LECTURE 26: LEAST-SQUARES

Monday, November 25, 2019 3:36 PM

Welcome to the culmination of our Math 3A adventure, the reason why we've been working so hard all quarter!
Because so far we have done the improbable, but today we will do the impossible: we'll solve systems... that have no solutions!

I- PRELUDE

Example: Consider the system

$$\underbrace{\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} 5 \\ 6 \end{bmatrix}}_{\mathbf{b}}$$

$$\Leftrightarrow \quad x\begin{bmatrix} 1 \\ 2 \end{bmatrix} + y\begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix}$$

$$\Leftrightarrow \quad \mathbf{b} = \begin{bmatrix} 5 \\ 6 \end{bmatrix} \text{ is in Span} \left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \end{bmatrix} \right\} = \text{Col}(A)$$

$\boxed{\text{Fact: } A\mathbf{x} = \mathbf{b} \text{ is consistent} \Leftrightarrow \mathbf{b} \text{ is in Col}(A)}$

(And this fact will help us solve inconsistent systems)

Today: Will present you 2 methods for solving inconsistent systems, need to know both

II- THE SLOW WAY (METHOD 1)

(Slow, but more intuitive)

Example: "Solve" $Ax = b$

$$A = \begin{array}{c} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} \\ \begin{matrix} u & \;\; v \end{matrix} \end{array} \qquad b = \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix}$$

Need to be orthogonal!!!

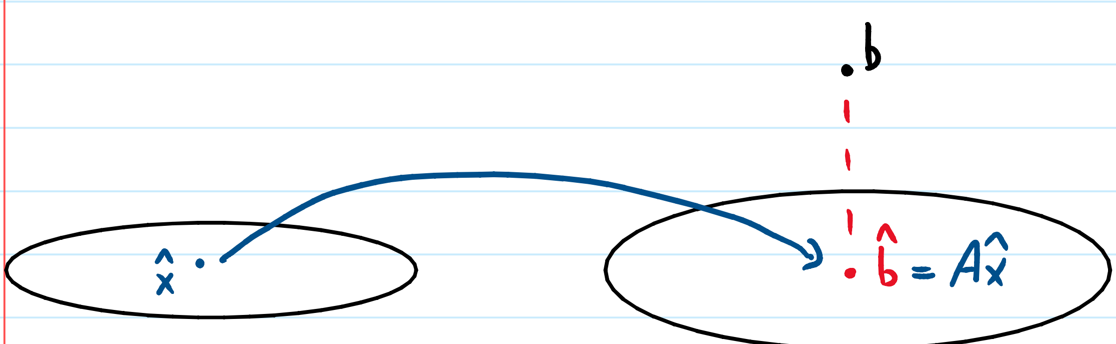
NO SOLUTION!!! 😨

But what if you have a job and your boss tells you "Solve this system or you'll be fired" ??? Than what could you do?

Goal: Find the best non-solution! (= best substitute for a solution)

Important observation:

The reason $Ax = b$ has no solution is because b is not in Col(A)



Col(A)

So just force b to be in Col(A) by calculating $\hat{b}$!

STEP 1

Calculate $\hat{b}$ = OP of b on Col(A) = Span{u,v}

$$\hat{b} = \left(\frac{b \cdot u}{u \cdot u}\right) u + \left(\frac{b \cdot v}{v \cdot v}\right) v$$ (because u and v are $\perp$)

$$= \frac{\begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix}} \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} + \frac{\begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix}} \begin{bmatrix} 1 \\ -1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \underbrace{\left(\frac{8}{4}\right)}_{2} \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix} + \underbrace{\frac{12}{4}}_{3} \begin{bmatrix} 1 \\ 1 \\ -1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 5 \\ -1 \\ 1 \\ 5 \end{bmatrix}$$

Upshot: Since $\hat{b}$ is in Col(A), $Ax = \hat{b}$ HAS a solution!

STEP 2 Solve $Ax = \hat{b}$

$$\begin{array}{c} \begin{array}{cc|c} \quad A & & \hat{b} \end{array} \\ \left[\begin{array}{cc|c} 1 & 1 & 5 \\ 1 & -1 & -1 \\ -1 & 1 & 1 \\ 1 & 1 & 5 \end{array}\right] \end{array} \dashrightarrow \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

$$\boxed{\hat{x} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}}$$

-> LEAST SQUARES SOLUTION OF $Ax = b$
(= "best" non-solution, in the following sense:)

(b) Find the least-squares error

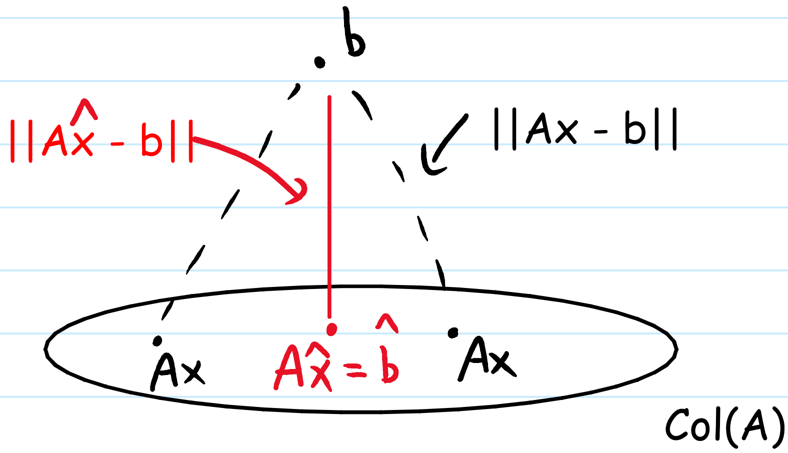
Answer: $||A\hat{x} - b|| = ||\hat{b} - b|| = \left|\left| \begin{bmatrix} 5 \\ -1 \\ 1 \\ 5 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \\ 6 \\ 8 \end{bmatrix} \right|\right|$

$$= \left|\left| \begin{bmatrix} 3 \\ -5 \\ -5 \\ -3 \end{bmatrix} \right|\right|$$

$$= \sqrt{68}$$

Interpretation: Among all the possible values of x, it's $\hat{x}$ that makes the error $||Ax - b||$ the smallest

Picture:



In fact, if $Ax = b$ actually has a solution, then $Ax - b = 0$
=> $||Ax - b|| = 0$ (no error)

III- THE FAST WAY (METHOD 2)

Works for ANY A (not just with orthogonal columns), but not 100 % clear why it works (see end of notes if you're curious why it works)

Example: (a) Find the least-squares solution to $Ax = b$

$$A = \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \qquad b = \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix}$$

TRICK: Multiply $Ax = b$ by A^T

$$A^T A\, \mathbf{x} = A^T b$$

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

$$\begin{bmatrix} 3 & 3 \\ 3 & 11 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 6 \\ 14 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} 3 & 3 & 6 \\ 3 & 11 & 14 \end{array}\right] \dashrightarrow \left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right]$$

$$\hat{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$ WOW!!!

(b) Find the least-squares error

$$||A\hat{x} - b|| = \left|\left| \begin{bmatrix} 1 & 3 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix} \right|\right|$$

$$= \left|\left| \begin{bmatrix} 4 \\ 0 \\ 2 \end{bmatrix} - \begin{bmatrix} 5 \\ 1 \\ 0 \end{bmatrix} \right|\right|$$

$$= \left|\left| \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix} \right|\right|$$

$$= \sqrt{6}$$

Do we always find a solution with this method? Not always, but:

Fact: If the columns of A are LI, then there is always a unique least-squares solution.

IV- LEAST-SQUARES AND QR

Last time we found the QR decomposition of a matrix, and now we can finally see why it's useful!

Example: Use $A = QR$ to find the least-squares solution of $Ax = b$

$$A = \begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 1 & 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 2/3 & -1/3 \\ 2/3 & 2/3 \\ 1/3 & -2/3 \end{bmatrix}}_{\substack{Q \\ \text{(orthonormal} \\ \text{columns)}}} \underbrace{\begin{bmatrix} 3 & 5 \\ 0 & 1 \end{bmatrix}}_{\substack{R \\ \text{(upper} \\ \text{triangular)}}} \qquad b = \begin{bmatrix} 9 \\ 9 \\ 9 \end{bmatrix}$$

(We found Q by applying the G-S process to the columns of A, and we would R using $R = Q^T A$)

Notice: If $A = QR$, then

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

$$\Rightarrow (QR)^T (QR) \hat{x} = (QR)^T b$$

$$\Rightarrow R^T \underbrace{Q^T Q}_{I} R \hat{x} = R^T Q^T b$$

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

But $R^T = \begin{bmatrix} 3 & 0 \\ 5 & 1 \end{bmatrix}$ invertible! (not always, but usually the case)

$$\Rightarrow \cancel{R^T} R \hat{x} = \cancel{R^T} Q^T b$$

$$\Rightarrow \boxed{R \hat{x} = Q^T b}$$

$$\Rightarrow \boxed{R\hat{x} = Q^T b}$$

$$\begin{bmatrix} 3 & 5 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2/3 & 2/3 & 1/3 \\ -1/3 & 2/3 & -2/3 \end{bmatrix} \begin{bmatrix} 9 \\ 9 \\ 9 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 5 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 15 \\ -3 \end{bmatrix}$$

SUPER easy to solve! (because R is upper triangular), don't even need row-reductions!

$$\begin{cases} 3x + 5y = 15 \quad \Rightarrow 3x = 15 - 5y = 15 - 5(-3) = 30 \Rightarrow x = 10 \\ y = -3 \end{cases}$$

$$\boxed{\hat{x} = \begin{bmatrix} 10 \\ -3 \end{bmatrix}}$$

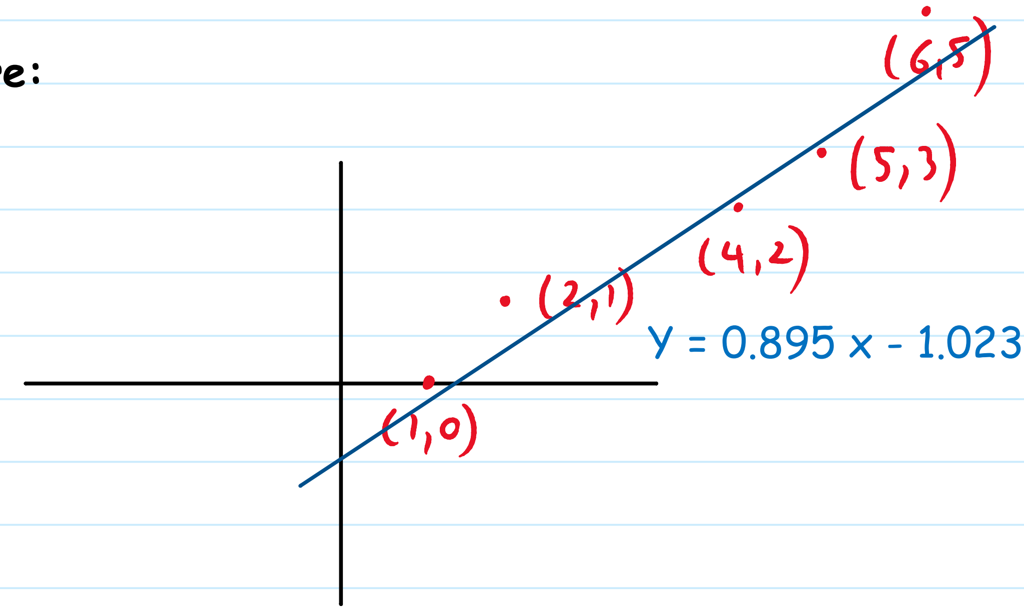
So QR helps us solve least-squares in an easy way!
(And that's in fact how computers solve least-squares problems)

V- LINEAR MODELS

I've saved the best application for last, because here's the single most useful application in linear algebra. You literally see this everywhere in your life!

Example: Find the equation of the line that best fits the points (1,0), (2,1), (4,2), (5,3), (6,5)

Picture:



(This is the line that Excel or R gives you when you ask to fit the points on a line)

Suppose the points actually **are** on the line y = ax + b

Then
$$\begin{cases} a(1) + b = 0 & \rightsquigarrow (1,0) \\ a(2) + b = 1 & \rightsquigarrow (2,1) \\ a(4) + b = 2 \\ a(5) + b = 3 \\ a(6) + b = 5 \end{cases}$$

$$\underbrace{\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \end{bmatrix}}\underbrace{\begin{bmatrix} a \\ b \end{bmatrix}} = \underbrace{\begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 5 \end{bmatrix}}$$

$A \qquad\qquad x \qquad\qquad b$

Solve using least-squares:

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

...

$$\Longrightarrow \quad \begin{bmatrix} 82 & 18 \\ 18 & 82 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 55 \\ 11 \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 77/86 \\ -44/43 \end{bmatrix} \approx \begin{bmatrix} 0.895 \\ -1.023 \end{bmatrix}$$

Answer: $\boxed{y = ax + b \approx 0.895x - 1.023}$

And with this we're officially done with the material of the course!!!

Congratulations, you made it!!!

The End

OPTIONAL: Why the Fast Method Works:

Suppose $A = QR$

Then $\hat{b}$ = OP of b on Col(A)   (= Col(Q), by construction of Q)

$= QQ^T b$ (since the columns of Q are orthonormal)

So $\hat{b} = QQ^T b$

So $A\hat{x} = \hat{b}$ (METHOD 1)

$\Rightarrow A\hat{x} = QQ^T b$

$\Rightarrow A^T A\hat{x} = A^T QQ^T b$

But $A^T QQ^T b = (QR)^T QQ^T b$

$$= R^T \underbrace{Q^T Q}_{I} Q^T b$$

$$= R^T Q^T b$$

$$= (QR)^T b$$

$$= A^T b$$

$\Rightarrow \boxed{A^T A\hat{x} = A^T b}$ (METHOD 2)