Adaptive Power Techniques for Blind Channel Estimation in CDMA Systems

Xenofon G. Doukopoulos, *Student Member, IEEE,* and George V. Moustakides, *Senior Member, IEEE*

***Abstract*—The problem of blind adaptive channel estimation in code-division multiple access (CDMA) systems is considered. Motivated by the iterative power method, which is used in numerical analysis for estimating singular values and singular vectors, we develop recursive least squares (RLS) and least mean squares (LMS) subspace-based adaptive algorithms in order to identify the impulse response of the multipath channel. The schemes proposed in this paper use only the spreading code of the user of interest and the received data and are therefore blind. Both versions (RLS and LMS) exhibit rapid convergence combined with low computational complexity. With the help of simulations, we demonstrate the improved performance of our methods as compared with the already-existing techniques in the literature.**

***Index Terms*—Blind adaptive algorithms, CDMA, channel estimation, LMS, RLS.**

I. Introduction

CODE-DIVISION multiple-access (CDMA) implemented with direct-sequence (DS) spread spectrum constitutes one of the most important emerging technologies in wireless communications. It is well known that CDMA has been selected as the base for the third-generation mobile telephone systems. In a CDMA system, users are capable of simultaneously transmitting in time, while occupying the same frequency band, by using a unique signature waveform assigned to each one of them. This important advantage also constitutes its principal weakness, since it is the main source of performance degradation. Indeed for every user, all other users play the role of (multiuser) interference.

When no multipath is present, numerous offline as well as adaptive detection schemes have been proposed and extensively analyzed in the literature [1]–[5]. These detectors, in order to be practically implementable, require at least knowledge of the signature waveform of the user of interest. Assuming availability of this information is, in fact, quite reasonable. Whenever CDMA signals propagate through a multipath environment, the received signal has the same form as in the nondispersive case, except that, in place of the signatures, we now have their convolution with the channel impulse response (also known as *composite signature*). This remarkable property allows for the employment of exactly the same detection structures cited above after simply replacing the initial signature of the user of interest with its corresponding composite signature. Of course, for this to be possible, we need knowledge or efficient estimates of the channel impulse response.

Due to their self-sufficiency with respect to training, blind estimation methods tend nowadays to be the most frequent candidates for channel estimation. The blind channel estimation literature for CDMA is rather limited. In [6] and [7], the (practically offline) methods that are proposed involve a large singular value decomposition (SVD) for estimating the noise subspace of the received data. This part is computationally intense, not to mention the fact that SVD presents no particular repetitive structure suitable for online processing. In [8], we proposed an alternative *offline* scheme that replaces the SVD with a simple matrix power. This resulted in a substantial computational gain, compared with the previous two methods, without any significant performance loss. Another advantage of the method in [8] is the fact that it does not require knowledge (or estimates) of the signal subspace rank in contrast with [6] and [7], where such information is indispensable. A similar approach was presented independently in [9], emphasizing its close connection to the subspace method of [6].

Blind adaptive channel estimation techniques were proposed in [9]–[11]. Specifically, in [10], based on the analytic results offered in [12], recursive least squares (RLS) and least mean squares (LMS) type versions were developed that will serve for comparison against our schemes in the simulations section. A variant of the work in [10], which is reported in [11], consists of using higher order cumulants. This approach, however, suffers from slow convergence even for small codes, whereas its success relies on the Gaussian noise assumption and, in particular, the fact that higher order cumulants of Gaussian random variables are zero. Finally, in [9], an RLS algorithm is proposed for the adaptive implementation of the power method suggested in [8] and [9]. We should also mention that in [9], the channel estimates are obtained through an SVD of the size of the channel impulse response.

In this work, we are going to extend the power method proposed in [8] and use it to develop RLS and, *more importantly*, LMS adaptive algorithms. In particular, we are going to introduce two versions of the power method suitably tuned to the CDMA channel-estimation problem. With this theory at hand, we will then develop RLS and LMS type adaptive algorithms that are characterized by high performance, even under very difficult signaling conditions. Although our RLS version resembles the one proposed in [9], there is a significant difference. We

Manuscript received August 6, 2003; revised March 4, 2004. This work was supported in part by the European Commission through TMR project System Identification (FMRX-CT98-0206). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Dennis R. Morgan.

X. G. Doukopoulos is with IRISA-INRIA and University of Rennes 1, Rennes, France (e-mail: xdoukopo@irisa.fr).

G. V. Moustakides is with IRISA-INRIA, Rennes, France, and also with the Departement of Computer and Communication Engineering, University of Thessaly, Volos, Greece (e-mail: moustaki@uth.gr).

Digital Object Identifier 10.1109/TSP.2004.842187

will be able to completely eliminate the SVD of the size of the channel needed in [9] by replacing it with a simple matrix-vector multiplication of the same size. This will result in the reduction of the corresponding computational complexity by one order of magnitude. As far as our LMS scheme is concerned, when it is compared with the corresponding version of [10], it can perform orders of magnitude better, at a similar computational cost level.

The rest of the paper is organized as follows. In Section II, we introduce the signal model for DS-CDMA in the presence of multipath, whereas in Section III, we present two subspace problems that constitute the heart of the blind channel estimation problem, along with a brief discussion on issues concerning identifiability. Section IV contains the power method and, in particular, two variants that are suitable for the solution of the two subspace problems introduced in Section III. In Section V, we develop blind adaptive RLS and LMS algorithms for the channel estimation problem. Simulations and comparisons are provided in Section VI, and finally, Section VII concludes our work.

II. System Model

Consider a K-user CDMA system with identical chip waveforms and signaling antipodally through a multipath channel in the presence of additive white (*but not necessarily Gaussian*) noise (AWN). Although CDMA systems are continuous in time, they can be adequately modeled by an equivalent discrete time system. Specifically, no information is lost if we limit ourselves to the output of a chip matched filter applied to the received analog signal and sampled at the chip rate [5].

Let N be the processing gain of the code and L the length of the channel impulse response. Moreover, let $\mathbf{s}_i = [s_i(0)s_i(1)\cdots s_i(N-1)]^t$ be the length N normalized signature waveform of User-i (i.e., $\|\mathbf{s}_i\| = 1$), and denote by $s_i(n)$ the sequence corresponding to this signature waveform zero-padded from both ends toward infinity. The transmitted signal due to User-i is given by

$$z_i(n) = a_i \sum_{k=-\infty}^{\infty} s_i(n-kN-\tau_i)b_i(k), \quad i=1,\ldots,K \tag{1}$$

where a_i is the amplitude of User-i, $b_i(n)$ the corresponding bit sequence, and τ_i the initial delay that can take any value in the set $\{0,\ldots,N-1\}$. The signal $z_i(n)$ propagates through a multipath AWN channel with impulse response $\mathbf{f}_i = [f_i(0)\cdots f_i(L-1)]^t$, and let $\tilde{\mathbf{s}}_i = [\tilde{s}_i(0)\tilde{s}_i(1)\cdots\tilde{s}_i(N+L-2)]^t$ be the composite signature waveform of User-i, i.e. $\tilde{\mathbf{s}}_i = \mathbf{s}_i \star \mathbf{f}_i$, where "$\star$" denotes convolution. Then, the received signal $y(n)$ can be written as

$$y(n) = \sum_{i=1}^{K}\sum_{k=-\infty}^{\infty} a_i\tilde{s}_i(n-kN-\tau_i)b_i(k) + \sigma w(n) \tag{2}$$

where $\tilde{s}_i(n)$ is the sequence corresponding to the composite signature waveform of User-i zero-padded from both ends toward infinity, and $w(n)$ is a unit variance i.i.d. noise sequence with σ^2 denoting its power.

The model given in (2) fully describes the uplink (mobile to base station) scenario of a multipath CDMA system. For the downlink, we simply need to select $\mathbf{f}_1 = \cdots = \mathbf{f}_K = \mathbf{f}$ and $\tau_1 = \cdots = \tau_K = \tau$ (since all users propagate through the *same* multipath channel and are completely synchronized). Although next we will consider the downlink case, we should keep in mind that with almost no modification, our methodology can be applied to the uplink as well in order to estimate the different channels *one-by-one*.

Without loss of generality, throughout this paper, we will assume that the user of interest is User-1. We will also assume that the initial delay τ is known, and therefore, we have exact synchronization with the user of interest. A simple synchronization technique, based on the same power method principle that we are going to use here, can be found in [13]. For the presentation of our method, it is more convenient to express the received signal in blocks of data. In particular, we are interested in blocks of size $mN+L-1$, where m is a positive integer. Consequently, let us consider the block

$$\mathbf{r}(n) = [y(nN)\cdots y((n-m)N-L+2)]^t \tag{3}$$

which, as we said, is assumed to be synchronized with the user of interest. Notice that due to synchronization, the block $\mathbf{r}(n)$ contains m entire copies of the composite signature of the user of interest. Specifically, $\mathbf{r}(n)$ can be decomposed as follows:

$$\mathbf{r}(n) = \sum_{l=1}^{m}\begin{bmatrix}\mathbf{0}_{(l-1)N\times 1}\\ \tilde{\mathbf{s}}_1\\ \mathbf{0}_{(m-l)N\times 1}\end{bmatrix} a_1 b_1(n-l+1) + \sum_{i=2}^{K}\sum_{l=1}^{m}\begin{bmatrix}\mathbf{0}_{(l-1)N\times 1}\\ \tilde{\mathbf{s}}_i\\ \mathbf{0}_{(m-l)N\times 1}\end{bmatrix} a_i b_i(n-l+1) + \text{ISI} + \sigma\mathbf{w}(n). \tag{4}$$

We observe in (4) that the sum of the first m terms involves the entire composite signature of the user of interest, then the multiaccess interference (MAI) part that contains terms similar to the first sum but coming from interfering users, and this is followed by the part that includes the intersymbol interference (ISI) of all users; finally, the last term is the AWN vector. All terms in (4), except the last one, are of the form $\mathbf{d}_l b_i(n-j)$, where $\mathbf{d}_l$ are deterministic vectors corresponding to shifted versions of composite signatures coming from the user of interest or MAI or shifted sections of composite signatures coming from ISI; $b_i(n)$ are binary data that are mutually independent and independent from the noise vector.

One final point we should make, before proceeding with the presentation of the two subspace problems, is the fact that the composite signature of User-1 can be written as

$$\tilde{\mathbf{s}}_1 = \mathbf{S}_1\mathbf{f} \tag{5}$$

where $\mathbf{S}_1$ is a convolution matrix of size $(N+L-1) \times L$, corresponding to the initial signature of User-1 and defined as

$$\mathbf{S}_1 = \begin{bmatrix} s_1(0) & 0 & \cdots & 0 \\ \vdots & s_1(0) & \ddots & \vdots \\ s_1(N-1) & \vdots & \ddots & 0 \\ 0 & s_1(N-1) & \ddots & s_1(0) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_1(N-1) \end{bmatrix}. \quad (6)$$

III. Two Subspace Problems

Let us first identify the channel impulse response $\mathbf{f}$, assuming availability of the data autocorrelation matrix and the initial signature waveform of the user of interest, i.e. $\mathbf{s}_1$. The data autocorrelation matrix satisfies

$$\mathbf{R} \triangleq \mathbb{E}\left\{\mathbf{r}(n)\mathbf{r}^t(n)\right\} = \mathbf{Q} + \sigma^2\mathbf{I} \quad (7)$$

where

$$\mathbf{Q} = \sum_l \mathbf{d}_l\mathbf{d}_l^t \quad (8)$$

is a symmetric, non-negative definite matrix, of dimensions $mN+L-1$, formed by the $\mathbf{d}_l$ vectors introduced in the signal model.

By applying an SVD on $\mathbf{R}$, we can write

$$\mathbf{R} = [\mathbf{U}_s \quad \mathbf{U}_w] \begin{bmatrix} \mathbf{\Lambda}_s + \sigma^2\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \sigma^2\mathbf{I} \end{bmatrix} [\mathbf{U}_s \quad \mathbf{U}_w]^t \quad (9)$$

where $\mathbf{U}_s$ and $\mathbf{U}_w$ are orthonormal bases for the signal and noise subspace, respectively. In particular, we should note that $\mathbf{U}_w$ spans the noise subspace, which corresponds to the *smallest* singular value of $\mathbf{R}$ (being equal to σ^2). Due to the orthogonality of the two subspaces, for any vector $\mathbf{d}_l$ in the signal subspace, we have

$$\mathbf{U}_w^t\mathbf{d}_l = 0. \quad (10)$$

As we can see from our signal model in (4), our data block contains m shifted copies of the composite signature of the user of interest that have the form

$$\mathbf{d}_l = \begin{bmatrix} \mathbf{0}_{(l-1)\times 1} \\ \tilde{\mathbf{s}}_1 \\ \mathbf{0}_{(m-l)\times 1} \end{bmatrix}, \quad l = 1, \ldots, m. \quad (11)$$

Since these vectors lie in the signal subspace, they satisfy the orthogonality condition (10), and moreover, the same property holds for their sum as well. The latter fact can be expressed with the following equation:

$$\mathbf{U}_w^t\mathbf{F}_1\mathbf{f} = 0 \quad (12)$$

with

$$\mathbf{F}_1 = \sum_{l=1}^{m} \begin{bmatrix} \mathbf{0}_{(l-1)N\times L} \\ \mathbf{S}_1 \\ \mathbf{0}_{(m-l)N\times L} \end{bmatrix} \quad (13)$$

and where we have used (5). $\mathbf{F}_1$ is a *known* matrix with a particularly simple structure. It is a convolution matrix as in (6), but of dimensions $(mN+L-1)\times L$, with the first column containing the signature $\mathbf{s}_1$ repeated m times, i.e., of the form $[\underbrace{\mathbf{s}_1^t \cdots \mathbf{s}_1^t}_{m \text{ times}} \ \mathbf{0}_{1\times L-1}]^t$. We should note that when $m=1$, then $\mathbf{F}_1$ reduces to $\mathbf{S}_1$. From (12), we can now conclude that

$$\mathbf{U}_w^t\mathbf{F}_1\mathbf{f} = \left(\mathbf{F}_1^t\mathbf{U}_w\mathbf{U}_w^t\mathbf{F}_1\right)\mathbf{f} = 0. \quad (14)$$

Equation (14) suggests the recovery of $\mathbf{f}$ as the singular vector corresponding, again, to the *smallest* singular value (which here is equal to zero) of the matrix $\mathbf{F}_1^t\mathbf{U}_w\mathbf{U}_w^t\mathbf{F}_1$.

As it becomes clear from the preceding discussion, there are two subspace problems involved in (14). The first concerns the estimation of $\mathbf{U}_w$ and the second the recovery of the channel impulse response $\mathbf{f}$. Let us present these two problems more explicitly.

Problem 1: If $\mathbf{R}$ satisfies the decomposition in (9), we are interested in estimating the *projection* $\mathbf{U}_w\mathbf{U}_w^t$, where $\mathbf{U}_w$ is an orthonormal basis for the (noise) subspace corresponding to the *smallest* singular value σ^2 of $\mathbf{R}$.

Problem 2: If $\mathbf{R}$ and $\mathbf{U}_w$ are as in Problem 1 and $\mathbf{F}_1$ the matrix defined in (13), we are interested in estimating the singular vector $\mathbf{f}$ corresponding to the *smallest* singular value of the matrix

$$\mathbf{W} = \mathbf{F}_1^t\mathbf{U}_w\mathbf{U}_w^t\mathbf{F}_1. \quad (15)$$

In [6], both problems are solved by direct SVD, whereas in [7], the first is solved with SVD and the second with QR decomposition, which consitutes an orthogonal triangularization process of a matrix [15]. It is clear that applying SVD on $\mathbf{R}$ (or more accurately to its estimate) to recover $\mathbf{U}_w$ is computationally intense and disqualifies these methods from online processing. We should also mention the need of these approaches in *knowing the noise subspace rank*. It turns out [8] that even slight errors in the estimate of this parameter can produce significant performance degradation in the schemes proposed in [6] and [7]. In [8] and [9], a power method was proposed to replace the large SVD of Problem 1 used in [6] and [7]. This idea will be fully exploited in the next section in a direction that is suitable for both subspace problems introduced previously, but before raising this issue, let us first briefly discuss the consistency of the estimates provided by (14).

A. Consistency

Let r_s and r_n denote the signal and noise subspace ranks, respectively; then, the matrix $\mathbf{U}_w^t\mathbf{F}_1$ in (15) is of dimensions $r_n \times L$. If $\mathbf{U}_w$ is the exact noise subspace then, due to (12), we conclude that the column rank of $\mathbf{U}_w^t\mathbf{F}_1$ can *at most* be equal to $L-1$. In order for (14) to have a unique solution (modulo a multiplicative constant-ambiguity), the column rank of $\mathbf{U}_w^t\mathbf{F}_1$ must be *exactly equal* to $L-1$. Since the column rank of a

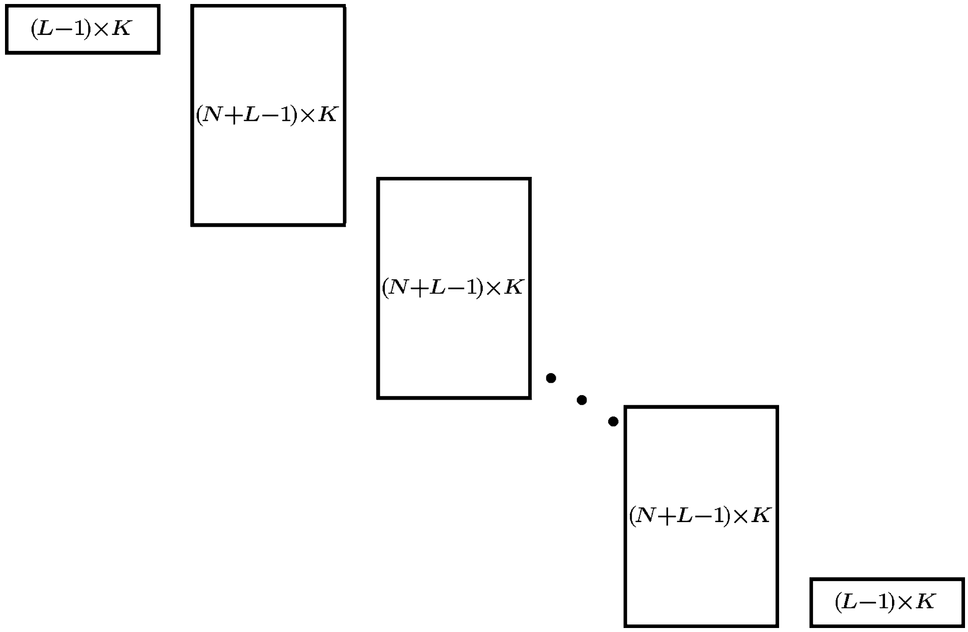

Fig. 1. Representation of the vectors composing the signal subspace.

matrix is equal to its row rank (and also equal to the rank of the matrix) in order to have a row rank equal to $L-1$, a *necessary* condition is to have at least $L-1$ rows, that is, $r_n \geq L-1$. Since $r_s + r_n = mN + L - 1$, this yields

$$r_s \leq mN. \tag{16}$$

Let us now specify, more precisely, the signal subspace rank. Notice that the number of columns of $\mathbf{U}_s$ is equal to r_s. In fact, $\mathbf{U}_s$ is an orthonormal basis for the subspace spanned by the vectors $\mathbf{d}_l$ introduced in (8). For the sake of clarity, we present these vectors in Fig. 1 for the downlink scenario. We recall that in this case, all K users are synchronized. As we can see, there are m big rectangles of dimensions $(N+L-1)\times K$ containing the entire composite signatures of all K users. The first such rectangle corresponds to the nth user-bits, whereas the last correspond to the $(n-m+1)$st. The two smaller rectangles, of dimensions $(L-1)\times K$, contain ISI coming from the $(n+1)$st and $(n-m)$th user-bits, respectively. Each rectangle has a rank that cannot exceed its *smallest* dimension. Assuming that the number of users K is smaller than the processing gain N, we conclude that

$$r_s \leq mK + 2\min\{L-1, K\}. \tag{17}$$

We therefore deduce that if we select m such that $mK + 2\min\{L-1, K\} \leq mN$, then the validity of the necessary condition (16) is guaranteed. This yields the following estimate for the number of blocks m:

$$m \geq \frac{2\min\{L-1, K\}}{N-K}. \tag{18}$$

Equivalently, for a given number of blocks m, we can obtain an upper bound for the maximum load of the system

$$K \leq N - 2\min\left\{\frac{N}{m+2}, \frac{L-1}{m}\right\}. \tag{19}$$

If we wish to follow the same analysis for the uplink scenario then, due to lack of synchronization, (17) becomes $r_s \leq (m+2)K$, yielding

$$m \geq \frac{2K}{N-K} \quad \text{or} \quad K \leq \frac{m}{m+2}N \tag{20}$$

as a possible estimate for m (for given K) or an upper bound for K (for given m). We must stress that the bounds introduced in (18) and (19) are by no means strict and must therefore be used with caution. We recall that they simply ensure validity of the *necessary condition* (16) and, thus, are not sufficient for identifiability. In numerous simulations, however, they turned out to be very accurate. In other words, whenever they were satisfied, the channel estimation was correct, whereas in the opposite case, examples appeared where identification failed. Unfortunately, we were not able to prove their sufficiency.

Finally, in a situation where the channel length is not available, we can assume that L plays the role of a known upper bound for the true parameter. In such a case, similarly to [14], an additional necessary condition for identifiability is needed. Specifically, the difference between the upper bound L and the true filter length L' must be strictly less than the processing gain N, i.e., $L - L' < N$. This is so because in the opposite case, one can easily produce two different solutions for (14), namely, $[\mathbf{f}^t\ \mathbf{0}_{1\times(L-L')}]$ and $[\mathbf{0}_{1\times N}\ \mathbf{f}^t\ \mathbf{0}_{1\times(L-L'-N)}]$, where $\mathbf{f}$ is the true channel impulse response, and L' is its corresponding length. Since *any linear combination* of these two solutions is also a solution of (14), we conclude that there is an infinite number of candidates for the role of the channel impulse response.

IV. Power Method Variants

The power method [15] is an *iterative* technique that is used to provide estimates of the subspace corresponding to the *largest* singular value of a matrix. Let us present two variants of this method that are appropriate for solving the two subspace problems of interest that will also serve as a starting point for developing our adaptive algorithms.

Lemma 1: Let $\mathbf{R}$ be as in (7) with an SVD as in (9) and $\rho \geq 0$ a nonnegative scalar. We then have

$$\lim_{k\to\infty}\left(\frac{\rho\mathbf{I}+\mathbf{R}}{\rho+\sigma^2}\right)^{-k} = \mathbf{U}_w\mathbf{U}_w^t. \tag{21}$$

Proof: The proof is straightforward. Using the decomposition of $\mathbf{R}$ defined in (9), we have the following limit as $k \to \infty$:

$$\begin{aligned}\left(\frac{\rho\mathbf{I}+\mathbf{R}}{\rho+\sigma^2}\right)^{-k} &= [\mathbf{U}_s\ \mathbf{U}_w]\begin{bmatrix}\left(\mathbf{I}+\frac{\mathbf{\Lambda}_s}{\rho+\sigma^2}\right)^{-k} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}[\mathbf{U}_s\ \mathbf{U}_w]^t\\ &\to [\mathbf{U}_s\ \mathbf{U}_w]\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}[\mathbf{U}_s\ \mathbf{U}_w]^t\\ &= \mathbf{U}_w\mathbf{U}_w^t\end{aligned} \tag{22}$$

with the last limit being true since $\mathbf{I}+(\mathbf{\Lambda}_s/(\rho+\sigma^2))$ is a diagonal matrix with diagonal elements strictly greater than unity. ∎

It is clear that Lemma 1 contributes to the solution of the first subspace problem, i.e., the estimation of the product $\mathbf{U}_w\mathbf{U}_w^t$ required in (14). The form of the power method stated in (21) is slightly more general (due to the existence of the parameter ρ) than the one used in [8] and [9]; the latter corresponds to $\rho = 0$. The extra degree of freedom provided by ρ will turn out to be very helpful in the development of our adaptive algorithms. To analyze the behavior of this more generalized form of the power method, we can apply the results contained in [9], replacing σ^2 with $\rho + \sigma^2$.

As seen from (14), the channel vector $\mathbf{f}$ is the singular vector corresponding to the smallest singular value (zero in the ideal case) of $\mathbf{W}$; consequently, we can again apply the power method for its estimation. We propose the following variant.

Lemma 2: Let $\mathbf{W}$ be the matrix defined in (15); suppose that the vector $\mathbf{f}$, which satisfies (14), is unique and of unit norm; then, with $\alpha = 1/\mathrm{tr}\{\mathbf{W}\}$, the sequence of vectors $\mathbf{f}(n)$ defined by the recursion

$$\mathbf{f}(n) = \frac{(\mathbf{I}-\alpha\mathbf{W})\mathbf{f}(n-1)}{\|(\mathbf{I}-\alpha\mathbf{W})\mathbf{f}(n-1)\|} \tag{23}$$

converges to the channel impulse response $\pm\mathbf{f}$ (modulo a sign ambiguity), provided that $\mathbf{f}(0)$ is not orthogonal to $\mathbf{f}$.

Proof: Again, the proof presents no particular difficulty. Using induction, we can show that

$$\mathbf{f}(n) = \frac{(\mathbf{I}-\alpha\mathbf{W})^n\mathbf{f}(0)}{\|(\mathbf{I}-\alpha\mathbf{W})^n\mathbf{f}(0)\|}. \tag{24}$$

Since $\mathbf{Wf} = 0$, this means that $\mathbf{f}$ is a singular vector for the matrix $\mathbf{I}-\alpha\mathbf{W}$ corresponding to the unit singular value (which is the largest since $\alpha\mathbf{W}$ is non-negative definite with all singular values smaller than unity). Using SVD, as in Lemma 1, we can show that

$$\lim_{n\to\infty}(\mathbf{I}-\alpha\mathbf{W})^n = \mathbf{f}\mathbf{f}^t \tag{25}$$

which yields

$$\lim_{n\to\infty}\mathbf{f}(n) = \mathrm{sgn}\left(\mathbf{f}^t\mathbf{f}(0)\right)\mathbf{f}. \tag{26}$$

This concludes the proof. ∎

Lemma 2 contributes to the solution of the second subspace problem and will provide the necessary channel impulse response estimates. From (23), we realize that this time, we did not apply the power method to the inverse of $\mathbf{W}$, but rather to its difference from the identity matrix. With this idea, we reduced the corresponding computational complexity by an order of magnitude since inversion requires $O(L^3)$ operations, whereas the proposed scheme requires $O(L^2)$. As we will see later, this simplification exhibits no significant performance loss when compared to direct SVD methods. At this point, we can make the following important remarks.

Remark 1: In Lemma 1, the convergence in (22) is exponential, and we observe that the corresponding rate is maximized when $\rho = 0$. Regardless of this fact, the employment of a $\rho > 0$ in the scheme will turn out to be particularly useful in the case of the LMS version since it will allow the algorithm to *forget past data* much more rapidly than the usual LMS with $\rho = 0$. In the exponentially windowed RLS version, on the other hand, we can select $\rho = 0$ since RLS has the inherent capability of forgetting past data through the forgetting factor.

Remark 2: A subtle and very important remark regarding Lemma 1 concerns the power k. Notice that the limit is correct, i.e., we obtain the projection $\mathbf{U}_w\mathbf{U}_w^t$ *only* when the singular values corresponding to the noise subspace are *exactly* equal. Unfortunately, in a realistic situation, when only estimates of $\mathbf{R}$ are available, this is rarely the case. This has a grave consequence since the corresponding limit instead of being the desired product will just become the rank-one matrix $\mathbf{u}\mathbf{u}^t$, where $\mathbf{u}$ is the singular vector corresponding to the smallest singular value of the *estimate* of $\mathbf{R}$. This in turn will make $\mathbf{W}$ a rank-one matrix as well, and thus, $\mathbf{f}$ will no longer be the only vector satisfying (14).

Fortunately, for CDMA signals, there is a simple remedy to this problem. In [8], it was observed that for offline processing, it was sufficient to use powers up to $k = 3$ and practically match the performance of the direct SVD-based techniques. We are going to follow the same idea here. In other words, we propose to approximate the product $\mathbf{U}_w\mathbf{U}_w^t$ as follows:

$$\widehat{\mathbf{U}_w\mathbf{U}_w^t} = \left(\frac{\rho\mathbf{I}+\mathbf{R}}{\rho+\sigma^2}\right)^{-k}, \quad k = 1, 2, 3. \tag{27}$$

Remark 3: Our final remark concerns the usage of (27). Notice that in approximating $\mathbf{U}_w\mathbf{U}_w^t$ this way, *we do not need any knowledge of the noise subspace rank*. This is particularly convenient since the slightest erroneous rank estimate can produce drastic performance degradation in direct SVD methods [8].

We are now ready to proceed with the presentation of our blind adaptive schemes.

V. Blind Adaptations for Channel Estimation

As stated in Problem 2, the channel impulse response can be recovered as the singular vector corresponding to the smallest singular value of the matrix $\mathbf{W} = \mathbf{F}_1^t \mathbf{U}_w \mathbf{U}_w^t \mathbf{F}_1$. Using the approximation proposed in (27), we have the following estimate for this matrix:

$$\mathbf{W}_k = \mathbf{F}_1^t (\rho \mathbf{I} + \mathbf{R})^{-k} \mathbf{F}_1 \tag{28}$$

where we have discarded the scalar quantity $(\rho + \sigma^2)^k$ since it does not affect the subspace determination problem.

When the autocorrelation matrix $\mathbf{R}$ is not available, we are interested in producing adaptive estimates $\mathbf{W}_k(n)$ of the matrix $\mathbf{W}_k$ defined in (28). There are different possibilities that we exploit next. Notice, however, that with the help of any such estimate $\mathbf{W}_k(n)$, the power method presented in Lemma 2 (which provides the final channel impulse response estimates) can be modified as follows:

$$\mathbf{f}(n) = \frac{(\mathbf{I} - \alpha(n)\mathbf{W}_k(n))\,\mathbf{f}(n-1)}{\|(\mathbf{I} - \alpha(n)\mathbf{W}_k(n))\,\mathbf{f}(n-1)\|} \tag{29}$$

where $\alpha(n) = 1/\mathrm{tr}\{\mathbf{W}_k(n)\}$. In other words, at every time step, we first apply a time adaptation of $\mathbf{W}_k(n)$ and then a *single iteration* of the power method. Let us now examine what possibilities exist for the estimate $\mathbf{W}_k(n)$.

A. Channel Estimation via RLS

As was mentioned previously, here, we select $\rho = 0$, and for the adaptive estimate $\mathbf{W}_k(n)$ of the matrix $\mathbf{W}_k$, we propose

$$\mathbf{W}_k(n) = \mathbf{F}_1^t \mathbf{P}^k(n) \mathbf{F}_1, \quad k = 1, 2, 3 \tag{30}$$

where $\mathbf{P}(n) = \mathbf{R}^{-1}(n)$, and $\mathbf{R}(n)$ is the exponentially windowed sample autocorrelation matrix of the data $\mathbf{r}(n)$, i.e., $\mathbf{R}(n) = \sum_{i=0}^{n} \lambda^{n-i} \mathbf{r}(i)\mathbf{r}^t(i)$, with $0 < \lambda < 1$: a forgetting factor. We recall the well-known RLS adaptation for $\mathbf{P}(n)$

$$\mathbf{k}(n) = \mathbf{P}(n-1)\mathbf{r}(n) \tag{31}$$

$$\gamma(n) = \frac{1}{(\lambda + \mathbf{r}^t(n)\mathbf{k}(n))} \tag{32}$$

$$\mathbf{P}(n) = \frac{1}{\lambda}\left(\mathbf{P}(n-1) - (\gamma(n)\mathbf{k}(n))\,\mathbf{k}^t(n)\right) \tag{33}$$

that has an overall complexity of $5(mN + L - 1)^2 + 3(mN + L - 1) + O(1)$ (counting together multiplications and additions). If we now compute $\mathbf{W}_k(n)$ as

$$\mathbf{W}_k(n) = \mathbf{F}_1^t \left(\mathbf{P}(n) \cdots (\mathbf{P}(n)\,(\mathbf{P}(n)\mathbf{F}_1)) \cdots\right) \tag{34}$$

then this part requires $2kL(mN+L-1)^2 + 2L^2(mN+L-1) + O(1)$ operations. Finally, once $\mathbf{W}_k(n)$ is available, the adaptation in (29) requires $2L^2 + 5L + O(1)$ operations. It is clear that the most computationally intense part is the computation of $\mathbf{W}_k(n)$ in (34).

We should mention that our RLS version is similar to the adaptive algorithm presented in [9], and when $k = 1$, it is similar to the RLS version of [10]. The advantage here is that we avoid the SVD on $\mathbf{W}_k(n)$ proposed in [9] and [10] since we replace it with *one iteration* of the power recursion (29). As far as the RLS version of [10] is concerned, we are going to see in our simulations that higher values of the power k can ameliorate performance significantly.

B. Channel Estimation via Leakage LMS

This is the most practically important part of our work. The LMS scheme we are going to present is computationally simple with performance that can be orders of magnitude better than the corresponding LMS adaptation of [10].

An alternative means to generate estimates for $\mathbf{W}_k$ consists of writing

$$\mathbf{W}_k = \mathbf{F}_1^t \mathbf{V}_k \tag{35}$$

where

$$\mathbf{V}_k = (\rho \mathbf{I} + \mathbf{R})^{-k} \mathbf{F}_1 \tag{36}$$

and produce estimates $\mathbf{V}_k(n)$ for $\mathbf{V}_k$. It turns out that LMS is particularly suited for this task. Consider first $k = 1$, and define the recursion

$$\mathbf{V}_1(n) = \lambda_{\mathrm{L}} \mathbf{V}_1(n-1) + \mu\left(\mathbf{F}_1 - \mathbf{r}(n)\mathbf{r}^t(n)\mathbf{V}_1(n-1)\right) \tag{37}$$

where $0 < \lambda_{\mathrm{L}} < 1$ is a leakage factor. By taking expectations and evoking the *Independence Assumption*, i.e., assuming that $\mathbf{V}_1(n-1)$ is independent from the data vector $\mathbf{r}(n)$, we can verify that

$$\lim_{n\to\infty} \mathbb{E}[\mathbf{V}_1(n)] = \left(\frac{1 - \lambda_{\mathrm{L}}}{\mu}\mathbf{I} + \mathbf{R}\right)^{-1} \mathbf{F}_1 \tag{38}$$

which is exactly (36) with $\rho = (1 - \lambda_{\mathrm{L}})/\mu$ and $k = 1$. Estimates for higher powers can be obtained with the following time and order recursion

$$\mathbf{V}_l(n) = \lambda_{\mathrm{L}} \mathbf{V}_l(n-1) + \mu\left(\mathbf{V}_{l-1}(n-1) - \mathbf{r}(n)\mathbf{r}^t(n)\mathbf{V}_l(n-1)\right) \tag{39}$$

for $l = 1, 2, \ldots, k$, where $\mathbf{V}_0(n) = \mathbf{V}_l(0) = \mathbf{F}_1$. With our next theorem, we analyze the mean behavior of the recursion in (39), thus generalizing (38).

Theorem 1: Let $\mathbf{V}_l(n)$, $l = 1, \ldots, k$ be as in (39), define $\rho = (1 - \lambda_{\mathrm{L}})/\mu$, and let $\mathbf{V}_l = (\rho \mathbf{I} + \mathbf{R})^{-l} \mathbf{F}_1$ be the expression defined in (36) corresponding to the power l; then, under the Independence Assumption, we have for $l = 1, \ldots, k$

$$\mathbb{E}[\mathbf{V}_l(n)] = \mathbf{V}_l + \sum_{j=0}^{l-1} \mu^j \binom{n}{j} (\lambda_{\mathrm{L}} \mathbf{I} - \mu \mathbf{R})^{n-j} (\mathbf{V}_{l-j}(0) - \mathbf{V}_{l-j}). \tag{40}$$

Proof: The proof can be found in the Appendix. ■

Although the Independence Assumption, strictly speaking, is erroneous, it has become a popular tool for analyzing adaptive algorithms. It turns out that the conclusions obtained by using it are correct at least up to a second-order approximation in the step size μ, this being true for a great variety of adaptive algorithms (including LMS and Leakage LMS) and rich classes of data models [16], [17].

From Theorem 1, we observe that due to the term $(\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^{n-j}$, we have an exponential convergence of $\mathbb{E}[\mathbf{V}_l(n)]$ toward the desired quantity $\mathbf{V}_l = (\rho \mathbf{I} + \mathbf{R})^{-l} \mathbf{F}_1$. In fact, the speed of convergence is governed by the largest eigenvalue of the matrix $\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R}$, which is equal to $\lambda_{\mathrm{L}} - \mu\sigma^2$. Notice that by using the Leakage LMS recursion in (39) with $0 < \lambda_{\mathrm{L}} < 1$, we assure an exponential convergence with a factor that is at least equal to λ_{L}, independently of the SNR level. If, on the other hand, one uses the regular LMS with $\lambda_{\mathrm{L}} = 1$, the corresponding factor becomes $1 - \mu\sigma^2$, which can induce an extremely small convergence rate in medium to high SNR. In the latter case, we cannot ameliorate the LMS convergence speed by simply increasing the step size μ (as it is the usual practice in adaptive algorithms). The reason is that when we increase μ, LMS becomes unstable well before we can reach any satisfactory convergence rate levels. It is therefore through Leakage LMS that we can bypass this serious handicap of the classical LMS algorithm.

Since, with Theorem 1, we have established that the recursion in (39) can provide proper estimates for $\mathbf{V}_k$ using $\mathbf{V}_k(n)$, we can now obtain estimates for $\mathbf{W}_k$ following (35) as

$$\mathbf{W}_k(n) = \mathbf{F}_1^t \mathbf{V}_k(n). \tag{41}$$

Finally, we apply one iteration of the power method in (29) to obtain the estimate $\mathbf{f}(n)$ of the channel impulse response.

The computational complexity of the proposed scheme is as follows. We need $7kL(mN+L-1)+O(1)$ operations for the adaptation in (39); $2L^2(mN+L-1)+O(1)$ for the computation of $\mathbf{W}_k(n) = \mathbf{F}_1^t\mathbf{V}_k(n)$ and, finally, $2L^2+5L+O(1)$ operations for the iteration in (29). As in the case of RLS, the most computationally heavy part is the one needed for $\mathbf{W}_k(n)$. It turns out that we can reduce this complexity by a factor that can be important. If we multiply the recursion in (39) from the left by $\mathbf{F}_1^t$, we obtain a time and order recursion for $\mathbf{W}_l(n)$, $l = 1, \ldots, k$

$$\begin{aligned}\mathbf{W}_l(n) &= \lambda_{\mathrm{L}}\mathbf{W}_l(n-1)\\ &\quad+\mu\left(\mathbf{W}_{l-1}(n-1) - \mathbf{F}_1^t\mathbf{r}(n)\mathbf{r}^t(n)\mathbf{V}_l(n-1)\right).\end{aligned} \tag{42}$$

Since all vectors $\mathbf{r}^t(n)\mathbf{V}_l(n-1)$ are available from (39), we only need to compute $\mathbf{F}_1^t\mathbf{r}(n)$ once and then form the $L\times L$ matrices $\mathbf{F}_1^t\mathbf{r}(n)\mathbf{r}^t(n)\mathbf{V}_l(n-1)$ appearing in (42). The total complexity for computing $\mathbf{W}_k(n)$ with this scheme is $2kL(mN+L-1)+5kL^2+O(1)$ operations, which should be compared to $2L^2(mN+L-1)+O(1)$ required by (41). Since, usually, $k < L$, this can result in a non-negligible computational gain. The price we pay for using (42) is the need to store the matrices $\mathbf{W}_l(n)$, $l = 1, \ldots, k$. This is clearly unnecessary when we use (41).

VI. Simulations—Comparisons

In this section, we provide several simulation results to demonstrate the performance of the blind adaptive schemes developed previously. In particular, we compare our RLS and LMS implementations with the corresponding schemes proposed in [10]. Before getting into our simulations, we must point out that we are going to examine the behavior of our algorithms under diverse signaling conditions with the received signal exhibiting drastic changes in its power. In such cases, it is advisable to use a *normalized* version of the data in order to account for signal power changes and obtain an algorithm that is relatively insensitive to them. We propose the following simple modification of (39):

$$\begin{aligned}\mathbf{V}_l(n) &= \lambda_{\mathrm{L}}\mathbf{V}_l(n-1)\\ &\quad+\mu\left(\mathbf{V}_{l-1}(n-1) - \frac{\mathbf{r}(n)\mathbf{r}^t(n)}{\beta(n)}\mathbf{V}_l(n-1)\right)\end{aligned} \tag{43}$$

where $\beta(n)$ is an estimate of the received signal power. A possible adaptive scheme for $\beta(n)$ is

$$\beta(n) = \nu\beta(n-1) + (1-\nu)\left\|\mathbf{r}(n)\right\|^2, \quad n > 1 \tag{44}$$

where $0 < \nu < 1$, and $\beta(1) = \|\mathbf{r}(1)\|^2$. We can now proceed with our simulations.

Randomly generated sequences of length $N = 128$ are used as spreading codes. Once generated, the codes are kept constant for the whole simulation set. Moreover, all graphs presented in the figures are the result of an average of 100 independent runs. In each run, we apply three different abrupt changes in order to observe the ability of the corresponding algorithms to follow them. Specifically, at bit 5000, we change the channel, and at bits 10 000 and 15 000, we change the number of users. For the multipath channel, we start with the length 3 "difficult" channel (containing a deep null) $\mathbf{f}_d = [0.407\ 0.815\ 0.407]^t$, and at 5000, we switch to the length 10 "easy" channel $\mathbf{f}_e = [0.04\ -0.05\ 0.07\ -0.21\ -0.5\ 0.72\ 0.36\ 0.21\ 0.03\ 0.07]^t$, which were both proposed in [18]. For our estimation, on the other hand, we assume that we have available only an upper bound for the channel length, which is $L = 10$. In other words even the length 3 channel is identified as being of length 10. We use only one data block, that is, $m = 1$.

The signaling conditions are as follows: We start with $K = 55$ users, under perfect power control. At bit 10 000, ten additional users enter the channel, five of them having power equal to the user of interest and the remaining five being 10 dB stronger. Finally, at bit 15 000, the last ten users along with five more exit the channel. As we can verify, the constraint in (19) is always satisfied. Finally, for ambient additive noise, we used zero-mean white Gaussian noise.

In our simulations, we have also included the *direct SVD approach*, that is, the solution of both subspace problems through direct SVD, where we have used the exact value of the noise subspace rank. This approach can be clearly regarded as a point of reference for all adaptive methods under comparison. However, due to its exceedingly large computational complexity, it was performed only once every 500 bits. We used the following parameter values in our algorithms: $\lambda = \lambda_{\mathrm{L}} = 0.998$, $\mu = 1$, and $\nu = 0.99$.

Fig. 2 depicts the mean square channel estimation error of the RLS schemes when the SNR of the user of interest is equal to 20 dB. We can see that our $k = 1$ version practically matches the RLS of [10] without needing an SVD on the matrix $\mathbf{W}_1(n)$ at each step. By employing higher powers $k = 2$,3, there is a slight performance improvement only in the beginning. After the channel changes at bit 5000, all RLS algorithms converge quickly to their new steady state. Moreover, the behavior of the RLS schemes is not affected by changes in the number of users (i.e. bits 10 000 and 15 000). It is clear that in this high SNR environment, selecting $k = 1$ is sufficient. We can also see that

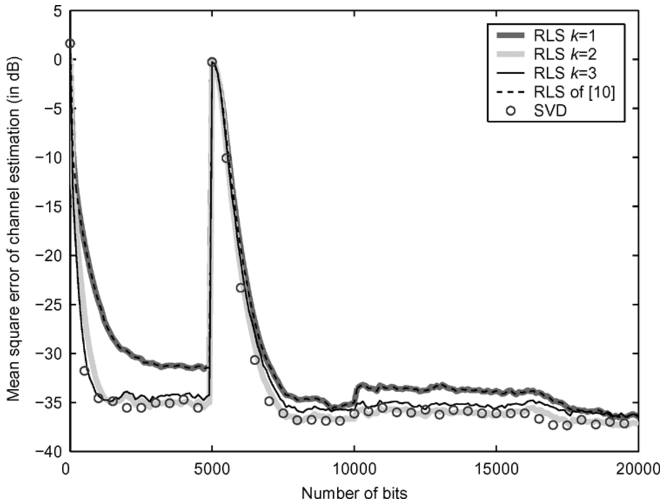

Fig. 2. Performance of the proposed RLS channel estimation scheme and the RLS version of [10]; noise power 20 dB.

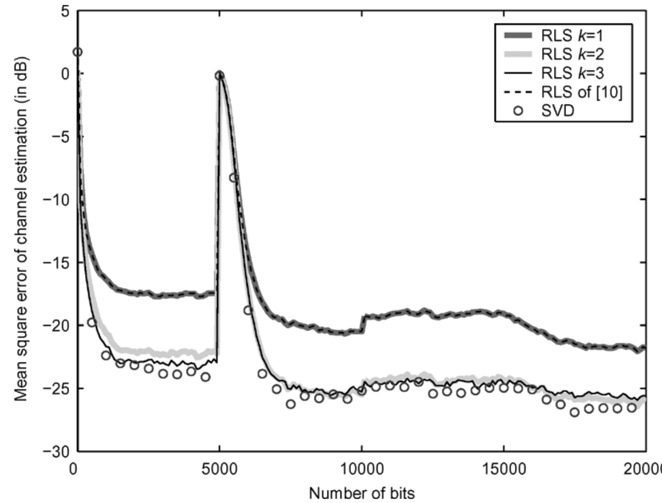

Fig. 4. Performance of the proposed RLS channel estimation scheme and the RLS version of [10]; noise power 10 dB.

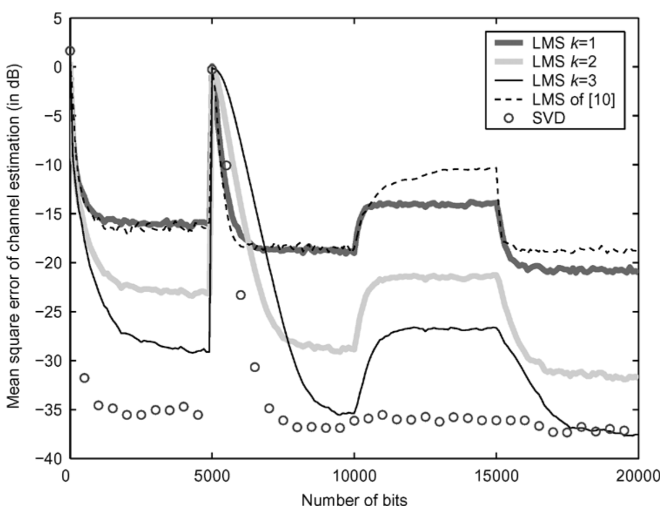

Fig. 3. Performance of the proposed LMS channel estimation scheme and the LMS version of [10]; noise power 20 dB.

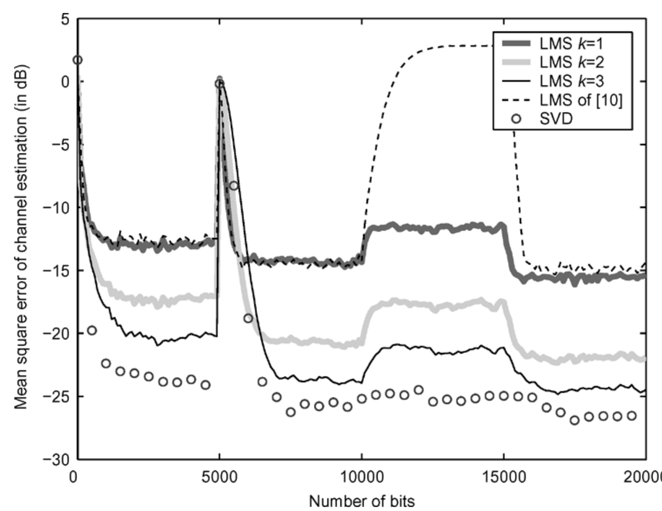

Fig. 5. Performance of the proposed LMS channel estimation scheme and the LMS version of [10]; noise power 10 dB.

our RLS with $k = 3$ practically matches the performance of the *direct SVD approach* at a significantly lower computational cost and without the need of knowledge of the noise subspace rank.

We continue with the presentation of the LMS schemes. In fact, we apply the LMS adaptation in (43) as well as the LMS algorithm of [10] after selecting its step size so that its steady-state performance matches our $k = 1$ version. Fig. 3 depicts the performance of all competing LMS schemes. The LMS algorithm of [10], as we can see, has similar performance to our $k = 1$ version but exhibits a smaller robustness to abrupt changes in the number of users. Furthermore, in contrast to RLS, here, we obtain substantial performance gains by employing higher powers in k. The difference between our $k = 3$ version and our $k = 1$ version (and therefore the LMS version of [10]) reaches almost 20 dB at steady state. Clearly, our LMS with $k = 3$ presents a slightly inferior performance as compared with its RLS counterpart or the direct SVD method, which ranges from 0 to 5 dB.

Next, we consider the same signaling scenario but with a significantly lower SNR. Specifically, we set the desired user's SNR to 10 dB. The performance of the RLS schemes is presented in Fig. 4. Again, the method of [10] is identical to our RLS $k = 1$ version. Here, however, in this low SNR environment, employing higher values of k ameliorates the overall RLS performance significantly, especially in the initial part, i.e., the case of the "difficult" channel. On the other hand, our $k = 3$ version again matches the performance of the direct SVD method. In Fig. 5, we have the corresponding LMS schemes. We observe that the LMS algorithm of [10] has similar performance to our LMS version with $k = 1$. As in the previous case, our $k = 3$ version outperforms our $k = 1$ version and, consequently, the LMS algorithm of [10], by more than 10 dB.

The next two figures (Figs. 6 and 7) concern the same scenario as in the previous two simulations, only here, the channel impulse response also experiences slow fading. At the beginning and at time instant 5000, the channels are initialized with the values $\mathbf{f}_d$ and $\mathbf{f}_e$, respectively; then, we use the Jakes-like model proposed in [19] in order to simulate a slowly fading multipath

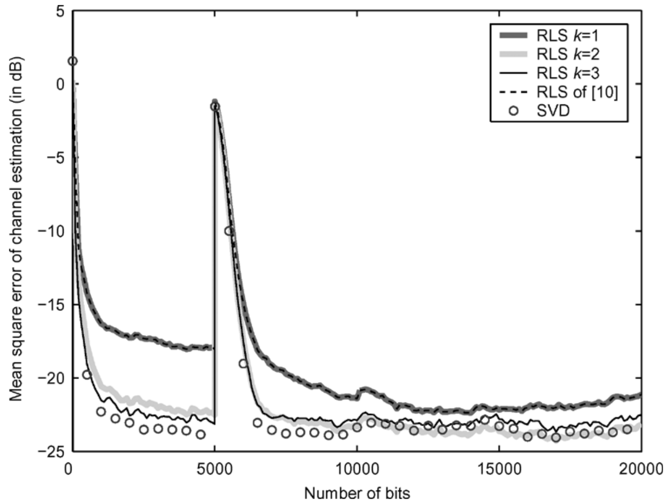

Fig. 6. Performance of the proposed RLS channel estimation scheme and the RLS version of [10] in a slowly fading environment; noise power 10 dB.

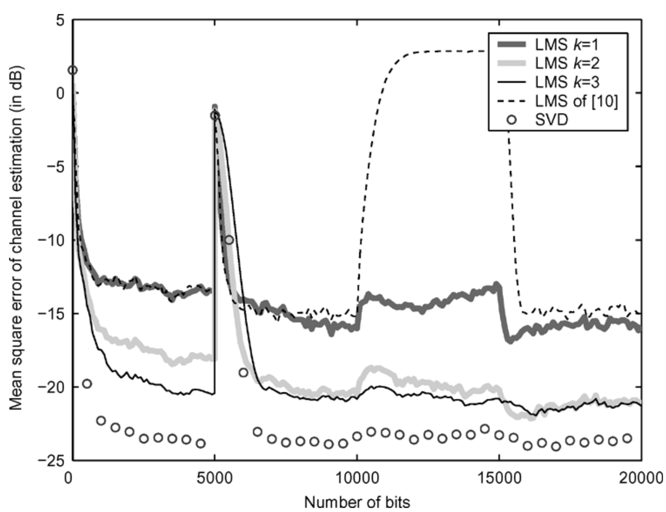

Fig. 7. Performance of the proposed LMS channel estimation scheme and the LMS version of [10] in a slowly fading environment; noise power 10 dB.

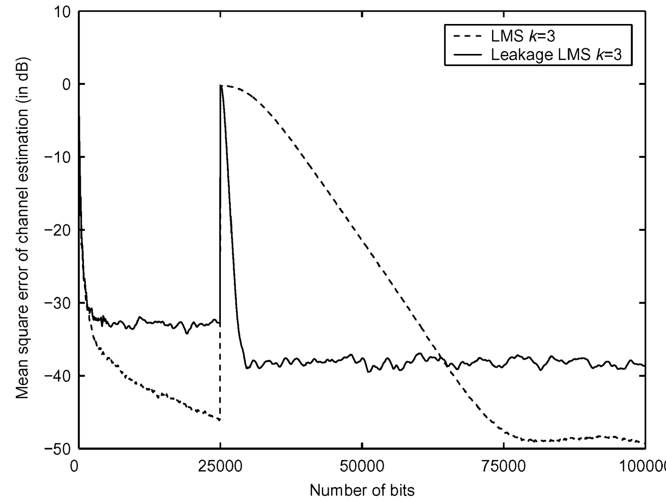

Fig. 8. Convergence characteristics of Leakage LMS (solid) and normal LMS (dashed) after an abrupt change in the channel; noise power 20 dB.

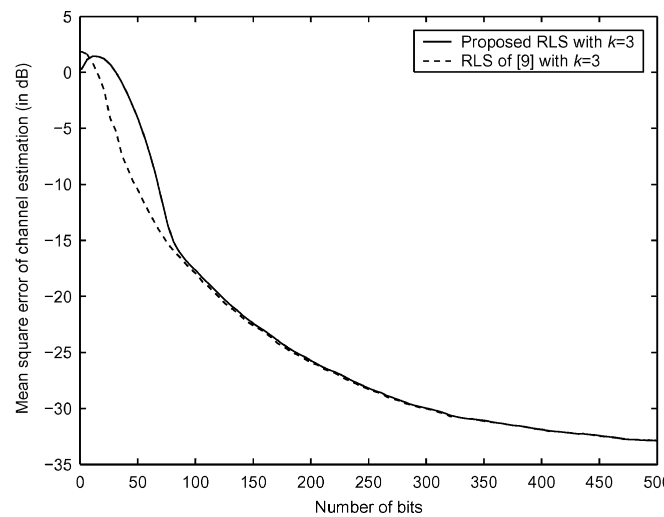

Fig. 9. Performance of the proposed RLS scheme versus the RLS of [9]; power parameter $k = 3$; noise power 20 dB.

channel. The parameters of this model are communication frequency carrier at 900 MHz, 15 scatterers for every channel coefficient (scatterers for different channel taps are independent), data rate of 2 Mbits/s, and SNR of 10 dB. Fig. 6 depicts the RLS versions and Fig. 7 the corresponding LMS. We can see that both proposed versions are capable of efficiently following variations in the channel impulse response. Again, our RLS $k = 3$ algorithm is very close to the direct SVD method. Compared with the nonfading case, we observe that all algorithms exhibit a 2-dB performance reduction due to tracking.

Comparing the RLS with the LMS schemes, we clearly observe the considerably more robust behavior of the former to changes in the number of users. This performance, unfortunately, comes at an increased computational cost.

A. Other Comparisons

Let us now compare the recursion in (43) for $0 < \lambda_{\mathrm{L}} < 1$ (Leakage LMS) with the case $\lambda_{\mathrm{L}} = 1$ (normal LMS). The signaling scenario depicted in Fig. 8 is similar to our first example. We start with the "difficult" channel $\mathbf{f}_d$, and at bit 25 000, we switch to the "easy" one $\mathbf{f}_e$. We run the simulation long enough to allow both algorithms to converge. We present only the case $k = 3$. The parameters of the algorithms are as before: $\mu = 1$; leakage factor $\lambda_{\mathrm{L}} = 0.998$; $\nu = 0.99$; and SNR level equal to 20 dB. The noticeable difference in convergence speed between the two versions is accompanied by an equivalent difference (in the opposite sense) in the steady-state behavior. As was mentioned previously, it is *not* possible to trade between steady-state performance and convergence speed in the usual $(\lambda_{\mathrm{L}} = 1)$ LMS algorithm by simply changing the step size μ. In fact, the value $\mu = 1$ used here is rather limiting since any slight increase in this parameter leads LMS to instability. Consequently, the performance of LMS depicted in Fig. 8 is the best this algorithm can offer, as far as convergence speed is concerned. It is through Leakage LMS that we can therefore obtain an improvement in the convergence speed of the algorithm with the analogous, of course, loss in steady-state behavior.

Finally, in Fig. 9, we depict the channel estimates of the proposed RLS scheme versus the RLS of [9], both with the power parameter set to $k = 3$. The dashed line corresponds to the method of [9] that applies an SVD at every time step on the matrix $\mathbf{W}_k(n)$, whereas the solid line corresponds to our method, where we apply the iteration in (29) only once. We can see that the corresponding channel estimates become indistinguishable after very few steps, with our method requiring an order of magnitude less operations than the direct SVD ($O(L^2)$ versus $O(L^3)$). This was the reason why the method of [9] was not included in our RLS simulations.

VII. Conclusion

In this work, we examined the blind adaptive channel estimation problem for CDMA in multipath additive white noise channels and considered a two-step methodology for its solution that is similar to [6] and [7]. The novelty of our approach consists of specifying two subspace problems, which we solve via two different variants of the power method. RLS and LMS algorithms are subsequently developed that implement adaptively the two power method variants providing efficient estimates for the channel impulse response. With a number of simulations, we demonstrate the satisfactory performance of our adaptive schemes in a dynamic environment that exhibits abrupt changes in the channel impulse response and the number of users. Compared with the adaptive methods proposed in [10], our schemes offer substantial performance gains at similar computational cost. Finally, the adaptive scheme of [9], although it is similar to our RLS version, still requires an SVD of the size of the channel length. In our case, this part is replaced by a single iteration of the power method that has an order-of-magnitude smaller computational complexity.

Appendix
Proof of Theorem 1

We will prove the validity of (40) by induction in the power l. Let us first show that (40) is true for $l = 1$. Taking expectation in (39), using the independence assumption, and recalling that $\mathbf{V}_0(n) = \mathbf{F}_1$, we obtain the following recursion:

$$\mathbb{E}\left[\mathbf{V}_1(n)\right] = (\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})\mathbb{E}\left[\mathbf{V}_1(n-1)\right] + \mu\mathbf{F}_1 \tag{45}$$

which readily leads to

$$\mathbb{E}\left[\mathbf{V}_1(n)\right] = \mu\sum_{j=0}^{n-1}(\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^j\mathbf{F}_1 + (\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^n\mathbf{V}_1(0). \tag{46}$$

Using the fact that $\sum_{j=0}^{n-1}\mathbf{A}^j = (\mathbf{I} - \mathbf{A}^n)(\mathbf{I} - \mathbf{A})^{-1}$, the previous equality becomes

$$\mathbb{E}\left[\mathbf{V}_1(n)\right] = (\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^n\mathbf{V}_1(0) + (\mathbf{I} - (\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^n)\underbrace{(\rho\mathbf{I} + \mathbf{R})^{-1}\mathbf{F}_1}_{\mathbf{V}_1} \tag{47}$$

which is exactly (40) for $l = 1$.

Let us now assume that (40) is true for $l = \kappa$, that is

$$\mathbb{E}\left[\mathbf{V}_\kappa(n)\right] = \mathbf{V}_\kappa + \sum_{j=0}^{\kappa-1}\mu^j\binom{n}{j}(\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^{n-j}\boldsymbol{\Delta}\mathbf{V}_{\kappa-j} \tag{48}$$

where $\boldsymbol{\Delta}\mathbf{V}_i$ is defined as

$$\boldsymbol{\Delta}\mathbf{V}_i = \mathbf{V}_i(0) - \mathbf{V}_i. \tag{49}$$

We will then prove that it is also true for $l = \kappa + 1$. Consider the recursion in (39) for $l = \kappa + 1$. After taking expectation and applying the independence assumption, i.e., that $\mathbf{V}_{\kappa+1}(n-1)$ is independent from the received data $\mathbf{r}(n)$, we obtain the recursion

$$\mathbb{E}\left[\mathbf{V}_{\kappa+1}(n)\right] = (\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})\mathbb{E}\left[\mathbf{V}_{\kappa+1}(n-1)\right] + \mu\mathbb{E}\left[\mathbf{V}_\kappa(n-1)\right] \tag{50}$$

which yields

$$\mathbb{E}\left[\mathbf{V}_{\kappa+1}(n)\right] = \mu\sum_{i=0}^{n-1}(\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^i\mathbb{E}\left[\mathbf{V}_\kappa(n-i-1)\right] + (\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^n\mathbf{V}_{\kappa+1}(0). \tag{51}$$

Substituting $\mathbb{E}[\mathbf{V}_\kappa(n-i-1)]$ with its equal from (48), we have

$$\mathbb{E}\left[\mathbf{V}_{\kappa+1}(n)\right] = +\sum_{i=0}^{n-1}\sum_{j=0}^{\kappa-1}\mu^{j+1}\binom{n-i-1}{j}(\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^{n-j-1}\boldsymbol{\Delta}\mathbf{V}_{\kappa-j} + \mu\sum_{i=0}^{n-1}(\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^i\mathbf{V}_\kappa + (\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^n\mathbf{V}_{\kappa+1}(0). \tag{52}$$

The sum of the last two terms in (52), using the fact that $(\rho\mathbf{I} + \mathbf{R})^{-1}\mathbf{V}_\kappa = \mathbf{V}_{\kappa+1}$, is equal to

$$\mathbf{V}_{\kappa+1} + (\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^n\boldsymbol{\Delta}\mathbf{V}_{\kappa+1}. \tag{53}$$

Furthermore, changing the order of summation in (52), using the property that $\sum_{i=0}^{m}\binom{j+i}{j} = \binom{j+m+1}{j+1}$ and changing variables $p = j + 1$, the double summation in (52) yields

$$\sum_{p=1}^{\kappa}\mu^p\binom{n}{p}(\lambda_{\mathrm{L}}\mathbf{I} - \mu\mathbf{R})^{n-p}\boldsymbol{\Delta}\mathbf{V}_{\kappa+1-p}. \tag{54}$$

By adding this to (53), we obtain the desired result.

References

[1] M. Honig, U. Madhow, and S. Verdú, "Blind adaptive multiuser detection," *IEEE Trans. Inf. Theory*, vol. 41, pp. 944–960, Jul. 1995.

[2] G. V. Moustakides, "Constrained adaptive linear multiuser detection schemes," *J. VLSI Signal Process., Special Issue on Multiuser Detection*, vol. 30, no. 1–3, pp. 293–309, 2002.

[3] J. B. Schodorf and D. B. Williams, "A constrained optimization approach to multiuser detection," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 258–262, Jan. 1997.

[4] ——, "Array processing techniques for multiuser detection," *IEEE Trans. Commun.*, vol. 45, no. 11, pp. 1375–1378, Nov. 1997.

[5] S. Verdú, *Multiuser Detection*. New York: Cambridge Univ. Press, 1998.

[6] X. Wang and H. V. Poor, "Blind equalization and multiuser detection in dispersive CDMA channels," *IEEE Trans. Commun.*, vol. 46, no. 1, pp. 91–103, Jan. 1998.

[7] Z. Pi and U. Mitra, "On blind timing acquisition and channel estimation for wideband multiuser DS-CDMA systems," *J. VLSI Signal Process., Special Issue on Multiuser Detection*, vol. 30, pp. 127–142, 2002.

[8] X. G. Doukopoulos and G. V. Moustakides, "Blind channel estimation for downlink CDMA systems," in *Proc. IEEE ICC*, Anchorage, AK, May 2003.

[9] Z. Xu, P. Liu, and X. Wang, "Blind multiuser detection: From MOE to subspace methods," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 510–524, Feb. 2004.

[10] Z. Xu and M. K. Tsatsanis, "Blind adaptive algorithms for minimum variance CDMA receivers," *IEEE Trans. Commun.*, vol. 49, no. 1, pp. 180–194, Jan. 2001.

[11] Z. Xu and P. Liu, "Kyrtosis based maximization/minimization approach to blind equalization for DS/CDMA systems in unknown multipath," in *Proc. IEEE ICASSP*, vol. III, Orlando, FL, 2002, pp. 2585–2588.

[12] M. K. Tsatsanis and Z. Xu, "Performance analysis of minimum variance CDMA receivers," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 3014–3022, Nov. 1998.

[13] X. G. Doukopoulos and G. V. Moustakides, "Blind Channel Estimation for DS-CDMA," INRIA, Rennes, France, Rep. 4781, Mar. 2003.

[14] P. Loubaton and E. Moulines, "On blind multiuser forward link channel estimation by the subspace method: identifiability results," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2366–2376, Aug. 2000.

[15] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Second ed. Baltimore, MD: Johns Hopkins Univ. Press, 1990.

[16] G. V. Moustakides, "Exponential convergence of products of random matrices, application to the study of adaptive algorithms," *International J. Adaptive Control Signal Process.*, vol. 2, no. 12, pp. 579–597, Dec. 1998.

[17] J. Y. Huang, I. Kontoyiannis, and S. P. Meyn, "The ODE method and spectral theory of Markov operators," *Lect. Notes Control Inf.*, vol. 280, pp. 205–221, 2002.

[18] J. G. Proakis, *Digital Communications*, Fourth ed. New York: McGraw-Hill, 2001.

[19] J. K. Cavers, *Mobile Channel Characteristics*. Boston, MA: Kluwer, 2000.

Xenofon G. Doukopoulos (S'03) was born in Grevena, Greece in 1978. He received the diploma in computer engineering and the M.Sc degree in signal and image processing systems, both from the Computer Engineering and Informatics Department (CEID) University of Patras, Patras, Greece, in 2000 and 2003, respectively.

Since November 2001, he has been pursuing Ph.D degree with the Institut National de Recherche en Informatique et Automatique (INRIA), Rennes, France. His research interests include channel estimation and multiuser detection for wireless telecommunication systems, signal processing for telecommunication applications, and adaptive subspace tracking techniques.

George V. Moustakides (SM'97) was born in Drama, Greece, in 1955. He received the diploma in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1979, the M.Sc. degree in systems engineering from the Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, in 1980, and the Ph.D. degree in electrical engineering and computer science from Princeton University, Princeton NJ, in 1983.

From 1983 to 1986, he was with Institut National de Recherche en Informatique et Automatique (INRIA), Rennes, France, and from 1987 to 1990, he held a research position at the Computer Technology Institute of Patras, Patras, Greece. In 1991, he joined the Computer Engineering and Informatics Department, University of Patras, as an Associate Professor, where in 1996, he became a Professor. From 2001 to 2004, he held a Senior Researcher position at INRIA. Since 2001, he has been a Professor with the Department of Computer and Communication Engineering, University of Thessaly, Volos, Greece. His interests include multiuser/multicarrier communications, adaptive signal processing, and sequential detection of changes.