

## Wiener Filters

### Linear Optimum Filtering: Problem Statement

Figure 5.1

**Problem Statement:** Design a linear, DT filter whose output  $y(n)$  provides an estimate of a desired response  $d(n)$ , given a set of input samples  $u(0), u(1), u(2), \dots$ . The estimate is such that the mean-square value of the estimation error  $e(n)$ , defined as the difference between the desired response  $d(n)$  and the actual response  $y(n)$ , is minimized.

We'll assume an IIR filter for now (FIR is the special case) and note that

$$y(n) = \sum_{k=0}^{\infty} \omega_k^* u(n-k)$$

and  $w_k = \alpha_k + \varphi_k \beta_k$ . We have as our estimation error

$$e(n) = \delta(n) - \psi(n)$$

and thus our cost function defined as the mean-squared error (MSE)

$$\begin{aligned} J &= \xi \\ &= E[\varepsilon(n)\varepsilon^*(n)] \\ &= E[|\varepsilon(n)|^2] \end{aligned}$$

We wish to determine  $w_0, w_1, w_2, \dots, K$  such that the MSE is minimized. We will solve for the coefficients using two methods (both ways lead to the same solution)

**Method 1:** Solve for  $w_k$  by completing the square (restricted to FIR).

**Method 2:** Take derivative (gradient) of  $J$  w.r.t. to  $w_k$  and set this to zero. This will lead us to orthogonality principles that govern optimal filter condition. Solve for  $w_k$  given orthogonality principles.

The latter method will lead to additional insight.

---

### Wiener Filter Solution (Method 1)

$$\begin{aligned} J(\mathbf{w}) &= E[\varepsilon(n)\varepsilon^*(n)] \\ &= E[(\delta(n) - \psi(n))(\delta(n) - \psi(n))^*] \\ &= E[(\delta(n) - \mathbf{w}^H \mathbf{v}(n))(\delta(n) - \mathbf{w}^H \mathbf{v}(n))^*] \\ &= E[(\delta(n) - \mathbf{w}^H \mathbf{v}(n))(\delta^*(n) - \mathbf{v}^H(n)\mathbf{w})] \\ &= E[|\delta^2(n)| - \mathbf{w}^H E[\mathbf{v}(n)\delta^*(n)] - E[\mathbf{v}^H(n)\delta(n)]\mathbf{w} + \mathbf{w}^H E[\mathbf{v}(n)\mathbf{v}^H(n)]\mathbf{w}] \\ &= \sigma_s^2 - \mathbf{w}^H \boldsymbol{\pi} - \boldsymbol{\pi}^H \mathbf{w} + \mathbf{w}^H \mathbf{P} \mathbf{w} \end{aligned}$$

where

$\sigma_d^2$  is the variance of the desired response

$\mathbf{w} = [\omega_0 \ \Lambda \ \omega_{M-1}]^T$  are the filter coefficients

$\mathbf{p} = E[\mathbf{u}(n)\delta^*(n)]$  is the  $M \times 1$  cross correlation vector between the input vector  $\mathbf{u}(n)$  and desired response  $d(n)$

$\mathbf{R} = E[\mathbf{u}(n)\mathbf{u}^H(n)]$  is the  $M \times M$  correlation matrix of the input vector

We complete the square of the MSE as follows.

$$\begin{aligned} J(\mathbf{w}) &= \sigma_\delta^2 - \boldsymbol{\omega}^H \boldsymbol{\pi} - \boldsymbol{\pi}^H \boldsymbol{\omega} + \boldsymbol{\omega}^H \mathbf{P} \boldsymbol{\omega} + \boldsymbol{\pi}^H \mathbf{P}^{-1} \boldsymbol{\pi} - \boldsymbol{\pi}^H \mathbf{P}^{-1} \boldsymbol{\pi} \\ &= \sigma_\delta^2 - \boldsymbol{\pi}^H \mathbf{P}^{-1} \boldsymbol{\pi} + (\boldsymbol{\omega} - \mathbf{P}^{-1} \boldsymbol{\pi})^H \mathbf{P} (\boldsymbol{\omega} - \mathbf{P}^{-1} \boldsymbol{\pi}) \end{aligned}$$

We can minimize  $J(\mathbf{w})$  by choosing

$$\boxed{\begin{aligned} \mathbf{w} &= \mathbf{P}^{-1} \boldsymbol{\pi} \\ &= \mathbf{w}_o \end{aligned}}$$

With this choice, we have

$$\boxed{J_{\min} = \sigma_\delta^2 - \boldsymbol{\pi}^H \mathbf{P}^{-1} \boldsymbol{\pi}}$$

The MSE (cost function)  $J$  is a second-order (quadratic) function of the filter coefficients  $\mathbf{w}$ . Consequently we may visualize the dependence of  $J$  on  $\mathbf{w}$  as a  $M + 1$  dimensional paraboloid or error performance surface. This surface has a unique minimum where  $J$  attains its minimum value using  $\mathbf{w}_o$ . We note that at the bottom of this surface, the gradient is zero.

**Figure 5.6**

---

### **Principle of Orthogonality (Method 2)**

In order to solve the statistical filtering problem, we minimize  $J$  (real function, same as  $\xi$ ) using standard calculus techniques (set first derivative or gradient of the cost function with respect to the filter coefficients equal to zero). Therefore we assume the  $k$ th filter coefficient is given by

$$w_k = \alpha_k + j\beta_k, \quad k = 0, 1, 2, \dots$$

and define the gradient operator as

$$\nabla_k = \frac{\partial}{\partial \alpha_k} + j \frac{\partial}{\partial \beta_k}$$

We apply the gradient operator to the cost function i.e. take multidimensional 1st derivative of  $J$  w.r.t. the filter coefficients. We then have the gradient vector  $\nabla J$  for which the  $k$ th element is defined as

$$\nabla_k J = \frac{\partial J}{\partial \alpha_k} + j \frac{\partial J}{\partial \beta_k}$$

For the cost function to be minimized we require

$$\nabla_k J = 0$$

1) We have as our cost function

$$J = E[e(n)e^*(n)]$$

2) Applying the gradient operator (using the derivative product rule) to our cost function we have

$$\begin{aligned} \nabla_k J &= E \left[ \frac{\partial e(n)}{\partial w_k} e^*(n) + \frac{\partial e^*(n)}{\partial w_k} e(n) \right] \\ &= E \left[ \frac{\partial e(n)}{\partial a_k} e^*(n) + \frac{\partial e^*(n)}{\partial a_k} e(n) + j \frac{\partial e(n)}{\partial b_k} e^*(n) + j \frac{\partial e^*(n)}{\partial b_k} e(n) \right] \end{aligned}$$

3) Since

$$\begin{aligned} e(n) &= \mathbf{w}^H \mathbf{u}(n) \\ &= \sum_k w_k u(n-k) \\ &= \sum_k (a_k + jb_k) u(n-k) \end{aligned}$$

we have

$$\begin{aligned} \frac{\partial e(n)}{\partial a_k} &= -u(n-k) \\ \frac{\partial e(n)}{\partial b_k} &= ju(n-k) \\ \frac{\partial e^*(n)}{\partial a_k} &= -u^*(n-k) \\ \frac{\partial e^*(n)}{\partial b_k} &= -ju^*(n-k) \end{aligned}$$

4) Substituting the above partial derivatives into our cost function gradient we arrive at

$$\nabla_k J = -2E[u(n-k)e^*(n)].$$

Let  $e_o$  denote the special value of the estimation error that results when the filter operates in its optimum condition. Then

$$E[u(n-\kappa)e_o^*(\nu)] = 0.$$

**Principle of orthogonality.** The necessary and sufficient condition for the cost function  $J$  to attain its minimum value is that the corresponding value of the estimation error  $e_o(n)$  is orthogonal to each input sample that enters into the estimation of the desired response at time  $n$ .

---

### Corollary to the Principle of Orthogonality

Consider

$$\begin{aligned} E[y(n)\varepsilon^*(\nu)] &= E \left[ \sum_{\kappa=0}^{\infty} \omega_{\kappa}^* u(\nu-\kappa)\varepsilon^*(\nu) \right] \\ &= \sum_{\kappa=0}^{\infty} \omega_{\kappa}^* E[u(\nu-\kappa)\varepsilon^*(\nu)] \end{aligned}$$

Let  $y_o$  denote the output produced by the filter optimized in the mean-square sense with  $e_o$  denoting the corresponding estimation error. Then using the Principle of orthogonality we also have

$$E[y_o(n)\varepsilon_o^*(\nu)] = 0 .$$

*Corollary to the Principle of Orthogonality.* When the filter operates in its optimum condition, the estimate of the desired response defined by the filter output,  $y_o(n)$ , and the corresponding estimation error,  $e_o$ , are orthogonal to each other.

---

**Geometric Interpretation of the Corollary to the Principle of orthogonality****Figure 5.3**

As an analogy, we see that for the optimum filter the vector representing the estimation error is normal (or perpendicular or orthogonal) to the vector representing the filter output.