What if an equation $A\mathbf{x}=\mathbf{b}$ has no solution?
Given a vector $\hat{\mathbf{x}}$ write

$$\mathbf{e} = \mathbf{b} - A\hat{\mathbf{x}}$$

for the error vector measuring how far $\hat{\mathbf{x}}$ is from being a solution.

EX: $\begin{bmatrix} 1 & -1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ has no solution.

- The error vector for $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is

$$\mathbf{e} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}$$

- The error vector for $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is

$$\mathbf{e} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & -1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ -2 \\ 0 \end{bmatrix}$$

---

Neither vector is a solution, but $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ is closer to being a solution than $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

The "best approximate solution" to $A\mathbf{x}=\mathbf{b}$ will be the $\hat{\mathbf{x}}$ which "minimizes" error.

Note: Error $\mathbf{e} = \mathbf{b} - A\hat{\mathbf{x}}$ is a vector. In this lecture, "minimize error" will mean "minimize length of $\mathbf{e}$" because this is actually very easy to do.
... But sometimes this is not the best minimization.
Minimizing length of $\mathbf{e}$ is called "least squares"

> **Formula:** The ("least squares") best approximate solution to the equation $A\mathbf{x}=\mathbf{b}$ is given by solving the Normal Equation
>
> $$(A^T A)\hat{\mathbf{x}} = (A^T \mathbf{b})$$

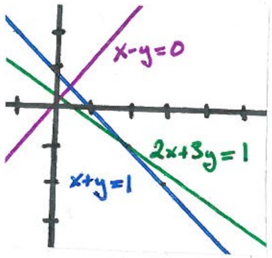
(We write $\hat{\mathbf{x}}$ instead of $\mathbf{x}$ to remind everyone that $\hat{\mathbf{x}}$ is (probably) not a solution to $A\mathbf{x}=\mathbf{b}$.)

Notes: The normal equation always has a solution — even when $A\mathbf{x}=\mathbf{b}$ doesn't.
If $A\mathbf{x}=\mathbf{b}$ has a solution, then the normal equation will have the same solution.
(i.e "best approx." solution = solution, if it exists.)

**Derivation / Example of normal equation:**

Find the best approximate solution to

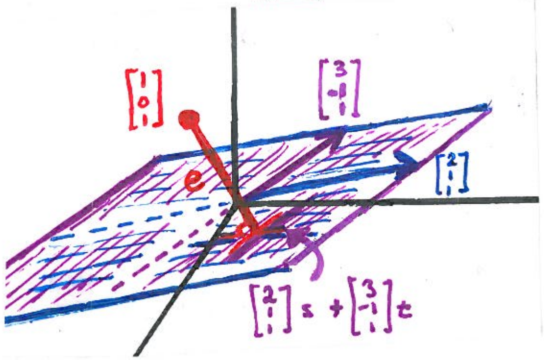
$$\left\{\begin{array}{l} 2x+3y=1 \\ x-y=0 \\ x+y=1 \end{array}\right\} \Longleftrightarrow \begin{bmatrix} 2 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$




{ No solution because these three lines do not have a shared point of intersection. }

Convert to vector equation:

$$\underbrace{\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} x + \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} y}_{\text{Considering different values of } x \,\&\, y \text{ this part gives a parametric plane in 3D space.}} = \underbrace{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}_{\text{This is a point in 3D space which is not on the parametric plane.}}$$




Choosing a value for $\hat{x}$ and $\hat{y}$ gives

$$\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}\hat{x} + \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}\hat{y}$$

a point in this plane.

The error vector

$$\mathbf{e} = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \left(\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}\hat{x} + \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}\hat{y}\right)$$

goes from this point to $\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$. Clearly the length of $\mathbf{e}$ is minimized when it is $\perp$ to the plane.

$\mathbf{e}$ will be $\perp$ to the plane when it is $\perp$ to the generating vectors $\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}$.

> Recall from MAT120: Two vectors $v$ & $w$ are $\perp$ when $v \bullet w = 0$
>
> → In matrix notation, dot products are $v^T w$
>
> EX
> $$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \bullet \begin{bmatrix} 1 \\ -1 \end{bmatrix} = (1)(1) + (2)(-1) \quad \text{(Dot product)}$$
> $$\begin{bmatrix} 1 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \end{bmatrix} = (1)(1) + (2)(-1) \quad \text{(Matrix mult.)}$$

To be $\perp$, need

$$\left.\begin{array}{l} \begin{bmatrix} 2 & 1 & 1 \end{bmatrix}\mathbf{e} = 0 \\ \begin{bmatrix} 3 & -1 & 1 \end{bmatrix}\mathbf{e} = 0 \end{array}\right\} \Longrightarrow \begin{bmatrix} 2 & 1 & 1 \\ 3 & -1 & 1 \end{bmatrix}\mathbf{e} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Simplify:

$$\begin{bmatrix} 2 & 1 & 1 \\ 3 & -1 & 1 \end{bmatrix}\left(\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 2 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix}\right) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 & 1 \\ 3 & -1 & 1 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 3 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

Normal Eqn: $(A^T \; A)\, \hat{\mathbf{x}} = (A^T \; \mathbf{b})$

$$\begin{bmatrix} 6 & 6 \\ 6 & 11 \end{bmatrix}\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$$

$$\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \underbrace{\frac{1}{66-36}\begin{bmatrix} 11 & -6 \\ -6 & 6 \end{bmatrix}}_{2\times 2 \text{ inverse matrix}}\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \frac{1}{30}\begin{bmatrix} 9 \\ 6 \end{bmatrix} = \begin{bmatrix} 3/10 \\ 2/10 \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

**Recall:** "$A^T$" is "transpose of A".

→ Matrix whose rows are the columns of A.

EX $\begin{bmatrix}1\\2\\3\end{bmatrix}^T = \begin{bmatrix}1 & 2 & 3\end{bmatrix}$

EX $\begin{bmatrix}1 & 4\\2 & 5\\3 & 6\end{bmatrix}^T = \begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\end{bmatrix}$  Column 1 ⟼ Row 1, Column 2 ⟼ Row 2

EX $\begin{bmatrix}1 & 3 & 4\\0 & 2 & 7\\0 & 0 & 5\end{bmatrix}^T = \begin{bmatrix}1 & 0 & 0\\3 & 2 & 0\\4 & 7 & 5\end{bmatrix}$  upper Δ ⟼ lower Δ (transpose does not change diagonal)

Multiplication is (rows)•(columns), but rows($A^T$) = cols($A$) so $A^TA$ is actually a very simple matrix:

$$A^TA = \begin{bmatrix}| & | & | & \\ c_1 & c_2 & c_3 & \cdots\\ | & | & | & \end{bmatrix}^T \begin{bmatrix}| & | & | & \\ c_1 & c_2 & c_3 & \cdots\\ | & | & | & \end{bmatrix}$$

$$= \begin{bmatrix}— c_1 —\\ — c_2 —\\ — c_3 —\\ \vdots\end{bmatrix}\begin{bmatrix}| & | & | & \\ c_1 & c_2 & c_3 & \cdots\\ | & | & | & \end{bmatrix} = \begin{bmatrix}c_1\cdot c_1 & c_1\cdot c_2 & c_1\cdot c_3 & \cdots\\ c_2\cdot c_1 & c_2\cdot c_2 & c_2\cdot c_3 & \cdots\\ c_3\cdot c_1 & c_3\cdot c_2 & c_3\cdot c_3 & \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

$\boxed{A^TA = \text{"dot products of columns of } A \text{ with each other"}}$

**EX:** $A = \begin{bmatrix}1 & 3\\2 & 0\\-1 & 2\end{bmatrix}$ ($c_1$, $c_2$) then $A^TA = \begin{bmatrix}\overset{c_1\cdot c_1}{1+4+1} & \overset{c_1\cdot c_2}{3+0-2}\\ \underset{c_2\cdot c_1}{3+0-2} & \underset{c_2\cdot c_2}{9+0+4}\end{bmatrix}$

Similarly, $A^Tb$ is the dot products of columns of A with b.

**EX:** Find the best approx. solution to $\underbrace{\begin{bmatrix}1 & -1\\1 & 0\\0 & 1\end{bmatrix}}_{A}\begin{bmatrix}x\\y\end{bmatrix} = \underbrace{\begin{bmatrix}3\\0\\3\end{bmatrix}}_{b}$

$A^TA = \begin{bmatrix}2 & -1\\-1 & 2\end{bmatrix}$  $A^Tb = \begin{bmatrix}3\\0\end{bmatrix}$  $\begin{bmatrix}1\\1\\0\end{bmatrix}\cdot\begin{bmatrix}3\\0\\3\end{bmatrix}$, $\begin{bmatrix}-1\\0\\1\end{bmatrix}\cdot\begin{bmatrix}3\\0\\3\end{bmatrix}$

**Normal Equation:** $\begin{bmatrix}2 & -1\\-1 & 2\end{bmatrix}\begin{bmatrix}\hat{x}\\\hat{y}\end{bmatrix} = \begin{bmatrix}3\\0\end{bmatrix}$

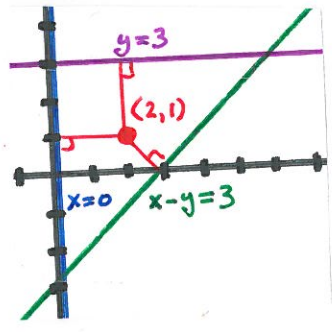
(solve using 2x2 inverse) $\begin{bmatrix}\hat{x}\\\hat{y}\end{bmatrix} = \frac{1}{4-1}\begin{bmatrix}2 & 1\\1 & 2\end{bmatrix}\begin{bmatrix}3\\0\end{bmatrix} = \begin{bmatrix}2\\1\end{bmatrix}$

---

Two ways to understand meaning of $\begin{bmatrix}\hat{x}\\\hat{y}\end{bmatrix}$:

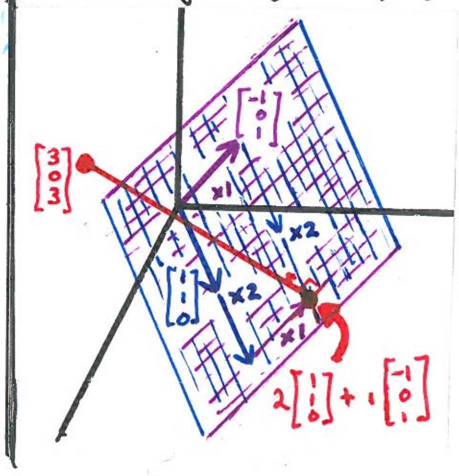
(1) The system of equations $\left\{\begin{matrix}x-y=3\\x=0\\y=3\end{matrix}\right\}$ has no common intersection point.

$(\hat{x},\hat{y}) = (2,1)$ minimizes distance to these three lines




(2) $\begin{bmatrix}3\\0\\3\end{bmatrix}$ is not on plane $\begin{bmatrix}1\\1\\0\end{bmatrix}x + \begin{bmatrix}-1\\0\\1\end{bmatrix}y$

$\hat{x}=2$ & $\hat{y}=1$ gives projection (which minimizes distance to plane)



**EX:** Find best approx. solution & error vector for

$$\begin{bmatrix} 1 & 1 \\ -1 & 3 \\ 0 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 4 \\ 6 \end{bmatrix}$$

$$A^TA = \begin{bmatrix} 1+1+0+4 & 1-3+0+0 \\ 1-3+0+0 & 1+9+4+0 \end{bmatrix} = \begin{bmatrix} 6 & -2 \\ -2 & 14 \end{bmatrix}$$

$$A^Tb = \begin{bmatrix} 2-1+0-2 \\ 2+3-6+0 \end{bmatrix} = \begin{bmatrix} 16 \\ 8 \end{bmatrix}$$

Normal Eqn: $\begin{bmatrix} 6 & -2 \\ -2 & 14 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} 16 \\ 8 \end{bmatrix}$

Approx. Solution: $\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \frac{1}{84-4} \begin{bmatrix} 14 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 16 \\ 8 \end{bmatrix} = \boxed{\begin{bmatrix} 3 \\ 1 \end{bmatrix}}$

Error: $e = \begin{bmatrix} 3 \\ -1 \\ 4 \\ 6 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ -1 & 3 \\ 0 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 4 \\ 6 \end{bmatrix} - \begin{bmatrix} 4 \\ 0 \\ 2 \\ 6 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 2 \\ 0 \end{bmatrix}$

Length of Error: $\sqrt{(-1)^2 + (-1)^2 + 2^2 + 0^2} = \sqrt{6}$

---

**Note:** The projection of $b$ onto the columns of $A$ is given by $A\hat{x}$ where $\hat{x}$ solves the normal equation.

**EX:** The projection of $\begin{bmatrix} 3 \\ -1 \\ 4 \\ 6 \end{bmatrix}$ onto $\begin{bmatrix} 1 & 1 \\ -1 & 3 \\ 0 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$ is $\begin{bmatrix} 4 \\ 0 \\ 2 \\ 6 \end{bmatrix}$

**EX:** Find best approx. solution and error vector for

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

Normal Eqn $\begin{bmatrix} 7 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \begin{bmatrix} 11 \\ -4 \end{bmatrix}$

Approx. Solution $\begin{bmatrix} \hat{x} \\ \hat{y} \end{bmatrix} = \frac{1}{13} \begin{bmatrix} 2 & -1 \\ -1 & 7 \end{bmatrix} \begin{bmatrix} 11 \\ -4 \end{bmatrix} = \boxed{\begin{bmatrix} 2 \\ -3 \end{bmatrix}}$

Error: $e = \begin{bmatrix} 2 \\ 1 \\ 2 \\ -5 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 2 \\ -5 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \\ 2 \\ -5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

---

Because the problem above actually had a solution, the best approximate solution was the solution. The minimum possible error was 0, so that is what we found!

In this case the vector $\begin{bmatrix} 2 \\ 1 \\ 2 \\ -5 \end{bmatrix}$ was already in the plane $\begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$, so it is equal to its own projection!

EX: Calculate the projection of

$$b = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \text{ onto the plane } \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} s + \begin{bmatrix} -1 \\ 2 \\ 0 \\ 1 \end{bmatrix} t$$

First solve the normal equation for

$$\begin{bmatrix} 1 & -1 \\ 0 & 2 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

Normal Equation: $\begin{bmatrix} 3 & 0 \\ 0 & 6 \end{bmatrix} \begin{bmatrix} \hat{s} \\ \hat{t} \end{bmatrix} = \begin{bmatrix} 8 \\ 7 \end{bmatrix}$

$$\begin{bmatrix} \hat{s} \\ \hat{t} \end{bmatrix} = \frac{1}{18} \begin{bmatrix} 6 & 0 \\ 0 & 3 \end{bmatrix} \cdot \begin{bmatrix} 8 \\ 7 \end{bmatrix} = \begin{bmatrix} 8/3 \\ 7/6 \end{bmatrix}$$

To get the projection, plug in for s & t

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} 8/3 + \begin{bmatrix} -1 \\ 2 \\ 0 \\ 1 \end{bmatrix} 7/6 = 1/6 \begin{bmatrix} 9 \\ 14 \\ 16 \\ 23 \end{bmatrix}$$

Summary:

- Normal Equation is useful when
  - → System has no solution
  - → System has #rows >> # columns
- $A^TA$ is
  - → square, symmetric matrix with same #cols as A (if A has many rows then $A^TA$ is smaller)
  - → elements are dot products of columns of A with each other
- $A^Tb$ is
  - → dot product of columns of A with b

Normal Equation is easy to compute & write.

Given an equation $A\underline{x} = \underline{b}$

→ Best approx. solution is $(A^TA)\hat{x} = A^Tb$

→ Projection of b onto Ax is $p = A\hat{x}$

→ Error of approx. solution is $e = b - A\hat{x}$