

MODULE 4

THE WIENER FILTER

MODULE 4: THE WIENER FILTER

- The Wiener filter addresses the additive noise MMSE signal estimation problem:

ASSUMPTION: There is a "desirable" statistical signal and an "undesirable" random noise.

- Both are modeled as stochastic processes.
- The autocorrelations, power spectral densities, crosscorrelation, and cross PSD are known.

GIVEN: The sum of the signal and noise.

FIND: The best estimate of the signal.

FIGURE OF MERIT: The best estimate of the Signal is the one that minimizes the mean (expected value) of the squared-error.

NOTE: If the signal and noise are spectrally disjoint, then they can be separated by a linear filter. Statistical techniques are not required in this case.

→ The interesting case is when the signal and noise have overlapping spectra. In this case, no linear filter can separate them perfectly.

\Rightarrow Thus, given that no linear filter can perform a perfect separation, we ask:

"What filter will do the best job in the MMSE sense?"

Symbology

$x(t)$: the desired signal.

$n(t)$: the corrupting noise.

$z(t) = x(t) + n(t)$: the given signal.

$y(t) = \hat{x}(t)$: our estimate of the signal $x(t)$.

Goal :

$$z(t) = x(t) + n(t) \rightarrow \boxed{G} \rightarrow y(t) = \hat{x}(t)$$

- Design the filter G so that $\hat{x}(t)$ is the optimal estimate in the MMSE sense.
- Since we seek to optimize the squared-error, we must first find an expression for it.

$$Z(s) = X(s) + N(s).$$

$$\hat{X}(s) = Y(s) = G(s) Z(s) = G(s) [X(s) + N(s)].$$

- The error in the estimate is given by

$$e(t) = x(t) - \hat{x}(t)$$

- So

$$\begin{aligned} E(s) &= X(s) - \hat{X}(s) \\ &= X(s) - G(s) [X(s) + N(s)] \\ &= X(s) - G(s)X(s) - G(s)N(s) \\ &= \underbrace{[1 - G(s)]X(s)}_{\text{Error due to filtering the signal}} - \underbrace{G(s)N(s)}_{\text{Error due to the noise.}} \end{aligned}$$

⇒ Until otherwise stated, we will assume that $x(t)$ and $n(t)$ are WSS and zero mean. *

→ Then $z(t)$ and $\hat{x}(t)$ are also WSS.

More Symbols:

$$S_x(s) = \mathcal{L}[R_x(\tau)] = \text{PSD of } x(t).$$

$$S_n(s) = \mathcal{L}[R_n(\tau)] = \text{PSD of } n(t).$$

etc...

* For WSS processes, the zero mean assumption is not too restrictive. Each process has a constant mean. If it is nonzero, it can be easily subtracted. PAGE 4.3

- Now for the mean squared error. It will take some work...

$$\begin{aligned}
 R_e(\tau) &= E\{e(t)e(t+\tau)\} \\
 &= E\{\left[x(t) - \hat{x}(t)\right]\left[x(t+\tau) - \hat{x}(t+\tau)\right]\} \\
 &= E\left\{x(t)x(t+\tau) - x(t)\hat{x}(t+\tau) - \hat{x}(t)x(t+\tau) + \hat{x}(t)\hat{x}(t+\tau)\right\} \\
 &= R_x(\tau) - R_{x,\hat{x}}(\tau) - R_{\hat{x},x}(\tau) + R_{\hat{x}}(\tau)
 \end{aligned}$$

$$\Rightarrow S_e(s) = S_x(s) - S_{x,\hat{x}}(s) - S_{\hat{x},x}(s) + S_{\hat{x}}(s). \quad (*)$$

$$\begin{aligned}
 R_z(\tau) &= E\{z(t)z(t+\tau)\} \\
 &= E\{\left[x(t) + n(t)\right]\left[x(t+\tau) + n(t+\tau)\right]\} \\
 &= E\left\{x(t)x(t+\tau) + x(t)n(t+\tau) + n(t)x(t+\tau) + n(t)n(t+\tau)\right\} \\
 &= R_x(\tau) + R_{x,n}(\tau) + R_{n,x}(\tau) + R_n(\tau)
 \end{aligned}$$

$$\Rightarrow S_z(s) = S_x(s) + S_{x,n}(s) + S_{n,x}(s) + S_n(s)$$

Now, using the Wiener-Kintchine relation from page 3.8

$$\begin{aligned}
 S_{\hat{x}}(s) &= S_z(s)G(-s)G(s) \\
 &= [S_x(s) + S_{x,n}(s) + S_{n,x}(s) + S_n(s)]G(-s)G(s) \quad (**)
 \end{aligned}$$

$$\begin{aligned}
R_{x,\hat{x}}(\tau) &= E \left\{ x(t) \hat{x}(t+\tau) \right\} \\
&= E \left\{ x(t) [z(t+\tau) + g(t+\tau)] \right\} \\
&= E \left\{ x(t) [(x(t+\tau) + n(t+\tau)) * g(t+\tau)] \right\} \\
&= E \left\{ x(t) [x(t+\tau) * g(t+\tau) + n(t+\tau) * g(t+\tau)] \right\} \\
&= E \left\{ x(t) \left[\int_{-\infty}^{\infty} x(t+\tau-\theta) g(\theta) d\theta + \int_{-\infty}^{\infty} n(t+\tau-\lambda) g(\lambda) d\lambda \right] \right\} \\
&= E \left\{ \int_{-\infty}^{\infty} x(t) x(t+\tau-\theta) g(\theta) d\theta + \int_{-\infty}^{\infty} x(t) n(t+\tau-\lambda) g(\lambda) d\lambda \right\} \\
&= \int_{-\infty}^{\infty} E[x(t)x(t+\tau-\theta)] g(\theta) d\theta + \int_{-\infty}^{\infty} E[x(t)n(t+\tau-\lambda)] g(\lambda) d\lambda \\
&= \int_{-\infty}^{\infty} R_x(\tau-\theta) g(\theta) d\theta + \int_{-\infty}^{\infty} R_{x,n}(\tau-\lambda) g(\lambda) d\lambda \\
&= R_x(\tau) * g(\tau) + R_{x,n}(\tau) * g(\tau)
\end{aligned}$$

$$\Rightarrow S_{x,\hat{x}}(s) = S_x(s) G(s) + S_{x,n}(s) G(s)$$

$$= G(s) [S_x(s) + S_{x,n}(s)] \quad (*)$$

$$\begin{aligned}
R_{\hat{x},x}(\tau) &= E\left\{\hat{x}(t)x(t+\tau)\right\} \\
&= E\left\{\left[z(t)+g(t)\right]x(t+\tau)\right\} \\
&= E\left\{\left[\left(x(t)+n(t)\right)+g(t)\right]x(t+\tau)\right\} \\
&= E\left\{\left[x(t)+g(t)+n(t)+g(t)\right]x(t+\tau)\right\} \\
&= E\left\{\left[\int_{-\infty}^{\infty} x(t-\theta)g(\theta)d\theta + \int_{-\infty}^{\infty} n(t-\lambda)g(\lambda)d\lambda\right]x(t+\tau)\right\} \\
&= E\left\{\int_{-\infty}^{\infty} x(t-\theta)x(t+\tau)g(\theta)d\theta + \int_{-\infty}^{\infty} n(t-\lambda)x(t+\tau)g(\lambda)d\lambda\right\} \\
&= \int_{-\infty}^{\infty} E[x(t-\theta)x(t+\tau)]g(\theta)d\theta + \int_{-\infty}^{\infty} E[n(t-\lambda)x(t+\tau)]g(\lambda)d\lambda \\
&= \int_{-\infty}^{\infty} R_x(\tau+\theta)g(\theta)d\theta + \int_{-\infty}^{\infty} R_{n,x}(\tau+\lambda)g(\lambda)d\lambda \\
&\quad \alpha = -\theta \qquad \beta = -\lambda \\
&= \int_{-\infty}^{\infty} R_x(\tau-\alpha)g(-\alpha)d\alpha + \int_{-\infty}^{\infty} R_{n,x}(\tau-\beta)g(-\beta)d\beta \\
&= R_x(\tau)*g(-\tau) + R_{n,x}(\tau)*g(-\tau) \\
\Rightarrow S_{\hat{x},x}(s) &= S_x(s)G(-s) + S_{n,x}(s)G(-s) \\
&= G(-s)[S_x(s) + S_{n,x}(s)] \quad (\star)
\end{aligned}$$

- Plug into (*) on PAGE 4.4 the following:

$$S_{\hat{x}}(s) \text{ from (***) on PAGE 4.4}$$

$$S_{x,n}(s) \text{ from (*) on PAGE 4.5}$$

$$S_{\hat{x},x}(s) \text{ from (*) on PAGE 4.6}$$

$$\Rightarrow S_e(s) = S_x(s) - G(s)[S_x(s) + S_{x,n}(s)] - G(-s)[S_x(s) + S_{n,x}(s)] \\ + G(s)G(-s)[S_x(s) + S_{x,n}(s) + S_{n,x}(s) + S_n(s)].$$

- In general, the MSE is given by

$$\begin{aligned} E[e^2(t)] &= R_e(0) \\ &= \mathcal{L}^{-1}\{S_e(s)\} \Big|_{t=0} \\ &= \frac{1}{2\pi j} \int_{-\infty}^{j\infty} S_e(s) e^{st} ds \Big|_{t=0} \\ &= \underline{\underline{\frac{1}{2\pi j} \int_{-\infty}^{j\infty} S_e(s) ds}} \quad (*) \end{aligned}$$

NOTE : Recall our assumptions that $x(t)$ and $n(t)$ are WSS and zero mean.

→ In the special case where $x(t)$ and $n(t)$ are uncorrelated (assumed often in practice), we have

$$R_{x,n}(\tau) = 0 \implies S_{x,n}(s) = 0$$

$$R_{n,x}(\tau) = 0 \implies S_{n,x}(s) = 0$$

→ In this case,

$$\begin{aligned} S_e(s) &= S_x(s) - G(s)S_x(s) - G(-s)S_x(s) + G(s)G(-s)[S_x(s) + S_n(s)] \\ &= S_x(s) - G(s)S_x(s) - G(-s)S_x(s) + G(s)G(-s)S_x(s) \\ &\quad + G(s)G(-s)S_n(s) \\ &= [1 - G(-s) - G(s) + G(s)G(-s)]S_x(s) + G(s)G(-s)S_n(s) \\ &= [1 - G(s)][1 - G(-s)]S_x(s) + G(s)G(-s)S_n(s). \end{aligned}$$



→ Plugging the last result on PAGE 4.8 into (+) on PAGE 4.7, we have

$$\begin{aligned} E[e^2(t)] &= \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} [1-G(s)][1-G(-s)] S_x(s) ds \\ &\quad + \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} G(s)G(-s) S_n(s) ds, \quad (+) \end{aligned}$$

→ which is Eq. (4.2.5) on page 162 of the book.

- Early attempts to minimize the MSE went like this:

1. Assume a parameterized form for $G(s)$.
2. Plug into (+) on PAGE 4.7 or (*) above.
3. Solve the integral for $E[e^2(t)]$ in terms of the parameter(s).
4. Use standard calculus to find the parameter value(s) to minimize $E[e^2(t)]$.

NOTE : This procedure optimizes only over the parameterized class of filters assumed.
see page 162 of the book for an example.

Wiener Filter Problem Formulation

$$z(t) = x(t) + n(t) \rightarrow \boxed{G} \rightarrow y(t) = \hat{x}(t+\alpha)$$

- Input $z(t)$ is the sum of the desired signal $x(t)$ and a corrupting noise $n(t)$.
- $x(t)$ and $n(t)$ are covariance stationary stochastic processes with known autocorrelations, crosscorrelations, spectral density, and cross power spectrum.
- G is an LSI filter.
- Output $y(t)$ is covariance stationary (any transients have died out).

Problem: Find $G(s)$ to minimize $E\{e^2(t)\}$, where

$$\begin{aligned} e(t) &= x(t+\alpha) - y(t) \\ &= x(t+\alpha) - \hat{x}(t+\alpha). \end{aligned}$$

Notes:

- When $\alpha > 0$, G tries to estimate future values of $x(t)$ based on the values of $z(t)$. This is the classical Prediction Problem.
- When $\alpha = 0$, G tries to estimate the current value of $x(t)$ from the values of $z(t)$. This is the classical Filtering Problem.

- When $\alpha < 0$, G tries to estimate past values of $x(t)$ from the values of $z(t)$. This is the classical Smoothing Problem.

- If the filter is allowed to consider all values of $z(t)$ in computing $y(t) = \hat{x}(t+\alpha)$, then it is not causal... i.e., it is a noncausal filter.
- If G is only allowed to consider past and present values of $z(t)$ in computing $y(t) = \hat{x}(t+\alpha)$, then it is a causal filter.
- In either case (causal or noncausal), the $G(s)$ that minimizes $E[e^2(t)]$ is called a Wiener Filter.
- ⇒ The stationary problem can be formulated in the time domain or in the frequency domain.
 - The frequency domain formulation is elegant. However, it does not generalize to the nonstationary case.
 - Therefore, we will formulate the problem in the time domain.

- The estimation error is given by

$$\begin{aligned} e(t) &= x(t+\alpha) - y(t) \\ &= x(t+\alpha) - \hat{x}(t+\alpha). \end{aligned}$$

- The MSE is

$$\begin{aligned} E[e^2(t)] &= E\left\{x^2(t+\alpha) - 2x(t+\alpha)y(t) + y^2(t)\right\} \\ &= E\left\{x^2(t+\alpha) - 2x(t+\alpha) \int_{-\infty}^{\infty} z(t-\theta)g(\theta)d\theta\right. \\ &\quad \left. + \int_{-\infty}^{\infty} z(t-\lambda)g(\lambda)d\lambda \int_{-\infty}^{\infty} z(t-\beta)g(\beta)d\beta\right\} \\ &= E\left\{x^2(t+\alpha) - 2 \int_{-\infty}^{\infty} z(t-\theta)x(t+\alpha)g(\theta)d\theta\right. \\ &\quad \left. + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} z(t-\lambda)z(t-\beta)g(\lambda)g(\beta)d\lambda d\beta\right\} \\ &= E\left\{x(t+\alpha)x(t+\alpha)\right\} - 2 \int_{-\infty}^{\infty} E\left\{z(t-\theta)x(t+\alpha)\right\}g(\theta)d\theta \\ &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} E\left\{z(t-\lambda)z(t-\beta)\right\}g(\lambda)g(\beta)d\lambda d\beta \\ &= R_x(0) - 2 \int_{-\infty}^{\infty} R_{z,x}(\alpha+\theta)g(\theta)d\theta + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\lambda-\beta)g(\lambda)g(\beta)d\lambda d\beta. \end{aligned}$$

(*)

- The problem is to find the $g(t)$ that minimizes $E[e^2(t)]$ in (*) on page 4.12.
- This is a problem in the calculus of variations.
- We are assuming that $g(t)$ is the impulse response of the optimal filter.
- For an arbitrary filter H with impulse response $h(t)$, we may write $h(t) = g(t) + \epsilon \eta(t)$, where
 - $\eta(t)$ is a perturbing function
 - ϵ is a perturbation scale factor.
- For the filter H , the MSE is

$$E[e^2(t)] = R_x(0) - 2 \int_{-\infty}^{\infty} R_{z,x}(\alpha+\theta) [g(\theta) + \epsilon \eta(\theta)] d\theta \\ + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\lambda-\beta) [g(\lambda) + \epsilon \eta(\lambda)] [g(\beta) + \epsilon \eta(\beta)] d\lambda d\beta \quad (*)$$

- In the limit as $\epsilon \rightarrow 0$, $h(t) \rightarrow g(t)$ and the above becomes the MSE of the optimal filter $g(t)$.
- The MSE above is a function of ϵ and $\eta(t)$.
- We can optimize with respect to ϵ by differentiating with respect to ϵ and setting the result equal to zero.

→ Since the optimum filter is obtained when $\epsilon=0$, we can obtain a constraint on $g(t)$ by evaluating the equation obtained in the previous step at $\epsilon=0$.

→ With Luck, this will give us a constraint that can be solved for $g(t)$. The solution must be valid for arbitrary choices of $\eta(t)$.

- Differentiate (*) on PAGE 4,13 with respect to ϵ :

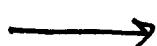
$$\begin{aligned} \frac{\partial}{\partial \epsilon} E[e^2(t)] &= \left[\right. \\ &= \frac{\partial}{\partial \epsilon} \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\lambda-\beta) [g(\lambda)g(\beta) + g(\lambda)\epsilon\eta(\beta) + \epsilon\eta(\lambda)g(\beta) + \epsilon^2\eta(\lambda)\eta(\beta)] d\lambda d\beta \right. \\ &\quad \left. - 2 \int_{-\infty}^{\infty} R_{z,x}(\alpha+\theta) [g(\theta+\epsilon\eta(\theta))] d\theta \right\} \\ &= \frac{\partial}{\partial \epsilon} \left\{ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\lambda-\beta) g(\lambda)g(\beta) d\lambda d\beta + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\lambda-\beta) g(\lambda) \epsilon\eta(\beta) d\lambda d\beta \right. \\ &\quad \left. + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\lambda-\beta) \epsilon\eta(\lambda)g(\beta) d\lambda d\beta + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\lambda-\beta) \epsilon^2\eta(\lambda)\eta(\beta) d\lambda d\beta \right. \\ &\quad \left. - 2 \int_{-\infty}^{\infty} R_{z,x}(\alpha+\theta) g(\theta) d\theta - 2 \int_{-\infty}^{\infty} R_{z,x}(\alpha+\theta) \epsilon\eta(\theta) d\theta \right\} \end{aligned}$$



$$\begin{aligned}
&= \frac{\partial}{\partial \varepsilon} \left\{ \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_2(\lambda - \beta) g(\lambda) g(\beta) d\lambda d\beta}_{\text{ZERO}} + \varepsilon \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_2(\lambda - \beta) g(\lambda) \eta(\beta) d\lambda d\beta \right. \\
&\quad + \varepsilon \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_2(\lambda - \beta) \eta(\lambda) g(\beta) d\lambda d\beta + \varepsilon^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_2(\lambda - \beta) \eta(\lambda) \eta(\beta) d\lambda d\beta \\
&\quad \left. - 2 \underbrace{\int_{-\infty}^{\infty} R_{2,x}(\alpha + \theta) g(\theta) d\theta}_{\text{ZERO}} - 2 \varepsilon \int_{-\infty}^{\infty} R_{2,x}(\alpha + \theta) \eta(\theta) d\theta \right\} \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_2(\lambda - \beta) g(\lambda) \eta(\beta) d\lambda d\beta + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_2(\lambda - \beta) \eta(\lambda) g(\beta) d\lambda d\beta \\
&\quad + 2\varepsilon \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_2(\lambda - \beta) \eta(\lambda) \eta(\beta) d\lambda d\beta - 2 \int_{-\infty}^{\infty} R_{2,x}(\alpha + \theta) \eta(\theta) d\theta \\
&= 0.
\end{aligned}$$

- For the optimal filter $g(t)$, $\varepsilon = 0$; this becomes

$$\begin{aligned}
0 &= \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_2(\lambda - \beta) g(\lambda) \eta(\beta) d\lambda d\beta}_{\text{Let } \varphi = \lambda, \varphi = \beta} + \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_2(\lambda - \beta) \eta(\lambda) g(\beta) d\lambda d\beta}_{\text{Let } \varphi = \lambda, \varphi = \beta} \\
&\quad - 2 \int_{-\infty}^{\infty} R_{2,x}(\alpha + \theta) \eta(\theta) d\theta \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_2(\varphi - \psi) g(\varphi) \eta(\psi) d\varphi d\psi + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_2(\varphi - \psi) g(\psi) \eta(\varphi) d\varphi d\psi \\
&\quad - 2 \int_{-\infty}^{\infty} R_{2,x}(\alpha + \theta) \eta(\theta) d\theta
\end{aligned}$$



- But $R_z(\tau) = R_z(-\tau)$, so this becomes

$$2 \underbrace{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\psi - \varphi) g(\psi) \eta(\varphi) d\psi d\varphi}_{\text{Let } \tau = \varphi} - 2 \underbrace{\int_{-\infty}^{\infty} R_{z,x}(\alpha + \theta) \eta(\theta) d\theta}_{\text{Let } \tau = \theta} = 0.$$

(divide both sides by 2)

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\psi - \tau) g(\psi) \eta(\tau) d\psi d\tau - \int_{-\infty}^{\infty} R_{z,x}(\alpha + \tau) \eta(\tau) d\tau = 0$$

$$\int_{-\infty}^{\infty} \eta(\tau) \left[-R_{z,x}(\alpha + \tau) + \int_{-\infty}^{\infty} R_z(\psi - \tau) g(\psi) d\psi \right] d\tau = 0. \quad (*)$$

\Rightarrow This is Eq. (4.3.7) on page 165 of the book.

\rightarrow The noncausal and causal Wiener filters are considered separately for the remainder of the solution.

Noncausal Solution

- Since (*) above must hold for arbitrary $\eta(t)$, we have

$$-R_{z,x}(\alpha + \tau) + \int_{-\infty}^{\infty} R_z(\theta - \tau) g(\theta) d\theta = 0$$

$$\int_{-\infty}^{\infty} R_z(\theta - \tau) g(\theta) d\theta = R_{z,x}(\alpha + \tau)$$



- Since $R_z(\tau) = R_z(-\tau)$, this becomes

$$\int_{-\infty}^{\infty} R_z(\tau - \theta) g(\theta) d\theta = R_{zx}(\tau + \alpha)$$

$$R_z(\tau) * g(\tau) = R_{zx}(\tau + \alpha)$$

$$S_z(s) G(s) = S_{zx}(s) e^{\alpha s}$$

$$G(s) = \frac{S_{zx}(s) e^{\alpha s}}{S_z(s)}$$

\Rightarrow This is the transfer function of the optimal
(minimum mean-squared error) noncausal filter.

\Rightarrow This filter is called the noncausal continuous-time
Wiener filter.

NOTE : This is an LSI filter. $\left\{ \begin{array}{l} \text{optimization over} \\ \text{the class of LSI} \\ \text{filters.} \end{array} \right.$

NOTE : By choosing $\alpha > 0$, $\alpha = 0$, or $\alpha < 0$, this
filter is made optimal for
prediction, filtering, and smoothing.

- What is the MSE of the optimal filter?

Eg. (*) on PAGE 4.12:

$$\begin{aligned} E[e^2(t)] &= R_x(0) - 2 \int_{-\infty}^{\infty} R_{z,x}(\alpha + \theta) g(\theta) d\theta \\ &\quad + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\lambda - \beta) g(\lambda) g(\beta) d\lambda d\beta \\ &= R_x(0) + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_z(\lambda - \beta) g(\lambda) g(\beta) d\lambda d\beta \\ &\quad - \int_{-\infty}^{\infty} R_{z,x}(\alpha + \beta) g(\beta) d\beta - \int_{-\infty}^{\infty} R_{z,x}(\alpha + \theta) g(\theta) d\theta \\ &= R_x(0) - \int_{-\infty}^{\infty} R_{z,x}(\alpha + \theta) g(\theta) d\theta + \int_{-\infty}^{\infty} g(\beta) \underbrace{\left[-R_{z,x}(\alpha + \beta) + \int_{-\infty}^{\infty} R_z(\lambda - \beta) g(\lambda) d\lambda \right]}_{\text{zero for the optimal filter, see bottom PAGE 4.16}} d\beta \end{aligned}$$

$$= R_x(0) - \int_{-\infty}^{\infty} R_{z,x}(\alpha + \theta) g(\theta) d\theta$$

→ This is the lowest MSE that can be achieved by any LSI filter, noncausal or causal.

→ Although the optimal filter will generally be noncausal, it may sometimes turn out to be causal.

Causal Solution

- We now return to Eq (*) on PAGE 4.1b and solve for the optimal causal LSI filter,
- Since the class of causal LSI filters is a subset of the class of noncausal LSI filters, the optimal causal filter can not be better than the optimal noncausal filter,
- For the optimal $g(t)$, we have again that

$$\int_{-\infty}^{\infty} \eta(\tau) \left[-R_{z,x}(\alpha+\tau) + \int_{-\infty}^{\infty} R_z(\theta-\tau) g(\theta) d\theta \right] d\tau = 0 \quad (*)$$

- In this case, $g(t)$ and $h(t) = g(t) + \epsilon \eta(t)$ are both causal.

→ Thus, $\eta(\tau) = 0 \quad \forall \tau < 0$.

→ So (*) above is satisfied for $\tau < 0$ independent of $g(t)$.

→ The constraint on $g(t)$ then becomes

$$\int_{-\infty}^{\infty} R_z(\theta-\tau) g(\theta) d\theta - R_{z,x}(\alpha+\tau) = 0, \quad \tau \geq 0. \quad (**)$$

⇒ This is the famous "Wiener-Hopf" equation.

- The classic solution to the Wiener-Hopf equation makes use of spectral factorization.
- The left side of (**) on page 4.19 must be zero for $\tau \geq 0$,
- for $\tau < 0$, however, it is unknown and may be nonzero.
- So we re-write (**) on page 4.19 as

$$\int_{-\infty}^{\infty} R_z(\theta - \tau) g(\theta) d\theta - R_{z,x}(\alpha + \tau) = a(\tau),$$

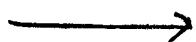
where $a(\tau) = \begin{cases} \text{unknown}, & \tau < 0 \\ 0, & \tau \geq 0. \end{cases}$

- Applying the Laplace transform, we have

$$G(s) S_z(s) - S_{z,x}(s) e^{\alpha s} = A(s)$$

$$G(s) S_z(s) = A(s) + S_{z,x}(s) e^{\alpha s}$$

- Now we spectrally factorize to obtain $S_z(s) = S_z^+(s) S_z^-(s)$, where $S_z^+(s)$ has left half-plane poles only and $S_z^-(s)$ has right half-plane poles only.



- This gives us

$$G(s)S_z^+(s)S_z^-(s) = A(s) + S_{z,x}(s)e^{xs}$$

where

$G(s)$ → left half-plane poles → transform of a causal signal

$S_z^+(s)$ → left half-plane poles → transform of a causal signal

$S_z^-(s)$ → right half-plane poles → transform of an anticausal signal

$A(s)$ → right half-plane poles → transform of an anticausal signal

$S_{z,x}(s)e^{xs}$ → may have poles in both half-planes

→ transform of a signal that is not restricted
to be causal or anticausal.

- Dividing through by $S_z^-(s)$, we obtain

$$\underbrace{G(s)S_z^+(s)}_{\text{Transform of a causal signal}} = \frac{A(s)}{\underbrace{S_z^-(s)}_{\text{Transform of an anticausal signal}}} + \frac{S_{z,x}(s)e^{xs}}{\underbrace{S_z^-(s)}_{\text{Transform of a two-sided signal}}} \quad (*)$$

Transform of a causal signal

Transform of an anticausal signal.

Transform of a two-sided signal

- For equality to hold in (+) on PAGE 4.21,

$$\mathcal{L}^{-1}\left\{\frac{A(s)}{S_z^-(s)}\right\} \text{ and } \mathcal{L}^{-1}\left\{\frac{S_{z,x}(s)e^{\alpha s}}{S_z^-(s)}\right\}$$

must cancel each other for $t < 0$.

- For $t > 0$, $\mathcal{L}^{-1}\left\{\frac{A(s)}{S_z^-(s)}\right\} = 0$.

- Thus, for $t \geq 0$,

$$\mathcal{L}^{-1}\left\{G(s)S_z^+(s)\right\} = \mathcal{L}^{-1}\left\{\frac{S_{z,x}(s)e^{\alpha s}}{S_z^-(s)}\right\}, \quad t \geq 0.$$

→ In other words,

$$\mathcal{L}^{-1}\left\{G(s)S_z^+(s)\right\} = \mathcal{L}^{-1}\left\{\frac{S_{z,x}(s)e^{\alpha s}}{S_z^-(s)}\right\} u(t)$$

- Now take the transform of both sides and solve for $G(s)$:

$$G(s)S_z^+(s) = \mathcal{L}\left\{\mathcal{L}^{-1}\left[\frac{S_{z,x}(s)e^{\alpha s}}{S_z^-(s)}\right]u(t)\right\}$$

$$\underline{\underline{G(s) = \frac{1}{S_z^+(s)} \mathcal{L}\left\{\mathcal{L}^{-1}\left[\frac{S_{z,x}(s)e^{\alpha s}}{S_z^-(s)}\right]u(t)\right\}}}$$

- In words:

1. Compute $\frac{S_{z,x}(s)e^{\alpha s}}{S_z^-(s)}$

2. Take the inverse transform of the result from step (1) and multiply by $u(t)$.

3. Take the transform of the result of step (2).

4. Divide the result of step (3) by $S_z^+(s)$ to obtain $G(s)$.

- Examples are given in the book on pages 169-172.

The Nonstationary Case

- We will not cover this in detail.

- As before, $x(t)$ and $n(t)$ are covariance stationary stochastic processes with known autocorrelations, power spectral densities, cross correlation, and cross power spectrum.

- In this case, the input $z(t) = [x(t) + n(t)] u(t)$ is applied at time $t=0$.

$$z(t) = x(t) + n(t) \xrightarrow[t=0]{X} [G] \rightarrow y(t) = \hat{x}(t+\alpha).$$

- A specialized variational approach for finding $g(\tau; t)$ is given in the book.

- The solution is a causal, time-varying linear filter.

- This leads to a treatment of optimal estimation in terms of the so-called "innovations process".

Orthogonality Principle

DEF: if X and Y are two RV's such that $E[XY] = 0$, then X and Y are called orthogonal.

- if X and Y are zero-mean, then orthogonality is equivalent to uncorrelated.
- If X and Y are orthogonal, we write $X \perp Y$.
- Suppose X is a random vector to be estimated.
Thus X can be a single RV, a vector of finite dimension, or a vector of infinite dimension.
- Let Z be an RV, random vector, or stochastic process containing observations that are related to X in some way:
$$Z = \{Z(t), t \in \mathcal{L}\}$$
- The indexing set \mathcal{L} could be finite, countably infinite, or uncountably infinite.
- A general function of the observations is denoted by $h(Z)$. This is any arbitrary function, and could be nonlinear.

- We want to estimate X from the observations Z .
- Let the optimal MMSE estimator be

$$\hat{X} = g(Z).$$

- The error in the optimal estimate is given by

$$e = X - \hat{X} = X - g(Z).$$

- The orthogonality principle states that the error in the optimal estimate is orthogonal to any function $h(Z)$ of the observations.

NOTE: if $h(Z)$ and $g(Z)$ come from a restricted class, linear functions for example, then the principle states that the error in the optimal MMSE estimator from the class is orthogonal to all functions in the class.

\Rightarrow We will prove this for unit-dimensional X and Z (scalar RVS). The proof can easily be extended to finite-dimensional vectors on a component-wise basis.

Theorem (orthogonality principle):

A function $g(z)$ of the observations is the MMSE estimator of X , i.e. $g(z) = \hat{X}$, if and only if

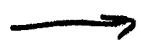
$$E\{h(z)[x-g(z)]\} = 0 \quad \forall h(z).$$

Proof:

A) Sufficiency: For a particular $g(z)$, suppose

$$E\{h(z)[x-g(z)]\} = 0 \quad \forall h(z). \text{ Then}$$

$$\begin{aligned} E\{[x-h(z)]^2\} &= E\left\{[x-g(z)+g(z)-h(z)]^2\right\} \\ &= E\left\{[x-g(z)]^2 + 2[x-g(z)][g(z)-h(z)] \right. \\ &\quad \left. + [g(z)-h(z)]^2\right\} \\ &= E\{[x-g(z)]^2\} + 2E\{[g(z)-h(z)][x-g(z)]\} \\ &\quad + E\{[g(z)-h(z)]^2\} \end{aligned}$$



Note that $[g(z) - h(z)]$ is a function of the observations.

Then by hypothesis

$$E\{[g(z) - h(z)][x - g(z)]\} = 0.$$

We have

$$\begin{aligned} E\{[x - h(z)]^2\} &= E\{[x - g(z)]^2\} + E\{[g(z) - h(z)]^2\} \\ \Rightarrow E\{[g(z) - h(z)]^2\} &\geq 0. \end{aligned}$$

Therefore, $E\{[x - h(z)]^2\} \geq E\{[x - g(z)]^2\}$,
i.e., the MSE of $g(z)$ is less than or equal
to the MSE of $h(z)$ for all $h(z)$.

\Rightarrow Then $g(z) = \hat{X}$, the optimal MMSE estimator.

B) Necessity: The proof will be by contrapositive.

That is, if we prove $\text{NOT}(B) \rightarrow \text{NOT}(A)$,
then we have also established that
 $A \rightarrow B$.

In this case, we will show that if any
 $h(z)$ is not orthogonal to $x - g(z)$, then
 $g(z)$ is not the MMSE estimator.

Suppose an $h(z)$ exists such that

$$E\{h(z)[X-g(z)]\} = \varepsilon \neq 0.$$

Assume without loss of generality that

$$E\{h^2(z)\} = 1,$$
 since if this is not true

appropriate scaling can be applied to $h(z)$ without changing the fact that $E\{h(z)[X-g(z)]\} \neq 0.$

Then

$$\begin{aligned} E\{[X-g(z) - \varepsilon h(z)]^2\} &= E\{[X-g(z)]^2\} \\ &\quad - 2\varepsilon E\{h(z)[X-g(z)]\} \\ &\quad + \varepsilon^2 E\{h^2(z)\} \\ &= E\{[X-g(z)]^2\} - 2\varepsilon^2 + \varepsilon^2 \\ &= E\{[X-g(z)]^2\} - \varepsilon^2 \\ &< E\{[X-g(z)]^2\}. \end{aligned}$$

This means that the function $g(z) + \varepsilon h(z)$ is a better estimator than $g(z).$

\Rightarrow Then $g(z)$ is not the optimal estimator $\hat{X}.$

QED.

The Discrete Case

- Suppose we have a vector of observations

$$\{z_1, z_2, \dots, z_N\}$$

from which we want to estimate a related sequence

$$\{y_1, y_2, \dots, y_N\}.$$

- We consider linear estimators for y_k of the form

$$\hat{y}_k = \sum_{i=1}^N h_{k,i} z_i. \quad (*)$$

↑ this set of N h 's is for $\underline{y_k}$.

→ Assume without loss of generality that all the y_k and z_k have zero mean.

→ By the orthogonality principle, the optimal estimator satisfies

$$E\{[y_k - \hat{y}_k] z_j\} = 0 \quad \forall j \in [1, N].$$

$$\begin{aligned} \rightarrow \text{But } E\{[y_k - \hat{y}_k] z_j\} &= E\{y_k z_j - \hat{y}_k z_j\} \\ &= E\{y_k z_j\} - E\{\hat{y}_k z_j\} = 0. \end{aligned}$$

$$\text{So } E\{y_k z_j\} = E\{\hat{y}_k z_j\} \quad \forall j \in [1, N]. \quad (**)$$

- Substituting (*) on page 4.29 into (**) on page 4.29, we have

$$\begin{aligned} E\{Y_k z_j\} &= E\left\{\sum_{i=1}^N h_{k,i} z_i z_j\right\} \\ &= \sum_{i=1}^N h_{k,i} E\{z_i z_j\}, \quad j \in [1, N] \end{aligned}$$

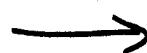
- In other words,

$$R_{Y,z}(k,j) = \sum_{i=1}^N h_{k,i} R_z(i,j), \quad j \in [1, N].$$

\Rightarrow For each Y_k , this gives a set of N equations ($j \in [1, N]$) that can be solved for the N filter coefficients $h_{k,i}$, $i \in [1, N]$, of the optimal estimator for Y_k .

\rightarrow If the processes $Y = \{Y_k\}$ and $Z = \{z_k\}$ are jointly W.S.S., this becomes

$$\begin{aligned} R_{Yz}(j-k) &= \sum_{i=1}^N h_{k,i} R_z(j-i) \\ &= \sum_{i=1}^N h_{k,i} R_z(i-j), \quad j \in [1, N] \end{aligned}$$



- Then

$$R_{ZY}(k-j) = \sum_{i=1}^N h_{k,i} R_Z(i-j), \quad j \in [1, N]$$

$$\text{let } m = k-i, \quad i = k-m$$

$$R_{ZY}(k-j) = \sum_{m=k-1}^{k-N} h_{k,k-m} R_Z(k-m-j), \quad j \in [1, N]$$

$$= \sum_{m=k-N}^{k-1} h_{k,k-m} R_Z(k-m-j)$$

$$\text{Let } l = k-j$$

$$R_{ZY}(l) = \sum_{m=k-N}^{k-1} h_{k,k-m} R_Z(l-m), \quad l \in [k-N, k-1] \quad (*)$$

→ For each k , this gives a set of N linear equations that can be solved for the optimal weights $h_{k,k-N} \dots h_{k,k-1}$.

- If the sequence Y and the observations Z are countably infinite in extent, then $(*)$ above becomes

$$R_{ZY}(l) = \sum_{m=-\infty}^{\infty} h_{k,k-m} R_Z(l-m), \quad l \in \mathbb{Z}. \quad (**)$$

⇒ The left side of $(**)$ does not depend on k .

→ Therefore, the right side cannot depend on k either.

→ i.e., the $h_{k,k-m}$ are independent of k ; they are the same for each k .



- Define $h_m = h_{k,k-m}$. Then (**) on PAGE 4.31 becomes

$$R_{zy}(l) = \sum_{m=-\infty}^{\infty} h_m R_z(l-m) = h_l * R_z(l), \quad l \in \mathbb{Z}. \quad (*)$$

- This is an infinite set of equations in the optimal filter weights h_m , thus we cannot solve for the h_m directly.
- However, we can use the Z-transform:

$$S_{zy}(z) = H(z) S_z(z)$$

- The optimal (noncausal) LSI estimator is then the filter with transfer function

$$H(z) = \frac{S_{zy}(z)}{S_z(z)}$$

(analogous to continuous-time case).

- If we restrict the optimization to causal LSI filters only, then (*) above becomes

$$R_{zy}(l) = \sum_{m=0}^{\infty} h_m R_z(l-m), \quad l \geq 0.$$

\Rightarrow This is the discrete Wiener-Hopf equation.

It can be solved by Z-domain spectral factorization techniques that are completely analogous to the continuous-time case.