NATURAL GRADIENT BASED BLIND MULTIUSER DETECTION

J.J. Murillo-Fuentes, F.J. González-Serrano, J. Ramos and A.J. Caamaño

Universidad Carlos III. DTSC.
Butarque, 15, 28911 Leganés, Madrid, Spain.
E-mail : ton@tsc.uc3m.es

ABSTRACT

In this paper, novel structures for dynamic removal of multiuser interference are proposed. The natural gradient (NG) is used either to compute whitening matrices in linear blind minimum MSE or to develop new structures. As a result, we propose a family of centralized and non-centralized multiuser detectors (MUD's). The NG provides the MUD's with the equivariant and superfficiency properties, making them near-far resistant by construction and their convergence optimal. Moreover, the complexity of these structures is significantly reduced. These novel solutions are successfully applied to the multiuser synchronous CDMA channel.

1. INTRODUCTION

Interference limitation due to the simultaneous access of multiple users in Code Division Multiple Access (CDMA) systems has been the stimulus to the development of a powerful family of Signal Processing techniques, namely Multiuser Detection (MUD). In the simple memoryless synchronous case where no training sequences are available, the algorithms should cope with noise and the near-far problem. In this sense, the minimum mean square error (MMSE) criteria provides a good linear solution to the problem. However, due to its computational complexity, several alternative algorithms have been proposed [1],[2],[3]. On the other hand, some algorithms have been proposed as *fully* blind MUD's, in the sense that spreading codes are unknown [4], [5], [6].

The natural or relative gradient (NG) has been widely used as a steepest descent algorithm in the blind separation of sources (BSS) [7], [8], [9]. However, these blind techniques are not usually noise-robust.

We propose in this article to use the structure of the MMSE MUD and its proven noise robustness together with the natural gradient algorithm. Noise robustness, superefficient convergence and near-far resistance of these new structures will be demonstrated.

2. PROBLEM STATEMENT

2.1. Signal Model

Let $b(t)$ be a vector of symbols transmitted by n independent finite-alphabet transmitter at time t. We then denote by $x(t)$ the $m \times 1$ vector corresponding to the receiver observation at time t,

$$x(t) = Sb(t) + n(t) \quad (1)$$

where H is an $m \times n$ memoryless channel matrix, and $n(t)$ the noise. This 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 model.

In synchronous CDMA communications, matrix S may be decomposed into $S = AH$ where A is a diagonal matrix with the amplitudes of each user and H is a matrix whose columns are the spreading codes.

2.2. Linear Multiuser Detection

A linear multiuser detector C gives an estimation of the original transmitted signals

$$y(t) = Cx(t) = CSb(t) + Cn(t) \quad (2)$$

The detector minimizing the mean-square-error (MSE) for each user k, $\mathrm{MSE}_k = \mathrm{E}\left[|y_k(t) - b_k(t)|^2\right]$, is the C_{MMSE}

$$\begin{aligned} C_{MMSE} &= R_{vv}^{-1}H^* = H^*HW_v^*W_vH^* \quad (3) \\ &= H^*R_{xx}^{-1} = H^*W_x^*W_x \quad (4) \end{aligned}$$

where $v(t) = H^*x(t)$. Here, as in the following, * denotes transpose-conjugate. Detectors as in (4) are usually referred as non-centralized, since they do not need the whole matrix H to receive one user k but only its column k. On the contrary, centralized detectors as in (3) need the whole matrix H.

In [4] the authors define a whitening-rotation detector (WR) as the C_{WR} matrix that minimizes the MSE sum of

MSE_k, $k = 1, ..., n$. A posible structure for this receiver was given as follows

$$C_{WR} = JQW \tag{5}$$

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

2.3. The Natural Gradient

The steepest descent method updates C according to the direction of the gradient ∇L of a loss function $L(C)$. The natural [9] or relative gradient [8] proposes to use $\widetilde{\nabla} L(C) = \widetilde{\nabla}\text{E}[l(C)] = \nabla L(C)C^{\text{T}}C$. The stochastic version uses the intantaneous value $\widetilde{\nabla} l(C)$. The learning law yields

$$C \longleftarrow C - \lambda \nabla l(C) C^{\text{T}} C \tag{6}$$

In the BSS problem it is usually assumed statistical independence at the outputs $y(t)$. Maximum likelihood (ML) is an extended technique to derive a loss function for this criteria [9], [8]. It can be shown that by forcing the estimating function $K(y) = \nabla l(C) \cdot C^T = \varphi(y)y^{\text{T}} - I$ to cancel, we achieve independence at the output. The learning law yields

$$C \longleftarrow C - \lambda K(y) C \tag{7}$$

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 [7]. By normalizing (7) the learning law may be written as follows

$$C \longleftarrow C - \lambda \frac{\varphi(y)y^{\text{T}} - I}{1 + \lambda|\varphi^{\text{T}}(y)y|} C \tag{8}$$

In [5] 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

$$K(y) = \frac{y\,\text{sgn}(y)^* - I}{1 + \lambda\, \text{sgn}(y)^* y} + \frac{1}{\alpha} \frac{\varphi\, \text{sgn}(y)^* - \text{sgn}(y)\varphi^*}{1 + \lambda|y^* \varphi|} \tag{9}$$

where $\text{sgn}(y) = \text{sgn}(\Re(y)) + j\text{sgn}(\Im(y))$, $y \in \mathbb{C}$. With this method we improve the stability of the algorithm [10], provide the method with phase recovering properties and make the method more robust against noise.

3. MULTIUSER DETECTORS BASED ON NATURAL GRADIENT

3.1. NG-MMSE Multiuser Detectors

We could use the natural gradient to compute a detector C by imposing some criteria such as the MSE. But we may directly exploit the MMSE estructure in (4) for the centralized (*i.e.*, base station, all codes available) and non-centralized case (*i.e.*, user equipment, only user code available).

3.1.1. Centralized MUD, the WWH detector

We propose to compute W_v in

$$C_{WWH} = H^* H W_v^* W_v H^* \tag{10}$$

by using (8) with $y = v = W_v H^* x$ and activation functions $\varphi(v) = v$. The learning law yields the decorrelating algorithm

$$W_v \longleftarrow W_v + \lambda \frac{I - (W_v v)(W_v v)^{\text{T}}}{1 + \lambda|(W_v v)^{\text{T}}(W_v v)|} W_v \tag{11}$$

3.1.2. Non-centralized MUD, the HWW detector

The non-centralized MMSE-MUD was given in (4) as

$$C_{HWW} = H^* W_x^* W_x \tag{12}$$

If we redefine $y = W_x x$ and rewrite the natural gradient blind source separator in (6) as we did in (11), it follows that

$$W_x \longleftarrow W_x + \lambda \frac{I - (W_x x)(W_x x)^{\text{T}}}{1 + \lambda|(W_x x)^{\text{T}}(W_x x)|} W_x \tag{13}$$

3.2. Blind Source Separation Based Detector

Following the same structure as before, we have centralized and non-centralized algorithms.

3.2.1. Centralized MUD, the BH detector

It is posible to substitute the matrix product $H^* H W_v^* W_v$ in (10) by a separating matrix B. The new detector yields

$$C_{BH} = BH^* = QW_v H_v^* \tag{14}$$

where 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 H^*, thus $J = I$ in (5). Matrix in B (14) is computed by using the M-EASI algorithm

$$B \longleftarrow B - \lambda K_B(y) \cdot B \tag{15}$$

where

$$K_B(y) = \frac{y\text{sgn}(y)^* - I}{1 + \lambda\,\text{sgn}(y)^* y} + \frac{1}{\alpha}\frac{\varphi\text{sgn}(y)^* - \text{sgn}(y)\varphi^*}{1 + \lambda|y^*\varphi|}, \quad (16)$$

$y = Bv = BH^*x$, and $\varphi(y)$ may be chosen as described in [7]. Thus, the MMSE structure is further enhanced by the properties of the M-EASI algorithm.

3.2.2. Non-centralized MUD, the B detector

It is not possible to define an architecture for the non-centralized MUD as in (14). Suppose the candidate now to be the detector

$$C_{HB} = H^*B = H^*QW_x \quad (17)$$

Here, matrix $C_B = B$ is already a solution, as independence is imposed at the outputs. Thus C_{HB} is not a detector as $H \neq I$. Computing matrix 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 [5], [6], [4], but rather constructs them.

4. THEORETICAL ANALYSIS

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

4.1. Interference Robustness

The main point in using the natural gradient in MUD is that it is equivariant [8], i.e., the convergence has a uniform performance. Let's define $D = CS$ and right multiply (7) by H to obtain

$$D \longleftarrow D + \lambda K(Db)D \quad (18)$$

Matrix $S = AH$ is only present at the initial value $D_0 = C_0AH$. Thus, the convergence does not depend on matrix AH. It can be concluded that the methods proposed in this paper are near-far resistance, as convergence is independent of the user's amplitudes.

4.2. Noise reduction

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,[4] ... 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 [3]. Besides, the structure of the MMSE detector has been used in (10) and (12) to cope with noise.

4.3. Superefficiency

Another important characteristic of natural gradient BSS techniques is that of superefficiency [11]. 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. Futhermore, $\lambda = 1/t$ 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.

4.4. 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.

5. EXPERIMENTAL RESULTS

In this section we face the MUD in a synchronous CDMA system. The BPSK symbols were spread using GOLD codes with spreading factor $L = m = 31$. The number of users in the simulations was $K = n = 8$. On the one hand, we study in Fig. 1 the convergence of the method by computing the signal to interference ratio (SIR) along the number of samples for a signal to noise ratio (SNR) of 15 dB and average multiaccess interference (MAI) of 30 dB. The learning rate was set to $\lambda = 1/t$ and the activation function used in the C_{BH} detector was the cubic function $\varphi(y) = y^3$. On the other hand, we simulated the same scenario to depict, in Fig. 2, the bit error rate (BER) for different signal to noise ratios (SNR). This BER was computed at $t = 6000$. The MF is included as reference. However, as this detector is not near-far resistance, we have computed the exact MMSE solution at every time t as described in Subsection 4.4. The PASTd [2] algorithm is also included in Fig. 1. The methods presented within this paper exhibit a good performance, in comparison to the MMSE. Under these conditions, other methods such as the PASTd have a poor behaviour [2],[3]. Besides, for a large enough number of training samples the methods have a similar BER compared to that of the MMSE. Notice that although the HWW detector converges to the MMSE solu-

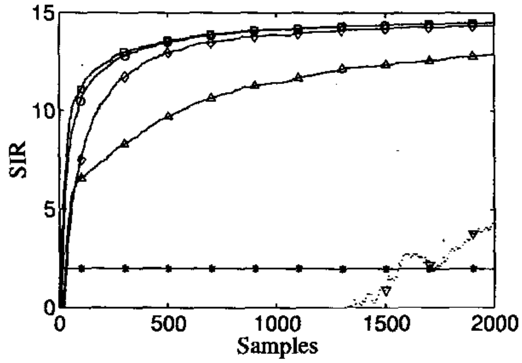

Fig. 1. Convergence for MMSE (○), BH (□), WWH (◇), HWW (△), MF (*) and PASTd (▽) for synchronous CDMA with $n = 8$ users, MAI=30dB and SNR=15dB.

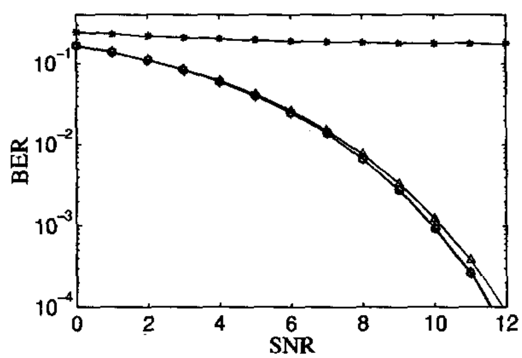

Fig. 2. Bit Error Rate for MMSE (○), BH (□), WWH (◇), HWW (△), MF (*) for synchronous CDMA with $n = 8$ users and MAI=30dB.

tion, its convergence is slower due to the higher dimension of the matrix to estimate, W_x.

It is interesting to notice in Fig 1 how the BH detector exhibits better convergence than the MMSE MUD. This is the effect of the superefficient convergence inherent to the natural gradient.

6. CONCLUSIONS

The natural gradient exploits the matrix estructure of the parameters of a loss function in the design of steepest descent algorithms. As a result, the algorithm is equivariant and the convergence superefficient. Thus, it should be taken into account in narrowband m-sensor linear-array applications. Here, we introduce it in the structure of the blind MMSE MUD to remove the noise weakness the NG suffers. Then we propose a novel BSS based centralized structure where the steering vectors are the basis of the subspace projection. Fully blind algorithms are presented as a particular case of the proposed family of MUD's. The results included here shows a good near-far resistance performance and fast convergence in synchronous CDMA.

7. REFERENCES

[1] Bin Yang, "Projection approximation subspace tracking," *IEEE Transactions in Signal Processing*, vol. 43, no. 1, pp. 95–107, January 1995.

[2] X. Wang and H.V. Poor, "Blind multiuser detection: A subspace approach," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 677–690, March 1998.

[3] A.J. Caama no, D. Segovia-Vargas, and J. Ramos, "Blind adaptive krylov subspace multiuser detection," *IEEE 54th Vehicular Technology Conference*, October 2001.

[4] R. T. Causey and J. R. Barry, "Blind multiuser detection using linear prediction," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 9, pp. 1702–1710, December 1998.

[5] J. J. Murillo-Fuentes, M. Sánchez-Fernández, A. Caamano-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.

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

[7] S. I. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251–276, Jan 1998.

[8] J. F. Cardoso, "Blind signal separation: Statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct 1998.

[9] S. I. Amari and A. Cichocki, "Adaptive blind signal processing - neural network approaches," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2026–2048, Oct 1998.

[10] J.J. Murillo-Fuentes and F.J. González-Serrano, "Improving stability in blind source separation with the stochastic median gradient," *Electronic Letters*, vol. 36, no. 19, pp. 1662–1663, September 2000.

[11] S. I. Amari, "Superefficiency in blind source separation," *IEEE Transactions on Signal Processing*, vol. 47, no. 4, pp. 936–944, 1999.