Linear Estimation

Note Title 11-08-2014

$\underline{X} \rightarrow$ random vector $\quad X_i$'s are complex random variables

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_M \end{bmatrix} \qquad \text{Mean vector } E(\underline{X}) = \begin{bmatrix} E(x_1) \\ \vdots \\ E(x_M) \end{bmatrix}$$

Covariance Matrix $\underset{(M \times M)}{R_x} = E\left\{ (\underline{X} - E(\underline{X}))(\underline{X} - E(\underline{x}))^* \right\}$

$* \rightarrow$ conjugate & transpose

$$R_x = E\left\{ \begin{bmatrix} x_1 - E(x_1) \\ \vdots \\ x_M - E(x_M) \end{bmatrix} \left[(x_1 - E(x_1))^* \; \dots \; (x_M - E(x_M))^* \right] \right\}$$

$$= \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1M} \\ r_{21} & & & \vdots \\ \vdots & & & \\ r_{2M} & & \dots & r_{MM} \end{bmatrix}$$

where $r_{\ell k} = E\left\{ (x_\ell - E(x_\ell))(x_k - E(x_k))^* \right\}$

Properties of Covariance Matrix R_x

① R_x is conjugate symmetric (also called hermitian)

$R_x^* = R_x$

ie) $R_x = R_x^*$; ie) $r_{\ell k} = r_{k\ell}^*$

Proof: $r_{\ell k} = E\left\{ (x_\ell - E(x_\ell))(x_k - E(x_k))^* \right\}$

$$= E\left\{ \left[(x_k - E(x_k))(x_\ell - E(x_\ell))^* \right]^* \right\}$$

interchanging expectation & conjugation $\rightarrow$ $= \left[E\{ (x_k - E(x_k)) (x_l - E(x_l))^* \} \right]^*$

$= r_{kl}^*$

conjugation and transpose = hermitian

12/8 (2) Eigen vectors and eigen values of R_x

(spectral Theorem: Valid for any conjugate symmetric matrix)

Let $\underline{q_i}$ be $M \times 1$ vector such that

$$R_x \underline{q_i} = \lambda_i \underline{q_i}$$

$\lambda_i \rightarrow$ eigen value $\quad \underline{q_i} \rightarrow$ eigen vector

Thm: (1) All eigen values of R_x $\{\lambda_1, \lambda_2, .., \lambda_M\}$ are real

(2) There exist M orthonormal eigen vectors for R_x.

ie) there exist eigen vectors $\{\underline{q_1}, .., \underline{q_M}\}$ such that

$$\underline{q_i}^* \underline{q_j} = 0 \quad \forall i \neq j$$

$$\underline{q_i}^* \underline{q_i} = 1 \quad \forall i$$

We can rewrite spectral theorem as

Let $Q = [\underline{q}_1 \ \underline{q}_2 \ \cdots \ \underline{q}_M]$

$M \times M$ matrix with columns being orthonormal eigen vectors of R_x

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_M \end{bmatrix}$$

diagonal matrix containing eigen values of R_x

We can write

$$\boxed{R_x = Q \Lambda Q^*}$$

Q is <u>unitary</u> matrix

<u>Verify</u>:

Note $Q Q^* = I = Q^* Q$.

$$Q^* Q = \begin{bmatrix} \underline{q}_1^* \\ \underline{q}_2^* \\ \vdots \\ \underline{q}_M^* \end{bmatrix} [\underline{q}_1 \ \underline{q}_2 \ \cdots \ \underline{q}_M]$$

$$= I$$

$\underline{q}_i^* \underline{q}_j = 1$ if $i = j$

0 $i \neq j$

Consider $R_x Q = R_x [\underline{q}_1 \ \underline{q}_2 \ \cdots \ \underline{q}_M]$

$$= [R_x \underline{q}_1 \ \ R_x \underline{q}_2 \ \cdots \ R_x \underline{q}_M]$$

$$= [\lambda_1 \underline{q}_1 \quad \lambda_2 \underline{q}_2 \quad \dots \quad \lambda_M \underline{q}_M]$$

$$= [\underline{q}_1 \quad \underline{q}_2 \quad \dots \underline{q}_M] \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \ \lambda_M \end{bmatrix}$$

$$R_x Q = Q \Lambda$$

Post Multiply with Q^*,

$$R_x \underbrace{Q Q^*}_{I} = Q \Lambda Q^*$$

$$R_x = Q \Lambda Q^*$$

<u>Property 3)</u> R_x is positive semi-definite

ie) (a) All eigen values of R_x are non-negative

or equivalently

(b) for any complex $M \times 1$ vector $\underline{a}$,

we have $\underline{a}^* R_x \underline{a} \geq 0$

<u>Proof.</u> (b) First note $\underline{a}^* R_x \underline{a}$ is always real for any $\underline{a}$.

$$(\underline{a}^* R_x \underline{a})^* = \underline{a}^* R_x^* (\underline{a}^*)^*$$

$$= \underline{a}^* R_X \underline{a}$$

$$\Rightarrow \underline{a}^* R_X \underline{a} \text{ is } \underline{\text{real}}.$$

Now let $\underset{\text{random variable}}{Y} = \underline{a}^* (\underline{X} - E(\underline{X}))$

$1 \times M \quad M \times 1 = 1 \times 1$

We have

$$0 \leq E\{|Y|^2\} = E\{Y Y^*\}$$

$$= E\{\underline{a}^* (\underline{X} - E(\underline{X})) [\underline{a}^* (\underline{X} - E(\underline{X}))]^*\}$$

$$= E\{\underline{a}^* (\underline{X} - E(\underline{X})) [\underline{X} - E(\underline{X})]^* (\underline{a}^*)^*\}$$

$$= \underline{a}^* \underbrace{E\{(\underline{X} - E(\underline{X}))(\underline{X} - E(\underline{X}))^*\}}_{R_X} \underline{a}$$

$$= \underline{a}^* R_X \underline{a}$$

So $\underline{a}^* R_X \underline{a} = E\{|Y|^2\} \geq 0.$

Proof (Part a)

$$\underline{a}^* R_X \underline{a} \geq 0 \quad \text{for any } \underline{a}$$

Substitute $R_X = Q \Lambda Q^*$

$$\underbrace{\underline{a}^* Q}_{\underline{b}^*} \Lambda \underbrace{Q^* \underline{a}}_{\underline{b}_{M\times 1}} \geq 0 \quad \text{for any } \underline{a}$$

$$\underbrace{\underline{b}^* \Lambda \underline{b}} \geq 0 \quad \text{for any } \underline{b}.$$

$$\underline{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix}$$

$$|b_1|^2 \lambda_1 + |b_2|^2 \lambda_2 + \cdots + |b_M|^2 \lambda_M \geq 0$$

for any $\underline{b}$

$\Rightarrow$ All λ_i's are non-negative.

x ———————— x

13/8

Case 1: Suppose $\underline{a}^* R_x \underline{a} > 0$ for all non-zero $\underline{a}$

(Positive definiteness).

Then R_x is invertible (full rank)

All eigen values are positive. $\lambda_i > 0$

$$R_x^{-1} = Q \Lambda^{-1} Q^*$$

$$\underline{\underline{\Lambda}}^{-1} = \begin{bmatrix} 1/\lambda_1 & & & 0 \\ & 1/\lambda_2 & & \\ & & \ddots & \\ 0 & & & 1/\lambda_M \end{bmatrix}$$

Proof: We have $R_x = Q \Lambda Q^*$

Consider $R_x R_x^{-1} = (Q \Lambda \underbrace{Q^*)(Q}_{I} \Lambda^{-1} Q^*)$

$=$

$$= Q \underbrace{\Lambda \, \Lambda^{-1}}_{I} Q^*$$

$$= Q Q^* = I$$

$\times$ ———————— $\times$

Case 2: Suppose $\underline{a}^* R_x \underline{a} = 0$ for some non-zero $\underline{a}$.

$\underline{x} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix}$

Then R_x is not invertible

Also, there exists scalars $\alpha_1, \alpha_2, \dots, \alpha_m$ (not all zeros)

such that $\alpha_1 (X_1 - E(X_1)) + \alpha_2 (X_2 - E(X_2)) + \dots + \alpha_n (X_M - E(X_M)) = 0$

Proof. We know that $\exists$ non zero $\underline{a}$ such that $\underline{a}^* R_x \underline{a} = 0$

Pick one such $\underline{a}$.

Let $Y = \underline{a}^* (\underline{X} - E(\underline{X}))$,

Will show that $Y = 0$

$$E\{|Y|^2\} = E\{Y Y^*\} = \underline{a}^* R_x \underline{a} = 0.$$

Also $E(Y) = 0$. So $Y = 0$.

$\times$ ———————— $\times$

Cross Covariance Matrix

$\underline{X}_{L\times 1}$ random vector

$\underline{Y}$ $M\times 1$ random vector

$$R_{XY} = E\left\{ \left(\underline{X} - E(\underline{X})\right)\left(\underline{Y} - E(\underline{Y})\right)^{*} \right\}$$

definition

$\Downarrow$

$L\times M$ matrix

$$R_{YX} = E\left\{ \left(\underline{Y} - E(\underline{Y})\right)\left(\underline{X} - E(\underline{X})\right)^{*} \right\}$$

$\Downarrow$

$M\times L$ matrix

we have $R_{XY}^{*} = R_{YX}$

x ———————— x

Linear MMSE Estimation

Let $\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_L \end{bmatrix}$ be (unknown) random vector

$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_M \end{bmatrix}$ be (observed) random vector.

Consider estimation of $\underline{X}$ given $\underline{Y}$

$$\hat{\underline{X}} = \begin{bmatrix} \hat{X}_1 \\ \vdots \\ \hat{X}_L \end{bmatrix} = g(\underline{Y})$$

Restrict $g(.)$ to be a linear function of $\underline{Y}$

$$\underline{\hat{X}} = K\underline{Y} + \underline{b}$$

$$\left.\begin{array}{lll} K \rightarrow & L \times M & \text{matrix} \\ \underline{b} \rightarrow & L \times 1 & \text{vector} \end{array}\right\} \text{fixed.}$$

Linear MMSE problem:

$$\text{Error } \underline{\tilde{X}} = \begin{bmatrix} X_1 - \hat{X}_1 \\ \vdots \\ X_L - \hat{X}_L \end{bmatrix} = \underline{X} - \underline{\hat{X}}$$

Find K and $\underline{b}$ such that

$$\tilde{X}_i = X_i - \hat{X}_i$$

$$E\{\underline{\tilde{X}}^* \underline{\tilde{X}}\} \text{ is minimum.}$$

$$= E\{|\tilde{X}_1|^2 + |\tilde{X}_2|^2 + \dots + |\tilde{X}_L|^2\}$$

x ——— x

We consider zero-mean case where

$$E(\underline{X}) = \underline{0} \text{ and } E(\underline{Y}) = \underline{0}$$

Theorem: In the zero-mean case,

optimal value of $\underline{b} = \underline{0}$

Proof:

$$\underline{\tilde{X}} = \underline{X} - (K\underline{Y} + \underline{b})$$

$$E\{\tilde{\underline{X}}^*\tilde{\underline{X}}\} = E\{(\underbrace{\underline{X}}-K\underline{Y}-\underline{b})^*(\underbrace{\underline{X}-K\underline{Y}}-\underline{b})\}$$

$$= E\{(\underline{X}-K\underline{Y})^*(\underline{X}-K\underline{Y}) + \underline{b}^*\underline{b} - (\underline{X}-K\underline{Y})^*\underline{b} - \underline{b}^*(\underline{X}-K\underline{Y})\}$$

$$= E\{(\underline{X}-K\underline{Y})^*(\underline{X}-K\underline{Y})\} + \underbrace{\underline{b}^*\underline{b}}_{\geq 0} - \underbrace{E(\underline{X}-K\underline{Y})^*}_{0}\underline{b} - \underline{b}^*\underbrace{E(\underline{X}-K\underline{Y})}_{0}$$

$$\geq E\{(\underline{X}-K\underline{Y})^*(\underline{X}-K\underline{Y})\}$$

equality if and only if $\underline{b} = 0$

We need to find optimal K such that

$E\{(\underline{X}-K\underline{Y})^*(\underline{X}-K\underline{Y})\}$ is minimum

$$\begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_L \end{bmatrix} = K\underline{Y} \qquad K = \begin{bmatrix} \underline{k}_1^* \\ \underline{k}_2^* \\ \vdots \\ \underline{k}_L^* \end{bmatrix}$$

$\underline{k}_i^* \rightarrow i^{th}$ row of K.

$$\hat{X}_i = \underline{k}_i^* \underline{Y}$$

$$\tilde{X}_i = X_i - \hat{X}_i = X_i - \underline{k}_i^* \underline{Y}$$

We want to minimize $E\{ |\tilde{X}_1|^2 + |\tilde{X}_2|^2 + \dots + |\tilde{X}_L|^2 \}$

LMMSE Problem decouples as

Choose each $\underline{k}_i^*$ seperately such that

$E|X_i - \underline{k}_i^* \underline{Y}|^2$ is minimum.

So, Each row of K can be optimized separately

x ———————— x

18/8 Normal Equation

Restate Problem:

X is a random Variable (scaler) $\to$ Zero mean

$\underline{Y}_{M\times 1}$ is a random vector $\to$ Zero mean

Want estimator $\hat{X} = \underline{a}^* \underline{Y} = a_1^* Y_1 + a_2^* Y_2 + \dots + a_M^* Y_M$

where $\underline{a}$ is $M \times 1$ complex vector (weights)

Find $\underline{a}$ such that $E\{ |X - \underline{a}^* \underline{Y}|^2 \}$

is minimum

We have $\underset{M\times M}{R_Y} = E\{\underline{Y}\,\underline{Y}^*\}$ → covariance matrix of $\underline{Y}$

$\underset{M\times 1}{R_{YX}} = E\{\underline{Y}X^*\}$ → cross covariance vector.

Theorem: Optimal weight vector $\underline{a}$ which minimizes $E\{|X - \underline{a}^*\underline{Y}|^2\}$ satisfies

$$\boxed{R_Y\,\underline{a} = R_{YX}}$$ → Normal Equation.

Proof:

$$\tilde{X} = X - \hat{X} = X - \underline{a}^*\underline{Y}$$

$$E\{|\tilde{X}|^2\} = E\{|X - \underline{a}^*\underline{Y}|^2\}$$

$$= E\{(X - \underline{a}^*\underline{Y})(X - \underline{a}^*\underline{Y})^*\}$$

$$= E\{XX^* + \underline{a}^*\underline{Y}\,\underline{Y}^*\underline{a} - \underline{a}^*\underline{Y}X^* - X\underline{Y}^*\underline{a}\}$$

$$= \sigma_X^2 + \underline{a}^* E\{\underline{Y}\,\underline{Y}^*\}\underline{a} - \underline{a}^* E\{\underline{Y}X^*\} - \underbrace{E\{X\underline{Y}^*\}}_{R_{XY}}\underline{a}$$

MSE

$$J(\underline{a}) = \sigma_X^2 + \underline{a}^* R_Y\,\underline{a} - \underline{a}^* R_{YX} - R_{YX}^*\,\underline{a}$$

cost as a
function of $\underline{a}$

$$J[\underline{a}] = \begin{bmatrix} 1 & \underline{a}^* \end{bmatrix} \begin{bmatrix} \sigma_x^2 & -R_{Yx}^* \\ -R_{Yx} & R_Y \end{bmatrix} \begin{bmatrix} 1 \\ \underline{a} \end{bmatrix}$$

Two approaches to minimize
$J(\underline{a})$ wrt $\underline{a}$

① complex Gradient Approach

② completion of Squares Approach

We will start with (2)

Consider $\begin{bmatrix} \sigma_x^2 & -R_{Yx}^* \\ -R_{Yx} & R_Y \end{bmatrix}$ is conjugate symmetric matrix

Let $V = \begin{bmatrix} A & B \\ B^* & C \end{bmatrix}$ with $A = A^*$, $C = C^*$

If C is invertible, then we can write

$$V = \begin{bmatrix} I & BC^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} I & 0 \\ C^{-1}B^* & I \end{bmatrix}$$

where $\Sigma = A - B C^{-1} B^*$

In our case, $C = R_Y \Rightarrow$ may not be invertible.

We go for more general factorization

$$V = \begin{bmatrix} I & -D^* \\ 0 & I \end{bmatrix} \begin{bmatrix} \Sigma & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} I & 0 \\ -D & I \end{bmatrix}$$

where $\Sigma = A + BD$ and

D is any matrix satisfying:

the equation $-D^* C = B$.

Map this to our situation

$$A = \sigma_X^2$$
$$C = R_Y$$
$$B = -R_{YX}^*$$

$$V = \begin{bmatrix} \sigma_X^2 & -R_{YX}^* \\ -R_{YX} & R_Y \end{bmatrix}$$

Take $\underline{d}$ such that

$$-\underline{d}^* R_Y = -R_{YX}^*$$

Take $(\,)^*$ on both sides

or $\boxed{R_Y \underline{d} = R_{YX}}$

$$V = \begin{bmatrix} 1 & -\underline{d}^* \\ 0 & I \end{bmatrix} \begin{bmatrix} \sigma_X^2 - R_{YX}^* \underline{d} & 0 \\ 0 & R_Y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \underline{d} & I \end{bmatrix}$$

$$J(\underline{a}) = \begin{bmatrix} 1 & \underline{a}^* \end{bmatrix} [V] \begin{bmatrix} 1 \\ \underline{a} \end{bmatrix}$$

MSE

$$= \begin{bmatrix} 1 & \underline{a}^* \end{bmatrix} [\quad] [\quad] [\quad] \begin{bmatrix} 1 \\ \underline{a} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \underline{a}^* - \underline{d}^* \end{bmatrix} \begin{bmatrix} \sigma_x^2 - R_{Yx}^* \underline{d} & 0 \\ 0 & R_Y \end{bmatrix} \begin{bmatrix} 1 \\ \underline{a} - \underline{d} \end{bmatrix}$$

$$J(\underline{a}) = \underbrace{\sigma_x^2 - R_{Yx} \underline{d}}_{\text{does not depend on } \underline{a}} + \underbrace{(\underline{a} - \underline{d})^* R_Y (\underline{a} - \underline{d})}_{\geq 0}$$

since R_Y is positive semi definite

$$J(\underline{a}) \geq \sigma_x^2 - R_{Yx} \underline{d}$$

equality happens if $\underline{a} = \underline{d}$.

So $\boxed{R_Y \underline{a} = R_{Yx}}$ $\rightarrow$ Normal Equation.

If R_Y is invertible,

optimal $\boxed{\underline{a} = R_Y^{-1} R_{Yx}}$

Even if R_Y is not invertible, Solution for normal equation always exist.

→ will prove this later

x ——— x

9/8 Alternative derivation of Normal Equations

$$J(\underline{a}) = E\{|\tilde{x}|^2\}$$

MSE cost

$$= \sigma_x^2 + \underline{a}^* R_Y \underline{a} - \underline{a} R_{Yx} - R_{Yx}^* \underline{a}$$

Scalar valued function with complex vector argument

Complex gradient approach is to differentiate $J(\underline{a})$ with respect to $\underline{a}$ and equate to zero

Review of Complex Gradients

Let $g(z)$ be a scalar valued function of a complex variable $z = x + jy$

Let $g(z) = u(x,y) + j\,v(x,y)$

u and v are real valued functions

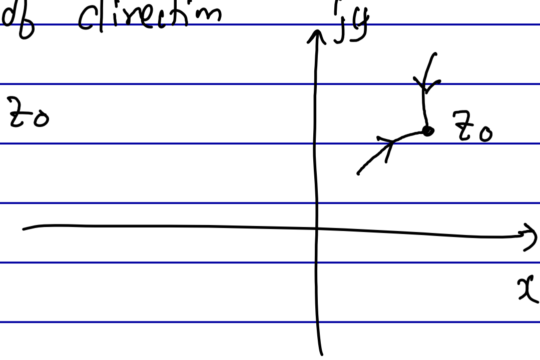
Derivative of $g(\cdot)$ at $z_0 = x_0 + j y_0$

is defined as $\quad \Delta z = \Delta x + j \Delta y$

$$\frac{dg}{dz} = \lim_{\Delta z \to 0} \frac{g(z_0 + \Delta z) - g(z_0)}{\Delta z}$$

For $g(\cdot)$ to be differentiable at z_0
(analytic at z_0)

If above limit exist and
be same irrespective of direction
from which z approaches z_0



Suppose $\Delta y = 0$ and $\Delta x \to 0$
(one particular direction)

$$\frac{dg}{dz} = \frac{du}{dx} + j \frac{dv}{\partial x} \longrightarrow (\$)$$

Suppose $\Delta x = 0$ and $\Delta y \to 0$

$$\frac{dg}{dz} = \frac{\partial v}{\partial y} - j \frac{\partial u}{\partial y} \longrightarrow (\$\$)$$

If $g(\cdot)$ is differentiable, $(\$)$ & $(\$\$)$ should
be same

So $\boxed{\frac{du}{dx} = \frac{\partial v}{\partial y}}$ and $\boxed{\frac{\partial v}{\partial x} = -\frac{\partial u}{\partial y}}$

These are called Cauchy - Riemann Conditions.

They are necessary and sufficient for $g(\cdot)$ to be analytic at z_0 if partial derivatives of $u(\cdot,\cdot)$ & $v(\cdot,\cdot)$ are continuous.

Note that
$$\frac{\partial g}{\partial x} = \frac{\partial u}{\partial x} + j\frac{\partial v}{\partial x}$$

$$\frac{\partial g}{\partial y} = \frac{\partial u}{\partial y} + j\frac{\partial v}{\partial y}$$

Definition
$$\frac{dg}{dz} = \frac{1}{2}\left\{\frac{\partial g}{\partial x} - j\frac{\partial g}{\partial y}\right\}$$

$\Rightarrow$ Average of ($) & ($$)

If g is a function of both z and z^*

we define
$$\frac{\partial g}{\partial z} = \frac{1}{2}\left\{\frac{\partial g}{\partial x} - j\frac{\partial g}{\partial y}\right\}$$

$$\frac{\partial g}{\partial z^*} = \frac{1}{2}\left\{\frac{\partial g}{\partial x} + j\frac{\partial g}{\partial y}\right\}$$

If $g(\cdot)$ is a function of z and analytic (differentiable)

then from Cauchy Riemann conditions it follows that $\frac{\partial g}{\partial z^*} = 0$

Examples:

(1) $g(z) = z = x + jy$

$$\frac{\partial g}{\partial z} = \frac{1}{2}\left\{ \frac{\partial g}{\partial x} - j\frac{\partial g}{\partial y} \right\}$$

$$= \frac{1}{2}\left\{ 1 - j(j) \right\} = 2/2 = 1$$

$$\frac{\partial g}{\partial z^*} = \frac{1}{2}\left\{ 1 + j^2 \right\} = 0$$

(2) $g(z) = z^2 = (x+jy)^2$

$$\frac{\partial g}{\partial z} = \frac{1}{2}\left\{ 2x + j2y - j\,(2x + j2y)\,j \right\}$$

$$= 2z$$

$$\frac{\partial g}{\partial z^*} = 0.$$

(3) $g(z) = |z|^2 = (z)(z^*)$

$$\frac{\partial g}{\partial z} = z^*$$
$$\frac{\partial g}{\partial z^*} = z$$

Verify yourselves.

(4) $g(z) = \lambda + \alpha z + \beta z^* + \gamma |z|^2$

$$\frac{\partial g}{\partial z} = \alpha + \gamma z^*$$
$$\frac{\partial g}{\partial z^*} = \beta + \gamma z$$

Verify

Now let $g(\underline{z})$ be a scalar valued function of complex vector $\underline{z}$

$$\underline{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \qquad z_i = x_i + j y_i$$

Complex gradient of $g(\cdot)$ wrt $\underline{z}$ is defined as

$$\nabla_{\underline{z}}(g) = \left[\frac{\partial g}{\partial z_1} \quad \frac{\partial g}{\partial z_2} \quad \cdots \quad \frac{\partial g}{\partial z_n} \right]$$

$\Downarrow$ $1 \times n$ row vector.

Complex gradient wrt $\underline{z}^*$

$$\nabla_{\underline{z}^*}(g) = \begin{bmatrix} \frac{\partial g}{\partial z_1^*} \\ \frac{\partial g}{\partial z_2^*} \\ \vdots \\ \frac{\partial g}{\partial z_n^*} \end{bmatrix} \rightarrow n \times 1 \text{ column vector}$$

Examples: ① $g(\underline{z}) = \underline{\alpha}^* \underline{z}$ $\underline{\alpha} \rightarrow$ constant vector.

$$\left.\begin{aligned} \nabla_{\underline{z}}(g) &= \underline{\alpha}^* \\ \nabla_{\underline{z}^*}(g) &= \underline{0} \end{aligned}\right\} \text{verify}$$

② $g(\underline{z}) = \underline{z}^* \underline{\beta}$

$$\nabla_{\underline{z}}(g) = \underline{0}$$

$$\nabla_{\underline{z}^*}(g) = \underline{\beta}$$

③ $g(\underline{z}) = \underline{z}^* \underline{z}$

$$\nabla_{\underline{z}}(g) = \underline{z}^* \qquad \nabla_{\underline{z}^*}(g) = \underline{z}$$

(4) $$g(\underline{z}) = \underline{\lambda} + \underline{\alpha}^* \underline{z} + \underline{z}^* \underline{\beta} + \underline{z}^* A \underline{z}$$

A is $n \times n$ matrix

$$\nabla_{\underline{z}}(g) = \underline{\alpha}^* + \underline{z}^* A$$

$$\nabla_{\underline{z}^*}(g) = \underline{\beta} + A\underline{z}$$

Back to our MSE expression

$$J(\underline{a}) = \sigma_x^2 + \underline{a}^* R_Y \underline{a} - \underline{a}^* R_{YX} - R_{YX}^* \underline{a}$$

Take gradient of $J(\cdot)$ wrt $\underline{a}$

$$0 = \nabla_{\underline{a}}(J) = \underline{a}^* R_Y - R_{YX}^*$$

$$\Rightarrow \boxed{R_Y \underline{a} = R_{YX}} \rightarrow \text{Normal Equations}$$

20/8

Suppose we use optimal LMMSE estimator

$$\hat{x} = \underline{a}^* \underline{Y}$$

where $\boxed{R_Y \underline{a} = R_{YX}}$

$$\text{M.M.SE} = E\left\{ |x - \hat{x}|^2 \right\}$$

$$= E\left\{ (x - \hat{x})(x - \hat{x})^* \right\}$$

$$= E\left\{ (x - \underline{a}^\dagger \underline{Y})(x - \underline{a}^\dagger \underline{Y})^* \right\}$$

$$= E\left\{ x x^* + \underline{a}^\dagger \underline{Y}\,\underline{Y}^\dagger \underline{a} - x \underline{Y}^\dagger \underline{a} - \underline{a}^\dagger Y x^* \right\}$$

$$= E|x|^2 + \underline{a}^\dagger R_Y \underline{a} - \overbrace{R_{Yx}^*}^{\underline{a}^\dagger R_Y} \underline{a} - \underline{a}^\dagger \underbrace{R_{Yx}}_{R_Y \underline{a}}$$

$$= \sigma_x^2 + \cancel{\underline{a}^\dagger R_Y \underline{a}} - \cancel{\underline{a}^\dagger R_Y \underline{a}} - \underline{a}^\dagger R_Y \underline{a}$$

$$\text{M.M.SE} = \sigma_x^2 - \underline{a}_{opt}^* R_Y \underline{a}_{opt}$$

$\underline{a}_{opt}$ is optimal LMMSE estimator

Suppose R_Y is invertible.

$$\underline{a}_{opt} = R_Y^{-1} R_{Yx} \qquad \underline{a}_{opt}^* = R_{Yx}^\dagger R_Y^{-1}$$

$$\text{MMSE} = \sigma_x^2 - R_{Yx}^* R_Y^{-1} R_Y R_Y^{-1} R_{Yx}$$

$$\boxed{\text{MMSE} = \sigma_x^2 - R_{Yx}^* R_Y^{-1} R_{Yx}}$$

Examples

① BPSK $Y = X + V$

X is ± 1 equally likely.

V is zero mean, real, variance σ_v^2

X and V are uncorrelated.

Find LMMSE $\hat{X}$ given Y.

Note: $E(Y) = 0$

$$R_Y = Var(Y) = \sigma_Y^2 = E\left((X+V)^2\right)$$

$$= E(X^2) + E(V^2)$$

$$= 1 + \sigma_V^2$$

$$R_{YX} = E(YX) = \sigma_{XY} = E\left((X+V)X\right) = E(X^2)$$

$$= 1$$

$$\sigma_Y^2 a = \sigma_{XY} \Rightarrow a = \frac{\sigma_{XY}}{\sigma_Y^2}$$

$$\boxed{\hat{X}_{LMMSE} = \frac{\sigma_{XY}}{\sigma_Y^2} Y = \frac{1}{1+\sigma_V^2} Y}$$

$$MMSE = \sigma_X^2 - \sigma_{XY} \cdot \frac{1}{\sigma_Y^2} \cdot \sigma_{XY}$$

$$= 1 - 1 \cdot \frac{1}{1+\sigma_v^2} \cdot 1$$

$$= 1 - \frac{1}{1+\sigma_v^2} = \frac{\sigma_v^2}{1+\sigma_v^2}$$

$x \longrightarrow$

(2) (Real) Jointly Gaussian X and Y

with zero mean, Variances σ_X^2, σ_Y^2

and cross covariance σ_{XY}

$$\left.\begin{aligned} \hat{X}_{LMMSE} &= \frac{\sigma_{XY}}{\sigma_Y^2} Y \\ \hat{X}_{MMSE} &= \frac{\sigma_{XY}}{\sigma_Y^2} Y \end{aligned}\right\}$$

for jointly Gaussian random variables MMSE & LMMSE are identical.

(3)

$$Y_1 = X + V_1$$

$$Y_2 = X + V_2$$

X, V_1, V_2 are uncorrelated.

zero mean

$Var(X) = \sigma_X^2$

$Var(V_1) = \sigma_1^2$

$Var(V_2) = \sigma_2^2$

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

Note $E(\underline{Y}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$$R_Y = E\{\underline{Y}\,\underline{Y}^*\} = \begin{bmatrix} E\{Y_1 Y_1^*\} & E\{Y_1 Y_2^*\} \\ E\{Y_2 Y_1^*\} & E\{Y_2 Y_2^*\} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_x^2 + \sigma_1^2 & \sigma_x^2 \\ \sigma_x^2 & \sigma_x^2 + \sigma_2^2 \end{bmatrix}$$

$$R_{Yx} = E\{\underline{Y} x^*\} = \begin{bmatrix} E(Y_1 x^*) \\ E(Y_2 x^*) \end{bmatrix} = \begin{bmatrix} \sigma_x^2 \\ \sigma_x^2 \end{bmatrix}$$

If $\sigma_1^2 \neq 0$ & $\sigma_2^2 \neq 0$ then R_Y^{-1} exist

$$\underline{a}_{opt} = R_Y^{-1} R_{Yx}$$

$$R_Y^{-1} = \frac{1}{\det R_Y} \begin{bmatrix} \sigma_x^2 + \sigma_2^2 & -\sigma_x^2 \\ -\sigma_x^2 & \sigma_x^2 + \sigma_1^2 \end{bmatrix}$$

$$\det R_Y = \sigma_x^2 (\sigma_1^2 + \sigma_2^2) + \sigma_1^2 \sigma_2^2$$

$$\underline{a}_{opt} = \frac{\sigma_x^2}{\det R_Y} \begin{bmatrix} \sigma_2^2 \\ \sigma_1^2 \end{bmatrix}$$

$$= \underbrace{\frac{\sigma_x^2 \sigma_1^2 \sigma_2^2}{\det R_Y}}_{\alpha} \begin{bmatrix} \frac{1}{\sigma_1^2} \\ 1/\sigma_2^2 \end{bmatrix}$$

$$\hat{X}_{LMMSE} = \alpha . \left[\frac{1}{\sigma_1^2} \quad \frac{1}{\sigma_2^2} \right] \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix}$$

$$\hat{X}_{LMMSE} = \alpha . \left(\frac{Y_1}{\sigma_1^2} + \frac{Y_2}{\sigma_2^2} \right)$$

Find M.M.S.E. yourself.

x ———————————— x

Orthogonality Principle (LMMSE)

$$\hat{X} = \underline{a}_{opt} \underline{Y}$$

where

$$R_Y \underline{a}_{opt} = R_{YX}$$

or

$$\underline{a}_{opt}^{*} R_Y = R_{YX}^{*} = R_{XY}$$

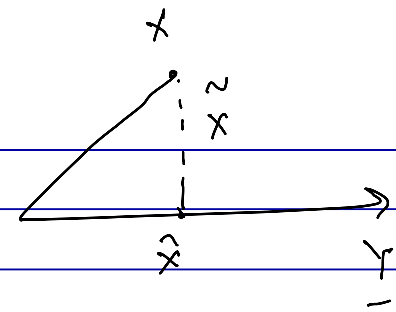
$$\underline{a}_{opt}^{*} E\{ \underline{Y} \underline{Y}^{*} \} = E\{ X \underline{Y}^{*} \}$$

$$E\{ \underbrace{\underline{a}_{opt}^{*} \underline{Y}}_{\hat{X}} \underline{Y}^{*} \} = E\{ X \underline{Y}^{*} \}$$

$$E(\hat{X} \underline{Y}^{*}) = E(X \underline{Y}^{*})$$

$$\Rightarrow \quad E(\underbrace{(X - \hat{X})}_{\tilde{X}} \underline{Y}^{*}) = 0 .$$

Orthogonality Principle. $\boxed{\tilde{X} \perp \underline{Y}}$



Error

$\tilde{X}$ is orthogonal to any linear function of $\underline{Y}$.

ie $\tilde{X} \perp B\underline{Y}$ where B is any complex matrix

$$E\{\tilde{X} \cdot (B\underline{Y})^*\} = E\{\tilde{X}\,\underline{Y}^*\,B^*\}$$

$$= \underbrace{E\{\tilde{X}\,\underline{Y}^*\}}_{\underline{0}}\,B^*$$

$$= \underline{0}$$

So $\tilde{X} \perp B\underline{Y}$

As a special case $\tilde{X} \perp \hat{X}$.

Since $\hat{X}$ is linear function of $\underline{Y}$

22/8 Existence of Solution for Normal Equations

X is random Variable,

$\underline{Y}$ is random vector of size $M \times 1$

$R_Y = E\{\underline{Y}\,\underline{Y}^{*}\} \rightarrow M \times M$ matrix

$R_{YX} = E\{\underline{Y} X^{*}\} \rightarrow M \times 1$ vector.

Normal Equation $\underset{M \times M}{R_Y} \; \underset{M \times 1}{\underline{a}} = \underset{M \times 1}{R_{YX}}$

Given R_Y & R_{YX}, need to find weight vector $\underline{a}$.

Normal equations is a linear system of equations.

M equations in M unknowns (variables)

<u>Theorem</u>: Normal equations <u>always</u> have a solution.

(a) We have <u>unique</u> solution iff R_Y is invertible.

(2) We have <u>multiple</u> solutions if R_Y is not invertible.

<u>Proof</u>: If R_Y is invertible then

$$\underline{a} = R_Y^{-1} R_{YX}$$

It will be a unique solution since R_Y is a one-one / onto linear transformation

$$R_Y \underline{a}_1 \neq R_Y \underline{a}_2 \quad \text{if } \underline{a}_1 \neq \underline{a}_2$$

If R_Y is not invertible:

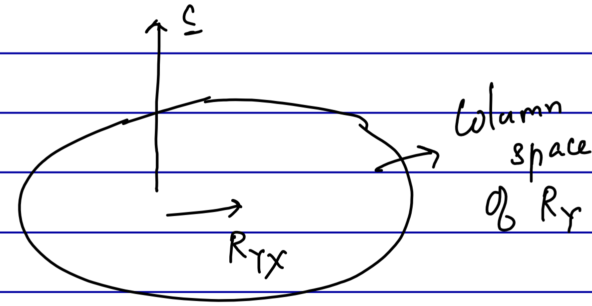
$$R_Y \underline{a} = R_{YX}$$ will

have a solution if R_{YX} lies in

column space of R_Y (space spanned by columns of R_Y)

$\Updownarrow$

Equivalently,



any vector $\underline{c}$ which is orthogonal to all columns of R_Y, is also orthogonal to R_{YX}.

$\underline{u}, \underline{v}$ be $M\times 1$ complex vectors

Inner product $\langle \underline{u}, \underline{v} \rangle = \underline{u}^* \underline{v}$

$\underline{u} \perp \underline{v}$ if $\underline{u}^* \underline{v} = 0$

Since $\underline{c}$ is orthogonal to all columns of R_Y,

$$\Rightarrow \underline{c}^* R_Y = \underline{0}$$

Need to show $\underline{c}^* R_{YX} = 0$

$$\underline{c}^* R_Y = \underline{0} \Rightarrow \underline{c}^* R_Y \underline{c} = 0$$

$$\Rightarrow \underline{c}^* E\{\underline{Y}\,\underline{Y}^*\} \underline{c} = 0$$

$$\Rightarrow E\{|\underline{c}^* \underline{Y}|^2\} = 0 \Rightarrow \underline{c}^* \underline{Y} = 0$$

Now $\underline{c}^* R_{YX} = \underline{c}^* E\{\underline{Y} X^*\}$

$$= E\{\underbrace{\underline{c}^* \underline{Y}}_{0} X^*\} = E\{0 X^*\} = 0$$

$\Rightarrow$ R_{YX} lies in column space of R_Y

$\Rightarrow$ solution always exist for Normal Equations

If R_Y is not invertible, you have multiple solutions.

Non-zero Mean Random Variables.

X is rand. var. with mean $E(X)$

$\underline{Y}$ is rand. vector with mean vector $E(\underline{Y})$

Want $\hat{X} = \underline{a}^* \underline{Y} + \underline{b}$ such that

$E\{|X-\hat{X}|^2\}$ is minimum

Solution: Define new random variables (with zero mean)

$$X_0 = X - E(X)$$

$$\underline{Y}_0 = \underline{Y} - E(\underline{Y})$$

Given $\underline{Y}_0$, optimal estimator for X_0 is

$$\hat{X}_0 = \underline{a}_0^* \underline{Y}_0 \quad \text{where} \quad R_{Y_0} \underline{a}_0 = R_{Y_0 X_0}$$

Since X and X_0 differ only by a constant (mean) we have

$$\hat{X} = \hat{X}_0 + E(X)$$

$$= \underline{a}_0^* \underline{Y}_0 + E(X)$$

$$= \underline{a}_0^* (\underline{Y} - E(\underline{Y})) + E(X)$$

Now,

$$R_Y = E\left\{ \underbrace{(\underline{Y} - E(\underline{Y}))}_{\underline{Y}_0} \underbrace{(\underline{Y} - E(\underline{Y}))^*}_{\underline{Y}_0^*} \right\}$$

$$= R_{Y_0}$$

Also $R_{Y_0 X_0} = R_{YX}$ (verify)

So

$$\hat{X} = \underline{a}^* (\underline{Y} - E(\underline{Y})) + E(X)$$

where $R_Y \underline{a} = R_{YX}$

For the rest of course, all random variables have zero mean unless stated otherwise.

Recall Vector Estimation Problem

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_L \end{bmatrix} \quad L \times 1 \text{ vector}$$

$$\underline{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_M \end{bmatrix} \quad M \times 1$$

$$\begin{bmatrix} \hat{X}_1 \\ \vdots \\ \hat{X}_L \end{bmatrix} = \hat{X} = K^* \underline{Y} \qquad K^* = \begin{bmatrix} \underline{k}_1^* \\ \underline{k}_2^* \\ \vdots \\ \underline{k}_L^* \end{bmatrix}$$

$$\hat{X}_i = \underline{k}_i^* \underline{Y}$$

$\underline{k}_i^*$ is ith row of K^*

We know $R_Y \underline{k}_i = E\{\underline{Y} X_i^*\} = R_{YX}^{(i)}$

Now $\underbrace{R_Y [\underline{k}_1 \ \underline{k}_2 \ \cdots \ \underline{k}_L]}_{K} = \underbrace{[R_{YX}^{(1)} \ R_{YX}^{(2)} \ \cdots \ R_{YX}^{(L)}]}$

$$R_{YX} = E\{\underline{Y}\,\underline{X}^*\}$$

$$\boxed{R_Y K = R_{YX}}$$

$M \times M \quad M \times L \quad M \times L$

Error Covariance Matrix

$$\underline{\hat{X}} = K^{*} \underline{Y}$$

Error vector $\underline{\tilde{X}} = \underline{X} - \underline{\hat{X}}$

Error Covariance $R_{\tilde{X}} = E\left\{ \underline{\tilde{X}}\, \underline{\tilde{X}}^{*} \right\}$

$M \times M$

$$= E\left\{ (\underline{X} - K^{*}\underline{Y})(\underline{X} - K^{*}\underline{Y})^{*} \right\}$$

$$= R_X + K^{*} R_Y K$$

$$- \underbrace{R_{XY} K}_{= R_{YX}^{*} = K^{*} R_Y} - K^{*} \underbrace{R_{YX}}_{R_Y K}.$$

for optimal LMMSE estimator

$$\boxed{R_{\tilde{X}} = R_X - K^{*} R_Y K}$$

$$= R_X - R_{YX}^{*} R_Y^{-1} R_{YX}$$

if R_Y is invertible

25/8

Application in Linear Models

$$\underline{Y} = H\underline{X} + \underline{V}$$

$\underline{X}_{L\times 1}$ random vector

$\underline{Y}_{M\times 1}$ random vector

$\underline{V}_{M\times 1}$ noise vector

} Zero mean

$H_{M\times L} \rightarrow$ known matrix

$\underline{X}$ and $\underline{V}$ are assumed to be uncorrelated

R_X be covariance of $\underline{X}$

R_V be covariance of $\underline{V}$

LMMSE estimation of $\underline{X}$ given $\underline{Y}$

$$\hat{\underline{X}}_{LMMSE} = K^* \underline{Y} \qquad \text{where}$$

$$R_Y K = R_{YX}$$

$$R_Y = E\{\underline{Y}\underline{Y}^*\} = E\{(H\underline{X}+\underline{V})(H\underline{X}+\underline{V})^*\}$$

$$= H R_X H^* + R_V$$

$$R_{YX} = E\{\underline{Y}\underline{X}^*\} = H R_X$$

If R_X and R_V are invertible

R_Y is also invertible

$$K = R_Y^{-1} R_{YX}$$

$$\hat{\underline{X}}_{LMMSE} = K^* \underline{Y}$$

(1) — $\hat{\underline{X}}_{LMMSE} = R_X H^* (H R_X H^* + R_V)^{-1} \underline{Y}$

Matrix inversion lemma (if A & C are invertible)

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B\left(C^{-1} + DA^{-1}B\right)^{-1} DA^{-1}$$

Applying this formula for

$$\left(H R_X H^* + R_V\right)^{-1} = R_V^{-1} - R_V^{-1} H \left(R_X^{-1} + H^* R_V^{-1} H\right)^{-1} H^* R_V^{-1}$$

(1) can also be written as

(2) $\longrightarrow$ $\hat{\underline{X}}_{LMMSE} = \left(R_X^{-1} + H^* R_V^{-1} H\right)^{-1} H^* R_V^{-1} \underline{Y}$

(1) & (2) are equivalent forms

Error Covariance Matrix

$$R_{\tilde{X}} = E\left\{ \tilde{\underline{X}} \tilde{\underline{X}}^* \right\}$$

$$= R_X - K^* R_Y K$$

$$= R_X - R_X H^* (R_Y + H R_X H^*)^{-1} H R_X$$

Matrix inversion lemma $\rightarrow$ $$= (R_X^{-1} + H^* R_Y^{-1} H)^{-1}$$

x ———————————— x

Kalman Filtering

- Solves linear MMSE estimation problem for a specific type of linear models

Properties about linear MMSE estimation

(1) Suppose $\underline{X} = A \underline{X}_1 + B \underline{X}_2$

$\underline{X}_1$ & $\underline{X}_2$ are random vectors

A and B are complex matrices

Let $\underline{Y}$ be observed vector

Let $\hat{\underline{X}}_1|\underline{Y}$ $\rightarrow$ LMMSE estimate of $\underline{X}_1$ given $\underline{Y}$

$\hat{\underline{X}}_2|\underline{Y}$ $\rightarrow$ LMMSE estimate of $\underline{X}_2$ given $\underline{Y}$

Thm: $\hat{\underline{X}}|\underline{Y} = A \hat{\underline{X}}_1|\underline{Y} + B \hat{\underline{X}}_2|\underline{Y}$

Proof:

$$\underline{\hat{X}}_1 = \left(R_{YX_1}\right)^* R_Y^{-1} \underline{Y}$$

$$\underline{\hat{X}}_1 = R_{X_1Y} R_Y^{-1} \underline{Y}$$

Similarly $$\underline{\hat{X}}_2 = R_{X_2Y} R_Y^{-1} \underline{Y}$$

So $$\underline{\hat{X}} = R_{XY} R_Y^{-1} \underline{Y}$$

Now $$R_{XY} = E\{\underline{X}\,\underline{Y}^*\}$$

$$= E\left\{\left(A\underline{X}_1 + B\underline{X}_2\right)\underline{Y}^*\right\}$$

$$= A\,E\{\underline{X}_1 \underline{Y}^*\} + B\,E\{\underline{X}_2 \underline{Y}^*\}$$

$$= A R_{X_1Y} + B R_{X_2Y}$$

So $$\underline{\hat{X}} = \left(A R_{X_1Y} + B R_{X_2Y}\right) R_Y^{-1} \underline{Y}$$

$$= A R_{X_1Y} R_Y^{-1} \underline{Y} + B R_{X_2Y} R_Y^{-1} \underline{Y}$$

$$= A \underline{\hat{X}}_1 + B \underline{\hat{X}}_2$$

x ——— x

Property ②

$\underline{X}$ is random vector

let $\underline{Z} = A\,\underline{Y}$

and A is invertible matrix

Theorem
$$\hat{\underline{X}}_{LMSSE}|\underline{Y} = \hat{\underline{X}}_{LMMSE}|\underline{Z}$$

Proof:
$$\hat{\underline{X}}_{LMMSE}|\underline{Y} = R_{XY}\,R_Y^{-1}\,\underline{Y}$$

$$\hat{\underline{X}}_{LMMSE}|\underline{Z} = R_{XZ}\,R_Z^{-1}\,\underline{Z}$$

$$R_Z = E\{\underline{Z}\,\underline{Z}^*\} = A\,R_Y\,A^*$$

$$R_{XZ} = E\{\underline{X}(A\underline{Y})^*\} = R_{XY}\,A^*$$

$$\hat{X}_{LMMSE}|\underline{Z} = R_{XZ}\,R_Z^{-1}\,\underline{Z}$$

$$= R_{XY}\,A^*\,\underbrace{(A\,R_Y\,A^*)^{-1}}_{(A^*)^{-1}\,R_Y^{-1}\,A^{-1}}\,A\,\underline{Y}$$

$$= R_{XY}\,R_Y^{-1}\,\underline{Y} = \hat{\underline{X}}_{LMMSE}|\underline{Y}$$

Property ③ : $\underline{Y}_1$, $\underline{Y}_2$ are two observed vectors

$\underline{Y}_1$ & $\underline{Y}_2$ are uncorrelated

Let $\hat{\underline{X}}|\underline{Y}_1$ be LMMSE estimator of $\underline{X}$ given $\underline{Y}_1$

$\hat{\underline{X}}|\underline{Y}_2$ be " " of $\underline{X}$ given $\underline{Y}_2$

Given both $\underline{Y}_1$ & $\underline{Y}_2$ what is the LMMSE estimator of X ?

26/8

Theorem $\hat{\underline{X}}|\underline{Y}_1, \underline{Y}_2 = \hat{\underline{X}}|\underline{Y}_1 + \hat{\underline{X}}|\underline{Y}_2$

Let $\underline{Y} = \begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \end{bmatrix}$

Note $\hat{\underline{X}}|\underline{Y}_1, \underline{Y}_2 = \hat{\underline{X}}|\underline{Y}$

$$\hat{\underline{X}}|\underline{Y} = R_{XY} R_Y^{-1} \underline{Y}$$

$$R_{XY} = E\{\underline{X}\,\underline{Y}^*\} = E\left\{\underline{X}\,[\underline{Y}_1^* \;\; \underline{Y}_2^*]\right\}$$

$$= [E\{\underline{X}\,\underline{Y}_1^*\} \;\; E\{\underline{X}\,\underline{Y}_2^*\}]$$

$$= [R_{XY_1} \;\; R_{XY_2}]$$

$$R_Y = E\left\{ \begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \end{bmatrix} \begin{bmatrix} \underline{Y}_1^* & \underline{Y}_2^* \end{bmatrix} \right\}$$

$$= \begin{bmatrix} R_{Y_1} & \overbrace{E\{\underline{Y}_1 \underline{Y}_2^*\}}^{0} \\ \underbrace{E\{\underline{Y}_2 \underline{Y}_1^*\}}_{0} & R_{Y_2} \end{bmatrix}$$

$$= \begin{bmatrix} R_{Y_1} & 0 \\ 0 & R_{Y_2} \end{bmatrix}$$

$$R_Y^{-1} = \begin{bmatrix} R_{Y_1}^{-1} & 0 \\ 0 & R_{Y_2}^{-1} \end{bmatrix}$$

$$\hat{\underline{X}}|_Y = R_{XY} R_Y^{-1} \underline{Y}$$

$$= \begin{bmatrix} R_{XY_1} & R_{XY_2} \end{bmatrix} \begin{bmatrix} R_{Y_1}^{-1} & 0 \\ 0 & R_{Y_2}^{-1} \end{bmatrix} \begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \end{bmatrix}$$

$$= R_{XY_1} R_{Y_1}^{-1} \underline{Y}_1 + R_{XY_2} R_{Y_2}^{-1} \underline{Y}_2$$

$$= \hat{\underline{X}}|_{Y_1} + \hat{\underline{X}}|_{Y_2}$$

Summary

(P1) If $\underline{X} = A\underline{X}_1 + B\underline{X}_2$ then

$$\hat{\underline{X}}_{LMMSE} = A\hat{\underline{X}}_{1\,LMMSE} + B\hat{\underline{X}}_{2\,LMMSE}$$

(P2) If $\underline{Z} = A\underline{Y}$ with invertible A

then $\hat{\underline{X}}\big|_{\underline{Y}} = \hat{\underline{X}}\big|_{\underline{Z}}$

(P3) If $\underline{Y}_1$ & $\underline{Y}_2$ are uncorrelated

then $\hat{\underline{X}}\big|_{\underline{Y}_1,\underline{Y}_2} = \hat{\underline{X}}\big|_{\underline{Y}_1} + \hat{\underline{X}}\big|_{\underline{Y}_2}$

$\downarrow$

LMMSE

x ———————— x

Innovation Process

We have observation vectors $\{\underline{Y}_1, \underline{Y}_2, \ldots, \underline{Y}_N\}$ (may be correlated)

From $\{\underline{Y}_1, \underline{Y}_2, \ldots, \underline{Y}_N\}$

$\Downarrow$ linear transformation (want invertibility)

Get $\{\underline{Z}_1, \underline{Z}_2, \ldots, \underline{Z}_N\}$ such that

Z_i's are uncorrelated.

Take $N = 2$

Here $\{\underline{Y}_1 \ \& \ \underline{Y}_2\}$

$$\begin{bmatrix} \underline{Z}_1 \\ \underline{Z}_2 \end{bmatrix} = A \begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \end{bmatrix}$$

Want invertible A

Want $\underline{Z}_1$ & $\underline{Z}_2$ to be uncorrelated.

- Gram-Schmidt Orthogonalization Procedure.

Take $\underline{Z}_1 = \underline{Y}_1$

$$\underline{Z}_2 = \underline{Y}_2 - \hat{\underline{Y}}_2 | \underline{Y}_1$$

lmmse estimate of $\underline{Y}_2$ given $\underline{Y}_1$

Innovation
(extra information present in $\underline{Y}_2$ which is not there in $\underline{Y}_1$.

$\underline{Z}_2$ is error corresponding to lmmse estimation of $\underline{Y}_2$ given $\underline{Y}_1$

From orthogonality principle,

$$\underset{\text{Error}}{\underline{Z}_2} \perp \underset{\text{observation}}{\underline{Y}_1} \Rightarrow \underline{Z}_2 \perp \underline{Z}_1$$

Note: $\underline{Z}_2 = \underline{Y}_2 - R_{Y_2 Y_1} R_{Y_1}^{-1} \underline{Y}_1$

$$\begin{bmatrix} \underline{Z}_1 \\ \underline{Z}_2 \end{bmatrix} = \underbrace{\begin{bmatrix} I & 0 \\ -R_{Y_2 Y_1} R_{Y_1}^{-1} & I \end{bmatrix}}_{A} \begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \end{bmatrix}$$

is lower triangular matrix

All diagonal entries are one.

$\det(A) = 1 \neq 0$

A is invertible

For general N

$$\underline{Z}_i = \underline{Y}_i - \hat{\underline{Y}}_i \mid \{\underline{Y}_1, \dots, \underline{Y}_{i-1}\}$$

lmmse estimate of $\underline{Y}_i$ given $\underline{Y}_1, \underline{Y}_2, \dots, \underline{Y}_{i-1}$

Kalman Filter

discrete time index k

State Process :

$k \geq 0$

$\underline{X}_k$ is $L \times 1$ vector

A_k is $L \times L$ matrix

D_k is $L \times L$ matrix

$\underline{W}_k$ is $L \times 1$

$$\underbrace{\underline{X}_{k+1}}_{\text{next state}} = A_k \underbrace{\underline{X}_k}_{\text{current state}} + D_k \underbrace{\underline{W}_k}_{\text{process noise}}$$

A_k, D_k are known matrices.

Measurement Process

$\underline{Y}_k$ is $M \times 1$ vector

C_k is $M \times L$ matrix

$\underline{V}_k$ is $M \times 1$

$$\underbrace{\underline{Y}_k}_{\text{observed vector at time } k.} = C_k \underline{X}_k + \underbrace{\underline{V}_k}_{\text{measurement noise}}$$

Problem Statement

(predictor)

Find optimal LMMSE estimation for

$\underline{X}_{k+1}$ given $\{\underline{Y}_1, \underline{Y}_2, \ldots, \underline{Y}_k\}$

Kalman filter solves the above problem using a recursive formulation

Assumptions:

* All random vectors are zero mean
* $\underline{W}_l$ and $\underline{W}_k$ are uncorrelated if $l \neq k$.
* $\underline{V}_l$ and $\underline{V}_k$ are uncorrelated if $l \neq k$

- $E\{\underline{W}_k \underline{W}_k^*\} = Q_k$
- $E\{\underline{V}_k \underline{V}_k^*\} = S_k$
- $E\{\underline{V}_k \underline{W}_l^*\} = 0$ for all l & k

(can be relaxed)

Initialialization

- $\underline{x}_0$ is zero mean with Covariance R_0
- $\underline{x}_0$ is uncorrelated $\underline{W}_k$ & $\underline{V}_k$ for all k.

Kalman Recursion Expressions

- Let $\hat{\underline{X}}_{k+1|k}$ denote LMMSE estimator (predictor) of $\underline{X}_{k+1}$ given all measurements upto time k, ie)

$$\{ \underline{Y}_1, \underline{Y}_2, \ldots, \underline{Y}_k \}$$

- $\tilde{\underline{X}}_{k+1|k} = \underline{X}_{k+1} - \hat{\underline{X}}_{k+1|k}$

  " prediction error

- Let P_{k+1} denote Covariance of prediction error

$$P_{k+1} = \text{Cov}(\tilde{\underline{X}}_{k+1|k})$$

Note $P_k = \text{Cov}(\tilde{\underline{X}}_{k|k-1})$

Theorem:

Previous prediction

$$\hat{\underline{X}}_{k+1|k} = A_k \hat{\underline{X}}_{k|k-1} + L_k(\underline{Y}_k - C_k \hat{\underline{X}}_{k|k-1})$$

where $L_k = A_k P_k C_k^* [C_k P_k C_k^* + S_k]^{-1}$

and P_k is computed recursively as

$$P_{k+1} = A_k P_k A_k^* + D_k Q_k D_k^* - A_k P_k C_k^* [C_k P_k C_k^* + S_k]^{-1} C_k P_k A_k^*$$

with initializations $\hat{\underline{x}}_{0|-1} = \underline{0}$

and $P_0 = Cov(\tilde{\underline{x}}_{0|-1}) = R_0$

Proof: Want to predict $\underline{x}_{k+1}$ given

$$\{\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_k\}$$

$$\|$$

We have $\{\underline{y}_1, \ldots, \underline{y}_{k-1}\} \cup \{\underline{y}_k\}$

Innovation in $\underline{y}_k$ is

$$\underline{z}_k = \underline{y}_k - \hat{\underline{y}}_{k|k-1}$$

$\Downarrow$ LMMSE estimate of $\underline{y}_k$ given $\{\underline{y}_1, \ldots, \underline{y}_{k-1}\}$

(P_1) property

$$\hat{\underline{y}}_{k|k-1} = C_k \hat{\underline{x}}_{k|k-1} + \underbrace{\hat{\underline{v}}_{k|k-1}}_{\underline{0}}$$

$\underline{0}$ since $\underline{v}_k$ is uncorrelated with $\{\underline{y}_1, \ldots, \underline{y}_{k-1}\}$

$$= C_k \hat{\underline{x}}_{k|k-1}$$

So innovation

$$\underline{z}_k = \underline{y}_k - C_k \hat{\underline{x}}_{k|k-1}$$

$$= C_k \underline{x}_k + \underline{v}_k - C_k \hat{\underline{x}}_{k|k-1}$$

$$= C_k \tilde{\underline{x}}_{k|k-1} + \underline{v}_k$$

and $\underline{Z}_k$ is uncorrelated with $\{\underline{Y}_1, \ldots, \underline{Y}_{k-1}\}$

(from orthogonality principle)

Now, using property (P_3)

$$(*) \quad \hat{\underline{X}}_{k+1|k} = \hat{\underline{X}}_{k+1|k-1} + \hat{\underline{X}}_{k+1|\underline{Z}_k}$$

$$\hat{\underline{X}}_{k+1|k-1} = A_k \cdot \hat{\underline{X}}_{k|k-1} + D_k \underbrace{\hat{\underline{W}}_{k|k-1}}_{\underline{0}\text{ since}}$$

$\underline{W}_k$ is uncorrelated with $\{\underline{Y}_1, \ldots, \underline{Y}_{k-1}\}$

$$(a) \quad \hat{\underline{X}}_{k+1|k-1} = A_k \hat{\underline{X}}_{k|k-1}$$

$$\hat{\underline{X}}_{k+1|\underline{Z}_k} = R_{X_{k+1} Z_k} R_{Z_k}^{-1} \underline{Z}_k$$

$$R_{X_{k+1} Z_k} = E\{\underline{X}_{k+1} \cdot \underline{Z}_k^*\}$$

$$= E\{\underline{X}_{k+1} [C_k \tilde{\underline{X}}_{k|k-1} + \underline{V}_k]^*\}$$

$$= E\{\underline{X}_{k+1} (\tilde{X}_{k|k-1}^* C_k^*)\}$$

$$= E\left\{ \left(A_k \underline{X}_k + D_k \underline{W}_k\right) \left[\underline{\tilde{X}}^*_{k|k-1} C_k^* \right] \right\}$$

$$= E\left\{ A_k X_k \tilde{X}^*_{k|k-1} C_k^* \right\}$$

$$= E\left\{ A_k \left(\underline{\hat{X}}_{k|k-1} + \tilde{X}_{k|k-1}\right) \underline{\tilde{X}}^*_{k|k-1} C_k^* \right\}$$

$$= E\left\{ A_k \underline{\tilde{X}}_{k|k-1} \underline{\tilde{X}}^*_{k|k-1} C_k^* \right\}$$

$$= A_k P_k C_k^*$$

$$R_{2k} = E\left\{ \underline{Z}_k \underline{Z}_k^* \right\} = Cov\left(\underline{Z}_k\right)$$

$$= Cov\left(\underline{Y}_k - \underline{\hat{Y}}_{k|k-1} \right)$$

$$= Cov\left(C_k \underline{\tilde{X}}_{k|k-1} + \underline{V}_k \right)$$

$\underline{V}_k$ is uncorrelated with $\underline{\tilde{X}}_{k|k-1}$

$$= Cov\left(C_k \underline{\tilde{X}}_{k|k-1}\right) + Cov\left(\underline{V}_k\right)$$

$$= C_k Cov\left(\underline{\tilde{X}}_{k|k-1}\right) C_k^* + S_k$$

$$R_{2k} = C_k P_k C_k^* + S_k$$

So $\quad \hat{\underline{X}}_{k+1}\Big|_{Z_k} = A_k P_k C_k^* \left[C_k P_k C_k^* + S_k \right]^{-1} \underline{Z}_k$

(b) —

Put (a) & (b) in (*)

$$\hat{\underline{X}}_{k+1|k} = A_k \hat{\underline{X}}_{k|k-1} + L_k \left[\underline{Y}_k - C_k \hat{\underline{X}}_{k|k-1} \right]$$

where $L_k = A_k P_k C_k^* \left[C_k P_k C_k^* + S_k \right]^{-1}$

We need a recursion to compute error covariance P_k.

Note $\quad P_k = \mathrm{Cov}\left(\tilde{X}_{k|k-1}\right)$; $\underline{P_{k+1}} = \mathrm{Cov}\left(\tilde{\underline{X}}_{k+1|k}\right)$

We will compute P_{k+1} in terms of P_k.

Note $\quad \underline{X}_{k+1} = \hat{\underline{X}}_{k+1|k} + \tilde{\underline{X}}_{k+1|k}$

$\hat{\underline{X}}$ & $\tilde{\underline{X}}$ are uncorrelated

(c) — $\quad \mathrm{Cov}\left(\underline{X}_{k+1}\right) = \mathrm{Cov}\left(\hat{\underline{X}}_{k+1|k}\right) + \underbrace{\mathrm{Cov}\left(\tilde{\underline{X}}_{k+1|k}\right)}_{P_{k+1}}$

We have

$$\underline{X}_{k+1} = A_k\left(\hat{\underline{X}}_{k|k-1} + \tilde{\underline{X}}_{k|k-1}\right) + D_k \underline{W}_k$$

$$Cov(\underline{X}_{k+1}) = Cov(A_k \hat{\underline{X}}_{k|k-1}) + Cov(A_k \tilde{X}_{k|k-1}) + cov.(D_k \underline{W}_k)$$

$$Cov(\underline{X}_{k+1}) = Cov(A_k \hat{\underline{X}}_{k|k-1}) + A_k P_k A_k^* + D_k Q_k D_k^*$$

(2)

Note: $\hat{\underline{X}}_{k+1|k} = A_k \hat{\underline{X}}_{k|k-1} + L_k \underline{Z}_k$

$$Cov(\hat{\underline{X}}_{k+1|k}) = A_k Cov(\hat{\underline{X}}_{k|k-1}) A_k^* + L_k Cov(\underline{Z}_k) L_k^*$$

(3)

Using (1), (2) & (3)

$$P_{k+1} = A_k P_k A_k^* + D_k Q_k D_k^* - A_k P_k C_k^* \left[C_k P_k C_k^* + S_k\right]^{-1} C_k P_k A_k$$