Semiblind linear multiuser interference cancellation: a maximum likelihood approach

Mónica F. Bugallo *, Joaquín Míguez, Luis Castedo

Departamento de Electrónica e Sistemas, Universidade da Coruña, Facultade de Informática (lab. 1.2), Campus de Elviña s/n, 15071 A Coruña, Spain

Received 3 April 2000; received in revised form 9 April 2001

Abstract

This paper introduces a linear interference suppression algorithm based on the maximum likelihood principle. The method is termed semiblind because it simultaneously exploits the transmission of training sequences and the statistics of the unknown transmitted symbols. We propose both iterative and adaptive implementations of the receiver using the expectation maximization algorithm and the matrix inversion lemma, respectively. Computer simulations show that the resulting multiuser receivers attain practically the same performance as the linear minimum mean square error multiuser detector. © 2001 Elsevier Science B.V. All rights reserved.

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

1. Introduction

Code division multiple access (CDMA) has been chosen as the basic multiple access scheme of third generation wideband mobile communication systems. The most outstanding advantages of CDMA over time division multiple access (TDMA) schemes employed in second generation systems are its higher spectral efficiency and superior flexibility in the radio interface [1,3–5]. Nevertheless, important limitations still exist for the capacity of practical CDMA systems. More specifically, multiple access interference (MAI), due to the different users simultaneously using the channel, and inter-symbol interference (ISI), due to multipath propagation, are major problems when high data rates must be attained. In order to mitigate both MAI and ISI, sophisticated signal processing capabilities have to be incorporated into wireless receivers.

Due to the prohibitive computational and informative requirements of the optimum CDMA multiuser receiver [24] alternative approaches based on linear filtering have been investigated. The decorrelating receiver [24] requires a perfect knowledge of the desired user signature waveform. This signature is not likely to be available because it depends on the channel distortion, which is usually unknown. To avoid this drawback, minimum mean square error (MMSE) detectors [16,24] resort to the transmission

* Corresponding author. Tel.: +34-981-167000; fax: +34-981-167160.

E-mail addresses: monica@des.fi.udc.es (M.F. Bugallo), miguez@des.fi.udc.es (J. Míguez), luis@des.fi.udc.es (L. Castedo).
**Nomenclature**

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b_i(n)$</td>
<td>$i$th user transmitted symbol</td>
</tr>
<tr>
<td>$b(n)$</td>
<td>transmitted symbols vector</td>
</tr>
<tr>
<td>$b_t$</td>
<td>training sequence</td>
</tr>
<tr>
<td>$D$</td>
<td>received code matrix</td>
</tr>
<tr>
<td>$g(n)$</td>
<td>noise vector</td>
</tr>
<tr>
<td>$x(n)$</td>
<td>observation vector</td>
</tr>
<tr>
<td>$w$</td>
<td>filter coefficient vector</td>
</tr>
<tr>
<td>$w_o$</td>
<td>optimum filter coefficient vector</td>
</tr>
<tr>
<td>$y$</td>
<td>soft estimate</td>
</tr>
<tr>
<td>$f_u(\cdot)$</td>
<td>probability density function (p.d.f.) of the random vector $u$</td>
</tr>
<tr>
<td>$f_{u,\Theta}(\cdot)$</td>
<td>p.d.f. of the random vector $u$ that depends on the parameter vector $\Theta$</td>
</tr>
<tr>
<td>$f_{u</td>
<td>v,\Theta}(\cdot)$</td>
</tr>
<tr>
<td>$E_u[\cdot]$</td>
<td>statistical expectation with respect to (w.r.t.) the random vector $u$</td>
</tr>
<tr>
<td>$E_{u,\Theta}[\cdot]$</td>
<td>statistical expectation w.r.t. the random vector $u$ depending on the parameter vector $\Theta$</td>
</tr>
<tr>
<td>$E_{u</td>
<td>v,\Theta}[\cdot]$</td>
</tr>
<tr>
<td>$E_{u</td>
<td>v,z,\Theta}[\cdot]$</td>
</tr>
<tr>
<td>$L(w_o)$</td>
<td>log-likelihood of $w_o$ w.r.t. the observed data $y$</td>
</tr>
<tr>
<td>$L(w_o)</td>
<td>b_t$</td>
</tr>
<tr>
<td>$T(\cdot)$</td>
<td>linear invertible transformation of $\cdot$</td>
</tr>
<tr>
<td>$J_T$</td>
<td>Jacobian of the transformation $T(\cdot)$</td>
</tr>
</tbody>
</table>

of training sequences. Such sequences, however, are not always available in many practical applications. Therefore, alternative blind implementations have been studied [6,9,15,18,19,23,25]. Most blind schemes based on the linearly constrained minimum variance (LCMV) criterion [9,15] require a precise knowledge of the desired user signature which is not likely to be available in practice. This drawback is overcome in [23], which only employs the transmitted spreading code. Nevertheless, all LCMV multiuser receivers suffer from a slow convergence rate and steady-state performance degradation [7]. Constant modulus (CM) receivers [18,19] are more robust to code mismatches but do not improve the convergence speed. Subspace techniques with a somewhat faster convergence rate have also been suggested [2,6,14,25] but they are limited by their high computational complexity and their poor performance in the low signal to noise ratio (SNR) region.

In third generation communication systems information symbols are transmitted in a burst way. Data frames are constructed according to communication protocols that typically add a priori known synchronization bits and training sequences to the unknown information symbols for many different purposes. In this paper, we introduce a novel semiblind technique for linear multiuser interference cancellation that exploits both the a priori known transmitted symbols and the statistical features of the unknown symbols transmitted by the desired user. No knowledge of the user codes or the channel response is assumed. Our approach consists in applying the maximum likelihood (ML) principle in order to estimate the coefficients of the linear receiver that suppresses both MAI and ISI. We show how this estimation
problem can be solved using the expectation maximization (EM) method. We also derive a computationally efficient adaptive implementation of the EM algorithm that attains fast convergence. It is interesting to note that since the proposed receiver jointly exploits the knowledge of the training sequences and the statistical properties of the unknown symbols, it achieves a significant performance gain with respect to (w.r.t.) conventional linear detectors that rely exclusively on the training sequence and blind linear receivers that ignore the known symbols in the data frames.

The remainder of this paper is organized as follows. Section 2 introduces the baseband discrete-time equivalent signal model of an asynchronous CDMA communication system with time-dispersive channels. In Section 2 we introduce the ML linear multiuser receiver. Section 4 describes the optimization algorithms used to compute the filter coefficients. Section 5 presents simulation results and Section 6 is devoted to the conclusions.

2. Signal model

Fig. 1 depicts the signal flow diagram of a discrete-time baseband equivalent model of an asynchronous time-dispersive direct sequence (DS) CDMA communication system. The \( i \)th user \( n \)th transmitted symbol, \( b_i(n) \), is multiplied by a unique binary-valued spreading sequence with \( L \) chips per symbol, \( c_i(k), k = 0, \ldots, L - 1 \). The resulting transmitted signal passes through a linear time-dispersive channel, \( h_i(k), k = 0, \ldots, P - 1 \), that accounts for the continuous-time channel response, the relative time delays of the different users and the transmitter and receiver front-end filters. The received signal is the superposition of the signals from the \( N \) users plus the additive white Gaussian noise (AWGN) sequence \( g(k) \), i.e.,

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

The sequence \( d_i(k) = c_i(k) * h_i(k) = \sum_{p=0}^{P-1} h_i(p)c_i(k - p) \), where \( * \) denotes convolution, has length \( L + P - 1 \) and will be termed the \( i \)th user signature waveform. Due to the channel time scattering effect, when a symbol stream is transmitted the symbol \( b_i(n) \) interferes with \( b_i(n - 1), \ldots, b_i(n - m + 1) \), where \( m = [(L + P - 1)/L] \) is the channel memory size (\( \lfloor \cdot \rfloor \) denotes upper integer part). Therefore, the overall received sequence during the \( n \)th symbol period is

\[
x(n,k) = x(nL + k) = \sum_{i=1}^{N} \sum_{r=0}^{m-1} d_i(r,k) b_i(n - r) + g(k), \quad k = 0, \ldots, L - 1,
\]
where \( d_i(r,k) = d_i(rL+k) \) is the \( r \)th segment of length \( L \) in \( d_i(k) \). Bringing together all the observations that involve the symbol of interest, \( b_i(n) \), we obtain an expression for the overall received sequence using the vector notation

\[
x(n) = D b(n) + g(n),
\]

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,

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

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

The linear multiuser receiver consists of a finite impulse response (FIR) filter, \( w = [w_1, \ldots, w_{Lm}]^T \), followed by a threshold detector as shown in Fig. 2. The soft estimate corresponding to the \( n \)th symbol period can be written as

\[
y(n) = w^H x(n),
\]

where the superindex H denotes Hermitian transposition.

3. Selection of the receiver coefficients

In this section we derive a novel statistical criterion to select the receiver coefficients in order to obtain MAI and ISI free estimates of the desired user symbols. The criterion is based on the fact that, when the MAI and the ISI are totally suppressed, the soft symbol estimate, \( y(n) \), consists of just two components:
the desired user symbol, \( b_1(n) \), and an additive Gaussian noise term, \( g_f(n) \). Indeed, let \( w_0 \) denote the optimum value of the filter coefficients that completely eliminates the MAI and the ISI. Then, we can write

\[
y(n) = w_0^H x(n) = b_1(n) + g_f(n),
\]

where \( b_1(n) \) is the desired user symbol and \( g_f(n) \) is a complex Gaussian random variable with zero mean and variance \( \sigma^2_f = \sigma^2_w w_0^H w_0 \). In this section, the filtered noise variance, \( \sigma^2_f \), 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 \( \sigma^2_f \). It is straightforward to show that the probability density function (p.d.f.) of \( y(n) = b_1(n) + g_f(n) \) is given by

\[
f_{y,w_0}(y(n)) = f_{y+g_f}(y(n)) = \frac{1}{\pi \sigma^2_f} E_b [e^{-|y(n) - b_1|^2/\sigma^2_f}],
\]

where \( E_b[\cdot] \) denotes statistical expectation w.r.t. the desired user transmitted symbols.

In digital communications, the transmitted symbols are usually modelled as discrete i.i.d. random variables with known p.d.f. and finite alphabet. As a consequence, the soft estimates obtained with the optimum filter \( w_0 \) can also be considered as i.i.d. random variables and, 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

\[
f_{y,w_0}(y) = \prod_{n=0}^{K-1} f_{y, w_0}(y(n)) = \left( \frac{1}{\pi \sigma^2_f} \right)^K \prod_{n=0}^{K-1} E_b [e^{-|y(n) - b_1|^2/\sigma^2_f}],
\]

that can be analytically calculated because a statistical expectation w.r.t. a finite alphabet discrete random variable reduces to a simple summation. Note that the p.d.f. of \( y \) given by Eq. (7) depends on the unknown parameter \( w_0 \). Thus, the ML estimate of \( w_0 \) corresponds to

\[
\hat{w} = \arg \max_{w_0} \left\{ \mathcal{L}(w_0) = \log \prod_{n=0}^{K-1} E_b [e^{-|y(n) - b_1|^2/\sigma^2_f}] = \sum_{n=0}^{K-1} \log E_b [e^{-|y(n) - b_1|^2/\sigma^2_f}] \right\},
\]

where \( \mathcal{L}(w_0) \) is the log-likelihood of \( w_0 \) w.r.t. the observed data \( y = [y(0), \ldots, y(K-1)] \). Note that the expression of the log-likelihood, \( \mathcal{L}(w_0) \), involves averaging over the p.d.f. of the unknown transmitted symbols. Therefore, the proposed approach can be classified within the family of stochastic ML methods [22]. Nevertheless, such classification only applies to the computation of the detector coefficients, since the receiver itself is a linear processor. As a consequence, it is suboptimum in terms of bit error rate (BER) but it presents a very reduced computational complexity when compared to true ML detectors [24].

The log-likelihood \( \mathcal{L}(w_0) \) is a nonquadratic function with several maxima. In particular, the solutions to problem (8) guarantee that the soft estimates, \( y(n) \), have a p.d.f. close to \( f_{b+g_f}(\cdot) \) but this is not enough to ensure that the desired user is extracted. Since in CDMA all users transmit symbols with the same modulation format, the p.d.f. of the \( i \)th interference at the receiver is \( f_{b+g_f}(\cdot) \) which does not differ from the desired user p.d.f.. Therefore, solving the optimization problem (8) may lead to the capture of the \( i \)th interference.

In order to alleviate this limitation we propose to exploit the transmission of a short training sequence of \( M < K \) symbols. Such sequences are usually employed for different purposes in currently
standardized mobile communication systems. Indeed, let us assume that the first $M$ symbols (i.e., $b = [b_1(0), \ldots, b_1(M-1)]^T$) are known a priori by the receiver. Conditioning the expectations in (8) w.r.t. the known symbols, $b$, we arrive at a semiblind receiver where the filter coefficients are computed as the solution to

$$\hat{w} = \arg \max_{w_o} \left\{ \mathcal{L}(w_o) \big|_{b} = \sum_{n=0}^{K-1} \log E_{b_1(n)}[e^{-|x(n)-b_1(n)|^2/s_1^2}] \right\}$$

$$= \arg \min_{w_o} \left\{ \sum_{n=0}^{M-1} |y(n) - b_1(n)|^2 - \sum_{n=M}^{K-1} \log E_{b_1}[e^{-|y(n)-b_1|^2/s_1^2}] \right\}. \quad (9)$$

The computer simulations in Section 5 show that relatively short training sequences are enough to avoid the capture problem. This is because the first term in (9) is a purely quadratic form with a single maximum that corresponds to the rough extraction of the desired user. Note that, since $\hat{w}$ is still computed according to the ML principle, all the available statistical information is employed to obtain the filter coefficients, and hence the proposed semiblind ML linear receiver outperforms conventional MMSE multiuser detectors [16,24] that only exploit the training sequence $b$.

It is interesting to remark that the semiblind formulation of the linear receiver in (9) is just a particular case of the more general ML approach that led to problem (8) (which is a totally blind formulation). On the contrary, most semiblind receivers proposed so far in the literature [11–13] result from an ad hoc approach that consists of combining two different criteria, one for the training sequence and another one for the unknown symbols.

### 3.1. Existence of $w_o$

According to Eq. (5), $w_o$ is the decorrelating solution for user 1, i.e., it is the first row in the left pseudoinverse of the distorted code matrix, $D$, that we denote as $D^\dagger$. It is apparent that matrix $D^\dagger$ only exists when $N$, $L$ and $m$ verify the inequality

$$N(2m-1) \leq Lm. \quad (10)$$

Nevertheless, even when (10) does not hold, the existence of the Wiener solution [8]

$$w_W = \left( E[x(n)x^H(n)] \right)^{-1} E[x(n)b_1^*(n)] \quad (11)$$

is guaranteed. The soft estimates resulting from Wiener filtering can be written as

$$y(n) = w_W^H x(n) = b_1(n) + g_{\ell}(n) + r(n), \quad (12)$$

where $r(n)$ is the residual MAI and ISI. Since the interference power, $E[r_{\text{II}}(n)r(n)]$, is very low (specially for medium and high values of the SNR) the associated p.d.f. $f_{y,w_W}(\cdot)$ is very close to Eq. (6). Indeed, it can be shown (see Appendix A) that the solution to problem (9) has the same structure as the Wiener filter.
4. Implementation

4.1. Iterative-block implementation

Unfortunately, it is not possible to find a closed form solution to problem (9) and, therefore, some optimization algorithm must be used to obtain the filter coefficients. In this work, we propose to compute \( w_o \) using the EM algorithm [20] that provides an iterative procedure to perform ML estimation when direct maximization of the likelihood is not feasible. The EM approach postulates the existence of some missing (unobserved) data that, if known, would aid in the estimation problem. The algorithm consists of a two-step iteration: use the incomplete (observed) data and the current parameter estimates to compute sufficient statistics of the complete data (E-step) and re-estimate the parameters using the computed complete data sufficient statistics (M-step). The sequence of estimates thus obtained exhibits the desirable property of being monotonically nondecreasing in likelihood.

In our problem, the incomplete-data set is given by the soft estimates, \( \mathbf{y}(n), n = 0, \ldots, K-1 \), whereas the complete-data set is given by the extended vectors \( \mathbf{y}_c(n) = [y(n), b_1(n)]^T, n = 0, \ldots, K-1 \). Let us build the complete-data block \( \mathbf{Y}_c = [\mathbf{y}_c(0), \ldots, \mathbf{y}_c(K-1)] \), with joint p.d.f. \( f_{\mathbf{Y}_c,w_o}(\cdot) \). It is easy to decompose \( f_{\mathbf{Y}_c,w_o}(\mathbf{y}) \) as

\[
 f_{\mathbf{Y}_c,w_o}(\mathbf{Y}_c) = f_{\mathbf{Y}_c|\mathbf{y}_c,w_o}(\mathbf{Y}_c) \cdot f_{\mathbf{y}_c}(\mathbf{y}). \tag{13} \]

Taking logarithms and conditional expectations on both sides of (13) yields

\[
 E_{\mathbf{Y}_c|\mathbf{y}_c,w_o}(\log f_{\mathbf{y}_c}(\mathbf{y})) = E_{\mathbf{Y}_c|\mathbf{y}_c,w_o}(\log (f_{\mathbf{Y}_c,w_o}(\mathbf{Y}_c)) - \log (f_{\mathbf{Y}_c|\mathbf{y}_c,w_o}(\mathbf{Y}_c))) \tag{14} \]

and it is straightforward to realize that

\[
 E_{\mathbf{Y}_c|\mathbf{y}_c,w_o}(\log f_{\mathbf{y}_c}(\mathbf{y})) = \log f_{\mathbf{y}_c}(\mathbf{y}). \tag{15} \]

Let us now introduce the following notation:

\[
 U(\mathbf{w}_o, \hat{\mathbf{w}}(i)) = E_{\mathbf{Y}_c|\mathbf{y}_c,w_o}(\log (f_{\mathbf{Y}_c,w_o}(\mathbf{Y}_c))), \tag{16} \]

\[
 V(\mathbf{w}_o, \hat{\mathbf{w}}(i)) = E_{\mathbf{Y}_c|\mathbf{y}_c,w_o}(\log (f_{\mathbf{Y}_c|\mathbf{y}_c,w_o}(\mathbf{Y}_c))). \tag{17} \]

Substituting (15)–(17) into (14) we arrive at the relationship

\[
 \log f_{\mathbf{y}_c}(\mathbf{y}) = U(\mathbf{w}_o, \hat{\mathbf{w}}(i)) - V(\mathbf{w}_o, \hat{\mathbf{w}}(i)), \tag{18} \]

which provides a tractable form of the log-likelihood function \( \log f_{\mathbf{y}_c}(\cdot) \). Indeed, an application of Jensen’s inequality allows us to show that [17]

\[
 V(\mathbf{w}_o, \hat{\mathbf{w}}(i)) \leq V(\hat{\mathbf{w}}(i), \hat{\mathbf{w}}(i)) \tag{19} \]

for any value of \( \mathbf{w}_o \) and, as a consequence, the sequence of estimates

\[
 \hat{\mathbf{w}}(i+1) = \arg \max_{\mathbf{w}_o} U(\mathbf{w}_o, \hat{\mathbf{w}}(i)) \tag{20} \]

is nondecreasing in likelihood since

\[
 \log f_{\mathbf{y}_c}(\mathbf{y}) \geq U(\hat{\mathbf{w}}(i), \hat{\mathbf{w}}(i)) - V(\hat{\mathbf{w}}(i), \hat{\mathbf{w}}(i)) = \log f_{\mathbf{y}_c}(\mathbf{y}). \tag{21} \]

Substituting the joint p.d.f.

\[
 f_{\mathbf{Y}_c,w_o}(\mathbf{Y}_c) = \prod_{n=0}^{K-1} \frac{1}{\pi \sigma^2_1} e^{-|y(n) - b_1(n)|^2/\sigma^2_1} \cdot f_b(b_1(n)) \tag{22} \]
(see Appendix B) into (20) leads to the single iterative rule
\[
\hat{\mathbf{w}}(i + 1) = \arg \max_{\mathbf{w}_o} \{ U(\mathbf{w}_o, \hat{\mathbf{w}}(i)) \}
\]
\[
= \arg \min_{\mathbf{w}_o} \left\{ \sum_{n=0}^{M-1} |y(n) - b_1(n)|^2 + \sum_{n=M}^{K-1} E_{y(n) | y(n); \hat{\mathbf{w}}(i)} \left[ |y(n) - b_1(n)|^2 \right] \right\},
\]
(23)
where we have neglected all terms that are constant w.r.t. \( \mathbf{w}_o \).

At this point, we have cast problem (9), which does not have a closed form solution, into a sequence of quadratic problems that can be analytically solved. It can be demonstrated that the solution to (23) is given by
\[
\hat{\mathbf{w}}(i + 1) = \left( \sum_{n=0}^{K-1} \mathbf{x}(n) \mathbf{x}^H(n) \right)^{-1} \left( \sum_{n=0}^{M-1} \mathbf{x}(n) b_1^*(n) + \sum_{n=M}^{K-1} E_{b_1(n) | y(n); \hat{\mathbf{w}}(i)} [b_1^*(n)] \mathbf{x}(n) \right),
\]
(24)
where we have also used the fact that the only random part in \( y_o(n) | y(n) \) is \( b_1(n) \). Notice that, for \( M = K \), the previous algorithm reduces to the least squares (LS) filter. Appendix C shows the derivation of the conditioned expectation \( E_{b_1(n) | y(n); \hat{\mathbf{w}}(i)} [ \cdot ] \) \( n \geq M \) using the Bayes theorem.

Finally, note that the filtered noise variance parameter, \( \sigma^2_t \), is required in order to compute \( E_{b_1(n) | y(n); \hat{\mathbf{w}}(i)} [ \cdot ] \), \( n \geq M \), in Eq. (24). Since \( \sigma^2_t = \mathbf{w}_o^H \mathbf{w}_o \sigma^2_g \) is not known a priori, it must be estimated. A very simple estimation method consists of iteratively updating \( \sigma^2_t \) using the estimates of \( \mathbf{w}_o \) obtained from (24), i.e.,
\[
\hat{\sigma}^2_t(i + 1) = \mathbf{w}_o^H(i) \hat{\mathbf{w}}(i) \hat{\sigma}^2_g(i).
\]
(25)
Thus, the updated value of \( \hat{\sigma}^2_t(i + 1) \) can be used to compute \( \hat{\mathbf{w}}(i + 1) \). According to our computer simulations, an adequate initialization of (25) is \( \hat{\sigma}^2_t = \hat{\sigma}^2_g \), where \( \hat{\sigma}^2_g \) is a rough estimate of the AWGN variance.

4.2. Adaptive implementation

In order to compute \( \hat{\mathbf{w}}(i) \) from \( \hat{\mathbf{w}}(i - 1) \) according to (24), a system of \( L_m \) linear equations with \( L_m \) 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. We would also like to compute the estimate \( \hat{\mathbf{w}}(i) \) recursively in order to minimize the number of operations per iteration. Both objectives can be realized by using the matrix inversion lemma [21]. Applying this result in (24) and denoting
\[
\mathbf{P}(0) = \mathbf{I}_{L_m},
\]
\[
\mathbf{P}(i) = \left( \sum_{n=0}^{i} \mathbf{x}(n) \mathbf{x}^H(n) \right)^{-1} = \mathbf{P}(i - 1) - \mathbf{k}(i) \mathbf{x}^H(i) \mathbf{P}(i - 1), \quad i > 0
\]
(26)
and
\[
\mathbf{k}(i) = \frac{\mathbf{P}(i - 1) \mathbf{x}(i)}{1 + \mathbf{x}^H(i) \mathbf{P}(i - 1) \mathbf{x}(i)},
\]
(27)
Table 1

Adaptive EM receiver

Initialize the algorithm by setting

\[ P(0) = I_{Lm} \]

\[ \sigma_f^2(0) \approx \sigma_f^2 \]

For each available observation vector \( i = 1, 2, \ldots \), compute

\[ k(i) = \frac{P(i-1)x(i)}{1 + x^H(i)P(i-1)x(i)} \]

\[ \sigma_f^2(i) = \sigma_f^2(i-1)\hat{w}^H(i-1)\hat{w}(i-1), \quad (i \geq M) \]

\[ d(i) = \begin{cases} b^*(i), & 0 \leq i \leq M-1 \\ E_{\hat{b}(i)\hat{w}(i)}b^*(i), & M \leq i \leq K-1 \end{cases} \]

\[ \hat{w}(i) = \hat{w}(i-1) + k(i)[d(i) - x^H(i)\hat{w}(i-1)] \]

\[ P(i) = P(i-1) - k(i)x^H(i)P(i-1) \]

where \( I_{Lm} \) is the \( Lm \times Lm \) identity matrix, we get the desired equation for updating the filter coefficients,

\[ \hat{w}(i) = \hat{w}(i-1) + k(i)[d(i) - x^H(i)\hat{w}(i-1)] \]

(28)

where

\[ d(i) = \begin{cases} b^*(i), & 0 \leq i \leq M-1 \\ E_{\hat{b}(i)\hat{w}(i)}b^*(i), & M \leq i \leq K-1 \end{cases} \]

(29)

We would like to remark that the adaptive EM algorithm (28) is a generalization of the well-known recursive least squares (RLS) method [8]. Actually, (28) reduces to the RLS rule when \( M = K \). Table 1 summarizes the proposed adaptive algorithm, including the update of the filtered noise variance estimate, \( \sigma_f^2 \).

5. Computer simulations

Finally, we present computer simulations that illustrate the validity of our approach. We have considered an asynchronous DS CDMA communication system with \( N = 4 \) users transmitting QPSK symbols and length \( L = 8 \) random binary spreading codes,

\[ C^T = \begin{bmatrix} -1 & 1 & 1 & 1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & 1 & 1 \end{bmatrix} \]
The user signals propagate through a randomly chosen unknown channel with $P = 16$ complex coefficients per user. The resulting multiuser channel matrix is

$$
\begin{bmatrix}
    h_1(0) & \cdots & h_N(0) \\
    \vdots & \ddots & \vdots \\
    h_1(P-1) & \cdots & h_N(P-1)
\end{bmatrix}
\begin{bmatrix}
    0.23 + 0.26i & 0.01 - 0.24i & 0.25 - 0.36i & -0.00 - 0.17i \\
    0.63 + 0.09i & 0.53 - 0.26i & -0.01 - 0.13i & 0.44 - 0.06i \\
    -0.33 + 0.15i & 0.57 - 0.13i & -0.30 - 0.94i & 1.14 - 0.39i \\
    -0.64 - 0.08i & -0.00 + 0.54i & -0.06 - 0.47i & -0.12 + 0.03i \\
    0.17 - 0.02i & -0.50 + 0.08i & 0.41 + 0.94i & 0.23 - 0.60i \\
    -0.34 + 0.38i & 0.28 + 0.75i & -0.09 + 0.10i & 0.17 - 0.57i \\
    -0.18 + 0.61i & -0.21 + 0.22i & -0.18 + 0.74i & -0.64 - 0.21i \\
    0.71 + 0.36i & -0.43 + 0.54i & 1.08 + 0.25i & -0.31 + 0.59i \\
    0.46 - 1.10i & 0.39 + 0.07i & -0.93 - 0.18i & 0.24 - 0.01i \\
    0.12 + 0.45i & -0.27 - 0.09i & 0.05 - 0.16i & 0.12 - 0.55i \\
    -0.37 + 0.30i & 0.26 - 0.47i & -0.48 + 0.23i & 0.70 - 0.38i \\
    -0.19 - 0.80i & 0.22 + 0.62i & 0.58 - 0.86i & 0.29 + 0.68i \\
    -0.90 - 1.03i & -0.01 - 0.63i & -0.12 + 0.15i & -0.75 - 1.36i \\
    -1.06 - 0.47i & 0.04 + 0.20i & 0.42 + 0.04i & 0.38 - 0.19i \\
    0.63 + 0.00i & 1.35 + 0.17i & 0.68 + 0.51i & -0.43 + 0.64i \\
    0.72 - 0.13i & 0.34 + 0.48i & 0.39 + 1.10i & 0.62 - 0.62i
\end{bmatrix}
$$

where $i = \sqrt{-1}$.

Fig. 3 (left) plots the mean square error (MSE) attained by the proposed iterative EM receiver (labelled EM) for several values of the input SNR, defined as

$$
\text{SNR} = 10 \log_{10} \frac{E[|b_1|^2]|d|^2}{\sigma_g^2} \text{ (dB)}
$$

(30)
when the number of observation vectors available to estimate the receiver coefficients is $K = 300$, including a training sequence of length $M = 15$. It is apparent that the proposed iterative-block algorithm performs close to the theoretical MMSE receiver (labelled MMSE) constructed with perfect knowledge of all user codes and channel coefficients. In this figure we have also plotted the MSE achieved by the conventional LS receiver (labelled LS) implemented with the RLS algorithm (as defined in [8]) using the training sequence, $b_s$, alone, as it is done in conventional supervised multiuser receivers. It can be seen that this detector performs considerably worse than the theoretical MMSE because the training sequence is too short to obtain an adequate estimation of the filter coefficients.

Fig. 3 (right) illustrates the performance of the adaptive EM receiver (labelled AEM) in terms of MSE using $K = 600$ observation vectors and a training sequence of $M = 30$ symbols. The reduction in computational complexity of the AEM algorithm w.r.t. its iterative counterpart is achieved at the expense of an increase in the number of observation vectors required to attain an adequate performance. The AEM receiver is compared, in terms of MSE, with the conventional RLS receiver that exploits only the training sequence, $b_s$. It is clear that, for practical values of the SNR, the proposed adaptive detector performs very close to the theoretical MMSE and achieves a clear improvement w.r.t. the conventional RLS detector.

Fig. 4 (left) depicts the MSE of the theoretical MMSE, LS and the iterative EM linear receivers in a near-far environment (with $K = 300$ and $M = 15$). Let us define the signal to interference ratio (SIR) of the desired user w.r.t. the $j$th interference as

$$\text{SIR}_j = 10 \log_{10} \frac{d_j^4 d_j E[|b_j|^2]}{d_j^4 d_j E[|b_j|^2]}$$

and let us choose the value of $E[|b_j|^2]$ so that $\text{SIR}_j = -5$ dB $\forall j$. Simulation results show that all detectors suffer a moderate degradation in performance but the proposed iterative receiver still approaches the theoretical limit for moderate and high SNR values. Similar results are obtained with the AEM receiver, as shown in Fig. 4 (right) with parameters $K = 600$ and $M = 30$.

Another important measure of the receiver performance is the MSE achieved for different system loads. Fig. 5 (left) plots the MSE for several values of the number of users, $N$, with $K = 300$, $M = 15$, $\text{SNR} = 12$ dB and $\text{SIR}_j = 0$ dB $\forall j$. The resulting curve shows that the performance degradation of the iterative EM receiver with increasing system load is the same one suffered by the theoretical MMSE receiver, whereas the conventional LS receiver performance is considerably worse. The MSE of the iterative receiver becomes considerably higher than the theoretical MMSE only when the system is heavily loaded. Fig. 5 (right) shows the corresponding results for the AEM receiver (with $K = 600$, $M = 30$).
Fig. 5. MSE vs. number of users in a dispersive asynchronous DS CDMA system with length \( L = 8 \) random binary spreading codes, length \( P = 16 \) discrete channels, SNR = 12 dB and SIR\(_j\) = 0 dB \( \forall j \). Left: \( K = 300, M = 15 \). Right: \( K = 600, M = 30 \).

Fig. 6. SER for several values of the SNR in a dispersive asynchronous DS CDMA system with \( N = 4 \) users, length \( L = 8 \) random binary spreading codes, length \( P = 16 \) discrete channels and SIR\(_j\) = 0 dB \( \forall j \). Left: \( K = 300, M = 15 \). Right: \( K = 600, M = 30 \).

Fig. 6 (left) depicts the symbol error rate (SER) attained by the proposed iterative EM algorithm for several values of the input SNR with system parameters \( N = 4, K = 300, M = 15 \) and SIR\(_j\) = 0 dB \( \forall j \). The resulting curves show that the proposed receiver performs close to the theoretical MMSE detector, whereas the performance of the LS receiver is considerably worse. The same conclusion is drawn from the simulation results shown in Fig. 6 (right) with \( K = 600 \) and \( M = 30 \).

Fig. 7 illustrates the fast convergence speed of the proposed algorithms. With the same simulation parameters as in the previous experiments and SNR = 12 dB, we have plotted the MSE at the receiver output as a function of the number of iterations (left) and the number of observation vectors used in the estimation of the filter coefficients (right) for the iterative and adaptive algorithms, respectively. It is apparent that the iterative receiver converges in less than 20 iterations and the recursive detector attains the theoretical limit with a moderate number of observation vectors. This may be an important advantage when time or computational load constraints have to be fulfilled.

Finally, we have carried out computer simulations in order to evaluate the relative merit of the proposed EM-based receivers when compared with the other blind and semiblind multiuser detectors. In particular, we have considered the following receivers:
- A block-iterative decision directed LS (DD LS) receiver.
Fig. 7. Convergence speed of the proposed algorithms in a dispersive asynchronous DS CDMA system with $N = 4$ users, length $L = 8$ random binary spreading codes, length $P = 16$ discrete channels, SNR = 12 dB and SIR$_j = 0$ dB. ∀$j$ Left: MSE vs. no. of iterations ($K = 300$ and $M = 15$). Right: MSE vs. no. of observation vectors ($K = 600$ and $M = 30$).

Fig. 8. Comparison with other methods: MSE for several values of the SNR in a dispersive asynchronous DS CDMA system with $N = 4$ users, length $L = 8$ random binary spreading codes, length $P = 16$ discrete channels and SIR$_j = 0$ dB. ∀$j$. Left: $K = 300$, $M = 15$. Right: $K = 600$, $M = 30$.

- A closed-form solution implementation of the blind LCMV (B LCMV) multiuser detector [9] which requires perfect knowledge of the desired user channel coefficients, $h_1(0), \ldots, h_1(P - 1)$.
- The blind Subspace Algorithm (BSA) proposed by Wang and Poor [26] with a smoothing factor of 8.
- A decision directed implementation (hard decisions are employed) of the RLS (DD RLS) algorithm [8].

The DDLS, SB CMA, B LCMV and BSA approaches are compared with the iterative EM receiver in Fig. 8 (left). It is apparent that the proposed method practically matches the theoretical MMSE whereas the remaining receivers present a significantly poorer performance. The simulation parameters are $K = 300$ and $M = 15$.

The AEM algorithm is compared with the DD RLS approach in Fig. 8 (right). The simulation parameters are $K = 600$ and $M = 30$. In this case, there is only a slight performance advantage of the AEM algorithm w.r.t. the conventional DD RLS approach. This result is actually a rather intuitive one, since the AEM algorithm can also be seen as a decision directed approach where symbol estimates are obtained according to the nonlinear mean-squared method [21] (see Eq. (29)).
6. Conclusions

We have introduced a new semiblind approach to linear interference cancellation in DS CDMA. The proposed receiver relies on the ML principle to estimate the coefficients of a linear FIR filter that suppresses both MAI and ISI. The method exploits both the transmission of short training sequences and the knowledge of the transmitted symbols and noise p.d.f. Two possible implementations are suggested: an iterative-block one based on the EM algorithm and an adaptive one that uses the matrix inversion lemma in order to reduce the computational complexity. Finally, computer simulations show that, using relatively short training sequences and small blocks of observations, the resulting semiblind receivers attain the same performance as the theoretical linear MMSE multiuser detector.

Acknowledgements

This work has been supported by FEDER funds (grant number 1FD97-0082) and Xunta de Galicia (PGIDT00PXI10504PR). We would also like to thank the anonymous reviewers for their comments (specially reviewer A) that have helped to significantly improve the final version of the paper.

Appendix A. Relationship between the proposed semiblind receiver and the linear Wiener solution

Let us consider the linear MMSE multiuser receiver, i.e.,

\[ w = \arg \min_w \{E_x[n][|w^H x(n) - b_1(n)|^2]\}. \] (A.1)

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

\[ w = R^{-1} p, \] (A.2)

where \( R = E_x[n][x(n)x^H(n)] \) and \( p = E_x[n][x(n)b_1^*(n)] \). In this appendix we will show the close relationship between the proposed receiver and the Wiener solution given by (A.2). With this aim, let us characterize the local maxima of the log-likelihood function

\[ \mathcal{L}(w) \mid b_t = \sum_{n=0}^{K-1} \log E_{b_t}[e^{-|y(n) - b|^2/a^2}] \] (A.3)

w.r.t. the filter coefficients \( w \). The stationary points of \( \mathcal{L}(w) \) are found on calculating the gradient \( \nabla_w \mathcal{L} \) and equalling it to zero:

\[ \nabla_w \mathcal{L} = \sum_{n=0}^{K-1} x(n)E_{b_t}[y^*(n) - b_t^*(n)] = 0. \] (A.4)

Taking into account that \( y(n) = w^H x(n) \), the previous equation can be elaborated to yield

\[ \sum_{n=0}^{K-1} x(n)x^H(n)w = \sum_{n=0}^{K-1} x(n)d(n), \] (A.5)

where

\[ d(n) = E_{b_t}|y(n); b_t; w[b_t^*(n)] = \begin{cases} b^*(n), & 0 \leq n \leq M - 1, \\ E_{b_t}|y(n); w b^*(n), & M \leq n \leq K - 1 \end{cases} \] (A.6)
is the nonlinear mean-squared estimate of $b_1^*(n)$ [21]. Solving for $w$ we arrive at
\[ w = \hat{R}_x^{-1} \hat{p}, \] (A.7)
where
\[ \hat{R}_x^{-1} = \sum_{n=0}^{K-1} x(n)x^H(n) \] (A.8)
is the empirical autocorrelation matrix and
\[ \hat{p} = \sum_{n=0}^{K-1} x(n)d(n) \] (A.9)
is an empirical cross-correlation vector where the transmitted symbols are substituted by their mean-squared estimates (when $n \geq M$). It is apparent that (A.7) converges to the Wiener solution (A.2) when the block size, $K$, is large enough.

Note that Eq. (A.7) is not a useful result from a practical point of view since it is not an explicit expression of $w$ because $p$ depends on $w$. This coefficient vector must be known in order to compute the mean-squared estimates of the symbols. The EM algorithm proposed in this paper is actually an iterative method to numerically approximate (A.7).

**Appendix B. Derivation of $f_{y_e w_e} (\cdot)$**

When $w$ is adequately chosen, i.e., $w = w_o$, $y(n) = w_o^H x(n) = b_1(n) + g_1(n)$ the extended vector $y_e(n) = [y(n)\ b_1(n)]^T$ is easily obtained through a linear invertible transformation of the extended symbol vector $b_e(n) = [g_1(n)\ b_1(n)]^T$ as
\[ y_e(n) = T(b_e(n)) = \begin{bmatrix} b_1(n) + g_1(n) \\ b_1(n) \end{bmatrix}. \] (B.1)

It is well known that the p.d.f.'s of $y_e(n)$ and $b_e(n)$ are related by the following expression [10]:
\[ f_{y_e w_e} (y_e(n)) = \frac{f_{b_e} (T^{-1}(y_e(n)))}{|J_T|}, \] (B.2)
where $J_T$ is the Jacobian of the transformation and $| \cdot |$ denotes absolute value. It is straightforward to show that
\[ J_T = \det \begin{bmatrix} \frac{\partial y(n)}{\partial g_1} & \frac{\partial b_1(n)}{\partial g_1} \\ \frac{\partial y(n)}{\partial b_1(n)} & \frac{\partial b_1(n)}{\partial b_1(n)} \end{bmatrix} = \begin{vmatrix} 1 & 0 \\ 1 & 1 \end{vmatrix} = 1 \] (B.3)

and, assuming $g_1(n)$ is statistically independent of $b_1(n),
\[
\begin{align*}
  f_{y_e w_e} (y_e(n)) &= f_{b_e} (T^{-1}(y_e(n))) \\
  &= f_b(b_1(n)) \cdot f_{g_1}(y(n) - b_1(n)) = f_b(b_1(n)) \cdot \frac{1}{\pi \sigma_1^2} e^{-|y(n) - b_1(n)|^2/\sigma_1^2}.
\end{align*}
\] (B.4)

Since the transmitted symbols are i.i.d., the joint p.d.f. of the $K$ extended vectors, $Y_e = [y_e(0), \ldots, y_e(K-1)]$, can be written as
\[ f_{Y_e w_e} (Y_e) = \prod_{n=0}^{K-1} \frac{1}{\pi \sigma_1^2} e^{-|y(n) - b_1(n)|^2/\sigma_1^2} \cdot f_b(b_1(n)). \] (B.5)
Appendix C. Derivation of the expectation $E_{b|y;\hat{\mathbf{x}}(i)}[\cdot]$ 

Let us consider the general expression for $E_{b|y;\hat{\mathbf{x}}(i)}[\cdot]$, 

$$E_{b|y;\hat{\mathbf{x}}(i)}[\Phi(b)] = \sum_b f_{b|y;\hat{\mathbf{x}}(i)}(b) \cdot \Phi(b), \quad (C.1)$$

where the summation is over all possible values of a single symbol, $f_{b|y;\hat{\mathbf{x}}(i)}$ is the p.d.f. of symbol $b$ conditioned to the output $y$ obtained with filter coefficients $\hat{\mathbf{x}}(i)$ and $\Phi(\cdot)$ is an arbitrary function. Using the Bayes theorem we can express the a posteriori density $f_{b|y;\hat{\mathbf{x}}(i)}(\cdot)$ in terms of $f_{y;\hat{\mathbf{x}}(i)}(\cdot)$, $f_{y|b;\hat{\mathbf{x}}(i)}(\cdot)$ and $f_b(\cdot)$ as 

$$f_{b|y;\hat{\mathbf{x}}(i)}(b(n)) = \frac{f_{y|b;\hat{\mathbf{x}}(i)}(y) \cdot f_b(b(n))}{f_{y;\hat{\mathbf{x}}(i)}(y)} = \frac{f_b(b(n))e^{-|y-b(n)|^2/\sigma_n^2}}{E_b[e^{-|y-b|/\sigma_n^2}]} = \frac{e^{-|y-b(n)|^2/\sigma_n^2}}{\sum_b e^{-|y-b|^2/\sigma_n^2}} \quad (C.2)$$

where $f_{y|b;\hat{\mathbf{x}}(i)}(\cdot)$ is obtained from $f_{y;\hat{\mathbf{x}}(i)}(\cdot)$ by simply dropping the expectation (see Eq. (6)). Substituting (C.2) into (C.1), the following relationship is obtained:

$$E_{b|y;\hat{\mathbf{x}}(i)}[\Phi(b)] = \frac{E_b[e^{-|y-b|/\sigma_n^2} \Phi(b)]}{E_b[e^{-|y-b|^2/\sigma_n^2}]} = \frac{\sum_b e^{-1/\sigma_n^2}|y-b|^2 \Phi(b)}{\sum_b e^{-1/\sigma_n^2}|y-b|^2}. \quad (C.3)$$

References