Iterate-Averaging Sign Algorithms for Adaptive Filtering With Applications to Blind Multiuser Detection

G. George Yin, *Fellow, IEEE*, Vikram Krishnamurthy, *Senior Member, IEEE*, and Cristina Ion

***Abstract*—Motivated by the recent developments on iterate averaging of recursive stochastic approximation algorithms and asymptotic analysis of sign-error algorithms for adaptive filtering, this work develops two-stage sign algorithms for adaptive filtering. The proposed algorithms are based on constructions of a sequence of estimates using large step sizes followed by iterate averaging. Our main effort is devoted to improving the performance of the algorithms by establishing asymptotic normality of a suitably scaled sequence of the estimation errors. The asymptotic covariance is calculated and shown to be the smallest possible. Hence, the asymptotic efficiency or asymptotic optimality is obtained. Then variants of the algorithm including sign-regressor procedures and constant-step algorithms are studied. The minimal window width of averaging is also dealt with. Finally, iterate-averaging algorithms for blind multiuser detection in direct sequence/code-division multiple-access (DS/CDMA) systems are proposed and developed, and numerical examples are examined.**

***Index Terms*—Adaptive filtering, asymptotic efficiency, iterate average, rate of convergence, sign algorithm.**

I. INTRODUCTION

MOTIVATED by the ingenious procedure of *iterate averaging* for accelerating convergence rates of stochastic approximation algorithms, proposed independently by Polyak [28] and Ruppert [32], this work is devoted to adaptive filtering algorithms using sign operators. We show that the convergence rates of such adaptive filtering algorithms can also be accelerated by iterate averaging and that the resulting algorithms have optimal convergence rates. Furthermore, we develop iterate-averaging algorithms for blind multiuser detection in direct sequence/code-division multiple-access (DS/CDMA) systems and provide promising numerical results.

Manuscript received December 1, 2000; revised November 18, 2002. The work of G. G. Yin was supported in part by the National Science Foundation under Grants DMS-9877090 and DMS-9971608. The work of V. Krishnamurthy Research was supported in part by the ARC Special Research Center for Ultra-Broadband Information Networks (CUBIN), University of Melbourne, Australia. The work of C. Ion was supported in part by the National Science Foundation under Grant DMS-9877090 and in part by Wayne State University.

G. G. Yin and C. Ion are with the Department of Mathematics, Wayne State University, Detroit, MI 48202 USA (e-mail: gyin@math.wayne.edu; cion@math.wayne.edu).

V. Krishnamurthy is with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver BC V6T 1Z4, Canada (e-mail: vikramk@ece.ubc.ca) and the University of Melbourne, Melbourne, Australia.

Communicated by J. A. O'Sullivan, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2002.808100

Owing to its importance and various applications in adaptive signal processing and learning, adaptive filtering algorithms have received much attention; see [1], [2], [9]–[12], [25], [33], [35], [37], [38], among others. Suppose that $\varphi_n \in \mathbb{R}^r$ and $y_n \in \mathbb{R}$ are sequences of measured output and reference signals, respectively. Assuming the sequence $\{(y_n, \varphi_n)\}$ is stationary, by adjusting the system parameter θ, adaptive filtering algorithms aim to make the weighted output $\theta' \varphi_n$ match the reference signal y_n as well as possible in the sense that a cost function is minimized. If a mean square error cost

$$L(\theta) = \mathbf{E}|y_n - \theta'\varphi_n|^2/2 = \mathbf{E}|y_1 - \theta'\varphi_1|^2/2$$

is used, the gradient of $L(\theta)$ is given by

$$L_\theta(\theta) = -\mathbf{E}\varphi_1(y_1 - \theta'\varphi_1)$$

and the recursive algorithm is of the form

$$\theta_{n+1} = \theta_n + a_n \varphi_n (y_n - \theta_n' \varphi_n) \tag{1}$$

where $0 \leq a_n \to 0$ as $n \to 0$ and $\sum_n a_n = \infty$. If the cost function is $L(\theta) = \mathbf{E}|y_1 - \theta'\varphi_1|$, the gradient becomes

$$L_\theta(\theta) = -\mathbf{E}(\varphi_1 \mathrm{sgn}\,(y_1 - \theta'\varphi_1)) \stackrel{\text{def}}{=} -f(\theta)$$

and a recursive algorithm takes the form

$$\theta_{n+1} = \theta_n + a_n \varphi_n \mathrm{sgn}\,(y_n - \theta_n'\varphi_n)\,, \tag{2}$$

where for any $y \in \mathbb{R}$ $\mathrm{sgn}(y) = 1_{\{y>0\}} - 1_{\{y<0\}}$ (1_A is the indicator of A). Algorithm (1) is commonly referred to as a least mean square (LMS) algorithm, whereas (2) is called a sign-error algorithm. Compared with (1), algorithm (2) has reduced complexity. Because of the use of the sign operator, the algorithms are easily implementable and multiplications in (2) can be replaced by simple bit shifts. As a result, it becomes appealing in various applications; see [9], [10], [12], [35] and the references therein. However, for each n, as a function of θ, $f_n(\theta) = \varphi_n \mathrm{sgn}(y_n - \theta'\varphi_n)$ is not continuous. Thus, the analysis of such an algorithm is more difficult than that of (1). Much effort has been devoted to the improvement of sufficient conditions for convergence of such algorithms. Recently, in [5], by treating an algorithm with randomly generated truncation bounds, we obtained that the recursive algorithm converges with probability 1 (w.p. 1) by assuming only stationarity and finite second moments of the signals, which is close to the minimal requirement needed. In addition, we also examined rate of convergence of the algorithm by weak convergence methods. A crucial observation is that although the functions $f_n(\theta)$ are not continuous in θ, $L_\theta(\theta)$ can be a smooth function thanks to the smoothing effect provided by taking expectation. Note that Gaussian approximation and central limit results for adaptive signal process algorithms have also been considered, for example, in [2], [33] among others. Notably, Markovian-type processes are treated in [2] and stochastic averaging ideas are used in [33].

In this paper, in addition to (1) and (2), an algorithm known as a sign-regressor algorithm used frequently in applications, will also be considered. In this case, in lieu of (2), one uses sign operator only for the regressor φ_n by taking the sign of φ_n componentwise. Experience with numerical examples shows that the sign-regressor algorithm often outperforms (2). The rationale for using sign-regressor algorithms is to take advantages of both LMS and sign-error algorithms and to have the performance close to that of (1) with less complexity. Devoted to (2) and its variations such as sign-regressor algorithms and algorithms with constant step size, in comparison to the recent study on the sign-error algorithms, we shift gear and emphasize the asymptotic efficiency issues. Our plan is as follows. We first develop the iterate-averaging sign-error algorithms. Then we proceed with the analysis of sign-regressor algorithms without providing verbatim proofs since they can be carried out similarly to those of sign-error algorithms with weaker conditions and simpler proofs. An alternative method for analyzing the averaging algorithms is along the line of strong approximation. We refer the reader to [26], [27] for related references and further study.

Inspired by the recent work on iterate averaging of stochastic approximation algorithms [28], [32], [21], we propose several iterate-averaging algorithms for sign adaptive filtering algorithms. The motivation behind the averaging approach can be traced back to the work of Chung [7] and many subsequent papers on adaptive stochastic approximation. Nevertheless, it has been shown that the iterate-averaging approach leads to asymptotic optimality (the best scaling factor and the minimal variance) and has advantages for various applications. First, its initial approximation uses slowly varying step sizes larger than $O(1/n)$ to get rough estimates, which enables the iterates to get to a neighborhood of the minimizer θ_* faster than that of a small step-size procedure. Then, by averaging the iterates, the resulting sample path possesses the minimal variance. Our effort in what follows is to prove that the iterate-averaging adaptive filtering algorithms are asymptotically optimal.

The rest of the paper is arranged as follows. Section II is devoted to the iterate-averaging of sign-error algorithm. It provides the convergence of algorithm (2) and obtains the convergence of $\overline{\theta}_n$. The asymptotic efficiency issue is then studied. Section III proceeds with the ramifications and variations of the iterate-averaging approach. We study averaging in sign-regressor algorithms, algorithms with constant step size, and minimal window width of averaging. To demonstrate the performance of the algorithms, a case study of blind interference suppression in DS/CDMA spread-spectrum telecommunication systems is provided in Section IV. Section V gives further remarks. Finally, an appendix containing the proofs of some technical results, concludes the paper. Throughout the paper, we use z' to denote the transpose of $z \in \mathbb{R}^{\ell\times r}$ for $\ell, r \geq 1$, and use $|z|$ to denote the norm of z. For notational simplicity, C denotes a generic positive constant whose values may vary for different usage. For a square matrix B, by $B > 0$ we mean that it is positive definite.

II. Iterate-Averaging Sign-Error Algorithms

A. Convergence of Sign-Error Algorithm

Consider the two-stage sign-error algorithm

$$\begin{aligned}\theta_{n+1} &= \theta_n + \frac{1}{n^\gamma}\varphi_n \operatorname{sgn}(y_n - \theta_n'\varphi_n), \qquad \frac{2}{3} < \gamma < 1\\ \overline{\theta}_{n+1} &= \overline{\theta}_n - \frac{1}{n+1}\overline{\theta}_n + \frac{1}{n+1}\theta_{n+1}. \end{aligned} \tag{3}$$

In what follows, we use $\mathbf{E}_n$ to denote the conditional expectation with respect to $\mathcal{F}_n$, the σ-algebra generated by $\{\theta_1, y_k, \varphi_k,\ k < n\}$. Define

$$f_n(\theta) \stackrel{\text{def}}{=} \varphi_n \operatorname{sgn}(y_n - \theta'\varphi_n)$$

and

$$\widetilde{f}_n(\theta) \stackrel{\text{def}}{=} \mathbf{E}_n f_n(\theta).$$

To proceed, we state the conditions needed.

(**A**) $L(\theta)$ has a unique minimizer, denoted by θ_*. $\{(y_n, \varphi_n)\}$ is a stationary sequence with

$$\mathbf{E}\begin{pmatrix} y_1\\ \varphi_1\end{pmatrix}(y_1\ \varphi_1') = \widetilde{R} > 0.$$

For each n, $\mathbf{E}_n f_n(\theta) f_n'(\theta)$ is continuous; there is an $A_n \in \mathbb{R}^{r\times r}$ such that for each θ

$$\widetilde{f}_n(\theta) - \widetilde{f}_n(\theta_*) = A_n(\theta - \theta_*) + O(|\theta-\theta_*|^2) \text{ w.p. 1} \quad \text{and } \mathbf{E}A_n = D_{\rm SE} \tag{4}$$

where $D_{\rm SE}$ is Hurwitz, i.e., all of its eigenvalues have negative real parts. Either $\{(y_n, \varphi_n)\}$ is a martingale difference sequence satisfying $\mathbf{E}|y_1|^{4+\delta} < \infty$, $\mathbf{E}|\varphi_1|^{4+\delta} < \infty$ for some $\delta > 0$, or it is a bounded uniformly mixing sequence such that there is a deterministic sequence of real numbers $\{\eta(n)\}$ satisfying $\eta(n) \geq 0$ for each n, $\sum_k \eta^{1/2}(k) < \infty$, and for each $j \geq n$ and some $C > 0$

$$|\mathbf{E}_n A_j - D_{\rm SE}| \leq C\eta^{1/2}(j-n)$$

and

$$|\mathbf{E}_n f_j(\theta_*)| \leq C\eta^{1/2}(j-n). \tag{5}$$

Remark 2.1: We have collected the conditions needed for both convergence and rate of convergence in (**A**). As far as convergence alone is concerned, not all aspects of the assumptions are needed.

By (4), $\widetilde{f}_n(\theta)$ is locally (near θ_*) linearizable. To see this, suppose that the joint density of y_n and φ_n exists. Denote by $F_{n,s|\varphi}(\cdot)$ the conditional distribution with respect to $\{y_j, \varphi_j, j < n, \varphi_n\}$ and by $g_{n,s|\varphi}(\cdot)$ the corresponding conditional density. It can be seen that

$$\begin{aligned}\widetilde{f}_n(\theta) &= -(\partial/\partial\theta)(\mathbf{E}_n|Y_n - \theta'\varphi_n|)\\ &= \mathbf{E}_n(\varphi_n(2F_{n,s|\varphi}(\theta'\varphi_n) - 1)).\end{aligned}$$

Moreover, for any θ, the partial derivatives of $\widetilde{f}_n(\theta)$ exist. Using θ^i to denote the ith component of θ, and using $\widetilde{f}_{n,i}(\theta)$ and $\varphi_{n,i}$ to denote the i th components of $\widetilde{f}_n(\theta)$ and φ_n, respectively, we have

$$(\partial/\partial\theta^j)\widetilde{f}_{n,i}(\theta) = -(\partial^2/\partial\theta^i\partial\theta^j)\left(\mathbf{E}_n|y_n - \theta'\varphi_n|\right) = 2\mathbf{E}_n\left(\varphi_{n,i}\varphi_{n,j}g_{n,s|\varphi}(\theta'\varphi_n)\right)$$

see also [12, p. 195] and [5]. Thus, a sufficient condition for (4) is that $g_{n,s|\varphi}(\cdot)$ is differentiable with bounded derivatives. Moreover, if $\{y_n\}$ and $\{\varphi_n\}$ are independent and identically distributed (i.i.d.) random variables or martingale difference sequences, $A_n = D_{\rm SE} = \mathbf{E}(\varphi_n\varphi_n' g(\theta'\varphi_n))$ where $g(\cdot)$ is the density of Y_1.

Condition (5) requires the signals $\{(y_j,\varphi_j),\ j < k\}$ and $\{(y_j,\varphi_j), j \geq k+n\}$ having decreasing dependence as $n\to\infty$. The $\eta(n)$ is referred to as a uniform mixing measure in [8, p. 348]; see also [3, p. 200]. The inequality is reminiscent of the well-known mixing inequalities (see [3], [8], [19]).

The noise sequences covered by the conditions include bounded and uniform mixing sequences, or uncorrelated signals with finite $(4+\delta)$th moment, or combination of them. Note that for uncorrelated signals, (5) is trivially satisfied and the conditional expectation is replaced by expectation. The conditions for an $\mathrm{MA}(p)$ moving average process of order p driven by a martingale difference noise are similar to those of the martingale difference noise; we need only place the conditions on the driving noise instead of on y_n and φ_n (The analysis can be carried out as in [38].) If the sequence $\{(y_n,\varphi_n)\}$ is bounded and uniform mixing with mixing rate $\eta(n)$ (see [19, p. 82] and [8, p. 349]), then so are $\{A_n\}$ and $\{f_n(\theta)\}$. By the stationarity, $\mathbf{E}f_j(\theta_*) = f(\theta_*) = 0$, and it follows from (5) that if $j<k$

$$\begin{aligned}|\mathbf{E}f_j(\theta_*)f_k'(\theta_*)| &= |\mathbf{E}f_j(\theta_*)\left[\mathbf{E}_j f_k'(\theta_*)\right]| \\ &\leq \mathbf{E}^{1/2}|f_j(\theta_*)|^2\mathbf{E}^{1/2}\left|\mathbf{E}_j f_k'(\theta_*)\right|^2 \\ &\leq C\eta^{1/2}(k-j) \\ |\mathbf{E}A_jA_k' - D_{\rm SE}D_{\rm SE}'| &= |\mathbf{E}A_jA_k' - \mathbf{E}a_j\mathbf{E}a_k| \\ &\leq C\eta^{1/2}(k-j)\end{aligned}$$

and

$$|\mathbf{E}(A_j - D_{\rm SE})(A_k - D_{\rm SE}')'| \leq C\eta^{1/2}(k-j). \tag{6}$$

Note that the bounded mixing signal is not restrictive. In practice, one often wants to avoid excessively large values of the observation. Although modeling at large values often follows from traditional setup (such as Gaussian assumptions), it is undesirable for single observation to have significant effect on the iterations. Thus, one often uses a robust algorithm. For the sign-error algorithm that we are interested in, we can use

$$\theta_{n+1} = \theta_n + \frac{1}{n^\gamma}\Xi\left(\varphi_n \mathrm{sgn}\left(y_n - \theta_n'\varphi_n\right)\right) \tag{7}$$

where $\Xi(x) = (\Xi_1(x_1),\ldots,\Xi(x_r))'$ for a vector $x\in\mathbb{R}^r$ such that for $i\leq r$, $\Xi_i(\cdot)$ are bounded real-valued functions on the real line that are nondecreasing and that satisfy $\Xi_i(0) = 0$, $\Xi_i(x_i) = -\Xi_i(-x_i)$, and $\Xi_i(x_i)/x_i \to 0$ as $x_i\to\infty$. For further discussions on the use of such functions and robust algorithms, see [29] (also [22, Sec. 1.3.4, p. 22]). For the sign-error algorithm, due to the boundedness of $\mathrm{sgn}(y_n - \theta_n'\varphi_n)$, the use of the function $\Xi(\cdot)$ is equivalent to the truncation of φ_n. However, for notational simplicity, we choose to use the bounded mixing condition here. Moreover, an alternative procedure projects the iterates into a bounded region (e.g., a hyperrectangle); see [22] for more discussion.

Theorem 2.2: Assume (**A**) and θ_* is the global asymptotic stable point of the ordinary differential equation (ODE) $\dot{\theta} = f(\theta)$, where $f(\theta)$ is an average of $f_n(\theta)$. Then $\theta_n\to\theta_*$ w.p. 1, and $\overline{\theta}_n\to\theta_*$ w.p. 1.

Remark 2.3: In lieu of an algorithm with expanding random truncation bounds as in [5], we examine the algorithms directly. Using the treatment of stochastic approximation algorithms of [22], the proof of convergence is converted to the verification of a recurrence condition by using [22, Theorem 7.1, p. 163]. In fact, we need only verify that the recurrence condition, namely, "for each $0<\rho<1$, let there be a compact set R_ρ such that $\theta_n\in R_\rho$ infinitely often (i.o.) with probability at least ρ" is verified, then as in the argument of [22, p. 164], $\limsup_n|\theta_n|<\infty$ w.p. 1. As a result, using the ODE method, a sequence of piecewise-constant interpolation $\{\theta^n(\cdot)\}$ of the iterates is uniformly bounded and equicontinuous in the extended sense (see [22, p. 73]). for a definition). By virtue of the Ascoli–Arzelá theorem, we obtain that any convergent subsequence has a limit $\theta(\cdot)$ satisfying $\dot{\theta} = f(\theta)$. A stability argument then implies $\theta_n\to\theta_*$ w.p. 1. Therefore, only the recurrence needs to be verified. By [22, Theorem 7.2, p. 164], a sufficient condition that guarantees the recurrence is: $\{\theta_n\}$ is bounded in probability. That is, for any $\varepsilon>0$, there is a $C_\varepsilon>0$ such that $\sup_n P(|\theta_n|\geq C_\varepsilon)\leq\varepsilon$. Since by Chebyshev's inequality

$$P(|\theta_n|\geq C_\varepsilon)\leq \mathbf{E}|\theta_n|^2/C_\varepsilon^2$$

which can be made $\leq\varepsilon$ if $\sup_n\mathbf{E}|\theta_n|^2<\infty$, and $1/C_\varepsilon^2\leq\varepsilon$ (or $C_\varepsilon\geq 1/\sqrt{\varepsilon}$), which can be established via a Liapunov function argument. Since we will prove a result with a sharper bound on $\theta_n - \theta_*$ in Theorem 2.4 using similar techniques, we omit the details here.

B. Asymptotic Efficiency

This subsection is devoted to the asymptotic efficiency of the sign algorithm. As was mentioned, the heart of the problem is to show that $\sqrt{n}(\overline{\theta}_n - \theta_*)$ is asymptotically normal with the optimal covariance matrix. In fact, we obtain a more interesting functional invariance theorem.

Define $\widetilde{\theta} = \theta - \theta_*$ and $\widetilde{\theta}_n = \theta_n - \theta_*$. Then (3) can be rewritten as

$$\widetilde{\theta}_{n+1} = \widetilde{\theta}_n + \frac{1}{n^\gamma}\left(\varphi_n\mathrm{sgn}\left(y_n - \theta_n'\varphi_n\right)\right). \tag{8}$$

The proof of the following bounds via Liapunov theory is included in the Appendix.

Theorem 2.4: Under (**A**), for sufficiently large n, $\mathbf{E}|\widetilde{\theta}_n|^2 = o(1/n^\gamma)$, and the bounds hold uniform in n.

Much effort has been devoted to improving the rate of convergence and to reduce the asymptotic variance in the adaptive estimation problems. Consider (2) with $a_n = O(1/n^\gamma)$, $2/3<\gamma\leq 1$. Under suitable conditions, it can be shown that

$n^{\gamma/2}(\theta_n - \theta_*)$ converges in distribution to a normal random variable $N(0, S)$ as $n \to \infty$. It is clear that among the possible γ's with $2/3 < \gamma \le 1$, $\gamma = 1$ gives us the best scaling factor. Since in evaluating rates of convergence, one uses the scaling factor together with the asymptotic covariance matrix S, for different algorithms with $a_n = O(1/n)$, we wish to find the one with minimal variance. The idea outlined in [7] is to consider (2) with $a_n = \Gamma/n$, where Γ is a (matrix-valued) parameter. It follows that the asymptotic covariance $S = S(\Gamma)$ is a smooth function of Γ. Minimizing $S(\Gamma)$ with respect to (w.r.t.) Γ leads to the choice $\Gamma^* = -D_{\rm SE}^{-1}$ and the optimal variance $S_{\rm SE}^* = D_{\rm SE}^{-1}\Sigma_{\rm SE}(D_{\rm SE}^{-1})'$, where $\Sigma_{\rm SE}$ is the noise covariance and $D_{\rm SE}$ is defined in (4). Although $S_{\rm SE}^*$ is explicitly given, $D_{\rm SE}$ is virtually unknown. To circumvent such a difficulty, researchers developed step-size-adaptation algorithms. In the context of adaptive filtering, this amounts to constructing another sequence $\widehat{A}_n$, estimates of $D_{\rm SE}^{-1}$, on top of the adaptive filtering estimate. Then use a sequence of matrix-valued step-size $a_n = \widehat{A}_n/n$ in the actual estimation, denoted by $\widehat{\theta}_n$. It can be shown that such a recursive least squares (RLS) type algorithm is convergent and $\sqrt{n}(\widehat{\theta}_n - \theta_*) \sim N(0, S_{\rm SE}^*)$. Although optimality is obtained, the RLS algorithm has computational complexity $O(r^2)$ compared to the order $O(r)$ complexity of a scalar step size stochastic approximation algorithm. A new approach, initiated in the late 1980s [28], provides a much better alternative (see also a scalar version of the algorithm in [32]). Instead of adaptively generating the matrix-valued estimates, a simple iterate-averaging approach is used leading to the desired asymptotic optimality. The corresponding problems for adaptive filtering under quadratic cost functions were treated in [38] among others. We will show that the averaging approach for the sign algorithms of adaptive filtering also leads to asymptotic optimality. Rather than dealing with the iterates as in [38], we work with suitably interpolated sequences. As a preparation, we first derive an asymptotic equivalence. Then we proceed with an invariance theorem. In order not to disrupt the flow of presentation, we relegate their proofs to the Appendix.

Using (**A**) and

$$f_n(\theta) = f_n(\theta_*) + [f_n(\theta) - f_n(\theta_*)]$$

rewrite the first equation in (3) as follows:

$$\widetilde{\theta}_{n+1} = \widetilde{\theta}_n + \frac{1}{n^\gamma} f_n(\theta_*) + \frac{1}{n^\gamma} D_{\rm SE}\widetilde{\theta}_n + \frac{1}{n^\gamma}(A_n - D_{\rm SE})\widetilde{\theta}_n + \frac{1}{n^\gamma}\delta M_n + \frac{1}{n^\gamma} O(|\widetilde{\theta}_n|^2) \quad (9)$$

where $D_{\rm SE}$ is defined in (4) and

$$\delta M_n = [f_n(\theta_n) - f_n(\theta_*)] - [\widetilde{f}_n(\theta_n) - \widetilde{f}_n(\theta_*)]. \quad (10)$$

Note that $\mathbf{E}_n \delta M_n = 0$ and hence $\{\delta M_n\}$ is a martingale difference sequence. Define

$$\Psi(n|j) = \begin{cases} \prod_{l=j+1}^{n} \left(\frac{I + D_{\rm SE}}{l^\gamma}\right), & \text{if } n \ge j+1 \\ I, & \text{if } n = j. \end{cases}$$

It follows from (9) that for any integer $m > 0$ and $n > m$

$$\begin{aligned} \widetilde{\theta}_{n+1} =& \Psi(n|m-1)\widetilde{\theta}_m + \sum_{j=m}^{n} \frac{1}{j^\gamma}\Psi(n|j) f_j(\theta_*) \\ &+ \sum_{j=m}^{n} \frac{1}{j^\gamma}\Psi(n|j)(A_j - D_{\rm SE})\widetilde{\theta}_j \\ &+ \sum_{j=m}^{n} \frac{1}{j^\gamma}\Psi(n|j) O(|\widetilde{\theta}_j|^2) \\ &+ \sum_{j=m}^{n} \frac{1}{j^\gamma}\Psi(n|j)\delta M_j. \end{aligned} \quad (11)$$

Next, consider a continuous-time interpolation

$$w_{n+1}(t) = \frac{\lfloor nt \rfloor + 1}{\sqrt{n+1}}(\overline{\theta}_{\lfloor nt \rfloor + 1} - \theta_*), \qquad \text{for } t \in [0, 1] \quad (12)$$

where $\lfloor z \rfloor$ denotes the integer part of z. Then we have

$$w_{n+1}(t) = \sum_{i=0}^{5} e_n^i(t)$$

where

$$\begin{aligned} e_n^0(t) =& \frac{1}{\sqrt{n+1}} \sum_{k=m}^{\lfloor nt \rfloor} \sum_{j=m}^{k} \frac{1}{j^\gamma}\Psi(k|j) f_j(\theta_*) \\ e_n^1(t) =& \frac{1}{\sqrt{n+1}} \sum_{k=1}^{m-1} \widetilde{\theta}_k \\ e_n^2(t) =& \frac{1}{\sqrt{n+1}} \sum_{k=m}^{\lfloor nt \rfloor} \Psi(k|m-1)\widetilde{\theta}_m \\ e_n^3(t) =& \frac{1}{\sqrt{n+1}} \sum_{k=m}^{\lfloor nt \rfloor} \sum_{j=m}^{k} \frac{1}{j^\gamma}\Psi(k|j) O(|\widetilde{\theta}_j|^2), \\ e_n^4(t) =& \frac{1}{\sqrt{n+1}} \sum_{k=m}^{\lfloor nt \rfloor} \sum_{j=m}^{k} \frac{1}{j^\gamma}\Psi(k|j)(A_j - D_{\rm SE})\widetilde{\theta}_j \\ e_n^5(t) =& \frac{1}{\sqrt{n+1}} \sum_{k=m}^{\lfloor nt \rfloor} \sum_{j=m}^{k} \frac{1}{j^\gamma}\Psi(k|j)\delta M_j. \end{aligned}$$

Since $D_{\rm SE}$ is nonsingular

$$\Psi(k|j) = \Psi(k-1|j) + \frac{D_{\rm SE}}{k^\gamma}\Psi(k-1|j)$$

and

$$D_{\rm SE}^{-1}\Psi(n|j) = D_{\rm SE}^{-1} + \sum_{l=j+1}^{n} \frac{1}{l^\gamma}\Psi(l-1|j).$$

Thus, for each n

$$\frac{1}{j^\gamma}\sum_{k=j+1}^{n} \Psi(k-1|j) = -D_{\rm SE}^{-1} + \Phi(n|j)$$

where

$$\Phi(n|j) = D_{\rm SE}^{-1}\Psi(n|j) + \sum_{k=j+1}^{n} \left(\frac{1}{j^\gamma} - \frac{1}{k^\gamma}\right)\Psi(k-1|j). \quad (13)$$

Lemma 2.5: The following estimates hold:

$$\sum_{j=m}^{k} \frac{1}{j^\gamma}|\Psi(k|j)| \le C < \infty, \qquad \text{for each } m \text{ and } k \ge m$$

$$\frac{1}{n}\sum_{j=1}^{n} |\Phi(n|j)| \to 0, \qquad \text{as } n \to \infty. \quad (14)$$

The proof of the lemma is in the Appendix. To proceed, choose $m = m(n)$ such that

$$m(n) \to \infty \text{ but } \frac{m(n)}{\sqrt{n}} \to 0, \qquad \text{as } n \to \infty. \tag{15}$$

We further derive the following lemma; its proof is in the Appendix.

Lemma 2.6:

$$w_{n+1}(t) = -\frac{D_{\rm SE}^{-1}}{\sqrt{n+1}} \sum_{j=1}^{\lfloor nt \rfloor} f_j(\theta_*) + o(1)$$

as $n \to \infty$, where $o(1) \to 0$ in probability uniformly in t.

We proceed to obtain a functional central limit theorem or invariance theorem. The proof is standard; see, for example, [3], [8], [22]. In fact, under (**A**)

$$\frac{1}{\sqrt{n+1}} \sum_{j=1}^{\lfloor nt \rfloor} f_j(\theta_*) \text{ converges weakly to a Brownian motion} \tag{16}$$

with covariance $\Sigma_{\rm SE} t$, where

$$\Sigma_{\rm SE} = \mathbf{E} f_1(\theta_*) f_1'(\theta_*) + \sum_{j=2}^{\infty} \mathbf{E} f_1(\theta_*) f_j'(\theta_*) + \sum_{j=2}^{\infty} \mathbf{E} f_j(\theta_*) f_1'(\theta_*). \tag{17}$$

The proof of the following theorem is also in the Appendix.

Theorem 2.7: Under (**A**), $\{w_{n+1}(\cdot)\}$ defined in (12) is tight in $D^r[0, 1]$, and it converges weakly to a Brownian motion with covariance $S_{\rm SE}^* t$, where $S_{\rm SE}^* = D_{\rm SE}^{-1} \Sigma_{\rm SE} (D_{\rm SE}^{-1})'$ with $D_{\rm SE}$ and $\Sigma_{\rm SE}$ defined by (4) and (17), respectively.

III. Iterate-Averaging Sign-Regressor Algorithms

A. Sign-Regressor Algorithm With Iterate Averaging

In lieu of (2), by taking sign componentwise, we obtain the so-called sign-regressor algorithm. In this section, we consider an iterate-averaging sign-regressor algorithm

$$\begin{aligned} \theta_{n+1} &= \theta_n + \frac{1}{n^\gamma} \mathrm{Sgn}(\varphi_n)(y_n - \varphi_n' \theta_n), \qquad \frac{2}{3} < \gamma < 1 \\ \overline{\theta}_{n+1} &= \overline{\theta}_n - \frac{1}{n+1} \overline{\theta}_n + \frac{1}{n+1} \theta_{n+1} \end{aligned} \tag{18}$$

where $\mathrm{Sgn}(\cdot)$ denotes $\mathrm{Sgn}(\varphi) = (\mathrm{sgn}(\varphi_1), \ldots, \mathrm{sgn}(\varphi_r))'$ for $\varphi \in \mathbb{R}^r$. To carry out the asymptotic analysis, we need the following conditions.

(**B**) $\{(y_n, \varphi_n)\}$ is stationary with $\mathbf{E}\mathrm{Sgn}(\varphi_1)\varphi_1' = D_{\rm SR}$, and $\mathbf{E}\mathrm{Sgn}(\varphi_1) y_1 = b$, where $-D_{\rm SR}$ is Hurwitz. Either $\{(y_n, \varphi_n)\}$ is a martingale difference sequence satisfying $\mathbf{E}|y_1|^{4+\delta} < \infty$, $\mathbf{E}|\varphi_1|^{4+\delta} < \infty$ for some $\delta > 0$, or it is bounded and uniformly mixing with mixing rate $\eta(n)$ satisfying $\sum_j \eta^{1/2}(j) < \infty$.

Remark 3.1: It is easily seen that the conditions are much weaker than (**A**) used before. The sequence $\{(y_n, \varphi_n)\}$ is stationary, so are $\{\mathrm{Sgn}(\varphi_n)\}$, $\{y_n \mathrm{Sgn}(\varphi_n)\}$, and $\{\mathrm{Sgn}(\varphi_n)\varphi_n'\}$. Moreover, $\{\mathrm{Sgn}(\varphi_n)\}$ is bounded by 1 w.p. 1. Since we only take the sign of φ_n componentwise, the nonsmoothness of $f_n(\theta)$ in the sign-error algorithm (3) is removed. As a result, the analysis is simpler. In addition, conditions for an $\mathrm{MA}(p)$ process driven by a martingale difference noise can also be provided (see Remark 2.1).

Proceeding as in Section II, under (**B**), we can verify the recurrence condition, and show that θ_n defined in (18) converges w.p. 1. Define $t_n = \sum_{i=1}^n 1/n^\gamma$, take a continuous-time interpolation of the iterates $\theta^n(t) = \theta_{n+i}$, for $t \in [t_{n+i} - t_n, t_{n+i+1} - t_n)$, $i \geq 0$, and let $m(t)$ be the unique n such that $t_n \leq t < t_{n+1}$. Using the ODE methods [22, Chs. 5 and 6], we can show that $\{\theta^n(\cdot)\}$ is uniformly bounded and equicontinuous in the extended sense. Then the Ascoli–Arzelà theorem implies that any convergent subsequence has a limit $\theta(\cdot)$ satisfying the limit ODE

$$\dot{\theta}(t) = b - D_{\rm SR} \theta(t) \tag{19}$$

with the unique stationary point $\theta_* = D_{\rm SR}^{-1} b$. Moreover, (**B**) implies that (19) is asymptotically stable. We then obtain the following result.

Theorem 3.2: Under condition (**B**), $\{\theta_n\}$ defined by (18) converges w.p. 1 to θ_* .

Remark 3.3: It is interesting to compare (18) with the algorithm (1). Under stationarity of the signals and assuming $\mathbf{E}\varphi_1\varphi_1' > 0$, the limit of the ODE for (1) and the unique minimizer θ_* of the quadratic cost functions are

$$\dot{\theta}(t) = \mathbf{E} y_1 \varphi_1 - [\mathbf{E}\varphi_1\varphi_1']\theta(t), \; \theta_* = [\mathbf{E}\varphi_1\varphi_1']^{-1} \mathbf{E} y_1 \varphi_1 \tag{20}$$

respectively. They are similar to that of (19). As a result, the two algorithms have similar asymptotic behavior. The difference is that $\mathbf{E}\varphi_1\varphi_1'$ is symmetric, whereas in (18), the symmetry is lost. We only assume the eigenvalues of $\mathbf{E}\mathrm{Sgn}(\varphi_n)\varphi_n'$ have positive real parts. To some extent, the sign-regressor algorithm is one "between" the LMS algorithm and the sign-error algorithm. As a result, its performance is close to LMS algorithm and its complexity is similar to the sign-error algorithm.

To proceed, define

$$\xi_n = \mathrm{Sgn}(\varphi_n) y_n - \mathrm{Sgn}(\varphi_n)\varphi_n' \theta_*.$$

Denoting $\widetilde{\theta}_n = \theta_n - \theta_*$ as before, we have

$$\widetilde{\theta}_{n+1} = \widetilde{\theta}_n + \frac{1}{n^\gamma}\xi_n - \frac{1}{n^\gamma} D_{\rm SR} \widetilde{\theta}_n - \frac{1}{n^\gamma}[\mathrm{Sgn}(\varphi_n)\varphi_n' - D_{\rm SR}]\widetilde{\theta}_n.$$

Define

$$\widetilde{\Psi}(n|j) = \begin{cases} \prod_{l=j+1}^{n} (\frac{I - D_{\rm SR}}{l^\gamma}), & \text{if } n \geq j+1 \\ I, & \text{if } n = j. \end{cases}$$

Similar to (11), we arrive at for any $m \geq 1$

$$\widetilde{\theta}_{n+1} = \widetilde{\Psi}(n|m-1)\widetilde{\theta}_m + \sum_{j=m}^{n} \frac{1}{j^\gamma} \widetilde{\Psi}(n|j)\xi_j - \sum_{j=m}^{n} \frac{1}{j^\gamma} \widetilde{\Psi}(n|j)[\mathrm{Sgn}(\varphi_j)\varphi_j' - D_{\rm SR}]\widetilde{\theta}_j. \tag{21}$$

Next define

$$\widetilde{w}_{n+1}(t) = ((\lfloor nt \rfloor + 1)/\sqrt{n+1})(\overline{\theta}_{n+1} - \theta_*)$$

for $t \in [0, 1]$, where $\overline{\theta}_n$ is given by (18). Similarly as in Section II, we obtain the following.

Theorem 3.4: Under (**B**), $\widetilde{w}_{n+1}(\cdot)$ converges weakly to a Brownian motion whose covariance is

$$S_{\rm SR}^{*}t = D_{\rm SR}^{-1}\Sigma_{\rm SR}(D_{\rm SR}^{-1})'t$$

where $D_{\rm SR}$ is given in condition (**B**) and

$$\Sigma_{\rm SR} = \mathbf{E}\xi_1\xi_1' + \sum_{j=2}^{\infty}\mathbf{E}\xi_1\xi_j' + \sum_{j=2}^{\infty}\mathbf{E}\xi_j\xi_1'. \tag{22}$$

B. Minimal Window of Averaging

So far, we have only considered the averaging with window width $O(n)$. In [21], averaging with "minimal" window width, the smallest window width needed to be effective for improving the performance, was considered. From an application point of view, the minimal window of averaging provides a useful insight. Following the approach outlined in [22, Ch. 11.1], let us illustrate the idea by use of the sign-regressor algorithm. For iterates θ_j given by (18), for $t \in [t_{n+i} - t_n, t_{n+i+1} - t_n)$ and $i \geq 0$, define

$$U_j = j^{\gamma/2}(\theta_j - \theta_*), \qquad U^n(t) = U_{n+i}. \tag{23}$$

Taking the averaging window width to be $\lfloor tn^{\gamma}\rfloor$ rather than $O(n)$ as before, for any $t > 0$, define

$$\begin{aligned}\widehat{\theta}^n(t) &= \frac{1}{\lfloor tn^{\gamma}\rfloor}\sum_{i=n}^{n+\lfloor tn^{\gamma}\rfloor-1}\theta_i\\ \widehat{U}^n(t) &= \frac{1}{\lfloor tn^{\gamma/2}\rfloor}\sum_{i=n}^{n+\lfloor tn^{\gamma}\rfloor-1}[\theta_i - \theta_*].\end{aligned} \tag{24}$$

It follows that

$$\begin{aligned}U^n(t) = U^n(0) &+ \sum_{i=n}^{n+m(t_n+t)-1}\frac{1}{i^{\frac{\gamma}{2}}}\xi_i - \sum_{i=n}^{n+m(t_n+t)-1}\frac{D_{\rm SR}}{i^{\gamma}}U_i\\ &- \sum_{i=n}^{n+m(t_n+t)-1}\frac{1}{i^{\gamma}}[{\rm Sgn}(\varphi_i)\varphi_i' - D_{\rm SR}]U_i + o(1)\end{aligned}$$

where $o(1) \to 0$ in probability uniformly in t. Using the weak convergence method (see [22, Chs. 8 and 10]), we establish that $U^n(\cdot)$ converges weakly to $U(\cdot)$, which is the stationary solution to

$$dU(t) = -D_{\rm SR}U(t)dt + \Sigma_{\rm SR}^{1/2}dw(t) \tag{25}$$

where $\Sigma_{\rm SR}^{1/2}$ is the "square root" of $\Sigma_{\rm SR}$ given in (22). By invoking [22, Theorem 1.1, p. 331], we obtain the following.

Theorem 3.5: Assume (**B**). For each $t > 0$, treat $\{\widehat{U}^n(t)\}$ as a sequence of random variables. Then $\widehat{U}^n(t)$ converges in distribution to a normal random vector $\widehat{U}(t)$ with mean 0 and covariance $V_t = S_{\rm SR}^{*}/t + O(1/t^2)$ where $S_{\rm SR}^{*}$ is defined in Theorem 3.4.

C. Constant-Step-Size Algorithms

In many practical applications of adaptive filtering such as the interference suppression example discussed in Section IV, constant-step-size algorithms are required for tracking slowly varying parameter variations. This subsection considers iterate averaging for the *constant step size* sign-regressor algorithm. The algorithm is

$$\begin{aligned}\theta_{n+1}^{\varepsilon} &= \theta_n^{\varepsilon} + \varepsilon{\rm Sgn}(\varphi_n)(y_n - \theta_n^{\varepsilon,\prime}\varphi_n)\\ \overline{\theta}_{n+1} &= (1-\rho_n)\overline{\theta}_n + \rho_n\theta_n^{\varepsilon}\end{aligned} \tag{26}$$

where $0 < \rho_n < 1$ denotes a forgetting factor applied to the averaging procedure. Since the minimal window of averaging is of particular importance, the following discussion is devoted to such cases.

Case i): Decreasing Forgetting Factor $\rho_n = 1/(n+1)$: Using (**B**) and for definiteness, let us concentrate on the case of bounded mixing condition. Then [3, p. 197] and [8, Ch. 7.3] imply that $\sqrt{\varepsilon}\sum_{i=n}^{n+\lfloor t/\varepsilon\rfloor-1}\xi_i$ converges weakly to a Brownian motion with covariance $\Sigma_{\rm SR}t$, where $\Sigma_{\rm SR}$ is given by (22). The sequence $\{(y_n,\varphi_n)\}$ is uniform mixing, so are $\{y_n{\rm Sgn}(\varphi_n)\}$ and $\{{\rm Sgn}(\varphi_n)\varphi_n'\}$. Therefore, they are strongly ergodic. Consequently, for any $m \geq 1$

$$\frac{1}{n}\sum_{j=m}^{n+m-1}{\rm Sgn}(\varphi_j)\varphi_j' \to D_{\rm SR}, \quad \frac{1}{n}\sum_{j=m}^{n+m-1}{\rm Sgn}(\varphi_j)y_j \to b \tag{27}$$

both in probability. (In fact, they converge w.p. 1, but for our analysis, convergence in probability is sufficient.) Define $\theta^{\varepsilon}(t) = \theta_n^{\varepsilon}$ for $t \in [n\varepsilon, n\varepsilon+\varepsilon)$. Similar to [20], we obtain the following. Assume θ_0^{ε} converges to θ_0 and (**B**). Then $\theta^{\varepsilon}(\cdot)$ converges weakly to $\theta(\cdot)$, which is a solution to the differential equation (19). Furthermore, for any $t_{\varepsilon} \to \infty$ as $\varepsilon \to 0$, $\theta^{\varepsilon}(t_{\varepsilon}+\cdot)$ converges weakly to θ_* given by (19).

Define $u_n = (\theta_n - \theta_*)/\sqrt{\varepsilon}$ and $u^{\varepsilon}(t) = u_n$ for $t \in [\varepsilon n, \varepsilon n + \varepsilon)$. Then it can be shown that $u^{\varepsilon}(t_{\varepsilon}+\cdot)$ converges weakly to $u(\cdot)$, a solution of (25), as $\varepsilon \to 0$ and $t_{\varepsilon} \to \infty$. To proceed, define

$$\widehat{U}^{\varepsilon}(t) = (\sqrt{\varepsilon}/t)\sum_{i=\lfloor t_{\varepsilon}/\varepsilon\rfloor}^{\lfloor(t_{\varepsilon}+t)/\varepsilon\rfloor}(\theta_i^{\varepsilon} - \theta_*).$$

Using the argument of [22, p. 333], we obtain the following.

Theorem 3.6: Assume θ_0^{ε} converges to θ_0 and (**B**). Then, for each fixed $t > 0$, $\widehat{U}^{\varepsilon}(t)$ converges in distribution to a normal random vector $\widehat{U}(t)$ with mean 0, covariance $S_{\rm SR}^{*}/t + O(1/t^2)$, and $S_{\rm SR}^{*}$ defined in Theorem 3.4.

Case ii): Constant Forgetting Factor ρ: Here we take a constant forgetting factor $\rho_n = \rho$ with $0 < \rho < 1$. In the analysis, we examine the asymptotic properties of the dynamic system given by (26) as $\varepsilon \to 0$ and $\rho = \rho_{\varepsilon} \to 0$, whereas in the implementation, ρ and ε are kept as constants. Define

$$\widetilde{U}^{\varepsilon}(t) = \frac{\sqrt{\rho}}{t}\sum_{j=t_{\varepsilon}/\rho}^{(t_{\varepsilon}+t)/\rho-1}(\theta_j - \theta_*). \tag{28}$$

By using the interpolations $\overline{u}^{\varepsilon}(\cdot)$ and $u^{\varepsilon}(\cdot)$, a similar argument as in Theorem 3.6 leads to the following result.

Theorem 3.7: Assume (**B**). For each fixed $t > 0$, $\widetilde{U}^{\varepsilon}(t)$ converges in distribution to a normal random vector $\widetilde{U}(t)$ with mean 0, covariance $S_{\rm SR}^{*}/t + O(1/t^2)$, and $S_{\rm SR}^{*}$ defined in Theorem 3.4.

D. Remarks on Averaged Sign-Error and LMS Algorithms

The discussions thus far readily carry over to minimal windows of averaging for the following decreasing step-size and constant step-size sign-error algorithms:

$$\begin{aligned}\theta^{\varepsilon}_{n+1} &= \theta^{\varepsilon}_{n} + \varepsilon\varphi_n \operatorname{sgn}(y_n - \theta^{\varepsilon,\prime}_{n}\varphi_n)\\ \overline{\theta}_{n+1} &= (1-\rho_n)\overline{\theta}_n + \rho_n\theta^{\varepsilon}_n\end{aligned} \tag{29}$$

where $0 < \rho_n < 1$. Moreover, either a constant forgetting factor ρ or a sequence of decreasing forgetting factors can be included. Results similar to Theorems 3.5, 3.6, and 3.7 can be obtained with the use of condition (**A**) in lieu of (**B**) and with $S^*_{\rm SR}$ replaced by $S^*_{\rm SE}$ given in Theorem 2.7.

Similar results for minimal windows for averaging and constant step size LMS algorithms with averaging can be established by using the techniques of [22]; we summarize the results as follows. For fixed t, $\widetilde{U}^{\varepsilon}(t)$, as defined in (28), converges in distribution to $\widetilde{U}(t)$ a normal random variable with mean 0 and covariance $S^*_{\rm LMS}/t + O(1/t^2)$, where

$$S^*_{\rm LMS} \stackrel{\rm def}{=} D^{-1}_{\varphi\varphi}\Sigma_{\rm LMS}D^{-1}_{\varphi\varphi} \tag{30}$$

and

$$\begin{aligned}\Sigma_{\rm LMS} &= \mathbf{E}\xi_1\xi_1' + \sum_{j=2}^{\infty}\mathbf{E}\xi_1\xi_j' + \sum_{j=2}^{\infty}\mathbf{E}\xi_j\xi_1'\\ D_{\varphi\varphi} &= \mathbf{E}\{\varphi_1\varphi_1'\}, \qquad \xi_n \stackrel{\rm def}{=} \varphi_n(y_n - \varphi_n'\theta_*).\end{aligned} \tag{31}$$

IV. Case Study: Sign Algorithms for Blind Multiuser Detection in DS/CDMA Systems

DS/CDMA is among the most promising multiplexing technologies for cellular telecommunications services such as personal communications, mobile telephony, and indoor wireless networks. Demodulating a given user in a DS/CDMA network requires processing the received signal to minimize two types of interference, namely, narrow-band interference (NBI) and wide-band multiple-access interference (MAI) caused by other spread-spectrum users in the channel—as well as ambient channel noise [15]. NBI is caused by the coexistence of spread-spectrum signals with conventional communications; see [15] and [17] for a recent review of active NBI suppression methods that have resulted in substantial gains in DS/CDMA systems. MAI arises in DS/CDMA systems due to the fact that all users communicate through the same physical channel using *nonorthogonal* multiplexing, which has many advantages in wireless CDMA systems such as greater bandwidth utilization under conditions of channel fading and bursty traffic.

Recently, *blind multiuser detection* techniques [14], [30], [31] have been developed that allow one to use a linear multiuser detector for a given user with no knowledge beyond that required for implementation of the conventional detector for that user. Blind multiuser detection is useful in mobile wireless channels when the desired user can experience a deep fade or if a strong interferer suddenly appears. In [14] a blind LMS algorithm is given for linear minimum mean-square error (MMSE) detection. In [31], a code-aided blind RLS algorithm for jointly suppressing MAI and NBI is given. More recently, in [16], a blind averaged LMS algorithm is presented with a heuristic mean-square error convergence analysis in the same spirit as [14] and [31].

The objective of this section is to use the averaged sign-error LMS and the sign-regressor LMS algorithms analyzed in Sections II and III of this paper to the MMSE detection scheme for multiuser detection in a DS/CDMA system. The performance of the sign algorithms will be studied and compared with that of the standard LMS.

A. DS/CDMA Signal Model

Consider a synchronous K-user binary DS/CDMA communication system. Assume that this system transmits through an additive white Gaussian noise channel. After the received continuous-time signal is preprocessed and sampled at the CDMA receiver (the received signal is passed through a chip-matched filter followed by a chip-rate sampler), the resulting discrete-time received signal at time n, denoted by r_n, is given by (see [31] for details)

$$r_n = \sum_{k=1}^{K}\sqrt{P_k}\,b_k(n)s_k + \varsigma_n + \sigma\varpi_n. \tag{32}$$

Here r_n is an N-dimensional vector; N is called the processing (spreading) gain; s_k is an N-vector denoting the normalized signature sequence of the kth user, i.e., each element $s_{ki} \in \{-1/\sqrt{N}, +1/\sqrt{N}\}$ for $i = 1, 2, \ldots, N$, so that $s_k's_k = 1$; $b_k(n)$ denotes the data bit of the kth user transmitted at time n; $P_k = A_k^2$ is the received power of the kth user; ς_n is the NBI signal N-vector, which is assumed to be a bounded stationary autoregressive (AR) process with mean zero and covariance matrix R_ς; σ is the standard deviation of the noise samples; ϖ_n is a white Gaussian vector with mean zero and covariance matrix I, where I denotes the $N\times N$ identity matrix. It is assumed that the discrete-time stochastic processes $\{b_k(n)\}$, $\{\varsigma_n\}$, and $\{\varpi_n\}$ are mutually independent, and that $\{b_k(n)\}$ is a collection of independent equiprobable ± 1 random variables.

We assume that user 1 is the user of interest. Following the definition of s_k, s_1 denotes the normalized signature sequence of user 1. For user 1, the term $\sum_{k=2}^{K}\sqrt{P_k}\,b_k(n)s_k$ in (32) is termed MAI. The aim of a multiuser detector is to suppress the MAI and adaptively estimate (demodulate) the bit sequence $b_1(n)$ given the observation sequence r_n. A linear blind multiuser detector demodulates the bits of user 1 according to (see [31] for details) $\widehat{b}_1(n) = \operatorname{sgn}(c_*'r_n)$, where $\widehat{b}_1(n)$ denotes the estimate of the transmitted bit $b_1(n)$ at time n, and c_* denotes an appropriately chosen "weight vector." In this section, we focus on the widely used code-aided blind linear mean output error (MOE) detector [14], [31] which chooses the "weight vector" c so as to minimize the MOE cost function

$$\zeta_n \stackrel{\rm def}{=} \mathbf{E}\{(c'r_n)^2\} \text{ subject to } c's_1 = 1. \tag{33}$$

The constraint ensures that the received energy from the user of interest is equal to 1. Thus, the above is a minimization of the energy from the interferers. Furthermore, as shown in [14], the MOE cost function has a unique global minimum (with respect to c). The blind MOE detector yields the following estimate $\widehat{b}_1(n)$ of the transmitted signal (see [31] for details):

$$\widehat{b}_1(n) = \operatorname{sgn}(c_*{}'r_n), \qquad \text{where } c_* = \frac{R^{-1}s_1}{s_1'R^{-1}s_1} \tag{34}$$

and $R=\mathbf{E}\{rr'\}$ denotes the autocorrelation matrix of the received signal r. In the preceding equation, c_* is the optimal linear MOE "weight vector." Such a detector is "blind" since it does not assume any knowledge of the data symbols $b_1(n)$ and signature sequences of other users.

The output signal-to-interference ratio (SIR) is widely used to characterize the performance of a linear multiuser receiver. The SIR for an arbitrary weight vector c is defined as

$$\text{SIR} \stackrel{\text{def}}{=} \frac{P_1}{\sigma^2 c'c + \sum_{k=2}^{K} P_k (c's_k)^2}. \tag{35}$$

The SIR of the optimal weight vector c_* and the MOE of c_* are given by [31, eqs. 5 and 7, respectively]

$$\begin{aligned} \text{SIR}^* &= P_1 \left[s_1'(R - P_1 s_1 s_1')^{-1} s_1 \right] \\ \overline{\zeta} &= \mathbf{E}\{(c_*' r)^2\} = c_*' R c_* = \frac{1}{s_1' R^{-1} s_1}. \end{aligned} \tag{36}$$

B. Adaptive Sign Algorithms for Blind Multiuser Detection

In *adaptive* blind multiuser detection problems, we are interested in recursively adapting the weight vector c_n to minimize ζ_n, the MOE given by (33). In particular, it is often necessary to use a constant step-size tracking algorithm due to the time-varying nature of c_* caused by the birth and death of users (MAI interferers). We now present constant step-size versions of the sign-regressor and sign-error algorithms for blind adaptive multiuser detection.

In presenting the sign algorithms for blind adaptive multiuser detection, it is convenient to work with an unconstrained optimization problem rather than (33). Let $c_{n,i}$, for $i=1,\dots,N$ denote the components of c_n. The constrained optimization problem (33) may be transformed into an unconstrained optimization problem by solving for one of the elements $c_{n,i}$, $i \in [1,\dots,N]$ using the constraint (33). With no loss of generality, we solve for the first element $c_{n,1}$ and obtain

$$c_{n,1} = \left(1 - \sum_{i=2}^{N} s_{1,i} c_{n,i}\right) / s_{1,1}.$$

By defining the $(N-1)$-dimensional vector

$$\theta_n = (c_{n,2},\dots,c_{n,N})'$$

we obtain the equivalent unconstrained optimization problem

$$\text{Compute } \min_\theta J_n \text{ where } J_n = \mathbf{E}(y_n - \theta'\varphi_n)^2. \tag{37}$$

Here, $y_n = -r_{n,1}/s_{1,1}$ and φ_n denotes the $(N-1)$-dimensional vector

$$\varphi_n = (r_{n,2} - r_{n,1}s_{1,2}/s_{1,1}, \dots, r_{n,N} - r_{n,1}s_{1,N}/s_{1,1})'.$$

As in (20), let θ_* denote the MMSE solution

$$\theta_* = \mathbf{E}\{\varphi_n\varphi_n'\}^{-1}\mathbf{E}\{\varphi_n y_n\}.$$

It is straightforward but tedious to show that the components of θ_* are indeed the last $(N-1)$ elements of optimal weight vector c_* defined in (34). Using the φ_n defined above, we call the constant step-size sign-regressor algorithm (26) with fixed forgetting factor ρ operating on the DS/CDMA signal model (32) as the *blind averaged sign-regressor algorithm*. Similarly, we call the constant step-size sign-error algorithm (29) as the *blind averaged sign-error algorithm*.

Remark 4.1: When $s_{k,1}$ is small, computations using φ_n and y_n may become ill-conditioned. This is trivially taken care of as follows. because $s_{1,1}^2 = 1/N$, $1/s_{1,1} = Ns_{1,1}$ in φ_n and y_n.

Canonical Coordinates. In [14], constraint (33) for the blind LMS algorithm is taken care of by introducing canonical coordinates together with a MSE analysis. The essential idea is to replace the unconstrained gradient of the MOE in (33), namely, $r_n' c_n r_n$, by its component orthogonal to s_1, namely, $r_n' c_n (r_n - r_n' s_1 s_1)$. The blind averaged LMS and blind averaged sign-error algorithm can be expressed in canonical coordinates as

$$\begin{cases} \widehat{c}_{n+1} = \widehat{c}_n - \mu r_n' \widehat{c}_n (r_n - r_n' s_1 s_1), & \widehat{c}_0 = s_1 \\ c_{n+1} = (1-\rho)c_n + \rho \widehat{c}_{n+1} \end{cases}$$

$$\begin{cases} \widehat{c}_{n+1} = \widehat{c}_n - \varepsilon \operatorname{sgn}(r_n' \widehat{c}_n)(r_n - r_n' s_1 s_1), & \widehat{c}_0 = s_1 \\ c_{n+1} = (1-\rho)c_n + \rho \widehat{c}_{n+1} \end{cases}$$

respectively. It is easily seen that the estimates $\widehat{c}_n$ in the above two algorithms automatically satisfy constraint (33). However, it is not possible to derive a sign-regressor algorithm in canonical coordinates that satisfies constraint (33). For example, the sign-regressor algorithm in canonical coordinates

$$\begin{cases} \widehat{c}_{n+1} = \widehat{c}_n - \varepsilon r_n' \widehat{c}_n \operatorname{Sgn}(r_n - r_n' s_1 s_1), & \widehat{c}_0 = s_1 \\ c_{n+1} = (1-\rho)c_n + \rho \widehat{c}_{n+1} \end{cases}$$

does *not* satisfy constraint (33). In the numerical examples presented later, we found the performance of the blind averaged LMS and sign-error algorithms in canonical coordinates are identical to the corresponding algorithms derived for the unconstrained cost function. However, it is more convenient to work with the equivalent algorithm derived for the unconstrained cost function.

C. Performance Analysis of Averaged Algorithms

Note that we have assumed that $\{\varsigma_n\}$ is a bounded sequence of regressive process (e.g., stationary truncated Gaussian autoregreesive process), and that $\{b_k(n)\}$ and $\{\varpi_n\}$ are i.i.d. processes. It follows that φ_n is a sum of bounded mixing sequence and martingale difference sequence, so the noise condition in (**A**) is satisfied. Thus, all the convergence and asymptotic optimality results derived in Sections III-C and -D for the averaged sign-error and sign-regressor algorithms hold. To proceed, we derive approximate expressions for S_{LMS}, S_{SE}, and S_{SR} and the asymptotic excess mean-square error and SIR of the averaged and un-averaged sign LMS algorithms for the DS/CDMA signal model. These are commonly used performance measures for adaptive filtering algorithms in the signal processing and CDMA literature; see [14] or [31]. In what follows, we use $D_{\varphi\varphi}$ and D_{SR}, the covariance matrices defined in condition (**B**) and (31).

To obtain expressions for the asymptotic excess mean-square error, we first note that the zero mean estimation error e_n^* of the MMSE (Wiener) solution θ_*, given by

$$e_n^* = y_n - \varphi_n'\theta_* \tag{38}$$

is uncorrelated with φ_n—this is the principle of orthogonality [13] for the MMSE solution θ_*, which is easily verified. Note that for the DS/CDMA signal model (32), using (36), and the definition of the equivalent unconstrained problem (37), we have $\mathbf{E}{e_n^*}^2/2 = \mathbf{E}[y_n - {\theta_*}'\varphi_n]^2/2 = \mathbf{E}({c_*}'r)^2/2 = \overline{\zeta}/2$.

We need the following additional assumptions:

i) e_n^* and φ_n are independent.

ii) The input data r_n and the previous weight vector c_{n-1} are statistically independent [13, Ch. 9].

These two assumptions are not needed for the weak convergence analysis presented earlier. They are introduced only to give simplified closed-form expressions for the weighted error correlation, excess mean square error, and steady-state SIR. Without these assumptions, the expressions would involve fourth-order moments—while these can be computed, the resulting expressions are messy and yield little insight (see also [14]).

Assumption i) is justified when for fixed processing gain N, the number of users K is large. Assuming the binary signature sequences are chosen randomly (equiprobably over all choices) and the amplitudes are identical, one can apply the i.i.d. version of the central limit theorem to (32). Alternatively, if the amplitudes $A_k = o(K)$, $k = 1, \ldots, K$, the Lindberg–Feller central limit theorem, see [36, pg. 150], can be applied. The central limit theorem implies r_n is asymptotically a zero-mean N-dimensional Gaussian random vector. This in turn implies $\{\varphi_n\}$ is approximately an $(N-1)$-dimensional zero-mean Gaussian random vector, and e_n^* and $\{y_n\}$ are scalar zero-mean Gaussian random variables. Since orthogonality for Gaussian random variables is equivalent to independence, e_n^* and φ_n are asymptotically independent.

Note that assumption ii) is satisfied if the interference consists only of MAI and white noise. This assumption is used in [31] for analyzing the blind RLS algorithm; it is also commonly used in deriving closed-form expressions for the performance of adaptive filtering algorithms (see [13]).

Weighted Error Correlation: The weighted error correlation matrices for the various averaged algorithms can be obtained from the analysis of Sections IV-A and –B as follows.

`Averaged LMS Algorithm`. Consider S^*_{LMS} given in (30). Because of i) and ii) above, Σ_{LMS} in (31) can be computed as

$$\Sigma_{\mathrm{LMS}} = \mathbf{E}\{\varphi_n\varphi_n'(y_n - \varphi_n'\theta_*)^2/2\} = D_{\varphi\varphi}\overline{\zeta}/2.$$

Let $K(n) \stackrel{\mathrm{def}}{=} \mathbf{E}\{\tilde{\theta}_n\tilde{\theta}_n'\}$ denote the weighted error correlation matrix. Note that $K(n)$ is an $(N-1)\times(N-1)$ positive-definite matrix. Then (30) implies that

$$K(\infty) = \rho S^*_{\mathrm{LMS}} = \frac{\rho D_{\varphi\varphi}^{-1}\overline{\zeta}}{2}. \tag{39}$$

`Averaged Sign-Regressor Algorithm`. Consider S^*_{SR} given in Theorem 3.4. Using results i) and ii), Σ_{SR} in (22) is

$$\begin{aligned}\Sigma_{\mathrm{SR}} =& \mathbf{E}\left\{\mathrm{Sgn}(\varphi_n)\mathrm{Sgn}(\varphi_n)'\frac{(y_n - \varphi_n'\theta_*)^2}{2}\right\}\\ =& \mathbf{E}\left\{\mathrm{Sgn}(\varphi_n)\mathrm{Sgn}(\varphi_n)'\right\}\frac{\overline{\zeta}}{2} = \frac{\overline{\zeta}}{2}I_{(N-1)\times(N-1)}\end{aligned}$$

since $\mathbf{E}\{\mathrm{Sgn}(\varphi_n)\mathrm{Sgn}(\varphi_n)'\} = I_{(N-1)\times(N-1)}$. We have the weighted error correlation matrix

$$K_{\mathrm{SR}}(\infty) = \rho S^*_{\mathrm{SR}} = \frac{\rho\overline{\zeta}D_{\mathrm{SR}}^{-1}D_{\mathrm{SR}}^{-1\prime}}{2}. \tag{40}$$

Next, as in [10], we use assumption ii) above which implies that φ_n is approximately a Gaussian sequence. Then Price's formula yields

$$D_{\mathrm{SR}} = \sqrt{\frac{2}{\pi}}\mathcal{D}D_{\varphi\varphi}$$

where

$$\begin{aligned}\mathcal{D} \stackrel{\mathrm{def}}{=} \mathrm{diag}\Bigg[&\frac{1}{\sqrt{\mathbf{E}\{\varphi_n^2(1,1)\}}}, \ldots,\\ &\frac{1}{\sqrt{\mathbf{E}\{\varphi_n^2(N-1,N-1)\}}}\Bigg].\end{aligned} \tag{41}$$

`Averaged Sign-Error Algorithm`. With S^*_{SE} defined in Theorem 2.7, using results i) and ii) above, Σ_{SE} in (17) is given by

$$\begin{aligned}\Sigma_{\mathrm{SE}} &= \mathbf{E}\left\{\varphi_n\varphi_n'\frac{(\mathrm{sgn}(y_n - \varphi_n'\theta_*))^2}{2}\right\}\\ &= \mathbf{E}\frac{\{\varphi_n\varphi_n\}'}{2} = \frac{D_{\varphi\varphi}}{2}.\end{aligned}$$

Thus, the weighted error correlation matrix satisfies

$$K_{\mathrm{SE}}(\infty) = \rho S^*_{\mathrm{SE}} = \rho D_{\mathrm{SE}}^{-1}D_{\varphi\varphi}\frac{D_{\mathrm{SE}}^{-1}}{2}. \tag{42}$$

Under the Gaussian assumption, [6, eq. (39)] shows that $D_{\mathrm{SE}} = CD_{\varphi\varphi}$ where $C = (\sqrt{2/\pi})(1/\sigma_e(n))$.

Asymptotic Excess Mean Square Error: The MOE ζ_n defined in (33) can be expressed again as

$$\begin{aligned}\zeta_n &= \mathbf{E}\frac{[y_n - \varphi_n'(\theta_* - \tilde{\theta}_n)]^2}{2}\\ &= \frac{\overline{\zeta}}{2} + \frac{\mathrm{tr}[D_{\varphi\varphi}K(n)]}{2} - \mathbf{E}\left\{(y_n - \varphi_n'\theta_*)\varphi_n'\tilde{\theta}_n\right\}.\end{aligned}$$

Since $\mathbf{E}\{\tilde{\theta}_n\} \to 0$ as $n \to \infty$, the last term is asymptotically unimportant. Hence, for large n, $\zeta_n = \overline{\zeta}/2 + \epsilon_{\mathrm{ex}}(n)$ where the excess mean square error is defined as

$$\epsilon_{\mathrm{ex}}(n) = \frac{\mathrm{tr}[D_{\varphi\varphi}K(n)]}{2}. \tag{43}$$

In what follows, we compute expressions for the asymptotic excess mean square error $\epsilon_{\mathrm{ex}}(\infty)$.

`Averaged LMS Algorithm`. It follows from (39) that

$$\epsilon_{\mathrm{ex}}(\infty) = \frac{\mathrm{tr}[D_{\varphi\varphi}K(\infty)]}{2} = \frac{\rho\overline{\zeta}(N-1)}{2}. \tag{44}$$

Note that the above equation is identical to that of the blind RLS, see [31, eq. 40]. As for blind RLS, the steady-state misadjustment of the averaged LMS algorithm is independent of the eigenvalue distribution of the data autocorrelation matrix.

`Averaged Sign-Regressor Algorithm`. Using the Gaussian assumption which implies (41) together with (40) and (43) yields

$$\frac{\epsilon_{\mathrm{ex}}(\infty)}{2} = \frac{\mathrm{tr}[D_{\varphi\varphi}K_{\mathrm{SR}}(\infty)]}{2} = \frac{\rho}{2}\overline{\zeta}\frac{\pi}{2}\mathrm{tr}\left[\mathcal{D}D_{\varphi\varphi}\mathcal{D}\right]. \tag{45}$$

We can easily compute a lower bound for $\epsilon_{\mathrm{ex}}(\infty)$ by bounding $\mathrm{tr}\left[\mathcal{D}D_{\varphi\varphi}\mathcal{D}\right]$ in terms of $\mathrm{tr}\left[\mathcal{D}D_{\varphi\varphi}\mathcal{D}\right]^{-1}$ as follows. Let λ_i, $i = 1, \ldots, N-1$, be the eigenvalues of the positive-definite symmetric matrix $\left[\mathcal{D}D_{\varphi\varphi}\mathcal{D}\right]^{-1}$. Since all the diagonal elements of this matrix are 1

$$\mathrm{tr}\left[\mathcal{D}D_{\varphi\varphi}\mathcal{D}\right]^{-1} = \sum_{i=1}^{N-1}\lambda_i = (N-1).$$

TABLE I
ASYMPTOTIC EXCESS MEAN SQUARE ERRORS $\epsilon_{\rm ex}(\infty)$

Algorithm	Standard	Averaged
blind LMS	$\frac{\varepsilon}{2}\zeta \operatorname{tr}[D_{\varphi\varphi}]$	$\frac{\rho}{2}\bar{\zeta}(N-1)$
sign regressor[1]	$\frac{\varepsilon}{2}\bar{\zeta}\sqrt{\frac{\pi}{2}}(N-1)\max_i \sqrt{D_{\varphi\varphi}(i,i)}$	$\frac{\rho}{2}\frac{\pi}{2}\bar{\zeta}(N-1)$
sign error	$\frac{\varepsilon}{2}\sqrt{\frac{\pi}{2}}\operatorname{tr}[D_{\varphi\varphi}]$	$\frac{\rho}{2}\frac{\pi}{2}\bar{\zeta}(N-1)$

[1] lower bound

Using the well-known inequality that the harmonic mean is less than the arithmetic mean, we obtain

$$\frac{N-1}{\operatorname{tr}[\mathcal{D}D_{\varphi\varphi}\mathcal{D}]} = \frac{N-1}{\sum_{i=1}^{N}\frac{1}{\lambda_i}} \le \frac{1}{N-1}\sum_{i=1}^{N}\lambda_i = 1$$

which implies that

$$\operatorname{tr}[\mathcal{D}D_{\varphi\varphi}\mathcal{D}] \ge N-1 \quad \text{and} \quad \epsilon_{\rm ex}(\infty) \ge [\rho/2]\bar{\zeta}(\pi/2)(N-1).$$

`Averaged Sign-Error Algorithm` . It follows from (42) that

$$\epsilon_{\rm ex}(\infty) = \frac{\operatorname{tr}[D_{\varphi\varphi}K_{\rm SE}(\infty)]}{2} = \frac{\rho}{2}\frac{\pi}{2}\bar{\zeta}(N-1). \quad (46)$$

Just like the blind RLS algorithm and the blind averaged LMS algorithm analyzed earalier, the steady-state misadjustment of the averaged sign-error algorithm is independent of the eigenvalue distribution of the data autocorrelation matrix.

It is illustrative to compare the asymptotic excess mean-square error of the averaged sign algorithms with their standard (unaveraged) counterparts. Expressions for the asymptotic excess mean-square error of the standard sign-error algorithms have been derived in [6] and for the sign-regressor algorithm in [10]. Table I summarizes the results.

Remark 4.2: i) All the expressions for the standard algorithms above assume that $\varepsilon \ll 1$. In particular, terms involving ε^2 are negligible. More precise expressions are available in [6] and [10]. The expressions for the sign-regressor algorithm given are lower bounds. ii) $\epsilon_{\rm ex}(\infty)$ for the averaged algorithms do not depend on the eigendistribution of $D_{\varphi\varphi}$. This is particularly useful in dynamic mobile environments where the eigenstructure of $D_{\varphi\varphi}$ can change rapidly. In [31], a similar property is shown for the blind RLS algorithm. It only remains to give tractable expressions for $\operatorname{tr}[D_{\varphi\varphi}]$. It is tedious but straightforward to show that

$$\operatorname{tr}[D_{\varphi\varphi}] = 2\left(\sum_{k=2}^{k} P_k(1-s_1's_k)\frac{s_{k1}}{s_{11}} + (N-1)\sigma^2\right).$$

SIR: The SIR Defined in (35) can be reformulated in terms of the asymptotic excess mean-square error as

$$\mathrm{SIR} = \mathrm{SIR}^*/[1+(\epsilon_{\rm ex}/p_1)\mathrm{SIR}^*].$$

D. Numerical Examples

In this section, computer simulations are presented that illustrate the performance of the averaged sign algorithms. For a detailed numerical study of the averaged LMS algorithm in blind-multiuser detection, please refer to [16]. As is common in the CDMA literature, we use the steady-state SIR as the figure of merit for assessing the interference suppression capability of the various algorithms. All the signal and noise powers are given in dB relative to the channel noise variance σ_n^2, see (32). The simulations below assume a synchronous DS/CDMA system with processing gain $N = 31$. The desired user's signature is generated as an m-sequence. The signature sequences of the other MAI's are generated randomly.

Example 1 (MAI Suppression): The user of interest has SNR of 20 dB. There are 7 multiple access interferers: 5 users each of SNR 20 dB, and two users of SNR 40 dB. Fig. 1 shows the SIR versus time for the following six algorithms, averaged over 100 independent simulations: a) blind LMS versus blind averaged LMS; b) blind sign regressor versus blind averaged sign regressor; c) blind sign error versus blind averaged sign error.

In addition, we also simulated the blind RLS algorithm given in [31]. The blind RLS algorithm and averaged blind LMS algorithm yielded virtually indistinguishable SIR plots. It is seen from Fig. 1 that the averaged LMS and averaged sign algorithms exhibit faster convergence than the unaveraged algorithms.

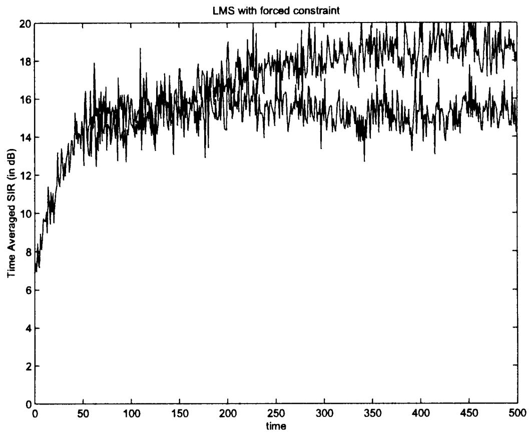
Example 2(Dynamic Environment—MAI and NBI With Time-Varying Statistics): The simulation starts with one desired user's signal and 6 MAI signals each of 20 dB. At time 500, a 10-dB NBI interferer is added to the system. The NBI signal is a bounded stationary AR signal with both poles at 0.99. At time 1000, another strong MAI signal of 40 dB is added. At time 1500, three of the original 20-dB MAI signals are removed from the system. Fig. 2 shows SIR versus time for the six algorithms averaged over 100 independent simulations:

a) blind LMS with step size $\varepsilon = 10^{-3}$;
b) blind averaged LMS with $\varepsilon = 10^{-3}$, $\rho = 0.01$;
c) blind sign regressor with $\varepsilon = 2\times 10^{-3} \approx 2^{-9}$ (for implementation using binary shifts);
d) blind averaged sign regressor with $\varepsilon = 2\times 10^{-3} \approx 2^{-9}$; $\rho = 0.01$.
e) blind sign error with $\varepsilon = 2\times 10^{-3} \approx 2^{-9}$;
f) blind averaged sign error with $\varepsilon = 2\times 10^{-3} \approx 2^{-9}$, $\rho = 0.01$.

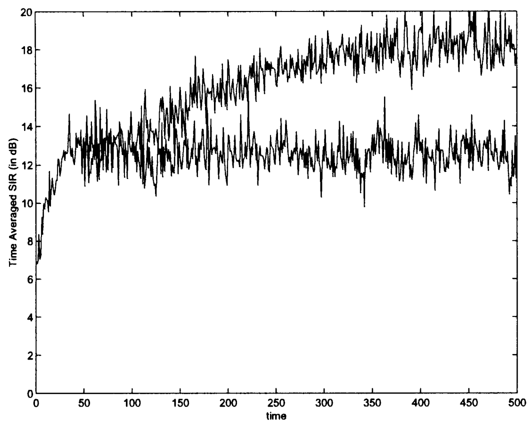
It is seen that in all cases, the averaged algorithms have better convergence properties than the algorithms without averaging. Also, it is interesting to note that the sign-regressor algorithm performs similarly to the LMS algorithm whereas the sign-error algorithm performs worse.

V. FURTHER REMARKS

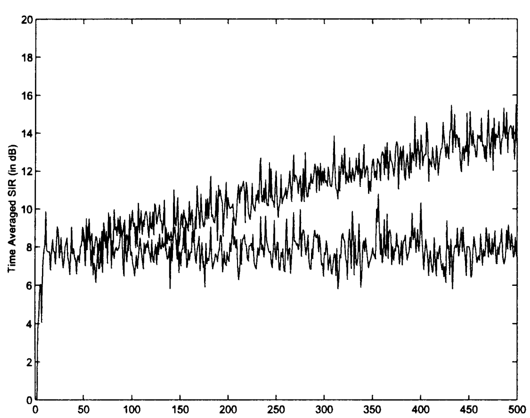
Iterate-averaging algorithms have been developed in this paper, and have been shown to be asymptotically efficient in the sense that $\sqrt{n}(\overline{\theta}_n - \theta_*) \sim N(0, S^*)$, where S^* (with $S^* = S_{\rm SE}$, or $S_{\rm SR}$, or $S_{\rm LMS}$ depending on the type of algorithms) is the optimal asymptotic covariance. In fact, a functional central limit theorem is obtained and the usual central limit theorem becomes a corollary. As pointed out in [38], the asymptotic optimality cannot be improved by placing a constant in the



(a) Blind LMS and Averaged LMS algorithms with $\varepsilon = 10^{-4}$, $\rho = 5 \times 10^{-3}$



(b) Blind Sign regressor and Averaged Sign Regressor algorithms with $\varepsilon = 10^{-4}$, $\rho = 5\times10^{-3}$



(c) Blind Sign regressor and Averaged Sign Regressor algorithms with $\varepsilon = 10^{-4}$, $\rho = 5\times10^{-3}$

Fig. 1. Average SIR versus time for MAI suppression. The parameters are specified in Section IV-D.

gain. That is, if we replace $1/n^\gamma$ by a/n^γ for some constant a, the a will not show up in the asymptotic covariance.

Using essentially the same analysis but with more complex notation, we can obtain similar results with more general step size a_n in lieu of the slowly varying step size $O(1/n^\gamma)$ used in this paper. For some of the related references, we refer the reader to [22, Ch. 11].

In a recent work [18], we have applied the sign algorithm to discrete stochastic approximation for optimization of spreading codes. For future study, one may consider further properties of such algorithms. In addition, one may consider an averaging algorithm with feedback; see [22, p. 60] and the references therein. One may also study algorithms using averaging in both iterates and observations. Another interesting problem is to consider the associated adaptive step-size algorithms (see [2] and [22, p. 53]).

APPENDIX

A. Proof of Theorem 2.4

We use the techniques of perturbed test function to obtain the estimate. Define $V(\widetilde{\theta}) = (\widetilde{\theta}'\widetilde{\theta})/2$. Note that θ_n is $\mathcal{F}_n$-measurable so is $\widetilde{\theta}_n$. By virtue of condition (**A**), for sufficiently large n,

$$\begin{aligned}\mathbf{E}_n V(\widetilde{\theta}_{n+1}) - V(\widetilde{\theta}_n) &= \frac{1}{n^\gamma}\widetilde{\theta}_n'\mathbf{E}_n[f_n(\theta_n) - f_n(\theta_*)]\\ &\quad + \frac{1}{n^\gamma}\widetilde{\theta}_n'\mathbf{E}_n f_n(\theta_*) + \frac{1}{2n^{2\gamma}}\mathbf{E}_n\varphi_n'\varphi_n\\ &= \frac{1}{n^\gamma}\widetilde{\theta}_n' D_{\rm SE}\widetilde{\theta}_n + \frac{1}{n^\gamma}\widetilde{\theta}_n'(A_n - D_{\rm SE})\widetilde{\theta}_n\\ &\quad + \frac{1}{n^\gamma}o(|\widetilde{\theta}_n|^2) + \frac{1}{n^\gamma}\mathbf{E}_n\widetilde{\theta}_n' f_n(\theta_*)\\ &\quad + \frac{1}{2n^{2\gamma}}\mathbf{E}_n\varphi_n'\varphi_n. \end{aligned} \tag{47}$$

To proceed, we introduce the perturbations and define

$$\begin{aligned}\widetilde{V}(\widetilde{\theta}, n) &= \sum_{j=n}^{\infty}\frac{1}{j^\gamma}\mathbf{E}_n\widetilde{\theta}'[A_j - D_{\rm SE}]\widetilde{\theta}\\ \widehat{V}(\widetilde{\theta}, n) &= \sum_{j=n}^{\infty}\frac{1}{j^\gamma}\mathbf{E}_n\widetilde{\theta}' f_j(\theta_*)\\ W(n) &= V(\widetilde{\theta}_n) + \widetilde{V}(\widetilde{\theta}_n, n) + \widehat{V}(\widetilde{\theta}_n, n). \end{aligned} \tag{48}$$

By virtue of (5)

$$|\widetilde{V}(\widetilde{\theta}, n)| \le C\frac{1}{n^\gamma}(1 + V(\widetilde{\theta}))$$

and

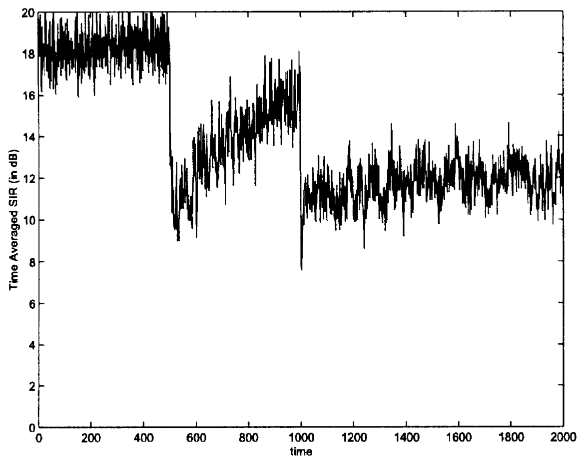
$$|\widehat{V}(\widetilde{\theta}, n)| \le C\frac{1}{n^\gamma}(1 + V(\widetilde{\theta})) \tag{49}$$

so the perturbations are small. We show that they also lead to desired cancellations. Direct computation yields that

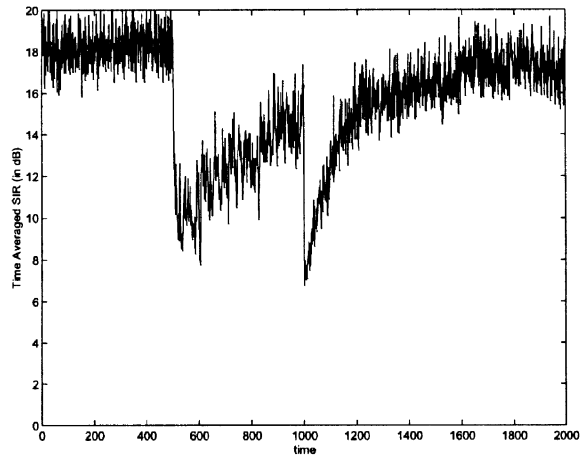
$$\begin{aligned}&\mathbf{E}_n\widetilde{V}(\widetilde{\theta}_{n+1}, n+1) - \widetilde{V}(\widetilde{\theta}_n, n)\\ &\quad = \mathbf{E}_n\widetilde{V}(\widetilde{\theta}_{n+1}, n+1) - \mathbf{E}_n\widetilde{V}(\widetilde{\theta}_n, n+1)\\ &\qquad + \mathbf{E}_n\widetilde{V}(\widetilde{\theta}_n, n+1) - \widetilde{V}(\widetilde{\theta}_n, n)\\ &\quad = -\frac{1}{n^\gamma}\widetilde{\theta}_n'(A_n - D_{\rm SE})\widetilde{\theta}_n + O\left(\frac{1}{n^{2\gamma}}\right)(1 + V(\widetilde{\theta}_n)). \end{aligned} \tag{50}$$

Similarly, we obtain

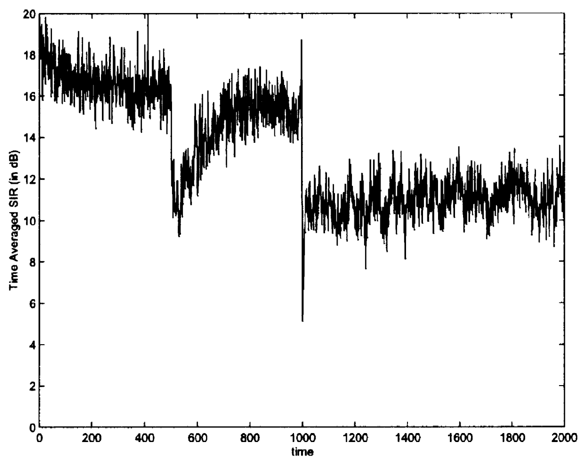
$$\begin{aligned}&\mathbf{E}_n\widehat{V}(\widetilde{\theta}_{n+1}, n+1) - \widehat{V}(\widetilde{\theta}_n, n)\\ &\quad = -\left(\frac{1}{n^\gamma}\right)\widetilde{\theta}_n'\mathbf{E}_n f_n(\theta_*) + O\left(\frac{1}{n^{2\gamma}}\right)(1 + V(\widetilde{\theta}_n)). \end{aligned} \tag{51}$$



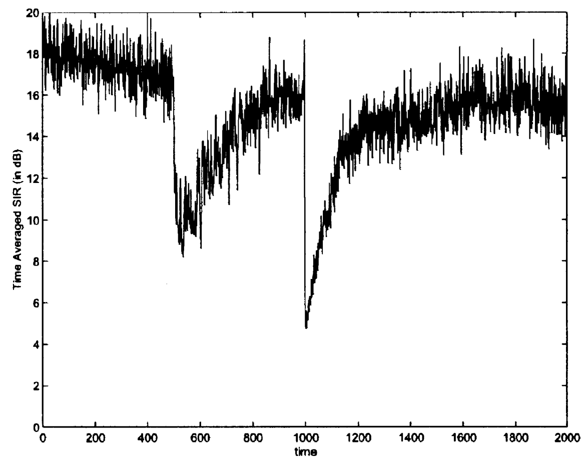
(a) Blind LMS with $\varepsilon = 10^{-3}$



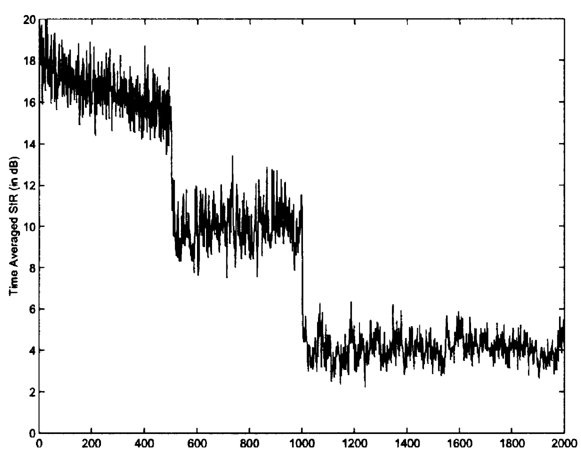
(b) Blind Averaged LMS $\varepsilon = 10^{-3}$, $\rho = 10^{-2}$



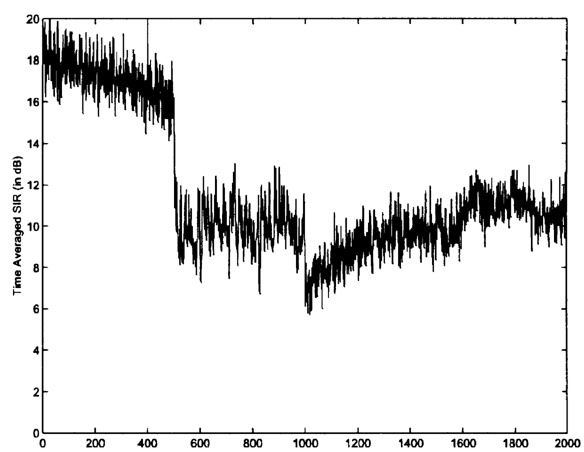
(c) Blind Sign Regressor with $\varepsilon = 2 \times 10^{-3}$



(d) Blind Averaged Sign Regressor with $\varepsilon = 2 \times 10^{-3}$, $\rho = 10^{-2}$.



(e) Blind Sign Error $\varepsilon = 2 \times 10^{-3}$



(f) Blind Averaged Sign Error $\varepsilon = 2 \times 10^{-3}$, $\rho = 10^{-2}$

Fig. 2. Average SIR versus time for time-varying NBI and MAI suppression. The parameters are specified in Section IV-D.

The Hurwitz assumption on $D_{\rm SE}$ implies that there is a $\lambda > 0$ such that $\widetilde{\theta}'_n D_{\rm SE}\widetilde{\theta}_n \le -\lambda V(\widetilde{\theta}_n)$ for some $\lambda > 0$. It follows that there is a λ_0 with $0 < \lambda_0 < \lambda$ such that

$$\frac{1}{n^\gamma}\widetilde{\theta}'_n D_{\rm SE}\widetilde{\theta}_n + \left(\frac{1}{n^\gamma}\right) o(|\widetilde{\theta}_n|^2) + O\left(\frac{1}{n^{2\gamma}}\right) V(\widetilde{\theta}_n) \le -\frac{\lambda_0}{n^\gamma} V(\widetilde{\theta}_n). \quad (52)$$

Using (47)–(52)

$$\mathbf{E}_n W(n+1) - W(n) \le -(\lambda_0/n^\gamma)V(\widetilde{\theta}_n) + O(1/n^{2\gamma})(1 + V(\widetilde{\theta}_n) + \mathbf{E}_n \varphi'_n \varphi_n).$$

It follows from (49) that for some $0 < \lambda_1 < \lambda_0$

$$\mathbf{E}_n W(n+1) \le \left(1 - \frac{\lambda_1}{n^\gamma}\right) W(n) + O\left(\frac{1}{n^{2\gamma}}\right)(1 + \mathbf{E}_n \varphi'_n \varphi_n). \quad (53)$$

Taking expectation in (53) and iterating on the resulting inequality

$$\begin{aligned}\mathbf{E}W(n+1) \le& \prod_{j=1}^{n}\left(1 - \frac{\lambda_1}{j^\gamma}\right)\mathbf{E}W(1) \\ &+ \sum_{j=1}^{n}\prod_{i=j}^{n}\left(1 - \frac{\lambda_1}{i^\gamma}\right) O\left(\frac{1}{j^{2\gamma}}\right) \\ =& O\left(\frac{1}{n^\gamma}\right).\end{aligned}$$

Moreover, by using (49), we also have $\mathbf{E}V(\widetilde{\theta}_n) = O(1/n^\gamma)$. Furthermore, the bounds derived are uniform in n. This concludes the proof. □

B. Proof of Lemma 2.5

Using telescoping

$$\begin{aligned}\sum_{j=m}^{k}\Psi(k|j)\frac{1}{j^\gamma} =& -\sum_{j=m}^{k}[\Psi(k|j) - \Psi(k|j-1)] \\ =& \Psi(k|m-1)\frac{1}{m^\gamma} - I. \quad (54)\end{aligned}$$

Since $D_{\rm SE}$ is stable, there is a $\lambda > 0$ such that

$$|\Psi(k|j)| \le \prod_{l=j+1}^{k}(I - \lambda/l^\gamma), \qquad \text{for } k \ge j+1.$$

Thus, $\Psi(k|m-1)$ is bounded yielding the first inequality in (14). The second equation in (14) is proved in [4, p. 9]. □

The following proofs are carried out by using bounded mixing conditions. The proofs under martingale difference signals or $\mathrm{MA}(p)$ processes are much simpler.

C. Proof of Lemma 2.6

1) We first show that under (**A**) and (15), $e_n^1(t) + e_n^2(t) + e_n^3(t) \to 0$ in probability as $n \to \infty$ uniformly in t. Since $\{\widetilde{\theta}_n\}$ is bounded w.p. 1 (Theorem 2.2), $e_n^1 \to 0$ w.p. 1 by (15). By (13)

$$\begin{aligned}&\frac{1}{\sqrt{n+1}}\sum_{k=m}^{\lfloor nt\rfloor}\psi(k|m-1) \\ &= \frac{(m-1)^{\frac{\gamma}{2}}}{\sqrt{n+1}}\Big[-D_{\rm SE} + \Phi(\lfloor nt\rfloor|m-1)\Big]\left((m-1)^{\gamma/2}\widetilde{\theta}_m\right).\end{aligned}$$

Since

$$E|(m-1)^{\gamma/2}\widetilde{\theta}_m| = O(1)$$

by Theorem 2.4, and $(m-1)^{\gamma/2}/\sqrt{n+1} \to 0$, $e_n^2 \to 0$ in probability.

Using $E|\widetilde{\theta}_j|^2 = O(j^{-\gamma})$ and interchanging the orders of summations, we obtain

$$\begin{aligned}E\frac{C}{\sqrt{n+1}}&\sum_{k=m}^{\lfloor nt\rfloor}\sum_{j=m}^{k}\frac{1}{j^\gamma}|\Psi(k|j)||\widetilde{\theta}_j|^2 \\ &\le \frac{C}{\sqrt{n+1}}\sum_{j=m}^{\lfloor nt\rfloor}\frac{1}{j^{2\gamma}}\sum_{k=j}^{\lfloor nt\rfloor}|\Psi(k|j)| \\ &= O(n^{(1/2)-\gamma}) \to 0\end{aligned}$$

by (13) and (14).

2) We show that for $i = 0{,}4$, and 5

$$e_n^i(t) = -\frac{D_{\rm SE}^{-1}}{\sqrt{n+1}}\sum_{j=m}^{\lfloor nt\rfloor}\zeta_j^i + o(1)$$

where $\zeta_j^0 = f_j(\theta_*)$, $\zeta_j^4 = (A_j - D_{\rm SE})\widetilde{\theta}_j$, $\zeta_j^5 = \delta M_j$, and $o(1) \to 0$ in probability uniformly in t. In fact, using the Dirichlet formula to interchange the order of summations in $e_n^0(t)$, $e_n^4(t)$, and $e_n^5(t)$, we obtain

$$\begin{aligned}e_n^0(t) &+ e_n^4(t) + e_n^5(t) \\ &= \frac{1}{\sqrt{n+1}}\sum_{j=m}^{\lfloor nt\rfloor}\sum_{k=j}^{\lfloor nt\rfloor}\frac{1}{j^\gamma}\Psi(k|j)f_j(\theta_*) \\ &\quad + \frac{1}{\sqrt{n+1}}\sum_{j=m}^{\lfloor nt\rfloor}\sum_{k=j}^{\lfloor nt\rfloor}\frac{1}{j^\gamma}\Psi(k|j)(A_j - D_{\rm SE})\widetilde{\theta}_j \\ &\quad + \frac{1}{\sqrt{n+1}}\sum_{j=m}^{\lfloor nt\rfloor}\sum_{k=j}^{\lfloor nt\rfloor}\frac{1}{j^\gamma}\Psi(k|j)\delta M_j.\end{aligned}$$

Note that by virtue of (13), for $i = 0,\ 4,\ 5$

$$e_n^i(t) = -\frac{D_{\rm SE}}{\sqrt{n+1}}\sum_{j=m}^{\lfloor nt\rfloor}\zeta_j^i + \frac{1}{\sqrt{n+1}}\sum_{j=m}^{\lfloor nt\rfloor}\Phi(\lfloor nt\rfloor|j)\zeta_j^i. \quad (55)$$

First, we have that as $n \to \infty$

$$\begin{aligned}E&\left|\frac{1}{\sqrt{n+1}}\sum_{j=m}^{\lfloor nt\rfloor}\Phi(\lfloor nt\rfloor|j)\zeta_j^0\right|^2 \\ &\le \frac{C}{n+1}\sum_{j=m}^{\lfloor nt\rfloor}|\Phi(\lfloor nt\rfloor|j)|\sum_{k>j}^{\lfloor nt\rfloor}|\Phi(\lfloor nt\rfloor|k)||Ef'_j(\theta_*)f_k(\theta_*)| \\ &\le \frac{C}{n+1}\sum_{j=m}^{\lfloor nt\rfloor}|\Phi(\lfloor nt\rfloor|j)|\sum_{k>j}^{\lfloor nt\rfloor}\eta^{1/2}(k-j) \to 0 \quad (56)\end{aligned}$$

by virtue of the mixing property of $\{f_j(\theta_*)\}$, and by Lemma 2.5.

For the corresponding term in $e_n^5(t)$, since $\{\delta M_n\}$ is a martingale difference sequence, it is uncorrelated and by Lemma 2.5, as $n \to \infty$

$$E\left|\frac{1}{\sqrt{n+1}}\sum_{j=m}^{\lfloor nt\rfloor}\Phi(\lfloor nt\rfloor|j)\delta M_j\right|^2 \le \frac{C}{n+1}\sum_{j=m}^{\lfloor nt\rfloor}|\Phi(\lfloor nt\rfloor|j)|^2 E|\delta M_j|^2 \to 0. \quad (57)$$

Finally, we come to the terms in $e_n^4(t)$. By using Theorem 2.4, $E|\widetilde{\theta}_j| = O(j^{-\gamma/2})$. This together with Lemma 2.5 and the boundedness of the signals yields that as $n \to \infty$

$$\begin{aligned} E|e_n^4(t)| \le& \frac{1}{n+1}\sum_{j=m}^{\lfloor nt\rfloor}|\Phi(\lfloor nt\rfloor|j)|\sum_{k>j}|\Phi(\lfloor nt\rfloor|k)| \\ &\times E^{1/2}|(A_j - D_{\rm SE})(A_k - D_{\rm SE})|^2 E^{1/2}|\widetilde{\theta}_j|^2 \\ &\times E^{1/2}|\widetilde{\theta}_k|^2 \\ \le& C\sum_{j=m}^{\lfloor nt\rfloor}\frac{1}{j^\gamma}|\Phi(\lfloor nt\rfloor|j)| \\ &\times\left(\frac{1}{n+1}\sum_{k=m(t+t_n)}^{m(t+s+t_n)-1}|\Phi(\lfloor nt\rfloor|k)|\right) \\ \to& 0. \end{aligned} \quad (58)$$

Combining (56)–(58), the desired result follows.

3) We next show that $e^4(t)$ and $e_n^5(t)$ contribute nothing to the limit, so only $e_n^0(t)$ is asymptotically important. To prove $e_n^4(t) \to 0$ in probability uniformly in t, it suffices to consider $(1/\sqrt{n+1})\sum_{j=m}^{\lfloor nt\rfloor}(A_j - D_{\rm SE})\widetilde{\theta}_j$ in accordance with step 2) above. First, note (59) at the bottom of the page. By using the mixing inequality (6), we have

$$\begin{aligned} &\frac{C}{n+1}\left|\sum_{j=m}^{\lfloor nt\rfloor}\sum_{k\ge j}^{\lfloor nt\rfloor}E\widetilde{\theta}_j'(A_j - D_{\rm SE})'(A_k - D_{\rm SE})\widetilde{\theta}_m\right| \\ &\le \frac{C}{n+1}\sum_{j=m}^{\lfloor nt\rfloor}E|\widetilde{\theta}_j||A_j - D_{\rm SE}| \\ &\quad\times\left|\sum_{k\ge j}^{\lfloor nt\rfloor}E_j(A_k - D_{\rm SE})\right||\widetilde{\theta}_m| \\ &\le \frac{C}{n+1}\sum_{j=m}^{\lfloor nt\rfloor}\frac{1}{j^{\gamma/2}} = O(n^{-\gamma/2}) \to 0, \qquad \text{as } n \to \infty. \end{aligned}$$

Moreover

$$\begin{aligned} &\frac{C}{n+1}\left|\sum_{j=m}^{\lfloor nt\rfloor}\sum_{k\ge j}^{\lfloor nt\rfloor}E\widetilde{\theta}_j'(A_j - D_{\rm SE})'(A_k - D_{\rm SE})\sum_{l\ge j}^{k-1}\frac{1}{l^\gamma}f_l(\theta_l)\right| \\ &\le \frac{C}{n+1}\sum_{j=m}^{\lfloor nt\rfloor}E|\widetilde{\theta}_j||A_j - D_{\rm SE}| \\ &\quad\times\sum_{l\ge j}^{\lfloor nt\rfloor}\left|\sum_{k=l}^{\lfloor nt\rfloor}E_l(A_k - D_{\rm SE})\right|\frac{1}{l^\gamma}|f_l(\theta_l)| \\ &\le \frac{C}{n+1}\sum_{j=m}^{\lfloor nt\rfloor}\frac{1}{j^{\gamma/2}}\sum_{l=m}^{\lfloor nt\rfloor}\frac{1}{l^\gamma} \\ &= O(n^{1-3\gamma/2}) \to 0, \qquad \text{as } n \to \infty. \end{aligned}$$

The above estimates and (59) then lead to that $e_n^4(t) \to 0$ as $n \to \infty$.

Finally, for $e_n^5(t)$, since $\{\delta M_n\}$ is a martingale difference sequence

$$\begin{aligned} E\left|\frac{1}{\sqrt{n+1}}\sum_{j=m}^{\lfloor nt\rfloor}\delta M_j\right|^2 &= \frac{1}{n+1}\sum_{j=m}^{\lfloor nt\rfloor}E(\delta M_j)'\delta M_j \\ &= \frac{1}{n+1}\sum_{j=m}^{\lfloor nt\rfloor}E[E_j(\delta M_j)'\delta M_j]. \end{aligned} \quad (60)$$

Note that $\widetilde{\theta}_n \to 0$ as $n \to \infty$, that $m = m(n) \to \infty$ as $n \to \infty$, that $j \ge m = m(n)$, and that $E_j(\delta M_j)'\delta M_j$ is continuous in θ by condition (**A**). Note also that as a function of θ, $f_n(\theta) - f_n(\theta_*)$ is bounded and $\widetilde{f}_n(\theta) - \widetilde{f}_n(\theta_*)$ is dominated by a linear function of $\theta - \theta_*$. In view of (10), the dominated convergence theorem and $E_j(\delta M_j)'\delta M_j \to 0$ as $n \to \infty$ yield that the last term in (60) goes to 0 uniformly in t. □

D. Proof of Theorem 2.7

By virtue of Lemma 2.6, and the choice of $m = m(n)$,

$$(1/\sqrt{n+1})\sum_{j=m}^{\lfloor nt\rfloor}f_j(\theta_*) = (1/\sqrt{n+1})\sum_{j=1}^{\lfloor nt\rfloor}f_j(\theta_*) + o(1)$$

as $n \to \infty$, where $o(1) \to 0$ in probability. Thus, $w_{n+1}(\cdot)$ is also tight in $D^r[0, 1]$. Moreover, (16) and Slutsky's theorem yield the desired result. □

$$\begin{aligned} E\left|\frac{1}{\sqrt{n+1}}\sum_{j=m}^{\lfloor nt\rfloor}(A_j - D_{\rm SE})\widetilde{\theta}_j\right| &= \frac{1}{n+1}\sum_{j=m}^{\lfloor nt\rfloor}\sum_{k=m}^{\lfloor nt\rfloor}E\widetilde{\theta}_j'(A_j - D_{\rm SE})'(A_k - D_{\rm SE})\widetilde{\theta}_k \\ &\le \frac{C}{n+1}\left|\sum_{j=m}^{\lfloor nt\rfloor}\sum_{k\ge j}^{\lfloor nt\rfloor}E\widetilde{\theta}_j'(A_j - D_{\rm SE})'(A_k - D_{\rm SE})\sum_{l\ge j}^{k-1}\frac{1}{l^\gamma}f_l(\theta_l)\right| \\ &\quad + \frac{C}{n+1}\left|\sum_{j=m}^{\lfloor nt\rfloor}\sum_{k\ge j}^{\lfloor nt\rfloor}E\widetilde{\theta}_j'(A_j - D_{\rm SE})'(A_k - D_{\rm SE})\widetilde{\theta}_m\right|. \end{aligned} \quad (59)$$

ACKNOWLEDGMENT

The authors are grateful to the referees and the Associate Editor. They thank them for the detailed comments, suggestions, and corrections on an early version of the manuscript.

REFERENCES

[1] A. Benveniste, M. Gooursat, and G. Ruget, "Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms with applications to data communication algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 1042–1058, Dec. 1980.

[2] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. Berlin, Germany: Springer-Verlag, 1990.

[3] P. Billingsley, *Convergence of Probability Measures*, 2nd ed. New York: Wiley, 1999.

[4] H. F. Chen, "Asymptotically efficient stochastic approximation," *Stochastics Stochastics Reports*, vol. 43, pp. 1–16, 1993.

[5] H. F. Chen and G. Yin, "Asymptotic properties of sign algorithms for adaptive filtering," preprint, 2000.

[6] S. H. Cho and V. J. Mathews, "Tracking analysis of the sign algorithm in nonstationary environments," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 2046–2057, Dec 1990.

[7] K. L. Chung, "On a stochastic approximation method," *Ann. Math. Statist.*, pp. 463–483, 1954.

[8] S. N. Ethier and T. G. Kurtz, *Markov Processes—Characterization and Convergence*. New York: Wiley, 1986.

[9] E. Eweda, "A tight upper bound of the average absolute error in a constant step size sign algorithm," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, pp. 1774–1776, Nov. 1989.

[10] ——, "Convergence of the sign algorithm for adaptive filtering with correlated data," *IEEE Trans. Inform. Theory*, vol. 37, pp. 1450–1457, Mar. 1991.

[11] E. Eweda and O. Macchi, "Quadratic mean and almost-sure convergence of unbounded stochastic approximation algorithms with correlated observations," *Ann. Henri Poincaré*, vol. 19, pp. 235–255, 1983.

[12] A. Gersho, "Adaptive filtering with binary reinforcement," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 191–199, Mar. 1984.

[13] S. Haykin, "Adaptive filter theory," in *Information and System Sciences Series*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1991.

[14] M. L. Honig, U. Madhow, and S. Verdu, "Adaptive blind multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944–960, July 1995.

[15] M. L. Honig and H. V. Poor, "Adaptive interference suppression in wireless communication systems," in *Wireless Communications: Signal Processing Perspectives*, H. V. Poor and G. W. Wornell, Eds. Englewood Cliffs, NJ: Prentice-Hall, 1998.

[16] V. Krishnamurthy, "Averaged stochastic gradient algorithms for adaptive blind multiuser detection in DS/CDMA systems," *IEEE Trans. Commun.*, vol. 48, pp. 125–134, Jan. 2000.

[17] V. Krishnamurthy and A. Logothetis, "Adaptive nonlinear filters for narrowband interference suppression in spread spectrum CDMA systems," *IEEE Trans. Commun.*, vol. 47, pp. 742–753, May 1999.

[18] V. Krishnamurthy, X. Wang, and G. Yin, "Spreading code optimization and adaptation in CDMA via discrete stochastic approximation," preprint, vol. 2002.

[19] H. J. Kushner, *Approximation and Weak Convergence Methods for Random Processes, With Applications to Stochastic Systems Theory*. Cambridge, MA: MIT Press, 1984.

[20] H. J. Kushner and A. Shwartz, "Weak convergence and asymptotic properties of adaptive filters with constant gains," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 177–182, Mar. 1984.

[21] H. J. Kushner and J. Yang, "Stochastic approximation with averaging of iterates: Optimal asymptotic rate of convergence for general processes," *SIAM J. Contr. Optimiz.*, vol. 31, no. 4, pp. 1045–1062, 1993.

[22] H. J. Kushner and G. Yin, *Stochastic Approximation Algorithms and Applications*. New York: Springer-Verlag, 1997.

[23] L. Ljung, "Aspects on accelerated convergence in stochastic approximation schemes," in *Proc. 33rd IEEE Conf. Decision and Control*, Lake Buena Vista, FL, 1994, pp. 1649–1652.

[24] L. Ljung and S. Gunnarsson, "Adaptive tracking in system identification—A survey," *Automatica*, vol. 26, no. 1, pp. 7–22, 1990.

[25] O. Macchi and E. Eweda, "Convergence analysis of self-adaptive equalizers," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 161–176, Mar. 1984.

[26] M. Peletier, "Asymptotic almost efficiency of averaged stochastic approximation algorithms," *SIAM J. Contr. Optimiz.*, vol. 39, pp. 49–72, 2000.

[27] H. Pezeshki-Esfahani and A. J. Heunis, "Strong diffusion approximations for recursive stochastic algorithms," *IEEE Tran. Inform. Theory*, vol. 43, pp. 512–523, Mar. 1997.

[28] B. T. Polyak, "New method of stochastic approximation type," *Automat. Remote Control*, vol. 51, pp. 937–946, 1990.

[29] B. T. Polyak and Y. Z. Tsypkin, "Robust pseudogradient adaptation algorithms," *Automat. Remote Control*, vol. 41, pp. 1404–1409, 1981.

[30] H. V. Poor and X. Wang, "Code-aided interference suppression for DS/CDMA communications-Part I: Interference suppression capability," *IEEE Trans. Commun.*, vol. 45, pp. 1101–1111, Sept. 1997.

[31] ——, "Code-aided interference suppression for DS/CDMA communications-Part II: Parallel blind adaptive implementations," *IEEE Trans. Commun.*, vol. 45, pp. 1112–1122, Sept. 1997.

[32] D. Ruppert, "Stochastic approximation," in *Handbook in Sequential Analysis*, B. K. Ghosh and P. K. Sen, Eds. New York: Marcel Dekker, 1991, pp. 503–529.

[33] V. Solo and X. Kong, *Adaptive Signal Processing Algorithms*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[34] S. Verdú, "Multiuser detection," in *Advances in Statistical Signal Processing*, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT: JAI, 1993.

[35] N. A. M. Verhoeckx, H. C. van den Elzen, F. A. M. Snijders, and P. J. van Gerwen, "Digital echo cancellation for baseband data transmission," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 27, pp. 761–781, June 1979.

[36] S. Verdú, *Multiuser Detection*. Cambridge, MA: MIT Press, 1998.

[37] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[38] G. Yin, "Adaptive filtering with averaging," in *IMA Volumes in Mathematics and Its Applications*, G. Goodwin, K. Aström, and P. R. Kumar, Eds. Berlin, Germany: Springer-Verlag, 1995, vol. 74, pp. 375–396.