DOUBLY ADAPTIVE FILTERS FOR NONSTATIONARY APPLICATIONS

by

S. Douglas Peters
B.Eng., TUNS, 1986
M.Sc.(Eng.), Queen's, 1988

Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY
in the Department of
Electrical and Computer Engineering

We accept this dissertation as conforming
to the required standard

Dr. A. Antoniou, Supervisor (ECE Dept.)

Dr. P. Agathoklis, Departmental Member (ECE Dept.)

Dr. V. K. Bhargava, Departmental Member (ECE Dept.)

Dr. Z. Dong, Outside Member (ME Dept.)

Dr. W. B. Mikhael, External Examiner (Univ. of Central Florida, EE Dept.)

Supervisor: Dr. A. Antoniou

ABSTRACT

This dissertation examines the performance of self-tuning adaptive filters in nonstationary environments and deals with extensions to conventional adaptive filters that lead to enhanced performance. A number of the available self-tuning adaptive filters, called doubly adaptive filters for the present purposes, are critically examined and three new schemes are proposed. The first and second are based on the normalized least-mean-squares (NLMS) adaptive filter, and their formulations are contrived to minimize the misadjustment in a convergent scenario and random walk scenario, respectively. The first of these filters, called reduced adaptation state estimation (RASE), achieves performance near that of the recursive-least squares (RLS) algorithm under known additive noise statistics and moderately correlated input samples. The development of the second proposed filter introduces the idea of having more than one adaptive filter applied in parallel to the same input and desired signals. This concept, called parallel adaptation (PA), is applied in both NLMS and RLS contexts in order to achieve optimal steady-state misadjustment in a random walk scenario. Numerous simulation results are presented that support the present analysis and demonstrate the effectiveness of the proposed algorithms in a number of different nonstationary environments.

Examiners:

Dr. A. Antoniou, Supervisor (ECE Dept.)

Dr. P. Agathoklis, Departmental Member (ECE Dept.)

Dr. V. K. Bhargava, Departmental Member (ECE Dept.)

Dr. Z. Dong, Outside Member (ME Dept.)

Dr. W. B. Mikhael, External Examiner (Univ. of Central Florida, EE Dept.)

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AGC	Automatic gain control
DA	Doubly adaptive
DCF	Damped convergence factor
DSA	Dual-sign algorithm
HCB	Harris, Chabries, and Bishop
GAS	Gradient adaptive stepsize
EKF	Extended Kalman filter
FIR	Finite-duration impulse response
FKY	Fortescue, Kershenbaum, and Ydstie
KJ	Kwong and Johnson
LMS	Least-mean squares
MAC	Misadjustment at convergence
MSE	Mean-squared error
MMSE	Minimum mean-squared error
NLMS	Normalized least-mean squares
NNR	Normalized nonstationarity to noise ratio
PA	Parallel adaptation
RASE	Reduced adaptation state estimation
RLS	Recursive-least squares
SKYW	Shan, Kailath, Ye, and Wu
VS	Variable step

The sunshine promises a bounteous feast
Come harvest-time. Look, see the ripening!
The tiny seeds have now been much increased,
By toils of Summer and the rains of Spring.

The fruit well-hidden 'midst the leaves above
Recalls the contributions of those who
Sowed generously, giving time and love,
As soil was tilled and fragile seedlings grew.

I thank, in consequence, my patient wife,
Dear Sharon, who still helps my mind to grow;
Dear Frank and Alice, who gave more than life;
And supervisor, A. Antoniou.

But most of all, I humbly thank the One
True God, who bids all flourish through His Son.

May, 1993

S. D. Peters

To Grace & Meg,

who keep me

wondering

Chapter 1

Introduction

1.1 Adaptive Filters

1.1.1 Introduction

An adaptive filter is an adjustable filter equipped with a mechanism for the purpose of minimizing either explicitly or implicitly, the mean-squared error (MSE) or the variance of the difference between the filter output and some reference (so-called desired) signal. The basic adaptive filter is depicted in Figure 1.1. Finite-duration impulse response (FIR) adjustable filters, and in particular transversal-structure filters are treated. While other structures have been recommended for adaptive filters for various reasons, the object of this thesis is to introduce an approach to the treatment of adaptive filters in nonstationary environments. In consequence, the most common and simple filter structure is adopted from which the present development can proceed without undue clutter.

The essence of an adaptive filter is embodied in the adaptation scheme by which the adjustable filter coefficients are modified. Not surprisingly, these methods are generally analogous to optimization techniques. Specifically, Cauchy's steepest-descent method [1] and Newton's method (see, e.g., [2]) are represented in common adaptive filters. The former technique has an adaptive filter counterpart in the so-called

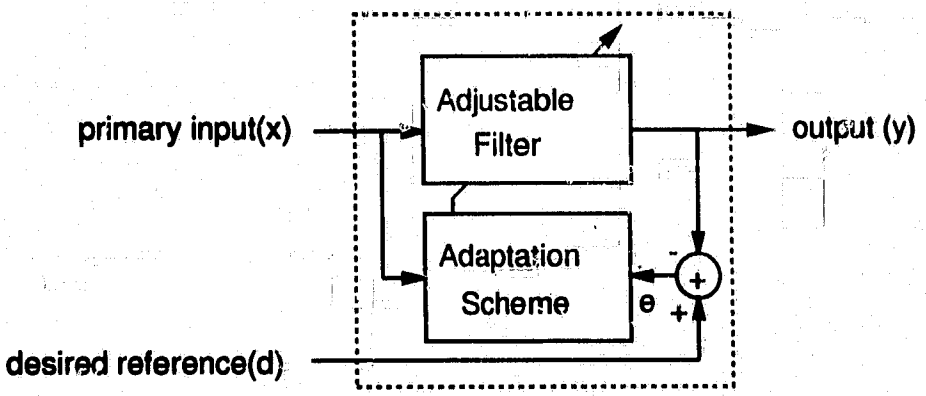

Figure 1.1: Adaptive filter

least-mean-squares (LMS) algorithm, first introduced by Widrow and Hoff in 1960 [3]. The latter optimization scheme is perhaps best represented in the adaptive filtering literature by the so-called recursive-least squares (RLS) algorithm. This scheme, a simplification of the Kalman filter [4], follows directly from Woodbury's identity [5].

There are a number of good textbooks providing excellent introductions to the subject of adaptive filtering [6] [7] [8] [9]. In consequence, the present introduction mainly concerns itself with the presentation of necessary notation and terminology.

Throughout this work, the attempt is made to examine adaptive filters in an application independent manner. Though simulations will necessarily involve an application model and comments will be made with regard to certain practical considerations, the applicability of the algorithms to be discussed will be left to the references and subsequent research.

Consider, then, an adaptive filter applied to a transversal filter of order $N-1$. The filter tap weights are denoted by the vector $\hat{\mathbf{w}}_k$. The symbol $\mathbf{x}_k$ is used for the filter input, i.e.,

$$\mathbf{x}_k = [x(k-N+1)\ \ x(k-N+2)\ \cdots\ x(k)]^T \tag{1.1}$$

where $x(k)$ is the filter input at sample k. The filter input is taken to come from a

zero-mean random process with variance σ_x^2 and for which the probability of $\| \mathbf{x}_k \| = 0$ is zero for all k. At every iteration, the output of this filter is given by $y_k = \hat{\mathbf{w}}_k^T \mathbf{x}_k$ where superscript T denotes transpose. The error signal is given by

$$e_k \triangleq d_k - y_k = \mathbf{x}_k^T(\mathbf{w}_k - \hat{\mathbf{w}}_k) + \epsilon_k, \tag{1.2}$$

where d, the desired response, is taken to be the sum of the output of a nonrecursive target filter (the Wiener filter) of order $N-1$ having tap weights $\mathbf{w}_k$ and additive noise ϵ which is assumed to be stationary, white, Gaussian with unknown variance σ_ϵ^2, and independent of the filter input. This simplification, which is common in the literature, is equivalent to taking the differences between the Wiener filter outputs and the desired response to be subsumed in the additive noise. This can be quite a poor assumption in some important adaptive filter applications, but continues to be used for the tractability that it affords. Finally, the weight error vector, $\mathbf{v}_k = \mathbf{w}_k - \hat{\mathbf{w}}_k$, is defined for convenience.

1.1.2 The LMS Algorithm

Optimization with respect to the mean-squared error is complicated by the fact that neither the objective function nor its derivatives with respect to the filter tap weights are available. The standard LMS solution to this problem is to use an instantaneous estimate of the gradient of the MSE, namely, $\nabla_{\hat{\mathbf{w}}} E(e_k^2) \approx \nabla_{\hat{\mathbf{w}}} e_k^2 = -2e_k\mathbf{x}_k$ in a steepest-descent manner [3]. That is, the filter weights are modified at each iteration by a small step in the opposite direction of this gradient estimate, i.e.,

$$\hat{\mathbf{w}}_{k+1} = \hat{\mathbf{w}}_k + \mu_k e_k \mathbf{x}_k \tag{1.3}$$

where μ_k, the so-called stepsize or convergence-controlling parameter, is used because our inability to measure the objective function precludes the use of any line search.

For standard LMS adaptation, μ_k is fixed. Variants on this include a large number of LMS-based algorithms for which the explicit sample dependence is necessary. For example, the normalized LMS (NLMS) method [10] [11], in which $\mu_k = \bar{\mu}/(\gamma + \mathbf{x}_k^T \mathbf{x}_k)$, will soon be examined. The small constant γ is used in practice to avoid division by zero, but for the present purposes this quantity is assumed to be zero. The quantity $\bar{\mu}$ is then considered to be the NLMS convergence-controlling parameter. The advantage of the NLMS algorithm is that no knowledge of the input statistics is necessary in order to guarantee convergence under persistent excitation [12]. The term *persistent excitation* refers to the property of an input signal that permits the unknown modes of the adaptive filter to be tested and subsequently modified, providing a unique Wiener filter. If the two-sided spectrum of a signal is nonzero at n points, the signal is said to be weakly persistently exciting of order n [13]. If an input signal is not persistently exciting, there can be no guarantee of convergence (for a more thorough discussion of this topic, see [6]).

1.1.3 The RLS Algorithm

The conventional RLS adaptive filter (see, e.g., [6]) recursively minimizes the objective function

$$\sum_{j=0}^{k} \lambda^{k-j} \varepsilon_j^2$$

where λ is the so-called forgetting factor, and

$$\varepsilon_k = d_k - \hat{\mathbf{w}}_{k+1}^T \mathbf{x}_k$$

are the *a posteriori* output errors. This minimization is accomplished via the standard Wiener-Hopf solution [14]

$$\hat{\mathbf{w}}_{k+1} = \left(\sum_{j=0}^{k} \lambda^{k-j} \mathbf{x}_j \mathbf{x}_j^T \right)^{-1} \left(\sum_{j=0}^{k} \lambda^{k-j} d_j \mathbf{x}_j \right).$$

This solution is obtained recursively using the standard RLS iteration

$$\hat{\mathbf{w}}_{k+1} = \hat{\mathbf{w}}_k + \frac{\mathbf{P}_k \mathbf{x}_k e_k}{\lambda + \mathbf{x}_k^T \mathbf{P}_k \mathbf{x}_k}$$
$$\mathbf{P}_{k+1} = \frac{1}{\lambda}\left(\mathbf{P}_k - \frac{\mathbf{P}_k \mathbf{x}_k \mathbf{x}_k^T \mathbf{P}_k}{\lambda + \mathbf{x}_k^T \mathbf{P}_k \mathbf{x}_k}\right) \tag{1.4}$$

where e_k are the previously defined *a priori* output errors. While the RLS procedure explicitly minimizes a weighted sum of the *a posteriori* errors, the performance criterion of interest is still the mean-squared *a priori* error (MSE), which measures the adaptive filter's ability to anticipate a desired behaviour. The matrix $\mathbf{P}$ is a scaled estimate of the inverse of the input covariance matrix, i.e.,

$$\mathbf{R} = E\left(\mathbf{x}_k \mathbf{x}_k^T\right).$$

For the benefit of a unified terminology, the RLS forgetting factor is referred to as that algorithm's convergence-controlling parameter. While the choice of λ is not usually made with convergence in mind, the ability of the RLS algorithm to reconverge after a sudden change in the statistics of the desired signal depends directly on its value.

As one might expect from basic optimization theory, the quasi-Newton-like RLS procedure provides better convergence behaviour than the steepest-descent-like LMS algorithm. This is especially true in the case when the contours of the objective MSE are not hyper-spherical [7]. This occurs when the samples of the input signal are correlated.

1.2 Adaptation State and Adaptation Environment

The term *adaptation state* is used to describe the operating state of an adaptive filter and consists of the instantaneous weight error vector $\mathbf{v}_k$. In practice, however,

a derivation of a reasonable estimate of this quantity cannot be expected, and so one must settle for an estimate of its second order statistics, $E(\mathbf{v}_k\mathbf{v}_k^T)$. Further, the estimation of this entire matrix is impractical. As a result, the *reduced adaptation state* to be considered consists of its trace.

The *adaptation environment* consists of the statistical properties of the input and desired signals. While much of the analysis of adaptive filters in the literature is based upon stationary environments, the application of adaptive filters is especially advantageous when the environment is nonstationary. In general, environmental nonstationarities can be categorized according to whether the statistics of the filter input, the desired response or both change with respect to time. Throughout this work, the filter input is taken to be stationary, although this is not necessary for the applicability of the following algorithms. Further, the additive noise is taken to be relatively stationary by comparison to the desired signal. Models of nonstationarity, in consequence, become models of target filter weight behaviour. In the literature, the first-order Markov model is most common. Moreover, this model is often simplified to become the random walk model, following [7], in order to improve the tractability of the problem.

This approach immediately has consequences in the analysis of the two algorithms to be considered. For example, the time-dependence of μ in the NLMS procedure provides robustness in the presence of unknown or nonstationary input rather than improved performance in an otherwise nonstationary environment. The possibility of an NLMS filter for which $\bar{\mu}$ is time-dependent in an attempt to provide performance enhancement in desired signal nonstationarities will be considered shortly. Further, the RLS $\hat{\mathbf{w}}$-update and $\mathbf{P}$-update equations (1.4) can clearly be decoupled since the latter is independent of the desired signal. In consequence, the distinction between a

steady state with respect to the input and that with respect to the desired signal is made. Without qualification, the term *steady state* denotes the latter since the input is taken to be stationary.

1.3 Performance Evaluation

As one might expect, the MSE resulting from the application of an adaptive filter can never be less than the variance of the additive noise. Of course, this quantity is also the MSE resulting from the substitution of the adaptive filter by the Wiener filter, and is often referred to as the minimum MSE (MMSE). The *excess* MSE, in consequence, is simply the difference between the actual and minimum MSE's. Moreover, the *misadjustment*, one of the most useful quantities by which to gauge the performance of an adaptive filter, is simply a normalized excess MSE: the ratio of excess MSE to MMSE. The *learning curve* of an adaptive filter is simply the curve representing the expected misadjustment against time.

When adaptive filters are applied in unknown stationary environments, the most meaningful measures of performance are the time that it takes the adaptation algorithm to converge and the misadjustment at convergence (MAC). Indeed, a trade-off exists in conventional adaptive filters between these two performance criteria. In consequence, the learning curves for conventional adaptive filters (neglecting initialization effects for RLS adaptation) belong qualitatively to the set shown by the solid lines in Figure 1.2: either they both converge quickly and level off at some high value of MAC, or they provide good MAC performance after taking some time to achieve that level.

In nonstationary environments, on the other hand, there is a performance trade-off between the so-called estimation and lag errors, the two components of excess MSE

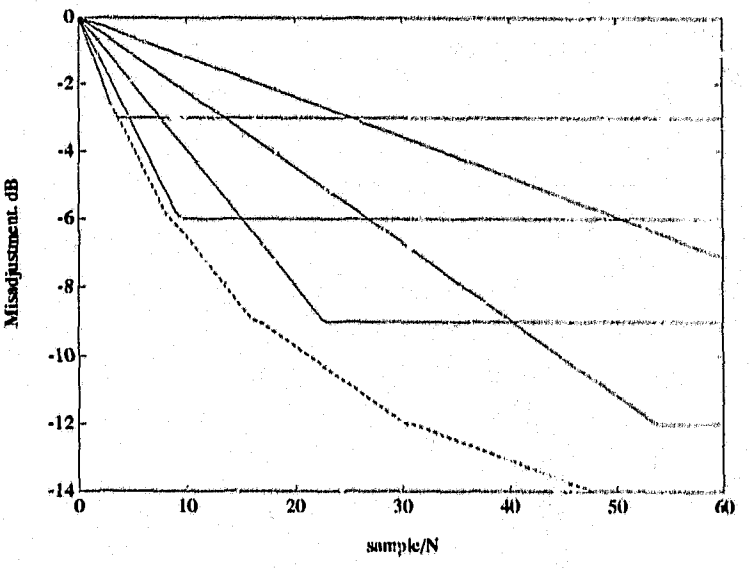

Figure 1.2: Adaptive filter learning curves

in the presence of a nonstationarity. In effect, adaptive filters that are designed to perform well in quickly varying situations perform poorly when their environment is stationary and *vice versa*. The estimation error is a measure of how well the adaptive filter performs in a stationary environment while the lag error measures the ability of the adaptive filter to respond to a highly nonstationary environment. The minimum mean-squared-error, the sum of these error quantities, can, in principle, be obtained by tuning the convergence-controlling parameter of the conventional adaptive filter [15] [16]. Unfortunately, this procedure is rarely practical. Real nonstationarities are unlikely to comply with an experimental tuning procedure. On the other hand, even if an accurate model of the nonstationarity in question were available, an analytic tuning method is difficult to obtain and may not be forthcoming. Indeed, such methods exist only for most simple and artificial nonstationarities. In particular, the random walk limit of the first-order Markov model of nonstationarity is the most frequently examined [15] [16] [17].

1.4 Doubly Adaptive Filters

The introduction of the term *doubly adaptive* (DA) filter to refer to adaptive filters whose convergence-controlling parameter is permitted to vary with time is problematic. The difficulty arises due to the fact that this is precisely the description of what used to be called "adaptive filters." Indeed, the term "filter" (usually meaning a Kalman-type filter) was used in the literature of a number of decades ago to refer to what is presently called an adaptive filter. For the present, however, the term will be adopted.

The level at which a doubly adaptive filter adjusts its tap-weights is termed *primary adaptation.* The adjustment of the convergence-controlling parameter of the primary adaptation process, on the other hand, is referred to as *secondary adaptation.* Further, only certain algorithms in which secondary adaptation involves memory will be considered. For example, the NLMS algorithm itself could be regarded as a doubly adaptive filter with the secondary adaptation taking the form of normalization. This and related algorithms, including those due to Mikhael, Diniz and their respective co-workers [18] [19], have no memory associated with their secondary adaptation and consequently behave like singly adaptive algorithms in nonstationary environments. That is, they are robust and predictable, but subject to the same convergence-MAC performance tradeoff as more conventional algorithms. As a result, these methods will be treated as singly adaptive for the present purposes. Moreover, only those algorithms whose secondary adaptation is principly data-dependent will be considered. The convergence-controlling parameters in the algorithms of [20] and [21] are governed in a deterministic and semi-deterministic manner respectively and, in consequence, algorithms of this type will not be addressed.

1.5 Outline of Dissertation

The present work is organized as follows: Chapter 2 summarizes current doubly adaptive filters and their shortcomings. Chapters 3, 4, and 5 present new DA algorithms based on conventional NLMS and RLS adaptive filters. These schemes govern the convergence-controlling parameters ($\bar{\mu}_k$ and λ_k, respectively) of those algorithms in an attempt to minimize the excess MSE in a model environment.

The first adaptation environment model to be considered is one in which the desired signal is piecewise stationary. That is, sudden changes in target filter are considered. In this environment, the attempt is to achieve a learning curve following the dashed line in Figure 1.2. Indeed, if the convergence-controlling parameter were governed properly, this curve is clearly obtainable. As will be seen, the RLS adaptive filter can attain such performance by a number of means. An NLMS-based algorithm that approximates this learning curve is presented in Chapter 3. This scheme, involving an explicit estimate of the trace of the instantaneous weight error covariance, is called *reduced adaptation state estimation* (RASE).

Subsequently, the random walk model of the environment is considered. In this event, the convergence-controlling parameter of the conventional algorithm is governed so as to "tune" the algorithm to provide the optimum steady-state misadjustment such that the sum of estimation and lag errors is minimized. Two algorithms that approximately achieve this optimal tracking are presented in Chapters 4 and 5. These schemes require additional complexity in the form of independent or semi-independent adaptive filters applied in parallel to the same signals. In consequence, they will be termed *parallel adaptation* (PA), and the PA-NLMS and PA-RLS algorithms will be based on NLMS and RLS conventional adaptive filters respectively.

In Chapter 6, the performance of the proposed algorithms is compared to that of a

number of existing DA methods via a suite of simulations. All aspects of adaptive filter performance are examined using a number of simulated nonstationary environments.

Chapter 2

Existing Doubly Adaptive Filters

2.1 Introduction

A number of doubly adaptive filters have been proposed in the literature in the last decade or so. Unfortunately, it is not entirely obvious from these papers what their advantages or shortcomings might be. Indeed, the comparison of DA algorithms with conventional adaptive filters or other DA schemes is somewhat problematic due to the fluidity of the various performance evaluation criteria that have been used. After a brief discussion of these difficulties, a qualitative examination of some of the more recent DA methods will be presented.

2.1.1 Sensitivity to Algorithmic Parameters

Doubly adaptive filters govern the convergence-controlling parameter of the "parent" adaptive filter on which they are based. Presumably, this "tuning" takes place because the engineer who designed the filter wanted the convergence-controlling parameter to be continually optimal for the environment at hand. In other words, since this engineer could not foresee what value, or sequence of values, would be needed for the best performance of the adaptive filter, a DA algorithm was selected in order to approximate this best sequence according to some sensible scheme.

Unfortunately, all doubly adaptive methods also have algorithmic parameters for which values need to be selected. If the selection of these values is as or more difficult than the selection of the best sequence of convergence-controlling parameter values, then there is no advantage to using the DA scheme in the first place. In consequence, it is important that DA methods be introduced with guidelines by which to choose algorithmic parameters and that some investigation into the sensitivity of the proposed algorithm to those choices be supplied. Unfortunately, a number of existing DA methods have been presented without any attempt to provide such information. In these cases, the simulation results that are used to demonstrate the performance of the given DA method may have involved careful tuning of a number of parameters. In some cases, the success of the method was entirely dependent on this careful tuning due to a high degree of sensitivity to parameter values. Unfortunately, this tuning procedure is rarely feasible in practice, and so its use gives the DA algorithm an unfair advantage against conventional schemes for which selection of the convergence-controlling parameter is usually rather straightforward. More often than not, the choice of DA algorithmic parameters depends on some *a priori* knowledge of the adaptation environment. However, the implications of this necessary knowledge are sometimes left unpublished.

2.1.2 Sensitivity to Environment

Another difficulty with the comparison between DA methods is their range of sensitivity to the adaptation environment. For example, it is well known that LMS-based conventional algorithms are quite sensitive to the correlatedness of the filter input samples. Can one expect the sensitivity of LMS-based doubly adaptive methods to this condition to be equivalent to that of their parent algorithm? For some simple DA methods, it is clear that this is the case. For more involved DA methods, however,

the answer to this question is less clear. In fact, this issue has not been addressed in any of the appropriate references. As shall be demonstrated in the simulations, there is often no advantage in the existing DA procedures for highly coloured input. The correlatedness of the input signal is just one aspect of the adaptation environment for which the sensitivity of DA algorithms is both important and unknown.

Due to the latitude available in the specification of a nonstationary environment, an apparent advantage of one adaptive filter over another can be manufactured by simply choosing to fix those aspects of the environment to which the algorithm of choice is most sensitive. In this manner, any of the DA algorithms to be considered seem to deliver performance improvements over their parent filter. In most cases, the performance advantage of the DA method is legitimate. As shall be demonstrated, however, there is no real benefit to using some of the existing DA methods over conventional adaptive filters. Of course, it is just as true that a given algorithm can be shown to perform poorly by focusing on those environmental aspects to which that method is most sensitive. Resisting that temptation, the attempt has been made in Chapter 6 to provide a reasonably fair comparison between existing and proposed DA methods.

2.1.3 DA Performance Evaluation Difficulties

As has been discussed previously, adaptive filter performance is measured by a number of criteria: convergence, MAC, tracking and reconvergence. Unfortunately, not all of these have universally accepted definitions. Only misadjustment at convergence needs no clarification.

The term "tracking", for example, depends on the nonstationarity to be tracked. Since the simplest nonstationarities are the most artificial, it comes as no surprise that researchers cannot agree on how exactly an adaptive filter's tracking performance

is to be measured.

An examination of the term "convergence" reveals another dilemma. Conventionally, this term has meant the time it takes the adaptive filter to attain its MAC in expectation. From Figure 1.2, it is clear that this is most appropriate for conventional algorithms for which the learning curve is the sum of constant and exponential terms. On the other hand, such a definition for convergence is quite inappropriate for the "optimal" DA filter whose learning curve is stylized by the dashed line in Figure 1.2, since MAC, zero in this case, is only achieved asymptotically.

A further difficulty is introduced by the effects of algorithm initialization. For conventional LMS-based filters, convergence and reconvergence behaviours are identical. For RLS and many DA filters, on the other hand, the convergence behaviour (from $k = 0$) is much better than the algorithm's ability to reconverge after a sudden change in the statistics of the desired signal. In consequence, the distinction is made between *first convergence* and *initial convergence* as follows. The term *first convergence* denotes all that occurs from $k = 0$ until the steady state including initialization effects. Since nonstationary environments are the present concern, little credit is given here to algorithms with good first convergence properties. Further, since the initialization of DA schemes often requires addition parameter selection, the elimination of first convergence as a performance criterion helps to mitigate the problem discussed in section 2.1.1. Since the response of an adaptive filter to a sudden change in the statistics of the desired signal is to be given greater consideration than first convergence, the term *initial convergence* is used to denote the slope of the learning curve directly after such an occurrence.

2.2 Existing LMS-based Algorithms

2.2.1 Algorithm of Harris Chabries and Bishop

The variable step (VS) algorithm proposed by Harris Chabries and Bishop in [22] is, perhaps, the most widely cited of all DA algorithms. To avoid confusion between this and other "variable step" algorithms to be considered, this scheme is referred to as the HCB algorithm. This DA filter is based on rather old stochastic approximation methods [23].

Summary

First of all, the HCB scheme maintains a separate convergence-controlling parameter for each tap-weight of the adaptive filter. In effect, the μ_k in (1.3) is replaced by a matrix whose diagonal entries are μ_k^i. The sign changes of the individual components of the gradient estimate are then monitored. If the sign changes in a given number of consecutive samples, the step-size corresponding to that component is decreased. On the other hand, the convergence-controlling parameter for any component is increased if the corresponding component of the gradient estimate has the same sign over a number of consecutive samples. Variants of this idea have been proposed in [24] and [25].

Critique

Unlike most of the other DA algorithms to be considered, the HCB scheme does not require any knowledge of its environment to choose values for its various algorithmic parameters. While this is a definite point in its favour, the algorithm remains quite sensitive to these choices. While reasonable recommendations for the values of these parameters are offered in [22], the applicability of this algorithm is still limited by its

sensitivity to them.

The most problematic aspect of this algorithm is the fact that it can be particularly unrobust in the presence of correlated inputs, in direct contradiction to the conjecture of Harris et al. [22]. This may be easily demonstrated by the examination of a limiting case. Consider the first-order (two tap-weights) adaptive transversal filter whose input samples are highly correlated such that the dominant eigenvector of the input covariance matrix is $[1 \quad -1]^T$. Further, let the weight error vector at some sample be in the direction $[1 \quad 1]^T$ with some large magnitude. Under these conditions, the input vector may be expressed as a sum of the input covariance eigenvectors, viz.,

$$\mathbf{x}_k = x_{1,k} \begin{bmatrix} 1 \\ -1 \end{bmatrix} + x_{2,k} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

where $E|x_{1,k}| \gg E|x_{2,k}|$. The individual components of the standard LMS gradient estimate are then approximately some scaled form of the product $x_{1,k}x_{2,k}$. For a transversal filter, the sign of $x_{2,k}$ is expected to be the same between adjacent samples. The sign of $x_{1,k}$, on the other hand, will almost certainly change from sample to sample. In consequence, each component of the gradient estimate will continually be changing sign, resulting in lower values of μ_k^i for all i. Unfortunately, since $\mathbf{v}$ has a large magnitude, one would prefer the opposite effect to take place. In consequence, in the presence of highly correlated inputs, the convergence behaviour of the HCB algorithm applied to a transversal filter is very poor indeed.

Another difficulty with the papers that present the HCB algorithm and its variants is the way in which these algorithms are compared to the standard LMS adaptive filter [22] [24] [25]. In all of the simulations in these papers, the convergence-controlling parameter for the LMS algorithm is taken to equal the lower limit set for those of the VS algorithm. This would, of course, be fair if the variable convergence-controlling parameters would converge to this quantity in a stationary steady state.

Unfortunately, they do not unless the input samples are highly correlated as shown above. In fact, with uncorrelated input samples, the expected value of the variable convergence-controlling parameters is quite a bit larger than this lower limit. In consequence, the LMS stationary steady-state misadjustment is always lower than that of the variable step algorithms and it comes as no surprise that the variable step convergence is superior.

The same difficulty occurs toward the upper limit of the convergence-controlling parameters. The only reason that the variable step algorithm converges quickly is that these parameters are set to their upper limit initially. In the presence of a sudden change in the statistics of the desired signal, the reconvergence of the variable step algorithm is, in fact, worse than that of the basic LMS algorithm providing the same steady-state misadjustment.

2.2.2 Algorithm of Kwong

Summary

The dual-sign algorithm (DSA) proposed by Kwong in [26] is a simple proto-DA method that permits the convergence-controlling parameter of the sign-error primary adaptation scheme to take one of two values depending on the magnitude of the error signal at any given iteration.

Critique

The HCB algorithm considered above is qualitatively between the DSA algorithm and the DA methods to follow in that it permits the convergence-controlling parameters to take one of a finite set of values. The DSA scheme, on the other hand, limits its convergence-controlling parameter to one of two values while subsequently considered

algorithms permit a continuous variation in their respective convergence-controlling parameters.

The major advantage of the DSA procedure is its simplicity and (lack of) computational complexity. Its immediate drawback, however, is the necessity to have a good estimate of the additive noise power. Indeed, to follow the recommendations on the choice of other algorithmic parameters, it is necessary to have some estimate of the initial adaptation state. These requirements have been sufficiently documented in [26], but the sensitivity of DSA performance to the estimates of these quantities is a distinct drawback. In Chapter 3, a new algorithm is developed that also requires such estimates. It shall be demonstrated, however, that the resulting algorithm is relatively insensitive to the initial estimate of the adaptation state and provides near-optimal performance when a reasonable estimate of the additive noise power is available.

2.2.3 Algorithm of Shan and Kailath

The so-called automatic gain control (AGC) algorithm proposed by Shan and Kailath in [27] has received some attention in the literature. A number of corrections are documented in [28] and [29].

Summary

This algorithm is an *ad hoc* method of reducing the step size when the cross-correlation between input and error signal is low, and increasing it when this quantity is high. Unfortunately, the estimate of cross-correlation is not particularly good, resulting in a modification of this procedure in [30]. A variant of the algorithm found in [30], which performs considerably better than either of its forebears, maintains exponentially weighted estimates of the quantities that are necessary to calculate the correlation coefficient between the error signal and each element of the input vector. The average

of the absolute values of the resulting correlation coefficient estimates is then used as the convergence-controlling parameter in a primary NLMS process. In effect,

$$\bar{\mu}_k = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{\hat{E}_k(e\mathbf{x}_i)}{\sqrt{\hat{E}_k(e^2)\hat{E}_k(\mathbf{x}_i^2)}}\right|$$

where

$$\hat{E}_k(e^2) = \lambda_{\text{SKYW}}\hat{E}_{k-1}(e^2) + (1-\lambda_{\text{SKYW}})e_k^2$$

$$\hat{E}_k(\mathbf{x}_i^2) = \lambda_{\text{SKYW}}\hat{E}_{k-1}(\mathbf{x}_i^2) + (1-\lambda_{\text{SKYW}})\mathbf{x}_{k,i}^2$$

$$\hat{E}_k(e\mathbf{x}_i) = \lambda_{\text{SKYW}}\hat{E}_{k-1}(e\mathbf{x}_i) + (1-\lambda_{\text{SKYW}})e_k\mathbf{x}_{k,i}$$

This is the method that is used in the simulations to examine the behaviour of the AGC idea, and is referred to as the SKYW algorithm in honor of the original authors. The computational complexity of this approach is comparable to that of the algorithm to be proposed in Chapter 4.

Critique

The most problematic aspect of the original automatic gain control algorithm is its implicit dependence on the statistics of the additive noise, in spite of the authors' claims to robustness to that very thing. This dependence is evident when considering what values should be used for the various algorithm parameters. Further, there are no guidelines given for the choice of algorithm parameters. In fact, the values used in the various simulations of [27] are not even reported. The immediate advantages of the AGC modifications of [30] are that no knowledge of the environment is necessary due to the normalization of the cross-correlation estimate and that only one algorithmic parameter, namely, λ_{SKYW} remains. In consequence, the above SKYW algorithm can be fairly compared to other methods with relative ease.

Remarkably, the simulation chosen to demonstrate the tracking properties of the algorithm in [27] actually demonstrates that its reconvergence properties are worse than that of the NLMS algorithm having the same steady-state misadjustment. In [30], on the other hand, the variable step size algorithm is more favourably compared with the standard NLMS adaptive filter. The further modification presented above results in further performance improvements for this approach.

2.2.4 Algorithm of Karni and Zeng

In this section, the damped convergence factor (DCF) algorithm proposed by Karni and Zeng in [31] is briefly treated.

Summary

This algorithm is another *ad hoc* attempt to reduce the step-size as the norm of the gradient estimate decreases and *vice versa.* In this case, however, an exponential relation between the gradient estimate magnitude and the convergence-controlling parameter is maintained.

Critique

Unfortunately, all of the criticisms appropriate for the original AGC algorithm apply equally to the DCF scheme. We might point out, however, that additional difficulties arise when dealing with a variable step algorithm that is based on a basic LMS rather than an NLMS filter. In particular, the upper limit on the convergence-controlling parameter is based on stability constraints. For the NLMS adaptive filter, algorithm stability is assured for positive values of convergence-controlling parameter less than two although there is no performance advantage in increasing $\bar{\mu}$ beyond unity. For the LMS filter, on the other hand, this upper limit depends on the statistics of the

input signal and no necessary conditions for instability exist. Sufficient conditions are given in [32]. These difficulties have not been addressed sufficiently in [31]. As a result, the application of the DCF algorithm in an unknown environment is too problematic to warrant further consideration.

2.2.5 Algorithm of Mathews and Xie

In this section, the gradient adaptive step-size (GAS) algorithm proposed by Mathews and Xie in [33] is treated.

Summary

The GAS method abstracts the LMS idea of using an instantaneous gradient estimate one step further: an instantaneous estimate of the gradient of the MSE with respect to the convergence-controlling parameter is utilized to maintain an estimate of the optimal parameter value for use in the primary adaptation process.

Critique

While this technique is perhaps the most creative of those to be considered, the GAS scheme does not deliver significant improvement in performance over the LMS filter on which it is based. While the secondary adaptation mechanism of this algorithm can succeed in providing the best possible convergence-controlling parameter in non-stationary steady-state conditions, its trade-off between convergence and steady-state performance is worse than that of the standard LMS filter. In other words, if the variance of the GAS convergence-controlling parameter is within acceptable limits, the time required for it to converge can be considerable.

2.2.6 Algorithm of Kwong and Johnston

In this section, the variable stepsize algorithm proposed by Kwong and Johnston in [34] is discussed. This algorithm is referred to as the KJ algorithm to avoid confusion.

Summary

The KJ algorithm simply takes the LMS convergence-controlling parameter to be the output of a one-pole filter whose inputs are given by e_k^2. If the error signal increases, so does the value of μ_k and *vice versa*.

Critique

The paper under consideration provides a clear demonstration of the power of the standard independence assumptions for the analysis of adaptive filters. Formulas are provided so that the algorithmic parameters can be selected to achieve certain desired specifications. Unfortunately, these formulas require *a priori* knowledge of the powers of the additive noise and input signal. Unfortunately, the KJ method is also quite sensitive to input correlatedness.

2.3 Existing RLS-based Algorithms

2.3.1 Algorithm of Fortescue, Kershenbaum and Ydstie

The most significant attempt to govern the forgetting factor of the RLS algorithm to enhance RLS performance is due to Fortescue, Kershenbaum and Ydstie in [35] and is referred to as the FKY algorithm. While other RLS-based DA methods are hinted at in the literature, the FKY approach is certainly the most cited. By comparison, the proposal to extend AGC ideas to RLS primary adaptation in [27] remains untested.

Summary

The FKY algorithm attempts to keep a measure of the information content of the filter constant at each sample. The mechanism by which the forgetting factor of the present sample is chosen depends on a ratio of the present MSE to an estimate of the additive noise power.

Critique

This algorithm was originally conceived to provide greater robustness in the presence of impersistently exciting inputs. Its application to nonstationary environments, however, can be quite successful. The only drawback to the FKY approach is that an estimate of an environmental quantity is required. In Chapter 5, it shall be demonstrated that such an estimate can be exploited to better advantage using a different criterion for the handling of the variable forgetting factor.

2.4 Summary

With the notable exception of the HCB algorithm, the other DA algorithms of interest require some *a priori* knowledge of the adaptation environment in the form of an estimate of the additive noise power. In the following chapters, three new DA algorithms are presented. In each case, such information is not explicitly required. If this estimate exists, however, then each of these forthcoming algorithms can be easily modified to accommodate this information with improved performance.

Chapter 3

RASE – Toward Optimal Convergence

3.1 Introduction

This chapter concerns itself with the development of an algorithm that attempts to govern its convergence-controlling parameter such that the misadjustment in the subsequent sample is minimized.

In Figure 1.2, the learning curve of an imaginary DA algorithm is shown by the dashed line. The time-constants of each section of this piecewise exponential curve are inherited from the conventional "parent" algorithm whose learning curves are provided by the solid lines in the same figure. Whenever the DA method arrives at a conventional misadjustment barrier, it "shifts gears" — modifying its convergence-controlling parameter to provide another 3-dB improvement in steady-state misadjustment. We note that the resulting DA performance is considerably better than that of its parent algorithm: the trade-off between convergence and MAC has been mitigated substantially.

It is quite clear that to implement such an algorithm would necessarily require some knowledge of the adaptation environment. In order to know when to "shift gears" one must either have some method by which the arrival at a barrier may be

ascertained, or have both knowledge of the MMSE and the quantities at which the conventional barriers occur. The HCB algorithm attempts the former option, while the other existing DA methods utilize environmental information in order to govern the convergence-controlling parameter of the primary adaptation process. Unfortunately, none of these other methods make use of knowledge about where the conventional performance barriers occur. This chapter will demonstrate how such knowledge can be obtained and utilized in a powerful doubly adaptive framework.

3.2 NLMS Preliminaries

In general, the potential for using a conventional algorithm as the basis for a DA scheme is directly proportional to the confidence with which one can predict the performance of that algorithm. In consequence, the performance of the NLMS adaptive filter is examined in detail. To facilitate the further development of NLMS-based DA methods in the subsequent chapter, the present analysis is made to be quite general. Let us consider, then, the nonstationary case in which the target filter weights vary at each iteration according to a random walk such that $\mathbf{w}_{k+1} = \mathbf{w}_k + \mathbf{z}_k$ where $\mathbf{z}$ is a Gaussian vector independent of the additive noise and filter inputs with $E(\mathbf{z}_k \mathbf{z}_j^T) = \delta_{kj}\sigma_z^2 \mathbf{I}_N$. $\mathbf{I}_N$ is the $N \times N$ identity matrix and

$$\delta_{kj} = \begin{cases} 1 & k = j \\ 0 & k \neq j \end{cases}.$$

While the fact that the random walk nonstationarity is an artificial model has been discussed in the literature (e.g., [36] [37]), the tractability that it affords permits the comparison of numerous algorithms in a nonstationary environment. Moreover, an algorithm that performs well in a random walk scenario, however artificial that may be, is likely to outperform a conventional algorithm in an arbitrary nonstationary environment.

In the case of randomly walking target filter coefficients, then, (1.3) suggests an NLMS weight error vector update of the form

$$\mathbf{v}_{k+1} = \mathbf{v}_k + \mathbf{z}_k - \frac{\bar{\mu} e_k \mathbf{x}_k}{\mathbf{x}_k^T \mathbf{x}}. \tag{3.1}$$

If we take the inner product of this result with itself, we obtain

$$\mathbf{v}_{k+1}^T \mathbf{v}_{k+1} = \mathbf{v}_k^T \mathbf{v}_k + \mathbf{z}_k^T \mathbf{z}_k + \frac{\bar{\mu}^2 e_k^2 + 2\mathbf{v}_k^T \mathbf{z}_k - 2\bar{\mu} e_k \mathbf{x}_k^T \mathbf{z}_k - 2\bar{\mu} e_k (e_k - \epsilon_k)}{\mathbf{x}_k^T \mathbf{x}_k}. \tag{3.2}$$

We now take the expected values of both sides of this result in the steady state, i.e., as $k \to \infty$, such that $E(\mathbf{v}_{k+1}^T \mathbf{v}_{k+1}) = E(\mathbf{v}_k^T \mathbf{v}_k)$. Further, we simplify the result obtained, namely,

$$\left(1 - \frac{\bar{\mu}}{2}\right) E\left(\frac{e^2}{\mathbf{x}^T \mathbf{x}}\right) = \frac{N\sigma_z^2}{2\bar{\mu}} + \sigma_\epsilon^2 E\left(\frac{1}{\mathbf{x}^T \mathbf{x}}\right)$$

in two ways. In particular, we need

$$E\left(\frac{1}{\mathbf{x}^T \mathbf{x}}\right) \approx \frac{1}{E\left(\mathbf{x}^T \mathbf{x}\right)} = \frac{1}{N\sigma_x^2}$$

and

$$E\left(\frac{e^2}{\mathbf{x}^T \mathbf{x}}\right) \approx E\left(e^2\right) E\left(\frac{1}{\mathbf{x}^T \mathbf{x}}\right).$$

The first of these approximations is valid for moderate to large values of N. The second, on the other hand, is more problematic. An assumption will be made that is sufficient to justify this second approximation. The quantity $v_k = \mathbf{v}_k^T \mathbf{x}_k$ is taken to be independent of $\mathbf{x}_k$. This is a modification of one of the independence assumptions common in the literature for the analysis of the standard LMS algorithm (see, for example, [38] [6]), namely, that the weight error vector $\mathbf{v}_k$ is independent of the input vector $\mathbf{x}_k$. If we can, in fact, take $\mathbf{v}_k$ to be independent of $\mathbf{x}_k$, then an independence between v and $\mathbf{x}$ develops in the steady state where we have $E(\mathbf{v}) = \mathbf{0}$ as the value of N becomes large relative to the input autocorrelation.

It is interesting that the invocation of the classical independence assumptions often includes some comment to the effect that "it has been shown to provide valuable results in the past" rather than arguing its merits on an *a priori* basis or appealing to work intended to establish their validity [39]. Indeed, we apply the present modified independence assumption on the support of the accuracy with which it allows adaptive filter performance to be predicted, in keeping with the way in which the original assumption is handled in the literature.

While other researchers have examined the dependence of NLMS performance on the input correlatedness (e.g., [40]), only the limit having autocorrelation width smaller than the filter length is considered. Apart from the support that this limit gives to the present independence assumption, its consideration is warranted since the fact that LMS-based adaptive filter algorithms perform poorly when input samples are highly correlated is widely known. As a result, it is unlikely that these algorithms will be applied in conditions far from the limit of input uncorrelatedness.

The application of the above pair of approximations gives

$$E(e^2) = \frac{2 + \frac{q}{\bar{\mu}}}{2 - \bar{\mu}} \sigma_\epsilon^2. \tag{3.3}$$

The quantity q is the normalized nonstationarity to additive noise ratio and is defined as

$$q \triangleq \frac{N^2 \sigma_z^2 \sigma_x^2}{\sigma_\epsilon^2}, \quad \text{NNR} \triangleq 10 \log_{10}(q).$$

The corresponding misadjustment, $\mathcal{M}$, is defined, given the accepted composition of the desired signal, as $E(e^2)/\sigma_\epsilon^2 - 1$. Consequently, the misadjustment at convergence, $\mathcal{M}_\infty$, is given by

$$\text{MAC} \triangleq \mathcal{M}_\infty = \frac{\bar{\mu} + q/\bar{\mu}}{2 - \bar{\mu}}. \tag{3.4}$$

We note that the quantity q represents a limit on the misadjustment in the nonstationary situation. The component of the *a priori* error due to the random walk

is simply $\mathbf{z}_k^T\mathbf{x}_k$. The normalized variance of this entirely unpredictable quantity is simply q. In consequence, we expect that the partial derivative of the steady-state misadjustment with respect to q will always be greater than unity as is true in the case of (3.4).

At this point, we have a fairly accurate expression for the steady-state performance of the NLMS adaptive filter. Next, we consider that algorithm's convergence behaviour in an unknown stationary environment. In [20], Slock suggests the use of a simplified model for the distribution of the input vector sequence to the adaptive filter. By maintaining the second-order statistics of the actual inputs, the steady-state MSE is successfully approximated. A slight modification of his model is as follows. The covariance matrix, $\mathbf{R}$, of the input vectors, $\mathbf{x}_k$, may be written as

$$\mathbf{R} = E(\mathbf{x}_k\mathbf{x}_k^T) = \sum_{i=1}^{N} r_i\mathbf{u}_i\mathbf{u}_i^T.$$

where r_i are the eigenvalues, and $\mathbf{u}_i$ are the corresponding orthonormal eigenvectors of $\mathbf{R}$. The simplest possible distribution that demonstrates this second-order behaviour is now constructed. The vectors $\mathbf{x}_k$ are then taken to be independent and identically distributed from a discrete distribution in which the vectors $\sqrt{r_i N}\mathbf{u}_i$ occur with equal likelihood. Though this distribution may be considered far too simple, the results which follow are worthy of consideration.

Consider the adaptation state embodied in the matrix $\mathbf{S}_k \triangleq E(\mathbf{v}_k\mathbf{v}_k^T)$, which is independent of $\mathbf{x}_k$ from previous assumptions, viz.

$$\mathbf{S}_{k+1} = E\left[\left(I - \frac{\bar{\mu}\mathbf{x}_k\mathbf{x}_k^T}{\mathbf{x}_k^T\mathbf{x}_k}\right)\mathbf{S}_k\left(I - \frac{\bar{\mu}\mathbf{x}_k\mathbf{x}_k^T}{\mathbf{x}_k^T\mathbf{x}_k}\right)\right] + \bar{\mu}^2\sigma_\epsilon^2 E\left[\frac{\mathbf{x}_k\mathbf{x}_k^T}{(\mathbf{x}_k^T\mathbf{x}_k)^2}\right]. \qquad (3.5)$$

Following Slock, the quantities defined by $s_{k,i} = \mathbf{u}_i^T\mathbf{S}_k\mathbf{u}_i$ are examined under the simplified distribution outlined above. We find that

$$s_{k+1,i}r_i = \left[1 - \frac{\bar{\mu}(2-\bar{\mu})}{N}\right]s_{k,i}r_i + \left(\frac{\bar{\mu}}{N}\right)^2\sigma_\epsilon^2.$$

Now, recalling that for i.i.d. $\mathbf{x}_k$ the MSE may be written as

$$E(e_k^2) = \text{tr}\,(\mathbf{R}\mathbf{S}_k) + \sigma_\epsilon^2 = \sigma_\epsilon^2 + \sum_{i=1}^{N} s_{k,i} r_i$$

where tr $(\cdot)$ denotes trace, we have

$$\mathcal{M}_{k+1} = \left[1 - \frac{\bar{\mu}(2-\bar{\mu})}{N}\right] \mathcal{M}_k + \frac{\bar{\mu}^2}{N}. \tag{3.6}$$

From this result, a number of observations can be made. We see, for example, that NLMS convergence is guaranteed for $0 < \bar{\mu} < 2$, in keeping with [12]. These limits are also explicit in the development of (3.3). Further, the steady state misadjustment, which may be obtained by letting $k \to \infty$, clearly matches (3.4) for $q = 0$. On the other hand, it is clear that the convergence behaviour provided by the above simplifications is not entirely satisfactory (in contrast with Slock's model), since there is no apparent mechanism by which NLMS convergence can degrade in the presence of correlated inputs. In order to account for this important behaviour without complicating matters unduly, we modify (3.6) as

$$\mathcal{M}_{k+1} = [1 - \xi\bar{\mu}(2-\bar{\mu})]\mathcal{M}_k + \xi\bar{\mu}^2 \tag{3.7}$$

where ξ, whose nominal value is N^{-1} and is always positive, provides for different convergence behaviour under different input regimes. Note that the MAC performance, which is independent of the input statistics, is also independent of ξ. A noise-free expression equivalent to (3.7) appears in [40].

3.3 Algorithm Development

At this point, we consider the possibility of continuous "gear-shifting" the NLMS convergence-controlling parameter in order to minimize the misadjustment in the

subsequent sample. Differentiating (3.7) with respect to $\bar{\mu}$ and setting the result to zero, we find that the best choice of $\bar{\mu}$ is given by

$$\bar{\mu}_k^* = \frac{\mathcal{M}_k}{1+\mathcal{M}_k}. \tag{3.8}$$

We observe, again, that this result is independent of ξ. Unfortunately, this optimal value depends on knowledge of the adaptation environment. The instantaneous misadjustment, which is unavailable to the adaptive filter, represents the reduced adaptation state of the adaptive filter. Without an ability to predict the NLMS behaviour, no estimate of this quantity could be made. Using the performance model provided by (3.7), however, we can obtain a reasonable estimate of the instantaneous misadjustment as follows. We construct the unknown-parameter state estimation model

$$\begin{aligned} \mathbf{y}_{k+1} &= \mathbf{F}_k(\xi)\mathbf{y}_k \\ E(e_k^2) &= \mathbf{h}\mathbf{y}_k \end{aligned} \tag{3.9}$$

where the vector $\mathbf{y}$ consists of

$$\mathbf{y} \triangleq [\mathcal{M}\sigma_\epsilon^2 \;\; \sigma_\epsilon^2]^T$$

the transition matrix, $\mathbf{F}_k(\xi)$, is given by

$$\begin{bmatrix} 1-\xi\bar{\mu}_k(2-\bar{\mu}_k) & \xi\bar{\mu}_k^2 \\ 0 & 1 \end{bmatrix}$$

and the measurement matrix, $\mathbf{h}$, is simply [1 1]. We remark that the adaptive filter has a readily available measurement of $E(e_k^2)$, namely, e_k^2.

Taking e_k to be approximately Gaussian for moderate to large values of N, we deduce the shifted-χ_1^2 nature of the measurement noise. While this measurement noise may not be symmetrically distributed, the facts that it is relatively white and

that its mean is zero are sufficient to encourage the use of an extended Kalman filter for the joint estimation of $\mathbf{y}$ and ξ. Further, the knowledge of the underlying noise distributions and the absence of plant noise also contribute to the confidence with which we apply this method [41].

Applying the nonlinear extensions to the Kalman filter found in [13], we obtain the following estimation method:

$$\begin{aligned}
\tilde{\mathbf{y}}_{k+1} &= \varphi\left[\tilde{\mathbf{F}}_k(\hat{\xi}_k)\tilde{\mathbf{y}}_k + \mathbf{g}_k(e_k^2 - \tilde{\mathbf{h}}\tilde{\mathbf{y}}_k)\right] \\
\tilde{\mathbf{P}}_{k+1} &= \check{\mathbf{F}}_k(\tilde{\mathbf{y}})\tilde{\mathbf{P}}_k\check{\mathbf{F}}_k^T(\tilde{\mathbf{y}}) + \mathbf{Q} - \mathbf{g}_k w_k \mathbf{g}_k^T \\
\mathbf{g}_k &= \tilde{\mathbf{F}}_k(\tilde{\mathbf{y}})\tilde{\mathbf{P}}_k\tilde{\mathbf{h}}^T/w_k \\
w_k &= \tilde{\mathbf{h}}\mathbf{P}_k\tilde{\mathbf{h}}^T + u(\tilde{\mathbf{y}}_k).
\end{aligned} \tag{3.10}$$

The estimator's state vector, $\tilde{\mathbf{y}}$, is given by $\tilde{\mathbf{y}} \triangleq [\hat{\mathbf{y}}^T \quad \hat{\xi}]^T$ (the circumflex accent denoting estimate). We also define $\tilde{\mathbf{h}} \triangleq [\mathbf{h} \;\; 0]$,

$$\tilde{\mathbf{F}}_k(\hat{\xi}) \triangleq \begin{bmatrix} \mathbf{F}(\hat{\xi}) & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}.$$

and

$$\check{\mathbf{F}}_k(\tilde{\mathbf{y}}) \triangleq \begin{bmatrix} \mathbf{F}(\hat{\xi}) & \frac{\partial}{\partial\xi}\mathbf{F}(\hat{\xi})\hat{\mathbf{y}} \\ \mathbf{0} & 1 \end{bmatrix}.$$

Further, φ is a projection facility to ensure that the state estimates remain meaningful (i.e., keeping the elements of $\tilde{\mathbf{y}}$ positive) [41], $\mathbf{Q}$, with small diagonal entries, provides for nonstationarities by the common strategy of pseudo plant noise and $u(\hat{\mathbf{y}}) = 2\tilde{\mathbf{h}}\hat{\mathbf{y}}\tilde{\mathbf{y}}^T\tilde{\mathbf{h}}^T = 2\mathbf{h}\hat{\mathbf{y}}\hat{\mathbf{y}}^T\mathbf{h}^T$ is an estimate of the measurement noise variance. Finally, a one-pole filter and lower bound are both imposed on the estimated optimal instantaneous convergence-controlling parameter using (3.8) to arrive at a practical $\bar{\mu}$ sequence, i.e.,

$$\hat{\mu}_k^* = \lambda_\mu \bar{\mu}_{k-1} + \frac{(1-\lambda_\mu)\hat{\mathcal{M}}_k}{1+\hat{\mathcal{M}}_k}$$

$$\bar{\mu}_k = \begin{cases} \hat{\mu}_k^* & \hat{\mu}_k^* \geq \check{\mu} \\ \check{\mu} & \text{otherwise} \end{cases} . \tag{3.11}$$

The latter bound is intended to both ensure that the adaptation remains alive and that the reduced adaptation state remains observable. The values $\lambda_\mu = 0.9$ and $\check{\mu} = 0.0198$ are recommended, the latter corresponding to a theoretical steady-state misadjustment of -20dB.

Finally, the EKF estimator needs to be initialized appropriately. To this end, $\hat{\xi}_0 = N^{-1}$, $\hat{\mathbf{y}}_0 = [e_0^2 \;\; \hat{\sigma}_{\epsilon,0}^2]^T$ are used, and $\tilde{\mathbf{P}}_0$ is a diagonal matrix with entries equal to twice the squares of the corresponding elements of $\tilde{\mathbf{y}}$. The algorithm as presented will be referred to as reduced adaptation state estimation (RASE).

3.4 Practical Considerations

The convergence of the EKF has been shown to be dependent on its initialization. Intuitively, a recursive estimator that makes use of a linear approximation in a nonlinear situation will converge if it starts close enough to the desired state. If it is initialized too far away from the global minima it may be susceptible to local minima in the vicinity of its initial state. Indeed, such is the case with this application of an EKF. In particular, the RASE EKF is susceptible to local minima when initial estimates of the additive noise variance are poor. This represents the greatest limitation of the present method. As has been noted, however, a number of DA algorithms exist in the literature for which such an estimate is explicitly or implicitly required. If some reasonable *a priori* knowledge of this quantity exists, the component of $\tilde{\mathbf{P}}_0$ corresponding to the additive noise power should be reduced accordingly. If perfect environmental knowledge were available, for example, one would use $\tilde{\mathbf{P}}_{0,22} = 0$. This would provide adaptation equivalent to that of a two-component-state RASE-like filter formulated under the assumption that σ_ϵ^2 is known.

When the filter inputs are uncorrelated, a reasonable estimate of the additive noise power is available, and the environment is stationary, the performance of RASE using $\mathbf{Q} = \mathbf{0}$ is excellent. In order to provide for input correlatedness and environmental nonstationarities, however, we choose the diagonal elements of $\mathbf{Q}$ to be positive in keeping with common pseudo-plant-noise strategies. In particular, q_{11} governs the ability of RASE to handle target filter weight nonstationarities and q_{33} allows for input correlatedness. In principle, positive choices of q_{22} might be made in order to handle slow variations in additive noise power. Unfortunately, the sensitivity of the RASE algorithm to this quantity does not permit much optimism in this regard. Although a positive $\mathbf{Q}$ permits the EKF to track variation in the appropriate environmental quantities, the steady-state performance is reduced accordingly. As a result, one would like the diagonal values of $\mathbf{Q}$ to be small.

The results of simulations which qualitatively address the sensitivity of RASE performance to its parameter set and provide comparison between the RASE adaptive filter and other related algorithms are presented in Chapter 6. Due to the relative sensitivity of the RASE algorithm to its parameters and initial estimates of its environment, it will not be examined in greater detail. It is worth mentioning, however, that the problems RASE has in these regards are comparable to those of the existing LMS-based DA algorithms while RASE performance can be significantly better. As the subsequent chapters shall bear out, DA methods that both perform well and are largely immune to these problems can be conceived.

Of course, the RASE method can be modified by the replacement of the basic EKF as presented here by some other, more reliable, estimator. The EKF modifications suggested in [41], for example, have been shown to guarantee EKF convergence is certain circumstances. The literature on so-called "adaptive" Kalman filtering

also provides options for enhanced environmental estimation. The *partitioned* and *multiple-model* estimators of, for example, [42] [43] [44] [45] are worthy of investigating for our dimensionally small application. For the present purposes, however, these possibilities for further research are mentioned in passing. The approach of the following chapter and especially its application in Chapter 5 seem to carry more promise than the relatively sensitive RASE method.

Chapter 4

PA-NLMS – Toward Optimal Tracking

4.1 Introduction

In this chapter, another approach to the handling of the time variation of the NLMS convergence-controlling parameter is presented. Instead of optimizing this quantity with respect to the misadjustment in the subsequent sample, the optimization is performed to minimize the steady-state misadjustment in a random walk environment. The advantage of this approach is the relative simplicity of estimating the necessary environmental quantities. The apparent disadvantage is the weight that it gives to an artificial model of nonstationarity.

4.1.1 Parallel Adaptation

In order to find the best possible NLMS convergence-controlling parameter for steady-state random walk performance, (3.4) is differentiated with respect to $\tilde{\mu}$. Setting the right-hand side of this result equal to zero and taking the more meaningful solution in the resulting quadratic equation, the optimal convergence-controlling parameter, viz.,

$$\tilde{\mu}^*(q) = \frac{1}{2}\left[\sqrt{q(q+4)} - q\right] \tag{4.1}$$

is obtained. For small q, $\bar{\mu}^* \approx \sqrt{q}$. On the other hand, as q gets large, $\bar{\mu}^*$ approaches unity. This is the value of $\bar{\mu}$ which provides fastest NLMS convergence in a stationary scenario [20]. Recalling the quadratic nature of the MSE performance surface with respect to the weight error vector for a transversal filter, it comes as no surprise that the fastest convergence occurs for a value of $\bar{\mu}$ midway between convergence limits. Values of $\bar{\mu} > 1$ in NLMS adaptation result in a poorer convergence *and* poorer MAC.

If a reliable estimate of the quantity q were available, optimal steady-state random walk NLMS adaptation would be readily available. Unfortunately, q is practically unobservable for a single NLMS procedure. The use of two independent NLMS processes with sufficiently different values of $\bar{\mu}$ running in parallel is therefore proposed. Using two instances of (3.3) of different $\bar{\mu}$, the quantity q becomes observable. The explicit use of this principle will be referred to as *parallel adaptation* (PA). The application of this idea, is not, of course, limited to NLMS adaptation. In Chapter 5, for example, the following development will be applied to RLS procedures. One of many possible PA algorithms will now be presented.

Consider a variable-$\bar{\mu}$ NLMS primary process and two fixed-$\bar{\mu}$ (one high and one low) indicator NLMS processes all running in parallel which will be denoted by the subscripts v, h, and l, respectively. Let us take $\bar{\mu}_h = 1$ and $\bar{\mu}_l = \alpha$, $0 < \alpha \ll 1$. The two indicator processes will provide for the estimation of the nonstationarity to additive noise ratio on which $\bar{\mu}_v$ will depend.

The natural approach is to solve two instances of (3.3) for q at each iteration, that is,

$$\tilde{q}_k = \frac{\tilde{q}_{r,k}}{\tilde{q}_{\epsilon,k}} = \frac{2\alpha[(2-\alpha)e_{l,k}^2 - e_{h,k}^2]}{e_{h,k}^2 - \alpha(2-\alpha)e_{l,k}^2}$$

and exponentially smooth the resulting sequence:

$$\hat{q}_k = \lambda_q \hat{q}_{k-1} + (1-\lambda_q)\tilde{q}_k.$$

As shall be seen, however, there is a performance advantage and an analysis simplification in the independent smoothing of the numerator and denominator of $\tilde{q}$, viz.,

$$\hat{q}_k = \frac{\hat{q}_{r,k}}{\hat{q}_{\epsilon,k}} = \frac{\lambda_z \hat{q}_{r,k-1} + (1-\lambda_z)2\alpha\left[(2-\alpha)e_{l,k}^2 - e_{h,k}^2\right]}{\lambda_\epsilon \hat{q}_{\epsilon,k-1} + (1-\lambda_\epsilon)\left[e_{h,k}^2 - \alpha(2-\alpha)e_{l,k}^2\right]} \tag{4.2}$$

where the values of λ_ϵ and λ_z depend on the application, but will, in general, be close to unity.

The instantaneous estimate of the optimal convergence-controlling parameter, $\tilde{\mu}_k^*$, may then be obtained using

$$\tilde{\mu}_k^* = \begin{cases} \bar{\mu}^*(\hat{q}_k), & \hat{q}_k > 0 \\ 0, & \text{otherwise} \end{cases} \tag{4.3}$$

Since a number of nonlinearities are involved it seems prudent to also smooth this result

$$\bar{\mu}_{v,k+1} = \hat{\mu}_k^* = \lambda_\mu \hat{\mu}_{k-1}^* + (1-\lambda_\mu)\tilde{\mu}_k^*. \tag{4.4}$$

The parallel adaptation algorithm using normalized LMS branches (PA-NLMS), capable of self-tuning the value of $\bar{\mu}$ toward the best possible NLMS steady-state random walk MSE has now been outlined. Before considering how this algorithm will perform in an unknown stationary adaptation scenario, let us consider the expected steady-state random walk performance of the PA-NLMS method as given.

4.2 Performance Evaluation

4.2.1 Random Walk Performance

In order to examine the performance of any LMS-based DA scheme, a number of simplifying assumptions are required. For the PA-NLMS algorithm, $\bar{\mu}_k$ is taken to be independent of e_k and the filter input. For $\lambda_\mu \approx 1$, this is a fair supposition since this will mean a slow variation of $\bar{\mu}$. Under this assumption, the MAC for the PA-NLMS

algorithm becomes

$$\mathrm{MAC}_{\mathrm{PA-NLMS}} = \frac{E(\bar{\mu}_v^2) + q}{2E(\bar{\mu}_v) - E(\bar{\mu}_v^2)} \tag{4.5}$$

In order to find $E(\bar{\mu}_v)$ and $E(\bar{\mu}_v^2)$, it is noted that the sequences $\hat{q}_{r|\epsilon,k}$ become jointly Gaussian as $\lambda_{\epsilon|r} \to 1$ by virtue of the central limit theorem. The steady-state variances of these sequences may then be obtained under the assumption that the sequences $v_{h|l,k}$ are mutually independent and sufficiently near Gaussian that the indicator error sequences, $e_{h|l,k}$, are approximately Gaussian. It is clear that any dependence between v_h and v_l will disappear in the steady state. Further, for $\bar{\mu} \leq 1$, at least $(2+q)^{-1}$ of this MSE will be independent of v. For small values of q, which can be expected in steady-state conditions, this is a significant proportion. One can expect, therefore, that the v sequences will be sufficiently well-behaved for practical values of N to render this assumption acceptable. Under these circumstances, the useful relations

$$\begin{aligned}
\mathrm{cov}\,[e_h^2, e_l^2] &= 2\sigma_\epsilon^4, \\
\mathrm{var}\,[e_h^2] &= 2(2+q)^2\sigma_\epsilon^4, \\
\mathrm{var}\,[\alpha(2-\alpha)e_l^2] &= 2(2\alpha+q)^2\sigma_\epsilon^4,
\end{aligned}$$

can be obtained.

It is easily shown that the steady-state expected values of $\hat{q}_z$ and $\hat{q}_\epsilon$ are $q(1-\alpha)\sigma_\epsilon^2$ and $(1-\alpha)\sigma_\epsilon^2$, respectively. Their variances are given by

$$\begin{aligned}
\mathrm{cov}\,[\hat{q}_z, \hat{q}_\epsilon] &= \frac{4\sigma_\epsilon^4(1-\lambda_z)(1-\lambda_\epsilon)}{\lambda_z\lambda_\epsilon - 1}\left[(1+\alpha)q^2 + 8\alpha q + \alpha(1+\alpha)(2+\alpha)\right] \\
\mathrm{var}\,[\hat{q}][\hat{q}_\epsilon] &= \frac{8\sigma_\epsilon^4(1-\lambda_z)}{1+\lambda_z}\left[(1+\alpha^2)q^2 + 4\alpha(1+\alpha)q + 2\alpha^2(2+\alpha)\right] \\
\mathrm{var}\,[\hat{q}_\epsilon] &= \frac{4\sigma_\epsilon^4(1-\lambda_\epsilon)}{1+\lambda_\epsilon}\left[q^2 + 2(1+\alpha)q + 2(1-\alpha) + 3\alpha^2\right].
\end{aligned} \tag{4.6}$$

The quantity $p_{\hat{q}}$, the probability distribution of $\hat{q}$ in the steady-state, can now be determined by integrating that of its numerator and denominator over an appropriate

slice of the $\hat{q}_z - \hat{q}_\epsilon$ plane, viz.,

$$p_{\hat{q}}(\hat{q}) = \int_{-\infty}^{\infty} |x| p_{\hat{q}_z,\hat{q}_\epsilon}(\hat{q}x, x)dx. \tag{4.7}$$

This can be solved explicitly for jointly Gaussian $\hat{q}_z$ and $\hat{q}_\epsilon$ using the identity

$$K\int_{-\infty}^{\infty} |x| \exp(-ax^2 + bx - c)dx = \frac{K}{a}\left[\exp(-c) + \frac{b}{2}\sqrt{\frac{\pi}{a}}\exp\left(\frac{b^2}{4a} - c\right)\operatorname{erf}\left(\frac{b}{2\sqrt{a}}\right)\right]$$

where, for our purposes, the various parameters are given by

$$\begin{aligned}
a &= \frac{\hat{q}^2 \text{ var } (\hat{q}_\epsilon) - 2\hat{q} \text{ cov } (\hat{q}_z, \hat{q}_\epsilon) + \text{var } (\hat{q}_z)}{2[\text{var } (\hat{q}_z)\text{var } (\hat{q}_\epsilon) - \text{cov}^2(\hat{q}_z, \hat{q}_\epsilon)]} \\
b &= 2(1-\alpha)\sigma_\epsilon^2 \frac{q\hat{q} \text{ var } (\hat{q}_\epsilon) - (q+\hat{q})\text{cov } (\hat{q}_z, \hat{q}_\epsilon) + \text{var } (\hat{q}_z)}{\text{var } (\hat{q}_z)\text{var } (\hat{q}_\epsilon) - \text{cov}^2(\hat{q}_z, \hat{q}_\epsilon)} \\
c &= 2(1-\alpha)^2\sigma_\epsilon^4 \frac{q^2 \text{ var } (\hat{q}_\epsilon) - 2q \text{ cov } (\hat{q}_z, \hat{q}_\epsilon) + \text{var } (\hat{q}_z)}{\text{var } (\hat{q}_z)\text{var } (\hat{q}_\epsilon) - \text{cov}^2(\hat{q}_z, \hat{q}_\epsilon)} \\
K &= \frac{1}{2\pi\sqrt{\text{var } (\hat{q}_z)\text{var } (\hat{q}_\epsilon) - \text{cov }^2(\hat{q}_z, \hat{q}_\epsilon)}}.
\end{aligned} \tag{4.8}$$

Applying (4.6), one can observe that $p_{\hat{q}}$ is independent of the additive noise power statistics. In consequence, one can expect that the PA-NLMS scheme will be as insensitive to nonstationary additive noise as the standard NLMS algorithm.

Due to the complexity of $p_{\hat{q}}$, we evaluate the expression

$$\int_0^{\infty} [\tilde{\mu}^*(q)]^n p_{\hat{q}}(q)dq \tag{4.9}$$

for $n = 1, 2$ numerically in order to find $E(\tilde{\mu}_v)$ and $E(\tilde{\mu}^2)$. Finally, the first two moments of $\tilde{\mu}_v$ can be determined from the results of (4.9), under the assumption that $E(\hat{\mu}\tilde{\mu}) \approx E(\hat{\mu})E(\tilde{\mu})$, which is valid for $\lambda_\mu \approx 1$, as

$$\begin{aligned}
E(\tilde{\mu}_v) &= E(\tilde{\mu}) \\
E(\tilde{\mu}_v^2) &= \frac{(1-\lambda_\mu)E(\tilde{\mu}^2) + 2\lambda_\mu E^2(\tilde{\mu})}{1+\lambda_\mu}.
\end{aligned} \tag{4.10}$$

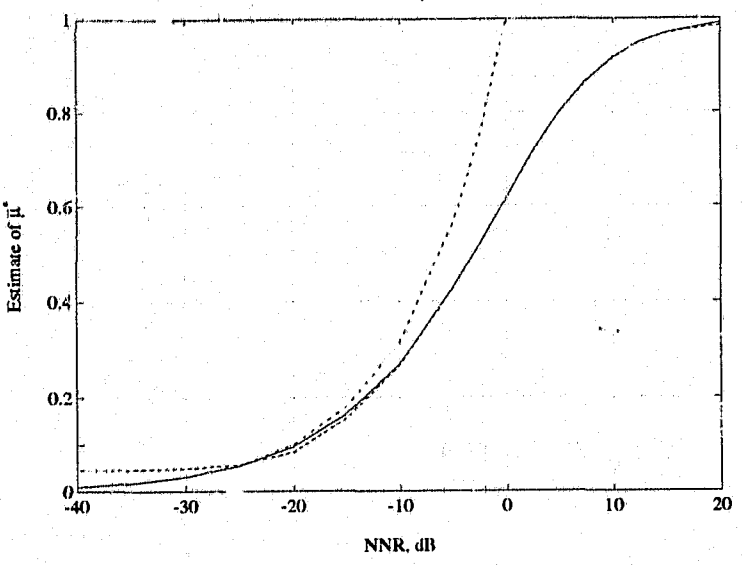

Figure 4.1: Estimates of optimal random walk $\bar{\mu}$
——————— present analysis (4.1)
– – – – – – – – PA-NLMS $\bar{\mu}_v$
– · – · – · – · – previous analysis [15]

This gives us the steady-state PA-NLMS MSE by way of (4.5).

Figure 4.1 shows the expected steady-state value of $\bar{\mu}_v$ versus NNR for typical choices of PA-NLMS parameters along with the optimal values given by (4.1). Here, $\lambda_\epsilon = 0.999$, $\lambda_z = 0.99$, $\lambda_\mu = 0.975$ and $\alpha = 0.0625$. The effects of the discontinuity of (4.3) are evident, since as $q \to 0$, the estimate diverges from its desired behaviour, that is, $E(\hat{\mu}_v) > \bar{\mu}^*[E(\hat{q})] \approx \sqrt{q}$. For relatively quickly varying scenarios, on the other hand, $E(\bar{\mu}_v) \approx \bar{\mu}^*(q)$. For purposes of comparison, the asymptotic result of Widrow et al. is also shown in Figure 4.1 under the relation $N\sigma_x^2\mu \sim \bar{\mu}$ [15]. The recent improvement due to Kwong [46] is indistinguishable from the asymptotic result of Widrow for reasonable values of N in this practical range of $\bar{\mu}$.

Figure 4.2 helps us examine the effects of the discontinuity of (4.3) in greater detail by displaying some contours of the stationary ($q = 0$) PA-NLMS MAC surface

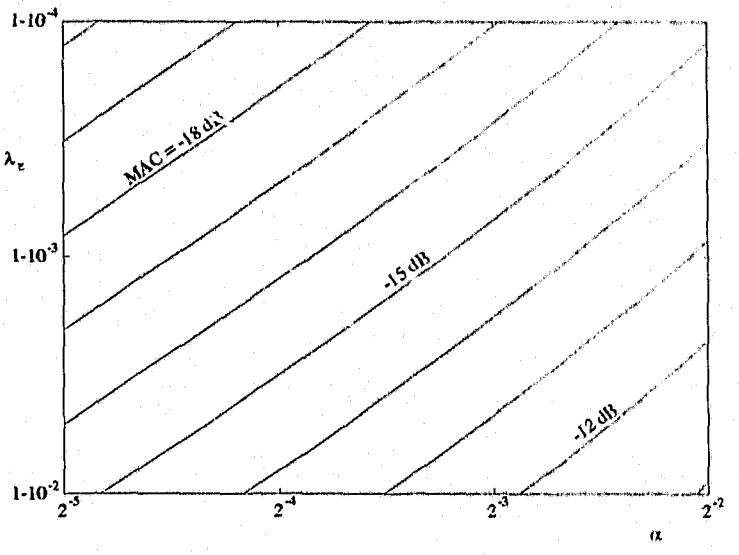

Figure 4.2: Theoretical stationary PA-NLMS performance

for a specific slice of the PA-NLMS parameter space, namely, $\lambda_\mu = 0.975$ and $\lambda_z = 1 - 10(1 - \lambda_\epsilon)$. Here, the explicit relation between the PA-NLMS parameters and the stationary MAC is shown. In consequence, some implicit design guidelines are provided assuming a stationary MAC specification. Further, the sensitivity of the stationary MAC to PA-NLMS parameters is quite low. That is, for a wide range of parameters, a rather slow and smooth variation in this performance measure is exhibited. Consequently, one can expect that any necessary tuning of the PA-NLMS parameters will be minimal.

If the application to be considered were actually a random walk, one would simply take $\alpha \approx 0$ and $\lambda_{\epsilon|r} \approx 1$. In practice, however, such an application cannot be expected, and we wish to have acceptable performance in a number of practical adaptation regimes. Moreover, it is not imagined that the PA-NLMS algorithm will successfully self-tune for every nonstationary application, but expect that it will outperform adaptive filters whose design is predicated on assumptions of stationar-

ity. In the following sections, the behaviour of the PA-NLMS scheme in an unknown stationary environment is examined as it *interprets* this environment as a random walk.

4.2.2 Performance Analysis, Stationary Adaptation

Since the PA-NLMS algorithm and the preceding analysis are based entirely on steady-state assumptions, the performance of the algorithm in typical unknown stationary applications needs to be addressed. Unfortunately, the convergence behaviour of this algorithm is particularly difficult to analyze, and so a rather less formal examination of this problem is provided.

In order for the PA-NLMS algorithm to perform satisfactorily in unknown stationary applications, it is necessary that a far-from-convergence state be interpreted as a quickly-varying random walk. Then, as convergence takes place, it is required that the scheme interprets its environment as a slowly-varying one. It is clear, therefore, that using $\lambda_z = \lambda_\epsilon \approx 1$ will not do for unknown stationary adaptation. At least one of these parameters must be reduced in order to allow a change in environment to register in a reasonably short length of time. As shall be seen, it is to our advantage to set $\lambda_z < \lambda_\epsilon \approx 1$. Consider, then, PA-NLMS convergence in the artificial asymptotic case in which $\lambda_\epsilon \to 1$ and $\lambda_z = 0$ assuming that any smoothing due to λ_z can be subsumed by that of λ_μ. Finally, we will examine the behaviour of $E(\hat{q})$ during noise-free PA-NLMS convergence under the questionable but qualitatively valuable assumption that $E(\hat{q}_z/\hat{q}_\epsilon) \approx E(\hat{q}_z)/E(\hat{q}_\epsilon)$. The elimination of additive noise from this analysis is justified since the effect of additive noise is reduced the further an adaptive filter is from convergence.

Using independence and input-distribution assumptions and a representative filter initialization, Slock has provided an expression for the NLMS learning curve in the

noise-free case that is particularly useful for our purposes, namely,

$$E(e_k^2) = \sum_{i=1}^{N} \left[1 - \frac{\bar{\mu}(2-\bar{\mu})r_i}{\mathrm{tr}\,\mathbf{R}}\right]^k r_i$$

where r_i, as defined previously, are the eigenvalues of the input correlation matrix $\mathbf{R}$ [20]. Using this result, one can obtain directly,

$$E(\hat{q}_{r,k}) = \alpha \sum_{i=1}^{N} r_i \left[\left(1 - \frac{r_i}{\mathrm{tr}\,\mathbf{R}}\right)^k - (2-\alpha)\left(1 - \frac{\alpha(2-\alpha)r_i}{\mathrm{tr}\,\mathbf{R}}\right)^k\right]$$

for $\lambda_z = 0$, and

$$E(\hat{q}_{\epsilon,k}) = \mathrm{tr}\,\mathbf{R}(1-\lambda_\epsilon) \sum_{i=1}^{N} \left\{\left[1 - \frac{\alpha(2-\alpha)r_i}{\mathrm{tr}\,\mathbf{R}}\right]^{k+1} - \left(1 - \frac{r_i}{\mathrm{tr}\,\mathbf{R}}\right)^{k+1}\right\}.$$

as $\lambda_\epsilon \to 1$. Further, when the input samples are relatively uncorrelated, and N is large, $r_i \ll \mathrm{tr}\,\mathbf{R}$ for all i. In consequence, one can approximate each exponential term by the first order polynomial resulting from their respective Taylor series. Applying this approximation, we arrive at

$$E(\hat{q}_k) \approx \frac{E(\hat{q}_{r,k})}{E(\hat{q}_{\epsilon,k})} \approx \frac{\alpha\,[1 - k\eta\alpha(2-\alpha)]}{(1-\lambda_\epsilon)(1-\alpha)(k+1)},$$

where $\eta = \mathrm{tr}\,(\mathbf{R}^2)/\mathrm{tr}^2\,\mathbf{R}$. From this result, a number of observations can be made. First, the initial value of $\hat{q}$ is $\alpha/[(1-\lambda_\epsilon)(1-\alpha)]$. More generally, for $\lambda_z > 0$ and $\hat{q}_{\epsilon|r,-1} = 0$, we have, without approximation,

$$\hat{q}_0 = \frac{(1-\lambda_z)\alpha}{(1-\lambda_\epsilon)(1-\alpha)}.$$

Since one would like this quantity to be large, we take $\lambda_z < \lambda_\epsilon$, and are inclined to increase α. Since the observability of $\hat{q}$ depends on a small α, however, this temptation is resisted. Second, we claim that the time constant of the $\hat{q}_k$ sequence for this asymptotic case will be approximately $\tau_q = [\eta\alpha(2-\alpha)]^{-1}$. For uncorrelated input samples, we have $\eta = 1/N$, and $\tau_q = \tau_l$, the time constant of tap-weight convergence

for the low-$\bar{\mu}$ indicator filter. For the more general case, we can expect this convergence to be postponed by a number of samples in the vicinity of $(1-\lambda_z)^{-1}$, and that convergence of the $\hat{\mu}_k^*$ sequence will be further delayed for about $(1-\lambda_\mu)^{-1}$ samples. Note, however, that the convergence of the $\hat{\mu}^*$ sequence represents the transition between the artificial initial interpretation of the convergence regime as a quickly-varying random walk to the more meaningful interpretation of a steady-state regime as a slowly-varying random walk. The more important tap-weight convergence, on the other hand, is difficult to quantify. Since, however, the initial $\hat{q}$ can be quite large, we have an initial $\hat{\mu}^*$ very near to unity. In consequence, initial tap weight convergence will be comparable to that of the fastest possible NLMS process. Moreover, we can be assured that convergence in the mean square will take place since the limitation of $0 < \bar{\mu}_v < 1$ establishes the convergence condition $2E(\bar{\mu}_v) > E(\bar{\mu}_v^2)$.

While the performance of the PA-NLMS algorithm with respect to the standard stationary measures has been shown to be quite satisfactory, its learning curve has the drawback of many DA schemes, that is, it is significantly different from the well-behaved component-wise exponential learning curve of LMS and NLMS adaptive filters. In fact, the learning curve of the PA-NLMS algorithm as given approximately follows that of its high-$\bar{\mu}$ indicator process until such time as its low-$\bar{\mu}$ indicator process nears convergence. At this point, the environmental estimation mechanism recognizes the stationary condition, and convergence in the mean square proceeds in a similar manner to the low-$\hat{\mu}$ indicator process. Though this may provide acceptable adaptation, a simple heuristic modification can result in significantly better convergence behaviour in many circumstances.

Since the final convergence of the PA-NLMS algorithm depends on the convergence of its slowest indicator filter, we propose to artificially speed up this process by

"leaking" some of the primary tap-weight vector to the low-$\bar{\mu}$ indicator process. This may be accomplished by modifying the appropriate indicator update equation from the standard

$$\hat{\mathbf{w}}_{l,k+1} = \hat{\mathbf{w}}_{l,k} + \frac{\alpha e_{l,k}^2 \mathbf{x}_k}{\mathbf{x}^T\mathbf{x}}$$

to

$$\hat{\mathbf{w}}_{l,k+1} = (1-\zeta)\left(\hat{\mathbf{w}}_{l,k} + \frac{\alpha e_{l,k}^2 \mathbf{x}_k}{\mathbf{x}^T\mathbf{x}}\right) + \zeta\hat{\mathbf{w}}_{v,k+1}$$

where $\zeta \ll 1$. Using this *ad hoc* modification, we have found that the convergence behaviour of the PA-NLMS algorithm can be improved without significantly changing its steady-state performance. A suitable range for ζ is in the order of $0.25/N$.

The classic adaptive filter trade-off between convergence and MAC has again been mitigated. For fixed MAC, the PA-NLMS scheme provides convergence improvement over the parent NLMS algorithm. The trade-off still persists, however, and is manifest in the effects of the various PA-NLMS parameters on that algorithm's performance. Small values of α, for example, provide low MAC but slow convergence. The converse takes place for large values of α. For improved convergence, it is also necessary that not both of λ_z and λ_ϵ be too close to unity, so that the estimate of q can vary in keeping with the operating point of the converging filter.

The PA-NLMS algorithm is only initially sensitive to estimates of environmental quantities. For example, the effects of poor knowledge of the additive noise power exponentially decay with time constant $(1-\lambda_\epsilon)^{-1}$. In the case where confidence exists in an initial estimate of a stationary noise power, however, there is no need for parallel adaptation. Indeed, an estimate of q can come directly from (3.3) based on the current value of $\bar{\mu}_k$, i.e.,

$$\hat{q}_k = \lambda_q\hat{q}_{k-1} + (1-\lambda_q)\bar{\mu}_k\left[\frac{e_k^2}{\hat{\sigma}_\epsilon^2}(2-\bar{\mu}_k) - 2\right].$$

This approach will not be pursued here. Rather, the success of the PA-NLMS algorithm as a DA scheme in an unknown environment will be examined experimentally in Chapter 6.

To summarize the PA-NLMS algorithm, Figure 4.3 provides a block outline of its implementation, and a step-by-step summary follows:

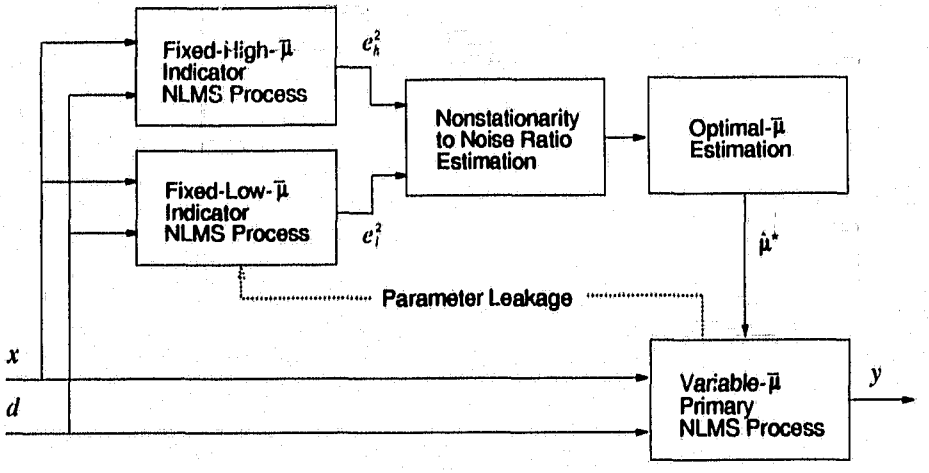

Figure 4.3: Block diagram of PA-NLMS adaptation

1. NLMS update (in parallel)

$$\hat{\mathbf{w}}_{v,k+1} = \hat{\mathbf{w}}_{v,k} + \bar{\mu}_{v,k}(\mathbf{x}_k^T\mathbf{x})^{-1}e_{v,k}\mathbf{x}_k$$

$$\hat{\mathbf{w}}_{h,k+1} = \hat{\mathbf{w}}_{h,k} + \bar{\mu}_{h,k}(\mathbf{x}_k^T\mathbf{x})^{-1}e_{h,k}\mathbf{x}_k$$

$$\hat{\mathbf{w}}_{l,k+1} = (1-\zeta)\left[\hat{\mathbf{w}}_{l,k} + \bar{\mu}_{l,k}(\mathbf{x}_k^T\mathbf{x})^{-1}e_{l,k}\mathbf{x}_k\right] + \zeta\hat{\mathbf{w}}_{v,k+1}$$

$$y_{\text{PA-NLMS}} = y_v.$$

2. Obtain $\hat{q}_k$

$$\hat{q}_{r,k} = \lambda_z\hat{q}_{r,k-1} + (1-\lambda_z)\alpha\left[(2-\alpha)e_{l,k}^2 - e_{h,k}^2\right]$$

$$\hat{q}_{e,k} = \lambda_e\hat{q}_{e,k-1} + (1-\lambda_e)\left[e_{h,k}^2 - \alpha(2-\alpha)e_{l,k}^2\right]$$

$$\hat{q}_k = \hat{q}_{r,k}/\hat{q}_{e,k}$$

3. Calculate $\hat{\mu}_k^* = \bar{\mu}_{v,k+1}$

$$\tilde{\mu}_k^* = \begin{cases} \frac{1}{2}\left[\sqrt{\hat{q}_k(\hat{q}_k+4)} - \hat{q}_k\right], & \hat{q}_k \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

$$\hat{\mu}_k^* = \lambda_\mu \hat{\mu}_{k-1}^* + (1-\lambda_\mu)\tilde{\mu}_k.$$

In Chapter 6, the performance of the PA-NLMS algorithm is demonstrated by way of simulations. In order to demonstrate the insensitivity of this algorithm to the initial estimates of the environmental quantities, we will use $\hat{q}_{z,-1} = \hat{q}_{e,-1} = 0$ in these experiments.

Chapter 5

PA-RLS – Toward Optimal Adaptation

5.1 Introduction

In the vast literature on "adaptive" Kalman filtering there are numerous attempts to use a number of filters in parallel in order to estimate certain parameters on which the performance of the filter bank depends. For example, the *partitioned* and *multiple-model* methods found in [42] [44] [45] are attempts to achieve optimal performance in a manner analogous to the parallel adaptation presented in the previous chapter. While the methods found in these papers and their references are developed more rigorously, the resulting algorithms are often quite involved. In general, this is due to the fact that the general Kalman filtering problem depends on a large number of unknown environmental quantities. For the present adaptive filtering problem, we found in the previous chapters that performance improvement may be forthcoming by considering only a few select aspects of the environment. Intuitively, this approach seems justified. If conventional adaptive filters operate without reference to their environment, then one would expect that modifications to these algorithms that include reasonable estimates of significant environmental quantities could very well result in performance improvements.

In this chapter, a DA extension to the RLS algorithm will be developed. Rather than simplifying the existing "adaptive" Kalman filters, we will apply the approach of the previous chapter directly to the RLS algorithm, which is, in itself, a simplification of the general Kalman filtering problem. To proceed, the existing expressions for steady-state RLS misadjustment in a random walk scenario will be examined and corrected. Subsequently, the PA-RLS algorithm will follow easily from the resulting expressions.

5.2 RLS Preliminaries

Based on the RLS formulation as given in Chapter 1, the derivation of the steady-state RLS MSE under the random walk environment model proceeds by an examination of the weight error update using the standard RLS iteration, viz.,

$$\mathbf{v}_{k+1} = \mathbf{v}_k + \mathbf{z}_k - \frac{\mathbf{P}_k \mathbf{x}_k e_k}{\lambda + \mathbf{x}_k^T \mathbf{P}_k \mathbf{x}_k}. \tag{5.1}$$

Taking the expected value of the outer product of this expression with itself, we obtain

$$\begin{aligned} E\left(\mathbf{v}_{k+1}\mathbf{v}_{k+1}^T\right) = E\left(\mathbf{v}_k\mathbf{v}_k^T\right) + \sigma_z^2\mathbf{I} + E\left[\frac{\mathbf{P}_k\mathbf{x}_k\mathbf{x}_k^T\mathbf{P}_k e_k^2}{\left(\lambda + \mathbf{x}_k^T\mathbf{P}_k\mathbf{x}_k\right)^2}\right] \\ -E\left(\frac{\mathbf{P}_k\mathbf{x}_k\mathbf{x}_k^T\mathbf{v}_k\mathbf{v}_k^T + \mathbf{v}_k\mathbf{v}_k^T\mathbf{x}_k\mathbf{x}_k^T\mathbf{P}_k}{\lambda + \mathbf{x}_k^T\mathbf{P}_k\mathbf{x}_k}\right). \end{aligned} \tag{5.2}$$

Taking the limit as $k \rightarrow \infty$ such that $E\left(\mathbf{v}_{k+1}\mathbf{v}_{k+1}^T\right) = E\left(\mathbf{v}_k\mathbf{v}_k^T\right)$ and dropping the subscript, we have

$$\sigma_z^2\mathbf{I} + E\left[\frac{\mathbf{P}\mathbf{x}\mathbf{x}^T\mathbf{P}e^2}{\left(\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x}\right)^2}\right] = E\left(\frac{\mathbf{P}\mathbf{x}\mathbf{x}^T\mathbf{v}\mathbf{v}^T + \mathbf{v}\mathbf{v}^T\mathbf{x}\mathbf{x}^T\mathbf{P}}{\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x}}\right). \tag{5.3}$$

A number of assumptions are now required to simplify this expression. The assumption set that we shall adopt will provide accuracy in a different limiting case than that used in [16]. In particular, the asymptotic assumption will be modified.

This assumption, commonly understood to have $\lambda \to 1$, is more accurately described as having $\beta \triangleq (1-\lambda)N \to 0$. Our approach will be to take N to be sufficiently large so that this limitation on β may be relaxed. First, however, we take $\mathbf{P}$ to be independent of $\mathbf{x}$, as in [16]. This is certainly valid for independent and identically distributed input samples. For the case of a transversal filter with correlated input samples, however, this remains a good assumption in most cases since the RLS window is, in general, wider than the autocorrelation of the input sequence. Further, the independence of v_k and $\mathbf{x}_k$ is once again assumed as in Chapter 3. The independence of e and $\mathbf{x}$ follows. Finally, we take $E(\mathbf{P}) = c\mathbf{R}^{-1}$. Following [16], we have,

$$\mathbf{PR} \approx c(\mathbf{I}_N + \mathbf{T})$$

where $\mathbf{T}$ is a zero-mean perturbation matrix which is assumed to be independent of $\mathbf{x}$. In [16] and elsewhere (e.g., [47]), c is taken to be its nominal value of $c_0 = 1-\lambda$ since $\mathbf{P}_\infty$ is the inverse of $\hat{\mathbf{R}}_\infty$ and

$$E\left(\hat{\mathbf{R}}_\infty\right) = \lim_{n\to\infty} \sum_{j=0}^{n} \lambda^j E\left(\mathbf{x}_{n-j}\mathbf{x}_{n-j}^T\right) = \mathbf{R}/(1-\lambda).$$

Unfortunately, this value is known to be a poor approximation for $\beta \not\approx 0$ [48]. We shall examine the implications of the value of c presently.

We now return to (5.3) and take the trace after pre-multiplying by $\mathbf{R}$. Under the above assumption set, this gives

$$N\sigma_x^2\sigma_z^2 + \psi\beta c E\left[\left(\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x}\right)^{-1}\right] E\left(e^2\right) = 2cE\left[\left(\mathbf{x}^T\mathbf{v}\right)^2\right] E\left[\left(\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x}\right)^{-1}\right] \quad (5.4)$$

where we introduce

$$\psi = \frac{\operatorname{tr} E\left[\mathbf{RP}\mathbf{x}\mathbf{x}^T\mathbf{P}(\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x})^{-2}\right]}{\operatorname{tr} E\left[\mathbf{RP}\mathbf{x}\mathbf{x}^T\mathbf{P}(\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x})^{-1}\right] E\left[(\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x})^{-1}\right]}$$

so that we can make use of the identity

$$c\beta = \operatorname{tr} E\left(\frac{\mathbf{RP}\mathbf{x}\mathbf{x}^T\mathbf{P}}{\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x}}\right) \quad (5.5)$$

which is simply the expected steady-state (with respect to the input) RLS **P**-update equation. To achieve (5.4), we also require

$$E\left[\nu \mathbf{v}^T \mathbf{T}\mathbf{x}\left(\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x}\right)^{-1}\right] = \mathbf{0}$$

which is not unreasonable considering the above assumptions. We now divide (5.4) by the minimum MSE (σ_ϵ^2). Solving for the steady-state misadjustment we obtain

$$\mathcal{M}_\infty = \frac{2q/\nu + \psi\beta}{2 - \psi\beta} \tag{5.6}$$

where

$$\nu \triangleq cNE\left[(\lambda + \mathbf{x}_k^T\mathbf{P}_k\mathbf{x}_k)^{-1}\right]$$

and the normalized nonstationarity to noise ratio, q, is given by $N^2\sigma_w^2\sigma_z^2/\sigma_\epsilon^2$, as before.

At this point, we simply require reasonable expressions for ψ and ν. To this end, we compare (5.6) with the misadjustment formula found in [16], viz.,

$$\mathcal{M}_\infty = \frac{q/\beta + \beta}{1 + \lambda}. \tag{5.7}$$

Note that the effects of the fourth-order statistics explored in [16] become insignificant for moderate to large values of N, and they are omitted. For the stationary case ($q = 0$), (5.7) is reasonably accurate. In consequence, we make use of this result, which occurred previously in [49] [50] and [51], and solve for ψ, giving

$$\psi = \frac{2}{1 + \lambda + \beta}.$$

We now examine the possibilities for ν. To this end we consider the approximation,

$$\nu = E\left(\frac{cN}{\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x}}\right) \approx \frac{cN}{\lambda + E\left(\mathbf{x}^T\mathbf{P}_k\mathbf{x}_k\right)} = \frac{cN}{\lambda + cN}, \tag{5.8}$$

which is valid for moderate to large values of N. Using the nominal value of $c = c_0 = 1 - \lambda$, we have $\nu = \nu_0 = \beta/(\lambda + \beta)$. While this case represents a common assumption,

the approximation c_0 is known to be significantly less than c for large values of β. On the other hand, if we revisit (5.5) under assumptions similar to those above, we have,

$$\beta = \frac{1}{c}\,\mathrm{tr}\left[E\left(\frac{\mathbf{R}\mathbf{P}\mathbf{x}\mathbf{x}^T\mathbf{P}}{\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x}}\right)\right] = \rho\nu \tag{5.9}$$

where ρ, similar to ψ, is defined as

$$\rho = \frac{\mathrm{tr}\,E\left[\mathbf{R}\mathbf{P}\mathbf{x}\mathbf{x}^T\mathbf{P}(\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x})^{-1}\right]}{\mathrm{tr}\,E\left(\mathbf{R}\mathbf{P}\mathbf{x}\mathbf{x}^T\mathbf{P}\right)E\left[(\lambda + \mathbf{x}^T\mathbf{P}\mathbf{x})^{-1}\right]}.$$

We now have $\nu = \beta/\rho$ and $cN = \lambda\beta/(\rho - \beta)$ from (5.8) and (5.9). The value of ρ corresponding to the nominal value of c is $\rho_0 = \lambda + \beta$. Since we require a value of c that is larger than the nominal value for large β, however, let us consider an expression for ρ of the form

$$\rho_{\beta_{\mathrm{crit}}^{-1}} = \lambda + (1 - \frac{\lambda}{\beta_{\mathrm{crit}}})\beta \tag{5.10}$$

where $0 \ll \beta_{\mathrm{crit}} < N$ corresponding to $0 < \lambda_{\mathrm{crit}} \ll 1$. Thus,

$$cN = \frac{\beta}{1 - \beta/\beta_{\mathrm{crit}}}. \tag{5.11}$$

The value of N/β_{crit} is related to the minimum time constant of the exponential window that is capable of maintaining a full-rank estimate of $\mathbf{R}$. The linear approximation implicit in (5.10) does not, of course, provide the predictive power to claim that $\mathrm{tr}\,E(\mathbf{P})$ does not exist for $\beta = \beta_{\mathrm{crit}}$, but it does provide a qualitatively valuable interpretation for the behaviour under consideration. We might expect, for example, that the value of β_{crit} depends on the eigen-structure of $\mathbf{R}$. This implies that while the convergence properties and estimation misadjustment for the RLS process are independent of the input eigen-structure, its tracking performance (lag misadjustment) is not. An increase in β_{crit} corresponds to an increase in the lag misadjustment. As we shall see subsequently, this behaviour can be demonstrated via numerical experiments. For the present, however, we will take $\beta_{\mathrm{crit}} = 2$ to be a fair approximation.

This choice has a number of nice properties. First, it is in reasonable agreement with a rather wide range of numerical experiments. Also noteworthy is the direct comparison it allows with the normalized least-mean-squares (NLMS) algorithm. For $\beta_{\text{crit}} = 2$ and moderate to large values of N, we have $\rho \approx \psi^{-1}$ and, consequently, $\nu \approx \tilde{\nu} = \beta\psi$. Moreover, the NLMS steady-state misadjustment (3.4) is evidently similar to (5.6) for $\nu \approx \tilde{\nu}$. Finally, the minimization of (5.6) is greatly simplified for $\nu = \tilde{\nu}$.

5.3 RLS vs. NLMS Performance

At this point, we briefly examine the convergence behaviour of the RLS algorithm. While the independence of $\mathbf{v}^T\mathbf{x}$ and $\mathbf{x}$ is only a good approximation in the steady state, we may still reasonably take the weight error vector, $\mathbf{v}$, to be independent of the input when the adaptive filter is far from convergence. In this event, (5.1) becomes

$$\mathbf{v}_{k+1} = \left(\mathbf{I} - \frac{\mathbf{P}_k\mathbf{x}_k\mathbf{x}_k^T}{\lambda + \mathbf{x}_k^T\mathbf{P}_k\mathbf{x}_k}\right)\mathbf{v}_k + \mathbf{z}_k - \frac{\mathbf{P}_k\mathbf{x}_k}{\lambda + \mathbf{x}_k^T\mathbf{P}_k\mathbf{x}_k}e_k$$

or, equivalently,

$$\mathbf{v}_{k+1} = \lambda\mathbf{P}_{k+1}\mathbf{P}_k^{-1}\mathbf{v}_k + \mathbf{z}_k - \frac{\mathbf{P}_k\mathbf{x}_k}{\lambda + \mathbf{x}_k^T\mathbf{P}_k\mathbf{x}_k}e_k$$

In consequence, the misadjustment, which is now given by

$$\mathcal{M}_k = \text{tr}\left[\mathbf{R}E\left(\mathbf{v}_k\mathbf{v}_k^T\right)\right],$$

is updated as

$$\mathcal{M}_{k+1} \approx \lambda^2\mathcal{M}_k + q/N + \psi\beta\nu\sigma_e^2/N \tag{5.12}$$

given that

$$\text{tr}\, E\left(\mathbf{R}\mathbf{P}_{k+1}\mathbf{P}_k^{-1}\mathbf{v}_k\mathbf{v}_k^T\mathbf{P}_k^{-1}\mathbf{P}_{k+1}\right) \approx \text{tr}\, E\left(\mathbf{R}\mathbf{v}_k\mathbf{v}_k^T\right).$$

We note, in passing, that for $\nu = \beta\psi^{-1}$, expression (5.12) can be solved for the steady-state misadjustment which becomes equivalent to (5.6) and (5.7). Unfortunately, this

implies that $\sqrt{3}-1 < \beta_{\text{crit}} < 1$, which is far too low. At this point, however, we are more interested in the apparent time constant for the convergence of the RLS misadjustment sequence, viz.,

$$\tau_{\text{RLS}} = \frac{1}{1-\lambda^2}.$$

As shown in [40], [20] and Chapter 3, the corresponding value for the NLMS algorithm with uncorrelated inputs is

$$\tau_{\text{NLMS}} = \frac{N}{\bar{\mu}(2-\bar{\mu})}.$$

We note that the NLMS time constant cannot become less than N. Indeed, when the filter inputs are correlated, even this limit cannot be achieved. On the other hand, the misadjustment time constant of the RLS algorithm is, in principle, limited by unity. In practice, of course, we require that the windowed input be sufficiently persistently exciting so that $\mathbf{P}$ remains positive definite. In consequence, the reduction of τ may result in instability due to insufficient excitation. While it is tempting to limit the value of τ to $N/(1+\lambda)\beta_{crit}$, we reiterate that the introduction of β_{crit} involves a linear approximation that is not strong enough to limit RLS behaviour. On the other hand, reduced-τ adaptation provides greater robustness when finite precision arithmetic is used [52].

We are now in a position to compare the lag and estimation errors for the RLS and NLMS algorithms in a fair manner by equating their time constants under the assumption of uncorrelated inputs, as in [16]. Taking $\nu = \tilde{\nu}$ and $\tau_{\text{NLMS}} = \tau_{\text{RLS}} = \tau$, we have

$$\bar{\mu} = 1 - \sqrt{1 - N\tau^{-1}}$$

and

$$\tilde{\nu} = \frac{2N}{\frac{1+\sqrt{1-\tau^{-1}}}{1-\sqrt{1-\tau^{-1}}} + N}.$$

The resulting estimation misadjustments, namely, $\mathcal{M}_\infty^e \triangleq \mathcal{M}_\infty|_{q=0}$, for the two algorithms are now

$$\begin{aligned}\mathcal{M}^e_{\infty,\text{NLMS}} &= \frac{1-\sqrt{1-N\tau^{-1}}}{1+\sqrt{1-N\tau^{-1}}} \\ \mathcal{M}^e_{\infty,\text{RLS}} &= N\frac{1-\sqrt{1-\tau^{-1}}}{1+\sqrt{1-\tau^{-1}}}.\end{aligned} \tag{5.13}$$

The corresponding lag errors, namely, $\mathcal{M}_\infty^l \triangleq \lim_{q\to\infty}\mathcal{M}_\infty$, on the other hand, are given by

$$\begin{aligned}\mathcal{M}^l_{\infty,\text{NLMS}} &= \frac{q\tau}{N} \\ \mathcal{M}^l_{\infty,\text{RLS}} &= \frac{q}{N}\left(\frac{1+\sqrt{1-\tau^{-1}}}{1-\sqrt{1-\tau^{-1}}}+N\right)^2\left(\frac{1-\sqrt{1-\tau^{-1}}}{1+\sqrt{1-\tau^{-1}}}\right).\end{aligned} \tag{5.14}$$

The large-τ and large-N approximations of these equations are

$$\begin{aligned}\mathcal{M}^e_{\infty,\text{NLMS}} &\approx \left(\frac{4\tau}{N}-1\right)^{-1} \\ \mathcal{M}^e_{\infty,\text{RLS}} &\approx \frac{N}{4\tau} \\ \mathcal{M}^l_{\infty,\text{NLMS}} &= \frac{q\tau}{N} \\ \mathcal{M}^l_{\infty,\text{RLS}} &\approx \frac{q\tau}{N}+\frac{q}{2}+\frac{qN}{4\tau}\end{aligned} \tag{5.15}$$

We observe that the RLS estimation error is always better than that of the NLMS algorithm for equal convergence properties while the NLMS lag error is superior. For small values of $N/\tau \approx 2\beta$, however, these errors become asymptotically equivalent, in agreement with [16]. It is also interesting that the minimum achievable MSE, requiring precise tuning of their respective convergence-controlling parameters, is identical for RLS and NLMS adaptive filters under the above approximations.

5.4 Algorithm Development

In this section, we introduce a new scheme by which to govern the forgetting factor of the RLS algorithm in a time-variable manner in accordance with the parallel adapta-

tion (PA) ideas introduced previously. Under the interpretation of the environment as a random walk, the nonstationarity to noise ratio is estimated. This estimate is then used to obtain an estimate of the "optimal" forgetting factor. The optimality attempted here is again with respect to the steady-state misadjustment. To this end, we substitute $\tilde{\nu}$ for $\tilde{\mu}$ in (4.1). The optimal forgetting factor is then obtained by

$$\lambda^*(q) = \frac{(2-\tilde{\nu}^*)N - \tilde{\nu}^*}{(2-\tilde{\nu}^*)N + \tilde{\nu}^*} = \frac{\left[q+4-\sqrt{q(q+4)}\right]N - \sqrt{q(q+4)} - q}{\left[q+4-\sqrt{q(q+4)}\right]N + \sqrt{q(q+4)} - q}. \tag{5.16}$$

We observe that as q vanishes, λ^* approaches unity. As q becomes large, on the other hand, λ^* approaches the value $(N-1)/(N+1)$. It is interesting to note that this limit corresponds to that of the estimation-error-equivalent sliding-rectangular-window RLS process [51].

Before we proceed, it is of interest to compare the result above with other "optimal" forgetting-factor results. In particular, we compare the RLS adaptive filter with the optimal Kalman one-step predictor, that is, the Kalman filter that explicitly minimizes the mean squares *a priori* error, having precise knowledge of the statistics of the plant and measurement noises. In a steady-state random walk scenario with uncorrelated inputs, the RLS adaptive filter with optimal forgetting factor is equivalent to this optimal Kalman filter. This can be easily seen as follows. The standard Kalman state-covariance ($\tilde{\mathbf{P}}$) update equation is

$$\tilde{\mathbf{P}}_{k+1} = \tilde{\mathbf{P}}_k - \frac{\tilde{\mathbf{P}}_k \mathbf{x}_k \mathbf{x}_k^T \tilde{\mathbf{P}}}{\sigma_\epsilon^2 + \mathbf{x}_k \tilde{\mathbf{P}}_k \mathbf{x}_k} + \sigma_z^2 \mathbf{I}_N.$$

Matching this equation with RLS $\mathbf{P}$-update in the steady state using the relation

$$\mathbf{P} = \frac{\lambda}{\sigma_\epsilon^2}\tilde{\mathbf{P}},$$

gives

$$\frac{\lambda^2 \sigma_z^2}{\sigma_\epsilon^2}\mathbf{I}_N = (1-\lambda)E\mathbf{P}_\infty$$

Finally, for $\lambda \approx 1$ (but not necessarily $\beta \approx 0$), this gives $q \approx \beta c N$. Using the approximation $c = c_0$, the optimal value for the forgetting factor becomes

$$\lambda_0^* = 1 - \frac{\sqrt{q}}{N}, \tag{5.17}$$

which can also be obtained by finding the minimum of (5.7) under the same asymptotic condition. On the other hand, if we permit the linear approximation for ρ above, we have

$$\lambda_1^* = 1 - \frac{q}{2\beta_{\text{crit}} N}\left(\sqrt{1 + 4\beta_{\text{crit}}^2/q} - 1\right). \tag{5.18}$$

We see that as q approaches zero, (5.18) and (5.16) approach (5.17). As q increases, on the other hand, both of these expressions approach $1 - \beta_{\text{crit}}/N$.

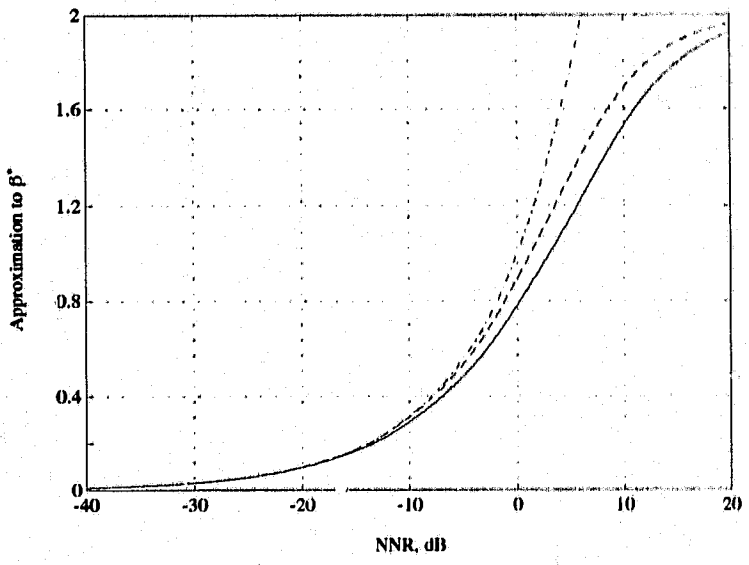

Figure 5.1: Estimates of optimal random walk β

———— $\beta^* = (1 - \lambda^*)N$

– – – – – – – – $\beta^* = (1 - \lambda_1^*)N$

– · – · – · – · – · $\beta^* = (1 - \lambda_0^*)N$

In Figure 5.1, the corresponding values of β^* for these three approximations for the optimal forgetting factor are compared against NNR graphically. The value of

$\beta_{\text{crit}} = 2$ is used for (5.18). We remark that the difference between (5.16) and (5.18) is not very significant with respect to the steady-state misadjustment as will be made clear in the experiments.

We now require an estimate of q. Unfortunately, as in the case of PA-NLMS, this quantity is unobservable for one RLS process. Nevertheless, let us proceed at this point under the assumption that some *a priori* knowledge about the additive noise power in the form of an estimate, say $\hat{\sigma}_e^2$, exists. Indeed, the variable-forgetting-factor algorithm due to Fortescue, Kershenbaum and Ydstie (FKY hereafter) also requires this knowledge of the additive noise power [35]. In this event, we can estimate the nonstationarity to noise ratio using

$$\hat{q}_{k+1} = \max\left\{0, \lambda_q \hat{q}_k + (1-\lambda_q)\tilde{\nu}_k \left[\left(e_{k+1}^2/\hat{\sigma}_e^2 - 1\right)(2-\tilde{\nu}_k) - \tilde{\nu}_k\right]\right\}. \tag{5.19}$$

Using (5.16), we can now estimate the optimal forgetting factor for use in the RLS equations in the next iteration, i.e. $\lambda_{k+1} = \lambda^*(\hat{q}_{k+1})$.

We now consider the choice of λ_q. In general, we require a high quality estimator when q is small so that the misadjustment due to its estimation, which includes the nonlinearity of (5.19), is minimized. When q is large, however, it is not so essential that our estimate of q be of high quality since all large estimates of q will result in roughly equal values of λ. With this in mind, we take $\lambda_q = \lambda_k$. While this choice provides exceptional performance for unknown stationary adaptation and random walk scenarios, the adaptive filter behaves poorly in another important scenario of interest, namely, when there is an abrupt change in the target filter after a period of stationarity. In this event, the adaptive filter gets "trapped" in a $\lambda \approx 1$ state. In order to work around this problem, we use a slightly modified version of (5.16), namely,

$$\tilde{\lambda}^*(q) = \lambda^*(q) - \beta_{\min}/N \tag{5.20}$$

where $\beta_{\min}$ is a small constant chosen to approximate a stationary steady-state misadjustment specification. For example, if -30 dB misadjustment were required, we would take $\beta_{\min} = 10^{-3}$ which permits a roughly 3 dB discrepancy between actual and nominal misadjustments.

In many applications, an *a priori* estimate of the additive noise power does not exist. An addition to the above algorithm will be necessary in order to successfully deal with these scenarios. We propose a reliable, if somewhat expensive, estimator of the additive noise power: a fixed-forgetting-factor RLS "indicator" process. If the forgetting factor of this process is sufficiently low, the estimation misadjustment will dominate under most practical conditions, providing an estimate of the form

$$\hat{\sigma}^2_{\epsilon,k+1} = \lambda_\epsilon \hat{\sigma}^2_{\epsilon,k} + (1-\lambda_\epsilon)(1-\tilde{\nu}_i/2)e^2_{i,k+1}$$

where the subscript i denotes that the respective quantities are due to the indicator process. For the same reasons as discussed for our q estimator, we take $\lambda_\epsilon = \lambda_k = \tilde{\lambda}^*(\hat{q}_k)$.

The resulting variable forgetting factor RLS procedure will be called PA-RLS by analogy to the PA-NLMS algorithm described in the previous chapter. The PA-RLS algorithm is summarized as follows:

1. Initialize primary and indicator RLS processes.

2. Obtain λ_k as

$$\hat{\sigma}^2_{\epsilon,k} = \lambda_{k-1}\hat{\sigma}^2_{\epsilon,k-1} + (1-\lambda_k)(1-\tilde{\nu}_{i,k-1}/2)e^2_{i,k}$$

$$\hat{q}_k = \max\left\{0, \lambda_{k-1}\hat{q}_{k-1} + (1-\lambda_{k-1})\tilde{\nu}_{k-1}\left[\left(e^2_{k-1}/\hat{\sigma}^2_{\epsilon,k-1} - 1\right)(2-\tilde{\nu}_{k-1}) - \tilde{\nu}_{k-1}\right]\right\}$$

$$\lambda_k = \frac{\left[\hat{q}_k + 4 - \sqrt{\hat{q}_k(\hat{q}_k+4)}\right]N - \sqrt{\hat{q}_k(\hat{q}_k+4)} - \hat{q}_k}{\left[\hat{q}_k + 4 - \sqrt{\hat{q}_k(\hat{q}_k+4)}\right]N + \sqrt{\hat{q}_k(\hat{q}_k+4)} - \hat{q}_k} - \frac{\beta_{\min}}{N}.$$

3. RLS update (in parallel):

$$\mathbf{P}_{k+1} = \frac{1}{\lambda_k}\left(\mathbf{P}_k - \frac{\mathbf{P}_k\mathbf{x}_k\mathbf{x}_k^T\mathbf{P}_k}{\lambda_k + \mathbf{x}_k^T\mathbf{P}_k\mathbf{x}_k}\right)$$

$$\mathbf{P}_{i,k+1} = \frac{1}{\lambda_i}\left(\mathbf{P}_{i,k} - \frac{\mathbf{P}_{i,k}\mathbf{x}_k\mathbf{x}_k^T\mathbf{P}_{i,k}}{\lambda_i + \mathbf{x}_k^T\mathbf{P}_{i,k}\mathbf{x}_k}\right)$$
$$\hat{\mathbf{w}}_{k+1} = \hat{\mathbf{w}}_k + \frac{\mathbf{P}_k\mathbf{x}_k e_k}{\lambda_k + \mathbf{x}_k^T\mathbf{P}_k\mathbf{x}_k}$$
$$\hat{\mathbf{w}}_{i,k+1} = \hat{\mathbf{w}}_{i,k} + \frac{\mathbf{P}_{i,k}\mathbf{x}_k e_{i,k}}{\lambda_i + \mathbf{x}_k^T\mathbf{P}_{i,k}\mathbf{x}_k}.$$

4. Repeat from 2.

5.5 Discussion

The correction for fourth-order statistics found in [16] provides improved estimates of the steady-state RLS misadjustment quantities for low values of N. It is, of course, possible to solve for ψ taking into consideration this correction if necessary.

The FKY algorithm as examined in Chapter 2 was originally intended to provide greater robustness in the presence of impersistently exciting inputs. In effect, it is built upon an environmental model that has stationary additive noise, a very slowly varying desired signal, and nonstationary input. In this environment, as the excitation diminishes, the FKY forgetting factor approaches $1 - \beta_0/N$. For the PA-RLS algorithm, on the other hand, the forgetting factor approaches $1 - \beta_{\min}/N$ regardless of the excitation. In consequence, we expect that the PA-RLS algorithm will also provide improved robustness in the presence of impersistently exciting inputs.

A qualitative look at the behaviour of a fixed-point implementation of the PA-RLS algorithm is in order. In [52], an expression similar to (5.7) for the errors due to fixed-point arithmetic has been derived in which the quantization effects of computing the *a priori* errors correspond to the additive noise power and those from the computation of the so-called Kalman gain correspond to the random walk power. Further, the assumption that $c = c_0$ is made in that work. Indeed, it has been shown that the finite precision quantization error system is of the random walk type [53]. We might expect,

in consequence, that the effects of quantization errors are similar to the effects of additive noise and random walk target filter fluctuations. In effect, even though there are four sources of error when the fixed-point RLS algorithm is applied to a random walk scenario, their dependence on the forgetting factor permits them to be grouped into two pairs. In consequence, if the PA-RLS algorithm were implemented with fixed-point arithmetic, while the meanings of the estimated quantities would change, the effect of approximating the optimal forgetting factor would remain. And while the resulting variable-forgetting-factor behaviour may not be exactly optimal, it would provide improved performance in fixed-point implementation and in nonstationary environments over a fixed-forgetting-factor algorithm.

The PA-RLS algorithm will still have all of the classical RLS problems with explosive divergence. For large N and highly correlated inputs, this problem can show up in even relatively long word-length implementations. To date, analysis of the explosive divergence problem has made the assumption that $c = c_0$, and proposed solutions introduce biases into the RLS process [54]. Since these biases are usually at the least-significant bit level, direct application of the PA-RLS algorithm in stabilized form is conceivable.

Unfortunately, the proposed variable-forgetting-factor algorithm is impossible to implement directly in an efficient (so-called fast [55]) manner since the shifting properties necessary for such a formulation no longer exist when the forgetting factor is time-variable [56]. Since the optimality considered above is with respect to the steady state the applicability of the PA-RLS scheme in an alternate form, such as those found in [57], deserves consideration.

In the following chapter, the performance of the PA-RLS algorithm will be examined by way of simulations of a number of different nonstationary environments.

Chapter 6

Simulations

6.1 Preliminaries

In this chapter, the results of a number of simulations will be presented in order to test the previous analysis and examine the performance of existing and proposed algorithms. First, results of experiments measuring the value of c in conventional RLS adaptation will be presented. The sensitivity of the RASE and PA methods to their algorithmic parameters will then be investigated. Finally, the comparative performance between these algorithms and the current DA schemes will be examined in a variety of nonstationary environments.

In order to facilitate the numerous experiments to follow, we construct a simple application framework into which all of the subsequent simulations will fit. To this end, we make use of the construct illustrated in Figure 6.1, often referred to as a system identification configuration.

In all of the experiments, the additive noise was modeled to be stationary and white. This immediately gives an unfair advantage to the DA methods over the conventional algorithms. The conventional RLS and NLMS adaptive filters are almost immune to the statistics of a zero-mean and not extremely correlated additive noise. Of existing DA methods, on the other hand, only the HCB algorithm is immune

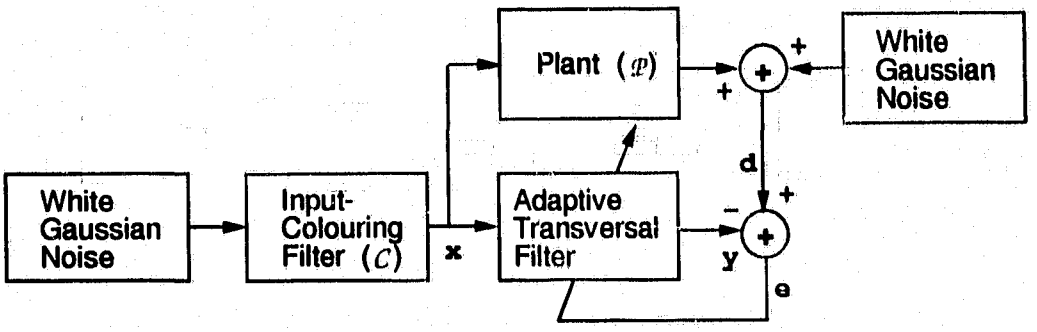

Figure 6.1: A system identification configuration.

to additive noise nonstationarities. It is, however, quite sensitive to additive noise correlatedness. So as not to bias the results either for or against this algorithm, a stationary and white additive noise will be adopted.

In accordance with Figure 6.1, the adaptive filter input will be taken to be the response of a fixed input-colouring filter, $\mathcal{C}$, to white Gaussian noise. In order to simplify the filter specification procedure, we consider the family of even-order (odd-length) FIR filters denoted by $\mathcal{F}^{\star}_{L,p}$ having an impulse response given by

$$h(n) = \kappa^p (\star 1)^n \begin{cases} (n+1)^p, & 0 \le n \le \frac{L-1}{2} \\ (L-n)^p, & \frac{L-1}{2} \le n \le L-1 \\ 0, & \text{otherwise} \end{cases}$$

where $\star \in \{+,-\}$ and κ is chosen such that

$$\sum_{n=-\infty}^{\infty} |h(n)|^2 = 1.$$

Filters from this family will be used in the specification of both the input-colouring filter, $\mathcal{C}$, and plant model, $\mathcal{P}$, in the system identification configuration.

6.2 Measuring $cN = \mathrm{tr}\, E(\mathbf{PR})$

Simulated measurements of the steady-state value of $\mathrm{tr}\, E(\mathbf{RP}_k)$ for conventional RLS adaptation will now be presented and compared with the various candidates

we have for cN. For this experiment, there is no need to have a full adaptive filter configuration. All that is necessary is to apply the RLS **P**-update equation to the input vector at each iteration. We will, however, take these input vectors to be as if they were the input of an adaptive transversal filter, according to (1.1). This experiment was performed with two choices of input colouring filters, representing uncorrelated and correlated input values. The first of these, $\mathcal{C}_u = \mathcal{F}^{+}_{9,\infty}$, simply passes its input. The second, $\mathcal{C}_c = \mathcal{F}^{-}_{9,1}$, results in an input covariance eigenvalue ratio of 3.09×10^3 for $N = 10$ and 3.76×10^6 for $N = 100$. The ratio of the largest eigenvalue of **R** to its smallest is considered to be a rough measure of the correlatedness of the input signal [6]. For the uncorrelated input case, this measure attains its minimum of unity. These two input colouring filters are, as a consequence, taken to represent the two extremes of input correlatedness. We also perform this experiment for both $N = 10$ and $N = 100$, so that we may examine the dependence of c on the filter order. The trace of **RP** was then averaged over $K = 10^5$ steady-state (with respect to the input) samples, and the results are tabulated in Table 6.1. The theoretical values are defined as follows: $c_{\beta^{-1}_{\text{crit}}} \triangleq \beta/(1 - \beta/\beta_{\text{crit}})$ and $\tilde{c} \triangleq \lambda\tilde{\mu}/(1 - \tilde{\mu})$.

6.2.1 Discussion

A number of interesting observations can be made from these results. First, we see that the value of c is not particularly sensitive to the correlatedness of the input samples. It is also interesting that for $N = 10$, c increases with input correlatedness while the converse takes place for the larger N. We would expect, in consequence, that the lag misadjustment for correlated input values will be much lower than that for uncorrelated inputs for low values of N. Of course, the behaviour of c is not the only important factor in the steady-state misadjustment. The assumption that v is independent of $\mathbf{x}$ apparently required that N be large relative to the input

Table 6.1: Estimates of $E[\mathrm{tr}\,(\mathbf{RP})]$.

$\mathcal{C}$	N	$\hat{E}\,[\mathrm{tr}\,(\mathbf{RP})]$	$c_0 N$	$c_{0.4} N$	$\tilde{c} N$
$\mathcal{C}_u$	10	3.64	1.50	3.75	7.29
$\mathcal{C}_u$	10	2.55	1.25	2.50	3.50
$\mathcal{C}_u$	10	1.73	1.00	1.67	2.00
$\mathcal{C}_u$	10	1.12	0.75	1.07	1.18
$\mathcal{C}_u$	10	0.65	0.50	0.63	0.66
$\mathcal{C}_u$	10	0.29	0.25	0.28	0.28
$\mathcal{C}_u$	100	3.26	1.50	3.75	6.09
$\mathcal{C}_u$	100	2.36	1.25	2.50	3.35
$\mathcal{C}_u$	100	1.66	1.00	1.67	2.00
$\mathcal{C}_u$	100	1.10	0.75	1.07	1.20
$\mathcal{C}_u$	100	0.66	0.50	0.63	0.67
$\mathcal{C}_u$	100	0.29	0.25	0.28	0.29
$\mathcal{C}_c$	10	4.70	1.50	3.75	7.29
$\mathcal{C}_c$	10	3.21	1.25	2.50	3.50
$\mathcal{C}_c$	10	2.08	1.00	1.67	2.00
$\mathcal{C}_c$	10	1.29	0.75	1.07	1.18
$\mathcal{C}_c$	10	0.72	0.50	0.63	0.66
$\mathcal{C}_c$	10	0.30	0.25	0.28	0.28
$\mathcal{C}_c$	100	3.15	1.50	3.75	6.09
$\mathcal{C}_c$	100	2.28	1.25	2.50	3.35
$\mathcal{C}_c$	100	1.56	1.00	1.67	2.00
$\mathcal{C}_c$	100	1.03	0.75	1.07	1.20
$\mathcal{C}_c$	100	0.60	0.50	0.63	0.67
$\mathcal{C}_c$	100	0.27	0.25	0.28	0.29

$\hat{E}\,[\mathrm{tr}\,(\mathbf{RP})] = 10^{-5} \sum_{k=1}^{10^5} \mathrm{tr}\,(\mathbf{RP}_k)$

autocorrelation. For $C = C_c$, this would suggest that the assumption is not valid for $N \not> 20$. In a following experiment, however, we find that the lag misadjustment is, in fact, lower for correlated inputs, and further that this effect diminishes as the value of N increases, as the measurements of cN would suggest. We also see that both $c_{0.4}$ and $\tilde{c} \approx c_{0.5}$ provide reasonable approximations to the behaviour of c, especially for low values of β.

6.3 Sensitivity Experiments

6.3.1 Introduction

The present section attempts a modest enquiry into the sensitivity of the RASE and PA algorithms to those components of their respective parameters and initializations that have not been acceptably addressed in the previous analysis. The need for such an examination has been emphasized previously. Rather than providing an comprehensive mapping of the algorithm performances with respect to their parameters in the manner of Figure 4.2, the present experiments will make an effort to answer the following questions:

(a) Is the algorithm reasonably insensitive to a given parameter?

(b) If not, can an acceptable value for said parameter be known *a priori*?

In this manner, the practicality of the algorithm under consideration may be assessed. If an affirmative answer to either of the above questions is forthcoming, we can expect that the choice of parameters and initialization will not unduly degrade the performance of the method at hand.

6.3.2 RASE Sensitivity

The practicality of the reduced adaptation state estimation algorithm is immediately called into question by the relatively large number of parameters needed to specify its behaviour. Initial estimates of the instantaneous misadjustment and the additive noise power, as well as the three diagonal entries of the pseudo plant noise covariance $\mathbf{Q}$, are required. To be complete, the pole placement of the μ-smoothing filter and the lower limit to the convergence-controlling parameter should also be addressed. In the following simulations, however, these quantities will be fixed at the values recommended in Chapter 3. In other words, in spite of the answer to question (a), above, with respect to these last two quantities, the answer to question (b) is taken to be positive. As a result, this subsection will consider RASE sensitivity to $\mathbf{Q}$-parameter selection and initialization.

Sensitivity to parameters

Attention will now be given to the sensitivity of RASE performance to the choices made for the diagonal elements of $\mathbf{Q}$, the pseudo plant noise covariance. To begin, the ability of the RASE algorithm to respond to a sudden change in desired signal statistics is considered. Such a change was simulated by having the plant change from $\mathcal{F}^{+}_{21,1}$ to $\mathcal{F}^{+}_{21,\infty}$ at sample $k = 8000$. Estimates of the learning curve in this and subsequent experiments were acquired by averaging over 50 independent trials. These learning curve estimates were also smoothed, resulting in an apparent anticipation of the discontinuity and an artificial dip at the end of the experiments. The input samples are uncorrelated, the balance of $\mathbf{Q}$ is zero, and the initial estimate of the additive noise power is known to be correct.

The results of this experiment are shown in Figure 6.2. For $q_{11} = 0$, RASE

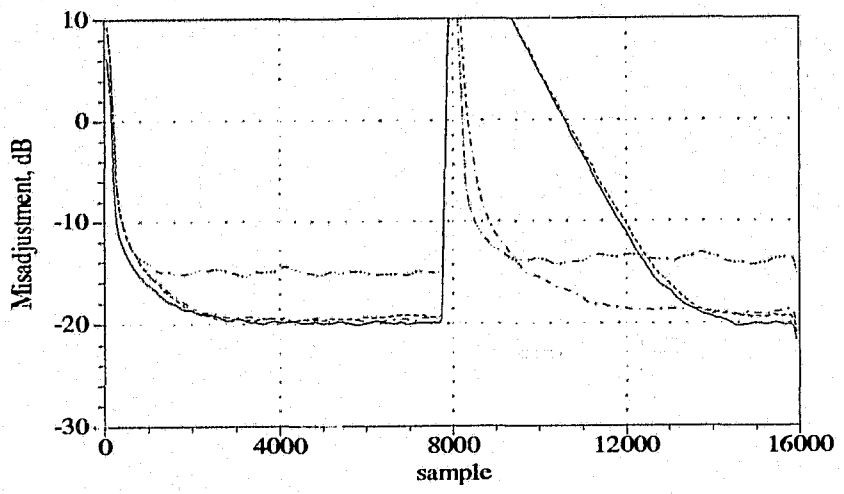

Figure 6.2: RASE sensitivity to q_{11}.

——————— RLS: $\lambda = 0.99905$
– – – – – – – – RASE: $q_{11} = 0$
– · – · – · – · – RASE: $q_{11} = 10^{-10}$
– ··· – ··· – ··· – RASE: $q_{11} = 10^{-8}$

performance is comparable to RLS with a forgetting factor chosen to match the steady-state misadjustment. An estimate of the RLS learning curve is also shown for comparison. For $q_{11} = 10^{-10}$, we see that the RASE reconvergence properties are enhanced without significant degradation in MAC. As q_{11} increases, however, an eventual loss in MAC performance is realized while the reconvergence curve matches first convergence.

The same simulation was used in order to examine the dependence of RASE performance on the parameter q_{22}. Moreover, an inkling is given of how the RASE parameters interact by changing both q_{11} and q_{22}. Perfect initialization was used once again. The results of this experiment are shown in Figure 6.3. It can be seen that the reconvergence behaviour of the RASE algorithm is much poorer for values of $q_{22} > 0$. The magnitude of q_{22} with respect to that of q_{11} reflects the *a priori* expectation

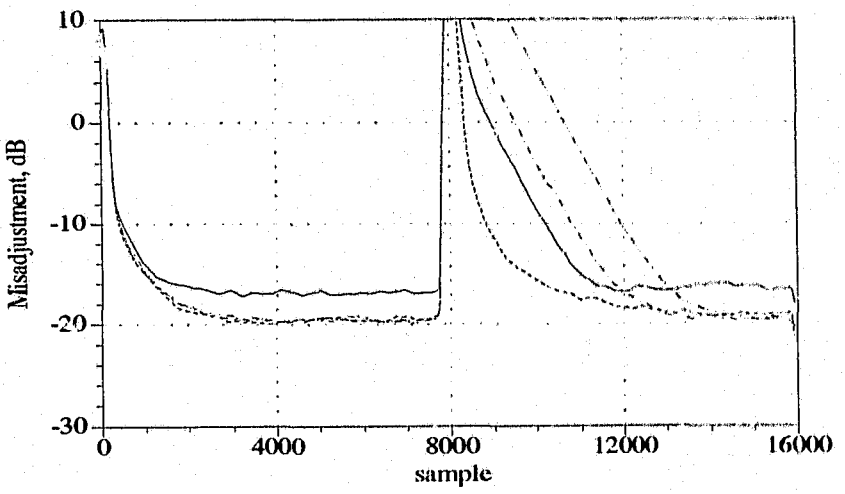

Figure 6.3: RASE sensitivity to q_{22}.

———— RASE: $q_{22} = 10^{-8}$; $q_{11} = 10^{-8}$
– – – – – – – – RASE: $q_{22} = 0$; $q_{11} = 10^{-10}$
– · – · – · – · – RASE: $q_{22} = 10^{-10}$; $q_{11} = 10^{-10}$
– ·· – ·· – ·· – RASE: $q_{22} = 10^{-8}$; $q_{11} = 10^{-10}$

that a sudden change in error power is due to a target discontinuity or a increase in MMSE. For example, if q_{22} is relatively large, an additive noise power fluctuation is preferred over the target discontinuity. On the other hand, if $q_{22} = 0$, as shall be the case in all subsequent simulations, this results in the opposite EKF behaviour.

A similar experiment was performed in order to test the sensitivity of RASE performance to the parameter q_{33}. In this simulation, however, the input samples are highly correlated via an input colouring filter of the form $\mathcal{F}_{9,1}^{-}$. This results in an input covariance eigenvalue ratio of 19561. Further, two discontinuities in the plant occur at $k = 5000$ and $k = 9000$, the latter having the plant revert to its original $\mathcal{F}_{21,1}^{+}$. The reason for this modification is that the behaviour of the NLMS algorithm depends on the interaction between input correlatedness and adaptation state. The pair of plant discontinuities provide a greater range for this phenomenon. The parameter q_{11} took

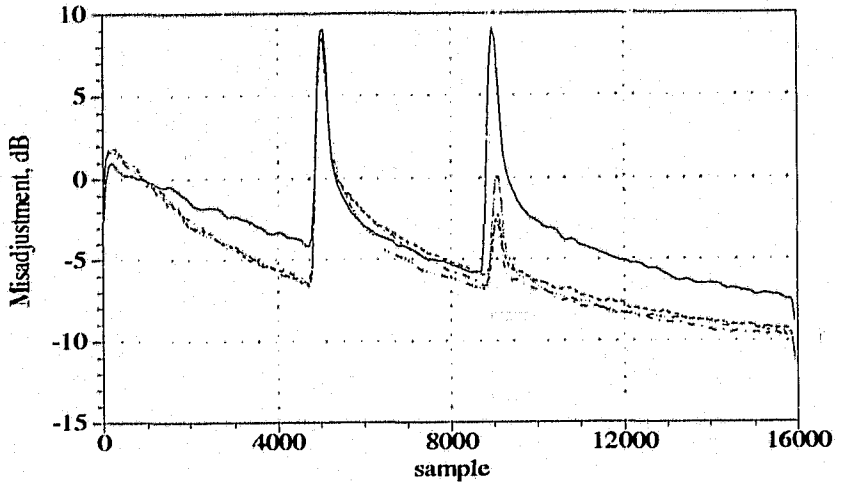

Figure 6.4: RASE sensitivity to q_{33}.

NLMS: $\bar{\mu} = 0.182$

RASE: $q_{33} = 0$

RASE: $q_{33} = 10^{-10}$

RASE: $q_{33} = 10^{-8}$

the value 10^{-10} for this simulation and the additive noise power was again taken to be known.

The results of this experiment are shown in Figure 6.4. It is seen that while modest improvement is possible by increasing the value of q_{33} in the presence of correlated input samples, the RASE algorithm is not particularly dependent on the choice of this parameter. The estimate of an NLMS learning curve with similar MAC performance is also shown in Figure 6.4 for comparison.

Sensitivity to initialization

As has been suggested in Chapter 3, the performance of the RASE algorithm depends heavily on its initialization. In particular, the initial additive noise power estimate and the quality assigned to it in the choice of $\mathbf{P}_{0,22}$ have a significant effect on the behaviour

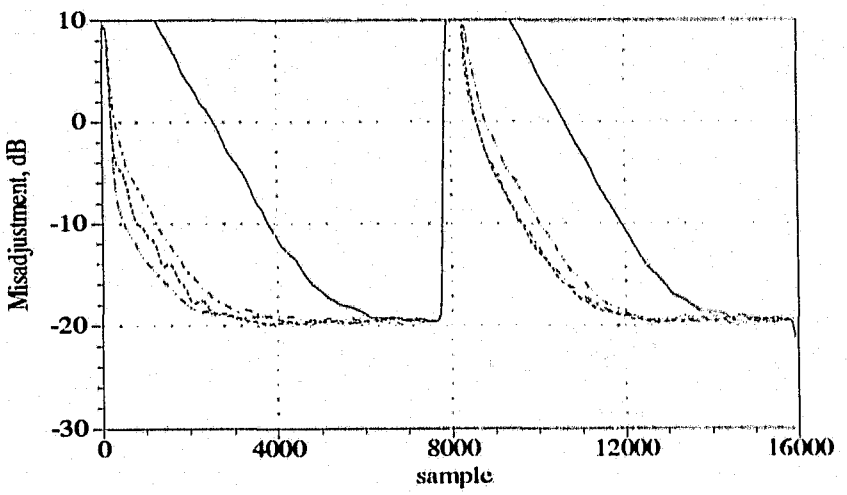

Figure 6.5: RASE sensitivity to $\hat{\sigma}^2_{\epsilon,0}$.

——————— NLMS: $\tilde{\mu} = 0.0198$
– – – – – – – – – RASE: $\hat{\sigma}^2_{\epsilon,0} = \sigma^2_{\epsilon,0}$
– · – · – · – · – RASE: $\hat{\sigma}^2_{\epsilon,0} = 5\sigma^2_{\epsilon,0}$
– ··· – ··· – ··· – RASE: $\hat{\sigma}^2_{\epsilon,0} = \sigma^2_{\epsilon,0}/5$

of the method. In this section, this dependence will be examined qualitatively by way of simulations. The same scenario as in the previous subsection was used for this purpose. Once again, input samples are uncorrelated, and $\mathbf{Q}$ is zero except for $q_{11} = 10^{-10}$. The initial additive noise power estimate and the corresponding initial state covariance entry were modified from the optimal $(\sigma^2_\epsilon\ ,\ 0)$.

The results of this experiment are found in Figures 6.5 and 6.6. In the former, we see the degradation in RASE performance when $\hat{\sigma}^2_\epsilon = f\sigma^2_\epsilon$ for $f = 5$ and $f = 1/5$. The value of $\mathbf{P}_{0,22}$ was 10^{-5} for these experiments. In Figure 6.6, on the other hand, $\hat{\sigma}^2_\epsilon = \sigma^2_\epsilon$, but $\mathbf{P}_{0,22}$ takes the values 10^{-3}, 10^{-4}, and 10^{-5}. It is clear that uncertainty about the additive noise power reduces the quality of RASE adaptation.

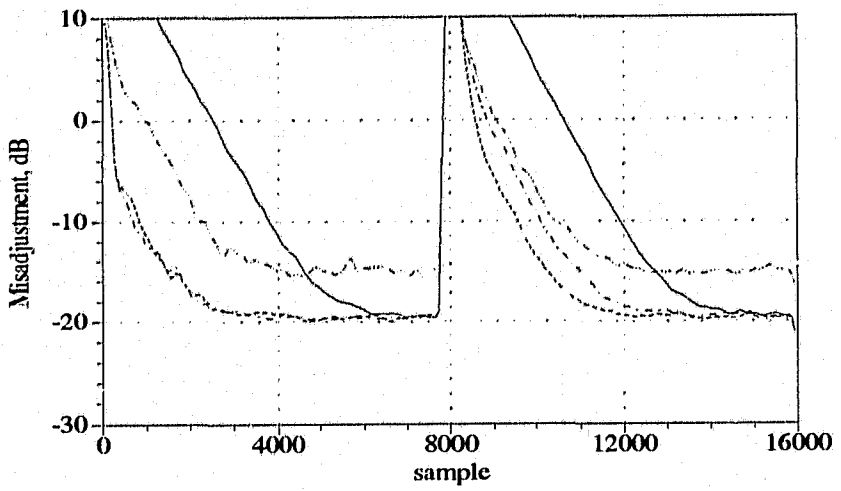

Figure 6.6: RASE sensitivity to $\mathbf{P}_{0,22}$.

——————— NLMS: $\bar{\mu} = 0{,}0198$
– – – – – – – – – RASE: $\mathbf{P}_{0,22} = 10^{-5}$
– · – · – · – · – · RASE: $\mathbf{P}_{0,22} = 10^{-4}$
– ··· – ··· – ··· – RASE: $\mathbf{P}_{0,22} = 10^{-3}$

6.3.3 PA-NLMS Sensitivity

In Chapter 4, provision was given to examine the steady-state performance of the PA-NLMS algorithm with respect to the parameters α, λ_ϵ, λ_z, and λ_μ. For the nonstationary case, this algorithm was shown to be insensitive to variations in these parameters over a typical slice of their parameter space. Since the construction of figures like Figure 4.2 is reasonably straightforward due to the formulas available in Chapter 4, this section will consider the manner in which the PA-NLMS algorithm depends upon the remaining parameter, namely ζ, and its initialization.

Sensitivity to parameters

Using the same simulation that was used previously with uncorrelated input samples, the convergence behaviour of the PA-NLMS algorithm is tested with respect to its

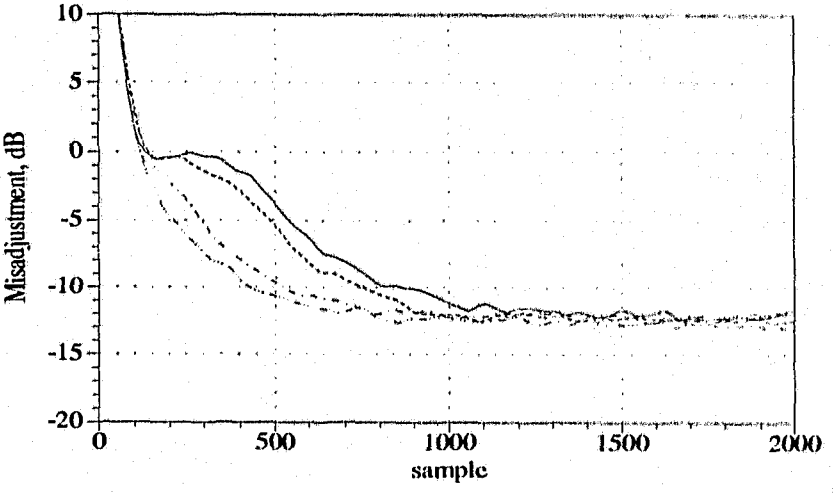

Figure 6.7: PA-NLMS sensitivity to ζ.
——— PA-NLMS: $\zeta = 0$
– – – – – – – – PA-NLMS: $\zeta = 0.001$
– · – · – · – · – PA-NLMS: $\zeta = 0.01$
– ··· – ··· – ··· – PA-NLMS: $\zeta = 0.1$

sensitivity to the parameter ζ. The balance of the PA-NLMS parameter set took on values $\alpha = 0.125$, $\lambda_c = 0.9975$, $\lambda_z = 0.975$, and $\lambda_\mu = 0.975$, and the initial estimates of the $\hat{q}$ components were zero. The estimated learning curves due to this algorithm for four different values of ζ are provided in Figure 6.7. It is clear that no significant degradation in the steady-state misadjustment is incurred by the utilization of $\zeta > 0$ — even for values exceeding sensible limits. Indeed, there seems to be no penalty for the improved convergence that results.

Sensitivity to initialization

In the previous section, the sensitivity of the RASE algorithm to the quality of its initial estimates of environmental quantities was demonstrated. The relative insensitivity of the PA-NLMS scheme to the analogous quantities will now be shown. The

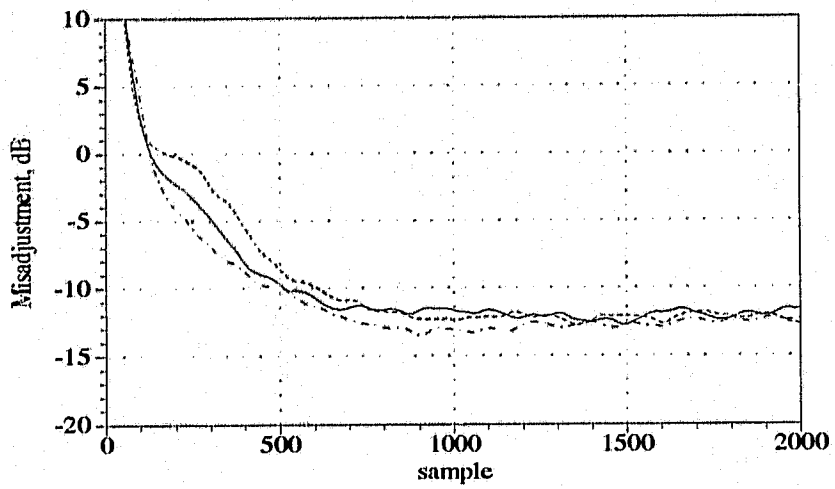

Figure 6.8: PA-NLMS sensitivity to $\hat{q}_{\epsilon,0}$ and $\hat{q}_{z,0}$.
——————— PA-NLMS: $\hat{q}_{\epsilon,0} = \hat{q}_{z,0} = 0$
– – – – – – – – PA-NLMS: $\hat{q}_{\epsilon,0} = 0.015$, $\hat{q}_{z,0} = 10$
– · – · – · – · PA-NLMS: $\hat{q}_{\epsilon,0} = 0.1$, $\hat{q}_{z,0} = 0$

experiment is the same as the previous one, except that the parameter ζ is fixed at 0.01 and the values of $\hat{q}_{\epsilon,0}$ and $\hat{q}_{z,0}$ are permitted to vary. As is evidenced in Figure 6.8, which summarizes the results of this experiment, the behaviour of the PA-NLMS algorithm is relatively fixed over a considerable range of initializations.

6.3.4 PA-RLS Sensitivity

The sensitivity of the PA-RLS algorithm to the choices of its parameters is now investigated. It shall be demonstrated that this adaptive filter is particularly robust toward the values chosen for β_i, $\beta_{\min}$, and $\hat{\sigma}_\epsilon^2$.

Sensitivity to parameters

The choice of β_i for the PA-RLS algorithm depends on the application. For example, if a very fast response were required to sudden changes in the desired signal statis-

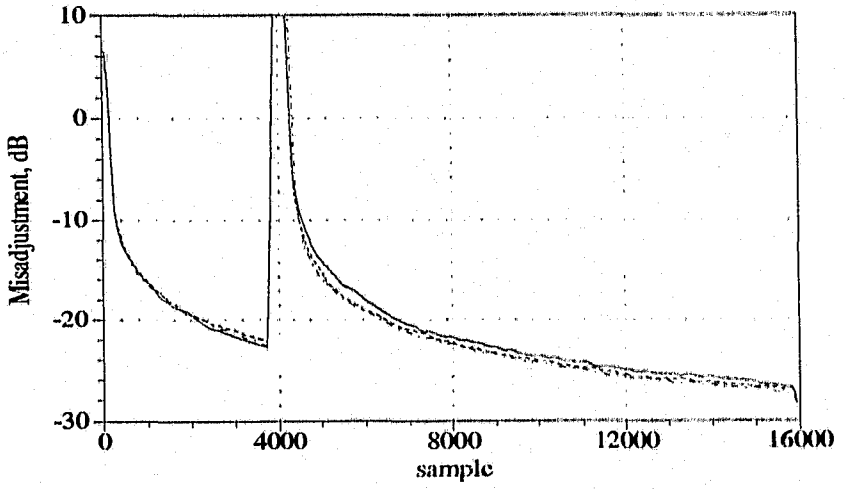

Figure 6.9: PA-RLS sensitivity to β_i.
——————— PA-RLS: $\beta_i = 1$
– – – – – – – – PA-RLS: $\beta_i = 5$
– · – · – · – · – PA-RLS: $\beta_i = 0.2$

tics, then a high value of β_i would be used. On the other hand, if the persistence of the system excitation were in question a more modest choice would be made. In Figure 6.9, the estimated learning curve for PA-RLS filters for various values of β_i are shown. The choice of $\beta_{\min} = 0.0015$ was made. While a slightly quicker reconvergence is available by the use of higher values of β_i, in all subsequent simulations this parameter will be taken to be unity.

The sensitivity of the PA-RLS algorithm to the parameter $\beta_{\min}$ is concerned with the "trapping" that occurs when the forgetting factor approaches unity. This effect is displayed in Figure 6.10. At some threshold $6 \times 10^{-7} < \beta_{\min}^{th} < 7 \times 10^{-7}$, this trapping behaviour occurs. Fortunately, the design criteria for this parameter is the desired steady-state misadjustment, which is unlikely to be in the range of -60dB. Taking a more realistic specification of -30dB, the value 0.0015 will be used for $\beta_{\min}$ in subsequent experiments.

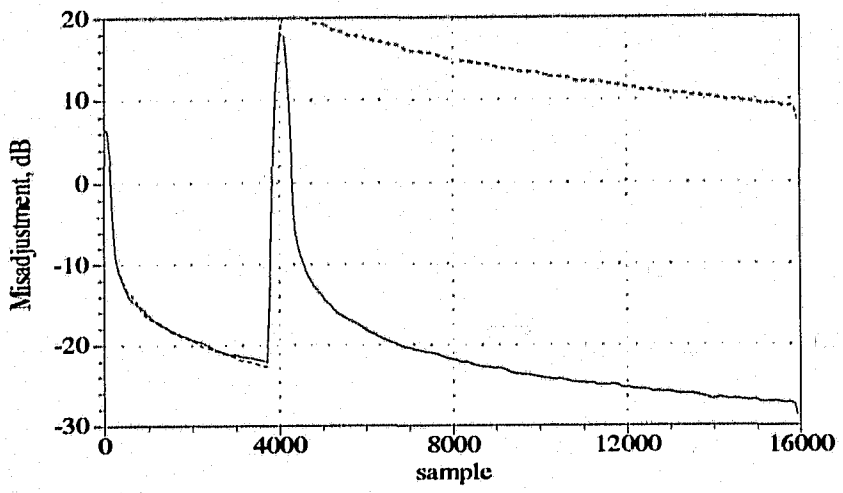

Figure 6.10: PA-RLS sensitivity to $\beta_{\min}$.
——————— PA-RLS: $\beta_{\min} = 6 \times 10^{-7}$
– – – – – – – – PA-RLS: $\beta_{\min} = 7 \times 10^{-7}$

Sensitivity to initialization

Having fixed the values of the PA-RLS parameters, the sensitivity of the algorithm under consideration to initialization was examined. In particular, three values were chosen for the initial estimate of the MMSE, $\hat{\sigma}^2_{\epsilon,0}$. Figure 6.11 demonstrates that the algorithm is acceptably insensitive to this choice

6.3.5 Discussion

We see that the sensitivity of the three proposed adaptive filters to their respective algorithmic parameters decreases in the order in which they have been treated. The RASE algorithm, though shown to perform exceptionally well when carefully tuned, is quite susceptible to poorly chosen parameters or initial environmental estimates. Since there are no suitable guidelines with which to choose these quantities in a general manner, this algorithm requires further attention before it may be recommended

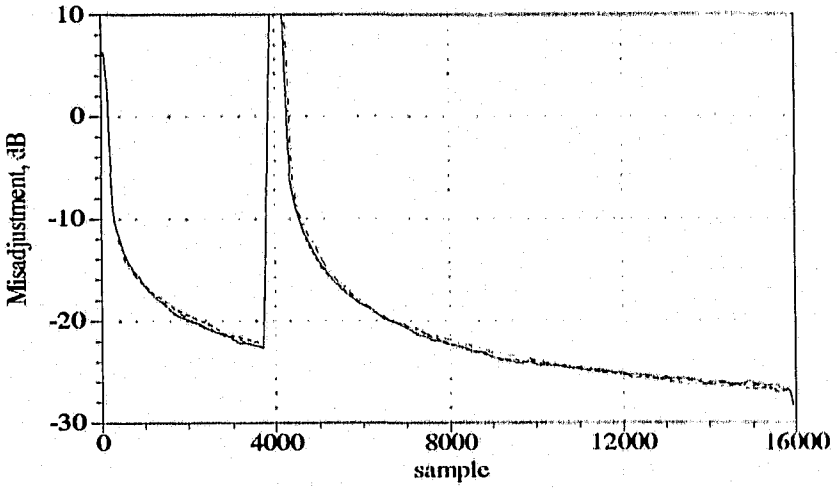

Figure 6.11: PA-RLS sensitivity to $\hat{\sigma}^2_{\epsilon,0}$.
——— PA-RLS: $\hat{\sigma}^2_{\epsilon,0} = \sigma^2_{\epsilon,0}$
– – – – – – – – PA-RLS: $\hat{\sigma}^2_{\epsilon,0} = 0$
– · – · – · – · – PA-RLS: $\hat{\sigma}^2_{\epsilon,0} = 100\sigma^2_{\epsilon,0}$

for use. The PA-NLMS scheme, though less sensitive to the choices made for its parameters, still suffers from a relatively large number of parameters. While the method for predicting PA-NLMS performance on the basis of the choice of parameters has been established, the more valuable converse design rule does not yet exist. That is, given a performance specification, it is a matter of trial and error to choose acceptable PA-NLMS parameters. Neither of these difficulties apply to the PA-RLS method, however. Indeed, it has been shown that this algorithm is particularly insensitive to either of its two parameters, and that the choice of these quantities can be made with little difficulty.

6.4 Performance Experiments

6.4.1 Introduction

In this section the performance of conventional adaptive filters and selected DA methods in nonstationary environments will be investigated. The principle difficulty with this endeavour is the inescapable artificiality of experiments to be performed. The approach throughout this work has been to try to examine the behaviour of adaptive filters in an application-independent manner. While this approach has been successful for the consideration of adaptive filters in stationary scenarios, its success in the presence of environmental nonstationarities is not so clear. There is no assurance, for example, that an algorithm that performs well in the presence of one kind of nonstationarity will deliver acceptable performance in another. Indeed, once the environment is permitted to behave in a nonstationary manner, there are simply too many variables to take into consideration. In effect, there are too many degrees of freedom in the specification of the statistics of the filter input, additive noise, and target filter weights. In order to fairly compare a number of algorithms in a nonstationary environment, care must be taken not to choose these statistics to suit a specific algorithm. Since the sensitivities of DA methods to various environmental statistics are unknown, the best strategy is to test the algorithms under consideration in a number of different situations. In the interest of simplicity, generality and fairness, the following experiments will involve easily described and simulated nonstationarities in spite of their artificiality. To be assured of acceptable adaptation performance in a practical nonstationary application, one must carefully model the nonstationary behaviour of a given application, and simulate this application by means of this model. Such work, which is considerable in itself, is beyond the scope of this dissertation.

The focus of the following simulations will be to investigate the performance of various DA methods in the presence of target filter weight nonstationarities. To this end, three models of target filter will be used. In the first case, the plant will change suddenly from the lowpass $\mathcal{F}_{N,1}^{+}$ to the allpass $\mathcal{F}_{N,\infty}^{+}$ as in the previous simulations. Of the three plant models to be simulated, only this one represents a practical situation in some applications. Secondly, the random walk plant model will be considered. The initial plant specification will be given by $\mathcal{F}_{N,1}^{+}$. The random walk simulations will be primarily for the purpose of testing the predictive power of the analysis in the previous chapters. Finally, a Markovian nonstationarity will be simulated in order to examine the performance of the PA methods in a continuously varying scenario that is sufficiently distinct from the random walk on which those algorithms are based. Under this nonstationarity, the plant will be modeled as the parallel combination of a fixed $\mathcal{F}_{N,1}^{+}$, and a time-varying FIR filter whose weights are modeled as the response of an N-input initially-relaxed one-pole Markov filter to N independent white zero-mean Gaussian sequences. If the pole of the Markov filter were at unity and the variances of its inputs were identical, a random walk would result. In order to provide a simulation that is sufficiently different from a random walk, however, we take the Markov pole to be at 0.9, and its inputs to have distinct variances. In particular, let us consider the case in which the Markov filter inputs have variances in proportion to the square of the corresponding nominal plant filter weight. In effect, each plant filter weight is updated at each sample according to

$$w_{i,k+1} = w_i^{(f)} + w_{i,k+1}^{(m)}$$
$$w_{i,k+1}^{(m)} = 0.9 w_{i,k}^{(m)} + w_i^{(f)} z_i$$

where the weights $w_i^{(f)}$ are the coefficients of the fixed filter, $\mathcal{F}_{N,1}$, and the noise z_i is zero-mean, temporally and spatially white, and Gaussian having variance σ_z^2. By

analogy to the random walk nonstationarity-to-noise ratio, the corresponding quantity for this Markov nonstationarity is $q^{(m)} = \sigma_z^2\sigma_x^2/\sigma_\epsilon^2$.

6.4.2 Discontinuous Target Filter Simulations

In this section, we examine the performance of a number of doubly adaptive filters in the presence of a sudden discontinuity in the target filter weights. We will present the results of these experiments comparing six LMS-based DA algorithms with NLMS as a benchmark. The results of the same experiment applied to RLS-based methods will then be given. For the LMS-based algorithms, simulations will be performed for three input-colouring filters, namely, the lowpass $\mathcal{F}_{5,1}^{+}$, the allpass $\mathcal{F}_{5,\infty}^{+}$, and the highpass $\mathcal{F}_{5,1}^{-}$ in order to observe the dependence of the various DA methods on the input correlatedness. This is not necessary for the RLS-based methods since the RLS algorithm is robust toward this condition. Moreover, in the interest of fairness, the parameters of these LMS-based algorithms will assume a constant value over the three different inputs – being "tuned" for best performance when the input is uncorrelated. Each of the simulated learning curves in this section represent averages over 200 independent trials.

In Figures 6.12, 6.13, and 6.14, estimates of PA-NLMS and SKYW learning curves are displayed for the allpass, lowpass and highpass input colouring filter respectively. These are compared with a benchmark NLMS providing comparable MAC performance in each case. These two DA methods have similar computational complexity, and each perform better than their parent filter for an uncorrelated input signal. For correlated inputs, however, little or no advantage is offered by the SKYW scheme. For the PA-NLMS algorithm, the parameters were set as follows: $\lambda_\epsilon = 0.999$, $\lambda_z = 0.99$, $\lambda_\mu = 0.98$, $\alpha = 0.125$, and $\zeta = 0.005$. The SKYW parameter, λ_{SKYW}, on the other hand was given the value 0.99. The benchmark NLMS filter used $\bar{\mu} = 0.077$.

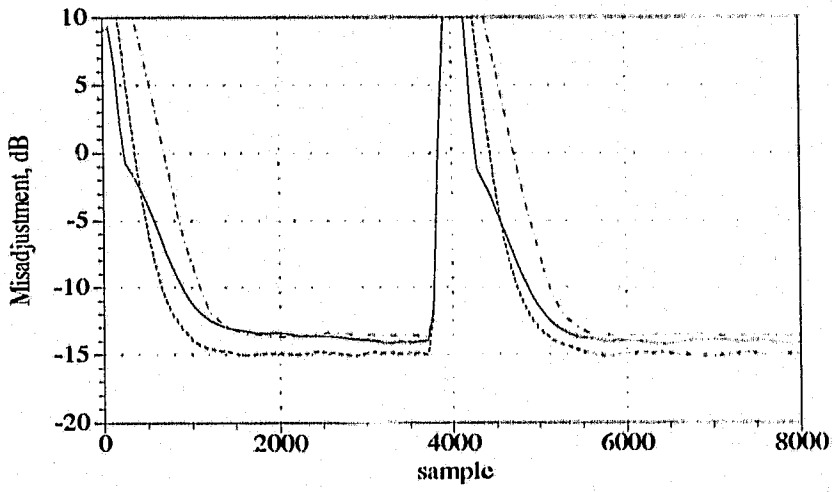

Figure 6.12: PA-NLMS and SKYW performance – uncorrelated input.

PA-NLMS
SKYW
NLMS

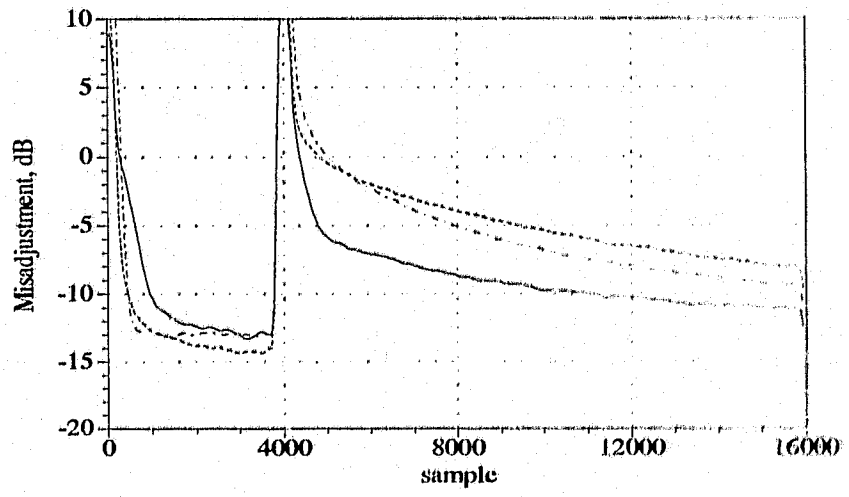

Figure 6.13: PA and SKYW performance – correlated input I.

PA-NLMS
SKYW
NLMS

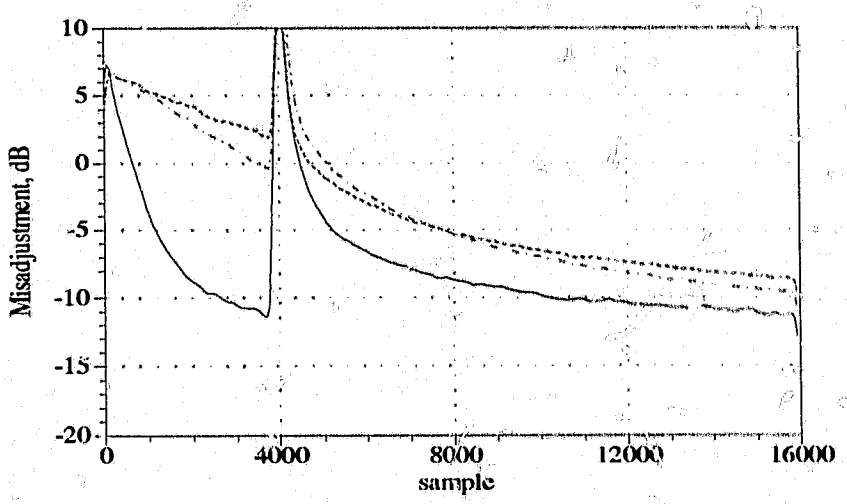

Figure 6.14: PA and SKYW performance – correlated input II.

PA-NLMS
SKYW
NLMS

Estimates of the learning curves for the RASE and KJ algorithms for the same three experiments are shown in Figures 6.15, 6.16, and 6.17. The RASE parameters were as follows: $\lambda_\mu = 0.9$, $q_{11} = 10^{-9}$, $q_{22} = 0$, $q_{33} = 10^{-6}$, and $\tilde{\mu}_{\min} = 0.0198$. The MMSE was assumed to be known for both of these algorithms. While the quality of RASE performance depends on this knowledge, the KJ algorithm requires this knowledge in order to choose its parameters. In this case, referring to [34], $\alpha_{KJ} = 0.97$ and $\gamma_{KJ} = 0.001$. The performance of the KJ algorithm is comparable to that of RASE with suboptimal initialization (cf. Figures 6.5 and 6.6) for uncorrelated inputs. In the case of correlated inputs, however, the advantage of the KJ algorithm is not nearly as significant. The convergence-controlling parameter of the benchmark NLMS filter assumed the value $\tilde{\mu} = 0.0198$ in this case.

The final two LMS-based DA methods that were tested in this scenario are the GAS and HCB algorithms, and the results of those tests are given in Figure 6.18,

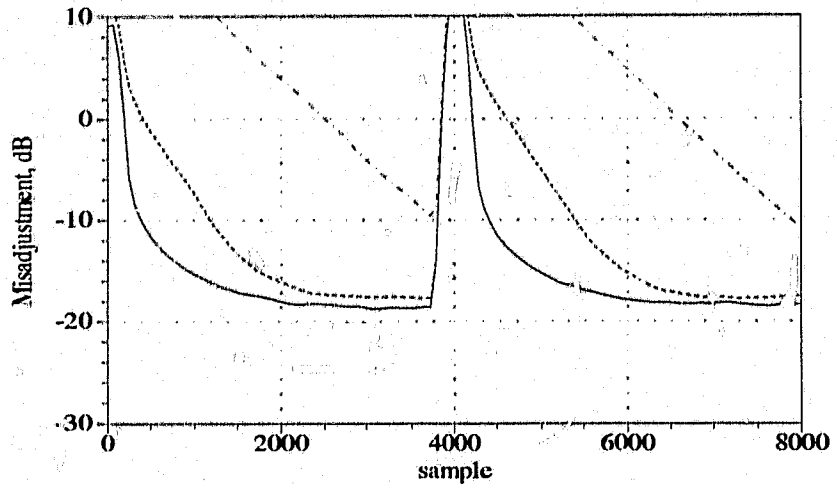

Figure 6.15: RASE and KJ performance – uncorrelated input.

RASE
KJ
NLMS

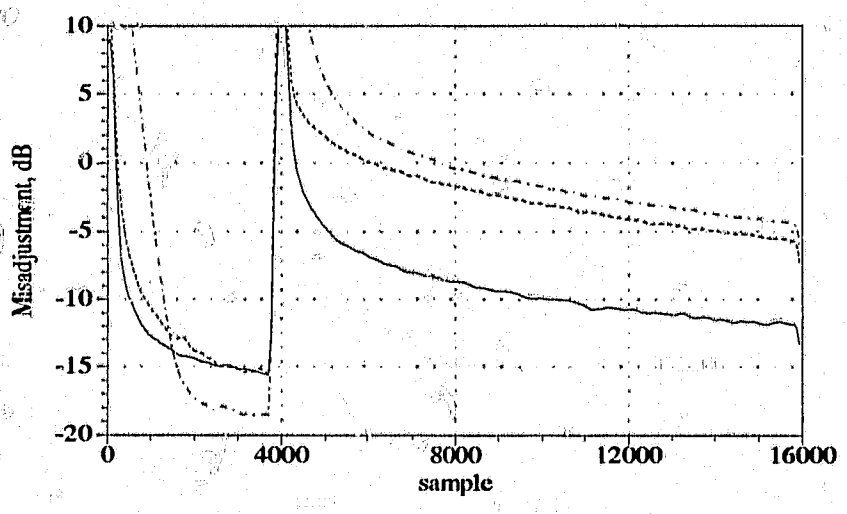

Figure 6.16: RASE and KJ performance – correlated input I.

RASE
KJ
NLMS

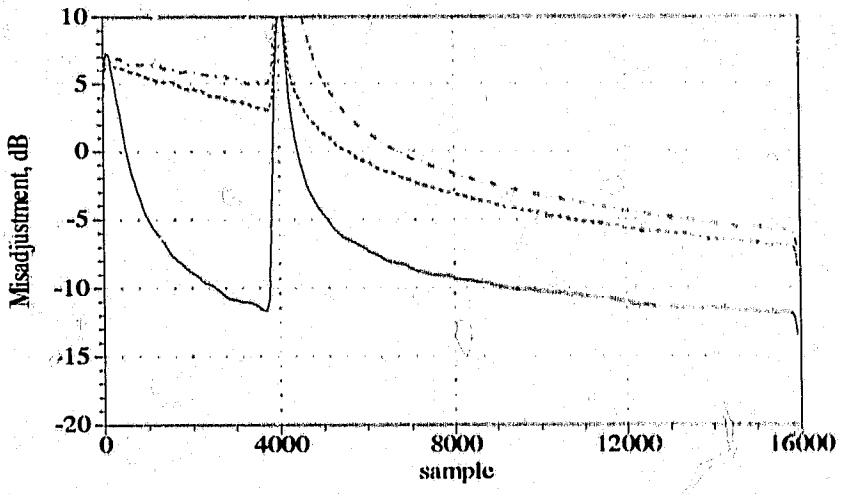

Figure 6.17: RASE and KJ performance - correlated input II.

RASE
KJ
NLMS

6.19 and 6.20. These methods are based on an LMS rather than an NLMS parent. Again, an estimated learning curve for an NLMS filter, with $\hat{\mu} = 0.182$ for the present experiment, is provided. In this case, however, it is unfair to suggest that this benchmark process provides comparable MAC performance to both DA algorithms. While it does approximate the MAC of the HCB filter, the GAS scheme provides much lower MAC. Unfortunately, it takes a considerable number of samples to achieve this state. In consequence, the comparison between GAS performance and that of other algorithms is qualitative at best. It is clear, however, that there is little or no advantage in the use of these two DA methods over the conventional algorithm for nonstationary environments and correlated input samples. The parameters of the DA algorithms for these simulations were chosen in keeping with the original papers. Although no guidelines were provided, representative choices of the algorithmic parameters were made for simulation purposes [33] [22]. For both algorithms, μ_{max} was taken to be

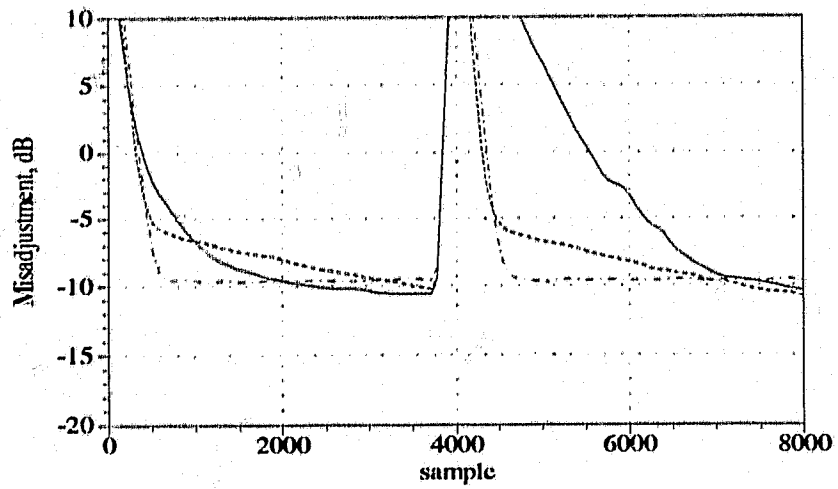

Figure 6.18: HCB and GAS performance – uncorrelated input.

HCB
GAS
NLMS

0.025, which is, for stability reasons, somewhat less than the LMS stability limit of $\frac{2}{3}(\mathrm{tr}\,\mathbf{R})^{-1}$ [32]. For the GAS scheme, $\rho_{\mathrm{GAS}} = \mu_{\max}/10$ was chosen. The HCB parameters were assigned the values $m_0 = m_1 = 5$, $\alpha_{\mathrm{HCB}} = 2$ and $\mu_{\min} = \mu_{\max}/1024$.

Finally, Figure 6.21 presents the resulting learning curve estimates of the RLS-based algorithms in the same scenario. The two DA methods and the benchmark RLS process were tuned to provide -30 dB MAC. This was accomplished by using a forgetting factor of $\lambda = 0.9999$ for the RLS filter, and choosing $\beta_0 = 0.002$ for the FKY scheme. One and a quarter dB discrepancy between actual and nominal PA-RLS MAC was assumed in order to choose $\beta_{\min} = 0.0015$. Further, β_i and $\beta_{\max}$ were both taken to be unity for the PA-RLS and FKY algorithms respectively.

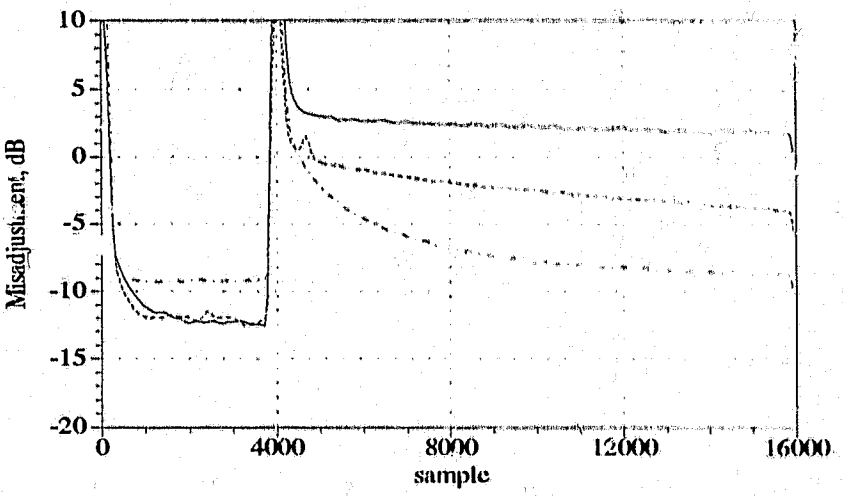

Figure 6.19: HCB and GAS performance – correlated input I.

HCB
GAS
NLMS

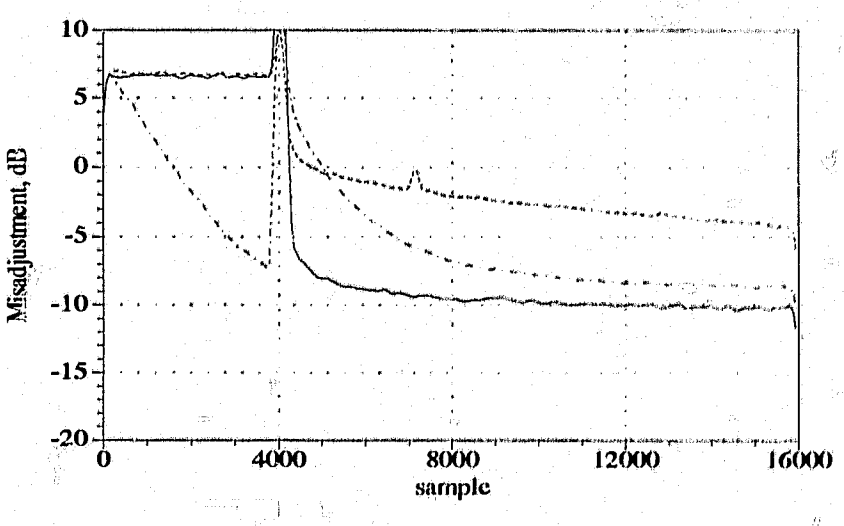

Figure 6.20: HCB and GAS performance – correlated input II.

HCB
GAS
NLMS

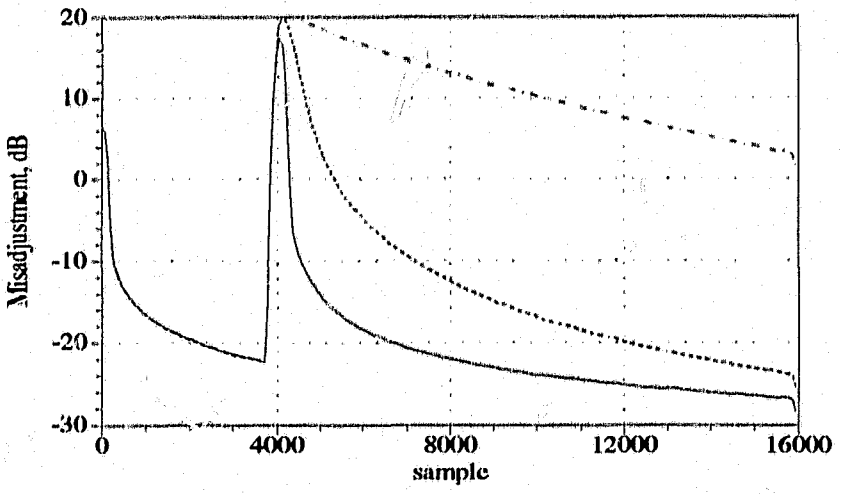

Figure 6.21: RLS-based DA filter performance.
PA-RLS
FKY
RLS

6.4.3 Random Walk Simulations

We now consider a simulated random walk scenario. The simulations from which the following results were obtained were performed for two reasons. First, it is of interest to determine how accurate the predictive formulas for conventional algorithm performance in this environment are. Second, it is of interest to compare the performance of a number of DA methods in an easily described continuously varying situation. We would expect, however, that the PA methods would be capable of performing very well in these experiments. In that respect, these simulations may seem unfairly weighted toward those methods. As we shall see, the performance of PA-NLMS is comparable to that of other LMS-based DA algorithms.

In testing the accuracy of the expressions that predict the steady-state misadjustment for the conventional adaptation algorithms in a random walk scenario, it is of

interest to consider how low N must be before the modified independence assumption that was necessary to derive these expressions becomes invalid. In consequence, random walk experiments were performed for filters of length $N = 5$, $N = 21$, and $N = 101$. In Figures 6.22 through 6.33, estimates of NLMS and RLS steady-state misadjustment in a simulated random walk scenario are displayed beside the respective theoretical curves. Further, the effects of input correlatedness are illustrated in the case that it is most significant: high NNR and low τ. The misadjustment measurements under the highpass input-colouring filter $\mathcal{C}_c$ are given as $*$'s in the appropriate figures. The quantity β_{crit} was taken to have the value two for the theoretical misadjustment given by (5.6). The normalized convergence time is simply the ratio of convergence time constant to filter length. In Tables 6.2, 6.3, and 6.4, the misadjustment results for the DA methods are provided. The experimentally achieved minimum misadjustments for the conventional algorithms after careful tuning are also given in these tables. All measured values represent averages over 10^5 steady-state samples. For the PA-NLMS algorithm, numerically calculated predictions are also provided as theoretical values. These data are not plotted on the respective figures because of the difficulty with which convergence time is determined for doubly adaptive methods.

6.4.4 Markov Simulations

In this section, the steady-state misadjustment measurements for the algorithms under consideration in the presence of the previously described Markov nonstationarity are presented. For all of the DA methods, algorithmic parameters are kept at the same values as in the previous performance experiments. For the conventional adaptive filters, however, numerous values of convergence-controlling parameter were attempted to achieve an estimate of the best-possible performance in this scenario. The experiment was performed for both $N = 5$ and $N = 21$, and the results are

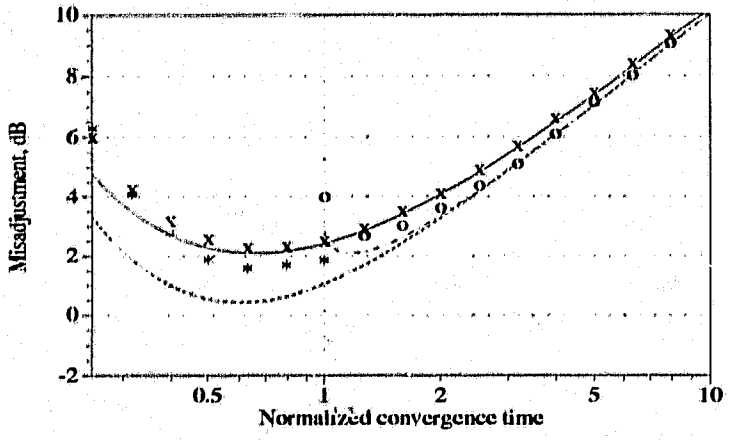

Figure 6.22: Random walk MAC performance: N=5; NNR=0 dB.

theoretical RLS: (5.6)
theoretical RLS: (5.7)
theoretical NLMS: (3.4)
x x x RLS;
o o o NLMS;
* * * RLS ($\mathcal{C} = \mathcal{C}_c$);

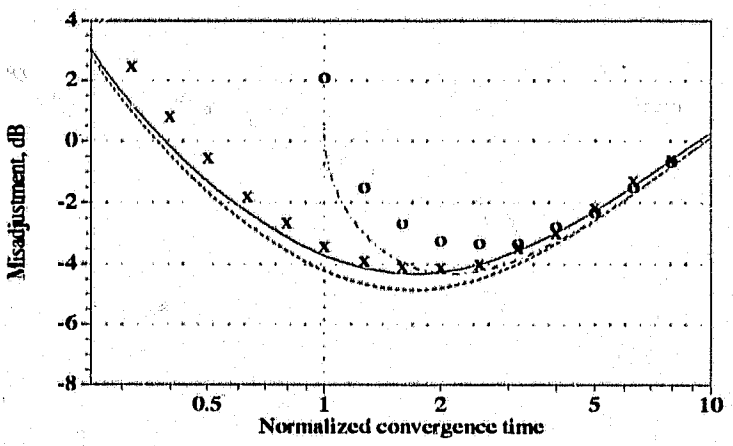

Figure 6.23: Random walk MAC performance: N=5; NNR=−10 dB.

theoretical RLS: (5.6)
theoretical RLS: (5.7)
theoretical NLMS: (3.4)
x x x RLS;
o o o NLMS;

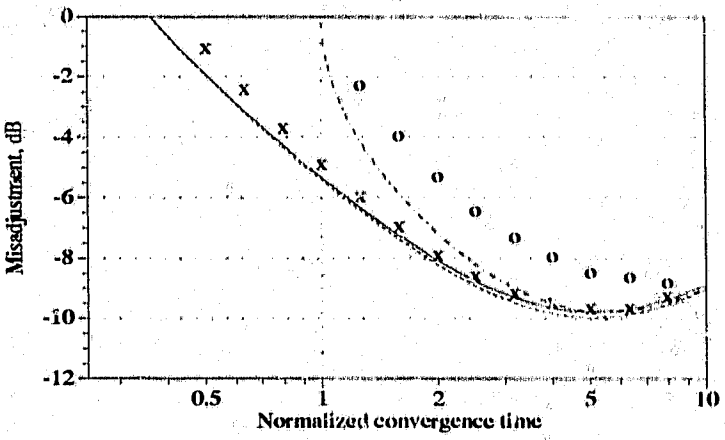

Figure 6.24: Random walk MAC performance: N=5; NNR=−20 dB.

theoretical RLS: (5.6)
theoretical RLS: (5.7)
theoretical NLMS: (3.4)
x x x RLS;
o o o NLMS;

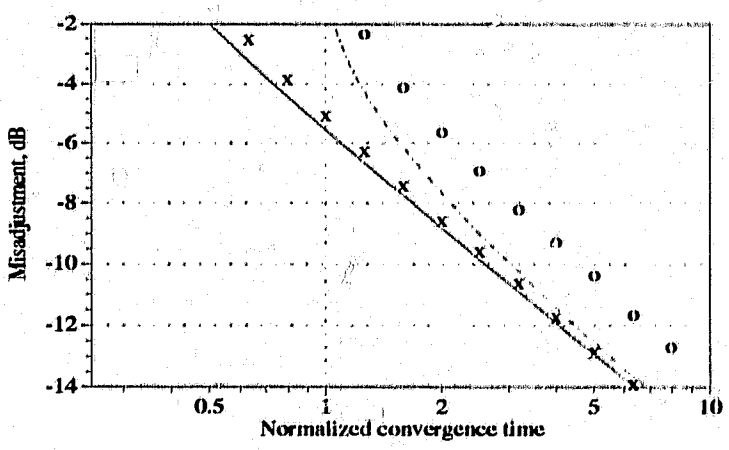

Figure 6.25: Random walk MAC performance: N=5; NNR=−100 dB.

theoretical RLS: (5.6)
theoretical RLS: (5.7)
theoretical NLMS: (3.4)
x x x RLS;
o o o NLMS;

Table 6.2: Random walk MAC estimates (dB): N=5.

Algorithm	NNR=0	NNR=−10	NNR=−20	NNR=−100
NLMS (tuned)	2.65	−3.36	−8.84	< -30.00
PA-NLMS (ex.)	2.70	−2.69	−7.66	−13.57
PA-NLMS (th.)	2.09	−4.31	−9.76	−14.65
SKYW	5.76	−2.50	−8.74	−13.36
RASE	2.71	−3.19	−8.16	−17.57
KJ	11.31	1.74	−7.61	−17.91
GAS	3.11	−3.95	−9.18	−13.44
HCB	4.96	−0.98	−5.83	−10.76
RLS (tuned)	2.20	−4.18	−9.72	< -30.00
FA-RLS	3.97	−0.64	−4.15	−28.99
FKY	12.00	6.33	0.34	−29.40

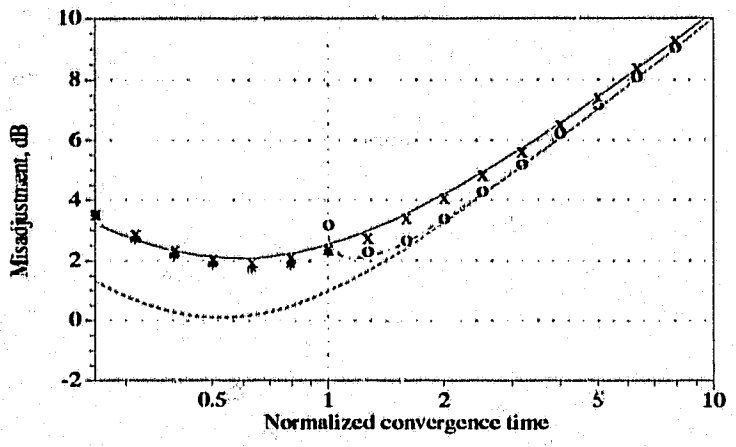

Figure 6.26: Random walk MAC performance: N=21; NNR=0 dB.

theoretical RLS: (5.6)
theoretical RLS: (5.7)
theoretical NLMS: (3.4)
RLS;
NLMS;
RLS ($C = C_c$);

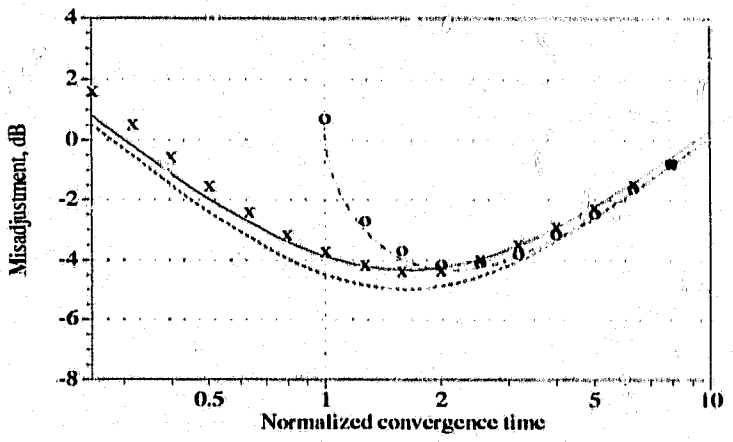

Figure 6.27: Random walk MAC performance: N=21; NNR=−10 dB.

theoretical RLS: (5.6)
theoretical RLS: (5.7)
theoretical NLMS: (3.4)
x x x RLS;
o o o NLMS;

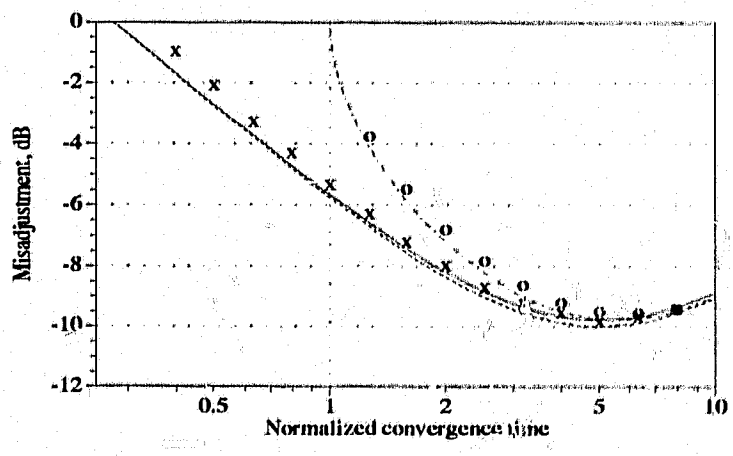

Figure 6.28: Random walk MAC performance: N=21; NNR=−20 dB.

theoretical RLS: (5.6)
theoretical RLS: (5.7)
theoretical NLMS: (3.4)
x x x RLS;
o o o NLMS;

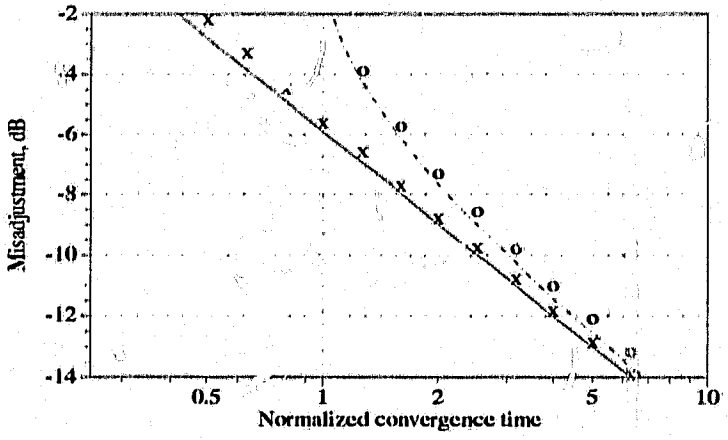

Figure 6.29: Random walk MAC performance: N=21; NNR=−100 dB.

———	theoretical RLS: (5.6)
– – – – – – – –	theoretical RLS: (5.7)
– · – · – · – ·	theoretical NLMS: (3.4)
x x x	RLS;
o o o	NLMS;

Table 6.3: Random walk MAC estimates (dB): N=21.

Algorithm	NNR=0	NNR=−10	NNR=−20	NNR=−100
NLMS (tuned)	2.26	−4.18	−9.57	< −30.00
PA-NLMS (ex.)	2.28	−3.99	−9.07	−13.96
PA-NLMS (th.)	2.09	−4.31	−9.76	−14.65
SKYW	6.26	−2.18	−9.35	−15.12
RASE	2.24	−3.86	−8.83	−18.86
KJ	9.09	1.70	−7.71	−17.75
GAS	2.54	−3.94	−8.32	−12.16
HCB	8.08	0.60	−5.44	−9.70
RLS (tuned)	1.39	−4.43	−9.88	< −30.00
PA-RLS	2.65	−3.49	−6.92	−26.23
FKY	11.93	6.51	0.32	−26.33

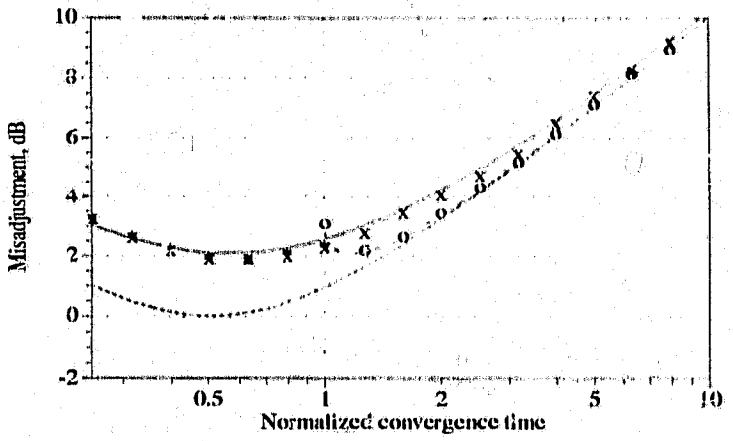

Figure 6.30: Random walk MAC performance: N=101; NNR=0 dB.

theoretical RLS: (5.6)
theoretical RLS: (5.7)
theoretical NLMS: (3.4)
RLS;
NLMS;
RLS ($C = C_0$);

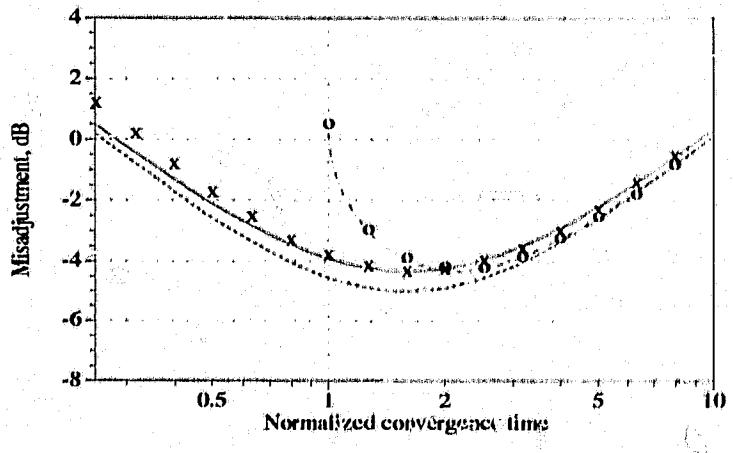

Figure 6.31: Random walk MAC performance: N=101; NNR=−10 dB.

theoretical RLS: (5.6)
theoretical RLS: (5.7)
theoretical NLMS: (3.4)
RLS;
NLMS;

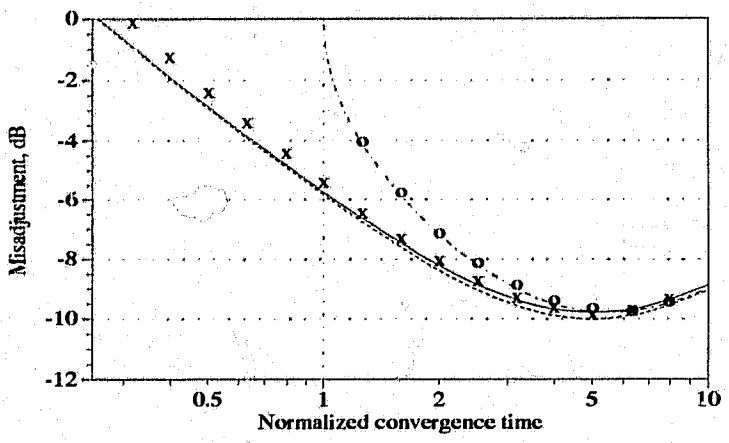

Figure 6.32: Random walk MAC performance: N=101; NNR=−20 dB.

theoretical RLS: (5.6)
theoretical RLS: (5.7)
theoretical NLMS: (3.4)
RLS;
NLMS;

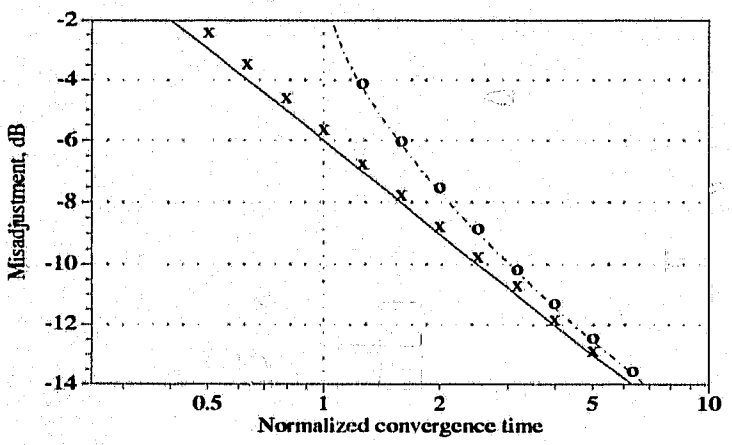

Figure 6.33: Random walk MAC performance: N=101; NNR=−100 dB.

theoretical RLS: (5.6)
theoretical RLS: (5.7)
theoretical NLMS: (3.4)
RLS;
NLMS;

Table 6.4: Random walk MAC estimates (dB): N=101.

Algorithm	NNR=0	NNR=−10	NNR=−20	NNR=−100
NLMS (tuned)	2.16	−4.23	−9.76	< -30.00
PA-NLMS (ex.)	2.43	−4.12	−9.40	−12.89
PA-NLMS (th.)	2.09	−4.31	−9.76	−14.65
SKYW	7.79	−1.24	−9.26	−15.23
RASE	2.14	−4.04	−9.17	−17.39
KJ	5.56	−0.71	−8.35	−16.78
GAS	2.90	−3.41	−7.05	−8.72
HCB	8.93	3.69	−5.45	−7.40
RLS (tuned)	1.84	−4.38	−9.91	< -30.00
PA-RLS	2.43	−4.15	−8.89	−20.37
FKY	12.02	5.73	−3.89	−21.19

tabulated in Tables 6.5 and 6.6, respectively. Measurements represent averages over 10^5 steady-state samples.

6.4.5 Discussion

It is clearly seen via the results from the discontinuous target filter simulations that the sensitivity of the LMS-based DA methods is generally considerably greater than that of the parent algorithm. The exceptions to this rule seem to be the PA-NLMS and RASE algorithms. At the other extreme is the HCB scheme. In Figures 6.19 and 6.20, this algorithm demonstrates alternately much better and much worse convergence in comparison to an NLMS filter. The remarkable reconvergence in the latter figure is due to the fact that only one of the components of the weight error vector is significant in this case. As a result, the μ^i corresponding to that component is successfully increased independent of the others. Unfortunately, this scenario is no more practical

Table 6.5: Markov nonstationarity MAC estimates (dB): N=5.

Algorithm	$q^{(m)} = 1$	$q^{(m)} = 10^{-1}$	$q^{(m)} = 10^{-2}$	$q^{(m)} = 10^{-10}$
NLMS (tuned)	6.32	−2.80	−12.85	< -30.00
PA-NLMS	6.35	−2.44	−10.52	−13.77
SKYW	6.99	−2.52	−9.84	−13.10
RASE	6.76	−1.32	−10.85	−17.57
KJ	7.06	−2.87	−11.72	−17.92
GAS	6.74	−1.97	−9.77	−13.17
HCB	6.58	−2.15	−8.49	−11.15
RLS (tuned)	6.23	−2.82	−12.77	< -30.00
PA-RLS	6.99	−2.82	−12.63	−27.98
FKY	7.16	−2.77	−12.74	−28.96

Table 6.6: Markov nonstationarity MAC estimates (dB): N=21.

Algorithm	$q^{(m)} = 1$	$q^{(m)} = 10^{-1}$	$q^{(m)} = 10^{-2}$	$q^{(m)} = 10^{-10}$
NLMS (tuned)	7.24	−2.79	−12.84	< -30.00
PA-NLMS	7.48	−2.09	−10.19	−14.05
SKYW	7.31	−2.48	−10.73	−14.94
RASE	8.93	−0.75	−11.03	−18.33
KJ	7.24	−2.65	−11.55	−17.92
GAS	8.12	−0.82	−8.41	−11.42
HCB	7.72	−1.58	−7.70	−9.78
RLS (tuned)	7.25	−2.76	−12.63	< -30.00
PA-RLS	7.20	−2.82	−12.63	−27.98
FKY	7.23	−2.75	−12.48	−24.27

than that of the former figure, in which HCB reconvergence is much worse than that of NLMS. It does, however, serve to demonstrate that the choice of simulation can make a good deal of difference when testing a doubly adaptive filter.

Of the RLS-based methods, the performance of PA-RLS is markedly superior to both the conventional algorithm and the FKY scheme. As shown in Figure 6.21, the reconvergence of the PA-RLS method is almost as good as its first convergence apart from a small delay. Since the PA-RLS first convergence is only slightly poorer than that of the RLS filter, which represents the optimal estimator over the first section of this experiment, the PA-RLS behaviour is similar to that of an RLS filter with re-initialization when a discontinuity in the squared error is observed. In contrast to such a filter, however, the PA-RLS method is capable of tracking slowly varying nonstationarities.

The random walk simulations demonstrate a good match between the analytic steady-state misadjustment values and the simulated counterparts for moderate to large values of N and τ. The discrepancy between theory and measurements for small τ for the RLS algorithm is due to the matching that was instilled between (5.6) and (5.7) for $q = 0$. If some other expression, perhaps making use of fourth-order statistics, were to provide improvement in this area, the expression (5.6) could be modified accordingly. The discrepancy between nominal and measured steady-state misadjustment for the RLS-based algorithms in the NNR$=-100$ dB case is due primarily to the difficulty with which this measurement is made: the nominal steady-state misadjustment is only attained after a considerable number of samples. Further, this convergence time is proportional to the length of the adaptive filter. As a result, this difficulty is exacerbated for the $N = 101$ case, as is evident from the tables.

It is interesting to note from Figures 6.22, 6.26, and 6.30 the sensitivity of the

RLS lag misadjustment on the input correlatedness. In keeping with the results of the first experiment of the present chapter, these effects are greater for smaller values of N. This suggests, perhaps, that the common use of the input covariance eigenvalue ratio gives a poor indication of the effects of input correlatedness and that the ratio of normalized autocorrelation power to filter length might be more informative.

The modified Markov experiments highlight once again the fact that a given algorithm may perform well in one nonstationarity while not in another. Indeed, the KJ and FKY methods, which performed quite poorly in a random walk scenario, demonstrate excellent tracking capabilities under the Markov regime. Note that it is difficult to tune the conventional methods in order to determine their optimal performance in a near-stationary scenario, and consequently, the corresponding entries in the tables reflect this.

Chapter 7

Conclusions

7.1 Contribution Summary

In Chapter 2, existing doubly adaptive filters were catalogued. It was shown that the most problematic aspect of most of these algorithms is the requirement to have a good estimate of an environmental quantity, the additive noise power. In the case that such an estimate was unnecessary, other difficulties with the algorithm were shown to make it unsuitable for nonstationary environments or adaptation with correlated input samples.

In Chapter 3, the nonstationary aspects of the adaptation environment were set aside in order to examine the dependence of the NLMS adaptive filter to its operating point. The attempt was made to make use of estimates of the additive noise power and the instantaneous misadjustment in order to optimize the convergence behaviour of the DA filter. As was demonstrated in the previous chapter, this approach is quite successful when the initial estimate of the additive noise power is quite accurate and the filter input samples is not particularly correlated. Unfortunately, the success of the RASE algorithm is quite sensitive to these two conditions.

In Chapter 4, the principle of parallel adaptation for the continuous estimation of environmental quantities was introduced. To facilitate the development of the

resulting PA-NLMS algorithm, the concept of adaptation state was set aside. As a result, the NLMS dependence on environmental nonstationarity was examined. The detailed performance analysis and subsequent simulations both demonstrated that the PA-NLMS method is quite successful even when no initial estimate of the environment exists. Unfortunately, the application of this algorithm does depend on the relative stationarity of the additive noise to the target filter weights. Further, the details of PA-NLMS parameter selection are not altogether obvious. However, PA-NLMS performance seems to be largely insensitive to the choices made for its parameter values.

In Chapter 5, the parallel adaptation principle was extended to RLS adaptive filters. In the context of RLS filters, the problems inherent in PA-NLMS seem to vanish. Since the convergence of an RLS filter with unity forgetting factor is optimal, the RLS operating point is largely irrelevant. In consequence, the application of an environmental model that consists solely of the additive noise power and nonstationarity-to-noise ratio is particularly effective. As demonstrated in the simulations of Chapter 6, the PA-RLS algorithm provides near-optimal adaptation in a discontinuous-target-weight environment even though it is built upon a random walk assumption. Further, the ability of this algorithm to handle sudden changes in additive noise statistics recommend it for applications in which the additive noise is nonstationary.

7.2 Recommendations for Further Study

Perhaps it is common for dissertations such as this one to attempt to say the last word on the topic at hand. In this case, however, no such comprehension is even attempted. In fact, the present work seems to open up a great number of opportunities for

further study. Once it is permitted that a number of adaptive filters can be operating simultaneously on the same inputs, the possibilities seem limitless.

In general, there are two basic approaches that can be made. First, the simplified environmental/adaptation state objective functions used in Chapters 3, 4, and 5 could be extended. For example, an amalgam of the principles set forth in Chapters 3 and 4 seems possible. Using a number of fixed-convergence-controlling-parameter indicator processes, one should be able to estimate both the adaptation state of the primary adaptive filter and a number of necessary environmental quantities. Other, more involved, environmental models could also be considered. Unfortunately, the greater the complexity of the model to be used, the more difficult it is to estimate the necessary quantities. There are likely to be application-specific environmental models that are more complex than those considered here while still permitting tractable estimation and optimization.

Second, the conversion of existing "adaptive" Kalman algorithms to be applicable in the adaptive filtering problem will result in considerable simplification. The resulting algorithms, while likely more complex than other algorithms, will explicitly involve a more sophisticated environmental model.

Moreover, the extension of the present work to lattice filters should be reasonably straightforward. Indeed, the principles developed herein could potentially offer improved adaptation for a number of different filter structures.

Finally, the construction of a so-called fast, or efficient implementation, or even approximate implementation, of the variable-forgetting-factor FA-RLS algorithm would be valuable. The effects of finite-precision arithmetic on this, and all of the current DA algorithms would also be of interest.

References

[1] A. Cauchy, "Méthode générale pour la résolution des systèmes d'équations simulatanées," *Compt. Rend. Acad. Sci.*, vol. 25, pp.536–538, 1847.

[2] R. Fletcher, *Practical Methods of Optimization*, 2nd Ed., Wiley, New York, 1987.

[3] B. Widrow and M. E. Hoff, Jr., "Adaptive switching circuits," *1960 Wescon Conf. Rec., Pt 4*, pp. 96–104, 1960.

[4] R. E. Kalman, "A new approach to linear and prediction problems," *Trans. ASME J. Basic Eng.*, vol. 82, pp. 35–45, 1960.

[5] M. Woodbury, *Inverting Modified Matrices*, Mem. Rep. 42, Statistical Research Group, Princeton University, Princeton, N.J., 1950.

[6] S. Haykin, *Adaptive Filter Theory.*, 2nd Ed., Prentice-Hall, Englewood-Cliffs, New Jersey, 1986.

[7] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Prentice-Hall, Englewood-Cliffs, New Jersey, 1985.

[8] M. L. Honig and D. Messerschmitt, *Adaptive Filters: Structures, Algorithms and Applications.* Kluwer, Boston, 1984.

[9] S. T. Alexander, *Adaptive Signal Processing: Theory and Applications.* Springer-Verlag, New York, 1986.

[10] A. E. Albert and L. S. Gardner, Jr., *Stochastic Approximation and Nonlinear Regression.* MIT Press, Cambridge, Mass., 1967

[11] J. I. Nagumo and A. Noda, "A learning method for system identification," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 282-287, 1967.

[12] A. Weiss and D. Mitra, "Digital adaptive filters: conditions for convergence, rates of convergence, effects of noise and errors arising from the implementation," *IEEE Trans. Inf. Theory*, vol. IT-25, pp. 637–652, 1979.

[13] G. C. Goodwin and S. K. Sin, *Adaptive Filtering Prediction and Control.* Prentice-Hall, Englewood-Cliffs, New Jersey, 1984.

[14] N. Wiener and E. Hopf, "On a class of singular integral equations," *Proc. Prussian Acad. Math-Phys. Ser.* p. 696, 1931.

[15] B. Widrow, J. M McCool, M. G. Larimore, and C. R. Johnson Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filter," *Proc. of IEEE*, vol. 64, pp. 1151–1162, Aug. 1976.

[16] E. Eleftheriou and D. D. Falconer, "Tracking properties and steady-state performance of RLS adaptive filter algorithms," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-34, pp. 1097–1110, Oct. 1986.

[17] L. Ljung and S. Gunnarsson, "Adaptation and tracking in system identification — a survey," *Automatica*, vol. 26, pp. 7–21, Jan. 1990.

[18] W. B. Mikhael, F. H. Wu, L. G. Kazovsky, G. S. Kang, and L. J. Fransen, "Adaptive Filters with Individual Adaptation of Parameters," *IEEE Trans. Circuits and Systems*, vol. CAS-33, pp. 677–686, July 1986.

[19] P. S. R. Diniz and L. W. P. Biscainho, "Optimal variable step size for the LMS/Newton algorithm with application to subband adaptive filtering," *IEEE Trans. Signal Processing*, vol. 40, pp. 2825–2829, Nov. 1992.

[20] D. T. Slock, "On the convergence behavior of the LMS and NLMS algorithms," *Proc. 5th European Signal Processing Conference*, Barcelona, Spain, pp. 197–200, Sept. 1990.

[21] C.-L. Wang and R.-Y. Chen, "Optimum design of the LMS algorithm using two step sizes for adaptive FIR filtering," *Signal Processing*, vol. 26, pp. 197–204, 1992.

[22] R. W. Harris, D. M. Chabries and F. A. Bishop, "A variable step (VS) adaptive filter algorithm," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-34, pp. 309–316, April 1986.

[23] H. Kesten, "Accelerated stochastic approximation," *Ann. Math. Stat.*, vol. 29, pp. 41–59, Mar. 1958

[24] J. B. Evans and B. Liu, "Variable step size methods for the LMS adaptive algorithm," in *Proc. International Symposium of Circuits and Systems*, Philadelphia, USA, pp. 422–425, 1987.

[25] J. B. Evans, "A new variable step size method suitable for efficient VLSI implementation," in *Proc. International Conference of Acoustics, Speech, and Signal Processing*, Toronto, Canada, pp. 2105–2108, 1991.

[26] C. P. Kwong, "Dual sign algorithm for adaptive filtering," *IEEE Trans. on Commun.*, vol. COM-34, pp. 1272–1275, Dec. 1986.

[27] T. J. Shan and T. Kailath, "Adaptive algorithms with an automatic gain control feature," *IEEE Trans. Circuits and Systems*, vol. 35, pp. 122–127, Jan. 1988.

[28] S. Karni and G. Zeng, "Comments on 'adaptive algorithms with an automatic gain control feature'," *IEEE Trans. Circuits and Systems*, vol. 37, pp. 974–975, July 1990.

[29] T. J. Shan and T. Kailath, "Comments on 'comments on "adaptive algorithms with an automatic gain control feature" '," *IEEE Trans. Circuits and Systems*, vol. 37, p. 975, July 1990.

[30] H. Ye and B.-X. Wu, "A robust adaptive filtering algorithm with variable step size," *Proc. of IEEE MILCOM-89*, Boston, MA, pp. 28.8.1–28.8.5, 1989.

[31] S. Karni and G. Zeng, "A new convergence factor for adaptive filters," *IEEE Trans. Circuits and Systems*, vol. 36, pp. 1011–1012, July 1989.

[32] A. Feuer and E. Weinstein, "Convergence analysis of LMS filters with uncorrelated Gaussian data," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-33, pp. 222–230, Feb. 1985.

[33] V. J. Mathews and Z. Xie, "Stochastic gradient adaptive filters with gradient adaptive step sizes," in *Proc. International Conference of Acoustics, Speech, and Signal Processing*, Albuquerque, USA, pp. 1385–1388, 1990.

[34] R. H. Kwong and E. W. Johnston, "A variable step size LMS algorithm," *IEEE Trans. Signal Processing*, vol. 40, pp. 1633–1642, July 1992.

[35] T. R. Fortescue, L. S. Kershenbaum, and B. E. Ydstie, "Implementation of self-tuning regulators with variable forgetting factors," *Automatica*, vol. 17, pp. 831–835, 1981.

[36] V. Solo, "The limiting behavior of LMS," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-37, pp. 1909–1922, Dec. 1989.

[37] V. Solo, "The error variance of LMS with time-varying weights," *IEEE Trans. Signal Processing*, vol. ASSP-40, pp. 803–813, April 1992.

[38] L. L. Horowitz and K. D. Senne, "Performance advantage of complex LMS for controlling narrow-band adaptive arrays," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-29, pp. 722–736, June 1981.

[39] J. E. Mazo, "On the independence theory of equalizer convergence," *The Bell System Technical Journal*, vol. 58, pp. 962–993, May–June 1979.

[40] S. Abu el Ata-Doss, J. L. Esteval and J. Richalet, "Dynamics of identification for the normalized least mean square algorithm," *Proc. IFAC Symp. on Ident. and Syst. Param. Est.*, York, U.K., pp. 1237-1242, 1985.

[41] L. Ljung, "Asymptotic behavior of the extended Kalman filter as a parameter estimator for linear systems," *IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 36–50, Jan. 1979.

[42] D. G. Lainiotis, "Partitioned estimation algorithms, I: Nonlinear estimation," *J. Inform. Sciences*, vol. 7, pp. 203–255, 1974.

[43] D. G. Lainiotis, "Partitioned estimation algorithms, I: Linear estimation," *J. Inform. Sciences*, vol. 7, pp. 317–340, 1974.

[44] P. R. Lamb and L. C. Westphal, "Simplex-directed partitioned adaptive filters," *Int. J. Control*, vol. 30, pp. 617–627, 1979.

[45] P. S. Maybeck, and K. P. Hentz, "Investigation of Moving-Bank Multiple Model Adaptive Algorithms," *J. Guidance*, vol. 10, pp. 90–96, 1987.

[46] C. P. Kwong, "Control theoretic design of the LMS and the sign algorithms in nonstationary environments," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-38, pp. 253–259, Feb. 1990.

[47] S. McLaughlin, B. Mulgrew, and C. F. N. Cowan, "Performance bounds for exponentially windowed RLS algorithm in a nonstationary environment," in *Proc. IMA Conf. on Math. in Signal Processing*, Warwick, England, pp. 449–464, Dec. 1988.

[48] M. L. R. de Campos and P. S. R. Diniz, personal communication, Jan 1992.

[49] R. S. Medaugh and L. J. Griffiths, "A comparison of two linear predictors," *IEEE Int. Conf. on Acoust., Speech and Signal Processing*, pp. 293–296, Atlanta, GA, 1981.

[50] F. Ling and J. G. Proakis, "Nonstationary learning characteristics of least squares adaptive estimation algorithms," *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Processing*, San Diego, CA, 1984.

[51] B. Porat, "Second-order equivalence of rectangular and exponential windows in least-squares estimation of Gaussian autoregressive processes," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-33, pp. 1209–1212, Oct. 1985.

[52] S. H. Ardalan and S. T. Alexander, "Fixed-point roundoff error analysis of the

exponentially windowed RLS algorithm for time-varying systems," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-35, pp. 770–783, June 1987.

[53] D. T. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least squares adaptive filtering," *IEEE Trans. on Signal Processing*, vol. 39, pp. 92–113, Jan. 1991.

[54] G. E. Bottomley, *Roundoff Error Problems and Solutions for Conventional Recursive Least Squares Filters*, Ph.D. dissertation, North Carolina State Univ., Raleigh NC, 1989.

[55] D. D. Falconer and L. Ljung, "Application of fast Kalman estimation to adaptive equalization," *IEEE Trans. Commun.*, vol. COM-26, pp. 1439–1446, Oct. 1978.

[56] J. M. Cioffi, *Fast Transversal Filters for Communications Applications*, Ph.D. dissertation, Stanford Univ., Stanford CA, Feb. 1984.

[57] B. Toplis and S. Pasupathy, "Tracking improvements in fast RLS algorithms using a variable forgetting factor," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, pp. 206–227, Feb. 1988.