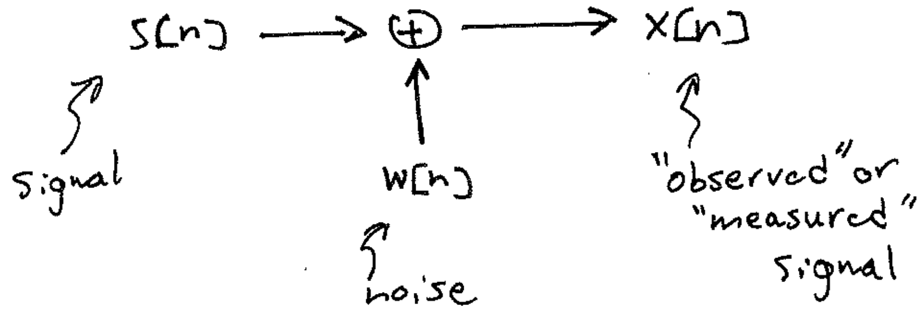
Optimal Filter Design from a Statistical Viewpoint

Basic Problem:



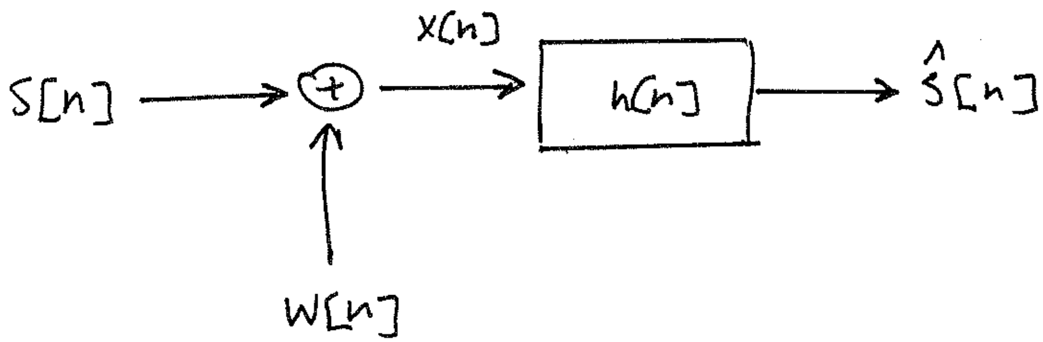
Can we recover s[n] from x[n]?

Applications:

- Comm systems
- control systems
- geophysics
- image processing
- speech enhancement

Goal:

Design a linear time-invariant filter to "reduce" the noise or (equivalently) "estimate" the signal.



We want to design $h[n]$ so that $\hat{s}[n]$ is as close as possible to $s[n]$.

Partial Knowledge:

If we know nothing at all about the signal or noise, then we have no basis on which to begin.

On the other hand, if we know the signal or noise exactly, then the problem is trivial.

Ex. If we know $w[n]$ (for example, if we could somehow measure it alone), then we have

$$s[n] = x[n] - w[n]$$

A reasonable and workable compromise between the two extremes above is to assume some partial knowledge of the signal and/or noise characteristics.

Here, we will model the signal and noise as realizations of stationary random processes. Moreover, we will assume:

1. $s[n]$ and $w[n]$ are statistically independent

2. the autocorrelation functions $R_{ss}[n]$ and $R_{ww}[n]$ are known

3. both $s[n]$ and $w[n]$ are zero-mean processes

<u>Remark 1</u>: It is not unreasonable to suppose that the signal and noise arise from separate and unrelated physical mechanisms. This supports assumption 1.

<u>Remark 2:</u> Knowledge of the autocorrelation functions is equ valent to knowing the power spectral densities:

$$S_{ss}(\omega) \overset{DTFT}{\longleftrightarrow} R_{ss}[n]$$

Thus, assumption 2 is essentially saying that we know how the signal or noise energy is distributed in frequency (on average).

<u>Remark 3:</u> Knowledge of the autocorrelation functions of power spectra may be based on a priori physical models or derived from previous measurements of similar signals.

<u>Ex.</u> Suppose we observe many realizations of the noise, say $w_1[n], \ldots, w_M[n]$. Then

$$R_{ww}[k] \approx \frac{1}{M} \sum_{i=1}^{M} w_i[n] w_i[n+k]$$

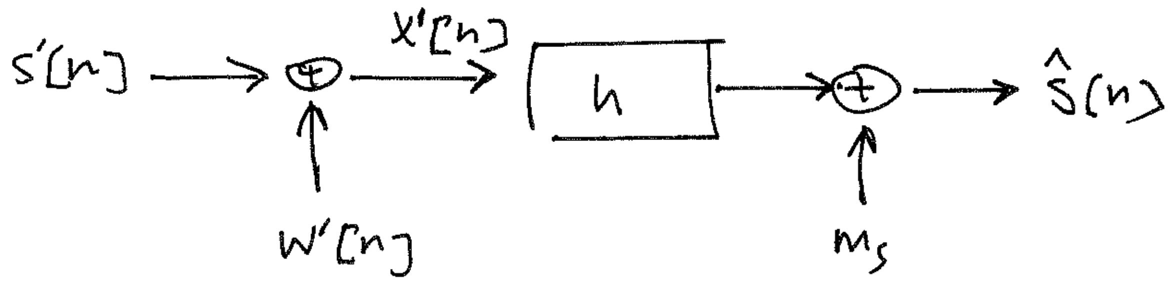
Remark 4: If the signal and/or noise are not zero-mean, but have known mean values, the we can simply redefine them:

$$s'[n] = s[n] - m_s$$

$$w'[n] = w[n] - m_w$$

zero-mean signals

Then, with $x'[n] = x[n] - m_s - m_w$, we have



So, as long as the means are known, without loss of generality we may assume that $s[n]$ and $w[n]$ are zero-mean.

Optimality Criterion:

Based on the partial knowledge we have assumed, we are now in a position to quantify the performance of a given filter $h[n]$. The partial knowledge is based on the "average" signal and noise characteristics, so it is natural to measure the average or expected error of the filter.

$$\left.\begin{aligned} e[n] &= s[n] - \hat{s}[n] \\ &= s[n] - h[n] * x[n] \end{aligned}\right\} \text{error}$$

Mean-square error:

$E\left[e^2[n]\right]$, average square error

The MSE is a function of $h[n]$:

$$MSE(h) = E\left[\left(s[n] - h[n] * x[n]\right)^2\right]$$

Minimizing the MSE:

The optimum lineare filter, in the sense of minimum mean-square error (MMSE), is called the Wiener filter.

Nobert Wiener

Born: 1894 in Columbia, Missouri

Received: Ph.D. math from Harvard at age 18.

with MIT, 1919–

Died: 1964

Let's look more closely at the MSE.

$$MSE(h) = E\left[\left(s[n] - h[n] * x[n]\right)^2\right]$$

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

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

Note that MSE(h) is quadratic in each $h[m]$. Therefore, MSE(h) has a unique minimum at

$$0 = \frac{\partial}{\partial h[m]} E\left[s^2[n] - 2\sum_{k=-\infty}^{\infty} h[k]\, x[n-k]\, s[n] + \left(\sum_{k=-\infty}^{\infty} h[k]\, x[n-k]\right)^2\right]$$

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

$$= E\Big[\, 0 \;-\; 2\, x[n-m]\, s[n]$$

$$+\; 2\left(\sum_{k=-\infty}^{\infty} h[k]\, x[n-k]\right)\cdot x[n-m] \,\Big]$$

$$= -2R_{xs}[m] + 2\sum_{k=-\infty}^{\infty} h[k]\, R_{xx}[m-k]$$

or

$$\boxed{\sum_{k=-\infty}^{\infty} h[k]\, R_{xx}[m-k] \;=\; R_{xs}[m]}$$

The MMSE-optimal filter satisfies the above system of equations known as the <u>Wiener-Hopf Equation</u>.

Note that assumptions 1 and 2 imply

$$R_{xx}[\ell] = R_{ss}[\ell] + R_{ww}[\ell]$$

$$R_{xs}[\ell] = R_{ss}[\ell]$$

are known sequences.

Resulting MMSE:

$h[n]$ satisfies

$$\sum_{k=-\infty}^{\infty} h[k]\, R_{xx}[m-k] = R_{xs}[m]$$

$$MSE(h) = E\Big[s^2[n] - 2\sum_{k=-\infty}^{\infty} h[k]\, x[n-k]\, s[n] + \sum_{k=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty} h[k]\, h[\ell]$$

$$= R_{ss}[0] - 2\sum_{k=-\infty}^{\infty} h[k]\, R_{xs}[k]$$

$$+ \underbrace{\sum_{k=-\infty}^{\infty}\sum_{\ell=-\infty}^{\infty} h[k]\, h[\ell]\, R_{xx}[k-\ell]}_{\sum_{k} h[k] \underbrace{\sum_{\ell} h[\ell]\, R_{xx}[k-\ell]}_{R_{xs}[k]}}$$

$$= R_{ss}[0] - \sum_{k=-\infty}^{\infty} h[k]\, R_{xs}[k]$$

Orthogonality Principle:

Intuitively, we expect that the Wiener filter (mmse filter) "extracts" the maximal amount of signal information from the noisy observation.

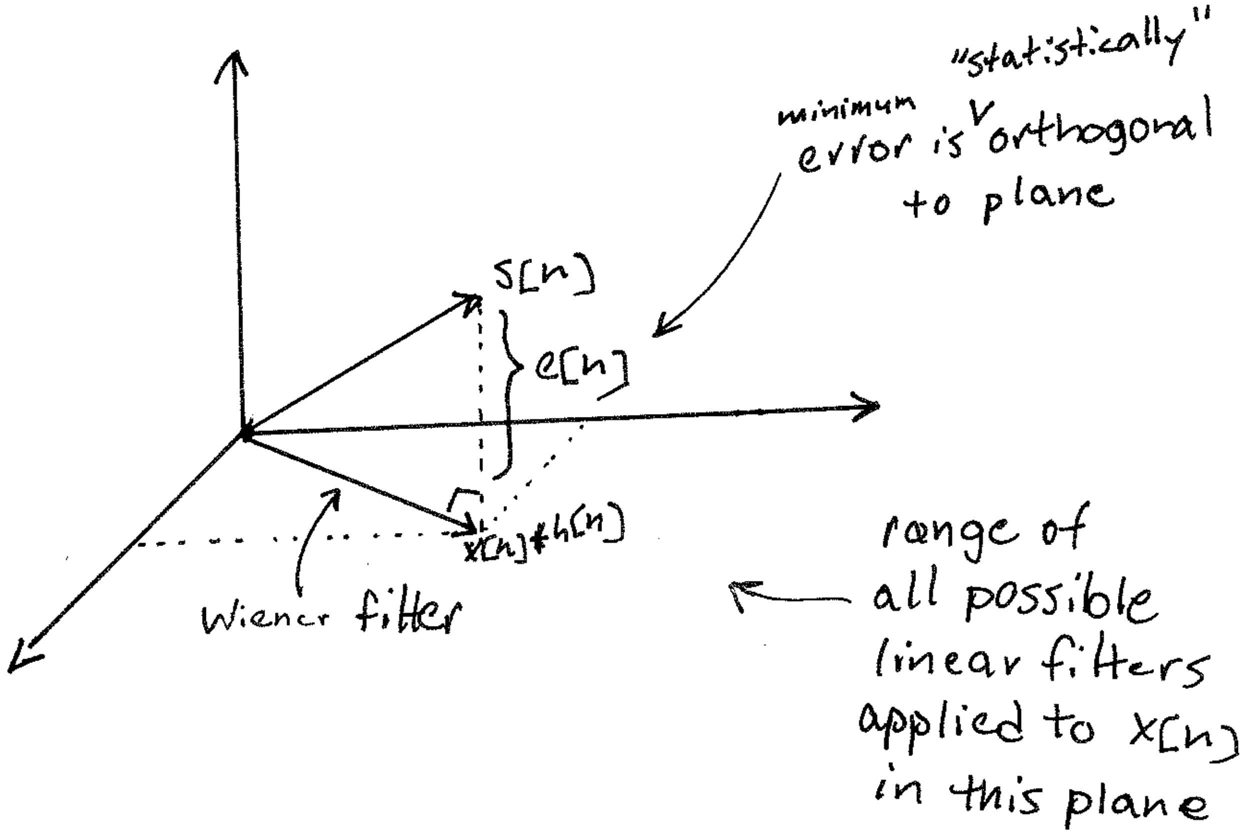
This intuition is supported by the so-called orthogonality principle, which states that the error of the Wiener filter, $s[n] - h[n] * x[n]$, is orthogonal to the measurement $x[n]$.

Indeed, for every $m \in \mathbb{Z}$ we have

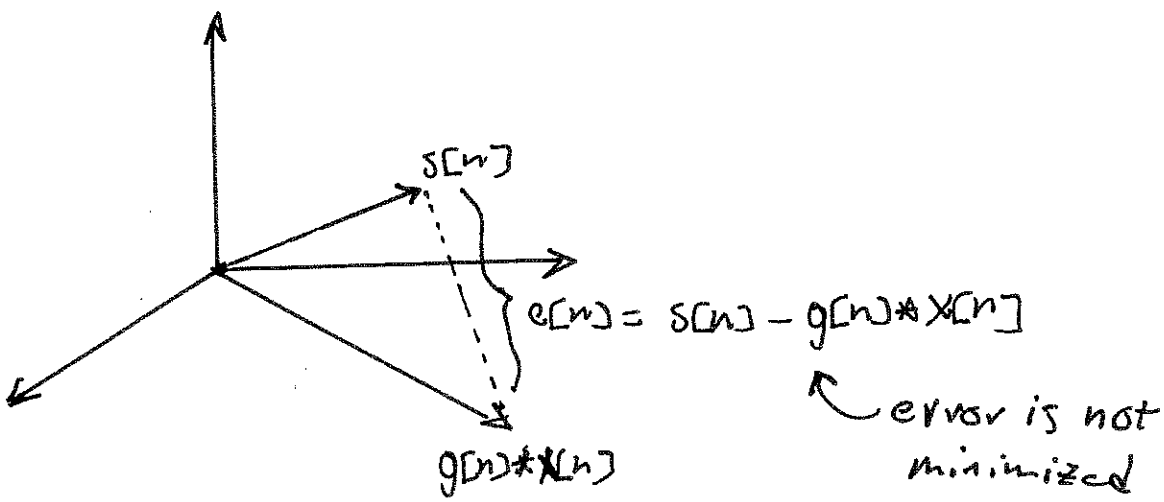
$$E\left[\left(s[n] - \sum_k h[k]\, x[n-k]\right) \cdot x[n-m]\right]$$

$$= R_{xs}[m] - \sum_k h[k]\, R_{xx}[m-k]$$

$$= 0$$



— suboptimal filter



Frequency Domain Interpretation

MMSE/Wiener filter

$$\sum_{k=-\infty}^{\infty} h[k]\, R_{XX}[m-k] = R_{XS}[m]$$

or equivalently

$$\sum_{k=-\infty}^{\infty} h[k] \left(R_{SS}[m-k] + R_{WW}[m-k] \right) = R_{SS}[m]$$

convolution

Take DTFT of both sides:

$$H(\omega)\left(S_{SS}(\omega) + S_{WW}(\omega) \right) = S_{SS}(\omega)$$

$$\Rightarrow \quad \boxed{H(\omega) = \frac{S_{SS}(\omega)}{S_{SS}(\omega) + S_{WW}(\omega)}}$$

$$H(\omega) = \frac{\text{signal power @ } \omega}{\text{signal + noise power @ } \omega}$$

$$S_{ss}(\omega) \gg S_{ww}(\omega)$$

$$\Rightarrow \quad H(\omega) \approx 1$$

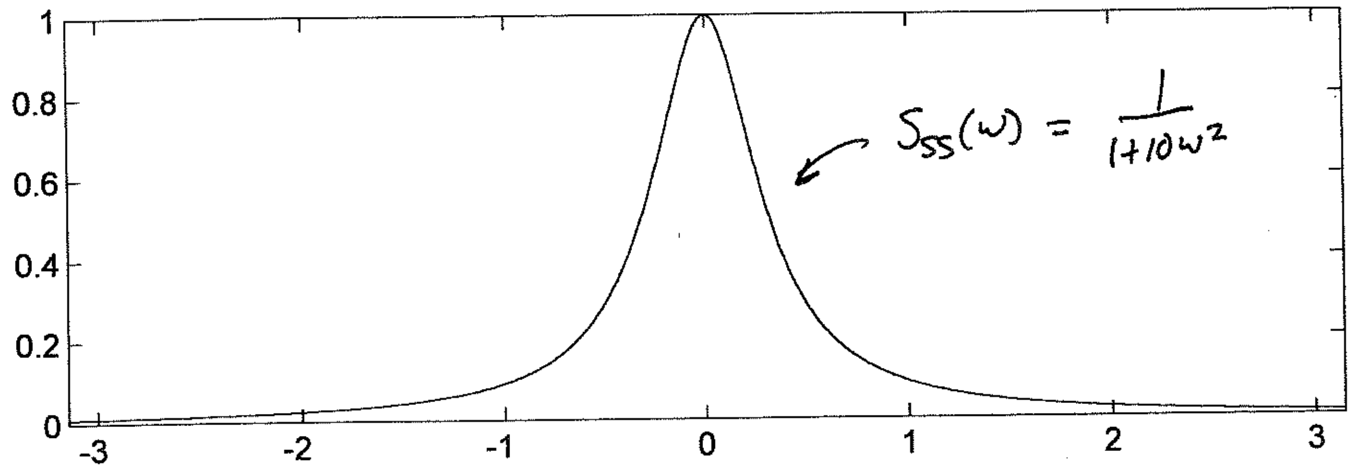
$$S_{ss}(\omega) \ll S_{ww}(\omega)$$

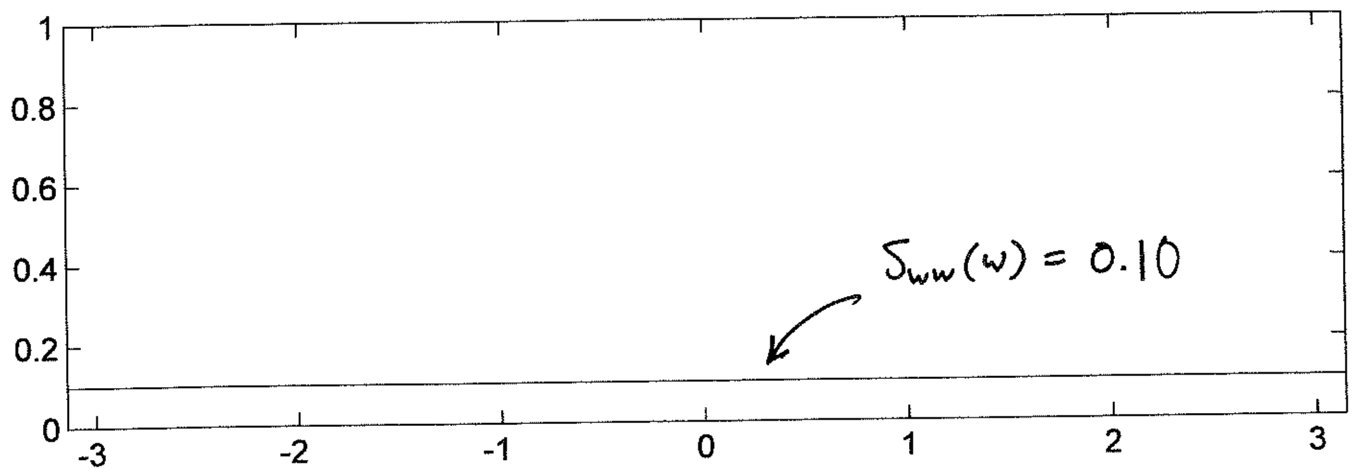
$$\Rightarrow \quad H(\omega) \approx 0$$

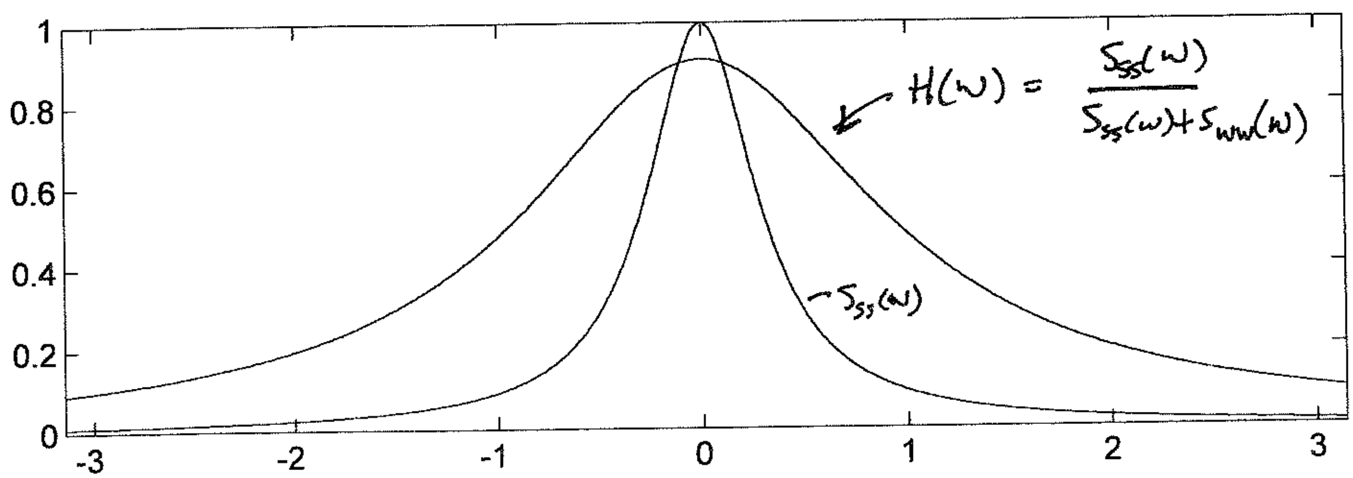
If the signal is strong (relative to noise) at frequency ω, then keep that frequency component...

Otherwise attenuate that component.

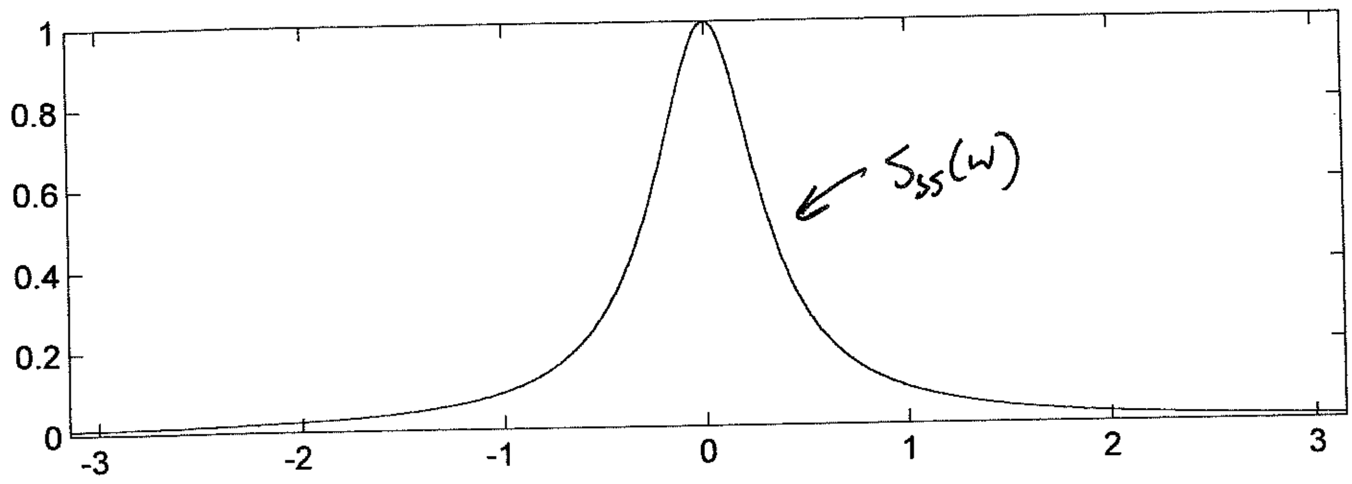
Ex.

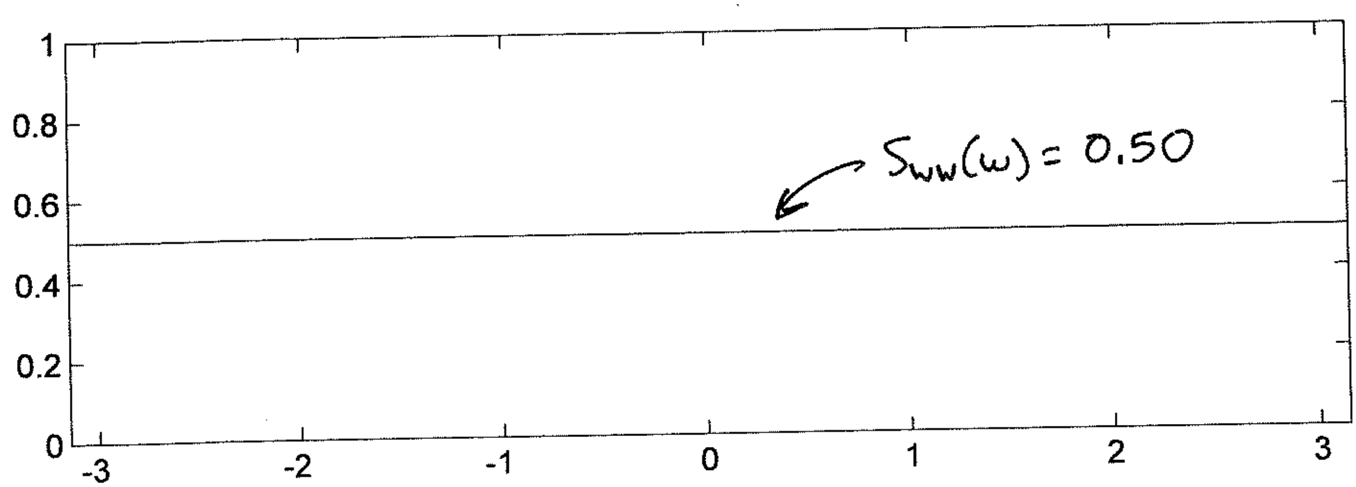

$S_{ss}(\omega) = \frac{1}{1+10\omega^2}$

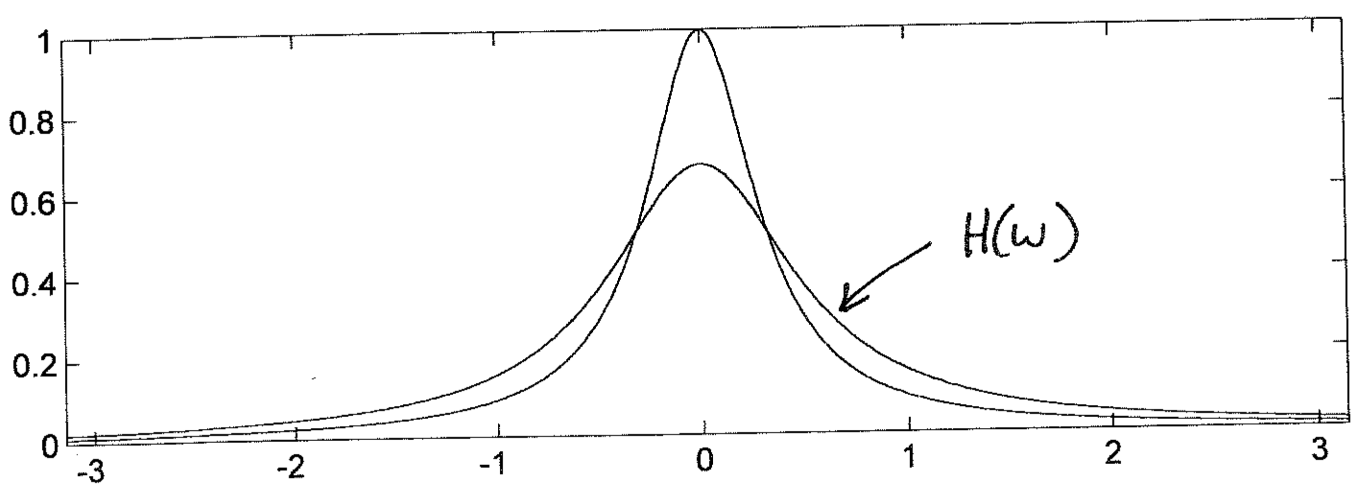

$S_{ww}(\omega) = 0.10$


$H(\omega) = \frac{S_{ss}(\omega)}{S_{ss}(\omega)+S_{ww}(\omega)}$
$S_{ss}(\omega)$

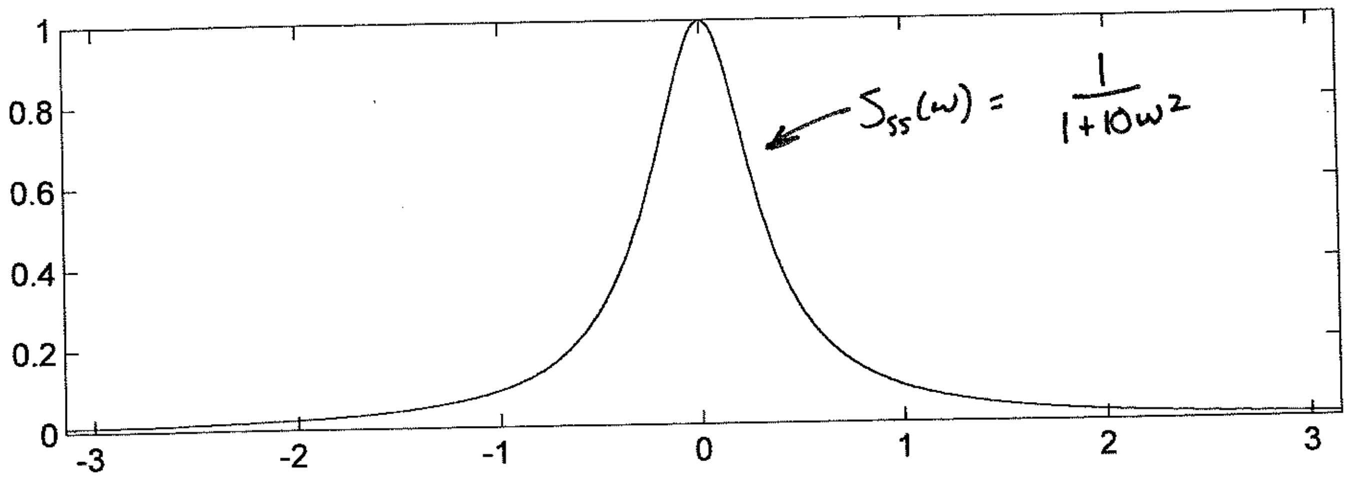
Ex.

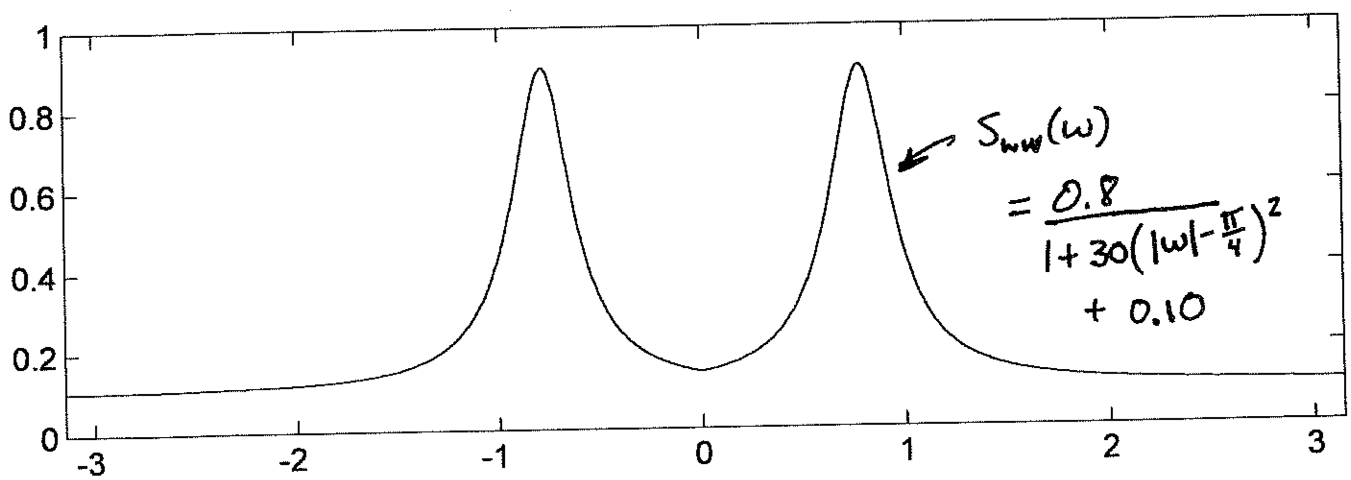


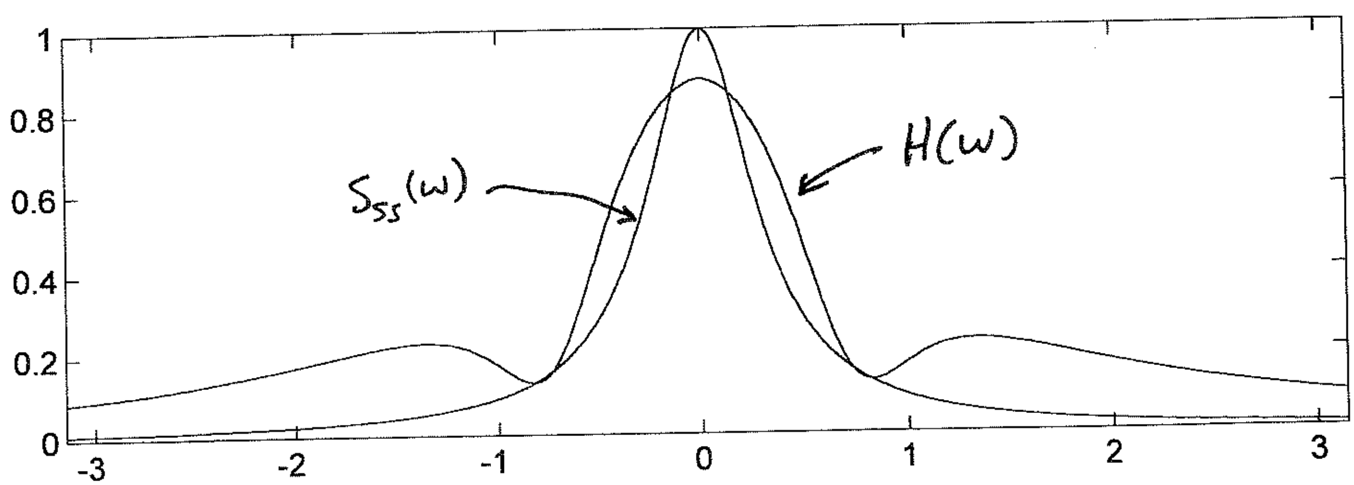




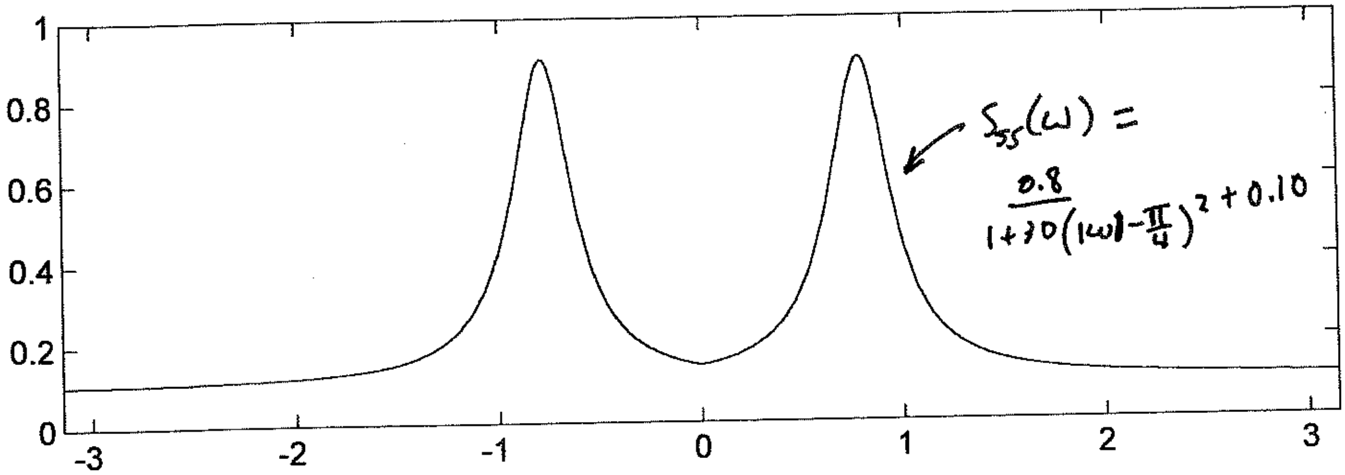
EX.

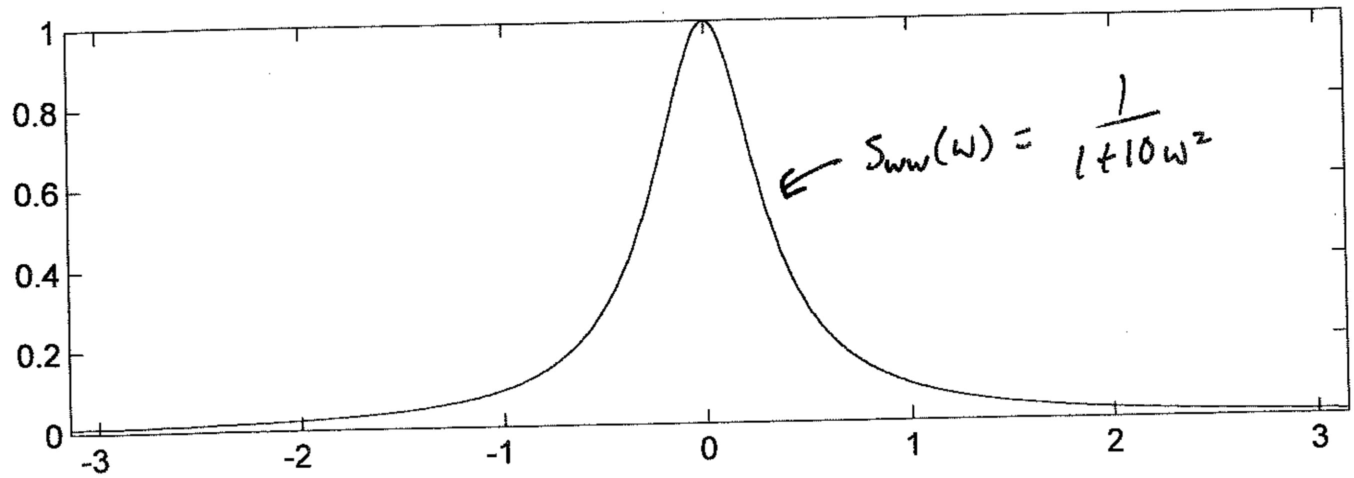

$S_{ss}(\omega) = \frac{1}{1+10\omega^2}$
1
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3

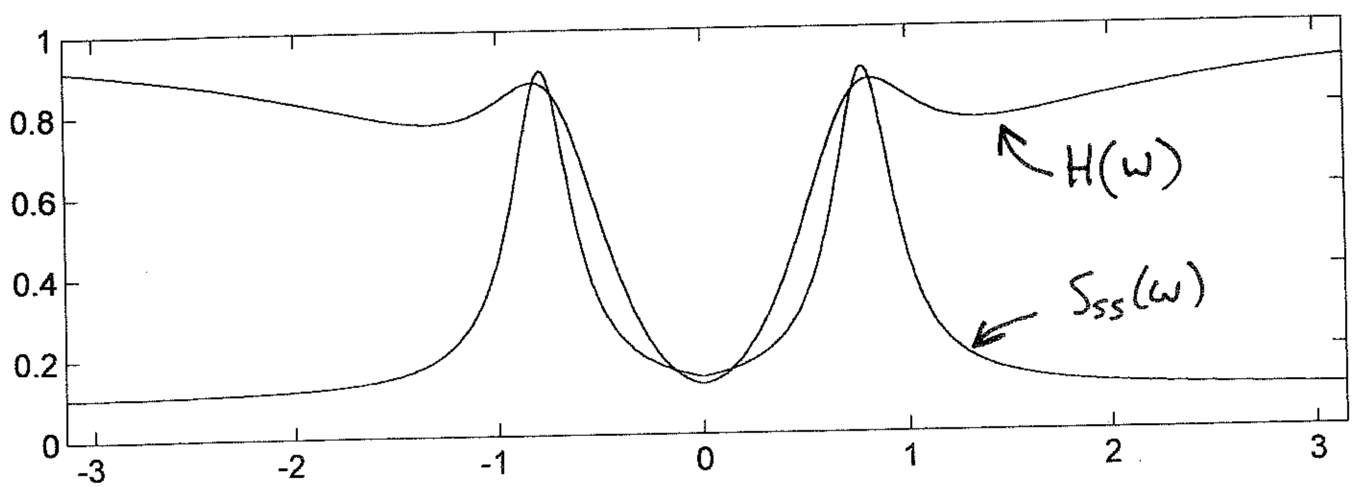

$S_{ww}(\omega)$
$= \frac{0.8}{1+30(|\omega|-\frac{\pi}{4})^2}$
$+ 0.10$
1
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3


$S_{ss}(\omega)$
$H(\omega)$
1
0.8
0.6
0.4
0.2
0
-3
-2
-1
0
1
2
3

Ex. "One person's signal is another's noise"







FIR Wiener Filters

In general, the optimum Wiener filter is IIR. That is, its time support is unlimited. This may mean that the filter is not easily realizable. Instead, let's try to find the best FIR filter; a sort of constrained Wiener filter.

We again choose MSE as our optimality criterion, but force the filter $h[n]$ to be a length M FIR filter.

$$\hat{s}[n] = \sum_{k=0}^{M-1} h[k]\, x[n-k]$$

$$MSE(h) = E\left[\left(s[n] - \sum_{k=0}^{M-1} h[k]\, x[n-k]\right)^2\right]$$

Again, the MSE is a quadratic function of $h[n]$ and the optimal filter satisfies the Wiener-Hopf equation:

$$\sum_{k=0}^{M-1} h[k]\, R_{xx}[m-k] = R_{xs}[m]$$

These equations can be written in matrix form as

$$\underset{(M\times M)}{\underline{R}_{xx}}\ \underset{(M\times 1)}{\underline{h}} = \underset{(M\times 1)}{\underline{R}_{xs}}$$

with

$$\underline{R}_{xx} = \begin{bmatrix} R_{xx}[0] & R_{xx}[1] & \cdots & R_{xx}[M-1] \\ R_{xx}[1] & \ddots & & \vdots \\ \vdots & & \ddots & R_{xx}[1] \\ R_{xx}[M-1] & \cdots & R_{xx}[1] & R_{xx}[0] \end{bmatrix}$$

where I used the fact that
$R_{xx}[k] = R_{xx}[-k]$ for
stationary processes

$$\underline{h} = \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[M-1] \end{bmatrix} \qquad \underline{R}_{xs} = \begin{bmatrix} R_{xs}[0] \\ R_{x\cdot}[1] \\ \vdots \\ R_{xs}[M-1] \end{bmatrix}$$

Thus, the optimal Wiener filter is given by

$$\boxed{\underline{h} = \underline{R}_{xx}^{-1} \underline{R}_{xs}}$$

Orthogonality Revisited:

Suppose $M = 2$