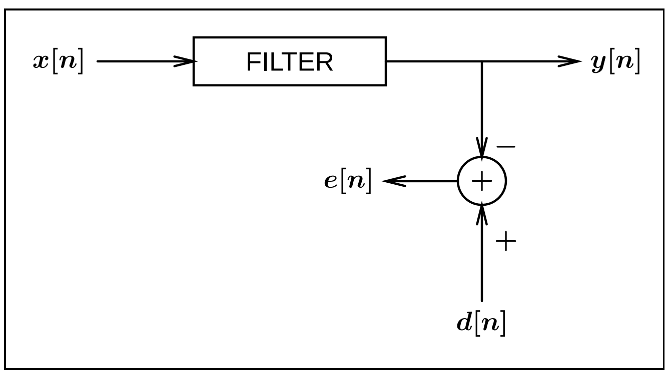
Wiener Filtering

In this lecture we will take a different view of filtering. Previously, we have depended on *frequency-domain* specifications to make some sort of LP/ BP/ HP/ BS filter, which would extract the desired information from an input signal.

Now, we wish to filter a signal $x[n]$ to modify it such that it approximates some other signal $d[n]$ in some statistical sense. That is, the output of the filter $y[n]$ is a good estimate of $d[n]$. The output error $e[n]$ represents the mismatch between $y[n]$ and $d[n]$. This can be considered a *time-domain* specification of the filter.



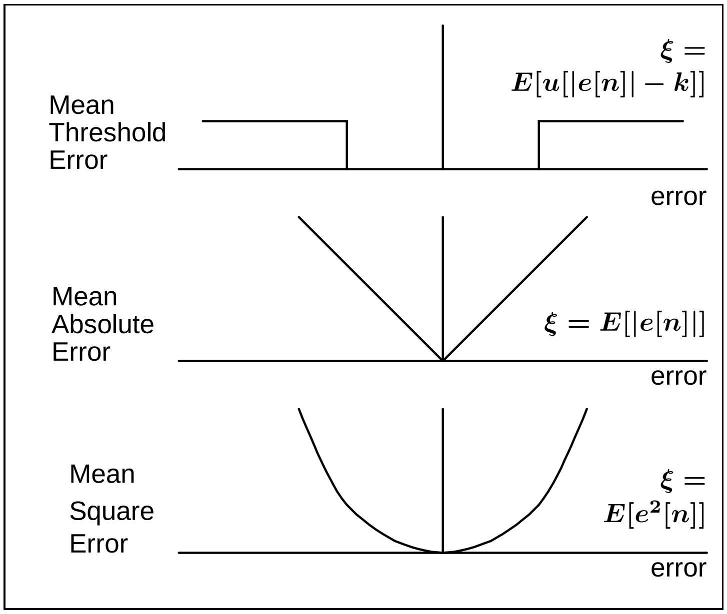
Why would we want to do this? Well, consider the problem of predicting the stock market. Assume that $x[n]$ represents the price of a particular stock on day n and $d[n]$ is the price of that stock one day in the future. The goal is to find a filter that will predict $d[n]$ given $x[n]$. Moreover, we would like to find the *best* such linear predictor.

Q: Is there an obvious solution for this case?

A: $\boxed{\text{Sure, pick } H(z) = z\text{, but this is worthless for this real-time example!}}$

In order to talk about an "optimal" filter which estimates $d[n]$ from $x[n]$, we must have a method of measuring how good a job the filter does. A "cost function" is used to judge the performance, and could take on many different forms.

For example,



We most commonly use the mean square error (MSE) as our cost function.

$$\xi = E[e^2[n]]$$

where $E[\cdot]$ represents statistical expectation, and

$$E[e^2[n]] = \int_{-\infty}^{\infty} x^2 p_e(x)\, dx$$

where $p_e(x)$ is the probability density function of the error. The filter that is optimum in the MSE sense is called a Wiener filter.
In all our analyses here we will assume:

1. $x[n]$ is *wide-sense stationary*, i.e. it has constant (and finite) mean and variance, and a correlation function which is a function of time shift only:

$$E[x[m_1] \cdot x[m_1 + n]] = E[x[m_2] \cdot x[m_2 + n]] \quad \forall\, m_1, m_2$$

Incidentally, a stronger condition would be to assume that $x[n]$ is *strong-sense stationary*, i.e. all of its moments are constant and finite. An even stronger condition would be to assume that $x[n]$ is *ergodic*, i.e. all of its sample averages converge to the true moments. Fortunately, these latter two conditions are not necessary for the present development.

2. All of the signals are zero-mean.

3. We use MSE as our error criterion.

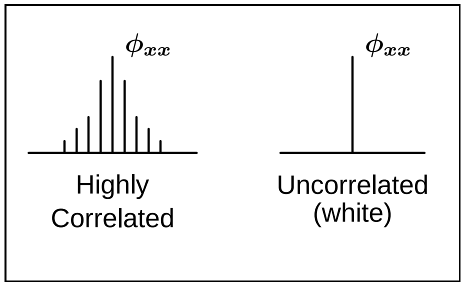
Correlation

The definition of correlation between two random processes $x[n]$ and $y[n]$ (when x and y are wide sense stationary) is:

$$\phi_{xy}[k] = E[x[n]y[n+k]]$$

It is a measure of the similarity between two random variables, and is related to the amount of information that one random variable contains about another.

Example:



Two-Sided (Unconstrained) Wiener Filters

$$\begin{aligned}\xi &= E[e^2[n]] \\ &= E[(d[n]-y[n])^2] \\ &= E[d^2[n]] - 2E[y[n]d[n]] + E[y^2[n]] \\ &= \phi_{dd}[0] - 2\phi_{yd}[0] + \phi_{yy}[0]\end{aligned}$$

So, the cost function is itself a function of the correlation and cross correlation of the output of the filter and the desired output. We know that:

$$y[n] = \sum_{m=-\infty}^{\infty} h[m]x[n-m]$$

Observe that, in general, $h[m]$ represents a non-causal filter with infinite impulse response (IIR). (Later, we consider the special case where $h[m]$ is either a causal IIR filter or a causal FIR filter.) Substituting the definition for $y[n]$ into the previous result, we get:

$$\begin{aligned}\phi_{yd}[0] &= E[y[n]d[n]] \\ &= E\left[\sum_{m=-\infty}^{\infty} h[m]x[n-m]d[n]\right] \\ &= \sum_{m=-\infty}^{\infty} h[m]\phi_{xd}[m] \\ \phi_{yy}[0] &= E[y[n]y[n]] \\ &= E\left[\sum_{m=-\infty}^{\infty} h[m]x[n-m]\sum_{l=-\infty}^{\infty} h[l]x[n-l]\right] \\ &= \sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} h[m]h[l]\phi_{xx}[l-m] \\ \xi &= \phi_{dd}[0] - 2\left[\sum_{m=-\infty}^{\infty} h[m]\phi_{xd}[m]\right] + \left[\sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} h[m]h[l]\phi_{xx}[l-m]\right]\end{aligned}$$

Our goal is to find $h[n]$ which minimizes ξ. The solution is to take the derivative $\frac{d\,\xi}{d\,h[n_i]}$ and set it to zero:

$$\frac{d\,\xi}{d\,h[n_i]} = 0 - 2\phi_{xd}[n_i] + 2\sum_{m=-\infty}^{\infty} h_{\text{opt}}[m]\phi_{xx}[n_i-m] = 0$$

$$\sum_{m=-\infty}^{\infty} h_{\text{opt}}[m]\phi_{xx}[n_i - m] \;=\; \phi_{xd}[n_i]$$

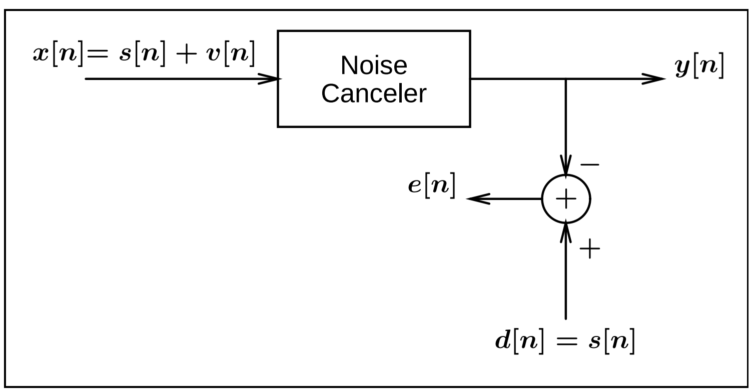
which must be satisfied for all n_i.
This equation specifies the impulse response of the optimal filter. Observe that it depends on the cross correlation between $x[n]$ and $d[n]$ and the autocorrelation function of $x[n]$.
The optimal filter is known as the Wiener solution. Observing that the left hand side of the equation is equal to the convolution of $h_{\text{opt}}[n]$ and $\phi_{xx}[n]$,

$$\begin{aligned} H_{\text{opt}}(z)\Phi_{xx}(z) &= \Phi_{xd}(z) \\ H_{\text{opt}}(z) &= \Phi_{xd}(z)\Phi_{xx}^{-1}(z) \end{aligned}$$

Note:

- The filter obtained may be non-causal!
- $\xi_{\min} = \phi_{dd}[0] - \sum_{n=-\infty}^{\infty} h_{\text{opt}}[n]\phi_{xd}[n] \neq 0$ in general.

Example: Noise Canceling:



Assuming that $s[n]$ and $v[n]$ are uncorrelated, zero mean:

$$\begin{aligned} E[x[n]x[n+k]] &= E[(s[n]+v[n])(s[n+k]+v[n+k])] \\ &= \phi_{ss}[k] + \phi_{vv}[k] \\ \Phi_{xx}(z) &= \Phi_{ss}(z) + \Phi_{vv}(z) \\ E[x[n]d[n+k]] &= E[(s[n]+v[k])s[n+k]] \\ &= \phi_{ss}[k] \\ \Phi_{xd}(z) &= \Phi_{ss}(z) \\ \text{Thus,}\quad H_{\text{opt}}(z) &= \frac{\Phi_{ss}(z)}{\Phi_{ss}(z)+\Phi_{vv}(z)} \end{aligned}$$

We notice two properties from this solution:

1. At “frequencies” where $\Phi_{vv}(z) \to 0$ (i.e. the noise is zero) there is no need to filter the signal and $H_{\text{opt}}(z) \to 1$.
2. At “frequencies” where $\Phi_{vv}(z) \to \infty$ (i.e. where the noise dominates) the filter is “turned off” and $H_{\text{opt}}(z) \to 0$.

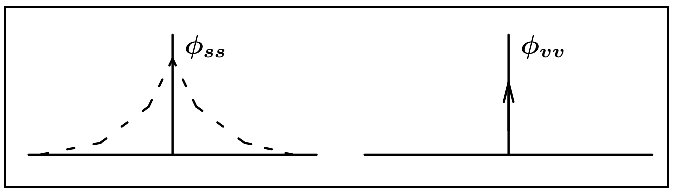
Numerical Example: Assume that the autocorrelation function of the input signal is:

$$\phi_{ss}[k] = \frac{10}{27}\left(\frac{1}{2}\right)^{|k|}$$

and that the autocorrelation of the (white) noise source is:

$$\phi_{vv}[k] = \frac{2}{3}\delta[k]$$

The signal and noise are assumed to be uncorrelated with each other.



To find the Wiener filter, we need to obtain $\Phi_{xd}(z)$ and $\Phi_{xx}(z)$. We can first find $\phi_{xd}[k]$ and $\phi_{xx}[k]$, and then z-transform them. Since the signal and noise are uncorrelated with each other,

$$\begin{aligned}\phi_{xx}[k] &= \phi_{ss}[k] + \phi_{vv}[k]\\ &= \frac{10}{27}\left(\frac{1}{2}\right)^{|k|} + \frac{2}{3}\delta[k],\\ \Phi_{xx}(z) &= \Phi_{ss}(z) + \Phi_{vv}(z).\end{aligned}$$

The z-transform of the noise autocorrelation is

$$\Phi_{vv}(z) = \frac{2}{3}.$$

The z-transform of the signal autocorrelation (after some math) is

$$\Phi_{ss}(z) = \frac{5/18}{(1-\frac{1}{2}z^{-1})(1-\frac{1}{2}z)}.$$

Combining these:

$$\begin{aligned}\Phi_{xx}(z) &= \frac{5/18}{(1-\frac{1}{2}z^{-1})(1-\frac{1}{2}z)} + \frac{2}{3}\\ &= \frac{20-6z-6z^{-1}}{18(1-\frac{1}{2}z^{-1})(1-\frac{1}{2}z)}.\end{aligned}$$

We saw earlier,

$$\begin{aligned}\phi_{xd}[k] &= \phi_{ss}[k],\\ \Phi_{xd}(z) &= \Phi_{ss}(z).\end{aligned}$$

Combining these results:

$$H_{\text{opt}}(z) = \frac{5/18}{(1-\frac{1}{3}z^{-1})(1-\frac{1}{3}z)}$$

$$h_{\text{opt}}[n] = \frac{5}{16}\left(\frac{1}{3}\right)^{|n|}, \quad \forall\, n$$

Q: How did we pick this $h_{\text{opt}}[n]$ without knowing the ROC of $H_{\text{opt}}(z)$?
A: $\boxed{\text{We need a stable noise canceller.}}$

Clearly this resulting filter is non-causal. We have two recourses:

1. Limit the filter to be a causal IIR filter.

2. Compute the optimal FIR filter instead.

Causal (Shannon-Bode) Wiener Filters

Our interest now focuses on the realization of *causal* Wiener filters, whose impulse responses are constrained to be zero for negative time. The optimal causal impulse response has zero response for negative time and has zero derivatives of ξ with respect to impulse response for all times equal to and greater than zero. The causal Wiener equation becomes:

$$\sum_{m=0}^{\infty} h_{\text{opt causal}}[m]\phi_{xx}[n-m] \;=\; \phi_{xd}[n], \quad n \geq 0$$
$$h_{\text{opt causal}}[n] \;=\; 0, \quad n < 0.$$

This is not a simple convolution like the unconstrained Wiener equation, and special methods will be needed to find useful solutions. The approach developed by Shannon and Bode will be used. We begin with a simple case. Let the filter input be white with zero-mean and unit variance, so that

$$\phi_{xx}[n] = \delta[n].$$

For this input, the causal Wiener equations become

$$\sum_{m=0}^{\infty} h_{\text{opt causal}}[m]\phi_{xx}[n-m] = h_{\text{opt causal}}[n] \;=\; \phi_{xd}[n], \quad n \geq 0$$
$$h_{\text{opt causal}}[n] \;=\; 0, \quad n < 0$$

With the same white input, but without the causality constraint, the Wiener equation would be

$$\sum_{m=-\infty}^{\infty} h_{\text{opt}}[m]\phi_{xx}[n-m] = h_{\text{opt}}[n] = \phi_{xd}[n], \quad \forall n.$$

From this we may conclude that when the input to the Wiener filter is white, the optimal solution with a causality constraint is the same as the optimal solution without the constraint, except that with the causality constraint the impulse response is set to zero for negative time. With a white input, the causal solution is easy to obtain, and it is key to Shannon-Bode. You find the unconstrained two-sided Wiener solution and remove the non-causal part in the time domain. *You don't do this if the input is not white.*

Usually, the input to the Wiener filter is not white. Therefore the first step in the Shannon-Bode realization is to whiten the input signal. A whitening filter can always be designed to do this, using a priori knowledge of the input autocorrelation function or its z-transform.

Assume that the z-transform of the autocorrelation function is a rational function of z that can be written as the ratio of a numerator polynomial in z to a denominator polynomial in z. Factor both the numerator and denominator polynomials. The autocorrelation function is symmetric, and its z-transform has the following symmetry:

$$\phi_{xx}[n] \;=\; \phi_{xx}[-n],$$
$$\Phi_{xx}(z) \;=\; \Phi_{xx}(z^{-1}).$$

Accordingly, there must be symmetry in the numerator and denominator factors:

$$\Phi_{xx}(z) = A\frac{(1-az^{-1})(1-az)(1-bz^{-1})(1-bz)\ldots}{(1-\alpha z^{-1})(1-\alpha z)(1-\beta z^{-1})(1-\beta z)\ldots}.$$

With no loss in generality,[1] assume that all the parameters $a,\ b,\ c,\ \ldots,\ \alpha,\ \beta,\ \ldots$ have magnitudes less than one. We can then factor $\Phi_{xx}(z)$ as

$$\Phi_{xx}(z) = \Phi_{xx}^{+}(z)\cdot\Phi_{xx}^{-}(z),$$

where

$$\begin{aligned}\Phi_{xx}^{+}(z) &= \sqrt{A}\frac{(1-az^{-1})(1-bz^{-1})\ldots}{(1-\alpha z^{-1})(1-\beta z^{-1})\ldots}\\ \Phi_{xx}^{-}(z) &= \sqrt{A}\frac{(1-az)(1-bz)\ldots}{(1-\alpha z)(1-\beta z)\ldots}.\end{aligned}$$

All poles and zeros of $\Phi_{xx}^{+}(z)$ will be inside the unit circle in the z-plane. All poles and zeros of $\Phi_{xx}^{-}(z)$ will be outside the unit circle in the z-plane. Furthermore,

$$\Phi_{xx}^{+}(z) = \Phi_{xx}^{-}(z^{-1}),\ \ \Phi_{xx}^{+}(z^{-1}) = \Phi_{xx}^{-}(z).$$

The whitening filter can now be designed. Let it have the transfer function

$$H_{\text{whitening}}(z) = \frac{1}{\Phi_{xx}^{+}(z)}.$$

To verify its whitening properties, let its input be $x[n]$ with autocorrelation function $\phi_{xx}[n]$, and let its output be $\tilde{x}[n]$ with autocorrelation function $\phi_{\tilde{x}\tilde{x}}[n]$. The transform of the output autocorrelation function is

$$\begin{aligned}\Phi_{\tilde{x}\tilde{x}}(z) &= H_{\text{whitening}}(z^{-1})\cdot H_{\text{whitening}}(z)\cdot\Phi_{xx}(z)\\ &= \frac{1}{\Phi_{xx}^{+}(z^{-1})}\cdot\frac{1}{\Phi_{xx}^{+}(z)}\cdot\Phi_{xx}(z)\\ &= \frac{1}{\Phi_{xx}^{-}(z)}\cdot\frac{1}{\Phi_{xx}^{+}(z)}\cdot\Phi_{xx}(z)\\ &= 1.\end{aligned}$$

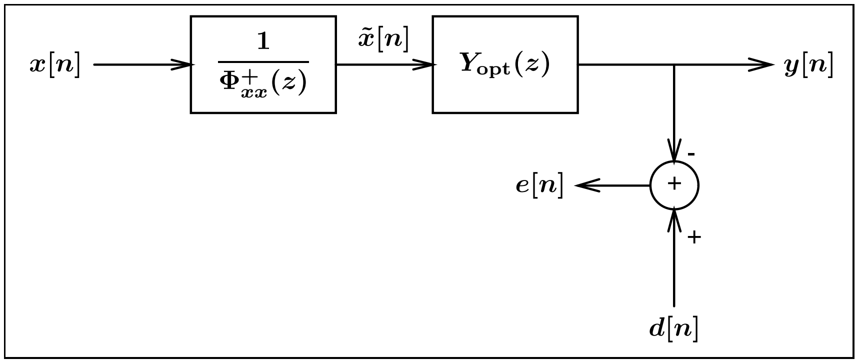
The autocorrelation function of the output is

$$\phi_{\tilde{x}\tilde{x}}[n] = \delta[n].$$

Therefore the output is white with unit mean square.

It is important to note that the whitening filter is both causal and stable, since the zeros of $\Phi_{xx}^{+}(z)$ are all inside the unit circle in the z-plane. Furthermore, using this whitening filter does nothing irreversible to the signal. The inverse of the whitening filter has a transfer function equal to $\Phi_{xx}^{+}(z)$, which is also stable and causal because $\Phi_{xx}^{+}(z)$ has all of its poles inside the unit circle. The whitening filter can be readily utilized since it is causal, stable, and invertible with a causal, stable filter.

[1] The only case not included involves zeros of $\Phi_{xx}(z)$ exactly on the unit circle. This is a special case and it requires special treatment. These zeros are assumed to be somewhere off the unit circle, and they are placed back on it by a limiting process.



The subsequent filter is easily designed, since its input is white. We represent its transfer function by $Y_{\text{opt causal}}(z)$ and can design it by first obtaining $Y_{\text{opt}}(z)$ disregarding causality, and then removing the non-causal part of its impulse response.

We wish to find $Y_{\text{opt}}(z)$ disregarding causality in terms of the autocorrelation of $x[n]$ and crosscorelation with $d[n]$. To do so, we could work out expressions for $\Phi_{\tilde{x}\tilde{x}}(z)$ and $\Phi_{\tilde{x}d}(z)$ in terms of $\Phi_{xx}(z)$ and $\Phi_{xd}(z)$ and then use the formula for the uncostrained Wiener filter. However, there is an easier way; notice that the cascading of $\frac{1}{\Phi^+_{xx}(z)}$ and $Y_{\text{opt}}(z)$ should be equal to the unconstrained Wiener filter given by $\frac{\Phi_{xd}(z)}{\Phi_{xx}(z)}$, hence

$$\begin{aligned} Y_{\text{opt}}(z) &= \frac{\Phi_{xd}(z)}{\Phi_{xx}(z)} \cdot \Phi^+_{xx}(z) \\ &= \frac{\Phi_{xd}(z)}{\Phi^-_{xx}(z)} \end{aligned}$$

To obtain $Y_{\text{opt causal}}(z)$, we first inverse transform $Y_{\text{opt}}(z)$ into the time domain, remove the non-causal part, and then z-transform the causal part obtained. This operation cannot be done with z-transforms alone. A special notation has been devised to represent taking the causal part:

$$Y_{\text{opt causal}}(z) = [Y_{\text{opt}}(z)]_+ = \left[\frac{\Phi_{xd}(z)}{\Phi^-_{xx}(z)}\right]_+ .$$

The Shannon-Bode realization of the causal Wiener filter can now be formulated:

$$H_{\text{opt causal}}(z) = \frac{1}{\Phi^+_{xx}(z)}\left[\frac{\Phi_{xd}(z)}{\Phi^-_{xx}(z)}\right]_+ .$$

Example:

An example helps to illustrate how this formula is used. We will rework the above noise filtering example, only in this case the Wiener filter will be designed to be causal.

The first step is to factor Φ_{xx}:

$$\Phi_{xx}(z) = \frac{(1-\frac{1}{3}z)(1-\frac{1}{3}z^{-1})}{(1-\frac{1}{2}z)(1-\frac{1}{2}z^{-1})}.$$

Therefore,

$$\Phi^+_{xx} = \frac{(1-\frac{1}{3}z^{-1})}{(1-\frac{1}{2}z^{-1})},\ \ \Phi^-_{xx} = \frac{(1-\frac{1}{3}z)}{(1-\frac{1}{2}z)}.$$

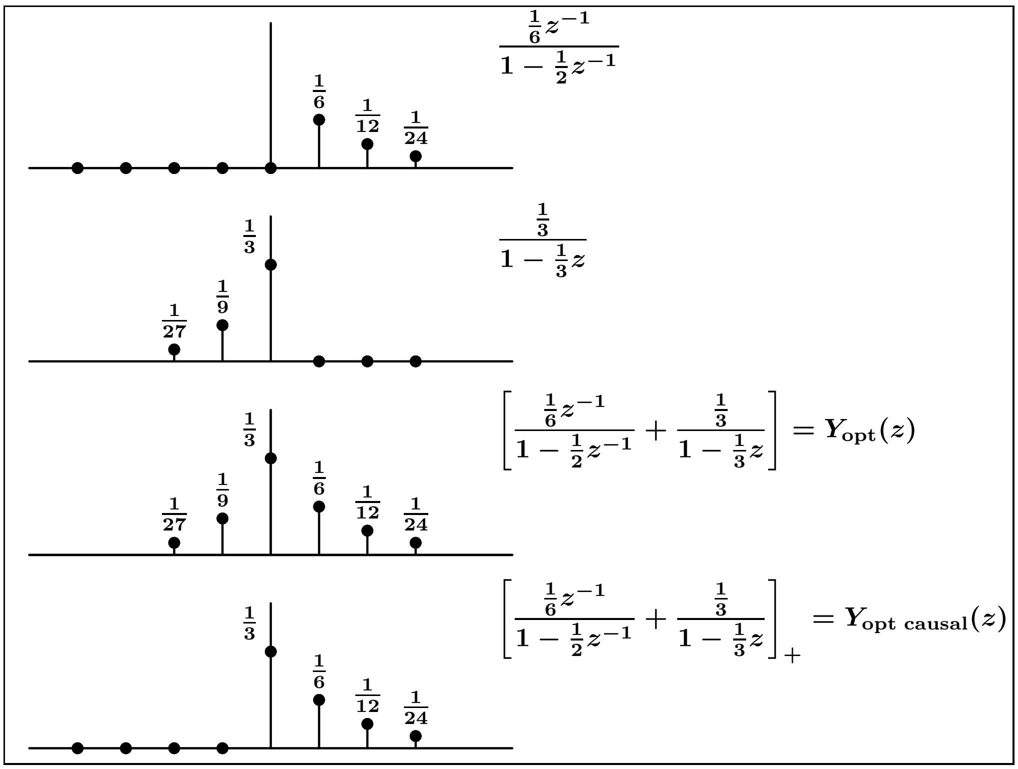
$\Phi_{xd}(z)$ was found before. Thus:

$$\Phi_{xd}(z) = \frac{5/18}{(1-\frac{1}{2}z^{-1})(1-\frac{1}{2}z)}.$$

$$
\begin{aligned}
Y_{\text{opt}}(z) = \frac{\Phi_{xd}(z)}{\Phi_{xx}^{-}(z)} &= \frac{5/18}{(1-\frac{1}{2}z^{-1})(1-\frac{1}{3}z)} \\
&= \left[\frac{\frac{1}{6}z^{-1}}{1-\frac{1}{2}z^{-1}} + \frac{\frac{1}{3}}{1-\frac{1}{3}z}\right].
\end{aligned}
$$

The filter $Y_{\text{opt}}(z)$ is generally two-sided in the time domain. In order for it to be stable, its two terms must correspond to time domain components as follows:

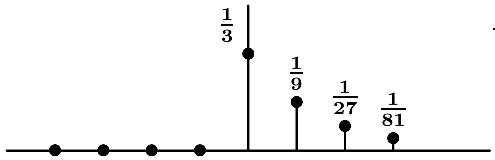
$$
\begin{aligned}
\frac{\frac{1}{6}z^{-1}}{1-\frac{1}{2}z^{-1}} &\rightarrow (\text{right-handed time function}). \\
\frac{\frac{1}{3}}{1-\frac{1}{3}z} &\rightarrow (\text{left-handed time function}).
\end{aligned}
$$



$$
\begin{aligned}
Y_{\text{opt causal}}(z) &= \frac{1}{3} + \frac{1}{6}z^{-1} + \frac{1}{12}z^{-2} + \ldots \\
&= \left(\frac{\frac{1}{3}}{1-\frac{1}{2}z^{-1}}\right).
\end{aligned}
$$

Including the whitening filter, the transfer function of the causal Wiener filter is

$$
\begin{aligned}
H_{\text{opt causal}}(z) &= \left(\frac{1-\frac{1}{2}z^{-1}}{1-\frac{1}{3}z^{-1}}\right)\left(\frac{\frac{1}{3}}{1-\frac{1}{2}z^{-1}}\right) \\
&= \frac{\frac{1}{3}}{1-\frac{1}{3}z^{-1}}. \\
h_{\text{opt causal}}[n] &= \frac{1}{3}\left(\frac{1}{3}\right)^n, \qquad n \geq 0.
\end{aligned}
$$



$$\frac{\frac{1}{3}z^{-1}}{1-\frac{1}{3}z^{-1}} = H_{\text{opt causal}}(z)$$

It is expected that the causal Wiener filter would not perform as well as the unconstrained non-causal Wiener filter. This is indeed the case, and can be verified with a little math.

$\xi_{\text{min causal}} = \phi_{dd}[0] - \sum_{n=0}^{\infty} h_{\text{opt causal}}[n]\phi_{xd}[n].$

Optimal FIR Wiener Filters

For an FIR realization, we make the following changes in the derivation:

$$y[n] = \sum_{k=0}^{N-1} h[k]x[n-k]$$

This time, we will solve the system using matrix notation:

Filter	$H = [h_0, h_1, \ldots, h_{N-1}]^T$
Input	$X = [x_n, x_{n-1}, \ldots, x_{n-N+1}]^T$
Output	$y[n] = H^T X = X^T H$ (scalar)
Error	$e[n] = d[n] - y[n] = d[n] - H^T X$

$$\begin{aligned}
\xi &= E[e^2[n]] \\
&= E[d^2[n]] - 2E[H^T X d[n]] + E[H^T X X^T H] \\
&= \underbrace{\phi_{dd}[0]}_{\sigma_d^2} - 2H^T \underbrace{E[Xd[n]]}_{P} + H^T \underbrace{E[XX^T]}_{R} H
\end{aligned}$$

Note: ξ is a quadratic function of the filter H. This implies (among other things) that there is a unique global optimum.

$$\begin{aligned}
P &= E[Xd[n]] \\
&= \begin{bmatrix} E[x[n]d[n]] \\ \vdots \\ E[x[n-N+1]d[n] \end{bmatrix} \\
&= \begin{bmatrix} p[0] \\ \vdots \\ p[N-1] \end{bmatrix}
\end{aligned}$$

If X and $d[n]$ are uncorrelated, then $P = \underline{0}$.

$$\begin{aligned}
R &= E[X(X)^T] \\
&= \begin{bmatrix} E[x[n]x[n]] & \cdots & E[x[n-N+1]x[n]] \\ \vdots & \ddots & \vdots \\ E[x[n]x[n-N+1] & \cdots & E[x[n-N+1]x[n-N+1]] \end{bmatrix}
\end{aligned}$$

$$= \begin{bmatrix} r[0] & \cdots & r[N-1] \\ \vdots & \ddots & \vdots \\ r[N-1] & \cdots & r[0] \end{bmatrix}$$

If $x[n]$ is a white noise process with unit variance, then $R = I$, the identity matrix.
Solution: The minimum error location will be where:

$$\nabla\xi(H_{\text{opt}}) = \underline{0} \qquad \text{where} \qquad \nabla\xi(H) = \left[\frac{\partial\,\xi}{\partial\,h[0]}, \ldots, \frac{\partial\,\xi}{\partial\,h[N-1]}\right]^T$$

As a result,

$$\begin{aligned} -2P + 2RH_{\text{opt}} &= \underline{0} \\ RH_{\text{opt}} &= P \end{aligned}$$

A correlation matrix R is always positive semi-definite, so unless $R = \underline{0}$, its inverse will exist:

$$H_{\text{opt}} = R^{-1}P$$

The minimum error can be found too:

$R^{-T} = R^{-1}$ since R is a cov. matrix

$$\begin{aligned} \xi_{\min} &= \xi(H_{\text{opt}}) \\ &= \sigma_d^2 - 2P^TR^{-T}P + P^TR^{-T}RR^{-1}P \\ &= \sigma_d^2 - P^TR^{-1}P \\ &= \sigma_d^2 - P^TH_{\text{opt}} \end{aligned}$$

and since R is positive semi-definite,

$$\xi_{\min} \le \sigma_d^2$$

Without filtering, the error is σ_d^2, so we do better!