7.1.1 Consistency in a System of Equations

- $Ax = b$

least squares : find the best x that approximates B

note : $A \in R^{M \times N}$

M : row / no of eq^n

N : col / no of variables

$x \in R^N$

$b = R^M$

- consistent : $Ax = b$ ← has soln "agrees"

inconsistent : no soln

- M & N

$M > N$	$M = N$	$M < N$
tall, thin	square	under-determined

↑

more eq^n than unknowns

overdetermined system → often inconsistent

$Ax = b$

consistent

- b is in col space of A
  $\Rightarrow$ b is formed by linear combination of A's columns

no of independent cols or rows

aug matrix

- $\text{Rank}(A) = \text{Rank}(A|b)$
  rank of A is the same as that of the aug. matrix

unique solution

$\text{rank}(A) = N = \text{rank}(A|b)$

infinite solution

$\text{rank}(A|b) = \text{rank}(A) < N$

inconsistent

not

- occurs when $M > N$ ie. over-determined matrix
- rows of A are dependent but not consistent w/ b values
- $\text{rank}(A) < \text{Rank}(A|b)$

least square

- when system is inconsistent, we try to find x $(\hat{x})$:

Ax is as close to b as possible

$$\|b - A\bar{x}\| \leq \|b - Ax\|$$

approximation error

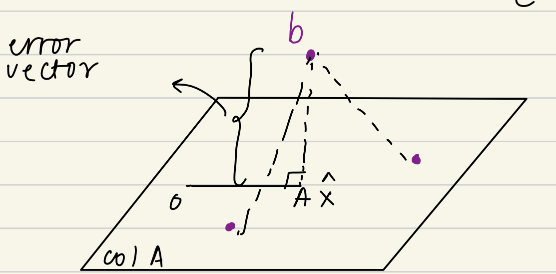
best smallest error

compare to error

distant b/w $A\hat{x}$ and b

least squares comes from the fact that $\|b - Ax\|$ is the square root of a sum of squares

$$e = b - A\hat{x} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix}$$



- least squares soln $\hat{x}$ minimises $f(x) = \|Ax - b\|^2$

norm of error squared

1. partial derivative

$e = Ax - b$

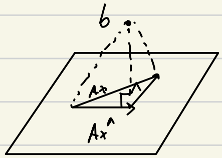
$\|Ax - b\|^2 = \|e\|^2 = e \cdot e = \sum e_i^2$

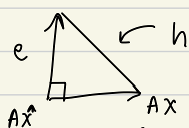
then n (no of unknown) partial derivatives

② . $(b - \text{proj } b \text{ on } A) = e$

$A\hat{x} = \text{proj}_{col(A)} b$

is orthogonal to cols of A
$\therefore$ any vector in col A is orthogonal to e





hyp > e always.

③ . Normal Eqⁿ : $A^TAx = A^Tb$

system of eqⁿ and least sq soln of $Ax=b$ satisfies it.

$e = (b - A\hat{x}) \perp$ to each col of A $\therefore$ $a_j \cdot (b - A\hat{x}) = 0$

and $a_j^T (b - A\hat{x})$ (goto pg7)

$\because a_j^T$ is a row of A^T :

$A^T(b - A\hat{x}) = 0 \quad A^Tb - A^TA\hat{x} = 0 \quad , \underline{\underline{A^TA}}\, x = A^Tb$

square

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

$x = A^{-1}A^{T^{-1}}A^Tb$

if A^TA is not inv. psuedo inverse is used.

let A : $m \times n$ matrix :

a. equation $Ax = b$ has a unique least squares soln for each b in R^m

b. columns of A are linearly independent

c. matrix A^TA is invertible

d. least sq. soln $\hat{x} = (A^TA)^{-1}A^Tb$

v. imp concept : pg 9 of orthogonality notes

$$\boxed{\text{proj}_v u = v \left(\frac{\overbrace{u^T v}^{u \cdot v}}{\|v\|^2} \right)}$$

$$\text{proj}_v u = \frac{v \cdot u}{v \cdot v} v$$

$$= \frac{v (v^T u)}{v^T v}$$

$$= \frac{(v v^T) u}{v^T v}$$

$$= \frac{v v^T}{v^T v} u \rightarrow R^{N \times 1}$$

outer product

$R^{N \times 1} \times R^{1 \times N} = N \times N$ matrix

Rank = 1

1×1 matrix = scalar.

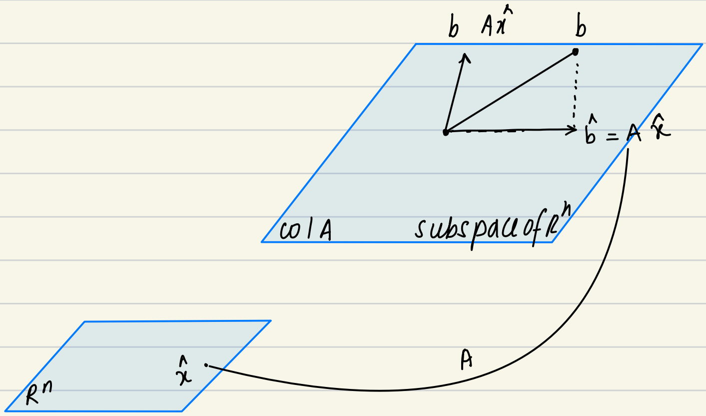
$$\frac{N \times N \quad N \times 1}{\text{scalar}} = N \times 1$$

$$\text{proj}_v u = \overbrace{\frac{v v^T}{v^T v}}^{N \times N} \overbrace{u}^{N \times 1}$$

projection matrix

$$\text{proj}_v y = P y \qquad P = \frac{v v^T}{v^T v}$$

projection matrix and least square solutions



$\hat{x} = (A^T A)^{-1} A^T b$ (if $A^T A$ is inv.)

$A\hat{x} = \hat{b}$

$\hat{b} = \text{proj}_{\text{col}A}\, b = A\hat{x}$

$\Updownarrow$

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

$\therefore \; A(A^T A)^{-1} A^T \leftarrow$ projection matrix

$\hat{b} = Pb$ (pg 26 of orthogonality)

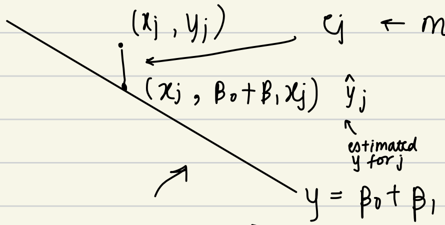
$\downarrow$

QQ^T

- P is a square matrix. $P = P^T$
- $P^N = P \leftarrow$ idempotent property

applications of least square to linear model

fitting a line



e_j ← minimise

$\sum_{j=1}^{n} e_j^2$ = sum of residual error square must be minimised

least squares line
line of regression of y on x

β_0, β_1 ← regression coeff.

$\varepsilon = y - X\beta$ ⟶ X is info matrix or design ; β is parameters

↑ residual vector

↖ actual y

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

$$X\beta = y$$
$$X^T X \beta = X^T y$$
$$(X^T X)^{-1} X^T X \beta = (X^T X)^{-1} X^T y$$
$$\beta = (X^T X)^{-1} X^T y$$

Fitting a curve

$$y = \beta_0 f_0(x) + \beta_1 f_1(x) + \dots + \beta_k f_k(x)$$

linear model as it is linear in unknown parameters

f_i are known parameters
β_0 must be determined

design matrix : $\begin{bmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_n(x_1) \\ f_0(x_2) & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \end{bmatrix}$