Hybrid Higher Order Statistics Learning in Multiuser Detection

Antonio J. Caamaño *Member, IEEE*, Rafael Boloix-Tortosa, Javier Ramos *Member, IEEE*,
Juan J. Murillo-Fuentes *Member, IEEE*,

Abstract—In this work we explore the significance of second and higher order statistics learning in communications systems. The final goal in spread spectrum communication systems is to receive a signal of interest completely free from interference caused by other concurrent signals. To achieve this end we exploit the structure of the interference, designing Second Order Statistics (SOS) detectors, such as the Minimum Square Error (MMSE) (known to be optimum linear detectors in noisy environments) in conjunction with Higher Order Statistics (HOS) techniques, such as the Blind Source Separation (BSS) and the Independent Component Analysis (ICA) (known to exhibit such remarkable properties in learning such as *scale invariance* and *superefficiency*). Such Hybrid Higher Order Statistics (HyHOS) approach enables us to alleviate BSS/ICA algorithms of one of their main problems, that is, their sensitiveness to high levels of noise as we benefit from their speed of learning and independence of the initial settings of the problem. We successfully applied the results of this approach to the design of multiuser detectors in Code Division Multiple Access (CDMA) channels.

Index Terms—Array Signal Processing, blind source separation, higher order statistics, independent component analysis, unsupervised learning, CDMA.

I. INTRODUCTION

INTERFERENCE limitation due to the simultaneous access of multiple users systems has been the stimulus to the development of a powerful family of Signal Processing techniques, namely Multiuser Detection (MUD). This techniques have been extensively applied to Direct Sequence Code Division Multiple Access (DS-CDMA) systems. Thus, most of last generation digital communication systems such as GPS, wireless 802.11b, UMTS, etc, may take advantage of any improvement on this topic. In the blind case the algorithms should cope not only with noise and the near-far problem but with no training sequence available. In this sense, the minimum mean square error (MMSE) criteria [1] provides a good blind linear solution to the problem. Due to its computational complexity, some other alternative algorithms have been proposed [2], [3], [4].

On the other hand, blind source separation (BSS) and independent component analysis (ICA) have lately been main fields of research due to their great number of applications

in different areas. ICA and BSS have been usually applied (see [5] and references therein) to communications signals (mainly in array processing), biomedical signals such as ECG or EEG, monitoring, as an alternative to principal component analysis in image or financial data processing, in monitoring, etc... Besides, they have been lately applied to other problems such as spread-spectrum based communications [6], [7], digital watermarking of images [8], audio spectrum basis functions computation [9], image classification, encoding or compression [10],...

Several methods have been proposed as solutions to the BSS/ICA problem. We bring out here two main approaches. On the one hand, we have techniques based on the cancellation of estimation equations by using the steepest descent algorithm [11], [12], [13], [7]. These methods usually have the maximum likelihood (ML) approach as starting point and are not usually robust as the probability density functions (p.d.f.) are given *a priori* through the separation [14], also score [12], functions. Here robustness stands for the methods to be available for mixtures of any statistical distributions. On the other hand, we have contrast functions. Contrast are cost functions whose minimization yields the solution to the BSS/ICA. These methods usually face the BSS/ICA problem starting from lower to higher order moments analysis. In this sense the mixtures are first reduced to zero mean signals. Then a whitening process is carried out to achieve decorrelation. Finally, higher order moments are used to compute a unitary transformation to diagonalize the associated higher order tensor of the whitened outputs. If at most one marginal fourth-order cumulant is null, this unitary matrix may be computed by diagonalizing the fourth order tensor. This tensor may be diagonalized by either canceling the whole set of cumulants out of the diagonal (minimization of the mutual information, MI), or by the maximization of the diagonal entries (minimization of the marginal entropy, ME [15]) as the SICA method used in this paper.

Some blind source separation (BSS) algorithms have been proposed as *fully* blind MUD's, in the sense that spreading codes are unknown [16], [17], [7], [18]. However this blind techniques are not usually noise robust. We propose in this paper to introduce a similar technique to BSS such as independent component analysis (ICA) [19], [5] into the structure of the MMSE MUD to exploit its proven noise robustness. We will compare the performance of a block ICA algorithm such as SICA [19] to an adaptive method to compute the matrix such as M-EASI algorithm [16].

The paper is organized as follows. Section II summarizes

A.J. Caamaño and Javier Ramos are with the DTSC, EPS-Telecomunicaciones, Universidad Carlos III de Madrid. Butarque, 15, 28911 Leganés, Madrid, Spain. Tl: +34 916248736. Fax: +34 916248749. E-mail: ton@tsc.uc3m.es

J.J. Murillo-Fuentes and R. Boloix-Tolosa are with ATSC. Escuela Superior de Ingenieros. Universidad de Sevilla. Paseo de los Descubrimientos sn. 41092 Sevilla. Spain. Tl: +34 954487333. Fax: +34 954487372. E-mail: murillo@esi.us.es.

the matrix model, main assumptions and definitions for a DS-CDMA communication system. The MMSE detector will be also addressed. Then, BSS/ICA techniques are introduced in Section III, where we will focus on the SICA algorithm. Section IV is devoted to relate both of these approaches BSS/ICA and the MMSE to propose a new HOS based solution, the so called BH MUD detector. These are referred to throughout this paper as HyHOS I and II respectively. Section V includes experiments to compare MMSE with the Hybrid High Order Statistics detectors developed in this work and to show noise resistance en superefficient learning. Section VI is devoted to conclusions.

II. COMMUNICATION SYSTEM MODEL: THE MMSE

Consider a CDMA transmitter [1]. The transmitted baseband signal, assuming that the propagation channel of each user is memoryless, can be represented as

$$x(t) = \sum_{l=1}^L \sum_{i=1}^n \chi_i s_i c_i(l - \tau_i) g(t - lT_c - \tau_i), 0 \leq t \leq T_s \quad (1)$$

where L is the length of the spreading code, n is the number of active users, χ_n is a scaling factor specified by the power control loop, s_n is the symbol transmitted by user n , $g(t)$ represents the transmitter pulse shaping filter, T_c is the chip period, and τ_n is the delay of each user. The spreading code of user n is denoted by c_n . Note that individual chips in c_n are denoted by the elements in $[c_n(1), \dots, c_n(L)]^T$ with each $c_n(\cdot) \in \pm 1$. Equation 1 usually is expressed as:

$$x(t) = \sum_{n=1}^N \chi_n s_n c_n(t - \tau_n) \quad (2)$$

where $c_n(t)$ is the normalized signature waveform assigned to the n -th user. If the transmitted pulse-shape filters are Nyquist, we assume the synchronous case and the receiver uses a chip-rate sampled matched filter followed by a serial to parallel converter, the matrix form of the system equation is as follows:

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{s}_t + \mathbf{n}_t \quad (3)$$

where $\mathbf{x}_t$ is a $L \times 1$ vector, $\mathbf{A}_t$ is a $L \times n$ matrix representing the channel response and $\mathbf{n}_t$ is a noise vector. The memoryless easiest solution is a synchronized matched filter (MF). However, if the difference between powers of users is high enough and user's codes are not orthogonal, we have the near-far problem and other well-known SOS based detectors such as the MMSE have much better a performance.

A. Linear Multiuser Detection

A linear multiuser detector $\mathbf{C}$ gives an estimation of the original transmitted signals as

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) = \mathbf{C}\mathbf{A}\mathbf{s}(t) + \mathbf{C}\mathbf{n}(t) \quad (4)$$

The centralized detector minimizing the mean-square-error (MSE) for each user k , $\text{MSE}_k = \text{E} [|\mathbf{y}_k(t) - \mathbf{s}_k(t)|^2]$, is the $\mathbf{C}_{MMSE}$

$$\mathbf{C}_{MMSE} = \mathbf{R}_{vv}^{-1} \mathbf{H}^* = \mathbf{H}^* \mathbf{H} \mathbf{W}_v^* \mathbf{W}_v \mathbf{H}^* \quad (5)$$

where $\mathbf{v}(t) = \mathbf{H}^* \mathbf{x}(t)$ and $\mathbf{W}_v$ is the decorrelating matrix for vector $\mathbf{v}$. Here, as in the following, $*$ denotes transpose-conjugate.

In [17] the authors define a whitening-rotation detector (WR) as the $\mathbf{C}_{WR}$ matrix that minimizes the MSE sum of MSE_k , $k = 1, \dots, n$. A possible structure for this receiver was given as follows

$$\mathbf{C}_{WR} = \mathbf{J}\mathbf{Q}\mathbf{W} \quad (6)$$

where $\mathbf{W}$ is a $m \times m$ whitener, $\mathbf{Q}$ a $m \times m$ rotation matrix, and $\mathbf{J} = [\mathbf{I} \ 0]$ is $n \times m$. In the context of the BSS, the problem consist of computing the matrix $\mathbf{W}_B$ that minimizes statistical dependence at the output.

Other previous BSS approaches to fully blind MUD's use a SVD decomposition. However, this involves a higher computational complexity. The natural or relative gradient (NG) has been also used as a steepest descent algorithm for adaptive approach in both BSS and MUD [20]. We face in this paper the development of a HOS based detector. Let's first introduce the HOS-BSS/ICA tools we are going to use.

III. BSS AND SICA

A. Main assumptions and model

Blind separation of sources (BSS) involve the task of obtaining a non-observable set of signals, the so-called sources, from another set of observable signals regarded as mixtures. Here, the adjective "blind" stands for the fact that neither the original sources nor the mixture itself are known. Usually, and in the context of this paper, the assumption of spatial statistical independence is the key to achieve separation. In this sense, the related [21] and more general technique of independent component analysis (ICA) computes the projection of a set of components (mixtures in BSS) that minimizes the statistical dependence of the outputs [15].

In its simplest form, the BSS reduces to the following matrix form. The $m = n$ mixtures sampled at time t , $\mathbf{x}_t$, are instantaneous linear combinations of the n sources $\mathbf{s}_t$, i.e.,

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad (7)$$

Usually, and in the context of this paper, statistical independence of the inputs is assumed. Thus, if $\mathbf{x}_t$ is a stationary ergodic random sequence and the mixing matrix $\mathbf{A}$ is non-singular, it is possible to estimate a separation matrix $\mathbf{B}$ to obtain the sources as $\mathbf{y} = \mathbf{B}\mathbf{x} = \mathbf{B}\mathbf{A}\mathbf{s}$.

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t = \mathbf{B}\mathbf{A}\mathbf{s}_t = \mathbf{C}\mathbf{s}_t \quad t = 1, 2, \dots \quad (8)$$

In Blind Source Separation (BSS) we compute matrix $\mathbf{B}$ so that $\mathbf{C}$ is ideally the identity matrix. However, the original scaling and arrangement cannot be estimated from the independence assumption. In this sense, $\mathbf{C}$ is a non mixing matrix if it has one and only one non-zero entry in each column and each row. The related [21] and more general problem of independent component analysis (ICA) [15] consists of obtaining the change of basis $\mathbf{B}$ that projects components $\mathbf{x}_t$ into a set $\mathbf{y}_t$ as statistically independent as possible (sources). As matrix $\mathbf{B}$ can be decomposed into the product of a

whitening $\mathbf{W}$ and a unitary $\mathbf{V}$ matrix, this sphering stage gives us signals $z_i(t)$:

$$\mathbf{y}_t = \mathbf{B}\mathbf{x}_t = \mathbf{V}\mathbf{W}\mathbf{A}\mathbf{s}_t = \mathbf{V}\mathbf{z}_t \quad t = 1, 2, \dots \quad (9)$$

In addition to the source independence hypothesis, there are other minor considerations. First, although the instantaneous mixture model considered here applies to many of the applications referred at the Introduction, it may be extended to the convolutive case [22], [23] when needed. Besides, we may consider here the case where the number of mixtures m is greater or equal to the number of sources n , i.e. $m \geq n$. In the case $m > n$, a subspace approach such as a singular value decomposition may be used to project the mixture space onto the signal subspace, reducing the effect of weak noises. For noisy mixtures see [24], [25], [26]. We thus assume the noiseless case with $m = n$. In addition, a necessary and sufficient condition for the waveform-preserving source estimation to be feasible is that no more than one gaussian distributed source be present in the mixture [24], [15], [27]. This model in (7) may be identified with the narrowband m-sensor linear-array application, the synchronous-CDMA case with spreading factor $L = m$, or the general instantaneous BSS/ICA model. In synchronous CDMA communications, matrix $\mathbf{A}$ may be decomposed into $\mathbf{A} = \mathbf{H}\mathbf{\chi}$, where $\mathbf{\chi}$ is a diagonal matrix with the amplitudes of each user and $\mathbf{H}$ is a matrix whose columns are the spreading codes. Notice that BSS/ICA applied to CDMA would lead to a detector where neither a training sequence nor the spreading codes themselves are needed. However, noise vector in (3) is not included in the BSS/ICA model in (7). Thus, if noise is high enough, the performance of the BSS/ICA model deteriorates.

Adaptive solutions to this problem are usually based on maximum likelihood and the natural (or relative) gradient [11], [28], [29]. On the contrary, most of the off-line solutions to ICA [15], [30] minimize one criterion, contrast function, or cancellation of multiple criteria. For instantaneous linear mixtures of two sources, direct methods, which consist of directly estimating the mixing matrix from the mixtures, are also possible. Direct methods suggested in [31], [32], [33] solve polynomial equations based on cumulants up to fourth-order. The solutions of the equations provide the entries of the mixing matrix. However, methods based on the optimization of criteria usually have a better performance. In this sense we consider here the contrast function based on the minimization of the marginal entropies.

B. Minimum marginal entropy based contrast

A contrast function $\phi(\cdot)$ maps $\mathcal{Y}$, the set of random vectors $\mathbf{y}$ (multivariate p.d.f.'s), on $\mathbb{R}$. It has a minimum when the entries of $\mathbf{y}$ are statistically independent, i.e. if $\mathbf{y}$ has independent component $\phi(\mathbf{y}) \leq \phi(\mathbf{A}\mathbf{y}) \quad \forall \mathbf{A}$ non-singular. Thus, the minimization of a contrast function yields the solution to the BSS/ICA problem. If the number of mixtures $m > 2$, these contrasts are complex functions to minimize. The 'Jacobi optimization' [15] consist of solving the n -dimensional problem by decomposing it in a set of 2-dimensional optimizations. The

rotation matrix $\mathbf{V}$ is decomposed into $g = m(m-1)/2$ Givens rotations. Thus, minimization of the contrasts for each angle θ_h , $h = 1, \dots, g$, drives the system to the solution. We will first focus on a 2-dimensional contrast, the SICA. Under the whiteness constraint, i.e. $\mathbb{E}[\mathbf{y}\mathbf{y}^T] = \mathbf{I}$, it yields the following orthogonal (denoted by ϕ_2) contrast [21],

$$\phi_2^{ME}(\mathbf{y}) \stackrel{c}{=} \sum_{i=1}^n H[y_i] \quad (10)$$

where the criteria to minimize is the marginal entropy $H[\mathbf{y}]$. Here, as in the following, $\stackrel{c}{=}$ means equality up to an additive constant c . Besides, we will denote by $\mu_{ijkl}^y = \mathbb{E}[y_i y_j^* y_k y_l^*]$ and $\mu_{ij}^y = \mathbb{E}[y_i y_j^*]$ the fourth order and second-order moments. If we approximate the possible distributions for s by and Edgeworth expansion, rewrite the result in terms of second-order and fourth-order cumulants, $C_{ijkl}^y = \mu_{ijkl}^y - \mu_{ij}^y \mu_{kl}^y - \mu_{ik}^y \mu_{jl}^y - \mu_{il}^y \mu_{jk}^y$ and then minimize it for all possible distributions [15], it follows that

$$\phi_2^{ME}(\mathbf{y}) \approx \frac{1}{48} \phi_{24}^{ME}(\mathbf{y}) = -\frac{1}{48} \sum_i (C_{iiii}^y)^2 \quad (11)$$

where $C_{iiii}^y = \mu_{iiii}^y - 3\mu_{ii}^y{}^2$ is the autocumulant of output y_i , symbol $*$ denoting complex conjugation. Thus, the minimization of the fourth-order cumulant based ME contrast yields the diagonalization of the associated tensor when at most one marginal fourth-order cumulant is null. In the following, we will face the real case. The extension to complex-valued sources is somehow immediate. By using the complex notation and under the whitening constraint, the independent components $y_p(t)$ and $y_q(t)$ in the two-dimensional approach yield

$$\mathbf{y}_*(t) = r(t)e^{j\theta_{pq}(t)} = r(t)e^{j(\theta_{pq} + \beta_{pq}(t))} = e^{j\theta_{pq}} z_*(t) \quad (12)$$

where $j = \sqrt{-1}$. The whitenned mixture vector $\mathbf{z}_*(t) = z_p(t) + jz_q(t)$ are a simple rotation of the normalized (unit variance) sources $\bar{\mathbf{s}}_*(t) = \bar{s}_p(t) + j\bar{s}_q(t) = r(t)e^{j\alpha}$. Notice that at the solution $\rho(t) = \theta + \beta(t) = \alpha(t) + k\pi/2$, $k = 0, 1, 2, \dots$. The associated estimation of the rotation angle θ in the optimization of 2-dimensional contrasts may be easily expressed as a closed function of the following complex-valued linear combinations (*centroids*) of the statistics of the outputs [34]

$$\xi_\gamma = \mathbb{E}[r^4(t)e^{j4\beta_{pq}(t)}] \quad (13)$$

$$\xi_\eta = \mathbb{E}^2[r^4(t)e^{j2\beta_{pq}(t)}] \quad (14)$$

$$\gamma = \kappa_r - 8 \quad (15)$$

Based on the so called 'weighted estimators' (WE) or WAML [35], [36] a general 2-dimensional estimator yields

$$\begin{aligned} \hat{\theta}_{GWE}(\omega_\gamma, \omega_\xi) &= \frac{1}{4} \angle(\omega_\xi \omega_\gamma \xi_\gamma + (1 - \omega_\xi) \xi_\eta) \\ 0 &< \omega_\xi < 1, \quad \omega_\gamma = \pm 1, \gamma \end{aligned} \quad (16)$$

This estimates yields EML estimator in [34] $\hat{\theta}_{EML} = \hat{\theta}_{GWE}(\text{sign}(\gamma), 1)$ or the [37], MK [38], [30], SKSE and SKDE [39], ML [40] estimates $\hat{\theta}_{GWE}(\pm 1, 1)$, $\hat{\theta}_{MASSFOC} = \hat{\theta}_{GWE}(\gamma, 1/2)$ in [39], or the $\hat{\theta}_{AML} = \hat{\theta}_{GWE}(\gamma, 1/3)$ [35]. The discussion on the optimum ω_ξ [36] is open as it is difficult

to compute the statistics of these estimates and analyze them for all possible p.d.f's. In this paper, we propose to use

$$\theta_{SICA} = \hat{\theta}_{GWE}(\gamma, 3/7) \quad (17)$$

as it may be proved that the SICA approach is as robust as the $\phi_{24}^{ME}(\mathbf{y})$, i.e., it provides a solution to the ICA problem with the only condition of no more than one source with null fourth order statistics in the mixture. The minimization of $\phi_{SICA}(\theta)$ is immediate as the solution is minus four times the phase of the resulting function, a sinusoid. This may be easily implemented by a look-up table so that we minimize the number of operations needed.

C. n-dimensional case

The 2-dimensional case may be easily extended to n dimensions by using the 'Jacobi Optimization' [15]. We have here rewritten the algorithm using $\phi_{SICA}(\theta)$. Such an algorithm can be summarized as follows.

Algorithm 1: n-dimensional SICA using Jacobi Optimization.

- 1) *Whitening.* Compute a whitening matrix $\mathbf{W}$ and the output vector $\mathbf{z} = \mathbf{W}\mathbf{x}$. Set $c = 1$ and $\mathbf{y} = \mathbf{z}$.
- 2) *Sweep c.* For all $g = n(n-1)/2$ pairs, i.e., for $1 \leq p < q \leq n$, do
 - (a) Compute the Givens angle θ_{pq} $\theta_{SICA}(\mathbf{y})$ in (17) with $[z_p, z_q]^T = [y_p, y_q]^T$.
 - (b) if $\theta_{pq} > \theta_{min}$, do rotate the pair (y_p, y_q) by θ_{pq} according to (12).
- 3) *End?* If the number of sweeps c satisfies $c \geq K = 1 + \sqrt{n}$ or no angle θ_{pq} has been updated, stop. Otherwise go to step 2 for another sweep with $c = c + 1$.

D. The Natural Gradient

The steepest descent method updates $\mathbf{C}$ according to the direction of the gradient ∇L of a loss function $L(\mathbf{C})$. The natural [23] or relative gradient [5] proposes to use $\tilde{\nabla} L(\mathbf{C}) = \tilde{\nabla} E[l(\mathbf{C})] = \nabla L(\mathbf{C})\mathbf{C}^T\mathbf{C}$. The stochastic version uses the instantaneous value $\tilde{\nabla} l(\mathbf{C})$. The learning law yields

$$\mathbf{C} \leftarrow \mathbf{C} - \lambda \nabla l(\mathbf{C})\mathbf{C}^T\mathbf{C} \quad (18)$$

In the BSS problem it is usually assumed statistical independence at the outputs $\mathbf{y}(t)$. Maximum likelihood (ML) is an extended technique to derive a loss function for this criteria [23], [5]. It can be shown that by forcing the estimating function $\mathbf{K}(\mathbf{y}) = \nabla l(\mathbf{C}) \cdot \mathbf{C}^T = \varphi(\mathbf{y})\mathbf{y}^T - \mathbf{I}$ to cancel, we achieve independence at the output. The learning law yields

$$\mathbf{C} \leftarrow \mathbf{C} - \lambda \mathbf{K}(\mathbf{y})\mathbf{C} \quad (19)$$

where in the ML approach $\varphi_i(y_i) = -q'_i(y_i)/q_i(y_i)$, being $q_i(\cdot)$ the probability density function of source b_i . However, as source distributions are supposed unknown, each author introduces his own activation function $\varphi_i(y_i)$. A family of them may be found, for sources with negative or positive

kurtoses, in [41]. By normalizing (19) the learning law may be written as follows

$$\mathbf{C} \leftarrow \mathbf{C} - \lambda \frac{\varphi(\mathbf{y})\mathbf{y}^T - \mathbf{I}}{1 + \lambda|\varphi^T(\mathbf{y})\mathbf{y}|} \mathbf{C} \quad (20)$$

In [16] the authors proposed a NG based algorithm to separate signals in digital communication, the M-EASI (Median-Equivariant Adaptive Separation via Independence). This method assumes zero-mean, symmetric, 'circularly distributed' signals and introduces the sign function to reduce the bias introduced by the noise. It estimating function yields

$$\mathbf{K}(\mathbf{y}) = \frac{\mathbf{y}\text{sgn}(\mathbf{y})^* - \mathbf{I}}{1 + \lambda|\text{sgn}(\mathbf{y})^*\mathbf{y}|} + \frac{1}{\alpha} \frac{\varphi\text{sgn}(\mathbf{y})^* - \text{sgn}(\mathbf{y})\varphi^*}{1 + \lambda|\mathbf{y}^*\varphi|} \quad (21)$$

where $\text{sgn}(\mathbf{y}) = \text{sgn}(\Re(\mathbf{y})) + j\text{sgn}(\Im(\mathbf{y}))$, $\mathbf{y} \in \mathbb{C}$. With this method we improve the stability of the algorithm [29], provide the method with phase recovering properties and make the method more robust against noise.

IV. HYBRID HIGH ORDER STATISTICS MULTIUSER DETECTORS

A. Natural Gradient Based HyHOS

We could use the natural gradient to compute a detector $\mathbf{C}$ by imposing some criteria such as the MSE. But we may directly exploit the MMSE structure in 5 for the centralized (i.e., base station, all codes available) and non-centralized case (i.e., user equipment, only user code available).

First, we propose the *WWH, centralized detector* by computing $\mathbf{W}_v$ in

$$\mathbf{C}_{WWH} = \mathbf{H}^* \mathbf{H} \mathbf{W}_v^* \mathbf{W}_v \mathbf{H}^* \quad (22)$$

by using (20) with $\mathbf{y} = \mathbf{v} = \mathbf{W}_v \mathbf{H}^* \mathbf{x}$ and activation functions $\varphi(\mathbf{v}) = \mathbf{v}$. The learning law yields the decorrelating algorithm

$$\mathbf{W}_v \leftarrow \mathbf{W}_v + \lambda \frac{\mathbf{I} - (\mathbf{W}_v \mathbf{v})(\mathbf{W}_v \mathbf{v})^T}{1 + \lambda|(\mathbf{W}_v \mathbf{v})^T(\mathbf{W}_v \mathbf{v})|} \mathbf{W}_v \quad (23)$$

Now, we propose a non-centralized MUD, the *HWW detector*. The non-centralized MMSE-MUD was given in (5) as

$$\mathbf{C}_{HWW} = \mathbf{H}^* \mathbf{W}_x^* \mathbf{W}_x \quad (24)$$

If we redefine $\mathbf{y} = \mathbf{W}_x \mathbf{x}$ and rewrite the natural gradient blind source separator in (18) as we did in (23), it follows that

$$\mathbf{W}_x \leftarrow \mathbf{W}_x + \lambda \frac{\mathbf{I} - (\mathbf{W}_x \mathbf{x})(\mathbf{W}_x \mathbf{x})^T}{1 + \lambda|(\mathbf{W}_x \mathbf{x})^T(\mathbf{W}_x \mathbf{x})|} \mathbf{W}_x \quad (25)$$

Following the same structure as before, we have centralized and non-centralized BSS based detector algorithms. In the case of the BSS centralized MUD, the *BH detector*, we can proceed as follows. It is possible to substitute the matrix product $\mathbf{H}^* \mathbf{H} \mathbf{W}_v^* \mathbf{W}_v$ in (22) by a separating matrix $\mathbf{B}$. The new detector yields

$$\mathbf{C}_{BH} = \mathbf{B} \mathbf{H}^* = \mathbf{Q} \mathbf{W}_v \mathbf{H}_v^* \quad (26)$$

where $\mathbf{B}$ is computed as a blind source separator, that is, a whitening-rotator. Notice that in the centralized case the dimensional reduction is carried out by the matched filter $\mathbf{H}^*$,

thus $\mathbf{J} = \mathbf{I}$ in (6). Matrix in $\mathbf{B}$ (31) is computed by using the M-EASI algorithm

$$\mathbf{B} \leftarrow \mathbf{B} - \lambda \mathbf{K}_B(\mathbf{y}) \cdot \mathbf{B} \quad (27)$$

where

$$\mathbf{K}_B(\mathbf{y}) = \frac{\mathbf{y} \text{sgn}(\mathbf{y})^* - \mathbf{I}}{1 + \lambda \text{sgn}(\mathbf{y})^* \mathbf{y}} + \frac{1}{\alpha} \frac{\boldsymbol{\varphi} \text{sgn}(\mathbf{y})^* - \text{sgn}(\mathbf{y}) \boldsymbol{\varphi}^*}{1 + \lambda |\mathbf{y}^* \boldsymbol{\varphi}|}, \quad (28)$$

$\mathbf{y} = \mathbf{B}\mathbf{v} = \mathbf{B}\mathbf{H}^* \mathbf{x}$, and $\boldsymbol{\varphi}(\mathbf{y})$ may be chosen as described in [41]. Thus, the MMSE structure is further enhanced by the properties of the M-EASI algorithm.

On the other hand, it is not possible to define an architecture for the non-centralized MUD as in (31). Suppose the candidate now to be the detector

$$\mathbf{C}_{HB} = \mathbf{H}^* \mathbf{B} = \mathbf{H}^* \mathbf{Q} \mathbf{W}_x \quad (29)$$

Here, matrix $\mathbf{C}_B = \mathbf{B}$ is already a solution, as independence is imposed at the outputs. Thus $\mathbf{C}_{HB}$ is not a detector as $\mathbf{H} \neq \mathbf{I}$. Computing matrix $\mathbf{C}_B$ is basically a blind separation problem and it is out of the scope of this paper. Notice that this detector is fully blind as it does not use the spreading codes [16], [7], [17], but rather constructs them.

Now, we include a discussion on some theoretical aspects of the methods above. We focus on the near-far problem, noise, convergence and complexity.

The main point in using the natural gradient in MUD is that it is equivariant [5], i.e., the convergence has a uniform performance. Let's define $\mathbf{D} = \mathbf{C}\mathbf{S}$ and right multiply (19) by $\mathbf{H}$ to obtain

$$\mathbf{D} \leftarrow \mathbf{D} + \lambda \mathbf{K}(\mathbf{D}\mathbf{b})\mathbf{D} \quad (30)$$

Matrix $\mathbf{S} = \mathbf{A}\mathbf{H}$ is only present at the initial value $\mathbf{D}_0 = \mathbf{C}_0 \mathbf{A}\mathbf{H}$. Thus, the convergence does not depend on matrix $\mathbf{A}\mathbf{H}$. It can be concluded that the methods proposed in this paper are near-far resistance, as convergence is independent of the user's amplitudes.

Algorithms in BSS usually do not cope with the noisy case whenever the number of sources and mixtures are the same $m = n$. By using the M-EASI algorithm we combat the effect of noise in the separation process for digital communication. On the other hand if $m > n$ a signal subspace projection allows noise reduction. Previous BSS approaches to fully blind MUD's use a SVD decomposition, a MPLL, [17] ... However, this involves a higher computational complexity. As the spreading codes are usually available (at least at the base station), it is straightforward to introduce them as a subspace algorithm at a null complexity cost [4]. Besides, the structure of the MMSE detector has been used in (22) and (24) to cope with noise.

Another important characteristic of natural gradient BSS techniques is that of superefficiency [42]. In this sense, provided $E[\varphi(y)] = 0$ (an usual case), the covariance between two outputs decreases of the order of $1/t^2$ in batch estimation and of the order of λ^2 in on-line learning. Furthermore, $\lambda = 1/t$ gives, asymptotically, the best performance, which is the same as the optimal batch estimator. Thus, with $\lambda = 1/t$ we achieve an on-line algorithm with batch features at every t . On the

other hand, if adaptive features are needed, we may use $\lambda < 1$ to achieve an output covariance decreasing as λ^2 .

On the question of complexity, the exact MMSE solution at every time t_e involves computing the eigenvalues of the autocorrelation matrix. Thus, the computational resources needed are significant. The methods proposed in this paper allow computing, as described in the last section devoted to superefficiency, the batch solution as an on-line algorithm at a low computational burden.

B. SICA Based HyHOS

In the section above, it is proposed to substitute the matrix product $\mathbf{H}^* \mathbf{H} \mathbf{W}_v^* \mathbf{W}_v$ in (5) by a matrix $\mathbf{B}$ that makes the outputs y_i as statistically independent as possible. The new detector yields

$$\mathbf{C}_{BH} = \mathbf{B}\mathbf{H}^* = \mathbf{Q}\mathbf{W}_v \mathbf{H}_v^* \quad (31)$$

where $\mathbf{B}$ is a whitening-rotator computed by using the SICA algorithm [19]. Notice that in the centralized case the dimensional reduction is carried out by the matched filter $\mathbf{H}^*$, thus $\mathbf{J} = \mathbf{I}$ in (6).

In [19] authors prove that under whiteness constrains some fourth-order contrasts may be approximated by a sinusoid. Thus, the minimization of the contrast reduces to computing its phase. The starting point is the Minimum Entropy (ME) ICA contrast given by Comon in [15] so that we use the 'Jacobi optimization' to cope with higher dimensions. This method, called SICA (Sinusoidal ICA,) has a good performance along with a low computational cost, outperforming the ME by Comon and the JADE methods [43].

Applying SICA to our problem, assuming inputs v_i have been decorrelated, this algorithm computes $\mathbf{y} = \mathbf{B}\mathbf{v}$ minimizing the following function, denoted as contrast,

$$\phi_{ME}[\mathbf{y}] \approx \phi_4[\mathbf{y}] = \frac{1}{48} \sum_{ijkl \neq iiii} (\text{Cum}_{ijkl}^{\mathbf{y}})^2 \quad (32)$$

where $\text{Cum}_{ijkl}^{\mathbf{y}}$ is the fourth order cumulant of outputs y_i, y_j, y_k and y_l .

Thus, this ICA algorithm may be seen as a fourth order decorrelation, and we go further than the second order based MMSE to find a new solution to the problem. The algorithm SICA is a block method that computes matrix $\mathbf{B}$ from a whole set of observations, similar to the MMSE method.

As the spreading codes are usually available (at least at the base station), it is straightforward to introduce them as a subspace algorithm at a null complexity cost [3]. Besides, the structure of the MMSE detector has been used to cope with noise.

We will compare the performance of this algorithm to that of the M-EASI algorithm, an adaptive method to compute matrix $\mathbf{B}$ that try to minimize a similar fourth order based contrast by using the natural gradient [16].

V. EXPERIMENTAL RESULTS

The test environment used here is a synchronous CDMA system in which the users here are spread using Gold sequences with spreading factor 31. We will present here two

main results: first, the typical evaluation of the performance in a digital communications system, i.e., Bit Error Rates from 10^{-3} to 10^{-6} and the equivalent $\text{SNR} \geq 8$ dBs, a Low Noise Environment (LNE); second, an evaluation in a High Noise Environment (HNE) where we test the *noise resistance* capabilities of the HyHOS algorithm with Bit Error Rates as low as $2 \times 10^{-1} \leq \text{BER} < 5 \times 10^{-1}$ and equivalent Signal to Noise Ratio of $-15 \leq \text{SNR} \leq 0$ dBs. As the BSS problem is posed in the *absence* of noise, one of the main issues is to test the performance of the resulting algorithms in noisy environments. The comparison of the various algorithms presented here must be done with equal convergence speed (i.e. learning speed) on one hand, and equal final variance on the other, to faithfully characterize their performance. The methods described in [44], [45] and [46] are to be used to analyze the final variance and the convergence speed of each algorithm. Those methods are useful and straightforward in the case of linear algorithms such as the MMSE but, in the case of the nonlinear functionals of the HyHOS algorithms, this is a much difficult task and, as such, merits a separate analysis which is currently under way. In the present work, and as a means to compare the algorithms on an equal footing, the training length for each of the detectors is 2000 samples, the convergence speed (i.e. step-size) and the final variance are fixed throughout the experiments. The number of Monte Carlo simulations averaged is 100 in the HNE and 500 in the LNE, which amounts to a confidence margin of at least 95% for any BER estimation presented in this work.

A. Low Noise Environment

In Low Noise Environment we present, for clarity, the convergence performance of the HyHOS I and the MMSE algorithms both with 15 and 20 users. We see a clear difference in the convergence behaviour of the HyHOS I and the MMSE algorithm. That of the HyHOS II algorithms closely resembles that of the HyHOS I. The average decrease in SINR of the HyHOS I algorithm with an increase of 5 users is half of the decrease suffered by the MMSE. In Figure 1 we can see that 2000 samples is more than enough training length for the HyHOS I algorithm whilst for the MMSE, convergence cannot be achieved in a highly occupied channel.

In the evaluation of the Bit Error Rate (see Figure 2), this speed of convergence of the different algorithms is reflected in the performance of both algorithms, HyHOS I and II and that of the MMSE. In this Low Noise Environment (LNE), the sought features in multiuser detector is that of a controlled degradation of Bit Error Rate with an increase in the number of users. Both the SICA based HyHOS algorithm (HyHOS I) and the Natural Gradient based HyHOS (HyHOS II) show a performance close to that of the MMSE (if not better in the HyHOS I case) with a low occupation of the channel (5 users) but the difference in behaviour arises in a highly occupied channel (25 users), where the BER of the MMSE degrades from 10^{-6} at approximately 14 dB of SNR to 10^{-5} , while both HyHOS algorithms keep the BER to approximately 10^{-6} in the same conditions.

Fig. 1. Signal to Interference plus Noise Rate, with SNR of the User of Interest (UOI) = 15 dB, for the HyHOS detector and the MMSE detector, both with 15 and 20 users. The powers of the interfering users is distributed homogeneously between 0 and 30 dB above that of the UOI.

Fig. 2. Bit Error Rate in Low Noise Environment (LNE), i.e. SNR of the UOI ≥ 8 dB, for the HyHOS detector and the MMSE detector, both with 5 and 25 users. The powers of the interfering users is distributed homogeneously between 0 and 30 dB above that of the UOI.

B. High Noise Environment

In the case of High Noise Environment, or HNE, the convergence speed (see Figure 3) of both HyHOS algorithms is completely similar but for higher variance in the HyHOS I algorithm. Both in a low or high occupation channel, the convergence is fairly similar. The MMSE shows a lack of convergence in a high-noise, low-interference channel, quit expected for the MMSE because the interference noise structure is completely masked by gaussian noise, avoiding this algorithm to “lock” on to the signal of the User of Interest. Quite the contrary is the case of the HyHOS algorithms. Their intrinsic nonlinear properties show here the limitations of linear functionals in linear detectors.

The Bit Error Rate of the HyHOS algorithms show the noise resistance proper to the MMSE without showing its demeanors. The convergence to a Matched Filter solution in the asymptotic limit of $\sigma \rightarrow \infty$ is shown experimentally in Figure 4.

Fig. 3. Signal to Interference plus Noise Rate, with SNR of the UOI = -15 dB, for the HyHOS detector and the MMSE detector, both with 5 and 25 users. The powers of the interfering users is distributed homogeneously between 0 and 30 dB above that of the UOI.

Fig. 4. Bit Error Rate in High Noise Environment (HNE), i.e. SNR of the UOI ≤ 0 dB, for the HyHOS detector and the MMSE detector, both with 5 and 25 users. The powers of the interfering users is distributed homogeneously between 0 and 30 dB above that of the UOI.

VI. CONCLUSION

In this paper the authors proposed both ICA and NG as valid techniques to exploit the matrix structure of the parameters involved in the MUD problem, an application of the more general narrowband m-sensor linear-array case. Here, we propose a novel Hybrid MUD by introducing an ICA matrix into the structure of the blind centralized MMSE MUD. This way we palliate one of the main problems of the BSS/ICA algorithms, their sensitiveness to noise. The results included here show a good near-far resistant performance in synchronous CDMA.

ACKNOWLEDGMENTS

The authors would like to acknowledge Dr. Zarzoso and Dr. Martínez-Ramón for their valuable comments.

REFERENCES

[1] S. Verdú, *Multuser Detection*. Cambridge University Press, 1998.

[2] B. Yang, "Projection approximation subspace tracking," *IEEE Transactions on Signal Processing*, vol. 43, pp. 95–107, January 1995.

[3] X. Wang and H. V. Poor, "Blind multiuser detection: A subspace approach," *IEEE Transactions on Information Theory*, vol. 44, pp. 677–690, March 1998.

[4] D. S.-V. A. J. Caamaño and J. Ramos, "Blind adaptive krylov subspace multiuser detection," in *IEEE 54th Vehicular Technology Conference*, October 2001.

[5] J. F. Cardoso, "Blind signal separation: Statistical principles," *Proceedings of the IEEE*, vol. 86, pp. 2009–2025, Oct 1998.

[6] A. J. Caamaño, J. J. Murillo-Fuentes, F. J. González-Serrano, and J. Ramos, "Natural gradient based multiuser detection," in *PIRMC'02*, (Lisbon), Sep 2002.

[7] J. Murillo-Fuentes and F. González-Serrano, "Median equivariant adaptive separation via independence: application to communications," *Neurocomputing*, vol. (Accepted for publication), 2001.

[8] J. Murillo-Fuentes, H. Molina-Bulla, and F. González-Serrano, "Independent component analysis applied to digital image watermarking," in *International Conference on Audio, Speech and Signal Processing*, vol. III, (Salt Lake City, USA), pp. 1997–2000, May 2001.

[9] MPEG7, "Working draft information technology multimedia content description interface," Tech. Rep. ISO/IEC CD 15938-4, ISO/IEC, 2000.

[10] T. Lee, M. Lewicki, and T. Sejnowski, "Unsupervised classification with non-gaussian mixture models using ICA," in *Advances in Neural Information Processing Systems* (D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, eds.), vol. 11, (Cambridge, MA), pp. 58–64, The MIT Press, 1999.

[11] J. F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, pp. 3017–3030, Dec 1996.

[12] S. I. Amari and J. Cardoso, "Blind source separation-semiparametric statistical approach," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2692–2700, Nov. 1997.

[13] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[14] D. T. Pahn, P. Garrat, and C. Jutten, "Separation of a mixture of independent sources through a ML approach," in *EUSIPCO*, pp. 771,774, 1992.

[15] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, Apr. 1994.

[16] J. J. Murillo-Fuentes, M. Sánchez-Fernández, A. Caamaño-Fernández, and F. J. González-Serrano, "Adaptive blind joint source-phase separation in digital communications," in *IEEE International Conference on Communications*, (Helsinki, Finland), Accepted, June 2001.

[17] R. T. Causey and J. R. Barry, "Blind multiuser detection using linear prediction," *IEEE Journal on Selected Areas in Communications*, vol. 16, pp. 1702–1710, December 1998.

[18] J. Joutsensalo and T. Ristaniemi, "Learning algorithms for blind multiuser detection in CDMA downlink," in *PIRMC*, 1998.

[19] J. Murillo-Fuentes and F. González-Serrano, "Independent component analysis with sinusoidal fourth-order contrast," in *International Conference on Audio, Speech and Signal Processing*, vol. V, (Salt Lake City, USA), pp. 2785–2788, May 2001.

[20] F. G.-S. A. J. Caamaño, J. J. Murillo-Fuentes and J. Ramos, "Natural gradient based multiuser detection," in *IEEE Proc PIRMC'2002 Lisbon*, September 2002.

[21] D. Pham, "Blind separation of instantaneous mixture of sources via an independent component analysis," *IEEE Transactions on Signal Processing*, vol. 44, pp. 2668–2779, Nov 1996.

[22] P. Comon, "Contrast for multichannel blind deconvolution," *IEEE Signal Processing Letters*, vol. 3, pp. 209–211, Jul 1996.

[23] S. I. Amari and A. Cichocki, "Adaptive blind signal processing - neural network approaches," *Proceedings of the IEEE*, vol. 86, pp. 2026–2048, Oct 1998.

[24] L. Tong, Y. Inouye, and R. Liu, "Waveform preserving blind estimation of multiple independent sources," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2461–2470, Jul 1993.

[25] A. Cichocki, S. Douglas, and S. Amari, "Robust techniques for independent component analysis (ICA) with noisy data," *Neurocomputing*, no. 22, pp. 113–129, 1998.

[26] S. Douglas, A. Cichocki, and S. Amari, "Self-whitening algorithms for adaptive equalization and deconvolution," *IEEE Transaction on Signal Processing*, vol. 47, pp. 1161–1165, April 1999.

[27] X. R. Cao and R. W. Liu, "General approach to blind source separation," *IEEE Transactions on Signal Processing*, vol. 44, pp. 562–571, Mar 1996.

- [28] S. I. Amari, "Neural learning in structured parameter spaces — natural riemannian gradient," in *Advances in Neural Information Processing Systems* (M. C. Mozer, M. I. Jordan, and T. Petsche, eds.), vol. 9, p. 127, The MIT Press, 1997.
- [29] J. Murillo-Fuentes and F. González-Serrano, "Improving stability in blind source separation with the stochastic median gradient," *Electronic Letters*, vol. 36, pp. 1662–1663, September 2000.
- [30] J. F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, pp. 157–192, Jan 1999.
- [31] P. Comon, "Separation of stochastic processes," in *Workshop Higher Order Spectral Anal.*, (Va, CO), June 1989.
- [32] A. Mansour and C. Jutten, "A direct solution for blind separation of sources," *IEEE Transactions on Signal Processing*, vol. 44, pp. 746–748, Mar 1996.
- [33] R. Martín and J. Acha, "Blind separation of sources using a new polynomial equation," *Electronic Letters*, vol. 33, pp. 176–177, Jan 1997.
- [34] V. Zarzoso and A. K. Nandi, "Blind separation of independent sources for virtually any source probability density function," *IEEE Transactions on Signal Processing*, vol. 47, pp. 2419–2432, September 1999.
- [35] M. Ghogho, A. Swami, and T. Durrani, "Approximate maximum likelihood blind source separation with arbitrary source pdfs," in *IEEE Workshop on Statistical Signal and Array Processing (SSAP'00)*, (Pocono Manor Inn, Pennsylvania), Aug 2000.
- [36] V. Zarzoso, F. Herrmann, and K. Nandi, "Weighted closed-form estimators for blind source separation," in *11th International Workshop on Statistical Signal Processing*, (Singapore), August 2001.
- [37] P. Comon and E. Moreau, "Improved contrast dedicated to blind separation in communications," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. V, (Munich, Germany), pp. 3453–3456, 1997.
- [38] E. Moreau and O. Macchi, "Higher order contrast for self-adaptive source separation," *International Journal of Adaptive Control and Signal Processing*, vol. 10, pp. 19–46, Jan 1996.
- [39] F. Herrmann and A. Nandi, "Blind separation of linear instantaneous mixture using close forms estimators," *Signal Processing*, vol. 81, pp. 1537–1556, 2001.
- [40] F. Harroy and J.-L. Lacoume, "Maximum likelihood estimators and cramer-rao bounds in source separation," *Signal Processing*, vol. 55, no. 2, pp. 167–177, 1996.
- [41] S. I. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, pp. 251–276, Jan 1998.
- [42] S. Amari, "Superefficiency in blind source separation," *IEEE Transactions on Signal Processing*, vol. 47, pp. 936–944, Apr 1999.
- [43] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signals," *Proceedings IEE F*, vol. 140, pp. 362–370, Dec 1993.
- [44] N. Bershad, P. Celka, and J. Vesin, "Stochastic analysis of gradient adaptive identification of nonlinear systems with memory for gaussian data and noisy input and output measurements," *IEEE Transactions on Signal Processing*, vol. 47, pp. 675–689, March 1999.
- [45] J. Bucklew, T. Kurtz, and W. Sethares, "Weak convergence and local stability properties of fixed step size recursive algorithms," *IEEE*, 2000.
- [46] M. Martínez-Ramón and A. Figueiras-Vidal, "Artículo logit/lms," *IEEE*, 2000.