Least Squares and <u>functional</u> data fitting

PROBLEM: Given data
$(x_1, y_1), \ldots, (x_n, y_n)$

Find a function f, $y = f(x)$ which successfully predicts y as a function of x for x <u>not</u> in the data (or training) set.

<u>Method 1</u>: We have n-data points so

find $p(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}$ n variables c_i

with $p(x_i) = y_i$ n data points.

in terms of linear equations

$$p(x_1) = c_0 + c_1 x_1 + \ldots + c_{n-1} x_1^{n-1} = y_1$$

$$\vdots$$

$$p(x_n) = c_0 + c_1 x_n + \ldots + c_{n-1} x_n^{n-1} = y_n$$

as a matrix

$$\begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ & \vdots & & \\ 1 & x_n & \cdots & x_n^{n-1} \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\mathbb{X} \vec{c} = \vec{y}, \quad \mathbb{X} = \text{Vandermode matrix}$$

$$\mathbb{X}\vec{c} = y$$

<u>FACT</u>: if all the x_i's are distinct

$\Rightarrow$ $\mathbb{X}$ is invertible

$\Rightarrow$ unique soln $\vec{c}$ $\Rightarrow$ unique poly $P(x)$ which exacts fits the data.

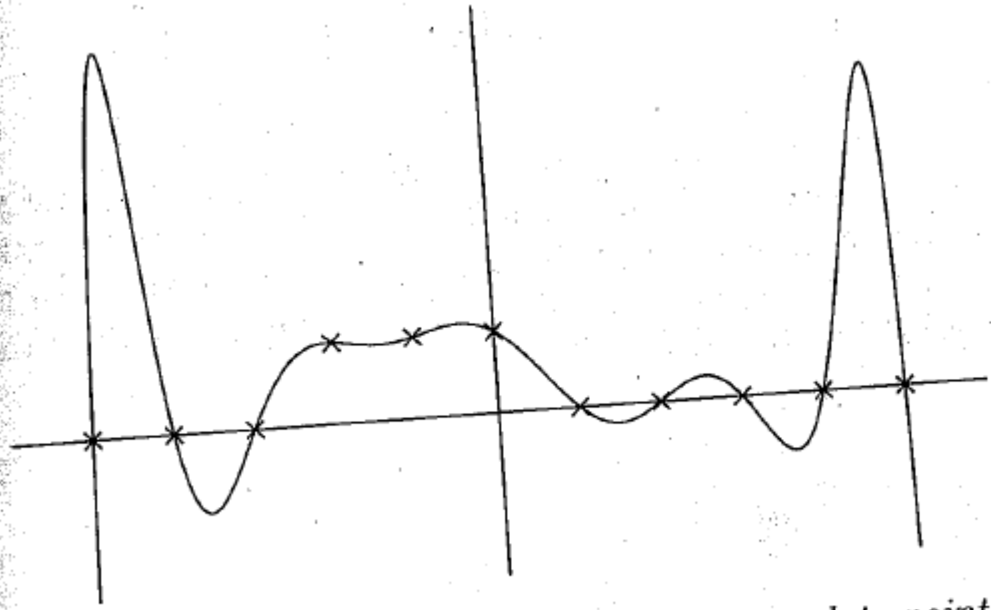
This approach overfits the data and doesn't generalize well to non-data point. — Pix

<u>Method 2</u>: Reduce the degree of the polynomial and do a best fit

degree of $p(x) = n$

data set has $m > n$ points

e reflection of the data.



11.1. *Degree* 10 *polynomial interpolant to eleven data points. The axis* **are** *not given, as these have no effect on the picture.*

the polynomial of Example 11.1. Though one cannot see this ... □
is also less sensitive to perturbations.

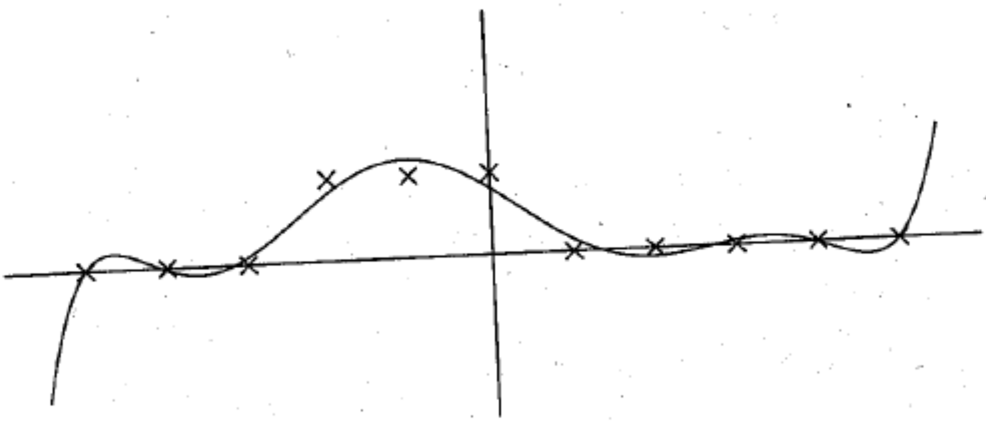


Figure 11.2. *Degree* 7 *polynomial least squares fit to the same eleven data points.*

ffrom Trefethen and Bau

As a matrix equation

$$m \begin{bmatrix} 1 & x_1 & \cdots & x_1^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \cdots & x_m^{n-1} \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_{n-1} \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$X \vec{c} = \vec{y}$$

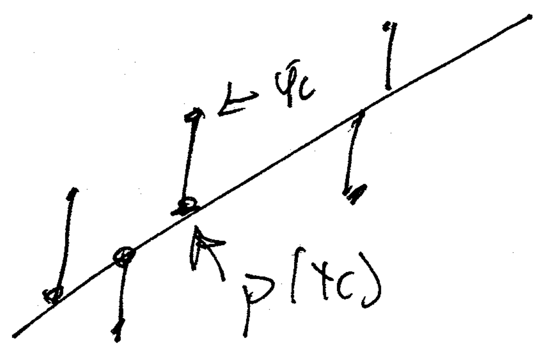
X is $m \times n$, <u>no soln</u> usually

$m > n$

Find $\vec{c}$ which minimizes $\| X\vec{c} - \vec{y} \|_2^2$

OR Find a $p(x)$ which minimizes

$$\sum_{i=1}^{m} |p(x_i) - y_i|^2 \iff \|X\vec{c} - \vec{y}\|_2^2$$

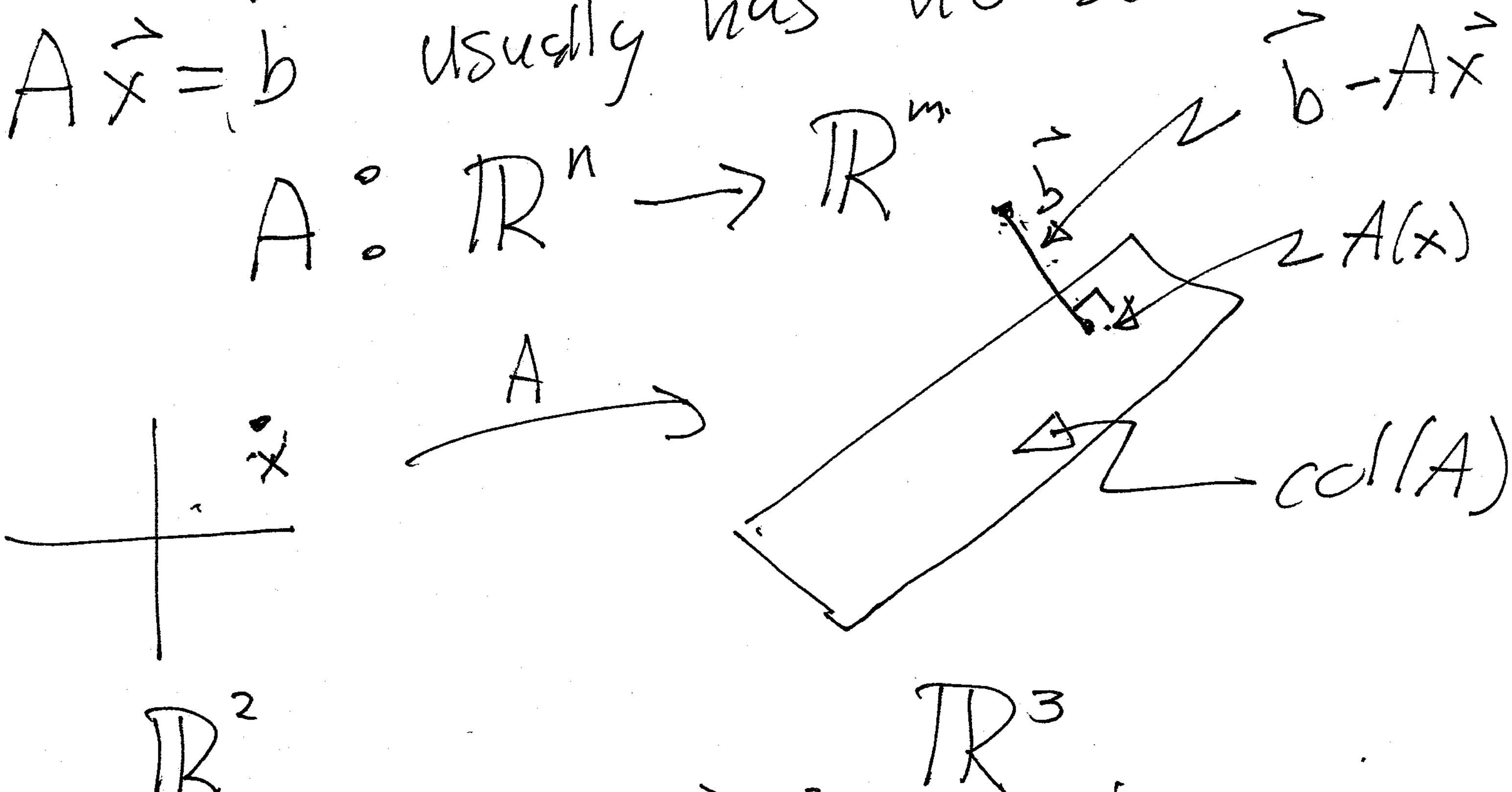


This is an example of the general
Linear least squares problem

A is $m \times n$ $\quad m \geq n$

$A\vec{x} = \vec{b}$ usually has no soln.

$A : \mathbb{R}^n \rightarrow \mathbb{R}^m$



$\|A\vec{x}-\vec{b}\|_2^2$ will be min.

Theorem: If A has rank n (i.e. lin ind col.). $\Rightarrow$ the $\vec{x}$ that minimize $\|A x - \vec{b}\|_2^2$ is the unique soln to the normal equations

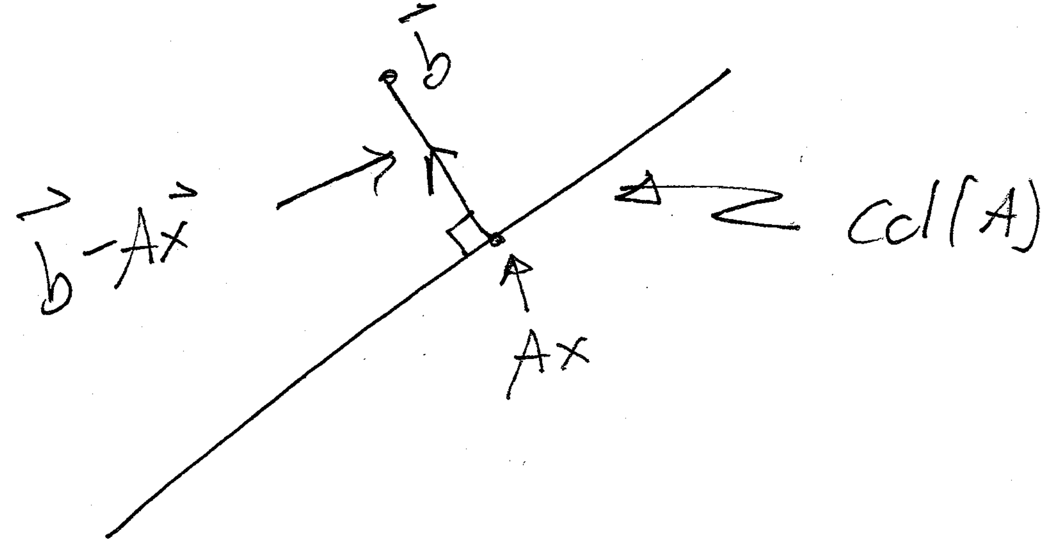
m >= n

$$\boxed{A^T A \vec{x} = A^T \vec{b}}$$

Proof: one method, use calculus to minimize the cost function

$$\Phi(\vec{x}) = \|\vec{b} - A\vec{x}\|_2^2$$

Geometric proof



$\vec{r} = \vec{b} - A\vec{x}$ is the <u>residual</u>

At the soln, $\vec{r} \perp \text{col}(A)$

if $A = [\vec{a}_1 \ldots \vec{a}_n] \Rightarrow$

$0 = \vec{r}\cdot\vec{a}_1 = \vec{r}\cdot\vec{a}_2 = \ldots = \vec{r}\cdot\vec{a}_n$

in terms of matrices

$$\begin{bmatrix} \vec{a}_1^{\,T} \\ \vdots \\ \vec{a}_n^{\,T} \end{bmatrix} \vec{r} = \begin{bmatrix} \vec{a}_1 \cdot \vec{r} \\ \vdots \\ \vec{a}_n \cdot \vec{r} \end{bmatrix} = \vec{0}$$

or $$A^T \vec{r} = \vec{0}$$

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

$$\boxed{A^T \vec{b} = A^T A \vec{x}}$$

end of proof

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

<u>Last time</u>: $A^T A$ is invertible $\Longleftrightarrow$
A has rank n (i.e. lin. ind col.).

When that is true, mult both sides
of the normal eq by $(A^T A)^{-1}$

$$\vec{x} = \underbrace{(A^T A)^{-1} A^T}_{\text{pseudo inverse of } A} \vec{b}$$, <u>unique soln</u>.

written $A^+ = (A^TA)^{-1}A^T$ when A has full rank.. So the unique least squares soln is $\boxed{A^+\vec{b} = \vec{x}}$

Notes on A^+:

(0) A^+ is $n \times m$.

(1) It is a left inverse

$$A^+A = \left[(A^TA)^{-1}A^+\right]A$$
$$= (A^TA)^{-1}(A^+A) = I.$$

(2) usually not Right inverse

(3) A^+ is computable from the SVD.

if $A = U \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_n \\ \hline & 0 & \end{bmatrix} V^T$ ($\sigma_n \neq 0$, since A is full rank by assumption)

$$\Rightarrow A^+ = V \underbrace{\left[\begin{array}{ccc|c} \sigma_1^{-1} & & & \\ & \ddots & & 0 \\ & & \sigma_n^{-1} & \end{array}\right]}_{\Sigma^+} U^T$$

more compactly, if $A = U \Sigma V^T$

$\Rightarrow A^+ = V \Sigma^+ U^T$

(4) So once you have the SVD of $A \Rightarrow$ can solve the normal equations

As $\vec{x} = V\Sigma^{+}U^{T}\vec{b}$

---

other methods of solving the normal equations.

(1) $\boxed{QR}$ decomposition of A.

(2) Cholesky decomposition of

$A^TA \leftarrow$ symmetric matrix

$A^TA = LL^T$ with L lower triangular