# Lecture 12

## Least Squares

## and Regularization

- OFFICE HOURS FRIDAY 11AM ! (P.)

- PSET 2

  Due Mon Oct 5

- Midterm : Wed Oct 7

  (no lecture that day)

  (material up to Lecture 11)

# THE INVERSE: SQUARE $\boxed{A \in \mathbb{R}^{n \times n}}$

A has a left inverse $\Leftrightarrow$ A nonsingular (invertible) $\Leftrightarrow$ A has a right inverse

$\Updownarrow$

rows linearly independent

$\Updownarrow$

columns linearly independent

$\Updownarrow$

$$\boxed{\det A \neq 0}$$

$\Updownarrow$

$\text{rank}(A) = n$ (Full rank)

$\Updownarrow$

$N(A) = \{0\}$

$\Updownarrow$

$C(A) = \mathbb{R}^n$

$\Updownarrow$

$Ax = b$ always has a unique sol. $(A^{-1}b)$

## Computation

Gaussian elimination!

Apply to

$$\begin{bmatrix} A & I \end{bmatrix} \xrightarrow{\text{G.E.}} \overbrace{\begin{bmatrix} I & B \end{bmatrix}}^{\text{RREF}}$$

Then, $\boxed{B = A^{-1}}$

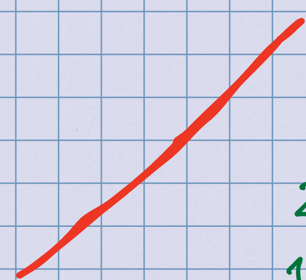
# $Ax = b$

**Q:** What happens when the inverse does not exist, but we still need a solution?

E.g.:

- Undetermined

  $[\ A\ ]$ "wide"



2 VARS
1 EQ

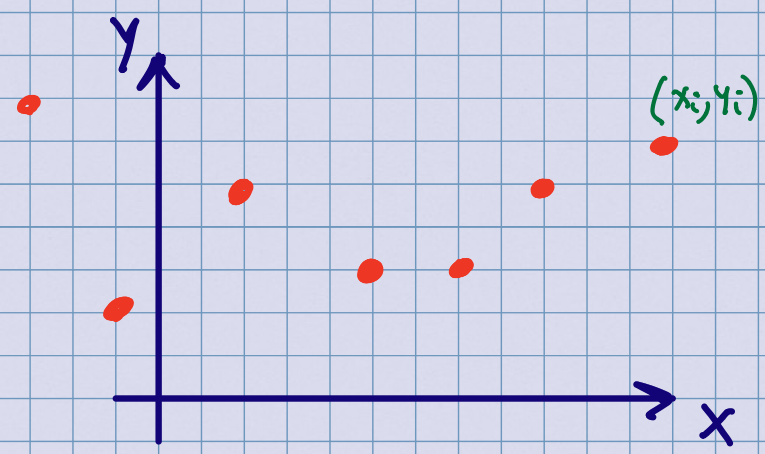
- Overdetermined

  $\begin{bmatrix} \\ A \\ \\ \end{bmatrix}$ "tall"

2 VARS
3 EQS

## Ex 1: Data Fitting

"regression"




- Fit given data $(x_i, y_i)$ with a linear combination of given functions $f_i(x)$

$$c_1 f_1(x) + \cdots + c_k f_k(x)$$

(e.g., polynomials)

- Yields the system

$$\begin{bmatrix} f_1(x_1) & f_2(x_1) \\ f_1(x_2) & f_2(x_2) \\ f_1(x_3) & f_2(x_3) \\ f_1(x_4) & f_2(x_4) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}$$

# parameters

← each row is a data point

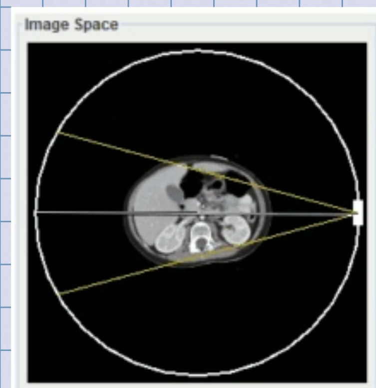
- Typically, $\boxed{\#\text{ data points} > \#\text{ parameters}}$

$\Rightarrow$ OVERDETERMINED SYSTEM !

# Ex 2: Tomography

CT SCAN / RECONSTRUCTION

"computed tomography"



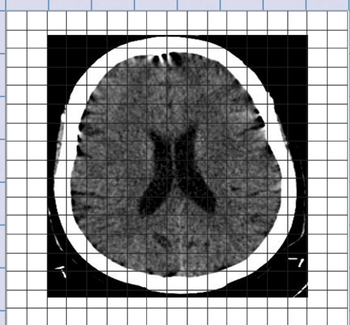

AGGREGATE MULTIPLE VIEWS

(E.G., X-RAY) IN A SINGLE IMAGE/MODEL

(VERY) SIMPLIFIED MODEL

pixels (2D)

voxels (3D)



density.

| | | | | |
|---|---|---|---|---|
| 0.0 | 0.1 | | | |
| 0.1 | 0.3 | 0.5 | 0.4 | |
| 0.2 | 0.4 | 0.6 | 0.5 | |
| 0.1 | 0 | | | |

- Each snapshot provides multiple equations.
  - Many more equations than variables.

    $\Rightarrow$ OVERDETERMINED

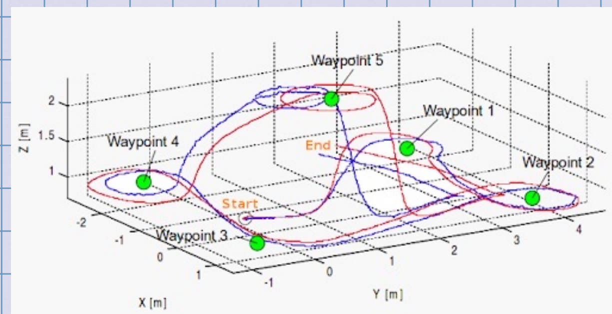
- Very large scale (3D, high resolution)

  $(2000)^3 = \underbrace{8\,000\,000\,000}_{\text{8 billion variables}}$

- Need good algorithms!

# Ex3: Control / Robotics.

## Trajectory generation/planning




Linearized dynamical model

$$x_{k+1} = A x_k + B u_k \qquad k=0,\dots,N$$

$x$: state
$u$: input

$x_0$ given

**Goal:** Design "low cost" input sequence $\{u_1, \dots, u_N\}$ to reach desired final state $x_N$ (given)

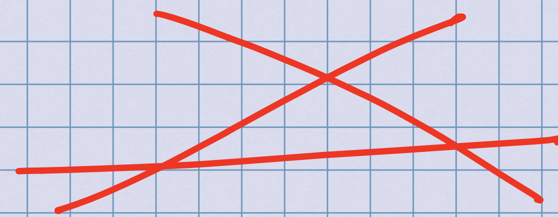
$$\left[\begin{array}{cccc|cccc} -I & 0 & \cdots & 0 & B & & & 0 \\ A & -I & & 0 & & B & & \\ & A & -I & & & & \ddots & \\ 0 & & \ddots & A & & 0 & & B \end{array}\right] \left[\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_{N-1} \\ \hline u_0 \\ u_1 \\ \vdots \\ u_{N-1} \end{array}\right] = \begin{bmatrix} -Ax_0 \\ 0 \\ 0 \\ \vdots \\ x_N \end{bmatrix}$$

e.g. minimize fuel, or energy, ...

Model is linear, infinitely many solutions

$\Longrightarrow$ UNDERDETERMINED

# $Ax = b$



## $\boxed{\text{OVERDETERMINED CASE}}$

$$\left[\, A \,\right] x = b$$

$A \in \mathbb{R}^{m \times n}$

- "tall matrix"
- "more equations than unknowns"

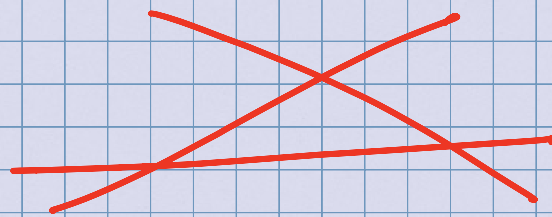
typically, no $x$ satisfies <u>all</u> equations

<u>E.g.</u> What is the "solution" of the linear system

$$\begin{cases} x = 1 \\ x = 2 \\ x = 5 \\ x = 4 \end{cases} \quad ?$$

Think, e.g., of different measurements of the same object.

# $Ax = b$



## OVERDETERMINED CASE

$$\left[\, A \,\right] x = b \qquad A \in \mathbb{R}^{m \times n}$$

- "tall matrix"
- "more equations than unknowns"

typically, no $x$ satisfies <u>all</u> equations

---

**Q:** What is a "good" approximate solution?

$\rightarrow$ Minimize "size" of residual (equation error)

$$\boxed{r = Ax - b}$$

We will measure "size" as

$$\|r\|^2 = r^T r = \sum r_i^2$$

(Euclidean norm, 2-norm)

<u>Why?</u> E.g., interpret $r$ as <u>error</u> or "<u>noise</u>"

$$Ax = b + r$$

"noise" $r_i$ added to each equation

Find $x$ that minimizes the total amount of noise / error. $\sum r_i^2$

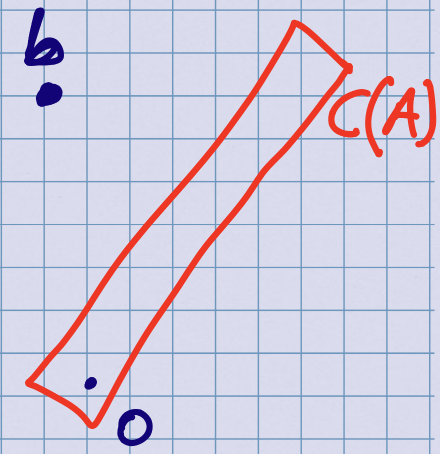
<u>Assumption</u>: $Ax = b$ (overdetermined) $A \in \mathbb{R}^{m \times n}$

- <u>Columns</u> of $A$ are <u>linearly indep.</u>
- Equivalently, $\underline{\operatorname{rank}(A) = n}$

## APPROACH I : GEOMETRY / ORTHOGONALITY.

We want to solve $A\mathbf{x} = \mathbf{b}$,

but b is not in column space of A




Q: How to change b (as little as possible) so that it is in $C(A)$?

$\Rightarrow$ replace b with its orthogonal projection onto $C(A)$

Hey, we've done this before!

(last lecture).

$$\boxed{P = A(A^TA)^{-1}A^T}$$ ← orthogonal projection onto $C(A)$

$$\Rightarrow \quad \tilde{b} = A(A^TA)^{-1}A^Tb$$

so now want to solve $Ax = \tilde{b}$, i.e.,

$$Ax = \overbrace{A(A^TA)^{-1}A^Tb}^{\tilde{b}}$$

Easy...

$$\Rightarrow \quad \boxed{x = (A^TA)^{-1}A^Tb}$$

this is the <u>least-squares approximate solution</u> of $Ax = b$.

## APPROACH II: LEFT INVERSE

(or, plug in and verify.)

$$Ax = b$$

$A \in \mathbb{R}^{m \times n}$

If columns of A are LI ("full column rank")
then it has a left inverse

In particular,

$$\boxed{\underbrace{(A^T A)}_{\text{rank } n,\ \text{invertible}}{}^{-1} A^T}$$

is a left inverse.

$\Rightarrow$ gives a candidate

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

What makes this point "special"?

"completion of squares"

$Ax = b$

Take any solution $x$, and write

$$x = x^* + d$$

Then,

$$\|Ax-b\|^2 = \|Ax^* + Ad - b\|^2$$
$$= \|Ax^*-b\|^2 + \|Ad\|^2 + \underbrace{2d^T A^T (Ax^*-b)}_{0}$$

why?

$$\Rightarrow \|Ax-b\|^2 = \|Ax^*-b\|^2 + \|Ad\|^2$$

$$\|Ax-b\|^2 \geq \underbrace{\|Ax^*-b\|^2}$$

$x^*$ gives "solution" with smallest residual

# APPROACH III: OPTIMIZATION (CALCULUS)

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

$$R = \|Ax-b\|^2 = (Ax-b)^T(Ax-b)$$
$$= x^TA^TAx - 2x^TA^Tb + b^Tb$$

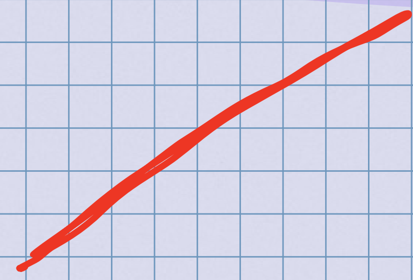
- Quadratic function of $x$ "strictly convex"
- Take gradient (derivative), and set to 0.

$$\frac{\partial R}{\partial x} = 2A^TAx - 2A^Tb = 0$$

$$\Rightarrow \quad A^TAx = A^Tb$$ "normal equations"

$$\Rightarrow \quad \boxed{x = (A^TA)^{-1}A^Tb}$$

# $Ax = b$



## UNDERDETERMINED CASE

"wide matrix"

"fewer equations than unknowns"!

$[\quad A \quad ]\, x = b$

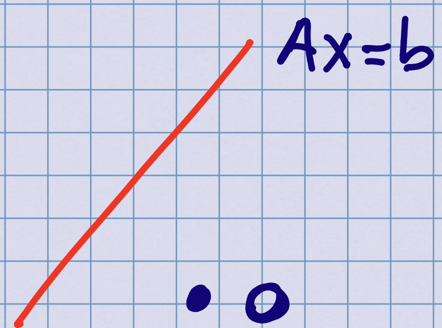
$A \in \mathbb{R}^{m \times n}$

typically, infinitely many $x$ satisfy equations

**Q:** What is a "good" point $x$?

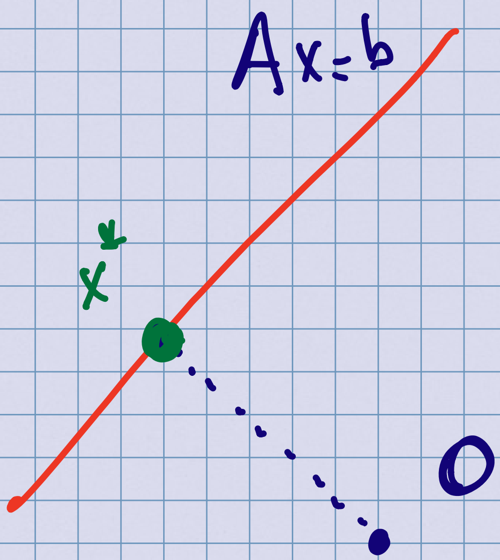
A solution $x$ of minimal "size",

e.g. Euclidean norm.



Assumption: Rows are LI (i.e., rank(A) = m)

Claim: (verify!)




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

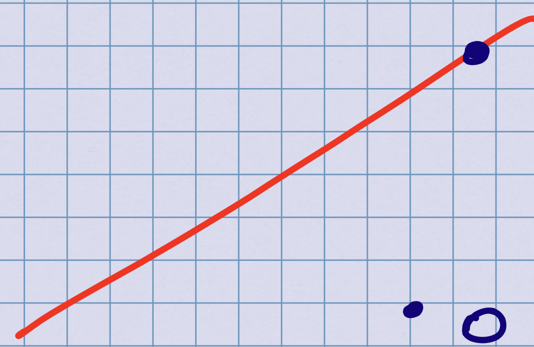
is the minimal norm (least squares) solution of the underdetermined system.

# REGULARIZATION

A slightly different way of handling $\left\{ \begin{matrix} \text{overdet} \\ \text{underdet} \end{matrix} \right\}$ linear systems.

Just like before, want to solve

$$Ax = b$$




Now, I'll choose a solution that minimizes

$$T = \underbrace{\|Ax - b\|^2}_{\text{equation error}} + \lambda \|x\|^2$$

$T$ (for Tikhonov)

$\lambda$: parameter ($\lambda > 0$)

$\|x\|$: norm of solution.

Typically, choose $\lambda$ based on prior information.

Can show that

(more about this later in the course)

$$\frac{\partial J}{\partial x} = 2A^T(Ax-b) + 2\lambda x = 0$$

$$\Rightarrow (A^TA + \lambda I)x = A^Tb$$

$$\Rightarrow \boxed{x^* = (A^TA + \lambda I)^{-1} A^T b}$$

This is the <u>$\lambda$-regularized solution.</u>

- Solution <u>depends</u> on $\lambda$
- $\lambda$ gives <u>tradeoff</u> between error and size of solution
- As $\lambda \to 0$, we recover the previous cases.