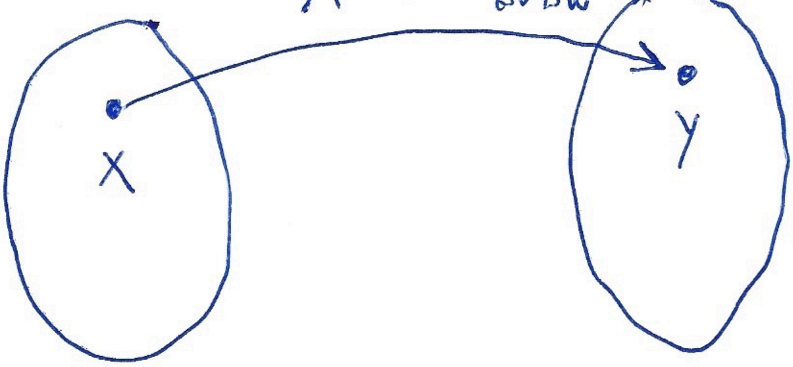
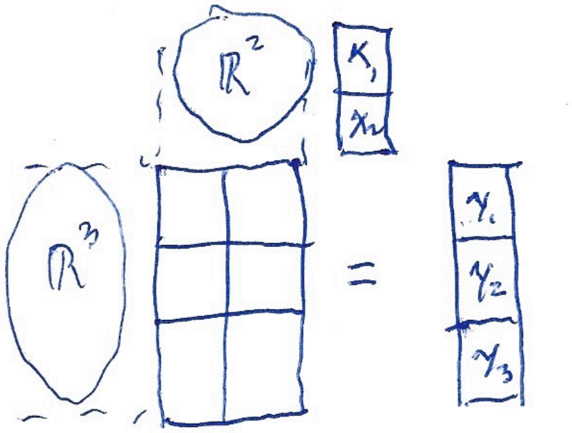
5. LEAST – SQUARES

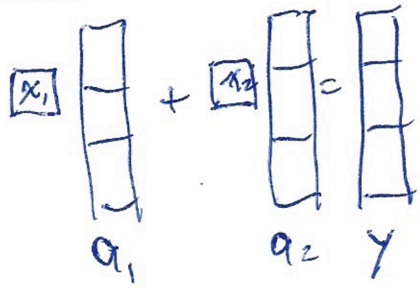
$A_{m \times n} \, x_{n \times 1} = b_{m \times 1}$

Interpreting $Ax = b$

$A = [\mathcal{L}]_{B_V B_W}$







$(V, \mathbb{F})$

B_V

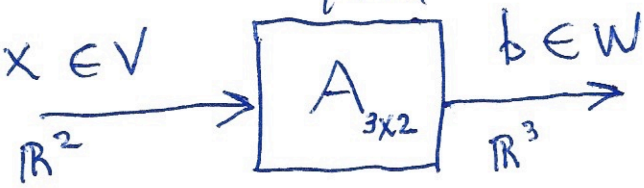
$[x]_{B_V} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

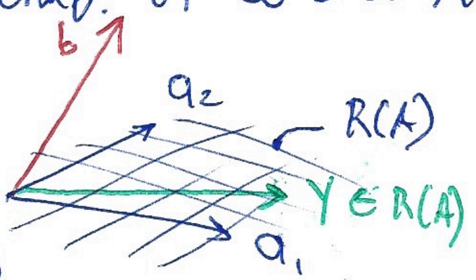
$(W, \mathbb{F})$

B_W

$[y]_{B_W} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$

$R(A) = \{ y \mid Ax = y \}$ = space of reachable vecs in W via Lin Comb. of cols of A





$\exists \, x \xrightarrow{A} y \in R(A) \subset \mathbb{R}^3$

$\nexists \, x \xrightarrow{A} b \in W = \mathbb{R}^3$

$\dim R(A) = 2 < \dim W = 3$

thus $Ax \neq b$

or

$Ax \cong b$

5.1. the LEAST Squares problem

When $b \notin R(A)$, the lin sys is inconsistent

$$\nexists x \mid Ax = b, \text{ denoted } Ax \cong b$$

means there may be no exact solutions

the LS formulation seeks an approximate solution $\hat{x} \mid A\hat{x} \approx b$. this can be written as

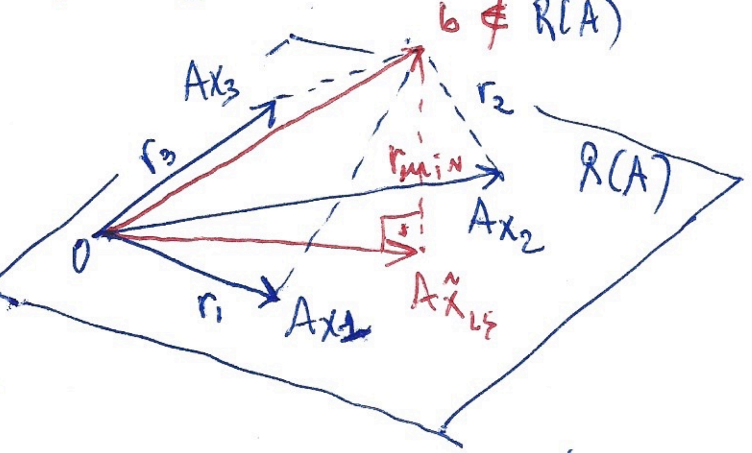
$$\|b - A\hat{x}\|^2 \leq \|b - Ax\|^2 \quad \forall\, x \in V$$

which, in optimization form

$$\boxed{\hat{x}_{LS} = \underset{x}{\arg\min} \|b - Ax\|^2}$$

5.2. Orthogonality principle & NORMAL EQUATIONS

Define the residue $r \triangleq b - Ax$, then $\min_x \|r\|^2$



$\forall\, p \in V \quad Ap \perp r_{min}$

$(Ap)^* r_{min} = 0 \qquad \langle u, v\rangle = u^* v$

$p^* \underbrace{A^*(b - A\hat{x}_{LS})}_{q} = 0$

$p^* q = 0 \;\; \forall p \text{ iff } q = 0$

r is an error vector

Orthogonality cond: $\boxed{A^* \underbrace{(b - A\hat{x}_{LS})}_{r_{min}} = 0}$ $\quad r_{min} \perp$ Cols of A

or $A^*(b - A\hat{x}_{LS}) = 0$

$$\boxed{A^*A\hat{x}_{LS} = A^*b}$$

Normal Equations
- Always consistent
$R(A^*A) = R(A^*)$, $A^*b \in R(A^*)$

- unique solution if A is full col rank, which means $\exists (A^*A)^{-1}$

If A is full col rank

$$\boxed{\hat{X}_{LS} = (A^*A)^{-1}A^*b}$$

the least-squares solution for $Ax \cong b$.

If A is not full col rank, or it is a fat matrix, then $\nexists (A^*A)^{-1}$. How to proceed? Form an auxiliary system and find the solution set, for example, via triangularization

$$A^*A\bar{X}_{LS} = A^*b \quad \Rightarrow \quad \boxed{C\bar{X}_{LS} = d}$$

Auxiliary Lin Sys

Applying GE method on it, we arrive at

$$\bar{X}_{LS} = \underbrace{X^0_{LS}}_{\substack{\text{particular} \\ \text{solution} \\ \text{(minimum norm solution)}}} + \underbrace{z}_{\substack{\text{Homogeneous} \\ \text{solution}}}$$

where $X^0_{LS} \in R(A^*)$
$z \in N(A)$

recall decomp. thm
$\bar{X}_{LS} \in \mathbb{F}^n = R(A^*) \oplus N(A)$

Let's check: $\bar{x}_{LS} = x^0_{LS} + z$

$$A^*A\bar{x}_{LS} = A^*b$$

$$A^*A(x^0_{LS} + z) = A^*b$$

$$A^*Ax^0_{LS} + A^*Az = A^*b$$

$$A^*Ax^0_{LS} + A^*\underbrace{(Az)}_{0} = A^*b \quad \therefore \quad \boxed{A^*Ax^0_{LS} = A^*b}$$

minimum norm solution

Recall $R(A^*) \perp N(A)$, or $x^{0*}_{LS}z = 0$

then

$$\|\bar{x}_{LS}\|^2 = \|x^0_{LS} + z\| = (x^0_{LS} + z)^*(x^0_{LS} + z)$$

$$= x^{0*}_{LS}x^0_{LS} + \underbrace{z^*x^0_{LS}} + \underbrace{x^{0*}_{LS}z}_{0} + z^*z$$

$$\boxed{\|\bar{x}_{LS}\|^2 = \|x^0_{LS}\|^2 + \|z\|^2}$$

therefore, the norm of $\bar{x}_{LS}$ is minimum when $z = 0$, and the minimum norm solution is

$$\bar{x}_{LS} = x^0_{LS}.$$

Another way to see this: assume X_1, X_2 are two different solutions

$$A^*AX_1 = A^*b \quad \text{and} \quad A^*AX_2 = A^*b$$

subtracting both equations yields

$$A^*AX_1 - A^*AX_2 = 0$$

$$A^*A\underbrace{(X_1 - X_2)}_{z} = 0 \quad \text{or} \quad A^*Az = 0$$

$\therefore$ $z \in \mathcal{N}(A^*A)$. But $z^*A^*Az = z^*0$

$$z^*A^*Az = 0 \iff (Az)^*(Az) = 0$$

that is, $\|Az\|^2 = 0 \iff Az = 0$.

thus, since $A \neq 0$ and $z \neq 0$, then

$\boxed{z \in \mathcal{N}(A)}$. since $X_1 \in R(A^*)$ and $z \in \mathcal{N}(A)$, $z \perp X_1$

therefore $z = X_1 - X_2$ or $X_2 = X_1 + z$, so that $\|X_2\|^2 = \|X_1\|^2 + \|z\|^2$ $\therefore$ Minimum norm sol for $z = 0$

5.3. REGULARIZED LEAST SQUARES

In cases where A is not full col rank, or it is ill-conditioned (Finite precision issues), regularization yields a unique regularized LS solution

$$\boxed{(\Pi + A^*A)\,\hat{x}_\Pi = A^*b} \qquad \Pi > 0$$

(positive definite)

Typically $\Pi = \varepsilon I$, $0 < \varepsilon \ll 1$.

We solve a slightly perturbed lin sys with good numerical properties, not far from the original system.

<u>Remark</u>: In cases A is not full rank, $\nexists\,(A^*A)^{-1}$, Π guarantees invertibility.

A positive definite matrix is always nonsing.

$$p^*(\varepsilon I + A^*A)p = \varepsilon p^*p + p^*A^*Ap$$

$$= \underbrace{\varepsilon \|p\|^2}_{>0} + \underbrace{\|Ap\|^2}_{\geq 0} > 0$$

≥ 0 (if $p \in \mathcal{N}(A)$) $Ap = 0$

$\therefore\ \Pi + A^*A > 0$ always invertible

5.4. RECURSIVE LEAST SQUARES

Very useful for real time operation and sequential data. Let A_i and b_i be

$$A_i = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{i-1} \\ a_i \end{bmatrix} \qquad b_i = \begin{bmatrix} b(0) \\ b(1) \\ \vdots \\ b(i-1) \\ b(i) \end{bmatrix}$$

$a_i = i^{th}$ row $\qquad$ $b(i) = i^{th}$ row

We may build the normal eqs and solve, at every iteration i:

$$\boxed{(\Pi + A_i^* A_i)\hat{X}_i = A_i^* b_i} \qquad (I)$$

Problem: A_i and b_i grow in size as $i \to \infty$ and solving (I) is progressively more computationally costly.

Let us pose (I) as follows

$$\boxed{\Phi_i \hat{X}_i = S_i} \,, \qquad \Phi_i \triangleq A_i^* A_i + \Pi$$
$$S_i \triangleq A_i^* b_i$$

And partition row wise A_i and b_i

$$A_i = \left[\begin{array}{c} A_{i-1} \\ \hline a_i \end{array}\right] \,, \qquad b_i = \left[\begin{array}{c} b_{i-1} \\ \hline b(i) \end{array}\right] \,,$$

now rewrite Φ_i and s_i via partitioning

$$\Phi_i = A_i^* A_i + \Pi = [A_{i-1}^* \mid a_i^*] \left[\frac{A_{i-1}}{a_i}\right] + \Pi$$

$$= A_{i-1}^* A_{i-1} + a_i^* a_i + \Pi$$

$$= \Phi_{i-1} + a_i^* a_i + \Pi \quad \text{or}$$

$$\boxed{\Phi_i = \Phi_{i-1} + a_i^* a_i} \qquad \boxed{\Phi_{-1} = \Pi}$$

$i \geq 0$

$$s_i = [A_{i-1}^* \mid a_i^*] \left[\frac{b_{i-1}}{b(i)}\right] = A_{i-1}^* b_{i-1} + a_i^* b(i)$$

$$\boxed{s_i = s_{i-1} + a_i^* b(i)} \qquad \boxed{s_{-1} = 0} \quad i \geq 0$$

Grouping:

$$\boxed{\begin{aligned} \Phi_i &= \Phi_{i-1} + a_i^* a_i && (1) \\ s_i &= s_{i-1} + a_i^* b(i) && (2) \\ \Phi_i \hat{x}_i &= s_i && (3) \end{aligned}}$$

Algorithm (1) – (3) has good numerical properties. it has two steps

a) update (1) and (2) as a_i and $b(i)$ are available

b) Solve the consistent lin sys (3)

b1) via factorizations, say QR, or LU

b2) Directly $\hat{x}_i = \Phi_i^{-1} s_i$ (numerically worse than b1)

Algorithm (1) – (3) can be further simplified: Can we relate $\hat{x}_i$ to the previous solution $\hat{x}_{i-1}$? Yes! Start with the direct solution

$$\hat{x}_i = \Phi_i^{-1} s_i,$$

then define

$$P_i \triangleq \Phi_i^{-1} = (\Phi_{i-1} + a_i^* a_i)^{-1}$$

$$P_i = (\underbrace{\Phi_{i-1}}_{A} + \underbrace{a_i^*}_{B} \underbrace{\wedge}_{C=I} \underbrace{a_i}_{D})^{-1} \quad \text{(use Matrix inversion Lemma)}$$

$$\boxed{P_i = P_{i-1} - \frac{P_{i-1} a_i^* a_i P_{i-1}}{1 + a_i P_{i-1} a_i^*}}, \quad P_{-1} = \Pi^{-1}.$$

$$\text{Now, } \hat{x}_i = P_i s_i = \left(P_{i-1} - \frac{P_{i-1} a_i^* a_i P_{i-1}}{1 + a_i P_{i-1} a_i^*}\right)\left(A_{i-1}^* b_{i-1} + a_i^* b(i)\right)$$

$$\hat{x}_i = \underbrace{\boxed{P_{i-1} A_{i-1}^* b_{i-1}}}_{\hat{x}_{i-1}} + P_{i-1} a_i^* b(i) - \frac{P_{i-1} a_i^* a_i \overbrace{\boxed{P_{i-1} A_{i-1}^* b_{i-1}}}^{\hat{x}_{i-1}}}{1 + a_i P_{i-1} a_i^*} - \frac{P_{i-1} a_i^* a_i P_{i-1} a_i^* b(i)}{1 + a_i P_{i-1} a_i^*}$$

$$\hat{x}_i = \hat{x}_{i-1} + P_{i-1} a_i^* b(i) - P_{i-1} a_i^* a_i \hat{x}_{i-1} + P_{i-1} a_i^* b(i) a_i P_{i-1} a_i^* - \frac{P_{i-1} a_i^* a_i P_{i-1} a_i^* b(i)}{1 + a_i P_{i-1} a_i^*}$$

$$\hat{x}_i = \hat{x}_{i-1} + \frac{P_{i-1} a_i^* b(i) - P_{i-1} a_i^* a_i \hat{x}_{i-1}}{1 + a_i P_{i-1} a_i^*}$$

$$\hat{x}_i = \hat{x}_{i-1} + \frac{P_{i-1} a_i^* (b(i) - a_i \hat{x}_{i-1})}{1 + a_i P_{i-1} a_i^*}$$

$$\boxed{\begin{aligned} e(i) &= b(i) - a_i \hat{x}_{i-1} \\ \hat{x}_i &= \hat{x}_{i-1} + \frac{P_{i-1} a_i^* e(i)}{1 + a_i P_{i-1} a_i^*} \end{aligned}}$$

RLS Algorithm $\quad P_{-1} = \Pi^{-1}$.

PSI5794 - Matrix Analysis - 2021

Homework 7 - Least Squares

1. A certain process $f\colon \mathbb{R} \to \mathbb{R}$ takes the form

$$f(x) = ax^2 + bx + c,$$

for some $a = 0.1$, $b = 1.0$ and $c = 1.5$. We take noisy measurements of this process and construct the following table:

x	0.1	0.5	1.0	2.0	2.5	3.0
$f(x)$	1.6912	1.9562	2.7460	3.9765	4.4972	5.3141

We want to use a mean-square framework to model this process.

(a) We know that the constant a is small, therefore we can approximate this s an affine function. Formulate the mean-square problem of finding the best polinomial $g(x) = px + q$ that approximate the process and find its solution. Compute the error.

(b) Now we want to model the system as a full quadratic function. Again, formulate the mean-square problem of finding the best polynomial $g(x) = rx^2 + sx + t$ that approximates the process and find its solution. Compute the error.

(c) Plot the graphs of the two solutions in a single figure. Also, plot a scatter graph of the set of measured points. Compare the results. (You may use software such as MATLAB or Octave to do this.)

(d) We make another mesurement and get $f(3.5) = 6.2250$. Find a way to compute the new solution from the previous one. Do so for both the affine case and the quadratic case. (Hint: the deterministic RLS algorithm.)

(e) Find the best (in a least-squares sense) degree 5 polynomial that approximates $f(x)$ using only the points in the table. Compute the error and plot the graph of the solution. Is it a good idea to use this solution to model $f(x)$?