# *OPTIMUM FILTERS (1)*

## *INTRODUCTION*

Signals and measurements:

Signal to noise ratio (SNR):

The *q*-transform:

Optimum filtering and the mean square error minimization:

Issues in Wiener filtering:

Types of Wiener filters – causal and noncausal FIR and IIR filters:

## *FIR WIENER FILTERS*

FIR Wiener filtering problems in a general way:

Wiener filtering:

Wiener prediction:

## *INTRODUCTION*

A filter is a system that is designed to process signals into those we desire. The purpose of using a filter can be various, to extract a desired signal from noisy data (measurement), to transform signals, to suppress noise, to separate two signals that are mixed in one measurement, *etc*. An optimum filter is such a filter used for acquiring a best estimate of desired signal from noisy measurement. It is different from the classic filters like lowpass, highpass and bandpass filters. Optimal filters are optimum because they are designed based on optimization theory to minimize the mean square error between a processed signal and a desired signal, or equivalently provides the best estimation of a desired signal from a measured noisy signal. The optimal filters studied in this chapter are *linear optimum discrete-time filters*, which include discrete Wiener filters and discrete Kalman filters. All of the topics in (linear) optimum filtering can be developed based on a single fact known as the *orthogonality principle*, which is the consequence of applying the optimization theory.

### Signals and measurements:

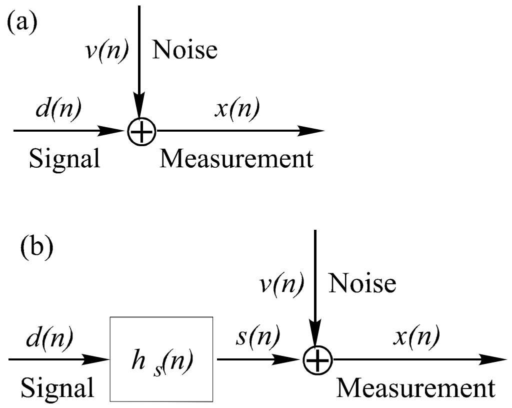



**Fig. 1.** Measurements of (a) a signal with noise, and (b) a signal with distortion and noise.

It is pervasive that when we measure a (desired) signal $d(n)$, noise $v(n)$ interferes with the signal so that a measured signal (Fig. 1(a)) becomes a noisy signal $x(n)$

$$x(n)=d(n)+v(n) \tag{1}$$

It is also very common that a signal $d(n)$ is distorted in its measurement (e.g., an electromagnetic signal distorts as it propagates over a radio channel). Assuming that the system causing distortion is characterized by an impulse response of $h_s(n)$, the measurement of $d(n)$ (Fig. 1(b)) can be expressed by the sum of distorted signal $s(n)$ and noise $v(n)$

$$x(n)=s(n)+v(n)=h_s(n)*d(n)+v(n) \tag{2}$$

where

$$s(n)=h_s(n)*d(n). \tag{3}$$

If both $d(n)$ and $v(n)$ are assumed to be wide-sense stationary (WSS) random processes, then $x(n)$ is also a WSS process. The signals that we discuss in this chapter will be WSS if they are not specially specified.

If signal $d(n)$ and measurement noise $v(n)$ are assumed to be uncorrelated (this is true in many practical cases), then $r_{dv}(k)=r_{vd}(k)=0$ . In this case, the noisy signal, $x(n)=h_s(n)*d(n)+v(n)$, in Eq. (2) may have the relation of $r_x(k)$ with $r_d(k)$ and $r_v(k)$ (the autocorrelations of $x(n)$, $d(n)$ and $v(n)$, respectively) as follows,

$$\boxed{r_x(k)=E\{x(n+k)x^*(n)\}=r_s(k)+r_v(k)=h_s(k)*h_s^*(-k)*r_d(k)+r_v(k)} \tag{4}$$

and the crosscorrelation between $d(n)$ and $x(n)$, $r_{dx}(k)$, given by

$$\boxed{r_{dx}(k)=E\{d(n+k)x^*(n)\}=E\{d(n+k)[s(n)+v(n)]^*\}=r_{ds}(k)=h_s^*(-k)*r_d(k)} \tag{5}$$

For the noisy signal of the form $x(n)=d(n)+v(n)$ in Eq. (1), a special case of Eq. (2) where $h_s(n)=\delta(n)$ and no distortion happens to $d(n)$ in its measurement, we have

$$\boxed{r_x(k)=r_d(k)+r_v(k)} \tag{6}$$

$$\boxed{r_{dx}(k) = r_d(k)} \tag{7}$$

The *z*-transforms of Eqs. (4) to (7) are given, respectively, by

$$\boxed{P_x(z) = H_s(z)H_s^*(1/z^*)P_d(z) + P_v(z)} \tag{8}$$

$$\boxed{P_{dx}(z) = H_s^*(1/z^*)P_d(z)} \tag{9}$$

$$\boxed{P_x(z) = P_d(z) + P_v(z)} \tag{10}$$

$$\boxed{P_{dx}(z) = P_d(z)} \tag{11}$$

The relations in Eqs. (4) to (11) will be useful below in the discussion of various Wiener filters.

## Signal to noise ratio (SNR):

To determine how large a desired signal is in a noisy measurement, the signal to noise ratio (SNR) is used that is defined as the ratio of the signal power $P_d$ to the noise power $P_v$,

$$SNR = 10\log_{10}\left(P_d / P_v\right) \text{ (dB)} \tag{12}$$

When $d(n)$ and $v(n)$ both are *zero mean* processes with variances $\sigma_d^2$ and $\sigma_v^2$, respectively, then $r_d(0) = \sigma_d^2 = E\left\{|d(n)|^2\right\}$ and $\sigma_v^2 = E\left\{|v(n)|^2\right\}$, and the SNR becomes

$$SNR = 10\log_{10}\left(\sigma_d^2 / \sigma_v^2\right) \text{ (dB)} \tag{13}$$

When $d(n) = A\sin(\omega n + \varphi)$ is a deterministic sinusoid and $v(n)$ is white nose, then $P_d = A^2/2$ and $P_v = \sigma_v^2$

$$SNR = 10\log_{10}\left(A^2/(2\sigma_v^2)\right) \text{ (dB)} \tag{14}$$

For digital signals given over interval $0 \le n \le N-1$, the SNR can be calculated by

$$SNR = 10\log_{10}\left(\sum_{n=0}^{N-1}|d(n)|^2 \Big/ \sum_{n=0}^{N-1}|v(n)|^2\right) \tag{15}$$

## The *q*-transform:

The *q*-transform is one that is based on a time shift operator, called *q*-operator, in the time domain. *q* is defined as the forward shift operator that yields

$$qy(n) = y(n+1), \tag{16}$$

and $q^{-1}$ is defined as the backward shift operator that makes

$$q^{-1}y(n) = y(n-1). \tag{17}$$

The *q*-transform applies to both deterministic and random processes whereas the *z*-transform only applies to deterministic signals. Since it can be directly used in the time domain, the *q*-transform is convenient to use to find a system function, which is equivalent to the *z*-transform. This is shown in the following example.

**Example. *q*-transform**

For a difference equation, $y(n)-0.5y(n-1)=x(n)+0.25x(n-1)$, we can write it in the *q*-transform as follows

$$y(n)(1-0.5q^{-1}) = x(n)(1+0.25q^{-1})$$

and dividing $(1-0.5q^{-1})$ on both sides of the above equation, we have

$$y(n) = \frac{1+0.25q^{-1}}{1-0.5q^{-1}}x(n)$$

Denoting $H(q^{-1}) = (1+0.25q^{-1})/(1-0.5q^{-1})$

we may express $y(n)$ as

$$y(n) = H(q^{-1})x(n)$$

which is equivalent to the convolution

$y(n) = h(n) * x(n)$.

As we can see, $H(q^{-1})$ and $H(z) = (1+0.25z^{-1})/(1-0.5z^{-1})$ are of the same form. Thus, in sequel we will use either $H(q^{-1})$ or $H(z)$ to express the system function, an equivalent transform of $h(n)$.

Noting that $(q^{-1})^* = q$ or $(q)^* = q^{-1}$, we can calculate the autocorrelation of $y(n)$ in the following manner

$$r_y(k) = E\{y(n+k)y^*(n)\} = E\left\{\frac{1+0.25q^{-1}}{1-0.5q^{-1}}x(n+k)\left[\frac{1+0.25q^{-1}}{1-0.5q^{-1}}x(n)\right]^*\right\}$$

$$= \frac{1+0.25q^{-1}}{1-0.5q^{-1}}\frac{1+0.25q}{1-0.5q}E\{x(n+k)x^*(n)\} = H(q^{-1})H^*(q)r_x(k)$$

where $H^*(q)$ is of the same form as $H^*(1/z^*) = (1+0.25z)/(1-0.5z)$, and which is equivalent to

$$r_y(k) = h(k) * h^*(-k) * r_x(k)$$

## Optimum filtering and the mean square error minimization:

Optimum filtering is to acquire the *best linear estimate* of a desired signal from a measurement. The main issues in optimal filtering contain

- filtering that deals with recovering a desired signal $d(n)$ from a noisy signal (or measurement) $x(n)$;
- prediction that is concerned with predicting a signal $d(n+m)$ for $m>0$ from observation $x(n)$;
- smoothing that is an a posteriori form of estimation, i.e., estimating $d(n+m)$ for $m<0$ from data $x(n)$;
- deconvolution that is to deal with finding the unit sample response (or a system function) of a LSI filter.

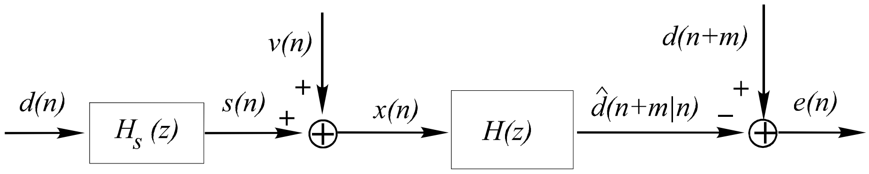



**Fig. 2.** General Wiener filtering problem

Let us look at an *optimum* LSI filter with a system function $H(z)$ or a unit sample response $h(n)$ (see Fig. 2). It is designed for the *best linear estimate* of a desired signal $d(n+m)$ from the measured signal (the input to the filter) $x(n) = h_s(n) * d(n) + v(n)$, that contains a desired signal $d(n)$ that is distorted due to $h_s(n)$, and measurement noise $v(n)$. That is, the output of the optimum filter is the best linear estimate of $d(n+m)$, denoted by $\hat{d}(n+m\,|\,n)$, which means the *best linear estimate* of $d(l)$ at time $l=n+m$ based on the values of *input signal* $x(l)$ up to time $l=n$, e.g., $x(0)$, $x(1)$, $x(2)$, …, $x(n)$.

It should be noted that the first index $n+m$ in $\hat{d}(n+m\,|\,n)$ (i.e., the index on the left side of vertical bar "|") is related to one in the estimated signal $d(n+m)$, and the second index $n$ (on the right side of "|") is related to one in the input signal $x(n)$ used for estimating $d(n+m)$. For $m=0$, we have $\hat{d}(n\,|\,n)$, simply denoted by $\hat{d}(n)$. Note that in the Hayes' textbook the notation $\hat{d}(n+m)$ (or $\hat{x}(n+m)$, …) is used instead of $\hat{d}(n+m\,|\,n)$ (or $\hat{x}(n+m\,|\,n)$, …).

An optimum filter for a best estimation is achieved at by minimizing <u>*the mean square error (MSE)*</u>

$$\boxed{\xi = E\{|e(n)|^2\}} \qquad (18)$$

where

$$\boxed{e(n) = d(n+m) - \hat{d}(n+m\,|\,n)} \qquad (19)$$

is the error between signal $d(n+m)$ and its estimate $\hat{d}(n+m\,|\,n)$.

The filter's output $\hat{d}(n+m\,|\,n)$ can be expressed as a convolution as follows

$$\hat{d}(n+m\,|\,n) = h(n) * x(n) \qquad (20)$$

Introducing the $q$-transform of $h(n)$, $\hat{d}(n+m\,|\,n)$ can be written as

$$\hat{d}(n+m\,|\,n) = H(q^{-1})x(n)\ . \qquad (21)$$

Issues in Wiener filtering:

Depending on what is $m$ in Fig. 2 and how desired signal $d(n)$ and measurement $x(n)$ that contains $d(n)$ plus noise $v(n)$ are related to each other, a number of important problems may be cast into a Wiener filtering framework.

1. $m$=0: it is the *filtering* problem. The filters used are causal and the goal is to estimate $d(n)$ from the current and past values of $x(n)$. The estimate of $d(n)$ is $\hat{d}(n)$.
2. $m$>0: it is the *prediction* problem. The filters are trying to predict (estimate) $d(n+m)$ using a linear combination of previous values of $x(n)$. The filters are causal. The estimate of $d(n+m)$ is denoted by $\hat{d}(n+m\,|\,n)$.
3. $m$<0: it is the *smoothing* problem that is the same as the filtering problem except estimating $d(n+m)$ is allowed to use noncausal filters, or using all $x(n)$, i.e., the past values for $n<n+m$, the current values for $n=n+m$, and the future values for $n>n+m$. The estimate of $d(n+m)$ is $\hat{d}(n+m\,|\,n)$.
4. When the signals $x(n)$ and $d(n)$ are related by $x(n) = d(n) * g(n) + v(n)$ where $g(n)$ is the unit sample response of a linear shift-invariant filter, estimating $g(n)$ is the *deconvolution* problem.

Types of Wiener filters – causal and noncausal FIR and IIR filters:

Consider an LSI Wiener filter with a system function $H(z) = \sum_{k=M}^{N} h(k)z^{-k}$, ($N$>$M$). Depending on the values of $N$ and $M$, the Wiener filters may be classified as

*causal FIR Wiener filters* for $M$=0, $N$>0, with $H(z) = \sum_{k=0}^{N} h(k)z^{-k}$; (22a)

*noncausal FIR Wiener filters* for $M$<0, $N$>0, with $H(z) = \sum_{k=M}^{N} h(k)z^{-k}$; (22b)

*causal IIR Wiener filters* for $M$=0, $N=\infty$, with $H(z) = \sum_{k=0}^{\infty} h(k)z^{-k}$; (22c)

*noncausal IIR Wiener filters* for $M$<0 (or $M=-\infty$), $N=\infty$, with $H(z) = \sum_{k=M}^{\infty} h(k)z^{-k}$. (22d)

One of the central issues in designing Wiener optimum filters is to find the coefficients of the filter that create the minimum mean square error $\xi = E\left\{|\,e(n)\,|^2\right\}$.

## *FIR WIENER FILTERS*

FIR Wiener filtering problems in a general way:

Let us first discuss a FIR Wiener filtering problem in a general way so that we can apply it to filtering, predicting and smoothing problems.

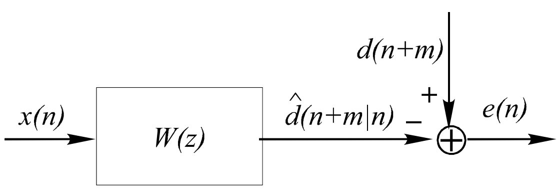



**Fig. 3.** FIR Wiener filtering problem in general.

Consider a noisy signal $x(n)$ that contains desired signal $d(n)$ plus noise $v(n)$. Here $x(n)$ and $d(n)$ are assumed to be jointly wide-sense stationary random processes with known autocorelations, $r_x(k)$ and $r_d(k)$, and known cross-correlation $r_{dx}(k)$. The specific relation between *x*(*n*) and *d*(*n*) is not specified yet at the moment. We design an FIR Wiener filter that provides the minimum mean-square error (MMSE) estimate of a desired signal *d*(*n*+*m*) from *x*(*n*), as shown in Fig. 3.

The system function of an FIR filter of (*p*–1)*th*-order is given by

$$W(z)=\sum_{k=0}^{p-1}w(k)z^{-k}\ . \tag{23}$$

The input to the filter is *x*(*n*). The output of the filter is $\hat{d}(n+m\,|\,n)$, estimate of *d*(*n*+*m*), given by

$$\hat{d}(n+m\,|\,n)=w(n)*x(n)=\sum_{k=0}^{p-1}w(k)x(n-k)\,, \tag{24}$$

which shows that the *p* values of *x*(*l*), from time *l*= *n*–*p*+1 up to time *l*=*n*, are used to estimate *d*(*n*+*m*). The mean square error to be minimized for finding the filter's coefficients is

$$\xi=E\{|\,e(n)\,|^2\} \tag{25}$$

where *e*(*n*) is the estimate error expressed as

$$e(n)=d(n+m)-\hat{d}(n+m\,|\,n)=d(n{+}m)-\sum_{k=0}^{p-1}w(k)x(n-k)\,. \tag{26}$$

The coefficients *w*(*n*) of the filter are determined by minimizing the MSE $\xi=E\{|\,e(n)\,|^2\}$, which is done by setting $\partial\xi\big/\partial w_p^*(k)=0$ for *k*=0, 1, …, *p*–1 (*k* starts from 0!) so that we have

$$\begin{aligned}\frac{\partial\xi}{\partial w^*(k)}&=\frac{\partial}{\partial w^*(k)}E\left\{|\,e(n)\,|^2\right\}=\frac{\partial}{\partial w^*(k)}E\left\{e(n)e^*(n)\right\}=E\left\{e(n)\frac{\partial e^*(n)}{\partial w^*(k)}\right\}\\&=E\left\{e(n)\frac{\partial}{\partial w^*(k)}\left[d^*(n+m)-\sum_{k=0}^{p-1}w^*(k)x^*(n-k)\right]\right\}=-E\left\{e(n)x^*(n-k)\right\}=0\end{aligned} \tag{27}$$

that is

$$\boxed{E\left\{e(n)x^*(n-k)\right\}=0;\quad k=0,1,...,p-1} \tag{28}$$

which is known as the *orthogonality principle* or the *projection theorem*.

Using Eq. (26) in Eq. (28) may give us

$$E\left\{\left[d(n+m)-\sum_{l=0}^{p-1}w(l)x(n-l)\right]x^*(n-k)\right\}=E\left\{d(n+m)x^*(n-k)\right\}-\sum_{l=0}^{p-1}w(l)E\left\{x(n-l)x^*(n-k)\right\}=0 \tag{29}$$

Since *x*(*n*) and *d*(*n*) are assumed to be jointly WSS processes, we have autocorrelation, $r_x(k-l)=E\left\{x(n-l)x^*(n-k)\right\}$, and cross-correlation $r_{dx}(k+m)=E\left\{d(n+m)x^*(n-k)\right\}$. Then, Eq. (29) reduces to

$$\boxed{\sum_{l=0}^{p-1}w(l)r_x(k-l)=r_{dx}(k+m);\quad k=0,1,\cdots,p-1} \tag{30}$$

which are known as the <u>*Wiener-Hopf equations*</u>, a system of *p* equations with *p* unknowns, *w*(*k*) for *k*=0, 1, …, *p*–1, so that the filter's coefficients can be determined. Eq. (30) can be expressed in the matrix form as

$$\boxed{\begin{bmatrix}r_x(0)&r_x^*(1)&\dots&r_x^*(p-1)\\r_x(1)&r_x(0)&\dots&r_x^*(p-2)\\\dots&\dots&\dots&\dots\\r_x(p-1)&r_x(p-2)&\dots&r_x(0)\end{bmatrix}\begin{bmatrix}w(0)\\w(1)\\\dots\\w(p-1)\end{bmatrix}=\begin{bmatrix}r_{dx}(m)\\r_{dx}(m+1)\\\dots\\r_{dx}(m+p-1)\end{bmatrix}} \tag{31}$$

in which the relation $r_x^*(k)=r_x(-k)$ is used since *x*(*n*) is WSS. Eq. (31) can be written in a compact form as

$$\mathbf{R}_x\mathbf{w}=\mathbf{r}_{dx} \tag{32}$$

where $\mathbf{R}_x$ is a $p\times p$ Hermitian Toeplitz matrix of autocorrelation, **w** is the vector of the filter coefficients, and $\mathbf{r}_{dx}$ is the vector of cross-correlation between $d(n+m)$ and $x(n)$. The matrix form is easy to implement in MATLAB.

Since the Wiener-Hopf equations are derived from the orthogonality principle by using the relations $E\{x(n-l)x^*(n-k)\}=r_x(k-l)$ and $E\{d(n+m)x^*(n-k)\}= r_{dx}(k+m)$ (see Eqs. (28)–(30)), then we may say that the Wiener-Hopf equations and the orthogonality principle are equivalent in designing optimum filters.

Applying the orthogonality principle (Eq. (28)) to calculating the MSE (Eq. (25)), we may have

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

$$= E\{e(n)d^*(n+m)\}-\sum_{k=0}^{p-1}w^*(k)E\{e(n)x^*(n-k)\}= E\{e(n)d^*(n+m)\}$$

$$= E\left\{\left[d(n+m)-\sum_{k=0}^{p-1}w(k)x(n-k)\right]d^*(n+m)\right\} \tag{33}$$

Since $E\{d(n+m)d^*(n+m)\}=r_x(0)$ and $E\{x(n-k)d^*(n+m)\}=\left[E\{d(n+m)x^*(n-k)\}\right]^*=r_{dx}^*(k+m)$, we have

$$\boxed{\xi_{\min} = r_d(0)-\sum_{k=0}^{p-1}w(k)r_{dx}^*(k+m)} \tag{34}$$

which is a reduced MSE, called *minimum mean-square error* (MMSE).

Note that Eq. (34) holds only if *all* the coefficients, $w(k)$ for $k$=0, 1, …, $p$–1, are determined by Wiener-Hopf equations (Eq. (30) or (31)), which can be seen in problem 7.4 in the Hayes' textbook.

If the measurement $x(n)$ is of the form $x(n)=h_s(n)*d(n)+v(n)$ where $d(n)$ and $v(n)$ are uncorrelated and their autocorrelations, $r_d(k)$ and $r_v(k)$, are known, then we can use Eq. (4) and (5) (namely, $r_x(k)=h_s(k)*h_s^*(-k)*r_d(k)+r_v(k)$ and $r_{dx}(k)=h_s^*(-k)*r_d(k)$) in Eqs. (30) and (34) to determine $w(k)$ and the MMSE $\xi_{\min}$. For $x(n)=d(n)+v(n)$, we should use $r_x(k)=r_d(k)+r_v(k)$ and $r_{dx}(k)=r_d(k)$ instead.

The results that we have obtained from the above study can be applied to filtering ($m$=0), predicting ($m$>0) and smoothing ($m$<0) problems.

***Summary***

To determine the coefficients $w(k)$ for an optimum filter, we first set $\partial\xi/\partial w_p^*(k)=0$ (optimization), then have $E\{e(n)x^*(n-k)\}=0$ (the orthogonality principle), and finally obtain $\sum_{l=0}^{p-1}w(l)r_x(k-l)=r_{dx}(k+m)$ (the Wiener-Hopf equations). The three-step procedure, i.e., from optimization to orthogonality principle, and then to Wiener-Hopf equations, results in an optimum Wiener filter, will be called here <u>*three-step optimization.*</u>

The orthogonality principle, $E\{e(n)x^*(n-k)\}=0$, results in a reduced MSE, the minimal MSE $\xi_{\min}=E\{e(n)d^*(n)\}=r_d(0)-\sum_{k=0}^{p-1}w(k)r_{dx}^*(k+m)$. In other words, a filter determined from Wiener-Hopf equations is an optimum filter in the MSE sense (with a reduced MSE).

By setting $m$ be different values, the Wiener-Hopf equations in Eq. (30) and the expression for $\xi_{\min}$ in Eq. (34) obtained here can be applied to filtering ($m$=0), predicting ($m$>0) and smoothing ($m$<0) problems.

Wiener filtering:

For Wiener filtering problem, all we need to do is to set $m$=0 in Eqs. (30) and (34), and then we have the Wiener-Hopf equations are

$$\boxed{\sum_{l=0}^{p-1} w(l) r_x(k-l) = r_{dx}(k); \quad k = 0, 1, ..., p-1}, \tag{35}$$

or in the matrix form

$$\boxed{\begin{bmatrix} r_x(0) & r_x^*(1) & ... & r_x^*(p-1) \\ r_x(1) & r_x(0) & ... & r_x^*(p-2) \\ ... & ... & ... & ... \\ r_x(p-1) & r_x(p-2) & ... & r_x(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ ... \\ w(p-1) \end{bmatrix} = \begin{bmatrix} r_{dx}(0) \\ r_{dx}(1) \\ ... \\ r_{dx}(p-1) \end{bmatrix}} \tag{36}$$

and the minimum mean-square error is

$$\xi_{\min} = E\{e(n) d^*(n)\} = r_d(0) - \sum_{k=0}^{p-1} w(k) r_{dx}^*(k) \tag{37}$$

**Example 1. Wiener Filtering** (Example 7.2.1 in the textbook)

**Example. FIR Wiener Filter for radio channel equalization**

Consider a communication signal $d(n)$ that is transferred using BPSK-modulation, in which only the symbols +1 and –1 are used. The signal $d(n)$ in this case can be considered as an uncorrelated, zero-mean random process, namely, its autocorrelation is $r_d(k) = \delta(k)$. Due to multi-path fading, the received signal through a radio channel is of the form

$$x(n) = \frac{1}{1+0.5q^{-1}} d(n) + v(n)$$

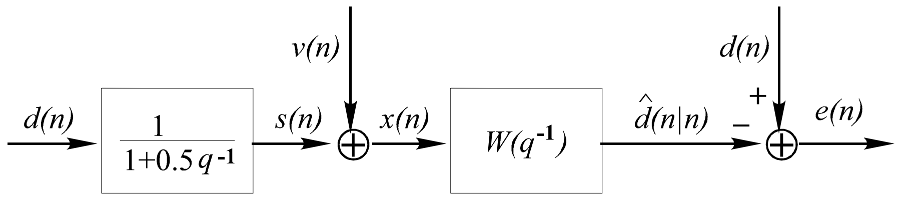
where $1/\left(1+0.5q^{-1}\right)$ is the channel effect, and $v(n)$ is a white noise process that has a zero mean and a variance $\sigma_v^2$ and is uncorrelated with $d(n)$.

Find an FIR Wiener filter of order one

$$W(q^{-1}) = w(0) + w(1)q^{-1}$$

that gives an optimum estimation of $d(n)$, i.e., $\hat{d}(n\,|\,n) = W(q^{-1})x(n)$.

*Solution*




The FIR Wiener filter can be found from the Wiener-Hopf equations

$$\begin{bmatrix} r_x(0) & r_x(1) \\ r_x(1) & r_x(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} r_{dx}(0) \\ r_{dx}(1) \end{bmatrix}$$

where the autocorrelation is $r_x(k) = r_s(k) + r_v(k)$ and the cross-correlation is

$$r_{dx}(k) = E\{d(n)x^*(n-k)\} = E\left\{ d(n)\left[ \frac{1}{1+0.5q^{-1}} d(n-k) + v(n-k) \right]^* \right\}$$

$$= E\left\{ d(n)\frac{1}{1+0.5q} d^*(n-k) \right\} = \frac{1}{1+0.5q} r_d(k) = \left[ \sum_{m=0}^{\infty} (-0.5q)^m \right] \delta(k) = \delta(k) - 0.5\delta(k+1) + 0.25\delta(k+2)\cdots$$

Thus, we have $r_{dx}(0) = 1$, and $r_{dx}(k) = 0$ for $k \ge 1$.

Since $P_x(z) = P_s(z) + P_v(z) = \dfrac{1}{\left(1+0.5z^{-1}\right)\left(1+0.5z\right)} + \sigma_v^2 = \dfrac{1+1.25\sigma_v^2 + 0.5\sigma_v^2 z^{-1} + 0.5\sigma_v^2 z}{\left(1+0.5z^{-1}\right)\left(1+0.5z\right)}$,

we have $r_x(k) = Z^{-1}\{P_x(z)\} = \dfrac{1}{1-0.5^2}\left[ \left(1+1.25\sigma_v^2\right)(-0.5)^{|k|} + 0.5\sigma_v^2(-0.5)^{|k-1|} + 0.5\sigma_v^2(-0.5)^{|k+1|} \right]$

$$r_x(0) = \frac{4}{3}\left(1+0.75\sigma_v^2\right) = \frac{4}{3} + \sigma_v^2, \text{ and } r_x(1) = \frac{4}{3}(-0.5) = -\frac{2}{3}$$

$$\begin{bmatrix} 4/3+\sigma_v^2 & -2/3 \\ -2/3 & 4/3+\sigma_v^2 \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} 4/3+\sigma_v^2 & -2/3 \\ -2/3 & 4/3+\sigma_v^2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

For $\sigma_v^2 = 0$, $\begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$, that is $W(q^{-1}) = 1+0.5q^{-1}$, and $\hat{d}(n\,|\,n) = d(n)$.

For $\sigma_v^2 = 2$, $\begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} 4/3+2 & -2/3 \\ -2/3 & 4/3+2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.3125 \\ 0.0625 \end{bmatrix}$, i.e., $W(q^{-1}) = 0.3125 + 0.0625q^{-1}$.

## Wiener prediction:

For the predicting problem, we can directly use all the results (i.e., the Wiener-Hopf equations in Eqs. (30)–(32) and the MMSE in Eq. (34)) obtained in the previous section on a FIR Wiener filtering problem in a general way, just by setting $m>0$. Here we discuss the prediction issue in the absence and presence of noise.

*A. Wiener prediction in the absence of noise:*

For the prediction in the absence of noise, the observation is noise-free so that $x(n)=h_s(n)*d(n)$. If we consider the case of $h_s(n)=\delta(n)$, i.e., $x(n)=d(n)$, then we have $r_x(k)=r_d(k)$ and $r_{dx}(k)=r_x(k)$. The $m$-step prediction of $x(n+m)$ is made from a linear combination of the current and previous values of $x(n)$ (Fig. 5).

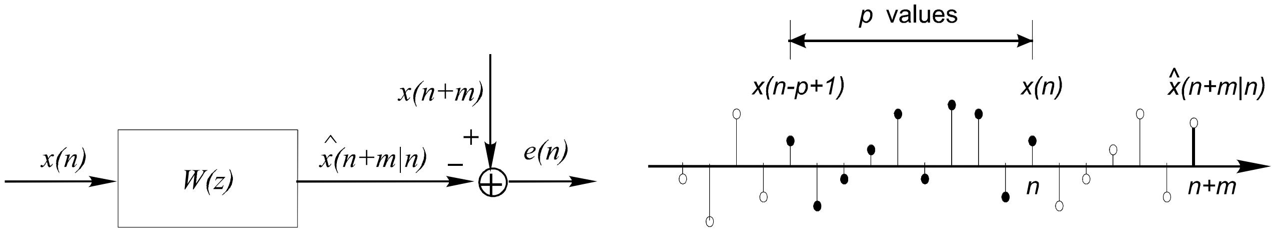



**Fig. 5.** The $m$-step prediction in the absence of noise.

The Wiener-Hopf equations for the $m$-step prediction in the absence of noise become

$$\boxed{\sum_{l=0}^{p-1} w(l) r_x(k-l) = r_x(k+m); \quad k=0,1,\ldots,p-1} \tag{38}$$

and in the matrix form

$$\begin{bmatrix} r_x(0) & r_x^*(1) & \cdots & r_x^*(p-1) \\ r_x(1) & r_x(0) & \cdots & r_x^*(p-2) \\ \cdots & \cdots & \cdots & \cdots \\ r_x(p-1) & r_x(p-2) & \cdots & r_x(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \\ \cdots \\ w(p-1) \end{bmatrix} = \begin{bmatrix} r_x(m) \\ r_x(m+1) \\ \cdots \\ r_x(m+p-1) \end{bmatrix} \tag{39}$$

The minimum mean-square error is

$$\boxed{\xi_{\min} = E\{e(n)x^*(n+m)\} = r_x(0) - \sum_{l=0}^{p-1} w(l) r_x^*(l+m)} \tag{40}$$

**Example 2. Linear Prediction in the absence of noise**

*B. Wiener prediction in the presence of noise:*

In the presence of noise $v(n)$, the $m$-step prediction of $d(n)$ is made from the noisy measurement $x(n) = h_s(n) * d(n) + v(n)$ (or $x(n) = d(n) + v(n)$ if $h_s(n) = \delta(n)$), and the FIR Wiener predictor for the prediction is determined from the Wiener-Hopf equations in Eqs. (30) or (31), and the MMSE is calculated using Eq. (34).

**Example 3. One-step linear Prediction in the presence of noise**

Design a FIR Wiener predictor of the first order

$$\hat{d}(n+1|n) = w(0)x(n) + w(1)x(n-1)$$

for one-step linear prediction of an AR(1) process $d(n)$ in a noise measurement as follows

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

where $d(n)$ is assumed to have $r_d(k) = \alpha^{|k|}$ ($|\alpha| < 1$), and $v(n)$ is zero mean white noise with a variance of $\sigma_v^2$ and uncorrelated with $d(n)$.

*Solution.*

The first-order FIR Wiener filter for the one-step prediction can be found from the Wiener-Hopf equations

$$\begin{bmatrix} r_x(0) & r_x(1) \\ r_x(1) & r_x(0) \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} r_{dx}(1) \\ r_{dx}(2) \end{bmatrix}$$

Since $d(n)$ is uncorrelated with $v(n)$, then $r_x(k) = \alpha^{|k|} + \sigma_v^2 \delta(k)$ and $r_{dx}(k) = r_d(k) = \alpha^{|k|}$. The Wiener-Hopf equations become

$$\begin{bmatrix} 1+\sigma_v^2 & \alpha \\ \alpha & 1+\sigma_v^2 \end{bmatrix} \begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \begin{bmatrix} \alpha \\ \alpha^2 \end{bmatrix}$$

Solving for $w(0)$ and $w(1)$ yields

$$\begin{bmatrix} w(0) \\ w(1) \end{bmatrix} = \frac{1}{(1+\sigma_v^2)^2 - \alpha^2} \begin{bmatrix} 1+\sigma_v^2 & -\alpha \\ -\alpha & 1+\sigma_v^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \alpha^2 \end{bmatrix} = \frac{\alpha}{(1+\sigma_v^2)^2 - \alpha^2} \begin{bmatrix} 1+\sigma_v^2 - \alpha^2 \\ \sigma_v^2 \alpha \end{bmatrix}$$

Therefore, the one-step predictor is

$$\hat{d}(n+1|n) = \frac{(1+\sigma_v^2 - \alpha^2)\alpha}{(1+\sigma_v^2)^2 - \alpha^2} x(n) + \frac{\sigma_v^2 \alpha^2}{(1+\sigma_v^2)^2 - \alpha^2} x(n-1)$$

As $\sigma_v^2 \to 0$, the predictor becomes the noise-free solution as in the previous example.

The minimum mean square error (MMSE) is

$$\xi_{\min} = r_d(0) - w(0) r_{dx}^*(1) - w(1) r_{dx}^*(2) = r_d(0) - w(0) r_d(1) - w(1) r_d(2) = 1 - \frac{(1+2\sigma_v^2 - \alpha^2)\alpha^2 + \sigma_v^2 \alpha^4}{(1+\sigma_v^2)^2 - \alpha^2}$$

The MMSE increases as $\sigma_v^2$ increases. In the limiting case that $\sigma_v^2 \to \infty$, $\xi_{\min} \to 1$.