Least Squares

Background: $Hx \cong y$ (symbol for inconsistency) $\quad H \in \mathbb{R}^{N\times n} \quad N \geq n$

$$\underbrace{[\ ]}_{N\times n}\underbrace{[\ ]}_{n\times 1} = \underbrace{[\ ]}_{N\times 1}$$

→ Overdetermined
inconsistent

$y \notin \mathcal{R}(H)$ [$\mathcal{R}(H)$: column space of H]

So $y = Hx + \underbrace{v}_{\text{residual}}$ for some $v \in \mathbb{R}^{N\times 1}$

Least square soln: $\hat{x}$ is one that minimizes $\|v\|^2$.

i.e. $\|y - H\hat{x}\|^2 \leq \|y - Hx\|^2 \quad \forall x \in \mathbb{R}^n$

If $y \in \mathcal{R}(H)$, then $\hat{x}$ is an exact soln
Then $\exists$ infinitely many ($\hat{x}$) exact solutions. [↳ not unique

Q. If $y \notin \mathcal{R}(H)$ then is $\hat{x}$ unique?

FACT: A vector $\hat{x}$ is a minimizer of the cost function $J(x) = \|y - Hx\|^2$ iff it satisfies the (always consistent) normal eqns:

$$H^T H \hat{x} = H^T y$$

The minimum is $J(\hat{x}) = \|y\|^2 - \|H\hat{x}\|^2$

Proof: $0 = \frac{\partial}{\partial x} J = \frac{\partial}{\partial x}\left(x^T H^T H x - x^T H^T y - y^T H x + y^T y\right)$

$$= \hat{x}^T H^T H - y^T H = 0$$

Also $\frac{\partial^2}{\partial x^2} J = H^T H \geq 0 \Rightarrow$ So minimum.

$J(\hat{x}) = \|y\|^2 - \|H\hat{x}\|^2 \longrightarrow$ Exercise

Q. Why are the normal equations always consistent?
→ i.e. why will it have solutions always?
FACT: $H \in \mathbb{R}^{N \times n}$. Then $R(H^T H) = R(H^T)$: Proof. Exercise

FACT: Consider $H^T H \hat{x} = H^T y$

a) When H is full rank, the unique solution is given by $\hat{x} = (H^T H)^{-1} H^T y$

b) When H is not full-rank, the normal eqns always have more than one solution, where any two solutions $\hat{x}_1$ & $\hat{x}_2$ satisfy:

$$H(\hat{x}_1 - \hat{x}_2) = 0$$

c) The projection of y onto $R(H)$ is unique & is defined as $\hat{y} := H\hat{x}$, where $\hat{x}$ is any solution to the normal eqns. When H has full rank, $\hat{y} = H(H^T H)^{-1} H^T y$

Proof: (a) → H full rank $\Leftrightarrow H^T H$ is non-singular
(Proof: Exercise)

(b) $\mathcal{R}(H^TH) = \mathcal{R}(H)$
$\Leftrightarrow \mathcal{N}(H^TH) = \mathcal{N}(H)$] so $\begin{cases} H^TH(\hat{x}_1 - \hat{x}_2) = 0 \\ \Leftrightarrow H(\hat{x}_1 - \hat{x}_1) = 0 \end{cases}$

(C) From (b) $\hat{y}_1 = H\hat{x}_1$ $\qquad \hat{y}_2 = H\hat{x}_2$

$$\hat{y}_1 - \hat{y}_2 = H(\hat{x}_1 - \hat{x}_2) = 0.$$

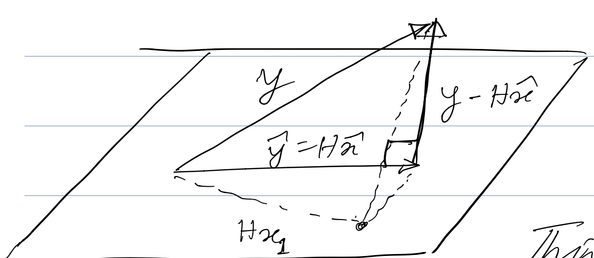
Geometry: Recall: $\min J = \|y - H\hat{x}\|^2 = \|y\|^2 - \|H\hat{x}\|^2$

Pythagoras thm $\Leftarrow \|y - H\hat{x}\|^2 + \|H\hat{x}\|^2 = \|y\|^2$

OR

$H^T(y - H\hat{x}) = 0$

i.e. $(y - H\hat{x})$ is $\perp$ to $\mathcal{R}(H)$



$\mathcal{R}(H)$

This arguments extend for arbitrary dimensions for inner-product spaces:

Defn: Let L be a linear subspace of a inner product space V & let $y \in V$. The projection of y onto L, denoted by the unique $\hat{y}_L \in L$ s.t. $\langle y - \hat{y}_L, a \rangle = 0$ $\forall\, a \in L$.

Proof of Uniqueness & Existence: Exercise.

FACT: $\|y - \hat{y}_L\|^2 \leq \|y - a\|^2 \;\; \forall\, a \in L$

Proof: $\|y-a\|^2 = \|y - \bar{y}_\perp + \bar{y}_\perp - a\|^2$ ..... Rest: Exercise.

If H has full rank, an explicit projection matrix can be written:

$$\hat{y} = H\hat{x} = \underbrace{H(H^TH)^{-1}H^T}\, y =: P_H y$$

Verify: (i) $P_H^T = P_H$ (ii) $P_H = P_H^2$ (3) $P_H^\perp := I - P_H$ projects only $R^\perp(H)$

Application of Geometric approach: Order Recursion (Size of x increases)

Let $\hat{x}_{n,N} \in \mathbb{R}^{n\times 1}$ denote the LS sol$\underline{\underline{n}}$ for $Hx \cong y$ where $H \in \mathbb{R}^{N\times n}$ has full column rank.

Suppose we add one column to H & one entry to x.

$$\begin{bmatrix} H & \underline{h}_n \end{bmatrix} \begin{bmatrix} x \\ x(n) \end{bmatrix} \cong y$$

Assume full col. rank.

Denote sol$\underline{\underline{n}}$. by $\hat{x}_{n+1,N} \in \mathbb{R}^{(n+1)\times 1}$

Q. Is $\hat{x}_{n+1,N}$ related to $\hat{x}_{n,N}$?

$$\hat{x}_{n+1,N} = \left(\begin{bmatrix} H^T \\ \underline{h}_n^T \end{bmatrix} \begin{bmatrix} H & \underline{h}_n \end{bmatrix} \right)^{-1} \begin{bmatrix} H^T \\ \underline{h}_n^T \end{bmatrix} y$$

Denote $\hat{h}_n = P_H \underline{h}_n = H\underbrace{(H^TH)^{-1}H^T\underline{h}_n}_{a}$ (Projection onto $R(H)$)

$= Ha$

Residual: $\tilde{h}_n = h_n - \hat{h}_n = h_n - Ha$

$[H\ \tilde{h}_n]$ provides a new basis for $R\{[H\ h_n]\}$

But $\tilde{h}_n$ is $\perp$ to H

Projection of y onto $R\{[H\ h_n]\}$ $\equiv$ Projection of y onto $R\{[H\ \tilde{h}_n]\}$

$$\hat{y}_{n+1} = \underbrace{P_H y + P_{\tilde{h}_n} y}_{\text{since } \tilde{h}_n \perp H} = H\hat{x}_{n,N} + P_{\tilde{h}_n} y \quad \text{—} (1)$$

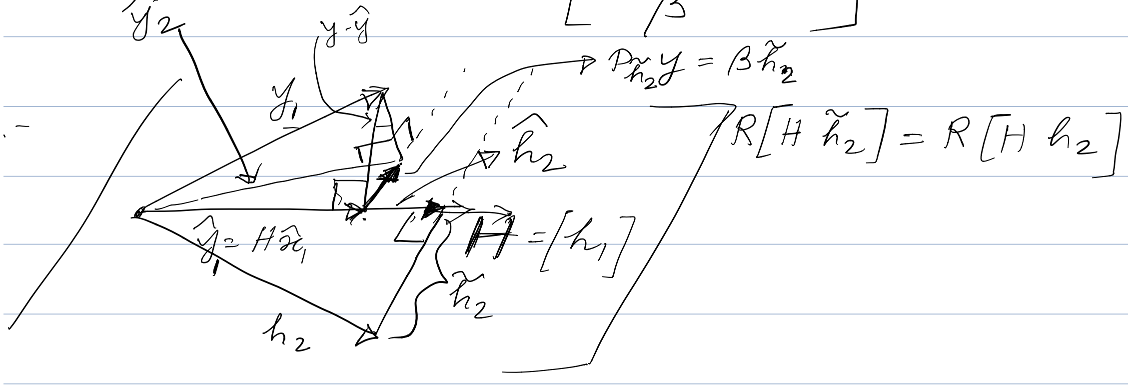
Now, $P_{\tilde{h}_n} y = \dfrac{\tilde{h}_n^T y}{\|\tilde{h}_n\|^2}\tilde{h}_n = \beta\tilde{h}_n$, $\beta \longrightarrow$ scalar

So from (1), $\hat{y}_{n+1} = H\hat{x}_{n,N} + \beta\tilde{h}_n = H\hat{x}_{n,N} + \beta[h_n - Ha]$

$$= [H\ \ h_n]\begin{bmatrix}\hat{x}_{n,N} - \beta a \\ \beta\end{bmatrix}$$

But $\hat{y}_{n+1} = [H\ \ h_n]\,\hat{x}_{n+1,N}$ & $\hat{x}_{n+1,N}$ is unique.

Hence $\hat{x}_{n+1,N} = \begin{bmatrix}\hat{x}_{n,N} - \alpha\beta \\ \beta\end{bmatrix}$



Regularized Least Squares

$$J(x) = (x - x_0)^T \Pi_0^{-1}(x - x_0) + \|y - Hx\|^2$$

$\Pi_0 > 0$. Clearly, if $\Pi_0 = \infty I$, then we recover earlier problem.

$\{\Pi_0, x_0\}$ can be tuned to factor in prior information about solution.
unique $\hat{x}$ guaranteed, even when H is rank deficient
If H is already full rank, Π_0 can be used to improve condition no. (in normal eqns)

Define $x' = x - x_0$, $y' \triangleq y - Hx_0$

Then $\min_{x'} \left[x'^T \Pi_0^{-1} x + \|y' - Hx'\|^2 \right]$

$$= \min_{x'} \underbrace{\left\| \begin{bmatrix} 0 \\ y' \end{bmatrix} - \begin{bmatrix} \Pi_0^{-1/2} \\ H \end{bmatrix} x' \right\|^2}_{\text{Same form as before}} \qquad \boxed{\Pi_0 = \Pi_0^{1/2} \Pi_0^{1/2}}$$

<u>Solution</u>

$$\hat{x} = x_0 + [\underbrace{\Pi_0^{-1} + H^T H}_{\text{Always invertible – so explicit form always possible.}}]^{-1} H^T [y - Hx_0]$$

<u>Recursive Least Square</u>: N increases sequentially.

$H_{i-1}\, x \cong y_{i-1}$

Let $x_0 = 0$. Then LS at $i-1$:

$\min_x \left[x^T \Pi_0^{-1} x + \|y_{i-1} - H_{i-1} x\|^2 \right]$

Say sol^n is $\hat{x}_{i-1}$.

$$H_{i-1} = \begin{bmatrix} h_0 \\ \vdots \\ h_{i-1} \end{bmatrix} \in \mathbb{R}^{i \times n} \qquad y_{i-1} = \begin{bmatrix} y(0) \\ \vdots \\ y(i-1) \end{bmatrix}$$

Now, we get another data pt. h_i, $y(i)$: we aim to find

$\min_x \left[x^T \Pi_0^{-1} x + \|y_i - H_i x\|^2 \right] \to$ sol^n is $\hat{x}_i$

Q. Is it possible to calculate $\hat{x}_i$ without re-inverting the matrix $(\Pi_0^{-1} + H_i^T H)$ again?

$$\hat{x}_i = \big(\underbrace{\Pi_0^{-1} + H_i^T H_i}_{P_i}\big)^{-1} H_i^T y_i$$
$$= \left(\Pi_0^{-1} + H_{i-1}^T H_{i-1} + h_i^T h_i\right)^{-1} \Big[H_{i-1}^T y_{i-1} + h_i^T y(i)\Big] \quad (1)$$

Define: $P_i = \left(\Pi_0^{-1} + H_i^T H_i\right)^{-1}$ $\qquad P_{-1} = \Pi_0$

Then: $P_i^{-1} = P_{i-1}^{-1} + h_i^T h_i \qquad P_{-1}^{-1} = \Pi_0^{-1}$

$$\Big[\text{Use: } [A + BCD]^{-1} = A^{-1} - A^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1}\Big]$$

Identify: $A = P_{i-1}^{-1}$; $B = h_i^T$; $C = 1$; $D = h_i$

Then;
$$P_i = P_{i-1} - \frac{P_{i-1} h_i^T h_i P_{i-1}}{1 + h_i P_{i-1} h_i^T} \qquad P_{-1} = \Pi_0$$

So, from (1): $\hat{x}_i = P_{i-1} H_{i-1}^T y_{i-1} - \ldots \ldots$ Exercise
$$= \hat{x}_{i-1} + \frac{P_{i-1} h_i^T}{1 + h_i P_{i-1} h_i^T}\left(y(i) - h_i \hat{x}_{i-1}\right)$$

Lemma: Solution of RLS $\hat{x}_i$ satisfies:
$$\hat{x}_i = \hat{x}_{i-1} + K_{p,i}\left(y(i) - h_i \hat{x}_{i-1}\right) \qquad \hat{x}_{-1} = 0$$

where $\quad K_{p,i} = P_{i-1} h_i^T r_e^{-1}(i)$;
$$r_{e,i} = 1 + h_i P_{i-1} h_i^T$$
$$P_i = P_{i-1} - P_{i-1} h_i^T \left(1 + h_i P_{i-1} h_i^T\right)^{-1} h_i P_{i-1}$$
$$P_{-1} = \Pi_0$$

Effort Required for 1-step recursion is $O(n^2)$ flops. (as compared to $O(n^3)$ for the matrix inversion)

<u>Surprising observation</u>: The above sol^{n} is the (Kalman filter) solution for:

$$x_{j+1} = x_j \qquad x_0 = x \in \mathbb{R}^n$$

$$y_j = h_j x_j + v(j)$$

$$E x_0 x_0^T = \Pi_0 \quad ; \quad E v(i) v^T(j) = \delta_{ij}$$

Q. Why? $\rightarrow$ $y_i = \begin{bmatrix} y(0) \\ \vdots \\ y(i) \end{bmatrix} \quad v_i = \begin{bmatrix} v(0) \\ \vdots \\ v(i) \end{bmatrix} \Rightarrow y_i = Hx + v_i$

STOCHASTIC LEAST SQUARES

X, Y are two vector valued random variables with joint density $f_{x,y}(\cdot,\cdot)$.

<u>Problem</u>: Given $Y = y$, find an estimate $\hat{x}$ of x. ie. find $h(\cdot)$ s.t. $\hat{x} = h(y)$ is a "good" estimate of x.

FACT: The estimator $\hat{x} = h(y)$ that solves:

$$\min_{h(\cdot)} E\left([x-\hat{x}][x-\hat{x}]^T\right)$$

(least mean square (lms) estimator of r.v. x given the value of y is $\hat{x} = E(x/y)$.

Note: As opposed to the parameter estimation problem, x & y are both random variables.

Assume $g(y)$

Proof: $E\left[(x-\hat{x})(x-\hat{x})^T\right] = E\left[\left(x - E(x/y) + \overbrace{E(x/y) - \hat{x}}^{g(y)}\right)\left(x - E(x/y) + E(x/y) - \hat{x}\right)^T\right]$

$\boxed{\begin{array}{l} \text{Recall } E[x g^T(y)] = E\{E\{x g^T(y)/y\}\} = E\{E\{x/y\} g^T(y)\} \\ \quad \text{for any } g(y) \\ \text{Hence here: } E\left[(x - E(x/y))\, g^T(y)\right] = E[x g^T(y)] - E\left[E(x/y) g^T(y)\right] \\ \qquad = 0 \end{array}}$

$$= E\left[x - E(x/y)\right]\left[x - E(x/y)\right]^T + E\left[E(x/y) - \hat{x}\right]\left[E(x/y) - \hat{x}\right]^T$$

So minimum is achieved by $\hat{x} = E(x/y)$

But we usually don't know $f(x/y)$.

So we restrict ourselves to linear estimators of the form $\hat{x} = K_0 y$

$$\begin{bmatrix} \hat{x}_1 \\ \vdots \\ \hat{x}_n \end{bmatrix} = \underbrace{\begin{bmatrix} K_0 \end{bmatrix}}_{n \times p(N+1)} \underbrace{\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_N \end{bmatrix}}_{p(N+1) \times 1} \begin{array}{l} \}p \\ \}p \end{array}$$

$y \to N$ observations of a p-dim random variable.

Find K_0 such that for every $K \in \mathbb{R}^{n \times p(N+1)}$

Compare with det. L.S.: $y \cong Hx \to H, y$ - known find $\hat{x}$ s.t. $\|y - H\hat{x}\|$ min.

Here $x \cong K_0 y$, some information about x, y known. Find K_0 s.t. $\|x - K_0 y\|$ is min.

$$\underbrace{P(K)}_{n\times n} := E[x - Ky][x - Ky]^T \underset{\text{P.S.D.}}{\geq} P(K_0) := E\left[(x-\hat{x})(x-\hat{x})^T\right] \quad (1)$$

Thm: (Optimal LMS): Given two zero mean r.v.s x & y, the l.l.m.s estimator of x given y (i.e. satisfying (1)) is given by any solution K_0 of the "normal eqs": $\quad K_0 R_y = R_{xy}$

where $R_y = Eyy^T$ & $R_{xy} = Exy^T = R_{yx}^T$

The merr $\quad P(K_0) = R_x - K_0 R_{yx} = R_x - R_{xy} K_0^T$

Proof: (1) above $\Leftrightarrow \; aP(K)a^T \geq aP(K_0)a^T \quad \forall a \in \mathbb{R}^n$

$$\overbrace{aP(K)a^T}^{\text{scalar}} = aE\left[(x-Ky)(x-Ky)^T\right]a^T = a\left[R_x - R_{xy}K^T - KR_{yx} + KR_yK^T\right]a^T$$

$$= aR_x - aR_{xy}(aK)^T - (aK)R_{yx}a^T + (aK)R_y(aK)^T$$

$$\frac{\partial aP(K)a^T}{\partial (aK)} = 0 \Leftrightarrow -aR_{xy} - aR_{xy} + 2(aK_0)R_y = 0 \quad \forall a$$

$$\Leftrightarrow \; a\left[R_{xy} - K_0R_y\right] = 0 \quad \forall a$$

$$\Leftrightarrow \; R_{xy} = K_0R_y$$

Corresponding mmse: $P(K_0) = \cdots = R_x - K_0 R_{yx}$

The optimal K_0 also minimizes mse in the estimator of each component of x.

If $R_y > 0$, $\; K_0 = R_{xy}R_y^{-1}$ & $P(K_0) = R_x - R_{xy}R_y^{-1}R_{yx}$

R_y singular is "unusual" since $\Rightarrow \exists\, c \in \mathbb{R}^{(N+1)p}, c \neq 0$ s.t. $c^TR_y = 0 \;\Leftrightarrow\; 0 = c^TR_yc = c^T(Eyy^T)c = \underbrace{E[c^Ty]^2}_{=0}$

i.e. $\quad \text{Mean}(c^Ty) = 0$ & $\text{Var}(c^Ty) = 0$

$\Leftrightarrow c^Ty = 0$ almost surely.

This usually means, something is wrong with the problem formulation.

Even then if R_y is assumed to be singular:

<u>Thm:</u> Even if R_y is singular, the normal equations $K_0 R_y = R_{xy}$ will be consistent, and there will be many solutions. No matter which solution K_0 is used, the corr. llms estimator $\hat{x} = K_0 y$ & $P(K_0)$ will each be unique.

<u>A geometric interpretation:</u> Consider the normal eqns:

$$K_0 R_y = R_{xy} \Leftrightarrow K_0 E y y^T = E x y^T$$
$$\Leftrightarrow E(x - K_0 y) y^T = 0 \quad \text{—— (1)}$$

Q. Can we consider x, y as vectors in an inner product space? with $\langle x, y \rangle := E x y^T$

Then (1) is eq. to $x - K_0 y \perp y$

<u>Potential problems:</u> $x : \Omega \to \mathbb{R}^n$, $y : \Omega \to \mathbb{R}^p$

$E x y^T \in \mathbb{R}^{n \times p}$ rect. matrix (not scalar)

<u>Check:</u> 1) Linearity: $\langle a_1 x_1 + a_2 x_2, y \rangle = a_1 \langle x_1, y \rangle + a_2 \langle x_2, y \rangle$ $\forall a_1, a_2 \in \mathbb{R}$

2) Reflexivity: $\langle x, y \rangle = \langle y, x \rangle^T$ $\longrightarrow$ Easy to check

3) Non-degeneracy: $\|x\|^2 = \langle x, x \rangle = 0 \Leftrightarrow x = 0$ $\longrightarrow$ Equality almost surely.

(Additional assumption: $Ex = 0$ req.)

So $\langle x, y\rangle = E\, x y^T$ is a valid inner product.

Then given x & $y = \begin{bmatrix} y_0 \\ \vdots \\ y_N \end{bmatrix} \to \mathbb{R}^{p\times 1}$, find $\hat{x} = K y$ s.t.

$E[x-\hat{x}][x-\hat{x}]^T = \langle\langle (x-\hat{x}), (x-\hat{x})\rangle\rangle$ $\underbrace{\text{is minimized}}_{\text{PSD}}$.

Recall: V, lin subspace $- L$.

$\|y - \hat{y}_L\| < \|y - a\| \ \forall a \in L$

iff $y - \hat{y}_L \perp L \iff \langle y - \hat{y}_L, a\rangle = 0 \ \forall a \in L$

Here $\hat{x}$ is Ky i.e. $\hat{x} \in L :=$ col sp. of y

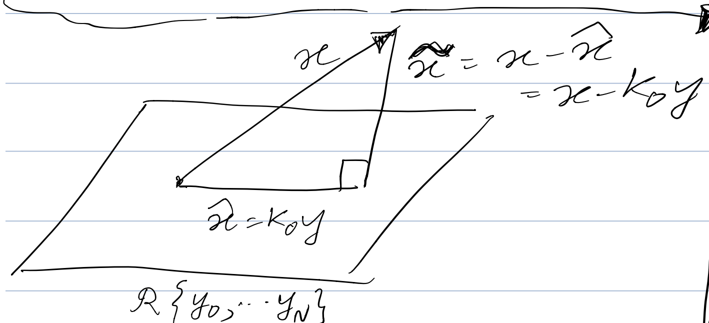
So $\|x - \hat{x}\|$ is minimized $\iff x - \hat{x} \perp \operatorname{span}\{Ky\}$

$\iff \langle x - K_0 y, Ky\rangle = 0$; $\langle x - K_0 y, y\rangle K^T = 0 \ \forall K$

$\iff \langle x, y_i\rangle = K_0 \langle y, y_i\rangle \quad \forall i = 0, \dots, N$.

$\iff R_{xy} = K_0 R_y \to$ Existence & uniqueness of projection is guaranteed by inner prod. sp. properties (complete? Hilbert sp.?)

Valid even if R_y singular



$\langle x - K_0 y, y\rangle = E \underbrace{\left[\tilde{x}\right]}_{n\times 1} \underbrace{[\underbrace{y_0}_{p}, \underbrace{y_1}_{p}, \dots]}_{(N+1)p}$

$= \left[\begin{array}{c|c|c} \tilde{x} y_0 & \tilde{x} y_1 & \dots \\ n\times p & n\times p & \end{array}\right] = 0$

$\iff$ Each $\langle x - K_0 y, y_i\rangle = 0$

FACT: The llms estimator of r.v. x given a set of other r.v. y is characterized by the error

$\tilde{x}$ being orthogonal (uncorrelated with) each of the r.v.'s used to form the estimate.

Proof: True since $(x - K_0 y) \perp y_i \quad \forall\, i = 0, \dots, N$.

Examples: Consider zero-mean stationary process $\{y(t)\}$ with autocovariance fn. $\langle y(t), y(t-\tau)\rangle = R_y(\tau)$

Find llms estimator of $\int_0^T y(t)\,dt$ in term of $y(0)$ & $y(T)$.

Let $Z = \int_0^T y(t)\,dt$ & $\hat{Z} = a y(0) + b y(T)$ (we find a & b)

The orthogonality condition:

$$Z - \hat{Z} \perp sp.\{y(0), y(T)\}$$

i.e $\left[\int_0^T y(t)dt - a y(0) - b y(T)\right] \perp sp\{y(0), y(T)\}$

$$\left\langle \int_0^T y(t)dt - ay(0) - by(T),\; y(0)\right\rangle = 0$$

$$\Leftrightarrow \int_0^T R_y(t)dt - aR_y(0) - bR_y(T) = 0 \quad \text{—— (1)}$$

Also;

$$\left\langle \int_0^T y(t)dt - ay(0) - by(T),\; y(T)\right\rangle = 0$$

$$\Leftrightarrow \int_0^T R_y(t-T)\,dt - aR_y(-T) - bR_y(0) = 0$$

$$= \int_0^T R_y(t)dt - aR_y(T) - bR_y(0) = 0 \quad \text{—— (2)}$$

$$\begin{bmatrix} R_y(0) & R_y(T) \\ R_y(T) & R_y(0) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \int_0^T R_y(t)\,dt$$

Solving: $\hat{z} = \dfrac{\int_0^T R_y(t)\,dt}{R_y(0) + R_y(T)}\left[y(0) + y(T)\right]$

<u>Linear Models</u>

$y = Hx + v$ $\qquad y \in \mathbb{R}^p$, $H = \mathbb{R}^{p \times n}$

$x \in \mathbb{R}^n$ $\qquad \hookrightarrow$ <u>known</u>

v is zero mean random noise

<u>Known:</u> H, R_x, R_v, x & v are uncorrelated.

So $R_y = HR_xH^T + R_v$ (Assume $R_y > 0$)

Then, llmse $\hat{x} = K_0 y$, $K_0 = R_xH^T\left[HR_xH^T + R_v\right]^{-1}$

Also $P(K_0) = R_x - R_xH^T\left[R_v + HR_xH^T\right]^{-1}HR_x$

Equivalent forms: $K_0 = \left(R_x^{-1} + H^TR_v^{-1}H\right)^{-1}H^TR_v^{-1}$

$P(K_0) = \left(R_x^{-1} + H^TR_v^{-1}H\right)^{-1}$

(derived by using $(A+BCD)^{-1} = \dots$ formula)

Sometime useful

<u>Gauss Markov Thm:</u>

x is not a r.v. but <u>unknown</u> constant.

Equivalently, $R_x = \alpha R$ with $\alpha \to \infty$.

From above, $\hat{x}_\infty := \left(H^TR_v^{-1}H\right)^{-1}H^TR_v^{-1}y$

$= \left(H^TH\right)^{-1}H^Ty$ if $R_v = I$

$P_\infty(K_0) := \left(H^TR_v^{-1}H\right)^{-1} = \left(H^TH\right)^{-1}$

(Assume H full rank)

<u>G-M Thm</u> (again) Consider $y = Hx + v$, where v is zero mean r.v. with unit variance $\langle v, v\rangle = I$

x is a deterministic vector, & H has full column rank. Then $\hat{x}_{\infty} = (H^T H)^{-1} H^T y$ is the optimum unbiased linear least-mean-square estimator of x.

Proof: Exercise (without this trick of $R_x = \alpha R$ with $\alpha \to \infty$)

Q. Is it MVUE? Refer to previous statement of GM thm.