Adaptive Power Techniques for Blind Channel Estimation in CDMA Systems

Xenofon Doukopoulos, George Moustakides

INRIA

Adaptive Power Techniques for Blind Channel Estimation in CDMA Systems

Xenofon G. Doukopoulos

and

George V. Moustakides

N° 4963

THÈME 1

Rapport de recherche

Adaptive Power Techniques for Blind Channel Estimation in CDMA Systems

Xenofon G. Doukopoulos
and
George V. Moustakides

Abstract: The problem of blind adaptive channel estimation in code-division multiple access systems is considered. Motivated by the iterative power method, used in Numerical Analysis for estimating singular values and singular vectors, we develop RLS and 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 to the already existing techniques in the literature.

Key-words: Channel estimation, CDMA, Blind adaptive algorithms, RLS, LMS.

Méthodes adaptatives de la puissance pour l'estimation aveugle de canal dans des systèmes CDMA

Résumé : Nous traitons le problème de l'estimation aveugle adaptatif de canal pour le système CDMA (Code Division Multiple Access). Motivé par la méthode itérative de puissance, utilisée dans l'analyse numérique pour estimer des valeurs singulières et des vecteurs singuliers, nous développons des algorithmes adaptatifs basés sur des sous-espaces RLS et LMS afin d'identifier la réponse impulsionelle d'un canal à trajets multiples. Les schémas proposés dans cet article utilisent seulement la signature de l'utilisateur d'intérêt et les données reçues, et sont donc aveugles. Les deux versions (RLS et LMS) montrent une convergence rapide, ainsi qu'une faible complexité de calcul. À l'aide de simulations nous démontrons que l'exécution de nos méthodes est améliorée par rapport aux techniques déjà existantes dans la littérature.

Mots-clés : Estimation de canal, CDMA, Algorithmes adaptatifs aveugles, RLS, LMS.

Contents

1 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 3-rd 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 off-line as well as adaptive detection schemes have been proposed and extensively analyzed in the literature [4, 7, 11, 12, 14]. 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 non-dispersive case, only 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 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 selfsufficiency with respect to training, blind estimation methods tend to be, nowadays, the most frequent candidates for channel estimation. The blind channel estimation literature for CDMA is rather limited. In [15, 9] the (practically off-line) methods that are proposed involve a large 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 on-line processing. In [1] we proposed an alternative *off-line* scheme that replaces the SVD with a simple matrix power. This resulted in a substantial computational gain as compared to the previous two methods, without any significant performance loss. Another advantage of the method in [1] is the fact that it does not require knowledge (or estimates) of the signal subspace rank, in contrast to [15, 9] where such information is indispensable. A similar approach was presented independently in [17] emphasizing its close connection to the subspace method of [15].

Blind adaptive channel estimation techniques were proposed in [18, 16, 17]. Specifically in [18], based on the analytic results offered in [13], several RLS and LMS type versions were developed, that will serve for comparison against our schemes in the simulations section. A variant of the work in [18], reported in [16], consists in using higher order cumulants. This approach however suffers from slow convergence even for small

codes, while 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 [17] an RLS algorithm is proposed for the adaptive implementation of the power method suggested in [1, 17]. We should also mention that in this last work the channel estimates are obtained through an SVD of the size of the channel impulse response length.

In this work we are going to extend the power method proposed in [1] 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 [17], there is a significant difference. We will be able to completely eliminate the SVD, of the size of the channel, needed in [17], 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 to the corresponding version of [18], 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, while 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.

2   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 [14].

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 towards 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 its initial delay that can take any value in the set $\{0, \ldots, N-1\}$. When $z_i(n)$ propagates through a multipath AWN channel with impulse response $\mathbf{f}_i = [f_i(0) \ \cdots \ f_i(L-1)]^t$, then the received signal $y(n)$ can be written as

$$\begin{aligned} y(n) &= \sum_{i=1}^{K} z_i(n) \star f_i(n) + \sigma w(n) \\ &= \sum_{i=1}^{K} \sum_{k=-\infty}^{\infty} a_i \tilde{s}_i(n - kN - \tau_i)b_i(k) + \sigma w(n), \end{aligned} \tag{2}$$

where $\star$ denotes convolution; $\tilde{s}_i(n) = s_i(n) \star f_i(n)$ is the convolution between the sequence $s_i(n)$ and the channel impulse response $\mathbf{f}_i$ (i.e. the composite signature of User-i zero-padded from both ends); and $w(n)$ is a unit variance i.i.d. noise sequence with σ^2 denoting the power of the AWN.

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 $\mathbf{f}_1 = \cdots = \mathbf{f}_K = \mathbf{f}$ (since all users propagate through the *same* multipath channel) and $\tau_1 = \cdots = \tau_K = \tau$ (since the users 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 article, 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 [2]. 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

$$
\begin{aligned}
\mathbf{r}(n) \;=\; & \sum_{l=1}^{m}
\left[
\begin{array}{c}
\mathbf{0}_{(l-1)N \times L} \\
\tilde{\mathbf{s}}_1 \\
\mathbf{0}_{(m-l)N \times L}
\end{array}
\right] a_1 b_1(n-l+1) \\
& + \sum_{i=2}^{K} \sum_{l=1}^{m}
\left[
\begin{array}{c}
\mathbf{0}_{(l-1)N \times L} \\
\tilde{\mathbf{s}}_i \\
\mathbf{0}_{(m-l)N \times L}
\end{array}
\right] a_i b_i(n-l+1) + \mathrm{ISI} + \sigma \mathbf{w}(n),
\end{aligned}
\tag{4}
$$

where $\tilde{\mathbf{s}}_i = [\tilde{s}_i(0) \cdots \tilde{s}_i(N+L-2)]^t$ is the composite signature of User-i. We observe in (4) that the sum of the first m terms involves the entire composite signature of the user of interest; then follows the multi-access interference (MAI) part that contains terms similar to the first sum but coming from interfering users; then follows the part that includes the inter-symbol interference (ISI) of all users and 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; and $b_i(n)$ 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

$$
\mathbf{S}_1 =
\left[
\begin{array}{cccc}
s_1(0) & 0 & \cdots & 0 \\
\vdots & s_1(0) & \ddots & \vdots \\
s_1(N-1) & \vdots & \ddots & 0 \\
0 & s_1(N-1) & \ddots & s_1(0) \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & s_1(N-1)
\end{array}
\right].
\tag{6}
$$

3 Two Subspace Problems

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

autocorrelation matrix satisfies

$$\mathbf{R} \stackrel{\triangle}{=} \mathbb{E}\{\mathbf{r}(n)\mathbf{r}^t(n)\} = \mathbf{Q} + \sigma^2\mathbf{I}, \tag{7}$$

where

$$\mathbf{Q} = \sum_l \mathbf{d}_l\mathbf{d}_l^t \tag{8}$$

is a symmetric, nonnegative 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

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

where $\mathbf{U}_s, \mathbf{U}_w$ 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

$$\mathbf{U}_w^t\mathbf{d}_l = 0. \tag{10}$$

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

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

These vectors, since they lie in the signal subspace, they satisfy the orthogonality relation (10) and the same property is true for their sum. The latter fact can be expressed with the following equation

$$\mathbf{U}_w^t\mathbf{F}_1\mathbf{f} = 0, \tag{12}$$

where

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

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

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

Equ. (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

$$\mathbf{W} = \mathbf{F}_1^t \mathbf{U}_w \mathbf{U}_w^t \mathbf{F}_1. \tag{15}$$

In [15] both problems are solved by direct SVD, while in [9] the first with SVD and the second with QR. 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 on-line processing. We should also mention the need of these approaches in *knowing the noise subspace rank*. It turns out [1] that even slight errors in the estimate of this parameter, can produce significant performance degradation in the schemes proposed in [15, 9]. In [1, 17] a power method was proposed to replace the large SVD of Problem 1 used in [15, 9]. 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 subject let us first briefly discuss the consistency of the estimates provided by (14).

3.1 Consistency

Let r_s, 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 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
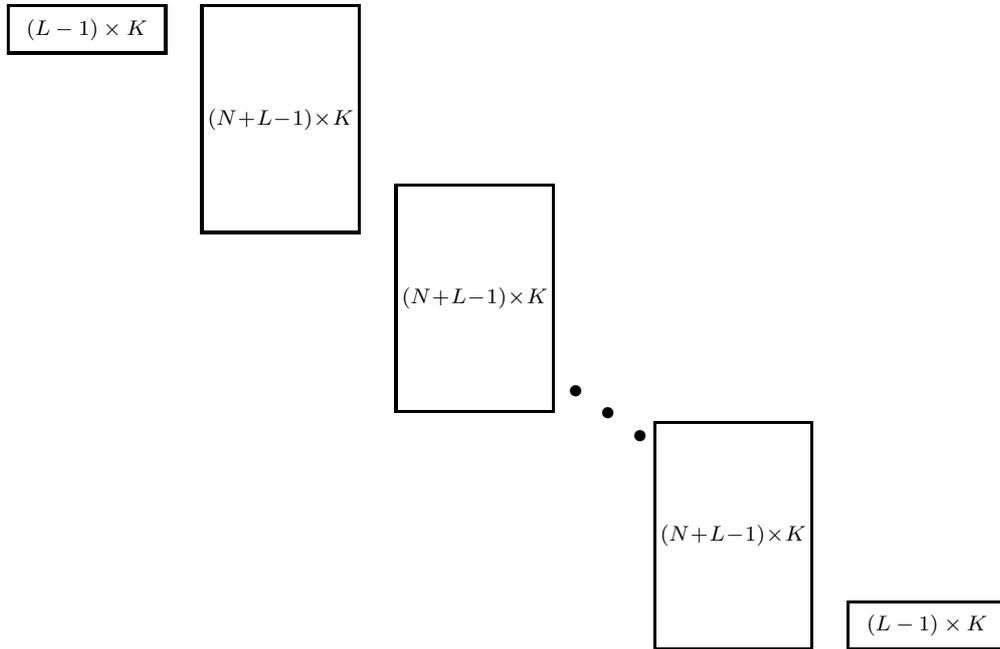

Figure 1: Representation of the vectors composing the signal subspace.

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 n-th user-bits whereas the last 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 like to follow the same analysis for the uplink scenario then, due to lack of synchronization, relation (17) becomes $r_s \leq (m + 2)K$, yielding

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

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 are thus 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 there appeared examples 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 [6], 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, that is, $L - L' < N$. 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' 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. We are now ready to proceed with the presentation of our blind adaptive schemes.

4   Power Method Variants

The power method [3] 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, 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 $k \to \infty$

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

$$\to [\mathbf{U}_s \quad \mathbf{U}_w] \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} [\mathbf{U}_s \quad \mathbf{U}_w]^t = \mathbf{U}_w \mathbf{U}_w^t, \tag{22}$$

with the last limit being true since $\mathbf{I} + \frac{\mathbf{\Lambda}_s}{\rho + \sigma^2}$ 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 Equ. (21) is slightly more general (due to the existence of the parameter ρ) than the one used in [1, 17]; 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.

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 apply again 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 $\mathbf{W}\mathbf{f} = 0$ 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 nonnegative definite with all singular values smaller than unity). Using singular value decomposition, 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

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

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, like direct SVD, requires $O(L^3)$ operations while the proposed scheme $O(L^2)$. As we will see 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 in this case we can efficiently control the forgetting of past data through the forgetting factor of the RLS algorithm.

Remark 2: A subtle and very important remark regarding Lemma 1 concerns 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 become just 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 solution to this problem. In [1] it was observed that, for off-line 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

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

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 the subspace channel estimation methods [1].

5 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

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

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, adaptively, estimates $\hat{\mathbf{W}}_k(n)$ of the matrix $\hat{\mathbf{W}}_k$ defined in (28). There are different possibilities that we exploit next. Notice however that with the help of any such estimate $\hat{\mathbf{W}}_k(n)$ the power method presented in Lemma 2 (which provides the final channel impulse response estimates) can be modified as follows

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

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

5.1 Channel Estimation via RLS

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

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

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

$$
\begin{align}
\mathbf{k}(n) &= \mathbf{P}(n-1)\mathbf{r}(n) \tag{31} \\
\gamma(n) &= 1/(\lambda + \mathbf{r}^t(n)\mathbf{k}(n)) \tag{32} \\
\mathbf{P}(n) &= \frac{1}{\lambda}\left(\mathbf{P}(n-1) - (\gamma(n)\mathbf{k}(n))\mathbf{k}^t(n)\right), \tag{33}
\end{align}
$$

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 $\hat{\mathbf{W}}_k(n)$ as

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

then this part requires $2kL(mN + L - 1)^2 + 2L^2(mN + L - 1) + O(1)$ operations. Finally, once $\hat{\mathbf{W}}_k(n)$ 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 $\hat{\mathbf{W}}_k(n)$ in (34).

We should mention that our RLS version is similar to the adaptive algorithm presented in [17] and when $k = 1$ to the RLS version of [18]. The advantage here is that we avoid the SVD on $\hat{\mathbf{W}}_k(n)$ proposed in [17, 18], since we replace it with *one iteration* of the power recursion (29). As far as the RLS version of [18] is concerned we are going to see in our simulations that higher values of the power k can ameliorate performance significantly.

5.2   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 [18].

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

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

where

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

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

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

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

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

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

$$\hat{\mathbf{V}}_l(n) = \lambda \hat{\mathbf{V}}_l(n-1) + \mu \left(\hat{\mathbf{V}}_{l-1}(n-1) - \mathbf{r}(n)\mathbf{r}^t(n)\hat{\mathbf{V}}_l(n-1) \right), \quad l = 1, 2, \ldots, k, \tag{39}$$

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

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

$$\mathbb{E}[\hat{\mathbf{V}}_l(n)] = \hat{\mathbf{V}}_l + \sum_{j=0}^{l-1} \mu^j \binom{n}{j} (\lambda \mathbf{I} - \mu \mathbf{R})^{n-j} \left(\hat{\mathbf{V}}_{l-j}(0) - \hat{\mathbf{V}}_{l-j} \right), \quad l = 1, \ldots, k. \tag{40}$$

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 (that includes LMS and Leakage LMS) and rich classes of data models [8, 5].

From Theorem 1 we observe that, due to the term $(\lambda \mathbf{I} - \mu \mathbf{R})^{n-j}$ we have an exponential convergence of $\mathbb{E}[\hat{\mathbf{V}}_l(n)]$ towards the desired quantity $\hat{\mathbf{V}}_l = (\rho \mathbf{I} + \mathbf{R})^{-l} \mathbf{F}$. In fact the speed of convergence is governed by the largest eigenvalue of the matrix $\lambda \mathbf{I} - \mu \mathbf{R}$ which is equal to $\lambda - \mu \sigma^2$. Notice that by using the Leakage LMS recursion in (39) with $0 < \lambda < 1$, we assure an exponential convergence with a factor that is at least equal to λ, independently of the SNR level. If on the other hand one uses the regular LMS with $\lambda = 1$ the corresponding factor becomes $1 - \mu \sigma^2$ which can induce an extremely small convergence rate in medium to high SNR. Unfortunately here 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 $\hat{\mathbf{V}}_k$, using $\hat{\mathbf{V}}_k(n)$ we can now obtain estimates for $\hat{\mathbf{W}}_k$ following (35) as

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

Finally we apply one iteration of the power method in (29) to obtain the estimate $\hat{\mathbf{f}}(n)$ 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 $\hat{\mathbf{W}}_k(n) = \mathbf{F}_1^t \hat{\mathbf{V}}_k(n)$ 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 $\hat{\mathbf{W}}_k(n)$. It turns out that we can reduce this complexity by a factor which 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 $\hat{\mathbf{W}}_l(n)$, $l = 1, \dots, k$

$$\hat{\mathbf{W}}_l(n) = \lambda \hat{\mathbf{W}}_l(n-1) + \mu \left(\hat{\mathbf{W}}_{l-1}(n-1) - \mathbf{F}_1^t \mathbf{r}(n) \mathbf{r}^t(n) \hat{\mathbf{V}}_l(n-1) \right). \tag{42}$$

Since all vectors $\mathbf{r}^t(n) \hat{\mathbf{V}}_l(n-1)$ 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 $\mathbf{F}_1^t \mathbf{r}(n) \mathbf{r}^t(n) \hat{\mathbf{V}}_l(n-1)$ appearing in (42). The total complexity for computing $\hat{\mathbf{W}}_k(n)$ 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 nonegligible computational gain. The price we pay for using (42) is the need to store the matrices $\hat{\mathbf{W}}_l(n)$, $l = 1, \dots, k$. This is clearly unnecessary when we use (41).

6   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 [18]. 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 the signal power changes and obtain an algorithm which is relatively insensitive to them. We propose the following simple modification of (39)

$$\hat{\mathbf{V}}_l(n) = \lambda \hat{\mathbf{V}}_l(n-1) + \mu \left(\hat{\mathbf{V}}_{l-1}(n-1) - \frac{\mathbf{r}(n)\mathbf{r}^t(n)}{\beta(n)} \hat{\mathbf{V}}_l(n-1) \right), \ l = 1, 2, \dots, k; \tag{43}$$

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

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

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 10000 and 15000 the number of users. For the multipath channel we start with the length 3 "difficult" channel and at 5000 we switch to the length 10 "easy" channel of [10]. 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 10000, 10 additional users enter the channel, 5 of them having power equal to the user of interest and the remaining 5 being 10 db stronger. Finally, at bit 15000 the last 10 users along with 5 more exit the channel. As we can verify, the constraint in (19) is always satisfied.

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 [18] without needing an SVD on the matrix $\mathbf{W}(n)$ 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 10000 and 15000). It is clear that in this high SNR environment selecting $k = 1$ is sufficient.

We continue with the presentation of the LMS schemes. We apply the LMS adaptation in (43) with $\lambda = 0.998$, $\mu = 1$ and $\nu = 0.99$. We also apply the LMS algorithm of [18] 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 [18], 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 [18]) reaches almost 20 db at steady state.
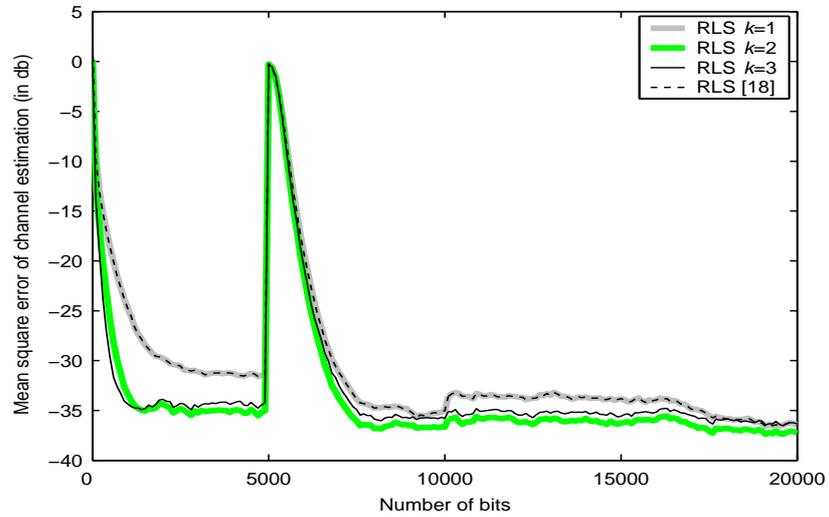
Figure 2: Performance of the proposed RLS channel estimation scheme and the RLS version of [18]; noise power 20 db.
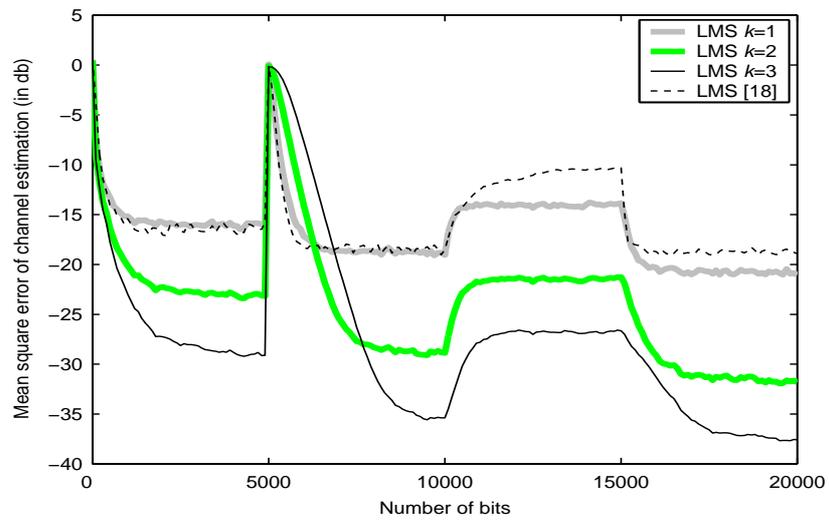

Figure 3: Performance of the proposed LMS channel estimation scheme and the LMS version of [18]; noise power 20 db.
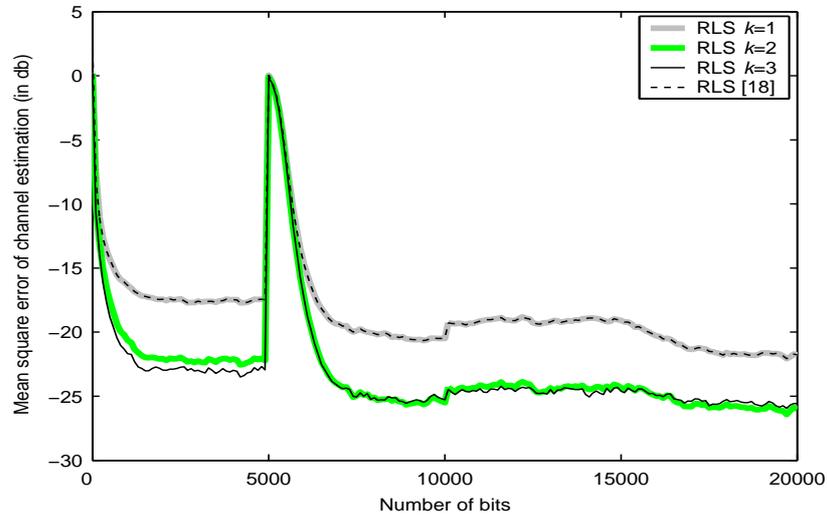
Figure 4: Performance of the proposed RLS channel estimation scheme and the RLS version of [18]; noise power 10 db.

Next we consider the same signaling scenario but with a significantly lower SNR. Specifically we set the desired user's SNR at 10 db. The performance of the RLS schemes is presented in Fig. 4. Again the method of [18] is identical to our RLS $k = 1$ version. Here however, in this low SNR environment, employing higher orders of k ameliorates the overall RLS performance significantly especially in the initial part, i.e. in the case of the "difficult" channel.

Finally in Fig. 5, the LMS schemes are presented for an SNR of 10 db. Again the LMS algorithm of [18] has similar performance to our LMS version with $k = 1$, exhibiting the same sensitivity to changes in the number of users as in the previous example. As before, our $k = 3$ version outperforms our $k = 1$ version, and consequently the LMS algorithm of [18], by more than 10 db.

A point that should be mentioned concerns the initial performance (up to bit 5000) of all channel estimation schemes. In fact, if we had exact knowledge of the filter length, that is, if we had used $L = 3$ instead of $L = 10$, both LMS and RLS versions would have attained better performance levels than the ones depicted in the corresponding figures. Finally, 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. A performance which, unfortunately, comes at a much greater computational cost.
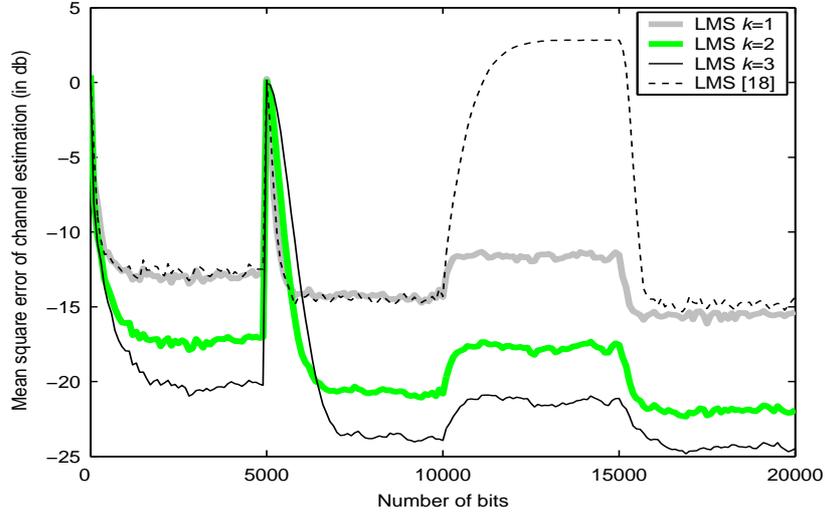
Figure 5: Performance of the proposed LMS channel estimation scheme and the LMS version of [18]; noise power 10 db.

6.1 Other Comparisons

Let us now compare the recursion in (43) for $0 < \lambda < 1$ (Leakage LMS) with the case $\lambda = 1$ (normal LMS). The signaling scenario depicted in Fig. 6 is similar to our first example. We start with the "difficult" channel and at bit 25000 we switch to the "easy" channel. 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 $\lambda = 0.998$; $\nu = 0.99$ and SNR level equal to 20 db. The difference in convergence speed between the two versions is noticeable, accompanied by an equivalent difference (in the opposite direction) in steady state behavior. As was mentioned previously, it is *not* possible to trade between steady state performance and convergence speed in the normal ($\lambda = 1$) LMS algorithm by simply increasing the step size μ. In fact the value $\mu = 1$, used here, is rather limiting since even the slightest increase in this parameter leads LMS to instability. Consequently the performance of LMS depicted in Fig. 6 is the best this algorithm can offer, as far as convergence speed is concerned. It is through Leakage LMS 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. 7 we depict the channel estimates of the RLS scheme with the power parameter set to $k = 3$. The dashed line corresponds to channel estimates obtained by
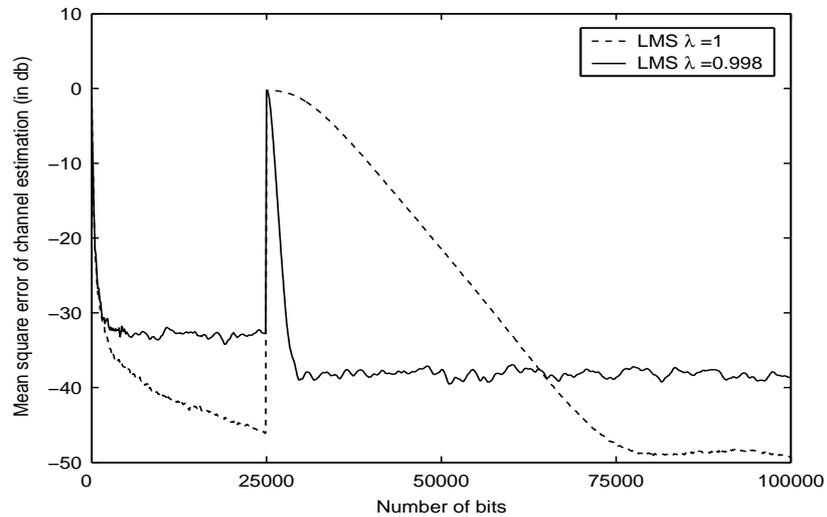
Figure 6: Convergence characteristics of Leakage LMS (solid) and normal LMS (dashed) after an abrupt change in the channel, SNR=20 db.
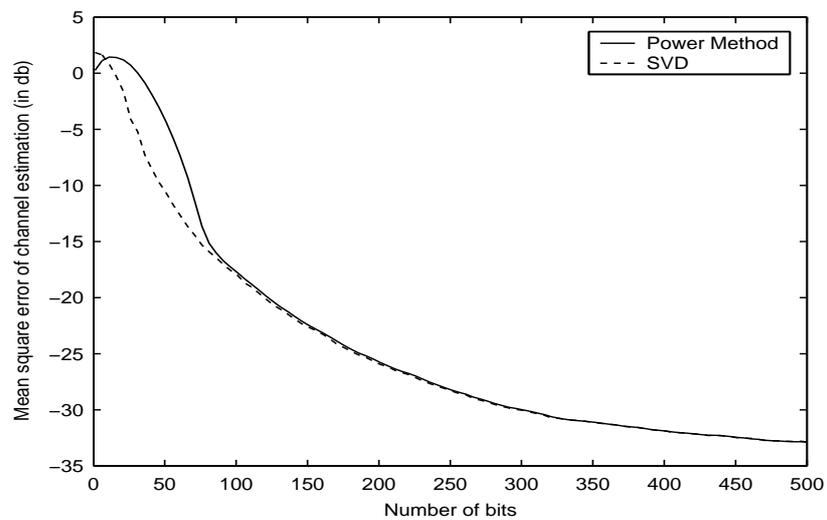

Figure 7: Performance of the proposed RLS scheme when the channel is estimated by applying one step of the power method (solid) and by SVD (dashed).

applying an SVD at every time step on the matrix $\hat{\mathbf{W}}_k(n)$ as suggested in [18, 17], whereas the solid to our method, that is, by applying the iteration in (29) only once. We can see that the two 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 $O(L^3)$).

7 Conclusion

In this work we examined the blind adaptive channel estimation problem for CDMA in multipath additive white noise channels and considered a similar to [15, 9] two-step methodology for its solution. The novelty of our approach consists in 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 simulation 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 to the adaptive methods proposed in [18] our schemes offer substantial performance gains at similar computational cost. Finally the adaptive scheme of [17], although it is similar to our RLS version, it 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 $\hat{\mathbf{V}}_0(n) = \mathbf{F}_1$, we obtain the following recursion

$$\mathbb{E}[\hat{\mathbf{V}}_1(n)] = (\lambda\mathbf{I} - \mu\mathbf{R})\mathbb{E}[\hat{\mathbf{V}}_1(n-1)] + \mu\mathbf{F}_1 \tag{45}$$

which readily leads to

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

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

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

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

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

$$\mathbb{E}[\hat{\mathbf{V}}_\kappa(n)] = \hat{\mathbf{V}}_\kappa + \sum_{j=0}^{\kappa-1} \mu^j \binom{n}{j} (\lambda\mathbf{I} - \mu\mathbf{R})^{n-j} \, \mathbf{\Delta}\hat{\mathbf{V}}_{\kappa-j}, \tag{48}$$

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

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

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 $\hat{\mathbf{V}}_{\kappa+1}(n-1)$ is independent from the received data $\mathbf{r}(n)$, we obtain the recursion

$$\mathbb{E}[\hat{\mathbf{V}}_{\kappa+1}(n)] = (\lambda\mathbf{I} - \mu\mathbf{R})\mathbb{E}[\hat{\mathbf{V}}_{\kappa+1}(n-1)] + \mu\mathbb{E}[\hat{\mathbf{V}}_\kappa(n-1)] \tag{50}$$

which yields

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

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

$$\mathbb{E}[\hat{\mathbf{V}}_{\kappa+1}(n)] = \mu\sum_{i=0}^{n-1}(\lambda\mathbf{I} - \mu\mathbf{R})^i \hat{\mathbf{V}}_\kappa + (\lambda\mathbf{I} - \mu\mathbf{R})^n \hat{\mathbf{V}}_{\kappa+1}(0) \tag{52}$$

$$+ \sum_{i=0}^{n-1}\sum_{j=0}^{\kappa-1} \mu^{j+1} \binom{n-i-1}{j} (\lambda\mathbf{I} - \mu\mathbf{R})^{n-j-1} \, \mathbf{\Delta}\hat{\mathbf{V}}_{\kappa-j}. \tag{53}$$

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

$$\hat{\mathbf{V}}_{\kappa+1} + (\lambda\mathbf{I} - \mu\mathbf{R})^n \, \mathbf{\Delta}\hat{\mathbf{V}}_{\kappa+1}. \tag{54}$$

Changing the order of summation in (53), 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 (53) yields

$$\sum_{p=1}^{\kappa} \mu^p \binom{n}{p} (\lambda\mathbf{I} - \mu\mathbf{R})^{n-p} \, \mathbf{\Delta}\hat{\mathbf{V}}_{\kappa+1-p}. \tag{55}$$

By adding this to (54) we obtain the desired result.

References

[1] X.G. Doukopoulos and G.V. Moustakides, "Blind channel estimation for downlink CDMA systems," IEEE ICC2003, Alaska, Anchorage, May 2003.

[2] X.G. Doukopoulos and G.V. Moustakides, "Blind channel estimation for DS-CDMA," *INRIA Report*, no. 4781, March 2003.

[3] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 2nd edn, The John Hopkins University Press, 1990.

[4] M. Honig, U. Madhow and S. Verdú, "Blind adaptive multiuser detection," *IEEE Trans. on Information Theory,* vol. 41, pp. 944-960, July 1995.

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

[6] P. Loubaton and E. Moulines, "On blind multiuser forward link channel estimation by the subspace method: Identifiability results," *IEEE Trans. on Signal Processing*, vol. 48, pp. 2366-2376, Aug. 2000.

[7] G.V. Moustakides, "Constrained adaptive linear multiuser detection schemes," invited paper, special issue on multiuser detection, *Journal of VLSI Signal Processing*, vol. 30, no 1-3, pp. 293-309, 2002.

[8] G.V. Moustakides, "Exponential convergence of products of random matrices, application to the study of adaptive algorithms," *International Journal of Adaptive Control and Signal Processing*, vol. 2, no. 12, pp. 579-597, Dec. 1998

[9] Z. Pi and U. Mitra, "On blind timing acquisition and channel estimation for wideband multiuser DS-CDMA systems," special issue on multiuser detection, *Journal of VLSI Signal Processing*, vol. 30, pp. 127-142, 2002.

[10] J.G. Proakis, *Digital Communications*, 4-th Edition, McGraw-Hill, New York, 2001.

[11] J.B. Schodorf and D.B. Williams, "A constrained optimization approach to multiuser detection," *IEEE Trans. on Signal Processing*, vol. 45, pp. 258-262, Jan. 1997.

[12] J.B. Schodorf and D.B. Williams, "Array processing techniques for multiuser detection," *IEEE Trans. on Communications*, vol. 45, pp. 1375-1378, Nov. 1997.

[13] M.K. Tsatsanis and Z. Xu. "Performance analysis of minimum variance CDMA receivers," *IEEE Trans. on Signal Processing*, vol. 46, pp. 3014-3022, Nov. 1998.

[14] S. Verdú, *Multiuser Detection*, Cambridge University Press, New York, 1998.

[15] X. Wang and H.V. Poor, "Blind equalization and multiuser detection in dispersive CDMA channels," *IEEE Trans. on Communications*, vol. 46, pp. 91-103, Jan 1998.

[16] Z. Xu and P. Liu, "Kyrtosis based maximization/minimization approach to blind equalization for DS/CDMA systems in unknown multipath," IEEE ICASSP2002, Orlando, vol. III, pp. 2585-2588, 2002.

[17] Z. Xu, P. Liu and X. Wang, "Blind multiuser detection: From MOE to subspace methods," *IEEE Trans. on Signal Processing*, accepted for publication, to appear.

[18] Z. Xu and M.K. Tsatsanis, "Blind adaptive algorithms for minimum variance CDMA receivers," *IEEE Trans. on Communications*, vol. 49, pp. 180-194, Jan. 2001.