Statistical Signal Processing

Lecture 6: Optimal Filtering and Applications

Carlas Smith & Peyman Mohajerin Esfahani

Delft Center for Systems and Control

Delft University of Technology

Recall A Note on Optimization

Example: Let $e(n,a)$ be affine in a, for example given as:

$$e(n,a) = d(n) + ax(n) \quad d(n), x(n), a \in \mathbb{C},$$

then the necessary condition for solving the following optimization problem,

$$\min_{a^*,(a)} |e(n,a)|^2 = \min_{a^*,(a)} e(n,a)e^*(n,a)$$

is given by,

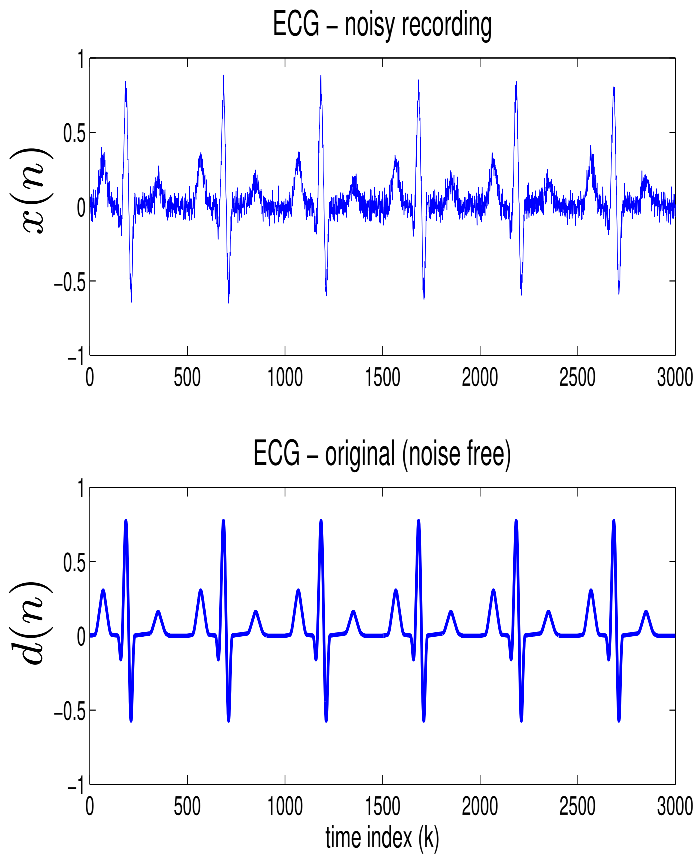
$$\boxed{e(n,a)\frac{\partial e^*(n,a)}{\partial a^*} = 0} \quad \text{or} \quad \left(\frac{\partial e(n,a)}{\partial a} e^*(n,a) = 0\right)$$

Optimal FIR filtering

1. **Recap 3 filtering problems**
2. A generic framework
3. The FIR Wiener filter
4. Three specific applications
5. The IIR Wiener filter
6. The denoising problem

Denoising of signals

ECG recordings



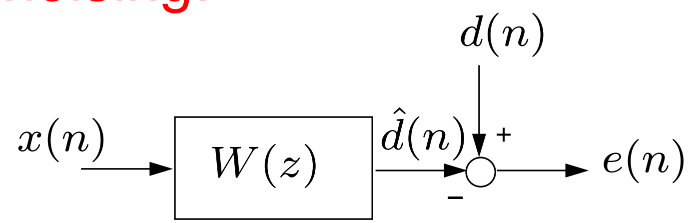
Observation model

$$x(n) = d(n) + v(n)$$

$d(n)$ — "desired" - signal of interest

$v(n)$ — "noise" - disturbance (additive)!

Denoising:



Determine $W(z)$ by minimizing $E[|e(n)|^2]$.

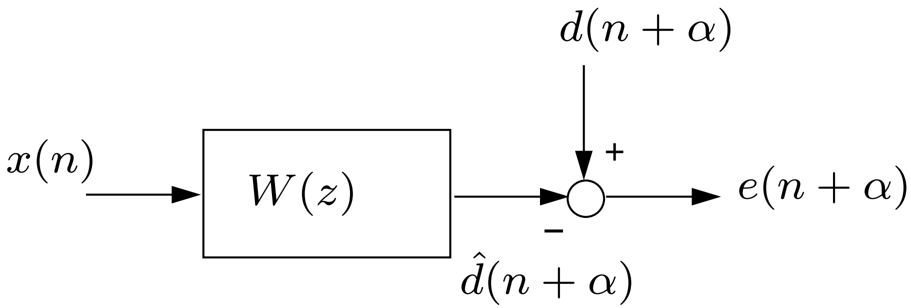
Prediction

Signal Modeling is crucial

Let $x(n) = d(n)$ with $d(n)$ modelled as (ARMA(p,q)):

$$d(n)+\sum_{\ell=1}^{p} a(\ell)d(n-\ell) = \sum_{\ell=0}^{q} b(\ell)w(n-\ell)$$

with $w(n)$ a zero-mean, white noise sequence, then the goal is to predict $d(n+\alpha)$ for $\alpha \in \mathbb{N}_0$ using $d(n), d(n-1), \cdots$.



Prediction

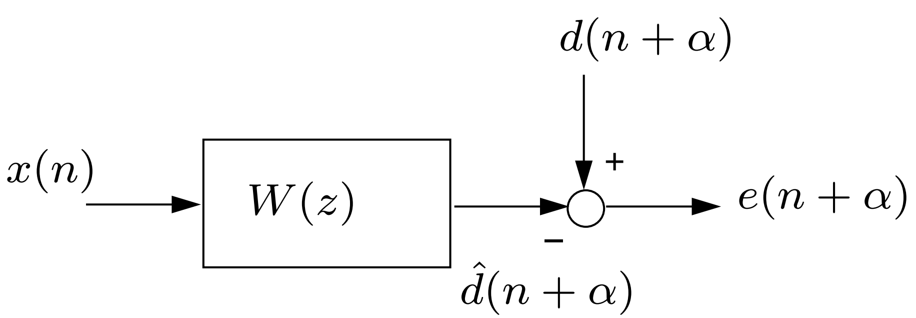
Signal Modeling is crucial

Let $x(n) = d(n)$ with $d(n)$ modelled as (ARMA(p,q)):

Extension: $x(n) = d(n) + v(n)$.

$$d(n)+\sum_{\ell=1}^{p} a(\ell)d(n-\ell) = \sum_{\ell=0}^{q} b(\ell)w(n-\ell)$$

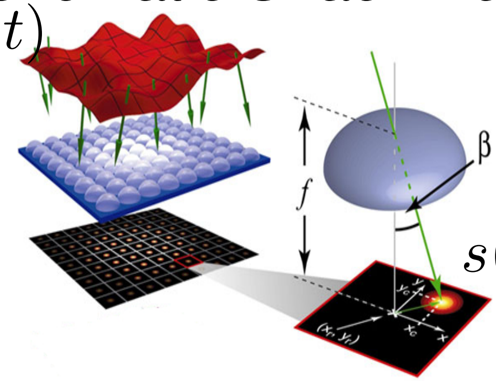
with $w(n)$ a zero-mean, white noise sequence, then the goal is to predict $d(n+\alpha)$ for $\alpha \in \mathbb{N}_0$ using $d(n), d(n-1), \cdots$.



Example: Verhaegen's CSI lab

Schematic Shack-Hartmann Sensor

$\phi(x, y, t)$

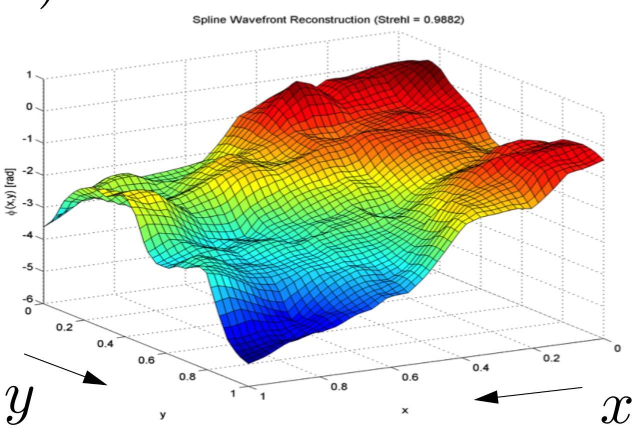


$s(x_i, y_i, t)$

[From M. Konnik, 2010]

Video Demo: Gemini Telescope (2:20)

$\phi(x, y, t)$



y $\quad$ x

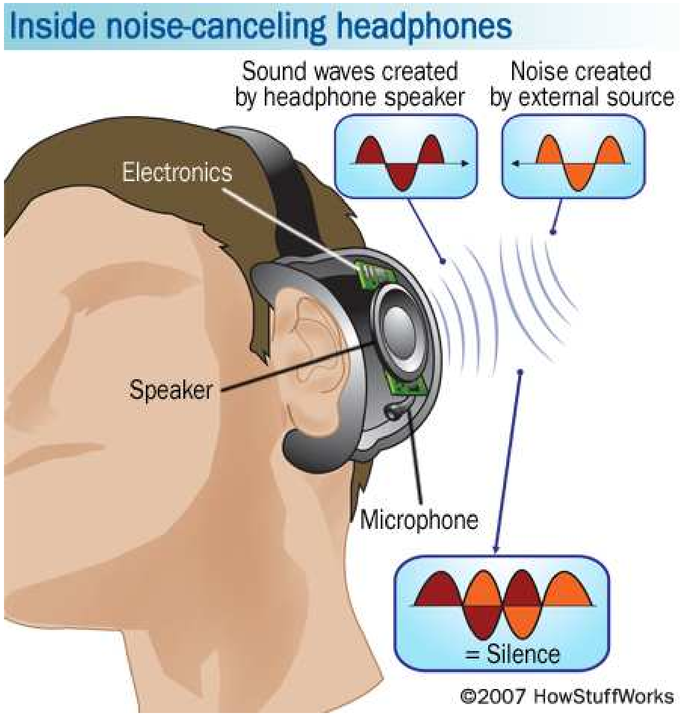
Challenge: Say n is the current time instant, can we predict $s(n+1)$ such that

$$E[\|s(n+1) - \hat{s}(n+1)|^2]$$

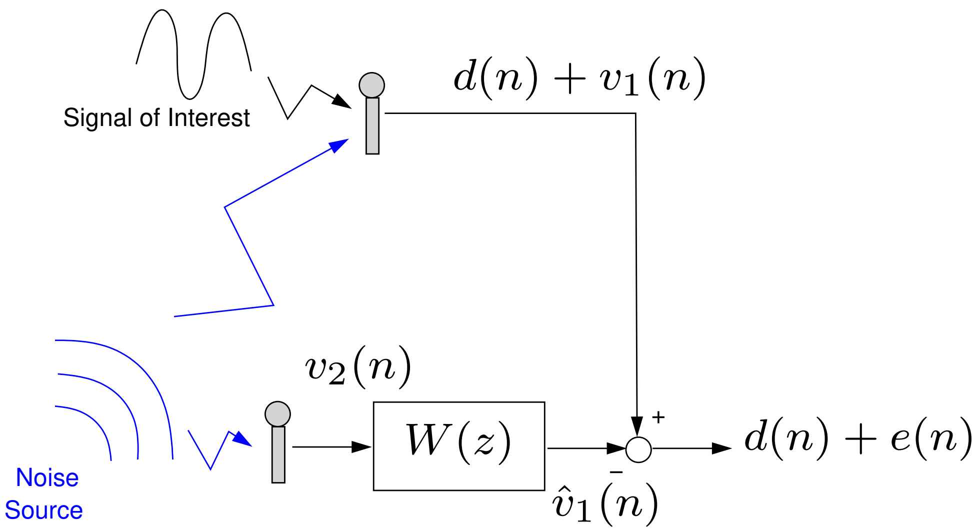
is minimized?

Active Noise Cancellation

Communicating in a “noisy” environment



Challenge: Signal modeling AND cancelling



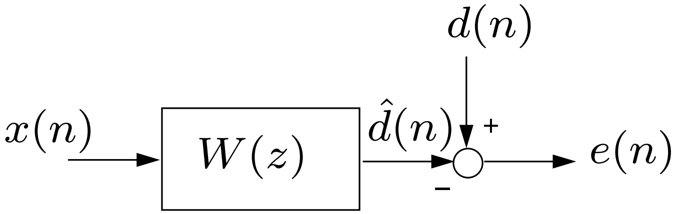
with $e(n) = v_1(n) - \hat{v}_1(n)$.

Optimal FIR filtering

1. Recap 3 filtering problems
2. **A generic framework**
3. The FIR Wiener filter
4. Three specific applications
5. The IIR Wiener filter
6. The denoising problem

A generic problem formulation

The generic problem deals with the (optimal [a]) estimation of one signal (denoted by $d(n)$) from another signal (denoted by $x(n)$.

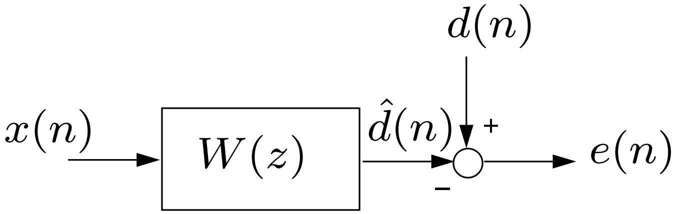


- *Filtering:* estimate $d(n)$ from $x(n) = d(n) + v(n)$.

[a] as specified by the error criterium on $e(n)$

A generic problem formulation

The generic problem deals with the (optimal[1]) estimation of one signal (denoted by $d(n)$) from another signal (denoted by $x(n)$).

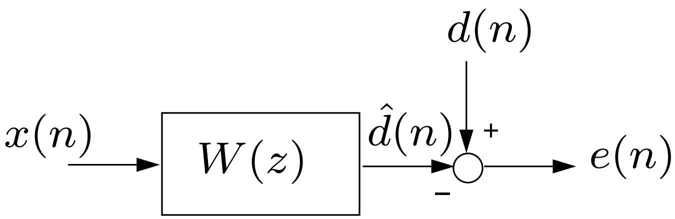


- *Filtering:* estimate $d(n)$ from $x(n) = d(n) + v(n)$.
- *Prediction:* estimate $d(n+\alpha)$ from $x(n), x(n-1), x(n-2), \cdots$ using signal model of $x(n)$ and its relations to $d(n)$.

[1] as specified by the error criterium on $e(n)$

A generic problem formulation

The generic problem deals with the (optimal) estimation of one signal (denoted by $d(n)$) from another signal (denoted by $x(n)$.



- *Filtering:* estimate $d(n)$ from $x(n) = d(n) + v(n)$.
- *Prediction:* estimate $d(n + \alpha)$ from $x(n), x(n-1), x(n-2), \cdots$ using signal model of $x(n)$ and its relations to $d(n)$.
- *Noise cancellation:* estimate $v_1(n)$ from $v_2(n)$ (and subtract it from $d(n) + v_1(n)$) using signal model $v_2(n)$ and its relation to $v_1(n)$.

---

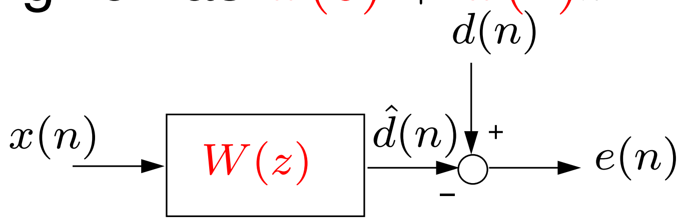
as specified by the error criterium on $e(n)$

Optimal FIR filtering

1. Recap 4 filtering problems
2. A generic framework
3. **The FIR Wiener filter**
4. Three specific applications
5. The IIR Wiener filter
6. The denoising problem

The FIR Wiener filter problem

Let the filter $W(z)$ be given as $w(0) + w(1)z^{-1} + \cdots + w(m-1)z^{m-1}$.



Then the estimate $\hat{d}(n)$ is given by:

$$\hat{d}(n) = \sum_{\ell=0}^{m-1} w(\ell)x(n-\ell) = \begin{bmatrix} w(0) & \cdots & w(m-1) \end{bmatrix} \begin{bmatrix} x(n) \\ \vdots \\ x(n-m+1) \end{bmatrix} = \mathbf{w}^T \mathbf{x}(\mathbf{n}$$

The optimality to find the coefficients $w(i)$ is expressed by minimizing the mean-square error:

$$\xi(\mathbf{w}) = E[|e(n)|^2] = E\left[\left|d(n) - \hat{d}(n)\right|^2\right] = E\left[\left|d(n) - \sum_{\ell=0}^{m-1} w(\ell)x(n-\ell)\right|^2\right]$$

The solution to the FIR Wiener filter problem

THEOREM: Let the conditions stipulated in the previous slide hold, let in addition $\{x(n), d(n)\}$ be jointly WSS and the following covariance matrices be given:

$$\mathbf{R_x} = E[\mathbf{x}^*(\mathbf{n})\mathbf{x}(\mathbf{n})^T] > 0 \quad \mathbf{r_{dx}} = E[d(n)\mathbf{x}^*(\mathbf{n})]$$

then the solution to estimate the signal $d(n)$ from $x(n)$ is derived from

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \xi\Big(\mathbf{w}\Big) \quad \text{and given as} \quad \hat{d}(n) = \hat{\mathbf{w}}^T\mathbf{x}(\mathbf{n})$$

The filter coefficients $\hat{\mathbf{w}}$ satisfy,

$$\begin{aligned} \mathbf{R_x}\hat{\mathbf{w}} &= \mathbf{r_{dx}} \quad \text{("The Wiener-Hopf equations")} \\ \xi_{\min} = \xi\Big(\hat{\mathbf{w}}\Big) &= r_d(0) - (\mathbf{r}^*_{\mathbf{dx}})^T\mathbf{R}_{\mathbf{x}}^{-1}\mathbf{r_{dx}} \end{aligned}$$

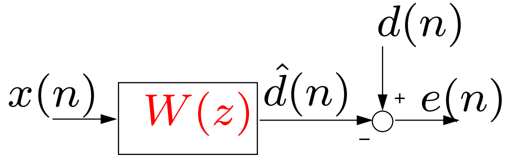
Proof of the solution to the FIR Wiener filter problem

Will done in three steps:

1. The orthogonality condition (principle)
2. The characterization of the solution with the Wiener-Hopf equations
3. The optimal residual $\xi_{\min}$.

[This will be done for case $m = 2$ i.e. $\hat{d}(n) = w(0)x(n) + w(1)x(n-1)$]

(1) The orthogonality principle

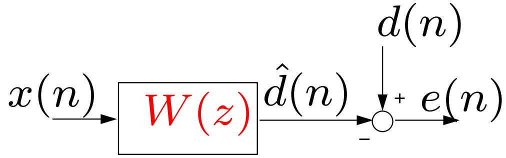


$$e(n) = d(n) - w(0)x(n) - w(1)x(n-1)$$

Then the necessary (and sufficient) condition to minimize

$\xi(\mathbf{w}) = E[|e(n)|^2] = E[e(n)e^*(n)]$:

(1) The orthogonality principle



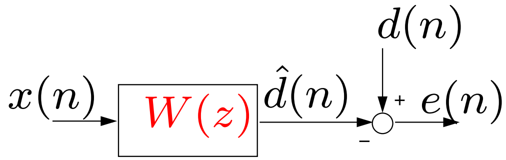
$$e(n) = d(n) - w(0)x(n) - w(1)x(n-1)$$

Then the necessary (and sufficient) condition to minimize $\xi(\mathbf{w}) = E[|e(n)|^2] = E[e(n)e^*(n)]$:

$$\begin{bmatrix} \dfrac{\partial \xi(\mathbf{w})}{\partial w^*(0)} \\ \dfrac{\partial \xi(\mathbf{w})}{\partial w^*(1)} \end{bmatrix}_{\mathbf{w}=\hat{\mathbf{w}}} = 0$$

Using the expression for $\xi(\mathbf{w})$ this equals:

(1) The orthogonality principle



$$e(n) = d(n) - w(0)x(n) - w(1)x(n-1)$$

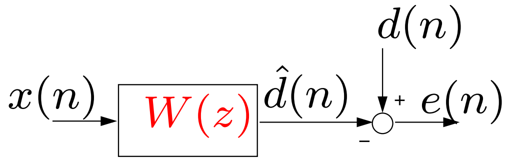
Then the necessary (and sufficient) condition to minimize $\xi(\mathbf{w}) = E[|e(n)|^2] = E[e(n)e^*(n)]$:

$$\left[\begin{array}{c} \frac{\partial \xi(\mathbf{w})}{\partial w^*(0)} \\ \frac{\partial \xi(\mathbf{w})}{\partial w^*(1)} \end{array}\right]_{\mathbf{w}=\hat{\mathbf{w}}} = 0$$

Using the expression for $\xi(\mathbf{w})$ this equals:

$$\boxed{E\left[\begin{array}{c} e_{\min}(n)\frac{\partial e^*(n)}{\partial w^*(0)} \\ e_{\min}(n)\frac{\partial e^*(n)}{\partial w^*(1)} \end{array}\right]_{\mathbf{w}=\hat{\mathbf{w}}} = E\left[\begin{array}{c} e_{\min}(n)x^*(n) \\ e_{\min}(n)x^*(n-1) \end{array}\right] = 0} \qquad \text{(O.C.)}$$

(2) The Wiener-Hopf Equations



$$e(n) = d(n) - w(0)x(n) - w(1)x(n-1)$$

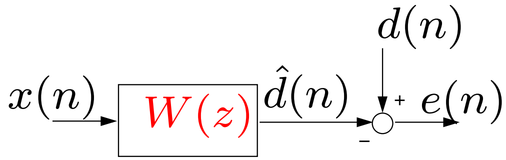
The orthogonality condition (O.C.) yields:

$$E[\begin{bmatrix} x^*(n) \\ x^*(n-1) \end{bmatrix} e_{\min}(n)] = 0 \Rightarrow$$

$$E[\begin{bmatrix} x^*(n) \\ x^*(n-1) \end{bmatrix} \Big(d(n) - \hat{w}(0)x(n) - \hat{w}(1)x(n-1)\Big)] = 0 \Rightarrow$$

$$\boxed{E[\begin{bmatrix} x^*(n) \\ x^*(n-1) \end{bmatrix} d(n)] - E[\begin{bmatrix} x^*(n) \\ x^*(n-1) \end{bmatrix} \Big[\hat{w}(0)x(n) + \hat{w}(1)x(n-1)\Big]] = 0}$$

(2) The Wiener-Hopf Equations



$$e(n) = d(n) - w(0)x(n) - w(1)x(n-1)$$

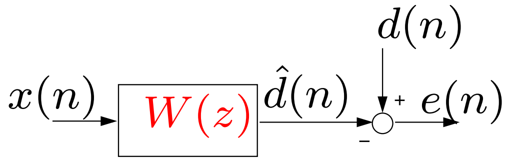
The orthogonality condition (O.C.) yields:

$$E\left[\begin{bmatrix} x^*(n) \\ x^*(n-1) \end{bmatrix} e_{\min}(n)\right] = 0 \Rightarrow$$

$$E\left[\begin{bmatrix} x^*(n) \\ x^*(n-1) \end{bmatrix} \Big(d(n) - \hat{w}(0)x(n) - \hat{w}(1)x(n-1)\Big)\right] = 0 \Rightarrow$$

$$\boxed{\underbrace{E\left[\begin{bmatrix} x^*(n) \\ x^*(n-1) \end{bmatrix} d(n)\right]}_{\mathbf{r_{dx}}} - \underbrace{E\left[\begin{bmatrix} x^*(n) \\ x^*(n-1) \end{bmatrix} \begin{bmatrix} x(n) & x(n-1) \end{bmatrix}\right]}_{\mathbf{R_x}} \underbrace{\begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix}}_{\hat{\mathbf{w}}} = 0}$$

(3) The optimal residual $\xi_{\min}$

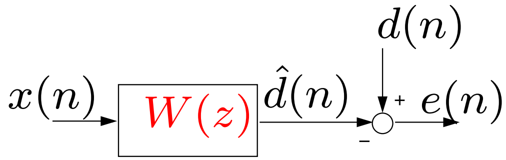


$$e(n) = d(n) - w(0)x(n) - w(1)x(n-1)$$

Using the orthogonality principle and the Wiener-Hopf equations:

$$\begin{aligned}
\xi_{\min} &= E[e_{\min}(n)e^*_{\min}(n)] \quad \text{for} \;\; e_{\min}(n) = d(n) - \begin{bmatrix} x(n) & x(n-1) \end{bmatrix} \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} \\
&= E[e_{\min}(n)d^*(n)] - E[e_{\min}(n) \begin{bmatrix} x^*(n) & x^*(n-1) \end{bmatrix}] \begin{bmatrix} \hat{w}^*(0) \\ \hat{w}^*(1) \end{bmatrix} \\
&= \\
&= \\
&=
\end{aligned}$$

(3) The optimal residual $\xi_{\min}$

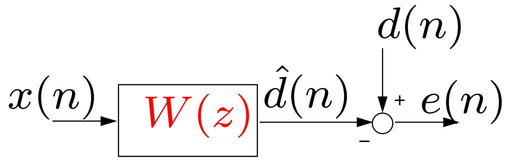


$$e(n) = d(n) - w(0)x(n) - w(1)x(n-1)$$

Using the orthogonality principle and the Wiener-Hopf equations:

$$\begin{aligned}
\xi_{\min} &= E[e_{\min}(n)e^*_{\min}(n)] \quad \text{for} \;\; e_{\min}(n) = d(n) - \begin{bmatrix} x(n) & x(n-1) \end{bmatrix} \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} \\
&= E[e_{\min}(n)d^*(n)] - E[e_{\min}(n) \begin{bmatrix} x^*(n) & x^*(n-1) \end{bmatrix}] \begin{bmatrix} \hat{w}^*(0) \\ \hat{w}^*(1) \end{bmatrix} \\
&= E[\Big(d(n) - \begin{bmatrix} x(n) & x(n-1) \end{bmatrix} \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} \Big) d^*(n)] \\
&= \\
&=
\end{aligned}$$

(3) The optimal residual $\xi_{\min}$

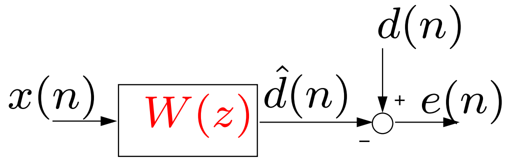


$$e(n) = d(n) - w(0)x(n) - w(1)x(n-1)$$

Using the orthogonality principle and the Wiener-Hopf equations:

$$\begin{aligned}
\xi_{\min} &= E[e_{\min}(n)e^*_{\min}(n)] \quad \text{for} \;\; e_{\min}(n) = d(n) - \begin{bmatrix} x(n) & x(n-1) \end{bmatrix} \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} \\
&= E[e_{\min}(n)d^*(n)] - E[e_{\min}(n) \begin{bmatrix} x^*(n) & x^*(n-1) \end{bmatrix}] \begin{bmatrix} \hat{w}^*(0) \\ \hat{w}^*(1) \end{bmatrix} \\
&= E[\Big(d(n) - \begin{bmatrix} x(n) & x(n-1) \end{bmatrix} \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} \Big) d^*(n)] \\
&= E[d(n)d^*(n)] - E\Big[d^*(n) \begin{bmatrix} x(n) & x(n-1) \end{bmatrix} \Big] \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} \\
&=
\end{aligned}$$

(3) The optimal residual $\xi_{\min}$



$$e(n) = d(n) - w(0)x(n) - w(1)x(n-1)$$

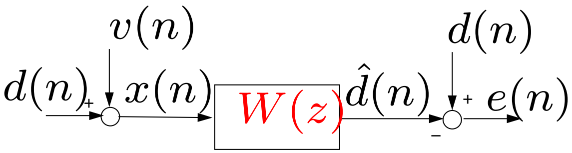
Using the orthogonality principle and the Wiener-Hopf equations:

$$\begin{aligned}
\xi_{\min} &= E[e_{\min}(n)e^*_{\min}(n)] \quad \text{for} \;\; e_{\min}(n) = d(n) - \begin{bmatrix} x(n) & x(n-1) \end{bmatrix} \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} \\
&= E[e_{\min}(n)d^*(n)] - E[e_{\min}(n) \begin{bmatrix} x^*(n) & x^*(n-1) \end{bmatrix}] \begin{bmatrix} \hat{w}^*(0) \\ \hat{w}^*(1) \end{bmatrix} \\
&= E[\Big(d(n) - \begin{bmatrix} x(n) & x(n-1) \end{bmatrix} \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} \Big) d^*(n)] \\
&= E[d(n)d^*(n)] - E\Big[d^*(n) \begin{bmatrix} x(n) & x(n-1) \end{bmatrix} \Big] \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} \\
&= r_d(0) - (\mathbf{r}^*_{\mathbf{dx}})^T \mathbf{R}_{\mathbf{x}}^{-1} \mathbf{r}_{\mathbf{dx}}
\end{aligned}$$

Optimal FIR filtering

1. Recap 4 filtering problems
2. A generic framework
3. The FIR Wiener filter
4. Three specific applications
   - **Denoising**
   - Multi-step prediction
   - Active noise cancellation
5. The IIR Wiener filter
6. The denoising Problem

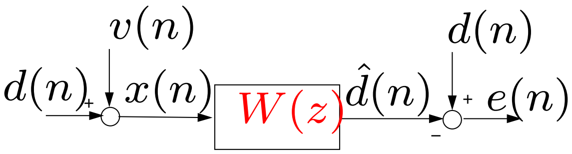
Example 1: Denoising



$$x(n) = d(n) + v(n) \quad E[d(n)v(\ell)] = 0 \ \ \forall n, \ell$$
$$\hat{d}(n) = \hat{w}(0)x(n) + \hat{w}(1)x(n-1)$$

The Wiener-Hopf equations read for this case (all signals real and WSS)

Example 1: Denoising



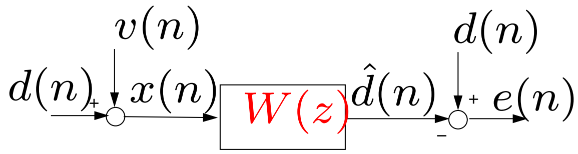
$$x(n) = d(n) + v(n) \quad E[d(n)v(\ell)] = 0 \;\; \forall n, \ell$$
$$\hat{d}(n) = \hat{w}(0)x(n) + \hat{w}(1)x(n-1)$$

The Wiener-Hopf equations read for this case (all signals real and WSS)

$$\begin{bmatrix} E[x(n)^2] & E[x(n)x(n-1)] \\ E[x(n-1)x(n)] & E[x(n-1)^2] \end{bmatrix} \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} = \begin{bmatrix} E[d(n)x(n)] \\ E[d(n)x(n-1)] \end{bmatrix}$$

To determine $r_x(0), r_x(1), r_{dx}(0), r_{dx}(1)$

Example 1: Denoising



$$x(n) = d(n) + v(n) \quad E[d(n)v(\ell)] = 0 \;\; \forall n, \ell$$
$$\hat{d}(n) = \hat{w}(0)x(n) + \hat{w}(1)x(n-1)$$

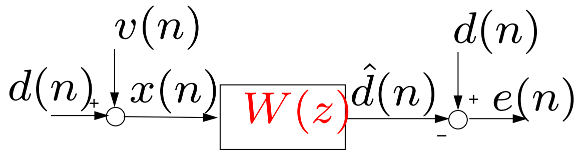
The Wiener-Hopf equations read for this case (all signals real and WSS)

$$\begin{bmatrix} E[x(n)^2] & E[x(n)x(n-1)] \\ E[x(n-1)x(n)] & E[x(n-1)^2] \end{bmatrix} \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} = \begin{bmatrix} E[d(n)x(n)] \\ E[d(n)x(n-1)] \end{bmatrix}$$

To determine $r_x(0), r_x(1), r_{dx}(0), r_{dx}(1)$ a signal model for $x(n)$ (and $d(n)$) is required.

$$\begin{aligned} d(n) &= ad(n-1) + w(n) \quad w(n)\ \mathrm{ZMWN}((1-a^2)) \\ x(n) &= d(n) + v(n) \quad E[w(n)v(\ell)] = 0 \; \forall n, \ell \Rightarrow E[d(n)v(\ell)] = 0 \end{aligned}$$

Example 1: Denoising



$x(n) = d(n) + v(n) \quad E[d(n)v(\ell)] = 0 \;\; \forall n, \ell$
$\hat{d}(n) = \hat{w}(0)x(n) + \hat{w}(1)x(n-1)$

The Wiener-Hopf equations read for this case (all signals real and WSS)

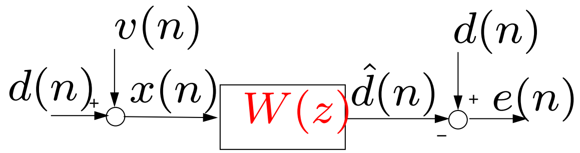
$$\begin{bmatrix} E[x(n)^2] & E[x(n)x(n-1)] \\ E[x(n-1)x(n)] & E[x(n-1)^2] \end{bmatrix} \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} = \begin{bmatrix} E[d(n)x(n)] \\ E[d(n)x(n-1)] \end{bmatrix}$$

To determine $r_x(0), r_x(1), r_{dx}(0), r_{dx}(1)$ a signal model for $x(n)$ (and $d(n)$) is required.

$$\begin{aligned} d(n) &= ad(n-1) + w(n) \quad w(n)\ \mathrm{ZMWN}((1-a^2)) \\ x(n) &= d(n) + v(n) \quad E[w(n)v(\ell)] = 0\ \forall n, \ell \Rightarrow E[d(n)v(\ell)] = 0 \end{aligned}$$

By the last assumption we have $E[d(n)x^*(n-\ell)] = E[d(n)d^*(n-\ell)] = r_d(\ell)$ and the Wiener-Hopf equations reduces to

Example 1: Denoising



$x(n) = d(n) + v(n) \quad E[d(n)v(\ell)] = 0 \ \ \forall n, \ell$
$\hat{d}(n) = \hat{w}(0)x(n) + \hat{w}(0)x(n-1)$

The Wiener-Hopf equations read for this case (all signals real and WSS)

$$\begin{bmatrix} r_x(0) & r_x(1) \\ r_x(1) & r_x(0) \end{bmatrix} \begin{bmatrix} \hat{w}(0) \\ \hat{w}(1) \end{bmatrix} = \begin{bmatrix} r_d(0) \\ r_d(1) \end{bmatrix}$$

To determine $r_x(0), r_x(1), r_{dx}(0), r_{dx}(1)$ a signal model for $x(n)$ (and $d(n)$) is required.

$$\begin{aligned} d(n) &= ad(n-1) + w(n) \quad w(n) \text{ ZMWN}((1-a^2)) \\ x(n) &= d(n) + v(n) \quad E[w(n)v(\ell)] = 0 \ \forall n, \ell \Rightarrow E[d(n)v(\ell)] = 0 \end{aligned}$$

By the last assumption we have $E[d(n)x^*(n-\ell)] = E[d(n)d^*(n-\ell)] = r_d(\ell)$ and the Wiener-Hopf equations reduces to

Example 1: Denoising (Ct'd)

From the model information

$$\begin{aligned} d(n) &= ad(n-1)+w(n) \quad w(n)\ \mathrm{ZMWN}((1-a^2)) \\ x(n) &= d(n)+v(n) \quad E[w(n)v^*(\ell)]=0\ \forall n,\ell \Rightarrow E[d(n)v^*(\ell)]=0 \end{aligned}$$

The Auto-correlation function of $x(n)$ is (Lecture 4)

$$r_x(k) = \underbrace{|a|^k}_{r_d(k)} + \sigma_v^2 \delta(k)$$

And the Wiener-Hopf equations become $\begin{bmatrix} 1+\sigma_v^2 & a \\ a & 1+\sigma_v^2 \end{bmatrix} \begin{bmatrix} \hat{w}_0 \\ \hat{w}_1 \end{bmatrix} = \begin{bmatrix} 1 \\ a \end{bmatrix}$

Resulting in the filter $W(z) = \frac{1}{(1+\sigma_v^2)^2-a^2}[(1+\sigma_v^2-a^2)+a\sigma_v^2 z^{-1}]$

and $\xi_{\min} = 1 - \begin{bmatrix} 1 & a \end{bmatrix} \begin{bmatrix} \hat{w}_0 \\ \hat{w}_1 \end{bmatrix}$

`Ex1Wf.m`

Optimal FIR filtering

1. Recap 4 filtering problems
2. A generic framework
3. The FIR Wiener filter
4. Three specific applications
   - Denoising
   - **Multi-step prediction**
   - Active noise cancellation
5. The IIR Wiener filter
6. The denoising Problem

Multi-step Prediction of $d(n+\alpha)$ from $d(n), d(n-1), \cdots, d(n-m$

The cost function we seek to optimize is:

$$\xi(\mathbf{w}) = E[\left|\mathbf{w}^T\mathbf{x}(\mathbf{n}) - d(n+\alpha)\right|^2]$$

with $\mathbf{x}(\mathbf{n}) = [\, d(n) \;\; d(n-1) \;\; \cdots \;\; d(n-m+1) \,]^T$. Then,
$E[d(n+\alpha)\mathbf{x}^*(\mathbf{n})]$ equals $[r_d(\alpha) \;\; r_d(\alpha+1) \;\; \cdots \;\; r_d(\alpha+m-1)]$,
and the WH-equations become:

$$\begin{bmatrix} r_d(0) & r_d^*(1) & \cdots & r_d^*(m-1) \\ r_d(1) & r_d(0) & & r_d(m-2) \\ \vdots & & \ddots & \\ r_d(m-1) & r_d(m-2) & \cdots & r_d(0) \end{bmatrix} \begin{bmatrix} \hat{w}_0 \\ \hat{w}_1 \\ \vdots \\ \hat{w}_{m-1} \end{bmatrix} = \begin{bmatrix} r_d(\alpha) \\ r_d(\alpha+1) \\ \vdots \\ r_d(\alpha+m-1) \end{bmatrix}$$

Extension: Multistep prediction in case of noisy measurements:

$$x(n) = d(n) + v(n) \quad v(n) \text{ ZMWN independent from } d(n)$$

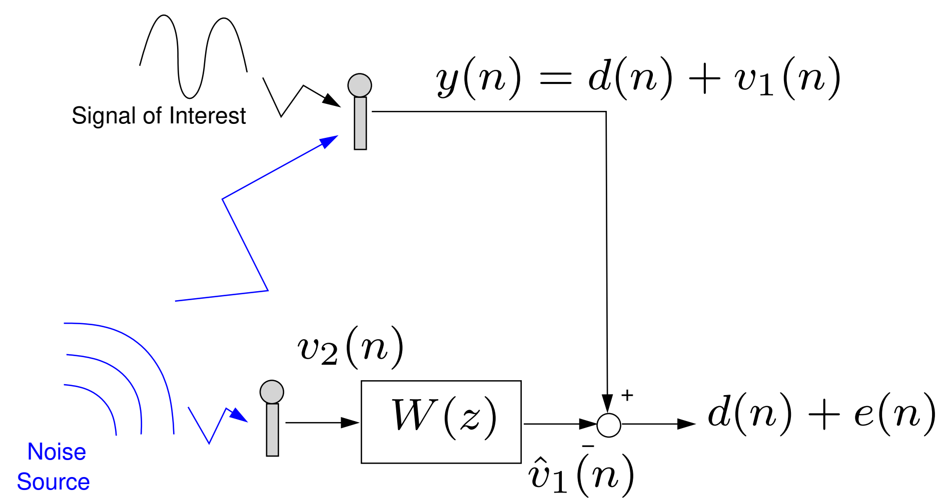
Optimal FIR filtering

1. Recap 4 filtering problems
2. A generic framework
3. The FIR Wiener filter
4. Three specific applications
   - Denoising
   - Multi-step prediction
   - **Active noise cancellation**
5. The IIR Wiener filter
6. The denoising Problem

Active Noise cancellation

We observe the 2 signals $y(n)$, $v_2(n)$ with:

$$y(n) = d(n) + v_1(n)$$



And define (for simplicity) $\mathbf{v_2(n)} = [v_2(n) \;\; v_2(n-1)]^T$.

Then predicting $v_1(n)$ from $v_2(n)$ is done via:

$$\hat{v}_1(n) = \hat{\mathbf{w}}^T \mathbf{v_2(n)} \qquad \hat{\mathbf{w}}^T = \text{arg min} E[|v_1(n) - \begin{bmatrix} v_2(n) & v_2(n-1) \end{bmatrix} \mathbf{w}|^2]$$

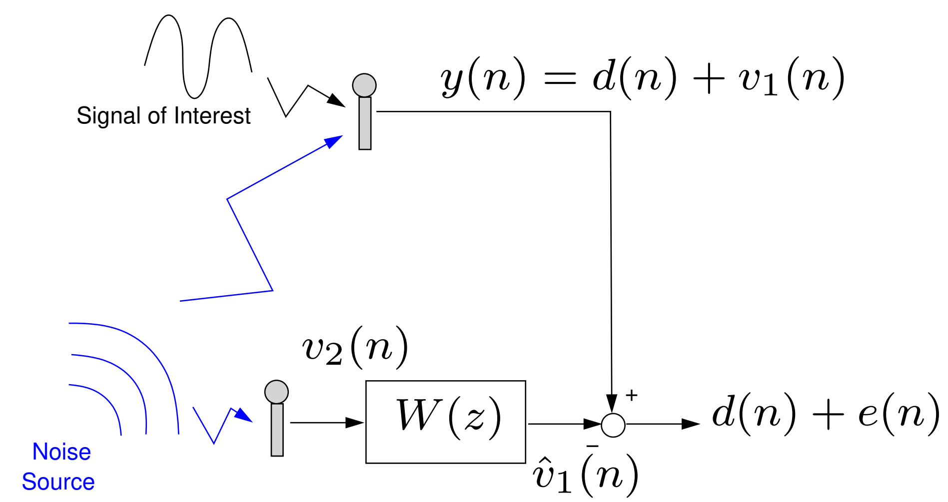
The Wiener-Hopf equations are:

$$\begin{bmatrix} E[v_2^*(n)v_2(n)] & E[v_2^*(n)v_2(n-1)] \\ E[v_2^*(n-1)v_2(n)] & E[v_2^*(n-1)v_2(n-1) \end{bmatrix} \underbrace{\begin{bmatrix} \hat{w}_0 \\ \hat{w}_1 \end{bmatrix}}_{\hat{\mathbf{w}}} = \begin{bmatrix} E[v_2^*(n)v_1(n)] \\ E[v_2^*(n-1)v_1(n)] \end{bmatrix}$$

Active Noise cancellation

We observe the 2 signals $y(n)$, $v_2(n)$ with:

$$y(n) = d(n) + v_1(n)$$



And define (for simplicity) $\mathbf{v_2(n)} = [v_2(n) \;\; v_2(n-1)]^T$.

Then predicting $v_1(n)$ from $v_2(n)$ is done via:

$$\hat{v}_1(n) = \hat{\mathbf{w}}^T \mathbf{v_2(n)} \qquad \hat{\mathbf{w}}^T = \text{arg min} E[|v_1(n) - \begin{bmatrix} v_2(n) & v_2(n-1) \end{bmatrix} \mathbf{w}|^2]$$

The Wiener-Hopf equations are (assuming $E[d(n)v_2^*(\ell)] = 0 \;\; \forall n, \ell$):

$$\begin{bmatrix} E[v_2^*(n)v_2(n)] & E[v_2^*(n)v_2(n-1)] \\ E[v_2^*(n-1)v_2(n)] & E[v_2^*(n-1)v_2(n-1) \end{bmatrix} \underbrace{\begin{bmatrix} \hat{w}_0 \\ \hat{w}_1 \end{bmatrix}}_{\hat{\mathbf{w}}} = \begin{bmatrix} E[v_2^*(n)y(n)] \\ E[v_2^*(n-1)y(n)] \end{bmatrix}$$

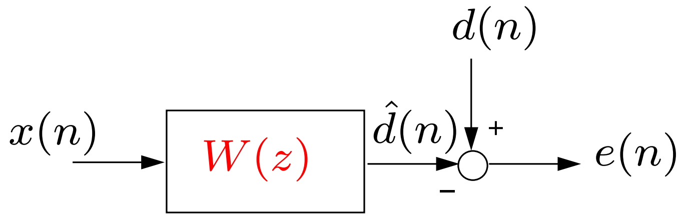
Optimal FIR filtering

1. Recap 4 filtering problems
2. A generic framework
3. The FIR Wiener filter
4. Three specific applications
5. **The IIR Wiener filter**
6. The denoising Problem

The IIR Wiener filter problem

Let the filter $W(z)$ be LSI with a double sided impulse response.

Let $x(n), d(n)$ be WSS with mean zero.



Then the estimate $\hat{d}(n)$ is given by:

$$\hat{d}(n) = \sum_{\ell=-\infty}^{\infty} w(\ell)x(n-\ell) = w(n) \star x(n)$$

The optimality to find the coefficients $w(\ell)$ is expressed by minimizing the mean-square error:

$$\xi(\mathbf{w}) = E[|e(n)|^2] = E\left[\left|d(n) - \hat{d}(n)\right|^2\right] = E\left[\left|d(n) - w(n) \star x(n)\right|^2\right]$$

The solution to the IIR Wiener filter problem

THEOREM: Let the conditions stipulated in the previous slide hold, let in addition $\{\mathbf{x}(\mathbf{n}), d(n)\}$ be jointly WSS and the following power and cross-spectra be given:

$$P_{\mathbf{x}}(e^{j\omega}) > 0 \quad P_{d\mathbf{x}}(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_{d\mathbf{x}}(k) e^{-j\omega k}$$

then the estimate of the signal $d(n)$ from $x(n)$ is derived from

$$\hat{W}(z) = \arg \min_{W(z)} \xi\Big(W(z)\Big) \quad \text{and given as} \quad \hat{D}(z) = \hat{W}(z) X(z)$$

with the optimal filter $\hat{W}(z)$ given by,

The solution to the IIR Wiener filter problem

THEOREM: Let the conditions stipulated in the previous slide hold, let in addition $\{\mathbf{x}(\mathbf{n}), d(n)\}$ be jointly WSS and the following power and cross-spectra be given:

$$P_{\mathbf{x}}(e^{j\omega}) > 0 \quad P_{d\mathbf{x}}(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_{d\mathbf{x}}(k) e^{-j\omega k}$$

then the estimate of the signal $d(n)$ from $x(n)$ is derived from

$$\hat{W}(z) = \arg\min_{W(z)} \xi\Big(W(z)\Big) \quad \text{and given as} \quad \hat{D}(z) = \hat{W}(z) X(z)$$

with the optimal filter $\hat{W}(z)$ given by,

$$\hat{W}(z) \quad = \quad P_{d\mathbf{x}}(z) P_{\mathbf{x}}(z)^{-1} \quad (\text{"The Wiener-Hopf equations (WH)"})$$

The solution to the IIR Wiener filter problem

THEOREM: Let the conditions stipulated in the previous slide hold, let in addition $\{\mathbf{x}(\mathbf{n}), d(n)\}$ be jointly WSS and the following power and cross-spectra be given:

$$P_{\mathbf{x}}(e^{j\omega}) > 0 \quad P_{d\mathbf{x}}(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_{d\mathbf{x}}(k)e^{-j\omega k}$$

then the estimate of the signal $d(n)$ from $x(n)$ is derived from

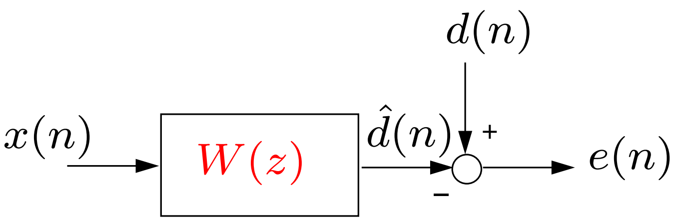
$$\hat{W}(z) = \arg\min_{W(z)} \xi\Big(W(z)\Big) \quad \text{and given as} \quad \hat{D}(z) = \hat{W}(z)X(z)$$

with the optimal filter $\hat{W}(z)$ given by,

$$\hat{W}(z) = P_{d\mathbf{x}}(z)P_{\mathbf{x}}(z)^{-1} \quad (\text{"The Wiener-Hopf equations (WH)"})$$

$$\xi\Big(\hat{W}(z)\Big) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_d(e^{j\omega}) - P_{d\mathbf{x}}(e^{j\omega})P_{\mathbf{x}}(e^{j\omega})^{-1}P_{d\mathbf{x}}^{*}(e^{j\omega})d\omega$$

(1) Solution via The orthogonality condition (OC)



$$
\begin{aligned}
e(n) &= d(n) - \sum_{\ell=-\infty}^{\infty} w(\ell)x(n-\ell) \\
\xi(\mathbf{w}) &= E[e(n)e^*(n)]
\end{aligned}
$$

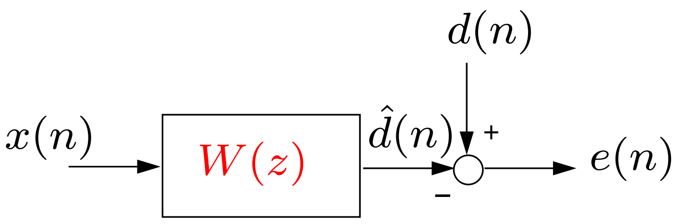
The OC reads: $\boxed{E[e(n)x^*(n-k)] = 0 \;\; \forall k : -\infty < k < \infty}$

$\Rightarrow$

$\Rightarrow$

$\Rightarrow$

(1) Solution via The orthogonality condition (OC)



$$
\begin{aligned}
e(n) &= d(n) - \sum_{\ell=-\infty}^{\infty} w(\ell) x(n-\ell) \\
\xi(\mathbf{w}) &= E[e(n)e^*(n)]
\end{aligned}
$$

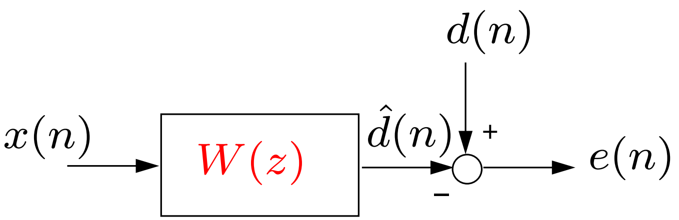
The OC reads: $\boxed{E[e(n)x^*(n-k)] = 0 \;\; \forall k : -\infty < k < \infty}$

$$\Rightarrow \quad E[d(n)x^*(n-k)] - E[\sum_{\ell=-\infty}^{\infty} \hat{w}(\ell) x(n-\ell) x^*(n-k)] = 0$$

$$\Rightarrow$$

$$\Rightarrow$$

(1) Solution via The orthogonality condition (OC)



$$\begin{aligned} e(n) &= d(n) - \sum_{\ell=-\infty}^{\infty} w(\ell)x(n-\ell) \\ \xi(\mathbf{w}) &= E[e(n)e^*(n)] \end{aligned}$$

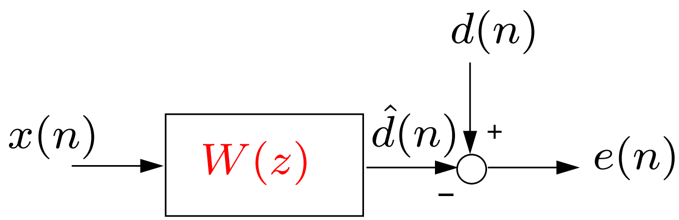
The OC reads: $\boxed{E[e(n)x^*(n-k)] = 0 \;\; \forall k : -\infty < k < \infty}$

$$\Rightarrow \quad E[d(n)x^*(n-k)] - E[\sum_{\ell=-\infty}^{\infty} \hat{w}(\ell)x(n-\ell)x^*(n-k)] = 0$$

$$\Rightarrow \quad r_{dx}(k) - \sum_{\ell=-\infty}^{\infty} \hat{w}(\ell)r_x(k-\ell) = 0$$

$$\Rightarrow$$

(1) Solution via The orthogonality condition (OC)



$$
\begin{aligned}
e(n) &= d(n) - \sum_{\ell=-\infty}^{\infty} w(\ell)x(n-\ell) \\
\xi(\mathbf{w}) &= E[e(n)e^*(n)]
\end{aligned}
$$

The OC reads: $\boxed{E[e(n)x^*(n-k)] = 0 \;\; \forall k : -\infty < k < \infty}$

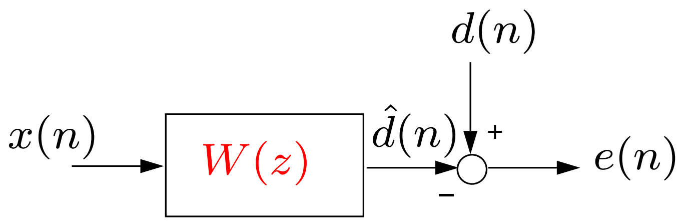
$$\Rightarrow \quad E[d(n)x^*(n-k)] - E[\sum_{\ell=-\infty}^{\infty} \hat{w}(\ell)x(n-\ell)x^*(n-k)] = 0$$

$$\Rightarrow \quad r_{dx}(k) - \sum_{\ell=-\infty}^{\infty} \hat{w}(\ell)r_x(k-\ell) = 0$$

$$\Rightarrow \quad r_{dx}(k) - \hat{w}(k) \star r_x(k) = 0$$

$$WH \quad \Rightarrow \quad \boxed{P_{dx}(e^{j\omega}) - \hat{W}(e^{j\omega})P_x(e^{j\omega}) = 0 \Rightarrow \hat{W}(z) = P_{dx}(z)P_x(z)^{-1}}$$

(2) The optimal value of the criterium $\xi_{\min}$

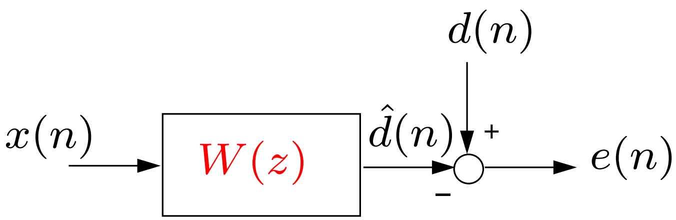


$$e(n) = d(n) - \sum_{\ell=-\infty}^{\infty} w(\ell)x(n-\ell)$$

$$\xi(\mathbf{w}) = E[e(n)e^*(n)]$$

$$\xi_{min} = E[e_{min}(n)d^*(n)]$$

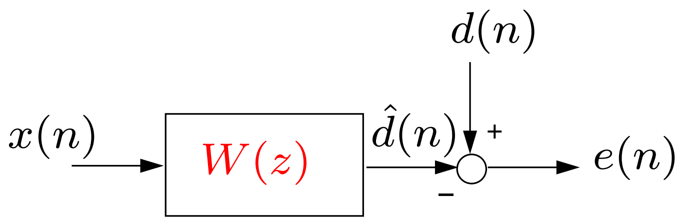
(2) The optimal value of the criterium $\xi_{\min}$



$$
\begin{aligned}
e(n) &= d(n) - \sum_{\ell=-\infty}^{\infty} w(\ell)x(n-\ell) \\
\xi(\mathbf{w}) &= E[e(n)e^*(n)]
\end{aligned}
$$

$$
\begin{aligned}
\xi_{min} &= E[e_{min}(n)d^*(n)] \\
&= E\Big[\Big(d(n) - \sum_{\ell=-\infty}^{\infty} \hat{w}(\ell)x(n-\ell)\Big)d^*(n)\Big]
\end{aligned}
$$

(2) The optimal value of the criterium $\xi_{\min}$

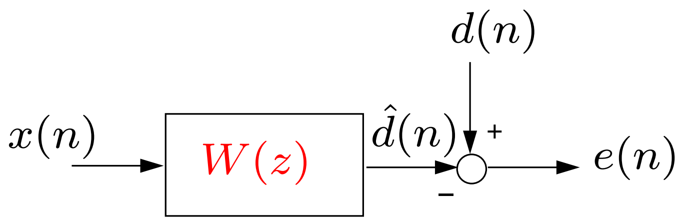


$$e(n) = d(n) - \sum_{\ell=-\infty}^{\infty} w(\ell)x(n-\ell)$$

$$\xi(\mathbf{w}) = E[e(n)e^*(n)]$$

$$\begin{aligned}
\xi_{min} &= E[e_{min}(n)d^*(n)] \\
&= E\left[\left(d(n) - \sum_{\ell=-\infty}^{\infty} \hat{w}(\ell)x(n-\ell)\right)d^*(n)\right] \\
&= r_d(0) - \sum_{\ell=-\infty}^{\infty} \hat{w}(\ell)\underbrace{E[d(n)x^*(n-\ell)]^*}_{r^*_{dx}(\ell)}
\end{aligned}$$

(2) The optimal value of the criterium $\xi_{\min}$



$$e(n) = d(n) - \sum_{\ell=-\infty}^{\infty} w(\ell)x(n-\ell)$$

$$\xi(\mathbf{w}) = E[e(n)e^*(n)]$$

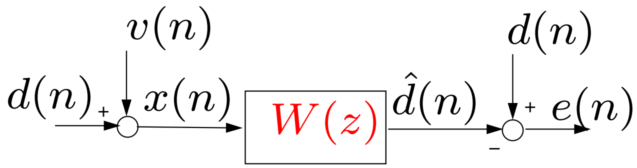
$$\begin{aligned}
\xi_{min} &= E[e_{min}(n)d^*(n)] \\
&= E\Big[\Big(d(n) - \sum_{\ell=-\infty}^{\infty} \hat{w}(\ell)x(n-\ell)\Big)d^*(n)\Big] \\
&= r_d(0) - \sum_{\ell=-\infty}^{\infty} \hat{w}(\ell)\underbrace{E[d(n)x^*(n-\ell)]^*}_{r_{dx}^*(\ell)}
\end{aligned}$$

$$\boxed{\xi_{\min} \overset{\text{Parseval}}{=} \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[P_d(e^{j\omega}) - \hat{W}(e^{j\omega})P_{dx}^*(e^{j\omega})\right]d\omega}$$

Optimal FIR filtering

1. Recap 4 filtering problems
2. A generic framework
3. The FIR Wiener filter
4. Three specific applications
5. The IIR Wiener filter
6. **The denoising Problem**

Denoising signals



$$x(n) = d(n) + v(n) \quad E[d(n)v^*(\ell)] = 0 \ \ \forall n, \ell$$

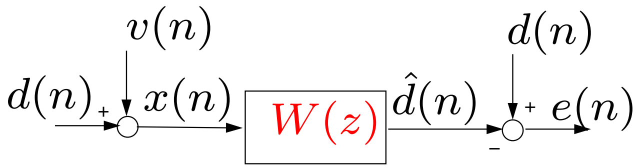
$$\hat{d}(n) = \sum_{\ell=-\infty}^{\infty} w(\ell)x(n-\ell)$$

$$\hat{W}(z) = P_{dx}(z)P_x(z)^{-1}$$

$$\begin{aligned} r_{dx}(k) &= E[d(n)x^*(n-k)] = E[d(n)d^*(n-k)] = r_d(k) \\ r_x(k) &= E[(d(n)+v(n))(d^*(n-k)+v^*(n-k))] = r_d(k) + r_v(k) \end{aligned}$$

$$\Rightarrow \hat{W}(z) = \frac{P_d(z)}{P_d(z) + P_v(z)}$$

Denoising signals



$x(n) = d(n) + v(n) \quad E[d(n)v^*(\ell)] = 0 \ \ \forall n, \ell$

$\hat{d}(n) = \sum_{\ell=-\infty}^{\infty} w(\ell) x(n-\ell)$

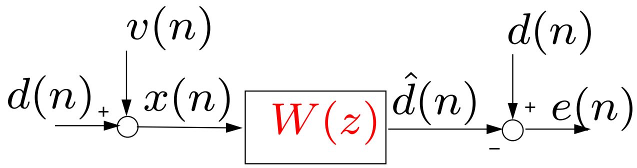
$$\xi_{\min} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[P_d(e^{j\omega}) - \hat{W}(e^{j\omega})P_{dx}^*(e^{j\omega})\right]d\omega \text{ and } \begin{cases} \hat{W}(z) = \frac{P_d(z)}{P_d(z)+P_v(z)} \\ r_{dx}(k) = r_d(k) \end{cases}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[P_d(e^{j\omega}) - \frac{P_d(e^{j\omega})}{P_d(e^{j\omega}) + P_v(e^{j\omega})}P_d^*(e^{j\omega})\right]d\omega$$

$$=$$

$$=$$

Denoising signals



$$x(n) = d(n) + v(n) \quad E[d(n)v^*(\ell)] = 0 \;\; \forall n, \ell$$

$$\hat{d}(n) = \sum_{\ell=-\infty}^{\infty} w(\ell)x(n-\ell)$$

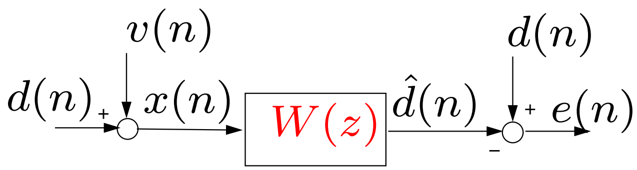
$$\xi_{\min} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[P_d(e^{j\omega}) - \hat{W}(e^{j\omega})P_{dx}^*(e^{j\omega})\right]d\omega \text{ and } \begin{cases} \hat{W}(z) = \frac{P_d(z)}{P_d(z)+P_v(z)} \\ r_{dx}(k) = r_d(k) \end{cases}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[P_d(e^{j\omega}) - \frac{P_d(e^{j\omega})}{P_d(e^{j\omega}) + P_v(e^{j\omega})}P_d^*(e^{j\omega})\right]d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{P_d(e^{j\omega})^2 + P_d(e^{j\omega})P_v(e^{j\omega}) - P_d(e^{j\omega})^2}{P_d(e^{j\omega}) + P_v(e^{j\omega})}\right]d\omega$$

$$=$$

Denoising signals



$x(n) = d(n) + v(n) \quad E[d(n)v^*(\ell)] = 0 \ \ \forall n, \ell$

$\hat{d}(n) = \sum_{\ell=-\infty}^{\infty} w(\ell) x(n-\ell)$

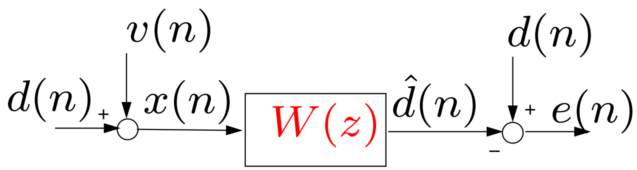
$$\xi_{\min} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[P_d(e^{j\omega}) - \hat{W}(e^{j\omega})P_{dx}^*(e^{j\omega})\right]d\omega \text{ and } \begin{cases} \hat{W}(z) = \frac{P_d(z)}{P_d(z)+P_v(z)} \\ r_{dx}(k) = r_d(k) \end{cases}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[P_d(e^{j\omega}) - \frac{P_d(e^{j\omega})}{P_d(e^{j\omega}) + P_v(e^{j\omega})}P_d^*(e^{j\omega})\right]d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{P_d(e^{j\omega})^2 + P_d(e^{j\omega})P_v(e^{j\omega}) - P_d(e^{j\omega})^2}{P_d(e^{j\omega}) + P_v(e^{j\omega})}\right]d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[P_v(e^{j\omega})\frac{P_d(e^{j\omega})}{P_d(e^{j\omega}) + P_v(e^{j\omega}}\right]d\omega$$

Denoising signals



$x(n) = d(n) + v(n) \quad E[d(n)v^*(\ell)] = 0 \ \ \forall n, \ell$

$\hat{d}(n) = \sum_{\ell=-\infty}^{\infty} w(\ell) x(n-\ell)$

$$\xi_{\min} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[P_d(e^{j\omega}) - \hat{W}(e^{j\omega})P_{dx}^*(e^{j\omega})\right]d\omega \text{ and } \begin{cases} \hat{W}(z) = \frac{P_d(z)}{P_d(z)+P_v(z)} \\ r_{dx}(k) = r_d(k) \end{cases}$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[P_d(e^{j\omega}) - \frac{P_d(e^{j\omega})}{P_d(e^{j\omega}) + P_v(e^{j\omega})}P_d^*(e^{j\omega})\right]d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[\frac{P_d(e^{j\omega})^2 + P_d(e^{j\omega})P_v(e^{j\omega}) - P_d(e^{j\omega})^2}{P_d(e^{j\omega}) + P_v(e^{j\omega})}\right]d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left[P_v(e^{j\omega})\frac{P_d(e^{j\omega})}{P_d(e^{j\omega}) + P_v(e^{j\omega}}\right]d\omega$$

Conclusion: If $v(n)$ and $d(n)$ have spectra that do not overlap, their product is $0 \ \forall\omega \Rightarrow \xi_{\min} = 0$.

Example 1 (Ct'd): Denoising (real case)

Consider the AR(1) process $d(n)$ given by ($a = 0.8$):
$d(n+1) = ad(n) + r(n)$ for $r(n)$ ZMWN($\sigma_r^2 = 1 - a^2$) and let the noise $v(n)$ in $x(n) = d(n) + v(n)$ to be also ZMWN($\sigma_v^2 = 1$), then the optimal IIR Wiener filter is:

$$\begin{aligned} \hat{W}(z) &= \frac{P_d(z)}{P_d(z) + P_v(z)} = \frac{(1-a^2)}{(1-a^2) + (1-az^{-1})(1-az)} \\ &= \frac{0.225}{(1-0.5z^{-1})(1-0.5z)} \end{aligned}$$

Resulting in the value of the cost function

$$\xi_{\min} = \sigma_v^2 \hat{w}(0) = 0.3$$

Ex1Wf.m