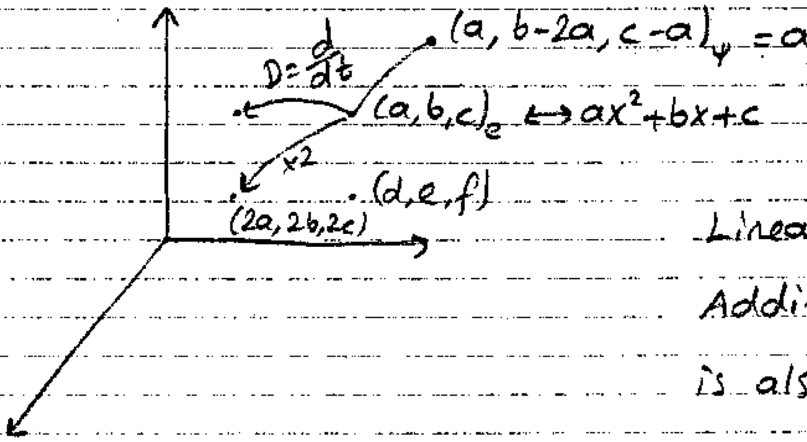
EE 503 2011/2012 Fall Osman Tayfun Biskin ①

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

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

Finite Energy

$$E = \int_{-\infty}^{\infty} |f(t)|^2 dt < \infty$$ Linear Space



$$(a, b-2a, c-a)_\psi = a\underbrace{(x+1)^2}_{\psi_2} + (b-2a)\underbrace{x}_{\psi_1} + (c-a)\underbrace{1}_{\psi_0}$$

Linear Space

Addition of two elements is also in Linear space.

$L\{.\}$: Linear operator in linear space.

$L\{.\} \leftrightarrow$ can be represented with matrices

$a(x+1)^2 + (b-2a)x + (c-a) \rightarrow 2a(x+1) + (b-2a) = (2a)x + (b)1$

$$\rightarrow ax^2 + bx + c \xrightarrow{D} 2ax + b$$

point of Linear space $\downarrow$

$$(a, b, c) \xrightarrow[\underline{\underline{m}}_D]{} (0, 2a, b)$$

$$\begin{bmatrix} 0 \\ 2a \\ b \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}}_{\underline{\underline{m}}_D \;=\, \text{matrix}} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \qquad \begin{bmatrix} \\ \\ \\ \end{bmatrix} = m_D \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

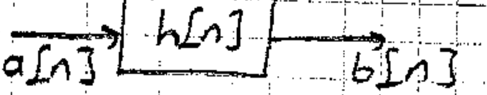
$D \leftrightarrow m_D^\psi$

$\leftrightarrow m_D^e$

Finite Energy Sequences

$$\sum_k |a_k|^2 < \infty$$

$$b[n] = a[n] * h[n]$$



$$b[n] = a[n] * h[n]$$
$$b[n] = \sum_{k=-\infty}^{\infty} a[k]\, h[n-k]$$

$$b[n] \longleftrightarrow B(e^{j\omega})$$
$$h[n] \longleftrightarrow H(e^{j\omega})$$
$$B(e^{j\omega}) = H(e^{j\omega})\, A(e^{j\omega})$$

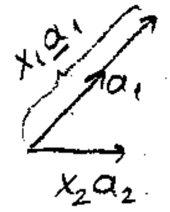
$\underline{\underline{M}}$ :

$\underline{X}$ : column vector $\begin{bmatrix} \vdots \end{bmatrix}_{N\times 1}$

$$\underline{y} = \underline{\underline{M}}\,\underline{X} \longrightarrow \underline{\underline{A}}\,\underline{X} = \underline{b}$$

$$\underline{\underline{A}} = [\, \overset{\nearrow 1^{st}\,\text{col.}}{\underline{a}_1} \;\; \underline{a}_2 \; --- \; \overset{\nearrow N^{th}\,\text{column}}{a_N} \,]$$

$$\underline{\underline{A}} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = x_1 \underline{a}_1 + x_2 \underline{a}_2 + \_\_\_\_ + x_N \underline{a}_N = \underline{b}$$



$$\left( \begin{array}{c} \underline{b} \in \text{Range}\{A\} \\ \Updownarrow \\ \text{column space}\{A\} \end{array} \right)$$

$\underline{\underline{A}}\,\underline{x} \overset{\downarrow}{=} \underline{b}$ has a solution

$$\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 10 \end{bmatrix}$$

$\underline{a_1}$ $\underline{a_2}$

$\begin{bmatrix} 5 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ infinite solution

$$\underbrace{\alpha_1 \underline{a_1} + \alpha_2 \underline{a_2} + \_\_\_\_ + \alpha_N a_N}= 0$$

$$\underline{X} = 5a_1 + 2a_1 + 0$$

$\int |f(t) + g(t)|^2 < M$

$|f(t)|^2 + |g(t)|^2$

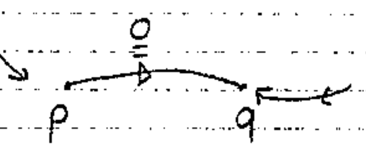
$\int |\alpha f(t)|^2$ linear space

$M = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 6 & 0 \\ 3 & 4 & 5 \end{bmatrix} \leftarrow P$ Lower triangular ile toplansa veya scalar ile çarpılırsa yine lower triangula olur. Bundan dolayı linear space'dir.

$f(t) \approx (a_0 + a_1 t + a_2 t^2)$ şeklinde yazılmaya çalışılıyor.

$\underline{P} = \alpha_0 e_0 + \alpha_1 e_1 + \_\_\_\_ + \alpha_N e_N$

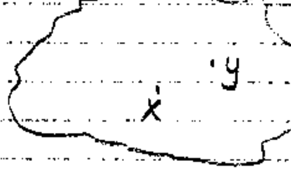
Linear space'de basislerden herhangi bir nokta yazılabilir.



$q = \beta_0 e_0 + \beta_1 e_1 + \_\_\_\_ + \beta_N e_N$

$$\underline{\underline{o}} \longrightarrow \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_N \end{bmatrix} = \underline{\underline{M}}e \begin{bmatrix} \alpha_0 \\ \vdots \\ \alpha_N \end{bmatrix}$$

Norm, Inner Product



x ve y'nin Linear space'de birbirlerine uzaklıkları

Norm: 1- $\|\underline{X}\| = 0 \longleftrightarrow \underline{X} = 0$

2- $\|X + y\| \leq \|X\| + \|y\|$

3- $\|\underset{\uparrow \text{ scalar}}{\underline{\alpha}} \underline{X}\| = |\alpha| \|\underline{X}\|$

Linear Space norm ile bulunur.

N-Dim: 1- Euclidean Norm: $\|x\| = \sqrt{\sum_{k=1}^{N} |x_k|^2} = \sqrt{\underline{x}^H \underline{x}}$

$$[\,-\,-\,-\,]_{N\times 1} \qquad \begin{bmatrix} | \\ | \end{bmatrix}_{N\times 1}$$

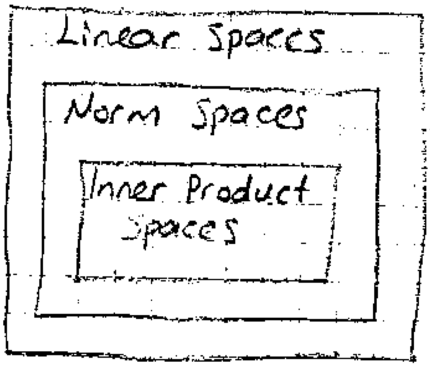
2- $\|x\|_p = \left(\sum_{k=1}^{N} |x_k|^p\right)^{1/p}$ ; $p \geq 1$

Inner Products: 1- $\langle x, x \rangle \geq 0$

$\langle \cdot , \cdot \rangle$ 2- $\langle x, y+z \rangle = \langle x, y \rangle + \langle x, z \rangle$

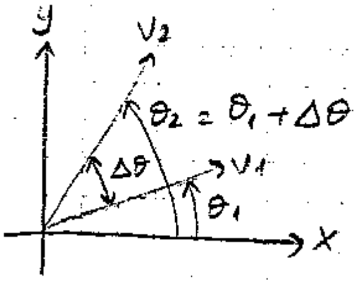
$( \cdot , \cdot )$ 3- $\langle \underline{x}, \alpha \underline{y} \rangle = \alpha \langle x, y \rangle$

Assume we operate in real field (no complex) multiplications or complex valued vectors.



If $\langle \cdot , \cdot \rangle$ is given $\|x\|^2 = \langle x, x \rangle$

2- Dim Space:



$$\langle \underline{v_1}, \underline{v_2} \rangle = x_1 x_2 + y_1 y_2 = v_1^T . v_2 \;\; (= v_2^T v_1)$$

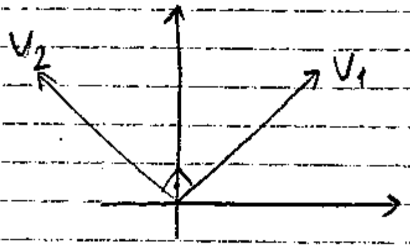
$$\underline{v_1} \rightarrow \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \qquad \underline{v_2} \rightarrow \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$$

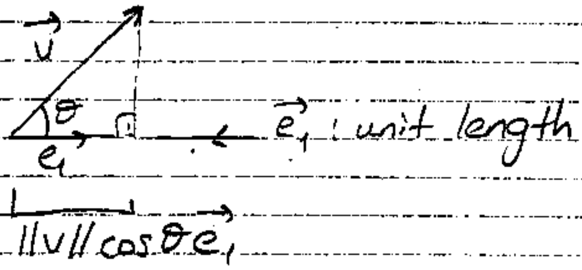
$$\langle v_1, v_2 \rangle = (\|v_1\| \cos\theta_1)(\|v_2\| \cos\theta_2) + (\|v_1\| \sin\theta_1)(\|v_2\| \sin\theta_2)$$

$$= \|v_1\| \, \|v_2\| \cos\underbrace{(\theta_1 - \theta_2)}_{\Delta\theta}$$

$$\cos(\theta_1 - \theta_2) = \frac{\langle v_1, v_2 \rangle}{\|v_1\| \, \|v_2\|}$$

$\langle v_1, v_2 \rangle = 0 \longrightarrow v_1 \perp v_2$ ($v_1$ is orthogonal to $v_2$)





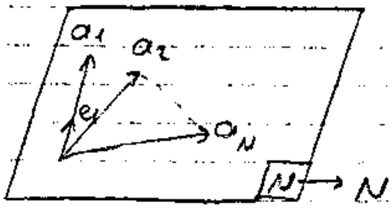
$\|v\| \cos\theta \, \vec{e}_1$

$\langle \vec{v}, \vec{e}_1 \rangle \vec{e}_1$

$\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle \triangleq E\{\underset{\sim}{x}\,\underset{\sim}{y}\} \longrightarrow E\{\underset{\sim}{x}, \underset{\sim}{y}\}$ $\underset{\sim}{x}$ is orth. to $\underset{\sim}{y}$

$\|\underset{\sim}{x}\|^2 \triangleq \sigma_x^2$ $E\{(x-\bar{x})(y-\bar{y})\} = 0$ $\underset{\sim}{x}$ is uncorrelated $\underset{\sim}{y}$

Gram - Schmidt Orthogonalization Procedure



Span$\{a_1, a_2, ___, a_N\}$ are linearly independent

Span$\{e_1, e_2, ___, e_N\}$. $e_k \perp e_l$, $k \neq l$, $\|e_k\| = 1$ } orthonormal

Gram - Schmidt: 1- $\hat{e}_1 = a_1$ 2- $\hat{e}_2 = a_2 - \langle e_1, a_2 \rangle e_1$

$$e_1 = \frac{\hat{e}_1}{\|e_1\|}$$

Check: $\langle e_1, \hat{e}_2 \rangle \overset{?}{=} 0 \rightarrow \langle e_1, a_2 - (\langle e_1, a_2 \rangle) e_1 \rangle$

$$= \langle e_1, a_2 \rangle - \langle e_1, a_2 \rangle \underbrace{\langle e_1, e_1 \rangle}_{1}$$

$$= 0$$

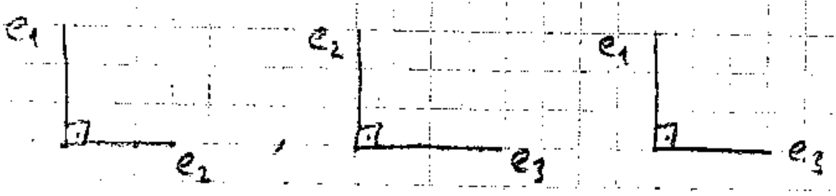
$$e_2 = \frac{\hat{e}_2}{\|\hat{e}_2\|}$$

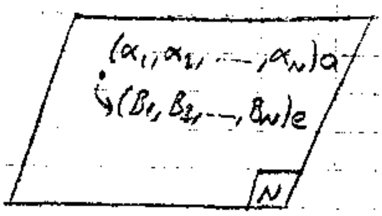
$$3-\ \hat{e}_3 = a_3 - \langle e_1, a_3\rangle e_1 - \langle e_2, a_3\rangle e_2$$

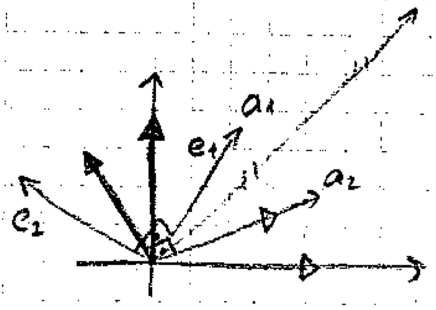
$$e_3 = \frac{\hat{e}_3}{\|\hat{e}_3\|}$$

At the end of the $3^{rd}$ s.t.p

$$\text{span}\{e_1, e_2, e_3\} = \text{span}\{a_1, a_2, a_3\}$$







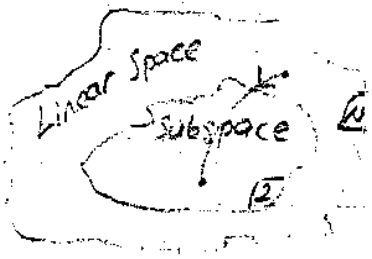
$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2a \\ 2b \end{bmatrix}$$

Projection Matrices

$S = \text{Span}\{a_1, a_2\}$

$\underline{x} \longrightarrow$ project $\underline{x}$ to $S$

P: Projection Matrix
① $\underline{\underline{P}}^2 = \underline{\underline{P}}$
② $\underline{\underline{P}}^T = \underline{\underline{P}}$
orthogonal Projectors.



$$\underline{X} \xrightarrow[\underline{\underline{P}}]{} \underline{\hat{X}} \in \text{Span}\{a_1, a_2\} = \text{Span}\{e_1, e_2\}$$

By Gram-Schmidt, I can find $e_1, e_2$ s.t. their span is same as span $\{a_1, a_2\}$

$$\underline{X} = \underbrace{\alpha_1 \underline{e_1} + \alpha_2 \underline{e_2}} + \_\_\_\_ + \alpha_N \underline{e_N}$$

$\underline{X} \in R^N$

$S: \text{Span}\{e_1, e_2\}$ $\quad \langle x, e_1\rangle = \alpha_1$ $\quad \langle x, e_2\rangle = \alpha_2$

$\hat{X} = \langle x, e_1\rangle e_1 + \langle x, e_2\rangle e_2$

$$= (\underline{e}_1^T \underline{x}) e_1 + (\underline{e}_2^T \underline{x}) e_2$$

⑦

$$= \underbrace{(\underline{e}_1 \underline{e}_1^T + \underline{e}_2 \underline{e}_2^T)}\ \underline{x}$$

$$\left[\ \right]\left[\quad\right]^{P}\ \left[\ \right]\left[\quad\right]$$

?

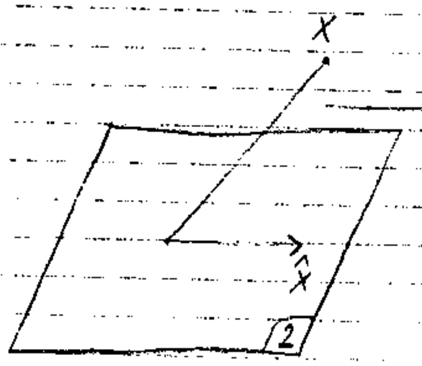
$$\underline{\underline{P}} = \underline{e}_1 \underline{e}_1^T + \underline{e}_2 \underline{e}_2^T \rightarrow ① \ \underline{\underline{P}}^2 \overset{?}{=} \underline{\underline{P}}$$

$$(e_1 e_1^T + e_2 e_2^T)(e_1 e_1^T + e_2 e_2^T)$$

$$= e_1 e_1^T + e_2 e_2^T$$

$$\underline{\underline{P}} = \sum_{k=1}^{M} e_k e_k^T$$

$$② \ P^T \overset{?}{=} P$$

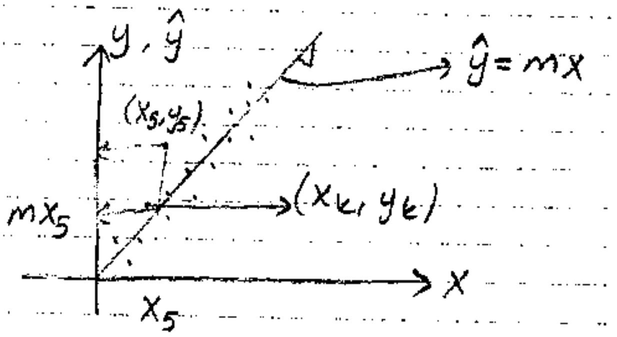


$$\rightarrow \epsilon = X - \hat{X}$$

Note $\epsilon \perp \underbrace{\text{span}\{e_1, e_2\}}_{S}$

Optimization:

? → [illegible]



$$\text{Cost}(m) = \sum_{k=1}^{N} (y_k - \hat{y}_k)^2$$

$$\searrow \hat{y}_k = m_k$$

Ex: $\text{Cost} = f(x,y) = x^2 + y^2 + 4xy + 2x + 5y + 1$

$$= [x \ \ y] \begin{array}{c|cc} & x & y \\ \hline x & \overset{\alpha}{1} & 2\beta \\ y & \gamma 2 & 1\theta \end{array} \begin{bmatrix} x \\ y \end{bmatrix} + [2 \ \ 5]_{2\times1} \begin{bmatrix} x \\ y \end{bmatrix} + 1$$

$$= \underline{X}^T \underline{A} X + \underline{b}^T \underline{x} + c$$

$$\nabla_x f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} \qquad 1 - \nabla_x (b^T x) = \begin{bmatrix} \ \\ \ \end{bmatrix} = b$$

$$2- \nabla_{\underline{x}} \left(\underline{x}^T \underline{\underline{A}}\, \underline{x}\right) = \left(\underline{\underline{A}} + \underline{\underline{A}}^T\right) \underline{x}$$

$$\nabla_x f(x,y) = \left(\underline{\underline{A}} + \underline{\underline{A}}^T\right) \underline{x} + \underline{b} = 0$$

$$= 2 \underline{\underline{A}}\, \underline{x} + \underline{b} = 0$$

$$\underline{2} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = - \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & -2 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2{,}5 \end{bmatrix} \quad \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -4/3 \\ 1/6 \end{bmatrix}$$

$$f(x,y) = x^2 + y^2 + 4xy + 2x + 5y + 1$$

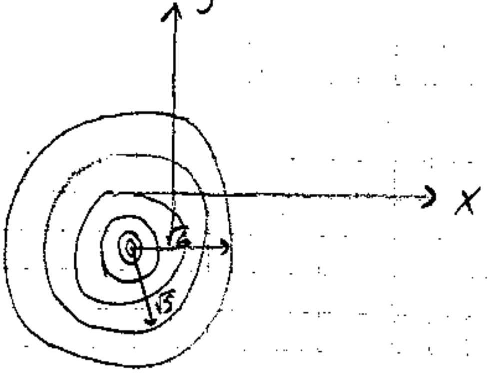
$$= \underline{x}^T \underline{\underline{A}}\, \underline{x} + \underline{b}^T \underline{x} + c$$

$$\nabla_{\underline{x}} f(x,y) = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \underline{0}$$

$$= \left(\underline{\underline{A}}^T + \underline{\underline{A}}\right) \underline{x} + \underline{b} = 0$$

$$\left(\underline{\underline{A}}^T + \underline{\underline{A}}\right) \begin{bmatrix} x \\ y \end{bmatrix} = -\underline{b}$$

$$\begin{bmatrix} 2 & 4 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = - \begin{bmatrix} 2 \\ 5 \end{bmatrix}$$



$$f(x,y) = c$$

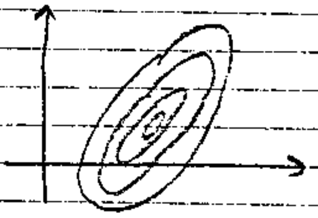
$$x^2 + y^2 + 4xy + 2x + 5y = 1$$

$$(x+1)^2 + \left(y + \frac{5}{2}\right)^2 + 4xy - \left(\frac{5}{2}\right) = c$$

$$(x+1)^2 + \left(y + \frac{5}{2}\right)^2 = c + \frac{25}{4} \Big\downarrow = 6$$

$$c = -1/4$$

$$c = -\frac{1}{4} - 1$$



Ex) $f(z_1, z_2) = \|z_1\|^2 + \|z_2\|^2 + Re\{z_1 z_2 - 1\}$

$\hookrightarrow z_1, z_2 \in \mathbb{C} \leftarrow$ Complex Numbers

$$f(\underbrace{z_{1R}, z_{1I}}_{\downarrow}, z_{2R}, z_{2I}) = z_{1R}^2 + z_{1I}^2 + z_{2R}^2 + z_{2I}^2 + (z_{1R} z_{2R} - z_{1I} z_{2I})$$

$z_1 = z_{1R} + j z_{1I}$

Complex variables
optimal chap. Hayes

$$f(z_1, z_2) = z_1 z_1^* + z_2 z_2^* + \frac{z_1 z_2 + z_1^* z_2^*}{2} - 1$$

$\frac{\partial f}{\partial z_1} = 0 \longrightarrow z_1^* + \frac{z_2}{2} = 0$ $\qquad$ $\frac{\partial f}{\partial z_2} = 0 \rightarrow z_2^* + \frac{z_1}{2} = 0$

$\updownarrow$ conjugate of each other

$\boxed{\frac{\partial f}{\partial z_1} = 0 \longrightarrow z_1 + \frac{z_2^*}{2} = 0}$ $\qquad$ $\frac{\partial f}{\partial z_2^*} = 0 \rightarrow z_2 + \frac{z_1^*}{2} = 0$

It can be satisfied $\rightarrow \frac{\partial f}{\partial z_1} = 0$

$f(z_1, z_2)$ is a complex variable function with real "output".

Then $\frac{\partial f}{\partial z_1^*} = 0$

$\frac{\partial f}{\partial z_2^*} = 0$ is sufficient to find its maxima/minima points;

$f(\underline{z}) = \underline{z}^H \underline{\underline{A}} \underline{z} + \underline{b}^H \underline{z} + c \qquad \underline{z}_{N\times 1} \in \mathbb{C}^N$

$$\nabla_{\underline{z}^*} f(z) = \begin{bmatrix} \frac{\partial f(z)}{\partial z_1^*} \\ \vdots \\ \frac{\partial f(z)}{\partial z_N^*} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

$\longrightarrow (A^H + A) z = -b$

## Least Squares Approximation

$\underline{\underline{A}}_{M\times N}\,\underline{x}_{N\times 1} = \underline{b}_{M\times 1} \longleftarrow$ Eq. system is incosistent (No solution)

$$\left[\, \underline{a}_1 \;\; \underline{a}_2 \; \cdots \; a_N \,\right]_{M\times N} \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ \\ b_n \end{bmatrix}$$

$(M > N)$

More eq. than unknowns

$$J(\underline{x}) = \| \underline{\underline{A}}\underline{x} - \underline{b} \|^2 = \overset{\boxed{x^H A^H - b^H}}{(\underline{\underline{A}}\underline{x} - \underline{b})^H} (\underline{A}\underline{x} - \underline{b}) = \underline{x}^H \underline{\underline{A}}^H \underline{\underline{A}}\underline{x} - \underline{x}^H \underline{\underline{A}}^H \underline{b}$$

$$- \underline{b}^H \underline{\underline{A}}\, x + b^H b$$

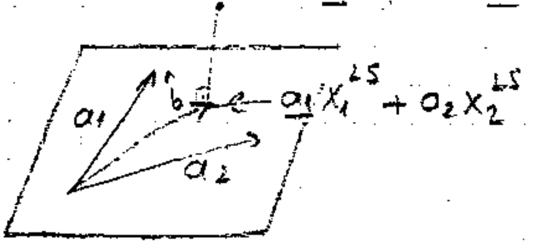
$$\nabla_{\underline{x}^*} J(x) = (A^H A)\underline{x} - \underline{\underline{A}}^H \underline{b} = \underline{0} \longrightarrow (\underline{\underline{A}}^H \underline{\underline{A}})\, \underline{x}_{LS} = \underline{\underline{A}}^H \underline{b}$$

$$\boxed{x_{LS} = (\underline{\underline{A}}^H \underline{\underline{A}})^{-1} \underline{\underline{A}}^H \underline{b}}$$

$\uparrow$

LS Solution

$\underline{b} = \underline{a}_1 x_1^{True} + \underline{a}_2 x_2^{True} + \text{noise}$



$$\underline{\underline{A}}\,\underline{x} = \underline{b} \longrightarrow (\underline{\underline{A}}^H \underline{\underline{A}})\, \underline{x}_{LS} = \underline{\underline{A}}^H \underline{b}$$

$$x_{LS} = (A^H A)^{-1} A^H b$$

$$\downarrow$$

$$\hat{\underline{b}} = \underline{\underline{A}}\, \underline{x}_{LS} = \underbrace{\underline{\underline{A}} (\underline{\underline{A}}^H \underline{\underline{A}})^{-1} \underline{\underline{A}}^H}\, \underline{b}$$

$$= \underline{P}\; \underline{b}$$

$$\underline{P} \overset{?}{=} \text{Projection Matrix} \rightarrow ① \ \underline{\underline{P}}^2 \overset{?}{=} \underline{\underline{P}} \rightarrow A(A^H A)^{-1} A^H A (A^H A)^{-1} A^H$$

$$= A(A^H A)^{-1} A^H$$

$$\underline{\underline{P}}^2 = \underline{\underline{P}}$$

$$② \ \underline{\underline{P}}^H = \underline{\underline{P}} \longrightarrow (A(A^H A)^{-1} A^H)^H = P$$

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

$f(0) = 10 \longrightarrow c = 10$

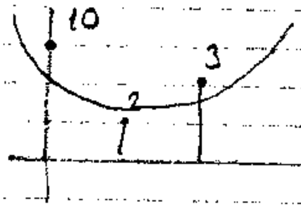
$f(1) = 2 \qquad a + b + c = 2$

$f(2) = 3 \longrightarrow 4a + 2b + c = 3$

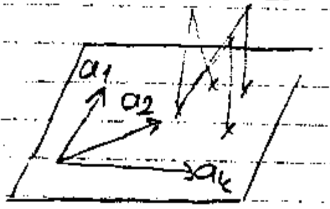
$f(3) = 40 \longrightarrow 9a + 3b + c = 40$

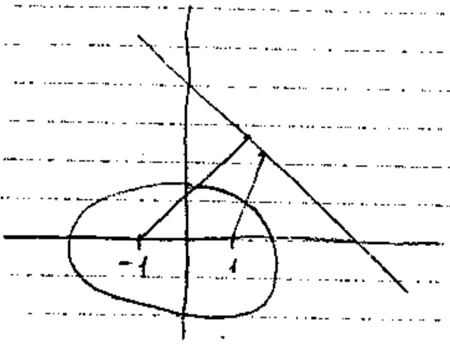
$$\underbrace{\begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{bmatrix}}_{\underline{\underline{A}}\,\underline{x} = \underline{b}} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 10 \\ 2 \\ 3 \\ 40 \end{bmatrix}$$

$$\begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{bmatrix} = (\underline{\underline{A}}^T \underline{\underline{A}})^{-1} \underline{\underline{A}}^T \underline{b}$$

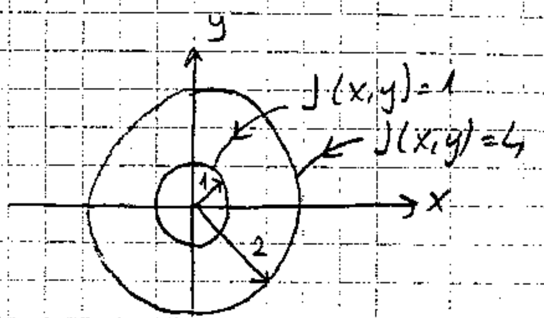


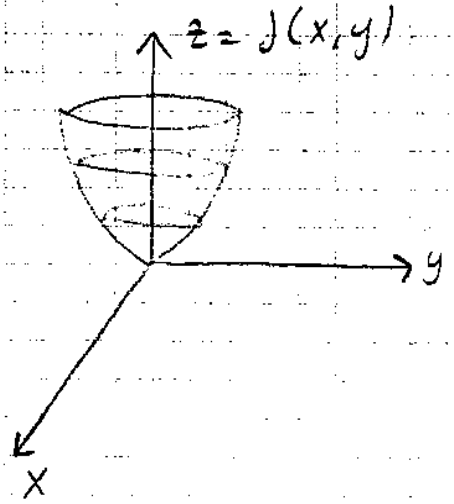
$c = d_1^2 + d_2^2 + d_3^2$



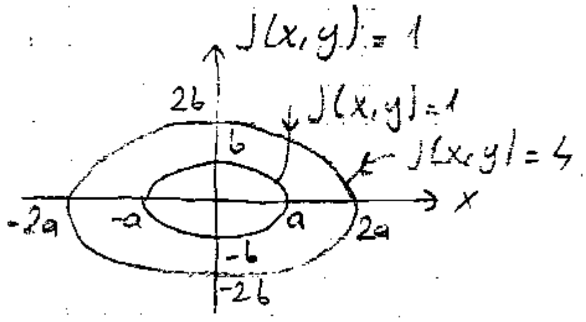


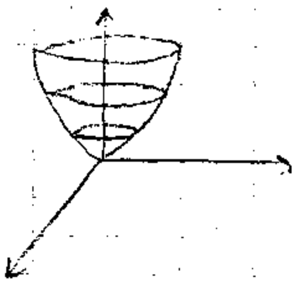
$J(x,y) = x^2 + y^2$





$J(x,y) = \frac{x^2}{a^2} + \frac{y^2}{b^2}$





J(x,y)'s in these examples are valid cost functions; since they have a unique minima.

$J(x,y) = x^T A x$

$J(x,y) \geq \; \forall\, x,y \rightarrow J(x,y)$ is a "proper" cost function

$$[x \;\; y]\left[\begin{array}{c|c} \cdot & \cdot \\ \hline \cdot & \cdot \end{array}\right]\begin{bmatrix} x \\ y \end{bmatrix} \geq 0$$

Def: $x^T A x > 0 \quad \forall\, \underline{x} \in R^N$

$\downarrow$

$\underline{\underline{A}}$ is positive definite, $\underline{\underline{A}} > 0$

$x^T \underline{\underline{A}} x \geq 0 \quad \forall x \in R^N$

$\downarrow$

$A$ is positive semi definite, $A \geq 0$

Result: $A \geq 0 \rightleftarrows$ eig. values of $A$ are non-negative

$\text{eig}(A) \geq 0$

(Since, det $e_k$ be an eigenvector

$$e_k^T \underbrace{\underline{\underline{A}}\, e_k}_{\lambda_k e_k} = \lambda e_k^T e_k = \lambda_k \quad (\|e_k\|^2 = 1)$$

$$\downarrow$$

$$\lambda_k \geq 0$$

Let's show that

$J(x) = \|\underline{\underline{A}}\,\underline{x} - \underline{b}\|^2$ is a "proper" cost function.

$J(x) = (\underline{\underline{A}}\,\underline{x} - \underline{b})^T (A x - b)$

$= \underline{x}^T \underline{\underline{A}}^T \underline{\underline{A}}\,\underline{x} - \underline{x}^T \underline{\underline{A}}\,\underline{b} - \underline{b}^T \underline{\underline{A}}\,\underline{x} + b^T b$

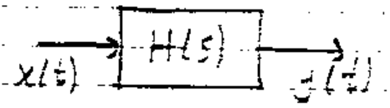
is $\underline{x}^T \underline{\underline{A}}^T \underline{\underline{A}}\,\underline{x} \geq 0 \qquad (Ax)^T (Ax) = \|(Ax)\|^2 \geq 0$

$\uparrow$

$(x^T A^T A x \geq 0)$

DSP Review

Linear System Theory

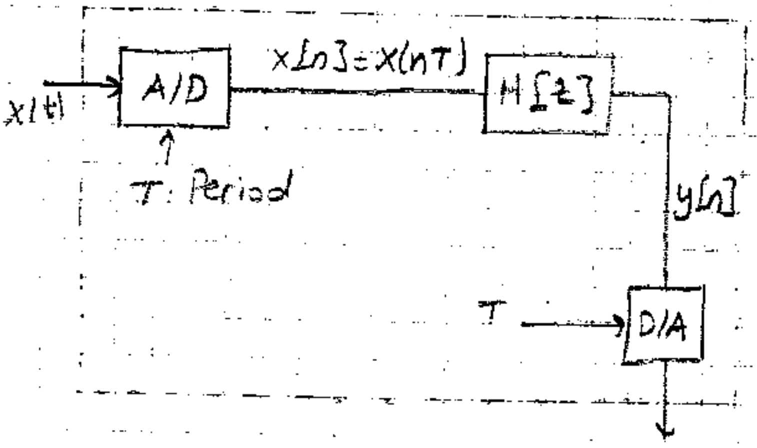


LTI system whose impulse resp. h(t) causal, h(t) is right-sided and H(s) has ROC extending to the right half plane; x(t) is the input y(t) is the output/ input/output are cont. variables.

$$H(s) = \frac{s+1}{s^2+3s+2} = \frac{Y(s)}{X(s)} \to Y(s)[s^2+3s+2] = (s+1)X(s)$$

$$\Big\downarrow \mathcal{L}^{-1}\{.\}$$

$$\boxed{(D^2+3D+2)\underset{\substack{\uparrow \\ \text{output}}}{y(t)} = (D+1)\underset{\substack{\uparrow \\ \text{input}}}{x(t)}}$$



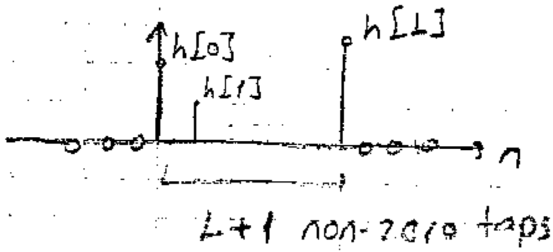
Sampling Theorem:

x(t) is bandlimited to $f_{max}$ Hertz.

then $x[n] = x(T_s n)$ ; $T_s < \frac{1}{2f_{max}}$

Then from $x[n] \xrightarrow[\text{(exactly)}]{\text{reconstruct}} x(t)$

$$x[n] * h[n] = y[n] \to y[n] = \sum_{k=-\infty}^{\infty} x[k]\,h[n-k]$$



$$= \sum_{k=-\infty}^{\infty} x[k]\, h[n-k]$$

$$= \sum_{k=n-L}^{n} x[k]\, h[n-k]$$

$$= \sum_{k'=L}^{0} x[n-k']\, h[k'] \qquad k' = n-k \;\; (k = n-k')$$

$$(a_1 + a_2 + \_\_\_ + a_L)$$

$$= \underbrace{\sum_{k=0}^{L} x[n-k]\, h[k]}_{y[n]}$$

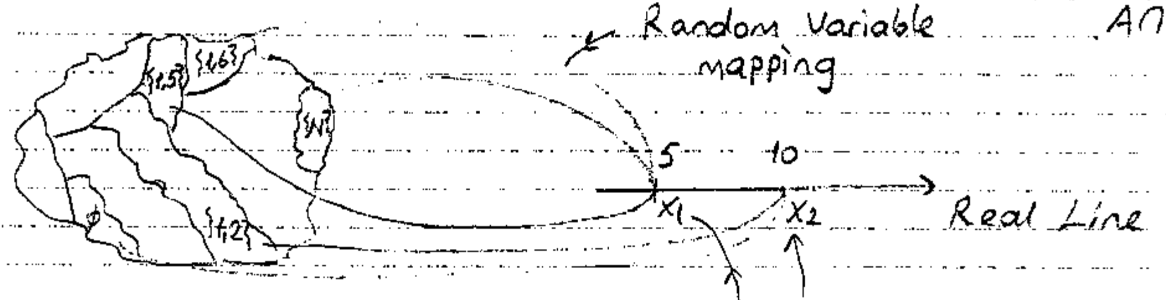
$$y[n] = h[0]\,x[n] + h[1]\,x[n-1] + h[2]\,x[n-2]$$

$$\begin{bmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \end{bmatrix} = \underbrace{\begin{bmatrix} h(0) & & & 0 \\ h(1) & h(0) & & \\ h(2) & h(1) & h(0) & \\ h(3) & h(2) & h(1) & h(0) \\ & & & \end{bmatrix}}_{\substack{\text{Convolution Matrix} \\ \text{(Toeplitz Matrix)}}} \begin{bmatrix} x(0) \\ x(1) \\ x(2) \\ x(3) \\ \end{bmatrix}$$

Probability Theory:

1- $P(A) \geqslant 0$  2- $P(\text{Universal set}) = 1$  3- $P(A \cup B) = P(A) + P(B)$

$A \cap B = \emptyset$



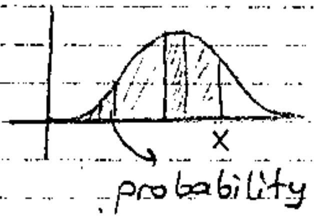
Sample Space = {1, 2, 3, ___, N}   Random Variables

$P\{x = 5\} = 0{,}2$

$P\{x = 10\} = 0{,}8$

$$\text{p.d.f} \longrightarrow \int_{-\infty}^{\infty} f_{\underset{\sim}{x}}(x)\, dx = 1$$

$$f_{\underset{\sim}{x}}(x) = \text{p.d.f}$$

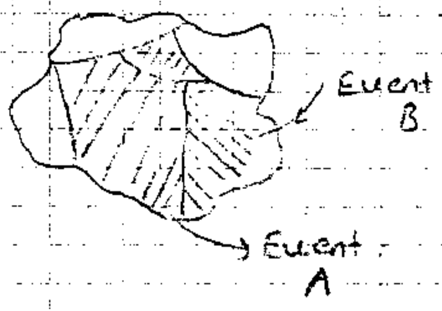


$$\text{cdf} \longrightarrow F_X(x) = \int_{-\infty}^{x} f_{\underset{\sim}{X}}(x)\,dx$$

$$P(\boxed{A|B}) = \xrightarrow{\qquad\qquad} P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}$$

$A|B$ : A given B

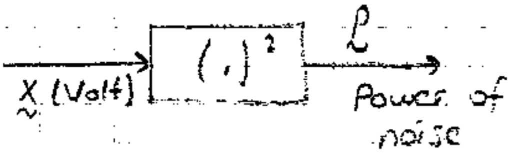
Baye's Theorem



Functions of Random Variables:

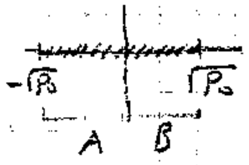
$\underset{\sim}{X}$ : r.v

$\underset{\sim}{y} = f(\underset{\sim}{X}) \longrightarrow$ p.d.f of $\underset{\sim}{y} = ?$



$$P\{\underset{\sim}{P} \leq P_0\} = P\{(\underset{\sim}{X})^2 \leq P_0\} = P\{|\underset{\sim}{X}|^2 \leq P_0\}$$

$$= P\{-\sqrt{P_0} \leq \underset{\sim}{X} \leq \sqrt{P_0}\}$$



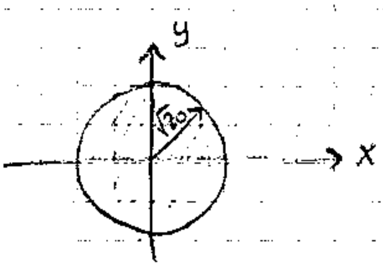
$$= P\{-\sqrt{P_0} \leq \underset{\sim}{X} \leq 0\} + P\{0 < x \leq \sqrt{P_0}\}$$

$$= F_X(0) - F_X(-\sqrt{P_0}) + \left(F_X(P_0) - F_X(0)\right)$$

$$F_P(P_0) = F_X(\sqrt{P_0}) - F_X(-\sqrt{P_0})$$

$\underset{\sim}{z} = \underset{\sim}{X}^2 + \underset{\sim}{y}^2$

$P(\underset{\sim}{z} \leq z_0)$

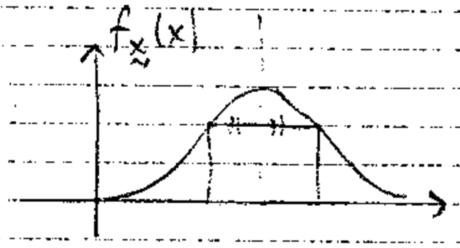


Expectation:

$$E\{g(x)\} = \int_{-\infty}^{\infty} g(x) f_{\underset{\sim}{x}}(x)\,dx = \lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} g(x_k)$$

$x_k$ : $k^{th}$ trial of the experiment
producing r.v. $\underset{\sim}{x}$

$$\bar{x} = E\{\underset{\sim}{x}\} = \int_{-\infty}^{\infty} x\, f_{\underset{\sim}{x}}(x)\,dx$$



$= E\{x^2\}$

$m_k = E\{x^k\}$ moments

$\bar{x}_1 \,\Big|\, \sigma_x^2 = E\{(x-\bar{x})^2\} = E\{x^2\} - (E\{x\})^2$

Moment Generating Function

$f(x) \longrightarrow$ Moment Gen Func.

$$\Phi(s)$$

$$\Phi(s) = E\{e^{sx}\} = \int_{-\infty}^{\infty} e^{sx} f_{\underset{\sim}{x}}(x)\,dx$$

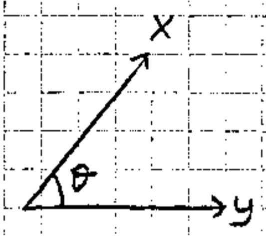
$$m_k = \frac{d^k}{ds^k}(\Phi(s))\Big|_{s=0} = \left(\frac{d^k}{ds^k}\{\overset{E\{e^{sx}\}}{\Phi(s)}\}\right)\Big|_{s=0} = \left(E\{\frac{d^k}{ds^k}e^{sx}\}\right)\Big|_{s=0}$$

$$= E\left\{\left(\frac{d}{ds^k}e^{sx}\right)\Big|_{s=0}\right\} = E\{(x^k e^{sx})\big|_{s=0}\} = E\{x^k\}$$

$$\Phi(s) \underset{\text{series at}}{\overset{\text{Taylor}}{=}} \sum_{k=0}^{\infty} \Phi^{(k)}(0)\frac{s^k}{k!} = \sum_{k=0}^{\infty} m_k \frac{s^k}{k!}$$

knowing all moments $E\{x^k\}$ can lead to p.d.f

# Correlation Coefficients



$$\underline{x}, \underline{y} \in R^N$$

$$\cos\theta = \frac{\underline{x}^T\underline{y}}{\|x\|\|y\|} = \frac{\sum_{k=1}^{N} x_k y_k}{\sqrt{\left(\sum_{k=1}^{N} x_k^2\right)\sum_{k=1}^{N} y_k^2}}$$

$$\theta = 0 \rightarrow \cos\theta = 1 \rightsquigarrow \underline{x} = \alpha\,\underline{y}$$

$\underset{\sim}{x}$, $\underset{\sim}{y}$ : random variables

Assume $E\{\underset{\sim}{x}\} = E\{\underset{\sim}{y}\} = 0$ (zero mean r.v)

$$\rho_{xy} = \frac{E\{x,y\}}{\sqrt{E\{x^2\}E\{y^2\}}} = \frac{\int_y\int_x xy\, f_{xy}(x,y)\,dx\,dy}{\sqrt{\left(\int_x x^2 f_x(x)\,dx\right)\left(\int_y y^2 f_y(y)\,dy\right)}}$$

↗ correlation coefficient

for $E\{x\} = E\{y\} = 0$

Properties:

1- $|\rho_{xy}| \leqslant 1$

2- $\rho_{xy} = 0 \rightsquigarrow \underset{\sim}{x}$ and $\underset{\sim}{y}$ are called uncorrelated

For zero mean r.v.'s

$\underset{\text{r.v}}{\text{uncorrelated}} \equiv \underset{\text{r.v.}}{\text{orthogonal}}$

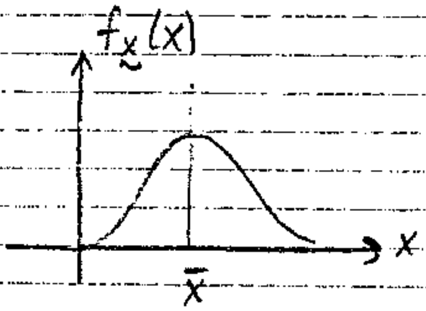
3- $\rho_{xy} = \mp 1 \rightarrow \underset{\sim}{x}$ and $\underset{\sim}{y}$ are "in the same direction", that is $\underset{\sim}{y} = a\underset{\sim}{x} + b$

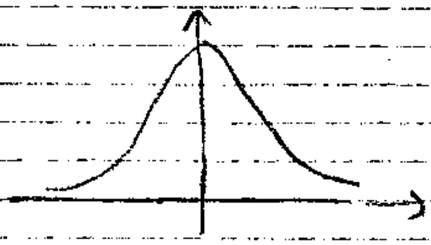
4- $\underset{\sim}{y} = a\underset{\sim}{x} + b \rightarrow |\rho_{xy}| = 1$

$$\rho_{xy} = \frac{E\{(x-\bar{x})(y-\bar{y})\}}{\sqrt{E\{(x-\bar{x})^2\}E\{(y-\bar{y})^2\}}}$$

; for non-zero mean r.v.'s general cor. coefficient definition

$\rho_{xy}$ does not depend on $\bar{x}$ or $\bar{y}$





$$\frac{\sum x_k y_k}{\sqrt{\sum (x_k)^2 \sum (y_k)^2}} \leqslant 1$$

$$\left(\sum x_k y_k\right)^2 \leqslant \left(\sum x_k^2\right)\left(\sum y_k^2\right)$$

Cauchy-Schwartz Ineq.

Ex: $\underset{\sim}{y} = a\underset{\sim}{x} + \underset{\sim}{n}$ (Model for $\underset{\sim}{y}$)

observation ↑ r.v ↖ noise

$\underset{\sim}{n}, \underset{\sim}{x}$ are ind.

$\rho_{xy} = ?$

$$\rho_{xy} = \frac{E\{\underset{\sim}{x}\underset{\sim}{y}\}}{\sqrt{E\{x^2\}E\{y^2\}}} \leftarrow$$

by assuming $\bar{x} = \bar{n} = 0$ without any loss generality

$$E\{xy\} = E\{\underset{\sim}{x}(a\underset{\sim}{x} + \underset{\sim}{n})\} = a\underbrace{E\{x^2\}}_{\sigma_x^2} + \underbrace{E\{xn\}}_{E\{x\}E\{n\}} = a\sigma_x^2$$

$$E\{x^2\} = \sigma_x^2$$

$$E\{y^2\} = E\{(ax+n)^2\} = a^2\sigma_x^2 + \sigma_n^2$$

$$\rho_{xy} = \frac{a\sigma_x^2}{\sqrt{\sigma_x^2(a^2\sigma_x^2 + \sigma_n^2}} = \frac{a\sigma_x^2}{\sqrt{\sigma_x^2 a^2 \sigma_x^2 \left(1 + \frac{\sigma_n^2}{a^2\sigma_x^2}\right)}}$$

$$= \frac{1}{\sqrt{1+\frac{\sigma_n^2}{a^2\sigma_x^2}}} = \frac{1}{\sqrt{1+\frac{1}{SNR}}} \; ; \; SNR = \frac{a^2\sigma_x^2}{\sigma_n^2}$$

SNR = Signal to Noise Ratio

$$(SNR)_{observation} = \frac{E\{(\overset{ax}{signal})^2\}}{E\{(\underset{n}{noise})^2\}} = \frac{a^2\sigma_x^2}{\sigma_n^2}$$

$\rho_{xy} \to 0$ when $SNR \to 0$

$\rho_{xy} \to 1$ when $SNR \to \infty$

$x_1, x_2, x_3$

$$E\left\{ x \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \right\} = E\left\{ \begin{bmatrix} xy_1 \\ xy_2 \\ xy_3 \end{bmatrix} \right\} = \begin{bmatrix} 0{,}75 \\ 0{,}1 \\ 0{,}001 \end{bmatrix}$$

$$\hat{x} = w_1 y_1 + w_2 y_2 + w_3 y_3$$

Review: $$\underset{\substack{\uparrow \\ \text{"}\cos\theta\text{"}}}{\rho_{xy}} = \frac{E\{(x-\bar{x})(y-\bar{y})\}}{\sigma_x \sigma_y}$$

$\uparrow$ standard dev. ($\sigma_x^2$: variance)

$\to \sqrt{E\{x^2\}}$?

$$= \frac{E\{xy\} - \bar{x}\bar{y}}{\sigma_x \sigma_y} = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

$$Cov(x,y) = E\{(x-\bar{x})(y-\bar{y})\} = E\{xy\} - \bar{x}\bar{y}$$

Then $x, y$ uncorrelated $Cov(x,y) = 0$ $E\{xy\} = E\{x\}E\{y\}$

$x, y$ independent $\longrightarrow f_{\underset{\sim}{xy}}(x,y) = f_{\underset{\sim}{x}}(x) \cdot f_{\underset{\sim}{y}}(y)$

$$\int_y \int_x xy f_{XY}(x,y)\,dxdy = \left(\int_x x f_X(x)dx\right)\left(\int_y y f_Y(y)dy\right)$$

$$E\{xy\} = E\{x\}E\{y\}$$

Ex) $x = \begin{cases} 1 & A \text{ occurs} \\ 0 & \text{other} \end{cases}$ $y = \begin{cases} 1 & B \text{ occur} \\ 0 & \text{other} \end{cases}$

$Cov(x,y) = ?$

$= E\{xy\} - E\{x\}E\{y\}$

$$E\{x\} = \sum_{k=0}^{1} k\, p(x=k) = p(x=1) = p(A \text{ occurs})$$

$$E\{y\} = P(B \text{ occurs})$$

$$E\{xy\} = \sum_{k=0}^{1}\sum_{l=0}^{1} kl\, p(x=k, y=l) = P(\underset{\text{together}}{A \text{ and } B \text{ occur}})$$

Case 1: $Cov(x,y) > 0$

$P(A \text{ and } B) - P(A)P(B) > 0$

$$\begin{tikzcd} & P(A \text{ and } B) > P(A)P(B) \arrow[dl] \arrow[dr] & \\ \frac{P(A \text{ and } B)}{P(B)} > P(A) & & \frac{P(A \text{ and } B)}{P(A)} > P(B) \end{tikzcd}$$

$P(A|B) > P(A)$ $P(B|A) > P(B)$

$Cov(x,y) > 0 \xrightarrow[\text{conclusion}]{}$ If A has occurred, B is more likely to occur! (and vice-versa)

(they occur together

Case 2: $Cov(x,y) < 0$

$$\begin{tikzcd} & \downarrow \arrow[dl] \arrow[dr] & \\ P(A|B) < P(A) & & P(B|A) < P(B) \end{tikzcd}$$

A has occurred → B is less likely to occur.

**Case3:** $Cov(x,y)=0$

$P(A \text{ and } B) = P(A)\,P(B)$ satisfied

Then it is possible that A and B are independent

Properties of $Cov(x,y)$

1- $Cov(X,X) = Var(X)$

2- $Cov(x,y) = Cov(y,x) \leftarrow E\{x.y\} - E\{x\}E\{y\}$

3- $Cov(cX,Y) = c\,Cov(X,Y)$

4- $Cov(X,Y+Z) = Cov(X,Y) + Cov(X,Z)$

$$\downarrow$$

$$E\{X(y+z)\} - \bar{x}(\bar{y}+\bar{z}) \nearrow$$

$$(E\{xy\} - \bar{x}\bar{y}) + (E\{xz\} - \bar{x}\bar{z})$$

Ex: $Var(\sum_{k=1}^{N} x_k) = ?$ $\quad Var(z) = E\{z^2\} - (\bar{z})^2$

$$1\text{-}\ Cov\left(\sum_{k=1}^{N} x_k, \sum_{k=1}^{N} x_l\right)$$

$$4\text{-}\ \sum_{l} Cov\left(\sum_{k=1}^{N} x_k, x_l\right)$$

$$2\text{-}\ \sum_{l} Cov\left(x_l, \sum_{k=1}^{N} x_k\right)$$

$$4\text{-}\ \sum_{k}\sum_{l} cov(x_l, x_k)$$

$$= \sum_{\substack{k \\ (k\neq l)}}\sum_{l} Cov(x_l, x_k) + \sum_{k=1}^{N} \underbrace{Cov(x_k, x_k)}_{var(x_k)}$$

$$= 2\sum_{\substack{k=0 \\ l>k}}^{N}\sum_{l=(k+1)}^{N} Cov(x_l, x_k) + \sum_{k=1}^{N} Var(x_k)$$

observation: If r.v. $x_k$'s are uncorrelated,

$$\mathrm{Cov}(x_k, x_\ell) = 0 \quad k \neq l$$

$$\mathrm{Var}\left(\sum x_k\right) = \sum_{k=1}^{N} \mathrm{Var}(x_k)$$

$$\downarrow$$

zero mean $\rightarrow$ [illegible]

$$E\{(x_1 + x_2 + \_\_ + x_N)^2\} = E\{x_1^2\} + \_\_ + E\{x_N^2\}$$

Random Vectors

$$\underline{x} = \begin{bmatrix} \underset{\sim}{x_1} \\ \underset{\sim}{x_2} \\ \vdots \\ \underset{\sim}{x_N} \end{bmatrix} \longrightarrow \underline{x}: \text{ defined through joint pdf of } \{x_1, \_\_, x_N\}$$

vector whose entries are random variables

$$E\{\underline{x}\} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_N \end{bmatrix} ;$$

$$\begin{bmatrix} \vdots \end{bmatrix} \times [\,-\,-\,]$$

Covariance Matrix: $\underline{\underline{C}}x = E\{(x - E(x))(x - E(x))^T\}$

$$\underline{\underline{C}}x = \begin{bmatrix} E\{(x_1-\bar{x}_1)(x_1-\bar{x}_1)\} & E\{(x_1-\bar{x}_1)(x_2-\bar{x}_2)\} & \cdots & \mathrm{Cov}(x_1,x_N) \\ \mathrm{Cov}(x_2,x_1) & \mathrm{Cov}(x_2,x_2) & & \mathrm{Cov}(x_2,x_N) \\ \vdots & & & \\ \mathrm{Cov}(x_N,x_1) & & & \mathrm{Cov}(x_N,x_N) \end{bmatrix}$$

entries are the covariance values of pairwise combination of random variables

$$\rho \quad \rho_{xy} = \frac{\mathrm{Cov}(x,y)}{\sigma_x \sigma_y} \quad \text{when } \sigma_x = \sigma_y = 1 \longrightarrow \rho_{x,y} = \mathrm{Cov}(x,y)$$

"$\cos\theta$" "$\cos\theta$"

## Properties of Cov. Matrix

1- $\underline{\underline{C}}x$ is symmetric

$$\underline{\underline{C}}x = E\{(x-\bar{x})(x-\bar{x})^T\}$$

$$\underline{\underline{C}}x^T \overset{?}{=} C_x$$

for complex valued r.v.'s.

$$C_x = E\{(x-\bar{x})(x-\bar{x})^H\}$$

$C_x$ = Hermitian matrix

2- diag( $\underline{\underline{C}}x$ ) ⇒ variance of random variable

3- $\underline{\underline{C}}x \geq 0 \quad (\underline{\alpha}^H \underline{\underline{C}}x \underline{\alpha} \geq 0 \ \forall \underline{\alpha})$ $\quad [..][:] : A$

$$\alpha^H \underline{\underline{C}}x\alpha = \alpha^H E\{(x-\bar{x})(x-\bar{x})^H\}\alpha = E\{\underbrace{\underline{\alpha}^H(x-\bar{x})}_{A}\underbrace{(x-\bar{x})^H\underline{\alpha}}_{A^*}\}$$

$$= E\{|A|^2\} \geq 0$$

$$\downarrow$$

$$\underline{\underline{C}}x \geq 0$$

$\underline{\underline{C}}x$ : Hermitian - Symmetric Matrix and positive semi-definite

→ eig. values are orthogonal set (Hermitian)

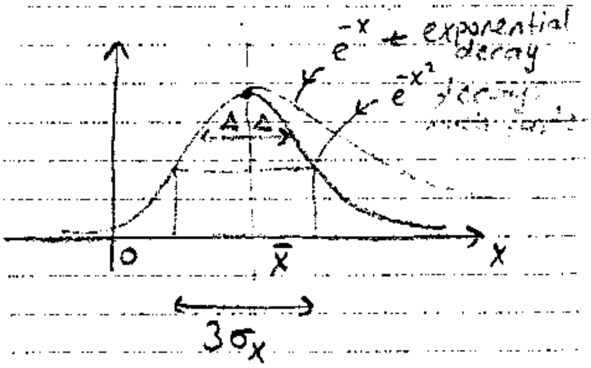
eig. values are real (Hermitian)

↓

eig. values are non-negative (P.S.D)

Gaussian Distribution

$$f_{\underset{\sim}{x}}(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\bar{x})^2}{2\sigma_x^2}}$$



## Central Limit Theorem:

$$\underset{\sim}{z}_N = \sum_{k=1}^{N} \underset{\nwarrow\text{random var}}{x_k}$$

1- $\underset{\sim}{x}_k$'s i.i.d (independent identically dist.)

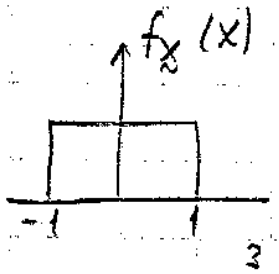
2- $\underset{\sim}{x}_k$'s have finite variance

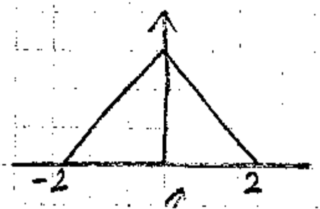
$\downarrow$

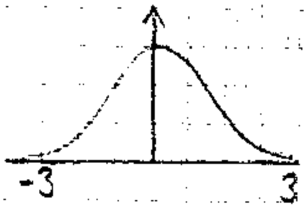
$$\overset{\text{c.d.f}\nearrow}{F_{z_N}}(z_N) \xrightarrow[N\to\infty]{}$$ c.d.f of Gaussian dist with mean $\to \sum_{k=1}^{N} \bar{x}_k$

var $\to \sum_{k=1}^{N} \sigma_{x_k}^2$

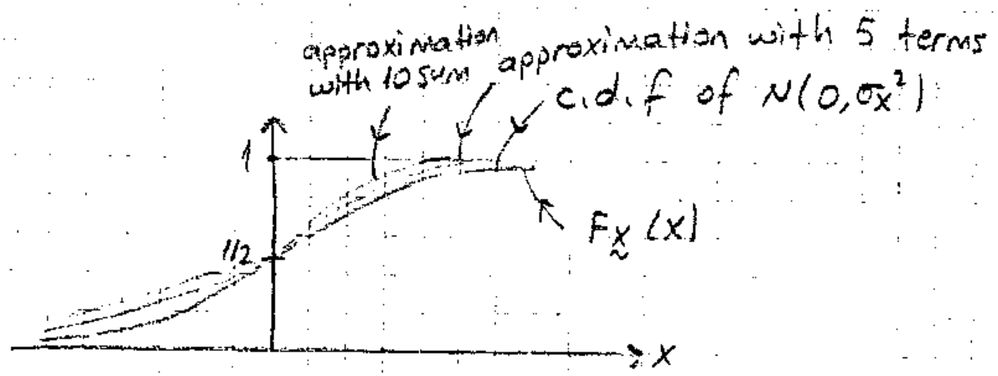
as $N \to \infty$







$$X_3 = \sum_{k=1}^{3} X_k = \overbrace{(X_1 + X_2)} + X_3$$



2-D Gaussian Dist.

$$N\text{-}D: \quad f_{\underline{x}}(x_1, x_2, \ldots, x_N) = \frac{1}{(2\pi)^{N/2}(\det(\underline{\underline{C}}_x))^{1/2}} e^{-\frac{1}{2}(x-\bar{x})^T \underline{\underline{C}}_x^{-1}(x-\bar{x})}$$

$$C_x \geq 0 \longleftrightarrow \underline{\underline{C}}_x^{-1} \geq 0$$

$$2\text{-D}:\quad \underline{\underline{C}}X = \begin{bmatrix} \sigma_{X_1}^2 & Cov(X_1,X_2 \\ \bigcirc & \sigma_{X_2}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{X_1}^2 & r_{xy}\sigma_x\sigma_y \\ r_{xy}\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

$$\underset{\substack{\uparrow \\ \text{correlation} \\ \text{coef.}}}{r_{xy}} = \frac{Cov(X_1,X_2)}{\sigma_x\sigma_y}$$

$$f_{\underset{\sim}{X}}(X_1,X_2) = \frac{1}{2\pi\,|\underline{\underline{C}}x|^{1/2}}\, e^{-\frac{1}{2}[X_1\ X_2]\begin{bmatrix} \sigma_{X_1}^2 & r_{xy}\sigma_x\sigma_y \\ r_{xy}\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}^{-1}\underset{\substack{\uparrow \\ \left(\begin{smallmatrix}\bar{X}_1=0 \\ \bar{X}_2=0\end{smallmatrix}\right)}}{\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}}}$$

Jointly Gaussianity: $X_1, X_2, \_\_, X_N$ are said to be jointly Gaussian if their joint pdf is in the form of N-Dim. Gaussian pdf.

Ex) $X_1$: Gaussian Dist $\rightarrow N(0,1)$ ← marginal of $f(X_1,X_2)$ is Gaussian

$X_2$: Gaussian Dist $\rightarrow N(0,5)$ ← $f_{X_1X_2}(X_1,X_2)$

Ex)

$X_1: N(\mu_1, \sigma_{X_1}^2)$

$X_2: N(\mu_2, \sigma_{X_2}^2)$

$X_1, X_2$ are independent.

$X_1, X_2$: Jointly Gaussian $\Rightarrow$ ?

$$f_{X_1X_2}(X_1,X_2) = f_{X_1}(X_1)f_{X_2}(X_2) = \left[\frac{1}{\sqrt{2\pi\sigma_{X_1}^2}}\, e^{-\frac{(X_1-\mu_1)^2}{2\sigma_{X_1}^2}}\right]\left[\frac{1}{\sqrt{2\pi\sigma_{X_2}^2}}\, e^{-\frac{(X_2-\mu_2)^2}{2\sigma_{X_2}^2}}\right.$$

$$= \frac{1}{2\pi\sigma_{X_1}\sigma_{X_2}}\, e^{-\frac{1}{2}\{(X_1-\mu_1)(X_2-\mu_2)\}\begin{bmatrix} \sigma_{X_1}^2 & 0 \\ 0 & \sigma_{X_2}^2 \end{bmatrix}^{-1}\begin{bmatrix} X_1-\mu_1 \\ X_2-\mu_2 \end{bmatrix}}$$

# Level Curves of 2-D Gaussians

N-Dim Gaussian Dist

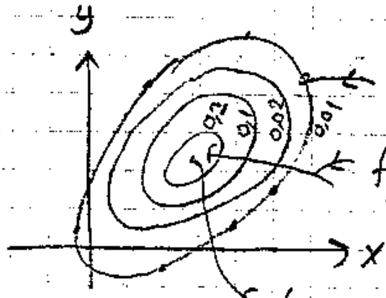
2-D Gaussian Dist

$$f_{\underset{\sim}{x},\underset{\sim}{y}}(x,y) = \frac{1}{(2\pi)|\underline{\underline{C}}_x|^{1/2}} e^{-\frac{1}{2}[(x-\mu_x)(y-\mu_y)]\underline{\underline{C}}_x^{-1}\begin{bmatrix}(x-\mu_x)\\(y-\mu_y)\end{bmatrix}}$$

Level curves of $f_{\underset{\sim}{x},y}(x,y)$

$$f_{xy}(x,y) = c$$

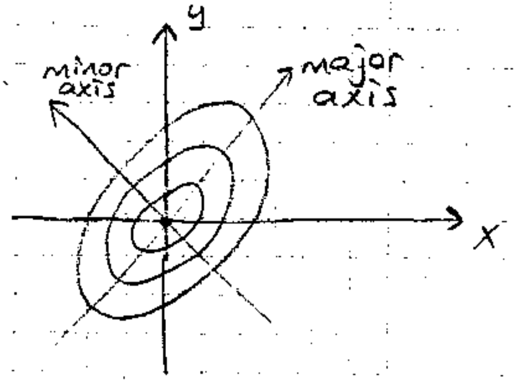
$$[(x-\mu_x)(y-\mu_y)]\underline{\underline{C}}_x^{-1}\begin{bmatrix}(x-\mu_x)\\(y-\mu_y)\end{bmatrix} = A$$



$\leftarrow f_{xy}(x,y) = \frac{1}{(2\pi)|C_x|^{1/2}} e^{-1/2 A} = 0{,}01$

$\leftarrow$ fxy increasing

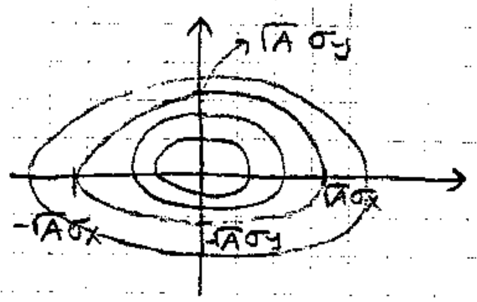
$\leftarrow (\mu_x, \mu_y)$



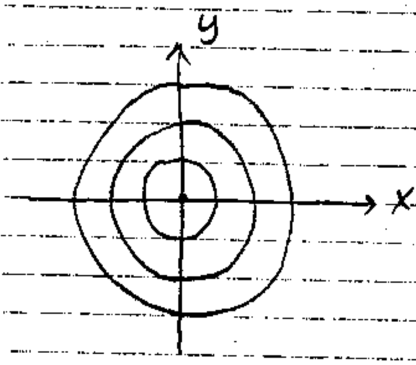
$$\mu_x = \mu_y = 0$$

$$\underline{\underline{C}}_x = \mathrm{diag}(\sigma_x^2, \sigma_y^2) \quad \mu_x = \mu_y = 0$$

$$[x \; y]\underbrace{\begin{bmatrix}1/\sigma_x^2 & 0\\ 0 & 1/\sigma_y^2\end{bmatrix}}_{C_x^{-1}}\begin{bmatrix}x\\y\end{bmatrix} = A$$



$$\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} = A$$



$\mu_X = \mu_y = 0$

$C_X = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \sigma^2$

## Linear Transformations on Random Vectors

$\underline{y} = \underline{\underline{A}}\, \underline{X}_{N\times 1}$ ← $\underline{X}$: random vector

↖ The mapped "$\underline{X}$".

$E\{\underline{X}\}$ and $\underline{\underline{C}}_X$ is known;
what is $E\{\underline{y}\}$ and $\underline{\underline{C}}_y$?

1- $\underline{y} = \underline{\underline{A}}\,\underline{X} \rightarrow E\{\underline{y}\} = E\{\underline{\underline{A}}\,\underline{X}\} = \underline{\underline{A}}\, E\{\underline{X}\}$

2- $\underline{\underline{C}}_X = E\{(X-\mu_X)(X-\mu_X)^T\} \rightarrow \underline{\underline{C}}_y = E\{(\underset{\underline{\underline{A}}\,\underline{X}}{y} - \overset{\underline{\underline{A}}\underline{\mu}_X}{\mu_y})(y-\mu_y)^T\}$

$= E\{\underline{\underline{A}}(X-\mu_X)(X-\mu_X)^T \underline{\underline{A}}^T\}$

$= \underline{\underline{A}}\, \underline{\underline{C}}_X\, \underline{\underline{A}}^T$

## Diagonalization of Auto-Corr. Matrices through Linear Processing (Whitening).

$\underline{X}$: $N\times 1$ random vector

without loss of generality $\mu_X = 0$

$\underline{\underline{C}}_X$ is given $\underline{\underline{C}}_X = E\{(X-\mu_X)(X-\mu_X)^T\}$

$= E\{X X^T\} \quad (\mu_X = 0)$

Finding a $\underline{\underline{T}}$ matrix such that

$\underline{y} = \underline{\underline{T}}\,\underline{X}$

and $\underline{y}$ has a diagonal auto-cor. matrix

1- Solution by Orthogonal (Unitary) Transformation

$$\underline{\underline{C}}_y = \underline{\underline{T}}\,\underline{\underline{C}}_x\,\underline{\underline{T}}^H$$

Remember:

$$\underbrace{\underline{\underline{C}}_x}_{N\times N}^{\swarrow \text{Hermitian matrix!}} = \underline{\underline{E}}\,\underline{\underline{\Lambda}}\,\underline{\underline{E}}^{-1}$$

(eig. decom.)

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \_\_, \lambda_N)$$

$$\underline{\underline{E}} = [\underline{e}_1\ \underline{e}_2\ \ldots\ \underline{e}_N]$$

and

$$\underline{\underline{C}}_x\,\underline{e}_k = \lambda_k\,\underline{e}_k$$

Since

$\underline{\underline{C}}_x$ : Hermitian

$e_k \perp e_l$

$(k \neq l)$

$$\underline{\underline{E}}^H \cdot \underline{\underline{E}} = \begin{bmatrix} \underline{e}_1^H \\ \underline{e}_2^H \\ \vdots \\ e_N^H \end{bmatrix} [\underline{e}_1 \ \ldots\ \underline{e}_N] = \begin{bmatrix} \overset{1}{\underline{e}_1^H \underline{e}_1} & \overset{0}{\underline{e}_1^H \underline{e}_2} & \cdots & \overset{0}{\underline{e}_1^H \underline{e}_N} \\ \underline{e}_2^H \underline{e}_1{}^{0} & & & \\ \vdots & & & \overset{1}{e_N^H e_N} \end{bmatrix}$$

$= I$ provided $\|e_k\|^2 = 1$

$$\underline{\underline{E}}^H \underline{\underline{E}} = I \longrightarrow \underline{\underline{E}}^{-1} = \underline{\underline{E}}^H$$

Then

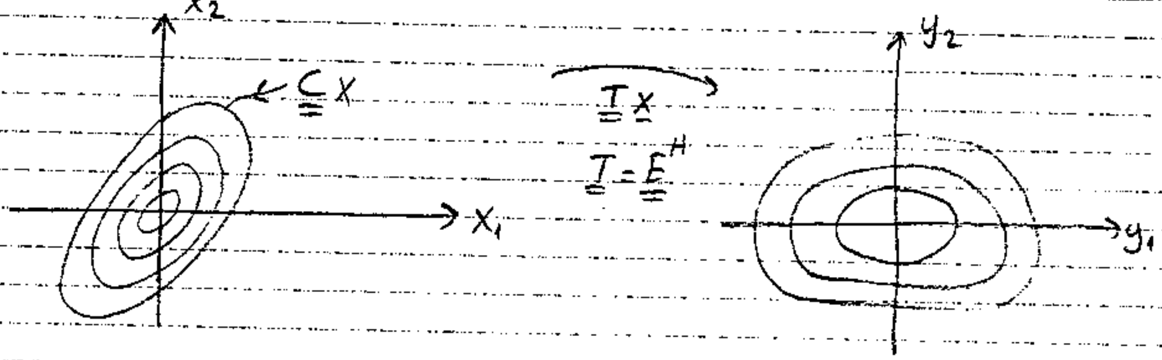
$$\underline{\underline{C}}_x = \underline{\underline{E}}\,\Lambda_x\,\underline{\underline{E}}^H$$

Finding $\underline{\underline{T}}$.

$$\underline{\underline{T}} = \underline{\underline{E}}^H \rightarrow \underline{\underline{C}}_y = \underline{\underline{E}}^H \underline{\underline{C}}_x \underline{\underline{E}}$$

$$= \Lambda_x \Leftarrow \text{diagonal matrix of eigenvalues}$$

$$\underline{\underline{T}} = E^H = \begin{bmatrix} \underline{e}_1^H \\ \underline{e}_2^H \\ \vdots \\ \underline{e}_N^H \end{bmatrix}$$

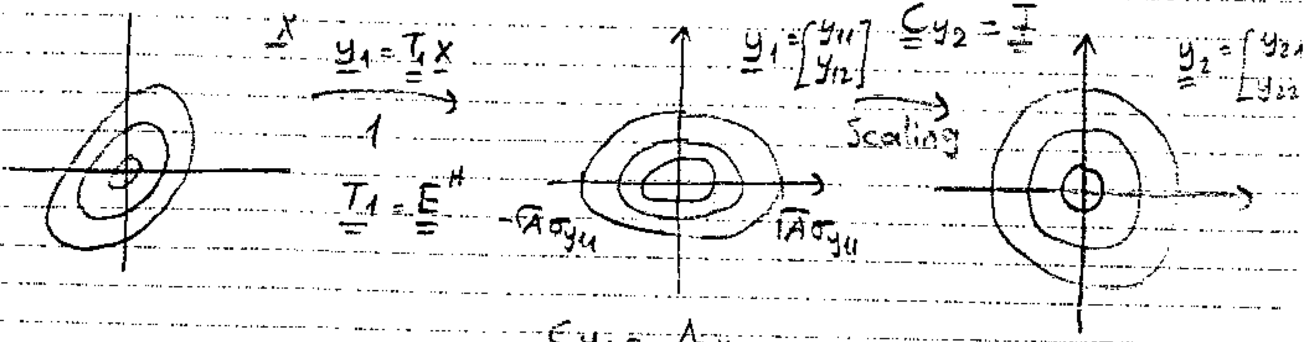


$(y_1, y_2)$ are uncorrelated (since $\underline{\underline{C}}_y \Rightarrow$ diagonal!)

If $\underline{x}$ : Gaussian $\rightarrow \underline{y} = \underline{\underline{T}}\underline{x}$ also Gaussian (for all $\underline{\underline{T}}$ matrices)

then $y_1, y_2$ independent.

2- Finding $\underline{\underline{T}}_2$ matrix such that $\underline{y}_2 = \underline{\underline{T}}_2 \underline{x}$ has covariance matrix of Identity.



$$\underline{\underline{C}}_{y_1} = \underline{\underline{\Lambda}}_x = \begin{bmatrix} \lambda_1 \lambda_2 & 0 \\ 0 & \lambda_N \end{bmatrix}$$

$$\underline{\underline{C}}_{y_2} = \underline{\underline{I}}$$

$$\underline{\underline{C}}_{y_2} = \underline{\underline{M}}\,\underline{\underline{C}}_{y_1}\underline{\underline{M}}^H$$

$$M = \begin{bmatrix} 1/\sqrt{\lambda_1} & & & 0 \\ & 1/\sqrt{\lambda_2} & & \\ & & \ddots & \\ 0 & & & 1/\sqrt{\lambda_N} \end{bmatrix}$$

$$\underline{\underline{T}}_2 = \begin{bmatrix} 1/\sqrt{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & 1/\sqrt{\lambda_N} \end{bmatrix} \begin{bmatrix} \underline{e}_1^H \\ \vdots \\ \underline{e}_N^H \end{bmatrix}$$

$\underline{\underline{R}}x$ : Auto-correlation matrix of $\underline{x}$

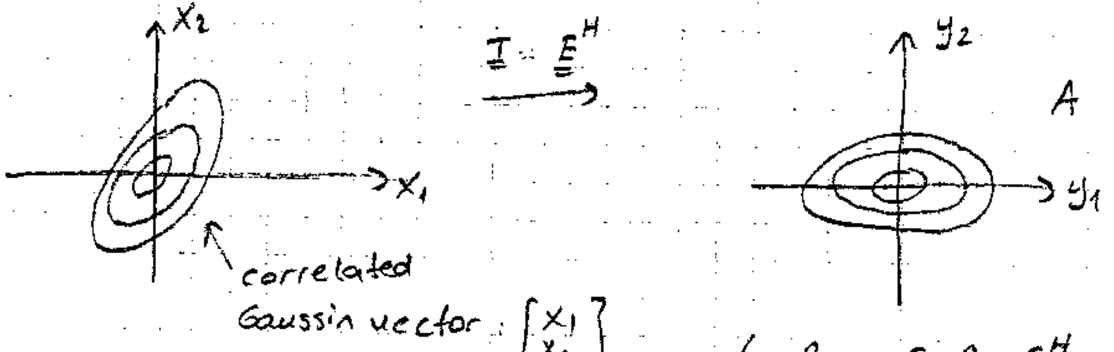
$\underline{y} = \underline{T}\,\underline{x}$

s.t. $\underline{\underline{R}}_y$ is diagonal matrix

special, v: Gaussian vector $\longrightarrow$ $\underline{y} = \underline{\underline{T}}\,\underline{x}$ satisfying.

This condition becomes a random vector with independent entries.

1-



correlated Gaussin vector : $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

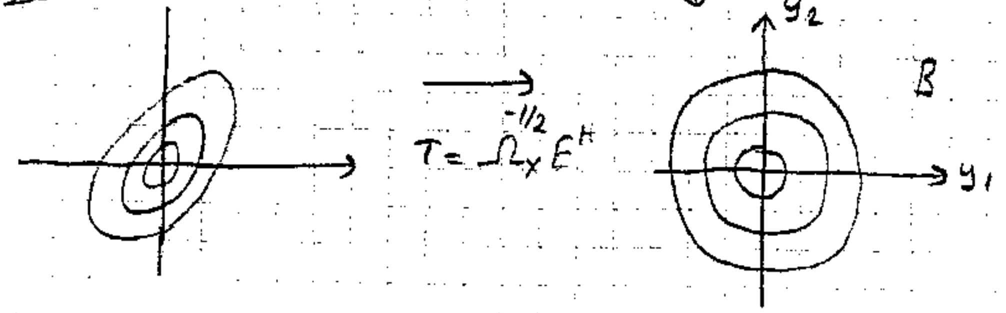
$\underline{\underline{R}}x = E\,\Omega_x\,E^H$

$\underline{\underline{E}} = [e_1\ e_2\ —\ e_N]$

$\underline{\underline{\Omega}}_x^{-1/2}$

$$y = \underline{\underline{T}}x \rightarrow \underline{y} = \underline{E}^H x = \begin{bmatrix} \underline{e}_1^H x \\ \underline{e}_2^H x \\ \vdots \\ \underline{e}_N^T \underline{x} \end{bmatrix}$$

2-



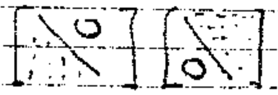
$R_y = \alpha\,\underline{\underline{I}}$

$$\underline{y}_B = \underline{\underline{\Omega}}_x^{-1/2}\, \underline{y}_A \longrightarrow \underline{\underline{R}}_{y_B} = \Omega_x^{-1/2}\, \underline{\underline{R}}_{y_A} (\Omega_x^{-1/2}) = I$$

3- Diagonalization by LU decomposition:

LU: Lower Upper Decomposition

$$\underline{\underline{A}} = \underline{\underline{L}}\,\underline{\underline{U}}$$



$$\underline{\underline{A}} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \longrightarrow \begin{bmatrix} 1 & 0 \\ -\frac{c}{a} & 1 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & -\frac{bc}{a} + d \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -\frac{c}{a} & 1 \end{bmatrix}^{-1} \begin{bmatrix} a & b \\ 0 & -\frac{bc}{a} + d \end{bmatrix}$$

$$\underline{\underline{A}} = \underline{\underline{L}}\,\underline{\underline{U}}$$

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 \\ -\frac{d}{a} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\underline{\underline{L}}_1} \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} = \begin{bmatrix} a & b & c \\ 0 & -\frac{bd}{a} + e & -\frac{cd}{a} + f \\ g & h & i \end{bmatrix}$$

$$\underline{\underline{L}}_1 A = \begin{bmatrix} \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}$$

$$\underline{\underline{L}}_2\, \underline{\underline{L}}_1\, A = \begin{bmatrix} \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot \\ 0 & \cdot & \cdot \end{bmatrix}$$

$$\underbrace{(L_3 \, L_2 \, L_1)}_{\text{Lower } \Delta} A = \begin{bmatrix} \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot \\ 0 & 0 & \cdot \end{bmatrix} \rightarrow \underline{\underline{A}} = \underline{\underline{L}} \, \underline{\underline{U}}$$

Unit Lower Δ Matrix: Lower Δ matrix with 1's on the diagonal.

$$\begin{bmatrix} 10 & 0 \\ 5 & 20 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 0 \\ \frac{5}{10} & 1 \end{bmatrix}}_{\text{Unit Lower } \Delta} \underbrace{\begin{bmatrix} 10 & 0 \\ 0 & 20 \end{bmatrix}}_{\text{Diagonal matrix}}$$

↙ Hermitian

$$\underline{\underline{R}}_x = \underline{\underline{L}} \, \underline{\underline{U}}$$

$$= \underline{\underline{L}} \, \underline{\underline{L}}^H = (\underline{\underline{L}}_u \underline{\underline{D}})(\underline{\underline{L}}_u \underline{\underline{D}})^H = \underline{\underline{L}}_u \underline{\underline{D}}^2 \underline{\underline{L}}_u^H$$ ↖ unit Lower Δ matrix

Then

$$\underline{y} = \underline{\underline{T}} \, \underline{x}$$

select $\underline{\underline{T}} = \underline{\underline{L}}_u^{-1}$

Then $y = \underline{\underline{L}}_u^{-1} \underline{X}$

$$R_y = \underline{\underline{L}}_u^{-1} \, \underline{\underline{R}}_x (\underline{\underline{L}}_u^{-1})^H$$

$$= D^2$$ ↙ Diagonal matrix

```
>> cholesky(A)
```

$$\underline{X} = \begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ \vdots \\ x[N-1] \end{bmatrix}$$ ← has auto-cor. $\underline{\underline{R}}_x$

$$\begin{bmatrix} y[0] \\ y[1] \\ y[2] \\ \vdots \\ y[N-1] \end{bmatrix} = \underline{y} = \underline{\underline{L}}_u \underline{x} = \begin{bmatrix} 1 & & & 0 \\ \alpha & 1 & & \\ \beta & \gamma & \ddots & \\ \theta & \sigma & & \\ & & & 1 \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ \vdots \\ x[N-1] \end{bmatrix}$$

Causal Transformation

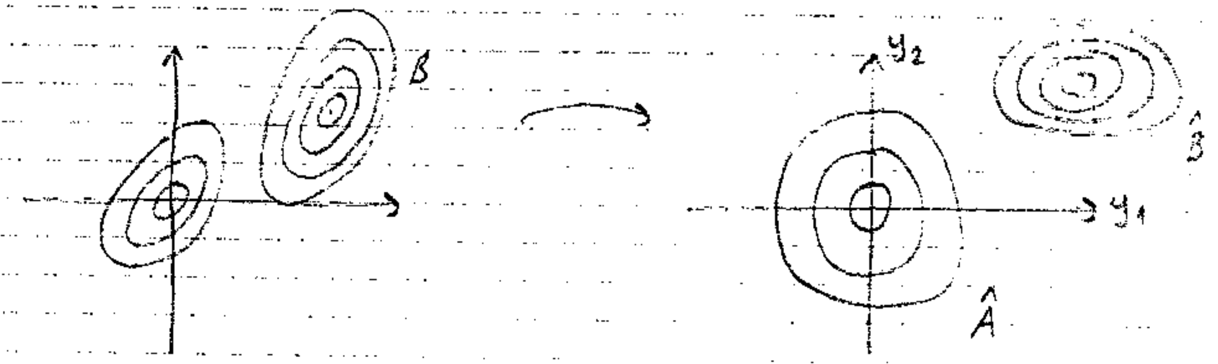
Find an $\underline{\underline{L}}_u$ matrix s.t.

$\underline{\underline{R}}_y$ is diagonal

4- $\underline{\underline{R}}_A$ and $\underline{\underline{R}}_B$ Auto-cov. matrices

and Find $\underline{\underline{T}}$ s.t. $\hat{\underline{a}} = \underline{\underline{T}}\, \underline{a}$, $\hat{\underline{b}} = \underline{\underline{T}}\, \underline{b}$, $\underline{\underline{R}}_{\hat{A}}$ and $R_{\hat{B}}$ are both diagonal.

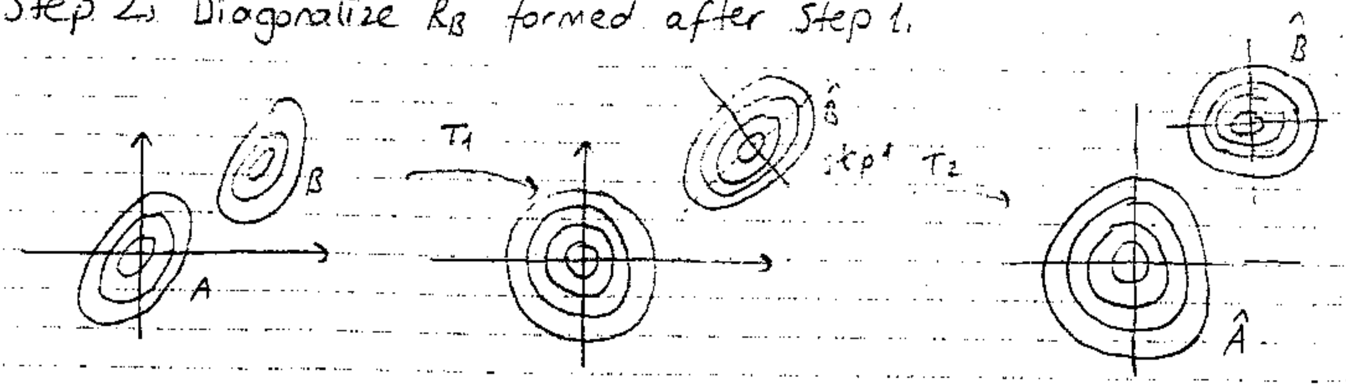
Joint Diagonalization of two cov matrices (Therrien)



Step 1: Transform $R_A$ into $\underline{\underline{I}}$ matrix.

check what happens to $R_B$ after this transformation.

Step 2: Diagonalize $R_B$ formed after Step 1.



Step 1: $\underline{\underline{T}}_1 = \underline{\underline{\Omega}}_A^{-1/2} E_A^H \longrightarrow \underline{\underline{R}}_A \xrightarrow{T_1} \underline{\underline{R}}_{\hat{A}} = T_1 R_A T_1^H = \underline{\underline{I}}$

$(R_A = E_A \underline{\underline{\Omega}}_A E_A^H)$

$$\underline{\underline{R}}_B \xrightarrow{T_1} \underline{\underline{R}}_{\hat{B}} = \underline{\underline{T}}_1 \underline{\underline{R}}_B \underline{\underline{T}}_1^H$$

Step 2: (Diagonalize $\underline{\underline{R}}_{\hat{B}}$ using eigenvectors)

Find eigenvectors of $R_{\hat{B}}$

$$\overset{\swarrow T_1 R_B T_1^H}{\underline{\underline{R}}_{\hat{B}}} \underline{e}_k = \lambda_k \underline{e}_k$$

$$\underline{\underline{R}}_B \underline{\underline{T}}_1^H \underline{e}_k = \lambda_k \underline{\underline{T}}_1^{-1} \underline{e}_k$$

$$= \lambda_k \underbrace{(\underline{\underline{E}}_A \underline{\underline{\Omega}}_A^{1/2})\, \overbrace{\underline{\underline{\Omega}}_A^{1/2} \underline{\underline{E}}_A^H}^{\underline{\underline{I}}}}_{\underline{\underline{R}}_A} \underbrace{\underline{\underline{E}}_A \underline{\underline{\Omega}}_A^{-1/2}}_{T_1^H} \underline{e}_k$$

$$\underline{\underline{R}}_B \underbrace{\underline{\underline{T}}_1^H \underline{e}_k}_{\underline{f}_k} = \lambda_k \underline{\underline{R}}_A \underbrace{\underline{\underline{T}}_1^H \underline{e}_k}_{\underline{f}_k} \longrightarrow \underline{\underline{R}}_B \underline{f}_k = \lambda_k \underline{\underline{R}}_A \underline{f}_k$$

($f_k$: The generalized eigenvector of $\underline{\underline{R}}_A$ and $\underline{\underline{R}}_B$)

At the end. Step 1 and Step 2

$$\underline{\underline{I}} = \begin{bmatrix} e_1^H \\ e_2^H \\ \vdots \\ e_N^H \end{bmatrix} \overset{\swarrow \text{Transformation of first step}}{(T_1)} = \begin{bmatrix} f_1^H \\ f_2^H \\ \vdots \\ f_N^H \end{bmatrix}$$

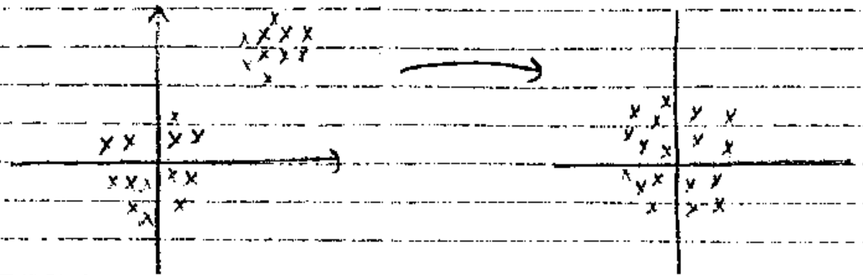
↖ cascade of step
step-2

Check $\hat{A}$ after 2nd step

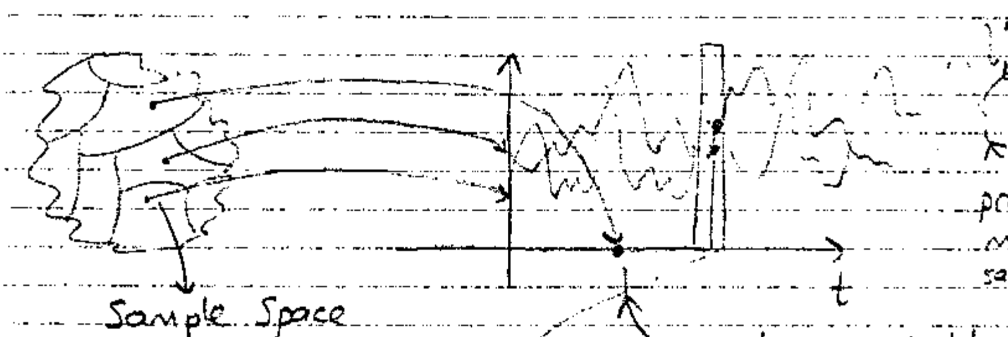
$$\underline{\underline{R}}_{\hat{A}} = \underline{\underline{I}}$$

$$\underline{\hat{\hat{a}}} = \begin{bmatrix} e_1^H \\ e_2^H \\ \vdots \\ e_N^H \end{bmatrix} \underline{\hat{a}} = \underline{\underline{E}}^H_{2^{nd}\,stage} \hat{a} \qquad \underline{\underline{R}}_{\hat{\hat{A}}} = E^H_{2^{nd}\,stage} \underline{\underline{I}} E_{2^{nd}\,stage}$$

$$= \underline{\underline{I}}$$



## Random Processes:



a random process is a mapping from sample space to "functions"

Sample Space

a random variable is a mapping from sample space to real line.

Evaluation of a random process at time $t_x$.

$$\underset{\sim}{x}(t)\Big|_{t=t_x} = \underset{\sim}{z} \leftarrow \text{random variable}$$

random process

Ex: $\underset{\sim}{x}(t) = A\cos(\underset{\sim}{\omega}t + 45^\circ)$

$$\hat{x}(t_y) = \underset{\text{combination}}{\text{Linear}} \{x(t_{x_1}), x(t_{x_2}), x(t_{x_3})\}$$

$$t_y > t_{x_k} \quad k = \{1, 2, 3\}$$

$$\underset{\sim}{x}(t) = \underset{\sim}{s}(t) + \underset{\sim}{w}(t) \leftarrow \text{noise}$$

signal

$\hat{s}(t_x)$ using the following system

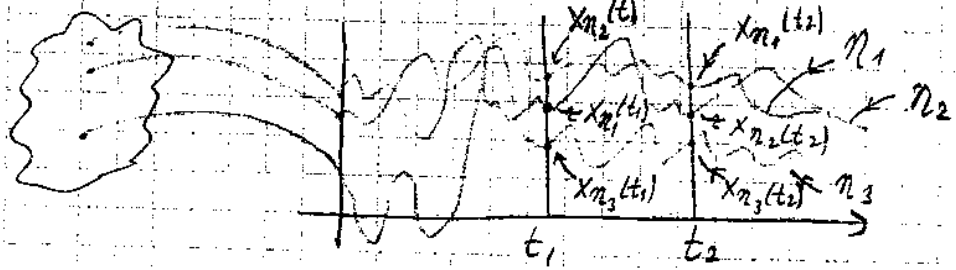
$$\hat{s}(t_x) = \int_{-\infty}^{t_x} \underset{\sim}{x}(t)\,h(t)\,dt \quad \leftarrow h(t)\text{: weighting function}$$

$E\{(\hat{s}(t_x)) - s(t_x))^2\}$

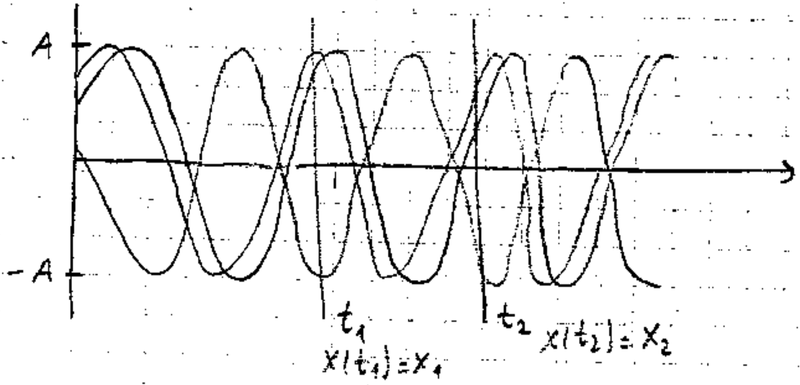
$\min_{h(t)} \; E\{(\hat{s}(t_x) - s(t_x))^2\}$

## Description of Random Processes

1- Joint pdf Description of a R.P.



$\underset{\frown}{Ex}$: $x(t) = A\cos(2\pi f_1 t + \underset{\sim}{\theta})$ ← $\theta_n$ : unit $[0, 2\pi]$



$1^{st}$ order pdf description: $f_{\underset{\sim}{x}(t_1)}(x_1) \; \forall t_1$

$2^{nd}$ order pdf description: $f_{\underset{\sim}{x}(t_1), \underset{\sim}{x}(t_2)}(x_1, x_2) \quad \forall (t_1, t_2)$

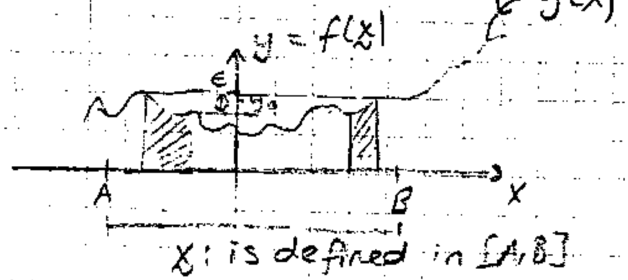
$1^{st}$ order description:

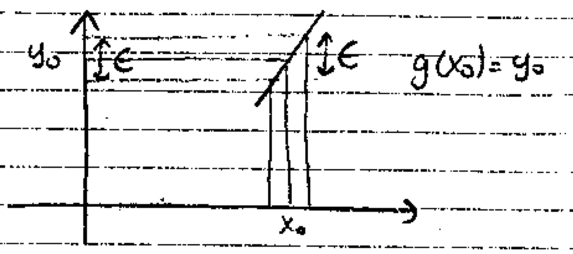
$x_1 = x(t_1) = A\cos(2\pi f t_1 + \underset{\sim}{\theta})$

$\underset{\sim}{\theta} \rightarrow \underset{\sim}{y} = f(\underset{\sim}{\theta})$

find p.d.f of y (Papouliss' Prob. Book)

1 func. of 1. r.v





$$\rightarrow \quad f_{X(t_1)}(x_1) = \begin{cases} \dfrac{1/\pi}{\sqrt{A^2 - x_1^2}} & |x_1| < A \\ 0 & |x_1| > A \end{cases} \qquad \forall\, t_1$$

is constant

2nd Order Description:

$$f_{X(t_1), X(t_2)}(x_1, x_2) = f_{X(t_2)|X(t_1)}(x_2 | x_1)\, f_{X_1(t_1)}(x_1)$$

$$x_2 = A\cos(2\pi f_1 t_2 + \underset{\sim}{\theta})$$

$$= A\cos(2\pi f_1 t_2 + 2\pi f(t_1 - t_1) + \theta)$$

$$= A\cos(\underbrace{2\pi f t_1 + \theta}_{I} + \underbrace{2\pi f(t_2 - t_1)}_{II})$$

$$= A\cos(I)\cos(II) - A\sin(I)\sin(II)$$

$$= \underset{\text{known}}{x_1}\cos(II) - (\mp\sqrt{A^2 - x_1^2})\,\underset{\text{known}}{\sin(II)}$$

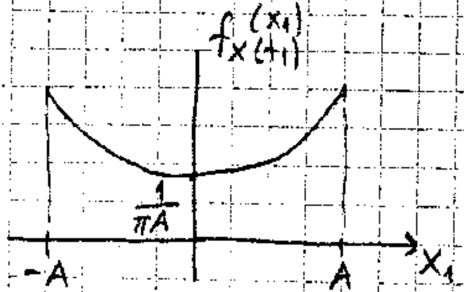
$\{x_{21}, x_{22}\}$ corresponds to $\mp$ selection result.

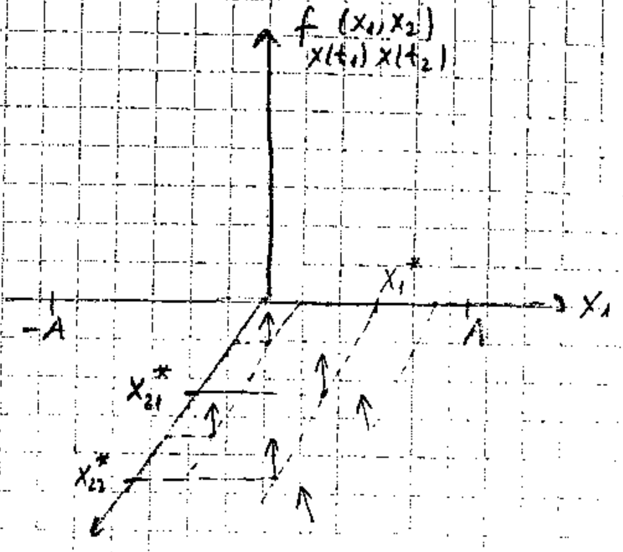
Then $x_2$ can take only two possible values if $x_1 = x(t_1)$ is known.

I will call these values as $x_{21}, x_{22}$ for these values.
possible $x_2$ values

$$\rightarrow \quad f_{X(t_2)|X(t_1)}(x_2 | x_1) = \frac{1}{2}\left\{\delta(x_2 - x_{21}) + \delta(x_2 - x_{22})\right\}$$

$$f_{X(t_2)|X(t_1)}(x_2, x_1) = \begin{cases} \dfrac{1}{2}\left\{\delta(x_2 - x_{21}) + \delta(x_2 - x_{22})\right\} \dfrac{\frac{1}{\pi}}{\sqrt{A^2 - x_1^2}} & |x_1| < A \\ 0 & |x_1| > A \end{cases}$$





$3^{rd}$ Order Description:

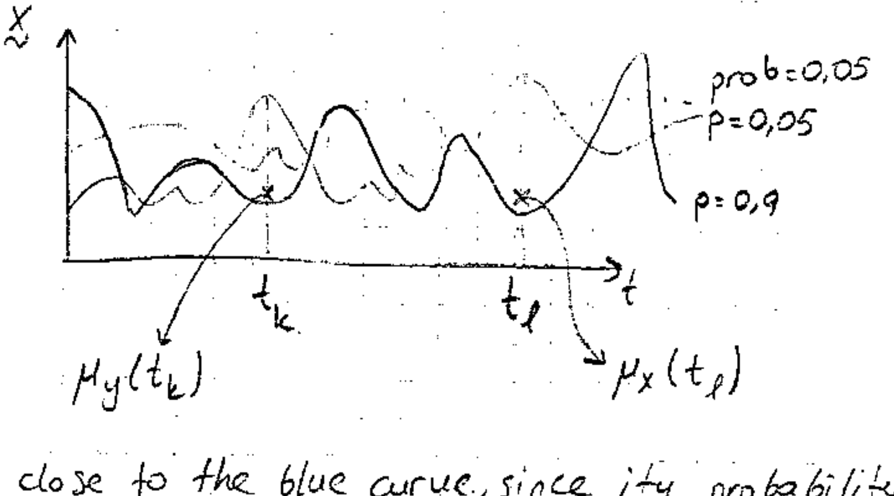
If you know $\left.\begin{matrix} x(t_1)=x_1 \\ x(t_2)=x_2 \end{matrix}\right\}$ $x(t_3)$ is fixed

$$f_{x(t_1),x(t_2),x(t_3)}(x_1,x_2,x_3) = \underbrace{f_{x(t_3)|x(t_2),x(t_1)}(x_3|x_2,x_1)}_{\delta(x_3-x_{31})} \cdot \underbrace{f_{x(t_2),x(t_1)}(x_2,x_1)}_{\text{the same relation found.}}$$

Description of Random Processes

1- pdf description

2- Moment description



Mean is very close to the blue curve, since ity probability is the most.

$1^{st}$ order moment description

$\mu_x(t_k) = E\{x(t_k)\} \quad \forall t_k$

$2^{nd}$ order moment description

$R_x(t_k,t_l) = E\{x(t_k)x^*(t_l)\} \quad \forall t_k, t_l$

Gaussian distribution

$x_k = x(t_k)$ , $x_l = x(t_l)$

$$f_{x(t_k)x(t_l)}(x_k,x_l) = \frac{1}{2\pi|\underline{\underline{C}}_x(t_k,t_l)|^{1/2}} e^{-\frac{1}{2}\begin{bmatrix}x_k-\mu_k\\x_l-\mu_l\end{bmatrix}^T \underbrace{\underline{\underline{C}}_x^{-1}}_{\Downarrow} \begin{bmatrix}x_k-\mu_k\\x_l-\mu_l\end{bmatrix}}$$

related to $1^{st}$ and $2^{nd}$ moments.

$C_x(t_k,t_l) = R_x(t_k,t_l) - \mu_k\mu_l$

Gaussian Process

Every set of samples of the process is jointly Gaussian distributed.

So knowing the $1^{st}$ and $2^{nd}$ order moments of the Gaussian process is equivalent to joint pdf description.

Ex: Let $\underset{\sim}{x}(t)$ is a random process whose $\mu_x(t) = 3$ and $R_x(t_1,t_2) = 9 + 4e^{-0.2|t_1-t_2|}$, $\underset{\sim}{z} = \underset{\sim}{x}(5)$, $\underset{\sim}{w} = \underset{\sim}{x}(8)$

Find $E\{z\}$, $E\{w\}$, $E\{z^2\}$, $E\{w^2\}$, $E\{zw\}$

a) $E\{z\} = E\{x(5)\} = \mu_x(5) = 3$

b) $E\{w\} = E\{x(8)\} = \mu_x(8) = 3$

c) $E\{z^2\} = E\{x(5)x(5)\} = R_x(5,5) = 13$

d) $E\{w^2\} = E\{x(8)x(8)\} = R_x(8,8) = 13$

e) $E\{zw\} = E\{x(5)x(8)\} = R_x(5,8) = 9 + 4e^{-0.6}$

Ex: $\underset{\sim}{z} = x(t_1) + x(t_2)$ $\quad E\{z^2\} = ?$

like averaging
(Low pass filtering)

$$E\{z^2\} = E\{[x(t_1)+x(t_2)]^2\} = R_x(t_1,t_1) + \underbrace{2R_x(t_1,t_2)}_{\Downarrow} + R_x(t_2,t_2)$$

Correlation of $x(t_1)$ and $x(t_2)$
is needed for output $\underset{\sim}{z}$
mean power calculation

Ex: $S = \int_a^b \underset{\sim}{x}(t)\,dt$ a) $E\{S\} = ?$

b) $E\{S^2\} = ?$

$$E\{S\} = E\{\int_a^b x(t)\,dt\} = \int_a^b E\{x(t)\}\,dt = \int_a^b \mu_x(t)\,dt$$

$$E\{S^2\} = E\{\int_a^b x(t)\,dt \int_a^b x(\tau)\,d\tau\} = \int_a^b\int_a^b E\{x(t)x(\tau)\}\,dt\,d\tau$$

$$= \int_a^b\int_a^b R_x(t,\tau)\,dt\,d\tau$$

Ex: fundamentally important

$\underset{\sim}{x}(t) = \underset{\sim}{\gamma}\cos(\omega t + \underset{\sim}{\theta})$

$\underset{\sim}{\gamma}$ and $\underset{\sim}{\theta}$ are independent

$\underset{\sim}{\theta}$: uniform in $[-\pi, \pi]$

$\underset{\sim}{\gamma}$: pdf not given.

a) $\mu_x(t) = ?$

$$\mu_x(t) = E_{\gamma,\theta}\{\underset{\sim}{\gamma}\cos(\omega t + \underset{\sim}{\theta})\} \underset{\substack{\downarrow \\ \text{independence}}}{=} E_{\underset{\sim}{\gamma}}(\underset{\sim}{\gamma})\, E_{\theta}\{\cos(\omega t + \underset{\sim}{\theta})\}$$

$$= \underbrace{\left[\int_{-\pi}^{\pi} \frac{1}{2\pi}\cos(\omega t + \underset{\sim}{\theta})\,d\underset{\sim}{\theta}\right]}_{\text{zero}} \left[E_{\underset{\sim}{\gamma}}(\underset{\sim}{\gamma})\right] = 0$$

b) $R_x(t_1, t_2) = ?$

$$R_x(t_1,t_2) = E\{\underset{\sim}{\gamma}^2\cos(\omega t_1 + \underset{\sim}{\theta})\cos(\omega t_2 + \underset{\sim}{\theta})$$

$$= E\{\underset{\sim}{\gamma}^2\}\,E\{\cos(\omega t_1 + \underset{\sim}{\theta})\cos(\omega t_2 + \underset{\sim}{\theta})$$

$$= E\{\underset{\sim}{\gamma}^2\}\,E\{\frac{1}{2}\cos(\omega t_1 + \omega t_2 + 2\underset{\sim}{\theta}) + \frac{1}{2}\cos(\omega(t_1 - t_2))\}$$

$$= \frac{1}{2}E\{\underset{\sim}{\gamma}^2\}\cos(\omega(t_1 - t_2)) + \frac{1}{2}E\{\underset{\sim}{\gamma}^2\}\underbrace{E\{\cos(\omega(t_1 + t_2 + \underset{\sim}{\theta})}_{0}$$

$$= \frac{1}{2}E\{\underset{\sim}{\gamma}^2\}\cos(\omega(t_1 - t_2))$$

$$\underbrace{R_x(t_1,t_2)}_{\Downarrow \text{ two-variables}} = \frac{1}{2} E(\underset{\sim}{\gamma}^2] \cos(\underbrace{\omega(t_1-t_2)}_{D=t_1-t_2 \text{ one variable}})]$$ is a function of $t_1-t_2$ only.

$\Rightarrow R_x(t_1+A, t_2+A) = R_x(t_1,t_2)$

In discrete time

$\rightarrow$ $R_x[n,k] = r_x[n-k]$

$x(0)$ 1) ?

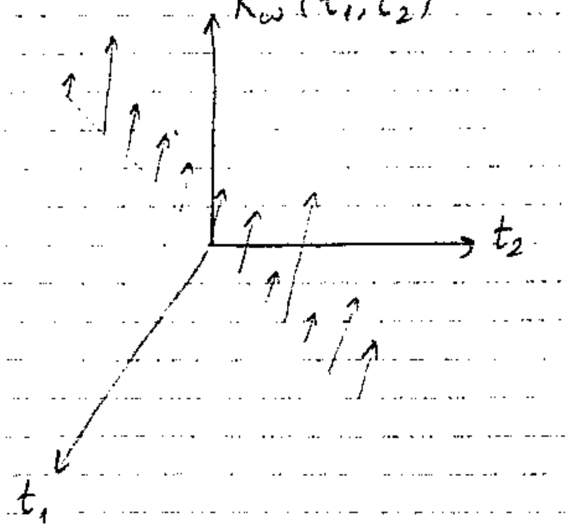
$$\begin{array}{c|ccc} & k=0 & k=1 & k=2 \\ \hline n=0 & R_x(0,0) & R_x(0,1) & R_x(0,2) \\ n=1 & R_x(1,0) & R_x(1,1) & R_x(1,2) \\ n=2 & R_x(2,0) & R_x(2,1) & R_x(2,2) \end{array} = \begin{bmatrix} r_x(0) & r_x(-1) & r_x(-2) \\ r_x(1) & r_x(0) & r_x(-1) \\ r_x(2) & r_x(1) & r_x(0) \end{bmatrix}$$

Toeplitz matrix

White Noise

$\underset{\sim}{\omega}(t)$ is called white noise if $E\{\underset{\sim}{\omega}(t)\} = 0$ zero mean

$E\{\omega(t_1)\omega(t_2)\} = R_\omega(t_1,t_2) = \sigma^2_{\omega(t)} \delta(t_1-t_2)$

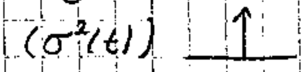


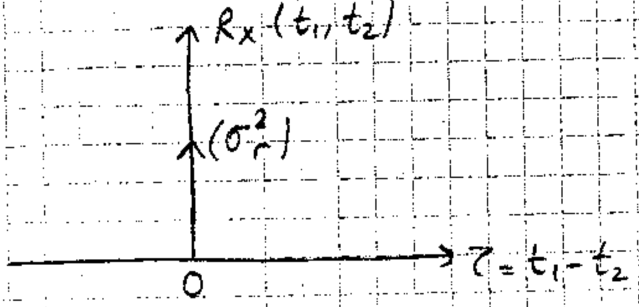
$\omega(t)$ is called stationary white noise if

$E\{\omega(t)\} = 0$

$E\{\omega(t_1)\omega(t_2)\} = \underset{\substack{\downarrow \\ \text{not a function of time} \\ \text{it is constant}}}{\sigma_\omega^2} \delta(t_1-t_2)$

→ Yeterlilik impulse'ın weight'i parantez içinde gösterilir.

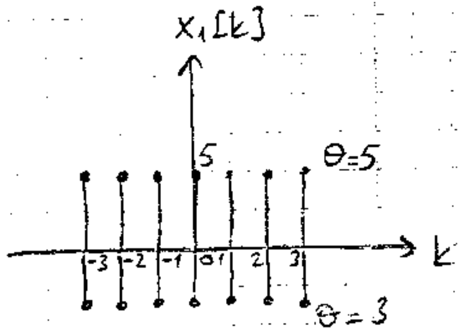


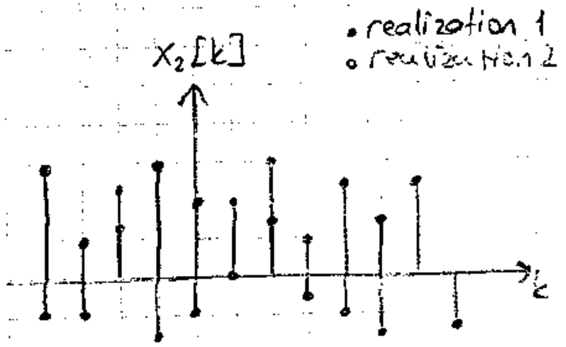


Ex) Two random processes $\underset{\sim}{X}_1[k]$ and $\underset{\sim}{X}_2[k]$ are given

a) $\underset{\sim}{X}_1[k] = \theta$ where $\theta$ is Gaussian dist. with $N(0,\sigma_\omega^2)$

b) $\underset{\sim}{X}_2[k] = \underset{\sim}{w}_k$ where $\underset{\sim}{w}_k$ i.i.d gaussian dist. with $N(0,\sigma_\omega^2)$





$1^{st}$ order pdf

$x_k = x_1[k] \Rightarrow f_{X_1[k]}(x_k) = N(0,\sigma_\omega^2)$ ↑ pdf of $\theta$

$x_k = X_2[k] \Rightarrow f_{X_2[k]}(x_k) = N(0,\sigma_\omega^2)$ ↑ pdf of $\underset{\sim}{w}[k]$

$\mu_{X_1[k]} = 0$

$\mu_{X_2[k]} = 0$

$2^{nd}$ order pdf

$$f_{X_1[k]X_2[l]}(x_k,x_l) = \underbrace{N(0,\sigma_\omega^2)}\,\underbrace{\delta[x_k - x_l]}$$

$$\downarrow$$

$$= f_{X_1[k]}(x_k) \cdot f_{X_1[l]|X_1[k]}(x_l|x_k)$$

iid (seperable)

$$f_{X_2[k]X_2[l]}(x_k,x_l) = f_{X_2[k]}(x_k)\, f_{X_2[l]}(x_l)$$

$$= N(0,\sigma_\omega^2) . N(0,\sigma_\omega^2)$$

## $2^{nd}$ moments

$R_{X_1}[k,l] = E\{X_1[k]X_2[l]\}$
$= E\{\theta^2\} = \sigma_\omega^2$

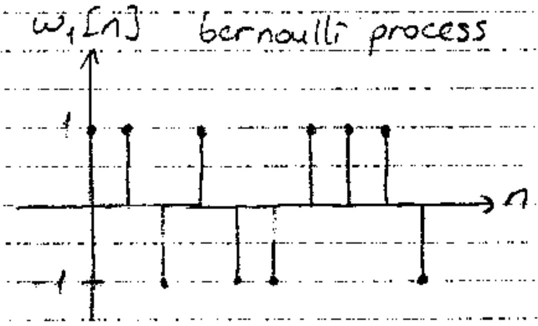
$R_{X_2}[k,l] = E\{X_2[k]X[l]\}$

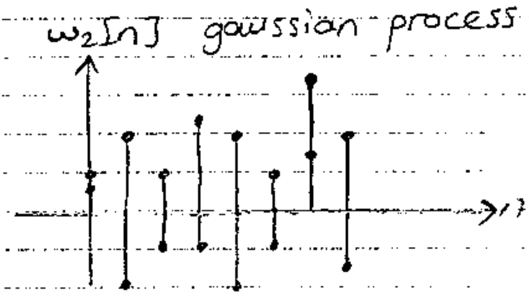
$$= \begin{cases} E\{X_2[k]\}E\{X_2[l]\} & k \neq l \\ E\{X_2^2[k]\} & k = l \end{cases}$$

$$= \begin{cases} 0 & k \neq l \\ \sigma_\omega^2 & k = l \end{cases}$$

$$= \sigma_\omega^2 \delta[k-l]$$

Ex: $\omega_1[n] = \pm 1$ i.i.d with equal probability for each $n$.
$\omega_2[n] = N(0,1)$ i.i.d for every $n$.





$$f_{\omega_1[n_1]\omega_1[n_2]\ldots\omega_1[n_L]} = \prod_{k=1}^{L} \underbrace{f_{\omega_1[n_k]}(x_k)}$$
$$\Downarrow$$
$$\frac{1}{2}\delta[x_k-1] + \frac{1}{2}\delta[x_k+1]$$

$$f_{\omega_2[n_1]\omega_2[n_2]\ldots\omega_2[n_L]} = \prod_{k=1}^{L} f_{\omega_2[n_k]}(x_k)$$
$$\Downarrow$$
$$N(0,1)$$

$\mu_{\omega_1}[k] = 0$

$R_{\omega_1}(n,k) = E\{\omega_1[n]\omega_1[k]\}$
$= \delta[n-k]$

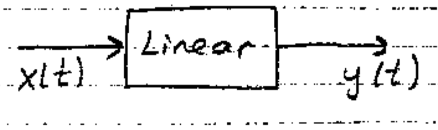
$\Rightarrow$ Stationary white noise

$\mu_{\omega_2}[k] = 0$

$R_{\omega_2}(n,k) = E\{\omega_2[n]\omega_2[k]\}$
$= \delta[n-k]$

$\Rightarrow$ Stationary white noise.

### Linear Systems with Stochastic Inputs (Papoulis):



output produced by an impulse at time $\tau$.

$$y(t) = \int_{-\infty}^{\infty} h(t,\tau)\,x(\tau)\,d\tau \xrightarrow{(LTI)} h(t,\tau) = h(t-\tau)$$

$$\int_{-\infty}^{\infty} h(t-\tau)\,x(\tau)\,d\tau \quad \text{(conv.)}$$

$$\left( \underline{y} = \underline{\underline{H}}\,\underline{x} \longrightarrow y_k = \sum_{n=1}^{N} \underbrace{H_{kn}}_{k^{th}\text{ row},\ n^{th}\text{ col. of } \underline{\underline{H}}} x_n \right)$$

My goal: Given moment descriptions of $x(t)$; Find moment description at output $(y(t))$.

$$x(t) \begin{cases} \nearrow \mu_x(t) \\ \searrow R_x(t_1,t_2) = E\{x(t_1)\,x(t_2)\} \quad \forall\,(t_1,t_2) \end{cases}$$

$y(t) \longrightarrow E\{y(t)\}$

Basic assumption

$$\underset{\uparrow \text{Linear system}}{E\{L(f(t))\}} = L\{E\{f(x)\}\} \qquad \underline{\underline{A}}\,\underline{\underline{B}} = \underline{\underline{B}}\,\underline{\underline{A}} \quad A, B \text{ commute}$$

$$y(t) \to \mu_y(t) = E\{y(t)\} = E\{L\{x(t)\}\} = L\{\mu_x(t)\} = \int_{-\infty}^{\infty} h(t,\tau)\,\mu_x(\tau)\,d\tau$$

$\hookrightarrow R_y(t_1,t_2) = ?$

Step 1: $R_{xy}(t_1,t_2) = E\{x(\overset{\text{fixed}}{t_1})\,y(t_2)\}$

$$= E\{x(t_1)\,L\{x(t)\}\big|_{t=t_2}\} = E\{L\{x(t_1)x(t)\}\big|_{t=t_2}\}$$

$$= L\{\underbrace{E\{x(t_1)x(t)\}}_{R_x(t_1,t)}\}\big|_{t=t_2}$$

$$= L\{R_x(t_1,t)\}\big|_{t=t_2}$$

$$= \int_{-\infty}^{\infty} h(t_2,\tau)\,R_x(t_1,\tau)\,d\tau$$

$$[R_x H^T]_{l,m} = \sum_{l'=1}^{N} R_x(l,l') \underbrace{H^T(l',m)}_{H(m,l')}$$

$$[R_y]_{n,k} = \sum_{k'=1}^{N} \sum_{l'=1}^{N} H[n,k'] R_x(k',l') H(k,l')$$

$$[R_y]_{5,2} = (\underline{\underline{H}})_{5,*} \underline{\underline{R}}x (H^T)_{*,2}$$

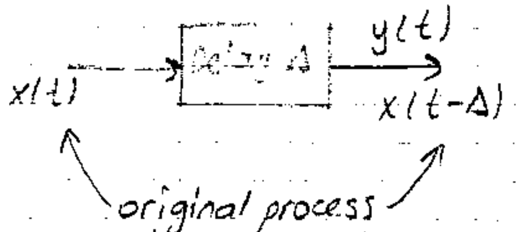
$$[h_{51} h_{52} \ldots] Rx \begin{bmatrix} h_{21} \\ h_{22} \\ \vdots \\ h_{2N} \end{bmatrix}$$

Ex: from Papoulis

- $\mu_y(t) = \frac{d}{dx} \mu_x(t)$
- $R_{yy}(t_1,t_2) = ?$

## Stationary Random Processes

If the time origin of the process is arbitrary; then the process is called a stationary process.



original process and delayed process have the same characteristics.

x(t), y(t) have the same

① joint pdf

② moment characterization

Step 2: $R_y(t_1,t_2) = E\{y(t_1)\overset{\text{fixed}}{y(t_2)}\}$ ← fixed

$$= E\{L\{x(t)\}\big|_{t=t_1}\overset{\text{fixed}}{y(t_2)}\}$$

$$= E\{L\{x(t)\,y(t_2)\}\big|_{t=t_1}\}$$

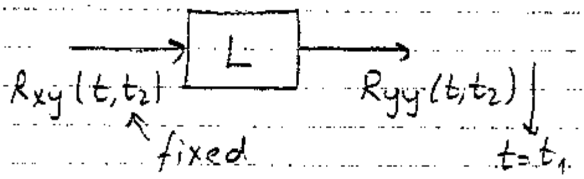
$$= L\{\underbrace{E\{x(t)\,y(t_2)\}}_{R_{xy}(t,t_2)}\}\big|_{t=t_1}$$

$$= \int_{-\infty}^{\infty} h(t_1,\tau)\,R_{xy}(\tau,t_2)\,d\tau$$

$$= \int_{-\infty}^{\infty} h(t_1,\tau')\left(\int h(t_2,\tau)\,R_x(\tau',\tau)\,d\tau\right)d\tau'$$

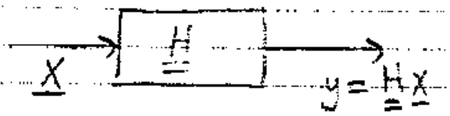
$$= \underset{\tau}{\int_{-\infty}^{\infty}}\underset{\tau'}{\int_{-\infty}^{\infty}} h(t_1,\tau')\,R_x(\tau',\tau)\,h(t_2,\tau)\,d\tau\,d\tau'$$

$$\xrightarrow{LTI} \begin{matrix} h(t_1,\tau') = h(t_1-\tau') \\ h(t_2,\tau) = h(t_2-\tau) \end{matrix} \longrightarrow \iint h(t_1-\tau')\,R_x(\tau',\tau)\,h(t_2-\tau)\,d\tau\,d\tau'$$



Discrete-time

$\underline{X}$: Random vector of dim N



$$R_y = \underset{\text{Matrix notation}}{\underline{\underline{H}}\,\underline{\underline{R}}_x\,\underline{\underline{H}}^T} \longrightarrow [R_y]_{n,k} = \sum_{k'=1}^{N} \boxed{H_{nk'}}\,[\underline{\underline{R}}_x\,\underline{\underline{H}}^T]_{k',k}$$

$n^{th}$ row, $k_1^{th}$ column

$$= \sum_{k'=1}^{N} H_{nk'}\left[\sum_{l'=1}^{N} R_x(\overset{l=k'}{l},l')\,H(\underset{m=k}{m},l')\right]_k$$

Two Types;

1- Stationary in pdf description (Strict Sense Stationary)

# random variables = 1 $\rightarrow f_{x(t_1)}(x) = f_{x(t_1+\Delta)}(x) \quad \forall (\Delta, t)$

# r.v = 2 $\rightarrow f_{x(t_1)x(t_2)}(x_1,x_2) = f_{x(t_1+\Delta,\, t_2+\Delta)}(x_1,x_2) \quad \forall (\Delta, t_1)$

# r.v = N $\rightarrow f_{x(t_1)\dots x(t_N)}(x_1 \dots x_N) = f_{x(t_1+\Delta,\dots,t_N+\Delta)}(x_1,\dots,x_N) \quad \forall(\Delta, t_1,\dots,t_N)$

2- Stationary in moments (WSS, Wide Sense Stationary)

$$\left.\begin{aligned} &1-\ \mu_x(t_1) = \mu_x(t_1+\Delta) = c \quad \forall(\Delta, t_1) \\ &2-\ R_x(t_1,t_2) = R_x(t_1+\Delta, t_2+\Delta) \quad \forall(\Delta, t_1, t_2) \end{aligned}\right\} \text{WSS}$$

$$= f(t_2-t_1)$$

$$R_x(t_1,t_2) = R_x(t_1+\Delta, t_2+\Delta) \Big\downarrow_{\Delta \rightarrow -t_1}$$

$$= R_x(0, t_2-t_1)$$

$$= f(t_2-t_1)$$

$$= f(\tau)$$

$t_2 - t_1 = \tau$ ← lag parameter

obviously

SSS → WSS

WSS → SSS

$$\text{If} \begin{pmatrix} x(t):\text{WSS} \\ \text{and} \\ \text{Gaussian} \end{pmatrix} \rightarrow \text{SSS}$$

Ex) $x(t) = \underset{\sim}{a}\cos\omega t + \underset{\sim}{b}\sin\omega t \quad (\omega \neq 0)$

Find conditions on $\underset{\sim}{a}$ and $\underset{\sim}{b}$ s.t. $x(t)$ is WSS.

1- $\mu_x(t) = c$

$\hookrightarrow E\{a\}\cos wt + E\{b\}\sin wt \Big\downarrow_{t=\frac{2\pi}{w}} = c$

$E\{a\} = c$

$E\{b\} = c \Big\downarrow_{t=\frac{\pi/2}{\omega}}$

$c(\cos wt + \sin wt) = c \longrightarrow c = 0$

2- $R_x(t_1, t_2) = f(t_2 - t_1)$

$R_x(t,t) = \text{constant} = f(0) \quad \forall t$

$R_x(t,t) = E\{(x(t))^2\} = E\{a^2\}\cos^2 wt + E\{ab\}\sin(2\omega t)$
$+ E\{b^2\}\sin^2 wt = \text{constant}$

$R_x(t,t)\Big\downarrow_{t=\frac{2\pi}{\omega}} = E\{a^2\}$

$E\{a^2\} = E\{b^2\}$

$R_x(t,t)\Big\downarrow_{t=\frac{\pi/2}{\omega}} = E\{b^2\}$

$R_x(t, t-\tau) \longleftarrow f(\tau)$ should not has an "t" dependence.

$\downarrow$

$E\{x(t)x(t-2)\} = E\{(a\cos wt + b\sin \omega t)(a\cos w(t-2) + b\sin w(t-2)$
$= E\{a^2\}(\overbrace{\cos wt\cos w(t-2) + \sin wt\sin w(t-2)}^{\cos(w\tau)})$
$+ E\{ab\}(\sin(wt + w(t-2)))$
$\Rightarrow E\{ab\} = 0 \quad \Rightarrow$ 2. condition

$\mu_x(t) = 0$
$R_x(t, t-\tau) = E\{a^2\}\cos(w\tau)$

See p 301 of Papoulis for the same example on the conditions for SSS.

Ex) $x[n]$ is a discrete r.p.

At even indexed samples, $x[2k]$: Unif$[-\sqrt{3}, \sqrt{3}]$

At odd sample, $x[2k+1]$: $N(0,1)$.

All samples are independent from each other

Q: $x[n]$: WSS or SSS or not stationary?

1- WSS: 1- $E\{x[n]\} = \begin{cases} E\{x[n]\} & n: \text{even} \\ E\{x[n]\} & n: \text{odd} \end{cases} = 0$

2- $R_x[n,m] = f(k)$ where $k=n-m$

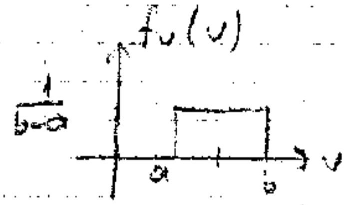
$R_x[n,m] = E\{x[n]\,x[m]\}$

$$= \begin{cases} E\{(x[n])^2\} = \sigma^2_{[n]} & n=m \\ E\{x[n]x[m]\} \quad 0 & n \neq m \end{cases}$$

$\searrow E\{x[n]\}E\{x[m]\}$

$$\sigma^2_{x[n]} = \begin{cases} \text{var}(\text{unif}[-\sqrt{3}, \sqrt{3}]) = 1 & n: \text{even} \\ \text{var}(N(0,1)) = 1 & n: \text{odd} \end{cases}$$

$R_x[n,m] = \delta[n,m]$



$\Delta = b-a$

$V_a(v) = \frac{\Delta^2}{12}$

WSS.

2- SSS: 1st order: $f_{x[2k]}(x) \overset{?}{=} f_{x[2k+1]}(x)$

Not even

First order stationary in pdf

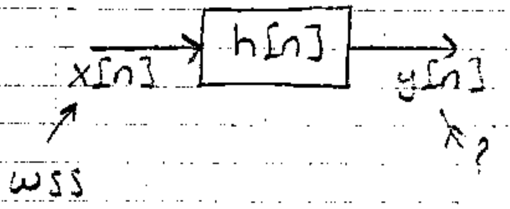
Jointly WSS Processes:

x[n], y[n] are called jointly WSS.

1- x[n] and y[n] are both WSS.

2- $R_{xy}[n,m] = E\{x[n]y^*[m]\} = f(n-m)$

Linear-Time Invariant Processing of WSS inputs:



$$y[n] = \sum_{k=-\infty}^{\infty} h[k]\,x[n-k]$$

Check WSS:

1- $E\{y[n]\} \overset{?}{=} c \rightarrow E\{\sum h[k]x[n-k]\} = \sum h[k].\mu_x[k] = \mu_x . H(0)$ ⇒ constant

$$\underset{\downarrow\; w=0}{H(e^{j\omega})} = \underset{\downarrow\; w=0}{\sum h[n]\, e^{j\omega n}}$$

2- $R_y[n,m] \overset{?}{=} f(n-m)$

Step 1:

$$R_{xy}[n,m] = E\{x[n]\overbrace{\sum h^*[k]\,x^*[m-k]}^{y^*[m]}\}$$

$$= \sum_{k=-\infty}^{\infty} h^*[k]\, R_x[n, m-k]$$

↓ x: WSS

$$= \sum_{k=-\infty}^{\infty} h^*[k]\, r_x(n-m+k)$$

↓ $l = -k$

$$= \sum_{l=-\infty}^{\infty} h^*[-l]\, r_x(n-m-l)$$

* lag = k

$$= r_x(k) * h^*[-k]$$

Step 2:

$$R_y[n,\overset{m}{\boxed{n-k}}] = E\{y[n]\,y^*[n-k]\}$$

$$= E\{(\sum_{t=-\infty}^{\infty} h[t]\,x[n-t'])\,y^*[n-k]\}$$

$$= \sum_{t'=-\infty}^{\infty} h[t']\,\overbrace{R_{xy}(n-t', n-k)}^{r_{xy}(k-t')} = h[k] * r_{xy}(k)$$

y[n] is also WSS!

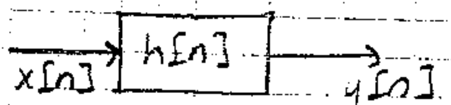
x[n] and y[n] are jointly WSS!

$r_y(k) = r_x(k) * h(k) * h^*(-k)$

Power Spectral Density

$S_x(e^{j\omega}) = DTFT\{r_x[k]\}$

$$= \sum_{k=-\infty}^{\infty} r_x[k]\, e^{-j\omega k} \qquad 0 \leqslant \omega \leqslant 2\pi$$



WSS

$r_x(k)$

$\updownarrow$ DTFT

$S_x(e^{j\omega})$

"Power Spectral Density"

WSS

$r_y(k)$

DTFT

$$S_y(e^{j\omega}) = DTFT\{r_x(k) * h^*(-k) * h(k)\}$$
$$= S_x(e^{j\omega}) \underbrace{DTFT\{h^*(-k)\}}_{H^*(e^{j\omega})} H(e^{j\omega})$$
$$= S_x(e^{j\omega}) |H(e^{j\omega})|^2$$

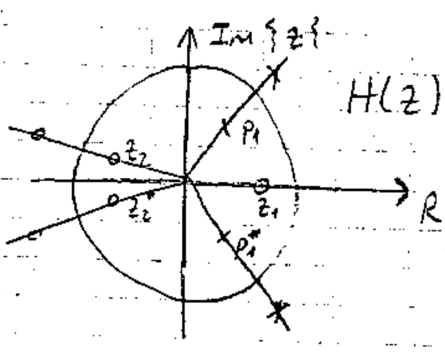
Since

$$H^*(e^{j\omega}) = \sum_{n=-\infty}^{\infty} h^*[n]\, e^{j\omega n} = \sum_{m=-\infty}^{\infty} h^*[-m]\, e^{-j\omega m} = DTFT\{\underline{h^*[-n]}\}$$

$$r_y(k) = ? \leftarrow (DTFT)^{-1}\{ S_x(e^{j\omega}) |H(e^{j\omega})|^2 \}$$

$$r_y(k) \xleftrightarrow[z]{} S_y(z) \xrightarrow[z=e^{j\omega}]{} S_y(e^{j\omega}) \qquad S_y(z) = \sum_{n=-\infty}^{\infty} r_y[n]\, z^{-n}$$

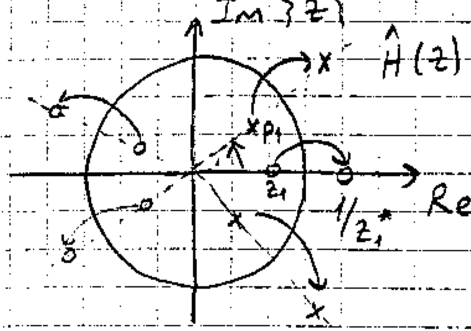
$$r_y(k) \xrightarrow[z]{} S_y(z) = S_x(z) \underbrace{\boxed{H(z) H^*(1/z^*)}}_{H_{combi}(z)}$$



$H(z)$ $\quad H(z_1) = 0$

Blue denotes poles and zeros of $H_{combi}(z)$.

poles-zeros appear in conjugate reciprocal pairs.



$$\hat{H}(z) = H^*(1/z^*)$$

$$\hat{H}(1/z_1^*) = H^*(z_1)$$
$$= 0^* = 0$$

$$p_1 \rightarrow |p_1| e^{j\angle p_1}$$

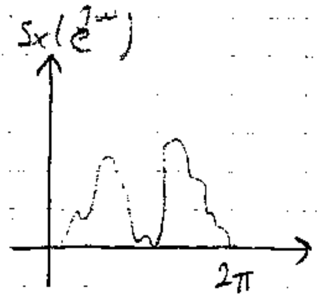
$$\frac{1}{p_1^*} \longrightarrow \frac{1}{|p_1| e^{-j\angle p_1}} = \frac{1}{|p_1|} e^{j\angle p_1}$$

## Power Spectral Density

Wiener-Klein Theorem

$$\underbrace{S_x(e^{j\omega})}_{\text{power spectral density}} = \underbrace{\sum_{k=-\infty}^{\infty} r_x[k] e^{-j\omega k}}_{DTFT\{r[k]\}}$$

what is $S_x(e^{j\omega})$ physically?



Some properties

1- $S_x(e^{j\omega})$ is real valued (since $r_x(k) = r^*(-k)$)

2- $S_x(e^{j\omega})$ is positive (more on this later)

3- $\underline{\frac{\text{Area under } S_x(e^{j\omega})}{2\pi}} = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(e^{j\omega})\,d\omega \overset{?}{=} r_x[0]$
$= \underset{\substack{\uparrow \\ \text{power "}x[n]\text{"}}}{E\{x^2[n]\}}$

(Since $r_x(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(e^{j\omega}) e^{j\omega k}\,d\omega$

$$\begin{tikzcd} x[n] \arrow[r] & \boxed{h[n]} \arrow[r] & y[n] \end{tikzcd} \quad S_y(e^{j\omega}) = |H(e^{j\omega})|^2 S_x(e^{j\omega})$$

$$\begin{tikzcd} WSS \arrow[rr] \arrow[d, leftrightarrow] & & WSS \arrow[d, leftrightarrow] \\ r_x[k] & & r_y[k] \\ S_x(e^{j\omega}) \arrow[rr] & & S_y(e^{j\omega}) \end{tikzcd}$$

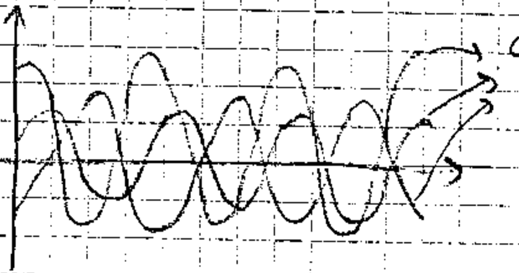
$j(\omega_0 n + \phi$

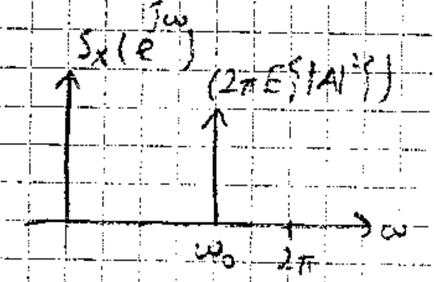
Ex: $x[n] = \underset{\sim}{A} e^{j\omega_0 n + \underset{\sim}{\phi}}$ $\quad A$: a r.v

$\phi$: unif $[0, 2\pi]$

$r_x(k) = E\{|A|^2\} e^{j\omega_0 k} \xleftrightarrow[DTFT]{} S_x(e^{j\omega}) = 2\pi E\{|A|^2\} \delta(\omega - \omega_0)$

$\searrow E\{x[n] x^*[n-k]\}$

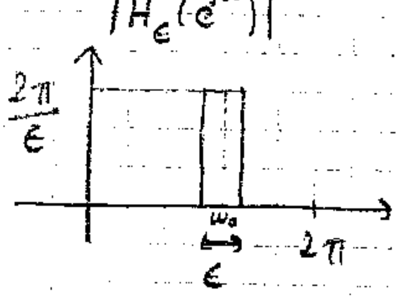


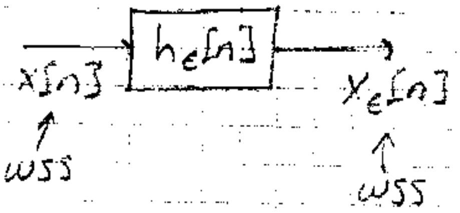


Power is concentrated at a single frequency $\omega_0$

$x[n]$: $\begin{pmatrix} \text{random amplitude} \\ \text{random initial phase} \end{pmatrix}$ complex expo. with freq. $\omega_0$ (rad/sample)

Definition: Let $h_\epsilon[n]$ be the impulse response of a filter whose freq. response is





$S_x(e^{j\omega}) = \lim_{\epsilon \to 0} \text{var}(x_\epsilon[n])$

p.s.d

$x_\epsilon[n] = x[n] * h_\epsilon[n]$

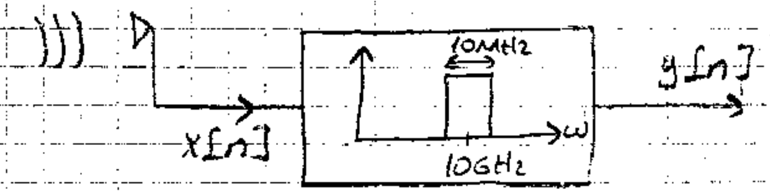
$r_{x_\epsilon}(k) = x[n] * h_\epsilon[n] * h_\epsilon[-n]$

$S_{x_\epsilon}(e^{j\omega}) = S_x(e^{j\omega}) |H_\epsilon(e^{j\omega})|^2$

$$\mathrm{Var}(x_\epsilon[n]) = r_{x_\epsilon}(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_{x_\epsilon}(e^{j\omega})\,d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_x(e^{j\omega})\,|H_\epsilon(e^{j\omega})|^2\,d\omega$$

$$= \frac{1}{2\pi}\int_{|\omega-\omega_0|<\frac{\epsilon}{2}} S_x(e^{j\omega})\,\frac{2\pi}{\epsilon}\,d\omega \cong S_x(e^{j\omega_0})$$



$$\hat{r}_y(0) = \left(\frac{1}{N}\sum_{n=1}^{N} |y[n]|^2\right)$$ estimated output average power

Some Important Facts:

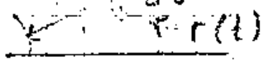
$$\underset{\substack{\text{is}\\ \text{a valid}\\ \text{auto-corr seq.}}}{r_x(k)} \xleftrightarrow{\text{iff}} \underset{\forall N}{\underline{\underline{R}}_{x\,(N\times N)}} \geq 0 \xleftrightarrow{\text{iff}} \underset{\substack{\uparrow\\ \mathrm{DTFT}\{r_x(k)\}\geq 0}}{S_x(e^{j\omega})} \geq 0 \quad \forall\omega$$

$$\underset{\substack{\uparrow\\ E\{(x[n])^2\}\geq 0}}{r_x(0)} = -1$$

Q: Given a positive valued function, can I always construct a random process whose psd is that function?

Papoulis 10.24 p.322 → Ex: $s(t) = a e^{j\omega_0\left(t - \frac{r(t)}{c}\right)}$

↑ signal received by "tower"



$r(t) = r_0 + \underset{\sim}{v}t$

c: speed of light

$f_{\underset{\sim}{v}}(v)$: density of velocity is given.

Find P.S.D of s(t)

$$r_s(\tau) = E\{s(t)\,s^*(t-\tau)\} = a^2 E\{e^{j\omega_0\tau} \cdot e^{-j\omega_0\left(\frac{\overbrace{r(t)-r(t-\tau)}^{\underset{\sim}{v}\tau}}{c}\right)}\}$$

$$= a^2 E_{\underset{\sim}{v}}\{e^{j\omega_0\tau(1-\frac{v}{c})}\}$$

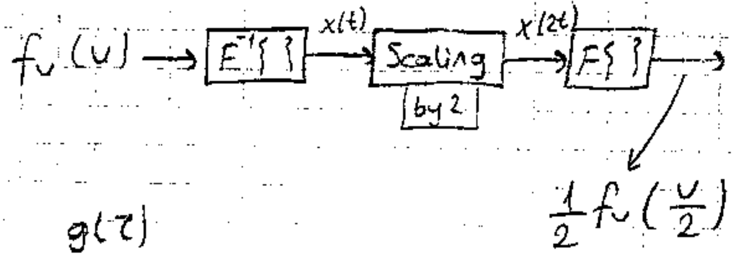
$$= a^2 E\{e^{j\omega_0\tau} \cdot e^{-j\frac{\omega_0\tau}{c}\cdot\underset{\sim}{v}}\}$$

$$= a^2 e^{j\omega_0\tau} E\{e^{-j\frac{\omega_0\tau}{c}\underset{\sim}{v}}\} = a^2 e^{j\omega_0\tau}\left(\int_{-\infty}^{\infty} e^{j(-\frac{\omega_0\tau}{c})\underset{\sim}{v}} f_{\underset{\sim}{v}}(v)\,dv\right)$$

$$= a^2 e^{j\omega_0\tau} F^{-1}\{f_v(v)\}\left(\frac{-\omega_0\tau}{c}\right)$$

$$S_s(j\omega) = F\{r_s(\tau)\}$$

$$\underbrace{\underset{\tau\to\omega}{F}\{F^{-1}\{f_v\}\left(\frac{-\omega_0\tau}{c}\right)\}(\omega)}_{}$$



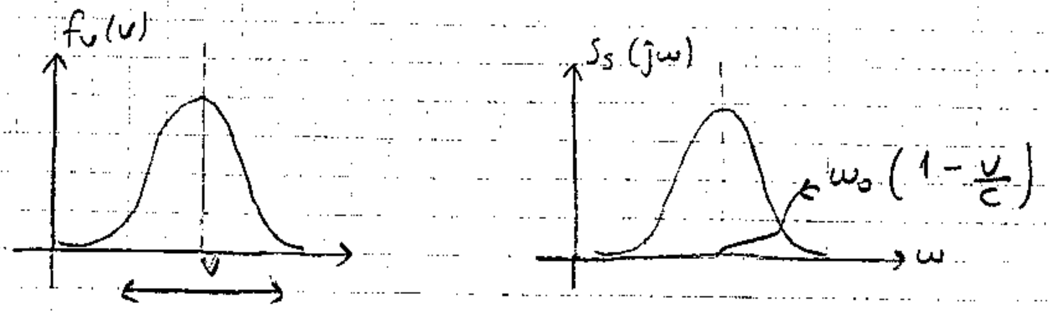
$$\text{Note: } \underset{\tau\to\omega}{F}\{\overbrace{F^{-1}\{f_v\}\left(-\frac{\omega_0\tau}{c}\right)}^{g(\tau)}\}(\omega)$$

$$\frac{c}{\omega_0} f\left(-\omega\frac{c}{\omega_0}\right)$$

$$S_s(j\omega) = F\{r_s(\tau)\}$$

$$= \underset{\tau\to\omega}{F}\{a^2 e^{j\omega_0\tau} g(\tau)\}$$

$$= a^2 G(j(\omega-\omega_0)) = \frac{a^2 c}{\omega_0} f_v\left((\omega_0-\omega)\frac{c}{\omega_0}\right)$$



Doppler effect

$$f_v\left(\overbrace{(\omega_0-\omega)\frac{c}{\omega_0}}^{\bar{v}}\right) \qquad (\omega_0-\omega)\frac{c}{\omega_0}=v$$

$$c-\frac{\omega c}{\omega_0}=v$$

$$c-v=\frac{\omega}{\omega_0}c \rightarrow \omega=\omega_0\left(1-\frac{v}{c}\right)$$

Ex: Hayes 3.4.1

$$x[n]\rightarrow\boxed{h[n]}\rightarrow y[n] \qquad H(z)=\frac{1}{1-\frac{1}{4}z^{-1}} \quad z\in ROC$$

x[n]: white noise, unit variance $r_w[k]=\delta[k]$

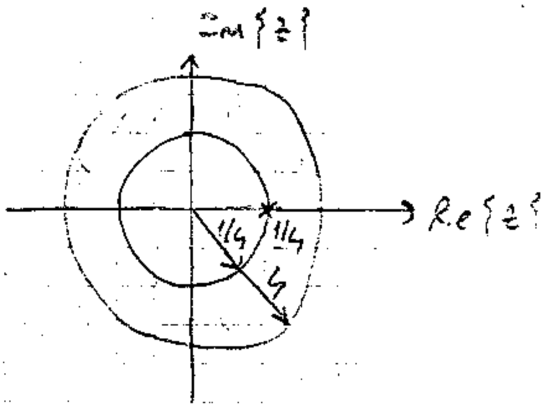
Find $r_y[k]$

$$r_y[k]=r_x[k]*h[n]*h[-n]$$

$$S_y(z)=\underset{1}{S_x(z)}\cdot H(z)\,H^*(1/z^*)$$

$$S_y(z)=\underset{\substack{ROC\\|z|>1/4}}{H(z)}\,\underset{\substack{ROC\\|z|<4}}{H(1/z)}=\frac{1}{1-\frac{1}{4}z^{-1}}\cdot\frac{1}{1-\frac{1}{4}z}$$

$$r_y[k]=\mathcal{Z}^{-1}\left\{\frac{1}{1-\frac{1}{4}z^{-1}}\cdot\frac{1}{1-\frac{1}{4}z}\right\}$$



h[n]: causal filter

(ROC extends up to ∞
ROC $|z|>1/4$)

$$r_y[k]=\mathcal{Z}^{-1}\left\{\frac{1}{1-\frac{1}{4}z^{-1}}\cdot\frac{1}{1-\frac{1}{4}z}\right\} \qquad \begin{matrix}ROC\\ \frac{1}{4}<|z|<4\end{matrix}$$

$$=\mathcal{Z}^{-1}\left\{\frac{16/15}{1-\frac{1}{4}z^{-1}}+\frac{4/15}{z^{-1}-1/4z}\right\}=\underbrace{\mathcal{Z}^{-1}\left\{\frac{16/15}{1-\frac{1}{4}z^{-1}}\right\}}_{ROC\ |z|>1/4}+\underbrace{\mathcal{Z}^{-1}\left\{\frac{4/15}{z^{-1}-1/4}\right\}}_{ROC\ |z|<4}$$

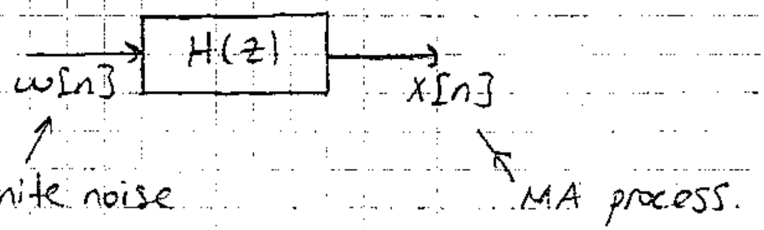
$$\left(\frac{1}{4}\right)^k \cdot \frac{16}{15} u(k) + \frac{16}{15} 4^k u(-k-1)$$

$$= \frac{16}{15} \frac{1}{4}^{|k|}$$

Hayes:

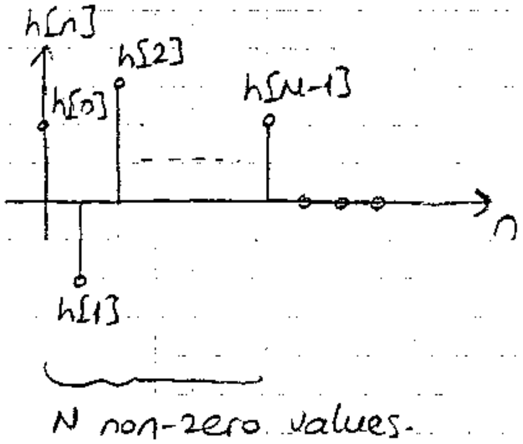
Types of Random Processes

1- MA Process (Moving Average)



H(z): All zero filter (no poles)

H(z): N-tap Causal filter.



$$x[n] = \sum_{k=0}^{N-1} w[n] h[n-k]$$

$$= w[n] h[0] + w[n-1] h[1]$$

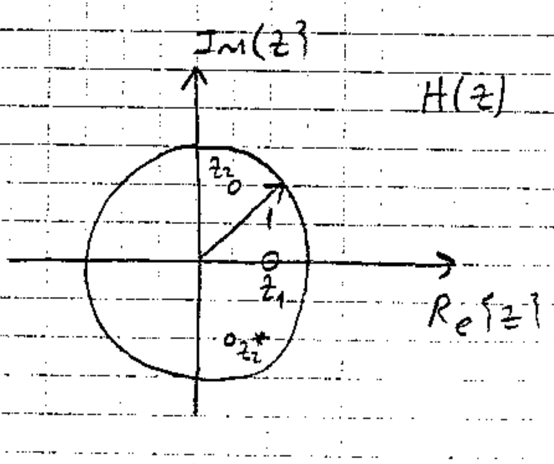
$$+ w[n-2] h[2]$$

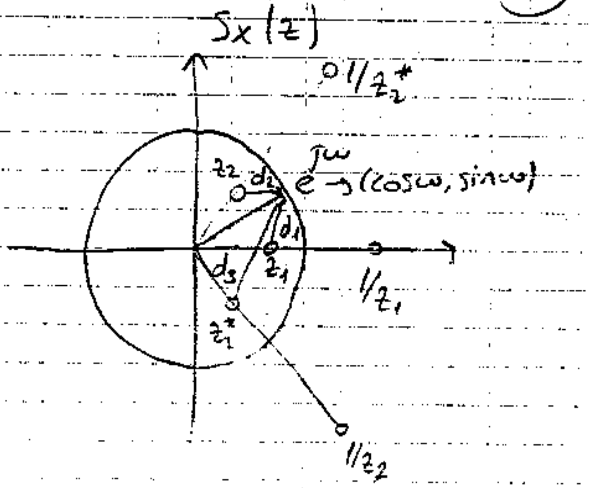
$$+$$

$$+ w[n-(N-1)] h[N-1]$$

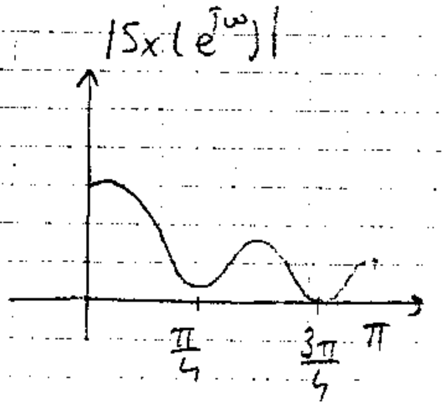
## $S_x(e^{j\omega})$ : Power Spectral Density of MA

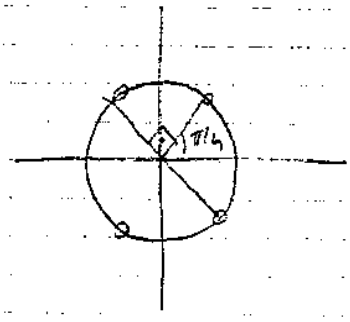
$$S_x(e^{j\omega}) = |H(e^{j\omega})|^2 \, \sigma_w^2 \quad \text{or} \quad S_x(z) = H(z)\, H^*(1/z^*)\, \sigma_w^2$$





$$H(z) = K \prod_{k=1}^{N} (z - z_k)$$

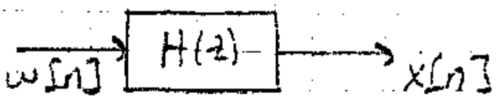




MA process do not have peaky power spectrum density; but on the contrary the p.s.d is broad and possibly have nulls.

Auto-correlation of MA process:

$H(z) = h_0 + h_1 z^{-1} + h_2 z^{-2}$



$r_x(k) = (h(k) * h(-k))\, \sigma_w^2$

$$r_x(k) = \sigma_\omega^2 \sum_{\ell=-\infty}^{\infty} h(-\ell)h(k-\ell) = \sigma_\omega^2 \sum_{\ell=-\infty}^{\infty} h(\ell)h(\ell+k)$$

Since

$$r_x(k) = r_x(-k) = \sigma_\omega^2 \underbrace{\sum_{\ell=-\infty}^{\infty} h(\ell)h(\ell-k)}_{\text{deterministic auto-correlation!}}$$

$$r_x(k) = \sigma_\omega^2 \sum_{\ell=0}^{2} h(\ell)h(\ell-k)$$

$r_x(0) = \sigma_\omega^2 (h_0^2 + h_1^2 + h_2^2)$

$r_x(1) = \sigma_\omega^2 (h_1 h_0 + h_2 h_1)$ (Since $h[-1] = 0$)

$r_x(2) = \sigma_\omega^2 (h_2 h_0)$

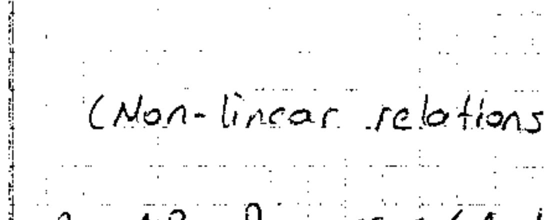
$r_x(3) = 0$

$h[n] = [\overset{\downarrow n=0}{h_0} \quad h_1 \quad h_2 \quad 0 \quad 0 \quad 0]$

$h[n] = [h_0 \quad h_1 \quad h_2 \quad 0 \quad 0 \quad 0]$

$h[n-1] = [0 \quad h_0 \quad h_1 \quad h_2 \quad 0]$

$h[n-2] = [0 \quad 0 \quad h_0 \quad h_1 \quad h_2]$

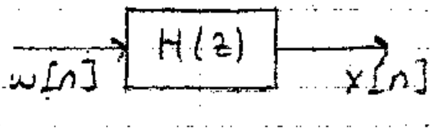


Given $r_x(k)$; can I find a filter excited with noise whose auto-correlation is the given $r_x(k)$?

Not easy for MA filters.

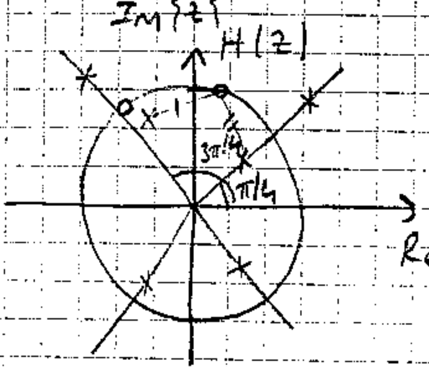
(Non-linear relationship between $r_x(k)$ and filter coef.)

## 2- AR Processes (Auto-regressive procession)

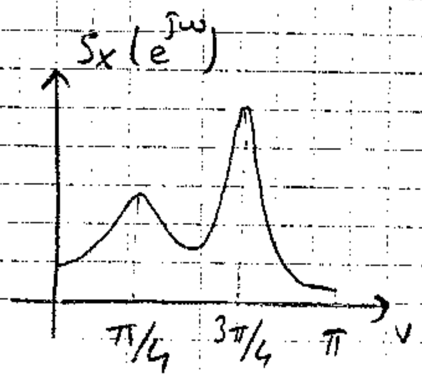


$H(z)$: All pole filter.

$$H(z) = \frac{K}{\prod_{k=1}^{N} (z - p_k)}$$

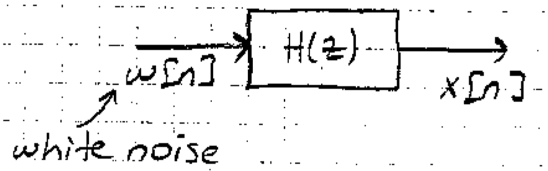


x: pole



$$S_x(z) = H(z)\, H^*(1/z^*)\, \sigma_\omega^2$$

AR process: peaky response ("resonounce type")



H(z): All-pole

$$H(z) = \frac{b_0}{1 + \sum_{k=1}^{p} a_k z^{-k}}$$

← Causal system IIR

↗ p poles

Auto-correlation Calculation:

$$r_x(k) = h(k) * h(-k) * \underset{\sigma_\omega^2 \delta[k]}{r_\omega(k)} \longrightarrow \sigma_\omega^2\, h(k) * h(-k)$$

$$\frac{X(z)}{\omega(z)} = \frac{b_0}{1 + \sum_{k=1}^{p} a_k z^{k}} \xrightarrow[z^{-1}\{.\}]{} x[n] + \sum_{k'=1}^{p} a_k\, x[n-k'] = b_0\, \omega[n]$$

$$\underset{(k \geq 0)}{r_x(k)} = E\{x[n]\, x^*[n-k]\} \longrightarrow r_x(k) + \sum_{k'=1}^{p} a_k\, r_x(k-k') = b_0 E\{\omega[n]\, x^*[n-k]\}$$

$$= \begin{cases} \overset{\text{causal system}}{0} & k > 0 \\ b_0 E\{\omega[n]\, x^*[n]\} & k = 0 \end{cases}$$

$$b_0 \overset{b_0 \sigma_\omega}{\boxed{E\{\omega[n]\, x^*[n]\}}} = |b_0|^2 \sigma_\omega^2$$

$$\uparrow \left(b_0 \omega[n] - \sum_{k'=1}^{p} a_k x[n-k']\right)^*$$

$$r_x(k) + \sum_{k'=1}^{p} a_k r_x(k-k') = |b_0|^2 \sigma_\omega^2 \delta[k] \quad \forall k$$

Then

$$\begin{array}{l} k=0 \rightarrow \\ k=1 \rightarrow \\ k=2 \rightarrow \\ \\ k=p \rightarrow \\ k=p+1 \rightarrow \end{array}
\left[\begin{array}{ccccc} r_x(0) & r_x(-1) & r_x(-2) & \cdots & r_x(-p) \\ r_x(1) & r_x(0) & r_x(-1) & \cdots & r_x(-p+1) \\ r_x(2) & r_x(1) & r_x(1) & \cdots & r_x(-p+2) \\ \vdots & & & & \\ r_x(p) & r_x(p-1) & r_x(p-2) & \cdots & r_x(0) \\ r_x(p+1) & r_x(p) & r_x(p-1) & \cdots & r_x(1) \end{array}\right]
\begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} |b_0|^2 \sigma_\omega^2 \\ 0 \\ \vdots \\ \\ 0 \\ 0 \end{bmatrix}$$

↗ Yule - Walker Equations

↖ (Yule Walker Matrix: Auto-corr. matrix of $x[n]$ is Hermitian and Toeplitz $\underline{\underline{R}}_x \geq 0$

AR (P): Auto-regressive $p^{th}$ order (P poles)

Ex: For AR(2) process $H(z) = \dfrac{b_0}{1 + a_1 z^{-1} + a_2 z^{-2}}$, Yule-Walker Eqn's.

$$\begin{bmatrix} r_x(0) & r_x(-1) & r_x(-2) \\ r_x(1) & r_x(0) & r_x(1) \\ r_x(-2) & r_x(1) & r_x(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sigma_\omega^2 |b_0|^2 \\ 0 \\ 0 \end{bmatrix}$$

AR Modelling:

1- Assume $r_x(k)$ is given for all $k$, and you want to fit AR(2) model to $r_x(k)$.

$$\begin{bmatrix} \underline{col}_1 & \underline{col}_2 & \underline{col}_3 \\ & & \\ & & \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \end{bmatrix} = \sigma_\omega^2 |b_0|^2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \rightarrow \underline{col}_1 + \begin{bmatrix} \underline{col}_2 & \underline{col}_3 \\ & \\ & \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} |b_0|^2 \\ 0 \\ 0 \end{bmatrix}$$

$\sigma_\omega^2 : 1$: unit var. white noise excites the All pole filter.

$$\left[\begin{array}{cc} r_x(-1) & r_x(-2) \\ \hline r_x(0) & r_x(-1) \\ r_x(1) & r_x(0) \end{array}\right] \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \left[\begin{array}{c} |b_0|^2 - r_x(0) \\ \hline 0 - r_x(1) \\ 0 - r_x(2) \end{array}\right]$$

Then using the last 2 rows (below blue dotted line) $\to a_1, a_2$ can be found

and then substitude the values for $a_1, a_2$ in the first eqn. $\longrightarrow |b_0|^2$ is found.

2- Auto-Correlation finding

Given filter coef $\longrightarrow r_x(k) = ?$

$$\underset{\nearrow \text{Given}}{\begin{bmatrix} 1 & a_1 & a_2 \\ a_1 & 1+a_2 & 0 \\ a_2 & a_1 & 1 \end{bmatrix}} \begin{bmatrix} r_x(0) \\ r_x(1) \\ r_x(2) \end{bmatrix} = \underset{\nwarrow \text{Given}}{\begin{bmatrix} \sigma_\omega^2 |b_0|^2 \\ 0 \\ 0 \end{bmatrix}}$$

Here $r_x(k)$ is assumed to be real.

$$r_x(3) = ? \to r_x(k) + \sum_{k'=1}^{\overset{(p=2 \text{ for } AR(2))}{p}} a_k r_x(k-k') = 0 \qquad k > 0$$

Ex) AR(1) (Causal System) Hayes' Book

$r_x(k) = ?$

$x[n] = ax[n-1] + \omega[n]$ ← unit var. white noise

$x[n-k] \; (k \geq 1)$

$\to r_x(k) = a r_x(k-1) \quad k \geq 1$

$r_x(1) = a r_x(0)$

$r_x(2) = a r_x(1) = a^2 r_x(0)$

$\vdots$

$r_x(k) = a^k r_x(0) \quad k \geq 1$

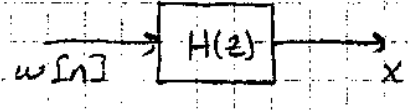
$$(x[n])^2 = (a x[n-1] + w[n])^2$$

$E\{.\}$

$$r_x(0) = a^2 r_x(0) + \sigma_w^2 + 2E\{a \overset{\to 0}{\cancel{x[n-1]w[n]}}\}$$ $x[n]$: causal filter output.

$$r_x(0) = \frac{\sigma_w^2}{1-a^2} \longrightarrow r_x(k) = \frac{\sigma_w^2}{1-a^2} a^{|k|} \quad \forall k$$

ARMA (Auto-Regressive Moving Average) Processes:



$$H(z) = \frac{B_q(z)}{A_p(z)}$$

$$\uparrow$$

$$\frac{X(z)}{W(z)}$$

$B_q(z) = b_q(0) + b_q(1) z^{-1} + \dots + b_q(Q) z^{-q}$ Q: zeros

$A_p(z) = 1 + a_p(1) z^{-1} + \quad + a_p(P) z^{-P}$ P: Poles

$$x[n] + \sum_{l=1}^{P} a_p(l) x[n-l] = \sum_{l=0}^{Q} b_q(l) w[n-l]$$

$k \geq 0$

$$r_x(k) + \sum_{l=1}^{P} a_p(l) r_x(k-l) = \sum_{l=0}^{Q} b_q(l) \underset{\nearrow}{r_{wx}(k-l)}$$

$$r_{wx}(k) = r_w(k) * h^*(-k) = \sigma_w^2 h^*(-k)$$

$$= \sigma_w^2 \sum_{l=0}^{Q} b_q(l) h^*(1-k) \leftarrow f[l]$$

$$= \sigma_w^2 \sum_{l=-k}^{Q-k} b_q(l+k) h^*(l) \leftarrow f[l+k]$$

$$= \sigma_w^2 \sum_{l=-k}^{Q-k} b_q(l+k) h^*(l)$$

$$= c_q(k) = \begin{cases} c_q(k) & k \leq Q \\ 0 & k > Q \end{cases}$$

Hayes' Book

$$r_x(k) + \sum_{l=1}^{p} a_p(l)\, r_x(k-l) = c_q(k)$$

65

$$\begin{array}{l} k=0 \rightarrow \\ k=1 \rightarrow \\ \\ k=q \\ k=q+1 \\ \\ \end{array}
\begin{bmatrix} r_x(0) & r_x(-1) & r_x(-2) & \cdots & r_x(-p) \\ r_x(1) & r_x(0) & r_x(-1) & \cdots & r_x(-p+1) \\ \vdots & & & & \\ r_x(q) & r_x(q-1) & r_x(q-2) & \cdots & r_x(q-p) \\ r_x(q+1) & r_x(q) & r_x(q-1) & \cdots & r_x(q-p+1) \\ \vdots & \vdots & \vdots & & \vdots \end{bmatrix}
\begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \\ \vdots \\ a_p(q) \end{bmatrix} = \begin{bmatrix} c_q(0) \\ c_q(1) \\ c_q(2) \\ \vdots \\ c_q(q) \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \end{bmatrix}$$

ARMA Process

Difficult to synthesize a filter for a given ARMA $r_x(k)$

4- Periodic Processes

Remember $|r_x(k)| \leqslant r_x(0)$ $\quad , E\{\underset{z}{x[n]}\,\underset{w}{x[n-k]}\}$ | $|\rho_{zw}| < 1$

$E\{zw\} \leqslant \sigma_z \sigma_w$

$r_x(k) \leqslant \sqrt{E\{x^2[n]\}}\sqrt{E\{x^2[n-k]\}}$

$\leqslant r_x(0)$

$$|r_x(k)| = \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(e^{j\omega})\, e^{j\omega k}\, d\omega \right|$$

$$\leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} |S_x(e^{j\omega})\, e^{j\omega k}|\, d\omega$$

$$\left( \frac{1}{2\pi} \int_{-\pi}^{\pi} S_x(e^{j\omega})\, d\omega \right) \rightarrow r_x(0)$$

MS Periodic: (Mean-Square Periodic)

$E\{(x[n+T] - x[n])^2\} = 0 \qquad \forall n,\ \exists T$

↳ period

Claim: If $r_x(k) = r_x(k+T)$ $\begin{pmatrix} \text{for some } T \\ \forall k \end{pmatrix}$ ; $r_x(k)$ is a periodic sequence of k.

$r_x(0) = r_x(T)$  $k=0$

Proof:

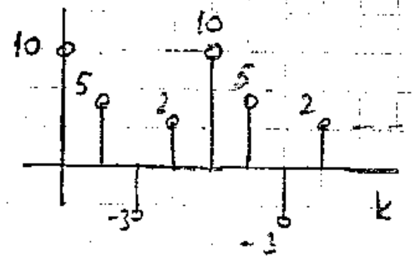
$$(E\{zw\})^2 \leqslant E\{z^2\}\, E\{w^2\} \qquad |\rho_{zw}| = 1$$

$z = x[n+k+T] - x[n+k]$

$w = x[n]$

$$\left(r_x(k+T) - r_x(k)\right)^2 \leqslant \left(2r_x(0) - 2r_x(T)\right) r_x(0)$$

If $r_x(0) = r_x(T) \rightarrow r_x(k+T) = r_x(k)$ $\forall k$



$R_x$ : Auto-corr. matrix for periodic process.

$R_x \geqslant 0$ in general

T: Period of Period Process $\rightarrow T = 3$

are identical for MS Periodic Process

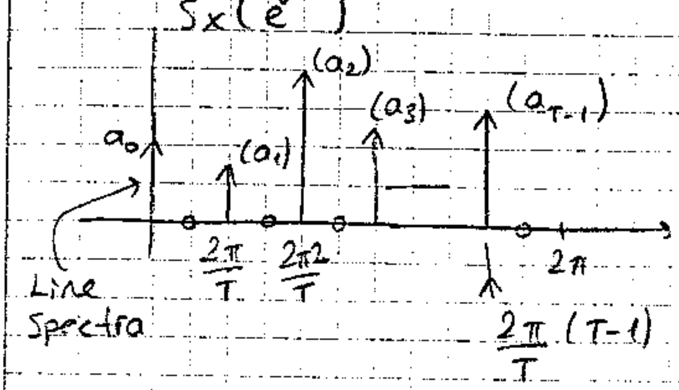
$$R_x = \begin{bmatrix} r_x(0) & r_x(-1) & r_x(-2) & r_x(-3) \\ r_x(1) & r_x(0) & r_x(-1) & r_x(-2) \\ r_x(2) & r_x(1) & r_x(0) & r_x(-1) \\ r_x(3) & r_x(2) & r_x(1) & r_x(0) \end{bmatrix}_{4\times4} \quad (T=3)$$

$|R_x| = 0$   $(R \not> 0)$   $(R_x \geqslant 0)$

$$r_x(k) = \sum_{l=0}^{T-1} a_k e^{j\frac{2\pi}{T} lk} \longrightarrow S_x(e^{j\omega}) = \sum_{l=0}^{T-1} a_k \delta\left(\omega - \frac{2\pi}{T} l\right)$$

$\uparrow$

( Periodic by $T$ )



Ex: $x[n] = \sum_{l=1}^{L} \underset{\sim}{a_l}\, e^{j\omega_l n + \underset{\sim}{\theta_l}}$

$a_l$ and $\theta_l$'s are independent, $\theta_{l_1}$ and $\theta_{l_2}$ are ind. $l_1 \neq l_2$;

$\theta_{l_1} = \text{unif}\,[0, 2\pi]$

$$r_x(k) = E\left\{\sum_{l_1}\sum_{l_2} a_{l_1} a_{l_2}^* \underbrace{\left(e^{j\omega_{l_1} n + \theta_{l_1}}\right)\left(e^{-j\omega_{l_2}(n-k) - \theta_{l_2}}\right)}_{e^{j(\omega_{l_1} - \omega_{l_2})n} \,.\, e^{j\omega_{l_2} k} \,.\, e^{j(\theta_{l_1} - \theta_{l_2})}}\right\}$$

$$= \sum_{l_1, l_2} e^{j(\omega_l - \omega_{l_2})n} \,.\, e^{j\omega_{l_2} k}\, E\{a_{l_1} a_{l_2}^*\}\, \underbrace{E\{e^{j(\theta_{l_1} - \theta_{l_2})}\}}_{\delta[l_1 - l_2]}$$

$$= \sum_{l_1} e^{j\omega_{l_2} k} \,.\, E\{|a_{l_1}|^2\}$$

## Spectral Factorization

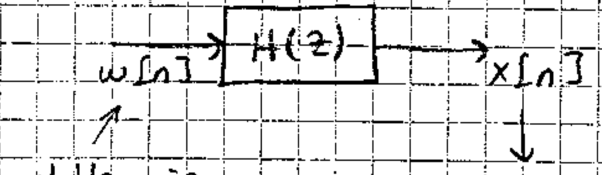
Given a random process whose Power Spectral Density can be written as ratio of trigonometric func. such as

$$P_x(e^{j\omega}) = \frac{5 - 4\cos\omega}{10 - 6\cos\omega};$$

find a $H(z)$ [LTI filter] excited with noise input generating the process.

1-

Process x[n] generation



Synthesis

white noise

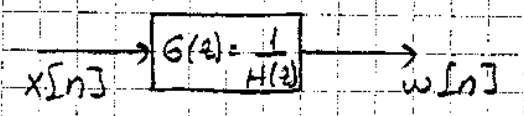
$P_w(z)=\sigma_w^2=1$

$$P_x(z)=H(z)\,H^*(1/z^*)\,\underset{1}{\sigma_w^2}$$

2-

Whitening

G(z): whitening filter.



$P_x(z)$ ← given

white

Ex: $P_x(e^{j\omega}) = \frac{5-4\cos\omega}{10-6\cos\omega}$

1- Find a synthesis filter generating x[n].

2- Find the whitening filter.

Method: $P_x(z)$ is factorized into $H(z).H^*(1/z^*)$

$$P_x(e^{j\omega}) = \frac{5-4\,\frac{e^{j\omega}+e^{-j\omega}}{2}}{10-6\,\frac{e^{j\omega}+e^{-j\omega}}{2}} = \frac{5-2z-2z^{-1}}{10-3z-3z^{-1}}$$

$\swarrow$ $P_x(z)$ $\quad z=e^{j\omega}$

$$P_x(z) = \frac{5-2z-2z^{-1}}{10-3z-3z^{-1}}\cdot\frac{-z}{-z}$$

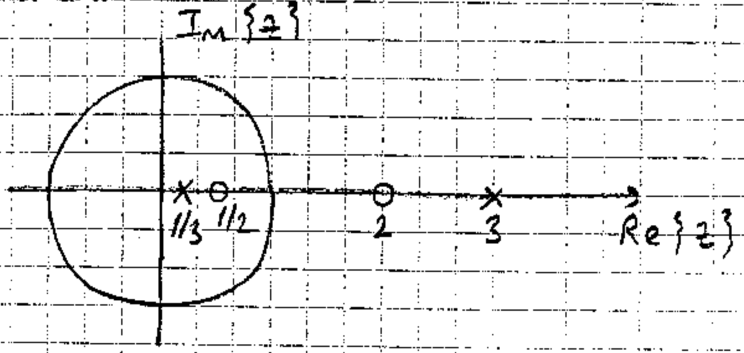
$$= \frac{2z^2-5z+2}{3z^2-10z+3}$$

$$= \frac{(2z-1)(z-2)}{(3z-1)(z-3)}\,\frac{-z^{-1}}{-z^{-1}}$$

$$= \frac{(2z-1)\,(2z^{-1}-1)}{(3z-1)\,(3z^{-1}-1)}$$

← 1- Synthesis $P_x(z)=H(z)H(1/z)$

$$H(z)=\left\{\frac{2z-1}{3z-1},\ \frac{2z-1}{3z^{-1}-1},\ \frac{2z^{-1}-1}{3z-1},\ \frac{2z^{-1}-1}{3z^{-1}-1}\right\}$$



$x[n] = \frac{1}{n!} \quad n \geq 0$

$$X(z) = \sum_{n=0}^{\infty} \frac{1}{n!} z^{-n}$$

$$= \sum_{n=0}^{\infty} \frac{(1/z)^n}{n!} \qquad e^{w} = \sum_{n=0}^{\infty} \frac{w^n}{n!}$$

$$= e^{1/z}$$

causal, stable and causally invertible (min. phase)

causal and stable

$$H(z) = \left\{ \frac{2z-1}{3z-1}, \frac{2z-1}{3z^{-1}-1}, \frac{2z^{-1}-1}{3z-1}, \frac{2z^{-1}-1}{3z^{-1}-1} \right. ?$$

can not be both causal and stable

min. phase system:

If H(z) has all poles and zeros inside unit circle → min. phase system (Oppenheim's Book)

Comment: Whenever I have R.P whose Power Spectral Density can be factorized (Spectral factorization); I can always apply a whitening filter and use its output without any loss of information due to causally invertible $H(z), \frac{1}{H(z)}$

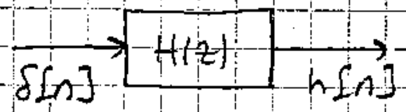
$$H(z) = \frac{2z-1}{3z-1} \qquad \frac{1}{H(z)} = \frac{3z-1}{2z-1}$$

again both poles, zeros inside unit circle.

# Signal Modeling (Hayes)

## Deterministic Signal Modeling

Problem: Given a time sequence $x[n]$, how can I find a LTI filter whose impulse resp. $h[n]$ approximates $x[n]$?



$h[n] \approx x[n]$ ← given

Ex: $H(z) = \dfrac{b_0}{1-a_1 z^{-1}} \longrightarrow h[n] = b_0 a_1^n u[n]$

given $x[0], x[1] \rightarrow x[0] = h[0] = b_0$, $x[1] = h[1] = b_0 a_1$ } Equate first two samples using $(b_0, a_1)$, then hope that the rest is OK.

Ex: $H(z) = \dfrac{b_0 + b_1 z^{-1}}{1-a_1 z^{-1}} \longrightarrow h[n] = b_0 a_1^n u[n] + b_1 a_1^{n-1} u[n-1]$

Set $(b_0, b_1, a_1)$ s.t. $h[n] = x[n]$ $n = \{0, 1, 2\}$

$h[0] = b_0 \leftarrow x[0]$

$h[1] = b_0 a_1 + b_1 \leftarrow x[1]$

$h[2] = b_0 a_1^2 + b_1 a_1 \leftarrow x[2]$

## Pade's Approximation

$$H(z) = \frac{B_q(z)}{A_p(z)} = \frac{\sum_{k=0}^{q} b_q(k) z^{-k}}{1 + \sum_{k=1}^{p} a_p(k) z^{-k}} = X(z)$$

$$\left(1 + \sum_{k=1}^{p} a_p(k) z^{-k}\right) X(z) = \sum_{k=0}^{q} b_q(k) z^{-k}$$

$$z^{-1}\{.\} \rightarrow x[n] * [1 \;\; a_p(1) \;\; a_p(2) \; \dots \; a_p(p)] = [b_q(0) \;\; b_q(1) \; \dots \; b_q(q) \; 00000]$$

$$x[n] + \sum_{k=1}^{p} a_p(k)\, x[n-k] = b_q[n]$$

$$\begin{array}{l} n=0 \rightarrow \\ n=1 \rightarrow \\ \\ \\ n=q \rightarrow \\ \\ \\ \\ \end{array}
\begin{bmatrix} x(0) & 0 & 0 & \cdots & 0 \\ x(1) & x(0) & 0 & \cdots & 0 \\ x(2) & x(1) & x(0) & \cdots & 0 \\ \vdots & & & & \\ x(q) & x(q-1) & x(q-2) & \cdots & \\ x(q+1) & x(q) & x(q-1) & \cdots & x(q-p) \\ x(q+2) & x(q+1) & x(q) & \cdots & x(q-p+1) \\ \vdots & & & & \end{bmatrix}
\begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \\ \vdots \\ a_p(p) \end{bmatrix} = \begin{bmatrix} b_q(0) \\ b_q(1) \\ \vdots \\ b_q(q) \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Using the bottom part of matrix, I can write.

$$\underbrace{\begin{bmatrix} x(q) & x(q-1) & \cdots & x(q-p+1) \\ x(q+1) & x(q) & & x(q-p+2) \\ x(q+2) & & & \\ & & & \\ x(q+p-1) & & & \end{bmatrix}}_{p\times p}
\underbrace{\begin{bmatrix} a_p(1) \\ \vdots \\ \\ a_p(p) \end{bmatrix}}_{p\times 1} = \underbrace{\begin{bmatrix} -x(q+1) \\ -x(q+2) \\ -x(q+3) \\ \vdots \\ -x(q+p) \end{bmatrix}}_{p\times 1}$$

Pade: Step1: Using bottom part of conv. matrix relation
Find $a_k$'s.
Step2: Insert $a_k$'s in the top part and get $b_k$'s.

Ex: $H(z) = \dfrac{b_0}{1+a_1 z^{-1}+a_2 z^{-2}}$ $\quad p=2 \quad q=0$

$$\begin{bmatrix} x(0) & & \\ x(1) & x(0) & 0 \\ x(2) & x(1) & x(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \end{bmatrix} = \begin{bmatrix} b_q(0) \\ 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} x(0) & 0 \\ x(1) & x(0) \end{bmatrix} \begin{bmatrix} a_p(1) \\ a_p(2) \end{bmatrix} = -\begin{bmatrix} x(1) \\ x(2) \end{bmatrix}$$

$$a_p(1) = -x(1)/x(0)$$

$$a_p(2) = \frac{-x(2)+x^2(1)/x(0)}{x(0)}$$

Warning: Pade matches first $(q+P+1)$ samples using

$\{b_0, \_\_, b_q\}$ and

$\{a_1, \_\_, a_p\}$.

But, does not say anything for other samples. The performance can be poor and even designed $H(z)$ can be unstable.

$$\begin{bmatrix} x(0) & & & & \\ x(1) & x(0) & & & \\ x(2) & x(1) & x(0) & & \\ \vdots & & & & \\ x(q) & x(q-1) & x(q-2) & & \\ x(q+1) & x(q) & x(q-1) & \dots & x(q-P+1) \\ x(q+2) & x(q+1) & \_\_ & \dots & x(q-P+2) \\ x(q+3) & x(q+2 & & & \\ x(q+4) & & & & \\ \vdots & & & & \\ x(N) & x(N-1) & & \dots & x(N-P) \end{bmatrix} \begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \\ \vdots \\ a_p(p) \end{bmatrix} = \begin{bmatrix} b_q(0) \\ b_q(1) \\ \vdots \\ b_q(q) \\ 0 \\ 0 \\ 0 \\ \\ \end{bmatrix}$$

Prony Step 1: Use the bottom part of the matrix to find the best $a_p(k)$'s in the LS sense.

bottom part of conv. matrix

$$\underline{\underline{X}}_{bot} \begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \\ \vdots \\ a_p(P) \end{bmatrix} = \underline{0} \qquad [\underset{\text{first column}}{\underline{X}_1} \quad \underline{\underline{X}}_{other}] \begin{bmatrix} 1 \\ a_p(1) \\ \vdots \\ a_p(P) \end{bmatrix} = \underline{0}$$

$$\underline{\underline{X}}_{other} \, \underline{a}_p = -\underline{X}_1$$

$$\underline{\hat{a}}_p^{LS} = (X_{other}^T \underline{\underline{X}}_{other})^{-1} \underline{\underline{X}}_{other}^T X_1$$

Step 2: Insert $\hat{a}_p^{LS}$ in the top part and find $b_\varphi(k)$.

## Deterministic Signal Modeling

Problem:

$x[n]$ : causal

$x[n] \approx h[n] \leftarrow$ LTI filter impulse response.

$$\downarrow$$

$$H(z) = \frac{B_\varphi(z)}{A_p(z)} = \frac{b_\varphi(0) + b_\varphi(1) z^{-1} + \dots + b_\varphi(\varphi) z^{-\varphi}}{1 + a_p(1) z^{-1} + \dots + a_p(p) z^{-p}}$$

Prony: (P. 144 Hayes)

$$e'[n] = x[n] - h[n] \longrightarrow E'(z) = X(z) - H(z) = X(z) - \frac{B_\varphi(z)}{A_p(z)}$$

$$\downarrow$$

$$\underbrace{E'(z) A_p(z)}_{} = X(z) A_p(z) - B_\varphi(z)$$

$$E(z) = X(z) A_p(z) - B_\varphi(z)$$

Prony Error

$$e[n] = x[n] * a_p[n] - b_\varphi[n]$$

$$e[n] = \begin{cases} x[n] + \sum_{l=1}^{p} a_p[l] x[n-l] - b_\varphi[n], & 0 \leqslant n \leqslant \varphi \\ x[n] + \sum_{l=1}^{p} a_p(l) x[n-l] & n \geqslant \varphi + 1 \end{cases}$$

$$J(\underline{a}_p) = \sum_{n=\varphi+1}^{\infty} |e[n]|^2$$

Idea: For every set of $a_p[n]$'s ; by correctly adjusting $b_\varphi[n]$'s,

$e[n] = 0 \qquad 0 \leqslant n \leqslant \varphi$

$$\frac{\partial}{\partial a_p^*[k]} \longrightarrow \frac{\partial J}{\partial a_p^*[k]} = \frac{\partial}{\partial a_p^*[k]} \sum e[n] e^*[n] \qquad 1 \leqslant k \leqslant p$$

$$= \sum_{n=\varphi+1}^{\infty} e[n] \underbrace{\frac{\partial}{\partial a_p[k]} e^*[n]}_{x^*[n-k]}$$

$$= \sum_{n=\varphi+1}^{\infty} \left( x[n] + \sum a_p(l)\, x[n-l] \right) x^*[n-k]$$

$$= r_x(k,0) + \sum_{l=1}^{p} r_x(k,l)\, a_p(l)$$

$$r_x(k,l) = \sum_{n=\varphi+1}^{\infty} x[n-l]\, x^*[n-k] \leftarrow \text{deterministic auto-correlation}$$

$$\begin{array}{l} k=1 \rightarrow \\ \uparrow \\ \frac{\partial J}{\partial a_p^*[1]} = 0 \end{array} \begin{bmatrix} r_x(1,1) & r_x(1,2) & \dots & r_x(1,p) \\ r_x(2,1) & r_x(2,2) & \dots & r_x(2,p) \\ \vdots & & & \\ r_x(p,1) & r_x(p,2) & \dots & r_x(p,p) \end{bmatrix} \begin{bmatrix} a_p(1) \\ \\ \\ a_p(p) \end{bmatrix} = \begin{bmatrix} -r_x(1,0) \\ -r_x(2,0) \\ \\ -r_x(p,0) \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} x(0) & 0 & 0 & \dots & 0 \\ x(1) & x(0) & 0 & \dots & 0 \\ x(2) & x(1) & x(0) & & \\ \vdots & & & & \\ x(\varphi) & x(\varphi-1) & & & \\ x(\varphi+1) & x(\varphi) & \dots & & x(\varphi-p) \\ & & & & \\ x(N) & x(N-1) & \dots & & x(N-p-1) \end{bmatrix}}_{} \begin{bmatrix} 1 \\ a_p(1) \\ a_p(2) \\ \vdots \\ a_p(p) \end{bmatrix} = \begin{bmatrix} b_\varphi(0) \\ b_\varphi(1) \\ \\ \\ b_q(\varphi) \\ \vdots \\ \\ 0 \end{bmatrix}$$

$$\uparrow$$

$$e[n] = x[n] + a_p[n] - b_\varphi[n]$$

↗

Prony Error

$$\begin{bmatrix} x(\varphi) & x(\varphi-1) & \dots & x(\varphi-p) \\ & & & \\ x(N-1) & x(N-2) & & x(N-p-1) \\ \vdots & \vdots & & \end{bmatrix} \begin{bmatrix} a_p(1) \\ \vdots \\ \\ a_p(p) \end{bmatrix} = -\begin{bmatrix} x(\varphi+1) \\ \\ \\ x(N) \end{bmatrix}$$

$$(\underline{\underline{A}}^H \underline{\underline{A}}) x^{LS} = \underline{\underline{A}}^H \underline{b}$$

$$x^{LS} = (\underline{\underline{A}}^H \underline{\underline{A}})^{-1} \underline{\underline{A}}^H b \leftarrow \underline{\underline{A}} \underline{x} = \underline{b}$$

## All pole Modeling (Special Case of Prony (Q=0))

$$H(z) = \frac{b_Q(0)}{1 + a_p(1) z^{-1} + \dots + a_p(p) z^{-p}}$$

$$\overset{\text{Prony}}{J}(\underline{a}_p) = \sum_{n=Q+1}^{\infty} |e[n]|^2 = \sum_{n=1}^{\infty} |e[n]|^2$$

$Q=0$

$e[0] = x(0) - b_Q(0)$ specific for all pole modeling

$$J'(\underline{a}_p) = \sum_{n=0}^{\infty} |e[n]|^2 \dots \rightarrow r_x(k,l) = \sum_{n=0}^{\infty} x[n-l]\, x^*[n-k]$$

$$r_x(k+\Delta, l+\Delta) \overset{?}{=} r_x(k,l)$$

$\Delta$ positive integer.

$$r_x(k+\Delta, l+\Delta) = \sum_{n=0}^{\infty} x[n-l-\Delta]\, x^*[n-k-\Delta]$$

$$= \sum_{n=-\Delta}^{\infty} x[n-l]\, x^*[n-k] \qquad x[n] : \text{causal}$$

$$= \sum_{n=0}^{\infty} x[n-l]\, x^*[n-k] \qquad k, l \text{ are positive}$$

$$= r_x(k,l)$$

$r_x(k,l)$

$1 \leq k \leq p$

$1 \leq l \leq p$

$$r_x(k,l) = \text{func.}(k-l)$$

$$\begin{bmatrix} \overset{r_x^*(1,1)\,\swarrow}{r_x(0)} & \overset{r_x^*(1,2)\,\swarrow}{r_x(1)} & r_x(2) & \cdots & r_x(P-1) \\ \\ \\ r_x(P-1) & r_x(P-2) & & & r_x(0) \end{bmatrix} \begin{bmatrix} a_p(1) \\ \\ \\ a_p(P) \end{bmatrix} = - \begin{bmatrix} \overset{\swarrow\, r_x(1,0)}{r_x(-1)} \\ r_x(-2) \\ \vdots \\ r_x(-P) \end{bmatrix}$$

$$r_x(\gamma) = \sum_{n=0}^{\infty} x[n]\, x^*[n-\gamma]$$

$\uparrow$ det. auto cor

$$r_x(-\gamma) = \sum_{n=0}^{\infty} x[n-\gamma]\, x^*[n] \qquad r_x^*(\gamma)$$

$$= \sum_{n=\gamma}^{\infty} x[n]\, x^*[n-\gamma] = \overset{\|}{\sum_{n=0}^{\infty}} x[n]\, x^*[n-\gamma]$$

Finite Data Records and Auto-Correlation Estimates:

1- Auto-Correlation Method

$x[n]$ : given for $n = \{0, \dots, N-1\}$

Assume $x[n]=0$ whenever it is not available.

Bottom part of the matrix for the Prony method. $\nearrow$

$$\underbrace{\overset{\underline{\underline{A}}}{\begin{bmatrix} x(0) & 0 & \cdots & 0 \\ x(1) & x(0) & \cdots & 0 \\ x(2) & x(1) & \cdots & 0 \\ \vdots & & & \\ x(N-1) & x(N-2) & \cdots & \\ x(N) & x(N-1) & \cdots & 0 \\ 0 & x(N) & x(N-2) & \\ 0 & 0 & x(N) & \end{bmatrix}}}_{\underline{\underline{X}}} \begin{bmatrix} a_p(1) \\ \vdots \\ a_p(P) \end{bmatrix} = - \begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(N) \\ x(N-1) \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$P \times P \rightarrow \hat{R}_x = \frac{\underline{\underline{X}}^T \underline{\underline{X}}}{N} \; : \text{Auto-cor. } \underline{\text{matrix}} \text{ estimate}$$

$r_x(k)$ is calculated with this assumption → auto-correlation method.

$$[\underline{\underline{A}}^T \underline{\underline{A}}]_{l,k} = r_x(k,l) = r_x(l-k)$$

$$r_x(k) = \sum_{n=0}^{\infty} x[n]\,x[n-k] \qquad \text{Assume } x[n]=0 \begin{Bmatrix} n>N \\ n<0 \end{Bmatrix}$$

$$= \sum_{n=k}^{N} x[n]\,x[n-k]$$

2- Covariance Method: You only make use the given data without any assumption.

$$\begin{array}{l} x(\text{negative index})=0 \\ \\ P-1 \rightarrow \\ \text{middle} \\ N \rightarrow \\ N+1 \rightarrow \end{array} \begin{bmatrix} x(0) & 0 & \cdots & 0 \\ x(1) & x(0) & \cdots & 0 \\ \vdots & & & \\ x(P-1) & x(P-2) & \cdots & x(0) \\ \vdots & & & \\ x(N) & x(N-1) & \cdots & x(N-P) \end{bmatrix}_{\underline{\underline{A}}_M} \begin{bmatrix} a_p(1) \\ \vdots \\ a_p(P) \end{bmatrix} = - \begin{bmatrix} x(1) \\ \vdots \\ x(P) \\ \vdots \\ x(N+1) \end{bmatrix}$$

If I use only the middle part for $\hat{r}_x(k)$ estimation; then there're no assumption introduced.

$$\underset{\substack{\uparrow \\ P\times P}}{\underline{\underline{R}}_x} = \frac{1}{N-(P-1)} \underline{\underline{A}}_M^T \underline{\underline{A}}_M \leftarrow \text{Matrix estimate with covariance method!}$$

Properties: Hayes

1- $\hat{\underline{\underline{R}}}_x$ matrix formed by $\hat{r}_x(k)$ according to auto-corr. method is positive def. $\hat{R}_x \geq 0$

2- Solution of $\hat{\underline{\underline{R}}}_x \underline{a} = \underline{b}$ given you an $1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_p z^{-p}$ $A_p(z)$ with poles inside the unit circle → when $\hat{r}_x(k)$ is calculated with auto-cor. method.

$\hat{\underline{\underline{R}}}_X$ estimation: Therrien
Hayes

$\underline{X}$: a random vector.

$\underline{\underline{R}}_X = E\{XX^H\}$ ; How to estimate $\underline{\underline{R}}_X$ ?

Auxilary Vectors $= \{\underline{t}_1, t_2, \ldots, \underline{t}_L\}$

$$\hat{\underline{\underline{R}}}_X = \frac{1}{L}\sum_{k=1}^{L} t_k t_k^H \quad \begin{array}{l} \rightarrow E\{\hat{\underline{\underline{R}}}_X\} = R_X \\ \rightarrow E\{\|\underline{\underline{R}}_X - \hat{\underline{\underline{R}}}_X\|\} \leftarrow \text{error variance} \end{array}$$

$\hat{\underline{\underline{R}}}_X$ : Maximum Likelihood (ML) est. of $R_X$ for Gaussian Vectors.

$$\underset{M\times M}{\hat{\underline{\underline{R}}}_X} = \frac{1}{L} \overset{\underline{\underline{X}}}{\underset{M\times L}{[\underline{t}_1 \ \underline{t}_2 \ \ldots \ t_L]}} \overset{\underline{\underline{X}}^H}{\underset{L\times M}{\begin{bmatrix} t_1^H \\ t_2^H \\ \vdots \\ t_L^H \end{bmatrix}}} = \frac{1}{L} \underline{\underline{X}} \cdot \underline{\underline{X}}^H$$

$\underline{t}_1$ ← column vectors $M\times 1$

$$Hw-0 \rightarrow (\underline{\underline{A}}\,\underline{\underline{B}})_{kl} = [\xleftrightarrow{k^{th}}] \left[\ \updownarrow l^{th} \text{ column}\right] = [\ \ a_{(k,l)}]$$

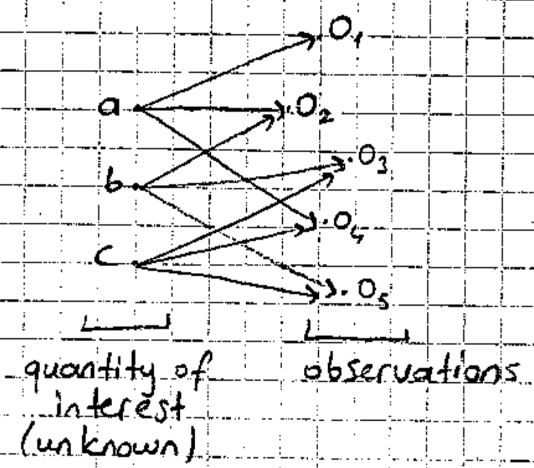
$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \end{bmatrix} [b_{11} \ b_{12}] + \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} [b_{21} \ b_{..}]$$

$$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

$$\underline{\underline{A}} = [\underline{e}_1 \ \underline{e}_2 \ \ldots \ e_N] \begin{bmatrix} \lambda & 0 & 0 & 0 & 0 \\ & & & & \\ 0 & & & & \lambda_N \end{bmatrix} \begin{bmatrix} e_1^T \\ \vdots \\ e_N^T \end{bmatrix}$$

$$[e_1 \ \ldots \ e_n] \begin{bmatrix} \lambda_1 e_1^T \\ \lambda_2 e_2^T \\ \vdots \\ \lambda_N e_N^T \end{bmatrix} = \sum_{k=1}^{N} \lambda_k \underline{e}_k \underline{e}_k^T$$

# Estimation:



$r = c + n$ ← noise: $N(0, \sigma_n^2)$

observation → $r$; $c$ ← quantity of interest (real)

$$\underline{r} = \begin{bmatrix} c \\ c \\ c \\ \vdots \\ c \end{bmatrix} + \underline{n} \leftarrow N: (0, \sigma_n^2 \underline{\underline{I}})$$

observation vector

Estimator: $\text{func}(\underline{r}) = \hat{\theta}$ ← $\theta$: quantity of interest

observation ↑ ; ↑ estimate

Estimation Theory

$\theta$: Quantity Interest

- Random Par. $f_{\underline{\theta}}(\theta)$
  - Bayesian Estimator: E{cost} is minimized
  - MAP: $\max\limits_{\theta} f(\underset{\sim}{\theta} | \underline{r})$
- Non-random (deterministic)
  - ML: Maximum Likelihood: $\max\limits_{\theta} f(\underline{r} | \theta)$
  - LS estimate

## Non-Random Par. Est.

Ex) $x_k = c + n_k$ , $n_k$ i.i.d $N(0, \sigma_n^2)$

$\{x_1, \_\_, x_N\}$ : N observations are provided.

Maximum Likelihood Est.

$$\underset{(\hat{c})}{\overset{\uparrow}{\hat{\theta}}} = \arg\max_{\theta} f(\underline{r}; \theta) \qquad \hat{c} = \arg\max_{c} f(x_1, x_2, \dots, x_N; c)$$

$$f(x_1, \_\_, x_N; c) = \left(\frac{1}{\sqrt{2\pi\sigma_n^2}}\right)^N \prod_{k=1}^{N} e^{-\frac{(\overset{\text{given observation}}{\overset{\downarrow}{x_k}} - c)^2}{2\sigma_n^2}}$$

$$\frac{\partial}{\partial c} f(x_1, \_\_, x_N; c)\Big|_{\underset{c=\hat{c}}{\downarrow}} = 0 \rightarrow \hat{c} = \arg\max_{c} f(x_1, \_\_, x_N; c)$$

$$= \arg\max_{c} \ln(f(x_1, \dots, x_N; c))$$

$$\frac{\partial}{\partial c} \ln f(x_1, \_\_, x_N; c) = 0$$

$$-\frac{N}{2} \ln 2\pi\sigma_n^2 - \sum_{k=1}^{N} \frac{(x_k - c)^2}{2\sigma_n^2}$$

$\downarrow$

$$\frac{-2}{2\sigma_n^2} \sum_{k=1}^{N} (x_k - c) = 0$$

$$\hat{c} = \frac{1}{N} \sum_{k=1}^{N} x_k$$

↑ ML estimate.

estimator.

Properties of Estimators.

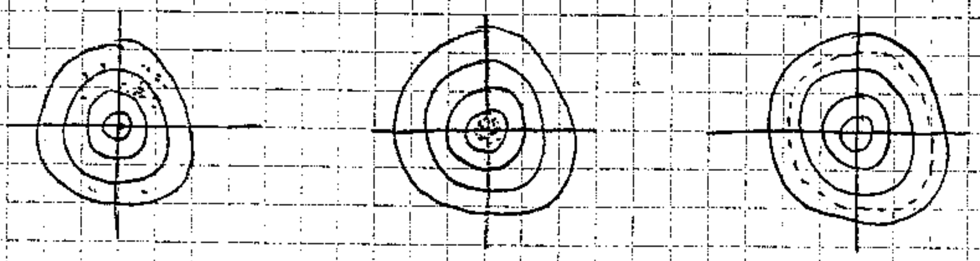
1- Bias

$E\{\hat{\theta}\} = \theta \rightarrow$ unbiased estimator.

otherwise $\rightarrow$ biased.

EX: Earlier example

$$E\{\hat{c}\} = \frac{1}{N}\sum_{k=1}^{N} E\{x_k\} = c \rightarrow \text{unbiased}$$



2- Consistency: An estimator is consistent if

$$E\{(\theta-\hat{\theta})^2\} \rightarrow 0 \quad \text{as} \quad \underset{\substack{\uparrow \\ \text{number of observations}}}{N} \rightarrow \infty$$

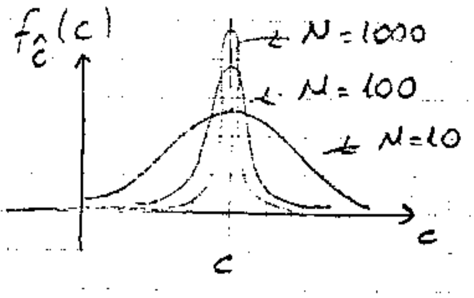
EX: Earlier example.

$$E\{(c-\hat{c}_N)^2\} = E\{(c-\frac{1}{N}\sum_{k=1}^{N} \underset{\leftarrow c+n_k}{x_k})^2\}$$

$$= E\left\{\left(-\frac{1}{N}\sum_{k=1}^{N} n_k\right)^2\right\}$$

$$= \frac{1}{n^2} \cdot N\sigma_n^2 = \frac{1}{N}\sigma_n^2 \rightarrow \text{consistent Estimator.}$$

$$\hat{c} = \frac{1}{N}\sum_{k=1}^{N} x_k \qquad \hat{c} \sim N(c, \frac{\sigma_n^2}{N})$$



3- Efficiency: An estimator is called efficient if it reaches the Cramer-Rao Bound

Cramer-Rao Lower Bound $\Rightarrow E\{(\theta-\theta)^2\} \geq CRB(\theta)$

For any unbiased estimator CRB is valid.

$$CRB(\theta) = \frac{1}{E_r\left\{\left(\frac{\partial}{\partial\theta}\ln f(r,\theta)\right)^2\right\}}$$

Ex) Earlier Example

$\hat{c}$: efficient

$$E_r\left\{\underbrace{\left[\frac{-2\sum(x_k-c)}{2\sigma_n^2}\right]}_{\frac{\partial}{\partial\theta}\ln(f(r;\theta))}^2\right\} = \frac{E\left\{\left(\sum_{k=1}^{N} n_k\right)^2\right\}}{\sigma_n^4}$$

$$= \frac{N\sigma_n^2}{\sigma_n^4}$$

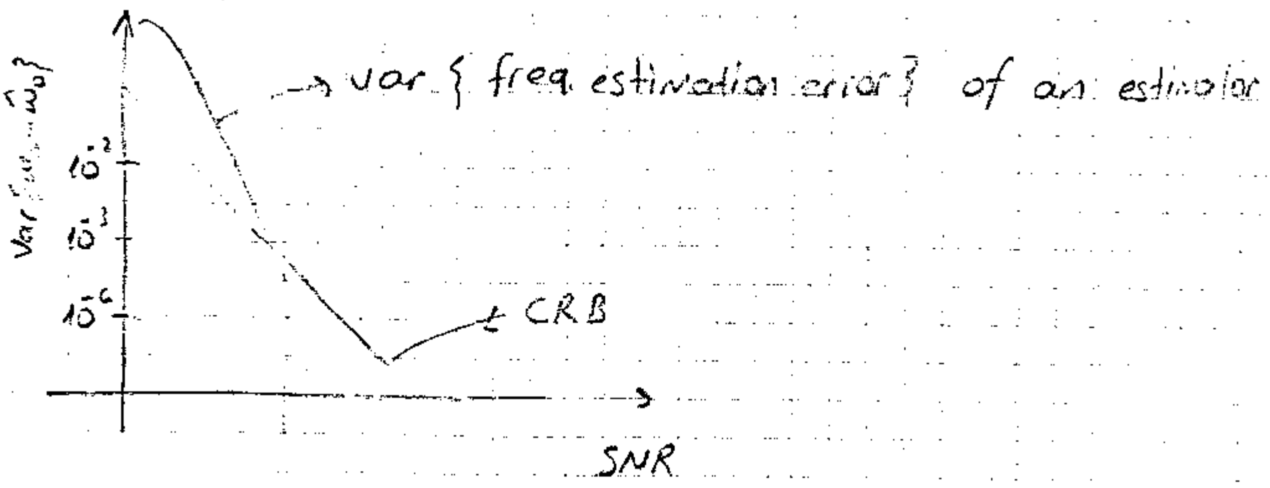
$$= \frac{N}{\sigma_n^2}$$

$$CRB(\theta) = \frac{\sigma_n^2}{N}$$

Ex) $\underline{r} = Ae^{j\emptyset}e^{j\omega_0 n} + \underline{n} \leftarrow N(0, \sigma_n^2 I)$

$$\left.\begin{array}{l} A: \\ \emptyset: \\ \omega_0: \end{array}\right\} \text{non-random par.}$$

$$SNR = \frac{A^2}{\sigma_n^2}$$



Asymptotic Efficiency: As $\begin{pmatrix} SNR \to \infty \\ \text{or} \\ N \to \infty \end{pmatrix}$ if CRB is achieved $\downarrow$ estimator asymptotically efficient.

Estimation:

2- Random Parameter Estimation:

$\underset{\sim}{\theta}$: $\leftarrow$ unknown quantity from a sample space

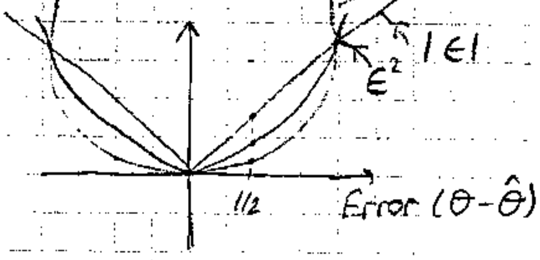
$\underline{X}$: observation vector

$\hat{\theta} = \overset{\leftarrow \text{estimator}}{\hat{\theta}(\underline{x})}$

Bayesian Cost Function for Estimation

$\text{Cost} = (\theta - \hat{\theta})^2$

$\text{Cost} = \| \theta - \hat{\theta} \|$



The goal is minimizing the average cost (penalty, risk):

$$E_{\theta,x} \{ \text{cost}(\underset{\sim}{\theta} - \underbrace{\hat{\theta}}_{f(\underline{x})}) \}$$

Let's adopt $\text{cost} = \epsilon^2$

$$E_{\theta,x} \{ (\theta - \hat{\theta})^2 \} = \int_{\theta} \int_{\underline{X}} (\theta - \hat{\theta}(\underline{x}))^2 f(\theta, \underline{x}) \, dx \, d\theta$$

$$= \int_{\underline{x}} \overline{\int_{\theta} (\theta - \hat{\theta}(x))^2 f(\theta | x) \, d\theta} \, f(x) \, dx$$

$$= \int_{\underline{x}} \overset{\leftarrow \text{positive quantity}}{\overline{I(\theta, \hat{\theta}(x))}} f_x(\underline{x}) \, dx$$

Then to minimize

$E\{(\theta-\hat{\theta})^2\}$ ; we can focus on minimizing $I(\theta-\hat{\theta}(x)]$

$$I(\theta,\hat{\theta}(x)]=\int(\theta-\hat{\theta}(x))^2 f(\theta|x)\,d\theta$$

$\underline{x}$ : given

$$\hat{\theta}(\underline{x})=c$$

$$=\int(\theta-c)^2 f(\theta|x)\,d\theta$$

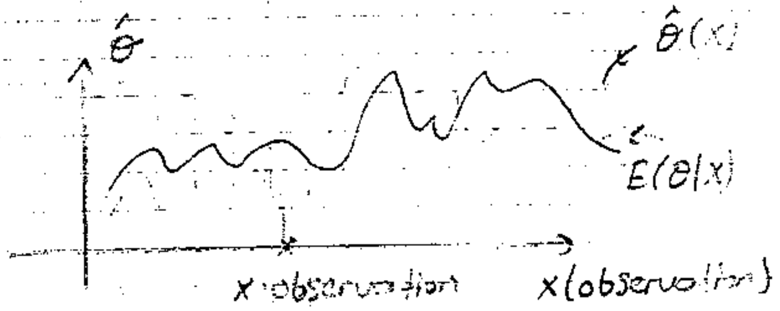
$$\frac{\partial}{\partial c} I(\theta,\hat{\theta}(x))=0 \longrightarrow \int -2(\theta-c) f(\theta|x)\,d\theta=0$$

$$\int \theta f(\theta|\underline{x})\,d\theta=\int c f(\theta|\underline{x})\,d\theta$$

$$E\{\theta|\underline{x}\}=c=\hat{\theta}(x)$$

Result: To minimize MSE of random par. est ; the optimal estimator is the conditional mean $E\{\theta|x\}=\hat{\theta}(x)$

Sample space for $(\theta,\underline{x})$



Problem: You need joint pdf of $(\theta,\underline{x})$ to calculate $E\{\theta|x\}$

Estimator optimal but too general

We suggest to impose a constraint on the estimator and solve the optimal estimator with in the constraints.

So a structure on the estimator will be imposed.

Linear / Affine Estimator: $$\left.\begin{aligned}\hat{\theta}&=ax+b \leftarrow \text{Affine func.}\\ &=ax \leftarrow \text{linear func.}\end{aligned}\right\}$$ what is the best a, b for the estimator?

Sec. 3.26 from Hayes

Ex: Goal minimize MSE for $\underset{\text{estimator}}{\hat{y}} = a\overset{\text{obs}}{x} + b$ , $J(a,b) = E\{(y-\hat{y})^2\}$

by selecting a and b properly.

$$\frac{\partial J}{\partial a} = 0 \rightarrow \frac{\partial}{\partial a} E\{(\underbrace{y-(ax+b)}_{e})^2\} = E\{-2ex\} = 0$$

$$\frac{\partial J}{\partial b} = 0 \longrightarrow \frac{\partial}{\partial b} E\{e^2\} = E\{-2e \cdot 1\} = 0$$

$$E\{ex\} = 0 \longrightarrow E\{(y-(ax+b))x\} = E\{yx\} - aE\{x^2\} - bE\{x\} = 0$$

$$E\{e.1\} = 0 \longrightarrow E\{y\} - aE\{x\} - b = 0$$

$$\begin{bmatrix} E\{x^2\} & E\{x\} \\ E\{x\} & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} E\{xy\} \\ E\{y\} \end{bmatrix}$$

$$\begin{bmatrix} a \\ b \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} 1 & -\mu_x \\ -\mu_x & E\{x^2\} \end{bmatrix} \begin{bmatrix} E\{xy\} \\ E\{y\} \end{bmatrix}$$

$$E\{x^2\} - \mu_x^2 = \sigma_x^2$$

$$= \frac{1}{\sigma_x^2} \begin{bmatrix} E\{xy\} - \mu_x E\{y\} \\ -\mu_x E\{xy\} + E\{x^2\}E\{y\} \end{bmatrix}$$

$$\hat{y} = ax + b = \frac{1}{\sigma_x^2}\left((\overbrace{E\{xy\} - \mu_x\mu_y}^{a})x + \overbrace{\overset{\sigma_x^2+\mu_x^2}{E\{x^2\}}\mu_y - \mu_x \overset{E\{xy\} = \rho_{xy}\sigma_x\sigma_y + \mu_x\mu_y}{E\{xy\}}}^{b}\right)$$

$$E\{(x-\mu_x)(y-\mu_y)\} = \rho_{xy}\sigma_x\sigma_y$$

cor. coef $|\rho_{xy}| \leq 1$

$$= \frac{\sigma_y}{\sigma_x}\rho_{xy}x + \frac{(\sigma_x^2+\mu_x^2)\mu_y}{\sigma_x^2} - \frac{\mu_x\rho_{xy}\sigma_x\sigma_y}{\sigma_x^2} - \frac{\mu_x^2\mu_y}{\sigma_x^2}$$

$$\hat{y} = \frac{\sigma_y}{\sigma_x} \rho_{xy} (x - \mu_x) + \mu_y$$

$J = E\{(y-\hat{y})^2\}$ what is the minimum value of J achieved by the optimal estimator?

$$J(a_{opt}, b_{opt}) = E\{(y-\hat{y})^2\} = E\{e(y-\hat{y})\}$$

$$= E\{ey\} = E\{e(a_{opt}x + b_{opt})\}$$

$$= E\{ey\} - a_{opt}\overset{0}{\cancel{E\{ex\}}} - b_{opt}\overset{0}{\cancel{E\{e\}}}$$

$$= E\{[y - (a_{opt}x + b_{opt})]y\}$$

$$= E\{y^2\} - a_{opt}E\{xy\} - b_{opt}E\{y\}$$

min LMMSE (min; Linear)

$$E\{ey\} = E\{e(y-\mu_y)\} = E\{ey\} - \overset{0}{E\{e\}}\mu_y$$

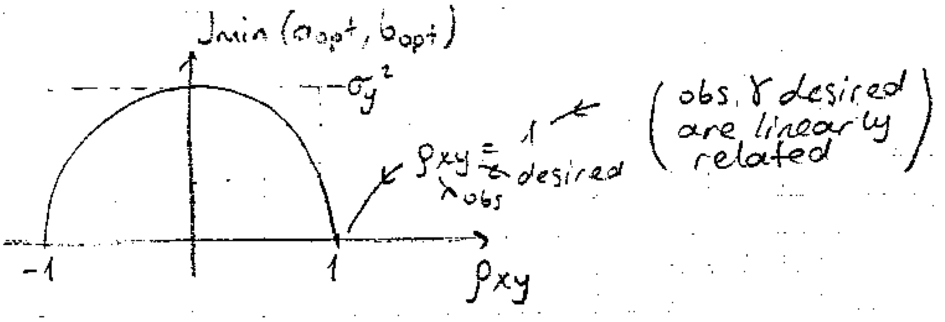
$$= E\left\{\left(y - \frac{\sigma_y}{\sigma_x}\rho_{xy}(x-\mu_x) - \mu_y\right)(y-\mu_y)\right\}$$

$$= E\{(y-\mu_y)^2\} - \frac{\sigma_y}{\sigma_x}\rho_{xy}\frac{E\{(x-\mu_x)(y-\mu_y)\}}{\rho_{xy}\sigma_x\sigma_y}$$

$$= \sigma_y^2 - \sigma_y^2\rho_{xy}^2$$

$$= \sigma_y^2(1-\rho_{xy}^2) \longleftarrow \text{min avg. error of the structured estimator.}$$

$\uparrow$ opt. coef.



$$\rho_{ab} = \mp 1$$

$$\updownarrow$$

$$\underset{\sim}{a} = \gamma \underset{\sim}{b} + \alpha$$

# Multiple Observations and LMMSE Estimation

$\underline{x}$ : obs vector $(N \times 1)$

$\hat{y} = \underline{w}^T . \underline{x}$

$\underline{w}$ : $N \times 1$ vector linear comb. coef.

Goal: $E\{(y-\hat{y})^2\}$ minimized wrt $\underline{w}$.

$$J(\underline{w}) = E\{(y - w^T x)^2\}$$

$$\nabla_{\underline{w}} J = \begin{bmatrix} \frac{\partial J}{\partial w_1} \\ \vdots \\ \frac{\partial J}{\partial w_N} \end{bmatrix} = \underline{0}$$

$$\nabla_{\underline{w}} J(\underline{w}) = E\{\nabla_w ((y - w^T x)^2)\}$$

$$= E\{\nabla_w (\quad)\} = E\{-2\underline{x}y + 2\underline{x}\underline{x}^T w\} = 0$$

$$(y - \underline{w}^T\underline{x})^2 = (y - \underline{w}^T\underline{x})^T (y - \underline{w}^T\underline{x})$$

$$= (y - w^T x)(y - w^T x)$$

$$= y^2 - 2\underline{w}^T\underline{x}y + w^T x x^T w \rightarrow \nabla_w (\downarrow) = -2\overset{\text{obs. vec. } N\times 1}{\underline{x}}y + 2\underline{x}\underline{x}^T\underline{w}$$

$$E\{\underline{x}\underline{x}^T\}\underset{\uparrow w_{opt}}{\underline{w}} = E\{\underline{x}y\}$$

$\underline{\underline{R}}_x \underline{w} = r_{\underline{x}y}$ ← Eqn. set for opt. est.

cross corr of each obs. and desired quantity listed in a vector.

Auto-cor. matrix of observation

Ex: Revisit the earlier problem

$$\hat{y} = ax + b = [w_0 \ \ w_1]\begin{bmatrix} x \\ 1 \end{bmatrix} = \underline{w}^T x_\oplus$$

Opt. $\underline{w}$'s are the solution of $\underline{\underline{R}}_{x_\oplus} \underline{w}_{opt} = r_{x_\oplus y}$

$$E\{x_{\oplus} x_{\oplus}^T\} = E\left\{\begin{bmatrix} x \\ 1 \end{bmatrix}[x \ \ 1]\right\} = \begin{bmatrix} E\{x^2\} & \mu_x \\ \mu_x & 1 \end{bmatrix}$$

$$\downarrow$$

$$\begin{bmatrix} & \\ & \end{bmatrix}\begin{bmatrix} \omega_0 \\ \omega_1 \end{bmatrix} = \begin{bmatrix} \\ \\ \end{bmatrix}$$

$$r_{x_{\oplus} y} = E\left\{\begin{bmatrix} x \\ 1 \end{bmatrix} y\right\} = \begin{bmatrix} E\{xy\} \\ \mu_y \end{bmatrix}$$

The same equation system is retrieved.

Important Note:

Orthogonality principle

$$\nabla_\omega J(\underline{\omega}) = \underline{0}$$

$$E\{\nabla_{\underline{\omega}} e^2\} = 0$$

$$E\left\{\begin{bmatrix} \frac{\partial}{\partial \omega_1} \\ \vdots \\ \frac{\partial}{\partial \omega_N} \end{bmatrix} e^2\right\}$$

$$E\left\{2\begin{bmatrix} e\frac{\partial e}{\partial \omega_1} \\ e\frac{\partial e}{\partial \omega_2} \\ \vdots \\ \end{bmatrix}\right\} = 0$$

$$e = y - \underline{\omega}^T \underline{x} = y - [\omega_1 \ — \ \omega_N]\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}$$

$$\frac{\partial e}{\partial \omega_k} = x_k$$

$$\downarrow$$

$$E\left\{2\begin{bmatrix} e x_1 \\ e x_2 \\ \vdots \\ e x_N \end{bmatrix}\right\} = \underline{0}$$

$E\{e\underline{x}\} = \underline{0}$ ← orthogorality cond.

the error of the opt. estimator is $\perp$ the observations (obs. vector).

Result: $E\{e\,x_{\oplus}\} = \underline{0}$

$$E\left\{e\begin{bmatrix} x \\ 1 \end{bmatrix}\right\} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Linear Min. MSE Estimation

$\underline{x}$: observation vector

$\theta$: random quantity to be estimated

$$\hat{\theta} = \underbrace{\underline{w}^T}_{\text{combination weights}} \cdot \underline{x} \longrightarrow \underline{w} = ? \qquad J(w) = E\{(\theta - \underbrace{\hat{\theta}}_{\underline{w}^T x})^2\}$$

$$\min_{\underline{w}} J(w) = \min_{\underline{w}} E\{(\theta - \underline{w}^T x)^2\}$$

$$\underbrace{\underline{\underline{R}}_x}_{E\{x^T x\}} w_{opt} = \underbrace{r_{\theta x}}_{E\{\theta \underline{x}\}} \qquad \hat{\theta} = E\{\theta / \underline{x}\}$$

Joint pdf $\{\underset{\sim}{\theta}, \underline{x}\}$

Ex: $x_n = \underline{c} + w_n$, $n = \{1, \ldots, N\}$

$w_n$: Independent Gaussian r.v. $N(0, \sigma_{w_n}^2)$

$$\underline{x} = \begin{bmatrix} \underline{x}_1 \\ \\ \\ \underline{x}_N \end{bmatrix} ; \quad \hat{c} = \underline{w}^T \cdot \underline{x}$$ , Find LMMSE estimator.

$$\underline{\underline{R}}_x \underline{w} = r_{c\underline{x}}$$

$$\underline{\underline{R}}_x = E\left\{\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} [x_1 \ldots x_N]\right\} = \begin{bmatrix} E\{x_1^2\} & E\{x_1 x_2\} & & E\{x_1, x_N\} \\ E\{x_2 x_1\} & & & E\{x_2, x_N\} \\ & & & \\ E\{x_N x_1\} & & & E\{x_N x_N\} \end{bmatrix}$$

$$E\{x_k^2\} = E\{(c + w_k)^2\} = \sigma_c^2 + \sigma_{w_k}^2$$

$$E\{x_k x_l\} = E\{(c + w_k)(c + w_l)\}$$
$$k \neq l \quad = \sigma_c^2$$

$$r_{c\underline{x}} = E\left\{ c \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \right\} = \begin{bmatrix} \sigma_c^2 \\ \vdots \\ \sigma_c \end{bmatrix}$$

$$\frac{1}{\sigma_c^2} \begin{bmatrix} \sigma_c^2 + \sigma_{w_1}^2 & \sigma_c^2 & & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_{w_2}^2 & & \\ & & & \\ \sigma_c^2 & \sigma_c^2 & \dots & \sigma_c^2 + \sigma_{w_N}^2 \end{bmatrix} \begin{bmatrix} w_1 \\ \\ \\ w_N \end{bmatrix} = \frac{1}{\sigma_c^2} \begin{bmatrix} \sigma_c^2 \\ \sigma_c^2 \\ \\ \sigma_c^2 \end{bmatrix}$$

$$(SNR)_{observation_k} = \frac{E\{(\text{signal term of } k^{th} \text{ observation})^2\}}{E\{(\text{noise term of } k^{th} \text{ observation})^2\}}$$

$$\begin{bmatrix} 1 + 1/SNR_1 & 1 & \dots & 1 \\ 1 & 1 + 1/SNR_2 & . & 1 \\ 1 & 1 & 1 + 1/SNR_3 & 1 \\ \vdots & & & \vdots \\ 1 & 1 & 1 & 1 + 1/SNR_N \end{bmatrix} \begin{bmatrix} w_1 \\ \\ \\ \\ w_N \end{bmatrix} = \begin{array}{c|l} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{bmatrix} & \begin{array}{l} \to 1^{st} \text{ equation} \\ \\ \to k^{th} \text{ equation} \\ \\ \\ \end{array} \end{array}$$

$$\to k^{th} \text{ equation} \quad \sum_{l=1}^{N} w_l + \frac{1}{SNR_k} \cdot w_k = 1$$

$$\rightarrow w_k = (1-S)\, SNR_k$$

$$S = \sum_{k=1}^{N} w_k = \sum_{k=1}^{N} (1-S)\, SNR_k$$

$$S = \frac{\sum_{k=1}^{N} SNR_k}{1 + \sum_{k=1}^{N} SNR_k} \qquad w_k = \frac{(SNR)_k}{1 + \sum_{k=1}^{N} (SNR)_k}$$

$$\underline{w}^T = \left[ \frac{SNR_1}{1 + \sum_k SNR_k}, \frac{SNR_2}{1 + \sum_k SNR_k}, \dots\dots, \frac{SNR_N}{1 + \sum SNR_k} \right]$$

# Properties of LMMSE estimators:

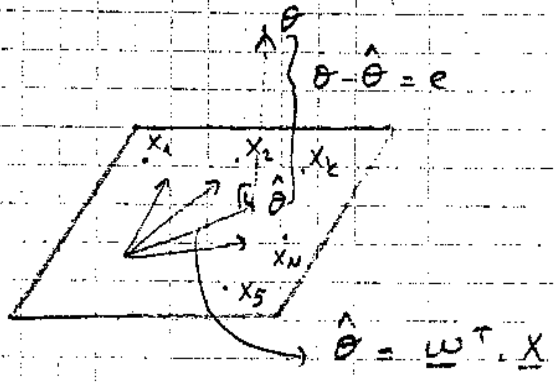
1- Geometric Interpretation:

$\theta$: desired variable

$\underline{X}$: composed of observations

$\hat{\theta} = \underline{w}^T x \rightarrow$ linear combination of observation

$E\{\underset{\underset{(\theta - \hat{\theta}) \text{ for opt (LMMSE est.)}}{\uparrow}}{e}\,\underline{x}\} = \underline{0}$ (orthogonality Principle for LMMSE est.)



$E\{xy\} = 0 \quad x \perp y$

$E\{xy\} = \langle x, y \rangle$

$E\{x^2\} = \|x\|^2$

review

$\hat{y} = \underline{w}^T \underset{\underset{\text{obs.}}{\uparrow}}{\underline{X}}$

min

$E\{(y-\hat{y})^2\}$ : MSE

$\underline{\underline{R}}_x \underset{\underset{\text{opt } \underline{w}}{\uparrow}}{\underline{w}} = r_{y\underline{x}}$

2- Multiple par. estimation from $\underline{X}$

$y_1, y_2, ----, y_N$ : N r.v.'s of interests.

$\underline{X} = \underline{\underline{H}}\,\underline{y} + \underline{\eta}$

$$\text{Goal: } \min\Big(\sum_{k=1}^{N} E\{(y_k - \underset{\underset{\underline{w}_k^T . \underline{X}}{\uparrow}}{\hat{y}_k})^2\} = E\{\|\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}}_{\underline{y}} - \underbrace{\begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_N \end{bmatrix}}_{\hat{\underline{y}}}\|^2\}$$

$$= E\{ \|\underbrace{\underline{y}-\underline{\hat{y}}}_{\epsilon}\|^2 \}$$

$$tr\{\underline{\underline{A}}\,\underline{\underline{B}}\} = tr\{\underline{\underline{B}}\,\underline{\underline{A}}\}$$

$$tr\{\underbrace{\epsilon^T\epsilon}_{1\times 1}\} = tr\{\underbrace{\epsilon\epsilon^T}_{N\times N}\}$$

$$= E\{ tr\{ (\underline{y}-\underline{\hat{y}})(\underline{y}-\underline{\hat{y}}) \} \}$$

$$= tr\{ E\{ \epsilon\epsilon^T \}$$

$$= tr\{ \underline{\underline{R}}_\epsilon \} \qquad \epsilon = \underline{y}-\underline{\hat{y}}$$

Let's find the LMMSE estimator for $y_k$

$$\underline{\underline{R}}_x \underline{\underline{w}}_k = r_{y_k \underline{x}} \qquad k = \{1, \ldots, N\}$$

$$\underline{\underline{R}}_x \underbrace{[\underline{w}_1 \ \underline{w}_2 \ \ldots \ \underline{w}_N]}_{\underline{\underline{w}}} = [r_{y_1\underline{x}} \ r_{y_2\underline{x}} \ \ldots \ r_{y_N\underline{x}}]$$

$$E\{\underline{x}\,\underline{y}^T\} = \underline{\underline{R}}_{xy}$$

$$\begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} [y_1 \ \ldots \ y_N]$$

$$\underline{\underline{w}} = \underline{\underline{R}}_x^{-1} \underline{\underline{R}}_{xy} \longrightarrow \underline{\hat{y}} = \begin{bmatrix} w_1^T \\ \vdots \\ w_N^T \end{bmatrix} \underline{x} = \underline{\underline{w}}^T \underline{x}$$

$$R_{xy}^T \qquad \hat{y} = R_{\underline{y}\underline{x}} \underline{\underline{R}}_x^{-1} \underline{x}$$

Special case: $\overset{\text{obs}}{\underline{x}} = \underline{\underline{H}}\,\underline{y} + \underline{n}$ ← Linear observation ($\underline{\underline{H}}$) + noise

$\underline{y}, \underline{n}$ are independent

$$\hat{y} = ? \longrightarrow \hat{y} = R_{\underline{y}\underline{x}} \underline{\underline{R}}_x^{-1} \underline{x}$$

$$\underline{\underline{R}}_x = \underline{\underline{H}}\,\underline{\underline{R}}_y \underline{\underline{H}}^T + \underline{\underline{R}}_n$$

$$\underline{\underline{R}}_{\underline{y}\underline{x}} = E\{ \underline{y} (\underline{\underline{H}}\,\underline{y} + n)^H \} = \underline{\underline{R}}_y \underline{\underline{H}}^H$$

$$\hat{y} = \underline{\underline{R}}_y \underline{\underline{H}}^H (\underline{\underline{H}}\,\underline{\underline{R}}_y \underline{\underline{H}}^H + \underline{\underline{R}}_n)^{-1} x$$

Linear MMSE for Linear obs. Estimator model.

If $\underset{\underset{\text{obs.}}{\uparrow}}{\underline{x}}$ and $\underset{\underset{\text{desired}}{\uparrow}}{\underline{y}}$ are jointly Gaussian then LMMSE is the

best estimator among all estimators (not only Linear) in MSE sense.

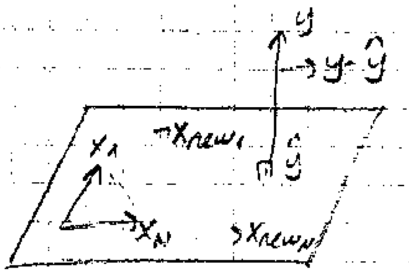
3- Orthogonality

Assume instead of $\underline{x}$, we are $\underline{\underline{M}}\underline{x}$ as the observation vector; Then ⟶ new observations are also satisfy the orthogonality condition for LMMSE ($E\{e\underline{x}\}=0$)

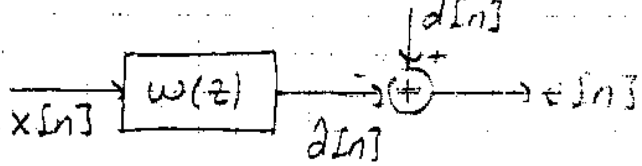
$$\underline{x} \longrightarrow \text{Opt. Linear Estimator} \rightarrow \hat{y} \rightarrow E\{\underbrace{e}_{(y-\hat{y})}\underline{x}\} = 0 \qquad \hat{y} = w_{opt}^T x$$

of the original estimator

$$\text{Then } \underbrace{\underline{\underline{M}}\underline{x}}_{x_{new}} \longrightarrow E\{e\underline{x}_{new}\} = E\{e\underline{\underline{M}}\underline{x}\} = \underline{\underline{M}}\underbrace{E\{e\underline{x}\}}_{=\underline{0}}$$



FIR Wiener Filtering (Hayes)



$x[n]$ : WSS
$d[n]$ : WSS
} Jointly WSS

$$w(z) = w_0 + w_1 z^{-1} + \ldots + w_{p-1} z^{-(p-1)} \Big\} \text{ P Tap}$$

$(P-1)^{th}$ order

$(P-1)^{th}$ zeros

Goal: Find a filter s.t. $E\{(d[n]-\hat{d}[n])^2\}$ is minimized.

$$\hat{d}[n] = \sum_{k=0}^{P-1} x[k]\,w[n-k] = \sum_{k=0}^{P-1} w[k]\,x[n-k]$$

$$= \underset{[w[0]\ \ldots\ w[P-1]]}{\underline{w}^T} \cdot \begin{bmatrix} x[n] \\ x[n-1] \\ \\ x[n-(P-1)] \end{bmatrix}$$

$$= \underline{w}^T \cdot \underline{x}[n]$$

most recent P samples.

Goal. $E\{(d[n] - w^T \underline{x}[n])^2\} \longrightarrow \underline{\underline{R}}_{x[n]}\,\underline{w} = r_{d[n]\underline{x}[n]}$

$$\underline{\underline{R}}_{x[n]} = E\left\{ \begin{bmatrix} x[n] \\ \\ x[n-(P-1)] \end{bmatrix} [x[n] \ \ldots\ x[n-(P-1)]] \right\}$$

$$= \underline{\underline{R}}_{x[n]}\,\underline{w} = E\left\{ d[n] \begin{bmatrix} x[n] \\ \vdots \\ x[n-(P-1)] \end{bmatrix} \right\} = r_{dx} \rightarrow \underline{\underline{R}}_x\,\underline{w} = r_{d\underline{x}}$$

$$\underline{w}^{FIR-wiener} = \underline{\underline{R}}_x^{-1}\, r_{d\underline{x}}$$

$$J_{min} = E\{(d[n] - \underline{w}_{opt}^T\, \underline{x}[n])^2\}$$

$$= E\{e[n](d[n] - w_{opt}^T\, \underline{x}[n])\}$$

$$= E\{\underset{\uparrow\, d[n] - w_{opt}^T \underline{x}[n]}{e(n)}\, d(n)\} - w_{opt}^T\, \overset{\to 0}{E\{e[n]\,\underline{x}[n]\}}$$

$$= r_d(0) - w_{opt}^T\, r_{dx}$$

$$J_{min} = r_d(0) - r_{d\underline{x}}^T\, \underline{\underline{R}}_x^{-1}\, r_{d\underline{x}}$$

Ex: $r_d(k) = \alpha^{|k|}$

$r_v(k) = \sigma_v^2 \delta(k)$ (white noise)

$x[n] = d[n] + v[n]$ ; noise and desired are uncorrelated. Find $\hat{d}[n]$ using two-tap FIR Wiener filter.

$$\hat{d}[n] = \underset{?}{w_0}\, x[n] + \overset{?}{w_1}\, x[n-1]$$

$$\underline{\underline{R}}_{x_{2\times2}}\, \underline{w} = r_{d\underline{x}_{2\times1}} \leftarrow r_{dx} = E\{d[n]\,x[n-k]\} = r_d(k)$$

$$\nwarrow r_x(k) = r_d(k) + r_v(k) = \alpha^{|k|} + \sigma_v^2 \delta[k]$$

$$\underline{\underline{R}}_x\, \underline{w} = r_{dx}$$

$$\begin{bmatrix} r_x(0) & r_x(1) \\ r_x(1) & r_x(0) \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} r_{dx}(0) \\ r_{dx}(1) \end{bmatrix}$$

$$\begin{bmatrix} 1+\sigma_v^2 & \alpha \\ \alpha & 1+\sigma_v^2 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha \end{bmatrix}$$

$\alpha = 0,8$ ; $\sigma_v^2 = 1$

$w_{opt}(z) = 0,4048 + 0,2381 z^{-1}$

$J_{min} = 1 - w_{opt}^T \cdot r_{dx} = 0,4048$

Comparison:

a) No filtering

$x[n] = d[n] + v[n]$

$\hat{d}[n] = E\{d[n]\} = \mu_d$

→ $J = E\{(d - \hat{d}[n])^2\} = r_d(0) = 1$

b) 1-Tap Wiener Filter

$\hat{d}[n] = w_0\, x[n]$

$$\underset{1\times1}{\underline{\underline{R}}_x}\,\underline{w} = \underset{1\times1}{r_{d\underline{x}}} \longrightarrow \sigma^2_{x[n]}\,w_0 = \overset{\overset{E\{d[n]x[n]\}}{\downarrow\quad\swarrow 1}}{r_d(0)}$$

$$w_0 = \frac{1}{1+\sigma_v^2}$$

$$J_{min}^{1-Tap} = r_d(0) - w_{opt}^T \cdot r_{dx}$$

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

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

$$\sigma_v^2 = 1 \longrightarrow J_{min} = \frac{1}{2} = 0{,}5$$

g) $\text{SNR (before filtering)} = \dfrac{E\{d[n]^2\}}{E\{v^2[n]\}} = \dfrac{r_d(0)}{r_v(0)} = \dfrac{1}{\sigma_v^2}\Big|_{\underset{\sigma_v^2=1}{\downarrow}} = 1$

$$10\log_{10}(SNR) = 0\ dB$$

$$\text{SNR (after filtering]} \Rightarrow \underset{\underset{\text{output}}{\uparrow}}{o[n]} = \underline{w}^T\underline{x}[n] = \underbrace{\underline{w}^T \cdot \underline{d}[n]}_{\text{signal}} + \underbrace{\underline{w}^T\underline{v}[n]}_{\text{noise}}$$

$$SNR(\text{output}) = \frac{E\{(\underline{w}^T\underline{d}[n])^2\}}{E\{(w^T\underline{v}[n])^2\}}$$

$$= \frac{\underline{w}^T\,\underline{\underline{R}}_d\,\underline{w}}{\underline{w}^T\,\underline{\underline{R}}_v \cdot \underline{w}}$$

$$SNR^{1-Tap}(\text{out}) = \frac{1/2 \cdot r_d(0) \cdot 1/2}{1/2 \cdot r_v(0) \cdot 1/2}$$

$$= SNR(\text{input}) \quad \text{OK since} \quad \hat{d}[n] = \frac{1}{2}x[n]$$

$$SNR^{2-Tap}(\text{out}) = \frac{[0{,}4048\ \ 0{,}2381]\begin{bmatrix}1 & 0{,}8\\ 0{,}8 & 1\end{bmatrix}\begin{bmatrix}0{,}4048\\ 0{,}2381\end{bmatrix}}{[0{,}4048\ \ 0{,}2381]\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\begin{bmatrix}0{,}4048\\ 0{,}2381\end{bmatrix}} \cong 2\,dB$$

what's the $\underline{w}$ filter maximizing SNR(out)?

$$\max_{\underline{w}} SNR(output) = \max_{\underline{w}} \frac{\underline{w}^T \underline{\underline{R}}_d \underline{w}}{\underline{w}^T \underline{\underline{R}}_v \underline{w}}$$

$$\downarrow$$

maximized by the generalized eigenvector of $\underline{\underline{R}}_d$ and $\underline{\underline{R}}_v$

$\underline{w}^{best}$ eigenvector of $\{\underline{\underline{R}}_v^{-1} \underline{\underline{R}}_d\}$ with max. eigenvalue.

$$SNR^{2\text{-}tap}(out) = \frac{\underline{w}^T \begin{bmatrix} 1 & 0{,}8 \\ 0{,}8 & 1 \end{bmatrix} \underline{w}}{\underline{w}^T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \underline{w}} = \frac{\underline{w}^T \begin{bmatrix} 1 & 0{,}8 \\ 0{,}8 & 1 \end{bmatrix} \underline{w}}{\|w\|^2}$$

$$eig\left\{\begin{bmatrix} 1 & 0{,}8 \\ 0{,}8 & 1 \end{bmatrix}\right\} = \left\{\frac{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}{\sqrt{2}}, \frac{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}{\sqrt{2}}\right\}$$

$$\underset{\lambda = 1{,}8}{\downarrow} \qquad \underset{\lambda = 0{,}2}{\downarrow}$$

$$\rightarrow \underline{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} 1/\sqrt{2} \rightarrow SNR(out) = \frac{w^T \lambda_{max} w}{1} = \lambda_{max} = 1{,}8$$

$$\underline{w} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} 1/\sqrt{2} \qquad = \quad \ldots \quad = \lambda_{min} = 0{,}2$$

$$\underline{w} = \alpha \begin{bmatrix} 1 \\ 1 \end{bmatrix} 1/\sqrt{2} + (1-\alpha^2)^{1/2} \begin{bmatrix} 1 \\ -1 \end{bmatrix} 1/\sqrt{2} \rightarrow = \alpha \lambda_{max} + (1-\alpha^2) \lambda_{min}$$

$$SNR^{Max\text{-}SNR\text{-}2\,Tap} = 1.8 = 2{,}55\ dB$$

In the notes for LMMSE given on the web; there are some further details.

Ergodicity: (Hayes)

$x[n]$ : WSS process.

Estimation of $\mu_x, r_x[k]$ of $x[n]$ → ?

Provided with a single realization

Mean Ergodicity.

$$\hat{\mu}_x = \frac{1}{N}\sum_{n=1}^{N} x[n]$$

A process is mean ergodic if

1- $E\{\hat{\mu}_x\} \to \mu_x$ (as $N \to \infty$)

2- $var\{\hat{\mu}_x - \mu_x\} \to 0$ (as $N \to \infty$)

Check:

$$1- \quad E\{\hat{\mu}_x\} = \frac{1}{N}\sum_{n=1}^{N} \underbrace{E\{x[n]\}}_{\mu_x} = \mu_x$$

$$2- \quad var\{\hat{\mu}_x\} = E\{(\hat{\mu}_x - \underset{\substack{\uparrow \\ E\{\hat{\mu}_x\}\text{ from (1)} \\ \| \\ \mu_x}}{\mu_x})^2\} = E\{(\frac{1}{N}\sum_{n=1}^{N}(x[n]-\mu_x))^2\}$$

$$= \frac{1}{N^2} E\{(\overbrace{[1\ 1\ 1 \ldots 1]}^{\underline{1}^T} \begin{bmatrix} x[1]-\mu_x \\ \vdots \\ x[N]-\mu_x \end{bmatrix})^2\}$$

$$= \frac{1}{N^2} E\{(\underline{1}^T \cdot \underline{x})(\underline{x}^T \cdot \underline{1})\}$$

$$= \frac{1}{N^2}\, \underline{1}^T \cdot \underline{\underline{C}}_x \cdot \underline{1}$$

$$= \frac{1}{N^2}[1\ 1 \ldots 1] \begin{bmatrix} C_x(0) & C_x(1) & \ldots & C_x(N-1) \\ C_x(1) & C_x(0) & \ldots & C_x(N-2) \\ \vdots & & & \\ C_x(N-1) & C_x(N-2) & & C_x(0) \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\frac{1}{N^2}\sum_{k=-(N-1)}^{N-1}(N-|k|)c_x[k] \to \frac{1}{N}\sum_{k=-(N-1)}^{N-1}\left(1-\frac{|k|}{N}\right)c_x[k]$$

Mean Ergodicity:

If $\frac{1}{N}\sum_{k=-(N-1)}^{N-1}\left(1-\frac{|k|}{N}\right)c_x[k] \xrightarrow[N\to\infty]{} 0$

iff mean ergodic.

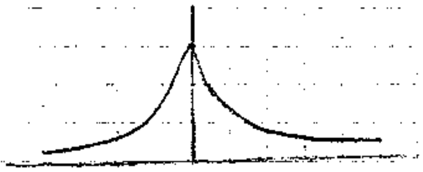
An equivalent Theorem (Papoulis) (Necessary and sufficient)

$$\frac{1}{N}\sum_{k=0}^{N-1}c_x[k] \xrightarrow[N\to\infty]{} 0 \xleftrightarrow{iff} \text{mean ergodic}$$

A sufficient condition (Papoulis]

$$\lim_{k\to\infty} C_x[k]=0 \longrightarrow \text{mean ergodic}$$

AR(1): $r_x(k)=\alpha^{|k|}$



Ergodicity in the auto-correlation:

$$\hat{r}_x[k]=\frac{1}{N}\sum_{n=0}^{N-1}\underbrace{(x[n]x[n-k])}_{z_k[n]}$$

$\underset{\uparrow}{z_k}[n]=x[n]x[n-k] \longrightarrow$ If $z_k[n]$ is mean ergodic

a new process

$x[n]$: auto-correlation ergodic

$$k=0 \to r_x[0]; \longrightarrow \hat{r}[0]=\frac{1}{N}\sum_{n=0}^{N-1}\underbrace{(x[n])^2}_{z[n]}$$

A necessary and sufficient cond.

$$\frac{1}{N}\sum_{k=0}^{N-1} C_2[k] \longrightarrow 0$$

$$\nwarrow$$

$$E\{x^2[n]\,x^2[n-k]\} \leftarrow 4^{th} \text{ order moments}$$

for Gaussian processes: $1^{st}, 2^{nd}$ order moments $\xrightarrow[\text{determine}]{}$ other moment

Then for Gaussian (Normal) Processes:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} C_x^2[k] = 0$$

$$\xrightarrow[\text{implies}]{} \text{ergodicity in the autocorrelation}$$

$$E\{x_1x_2x_3x_4\} = E\{x_1x_2\}E\{x_3x_4\} + E\{x_1x_3\}E\{x_2x_4\} + E\{x_1x_4\}E\{x_2x_3\}$$

Filtering: Causal operation, $(x[0], x[1], \_\_, x[N])$ is used to estimate a $d[N]$

Prediction: Estimation of a future sample from past samples.

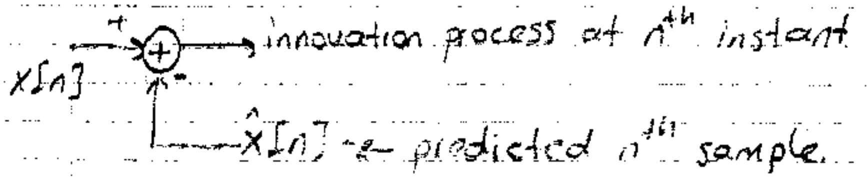
$$\hat{x}[n] = w_0 x[0] + w_1 x[1] + \_\_\_ + w_{n-1} x[n-1]$$

$\nearrow$ estimate of the next sample

Smoothing: Non-causal operation $\hat{d}[n]$ using samples

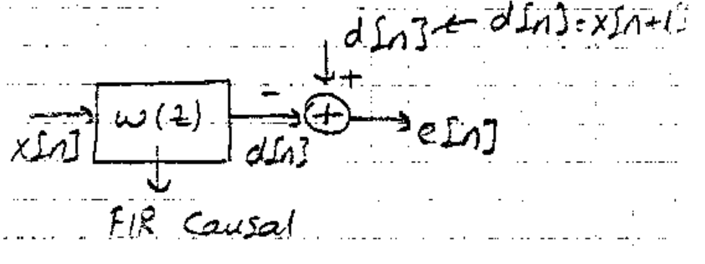
$$\{x[0], x[1], \_\_, x[n], x[n+1], \_\_, x[2n]\}$$

Recursive Est:



Linear Predictors:

$$d[n] = x[n+1];$$



$$E\{x[n]x[n]^T\} \to \underline{R}_x \overset{[w_0 \cdots w_{N-1}]}{\underline{w}} = r_{d\underline{x}}$$

$$\begin{bmatrix} r_x(0) & r_x(-1) & \cdots & r_x(-(N-1)) \\ r_x(1) & r_x(0) & \cdots & r_x(-(N-2)) \\ \vdots & & & \\ r_x(N-1) & r_x(N-2) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} w_0 \\ \vdots \\ w_{N-1} \end{bmatrix}_{N\times 1} = E\left\{ \overset{x[n+1]}{d[n]} \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-(N-1)] \end{bmatrix}_{N\times 1} \right\}$$

$$= \begin{bmatrix} r_x(1) \\ \vdots \\ r_x(N) \end{bmatrix}$$

Notes: 1- Single Step Predictor: $d[n] = x[n+1]$

Two Step Predictor: $d[n] = x[n+2] \to$ (only $r_{d\underline{x}}$ change

Multi Step Predictor - -

2- $\hat{x}[n] = \underset{?}{w_0} x[0] + \underset{?}{w_1} x[1] + \cdots + \underset{?}{w_{N-1}} x[n-1]$

So LMMSE predictor is constructed; that is

$J = E\{(d[n] - \hat{d}[n])^2\}$

is minimized for this problem $J = E\{(x[n+\underset{\text{Multi-step predicta}}{M}] - \underline{w}^T\underline{x}[n])^2\}$

is minimized

$J_{min} = r_d(0) - w_{opt}^T \cdot r_{dx}$ (General)

$\downarrow$ for prediction

$J_{min} = r_x(0) - w_{opt}^T r_{dx}$ $\to$ In some books, the prediction problem is constructed as

$\hat{x}[n] = -w_0 x[0] - w_1 x[1] \cdots - w_{N-1} x[n-1]$

for such a definition

Prediction error $\Rightarrow x[n] - \hat{x}[n] = x[n] + w_0 x[0] + w_1 x[1] + \cdots + w_{N-1} x[n-1]$

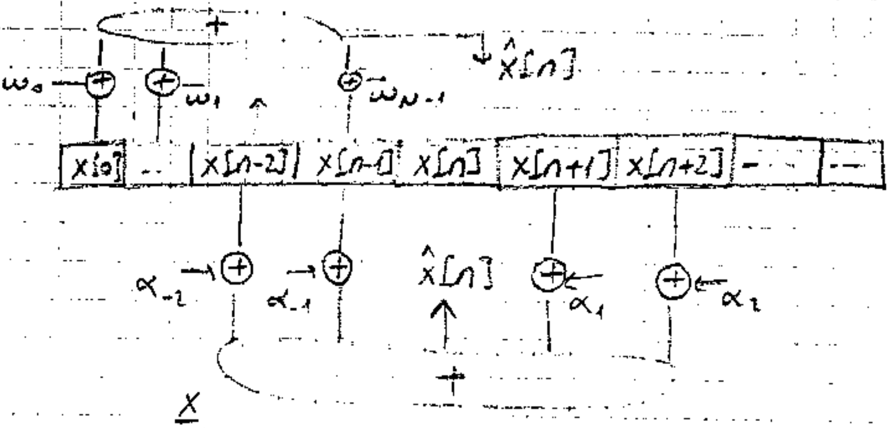
Prediction Error Filter $= x[n] + w_{n-1}x[n-1] + w_{n-2}x[n-2] + \_\_\_ + w_0 x[0]$

$$\hat{x}[n] - x[n] = \epsilon[n]$$

$$x[n] \longrightarrow \boxed{1 + w_{n-1}z^{-1} + \_\_\_ + w_0 z^{-n}} \longrightarrow \epsilon[n]$$

↖ prediction error of linear predictor.

Handout: (Manolakis' Book)



$$= \alpha^T \overbrace{[x[n-2]; x[n-1]; x[n+1]; x[n+2]]}^{\underline{x}}$$

$d[n] = x[n]$ $\quad \hat{d}[n] = \alpha_{-2}x[n-2] + \alpha_{-1}x[n-1] + \alpha_1 x[n+1] + \alpha_2 x[n+2]$

$$\downarrow$$

$$\underline{\underline{R}}_x \underline{\alpha} = r_{dx} \longrightarrow E\{xx^T\}\underline{\alpha} = E\{d[n]\underline{x}\}$$

$$\underline{\underline{R}}_x = \begin{bmatrix} r_x(0) & r_x(1) & r_x(3) & r_x(4) \\ r_x(1) & r_x(0) & r_x(2) & r_x(3) \\ r_x(3) & r_x(2) & r_x(0) & r_x(1) \\ r_x(4) & r_x(3) & r_x(1) & r_x(0) \end{bmatrix}$$

$$\uparrow$$

$$E\{x[n+1]x[n+2]\}$$

$$r_{d\underline{x}} = E\left\{x[n]\begin{bmatrix} x[n-2] \\ x[n-1] \\ x[n+1] \\ x[n+2] \end{bmatrix}\right\} = \begin{bmatrix} r_x(2) \\ r_x(1) \\ r_x(-1) \\ r_x(-2) \end{bmatrix}$$

$$\underline{\underline{R}}x \begin{bmatrix} \alpha_{-2} \\ \alpha_{-1} \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = r_{d\underline{x}} \longrightarrow \text{The solution has the form}$$

$$\left.\begin{matrix} \alpha_{-1} = \alpha_1 \\ \alpha_{-2} = \alpha_2 \end{matrix}\right\} \text{Let assume this is correct}$$

$$\underline{\underline{R}}x \begin{bmatrix} \alpha_{-2} \\ \alpha_{-1} \\ \alpha_1 \\ \alpha_2 \end{bmatrix} = r_{d\underline{x}}$$

$$\downarrow \text{ by assumption}$$

$$\underline{\underline{R}}x \begin{bmatrix} \alpha_2 \\ \alpha_1 \\ \alpha_{-1} \\ \alpha_{-2} \end{bmatrix} = r_{d\underline{x}}$$

$$\left.\underline{\underline{R}}x : \text{centro-symmetric matrix.}\right\} \rightarrow \text{Centro-symmetric matrices have efficient solution using Levinson Recursion}$$

$$\underline{\underline{R}}x = \underline{\underline{J}}\,\underline{\underline{R}}x\,\underline{\underline{J}} \swarrow J = \begin{bmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{bmatrix}$$

$$(\underline{\underline{J}}\,\underline{\underline{R}}_x J)\alpha = r_{dx} \rightarrow \underline{\underline{R}}_x \underline{\underline{J}}\alpha = \underline{\underline{J}} r_{dx} = r_{dx}$$

3- Backward Prediction:

$$\hat{x}[n] = [B_1 \;\; B_2 \;\; \ldots \;\; B_N] \begin{bmatrix} x[n+1] \\ x[n+2] \\ \vdots \\ x[n+N] \end{bmatrix} \; ; \quad J = E\{(x[n] - \hat{x}[n])^2$$

$$\underline{\underline{R}}x \, \underline{B} = r_{d\underline{x}}$$

$$\begin{bmatrix} r_x(0) & r_x(1) & \cdots & r_x(N-1) \\ r_x(1) & r_x(0) & \cdots & r_x(N-2) \\ \vdots & & & \vdots \\ r_x(N-1) & \cdots & & r_x(0) \end{bmatrix} \begin{bmatrix} B_1 \\ \\ \vdots \\ B_N \end{bmatrix} = \begin{bmatrix} r_x(1) \\ \vdots \\ \\ r_x(N) \end{bmatrix} \leftarrow$$

Equation system is identical to the earlier 1 step Forward Prediction

$x[n]$ : wss

$x[-n] = y[n] \rightarrow$ wss

IIR Wiener Filtering

$x[n]$ : wss process }
$d[n]$ : desired process } jointly wss.

$$\hat{d}[n] = \sum_{k'=-\infty}^{\infty} h[k']\, x[n-k'] \longleftarrow$$ Structure for estimator.

$$J = E\left\{ \underbrace{(d[n]-\hat{d}[n])}_{e[n]}{}^2 \right\} \rightarrow \frac{\partial J}{\partial h[k]} = 0$$

$k \in (-\infty, \infty)$

$$E\left\{ 2\, e[n] \underbrace{\frac{\partial e[n]}{\partial h[k]}}_{x[n-k]} \right\} = 0 \rightarrow E\{ e[n]\, x[n-k] \} = 0 \swarrow \text{Orthogonal}$$

$$\downarrow$$

$$r_{dx}(k) = \sum_{k'=-\infty}^{\infty} h[k']\, r_x(k-k') = 0 \quad \forall k \quad k \in (-\infty, \infty)$$

(DTFT
↓

$$S_{dx}(e^{j\omega}) - H(e^{j\omega}) \,.\, S_x(e^{j\omega}) = 0$$

$$\downarrow$$

$$H(e^{j\omega}) = \frac{S_{dx}(e^{j\omega})}{S_x(e^{j\omega})}$$

↗ IIR (Non-causal) Wiener Filter Freq. Response.

Filtering Application:

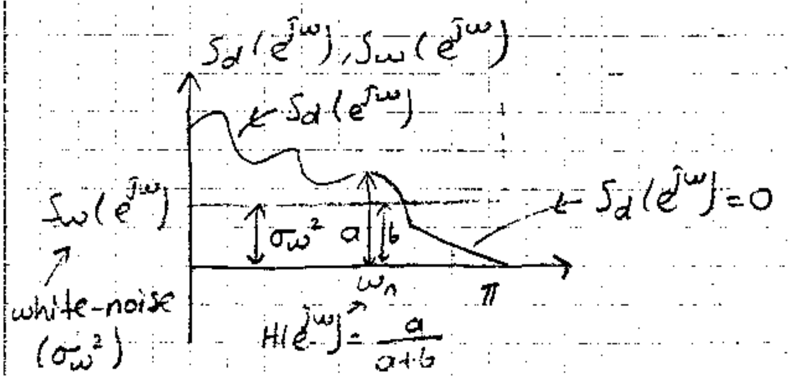
$$x[n] = d[n] + w[n] \leftarrow \begin{matrix} d[n] - \text{desired} \\ w[n] \leftarrow \text{noise} \end{matrix} \Bigg\} \text{Jointly WSS}$$

d[n], w[n] : uncorrelate

$$S_x(e^{j\omega}) = DTFT\{\underbrace{r_x(k)}_{r_d(k) + r_w(k)}\} = S_d(e^{j\omega}) + S_w(e^{j\omega})$$

$$S_{dx}(e^{j\omega}) = DTFT\{\underbrace{r_{dx}(k)}_{E\{d[n]\underbrace{x[n-k]}_{d[n-k]+w[n-k]}\} = r_d(k)}\} = S_d(e^{j\omega})$$

$$H^{IIR\text{-wiener Non-causal}}(e^{j\omega}) = \frac{S_d(e^{j\omega})}{S_d(e^{j\omega}) + S_w(e^{j\omega})} \leftarrow h^{IIR\text{-wiener-Non-causal}}[n] : \text{Even sequenc}$$



$$r_w(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \underbrace{S_w(e^{j\omega})}_{\sigma_w^2} d\omega$$

Let's calculate $J_{min}$ for IIR wiener estimator for filterin problem

IIR Non-Causal Wiener Filter

FIR Wiener

$\underline{R}_x \underline{w} = r_{d\underline{x}}$

$E\{ex\} = 0$

$$\underline{R}_x \underline{w} = r_{dx} \qquad \underline{w} = \begin{bmatrix} w(\infty) \\ \vdots \\ w(0) \\ \vdots \\ w(-\infty) \end{bmatrix}$$

$$H^{IIR-Non-Causal}(e^{j\omega}) = \frac{P_{dx}(e^{j\omega})}{P_x(e^{j\omega_0})}$$

Error Calculation for IIR Non-Causal Wiener Filter

$$J_{min} = E\{|d[n]-\hat{d}[n]|^2\}$$

$$\nwarrow \hat{d}[n] = h^{IIR-Non-causal}[n] * x[n] \quad ?$$

$$= E\{e[n](d[n]-\hat{d}[n])\} = E\{e[n]d[n]\} - E\{e[n]\hat{d}[n]\}$$

since

$E\{e[n]x[n-\Delta]\}=0 \ \forall \Delta$

$$= r_d(0) - E\{(\sum_{k=-\infty}^{\infty} h[k]x[n-k])d[n]\}$$

$$J_{min}^{NC-IIR} = r_d(0) - \sum_{k=-\infty}^{\infty} h^{IIR-NC}[k]\, r_{dx}(k)$$

$\int x(t)y^*(t) = \int X(f)Y^*(f$

$$\nearrow$$

$$J_{min}^{NC-IIR} = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_d(e^{j\omega})\,d\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi} H^{IIR-NC}(e^{j\omega})\,P_{dx}^*(e^{j\omega})\,d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(P_d(e^{j\omega}) - H^{IIR-NC}(e^{j\omega})\,P_{dx}^*(e^{j\omega})\right)d\omega$$

For Filtering Problems that is $x[n] = d[n] + \underset{\nwarrow \text{noise}}{v[n]}$

$$P_x(e^{j\omega}) = P_d(e^{j\omega}) + P_v(e^{j\omega})$$

and

$$P_{dx}(e^{j\omega}) = P_d(e^{j\omega})$$

$$H^{IIR-NC}(e^{j\omega}) = \frac{P_d(e^{j\omega})}{P_d(e^{j\omega}) + P_v(e^{j\omega})}$$

Then $J_{min}$ for the filtering problem

$$J_{min}^{NC-IIR} = \frac{1}{2\pi}\int \left(P_d(e^{j\omega}) - \frac{P_d(e^{j\omega})\,P_d(e^{j\omega})}{P_d(e^{j\omega}) + P_v(e^{j\omega})}\right)d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{P_d(e^{j\omega})\,P_v(e^{j\omega})}{P_d(e^{j\omega})+P_v(e^{j\omega})}\,d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} H^{NC-IIR}(e^{j\omega})\,P_v(e^{j\omega})\,d\omega$$

Special Case: $P_v(e^{j\omega}) = \sigma_v^2$ (white noise)

$$= \sigma_v^2 \cdot \frac{1}{2\pi}\int_{-\pi}^{\pi} H^{NC-IIR}(e^{j\omega})\,d\omega = \sigma_v^2\, h^{NC-IIR}[0]$$

Ex) Previously, $x[n] = d[n] + v[n]$, $r_d(k) = 0{,}8^{|k|}$

$r_v(k) = \delta[k]$

and $\hat{d}[n]$ is estimated using

a) 2-Tap FIR filter, $\underset{\nearrow \left(\text{2-Tap LMMSE est}\right)}{\hat{d}[n]} = 0{,}4048\, x[n] + 0{,}2381\, x[n-1]$

$J_{min} = 0{,}4048$

b) 1-Tap Filter $\underset{\nearrow \text{1-Tap wiener Filter}}{\hat{d}[n]} = 0{,}5\, x[n]$, $J_{min} = 0{,}5$

Then, how about the error of $h^{NC-IIR}[n]$?

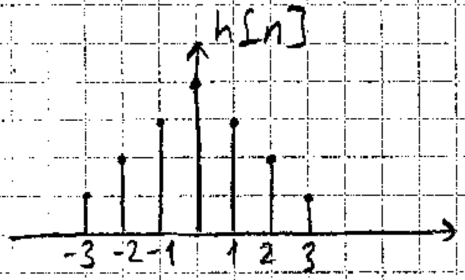
$$H^{NC-IIR}(e^{j\omega}) = \frac{P_d(e^{j\omega})}{P_d(e^{j\omega}) + P_v(e^{j\omega})}$$

$$H^{NC-IIR}(e^{j\omega}) = \frac{P_d(z)}{P_d(z)+P_v(z)} = \frac{Z\{0{,}8^{|k|}\}}{Z\{0{,}8^{|k|}\}+1} \qquad Z\{\alpha^{|k|}\} = \frac{1-\alpha^2}{(1-\alpha z^{-1})(1-\alpha z)}$$

$$= \left.\frac{1-\alpha^2/(1-\alpha z^{-1})(1-\alpha z)}{\dfrac{1-\alpha^2}{(1-\alpha z^{-1})(1-\alpha z)} + 1}\right|_{\alpha=0{,}8}$$

$$= 0{,}3\,\frac{(1-1/4)}{(1-\frac{1}{2}z^{-1})(1-\frac{1}{2}z)}$$

$$\overset{NC-IIR}{h[n]} = 0,3\left(\frac{1}{2}\right)^{|k|}$$



$$\longrightarrow \overset{IIR-NC}{J_{min}} = \overset{NC-IIR}{h[0]} = 0,3$$

$$\overset{NC-IIR}{J_{min}} = \sigma_v^2 \overset{NC-IIR}{h[0]}$$

The Causal IIR wiener Filter

$$\hat{d}[n] = \sum_{l=0}^{\infty} h[l]\, x[n-l]$$

$$J = E\{(d[n]-\hat{d}[n])^2\} \rightarrow \underline{\underline{R}}_x\, \underline{w} = r_{dx}$$

$$E\{e[n]x[n-k]\}=0 \quad \rightarrow \underbrace{\sum_{l=0}^{\infty} h[l]}_{\text{opt. Causal IIR filter coef.}} r_x(k-l) = r_{dx}(k) \quad k=\{0,1,\ldots\infty\}$$

$$k=\{0,1,\ldots,\infty\}$$

Special case: $x[n] = \epsilon[n]$ , $\epsilon$ : white noise unit variance

$$\downarrow$$

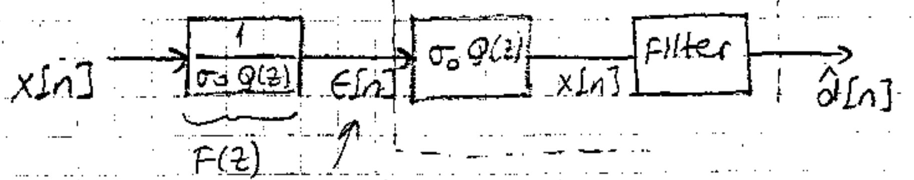
$$r_\epsilon[k] = \delta[k]$$

Then for this input $\rightarrow \overset{Causal\ IIR}{h[k]} = r_{d\epsilon}[k] \quad k=\{0,1,\ldots,\infty\}$

↑ Problem solved

More General Case: Assume it is possible to write $P_x(e^{j\omega})$ as a ratio of two trigonometric polynomials. ($P_x(z) = \frac{B_\varphi(z)}{A_\varphi(z)}$)

$$P_x(z) = \sigma_0^2\, \varphi(z)\, \varphi^*(1/z^*)$$



white-noise
unit var
as in special case

$$\xrightarrow{\text{Then}} \overset{\text{causal IIR}}{H(z)} = \frac{1}{\sigma_0\, Q(z)}\overset{F(z)}{\ }$$

$$r_{d\epsilon}(k) = E\{d[n]\,\overset{f[n]*x[n]}{\epsilon}[n-k]\}$$

$$= E\{d[n]\sum_{\ell} f(\ell)\,x[n-k-\ell]\}$$

$$r_{d\epsilon}(k) = \sum_{\ell=0}^{\infty} f[\ell]\,r_{dx}(k+\ell)$$

$$r_{d\epsilon}(k) = \sum_{\ell=-\infty}^{0} f^*[-\ell]\,r_{dx}(k-\ell]$$

$$r_{d\epsilon}(k) = f^*[-k] * r_{dx}[k]$$

$$P_{d\epsilon}(z) = F^*(1/z)\,.\,P_{dx}(z)$$

$$\overset{\text{causal-IIR}}{h[n]} = f[n] * \text{causal part of } (r_{d\epsilon}[n])$$

$$\overset{\text{causal IIR}}{H(z)} = F(z)\left[F^*(1/z)\,.\,P_{dx}(z)\right]_{+}$$

causal part

$$[1 + z^{-1} + 5z + 12z^{2}]_{+} = 1 + z^{-1}$$

Then

$$\overset{\text{causal IIR}}{H(z)} = \frac{1}{\sigma_0\,\varphi(z)}\left[\frac{P_{dx}(z)}{\sigma_0\,\varphi^*(1/z^*)}\right]_{+}$$

Note: $\varphi(z) = 1 + q_1 z^{-1} + q_2 z^{-2} + \ldots$ : monic polynomial in z.

Let's compare the filter with non-causal counter-part:

$$\overset{\text{non-causal-IIR}}{H(z)} = \frac{P_{dx}(z)}{P_x(z)} = \frac{P_{dx}(z)}{\sigma_0\,\varphi(z)\,\varphi^*(1/z^*)}$$

Ex: $x[n] = d[n] + v[n]$ $\quad r_d(k) = 0.8^{|k|}$

$r_v(k) = \delta[k]$

Min MSE

1- Tap sol. ⟶ 0,5

2- Tap sol. ⟶ 0,4048

$(0,\infty) \to$ Causal IIR $\to$ ? ⟶ 0,375

$(-\infty,\infty) \to$ Non-Causal IIR $\to$ 0,3

$$r_d(k) = 0,8^{|k|} \qquad d[n] = 0,8\, d[n-1] + \underset{\text{white-noise}}{w[n]} \qquad \sigma_w^2 = 0,36$$

$$r_d(k) = \frac{\sigma_w^2 \cdot 0,8^{|k|}}{1-(0,8)^2} = \frac{\sigma_w^2\, 0,8^{|k|}}{\underset{\underset{(0.6)^2}{\uparrow}}{0,36}} \Bigg\}\underset{\sigma_w^2}{=} 0,8^{|k|}$$

$$P_d(z) = \mathcal{Z}\{0,8^{|k|}\} = \frac{0,36}{(1-0,8z^{-1})(1-0,8z)}$$

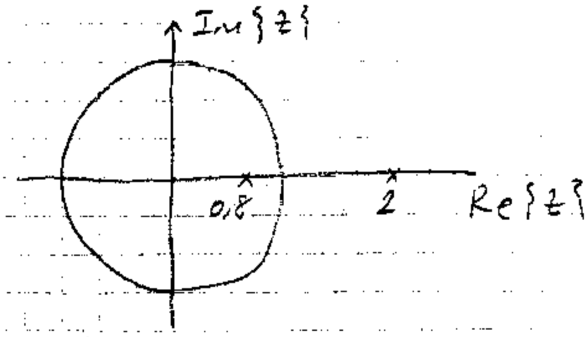
$$P_x(z) = P_d(z) + P_v(z) = 1,6\, \frac{(1-0,5z^{-1})(1-0,5z)}{(1-0,8z^{-1})(1-0,8z)}$$

$$P_x(z) = \sigma_0^2\, Q(z)\, Q^*(1/z^*)$$

causal IIR

$$H(z) = \frac{1}{\sigma_0^2 Q(z)} \left[\frac{P_{dx}(z)}{Q^*(1/z^*)}\right]_+ = \frac{1}{1,6}\, \frac{1-0,8z^{-1}}{1-0,5z^{-1}} \left[\frac{\frac{36}{(1-0,8z^{-1})(1-0,8z)}}{\frac{(1-0,5z)}{(1-0,8z)}}\right]$$

$$\frac{0,36}{(1-0,8z^{-1})(1-0,5z)} = \frac{\overset{0,6}{K}}{1-0,8z^{-1}} + \frac{L}{1-2z^{-1}}$$



$$\frac{1}{1,6} \cdot \frac{1-0,8z^{-1}}{1-0,5z^{-1}} \cdot \frac{0,6}{1-0,8z^{-1}} = \frac{3}{8}\, \frac{1}{1-0,5z^{-1}}$$

$$\overset{\text{causal-IIR}}{h[n]} = 0{,}375\left(\frac{1}{2}\right)^n u[n] \rightarrow \overset{\text{causal-IIR}}{H(z)} = \frac{3}{8}\,\frac{1}{1-0{,}5z^{-1}} = \frac{\hat{D}(z)}{X(z)}$$

$$\overset{\text{causal-IIR}}{J_{min}} = r_d(0) - \sum_{l=0}^{\infty} r_{dx}(l)\,h(l)$$

For our problem

$$J_{min} = 1 - \frac{3}{8}\sum_{l=0}^{\infty} 0{,}8^l \cdot 0{,}5^l = \frac{3}{8}$$

Summary of Last example

$$\left.\begin{aligned} d[n] &= 0{,}8\,d[n-1] + 0{,}6\,\overset{\swarrow \text{unit var. white noise}}{w[n]} \quad (r_d(k) = 0{,}8^{|k|}) \end{aligned}\right] \text{Process}$$

$$\left.\begin{aligned} x[n] &= d[n] + v[n] \quad (r_v(k) = \delta[k]) \end{aligned}\right] \text{Measurement process}$$

$$\left.\begin{aligned} \hat{d}[n] &= 0{,}375\sum_{l=0}^{\infty}\left[\frac{1}{2}\right]^l x[n-l] \\ \hat{d}[n] &= 0{,}5\,\hat{d}[n-1] + 0{,}375\,x[n] \end{aligned}\right\} \text{Estimator}$$

$$\hat{d}[n] = \underbrace{0{,}8\,\hat{d}[n-1]}_{\substack{\text{Prediction}\\ \text{of next sample}\\ \text{before } x[n] \text{ arrives}}} + \underset{\searrow \underline{\underline{K(n)}}}{\overset{\swarrow\text{weight}}{\boxed{\frac{3}{8}}}}\underbrace{(x[n] - 0{,}8\,\hat{d}[n-1])}_{\substack{\text{innovation of } x[n] \text{ given}\\ \text{all earlier } x[n] \text{ samples}\\ (\text{i.e } x[0], x[1], --, x[n-1])}}$$

They are identically the same

Causal Linear Prediction

Let's predict next sample of $x[n]$

$$\hat{x}[n+1] = \sum_{k=0}^{\infty} h[k]\,x[n-k]$$

$$d[n] = x[n+1] \rightarrow r_{dx}(k) = E\{x[n+1]\,x[n-k]\} = r_{dx}(k+1)$$

$$\rightarrow P_{dx}(z) = z \cdot P_x(z)$$

$$H(z) = \frac{1}{\sigma_0^2\,\varphi(z)}\left[\frac{P_{dx}(z)}{\varphi^*(1/z^*)}\right]_+ \qquad P_{dx}(z) = z\,(\sigma_0^2\,\varphi(z)\,\varphi^*(1/z^*))$$

$$= \frac{1}{\varphi(z)}\left[z\,\varphi(z)\right]_+ = \frac{z\,(\varphi(z) - 1)}{\varphi(z)} = z\left[1 - \frac{1}{\varphi(z)}\right]$$

$$z\,\varphi(z) = z\,(1+q_1 z^{-1} + q_2 z^{-2})$$

$$= z + q_1 + q_2 z^{-1} + \_\_\_$$

$$[z\,\varphi(z)]_+ = q_1 + q_2 z^{-1} + \_\_\_ = z\,(\varphi(z)-1)$$

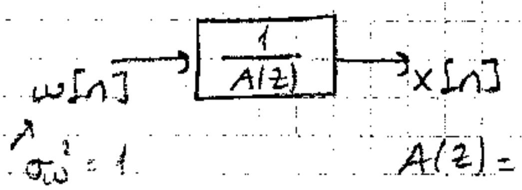
Ex: 7.3.3

$$x[n] = 0{,}9\,x[n-1] - 0{,}2\,x[n-2] + \overset{\text{white-noise } \sigma_w^2 = 1}{w[n]}$$

Find $\hat{x}[n+1]$ using all previous samples in history, and minimize the prediction error var.

$$P_x(z) = \sigma_0^2\,\varphi(z)\,\varphi^*(1/z^*) = \frac{1}{A(z)} \cdot \frac{1}{A^*(1/z^*)}$$

Process $x[n]$



$$A(z) = 1 - 0{,}9 z^{-1} + 0{,}2 z^{-2}$$

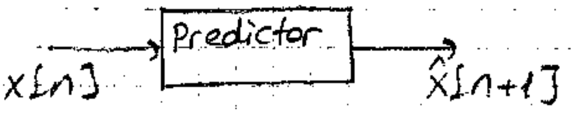
Predictor

$$\to H(z) = z\left(1 - \frac{1}{\varphi(z)}\right)$$

$$= z\,(1 - A(z))$$

$$= 0{,}9 - 0{,}2 z^{-1}$$

$$\hat{x}[n+1] = 0{,}9\,x[n] - 0{,}2\,x[n-1]$$



$$\hat{x}[n+1] = \sum_{l=0}^{N} h[l]\,x[n-l]$$

$$H(z) = \frac{1}{\sigma_0^2\,\varphi(z)} \left[\frac{P_{dx}(z)}{\varphi^*(1/z^*)}\right]_+$$

$$= \frac{1}{\varphi(z)}\left[z\,\varphi(z)\right] = \frac{z\,(\varphi(z)-1)}{\varphi(z)} = z\left[1 - \frac{1}{\varphi(z)}\right]$$

Reading Assignment: Read min. error calculation of Causal Linear Predictor at p. 365

$$\frac{1}{Q(z)}\left[z^{p}\,Q(z)\right]$$

Estimation of Non-Random Vector Using LS approach:

Linear observation Model:

$$\underbrace{\underline{x}}_{\substack{\text{observation}\\ \text{(also random)}}} = \overset{\underline{\underline{W}}\text{: known matrix}}{\underline{\underline{W}}}\;\underbrace{\underline{y}}_{\text{unknown non-random vector}} + \underbrace{\underline{\epsilon}}_{\text{noise: random noise vector}}$$

Let's propose the following estimator

$$\underline{\hat{y}} = \underline{\underline{K}}\,\underline{x} \quad \leftarrow ?$$

K should be selected to minimize $E\{\|\underline{y}-\underline{\hat{y}}\|^2\}$

$$\underline{\hat{y}}-\underline{y} = \underline{\underline{K}}\underline{x}-\underline{y} = \underline{\underline{K}}\,\underline{\underline{W}}\,\underline{y}+\underline{\underline{K}}\,\underline{\epsilon}-\underline{y} = (\underline{\underline{K}}\,\underline{\underline{W}}-\underline{\underline{I}})\underline{y}+\underline{\underline{K}}\,\underline{\epsilon}$$

$$\|\underbrace{\underline{\hat{y}}-\underline{y}}_{\underline{e}}\|^2 = \underline{e}^T\underline{e} = \|(\underline{\underline{K}}\,\underline{\underline{W}}-\underline{\underline{I}})\underline{y}+\underline{\underline{K}}\,\underline{\epsilon}\|^2$$

$E\{\|\underline{\hat{y}}-\underline{y}\|^2\}$ should be minimized — but this is a function of $\underline{y}$ unless $\underline{\underline{K}}\,\underline{\underline{W}}=\underline{\underline{I}}$

Then let's enforce $\underline{\underline{K}}\,\underline{\underline{W}}=\underline{\underline{I}}$ as a constraint.

$$E\{\|\underline{\hat{y}}-\underline{y}\|^2\} = E\{\|\underline{\underline{K}}\,\underline{\epsilon}\|^2\} = E\{(\underline{\underline{K}}\,\underline{\epsilon})^T(\underline{\underline{K}}\,\underline{\epsilon})\}$$
$$= E\{Tr\{(\underline{\underline{K}}\,\underline{\epsilon})(\underline{\underline{K}}\,\underline{\epsilon})^T\}\}$$
$$= \underbrace{Tr\{\underline{\underline{K}}\,\underline{\underline{R}}_{\epsilon}\,\underline{\underline{K}}^T\}}_{\text{cost}}$$

Then problem is

$$\min_{K}\; Tr\{\underline{\underline{K}}\,\underline{\underline{R}}_{\epsilon}\,\underline{\underline{K}}^T\} \quad \text{s.t.} \quad \underline{\underline{K}}\,\underline{\underline{W}}=\underline{\underline{I}}$$

Let's assume problem is solved for $\underline{\underline{K}}$.

Then

$$\underline{\hat{y}} = \underline{\underline{K}}\,\underline{x} \longrightarrow E\{\hat{y}\} = \overset{I}{\underline{\underline{K}}\,\underline{\underline{W}}}\,\underline{y} + \underline{\underline{K}}\,\overset{0}{E\{\epsilon\}}$$

$$= \underline{y} \rightarrow \text{unbiased estimator.}$$

Also

$$\begin{pmatrix}\text{Total Error}\\ \text{Variance}\end{pmatrix} = E\{\|\overset{K\begin{bmatrix}\epsilon_1\\ \epsilon_N\end{bmatrix}}{\hat{y}-y}\|^2\} = Tr\{\underline{\underline{K}}\,\underline{\underline{R}}_\epsilon K^T\}$$

minimized by "K".

So I have the best linear unbiased min. error estimator

(BLUE)

↑ ↑

Best Estimator

The solution of the problem can be found

1- Lagrange multiplier.

2- Matrix dif. formulation

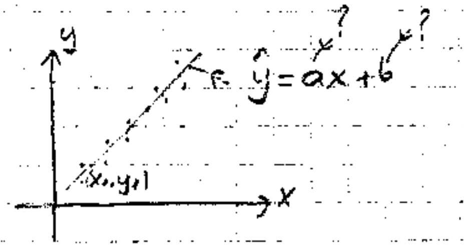
you get

$$\underline{\underline{K}} = (\underline{\underline{W}}^T \underline{\underline{R}}_\epsilon^{-1} \underline{\underline{W}})^{-1} \underline{\underline{W}}^T \cdot R_\epsilon^{-1}$$

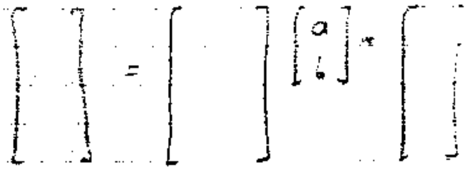
Special Cases:

1- $\underline{\underline{R}}_\epsilon = \underline{\underline{I}} \longrightarrow \underline{\underline{K}} = (\underline{\underline{W}}^T \underline{\underline{W}})^{-1} \underline{\underline{W}}^T$

estimator



W: tall-thin



$$\underline{\underline{\hat{y}}}_{LS} = \underline{\underline{K}}\,\underline{x} = (\underline{\underline{W}}^T \underline{\underline{W}})^{-1} \underline{\underline{W}}^T \underline{x}$$

$$\underline{\underline{X}} = \underline{\underline{W}}\,\underline{y} + \underline{\epsilon}$$

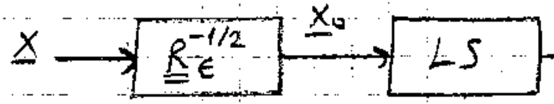
⟶ So LS solution is BLUE estimator for $\underline{\underline{R}}_\epsilon = \underline{\underline{I}}$ white noise.

2- $\underline{\underline{R}}_\epsilon = \underset{\substack{\uparrow \\ \text{positive scalar}}}{\alpha} \underline{\underline{I}} \longrightarrow \epsilon$ : white noise with variance $= \alpha$

$$\longrightarrow K = (\underline{\underline{W}}^T \underset{\frac{1}{\alpha}\underline{\underline{I}}}{\boxed{\underline{\underline{R}}_\epsilon^{-1}}} \underline{\underline{W}})^{-1} W^T \underset{1/\alpha \underline{\underline{I}}}{\boxed{\underline{\underline{R}}_\epsilon^{-1}}} = (\underline{\underline{W}}^T \underline{\underline{W}})^{-1} \underline{\underline{W}}^T$$

$\hookrightarrow$ So LS solution (as in ①) is the BLUE for stationary white noise.

3- $\underline{\underline{R}}_\epsilon \neq \alpha \underline{\underline{I}} \rightarrow \underline{\underline{R}}_\epsilon = \underline{\underline{R}}_\epsilon^{1/2} R_\epsilon^{1/2}$



$\hat{\underline{y}} = (\underline{\underline{W}}_0^T \underline{\underline{W}})^{-1} \underline{\underline{W}}_0 \underline{x}_0^T$

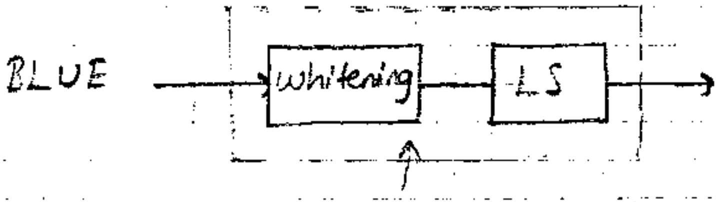
$= (\underline{\underline{W}}^T \underline{\underline{R}}_\epsilon^{-1} \underline{\underline{W}})^{-1} \underline{\underline{W}}^T \underline{\underline{R}}_\epsilon^{-1} \underline{X}$

$\uparrow$ The BLUE for $\underline{X}$ with $\underline{\underline{R}}_\epsilon$

$\underline{x}_0 = \underline{\underline{R}}_\epsilon^{-1/2} \underline{X} = \underline{\underline{R}}_\epsilon^{-1/2} \underline{\underline{W}} \underline{y} + \underline{\underline{R}}_\epsilon^{-1/2} \underline{\epsilon}$

$X = \underline{\underline{W}} \underline{y} + \underline{\epsilon}$

$\underline{\underline{W}}_0$ $\quad \underline{\epsilon}_0$ : unit var. white noise vector



Comparison of Non-Random and Random Vector Estimation

| | |
|---|---|
| $\underline{X}$ : Non-Random | $\underline{X}$ : Random Vector |
| $\hat{\underline{y}} = (\underline{\underline{W}}^T \underline{\underline{R}}_\epsilon^{-1} \underline{\underline{W}})^{-1} \underline{\underline{W}}^T \underline{\underline{R}}_\epsilon^{-1} X$ | $\left.\begin{matrix} \underline{\underline{R}}_x \\ \underline{\underline{R}}_y, \underline{\underline{R}}_{xy} \end{matrix}\right\}$ known $\Big\vert$ $\min E\{\lVert \underline{y} - \underset{\substack{\uparrow \\ \underline{\underline{K}}_{MMSE}^H \underline{X}}}{\hat{\underline{y}}} \rVert^2\}$ |
| | $\underline{\underline{R}}_x \underline{\underline{K}}_{MMSE} = \underline{\underline{R}}_{xy}$ |

$$\underline{X} = \underline{\underline{W}} \underline{y} + \underline{\epsilon} \longrightarrow \underline{\underline{R}}_X = \underline{\underline{W}} R_y \underline{\underline{W}}^T + \underline{\underline{R}}_\epsilon$$

$$\hookrightarrow \underline{\underline{R}}_{xy} = E\{\underline{X} y^T\} = \underline{\underline{W}} \underline{\underline{R}}_y$$

$$\underline{\underline{R}} X \underline{\underline{K}}_{MMSE} = \underline{\underline{R}}_{xy}$$

$$\underline{\underline{K}}_{MMSE} = (\underline{\underline{W}}\,\underline{\underline{R}}_y \underline{\underline{W}}^T + \underline{\underline{R}}_\epsilon)^{-1} \underline{\underline{W}}\,\underline{\underline{R}}_y$$

$$\hat{\underline{y}}_{MMSE} = \underline{\underline{K}}^H_{MMSE} \underline{X} = \underline{\underline{R}}_y \underline{\underline{W}}^T (\underline{\underline{W}}\,\underline{\underline{R}}_y \underline{\underline{W}}^T + R_\epsilon)^{-1} X$$

$$= (\underline{\underline{W}}^T \underline{\underline{R}}_\epsilon^{-1} \underline{\underline{W}} + \underline{\underline{R}}_y^{-1})^{-1} \underline{\underline{W}}^T \underline{\underline{R}}_\epsilon^{-1} \underline{X}$$

Apply matrix inversion lemma

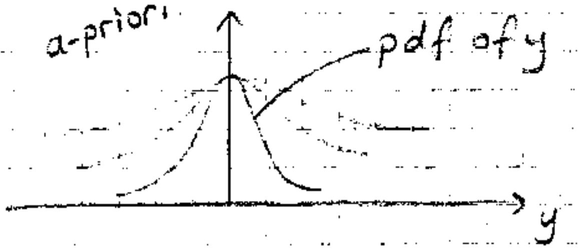
If $\underline{\underline{R}}_y^{-1} \to \underline{\underline{0}}$ or $\underline{\underline{R}}_y \to \infty$

$$\begin{bmatrix} \infty & \infty & \infty \\ \infty & \infty & \infty \\ \infty & \infty & \infty \end{bmatrix}$$

Then $\hat{y}_{MMSE} \to \hat{y}_{non-random}$

Ex: $y: 1\times1 \to \underline{\underline{R}}_y = \sigma_y^2 \to \infty$

$\hat{y}_{MMSE} \to \hat{y}_{non-random}$



$\hat{y}_{MMSE} \to \hat{y}_{non-random}$ when signal have high power

Karhunen-Loeve Transform (KL Transform)

Remember Orthogonal expansion in N-dim space

$$\underset{N\times1}{\underline{X}} = \sum_{k=1}^{N} \alpha_k \underline{\phi}_k \qquad \phi_k : k = \{1, \_\_, N\}$$

$\underline{\phi}_k \perp \underline{\phi}_l \to \underline{\phi}_k^T \cdot \phi_l = 0,\ k = l$

$\| \phi_k \|^2 = \phi_k^T \cdot \phi_k = 1$

$\to \phi_k$'s form an orthonormal set.

$$\underline{\phi}_l^T \underline{X} = \sum_{k=1}^{N} \alpha_k \underbrace{\phi_l^T \phi_l}_{\delta[k-l]} = \alpha_l$$

$$\alpha_\ell = \underline{\phi}_\ell^T \underline{x} = \langle \phi_\ell , \underline{x} \rangle$$

$$\underline{x} = \underbrace{[\, \underline{\phi}_1 \ \ \underline{\phi}_2 \ \dots \ \underline{\phi}_N \,]}_{\underline{\underline{\phi}}} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}$$

$$\underline{\alpha} = \underline{\underline{\phi}}^{-1} \underline{x} \longrightarrow \underline{\alpha} = \underline{\underline{\phi}}^T . x$$

$$= \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \vdots \\ \phi_N^T \end{bmatrix} \underline{x} = \begin{bmatrix} \underline{\phi}_1^T \underline{x} \\ \vdots \\ \underline{\phi}_N^T \underline{x} \end{bmatrix} \begin{matrix} \leftarrow \alpha_1 \\ \\ \rightarrow \alpha_N \end{matrix}$$

$\underline{x}$ is to be expanded in an L-dim space, $L < N$. It's clear that there is an approximation error when $\underline{x}$ projected on an L-dim subspace.

$$\underset{\overset{\nearrow}{N \times 1}}{\underline{x}} \cong [\, \underline{\psi}_1 \ \ \underline{\psi}_2 \ --- \ \underline{\psi}_L \,] \begin{bmatrix} B_1 \\ \vdots \\ B_L \end{bmatrix}$$

$\underline{x} = \underline{\underline{\psi}} \, \underline{B}$ → $\underline{x} \in$ Range ( $\psi$ )

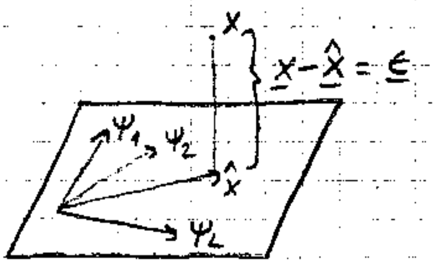
no problems

equation is exactly satisfied

→ inconsistent eq. system.

For the inconsistent case, we may use LS solution which is

$$B_{LS} = ( \underline{\underline{\psi}}^T \underline{\underline{\psi}} )^{-1} \underline{\underline{\psi}}^T \underline{x}$$



Then LS solution minimizes

$$\| x - \hat{x} \|^2$$

by chosing $\underline{B}$ for a given $\underline{\underline{\psi}}$.

In the L-dim app. problem, assume that $\underline{\Psi}_k$'s are not provided.

How to select a good $\underline{\Psi}_k$ $k = \{1, \_\_, L\}$ so that L-Dim projection error is minimized in Euclidean norm sense.

In the deterministic problem,

$$\hat{X} = \sum_{k=1}^{L} B_k \underline{\Psi}_k \quad \Big\} \ \underline{\Psi}_k \text{ is an orthonormal set.}$$

$$\|\underline{X} - \hat{\underline{X}}\|^2 = \|\underline{X} - \underline{\underline{\Psi}}\,\underline{B}\|^2 \qquad \underline{B} = B_{LS}$$

$$= \|\underline{X} - \underline{\underline{\Psi}}\,\underline{\underline{\Psi}}^T \underline{X}\|^2$$

$$= \|\underbrace{(\underline{\underline{I}} - \underline{\underline{\Psi}}\,\underline{\underline{\Psi}}^T)}_{\underline{\underline{P}}_{\Psi\perp}} \underline{X}\|^2 \leftarrow \text{select } \Psi \text{ s.t. the projection error is minimized.}$$

$$= \|\underline{\underline{P}}_{\Psi\perp} \underline{X}\|^2$$

$$\Psi_1 = \underline{X} \longrightarrow \underbrace{\underline{e} = 0}_{\text{projection error}}$$

So this projection error containing $\underline{X}$ has zero error; but has no practical use in signal approximation.

$$E\{\|\underline{X} - \underline{\Psi}\,\underline{B}\|^2\} \qquad \underline{X} : X \text{ is a random vector}$$

$$= E\{\|\underline{\underline{P}}_{\Psi_k\perp} \underline{X}\|^2\} \qquad \underline{\underline{P}}: \text{Projection matrix}$$

$$= E\{\underline{X}^T \underline{\underline{P}}_{\Psi}^T \underline{\underline{P}}_{\Psi} \underline{X}\} \qquad \underline{\underline{P}}^2 = \underline{\underline{P}}$$

$$= E\{\underline{X}^T \underline{\underline{P}}_{\Psi_1\perp} X\} \qquad \underline{\underline{P}}^T = \underline{\underline{P}} \searrow (I - \Psi\Psi^T)^T$$

$$\underline{\underline{P}}_{\Psi\perp} : \text{span}\{\Psi_1^{\perp}, \Psi_2^{\perp}, \_\_, \Psi_{N-L}^{\perp}\}$$

$\Psi$ : L Dim space

$\Psi^{\perp}$ : (N-L) Dim space

$P_{\Psi\perp}$ : span $\{e_1, \_\_, e_{N-L}\}$ → $e_k$'s are orthonormal

$$\underline{P}_{\psi\perp} = [\underline{e}_1, \text{——}, \underline{e}_{N-L}] \begin{bmatrix} \underline{e}_1 \\ | \\ \underline{e}_{N-L} \end{bmatrix}$$

$$\underline{\underline{P}}_{\psi\perp} = \sum_{k=1}^{N-L} \underline{e}_k \, \underline{e}_k^T$$

Inserting the def. for $\underline{P}_{\psi\perp}$

$$= E\left\{ \sum_{k=1}^{N-L} |\underline{e}_k^T \underline{x}|^2 \right\}$$

$$= \sum_{k=1}^{N-L} E\left\{ |\underline{e}_k^T \underline{x}|^2 \right\}$$

$$= \sum_{k=1}^{N-L} \underline{e}_k^T \underline{\underline{R}}_x \underline{e}_k$$

$\uparrow$

By selecting $e_k$'s, this should be minimized.

So Let's say $e_k$'s are eigenvectors of $\underline{\underline{R}}_x$ $(N \times N)$
$\hookrightarrow$orthonormal

$$e_1, e_2, e_3, \text{——}, e_N \longrightarrow \lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_N \geq 0$$

$$\text{then} \quad \underline{e}_k^T \cdot (\underbrace{\underline{\underline{R}}_x e_k}_{\lambda_k e_k}) = \lambda_k \underbrace{\|e_k\|}_{1}{}^2 = \lambda_k$$

So to minimize the app. error; $P_{\psi\perp} = \text{span}\{e_1, \text{—}, e_{N-L}\}$ should be formed by the eigenvectors having the smallest eigenvalues.

Or $\underline{\underline{P}}_\psi$ space should be formed the eigenvalues of $\underline{\underline{R}}_x$ having largest eigenvalues.

If $\underline{x}$ : r. vector having $\underline{\underline{R}}_x$ matrix, then optimal $\hat{\underline{x}}$

i) in 1-Dim space should be generated using $\underline{e}_1$

ii) in 2-Dim space " " " " $\underline{e}_1, \underline{e}_2$

⋮

where $\underline{\underline{R}}_x e_k = \lambda_k \underline{e}_k$

and Then minimized error for L-Dim approximation

$$\text{Error} = \sum_{k=L+1}^{N} \lambda_k$$

Ex) Therrien Section 4.7
?

$$\underline{\underline{R}}_x = \begin{bmatrix} 4 & 2.4 & 1.4 & 0.8 \\ \vdots & \vdots & \vdots & \vdots \\ & & & \end{bmatrix}_{4\times4}$$

$\underline{X}_{4\times1}$

$$\underline{\underline{R}}_x \cong \frac{1}{S} \sum_{k=1}^{S} X_k X_k^T$$

1- Term approximation (1-D)

2- Term

$\vdots$

4- Term (4-D)