A Globally Convergent Approach for Blind MIMO Adaptive Deconvolution

Azzédine Touzni, Inbar Fijalkow, *Member, IEEE*, Michael G. Larimore, *Member, IEEE*, and John R. Treichler, *Fellow, IEEE*

***Abstract*—We discuss the blind deconvolution of multiple input/multiple output (MIMO) linear convolutional mixtures and propose a set of *hierarchical* criteria motivated by the maximum entropy principle. The proposed criteria are based on the constant–modulus (CM) criterion in order to guarantee that all minima achieve perfectly restoration of different sources. The approach is moreover robust to errors in channel order estimation. Practical implementation is addressed by a stochastic adaptive algorithm with a low computational cost. Complete convergence proofs, based on the characterization of all extrema, are provided. The efficiency of the proposed method is illustrated by numerical simulations.**

***Index Terms*—Blind adaptive source separation, constant modulus criterion, multiple input/multiple output convolutional systems.**

I. Introduction

A. Problem Formulation

THE SO-CALLED signal deconvolution problem for multiple inputs/multiple output (MIMO) linear convolutional mixtures arises in a wide variety of signal processing and communications applications. It is a crucial issue in wireless multiuser digital communication systems, for instance, when the different users share parts of the same frequency band and are received on an omni-directional antenna. The restoration of multiple input signals is also required when two orthogonally polarized sources are mixed by multipath propagation with finite delay spread (see [27]). In this case, we know that the two sources are temporally and spatially mixed. The resulting undesirable effects, which must be suppressed, are, respectively, known as inter-symbol interference (ISI), corresponding to the perturbation from a delayed and scaled versions of the same signal, and inter-user interference (iui) for the additive perturbations occurring from the other incoming signals. In this paper, we consider the extraction of the input signal from the only knowledge of the output mixture, i.e., when weak *a priori* information is available on the channel (in particular, the geometry of the antenna manifold is supposed unknown) and where only some mild assumptions on the input signal are considered. Specifically, we do not consider a training approach using knowledge of part of the input sequence; see [16]. This problem, which is referred to as blind MIMO deconvolution or blind source separation of convolutional MIMO mixtures, is crucial for eavesdropping and when the training sequence is too short (semi-blind approaches).

B. Data Model

In order to address more precisely the problem, let us first introduce the channel propagation model and eight assumptions denoted A1)–A8). The justification of these assumptions will appear in the sequel.

Throughout this paper, we consider the MIMO convolutional linear mixture, for which the received signal on the kth sensor is

$$\tilde{y}_k(t) = \sum_{l=1}^{P} \sum_{n\in\mathbb{Z}} \tilde{h}_l^{(k)}(t-nT)s_l(n) + \tilde{w}_k(t), \qquad t \in \mathbb{R} \quad (1)$$

where $1/T$ is the symbol rate of all input signals. We suppose that each sensor receives a contribution from each of the P sources. The lth source contribution on the kth sensor is the convolutional mixture of the input sequence $s_l(n)$, drawn from a discrete alphabet, with the unknown impulse response $\tilde{h}_l^{(k)}(t)$. The impulse responses $\tilde{h}_l^{(k)}(t), l = 1 : P$ are assumed to have a finite time span and to incorporate the propagation channel effects due to various multipath propagation caused by obstacles or nonhomogeneities of the propagation media as well as pulse shaping and receiver filters. We denote by $\tilde{w}_k(t)$ the additive noise received at sensor k.

After sampling at baud rate[1] $1/T$, we obtain the discrete-time signal $y_k(n) = \tilde{y}_k(nT)$. The received signal collected on L sensors $\mathbf{y}(n) = (y_1(n), \ldots, y_L(n))^\top$ can now be written as

$$\mathbf{y}(n) = \sum_{l=1}^{P} \sum_{m=0}^{Q_l} \mathbf{h}_{l,m} s_l(n-m) + \mathbf{w}(n). \quad (2)$$

Let us now define $\mathbf{h}_l(z)$ ($1 \le l \le P$), which is a $L \times 1$ polynomial function, as

$$\mathbf{h}_l(z) = \sum_{p=0}^{Q_l} \mathbf{h}_{l,p} z^{-p} \quad (3)$$

where the kth component of $\mathbf{h}_l(z)$ ($1 \le k \le L$) $h_l^{(k)}(z) = \sum_{p=0}^{Q_{l,k}} h_{l,p}^k z^{-p}$ with $h_{l,p}^k = \tilde{h}_l^k(pT)$ for $p = 0, \ldots, Q_{l,k}$.

[1]Sampling at an higher rate can be considered to add temporal diversity to the spatial diversity of factor L.

Manuscript received May 20, 1998; revised July 14, 2000. The associate editor coordinating the review of this paper and approving it for publication was Prof. Arnab K. Shaw.

A. Touzni and I. Fijalkow are with Équipe de Traitement du Signal et des Images (ETIS), ENSEA—University de Cergy-Pontoise, Cergy-Pontoise, France (e-mail: fijalkow@ensea.fr).

M. G. Larimore and J. R. Treichler are with Applied Signal Technology Inc., Sunnyvale, CA 94086 USA (e-mail: mgl@appsig.com; jrt@appsig.com).

Publisher Item Identifier S 1053-587X(01)03888-0.

From the finite time-span of $\tilde{h}_l^{(k)}(t)$, it is understood that $Q_l \stackrel{\text{def}}{=} \max_{1\leq k\leq L} Q_{l,k}$ is finite, and it is called the degree of $\mathbf{h}_l(z)$. Then, the transfer function $\mathbf{H}(z) = (\mathbf{h}_1(z), \ldots, \mathbf{h}_P(z))$ is a $L\times P$ matrix of polynomials, and (2) can have the following compact notation:

$$\mathbf{y}(n) = [\mathbf{H}(z)]\mathbf{s}(n) + \mathbf{w}(n) \quad (4)$$

where $[\mathbf{H}(z)]\mathbf{s}(n)$ stands for the transfer function $\mathbf{H}(z)$ applied to the P-dimensional signal of interest $\mathbf{s}(n) = (s_1(n), \ldots, s_P(n))^\top$. The L-dimensional noise contribution $\mathbf{w}(n) = (w_1(n), \ldots, w_L(n))^\top$ is defined from $w_l(n) = \tilde{w}_l(nT)$.

We assume the following in the sequel.

A1) $L > P$ (strictly more sensors than sources).

A2) $\mathtt{Span}\{\mathbf{h}_1(z), \ldots, \mathbf{h}_P(z)\} = P$ for all $z \in \mathbb{C}$, where $\mathtt{Span}$ means the subspace spanned by the vectors between brackets.

A3) $\mathbf{H}(z)$ is a column reduced matrix, i.e., $\mathrm{Rank}\{\mathbf{h}_{1,Q_1}, \ldots, \mathbf{h}_{P,Q_P}\} = P$, where $\mathbf{h}_{l,Q_l}$ denotes the highest degree term of the polynomial $\mathbf{h}_l(z)$.

A4) Each $s_l(n)$ is an independent and identically distributed sequence with $\mathbb{E}\{s_l(n)\} = 0$, $\mathbb{E}\{s_l^2(n)\} = \sigma_l^2$. $s_l(n)$ is circular and sub-Gaussian, i.e., its normalized kurtosis $\rho_l \stackrel{\text{def}}{=} \mathbb{E}\{|s_l(n)|^4\}/\sigma_l^4$ satisfies $\rho_l < 3$ in the real-valued case.

A5) The sources are mutually independent.

A6) The sources are independent from the noise vector $\mathbf{w}(n)$.

A7) $N \geq \sum_{k=1}^{P} Q_k$, where N will define the length of received data processed per equalizer output.

A8) $\beta_k \geq 2\max_{m=1,P} \rho_m$ (will be needed in Section IV).

C. Related Work

Next, we give an overview of the main contributions in blind source separation of convolutional MIMO mixtures and introduce the motivation for the proposed approach.

1) Second-Order Statistics Methods: Most recent methods on blind deconvolution of MIMO convolutional mixtures are based on second-order statistics (SOS) and can be viewed as extensions of blind channel identification approaches for the single-input/multiple-output (SIMO) case. In [30], a solution based on cancellation of all covariance matrices of the observation $\mathbf{y}(n)$ corresponding to all possible delays is proposed (see also [17] for a left inverse identification and the pioneering contributions of Gardner [10] and Tong [25] for the SIMO case). From the covariance matrix of a vector collecting the multiples observations $(\mathbf{y}(n-k))_{0\leq k\leq K}$, a signal-subspace approach is proposed in [12], generalizing the SIMO subspace approach of Moulines *et al.* [19]. A linear prediction method is also proposed in [12] in order to estimate a left inverse (see [1] for the SIMO case).

Unfortunately, it is known that all these methods suffer from lack of robustness. In particular, they rely on a perfect knowledge of the degrees of the columns of the transfer functions in order to ensure uniqueness of the solution. Furthermore, SOS methods cannot solve the mixture separation problem. They reduce the problem to the instantaneous source separation problem, which requires more than SOS knowledge. This two-stage estimation approach results most often in a nonadaptive implementation. In order to avoid a two-stage procedure, we consider in the sequel one-stage HOS-based approaches.

2) High-Order Statistics Methods: Although many approaches based on HOS exist for the specific problem of instantaneous mixture separation [i.e., when $H(z) = H(0)$], very few results are available for convolutional mixtures. One can, however, mention the significant contribution of Yellin and Weinstein for a 2×2 channel model. They show that under mild assumptions on the input signal statistical distribution, cancellation of outputs' cross-cumulants leads to the separation of a convolutional mixture [31]. Independently, Thi *et al.* proposed algorithms based on the cancellation of high-order moments generated by nonlinear functions [15], [21]). Unfortunately, the resulting cost functions often suffer from undesired local minima. This is a major drawback for HOS-based methods, in particular when we are interested in the development of adaptive algorithms.

Extensions of equalization approaches using HOS were also proposed based on successive restoration of the sources. A *deflation* approach is proposed by Delfosse and Loubaton in [7]: a multistage maximization procedure using the Shalvi-Weinstein criterion in [13]. A subtraction approach based on the *constant modulus* (CM) cost function was first investigated in [27] and updated by Tugnait in [29] using some invertibility properties. The approach consists of subtracting the contribution of the estimated signal from the mixtures in order to estimate next a different source. Unfortunately, derived adaptive algorithms suffer from highly increasing variance of estimation as the number of sources increases.

More recently, some constrained approaches relying on the CM ability to capture one source from convolutive mixtures (to be recalled in Sections IV and V) have been proposed independently in [22] and [24] (see [4] for the instantaneous mixtures case). They use decorrelation constraints in order to tune several filters to simultaneously capture different sources. However, spurious minima may be caused by the additive decorrelation constraints.

In this paper, we propose a set of composite *hierarchical* criteria motivated by the maximum entropy principle in order to guarantee that all minima achieve perfectly restoration of different sources. Practical implementation is addressed with an efficient stochastic adaptive algorithm.

D. Organization

The paper is organized as follows. In Section II, we recall results on invertibility of MIMO transfer functions. In Section III, we introduce the family of hierarchical criteria resulting from the maximum entropy principle and the associated adaptive optimization algorithm. In Section IV, we characterize the stationary points and, in Section V, the stable extrema. Illustrative simulations are reported in Section VI. The conclusion is given in Section VII.

Note that all the analysis is performed in the noise-free context. Moreover, for reasons of simplicity, we have chosen to derive all proofs in the real-valued case.

II. Channel Invertibility

In this section, we formulate the conditions for the left-invertibility of the channel $\mathbf{H}(z)$. The observations $(\mathbf{y}(n-m))_{0\le m\le N-1}$ are collected in the regressor vector $\mathbf{Y}_N(n)=(\mathbf{y}(n)^\top, \mathbf{y}(n-1)^\top, \ldots, \mathbf{y}(n-(N-1))^\top)^\top$ of dimension NL. Using the model as in (4)

$$\mathbf{Y}_N(n)=\sum_{l=1}^{P}\mathcal{T}(\mathbf{h}_l)\mathbf{S}_l(n)=\mathcal{T}(\mathbf{H})\mathbf{S}(n) \tag{5}$$

where the input signal $\mathbf{S}(n)=(\mathbf{S}_1(n)^\top, \ldots, \mathbf{S}_P(n)^\top)^\top$ is of dimension $NP+\sum_{k=1}^{P}Q_k$ with entry $\mathbf{S}_l(n)=(s_l(n), \ldots, s_l(n-(N+Q_l+1)))^\top$, which is a vector of dimension $N+Q_l$ that contains the contributions of the source $s_l(n)$ for different delays. For convenience, we introduce the notation $K\overset{\text{def}}{=}NP+\sum_{k=1}^{P}Q_k$. The MIMO convolution matrix, of dimension $NL\times K$, is written as $\mathcal{T}(\mathbf{H})=(\mathcal{T}(\mathbf{h}_1), \ldots, \mathcal{T}(\mathbf{h}_P))$. The matrix $\mathcal{T}(\mathbf{h}_l)$ denotes the SIMO Sylvester convolution matrix of dimension $NL\times(N+Q_l)$ between the source $s_l(n)$ and the L-dimensional sensors

$$\mathcal{T}(\mathbf{h}_l)=\begin{pmatrix}\mathbf{h}_{l,0} & \cdots & \mathbf{h}_{l,Q_l} & \mathbf{0}_{L\times 1} & \cdots\mathbf{0}_{L\times 1}\\ & \ddots & & \ddots & \\ \mathbf{0}_{L\times 1}\cdots & \mathbf{0}_{L\times 1} & \mathbf{h}_{l,0} & \cdots & \mathbf{h}_{l,Q_l}\end{pmatrix}.$$

Restoring M input sources $s_k(n)$ (with $1\le M\le P$) leads to estimate a set of M polynomial vectors $\mathbf{g}_k(z)=\sum_{p=0}^{N-1}\mathbf{g}_{k,p}z^{-p}$ of dimension $L\times 1$, which are referred to as *separating* vectors, according to the following definition.

Definition 1: The polynomial filter $\mathbf{g}_k(z)=\sum_{p=0}^{N-1}\mathbf{g}_{k,p}z^{-p}$ of dimension $L\times 1$ and degree $N-1$ is a separating filter if and only if

$$[\mathbf{g}_k(z)^\top]\mathbf{y}(n)=\sum_{p=0}^{N-1}\mathbf{g}_{k,p}^\top\mathbf{y}(n-p)=\lambda_{\sigma(k)}s_{\sigma(k)}(n-\nu_{\sigma(k)})$$

where $\sigma(k)$ defines any permutation between the indices. λ_k is any nonzero scalar factor, and ν_k denotes any possible delay. A separating filter is defined equivalently if and only if

$$\mathbf{g}_k^\top\mathcal{T}(\mathbf{H})\overset{\text{def}}{=}(f_{(1),k}^\top, \ldots, f_{(l),k}^\top, \ldots, f_{(P),k}^\top)^\top=\lambda_{\sigma(k)}u_{\sigma(k)}$$

where $\mathbf{g}_k$ is the NL-long impulse response defined as $\mathbf{g}_k=(\mathbf{g}_{k,0}^\top, \ldots, \mathbf{g}_{k,N-1}^\top)^\top$. u_k is a canonical vector of length K selecting the kth source[2] with delay ν_k

$$u_k=\begin{cases} f_{(l),k}=\delta_{(l)}^{\nu_k}=(0\cdots010\cdots0)^\top, & \text{for } l=k\\ f_{(l),k}=\mathbf{0}_{N+Q_l}, & \text{for } l\ne k\end{cases}$$

(the 1 is the ν_k+1 component). Notation $f_{(l)}$ stands for the vector of length $N+Q_L$ corresponding to the contribution of source s_l.

[2]Note that if $k\ne k'$, the corresponding separating vector $\mathbf{g}_k$ and $\mathbf{g}_{k'}$ always select different sources.

An important feature concerning the restoration of the input signal $(s(n))_{n\in\mathbb{Z}}$ relies on the left invertibility of the matrix $\mathcal{T}(\mathbf{H})$.

Lemma 1 [12]: Under A-1), A-2), A-3), and A-7), $\mathcal{T}(\mathbf{H})$ is full-column rank. These assumptions are referred as the *left-invertibility conditions* of the convolutional mixing transfer function $\mathbf{H}(z)$.

Under the system left-invertibility conditions, all combined channel-receiver impulse responses $f_k^\top=\mathbf{g}_k^\top\mathcal{T}(\mathbf{H})$ are achievable, in particular, the separating solutions leading to the exact solution $f_{(l)}=\delta_{(l)}^{\nu_k}$ for $l=k$ and $f_{(l)}=\mathbf{0}_{N+Q_l}$ for $l\ne k$, corresponding to the vector u_k previously defined.

The estimation of separating vectors $(\mathbf{g}_k^\star)_{1\le k\le P}$, under the left invertibility assumption of $\mathbf{H}(z)$, is investigated in the following sections.

III. Hierarchical Criteria

A. Maximum Entropy Principle

In this subsection, we give preliminary results connected to the maximum entropy principle, and we define the notion of *hierarchical* criteria.

Let $z_A=(z_k(n))_{k\in A}$ be a set of filtered outputs with $z_k(n)=\mathbf{g}_k^\top\mathbf{Y}_N(n)$, where $A\subset\{1, \ldots, P\}$. The maximum (Shannon) entropy principle [14] consists of looking for a set of vectors $\mathbf{g}_A^\star$, where $\mathbf{g}_A=(\mathbf{g}_k)_{k\in A}$, such that

$$\begin{aligned}\mathbf{g}_A^\star&=\arg\max_{\mathbf{g}_A}\mathcal{H}_A(\varphi(z_A); \mathbf{g}_A)\\ \varphi(z_A)&=(\varphi_1(z_1), \ldots, \varphi_{|A|}(z_{|A|}))\end{aligned} \tag{6}$$

where we suppose that $\varphi_k\colon\mathbb{R}\to[0, 1]$ is a given real nonlinear function differentiable and monotonously increasing. Note that the index k does not refer specifically to the source s_k. The entropy can be expressed as

$$\begin{aligned}\mathcal{H}_A(\varphi(z_A); \mathbf{g}_A)&=-\mathbb{E}\{\log p(\varphi(z_A)\mathbf{g}_A)\}\\ &=-\int p(\varphi(z_A); \mathbf{g}_A)\\ &\qquad\log p(\varphi(z_A); \mathbf{g}_A)\, d\varphi(z_A).\end{aligned}$$

Using entropy additivity for monotonically transformed vectors as in [2], one may check that for any set of indices A', the entropy can be rewritten as

$$\begin{aligned}&\mathcal{H}_{A'\cup A}(\varphi(z_{A'\cup A}); \mathbf{g}_{A'}, \mathbf{g}_A)\\ &\quad=\mathcal{H}_A(\varphi(z_A); \mathbf{g}_A)-\sum_{k\in A'}\mathcal{K}(p(z_k; \mathbf{g}_k)\|\varphi_k'(z_k); \mathbf{g}_k)\\ &\qquad-\mathcal{K}(p(z_{A'}; \mathbf{g}_{A'}|z_A)\|p_\#(z_{A'}; \mathbf{g}_{A'}))\end{aligned} \tag{7}$$

where $\mathcal{K}(p\|q)=\int p(z)\log(p(z)/q(z))\, dz$ denotes the Kullback–Leibler divergence between the probability density p and q. $p_\#(z_{A'}; \mathbf{g}_{A'})$ is the product of the marginal distributions $\prod_{k\in A'}p(z_k; (\mathbf{g}_k)_{k\in A'})$.

Proof: From [2], $\mathcal{H}_A(\varphi(z_A))=\mathcal{H}_A(z_A)+\sum_{k\in A}\mathbb{E}\{\log\varphi_k'(z_k)\}$ (we omit to note the filters $\mathbf{g}$ in the proof).

Then, using Bayes's rule, and the definition of $\mathcal{K}$

$$\begin{aligned}\mathcal{H}_{A'\cup A}(\varphi(z_{A'\cup A})) &= -\mathcal{K}(p(z_{A'}|z_A)\|p_\#(z_{A'})) \\ &\quad \underbrace{-\int p(z_A)\log(p(z_A))\,dz_A + \sum_{k\in A}\mathbb{E}\{\log\varphi'_k(z_k)\}}_{\mathcal{H}_A(\varphi(z_A))} \\ &\quad \underbrace{-\int p(z_{A'})\log(p_\#(z_{A'}))\,dz_{A'} + \sum_{k\in A'}\mathbb{E}\{\log\,\varphi'_k(z_k)\}}_{-\sum_{k\in A'}\mathcal{K}(p(z_k)\|\varphi'_k(z_k))}.\end{aligned}$$

■

Expression (7) implies that if $(\mathbf{g}_k^\star)_{k\in A}$, with $|A| = K$, is a subset of K separating vectors, the estimation of a new separating vector $\mathbf{g}_{A_1}^\star$ (with $|A_1| = 1$) corresponds to

$$\mathbf{g}_{A_1}^\star = \arg\min_{\mathbf{g}_{A_1}}\{\mathcal{K}(p(z_{A_1};\mathbf{g}_{A_1})\|\,\varphi'_{A_1}(z_{A_1});\mathbf{g}_{A_1}) + \mathcal{K}(p(z_{A_1};\mathbf{g}_{A_1}|z_A^\star)\|p(z_{A_1};\mathbf{g}_{A_1}))\}. \quad (8)$$

Note that since $|A_1| = 1$, $p_\#(z_{A_1};\mathbf{g}_{A_1}) = p(z_{A_1};\mathbf{g}_{A_1})$. $\mathcal{K}(p(z_{A_1};\mathbf{g}_{A_1})\|\varphi'_{A_1}(z_{A_1});\mathbf{g}_{A_1})$ corresponds to the extraction of a source provideing a good choice of φ_{A_1}. From the instantaneous source separation case, it is known that a good choice for the nonlinearity is given by the cumulative distribution function of the source to be extracted (i.e., φ'_{A_1} is the source pdf); see, for instance, [23]. In the convolutive case, if φ'_{A_1} is the source pdf, it can be shown that the entropy global maxima are separating sources. The proof of this claim can be deduced from [23]; see [28]. The second term $\mathcal{K}(p(z_{A_1};\mathbf{g}_{A_1}|z_A^\star)\|p(z_{A_1};\mathbf{g}_{A_1}))$ is understood to be a measure of independence between the $z_A^\star$ and z_{A_1} since $\mathcal{K}(p(z_{A_1};\mathbf{g}_{A_1}|z_A^\star)\|p(z_{A_1};\mathbf{g}_{A_1})) = 0$ if z_{A_1} is independent of $z_A^\star$. In other words, the minimization of the above criterion (8) results in the minimization of a cost function $\mathcal{K}(p(z_{A_1};\mathbf{g}_{A_1})\|\varphi'_{A_1}(z_{A_1});\mathbf{g}_{A_1})$, leading to recover an input source, under independence constraints related to the other sources "previously" estimated and given by $\mathcal{K}(p(z_{A_1};\mathbf{g}_{A_1}|z_A^\star)\|p(z_{A_1};\mathbf{g}_{A_1}))$. The point of interest is that the estimated vector $\mathbf{g}_{A_1}^\star$ leads necessarily to the estimation of a new source $s_{A_1}(n)$ with $A_1 \notin A$.

According to this result, a general guiding principle is provided for the restoration of input signals by the following procedure:

$$\begin{cases}\mathbf{g}_1^\star = \arg\min_{\mathbf{g}_1}\mathcal{K}(p(z_1;\mathbf{g}_1)\|\varphi'_1(z_1);\mathbf{g}_1) \\ \mathbf{g}_k^\star = \arg\min_{\mathbf{g}_k}\{\mathcal{K}(p(z_k;\mathbf{g}_k)\|\varphi'_k(z_k);\mathbf{g}_k) \\ \qquad\qquad + \mathcal{K}(p(z_k;\mathbf{g}_k|(z_j^\star)_{j<k})\|p(z_k;\mathbf{g}_k))\}.\end{cases} \quad (9)$$

This set of criteria is based on a *hierarchical* principle since each new vector $\mathbf{g}_k^\star$ is estimated with a new cost function that depends on the previous criteria minima.

B. New Cost Function Definition

In this subsection, we introduce a set of hierarchical criteria motivated by the structure in (9). Since we do not know *a priori* the sources pdfs (in the digital communication case, the sources alphabets), the choice of the proposed set of criteria is motivated by the thought that the hierarchical structure will force the kth filter to be a separating one selecting a different source if we can be sure that $(\mathbf{g}_s)_{s<k}$ are separating filters. Therefore, the first criterion needs only to be able to extract one source from the mixtures. To do so, we choose the CM criterion because it is known (see [27], for a first statement) to be able to capture one source from mixtures. The exact conditions for a capture are derived in Sections IV and V, which is an original result when the sources have different kurtosis ρ_j.

We propose the following set of hierarchical criteria $(\Phi_k)_{1\le k\le M}$:

$$\begin{aligned}\Phi_1(\mathbf{g}_1) &= \Phi_c(\mathbf{g}_1) \\ \Phi_k(\mathbf{g}_k/(\mathbf{g}_s)_{s<k}) &= \Phi_c(\mathbf{g}_k) + \beta_k\sum_{j=1}^{k-1}\sum_{m=-T}^{+T}\mathcal{E}_m(\mathbf{g}_k,\,\mathbf{g}_j)\end{aligned} \quad (10)$$

where $\Phi_c(\mathbf{g}_k)$ denotes the CM cost function [11], [26] defined in the real-valued case as

$$\Phi_c(\mathbf{g}_k) = \mathbb{E}\{((z_k(n))^2 - r)^2\}$$

where r is an *a priori* constant dispersion. The second term of (10) $\mathcal{E}_m(\mathbf{g}_k,\,\mathbf{g}_j) = \mathbb{E}\{z_k(n)z_j(n-m)\}^2$ is a quadratic function in terms of $\mathbf{g}_k$ and $\mathbf{g}_j$ based on SOS of the observations. It corresponds to a decorrelation constraint between $z_k(n)$ and $z_j(n-m)$, where the delay m covers $[-T, +T]$, with $T \ge \max_k N + Q_k$. If the filters outputs $(z_j(n))_{j=1,\ldots,k}$ are true input sources, the decorrelation term becomes a measure of independence. In the sequel, we show that the decorrelation term is a simple and sufficient constraint, requiring no *a priori* knowledge as in $\mathcal{K}(p(z_k|(z_j^\star)_{j<k})\|p_\#(z_k))$ to guarantee that each criterion Φ_k leads to the selection of a the different source when A-8) is satisfied. β_k is a positive constant that was introduced to control the constraint level. Its effect on (10) is studied in the sequel.

C. Adaptive Optimization

For each function $\Phi_k(\mathbf{g}_k/(\mathbf{g}_s)_{s<k})$, we propose to derive a simple stochastic gradient descent algorithm for minimizing simultaneously the criteria $(\Phi_k)_{1\le k\le M}$

$$\mathbf{g}_k^{(n+1)} = \mathbf{g}_k^{(n)} - \mu\tfrac{1}{4}\nabla_{g_k}\Phi_k\left(\left(\mathbf{g}_k\Big/\left(\mathbf{g}_s^{(n+1)}\right)_{s<k}\right)\right|_{\mathbf{g}_k=\mathbf{g}_k^{(n)}} \quad (11)$$

for $k = 1,\ldots,M$, with μ a small positive step-size. The stochastic gradient is given by the expression $\nabla_{g_k}\Phi_k(\mathbf{g}_k) = \nabla_{g_k}\Phi_c(\mathbf{g}_k) + \beta_k\sum_{j=1}^{k-1}\sum_m\nabla_{g_k}\mathcal{E}_m(\mathbf{g}_k,\,\mathbf{g}_j)$ where $1/4\nabla_{g_k}\Phi_c(\mathbf{g}_k) = (z_k(n)^2 - r)z_k(n)\mathbf{Y}_N(n)$, $z_k(n) = \mathbf{g}_k^{(n)\top}\mathbf{Y}_N(n)$, and where the second term is $1/2\nabla_{g_k}\mathcal{E}_m(\mathbf{g}_k,\,\mathbf{g}_j) = (\mathbf{g}_k^\top\hat{\mathbf{R}}_Y(m)\mathbf{g}_j)\hat{\mathbf{R}}_Y(m)\mathbf{g}_j$. $\hat{\mathbf{R}}_Y(m)$ is the estimation of covariance matrix $\mathbb{E}\{\mathbf{Y}_N(n)\mathbf{Y}_N(n-m)^\top\}$, which can be estimated recursively by

$$\hat{\mathbf{R}}_Y(m)^{(n+1)} = \hat{\mathbf{R}}_Y(m)^{(n)} + \lambda\left(\hat{\mathbf{R}}_Y(m)^{(n)} - \mathbf{Y}_N(n)\mathbf{Y}_N(n-m)^\top\right) \quad (12)$$

where λ is a small positive constant. An important point is that for small enough μ and λ, the asymptotic convergence points (in mean) of the algorithm (11) are exactly the minima of the criteria

$(\Phi_k)_{1\le k\le M}$ since the estimator of $\mathbf{R}_Y$ in (12) is unbiased; see [3]. In particular, convergence to a saddle point is not possible.

In the following sections, we propose an analysis of the extrema of $(\Phi_k)_{1\le k\le M}$. We will show in particular that the proposed criteria associated with an appropriate choice of β_k do not admit spurious local minima, i.e., all minima correspond to separating filters restoring different signals, in contrast to the symmetrical criteria used in [22] and [24].

IV. Extrema Analysis

Because of the hierarchical structure, the extrema solutions of the cost function $\Phi_k(\mathbf{g}_k/(\mathbf{g}_s)_{s<k})$ depend on the extrema solutions of $(\Phi_s)_{s<k}$. Thus, we need to substitute all extrema solutions of $(\Phi_s)_{s<k}$ into $\Phi_k(\mathbf{g}_k/(\mathbf{g}_s)_{s<k})$ in order to find the extrema in terms of $\mathbf{g}_k$. The analysis of the extrema of Φ_k is therefore derived by induction. We first establish that $\Phi_1=\Phi_c$ minima are separating solutions. Then, assuming that $(\mathbf{g}_j^\star)_{j<k}$ is a set of separating solutions corresponding to the minima of the criteria $(\Phi_k)_{k<s}$, we investigate the extrema of $\Phi_k(\mathbf{g}_k/(\mathbf{g}_j^\star)_{j<k})$ in terms of $\mathbf{g}_k$. We derive a condition on β_k in order to prevent ourselves from capturing a source restored by $(\Phi_j)_{j<k}$. Finally, we address the stability of the separating solution $\mathbf{g}_k^*$ and the stability of other possible extrema.

A. Property of the $\Phi_c(\mathbf{g})$ *Extrema*

In this subsection, we study the extrema of Φ_c: the CM cost-function under the invertibility conditions stated in Lemma 1 [see A-1)–A-3), and A-7)]. The analysis is derived from calculation of the gradient with respect to the vector g and expressed in terms of the overall impulse response $f=\mathcal{T}(\mathbf{H})^\top\mathbf{g}$.

Proposition 1: If we denote $f^c=(f_{(1)}^{c\top},\ldots,f_{(l)}^{c\top},\ldots,f_{(P)}^{c\top})^\top$, the extrema points of Φ_c

$$f_{(l)}^c=\begin{cases}\displaystyle\sum_{p\in I_l}\pm\,\omega_{c,l}\delta_{(l)}^p\\ \text{with } \omega_{c,l}^2=\dfrac{r}{\sigma_l^2(\rho_l-3)+3\displaystyle\sum_k\nu_k\sigma_l^2\,\dfrac{(\rho_l-3)}{(\rho_k-3)}}\\ \text{or } \mathbf{0}_{N+Q_l}\end{cases}\tag{13}$$

where $I_l\subseteq\{0,1,\ldots,N+Q_l-1\}$ with $|I_l|=\nu_l$, where ν_l denotes the number of nonzero components in each subvector $f_{(l)}^c$. Moreover, it is understood that $\delta_{(l)}^p$ and $\mathbf{0}_{N+Q_l}$ are, respectively, a canonical vector [with nonzero components at the $(p+1)$th entry] and a null vector of dimension $N+Q_l$.

The set of all extrema points corresponding to the points $f^c\neq\mathbf{0}$ verify

$$\frac{r}{3}\le\|f^c\|_s^2\le\min_l\frac{r}{r_l}\tag{14}$$

where $\|f^c\|_s^2\overset{\text{def}}{=}\sum_{j=1}^P\sigma_j^2\|f_{(j)}^c\|_2^2$, where it is understood that $\|f_{(j)}^c\|_2^2=\sum_{p=1}^{N+Q_j}f_{(j)}^c(p)^2$.

Proof: See Appendix A.

The extrema in (13) give a mixture of the sources for $\nu_l>1$, and we will show in Section V that they are unstable. The only solutions corresponding to the (perfect) restoration of one source are of the form

$$f^\star=\pm\sqrt{\frac{r}{r_l}}(0\cdots0\cdots\underbrace{010}_{l}\cdots0\cdots0)=\pm\sqrt{\frac{r}{r_l}}u_l$$

where $r_l\overset{\text{def}}{=}\mathbb{E}\{s_l^4\}/\sigma_l^2=\rho_l\sigma_l^2$ is the dispersion constant associated with the lth source.

In conclusion, under the conditions defined above, $\Phi_c(\mathbf{g})$ admits two sets of extrema: a set of separating solutions $\mathbf{g}^\star=\mathbf{g}_1^\star$ leading to a global impulse response of the form $f=f_1^\star$ and a set of solutions given by (13), selecting linear combinations of the sources. In the next subsection, we characterize the minima of (10) in terms of β_k and exhibit a simple condition to suppress the undesired solutions.

B. Extrema Setting of $\Phi_k(\mathbf{g}_k/(\mathbf{g}_j^\star)_{j<k})$

An analytical analysis of the $(\Phi_k)_{1\le k\le M}$ extrema is proposed in this section. The extrema are defined by zeroing the gradient $\nabla_{g_k}\Phi_k$. If we introduce separating solutions $(\mathbf{g}_j^\star)_{j<k}$ in Φ_k, characterization of the extrema settings of cost function Φ_k in terms of $\mathbf{g}_k$ becomes easily derivable. For each criterion Φ_k, we may *a priori* classify the extrema in separating solutions in terms of overall impulse response $f_k^\star=\mathcal{T}(\mathbf{H})^\top\mathbf{g}_k^\star$ belonging to the subset $\mathcal{F}_k^*$ of separating vectors and other solutions of the form $\overline{f}_k=\mathcal{T}(\mathbf{H})^\top\overline{\mathbf{g}}_k$ belonging to $\overline{\mathcal{F}}_k$ as the subset of nonseparating extrema. We denote $\mathcal{F}_k=\mathcal{F}_k^\star\cup\overline{\mathcal{F}}_k$ as the set of all extrema of $\Phi_k(\mathbf{g}_k/(\mathbf{g}_j^\star)_{j<k})$ in terms of g_k. We give, in (15), the mean gradient equation that must be solved to define these extrema.

For a given set of vectors $f_j^\star\in\mathcal{F}_j^\star$ such that $1\le j\le k-1$, the extrema $f_k\in\mathcal{F}_k$ of Φ_k are solutions of the following equation, which is derived in Appendix B:

$$\mathcal{P}_s\Delta_s(f_k)f_k+r\frac{\beta_k}{2}\sum_{l\in I_{k-1}}\sum_{\epsilon=1}^{N+Q_l}\frac{\sigma_l^4}{r_l}f_{(l),k}(\epsilon)\delta_{l+\epsilon}=0_K\tag{15}$$

where $\mathcal{P}_s\Delta_s(f_k)f_k$ is the gradient of the CM cost derived in Appendix A. I_{k-1} is a subset of dimension $k-1$, whichs contain the subscripts of all sources selected by criteria $(\Phi_s)_{s<k}$. δ_ϵ denotes a canonical vector of length K, with 1 on the $(\epsilon+1)$th entry and 0_K a null vector of the same dimension. The first term of (15) corresponds to the CM criterion (see Appendix B to get the full expressions) and the second term to the additive constraints. The solutions of equation (15) are now described.

Proposition 2: The extrema $f_k=(f_{(1),k}^\top,\ldots,f_{(l),k}^\top,\ldots,f_{(P),k}^\top)^\top\in\mathcal{F}_k$ of $\Phi_k(\mathbf{g}_k/(\mathbf{g}_j^\star)_{j<k})$ verify

$$f_{(l),k}=\sum_{p\in I_l}\pm\omega_l\delta_{(l)}^p\qquad\text{or}\qquad f_{(l),k}=0_{N+Q_l}\tag{16}$$

where

$$\omega_l^2=\begin{cases}\omega_{c,l}^2\left(1+\dfrac{\beta_k}{2}\left(\displaystyle\sum_{j\in I_{k-1}}\frac{3\nu_j}{\rho_j(\rho_j-3)}-\sum_j\frac{3\nu_j}{\rho_l(\rho_j-3)}-\frac{1}{\rho_l}\right)\right)\\ \qquad\text{for } l\in I_{k-1},\\ \omega_{c,l}^2\left(1+\dfrac{\beta_k}{2}\displaystyle\sum_{j\in I_{k-1}}\frac{3\nu_j}{\rho_j(\rho_j-3)}\right),\qquad\text{for } l\in\overline{I}_{k-1}.\end{cases}\tag{17}$$

$\overline{I}_{k-1}$ is the complement of I_l. $\omega_{c,l}$ is the solution of the CM contribution given in Proposition 1. For convenience, we denote $\omega_l = \overline{\omega}_l$ as the components for which the subscripts $l \in \overline{I}_{k-1}$ and $\omega_l = \underline{\omega}_l$ as those for which $l \in I_{k-1}$.

Moreover, each extrema solution f_k verifies

$$\|f_k\|_s^2 = \|f_k^c\|_s^2 - \frac{\beta_k}{2}\,\frac{r\displaystyle\sum_{j\in I_{k-1}}\frac{\nu_j}{\rho_j(\rho_j-3)}}{1+3\displaystyle\sum_{j\in I_{k-1}\cup\overline{I}_{k-1}}\frac{\nu_j}{\rho_j-3}} \tag{18}$$

where $\|f_k^c\|_s^2$ denotes the norm of the CM solution given in Proposition 1.

Proof: The proof is similar to that of Proposition 1 (Appendix A). It is based on solving (15). If we denote f_k as the extrema solution, the components of each subvector $f_{(m),k}$ are solution of the equation $(\mathcal{P}_s\ \Delta_s(f_k)f_k)_{(m)} + (\beta_k/2)\sigma_m^4(r/r_m)f_{(m),k} = \mathbf{0}_{N+Q_m}$ if $m \in I_{k-1}$, and $(\mathcal{P}_s\Delta_s(f_k)f_k)_{(m)} = 0_{N+Q_m}$ if $m \in \overline{I}_{k-1}$, where $(\mathcal{P}_s\Delta_s(f_k)f_k)_{(m)} = \sigma_m^2(3\|f_k\|_s^2 - r) + \sigma_m^4(\rho_m-3)\omega_m^2$. Thus, $\omega_m^2 = 0$ or $\omega_m^2 = \underline{\omega}_m^2 = (r - 3\|f_k\|_s^2 - (\beta_k/2)\sigma_m^4(r/r_m))/(\sigma_m^2(\rho_m-3))$ for $m \in I_{k-1}$ or $\omega_m^2 = \overline{\omega}_m^2 = (r-3\|f_k\|_s^2)/(\sigma_m^2(\rho_m-3))$ for $m \in \overline{I}_{k-1}$. If we denote ν_m as the number of nonzero components in each block $f_{(m),k}$, we have $\|f_k\|_s^2 = \sum_{m\in I_{k-1}}\sigma_m^2\nu_m\underline{\omega}_m^2 + \sum_{m\in\overline{I}_{k-1}}\sigma_m^2\nu_m\overline{\omega}_m^2$. A straightforward calculation leads to (18). By introducing the result of $\|f_k\|_s^2$ in the previous expression $\overline{\omega}_m$ and $\underline{\omega}_m$, we arrive at (17). ■

The expression of the separating solution $f_k^\star \in \mathcal{F}_k^\star$ of Φ_k is then easily established. The solution is expressed in terms of the constraint level β_k in the next proposition.

Proposition 3: The separating solution $f_k^\star \in \mathcal{F}_k^\star$ of $\Phi_k(\mathbf{g}_k, (\mathbf{g}_j^\star)_{j<k})$ is described by

$$f_{(l),k}^\star = \begin{cases} \pm\sqrt{\dfrac{r}{r_l}\left(1-\dfrac{\beta_k}{2\rho_l}\right)}u_l, & \text{for } l \in I_{k-1} \\ \pm\sqrt{\dfrac{r}{r_l}}\,u_l, & \text{elsewhere.}\end{cases} \tag{19}$$

The selection of $l \in I_{k-1}$ means that the restored source at the kth step has already been selected once. According to the previous result, one can remark that it is always possible to select the level constraint β_k in order to avoid separating solutions given by previous criteria $(\Phi_j)_{j<k}$. We get the simple condition

$$\beta_k \geq 2\rho_l \tag{20}$$

where we recall that $\rho_l = \mathbb{E}\{s_l^4\}/\mathbb{E}\{s_l^2\}^2$ in the real case. Actually, to prevent selection of the same source, we only have to verify that each criterion has a level constraint such as $\beta_k \geq 2\max_{m=1,P}\rho_m$ [see A-8)]. In this case, $f_k^\star$ does not belong to the subset $(\mathcal{F}_j)_{j\in I_{k-1}}$. We would like to use the smallest possible value of β_k in order to not disturb the good behavior of the CMA. Roughly speaking, the knowledge about the minimal β_k is similar to the knowledge of the dispersion constant. Note that this condition is not too restrictive in practice since for QAM, sources $\rho < 2$ so that β_k can be set to 4.

Proof: The result is deduced from Proposition 3. Let l the subscript of the nonzero component ω_l. We have $\nu_k = 1$ if $k = l$ and $\nu_k = 0$ if $k \neq l$. For $l \in \overline{I}_{k-1}$ (we have $\nu_j = 0$ for all $j \in I_{k-1}$), $\overline{\omega}_l = \omega_{c,l}$, where $\omega_{c,l} = \pm\sqrt{r/r_l}$ (see Appendix A). Solution $f_k^\star$ turns to $f_{(m),k}^\star = 0_{N+Q_m}$ if $m \neq l$ and $f_{(m),k}^\star = \omega_{c,l}\delta_{(l)}^p$ if $m \neq l$ where p is an arbitrary element of $\{0, 1, \ldots, , N+Q_m-1\}$. The compact expression is $f_k^\star = \sqrt{r/r_l}\,u_l$. For $l \in I_{k-1}$, we have $\underline{\omega}_l^2 = \omega_{c,l}^2(1+\beta_k/2(3/\rho_l(\rho_l-3) - 3/\rho_l(\rho_l-3) - 1/\rho_l))$, which leads to $f_k^\star = \pm\omega_{c,l}\sqrt{(1-\beta_k/2\rho_l)}u_l$. ■

In the next section, we prove that the separating solutions belonging to $\mathcal{F}_k^\star$ are the only global minima. The others extrema belonging to $\overline{\mathcal{F}}_k$ are necessarily unstable extrema.

V. Stability Analysis

Following our proof by induction, we investigate the stability of the extrema f_k. The approach is based on a straightforward analysis of the sign definiteness of the Hessian matrix of Φ_k in terms of $\mathbf{g}_k$. Let us point out that according to the triangular form of the criteria, the only two situations that must be addressed correspond to vector $(f_j)_{1\leq j\leq k}$ belonging either to $\mathcal{F}_k \cup (\mathcal{F}_j^\star)_{j<k}$ or to $\overline{\mathcal{F}}_k \cup (\mathcal{F}_j^\star)_{j<k}$.

A. Stability Conditions

The stability conditions are derived in Appendix C. Proposition 6 sums up the main result.

Proposition 4: The extrema of criterion Φ_k corresponding to minima must verify

$$\mathcal{T}(\mathbf{H})\Psi_k((f_s)_{s\leq k})\mathcal{T}(\mathbf{H})^\top \geq 0 \tag{21}$$

with $\Psi_k((f_s)_{s\leq k}) \stackrel{\text{def}}{=} 2\Psi_c(f_k) + \beta_k\Psi_\epsilon((f_s)_{s<k})$, where $\Psi_c(.)$ and $\Psi_\epsilon(.)$ are symmetric matrices of dimension $K\times K$ defined, respectively, by

$$\Psi_c(f_k) = \mathcal{P}_s\Delta_s(f_k) + 6\mathcal{P}_s f_k f_k^\top\mathcal{P}_s + 2\mathcal{P}_s\mathcal{K}_s\mathrm{diag}(f_k f_k^\top) \tag{22}$$

$$\Psi_\epsilon((f_s)_{s<k}) = \sum_{m=-T}^{+T}\sum_{j<k} J_s^{(m)} f_j f_j^\top J_s^{(m)\top} \tag{23}$$

where $J_s^{(m)} = I_P \otimes (\sigma_1^2 J_{m,1}, \ldots, \sigma_P^2 J_{m,P})$ is a block-diagonal matrix of dimension $K\times K$, where $J_{m,k}$ is a Jordan matrix of dimension $(N+Q_k)\times(N+Q_k)$ ($(J_{m,k})_{a,b} = 1$ if $a-b=m$ and 0 elsewhere).

B. Stability of $(\mathbf{g}_j^\star)_{j\leq k}$

Let us consider first the case where all extrema f_j belong to $(\mathcal{F}_j^\star)_{j=1:k}$. In this case, it is straightforward to verify that Ψ_k is a diagonal matrix. Indeed, $\Psi_c(f_k^\star) = \mathcal{P}_s(3\|f\|_s^2 - r)I_K + 3\mathcal{P}_s^2\mathcal{K}_s f_k^\star f_k^{\star\top}$. Then, the sub-matrix $(\Psi_c(f_k^\star))_{(l),(l)}$ of dimension $(N+Q_l)\times(N+Q_l)$ is written as $(\Psi_c(f_k^\star))_{(l),(l)} = \sigma_l^2(3\|f_k\|_s^2 - r)I_{N+Q_l} + 3\sigma_l^4(\rho_l-1)f_{(l),k}^\star f_{(l),k}^{\star\top}$. Let us consider the separating solution of the form $f_{(m),k}^\star = \pm\sqrt{r/r_l}\,\delta_{(m)}^p$ for $m = l$ with $p \in I_l$ and 0_{N+Q_m} for $m \neq l$. Then, we obtain

$$(\Psi_c(f_k^\star))_{(l),(l)}^{i,j} = \begin{cases} \dfrac{r}{r_l}\sigma_l^4(3-\rho_l), & \text{if } i = j \neq p \\ 2\dfrac{r}{r_l}\sigma_l^4\rho_l, & \text{if } i = j \neq p \\ 0, & \text{if } i \neq j \end{cases} \tag{24}$$

and for $m \neq l$, we have

$$(\Psi_c(f_k^\star))^{i,j}_{(m),(m)} = \begin{cases} \dfrac{r}{r_l}\,\sigma_l^2\sigma_m^2(3-\rho_l), & \text{if } i=j \\ 0, & \text{if } i \neq j. \end{cases} \quad (25)$$

According to assumption A-4), $\rho_l < 3$, Ψ_c is a diagonal matrix with positive terms. Moreover, when $(f_s = f_s^\star)_{s<k}$, $\Psi_\epsilon((f_s^\star)_{s<k})$ is also a diagonal strictly positive matrix with $(\Psi_\epsilon((f_s^\star)_{s<k}))_{(m),(m)} = 0_{N+Q_m}$ for $m \in \overline{I}_{k-1}$ and $(r/r_m)\sigma_m^4 I_{N+Q_m}$ for $m \in I_{k-1}$. Thus, $\Psi_k((f_j^\star)_{s\leq k})$ is a diagonal positive matrix, and (21) is therefore always verified; the corresponding points $f_k^\star \in \mathcal{F}_k^\star$ are minima.

In the second case, we consider the analysis of the stability of the extrema setting $\overline{f}_k \in \overline{\mathcal{F}}_k$ with $(f_j^\star)_{j<k} \in (\mathcal{F}_j^\star)_{j<k}$. In this case, the analysis of the sign definition of Ψ_c is not straightforward, except in the trivial case where $\overline{f}_k = \mathbf{0}_K$, which corresponds to a maximum since $(\Psi_k(0, (f_j^\star)_{j<k}))_{(m),(m)} = -r\sigma_m^2$ for $m \in \overline{I}_{k-1}$ and $(\Psi_k(\mathbf{0}, (f_j^\star)_{j<k}))_{(m),(m)} = r\sigma_m^2(\beta_k 2\rho_m - 1) \leq 0$ [according to A-8)] for $m \in I_{k-1}$.

C. Instability of $\overline{\mathbf{g}}_k$ Given $(\mathbf{g}_j^\star)_{j<k}$

For the other extrema $\overline{f}_k \in \overline{\mathcal{F}}_k$ of subvector $\overline{f}_{(m),k}$ of the form $\overline{f}_{(m),k} = \sum_{p\in I_m} \pm\omega_m \delta^p_{(m)}$, the Hessian $\Psi_c(\overline{f}_k)$ is a sparse symmetric matrix. The diagonal contribution of Ψ_c is $\Psi_c(f_k))^{i,i}_{(m),(m)} = \sigma_m^2(3\|\overline{f}_k\|^2 - r)$ if $i \in \overline{I}_m$ and $\Psi_c(f_k))^{i,i}_{(m),(m)} = \sigma_m^2(3\|\overline{f}_k\|^2 - r) + 3\sigma_m^4(\rho_m - 1)\omega_m^2$ if $i \in I_m$, which leads to the global contribution

$$(\Psi_k(\overline{f}_k))^{i,i}_{(m),(m)} = \begin{cases} \sigma_m^4(3-\rho_m)\omega_m^2, & \text{if } i \in \overline{I}_m \\ 2\sigma_m^4\rho_m\omega_m^2, & \text{if } i \in I_m. \end{cases} \quad (26)$$

For the nondiagonal contribution, we get

$$\begin{aligned}(\Psi_k(\overline{f}_k))^{i,j}_{(m),(m)} &= 6\sigma_m^4 \overline{f}^i_{(m),k}\overline{f}^{j\top}_{(m),k} \\ &= 6\sigma_m^4 \begin{cases} \pm\omega_m^2, & \text{if } (i,j) \in I_m^2 \\ 0, & \text{if } i \in \overline{I}_m, \text{ or } j \in \overline{I}_m \end{cases}\end{aligned} \quad (27)$$

where $\omega_m = \overline{\omega}_m$ if $m \in \overline{I}_{k-1}$, and $\omega_m = \underline{\omega}_m$ if $m \in I_{k-1}$. For the subvector of $\overline{f}_k$ of the form $\overline{f}_{(n),k} = 0_{N+Q_n}$, we have $(\Psi_k(f_k))^{i,i}_{(n),(n)} = \sigma_n^2(3\|f_k\|^2 - r) \geq 0$ for $m \in \overline{I}_{k-1}$, and $\sigma_n^2(3\|f_k\|^2 - r) + (\beta_k/2)\,(r/r_n)\,\sigma_n^4 \geq 0$ for $m \in I_{k-1}$.

Let us consider now the contribution of $(\Psi_k)_{(m),(n)}$ for $m \neq n$. Since Ψ_ϵ is a diagonal matrix, the nonzero terms are given by $(\Psi_c)_{(m),(n)}$. For $\overline{f}_{(m),k} = 0_{N+Q_m}$ and/or $\overline{f}_{(n),k} = 0_{N+Q_n}$, we have $(\Psi(f_k))^{i,j}_{(m),(n)} = 0_{(N+Q_m)\times(N+Q_n)}$. For $\overline{f}_{(m),k} = \sum_{p\in I_m} \pm\omega_m\delta^p_{(m)}$ and $\overline{f}_{(n),k} = \sum_{q\in I_n} \pm\omega_n\delta^q_{|n|}$, we get

$$(\Psi_k(\overline{f}_k))^{i,j}_{(m),(n)} = 6\begin{cases} \pm\sigma_m^2\sigma_n^2\omega_m\omega_n, & \text{if } i \in I_m,\, j \in I_n \\ 0, & \text{if } i \in \overline{I}_m \text{ or } j \in \overline{I}_n \end{cases} \quad (28)$$

where $\omega_k = \overline{\omega}_k$ or $\underline{\omega}_k$, respectively, if $k \in \overline{I}_{k-1}$ and $k \in I_{k-1}$ (where $k = m, n$).

Due to this particular form of Ψ_k, we can derive an analysis of the sign of the associated quadratic form $x^\top \Psi_k x$. The result is given below.

D. Stability of $(\Phi_k)_{1\leq j\leq k}$ Extrema

Lemma 2: If $\overline{f}_k \in \overline{\mathcal{F}}_k$, for any vector x of dimension K, we have the following decomposition:

$$x^\top \Psi_k(\overline{f}_k, (f_s^\star)_{s<k})x = \sum_k \lambda_k x_k^2 + \sum_{i,j} x_{ij}^\top B_{ij} x_{ij} \quad (29)$$

where the λ_k (with $k \neq i, j$) denotes positive terms and B_{ij} a matrix of dimension 2×2 of negative determinant. The notation $x_{ij} \stackrel{\text{def}}{=} (x_i, x_j)$ corresponds to a vector of dimension 2 extracted from x.

Proof: See Appendix C.

We can easily verify that there is some vector $\overline{x} \in \mathbb{R}^K$ such that $\overline{x}^\top \Psi_k \overline{x} \stackrel{\text{def}}{=} \overline{x}^\top \Psi_k^A \overline{x} = \sum_k \lambda_k x_k^2 \geq 0$, where the lth entry of $\overline{x}$ is such that $(\overline{x})_l = 0$ if l is the subscript of a nonzero component of $\overline{f}_k$, and $(\overline{x})_l = x_l$ elsewhere. Introducing a complementary definition for $\underline{x} \in \mathbb{R}^K$, i.e., $(\underline{x})_l = x_l$ when $(\overline{x})_l = 0$ and $(\underline{x})_l = 0$ elsewhere, we get $\underline{x}^\top \Psi_k \underline{x} = \sum_{i,j} x_{ij}^\top B_{ij} x_{ij} \leq 0$, i.e., $\overline{f}_k$ is a saddle point. According to the result of Lemma 2, the extrema characterization of $(\Phi_k)_{1\leq k\leq M}$ can be easily derived. It is summarized in the next lemma.

Lemma 3: Under the assumptions A-1)–A-7), the extrema of $(\Phi_k)_{1\leq k\leq M}$ can be classified as

- $\underline{f}_k = 0_K \in \overline{\mathcal{F}}_k$ (maximum);
- $\overline{f}_k = \sum_{j\in I_{k-1}} \underline{\omega}_j u_j + \sum_{j\in \overline{I}_{k-1}} \overline{\omega}_j u_j \in \overline{\mathcal{F}}_k$ (saddle);
- $f_k^\star = \pm\sqrt{r/r_l}\,u_l \in \mathcal{F}_k^\star$ where $\mathcal{F}_k^\star \cap (\mathcal{F}_s^\star)_{s<k} = 0$ (global minima).

The minima of Φ_k are necessarily separating solutions, and each criterion selects a different source.

Lemma 3 proves, in particular for $k = 1$, the CM capture property, i.e., Φ_c minima are separating solutions. The separating solution $f_k^\star$ is given up to a scaling factor that depends on r and the constant dispersion of the selected source. In practical situations, we have to elaborate upon strategies to provide the user the best possible factor r, for instance, in terms of minimizing the remaining CM cost. If all the sources have similar known dispersion constant $r_i = r^*$, one should consider $r = r^*$. Any other constant will scale all the separating filters accordingly. When the sources have different dispersion constant, one can choose r so that all extracted sources have the closest possible mean energy: $r = \arg\min_r \sum_i (r/r_i - 1)^2$. This choice results in $r = \sum_i 1/r_i / \sum_i 1/r_i^2$, which requires the knowledge of all r_i. Indeed, when $r_i = r^*$ for all sources, the previous expression reduces to $r = r^*$.

VI. SIMULATIONS

We evaluate the performances of the proposed technique in two different settings. Simulations are performed in a scenario of two BPSK input sequences ($P = 2$) impinging on a three-sensor array ($L = 3$). In example I, we consider an "academic" random channel, where the columns are polynomials of degree 2. In simulation example II, the channel propagation model is simulated according to the Clarke model [5] in a context of wireless communications. In these examples, we are interested in the restoration of the two-input signals ($M = 2$). We quantify

TABLE I
ZEROS LOCATIONS OF CHANNEL #1

$h_{11}(z)$	$-22.997+i\,0.000$	$-0.401+i\,0.000$
$h_{21}(z)$	$-0.060-i\,0.615$	$-0.060+i\,0.615$
$h_{31}(z)$	$-0.261-i\,1.149$	$-0.261+i\,1.149$
$h_{12}(z)$	$-0.314-i\,0.875$	$-0.314+i\,0.875$
$h_{22}(z)$	$-0.601-i\,0.894$	$-0.601+i\,0.894$
$h_{32}(z)$	$-0.114-i\,0.442$	$-0.114+i\,0.442$

TABLE II
DETERMINANT ROOTS OF 2×2 MINORS

$\|H_{1,2}(z)\|$	$\|H_{1,3}(z)\|$	$\|H_{2,3}(z)\|$
$1.858+i\,0.000$	$-0.70+i\,1.283$	$0.259+i\,0.000$
$-0.487+i\,0.870$	$-0.7-i\,1.283$	$-1.082+i\,0.000$
$-0.487-i\,0.870$	$-0.379+i\,1.283$	$-0.262+i\,0.860$
$-0.166+i\,0.000$	$-0.379+i\,1.283$	$-0.262-i\,0.861$

performance of the input signal restoration through the global inter-symbol interference (GISI) index defined as

$$\mathrm{GISI}(f_k) = \frac{\sum_{j=1}^{P} \|f_{(j),\,k}\|^2}{\max f_k} - 1.$$

This measure takes into account both the ISI of the signal of interest and the IUI induced by the other sources.

Example I: In the first example, we consider the 3×2 channel $\mathbf{H}(z) = (\mathbf{h}_1(z), \mathbf{h}_2(z))$ for which the zero locations of functions $h_{ij}(z)$ with $1 \le i \le 3$ and $1 \le j \le 2$ are displayed in Table I. We check that $\mathbf{H}(z)$ satisfies A-3) since the maximal degree terms $\mathbf{h}_1(2) = (0.3686,\, 0.2895,\, 0.7802)^\top$ and $\mathbf{h}_2(2) = (0.6541,\, 0.9651,\, 0.1635)^\top$ are linearly independent, and A-2) Rank$(\mathbf{h}_1(z), \mathbf{h}_2(z)) = 2$ for all $z \in \mathbb{C}$ because all 2×2 matrix determinants do not share common roots (see Table II). None of these determinants is close to 0, showing, therefore, that the channel is not too difficult to equalize. All simulation parameters are summarized in Table III.

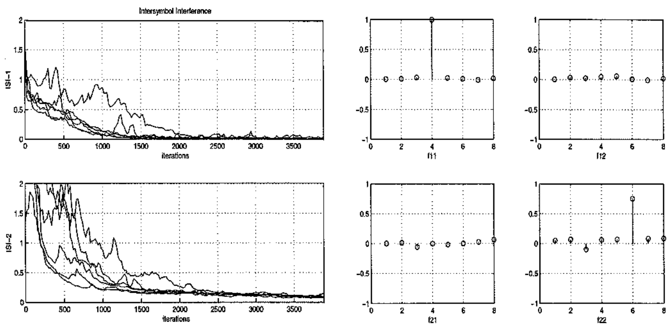
In Fig. 1, we plot the GISI, for each signal of interest, with respect to the number of iterations. Six different realizations with the same initialization (see Table III) are plotted in Fig. 1. At mean convergence of $\mathbf{g}_1^{(n)}$ and $\mathbf{g}_2^{(n)}$, we display the associated global impulse response $f_1^\star = (f_{(1),\,1}^{\star\top}, f_{(2),\,1}^{\star\top})$ and $f_2^\star = (f_{(1),\,2}^{\star\top}, f_{(2),\,2}^{\star\top})$ averaged over 500 iterations [Fig. 1(b)]. In this example, $\mathbf{g}_1^{\star\top}\mathbf{Y}_N(n)$ and $\mathbf{g}_2^{\star\top}\mathbf{Y}_N(n)$ achieve the restoration of $+s_1(n-4)$ and $+s_2(n-6)$, respectively.

Example II: In the second setting, the channel is simulated according to the model of Clarke, where for each sensor k and one source, the multipath time-continuous channel is obtained according to the relation,

$$h^{(k)}(t) = \sum_{p=1}^{nb_{paths}} A\delta(t-\tau_p) \sum_{p=1}^{nb_{rays}} \frac{1}{\sqrt{nb_{rays}}} e^{j(\phi_{n,\,p} + \Delta_{n,\,p}^k)} \tag{30}$$

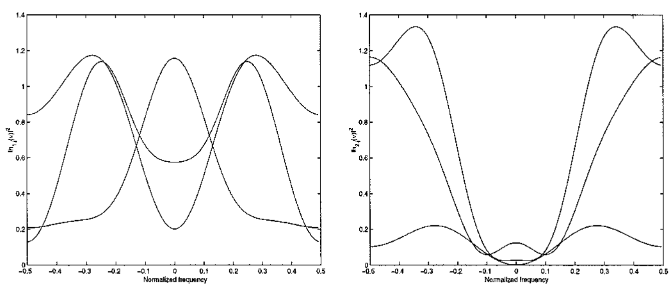
TABLE III
SIMULATION PARAMETERS

n_{ech}	N	P	L	T	μ	λ
4000	2	2	3	4	0.01	0.02
β_1	Q_1	Q_2	σ_w^2	$\mathbf{g}_1^{(0)}$	$\mathbf{g}_2^{(0)}$	
2	2	2	0.001	δ_3	δ_{11}	



(a) (b)

Fig. 1. (a) GISI versus number of iterations for different realizations, (b) overall impulse responses $f_1^\star$ and $f_2^\star$ at convergence of $g_1^{(n)}$ and $g_2^{(n)}$.



(a) (b)

Fig. 2. Frequency responses of $\mathbf{h}_i(z)$, for (a) $i=1$ and (b) $i=2$.

where path p has equal amplitude A and delay τ_p. The number of rays impinging on the sensor is $nb_{rays} = 20$ within each path. The sensors are uniformly distributed over a circular array. $\phi_{n,\,p}$ is an i.i.d. uniform process in $[0,\, 2\pi]$. Assuming planar wavefronts, $\Delta_{n,\,p}^i = 2\pi d/\lambda \cos(\theta_{n,\,p} - (i-1)2\pi/L)$ is the propagation delay of a ray with random angular incidence $\theta_{n,\,p}$ from the array origin to sensor i. Here, d denotes the array radius and λ the wavelength ($\simeq$33 cm) at 900 MHz. The symbols are of duration $3.7\ \mu$s. Square root raised cosine filters with a roll off of 0.5 are used for pulse shaping and reception filters. The overall channels are sampled at the baud rate and normalized to have unit gain; see Fig. 2 for the frequency responses of the columns of $\mathbf{H}(z)$. The other simulation parameters are described in Table IV.

We verify in Fig. 3, for three different realizations and initialization $\mathbf{g}_1^{(0)} = \delta_3$ and $\mathbf{g}_2^{(0)} = \delta_{14}$, that the behavior of the algorithm is similar to the behavior of the academic example I. This result is to be compared with Fig. 4 for which we use, for $\mathbf{g}_2^{(0)}$, the "bad initialization" [in the sense that it corresponds to the same basin of attraction as the initialization of $\mathbf{g}_1^{(0)}$] δ_4. In the first case, the convergence of f_1 and f_2 to separating solutions is slow but correct. In the second case, the constrain effect

TABLE IV
SIMULATION PARAMETERS

	T	μ	λ	N_{ech}	β_1
	8	0.01	0.2	8000	2
P	L	$Q_{1,2}$	σ_w^2	nb_{rays}	Roll-off
2	3	5	0.001	20	0.5

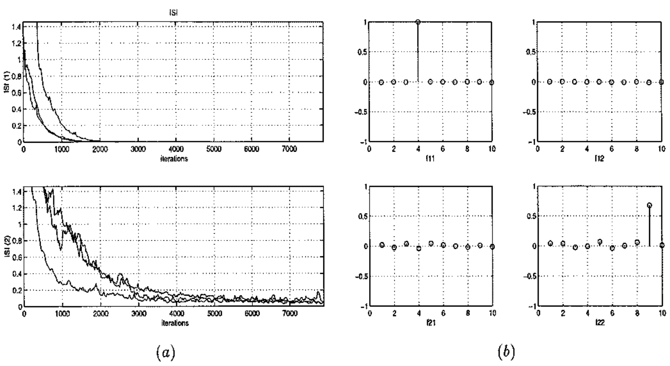


Fig. 3. (a) GISI for an initialization in two different basins of attraction $g_1^{(0)} = \delta_3$ and $g_2^{(0)} = \delta_{14}$ and (b) overall impulse response $f_1^\star$ and $f_2^\star$ at convergence of $g_1^{(n)}$ and $g_2^{(n)}$. The two different sources are selected.

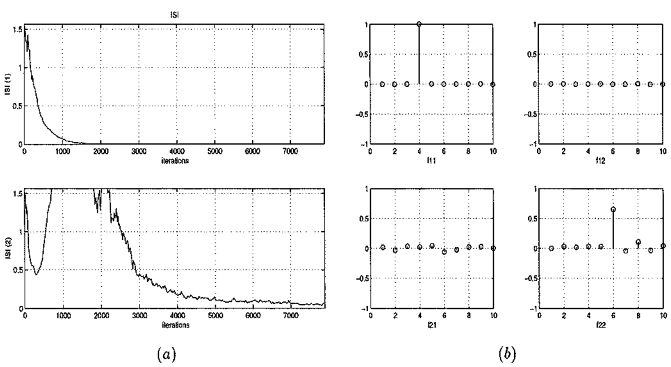


Fig. 4. (a) GISI for an initialization in the same basin of attraction $g_1^{(0)} = \delta_3$ and $g_2^{(0)} = \delta_4$ and (b) overall impulse response $f_1^\star$ and $f_2^\star$ at convergence of $g_1^{(n)}$ and $g_2^{(n)}$. The decorrelation constraint pushes g_2 to an other basin of attraction, selecting the second source.

pushes f_2 to another basin of attraction of the global cost function corresponding to the extraction of the source $s_2(n-6)$.

In Figs. 5 and 6, we investigate the effect of the parameter β_1. For the same initialization and the same input sequence s_1 and s_2, we consider two different penalty constraints. When β_1 is large enough, $\beta_1 \geq 2$, and the decorrelation constraint pushes the restoration filter $\mathbf{g}_2$ to a basin of attraction of the CM cost-function, which leads to the restoration of the second source (see Fig. 6). As predicted by our analysis, this is not the case when β_1 is too small, as in Fig. 5, where $\beta_1 = 0.2$).

VII. CONCLUSION

In this paper, we have proposed a new globally convergent approach for the multiple-input/multiple-output adaptive blind deconvolution problem when the number of outputs is strictly greater than the number of inputs. The originality of our work is to combine the CM criterion and triangular decorrelation constraints into a hierarchical set of criteria motivated by entropy maximization. Thanks to this new composite criteria design, we prove that each filter restores perfectly an arbitrary source different from the others, provided the sources are i.i.d., sub-Gaussian, and independent from each other. In particular, there are no local minima corresponding to a spatio-temporal mixture of the input signals or to selecting the same source several times. This result implies that the proposed simple adaptive gradient descent algorithm is guaranteed to asymptotically converge in the mean to the desired settings. Moreover, the use of the constant modulus criterion implies some robustness to noise and to the system invertibility conditions (see [9] in the single source case). Further work should imply the study of the robustness of the proposed algorithm.

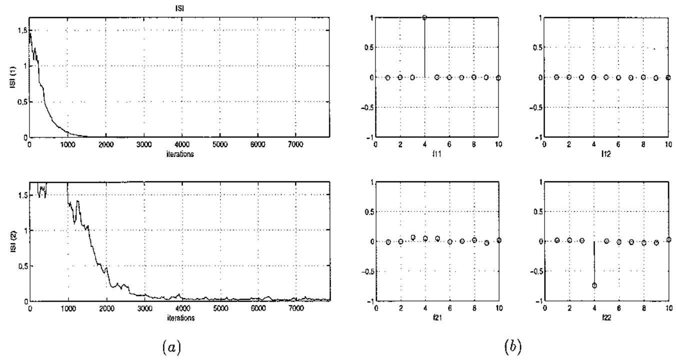


Fig. 5. (a) GISI for $g_1^{(0)}$ and $g_2^{(0)}$ initialized in the same basin of attraction and $\beta_1 = 0.2$ and (b) f_1 and f_2 overall impulse response after the convergence of g_1 and g_2. Because β_1 is too small, the second updating equation cannot select the other source.

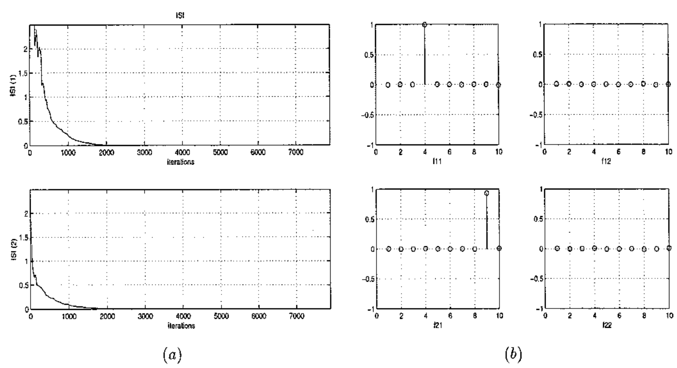


Fig. 6. (a) GISI for $g_1^{(0)}$ and $g_2^{(0)}$ initialized in the same basin of attraction and for $\beta_1 = 2$ and (b) overall impulse response estimated after convergence of g_1 and g_2. The two sources are restored.

APPENDIX A

A. Proof of Proposition 1

Expressed in terms of the global impulse response f, the gradient of the CM cost-function is given by

$$\nabla_g \Phi_c(\mathbf{g}) = 4\mathcal{T}(\mathbf{H})\mathbb{E}\{(\mathbf{S}(n)^\top f)(((f^\top \mathbf{S}(n))^2 - r)\mathbf{S}(n)\}. \tag{31}$$

A straightforward calculation of the lth subvector of the expectation contribution above is

$$\mathbb{E}\left\{\left(f_l^\top \mathbf{S}_l+\sum_{k\neq l}\mathbf{S}_k^\top f_k\right)\cdot\left(\sum_{k\neq l}(\mathbf{S}_k^\top f_k)^2+2\sum_{i\neq j}(\mathbf{S}_i^\top f_i)(\mathbf{S}_j^\top f_j)-r\right)\mathbf{S}_l\right\} \quad (32)$$

of dimension $N+Q_l$, which leads to the expression

$$\overset{(a)}{\mathbb{E}\{(f_l^\top \mathbf{S}_l)((f_l^\top \mathbf{S}_k)^2-r)\mathbf{S}_l\}} + \overset{(b)}{\mathbb{E}\left\{(f_l^\top \mathbf{S}_l)\sum_{k\neq l}(\mathbf{S}_k^\top f_k)^2\mathbf{S}_l\right\}} + 2\overset{(c)}{\mathbb{E}\left\{(f_l^\top \mathbf{S}_l)\sum_{i\neq j}(\mathbf{S}_i^\top f_i)(\mathbf{S}_j^\top f_j)\mathbf{S}_l\right\}} + \overset{(d)}{\mathbb{E}\left\{\sum_{k\neq l}(\mathbf{S}_k^\top f_k)\left(\sum_{k\neq l}(f_k^\top \mathbf{S}_k)^2-r\right)S_l\right\}} + 2\overset{(e)}{\mathbb{E}\left\{\sum_{k\neq l}(\mathbf{S}_k^\top f_k)\sum_{i\neq j}(\mathbf{S}_i^\top f_i)(\mathbf{S}_j^\top f_j)\mathbf{S}_l\right\}}.$$

Let us evaluate the different contribution of the previous expression under the i.i.d assumption and the statistical independence between the sources. Note that we may write (c) as

$$\mathbb{E}\left\{\sum_{j\neq l}(f_l^\top \underbrace{\mathbf{S}_l)^2(f_j^\top}_{0}\mathbf{S}_j)\mathbf{S}_l+\sum_{i\neq l}\sum_{i\neq j}(\mathbf{S}_i^\top \underbrace{f_i)(\mathbf{S}_j^\top f_j)}_{0}\mathbf{S}_l\right\} = \mathbf{0}_{N+Q_l}. \quad (33)$$

The expression (b) becomes $\mathbb{E}\{(f_l^\top \mathbf{S}_l)\sum_{k\neq l}(\mathbf{S}_k^\top f_k)^2\mathbf{S}_l\} = \sigma_l^2\sum_{k\neq l}\sigma_k^2\|f_k\|^2 f_l$. The contribution of (d) turns to

$$\mathbb{E}\left\{\sum_{k\neq l}(\mathbf{S}_k^\top f_k)\sum_{k\neq l}(f_k^\top \mathbf{S}_k)^2\mathbf{S}_l - r\sum_{k\neq l}(f_k^\top \underbrace{\mathbf{S}_k)\mathbf{S}_l}_{0}\right\}. \quad (34)$$

The development of the first term gives

$$\mathbb{E}\left\{\sum_{k\neq l}(f_k^\top \underbrace{\mathbf{S}_k)(\mathbf{S}_l^\top}_{0} f_l)\mathbf{S}_l\right\} + \mathbb{E}\left\{\sum_{k\neq l}(\mathbf{S}_k^\top f_k)\sum_{j\neq l}(f_j^\top \underbrace{\mathbf{S}_j)^2\mathbf{S}_l}_{0}\right\}. \quad (35)$$

Thus, (d) has a zero contribution. Let us consider now the last term (e). We obtain

$$\mathbb{E}\left\{(f_l^\top \mathbf{S}_l)\left(\sum_{k\neq l}(f_k^\top \mathbf{S}_k)^2-2\sum_{k\neq j}(f_k^\top \underbrace{\mathbf{S}_k)(f_j^\top}_{0}\mathbf{S}_j)\right)\mathbf{S}_l + \sum_{k\neq l}(\mathbf{S}_k^\top f_k)\sum_{i\neq j}(\mathbf{S}_i^\top \underbrace{f_i)(\mathbf{S}_j^\top}_{0} f_j)\mathbf{S}_l\right\}.$$

The contribution of (e) is equal to (b). For the contribution, we have $\mathbb{E}\{(f_l^\top \mathbf{S}_l)((f_l^\top \mathbf{S}_k)^2-r)\mathbf{S}_l\} = \sigma_l^4\Delta(f_l)f_l$, where $\Delta(f)=(3\|f\|^2-r/\sigma_l^2)I-(3-\rho_l)\mathrm{diag}(f_l f_l^\top)$. Note that this contribution corresponds to the gradient of the CM criterion for a single source and multiple outputs (see [9] for details). Finally, the contribution (32) becomes the expression

$$\sigma_l^4\Delta(f_l)f_l+3\sigma_l^2\sum_{k\neq l}\sigma_k^2\|f_k\|^2 f_l \qquad \text{for } k=1, P. \quad (36)$$

In order to write the global system, we introduce the diagonal matrix $\Delta_s(f)$ defined above. Then, it is easy to verify that the lth subvector $(\mathcal{P}_s\Delta_s(f)f)_{|l|}$ is equal to the expression (36). Thus, the gradient the CM in terms of the global impulse response $f=\mathcal{T}(\mathbf{H})^\top\mathbf{g}$ is written as

$$\nabla_g\Phi_c(\mathbf{g}) = 4\mathcal{T}(\mathbf{H})\mathcal{P}_s\Delta_s(f)f \quad (37)$$

where $\Delta_s(f)=(3\|f\|_s^2-r)I+\mathcal{K}_s\mathrm{diag}(ff^\top)$ is a diagonal matrix, where $\mathrm{diag}(A)$ denotes the matrix extracted from A with the same diagonal entries and 0 elsewhere. Herein, $\|f\|_s^2$ is understood as

$$\|f\|_s^2 \overset{\text{def}}{=} \sum_{j=1}^{P}\sigma_j^2\|f_{(j)}\|_2^2 \quad (38)$$

where $\|f_{(j)}\|_2^2=\sum_{p=1}^{N+Q_j} f_{(j)}(p)^2\ \mathcal{P}_s$, and $\mathcal{K}_s$ denotes two block-diagonal matrices of dimension $K\times K$, which are defined as $\mathcal{K}_s=I_P\otimes(\mathcal{K}_1,\ldots,\mathcal{K}_P)$ with $\mathcal{K}_l=(\rho_l-3)I_{N+Q_l}$. Likewise, we have $\mathcal{P}_s=I_P\otimes(\mathcal{P}_1,\ldots,\mathcal{P}_P)$ with $\mathcal{P}_l=\sigma_l^2 I_{N+Q_l}$. Note that the nonzero components of $f_{(l)}^c$ satisfy

$$f_{(l)}(p)^2 \overset{\text{def}}{=} \omega_{c,l}^2 = \frac{r-3\|f^c\|_s^2}{\sigma_l^2(\rho_l-3)} \quad (39)$$

by using the result of $f_{(l)}^c$ introduced in the previous expression. A straightforward calculation leads to $\omega_{c,l}^2=r/(\sigma_l^2(\rho_l-3)+3\sum_k \nu_k\sigma_l^2(\rho_l-3)/(\rho_k-3))$. Hence, $f_{(l)}^c$ is a vector of the form $f_{(l)}^c=(0,\ldots,\pm\omega_{c,l},\ldots,\pm\omega_{l,c},\ldots,0)^\top$, where $f_{(l)}^c(p)=\pm\omega_{c,l}$ if $p\in I_l$ and 0 elsewhere. A compact formulation is obtain by splitting $f_{(l)}^c$ on a basis of canonical vectors such that $f_{(l)}^c=\sum_{p\in I_l}\pm\omega_{c,l}\delta_{(l)}^p$. ■

Under the identifiability assumptions, the extrema f^c of Φ_c are solution of the equations $\mathcal{P}_s\Delta_s(f^c)f^c=\mathbf{0}_K$ (see the proof of Proposition 1), where we recall that $\mathcal{P}_s$ and Δ_s are diagonal matrices. The components of $f_{(l)}^c$, which are denoted $f_{(l)}^c(p)=$

$\omega_{c,l}$ for $p \in \{1, \ldots, N+Q_l\}$ and $p = 1, \ldots, P$, are solution of

$$\sigma_l^2 \left(3 \left\|f_{(l)}^c\right\|_s^2 - r\right) + \sigma_l^4(\rho_l - 3)\omega_{c,l}^2 = \mathbf{0}_K.$$

$\|f_{(l)}^c\|^2$ can be expressed as $\|f_{(l)}^c\|^2 = \nu_l \omega_{c,l}^2$, and we get $\|f^c\|_s^2 = \sum_k \sigma_k^2 \nu_k \omega_{c,k}^2$. By substituting the norm $\|f_{(l)}^c\|_s$ in the previous equation, a straightforward calculation leads to

$$\|f^c\|_s^2 = \sum_l \frac{r\nu_l \sigma_l^2}{3(\nu_l - 1)\rho_l}$$

with $\|f^c\|^2 \geq r/3$ since $\rho_k < 3$ A-4). If all components of f^c but one are equal to zeros, i.e., if $\nu_k = 1$ for an arbitrary $k = l$ and 0 for $k \neq l$, we have $\|f^c\|^2 = r/r_l$. If $\nu_k > 1$, we easily get $\|f^c\|_s^2 \leq \min_l r/r_l$. ■

APPENDIX B

A. Gradient of $\Phi_k(\mathbf{g}_k/(\mathbf{g}_k^\star)_{j<k})$

Equation (15) stems from a straightforward calculation of the Φ_k gradient. The CM contribution, which requires some lengthy calculation, is addressed in Appendix A. The gradient contribution of the constraint $\mathcal{E}_{i,j}(.)$ is given by $\nabla_{g_k}\mathcal{E}_{k,j}(m) = \nabla_{g_k}(\mathbf{g}_k^\top \mathbf{R}_Y(m)\mathbf{g}_j)^2 = 2(\mathbf{g}_k^\top \mathbf{R}_Y(m)\mathbf{g}_j)\ \mathbf{R}_Y(m)\mathbf{g}_j$, where $\mathbf{R}_Y(m) = \mathbb{E}\{\mathbf{Y}_N(n)\mathbf{Y}_N(n-m)^\top\} = \mathcal{T}(\mathbf{H})\mathbf{R}_S(m)\mathcal{T}(\mathbf{H})^\top$, and where $\mathbf{R}_S(m) \stackrel{\text{def}}{=} \mathbb{E}\{\mathbf{S}_N(n)\mathbf{S}_N(n-m)^\top\}$. One may check that according to the whiteness and independence sources assumptions, $(\mathbf{R}_S(m))_{(l),(l)} = \mathbb{E}\{\mathbf{S}_l(n)\mathbf{S}_l(n-m)^\top\} = \sigma_l^2 J_{m,n}$, where $J_{m,n}$ is a block-diagonal Jordan matrix of dimension $(N+Q_l) \times (N+Q_l)$, and $(.)_{(l),(l)}$ denotes the sub-block matrix of coordinates (l, l). It results that $\mathbf{R}_S(m) = J_s^{(m)}$. In terms of the global impulse response $f_k = \mathcal{T}(\mathbf{H})^\top \mathbf{g}_k$, we get $\nabla_{g_k}\mathcal{E}_{k,j}(m) = (f_k^\top J_s^{(m)} f_j)\mathcal{T}(\mathbf{H})J_s^{(m)} f_j$. Thus, the gradient of Φ_k, which is expressed in f_k space, is

$$2\mathcal{T}(\mathbf{H})\mathcal{P}_s\Delta_s(f_k)f_k + \beta_k\mathcal{T}(\mathbf{H})\sum_{j=1}^{k-1}\sum_{m=-T}^{+T}\left(f_k^\top J_s^{(m)} f_j\right) \cdot J_s^{(m)} f_j = \mathbf{0}_K \tag{40}$$

where $\Delta_s(f_k) = (3\|f_k\|_s^2 - r)I + \mathcal{K}_s\ \mathrm{diag}(f_k f_k^\top)$ is a diagonal matrix defined in Appendix A corresponding to the CM gradient. Note first that we may suppress the matrix $\mathcal{T}(\mathbf{H})$ from (40) since it is full column rank. Next, we focus on the expression of the constraint derivation for which the vectors f_j $(j < k)$ define the separating solution $f_j^\star$. If we denote I_{k-1} as the subset of the subscripts of all sources selected by the criteria $\Phi_{s<k}$, we get

$$\sum_{j=1}^{k-1}\sum_{m=-T}^{+T}\left(f_k^\top J_s^{(m)} f_j^\star\right) J_s^{(m)} f_j^\star = \sum_{l \in I_{k-1}}\sum_{m=-T}^{+T}\left(f_k^\top J_s^{(m)} u_l\right) J_s^{(m)} u_l. \tag{41}$$

$f_k^\top J_s^{(m)} u_l$ is the scalar and can be rewritten as $\sum_{n=1}^{P} \sigma_n^2 f_{(n),k}^\top J_{m,n} u_{(n),l}$, resulting in $\pm\sqrt{r/r_l}\sigma_l^2 f_{(l),k}^\top J_{m,l}\delta_{(l)}^p$, where $\delta_{(l)}^p$ is a canonical vector, and $p \in \Omega_l$, according to the notation introduced in Appendix A. Since we have $J_{m,n}\delta_{(l)}^p = \delta_{(l)}^{p-m}$ (41) becomes

$$\pm\sqrt{\frac{r}{r_l}}\sum_{m=-T}^{+T} f_{(l),k}^\top \delta_{(l)}^{p-m} J_s^{(m)} u_l = \frac{r}{r_l}\sum_{j=1}^{N+Q_l} f_{(l),k}(j)\sigma_l^2 \delta_{l+j}. \tag{42}$$

Finally, we get (16). ■

APPENDIX C

A. Proof of Proposition 3

For any extrema $f_k \in \mathcal{F}_k$ of criterion Φ_k, the Hessian results in the summation of the CM and constraint $\mathcal{E}_{i,j}(.)$ Hessian contribution. For $f_k = f$, we have $\partial/\partial \mathbf{g}(\mathcal{T}(\mathbf{H})\mathcal{P}_s\Delta_s(f)) = \mathcal{T}(\mathbf{H})\Psi_c(f)\mathcal{T}(\mathbf{H})^\top$, where $\Psi_c(f_k) = \partial(\mathcal{P}_s\Delta_s(f)f)/\partial f$ of component (i, j) of the form

$$(\Psi_c(f))_{i,j} = \frac{\partial(\mathcal{P}_s\Delta_s(f)f_i)}{\partial f_j} \tag{43}$$

where $(\mathcal{P}_s\Delta_s(f))_i = \sigma_i^2(3\|f\|^2 - r)f_i + \sigma_i^4(\rho_i - 3)f_i^3$.

For $i \neq j$, we have $(\Psi_c)_{i,j} = 3\sigma_i^2\partial/\partial f_j(\sum_{k,p}\sigma_k^2 f_{k,p}^2 f_i)$, i.e., $(\Psi_c)_{i,j} = 6\sigma_i^2\sigma_j^2 f_i f_j$. We may notice that when (i, j) belong to the same block (l), we have $(\Psi_c)_{i,j} = 6\sigma_l^4 f_i f_j$. For $i = j$, we have $(\Psi_c)_{i,i} = -\sigma_i^2 + 3(\rho_i - 3)\sigma_i^4 + 3\sigma_i^2\partial/\partial f_i(\|f\|_s^2 f_i)$. Since $\partial/\partial f_i(\|f\|_s^2 f_i) = \|f\|_s^2 + 2\sigma_i^2 f_i^2$, we obtain $(\Psi_c)_{i,i} = \sigma_i^2(3\|f\|_s^2 - r) + 3\sigma_i^4(\rho_i - 3)f_i^2 + 6\sigma_i^4 f_i^2$. By introducing $\Delta_s(f)$ in the previous expression, $\Psi_c(f)$ becomes the compact form (22). In the same way, the contribution of $\Psi_\epsilon((f_s)_{s<k})$ is given by

$$\sum_{j=1}^{k-1}\sum_{m=-T}^{+T}\mathcal{T}(\mathbf{H})\frac{\partial(f^\top J_*^{(m)} f_j)J_*^{(m)} f_j}{\partial f}\mathcal{T}(\mathbf{H})^\top \tag{44}$$

where $\partial/\partial f((f^\top J_*^{(m)} f_j)J_*^{(m)} f_j) = J_*^{(m)} f_j(\partial(f^\top J_*^{(m)} f_j)/\partial f) = f_j^\top J_*^{(m)\top}$. Finally, f is a minimum if the contribution of the Hessian matrix of the CM criterion plus the contribution of the Hessian of the constraint criterion are positive at the extrema point $f_k = f$. ■

B. Proof of Lemma 3

For $\overline{f}_k \in \overline{\mathcal{F}}_k$ extrema of Φ_k and a given $(f_s^\star)_{s<k}$, it is easy to show that Ψ_k can be decomposed as $\Psi_k = \Psi_k^A + \sum_{p=1}^{\nu(\nu-1)/2}\Psi_{k,p}^B$, where Ψ_k^A is a diagonal matrix of positive terms, and $\Psi_{k,p}^B$ is a sparse matrix for which the only nonzero entries form a 2×2 symmetric matrix $B_{m,n}$; it is understood that the subscripts m, n are referred to the block coordinates $(m), (n)$ of matrix Ψ_k. We see that $B_{m,n}^{i,i} = 2\sigma_m^4\omega_m^2(\rho_m/\nu - 1)$ and that $B_{m,n}^{j,j} = 2\sigma_n^4\omega_n^2(\rho_n/\nu - 1)$ for the diagonal entries if $\nu > 1$ with $i \in I_m$ and $j \in I_n$, where $\nu \stackrel{\text{def}}{=} \sum_{k=1,P}\nu_k$ is the number of the nonzero components in f_k and $B_{m,n}^{i,j} = \pm 6\sigma_m^2\sigma_n^2\omega_m\omega_n$ for the nondiagonal entries. Note that ω_l, for

$l = m, n$, is equal to $\overline{\omega}_l$ if $l \in \overline{I}_{k-1}$ and $\underline{\omega}_l$ if $l \in I_{k-1}$. The sign of the quadratic form $x^\top \Psi_k^A x + \sum_{i,j} x_{ij}^\top B_{ij} x_{ij}$ depends of the sign of the determinant of $B_{m,n}$, which is equal to $|B_{mn}| = 4\sigma_m^4 \sigma_n^4 \omega_m^2 \omega_n^2 (\rho_m \rho_n/(\nu-1)^2 - 9)$. $|B_{m,n}|$ is negative if $-3 \le \rho_k/(\nu-1) \le 3$ for $k = 1, P$. A sufficient condition is given by $|\rho_k| \le 3$. ■

ACKNOWLEDGMENT

The authors would like to thank reviewer #2 for help in improving the paper. A. Touzni would also like to thank P. Comon, E. Moreau, and A. Belouchrani for fruitful discussions and comments.

REFERENCES

[1] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problem," *IEEE Trans. Inform. Theory*, vol. 43, pp. 499–511, Mar. 1997.
[2] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1004–1034, 1995.
[3] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. New York: Springer-Verlag, 1990.
[4] L. Castedo, C. J. Escudero, and A. Dapena, "A blind signal separation method for multiuser communications," *IEEE Trans. Signal Processing*, vol. 45, pp. 1343–1348, May 1997.
[5] H. Clarke, "A statistical theory of mobile reception," *Bell Syst. Tech. J.*, vol. 47, pp. 987–1000, 1968.
[6] P. Comon, "Independent component analysis, A new concept?," *Signal Process.*, vol. 24, pp. 1–10, July 1991.
[7] N. Delfosse and P. Loubaton, "Adaptive separation of independent sources: A deflation approach," in *Proc. ICASSP*, vol. 4, 1994, pp. 41–44.
[8] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*, D. Findley, Ed. New York: Academic, 1981, pp. 556–608.
[9] I. Fijalkow, A. Touzni, and J. R. Treichler, "Fractionally-spaced equalization using CMA: Robustness to channel noise and lack of disparity," *IEEE Trans. Signal Processing*, vol. 45, pp. 56–66, Jan. 1997.
[10] W. Gardner, "A new method of channel identification," *IEEE Trans. Commun.*, vol. 39, pp. 813–817, Aug. 1991.
[11] D. Godard, "Self-recovering equalization and carrier tracking in two dimensional data communication systems," *IEEE Trans. Commun.*, vol. COMM-28, pp. 1867–1875, 1980.
[12] A. Gorokhov, P. Loubaton, and E. Moulines, "Second order blind equalization in multiple inputs multiple outputs FIR systems; a weighted squares approach," in *Proc. ICASSP*, 1996.
[13] Y. Inouye and T. Habe, "Multichannel blind equalization using second and fourth-order cumulants," in *Proc. HOS Workshop*, 1995.
[14] E. T. Jaynes, *Papers on Probability, Statistics and Statistical Phyisics*, R. D. Rosenkrantz and Reidel, Eds. Amsterdam, The Netherlands: Kluwer, 1982.
[15] C. Jutten and J. Hérault, "Blind separation of sources—Part I: An adaptive algorithm based on a neuromimetic architecture," *Signal Process.*, vol. 24, no. 1, pp. 1–10, 1991.
[16] E. Lee and D. Messerschmitt, *Digital Communication*. Boston, MA: Kluwer, 1988.
[17] A. Mansour, C. Jutten, and P. Loubaton, "Subspace method for blind separation of sources in convolutional mixture," in *Proc. EUSISPCO*, 1996.
[18] S. Mayrargue, "A blind spatio-temporal equalizer for a radio-mobile channel using the constant modulus algorithm (CMA)," in *Proc. ICASSP*, vol. 2, 1994, pp. 317–320.
[19] E. Moulines, P. Duhammel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, pp. 516–525, Feb. 1995.
[20] E. Moreau and O. Macchi, "High-order contrasts for self-adaptive source separation," *Int. J. Adapt. Contr. Signal Process.*, vol. 10, pp. 19–46, 1996.
[21] H. L. N. Thi and C. Jutten, "Blind sources separation for convolutives mixtures," *Signal Process.*, vol. 45, pp. 209–229, 1995.
[22] C. B. Papadias and A. Paulraj, "A constant Modulus algorithm for multi-user signal separation in presence of delay spread using antenna arrays," *IEEE Signal Processing Lett.*, vol. 4, pp. 178–181, June 1997.
[23] D. T. Pham, "Blind separation of instantaneous mixture of sources via independent component analysis," *IEEE Trans. Signal Processing*, vol. 44, pp. 2768–2779, Nov. 1996.
[24] A. Touzni and I. Fijalkow, "Blind adaptive equalization and simultaneous separation of convolutional mixtures," in *Proc. Int. DSP Conf.*, July 1997.
[25] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization of multipath channels," in *Proc. 25th Asilomar Conf.*, 1991, pp. 856–860.
[26] J. R. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP–31, pp. 459–472, 1983.
[27] J. R. Treichler and M. G. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP–33, pp. 420–431, Apr. 1985.
[28] A. Touzni, "Performances et robustesse en égalization avec diversit é spatio-temporelle," Ph.D. dissertation, Univ. Cergy-Pontoise, Cergy-Pontoise, France, Nov. 1998.
[29] J. K. Tugnait, "Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a Godard cost function," *IEEE Trans. Signal Processing*, vol. 45, pp. 268–271, Jan. 1997.
[30] S. Van Gerven and D. Van Campernolle, "Feedforward and feedback in a symmetric adaptive noise canceler: Stability analysis in a simplified case," *Signal Process. VI, Theories Applicat.*, pp. 1081–1084, Aug. 1992.
[31] D. Yellin and E. Weinstein, "Criteria for multichannel signal separation," *IEEE Trans. Signal Processing*, pp. 2158–2168, Aug. 1994.

Azzédine Touzni was born on July 12, 1969, in Paris, France. He received the Engineering degree in optic telecommunication from the Institut Galilé, Paris, in 1994 and the Master degree in signal processing from the University of Cergy-Pontoise, Cergy (UCP), Pointoise, France, in 1995. In 1998, he received the Ph.D degree from UCP.

In 1999, he was with the Institut National de Recherche en Informatique et Automatique (INRIA), Rocquencourt, France, as well as with Cornell University, Ithaca, NY, as a Research Associate with the School of Electrical Engineering. Since June 2000, he has been with Nxtwave Communications Inc., Langhorne, PA. His research interests focus on the area of adaptive signal processing applied to digital communications.

Inbar Fijalkow (M'96) was born in Haifa, Israel, on October 17, 1966. She received the Engineering and Ph.D. degrees from Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France, in 1990 and 1993, respectively.

From 1993 to 1994, she was a Research Associate at Cornell University, Ithaca, NY, in the School of Electrical Engineering. From 1994 to 1999, she was an Associate Professor at the Ecole Nationale Supérieure de l'Electronique et de ses Applications (ENSEA), Cergy-Pontoise, France, where she has been a Professor since 1999. In 1998, she spent four months at the Australian National University (ANU), Canberra, as a Visiting Researcher. Her current research interests are in statistical signal processing, particularly as applied to digital communications: adaptive and iterative (turbo) processing, blind deconvolution/equalization of multiple sources, and sensors systems. She is member a of the board of the GDR ISIS, which is the CNRS Research Group on signal and image processing. She is in charge of the working group on telecommunications and transmission.

Michael G. Larimore (M'77) was born in Cleveland Heights, OH, on January 7, 1949. He received the B.S. and M.S. degrees in electrical engineering from the University of Colorado, Boulder, in 1971 and 1972, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1977.

From 1977 to 1981, he was on the faculty at the University of Colorado, Colorado Springs. In 1981, he joined ArgoSystems, Inc., Sunnyvale, CA, where he worked in the area of adaptive techniques for interference mitigation. Since 1986, he has been with Applied Signal Technology, Inc., Sunnyvale, as Senior Scientist of the Multichannel Systems Division. His principal efforts have dealt with implementational aspects of adaptive techniques applied to demodulation and equalization of digital transmission. His research interests involve adaptive processing and its application to the transmission and storage of digital information. He has served as Lecturer at both Stanford University, Stanford, CA, and Santa Clara University, Santa Clara, CA.

Dr. Larimore and has served on the Digital Signal Processing Technical Committee of the Signal Processing Society.

John R. Treichler (F'91) was born in Velasco, TX, on September 22, 1947. He received the B.A. and M.S. degrees in electrical engineering from Rice University, Houston, TX, in 1970 and the Ph.D. degree from Stanford University, Stanford, CA, in 1977.

From 1970 to 1974, he was in the U.S. Navy. From 1977 to 1983, he was employed with ArgoSystems, Inc. (now a subsidiary of Boeing). He co-founded Applied Signal Technology, Inc., Sunnyvale, CA, in 1984 and is currently their Chief Technical Officer. He served as a Lecturer at Stanford from 1977 to 1983 and spent the 1983–1984 academic year at Cornell University, Ithaca, NY, as an Associate Professor with the School of Electrical Engineering. His current research interests are in the area of digital and adaptive signal processing, particularly as applied to communications problems.

Dr. Treichler was awarded the IEEE Signal Processing Society's Technical Achievement Award in 1999 and received the IEEE Third Millinnium Medal in 2000.