Linear Optimum Filtering
Wiener Filters

Problem statement

- Given the set of input samples $\{u(0), u(1), u(2), \ldots\}$ and the set of desired response $\{d(0), d(1), d(2), \ldots\}$
- In the family of filters computing their output according to

$$y(n) = \sum_{k=0}^{\infty} w_k u(n-k), \quad n = 0, 1, 2, \ldots \tag{1}$$

- Find the parameters $\{w_0, w_1, w_2, \ldots\}$ such as to minimize the mean square error defined as

$$J = E[e(n)^2] \tag{2}$$

where the error signal is

$$e(n) = d(n) - y(n) = d(n) - \sum_{l=0}^{\infty} w_l u(n-l) \tag{3}$$

□

The family of filters (1) is the family of linear discrete time filters (IIR or FIR).

Principle of ortogonality

Define the gradient operator ∇, having its k-th entry

$$\nabla_k = \frac{\partial}{\partial w_k} \tag{4}$$

and thus, the $k-$th entry of the gradient of criterium J is (remember, $e(n) = d(n) - \sum_{l=0}^{\infty} w_l u(n-l)$)

$$\nabla_k\ J \;=\; \frac{\partial J}{\partial w_k} = 2E\left[e(n)\frac{\partial e(n)}{\partial w_k}\right] = -2E\left[e(n)u(n-k)\right]$$

For the criterium to attain its minimum, the gradient of the criterium must be identically zero, that is

$$\nabla_k\ J = 0, \quad k = 0, 1, 2, \ldots$$

resulting in the fundamental

$$\textbf{Principle of ortogonality:} \quad E\left[e_o(n)u(n-k)\right] \;=\; 0, \quad k = 0, 1, 2, \ldots \tag{5}$$

Stated in words:

- The criterium J attains its minimum *iff*
- the estimation error $e_o(n)$ is orthogonal to the samples $u(i)$ which are used to compute the filter output.

We will index with o all the variables e.g. e_o, y_o computed using the optimal parameters $\{w_{o0}, w_{o1}, w_{o2}, \ldots\}$.
Let us compute the cross- correlation

$$E\left[e_o(n)y_o(n)\right] \;=\; E\left[e_o(n)\sum_{k=0}^{\infty} w_{ok}u(n-k)\right] = \sum_{k=0}^{\infty} w_{ok}E\left[u(n-k)e_o(n)\right] = 0 \tag{6}$$

Otherwise stated, in words, we have the following **Corolary of Orthogonality Principle:**

- When the criterium J attains its minimum then
- the estimation error $e_0(n)$ is orthogonal to the filter output $y_0(n)$.

Wiener – Hopf equations

From the orthogonality *estimation error – input window samples* we have

$$E\left[u(n-k)e_0(n)\right] = 0, \quad k = 0, 1, 2, \ldots$$
$$E\left[u(n-k)(d(n) - \sum_{i=0}^{\infty} w_{oi}u(n-i))\right] = 0, \quad k = 0, 1, 2, \ldots$$
$$\sum_{i=0}^{\infty} w_{oi}E\left[u(n-k)u(n-i)\right] = E\left[u(n-k)d(n)\right], \quad k = 0, 1, 2, \ldots$$

But

* $E\left[u(n-k)u(n-i)\right] = r(i-k)$ is the autocorrelation function of input signal $u(n)$ at lag $i-k$

* $E\left[u(n-k)d(n)\right] = p(-k)$ is the cross-correlation between the filter input $u(n-k)$ and the desired signal $d(n)$

and therefore

$$\sum_{i=0}^{\infty} w_{oi}r(i-k) = p(-k), \quad k = 0, 1, 2, \ldots \quad WIENER - HOPF \tag{7}$$

Solution of the Wiener – Hopf equations for linear transversal filters (FIR)

$$y(n) = \sum_{k=0}^{M-1} w_k u(n-k), \quad n = 0, 1, 2, \ldots \tag{8}$$

and since only $w_0, w_1, w_2, \ldots\ w_{M-1}$ are nonzero, Wiener-Hopf equations becomes

$$\sum_{i=0}^{M-1} w_{oi} r(i-k) = p(-k), \quad k = 0, 1, 2, \ldots, M-1 \quad WIENER-HOPF \tag{9}$$

which is a system of M equations with M unknowns: $\{w_{o0}, w_{o1}, w_{o2}, \ldots\}$.

Matrix formulation of Wiener – Hopf equations

Let us denote

$$\begin{aligned}
\underline{u}(n) &= \begin{bmatrix} u(n) & u(n-1) & u(n-2) & \ldots & u(n-M+1) \end{bmatrix}^T \\
R &= E[\underline{u}(n)\underline{u}^T(n)] = E \begin{bmatrix} u(n) \\ u(n-1) \\ u(n-2) \\ u(n-M+1) \end{bmatrix} \begin{bmatrix} u(n) & u(n-1) & u(n-2) & \ldots & u(n-M+1) \end{bmatrix} \qquad (10) \\
&= \begin{bmatrix} Eu(n)u(n) & Eu(n)u(n-1) & \ldots & Eu(n)u(n-M+1) \\ Eu(n-1)u(n) & Eu(n-1)u(n-1) & \ldots & Eu(n-1)u(n-M+1) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ Eu(n-M+1)u(n) & Eu(n-M+1)u(n-1) & \ldots & Eu(n-M+1)u(n-M+1) \end{bmatrix} = \begin{bmatrix} r(0) & r(1) & \ldots & r(M-1) \\ r(1) & r(0) & \ldots & r(M-2) \\ . & . & . & . \\ . & . & . & . \\ . & . & . & . \\ r(M-1) & r(M-2) & \ldots & r(0) \end{bmatrix} \\
\underline{p} &= E[\underline{u}(n)d(n)] = \begin{bmatrix} p(0) & p(-1) & p(-2) & \ldots & p(1-M) \end{bmatrix}^T \qquad (11) \\
\underline{w}_0 &= \begin{bmatrix} w_{o,0} & w_{o,1} & \ldots & w_{o,M-1} \end{bmatrix}^T \qquad (12)
\end{aligned}$$

then Wiener – Hopf equations can be written in a compact form

$$R\underline{w}_0 = \underline{p} \quad \text{with solution} \quad \underline{w}_o = R^{-1}\underline{p} \tag{13}$$

Mean square error surface

Let us define

$$e_{\underline{w}}(n) = d(n) - \sum_{k=0}^{M-1} w_k u(n-k) = d(n) - \underline{w}^T \underline{u}(n) \tag{14}$$

Then the cost function can be written as

$$\begin{aligned}
J_{\underline{w}} &= E[e_{\underline{w}}(n)e_{\underline{w}}(n)] = E[(d(n) - \underline{w}^T\underline{u}(n))(d(n) - \underline{u}^T(n)\underline{w})] \\
&= E[d^2(n) - d(n)\underline{u}^T(n)\underline{w} - \underline{w}^T\underline{u}(n)d(n) + \underline{w}^T\underline{u}(n)\underline{u}^T(n)\underline{w}] \\
&= E[d^2(n)] - E[d(n)\underline{u}^T(n)]\underline{w} - \underline{w}^T E[\underline{u}(n)d(n)] + \underline{w}^T E[\underline{u}(n)\underline{u}^T(n)]\underline{w} \\
&= E[d^2(n)] - 2E[d(n)\underline{u}^T(n)]\underline{w} + \underline{w}^T E[\underline{u}(n)\underline{u}^T(n)]\underline{w} \\
&= \sigma_d^2 - 2\underline{p}^T\underline{w} + \underline{w}^T R\underline{w} \\
&= \sigma_d^2 - 2\sum_{i=0}^{M-1} p(-i)w_i + \sum_{l=0}^{M-1}\sum_{i=0}^{M-1} w_l w_i R_{i,l}
\end{aligned} \tag{15}$$

Thus, we can proceed in another way to find the (same) optimal solution $\underline{w}_o$.

* $J_{\underline{w}}$ is a second order function of the parameters $\{\ w_0 \quad w_1 \quad \ldots \quad w_{M-1}\ \}$

* $J_{[\ w_0 \quad w_1 \quad \ldots \quad w_{M-1}\ ]}$ is a bowl shaped $M+1$- dimensional surface with M degrees of freedom.

* J attains the minimum, J_{min}, where the gradient is zero

$$
\begin{aligned}
\nabla_{\underline{w}} J &= 0 \\
\frac{\partial J}{\partial w_k} &= 0, \quad k = 0, 1, \ldots, M-1 \\
\frac{\partial J}{\partial w_k} &= -2p(-k) + 2\sum_{l=0}^{M-1} w_l r(k-l) = 0, \quad k = 0, 1, \ldots, M-1
\end{aligned}
$$

which finally gives the same Wiener – Hopf equations

$$
\sum_{l=0}^{M-1} w_l r(k-l) = p(-k) \tag{16}
$$

Minimum Mean square error

Using the form of the criterium

$$
J_{\underline{w}} = \sigma_d^2 - 2\underline{p}^T \underline{w} + \underline{w}^T R \underline{w} \tag{17}
$$

one can find the value of the minimum criterium (remember, $R\underline{w}_0 = \underline{p}$ and $\underline{w}_o = R^{-1}\underline{p}$):

$$
\begin{aligned}
J_{\underline{w}_o} &= \sigma_d^2 - 2\underline{p}^T \underline{w}_o + \underline{w}_o^T R \underline{w}_o = \sigma_d^2 - 2\underline{w}_o^T R \underline{w}_o + \underline{w}_o^T R \underline{w}_o \\
&= \sigma_d^2 - \underline{w}_o^T R \underline{w}_o \\
&= \sigma_d^2 - \underline{w}_o^T \underline{p} \\
&= \sigma_d^2 - \underline{p}^T R^{-1} \underline{p}
\end{aligned} \tag{18}
$$

Canonical form of the Error - performance surface

(Paranthesis: How to compute a scalar out of a vector $\underline{w}$, containing the entries of $\underline{w}$ at power one (linear combination) or at power two (quadratic form):

* linear combination (first order form) $\underline{a}^T\underline{w} = \sum_{l=0}^{M-1} a_l w_l$;

* quadratic form $\underline{w}^T R\underline{w} = \sum_{l=0}^{M-1}\sum_{i=0}^{M-1} w_l w_i R_{i,l} = w_0^2 R_{0,0} + w_0 w_1 R_{1,0} + \ldots w_{M-1}^2 R_{M-1,M-1}$)

How can we rewrite the criterium

$$J_{\underline{w}} \;=\; \sigma_d^2 - 2\underline{p}^T\underline{w} + \underline{w}^T R\underline{w} \tag{19}$$

in a quadratic form (how to complete a perfect "square", encompassing $-2\underline{p}^T\underline{w}$)?

Consider first the case when $\underline{w}$ is simply a scalar (resulting also in scalars $R, \underline{r}, \underline{p}$)

$$J_w = Rw^2 - 2pw + \sigma_d^2 = R(w^2 - 2w\frac{p}{R}) + \sigma_d^2 = R(w^2 - 2w\frac{p}{R} + \frac{p^2}{R^2}) - \frac{p^2}{R} + \sigma_d^2 = R(w - \frac{p}{R})^2 - \frac{p^2}{R} + \sigma_d^2$$

In the vector case for $\underline{w}$, the term corresponding to the one-dimensional $\frac{p^2}{R}$ is $\underline{p}^T R^{-1}\underline{p}$

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

$\square$ (This was the solution of exercise 5.5 page 182 in [Hayhin 91])

Let $\lambda_1, \lambda_2, \ldots, \lambda_M$ be the eigenvalues and (generally the complex) eigenvectors $\mu_1, \mu_2, \ldots, \mu_M$ of the matrix R, thus satisfying

$$R\mu_i = \lambda_i \mu_i \tag{20}$$

Then the matrix $Q = [\mu_1 \; \mu_2 \; \ldots \; \mu_M]$ can transform R to diagonal form Λ

$$R = Q\Lambda Q^H \tag{21}$$

where the superscript H means complex conjugation and transposition. Then

$$J_{\underline{w}} \;=\; J_{\underline{w}_0} + (\underline{w} - \underline{w}_0)^T R(\underline{w} - \underline{w}_0) = J_{\underline{w}_0} + (\underline{w} - \underline{w}_0)^T Q\Lambda Q^H (\underline{w} - \underline{w}_0)$$

Introduce now the transformed version of the tap vector w as

$$\underline{\nu} = Q^H(\underline{w} - \underline{w}_o) \tag{22}$$

Now the quadratic form can be put into its canonical form

$$\begin{aligned} J &= J_{\underline{w}_o} + \underline{\nu}^H \Lambda \nu \\ &= J_{\underline{w}_o} + \sum_{i=1}^{M} \lambda_i \nu_i \nu_i^* \\ &= J_{\underline{w}_o} + \sum_{i=1}^{M} \lambda_i |\nu_i|^2 \end{aligned}$$

Optimal Wiener Filter Design: Example

- *(Useful) Signal Generating Model* The model is given by the transfer function

$$H_1(z) = \frac{D(z)}{V_1(z)} = \frac{1}{1 + az^{-1}} = \frac{1}{1 + 0.8458z^{-1}}$$

  or the difference equation

$$d(n) + ad(n-1) = v_1(n) \qquad d(n) + 0.8458d(n-1) = v_1(n)$$

  where $\sigma_{v_1}^2 = r_{v_1}(0) = 0.27$

- *The channel (perturbation) model* is more complex. It involves a low pass filter with a transfer function

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

  leading for the variable $x(n)$ to the difference equation

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

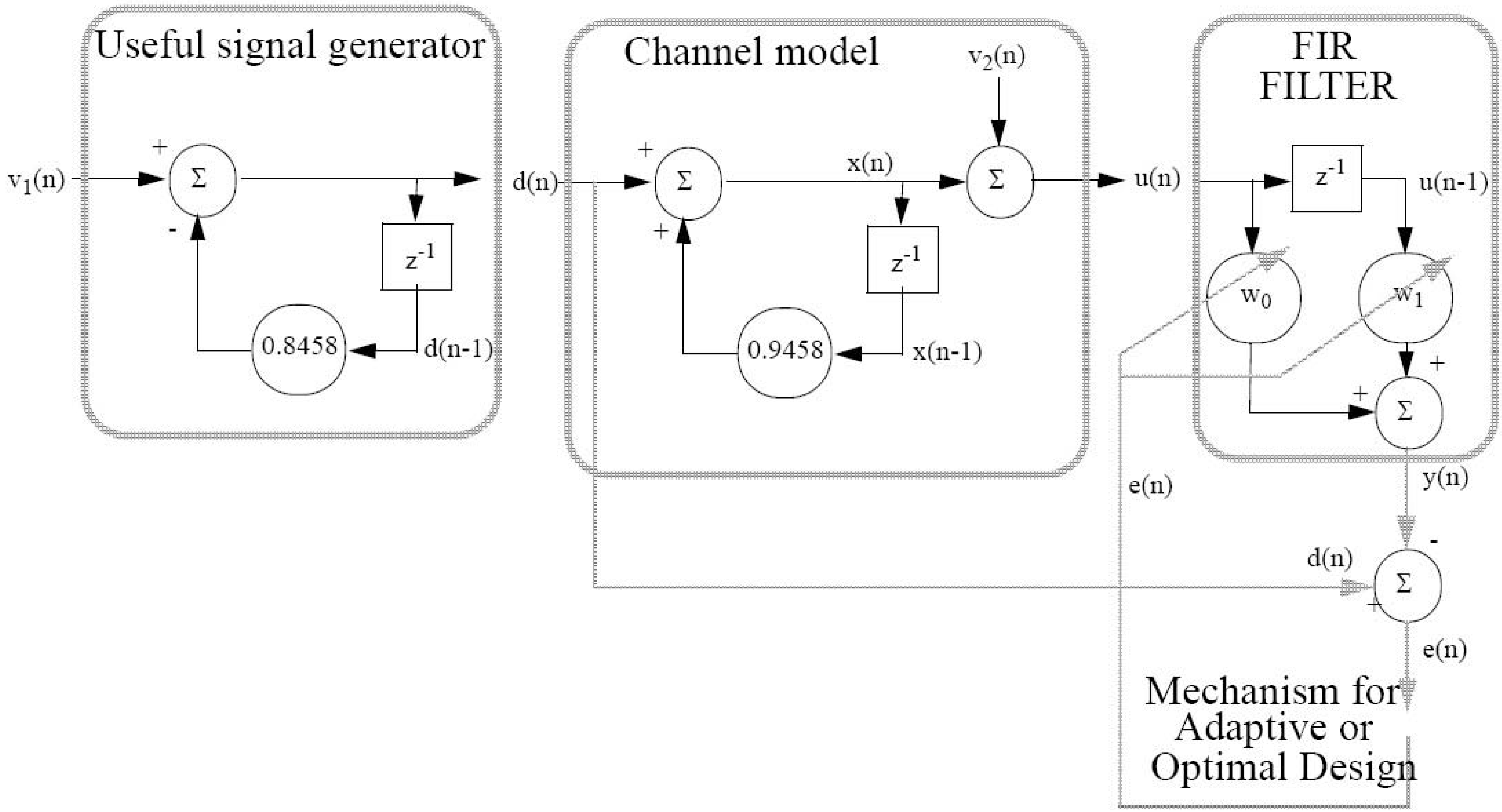
  and a white noise corruption ($x(n)$ and $v_2(n)$ are uncorrelated)

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

  with $\sigma_{v_2}^2 = r_{v_2}(0) = 0.1$ resulting in the final measurable signal $u(n)$.

- *FIR Filter* The signal $u(n)$ will be filtered in order to recover the original (useful) $d(n)$ signal, using the filter

$$y(n) = w_0u(n) + w_1u(n-1)$$


Useful signal generator
Channel model
FIR
FILTER
$v_1(n)$
Σ
+
-
z^{-1}
0.8458
$d(n-1)$
$d(n)$
$x(n)$
z^{-1}
0.9458
$x(n-1)$
$v_2(n)$
$u(n)$
$u(n-1)$
w_0
w_1
$y(n)$
$e(n)$
Mechanism for
Adaptive or
Optimal Design

We plan to apply the Wiener – Hopf equations

$$\begin{bmatrix} r_u(0) & r_u(1) \\ r_u(1) & r_u(0) \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} Ed(n)u(n) \\ Ed(n)u(n-1) \end{bmatrix}$$

The signal $x(n)$ obeys the generation model

$$H(z) = \frac{X(z)}{V_1(z)} = H_1(z)H_2(z) = \frac{1}{1+az^{-1}}\frac{1}{1+bz^{-1}} = \frac{1}{1+a_1z^{-1}+a_2z^{-2}} = \frac{1}{1-0.1z^{-1}-0.8z^{-2}}$$

and thus

$$x(n) + a_1x(n-1) + a_2x(n-2) = v_1(n)$$

Using the fact that $x(n)$ and $v_2(n)$ are uncorrelated and

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

it results

$$r_u(k) = r_x(k) + r_{v_2}(k)$$

and consequently, since for white noise $r_{v_2}(0) = \sigma_{v_2}^2 = 0.1$ and $r_{v_2}(1) = 0$ it follows

$$r_u(0) = r_x(0) + 0.1, \text{ and } \quad r_u(1) = r_x(1)$$

Now we concentrate to find $r_x(0), r_x(1)$ for the AR process

$$x(n) + a_1x(n-1) + a_2x(n-2) = v(n)$$

First multiply in turn the equation with $x(n)$, $x(n-1)$ and $x(n-2)$ and then take the expectation

$$Ex(n)\times \rightarrow \qquad Ex(n)x(n) + Ex(n)a_1x(n-1) + Ex(n)a_2x(n-2) = Ex(n)v(n)$$

$$\text{resulting in} \qquad r_x(0) + a_1r_x(1) + a_2r_x(2) = Ex(n)v(n) = \sigma_v^2$$

$$Ex(n-1)\times \rightarrow \qquad Ex(n-1)x(n) + Ex(n-1)a_1x(n-1) + Ex(n-1)a_2x(n-2) = Ex(n-1)v(n)$$

$$\text{resulting in} \qquad r_x(1) + a_1r_x(0) + a_2r_x(1) = Ex(n-1)v(n) = 0$$

$$Ex(n-2)\times \rightarrow \qquad Ex(n-2)x(n) + Ex(n-2)a_1x(n-1) + Ex(n-2)a_2x(n-2) = Ex(n-2)v(n)$$

$$\text{resulting in} \qquad r_x(2) + a_1r_x(1) + a_2r_x(0) = Ex(n-2)v(n) = 0$$

The equality $Ex(n)v(n) = \sigma_v^2$ can be obtained multiplying the AR model difference equation with $v(n)$ and then taking expectations

$$Ev(n)\times \rightarrow \qquad Ev(n)x(n) + Ev(n)a_1x(n-1) + Ev(n)a_2x(n-2) = Ev(n)v(n)$$

$$\text{resulting in} \qquad Ev(n)x(n) = \sigma_v^2$$

since $v(n)$ is uncorrelated with older values, $x(n-\tau)$. We obtained the most celebrated Yule Walker equations:

$$\begin{aligned} r_x(0) + a_1r_x(1) + a_2r_x(2) &= \sigma_v^2 \\ r_x(1) + a_1r_x(0) + a_2r_x(1) &= 0 \\ r_x(2) + a_1r_x(1) + a_2r_x(0) &= 0 \end{aligned}$$

or as usually given in matrix form

$$\begin{bmatrix} r_x(0) & r_x(1) & r_x(2) \\ r_x(1) & r_x(0) & r_x(1) \\ r_x(2) & r_x(1) & r_x(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sigma_v^2 \\ 0 \\ 0 \end{bmatrix}$$

But we need to use the equations differently:

$$\begin{bmatrix} 1 & a_1 & a_2 \\ a_1 & 1+a_2 & 0 \\ a_2 & a_1 & 1 \end{bmatrix} \begin{bmatrix} r_x(0) \\ r_x(1) \\ r_x(2) \end{bmatrix} = \begin{bmatrix} \sigma_v^2 \\ 0 \\ 0 \end{bmatrix}$$

Solving for $r_x(0), r_x(1), r_x(2)$ we obtain

$$r_x(0) = \left(\frac{1+a_2}{1-a_2}\right) \frac{\sigma_v^2}{(1+a_2)^2 - a_1^2}$$

$$r_x(1) = \frac{-a_1}{1+a_2} r_x(0)$$

$$r_x(2) = \left(-a_2 + \frac{a_1^2}{1+a_2}\right) r_x(0)$$

In our example we need only the first two values, $r_x(0), r_x(1)$, which results to be $r_x(0) = 1, \;\; r_x(1) = 0.5$,

Now we will solve for the cross-correlations $Ed(n)u(n), Ed(n)u(n-1)$. First observe

$$Eu(n)d(n) = E(x(n) + v_2(n))d(n) = Ex(n)d(n)$$
$$Eu(n-1)d(n) = E(x(n-1) + v_2(n-1))d(n) = Ex(n-1)d(n)$$

and now take as a "master" difference equation

$$x(n) + bx(n-1) = d(n)$$

and multiply it in turn with $x(n)$ and $x(n-1)$ and then take the expectation

$$\begin{aligned} Ex(n) \rightarrow \qquad Ex(n)x(n) + Ex(n)bx(n-1) &= Ex(n)d(n) \\ Ex(n)d(n) &= r_x(0) + br_x(1) \end{aligned}$$

$$\begin{aligned} Ex(n-1) \rightarrow \qquad Ex(n-1)x(n) + Ex(n-1)bx(n-1) &= Ex(n-1)d(n) \\ Ex(n-1)d(n) &= r_x(1) + br_x(0) \end{aligned}$$

Using the numerical values, one obtain

$$Eu(n)d(n) = Ex(n)d(n) = 0.5272 \qquad Eu(n-1)d(n) = Ex(n-1)d(n) = -0.4458$$

Now we have all necessary variables needed to write the Wiener – Hopf equations

$$\begin{bmatrix} r_u(0) & r_u(1) \\ r_u(1) & r_u(0) \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} Ed(n)u(n) \\ Ed(n)u(n-1) \end{bmatrix}$$

$$\begin{bmatrix} 1.1 & 0.5 \\ 0.5 & 1.1 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 0.5272 \\ -0.4458 \end{bmatrix}$$

resulting in

$$\begin{bmatrix} w_0 \\ w_1 \end{bmatrix} = \begin{bmatrix} 0.8360 \\ -0.7853 \end{bmatrix}$$