Decision-feedback interference suppression in CDMA systems: a ML-based semiblind approach

Mónica F. Bugallo\textsuperscript{a}, Joaquín Míguez\textsuperscript{b}, Luis Castedo\textsuperscript{b}

\textsuperscript{a}Department of Electrical and Computer Engineering, Stony Brook University, Stony Brook, NY 11794, USA
\textsuperscript{b}Departamento de Electrónica e Sistemas, Universidade da Coruña, Facultade de Informática, Campus de Elviña s/n, 15071 A Coruña, Spain

Received 11 April 2001; received in revised form 10 February 2003

Abstract

This paper addresses the problem of interference suppression in direct sequence code division multiple access systems. We propose a novel semiblind decision feedback (DF) receiver based on the maximum likelihood principle that simultaneously exploits the transmission of training sequences and the statistical information of the unknown transmitted symbols. Both iterative and adaptive implementations of the proposed receiver, derived within the framework of the expectation maximization algorithm, are presented. Computer simulations show that the resulting multiuser detectors attain practically the same performance as the theoretical DF minimum mean square error receiver.

\textcopyright{} 2003 Elsevier B.V. All rights reserved.

Keywords: Decision feedback; Maximum likelihood; Multiuser detection; Semiblind receivers; Interference suppression; CDMA

1. Introduction

Third generation wideband mobile communication systems rely on the direct sequence code division multiple access (DS CDMA) scheme because it provides an efficient and flexible use of the spectrum [5]. In practice, however, the capacity of CDMA systems is severely limited by the inter-symbol interference (ISI) and, specially, multiple access interference (MAI).

Optimum detection consists of the maximum a posteriori (MAP) estimation of the desired user data, which reduces to maximum likelihood (ML) sequence estimation when the transmitted symbols have equal probabilities [23]. Implementations of the ML detector proposed in the literature [23], however, are limited by the need of knowing the channel characteristics and their prohibitive computational complexity. Therefore, low-complexity alternative approaches based on linear filtering have been investigated. Conventional linear minimum mean square error (MMSE) receivers [23] implicitly estimate the channel parameters using a training sequence. Therefore, the longer this training sequence is the better the receiver performance is. However, in burst transmission systems each block of received data consists of a training part and a sequence of unknown symbols. In this context, it is desirable to minimize the length of the training part in order to use the channel efficiently and, for this reason, several semiblind receivers have been recently proposed [1,11–13,15]. The semiblind approach consists of exploiting the whole burst of data, including both the known and unknown symbols, in order
to select the receiver coefficients. This includes both conventional decision-directed methods [15], statistical approaches that rely on the independence of the transmitted symbols [11,12] and subspace techniques [1,13].

In this paper, we extend the ML-based criterion for MAI and ISI suppression, originally introduced in [2], to the case of a decision feedback (DF) receiver structure. The utilization of DF receiver structures has already been proposed in the context of multiuser detection [19,22] as an attractive solution that exhibits superior performance with respect to (w.r.t.) linear receivers at the cost of a small computational increase. The generalized expectation-maximization (EM) [14] approach is followed to derive both iterative and adaptive algorithms that compute the DF receiver parameters according to the ML principle. The resulting detectors are semiblind because they exploit both the transmission of training sequences and the known statistical features of the transmitted symbols and the noise in the channel. As a consequence, the proposed scheme allows to achieve a very advantageous trade-off between training sequence length and receiver performance when compared to conventional approaches.

It is important to remark that the proposed semiblind DF-EM receivers substantially differ from existing DF-ML sequence detectors (DF-MLSD) [8] because the latter are based on the Viterbi algorithm whereas the former simply consist of two linear filters (one forward filter and one backward filter) with an intercalated threshold detector. Thus, our approach is similar in complexity to conventional linear multiuser receivers and considerably less computationally demanding than DF-MLSD. Furthermore, computer simulations reveal that DF-EM receivers widely improve the theoretical performance limit of linear multiuser receivers due to the nonlinearity introduced by the threshold detector.

The remaining of the paper is organized as follows. Section 2 describes the signal model of an asynchronous time-dispersive DS CDMA system. Section 3 introduces the DF semiblind multiuser receiver based on the ML criterion. In Section 4 we
develop EM-based iterative and adaptive algorithms that solve the ML optimization problem. Section 5 presents simulation results and Section 6 is devoted to the conclusions.

2. Signal model

Fig. 1 shows the baseband discrete-time equivalent model of an asynchronous Direct Sequence (DS) CDMA system with time dispersive channels. When the \( i \)th user transmits an isolated symbol, \( b_i \), it is multiplied by a unique binary-valued spreading sequence with \( L \) chips per symbol, \( c_i(k), k=0,\ldots,L-1 \). The resulting signal passes through a linear time-dispersive channel, \( h_i(k), k=0,\ldots,P-1 \), that accounts not only for the channel response but also for the relative time delays of the different users and the transmitter and receiver front-end filters. The received sequence is the superposition of the transmitted signals from the \( N \) users plus the additive white gaussian noise (AWGN) sequence \( g(k) \), i.e.,

\[
x(k) = \sum_{i=1}^{N} d_i(k) b_i + g(k), \quad k = 0,\ldots,L + P - 2,
\]

where \( d_i(k) = c_i(k) \ast h_i(k) = \sum_{p=0}^{L-1} h_i(p) c_i(k-p) \), has length \( L + P - 1 \) and will be termed received code. Due to the channel time dispersive effect, when a symbol, \( b_i(n) \), is transmitted, there is ISI from \( b_i(n-1), b_i(n-2),\ldots,b_i(n-m+1) \), where \( m = \lfloor (L + P - 1)/L \rfloor \) is the channel memory size. The received sequence for the \( n \)th symbol period is

\[
x(n,k) = x(nL + k)
\]

\[
= \sum_{i=1}^{N} \sum_{r=0}^{m-1} d_i(r,k) b_i(n-r) + g(n,k),
\]

\( k = 0,\ldots,L - 1, \) (2)

where \( d_i(r,k) = d_i(rL + k) \). Bringing together all the observations where the symbol of interest, \( b_i(n) \), is involved, we obtain an expression for the overall received sequence using vector notation as

\[
x(n) = \text{Db}(n) + g(n), \quad (3)
\]

where \( x(n) = [x(n,0),\ldots,x(n,L-1),\ldots,x(n+m-1,0),\ldots,x(n+m-1,L-1)]^T \) is the \( Lm \times 1 \) observation vector (\( \cdot^T \) denotes transposition),

\[
D = \begin{bmatrix}
D_{m-1} & D_{m-2} & \cdots & D_0 & 0 & \cdots & 0 \\
0 & D_{m-1} & D_{m-2} & \cdots & D_0 & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \cdots & D_{m-1} & D_{m-2} & \cdots & D_0 
\end{bmatrix}
\]

is the \( Lm \times N(2m-1) \) received code matrix, composed of the submatrices \( D_r = [d_1(r),\ldots,d_N(r)]^T \), and \( d_i(r) = [d_i(r,0),\ldots,d_i(r,L-1)]^T \). The transmitted symbol vector is \( \text{b}(n) = [b_1(n-m+1),\ldots,b_N(n-m+1),\ldots,b_1(n+m-1),\ldots,b_N(n+m-1)]^T \) and \( \text{g}(n) = [g(n,0),\ldots,g(n+m-1,L-1)]^T \) is a vector of independent and identically distributed (i.i.d.) Gaussian random variables with zero mean and variance \( \sigma_g^2 \) (the power spectral density of the AWGN).

The DF multiuser receiver consists of a linear finite impulse response (FIR) forward filter, \( w = [w_1,\ldots,w_{Lm}]^T \), and a backward filter, \( v = [v_1,\ldots,v_{m-1}]^T \). Therefore, the \( n \)th symbol soft estimate is

\[
y(n) = w^H x(n) + v^H \hat{b}_1(n), \quad (4)
\]

where the superindex \( ^H \) denotes Hermitian transposition and \( \hat{b}_1(n) = [\hat{b}_1(n-1),\ldots,\hat{b}_1(n-m+1)]^T \) is an \( (m-1) \times 1 \) vector that corresponds to an estimate of the causal ISI. The symbols in \( b_1(n) \) are easily obtained from the available soft estimates, \( y(n-1),\ldots,y(n-m+1) \) using a threshold detector, as indicated in Fig. 1.

3. Selection of the receiver coefficients

Let us assume that both the MAI and the ISI are totally suppressed by the receiver. Then, the soft symbol estimate, \( y(n) \), consists of just two components: the desired user symbol and an additive Gaussian noise term. Denoting \( w_* \) and \( v_* \), as the filters that completely eliminate both MAI and ISI, we can write

\[
y(n) = w_*^H x(n) + v_*^H \hat{b}_1(n) = b_1(n) + g_f(n) \quad (5)
\]

where \( b_1(n) \) is the desired user symbol and \( g_f(n) \) is a complex Gaussian random variable with zero mean and variance \( \sigma_f^2 \). In this section, the filtered noise
variance, $\sigma_f^2$, will be considered a constant for the sake of simplicity. In Section 4, however, we will also propose an easy-to-implement updating rule for $\sigma_f^2$.

When the channel does not introduce ISI ($m = 1$), the estimates obtained with the optimum filter, $w_*$, are i.i.d. random variables according to assumption (5), because the filtered noise process, $\{g_f(n)\}_{n=0}^{\infty}$, is white. In this case, when a block of $K$ observation vectors is available, the joint p.d.f. of the resulting data frame $y = [y(0), \ldots, y(K-1)]^T$ is given by the following expression (see Appendix A.1),

$$f_{y,\Theta}(y) = \left(\frac{1}{\pi \sigma_f^2}\right)^K \prod_{n=0}^{K-1} E_b \left[ e^{-\frac{|y(n)-b|^2}{2\sigma_f^2}} \right],$$  \(6\)

where $\Theta = [w_*, v_*]$ is the parameter set of the p.d.f. and $E_b[\cdot]$ denotes statistical expectation w.r.t. the desired user symbol. Notice that $E_b[\cdot]$ can be analytically calculated because it reduces to a simple summation.

When the channel introduces ISI ($m > 1$), process $\{g_f(n)\}_{n=0}^{\infty}$ is not, in general, white. However, although there exists statistical dependence among the estimates $y(0), \ldots, y(K-1)$, it is a common practice in statistical signal processing (see examples in [7]) to approximate the joint p.d.f. by the product of the marginal densities when the latter is difficult (or impossible) to derive. Moreover, in our case the statistical dependence between estimates is only due to the noise term, $g_f(n)$, and therefore, the model p.d.f. given by (6) converges to the true density of $y$ when the optimal filters, $w_*$ and $v_*$, are used and the SNR grows. Hence, we adopt (6) in the sequel as an approximate model p.d.f. for the vector of symbol estimates, $y$.

Assuming (6), the p.d.f. of $y$ clearly depends on the parameter vector $\Theta$, that contains the desired forward and backward filter coefficients, $w_*$ and $v_*$. Therefore, following [2], we can apply the classical ML approach to estimate them as

$$\hat{\Theta} = \arg \max_{\Theta} \left\{ \mathcal{L}(\Theta) = \sum_{n=0}^{K-1} \log E_b \left[ e^{-\frac{|y(n)-b|^2}{2\sigma_f^2}} \right] \right\}$$  \(7\)

where $\hat{\Theta} = [\hat{w}, \hat{v}]$, and $\mathcal{L}(\Theta)$ is the log-likelihood of $\Theta$ w.r.t. the block of soft estimates, $y$, that derives from model (6).

Unfortunately, the log-likelihood $\mathcal{L}(\Theta)$ is a non-quadratic function with several maxima. In particular, the solutions to problem (7) guarantee that the soft estimates, $y(n)$, have a p.d.f. close to $f_{b+g_f}(\cdot)$, but this is not enough, however, to ensure that the desired user is extracted. Since it is assumed that all users transmit symbols with the same modulation format, the p.d.f. of the $i$th interference at the receiver is also $f_{b+g_f}(\cdot)$, which does not differ from the desired user p.d.f. Therefore, solving the optimization problem (7) may lead to the capture of an interference.

This capture problem can be considerably alleviated if we exploit the transmission of a short training sequence of $M < K$ symbols, as it is done in currently standardized mobile communication systems. Indeed, let us assume that the first $M$ symbols
(i.e., \( \mathbf{b}_t = [b_1(0), \ldots, b_1(M - 1)]^T \)) are known a priori by the receiver. Conditioning the expectations in (7) w.r.t. the known symbols, \( \mathbf{b}_t \), we arrive at a semiblind receiver where the filter coefficients are computed as the solution to

\[
\hat{\Theta} = \arg \max_{\Theta} \{ \mathcal{L}(\Theta) | \mathbf{b}_t \} = \arg \min_{\Theta} \left\{ \sum_{n=0}^{M-1} |y(n) - b_1(n)|^2 - \sum_{n=M}^{K-1} \log E_b \left[ e^{-|y(n) - b_1(n)|^2/\sigma^2} \right] \right\}, \tag{8}
\]

The computer simulations in Section 5 show that relatively short training sequences (\( M \approx 30 \) symbols) may be enough to avoid the capture problem. This is because the first term in (8) is a purely quadratic form with a single minimum that corresponds to the rough extraction of the desired user. Note that \( \Theta \) is still computed according to the ML principle, i.e., all the available statistical information is employed to obtain the filter coefficients and, hence, the proposed semiblind DF receiver outperforms the conventional DF-MMSE multiuser detector (i.e., the straightforward extension to the multiuser case of the DF equalizer described in [17]) that only exploits the training sequence \( \mathbf{b}_t \).

3.1. Discussion

It is important to remark that, rigorously speaking, criterion (8) is inherently unfeasible because it relies on the hypothesis that the soft estimates \( y(n) \) have the desired p.d.f. in order to carry out ML estimation. It is apparent that this assumption does not hold in practice because both \( \mathbf{w}_s \) and \( \mathbf{v}_s \) are a priori unknown. One further theoretical objection concerns the existence and unicity of \( \mathbf{w}_s \) and \( \mathbf{v}_s \) that comply with (5): it can be easily found that, depending on the value of the system parameters \( L, N \) and \( P \), complete suppression of both MAI and ISI may be feasible or not and, if feasible, \( \mathbf{w}_s \) and \( \mathbf{v}_s \) are not necessarily unique.

With respect to the issue of applicability of the criterion, let us point out that we have stated Eqs. (5) and (6) as working assumptions that lead to an approximate statistical characterization of the signal where the powerful ML principle can be easily applied. From that point of view, validation of the method is carried out by means of computer simulations in Section 5, that show a remarkably good performance.

Regarding the existence of the solution \(^2\) the analysis presented in Appendix A.2 shows that the proposed multiuser receiver, which is obtained as the solution to problem (8), is closely related to the conventional DF-MMSE receiver, which is well defined for any choice of \( N, L \) and \( P \). In accordance with this result, the latter theoretical detector is used as a benchmark in Section 5 and numerical results illustrate the convergence of the proposed receiver.

4. Implementation

Since it is not possible to find a closed form solution to problem (8), some optimization method must be used in order to compute the parameter estimates \( \hat{\Theta} \). The EM algorithm [14] provides a very general framework for the development of both iterative and adaptive algorithms that solve (8).

4.1. Iterative space alternating generalized EM algorithm

The EM approach postulates the existence of some missing (unobserved) data that, if known, would aid in the estimation problem. Using the common EM terminology, let the observed soft estimates, \( y(n) \), \( n = 0, \ldots, K - 1 \), be the incomplete data set and let the extended vectors \( \mathbf{y}_e(n) = [y(n), b_1(n)]^T \), \( n = 0, \ldots, K - 1 \), be the complete data set (where \( b_1(n) \) are the missing data).

Following the same procedure as in [11, Chapter 3], it can be shown that the complete data sufficient statistics are provided by the function

\[
U(\Theta, \hat{\Theta}_{t,i}) = \sum_{n=0}^{K-1} E_b[b_1(n)|y(n), \mathbf{b} \in \hat{\Theta}_{t,i}] [\log f_{\mathbf{y}_e(n)}(\mathbf{y}_e(n), \mathbf{b})]
\]

where \( \hat{\Theta}_{t,j} = [\hat{\mathbf{w}}(i), \hat{\mathbf{v}}(j)] \) and \( \mathbf{y}_e(n) | y(n), \mathbf{b} \) denotes conditioning of the complete data to the observed data and the training sequence, and, according to

\(^2\) The issue of unicity falls out of the scope of this paper. We assume all interference-mitigating solutions to be valid ones.
where the EM approach yields the single iterative rule 

$$f(y_i; \Theta_i(\theta)) = \frac{1}{\pi \sigma_f^2} f_{b_1}(b_1(n)) e^{-|y_i - b_1(n)|^2/\sigma_f^2}$$  

(10)

is the likelihood of $\Theta$ w.r.t. a single complete data vector (see Appendix A.3). Substituting (10) into (9), the EM approach yields the single iterative rule

$$\hat{\Theta}_{i+1} = \arg \max_{\Theta} \left\{ U(\Theta, \hat{\Theta}_i) \right\}$$

$$= \arg \min_{\Theta} \left\{ \sum_{n=0}^{M-1} |y(n) - b_1(n)|^2 + \sum_{n=M}^{K-1} \hat{E}_{b_1(n), y(n); \hat{\Theta}_i}(|y(n) - b_1(n)|^2) \right\}.$$  

(11)

Thus, we have cast problem (8), which does not have a closed form solution, into a sequence of quadratic problems that can be analytically solved.

Nevertheless, solving (11) w.r.t. the joint parameter vector $\Theta$ is rather involved. The Space Alternating Generalized EM (SAGE) algorithm [6] is a suitable modification of the conventional EM approach that consists of successively maximizing function $U(\cdot, \cdot)$ w.r.t. different parameter subsets [6]. In our case, it is straightforward to find separate updating rules for $\hat{w}(i+1)$ and $\hat{v}(i+1)$,

$$\hat{w}(i+1) = R_{x}^{-1} \left( \sum_{n=0}^{M-1} x(n)b_1^*(n) + \sum_{n=M}^{K-1} x(n)e_{i,n} \right)$$

$$\hat{v}(i+1) = R_{b_1}^{-1} \left( \sum_{n=0}^{K-1} \hat{b}_1(n)b_1^*(n) + \sum_{n=M}^{K-1} \hat{b}_1(n)e_{i+1,n} \right)$$  

(12)

where $R_x = \sum_{n=0}^{K-1} x(n)x^H(n)$, $R_{b_1} = \sum_{n=0}^{K-1} \hat{b}_1(n)\hat{b}_1^H(n)$ and $\hat{e}_{i,n} = \hat{E}_{b_1(n), y(n); \hat{\Theta}_i}(b_1(n))$.

The latter expression can be evaluated using the Bayes theorem and yields

$$\hat{e}_{i,n} = \frac{E_{b_1(n)}[e^{-|y(n) - b_1(n)|^2/\sigma_f^2} b_1(n)]}{E_{b_1(n)}[e^{-|y(n) - b_1(n)|^2/\sigma_f^2}]}.$$  

(14)

The hard symbol estimates $\hat{b}_1(n - q), q = 1, \ldots, m - 1,$ which are used to build vector $b_1(n)$, are computed from the corresponding soft estimates, $y(n - q)$, obtained using the previous iteration filter coefficients $\hat{w}(i)$ and $\hat{v}(i)$.

Finally, note that the filtered noise variance parameter, $\sigma_f^2$, is required in order to compute $\hat{e}_{i,n}$ in Eq. (14). The variance of the Gaussian component resulting from forward filtering is $\hat{\sigma}_y^2 = \hat{w}_i^H w_i N_0$, which is not known a priori and, therefore, must be estimated. A very simple estimation method consists of iteratively updating $\hat{\sigma}_y^2$ using the estimates of $w_i$ obtained from (12), i.e.,

$$\hat{\sigma}_y^2(i + 1) = \hat{w}_i^H(i)\hat{w}(i) N_0.$$  

(15)

Thus, the updated value of $\hat{\sigma}_y^2(i + 1)$ can be used to compute $\hat{w}(i+1)$ and $\hat{v}(i+1)$. According to our computer simulations, an adequate initialization of (15) is $\hat{\sigma}_y^2 = N_0$.

At a first glance, it may seem that the algorithm given by Eqs. (12)–(13) is computationally very demanding due to the need to obtain the inverse matrices $R_x^{-1}$ and $R_{b_1}^{-1}$, which involve $O((Lm)^3)$ and $O((m - 1)^3)$ operations, respectively. However, a closer look reveals that this is not the case in practice. Even if we decide to actually implement the inversion of $R_x$, this demanding operation needs only to be carried out once, before running the algorithm, and, subsequently, only a matrix–vector product is needed at each iteration, with a complexity $O((Lm)^2)$. Alternatively, we can choose not to perform the inversion of $R_x$. Instead, we can find $\hat{w}(i+1)$ at each algorithm iteration by solving the set of linear equations

$$R_x \hat{w}(i+1) = \left( \sum_{n=0}^{M-1} x(n)b_1^*(n) + \sum_{n=M}^{K-1} x(n)e_{i,n} \right)$$

$$\hat{e}_{i,n} = \sum_{n=0}^{K-1} x(n)\hat{b}_1^H(n)\hat{v}(i)$$  

(16)

which implies $O((Lm)^2)$ operations using the standard LU factorization, or other well-known methods.
The same argument applies for matrix $\mathbf{R}_b$, hence vector $\hat{v}(i+1)$ is computed as the solution to a set of linear equations with complexity $O((m-1)^2)$ (notice that, usually, $m \ll L$).

The third factor contributing to the algorithm complexity is the computation of the posterior expectations in the right-hand side of (12) and (13). These expectations consist of summations of exponentials, that can be implemented by adequate Taylor series. The number of terms in each summation is equal to the size of the modulation alphabet, hence the overall complexity is the computation of the posterior expectations given by:

\[ \mathcal{O}(m) \times \mathcal{O}(m) = \mathcal{O}(m^2) \]

4.2. Adaptive generalized EM algorithm

As indicated above, in order to compute $\hat{w}(i+1)$ and $\hat{v}(i+1)$ from $\hat{w}(i)$ and $\hat{v}(i)$ using (12)–(13), a system of $(L+1)m-1$ linear equations with $(L+1)m-1$ unknowns has to be solved. In practice, however, it may be desirable to avoid this operation, specially when the spreading code length, $L$, is very large. Thus, we investigate an adaptive algorithm to compute $\hat{w}(i)$ and $\hat{v}(i)$ with a lower number of operations per observation.

In order to derive an adaptive version of the proposed EM-based receiver, let us define the one sample cost function

\[ U_1(\Theta, \hat{\Theta}_{l,i}) = E_{b_1(n),y(n),b_i,\hat{\Theta}_{l,i}}[\log(\mathcal{N}(\hat{y}_e(n)))] \]

which is easily obtained from (9) by setting $K = 1$. We propose to compute the sequence of parameter estimates adaptively according to the simple gradient algorithm

\[ \hat{\Theta}_{l+1,i+1} = \hat{\Theta}_{l,i} - \mu \nabla_{\hat{\Theta}_{l,i}} U_1(\hat{\Theta}_{l,i}, \hat{\Theta}_{l,i}) \]

i.e.,

\[ \hat{w}(i+1) = \hat{w}(i) - \mu \mathbf{x}(i)(\hat{y}^*(i) - d^*(i)) \]

\[ \hat{v}(i+1) = \hat{v}(i) - \mu \hat{b}(i)(\hat{y}^*(i) - d^*(i)) \]

where $y(i) = \mathbf{w}^H(i)\mathbf{x}(i) + \mathbf{v}^H(i)\hat{b}(i)$,

\[ d(i) = \begin{cases} b_1(i), & 0 \leq i \leq M - 1, \\ E_{b_1(i),y(i),\hat{\Theta}_{l,i}}[b_1(i)], & M \leq i \leq K - 1 \end{cases} \]

and $0 < \mu \ll 1$ is the step-size parameter. The sequence $\{\hat{\Theta}_{l,i}\}_{i=0,1,\ldots}$ selected in this way verifies the generalized EM (GEM) criterion [14]

\[ U_1(\hat{\Theta}_{l+1,i+1}, \hat{\Theta}_{l,i}) \geq U_1(\hat{\Theta}_{l,i}, \hat{\Theta}_{l,i}). \]

We would like to remark that algorithm (19)–(20) can be seen as a generalization of the well-known least mean squares (LMS) method. Therefore, it shares the classical weaknesses of LMS-like adaptive techniques, namely a relatively slow convergence speed and performance losses when sudden changes occur in the environment (e.g., a new user enters the channel).

Table 1 summarizes the proposed adaptive DF-GEM algorithm, including the update of the filtered noise variance estimate, $\hat{\sigma}_v^2$. Similarly to the normalized LMS algorithm [9], we have chosen a variable step-size parameter that suppresses a fixed percentage (20%) of the instantaneous squared error $e(i) = |\hat{\mathbf{w}}^H(i+1)\mathbf{x}(i) + \hat{\mathbf{v}}^H(i+1)\hat{\mathbf{b}}(i) - d(i)|^2$.

5. Computer simulations

We have carried out computer simulations to illustrate the performance of the proposed semiblind receiver in an asynchronous DS CDMA system with $N$ users transmitting QPSK symbols, length $L = 6$ and
binary codes and length $P = 10$ complex unknown user channels (thus, the channel memory size is $m = 3$). Both the spreading codes and the channel coefficients of all users have been chosen randomly.

The channel impulse responses for the different users, depicted in Fig. 2, are typical realizations of a wireless channel with a long delay spread ($P = 10$ resolvable paths with chip-period spacing). To initialize the iterative algorithm, we set the coefficients in the forward filter, $w$, to the linear solution proposed in [3]. The feedback filter is initially set to 0. For the adaptive algorithm, we start with both filters set to 0.

Fig. 3a plots symbol error rate (SER) curves for several values of the signal to noise ratio (SNR), defined as

$$\text{SNR} = 10 \log_{10} \frac{\sigma_h^2}{N_0} \sum_{r=0}^{m-1} d_1^H(r)d_1(r)$$

with $\sigma_h^2 = E[|b_1|^2]$, when the number of users in the system is $N = 4$, the number of observation vectors available to estimate the receiver coefficients is $K = 300$ and the length of the training sequence is $M = 30$ symbols. In this figure we compare:

1. The theoretical linear MMSE (LMMSE) receiver,

$$w_{\text{LMMSE}} = R^{-1}p,$$

where $R = (\sigma_h^2 D D^H + N_0 I_{Nm})$, $p = \sigma_h^2 d_1(0)$ and $I_{Nm}$ is the $Nm \times Nm$ identity matrix.
2. The decision-feedback decorrelating (DFD) receiver that is constructed with perfect knowledge of the received code matrix $D$. In order to characterize this receiver, let us consider the following decomposition of matrix $D$,$$
abla = [\tilde{d}_1(1 - m) \ldots \tilde{d}_N(1 - m) \ldots \tilde{d}_1(0) \ldots \tilde{d}_N(0) \ldots \tilde{d}_1(m - 1) \ldots \tilde{d}_N(m - 1)]$$

that allows to write the observation vector as

$$x(n) = D(n)b(n) + g(n)$$

$$= \tilde{d}_1(0)b_1(n) + DI_{SI}b_{ISI}(n) + \hat{D}b(n),$$

where $\tilde{d}(0) = [d_1^T(0) \ldots d_1^T(m - 1)]^T$ is the $Lm \times 1$ vector corresponding to the desired user current transmitted symbol,

$$D_{SI} = [\tilde{d}_1(1 - m) \ldots \tilde{d}_1(-1)]$$

is the matrix with the received codes for the causal ISI, $b_{ISI}(n) = [b_1(n - m + 1) \ldots b_1(n - 1)]^T$ is the ISI vector,

$$\hat{D} = [\tilde{d}_2(1 - m) \ldots \tilde{d}_N(1 - m)\tilde{d}_2(2 - m) \ldots \tilde{d}_N(2 - m) \ldots \tilde{d}_2(-1) \ldots \tilde{d}_N(-1)$$

$$\tilde{d}_1(m - 1) \ldots \tilde{d}_N(m - 1)]$$

is the matrix containing the received codes associated to the interferent users and the non-causal ISI, and $\hat{b}(n) = [b_2(n - m + 1) \ldots b_N(n - m + 1) \ldots b_2(n) \ldots b_N(n)b_1(n + 1) \ldots b_N(n + 1) \ldots b_1(n + m - 1) \ldots b_1(n + m - 1)]^T$. The forward filter in the DFD, $w_{DFD}$, is given by the last row in the pseudoinverse of the extended matrix $[\hat{D} \tilde{d}_1(0)]$. This formulation guarantees the complete MAI suppression and the extraction of the desired user with unit amplitude. Finally, it is enough to define the backward filter as

$$v_{DFD} = -D_{SI}R^{-1}w_{DFD}$$

to eliminate the residual ISI.

3. The theoretical DF-MMSE detector,

$$v_{DFMMSE} = (Q^HR_x^{-1}Q - \sigma_b^2[I_{m-1}]^{-1}Q^HR_x^{-1}p)$$

$$w_{DFMMSE} = R_x^{-1}(p - Qv_{DFMMSE}),$$

where $Q = \sigma_b^2[\hat{D}_1I_{m-1} \text{and } \hat{D}_1 = [d_1(1), \ldots, d_1(m - 1)]$. Both the LMMSE and the DF-MMSE
receivers require perfect knowledge of the N users’ channels, which is not available in practice. Therefore, these detectors are not practical solutions, but are simply used for benchmarking.

4. The practical block implementation of the previous DF-MMSE receiver as described in Table 2. This receiver will be termed DD LS because it can also be derived as an iterative, decision directed solution to the following least squares problem [24],

\[ \begin{bmatrix} w_{DDLS} \\ \nu_{DDLS} \end{bmatrix} = \arg \min_{w, \nu} \left\{ \frac{1}{K} \sum_{n=0}^{K-1} |y(n) - b_1(n)|^2 \right\}. \] (27)

5. The proposed semiblind iterative DF-SAGE detector.

It is apparent that the proposed DF-SAGE receiver practically matches the performance limit of the DF-MMSE receiver and clearly outperforms the LMMSE, the DFD and the conventional DDLS receivers. The poor performance of the latter detector is due to the error propagation phenomenon. Note that the DFD suffers from considerable noise power enhancement due to the definition of \( \hat{w}_{DFD} \). For the low SNR region, however, its performance is similar to the other detectors because all of them, including the DF-MMSE, exhibit a high BER. Although it is not explicitly shown in the figure, the DFD converges to the DF-MMSE for SNR values in the range 30–40 dB for the considered channel.

In Fig. 3b, the SER performance of the adaptive DF-GEM multiuser receiver is considered. The simulation parameters are \( N = 4 \) users, \( K = 4000 \) observation vectors and \( M = 150 \) training symbols. For comparison purposes, we also plot the SER curves corresponding to:

1. The theoretical LMMSE and DF-MMSE receivers,
2. The DD LMS algorithm (see Table 3) [4].

The proposed receiver approaches the performance limit of the DF MMSE receiver while the DD LMS algorithm performs only near the LMMSE limit (due to slow convergence).

Fig. 4 shows the attained SER for several values of SNR for the proposed iterative semiblind algorithm in a near-far environment. We have chosen the value of...
$E[|b_j|^2]$ so that the Signal-to-Interference Ratio (SIR) w.r.t. the $j$th user is

$$\text{SIR}_j = 10 \log_{10} \left( \frac{E[|b_1|^2]}{E[|b_j|^2]} \right) = -5 \text{ dB} \quad \forall j.$$  

The resulting SER curves show a very slight performance loss and the proposed iterative receiver still approaches the theoretical limit.

Another important measure of the receiver performance is the SER achieved for different system loads. Fig. 5 shows the SER for several values of the number of users, $N$, when SNR=12 dB. Both the iterative and adaptive algorithms present a performance close to the theoretical DF-MMSE. Only when the system is heavily loaded, the proposed algorithms show a considerable degradation w.r.t. the theoretical limit, but they still outperform the conventional procedures.

The convergence speed of the DF-SAGE iterative algorithm is illustrated in Fig. 6a, in terms of the mean square error (MSE), for an SNR of 12 dB. It is observed that a few iterations ($\approx 15$) are enough for the linear filter to practically attain the minimum MSE imposed by the LMMSE. Once the initialization process has been finished, the iterative feedback semiblind receiver approximates the theoretical DF-MMSE limit in about 20 iterations.

The convergence speed of the adaptive DF-GEM algorithm is illustrated in Fig. 6b for a SNR of 12 dB. The MSE curve obtained with the DD LMS algorithm is also depicted for comparison. It is apparent that
DF-GEM outperforms DD LMS although it still exhibits a low speed of convergence that may prevent its utilization in certain applications.

Finally, Fig. 7 depicts the SER for several SNR values for the adaptive algorithm when the spreading code length, $L$, is larger. The obtained results show that the proposed receiver approaches the theoretical limit and it still outperforms the DD LMS algorithm specially for medium and high SNR values.

6. Conclusions

We have introduced a new semiblind nonlinear approach to MAI and ISI rejection in DS CDMA. The proposed method uses the ML principle to estimate the coefficients of a DF receiver. It is termed semiblind because it uses short training sequences but also exploits the statistical information of the unknown transmitted symbols and the AWGN in the channel. Both iterative and adaptive techniques based on the generalized EM approach have been suggested for practical implementations. Computer simulations show that the proposed semiblind DF receivers attain nearly the same performance as the theoretical DF MMSE multiuser receiver using relatively short training sequences. Thus, they clearly outperform linear multiuser receivers at the expense of a very modest increase in computational complexity.
Acknowledgements

Mónica F. Bugallo acknowledges the support of the National Science Foundation (Award CCR-0082607).
Joaquín Míguez and Luis Castedo acknowledge the support of Ministerio de Ciencia y Tecnología of Spain, Xunta de Galicia and FEDER funds of the European Union (project TIC2001-0751-C04-01).

Appendix A

A.1. Derivation of $f_{y,\theta}(\cdot)$

Let us assume that the $N$ system users employ the same modulation format with i.i.d. and equiprobable symbols. Thus, an arbitrary symbol $b_i(n)$ belongs to the finite alphabet $\mathcal{B} = \{b^{(1)}, \ldots, b^{(S)}\}$, where $\log_2 S$ is the number of bits per symbol, and its p.d.f. is

$$f_b(b_i(n)) = \frac{1}{S} \sum_{l=1}^{S} \delta(b_i(n) - b^{(l)}), \quad \text{(A.1)}$$

where $\delta(\cdot)$ is Kronecker's delta function. The p.d.f. of the noisy symbols $b_i(n) + g_f(n)$ is simply the convolution of $f_b(\cdot)$ and the Gaussian p.d.f. $f_{g_f}(\cdot)$, i.e.,

$$f_{b + g_f}(z) = f_b(z) * f_{g_f}(z) = \frac{1}{S} \sum_{l=1}^{S} f_{g_f}(z - b^{(l)})$$

$$= \frac{1}{S} \sum_{l=1}^{S} \frac{1}{\sigma_f^2} e^{-|z - b^{(l)}|^2 / \sigma_f^2}$$

$$= \frac{1}{\sigma_f^2} \sum_{l=1}^{S} E_b \left[ e^{-|y(z) - b^{(l)}|^2 / \sigma_f^2} \right]. \quad \text{(A.2)}$$

When a block of $K$ observation vectors is available at the $j$th receiver, and the symbols transmitted through the $j$th antenna are i.i.d., we can use a factorized model of the soft estimates joint pdf,

$$f_{y,\theta}(y) = \left(\frac{1}{\pi \sigma_f^2}\right)^{K-1} \prod_{n=0}^{K-1} E_b \left[ e^{-|y(n) - b^{(l)}|^2 / \sigma_f^2} \right]. \quad \text{(A.3)}$$

A.2. Relationship between the proposed semiblind receiver and the DF Wiener solution

Let us consider the DF-MMSE multiuser receiver, i.e.,

$$[w_W, v_W] = \arg \min_{w,v} \left\{ E_{x(n),b(n)} \left[ |w^H x(n) + v^H b_1(n) - b_1(n)|^2 \right] \right\}. \quad \text{(A.4)}$$

It is straightforward to show that the solution to the above problem is

$$w_W = R_x^{-1}(p - Q v_W),$$

$$v_W = (Q^H R_x^{-1} Q - \sigma_b^2 I_{m-1})^{-1} Q^H R_x^{-1} p,$$ \quad \text{(A.5)}

where

$$R_x = E_{x(n),b(n)}[x(n)x^H(n)],$$

$$p = E_{x(n),b(n)}[x(n)b_1^*(n)]$$

and

$$Q = E_{x(n)}[x(n)b_1^H(n)].$$

In this appendix we will show the close relationship between the proposed receiver and the Wiener solution given by (A.5). Towards this aim, let us characterize the local maxima of the log-likelihood function

$$\mathcal{L}(\Theta)|b_j = \sum_{n=0}^{K-1} \log \left( E_{b(n)|b_j} \left[ e^{-|y(n) - b(n)|^2 / \sigma_f^2} \right] \right), \quad \text{(A.6)}$$

w.r.t. the forward filter coefficients, $w$, and the backward filter coefficients, $v$. The stationary points of $\mathcal{L}(\Theta)$ are found calculating the gradients $\nabla_w \mathcal{L}$ and $\nabla_v \mathcal{L}$ and equaling them to zero,

$$\nabla_w \mathcal{L} = \sum_{n=0}^{K-1} x(n)E_{b(n)|y(n);\theta}[y^*(n) - b_1^*(n)] = 0,$$

$$\nabla_v \mathcal{L} = \sum_{n=0}^{K-1} b_1(n)E_{b(n)|y(n);\theta}[y^*(n) - b_1^*(n)] = 0.$$ \quad \text{(A.7)}
Taking into account that $y(n) = w^H x(n) + v^H \hat{b}_1(n)$, the previous equations can be elaborated to yield

$$
\sum_{n=0}^{K-1} x(n)x^H(n)w_* = -\sum_{n=0}^{K-1} x(n)\hat{b}^H_1(n)v_* + \sum_{n=0}^{K-1} x(n)d(n),
$$

where

$$
d(n) = E_{b_1(n), y(n)}b_*, \hat{b}^*_1(n)
$$

is the nonlinear mean-squared estimate of $b_1^*(n)$ [21]. Solving for $w_*$ and $v_*$ we arrive at

$$
w_* = \hat{R}_x^{-1}(\hat{p} - \hat{Q}v_*),
$$

$$
v_* = (\hat{Q}^H\hat{R}_x^{-1}\hat{Q} - \sigma_b^2 I_{M-1})\hat{Q}^H\hat{R}_x^{-1}\hat{p},
$$

(A.10)

where

$$
\hat{R}_x^{-1} = \sum_{n=0}^{K-1} x(n)x^H(n)
$$

is the empirical autocorrelation matrix,

$$
\hat{p} = \sum_{n=0}^{K-1} x(n)d(n)
$$

is an empirical cross-correlation vector where the transmitted symbols are substituted by their mean-squared estimates (when $n \geq M$) and

$$
\hat{Q} = \sum_{n=0}^{K-1} \hat{b}_1(n)x^H(n)
$$

(A.13)

is the correlation matrix between the ISI-vector and the observation vector. It is apparent that (A.10) converges to the Wiener solution (A.5) when the block size, $K$, is large enough.

Notice that the equations in (A.10) are not useful results from a practical point of view: they do not provide explicit expressions of $w_*$ and $v_*$ because $p$ depends on the filters, while this coefficient vector must be known in order to compute the mean-squared estimates of the symbols. The algorithm proposed in this paper can be seen as an iterative method to numerically approximate (A.10).

A.3. Derivation of $f_{y_c, \Theta}(\cdot)$

When $\Theta$ is adequately chosen (i.e., $y(n) = w^H x(n) + v^H \hat{b}_1(n) = b_1(n) + g_f(n)$), the extended vector

$$
y_c(n) = \begin{bmatrix} y(n) \\ b_1(n) \end{bmatrix}
$$

is easily obtained through a linear invertible transformation of the extended symbol vector

$$
b_c(n) = \begin{bmatrix} g_f(n) \\ b_1(n) \end{bmatrix},
$$

as

$$
y_c(n) = t(b_c(n)) = \begin{bmatrix} b_1(n) + g_f(n) \\ b_1(n) \end{bmatrix}.
$$

(A.14)

It is well known that the p.d.f.’s of $y_c(n)$ and $b_c(n)$ are related by the following expression [10]

$$
f_{y_c, \Theta}(y_c(n)) = \frac{f_{b_c}(t^{-1}(y_c(n)))}{|J_t|},
$$

(A.15)

where $J_t$ is the Jacobian of the transformation and $| \cdot |$ denotes absolute value. It is straightforward to show that

$$
J_t = \det \begin{bmatrix} \frac{\partial y(n)}{\partial g_f} & \frac{\partial b_1(n)}{\partial g_f} \\ \frac{\partial y(n)}{\partial b_1(n)} & \frac{\partial b_1(n)}{\partial b_1(n)} \end{bmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1 \quad (A.16)
$$

and, assuming $g_f(n)$ is statistically independent of $b_1(n)$,

$$
f_{y_c, \Theta}(y_c(n)) = f_{b_c}(t^{-1}(y_c(n))) = f_{b_c}(b_1(n)) \cdot \frac{1}{2\pi\sigma_f^2} e^{-\frac{|y_c(n) - b_1(n)|^2}{2\sigma_f^2}}.
$$

(A.17)
References


