

Example 32. (warmup) $\begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}$

Note that this means that the system of equations $\begin{matrix} x_1 + 2x_2 = 1 \\ 3x_1 + x_2 = 1 \\ 5x_2 = 1 \end{matrix}$ can also be written as $\begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

[This was the motivation for introducing matrix-vector multiplication.]

In the same way, any system can be written as $Ax = b$, where A is a matrix and b a vector. In particular, this makes it obvious that:

$$Ax = b \text{ is consistent} \iff b \text{ is in } \text{col}(A)$$

Recall that, by the FTLA, $\text{col}(A)$ and $\text{null}(A^T)$ are orthogonal complements.

Theorem 33. $Ax = b$ is consistent $\iff b$ is orthogonal to $\text{null}(A^T)$

Proof. $Ax = b$ is consistent $\iff b$ is in $\text{col}(A) \xrightarrow{\text{FTLA}} b$ is orthogonal to $\text{null}(A^T)$

Note. b is orthogonal to $\text{null}(A^T)$ means that $y^T b = 0$ whenever $y^T A = 0$. Why?!

Example 34. Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix}$. For which b does $Ax = b$ have a solution?

Solution. (old)

$$\left[\begin{array}{cc|c} 1 & 2 & b_1 \\ 3 & 1 & b_2 \\ 0 & 5 & b_3 \end{array} \right] \xrightarrow{R_2 - 3R_1 \Rightarrow R_2} \left[\begin{array}{cc|c} 1 & 2 & b_1 \\ 0 & -5 & -3b_1 + b_2 \\ 0 & 5 & b_3 \end{array} \right] \xrightarrow{R_3 + R_2 \Rightarrow R_3} \left[\begin{array}{cc|c} 1 & 2 & b_1 \\ 0 & -5 & -3b_1 + b_2 \\ 0 & 0 & -3b_1 + b_2 + b_3 \end{array} \right]$$

So, $Ax = b$ is consistent if and only if $-3b_1 + b_2 + b_3 = 0$.

Solution. (new) We determine a basis for $\text{null}(A^T)$:

$$\left[\begin{array}{ccc} 1 & 3 & 0 \\ 2 & 1 & 5 \end{array} \right] \xrightarrow{R_2 - 2R_1 \Rightarrow R_2} \left[\begin{array}{ccc} 1 & 3 & 0 \\ 0 & -5 & 5 \end{array} \right] \xrightarrow{-\frac{1}{5}R_2 \Rightarrow R_2} \left[\begin{array}{ccc} 1 & 3 & 0 \\ 0 & 1 & -1 \end{array} \right] \xrightarrow{R_1 - 3R_2 \Rightarrow R_1} \left[\begin{array}{ccc} 1 & 0 & 3 \\ 0 & 1 & -1 \end{array} \right]$$

We read off from the RREF that $\text{null}(A^T)$ has basis $\begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$.

b has to be orthogonal to $\text{null}(A^T)$. That means $b \cdot \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} = 0$. As above!

Least squares

Example 35. Not all linear systems have solutions.

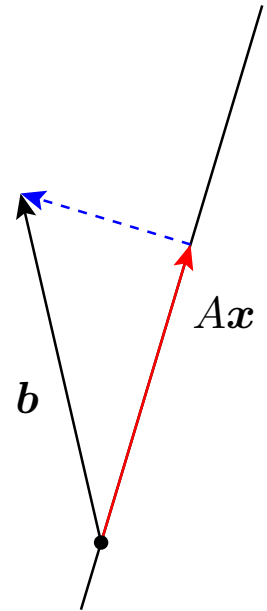
In fact, for many applications, data needs to be fitted and there is no hope for a perfect match.

For instance, $Ax = b$ with

$$\begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix} x = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

has no solution:

- $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ is not in $\text{col}(A)$ since $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} \neq 0$ (see previous example).
- Instead of giving up, we want the x which makes Ax and b as close as possible.
- Such x is characterized by the error $Ax - b$ being **orthogonal** to $\text{col}(A)$ (i.e. all possible Ax).



Definition 36. \hat{x} is a **least squares solution** of the system $Ax = b$ if \hat{x} is such that $A\hat{x} - b$ is as small as possible (i.e. minimal norm).

- If $Ax = b$ is consistent, then \hat{x} is just an ordinary solution (in that case, $A\hat{x} - b = 0$)
- Interesting case: $Ax = b$ is inconsistent. (in particular, if the system is overdetermined)

The normal equations

The following result provides a straightforward recipe (thanks to the FTLA) to find least squares solutions for any system $Ax = b$.

Theorem 37. \hat{x} is a least squares solution of $Ax = b$

$$\iff A^T A \hat{x} = A^T b \quad (\text{the normal equations})$$

Proof.

\hat{x} is a least squares solution of $Ax = b$

$$\iff A\hat{x} - b \text{ is as small as possible}$$

$$\iff A\hat{x} - b \text{ is orthogonal to } \text{col}(A)$$

$$\stackrel{\text{FTLA}}{\iff} A\hat{x} - b \text{ is in } \text{null}(A^T)$$

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

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

□

Example 38. Find the least squares solution to $A\mathbf{x} = \mathbf{b}$, where

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

Solution. First, $A^T A = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ and $A^T \mathbf{b} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

Hence, the normal equations $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$ take the form $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \hat{\mathbf{x}} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.

Solving, we immediately find $\hat{\mathbf{x}} = \begin{bmatrix} 1/2 \\ 3/2 \end{bmatrix}$.

Check. Since $A\hat{\mathbf{x}} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$, the error is $A\hat{\mathbf{x}} - \mathbf{b} = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$. Recall that the error must be orthogonal to $\text{col}(A)$!

This error is indeed orthogonal to $\text{col}(A)$ because $\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} = 0$ and $\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 0$.

Example 39. (extra) Find the least squares solution to $A\mathbf{x} = \mathbf{b}$, where

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

Solution. First, $A^T A = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 1 & 5 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 1 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 10 & 5 \\ 5 & 30 \end{bmatrix}$ and $A^T \mathbf{b} = \begin{bmatrix} 1 & 3 & 0 \\ 2 & 1 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$.

Hence, the normal equations $A^T A \hat{\mathbf{x}} = A^T \mathbf{b}$ take the form $\begin{bmatrix} 10 & 5 \\ 5 & 30 \end{bmatrix} \hat{\mathbf{x}} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$.

Since $\begin{bmatrix} 10 & 5 \\ 5 & 30 \end{bmatrix}^{-1} = \frac{1}{275} \begin{bmatrix} 30 & -5 \\ -5 & 10 \end{bmatrix} = \frac{1}{55} \begin{bmatrix} 6 & -1 \\ -1 & 2 \end{bmatrix}$, we find $\hat{\mathbf{x}} = \frac{1}{55} \begin{bmatrix} 6 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 8 \end{bmatrix} = \frac{1}{55} \begin{bmatrix} 16 \\ 12 \end{bmatrix}$.

Check. Since $A\hat{\mathbf{x}} = \frac{1}{55} \begin{bmatrix} 40 \\ 60 \\ 60 \end{bmatrix}$, the error $A\hat{\mathbf{x}} - \mathbf{b} = \frac{1}{55} \begin{bmatrix} -15 \\ 5 \\ 5 \end{bmatrix} = \frac{1}{11} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$ must be orthogonal to $\text{col}(A)$.

The error is indeed orthogonal to $\text{col}(A)$ because $\begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} \cdot \frac{1}{11} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} = 0$ and $\begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} \cdot \frac{1}{11} \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix} = 0$.