Statistical Digital Signal Processing

Week 6 Optimum Filters: FIR Wiener Filter

Lecturer: ***Zelalem Hailu (Assistant Prof.)***

Addis Ababa Science and Technology University,
Addis Ababa, Ethiopia

Previous Topic (Week-5)

Levinson-Durbin Recursion and Lattice Filters

- Basics of Levinson-Durbin Recursion
- Development of Levinson-Durbin Recursion
- Computational Complexity of Levinson-Durbin Recursion
- Lattice Filters
- Properties of The Reflection Coefficient

Lecture Learning Outcomes

1. Explain the fundamental concepts of optimum filtering, including the criteria for optimality and the role of statistical properties of signals and noise in filter design.
2. Analyze and derive the Wiener Filter, and interpret its significance in minimizing mean square error (MSE) for signal estimation.
3. Formulate and implement the FIR Wiener filter, including solving the Wiener–Hopf equations.
4. Apply concepts of linear prediction and noise cancellation, including predictive modeling of signals and filtering techniques for removing unwanted noise components.

Week 6: Optimum Filters: FIR Wiener Filter

Outline

- Optimum Filters: Introduction
- Wiener Filter
- The FIR Wiener Filter
- Filtering Application
- Linear Prediction
- Noise Cancellation

Optimum Filters: Introduction

- In most practical applications such as radar communication, wireless communication etc., the desired signals are not observed directly
- In such applications, the desired signal may be noisy or distorted
- Therefore, estimation of the desired signal from its noisy observation is the most important problem in Varity of engineering applications
- In a very simple ideal environment, it may be possible to design classical filters such as low-pass, band-pass, and high-pass filters to recover the desired signal
- However, in most practical cases those classical filters are not optimum in producing the best estimate of the desired signal from its noisy observations
- To solve such problems, Digital Optimum Filters such as Wiener Filter and Discrete Kalman Filter become very important

Optimum Filters: Introduction

- The signal estimation problem can be illustrated as follows:

Figure 1: Illustration of Optimum Filtering

- Our objective is to estimate the random signal x_n based on the available related signal y_n and the estimated signal $\hat{x}_n$ becomes the function of y_n
- We will focus Linear estimation which applies the linear mean square (LMS) criterion

Optimum Filters: Introduction

- The estimated signal is expressed as a linear combination of the observed data:

$$\hat{x}(n) = \sum_{i=n_a}^{n_b} h(n,i)\, y_i \qquad (1)$$

Where: the estimation interval is $\rightarrow n_a \leq i \leq n_b$

- The weight $h(n,i)$ which can minimize the following mean square estimation error will be selected [1]:

$$\xi = E\left\{e_n^2\right\} = E\left\{\left[x_n - \hat{x}_n\right]^2\right\} \qquad (2)$$

- Hence, the objective is to minimize the mean square error given on the above equation, eq(2)

Wiener Filter

- In the 1940s, Norbert Wiener pioneered the design of optimal filter to estimate signals from noisy observations
- The Wiener filtering problem, is to design a filter to recover a signal $d(n)$ from noisy measurement $x(n)$ given by:

$$x(n) = d(n) + v(n)$$

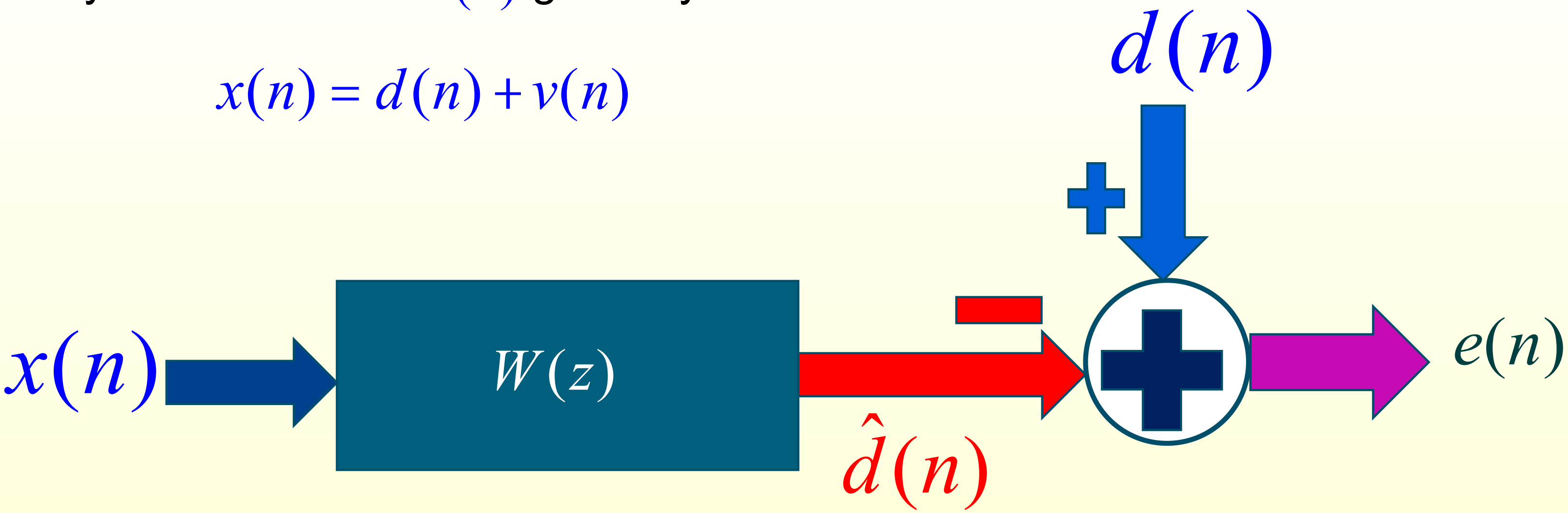

Figure 2: Illustration of Wiener Filtering

Wiener Filter

- Assuming $d(n)$ and $v(n)$ are wide-sense stationary random processes, Wiener formulated the design of a filter that minimizes the mean square error in estimating $d(n)$ from $x(n)$
- The mean square error term is also given by:

$$\xi = E\left\{| e(n) |^2\right\} \qquad (3)$$

Where:

$$e(n) = d(n) - \hat{d}(n) \qquad (4)$$

- Thus, the problem is to find the linear shift invariant (LSI) filter that minimizes the mean square error ξ
- The filter might be **Finite Impulse Response (FIR) filter** or **Infinite Impulse Response (IIR) filter**

Wiener Filter

- Depending on the relationship between $x(n)$ and $d(n)$, various important problems can be formulated within the Wiener filtering framework.
- Some of the problems which can be addressed by Wiener filtering are as follows:

I. Filtering

- Given that the signal $x(n) = d(n) + v(n)$ is available, the objective is to estimate $d(n)$ from current and past values of $x(n)$ using causal filter

II. Smoothing

- This is similar to the filtering problem except that the filter is allowed to be non-causal
- The Weiner smoothing filter may be designed to estimate $d(n)$ from all available data of $x(n) = d(n) + v(n)$

Wiener Filter

III. Prediction

- If the present and past values of $x(n)$ are available, the goal to produce the prediction of future value, $x(n+\alpha)$ for $\alpha > 0$

IV. Deconvolution

- When $x(n) = d(n) * g(n) + v(n)$ with $g(n)$ is the unit sample response of linear shift invariant filter, the winer filter which is used to estimate $d(n)$ from $x(n)$ is a deconvolution Wiener filter

V. Noise Cancelation

- Measure the noise corrupted signal $x(n) = d(n) + v_1(n)$ and the noise $v_2(n)$ using two separate sensors
- Then, estimate $v_1(n)$ from $v_2(n)$ and subtract from $x(n)$ to recover the target signal $d(n)$

The FIR Wiener Filter

- In this week lecture, we will focus on the design of FIR Wiener filter to estimate a given process $d(n)$ by filtering a statistically related observation $x(n)$
- The design should also ensure minimization of the mean square estimation error
- We will do also the following assumptions:

$x(n)$ & $d(n)$ **: are jointly wide-sense stationary**

$r_x(k),\ r_d(k),\ r_{dx}(k)$**: are known**

- Considering a $p-1$ order Wiener filter, the system function of the filter $W(z)$ is given by:

$$W(z) = \sum_{n=0}^{p-1} w(n) z^{-n} \qquad (5)$$

Where:

$w(n)$ is the unit sample response of the Wiener filter

The FIR Wiener Filter

- For the input $x(n)$, the output of the filter $\hat{d}(n)$ is given by:

$$\hat{d}(n) = w(n) * x(n) = \sum_{l=0}^{p-1} w(l)x(n-l) \quad (6)$$

- The Wiener filter design seeks finding the coefficients $w(k)$ that minimize the following mean-square error:

$$\xi = E\left\{ |e(n)|^2 \right\} = E\left\{ |d(n) - \hat{d}(n)|^2 \right\} \quad (7)$$

- Using the optimization steps for $k = 0, 1, \ldots, p-1$, we will have:

$$\frac{\partial \xi}{\partial w^*(k)} = \frac{\partial}{\partial w^*(k)} E\left\{ |e(n)|^2 \right\} = \frac{\partial}{\partial w^*(k)} E\left\{ e(n)e^*(n) \right\}$$

$$= E\left\{ e(n) \frac{\partial e^*(n)}{\partial w^*(k)} \right\} = 0 \quad (8)$$

The FIR Wiener Filter

- From eq(6) and eq(7), we have:

$$e(n) = d(n) - \hat{d}(n) = d(n) - \sum_{l=0}^{p-1} w(l)x(n-l) \qquad (9)$$

- Then, conjugating eq(9):

$$e^*(n) = d^*(n) - \sum_{l=0}^{p-1} w^*(l)x^*(n-l) \qquad (10)$$

- Taking the partial derivative of eq(10) for the k^{th} value or at the particular value of $l = k$, we will have:

$$\frac{\partial e^*(n)}{\partial w^*(k)} = \frac{\partial}{\partial w^*(k)}\left[d^*(n) - w^*(k)x^*(n-k)\right]$$

$$= -x^*(n-k) \qquad (11)$$

The FIR Wiener Filter

- Using eq(11), now eq(8) can be written in the following form :

$$\frac{\partial \xi}{\partial w^*(k)} = E\left\{ e(n) \frac{\partial e^*(n)}{\partial w^*(k)} \right\} = 0 \quad (12)$$

Substituting $\frac{\partial e^*(n)}{\partial w^*(k)} = -x^*(n-k)$

$$E\left\{ e(n) x^*(n-k) \right\} = 0 \quad ; k = 0,1,\ldots,p-1 \quad (13)$$

- Eq (13) is called ***Orthogonality principle*** or ***Projection Theorem***

- Then eq(13) can be written in the following form :

$$E\left\{ e(n) x^*(n-k) \right\} = E\left\{ \left[d(n) - \sum_{l=0}^{p-1} w(l) x(n-l) \right] x^*(n-k) \right\} = 0 \quad (14)$$

The FIR Wiener Filter

- Then:

$$E\left\{d(n)x^*(n-k)\right\} - \sum_{l=0}^{p-1} w(l) E\left\{x(n-l)x^*(n-k)\right\} = 0 \qquad (15)$$

- Recalling $x(n)$ and $d(n)$ jointly wide-sense stationary, we have :

$$E\left\{d(n)x^*(n-k)\right\} = r_{dx}(k) \qquad (16)$$

$$E\left\{x(n-l)x^*(n-k)\right\} = r_x(k-l) \qquad (17)$$

- Using eq(16 & 17), eq(15) can be written in the following form :

$$\sum_{l=0}^{p-1} w(l) r_x(k-l) = r_{dx}(k) \; ; \; k = 0,1,\ldots,p-1 \qquad (19)$$

The FIR Wiener Filter

- Using the following conjugate symmetry characteristics of the autocorrelation sequence:

$$r_x(k) = r_x^*(-k) \qquad (20)$$

- Writing eq(19) in matrix form:

$$\underbrace{\begin{bmatrix} r_x(0) & r_x^*(1) & \cdots & r_x^*(p-1) \\ r_x(1) & r_x(0) & \cdots & r_x^*(p-2) \\ r_x(2) & r_x(1) & \cdots & r_x^*(p-3) \\ \vdots & \vdots & & \vdots \\ r_x(p-1) & r_x(p-2) & \cdots & r_x(0) \end{bmatrix}}_{\mathbf{R}_x} \underbrace{\begin{bmatrix} w(0) \\ w(1) \\ w(2) \\ \vdots \\ w(p-1) \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} r_{dx}(0) \\ r_{dx}(1) \\ r_{dx}(2) \\ \vdots \\ r_{dx}(p-1) \end{bmatrix}}_{\mathbf{r}_{dx}} \qquad (21)$$

The FIR Wiener Filter

- Eq(21) is the matrix form of ***Wiener-Hopf Equations*** and writing it in compact form:

$$\mathbf{R}_x\mathbf{w} = \mathbf{r}_{dx} \qquad (22)$$

Where:

$\mathbf{R}_x$ is a $p \times p$ Hermitian Toeplitz matrix of autocorrelation

$\mathbf{w}$ is the vector of filter coefficients

$\mathbf{r}_{dx}$ is the vector cross correlation between $d(n)$ and $x(n)$

- The minimum mean square estimation error can also be evaluated as:

$$\xi = E\left\{|e(n)|^2\right\} = E\left\{e(n)\left[d(n) - \sum_{l=0}^{p-1} w(l)x(n-l)\right]^*\right\}$$

$$= E\left\{e(n)d^*(n)\right\} - \sum_{l=0}^{p-1} w(l)E\left\{e(n)x^*(n-l)\right\} \qquad (23)$$

The FIR Wiener Filter

- Recalling the orthogonality principle given in eq(13), i.e.: $\rightarrow E\left\{e(n)x^*(n-k)\right\}=0$, the minimum mean square error, $\xi_{\min}$, can be derived from eq(23) as:

$$\xi_{\min}=E\left\{e(n)d^*(n)\right\}=E\left\{\left[d(n)-\sum_{l=0}^{p-1}w(l)x(n-l)\right]d^*(n)\right\}$$

$$=E\left\{d(n)d^*(n)\right\}-\sum_{l=0}^{p-1}w(l)E\left\{x(n-l)d^*(n)\right\} \qquad (24)$$

- Taking the expected value, eq(24) boils down to:

$$\xi_{\min}=r_d(0)-\sum_{l=0}^{p-1}w(l)r_{xd}(-l)$$

$$=r_d(0)-\sum_{l=0}^{p-1}w(l)r_{dx}^*(l) \qquad (25)$$

Filtering Application

- In Filtering problem, the goal is to estimate $d(n)$ from noise corrupted observation $x(n)$

$$x(n) = d(n) + v(n) \qquad (26)$$

- Filtering (noise reduction) is widely used in:
  - Speech transmission in noisy environments
  - Data communication over noisy channels
  - Image enhancement
- Assuming the noise $v(n)$ is zero mean and uncorrelated with $d(n)$, hence:

$$r_{dv}(k) = E\left\{d(n)v^*(n-k)\right\} = 0 \qquad (27)$$

Filtering Application

- Then the cross correlation between $d(n)$ and $x(n)$ becomes:

$$\begin{aligned} r_{dx}(k) &= E\left\{d(n)x^*(n-k)\right\} \\ &= E\left\{d(n)\left[d^*(n-k)+v^*(n-k)\right]\right\} \\ &= E\left\{d(n)d^*(n-k)\right\}+E\left\{d(n)v^*(n-k)\right\} \\ &= r_d(k) \end{aligned} \tag{28}$$

- Then $r_x(k)$ also given by:

$$\begin{aligned} r_x(k) &= E\left\{x(n)x^*(n-k)\right\} \\ &= E\left\{\left[d(n)+v(n)\right]\left[d^*(n-k)+v^*(n-k)\right]\right\} \\ &= r_d(k)+r_v(k) \end{aligned} \tag{29}$$

Filtering Application

- In Filtering problem, ***Wiener-Hopf Equations*** given on eq(22), $\mathbf{R}_x \mathbf{w} = \mathbf{r}_{dx}$, becomes:

$$[\mathbf{R}_d + \mathbf{R}_v]\mathbf{w} = \mathbf{r}_d \qquad (30)$$

Where: $\mathbf{R}_d$ is the autocorrelation matrix for $d(n)$ $\quad$ $\mathbf{r}_d$ is the autocorrelation vector for $d(n)$

$\mathbf{R}_v$ is the autocorrelation matrix for $v(n)$

- In matrix form:

$$\underbrace{\begin{bmatrix} r_d(0)+r_v(0) & r_d^*(1)+r_v^*(1) & \cdots & r_d^*(p-1)+r_v^*(p-1) \\ r_d(1)+r_v(1) & r_d(0)+r_v(0) & \cdots & r_d^*(p-2)+r_v^*(p-2) \\ r_d(2)+r_v(2) & r_d(1)+r_v(1) & \cdots & r_d^*(p-3)+r_v^*(p-3) \\ \vdots & \vdots & & \vdots \\ r_d(p-1)+r_v(p-1) & r_d(p-2)+r_v(p-2) & \cdots & r_d(0)+r_v(0) \end{bmatrix}}_{\mathbf{R}_d+\mathbf{R}_v} \underbrace{\begin{bmatrix} w(0) \\ w(1) \\ w(2) \\ \vdots \\ w(p-1) \end{bmatrix}}_{\mathbf{w}} = \underbrace{\begin{bmatrix} r_d(0) \\ r_d(1) \\ r_d(2) \\ \vdots \\ r_d(p-1) \end{bmatrix}}_{\mathbf{r}_d} \qquad (31)$$

Linear Prediction in Absence of Noise

- Linear prediction is an important problem in variety of engineering disciplines signal processing applications:
- **Linear Prediction in Noise Free Environment:** Concerned with predicting $x(n+1)$ as a linear combination of current and past samples of $x(n)$

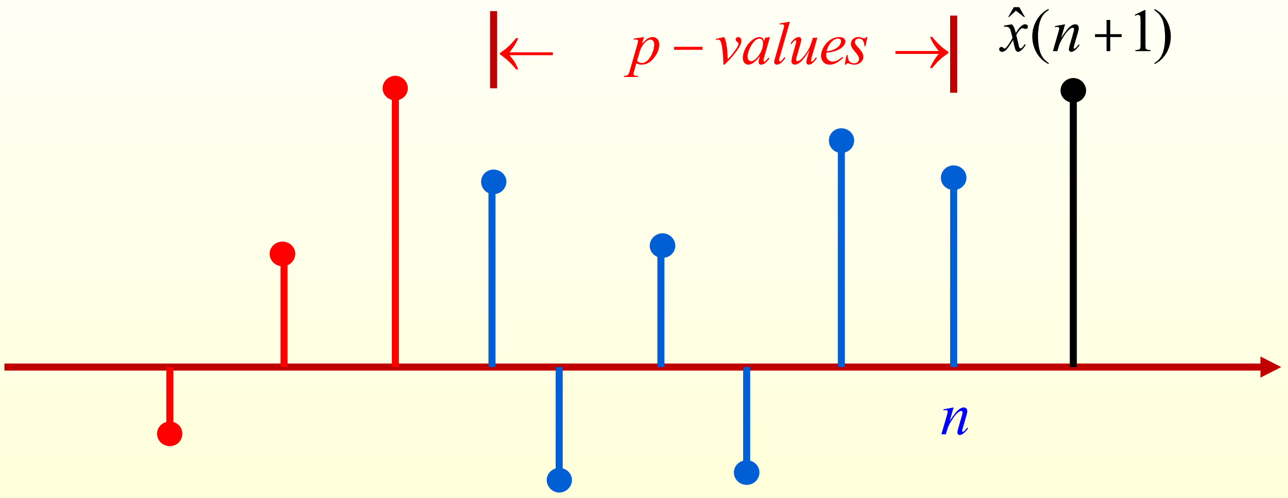

Figure 3: Illustration of Single-Step Linear Prediction

Linear Prediction in Absence of Noise

- An FIR linear predictor of order $p-1$ has the form:

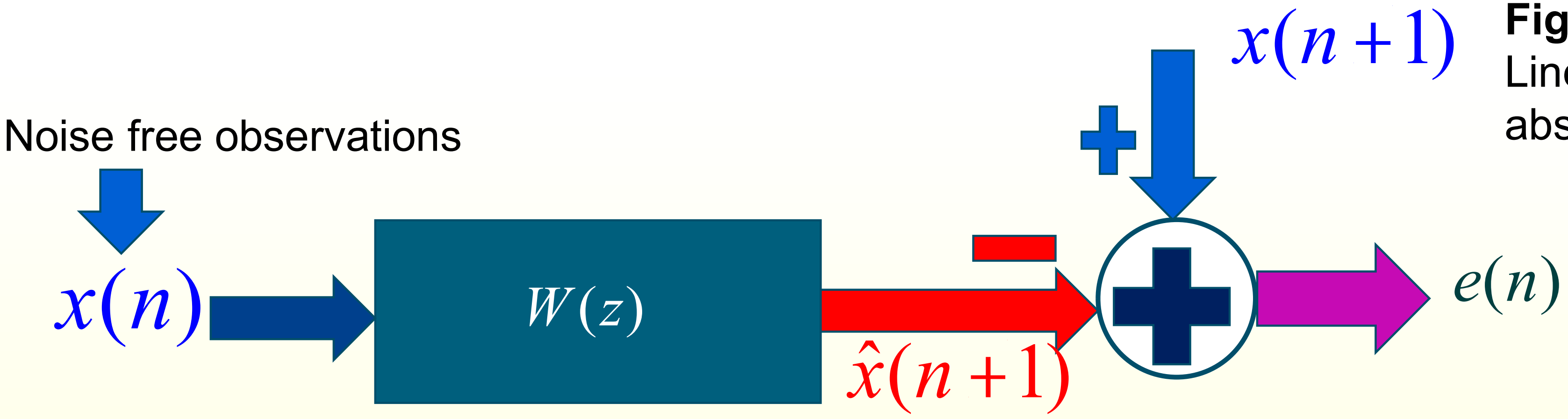

Figure 4: Single-Step Linear Prediction in absence of noise

- Compared to the conventional Wiener filter setup, the difference is:

$d(n)$ **Replaced by** $x(n+1)$

$\hat{d}(n)$ **Replaced by** $\hat{x}(n+1)$

- Therefore, we can cast this prediction problem in to conventional Winer Filter analysis

Linear Prediction in Absence of Noise

- Therefore the predicted signal can be written as the output of the predictor filter as:

$$\hat{x}(n+1) = \sum_{k=0}^{p-1} w(k)x(n-k) \qquad (32)$$

Where:

$w(k)$ is the coefficient of prediction filter

- Setting $d(n) = x(n+1)$, the $r_{dx}(k)$ terms in Wiener-Hopf Equations given at eq(22) becomes:

$$\begin{aligned} r_{dx}(k) &= E\left\{d(n)x^{*}(n-k)\right\} \\ &= E\left\{x(n+1)x^{*}(n-k)\right\} \\ &= r_{x}(k+1) \qquad (33) \end{aligned}$$

Linear Prediction in Absence of Noise

- Then, the **Wiener-Hopf Equations** for the optimum linear predictor becomes:

$$\begin{bmatrix} r_x(0) & r_x^*(1) & \cdots & r_x^*(p-1) \\ r_x(1) & r_x(0) & \cdots & r_x^*(p-2) \\ r_x(2) & r_x(1) & \cdots & r_x^*(p-3) \\ \vdots & \vdots & & \vdots \\ r_x(p-1) & r_x(p-2) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ w(2) \\ \vdots \\ w(p-1) \end{bmatrix} = \begin{bmatrix} r_x(1) \\ r_x(2) \\ r_x(3) \\ \vdots \\ r_x(p) \end{bmatrix} \qquad (34)$$

- And the minimum mean square prediction error:

$$\xi_{\min} = r_x(0) - \sum_{k=0}^{p-1} w(k) r_x^*(k+1) \qquad (35)$$

Linear Prediction in Prescence of Noise

- The **noise free assumption** we have seen shofar is not realistic and practical
- The realistic scenario is considering the noisy observation at the input of the predictor:

$$y(n) = x(n) + v(n) \qquad (36)$$

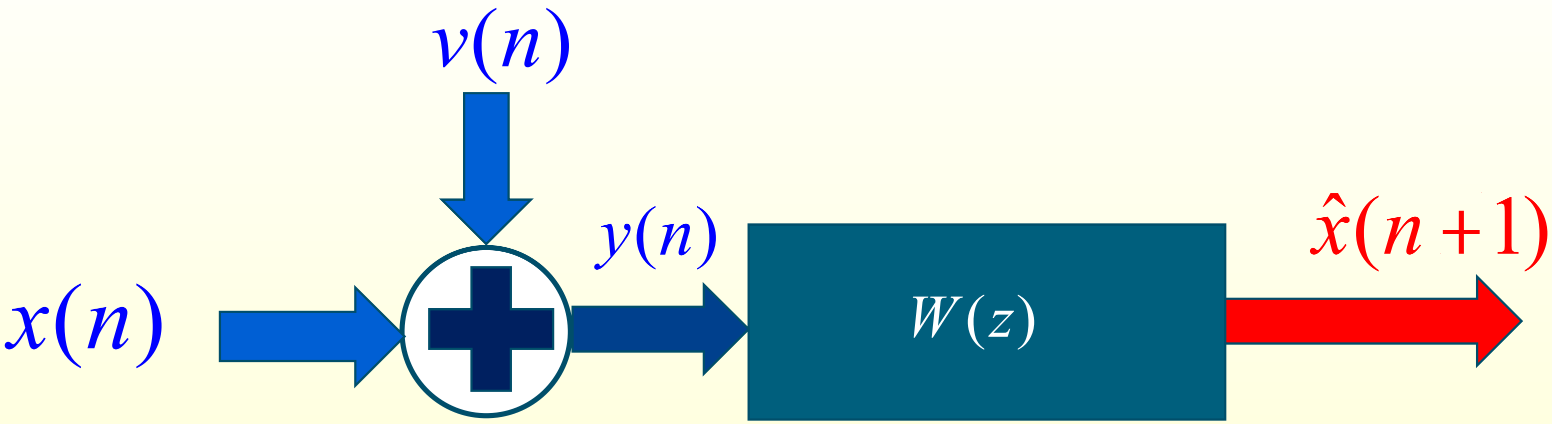

Figure 5: Single-Step Linear Prediction in Prescence of Noise

- The goal is to predict $x(n+1)$ as a linear combination of p values of $y(n)$

Linear Prediction in Prescence of Noise

- Therefore,

$$\hat{x}(n+1)=\sum_{k=0}^{p-1} w(k)y(n-k)=\sum_{k=0}^{p-1} w(k)\left[x(n-k)+v(n-k)\right] \qquad (37)$$

- The Wiener-Hopf Equations will have the following form:

$$\mathbf{R}_y\mathbf{w}=\mathbf{r}_{dy} \qquad (38)$$

- Assuming the noise $v(n)$ is uncorrelated with the signal $x(n)$, we have:

$$r_y(k)=E\left\{y(n)y^*(n-k)\right\}=E\left\{\left[x(n)+v(n)\right]\left[x^*(n-k)+v^*(n-k)\right]\right\}$$
$$=r_x(k)+r_v(k) \qquad (39)$$

- And,

$$r_{dy}(k)=E\left\{d(n)y^*(n-k)\right\}=E\left\{x(n+1)\left[x^*(n-k)+v^*(n-k)\right]\right\}$$
$$=r_x(k+1) \qquad (40)$$

Multi-Step Linear Prediction

- So far, we have seen prediction of $x(n+1)$ using current and previous values of $x(n)$ which is a One-step or Single-step linear prediction
- **Multi-Step Linear Prediction**: the goal is to predict $x(n+\alpha)$ for $\alpha > 1$ using current and previous values of $x(n)$

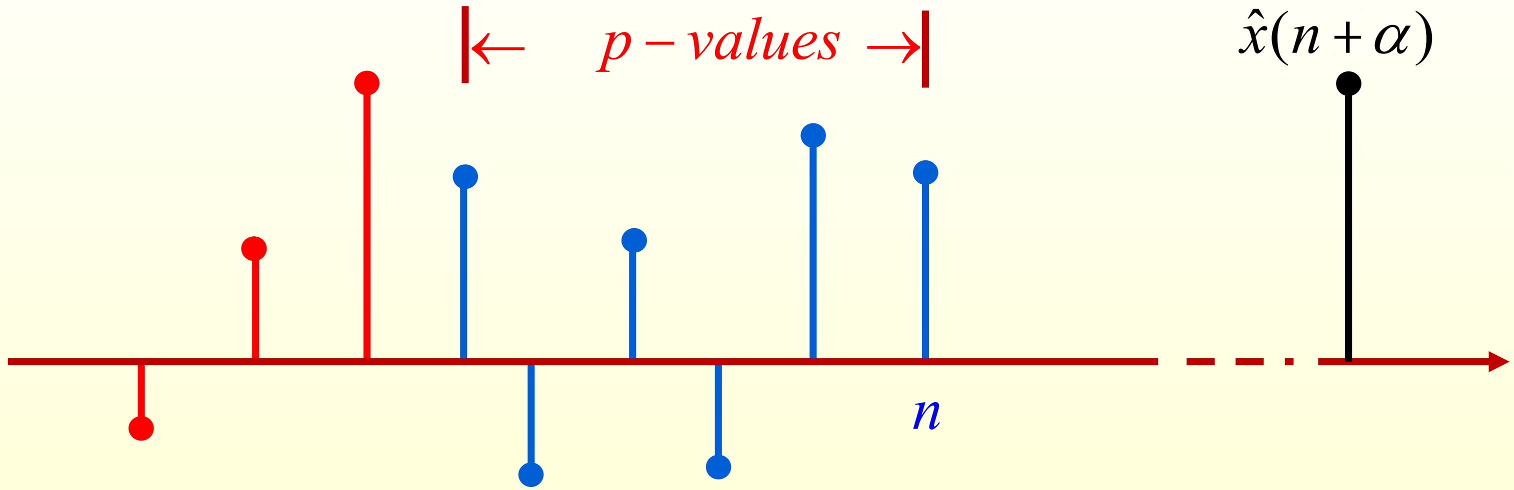

Figure 6: Illustration of Multi-Step Linear Prediction

Multi-Step Linear Prediction

- The predicted sample is given by:

$$\hat{x}(n+\alpha) = \sum_{k=0}^{p-1} w(k)x(n-k) \qquad (41)$$

- Similarly:

$$r_{dx}(k) = E\left\{d(n)x^*(n-k)\right\}$$

$$= E\left\{x(n+\alpha)x^*(n-k)\right\}$$

$$= r_x(k+\alpha) \qquad (42)$$

Multi-Step Linear Prediction

- The Wiener-Hopf Equations for optimum multi-step predictor in matrix form:

$$\begin{bmatrix} r_x(0) & r_x^*(1) & \cdots & r_x^*(p-1) \\ r_x(1) & r_x(0) & \cdots & r_x^*(p-2) \\ r_x(2) & r_x(1) & \cdots & r_x^*(p-3) \\ \vdots & \vdots & & \vdots \\ r_x(p-1) & r_x(p-2) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ w(2) \\ \vdots \\ w(p-1) \end{bmatrix} = \begin{bmatrix} r_x(\alpha) \\ r_x(\alpha+1) \\ r_x(\alpha+2) \\ \vdots \\ r_x(p-1+\alpha) \end{bmatrix} \qquad (43)$$

- And the minimum mean square prediction error:

$$\xi_{\min} = r_x(0) - \sum_{k=0}^{p-1} w(k) r_x^*(k+\alpha) \qquad (44)$$

Noise Cancellation

- The target of noise canceller is to estimate the signal $d(n)$ from noise corrupted observation, $x(n) = d(n) + v_1(n)$, recorded by primary sensor as shown in the figure below [2].

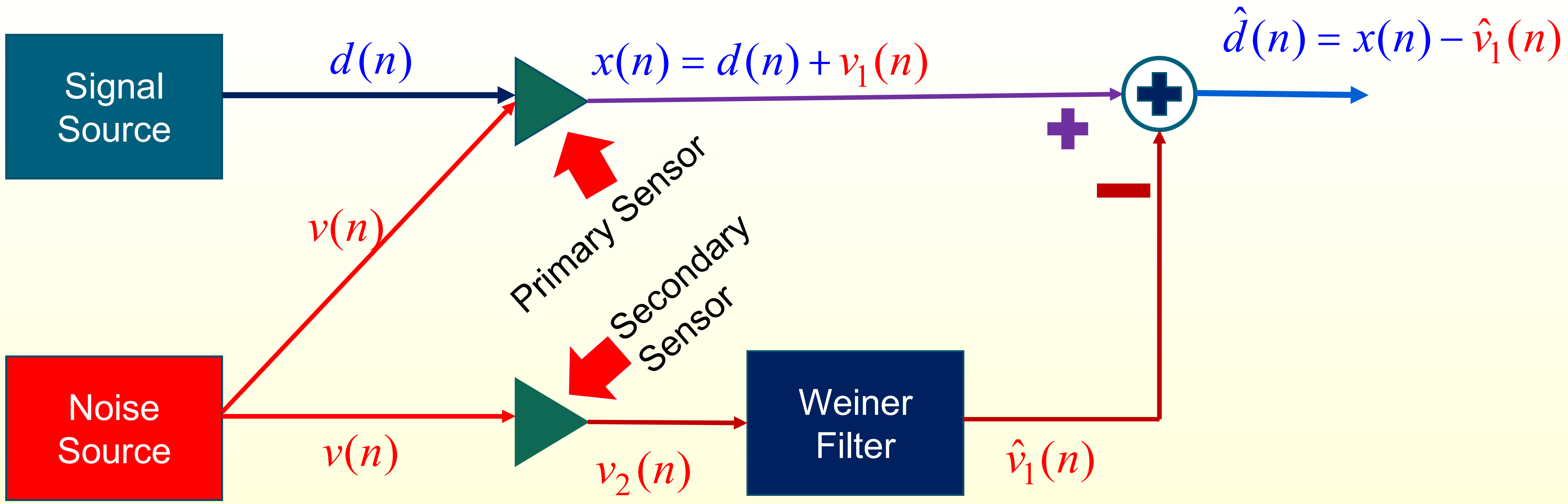

Figure 7: Illustration of Noise Cancellation

Noise Cancellation

- $v_1(n)$ and $v_2(n)$ are correlated but not similar
- It is possible to estimate $v_1(n)$ from $v_2(n)$ by using Wiener filter linear estimation problem
- Remembering the Winer filtering problem, $\mathbf{R}_x\mathbf{w} = \mathbf{r}_{dx}$, when the input $x(n)$ and the signal to be estimated is $d(n)$
- Similarly, for the current case, we have:

$$\mathbf{R}_{v_2}\mathbf{w} = \mathbf{r}_{v_1 v_2} \qquad (45)$$

- $r_{v_1 v_2}(k)$, the cross correlation between $v_1(n)$ and $v_2(n)$ is given as:

$$\begin{aligned} r_{v_1 v_2}(k) &= E\left\{v_1(n)v_2^*(n-k)\right\} = E\left\{\left[x(n)-d(n)\right]v_2^*(n-k)\right\} \\ &= E\left\{x(n)v_2^*(n-k)\right\} - \underbrace{E\left\{d(n)v_2^*(n-k)\right\}} \\ &= E\left\{x(n)v_2^*(n-k)\right\} \\ &= r_{xv_2}(k) \qquad (46) \end{aligned}$$

Equals to zero due to Orthogonality

Summary

- **Optimum Filters:**
  - ✓ For estimation of the desired signal from its noisy observation
  - ✓ Applicable to real-world applications, where classical filters often prove inadequate under practical conditions
- **Digital Optimum Filters**
  - ✓ **Wiener Filter**
    - ❖ FIR Wiener Filter
    - ❖ IIR Wiener Filter
  - ✓ **Discrete Kalman Filter**
- **FIR Wiener Filters**
  - ✓ **Applications:** Filtering (Estimation), Prediction, Noise Cancilation

References

[1] Charles W. Therrien,‘’ *Discrete Random Signals and Statistical Signal Processing”*, Prentice Hall, Pp.338, 1992.

[2] Monson H. Hayes, *“Statistical Digital Signal Processing and Modeling”,* John Wiley and sons, Pp.349, 1996.

Contents Here

Thank You!