L14: Optimal linear filtering – Wiener filtering

Tomas McKelvey

Department of Electrical Engineering
Chalmers University of Technology

Reviewing L11-L13

What have we done so far?

- **Signal models** (L11-L12)
  - Nonparametric models: ACF and PSD.
  - Parametric models: AR, MA and ARMA.
- **Signal model estimation** (L13)
  - Nonparametric spectral estimation: the periodogram.
    Pros:
    - fast to compute
    - asymptotically unbiased.

    Cons:
    - limited resolution for finite N:
      $\rightsquigarrow$ the modified periodogram improves this
    - large variance for all N:
      $\rightsquigarrow$ Blackman-Tukey's method lowers variance.
  - Parametric spectral estimation: AR-estimation.
    1. Estimate $r_x[k]$ from data.
    2. Reformulate Yule-Walker to get $\hat{a} = R_x^{-1} r_x$.

Learning objectives

After today's lecture you should be able to

- explain what type of problems **Wiener-filters** can solve.
- derive the **Wiener-Hopf** (WH) equations.
- use the WH-equations to derive a **causal FIR Wiener** filter.
- use the WH-equations to derive a **non-causal IIR Wiener** filter.
- Compute the **mean squared error** (MSE) of a Wiener-filter.

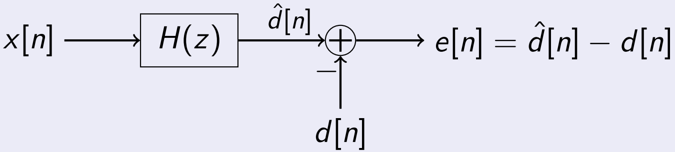
Filtering, smoothing and prediction

- Let $s[n]$ and $w[n]$ be zero mean, wide sense stationary processes and

$$x[n] = s[n] + w[n].$$

Objective

- Select $H(z)$ to make $e[n]$ as "small" as possible



- Small could mean different things. We use mean squared error

$$E\left\{e[n]^2\right\},$$

since this is easy to minimize.

Filtering, smoothing and prediction

Based on measurements collected up until now, we encounter three common problems ($k > 0$):

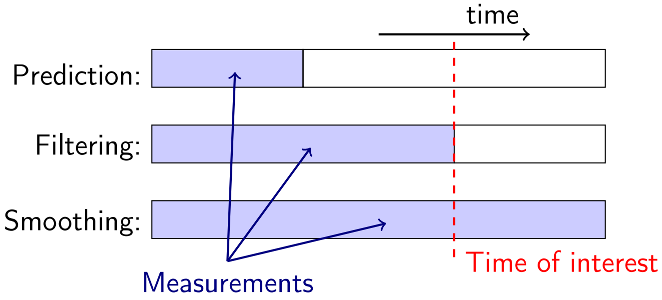
- **Filtering** – estimating current signal values, $d[n] = s[n]$. *Applications:* positioning, control systems, noise or echo cancellation, etc.

- **Smoothing** – estimating past signal values, $d[n] = s[n-k]$. *Applications:* signal analysis, image processing, system identification (modelling).

- **Prediction** – estimating future signal values, $d[n] = s[n+k]$. *Applications:* decision making, planning, weather forecasts, etc.

Filtering, smoothing and prediction

These problems can be illustrated as



Filtering, smoothing and prediction

- We seek a linear estimator (filter)

$$\hat{d}[n] = h[n] \star x[n] = \sum_k h[k]x[n-k]$$

  of $d[n]$.

- As mentioned above, we wish to minimize the mean square error (MSE),

$$\mathrm{MSE}(\mathbf{h}) = E\left\{\left(d[n] - \sum_k h[k]x[n-k]\right)^2\right\}$$

  where the vector $\mathbf{h}$ contains the impulse response coefficients $h[k]$.

- The resulting Wiener filter $\hat{d}[n]$ is a linear minimum mean square error (LMMSE) estimator.

Cross-correlation function

In this filtering case we have two signals x and d and when evaluating the MSE we will obtain terms $E\{d[n]x[n-k]\}$.

The *cross-correlation function* for signals d and x is defined as

$$r_{dx}[k] = E\{d[n]x[n-k]\}$$

and describes how the two signals co-vary.

Wiener-Hopf (W-H) equations

- The W-H equations are ***very important*** and can be used to solve all the problems mentioned above.

- **Objective**: (again) We wish to minimize

$$\text{MSE}(\mathbf{h}) = E\left\{\left(d[n] - \sum_k h[k]x[n-k]\right)^2\right\}$$

with respect to $\mathbf{h}$.

- **Derivation 1:** the function is quadratic in $\mathbf{h}$
  $\Rightarrow$ it is convex in $\mathbf{h}$
  $\Rightarrow$ no local optima (except for the global optimum)
  $\Rightarrow$ sufficient to differentiate and set to zero!

Wiener-Hopf (W-H) equations

- Differentiate the MSE:

$$\frac{\partial}{\partial h[t]}\mathrm{MSE}(\mathbf{h}) = \frac{\partial}{\partial h[t]} E\left\{\left(d[n] - \sum_k h[k]x[n-k]\right)^2\right\}$$
$$= E\left\{2\left(d[n] - \sum_k h[k]x[n-k]\right)(-x[n-t])\right\}$$
$$= -2r_{dx}[t] + 2\sum_k h[k]r_x[t-k]$$

- Setting this derivative to zero gives the

Wiener-Hopf (WH) equations

$$\sum_k h[k]r_x[t-k] = r_{dx}[t],$$

for all t. Optimal $h[t]$ must satisfy WH.

FIR filters

- Suppose $H(z)$ is a causal FIR filter:

$$\hat{d}[n] = \sum_{n=0}^{p-1} h[k]x[n-k].$$

- The W-H eq's can be written as

$$\underbrace{\begin{bmatrix} r_x[0] & r_x[1] & \dots & r_x[p-1] \\ r_x[1] & r_x[0] & \dots & r_x[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_x[p-1] & r_x|p-2] & \dots & r_x[0] \end{bmatrix}}_{\mathbf{R}_x} \underbrace{\begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[p-1] \end{bmatrix}}_{\mathbf{h}} = \underbrace{\begin{bmatrix} r_{dx}[0] \\ r_{dx}[1] \\ \vdots \\ r_{dx}[p-1] \end{bmatrix}}_{\mathbf{r}_{dx}}$$

which yields that

$$\mathbf{h}_{\text{opt}} = \mathbf{R}_x^{-1}\mathbf{r}_{dx}.$$

What is the minimum MSE?

- The minimum MSE can be calculated by plugging in $\mathbf{h}_{\text{opt}}$:

$$E\left\{e_{\min}^2[n]\right\} = E\left\{e_{\min}[n]\left(d[n] - \hat{d}_{\text{opt}}[n]\right)\right\} = \left\{\text{Note: } \hat{d}_{\text{opt}} \perp e[n]\right\}$$
$$= E\left\{\left(d[n] - \sum_{k=0}^{p-1} h_{\text{opt}}[k]x[n-k]\right) d[n]\right\}$$
$$= r_d[0] - \sum_{k=0}^{p-1} h_{\text{opt}}[k] r_{dx}[k] = r_d[0] - \mathbf{r}_{dx}^T \mathbf{R}_x^{-1} \mathbf{r}_{dx}$$

- **Special case**: if $d[n]$ and $x[n]$ are uncorrelated, then $\hat{d}[n] = 0$ and the MSE is $r_d[0]$.
- In general, the more correlated (similar) $x[n]$ is to $d[n]$ the better is the estimate $\hat{d}[n]$!

The LMMSE estimator

Consider two vectors z and v which are zero mean and have the joint covariance

$$E\left\{\begin{bmatrix} z \\ v \end{bmatrix}\begin{bmatrix} z \\ v \end{bmatrix}^T\right\} = \begin{bmatrix} E\left\{zz^T\right\} & E\left\{zv^T\right\} \\ E\left\{vz^T\right\} & E\left\{vv^T\right\} \end{bmatrix} = \begin{bmatrix} Q_{zz} & Q_{vz}^T \\ Q_{vz} & Q_{vv} \end{bmatrix}$$

Assume we want to estimate the value of z by forming a linear combination of v, i.e.

$$\hat{z} = Kv$$

such that

$$\text{MSE} = E\left\{\|z - \hat{z}\|^2\right\} = E\left\{(z - Kv)^T(z - Kv)\right\}$$

is minimized. *Linear Minimum Mean Squared Error Estimator (LMMSE)* .

Solution

Evaluating the MSE yields

$$\begin{aligned}\mathrm{MSE} =& E\left\{(z-Kv)^T(z-Kv)\right\} = \mathrm{tr} E\left\{(z-Kv)(z-Kv)^T\right\}\\ =& \mathrm{tr}(Q_{zz} - KQ_{vz} - Q_{vz}^T K^T + KQ_{vv}K^T)\\ =& \mathrm{tr}\left((K - Q_{vz}^T Q_{vv}^{-1})Q_{vv}(K - Q_{vz}^T Q_{vv}^{-1})^T\right) + \mathrm{tr}(Q_{zz} - Q_{vz}^T Q_{vv}^{-1} Q_{vz})\end{aligned}$$

The optimal K is hence

$$K_{opt} = Q_{vz}^T Q_{vv}^{-1}$$

and the optimal MSE is

$$\mathrm{MSE}_{opt} = \mathrm{tr}(Q_{zz} - Q_{vz}^T Q_{vv}^{-1} Q_{vz})$$

The Wiener filter is the LMMSE estimator

The FIR Wiener filter is obtained by setting

$$z = d[n] \quad \text{and} \quad v = \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[p-1] \end{bmatrix}$$

which imply

$$\begin{bmatrix} Q_{zz} & Q_{vz}^T \\ Q_{vz} & Q_{vv} \end{bmatrix} = \begin{bmatrix} r_d[0] & \mathbf{r}_{dx}^T \\ \mathbf{r}_{dx} & \mathbf{R}_x \end{bmatrix}$$

and hence,

$$K_{opt} = Q_{vz}^T Q_{vv}^{-1} = \mathbf{r}_{dx}^T \mathbf{R}_x^{-1} \quad \text{and} \quad \text{MSE}_{opt} = r_d[0] - \mathbf{r}_{dx}^T \mathbf{R}_x^{-1} \mathbf{r}_{dx}$$

$$\hat{d}[n] = \hat{z} = K_{opt} v = \mathbf{r}_{dx}^T \mathbf{R}_x^{-1} v = \sum_{k=0}^{p-1} h_{opt}[k] x[n-k]$$

Learning objectives

After today's lecture you should be able to

- explain what type of problems **Wiener-filters** can solve.
- derive the **Wiener-Hopf** (WH) equations.
- use the WH-equations to derive a **causal FIR Wiener** filter.
- use the WH-equations to derive a **non-causal IIR Wiener** filter.
- Compute the **mean square error** (MSE) of a Wiener-filter.