Today: • LS review & Examples
• Intro to Matching Pursuit.

Summary of LS:

The set of solutions to the LS problem:

$$\left\{ \min_{\vec{x}} \|A\vec{x} - \vec{b}\|^2 \right\}$$

is given by the set of solutions to:

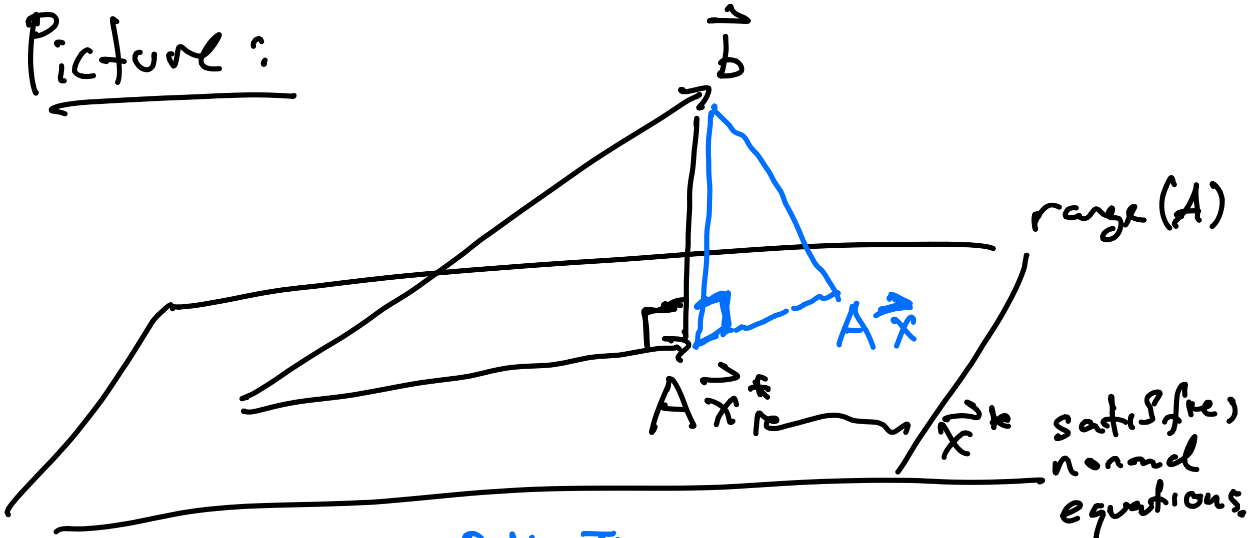
$$\left\{ (A^T A)\vec{x} = A^T \vec{b} \quad \left(\begin{array}{c}\text{normal Eqns} \\ \text{Always consistent}\end{array}\right) \right\}$$

In particular, if $N(A) = \{\vec{0}\}$, then

$$\vec{x}^* = (A^T A)^{-1} A^T \vec{b}.$$

Ex: $A = \vec{a}$, $x^* = \dfrac{\langle \vec{a}, \vec{b} \rangle}{\langle \vec{a}, \vec{a} \rangle}$ $\langle \vec{a}, \vec{b} \rangle \leftarrow A^T \vec{b}$, $\langle \vec{a}, \vec{a} \rangle \leftarrow (A^T A)^{-1}$

Picture:



$$\|A\vec{x}-\vec{b}\|^2 \overset{\text{Pyth Thm}}{=} \|A\vec{x}^*-\vec{b}\|^2 + \|A\vec{x}-A\vec{x}^*\|^2$$

$$\geq \|A\vec{x}^*-\vec{b}\|^2$$

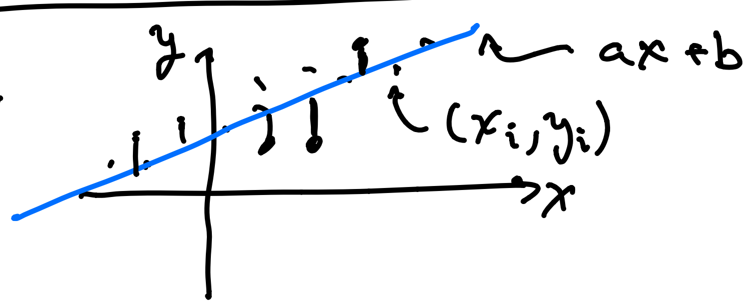
To verify desired orthogonality, just use normal equations:

$$\langle \vec{b}-A\vec{x}^*, A\vec{x}\rangle \overset{\text{def}}{=} \vec{x}^TA^T\vec{b} - \vec{x}^T\underbrace{A^TA\vec{x}^*}$$

$$\overset{\text{normal eqns}}{=} \vec{x}^TA^T\vec{b} - \vec{x}^T\underbrace{A^T\vec{b}}$$

$$= 0.$$

Ex 1:



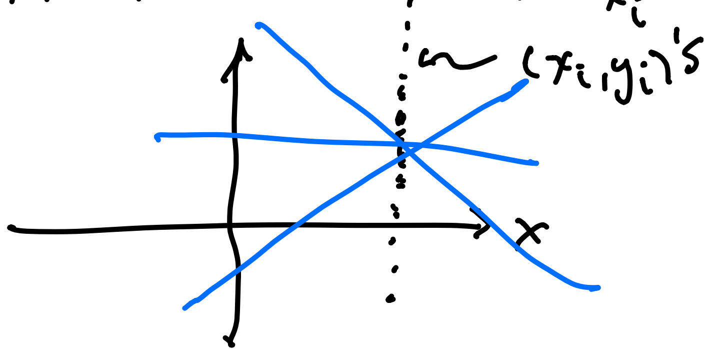
LS problem:

$$\min_{a,b} \left\| \underbrace{\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_m & 1 \end{bmatrix}}_{A} \begin{bmatrix} a \\ b \end{bmatrix} - \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}}_{\vec{b}} \right\|^2$$

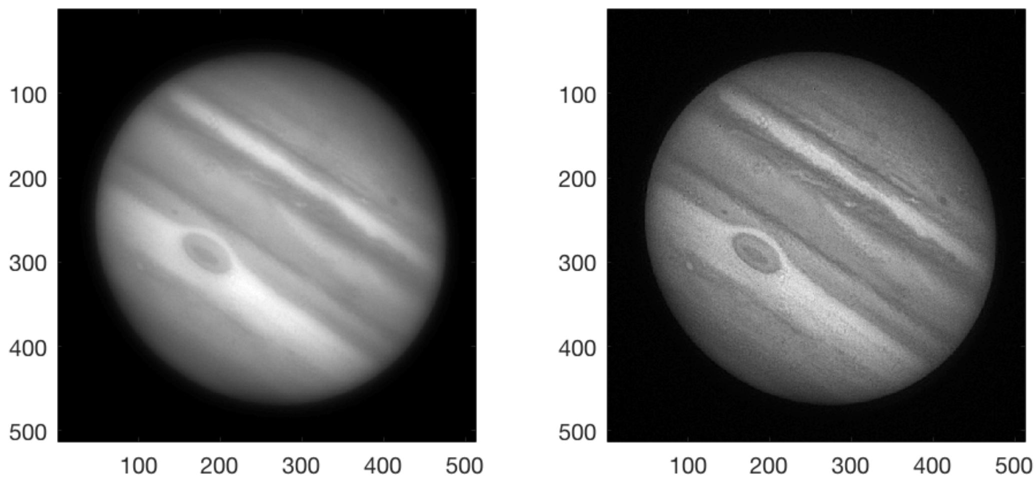
$$= \sum_{i=1}^{m} \underbrace{(ax_i + b - y_i)}^2$$

vertical distances btwn data points and line $ax+b$.

Rmk: To get some intuition about what LS problems look like when $N(A) \neq \{\vec{0}\}$, consider Linear Regression example. In this case, all x_i's are equal,



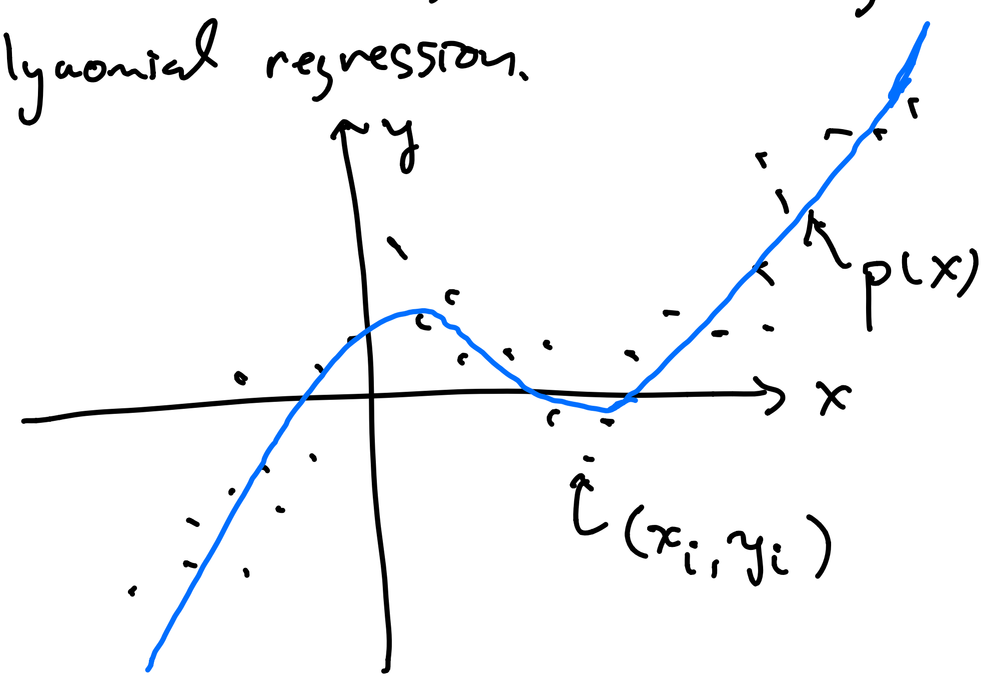
Ex: Hubble Space telescope



<u>New Examples</u>:

Common Misconception: LS problems look "linear". i.e, like linear regression.

<u>Ex</u>: polynomial regression.



$$p(x) = \underset{=}{p_0} + \underset{=}{p_1} x + \underset{=}{p_2} x^2 + \underset{=}{p_3} x^3$$

"for all"

Want:

$$p(x_i) = p_0 + p_1 x_i + p_2 x_i^2 + p_3 x_i^3 \approx y_i \quad \forall i.$$

variables

system of LEQs in variables $p_0 \cdots p_3$.

LS problem:

$$\min_{p_0, p_1, p_2, p_3} \underbrace{\left\| \overbrace{\begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_m & x_m^2 & x_m^3 \end{bmatrix}}^{A} \overset{\vec{p}}{\begin{bmatrix} p_0 \\ p_1 \\ p_2 \\ p_3 \end{bmatrix}} - \overset{\vec{y}}{\begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_m \end{bmatrix}} \right\|^2}_{\sum_{i=1}^{m} (p(x_i) - y_i)^2}$$

To solve for p_i's, just find solutions to:

$$A^T A \vec{p} = A^T \vec{y}$$

Ex: Piazzi discovery of Ceres
dwarf planet

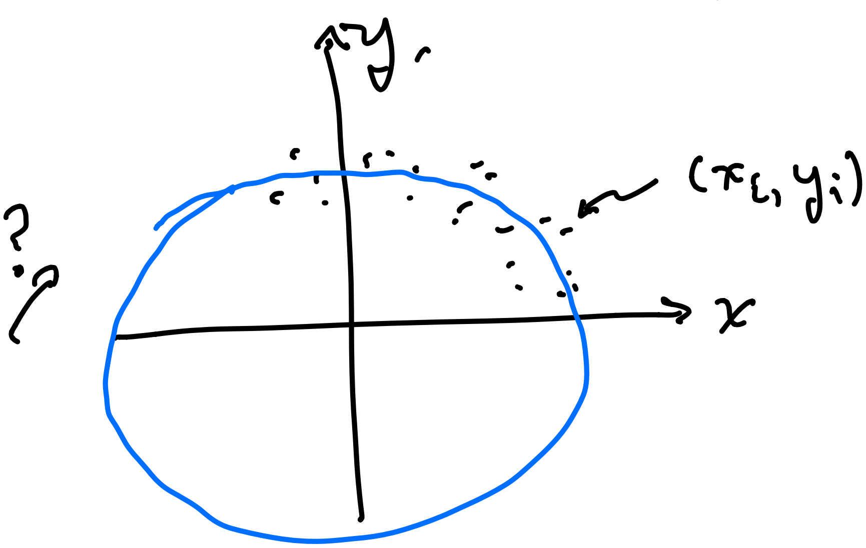
Kepler's Law of planetary motion:
planets have elliptical orbits.

So, in (x,y)-plane

$$ax^2 + by^2 + cxy + dx + ey = 1$$

Eqn for ellipse

coefficients (unknowns) determine shape of ellipse.



LS problem:

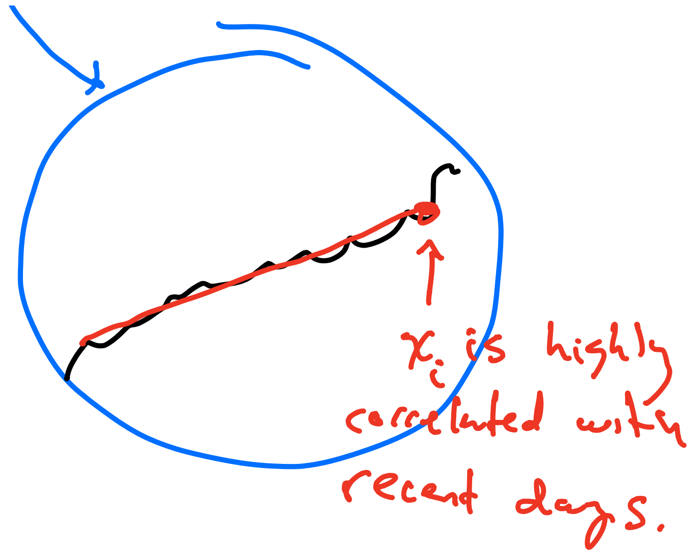
want $\quad ax_i^2 + by_i^2 + cx_iy_i + dx_i + ey_i \approx 1 \quad \forall i.$

$$\min_{a,b,c,d,e} \underbrace{\left\| \begin{bmatrix} x_1^2 & y_1^2 & x_1y_1 & x_1 & y_1 \\ \vdots & & & & \\ x_m^2 & y_m^2 & x_my_m & x_m & y_m \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \\ e \end{bmatrix} - \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right\|^2}_{\sum_{i=1}^{m} \left(ax_i^2 + by_i^2 + \cdots + ey_i - 1\right)^2}$$

<u>Ex</u>: Prediction of stock prices.

Let x_i, $i = 1 \cdots T$ be stock price on day i.

From these, we want to predict x_{T+1}.



Propose model:

$$(*) \quad a_1 x_{i+1} + a_2 x_{i+2} + \cdots + a_k x_{i+k} \approx x_{i+k+1}$$

i.e. stocks can be predicted by looking at prev k days (but there is noise!)

Goal: Learn model [i.e., a_i's] from past data.

Set up LS problem:

$$\min_{a_1 \cdots a_k} \left\| \begin{bmatrix} x_1 \cdots x_k \\ x_2 \cdots\cdots x_{k+1} \\ \\ x_{T-k} \cdots x_{T-1} \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix} - \begin{bmatrix} x_{k+1} \\ \vdots \\ \vdots \\ x_T \end{bmatrix} \right\|^2$$

$$\underbrace{\sum_{i=1}^{T-k}\left(a_1 x_i + \cdots a_k x_{i+k-1} - x_{i+k}\right)^2}$$

Bottom line: Trick is to recognize your problem can be cast as a LS problem. Then just use generic LS solution to solve.

Matching Pursuit:

Often times we want to solve a LS problem $A\vec{x} \approx \vec{b}$ in a way that gives a "sparse" solution.

(sparse = mostly zero entries).

Ex:

$$\underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ \vdots \\ x_5 \end{bmatrix}}_{\vec{x}} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

many solutions:

Examples of solutions

sparse solution

$$\vec{x} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} -1 \\ 2 \\ 2 \\ 2 \\ 2 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

More formally, want to solve "constrained" LS problem:

$$\min_{\vec{x}} \|A\vec{x} - \vec{b}\|^2 \quad \text{subject to } \|\vec{x}\|_0 \leq k.$$

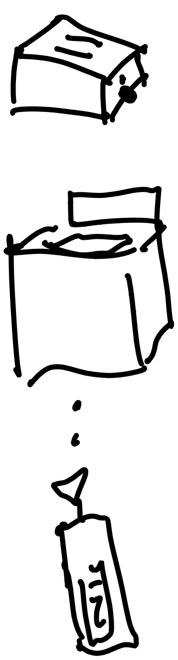
$\|\vec{x}\|_0$ = # of nonzero entries in $\vec{x}$.

This class of problems is computationally difficult to solve optimally.

So, we resort to heuristics that give good performance in practice. Matching Pursuit (next time) is one such heuristic.

Before introducing MP, let's see why sparsity is motivated in real-life.

<u>Ex</u>: Internet of Things.



$\vec{s}_1$

$\vec{s}_2$

$\vdots$

$\vec{s}_n$

signatures uniquely assigned to devices

all of same length

$$\vec{r} = \sum_{i=1}^{n} x_i \vec{s}_i$$

rcvd signal

$$x_i = \begin{cases} +1 & \text{if device } i \text{ Tx } +1 \\ -1 & \text{if device } i \text{ Tx } -1 \\ 0 & \text{if device } i \text{ not Txing.} \end{cases}$$

Assumption: a very small # of devices is Tx-ing at any given time.

Hence, to determine x_i's, can solve

$$\min_{\vec{x} \in \mathbb{R}^n} \| A \vec{x} - \vec{r} \|^2 \quad \text{s.t. } \|\vec{x}\|_0 \leq k.$$

$$A = \underbrace{\begin{bmatrix} | & & | \\ \vec{s}_1 & \cdots & \vec{s}_n \\ | & & | \end{bmatrix}}$$

in general signatures will be linearly dependent.

Ex: Classification Problem.

A proprietary pharmaceutical has chemical composition $\begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$, b_i = concentration of chem i.

There are a variety of generic compounds we can purchase, with compositions

$$\vec{a}_k = \begin{bmatrix} a_{1k} \\ \vdots \\ a_{nk} \end{bmatrix} = \text{chem composition of compound } k.$$

Goal: Solve:

$$\min \|\underbrace{A}_{\begin{bmatrix} | & & | \\ \vec{a}_1 & \cdots & \vec{a}_m \\ | & & | \end{bmatrix}}\vec{x} - \vec{b}\|^2 \quad \text{s.t. } \vec{x} \text{ sparse}$$

to find a few ingredients we can combine to make $\vec{b}$.