Abstract- Nongaussianity maximization based independent component analysis (ICA) algorithms are becoming increasingly popular in signal processing and communications. These algorithms exploit higher order statistics (HOS) of the data either by an explicit computation or implicitly by nonlinearly transforming the observations. Learning algorithms based on explicit HOS are simple to analyze, but exhibit slow convergence. Introduction of suitable nonlinearities in the algorithm increases the convergence rate and robustness to outliers. In this paper, we propose a general nonlinearity based algorithm for blind adaptive multiuser detection in multipath channels. Implementation of the proposed HOS based detector is done via an extension of a recently proposed constrained less-complete ICA framework. Computer simulations illustrate the improved performance over the previously proposed kurtosis based detector.

I. INTRODUCTION

Correlation and multipath propagation introduces multiple access interference (MAI) in a direct sequence-code division multiple access (DS-CDMA) system. Signature codes are designed to have controlled cross-correlation properties, however, multipath propagation destroys these properties and renders the system highly susceptible to MAI. Power imbalance between different users worsens the situation and gives rise to the near-far problem [1]. Conventional detector is not near-far resistant and exhibits serious degradation of performance under these conditions. Multiuser detection [1] techniques attempt to mitigate these problems and can substantially increase the capacity [2] of a CDMA system.

Blind multiuser detection algorithms are attractive because they need minimal information about the structure of MAI and eliminate the dependence on training symbols which are clearly a waste of precious bandwidth. In [3] a blind multiuser detector based on constrained minimization of output energy (MOE) was proposed. It was however observed that the performance of this detector was quite sensitive to the signature waveform mismatch. In a realistic multipath environment, where this mismatch is inevitable, the performance of this detector was shown to degrade considerably.

A different approach to deal with the multipath problem was proposed in [4], where the output variance was minimized under multiple constraints to null out the contributions from the delayed copies of desired signal. In doing so, however, this method does not exploit all the energy of desired user and results in suboptimal performance. Nevertheless, it was suggested by the same authors in [5], [6] to generalize the constraint vector to a general parameter vector, which can arbitrarily be chosen so as to maximize the contribution from the desired user after the MAI was suppressed. By doing so, they are able to simultaneously capture the channel variations. This class of detectors was termed as linearly constrained minimum variance (LCMV) detectors. However, these detectors are based on the second order statistics (SOS) of the data and tend to ignore the higher order statistical information.

ICA [7] is a recent, HOS based statistical technique that attempts to make the observations independent or as independent as possible. ICA algorithms possess the capability to separate the observed mixtures into constituent independent components under the assumption of an appropriate signal model. This is more commonly known as the blind source separation (BSS) problem. The problem of ICA or BSS is however, ill-posed, in the sense that solutions are obtained subject to a scaling and permutation ambiguity [8]. To resolve these ambiguities a multiuser detection approach was suggested in [9] where ICA was used as an add on to the SOS based detector. Other constrained approaches utilizing a priori information about the desired user’s signature codes were proposed in [10], [11].

A recent approach for joint blind multiuser detection and channel estimation based on HOS of the data, was suggested in [12], where the absolute value of the kurtosis of the data was minimized with respect to the detector and then maximized under multiple constraints. This detector, however, involves the explicit computation of kurtosis, which is known to be sensitive to outliers in data in addition to being computationally complex. The resulting algorithm suffers from a slow convergence problem.

In this paper, instead of computing kurtosis directly we propose to introduce a suitable nonlinearity in the learning algorithm, hence implicitly realizing a HOS based algorithm. The proposed algorithm is computationally simpler and more robust to outliers in the data. With a proper selection of nonlinearity, these algorithms are shown to converge faster.
than those based on kurtosis [13]. Another advantage of this approach is that we are less restricted in the choice of the contrast function. With a specific form of nonlinearity, the proposed algorithm reduces to the method reported in [12]. To resolve the ambiguities involved in the proposed algorithm, we propose to extend the recently proposed less-complete constrained ICA framework [14], [15] to include the constraints proposed in [5] in the context of LCMV detectors. These additional constraints are based on a priori information about the desired user’s signature code and timing.

II. SYSTEM MODEL

In this paper, we assume a signal model involving $p$ users, transmitting independent and identically distributed (iid) data. Let $b_i(n)$ denote the data of $i^{th}$ user at time $n$, where $1 \leq i \leq p$. We assume that all the $p$ users in signal model are mutually independent. This is a strong hypothesis but very plausible in practice for physically separated sources. If we denote multiuser data in a column vector $\mathbf{b}(n) = [b_1(n), b_2(n), \ldots, b_p(n)]^T$, the independence assumption states that joint pdf $r(\mathbf{b})$ of $\mathbf{b}(n)$ is a product of marginal pdfs of the individual sources

$$r(\mathbf{b}) = \prod_{i=1}^{p} r_i(b_i). \tag{1}$$

In DS-CDMA systems each user transmits its data in digital form after modulating it by a distinct signature (spreading) sequence $c_i(k)$ of length $P$. Let $i^{th}$ user’s modulated data be transmitted through a linear channel with a baseband impulse response $g_i(n)$. Assuming a chip-rate sampling at receiver the $i^{th}$ user’s received data can be expressed as

$$y_i(n) = \sum_{l=-\infty}^{\infty} b_i(l) h_i(n - d_i - 1P), \tag{2}$$

$$h_i(n) = \sum_{m=\infty}^{\infty} A_i g_i(m) c_i(n - m), \tag{3}$$

where $A_i$ is the amplitude of the $i^{th}$ user, $h_i(n)$ is the effective signature sequence for the $i^{th}$ user. $g_i(n)$ is assumed to have a finite impulse response of maximum order $q$ (typically $q \ll P$) which includes the effects of the transmitting and the receiving filters. $d_i$ is the channel delay parameter for $i^{th}$ user in chip periods $T_c$. We assume, without loss of generality that the delay $d_i$ is of the order of one symbol period, i.e., $0 \leq d_i < P$. The total received signal $y(n)$ at the receiver is given by the superposition of signals from all active users in the channel

$$y(n) = \sum_{i=1}^{p} y_i(n) + \xi(n), \tag{4}$$

where $\xi(n)$ is additive white Gaussian noise (AWGN) with zero-mean and variance $\sigma^2$. For comparison purposes, we adopt the signal model of [6]. Let us collect $L = P + q$ samples of the received data in a vector as $y(n) = [y(nP), \ldots, y(nP + L - 1)]^T$. Then the composite signal model in a matrix form can be represented as

$$y(n) = \mathcal{H} \mathbf{B}(n) + \Xi(n), \tag{5}$$

where $\Xi(n) = [\xi_0(n), \ldots, \xi_{L-1}(n)]^T$ is a zero mean, $L$ dimensional AWGN vector. Assuming the receiver is synchronized to user 1, then contribution to the composite signal due to user 1 ($d_1 = 0$) is given by

$$y_1(n) = h_1 b_1(n) + h_1 b_1(n - 1) + \tilde{h}_1 b_1(n + 1),$$

where $y_1(n) = [y_1(nP), \ldots, y_1(nP + L - 1)]^T$ is the collection of measurements of $y_1(n)$, $h_1 = [h_1(0), \ldots, h_1(P + q - 1)]^T$ is the effective signature code of user 1, $\tilde{h}_1 = [0, \ldots, 0, h_1(0), \ldots, h_1(q - 1)]^T$ and $h_1 = [0, \ldots, 0, h_1(0), \ldots, h_1(q - 1)]^T$ are the signatures of other $p - 1$ interfering users. The formulation given above dictates the form of $(P + q) \times (2p + 1)$ matrix $\mathcal{H}$ in (5) as

$$\mathcal{H} = [h_1, \tilde{h}_1, \tilde{h}_2, \ldots, \tilde{h}_p, \tilde{h}_p], \tag{6}$$

and $(2p + 1) \times 1$ vector $\mathbf{B}(n)$ in (5) takes the form of

$$\mathbf{B}(n) = [b_1(n), b_1(n - 1), b_1(n + 1), b_2(n), b_2(n - 1), \ldots, b_p(n), b_p(n - 1)]^T \tag{7}$$

At this time, let us assume knowledge of the spreading code and the timing of the desired user, similar assumptions were made in [3]–[6], [12]. This facilitates the description of the desired user’s effective signature sequence as

$$h_1 = \mathbf{C}_1 \mathbf{g}_1, \tag{9}$$

where

$$\mathbf{C}_1 = \begin{bmatrix}
\cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & \cdots & \cdots & \cdots \\
\end{bmatrix}_{(P+q)\times(q+1)}, \tag{10}$$

and

$$\mathbf{g}_1 = \begin{bmatrix}
g_1(0) \\
\vdots \\
g_1(q) \\
\end{bmatrix}_{(q+1)\times 1},$$

is the unknown multipath channel response for user 1. Finally, we would like to mention that in the specific case of the CDMA downlink, the channel response as seen by any active user in the channel is $\mathbf{g}_1 = \mathbf{g}_2, \ldots, \mathbf{g}_p = \mathbf{g}$ and the system model reduces to a synchronous system.
III. INDEPENDENT COMPONENT ANALYSIS

ICA [7] is a generalization of the SOS based principal component analysis (PCA) technique. A key assumption in ICA is that of the independence of the underlying sources. It is a blind statistical techniques that optimizes a so called contrast function [7] to make the observed mixtures independent or as independent as possible. Under an appropriate signal model ICA can be used to solve the BSS problem.

ICA algorithms that exploit non-gaussianity require that there exists at most one Gaussian source [7]. Algorithms that exploit temporal correlation structure of the sources are not bound to this assumption, while independence is still key. Non-gaussianity maximization based ICA algorithms exploit HOS of the data. These algorithms either compute explicit HOS, for example the fourth order cumulants (kurtosis), cumulant tensors etc. to extract independent components [16]–[18], or implicitly by non-linearly transforming the observations [13], [19]–[21]. Implicit HOS based algorithms are attractive by virtue of their computational efficiency and relative robustness to the outliers in the data.

The performance of ICA algorithms involving nonlinearities, however, critically depends on the choice of nonlinearity. In the ideal case, the nonlinearity should match the cumulative density function of the underlying source. In practice, most ICA algorithms choose a nonlinearity whose form depends on the sign of the kurtosis of the data. With this simplification, most ICA algorithms work well with either subgaussian or supergaussian sources. Recently a negentropy maximization based ICA algorithm was proposed in [13], [21], which was shown to be very robust and works well with almost all types of sources. The performance of this algorithm is somewhat less critically dependent on the choice of the nonlinearity albeit it has an effect on unimodality and convergence of the algorithm [22].

ICA algorithms yields solutions that are subject to a scaling and permutation ambiguity [8]. These ambiguities or indeterminacies might be tolerable in the cases where most of the information is carried in the waveform of the signal (rather than its amplitude) and ordering is not that critical, for example, in speaker separation problem or the ‘cocktail party’ problem. However, in a communication system ordering and scaling ambiguities cannot be tolerated. Scaling ambiguity can be resolved up to a sign ambiguity by constraining the input source variance to unity. To remove the permutation ambiguity further constraints have to be applied, which typically in a CDMA system can be formulated in terms of a priori knowledge about the desired user.

IV. NONLINEAR CONTRAST FUNCTION BASED CONstrained BLIND MULTIUSER DETECTION

Constrained ICA framework in its most general form is described in [14]. This framework has recently been extended for the less-complete ICA problem (extracting a subset of sources) in [15]. The less-complete ICA objective function with equality constraints is given as

\[ J(W, \mu) = J(z) + H(z, \mu), \]

where

\[ H(z, \mu) = \mu^T h(z) + \frac{1}{2} \gamma ||h(z)||^2, \]

where \( \gamma \) is a penalty parameter, \( \mu \) is a Lagrange multiplier [23] whose dimension depends on the number of sources to be separated. \( z \) is a vector containing recovered sources. It is however, noted that constraints \( H(z, \mu) \) in [15] account for constraining the output variance to unity and enforcing the decorrelation condition among the recovered components. They do not incorporate any other a priori information that might be available about the system. For the detection of a single user (independent component), we extend their framework and add further constraints based on a priori information of the desired user’s signature code and timing to the following general contrast function [21]

\[ J(z) = E\{F(|z|^2)\}, \]

where \( z = w^T y \) and \( w \) is a separating vector to be optimized for the detection of a single user. \( E \) is the statistical expectation operator. \( F : \mathbb{R}^+ \cup \{0\} \rightarrow \mathbb{R} \) is a smooth even function. The choice of this contrast function has some important benefits. First, this contrast function facilitates the less-complete ICA, where as other information-theoretic and maximum likelihood approaches do not [15]. Second, by choosing \( F \) appropriately, the nonlinearities can be embedded into the learning algorithm hence avoiding the explicit computations of HOS, which in some cases might be prohibitive. Another advantage is the robustness of the estimator, more slowly the \( F \) grows as its argument increases more robust is the estimator. Following choice of nonlinearities \( F \) were proposed in [21],

\[
\begin{align*}
F_1(y) &= \sqrt{a_1 + y}, \quad f_1(y) = \frac{1}{2\sqrt{a_1 + y}} \\
F_2(y) &= \log(a_2 + y), \quad f_2(y) = \frac{1}{a_2 + y} \\
F_3(y) &= \frac{1}{2} y^2, \quad f_3(y) = y,
\end{align*}
\]

where \( f \) is the first derivative of \( F \). \( a_1 \) and \( a_2 \) are arbitrary constants that are taken as \( a_1 = 0.1 \) and \( a_2 \approx 0.9 \) in this paper. It is to be noted that with \( F_3 \) the contrast function in (13) is similar to the criteria proposed in [12], \( F_1 \) and \( F_2 \) grow slowly than \( F_3 \) and are thus expected to give more robust estimators. Maximization or minimization of (13) depends on whether the underlying sources are supergaussian or subgaussian respectively [21]. In a communications environment all the signal sources and their filtered outputs can be shown to follow subgaussian distributions, hence our task in this paper is the following constrained optimization problem

\[
\text{minimize} \quad J(w) = E\{F(|w^H y|^2)\} \quad \text{with respect to} \quad w
\]

under the constraint \( C^H w = g, \)
where $C_1$ is given as in (10), $g$ is a general constraint vector which is adapted so as to maximize the contribution of the desired user after the MAI has been suppressed. Upon convergence this constraint vector gives an estimate of the desired user’s channel impulse response up to a phase ambiguity [5]. More details on the form of this constraint (15) can be found in [5], [12]. Again, using the method of Lagrange multipliers the objective function (15) is reduced to

$$
\max \min J(w) = E\{f(|w^H y|^2)\} + \lambda^H (C_1^H w - g) + (w^H C_1 - g^H) \lambda,
$$

(16)

where $\lambda$ is a complex Lagrange multiplier vector. It is to be noted that in order to get meaningful results from (16) we need to enforce unit norm constraint on $g$. It is however, simpler to update the norm constraints on $w$ and $g$ manually in each iteration of the algorithm rather than

The first step in deriving an adaptive algorithm for the task at hand is to compute the gradient of (16) with respect to $w$ and $g$ and equate it to zero and compute the optimal form of the detector $w$ as

$$
\nabla J = E\{f(|w^H y|^2 yy^H w\} + C_1 \lambda = 0.
$$

(17)

We can express (17) as

$$
E\{f(|w^H y|^2 yy^H w\} = -C_1 \lambda.
$$

Let us denote $R = E\{f(|w^H y|^2 yy^H\}$, we then have $Rw_o = -C_1 \lambda$, and obtain

$$
w_o = -R^{-1} C_1 \lambda.
$$

(18)

Now using the constraint $C_1^H w_o = g$ we obtain

$$
\lambda = - (C_1^H R^{-1} C_1)^{-1} g.
$$

(19)

Substituting this value of $\lambda$ in (18), for a given $g$ we obtain

$$
w_o = R^{-1} C_1 (C_1^H R^{-1} C_1)^{-1} g.
$$

(20)

It is noted that the optimal separating vector $w_o$ depends on the constraint vector $g$, hence a procedure to jointly update $w$ and $g$ is required. Apart from the least mean square (LMS) based algorithm to be discussed in the next section, one possible way to update $g$ is to use recently proposed power techniques for blind channel estimation problem [24]. According to this method, an estimate of the channel response can be obtained from the following optimization problem

$$
\hat{g} = \arg \min g \; g^H C_1 R^{-m} C_1 g
$$

under the constraint $||g|| = 1,$

(21)

where $m$ in (21) is a finite power and $R = E\{yy^H\}$. The accuracy of channel estimate increases for higher values of $m$. It is also clear from the above equation that the solution for $g$ could be obtained as the smallest eigenvector of the matrix $C_1^H R^{-m} C_1$. It is interesting to observe that with $m = 1$ the criteria in [24] reduces to the criteria presented in [5] while deriving LCMV detectors.

V. LEAST MEAN SQUARE (LMS) ALGORITHM

In this section, we derive the LMS based algorithm for the optimization task in (16). We form two update equations for $w$ and $g$ as

$$
w_{n+1} = w_n - \mu_w \nabla_w J,
$$

(22)

$$
g_{n+1} = g_n + \mu_g \nabla_g J,
$$

(23)

where $\mu_w$ and $\mu_g$ are step size parameters and $\nabla_w J$ and $\nabla_g J$ are the gradients of (16) with respect to $w$ and $g$ respectively. Following the developments in [6] we chose to project $\nabla_g J$ onto the space orthogonal to $g$ to obtain the following equation for the update of $g$ as

$$
g_{n+1} = g_n + \mu_g \left( I - \frac{g_n g_n^H}{g_n^H g_n} \right) \nabla_g J.
$$

(24)

The LMS based update equations are then obtained as

$$
w_{n+1} = \Pi_{C_1} [w_n - \mu_w \hat{R} w_n] + C_1 (C_1^H C_1)^{-1} g_n,
$$

(25)

$$
g_{n+1} = g_n + \mu_g \left( I - \frac{g_n g_n^H}{g_n^H g_n} \right) (C_1^H C_1)^{-1} [\mu_w C_1^H \hat{R} w_n + g_n - C_1^H w_n],
$$

(26)

where $\Pi_{C_1} = I - C_1 (C_1^H C_1)^{-1} C_1^H$. To constrain $g_{n+1}$ and $w_{n+1}$ to unit norm vectors we perform the following additional updates

$$
g_{n+1} \leftarrow \frac{g_{n+1}}{||g_{n+1}||}; \quad w_{n+1} \leftarrow \frac{w_{n+1}}{||w_{n+1}||}.
$$

(27)

In the actual implementation of the algorithm, the statistical expectation in $\hat{R}$ is replaced by its instantaneous value.

VI. CONVERGENCE PROPERTIES

The convergence properties of the contrast function in (13) are rather well studied in the unconstrained case. It can be shown that the extrema of (13) or its related gradient algorithm coincides with the underlying independent components [25]. In another paper [26], it has been proved that independent components are the only asymptotically stable points upon the convergence of the algorithm.

These convergence results [25, 26], are however for the unconstrained case and do not specify to which user the algorithm will converge. In the present paper, the convergence to the desired user is enforced by the constraints in (15). In fact the convergence to the desired user in the specific case of kurtosis based cost function was proved in [12], where it was also elaborated how the kurtosis minimization is able to cancel the MAI as well as inter-symbol interference. It was also shown that the constraint vector $g$ converges to the channel of the desired user up to a phase ambiguity [5], [6], [12]. Due to similarity of our learning equations (25) to the approach in [6, eq. 33-34], their convergence results apply to our case as well.

VII. SIMULATION RESULTS

In this section, we simulate the performance of the proposed detector and compare the results with the kurtosis based
detector of [12], LCMV detector of [6] and theoretical values for ideal minimum mean square error (MMSE) detector [3]. A DS-CDMA system was simulated with quadrature phase shift keying (QPSK) modulation and ten users in a 10-dB AWGN environment. Gold spreading sequences of length 31 were used as spreading codes for all the users in the channel. All inputs took values from \{±1, ±j\} where \(j^2 = -1\). User 1 was assumed to be the desired user, while simulating near-far effect which was 5-dB weaker than each of the other nine equal-power users. The receiver was synchronized to user 1. Signals from other users arrive at the receiver with arbitrary delays \(d_j (j = 2, \cdots, 10)\) between 4-31 chip periods. Channel vectors for all the users were also randomly generated. The channel vectors were all of length \(q + 1\), Gaussian distributed, and were changed in each realization. The step sizes \(\mu_w\) and \(\mu_g\) were taken to be 0.001 and 0.015 respectively. Fifty Monte Carlo runs were performed for each experiment with \(N = 5000\) bits/realization. The input sequences, the AWGN sequence, and the channel vectors changed from one realization to another, while other parameters were kept unchanged to obtain average results.

In Fig. 1(a)-(b) compare the mean square error (MSE) of the channel estimates obtained from the LCMV based method of [6], method proposed in [12] and the method proposed in this paper. It is noted that the convergence of the method proposed in this paper is slightly slower as compared to [6] due to the involvement of HOS based computations, although MSE converges to a slightly lower value, see Fig. 1(b). Results obtained by \(\mathcal{F}_1\) and \(\mathcal{F}_2\) are labeled as ‘HOS1’ and ‘HOS2’ respectively. On the other hand, the kurtosis based

---

**Fig. 1.** Comparison of mean square error of the channel estimate between the LCMV detector of [6], the kurtosis based method of [12] and our implicit HOS based method. SNR for the desired user is 10dB: (a) 10 Equal Energy Users, (b) Unequal Energy Scenario. Desired user’s signal power is 5 dB below to that of the interfering users.

**Fig. 2.** Comparison of achievable SINR between the LCMV detector of [6], the kurtosis based method of [12] and our implicit HOS based method. SNR for the desired user is 10dB: (a) 10 Equal Energy Users, (b) Unequal Energy Scenario. Desired user’s signal power is 5 dB below to that of the interfering users.
approach of [12], does not converge in 5000 data points taken in this simulation. This fact was also observed in [12] with the kurtosis based cost function. Also, note that kurtosis based method of [12] is computationally complex since the gradient of kurtosis based cost function cannot be computed by simple instantaneous approximations of statistical expectation. Instead, the ensemble averages are replaced by empirical averages that are then adaptively updated through the use of a forgetting factor $\alpha$ [27]. In this paper, the value of $\alpha$ used was 0.95 in accordance with [12]. We computed the approximate gradient of the kurtosis using the expression in [12], although it differs from the expression given in [27].

In Fig. 2(a)-(b) we show the comparison of the signal to interference plus noise ratio (SINR) achieved by the various methods considered in this paper. We computed the SINR using the general formula [5]

$$\text{SINR} = \frac{\sigma_2^2 H^H \mathbf{1}_1 \mathbf{1}_1^H w}{w^H (\sigma_2^2 \bar{H} \bar{H}^H + 1) w},$$ (27)

where $\bar{H}$ is the interference matrix which is obtained by removing the first column from (7). It is seen from the Fig. 2(a) that the non-linearities $F_1$ and $F_2$ give almost similar results, verifying the robustness of the proposed method in terms of nonlinearity selection, whereas with the kurtosis based approach of [12] we could not obtain good results, further, it suffers from slow convergence. Similar slow convergence was observed in [12]. For the ideal MMSE detector the SINR value of 11.9670 dB was computed for the perfect power control case [5, eq.22]. Fig. 2(b) depicts the same trend in the performances, in this case theoretical value of SINR was computed to be 10.64 dB for the ideal MMSE detector.

In the power imbalance case, the performance of the MOE based detector is severely affected, which is however, not the case with HOS-based detectors presented in this paper whose performance approaches that of the ideal MMSE detector.

VIII. CONCLUSIONS

In this paper, we have presented a general HOS based approach to blind multiuser detection in multipath channels. The main contribution of the paper lies in the generalization of previously proposed kurtosis based approach to a more general nonlinear approach which exhibits improved convergence properties and is less computationally complex. Another contribution lies in the extension of less-complete constrained ICA framework to DS-CDMA context by incorporating a priori information about desired user where only decorrelation and unit variance constraints were previously considered. We have also developed an adaptive LMS based algorithm for separating vector and parametric constraint vector.

REFERENCES