HIGHER ORDER MOMENTS ALGORITHMS FOR BLIND SIGNAL SEPARATION

J. J. Murillo-Fuentes, F. J. González-Serrano

Departamento de Teoría de la Señal y Comunicaciones.
Escuela Politécnica Superior. Universidad Carlos III de Madrid.
C/Butarque, 15. 28911 Leganés. Madrid. Spain.

Abstract— **An on-line learning algorithm, which minimizes a criterion based on geometrical properties, is derived for blind separation of mixed signals. This new contrast function focuses on the concept of center of masses and higher order moments (HOM) applied to the outputs. The source signals and the mixing matrix are unknown except for the number of sources. A set of estimating equations is obtained. The relative (natural) gradient is used as learning law. This new algorithm is related to the Maximum Likelihood approaches providing a new point of view for understanding them. It is concluded that HOM methods outperform them. Some results are included for audio and synthetic signals. These results show how the algorithm proposed presents better convergence in comparison to other well-known approaches.**

I. INTRODUCTION

The problem of blind signal separation arises in many areas such as speech recognition, data communication [3], sensor signal processing [10][11], finances [4] and medical science [13][14]. Several on-line algorithms have been proposed. Most of them may be rewritten in the form of relative gradient through the score functions. The performance of these algorithms is strongly affected by the selection of these functions. The optimum score functions are computed from the probability density functions (p.d.f.) of the sources [9]. As they are unknown, the score functions are usually chosen ad hoc or computed from the outputs [15]. We shall present a new contrast function based on the center of masses (CoM) and higher order moments (HOM) of the outputs. The score functions related to this estimating function are derived to compare this method to others found in the literature. Other approaches exist based upon geometrical properties [16]. Nevertheless, these are based on approaching the mixture by a set of uniform distributions. Thus, they are faraway from the one presented here. Next, we proceed to introduce the problem.

The simplest source separation model is that of an $n \times 1$ vector of observations with structure

$$\boldsymbol{x}(t) = A\boldsymbol{s}(t) \tag{1}$$

where A is an invertible $n \times n$ unknown matrix and $\boldsymbol{s}$ is an unobserved $n \times 1$ vector. The components of vector $\boldsymbol{s}$, the so-called 'sources', are statistically independent. The task is to recover the source signals and/or to identify matrix A using only the assumption of source independence. Only the case of real zero mean signals is considered here. In adaptive approaches, one explicitly updates an $n \times n$ 'separating' matrix B which yields an 'output vector' $\boldsymbol{y}=B\boldsymbol{x}=C\boldsymbol{s}$. The product $C\boldsymbol{s}$ is said to be a 'copy' of $\boldsymbol{s}$ [9] if matrix C is nonmixing, i.e, it has one and only one nonzero entry in each row and each column. Therefore the entries of $C\boldsymbol{s}$ are identical to those of $\boldsymbol{s}$ up to permutations and changes of scales and signs.

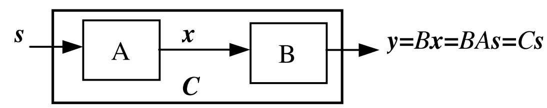

Fig. 1. Mixing and demixing matrix in a BSS problem.

Under the stationary assumption, it may be defined a contrast function [12]. These contrasts are of the form $\phi[y]$ with $y=Bx$ and should, for any matrix C, satisfy $\phi[Cs] \geq \phi[s]$ with equality only when $y=Cs$ is a copy of the source signals. Since the mixture can be reduced to a rotation matrix by enforcing the whiteness constraint $E[yy^T]=I$, one can also consider orthogonal contrast functions denoted herein by $\phi^{o}[y]$.

II. CoM BASED CONTRAST FUNCTION

We shall work on the 2-dimensional (2-d) case. The results will be carried over higher order cases. The center of mass of a body is the point about which its mass can be considered to act. If a 2-d body is defined by a set of points samples of the vector $\boldsymbol{s}(t)$, $\boldsymbol{s}(t)$ $t = 1,...,T$, its center of mass may be written as the pair

$$cm_1(\boldsymbol{s}) \triangleq \frac{1}{T}\sum_{t=1}^{T} s_1(t) \quad \text{and} \quad cm_2(\boldsymbol{s}) \triangleq \frac{1}{T}\sum_{t=1}^{T} s_2(t) \, . \tag{2}$$

Proposition 1: Under the whiteness constraint and for zero mean mixtures, it may be written the following contrast function:

$$\phi^{o}{}_{CoM}[\boldsymbol{y}] \triangleq \mathrm{abs}(\psi_{12}[\boldsymbol{y}]) + \mathrm{abs}(\psi_{21}[\boldsymbol{y}]) \tag{3}$$

with

$$\psi_{12}[\boldsymbol{y}] \triangleq \left(cm_1(\boldsymbol{y}, y_1>0, y_2>0) + cm_1(\boldsymbol{y}, y_1<0, y_2>0)\right) - \left(cm_1(\boldsymbol{y}, s_1>0, y_2<0) + cm_1(\boldsymbol{y}, y_1<0, y_2<0)\right) \quad (4)$$

and

$$\psi_{21}[\boldsymbol{y}] \triangleq \left(cm_2(\boldsymbol{y}, y_1>0, y_2>0) + cm_2(\boldsymbol{y}, y_1>0, y_2<0)\right) - \left(cm_2(\boldsymbol{y}, y_1<0, y_2>0) + cm_2(\boldsymbol{y}, y_1<0, y_2<0)\right) \quad (5)$$

Proof of Proposition 1: As the number of samples $T \mapsto \infty$, $\psi_{12}[\boldsymbol{y}]$ yields

$$\psi_{12}[\boldsymbol{y}] = \int_0^\infty\int_0^\infty y_1 P(y_1, y_2) dy_1 dy_2 + \int_{-\infty}^0\int_0^\infty y_1 P(y_1, y_2) dy_1 dy_2 - \int_0^\infty\int_{-\infty}^0 y_1 P(y_1, y_2) dy_1 dy_2 - \int_{-\infty}^0\int_{-\infty}^0 y_1 P(y_1, y_2) dy_1 dy_2 \quad (6)$$

As the mixture is unbiased, and under whiteness constraint, we impose independence between y_1 and y_2,

$$\psi_{12}[\boldsymbol{y}] = c_2^+ \int_0^\infty y_1 P(y_1) dy_1 + c_2^+ \int_{-\infty}^0 y_1 P(y_1) dy_1 - c_2^- \int_0^\infty y_1 P(y_1) dy_1 - c_2^- \int_{-\infty}^0 y_1 P(y_1) dy_1 = 0 \quad (7)$$

with

$$c_2^+ = \int_0^\infty P(y_2) dy_2 \quad \text{and} \quad c_2^- = \int_{-\infty}^0 P(y_2) dy_2 \quad (8)$$

Simple calculus yields:

$$\psi_{12}[\boldsymbol{y}] = (c_2^+ - c_2^-) \int_{-\infty}^\infty y_1 P(y_1) dy_1 = 0 \quad (9)$$

The same applies to $\psi_{21}[\boldsymbol{y}]$. Thus, a rotation angle θ_{dmx} that makes null both terms in (3) makes the output independent.

An interpretation of (7) and (9) is that the centers of mass for each quadrant must satisfy some symmetry rules at the optimum-demixing angle. Fig. 2 includes some different situations and its CoM. Fig. 2 (a) and (b) are mixtures of two sources with different p.d.f. We have drawn the projections onto the y_1 axis so $\psi_{21}[\boldsymbol{y}]$ is easily computed. The rotation angles are zero and a random one respectively. Fig. 2 (c) shows a maximum mixing angle ($\pi/4$) that is a solution to the contrast. In this case both sources have the same p.d.f.

We have demonstrated that the demixing angle is a solution to the contrast. Nevertheless, as in other contrast functions such as $\phi_{ML}[\boldsymbol{y}]$ (the based on the maximum likelihood, ML, theory), we cannot prove that the conditions provides are sufficient. In fact, sometimes the maximum mixing angle, θ_{mmx}, is also a solution to these contrasts, see **Fig. 2** (c). They may be discriminated by studying the sign of the first derivative of the sum $\psi_{12}[\boldsymbol{y}]+\psi_{21}[\boldsymbol{y}]$ along with the sign of the kurtosis of the sources.

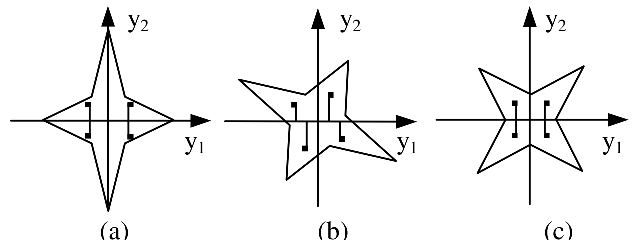

Fig. 2. CoM for different rotations: (a) independence, (b) any rotation, and (c) maximum mixing and same p.d.f's.

Finally, in non-very probable situations, may be more than 2 solutions. It takes place whenever the sum of CoM in (4) and (5) cancels at a rotation angle different from θ_{dmx} and θ_{mmx}. Thus, it depends on the specific structure of the data to separate.

As we focus on an on-line method, we devote the next section to relate this contrast function to the ML approach and the natural gradient. This way we will extend the CoM contrast function to the n-dimensional case under non-whiteness constraint.

III. Estimating And Score Functions

Before relating both theories, we first introduce the main concepts of the ML approach [9]. An estimating function for the BSS problem is a function H: R$\mapsto$ R^{nxn}. It is associated with an estimating equation [8]

$$\frac{1}{T}\sum_{t=1}^{T} H(\boldsymbol{y}(t)) = 0 \quad (10)$$

thus called because H being matrix valued. At the core of the BSS contrast function is $\phi_{ML}[\boldsymbol{y}]$, which is associated with the likelihood given the source densities $q_1, q_2, \ldots, q_n$. The ML principle suggest the specific form

$$H_\varphi(\boldsymbol{y}) \triangleq \varphi(\boldsymbol{y})\boldsymbol{y}^{\mathrm{T}} - I \quad (11)$$

with φ: R$\mapsto$ R^n the entry-wise nonlinear function

$$\varphi(\boldsymbol{y}) \triangleq \left[\varphi_1(y_1), \ldots, \varphi_n(y_n)\right] \quad (12)$$

collecting the score functions related to each source through

$$\varphi_1 \triangleq -(\log q_i)' \quad \text{or} \quad \varphi_1 = -q_i(\cdot)'/q_i(\cdot) \quad (13)$$

Once the estimation and score function has been defined, the contrast function in (3) may be rewritten to identify the score functions related to the CoM approach.

*Proposition 2***:** The contrast function $\phi_{CoM}[\boldsymbol{y}]$ is equivalent to the $\phi_{ML}[\boldsymbol{y}]$ with a priori p.d.f. $q_i(\cdot)=\exp(-|s_i|)$.

Proof of Proposition 2: The expressions for $\psi_{12}[\boldsymbol{y}]$ and $\psi_{21}[\boldsymbol{y}]$ are special cases of the general term

$$\psi_{ij}[\boldsymbol{y}] = \mathrm{sign}(y_i) \cdot y_j \quad (14)$$

Equation (3) may be rewritten as

$$\phi_{CoM}[\boldsymbol{y}] \triangleq \mathrm{abs}(\psi_{11}[\boldsymbol{y}]) + \mathrm{abs}(\psi_{12}[\boldsymbol{y}]) + \\ + \mathrm{abs}(\psi_{22}[\boldsymbol{y}]) + \mathrm{abs}(\psi_{21}[\boldsymbol{y}]) \quad (15)$$

Notice that the contrast function in this new expression is no more constrained to the whiteness condition. Since symmetry between quadrants may only be achieved if $\mathrm{E}[\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}]=I$.

From (11) and (15), the score functions for the CoM approach are found to be

$$\varphi_i = \mathrm{sgn}(y_i) \quad (16)$$

Simple calculus yields

$$q_i(s) = e^{-|s_i|} \quad (17)$$

We conclude that the $\phi_{CoM}[\boldsymbol{y}]$ is equivalent to $\phi_{ML}[\boldsymbol{y}]$ with supposed p.d.f as given in (17). Notice that, as for ML solutions, the sign of the kurtosis discriminates between the maximum and the minimum mixing solutions, if the first exists.

At this point, the extension to a higher number of sources is immediate. On the other hand, this result, along with the relative gradient definition [2], allow us to formulate the following learning law

$$B_{t+1} = (I - \mu_t H_\varphi(\boldsymbol{y}(t)) B_t \quad (18)$$

where H_φ was given in (11), $\varphi(\boldsymbol{y})$ is that of equation (16), μ_t is the learning rate and B is the separating matrix. Thus, this learning algorithm is equivariant [7].

Proposition 2 establishes the relation between CoM and ML approaches. Before drawing some conclusions about this relation, it is interesting to consider introducing higher order moments (HOM). This way, it will be possible to study its performance along with other well know ML based contrast functions.

IV. Higher Order Moments

If the CoM of a 2-d body was defined in (2), the moment of masses of order l yields

$$hom_1(\boldsymbol{s}) \triangleq \frac{1}{T}\sum_{t=1}^{T} s_1{}^l(t) \quad \text{and} \quad hom_2(\boldsymbol{s}) \triangleq \frac{1}{T}\sum_{t=1}^{T} s_2{}^l(t) \,. \quad (19)$$

In the following, the estimating matrix in (10) will be derived for odd and even orders.

A. Odd Order Moments

Proposition 3: The matrix function

$$H_\upsilon = \upsilon(\boldsymbol{y}).\varphi(\boldsymbol{y})^{\mathrm{T}} \quad (20)$$

where $\varphi(\boldsymbol{y})$ was defined in (16) and

$$\upsilon_i = y_i{}^l \quad l = 1,3,5,.. \quad (21)$$

is an estimating matrix, i.e., a solution to (10) if the p.d.f. of the sources, $q_i(\cdot)$, are symmetric.

Proof of Proposition 3: In Equation (9), there are diagonal

$$h_{ii} = \mathrm{sgn}(y_{i.}) \cdot y_i{}^l \quad l = 1,3,5,... \quad (22)$$

and crossed elements:

$$h_{ij,i\neq j} = \mathrm{sgn}(y_{i.}) \cdot y_j{}^l \quad l = 1,3,5,... \quad (23)$$

As the number of samples $T \mapsto \infty$, h_{ii} yields

$$\mathrm{E}[h_{ii}] = \int_0^\infty y_i{}^l P(y_i) dy_i - \int_{-\infty}^0 y_i{}^l P(y_i) dy_i = co_i{}^l \quad (24)$$

If we impose independence between y_1 and y_2, co_i^l is a constant. In addition, the crossed elements cancel:

$$\mathrm{E}[h_{ij}] = (c_j{}^+ - c_j{}^-) \int_{-\infty}^{\infty} y_i{}^l P(y_i) dy_i = 0 \quad (25)$$

Symmetry in the p.d.f. enforces both terms in (25) to cancel. Thus, Odd order moments may be included in the manifold of estimating matrices solutions to the BSS problem. In the following section we face the even order moments.

B. Even Order Moments

Similarly, it can be stated the following for even orders.

*Proposition 4***:** The estimating matrix

$$H_\upsilon = \upsilon(\boldsymbol{y}).\varphi(\boldsymbol{y})^{\mathrm{T}} \quad (26)$$

where $\varphi(\boldsymbol{y})$ was defined in (16) and

$$\upsilon_i = \mathrm{sgn}(y_i) \cdot y_i{}^l \quad l = 2,4,... \quad (27)$$

is an estimating matrix for sources with symmetric p.d.f.

Proof of Proposition 4: Again we study diagonal, h_{ii}, and cross elements, $h_{ij\ i\neq j}$, separately. Similarly, as the number of samples $T \mapsto \infty$ and under independence condition, h_{ii} yields

$$\mathrm{E}[h_{ii}] = \int_{-\infty}^{\infty} \mathrm{sgn}(y_i) \cdot \mathrm{sgn}(y_i) y_i{}^l P(y_i) dy_i = ce_i{}^l \quad (28)$$

where co_i^l is a constant. Again, the crossed terms cancel:

$$\mathrm{E}[h_{ij}] = (c_j{}^+ - c_j{}^-) \left(\int_0^\infty y_i{}^l P(y_i) dy_i - \int_{-\infty}^0 y_i{}^l P(y_i) dy_i \right) = 0 \quad (29)$$

In the same way as the odd case, equations (28) and (29) prove that even order moments may be included in the manifold of estimating matrices solutions to the BSS problem. Notice that the even moments have been artificially converted to odd symmetric functions. Thus, the distributions of moments for each quadrant are different and the gradient will perform better. Notice that some even ML score functions such as $\varphi(\boldsymbol{y})=\boldsymbol{y}^2$ are not usually used as they have poor performance.

C. Polynomial Estimating Matrices

Once the even and odds moments have been developed, we next define the manifold of possible estimating matrices as a linear combination of them.

Corollary 1: the matrix function

$$H_\upsilon = \upsilon(\boldsymbol{y}).\varphi(\boldsymbol{y})^{\mathrm{T}} \tag{30}$$

where

$$\upsilon_i = \sum_{l=1}^{L} \alpha_l \,\mathrm{sgn}(y_i)^{l-1} y_i^{\,l} \tag{31}$$

And $\varphi(\boldsymbol{y})=\mathrm{sgn}(\boldsymbol{y})$ is a valid estimating matrix for sources with symmetric p.d.f.

Proof of Corollary 1: If there is only one $\alpha_l \neq 0$ in (35), the Corollary reduces to Proposition 3 (l odd) or Proposition 4 (l even). Similarly to proofs of Proposition 3 and Proposition 4, as the number of samples $T\mapsto\infty$ and by imposing independence, a linear combination of the functions in (21) and (27) results in constants values for diagonal elements and zeros for the crossed ones.

The generalization of $\upsilon(\boldsymbol{y})$ to any function is, somehow, immediate. Nevertheless, it is not in the scope of this paper. On the other hand the estimating matrices presented in (30) does not cancel for non-symmetric p.d.f. However, in the results included in Section V.C. the natural gradient converges to the separation minimizing the associated contrast function. The generalization for non-symmetric p.d.f. it is considered here as a future line of work.

In the following section we include a discussion on the ML and HOM estimating matrices, Hφ and Hυ respectively.

D. Discussion

The set of ML score functions $\varphi(\boldsymbol{y})=y^l$ with l=1,3,5,..., are well known and exploited, [5][1]. We first analyze their estimating matrix H_φ in comparison to H_υ.

Similarly to previous proofs, we apply the same steps to H_φ. Again we study diagonal, h_{ii}, and cross terms, $h_{ij\; i\neq j}$, separately. As the number of samples $T\mapsto\infty$ and by imposing independence, h_{ii} yields

$$\mathrm{E}[h_{ii}] = \int_{-\infty}^{\infty} y_i^{\,l+1} P(y_i)dy_i = ct_i \tag{32}$$

where ct_i is a constant. On the other hand, crossed terms cancel:

$$\begin{aligned}\mathrm{E}[h_{ij}] &= c_j^{\,+}\int_0^{\infty} y_i^{\,l} P(y_i)dy_i + c_j^{\,+}\int_{-\infty}^{0} y_i^{\,l} P(y_i)dy_i \\ &+ c_j^{\,-}\int_0^{\infty} y_i^{\,l} P(y_i)dy_i + c_j^{\,-}\int_{-\infty}^{0} y_i^{\,l} P(y_i)dy_i\end{aligned} \tag{33}$$

Symple calculus yields

$$\mathrm{E}[h_{ij}] = (c_j^{\,+} + c_j^{\,-})\int_{-\infty}^{\infty} y_i^{\,l} P(y_i)dy_i = 0 \tag{34}$$

where now

$$c_j^{\,+} = \int_0^{\infty} y_j P(y_j)dy_j \quad \text{and} \quad c_j^{\,-} = \int_{-\infty}^{0} y_j P(y_j)dy_j \tag{35}$$

We first may deduce that if the p.d.f. are symmetric and l is even, only the first term in (34) cancels. That is, the proposed combination of four different moments in a step of the gradient reduces to two as the first two terms and the two last ones are the same. Thus, the combination of them provide no information to the gradient, since

$$\begin{aligned}\int_0^{\infty}\int_0^{\infty} y_i^{\,l} y_j P(y_i,y_j)dy_i &= \int_0^{\infty}\int_{-\infty}^{0} y_i^{\,l} y_j P(y_i,y_j)dy_i \\ \int_{-\infty}^{0}\int_0^{\infty} y_i^{\,l} y_j P(y_i,y_j)dy_i &= \int_{-\infty}^{0}\int_{-\infty}^{0} y_i^{\,l} y_j P(y_i,y_j)dy_i\end{aligned} \tag{36}$$

On the other hand, while in H_υ only the integration of the p.d.f. over the negative and positive parts is estimated, in H_φ the estimation of its mean is computed. Leaving the computational burden out of this discussion, it may be deduced that the error in the estimation of these values is bigger than in the HOM case. To accomplish this idea, we compute the variance of both estimators. Under independence constraint, the variance for the crossed element h_{ij} of the ML based estimating matrix H_φ at any step of the natural gradient yields

$$\sigma_{ML}^{\;2} = E\left[\left(\varphi(y_i)\cdot y_j\right)^2\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(y_i)^2 y_j^{\,2} P(y_i,y_j)dy_i dy_j \tag{37}$$

On the other hand,

$$\sigma_{HOM}^{\;2} = E\left[\left(\upsilon_i(y_i)\mathrm{sgn}(y_j)\cdot\right)^2\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \upsilon(y_i)^2 P(y_i,y_j)dy_i dy_j \tag{38}$$

Thus, the error of the ML estimation is affected by the term $y_j^{\,2}$. The convergence and stability of this new approach will improve in relation to this factor.

Finally we close this section remarking two important conclusions. Notice first that by introducing HOM in the estimating equation it is possible to use even functions as scores. Besides, the method described provides the ML approach with the basis to explain why a p.d.f. and its associated score function does, or not, performance adequately. Symmetry properties impose the necessary and sufficient conditions.

For each cross term in the estimating equation, the necessary conditions have been discussed through the latest sections. The sufficient condition, as discussed in Section II, depends on the structure of the sources. Recall that usually there are just one or two solutions to the estimating equation. These are discriminated by the sign of the kurtosis. If any other solution exists, the method may not converge to the separation. This, again, may be explained by the HOM approach, as this solution does meet the symmetry conditions enforced by the estimating equation elements.

We present next some experiments to compare the performance of our approach to some well-known contrast functions.

V. Performance Analysis.

We divide this section in various points containing some different experiments to study the performance of the HOM approach. They all have in common the index [1] used to measure the performance. It is defined as:

$$E_1 = \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{|p_{ij}|}{max_k\,|p_{ik}|}-1\right)+\sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{|p_{ij}|}{max_k\,|p_{kj}|}-1\right) \qquad (39)$$

where C was defined in section I as the product BA.

The mixing matrix for all subsections was randomly generated as

$$A=\begin{bmatrix}0.8264 & -0.2359\\ -0.7538 & 0.9722\end{bmatrix} \qquad (40)$$

A. Third order functions

In this subsection the estimating matrices H_φ and H_υ with $\varphi = y^3$ and $\upsilon = y^3$ are compared. Fig. 3 includes the performance index along time for the mixture of the sources in [17]. These signals were proposed as benchmark for the ICA 1999 workshop. The continuous and dashed lines correspond to the ML and HOM cases respectively. We studied first the learning rates, μ_{ML} and μ_{HOM}, making the algorithm unstable and then reduce them slightly (μ_{ML}=8e-5 and μ_{HOM}=10e-5) to smooth their learning curves. It can be remarked two aspects. First of all, the algorithm proposed is 2/3 times faster than the ML approach. Besides it is less sensitive to problems associated to the correlation and absence of data. At second 5, the first source has a set of samples with values near to 0. This set makes the ML algorithm to diverge from separation.

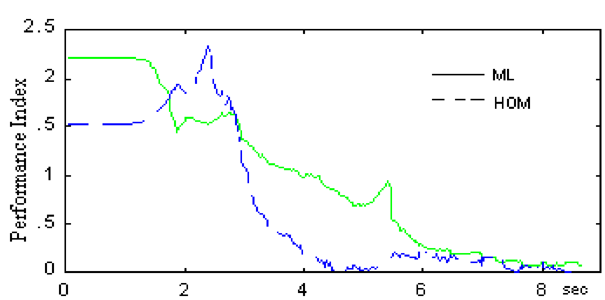

Fig. 3. ML and HOM convergence for mixed audio signals.

B. Uniform Performance: EASI Algorithm

The uniform performance [6] (UP), applied to the ML score functions [7] results in efficient algorithms whose behavior do not highly depends on the learning rate. They provide better stability features. We include here three methods: the score function $\varphi = y^3$ with uniform performance (EASI), the HOM with uniform performance and the HOM algorithm for $\upsilon = y^3$. They are included in Fig. 4. with continuous, dashed and dotted lines respectively. The sources in [17] were mixed with the matrix in (40). The learning rates μ_{EASI}=6.67e-5 (parameter λ in [6]) μ_{HUP}=6.67e-5 and μ_{HOM}=10e-5 were the maximum rates keeping the index performance under E_I=0.3. The HOM algorithm outperforms both, EASI and HOM UP methods. Besides, it reduced the number of float points operations in a factor η=1.89.

C. Polynomial Score Functions

In [1], a polynomial score function was derived obtaining:

$$\varphi(\mathbf{y}) = .75\mathbf{y}^{11} + 6.23\mathbf{y}^{9} + 4.67\mathbf{y}^{7} + 11.75\mathbf{y}^{5} + 14.5\mathbf{y}^{3} \qquad (41)$$

We included in Fig. 5 a comparison between the EASI algorithm with $\varphi = \mathbf{y}^3$ (continuous line), the polinomial score function in (41) (dashed) and the HOM with $\upsilon = y^3$ (dotted). We generated four pairs of random synthetic signals: beta and beta distributions, uniform and uniform, exponential and exponential, and gamma (with symmetry around origin) and exponential. We set the learning rates to μ_{EASI}=1.3e-4, μ_{Pol}=1.8e-4 and μ_{HOM}=3e-4. The conclusion is that for a

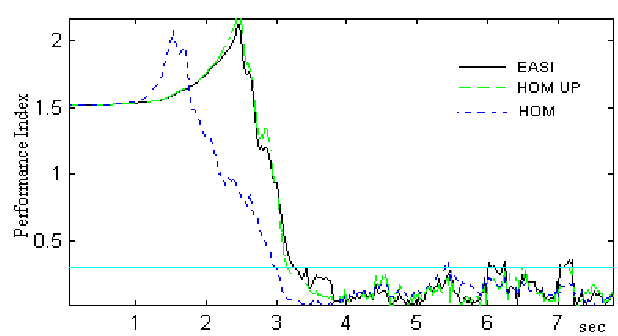

Fig. 4. Uniform performance in the HOM approach: EASI (continuous line), HOM with UP (dashed) and HOM (dotted).

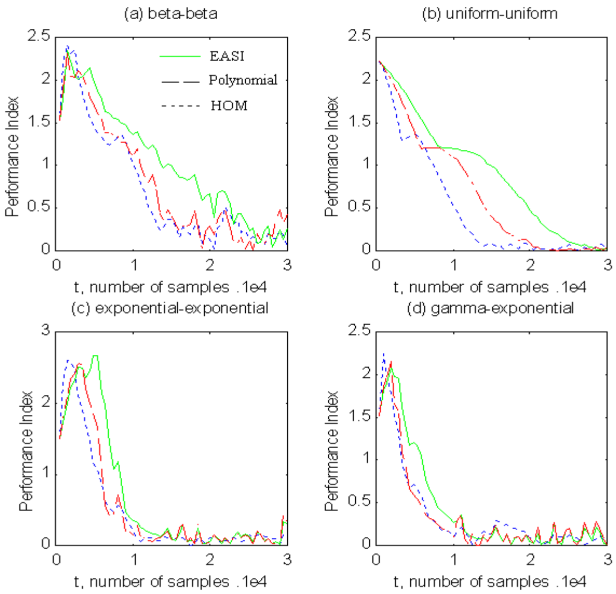

Fig. 5. Separation of mixtures of random synthetic signals with different p.d.f. with EASI, a polynomial score and HOM.

fixed learning rate the HOM algorithm presents a better and homogeneous behavior. Besides, and as it was remarked in section IV.C, the results for non-symetric p.d.f. are also succesful.

D. Standard Deviation of the Estimators

Finally, we show in Fig. 6 the standard deviations for one element of the estimators H_φ (continuous line) and H_υ (dotted) in Section IV.A. It can be observed that the error in in the ML approach is 3 times bigger than in the HOM case.

VI. SUMMARY

The theory on center of masses was the starting point to draw a new contrast function. This function was related to the ML approach. First of all, to obtain the corresponding supposed p.d.f. and, in second place, to use the natural gradient as on-line method.
Once the CoM was introduced we exploited higher order moments and its combinations resulting in a set of estimating matrices. The experiments included within this paper show how the convergence of this new approach improves in comparison to other well-known algorithms.

Ml approaches suppose an a priori p.d.f., $q(\cdot)$, for the sources. Only if the $q(\cdot)$ is the correct one, the separation is optimum in the ML sense. But it does not explains what happens when the p.d.f. are not $q(\cdot)$. At this point we first pointed out that symmetry is an important property to take into account. It states the necessary and sufficient conditions to allow separation. We have gone further and exploit it to present a new algorithm with better convergence features.

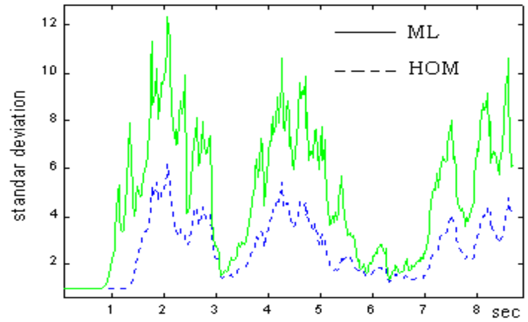

Fig. 6. Standard deviation of an element of the estimation equation for the ML and the HOM algorithms in Section V.A.

VII. REFERENCES

[1] S.I. Amari, A. Cichocki and H.H. Yang. "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, vol. 8, pp. 757- 763, MIT Press, 1996.

[2] S.I. Amari, "Natural Gradient works effciently in learning," *Neural Computation*, vol. 10, pp. 251-276, 1998.

[3] K. Anand, G. Mathew, and V. Reddy, "Blind separation of multiple co-channel BPSK signals arriving at an antenna array," *IEEE Signal Processing Letters*. , vol. 2, pp. 176-178, Sept. 1995.

[4] A. Back and A. Weigend, "A first aplication of independent component analysis to extracting structure from stock returns", *Int. J. Neural Sys*., vol. 8, no. 4, Aug. 1997.

[5] J.F. Cardoso and A. Souloumiac, "Blind beamforming for nongaussian signals," in *Proc. Inst. Elect. Eng.*, vol. 140, pt F., no. 6, pp. 395-401, Dec. 1993.

[6] B.Laheld and J.F. Cardoso., "Adaptive source separation with uniform performance, " *Proceedings of EUSIPCO-94*, Vol. 1, pp. 183-6, 1994.

[7] J.F. Cardoso, B. Laheld , "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 112-114, Apr 1997.

[8] J.F. Cardoso, "Estimating equations for source separation, " in *Proc. ICASSP*, vol. 5, pp. 3449-52, 1997.

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

[10] E. Chaumette, P. Comon and D. Muller, "ICA-based technique for radiating sources estimation: Aplication to airport surveillance," in *Proc. Inst. Electr. Eng*, vol. 140, pt F., no 6, pp. 395-401, Dec, 1993.

[11] A. Cichocki, W. Kasprazk and S. Amari, "Neural network approach to blind separation and enhancement of images," in *Proc. EUSIPCO-96*, Eighth European Signal Processing Conference, vol. 1, pp. 579-582, 1996.

[12] P. Comon "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287-314, April 1994.

[13] L. De Lathanuwer, B. De Moor and J. Vandewalle, "Fetal electro-cardiogram extraction by source separation", in *Proc. HOS'95*, Aiguabrava, Spain, pp 134-138, June 1995.

[14] S. Makeig, A. Bell, T.P. Jung and T.J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in Neural Information Processing Systems*, vol. 8, Denver, MIT Press, 1995.

[15] D.T. Pham, "Blind separation of instantaneous mixture of sources via an independent component analysis," *IEEE Trans. Signal Processing*, vol. 44, no. 11, pp. 2768-2779, nov. 1996.

[16] C.G. Puntonet and A. Prieto, "Neural net approach for blind separation of sources based on geometric properties," *Neurocomputing*, vol. 18, no 1-3, pp. 141-164, 1998.

[17] http://sound.media.mit.edu/ica-bench/sources/mike.wav and http://sound.media.mit.edu/ica-bench/sources/street.wav