Blind Adaptive Multiuser Detection for Cellular Systems Using Stochastic Approximation With Averaging

Deepak Das, *Member, IEEE* and Mahesh K. Varanasi, *Senior Member, IEEE*

***Abstract*—In this paper, we consider blind adaptive multiuser detection in correlated waveform multiple-access-based cellular radio networks. A common stochastic approximation (SA)-based framework is proposed from which three blind adaptive algorithms for linear minimum mean squared error detection are obtained. Two of them coincide with previously proposed algorithms and the third is shown to be best suited for implementation at a base station. The work here also improves these SA-based adaptation algorithms in the context of cellular radio networks, in terms of convergence properties by using the more recent results on the SA technique with *averaging*. Convergence issues of the different adaptations are investigated and numerical examples are presented to demonstrate the performance improvement due to averaging.**

***Index Terms*—Blind adaptive detectors, correlated waveform multiple-access, cellular wireless systems, multiuser detection, stochastic approximation.**

I. Introduction

CELLULAR systems, which have received much attention in recent years in the wireless communications literature, provide a natural application for multiuser detection. It is well known that multiuser detectors yield significant improvements in performance over single-user detectors in correlated waveform multiple access (CWMA) systems (cf. [1]). Since optimum multiuser detection is generally exponentially complex in the number of users, subsequent research has yielded several suboptimal detectors, including the two well known linear detectors—the decorrelating [2] and the minimum mean-squared-error (MMSE) [3] detectors. When all system parameters are known, these detectors, in a K-user system, would require the inversion of a $K \times K$ matrix.

The study of adaptive multiuser detection schemes is motivated mainly by incomplete knowledge of the system parameters. Such schemes generate estimates of the multiuser detectors stochastically from the received data. They can be designed to use computationally simple recursions based on only the available system parameters and, additionally, to proceed without the use of training sequences. Among several other schemes, a blind adaptive algorithm based on the minimization of the output energy was proposed in [4] (and was shown to adapt to the MMSE detector), and blind adaptive algorithms for decorrelation and MMSE detection were presented in [5]. These algorithms are guided by only the signal of the user of interest and we refer to them as the single-signal (SS) adaptations in this work. They are, hence, suitable for fully distributed implementation as at a mobile.

In this paper, we focus on the up-link and consider a common framework for stochastic approximation (SA)-based blind adaptive multiuser detectors at the base station. The SS blind adaptive detectors of [4] and [5] can be easily obtained from our common framework. In general, however, a base station will likely have more system information (such as the signal and the timing information of users in its own cell) and computational capacity than an individual user. Hence, we propose a new algorithm that takes particular advantage of this additional knowledge. This algorithm, therefore, converges more quickly to the MMSE solution than its SS counterparts. In the case where the base station has the required information for only a subset of all its in-cell users, the rest of this discussion can be equivalently done in terms of known and unknown users.

Although several existing adaptive multiuser detection algorithms are applicable to up-link detection (cf. [6]–[9]), we have limited a comparison of our algorithms with only those that use the subspace tracking technique [10]. Specifically, [11] and [12] have recently proposed adaptive detection schemes that are suitable for implementation at base stations. These papers present a hybrid multiuser detection scheme that combines a decorrelative projection out of the known signal space of the in-cell users, and blind adaptive MMSE detection to combat the unknown out-of-cell multiple-access interference (MAI). This hybrid solution is however constrained by the dimensional requirements that make the decorrelating projection feasible, and also by the computational complexity of subspace tracking. Of course, exploiting the knowledge of the entire in-cell signal space helps these blind multiuser detectors to show improved convergence properties. However, it is to be noted that the linear MMSE detector that accounts for all the users (both in-cell and out-of-cell) is most suitable for a cellular system because it achieves the maximum signal-to-interference ratio (SIR) among linear detectors (cf. [3]), and with a large number of in-cell users in a typical cellular network, even a partial decorrelation strategy may be dimensionally infeasible.

Manuscript received June 6, 2001; revised September 10, 2001. This work was supported in part by the National Science Foundation (NSF) under Grant NCR-9725778 and Army Research Office (ARO) under Grant DADD 19-99-1-0291. This paper was presented in part at the IEEE International Symposium of Information Theory, Sorrento, Italy, June 2000.

D. Das was with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425 USA, and is now with Lucent Technologies, Whippany, NJ 07981 USA (e-mail: deepak@dsp.colorado.edu; ddas2@lucent.com).

M. K. Varanasi is with the Department of Electrical and Computer Engineering, University of Colorado, Boulder, CO 80309-0425 USA (e-mail: varanasi@colorado.edu; varanasi@schof.colorado.edu).

Publisher Item Identifier S 0733-8716(02)00989-7.

The rate of convergence of SA methods has for a long time been of great interest both from a theoretical and a practical point of view, and has been largely addressed in terms of adaptively optimizing the step-size sequence used. In a couple of fundamental papers on the SA technique, Ruppert [13] and Polyak [14] showed under certain assumptions that with a decreasing step-size sequence that decays slower than the usual Robbins-Monro step-size sequence, significant improvement in convergence can be achieved if the output of the conventional SA algorithm is averaged. In this paper, we further improve the efficiency of blind adaptive multiuser detection on the up-link in a cellular radio network by using the concept of averaging. We analytically show the convergence of the "averaged" adaptive schemes, and numerically demonstrate the resulting improvement in performance.

II. System Model

We consider a cellular network model of [16] in which there are B base stations and K active users with K_j users assigned to base j. For simplicity of presentation, we assume that each base employs a common set of N matched filters matched to orthonormal basis functions that span the entire signal space. The discussion in this paper remains valid for the case where any base station, say j, uses a different set of $N_j \leq N$ matched filters matched to the orthonormal basis of some possibly "reduced" signal space that includes the signals received from the users within its cell. While the transmissions of the mobile users of a particular cell arrive at the bases of other cells symbol-asynchronously, it is assumed, for the sake of simplicity, that they arrive at their own base symbol-synchronously. It is also assumed that while the receiver at base j has knowledge of the common timing of the received signals of its own cell, it does not have the timing information associated with signals of other cells.

The discrete-time model for the N matched filter outputs at base j can be expressed as

$$\mathbf{y}_j = \sum_{i=1}^{K_j} \sqrt{w_{ij} g_{ijj}} \mathbf{s}_{ij} b_{ij} + \sum_{l \neq j}^{B} \sum_{i=1}^{K_l} \sqrt{w_{il} g_{ilj}} \times (\mathbf{s}_{il}^- b_{il}^- + \mathbf{s}_{il}^+ b_{il}^+) + \chi_j \quad (1)$$

where w_{il} and b_{il} denote the transmit power and the transmitted symbol (binary antipodal signalling, i.e., ± 1), respectively, of user i of base l. The channel gain of the ith user of base l to base j is denoted by g_{ilj}. The vectors $\mathbf{s}_{ij}$ denote the vector representations (the "signature sequence") of the signal of user i of base j. The vectors $\mathbf{s}_{il}^-$ and $\mathbf{s}_{il}^+$ denote the vector representations of the segments of the signals associated with the two symbols of user i of base l, b_{il}^- and b_{il}^+, respectively, that overlap with the symbol interval of interest at base j. χ_j is an N-dimensional zero-mean Gaussian random vector with a covariance matrix equal to $\sigma_j^2 \mathbf{I}$.

Let us define the signal matrix $\mathbf{S}_j = [\mathbf{s}_{1j}\ \mathbf{s}_{2j}\ \cdots\ \mathbf{s}_{K_j j}]$ whose columns are, therefore, the signature sequences of the users of base j. Also, for $l \neq j$, define the signal matrix $\mathbf{S}_l^- = [\mathbf{s}_{1l}^-\ \mathbf{s}_{2l}^-\ \cdots\ \mathbf{s}_{K_l l}^-]$ and $\mathbf{S}_l^+ = [\mathbf{s}_{1l}^+\ \mathbf{s}_{2l}^+\ \cdots\ \mathbf{s}_{K_l l}^+]$. Let the diagonal matrix of all the received user energies be defined as $\mathbf{W}_j = \mathrm{diag}[w_{1j} g_{11j} \ldots w_{K_1 1} g_{K_1 1 j}, \ldots, w_{1B} g_{1Bj} \ldots w_{K_B B} g_{K_B B j}]$.

III. Linear Multiuser Detection

Without loss of generality, we will consider detection at base 1. Let us denote the diagonal matrix of in-cell user energies as $\mathbf{W}_{11} = \mathrm{diag}[w_{11} g_{111}, \ldots, w_{K_1 1} g_{K_1 11}]$ and the diagonal matrix of the out-of-cell interfering user energies (as received at base 1) as $\bar{\mathbf{W}}_{11} = \mathrm{diag}[w_{12} g_{121}, \ldots, w_{K_2 2} g_{K_2 21}, \ldots, w_{1B} g_{1B1}, \ldots, w_{K_B B} g_{K_B B1}]$. A linear detector for user k in cell $1, 0 < k \leq K_1$, can be represented by the vector, π_{k1}, such that the decision for user k is given by

$$\hat{b}_{k1} = \mathrm{sgn}(\langle \mathbf{y}_1, \pi_{k1} \rangle) \quad (2)$$

where $\langle \psi_1, \psi_2 \rangle$ denotes the usual inner product of vectors ψ_1 and ψ_2. Let E denote the "Expectation" operator. Noting that

$$\begin{aligned} \mathbf{A} &\triangleq E\left[\mathbf{y}_1 \mathbf{y}_1^T\right] = \mathbf{S}_1 \mathbf{W}_{11} \mathbf{S}_1^T \\ &\quad + \sum_{l \neq 1}^{B} \sum_{i=1}^{K_l} w_{il} g_{il1} \mathbf{s}_{il}^- (\mathbf{s}_{il}^-)^T \\ &\quad + \sum_{l \neq 1}^{B} \sum_{i=1}^{K_l} w_{il} g_{il1} \mathbf{s}_{il}^+ (\mathbf{s}_{il}^+)^T + \sigma_1^2 \mathbf{I}, \\ &\triangleq \mathbf{S}_1 \mathbf{W}_{11} \mathbf{S}_1^T + \bar{\mathbf{S}}_1^- \bar{\mathbf{W}}_{11} (\bar{\mathbf{S}}_1^-)^T \\ &\quad + \bar{\mathbf{S}}_1^+ \bar{\mathbf{W}}_{11} (\bar{\mathbf{S}}_1^+)^T + \sigma_1^2 \mathbf{I} \end{aligned} \quad (3)$$

the solution over all linear detectors of the minimization of the mean squared error, i.e., $\min_{\pi_{k1}} E[(b_{k1} - \langle \mathbf{y}_1, \pi_{k1} \rangle)^2]$, is the linear MMSE multiuser detector. This is given with scaling as the unique solution to the equation (cf. [1])

$$\mathbf{A} \mathbf{c}_{k1} = \mathbf{s}_{k1} \quad (4)$$

where we note that positive scaling does not affect the decision rule in (2).

IV. Blind Adaptive Multiuser Detection

In this section, we will present a common framework for obtaining different blind adaptation rules to estimate $\mathbf{c}_{k1}$.

From the theory of iterative methods to solve linear equations [17], we can form the following general deterministic iteration:

$$\mathbf{c}_{k1}(n+1) = (\mathbf{I} - \mu_n \mathbf{Q} \mathbf{A}) \mathbf{c}_{k1}(n) + \mu \mathbf{Q} \mathbf{s}_{k1} \quad (5)$$

that converges to the desired solution $\mathbf{c}_{k1}$ of (4), where $\mathbf{Q}$ is a nonsingular matrix, whose inverse is called the splitting matrix. A stochastic version of (5) is obtained by replacing $\mathbf{A}$ by its instantaneous stochastic estimate $\mathbf{y}_1(n+1)\mathbf{y}_1(n+1)^T$, i.e.

$$\mathbf{c}_{k1}(n+1) = (\mathbf{I} - \mu_n \mathbf{Q} \mathbf{y}_1(n+1) \mathbf{y}_1(n+1)^T) \times \mathbf{c}_{k1}(n) + \mu \mathbf{Q} \mathbf{s}_{k1}. \quad (6)$$

Replacing $\mathbf{Q}$ by the identity matrix leads to the following simple SA-based algorithm [5]:

$$\mathbf{c}_{k1}(n+1) = (\mathbf{I} - \mu_n \mathbf{y}_1(n+1) \mathbf{y}_1(n+1)^T) \times \mathbf{c}_{k1}(n) + \mu_n \mathbf{s}_{k1} \quad (7)$$

where $\mu_n > 0$ is a suitably chosen fixed or decreasing step-size sequence, and the computation involved is $O(N)$. The choice of step-size will be dictated by the convergence analysis in Section VI.

Let $\mathbf{P_X} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ denote the projection operator onto the column space of matrix $\mathbf{X}$ and $\mathbf{P}_\mathbf{X}^\perp = \mathbf{I} - \mathbf{P_X}$ the projection operator onto the orthogonal subspace. Using the canonical representation for the (scaled) MMSE linear detector

$$\frac{1}{\mathbf{s}_{k1}^T\mathbf{A}^{-1}\mathbf{s}_{k1}}\mathbf{c}_{k1} = \mathbf{s}_{k1} + \mathbf{p}_{k1} \quad \text{where } \mathbf{p}_{k1} \perp \mathbf{s}_{k1} \tag{8}$$

we next ask whether there is any $\mathbf{Q}$ in our common framework (6) that will help us form an adaptation rule to estimate just $\mathbf{p}_{k1}$ (since the base station already knows $\mathbf{s}_{k1}$). We find that $\mathbf{P}_{\mathbf{s}_{k1}}^\perp$ is the right choice and this yields the following algorithm:

$$\mathbf{s}_{k1} + \mathbf{p}_{k1}(n+1) = (\mathbf{I} - \mu_n\mathbf{P}_{\mathbf{s}_{k1}}^\perp\mathbf{y}_1(n+1)\mathbf{y}_1^T(n+1)) \times (\mathbf{s}_{k1} + \mathbf{p}_{k1}(n)) + \mu_n\mathbf{P}_{\mathbf{s}_{k1}}^\perp\mathbf{s}_{k1} \tag{9}$$

$$\mathbf{c}_{k1}(n+1) = \mathbf{s}_{k1} + \mathbf{p}_{k1}(n+1) \tag{10}$$

where, of course, the second term in the right-hand side (r.h.s.) of (9), $\mu_n\mathbf{P}_{\mathbf{s}_{k1}}^\perp\mathbf{s}_{k1} = \mathbf{0}$. Further, we require that $\mathbf{P}_{\mathbf{s}_{k1}}^\perp\mathbf{p}_{k1}(0) = \mathbf{p}_{k1}(0)$ and note that $\mathbf{p}_{k1}(0) = \mathbf{0}$ satisfies this condition. The adaptation rule in (9) derived quite naturally from (6) can be shown to be identical to the one in [4], where it was derived differently by minimizing the output energy. An alternative derivation of this rule as a stochastic gradient algorithm to estimate the solution of an unconstrained optimization problem can be found in [20]. Expanding $\mathbf{P}_{\mathbf{s}_{k1}}^\perp = \mathbf{I} - \mathbf{s}_{k1}\mathbf{s}_{k1}^T$ in (9), we can verify that the recursion in (9) involves $O(N)$ computation.

Both the recursions in (7) and (9) are based on the knowledge of only each user's own signal, i.e., $\mathbf{s}_{k1}$. We will refer to these SS-based adaptation algorithms as the SS algorithms. The "simpler" recursion in (7) will be referred to as the 'S-SS' rule and the "canonical" decomposition based recursion in (9) as the 'C-SS' rule.

We next ask the question whether our common iterative framework does in fact suggest new algorithms to estimate the linear MMSE receiver that uses more of the information available at the base stations. The answer is "yes." One way would be to decompose the desired solution in terms of what is known and what is unknown (an extension of the canonical decomposition idea) and apply the common framework by finding a suitable $\mathbf{Q}$ to estimate only the unknown part. We explore a different approach by first noting from the least mean square (LMS) adaptation literature [18] that the more disparate the eigenvalues of the matrix $\mathbf{A}$, the slower the rate of convergence for the recursion in (7) with a fixed step-size μ. The goal, therefore, is to effectively "cancel" the correlatedness of $\mathbf{A}$. This is typically accomplished by what is called "orthogonalization" [19], which could be achieved, for instance, by replacing $\mathbf{Q}$ by some approximation of $\mathbf{A}^{-1}$ in (6). This approximation could be a fixed one or an estimate that is stochastically updated (and improved) as the adaptation proceeds. To obtain a candidate for a fixed approximation, we first note that with $\mathbf{B} \triangleq \mathbf{S}_1\mathbf{W}_{11}\mathbf{S}_1^T + \sigma_1^2\mathbf{I}$, we can write $\mathbf{A} = \mathbf{B} + \bar{\mathbf{S}}_1^-\bar{\mathbf{W}}_{11}(\bar{\mathbf{S}}_1^-)^T + \bar{\mathbf{S}}_1^+\bar{\mathbf{W}}_{11}(\bar{\mathbf{S}}_1^+)^T$. In a cellular system where the out-of-cell interferers are typically weak,

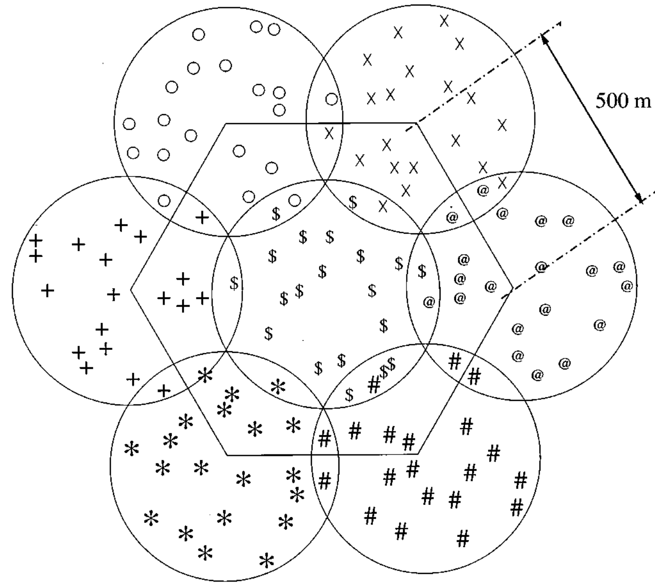

Fig. 1. Seven-cell cellular system with 12 users per cell.

$\mathbf{B}$ is the dominant part of $\mathbf{A}$, and can be considered as a course approximation of $\mathbf{A}$. In this case, $\mathbf{Q}$ can be replaced by $\mathbf{B}^{-1}$ in the stochastic version of (5), and as we will show later in this paper, this could considerably improve the rate of convergence over the SS algorithms, albeit, with some increase in computational overhead. More precisely, with the knowledge of the signals of the in-cell users, their energies, and the noise variance, we formulate the following stochastic adaptive rule for the estimation of $\mathbf{c}_{k1}$:

$$\mathbf{c}_{k1}(n+1) = (\mathbf{I} - \mu_n\mathbf{B}^{-1}\mathbf{y}_1(n+1)\mathbf{y}_1(n+1)^T) \times \mathbf{c}_{k1}(n) + \mu_n\mathbf{B}^{-1}\mathbf{s}_{k1} \tag{11}$$

where the matrix $\mathbf{B}^{-1}$ is known at the base station and, therefore, the vector $\mathbf{B}^{-1}\mathbf{s}_{k1}$ can be precomputed and stored. Note that $\mathbf{B}^{-1}\mathbf{s}_{k1}$ can be viewed as the "single-cell" MMSE receiver that ignores the out-of-cell interference. The multiplication by $\mathbf{B}^{-1}$ in the correction term in (11) makes the overall computational complexity of the recursion $O(N^2)$, which in a bandwidth efficient system with low processing gain, may be an acceptable burden. In any case, we can rewrite $\mathbf{B}^{-1}$ as $(1/\sigma_1^2)(\mathbf{I} - \mathbf{S}_1(\sigma_1^2\mathbf{W}_{11} + \mathbf{S}_1^T\mathbf{S}_1)^{-1}\mathbf{S}_1^T)$ in (11), to obtain an order of computation of $O(NK_1)$. We will refer to the "Orthogonalization"-based SA algorithm as the O-SA algorithm. Issues such as signal mismatch and imperfect noise variance estimate in the computation of $\mathbf{B}^{-1}$ can be investigated along the lines of [4] and [5], respectively, and will not be discussed in this work.

The following example compares the performance of the low-complexity SA-based techniques with the higher-complexity subspace-based techniques.

Example 1: We consider a synchronous seven-cell (12 users per cell) system as shown in Fig. 1, where the cell of interest is the cell in the center. The gains to every base station for each user is calculated using a $1/d^4$ path-loss law, where d is the distance to the base station of interest. The transmitted powers of all users are the same. The user of interest in cell 1 is received at the base station with a SNR of 28 dB. The randomly chosen unit-energy

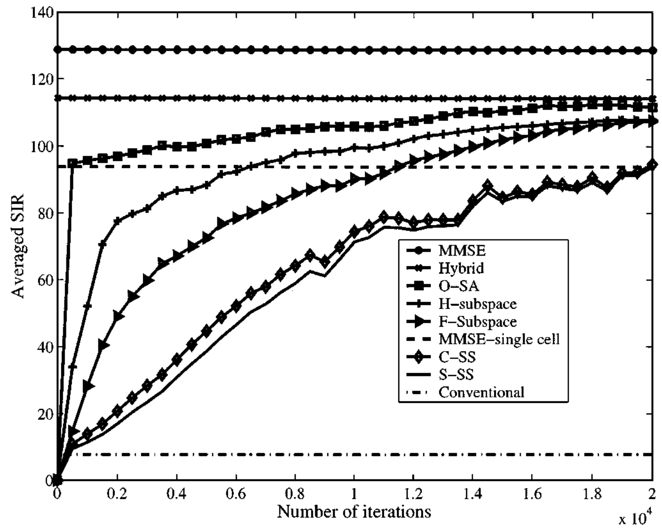

Fig. 2. SIR versus number of iterations for the S-SS, C-SS, O-SA, and the subspace-based F-subspace and H-subspace adaptive rules. Also shown are the SIRs for the deterministic detectors—MMSE, single cell MMSE, hybrid and conventional. (Every 500th iteration has been plotted for distinguishing line style.)

signals have processing gain of 31. The in-cell users are linearly independent. The starting vector for each recursion is chosen to be the zero vector. In Fig. 2 we have plotted (as horizontal lines) the SIR achieved by three deterministic receivers—the optimum SIR achieving MMSE receiver in (4), the single-cell MMSE receiver discussed in the preceding paragraph and the "hybrid" or "group" MMSE receiver of [11] that applies (to the receive signal) a decorrelative projection out of the known in-cell signal subspace and MMSE interference cancellation for out-of-cell interference. The figure also plots the SIR performance of the SA-based adaptations S-SS, C-SS and O-SA with fixed step-sizes, and of two "subspace"-based adaptations denoted as "full" or F-subspace and "hybrid" or H-subspace, respectively. The F-subspace algorithm is based on estimating the signal subspace that the received signal lies in and adapting blindly to the subspace version of the MMSE receiver in (4). A formulation of this is not included here and the reader is encouraged to refer to [10]. The H-subspace algorithm, described in [11], is also a subspace estimation-based adaptation that is designed to converge to the "hybrid" MMSE receiver. The subspace adaptations in this example have been "idealized" in the sense that the eigen-decomposition of the estimate of the correlation matrix (which is given by a running average of the term $\mathbf{y}_1\mathbf{y}_1^T$) is done at each iteration by the Matlab routine "eig." For a stationary environment as in this example, this sets an upper bound on the performance achievable with subspace tracking techniques such as projection approximation subspace tracing algorithm (PASTd), etc.

Note that subspace-based techniques perform better than the S-SS and C-SS rules. In fact, even though the F-subspace is the only adaptation here that is guaranteed to *aymptotically* reach the maximum SIR (in this simulation, convergence is not achieved even for 20 000 iterations), it converges slower than the hybrid technique which benefits noticeably from the added use of in-cell user information. The other in-cell information based algorithm O-SA also benefits significantly (in this case using the knowledge of $\mathbf{B}^{-1}$) and shows the best rate of convergence and out-performs both the subspace-based techniques in this example.

V. Blind Adaptation With Averaging

A recent fundamental development in stochastic approximation is the idea of averaging as introduced in [13] and [14]. In the former work, a linear algorithm for the one-dimensional case was considered, and asymptotic normality of the procedure was proved. Multidimensional problems were considered in [14] and under certain assumptions, mean square convergence for decreasing step-size adaptive algorithms followed by averaging was demonstrated. The improvement in convergence is essentially a result of averaging a SA algorithm that uses a step-size sequence that decays more slowly (or, is relatively "larger") than the a/n step-size that was used in the original Robbins-Monro formulation. It was further shown in [15], that almost sure (a.s.) convergence is achieved even for a suitable fixed step-size SA strategy with averaging, and this was "optimal" in terms of the convergence rate and the asymptotic error covariance matrix. A lot of work has since followed in stochastic approximation using the new ideas on averaging, (cf. [22] and [23]). Applying the averaging technique to the general SA rule in (6), i.e., including an averaging step after the "basic" recursion, we get

$$\mathbf{c}_{k1}(n+1) = (\mathbf{I} - \mu\mathbf{Q}\mathbf{y}_1(n+1)\mathbf{y}_1(n+1)^T) \times \mathbf{c}_{k1}(n) + \mu\mathbf{Q}\mathbf{s}_{k1} \quad (12)$$

$$\tilde{\mathbf{c}}_{k1}(n+1) = \frac{1}{n+1}\sum_{i=1}^{n+1}\mathbf{c}_{k1}(i), = \frac{1}{n+1}(n\tilde{\mathbf{c}}_{k1}(n) + \mathbf{c}_{k1}(n+1)) \quad (13)$$

where μ is a suitably chosen fixed step-size. In a slowly changing nonstationary environment one should implement the averaging over a finite window of past estimates for improved "tracking" properties. For the purpose of convergence analysis, however, it is more convenient to consider the averaging over all past estimates as in (13) above.

The three rules in the previous section combined with averaging will be referred to, respectively, as S-SS-AV, C-SS-AV, and O-SA-AV. Note that the system requirements and the computational complexity for the recursions with averaging remain the same as for the nonaveraged ones in (7), (9), and (11). In each of the rules above, the "smoothing" effect due to the averaging allows the basic recursion step to use relatively "larger" step-sizes than would be feasible for the nonaveraged adaptive rules. This is essentially the basis for the improvement in convergence that our numerical examples will later demonstrate.

Although our work was done independently, the authors have since discovered that the technique of averaging has been applied to blind multiuser detection in [20]. This paper investigates the averaged version of the minimum output energy (MOE)-based rule in a single-cell synchronous system and shows, using a result from [21], that the technique of averaging ensures performance comparable to the $O(N^2)$ recursive least squares (RLS) method with only $O(N)$ complexity. We note that this work addresses multiuser detection strategies for the more general, typically asynchronous, cellular system and, consequently, our proof techniques account for this (we also do not require the approximation of fourth-order statistics in terms of second-order statistics as in [20]). Moreover, the averaged MOE-based adaptation rule (derived differently in this paper than in [20]) is only one of three blind detection rules that we investigate.

The averaged adaptation rules discussed in this section can also be implemented with a suitable decreasing step size sequence (of the form $1/n^{\gamma}$, where $\gamma \in (0.5, 1)$, see [20], [15]), but we will consider only the fixed step size versions that are more suitable for tracking a time-varying environment. The tracking ability of the averaged adaptation rules can be improved by suitably changing the averaging step, and one such method using a "forgetting" factor has been suggested for the C-SS-AV rule in [20].

VI. Convergence Analysis

In our convergence analysis, we will primarily focus on the O-SA and O-SA-AV algorithms, the new algorithms in this paper that are suitable for up-link detection. First, we investigate the convergence of these algorithms for the general cellular network model of Section II where out-of-cell users are asynchronous. Note that in this case the received signal vector at a base station cannot be assumed to form an independent, identically distributed (i.i.d.) sequence. The i.i.d. assumption is sometimes made for simplicity of analysis and in certain cases, as we point out later, it is indeed valid. In Section VI-A, we do not make any such assumption and discuss the convergence of the O-SA algorithm by invoking the key result in [26]. Next, we assert the a.s. convergence of the O-SA-AV rule in Section VI-B using the main theorem on averaging in [23].

A special case of the asynchronous cellular model is the one where the received vector at a base station can be assumed to be i.i.d. We will refer to this as the "independence" assumption. This case is of interest in the single-cell (synchronous) model considered in [4], [5], or in a meulticell model where both the in-cell as well as the out-of-cell users are received synchronously at the base station. The independence assumption sometimes leads to stronger results than in the general asynchronous case. In the latter part of our convergence analysis, we will focus on this independence assumption case. We briefly outline in Section VI-C a proof of convergence for the fixed step-size C-SS algorithm that is different from that in [4]. Our analysis follows naturally from the derivation of the C-SS algorithm from the common framework of Section II. This analysis does not make the simplifying assumptions of [4] (the convergence analysis in [4] does, however, have the implementational advantage in that it enables the estimation of a suitable step size in terms of information available at a user). Next, in Section VI-D, we show the mean square convergence of the O-SA-AV rule.

A. Convergence of the O-SA Rule

For the sake of analysis, it is convenient to work with an equivalent symmetrized version of the O-SA rule. Based on the Cholesky factorization of $\mathbf{B}^{-1} = \mathbf{L}\mathbf{L}^T$ ($\mathbf{L}$ lower triangular), this is obtained as

$$\mathbf{d}_{k1}(n+1) = (\mathbf{I} - \mu_n \mathbf{L}^T \mathbf{y}_1(n+1)\mathbf{y}_1(n+1)^T \mathbf{L}) \times \mathbf{d}_{k1}(n) + \mu_n \mathbf{L}^T \mathbf{s}_{k1} \quad (14)$$

where we are trying to estimate $\mathbf{d}_{k1} = \mathbf{L}^{-1}\mathbf{A}^{-1}\mathbf{s}_{k1}$. The estimate of the MMSE solution is obtained as $\mathbf{c}_{k1}(n+1) = \mathbf{L}\mathbf{d}_{k1}(n+1)$. The symmetrization of the algorithm makes a direct application of the result in [24] possible, thereby enabling a more compact presentation here. With suitable modification, one can show a.s. convergence for the O-SA algorithm as stated in (11).

Due to asynchronocity of out-of-cell users, we cannot make the i.i.d. assumption. The received vector is, however, stationary and ergodic. Now, let the step size be chosen to be a decreasing sequence of the form $a_0/(a_1+n)$ and let

$$\mathbf{A}_n = \mathbf{y}_1(n)\mathbf{y}_1(n)^T. \quad (15)$$

Since $\{\mathbf{L}^T\mathbf{A}_n\mathbf{L}\}$ is an ergodic sequence and $E\|\mathbf{L}^T\mathbf{A}_n\mathbf{L}\|^2 < \infty$, by Proposition 4.3 [25, Ch. 1], we know that $\{\|\mathbf{L}^T\mathbf{A}_n\mathbf{L}\|^2\}$ is also an ergodic sequence. The matrix $\mathbf{L}^T\mathbf{A}\mathbf{L}$ with $\mathbf{A}$ as defined in (3) is clearly positive definite. By Theorem 1 in [24] (since all the conditions are satisfied a.s.) we can then say that $\lim_{n\to\infty} \mathbf{d}_{k1}(n) = \mathbf{L}^{-1}\mathbf{A}^{-1}\mathbf{s}_{k1} = \mathbf{d}_{k1}$ a.s. (the proof can be found in [26]). The last result implies that $\lim_{n\to\infty} \mathbf{c}_{k1}(n) = \mathbf{A}^{-1}\mathbf{s}_{k1} = \mathbf{c}_{k1}$ a.s.

B. Convergence of the O-SA-AV Rule

Once again, we find it convenient to work with a symmetrized version of the fixed step-size O-SA-AV rule presented in Section V so that we can easily apply the result in [23]. The symmetrized algorithm can be written as

$$\mathbf{d}_{k1}(n+1) = (\mathbf{I} - \mu \mathbf{L}^T \mathbf{y}_1(n+1)\mathbf{y}_1(n+1)^T \mathbf{L}) \times \mathbf{d}_{k1}(n) + \mu \mathbf{L}^T \mathbf{s}_{k1} \quad (16)$$

$$\tilde{\mathbf{d}}_{k1}(n+1) = \frac{1}{n+1}\sum_{i=1}^{n+1} \mathbf{d}_{k1}(i) \quad (17)$$

where $\mathbf{B}^{-1} = \mathbf{L}\mathbf{L}^T$ ($\mathbf{L}$ lower triangular) and what we are trying to estimate is $\mathbf{d}_{k1} = \mathbf{L}^{-1}\mathbf{A}^{-1}\mathbf{s}_{k1}$. The estimate of the MMSE solution is obtained as $\tilde{\mathbf{c}}_{k1}(n+1) = \mathbf{L}\tilde{\mathbf{d}}_{k1}(n+1)$. Applying [23, Thm. 2.1] we can state that.

Theorem 1: There exists a $\mu' > 0$ such that (s.t.) for all $0 < \mu < \mu'$

$$\tilde{\mathbf{d}}_{k1}(n) \to \mathbf{d}_{k1} + \delta_\mu \text{ as } n \to \infty \text{ a.s.} \quad (18)$$

Defining $\mathbf{A}_{n,k}(\mu) \triangleq (\mathbf{I} - \mu\mathbf{L}^T\mathbf{A}_n\mathbf{L}) \cdots (\mathbf{I} - \mu\mathbf{L}^T\mathbf{A}_k\mathbf{L}), k \le n$, and $\mathbf{A}_{n,n+1}(\mu) \triangleq \mathbf{I}$, the proof of the theorem above is given by a simple modification of that in [23, Thm. 2.1]. Clearly, it is important to ensure an acceptable asymptotic bias δ_μ and, therefore, we study its behavior further. We first note that $E\|\mathbf{L}^T\mathbf{A}_1\mathbf{L}\|^q < \infty$ and $\|\mathbf{L}^T\mathbf{s}_{k1}\|^q < \infty$ for all $q > 0$. Moreover, the sequence $\mathbf{A}_i$ is 2-dependent, i.e., the two sets of random variables $\{\ldots, \mathbf{A}_{n-1}, \mathbf{A}_n\}$ and $\{\mathbf{A}_{n+2}, \mathbf{A}_{n+3}, \ldots\}$ are statistically independent. Therefore, Assumption B1 in [23] is satisfied. It is suggested [23, Theorem 2.7, part b], then suggests that $\lim_n E\|\mathbf{d}_{k1}(n) - \mathbf{d}_{k1}\| = O(\mu^{1/2}), \delta_\mu = O(\mu^{1/2})$ as $\mu \to 0$. This result implies that there is a tradeoff between the rate of convergence and asymptotic bias. So although, one might employ larger step-sizes to benefit from the smoothing effect of averaging in pursuit of faster convergence, it comes at the price of larger asymptotic bias. As pointed out earlier it is sometimes possible to arrive at stronger results with the independence assumption, i.e., when the received vector sequence $\mathbf{y}_1$ can be assumed to be i.i.d. In this case, the problem of asymptotic bias disappears as shown (without proof here) by [23, Thm. 3.1]:

Theorem 2: There exists a μ' with $0 < \mu' < \mu^*$, s.t. $\forall 0 < \mu < \mu', \tilde{\mathbf{d}}_{k1}(n) \to \mathbf{d}_{k1}(n \to \infty)$ a.s.

Often in the adaptive literature, algorithms are analyzed for convergence in the mean square sense, i.e., the asymptotic behavior as $n \to \infty$ of the mean value of the error norm squared. In the following sections, we will investigate mean square convergence of the averaged and nonaveraged algorithms with the independence assumption.

C. Mean Square Convergence of the C-SS Rule With the Independence Assumption

In this and the following sections, we consider the case where the independence assumption is valid. We first investigate the mean square convergence of the nonaveraged algorithms with fixed step-size. The analysis for the S-SS rule is detailed in [5] based on earlier analyzes in [27] and [28]. A similar analysis can be done to show the mean square convergence of the symmetrized (as in Section VI-A) fixed step-size O-SA algorithm. We choose to outline the details of the analysis for the fixed step-size C-SS algorithm mainly to contrast it to the analysis in [4]. The proof of convergence here differs from that in [4] in the following ways: fourth-order statistics are not approximated in terms of second-order statistics and the signal vectors are not assumed to be approximately orthogonal. For notational simplicity, we shall omit the iteration index n on the received vector $\mathbf{y}_1$.

We first define the zero mean vector $\beta(n)$ as

$$\beta(n) \triangleq \mathbf{P}^{\perp}_{\mathbf{s}_{k1}} \left(\mathbf{y}_1\mathbf{y}_1^T - \mathbf{A}\right) \left(\mathbf{s}_{k1} + \mathbf{P}^{\perp}_{\mathbf{s}_{k1}}(\mathbf{p}_{k1} + \epsilon_{k1}(n))\right) \quad (19)$$

where $\epsilon_{k1}(n) = \mathbf{p}_{k1}(n) - \mathbf{p}_{k1}$.

After some manipulation, we obtain the following from (9):

$$\epsilon_{k1}(n+1) = \epsilon_{k1}(n) - \mu(\mathbf{P}^{\perp}_{\mathbf{s}_{k1}}\mathbf{A}\mathbf{P}^{\perp}_{\mathbf{s}_{k1}}\epsilon_{k1}(n) + \beta(n)). \quad (20)$$

Taking the norm squared of both sides of (20) and the conditional expectation, conditioned on $\epsilon_{k1}(n) = \epsilon$, and noting from (19) and the independence assumption that $E[\beta(n) \,|\, \epsilon_{k1}(n) = \epsilon] = \mathbf{0}$ we obtain that

$$\begin{aligned} &E[\|\epsilon_{k1}(n+1)\|^2 \,|\, \epsilon_{k1}(n) = \epsilon] \\ &\quad = \|\epsilon\|^2 - 2\mu\epsilon^T\mathbf{P}^{\perp}_{\mathbf{s}_{k1}}\mathbf{A}\mathbf{P}^{\perp}_{\mathbf{s}_{k1}}\epsilon \\ &\qquad + \mu^2\epsilon^T(\mathbf{P}^{\perp}_{\mathbf{s}_{k1}}\mathbf{A}\mathbf{P}^{\perp}_{\mathbf{s}_{k1}})^2\epsilon \\ &\qquad + \mu^2 E[\|\beta(n)\|^2 \,|\, \epsilon_{k1}(n) = \epsilon]. \end{aligned} \quad (21)$$

In order to develop bounds on the various terms in (21), we note, first, that $\mathbf{p}_{k1}(n)$ will a.s. have a nonzero projection on $\langle \mathbf{P}^{\perp}_{\mathbf{s}_{k1}}\mathbf{S}_1\rangle$ [24], [29]. It follows that $\epsilon_{k1}(n)$ will also a.s. have a nonzero projection on $\langle \mathbf{P}^{\perp}_{\mathbf{s}_{k1}}\mathbf{S}_1\rangle$, i.e., $\text{Prob}[\mathbf{P}_{\mathbf{P}^{\perp}_{\mathbf{s}_{k1}}\mathbf{S}_1}(\epsilon_{k1}(n)) = \mathbf{0}] = 0$. Along the lines of [5], we can show that with $\mathbf{G} \triangleq \mathbf{P}^{\perp}_{\mathbf{s}_{k1}}\mathbf{A}\mathbf{P}^{\perp}_{\mathbf{s}_{k1}}$, there exists $0 < k_0 \le k_1 < \infty$, such that $k_0\|\epsilon\|^2 \le \epsilon^T\mathbf{G}\epsilon \le k_1\|\epsilon\|^2$ and $k_0^2\|\epsilon\|^2 \le \epsilon^T\mathbf{G}^2\epsilon \le k_1^2\|\epsilon\|^2$. Further, there exists constants c_0, c_1 s.t. $0 \le E[\|\beta(n)\|^2 \,|\, \epsilon_{k1}(n) = \epsilon] \le c_0 + c_1\|\epsilon\|^2$.

Now, letting $v_n = E[\|\epsilon_{k1}(n)\|^2]$, we obtain the upper and lower bounds for v_{n+1} as

$$\begin{aligned} &\left(1 - 2\mu k_1 + k_0^2\mu^2\right) v_n \\ &\quad \le v_{n+1} \\ &\quad \le \left(1 - 2\mu k_0 + (k_1^2 + c_1)\mu^2\right) v_n + c_0\mu^2. \end{aligned} \quad (22)$$

By defining $\alpha_0 = 1-2\mu k_0+(k_1^2+c_1)\mu^2$ and $\alpha_1 = 1-2\mu k_1+k_0^2\mu^2$, we can rewrite (22) as $\alpha_1 v_n \le v_{n+1} \le \alpha_0 v_n + c_0\mu^2$. It is easily shown that we can always choose μ small enough so that $|\alpha_0| < 1$ and $|\alpha_1| < 1$, in which case the limiting MSE, i.e., $\lim_{n\to\infty} v_n$, has finite lower and upper bounds. If the step-size μ is chosen to be arbitrarily small, the MSE is seen to converge to zero as the number of iterations grows to infinity. This comes at a price though, since $\mu \to 0$ implies that the rate of convergence goes to zero. In other words, a large step-size will guarantee faster convergence but higher limiting MSE, whereas a small step-size will bring down the limiting MSE but adversely affect the convergence rate.

We note that a finite asymptotic MSE for a nonzero fixed step size would also be true for the O-SA algorithm. In the next section we show that with averaging, the O-SA-AV algorithm converges with *zero* asymptotic MSE.

D. Mean Square Convergence of the O-SA-AV Rule With the Independence Assumption

In this section, we show the mean square convergence with *zero* asymptotic MSE for the O-SA-AV algorithm

$$\begin{aligned} \mathbf{c}_{k1}(n+1) &= (\mathbf{I} - \mu\mathbf{B}^{-1}\mathbf{y}_1(n+1)\mathbf{y}_1(n+1)^T) \\ &\quad \times \mathbf{c}_{k1}(n) + \mu\mathbf{B}^{-1}\mathbf{s}_{k1} \end{aligned} \quad (23)$$

$$\tilde{\mathbf{c}}_{k1}(n+1) = \frac{1}{n+1}\sum_{i=1}^{n+1}\mathbf{c}_{k1}(i) \quad (24)$$

when the received vector sequence can be assumed to be i.i.d. Define $\mathbf{A}_{n,k}(\mu) \triangleq (\mathbf{I} - \mu\mathbf{B}^{-1}\mathbf{A}_n)\cdots(\mathbf{I} - \mu\mathbf{B}^{-1}\mathbf{A}_k), k \le n$, where $\mathbf{A}_n$ is as given in (15) and $\mathbf{A}_{n,n+1}(\mu) \triangleq \mathbf{I}$. We first concentrate on the basic recursion (23) to show that

$$E[\|\mathbf{A}_{n+1,1}(\mu)\|^2] = O(|\delta|^n). \quad (25)$$

To this end, we start by rewriting (23) in terms of the error $\epsilon(n) = \mathbf{c}_{k1}(n) - \mathbf{c}_{k1}$, as

$$\epsilon(n+1) = (\mathbf{I} - \mu\mathbf{B}^{-1}\mathbf{A})\epsilon(n) + \mu\eta(n) \quad (26)$$

where $\eta(n) \triangleq \mathbf{B}^{-1}(\mathbf{A} - \mathbf{A}_{n+1})\mathbf{c}_{k1}(n)$. Using the diagonalization $\mathbf{B}^{-1}\mathbf{A} \triangleq \mathbf{Z} = \mathbf{Q}_{\mathbf{Z}}\Lambda_{\mathbf{Z}}\,\mathbf{Q}_{\mathbf{Z}}^{-1}$, and defining $\mathbf{e}(n) \triangleq \mathbf{Q}_{\mathbf{Z}}^{-1}\epsilon(n)$ and $\nu(n) \triangleq \mathbf{Q}_{\mathbf{Z}}^{-1}\eta(n)$, we obtain, from (26), that

$$\mathbf{e}(n+1) = (\mathbf{I} - \mu\Lambda_{\mathbf{Z}})\mathbf{e}(n) + \mu\nu(n). \quad (27)$$

Note that using the Rayleigh–Ritz ratio for the positive definite matrix $(\mathbf{Q}_\mathbf{Z}^{-1})^T \mathbf{Q}_\mathbf{Z}^{-1}$

$$\frac{\|\mathbf{e}(n)\|^2}{\lambda_{\max}\left(\left(\mathbf{Q}_\mathbf{Z}^{-1}\right)^T \mathbf{Q}_\mathbf{Z}^{-1}\right)} \le \|\epsilon(n)\|^2 \le \frac{\|\mathbf{e}(n)\|^2}{\lambda_{\min}\left(\left(\mathbf{Q}_\mathbf{Z}^{-1}\right)^T \mathbf{Q}_\mathbf{Z}^{-1}\right)}. \quad (28)$$

Taking the expectation of the norm squared on both sides of (27), conditioned on $\mathbf{e}(n) = \mathbf{e}$, and since $E[\nu(n) \,|\, \mathbf{e}(n) = \mathbf{e}] = \mathbf{0}$, we obtain

$$E[\|\mathbf{e}(n+1)\|^2 \,|\, \mathbf{e}(n) = \mathbf{e}] = \mathbf{e}^T(\mathbf{I} - \mu\Lambda_\mathbf{Z})^2\mathbf{e} + \mu^2 E[\|\nu(n)\|^2 \,|\, \mathbf{e}(n) = \mathbf{e}]. \quad (29)$$

In order to bound the second term on the r.h.s. we note that

$$E[\|\nu(n)\|^2 \,|\, \mathbf{e}(n) = \mathbf{e}] \le 2E\left[\left\|(\mathbf{A} - \mathbf{A}_{n+1})\mathbf{B}^{-1} \times \left(\mathbf{Q}_\mathbf{Z}^{-1}\right)^T\right\|^2\right] (\|\mathbf{Q}_\mathbf{Z}\mathbf{e}\|^2 + \|\mathbf{c}_{k1}\|^2) \quad (30)$$

where we have used the fact that $\|\mathbf{Q}_\mathbf{Z}\mathbf{e} + \mathbf{c}_{k1}\|^2 \le 2\|\mathbf{Q}_\mathbf{Z}\mathbf{e}\|^2 + 2\|\mathbf{c}_{k1}\|^2$ (parallelogram inequality). Further, using the Rayleigh–Ritz ratio and the fact that $E[\|(\mathbf{A} - \mathbf{A}_{n+1})\mathbf{B}^{-1}(\mathbf{Q}_\mathbf{Z}^{-1})^T\|^2] \le \mathcal{K} < \infty$, we take the expected value with respect to $\mathbf{e}(n)$ of (29) to get

$$\begin{aligned} 0 &\le E[\|\mathbf{e}(n+1)\|^2] \\ &\le \left(1 - 2\mu\lambda_{\min}(\mathbf{Z}) + \mu^2\left(\lambda_{\max}^2(\mathbf{Z}) + 2\mathcal{K}\lambda_{\max}\left(\mathbf{Q}_\mathbf{Z}^T\mathbf{Q}_\mathbf{Z}\right)\right)\right) E[\|\mathbf{e}(n)\|^2] \\ &\quad + 2\mu^2\mathcal{K}\|\mathbf{c}_{k1}\|^2, \\ &\triangleq \delta(\mu)E[\|\mathbf{e}(n)\|^2] + 2\mu^2\mathcal{K}\|\mathbf{c}_{k1}\|^2 \end{aligned} \quad (31)$$

where $\lambda_{\min}(\mathbf{Z}) > 0$ by [30, Thm. 7.6.3]. We note that $\delta(0) = 1$ and $\delta'(0) = -2\lambda_{\min}(\mathbf{Z})$ is negative. This implies that for sufficiently small $\mu, |\delta(\mu)| \triangleq |\delta| < 1$. In the special case when $\mathbf{c}_{k1} = \mathbf{0}$, i.e., when we are stochastically estimating the solution of the equation $\mathbf{Z}\pi_{k1} = \mathbf{0}, E[\|\mathbf{e}(n+1)\|^2] = O(|\delta|^n)$, and this implies, from (28), that

$$E[\|\epsilon(n+1)\|^2] = O(|\delta|^n). \quad (32)$$

We now rewrite the error recursion in (26) as

$$\epsilon(n+1) = (\mathbf{I} - \mu\mathbf{B}^{-1}\mathbf{A}_{n+1})\epsilon(n) + \mu\mathbf{v}(n+1) \quad (33)$$

where $\mathbf{v}(n+1) \triangleq \mathbf{B}^{-1}(\mathbf{A} - \mathbf{A}_{n+1})\mathbf{c}_{k1}$. Note that the recursion in (33) estimates the solution of the equation $\mathbf{Z}\epsilon = \mathbf{0}$, which, since $\mathbf{Z}$ is nonsingular, we know to be the $\mathbf{0}$ vector. From (33), we also observe that if $\mathbf{c}_{k1} = \mathbf{0}$, then $\mathbf{v}(n) = \mathbf{0}\ \forall n$, and

$$E[\|\epsilon(n+1)\|^2] = E[\|\mathbf{A}_{n+1,1}(\mu)\epsilon(0)\|^2]. \quad (34)$$

From (32) and (34), we obtain that, for sufficiently small $\mu, E[\|\mathbf{A}_{n+1,1}(\mu)\epsilon(0)\|^2] = O(|\delta|^n)$. For any finite initial estimate for $\mathbf{c}_{k1}, \|\epsilon(0)\|^2 < \infty$, and in the the special case of $\epsilon(0)$ being N-dimensional unit vectors, (25) is true, i.e., $E[\|\mathbf{A}_{n+1,1}(\mu)\|^2] = O(|\delta|^n)$.

At this point we need to consider the effect of the "averaging" step (24). To do this we find that we can apply the analysis in [23, Thm. 5.3,] in a straightforward way. For the sake of completeness, an outline of the rest of the proof has been included here.

We first note that by definition, $E[\mathbf{v}(n)]$ is the $\mathbf{0}$ vector, and in view of (25), there is no loss in generality in assuming $\epsilon(0) = \mathbf{0}$. Then $\epsilon(n) = \sum_{i=1}^n \mathbf{R}_{ni}\mathbf{v}(i)$, where $\mathbf{R}_{ni} = \mu\mathbf{A}_{n,i+1}(\mu)$. It can be shown that

$$E[\epsilon(n)\epsilon(n)^T] = \sum_{i=1}^n E\left[\mathbf{R}_{ni}\Omega\mathbf{R}_{ni}^T\right] \triangleq \Xi_{\mu,n} \quad (35)$$

where $\Omega = E[\mathbf{v}(1)\mathbf{v}^T(1)]$. Using the facts that the matrix sequence $\{\mathbf{A}_n\}$ is i.i.d., that $E[\|\mathbf{B}^{-1}\mathbf{A}_n\|^2] < \infty$, and that the result in (25) holds, we can deduce that, for sufficiently small value of $\mu, \sum_{n=1}^\infty E[\|\mathbf{R}_{n1}\|^2] < \infty$. Consequently

$$\begin{aligned} \lim_{n\to\infty} \Xi_{\mu,n} &= \lim_{n\to\infty} \sum_{i=1}^n E\left[\mathbf{R}_{ni}\Omega\mathbf{R}_{ni}^T\right] \\ &= \lim_{n\to\infty} \sum_{i=1}^n E\left[\mathbf{R}_{i1}\Omega\mathbf{R}_{i1}^T\right] \\ &= \sum_{i=1}^\infty E\left[\mathbf{R}_{i1}\Omega\mathbf{R}_{i1}^T\right] = \Xi_\mu \end{aligned} \quad (36)$$

exists and is finite. The error recursion obtained by subtracting $\mathbf{c}_{k1}$ from both sides of the averaging step of the O-SA-AV rule in (23) (at time index n) is given by $\tilde{\epsilon}(n) = (1/n)\sum_{i=1}^n \epsilon(i)$, where, $\tilde{\epsilon}(n) = \tilde{\mathbf{c}}_{k1}(n) - \mathbf{c}_{k1}$. We now obtain

$$nE\left[\tilde{\epsilon}(n)\tilde{\epsilon}(n)^T\right] = \frac{1}{n}E\left[\sum_{k=1}^n \epsilon(k)\epsilon(k)^T + \sum_{k=1}^{n-1}\sum_{l=k+1}^n \epsilon(k)\epsilon(l)^T + \sum_{l=1}^{n-1}\sum_{k=l+1}^n \epsilon(k)\epsilon(l)^T\right]. \quad (37)$$

For the first term on the rhs, we can show the following:

$$\frac{1}{n}\left(\sum_{k=1}^n \epsilon(k)\epsilon(k)^T\right) = \frac{1}{n}\sum_{k=1}^n \Xi_{\mu,k} \to \Xi_\mu. \quad (38)$$

For $k < l$, $E[\epsilon(k)\epsilon(l)^T]$

$$= \sum_{i=1}^k E\left[\mathbf{R}_{ki}\Omega\mathbf{R}_{li}^T\right] \quad (39)$$

$$\begin{aligned} &= \sum_{i=1}^k E\left[\mathbf{R}_{ki}\Omega\mathbf{R}_{ki}^T\right](\mathbf{I} - \mu\mathbf{B}^{-1}\mathbf{A})^{l-k} \\ &= \Xi_{\mu,k}(\mathbf{I} - \mu\mathbf{B}^{-1}\mathbf{A})^{l-k}. \end{aligned} \quad (40)$$

Therefore, after some manipulation

$$\begin{aligned} &\frac{1}{n}E\left[\sum_{k=1}^{n-1}\sum_{l=k+1}^n \epsilon(k)\epsilon(l)^T\right] \\ &= \frac{1}{n}\sum_{k=1}^{n-1} \Xi_{\mu,k}(\mathbf{I} - (\mathbf{I} - \mu\mathbf{B}^{-1}\mathbf{A})^{n-k}) \\ &\quad \times (\mu\mathbf{B}^{-1}\mathbf{A})^{-1}(\mathbf{I} - \mu\mathbf{B}^{-1}\mathbf{A}) \\ &\to \Xi_\mu(\mu\mathbf{B}^{-1}\mathbf{A})^{-1}(\mathbf{I} - \mu\mathbf{B}^{-1}\mathbf{A}) \\ &= \Xi_\mu(\mu\mathbf{B}^{-1}\mathbf{A})^{-1} - \Xi_\mu. \end{aligned} \quad (41)$$

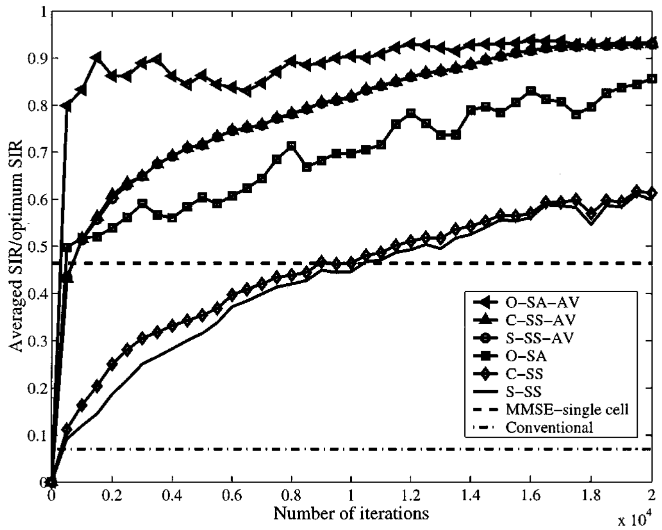

Fig. 3. SIR performance (relative to the deterministic MMSE solution) versus number of iterations with the independence assumption: high SNR case. (Every 500th iteration has been plotted.)

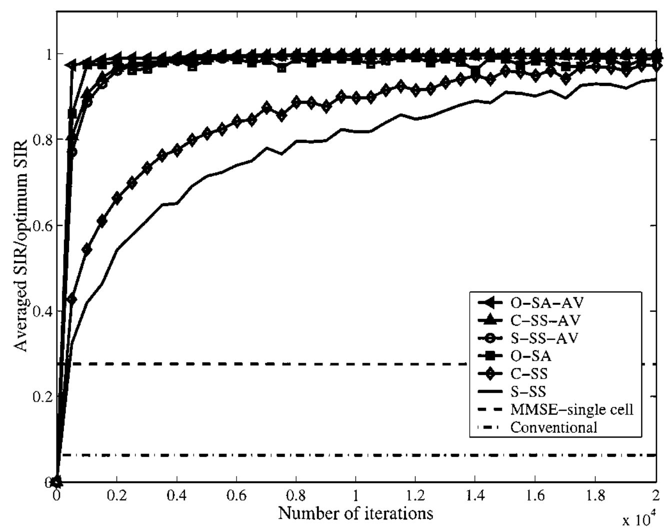

Fig. 4. SIR performance (relative to the deterministic MMSE solution) versus number of iterations with the independence assumption: low SNR case. (Every 500th iteration has been plotted.)

In the same way one obtains

$$\frac{1}{n}E\left[\sum_{l=l}^{n-1}\sum_{k=l+1}^{n}\epsilon(k)\epsilon(l)^T\right]\rightarrow(\mu\mathbf{B}^{-1}\mathbf{A})^{-1}\Xi_\mu-\Xi_\mu. \quad (42)$$

From (38), (41), and (42), therefore, we obtain that

$$\lim_{n\to\infty} nE[\tilde{\epsilon}(n)\tilde{\epsilon}(n)^T]=\Xi$$
$$=\frac{1}{\mu}(\Xi_\mu\mathbf{A}^{-1}\mathbf{B}+\mathbf{B}\mathbf{A}^{-1}\Xi_\mu)-\Xi_\mu \quad (43)$$

is finite. This implies that $E[\tilde{\epsilon}(n)^T\tilde{\epsilon}(n)] = O((1/n))$ or, $\lim_{n\to\infty} E[\tilde{\epsilon}(n)^T\tilde{\epsilon}(n)]=0$, i.e., the MSE is zero.

VII. Numerical Examples

For the following examples, we consider the seven-cell cellular system in Fig. 1 with 12 users in each cell. The cell of interest is the center cell, which we call cell 1. Once again, we use $1/d^4$ path-loss law and equal transmit powers. We consider detection on the up-link at base station 1. The users employ randomly chosen unit norm signature sequences, which once chosen, remain fixed throughout the adaptation. We demonstrate the convergence improvement due to averaging of the SS and the orthogonalization-based adaptation rules and provide comparison with the subspace-based methods.

Example 2: We consider first the simple model where all the users are received synchronously at base station 1 which has the knowledge of the common timing. The signalling is bandwidth-efficient in that the processing gain is only 12. The received SNR for in-cell users at base station 1 varies from 37 dB to 15 dB. Fig. 3 plots the averaged SIR (relative to the deterministic MMSE) for the highest SNR user. The initial guess for the vector being estimated by each recursion is chosen to be the zero vector. The fixed step-sizes have been chosen by trial and error to ensure good convergence for each recursion. The optimum SIR (corresponding to the MMSE detector) is significantly larger than those for the conventional and the single-cell MMSE detectors. The SIR performance for C-SS is marginally better than that for S-SS, and they are both bettered by the O-SA rule in terms of rate of convergence. If the eigen-spread of $\mathbf{A}$ and $\mathbf{B}^{-1}\mathbf{A}$ (where eigen-spread is the ratio of the maximum to the minimum eigenvalue) is taken as an indicator of the rate of convergence for the S-SS and the O-SA rules, respectively, (where the rate of convergence is inversely proportional to the eigen-spread), then we note for this example that with the out-of-cell users being sufficiently weaker than the in-cell users, the eigen-spread for $\mathbf{A}$ is $\approx$521 as compared with $\approx$11 for $\mathbf{B}^{-1}\mathbf{A}$. There is also a clear performance improvement for all three adaptive rules due to averaging. Fig. 4 plots the SIR perfromance for the lowest SNR user. This plot shows the same trends, although the O-SA, S-S-AV, C-SS-AV, and O-SA-AV algorithms are much closer to each other in their performance.

Example 3: We consider the same system as in the previous example, but allow symbol asynchronism at base station 1 as in (1). Fig. 5 plots the averaged SIR for the highest SNR user, for a random set of time lags assigned to out-of-cell users. The step-size sequence chosen for the S-SS, C-SS, and O-SA rules are of the form $a_0/(a_1+n)$ where the constants have been adjusted separately for each of the adaptation rules by trial and error to ensure good convergence performance. Even with symbol asynchronism, the eigen-spread for $\mathbf{A}(\approx 590)$ remains significantly higher than that for $\mathbf{B}^{-1}\mathbf{A}(\approx 10)$. Consequently, the almost identical SIR performance for the C-SS and S-SS is once again bettered by the O-SA rule in terms of rate of convergence. There is again a clear performance improvement for all three adaptive rules when using relatively larger fixed step-sizes followed by averaging.

Example 4: We consider the same model as in the previous example, with variable number of users. More specifically, we start with ten users in a cell, run the adaptation for several iterations, at which point we switch on two more users in each cell. Fig. 6 shows the change in the desired maximum SIR for the

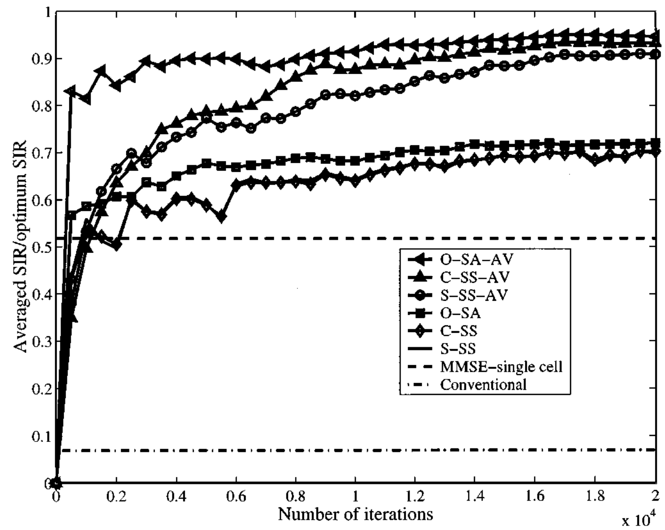

Fig. 5. SIR performance (relative to the deterministic MMSE solution) versus number of iterations for the *asynchronous* out-of-cell users case. (Every 500th iteration has been plotted.)

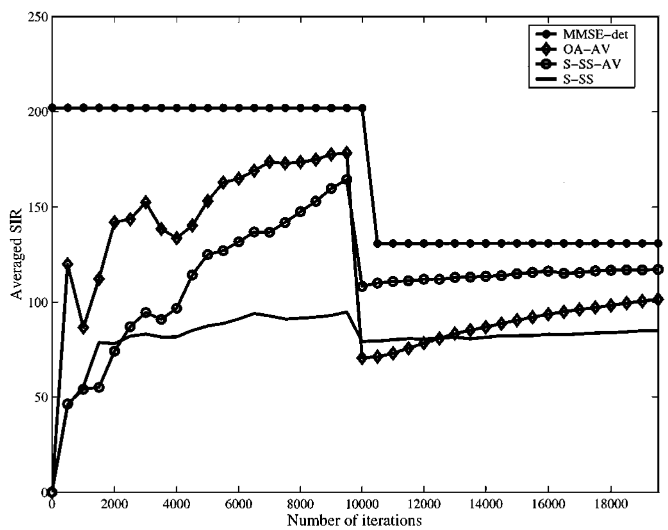

Fig. 6. SIR versus number of iterations for the S-SS, S-SS-AV, O-SA-AV, MMSE deterministic for initially ten and later 12 users per cell. (Every 500th iteration has been plotted.)

user of interest and the effect this has on the trajectories for the S-SS, S-SS-AV, and the O-SA-AV rules. The S-SS algorithm and the basic recursions for both the S-SS-AV and O-SA-AV [see (12)[, use fixed step sizes. Note that for this example the new users were comparatively weak and we did not re-initialize the averaging for the O-SA-AV rule or use any special modification for tracking. Even so, the averaging algorithms show reasonable tracking capability (compare with the nonaveraged algorithm). One simple way to improve the tracking capability of the averaging algorithms would be to use a finite window of iterates in the averaging step of (13).

Example 5: We return again to the synchronous model in Example 1. Fig. 7 plots the SIRs achieved by the SA-based adaptations S-SS, C-SS, S-SS-AV, C-SS-AV, O-SA, and O-SA-AV, and the subspace-based adaptations, F-subspace and H-subspace. The performance of the nonaveraged S-SS and C-SS rules can be seen to be comparable, with averaging, to the

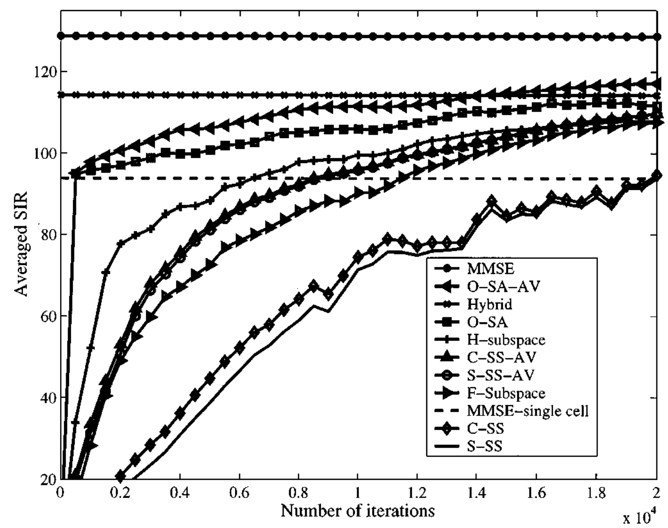

Fig. 7. SIR versus number of iterations for the S-SS, C-SS, O-SA, S-SS, C-SS-AV, O-SA-AV, the subspace-based F-subspace, and H-subspace adaptive rules. Also shown are the SIRs for the deterministic detectors—MMSE, MMSE single cell, and Hybrid. Only a portion of the y-axis has been shown for clarity. Refer to Fig. 2 for the conventional detector plot. (Every 500th iteration has been plotted.)

computationally more intensive subspace-based methods. Note that the H-subspace technique can, of course, at best achieve its deterministic performance and, therefore, asymptotically all the averaged rules are guaranteed to outperform it. The cell-based algorithm O-SA performs slightly better with averaging and in this example outperforms all the other rules.

VIII. Conclusion

Blind adaptive MMSE detection at the base station in a cellular system is studied in this paper. A cell-based adaptation strategy that outperforms the previously known SS-based blind adaptation rules in a typical cellular environment has been presented. The technique of averaging to improve the convergence of single-signal as well as cell-based blind adaptive detectors has been introduced and analyzed. It has been shown that the averaged version of the cell-based adaptive scheme converges a.s., and (with the independence assumption for the received vectors) in mean square with zero asymptotic bias. The superiority in terms of convergence of the averaged adaptations has been explored numerically.

Acknowledgment

D. Das would like to thank A. Kapur for useful discussions about the O-SA rule and its convergence analysis.

References

[1] S. Verdú, *Multiuser Detection*. New York: Cambridge Univ. Press, 1998.
[2] R. Lupas and S. Verdu, "Linear multiuser detectors for synchronous code-division multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35, pp. 123–136, Jan. 1989.
[3] U. Madhow and M. L. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 38, pp. 509–519, Apr. 1990.
[4] M. L. Honig, U. Madhow, and S. Verdu, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 994–960, July 1995.

[5] S. Ulukus and R. D. Yates, "A blind adaptive decorrelating detector for CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1530–1541, Oct. 1998.

[6] M. L. Honig and M. Tsatsanis, "Adaptive techniques for CDMA receivers," *IEEE Signal Proc. Mag.*, vol. 17, pp. 49–61, May 2000.

[7] P. Rapajic and B. Vucetic, "Adaptive receiver structures for asynchronous CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 685–697, May 1994.

[8] D. Pados and S. Batalama, "Low complexity blind detection of DS/CDMA signals: Auxilliary vector receivers," *IEEE Trans. Commun.*, vol. 45, pp. 1586–1594, Dec. 1997.

[9] D.-S. Chen and S. Roy, "An adaptive multi-user receiver for CDMA systems," *IEEE J. Select. Areas Commun.*, vol. 12, pp. 808–816, June 1994.

[10] X. Wang and V. Poor, "Blind multiuser detection: A subspace approach," *IEEE Trans. Inform. Theory*, vol. 44, pp. 677–689, Mar. 1998.

[11] J. Yu and A. Høst-Madsen, "Subspace tracking for group-blind multiuser detectors," presented at the Vehicular Technology Conf., Houston, TX, May 16–19, 1999.

[12] A. Høst-Madsen and X. Wang, "Group-blind multiuser detection for up-link CDMA," presented at the Vehicular Technology Conf., Houston, TX, May 16–19, 1999.

[13] D. Ruppert, "Efficient Estimators from a Slowly Convergent Robbins-Monro Procedure," School of Oper. Res. and Ind. Eng., Cornell Univ., Ithaca, NY, Tech. Rep. 781, 1988.

[14] B. T. Polyak, "New stochastic approximation type procedures," *Automat. i Telemekh.*, vol. 51, pp. 98–107, 1990. In Russian; translated in *Automatic Remote Control*, 1991.

[15] B. T. Polyak and A. B. Juditsky, "Acceleration of stochastic approximation by averaging," *SIAM J. Contr. Optim.*, vol. 30, pp. 838–855, 1992.

[16] M. Varanasi, "Stochastic Power Control for Nonlinear Multiuser Receivers in Cellular Radio Networks," presented at the IEEE Information Theory Workshop, South Africa, June 20–25, 1999.

[17] L. A. Hageman and D. M. Young, *Applied Iterative Methods*. New York: Academic, 1981.

[18] O. Macchi, *Adaptive Processing—The Least Mean Squares Approach with Applications in Transmission*. New York: Wiley, 1995.

[19] W. K. Jenkins *et al.*, *Advanced Concepts in Adaptive Signal Processing*. Norwell, MA: Kluwer, 1996.

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

[21] G. G. Yin and Y. Zhu, "Averaging procedures in adaptive filtering: An efficient approach," *IEEE Trans. Automat. Contr.*, vol. 37, pp. 466–475, 1992.

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

[23] L. Gyorfi and H. Walk, "On the averaged stochastic approximation for linear regression," *SIAM J. Contr. Optim.*, vol. 34, pp. 31–61, 1996.

[24] L. Gyorfi, "Adaptive linear procedures under general conditions," *IEEE Trans. Inform. Theory*, vol. IT–30, pp. 262–267, Mar. 1984.

[25] U. Krengel, *Ergodic Theorems*. New York: de Gruyter Studies in Math., 1985.

[26] L. Gyorfi, "Stochastic approximation from ergodic samples for linear regression," *Z. fur Wahrscheinlichkeitstheorie*, vol. 54, pp. 47–55, 1980.

[27] D. J. Sakrison, "Stochastic approximation: A recursive method for solving regression problems," in *Advances in Communication Systems 2*, A. V. Balakrishnan, Ed. New York: Academic, 1966.

[28] Y. Z. Tsypkin, *Adaptation and Learning in Automatic Systems*. New York: Academic, 1971.

[29] W. F. Stout, *Almost Sure Convergence*. New York: Academic, 1974.

[30] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1993.

Deepak Das (S'96–M'01) received the B.E. (with honors) degree in electrical and electronics engineering from Birla Institute of Technology and Science, Pilani, India, in 1992, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Colorado, Boulder, in 1997 and 2000, respectively.

From 1992 to 1995, he was a software Consultant at Mahindra-British Telecom, and worked with British Telcom, London, U.K. He is currently working with the Wireless Technology Laboratory of Lucent Technologies, NJ. His research interests have been primarily in the areas of multiuser detection, power control, and intelligent antennas.

Mahesh K. Varanasi (S'87–M'89–SM'95) received the B.E. degree in electronics and communication engineering from Osmania University, Hyderabad, India, in 1984, and the M.S. and Ph.D. degrees in electrical engineering from Rice University, Houston, TX, in 1987 and 1989, respectively.

In 1989, he joined the faculty of the College of Engineering and Applied Sciences at the University of Colorado at Boulder, CO, where he is currently Professor of Electrical and Computer Engineering. His teaching and research interests are in communication theory, information theory, and signal processing. His research has been in the areas of multiuser detection, space-time communications, equalization, signal design, and power- and bandwidth-efficient multiuser communications.