7

ADAPTIVE FILTERS

A daptive filters are used for non-stationary signals and environments, or in applications where a sample-by-sample adaptation of a process or a low processing delay is required. Applications of adaptive filters include multichannel noise reduction, radar/sonar signal processing, channel equalization for cellular mobile phones, echo cancellation, and low delay speech coding. This chapter begins with a study of the state-space Kalman filter. In Kalman theory a state equation models the dynamics of the signal generation process, and an observation equation models the channel distortion and additive noise. Then we consider recursive least square (RLS) error adaptive filters. The RLS filter is a sample-adaptive formulation of the Wiener filter, and for stationary signals should converge to the same solution as the Wiener filter. In least square error filtering, an alternative to using a Wiener-type closed-form solution is an iterative gradient-based search for the optimal filter coefficients. The steepest-descent search is a gradient-based method for searching the least square error performance curve for the minimum error filter coefficients. We study the steepest-descent method, and then consider the computationally inexpensive LMS gradient search method.

7.1 State-Space Kalman Filters

The Kalman filter is a recursive least square error method for estimation of a signal distorted in transmission through a channel and observed in noise. Kalman filters can be used with time-varying as well as time-invariant processes. Kalman filter theory is based on a state-space approach in which a state equation models the dynamics of the signal process and an observation equation models the noisy observation signal. For a signal $x(m)$ and noisy observation $y(m)$, the state equation model and the observation model are defined as

$$x(m) = \boldsymbol{\Phi}(m, m-1)x(m-1) + e(m) \tag{7.1}$$

$$y(m) = H(m)x(m) + n(m) \tag{7.2}$$

where

$x(m)$ is the P-dimensional signal, or the state parameter, vector at time m,
$\boldsymbol{\Phi}(m, m-1)$ is a $P \times P$ dimensional state transition matrix that relates the states of the process at times $m-1$ and m,
$e(m)$ is the P-dimensional uncorrelated input excitation vector of the state equation,
$\Sigma_{ee}(m)$ is the $P \times P$ covariance matrix of $e(m)$,
$y(m)$ is the M-dimensional noisy and distorted observation vector,
$H(m)$ is the $M \times P$ channel distortion matrix,
$n(m)$ is the M-dimensional additive noise process,
$\Sigma_{nn}(m)$ is the $M \times M$ covariance matrix of $n(m)$.

The Kalman filter can be derived as a recursive minimum mean square error predictor of a signal $x(m)$, given an observation signal $y(m)$. The filter derivation assumes that the state transition matrix $\boldsymbol{\Phi}(m, m-1)$, the channel distortion matrix $H(m)$, the covariance matrix $\Sigma_{ee}(m)$ of the state equation input and the covariance matrix $\Sigma_{nn}(m)$ of the additive noise are given.

In this chapter, we use the notation $\hat{y}(m|m-i)$ to denote a prediction of $y(m)$ based on the observation samples up to the time $m-i$. Now assume that $\hat{y}(m|m-1)$ is the least square error prediction of $y(m)$ based on the observations $[y(0), ..., y(m-1)]$. Define a so-called *innovation*, or prediction error signal as

$$v(m) = y(m) - \hat{y}(m|m-1) \tag{7.3}$$
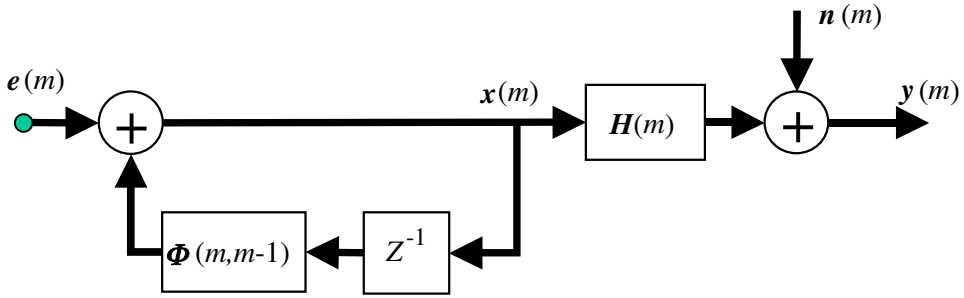
Figure 7.1 Illustration of signal and observation models in Kalman filter theory.

The innovation signal vector $v(m)$ contains all that is unpredictable from the past observations, including both the noise and the unpredictable part of the signal. For an optimal linear least mean square error estimate, the innovation signal must be uncorrelated and orthogonal to the past observation vectors; hence we have

$$\mathcal{E}\left[v(m)y^{\mathrm{T}}(m-k)\right] = 0, \qquad k > 0 \tag{7.4}$$

and

$$\mathcal{E}\left[v(m)v^{\mathrm{T}}(k)\right] = 0, \qquad m \neq k \tag{7.5}$$

The concept of innovations is central to the derivation of the Kalman filter. The least square error criterion is satisfied if the estimation error is orthogonal to the past samples. *In the following derivation of the Kalman filter, the orthogonality condition of Equation (7.4) is used as the starting point to derive an optimal linear filter whose innovations are orthogonal to the past observations.*

Substituting the observation Equation (7.2) in Equation (7.3) and using the relation

$$\hat{y}(m \mid m-1) = \mathcal{E}\left[y(m)\big|\hat{x}(m|m-1)\right]$$
$$= H(m)\hat{x}(m|m-1) \tag{7.6}$$

yields

$$v(m) = H(m)x(m) + n(m) - H(m)\hat{x}(m|m-1)$$
$$= H(m)\tilde{x}(m) + n(m) \tag{7.7}$$

where $\tilde{x}(m)$ is the signal prediction error vector defined as

$$\tilde{x}(m) = x(m) - \hat{x}(m|m-1) \tag{7.8}$$

From Equation (7.7) the covariance matrix of the innovation signal is given by

$$\boldsymbol{\Sigma}_{\boldsymbol{vv}}(m) = \mathcal{E}\!\left[\boldsymbol{v}(m)\boldsymbol{v}^{\mathrm{T}}(m)\right]$$
$$= \boldsymbol{H}(m)\boldsymbol{\Sigma}_{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}}(m)\boldsymbol{H}^{\mathrm{T}}(m) + \boldsymbol{\Sigma}_{\boldsymbol{nn}}(m) \tag{7.9}$$

where $\boldsymbol{\Sigma}_{\tilde{\boldsymbol{x}}\tilde{\boldsymbol{x}}}(m)$ is the covariance matrix of the prediction error $\tilde{\boldsymbol{x}}(m)$. Let $\hat{\boldsymbol{x}}(m+1|m)$ denote the least square error prediction of the signal $\boldsymbol{x}(m+1)$. Now, the prediction of $\boldsymbol{x}(m+1)$, based on the samples available up to the time m, can be expressed recursively as a linear combination of the prediction based on the samples available up to the time $m-1$ and the innovation signal at time m as

$$\hat{\boldsymbol{x}}(m+1|m) = \hat{\boldsymbol{x}}(m+1|m-1) + \boldsymbol{K}(m)\boldsymbol{v}(m) \tag{7.10}$$

where the $P \times M$ matrix $\boldsymbol{K}(m)$ is the Kalman gain matrix. Now, from Equation (7.1), we have

$$\hat{\boldsymbol{x}}(m+1|m-1) = \boldsymbol{\Phi}(m+1,m)\hat{\boldsymbol{x}}(m|m-1) \tag{7.11}$$

Substituting Equation (7.11) in (7.10) gives a recursive prediction equation as

$$\hat{\boldsymbol{x}}(m+1|m) = \boldsymbol{\Phi}(m+1,m)\hat{\boldsymbol{x}}(m|m-1) + \boldsymbol{K}(m)\boldsymbol{v}(m) \tag{7.12}$$

To obtain a recursive relation for the computation and update of the Kalman gain matrix, we multiply both sides of Equation (7.12) by $\boldsymbol{v}^{\mathrm{T}}(m)$ and take the expectation of the results to yield

$$\mathcal{E}\!\left[\hat{\boldsymbol{x}}(m+1|m)\boldsymbol{v}^{\mathrm{T}}(m)\right] = \mathcal{E}\!\left[\boldsymbol{\Phi}(m+1,m)\hat{\boldsymbol{x}}(m|m-1)\boldsymbol{v}^{\mathrm{T}}(m)\right] + \boldsymbol{K}(m)\mathcal{E}\!\left[\boldsymbol{v}(m)\boldsymbol{v}^{\mathrm{T}}(m)\right] \tag{7.13}$$

Owing to the required orthogonality of the innovation sequence and the past samples, we have

$$\mathcal{E}\!\left[\hat{\boldsymbol{x}}(m|m-1)\boldsymbol{v}^{\mathrm{T}}(m)\right] = 0 \tag{7.14}$$

Hence, from Equations (7.13) and (7.14), the Kalman gain matrix is given by

$$\boldsymbol{K}(m) = \mathcal{E}\!\left[\hat{\boldsymbol{x}}(m+1|m)\boldsymbol{v}^{\mathrm{T}}(m)\right]\boldsymbol{\Sigma}_{\boldsymbol{vv}}^{-1}(m) \tag{7.15}$$

The first term on the right-hand side of Equation (7.15) can be expressed as

$$
\begin{aligned}
\mathcal{E}\!\left[\hat{x}(m+1|m)\boldsymbol{v}^{\mathrm{T}}(m)\right] &= \mathcal{E}\!\left[(x(m+1)-\tilde{x}(m+1|m))\boldsymbol{v}^{\mathrm{T}}(m)\right] \\
&= \mathcal{E}\!\left[x(m+1)\boldsymbol{v}^{\mathrm{T}}(m)\right] \\
&= \mathcal{E}\!\left[(\boldsymbol{\Phi}(m+1,m)\boldsymbol{x}(m)+\boldsymbol{e}(m+1))(\boldsymbol{y}(m)-\hat{\boldsymbol{y}}(m|m-1))^{\mathrm{T}}\right] \\
&= \mathcal{E}\!\left[(\boldsymbol{\Phi}(m+1,m)(\hat{\boldsymbol{x}}(m|m-1)+\tilde{x}(m|m-1))](\boldsymbol{H}(m)\tilde{x}(m|m-1)+\boldsymbol{n}(m))^{\mathrm{T}}\right] \\
&= \boldsymbol{\Phi}(m+1,m)\mathcal{E}\!\left[\tilde{x}(m|m-1)\tilde{x}^{\mathrm{T}}(m|m-1)\right]\boldsymbol{H}^{\mathrm{T}}(m)
\end{aligned}
$$

(7.16)

In developing the successive lines of Equation (7.16), we have used the following relations:

$$
\mathcal{E}\!\left[\tilde{x}(m+1\,|\,m)\boldsymbol{v}^{\mathrm{T}}(m)\right]=0
\tag{7.17}
$$

$$
\mathcal{E}\!\left[\boldsymbol{e}(m+1)(\boldsymbol{y}(m)-\hat{\boldsymbol{y}}(m|\,m-1))^{\mathrm{T}}\right]=0
\tag{7.18}
$$

$$
\boldsymbol{x}(m)=\hat{\boldsymbol{x}}(m|\,m-1)+\tilde{x}(m|m-1)
\tag{7.19}
$$

$$
\mathcal{E}[\hat{\boldsymbol{x}}(m\,|\,m-1)\tilde{x}(m|\,m-1)]=0
\tag{7.20}
$$

and we have also used the assumption that the signal and the noise are uncorrelated. Substitution of Equations (7.9) and (7.16) in Equation (7.15) yields the following equation for the Kalman gain matrix:

$$
\boldsymbol{K}(m)=\boldsymbol{\Phi}(m+1,m)\boldsymbol{\Sigma}_{\tilde{x}\tilde{x}}(m)\boldsymbol{H}^{\mathrm{T}}(m)\left[\boldsymbol{H}(m)\boldsymbol{\Sigma}_{\tilde{x}\tilde{x}}(m)\boldsymbol{H}^{\mathrm{T}}(m)+\boldsymbol{\Sigma}_{nn}(m)\right]^{-1}
\tag{7.21}
$$

where $\boldsymbol{\Sigma}_{\tilde{x}\tilde{x}}(m)$ is the covariance matrix of the signal prediction error $\tilde{x}(m|m-1)$. To derive a recursive relation for $\boldsymbol{\Sigma}_{\tilde{x}\tilde{x}}(m)$, we consider

$$
\tilde{x}(m|m-1)=\boldsymbol{x}(m)-\hat{\boldsymbol{x}}(m|m-1)
\tag{7.22}
$$

Substitution of Equation (7.1) and (7.12) in Equation (7.22) and rearrangement of the terms yields

$$
\begin{aligned}
\tilde{x}(m\,|\,m-1) &= [\boldsymbol{\Phi}(m,m-1)\boldsymbol{x}(m-1)+\boldsymbol{e}(m)]-[\boldsymbol{\Phi}(m,m-1)\hat{\boldsymbol{x}}(m-1|m-2)+\boldsymbol{K}(m-1)\boldsymbol{v}(m-1)] \\
&= \boldsymbol{\Phi}(m,m-1)\tilde{x}(m-1)+\boldsymbol{e}(m)-\boldsymbol{K}(m-1)\boldsymbol{H}(m-1)\tilde{x}(m-1)+\boldsymbol{K}(m-1)\boldsymbol{n}(m-1) \\
&= [\boldsymbol{\Phi}(m,m-1)-\boldsymbol{K}(m-1)\boldsymbol{H}(m-1)]\tilde{x}(m-1)+\boldsymbol{e}(m)+\boldsymbol{K}(m-1)\boldsymbol{n}(m-1)
\end{aligned}
$$

(7.23)

From Equation (7.23) we can derive the following recursive relation for the variance of the signal prediction error

$$\boldsymbol{\Sigma}_{\tilde{x}\tilde{x}}(m)=\boldsymbol{L}(m)\boldsymbol{\Sigma}_{\tilde{x}\tilde{x}}(m-1)\boldsymbol{L}^{\mathrm{T}}(m)+\boldsymbol{\Sigma}_{ee}(m)+\boldsymbol{K}(m-1)\boldsymbol{\Sigma}_{nn}(m-1)\boldsymbol{K}^{\mathrm{T}}(m-1)$$

(7.24)

where the $P \times P$ matrix $\boldsymbol{L}(m)$ is defined as

$$\boldsymbol{L}(m) = \left[\boldsymbol{\Phi}(m,m-1) - \boldsymbol{K}(m-1)\boldsymbol{H}(m-1)\right]$$

(7.25)

Kalman Filtering Algorithm

Input: observation vectors $\{\boldsymbol{y}(m)\}$
Output: state or signal vectors $\{\hat{\boldsymbol{x}}(m)\}$
Initial conditions:

$$\boldsymbol{\Sigma}_{\tilde{x}\tilde{x}}(0) = \delta\boldsymbol{I}$$

(7.26)

$$\hat{\boldsymbol{x}}(0|-1) = 0$$

(7.27)

For $m = 0, 1, ...$
Innovation signal:

$$\boldsymbol{v}(m) = \boldsymbol{y}(m) - \boldsymbol{H}(m)\hat{\boldsymbol{x}}(m|m-1)$$

(7.28)

Kalman gain:

$$\boldsymbol{K}(m)=\boldsymbol{\Phi}(m+1,m)\boldsymbol{\Sigma}_{\tilde{x}\tilde{x}}(m)\boldsymbol{H}^{\mathrm{T}}(m)\left[\boldsymbol{H}(m)\boldsymbol{\Sigma}_{\tilde{x}\tilde{x}}(m)\boldsymbol{H}^{\mathrm{T}}(m)+\boldsymbol{\Sigma}_{nn}(m)\right]^{-1}$$

(7.29)

Prediction update:

$$\hat{\boldsymbol{x}}(m+1|m)= \boldsymbol{\Phi}(m+1,m)\hat{\boldsymbol{x}}(m|m-1)+ \boldsymbol{K}(m)\boldsymbol{v}(m)$$

(7.30)

Prediction error correlation matrix update:

$$\boldsymbol{L}(m+1) = [\boldsymbol{\Phi}(m+1,m) - \boldsymbol{K}(m)\boldsymbol{H}(m)]$$

(7.31)

$$\boldsymbol{\Sigma}_{\tilde{x}\tilde{x}}(m+1)=\boldsymbol{L}(m+1)\boldsymbol{\Sigma}_{\tilde{x}\tilde{x}}(m)\boldsymbol{L}(m+1)^{\mathrm{T}} +\boldsymbol{\Sigma}_{ee}(m+1)+\boldsymbol{K}(m)\boldsymbol{\Sigma}_{nn}(m)\boldsymbol{K}(m)$$

(7.32)

Example 7.1 Consider the Kalman filtering of a first-order AR process $x(m)$ observed in an additive white Gaussian noise $n(m)$. Assume that the signal generation and the observation equations are given as

$$x(m)= a(m)x(m-1)+ e(m)$$

(7.33)

$$y(m) = x(m) + n(m) \tag{7.34}$$

Let $\sigma_e^2(m)$ and $\sigma_n^2(m)$ denote the variances of the excitation signal $e(m)$ and the noise $n(m)$ respectively. Substituting $\Phi(m+1,m) = a(m)$ and $H(m) = 1$ in the Kalman filter equations yields the following Kalman filter algorithm:

Initial conditions:

$$\sigma_{\tilde{x}}^2(0) = \delta \tag{7.35}$$

$$\hat{x}(0|-1) = 0 \tag{7.36}$$

For $m = 0, 1, ...$
Kalman gain:

$$k(m) = \frac{a(m+1)\sigma_{\tilde{x}}^2(m)}{\sigma_{\tilde{x}}^2(m) + \sigma_n^2(m)} \tag{7.37}$$

Innovation signal:

$$v(m) = y(m) - \hat{x}(m \mid m-1) \tag{7.38}$$

Prediction signal update:

$$\hat{x}(m+1|m) = a(m+1)\hat{x}(m|m-1) + k(m)v(m) \tag{7.39}$$

Prediction error update:

$$\sigma_{\tilde{x}}^2(m+1) = [a(m+1) - k(m)]^2\,\sigma_{\tilde{x}}^2(m) + \sigma_e^2(m+1) + k^2(m)\,\sigma_n^2(m) \tag{7.40}$$

where $\sigma_{\tilde{x}}^2(m)$ is the variance of the prediction error signal.

Example 7.2 _Recursive estimation of a constant signal observed in noise._
Consider the estimation of a constant signal observed in a random noise. The state and observation equations for this problem are given by

$$x(m) = x(m-1) = x \tag{7.41}$$

$$y(m) = x + n(m) \tag{7.42}$$

Note that $\Phi(m,m-1) = 1$, state excitation $e(m) = 0$ and $H(m) = 1$. Using the Kalman algorithm, we have the following recursive solutions:

Initial Conditions:

$$\sigma_{\tilde{x}}^2(0) = \delta \tag{7.43}$$

$$\hat{x}(0|-1) = 0 \tag{7.44}$$

For $m = 0, 1, ...$
Kalman gain:

$$k(m) = \frac{\sigma_{\tilde{x}}^2(m)}{\sigma_{\tilde{x}}^2(m) + \sigma_n^2(m)} \tag{7.45}$$

Innovation signal:

$$v(m) = y(m) - \hat{x}(m|m-1) \tag{7.46}$$

Prediction signal update:

$$\hat{x}(m+1|m) = \hat{x}(m|m-1) + k(m)v(m) \tag{7.47}$$

Prediction error update:

$$\sigma_{\tilde{x}}^2(m+1) = [1 - k(m)]^2 \sigma_{\tilde{x}}^2(m) + k^2(m)\sigma_n^2(m) \tag{7.48}$$

7.2 Sample-Adaptive Filters

Sample adaptive filters, namely the RLS, the steepest descent and the LMS, are recursive formulations of the least square error Wiener filter. Sample-adaptive filters have a number of advantages over the block-adaptive filters of Chapter 6, including lower processing delay and better tracking of non-stationary signals. These are essential characteristics in applications such as echo cancellation, adaptive delay estimation, low-delay predictive coding, noise cancellation, radar, and channel equalisation in mobile telephony, where low delay and fast tracking of time-varying processes and environments are important objectives.

Figure 7.2 illustrates the configuration of a least square error adaptive filter. At each sampling time, an adaptation algorithm adjusts the filter coefficients to minimise the difference between the filter output and a desired, or target, signal. An adaptive filter starts at some initial state, and then the filter coefficients are periodically updated, usually on a sample-by-sample basis, to minimise the difference between the filter output and a desired or target signal. The adaptation formula has the general recursive form:

next parameter estimate = previous parameter estimate + update(error)

where the update term is a function of the error signal. In adaptive filtering a number of decisions has to be made concerning the filter model and the adaptation algorithm:
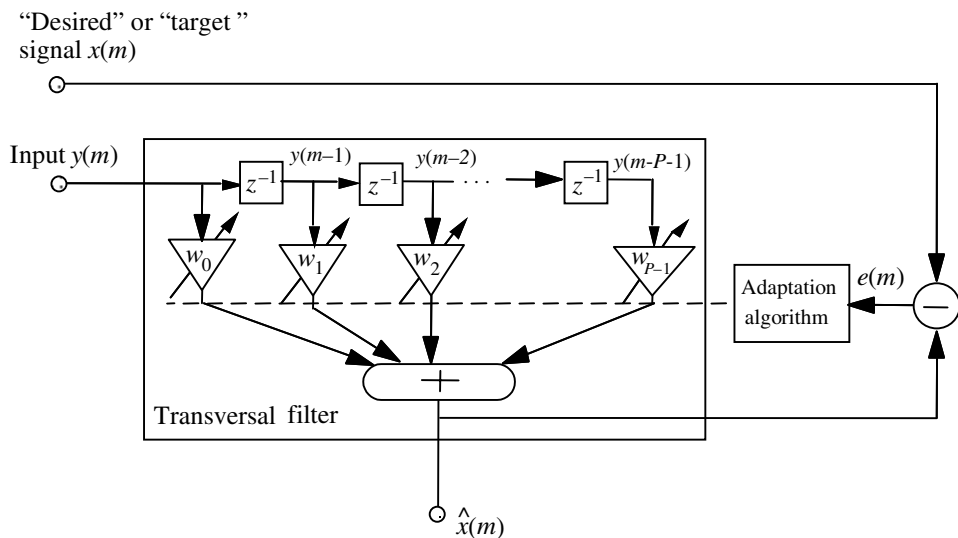
(a) Filter type: This can be a finite impulse response (FIR) filter, or an infinite impulse response (IIR) filter. In this chapter we only consider FIR filters, since they have good stability and convergence properties and for this reason are the type most often used in practice.

(b) Filter order: Often the correct number of filter taps is unknown. The filter order is either set using *a priori* knowledge of the input and the desired signals, or it may be obtained by monitoring the changes in the error signal as a function of the increasing filter order.

(c) Adaptation algorithm: The two most widely used adaptation algorithms are the recursive least square (RLS) error and the least mean square error (LMS) methods. The factors that influence the choice of the adaptation algorithm are the computational complexity, the speed of convergence to optimal operating condition, the minimum error at convergence, the numerical stability and the robustness of the algorithm to initial parameter states.

7.3 Recursive Least Square (RLS) Adaptive Filters

The recursive least square error (RLS) filter is a sample-adaptive, time-update, version of the Wiener filter studied in Chapter 6. For stationary signals, the RLS filter converges to the same optimal filter coefficients as the Wiener filter. For non-stationary signals, the RLS filter tracks the time variations of the process. The RLS filter has a relatively fast rate of convergence to the optimal filter coefficients. This is useful in applications such as speech enhancement, channel equalization, echo cancellation and radar where the filter should be able to track relatively fast changes in the signal process.

In the recursive least square algorithm, the adaptation starts with some initial filter state, and successive samples of the input signals are used to adapt the filter coefficients. Figure 7.2 illustrates the configuration of an adaptive filter where $y(m)$, $x(m)$ and $\mathbf{w}(m) = [w_0(m), w_1(m), ..., w_{P-1}(m)]$ denote the filter input, the desired signal and the filter coefficient vector respectively. The filter output can be expressed as

$$\hat{x}(m) = \mathbf{w}^{\mathrm{T}}(m)\,\mathbf{y}(m) \tag{7.49}$$

"Desired" or "target"
signal $x(m)$

Input $y(m)$

$y(m–1)$ $y(m–2)$ $\cdots$ $y(m\text{-}P\text{-}1)$

z^{-1} z^{-1} z^{-1}

w_0 w_1 w_2 w_{P-1}

Adaptation algorithm $e(m)$

$+$

Transversal filter

$\hat{x}(m)$

Figure 7.2 Illustration of the configuration of an adaptive filter.

where $\hat{x}(m)$ is an estimate of the desired signal $x(m)$. The filter error signal is defined as

$$e(m) = x(m) - \hat{x}(m)$$
$$= x(m) - \boldsymbol{w}^{\mathrm{T}}(m)\boldsymbol{y}(m) \tag{7.50}$$

The adaptation process is based on the minimization of the mean square error criterion defined as

$$\mathcal{E}[e^2(m)] = \mathcal{E}\left\{\left[x(m) - \boldsymbol{w}^{\mathrm{T}}(m)\boldsymbol{y}(m)\right]^2\right\}$$

$$= \mathcal{E}[x^2(m)] - 2\boldsymbol{w}^{\mathrm{T}}(m)\mathcal{E}[\boldsymbol{y}(m)x(m)] + \boldsymbol{w}^{\mathrm{T}}(m)\mathcal{E}[\boldsymbol{y}(m)\boldsymbol{y}^{\mathrm{T}}(m)]\boldsymbol{w}(m)$$

$$= r_{xx}(0) - 2\boldsymbol{w}^{\mathrm{T}}(m)\boldsymbol{r}_{yx}(m) + \boldsymbol{w}^{\mathrm{T}}(m)\boldsymbol{R}_{yy}(m)\boldsymbol{w}(m)$$

$$\tag{7.51}$$

The Wiener filter is obtained by minimising the mean square error with respect to the filter coefficients. For stationary signals, the result of this minimisation is given in Chapter 6, Equation (6.10), as

$$\boldsymbol{w} = \boldsymbol{R}_{yy}^{-1}\,\boldsymbol{r}_{yx} \tag{7.52}$$

where $\boldsymbol{R}_{yy}$ is the autocorrelation matrix of the input signal and $\boldsymbol{r}_{yx}$ is the cross-correlation vector of the input and the target signals. In the following, we formulate a recursive, time-update, adaptive formulation of Equation (7.52). From Section 6.2, for a block of N sample vectors, the correlation matrix can be written as

$$\boldsymbol{R}_{yy} = \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{Y} = \sum_{m=0}^{N-1} \boldsymbol{y}(m)\boldsymbol{y}^{\mathrm{T}}(m) \tag{7.53}$$

where $\boldsymbol{y}(m)=[y(m), ..., y(m–P)]^{\mathrm{T}}$. Now, the sum of vector product in Equation (7.53) can be expressed in recursive fashion as

$$\boldsymbol{R}_{yy}(m) = \boldsymbol{R}_{yy}(m-1) + \boldsymbol{y}(m)\boldsymbol{y}^{\mathrm{T}}(m) \tag{7.54}$$

To introduce adaptability to the time variations of the signal statistics, the autocorrelation estimate in Equation (7.54) can be windowed by an exponentially decaying window:

$$\boldsymbol{R}_{yy}(m) = \lambda\,\boldsymbol{R}_{yy}(m-1) + \boldsymbol{y}(m)\boldsymbol{y}^{\mathrm{T}}(m) \tag{7.55}$$

where λ is the so-called adaptation, or forgetting factor, and is in the range $0>\lambda>1$. Similarly, the cross-correlation vector is given by

$$\boldsymbol{r}_{yx} = \sum_{m=0}^{N-1} \boldsymbol{y}(m)x(m) \tag{7.56}$$

The sum of products in Equation (7.56) can be calculated in recursive form as

$$\boldsymbol{r}_{yx}(m) = \boldsymbol{r}_{yx}(m-1) + \boldsymbol{y}(m)x(m) \tag{7.57}$$

Again this equation can be made adaptive using an exponentially decaying forgetting factor λ:

$$\boldsymbol{r}_{yx}(m) = \lambda\boldsymbol{r}_{yx}(m-1) + \boldsymbol{y}(m)x(m) \tag{7.58}$$

For a recursive solution of the least square error Equation (7.58), we need to obtain a recursive time-update formula for the inverse matrix in the form

$$R_{yy}^{-1}(m)=R_{yy}^{-1}(m-1)+Update(m) \tag{7.59}$$

A recursive relation for the matrix inversion is obtained using the following lemma.

The Matrix Inversion Lemma Let A and B be two positive-definite $P \times P$ matrices related by

$$A = B^{-1} + CD^{-1}C^{T} \tag{7.60}$$

where D is a positive-definite $N \times N$ matrix and C is a $P \times N$ matrix. The matrix inversion lemma states that the inverse of the matrix A can be expressed as

$$A^{-1} = B - BC\left(D + C^{T}BC\right)^{-1}C^{T}B \tag{7.61}$$

This lemma can be proved by multiplying Equation (7.60) and Equation (7.61). The left and right hand sides of the results of multiplication are the identity matrix. The matrix inversion lemma can be used to obtain a recursive implementation for the inverse of the correlation matrix $R_{yy}^{-1}(m)$.

Let

$$R_{yy}(m) = A \tag{7.62}$$

$$\lambda^{-1}R_{yy}^{-1}(m-1) = B \tag{7.63}$$

$$y(m) = C \tag{7.64}$$

$$D = \text{identity matrix} \tag{7.65}$$

Substituting Equations (7.62) and (7.63) in Equation (7.61), we obtain

$$R_{yy}^{-1}(m) = \lambda^{-1}R_{yy}^{-1}(m-1) - \frac{\lambda^{-2}R_{yy}^{-1}(m-1)y(m)y^{T}(m)R_{yy}^{-1}(m-1)}{1+\lambda^{-1}y^{T}(m)R_{yy}^{-1}(m-1)y(m)} \tag{7.66}$$

Now define the variables $\Phi(m)$ and $k(m)$ as

$$\Phi_{yy}(m)= R_{yy}^{-1}(m) \tag{7.67}$$

and

$$k(m) = \frac{\lambda^{-1} R_{yy}^{-1}(m-1) y(m)}{1 + \lambda^{-1} y^{T}(m) R_{yy}^{-1}(m-1) y(m)} \qquad (7.68)$$

or

$$k(m) = \frac{\lambda^{-1} \Phi_{yy}(m-1) y(m)}{1 + \lambda^{-1} y^{T}(m) \Phi_{yy}(m-1) y(m)} \qquad (7.69)$$

Using Equations (7.67) and (7.69), the recursive equation (7.66) for computing the inverse matrix can be written as

$$\Phi_{yy}(m) = \lambda^{-1} \Phi_{yy}(m-1) - \lambda^{-1} k(m) y^{T}(m) \Phi_{yy}(m-1) \qquad (7.70)$$

From Equations (7.69) and (7.70), we have

$$\begin{aligned} k(m) &= \left[\lambda^{-1} \Phi_{yy}(m-1) - \lambda^{-1} k(m) y^{T}(m) \Phi_{yy}(m-1) \right] y(m) \\ &= \Phi_{yy}(m) y(m) \end{aligned} \qquad (7.71)$$

Now Equations (7.70) and (7.71) are used in the following to derive the RLS adaptation algorithm.

Recursive Time-update of Filter Coefficients The least square error filter coefficients are

$$\begin{aligned} w(m) &= R_{yy}^{-1}(m) \, r_{yx}(m) \\ &= \Phi_{yy}(m) r_{yx}(m) \end{aligned} \qquad (7.72)$$

Substituting the recursive form of the correlation vector in Equation (7.72) yields

$$\begin{aligned} w(m) &= \Phi_{yy}(m) \left[\lambda r_{yx}(m-1) + y(m) x(m) \right] \\ &= \lambda \Phi_{yy}(m) r_{yx}(m-1) + \Phi_{yy}(m) y(m) x(m) \end{aligned} \qquad (7.73)$$

Now substitution of the recursive form of the matrix $\Phi_{yy}(m)$ from Equation (7.70) and $k(m) = \Phi(m) y(m)$ from Equation (7.71) in the right-hand side of Equation (7.73) yields

$$w(m) = \left[\lambda^{-1} \boldsymbol{\Phi}_{yy}(m-1) - \lambda^{-1} \boldsymbol{k}(m) \boldsymbol{y}^{\mathrm{T}}(m) \boldsymbol{\Phi}_{yy}(m-1) \right] \lambda \boldsymbol{r}_{yx}(m-1) + \boldsymbol{k}(m) x(m)$$

(7.74)

or

$$w(m) = \boldsymbol{\Phi}_{yy}(m-1) \boldsymbol{r}_{yx}(m-1) - \boldsymbol{k}(m) \boldsymbol{y}^{\mathrm{T}}(m) \boldsymbol{\Phi}_{yy}(m-1) \boldsymbol{r}_{yx}(m-1) + \boldsymbol{k}(m) x(m)$$

(7.75)

Substitution of $w(m{-}1) = \boldsymbol{\Phi}(m{-}1) \boldsymbol{r}_{yx}(m{-}1)$ in Equation (7.75) yields

$$w(m) = w(m-1) - \boldsymbol{k}(m) \left[x(m) - \boldsymbol{y}^{\mathrm{T}}(m) w(m-1) \right]$$

(7.76)

This equation can be rewritten in the following form

$$w(m) = w(m-1) - \boldsymbol{k}(m) e(m)$$

(7.77)

Equation (7.77) is a recursive time-update implementation of the least square error Wiener filter.

RLS Adaptation Algorithm

Input signals: $y(m)$ and $x(m)$

Initial values: $\boldsymbol{\Phi}_{yy}(m) = \delta \mathbf{I}$

$\qquad\qquad\qquad w(0) = w_{\mathrm{I}}$

For $m = 1, 2, \ldots$

Filter gain vector:

$$\boldsymbol{k}(m) = \frac{\lambda^{-1} \boldsymbol{\Phi}_{yy}(m-1) \boldsymbol{y}(m)}{1 + \lambda^{-1} \boldsymbol{y}^{\mathrm{T}}(m) \boldsymbol{\Phi}_{yy}(m-1) \boldsymbol{y}(m)}$$

(7.78)

Error signal equation:

$$e(m) = x(m) - \boldsymbol{w}^{\mathrm{T}}(m-1) \boldsymbol{y}(m)$$

(7.79)

Filter coefficients:

$$w(m) = w(m-1) - \boldsymbol{k}(m) e(m)$$

(7.80)

Inverse correlation matrix update:

$$\boldsymbol{\Phi}_{yy}(m) = \lambda^{-1} \boldsymbol{\Phi}_{yy}(m-1) - \lambda^{-1} \boldsymbol{k}(m) \boldsymbol{y}^{T}(m) \boldsymbol{\Phi}_{yy}(m-1)$$

(7.81)

$$\mathcal{E}\left[e^2(m)\right]$$
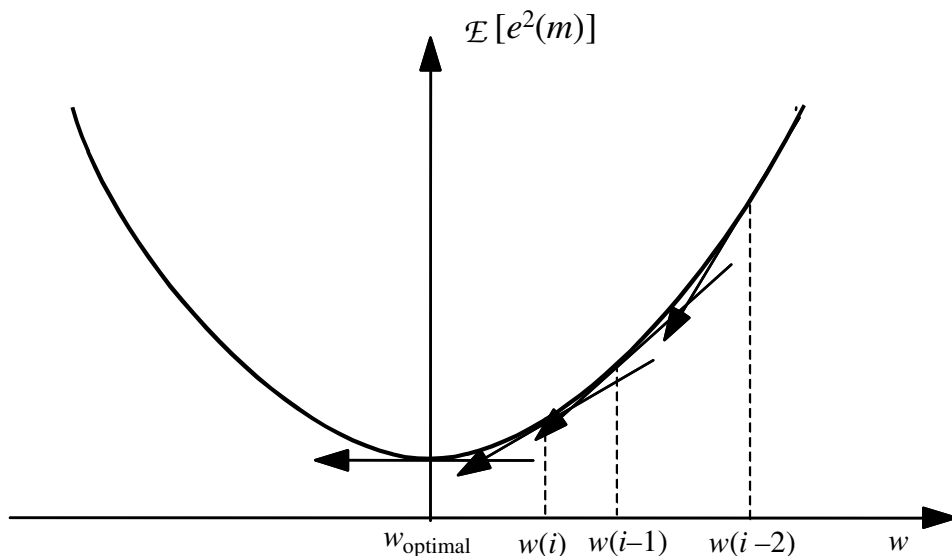
Figure 7.3 Illustration of gradient search of the mean square error surface for the minimum error point.

7.4 The Steepest-Descent Method

The mean square error surface with respect to the coefficients of an FIR filter, is a quadratic bowl-shaped curve, with a single global minimum that corresponds to the LSE filter coefficients. Figure 7.3 illustrates the mean square error curve for a single coefficient filter. This figure also illustrates the steepest-descent search for the minimum mean square error coefficient. The search is based on taking a number of successive downward steps in the direction of negative gradient of the error surface. Starting with a set of initial values, the filter coefficients are successively updated in the downward direction, until the minimum point, at which the gradient is zero, is reached. The steepest-descent adaptation method can be expressed as

$$w(m+1) = w(m) + \mu \left[-\frac{\partial\,\mathcal{E}[e^2(m)]}{\partial\,w(m)} \right] \tag{7.82}$$

where μ is the adaptation step size. From Equation (5.7), the gradient of the mean square error function is given by

$$\frac{\partial\, \mathcal{E}[e^2(m)]}{\partial\, w(m)} = -2r_{yx} + 2R_{yy}\,w(m) \tag{7.83}$$

Substituting Equation (7.83) in Equation (7.82) yields

$$w(m+1) = w(m) + \mu\left[r_{yx} - R_{yy}\,w(m)\right] \tag{7.84}$$

where the factor of 2 in Equation (7.83) has been absorbed in the adaptation step size μ. Let w_o denote the optimal LSE filter coefficient vector, we define a filter coefficients error vector $\tilde{w}(m)$ as

$$\tilde{w}(m) = w(m) - w_o \tag{7.85}$$

For a stationary process, the optimal LSE filter w_o is obtained from the Wiener filter, Equation (5.10), as

$$w_o = R_{yy}^{-1} r_{yx} \tag{7.86}$$

Subtracting w_o from both sides of Equation (7.84), and then substituting $R_{yy}w_o$ for r_{yx}, and using Equation (7.85) yields

$$\tilde{w}(m+1) = \left[\mathbf{I} - \mu R_{yy}\right]\tilde{w}(m) \tag{7.87}$$

It is desirable that the filter error vector $\tilde{w}(m)$ vanishes as rapidly as possible. The parameter μ, the adaptation step size, controls the stability and the rate of convergence of the adaptive filter. Too large a value for μ causes instability; too small a value gives a low convergence rate. The stability of the parameter estimation method depends on the choice of the adaptation parameter μ and the autocorrelation matrix. From Equation (7.87), a recursive equation for the error in each individual filter coefficient can be obtained as follows. The correlation matrix can be expressed in terms of the matrices of eigenvectors and eigenvalues as

$$R_{yy} = Q\Lambda Q^{\mathrm{T}} \tag{7.88}$$
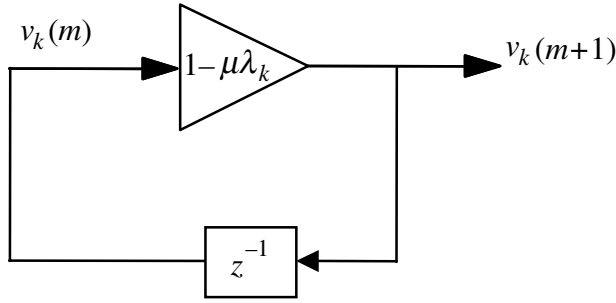
Figure 7.4 A feedback model of the variation of coefficient error with time.

where Q is an orthonormal matrix of the eigenvectors of R_{yy}, and Λ is a diagonal matrix with its diagonal elements corresponding to the eigenvalues of R_{yy}. Substituting R_{yy} from Equation (7.88) in Equation (7.87) yields

$$\tilde{w}(m+1)=\left[I-\mu Q\Lambda Q^{T}\right]\tilde{w}(m) \tag{7.89}$$

Multiplying both sides of Equation (7.89) by Q^{T} and using the relation $Q^{T}Q=QQ^{T}=I$ yields

$$Q^{T}\tilde{w}(m+1)=[I-\mu\Lambda]Q^{T}\tilde{w}(m) \tag{7.90}$$

Let

$$v(m)=Q^{T}\tilde{w}(m) \tag{7.91}$$

Then

$$v(m+1)=[I-\mu\Lambda]v(m) \tag{7.92}$$

As Λ and I are both diagonal matrices, Equation (7.92) can be expressed in terms of the equations for the individual elements of the error vector $v(m)$ as

$$v_{k}(m+1)=[1-\mu\lambda_{k}]v_{k}(m) \tag{7.93}$$

where λ_{k} is the kth eigenvalue of the autocorrelation matrix of the filter input $y(m)$. Figure 7.4 is a feedback network model of the time variations of the error vector. From Equation (7.93), the condition for the stability of the adaptation process and the decay of the coefficient error vector is

$$-1<1-\mu\lambda_{k}<1 \tag{7.94}$$

Let λ_{max} denote the maximum eigenvalue of the autocorrelation matrix of $y(m)$ then, from Equation (7.94) the limits on μ for stable adaptation are given by

$$0 < \mu < \frac{2}{\lambda_{max}} \tag{7.95}$$

Convergence Rate The convergence rate of the filter coefficients depends on the choice of the adaptation step size μ, where $0 < \mu < 1/\lambda_{max}$. When the eigenvalues of the correlation matrix are unevenly spread, the filter coefficients converge at different speeds: the smaller the kth eigenvalue the slower the speed of convergence of the kth coefficients. The filter coefficients with maximum and minimum eigenvalues, λ_{max} and λ_{min} converge according to the following equations:

$$v_{max}(m+1) = (1 - \mu \lambda_{max}) v_{max}(m) \tag{7.96}$$

$$v_{min}(m+1) = (1 - \mu \lambda_{min}) v_{min}(m) \tag{7.97}$$

The ratio of the maximum to the minimum eigenvalue of a correlation matrix is called the eigenvalue spread of the correlation matrix:

$$\text{eigenvalue spread} = \frac{\lambda_{max}}{\lambda_{min}} \tag{7.98}$$

Note that the spread in the speed of convergence of filter coefficients is proportional to the spread in eigenvalue of the autocorrelation matrix of the input signal.

7.5 The LMS Filter

The steepest-descent method employs the gradient of the *averaged* squared error to search for the least square error filter coefficients. A computationally simpler version of the gradient search method is the least mean square (LMS) filter, in which the gradient of the *mean* square error is substituted with the gradient of the *instantaneous* squared error function. The LMS adaptation method is defined as
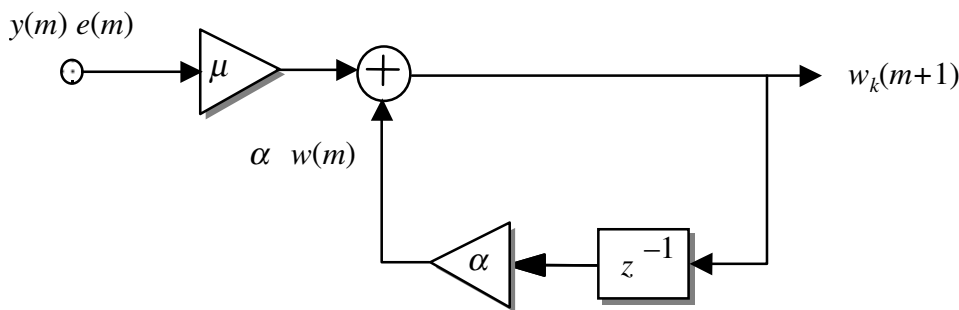
Figure 7.5 Illustration of LMS adaptation of a filter coefficient.

$$w(m+1) = w(m) + \mu \left(-\frac{\partial\, e^2(m)}{\partial\, w(m)} \right) \tag{7.99}$$

where the error signal $e(m)$ is given by

$$e(m) = x(m) - w^{\mathrm{T}}(m)x(m) \tag{7.100}$$

The instantaneous gradient of the squared error can be re-expressed as

$$\frac{\partial e^2(m)}{\partial w(m)} = \frac{\partial}{\partial w(m)}[x(m) - w^{\mathrm{T}}(m)y(m)]^2$$
$$= -2y(m)[x(m) - w^{\mathrm{T}}(m)y(m)]^2 \tag{7.101}$$
$$= -2y(m)e(m)$$

Substituting Equation (7.101) into the recursion update equation of the filter parameters, Equation (7.99) yields the LMS adaptation equation:

$$w(m+1) = w(m) + \mu\left[y(m)e(m)\right] \tag{7.102}$$

It can seen that the filter update equation is very simple. The LMS filter is widely used in adaptive filter applications such as adaptive equalisation, echo cancellation etc. The main advantage of the LMS algorithm is its simplicity both in terms of the memory requirement and the computational complexity which is $O(P)$, where P is the filter length.

Leaky LMS Algorithm The stability and the adaptability of the recursive LMS adaptation Equation (7.86) can improved by introducing a so-called leakage factor α as

$$w(m+1)=\alpha w(m)+\mu\left[y(m)e(m)\right] \qquad (7.103)$$

Note that the feedback equation for the time update of the filter coefficients is essentially a recursive (infinite impulse response) system with input $\mu y(m)e(m)$ and its poles at α. When the parameter $\alpha<1$, the effect is to introduce more stability and accelerate the filter adaptation to the changes in input signal characteristics.

Steady-State Error: The optimal least mean square error (LSE), $E_{\min}$, is achieved when the filter coefficients approach the optimum value defined by the block least square error equation $w_{\mathrm{o}}=R_{yy}^{-1}r_{yx}$ derived in Chapter 6. The steepest-decent method employs the average gradient of the error surface for incremental updates of the filter coefficients towards the optimal value. Hence, when the filter coefficients reach the minimum point of the mean square error curve, the *averaged* gradient is zero and will remain zero so long as the error surface is stationary. In contrast, examination of the LMS equation shows that for applications in which the LSE is non-zero such as noise reduction, the incremental update term $\mu e(m)y(m)$ would remain non-zero even when the optimal point is reached. Thus at the convergence, the LMS filter will randomly vary about the LSE point, with the result that the LSE for the LMS will be in excess of the LSE for Wiener or steepest-descent methods. Note that at, or near, convergence, a gradual decrease in μ would decrease the excess LSE at the expense of some loss of adaptability to changes in the signal characteristics.

7.6 Summary

This chapter began with an introduction to Kalman filter theory. The Kalman filter was derived using the orthogonality principle: for the optimal filter, the innovation sequence must be an uncorrelated process and orthogonal to the past observations. Note that the same principle can also be used to derive the Wiener filter coefficients. Although, like the Wiener filter, the derivation of the Kalman filter is based on the least squared error criterion, the Kalman filter differs from the Wiener filter in two respects.

First, the Kalman filter can be applied to non-stationary processes, and second, the Kalman theory employs a model of the signal generation process in the form of the state equation. This is an important advantage in the sense that the Kalman filter can be used to explicitly model the dynamics of the signal process.

For many practical applications such as echo cancellation, channel equalisation, adaptive noise cancellation, time-delay estimation, etc., the RLS and LMS filters provide a suitable alternative to the Kalman filter. The RLS filter is a recursive implementation of the Wiener filter, and, for stationary processes, it should converge to the same solution as the Wiener filter. The main advantage of the LMS filter is the relative simplicity of the algorithm. However, for signals with a large spectral dynamic range, or equivalently a large eigenvalue spread, the LMS has an uneven and slow rate of convergence. If, in addition to having a large eigenvalue spread a signal is also non-stationary (e.g. speech and audio signals) then the LMS can be an unsuitable adaptation method, and the RLS method, with its better convergence rate and less sensitivity to the eigenvalue spread, becomes a more attractive alternative.

Bibliography

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

BELLANGER M.G. (1988) Adaptive Filters and Signal Analysis. Marcel-Dekker, New York.

BERSHAD N.J. (1986) Analysis of the Normalised LMS Algorithm with Gaussian Inputs. IEEE Trans. Acoustics Speech and Signal Processing, **ASSP-34**, pp. 793–807.

BERSHAD N.J. and QU L.Z. (1989) On the Probability Density Function of the LMS Adaptive Filter Weights. IEEE Trans. Acoustics Speech and Signal Processing, **ASSP–37**, pp. 43–57.

CIOFFI J.M. and KAILATH T. (1984) Fast Recursive Least Squares Transversal Filters for Adaptive Filtering. IEEE Trans. Acoustics Speech and Signal Processing, **ASSP-32**, pp. 304–337.

CLASSEN T.A. and MECKLANBRAUKER W.F., (1985) Adaptive Techniques for Signal Processing in Communications. IEEE Communications, **23**, pp. 8–19.

COWAN C.F. and GRANT P.M. (1985) Adaptive Filters. Prentice-Hall, Englewood Cliffs, NJ.

EWEDA E. and MACCHI O. (1985) Tracking Error Bounds of Adaptive Non-sationary Filtering. Automatica, **21**, pp. 293–302.

GABOR D., WILBY W. P. and WOODCOCK R. (1960) A Universal Non-linear Filter, Predictor and Simulator which Optimises Itself by a Learning Process. IEE Proc. **108**, pp. 422–38.

GABRIEL W.F. (1976) Adaptive Arrays: An Introduction. Proc. IEEE, **64**, pp. 239–272.

HAYKIN S.(1991) Adaptive Filter Theory. Prentice Hall, Englewood Cliffs, NJ.

HONIG M.L. and MESSERSCHMITT D.G. (1984) Adaptive Filters: Structures, Algorithms and Applications. Kluwer Boston, Hingham, MA.

KAILATH T. (1970) The Innovations Approach to Detection and Estimation Theory, Proc. IEEE, **58**, pp. 680–965.

KALMAN R.E. (1960) A New Approach to Linear Filtering and Prediction Problems. Trans. of the ASME, Series D, Journal of Basic Engineering, **82**, pp. 34–45.

KALMAN R.E. and BUCY R.S. (1961) New Results in Linear Filtering and Prediction Theory. Trans. ASME J. Basic Eng., **83**, pp. 95–108.

WIDROW B. (1990) 30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Back Propagation. Proc. IEEE, Special Issue on Neural Networks I, **78**.

WIDROW B. and STERNS S.D. (1985) Adaptive Signal Processing. Prentice Hall, Englewood Cliffs, NJ.

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

ZADEH L.A. and DESOER C.A. (1963) Linear System Theory: The State-Space Approach. McGraw-Hill, New York.