6

WIENER FILTERS

Wiener theory, formulated by Norbert Wiener, forms the foundation of data-dependent linear least square error filters. Wiener filters play a central role in a wide range of applications such as linear prediction, echo cancellation, signal restoration, channel equalisation and system identification. The coefficients of a Wiener filter are calculated to minimise the average squared distance between the filter output and a desired signal. In its basic form, the Wiener theory assumes that the signals are stationary processes. However, if the filter coefficients are periodically recalculated for every block of N signal samples then the filter adapts itself to the average characteristics of the signals within the blocks and becomes block-adaptive. A block-adaptive (or segment adaptive) filter can be used for signals such as speech and image that may be considered almost stationary over a relatively small block of samples. In this chapter, we study Wiener filter theory, and consider alternative methods of formulation of the Wiener filter problem. We consider the application of Wiener filters in channel equalisation, time-delay estimation and additive noise reduction. A case study of the frequency response of a Wiener filter, for additive noise reduction, provides useful insight into the operation of the filter. We also deal with some implementation issues of Wiener filters.

6.1 Wiener Filters: Least Square Error Estimation

Wiener formulated the continuous-time, least mean square error, estimation problem in his classic work on interpolation, extrapolation and smoothing of time series (Wiener 1949). The extension of the Wiener theory from continuous time to discrete time is simple, and of more practical use for implementation on digital signal processors. A Wiener filter can be an infinite-duration impulse response (IIR) filter or a finite-duration impulse response (FIR) filter. In general, the formulation of an IIR Wiener filter results in a set of non-linear equations, whereas the formulation of an FIR Wiener filter results in a set of linear equations and has a closed-form solution. In this chapter, we consider FIR Wiener filters, since they are relatively simple to compute, inherently stable and more practical. The main drawback of FIR filters compared with IIR filters is that they may need a large number of coefficients to approximate a desired response.

Figure 6.1 illustrates a Wiener filter represented by the coefficient vector $\boldsymbol{w}$. The filter takes as the input a signal $y(m)$, and produces an output signal $\hat{x}(m)$, where $\hat{x}(m)$ is the least mean square error estimate of a desired or target signal $x(m)$. The filter input–output relation is given by

$$\hat{x}(m) = \sum_{k=0}^{P-1} w_k\, y(m-k)$$
$$= \boldsymbol{w}^{\mathrm{T}} \boldsymbol{y}$$

(6.1)

where m is the discrete-time index, $\boldsymbol{y}^{\mathrm{T}}=[y(m),\ y(m-1),\ ...,\ y(m-P-1)]$ is the filter input signal, and the parameter vector $\boldsymbol{w}^{\mathrm{T}}=[w_0,\ w_1,\ ...,\ w_{P-1}]$ is the Wiener filter coefficient vector. In Equation (6.1), the filtering operation is expressed in two alternative and equivalent forms of a convolutional sum and an inner vector product. The Wiener filter error signal, $e(m)$ is defined as the difference between the desired signal $x(m)$ and the filter output signal $\hat{x}(m)$:

$$e(m) = x(m) - \hat{x}(m)$$
$$= x(m) - \boldsymbol{w}^{\mathrm{T}} \boldsymbol{y}$$

(6.2)

In Equation (6.2), for a given input signal $y(m)$ and a desired signal $x(m)$, the filter error $e(m)$ depends on the filter coefficient vector $\boldsymbol{w}$.
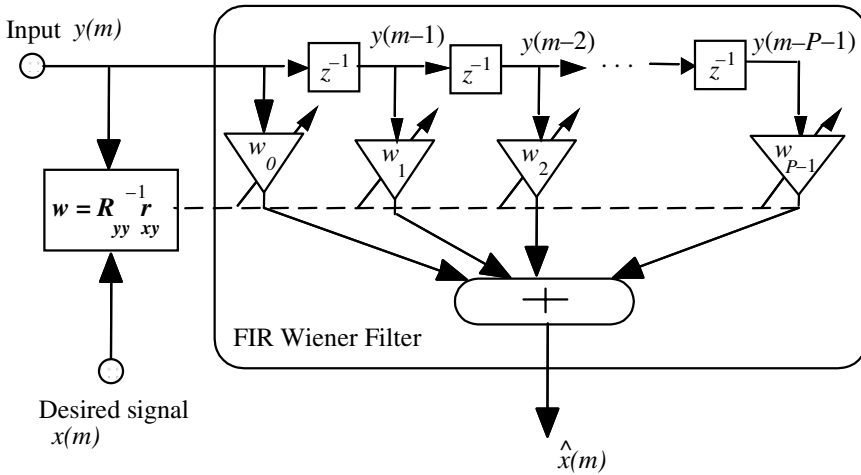
Figure 6.1 Illustration of a Wiener filter structure.

To explore the relation between the filter coefficient vector w and the error signal $e(m)$ we expand Equation (6.2) for N samples of the signals $x(m)$ and $y(m)$:

$$
\begin{pmatrix} e(0) \\ e(1) \\ e(2) \\ \vdots \\ e(N-1) \end{pmatrix} = \begin{pmatrix} x(0) \\ x(1) \\ x(2) \\ \vdots \\ x(N-1) \end{pmatrix} - \begin{pmatrix} y(0) & y(-1) & y(-2) & \dots & y(1-P) \\ y(1) & y(0) & y(-1) & \dots & y(2-P) \\ y(2) & y(1) & y(0) & \dots & y(3-P) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y(N-1) & y(N-2) & y(N-3) & \dots & y(N-P) \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{P-1} \end{pmatrix}
$$

$$(6.3)$$

In a compact vector notation this matrix equation may be written as

$$e = x - Yw \qquad (6.4)$$

where e is the error vector, x is the desired signal vector, Y is the input signal matrix and $Yw = \hat{x}$ is the Wiener filter output signal vector. It is assumed that the P initial input signal samples $[y(-1), \dots, y(-P-1)]$ are either known or set to zero.

In Equation (6.3), if the number of signal samples is equal to the number of filter coefficients $N=P$, then we have a square matrix equation, and there is a unique filter solution w, with a zero estimation error $e=0$, such

that $\hat{x} = Yw = x$. If $N < P$ then the number of signal samples N is insufficient to obtain a unique solution for the filter coefficients, in this case there are an infinite number of solutions with zero estimation error, and the matrix equation is said to be *underdetermined*. In practice, the number of signal samples is much larger than the filter length $N>P$; in this case, the matrix equation is said to be *overdetermined* and has a unique solution, usually with a non-zero error. When $N>P$, the filter coefficients are calculated to minimise an average error cost function, such as the average absolute value of error $\mathcal{E}[|e(m)|]$, or the mean square error $\mathcal{E}[e^2(m)]$, where $\mathcal{E}[.]$ is the expectation operator. The choice of the error function affects the optimality and the computational complexity of the solution.

   In Wiener theory, the objective criterion is the least mean square error (LSE) between the filter output and the desired signal. The least square error criterion is optimal for Gaussian distributed signals. As shown in the followings, for FIR filters the LSE criterion leads to a linear and closed-form solution. The Wiener filter coefficients are obtained by minimising an average squared error function $\mathcal{E}[e^2(m)]$ with respect to the filter coefficient vector w. From Equation (6.2), the mean square estimation error is given by

$$
\begin{aligned}
\mathcal{E}[e^2(m)] &= \mathcal{E}[(x(m) - w^{\mathrm{T}}y)^2] \\
&= \mathcal{E}[x^2(m)] - 2w^{\mathrm{T}}\mathcal{E}[yx(m)] + w^{\mathrm{T}}\mathcal{E}[yy^{\mathrm{T}}]w \\
&= r_{xx}(0) - 2w^{\mathrm{T}}r_{yx} + w^{\mathrm{T}}R_{yy}w
\end{aligned}
\tag{6.5}
$$

where $R_{yy} = \mathcal{E}[y(m)y^{\mathrm{T}}(m)]$ is the autocorrelation matrix of the input signal and $r_{xy} = \mathcal{E}[x(m)y(m)]$ is the cross-correlation vector of the input and the desired signals. An expanded form of Equation (6.5) can be obtained as

$$
\mathcal{E}[e^2(m)] = r_{xx}(0) - 2\sum_{k=0}^{P-1} w_k \, r_{yx}(k) + \sum_{k=0}^{P-1} w_k \sum_{j=0}^{P-1} w_j \, r_{yy}(k-j)
\tag{6.6}
$$

where $r_{yy}(k)$ and $r_{yx}(k)$ are the elements of the autocorrelation matrix R_{yy} and the cross-correlation vector r_{xy} respectively. From Equation (6.5), the mean square error for an FIR filter is a quadratic function of the filter coefficient vector w and has a single minimum point. For example, for a filter with only two coefficients (w_0, w_1), the mean square error function is a
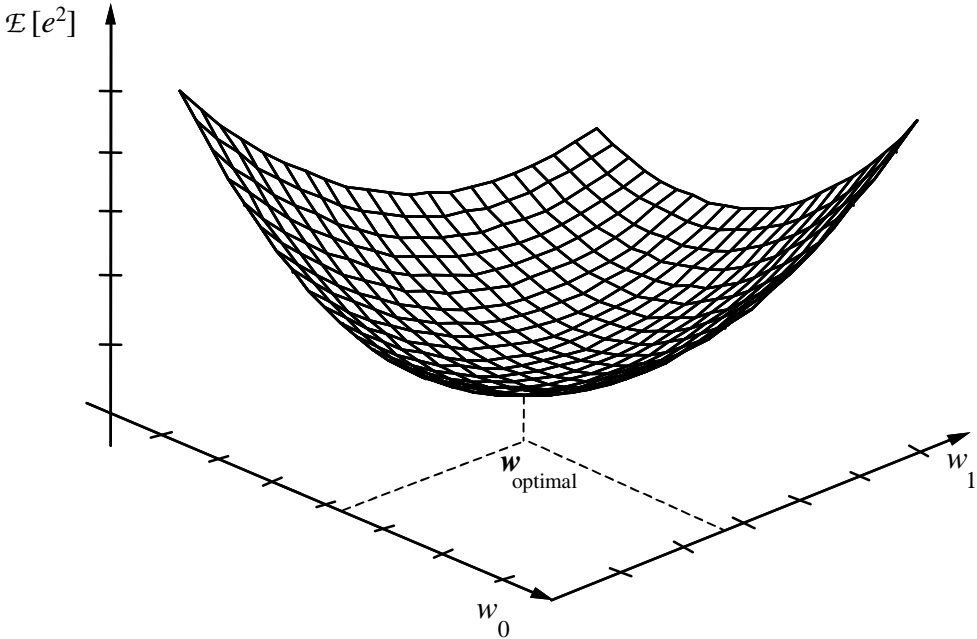
$\mathcal{E}[e^2]$

Figure 6.2 Mean square error surface for a two-tap FIR filter.

bowl-shaped surface, with a single minimum point, as illustrated in Figure 6.2. The least mean square error point corresponds to the minimum error power. At this optimal operating point the mean square error surface has zero gradient. From Equation (6.5), the gradient of the mean square error function with respect to the filter coefficient vector is given by

$$\frac{\partial}{\partial \mathbf{w}} \mathcal{E}[e^2(m)] = -2\mathcal{E}[x(m)\,\mathbf{y}(m)] + 2\mathbf{w}^{\mathrm{T}} \mathcal{E}[\mathbf{y}(m)\,\mathbf{y}^{\mathrm{T}}(m)]$$

$$= -2\mathbf{r}_{yx} + 2\mathbf{w}^{\mathrm{T}} \mathbf{R}_{yy} \tag{6.7}$$

where the gradient vector is defined as

$$\frac{\partial}{\partial \mathbf{w}} = \left[\frac{\partial}{\partial w_0}, \frac{\partial}{\partial w_1}, \frac{\partial}{\partial w_2}, \dots, \frac{\partial}{\partial w_{P-1}} \right]^{\mathrm{T}} \tag{6.8}$$

The minimum mean square error Wiener filter is obtained by setting Equation (6.7) to zero:

$$\mathbf{R}_{yy}\,\mathbf{w} = \mathbf{r}_{yx} \tag{6.9}$$

or, equivalently,

$$w = R_{yy}^{-1} \, r_{yx} \tag{6.10}$$

In an expanded form, the Wiener filter solution Equation (6.10) can be written as

$$
\begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{P-1} \end{pmatrix}
=
\begin{pmatrix}
r_{yy}(0) & r_{yy}(1) & r_{yy}(2) & \ldots & r_{yy}(P-1) \\
r_{yy}(1) & r_{yy}(0) & r_{yy}(1) & \ldots & r_{yy}(P-2) \\
r_{yy}(2) & r_{yy}(1) & r_{yy}(0) & \ldots & r_{yy}(P-3) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
r_{yy}(P-1) & r_{yy}(P-2) & r_{yy}(P-3) & \ldots & r_{yy}(0)
\end{pmatrix}^{-1}
\begin{pmatrix} r_{yx}(0) \\ r_{yx}(1) \\ r_{yx}(2) \\ \vdots \\ r_{yx}(P-1) \end{pmatrix}
$$

$$\tag{6.11}$$

From Equation (6.11), the calculation of the Wiener filter coefficients requires the autocorrelation matrix of the input signal and the cross-correlation vector of the input and the desired signals.

In statistical signal processing theory, the correlation values of a random process are obtained as the averages taken across the ensemble of different realisations of the process as described in Chapter 3. However in many practical situations there are only one or two finite-duration realisations of the signals $x(m)$ and $y(m)$. In such cases, assuming the signals are correlation-ergodic, we can use time averages instead of ensemble averages. For a signal record of length N samples, the time-averaged correlation values are computed as

$$r_{yy}(k) = \frac{1}{N} \sum_{m=0}^{N-1} y(m) \, y(m+k) \tag{6.12}$$

Note from Equation (6.11) that the autocorrelation matrix R_{yy} has a highly regular Toeplitz structure. A Toeplitz matrix has constant elements along the left–right diagonals of the matrix. Furthermore, the correlation matrix is also symmetric about the main diagonal elements. There are a number of efficient methods for solving the linear matrix Equation (6.11), including the Cholesky decomposition, the singular value decomposition and the QR decomposition methods.

6.2 Block-Data Formulation of the Wiener Filter

In this section we consider an alternative formulation of a Wiener filter for a block of N samples of the input signal $[y(0), y(1), ..., y(N–1)]$ and the desired signal $[x(0), x(1), ..., x(N–1)]$. The set of N linear equations describing the Wiener filter input/output relation can be written in matrix form as

$$
\begin{pmatrix} \hat{x}(0) \\ \hat{x}(1) \\ \hat{x}(2) \\ \vdots \\ \hat{x}(N-2) \\ \hat{x}(N-1) \end{pmatrix} = \begin{pmatrix} y(0) & y(-1) & y(-2) & \dots & y(2-P) & y(1-P) \\ y(1) & y(0) & y(-1) & \dots & y(3-P) & y(2-P) \\ y(2) & y(1) & y(0) & \dots & y(4-P) & y(3-P) \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ y(N-2) & y(N-3) & y(N-4) & \dots & y(N-P) & y(N-1-P) \\ y(N-1) & y(N-2) & y(N-3) & \dots & y(N+1-P) & y(N-P) \end{pmatrix} \begin{pmatrix} w_0 \\ w_1 \\ w_2 \\ \vdots \\ w_{P-2} \\ w_{P-1} \end{pmatrix}
$$

$$(6.13)$$

Equation (6.13) can be rewritten in compact matrix notation as

$$\hat{x} = Yw \tag{6.14}$$

The Wiener filter error is the difference between the desired signal and the filter output defined as

$$
\begin{aligned}
e &= x - \hat{x} \\
&= x - Yw
\end{aligned}
\tag{6.15}
$$

The energy of the error vector, that is the sum of the squared elements of the error vector, is given by the inner vector product as

$$
\begin{aligned}
e^{\mathrm{T}}e &= (x-Yw)^{\mathrm{T}}(x-Yw) \\
&= x^{\mathrm{T}}x - x^{\mathrm{T}}Yw - w^{\mathrm{T}}Y^{\mathrm{T}}x + w^{\mathrm{T}}Y^{\mathrm{T}}Yw
\end{aligned}
\tag{6.16}
$$

The gradient of the squared error function with respect to the Wiener filter coefficients is obtained by differentiating Equation (6.16):

$$\frac{\partial\, e^{\mathrm{T}}e}{\partial\, w} = -2x^{\mathrm{T}}Y + 2w^{\mathrm{T}}Y^{\mathrm{T}}Y \tag{6.17}$$

The Wiener filter coefficients are obtained by setting the gradient of the squared error function of Equation (6.17) to zero, this yields

$$\left(Y^{\mathrm{T}}Y\right)w = Y^{\mathrm{T}}x \qquad (6.18)$$

or

$$w = \left(Y^{\mathrm{T}}Y\right)^{-1}Y^{\mathrm{T}}x \qquad (6.19)$$

Note that the matrix $Y^{\mathrm{T}}Y$ is a time-averaged estimate of the autocorrelation matrix of the filter input signal R_{yy}, and that the vector $Y^{\mathrm{T}}x$ is a time-averaged estimate of r_{xy} the cross-correlation vector of the input and the desired signals. Theoretically, the Wiener filter is obtained from minimisation of the squared error across the ensemble of different realisations of a process as described in the previous section. For a correlation-ergodic process, as the signal length N approaches infinity the block-data Wiener filter of Equation (6.19) approaches the Wiener filter of Equation (6.10):

$$\lim_{N \to \infty}\left[w = \left(Y^{\mathrm{T}}Y\right)^{-1}Y^{\mathrm{T}}x\right] = R_{yy}^{-1}r_{xy} \qquad (6.20)$$

Since the least square error method described in this section requires a block of N samples of the input and the desired signals, it is also referred to as the block least square (BLS) error estimation method. The block estimation method is appropriate for processing of signals that can be considered as time-invariant over the duration of the block.

6.2.1 QR Decomposition of the Least Square Error Equation

An efficient and robust method for solving the least square error Equation (6.19) is the QR decomposition (QRD) method. In this method, the $N \times P$ signal matrix Y is decomposed into the product of an $N \times N$ orthonormal matrix Q and a $P \times P$ upper-triangular matrix $\mathcal{R}$ as

$$QY = \begin{pmatrix} \mathcal{R} \\ \mathbf{0} \end{pmatrix} \qquad (6.21)$$

where $\mathbf{0}$ is the $(N-P) \times P$ null matrix, $\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{Q} = \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}} = \mathbf{I}$, and the upper-triangular matrix $\mathcal{R}$ is of the form

$$\mathcal{R} = \begin{pmatrix} r_{00} & r_{01} & r_{02} & r_{03} & \cdots & r_{0P-1} \\ 0 & r_{11} & r_{12} & r_{13} & \cdots & r_{1P-1} \\ 0 & 0 & r_{22} & r_{23} & \cdots & r_{2P-1} \\ 0 & 0 & 0 & r_{33} & \cdots & r_{3P-1} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & r_{P-1P-1} \end{pmatrix} \tag{6.22}$$

Substitution of Equation (6.21) in Equation (6.18) yields

$$\begin{pmatrix} \mathcal{R} \\ \mathbf{0} \end{pmatrix}^{\mathrm{T}} \boldsymbol{Q}\boldsymbol{Q}^{\mathrm{T}} \begin{pmatrix} \mathcal{R} \\ \mathbf{0} \end{pmatrix} w = \begin{pmatrix} \mathcal{R} \\ \mathbf{0} \end{pmatrix}^{\mathrm{T}} \boldsymbol{Q}x \tag{6.23}$$

From Equation (6.23) we have

$$\begin{pmatrix} \mathcal{R} \\ \mathbf{0} \end{pmatrix} w = \boldsymbol{Q}x \tag{6.24}$$

From Equation (6.24) we have

$$\mathcal{R}w = x_Q \tag{6.25}$$

where the vector x_Q on the right hand side of Equation (6.25) is composed of the first P elements of the product $\boldsymbol{Q}x$. Since the matrix $\mathcal{R}$ is upper-triangular, the coefficients of the least square error filter can be obtained easily through a process of back substitution from Equation (6.25), starting with the coefficient $w_{P-1} = x_Q(P-1)/r_{P-1P-1}$.

The main computational steps in the QR decomposition are the determination of the orthonormal matrix $\boldsymbol{Q}$ and of the upper triangular matrix $\mathcal{R}$. The decomposition of a matrix into QR matrices can be achieved using a number of methods, including the Gram-Schmidt orthogonalisation method, the Householder method and the Givens rotation method.
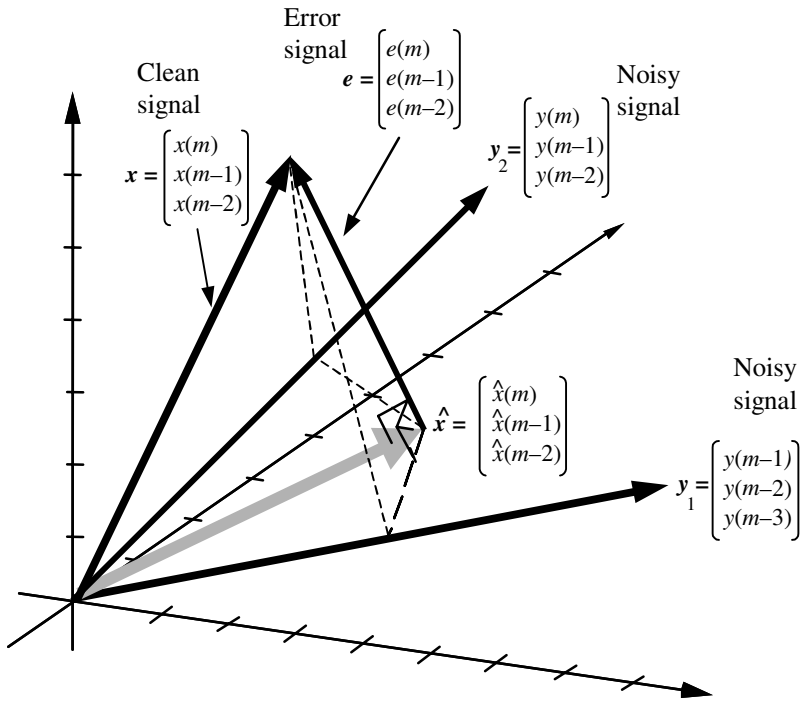
Figure 6.3 The least square error projection of a desired signal vector x onto a plane containing the input signal vectors y_1 and y_2 is the perpendicular projection of x shown as the shaded vector.

6.3 Interpretation of Wiener Filters as Projection in Vector Space

In this section, we consider an alternative formulation of Wiener filters where the least square error estimate is visualized as the perpendicular minimum distance *projection* of the desired signal vector onto the vector space of the input signal. A vector space is the collection of an infinite number of vectors that can be obtained from linear combinations of a number of independent vectors.

In order to develop a vector space interpretation of the least square error estimation problem, we rewrite the matrix Equation (6.11) and express the filter output vector $\hat{x}$ as a linear weighted combination of the column vectors of the input signal matrix as

$$\begin{pmatrix} \hat{x}(0) \\ \hat{x}(1) \\ \hat{x}(2) \\ \vdots \\ \hat{x}(N-2) \\ \hat{x}(N-1) \end{pmatrix} = w_0 \begin{pmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \\ y(N-2) \\ y(N-1) \end{pmatrix} + w_1 \begin{pmatrix} y(-1) \\ y(0) \\ y(1) \\ \vdots \\ y(N-3) \\ y(N-2) \end{pmatrix} + \cdots + w_{P-1} \begin{pmatrix} y(1-P) \\ y(2-P) \\ y(3-P) \\ \vdots \\ y(N-1-P) \\ y(N-P) \end{pmatrix}$$

$$(6.26)$$

In compact notation, Equation (6.26) may be written as

$$\hat{x} = w_0 y_0 + w_1 y_1 + \cdots + w_{P-1} y_{P-1} \qquad (6.27)$$

In Equation (6.27) the signal estimate $\hat{x}$ is a linear combination of P basis vectors $[y_0, y_1, \ldots, y_{P-1}]$, and hence it can be said that the estimate $\hat{x}$ is in the vector subspace formed by the input signal vectors $[y_0, y_1, \ldots, y_{P-1}]$.

In general, the P N-dimensional input signal vectors $[y_0, y_1, \ldots, y_{P-1}]$ in Equation (6.27) define the *basis* vectors for a subspace in an N-dimensional signal space. If P, the number of basis vectors, is equal to N, the vector dimension, then the subspace defined by the input signal vectors encompasses the entire N-dimensional signal space and includes the desired signal vector x. In this case, the signal estimate $\hat{x}=x$ and the estimation error is zero. However, in practice, $N>P$, and the signal space defined by the P input signal vectors of Equation (6.27) is only a subspace of the N-dimensional signal space. In this case, the estimation error is zero only if the desired signal x happens to be in the subspace of the input signal, otherwise the best estimate of x is the perpendicular projection of the vector x onto the vector space of the input signal $[y_0, y_1, \ldots, y_{P-1}]$., as explained in the following example.

Example 6.1 Figure 6.3 illustrates a vector space interpretation of a simple least square error estimation problem, where $y^T=[y(2), y(1), y(0), y(-1)]$ is the input signal, $x^T=[x(2), x(1), x(0)]$ is the desired signal and $w^T=[w_0, w_1]$ is the filter coefficient vector. As in Equation (6.26), the filter output can be written as

$$\begin{pmatrix} \hat{x}(2) \\ \hat{x}(1) \\ \hat{x}(0) \end{pmatrix} = w_0 \begin{pmatrix} y(2) \\ y(1) \\ y(0) \end{pmatrix} + w_1 \begin{pmatrix} y(1) \\ y(0) \\ y(-1) \end{pmatrix} \tag{6.28}$$

In Equation (6.28), the input signal vectors $y_1^T = [y(2), y(1), y(0)]$ and $y_2^T = [y(1), y(0), y(-1)]$ are 3-dimensional vectors. The subspace defined by the linear combinations of the two input vectors $[y_1, y_2]$ is a 2-dimensional plane in a 3-dimensional signal space. The filter output is a linear combination of y_1 and y_2, and hence it is confined to the plane containing these two vectors. The least square error estimate of x is the orthogonal projection of x on the plane of $[y_1, y_2]$ as shown by the shaded vector $\hat{x}$. If the desired vector happens to be in the plane defined by the vectors y_1 and y_2 then the estimation error will be zero, otherwise the estimation error will be the perpendicular distance of x from the plane containing y_1 and y_2.

6.4 Analysis of the Least Mean Square Error Signal

The optimality criterion in the formulation of the Wiener filter is the least mean square distance between the filter output and the desired signal. In this section, the variance of the filter error signal is analysed. Substituting the Wiener equation $R_{yy}w = r_{yx}$ in Equation (6.5) gives the least mean square error:

$$\mathcal{E}[e^2(m)] = r_{xx}(0) - w^T r_{yx}$$
$$= r_{xx}(0) - w^T R_{yy} w \tag{6.29}$$

Now, for zero-mean signals, it is easy to show that in Equation (6.29) the term $w^T R_{yy} w$ is the variance of the Wiener filter output $\hat{x}(m)$:

$$\sigma_{\hat{x}}^2 = \mathcal{E}[\hat{x}^2(m)] = w^T R_{yy} w \tag{6.30}$$

Therefore Equation (6.29) may be written as

$$\sigma_e^2 = \sigma_x^2 - \sigma_{\hat{x}}^2 \tag{6.31}$$

where $\sigma_x^2 = \mathcal{E}[x^2(m)], \sigma_{\hat{x}}^2 = \mathcal{E}[\hat{x}^2(m)]$ and $\sigma_e^2 = \mathcal{E}[e^2(m)]$ are the variances of the desired signal, the filter estimate of the desired signal and the error signal respectively. In general, the filter input $y(m)$ is composed of a signal component $x_c(m)$ and a random noise $n(m)$:

$$y(m) = x_c(m) + n(m) \tag{6.32}$$

where the signal $x_c(m)$ is the part of the observation that is correlated with the desired signal $x(m)$, and it is this part of the input signal that may be transformable through a Wiener filter to the desired signal. Using Equation (6.32) the Wiener filter error may be decomposed into two distinct components:

$$
\begin{aligned}
e(m) &= x(m) - \sum_{k=0}^{P} w_k\, y(m-k) \\
&= \left[x(m) - \sum_{k=0}^{P} w_k\, x_c(m-k) \right] - \sum_{k=0}^{P} w_k\, n(m-k)
\end{aligned} \tag{6.33}
$$

or

$$e(m) = e_x(m) + e_n(m) \tag{6.34}$$

where $e_x(m)$ is the difference between the desired signal $x(m)$ and the output of the filter in response to the input signal component $x_c(m)$, i.e.

$$e_x(m) = x(m) - \sum_{k=0}^{P-1} w_k\, x_c(m-k) \tag{6.35}$$

and $e_n(m)$ is the error in the output due to the presence of noise $n(m)$ in the input signal:

$$e_n(m) = -\sum_{k=0}^{P-1} w_k\, n(m-k) \tag{6.36}$$

The variance of filter error can be rewritten as

$$\sigma_e^2 = \sigma_{e_x}^2 + \sigma_{e_n}^2 \tag{6.37}$$

Note that in Equation (6.34), $e_x(m)$ is that part of the signal that cannot be recovered by the Wiener filter, and represents distortion in the signal output, and $e_n(m)$ is that part of the noise that cannot be blocked by the Wiener filter. Ideally, $e_x(m)=0$ and $e_n(m)=0$, but this ideal situation is possible only if the following conditions are satisfied:

(a) The spectra of the signal and the noise are separable by a linear filter.
(b) The signal component of the input, that is $x_c(m)$, is *linearly* transformable to $x(m)$.
(c) The filter length P is sufficiently large. The issue of signal and noise separability is addressed in Section 6.6.

6.5 Formulation of Wiener Filters in the Frequency Domain

In the frequency domain, the Wiener filter output $\hat{X}(f)$ is the product of the input signal $Y(f)$ and the filter frequency response $W(f)$:

$$\hat{X}(f)=W(f)Y(f) \tag{6.38}$$

The estimation error signal $E(f)$ is defined as the difference between the desired signal $X(f)$ and the filter output $\hat{X}(f)$,

$$\begin{aligned}E(f)&=X(f)-\hat{X}(f)\\&=X(f)-W(f)Y(f)\end{aligned} \tag{6.39}$$

and the mean square error at a frequency f is given by

$$\mathcal{E}\left[\left|E(f)\right|^2\right]=\mathcal{E}\left[\left(X(f)-W(f)Y(f)\right)^*\left(X(f)-W(f)Y(f)\right)\right] \tag{6.40}$$

where $\mathcal{E}[\cdot]$ is the expectation function, and the symbol $*$ denotes the complex conjugate. Note from Parseval's theorem that the mean square error in time and frequency domains are related by

$$\sum_{m=0}^{N-1} e^2(m) = \int_{-1/2}^{1/2} |E(f)|^2 \, df \tag{6.41}$$

To obtain the least mean square error filter we set the complex derivative of Equation (6.40) with respect to filter $W(f)$ to zero

$$\frac{\partial \, \mathcal{E}[|\,E(f)\,|^2]}{\partial W(f)} = 2W(f)P_{YY}(f) - 2P_{XY}(f) = 0 \tag{6.42}$$

where $P_{YY}(f) = \mathcal{E}[Y(f)Y^*(f)]$ and $P_{XY}(f) = \mathcal{E}[X(f)Y^*(f)]$ are the power spectrum of $Y(f)$, and the cross-power spectrum of $Y(f)$ and $X(f)$ respectively. From Equation (6.42), the least mean square error Wiener filter in the frequency domain is given as

$$W(f) = \frac{P_{XY}(f)}{P_{YY}(f)} \tag{6.43}$$

Alternatively, the frequency-domain Wiener filter Equation (6.43) can be obtained from the Fourier transform of the time-domain Wiener Equation (6.9):

$$\sum_m \sum_{k=0}^{P-1} w_k r_{yy}(m-k)e^{-j\omega m} = \sum_m r_{yx}(n)e^{-j\omega m} \tag{6.44}$$

From the Wiener–Khinchine relation, the correlation and power-spectral functions are Fourier transform pairs. Using this relation, and the Fourier transform property that convolution in time is equivalent to multiplication in frequency, it is easy to show that the Wiener filter is given by Equation (6.43).

6.6 Some Applications of Wiener Filters

In this section, we consider some applications of the Wiener filter in reducing broadband additive noise, in time-alignment of signals in multi-channel or multisensor systems, and in channel equalisation.
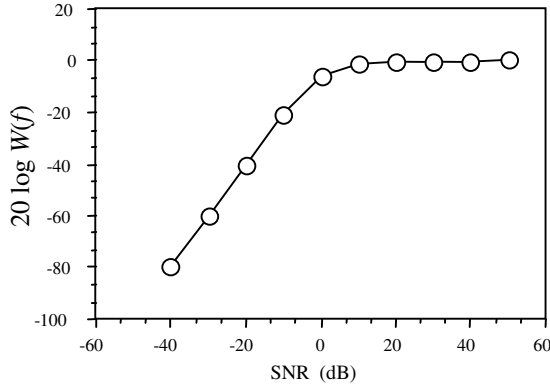
Figure 6.4 Variation of the gain of Wiener filter frequency response with SNR.

6.6.1 Wiener Filter for Additive Noise Reduction

Consider a signal $x(m)$ observed in a broadband additive noise $n(m)$., and model as

$$y(m) = x(m) + n(m) \tag{6.45}$$

Assuming that the signal and the noise are uncorrelated, it follows that the autocorrelation matrix of the noisy signal is the sum of the autocorrelation matrix of the signal $x(m)$ and the noise $n(m)$:

$$\boldsymbol{R_{yy}} = \boldsymbol{R_{xx}} + \boldsymbol{R_{nn}} \tag{6.46}$$

and we can also write

$$\boldsymbol{r_{xy}} = \boldsymbol{r_{xx}} \tag{6.47}$$

where $\boldsymbol{R_{yy}}$, $\boldsymbol{R_{xx}}$ and $\boldsymbol{R_{nn}}$ are the autocorrelation matrices of the noisy signal, the noise-free signal and the noise respectively, and $\boldsymbol{r_{xy}}$ is the cross-correlation vector of the noisy signal and the noise-free signal. Substitution of Equations (6.46) and (6.47) in the Wiener filter, Equation (6.10), yields

$$\boldsymbol{w} = (\boldsymbol{R_{xx}} + \boldsymbol{R_{nn}})^{-1} \boldsymbol{r_{xx}} \tag{6.48}$$

Equation (6.48) is the optimal linear filter for the removal of additive noise. In the following, a study of the frequency response of the Wiener filter provides useful insight into the operation of the Wiener filter. In the frequency domain, the noisy signal $Y(f)$ is given by
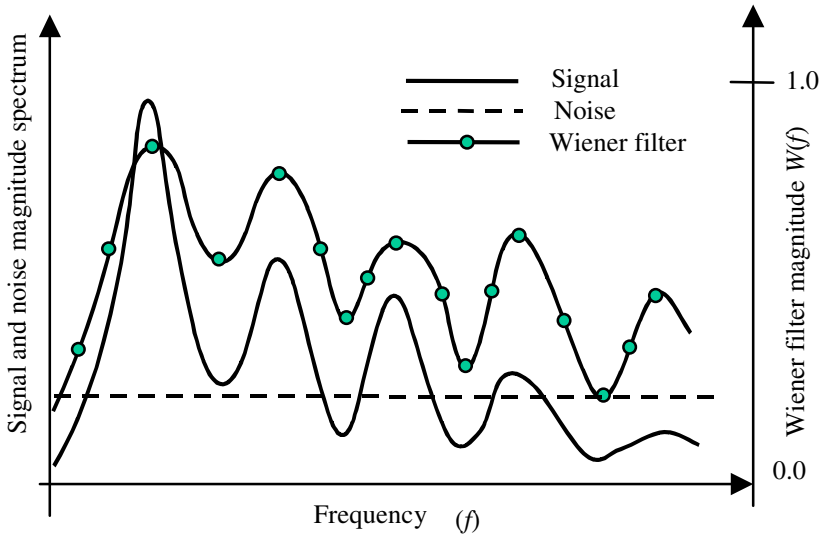
Figure 6.5 Illustration of the variation of Wiener frequency response with signal spectrum for additive white noise. The Wiener filter response broadly follows the signal spectrum.

$$Y(f) = X(f) + N(f) \tag{6.49}$$

where $X(f)$ and $N(f)$ are the signal and noise spectra. For a signal observed in additive random noise, the frequency-domain Wiener filter is obtained as

$$W(f) = \frac{P_{XX}(f)}{P_{XX}(f) + P_{NN}(f)} \tag{6.50}$$

where $P_{XX}(f)$ and $P_{NN}(f)$ are the signal and noise power spectra. Dividing the numerator and the denominator of Equation (6.50) by the noise power spectra $P_{NN}(f)$ and substituting the variable $SNR(f) = P_{XX}(f)/P_{NN}(f)$ yields

$$W(f) = \frac{SNR(f)}{SNR(f) + 1} \tag{6.51}$$

where SNR is a signal-to-noise ratio measure. Note that the variable, $SNR(f)$ is expressed in terms of the power-spectral ratio, and not in the more usual terms of log power ratio. Therefore $SNR(f) = 0$ corresponds to $-\infty$ dB.

From Equation (6.51), the following interpretation of the Wiener filter frequency response $W(f)$ in terms of the signal-to-noise ratio can be
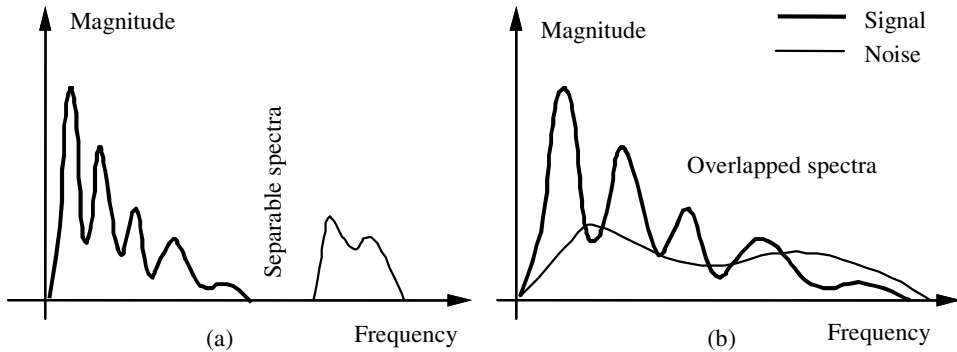
Figure 6.6 Illustration of separability: (a) The signal and noise spectra do not overlap, and the signal can be recovered by a low-pass filter; (b) the signal and noise spectra overlap, and the noise can be reduced but not completely removed.

deduced. For additive noise, the Wiener filter frequency response is a real positive number in the range $0 \leq W(f) \leq 1$. Now consider the two limiting cases of (a) a noise-free signal $SNR(f) = \infty$ and (b) an extremely noisy signal $SNR(f)=0$. At very high SNR, $W(f) \approx 1$, and the filter applies little or no attenuation to the noise-free frequency component. At the other extreme, when $SNR(f)=0$, $W(f)=0$. Therefore, *for additive noise, the Wiener filter attenuates each frequency component in proportion to an estimate of the signal to noise ratio*. Figure 6.4 shows the variation of the Wiener filter response $W(f)$, with the signal-to-noise ratio $SNR(f)$.

An alternative illustration of the variations of the Wiener filter frequency response with $SNR(f)$ is shown in Figure 6.5. It illustrates the similarity between the Wiener filter frequency response and the signal spectrum for the case of an additive white noise disturbance. Note that at a spectral peak of the signal spectrum, where the $SNR(f)$ is relatively high, the Wiener filter frequency response is also high, and the filter applies little attenuation. At a signal trough, the signal-to-noise ratio is low, and so is the Wiener filter response. Hence, for additive white noise, the Wiener filter response broadly follows the signal spectrum.

6.6.2 Wiener Filter and the Separability of Signal and Noise

A signal is completely recoverable from noise if the spectra of the signal and the noise do not overlap. An example of a noisy signal with separable signal and noise spectra is shown in Figure 6.6(a). In this case, the signal

and the noise occupy different parts of the frequency spectrum, and can be separated with a low-pass, or a high-pass, filter. Figure 6.6(b) illustrates a more common example of a signal and noise process with overlapping spectra. For this case, it is not possible to completely separate the signal from the noise. However, the effects of the noise can be reduced by using a Wiener filter that attenuates each noisy signal frequency in proportion to an estimate of the signal-to-noise ratio as described by Equation (6.51).

6.6.3 The Square-Root Wiener Filter

In the frequency domain, the Wiener filter output $\hat{X}(f)$ is the product of the input frequency $X(f)$ and the filter response $W(f)$ as expressed in Equation (6.38). Taking the expectation of the squared magnitude of both sides of Equation (6.38) yields the power spectrum of the filtered signal as

$$\mathcal{E}[|\hat{X}(f)|^2] = |W(f)|^2 \, \mathcal{E}[|Y(f)|^2]$$
$$= |W(f)|^2 \, P_{YY}(f) \tag{6.52}$$

Substitution of $W(f)$ from Equation (6.43) in Equation (6.52) yields

$$\mathcal{E}[|\hat{X}(f)|^2] = \frac{P_{XY}^2(f)}{P_{YY}(f)} \tag{6.53}$$

Now, for a signal observed in an uncorrelated additive noise we have

$$P_{YY}(f) = P_{XX}(f) + P_{NN}(f) \tag{6.54}$$

and

$$P_{XY}(f) = P_{XX}(f) \tag{6.55}$$

Substitution of Equations (6.54) and (6.55) in Equation (6.53) yields

$$\mathcal{E}[|\hat{X}(f)|^2] = \frac{P_{XX}^2(f)}{P_{XX}(f) + P_{NN}(f)} \tag{6.56}$$

Now, in Equation (6.38) if instead of the Wiener filter, the square root of the Wiener filter magnitude frequency response is used, the result is

$$\hat{X}(f) = |W(f)|^{1/2} Y(f) \tag{6.57}$$

and the power spectrum of the signal, filtered by the square-root Wiener filter, is given by

$$\mathcal{E}[|\hat{X}(f)|^2] = \left[|W(f)|^{1/2}\right]^2 \mathcal{E}[|Y(f)|^2] = \frac{P_{XY}(f)}{P_{YY}(f)} P_{YY}(f) = P_{XY}(f) \tag{6.58}$$

Now, for uncorrelated signal and noise Equation (6.58) becomes

$$\mathcal{E}[|\hat{X}(f)|^2] = P_{XX}(f) \tag{6.59}$$

Thus, for additive noise the power spectrum of the output of the square-root Wiener filter is the same as the power spectrum of the desired signal.

6.6.4 Wiener Channel Equaliser

Communication channel distortions may be modelled by a combination of a linear filter and an additive random noise source as shown in Figure 6.7. The input/output signals of a linear time invariant channel can be modelled as

$$y(m) = \sum_{k=0}^{P-1} h_k\, x(m-k) + n(m) \tag{6.60}$$

where $x(m)$ and $y(m)$ are the transmitted and received signals, $[h_k]$ is the impulse response of a linear filter model of the channel, and $n(m)$ models the channel noise. In the frequency domain Equation (6.60) becomes

$$Y(f) = X(f)H(f) + N(f) \tag{6.61}$$

where $X(f)$, $Y(f)$, $H(f)$ and $N(f)$ are the signal, noisy signal, channel and noise spectra respectively. To remove the channel distortions, the receiver is followed by an equaliser. The equaliser input is the distorted channel output, and the desired signal is the channel input. Using Equation (6.43) it is easy to show that the Wiener equaliser in the frequency domain is given by
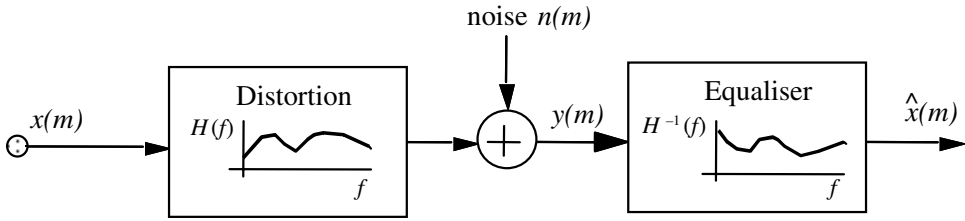
Figure 6.7 Illustration of a channel model followed by an equaliser.

$$W(f) = \frac{P_{XX}(f)H^*(f)}{P_{XX}(f)|H(f)|^2 + P_{NN}(f)} \tag{6.62}$$

where it is assumed that the channel noise and the signal are uncorrelated. In the absence of channel noise, $P_{NN}(f)=0$, and the Wiener filter is simply the inverse of the channel filter model $W(f)=H^{-1}(f)$. The equalisation problem is treated in detail in Chapter 15.

6.6.5 Time-Alignment of Signals in Multichannel/Multisensor Systems

In multichannel/multisensor signal processing there are a number of noisy and distorted versions of a signal $x(m)$, and the objective is to use all the observations in estimating $x(m)$, as illustrated in Figure 6.8, where the phase and frequency characteristics of each channel is modelled by a linear filter. As a simple example, consider the problem of time-alignment of two noisy records of a signal given as

$$y_1(m) = x(m) + n_1(m) \tag{6.63}$$

$$y_2(m) = Ax(m-D) + n_2(m) \tag{6.64}$$

where $y_1(m)$ and $y_2(m)$ are the noisy observations from channels 1 and 2, $n_1(m)$ and $n_2(m)$ are uncorrelated noise in each channel, D is the time delay of arrival of the two signals, and A is an amplitude scaling factor. Now assume that $y_1(m)$ is used as the input to a Wiener filter and that, in the absence of the signal $x(m)$, $y_2(m)$ is used as the "desired" signal. The error signal is given by
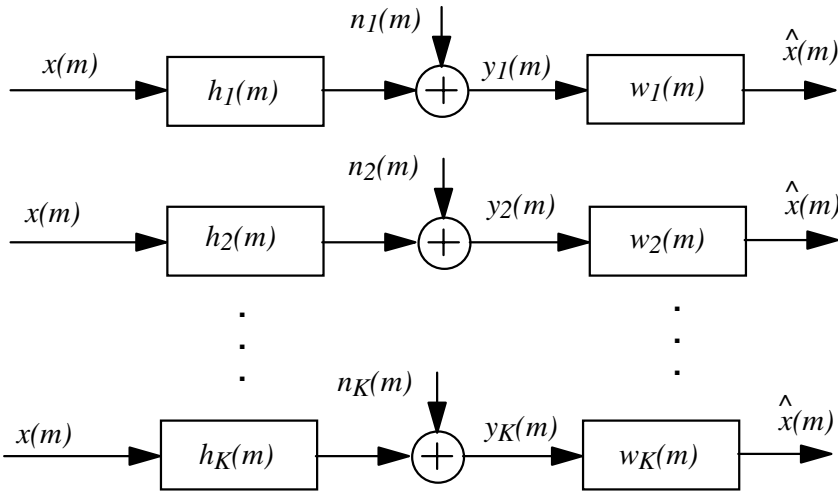
Figure 6.8 Illustration of a multichannel system where Wiener filters are used to time-align the signals from different channels.

$$e(m) = y_2(m) - \sum_{k=0}^{P-1} w_k \, y_1(m)$$

$$= \left(A x(m-D) - \sum_{k=0}^{P-1} w_k \, x(m) \right) + \left(\sum_{k=0}^{P-1} w_k \, n_1(m) \right) + n_2(m)$$

(6.65)

The Wiener filter strives to minimise the terms shown inside the square brackets in Equation (6.65). Using the Wiener filter Equation (6.10), we have

$$w = R_{y_1 y_1}^{-1} r_{y_1 y_2}$$

$$= \left(R_{xx} + R_{n_1 n_1} \right)^{-1} A r_{xx}(D)$$

(6.66)

where $r_{xx}(D) = \mathcal{E}\,[x(PD)x(m)]$. The frequency-domain equivalent of Equation (6.65) can be derived as

$$W(f) = \frac{P_{XX}(f)}{P_{XX}(f) + P_{N_1 N_1}(f)} A e^{-j\omega D}$$

(6.67)

Note that in the absence of noise, the Wiener filter becomes a pure phase (or a pure delay) filter with a flat magnitude response.
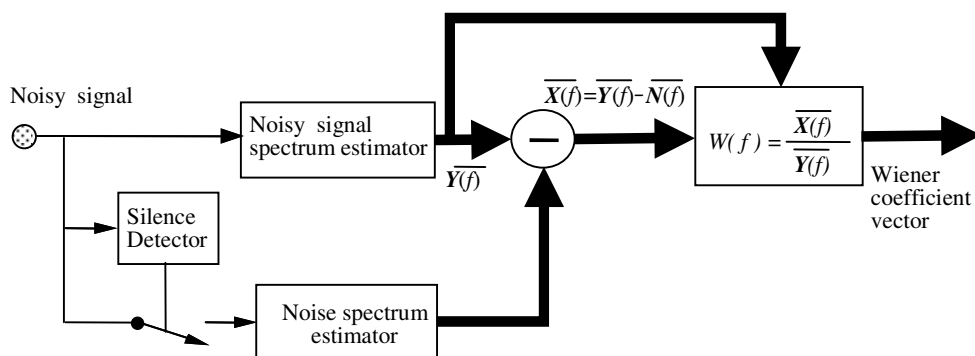
Figure 6.9 Configuration of a system for estimation of frequency Wiener filter.

6.6.6 Implementation of Wiener Filters

The implementation of a Wiener filter for additive noise reduction, using Equations (6.48)–(6.50), requires the autocorrelation functions, or equivalently the power spectra, of the signal and noise. The noise power spectrum can be obtained from the signal-inactive, noise-only, periods. The assumption is that the noise is quasi-stationary, and that its power spectra remains relatively stationary between the update periods. This is a reasonable assumption for many noisy environments such as the noise inside a car emanating from the engine, aircraft noise, office noise from computer machines, etc. The main practical problem in the implementation of a Wiener filter is that the desired signal is often observed in noise, and that the autocorrelation or power spectra of the desired signal are not readily available. Figure 6.9 illustrates the block-diagram configuration of a system for implementation of a Wiener filter for additive noise reduction. An estimate of the desired signal power spectra is obtained by subtracting an estimate of the noise spectra from that of the noisy signal. A filter bank implementation of the Wiener filter is shown in Figure 6.10, where the incoming signal is divided into N bands of frequencies. A first-order integrator, placed at the output of each band-pass filter, gives an estimate of the power spectra of the noisy signal. The power spectrum of the original signal is obtained by subtracting an estimate of the noise power spectrum from the noisy signal. In a Bayesian implementation of the Wiener filter, prior models of speech and noise, such as hidden Markov models, are used to obtain the power spectra of speech and noise required for calculation of the filter coefficients.
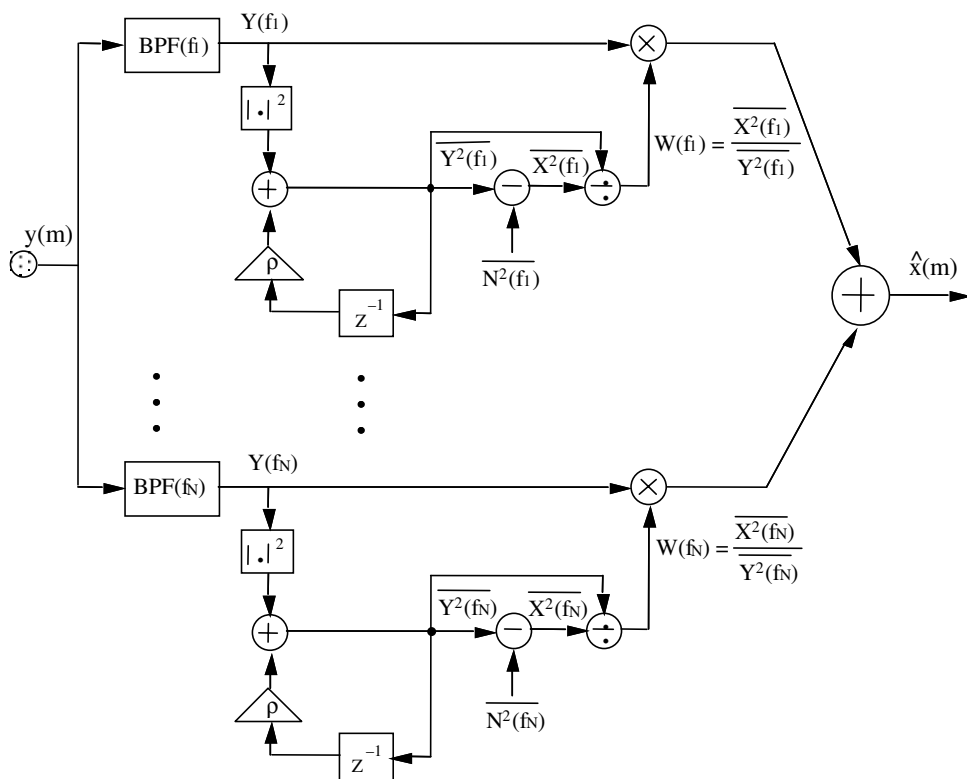
Figure 6.10 A filter-bank implementation of a Wiener filter.

6.7 The Choice of Wiener Filter Order

The choice of Wiener filter order affects:

(a) the ability of the filter to remove distortions and reduce the noise;
(b) the computational complexity of the filter; and
(c) the numerical stability of the of the Wiener solution, Equation (6.10).

The choice of the filter length also depends on the application and the method of implementation of the Wiener filter. For example, in a filter-bank implementation of the Wiener filter for additive noise reduction, the number of filter coefficients is equal to the number of filter banks, and typically the

number of filter banks is between 16 to 64. On the other hand for many applications, a direct implementation of the time-domain Wiener filter requires a larger filter length say between 64 and 256 taps.

A reduction in the required length of a time-domain Wiener filter can be achieved by dividing the time domain signal into N sub-band signals. Each sub-band signal can then be decimated by a factor of N. The decimation results in a reduction, by a factor of N, in the required length of each sub-band Wiener filter. In Chapter 14, a subband echo canceller is described.

6.8 Summary

A Wiener filter is formulated to map an input signal to an output that is as close to a desired signal as possible. This chapter began with the derivation of the least square error Wiener filter. In Section 6.2, we derived the block-data least square error Wiener filter for applications where only finite-length realisations of the input and the desired signals are available. In such cases, the filter is obtained by minimising a time-averaged squared error function. In Section 6.3, we considered a vector space interpretation of the Wiener filters as the perpendicular projection of the desired signal onto the space of the input signal.

In Section 6.4, the least mean square error signal was analysed. The mean square error is zero only if the input signal is related to the desired signal through a linear and invertible filter. For most cases, owing to noise and/or nonlinear distortions of the input signal, the minimum mean square error would be non-zero. In Section 6.5, we derived the Wiener filter in the frequency domain, and considered the issue of separability of signal and noise using a linear filter. Finally in Section 6.6, we considered some applications of Wiener filters in noise reduction, time-delay estimation and channel equalisation.

Bibliography

AKAIKE H. (1974) A New Look at Statistical Model Identification. IEEE Trans. Automatic Control, AC-19, pp. 716–23.

ALEXANDER S.T. (1986) Adaptive Signal Processing Theory and Applications. Springer-Verlag, New York.

ANDERSON B.D. and MOOR J.B. (1979) Linear Optimal Control. Prentice-Hall, Englewood Cliffs, NJ.

DORNY C.N. (1975) A Vector Space Approach to Models and Optimisation. Wiley, New York.

DURBIN J. (1959) Efficient Estimation of Parameters in Moving Average Models. Biometrica, **46**, pp. 306–16.

GIORDANO A.A. and HSU F.M. (1985) Least Square Estimation with Applications to Digital Signal Processing. Wiley, New York.

GIVENS W. (1958) Computation of Plane Unitary Rotations Transforming a General Matrix to Triangular Form. SIAM J. Appl. Math. **6**, pp. 26–50.

GOLUB G.H. and REINSCH (1970) Singular Value Decomposition and Least Squares Solutions. Numerical Mathematics, **14**, pp. 403–20.

GOLUB G.H. and VAN LOAN C.F. (1983). Matrix Computations. Johns Hopkins University Press, Baltimore, MD.

GOLUB G.H., and VAN LOAN C.F. (1980). An Analysis of the Total Least Squares Problem. SIAM Journal of Numerical Analysis, **17**, pp. 883–93.

HALMOS P.R. (1974) Finite-Dimensional Vector Spaces. Springer-Verlag, New York.

HAYKIN S. (1991) Adaptive Filter Theory, 2nd Ed. Prentice-Hall, Englewood Cliffs, NJ.

HOUSEHOLDER A.S.(1964) The Theory of Matrices in Numerical Analysis. Blaisdell, Waltham, MA.

KAILATH T. (1974) A View of Three Decades of Linear Filtering Theory. IEEE Trans. Info. Theory, **IT-20**, pp. 146–81.

KAILATH T. (1977) Linear Least Squares Estimation, Benchmark Papers in Electrical Engineering and Computer science. Dowden, Hutchinson &Ross.

KAILATH T. (1980) Linear Systems, Prentice-Hall, Englewood Cliffs, NJ.

KLEMA V.C. and LAUB A. J. (1980) The Singular Value Decomposition: Its Computation and Some Applications. IEEE Trans. Automatic Control, **AC-25**, pp. 164-76.

KOLMOGROV A.N. (1939) Sur l' Interpolation et Extrapolation des Suites Stationaires. Comptes Rendus de l'Academie des Sciences, **208**, pp. 2043-2046.

LAWSON C.L. and HANSON R.J. (1974) Solving Least Squares Problems. Prentice-Hall, Englewood Cliffs, NJ.

ORFANIDIS S.J. (1988) Optimum Signal Procesing: An Introduction, 2nd Ed. Macmillan, New York.

SCHARF L.L. (1991) Statistical Signal Processing: Detection, Estimation, and Time Series Analysis, Addison Wesley, Reading, MA.

STRANG G. (1976) Linear Algebra and Its Applications, 3rd Ed. Harcourt Brace Jovanovich, San Diego, California.

WIENER N. (1949) Extrapolation, Interpolation and Smoothing of Stationary Time Series. MIT Press Cambridge, MA.

WILKINSON J.H. (1965) The Algebraic Eigenvalue Problem. Oxford University Press.

WHITTLE P.W. (1983) Prediction and Regulation by Linear Least-Squares Methods. University of Minnesota Press, Minneapolis, Minnesota.

WOLD H. (1954) The Analysis of Stationary Time Series, 2nd Ed. Almquist and Wicksell, Uppsala.