Chapter 3 Optimal Signal Processing

I. Optimal Estimation

1) Applications of Estimation Theory:

   Estimate moments, spectra, signals, system model parameters, …

2) Problem Description:

   Estimate one or more parameters from a set of data measurements (or observations).

3) Classification of Estimation Problems:

   (a) Classical parameter estimation or deterministic parameter estimation – Estimate deterministic parameters (unknown constants) from observed random data (observations).

   (b) Bayesian parameter estimation or random parameter estimation – Estimate random parameters from observed random data (observations), with prior knowledge of an *a priori* density of the random parameters, $f(\theta)$.

4) Three Estimation Procedures to be Discussed:

   (a) Maximum A Posteriori (MAP) Estimation

   (b) Maximum Likelihood (ML) Estimation

   (c) Minimum Mean-Squared Error (MMSE) Estimation

5) MAP Estimation:

   (a) Single-parameter case:

   Estimate the value of a random parameter, θ, from a set of random observations, $x(n)$, for n = 0, 1, 2, …, M–1. Assume that the pdf of θ is $f(\theta)$. Find the value of θ that maximizes $f(\theta|\mathbf{x})$, where $\mathbf{x}=\left[x(0),x(1),\ldots,x(M-1)\right]^T$.

   $f(\theta)$: *a priori* density.

   $f(\theta|\mathbf{x})$: *a posteriori* density.

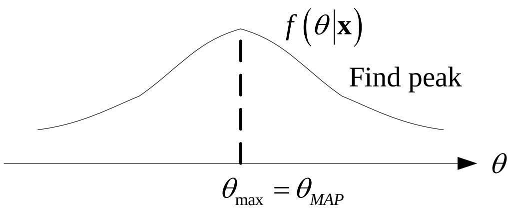
$$\frac{\partial f(\theta|\mathbf{x})}{\partial\theta}=0 \Rightarrow \text{the solution is } \theta_{MAP} \tag{1}$$

$$\text{maximize } f(\theta|\mathbf{x}) \Leftrightarrow \text{ maximize } \ln\left[f(\theta|\mathbf{x})\right] \tag{2}$$

$$\begin{aligned}\Rightarrow \frac{\partial\left[\ln f(\theta|\mathbf{x})\right]}{\partial\theta} &= \frac{\partial}{\partial\theta}\left[\ln\frac{f(\theta,\mathbf{x})}{f(\mathbf{x})}\right]=\frac{\partial}{\partial\theta}\left[\ln\frac{f(\mathbf{x}|\theta)f(\theta)}{f(\mathbf{x})}\right]\\ &=\frac{\partial}{\partial\theta}\left[\ln f(\mathbf{x}|\theta)+\ln f(\theta)-\ln f(\mathbf{x})\right]\end{aligned} \tag{3}$$

$\because f(\mathbf{x})$ is not dependent on θ

$$\therefore \frac{\partial}{\partial \theta}\left[\ln f(\mathbf{x}|\theta) + \ln f(\theta)\right] = 0 \tag{4}$$



■ **Figure 3.1** Graphic representation of MAP estimation.

(b) Multiple-parameter case:

Find a vector $\boldsymbol{\theta}_{MAP}$ that maximizes $f(\boldsymbol{\theta}|\mathbf{x})$, where $\boldsymbol{\theta} = [\theta_0, \theta_1, \ldots, \theta_{L-1}]^T$, $\mathbf{x} = [x(0), x(1), \ldots, x(M-1)]^T$, and the pdf of $\boldsymbol{\theta}$ is given by $f(\boldsymbol{\theta})$.

$$\frac{\partial f(\boldsymbol{\theta}|\mathbf{x})}{\partial \theta_i} = 0 \text{ for } i = 0,\ 1,\ 2, \ldots,\ L-1 \tag{5}$$

$$\Rightarrow \frac{\partial \ln\left[f(\boldsymbol{\theta}|\mathbf{x})\right]}{\partial \theta_i} = 0 \text{ for } i = 0,\ 1,\ 2, \ldots,\ L-1 \tag{6}$$

$$\Rightarrow \frac{\partial}{\partial \theta_i}\left[\ln f(\mathbf{x}|\boldsymbol{\theta}) + \ln f(\boldsymbol{\theta})\right] = 0 \text{ for } i = 0,\ 1,\ 2, \ldots,\ L-1 \tag{7}$$

6) ML Estimation

(a) Single-parameter case:

Estimate the value of an unknown (deterministic) constant θ from a set of random observations, $\mathbf{x} = [x(0), x(1), \ldots, x(M-1)]^T$.

Likelihood function: $L_\theta = f(\mathbf{x};\theta)$ (pdf of $\mathbf{x}$, given θ).

(For comparison with MAP estimation, we image that $f(\mathbf{x};\theta) = f(\mathbf{x}|\theta) f(\theta) = f(\mathbf{x}|\theta)\cdot 1$)

ML estimation can be described as follows:

Find the value of θ that maximizes $L_\theta = f(\mathbf{x};\theta)$.

$$\frac{\partial}{\partial \theta} L_\theta = 0 \tag{8}$$

$$\Rightarrow \frac{\partial}{\partial \theta}\ln(L_\theta) = \frac{\partial}{\partial \theta}\ln\left[f(\mathbf{x};\theta)\right] = 0 \tag{9}$$

$$\Leftrightarrow \frac{\partial}{\partial \theta}\ln\left[f(\mathbf{x}|\theta)\right] = 0 . \tag{10}$$

(b) Multiple-parameter case:

$$L_\theta = f(\mathbf{x};\boldsymbol{\theta}),\ \ \boldsymbol{\theta} = [\theta_0, \theta_1, \ldots, \theta_{L-1}]^T \tag{11}$$

$$\frac{\partial}{\partial \theta_i} L_\theta = 0,\ \ i = 0,\ 1,\ 2, \ldots,\ L-1. \tag{12}$$

(c) Properties of the ML estimation process:

(i) Sufficiency:

If a sufficient statistic for θ exists, then the ML estimate will be a sufficient statistic.

(ii) Efficiency:

- There exists a lower bound on the variance attainable by an unbiased estimator for an unknown θ.
- An estimator whose variance attains the lower bound and is sufficient is called fully efficient.
- If a fully efficient statistic for θ exists, then the ML estimate is that statistic.

(iii) Gaussianity:

ML estimates are asymptotically Gaussian.

7) Examples

Example 1: Single parameter, single observation

$$x = \theta + w,\ \text{ where } w \sim N(0, \sigma_w^2).$$

➢ ML estimation:

$$\frac{\partial}{\partial \theta} \ln L_\theta = 0$$

$$\frac{\partial}{\partial \theta}\left[\ln f(x;\theta)\right] = 0$$

$$\because f(x;\theta) = \frac{1}{\sqrt{2\pi\sigma_w^2}} e^{-\frac{(x-\theta)^2}{2\sigma_w^2}}$$

$$\therefore \ln f(x;\theta) = -\frac{1}{2}\ln\left(2\pi\sigma_w^2\right) - \frac{(x-\theta)^2}{2\sigma_w^2}$$

$$\Rightarrow \frac{\partial}{\partial \theta}\left[\ln f(x;\theta)\right] = \frac{x-\theta_{ML}}{\sigma_w^2} = 0$$

$$\Rightarrow \theta_{ML} = x$$

The best estimate in the ML sense is the raw data.

➢ MAP estimation: $\frac{\partial}{\partial \theta}\left[\ln f(x|\theta)+\ln f(\theta)\right]=0$

Assume

$$f(\theta)=\frac{1}{\sqrt{2\pi\sigma_\theta^2}}e^{-\frac{(\theta-\bar{\theta})^2}{2\sigma_\theta^2}}$$

and the conditional pdf of x given θ is

$$f(x|\theta)=\frac{1}{\sqrt{2\pi\sigma_w^2}}e^{-\frac{(x-\theta)^2}{2\sigma_w^2}}$$

$$\ln f(\theta)=-\frac{1}{2}\ln\left(2\pi\sigma_\theta^2\right)-\frac{(\theta-\bar{\theta})^2}{2\sigma_\theta^2}$$

$$\frac{\partial}{\partial \theta}\left[\ln f(x|\theta)+\ln f(\theta)\right]=0$$

$$\Rightarrow \frac{x-\theta_{MAP}}{\sigma_w^2}-\frac{\theta_{MAP}-\bar{\theta}}{\sigma_\theta^2}=0$$

$$\Rightarrow \theta_{MAP}=\frac{x\sigma_\theta^2+\bar{\theta}\sigma_w^2}{\sigma_w^2+\sigma_\theta^2}=\frac{x+\bar{\theta}\left(\frac{\sigma_w^2}{\sigma_\theta^2}\right)}{1+\frac{\sigma_w^2}{\sigma_\theta^2}}$$

$\boxed{\text{Case 1}}$: $\sigma_\theta^2 \uparrow$

$\bar{\theta}$ has low reliability, then the term $\bar{\theta}\left(\frac{\sigma_w^2}{\sigma_\theta^2}\right) \downarrow$

$\Rightarrow x$ is dominated $\Rightarrow$ $\theta_{MAP}\approx x$

$\boxed{\text{Case 2}}$: $\sigma_\theta^2 \downarrow$

$\bar{\theta}$ is dominated $\Rightarrow$ $\theta_{MAP}\approx \bar{\theta}$

(ML estimation can provide an optimal solution when no *a priori* information is given; however, MAP estimation is better than ML estimation when *a priori* information is given.) ■

Example 2: Single parameter, multiple observations

$$x(n) = \theta + w(n), \quad n = 0, 1, 2, \ldots, M-1$$

$w(n)$: iid Gaussian with zero mean and variance σ_w^2

- ML estimation:

$$\frac{\partial}{\partial\theta}\left[\ln f(x;\theta)\right] = 0$$

$$L_\theta = f(\mathbf{x};\theta) = \frac{1}{\left(2\pi\sigma_w^2\right)^{M/2}} e^{-\frac{1}{2\sigma_w^2}\sum_{i=0}^{M-1}\left[x(i)-\theta\right]^2}$$

$$\frac{\partial}{\partial\theta}\left[\ln L_\theta\right] = 0 \Rightarrow \theta_{ML} = \frac{1}{M}\sum_{i=0}^{M-1} x(i)$$

- MAP estimation

$$f(\theta) = \frac{1}{\sqrt{2\pi\sigma_\theta^2}} e^{-\frac{(\theta-\bar{\theta})^2}{2\sigma_\theta^2}}$$

$$f(\mathbf{x}|\theta) = \frac{1}{\left(2\pi\sigma_w^2\right)^{M/2}} e^{-\frac{1}{2\sigma_w^2}\sum_{i=0}^{M-1}\left[x(i)-\theta\right]^2}$$

$$\frac{\partial}{\partial\theta}\left[\ln f(\mathbf{x}|\theta) + \ln f(\theta)\right] = 0$$

$$\theta_{MAP} = \left[\frac{1}{M}\sum_{i=0}^{M-1} x(i) + \bar{\theta}\,\frac{\sigma_w^2}{M\sigma_\theta^2}\right] \Big/ \left[1 + \frac{\sigma_w^2}{M\sigma_\theta^2}\right]$$

Reduce the dependence on the priori density by a factor of *M*. ■

Note: Cramer-Rao Lower Bound (CRLB)

1. Parameter vector to be estimated:

$$\begin{cases} \boldsymbol{\theta} = \left[\theta_0, \theta_1, \ldots, \theta_{L-1}\right]^T : \text{an } L\times 1 \text{ vector} \\ \hat{\boldsymbol{\theta}}: \text{an unbiased estimator of } \boldsymbol{\theta} \end{cases}$$

$$Cov\left\{\hat{\boldsymbol{\theta}}\right\} = E\left\{\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^T\right\} \tag{13}$$

$$J(i,j) = -E_x\left\{\frac{\partial^2 \ln f(\mathbf{x};\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right\}, \; i, j = 0, 1, 2, \ldots, L-1 \tag{14}$$

$$J = \left[J(i,j)\right]_{L\times L} : \text{Fisher information matrix}$$

CRLB:

$Cov\left\{\hat{\boldsymbol{\theta}}\right\} - J^{-1}$ is positive semi-definite. (Diagonal elements are all non-negative)

$Cov\left\{\hat{\boldsymbol{\theta}}\right\} \geq \underbrace{J^{-1}}_{\text{Lower bound}}$ (Convenient notation)

When the ML estimator's variance attains a lower bound, the lower bound is the CRLB.

2. Minimum Variance Unbiased Estimator (MVUE)

An unbiased estimator which attains the CRLB is called an MVUE.

3. Variance bound of each component of $\hat{\theta}$:

$Var\{\hat{\theta}_i\} \geq J^{-1}(i,i), \quad i = 0,\ 1,\ 2, \ldots,\ L-1 \Rightarrow$ Diagonal terms

4. Special case:

$\boldsymbol{\theta} = [\theta] \rightarrow$ There is only one component in $\boldsymbol{\theta}$. (single-parameter case)

$$\begin{aligned} J &= -E\left\{\frac{\partial^2}{\partial\theta^2}\left[\ln f\left(x;\theta\right)\right]\right\} \\ &= -E\left\{\frac{\partial^2}{\partial\theta^2}\left[-\frac{1}{2}\ln\left(2\pi\sigma_w^2\right) - \frac{1}{2\sigma_w^2}\left(x-\theta\right)^2\right]\right\} \\ &= -E\left\{\frac{-2}{2\sigma_w^2}\right\} = \frac{1}{\sigma_w^2} \end{aligned} \tag{15}$$

$$x = \theta + w \tag{16}$$

$$\because \theta_{ML} = x \tag{17}$$

$$\therefore \sigma_{ML}^2 = \sigma_w^2 = Var\left\{\theta_{ML}\right\} \geq 1/J \tag{18}$$

Here $Var\left\{\theta_{ML}\right\} = 1/J$. So, the ML estimator attains the CRLB.

8) Minimum Mean-Squared Error (MMSE) Estimation (Random Parameter Estimation)

In practice, $f(\boldsymbol{\theta}|\mathbf{x})$ cannot be found easily, so we consider MMSE estimation, estimating a random parameter θ using a function of observed random data:

$$\theta_{MS} = h(x) \tag{19}$$

Cost function: $J = E\{[\theta - h(x)]^2\}$

MMSE estimation can be derived as follows:

Find $h(x)$ such that J is minimized.

$$J = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\theta - h(x)\right]^2 f(x,\theta)\,dx\,d\theta \tag{20}$$

$$\begin{aligned}\min_{h(x)} J &= \min_{h(x)} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\theta - h(x)\right]^2 f(\theta|x) f(x)\,dx\,d\theta \\ &= \min_{h(x)} \int_{-\infty}^{\infty} f(x)\left\{\int_{-\infty}^{\infty}\left[\theta - h(x)\right]^2 f(\theta|x)\,d\theta\right\}dx\end{aligned} \tag{21}$$

The outer integral is unaffected by the choice of $h(x)$.

$$\Rightarrow \min_{h(x)} \int_{-\infty}^{\infty}\left[\theta - h(x)\right]^2 f(\theta|x)\,d\theta \tag{22}$$

$$\frac{\partial}{\partial h(x)}\left\{\int_{-\infty}^{\infty}\left[\theta - h(x)\right]^2 f(\theta|x)\,d\theta\right\} = 0 \tag{23}$$

$$\Rightarrow 2\int_{-\infty}^{\infty}\left[\theta - h(x)\right] f(\theta|x)\,d\theta = 0 \tag{24}$$

$$\Rightarrow \int_{-\infty}^{\infty}\theta f(\theta|x)\,d\theta = \int_{-\infty}^{\infty} h(x) f(\theta|x)\,d\theta \tag{25}$$

$$\Rightarrow E\{\theta|x\} = h(x) \rightarrow \text{conditional mean} \tag{26}$$

i.e., the expectation of θ conditioned on the observed data.

$$\begin{cases}\text{MAP: Find the value of } \theta \text{ that maximizes } f\{\theta|x\} \\ \text{MMSE: Find the conditional mean } E\{\theta|x\}\end{cases}$$

$\Rightarrow$ MMSE estimation is a kind of Bayesian estimation

(a) For a symmetric pdf case, $\theta_{MAP} = \theta_{MS}$.

For a non-symmetric pdf case, $\theta_{MAP} \neq \theta_{MS}$.

(b) Other forms of cost function

$J = E\{|\theta - h(x)|\} \rightarrow$ minimum mean absolute error (MMAE).

$J = E\{[\theta - h(x)]^4\} \rightarrow$ minimum mean fourth error.

θ_{MMAE} : the median of $f(\theta|x)$.

(c) Extension to the multiple-parameter case

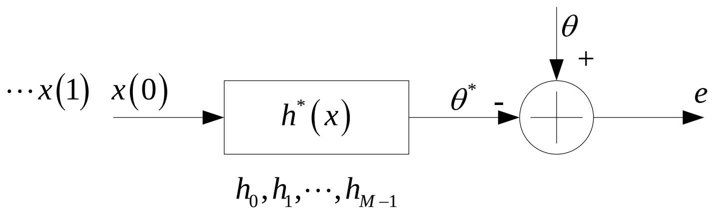
$$\begin{aligned} &\boldsymbol{\theta}_{MS} = h(\mathbf{x}) = E\{\boldsymbol{\theta}|\mathbf{x}\} \\ &\boldsymbol{\theta} = [\theta_0 \quad \theta_1 \quad \cdots \quad \theta_{L-1}]^T,\ \mathbf{x} = [x(0) \quad x(1) \quad \cdots \quad x(M-1)]^T \end{aligned} \tag{27}$$

(d) Linear MMSE estimation

$\theta_{MS} = h(x) = E\{\theta|x\}$ is in general a nonlinear function of **x**.

Simplified version:

Let $h(x)$ be a linear function of x.



■ **Figure 3.2** Graphic representation of linear MMSE.

$$\theta^* = \sum_{j=0}^{M-1} h_j x(j) = h(x) \tag{28}$$

$$J = E\{(\theta - \theta^*)^2\} = E\left\{\left[\theta - \sum_{j=0}^{M-1} h_j x(j)\right]^2\right\} = E\{e^2\} \tag{29}$$

$$\min_{\theta = h^*(x)} J \Rightarrow \min_{\substack{h_j \\ j=0,1,2,\ldots,M-1}} J \Rightarrow \frac{\partial J}{\partial h_j} = 0, \quad j = 0,\ 1,\ 2, \ldots,\ M-1 \tag{30}$$

$$J = E\{e^2\},\ e = \theta - \sum_{j=0}^{M-1} h_j x(j) \tag{31}$$

$$\frac{\partial J}{\partial h_j} = E\left\{2e\frac{\partial e}{\partial h_j}\right\} = 0 \tag{32}$$

$$\Rightarrow E\{ex(j)\} = 0 \tag{33}$$

The error is orthogonal to the data (measurements).

$\rightarrow$ "Orthogonality Principle"

(e) Estimation of signals

$$\hat{d}(n) = T\underbrace{\{x(n)\}}_{observations} \tag{34}$$

(i) MAP estimation:

maximize $f\left(d(n)\middle|x(n)\right)$ or

maximize $\ln f\left(x(n)\middle|d(n)\right) + \ln f\left(d(n)\right)$

(ii) ML estimation:

maximize $f\left(x(n);d(n)\right)$ or maximize $\ln f\left(x(n)\middle|d(n)\right)$

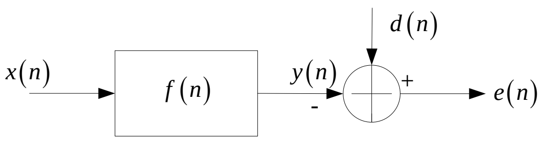
(iii) MMSE estimation:

$$\hat{d}(n) = E\left\{d(n)\middle|x(n)\right\} \tag{35}$$

(iv) Linear MMSE estimation:

$$\hat{d}(n) = \sum_{i=0}^{M-1} h_i x(i) \Rightarrow \hat{d}(n) = \sum_{i=0}^{M-1} f(i)x(n-i), \text{ in filter form} \tag{36}$$

II. Optimal Least-Squares Filter Design



■ **Figure 3.3** Least-squares optimal filtering [1, p.90].

1) Random Case: $J = E\left\{e^2(n)\right\}$

Deterministic Case: $J = \sum_n e^2(n)$

$$\begin{aligned} e(n) &= d(n) - y(n) \\ &= d(n) - \sum_i f(i)x(n-i) \end{aligned} \tag{37}$$

$$J = E\left\{e^2(n)\right\} \tag{38}$$

$$\frac{\partial J}{\partial f(j)} = 2E\left\{e(n)\frac{\partial e(n)}{\partial f(j)}\right\} = -2E\left\{e(n)x(n-j)\right\} = 0, \quad \forall j \tag{39}$$

$$E\left\{\left[d(n)-\sum_i f(i)x(n-i)\right]x(n-j)\right\}=0 \tag{40}$$

$$\Rightarrow E\left\{\sum_i f(i)x(n-i)x(n-j)\right\}=E\{x(n-j)d(n)\} \tag{41}$$

$$\Rightarrow \sum_i f(i)E\{x(n-i)x(n-j)\}=E\{x(n-j)d(n)\} \tag{42}$$

$$\Rightarrow \sum_i R_x(n-i,n-j)\, f(i)=g(n-j,n), \quad \forall j \tag{43}$$

where

$$R_x(n-i,n-j)=E\{x(n-i)x(n-j)\} \tag{44}$$

and

$$g(n-j,n)=E\{x(n-j)d(n)\} \tag{45}$$

This set of equations is called the "normal equations."

2) Linear MMSE Solution

The solution of the normal equations

≡ Least-squares solution

≡ Least-squares filter

(a) Orthogonality condition

The error is orthogonal to the input data.

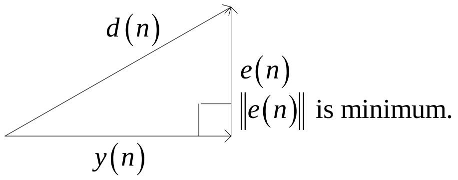
$$E\{e(n)x(n-j)\}=0 \tag{46}$$

(b) Orthogonality condition ⇔ least-squares solution.

(c) Orthogonality of the error and the output data at time n.

$$\because E\left\{e(n)\sum_i f(i)x(n-i)\right\}=\sum_i f(i)E\{e(n)x(n-i)\}=0 \tag{47}$$

$$\therefore E\{e(n)y(n)\}=0 \tag{48}$$



■ **Figure 3.4** Orthogonality condition.

3) Forms of Normal Equations

(a) Stationary formulation

$$\sum_i r(j-i) f(i) = g(j) \tag{49}$$

(b) Nonstationary formulation

$$\sum_i R_x(n-i, n-j) f(i) = g(n-j, n) \tag{50}$$

4) Solutions of the Normal Equations

(a) Infinite two-sided filter (noncausal IIR filter)

$$\sum_{i=-\infty}^{\infty} r(j-i) f(i) = g(j), \quad -\infty < j < \infty \tag{51}$$

$$r(j) * f(j) = g(j), \quad -\infty < j < \infty \tag{52}$$

Taking Z-transform on both sides,

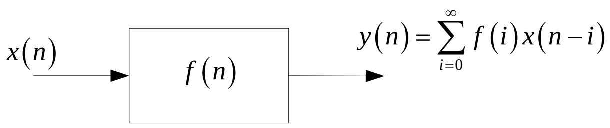
$$R(z)F(z) = G(z) \Rightarrow F^*(z) = \frac{G(z)}{R(z)}, \quad (\cdot)^* : \text{ optimal solution} \tag{53}$$

$$F^*\left(e^{j\omega}\right) = \frac{G\left(e^{j\omega}\right)}{R\left(e^{j\omega}\right)} \tag{54}$$

$$\begin{cases} r(j) = E\{x(n)x(n+j)\} = r_{xx}(j) \\ R(z) = R_{xx}(z) \\ g(j) = E\{x(n-j)d(n)\} = r_{xd}(j) \\ G(z) = R_{xd}(z) \end{cases} \tag{55}$$

(b) Infinite causal IIR filter

$$\sum_{i=0}^{\infty} r(j-i) f(i) = g(j), \quad 0 \le j < \infty \tag{56}$$



■ **Figure 3.5** Infinite causal IIR filter.

(i) The input $x(n)$ is white noise

$$r(j) = r_{xx}(j) = \sigma_x^2 \delta(j) \tag{57}$$

$$f(j) * r(j) = g(j), \quad 0 \le j < \infty \tag{58}$$

$$\therefore f(j) * \delta(j) = g(j) / \sigma_x^2, \quad 0 \le j < \infty \tag{59}$$

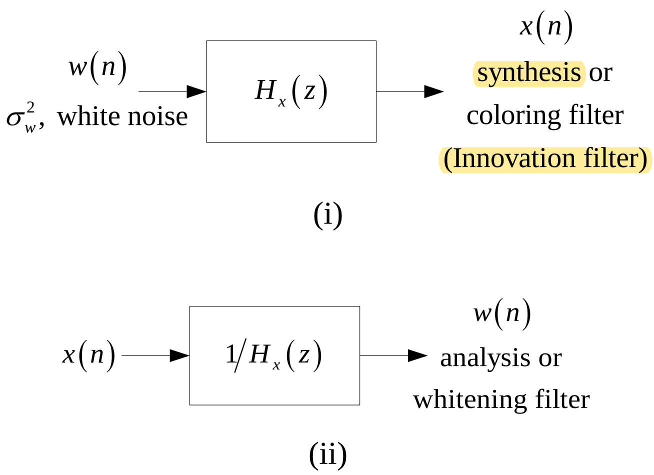
$$f(j) = \begin{cases} g(j)/\sigma_x^2, & 0 \le j < \infty \\ 0 & , \quad j < 0 \end{cases} \tag{60}$$

$$\Rightarrow F(z) = \frac{1}{\sigma_x^2}\left[G(z)\right]_+ = \frac{1}{\sigma_x^2}\left[R_{xd}(z)\right]_+ \tag{61}$$

where

$$\left[G(z)\right]_+ \triangleq \sum_{j=0}^{\infty} g(j) z^{-j} \quad (\text{one-sided ZT}) \tag{62}$$

(ii) The input $x(n)$ is a regular process (linear)

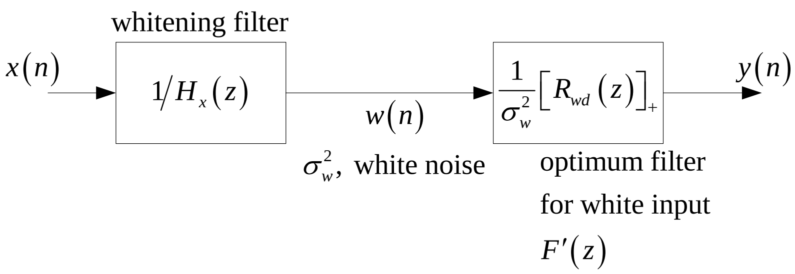


■ **Figure 3.6** (i) Synthesis (innovation) filter; (ii) Analysis (whitening) filter [2, p.152].

Assume $H_x(z)$ is both causal and stable (i.e., minimum phase),

$$H_x(z) \equiv \text{innovation filter}$$

$$= \sum_{k=0}^{\infty} h_x(k) z^{-k} \quad \text{with } h_x(0) = 1 \quad (\text{normalization})$$



■ **Figure 3.7** Signal pre-whitening.

$$F'(z)=\frac{1}{\sigma_w^2}\left[R_{wd}(z)\right]_+=\frac{1}{\sigma_w^2}\left[\frac{R_{xd}(z)}{H_x\left(z^{-1}\right)}\right]_+ \tag{63}$$

Note:

$$R_{xx}(z)H(z)H\left(z^{-1}\right)=R_{yy}(z) \tag{64}$$

$$R_{xx}(z)H(z)=R_{xy}(z) \tag{65}$$

$$R_{xx}(z)H\left(z^{-1}\right)=R_{yx}(z) \tag{66}$$

$$R_{wd}(z)=\frac{R_{xd}(z)}{H_x\left(z^{-1}\right)} \tag{67}$$

$$F(z)=\frac{1}{H_x(z)}F'(z)=\frac{1}{\sigma_w^2 H_x(z)}\left[\frac{R_{xd}(z)}{H_x\left(z^{-1}\right)}\right]_+ \tag{68}$$

(iii) finite causal filter (FIR)

$$\sum_{i=0}^{L-1} r(j-i)\,f(i)=g(j),\quad 0\le j\le L-1 \tag{69}$$

$$\begin{aligned} j=0,\quad & r(0)f(0)+r(-1)f(1)+r(-2)f(2)+\cdots \\ & +r(-L+1)f(L-1)=g(0) \end{aligned} \tag{70}$$

$$\begin{aligned} j=1,\quad & r(1)f(0)+r(0)f(1)+r(-1)f(2)+\cdots \\ & +r(-L+2)f(L-1)=g(1) \\ & \vdots \end{aligned} \tag{71}$$

The matrix form is

$$\begin{bmatrix} r(0) & r(-1) & \cdots & r(-L+1) \\ r(1) & r(0) & \cdots & r(-L+2) \\ \vdots & \vdots & \ddots & \vdots \\ r(-L+1) & r(-L+2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} f(0) \\ f(1) \\ \vdots \\ f(L-1) \end{bmatrix} = \begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ g(L-1) \end{bmatrix} \tag{72}$$

$$\Rightarrow \mathbf{Rf} = \mathbf{g}$$

Vector notation:

$$\mathbf{f} = \begin{bmatrix} f(0) & f(1) & \cdots & f(L-1) \end{bmatrix}^T \tag{73}$$

$$\mathbf{x}_n = \begin{bmatrix} x(n) & x(n-1) & \cdots & x(n-L+1) \end{bmatrix}^T \tag{74}$$

$$y(n) = \sum_{i=0}^{L-1} f(i)x(n-i) = \mathbf{f}^T\mathbf{x}_n = \mathbf{x}_n{}^T\mathbf{f} \tag{75}$$

$$e(n) = d(n) - y(n) = d(n) - \mathbf{f}^T\mathbf{x}_n \tag{76}$$

$$\begin{aligned} J &= E\left\{e^2(n)\right\} = E\left\{\left[d(n) - \mathbf{f}^T\mathbf{x}_n\right]\left[d(n) - \mathbf{x}_n{}^T\mathbf{f}\right]\right\} \\ &= E\left\{d^2(n)\right\} + \mathbf{f}^T E\left\{\mathbf{x}_n\mathbf{x}_n{}^T\right\}\mathbf{f} - 2\mathbf{f}^T E\left\{\mathbf{x}_n d(n)\right\} \\ &= \sigma_d^2 + \mathbf{f}^T\mathbf{Rf} - 2\mathbf{f}^T\mathbf{g} \end{aligned} \tag{77}$$

where $\mathbf{R} = E\left\{\mathbf{x}_n\mathbf{x}_n{}^T\right\}$ and $\mathbf{g} = E\left\{\mathbf{x}_n d(n)\right\}$.

$$\nabla J = \begin{bmatrix} \dfrac{\partial J}{\partial f(0)} & \dfrac{\partial J}{\partial f(1)} & \cdots & \dfrac{\partial J}{\partial f(L-1)} \end{bmatrix}^T = 2\mathbf{Rf} - 2\mathbf{g} = 0 \tag{78}$$

$$\Rightarrow \mathbf{Rf} = \mathbf{g} \Rightarrow \mathbf{f} = \mathbf{R}^{-1}\mathbf{g} \tag{79}$$

5) Computation of the Least-Squares (LS) Solution

(a) Computation of auto- and cross-correlation coefficients

$$r'(i) = \frac{1}{M}\sum_{n=0}^{M-i-1} x(n)x(n+i), \quad i = 0,\ 1,\ 2, \ldots,\ L-1 \tag{80}$$

$$g'(i) = \frac{1}{M}\sum_{n=0}^{M-i-1} x(n)d(n+i), \quad i = 0,\ 1,\ 2, \ldots,\ L-1 \tag{81}$$

(b) Solving the normal equations

$$\mathbf{R}'\mathbf{f}' = \mathbf{g}' \tag{82}$$

6) Error Energy in Finite Length Least-Squares Filtering

(a) The mean-squared error energy is

$$J = \sigma_d^2 + \mathbf{f}^T\mathbf{R}\mathbf{f} - 2\mathbf{f}^T\mathbf{g} \tag{83}$$

Let $\mathbf{f}^*$ be the least-squares solution (i.e., the linear MMSE solution)

$$J_{\min} = \sigma_d^2 - \mathbf{f}^{*T}\mathbf{g} \quad \text{or} \quad J_{\min} = \sigma_d^2 - \mathbf{f}^{*T}\mathbf{R}\mathbf{f}^* \tag{84}$$

$$\mathbf{v} = \mathbf{f} - \mathbf{f}^* \tag{85}$$

$$\begin{aligned} \mathbf{v}^T\mathbf{R}\mathbf{v} &= \left(\mathbf{f} - \mathbf{f}^*\right)^T \mathbf{R}\left(\mathbf{f} - \mathbf{f}^*\right) \\ &= \mathbf{f}^T\mathbf{R}\mathbf{f} + \mathbf{f}^{*T}\mathbf{R}\mathbf{f}^* - 2\mathbf{f}^{*T}\mathbf{R}\mathbf{f} \end{aligned} \tag{86}$$

where

$$\begin{aligned} 2\mathbf{f}^{*T}\mathbf{R}\mathbf{f} &= 2\left(\mathbf{R}^{-1}\mathbf{g}\right)^T \mathbf{R}\mathbf{f} \\ &= 2\mathbf{g}^T\mathbf{R}^{-1}\mathbf{R}\mathbf{f} \quad (\text{symmetric}) \\ &= 2\mathbf{g}^T\mathbf{f} = 2\mathbf{f}^T\mathbf{g} \end{aligned} \tag{87}$$

$$\Rightarrow \mathbf{v}^T\mathbf{R}\mathbf{v} = \mathbf{f}^T\mathbf{R}\mathbf{f} + \mathbf{f}^{*T}\mathbf{g} - 2\mathbf{f}^T\mathbf{g} \tag{88}$$

$J - J_{\min} = \mathbf{f}^T\mathbf{R}\mathbf{f} + \mathbf{f}^{*T}\mathbf{g} - 2\mathbf{f}^T\mathbf{g} = \mathbf{v}^T\mathbf{R}\mathbf{v} \geq 0$ ($\mathbf{R}$ is positive semi-definite.)

choose $\begin{cases} \mathbf{f}^* \to J_{\min} \\ \mathbf{f} \to J \end{cases}$.

For any $\mathbf{f}$ we choose, we can get the cost function $J \geq J_{\min}$; so $J_{\min}$ is the minimum value. The solution of the normal equations is called the linear MMSE solution.

(b) Monotonicity of $J_{\min}$ with L

The error energy $J_{\min}^{(L)}$ is a monotonically non-increasing function of L.

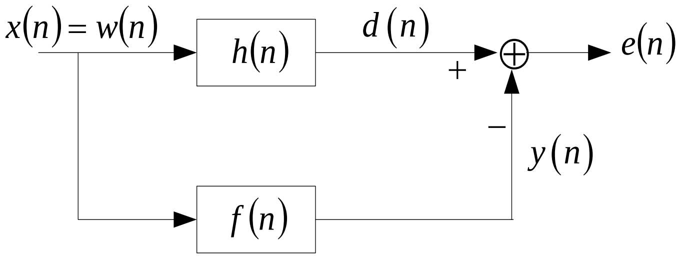
$$J_{\min}^{(L+1)} \leq J_{\min}^{(L)} \tag{89}$$

$$\begin{pmatrix} J_{\min}^{(5)} \to \begin{bmatrix} f_0^* & f_1^* & f_2^* & f_3^* & f_4^* \end{bmatrix} \\ J^{(6)} \to \begin{bmatrix} f_0^* & f_1^* & f_2^* & f_3^* & f_4^* & 0 \end{bmatrix} \\ J_{\min}^{(5)} = J^{(6)} \\ J_{\min}^{(6)} \leq J^{(6)} = J_{\min}^{(5)} \end{pmatrix} \tag{90}$$

Example 3:

$$d(n) = x(n) + 0.5x(n-1) + 0.7x(n-2),$$

given $h(0)=1$, $h(1)=0.5$, and $h(2)=0.7$.



■ **Figure 3.8** Block diagram for a system identification example [1, p.101].

Least-squares solution:

$$\mathbf{Rf}^* = \mathbf{g}$$

$$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(L-1) \\ r(1) & r(0) & r(1) & \cdots & r(L-2) \\ r(2) & r(1) & r(0) & \cdots & r(L-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(L-1) & r(L-2) & r(L-3) & \cdots & r(0) \end{bmatrix}_{L\times L}, \quad \mathbf{g} = \begin{bmatrix} g(0) \\ g(1) \\ \vdots \\ g(L-1) \end{bmatrix}_{L\times 1}$$

$$\because r(j) = E\{x(n)x(n+j)\} = E\{w(n)w(n+j)\} = \sigma_w^2 \delta(j)$$

$$\therefore \mathbf{R} = \begin{bmatrix} \sigma_w^2 & 0 & \cdots & 0 \\ 0 & \sigma_w^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \sigma_w^2 \end{bmatrix}_{L\times L}$$

$$\because g(j) = E\{x(n-j)d(n)\} = E\{x(n-j)[x(n)+0.5x(n-1)+0.7x(n-2)]\}$$

$$\therefore \begin{cases} j=0, & g(0)=\sigma_w^2 \\ j=1, & g(1)=0.5\sigma_w^2 \\ j=2, & g(2)=0.7\sigma_w^2 \\ \text{other } j, & g(j)=0 \end{cases} \Rightarrow \mathbf{g} = \sigma_w^2 \begin{bmatrix} 1 \\ 0.5 \\ 0.7 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{L\times 1}$$

$$\Rightarrow \mathbf{f}^* = \mathbf{R}^{-1}\mathbf{g} = \frac{1}{\sigma_w^2}\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0.5 \\ 0.7 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \sigma_w^2 = \begin{bmatrix} 1 \\ 0.5 \\ 0.7 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\begin{aligned} J_{\min}^{(1)} &= \sigma_d^2 - \mathbf{f}^{*T}\mathbf{g} = \left[1 + (0.5)^2 + (0.7)^2\right]\sigma_w^2 - 1 \cdot \sigma_w^2 = \left[(0.5)^2 + (0.7)^2\right]\sigma_w^2 \\ &= 0.74\sigma_w^2 \\ J_{\min}^{(2)} &= \sigma_d^2 - \mathbf{f}^{*T}\mathbf{g} = \left[1 + (0.5)^2 + (0.7)^2\right]\sigma_w^2 - \left[1 + (0.5)^2\right]\sigma_w^2 = 0.49\sigma_w^2 \\ J_{\min}^{(3)} &= \sigma_d^2 - \mathbf{f}^{*T}\mathbf{g} = \left[1 + (0.5)^2 + (0.7)^2\right]\sigma_w^2 - \left[1 + (0.5)^2 + (0.7)^2\right]\sigma_w^2 = 0 \end{aligned}$$

■

(c) Normalized mean-squared error (MSE)

$$\sigma_d^2 \geq J_{\min} = \sigma_d^2 - \mathbf{f}^{*T}\mathbf{g} = \sigma_d^2 - \mathbf{f}^{*T}\mathbf{R}\mathbf{f}^* \geq 0 \tag{91}$$

$$0 \leq J'_{\min} \triangleq \frac{J_{\min}}{\sigma_d^2} = 1 - \frac{1}{\sigma_d^2}\mathbf{f}^{*T}\mathbf{R}\mathbf{f}^* \leq 1 \tag{92}$$

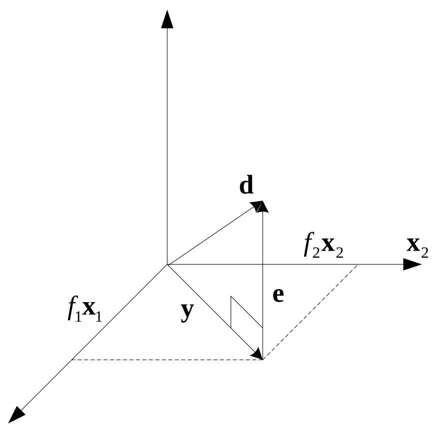
$$0 \leq J'_{\min} \leq 1 \tag{93}$$

(d) Performance of the filter

$$P \triangleq 1 - J'_{\min} \Rightarrow 0 \leq P \leq 1 \tag{94}$$

$$\begin{cases} P = 1, & \mathbf{d} \text{ is in the Span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots\} \\ P = 0, & \mathbf{d} \text{ is orthogonal to Span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots\} \end{cases}$$

(If $\mathbf{R}$ were singular, $\mathbf{f}^{*T}\mathbf{R}\mathbf{f}^*$ might be negative, which might make $J'_{\min} \triangleq \frac{J_{\min}}{\sigma_d^2}$ greater than 1. Hence, P would be negative.)

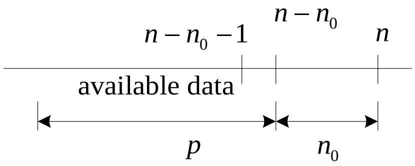


■ **Figure 3.9** Orthogonality principle [2, p.273].

III. Applications of Least-Squares Filters

1) Linear Prediction

$$\hat{x}(n) = \sum_{i=1}^{p} c_i x(n - n_0 - i + 1) \tag{95}$$



■ **Figure 3.10** The operation of linear prediction.

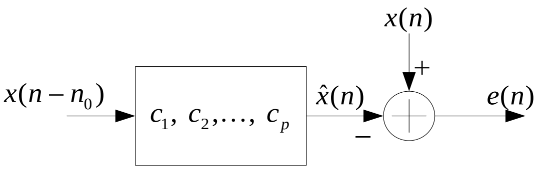
(a) A predictor of order p and prediction distance n_0.

single-step predictor: $n_0 = 1$,

$$\hat{x}(n) = \sum_{i=1}^{p} c_i x(n-i)\,. \tag{96}$$

(b) Prediction error:

$$\begin{aligned} e(n) &= x(n) - \hat{x}(n) \\ &= x(n) - \sum_{i=1}^{p} c_i x(n - n_0 - i + 1) \end{aligned}\,. \tag{97}$$



■ **Figure 3.11** Least-squares optimal prediction for prediction distance n_0 [1, p.106].

Cost function – $J = E\left\{e^2(n)\right\}$ $\Rightarrow$ minimize $J \Rightarrow$ find a set of c_i.

$$J = E\left\{\left[x(n) - \sum_{i=1}^{p} c_i x(n - n_0 - i + 1)\right]^2\right\} \tag{98}$$

$$\begin{aligned} &\frac{\partial}{\partial c_i} J = 0,\ \ i = 1,\ 2, \ldots,\ p \\ &\Rightarrow \mathbf{Rc} = \mathbf{g} \Rightarrow \mathbf{c} = \mathbf{R}^{-1}\mathbf{g} \end{aligned} \tag{99}$$

where

$$\mathbf{c} = \begin{bmatrix} c_1 & c_2 & \cdots & c_p \end{bmatrix}^T \tag{100}$$

$$\mathbf{g} = \begin{bmatrix} g(0) & g(1) & \cdots & g(p-1) \end{bmatrix}^T \tag{101}$$

We know

$$g(j) = E\{x(n-j)d(n)\} = E\{x(n-n_0-j)x(n)\}$$
$$= r(n_0+j),\ \ j = 0,\ 1,\ 2, \cdots,\ p-1 \tag{102}$$

$\Rightarrow$ change the old notations (in the least-squares optimal filter):

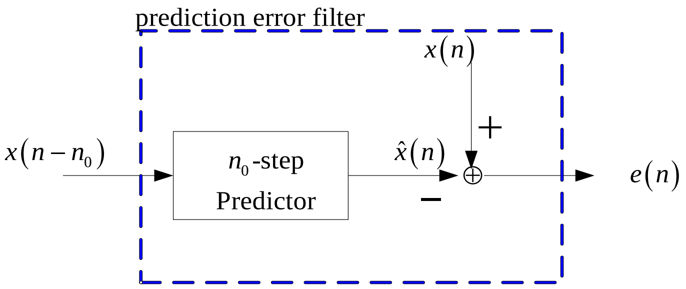
$x(n-n_0)$ for $x(n)$, $\hat{x}(n)$ for $y(n)$, and $x(n)$ for $d(n)$.

$$\therefore \mathbf{g} = \begin{bmatrix} r(n_0) & r(n_0+1) & \cdots & r(n_0+p-1) \end{bmatrix}^T \tag{103}$$

$$r(j) = E\{x(n)x(n+j)\} \tag{104}$$

$$\mathbf{R} = \begin{bmatrix} r(0) & r(1) & r(2) & \cdots & r(p-1) \\ r(1) & r(0) & r(1) & \cdots & r(p-2) \\ r(2) & r(1) & r(0) & \cdots & r(p-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r(p-1) & r(p-2) & r(p-3) & \cdots & r(0) \end{bmatrix} \tag{105}$$

2) Prediction Error Filter (PEF)



■ **Figure 3.12** Prediction error filter.

$$e(n) = x(n) - \hat{x}(n)$$
$$= x(n) - \sum_{i=1}^{p} c_i x(n-n_0-i+1) \tag{106}$$
$$= \sum_{i=-n_0+1}^{p} f_i x(n-n_0-i+1)$$

where

$$\mathbf{c} = \begin{bmatrix} c_1 & c_2 & \cdots & c_p \end{bmatrix}^T \tag{107}$$

$$\mathbf{f} = \left[\underset{i=-n_0+1}{\underset{\downarrow}{1}}, \underbrace{0,\cdots,0}_{n_0-1}, \underset{i=1}{\underset{\downarrow}{-c_1}}, -c_2, \cdots, \underset{i=p}{\underset{\downarrow}{-c_p}} \right]^T$$
$$= \left[\underset{1}{\underset{\|}{f_{-n_0+1}}}, \underset{0}{\underset{\|}{f_{-n_0+2}}}, \cdots, \underset{0}{\underset{\|}{f_{-1}}}, \underset{0}{\underset{\|}{f_0}}, \underset{-c_1}{\underset{\|}{f_1}}, \underset{-c_2}{\underset{\|}{f_2}}, \cdots, \underset{-c_p}{\underset{\|}{f_p}} \right]^T \tag{108}$$

Special case: $n_0 = 1$

$$\mathbf{f} = \left[1, \quad -c_1, \quad -c_2, \quad \cdots, \quad -c_p\right]^T \tag{109}$$

$$e(n) = \mathbf{f}^T \mathbf{x}_n \tag{110}$$

We can set up an augmented set of normal equations whose solutions will yield the PEF directly.

$$\mathbf{c} = \mathbf{R}^{-1}\mathbf{g} \tag{111}$$

$$\begin{bmatrix} r(0) & r(1) & \cdots & \cdots & r(p-1) \\ r(1) & r(0) & & & r(p-2) \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ r(p-1) & \cdots & \cdots & \cdots & r(0) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ \vdots \\ c_p \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ \vdots \\ r(p) \end{bmatrix} \tag{112}$$

$$\Rightarrow \begin{cases} r(0)c_1 + r(1)c_2 + \cdots + r(p-1)c_p = r(1) \\ r(1)c_1 + r(0)c_2 + \cdots + r(p-2)c_p = r(2) \\ \vdots \\ r(p-1)c_1 + r(p-2)c_2 + \cdots + r(0)c_p = r(p) \end{cases} \tag{113}$$

$$r(0) - r(1)c_1 - r(2)c_2 - \cdots - r(p)c_p = ?$$
$$\Rightarrow \begin{cases} r(1) - r(0)c_1 - r(1)c_2 - \cdots - r(p-1)c_p = 0 \\ r(2) - r(1)c_1 - r(0)c_2 - \cdots - r(p-2)c_p = 0 \\ \vdots \\ r(p) - r(p-1)c_1 - r(p-2)c_2 - \cdots - r(0)c_p = 0 \end{cases} \tag{114}$$

$$\Rightarrow \mathbf{R}_{(p+1)\times(p+1)}\mathbf{f} = \begin{bmatrix} ? \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{115}$$

MMSE solution:

$$\begin{aligned} J_{\min} &= \sigma_d^2 - \mathbf{f}^{*T}\mathbf{g} \\ &= E\left\{d^2(n)\right\} - \sum_{i=0}^{L-1} f(i)g(i) \end{aligned} \tag{116}$$

$$\begin{aligned} J &= E\left\{e^2(n)\right\} \\ &= E\left\{\left[x(n) - \sum_{i=0}^{p} c_i x(n-i)\right]^2\right\} \\ &= E\left\{e(n)\left[x(n) - \sum_{i=1}^{p} c_i x(n-i)\right]\right\} \end{aligned} \tag{117}$$

where $x(n-i)$ is the input signal.

Using the orthogonality principle, we have

$$J_{\min} = E\left\{e(n)x(n)\right\} \tag{118}$$

The least-squares solution occurs when $e(n)$ is orthogonal to $x(n-i)$.

$$\begin{aligned} J_{\min} &= E\left\{\left[x(n) - \sum_{i=1}^{p} c_i x(n-i)\right]x(n)\right\} \\ &= r(0) - \sum_{i=1}^{p} c_i r(i) \\ &= r(0) - r(1)c_1 - r(2)c_2 - \cdots - r(p)c_p \end{aligned} \tag{119}$$

Then, the augmented equation

$$\mathbf{R}_{(p+1)\times(p+1)}\mathbf{f} = \begin{bmatrix} J_{\min} \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{120}$$

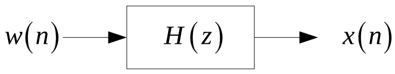
This augmented set of normal equations can be solved directly using the Levinson recursion.

3) Linear Prediction and the AR Model

AR(M):

$$H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_M z^{-M}} \tag{121}$$

$$x(n) = a_1 x(n-1) + a_2 x(n-2) + \cdots + a_M x(n-M) + w(n) \tag{122}$$



■ **Figure 3.13** System model.

Single-step predictor:

$$\hat{x}(n) = c_1 x(n-1) + c_2 x(n-2) + \cdots + c_p x(n-p) \tag{123}$$

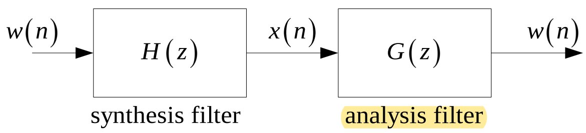
Assume $p = M$

$$\begin{aligned} J &= E\left\{e^2(n)\right\} = E\left\{\left[x(n) - \hat{x}(n)\right]^2\right\} \\ &= E\left\{\left[a_1 x(n-1) + a_2 x(n-2) + \cdots + a_p x(n-p) + w(n)\right.\right. \\ &\quad \left.\left. - c_1 x(n-1) - c_2 x(n-2) - \cdots - c_p x(n-p)\right]^2\right\} \\ &= E\left\{\left[(a_1 - c_1) x(n-1) + (a_2 - c_2) x(n-2) + \cdots + (a_p - c_p) x(n-p)\right]^2\right\} \\ &\quad + E\left\{w^2(n)\right\} \end{aligned} \tag{124}$$

$$J_{\min} = E\left\{w^2(n)\right\} = \sigma_w^2 \text{ when } c_1 = a_1,\ c_2 = a_2\ ,\cdots,\text{ and } c_p = a_p. \tag{125}$$

Note:

1. The single-step prediction error filter is identical to the analysis filter for the AR process.



■ **Figure 3.14** Synthesis and analysis filters.

$$H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_M z^{-M}} \tag{126}$$

$$\therefore G(z) = \frac{1}{H(z)} = 1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_M z^{-M} \tag{127}$$

Single-step predictor:

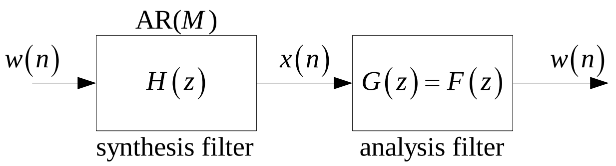
$$\begin{aligned} \mathbf{c} &= \begin{bmatrix} c_1 & c_2 & \cdots & c_p \end{bmatrix}^T \\ &= \begin{bmatrix} a_1 & a_2 & \cdots & a_p \end{bmatrix}^T = \begin{bmatrix} a_1 & a_2 & \cdots & a_M \end{bmatrix}^T, \text{ when } p = M \end{aligned} \tag{128}$$

Single-step prediction error filter

$$\mathbf{f} = \begin{bmatrix} 1 & -c_1 & -c_2 \cdots & -c_M \end{bmatrix}^T = \begin{bmatrix} 1 & -a_1 & -a_2 & \cdots & -a_M \end{bmatrix}^T \tag{129}$$

$$\therefore F(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_M z^{-M} = G(z) \tag{130}$$

2. The prediction error filter $F(z)$ is always minimum phase irrespective of the phase structure of $x(n)$.



■ **Figure 3.15** Synthesis and analysis filter for AR process.

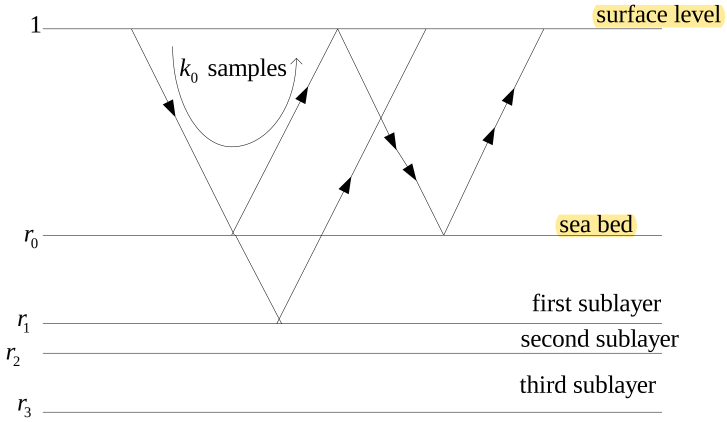
$$F(z) = G(z) = \frac{1}{H(z)} \tag{131}$$

Poles of $H(z)$ are all inside the unit circle.

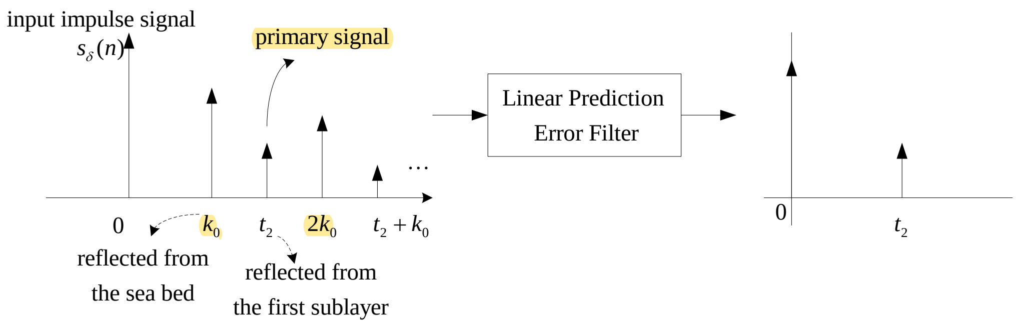
$\Rightarrow$ Zeros of $F(z)$ are all inside the unit circle.

Example 4: Application of Linear Prediction to "Geophysical Exploration."

Prediction error filters have been used for the purpose of removing, or at least attenuating, predictable components from seismic signals. The problem is particularly prevalent in marine seismic data.



■ **Figure 3.16** Reflection in an idealized marine seismic system [1, p.112].



■ **Figure 3.17** Geophysical exploration by using the linear prediction error filter.

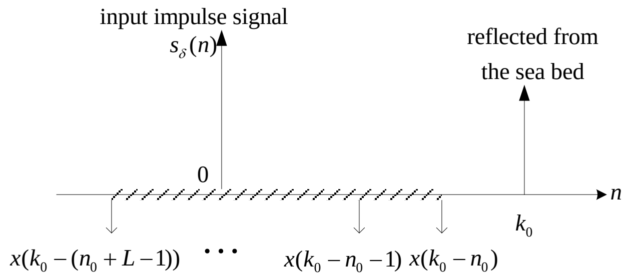
Consider the prediction of the first reflection signal from the input impulse signal

(1) $k_0 - n_0 \geq 0$

n_0 must be smaller than k_0 such that the first reflection signal can be predicted by using the input impulse signal, $s_\delta(n)$.

(2) $k_0 - n_0 - L + 1 \leq 0$

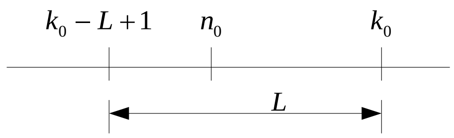
The input impulse signal, $s_\delta(n)$, must be one of the input data for use in the prediction of the first reflection signal. Thus, $n_0 + L - 1 \geq k_0$.



■ **Figure 3.18** The range of the input data for the linear prediction error filter.

From the conditions (1) and (2), the predictor will be effective for any choice of prediction distance in the range

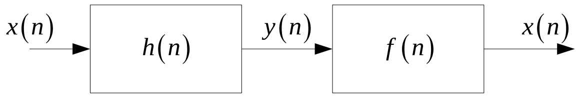
$$k_0 - L + 1 \le n_0 \le k_0 .$$



■ **Figure 3.19** The range of prediction distance.

■

4) Deconvolution (Inverse filtering)



■ **Figure 3.20** Deconvolution using an inverse filter $f(n)$ [1, p.116].

$$\left.\begin{aligned} x(n)*h(n)&=y(n) \\ y(n)*f(n)&=x(n) \end{aligned}\right\} \Rightarrow x(n)*h(n)*f(n)=x(n) \tag{132}$$

$$h(n)*f(n)=\delta(n) \tag{133}$$

$$\Rightarrow H(z)F(z)=1 \Rightarrow F(z)=\frac{1}{H(z)} \tag{134}$$

Example 5: $H(z)$: FIR, causal and stable.

$$H(z)=1-a_1z^{-1}, \qquad |a_1|>1$$

$$F(z) = \frac{1}{H(z)} = \frac{1}{1 - a_1 z^{-1}}$$

$$F(z) = 1 + a_1 z^{-1} + a_1^2 z^{-2} + a_1^3 z^{-3} + \cdots \ \rightarrow \text{causal but not stable.}$$

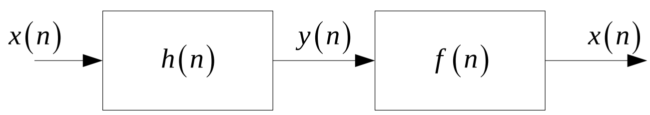
or

$$F(z) = -\frac{1}{a_1} z - \frac{1}{a_1^2} z^2 - \frac{1}{a_1^3} z^3 - \cdots \rightarrow \text{stable but noncausal.}$$ ■

How to find an appropriate solution that is both causal and stable?

(a) Optimal least-squares inverse filter design

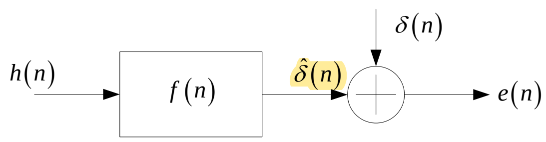
(i)



causal & stable

$h(n) = 0$ for $n < 0$

■ **Figure 3.21** Inverse filtering.



■ **Figure 3.22** Inverse filter design by LS approximation.

Cost function –

$$J = \sum_{n=-\infty}^{\infty} e^2(n) = \sum_{n=-\infty}^{\infty} \left[\delta(n) - \hat{\delta}(n)\right]^2 = \sum_{n=-\infty}^{\infty} \left[\delta(n) - h(n) * f(n)\right]^2 \tag{135}$$

$\Rightarrow$ LS solution

$$\sum_i r(j-i) f(i) = g(j) \tag{136}$$

$$r(j-i) = \sum_{n=-\infty}^{\infty} h(n-i) h(n-j) = \sum_{n=-\infty}^{\infty} h(n) h(n+i-j) \tag{137}$$

$$r(m) = \sum_{n=-\infty}^{\infty} h(n) h(n-m) = h(m) * h(-m) \tag{138}$$

$$g(j) = \sum_{n=-\infty}^{\infty} \delta(n) h(n-j) = h(-j) \tag{139}$$

(ii) solution of $\sum_i r(j-i)f(i) = g(j)$

- infinite unconstrained filter

$$\left.\begin{aligned} &r(j) * f(j) = g(j) \\ &R(z)F(z) = G(z) = H(z^{-1}) \\ &R(z) = H(z)H(z^{-1}) \end{aligned}\right\} \Rightarrow F(z) = \frac{1}{H(z)} \tag{140}$$

- infinite causal filter

$$\sum_{i=0}^{\infty} r(j-i)f(i) = \begin{cases} h(0) & , \quad j=0 \\ 0 & , \quad j=1,2,\cdots \end{cases} \tag{141}$$

If $H(z)$ is minimum phase, then

$$F(z) = \frac{1}{H(z)} \tag{142}$$

"analysis filter"

If $H(z)$ is not minimum phase, then

$$H(z) = H_{eqmin}(z)H_{ap}(z) \tag{143}$$

and

$$F(z) = \frac{c}{H_{eqmin}(z)} \tag{144}$$

where c is a constant.

$$\Rightarrow H(z)F(z) = cH_{ap}(z). \tag{145}$$

The amplitude of $x(n)$ can be recovered but the phase cannot.

Example 6:

$$H(z) = 1 + az^{-1}, \qquad |a| > 1$$

$$H(z) = H_{eqmin}(z) \cdot H_{ap}(z) = (a + z^{-1}) \cdot \frac{1 + az^{-1}}{a + z^{-1}}$$

$$F(z) = \frac{c}{H_{eq\min}(z)} = \frac{c}{a + z^{-1}}$$

- finite causal filter

$$\begin{aligned} \mathbf{Rf} &= \mathbf{g} \\ \mathbf{f} &= [f(0) \;\; f(1) \;\; \cdots \;\; f(L-1)]^T. \\ \mathbf{g} &= [h(0) \;\; 0 \;\; 0 \;\; \cdots \;\; 0]^T \end{aligned} \tag{146}$$

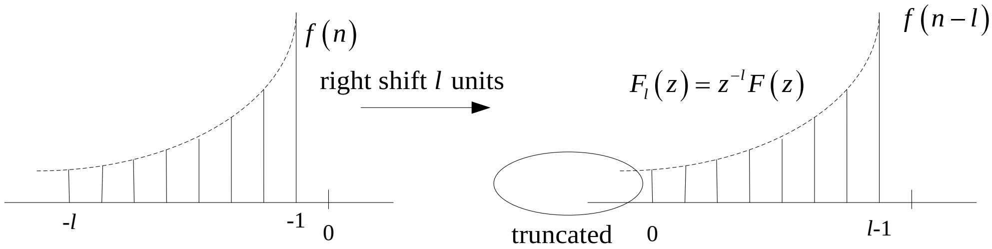
(b) The use of delay in inverse filtering

$$H(z) = 1 + az^{-1}$$
$$\Rightarrow F(z) = \frac{1}{1+az^{-1}} = \frac{z}{z+a} = \frac{z}{a}\left[\frac{1}{1+za^{-1}}\right], \ \left|za^{-1}\right| < 1 \text{ or } |z| < |a| \tag{147}$$

If $|a| > 1$,

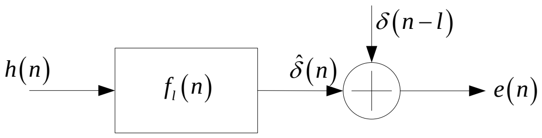
$$F(z) = \frac{1}{a}z - \frac{1}{a^2}z^2 + \frac{1}{a^3}z^3 - \cdots \tag{148}$$

The system is stable but noncausal.



■ **Figure 3.23** The filter impulse response with delay *l*.

(c) Least squares inverse filtering with delay



■ **Figure 3.24** Least-squares inverse filtering with delay *l* [1, p.120].

$$f(n) * h(n) = \delta(n) \tag{149}$$

$$f_l(n) * h(n) = \hat{\delta}(n) \rightarrow \delta(n-l) \tag{150}$$

where $f_l(n)$ refers to $f(n-l)$ with truncation for $n < 0$.

$$\sum_i r(j-i) f_l(i) = g(j) \tag{151}$$

$$g(j) = \sum d(n)x(n-j) = \sum \delta(n-l)h(n-j) = h(l-j) \tag{152}$$

$g(j)=0$ for $j>l$ when $h(n)$ is causal.

$$
\begin{aligned}
&j=0 \Rightarrow g(0)=h(l) \\
&j=1 \Rightarrow g(1)=h(l-1) \\
&j=2 \Rightarrow g(2)=h(l-2) \\
&\qquad\vdots \\
&j=l \Rightarrow g(l)=h(0) \\
&j=l+1 \Rightarrow g(l+1)=0 \\
&\qquad\vdots
\end{aligned} \tag{153}
$$

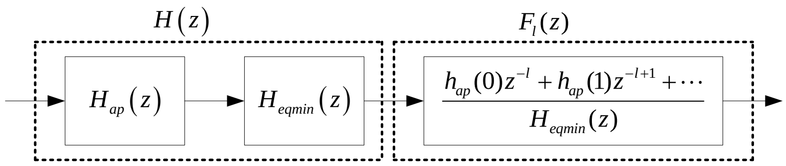
$$
\mathbf{g}=\begin{bmatrix} h(l) & h(l-1) & \cdots & h(0) & 0 & \cdots & 0 \end{bmatrix}^T \tag{154}
$$

Solution:

(i) infinite causal case

$$
F_l(z)=\frac{h_{ap}(0)z^{-l}+h_{ap}(1)z^{-l+1}+\cdots+h_{ap}(l)}{H_{eq\min}(z)} \tag{155}
$$

$$
H_{ap}(z)=h_{ap}(0)+h_{ap}(1)z+h_{ap}(2)z^2+\cdots+h_{ap}(l)z^l+\cdots \tag{156}
$$



■ **Figure 3.25** Least-squares inverse filtering with delay *l* and infinite filter length.

$l \to \infty$

$$
\begin{aligned}
F_l(z)&=\frac{h_{ap}(0)z^{-l}+\cdots+h_{ap}(l)}{H_{eqmin}(z)} \\
&=\frac{z^{-l}\cdot\left[h_{ap}(0)+h_{ap}(1)z+\cdots+h_{ap}(l)z^l\right]}{H_{eqmin}(z)}=\frac{z^{-l}\cdot H_{ap}(z)}{H_{eqmin}(z)}
\end{aligned} \tag{157}
$$

$$
\begin{aligned}
F_l(z)\cdot H(z)&=\frac{z^{-l}\cdot H_{ap}(z)}{H_{eqmin}(z)}\cdot H_{eqmin}(z)\cdot H_{ap}(z) \\
&=z^{-l}H_{ap}(z)\cdot H_{ap}(z)
\end{aligned} \tag{158}
$$

- The infinite causal least-square filter perfectly inverts $H(z)$ if the output is infinitely delayed.
- The error energy J_l is monotonically nonincreasing with l.

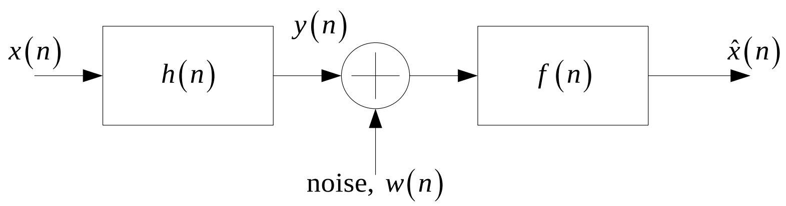
(ii) finite causal case

$$\mathbf{Rf} = \mathbf{g} \tag{159}$$

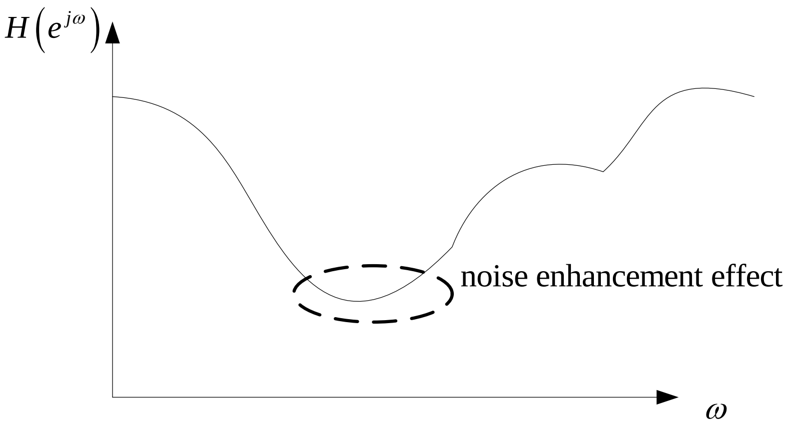
$$\begin{aligned}\mathbf{g} &= \left[g(0)\ \ g(1)\ \ g(2)\ \ \cdots\ \ g(l)\ \ 0\ \ 0\ \ \cdots\ \ 0\right]^T \\ &= \left[h(l)\ \ \ldots\ \ h(1)\ \ h(0)\ \ 0\ \ 0\ \cdots\ 0\right]^T\end{aligned} \tag{160}$$

Note: $\mathbf{R}^{-1}$ needs only to be calculated once, hence the complexity is the same.

(iii) deconvolution of noisy measured signals



■ **Figure 3.26** Deconvolution of a noisy measured signal.



■ **Figure 3.27** Frequency response of a filter.

$F(e^{jw}) = \dfrac{1}{H(e^{jw})}$: noise enhancement effect.

- reducing the noise effect (shaping)

$$F(e^{jw}) = \frac{D(e^{jw})}{H(e^{jw})} \tag{161}$$

where $D(e^{jw})$ is the Fourier transform of the desired signal.

- improving the numerical stability (prewhitening)

$$\mathbf{Rf} = \mathbf{g} \tag{162}$$

$$\mathbf{R} \to (\mathbf{R} + \sigma^2 \mathbf{I})$$
$$\Rightarrow \hat{\mathbf{R}} = \begin{bmatrix} r(0)+\sigma^2 & r(1) & r(2) & \cdots \\ r(1) & r(0)+\sigma^2 & r(1) & \vdots \\ \vdots & & \ddots & \vdots \\ \vdots & \cdots & \cdots & r(0)+\sigma^2 \end{bmatrix} \tag{163}$$

The numerical stability of $\mathbf{R}^{-1}$ is dependent on the value of $c = \lambda_{\max} / \lambda_{\min}$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum eigenvalue and the minimum eigenvalue of $\mathbf{R}$, respectively. The larger the value of *c*, the worse the condition of $\mathbf{R}$.

$$\mathbf{R} \to \{\lambda_1 \ , \lambda_2 \ , \ \ldots \ , \lambda_L\} \Rightarrow c = \frac{\lambda_{\max}}{\lambda_{\min}} \tag{164}$$

$$\mathbf{R} + \sigma^2 \mathbf{I} \to \{\lambda_1 + \sigma^2 \ , \lambda_2 + \sigma^2 \ , \ \ldots \ , \lambda_L + \sigma^2\} \tag{165}$$

$$\Rightarrow c' = \frac{\lambda_{\max} + \sigma^2}{\lambda_{\min} + \sigma^2} \le c \tag{166}$$

∗ Empirical values of σ^2: 0.5% ~ 5% of $r(0)$.

References

[1] P. M. Clarkson, *Optimal and Adaptive Signal Processing*. Boca Raton, FL: CRC, 1993.

[2] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing*. McGraw-Hill, 2000.