Blind Adaptive Multiuser Detection

Michael Honig, *Senior Member, IEEE*, Upamanyu Madhow, *Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

Abstract—**The decorrelating detector and the linear minimum mean-square error (MMSE) detector are known to be effective strategies to counter the presence of multiuser interference in code-division multiple-access channels; in particular, those multiuser detectors provide optimum near–far resistance. When training data sequences are available, the MMSE multiuser detector can be implemented adaptively without knowledge of signature waveforms or received amplitudes. This paper introduces an adaptive multiuser detector which converges (for any initialization) to the MMSE detector without requiring training sequences. This *blind* multiuser detector requires no more knowledge than does the conventional single-user receiver: the desired user's signature waveform and its timing. The proposed blind multiuser detector is made robust with respect to imprecise knowledge of the received signature waveform of the user of interest.**

Index Terms— **Multiuser detection, multiple-access channels, code-division multiple access, blind equalization, minimum mean-square error detection.**

I. Introduction

MULTIUSER detection deals with the demodulation of digitally modulated signals in the presence of multiaccess interference. Multiuser detection finds its major application in Code-Division Multiple-Access (CDMA) receiver design. A major technological hurdle of CDMA systems is the near–far problem: the bit-error-rate of the conventional receiver is so sensitive to differences between the received energies of the desired user and interfering users that reliable demodulation is impossible unless stringent power control is exercised. The optimum multiuser detector for asynchronous multiple-access Gaussian channels was obtained in [1] where it was shown that the near–far problem suffered by the conventional CDMA receiver (a matched filter for the user of interest) is overcome by a more sophisticated receiver which accounts for the presence of other interferers in the channel. This receiver was shown ([1] and [2]) to attain essentially single-user performance assuming that the receiver knows (or can acquire) the following.

1) The signature waveform of the desired user.
2) The signature waveforms of the interfering users.
3) The timing (bit-epoch and carrier phase) of the desired user.
4) The timing (bit-epoch and carrier phase) of each of the interfering users.
5) The received amplitudes of the interfering users (relative to that of the desired user).

The conventional receiver only requires 1) and 3), but it is severely limited by the near–far problem, and even in the presence of perfect power control, its bit-error-rate is orders of magnitude far from optimal. The decorrelating detector of [3] and [4] showed that a linear receiver (modified matched filter orthogonal to the multiaccess interference) is sufficient in order to achieve optimum resistance against the near–far problem (for high signal-to-background-noise ratios). At the expense of a (generally) slight increase over the minimum bit-error-rate, the decorrelating detector avoids the exponential complexity in the number of active users of the optimum multiuser detector. Moreover, it does not require knowledge of 5). Considerable work has been done in the last few years on other multiuser detectors not only for coherent detection in the Gaussian channel, but for noncoherent demodulation and for fading and multipath channels as well. We refer the reader to [5] for a tutorial survey.

Some attention has been focused recently on *adaptive multiuser detection* which eliminates the need to know the signature waveforms of the interferers 2), timing 4), and amplitudes 5). Note that even in systems where this knowledge is available (as in the case of a centralized multiuser receiver which demodulates every active user), it is usually computationally intensive to incorporate that knowledge into the receiver parameters, so an adaptive algorithm can be an attractive alternative even in such a situation. The adaptive multiuser detectors in [6]–[9] are based on the minimization of mean-square-error (MMSE) between the outputs and the data. For a survey of adaptive multiuser detection see [10]. The decorrelating detector (which can be seen as the conceptual counterpart to the zero-forcing equalizer in single-user demodulation of signals subject to intersymbol interference) can be considered an asymptotic form of the MMSE detector as the background noise level goes to zero [6], [11]. Both detectors exhibit the same near–far resistance, which is defined as the worst case asymptotic efficiency (slope of bit-error-rate curve at high SNR [5]) over all values of the interfering-to-desired user energies. However, the MMSE detector lends itself to adaptive implementation more readily than the decorrelating detector.

The adaptive MMSE detectors proposed recently in [6]–[8] substitute the need to know 2), 4), and 5) by the need to have

6) Training data sequences for every active user.

Manuscript received July 14, 1994; revised February 4, 1995. This work was supported by Bellcore and by the U.S. Army Research Office under Grant DAAH04-93-G-0219. The material in this paper was presented in part at the 1994 Globecom Conference, San Francisco, CA, November 30–December 2, 1994.

M. L. Honig is with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60208 USA.

U. Madhow is with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA.

S. Verdú is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA.

IEEE Log Number 9412246.

The typical operation of those adaptive multiuser detectors requires each transmitter to send a training sequence at start-up which the receiver uses for initial adaptation. After the training phase ends, adaptation during actual data transmission occurs in decision-directed mode. However, any time there is a drastic change in the interference environment (e.g., a deep fade or the powering on of a strong interferer) decision-directed adaptation becomes unreliable, and data transmission (of the desired user) must be temporarily suspended and yield to a fresh training sequence. Thus the reliance on training sequences is cumbersome in most CDMA systems, where one of the most important advantages is the ability to have completely asynchronous and uncoordinated transmissions that switch on and off autonomously.

The foregoing observations imply that the need for *blind* adaptive receivers is even more evident in multiaccess channels than in single-user channels subject to intersymbol interference. The goal of this paper is to obtain an adaptive receiver, which does not require training sequences and requires knowledge of only 1) and 3), that is, the same knowledge as the conventional receiver.

Note that it is possible to pose a problem incorporating no *a priori* information: demodulate all the active users signals without knowledge of any of the signature waveforms or training sequences. Eavesdropping is one of the main applications of such a problem, and in this context it is worth mentioning a generalization of the Sato blind equalizer to multidimensional systems [12]. In addition to spurious local minima, a penalty one would be expected to pay for not incorporating knowledge of the desired signature waveform is that in near–far situations the accuracy with which weak users are demodulated is much lower than that corresponding to the strong users.

The blind multiuser detector derived in this paper is reminiscent of the philosophy of (single-user) anchored minimum-energy adaptive equalization proposed in [13]. That equalizer overcomes some of the ill-convergence problems suffered by conventional Godard-type blind equalizers (see [14] for a survey) by using a very simple cost function: output energy. That cost function cannot be used with conventional equalizers, where all the taps are adjustable (or floating). The anchored equalizer maintains one of the filter tap coefficients constant. This could be viewed as decomposing the filter impulse response into two *orthogonal* components, one of which is one-dimensional and nonadaptive. The setting in our case is, as we shall see, fundamentally different from the single-user channel subject to intersymbol interference. However, we propose a related approach where the impulse response of the linear receiver is decomposed into the signature waveform of the desired user plus an orthogonal adaptive component. We show that the receiver that results from the minimization of the output energy is the MMSE multiuser detector. Thus we succeed in obtaining an adaptive MMSE multiuser detector that does not require training sequences. In contrast to existing gradient-based single-user blind equalization algorithms which are plagued by local minima, our blind multiuser detector exhibits global convergence. A related blind multiuser detector was presented in [15] concurrent with a conference version of the present paper [16]. The approach in [15] was inspired by the minimum variance technique of adaptive array processing where the direction of arrival of the desired signal is known [17], [18]. The major difference between our approach and that of [15] is that the latter assumes knowledge of the interfering signature waveforms 2) and the acquisition of their timing 4).

Section II is devoted to the derivation of the relationship between the MMSE receiver and the anchored minimum-energy multiuser receiver, as well as the derivation of a blind adaptation rule which implements the minimum-energy multiuser receiver. It is shown for the first time that it is possible to have optimum near–far resistance with no knowledge beyond that assumed by the conventional single-user detector.

If the ability of our blind multiuser detector to successfully combat multiuser interference were predicated on the *exact* knowledge of the signature waveform of the user of interest, its practical applicability would be compromised. This is because the transmitted waveforms undergo *a priori* unknown (and time-varying) channel distortion in many of the environments where CDMA is used, and in particular, in mobile cellular and other wireless communication systems. For example, in a multipath scenario the received waveform is rather different from the transmitted signature sequence, although its normalized crosscorrelation with the nominal signature sequence is (normally) still much higher than the crosscorrelation with any of the interfering signature waveforms. Therefore, it is important to obtain a blind multiuser detector which is robust against imperfect knowledge of the assumed waveform of the user of interest. We show in Section III that a very simple modification of the multiuser detector of Section II achieves that goal thereby requiring only *soft* knowledge of the received waveform of the user of interest. The modification of the algorithm in Section III makes the receiver robust with respect to nominal desired signature waveform mismatch but is not designed so that the receiver learns the actual received signature waveform. When the mismatch is large, this adaptive capability (possessed by the MMSE adaptive detector with training sequences) is desirable and can still be achieved without training sequences. To that end, one possibility suggested by the results of Section III and IV, is to switch to a different (decision-directed) adaptation strategy after the minimum-energy receiver succeeds in lowering the bit-error-rate to adequate levels. Another possibility [19] is to replace the energy cost function by other nonconvex cost functions such as those used in single-user blind equalization.

Simulations are illustrated in Section IV along with an analysis of the convergence rate of the blind multiuser detector and the steady-state mean-square error for a fixed algorithm step size.

II. Blind Minimum Output Energy Multiuser Detector

A. Channel Model

The antipodal K-user asynchronous CDMA white Gaussian channel is (e.g. [5])

$$y(t) = \sum_{i=-M}^{M} \sum_{k=1}^{K} A_k b_k[i] s_k(t - iT - \tau_k) + \sigma n(t) \quad (1)$$

where $n(t)$ is white Gaussian noise with unit spectral density, the data $b_k[i]$ are independent and equally likely to be -1 or $+1$, $s_k(t)$ is the kth signature waveform which is assumed to have unit energy ($||s_k|| = 1$), A_k is the received amplitude of the kth user, and τ_k are the relative offsets of the received asynchronous signals at the receiver. The adoption of the baseband model in (1) is customary and incurs no loss of generality. At this point, we call attention to the fact that the assumption that the background noise is Gaussian plays a very minor role in this paper; in fact, it is only used in connection with a few observations having to do with the behavior of bit-error-rate, and it will be evident at which points this assumption is not superfluous.

Even though *synchronous* CDMA systems are more the exception than the rule, it is beneficial as usual (cf. [5]), to carry out the development first in the synchronous case, and then to incorporate the changes necessary to accommodate the more general asynchronous case. When the users are synchronous, it is sufficient to consider the one-shot version of (1) where $\tau_1 = \cdots = \tau_K$

$$y(t) = \sum_{k=1}^{K} A_k b_k s_k(t) + \sigma n(t), \quad t \in [0, T]. \quad (2)$$

The discussion in Sections II-B through II-E will be circumscribed to the synchronous case. In Section II-F we will study the asynchronous case.

B. Canonical Representation of Linear Multiuser Detectors

As we mentioned in Section I, our approach will be based on the decomposition of the linear multiuser detector as the sum of two orthogonal components. One of those components is equal to the signature waveform of the desired user which is assumed known and fixed throughout this section. As we show in this subsection, this decomposition is canonical in the sense that any linear multiuser detector can be represented in that form.

For convenience, we will assume that the user of interest is $k = 1$. A linear detector for user 1 is characterized by the impulse response $c_1 \in L_2[0, T]$, such that the decision on b_1 is

$$\hat{b}_1 = \text{sgn}(< y, c_1 >) \quad (3)$$

where the inner product notation denotes

$$< x, y >= \int_0^T x(t) y(t)\, dt.$$

Note that in situations where several users are to be demodulated simultaneously it is equivalent to view a linear multiuser detector as a multidimensional linear transformation or as a bank of single-user detectors.

For the purposes of this paper it is important to introduce the following *canonical representation* for the linear detector of user 1:

$$c_1 = s_1 + x_1 \quad (4a)$$

where

$$< s_1, x_1 >= 0. \quad (4b)$$

To see why (4) is indeed a canonical representation for any linear multiuser detector for user 1, note first that the set of signals that can be written as in (4) are those that satisfy

$$< c_1, s_1 >= ||s_1||^2 = 1 \quad (5)$$

and there is no loss of generality in restricting attention to linear transformations whose inner product with the signature waveform of the user of interest is normalized to 1, because a) we can rule out linear transformations that are orthogonal to the desired signal (they result in error probability equal to $1/2$), and b) the decision (3) is invariant to (positive) scaling.

Given a desired (up to a scale factor) c_1, the corresponding component orthogonal to s_1 is

$$x_1 = \frac{1}{< c_1, s_1 >} c_1 - s_1. \quad (6)$$

The bit-error-rate of the linear detector defined by (3) is equal to

$$P_1 = 2^{1-K} \sum_{d_2 \in \{-1,1\}} \cdots \sum_{d_K \in \{-1,1\}} Q\left(\frac{A_1 < c_1, s_1 > + \sum_{k=2}^{K} A_k d_k < c_1, s_k >}{\sigma ||c_1||} \right). \quad (7)$$

In the high SNR region ($\sigma \to 0$), the bit-error-rate is dominated by the largest term in the sum in (7). The *asymptotic multiuser efficiency* [2] is

$$\eta_1 = \frac{\max^2 \left[0, 1 - \sum_{k=2}^{K} \frac{A_k}{A_1} | < s_k, s_1 > + < s_k, x_1 > | \right]}{1 + ||x_1||^2}. \quad (8)$$

If $\eta_1 > 0$, then (7) goes to zero as $\sigma \to 0$ with the same slope (in log scale) as that of a single-user system

$$y(t) = A_1 {\eta_1}^{1/2} b_1 s_1(t) + \sigma n(t). \quad (9)$$

If $\eta_1 = 0$, then (7) does not go to zero (or at least not exponentially in $-\sigma^{-2}$). Therefore, the bit-error-rate in the high SNR region is determined by the asymptotic efficiency η_1 which can be viewed as a normalized version of the *eye opening*. The minimum asymptotic efficiency over all $A_k/A_1, k = 2, \cdots K$ is called the *near–far resistance* of the detector c_1 [5] or [20]. Among all the detectors that are independent of $A_k/A_1, k = 2, \cdots K$, the decorrelating detector [3] is the only one that has nonzero near–far resistance, equal to

$$\eta_1 = \frac{1}{1 + ||x_1||^2} \quad (10)$$

where in addition to being orthogonal to s_1, x_1 satisfies, for $k = 2, \cdots, K$

$$< s_k, x_1 >= - < s_k, s_1 > . \quad (11)$$

In fact, the near–far resistance of the decorrelating detector is not only nonzero but optimal [5]. Other linear detectors (which depend on the received amplitudes and noise level) with optimum near–far resistance include the optimum linear detector [3] and the MMSE detector [6] (cf. Section II-C). It is interesting to note from (10) that χ_I, the energy of x_1 necessary to cancel the multiple-access interference (in the absence of noise) depends (monotonically) only on η_1

$$\chi_I = \eta_1^{-1} - 1.$$

Another performance measure that we will investigate is the signal-to-interference ratio (SIR) at the output of the linear transformation c_1, i.e., the energy in the decision statistic due to the desired signal divided by the energy due to the interfering users plus the background Gaussian noise. This is an intuitively useful measure of performance, particularly in situations where the background noise is not negligible with respect to the multiaccess interference. A linear detector in canonical form $c_1 = s_1 + x_1$ has the following SIR:

$$\begin{aligned}\text{SIR} &= \frac{A_1^2(< c_1, s_1 >)^2}{\sigma^2||c_1||^2 + \sum_{k=2}^{K} A_k^2(< c_1, s_k >)^2} \\ &= \left(\frac{\sigma^2}{A_1^2}\left(1 + ||x_1||^2\right) + \sum_{k=2}^{K}\frac{A_k^2}{A_1^2}(< s_1 + x_1, s_k >)^2\right)^{-1}.\end{aligned}$$

C. MMSE Linear Multiuser Detector

The minimum mean-square-error (MMSE) linear multiuser detector for user 1 is defined as the signal $c_1 \in L_2[0,T]$ that minimizes the MSE

$$E\left[(A_1 b_1 - < y, c_1 >)^2\right]. \tag{12}$$

This detector has been previously obtained in different forms in [11] and [6]. For the sake of completeness we will show a simple way to obtain a closed-form expression for c_1 . Define for an arbitrary $K \times K$ matrix, $\boldsymbol{C} = \{c_{kj}\}$, the following signals:

$$c_k(t) = \sum_{j=1}^{K} c_{kj} s_j(t). \tag{13}$$

Instead of minimizing (12) with respect to c_1, we will minimize the function in (14) below with respect to $\boldsymbol{C}$. Naturally, the desired c_1 is obtained as the linear combination of signature waveforms dictated by the first row of $\boldsymbol{C}$.

$$\begin{aligned}&\sum_{k=1}^{K} E\left[(A_k b_k - < y, c_k >)^2\right] \\ &\quad = \text{tr}\left[\boldsymbol{W}^{\frac{1}{2}}(\boldsymbol{I} - \boldsymbol{R}\boldsymbol{C}^T)(\boldsymbol{I} - \boldsymbol{C}\boldsymbol{R})\boldsymbol{W}^{\frac{1}{2}} + \sigma^2 \boldsymbol{C}\boldsymbol{R}\boldsymbol{C}^T\right]\end{aligned} \tag{14}$$

where $\boldsymbol{W} = \text{diag}\{A_1^2, \cdots A_K^2\}$ and

$$\boldsymbol{R} = \left\{r_{kj} = \int_0^T s_k(t) s_j(t)\right\}.$$

The matrix $\boldsymbol{C}$ that minimizes (14) is

$$\begin{aligned}C_{\text{MMSE}} &= \boldsymbol{W}\boldsymbol{R}\left[\boldsymbol{R}\boldsymbol{W}\boldsymbol{R} + \sigma^2\boldsymbol{R}\right]^{-1} \\ &= \left[\boldsymbol{R} + \sigma^2\boldsymbol{W}^{-1}\right]^{-1}\end{aligned} \tag{15}$$

because letting $\boldsymbol{Q} = \boldsymbol{R}\boldsymbol{W}\boldsymbol{R} + \sigma^2\boldsymbol{R}$ and $\boldsymbol{P} = \boldsymbol{W}\boldsymbol{R}$ we can use the following general result.

Fact: For any positive definite matrix $\boldsymbol{Q}$

$$\text{argmin}_{\boldsymbol{C}}\ \text{tr}\left[\boldsymbol{C}\boldsymbol{Q}\boldsymbol{C}^T - \boldsymbol{P}\boldsymbol{C}^T - \boldsymbol{C}\boldsymbol{P}^T\right] = \boldsymbol{P}\boldsymbol{Q}^{-1}. \tag{16}$$

Proof: Denote the function of $\boldsymbol{C}$ in the left side of (16) by $f(\boldsymbol{C})$. It is easy to check that

$$f(\boldsymbol{Q}^{-1}\boldsymbol{P} + \boldsymbol{Z}) = f(\boldsymbol{Q}^{-1}\boldsymbol{P}) + tr(\boldsymbol{Z}\boldsymbol{Q}\boldsymbol{Z}^T) \tag{17}$$

where the last term is nonnegative by nonnegative definiteness of $\boldsymbol{Z}\boldsymbol{Q}\boldsymbol{Z}^T$. ■

Note that as $\sigma \to 0$, (15) becomes the decorrelating detector $\boldsymbol{R}^{-1}$ [3]. Another characterization of the linear MMSE detector is given in Section III.

We would like to investigate the canonical form of the linear MMSE detector. The two-user solution does not appear to reveal any particular structure

$$x_1 = \frac{\rho^2}{1 + \sigma^2/A_2^2 - \rho^2}s_1 - \frac{\rho}{1 + \sigma^2/A_2^2 - \rho^2}s_2. \tag{18}$$

However, a nice general characterization of the canonical representation of the MMSE linear detector is found in the following subsection.

D. Minimum Output-Energy Linear Detector

We consider in this subsection the linear detector in canonical form $s_1 + x_1$ that minimizes (over all x_1 orthogonal to s_1) the mean output energy

$$E\left[(< y, s_1 + x_1 >)^2\right]$$

when the input y is given by (2). The terminology "output energy" is in keeping with [13]; note, however, that we are referring to the variance of the correlator output at time T, rather than the energy of the correlator output waveform $y(t)c_1(t)$. Note that it is important to restrict the detector to be in canonical form, for otherwise the output energy is trivially minimized with $c_1 = 0$. Aside from the aforementioned motivation from the anchored minimum output energy approach of [13], we can expect intuitively that minimizing the output energy of the canonical linear detector will be a sensible approach. This is because the energy at the output can be written as the sum of the energy due to desired signal plus the energy due to the interference (background noise plus multiaccess interference), and the energy due to the desired signal is transparent to the choice of x_1. However, the main motivation is that the canonical linear detector with minimum output energy is, in fact, the MMSE detector as the following almost trivial observation shows.

Proposition 1: Consider a linear multiuser detector for user 1 in canonical form (4). Denote the mean-output-energy and the (scaled) mean-square-error, respectively, by

$$\mathrm{MOE}(x_1) = E\left[(< y, s_1 + x_1 >)^2\right] \tag{19}$$

and

$$\mathrm{MSE}(x_1) = E\left[(A_1 b_1 - < y, s_1 + x_1 >)^2\right]. \tag{20}$$

Then

$$\mathrm{MSE}(x_1) = \mathrm{MOE}(x_1) - A_1^2. \tag{21}$$

Proof:

$$\mathrm{MSE}(x_1) = A_1^2 + \mathrm{MOE}(x_1) - 2A_1^2 < s_1, s_1 + x_1 > \tag{22}$$

and the result follows from the fact that s_1 is orthogonal to x_1 and has unit energy. ■

Note that in order to obtain Proposition 1 we have not made use of the structure of the interference in (2). It is sufficient to assume that it is uncorrelated with the desired signal.

The simple observation that the mean-square-error and the output energy differ by a constant in terms of the canonical representation of the linear detector has key consequences for its adaptive implementation. The arguments that minimize both functions are the same. This means that (in contrast to the MMSE criterion) it is not necessary to know the data in order to implement a gradient descent algorithm for the minimization of mean-square-error. This sidesteps the use of training sequences and leads to the blind adaptation rule presented in the next subsection. Can the same idea be used to eliminate the need for training sequences in MMSE equalization of single-user channels subject to intersymbol interference? The answer is negative, because the counterpart of s_1 is the *unknown* channel impulse response. However, we will reconsider this answer in Section III.

Since we will be minimizing the mean output energy it is interesting to study the shape of this function. It is easy to show that the function $\mathrm{MOE}(x_1)$ is strictly convex over the set of signals orthogonal to s_1 (a convex set)

$$\mathrm{MOE}\left(\alpha x_1^a + (1-\alpha)x_1^b\right) = \alpha\ \mathrm{MOE}(x_1^a) + (1-\alpha)\,\mathrm{MOE}\left(x_1^b\right) - \alpha(1-\alpha)E\left[\left(< y, x_1^a - x_1^b >\right)^2\right] \tag{23}$$

where the expectation in the last term in (23) is larger than or equal to $\sigma^2\|x_1^a - x_1^b\|^2$. Therefore, the output energy has no local minima other than the unique global minimum—a most desirable property for gradient adaptation.

The minimum output energy solution exists even in the case where s_1 is spanned by the interferers, because the MMSE solution always exists if $\sigma > 0$, as we can deduce from (15).

E. Blind Adaptation Rule

The output energy function MOE lends itself to a simple stochastic gradient-descent adaptation rule which we present in this subsection. Note that other, potentially faster, techniques can be used in lieu of gradient descent; for example, Recursive Least Squares [21].

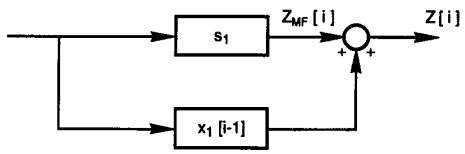

Fig. 1. Blind multiuser detector with $x_1[i-1]$ governed by (26).

We need to find the projection of the gradient of the output energy $\mathrm{MOE}(x_1)$ onto the linear subspace orthogonal s_1, so that the orthogonality condition (4b) is satisfied at each step of the algorithm. Note that the steepest descent line along the subspace orthogonal to s_1 is the projection of the gradient on that subspace. (The unconstrained gradient can be decomposed as the sum of its projections along s_1 and its orthogonal subspace; steepest descent requires steepest descent in each of those directions.)

Denote the observed waveform in the ith interval $[iT, iT + T]$ by $y[i] \in L_2[0,T]$. Let the ith output of the conventional single-user matched filter be the random variable

$$Z_{MF}[i] = < y[i], s_1 > . \tag{24}$$

Analogously, let the ith output of the proposed linear transformation be

$$Z[i] = < y[i], s_1 + x_1[i-1] > . \tag{25}$$

Recall that the output of the detector is $\hat{b}_1(i) = \mathrm{sgn}(Z[i])$. The linear transformation outputs $Z[i]$ and $Z_{MF}[i]$ are used to compute $x_1[i]$ (which therefore depends on the received waveforms $\cdots y[i-1], y[i]$). The derivation of the adaptation rule for $x_1[i]$ is very simple. The unconstrained gradient of the averaged random variable in

$$\mathrm{MOE}(x_1) = E\left[(< y, s_1 + x_1 >)^2\right]$$

is equal to a scaled version of the observations

$$2 < y, s_1 + x_1 > y.$$

The component of y orthogonal to s_1 is equal to

$$y - < y, s_1 > s_1.$$

Therefore, the stochastic gradient adaptation rule is

$$x_1[i] = x_1[i-1] - \mu Z[i](y[i] - Z_{MF}[i]s_1). \tag{26}$$

In practice, because of finite precision effects, the updated vector x_1 may not exactly satisfy the orthogonality condition (4b). It may, therefore, be necessary on occasion to replace $x_1[i]$ by its orthogonal projection onto s_1.

The foregoing derivation has been general enough to apply to any Hilbert space, not necessarily $L_2[0,T]$. In particular, in applications where there is a chip-matched filter at the front end (or some other sampling mechanism) the signals s_1 and x_1 should be viewed as belonging to a finite-dimensional Euclidean space whose dimensionality is equal to the number of samples used per symbol decision.

Using the general results of [22], it can be shown that the algorithm (26) converges regardless of the initial condition to the MMSE detector if the step size decreases as $\mu[i] = 1/i$. In practice, a lower bounded step size μ is often needed to track channel variations; the dynamics and excess MSE of the adaptation rule in (26) are studied in Section IV, for a fixed arbitrary step size μ .

F. Asynchronous Case

Even though the asynchronous channel (1) looks quite a bit more complicated than the synchronous channel (2), it turns out that we can extend previous results with little conceptual difficulty.

As in [5] and [20], we may view every bit as transmitted by a different fictitious user. Let us single out a particular bit of the desired user, say, $b_1(0)$, such that among all $(2M+1)K$ bits in (1) we take $b_1(0)$ to be the "user" of interest. Rather than restricting our observation interval to $[0,T]$ (which is where the desired signal lies, assuming without loss of generality that $\tau_1 = 0$) as in the synchronous case, we may take an interval (which we will refer to as the *processing window*) that supports all the bit periods in (1), e.g., $[-MT, MT+T]$. Now, our previous analysis carries over to the Hilbert space $L_2[-MT, MT+T]$, and in particular the adaptation rule in (26) is unchanged with the proper interpretations of the correlations in (24) and (25): x_1 is now defined over the entire processing window. Note that even though the signal of the desired "user" is identically zero outside the interval $[0,T]$, the minimum-energy x_1 need not be zero outside that interval. This is because the contributions to $< y, s_1 + x_1 >$ from inside and outside $[0,T]$ are correlated.

In practice, one would not implement the blind multiuser receiver with a processing window spanning the whole observation interval. Not only would that be impractical but a sufficiently long sliding processing window can achieve practically the same performance. In some cases, just the interval of the desired symbol is a sensible choice for the processing window [6]. The loss of near–far resistance caused by truncating the processing window to the interval of the desired symbol is studied in [20]. We note that the global convergence properties mentioned in the synchronous case can also be proven in the asynchronous case using the results of [22].

In the synchronous direct-sequence spread-spectrum case, sampling a chip-matched filter at the chip rate incurs no loss of information because all the signature waveforms can be decomposed with that basis. In other words, the chip-matched filter samples are sufficient statistics. In the asynchronous case, we would have to have a chip-matched filter synchronized with each of the interfering users. However, acquiring the timing of the interfering users (requirement 4) in Section I) is clearly undesirable. But if the chip waveforms are bandlimited to f_0, sufficient statistics are obtained by sampling at $2f_0$. If the results for the MMSE linear multiuser detector in [6] are to serve as an indication, we would expect good performance by sampling at the chip rate (synchronized with the user of interest).

In some CDMA applications of practical interest the channel model in (1) does not apply because the signature (pseudonoise) sequence spans L bits for each of the users. If L is low enough that the offsets and received waveforms remain (reasonably) constant, the blind adaptation algorithm can be extended to L independent algorithms running in parallel.

III. Blind Multiuser Detector with Mismatched Nominal

A. Mismatch and Surplus Energy

We assumed in Section II that the receiver has perfect knowledge of the signature waveform s_1 used to modulate the bits of the desired user. This may not be true in practice, e.g., the receiver may assume the original spreading waveform as its nominal, whereas the actual received waveform s_1 may include additional multipath components or other types of channel distortion. In this section, we evaluate the performance of the minimum energy detector under such a *mismatch*. Specifically, it is assumed that the linear detector for the desired user is $c_1 = \hat{s}_1 + x_1$, where $\hat{s}_1$ is the assumed *nominal*, and where $< \hat{s}_1, x_1 > = 0$. We assume that $\| \hat{s}_1 \| = 1$ without loss of generality. When the nominal $\hat{s}_1$ is not equal to the desired waveform s_1, minimizing the energy $E[< y, \hat{s}_1 + x_1 >^2]$ without additional constraints can cause cancellation of the desired signal, since x_1 is no longer constrained to be orthogonal to the desired waveform. We explore the effect of such mismatch in this section, starting with an example.

Henceforth, it is convenient to represent signals as finite-dimensional vectors with respect to some basis. We will denote such vectors in bold notation (e.g., $\boldsymbol{s}_k$ is the vector corresponding to s_k, $\tilde{\boldsymbol{s}}_1$ to $\hat{s}_1$, $\boldsymbol{c}_1$ to c_1, and $\boldsymbol{x}_1$ to x_1). Such a vector representation arises naturally in practice due to the conversion of the continuous-time received signal to discrete time by filtering and sampling. The inner product $< \cdot, \cdot >$ now denotes a conventional vector inner product.

Example 3.1: Consider a system with two users ($K = 2$). Since the desired signal, the interfering signal and the nominal can span a space of dimension at most three, we assume without loss of generality that these signals lie in $\mathbb{R}^3$. We set $\tilde{\boldsymbol{s}}_1^T = (1\ 0\ 0)$ and $\boldsymbol{s}_1^T = (1\ \epsilon\ 0)/\sqrt{1+\epsilon^2}$, where ϵ is a measure of the mismatch. We choose $\boldsymbol{s}_2{}^T = (\delta\ 0\ 1)/\sqrt{1+\delta^2}$. This last choice does involve some loss of generality, but it has the advantage of parametrizing $\boldsymbol{s}_2$ using a single parameter δ , which is a measure of the correlation of the interferer $\boldsymbol{s}_2$ with the nominal $\tilde{\boldsymbol{s}}_1$ and the desired signal $\boldsymbol{s}_2$. In the canonical form of the detector, the vector $\boldsymbol{x}_1$ is of the form $\boldsymbol{x}_1{}^T = (0\ a\ b)$, since it must be orthogonal to $\tilde{\boldsymbol{s}}_1$.

Assuming that the desired signal's amplitude is $A_1 = 1$, consider first a situation with zero thermal noise and interference amplitude $A_2 = 0$. Provided there is a mismatch ($\epsilon \neq 0$), a minimum output energy of zero is attained by choosing $\boldsymbol{x}_1$ to cancel the desired signal completely, i.e., for $\boldsymbol{x}_1$ such that $< \tilde{\boldsymbol{s}}_1 + \boldsymbol{x}_1, \boldsymbol{s}_1 >= 0$. The minimum-norm $\boldsymbol{x}_1$ achieving this is clearly $\boldsymbol{x}_1{}^T = (0\ -\epsilon^{-1}\ 0)$, which has energy $\|\boldsymbol{x}_1\|^2 = \epsilon^{-2}$. In order to prevent cancellation of the desired signal, therefore, we must force $\|\boldsymbol{x}_1\|^2$ to be smaller than ϵ^{-2}

($\epsilon = 0$ corresponds to no mismatch, the situation considered in Section II, where no constraint on $\|\boldsymbol{x}_1\|^2$ is needed).

Consider now a situation in which $A_2 \to \infty$. In this case, the minimum energy detector clearly needs to satisfy $< \hat{\boldsymbol{s}}_1 + \boldsymbol{x}_1, \boldsymbol{s}_2 > = 0$ (otherwise the output energy grows without bound as $A_2 \to \infty$), and the minimum-norm $\boldsymbol{x}_1$ achieving this is $\boldsymbol{x}_1{}^T = (0\ 0\ -\delta)$. The energy of $\|\boldsymbol{x}_1\|^2$ in this case is given by δ^2. It is therefore essential to allow $\|\boldsymbol{x}_1\|^2$ to be larger than δ^2 in order to preserve the ability of the minimum energy detector to suppress strong interference.

Clearly, the two conditions on $\|\boldsymbol{x}_1\|^2$ can both be satisfied only if $\delta^2 < \epsilon^{-2}$. If the latter does not hold, it is not possible to prevent signal cancellation while canceling interference. This is to be expected, however, since violation of this condition is equivalent to saying that the nominal is closer to the space spanned by the interferer than to the space spanned by the signal. We elaborate upon this condition in a more general setting in the following.

The preceding example illustrates how the presence of mismatch forces us to constrain $\|\boldsymbol{x}_1\|^2$, which is henceforth termed the *surplus energy* χ available to the detector. The term arises from the fact that the energy of the linear transformation in the detector is given by .

$$\|\boldsymbol{c}_1\|^2 = \|\hat{\boldsymbol{s}}_1\|^2 + \|\boldsymbol{x}_1\|^2 = 1 + \chi$$

so that χ is a measure of the extent to which $\boldsymbol{c}_1$ can be shaped to reduce the output energy ($\chi = 0$ corresponds to the conventional detector). In order to discuss the tradeoffs involved in choosing the surplus energy, it is convenient to define χ_S, the minimum value of surplus energy necessary for complete cancellation of the desired signal ($\chi_S = \epsilon^{-2}$ in Example 3.1), and χ_I, the minimum value of surplus energy necessary for complete cancellation of the multiple-access interference regardless of the amplitudes $A_2, \cdots A_K$ ($\chi_I = \delta^2$ in Example 3.1). As shown in Section III-B, these quantities depend only on the crosscorrelations of $\{\hat{\boldsymbol{s}}_1, \boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_K\}$.

In the presence of mismatch, choosing too large a value of surplus energy leads to cancellation of the desired signal. Further, for nonzero background noise, high values of surplus energy lead to noise enhancement at the output. On the other hand, choosing $\chi < \chi_I$ implies that the detector is unable to suppress strong interference. A surplus energy of approximately χ_I appears, therefore, to be the best choice for trading off interference suppression versus signal degradation and noise enhancement. However, this choice can still lead to significant cancellation of the desired signal *unless* $\chi_I < \chi_S$ (preferably $\chi_I \ll \chi_S$), especially when the interference is weak. The latter is therefore a *necessary* condition for obtaining near–far resistant performance without excessive signal cancellation, and is shown in Section III-B to be equivalent to the intuitively pleasing criterion that the nominal $\tilde{\boldsymbol{s}}_1$ is closer (in L_2 distance) to the subspace spanned by the desired signal $\boldsymbol{s}_1$ than to the *interference subspace* $\boldsymbol{S}_I$ spanned by the interfering signals $\boldsymbol{s}_2, \cdots, \boldsymbol{s}_K$. In two-user channels with sufficiently low background noise level, it can also be shown that the preceding condition is sufficient to ensure that, for any interference amplitude, there is a value of surplus energy χ for which the constrained minimum-energy detector gives rise to nonzero asymptotic efficiency, or equivalently, to an open eye.

While higher surplus energy permits more cancellation of both the desired signal and the interference, for nonzero background noise, it also increases the noise contribution at the output. Since the surplus energy for the minimum output energy detector is based on the preceding tradeoff, it is clear that higher values of background noise lead to smaller values of surplus energy. This *implicit* constraint on the surplus energy due to the background noise (see Section III-C for details) is important for many practical applications in which the receiver may not know the range of χ_I and χ_S, and therefore may have difficulty in choosing the constraint on χ . However, for relatively high signal-to-noise ratio (SNR), the constraint imposed on the surplus energy due to the background noise is not stringent enough to prevent significant cancellation of the desired signal. In such cases, it is necessary to impose a further *explicit* constraint on χ to prevent signal degradation (see the numerical results for Example 3.2 later in this section).

The rest of this section is organized as follows. In Section III-B, we compute the values of χ_S and χ_I in terms of crosscorrelation parameters. In Section III-C, we derive the solution to the problem of minimizing the output energy subject to a constraint on the surplus energy. Given the solution, performance measures like SIR and asymptotic efficiency can be computed using the definitions in Section II. We also give the modification to the adaptive implementation due to the constraint on the surplus energy. Numerical results are presented in Section III-D.

B. Values of Surplus Energy for Signal and Interference Cancellation

For $1 \le k \le K$, let $\hat{\rho}_k =< \boldsymbol{s}_k, \hat{\boldsymbol{s}}_1 >$ denote the crosscorrelation of the kth-signal waveform with the nominal. Let $\hat{\boldsymbol{p}}_S$ denote the projection of $\tilde{\boldsymbol{s}}_1$ orthogonal to the space spanned by $\boldsymbol{s}_1$. Note that $\|\hat{\boldsymbol{p}}_S\|$ is the L_2 distance of $\tilde{\boldsymbol{s}}_1$ from the subspace spanned by the desired signal $\boldsymbol{s}_1$, and is given by $\|\hat{\boldsymbol{p}}_S\|^2 = 1 - \hat{\rho}_1^2$. The contribution of the desired signal to the output can be canceled completely by choosing $\boldsymbol{x}_1$ such that $\tilde{\boldsymbol{s}}_1 + \boldsymbol{x}_1$ is a scaled version of $\hat{\boldsymbol{p}}_S$. Moreover, this choice of $\boldsymbol{x}_1$ attains the minimum surplus energy that cancels the desired signal, i.e., $\|\boldsymbol{x}_1\|^2 = \chi_S$. While χ_S can be computed algebraically based on the preceding observation, we attempt to add to our intuition by computing it via a simple geometric observation. Fig. 2 shows the direction of $\boldsymbol{x}_1$ for which the output signal energy decreases the fastest as a function of χ. This is also the asymptotic direction of $\boldsymbol{x}_1$ minimizing the output energy for small interference amplitudes. The surplus energy χ_S is clearly given by

$$\begin{aligned}\chi_S = \|\boldsymbol{x}_1\|^2 &= \|\hat{\boldsymbol{s}}_1\|^2 \tan^2\theta_S = \frac{1-\cos^2\theta_S}{\cos^2\theta_S} \\ &= \|\hat{\boldsymbol{p}}_S\|^{-2} - 1 = \hat{\rho}_1^2/(1-\hat{\rho}_1^2)\end{aligned} \tag{27}$$

since the angle θ_S between $\tilde{\boldsymbol{s}}_1$ and $\hat{\boldsymbol{p}}_S$ is given by

$$\cos\theta_S = \|\hat{\boldsymbol{p}}_S\|/\|\hat{\boldsymbol{s}}_1\| = \|\hat{\boldsymbol{p}}_S\|.$$

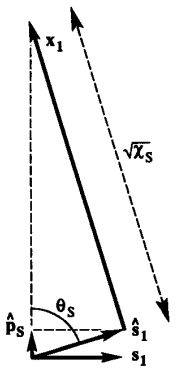

Fig. 2. Computation of surplus energy for complete signal cancellation.

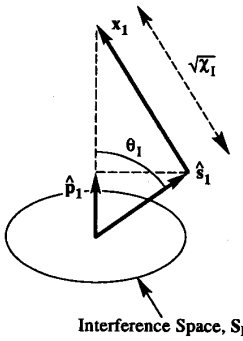

Fig. 3. Computation of surplus energy for complete interference cancellation.

The computation of χ_I, the minimum value of χ required to cancel *all* the interfering signals, is entirely similar. In this case, $\boldsymbol{x}_1$ is chosen so that $\tilde{\boldsymbol{s}}_1+\boldsymbol{x}_1$ is a multiple of the projection $\hat{\boldsymbol{p}}_I$ of $\tilde{\boldsymbol{s}}_1$ orthogonal to the interference space $\boldsymbol{S}_I$, as shown in Fig. 3. Note that $\|\hat{\boldsymbol{p}}_I\|$ is the L_2 distance of $\tilde{\boldsymbol{s}}_1$ from the interference space. Since $\cos\theta_I = \|\hat{\boldsymbol{p}}_I\|$ and $\chi_I = \tan^2\theta_I$, we have

$$\chi_I = 1/\|\hat{\boldsymbol{p}}_I\|^2 - 1 = 1/\hat{\eta}_I - 1 \tag{28}$$

where $\hat{\eta}_I = \|\hat{\boldsymbol{p}}_I\|^2/\|\hat{\boldsymbol{s}}_1\|^2 = \|\hat{\boldsymbol{p}}_I\|^2$ is the near–far resistance of $\tilde{\boldsymbol{s}}_1$ with respect to the interference space, and is given by [6]

$$\hat{\eta}_I = 1 - \hat{\boldsymbol{\rho}}_1^T \boldsymbol{R}_I{}^{-1}\hat{\boldsymbol{\rho}}_1 \tag{29}$$

where $\hat{\boldsymbol{\rho}}_1^T = (\hat{\rho}_2, \cdots, \hat{\rho}_K)$ is the crosscorrelation vector of $\tilde{\boldsymbol{s}}_1$ with the interfering signals, and $\boldsymbol{R}_I = (R_{kl})_{2\le k,l\le K}$ is the $(K-1)\times(K-1)$ matrix of crosscorrelations for the interfering signals.

Note that the choice of $\boldsymbol{x}_1$ in Fig. 3 is not necessarily that which minimizes the output interference energy for a given value of $\chi < \chi_I$, since the minimizing $\boldsymbol{x}_1$ depends on the amplitudes $A_2, \cdots, A_K$. However, the direction of $\boldsymbol{x}_1$ shown in Fig. 3 is an amplitude-insensitive choice which completely cancels the interference with the least possible surplus energy.

From (27) and (28), the condition $\chi_I < \chi_S$ is seen to be equivalent to $\|\hat{\boldsymbol{p}}_I\| > \|\hat{\boldsymbol{p}}_S\|$, i.e., to the nominal $\tilde{\boldsymbol{s}}_1$ being further from the interference space than from the space spanned by the desired signal.

It is worth noting that similar reasoning also gives the minimum surplus energy χ^* required for complete cancellation of both the desired signal and the multiple-access interference to be $\chi^* = 1/\hat{\eta} - 1$, where $\hat{\eta}$ is the near–far resistance of $\tilde{\boldsymbol{s}}_1$ with respect to the entire signal space, i.e., the space $\boldsymbol{S}$ spanned by $\boldsymbol{s}_1, \cdots, \boldsymbol{s}_K$, given by

$$\hat{\eta} = 1 - \hat{\boldsymbol{\rho}}^T \boldsymbol{R}^{-1}\hat{\boldsymbol{\rho}}$$

where $\hat{\boldsymbol{\rho}}^T = (\hat{\rho}_1, \hat{\rho}_2, \cdots, \hat{\rho}_K)$ and $\boldsymbol{R} = (R_{kl})_{1\le k,l\le K}$.

C. Minimum Output Energy Detector with Constrained Surplus Energy

The *constrained* minimum output energy detector minimizes (over $\boldsymbol{x}_1$) the cost function

$$\mathrm{MOE}(\boldsymbol{x}_1) = E\left[<\boldsymbol{y}, \hat{\boldsymbol{s}}_1 + \boldsymbol{x}_1>^2\right] = E\left[<\boldsymbol{y}, \boldsymbol{c}_1>^2\right]$$

subject to $\|\boldsymbol{x}_1\|^2 = \chi$ and $<\hat{\boldsymbol{s}}_1, \boldsymbol{x}_1> = 0$. Expressing the optimization problem in terms of $\boldsymbol{c}_1 = \hat{\boldsymbol{s}}_1 + \boldsymbol{x}_1$, we obtain, upon taking the derivative with respect to $\boldsymbol{c}_1$ of the associated Lagrangian, the following optimality condition:

$$E[<\boldsymbol{c}_1, \boldsymbol{y}>\boldsymbol{y}] + \nu_1\boldsymbol{c}_1 - \nu_2\hat{\boldsymbol{s}}_1 = 0 \tag{30}$$

where ν_1 and ν_2 are Lagrange multipliers chosen so that $\|\boldsymbol{c}_1\|^2 = \chi + 1$ and $<\boldsymbol{c}_1, \hat{\boldsymbol{s}}_1> = 1$. Assuming that the bits b_k are uncorrelated, the preceding condition can be rewritten as

$$\sum_{k=1}^{K} A_k^2 <\boldsymbol{c}_1, \boldsymbol{s}_k>\boldsymbol{s}_k + \left(\nu_1 + \sigma^2\right)\boldsymbol{c}_1 - \nu_2\hat{\boldsymbol{s}}_1 = 0. \tag{31}$$

Letting $\boldsymbol{A}$ denote the outer-product matrix

$$\boldsymbol{A} = \sum_{k=1}^{K} A_k^2 \boldsymbol{s}_k \boldsymbol{s}_k{}^T$$

the optimality condition (31) can be shown to yield

$$\boldsymbol{c}_1 = \nu_2(\boldsymbol{A} + \gamma\boldsymbol{I}_N)^{-1}\hat{\boldsymbol{s}}_1 = \frac{(\boldsymbol{A} + \gamma\boldsymbol{I}_N)^{-1}\tilde{\boldsymbol{s}}_1}{\hat{\boldsymbol{s}}_1^T(\boldsymbol{A} + \gamma\boldsymbol{I}_N)^{-1}\tilde{\boldsymbol{s}}_1} \tag{32}$$

where

$$\gamma = \nu_1 + \sigma^2 \tag{33}$$

$\boldsymbol{I}_N$ is the $N \times N$ identity, and where the value of

$$\nu_2 = \left[\hat{\boldsymbol{s}}_1^T(\boldsymbol{A} + \gamma\boldsymbol{I}_N)^{-1}\hat{\boldsymbol{s}}_1\right]^{-1}$$

is obtained using the condition $<\boldsymbol{c}_1, \hat{\boldsymbol{s}}_1> = 1$. The corresponding minimum value $\xi_{\min}$ of output energy is obtained by taking the inner product of (30) with $\boldsymbol{c}_1$

$$\xi_{\min} = \nu_2 - \nu_1(1 + \chi) \tag{34}$$

where the surplus energy χ is given by

$$\chi = \|\boldsymbol{c}_1\|^2 - 1 = \frac{\hat{\boldsymbol{s}}_1^T(\boldsymbol{A} + \gamma\boldsymbol{I}_N)^{-2}\tilde{\boldsymbol{s}}_1}{\left[\hat{\boldsymbol{s}}_1^T(\boldsymbol{A} + \gamma\boldsymbol{I}_N)^{-1}\hat{\boldsymbol{s}}_1\right]^2} - 1. \tag{35}$$

The preceding results can be easily specialized to the situation in which there is no explicit constraint on surplus energy by setting $\nu_1 = 0$. Defining

$$\boldsymbol{R}_y = E\left[\boldsymbol{y}\boldsymbol{y}^T\right] = \boldsymbol{A} + \sigma^2\boldsymbol{I}_N$$

we obtain

$$\boldsymbol{c}_1 = \xi_{\min} \boldsymbol{R}_y^{-1} \hat{\boldsymbol{s}}_1 \qquad \xi_{\min} = \left(\hat{\boldsymbol{s}}_1 \boldsymbol{R}_y^{-1} \hat{\boldsymbol{s}}_1\right)^{-1}.$$

This is further specialized to the solution without mismatch by setting $\hat{\boldsymbol{s}}_1 = \boldsymbol{s}_1$.

Having specified the detector $\boldsymbol{c}_1$ as in (32), we can now compute all the performance measures of interest, including SIR and asymptotic efficiency, using the definitions in Section II. The performance and the surplus energy χ are functions of γ, so that we can plot the performance as a function of χ by varying γ. Note that the value of γ completely determines the minimum-output energy solution. It follows from (33) that the Lagrange multiplier ν_1 plays precisely the same role as the noise variance σ^2, i.e., that imposing an explicit constraint on the surplus energy is equivalent to an implicit constraint due to excess background noise in terms of the minimum-output energy solution (the solution in the absence of any implicit or explicit constraint on the surplus energy corresponds to $\gamma = 0$). However, for a given value of γ, the performance will clearly be worse for a larger value of σ^2, since the noise contribution to the output is greater while the signal and interference contributions remain the same.

Since the rank of the outer product matrix $\boldsymbol{A}$ is bounded above by the number of signals K, inverting the $N \times N$ matrix $\boldsymbol{A} + \gamma \boldsymbol{I}_N$ may be an ill-conditioned problem for small γ if $N > K$, as is typically the case. These difficulties are overcome by expressing the minimum energy solution in terms of crosscorrelation parameters as follows. Without loss of generality, we write $\boldsymbol{c}_1$ as

$$\boldsymbol{c}_1 = \alpha \hat{\boldsymbol{s}}_1 + \sum_{k=1}^{K} z_k \boldsymbol{s}_k. \tag{36}$$

Any component of $\boldsymbol{c}_1$ orthogonal to the space spanned by $\hat{\boldsymbol{s}}_1, \boldsymbol{s}_1, \cdots, \boldsymbol{s}_K$ increases the contribution of the background noise to the output energy while not affecting the signal contribution. For a nonzero background noise level, therefore, no such component can appear in the minimum-output energy solution. In the adaptive implementation of the minimum-output energy detector, however, such components can cause the phenomenon of "tap wandering" [23] for low noise levels. This can be prevented, however, by an explicit constraint on the surplus energy.

It remains to compute α and $\boldsymbol{z}^T = (z_1, \cdots, z_K)$. The constraint $< \boldsymbol{c}_1, \hat{\boldsymbol{s}}_1 >= 1$ yields α in terms of $\boldsymbol{z}$

$$\alpha = 1 - \boldsymbol{z}^T \hat{\boldsymbol{\rho}}. \tag{37}$$

From (36) and (37), we obtain

$$\begin{aligned} < \boldsymbol{c}_1, \boldsymbol{s}_k > &= \alpha \hat{\rho}_k + (\boldsymbol{R}\boldsymbol{z})_k \\ &= \hat{\rho}_k + \left(\hat{\boldsymbol{R}}\boldsymbol{z}\right)_k, \quad 1 \le k \le K. \end{aligned} \tag{38}$$

where

$$\hat{\boldsymbol{R}} = \boldsymbol{R} - \hat{\boldsymbol{\rho}}\hat{\boldsymbol{\rho}}^T. \tag{39}$$

Using (36), (37), and (39), it is easy to show that the surplus energy is given by

$$\chi = \boldsymbol{z}^T \hat{\boldsymbol{R}} \boldsymbol{z}. \tag{40}$$

In order to find $\boldsymbol{z}$, we express the optimality condition (31) in terms of crosscorrelation parameters by taking the inner product of both sides with $\boldsymbol{s}_k$, $1 \le k \le K$. Using (38) and (39), the resulting K equations can be expressed in vector form as

$$\boldsymbol{R}\boldsymbol{W}\left(\hat{\boldsymbol{R}}\boldsymbol{z} + \hat{\boldsymbol{\rho}}\right) + \gamma\left(\hat{\boldsymbol{R}}\boldsymbol{z} + \hat{\boldsymbol{\rho}}\right) - \nu_2 \hat{\boldsymbol{\rho}} = 0. \tag{41}$$

An additional equation is obtained by taking the inner product of (31) with $\tilde{\boldsymbol{s}}_1$

$$\hat{\boldsymbol{\rho}}^T \boldsymbol{W}\left(\hat{\boldsymbol{R}}\boldsymbol{z} + \hat{\boldsymbol{\rho}}\right) + \gamma - \nu_2 = 0. \tag{42}$$

Eliminating ν_2 between (41) and (42), we obtain upon simplification that $\boldsymbol{z}$ must satisfy

$$\hat{\boldsymbol{R}}\left[\left(\boldsymbol{W}\hat{\boldsymbol{R}} + \gamma \boldsymbol{I}_K\right)\boldsymbol{z} + \boldsymbol{W}\hat{\boldsymbol{\rho}}\right] = 0 \tag{43}$$

where $\boldsymbol{I}_K$ denotes the $K \times K$ identity. The solution to (43) is unique if, and only if, $\boldsymbol{R}$ is nonsingular. It can be checked, however, that for fixed γ, all solutions to (43) lead to the same detector and the same value of surplus energy, so that it suffices to consider a specific solution

$$\boldsymbol{z} = -\left(\hat{\boldsymbol{R}} + \gamma \boldsymbol{W}^{-1}\right)^{-1} \hat{\boldsymbol{\rho}}, \tag{44}$$

where the inverse is replaced by a pseudoinverse if necessary (i.e., if $\hat{\boldsymbol{R}}$ is singular and $\gamma = 0$). Using (38), (40), and (44), we can now compute quantities such as SIR, asymptotic efficiency, and surplus energy in terms of γ and the crosscorrelation parameters. As before, $\gamma = 0$ corresponds to the decorrelating solution $\boldsymbol{z} = \hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{\rho}}$ for unconstrained surplus energy.

Stochastic Gradient Algorithm for Constrained Minimum Output Energy Detector

Finally, we give an adaptive algorithm for implementing the constrained minimum output energy detector. This is obtained by modifying the stochastic gradient algorithm in Section II to reflect that fact that the Lagrange multiplier $\nu_1 \ge 0$ simply adds a term $\nu_1 \boldsymbol{x}_1$ to the projection of the gradient of the output energy orthogonal to $\hat{\boldsymbol{s}}_1$.

$$\boldsymbol{x}_1[i+1] = (1 - \mu\nu_1)\boldsymbol{x}_1[i] - \mu Z[i](y[i] - Z_{MF}[i]\hat{\boldsymbol{s}}_1). \tag{45}$$

The results of Section IV show, however, that a good practical alternative to the preceding constrained adaptation is to adapt using an unconstrained output energy criterion at the beginning, starting from $\boldsymbol{x}_1 = 0$ (hence $\chi = 0$), then let χ grow, and finally switch to a decision-directed mode using a mean-squared-error criterion before the surplus energy becomes too large. One possibility for the value of surplus energy at which to switch is $\chi_I = \hat{\eta}_I^{-1} - 1$. While χ_I is typically unknown, a rough estimate for it may be obtained as follows. If all the signature waveforms are chosen to be independent random binary sequences, it has been shown in [24] that $E[\hat{\eta}_I] \approx 1 - (K-1)/N$ for synchronous CDMA and $E[\hat{\eta}_I] \approx 1 - 2(K-1)/N$ for asynchronous CDMA, where

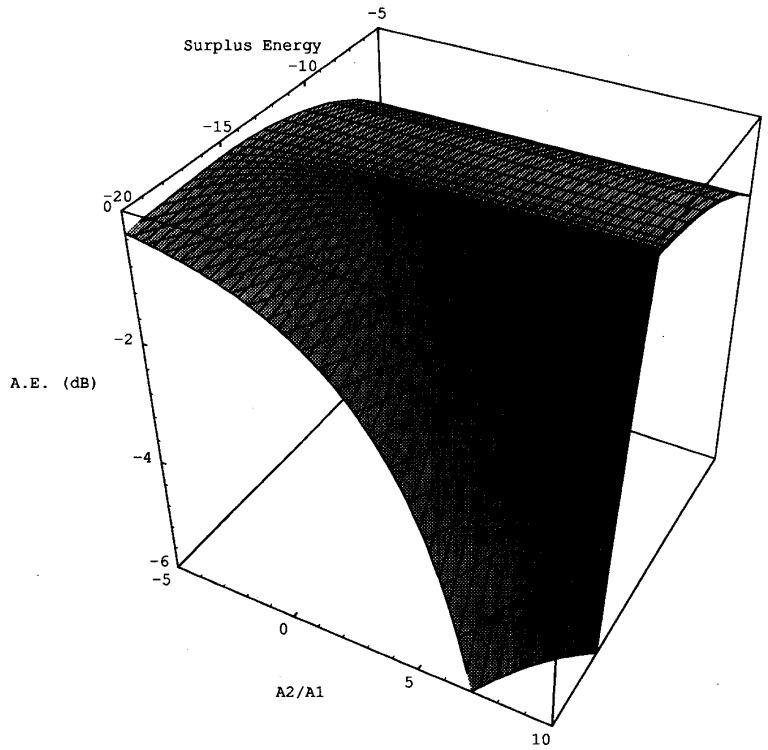

Fig. 4. Asymptotic multiuser efficiency as a function of surplus energy and interference ratio. Example 3.1 with $\epsilon = 0.001$ and $\delta = 0.2$.

these estimates are actually lower bounds. Replacing $\hat{\eta}_I$ by these estimates of $E[\hat{\eta}_I]$, we obtain

$$\chi_I \approx \begin{cases} \dfrac{K-1}{N-(K-1)}, & \text{synchronous CDMA} \\ \dfrac{2(K-1)}{N-2(K-1)}, & \text{asynchronous CDMA.} \end{cases}$$

D. Numerical Results

We consider two examples. In the first, we return to the two-user system described in Example 3.1, and study the effect of surplus energy on asymptotic efficiency, which, as described in Section II, is a measure of the detector performance relative to a single-user system. In the second example (Example 3.2), we consider a system with $K = 7$ and processing gain $N = 10$. The signature sequences are generated randomly, and the mismatch is generated by assuming the presence of multipath. The performance measure considered in Example 3.2 is the SIR. This second model is also used to generate the numerical results in Section IV on the performance of the adaptive algorithm.

Example 3.1 (Continued): For the two-user system considered at the beginning of this section, we have

$$\hat{\boldsymbol{\rho}}^T = \left(\frac{1}{\sqrt{1+\epsilon^2}} \quad \frac{\delta}{\sqrt{1+\delta^2}} \right) \qquad \boldsymbol{R} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

where $\rho = \delta/\sqrt{(1+\epsilon^2)(1+\delta^2)}$. The values of χ_S, χ_I, and χ^* are given by

$$\chi_S = \epsilon^{-2} \qquad \chi_I = \delta^2 \qquad \chi^* = \epsilon^{-2} + \delta^2.$$

As mentioned earlier, choosing $\chi = \chi_I = \delta^2$ balances the ability to suppress multiple-access interference with the need to avoid excessive signal cancellation and noise enhancement. The necessary condition $\chi_I < \chi_S$ for this approach to work translates to $(\epsilon\delta)^2 < 1$ (the smaller the left-hand side, the better the performance can be expected to be).

In Figs. 4–6 we show the asymptotic efficiency of the desired user as a function of the surplus energy and the ratio of interfering user amplitude to desired user amplitude. All three quantities are displayed in decibels; the value in decibels of the surplus energy can be thought of as being relative to the nominal signal energy. Figs. 4–6 correspond to the values: $(\epsilon, \delta) = \{(0.001, 0.2), (0.5, 0.1), (0.1, 1)\}$, respectively. Fig. 4 corresponds to a case with extremely small mismatch. Fig. 5 has relatively high mismatch but moderate crosscorrelation with the interfering waveform, and Fig. 6 depicts the case of unusually heavy crosscorrelation between both received signals The corresponding values of $\chi_I = \delta^2$ are -14 dB, -20 dB, and 0 dB, respectively. In all three cases, we can see that for low A_2/A_1 , the choice of surplus energy is relatively unimportant, unless it is much higher than χ_I, in which case the effect of desired-signal cancellation and noise enhancement is evident. As we would expect the sensitivity of asymptotic efficiency to surplus energy in the region of low interference increases with the degree of mismatch, quantified by ϵ. If the surplus energy is well below χ_I, the detector is not near–far resistant. In the high-interference region, we see that the sensitivity to the surplus energy is much higher below χ_I than above. In all cases considered, given an optimum

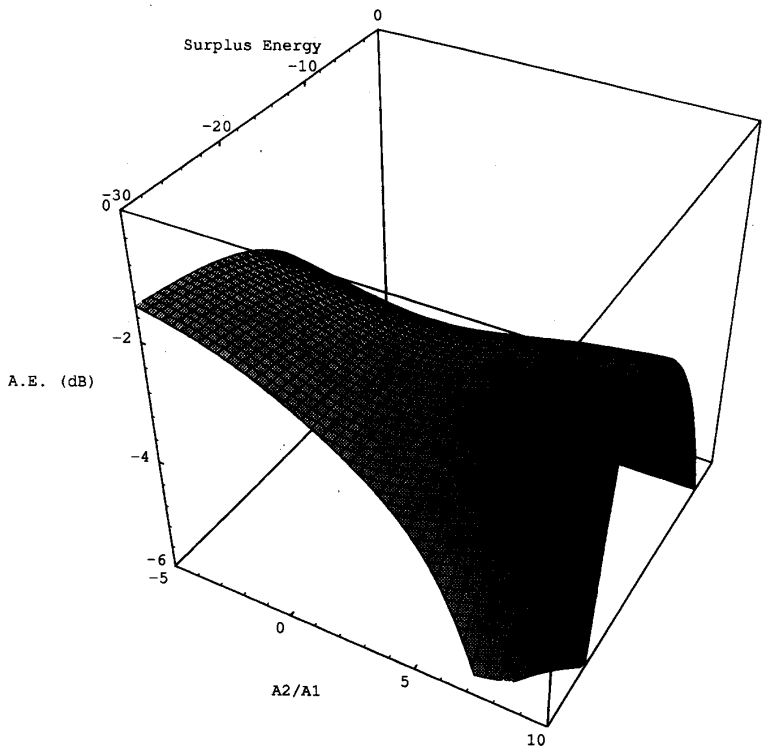

Fig. 5. Asymptotic multiuser efficiency as a function of surplus energy and interference ratio. Example 3.1 with $\epsilon = 0.5$ and $\delta = 0.1$.

choice of surplus energy the worst case asymptotic efficiency with respect to A_2/A_1 occurs roughly between -5 dB and 0 dB—a typical behavior in multiuser detection. We can see that in each of the cases we consider, values of surplus energy equal to or moderately above χ_I are excellent choices unless A_2/A_1 is extremely low.

Example 3.2: We now consider a somewhat larger system with $K = 7$ users and a processing gain $N = 10$. Each user uses a signature sequence of N chips to generate the symbol waveform. In order to avoid averaging over the relative delays of the interferers, we assume that the interferers are synchronous. On the other hand, we choose the signature sequences *randomly* rather than optimizing over deterministic sequences. We assume an observation interval of 1-bit period T_b for each bit decision. We assume that the received signal due to the desired user has two components, one main component which is aligned with the observation interval, and a multipath component which is offset by half a bit-interval ($N/2$ chips) from the observation interval. The nominal $\tilde{\boldsymbol{s}}_1$ is taken to be a scaled version of the desired spreading sequence, and the signal $\boldsymbol{s}_1$ modulating the desired bit within the observation interval $[0, T_b]$ is given by

$$s_1(t) = \frac{\hat{s_1}(t) + a\hat{s_1}(t - T_b/2)}{\|\hat{s_1}(t) + a\hat{s_1}(t - T_b/2)\|}, \qquad 0 \le t \le T_b$$

where a is the relative amplitude of the multipath component. The self-interference due to the multipath component (i.e., the part within $[0, T_b]$ of the multipath signal modulating a bit other than the desired bit) is modeled as an additional interferer with modulating waveform

$$A_{K+1}s_{K+1} = a\hat{s_1}(t + T_b/2), \qquad 0 \le t \le T_b.$$

The relative amplitude a of the multipath component dictates the extent of the mismatch between the desired signal and the nominal. In our numerical results, we consider $a = 1$, which causes a fairly large mismatch and corresponds to a minimum surplus energy for signal cancellation $\chi_S = 2.47$.

The $K-1$ interfering signals are taken to be scaled versions of randomly generated spreading sequences. For convenience, we assume that all interferers have the same amplitude A relative to the desired signal, i.e., that $A_k = A$ for $2 \le k \le K$. For the particular choice of signature sequences we consider, the minimum surplus energy for complete interference cancellation is $\chi_I = 0.6$. Since the self-interference due to multipath does not cause a near–far problem, it is ignored in the computation of χ_I (we do include this interference in computing performance measures such as SIR, however). Since $\chi_I \ll \chi_S$, a well-designed constrained minimum-energy detector is expected to perform well. In the following, we compare the SIR of the minimum-output-energy detector with mismatch (with and without explicit constraints on surplus energy) with that of the minimum-output-energy detector without mismatch and without an explicit constraint on surplus energy. As shown in Section II, the latter is also the MMSE detector.

In Fig. 7, we plot the SIR versus the SNR of the desired signal, given by $\|\boldsymbol{s}_1\|^2/\sigma^2 = \sigma^{-2}$. In the absence of mismatch (i.e., if the anchor takes into account the multipath component),

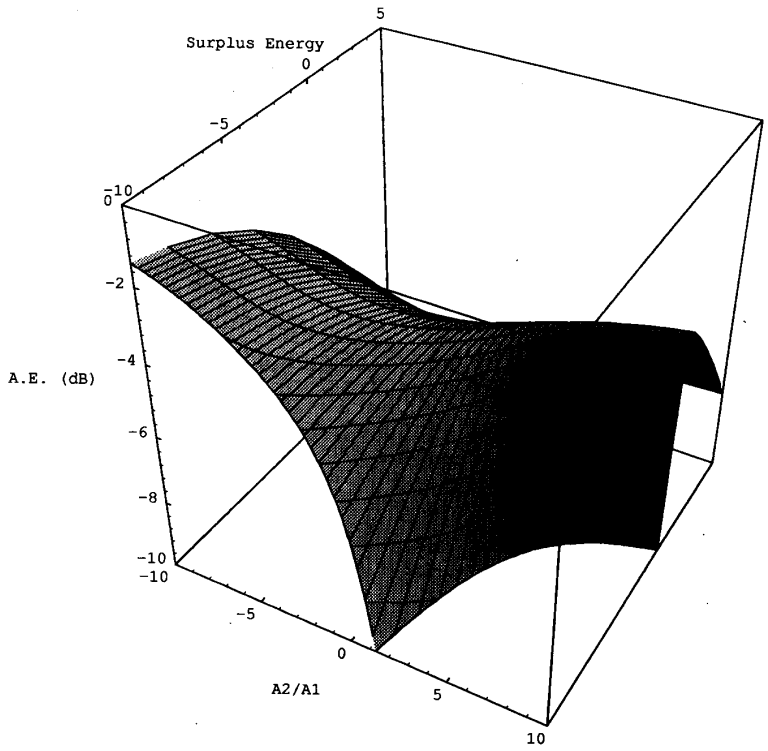

Fig. 6. Asymptotic multiuser efficiency as a function of surplus energy and interference ratio. Example 3.1 with $\epsilon = 0.1$ and $\delta = 1$.

the SIR increases almost linearly with SNR. However, with mismatch, the SIR of the minimum-energy detector degrades for SNR's beyond a certain range unless there is an explicit constraint on the surplus energy. This is because for high SNR, the low noise level allows a high value of surplus energy, which leads to signal degradation and noise enhancement. The performance improves substantially when we impose an explicit constraint on surplus energy by taking $\nu = 0.1$. This choice is motivated by the fact that $\sigma^2 = 0.1$ or an SNR of 10 dB gives reasonable SIR when no explicit constraint is used, so that an explicit constraint which maintains the same level for $\gamma = \nu_1 + \sigma^2$ as $\sigma^2 \to 0$ should prevent excessive signal degradation and noise enhancement for high SNR. Thus while mismatch does cause a deterioration in performance, the SIR is good enough to justify the use of the minimum energy algorithms as an initial blind adaptation mechanism, possibly to be followed by decision-directed adaptation based on an MMSE criterion. Fig. 8 shows the values of surplus energy χ for the constrained minimum energy detector as a function of SNR. In the absence of an explicit constraint on the surplus energy, the surplus energy grows with the SNR, leveling off at a value that permits almost complete cancellation of both desired and interfering signals. This is because for high SNR (i.e., low background noise), there is effectively no implicit constraint on the detector surplus energy. However, the surplus energy levels off much faster when an additional fictitious noise level is imposed via a Lagrange multiplier of $\nu = 0.1$ in the cost function. Thus the imposition of an appropriate explicit constraint on the surplus energy permits interference suppression without excessive signal degradation and noise enhancement.

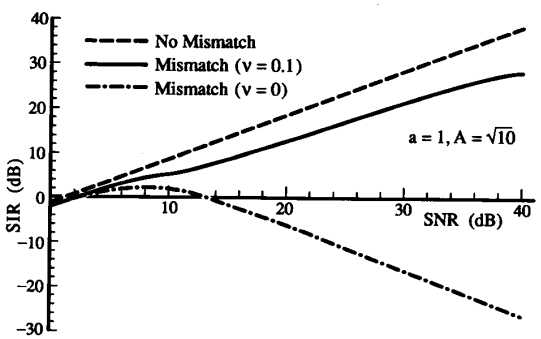

Fig. 7. Signal-to-interference ratio of blind minimum-output energy detector versus signal-to-noise ratio of desired user.

IV. Convergence Analysis of Stochastic Gradient Algorithm

In this section we analyze the convergence properties of the gradient algorithm (26). Our goal is to obtain expressions for the trajectories of the mean tap vector and the MSE as functions of the amplitudes, signature waveforms, and algorithm step-size μ (which we assume fixed throughout this section). Because the true gradient of the energy is approximated by its instantaneous value, algorithm "noise" contributes "excess" MSE beyond that achievable with a fixed optimal (minimum energy) tap vector $\boldsymbol{c}$. The asymptotic value of the MSE after convergence, together with a condition on the

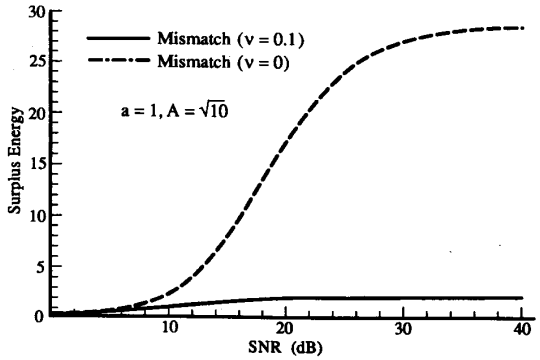

Fig. 8. Surplus energy of blind minimum output energy detector versus signal-to-noise ratio of desired user.

step size μ that guarantees convergence (i.e., finite asymptotic MSE), is therefore of interest. Throughout this section we assume a vector representation for the user waveforms, which results from projection onto a finite basis. For example, the signature signals assigned to each user can be viewed as vectors of samples of chip-matched filter outputs within a symbol period. Throughout this section lower case bold variables will denote vectors in $\mathbb{R}^N$. Here we assume a symbol-synchronous system, and no mismatch, so that $\hat{\boldsymbol{s}}_1 = \boldsymbol{s}_1$.

The analysis given here is analogous to that given in [21] for the conventional stochastic gradient (LMS) algorithm, which is patterned after the analysis given by Ungerboeck [25]. The approximations made in our analysis are similar to those made in the analysis of the LMS algorithm, and include the approximation of fourth-order statistics in terms of second-order statistics.[1] In addition, we obtain simple expressions for quantities of interest by approximating the eigenvalues of the outer product matrix $\boldsymbol{R}_y = E(\boldsymbol{y}[i]\boldsymbol{y}'[i])$. However, we also point out that the *independence assumption*, which states that the tap vector at time $i-1$ is independent of the data vector $\boldsymbol{y}[i]$, is in fact satisfied in the synchronous multiuser case considered. This is in contrast to the standard analysis of the single-user adaptive equalizer, where this assumption is not satisfied, but is assumed nevertheless for analytical tractability. We also point out that, under the independence assumption and assuming that the user signal vectors are appropriately modified, the following analysis also applies to asynchronous users.

A. Trajectory of the Mean Tap Vector

We start by computing the trajectory of the mean tap vector $E(\boldsymbol{c}[i])$. Adding $\boldsymbol{s}_1$ to both sides of (26) gives

$$\begin{aligned}\boldsymbol{c}[i] &= \boldsymbol{c}[i-1] - \mu(\boldsymbol{c}'[i-1]\boldsymbol{y}[i])\boldsymbol{v}[i] \\ &= (\boldsymbol{I} - \mu\boldsymbol{v}[i]\boldsymbol{y}'[i])\boldsymbol{c}[i-1] \end{aligned} \tag{46}$$

where

$$\boldsymbol{v}[i] = (\boldsymbol{I} - \boldsymbol{s}_1\boldsymbol{s}'_1)\boldsymbol{y}[i]. \tag{47}$$

[1] These fourth-order statistics can be computed exactly for the situation considered. However, this computation is quite involved, and approximations are still needed to derive a stability condition along with an expression for asymptotic MSE

Now define the vector $\boldsymbol{c}_{\text{opt}} = \xi_{\min}\boldsymbol{R}_y^{-1}\boldsymbol{s}_1$, and the tap vector error

$$\boldsymbol{e}[i] = \boldsymbol{c}[i] - \boldsymbol{c}_{\text{opt}}. \tag{48}$$

Rewriting (46) as

$$\begin{aligned}\boldsymbol{e}[i] &= (\boldsymbol{I} - \mu\boldsymbol{v}[i]\boldsymbol{y}'[i])\boldsymbol{e}[i-1] + (\boldsymbol{I} - \mu\boldsymbol{v}[i]\boldsymbol{y}'[i])\boldsymbol{c}_{\text{opt}} - \boldsymbol{c}_{\text{opt}} \\ &= (\boldsymbol{I} - \mu\boldsymbol{v}[i]\boldsymbol{y}'[i])\boldsymbol{e}[i-1] - \mu\boldsymbol{v}[i]\boldsymbol{y}'[i])\boldsymbol{c}_{\text{opt}} \end{aligned} \tag{49}$$

and taking expectation of both sides gives

$$E(\boldsymbol{e}[i]) = (\boldsymbol{I} - \mu\boldsymbol{R}_{vy})E(\boldsymbol{e}[i-1]) \tag{50}$$

where

$$\begin{aligned}\boldsymbol{R}_{vy} &= E(\boldsymbol{v}[i]\boldsymbol{y}'[i]) = (\boldsymbol{I} - \boldsymbol{s}_1\boldsymbol{s}'_1)\boldsymbol{R}_y \\ &= \sum_{k=1}^{K} A_k^2(\boldsymbol{s}_k - \rho_{1k}\boldsymbol{s}_1)\boldsymbol{s}'_k + \sigma^2(\boldsymbol{I} - \boldsymbol{s}_1\boldsymbol{s}'_1) \end{aligned} \tag{51}$$

where $\rho_{jk} = \boldsymbol{s}'_j\boldsymbol{s}_k$, and the fact that $\boldsymbol{R}_{vy}\boldsymbol{c}_{\text{opt}} = \boldsymbol{0}$ has been used.

We therefore conclude that $\boldsymbol{c}[i]$ converges to $\boldsymbol{c}_{\text{opt}}$ along N modes, each of which decays exponentially with parameter $1 - \mu\lambda_k^{(vy)}$, where $\lambda_k^{(vy)}$ is the kth eigenvalue of $\boldsymbol{R}_{vy}$. Since $\boldsymbol{R}_{vy}$ need not be symmetric, the eigenvalues $\lambda_k^{(vy)}$ may be complex. For stability, we must have

$$0 \le \mu < \min_k \frac{2}{|\lambda_k^{(vy)}|}. \tag{52}$$

To gain more insight into the convergence of the mean tap vector, it is necessary to study the eigenvalues of the matrix $\boldsymbol{R}_{vy}$. We first observe that $\boldsymbol{s}_1$ is an eigenvector of $\boldsymbol{R}_{vy}$ with eigenvalue $\lambda_1^{(vy)} = 0$. Consequently, the convergence of the mean tap vector is determined by the remaining $N-1$ modes. We next observe that K eigenvectors of $\boldsymbol{R}_{vy}$ lie in the space spanned by the signal vectors $\boldsymbol{s}_1, \cdots, \boldsymbol{s}_K$, and the remaining $N-K$ eigenvectors of $\boldsymbol{R}_{vy}$ are orthogonal to the signal space. The eigenvalue associated with these latter eigenvectors is σ^2. An approximation for the eigenvalues of $\boldsymbol{R}_{vy}$ corresponding to the remaining $K-1$ eigenvectors in the signal space can be obtained by observing that if the signal vectors are approximately orthogonal, then $\boldsymbol{u}'_k\boldsymbol{s}_j \approx 0$ for $k \neq j$, where $\boldsymbol{u}_k$ is the orthogonal projection of $\boldsymbol{s}_k$ onto $\boldsymbol{s}_1$, i.e., $\boldsymbol{u}_k = \boldsymbol{s}_k - \rho_{1k}\boldsymbol{s}_1$. We therefore have that

$$\boldsymbol{R}_{vy}\boldsymbol{u}_k \approx \left[A_k^2(1-\rho_{1k}^2) + \sigma^2\right]\boldsymbol{u}_k \tag{53}$$

so that the eigenvalues of $\boldsymbol{R}_{vy}$ can be approximated as

$$\lambda_k^{(vy)} \approx \begin{cases} A_k^2(1-\rho_{1k}^2) + \sigma^2, & k = 2, \cdots, K \\ \sigma^2, & k = K+1, \cdots, N. \end{cases} \tag{54}$$

Note that this approximation becomes exact as $\rho_{1k} \to 0$, $k = 1, \cdots, K$. There are, of course, other approximations for the eigenvectors of $\boldsymbol{R}_{vy}$ that could be used to obtain approximations for the corresponding eigenvalues, given that the signal vectors are approximately orthogonal. The reason for choosing the preceding approximation is that summing over the approximate eigenvalues, given by (54), gives the

correct value for $\operatorname{tr} \boldsymbol{R}_{vy}$, which will appear in the forthcoming analysis of MSE. Specifically

$$\operatorname{tr} \boldsymbol{R}_{vy} = \sum_{k=1}^{K} \lambda_k^{(vy)} = \sum_{k=2}^{K} A_k^2 \left(1 - \rho_{1k}^2\right) + (N-1)\sigma^2. \quad (55)$$

We also point out that according to the preceding approximation, taking $\mu < 2/\left(A_{\max}^2 + \sigma^2\right)$, where $A_{\max} = \max_k A_k$, satisfies the stability condition (52).

B. Trajectory of MSE

We now turn our attention to the convergence of output MSE. Let

$$\epsilon[i] = \operatorname{MSE}(\boldsymbol{x}[i]) \quad (56a)$$

and

$$\xi[i] = \operatorname{MOE}(\boldsymbol{x}[i]) \quad (56b)$$

that is, $\epsilon[i]$ and $\xi[i]$ are the MSE and mean output energy, respectively, at iteration i. First recall from (22) that

$$\begin{aligned} \epsilon[i] &= \xi[i] - 2E(\boldsymbol{c}'[i]\boldsymbol{s}_1) + E\left(b_k^2\right) \\ &= \epsilon_{\min} + \xi_{ex}[i] - 2E\left(\tilde{\boldsymbol{e}}'[i]\right)\boldsymbol{s}_1 \end{aligned} \quad (57)$$

where

$$\xi[i] = E\left(\boldsymbol{c}'[i-1]\boldsymbol{y}[i]\boldsymbol{y}[i]'\boldsymbol{c}[i-1]\right)$$

is the expected output energy at time i, $\epsilon_{\min}$ is the MSE with $\boldsymbol{c}_{opt} = \xi_{\min}\boldsymbol{R}_y^{-1}\boldsymbol{s}_1$, where $\xi_{\min} = 1/\left(\boldsymbol{s}_1'\boldsymbol{R}_y^{-1}\boldsymbol{s}_1\right)$ is the minimum output energy, and $\xi_{ex}[i] = \xi[i] - \xi_{\min}$ is the *excess* output energy due to adaptation at time i. Since $\lim_{i\to\infty} E(\tilde{\boldsymbol{e}}[i]) = \boldsymbol{0}$, we therefore have that

$$\lim_{i\to\infty} \epsilon[i] = \epsilon_{\min} + \lim_{i\to\infty} \xi[i] \equiv \epsilon_{\min} + \bar{\xi}_{ex}. \quad (58)$$

The asymptotic *excess* MSE due to adaptation is therefore equal to the asymptotic excess output energy.

We therefore focus on the trajectory of $\xi[i]$, and in particular, we are interested in the asymptotic excess energy $\bar{\xi}_{\text{ex}}$. First note that

$$\begin{aligned} \xi[i] &= E(\boldsymbol{y}[i]\boldsymbol{c}[i-1]\boldsymbol{c}'[i-1]\boldsymbol{y}[i]) \\ &= \operatorname{tr} E(\boldsymbol{c}[i-1]\boldsymbol{c}'[i-1]\boldsymbol{y}[i]\boldsymbol{y}'[i]) \\ &= \operatorname{tr} E(\boldsymbol{R}_c[i-1]\boldsymbol{R}_y) \end{aligned} \quad (59)$$

where $\boldsymbol{R}_{\boldsymbol{c}} \equiv E(\boldsymbol{c}\boldsymbol{c}')$. We therefore have that

$$\begin{aligned} \boldsymbol{R}_c[i] &= E\{(\boldsymbol{e}[i] + \boldsymbol{c}_{\text{opt}})(\boldsymbol{e}[i] + \boldsymbol{c}_{\text{opt}})'\} \\ &= \boldsymbol{R}_{\boldsymbol{c}_{\text{opt}}} + E(\boldsymbol{e}[i])\boldsymbol{c}_{\text{opt}}' + \boldsymbol{c}_{\text{opt}}E(\boldsymbol{e}'[i]) + \boldsymbol{R}_{\boldsymbol{e}}[i] \end{aligned} \quad (60)$$

where

$$\boldsymbol{R}_{\boldsymbol{c}_{\text{opt}}} = \boldsymbol{c}_{\text{opt}}\boldsymbol{c}_{\text{opt}}' = \xi_{\min}^2 \boldsymbol{R}_y^{-1}\boldsymbol{s}_1\boldsymbol{s}_1'\boldsymbol{R}_y^{-1}.$$

The following coordinate transformation will be useful. Since $\boldsymbol{R}_y$ is symmetric and nonnegative definite, we can write

$$\boldsymbol{R}_y = \Phi\Lambda\Phi' \quad (61)$$

where the columns of Φ are the orthonormal eigenvectors of $\boldsymbol{R}_y$, and Λ is the diagonal matrix of corresponding eigenvalues $\lambda_1, \cdots, \lambda_N$. Defining the rotated tap vector error

$$\tilde{\boldsymbol{e}}[i] = \Phi'\boldsymbol{e}[i] \quad (62)$$

and the rotated signal vectors

$$\tilde{\boldsymbol{y}}[i] = \Phi'\boldsymbol{y}[i] \quad \tilde{\boldsymbol{s}}_1 = \Phi'\boldsymbol{s}_1 \quad (63)$$

we have from (50),

$$E(\tilde{\boldsymbol{e}}[i]) = \left[\boldsymbol{I} - \mu\left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right)\Lambda\right]E(\tilde{\boldsymbol{e}}[i-1]). \quad (64)$$

Rewriting (60) in terms of the transformed vector $\tilde{\boldsymbol{e}}$, we first note that

$$\boldsymbol{R}_{\tilde{\boldsymbol{e}}} = E\left[\tilde{\boldsymbol{e}}\tilde{\boldsymbol{e}}'\right] = \Phi'\boldsymbol{R}_{\boldsymbol{e}}\Phi \quad (65)$$

and from (59) and (60)

$$\begin{aligned} \xi[i] &= \operatorname{tr} E(\Lambda\Phi'\boldsymbol{R}_c[i-1]\Phi) \\ &= \xi_{\min} + \operatorname{tr}\left\{E(\tilde{\boldsymbol{e}}[i])\tilde{\boldsymbol{s}}_1' + \tilde{\boldsymbol{s}}_1 E\left(\tilde{\boldsymbol{e}}'[i]\right) + \Lambda\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i]\right\}. \end{aligned} \quad (66)$$

Since $\lim_{i\to\infty} E(\tilde{\boldsymbol{e}}[i]) = 0$, it follows that $\bar{\xi}_{\text{ex}} = \lim_{i\to\infty} \operatorname{tr}\{\Lambda\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i]\}$.

The preceding results imply that to study the evolution of output MSE, it is sufficient to study the evolution of the covariance matrix $\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i]$. It is shown in the appendix that

$$\begin{aligned} \boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i] \approx\ & \boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i-1] - \mu\left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right)\Lambda\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i-1] \\ & - \mu\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i-1]\Lambda\left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right) \\ & + \mu^2\left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right)\Lambda\left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right) \\ & \cdot \left(\operatorname{tr}\left(\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i-1]\Lambda\right) + 2\xi_{\min}E\left(\tilde{\boldsymbol{e}}'[i-1]\right)\tilde{\boldsymbol{s}}_1\right) \\ & + \mu^2\xi_{\min}\left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right)\Lambda\left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right) \end{aligned} \quad (67)$$

We now observe that if the signal vectors are approximately orthogonal, then the first K eigenvectors of $\boldsymbol{R}_y$ can be approximated as $\boldsymbol{s}_1, \cdots, \boldsymbol{s}_K$. Since the remaining eigenvectors of $\boldsymbol{R}_y$ are orthogonal to the signal space, $[\tilde{\boldsymbol{s}}_1]_j \approx 0$, $j \neq 1$, and $[\tilde{\boldsymbol{s}}_1]_1 \approx 1$. To proceed, we therefore make the approximation that the matrix $\tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'$ is diagonal, so that $\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i]$ is approximately diagonal. Define the N-vector $\boldsymbol{r}_{\tilde{\boldsymbol{e}}}[i]$ with elements equal to the diagonal elements of $\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i]$. After some manipulation, we can rewrite (67) as

$$\begin{aligned} \boldsymbol{r}_{\tilde{\boldsymbol{e}}}[i] \approx\ & \boldsymbol{B}\boldsymbol{r}_{\tilde{\boldsymbol{e}}}[i-1] + \mu^2\xi_{\min}\left(2E\left(\tilde{\boldsymbol{e}}'[i-1]\right)\tilde{\boldsymbol{s}}_1 + 1\right) \\ & \cdot \left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right)^2\boldsymbol{\lambda} \end{aligned} \quad (68)$$

where

$$\boldsymbol{B} = \boldsymbol{I} - 2\mu\left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right)\Lambda + \mu^2\left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right)^2\boldsymbol{\lambda}\boldsymbol{\lambda}' \quad (69)$$

and $\boldsymbol{\lambda}$ is the N-vector containing the eigenvalues of $\boldsymbol{R}_y$.

Since $E(\tilde{\boldsymbol{e}}[i])$ converges to zero, to guarantee stability of the preceding difference equation it is sufficient that all eigenvalues of $\boldsymbol{B}$ have magnitude less than one, which is true if the row sums of $\boldsymbol{B}$ are less than one. This implies that for stability

$$\mu < \frac{2}{\sum_{k=1}^{N} \lambda_k} = \frac{2}{\sum_{k=1}^{K} A_k^2 + N\sigma^2} \quad (70)$$

which is the same stability condition for the conventional LMS algorithm, which could be used to adapt the vector $\boldsymbol{c}$ with a training sequence, and is a considerably more stringent condition than (52). Letting $i \to \infty$ in (68), using the fact that $\lim_{i\to\infty} \boldsymbol{\lambda}' \boldsymbol{r}_{\tilde{\boldsymbol{e}}}[i] = \bar{\xi}_{\text{ex}}$, and rearranging gives

$$\lim_{i\to\infty} \boldsymbol{r}_{\tilde{\boldsymbol{e}}}[i] \approx \frac{\mu}{2}\left(\xi_{\min} + \bar{\xi}_{ex}\right)\Lambda^{-1}\left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right)\boldsymbol{\lambda}. \tag{71}$$

Multiplying both sides by Λ and summing components gives

$$\bar{\xi}_{\text{ex}} \approx \frac{\mu}{2}\left(\xi_{\min} + \bar{\xi}_{\text{ex}}\right)\left[\mathbf{1}'\left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right)\boldsymbol{\lambda}\right] \tag{72}$$

where $\mathbf{1}$ is the N-vector with elements equal to one. Approximating

$$\begin{aligned}\mathbf{1}'\left(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\right)\boldsymbol{\lambda} &= \operatorname{tr}\boldsymbol{R}_y - \sum_{n=1}^{N}\sum_{m=1}^{N}\lambda_n\tilde{\boldsymbol{s}}_{1,n}\tilde{\boldsymbol{s}}_{1,m} \\ &\approx \operatorname{tr}\boldsymbol{R}_y - \sum_{n=1}^{N}\lambda_n\tilde{\boldsymbol{s}}_{1,n}^2 = \operatorname{tr}\boldsymbol{R}_{vy}\end{aligned} \tag{73}$$

we have that

$$\bar{\xi}_{\text{ex}} \approx \xi_{\min}\frac{\frac{\mu}{2}\operatorname{tr}\boldsymbol{R}_{vy}}{1 - \frac{\mu}{2}\operatorname{tr}\boldsymbol{R}_{vy}} \tag{74}$$

where $\operatorname{tr}\boldsymbol{R}_{vy}$ is given by (55).

C. Comparison with LMS Algorithm with Training Sequence

We now compare the preceding results with the analogous results for the conventional LMS algorithm with a training sequence, which is given by

$$\boldsymbol{c}[i] = \boldsymbol{c}[i-1] - \mu e_c[i]\boldsymbol{y}[i] \tag{75}$$

where the error $e_c[i] = b_1[i] - \boldsymbol{c}'[i-1]\boldsymbol{y}[i]$, and $b_1[i]$ is the transmitted symbol for user 1 at time i. It is well known [21] that the mean tap vector converges to $\boldsymbol{c}_{\text{opt}} = \boldsymbol{R}_y^{-1}\boldsymbol{s}_1$ along N normal modes, each corresponding to the eigenvalues of $\boldsymbol{I} - \mu\boldsymbol{R}_y$ (cf. [26]). In contrast, for the blind algorithm (26), we have shown that the mean tap vector converges to $\boldsymbol{c}_{\text{opt}} = \xi_{\min}\boldsymbol{R}_y^{-1}\boldsymbol{s}_1$ along N normal modes corresponding to the eigenvalues of $\boldsymbol{I} - \mu\boldsymbol{R}_{vy}$. If the signal vectors are approximately orthogonal, then according to the preceding discussion, $N-1$ eigenvalues of $\boldsymbol{R}_y$ and $\boldsymbol{R}_{vy}$ are given approximately by (54), where $\rho_{1k} \approx 0$, $k \neq 1$. However, for $\boldsymbol{R}_y$, $\lambda_1 \approx A_1^2 + \sigma^2$, whereas for $\boldsymbol{R}_{vy}$, $\lambda_1^{(vy)} = 0$.

The asymptotic excess MSE for the conventional LMS algorithm is given approximately by

$$\frac{1}{2}\mu\epsilon_{\min}(\operatorname{tr}\boldsymbol{R}_y)/\left(1 - \frac{1}{2}\mu\operatorname{tr}\boldsymbol{R}_y\right)$$

and for the same μ is significantly smaller than the asymptotic MSE for the blind algorithm, given by (74). This is because $\epsilon_{\min} \ll \xi_{\min}$ for small levels of background noise. Specifically, when the signal vectors are approximately orthogonal

$$\xi_{\min} = \frac{1}{\tilde{\boldsymbol{s}}_1'\Lambda^{-1}\tilde{\boldsymbol{s}}_1} \approx \lambda_1 \approx A_1^2 + \sigma^2$$

whereas it is easily shown that

$$\epsilon_{\min} = 1 - A_1^2\tilde{\boldsymbol{s}}_1'\Lambda^{-1}\tilde{\boldsymbol{s}}_1 \approx 1 - \frac{A_1^2}{A_1^2 + \sigma^2}.$$

Even though $\operatorname{tr}\boldsymbol{R}_{vy}$ and $\operatorname{tr}\boldsymbol{R}_y$ may be close, if σ^2 is close to zero, the difference between $\xi_{\min}$ and $\epsilon_{\min}$ is likely to be substantial. Consequently, (74) implies that the blind gradient algorithm (26) is quite "noisy," and it is therefore best to switch to a decision directed algorithm as soon as possible. This will be illustrated in Section IV-D, which contains a numerical example.

We observe from the preceding discussion that one way to improve upon the dynamics of the stochastic gradient algorithm (26) is to use a different cost function. Namely, each component of the stochastic driving term for the algorithm (26) has variance on the order of $\xi_{\min}$ whereas each component of the stochastic driving term for the conventional LMS algorithm has variance on the order of $\epsilon_{\min}$. This explains why the blind algorithm (26) performs worse than the conventional LMS algorithm with a training sequence. We can, however, replace the energy cost function by other cost functions which are driven close to zero when $\boldsymbol{c}$ is chosen optimally (i.e., $[\boldsymbol{c}'\boldsymbol{y} - \operatorname{sgn}(\boldsymbol{c}'\boldsymbol{y})]^2$). However, this may introduce local minima, i.e., $\boldsymbol{c}$ may adapt to an interferer rather than to the desired user. However, if the signal vectors are nearly orthogonal, then the orthogonal decomposition of the tap vector described in Section II guarantees that the $\boldsymbol{c}$ which achieves a local minimum must have a very large norm, and can therefore be rejected by an appropriate norm constraint.

D. Simulation Results

1) Stochastic Gradient Algorithm: Fig. 9 shows a plot of averaged SIR versus time assuming the algorithm (26) is used in a synchronous CDMA system with processing gain $N = 10$ and number of users $K = 7$. Averaged SIR at the ith iteration is given by

$$\text{SIR}_{\text{av}}[i] = \frac{\sum_{r=1}^{M}\left(\boldsymbol{c}_r'[i]\boldsymbol{s}_1\right)^2}{\sum_{r=1}^{M}\boldsymbol{c}_r'[i]\left(\boldsymbol{y}_r[i] - b_{1,r}[i]\boldsymbol{s}_1\right)]^2}$$

where $A_1 = 1$, the number of algorithm runs is $M = 100$, and the subscript r indicates that the associated variable depends on the particular run. The signature sequences are the same randomly picked sequences used to generate the numerical results for Example 3.2. As explained in Section III, there is a multipath component associated with the desired user. The signal power to background noise power is 20 dB.

The interfering amplitudes $A_2, \cdots, A_7$ are each 20 times A_1, representing an extreme near–far situation. Because of the strong interference, conventional single-user blind equalization algorithms (i.e., those discussed in [14]) do not succeed in isolating user 1. Two plots are shown in Fig. 9 corresponding to no mismatch and a mismatched nominal. The desired signal contains the same multipath component as that in Example 3.2. Namely, the nominal signal is the normalized sum of the spreading sequence of the desired user plus the part of

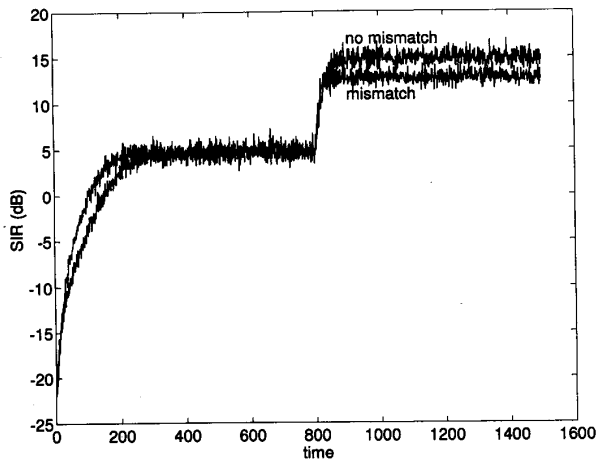

Fig. 9. Averaged SIR versus time for the stochastic gradient algorithm (26) with and without a mismatched nominal. The simulation parameters are specified in Section IV-D1.

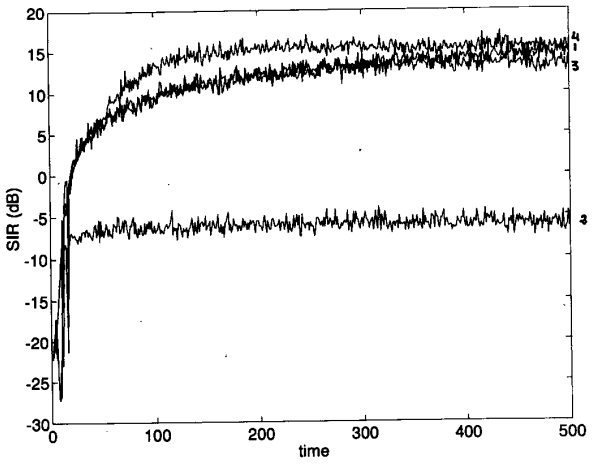

Fig. 10. Averaged SIR versus time for the least squares algorithm (79)–(81). The cases simulated are described in Section IV-D2.

the multipath component multiplied by the same bit. The plot with mismatch assumes that the nominal tap vector is equal to the spreading sequence of the desired user (neglecting the multipath component) plus an additive Gaussian perturbation where the variance of each component is 0.01. This latter type of mismatch models finite precision effects. The blind algorithm (26) is used for the first 800 iterations, and the conventional LMS algorithm in decision directed mode is used thereafter.

In both cases shown in Fig. 9 the blind algorithm succeeds in suppressing the strong interferers, and drives the SIR above 0 dB. What is interesting is that the mismatch creates an initial condition for the conventional LMS algorithm which leads to a lower SIR than the case without mismatch. Additional simulation results show that different mismatches lead to different SIR's. The explanation for this is that the mismatch causes the tap vector to wander outside the space spanned by the actual signal vectors, and thereby creates an orthogonal component to the signal space which takes an extremely long time to suppress with a training sequence if the background noise is very small. This is an inherent problem with the LMS algorithm with a training sequence, and can be handled by tap leakage [23].

2) Simulation Results—Least Squares Algorithm: As an alternative to the stochastic gradient algorithm (26), one could instead select the tap vector $\boldsymbol{c}$ that achieves

$$\min \sum_{j=1}^{i} (\boldsymbol{c}'\boldsymbol{y}[i])^2 \tag{76}$$

subject to

$$\boldsymbol{c}'\boldsymbol{s}_1 = 1. \tag{77}$$

In the presence of mismatch, we add the constraint

$$||\boldsymbol{c}||^2 = 1 + \chi \tag{78}$$

where χ is chosen according to the guidelines set in Section III. The solution to this optimization problem is

$$\boldsymbol{c}[i] = \hat{\xi}_{\min}[i]\hat{\boldsymbol{R}}_y^{-1}[i]\boldsymbol{s}_1 \tag{79}$$

where

$$\hat{\boldsymbol{R}}_y[i] = \sum_{j=0}^{i} \boldsymbol{y}[j]\boldsymbol{y}'[j] + \nu\boldsymbol{I} \tag{80}$$

$$\hat{\xi}_{\min}[i] = \left(\boldsymbol{s}_1'\hat{\boldsymbol{R}}_y^{-1}[i]\boldsymbol{s}_1\right)^{-1} \tag{81}$$

and ν is selected to satisfy the constraint (78). Note that as ν decreases, χ increases. Comparing (79)–(81) with (32), we observe that the least squares (LS) solution for $\boldsymbol{c}$ has the same form as the optimal solution (32) where expectations are replaced by time averages.

Fig. 10 shows averaged SIR versus time for the LS solution (79), assuming the same parameters as were used to generate Fig. 9. The following four cases were simulated:

Case 1 (No mismatch, $\nu = 0.01$): Since ν is very small, the surplus energy is very large. The LS algorithm drives the SIR to 5 dB in less than 50 iterations, which is roughly four times faster than the convergence time of the stochastic gradient algorithm shown in Fig. 9. Because the LS solution in (79) does not have a forgetting factor (i.e., does not exponentially weight the data), the tap vector converges to the MMSE solution, so that the asymptotic SIR is 20 dB.

Case 2 (Mismatch, $\nu = 0.01$): In this case the same mismatched nominal without the multipath component is used, as was assumed in Example 3.2. The steady-state SIR is -7 dB since the allowed surplus energy is large enough to suppress most of the desired signal.

Case 3 (Mismatch, $\nu = 100$): The surplus energy in this case is much smaller than for the preceding case. The performance of the blind LS algorithm is nearly identical to the performance shown in the first case without mismatch. The only difference is that without mismatch the tap vector converges to the MMSE solution, whereas with mismatch the

tap vector converges to another solution which lowers the asymptotic SIR. However, Fig. 10 indicates that this difference in SIR is very small after 500 iterations.

Case 4 (Mismatch, $\nu = 100$, Switch to decision-directed mode): This case is the same as that just considered except that the algorithm switches to an LMS algorithm used in decision-directed mode after 50 iterations. The steady-state performance is only slightly better than that of the blind LS algorithm.

V. Appendix
Derivation of (67)

Premultiplying both sides of (49) by Φ', we can compute

$$\begin{aligned}\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i] &= E\big(\tilde{\boldsymbol{e}}[i]\tilde{\boldsymbol{e}}'[i]\big)\\ &= E\big(\boldsymbol{P}[i]\tilde{\boldsymbol{e}}[i-1]\tilde{\boldsymbol{e}}'[i-1]\boldsymbol{P}[i]\big)\\ &\quad - \mu\xi_{\min}E\big(\boldsymbol{P}[i]\tilde{\boldsymbol{e}}[i-1]\tilde{\boldsymbol{s}}_1'\Lambda^{-1}\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{v}}'[i]\big)\\ &\quad - \mu\xi_{\min}E\big(\tilde{\boldsymbol{v}}[i]\tilde{\boldsymbol{y}}'[i]\Lambda^{-1}\tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{e}}'[i-1]\boldsymbol{P}[i]\big)\\ &\quad + \mu^2\xi_{\min}^2E\big(\tilde{\boldsymbol{v}}[i]\tilde{\boldsymbol{y}}'[i]\Lambda^{-1}\tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\Lambda^{-1}\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{v}}'[i]\big)\end{aligned} \tag{A1}$$

where $\boldsymbol{P} = \boldsymbol{I} - \mu\tilde{\boldsymbol{v}}[i]\tilde{\boldsymbol{y}}'[i]$, and variables with tildes indicate premultiplication by Φ'.

Examining the first term on the right,

$$\begin{aligned}E(\boldsymbol{P}[i]\tilde{\boldsymbol{e}}'[i-1]\tilde{\boldsymbol{e}}'[i-1]\boldsymbol{P}[i]) &= \boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i-1] - \mu E(\tilde{\boldsymbol{v}}[i]\tilde{\boldsymbol{y}}'[i])\\ &\quad \boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i-1] - \mu\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i-1]E(\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{v}}'[i])\\ &\quad + \mu^2E\big(\tilde{\boldsymbol{v}}[i]\tilde{\boldsymbol{y}}'[i]\tilde{\boldsymbol{e}}[i-1]\tilde{\boldsymbol{e}}'[i-1]\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{v}}[i]\big)\\ &= \boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i-1] - \mu\big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)\Lambda\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i-1]\\ &\quad - \mu\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i-1]\Lambda\big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)\\ &\quad + \big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)E\big(\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{y}}'[i]\\ &\quad \cdot \boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i-1]\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{y}}'[i]\big)\big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big).\end{aligned} \tag{A2}$$

Assuming that the correlations between different components of $\tilde{\boldsymbol{y}}$ and between components of $\tilde{\boldsymbol{e}}$ are small, the last expectation can be approximated as (see [21, eq. (7.1.26)])

$$E\big(\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{y}}'[i]\tilde{\boldsymbol{e}}[i-1]\tilde{\boldsymbol{e}}'[i-1]\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{y}}[i]\big) \approx \Lambda\,\mathrm{tr}\,(\boldsymbol{R}_{\tilde{\boldsymbol{e}}}[i-1]\Lambda). \tag{A3}$$

Examining the second term on the right of (A1)

$$\begin{aligned}E\big(\boldsymbol{P}\tilde{\boldsymbol{e}}[i-1]\tilde{\boldsymbol{s}}_1'\Lambda^{-1}\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{v}}'[i]\big) &= E(\tilde{\boldsymbol{e}}[i-1])\tilde{\boldsymbol{s}}_1'\Lambda^{-1}\Lambda\big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)\\ &\quad - \mu E\big(\tilde{\boldsymbol{v}}[i]\tilde{\boldsymbol{y}}'[i]\tilde{\boldsymbol{e}}[i-1]\tilde{\boldsymbol{s}}_1'\Lambda^{-1}\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{v}}[i]'\big)\\ &= -\mu\big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)E\big(\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{y}}'[i]\tilde{\boldsymbol{e}}[i-1]\\ &\quad \cdot \tilde{\boldsymbol{s}}_1'\Lambda^{-1}\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{y}}'[i]\big)\big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)\\ &\approx -\mu\big(E\big(\tilde{\boldsymbol{e}}'[i-1]\big)\tilde{\boldsymbol{s}}_1\big)\big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)\Lambda\big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)\end{aligned} \tag{A4}$$

where the last approximation is analogous to the approximation (A3). Finally, the last term on the right of (A1) can be approximated as

$$\begin{aligned}E\big(\tilde{\boldsymbol{v}}[i]\tilde{\boldsymbol{y}}'[i]\Lambda^{-1}\tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\Lambda^{-1}\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{v}}'[i]\big) &= \big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)\\ &\quad \cdot E\big(\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{y}}'[i]\Lambda^{-1}\tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\Lambda^{-1}\tilde{\boldsymbol{y}}[i]\tilde{\boldsymbol{y}}'[i]\big)\big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)\\ &\approx \big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)\Lambda\big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)\big(\tilde{\boldsymbol{s}}_1'\Lambda^{-1}\tilde{\boldsymbol{s}}_1'\big)\\ &= \frac{1}{\xi_{\min}}\big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big)\Lambda\big(\boldsymbol{I} - \tilde{\boldsymbol{s}}_1\tilde{\boldsymbol{s}}_1'\big).\end{aligned} \tag{A5}$$

Combining (A1)-(A5) gives (67).

References

[1] S. Verdú, "Minimum probability of error for asynchronous gaussian multiple-access channels," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 85–96, Jan. 1986.

[2] ——, "Optimum multiuser asymptotic efficienc," *IEEE Trans. Commun.*, vol. COM-34, pp. 890–897, Sept. 1986.

[3] R. Lupas and S. Verdú, "Linear multiuser detectors for synchronous code-divsion multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 35 pp. 123–136, Jan. 1989.

[4] ——, "Near-far resistance of multiuser detectors in asynchronous channels," *IEEE Trans. Commun.*, vol. 38, Mar. 1990.

[5] S. Verdú, "Multiuser detection," in *Advances in Detection and Estimation*. JAI Press, 1993.

[6] U. Madhow and M. Honig, "MMSE interference suppression for direct-sequence spread spectrum CDMA," *IEEE Trans. Commun*, vol. 42, pp. 3178–3188, Dec. 1994.

[7] P. Rapajic and B. Vucetic, "A linear adaptive fractionally spaced single user receiver for asynchronous CDMA systems," in *IEEE Int. Symp. on Information Theory* (San Antonio, TX, Jan. 1993), p. 45.

[8] M. Abdulrahman, D. D. Falconer, and A. U. H. Sheikh, "Equalization for interference cancellation in spread spectrum multiple access systems," in *Proc. IEEE Vehicular Technology Conf.* (Denver, CO, May, 1992).

[9] S. L. Miller, "An adaptive direct-sequence code-division multiple-access receiver for multiuser interference rejection," *IEEE Trans. Commun.*, to appear.

[10] S. Verdú, "Adaptive multiuser detection," in *Proc. IEEE Int. Symp. on Spread Spectrum Theory and Applications* (Oulu, Finland, July 1994).

[11] Z. Xie, R. T. Short, and C. K. Rushforth, "A family of suboptimum detectors for coherent multiuser communications," *IEEE J. Selected Areas Commun.*, pp. 683–690, May 1990.

[12] H. Oda and Y. Sato, "A method of multidimensional blind equalization," in *IEEE Int. Symp. on Information Theory* (San Antonio, TX, Jan. 1993), pp. 327.

[13] S. Verdú, B. D. O. Anderson, and R. Kennedy, "Blind equalization without gain identification," *IEEE Trans. Inform. Theory*, vol. 39, pp. 292–297, 1993.

[14] C. R. Johnson, "Admissibility in blind adaptive equalization," *IEEE Contr. Syst. Mag.*, vol. 11, pp. 3–15, Jan. 1991.

[15] K. Fukawa and H. Suzuki, "Orthogonalizing matched filter (OMF) detection for DS-CDMA mobile radio systems," in *Proc. 1994 Globecom* (San Francisco, CA, Nov. 28–Dec. 1, 1995), pp. 385-389.

[16] M. L. Honig, U. Madhow, and S. Verdú, "Blind adaptive interference suppression for near-far resistant CDMA," in *Proc. 1994 Globecom* (San Francisco, CA, Nov. 28–Dec. 1, 1995), pp. 379-384.

[17] C. A. Baird, Jr., "Recursive minimum variance estimation for adaptive sensor arrays," in *Proc. 1972 IEEE Int. Conf. on Cybernetics and Society* (Washington, DC, Oct. 9–12, 1972), pp. 412-414.

[18] D. H. Johnson, and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[19] M. L. Honig, "Orthogonally anchored interference suppression using the Sato cost criterion," in *1995 IEEE Int. Symp. on Information Theory*, to be published.

[20] S. Verdú, *Recent Progress in Multiuser Detection Advances in Communications and Control Systems*. Berlin-Heidelberg, Germany: Springer-Verlag, 1989, pp. 66–77. Reprinted in *Multiple Access Communications*, N. Abramson, Ed. Piscataway, NJ: IEEE Press, 1993.

[21] M. L. Honig, and D. G. Messerschmitt, *Adaptive Filters: Structures, Algorithms and Applications*. Boston, MA: Kluwer, 1984.

[22] L. Gyorfi, "Adaptive linear procedures under general conditions," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 262–267, Mar. 1984.

[23] R. D. Gitlin, H. C. Meadows, and S. B. Weinstein, "The tap-leakage algorithm: An algorithm for the stable operation of a digitally implemented, fractionally spaced, adaptive equalizer," *Bell Syst. Tech. J.*, vol. 61, pp. 1817–1839, Oct. 1982.

[24] U. Madhow and M. L. Honig, "MMSE detection of direct-sequence CDMA signals: analysis for random signature sequences," in *Proc. IEEE Int. Symp. on Information Theory* (San Antonio, TX, Jan. 1993).

[25] G. Ungerboeck, "Theory on the speed of convergence in adaptive equalizers for digital communications," *IBM J. Res. Devel.*, pp. 546–555, Nov. 1972.

[26] S. Miller, "Transient behavior of the minimum mean-squared error receiver for direct-sequence code-division multiple-access systems," presented at MILCOM '94. Oct. 1994.