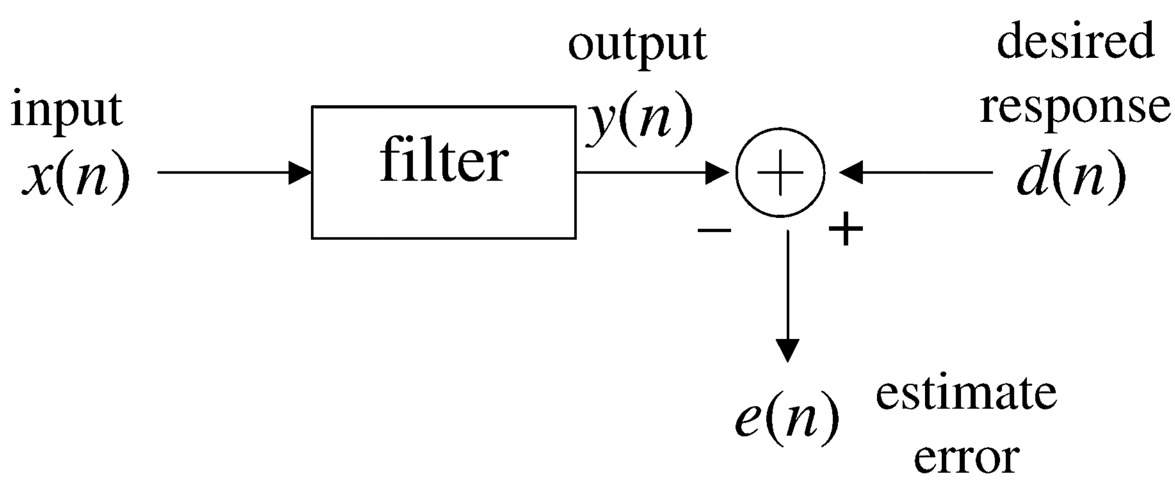
Weiner Filtering Theory

Problem: produce an estimate of a desired process



Restrictions placed on system

1. filter is linear
2. filter is discrete time

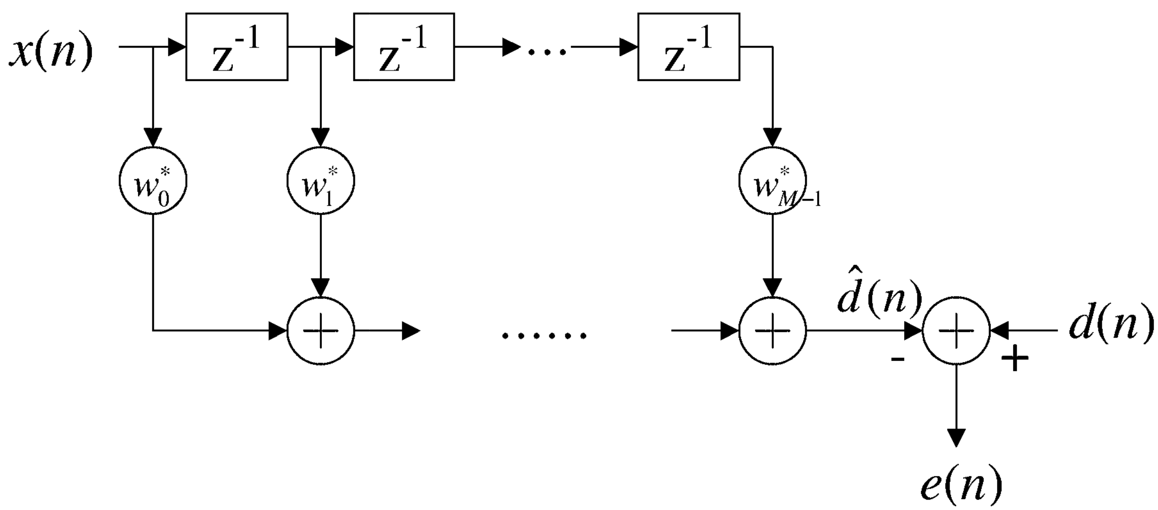
Also

1. Finite impulse response (FIR) is assumed
2. Statistical optimization is employed

The linear filtering problem was solved by

- Wiener in 1941 for continuous time
- Kolmogoror in 1938 for discrete time

For the discrete time case



The impulse response of the filter is

$$h_k = w_k^* \quad k = 0, 1, \cdots, M-1$$

and 0 for all other k

$\hat{d}(n)$ is the estimate of the desired signal $d(n)$.

Thus

$$\hat{d}(n) = \sum_{k=0}^{M-1} w_k^* x(n-k) = \mathbf{w}^H \mathbf{x}(n)$$

where

$$\mathbf{w} = [w_0, w_1, \cdots, w_{M-1}]^T$$

$$\mathbf{x} = [x(n), x(n-1), \cdots, x(n-M+1)]^T$$

The error can now be written as

$$e(n) = d(n) - \hat{d}(n) = d(n) - \mathbf{w}^H \mathbf{x}(n)$$

The performance criteria is chosen as the mean squared-error (MSE)

$$J(\mathbf{w}) = E\{e(n)e^*(n)\}$$

The $\mathbf{w}$ that minimizes $J(\mathbf{w})$ is the optimal (Wiener) filter.

Expanding the performance criteria,

$$\begin{aligned} J(\mathbf{w}) &= E\{e(n)e^*(n)\} \\ &= E\{(d(n) - \mathbf{w}^H\mathbf{x}(n))(d^*(n) - \mathbf{x}^H(n)\mathbf{w})\} \\ &= E\{|d(n)|^2 - d(n)\mathbf{x}^H(n)\mathbf{w} - \mathbf{w}^H\mathbf{x}(n)d^*(n) \\ &\quad +\mathbf{w}^H\mathbf{x}(n)\mathbf{x}^H(n)\mathbf{w}\} \\ &= E\{|d(n)|^2\} - E\{d(n)\mathbf{x}^H(n)\}\mathbf{w} - \mathbf{w}^H E\{\mathbf{x}(n)d^*(n)\} \\ &\quad +\mathbf{w}^H E\{\mathbf{x}(n)\mathbf{x}^H(n)\}\mathbf{w} \end{aligned}$$

Let

$$\mathbf{R} = E\{\mathbf{x}(n)\mathbf{x}^H(n)\} \quad \text{(autocorrelation of } \mathbf{x}(n))$$

$$\mathbf{p} = E\{\mathbf{x}(n)d^*(n)\} \quad \text{(cross correlation between } \mathbf{x}(n) \text{ and } d(n))$$

Then

$$J(\mathbf{w}) = \sigma_d^2 - \mathbf{p}^H\mathbf{w} - \mathbf{w}^H\mathbf{p} + \mathbf{w}^H\mathbf{R}\mathbf{w}$$

where we have assumed $x(n)$ is zero mean and stationary.

The error

$$J(\mathbf{w}) = \sigma_d^2 - \mathbf{p}^H\mathbf{w} - \mathbf{w}^H\mathbf{p} + \mathbf{w}^H\mathbf{R}\mathbf{w}$$

is a quadratic function of $\mathbf{w}$ and is thus a bowl-shaped function of $\mathbf{w}$ with a unique minimum.

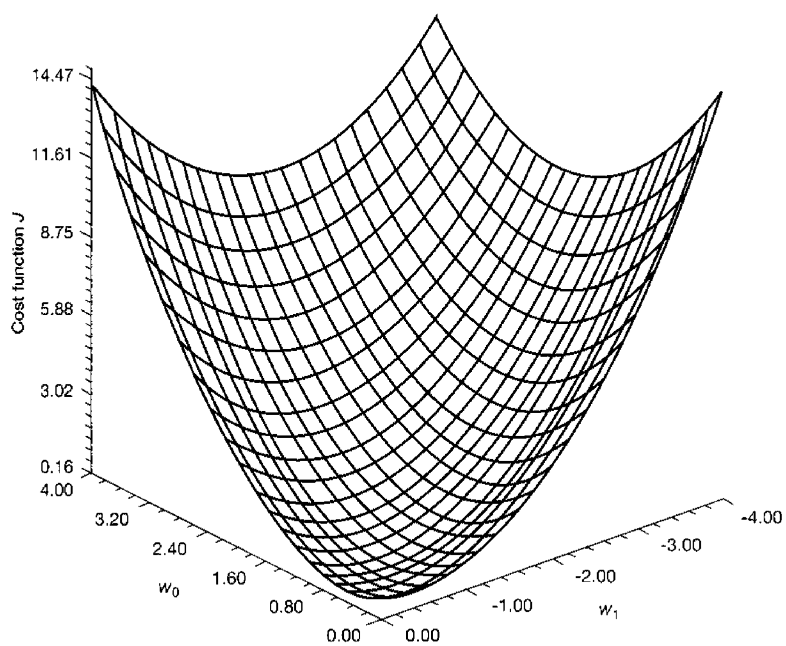


Figure 5.6 Error-performance surface of the two-tap transversal filter described in the numerical example.

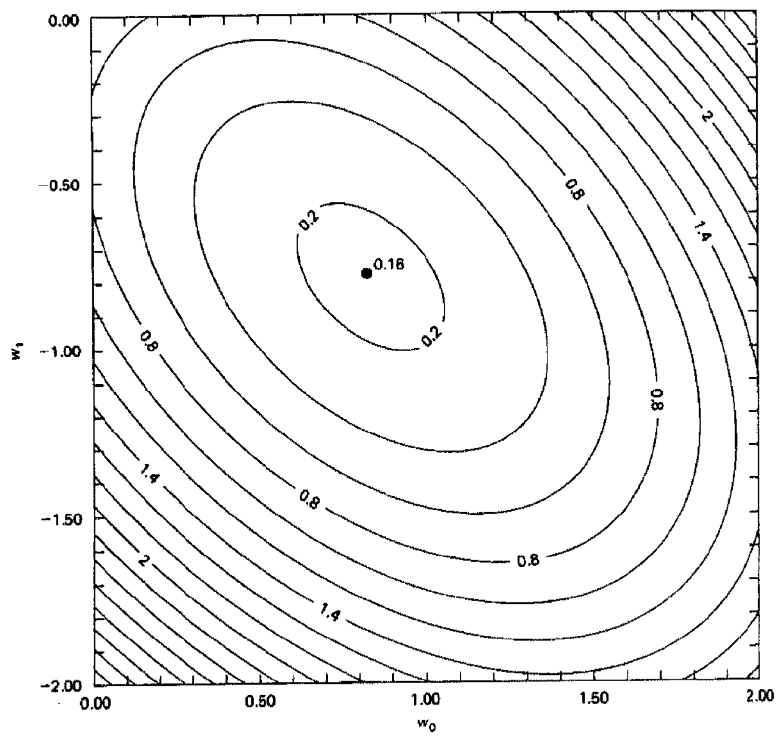


Figure 5.7 Contour plots of the error-performance surface depicted in Fig. 5.6.

We can find the optimal weight vector, $\mathbf{w}_0$, by differentiating $J(\mathbf{w})$ and setting this to zero

$$\nabla_{\mathbf{w}} J(\mathbf{w})|_{\mathbf{w}=\mathbf{w}_0} = 0$$

In general, for complex data,

$$w_k = a_k + jb_k \quad k = 0, 1, \cdots, M-1$$

the gradient, with respect to w_k, is

$$\nabla_k(J) = \frac{\partial J}{\partial a_k} + j\frac{\partial J}{\partial b_k} \quad k = 0, 1, \cdots, M-1$$

The complete gradient is given by

$$\nabla_{\mathbf{w}}(J) = \begin{bmatrix} \nabla_0(J) \\ \nabla_1(J) \\ \vdots \\ \nabla_{\text{M-1}}(J) \end{bmatrix} = \begin{bmatrix} \frac{\partial J}{\partial a_0} + j\frac{\partial J}{\partial b_0} \\ \frac{\partial J}{\partial a_1} + j\frac{\partial J}{\partial b_1} \\ \vdots \\ \frac{\partial J}{\partial a_{M-1}} + j\frac{\partial J}{\partial b_{M-1}} \end{bmatrix}$$

Examples of complex matrix differentiation: Let $\mathbf{c}$ and $\mathbf{w}$ be $M \times 1$ complex vectors.

For $g = \mathbf{c}^H \mathbf{w}$, find $\nabla_{\mathbf{w}}(g)$

$$g = \mathbf{c}^H \mathbf{w} = \sum_{k=0}^{M-1} \mathbf{c}_k^* \mathbf{w}_k = \sum_{k=0}^{M-1} \mathbf{c}_k^* (a_k + jb_k)$$

Thus

$$\begin{aligned} \nabla_k(g) &= \frac{\partial g}{\partial a_k} + j \frac{\partial g}{\partial b_k} \\ &= c_k^* + j(jc_k^*) \quad k = 0, 1, \cdots, M-1 \end{aligned}$$

Thus for $g = \mathbf{c}^H \mathbf{w}$

$$\nabla_{\mathbf{w}}(g) = \begin{bmatrix} \nabla_0(g) \\ \nabla_1(g) \\ \vdots \\ \nabla_{\text{M-1}}(g) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$$

Now suppose $g = \mathbf{w}^H\mathbf{c}$, then

$$g = \mathbf{w}^H\mathbf{c} = \sum_{k=0}^{M-1} \mathbf{c}_k \mathbf{w}_k^* = \sum_{k=0}^{M-1} \mathbf{c}_k (a_k - jb_k)$$

and

$$\begin{aligned}\nabla_k(g) &= \frac{\partial g}{\partial a_k} + j\frac{\partial g}{\partial b_k} \\ &= c_k + j(-jc_k) = 2c_k \quad k = 0, 1, \cdots, M-1\end{aligned}$$

Thus for $g = \mathbf{w}^H\mathbf{c}$

$$\nabla_{\mathbf{w}}(g) = \begin{bmatrix} \nabla_0(g) \\ \nabla_1(g) \\ \vdots \\ \nabla_{\text{M-1}}(g) \end{bmatrix} = \begin{bmatrix} 2c_0 \\ 2c_1 \\ \vdots \\ 2c_{M-1} \end{bmatrix} = 2\mathbf{c}$$

Lastly, suppose $g = \mathbf{w}^H \mathbf{Q} \mathbf{w}$, then

$$\begin{aligned} g &= \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} \mathbf{w}_i^* \mathbf{w}_j q_{i,j} \\ &= \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} (a_i - jb_i)(a_j + jb_j) q_{i,j} \end{aligned}$$

$$\begin{aligned} \nabla_k(g) &= \frac{\partial g}{\partial a_k} + j \frac{\partial g}{\partial b_k} \\ &= 2 \sum_{j=0}^{M-1} (a_j + jb_j) q_{k,j} + 0 \\ &= 2 \sum_{j=0}^{M-1} w_j q_{k,j} \end{aligned}$$

Thus for $g = \mathbf{w}^H \mathbf{Q} \mathbf{w}$

$$\nabla_{\mathbf{w}}(g) = \begin{bmatrix} \nabla_0(g) \\ \nabla_1(g) \\ \vdots \\ \nabla_{\text{M-1}}(g) \end{bmatrix} = 2 \begin{bmatrix} \sum_{i=0}^{M-1} q_{0,i} w_i \\ \sum_{i=0}^{M-1} q_{1,i} w_i \\ \vdots \\ \sum_{i=0}^{\text{M-1}} q_{M-1,i} w_i \end{bmatrix} = 2\mathbf{Q}\mathbf{w}$$

Returning to the MSE performance criteria

$$J(\mathbf{w}) = \sigma_d^2 - \mathbf{p}^H \mathbf{w} - \mathbf{w}^H \mathbf{p} + \mathbf{w}^H \mathbf{R} \mathbf{w}$$

differentiating with respect to $\mathbf{w}$ and using the above

$$\nabla_{\mathbf{w}}(J) = \mathbf{0} - \mathbf{0} - 2\mathbf{p} + 2\mathbf{R}\mathbf{w}$$

Setting this to zero gives the optimal weight vector, $\mathbf{w}_0$

$$\nabla_{\mathbf{w}}(J) = \mathbf{0}$$

$$\Downarrow$$

$$\mathbf{R}\mathbf{w}_0 = \mathbf{p} \quad \text{(normal equation)}$$

and the Wiener filter is defined by

$$\mathbf{w}_0 = \mathbf{R}^{-1}\mathbf{p}$$

Orthogonality Principle

Consider again the normal equation that defines the optimal solution

$$\mathbf{R}\mathbf{w}_0 = \mathbf{p}$$

$$E\{\mathbf{x}(n)\mathbf{x}^H(n)\}\mathbf{w}_0 = E\{\mathbf{x}(n)d^*(n)\}$$

Rearranging

$$E\{\mathbf{x}(n)d^*(n)\} - E\{\mathbf{x}(n)\mathbf{x}^H(n)\}\mathbf{w}_0 = \mathbf{0}$$

$$E\{\mathbf{x}(n)[d^*(n) - \mathbf{x}^H(n)\mathbf{w}_0]\} = \mathbf{0}$$

$$E\{\mathbf{x}(n)e_0^*(n)\} = \mathbf{0}$$

where $e_0^*(n)$ is the error when optimal weights are used.

Thus

$$E\{\mathbf{x}(n)e_0^*(n)\} = E\begin{bmatrix} x(n)e_0^*(n) \\ x(n-1)e_0^*(n) \\ \vdots \\ x(n-M+1)e_0^*(n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Thus for a filter to be optimal, a necessary and sufficient condition is that the estimate error, $e^*(n)$, is orthogonal to each input sample in $\mathbf{x}(n)$.

Having found the optimal filter, we can determine the minimum MSE. Recall

$$J(\mathbf{w}) = \sigma_d^2 - \mathbf{p}^H\mathbf{w} - \mathbf{w}^H\mathbf{p} + \mathbf{w}^H\mathbf{R}\mathbf{w}$$

Using the optimal weights $\mathbf{w}_0 = \mathbf{R}^{-1}\mathbf{p}$ in the above gives the minimum MSE.

$$\begin{aligned} J_{\min} &= \sigma_d^2 - \mathbf{p}^H\mathbf{w}_0 - \mathbf{w}_0^H\mathbf{p} + \mathbf{w}_0^H\mathbf{R}(\mathbf{R}^{-1}\mathbf{p}) \\ &= \sigma_d^2 - \mathbf{p}^H\mathbf{w}_0 - \mathbf{w}_0^H\mathbf{p} + \mathbf{w}_0^H\mathbf{p} \\ &= \sigma_d^2 - \mathbf{p}^H\mathbf{w}_0 \end{aligned}$$

or

$$J_{\min} = \sigma_d^2 - \mathbf{p}^H\mathbf{R}^{-1}\mathbf{p}$$

Next, consider

$$J(\mathbf{w}) - J_{\min} = -\mathbf{p}^H\mathbf{w} - \mathbf{w}^H\mathbf{p} + \mathbf{w}^H\mathbf{R}\mathbf{w} + \mathbf{p}^H\mathbf{w}_0 + \mathbf{w}_0^H\mathbf{p} - \mathbf{w}_0^H\mathbf{R}\mathbf{w}_0$$

Using the fact that

$$\mathbf{p} = \mathbf{R}\mathbf{w}_0 \quad \text{and} \quad \mathbf{p}^H = \mathbf{w}_0^H\mathbf{R}$$

in

$$
\begin{aligned}
J(\mathbf{w}) - J_{\min} &= -\mathbf{p}^H\mathbf{w} - \mathbf{w}^H\mathbf{p} + \mathbf{w}^H\mathbf{R}\mathbf{w} + \mathbf{p}^H\mathbf{w}_0 + \mathbf{w}_0^H\mathbf{p} - \mathbf{w}_0^H\mathbf{R}\mathbf{w}_0 \\
&= -\mathbf{w}_0^H\mathbf{R}\mathbf{w} - \mathbf{w}^H\mathbf{R}\mathbf{w}_0 + \mathbf{w}^H\mathbf{R}\mathbf{w} + \mathbf{w}_0^H\mathbf{R}\mathbf{w}_0 \\
&\quad +\mathbf{w}_0^H\mathbf{R}\mathbf{w}_0 - \mathbf{w}_0^H\mathbf{R}\mathbf{w}_0 \\
&= -\mathbf{w}_0^H\mathbf{R}\mathbf{w} - \mathbf{w}^H\mathbf{R}\mathbf{w}_0 + \mathbf{w}^H\mathbf{R}\mathbf{w} + \mathbf{w}_0^H\mathbf{R}\mathbf{w}_0 \\
&= (\mathbf{w} - \mathbf{w}_0)^H\mathbf{R}(\mathbf{w} - \mathbf{w}_0)
\end{aligned}
$$

$$\Downarrow$$

$$J(\mathbf{w}) = J_{\min} + (\mathbf{w} - \mathbf{w}_0)^H\mathbf{R}(\mathbf{w} - \mathbf{w}_0)$$

Finally, if we express $\mathbf{R}$ in terms of its eigenvalues and eigenvectors

$$J(\mathbf{w}) = J_{\min} + (\mathbf{w} - \mathbf{w}_0)^H\mathbf{Q}\boldsymbol{\Omega}\mathbf{Q}^H(\mathbf{w} - \mathbf{w}_0)$$

or defining the eigenvector transformed difference

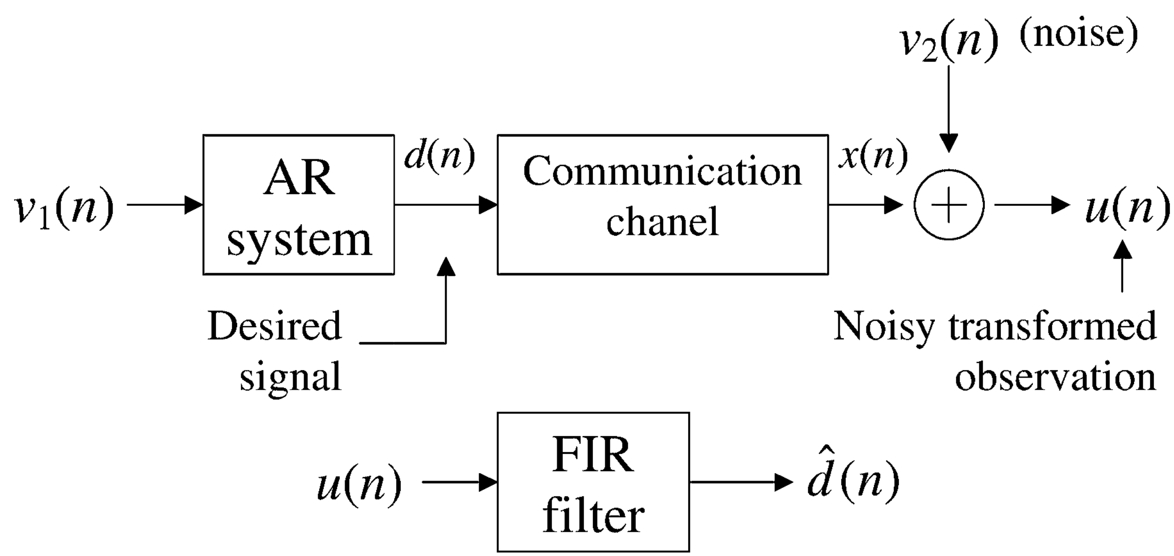
$$\mathbf{v} = \mathbf{Q}^H(\mathbf{w} - \mathbf{w}_0)$$

we have

$$J(\mathbf{w}) = J_{\min} + \mathbf{v}^H \mathbf{\Omega} \mathbf{v}$$
$$= J_{\min} + \sum_{k=1}^{M} \lambda_k v_k v_k^*$$
$$= J_{\min} + \sum_{k=1}^{M} \lambda_k |v_k|^2$$

where $\mathbf{v}_k$ is the difference $(\mathbf{w} - \mathbf{w}_0)$ projected onto eigenvector $\mathbf{q}_k$.

Example: Consider the following system



Goal: Determine the optimal order two filter weights, $\mathbf{w}_0$, for

$$H_1(z) = \frac{1}{1+0.8458z^{-1}} \quad \text{(AR process)}$$

$$H_2(z) = \frac{1}{1-0.9458z^{-1}} \quad \text{(communication channel)}$$

and $v_1(n)$ and $v_2(n)$ zero mean white noise with $\sigma_1^2 = 0.27$ and $\sigma_2^2 = 0.1$.

To determine $\mathbf{w}_0$, we need

$\mathbf{R_u}$ (auto-correlation of received signal)

$\mathbf{p}$ (cross correlation between received signal $\mathbf{u}$(n) and desired signal d(n))

Consider $\mathbf{R_u}$ first. Since $u(n) = x(n) + v_2(n)$, where $v_2(n)$ is uncorrelated with $x(n)$

$$\mathbf{R_u} = \mathbf{R_x} + \mathbf{R_{v_2}} = \mathbf{R_x} + \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}$$

Note that

$$X(z) = H_1(z)H_2(z)V_1(z)$$

where

$$H_1(z)H_2(z) = \frac{1}{(1+0.8458z^{-1})(1-0.9458z^{-1})}$$

or , $x(n)$ is an order 2 AR process

$$\begin{aligned} x(n) - 0.1x(n-1) - 0.8x(n-2) &= v_1(n) \\ x(n) + a_1 x(n-1) + a_2 x(n-2) &= v_1(n) \end{aligned}$$

Since $x(n)$ is a real valued order two AR process, the Yule-Walker equations are given by

$$\begin{bmatrix} r(0) & r(1) \\ r^*(1) & r(0) \end{bmatrix} \begin{bmatrix} -a_1 \\ -a_2 \end{bmatrix} = \begin{bmatrix} r^*(1) \\ r^*(2) \end{bmatrix}$$

$$\begin{bmatrix} r(0) & r(1) \\ r(1) & r(0) \end{bmatrix} \begin{bmatrix} -a_1 \\ -a_2 \end{bmatrix} = \begin{bmatrix} r(1) \\ r(2) \end{bmatrix}$$

which gives

$$-a_1 = \frac{r(1)[r(0) - r(2)]}{r^2(0) - r^2(1)}$$

$$-a_2 = \frac{r(0)r(2) - r^2(1)}{r^2(0) - r^2(1)}$$

or rearranging the noting $r(0) = \sigma_x^2$

$$\begin{aligned} r(1) &= \frac{-a_1}{1+a_2}\sigma_x^2 \\ r(2) &= \left(-a_2 + \frac{a_1^2}{1+a_2}\right)\sigma_x^2 \end{aligned}$$

The Yule-Walker equations also stipulate

$$\sigma_{v_1}^2 = r(0) + a_1 r(1) + a_2 r(2)$$

or rearranging and using the above

$$\sigma_x^2 = r(0) = \frac{1+a_2}{1-a_2}\frac{\sigma_{v_1}^2}{(1+a_2)^2 - a_1^2}$$

Using $a_1 = -0.1$, $a_2 = -0.8$, and $\sigma_{v_1}^2 = 0.27$, we have

$$r(0) = \sigma_x^2 = \frac{1+a_2}{1-a_2}\frac{\sigma_{v_1}^2}{(1+a_2)^2 - a_1^2} = 1$$

$$r(1) = \frac{-a_1}{1+a_2}\sigma_x^2 = 0.5$$

Thus

$$\mathbf{R_x} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$$

and

$$\mathbf{R_u} = \mathbf{R_x} + \mathbf{R_{v_2}} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} + \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} = \begin{bmatrix} 1.1 & 0.5 \\ 0.5 & 1.1 \end{bmatrix}$$

Recall that the Wiener solution is $\mathbf{w}_0 = \mathbf{R}^{-1}\mathbf{p}$.

Thus we must still determine

$$\mathbf{p} = E\left\{\begin{bmatrix} d(n)x(n) \\ d(n)u(n-1) \end{bmatrix}\right\}$$

Recall

$$X(z) = H_2(z)D(z) = \frac{D(z)}{1 - 0.9458z^{-1}}$$

or

$$x(n) - 0.9458x(n-1) = d(n)$$

and

$$u(n) = x(n) + v_2(n)$$

Thus

$$\begin{aligned} E\{u(n)d(n)\} &= E\{[x(n) + v_2(n)][x(n) - 0.9458x(n-1)]\} \\ &= E\{x^2(n)\} + E\{x(n)v_2(n)\} - 0.9458E\{x(n)x(n-1)\} \\ &\quad -0.9458E\{v_2(n)x(n-1)\} \end{aligned}$$

$$
\begin{aligned}
&= \sigma_x^2 + 0 - 0.9458r(1) - 0 \\
&= 1 - 0.9458\left(\frac{1}{2}\right) \\
&= 0.5272
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
E\{u(n-1)d(n)\} &= E\{[x(n-1) + v_2(n-1)][x(n) - 0.9458x(n-1)]\} \\
&= r(1) - 0.9458r(0) \\
&= -0.4458
\end{aligned}
$$

Thus

$$
\mathbf{p} = \begin{bmatrix} 0.5272 \\ -0.4458 \end{bmatrix}
$$

and

$$
\mathbf{w}_0 = \mathbf{R}^{-1}\mathbf{p}
$$

$$
\begin{aligned}
&= \begin{bmatrix} 1.1 & 0.5 \\ 0.5 & 1.1 \end{bmatrix}^{-1} \begin{bmatrix} 0.5272 \\ -0.4458 \end{bmatrix} \\
&= \begin{bmatrix} 0.8360 \\ -0.7853 \end{bmatrix}
\end{aligned}
$$