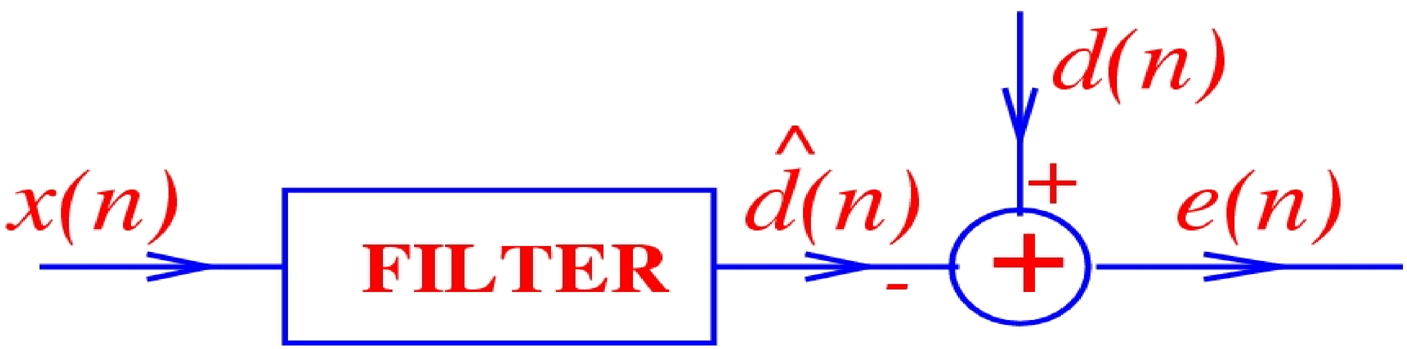
Optimal Filtering

Optimal filtering is a means of adaptive extraction of a *weak* desired signal in the presence of noise and interfering signals.

Mathematically: Given

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

estimate and extract $d(n)$ from the current and past values of $x(n)$.



Optimal Filtering (cont.)

Let the filter coefficients be

$$\boldsymbol{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{N-1} \end{bmatrix}.$$

Filter output:

$$y(n) = \sum_{k=0}^{N-1} w_k^* x(n-k) = \boldsymbol{w}^H \boldsymbol{x}(n) = \widehat{d}(n),$$

where

$$\boldsymbol{x}(n) = \begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-N+1) \end{bmatrix}.$$

Omitting index n, we can write

$$\widehat{d} = \boldsymbol{w}^H \boldsymbol{x}, \quad \boldsymbol{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix}.$$

Optimal Filtering (cont.)

Minimize the Mean Square Error (MSE)

$$\begin{aligned}\mathrm{MSE} &= \mathrm{E}\{|e|^2\} = \mathrm{E}\{|d - \boldsymbol{w}^H \boldsymbol{x}|^2\} \\ &= \mathrm{E}\{(d - \boldsymbol{w}^H \boldsymbol{x})(d^* - \boldsymbol{x}^H \boldsymbol{w})\} \\ &= \mathrm{E}\{|d|^2\} - \boldsymbol{w}^H \mathrm{E}\{\boldsymbol{x} d^*\} - \mathrm{E}\{d \boldsymbol{x}^H\} \boldsymbol{w} + \boldsymbol{w}^H \mathrm{E}\{\boldsymbol{x}\boldsymbol{x}^H\} \boldsymbol{w}.\end{aligned}$$

$$\min_{\boldsymbol{w}} \mathrm{MSE} \rightarrow \partial \mathrm{MSE} / \partial \boldsymbol{w} = \boldsymbol{0} \rightarrow \mathrm{E}\{\boldsymbol{x}\boldsymbol{x}^H\} \boldsymbol{w} - \mathrm{E}\{\boldsymbol{x} d^*\} = \boldsymbol{0}.$$

$$\begin{aligned} R &= \mathrm{E}\{\boldsymbol{x}\boldsymbol{x}^H\} \quad \text{correlation matrix} \\ \boldsymbol{r} &= \mathrm{E}\{\boldsymbol{x} d^*\}. \end{aligned}$$

Wiener-Hopf equation (see Hayes 7.2):

$$R\boldsymbol{w} = \boldsymbol{r} \quad \longrightarrow \quad \boldsymbol{w}_{\mathrm{opt}} = R^{-1} \boldsymbol{r}.$$

Wiener Filter

Three common filters:

1. general non-causal:

$$H(z) = \sum_{k=-\infty}^{\infty} h_k z^{-k},$$

2. general causal:

$$H(z) = \sum_{k=0}^{\infty} h_k z^{-k},$$

3. finite impulse response (FIR):

$$H(z) = \sum_{k=0}^{N-1} h_k z^{-k}.$$

Wiener Filter

Case 1: Non-causal filter:

$$\begin{aligned}\mathrm{MSE} &= \mathrm{E}\left[|e(n)|^2\right] = \mathrm{E}\left\{[d(n) - \sum_{k=-\infty}^{\infty} h_k x(n-k)][d(n) - \sum_{l=-\infty}^{\infty} h_l x(n-l)]^*\right\} \\ &= r_{dd}(0) - \sum_{l=-\infty}^{\infty} h_l^* r_{dx}(l) - \sum_{k=-\infty}^{\infty} h_k r_{dx}(k)^* + \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} r_{xx}(l-k) h_k h_l^*.\end{aligned}$$

Remark: for causal and FIR filters, only limits of sums differ.
Let $h_i = \alpha_i + j\beta_i$. $\partial\mathrm{MSE}/\partial\alpha_i = 0$, $\partial\mathrm{MSE}/\partial\beta_i = 0 \quad \forall i \Longrightarrow$

$$r_{dx}(i) = \sum_{k=-\infty}^{\infty} h_k r_{xx}(i-k) \quad \forall i.$$

In Z-domain

$$P_{dx}(z) = H(z)P_{xx}(z),$$

which is the optimal non-causal Wiener filter.

Example: $x(n) = d(n) + e(n)$,

$$\begin{aligned} P_{dd}(z) &= \frac{0.36}{(1-0.8z^{-1})(1-0.8z)}, \\ P_{ee}(z) &= 1, \end{aligned}$$

$d(n)$ and $e(n)$ uncorrelated.

Optimal filter?

$$
\begin{aligned}
P_{xx}(z) &= P_{dd}(z) + P_{ee}(z) \\
&= \frac{0.36}{(1-0.8z^{-1})(1-0.8z)} + 1 \\
&= 1.6\frac{(1-0.5z^{-1})(1-0.5z)}{(1-0.8z^{-1})(1-0.8z)}, \\
r_{dx}(z) &= \mathrm{E}\left[d(n+k)[d(n)^* + e(n)^*]\right] = r_{dd}(k) \\
P_{dx}(z) &= P_{dd}(z), \\
H(z) &= \frac{P_{dx}(z)}{P_{xx}(z)} = \frac{0.36}{1.6(1-0.5z^{-1})(1-0.5z)}, \\
h(k) &= 0.3(\tfrac{1}{2})^{|k|}.
\end{aligned}
$$

Case 2: Causal filter (7.3.2 in Hayes):

$$\begin{aligned}
\text{MSE} &= \mathrm{E}\left[|e(n)|^2\right] = \mathrm{E}\left\{[d(n) - \sum_{k=0}^{\infty} h_k x(n-k)][d(n) - \sum_{l=0}^{\infty} h_l x(n-l)]^*\right\} \\
&= r_{dd}(0) - \sum_{l=0}^{\infty} h_l^* r_{dx}(l) - \sum_{k=0}^{\infty} h_k r_{dx}(k)^* + \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} r_{xx}(l-k) h_k h_l^* \Longrightarrow \\
r_{dx}(i) &= \sum_{k=0}^{\infty} h_k r_{xx}(i-k) \quad \forall i.
\end{aligned}$$

Let

$$B(z)B^*(\frac{1}{z^*}) = \frac{1}{P_{xx}(z)}.$$

Pick $B(z)$ to be a stable, causal, minimum-phase system. Then

$$P_{dx}(z) = \underbrace{H(z)B^{-1}(z)}_{\text{causal}} B^{-*}(\frac{1}{z^*}) \Longrightarrow$$

$$\frac{H(z)}{B(z)} = [P_{dx}(z)B^*(\frac{1}{z^*})]_+,$$

where

$$[Y(z)]_+ = [\sum_{k=-\infty}^{\infty} y_k z^{-k}]_+ = \sum_{k=0}^{\infty} y_k z^{-k}.$$

$$\Longrightarrow H(z) = B(z)[P_{dx}(z)B^*(\frac{1}{z^*})]_+.$$

Causal Filter: Example

Same as before: $x(n) = d(n) + e(n)$,

$$
\begin{aligned}
P_{dd}(z) &= \frac{0.36}{(1 - 0.8z^{-1})(1 - 0.8z)}, \\
P_{ee}(z) &= 1,
\end{aligned}
$$

$d(n)$ and $e(n)$ uncorrelated.

Optimal filter?

$$\begin{aligned}
P_{dx}(z) &= P_{dd}(z), \\
P_{xx}(z) &= P_{dd}(z) + P_{ee}(z) \\
&= 1.6\frac{(1-0.5z^{-1})(1-0.5z)}{(1-0.8z^{-1})(1-0.8z)}, \\
B(z) &= \frac{1}{\sqrt{1.6}}\frac{1-0.8z^{-1}}{1-0.5z^{-1}} \quad \text{stable and causal,} \\
P_{dx}(z)B^*(\frac{1}{z^*}) &= \frac{0.36}{(1-0.8z^{-1})(1-0.8z)}\frac{1}{\sqrt{1.6}}\frac{1-0.8z}{1-0.5z} \\
&= \frac{0.36}{\sqrt{1.6}}\frac{1}{(1-0.8z^{-1})(1-0.5z)} \\
&= \frac{0.36}{\sqrt{1.6}}\Big[\frac{\frac{5}{3}}{1-0.8z^{-1}} + \frac{\frac{5}{6}z}{1-0.5z}\Big]
\end{aligned}$$

$$
\begin{aligned}
[P_{dx}(z)B^*(\frac{1}{z^*})]_+ &= \frac{0.36}{\sqrt{1.6}} \frac{\frac{5}{3}}{1-0.8z^{-1}} \\
H(z) &= B(z)[P_{dx}(z)B^*(\frac{1}{z^*})]_+ = \frac{1}{\sqrt{1.6}} \frac{1-0.8z^{-1}}{1-0.5z^{-1}} \frac{0.36}{\sqrt{1.6}} \frac{\frac{5}{3}}{1-0.8z^{-1}} \\
&= 0.375 \frac{1}{1-0.5z^{-1}} \Longrightarrow h(k) = \frac{3}{8}(\tfrac{1}{2})^k, \quad k = 0, 1, 2, \ldots
\end{aligned}
$$

Case 3: FIR filter (done before):

$$
r_{dx}(i) = \sum_{k=0}^{N-1} h_k r_{xx}(i-k) \quad \forall i.
$$

FIR Wiener Filter - Generalization

Theorem 1. *Assume that*

$$\begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{d} \end{bmatrix} \in \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_d \end{bmatrix} \begin{bmatrix} C_{xx} & C_{xd} \\ C_{dx} & C_{dd} \end{bmatrix} \right).$$

Then the posterior pdf of $\boldsymbol{d}$ *given* $\boldsymbol{x}$*,* $f_{d|x}(\boldsymbol{d}|\boldsymbol{x})$*, is also Gaussian, with moments given by*

$$\begin{aligned} \mathrm{E}\left[\boldsymbol{d}|\boldsymbol{x}\right] &= \boldsymbol{\mu}_d + \overbrace{C_{dx}C_{xx}^{-1}}^{W^H}(\boldsymbol{x} - \boldsymbol{\mu}_x) \\ C_{d|x} &= C_{dd} - C_{dx}C_{xx}^{-1}C_{xd}. \end{aligned}$$

FIR Wiener Filter - Generalization

$$\text{Consider} \quad \begin{aligned} \boldsymbol{d} &= A\boldsymbol{\beta} + \boldsymbol{\xi}, \\ \boldsymbol{y} &= B\boldsymbol{\beta} + \boldsymbol{\eta}, \quad \text{where} \end{aligned}$$

$$\mathrm{E}\left\{ \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{bmatrix} \right\} = \mathbf{0}, \quad \mathrm{cov}\left\{ \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\eta} \end{bmatrix} \right\} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}.$$

Also, assume $\mathrm{E}[\boldsymbol{\beta}] = \boldsymbol{\beta}_0$ and $\mathrm{cov}(\boldsymbol{\beta}) = \Gamma$. We wish to predict $\boldsymbol{d}$ given $\boldsymbol{y}$.

Theorem 2. Minimum Mean-Square Error (MMSE) Solution:

$$\widehat{\boldsymbol{d}} \quad = \quad A\boldsymbol{\beta}_0 + C(\boldsymbol{y} - B\boldsymbol{\beta}_0),$$

where $C = (A\Gamma B^H + V_{12})(B\Gamma B^H + V_{22})^{-1}$.

MMSE:

$$A\Gamma A^H + V_{11} + C(B\Gamma B^H + V_{22})C^H - (A\Gamma B^H + V_{12})C^H - C(B\Gamma A^H + V_{21}).$$

Kalman Filter

Consider a state-space model:

$$\begin{aligned} \boldsymbol{x}(t) &= A\boldsymbol{x}(t-1) + \boldsymbol{\xi}(t), \\ \boldsymbol{y}(t) &= B\boldsymbol{x}(t) + \boldsymbol{\eta}(t), \end{aligned}$$

where

- $\{\boldsymbol{\xi}(t)\}$ and $\boldsymbol{\eta}(t)$ are independent sequences of zero-mean random vectors with covariances Σ_ξ and $\Sigma_{\boldsymbol{\eta}}$, respectively;
- $\boldsymbol{\xi}(t)$ and $\boldsymbol{x}(u)$ are independent for $t > u$ and $\boldsymbol{\eta}(t)$ and $\boldsymbol{x}(u)$ are independent for $t \geq u$.

We wish to predict $\boldsymbol{x}(t)$ given $\boldsymbol{y}(1), \boldsymbol{y}(2), \ldots \boldsymbol{y}(t)$.

Idea: Theorem 2 can be used to predict $\boldsymbol{x}(s)$ given $\boldsymbol{y}(t)$. We write such predictor as $\widehat{\boldsymbol{x}}(s|t)$ and its (minimum) mean-square error as $P(s|t)$.

How to construct a recursive procedure?

Kalman Filter

Assume that we know $\widehat{\boldsymbol{x}}(t+1|t)$ and its MMSE $P(t+1|t)$. We wish to find $\widehat{\boldsymbol{x}}(t+1|t+1)$. Then, we can write the following model:

$$\begin{aligned}
\boldsymbol{x}(t+1) &= \boldsymbol{\beta} \\
\boldsymbol{y}(t+1) &= B\boldsymbol{\beta} + \boldsymbol{\eta},
\end{aligned}$$

where $\boldsymbol{\beta}_0 = \widehat{\boldsymbol{x}}(t+1|t)$ and $\Gamma = P(t+1|t)$. This implies

$$\widehat{\boldsymbol{x}}(t+1|t+1) = \widehat{\boldsymbol{x}}(t+1|t) + C(t+1) \cdot [\boldsymbol{y}(t+1) - B\widehat{\boldsymbol{x}}(t+1|t)],$$

where

$$C(t+1) = P(t+1|t)B^H[BP(t+1|t)B^H + \Sigma_\eta]^{-1}.$$

Also

$$P(t+1|t+1) = P(t+1|t) - C(t+1)[BP(t+1|t)B^H + \Sigma_\eta]^{-1}C(t+1)^H.$$

Kalman Filter

How about finding $\widehat{\boldsymbol{x}}(t+1|t)$ based on $\widehat{\boldsymbol{x}}(t|t)$? Then, we can write the following model:

$$\boldsymbol{x}(t+1) \quad = \quad A\boldsymbol{\beta} + \boldsymbol{\xi}(t+1),$$

where $\boldsymbol{\beta}_0 = \widehat{\boldsymbol{x}}(t|t)$ and $\Gamma = P(t|t)$. This implies

$$\begin{aligned} \widehat{\boldsymbol{x}}(t+1|t) &= A\widehat{\boldsymbol{x}}(t|t) \\ P(t+1|t) &= AP(t|t)A^H + \Sigma_\xi. \end{aligned}$$

Now we have all the equations for recursion!