Lecture 7: (More) Least Squares.

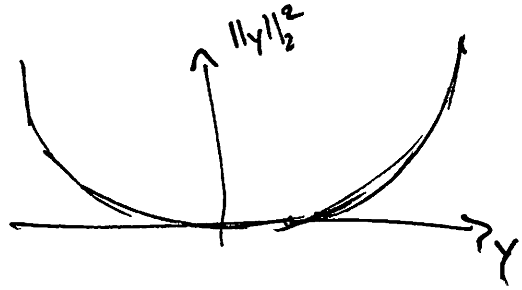
L2 norm of a vector.

$$\|y\|_2 = \sqrt{y^T y}$$

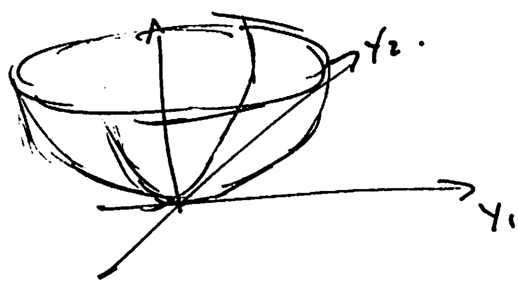
so

$$\|y\|_2^2 = y^T y = \sum_i y_i^2$$

if $\dim(y) = 1$. then y^2.

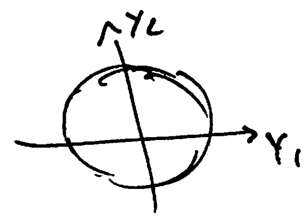


if $\dim(y) = 2$. $y_1^2 + y_2^2$



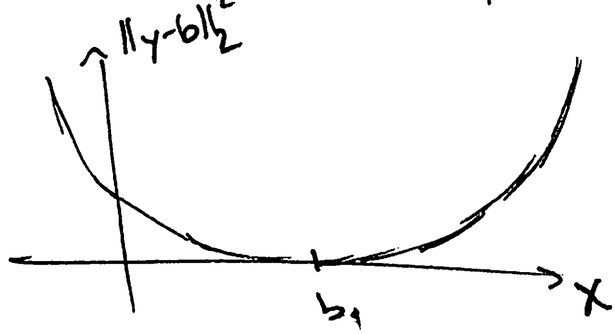
"quadratic bowl".

"unit-ball" of L2 norm was a circle.



level set of quadratic bowl.

what happens with $\|x - b\|_2^2$?

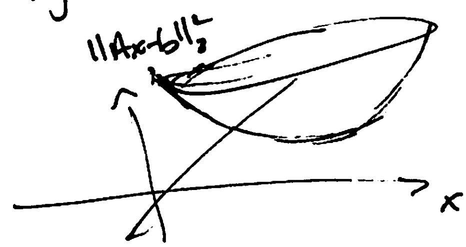


what about $\|Ax - b\|_2^2$

still quadratic in x. ~~$\alpha_1 x^2$~~

e.g. $\alpha_1 x^2 + \alpha_2 x + \alpha_3$

and always non-negative.

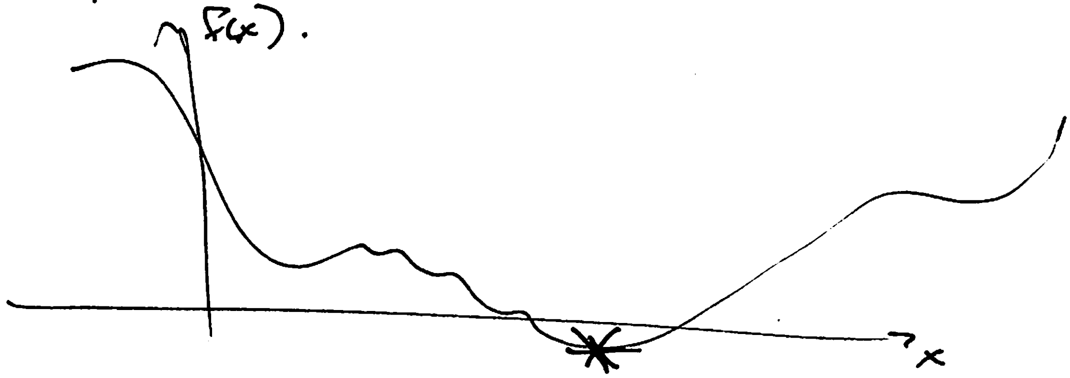


can really think of it as a "bowl".

Big idea: optimization.

Design the best input / system, etc by minimizing an "objective" function (or cost function).

$$\min_x f(x).$$



From calc, you know that if f is smooth, then $\frac{\partial f}{\partial x} = 0$ is necessary condition.

(Non-negative) Quadratic functions are special.

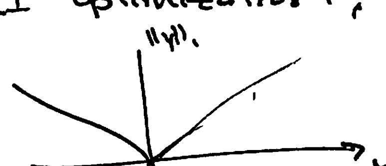
$\frac{\partial f}{\partial x} = 0$ is sufficient for optimality. (no local minima)

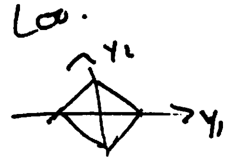
for $\min_x \|Ax-b\|_2^2$.

$\frac{\partial f}{\partial x} = 0 \Rightarrow x^* = (A^TA)^{-1}A^Tb$.

Compare this to L1 optimization, L_∞.

$\sum_i |y_i|$





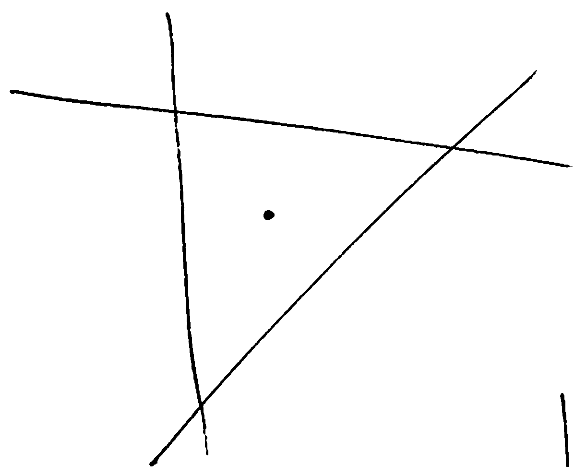
not quite as nice. but good algorithms exist.

Geometric interpretation.

$$\min_x \|Ax-b\|_2^2$$

is approximately solving. $Ax \approx b$.

minimize "error" vector. $\xrightarrow{\hspace{2em}}$ $Ax-b$.



(in matlab).

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 2 \\ 2 & -1 \end{bmatrix} \quad b = \begin{bmatrix} 1/2 \\ -1 \\ -1 \end{bmatrix}.$$

$$\left\| [1 \;\; 1]\,x - 1/2 \right\|_2^2$$

quadratic "trough".

$$\begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix} x - \begin{bmatrix} 1/2 \\ -1 \end{bmatrix}.$$

$$\| [1 \;\; 1]\,x - 1/2 \|_2^2 + \| [-1 \;\; 2]\,x + 1 \|_2^2.$$

Sum of quadratics is still a quadratic.

e.g. $(2x^2 + x + 4) + (4x^2 + 6x - 12) = 6x^2 + \ldots$

min at $\hat{A}x = \hat{b}$. (solution to 2 eqs.).

Now all three.

$$\|Ax-b\|_2^2.$$

~~Getting solution is easy. (A\b in matlab)~~.

(4) Equality constrained least-squares. (slice through quadratic is: guess what? quadratic)

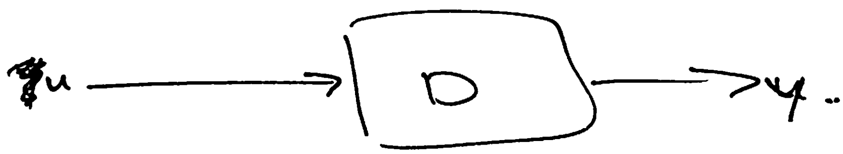
Getting solution is easy. (A\b in matlab).

Key skill: identify $\min_x \|Ax-b\|_2^2$ in your problem, and understand how to generate A + b.

Key: decision variable x enters linearly.

Least squares is everywhere.

Linear systems.



$y = Du$.

design u. to get desired y.

$$\min_u \| Du - y_d \|_2^2 .$$

or design D! from examples u, y.

Ex.: fit a line to data.

Example data. (u_i, y_i). say $\dim(u) = \dim(y) = 1$.

(2x2 in practice quiz).

want $y_i = D u_i$.

To use $Ax \approx b$, what is x? d!

what is A? u

what is b? y.

$$\underbrace{\begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_N \end{bmatrix}}_{A} d = \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ \\ y_N \end{bmatrix}}_{b}.$$

$d = A \backslash b$.

Can fit this example better w/

$$y = \cancel{d_1 u + d_2} \; d_1 u + d_2 .$$

not a linear system. but happens to still work for LS:

$$\underbrace{\begin{bmatrix} u_1 & 1 \\ \vdots & 1 \\ \vdots & 1 \\ u_m & 1 \end{bmatrix}}_{A} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \underbrace{\begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_m \end{bmatrix}}_{b} .$$

$$d = A \backslash b .$$

Core tool for machine learning / statistics.

Ex: Character recognition.

Can we use linear least-squares to build a "Two" detector?

$$28 \times 28 \text{ image} \xrightarrow{u} \boxed{\text{Two detector}} \xrightarrow{y} \begin{matrix} 1 & \text{if two} \\ 0 & \cancel{\text{if not two}}\ \text{otherwise.} \end{matrix}$$

784×1.

5000 examples (u_i, y_i) in training set.

Want $y = Du$. (what size is D?)

1×784.

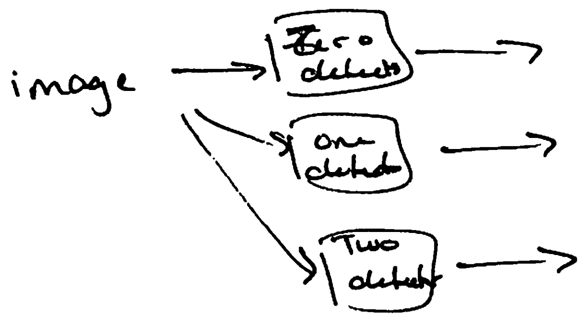
what does A look like?

$$\begin{bmatrix} - \text{image} - \\ - \text{image} - \\ - \text{image} - \\ \end{bmatrix} d = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad d = A \backslash b .$$

Turn this into a "classifier".

Build $A \backslash b$ for 0, 1, 2, … , 9.

New image.



Take the highest score.

Performance on training.

779 / 1000. in test set.

Can do a little better w/ baseline removed...

Input design.

$$u \longrightarrow \boxed{D} \longrightarrow y.$$

D is known.

find u to achieve y desired, y_d.

You've seen three examples:

$$\min_u \| Du - y_d \|_2^2 .$$ (e.g. for overconstrained)

$$\min_u \|u\|^2$$ (e.g. for underconstrained)

subject to. $y = Du$.

$y[N] = y_d$.

$$\min_u \left\| \begin{bmatrix} Du - y_d \\ u \end{bmatrix} \right\|_2^2 .$$

trade-off energy input with desired output.

Approximate inverse.

Kelly Snider.

make-up.

What is not (linear) least squares?
w/ equality constraints.

- L_1, L_∞ norms.

- $u \longrightarrow \boxed{f(u)} \longrightarrow y$ nonlinear systems.

- $\min_x \|Ax-b\|_2^2$ s.t. $Cx \le d$.

inequality constraints.
(e.g. for resonance ~~from~~ problem from homework).

⋮