

Three Cases

1. $D = 0$

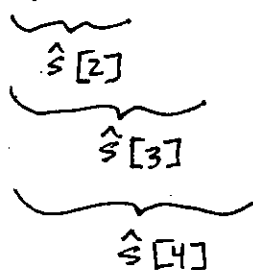
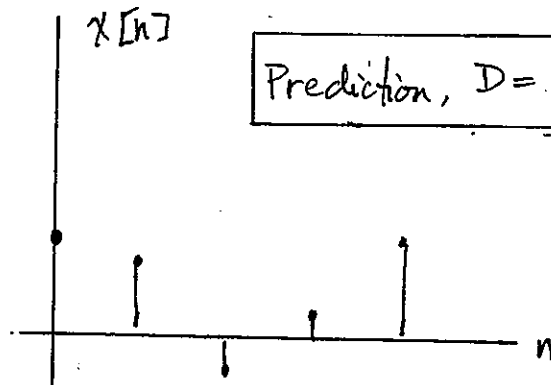
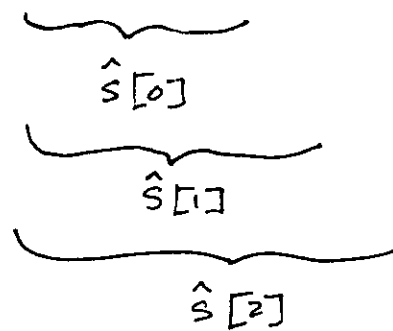
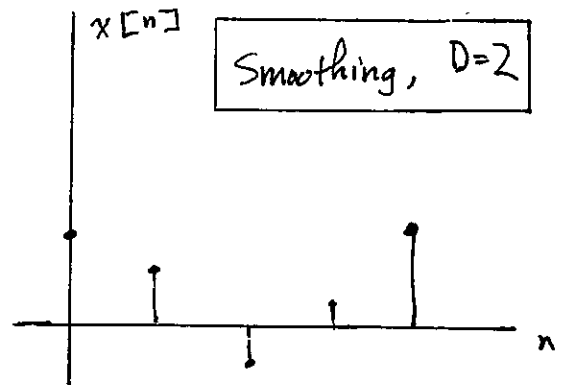
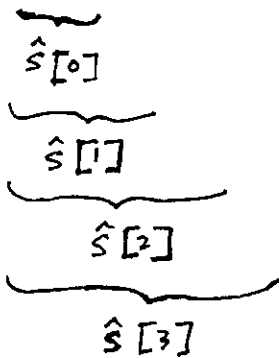
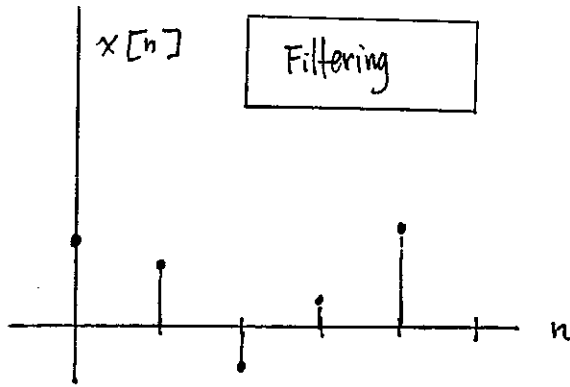
filtering

2. $D > 0$

signal prediction

3. $D < 0$

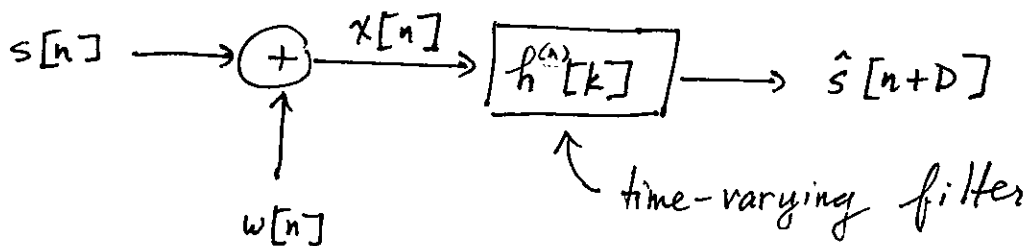
smoothing



Note | Signal prediction is different from the "measurement prediction" problem discussed previously:

$$\hat{s}[n+D] \neq \hat{x}[n+D].$$

Filtering Interpretation



Smoothing requires a noncausal filter in the sense that, to estimate a present signal value, you need data from the future.

Filtering and prediction, on the other hand, are causal operations.

Assumptions

We assume all first and second order moments are known, as required for LMMSE estimation. Furthermore, we assume

1. $s[n]$ and $w[n]$ are zero mean
2. $x[n]$ is wide-sense stationary (WSS) with autocorrelation

$$r_{xx}[k] = \mathbb{E} \{ x[n] x[n+k] \}$$

3. $x[n]$ and $s[n]$ are jointly WSS with cross-correlation

$$r_{xs}[k] = \mathbb{E} \{ x[n] s[n+k] \}$$

Example

These conditions hold when $s[n]$ and $w[n]$ are zero-mean, WSS, and uncorrelated.

Wiener - Hopf Equations

Let's focus on the filtering problem ($D=0$)

From LMMSE estimation theory we know the optimal filter satisfies the Wiener-Hopf equations:

$$R_{xx} \underline{h}_p = R_{xo}$$

where R_{xx} and R_{xo} are given in terms of $r_{xx}[k]$ and $r_{xs}[k]$.

So in theory, we can compute the Wiener filter. In practice, however, we want a fast, online algorithm for computing and updating \underline{h}_p as data streams in.

Generalized Levinson-Durbin Algorithm

The Wiener-Hopf equations are

$$\begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \dots & r_{xx}[p-1] \\ r_{xx}[1] & r_{xx}[0] & \dots & r_{xx}[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}[p-1] & r_{xx}[p-2] & \dots & r_{xx}[0] \end{bmatrix} \begin{bmatrix} h_p[0] \\ h_p[1] \\ \vdots \\ h_p[p-1] \end{bmatrix} = \begin{bmatrix} r_{xs}[0] \\ r_{xs}[1] \\ \vdots \\ r_{xs}[p-1] \end{bmatrix}$$

$$\begin{matrix} \downarrow & & \downarrow & & \downarrow \\ R_p & & \underline{h}_p & & \underline{b}_p \end{matrix}$$

This system of equations is very similar to the WH equation for linear (measurement) prediction.

The difference is that \underline{b}_p is not related to the rows/columns of R_p .

In linear prediction we had

$$R_p \cdot \underline{h}_p = \underline{r}_p, \quad \underline{r}_p = \begin{bmatrix} r_{xx}[1] \\ r_{xx}[2] \\ \vdots \\ r_{xx}[p] \end{bmatrix}$$

Let's try to update \underline{h}_p from \underline{h}_{p-1} :

$$\underline{h}_p = \begin{bmatrix} \underline{h}_{p-1} \\ 0 \end{bmatrix} + \begin{bmatrix} \underline{d}_{p-1} \\ k_p \end{bmatrix}$$

Recall

$$R_p = \begin{bmatrix} R_{p-1} & \tilde{\underline{r}}_{p-1} \\ \tilde{\underline{r}}_{p-1}^T & r_{xx}[0] \end{bmatrix}$$

$$\tilde{\underline{r}}_p = \underline{r}_p \text{ upside-down}$$

and also notice

$$\underline{b}_p = \begin{bmatrix} \underline{b}_{p-1} \\ r_{xs}[p-1] \end{bmatrix}$$

Thus we may write the WFI equations

$$\begin{bmatrix} R_{p-1} & \tilde{\underline{r}}_{p-1} \\ \tilde{\underline{r}}_{p-1}^T & r_{xx}[0] \end{bmatrix} \cdot \left\{ \begin{bmatrix} \underline{h}_{p-1} \\ 0 \end{bmatrix} + \begin{bmatrix} \underline{d}_{p-1} \\ k_p \end{bmatrix} \right\} = \begin{bmatrix} \underline{b}_{p-1} \\ r_{xs}[p-1] \end{bmatrix}$$



$$R_{p-1} \underline{h}_{p-1} + R_{p-1} \underline{d}_{p-1} + k_p \tilde{\underline{r}}_{p-1} = \underline{b}_{p-1} \quad (1)$$

$$\tilde{\underline{r}}_{p-1}^T \underline{h}_{p-1} + \tilde{\underline{r}}_{p-1}^T \underline{d}_{p-1} + k_p r_{xx}[0] = r_{xs}[p-1] \quad (2)$$

To simplify (1), observe

$$R_{p-1} \underline{h}_{p-1} = \underline{b}_{p-1}$$

which implies

$$\underline{d}_{p-1} =$$

How can this be simplified?

Recall the WH equations for linear prediction:

$$R_p \underline{g}_p = \underline{r}_p$$

where $\hat{x}[n] = \sum_{k=1}^p g_p[k] x[n-k]$ is the LMMSE predictor of $x[n]$.

Previously we used the fact that R_p is symmetric and Toeplitz to show

$$R_{p-1} \tilde{\underline{g}}_{p-1} = \tilde{\underline{r}}_{p-1}$$

$$\Rightarrow R_{p-1}^{-1} \tilde{\underline{r}}_{p-1} = \tilde{\underline{g}}_{p-1}$$

$$\Rightarrow \boxed{d_{p-1} = -k_p \cdot \tilde{\underline{g}}_{p-1}}$$

Exercise

Determine k_p

Solution | From equation (2), and plugging in

$$\underline{d}_{p-1} = -k_p \underline{\tilde{g}}_{p-1}, \text{ we obtain}$$

$$\underline{r}_{p-1}^T \underline{h}_{p-1} + k_p (r_{xx}[0] - \underline{r}_{p-1}^T \underline{\tilde{g}}_{p-1}) = r_{xs}[p-1]$$

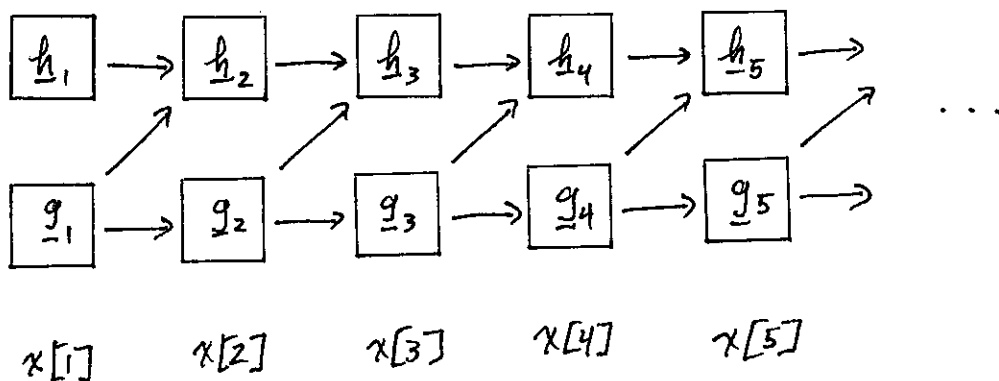
$$\Rightarrow \boxed{k_p = \frac{r_{xs}[p-1] - \underline{r}_{p-1}^T \underline{h}_{p-1}}{r_{xx}[0] - \underline{r}_{p-1}^T \underline{\tilde{g}}_{p-1}}$$

This leads to the Generalized Levinson-Durbin algorithm:

(a) Initialize $\underline{h}_1 =$ $\underline{g}_1 =$

Iterate

1. Update \underline{g}_p from \underline{g}_{p-1} using Levinson-Durbin recursion
2. Update \underline{h}_p from \underline{h}_{p-1} and \underline{g}_{p-1}



For prediction and smoothing, a similar algorithm can be derived. The WH equations are

$$R_p \underline{h}_p = \underline{b}_p$$

where

$$\underline{b}_p = \begin{bmatrix} r_{xs}[D] \\ r_{xs}[D+1] \\ \vdots \\ r_{xs}[D+p-1] \end{bmatrix}$$

The GLD recursion changes only slightly.

In general, the GLD algorithm requires $O(p^2)$ operations to compute $\{\underline{h}_1, \dots, \underline{h}_p\}$.

IIR Wiener Filtering

So far we have discussed "FIR Wiener Filtering," because $\hat{\theta} = \hat{s}[n+D]$ only depends on a finite number of observations $x[n], x[n-1], \dots, x[n-p+1]$, which implies \underline{h}_p has finitely many nonzero taps.

We will consider two IIR problems

- The causal, IIR Wiener filter

$$\hat{s}[n] = \sum_{k=0}^{\infty} h[k] x[n-k]$$

- The noncausal, IIR Wiener smoother

$$\hat{s}[n] = \sum_{k=-\infty}^{\infty} h[k] x[n-k]$$

Infinite Wiener Smoother

Given $\{x[n]\}_{n=-\infty}^{\infty}$ we seek the filter

$\{h[k]\}_{k=-\infty}^{\infty}$ such that

$$E \left\{ \left(s[n] - \sum_{k=-\infty}^{\infty} h[k] x[n-k] \right)^2 \right\}$$

is minimized.

Observe that $\hat{s}[n]$ is the projection of $s[n]$ onto the closed linear span of $\{x[n]\}_{n=-\infty}^{\infty}$.

By the orthogonality principle, we know

$$E \left\{ (s[n] - \hat{s}[n]) \cdot x[n-l] \right\} = 0 \quad \forall l$$

\Downarrow

$$E \left\{ \left(s[n] - \sum_{k=-\infty}^{\infty} h[k] x[n-k] \right) x[n-l] \right\} = 0 \quad \forall l$$

$$\Updownarrow$$

$$E \left\{ x[n-l] s[n] \right\} = \sum_{k=-\infty}^{\infty} h[k] E \left\{ x[n-l] x[n-k] \right\}$$

$$\Updownarrow$$

$$\boxed{\sum_{k=-\infty}^{\infty} h[k] r_{xx}[l-k] = r_{xs}[l] \quad \forall l \in \mathbb{Z}}$$

These are the Wiener-Hopf equations for the infinite Wiener smoother.

How can we solve for $\{h[k]\}_{k=-\infty}^{\infty}$?

Take the DTFT of both sides:

$$\text{DTFT} \{ h * r_{xx} \} = \text{DTFT} \{ r_{xs} \}$$

||

||

$$H(f) \cdot P_{xx}(f)$$

$$P_{xs}(f)$$

↑
spectral density

↑
cross spectral density

$$\Rightarrow \boxed{H(f) = \frac{P_{xs}(f)}{P_{xx}(f)}}$$

When signal and noise are uncorrelated

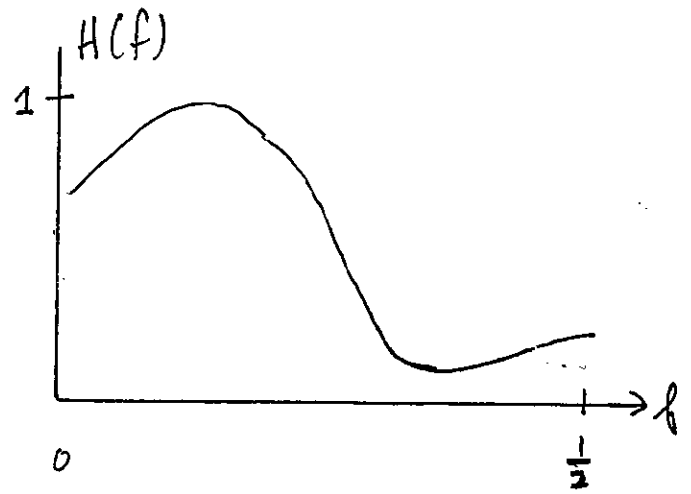
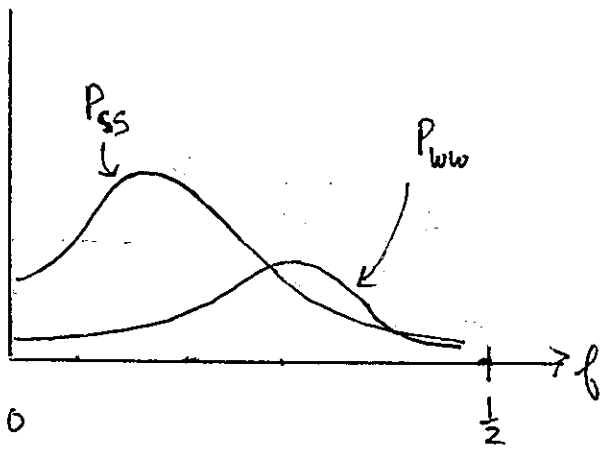
$$P_{xs}(f) =$$

$$P_{xx}(f) =$$

and therefore

$$\boxed{H(f) =}$$

Interpretation:



Infinite Wiener Filter

Now let's try to estimate $s[n]$ based on data from the present and infinite past.

$$\hat{s}[n] = \sum_{k=0}^{\infty} h[k] x[n-k]$$

As before, we may apply the orthogonality principle to arrive at the Wiener-Hopf equations:

$$\sum_{k=0}^{\infty} h[k] r_{xx}[l-k] = r_{xs}[l] \quad \forall l \geq 0$$

However, since these equations only hold for $l \geq 0$ (as opposed to all $l \in \mathbb{Z}$), it is not true that $h * r_{xx} = r_{xs}$.

Therefore, we can not solve for $h[k]$ or $H(f)$ by simply taking the DTFT.

It can be shown that

$$H(z) = \frac{1}{G(z)} \cdot \left[\frac{P_{xx}(z)}{G(z^{-1})} \right]_+$$

where

$$\bullet \quad P_{xx}(z) := \sum_{k=-\infty}^{\infty} r_{xx}[k] z^{-k}$$

$$= G(z) G(z^{-1})$$

spectral
factorization

↑ minimum phase, causal

$$\bullet \quad [Y(z)]_+ := \sum_{k=0}^{\infty} y[k] z^{-k},$$

the z -transform of the causal part

$$\text{of } \{y[k]\}_{k=-\infty}^{\infty}$$

These IIR Wiener estimators are not just theoretical curiosities. For large p , \hat{h}_p is well-approximated by its IIR counterpart. Therefore, when sufficient data are available, it is convenient to use the IIR estimators which have convenient frequency domain implementations.

Summary

- Wiener filtering \Leftrightarrow LMMSE recovery of signal in additive noise, assuming WSS
- Three basic problems: prediction, filtering, smoothing
- Generalized Levinson-Durbin \Rightarrow efficient algorithm for updating FIR filter for streaming apps.
- IIR estimators obtained by orthogonality principle and frequency/ z -transform domain techniques.

Key

$$a. \quad \underline{h}_1 = \frac{r_{xs}[0]}{r_{xx}[0]}, \quad \underline{g}_1 = \frac{r_{xx}[1]}{r_{xx}[0]}$$

b.

$$P_{ss}(f) = \text{DTFT} \{ r_{xs} \} = \text{DTFT} \{ r_{ss} \} = P_{ss}(f)$$

$$P_{xx}(f) = P_{ss}(f) + P_{ww}(f)$$

$$H(f) = \frac{P_{ss}(f)}{P_{ss}(f) + P_{ww}(f)}$$