 x_n = y_n$$

- The $\beta_0$ and $\beta_1$ satisfy the following

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}$$

- Or simply just $X\beta = y$.

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k$

- The $\beta_0$ and $\beta_1$ satisfy the following

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \\ 1 & x_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix}$$

- Or simply just $X\beta = y$.

- We can find the least-squares solution to $X\beta = y$ by solving the matrix equation $X^T X\beta = X^T y$.

# Example

- 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 $(8, 3)$.

# Example

- 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 $(8, 3)$.

Sol. We solve the equation $X^T X\beta = X^T y$, where

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

# Example

- 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 $(8,3)$.

Sol. We solve the equation $X^T X\beta = X^T y$, where

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

- The equation can be written as

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

# Example

- 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 $(8,3)$.

Sol. We solve the equation $X^T X \beta = X^T y$, where

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

- The equation can be written as

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

- So the solution is $[\beta_0 \ \beta_1] = [2/7 \ 5/14]$.

# Example

- 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 $(8, 3)$.

Sol. We solve the equation $X^T X\beta = X^T y$, where

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

- The equation can be written as

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

- So the solution is $[\beta_0\ \beta_1] = [2/7\ 5/14]$.
- The least-squares line is $y = 2/7 + 5/14x$.