

Self-Whitening Adaptive Equalization and Deconvolution Algorithms

Scott C. Douglas, Andrzej Cichocki, and Shun-ichi Amari



Technical Report # EE-99-001

Department of Electrical Engineering
School of Engineering and Applied Science
Southern Methodist University
Dallas, Texas 75275

January 1999

Affiliations

Scott C. Douglas: Department of Electrical Engineering, School of Engineering and Applied Science, Southern Methodist University, P.O. Box 750338, Dallas, Texas 75275 USA.

Andrzej Cichocki: Laboratory for Open Information Systems, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198 JAPAN, on leave from Warsaw University of Technology, Warsaw, POLAND.

Shun-ichi Amari: Laboratory for Information Synthesis, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198 JAPAN.

Acknowledgement

This work was supported in part by the U.S. Department of Justice under Contract No. JFBI-95107 at the University of Utah, Salt Lake City, Utah 84112 USA. The content of this document does not reflect the position and policy of the U.S. government, and no official endorsement should be inferred.

Portions of this work were first presented at the IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications, Paris, France, April 16-18, 1997 [Douglas *et al* 1997]. A shortened version of this document is scheduled for publication in *IEEE Transactions on Signal Processing* [Douglas *et al* 1999].

Abstract

In equalization and deconvolution tasks, the correlated nature of the input signal slows the convergence speeds of the least-mean-square (LMS) and other stochastic gradient adaptive filters. Prewhitening techniques have been proposed to improve convergence performance, but the additional coefficient memory and updates for the prewhitening filter can be prohibitive in some applications. In this report, we present two simple algorithms that employ the equalizer as a prewhitening filter within the gradient updates. A statistical analysis of these self-whitening algorithms indicates that they provide quasi-Newton convergence locally about the optimum coefficient solution for deconvolution and equalization tasks. Simulations indicate that the algorithms have excellent adaptation properties both for supervised and unsupervised (blind) adaptation criteria. Extensions of the techniques to multichannel deconvolution and equalization tasks are also described.

Contents

1	Introduction	4
2	Self-Whitening Algorithms	7
2.1	Prewhitening and Self-Whitening in Adaptive Filtering	7
2.2	The Filtered-Error/Regressor (FER) Algorithm	9
2.3	The Filtered-Regressor (FR) and Filtered-Error (FE) Algorithms	9
2.4	Algorithm Comparison	13
3	Analysis Summary	13
4	Multichannel Self-Whitening Algorithms	15
5	Simulations	18
5.1	Trained Adaptation Examples	18
5.2	Blind Adaptation Example	21
5.3	Multichannel Example	23
6	Conclusions	26
7	Appendix	27
7.1	Algebraic Comparison of the FER and FR Algorithms	27
7.2	Stationary Point Analysis	27
7.3	Analysis of Local Stability and Convergence Behavior	28
7.4	Normalized Step Sizes	30
7.5	Persistence of Excitation and Coefficient Initialization	31
8	References	32

1 Introduction

Adaptive channel equalization was one of the first widespread uses of adaptive filters [Lucky 1966, Sato 1975]. Today, channel equalization, deconvolution, and inverse modeling have numerous applications in communications, geophysical exploration, image processing, and inverse control [Widrow and Stearns 1985, Haykin 1994, Widrow and Walach 1996]. In these tasks, a linear, time-invariant filter generates a complex-valued signal $x(k)$ as

$$x(k) = \sum_{p=-\infty}^{\infty} h_p s(k-p), \quad (1)$$

where h_p , $-\infty \leq p \leq \infty$ is the impulse response of the unknown filter and $s(k)$ is an unknown zero-mean, i.i.d. source signal. The task is to process the signal $x(k)$ with a finite-impulse-response (FIR) filter with coefficient vector $\mathbf{w}(k) = [w_0(k) \cdots w_L(k)]^T$ such that $cs(k-\Delta)$ can be recovered, where c and Δ are scaling and time-delay factors, respectively. Depending on the application, training signal sets may or may not be available.

The simplest stochastic gradient algorithms for updating the coefficient vector are given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)\psi(e(k))\mathbf{x}^*(k) \quad (2)$$

$$e(k) = d(k) - y(k) \quad (3)$$

$$y(k) = \mathbf{x}^T(k)\mathbf{w}(k), \quad (4)$$

where $\mathbf{x}(k) = [x(k) \cdots x(k-L)]^T$ is the input signal vector, $d(k)$ is the desired response signal, $e(k)$ is the error signal, $\mu(k)$ is the step size at time k , and $\psi(e)$ is the derivative of the instantaneous-valued cost function $\phi(e)$. For example, if the squared-error cost function $\phi(e) = 0.5e^2$ is used, then $\psi(e) = e$ is obtained, and (2)–(4) becomes the well-known least-mean-square (LMS) adaptive filter [Widrow and Stearns 1985]. If the training signal $d(k)$ in (3) is not available, Bussgang methods such as the Sato algorithm [Sato 1975] or Godard/constant modulus algorithm (CMA) [Godard 1980, Treichler and Agee 1983] can be employed by replacing the error function $\psi(e(k))$ in (2) with the blind error function $f(y(k))$ whose form depends on the statistics of $s(k)$ [Benveniste and Goursat 1984, Bellini and Rocca 1996]. For example, if the CMA cost function is used, we select $f(y) = (A - |y|^2)y$, where A is the modulus of the source signal.

While useful, these stochastic gradient algorithms are known to have poor performance when the input signal $x(k)$ is highly correlated. In this situation, the input signal autocorrelation matrix $\mathbf{R}_{\mathbf{xx}} = E\{\mathbf{x}^*(k)\mathbf{x}^T(k)\}$, where $E\{\cdot\}$ denotes statistical expectation and $*$ denotes complex-conjugate, governs the speeds of adaptation of the modes of the system. When $\mathbf{R}_{\mathbf{xx}}$ has a large

eigenvalue spread, it is either difficult or impossible to choose a step size sequence that achieves the desired performance level in a reasonable number of iterations [Widrow and Stearns 1985, Shynk *et al* 1991].

Several techniques have been suggested to improve the convergence properties of stochastic gradient adaptive filters. Gauss-Newton methods [Ljung and Söderström 1983], of which the recursive least-squares (RLS) algorithm is an example, can be used to improve an equalizer’s convergence speed and estimation accuracy at convergence, but such systems are computationally-intensive and difficult to implement in hardware even in their most efficient forms [Slock and Kailath 1991]. More recently, quasi-Newton techniques that employ additional filters to effectively whiten the input signal within the adaptation algorithm have been proposed [Slock 1992, Mboup *et al* 1994, Gay and Tavathia 1995, Tanaka *et al* 1995]. Figure 1 shows the block diagram for this class of systems, where

$$W(z, k) = \sum_{l=0}^L w_l(k)z^{-l} \quad \text{and} \quad P(z, k) = \sum_{n=0}^N p_n(k)z^{-n} \quad (5)$$

are the z -transforms of $\mathbf{w}(k)$ and $\mathbf{p}(k)$, respectively, and the vector $\mathbf{p}(k) = [p_0(k) \cdots p_N(k)]^T$ contains the coefficients of a second FIR filter, usually a linear predictor with $p_0(k) = 1$, designed such that the signal $x_f(k) = \sum_{n=0}^N p_n(k)x(k-n)$ is approximately decorrelated. Since the statistics of $x(k)$ are often unknown or time-varying, however, $\mathbf{p}(k)$ must be adapted using either stochastic gradient [Mboup *et al* 1994] or least-squares [Slock 1992, Gay and Tavathia 1995, Tanaka *et al* 1995] techniques, adding additional complexity to the system.

In this report, we propose simple modifications to the system in Figure 1 that remove the need for a separate predictor filter with its associated coefficient memory and adaptive updates. The techniques are related to a method for improving the convergence performance of blind source separation algorithms for instantaneous additive mixtures [Amari *et al* 1996]. Simply put, our modified algorithms replace the adaptive predictor $\mathbf{p}(k)$ with the adaptive filter $\mathbf{w}(k)$ in this system. Using the equalizer as a prewhitening filter is motivated by the fact that the optimum equalizer filter also decorrelates the input signal. Thus, our proposed algorithms iteratively *self-whiten* the input signal as they converge to the optimum coefficient solution. The overall complexity of the most accurate and efficient version of the system is approximately 50% greater than that of the LMS adaptive filter and is less than half the complexity of the most-efficient stabilized RLS fast transversal scheme [Slock and Kailath 1991]. Thus, the proposed algorithms provide quasi-Newton behavior locally about an equalizer solution without the coefficient updates and memory for a second predictor filter [Slock 1992, Mboup *et al* 1994, Gay and Tavathia 1995,

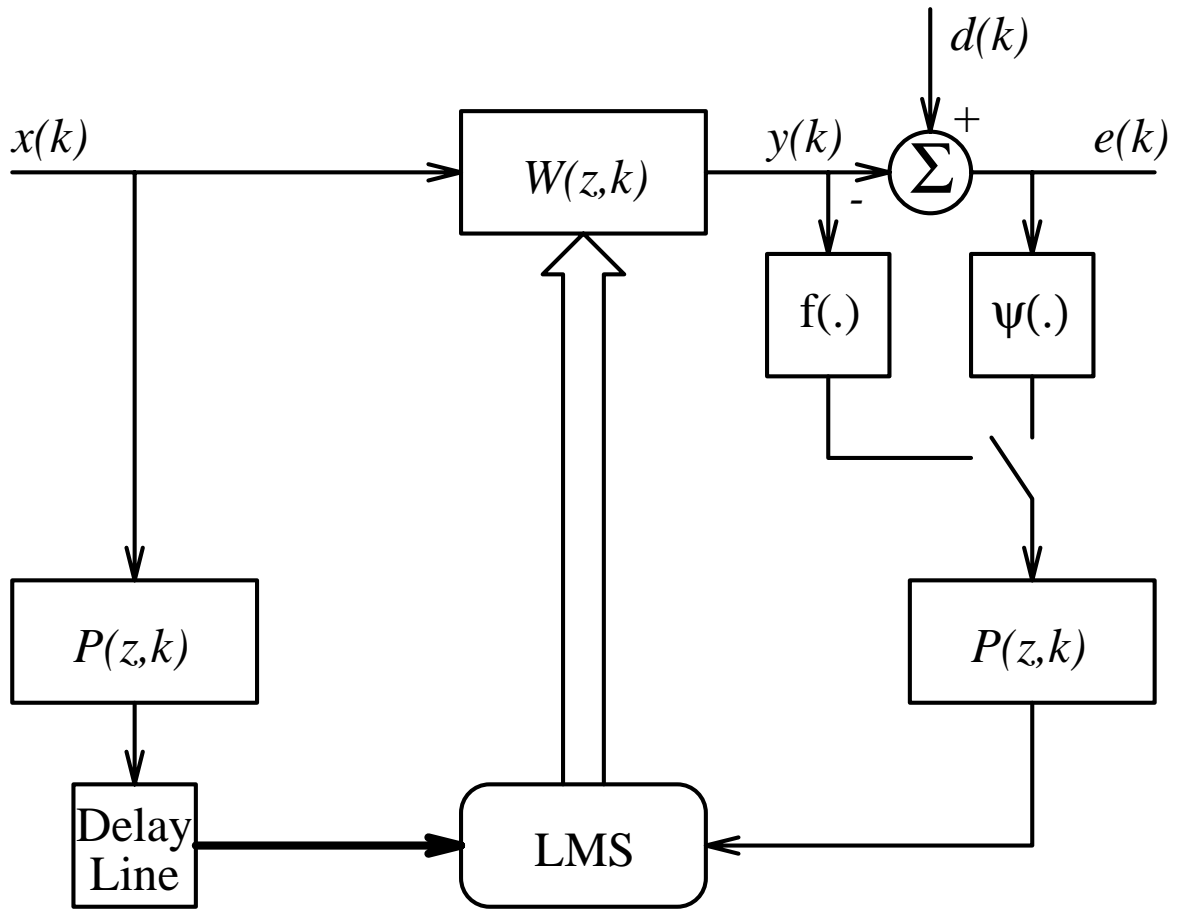


Figure 1: Adaptive filter employing prewhitening filters.

Tanaka *et al* 1995] and without the significantly-increased complexity of a Gauss-Newton or RLS-based scheme [Ljung and Söderström 1983, Slock and Kailath 1991].

The organization of this report is as follows. In Section 2, we describe the goal of prewhitening within stochastic gradient adaptive filtering, and we propose two self-whitening algorithms that use the equalizer as a prewhitening filter. We also compare the implementation complexity of the proposed schemes with an existing prewhitening method [Mboup *et al* 1994]. In Section 3, we summarize the results of several analytical studies of the proposed self-whitening schemes, the details of which appear in the Appendix. In Section 4, we present multichannel versions of both self-whitening schemes and relate their forms to existing algorithms for adaptive feedforward control. Simulations comparing the performance of the proposed schemes with existing prewhitening, Gauss-Newton, and RLS schemes in both trained and blind adaptation modes are provided in Section 5. Section 6 presents our conclusions and suggestions for future work.

2 Self-Whitening Algorithms

2.1 Prewhitening and Self-Whitening in Adaptive Filtering

It is well-known [Widrow and Stearns 1985, Ljung and Söderström 1983] that stochastic gradient adaptive filters trained with a known desired response signal $d(k)$ perform best when operating on stationary, zero-mean, uncorrelated input signals, such that $\mathbf{R}_{\mathbf{xx}} = \sigma_x^2 \mathbf{I}$, where $\sigma_x^2 = E\{|x^2(k)|\}$. In adaptive equalization and deconvolution tasks, however, $x(k)$ is a correlated sequence. For example, in situations where the signal model in (1) is valid, the (i, j) th entry of $\mathbf{R}_{\mathbf{xx}}$ is the $(i - j)$ th term of the autocorrelation function of the channel impulse response h_p , which is by definition not equal to a unit impulse function if deconvolution or equalization needs to be performed.

To obtain a system that operates on uncorrelated input signals, one can apply a linear transformation to the input signal $x(k)$ or the input signal vector $\mathbf{x}(k)$ to produce a whitened input signal vector. This transformation can take the form of a matrix $\mathbf{P}(k)$, such that $\mathbf{P}(k)\mathbf{x}(k)$ is the transformed vector, or it can take the form of a filter, as in the system in Figure 1 in which $x_f(k)$ is approximately uncorrelated. In either case, a new signal vector is obtained whose elements are uncorrelated with equal powers. If the filter $P(z, k)$ is used, then it should satisfy

$$|P(e^{j\Omega}, k)|^2 \approx |c_1|^2 S_{xx}^{-1}(\Omega), \quad (6)$$

where $S_{xx}(\Omega)$ is the power spectral density of the input signal $x(k)$ and c_1 is an arbitrary constant. The relationship in (6) is not unique; *i.e.*, there exist an infinite number of filters whose squared-magnitude frequency response is equal to the inverse of the input power spectral density. In practice,

any filter $P(z, k)$ that approximately satisfies (6) could be used when $d(k)$ is available, although prefilters with a lower group delay have better performance as they introduce less delay within the resulting system. In the case of the LMS algorithm in which $\psi(e) = e$, it can be shown that applying the filter $P(z, k)$ to both the input signal and the error signal within the update alone as shown in Figure 1 is equivalent to prefiltering both $x(k)$ and $d(k)$ prior to the application of the LMS adaptive filter to the prefiltered signals, so long as both $P(z, k)$ and $W(z, k)$ are slowly-varying. The filtered-X LMS algorithm for adaptive feedforward control [Widrow and Stearns 1985, Kuo and Morgan 1996] also uses this approximation and is known to have well-behaved convergence properties.

In blind deconvolution and equalization tasks, applying a prewhitening filter that satisfies (6) to $\mathbf{x}(k)$ and $f(y(k))$ within the coefficient updates is not sufficient to obtain significant improvements in the convergence behavior of the system. In filters employing blind adaptation criteria, the system's adaptation modes depend on higher-order statistics through the form of $f(\cdot)$, and these modes are not decoupled through simple prewhitening. In such cases, fast adaptation can only be realized when the transformed signal vector elements are i.i.d., and we provide a simulation example in Section 5 indicating this fact. Such will be the case in noiseless deconvolution and equalization tasks if $P(z, k)$ is of the form

$$P(z, k) = cH^{-1}(z)z^{-\Delta}, \quad (7)$$

where c and Δ are an arbitrary constant and integer, respectively.

In equalization and deconvolution tasks in which (1) is valid, we have that

$$S_{xx}(\Omega) = |H(e^{j\Omega})|^2, \quad (8)$$

where $H(z)$ is the z -transform of h_p . Assuming that the equalizer is adequate for equalizing the channel, we have that $y(k) \approx cs(k - \Delta)$ near convergence such that

$$W(z, k) \approx cH^{-1}(z)z^{-\Delta}. \quad (9)$$

Comparing (6), (8), and (9), we see that choosing $W(z, k)$ for $P(z, k)$ yields a prewhitening algorithm in the deconvolution or equalization task near convergence. Moreover, such a choice also satisfies (7) and thus is appropriate for blind adaptation criteria as well. We call this technique “self-whitening” because the equalizer is also used as the prewhitener, and to our knowledge, this technique has not been previously explored in the technical literature. How well such a procedure works away from the optimum coefficient solution clearly depends on the channel; however,

as we show via simulations, there typically exists a sizable region about the optimum coefficient solution whereby using the equalizer as a prewhitener significantly increases the system's overall performance.

2.2 The Filtered-Error/Regressor (FER) Algorithm

Given the previous arguments and discussions, we propose to replace $P(z, k)$ with $W(z, k)$ in Figure 1 to obtain the *filtered-error/regressor* (FER) algorithm given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)g(k)\mathbf{y}^*(k) \quad (10)$$

$$g(k) = \psi(\mathbf{e}^T(k))\mathbf{w}(k), \quad (11)$$

where $\mathbf{y}(k) = [y(k) \cdots y(k-L)]^T$ and $\psi(\mathbf{e}(k)) = [\psi(e(k)) \cdots \psi(e(k-L))]^T$, respectively. If blind adaptation is employed, the blind error signal $f(y(m))$ is used instead of $\psi(e(m))$ in (11). Figure 2 depicts the block diagram for this algorithm, where the portion of the system in dashed lines need not be implemented as $y(k)$ is already computed by the equalizer.

If the mean-squared error criterion is used such that $\psi(e) = e$, then the FER algorithm becomes

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)(d_f(k) - \mathbf{y}^T(k)\mathbf{w}(k))\mathbf{y}^*(k) \quad (12)$$

$$d_f(k) = \mathbf{d}^T(k)\mathbf{w}(k), \quad (13)$$

where $\mathbf{d}(k) = [d(k) \cdots d(k-L)]^T$. In this case, the FER algorithm is an LMS adaptive filter whose input and desired response signals are $y(k)$ and $d_f(k)$, respectively. For this reason, we are motivated to compute the step size in (10) as

$$\mu(k) = \frac{\mu}{\beta + \|\mathbf{y}(k)\|_q^q}, \quad (14)$$

where μ and β are positive parameters and q is a positive integer. For $q = 2$, Equations (12)–(14) describe a normalized LMS adaptive filter operating on filtered input and desired response signals. This version of the algorithm is particularly desirable, as NLMS adaptive filters have several useful convergence and stability properties [Slock 1993, Douglas and Meng 1994].

2.3 The Filtered-Regressor (FR) and Filtered-Error (FE) Algorithms

A comparison of the system in Figure 2 with similar systems for feedforward active noise control [Kuo and Morgan 1996] shows that it is identical to the filtered-X LMS algorithm when $\psi(e) = e$ if the plant and adaptive controller are identical. We can use this connection to develop a second approximate implementation of this system, as shown in Figure 3. This system filters the

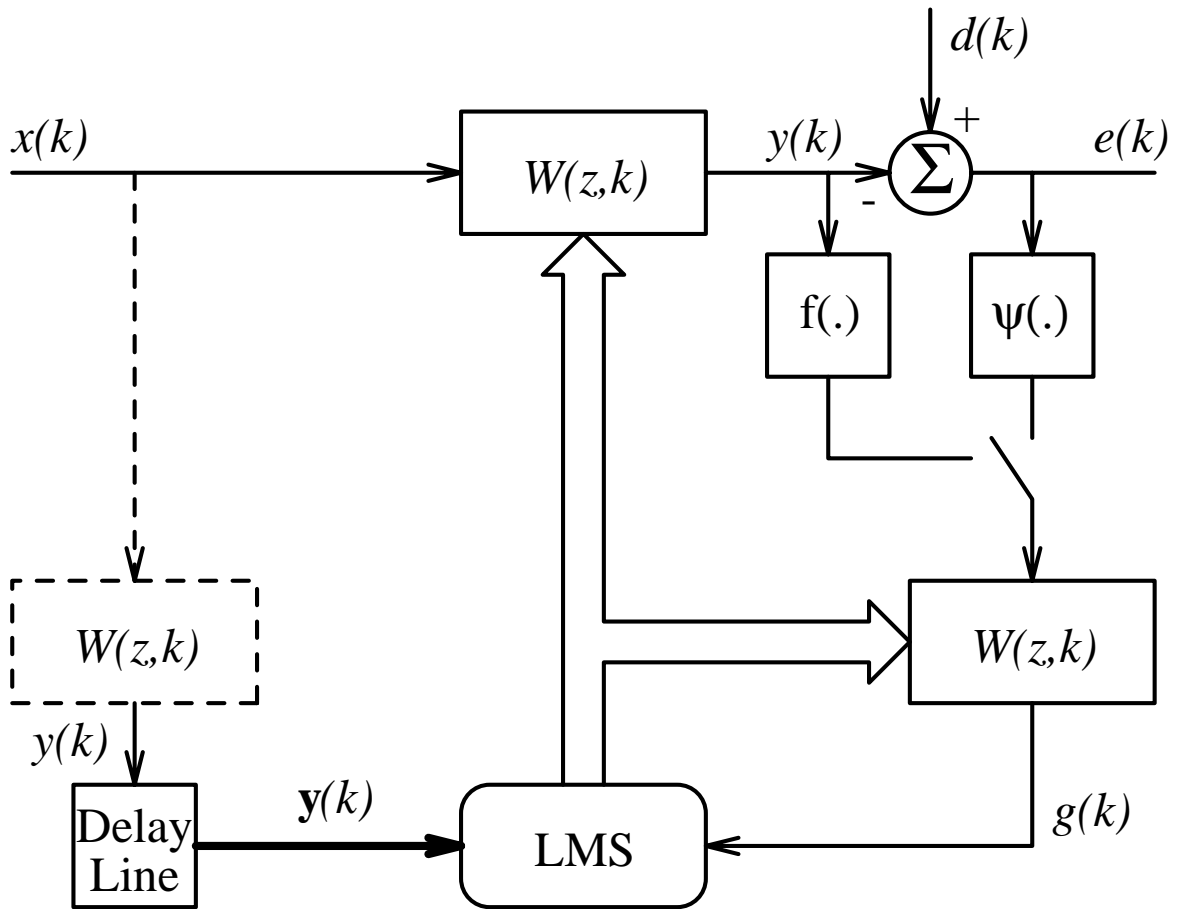


Figure 2: The self-whitening filtered-error/regressor (FER) algorithm.

regressor vector signal by $z^{-L}W^*(z^{-1}, k)W(z, k)$ before forming the coefficient updates using a delayed version of the error signal. This *filtered-regressor* algorithm is given by

$$\mathbf{w}(k+1) = \mathbf{w}(k) + \mu(k)\psi(e(k-L))\mathbf{u}^*(k), \quad (15)$$

where $\mathbf{u}(k) = [u(k) \cdots u(k-L)]^T$ and $u(k)$ is computed as

$$u(k) = \sum_{m=0}^L w_{L-m}^*(k)y(k-m). \quad (16)$$

The step size for (15) can be computed using either (14) or

$$\mu(k) = \frac{\mu}{(\beta + \|\mathbf{x}(k)\|_q^q)\|\mathbf{w}(k)\|_2^2}. \quad (17)$$

The FR algorithm is the dual of the adjoint LMS algorithm proposed in [Wan 1996] as applied to our equalization task. A version of (15) for blind equalization tasks was first presented in [Douglas *et al* 1996]. In addition, the modification is similar to that used in a recently-derived algorithm for multichannel blind source separation [Amari *et al* 1997a].

The FER and FR algorithms can be shown to be quite similar in form. In the Appendix, the algebraic forms of the two algorithms are compared, and it is shown that the only difference between the averaged forms of the two algorithms is due to the delays in the coefficient values employed within the updates. Analyses of the delayed LMS [Long *et al* 1989, Long *et al* 1992] and filtered-X LMS [Bjarnason 1995] algorithms indicate that both have similar average behaviors to their non-delayed counterparts for small step sizes. Similarly, our simulations of (10) and (15) indicate that they too have nearly-identical behaviors for suitably small step sizes. In situations where a small step size is required for both accurate estimation of the channel inverse and stability of the algorithm, either algorithm can be chosen. However, the FER algorithm can be expected to have somewhat better adaptation performance than the FR algorithm for larger step sizes, as the former algorithm has less coefficient delay in its updates than does the latter algorithm.

In addition to the FER and FR algorithms, a third algorithm can be developed in which the error signal is filtered by $z^{-L}W^*(z^{-1}, k)W(z, k)$ before being multiplied by the delayed regressor vector $\mathbf{x}(k-L)$ in the coefficient updates. This *filtered-error* (FE) algorithm is a version of the adjoint LMS algorithm as applied to our equalization task. However, the FE algorithm exhibited erratic and unstable behavior in simulations for numerous cases involving both training data and blind adaptation criteria. Attempts to stabilize its behavior using the techniques proposed in [Orgren *et al* 1991] caused a significant decrease in adaptation performance. For these reasons, we do not discuss the FE algorithm in the remainder of this report.

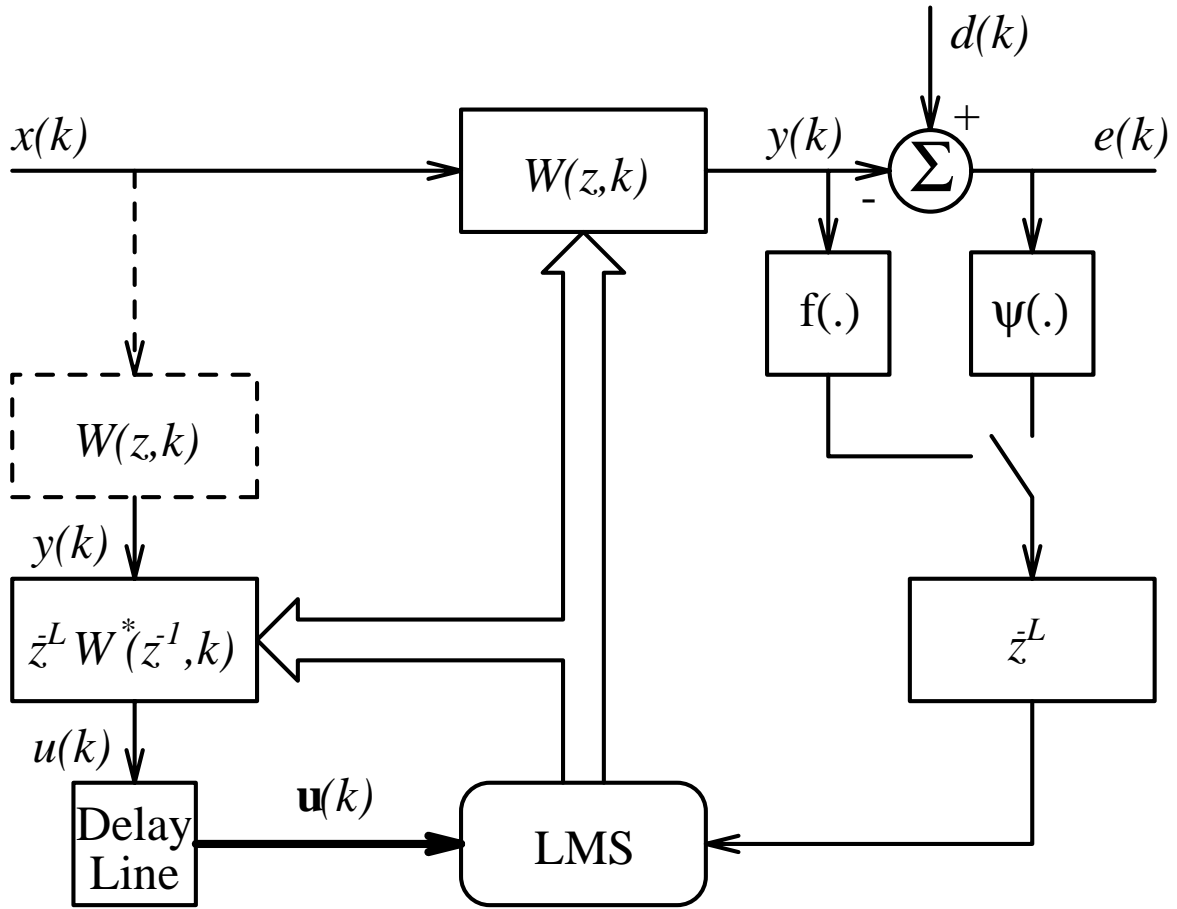


Figure 3: The self-whitening filtered-regressor (FR) algorithm.

Table 1: Comparison of Algorithm Complexities

<i>Quantity</i>	<i>LMS</i>	<i>FER</i>	<i>FR</i>	<i>Ref. [Mboup et al 1994]</i>
<i># of MACs</i>	$2L + 3$	$3L + 4$	$3L + 4$	$2L + 2N + 3$
<i># of Mem. Loc.</i>	$3L + 4$	$4L + 4$	$5L + 5$	$3L + 2N + 6$

2.4 Algorithm Comparison

We now compare the complexities of the FER and FR algorithms with the standard LMS adaptive equalizer and the prewhitening algorithm in [Mboup *et al* 1994] in the case where $\psi(e) = e$ and $\mu(k)$ is a fixed constant. Table 1 lists the number of multiply/accumulates (MACs) and memory locations required for each algorithm, where $L + 1$ is the equalizer length and N is the order of the LMS adaptive predictor filter in [Mboup *et al* 1994]. As can be seen, the proposed algorithms require about 50% more MACs than those for the LMS adaptive equalizer, and they require between 50% and 100% more memory locations. This increase in system complexity also brings improved performance, as we shall show. Moreover, one must select $N = (L + 1)/2$ for the algorithm in [Mboup *et al* 1994] if it is to have similar complexity and memory requirements as those of the FER and FR algorithms. Note that the complexities of the proposed algorithms are less than half that of the stabilized fast transversal filter in [Slock and Kailath 1991].

3 Analysis Summary

In this section, we summarize the results of several analyses of the behaviors and performances of the FER and FR algorithms. The details of these analyses appear in the Appendix.

Stationary Points of the Algorithms. The stationary points of an algorithm are the set of coefficient vectors \mathbf{w} defined by the relationship $E\{\mathbf{w}(k + 1)\} = E\{\mathbf{w}(k)\} = \mathbf{w}$. For the LMS algorithm, the only stationary point is the Wiener solution $\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{p}_{dx}$, where $\mathbf{p}_{dx} = E\{d(k)\mathbf{x}(k)\}$. The stationary points of the FER and FR algorithms are the same for both algorithms, but these stationary points generally differ from those of the stochastic gradient algorithm in (2)–(4). In noiseless deconvolution and equalization tasks where (1) is valid, however, the differences between the proposed algorithms' solutions and those of the standard stochastic gradient algorithm are small if the length of the equalizer is adequate to model the channel inverse with little error.

Local Convergence Behavior. A statistical analysis of either the FER or FR algorithm's convergence behavior is challenging due to the coefficient delays that appear within the updates. In the Appendix, we provide a statistical analysis of a similar algorithm whose updates depend on the coefficient values at time k , as such an algorithm can be expected to have similar behavior to those of the FER and FR algorithms when $\mu(k)$ is small. In stationary system identification tasks where $d(k)$ can be modeled as

$$d(k) = \hat{d}(k) + \varepsilon(k) \quad (18)$$

$$\hat{d}(k) = \mathbf{x}^T(k) \mathbf{w}_{opt}, \quad (19)$$

$\hat{d}(k)$ is the noiseless desired response signal, \mathbf{w}_{opt} is an optimum coefficient vector, and $\varepsilon(k)$ is a stationary noise signal that is independent of $x(k)$, then the mean behavior of the coefficient error vector $\tilde{\mathbf{w}}(k) = \mathbf{w}(k) - \mathbf{w}_{opt}$ when $\|\tilde{\mathbf{w}}(k)\|^2$ is small is

$$E\{\tilde{\mathbf{w}}(k)\} = (\mathbf{I} - \mu(k)E\{\psi'(\varepsilon(k))\})\mathbf{R}_{\hat{\mathbf{d}}\hat{\mathbf{d}}}E\{\tilde{\mathbf{w}}(k)\}, \quad (20)$$

where $\mathbf{R}_{\hat{\mathbf{d}}\hat{\mathbf{d}}}$ is the autocorrelation matrix of $\hat{d}(k)$. In equalization and deconvolution tasks, $\hat{d}(k) \approx cs(k - \Delta)$, where $E\{\mathbf{s}^*(k)\mathbf{s}^T(k)\} = \sigma_s^2\mathbf{I}$, and the form of the analysis equation in (20) becomes similar to that of the LMS/Newton algorithm [Widrow and Stearns 1985]. Thus, the FER and FR algorithms provide quasi-Newton adaptation behavior about the optimum coefficient solution in this situation, independent of the correlation statistics of the input signal.

Step Size Selection. Choosing a step size for the FER and FR algorithms to guarantee their stability is a challenging task due to the coefficient delays within the updates. In the Appendix, we show for a similar algorithm without coefficient delays and with $\psi(e) = e$ that choosing $\mu(k)$ as in (14) with $q = 2$ yields a normalized coefficient update. The normalized LMS adaptive filter is known to have a number of performance advantages over the LMS adaptive filter, and while (14) does not guarantee stability of the FER or FR algorithms for a fixed range of μ , it can provide proper scaling of the updates for particular choices of $\psi(e)$ and q . Moreover, for the FER algorithm with $\psi(e) = e$ and $\mu(k)$ chosen in (14) with $q = 2$, we obtain a normalized LMS adaptive filter that is guaranteed to be stable for all $0 < \mu < 2$ and $\beta > 0$, although convergence to the optimum coefficient vector is not guaranteed.

Coefficient Initialization. Both the FER and FR algorithms require a non-zero initial coefficient vector $\mathbf{w}(0)$, as $y(k) = 0$ for $k \geq 0$ if $\mathbf{w}(0) = \mathbf{0}$. Moreover, the coefficients of the systems need

to acquire a reasonably-accurate estimate of the inverse of the channel to engage their algorithms' self-whitening capabilities. In fact, it is the output signal $y(k)$, and not the input signal $x(k)$, that must satisfy a persistence of excitation condition to avoid ill-convergence of the self-whitening algorithms. Thus, coefficient initializations that cause the combined channel $W(z, k)H(z)$ to have deep spectral nulls should be avoided, as they could cause slow initial convergence behavior.

If training data are available, we can use the stochastic gradient algorithm in (2) as an initialization procedure to adapt the filter coefficients for k_0 iterations, at which point adaptation of $\mathbf{w}(k)$ is switched to the FER or FR algorithm to obtain fast and accurate convergence behavior. Typically, the normalized step size $\mu(k) = \mu/(\beta + \|\mathbf{x}(k)\|_2^2)$ is chosen in (2) to provide the fastest possible adaptation behavior so that k_0 can be as small as possible. When blind adaptation criteria are employed or when the above initialization scheme is not practical, a standard center-tap initialization scheme for the FER and FR algorithms, in which

$$w_l(0) = \alpha\delta(l - p) \quad (21)$$

for some $0 \leq p \leq L$ and $\alpha > 0$, can be used. For this initialization scheme, the initial convergence behaviors of the FER and FR algorithms are the same as that of the delayed-NLMS adaptive filter, which is known to have robust convergence properties [Rupp and Frenzel 1994, Ahn and Voltz 1996].

4 Multichannel Self-Whitening Algorithms

We now consider multichannel deconvolution and equalization tasks in which the received signals $\{x_j(k)\}$, $1 \leq j \leq J$, are produced from the source signals $\{s_i(k)\}$, $1 \leq i \leq I$, $I \leq J$, as

$$x_j(k) = \sum_{i=1}^I \sum_{p=-\infty}^{\infty} h_{jip} s_i(k-p), \quad (22)$$

where $\{h_{jip}\}$, $1 \leq i \leq I$, $1 \leq j \leq J$, $-\infty \leq p \leq \infty$ are a set of mixing coefficients. We calculate a set $\{y_i(k)\}$, $1 \leq i \leq I$ of scaled and/or delayed estimates of the source signals as

$$y_i(k) = \sum_{j=1}^J \mathbf{x}_j^T(k) \mathbf{w}_{ij}(k), \quad (23)$$

where $\mathbf{w}_{ij}(k) = [w_{ij0}(k) \cdots w_{ijL}(k)]^T$ contains the FIR filter coefficients for the (i, j) th channel of the multichannel equalizer and $\mathbf{x}_j(k) = [x_j(k) \cdots x_j(k-L)]^T$. The multichannel stochastic gradient algorithm for this system is

$$\mathbf{w}_{ij}(k+1) = \mathbf{w}_{ij}(k) + \mu(k)\psi_i(e_i(k))\mathbf{x}_j^*(k), \quad (24)$$

$$e_i(k) = d_i(k) - y_i(k), \quad (25)$$

where $e_i(k)$ and $d_i(k)$ are the i th error and desired response signals, respectively. A simplified normalized step size for (24) can be computed as

$$\mu(k) = \frac{\mu}{\beta + \sum_{j=1}^J \|\mathbf{x}_j(k)\|_q^q}. \quad (26)$$

We can easily extend the FER and FR algorithms to the multichannel case by exploiting their connection to existing multichannel active noise control algorithms [Kuo and Morgan 1996]. For a J -input, I -output, J -error multichannel active noise control system, each coefficient update consists of sums of products of the J filtered error terms at the output of the plant with IJ^2 filtered input terms. For a multichannel version of the system in Figure 2, the l th filtered error signal $g_l(k)$, $1 \leq l \leq J$, is

$$g_l(k) = \sum_{i=1}^I \psi_i(\mathbf{e}_i^T(k)) \mathbf{w}_{il}(k), \quad (27)$$

where $\psi_i(\mathbf{e}_i(k)) = [\psi_i(e_i(k)) \cdots \psi_i(e_i(k-L))]^T$. We compute the IJ^2 filtered regressor signals

$$y_{ijl}(k) = \mathbf{x}_j^T(k) \mathbf{w}_{il}(k). \quad (28)$$

Then, $\mathbf{w}_{ij}(k)$ is updated using the multichannel FER algorithm given by

$$\mathbf{w}_{ij}(k+1) = \mathbf{w}_{ij}(k) + \mu(k) \sum_{l=1}^J g_l(k) \mathbf{y}_{ijl}^*(k). \quad (29)$$

where $\mathbf{y}_{ijl}(k) = [y_{ijl}(k) \cdots y_{ijl}(k-L)]^T$. Alternatively, we can use the multichannel FR algorithm given by

$$\mathbf{w}_{ij}(k+1) = \mathbf{w}_{ij}(k) + \mu(k) \psi_i(e_i(k-L)) \mathbf{u}_j^*(k), \quad (30)$$

where $\mathbf{u}_j(k) = [u_j(k) \cdots u_j(k-L)]^T$ and $u_j(k)$ for $1 \leq j \leq J$ is computed as

$$u_j(k) = \sum_{i=1}^I \sum_{m=0}^L w_{ij(L-m)}^*(k) y_i(k-m). \quad (31)$$

In analogy with (14), a simplified normalized step size can be computed as

$$\mu(k) = \frac{\mu}{\beta + \sum_{i=1}^I \|\mathbf{y}_i(k)\|_q^q}, \quad (32)$$

Each of these algorithms are summarized in component form in Tables 2 and 3, respectively.

<i>Equation</i>	<i>MACs</i>
for $i = 1$ to I do $y_i(k) = \sum_{j=1}^J \sum_{m=0}^L w_{ijm}(k)x_j(k-m)$ $e_i(k) = d_i(k) - y_i(k)$ end for $l = 1$ to J do $u_j(k) = \sum_{i=1}^I \sum_{m=0}^L w_{ij(L-m)}^*(k)y_i(k-m)$ end for $i = 1$ to I do for $j = 1$ to J do for $m = 0$ to L do $w_{ijm}(k+1) = w_{ijm}(k) + \mu(k)\psi_i(e_i(k-L))u_j^*(k-m)$ end end end end	$IJ(L+1)$ I $IJ(L+1)$ $I(L+1)$
Total: $3IJ(L+1) + I$	

Table 3: The multichannel FR algorithm.

5 Simulations

We now explore the properties of the algorithms in (10) and (15) through simulations. In each case, we have used MATLAB's `rand` or `randn` and `filter` commands to generate the input and desired response signals. One hundred simulations have been run and the results averaged to obtain each plotted trajectory or convergence curve.

5.1 Trained Adaptation Examples

For our first example, we consider a two-coefficient equalization task in which $x(k)$ is defined as

$$x(k) = ax(k-1) + d(k), \quad (33)$$

where $a = 0.95$ and $d(k)$ is an i.i.d. binary- $\{\pm 1\}$ sequence. In this case, the optimum equalizer coefficients are $\mathbf{w}_{opt} = [1 \ -0.95]^T$, and $\mathbf{R}_{\mathbf{xx}}$ is

$$\mathbf{R}_{\mathbf{xx}} = \frac{1}{1 - (0.95)^2} \begin{bmatrix} 1 & 0.95 \\ 0.95 & 1 \end{bmatrix}. \quad (34)$$

Figure 4(a) shows the mean trajectories of the coefficients for the LMS adaptive equalizer in this situation for six different initial coefficient vectors $\mathbf{w}(0) = \mathbf{w}_{opt} + [\cos(i\pi/3) \ \sin(i\pi/3)]^T$, $i = \{1, 2, 3, 4, 5, 6\}$ with $\mu = 0.005$. Because $\mathbf{R}_{\mathbf{xx}}$ has a condition number of 39 in this case, the

coefficients converge slowly along the eigenvector direction of $[1 \ -1]/\sqrt{2}$ associated with the minimum eigenvalue of $\mathbf{R}_{\mathbf{x}\mathbf{x}}$. Figures 4(b) and (c) show the mean trajectories of the coefficients for the FER and FR algorithms, respectively, where $\psi(e) = e$ and $\mu(k)$ is computed according to (14) and (17) for the algorithms, respectively, with $\mu = 0.01$, $\beta = 0.1$, and $q = 2$. As can be seen, the filter coefficients take a more direct path to the optimum solution vector for each of the proposed algorithms. Figure 4(d) shows the corresponding mean coefficient trajectories for the LMS/Newton algorithm in which $\mathbf{x}(k)$ in (2) is replaced by $\mathbf{R}_{\mathbf{x}\mathbf{x}}^{-1}\mathbf{x}(k)$, where $\mu = 0.005$. The similarity of Figure 4(d) to Figures 4(b) and (c) indicate that the FER and FR algorithms behave like quasi-Newton algorithms locally about the optimum coefficient vector \mathbf{w}_{opt} .

In our second example, we consider an autoregressive system identification task in which

$$x(k) = x(k-1) - 0.98x(k-2) + 0.97x(k-3) + s(k) \quad (35)$$

$$d(k) = s(k) + \varepsilon(k), \quad (36)$$

where $s(k)$ and $\varepsilon(k)$ are zero-mean uncorrelated Gaussian signals with variances of 1.0 and 10^{-6} , respectively. Figure 5 shows the total coefficient error powers $\|\mathbf{w}(k) - \mathbf{w}_{opt}\|_2^2$ for four different algorithms for $L = 3$. The step sizes for the NLMS and FER adaptive filters were computed as $\mu(k) = \mu/(\beta + \|\mathbf{x}(k)\|_q^q)$ and (14), where $\beta = 0.1$ and $q = 2$ for both algorithms. In this case, we have chosen $\mu = 1$ for the NLMS algorithm to provide the fastest possible convergence behavior for this system, and we have selected $\mu = 0.035$ for the FER algorithm so that the total coefficient error powers are the same for both systems in steady-state. Here, we have chosen random vectors $\mathbf{w}(0)$ such that $\|\mathbf{w}(0) - \mathbf{w}_{opt}\|_2^2 = 1$ for each simulation run in our averaged results. As can be seen, the FER algorithm converges after about 1000 iterations on average, whereas the NLMS algorithm takes approximately 2000 iterations on average to converge in this situation. Also shown for comparison is the prewhitening algorithm of [Mboup *et al* 1994] in which a prewhitening prediction filter of $N = 2$ is chosen so that this system's complexity is similar to that of the FER algorithm. The step sizes for the equalizer and adaptive predictor were computed as $\mu(k) = 0.13/(0.1 + \sum_{p=0}^L x_f^2(k-p))$ and $\mu_P(k) = 0.1/(0.1 + \sum_{p=1}^N x^2(k-p))$, respectively, and these choices provided fastest stable behavior for this system. This algorithm does not perform well because the length of the prediction filter is not adequate to whiten the input signal in this case when the algorithm's complexity is similar to that of the FER algorithm. Finally, we show the performance of the $O(8L)$ stabilized FTF algorithm in this task [Slock and Kailath 1991]. While the FTF algorithm has the best performance in this task, it also is more than twice as complex as the FER algorithm. The behavior of the FR algorithm is not shown in this example as it could not provide fast convergence due to the coefficient

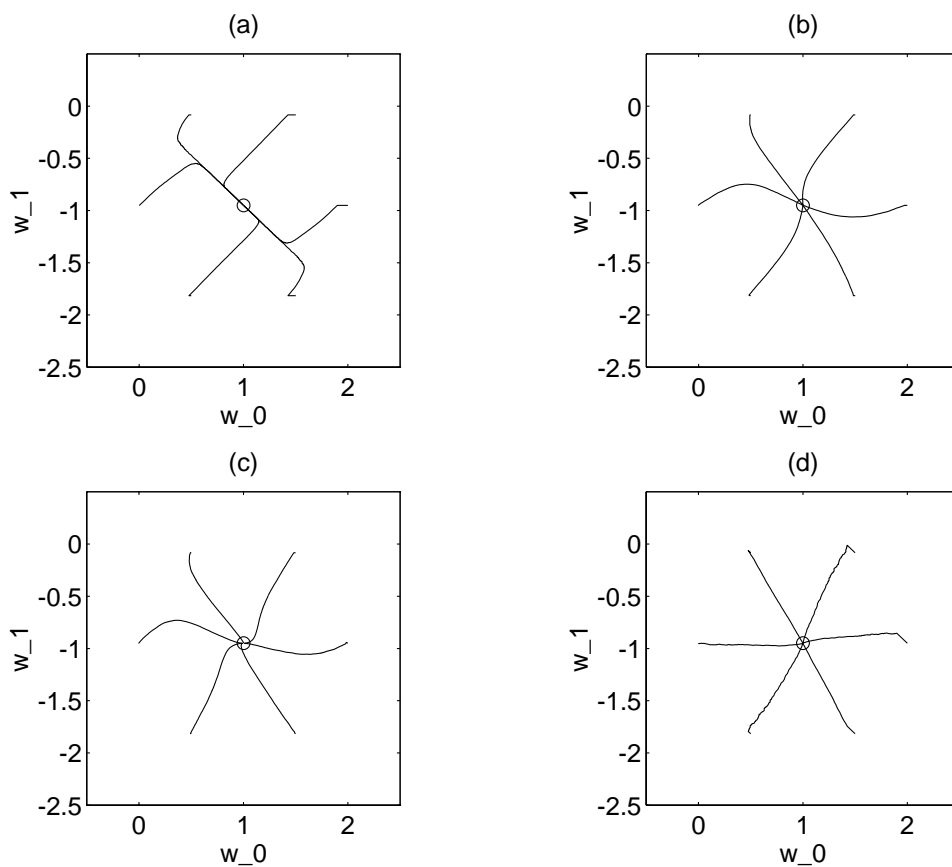


Figure 4: Averaged coefficient trajectories for the two-coefficient equalizer using (a) the LMS algorithm, (b) the FER algorithm, (c) the FR algorithm, and (d) the LMS-Newton algorithm in the first example.

delays within its updates, although both the FER and FR algorithms had similar average behaviors with these signals for smaller step sizes.

5.2 Blind Adaptation Example

We now present results from simulations of the FER and FR algorithms as applied to a blind equalization task using the constant modulus algorithm (CMA) error nonlinearity $f(y) = (A - |y|^2)y$. For these simulations, we have chosen $L + 1 = 11$ coefficient equalizers, and $s(k)$ is chosen as a quadrature-amplitude-modulated (QAM) signal with unity modulus ($A = 1$). The received signal $x(k)$ is generated using a non-minimum phase all-pass channel with transfer function

$$H(z) = e^{j\pi/4} \frac{0.7 - z^{-1}}{1 - 0.7z^{-1}}, \quad (37)$$

such that exact equalization with a finite-length equalizer is not possible, a realistic situation [Ding *et al* 1994]. For the standard CMA algorithm, we have chosen $\mu(k) = \mu/(\beta + \|\mathbf{x}(k)\|_2^2)$ where $\mu = 0.01$ and $\beta = 0.1$ provide the fastest-converging adaptation behavior for this system. For the FER CMA and FR CMA algorithms, we have chosen the step size normalization scheme in (14) with $\mu = \beta = 0.1$ and $q = 2$. The center tap of the initial coefficient vector $\mathbf{w}(0)$ is set to one for all equalizers.

Figure 6(a), (b), and (c) show the values of $y(k)$ over the ranges $k \in [1, 400]$, $k \in [1001, 1400]$, and $k \in [2001, 2400]$, respectively, for the standard CMA equalizer, where each dot indicates one output value. Clearly, the equalizer has not converged even after 2000 iterations of the equalizer for this channel. Figure 7 shows the corresponding output values for the FER CMA equalizer. The adaptive filter quickly equalizes the channel up to an unknown phase rotation of the QAM constellation. Figure 8 shows the corresponding output values for the FR CMA equalizer. A comparison of Figures 7 and 8 show that both algorithms have similar behavior. The improved performances of the self-whitening algorithms over that of standard CMA adaptation in this example also support our previous claim that self-whitening is better than prewhitening for blind adaptive equalizers, as the all-pass nature of $H(z)$ in this example implies that the elements of $\underline{\mathbf{x}}(k)$ are already uncorrelated for the standard CMA algorithm.

Shown in Figure 9 are the corresponding signal constellations for a Gauss-Newton version of the CMA equalizer, in which the regressor vector is computed as $\mathbf{R}_{\mathbf{xx}}^{-1}(k)\mathbf{x}^*(k)$, where

$$\mathbf{R}_{\mathbf{xx}}(k) = \lambda\mathbf{R}_{\mathbf{xx}}(k) + (1 - \lambda)\mathbf{x}^*(k)\mathbf{x}^T(k), \quad (38)$$

$\mathbf{R}_{\mathbf{xx}}(0) = 0.1\mathbf{I}$, $\lambda = 0.995$, and $\mu = 0.005$. Although the performance of this equalizer is the best of the three, we see that its initial acquisition of the channel inverse is similar to those of the FER

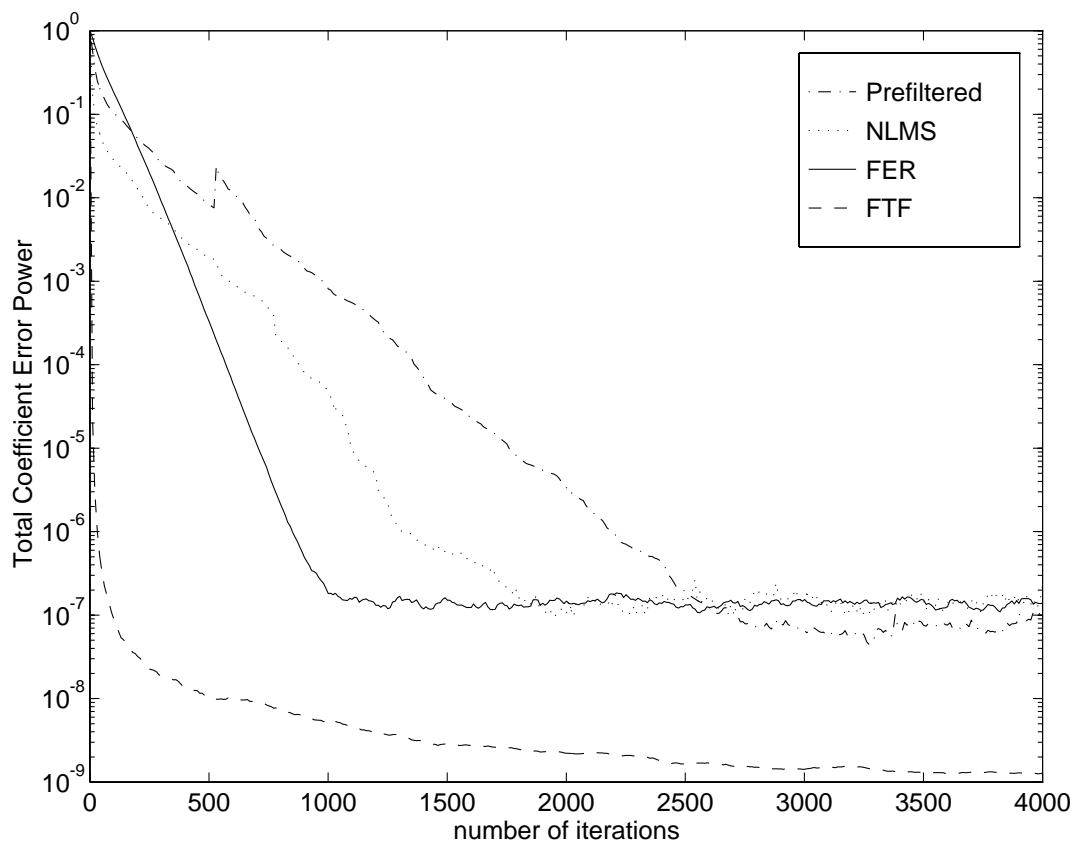


Figure 5: Evolution of total coefficient error powers for each of the algorithms in the inverse system identification example.

and FR algorithms. Since the step sizes for all of the algorithms are limited by the nature of the blind error nonlinearity $f(y)$ for stability, the self-whitening algorithms are extremely attractive for this and other blind equalization and deconvolution tasks.

5.3 Multichannel Example

We now illustrate the behavior and performance of the multichannel FR algorithm in (30)–(31) via a simulation example. We choose three i.i.d. binary- $\{\pm 1\}$ source signals for this experiment, and we generate the measured signals in $\underline{\mathbf{x}}(k) = [x_1(k) \ x_2(k) \ x_3(k)]^T$ from the source signals in $\underline{\mathbf{s}}(k) = [s_1(k) \ s_2(k) \ s_3(k)]^T$ as

$$\underline{\mathbf{x}}(k) = \mathbf{A}_1 \underline{\mathbf{x}}(k-1) + \mathbf{A}_2 \underline{\mathbf{x}}(k-2) + \mathbf{B}_0 \underline{\mathbf{s}}(k) + \mathbf{B}_1 \underline{\mathbf{s}}(k-1), \quad (39)$$

where

$$\mathbf{A}_1 = \begin{bmatrix} -0.06 & 0.24 & -0.04 \\ -0.18 & 0.18 & -0.26 \\ -0.08 & -0.20 & 0.06 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0.04 & 0.02 & 0.03 \\ 0.08 & 0.04 & 0 \\ 0.03 & 0.06 & 0.06 \end{bmatrix}, \quad (40)$$

$$\mathbf{B}_0 = \begin{bmatrix} 0.02 & 0.07 & 0.05 \\ 0 & 0.09 & 0.08 \\ 0.07 & 0.04 & 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{B}_1 = \begin{bmatrix} 0.1 & 0 & 0.4 \\ 0.5 & 0.4 & 0.7 \\ 0.7 & 0.1 & 0.6 \end{bmatrix}, \quad (41)$$

respectively. The multichannel system in (39) is non-minimum phase and cannot be perfectly equalized with a finite-length causal filter. We process the received vector sequence using a three-input, three-output multichannel equalizer with $L = 6$. For the first 300 iterations of the system, the equalizer is adapted using the LMS algorithm in (24)–(26) with $d_i(k) = s_i(k-4)$ and $\psi_i(e) = e$ for $1 \leq i \leq 3$, $\mu = 1$, $\beta = 0.1$, and $q = 2$, at which point adaptation is switched to the multichannel FR algorithm in (30)–(32) where $d_i(k) = s_i(k-4)$ and $\psi_i(e) = e$ for $1 \leq i \leq 3$, $\mu = 0.3$, $\beta = 0.1$, and $q = 2$.

Figure 10 shows the convergence of the total average squared error for the proposed system, computed using ensemble averages of $\sum_{i=1}^3 (s_i(k) - y_i(k+4))^2$. Also shown for comparison are the squared errors from the multichannel LMS equalizer with $\psi_i(e) = e$, $\mu = 1$, $\beta = 0.1$, and $q = 2$. The convergence speed of the FR algorithm after its activation at iteration $k_0 = 300$ is clearly much faster than that of the LMS algorithm in this case. Although these results do not guarantee that the self-whitening algorithms have better performance than that of the LMS algorithm in general, they are an indication of the improved convergence characteristics that the self-whitening algorithms can provide in the multichannel case.

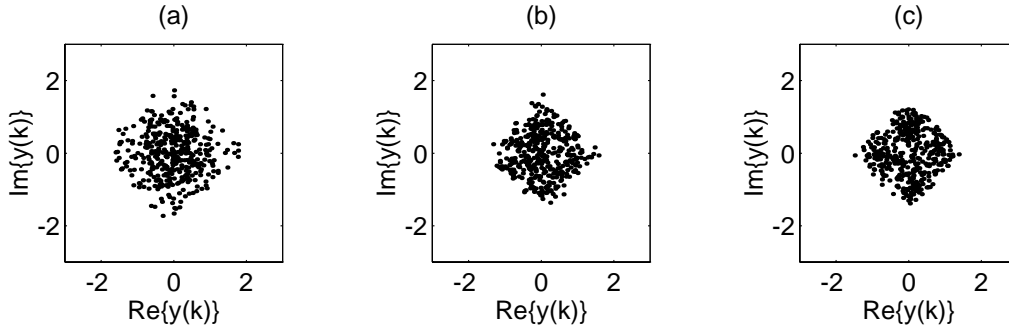


Figure 6: Output signal constellations for the CMA equalizer in the blind equalization example.

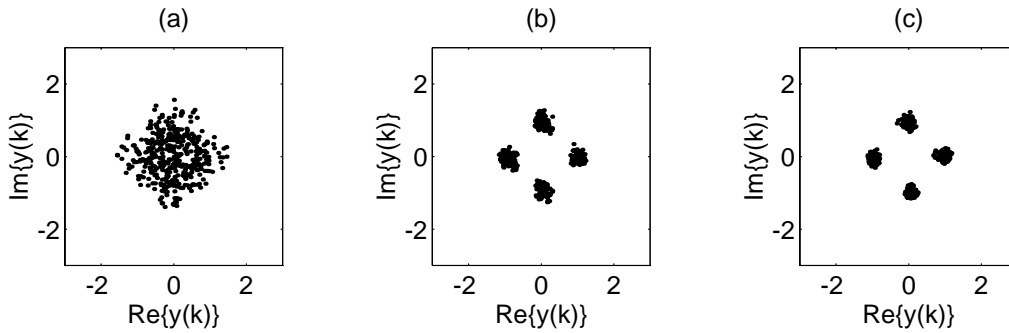


Figure 7: Output signal constellations for the FER CMA equalizer in the blind equalization example.

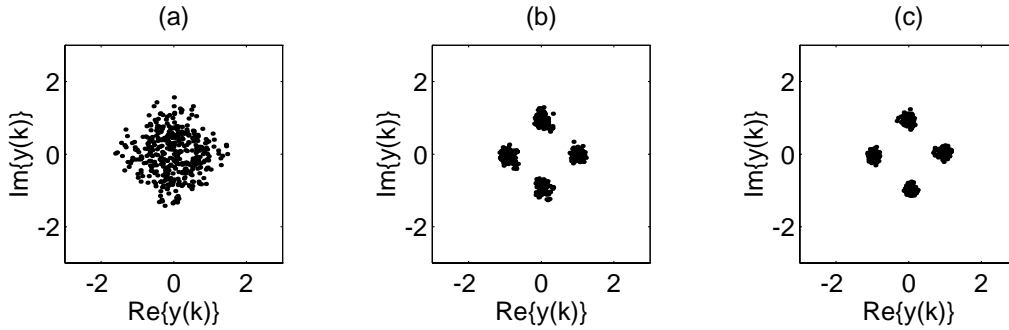


Figure 8: Output signal constellations for the FR CMA equalizer in the blind equalization example.

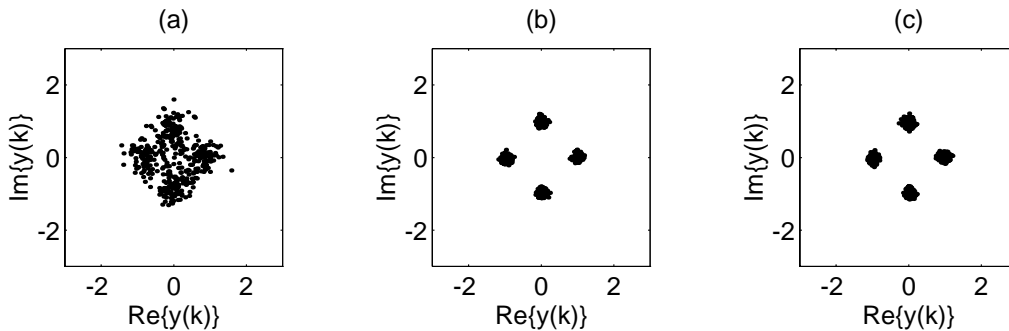


Figure 9: Output signal constellations for the Gauss-Newton CMA equalizer in the blind equalization example.

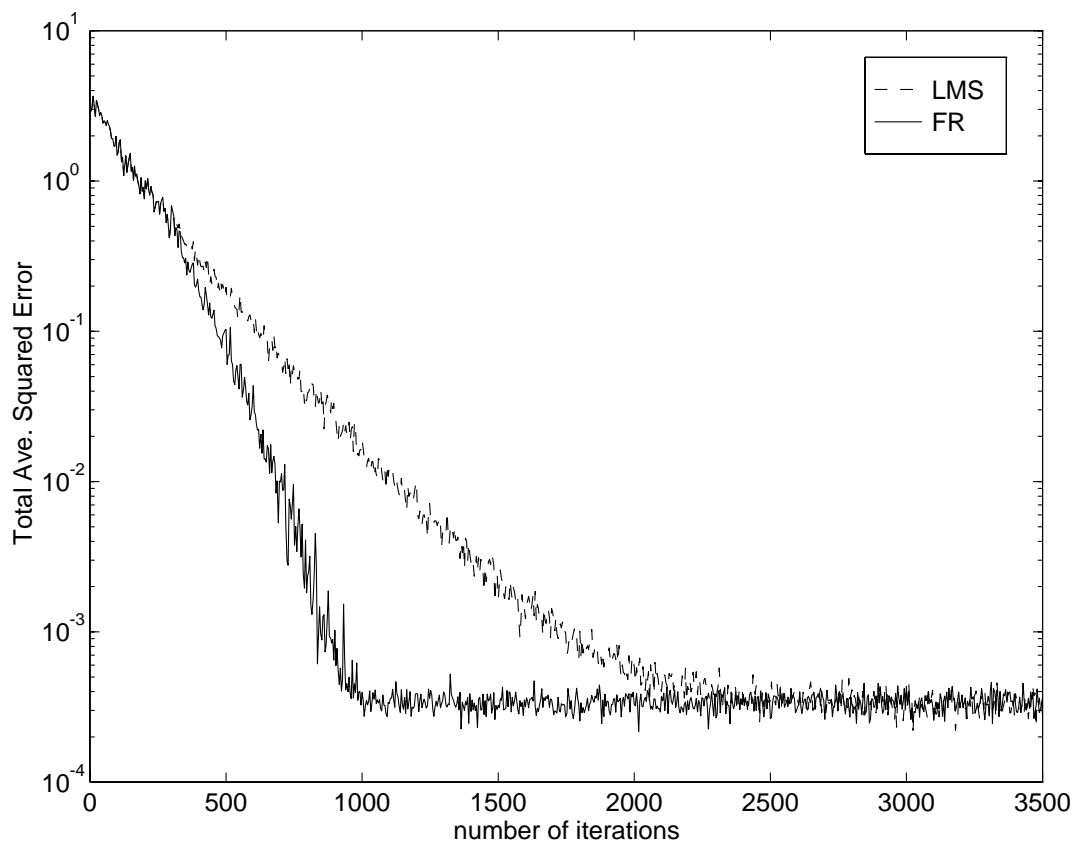


Figure 10: Evolution of total average squared error for the equalizers adapted using the LMS and FR algorithms in the multichannel example.

6 Conclusions

We have described two algorithms for single- and multichannel equalization and deconvolution tasks that self-whiten the input signal(s) within the coefficient updates. These algorithms are particular versions of prewhitening algorithms that do not require separate predictor filters to operate. Through our analyses and simulations we have shown that

- the algorithms exhibit quasi-Newton coefficient adaptation behavior locally about an equalizing or deconvolving solution,
- the complexities of the simplest algorithms are only 50% greater than those of the corresponding LMS algorithm and are typically much less than competing Gauss-Newton schemes, and
- they perform well in both trained and blind adaptation modes.

The proposed methods are expected to have many uses for problems in the communications, control, and signal processing areas.

Although our analyses and simulation results have indicated the usefulness of the self-whitening schemes, there are no existing proofs of convergence for these algorithms. Such proofs would be useful to obtain, in particular for the projection-based update of (12)–(14) with $q = 2$, as this system appears to be well-behaved, stable, and convergent for a wide range of operating conditions. This issue, as well as investigations of the performance behaviors of the multichannel algorithms for fractionally-spaced equalization, are the subject of current research.

7 Appendix

7.1 Algebraic Comparison of the FER and FR Algorithms

In this section, we show for stationary signals that any differences in the average behaviors of the FER and FR algorithms is due to the coefficient delays appearing within the updates.

Let the superscripts (er) and (r) denote quantities associated with the FER and FR algorithms, respectively. Consider the update for $w_i^{(er)}(k)$ in (10). Using the definitions of $y(k)$ and $g(k)$ in (4) and (11), respectively, the update for the i th filter coefficient of this algorithm is

$$w_i^{(er)}(k+1) = w_i^{(er)}(k) + \mu(k) \sum_{m=0}^L w_m^{(er)}(k) \sum_{n=0}^L w_n^{*(er)}(k-i) \psi(e^{(er)}(k-m)) x^*(k-i-n). \quad (42)$$

Similarly, using (16), the i th coefficient update for the FR algorithm in (15) is

$$w_i^{(r)}(k+1) = w_i^{(r)}(k) + \mu(k) \sum_{m=0}^L w_m^{(r)}(k-i) \sum_{n=0}^L w_n^{*(r)}(k+m-L-i) \psi(e^{(r)}(k-L)) x^*(k+m-L-i-n). \quad (43)$$

Letting $k_m = k + m - L$ in the summation on the RHS of (43), we can rewrite this update as

$$w_i^{(r)}(k+1) = w_i^{(r)}(k) + \mu(k) \sum_{m=0}^L w_m^{(r)}(k-i) \sum_{n=0}^L w_n^{*(r)}(k_m-i) \psi(e^{(r)}(k_m-m)) x^*(k_m-i-n). \quad (44)$$

Assume that the signals $d(k)$ and $x(k)$ are statistically-stationary. Then, we have that

$$E_{\mathbf{w}(k-m)} \{ \psi(e(k-m)) x^*(k-i-n) \} = E_{\mathbf{w}(l-m)} \{ \psi(e(l-m)) x^*(l-i-n) \} \quad (45)$$

if $\mathbf{w}(k) = \mathbf{w}(l)$, where $E_{\mathbf{w}}\{\cdot\}$ denotes statistical expectation conditioned on the coefficient vector \mathbf{w} . Thus, if we assume slow adaptation such that $\mathbf{w}(k) \approx \mathbf{w}(k-1) \approx \dots \approx \mathbf{w}(k-2L)$, then the right-hand-sides of (42) and (44) are identical for identical coefficient vectors, and thus the behaviors of the algorithms are expected to be similar for small step size values.

7.2 Stationary Point Analysis

In this section, we determine the system of equations that define the stationary points \mathbf{w} of the algorithms in (10) and (15). For this analysis, define $\nu(k)$ as

$$\nu(k) = d(k) - \mathbf{x}^T(k) \mathbf{w}, \quad (46)$$

and let $\mathbf{w} = \mathbf{w}(m)$, $k-L \leq m \leq k+1$ for the FER algorithm in (10). Then, by taking expectations of both sides of (10), we find after some algebra that \mathbf{w} is a stationary point of this algorithm if

$$\mathbf{w}^{*T} \mathbf{R}_{\mathbf{x}\psi}(\nu)(l) \mathbf{w} = 0, \quad 0 \leq l \leq L, \quad (47)$$

where

$$\mathbf{R}_{\mathbf{x}\psi(\boldsymbol{\nu})}(l) = E\{\mathbf{x}^*(k-l)\psi(\boldsymbol{\nu}^T(k))\}. \quad (48)$$

and $\boldsymbol{\nu}(k) = [\nu(k) \cdots \nu(k-L)]^T$. A similar analysis of the FR algorithm in (15) for $\mathbf{w} = \mathbf{w}(m)$, $k-2L \leq m \leq k+1$ yields the system of equations

$$\sum_{m=0}^L w_m E\{\psi(\nu(k-L))\mathbf{x}^{*T}(k+m-L-l)\}\mathbf{w}^* = 0, \quad 0 \leq l \leq L. \quad (49)$$

When $x(k)$ and $d(k)$ are stationary, (49) is identical to (47), and thus both algorithms have the same set of stationary points.

The stationary points for standard stochastic gradient algorithms that minimize $E\{\phi(e(k))\}$ are defined by

$$E\{\psi(\nu(k))x^*(k-l)\} = 0, \quad 0 \leq l \leq L, \quad (50)$$

such that $\psi(\nu(k))$ is uncorrelated with $x(k)$ over the time interval $[k-L, k]$. Since the (i, j) th entry of $\mathbf{R}_{\mathbf{x}\psi(\boldsymbol{\nu})}(l)$ is $E\{\psi(\nu(k))x^*(k-i+j-l)\}$ for stationary signals, the conditions in (50) do not guarantee that (47) is satisfied. Thus, the solutions obtained by the FER and FR algorithms do not minimize the value of $E\{\phi(e(k))\}$ in general. However, from (47), a set of sufficient conditions for \mathbf{w} to be a stationary point of either the FER or FR algorithms is

$$E\{\psi(\nu(k))x^*(k-l)\} = 0, \quad -L \leq l \leq 2L, \quad (51)$$

such that $\psi(\nu(k))$ is uncorrelated with $x(k)$ over the time interval $[k-2L, k+L]$. These conditions are similar to those in (50). Moreover, in situations where L is large enough and the time interval of the filter is properly chosen to provide an accurate estimate of the inverse of the channel, the coefficient values that satisfy (50) nearly satisfy (51), and thus the differences in the solutions obtained by the algorithms is small.

7.3 Analysis of Local Stability and Convergence Behavior

To study the local stability and local convergence behavior of (10) and (15), consider the algorithm given by

$$\bar{\mathbf{w}}(k+1) = \bar{\mathbf{w}}(k) + \bar{\mu}(k)(\psi(\bar{\mathbf{e}}^T(k))\bar{\mathbf{w}}(k))\bar{\mathbf{y}}^*(k) \quad (52)$$

$$\bar{\mathbf{e}}(k) = \mathbf{d}(k) - \bar{\mathbf{y}}(k) \quad (53)$$

$$\bar{\mathbf{y}}(k) = \mathbf{X}(k)\bar{\mathbf{w}}(k), \quad (54)$$

where $\bar{\mathbf{w}}(k)$ is the coefficient vector of this system and $\mathbf{X}(k)$ is an $(L + 1 \times L + 1)$ -dimensional matrix whose $(i + 1, j + 1)$ th element is $x(k - i - j)$. This algorithm's coefficient updates at time k depend on coefficient values at time k , unlike those in (10) and (15). Even so, the local stability behavior of this algorithm for small step sizes and stationary signals is the same as that in (10) and (15), and the transient behavior about any stationary point of the algorithm can be expected to be similar to those of (10) and (15) for small step size values.

By linearizing the coefficient updates about a stationary point \mathbf{w} , we can determine an approximate equation for the evolution of $E\{\tilde{\mathbf{w}}(k)\}$, where

$$\tilde{\mathbf{w}}(k) = \bar{\mathbf{w}}(k) - \mathbf{w} \quad (55)$$

is a small deviation of $\bar{\mathbf{w}}(k)$ about \mathbf{w} . For this analysis, we shall assume that both $d(k)$ and $x(k)$ are statistically stationary with finite second-order moments and that the pair $\{d(m), \mathbf{X}(m)\}$ is independent of $\{d(n), \mathbf{X}(n)\}$ if $m \neq n$. These assumptions are similar to the independence assumptions commonly employed in statistical analyses of adaptive filters [Haykin 1996]. For a discussion of the accuracy of such assumptions, see [Mazo 1979, Douglas and Pan 1995]. We shall also assume that $\psi(e)$ is continuous and differentiable at all points, although this assumption can be relaxed if the joint probability distributions of $d(k)$ and $\mathbf{X}(k)$ are suitably smooth.

Using (55), we can express (53) as

$$\bar{\mathbf{e}}(k) = \mathbf{d}(k) - \mathbf{X}(k)(\mathbf{w} + \tilde{\mathbf{w}}(k)) \quad (56)$$

$$= \boldsymbol{\nu}(k) - \mathbf{X}(k)\tilde{\mathbf{w}}(k), \quad (57)$$

where $\boldsymbol{\nu}(k)$ is defined in (46). Expand $\psi(\bar{\mathbf{e}}(k))$ in a Taylor series about $\boldsymbol{\nu}(k)$ as

$$\psi(\bar{\mathbf{e}}(k)) = \psi(\boldsymbol{\nu}(k)) - \text{diag}[\psi'(\boldsymbol{\nu}(k))]\mathbf{X}(k)\tilde{\mathbf{w}}(k) + O(\|\tilde{\mathbf{w}}(k)\|_2^2), \quad (58)$$

where $\psi'(e) = d\psi(e)/de$, $\text{diag}[\psi'(\boldsymbol{\nu}(k))]$ is a diagonal matrix whose $(i + 1, i + 1)$ th entry is $\psi'(\nu(k - i))$, and $O(\|\tilde{\mathbf{w}}(k)\|_2^2)$ denotes terms that are proportional to products of the elements of $\tilde{\mathbf{w}}(k)$.

Subtracting \mathbf{w} from both sides of (52), we have

$$\tilde{\mathbf{w}}(k + 1) = \tilde{\mathbf{w}}(k) + \bar{\boldsymbol{\mu}}(k)(\psi(\bar{\mathbf{e}}^T(k))\bar{\mathbf{w}}(k))\mathbf{X}^*(k)\bar{\mathbf{w}}^*(k). \quad (59)$$

We now substitute the expressions in (55) and (58) for $\bar{\mathbf{w}}(k)$ and $\psi(\bar{\mathbf{e}}(k))$ in (59), respectively.

Simplifying the result gives

$$\begin{aligned} \tilde{\mathbf{w}}(k + 1) &= (\mathbf{I} - \bar{\boldsymbol{\mu}}(k)\mathbf{X}^*(k)\mathbf{w}^*\mathbf{w}^T \text{diag}[\psi'(\boldsymbol{\nu}(k))]\mathbf{X}(k))\tilde{\mathbf{w}}(k) \\ &\quad + \bar{\boldsymbol{\mu}}(k)(\psi(\boldsymbol{\nu}^T(k))\mathbf{w}\mathbf{X}^*(k)\tilde{\mathbf{w}}^*(k) + \mathbf{X}^*(k)\mathbf{w}^*\psi(\boldsymbol{\nu}^T(k))\tilde{\mathbf{w}}(k)) + O(\|\tilde{\mathbf{w}}(k)\|_2^2). \end{aligned} \quad (60)$$

If the magnitudes of the elements of $\tilde{\mathbf{w}}(k)$ are small, the terms of $O(\|\tilde{\mathbf{w}}(k)\|_2^2)$ in (60) can be neglected with respect to the other terms in the equation. Taking expectations of both sides of the equation and using our assumptions gives

$$\begin{aligned} E\{\tilde{\mathbf{w}}(k+1)\} &= (\mathbf{I} - \bar{\mu}(k)E\{\mathbf{X}^*(k)\mathbf{w}^*\mathbf{w}^T \text{diag}[\psi'(\boldsymbol{\nu}(k))]\mathbf{X}(k)\})E\{\tilde{\mathbf{w}}(k)\} \\ &\quad + \bar{\mu}(k) \left(E\{\psi(\boldsymbol{\nu}^T(k))\mathbf{w}\mathbf{X}^*(k)\}E\{\tilde{\mathbf{w}}^*(k)\} + E\{\mathbf{X}^*(k)\mathbf{w}^*\psi(\boldsymbol{\nu}^T(k))\}E\{\tilde{\mathbf{w}}(k)\} \right) \end{aligned} \quad (61)$$

In this expression, the j th column vector of the matrix $E\{\psi(\boldsymbol{\nu}^T(k))\mathbf{w}\mathbf{X}^*(k)\}$ is equal to $\mathbf{R}_{\mathbf{x}\psi(\boldsymbol{\nu})}(j)\mathbf{w}$, and the i th row vector of the matrix $E\{\mathbf{X}^*(k)\mathbf{w}^*\psi(\boldsymbol{\nu}^T(k))\}$ is equal to $\mathbf{w}^{*T}\mathbf{R}_{\mathbf{x}\psi(\boldsymbol{\nu})}(i)$.

Now, we consider situations in which the adaptive filter can accurately model the inverse of the channel such that the conditions in (51) are satisfied for some coefficient vector \mathbf{w}_{opt} . One such situation is the system identification task in which $d(k)$ is given by (18)–(19). For $\mathbf{w} = \mathbf{w}_{opt}$, we have $\mathbf{R}_{\mathbf{x}\psi(\boldsymbol{\nu})}(l) = \mathbf{O}$ for $0 \leq l \leq L$. Thus, the last term on the RHS of (61) is zero, and thus

$$E\{\tilde{\mathbf{w}}(k+1)\} = (\mathbf{I} - \bar{\mu}(k)E\{\mathbf{X}^*(k)\mathbf{w}_{opt}^*\mathbf{w}_{opt}^T \text{diag}[\psi'(\boldsymbol{\nu}(k))]\mathbf{X}(k)\})E\{\tilde{\mathbf{w}}(k)\}, \quad (62)$$

which reduces to (20) if $\bar{\mu}(k) = \mu(k)$, as $\boldsymbol{\nu}(k) \approx \boldsymbol{\varepsilon}(k)$ when $\tilde{\mathbf{w}}(k)$ is small. Therefore, the behavior of (52) locally about the optimum coefficient solution is the same as that of an LMS adaptive filter about its optimum coefficient solution with a step size equal to $\bar{\mu}(k)E\{\psi'(\boldsymbol{\varepsilon}(k))\}$ and with $\hat{d}(k)$ as its input signal.

7.4 Normalized Step Sizes

We now consider the behavior of the algorithm in (52) from a deterministic standpoint when $\psi(e) = e$. We can write this algorithm as

$$\mathbf{w}(k+1) = \bar{\mathbf{w}}(k) + \bar{\mu}(k)\bar{\mathbf{y}}^*(k)\bar{\mathbf{w}}^T(k)\bar{\mathbf{e}}(k). \quad (63)$$

Suppose that $\mathbf{d}(k) = \mathbf{X}(k)\mathbf{w}_{opt}$ such that perfect system identification is possible. Using (55), we can write (63) in terms of $\tilde{\mathbf{w}}(k)$ as

$$\tilde{\mathbf{w}}(k+1) = (\mathbf{I} - \bar{\mu}(k)\bar{\mathbf{y}}^*(k)\bar{\mathbf{w}}^T(k)\mathbf{X}(k))\tilde{\mathbf{w}}(k) \quad (64)$$

$$= (\mathbf{I} - \bar{\mu}(k)\bar{\mathbf{y}}^*(k)\bar{\mathbf{y}}^T(k))\tilde{\mathbf{w}}(k). \quad (65)$$

Therefore, by choosing

$$\bar{\mu}(k) = \frac{\mu}{\beta + \|\bar{\mathbf{y}}(k)\|_2^2}, \quad (66)$$

we obtain the normalized LMS algorithm for (63). This update reduces the magnitude of the component of $\tilde{\mathbf{w}}(k)$ in the direction of $\bar{\mathbf{y}}(k)$ for all $0 < \mu < 2$ and $\beta > 0$ in the noiseless case.

The algorithms in (10) and (15) for $\psi(e) = e$ differ from (52) in that the current coefficient updates are computed using past coefficient values. Even if $\mu(k)$ is computed as in (14) with $q = 2$ for these algorithms, $\|\mathbf{w}(k) - \mathbf{w}_{opt}\|_2^2$ is not guaranteed to monotonically decrease to zero for all $0 < \mu < 2$ and $\beta > 0$. Even so, choosing $\mu(k)$ as in (14) or (17) improves the convergence behaviors of these schemes by properly scaling the magnitudes of the update terms if $\psi(e) = |e|^{q-2}e$. Moreover, choosing $\mu(k)$ as in (14) for (12) guarantees the stability of this projection-based algorithm, although its convergence to the optimum coefficient vector is not guaranteed even in the noiseless system identification task.

7.5 Persistence of Excitation and Coefficient Initialization

We now explore the persistence of excitation conditions needed for proper convergence of the FER and FR algorithms. Recall that the robust convergence behavior of the LMS algorithm is assured only if the input vector sequence $\{\mathbf{x}(k)\}$ satisfies a persistence of excitation condition given by the following [Sethares and Johnson 1991]: for all k , there exists $K < \infty$, $\delta_1 > 0$, and $\delta_2 > 0$ such that

$$\delta_1 \mathbf{I} < \frac{1}{K} \sum_{n=k}^{k+K} \mathbf{X}^*(n) \mathbf{X}(n)^T < \delta_2 \mathbf{I}. \quad (67)$$

Such a result cannot be easily extended to the FER and FR algorithms as the vectors $\mathbf{y}(k)$ and $\mathbf{u}(k)$ depend on current and past coefficient values which are time-varying and are not guaranteed to be asymptotically-bounded. However, for small step sizes, the elements of $\mathbf{y}(k+n)$ are approximately equal to $\mathbf{X}(k+n)\mathbf{w}(k)$ for $k \leq n \leq k+K$. If the value of $\mathbf{w}(k)$ causes the matrix

$$\hat{\mathbf{R}}_{\mathbf{y}\mathbf{y}}(k) = \frac{1}{K} \sum_{n=k}^{k+K} \mathbf{X}^*(k+n) \mathbf{w}^*(k) \mathbf{w}^T(k) \mathbf{X}(k+n) \quad (68)$$

to be nearly singular for reasonably large values of K , then the FER and FR algorithms exhibit slow convergence. Such a situation can occur if the short-term averaged power spectrum of $y(k)$ has deep nulls in its frequency response caused by the current filter coefficient settings, *e.g.*, one or more of the zeros of $W(z, k)$ lies near the unit circle in the z -plane at a frequency where the channel frequency response does not have a spectral peak. For this reason, care must be taken when selecting $\mathbf{w}(0)$ to avoid such initializations.

References

- [Ahn and Voltz 1996] S.-S. Ahn and P.J. Voltz, "Convergence of the delayed normalized LMS algorithm with decreasing step size, *IEEE Trans. Signal Processing*, vol. 44, pp. 3008-3016, Dec. 1996.
- [Amari *et al* 1996] S. Amari, A. Cichocki, and H.H. Yang, "A new learning algorithm for blind signal separation," *Adv. Neural Inform. Proc. Syst.* (Boston, MA: MIT Press, 1996), pp. 752-763.
- [Amari *et al* 1997a] S. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," *Proc. 1st IEEE Workshop Signal Processing Adv. Wireless Commun.*, Paris, France, pp. 101-104, Apr. 1997.
- [Amari *et al* 1997b] S. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang, "Novel on-line adaptive learning algorithms for blind deconvolution using the natural gradient approach," *Proc. 11th IFAC Symp. System Identification*, Kitakyushu City, Japan, vol. 3, pp. 1057-1062, July 1997.
- [Bellini and Rocca 1996] S. Bellini and F. Rocca, "Blind deconvolution: polyspectra or Bussgang techniques," in *Digital Communications*, E. Biglieri and G. Prati, eds. (North-Holland: Elsevier, 1986), pp. 251-263.
- [Benveniste and Goursat 1984] A. Benveniste and M. Goursat, "Blind equalizers," *IEEE Trans. Commun.*, vol. 32, pp. 871-883, Aug. 1984.
- [Bjarnason 1995] E. Bjarnason, "Analysis of the filtered-X LMS algorithm," *IEEE Trans. Speech Audio Processing*, vol. 3, pp. 504-514, Nov. 1995.
- [Ding *et al* 1994] Z. Ding, C.R. Johnson, Jr., and R.A. Kennedy, "Global convergence issues with linear blind adaptive equalizers," in *Blind Deconvolution*, S. Haykin, ed. (Englewood Cliffs, NJ: Prentice-Hall, 1994), pp. 60-120.
- [Douglas *et al* 1996] S.C. Douglas, A. Cichocki, and S. Amari, "Fast-convergence filtered regressor algorithms for blind equalisation," *Electron. Lett.*, vol. 32, pp. 2114-2115, 7th Nov. 1996.
- [Douglas *et al* 1997] S.C. Douglas, A. Cichocki, and S. Amari, "Quasi-Newton filtered-error and filtered-regressor algorithms for adaptive equalization and deconvolution," *Proc. 1st IEEE Workshop Signal Processing Adv. Wireless Commun.*, Paris, France, pp. 109-112, Apr. 1997.
- [Douglas *et al* 1999] S.C. Douglas, A. Cichocki, and S. Amari, "Self-whitening algorithms for adaptive equalization and deconvolution," *IEEE Trans. Signal Processing*, vol. 47, 1999 (in press).
- [Douglas and Meng 1994] S.C. Douglas and T.H.-Y. Meng, "Normalized data nonlinearities for LMS adaptation," *IEEE Trans. Signal Processing*, vol. 42, pp. 1352-1365, June 1994.
- [Douglas and Pan 1995] S.C. Douglas and W. Pan, "Exact expectation analysis of the LMS adaptive filter," *IEEE Trans. Signal Processing*, vol. 43, pp. 2863-2871, Dec. 1995.
- [Gay and Tavathia 1995] S. Gay and S. Tavathia, "The fast affine projection algorithm," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Detroit, MI, vol. 5, pp. 3023-3026, May 1995.
- [Godard 1980] D.N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. 28, pp. 1867-1875, Nov. 1980.

- [Haykin 1994] S. Haykin, ed., *Blind Deconvolution*, (Englewood Cliffs, NJ: Prentice-Hall, 1994).
- [Haykin 1996] S. Haykin, *Adaptive Filter Theory*, 3rd Ed. (Upper Saddle River, NJ: Prentice-Hall, 1996).
- [Ljung and Söderström 1983] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification* (Cambridge, MA: MIT Press, 1983).
- [Lucky 1966] R.W. Lucky, "Techniques for adaptive equalization of digital communication systems," *Bell Sys. Tech. J.*, vol. 45, pp. 255-286, Feb. 1966.
- [Kuo and Morgan 1996] S.M. Kuo and D.R. Morgan, *Active Noise Control Systems: Algorithms and DSP Implementations* (New York: Wiley, 1996).
- [Long *et al* 1989] G. Long, F. Ling, and J.G. Proakis, "The LMS algorithm with delayed coefficient adaptation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 1397-1405, Sept. 1989.
- [Long *et al* 1992] G. Long, F. Ling, and J.G. Proakis, "Corrections to 'The LMS algorithm with delayed coefficient adaptation,'" *IEEE Trans. Signal Processing*, vol. 40, pp. 230-232, Jan. 1992.
- [Mazo 1979] J.E. Mazo, "On the independence theory of equalizer convergence," *Bell Sys. Tech. J.*, vol. 58, pp. 963-993, May-June 1979.
- [Mboup *et al* 1994] M. Mboup, M. Bonnet, and N.J. Bershad, "LMS coupled adaptive prediction and system identification: a statistical model and transient analysis," *IEEE Trans. Signal Processing*, vol. 42, pp. 2607-2615, Oct. 1994.
- [Orgren *et al* 1991] A.C. Orgren, S. Dasgupta, C.E. Rohrs, and N.R. Malik, "Noise cancellation with improved residuals," *IEEE Trans. Signal Processing*, vol. 39, pp. 2629-2639, Dec. 1991.
- [Rupp and Frenzel 1994] M. Rupp and R. Frenzel, "The behavior of LMS and NLMS algorithms with delayed coefficient update in the presence of spherically invariant processes," *IEEE Trans. Signal Processing*, vol. 42, pp. 668-672, Mar. 1994.
- [Sato 1975] Y. Sato, "Two extensional applications of the zero-forcing equalization method," *IEEE Trans. Commun.*, vol. 23, pp. 684-687, June 1975.
- [Sethares and Johnson 1991] W.A. Sethares and C.R. Johnson, Jr., "A comparison of two quantized state adaptive algorithms," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 138-143, Jan. 1989.
- [Shynk *et al* 1991] J.J. Shynk, R.P. Gooch, G. Krishnamurthy, and C.K. Chan, "A comparative performance study of several blind equalization algorithms," *Proc. SPIE Conf. Adaptive Signal Processing*, San Diego, CA, vol. 1565, pp. 102-117, July 1991.
- [Slock 1992] D.T.M. Slock, "Underdetermined growing and sliding window fast transversal filter RLS algorithms," *Proc. Sixth European Signal Proc. Conf.*, Brussels, Belgium, vol. 2, pp. 1169-1172, Aug. 1992.
- [Slock 1993] D.T.M. Slock, "On the convergence behavior of the LMS and the normalized LMS algorithms," *IEEE Trans. Signal Processing*, vol. 41, pp. 2811-2825 Sept. 1993.

- [Slock and Kailath 1991] D.T.M. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least squares adaptive filtering," *IEEE Trans. Signal Processing*, vol. 39, pp. 92-114, Jan. 1991.
- [Tanaka *et al* 1995] M. Tanaka, Y. Kaneda, S. Makino, and J. Kojima, "Fast projection algorithm and its step size control," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Detroit, MI, vol. 2, pp. 945-948, May 1995.
- [Treichler and Agee 1983] J.R. Treichler and B.G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 349-372, Apr. 1983.
- [Wan 1996] E.A. Wan, "Adjoint LMS: An efficient alternative to the filtered-X LMS and multiple error LMS algorithms," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Atlanta, GA, vol. 3, pp. 1842-1845, May 1996.
- [Widrow and Stearns 1985] B. Widrow and S.D. Stearns, *Adaptive Signal Processing* (Englewood Cliffs, NJ: Prentice-Hall, 1985).
- [Widrow and Walach 1996] B. Widrow and E. Walach, *Adaptive Inverse Control* (Upper Saddle River, NJ: Prentice-Hall PTR, 1996).