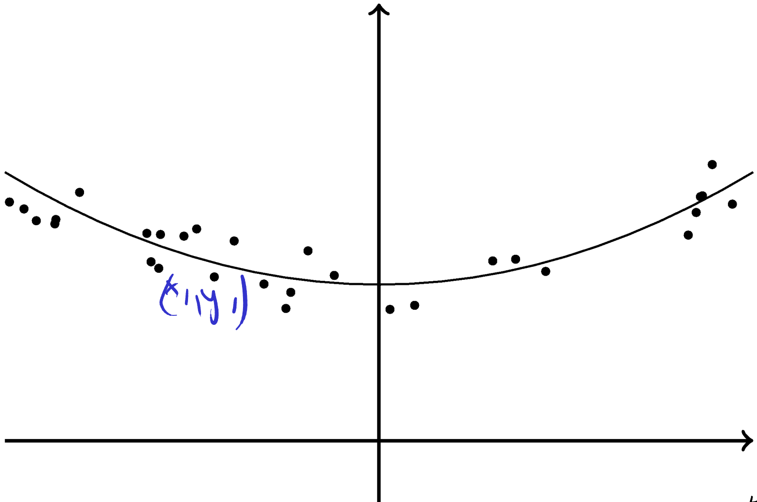
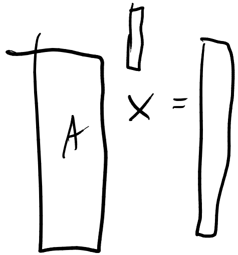
# Example: Data Fitting




Have data: $(x_i, y_i)$ and model:

$$y(x) = a + bx + cx^2$$




- Find data that best fit model:

$$\begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

$Ax \cong b$ $\qquad \|Ax - b\|_2^2$

## **Demo:** Interactive Polynomial Fitting

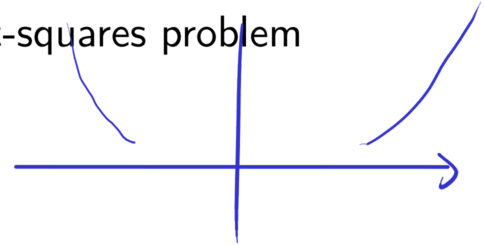
## Properties of Least-Squares

Consider the least squares problem $Ax \cong b$ and its associated objective function

$$\varphi(x) = \|b - Ax\|_2^2. \geq 0$$

- Is there always a solution to a linear least-squares problem

  $\varphi \geq 0 \quad \varphi \to \infty \ (\|x\| \to \infty) \Rightarrow \exists$ minimum



- Is it unique?

  If $A$ has a nullspace, then no.

- Examine the objective function, find its minimum.

$$\varphi(x) = (b - Ax)^T (b - Ax)$$

$$= b^T b - 2x^T A^T b + x^T A^T A x$$

$$\nabla\varphi(x) = 0 \quad -2A^T b + 2A^T A x \overset{!}{=} 0 \rightsquigarrow A^T A x = A^T b$$

Normal equations

**In-class activity:** Least Squares

**Demo:** Polynomial fit using the normal equations

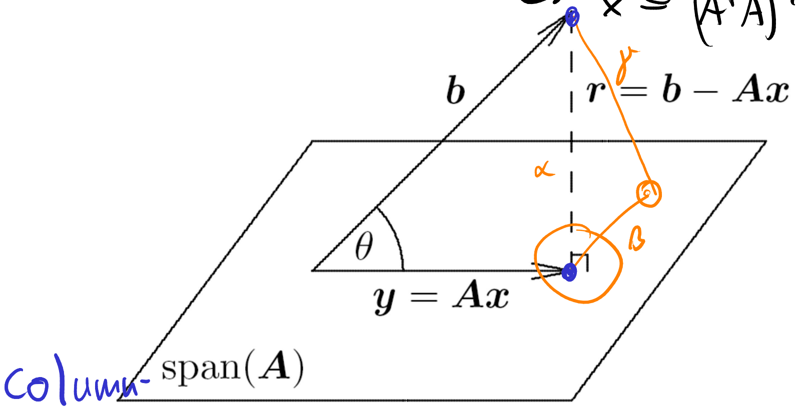
- What's the shape of $A^T A$?

**Demo:** Issues with the normal equations

## Least Squares, Viewed Geometrically

$Ay = y = Pb = A(A^TA)^{-1}A^Tb$

$x = (A^TA)^{-1}A^Tb$




$\alpha^2 + \beta^2 = \gamma^2$

- Why is $\boldsymbol{r} \perp \operatorname{span}(A)$ a good thing to require?

  (Pythagoras)

- Phrase that as an equation.

$$\underbrace{\begin{pmatrix} \equiv \end{pmatrix}}^{A^T} r = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \to 0 = A^T(b - Ax) = A^Tb - A^TAx$$

- Write that with an orthogonal projection matrix $P$.

$$Ax = Pb$$

## About Orthogonal Projectors

- What is a projector?

  $P^2 = P$

- What is an orthogonal projector?

  $(Py - y) \perp Py \rightarrow P$ symmetric, i.e. $P^T = P$

- How do I make one projecting onto span$\{\boldsymbol{q}_1, \boldsymbol{q}_2, \dots, \boldsymbol{q}_\ell\}$?

  Assume (or make) $q_1, \dots, q_\ell$ orthonormal. (Gram-Schmidt)

  $QQ^Tx$

- Check that $P = A(A^TA)^{-1}A^T$ is an orthogonal projector onto colspan$(A)$.

  $P^2 = A(A^TA)^{-1}\cancel{A^T A(A^TA)^{-1}}A^T = A(A^TA)^{-1}A^T = P$ ✓

- What assumptions do we need to define the $P$ from the last question?

  $(A) \qquad (\; A^T \;)(A) = (\quad)$

  $(A^TA)$ is only invertible if A has full rank.

## Pseudoinverse

- What is the pseudoinverse of $A$?

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

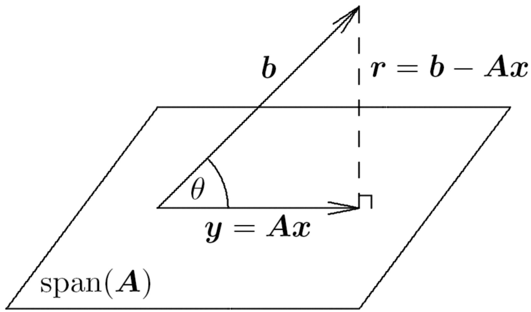
$(A^T A^{-1}) A^T$ solves $Ax \cong b$

- What can we say about the condition number in the case of a tall-and-skinny, full-rank matrix?

- What does all this have to do with solving least squares problems?

## 3.2 Sensitivity and Conditioning

## Sensitivity and Conditioning of Least Squares




- How sensitive is it?

- What values of $\theta$ are bad?

- Any comments regarding dependencies?