Lecture 17 3.21.2019

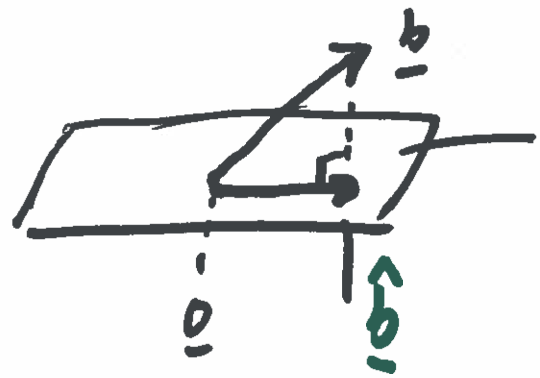
Recap:

Given $A \in \mathbb{R}^{m \times n}$ $(m > n)$

$\underline{b} \in \mathbb{R}^m$.

Consider $A\underline{x} = \underline{b}$

- $\underline{x} \in \mathbb{R}^n$ is not always solvable



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

when $\underline{b} \notin$ range(A),

we don't have solvability

Least squares solution

- project $\underline{b}$ to range(A), we get $\hat{\underline{b}} \in$ range(A)
- Solve $A\hat{\underline{x}} = \hat{\underline{b}}$

$\hat{\underline{x}}$: least squares solution of $A\underline{x} = \underline{b}$

If exists, $\hat{\underline{x}}$ will satisfy

the normal equation

$$\underbrace{A^T A}_{n\times n \text{ matrix}} \hat{\underline{x}} = A^T \underline{b}$$

The existence and uniqueness of $\hat{\underline{x}}$

$\Leftrightarrow$ $A^T A$ is invertible

$\Leftrightarrow$ columns of A are linearly independent.

How to understand $\hat{\underline{x}}$ is the best? in what sense?

Lemma: $\hat{\underline{b}} \in \text{range}(A)$, satisfying

$\Leftrightarrow$ $\underline{b} - \hat{\underline{b}} \perp \text{range}(A)$

$\Rightarrow$ $\|\underline{b} - \underbrace{\hat{\underline{b}}}_{\downarrow\, A\hat{\underline{x}}}\|_2 = \min_{\underline{y} \in \text{range}(A)} \|\underline{b} - \underline{y}\|_2$.

$\Leftrightarrow$ $\|\underline{b} - A\hat{\underline{x}}\|_2 = \min_{\underline{x} \in \mathbb{R}^n} \|\underbrace{\underline{b} - A\underline{x}}_{\text{residual}}\|_2$

'next best' : the residual

$$\underline{r} = \underline{b} - A\underline{x}$$ is minimized

in $\|\cdot\|_2$ sense

<u>Proof of Lemma.</u> :

let $S = \text{range}(A)$

let $\underline{\hat{b}} \in S$, and it satisfies

$$\underline{b} - \underline{\hat{b}} \perp S$$

we want to show

$$\|\underline{b} - \underline{\hat{b}}\|_2 = \min_{\forall \underline{y} \in S} \|\underline{b} - \underline{y}\|_2$$

Consider any $\underline{y} \in S$

$$\|\underline{b} - \underline{y}\|_2^2 = \|\underline{b} - \underline{\hat{b}} + \underline{\hat{b}} - \underline{y}\|_2^2$$

$$= (\underline{b} - \underline{\hat{b}} + \underline{\hat{b}} - \underline{y})^T(\underline{b} - \underline{\hat{b}} + \underline{\hat{b}} - \underline{y})$$

$$= (\underline{b} - \underline{\hat{b}})^T(\underline{b} - \underline{\hat{b}}) + (\underline{b} - \underline{\hat{b}})^T(\underline{\hat{b}} - \underline{y}) + (\underline{\hat{b}} - \underline{y})^T(\underline{b} - \underline{\hat{b}}) + (\underline{\hat{b}} - \underline{y})^T(\underline{\hat{b}} - \underline{y})$$

$$= \|\underline{b} - \underline{\hat{b}}\|_2^2 + 2(\underline{b} - \underline{\hat{b}})^T(\underline{\hat{b}} - \underline{y}) + \|\underline{\hat{b}} - \underline{y}\|_2^2$$

note that $\hat{\underline{b}} - \underline{y} \in S$ ... due to "S is a linear space"

hence $(\underline{b} - \hat{\underline{b}})^T (\hat{\underline{b}} - \underline{y}) = 0$ $\forall \underline{y} \in S$

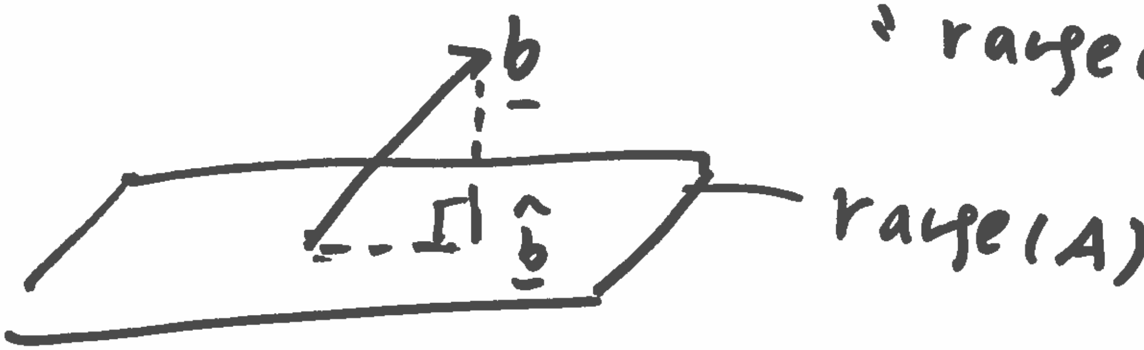
$$\Rightarrow \|\underline{b} - \underline{y}\|_2^2 = \|\underline{b} - \hat{\underline{b}}\|_2^2 + \|\hat{\underline{b}} - \underline{y}\|_2^2 \geq \|\underline{b} - \hat{\underline{b}}\|_2^2$$

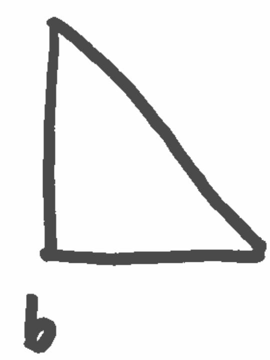
range (A)

$$\Rightarrow \boxed{\|\underline{b} - \underline{y}\|_2 \geq \|\underline{b} - \hat{\underline{b}}\|_2 \quad \forall \underline{y} \in S}$$

$$\Rightarrow \|\underline{b} - \hat{\underline{b}}\|_2 = \min_{\forall \underline{y} \in S} \|\underline{b} - \underline{y}\|_2$$

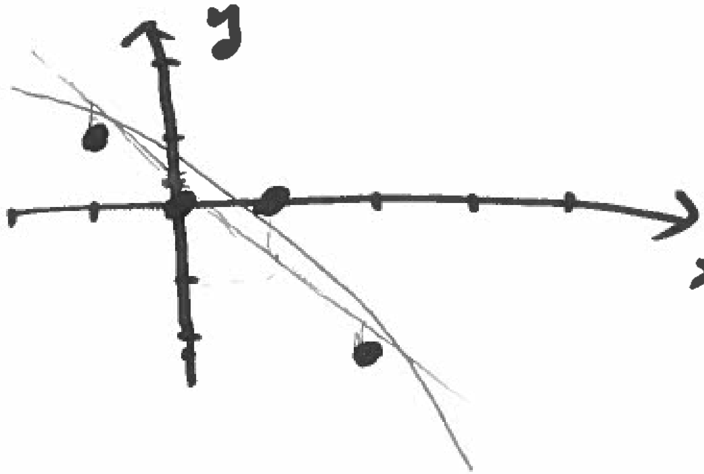
= range(A).





Application: data fitting

Example given 4 data point,



j	x_j	y_j
1	-1	1
2	0	0
3	1	0
4	2	-2

Find the best linear and best parabola to fit the data.

The best is in the 2-norm sense.

hence the least square sense.

Sol: 1) linear, find $f(x) = a_1 + a_2 x$

and "hopefully" $f(x_j) = a_1 + a_2 x_j = y_j \quad j = 1, \cdots 4$

In matrix vector form

$$A \underline{a} = \underline{b}. \qquad A \in \mathbb{R}^{4 \times 2}$$

$$A = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad \underline{b} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -2 \end{pmatrix}$$

$$\underline{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

This is over-determined, we want to find the best one in 2-norm

that is

$$\| \underbrace{\underline{b} - A\hat{\underline{a}}}_{\text{residual}} \|_2 = \min_{\forall \underline{a} \in \mathbb{R}^2} \| \underline{b} - A\underline{a} \|_2$$

and the residual $\underline{r} = \underline{b} - A\underline{a}$ is minimized in 2-norm.

This corresponds to the least squares solution.

The Solution $\hat{\underline{a}}$ satisfies the normal equation

$$A^T A \hat{\underline{a}} = A^T \underline{b}.$$

where $A^T A = \begin{pmatrix} 4 & 2 \\ 2 & 6 \end{pmatrix}$

$$A^T \underline{b} = \begin{pmatrix} -1 \\ -5 \end{pmatrix}$$

$$\Rightarrow \quad \hat{\underline{a}} = \begin{pmatrix} 0.2 \\ -0.9 \end{pmatrix}$$

$\Rightarrow$ The linear least squares fitting $f(x) = 0.2 - 0.9x$

2) quadratic: find $g(x) = a_1 + a_2 x + a_3 x^2$

such that $g(x_j) = a_1 + a_2 x_j + a_3 x_j^2 = y_j$

over-determined $\quad j = 1, 2, 3, 4$

we look for LS solution to minimize the residual

$$\underline{r} = \begin{bmatrix} y_1 - g(x_1) \\ y_2 - g(x_2) \\ y_3 - g(x_3) \\ y_4 - g(x_4) \end{bmatrix}$$

in 2-norm.

This corresponds to the least LS solution of $A\underline{a} = \underline{b}$.

where
$$A = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \\ 1 & x_4 & x_4^2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}$$

$$\underline{b} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ -2 \end{bmatrix} \qquad \underline{a} = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

LS solution $\hat{a}$ will satisfy the normal equation

$$A^T A \hat{\underline{a}} = A^T \underline{b}$$

$$\vdots$$

$$\hat{\underline{a}} = \begin{bmatrix} 0.45 \\ -0.65 \\ -0.25 \end{bmatrix}$$ hence

the quadratic least squares fit is

$$g(x) = 0.45 - 0.65x - 0.25x^2$$

Discussion. $A \in \mathbb{R}^{m \times n}$ $\underline{b} \in \mathbb{R}^m$.

Consider $A\underline{x} = \underline{b}$

1) Least squares solution $(m > n)$

$$\|\underbrace{\underline{b} - A\hat{\underline{x}}}_{\underline{r}:\text{residual}}\|_2 = \min_{\underline{x} \in \mathbb{R}^n} \|\underline{b} - A\underline{x}\|_2$$

$\Updownarrow$

the normal equation $A^T A \hat{\underline{x}} = A^T \underline{b}$

other norm can be used $\|\cdot\|_*$ instead of 2-norm.

2) when $m > n >> 1$

computationally solving the normal equation ?, not the robust way to find LS solution.

more robust algorithms are available: QR factorization, SVD decomposition.

3). what about

$$A\underline{x} = \underline{b} \quad A \in \mathbb{R}^{m \times n}$$

$m < n$.

additional constraints are needed

- $\|\underline{x}\|_*$: minimized
- fewest nonzero entries "sparsity"

which is better? 'linear or quadratic'

- intuitively: quadratic is no worse than linear.
- to measure

$$\underline{r} = \underline{b} - A\,\hat{\underline{a}} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{pmatrix}$$

squared error (SE): $= \|\underline{r}\|_2^2$

$$SE = \underbrace{r_1^2 + r_2^2 + r_3^2 + r_4^2}_{m}$$

root mean squared error (RMSE)

$$RMSE = \sqrt{\frac{SE}{4}} = \frac{\|\underline{r}\|_2}{\sqrt{m}}$$

$4 \rightarrow m.$

$$SE = \begin{cases} 0.7 & (\text{linear}) \\ 0.45 & (\text{quadratic}) \end{cases}$$

(for the data fitting example)

§4 Numerical Differentiation and integration.

Given $f(x)$,

$\begin{cases} \text{to approximate} & f'(x) \\ \text{or to} \quad : & \int_a^b f(x)\,dx \end{cases}$

One idea:

$$f(x) \sim p(x) \quad \text{(approximate)}$$

$$\begin{array}{ccc} f(x) & & \\ \downarrow & & \\ f'(x) & \overset{?}{\longleftarrow} & p'(x) \end{array}$$

$$\int_a^b f(x)\,dx \overset{?}{\longleftarrow} \int_a^b p(x)\,dx$$

§4.1 Numerical differentiation.

Given a function $f(x)$, we sample $\{(x_j, y_j)\}_{j=1}^{n+1}$, $y_j = f(x_j)$

we assume $x_{j+1} - x_j = h =$ Constant
(not essential).

$n=1$ (two points)

$P_1(x)$ is the linear interpolation

$$P_1(x) = y_1 + \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

$$\Rightarrow P_1'(x) = \frac{y_2 - y_1}{x_2 - x_1} = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

At x_1

$$f'(x_1) \approx P_1'(x_1) = \frac{f(x_1 + h) - f(x_1)}{h}$$

forward difference

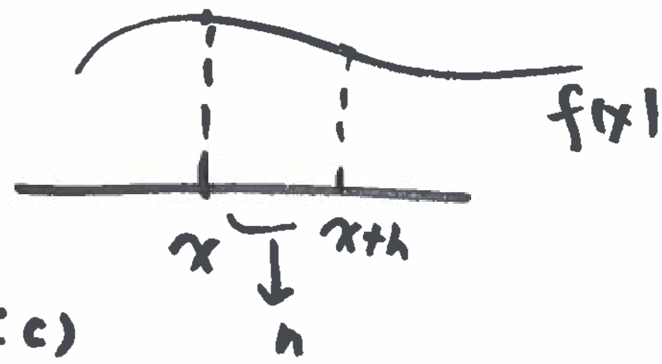
At x_2

$$f'(x_2) \approx P_1'(x_2) = \frac{f(x_2) - f(x_2 - h)}{h}$$

backward difference

$$f'(x) \approx \begin{cases} \dfrac{f(x+h) - f(x)}{h} & \text{forward difference} \\[2ex] \dfrac{f(x) - f(x-h)}{h} & \text{backward difference} \end{cases}$$

These approximations can also be derived based on Taylor's series expansion.



$f(x+h) - f(x)$

$= h f'(x) + \frac{h^2}{2} f''(c)$

c is some number between x and $x+h$.

$$\frac{f(x+h) - f(x)}{h} = f'(x) + \frac{h}{2} f''(c)$$

$\Rightarrow$ forward difference approximation

$$f'(x) \sim \frac{f(x+h) - f(x)}{h}$$

the error $\frac{h}{2} f''(c)$ is first order in h. or written as $O(h)$

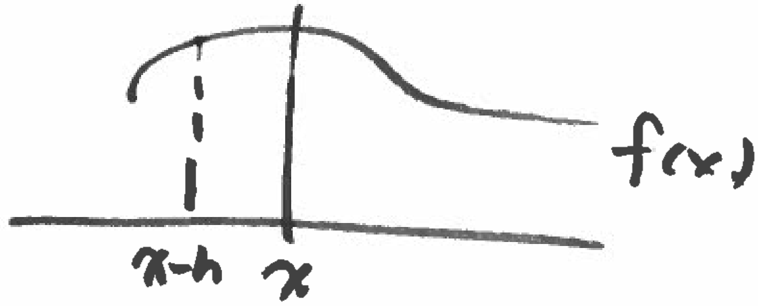
notation:

$O(h^n)$: for a quantity $Q(h)$, if $\exists M > 0$. s.t.

$$|Q(h)| \leq M h^n \quad \forall h > 0$$

then $Q(h) = O(h^n)$, it is said to be nth order of h.

Similarly



$$f(x-h) - f(x) = -h f'(x) + \frac{h^2}{2} f''(\hat{c})$$

$$\tilde{c} \in [x-h, x]$$

$$\Rightarrow \quad \frac{f(x-h) - f(x)}{-h} = \frac{f(x) - f(x-h)}{h}$$

$$= f'(x) + \underbrace{\frac{h}{2} f''(\hat{c})}_{O(h)}$$

$$\Rightarrow f'(x) \approx \frac{f(x) - f(x-h)}{h}$$

backward difference approximation

with a first order in h error

$n = 2 \qquad (x_j, y_j) \quad j = 1, 2, 3$

$P_2(x)$ is quadratic interpolation.

$x_{j+1} - x_j = h$

$$\begin{array}{llll}
x_1 & y_1 & & \\
 & & \dfrac{y_2 - y_1}{\underbrace{x_2 - x_1}_{= h}} & \\
x_2 & y_2 & & \dfrac{y_3 - 2y_2 + y_1}{2h^2} \\
 & & \dfrac{y_3 - y_2}{\underbrace{x_3 - x_2}_{= h}} & \\
x_3 & y_3 & &
\end{array}$$

$$P_2(x) = y_1 + \frac{y_2 - y_1}{h}(x - x_1) + \frac{y_3 - 2y_2 + y_1}{2h^2}(x - x_1)(x - x_2)$$

Take derivative

$$P_2'(x) = \frac{y_2 - y_1}{h} + \frac{y_3 - 2y_2 + y_1}{2h^2} \cdot (2x - x_1 - x_2)$$

$$P_2''(x) = \frac{y_3 - 2y_2 + y_1}{h^2}$$

at $x = x_1$

$$P_2'(x)\Big|_{x=x_1} = \frac{y_2-y_1}{h} + \frac{y_3-2y_2+y_1}{2h^2} \cdot (\underbrace{2x_1 - x_1}_{-h} - x_2)$$

$$= \frac{-3y_1 + 4y_2 - y_3}{2h}$$

$$= \frac{-3f(x_1) + 4f(x_1+h) - f(x_1+2h)}{2h}$$

at $x = x_2$

$$P_2'(x)\Big|_{x=x_2} = \frac{y_2-y_1}{h} + \frac{y_3-2y_2+y_1}{2h^2} \cdot h$$

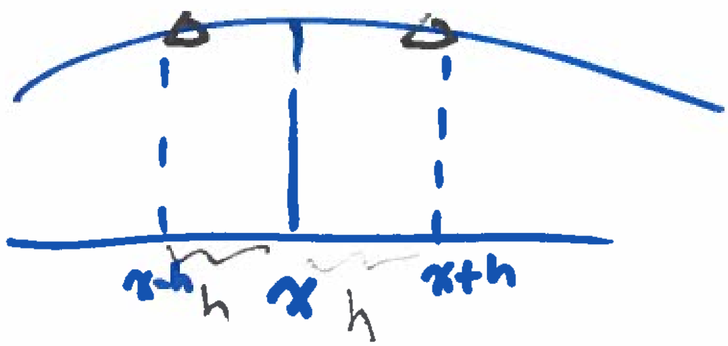
$$= \frac{y_3-y_1}{2h}$$

$$= \frac{f(x_2+h) - f(x_2-h)}{2h}$$

at $x = x_3$

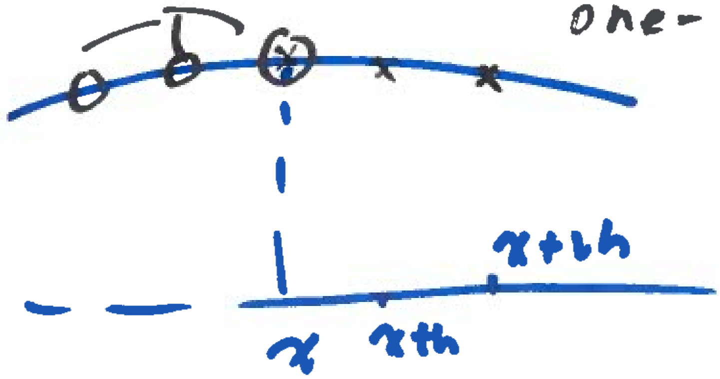
$$P_2'(x_3) = \frac{f(x_3-2h) - 4f(x_3-h) + 3f(x_3)}{2h}$$

This gives us two one-sided approximations and one central approximation



central

one-sidel



Similarly

$$f''(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}$$

Central approximation.