$y = Ax \longrightarrow x = A \backslash y$ ← introduced by MATLAB

m divide ..

Least-Squares

Gabriel Hugh Elkaim

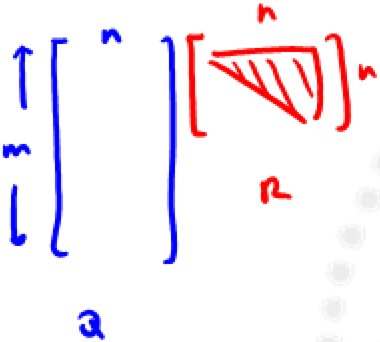
Winter 2016

Least Squares via QR Factorization

$A \in \mathbb{R}^{m \times n}$ skinny, full rank

$A = QR \qquad Q^TQ = I_n \qquad R \in \mathbb{R}^{n \times n}$

upper triangular



$$A^\dagger = (A^TA)^{-1}A^T = (R^T \underbrace{Q^TQ}_{} R)^{-1} R^T Q^T$$

$$= (R^TR)^{-1} R^T Q^T = R^{-1} \underbrace{R^{-T} R^T}_{I} Q^T$$

$$\boxed{A^\dagger = R^{-1}Q^T}$$

$$x_{ls} = R^{-1}Q^Ty$$

Not I

$$P_r(A) = A(A^TA)^{-1}A^T = AR^{-1}Q^T = QRR^{-1}Q^T = \boxed{QQ^T = P_r(A)}$$

$$P_r(A) \cdot y = \sum_{i=1}^{n} (q_i^Ty)\, q_i$$

Least-Squares via "Full" QR Factorization (1.3)

$$A = [Q_1 \; Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} \qquad [Q_1 \vdots Q_2] \in \mathbb{R}^{m \times m} \text{ orthogonal}$$

$R_1 \in \mathbb{R}^{n \times n}$ $\triangledown$, invertible

$$\|Ax - y\|_2^2 = \left\| [Q_1 \; Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} x - y \right\|^2 \qquad [Q_1 \; Q_2]^T$$

$$\left\| [Q_1 \; Q_2]^T [Q_1 \; Q_2] \begin{bmatrix} R_1 \\ 0 \end{bmatrix} x - [Q_1 \; Q_2]^T y \right\|^2$$

$$\left\| \begin{matrix} R_1 x - Q_1^T y \\ \hline 0 - Q_2^T y \end{matrix} \right\|^2 = \|R_1 x - Q_1^T y\|^2 + \|Q_2^T y\|^2$$

Least-Squares via "Full" QR Factorization (2.3)

$$\|Ax - y\|^2 = \|R_1 x - Q_1^T y\|^2 + \|Q_2^T y\|^2$$

$$x = x_{ls} = R_1^{-1} Q_1^T y \rightarrow \underbrace{\|\underbrace{R_1 R_1^{-1}}_{I} Q_1^T y - Q_1^T y\|^2}_{\phi} + \underline{\|Q_2^T y\|^2}$$

$$A x_{ls} - y = -Q_2 Q_2^T y$$

$Q_1 Q_1^T$ gives the projection onto $\mathcal{R}(A)$

$Q_2 Q_2^T$ gives the projection onto $\mathcal{R}(A)^\perp$

$\mathcal{N}(A^T)$

Least-Squares Estimation (1.2)

$y = Ax + v$ ← noise

↑ forward model

x – is what we want to reconstruct or estimate ($\hat{x}$)

y – is our sensor measurements

v – is an unknown measurement noise or sensor error

assume that v is small.

i^{th} row of A characterizes the i^{th} sensor

$\hat{x} \simeq x \qquad \hat{y} = A\hat{x} \rightarrow \boxed{\hat{v} = y - A\hat{x}}$

Least-Squares Estimation (2.2)

$\min_{\hat{x}} \|A\hat{x} - y\|$ minimize the deviation between

- what we observe (y)
- what we would observe if $x = \hat{x}$ and there was no noise.

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

BEST LINEAR Estimator

BLUE Property (1.2)

unbiased

$y = Ax + v \quad A \in \mathbb{R}^{m \times n}$ skinny, full rank

linear estimator: $\hat{x} = By \qquad \hat{x} = B(Ax + v) = BAx + Bv$

B must be a left inverse of A.

UNBIASED $\longrightarrow$ $\hat{x} = x$ if $v = 0$. $\longrightarrow$ $BA = I$.

any left inverse of A.

$$x - \hat{x} = x - B(Ax + v) = \cancel{x} - \underbrace{B\cancel{A}}_{I}\,x - Bv$$

$$= -Bv$$

would like B to be the smallest left inverse of A.

BLUE Property (2.2)

$A^+ = (A^TA)^{-1}A^T$ is <u>the</u> smallest left inverse of A.

for any $BA = I$ $\quad \sum_{i,j} B_{ij}^2 \geq \sum_{i,j} A^{+\,2}_{ij}$ ← Frobenius norm

$A^+A = I$

$(A^+ + \alpha F)A = I$

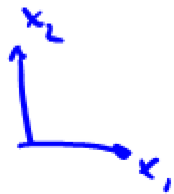
$\underline{\alpha FA = 0}$.

F_{ij} — should be <u>small</u>.

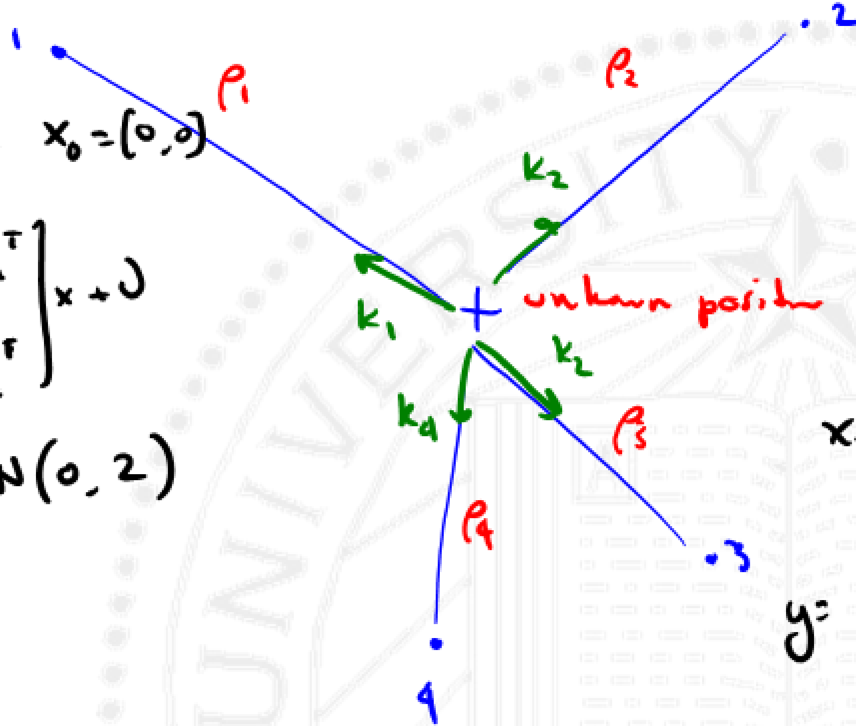
$F \in N(A^+)$

Rows of F are orthogonal to the columns of A (2nd half of Fund Thm)

Example: Navigation (ρ-ρ)



Initially $x_0 = (0,0)$



$$y = -\begin{bmatrix} k_1^T \\ \vdots \\ k_4^T \end{bmatrix} x + v$$

$$v \sim N(0, 2)$$

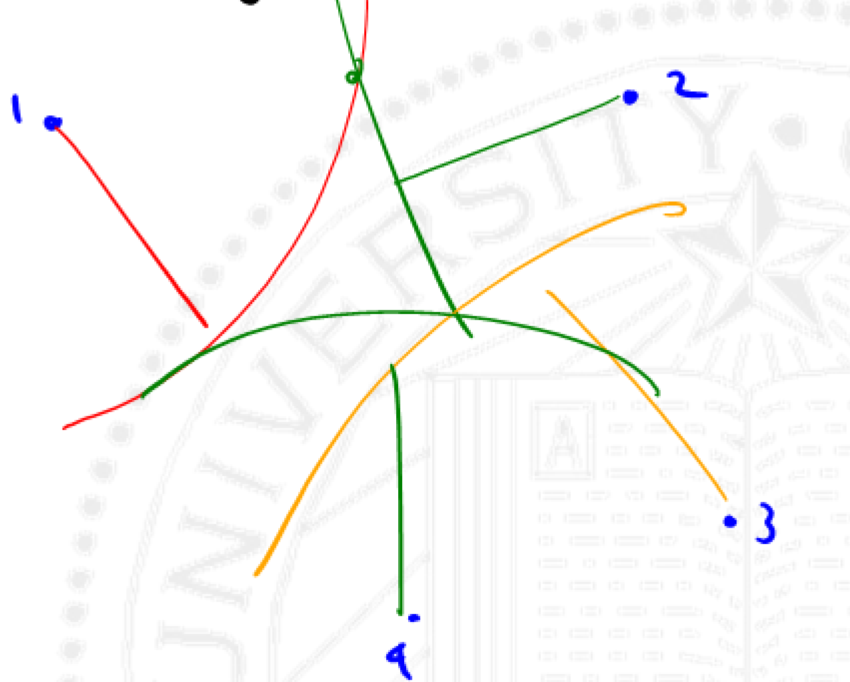
$$\boxed{x \in \mathbb{R}^2}$$

$$x_{true} = \begin{bmatrix} 5.59 \\ 10.55 \end{bmatrix}$$

$$y = \begin{bmatrix} -11.95 \\ -2.89 \\ -9.51 \\ 2.51 \end{bmatrix}$$

Example: Navigation (ρ-ρ)

given $y \rightarrow$ estimate x



R.A.I.M.

Just Enough Measurements

$$\hat{x} = By = \begin{bmatrix} 0 & -1 & \boxed{0} & \boxed{0} \\ -1.12 & 0.5 & \boxed{0} & \boxed{0} \end{bmatrix} y$$

$$\hat{x} = \begin{bmatrix} 2.89 \\ 4.9 \end{bmatrix} \qquad \|r\| = \underline{3.07}$$

Best case if Beacon 3 & 4 are known to be bad.

Least-Squares Method

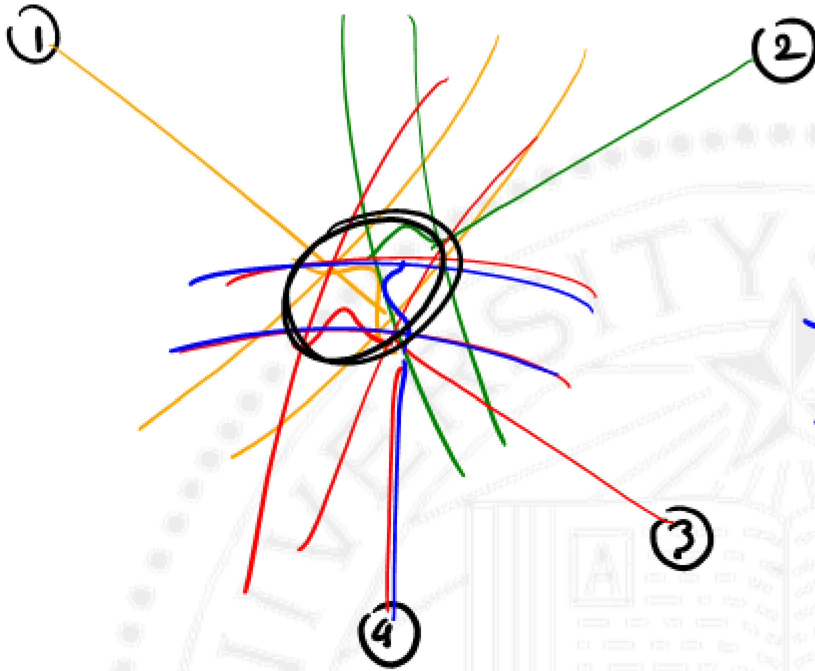
$$\hat{x} = (A^TA)^{-1}A^Ty = A^{\dagger}y = \underbrace{\begin{bmatrix} -.23 & -.48 & 0.04 & 0.94 \\ -.47 & -0.02 & -.51 & -.48 \end{bmatrix}}_{\text{BLENDING MATRIX}}$$

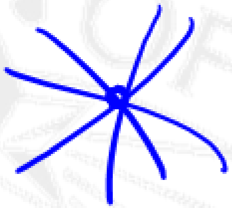
$$\hat{x} = \begin{bmatrix} 4.95 \\ 10.26 \end{bmatrix} \longrightarrow \|r\|_2 = 0.72$$

"SENSOR FUSION"

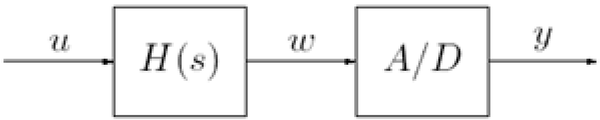
B and $A^{\dagger}$ are BOTH left inverses of A.

Larger entries of B_{ij} lead to larger estimation errors.


1
2
3
4



Example from Overview: Comm



u piecewise constant, period of 1 unit $0 \le t \le 10$.

$$w(t) = \int_0^t h(t-\tau)\, u(\tau)\, d\tau$$

sample y @ 10x $\quad \tilde{y}_i = w(0.1\, i) \quad i = 1..100$

3 bit A/D $\quad Q(a) = \frac{1}{4}\left(\text{round}(4a + \tfrac{1}{2}) - \tfrac{1}{2}\right)$

given $y \in \mathbb{R}^{100\times 1} \longrightarrow$ estimate $\hat{x} \in \mathbb{R}^{10\times 1}$

Comm. Example (1.3)

$y = Ax + v$

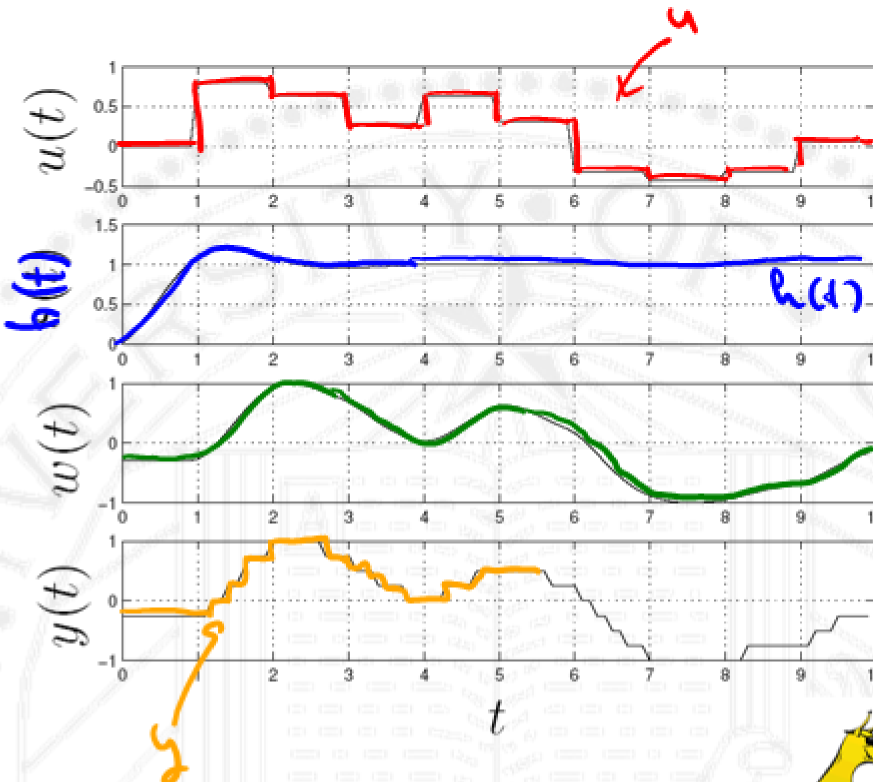
$A \in \mathbb{R}^{100 \times 10}$

$A_{ij} = \int_{j-1}^{j} h(0.1i - \tau)\, d\tau$

$v \in \mathbb{R}^{100}$ $\quad v_i = a(g_i)$

$|v_i| \le 0.125$.

$x_{ls} = (A^T A)^{-1} A^T y$



$x_{ls} = A^{+} y$

$\frac{\|x - x_{ls}\|}{\sqrt{10}} = 0.02$

↓

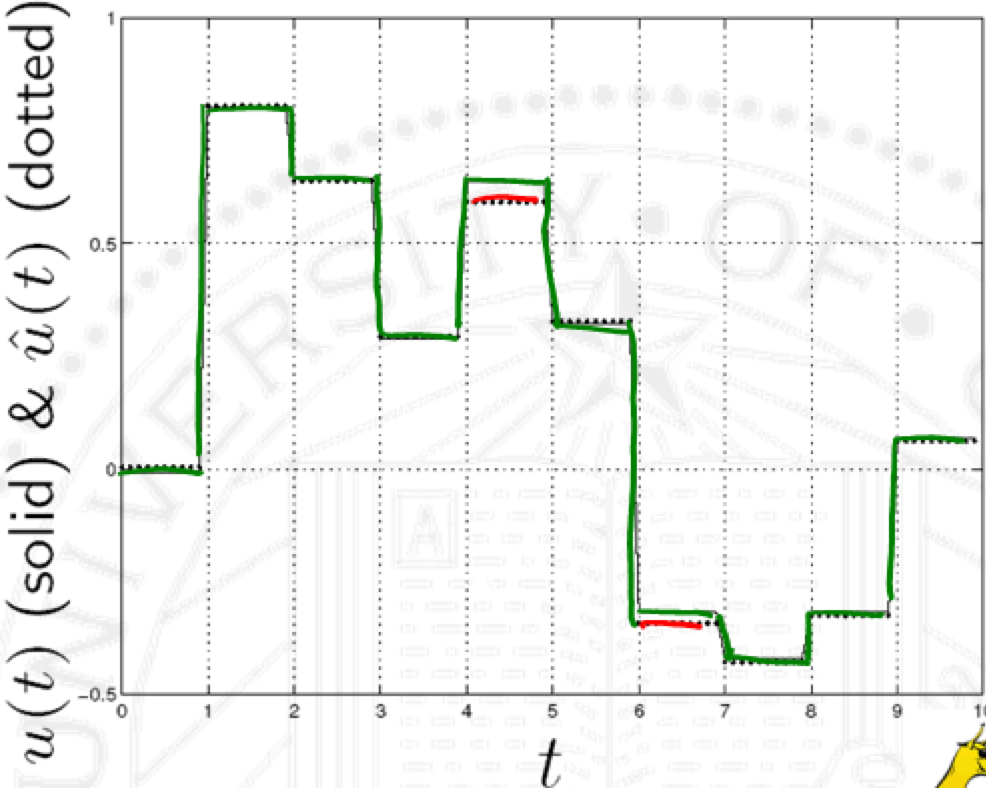
Better than if

$h(1) = 1$

↓

rms ~ 0.07

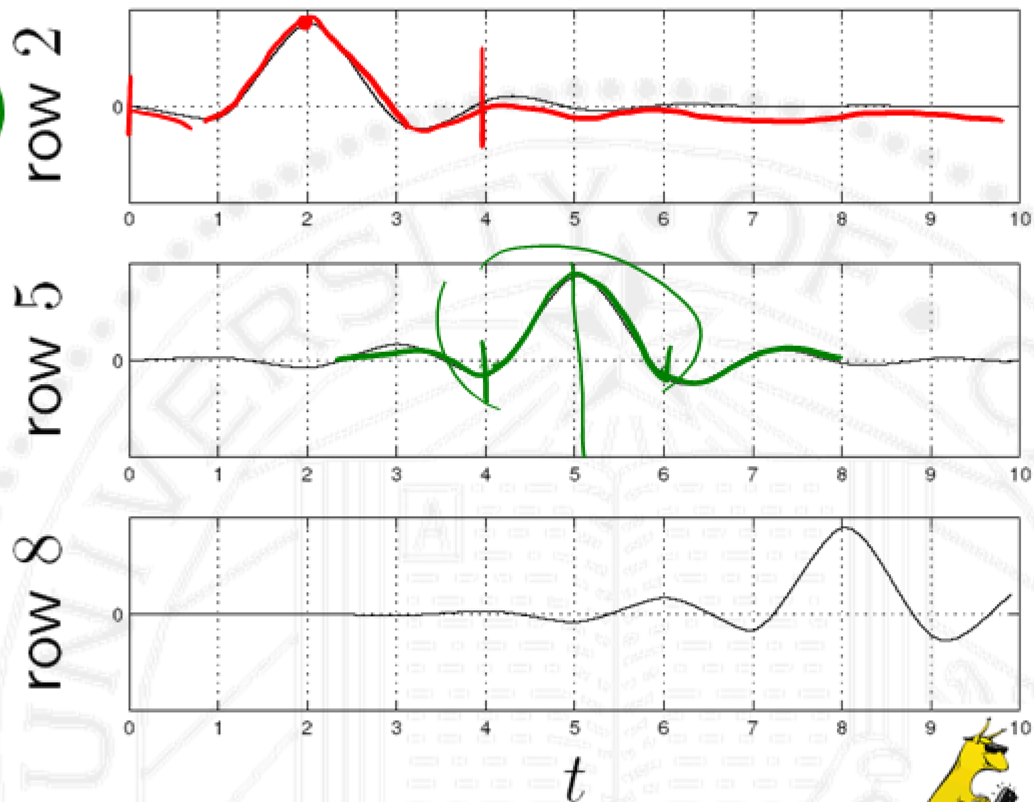
Comm. Example (2.3)



// Rows of $B_{ls} = A^{+}$ $\hat{x} = B_{ls}\, y$

Comm. Example (3.3)

$X_5 \sim f(y^{40-60})$



Questions?

$\langle u_i , u_j \rangle = u_i^T u_j$ — inner product

orthogonal vectors — $\langle u_i , u_j \rangle = 0 \;\; \forall \; i \neq j$

$a_1^T a_2$ or $\langle a_1 , a_2 \rangle$

normal vector $\langle u_i , u_i \rangle = 1$

Orthonormal vectors — both $\left\{ \begin{array}{l} \langle u_i , u_j \rangle = 0 \;\; i \neq j \\ = 1 \;\; i = j \end{array} \right.$

$U = [u_1 \dots u_k]$ <u>orthogonal</u>