$$E(\underline{y} \mid \underline{z}) = E(\underline{y}) + \underbrace{C_{\underline{y}\underline{z}} C_{\underline{z}\underline{z}}^{-1}}_{\to M} (\underline{z} - E(\underline{z}))$$

$\underline{\Theta}$

- optimal for jointly Gaussian RVs $(\underline{y}, \underline{z})$
- linear estimator $\sim M \cdot (\underline{z} - E(\underline{z}))$
- Complexity (computional) is finding C_{zz}^{-1}

linear estimator : $\Longleftrightarrow$ Gaussian $E[\underline{y} / \underline{z}]$

$\underline{\Theta}$

$$\hat{\Theta} = \sum_{i=1}^{n} a_i x_i + \underbrace{b}_{\text{bias } \mu_\theta}$$

$\boxed{\text{Scalar}}$

$$E((\hat{\theta} - \theta)^2) = E((\overbrace{\sum_{i=1}^{n} a_i x_i + b}^{\hat{\Theta}}) - \Theta)^2)$$

$$\frac{d}{db}(\ ) = \cancel{2} E((\hat{\theta} - \theta)) \cdot \overbrace{\frac{d}{db}\hat{\theta}}^{=1} = 0$$

$$= E\Big(\underbrace{\sum_{i=1}^{n} a_i x_i + b}\ \underbrace{-\theta}\Big) = 0$$

$$b = E[\theta] - \sum_{i=1}^{n} a_i E[x_i] \rightarrow 0$$

Find coefficients via

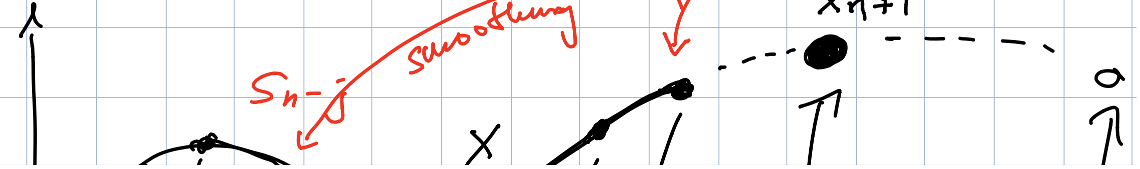
$$\frac{\partial}{\partial a_i} E\big((\hat{\theta}-\theta)^2\big) = \cancel{2}\, E\Big[(\hat{\theta}-\theta)\cdot \frac{\partial \hat{\theta}}{\partial a_i}\Big] = 0$$

$$= E\Big[\underbrace{\overbrace{\Big(\sum_{i=1}^{n} a_i x_i + b - \theta\Big)}^{\hat{\theta}\,-\,\theta}}_{\text{estimator error}} \cdot \underbrace{x_i}_{\text{data}}\Big] = 0$$

$$E[e \cdot x_i] = 0$$

orthogonality principle

<u>Linear</u> <u>predictor</u> (specialize)



$x_i = s_i + n_i$ $\longrightarrow$ 1 step - - - - - - n steps

1) $\Theta = x_{n+1}$

$f_{x_{n+1}, \underline{x}} \longrightarrow \boxed{f_{\Theta|\underline{x}}(\Theta, \underline{x})}$ $\quad \cancel{\boxed{\Theta} \to \underline{x} \to}$

2) Θ changes with $n \Rightarrow$ dynamic

OP:

$$E\left((\hat{x}_{n+1} - x_{n+1}) \cdot \underline{x} \right) = \underline{0}$$

$$E\left(\left(\sum_{i=0}^{N-1} a_i x_{n-i} - x_{n+1} \right) \cdot \underbrace{x_i }_{i = n-N+1,\, n} \right) = 0$$

predictor order

N unknowns

$$\underline{a} = R^{-1} \cdot \underline{r}$$

$R_{ij} = E[x_i x_j]$

$x_i = a + n_i$

$x_i = s_i + n_i$

$$\begin{bmatrix} R_{nn} & R_{n,n-1} & \cdots \end{bmatrix}$$

$$R = \begin{bmatrix} R_{n-1,n} & R_{n-1,n-1} & \cdots & \\ \vdots & & \ddots & \vdots \\ \vdots & & & \\ & \cdots & & R_{n-N+1,\,n-N+1} \end{bmatrix}$$

$E[X_i X_j]$

$n-N+1 < i,j < n$

$$\underline{r} = [R_{n+1,n} \quad R_{n+1,n-1} \quad \cdots\cdots \quad R_{n+1,\,n-N+1}]$$

$E[X_{n+1} \cdot X_i]$

Assume a stationary process:

$$\underbrace{E[X_i X_j]}_{R_\Delta} = E[X_{i+n} X_{j+n}]$$

$$= E[X_i X_{i+(j+i)}]$$

$$= E[X_i X_{i+\Delta}]$$

Wiener-Hopf Equations:

$$\underline{a} = R^{-1} \cdot \underline{r}$$

avoid at all cost!

$$R\underline{a} = \underline{r}$$

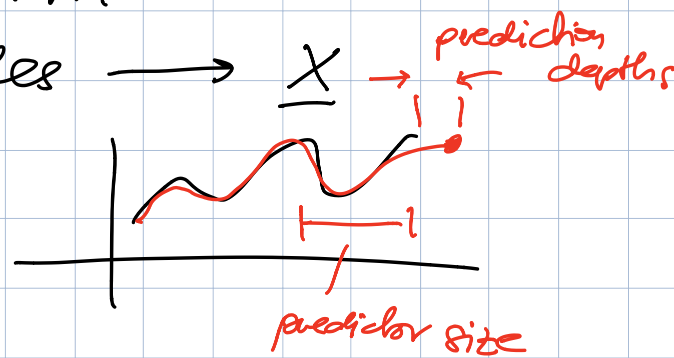
$$\begin{bmatrix} R_0 & R_1 & \cdots & R_{N-1} \\ R_1 & R_0 & R_1 & \cdots \\ R_2 & R_1 & R_0 & R_1 & \cdots \end{bmatrix} \cdot \underline{a} = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} \vdots & & \ddots & \\ R_{N-1} & \cdots & \cdots & R_0 \end{bmatrix}}_{\text{Levinson-Durbin Algorithm ??}} \qquad \begin{bmatrix} R_N \end{bmatrix}$$

complexity of $R^{-1} \sim O(N^3)$

Linear Predictor. m

fed samples $\longrightarrow \underline{X}$



How do we find

$$R_\Delta = E[X_i X_{i+\Delta}] = ?$$

1) assume process stationary

2) sample correlation

$$R_\Delta \cong \frac{1}{L} \sum_{i=0}^{L-1} X_i X_{i+\Delta} = \hat{R}_\Delta$$

learning phase

theory!

practice

Wiener-Hopf

$\Rightarrow a$

$\Rightarrow \hat{X}_{n+1} = \sum_{i=0}^{N-1} a_i X_{n-i}$