Minimum Probability of Error-Based Methods for Adaptive Multiuser Detection in Multipath DS-CDMA Channels

Aditya Dua, Student Member, IEEE, Uday B. Desai, Senior Member, IEEE, and Ranjan K. Mallik, Senior Member, IEEE

Abstract—Direct-sequence code-division multiple-access (DS-CDMA) is a popular multiple-access technology for wireless communications. However, its performance is limited by multiple-access interference and multipath distortion. Multiuser detection and space–time processing are two signal processing techniques employed to improve the performance of DS-CDMA. Two minimum probability of error-based space–time multiuser detection algorithms are proposed in this paper. The first algorithm, minimum joint probability of error (MJPOE), aims to minimize the joint probability of error for all users. The second algorithm, minimum conditional probability of error (MCPOE), minimizes the probability of error of each user conditioned on the transmitted bit vector, for each user individually. In both the algorithms, the optimal filter weights are computed adaptively using a gradient descent approach. The MJPOE algorithm is blind and offers a bit-error-rate (BER) performance better than the nonadaptive minimum mean squared error (MMSE) algorithm, at the cost of higher computational complexity. An approach for reducing the computational overheads of MJPOE using Gram–Schmidt orthogonalization is suggested. The BER performance of the MCPOE algorithm is slightly inferior to MMSE, however, it has a computational complexity linear in the number of users. Both blind and training-based implementations for MCPOE are proposed. Both MJPOE and MCPOE have a convergence rate much faster than earlier known adaptive implementations of the MMSE detector, viz. least mean square and recursive least squares. Simulation results are presented for synchronous single path channels as well as asynchronous multi-path channels, with multiple antennas employed at the receiver.

Index Terms—Direct-sequence code-division multiple-access (DS-CDMA), multiuser detection, minimum conditional probability of error (MCPOE), minimum joint probability of error (MJPOE), minimum mean squared error (MMSE), minimum probability of error (MPOE), space–time processing.

I. INTRODUCTION

DIRECT-SEQUENCE code-division multiple-access (DS-CDMA) is widely used for multiplexing users in a wireless scenario [1]. However, its performance is limited by multiple-access interference (MAI) and multipath channel distortion. The conventional DS-CDMA matched filter detector fails to combat these problems. Many advanced signal processing techniques have been proposed to enhance the performance of DS-CDMA systems, and these techniques fall into two broad categories: multiuser detection [2]–[4] and space–time processing [5], [6]. The former exploits the underlying statistical structure of MAI for interference cancellation, while the latter employs an antenna array at the receiver to optimally combine the different multipath signals of users. Combined multiuser detection and space–time processing has also been addressed in literature. The initial focus was on systems with a single transmit antenna and multiple receive antennas [7], [8]. Recently, much research has focused on systems with multiple transmit and receive antennas, employing space–time coding at the transmitter to achieve higher diversity gain [9], [10]. Blind multiuser detection for systems employing Alamouti space–time block codes at the transmitter has been investigated in [11].

Adaptive multiuser detectors are especially attractive because they can potentially adapt to unknown and time varying channel parameters [12]. Amongst these, blind adaptive multiuser detectors have the advantage that they eliminate the need for a training sequence in adaptation mode, which translates to savings in bandwidth [13]–[15]. Although not many wireless standards today use blind algorithms, it shall definitely be an advantage, in terms of lesser bandwidth consumption, if we can develop blind multiuser detection algorithms which have a performance comparable to training sequence-based algorithms. A space–time multiuser detector exploits the signal structure in both time domain and spatial domain for interference cancellation. Two adaptive space–time multiuser detectors based on the criterion of minimum probability of error (MPOE) are proposed in this paper, for multipath asynchronous DS-CDMA channels, with multiple antennas at the receiver.

The first algorithm, minimum joint probability of error (MJPOE), minimizes the joint probability of error for all users. The algorithm is blind, i.e., no training sequence is required in adaptation mode. MJPOE offers a bit-error-rate (BER) performance better than the nonadaptive minimum mean squared error (MMSE) detector [16]–[18]. It, however, has a computational complexity which is exponential in the number of users. A scheme based on the Gram–Schmidt orthogonalization procedure is proposed to reduce the computational burden of MJPOE. In order to further reduce the computational complexity, we propose the minimum conditional probability of error (MCPOE) algorithm, which minimizes the probability...
of error conditioned on the transmitted bit vector, for each user individually [19], [20]. It has a BER performance that is slightly inferior to the MMSE detector, however, at a convergence rate comparable to MJPOE and a computational complexity linear in the number of users. Both MJPOE and MCPOE have a convergence rate much faster than adaptive implementations. Both MJPOE and MCPOE have a convergence rate much faster than adaptive implementations of MMSE, viz. least mean square (LMS) and recursive least squares (RLS). In fact, MCPOE offers a convergence rate faster than RLS with computational complexity comparable to LMS. We also propose a blind MCPOE algorithm, whose performance is comparable to the training-based MCPOE algorithm.

The rest of this paper is organized as follows. Section II provides a mathematical description of an asynchronous multipath DS-CDMA channel with multiple antennas at the receiver. An expression for the received signal vector in terms of the system parameters is obtained. In Section III, the receiver structure is explained, probability distribution for the decision statistic vector is derived, and expressions for conditional and joint probability of error are also derived. The adaptive algorithms MJPOE and MCPOE are presented in Section IV. Adaptation equations for filter weights based on gradient descent are derived in this section. Section V presents simulation results, and comparisons of MJPOE and MCPOE with other multiuser detection algorithms proposed in literature. The paper concludes in Section VI.

II. SIGNAL MODEL

Consider a DS-CDMA channel with \( K \) users sharing the same bandwidth. A schematic of the channel model described in this section is depicted in Fig. 1. The signaling interval of each user is \( T \) seconds, and the input alphabet is antipodal binary: \( \{ -1, 1 \} \). During the \( i \)th signaling interval, the input bit vector\(^1\) \( \mathbf{x}_i = [b_{k1}(i), \ldots, b_{kK}(i)]^T \), where \( b_{ki}(i) \) denotes the \( i \)th input symbol of the \( k \)th user and \( (\cdot)^T \) denotes the transpose. User \( k \) is assigned a spreading waveform \( c_k(t) \) which is supported on \([0, T]\) and is normalized to one. Let \( \mathbf{s}_k = [s_{k1}, \ldots, s_{kN}]^T \) (\( N \times 1 \) vector) denote the corresponding spreading chip sequence, so that the spreading waveform can be expressed as

\[
c_k(t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} s_{kn} \text{rect}(t - (n - 1)T_c)
\]

where \( \text{rect}(t) \) is a rectangular waveform with unit amplitude in \([0, T_c]\), and \( s_{kn} \in \{-1, 1\} (1 \leq k \leq K, 1 \leq n \leq N) \). The transmission amplitude of the \( k \)th user is denoted by \( A_k \), which is real. The processing gain \( N \) is defined as the number of chips in a signaling period. The chip period \( T_c \) is, thus, equal to \( T/N \). The baseband signal of the \( k \)th user in the \( i \)th bit interval can now be expressed as

\[
x_k(t) = A_k b_k(i) c_k(t - iT), \quad iT \leq t < (i + 1)T.
\]

At the receiver, an array of \( P \) elements is employed. Assuming that each transmitter is equipped with a single antenna, the baseband signal between the \( k \)th user’s transmitter and the base station receiver can be modeled as a single-input multiple-output channel with the impulse response

\[
h_k(t) = \sum_{m=1}^{M} a_{km} g_{km} \delta(t - \tau_{km})
\]

where \( M \) is number of multipaths in each user’s channel, \( g_{km} \) and \( \tau_{km} \) are, respectively, the complex gain and delay of the \( m \)th multipath of the \( k \)th user’s signal, and \( \mathbf{a}_{km} = [g_{km1}, \ldots, g_{kmP}]^T \) is the array response vector corresponding to the \( m \)th path of the \( k \)th user’s signal. The total received signal \( \mathbf{r}(t) \) at the receiver is a superposition of the signal from the \( K \) users plus the additive noise given by

\[
\mathbf{r}(t) = \sum_{i} \sum_{k=1}^{K} x_k(t) \mathbf{h}_k(t) + \sigma \mathbf{n}(t)
\]

\[
= \sum_{i} \sum_{k=1}^{K} A_k b_k(i) \sum_{m=1}^{M} a_{km} g_{km} c_k(t - iT - \tau_{km}) + \sigma \mathbf{n}(t)
\]
where \( \star \) denotes the convolution operation, \( \mathbf{r}(i) = [r_1(i), \ldots, r_P(i)]^T \) is a \( P \times 1 \) vector of independent zero-mean complex Gaussian noise processes, each with unit variance, and \( \sigma^2 \) is the variance of the ambient noise at each antenna.

Let \( \mathbf{r}_p(i) \) denote the received signal vector at the \( p \)th antenna in the \( i \)th signaling interval, obtained by sampling \( \mathbf{r}_p(t) \) at chip rate. Since we have considered an asynchronous channel, \( \mathbf{r}_p(i) \) contains contributions from bits transmitted in the \((i - 1)\)th interval as well as the \( i \)th interval. Let us assume that the maximum multipath delay of any user does not exceed the symbol period, so that contributions from \((i - 2)\)th and all preceding signaling intervals are zero. Also, let us assume that multipath delays are resolvable up to the accuracy of a chip period, i.e., \( \tau_{kl} \) is an integral multiple of \( T_c/V_k \). Now define the following:

\[
\mathbf{s}_{kL}^{(n)} = [s_{kN-n+1}, \ldots, s_{kN}]^T \\
\mathbf{s}_{kR}^{(n)} = [0, \ldots, 0, s_{k1}, \ldots, s_{kN-n}]^T.
\]

(5)

(6)

Now, the received signal vector \( \mathbf{r}_p(i) \) can be expressed as

\[
\mathbf{r}_p(i) = \sum_{k=1}^{K} A_k \mathbf{b}_k(i - 1) + \sum_{m=1}^{M} \mathbf{a}_{kmq} \mathbf{g}_{km} \mathbf{s}_{kL}^{(n_{km})} + \sum_{m=1}^{M} \mathbf{a}_{kmq} \mathbf{g}_{km} \mathbf{s}_{kR}^{(n_{km})} + \sigma \mathbf{m}_p(i)
\]

(7)

where \( n_{km} = \tau_{km}/T_c \) and \( \mathbf{m}_p(i) \) denotes the sampled noise vector, obtained by chip rate sampling of the received signal in the \( i \)th bit interval. Since the multipath delays are assumed to be multiples of the chip period, \( n_{km} \) is an integer. In (7), the first term represents contribution from the previous bit interval or intersymbol interference, the second term represents contribution of the bits transmitted in the \( i \)th bit interval, whereas the last term represents additive white Gaussian noise. For ease of representation, define the following:

\[
\alpha_{kp,L} = \sum_{m=1}^{M} \mathbf{a}_{kmq} \mathbf{g}_{km} \mathbf{s}_{kL}^{(n_{km})}
\]

(8)

\[
\alpha_{kp,R} = \sum_{m=1}^{M} \mathbf{a}_{kmq} \mathbf{g}_{km} \mathbf{s}_{kR}^{(n_{km})}.
\]

(9)

Thus, \( \mathbf{r}_p(i) \) can be expressed as

\[
\mathbf{r}_p(i) = \sum_{k=1}^{K} A_k \mathbf{b}_k(i - 1) \alpha_{kp,L} + \sum_{k=1}^{K} A_k \mathbf{b}_k(i) \alpha_{kp,R} + \sigma \mathbf{m}_p(i).
\]

(10)

This can further be expressed using matrix notation as

\[
\mathbf{r}_p(i) = \mathbf{S}_{PL} \mathbf{A} \mathbf{b}(i - 1) + \mathbf{S}_{PR} \mathbf{A} \mathbf{b}(i) + \sigma \mathbf{m}_p(i)
\]

(11)

where \( \mathbf{S}_{PL} = [\alpha_{1PL}, \ldots, \alpha_{KPL}] \) \((N \times K)\) matrix, \( \mathbf{S}_{PR} = [\alpha_{1PR}, \ldots, \alpha_{KPR}] \) \((N \times K)\) matrix, and \( \mathbf{A} = \text{diag}(A_1, \ldots, A_K) \) \((K \times K)\) matrix.

III. PROBABILITY OF ERROR

A. Probability Density of the Decision Statistic

Let \( \mathbf{W}_k \) \((NP \times 1)\) vector denote the linear filter used to demodulate the bits transmitted by the \( k \)th user. \( \mathbf{W}_k \) operates on the augmented \( NP \times 1 \) signal vector \( \mathbf{r}(i) = [\mathbf{r}_1(i)^T, \ldots, \mathbf{r}_P(i)^T]^T \). Let \( \mathbf{w}_k \) be represented as \( \mathbf{w}_k = [\mathbf{w}_{k1}^T, \ldots, \mathbf{w}_{kP}^T]^T \), where each \( \mathbf{w}_{kp} \) \((1 \leq p \leq P)\) is an \( N \times 1 \) vector. The soft output of the filter \( \mathbf{y}_k(i) \) in the \( i \)th bit interval is given by

\[
\mathbf{y}_k(i) = \sum_{p=1}^{P} \mathbf{w}_{kp}^H \mathbf{r}(i)
\]

(12)

where \( (\cdot)^H \) denotes the Hermitian and \( \mathbf{u}(i) = [\mathbf{u}_1(i)^T, \ldots, \mathbf{u}_P(i)^T]^T \). The bit decision for the \( k \)th user is given by

\[
\hat{b}_k(i) = \text{sgn}[\Re(\mathbf{y}_k(i))]
\]

(13)

where \( \text{sgn}[\cdot] \) denotes the signum function and \( \Re(\cdot) \) denotes the real part.

For notational ease, we can drop the index \( i \) and represent \( \mathbf{b}(i) \) as \( \mathbf{b} \), \( \mathbf{y}_k(i) \) as \( \mathbf{y}_k \), and \( \mathbf{n}(i) \) as \( \mathbf{n} \). Thus

\[
\mathbf{y}_k = \sum_{p=1}^{P} \mathbf{w}_{kp}^H \mathbf{S}_{PL} \mathbf{A} \mathbf{b}(i - 1) + \sum_{p=1}^{P} \mathbf{w}_{kp}^H \mathbf{S}_{PL} \mathbf{A} \mathbf{b}(i) + \sigma \mathbf{w}_{kp}^H \mathbf{n}
\]

(14)

where

\[
\zeta_k = \sum_{p=1}^{P} \mathbf{w}_{kp}^H \mathbf{S}_{PL} \mathbf{A}
\]

(15)

represents the contribution of the \((i - 1)\)th bit to the decision statistic. It must be noted here that \( \zeta_k \) depends on \( \mathbf{b}(i - 1) \), which is not known to the receiver. However, since the receiver has estimated \( \mathbf{b}(i - 1) \) in the \((i - 1)\)th bit interval, we can form an estimate of \( \zeta_k \), denoted by \( \hat{\zeta}_k \) and given by

\[
\hat{\zeta}_k = \sum_{p=1}^{P} \mathbf{w}_{kp}^H \mathbf{S}_{PL} \mathbf{A}(i - 1)
\]

(16)

\[
+ \sum_{p=1}^{P} \mathbf{w}_{kp}^H \mathbf{S}_{PL} \mathbf{A} + \sigma \mathbf{w}_{kp}^H \mathbf{n}
\]
where \( \hat{\mathbf{b}}(i - 1) \) denotes the estimate of \( \mathbf{b}(i - 1) \) at the receiver, and
\[
\hat{\gamma}_k = \sum_{j=1}^{P} \mathbf{w}_{kj}^H \mathbf{S}_{pt} \mathbf{A} \hat{\mathbf{b}}(i - 1). \tag{17}
\]

Although, all subsequent results are derived in terms of \( \mathbf{y}_k \), it must be noted that in practice the receiver would use \( \hat{\mathbf{y}}_k \) for all its computations, because \( \hat{\mathbf{b}}(i - 1) \) is unknown at the receiver. The simulation results presented in Section V are also based on the estimate of \( \mathbf{y}_k \) given by (16).

Let \( \mathbf{y} = [\mathbf{y}_1, \ldots, \mathbf{y}_K]^T \) \((K \times 1)\) vector. Since the bit decisions depend on the real part of \( \mathbf{y} \), we are interested in the probability distribution of the decision statistic vector \( \Re(\mathbf{y}) \). Conditioned on vector \( \mathbf{b} \), \( \Re(\mathbf{y}) \) is Gaussian with mean given by
\[
\mu_k = \Re\left( \hat{\gamma}_k + \sum_{j=1}^{P} \mathbf{w}_{kj}^H \mathbf{S}_{pt} \mathbf{A} \hat{\mathbf{b}} \right), \tag{18}
\]
which is Gaussian with mean given by
\[
\mathbf{C}_{kl} = \text{ Cov}(\mathbf{y}_k, \mathbf{y}_l) = \left\{ \begin{array}{ll}
\rho_{kl} \sigma_{\mathbf{y}_k} \sigma_{\mathbf{y}_l} & \text{if } k = l \\
\sigma^2 & \text{if } k \neq l
\end{array} \right. \tag{19}
\]
where \( \sigma^2 = \text{Var}(\mathbf{y}) \) and \( \rho_{kl} = \left( \Re(\mathbf{w}_l^H \mathbf{w}_l) / (||\mathbf{w}_l||) \right) \).

### B. Joint Probability of Error

Let \( \hat{\mathbf{b}} = [\hat{b}_1, \ldots, \hat{b}_K]^T \) denote the estimated bit vector, where \( \hat{b}_k \), \( 1 \leq k \leq K \), is given by (13). Assuming all bit vectors are equally likely (with probability \((1/2^K)\)), the joint probability of error \( P_E \) can be expressed as
\[
P_E = \sum_{\mathbf{b}} \Pr(\hat{\mathbf{b}} \neq \mathbf{b} | \mathbf{b}) \Pr(\mathbf{b})
= \sum_{\mathbf{b}} \left[ 1 - \Pr(\hat{\mathbf{b}} = \mathbf{b} | \mathbf{b}) \right] \frac{1}{2^K}
= 1 - \frac{1}{2^K} \sum_{\mathbf{b}} \Pr(\hat{\mathbf{b}} = \mathbf{b} | \mathbf{b}) \tag{20}
\]
where the summation is over all possible bit vectors of length \( K \), with entries \( \pm 1 \). Since \( \Re(\mathbf{y}) \) conditioned on vector \( \mathbf{b} \) is Gaussian with mean \( \mathbf{\mu} \) and covariance matrix \( \mathbf{C} \), its probability density function is given by
\[
f_{\Re(\mathbf{y})}(\mathbf{x} | \mathbf{b}) = \frac{1}{\sqrt{(2\pi)^K |\mathbf{C}|}} \exp\left( -\frac{1}{2} (\mathbf{x} - \mathbf{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{\mu}) \right) \mathbf{x} \in (-\infty, \infty)^K \tag{21}
\]
where \( |\mathbf{C}| \) denotes the determinant of \( \mathbf{C} \).

The conditional probability of correct demodulation \( \Pr(\hat{\mathbf{b}} = \mathbf{b} | \mathbf{b}) \) can be expressed as the integral of the multivariate Gaussian distribution given by (21), with appropriately chosen integration limits. For instance, if \( \mathbf{b} = [1, \ldots, 1]^T \), \( \Pr(\hat{\mathbf{b}} = \mathbf{b} | \mathbf{b}) \) is given by
\[
\Pr(\hat{\mathbf{b}} = \mathbf{b} | \mathbf{b}) = \frac{1}{\sqrt{(2\pi)^K |\mathbf{C}|}} \int_0^{\infty} \cdots \int_0^{\infty} \exp\left( -\frac{1}{2} (\mathbf{x} - \mathbf{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{\mu}) \right) dx_1 \cdots dx_K. \tag{22}
\]
The joint probability of error \( P_E \) can be expressed as a weighted sum of \( 2^K \) integrals of the above form. If all bit vectors are equally likely to be transmitted, then each of the weights is \((1/2^K)\). Thus, using (20) and (22), \( P_E \) is given by
\[
P_E = 1 - \frac{1}{2^K} \sum_{\mathbf{b}} \int_0^{\infty} \cdots \int_0^{\infty} \exp\left( -\frac{1}{2} (\mathbf{x} - \mathbf{\mu})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{\mu}) \right) dx_1 \cdots dx_K \tag{23}
\]
where \( l_k = 0, u_k = \infty \), if \( l_k = 1 \) and \( l_k = -\infty, u_k = 0 \), if \( l_k = -1 \).

Now consider the situation when \( \rho_{kl} = \epsilon_k \epsilon_l \), where \( \epsilon_k \) and \( \epsilon_l \) are arbitrarily chosen real numbers at this moment and will be obtained as outputs of the adaptive algorithm, MJPOE, to be described in a subsequent section. Since the random variable \( \Re(\mathbf{y}_k) \) conditioned on \( \mathbf{b} \) is \( N(\mu_k, \sigma^2_k) \), it can be expressed as
\[
\Re(\mathbf{y}_k) = \mu_k + \epsilon_k V + \frac{1 - \epsilon_k^2}{\sigma_k^2} U_k, \quad k = 1, \ldots, K \tag{24}
\]
where \( U_1, \ldots, U_K \), and \( V \) are independent identically distributed \( \mathcal{N}(0, 1) \) random variables.

Next, consider the following joint probability \( P_J \):
\[
P_J = \Pr(\delta_{L_1} < \Re(\mathbf{y}_1) \leq \delta_{U_1}, \ldots, \delta_{L_K} < \Re(\mathbf{y}_K) \leq \delta_{U_K}). \tag{25}
\]
From (24) we have
\[
U_k = \frac{\Re(\mathbf{y}_k) - \mu_k - \epsilon_k V}{\epsilon_k \sigma_k}, \tag{26}
\]
Substituting (26) in (25), \( P_J \) can be expressed as
\[
P_J = \Pr\left\{ \frac{\delta_{U_k} - \mu_k - \epsilon_k V}{\sqrt{1 - \epsilon_k^2}} < U_k < \frac{\delta_{L_k} - \mu_k - \epsilon_k V}{\sqrt{1 - \epsilon_k^2}} \right\}. \tag{27}
\]
Since \( U_k \)'s and \( V \) are \( \mathcal{N}(0, 1) \), (27) can be simplified to yield
\[
P_J = \int_{-\infty}^{\infty} \prod_{k=1}^{K} \left[ Q\left( \frac{\delta_{L_k} - \mu_k - \epsilon_k \sigma_k V}{\sigma_k \sqrt{1 - \epsilon_k^2}} \right) - Q\left( \frac{\delta_{U_k} - \mu_k - \epsilon_k \sigma_k V}{\sigma_k \sqrt{1 - \epsilon_k^2}} \right) \right] \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \tag{28}
\]
where \( Q(x) \) is given by
\[
Q(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left( -\frac{t^2}{2} \right) dt. \tag{29}
\]
1) Illustration: For the purpose of illustration, consider the two user case \((K = 2)\). \{\delta_{L_2} \}_{k=1}^2 \text{ and } \{\delta_{U_2} \}_{k=1}^2 \text{ for different transmitted bit vectors } \mathbf{b} = [b_1, b_2]^T \text{ are shown in Table I.}
TABLE I
\( \delta_{L_k} \) and \( \delta_{U_k} \) for the Two-User Case for Different Transmitted Bit Vectors

<table>
<thead>
<tr>
<th>( b_1 )</th>
<th>( b_2 )</th>
<th>( \delta_{L_1} )</th>
<th>( \delta_{U_1} )</th>
<th>( \delta_{L_2} )</th>
<th>( \delta_{U_2} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>( \infty )</td>
<td>0</td>
<td>( \infty )</td>
</tr>
<tr>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>( \infty )</td>
<td>-1</td>
<td>0</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \infty )</td>
</tr>
<tr>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>( \infty )</td>
<td>( \infty )</td>
</tr>
</tbody>
</table>

The joint probability \( P_J \), evaluated by appropriately substituting \( \delta_{L_k} \) and \( \delta_{U_k} \) in (28), gives the probability of correct demodulation conditioned on \( b \), i.e., \( \Pr(\tilde{b} = b | b) \). For instance, in the two-user case, the conditional probability of correct demodulation given bit vector \( b = [1, 1]^T \) is transmitted is given by

\[
P_J(\delta_{L_1} = 0, \delta_{U_1} = \infty)^2 = \int_{-\infty}^{\infty} \left( -\frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x - \mu_1)^2}{2\sigma_1^2}} \right) \cdot \exp \left( \frac{x - \mu_2}{\sigma_2 \sqrt{2\pi}} \right) Q \left( \frac{x - \mu_2}{\sigma_2 \sqrt{2\pi}} \right) \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx 
\]

where the subscript of \( P_J \) denotes the values of \( \delta_{L_k} \) and \( \delta_{U_k} \) at which the expression for \( P_J \) given by (28) is evaluated, and

\[
\mu_{11}[1,1]^T = \Re \left( \zeta_1 + A_1 \sum_{p=1}^{P} \sum_{m=1}^{M} a_{1m,p}g_{1m,s_{11}^{(1)}} + A_2 \sum_{p=1}^{P} \sum_{m=1}^{M} a_{2m,p}g_{2m,s_{21}^{(1)}} \right) 
\]

\[
\mu_{21}[1,1]^T = \Re \left( \zeta_2 + A_1 \sum_{p=1}^{P} \sum_{m=1}^{M} a_{1m,p}g_{1m,s_{12}^{(1)}} + A_2 \sum_{p=1}^{P} \sum_{m=1}^{M} a_{2m,p}g_{2m,s_{22}^{(1)}} \right) 
\]

where \( s_{11}^{(1)} \) and \( s_{21}^{(1)} \) are given by (5) and (6), respectively. Expressions similar to (30) can be obtained for different transmitted bit vectors, and be summed up (with appropriate weights) to obtain the joint probability of error (20).

C. Conditional Probability of Error

The motivation for deriving an expression for the conditional probability of error will become clearer in Section IV-B. Let \( P_k | b \) denote the probability of error in demodulation of the bits transmitted by the \( k \)th user, conditioned on the transmitted bit vector \( b \). Assuming that bits +1 and -1 are transmitted with equal probability, \( P_k | b \) can be expressed as

\[
P_k | b = \frac{1}{2} \Pr \left( \Re(y_k) < 0 \mid b^{(k-)}, b_k = +1 \right) + \frac{1}{2} \Pr \left( \Re(y_k) \geq 0 \mid b^{(k-)}, b_k = -1 \right) 
\]

THE NATURE OF A TWO-USER CASE FOR DIFFERENT TRANSMITTED BIT VECTORS

where \( b^{(k-)} = [b_1, \ldots, b_{k-1}, b_{k+1}, \ldots, b_K]^T \). Since \( \Re(y_k) \) conditioned on \( b \) is Gaussian with mean \( \mu_k \) (18) and variance \( \sigma_k^2 \) (19), its probability density function is given by

\[
f_{\Re(y_k)}(x | b) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp \left( \frac{- (x - \mu_k)^2}{2\sigma_k^2} \right) 
\]

The conditioning on \( b \) is implicit in \( \mu_k \). Let \( \mu_k \) be represented as \( \mu_k|1 \) when \( b_k = +1 \) and as \( \mu_k|-1 \) when \( b_k = -1 \). We have

\[
\mu_k|1 = \Re \left( \zeta_k + \sum_{i=1}^{K} A_{di} \sum_{p=1}^{P} \sum_{m=1}^{M} a_{im,p}g_{im,s_{i1}^{(m)}} \right) + A_k \sum_{p=1}^{P} \sum_{m=1}^{M} a_{km,p}g_{km,s_{12}^{(m)}} 
\]

and

\[
\mu_k|-1 = \Re \left( \zeta_k + \sum_{i=1}^{K} A_{di} \sum_{p=1}^{P} \sum_{m=1}^{M} a_{im,p}g_{im,s_{i2}^{(m)}} \right) + A_k \sum_{p=1}^{P} \sum_{m=1}^{M} a_{km,p}g_{km,s_{22}^{(m)}} 
\]

From (33), the conditional probability of error for user \( k \) can be expressed as

\[
P_k | b = \frac{1}{2} \int_{-\infty}^{0} \frac{1}{\sigma_k \sqrt{2\pi}} \exp \left( \frac{- (y - \mu_k|1)^2}{2\sigma_k^2} \right) dy 
\]

\[
+ \frac{1}{2} \int_{0}^{\infty} \frac{1}{\sigma_k \sqrt{2\pi}} \exp \left( \frac{- (y - \mu_k|-1)^2}{2\sigma_k^2} \right) dy 
\]

which can be simplified to obtain

\[
P_k | b = \frac{1}{2} Q \left( \frac{\mu_k|1}{\sigma_k} \right) + \frac{1}{2} Q \left( \frac{-\mu_k|-1}{\sigma_k} \right) 
\]

where \( Q(x) \) is given by (29).

IV. ADAPTIVE ALGORITHMS

A. Minimizing Joint Probability—MIPOE

The aim of the MIPOE-based detector is to minimize the joint probability of error expression (23) with respect to the filter weights \( w_1, \ldots, w_K \). The filter weights are iteratively adapted using a gradient descent technique, until the minimum is attained. The minimization problem is subject to the constraint

\[
\Re(\Re(w_{1l}) | w_{1l} | w_{1l}) = \epsilon_{k1} |1| |1| |1| (1 \leq k, l \leq K) 
\]

This constraint results from the definition of \( \rho_{kl} \) in Section III-A, and from the assumption that \( \rho_{kl} = \epsilon_{k1} \). The expression for joint probability of error is, however, complex and its constrained minimization with respect to the filter weights is a computationally intensive task.

In order to overcome the computational overhead in minimizing the joint probability of error, we have explored possibilities of approximations and constraints. We first carried out
A stochastic gradient descent approach can now be used to minimize the joint probability of error expression given by (40) with respect to the filter weights. Orthogonality of filter weights can be executed at each step of the iteration by using the Gram–Schmidt orthogonalization procedure. Let $\mathbf{w}^{(i)}_k$ denote the filter of the $k$th user at the $i$th iteration of the algorithm. In gradient descent, the filter weights are updates according to the rule

$$\mathbf{w}^{(i+1)}_k = \mathbf{w}^{(i)}_k - \lambda_k \frac{\partial P_E}{\partial \mathbf{w}_k} |_{\mathbf{w}_i = \mathbf{w}^{(i)}_i}$$  \hspace{1cm} (41)$$

where $\lambda_k$ is an appropriately chosen step-size parameter and in general could be different for different users. $\mathbf{w}^{(i)}_k$ denotes the filter obtained after Gram–Schmidt orthogonalization. It can be shown that the partial derivative of $P_E$ (40) with respect to filter $\mathbf{w}_k$ is given by

$$\frac{\partial P_E}{\partial \mathbf{w}_k} = - \frac{1}{2 \pi \cdot 2K} \sum_{\mathbf{v}_b} b_{k_b} e^{-\frac{||\mathbf{v}_b||^2}{2 \sigma_k^2}} \Phi_k \prod_{i \neq k} \bigg( - \frac{b_{k_i} \mu_k}{\sigma_i} \bigg)$$  \hspace{1cm} (42)$$

where $\Phi_k = [\phi_{k1}, \ldots, \phi_{kP}]^T$ is an $NP \times 1$ vector, such that $\phi_{kp} (1 \leq p \leq P)$ is an $N \times 1$ vector given by

$$\phi_{kp} = \frac{1}{\sigma_k} \left( \Re(S_{PLAb}(i-1)) + \Re(S_{PPAb}) - \frac{\mu_k \mathbf{w}_k}{\mathbf{w}_k^H \mathbf{w}_k} \right)$$  \hspace{1cm} (43)$$

After gradient descent adaptation, the filters of different users are made orthogonal to each other using the Gram–Schmidt procedure. The set of nonorthogonal vectors $\{\mathbf{w}_k\}_{k=1}^K$ is transformed into a set of orthogonal vectors $\{\mathbf{w}_k\}_{k=1}^K$ using the relation

$$\mathbf{w}_k = \mathbf{w}_k - \sum_{i=1}^{k-1} \mathcal{P}(\mathbf{w}_k | \mathbf{w}_i)$$  \hspace{1cm} (44)$$

where $\mathcal{P}(x|y)$ denotes the projection of $x$ on $y$ and is given by

$$\mathcal{P}(x|y) = \left( \frac{\langle x, y \rangle}{\|y\|^2} \right) y.$$  \hspace{1cm} (45)$$

1) Further Simplification: The expression for joint probability of error (40) is a sum of $2^K$ terms (each term corresponding to one bit vector of length $K$). A further reduction in computational complexity can be made by noting that a term corresponding to a particular bit vector in this sum remains unchanged if $\mathbf{b}$ is replaced by $-\mathbf{b}$, i.e., all the +1s are changed to –1s and vice-versa. Thus, instead of summing over $2^K$ bit vectors, we need to sum over only $2^{K-1}$ bit vectors, which implies a reduction in computational complexity by a factor of two.

B. Minimizing Conditional Probability—MCPOE

Although forcing the orthogonality of filters of different users leads to a reduction in the computational complexity of the algorithm, it still continues to be exponential in the number of users. The computational complexity can be further reduced, albeit at

extent simulation studies for computing MJPOE using gradient-based techniques. We noticed an interesting occurrence: At convergence, the filters of different users were nearly orthogonal to each other, i.e., $\mathbf{w}_k^H \mathbf{w}_l \approx 0$. This is depicted in Fig. 2. To a certain extent, this result is intuitively expected. This is because at convergence we want the filters $\mathbf{w}_k$ and $\mathbf{w}_l$ to be well separated, so that we can detect the bits of the $k$th user and the $l$th user. In case $\mathbf{w}_k$ and $\mathbf{w}_l$ are close to each other (almost colinear), then in the limiting case $\mathbf{w}_k = \gamma \mathbf{w}_l$ (where $\gamma$ is a scalar), both of them will detect the same user. The observation $\mathbf{w}_k^H \mathbf{w}_l = 0$ can be incorporated as a priori into the minimization problem, and this then leads to significant reduction in computational overhead.

Forcing filters of different users to be orthogonal is equivalent to making $\epsilon_k = 0$. It is worth mentioning here that in the nonorthogonal filter case, $\epsilon_k$ ($1 \leq k \leq K$), which were arbitrarily chosen when first defined in Section III-B, are obtained via simulation. However, in the orthogonal filter case, we force them to be zero. Now, from (19), the covariance between the decision statistics of user $k$ and user $l$ is given by $C_{kl} = \sigma^2 \mathbf{w}_k^H \mathbf{w}_l$. Since $\mathbf{w}_k^H \mathbf{w}_l = 0$, the entries of the decision statistic vector become uncorrelated. Thus, we have from (25)

$$P_J = \Pr(\delta_{L1} < \Re(y_1) \leq \delta_{U1}, \ldots, \delta_{LK} < \Re(y_K) \leq \delta_{UK})$$

$$= \prod_{k=1}^{K} \Pr(\delta_{Lk} < \Re(y_k) \leq \delta_{Uk}).$$  \hspace{1cm} (39)$$

Thus, using (39), the joint probability of correct demodulation can be expressed as a product of individual probabilities of correct demodulation. The joint probability of error is then given by

$$P_E = 1 - \frac{1}{2K} \sum_{\mathbf{v}_b} \prod_{k=1}^{K} Q \left( - \frac{b_{k_b} \mu_k}{\sigma_k} \right)$$  \hspace{1cm} (40)$$

where $\mu_k$ and $\sigma_k$ are given by (18) and (19). The expression can be easily obtained by substituting $\epsilon_k = 0$ in (28) and choosing $\delta_{Lk}$ and $\delta_{Uk}$ suitably.

Fig. 2. Correlation between filters of two different users for MJPOE; the correlation decays to zero.
the cost of some degradation in BER performance, by mini-
mizing the conditional probability of error for each user
individually, the expression for which was derived in Section III-C.
The adaptive algorithm, thus, obtained (MCPOE) has a con-
vergence rate comparable to MJPOE and a computational com-
plexity linear in the number of users. Of course, there is a minor
degradation in BER performance as compared to MJPOE.

The aim of the MCPOE algorithm is to minimize the con-
tingual probability of error, \( P_k|b \) (38), conditioned on trans-
mited bit vector \( b \) for each user individually. Here too a gra-
dient descent approach is adopted for adaptively computing the
optimal weights which minimize the conditional probability of
error. Define the \( NP \times 1 \) vectors \( \mathbf{u}_k = [u_{k1}^T \ldots u_{kP}^T]^T \) and
\( \mathbf{v}_k = [v_{k1}^T \ldots v_{kP}^T]^T \) such that \( u_{kp} \) and \( v_{kp} (1 \leq p \leq P) \) are
\( N \times 1 \) vectors and

\[
\mathbf{u}_p = \frac{\partial \mu_p |1}{\partial \mathbf{u}_p} \quad \mathbf{v}_p = \frac{\partial \mu_p |_{-1}}{\partial \mathbf{u}_p}, \quad p = 1, \ldots, P. \tag{46}
\]

Then it can be shown that

\[
\mathbf{u}_p = \Re \left( S_p L A (i-1) + \sum_{j \neq k}^K A_j b_j \sum_{m=1}^M a_{jm,p} g_{jm} s_{j,m}^{(n_p,m)} \right) 
+ A_k \sum_{m=1}^M a_{km,p} g_{km} s_{k,R}^{(n_p,m)} \right)
\]

\[
\mathbf{v}_p = \Re \left( S_p L A (i-1) + \sum_{j \neq k}^K A_j b_j \sum_{m=1}^M a_{jm,p} g_{jm} s_{j,m}^{(n_p,m)} \right) 
- A_k \sum_{m=1}^M a_{km,p} g_{km} s_{k,R}^{(n_p,m)} \right). \tag{47}
\]

Also, it can be shown that

\[
\frac{\partial}{\partial \mathbf{u}_k} \left( \frac{\mu_k |1}{\sigma_k} \right) = \frac{|\mathbf{u}_k|^2 \mathbf{u}_k - \mu_k |1 \mathbf{u}_k}{\sigma_k |\mathbf{u}_k|^3/2} \]
\[
\frac{\partial}{\partial \mathbf{u}_k} \left( \frac{\mu_k |_{-1}}{\sigma_k} \right) = \frac{|\mathbf{u}_k|^2 \mathbf{u}_k - \mu_k |_{-1} \mathbf{u}_k}{\sigma_k |\mathbf{u}_k|^3/2}. \tag{48}
\]

Thus, the derivative of the conditional probability of error of the
kth user \( P_k | b \) can be computed using (47), (48), and the fol-
lowing:

\[
\frac{\partial P_k | b}{\partial \mathbf{u}_k} = \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{\mu^2_k |1}{\sigma_k^2} \right) \frac{\partial}{\partial \mathbf{u}_k} \left( \frac{\mu_k |1}{\sigma_k} \right) 
- \frac{1}{\sqrt{2\pi}} \exp \left( -\frac{\mu^2_k |_{-1}}{\sigma_k^2} \right) \frac{\partial}{\partial \mathbf{u}_k} \left( \frac{\mu_k |_{-1}}{\sigma_k} \right). \tag{49}
\]

1) Blind Adaptation: The MCPOE algorithm can be modi-
ified such that the adaptation of filter weights is blind, i.e., the
need for a training sequence to be transmitted by each user
is eliminated. Instead of using training bits from each user in
adaptation mode, the output of a space–time matched filter can
be used. The probability of error is now conditioned on the
output of a space–time matched filter rather than on the received
bits. This eliminates the need of having each user to transmit a
training sequence. The space–time matched filter can be thought
of as a space–time counterpart of the code matched filter. Details
of its implementation can be found in [8]. The detector, thus, be-
comes blind in the sense that it does not require a known bit se-
quence for training; a knowledge of spreading codes of the users
is still essential. This modification does not have a significant
impact on the performance of MCPOE. The same modification,
however, when extended to LMS or RLS, leads to a substantial
degradation in performance. The claim is substantiated by re-
sults presented in the following section.

V. SIMULATION RESULTS

Simulation results are presented both for synchronous
DS-CDMA channels and asynchronous DS-CDMA channels
with multiple antennas at the receiver. BPSK modulation
was assumed, with the bits +1 and −1 being equiprobable.
m-sequences with spreading gain 15 were used as spreading
sequences. The channel model used here has been primarily
taken from [8]. The channel for each user was assumed to have
three multipaths (\( M = 3 \)). The receiver was assumed to consist
of a linear antenna array with three elements (\( P = 3 \)) and half
wavelength spacing, with response given by

\[
a_{km,p} = \exp[j(p-1)\pi \sin(\phi_{km})] \tag{50}
\]

where \( j = \sqrt{-1} \) and \( \phi_{km} \) is the direction of arrival (DOA) of the
kth user along the mth path with respect to the antenna array.
The multipath gains, DOA, and propagation delays were ran-
domly generated for all users and kept fixed for the simulations.
All users were assumed to have equal transmit power. However,
the received signal powers are unequal due to unequal strength
of the multipath gain for each user. This creates a near–far situa-
tion at the receiver.

Fig. 3 depicts the convergence curves for the MCPOE algo-
rithm for different SNR levels, for an asynchronous multipath
Fig. 4. Convergence curves for the MJPOE algorithm for an asynchronous multipath channel, for two different SNRs.

Fig. 5. Convergence curves for LMS and RLS for an asynchronous multipath channel. The filter weights converge to a steady value in 10–15 iterations. Fig. 4 depicts the corresponding convergence plots for the MJPOE algorithm. The algorithm converges in 15–20 iterations. Convergence plots for training-based LMS and RLS [22] for a multipath channel are shown in Fig. 5 for comparison. The LMS algorithm converges in 300 iterations, whereas, the RLS algorithm requires roughly 50 iterations for convergence. Thus, the convergence rates of both MCPOE and MJPOE are significantly faster than other well known adaptive algorithms such as LMS and RLS.

Fig. 6. Comparison of BER performance of MCPOE, MJPOE, and MMSE for different SNRs.

Fig. 7. Convergence curves for training-based MCPOE and blind MCPOE because it accounts for the nonzero cross correlation between the decision statistics of various users by minimizing the joint probability of error.

Fig. 2 represents the cross correlation between the filters of two different users in MJPOE when Gram–Schmidt orthogonalization is not used, i.e., the filters of different users are assumed to be nonorthogonal in general. It is seen that the cross correlation dies down to zero as the filter weights attain a steady state, providing empirical proof for the assumption made in Section IV-A ($\epsilon_k = \theta_k$).

Fig. 7 depicts the comparison between training-based MCPOE and blind MCPOE (using a matched filter front end). It is seen that there is no substantial change in performance even if the output of a matched filter instead of a training sequence in adaptation mode. However, as shown in Fig. 8, the same modification when applied to the RLS adaptive algorithm yields poor results. The mean squared error at convergence for “blind RLS” is much higher than that for training-based RLS.
VI. CONCLUSION

Two new adaptive multiuser detectors, based on the criterion of MPOE, have been proposed in this paper. The former, MJPOE, aims to minimize the joint probability of error for all users simultaneously, while the latter, MCPOE, aims to minimize the probability of error conditioned on the transmitted bit vector for each user individually. Both detectors are comprised of a linear filter bank at the receiver, the weights of which are adapted using a stochastic gradient descent technique. The MJPOE algorithm is blind, whereas, both blind and training-based implementations are proposed for MCPOE. MJPOE offers a steady state BER performance better than the MMSE algorithm, although at a high computational complexity. A method to reduce the computational complexity of MJPOE, based on the Gram–Schmidt orthogonalization procedure is suggested. The BER performance of MCPOE is inferior to the MMSE algorithm but it has a computational complexity comparable to the LMS detector (linear in \( N \)). The convergence rates of both MCPOE and MJPOE are significantly faster than well-known adaptive algorithms such as LMS and RLS. Thus, the MCPOE algorithm has advantages of a fast convergence rate and low computational complexity, but the disadvantage of having a relatively poor BER performance. The MJPOE algorithm, on the other hand, has the advantages of a fast convergence rate and a BER performance better than the MMSE detector, but at the disadvantage of being computationally intensive. Ideally, one would like to achieve MJPOE-like performance at MCPOE-like complexity. Comparative results are summarized in Table II.

![Convergence curves for training-based RLS and blind RLS.](image)

**Fig. 8.** Convergence curves for training-based RLS and blind RLS.

**TABLE II**

<table>
<thead>
<tr>
<th>Comparison Metric</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steady state BER</td>
<td>MJPOE &lt; non-adaptive MMSE &lt; training based MCPOE &lt; blind MCPOE</td>
</tr>
<tr>
<td>Convergence rate</td>
<td>MJPOE \approx MCPOE (both blind and training based) &gt; RLS &gt; LMS</td>
</tr>
<tr>
<td>Computational complexity</td>
<td>MJPOE &gt; RLS &gt; MCPOE (both blind and training based) &gt; LMS</td>
</tr>
</tbody>
</table>

REFERENCES

Uday B. Desai (S’75–M’78–S’79–M’79–SM’96) received the B.Tech. degree from Indian Institute of Technology, Kanpur, India, in 1974, the M.S. degree from the State University of New York, Buffalo, in 1976, and the Ph.D. degree from The Johns Hopkins University, Baltimore, MD, in 1979, all in electrical engineering.

From 1979 to 1984 he was an Assistant Professor in the Electrical Engineering Department at Washington State University, Pullman, WA, and an Associate Professor at the university from 1984 to 1987. Since 1987, he has been a Professor in the Electrical Engineering Department at the Indian Institute of Technology—Bombay, Mumbai, India. He has held Visiting Associate Professor positions at Arizona State University, Tempe, Purdue University, West Lafayette, IN, and Stanford University, Stanford, CA. He was a Visiting Professor at EPFL, Lausanne during summer of 2002. From July 2002, he has been on leave from IIT-Bombay, and is the Director of HP-IITM Joint Laboratory at IIT-Madras. His research interest is in statistical signal processing as applied to wireless and wire-line communication. In particular, the focus has been on blind multiuser detection in a space–time framework. He is also interested in multimedia, image and video processing, artificial neural networks, computer vision, wavelet analysis. He is also interested in connectivity, particularly for rural India—the objective being to bring the advantages of modern day telecommunication to the Indian masses. He is the Editor of the book Modeling and Applications of Stochastic Processes (Norwell, MA: Kluwer, 1986). He is also a coauthor to the book A Bayesian Approach to Image Interpretation (Norwell, MA: Kluwer, 2001).

Dr. Desai is a Fellow of the Indian National Academy of Engineering (INAE) and a Fellow of The Institution of Electronic and Telecommunication Engineers (IETE). He is recipient of the Ram Lal Wadhwa Award from the Indian Institute of Electronics and Telecommunication Engineers for 2001. He was an Associate Editor of IEEE TRANSACTIONS ON IMAGE PROCESSING form January 1999 to December 2001. He is a Vice-President Indian Unit for Pattern Recognition and Artificial Intelligence.

Ranjan K. Mallik (S’88–M’93–SM’02) received the B.Tech. degree from the Indian Institute of Technology, Kanpur, India, in 1987 and the M.S. and Ph.D. degrees from the University of Southern California, Los Angeles, in 1988 and 1992, respectively, all in electrical engineering.

From August 1992 to November 1994, he was a Scientist at the Defence Electronics Research Laboratory, Hyderabad, India, working on missile and EW projects. From November 1994 to January 1996, he was a Faculty Member of the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, India. In January 1996, he joined the faculty of the Department of Electronics and Communication Engineering, Indian Institute of Technology, Guwahati, India, where he worked till December 1998. Since December 1998, he has been with the Department of Electrical Engineering, Indian Institute of Technology, Delhi, India, where he is an Associate Professor. His research interests are in communication theory and systems, difference equations, and linear algebra.

Dr. Mallik is a Member of Eta Kappa Nu. He is also a Member of the IEEE Communications and Information Theory Societies, the American Mathematical Society, and the International Linear Algebra Society, a Fellow of The Institution of Electronics and Telecommunication Engineers, a Life Member of the Indian Society for Technical Education, and an Associate Member of The Institution of Engineers (India).