Source Extraction by Maximizing the Variance in the Conditional Distribution Tails

Ronald Phlypo, *Member, IEEE*, Vicente Zarzoso, *Member, IEEE*, and Ignace Lemahieu, *Senior Member, IEEE*

Abstract—This paper presents a method for signal extraction based on conditional second-order moments of the output of the extraction filter. The estimator of the filter is derived from an approximate maximum likelihood criterion conditioned on a presence indicator of the source of interest. The conditional moment is shown to be a contrast function under the conditions that 1) all cross-moments of the same order between the source signal of interest and the other source signals are null and 2) that the source of interest has the largest conditional moment among all sources. For the two-source two-observation case, this allows us to derive the theoretical recovery bounds of the contrast when the conditional cross-moment does not vanish. A comparison with empirical results confirms these bounds. Simulations show that the estimator is quite robust to additive Gaussian distributed noise. Also through simulations, we show that the error level induced by a rough approximation of the presence indicator shows a strong similarity with that of additive noise. The robustness, with respect both to noise and to inaccuracies in the prior information about the source presence, guarantees a wide applicability of the proposed method.

Index Terms—Conditional likelihood, contrast functions, estimation, source extraction.

I. INTRODUCTION

SIGNAL extraction methods focus on the estimation of a specific source when only a linear combination of source signals is available on a sensor array. A classical approach to the estimation of a specific source from the observations is based on a complete separation of the observations into its sources, followed by a posterior selection of the source of interest. The first step is solved for with the class of blind source separation (BSS) algorithms. Their aim is to recover source signals when only a mixture of them is observed on a sensor array. This implicitly involves the inversion of an estimate of the linear mixture. This inverse of the estimate applied to the observations then yields outputs that are estimates of the source signals.

In the past two decades, the topic of blind separation has received growing interest, specially since the introduction of the quite natural model of independent sources, which seems to be an appropriate model for communications and biomedical signal analysis, to give a few examples. BSS under the aforementioned model can be achieved by the tool of independent component analysis (ICA) [1], [2]. The algorithms for ICA are based on the optimization of a contrast function, imposing a measure for independence on the separator outputs. It has been shown that the optimization of any such measure based on the independence of the outputs is sufficient to solve the separation of the observations into the independent sources, up to the inherent ambiguities of scaling, source permutation (order), and phase [3]. Since these ambiguities are waveform preserving, they are generally admissible. However, the above divide-and-conquer strategy presents a significant computational overload for the estimation of a single component, especially when large datasets are considered.

Recently, specific contrast functions for source extraction have been proposed in the literature, e.g., in [4] and [5], and valuable objective functions are known to depend (implicitly) on the marginal negentropy of the extractor output. Theoretically, the extraction order of the sources can be fixed, based upon their stochastic properties [6], allowing for the more informative sources (higher negentropy) to be extracted first. However, when the source of interest is not the source with highest negentropy, the extraction has to be pursuit, estimating source by source until the source of interest has been found. To prevent having twice an estimate of the same source, the observation space is deflated by the current source space before a next source has to be estimated [7], [8]. Unfortunately, this iterative estimation-deflation scheme engenders a propagating error that accumulates over the iterations [7], [9]. Since the source of interest is not always the source with highest entropy, and since the extraction order of the sources cannot always be fixed in practice, the source of interest accumulates an error in its estimate whenever it does not appear in the first extraction.

It is clear from the above that we cannot resolve for the permutation ambiguity without adding some discriminating information about the source of interest—other than negentropy—into the source extraction objective function. However, the prior information used to discriminate our source of interest from the other sources should be kept to a minimum if we want to keep the source extraction maximally blind. This is the aim of the class of constrained ICA (cICA) algorithms proposed in [10] and [11]. cICA introduces a constraint on the solution space of the (approximated) negentropy objective function, by means of a penalizing term, generally based on a maximally admissible distance measure between the output and a reference signal. In contrast to the solution obtained by minimizing the squared error between the filter output and the reference signal (the basis to the Wiener filter [12]), the solution to cICA is the output that has maximal negentropy among the solutions meeting the constraint on the distance measure. A closely related algorithm is BSS with a reference (BSSR) [13], based

R. Phlypo and I. Lemahieu are with MEDISIP/IBBT, Heymans Institute, University of Ghent, 9000 Ghent, Belgium (e-mail: ronald.phlypo@ugent.be; ignace.lemahieu@ugent.be).

V. Zarzoso is with UNS/CNRS, Les Algorithmes-Euclide B, 06903 Sophia Antipolis Cedex, France (e-mail: zarzoso@i3s.unice.fr).

on the higher order dependencies between the output signal and a reference. The BSSR algorithm differs from the Wiener filter mainly in the distance measure used. Because higher order moments are considered, BSSR offers a better performance when the reference signal has relatively few nonzero values [13]. The quadratic higher order criteria (QHOC) [14], [15] can be considered as a generalization of the BSSR approach. QHOC have been proven to be contrasts for source separation, but they have not been derived in a maximum likelihood sense, so that they do not inherit the estimation optimality of maximum likelihood estimators in the sense of Fisher's information. Since the reference signal is chosen arbitrary, there is no explicit control over the extractor output. The best one could do if a specific source of interest has to be estimated is the use of the QHOC with an estimation-deflation scheme until the source of interest has been found. But this estimation-deflation scheme suffers from error propagation and accumulation as mentioned above.

The method of conditional moments (MCM) has also been used for source separation [16], where a possible link with the theory of contrast functions has been evoked. In this paper, we will show that the use of well-chosen conditional moments indeed results in a contrast function for source extraction. But contrary to the method of Xerri *et al.*, our method is not limited to symmetrical distributions, does handle the extraction of Gaussian sources, and does not require a posteriori manipulations, such as correcting rotations for super-Gaussian distributed sources. Note also that the method in [16] does not envisage the estimation of a specific source but focuses solely on the full separation problem. On the other hand, although the approach of conditional moments differs from that of reference-based filtering, we will show that for certain well-defined cases, the above algorithms (BSSR, QHOC, and Wiener filtering) can be related to the theory presented in this paper.

This paper begins with an introduction on the signal model and the notational conventions in Section II. We then provide the theoretical aspects of the framework and present our method in Section III. Section IV places the presented method in perspective with respect to some competing models and algorithms found in literature. We show that under certain conditions, some explicit or implicit relations exist between these methods and the proposed method. Because of their similarities, these algorithms and models will be used in comparison studies in Section V after the performance bounds and some properties/characteristics of our model have been examined. This will be followed by a discussion in Section VI and a summary in Section VII.

II. Signal Model and Notation

A. Notational Conventions

Scalar variables, column vectors, and matrices are, respectively, given by lowercase lightface (u), lowercase boldface $(\mathbf{u})$, and uppercase boldface $(\mathbf{U})$ characters. Consistency of the notations then requires the jth entry of $\mathbf{u}$ to be denoted by u_j and the jth column of $\mathbf{U}$ by $\mathbf{u}_j$. The probability density function (pdf) associated to the random variable a will be denoted by p_a for continuous sample domains and P_a for discrete sample domains. The association is denoted as $a \sim p_a$ and $p_a(u) \equiv p(a = u)$. Realizations of random variables or vectors are,

respectively, given as scalars or vectors with an (arbitrary) indexing to identify the samples, e.g., $\mathbf{u}[k]$ stands for a sample of $\mathbf{u}$, referenced by the index k. Also, let constants be given as uppercase lightface characters (U), the set of real numbers as $\mathbb{R}$, and sets by calligraphic uppercase characters, such as $\mathcal{U}$, whose cardinal number is $\#(\mathcal{U})$. A set of K realizations from the random vector $\mathbf{u}$ (a population) is then defined as $\mathcal{U} = \{\mathbf{u}[k] | \mathbf{u} \sim p_{\mathbf{u}}, k = 1, 2 \ldots K\}$ and will be referred to by the shorthand notation $\{\mathbf{u}\}_K$, although with some abuse of notation we will commonly drop K as well as the accolades.

Furthermore, the mathematical expectation of a function f with respect to $\mathbf{u}$ defined as $\int p_{\mathbf{u}}(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}$ will be denoted by $\mathbb{E}\{f(\mathbf{u})\}$. Finally, the transpose of a column vector $\mathbf{u}$ is written as $\mathbf{u}^T$.

B. Signal Model and Contrast Functions

In this paper, we assume the generative linear model where an M-dimensional random observation vector $\mathbf{y}$ can be linked to the underlying N-dimensional random source vector $\mathbf{s}$ through the instantaneous linear relation

$$\mathbf{y} = \mathbf{A}\mathbf{s} \qquad (1)$$

with $M \geq N$. Contrary to classical algorithms, we assume that s_j is independently distributed with respect to $\tilde{\mathbf{s}} = [s_1, s_2, \ldots, s_{j-1}, s_{j+1}, \ldots, s_N]^T$, while mutual dependencies may exist between the entries of $\tilde{\mathbf{s}}$. We further assume that all random variables are zero-mean, without loss of generality. We denote by $x = \mathbf{h}^T \mathbf{y}$ an output of the filter $\mathbf{h}$ acting on $\mathbf{y}$.

Our goal is to estimate s_j from the observations $\mathbf{y}$. An appropriate strategy is to use the notion of a contrast function for the extraction of s_j.

Definition 1: A function $\Psi(x)$ is a contrast function for the extraction of the source s_j from the observations under the model (1) if it fulfils the following three properties.

P1) *Scaling Invariance:*

$$\Psi(x) = \Psi(\lambda x), \quad \forall \lambda \in \mathbb{R} \setminus \{0\}.$$

P2) *Domination:*

$$\Psi(s_j) \geq \Psi(\mathbf{g}^T \mathbf{s}), \quad \forall \mathbf{g} \in \mathbb{R}^N.$$

P3) *Discrimination:*

$$\Psi(s_j) = \Psi(\mathbf{g}^T \mathbf{s}) \Leftrightarrow \mathbf{g} = \boldsymbol{\lambda}_j.$$

$\boldsymbol{\lambda}_j$ is the jth column of a nonsingular diagonal scaling matrix $\boldsymbol{\Lambda} \in \mathbb{R}^{N \times N}$.

It follows that, if $\Psi(x)$ is a contrast function for the extraction of s_j from the observations, we have that $\hat{\mathbf{h}} = \arg\max_{\mathbf{h}} \Psi(\mathbf{h}^T \mathbf{y})$ is an extraction filter and $x = \hat{\mathbf{h}}^T \mathbf{y} = \hat{s}_j$. We also observe that, by fixing the index j, no permutation ambiguity exists with the above definition of contrast functions for source extraction, contrary to the definition of contrast functions for source separation [2] and the previous definitions of contrasts functions for source extraction [4], [5], [15]. This follows from the fact that previous definitions of contrast functions for source extraction are based on a source by

source extraction scheme to solve the full separation problem and do not consecrate more importance to s_j than to any other source.

III. METHODS

A. A Likelihood— Based Contrast

If we start from the distribution of $\mathbf{s}$ and we suppose $M = N$, we obtain the expected likelihood of the observations $\mathbf{y}$ as

$$\mathcal{L}_{\text{BSS}}(\mathbf{A}|\mathbf{y}) = \int_{\mathbb{R}^N} p_{\mathbf{A}^{-1}\mathbf{y}}(\mathbf{u}) \log p_{\mathbf{s}}(\mathbf{u}) d\mathbf{u} - \log|\det(\mathbf{A})|$$

where $\mathbf{A}$ has been supposed nonsingular. This is the maximum likelihood approach to BSS proposed in [17], [18]. By factorizing $p_{\mathbf{s}|\vartheta_{\mathbf{s}}}$ as $p_{s_j} p_{\bar{\mathbf{s}}}$ and retaining only the terms depending on the jth column of $\mathbf{H} = (\mathbf{A}^{-1})^T$, which we denote by $\mathbf{h}$, we obtain

$$\mathcal{L}_{\text{E}}(\mathbf{h}|\mathbf{y}) \propto \int_{\mathbb{R}} p_{\mathbf{h}^T\mathbf{y}}(u) \log p_{s_j}(u) du \qquad (2)$$

where we have discarded the last term of $\mathcal{L}_{\text{BSS}}$, since $\log|\det(\mathbf{H})| = -\log|\det(\mathbf{A})|$ can be considered constant if we absorb any changes in $\det(\mathbf{H})$ due to $\mathbf{h}$ in the remaining columns of $\mathbf{H}$. If we want to consider the above likelihood function as a contrast function for the extraction of the source s_j, it should satisfy properties P1)–P3) of Definition 1, which is not straightforward. It is obvious from (2) that two sources s_i and s_j for which $p_{s_i} = p_{s_j}$ cannot be distinguished with the above likelihood function.

Our goal is to find some adaptation of the likelihood $\mathcal{L}_{\text{E}}(\mathbf{h}|\mathbf{y})$ such that it can be used as a contrast function for the extraction of s_j but without having p_{s_j} (neither a parametrization) at its disposition. In what follows, we illustrate first the basis of the conditional expected log-likelihood as a contrast function for source extraction under the general model (1). Since the maximization of the proposed estimator will be shown to be combinatorial in nature, we next derive a practical algorithm based on conditional second-order statistics.

B. The Conditional Likelihood

We assume that we have a presence indicator $\mathbb{1}_{s_j}$ for s_j with respect to a threshold C, which is defined as

$$\begin{cases} \mathbb{1}_{s_j}, & \text{if } |s_j/\sigma_{s_j}| \geq C \\ \bar{\mathbb{1}}_{s_j}, & \text{otherwise.} \end{cases} \qquad (3)$$

If we define furthermore $\mathbb{B} = \mathbb{R} \setminus (-C, C)$ and $\bar{\mathbb{B}} = (-C, C)$, then we may rewrite the log-likelihood of (2) conditioned on $\mathbb{1}_{s_j}$ as

$$\mathcal{L}_{\text{E}}\left(\mathbf{h}|\mathbf{y}, \mathbb{1}_{s_j}\right) \propto \int_{\mathbb{B}} p_{x|\mathbb{1}_{s_j}}(u) \log p_{s_j|\mathbb{1}_{s_j}}(u) du$$

$$+ \int_{\bar{\mathbb{B}}} p_{x|\mathbb{1}_{s_j}}(u) \log p_{s_j|\mathbb{1}_{s_j}}(u) du. \qquad (4)$$

With the definition $P(\mathbb{1}_{s_j}) = \int_{\mathbb{B}} p_{s_j}(u) du$, we have

$$p_{s_j|\mathbb{1}_{s_j}}(u) = \begin{cases} p_{s_j}(u)/P\left(\mathbb{1}_{s_j}\right), & \forall u \in \mathbb{B} \\ 0, & \forall u \in \bar{\mathbb{B}} \end{cases}.$$

Let us admit the commonly accepted definition $0 \log 0 \triangleq 0$. The conditional log-likelihood of (4) is heavily penalized by the term $\int_{\bar{\mathbb{B}}} p_{x|\mathbb{1}_{s_j}}(u) \log p_{s_j|\mathbb{1}_{s_j}}(u) du$, which is zero if and only if $\mathbf{h}$ is such that $p_{x|\mathbb{1}_{s_j}}(u) = 0, \forall u \in \bar{\mathbb{B}}$.

Proposition 1: $\mathcal{L}(x) = \mathcal{L}_{\text{E}}(\mathbf{h}|\mathbf{y}, \mathbb{1}_{s_j})$ is a contrast for the source model (1) and this for any $C > 0$.

For a proof, we refer the reader to the Appendix.

We can follow an analogous reasoning for the conditional log-likelihood $\mathcal{L}_{\text{C}}(\mathbf{h}|\mathbf{y}, \bar{\mathbb{1}}_{s_j})$, from which we see that our goal is to obtain $p_{x|\bar{\mathbb{1}}_{s_j}}(u) = 0, \forall u \in \mathbb{B}$.

Since $\mathcal{L}(x)$ is a contrast under the above condition, $\hat{\mathbf{h}}$ (and thus $\hat{s}_j$) can be found through its maximization. However, the heavy penalization term makes the solution not tractable, and the maximization of $\mathcal{L}(x)$ is equivalent to an exhaustive search over all possible filters $\mathbf{h}$. This is an NP-hard problem.

C. Relaxation of the Conditional Likelihood Function

We have seen that a numerical optimization of the conditional log-likelihood in (4) is NP-hard. In this paragraph, we try to relax the condition $p_{x|\mathbb{1}_{s_j}}(u) = 0, \forall u \in \bar{\mathbb{B}}$ such that a numerically tractable solution exists. Consider the following relaxation:

$$\min_{\mathbf{h}} \int_{\bar{\mathbb{B}}} p_{x|\mathbb{1}_{s_j}}(u)\gamma(u) du \qquad (5)$$

where $\gamma(u)$ is any positive function. It is straightforward that $p_{x|\mathbb{1}_{s_j}}(u) = 0, \forall u \in \bar{\mathbb{B}} \Rightarrow \int_{\bar{\mathbb{B}}} p_{x|\mathbb{1}_{s_j}}(u)\gamma(u) = 0$. The inverse holds equally true for all (general) distributions (proof in Appendix). We thus have that $\min_{\mathbf{h}} \int_{\bar{\mathbb{B}}} p_{x|\mathbb{1}_{s_j}}(u)\gamma(u) du = 0$, which is equivalent to the minimization of $p_{x|\mathbb{1}_{s_j}}$. However, the above minimization remains NP-hard due to the integration over the posterior $\bar{\mathbb{B}}$.

To make the minimization in (5) numerically tractable, we take $\gamma(u) = u^2$ as a possible weighting function. The minimization reads

$$\min_{\mathbf{h}} \int_{\bar{\mathbb{B}}} p_{x|\mathbb{1}_{s_j}}(u)u^2 du$$

and would then intuitively be equivalent to a maximization problem

$$\max_{\mathbf{h}} \frac{\int_{\mathbb{R}} p_{x|\mathbb{1}_{s_j}}(u)u^2 du}{\sigma_x^2} \qquad (6)$$

where we used $\sigma_x^2 = \int_{\mathbb{R}} p_x(u)u^2 du$ in the denominator to satisfy the scale invariance P1). The equivalence can be seen from the equality

$$\frac{\int_{\mathbb{R}} p_{x|\mathbb{1}_{s_j}}(u)u^2 du}{\sigma_x^2} = \frac{\int_{\mathbb{B}} p_{x|\mathbb{1}_{s_j}}(u)u^2 du + \int_{\bar{\mathbb{B}}} p_{x|\mathbb{1}_{s_j}}(u)u^2 du}{\sigma_x^2}$$

where $\int_{\mathbb{B}} p_{x|\mathbb{1}_{s_j}}(u)u^2 du$ dominates the nominator due to the chosen weighting function. Maximizing the former means that the probability mass of $p_{x|\mathbb{1}_{s_j}}(u)$ would shift towards large values of x ($|x| \gg C$, bounded by the variance normalization), which can be regarded as a minimization of $p_{x|\mathbb{1}_{s_j}}(u), \forall u \in \overline{B}$, as required.

D. A Contrast With an Algebraic Optimum

The maximization in (6) is thus equivalent to the maximization of $\Psi(x) = \mathbb{E}_{s_j}\{x^2\}/\mathbb{E}\{x^2\}$, with $\mathbb{E}_{s_j}\{x^2\} = \int_{\mathbb{R}} p_{x|\mathbb{1}_{s_j}}(u)u^2 du$. Introducing the shorthand notations $\Phi_{\mathbf{u}}^{s_j} = \mathbb{E}_{s_j}\{\mathbf{u}\mathbf{u}^T\}$ and $\Phi_{\mathbf{u}} = \mathbb{E}\{\mathbf{u}\mathbf{u}^T\}$, we write

$$\Psi(x) = \frac{\Phi_x^{s_j}}{\Phi_x} = \frac{\mathbf{h}^T \Phi_{\mathbf{y}}^{s_j} \mathbf{h}}{\mathbf{h}^T \Phi_{\mathbf{y}} \mathbf{h}}. \tag{7}$$

This is a generalized Rayleigh quotient, and its maximization has as an algebraic solution; see, e.g., [19, Sec. 8.7.1].

The maximization of (7) can be done through the eigenvalue decomposition of $\Phi_{\mathbf{y}}^{-1}\Phi_{\mathbf{y}}^{s_j}$ (whenever $\Phi_{\mathbf{y}}$ is invertible) and choosing the major eigenvector/eigenvalue pair $\mathbf{q}, \lambda$ for which

$$\Phi_{\mathbf{y}}^{-1}\Phi_{\mathbf{y}}^{s_j}\mathbf{q} = \mathbf{q}\lambda. \tag{8}$$

Taking $\hat{\mathbf{h}} = \mathbf{q}$, we obtain $x = \hat{s}_j = \hat{\mathbf{h}}^T \mathbf{y}$ and $\Psi(x) = \lambda$.

We have already seen that the log-likelihood as defined in (4) is a contrast function for the extraction of a source s_j independently distributed with respect to $\tilde{\mathbf{s}}$. However, since $\Psi(x)$ in (7) is an approximation thereof, we need to investigate under what conditions the above approximate likelihood is indeed a contrast function.

Proposition 2:

$$\Psi(x) = \frac{\Phi_x^{s_j}}{\Phi_x} \text{ subject to } x = \mathbf{h}^T \mathbf{y} \tag{9}$$

is a contrast for the extraction of s_j under the following sufficient conditions $\forall i \neq j$:

C1) $\mathbb{E}\{s_j s_i\} = 0$;
C2) $\mathbb{E}_{s_j}\{s_j s_i\} = 0$;
C3) $\mathbb{E}_{s_j}\{s_j^2\}/\mathbb{E}\{s_j^2\} > \mathbb{E}_{s_j}\{s_i^2\}/\mathbb{E}\{s_i^2\}$.

For the proof, we refer the reader to the Appendix.

Remark that the statistical independence of s_j with respect to $\tilde{\mathbf{s}}$ is no longer a necessary condition and that this condition has been relaxed to second-order independence (decorrelation) only. Since the conditioning on $|s_j| \geq C$ is used for the calculation of the conditional variance, we refer to our method as that of *maximum variance in the tails* of the conditional distribution (*MaxViT*).

Remark that we can formulate a slightly adapted version for the MaxViT contrast function as

$$\Psi'(x) = \frac{\mathbb{E}_{s_j}\{x^2\} - \mathbb{E}\{x^2\}}{\mathbb{E}\{x^2\}} = \Psi(x) - 1.$$

This equation has the same maximizer $\hat{\mathbf{h}}$, but we now have that all eigenvalues—other than the major eigenvalue—equal zero under model (1).

IV. CONNECTION TO OTHER METHODS

While the starting point of our method is quite different from that of most of the methods that will be discussed below, certain connections exist with these methods. We insist on clarifying possible connections before the presentation of the results to motivate our choice of algorithms used in Section V.

A. ICA

In most practical cases, the mutual independence of the sources is an acceptable prior, which makes ICA one of the most popular source separation algorithms nowadays [2], [20]–[22]. We prove next that the independence of the source s_j with respect to $\tilde{\mathbf{s}}$ is a sufficient condition to be recovered by the approximate maximum likelihood estimator of MaxViT, under the assumption that the conditional covariance can be calculated, i.e., the set $\mathcal{C}_{s_j}$ should be available. As such, we show also that our assumptions are more general than those made to derive (4).

Since the independence of the entries already assures that C1) and C2) are met, we are only left to show the plausibility of C3) under the independence assumption. Independence means that $p_{z|s_j}(u) = p_z(u), \forall z \neq f(s_j)$ and where $f(\cdot)$ can be any function. We thus have $\mathbb{E}_{s_j}\{s_i^2\} = \mathbb{E}\{s_i^2\}, \forall i \neq j$ and $\mathbb{E}_{s_j}\{s_j^2\} > \mathbb{E}\{s_i^2\}$, where the last inequality is proven in the Appendix. In addition, the results obtained in the Appendix allow us to alter the condition in (5) to $\int_{(-C,C)} p_{s_j}(u)g(u)du \to 0$, which is a condition on the function $g(u) = \log \hat{p}_{s_j|\mathbb{1}_{s_j}}(u)$ but now directly in relation to $p_{s_j}(u)$.

B. Reference-Based Filtering

When a reference signal is available for the extraction of a source, one can use extraction filters such as obtained, amongst others, via the optimal Wiener filter estimate [12] or via BSSR [13]. In this section, we show that by choosing the right reference for the Wiener filter or the BSSR method, we obtain the same result as with the approximate maximum likelihood estimator of MaxViT under certain conditions.

Consider first the Wiener filter $\mathbf{h}_W = \mathbb{E}\{\mathbf{y}\mathbf{y}^T\}^{-1}\mathbb{E}\{\mathbf{y}r\}$, where r is the reference signal. Taking as a reference $r = s_j$, we have $\mathbf{h}_W = \Phi_{\mathbf{y}}^{-1}\mathbb{E}\{\mathbf{y}s_j\}$, and the variance of the output $x_W = \mathbf{h}_W^T \mathbf{y}$ is

$$\begin{aligned} \Phi_{x_W} &= \mathbb{E}\{s_j \mathbf{y}^T\}\Phi_{\mathbf{y}}^{-1}\mathbb{E}\{\mathbf{y}s_j\} \\ &= \mathbb{E}\{s_j \mathbf{s}^T\}\mathbf{A}^T(\mathbf{A}\Phi_{\mathbf{s}}\mathbf{A}^T)^{-1}\mathbf{A}\mathbb{E}\{\mathbf{s}s_j\} \\ &= \mathbb{E}\{s_j^2\}^2 \left[\Phi_{\mathbf{s}}^{-1}\right]_{jj} \end{aligned} \tag{10}$$

where the last equality follows from C1), from which follows that $\left[\Phi_{\mathbf{s}}^{-1}\right]_{jj} = (\det \Phi_{\mathbf{s}})^{-1}(\det \Phi_{\tilde{\mathbf{s}}}) = (\det \Phi_{\tilde{\mathbf{s}}})^{-1}[\Phi_{\mathbf{s}}]_{jj}^{-1}(\det \Phi_{\tilde{\mathbf{s}}}) = [\Phi_{\mathbf{s}}]_{jj}^{-1} = \mathbb{E}\{s_j^2\}^{-1}$. The conditional variance is analogously given by

$$\Phi_{x_W}^{s_j} = \mathbb{E}\{s_j^2\}^2 \left[(\Phi_{\mathbf{s}}^{s_j})^{-1}\right]_{jj} \tag{11}$$

where $[(\Phi_{\mathbf{s}}^{s_j})^{-1}]_{jj} = [(\Phi_{\mathbf{s}}^{s_j})]_{jj}^{-1} = \mathbb{E}_{s_j}\{s_j^2\}^{-1}$ if C2) is fulfilled.

The value of the solution to the Wiener filter in the contrast function can be given by combining (10) and (11) and putting them into (7), yielding

$$\Psi(x_W) = \frac{\Phi_{x_W}^{s_j}}{\Phi_{x_W}} = \frac{\Phi_{s_j}}{\Phi_{s_j}^{s_j}} \leq 1 \leq \frac{\Phi_{s_j}^{s_j}}{\Phi_{s_j}} \qquad (12)$$

with equalities if and only if $\Phi_{s_j} = \Phi_{s_j}^{s_j}$ or $P_{s_j}(\overline{\mathcal{B}}) = 0$. Unfortunately, we then no longer have the dominance of the source s_j in the contrast function since all sources now satisfy $\Phi_{s_i}^{s_j}/\Phi_{s_i} = 1$.

For the BSSR method, we have the more general objective function (defined for real variables)

$$\phi_{BSSR}^{(p)}(x) = \frac{1}{2n} \mathbb{E}\left\{(xr)^{2p}\right\} - \frac{\lambda}{2}(\mathbf{h}^T\mathbf{h} - 1) \qquad (13)$$

where r is an a-priori defined reference signal and $2p$ the order. An iterative fixed-point algorithm has been proposed in [13] to maximize this function, but algebraic solutions exist at orders $p = 1/2$ and $p = 1$. At $p = 1/2$, the BSSR cost function accepts the closed-form solution $\hat{\mathbf{h}} = \mathbb{E}\{\mathbf{y}r\}$ and is equivalent to the optimal Wiener filter associated with reference signal r if the observations $\mathbf{y}$ are spatially white ($\Phi_{\mathbf{y}} = \mathbf{I}_m$). At order $p = 1$, the cost function can also be solved algebraically; indeed, $\hat{\mathbf{h}}$ is then given by the dominant eigenvector of the reference-weighted covariance matrix $\mathbb{E}\{\mathbf{y}\mathbf{y}^T r^2\}$.

A similar approach can be found in the QHOC as developed in [14] and [15], where we have in the real, instantaneous case

$$\Psi_{QHOC}^{(R)}(x, \mathbf{r}) = \kappa_{R,\mathbf{r}}\{x\}, \text{ subject to } \Phi_x = 1 \qquad (14)$$

with $\kappa_{R,\mathbf{r}}\{x\} = \mathrm{Cum}\{x, x, r_1, r_2 \ldots r_{R-2}\}$. In practice, most often a single reference signal r is used, which is arbitrarily chosen (e.g., as an arbitrary linear combination of the observations). By the multilinearity of cumulants, one may then write $\kappa_{R,r}\{x\} = \mathbf{h}^T \Phi_{\mathbf{y},r}\mathbf{h}$, where $(\Phi_{\mathbf{y},r})_{i,j} = \mathrm{Cum}\{y_i, y_j, \underbrace{r, r \ldots r}_{\times(R-2)}\}$. Alternating between updates of $\mathbf{h} = \arg\max_{\mathbf{h}} \mathbf{h}^T \Phi_{\mathbf{y},r}\mathbf{h}$ (subjected to $\mathbf{h}^T \Phi_{\mathbf{y}}\mathbf{h} = 1$) and recalculating the reference as $r = \mathbf{h}^T\mathbf{y}$, one then obtains a source estimate $\hat{s}_i$. Contrary to the BSSR method, QHOC aims at estimating the full separation, and thus no source order has been fixed for the successive extractions (with possible deflation). Evidently, the reference could be chosen with respect to a specific source, similar to the BSSR method. If the observations are prewhitened, the BSSR objective [(13)] for $p = 1$ and the QHOC objective [(14)] for $R = 2$ are then essentially the same.

The BSSR and QHOC methods are closely related in the sense that their objective differs mainly in their choice of reference. Moreover, the BSSR at order $2p = 2$ and QHOC method for any pair order R are similar to the MaxViT method when using the following specific reference:

$$r[k] = \begin{cases} \mathrm{sign}\left(s_j[k]\right), & \text{if } |s_j[k]| \geq C \\ 0, & \text{otherwise} \end{cases} \qquad (15)$$

and for a spatially white observation vector $\mathbf{y}$. The latter is not explicitly required by MaxViT, which renders MaxViT less susceptible to the performance bounds imposed by a prewhitening

stage [9], [23]. In addition, the choice of a conditional probability (resulting in the reference signal of (15) for QHOC or BSSR) guarantees an algebraic solution for MaxViT and no cumulating errors due to successive estimation-deflation procedures as with a random reference in [15].

It should also be noted that the BSSR method in [13] has been proposed with a specific application in mind, and little research has been done on its convergence and robustness. In this paper, we make use of the connection between BSSR and our method to show the robustness with respect to arbitrary binary references, which we prefer to address as a conditional indicator function. Note that, whenever we will refer to BSSR in what follows, we refer explicitly to the original iterative implementation as can be found in [13].

C. ICA With a Reference

As we have shown in Section IV-A, the independent source model, which is the basis for ICA, is also a suitable MaxViT model under a not too restrictive condition [i.e., the conditions C1)–C3) are generally satisfied under the ICA model]. Within this perspective, MaxViT may be seen as a direct competitor to cICA [11]. While the methods of cICA are generally based on an augmented Lagrangian in the framework of constrained programming using iterative updating methods, the contrast function in MaxViT offers a closed-form estimator for the extraction filter. Contrary to the family of cICA algorithms, we can now guarantee a global optimizer in low-noise conditions. Moreover, in the noiseless case and for independent sources, MaxViT will provide a filter estimate from which we can obtain the independent source, under the condition that we can approximate the conditional set $\mathcal{C}_{s_j}$ sufficiently well. A simple indicator function $\mathbb{1}_{s_j}$ can be used to construct a simple binary reference signal as in (15) (see also [11], [24]), where we have now shown its relation to a maximum likelihood approach. It should be investigated whether another choice for the conditional probability function $\hat{p}_{s_j|\mathbb{1}_{s_j}}$ with an appropriate updating rule would yield better results.

D. Sparse Decompositions

Also interesting is the similarity between our method and the sparsity pursuit methods (e.g., [25]), where the objective is to have a low approximation error of the observations (with respect to some measure, generally ℓ_2) with as few representative basis functions as possible. This is similar to the objective in our MaxViT model, aimed at minimizing the approximation error (through a maximization of the explained variance of the observations) on a limited amount of samples (the basis functions being Dirac functions). While MaxViT needs a prior knowledge about the presence of s_j, which is reflected in the condition $\mathbb{1}_{s_j}$, sparsity pursuit for multidimensional signals aims at searching a combination of a minimum number of dictionary elements to approximate the observations [26] when the mixing matrix $\mathbf{A}$ is supposed known. Combining these two strategies would give a weighted conditional covariance $\mathbb{E}_{s_j}\{\mathbf{y}\mathbf{y}^T\} = \mathbb{E}\{\mathbf{y}\mathbf{D}(\mathbb{1}_{s_j})\mathbf{D}(\mathbb{1}_{s_j})^T\mathbf{y}^T\}$, where $\mathbf{D}$ can be any linear basis and $\mathbb{1}_{s_j}$ acts as an indicator function for those elements in that dictionary on which s_j has a significant presence. Maximizing the MaxViT contrast under a maximum sparsity constraint could

then be done jointly over $\mathbf{h}$ and $\mathbb{I}_{s_j}$, but this problem will be tackled in future research.

Within the framework of sparse component analysis, MaxViT—calculating the variance dominance on a subset of the observations—can also be seen in the category of algorithms based on piecewise linear source separation [26], [27]. The latter has the basic assumption that outside the support of the source of interest, its amplitude is zero or is captured in the background noise with a predefined (low) noise variance (our σ_ε). MaxViT has an equivalent assumption on the source presence, as may be seen from Section III-C.

E. The Method of Conditional Moments

At first glance, our approach seems similar to the MCM of [16]. MCM is based on a generalized eigenvalue decomposition of second-order moments conditioned on the half-space $x_i > 0$ of the current estimates $\mathbf{x} = \mathbf{H}^T \mathbf{y}$. As is the case for QHOC, an estimate is obtained by iterating over alternating updates, this time between $\mathbf{H} = \arg\max \mathbf{H}^T (\Phi_{\mathbf{x}}^{x_i})^{-1} \tilde{\Phi}_{\mathbf{x}}^{x_i} \mathbf{H}$ and $\mathbf{x} \leftarrow \mathbf{H}^T \mathbf{x}$, with $\Phi_{\mathbf{x}}^{x_i} = \mathbb{E}\{\mathbf{x}\mathbf{x}^T | x_i > 0\}$ and $\tilde{\Phi}_{\mathbf{x}}^{x_i} = \Phi_{\mathbf{x}}^{x_i} - \mathbb{E}\{\mathbf{x}|x_i > 0\}\mathbb{E}\{\mathbf{x}|x_i > 0\}^T$. As opposed to MCM, our method does not limit itself to Laplace or uniformly distributed sources and does not impose a correction on the estimation of Laplace (super Gaussian) sources. In addition, contrary to MCM, MaxViT does not limit its application to symmetrically distributed sources and can deal with Gaussian sources. Of course, these advantages go at the expense of the extra additional information of the presence indicator $\mathbb{I}_{s_j}$.

In this paper, we have also shown that methods based on conditional moments are derived from a conditional likelihood function and that a specific class of these conditional likelihood functions for which $\int_{(-C,C)} p_{s_j}(u) \log \hat{p}_{s_j | \mathbb{I}_{s_j}}(u) du \to 0$ (see Section IV-A) result in a contrast under model (1).

V. PERFORMANCE ANALYSIS

A. Theoretical Estimation Bounds of MaxViT

In this section, we establish the error bounds on the estimation of s_j in the model $x = \mathbf{h}^T \mathbf{A}\mathbf{s} = \mathbf{g}^T \mathbf{s}$. This error can be measured through the interference-to-signal ratio (ISR) defined as

$$\text{ISR} = \frac{\sum_{i \neq j} |g_i|^2}{(n-1)|g_j|^2} \qquad (16)$$

which is a measure for the average interference and takes the value zero if and only if the extraction filter is the jth canonical vector.

The filter $\mathbf{g}$ is the product of the dominant generalized eigenvector $\mathbf{h}$ associated to $\mathbf{A}\Phi_{\mathbf{s}}^{s_j}\mathbf{A}^T / \mathbf{A}\Phi_{\mathbf{s}}\mathbf{A}^T$ and $\mathbf{A}$. Here, we only consider the bias in the estimate of $\mathbf{h}$ as a consequence of the nonvanishing conditional covariance between s_j and s_i, $i \neq j$. The ISR as a function of this covariance can be calculated for $\mathbf{s} \in \mathbb{R}^2$ as

$$\text{ISR}(\rho, \delta) = \frac{\sqrt{\delta^2 + |\rho|^2} - \delta}{\sqrt{\delta^2 + |\rho|^2} + \delta} \stackrel{\delta \neq 0}{=} \frac{\sqrt{1 + (|\rho|/\delta)^2} - 1}{\sqrt{1 + (|\rho|/\delta)^2} + 1} \qquad (17)$$
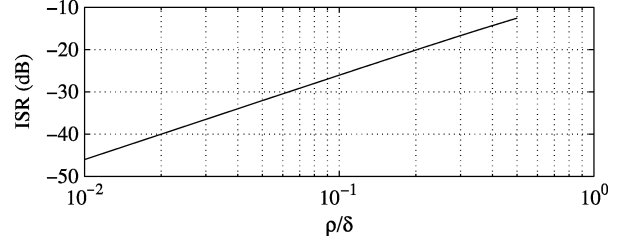

Fig. 1. Theoretical bounds for the value of ISR as a function of the conditional covariance ρ and the conditional variance domination δ.

	Uniform	Normal	Laplace
$C = 1$	0.12±0.09	0.09±0.07	0.07±0.06
$C = \sqrt{2}$	0.12±0.10	0.09±0.07	0.07±0.06
$C = \sqrt{3}$	N/A[1]	0.10±0.06	0.09±0.06

[1]For $c = \sqrt{3}$, we have $P_{s_j}(\mathcal{B}) = 0$ and our method is not applicable (N/A).

where

$$\delta = \left(\Phi_{s_j}^{s_j} - \Phi_{s_i}^{s_j} \right) / 2 \qquad (18)$$

$$\rho = \mathbb{E}_{s_j}\{s_j s_i\}. \qquad (19)$$

The calculations for the value of ISR are given in Appendixes D and E; and the relation between $|\rho|/\delta$ and the theoretical ISR value is given in Fig. 1. We can give an impression of the accuracy of this theoretical measure by comparing it with the obtained ISR as obtained through the relation of (16). We did this for 1000 Monte Carlo realizations of two independent identically distributed (i.i.d.) (respectively Laplacian, Gaussian, and uniformly distributed) unit variance, zero-mean source signals of $K = 1000$ samples each observed through an orthonormal mixing matrix $\mathbf{A}$. With this simulations, we obtained a maximum absolute error of 9.44 10^{-16}, confirming the accuracy of (17).

From both (17) and Fig. 1, we see that $x = \hat{s}_j \approx s_j$ if $|\rho| \ll \delta$, and a good estimation of the source s_j is guaranteed even if $|\rho|$ tends to δ (we still have a theoretical -7.7 dB if $|\rho| = \delta$), which is a reasonable assumption in many practical situations. It can be seen that the smaller the discrimination in the conditional variance becomes in C3, the more stringent the condition C2 on $|\rho|$ becomes [and thus automatically also C1]. Under the condition $|\rho| = 0$, we have already shown that (9) is a contrast for the separation of s_j from a mixture in Section III-D; and, indeed, we obtain ISR= 0 from (17), as long as $\delta \neq 0$ (for $\delta = 0$, we have that ISR= 1, since no discrimination is possible with the current contrast).

Table I gives the mean fraction of $|\rho|/\delta$ for three different distributions (uniform, Laplacian, and normal) and for different values of C. Note that the number of sample indexes in the set $\mathcal{C}_{s_j}$ differs according to the chosen distribution and consequently has a considerable influence on the variance of the statistics $\hat{\mathbb{E}}_{s_j}\{f(x)\} = \sum_{k \in \mathcal{C}_{s_j}} f(x[k])/\#\mathcal{C}_{s_j}$. Therefore, we decided to use K samples on a basis of K_b, where

$K = K_b/P(\mathbb{1}_{s_j})$. This brings the number of sample indexes in $\mathcal{C}_{s_j}$ from which $\mathbb{E}_{s_j}\{f(u)\}$ is estimated to an almost equal number, independent of the distribution used. The ISR or the fraction $|\rho|/\delta$ can now directly be compared for a given K_b.

Equation (17) is a compact expression for case $\mathbf{s} \in \mathbb{R}^2$, but for $\mathbf{s} \in \mathbb{R}^n$, $n > 2$ the calculations become more cumbersome. For $n = 3$, we turn to simulations on a synthetic dataset, for which we give the results below.

B. Experimental Performance Comparison

To compare the performance of our algorithm with respect to the related algorithms discussed in Section IV, a dataset has been created based on realizations of a source vector $\mathbf{s} \in \mathbb{R}^3$, for which we have $K = 1000$ realizations. The entries in $\{\mathbf{s}\}_{10^3}$ are samples from an i.i.d. unit-variance zero-mean Laplacian distribution. The so-obtained source signals are then transformed through a unitary matrix $\mathbf{A}$ to the observation space $\mathbf{y} = \mathbf{As}$. Without further specifications, we have set $C = 1$ to determine the conditional probabilities and let s_1 be the source of interest.

The algorithms of the Wiener filter and our MaxViT algorithm both have a closed-form solution, while the ICA algorithm (COM2 [2], without prewhitening, since we have a unitary mixture) and the BSSR algorithm (taken at higher order $2p = 4$ for the evident reason of avoiding similarity with our MaxViT contrast; see Section IV-B) are iterative. The COM2 algorithm has been run over the classical $\lfloor 1 + \sqrt{n} \rfloor$ sweeps over all the signal pairs, which guarantees (although heuristically) its convergence. The BSSR algorithm has either been run until convergence or over 10^3 iterations, whatever has been reached first. Since COM2 provides a separation rather than an extraction, we only retained the output x_i that had the highest correlation with s_1, the source of interest.

Both BSSR and the Wiener filter can be used with different reference signals. To restrict the wide scope of possibilities, we retain only those references that have a close resemblance with the conditional used for MaxViT, i.e., through the indicator function $\mathbb{1}_{s_1}$. The so-obtained reference signal r is then defined as

$$r[k] = \begin{cases} s_j[k], & \text{if } |s_j[k]| > C \\ 0, & \text{otherwise} \end{cases}. \qquad (20)$$

Derivations of this reference function defined as $b = \text{sign}(r)$ [see also (15)] or $|b| = |\text{sign}(r)|$ are also used, where we define $\text{sign}(0) = 0$. Similar reference functions have also been proposed, e.g., in [11] and [24]. All experiments are evaluated over 1000 Monte Carlo realizations of $\{\mathbf{s}\}_{10^3}$ and $\mathbf{A}$.

In Table II, we show the mean ISR value as defined in (16). The ISR is a measure that quantitatively measures the estimation of the filter $\mathbf{h}$, through an evaluation of $\mathbf{g} = \mathbf{A}^T\mathbf{h}$. Contrary to measures such as Pearson's correlation coefficient, it is an asymptotic evaluation of the ISR and does not make any assumption on the distribution of the error. Table II is organized in such a way that, reading it from left to right, the information content in the reference signal decreases. The values between brackets are obtained after a rotation of the i.i.d. vector $\mathbf{s}$ by a unitary matrix. This results in decorrelated entries of $\mathbf{s}$ that are no longer guaranteed independent.

TABLE II
ISR as a Measure for the Asymptotic Accuracy of the Source Estimate From a Synthetic Dataset of Three i.i.d. Laplacian Sources of $K = 10^3$ Samples for Different Algorithms and Different Information Feeds. Values Between Brackets Are Obtained From Uncorrelated Sources, Which Are Not Independent

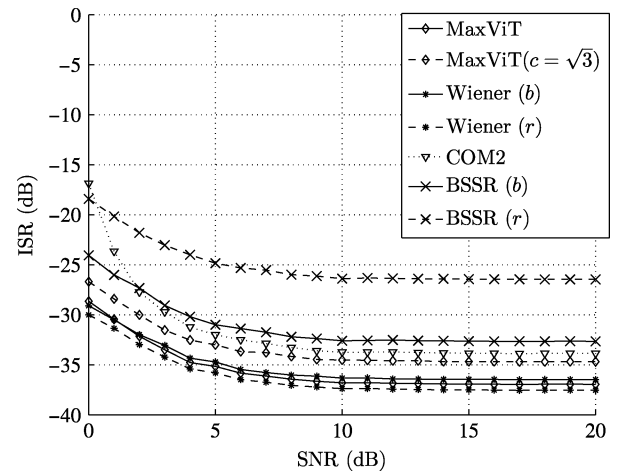|	r	b	$|b|$	no ref.
MaxViT(1)	.	.	-36.01	.
	.	.	(-28.62)	.
MaxViT($\sqrt{3}$)	.	.	-34.19	.
	.	.	(-28.58)	.
Wiener	-36.61	-35.80	17.31	.
	(-29.37)	(-28.91)	(17.08)	.
BSSR$_{p=2}$	-26.13	-31.91	-31.91	.
	(-25.37)	(-29.30)	(-29.30)	.
COM2	.	.	.	-33.80
	.	.	.	(-24.26)

Fig. 2. The source interference ISR (dB) as a function of the SNR (dB). The noise is normally distributed additive noise (see text for details). The method is compared with a classical ICA method, the BSSR solution, and the solution by a Wiener filter.

C. Influence of Additive Noise

We start from the same observations and source signals as defined above. To discard the influence of the parameter quotient $|\rho|/\delta$ on the ISR—see (17)—we assure that we have $p_{s_i}(|u| > C|\mathbb{1}_{s_j}) = p_{s_j}(|u| > C|\mathbb{1}_{s_i}) = 0, \forall i \neq j$ by permuting the samples of $\{s_j\}_K$ appropriately. To test the performance of the algorithm under noisy conditions, centered Gaussian noise $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \sigma_\eta^2\mathbf{I}_3)$ has been added to the observations $\mathbf{y}$. Since the observations are standardized and the noise is isotropic, the signal-to-noise ratio (SNR) can be given by the simple expression $\text{SNR} = \sigma_\eta^{-2}$. The model reads $\mathbf{y} = \mathbf{As} + \boldsymbol{\eta}$ and the estimate of s_j is $x = \mathbf{g}^T\mathbf{s} + \mathbf{g}^T\boldsymbol{\eta}$.

The influence of the SNR on the performance parameter ISR is shown in Fig. 2. Since in the case of additive noise an accurate estimate of the filter does not guarantee an accurate estimate of the source, we also give the value of $1 - |\hat{\rho}|$, with $\hat{\rho}$ the sample estimate of $\mathbb{E}\{xs_j\}/(\mathbb{E}\{x^2\}^{1/2}\mathbb{E}\{s_j^2\}^{1/2})$. This direct comparison between the source estimate x and the source s_1 can be found in Fig. 3. The comparison of MaxViT has been carried out with respect to the algorithms used in Table II, however, making a selection of reference signals that we judged most useful for
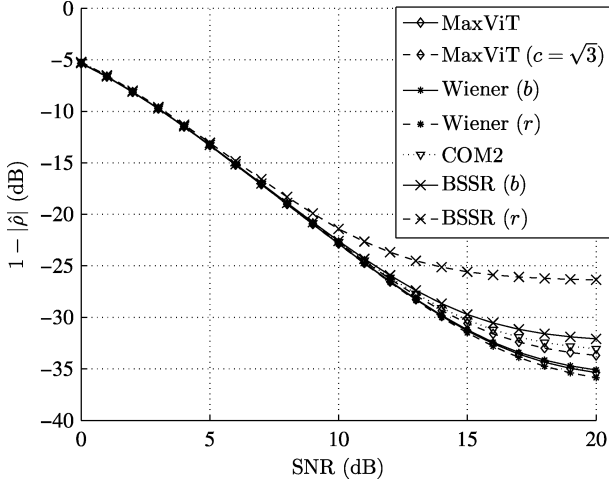
Fig. 3. The "correlation" $1 - \hat{\rho}(x, s_j)$ (dB) as a function of the SNR (dB). The noise is normally distributed zero-mean additive isotropic noise (see text for details). The method is compared with a classical ICA method, the BSSR solution, and the solution by a Wiener filter.

comparison. This includes the performance of a Wiener filter and the BSSR method with an unsigned reference $|b|$, adding exactly the same amount of information as is used in MaxViT.

To complete the performance picture, we also add MaxViT with $c = \sqrt{3}$ for comparison. Note that in Fig. 3, the Wiener solution has all of its performance values out of the range used for plotting (ISR (Wiener($|b|$)) $\in [10, 50]$dB).

D. Robustness With Respect to the Conditional Set

Assume we no longer have $p(\cdot | \ |s_j| > C\sigma_{s_j})$ but rather $p(\cdot | \ |s_j| + \eta > C\sigma_{s_j})$, where η is a nuisance parameter expressing the uncertainty we have about our initial condition. As before, let us denote by $\mathcal{C}_{s_j} = \{k | \ |s_j[k]| > C\sigma_{s_j}\}$ the conditional set of sample indexes. We can now suppose that the condition $|s_j| + \eta > C\sigma_{s_j}$ gives rise to an indicator $\hat{\mathbb{1}}_{s_j}$, which results in a mismatch in the conditional set $\mathcal{C}_{s_j}$. In what follows, we experimentally analyze the robustness of the algorithm with respect to a mismatch of the conditional set $\mathcal{C}_{s_j}$.

As above, we have $K = 1000$ realizations of three i.i.d. standardized Laplacian sources $\mathbf{s}$ observed in $\mathbf{y}$ through a unitary mixture $\mathbf{A}$. The samples of s_j have been permuted such that $\forall i \neq j, \mathcal{C}_{s_j} \bigcap (\bigcup_i \mathcal{C}_{s_i}) = \emptyset$, and thus the source s_j can be estimated since we have $\mathbb{E}\{s_i s_j\} \approx 0$, $\mathbb{E}_{s_j}\{s_i s_j\} \approx 0$ (i.i.d. variables) and $\Phi^{s_j}_{s_j} / \Phi_{s_j} > 1$ (see Sections IV-A and V-A). Note that we artificially lowered the theoretical ISR estimation bound by permuting the samples and thus augmenting δ. Also, define the following sets of sample indexes:

- $\mathcal{K} = \{k | k \in \mathbb{N}, 1 \leq k \leq K\}$;
- $\overline{\mathcal{C}}_{s_i} = \mathcal{K} \setminus \mathcal{C}_{s_i}$;
- $\mathcal{C}_{ne} = \bigcap_i \overline{\mathcal{C}}_{s_i}$;
- $\mathcal{C}_{co,j} = (\bigcup_{i \neq j} \mathcal{C}_{s_i}) \setminus \mathcal{C}_{s_j}$.

The latter two sets are, respectively, the neutral and the conflicting set with respect to s_j.

Consider also the following three set operations:

- $\mathcal{P}_2 = \text{Shrinkage}(\mathcal{P}_1, \alpha)$

$$\mathcal{P}_2 \subseteq \mathcal{P}_1 \text{ with } \#(\mathcal{P}_2) = (1 - \alpha) \times \#(\mathcal{P}_1)$$

- $\mathcal{P}_3 = \text{Inflation}(\mathcal{P}_1, \mathcal{P}_2, \alpha)$

$$\mathcal{P}_3 = \mathcal{P}_1 \cup \mathcal{Z}$$

with $\#(\mathcal{P}_3) = (1 + \alpha) \times \#(\mathcal{P}_1)$ and $\mathcal{Z} \subseteq \mathcal{P}_2$

- $\mathcal{P}_3 = \text{Interchange}(\mathcal{P}_1, \mathcal{P}_2, \alpha)$

$$\mathcal{P}_3 = \text{Inflation}\left(\text{Shrinkage}(\mathcal{P}_1, \alpha), \mathcal{P}_2, \frac{\alpha}{1 - \alpha}\right)$$

where # is the cardinal number of the set and $0 \leq \alpha \leq 1$.

By applying set operations to $\mathcal{C}_{s_j}$, we obtain an estimate of the perturbation of the conditional probability $p(\cdot | \ |s_j| + \eta > C\sigma_{s_j})$, as has been explained above. The results of this perturbation study can be found in Fig. 4, where we present the results of the above-defined set operations with $\mathcal{C}_{s_j}$ as the basis set. The set ($\mathcal{P}_2$) is chosen as $\mathcal{C}_{ne}$ or $\mathcal{C}_{co,j}$ for a neutral, respectively, a conflicting operation with respect to s_j. The influence of the set perturbation is expressed in terms of the source interference ratio ISR (16).

The following scenarios can now be investigated to observe the behavior in the most optimistic, a random, or, respectively, the most pessimistic scenario.

- The smart set choice [see Fig. 4(c)] has a shrinkage operator for which $\forall k_1 \in \mathcal{P}_2, k_2 \in \mathcal{P}_1 \setminus \mathcal{P}_2 : |s_j[k_1]| \geq |s_j[k_2]|$ and an inflation operator for which $\forall k_1 \in \mathcal{Z}, k_2 \in \mathcal{P}_2 \setminus \mathcal{Z} : |s_j[k_1]| \leq |s_j[k_2]|$.
- The random set choices [see Fig. 4(b)] does not make any assumption about a possible order and apply the rules naïvely.
- The worst case scenario [see Fig. 4(a)] has a shrinkage operator for which $\forall k_1 \in \mathcal{P}_2, k_2 \in \mathcal{P}_1 \setminus \mathcal{P}_2 : |s_j[k_1]| \leq |s_j[k_2]|$ and an inflation operator for which $\forall k_1 \in \mathcal{Z}, k_2 \in \mathcal{P}_2 \setminus \mathcal{Z} : |s_j[k_1]| \geq |s_j[k_2]|$.

VI. DISCUSSION

The performance of MaxViT in the noiseless case has shown competitive results with respect to the algorithms used in the comparison Table II. MaxViT even outperforms the reference-based algorithms BSSR ($2p = 4$) and the limited support Wiener filter that have access to a larger amount of information (b instead of $|b|$ makes a 1 bit per sample information gain). We also outperform a completely blind algorithm based on higher order statistics (COM2), showing the advantage of using a probability conditioned on the source of interest only. Moreover, the little performance gain that can be obtained by the Wiener filter is at the expense of a highly informative prior, using the waveform r from (20), which is generally not available. In an observation environment contaminated by additive isotropic Gaussian distributed zero-mean i.i.d. noise, the MaxViT estimator shows to be robust, being competitive with the methods used in the comparison, with a slight estimation gain over almost the whole SNR range used in the simulations (Figs. 2 and 3). The only competitor that outperforms MaxViT when additive noise is present is the Wiener estimate with reference r from (20).

We also observe from Table II that the performance of BSSR remains equal, whether a signed or unsigned binary reference is used. This is an immediate consequence of the limitation of the
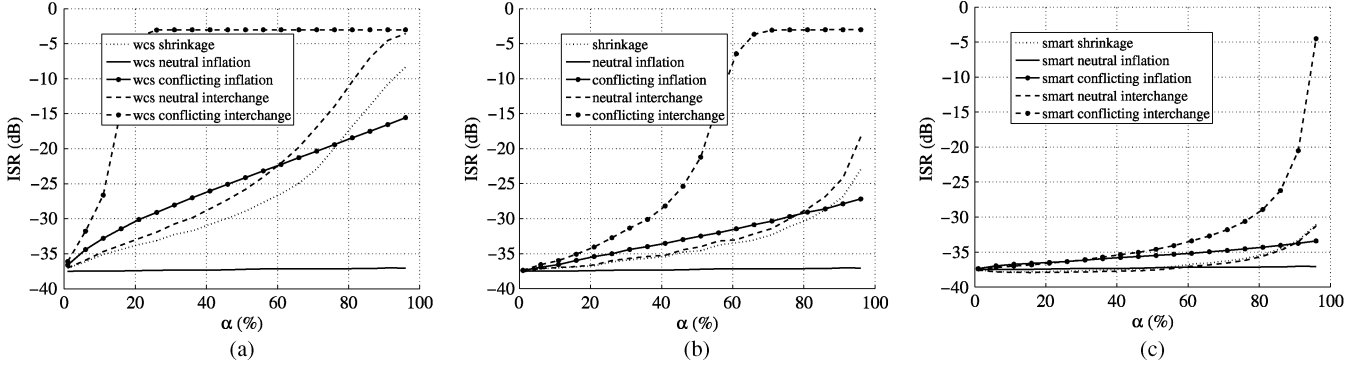
Fig. 4. The effect of a perturbation of the conditional set $\mathcal{C}_{s_j}$. The effect of the mismatch is measured through the source interference ISR (16) as a function of the relative number of samples α that are affected by the set operations. See text for more details.

BSSR algorithm to use even powers of the reference signal ($2p$) [13]. Surprisingly, as can be seen from the same Table II, the BSSR algorithm (and we may assume that the same would hold for the QHOC algorithm) does not yield a significant increase in estimation accuracy with an increase in available information, i.e., changing the reference signal from b to r. This points out that the conditional relative variance may be seen as a sufficient statistic to extract the source s_j from the mixture.

The MaxViT estimator also has been shown to be quite robust to mismatches with respect to the conditional set $\mathcal{C}_{s_j}$; see Fig. 4(b). This distinguishes our method from other works such as [10], [13], and [28], where the estimator is reported to be susceptible to mismatches between the used reference and $s[k]_j$, especially with respect to its phase. Empirically, BSSR has been shown robust to reference mismatches including time shift and sample omission (i.e., the equivalent of an interchange, respectively, a shrinkage operator acting on the set $\mathcal{C}_{s_j}$) [29]. By placing BSSR in the framework of MaxViT, this can now partly be explained by the robustness of MaxViT to the conditional set $\mathcal{C}_{s_j}$. This follows from the assumption that BSSR with a binary reference inherits certain properties of MaxViT while being equivalent to MaxViT for $p = 1$ and reference signal b defined in Section V-B.

The errors induced by the mismatch between the conditional set $\mathcal{C}_{s_j}$ and $\hat{\mathcal{C}}_{s_j}$ fed as a prior to the algorithm are comparable to those induced by additive noise, as has been suggested in Section V-D. However, notice that small errors in the filter estimate do induce a smaller error in the filter output than does the additive noise. This can be deduced from the fact that our filter output can be written as a function of the optimal filter $\mathbf{h}^\star$ and a perturbation $\varepsilon_\mathbf{h}$ as $x^\star + \varepsilon_x = (\mathbf{h}^\star + \varepsilon_\mathbf{h})^T \mathbf{y}$, whereas in the case of additive noise, the same error in the filter estimate results in $x^\star + \varepsilon_{x,2} = (\mathbf{h}^\star + \varepsilon_\mathbf{h})^T (\mathbf{y} + \boldsymbol{\eta}) = x^\star + \varepsilon_x + (\mathbf{h}^\star + \varepsilon_\mathbf{h})^T \boldsymbol{\eta}$. Thus for the same error in the filter estimate, we logically obtain a better estimate of the source if the error is due to the set mismatch only.

Note that despite the use of specially designed simulations to reduce the fraction $|\rho|/\delta$ and thus to minimize the ISR (by choosing $\mathcal{C}_{s_j} \bigcap (\bigcup_{i \neq j} \mathcal{C}_{s_i}) = \emptyset$), we may generalize our results to independently distributed sources that have not been corrected. This is because, from Table I, we have that the fraction $|\rho|/\delta$ generally remains acceptably small for i.i.d. Laplacian, Gaussian, and uniform sources.

A quick overview of the performance of the MaxViT algorithm can be given by evaluating under what conditions we obtain an acceptable ISR of -30 dB. It follows from Figs. 2 and 3 that we accept a signal to noise level no lower than 4 dB and [from Figs. 4(a)–(c)] a worst case interchange of indexes of $\mathcal{C}_{s_j}$ with $\mathcal{C}_{s_i}$ of up to 7% of $\#\mathcal{C}_{s_j}$. However, in practical situations, an estimation of the set $\mathcal{C}_{s_j}$ is usually done with more care and, even when unfortunately chosen, we would interchange samples between sets randomly rather than creating a worst case scenario. This random interchange can be done for up to some 30% to 70% of the samples of $\mathcal{C}_{s_j}$, depending on whether conflicting, respectively, neutral sample indexes have been involved. In practical situations, a set estimate $\hat{\mathcal{C}}_{s_j}$ offering a considerable performance should thus often be available, e.g., by using a threshold on the amplitude of the observations (as in [28]) or based upon prior knowledge of the support in the frequency domain (see, e.g., [30]).

The estimation of a source s_j from a set of observations $\mathbf{y}$ can be done for every source in the mixture (approximately) satisfying the sufficient conditions C1)–C3), and this whenever an approximation of its conditional set is available. When more than one source is of interest, we propose an iterative estimation without deflation, especially when $\#(\mathcal{C}_{s_i} \bigcap \mathcal{C}_{s_j})$ is relatively small. Avoiding the subtraction of the projection of $\mathbf{y}$ onto s_j from $\mathbf{y}$ prior to estimating s_i reduces the possible error propagation from which these deflation approaches suffer.

As explained in Section III, the only constant in MaxViT that has to be set is C, and its influence on different source distributions can be found in Table I. We see that its value is not critical, at least for large sample populations. In practical situations, where only a limited population sample is available, it should neither be taken too large nor too small because the conditional covariance with respect to $\mathbb{1}_{s_j}$, respectively (implictly), with respect to $\bar{\mathbb{1}}_{s_j}$, would be calculated on too small a sample set, and its estimation would suffer from an increase in variance.

VII. CONCLUSION

We have shown that maximizing the likelihood criterion conditioned on a presence indicator gives rise to a contrast function for the extraction of a source of interest. The filter corresponding to the optimum of the contrast function can be found algebraically, provided that the conditional second moment can be estimated from the observations. The MaxViT estimator has

interesting properties, such as robustness to noise or perturbations of the conditional set, making it a valuable alternative to constrained ICA algorithms.

APPENDIX

A. Proof of Proposition 1

P1) holds if we consider unit variance variables only. This can be done without loss of generality. Since the log-likelihood $\mathcal{L}(x)$ of (4) is either zero or tends to $-\infty$, and $\mathcal{L}(x) = 0$ holds if and only if we have $p_{x|\mathbb{I}_{s_j}}(u) = 0, \forall u \in \overline{\mathbb{B}}$ and $p_{x|\overline{\mathbb{I}}_{s_j}}(u) = 0, \forall u \in \mathbb{B}$ P2), it remains to prove that the latter two conditions imply the equality $x = \lambda_j s_j$ of P3).

Proof: Suppose that we have $x \neq \lambda_j s_j$, and thus $x = \mathbf{g}^T \mathbf{s} = g_j s_j + \sum_{i \neq j} g_i s_i = g_j s_j + \breve{s}$, where at least one g_i has a nonzero value and for which $p_{x|\mathbb{I}_{s_j}}(u) = 0, \forall u \in \overline{\mathbb{B}}$.

Since s_j is independently distributed with respect to $\tilde{\mathbf{s}}$ and thus with respect to all $s_i (i \neq j)$, we have that s_j is independently distributed with respect to $\breve{s}$. As a consequence, the distribution $p_{x|\mathbb{I}_{s_j}}(u)$ can be written as the convolution of the distributions $p_{s_j|\mathbb{I}_{s_j}}(u)$ and $p_{\breve{s}}(u)$, or

$$p_{x|\mathbb{I}_{s_j}}(u) = \int_{\mathbb{R}} p_{s_j|\mathbb{I}_{s_j}}(\tau - u) p_{\breve{s}}(\tau) d\tau. \qquad (21)$$

A necessary condition for x to yield $\mathcal{L}(x) = 0$ is that $p_{x|\mathbb{I}_{s_j}}(u) = 0, \forall u \in \overline{\mathbb{B}}$. However, if $\exists \varepsilon$ with nonzero Lebesgue measure for which the support set $\mathcal{S}$ of $p_{\breve{s}}$ has a measure $|\mathcal{S}| \geq \varepsilon$ and for which $p_{s_j|\mathbb{I}_{s_j}}(u) > 0, \forall u : |u| \in [C, C + |\varepsilon|]$, then, by (21), $\exists u : |u| < C, p_{x|\mathbb{I}_{s_j}}(u) > 0$. As a consequence, our initial supposition was wrong, and we must have $g_i = 0, \forall i \neq j$, i.e., $\mathcal{L}(x) = 0 \Rightarrow x = \lambda_j s_j$. ∎

An analogous reasoning can be used to prove that $p_{x|\overline{\mathbb{I}}_{s_j}}(u) = 0, \forall u \in \mathbb{B}$ implies $x = \lambda s_j$.

B. Proof of $\int_{\overline{\mathbb{B}}} p_{x|\mathbb{I}_{s_j}}(u) \gamma(u) = 0 \Rightarrow p_{x|\mathbb{I}_{s_j}}(u) = 0, \forall u \in \overline{\mathbb{B}}$

Proof: If $\int_{\overline{\mathbb{B}}} p_{x|\mathbb{I}}(u) \gamma(u) du = 0$, with $\gamma(u)$ a positive function taking $\gamma(u) = 0$ for $u = 0$ only. We must have

$$\begin{cases} p_{x|\mathbb{I}}(u) = 0, & \forall u \in \overline{\mathbb{B}} \setminus \{0\} \\ p_{x|\mathbb{I}}(0) = \alpha_k \delta_{u,0} \end{cases}$$

where $\delta_{u,0} = 1$ is Dirac's delta at the origin. Assuming sources with continuous distributions, the second option is not possible unless $\mathbf{h} = \mathbf{0}$. This trivial case is avoided by the constraints introduced later in the construction of the algorithm.

The assumption of continuous distributions can be relaxed if C is chosen such that $\overline{\mathbb{B}}$ and $\mathbb{B}$ both have a nonzero Lebesgue measure (for discrete distributions, we have, respectively, $\#\overline{\mathbb{B}} > 1$ and $\#\mathbb{B} > 1$). ∎

C. The Objective Function of (9) is a Contrast for the Extraction of s_j

To prove that Ψ_x is a contrast under the conditions C1)–C3) from Section III-D, we need to show that it has the properties P1')–P3') from Section II-B.

Proof: The indeterminacy of the source scaling has been taken care of by the denominator in (9), and thus P1') holds. Furthermore, we have

$$\frac{\mathbf{h}^T \Phi_{\mathbf{y}}^{s_j} \mathbf{h}}{\mathbf{h}^T \Phi_{\mathbf{y}} \mathbf{h}} = \frac{\mathbf{g}^T \Phi_{\mathbf{s}}^{s_j} \mathbf{g}}{\mathbf{g}^T \Phi_{\mathbf{s}} \mathbf{g}} = \frac{\sum_i |g_i|^2 \Phi_{s_i}^{s_j}}{\sum_i |g_i|^2}$$

since our sources are uncorrelated, both conditionally and unconditionally. Splitting up the sum in the different contributions gives us (up to a multiplicative positive constant)

$$|g_{jj}|^2 \Phi_{s_j}^{s_j} + \sum_{i \neq j} |g_i|^2 \Phi_{s_i}^{s_j} + \sum_{i \neq j} |g_i|^2 \Phi_{s_j}^{s_j} - \sum_{i \neq j} |g_i|^2 \Phi_{s_j}^{s_j}$$

which can be rewritten as

$$\Phi_{s_j}^{s_j} + \sum_{i \neq j} |g_i|^2 \left(\Phi_{s_i}^{s_j} - \Phi_{s_j}^{s_j} \right) \leq \Phi_{s_j}^{s_j}$$

where the inequality follows from $(\Phi_{s_i}^{s_j} - \Phi_{s_j}^{s_j}) < 0, \forall i \neq j$. This proves the domination.

We also have

$$\Phi_x^{s_j} = \Phi_{s_j}^{s_j} \Leftrightarrow \sum_{i \geq 2} |g_i|^2 \left(\Phi_{s_i}^{s_j} - \Phi_{s_1}^{s_j} \right) = 0.$$

Now, since $(\Phi_{s_i}^{s_j} - \Phi_{s_1}^{s_j}) < 0, \forall i \neq j$, we have the above equality if and only if $|g_i|^2 = 0, \forall i \geq 2$. Thus

$$\frac{\mathbf{g}^H \Phi_{\mathbf{s}}^{s_j} \mathbf{g}}{\mathbf{g}^H \Phi_{\mathbf{s}}^{\mathbb{R}} \mathbf{g}} = \frac{\Phi_{s_j}^{s_j}}{\Phi_{s_1}}.$$

This proves the discrimination and thus, together with the domination, P2') and P3') are fulfilled.

Since any objective function fulfilling P1')–P3') is a contrast function for source extraction, our function $\Psi(x)$ in (9) is a contrast under the conditions C1)–C3). ∎

Note that this could be extended to the case where the covariance $\mathbb{E}_{s_j}\{s_i s_k\} \neq 0, \forall i, k \neq ji \neq k$, as long as $\mathbb{E}_{s_j}\{s_j s_i\} = 0, \forall i \neq j$. For the proof, define $\tilde{\mathbf{s}} = [s_1, s_2 \ldots s_{j-1}, s_{j+1} \ldots s_N]^T$. Now take the eigenvalue decomposition of $\Phi_{\tilde{\mathbf{s}}}^{s_j}$ as $\mathbf{V}^T \Phi_{\tilde{\mathbf{s}}}^{s_j} \mathbf{V} = \Delta$, where Δ is a diagonal matrix with the eigenvalues λ_i on its diagonal and extend $\mathbf{V}$ to

$$\tilde{\mathbf{V}} = \begin{pmatrix} 1 & \mathbf{0}_{n-1}^T \\ \mathbf{0}_{n-1} & \mathbf{V} \end{pmatrix}$$

where $\mathbf{0}_{n-1}$ is a vector of zeros in $\mathbb{R}^{n-1}$. The proof then continues similarly as above but replacing $\Phi_{\tilde{\mathbf{s}}}^{s_j}$ by Δ and $\mathbf{g}$ by $\tilde{\mathbf{V}} \mathbf{g}$. As a consequence, condition C3) in Section III-D becomes $\Phi_{s_j}^{s_j} > \max \lambda_i$.

D. Algebraic Solution for the 2×2 Case

Suppose that $\mathbf{y}$ has uncorrelated, unit-variance, and zero-mean entries, without loss of generalization, since $\mathbf{y}$ can always be rendered so through whitening. Since $\Phi_{\mathbf{y}} = \mathbf{I}_2$, the eigenvector that would separate our source as $x = \mathbf{e}_1^T \mathbf{y}$ is the dominant eigenvector of the covariance matrix $\Phi_{\mathbf{y}}^{s_1}$, which has a general symmetric form

$$\Phi_{\mathbf{y}}^{s_1} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}. \qquad (22)$$

The above matrix has eigenvalues

$$\lambda_{1,2} = \frac{a+c}{2} \pm \sqrt{\left(\frac{a-c}{2}\right)^2 + |b|^2} \quad (23)$$

and thus, if $a \neq c$, has a largest eigenvector

$$\mathbf{e}_1 = \pm \begin{bmatrix} \frac{1}{\sqrt{2}}\sqrt{1 + \frac{\xi}{\xi^2 + |b|^2}} \\ \frac{1}{\sqrt{2}}\sqrt{1 - \frac{\xi}{\xi^2 + |b|^2}} \end{bmatrix} \quad (24)$$

with $\xi = (a-c)/2$.

E. Calculation of ISR

We induce the estimation bound in case the sources are not perfectly conditionally uncorrelated. Since we only consider unitary transformations $\mathbf{A} = \mathbf{Q}$ (for our $\mathbf{y}$ is or has been rendered spatially *white*), we know that the eigenvalues of $\Phi_{\mathbf{s}}^{s_j}$ and $\Phi_{\mathbf{y}}^{s_j}$ are equal. Moreover, the ith eigenvector $\mathbf{q}_i$ of $\Phi_{\mathbf{y}}^{s_j}$ equals $\mathbf{Q}\mathbf{e}_i$, where $\mathbf{e}_i$ is the ith eigenvector of $\Phi_{\mathbf{s}}^{s_j}$ (see also the equivariance property [9]). As a consequence, we only need to consider the simpler case where $\mathbf{Q} = \mathbf{I}_m$, without loss of generality.

Limiting $\mathbf{s}$ to belong to $\mathbb{R}^2$, the matrix $\Phi_{\mathbf{s}}^{s_1}$ takes the form

$$\Phi_{\mathbf{s}}^{s_1} = \begin{bmatrix} \Phi_{s_1}^{s_1} & \mathbb{E}_{s_1}\{s_1 s_2\} \\ \mathbb{E}_{s_1}\{s_1 s_2\} & \Phi_{s_2}^{s_1} \end{bmatrix}. \quad (25)$$

From (24), one can explicitly calculate the separation filter $\mathbf{h}$ associated to $\Phi_{\mathbf{s}}^{s_1}$ by the above (24). As such, we obtain for the ISR $(|g_2|^2/|g_1|^2 = |h_2|^2/|h_1|^2)$

$$\text{ISR} = \frac{\sqrt{\delta^2 + |\rho|^2} - \delta^2}{\sqrt{\delta^2 + |\rho|^2} + \delta^2} \quad (26)$$

with $\delta = (\Phi_{s_1}^{s_1} - \Phi_{s_2}^{s_1})/2$ and $\rho = \mathbb{E}_{s_1}\{s_1 s_2\}$.

F. Proof of the Inequality $\mathbb{E}_{s_j}\{s_j^2\} \geq \mathbb{E}\{s_j^2\}$

Proof: To prove the inequality, we prove the more general form $\mathbb{E}_{s_j}\{\gamma(s_j)\} \geq \mathbb{E}\{\gamma(s_j)\}$ for any positive valued function γ. We have

$$\mathbb{E}_{s_j}\{\gamma(s_j)\} = \int p_{s_j|\mathbb{1}_{s_j}}(u)\gamma(u)du$$
$$= \frac{\int_{\mathbb{B}} p_{s_j}(u)\gamma(u)du}{\int_{\mathbb{B}} p_{s_j}(u)du}$$
$$= \frac{\int p_{s_j}(u)\gamma(u)du - \int_{\overline{\mathbb{B}}} p_{s_j}(u)\gamma(u)du}{\int_{\mathbb{B}} p_{s_j}(u)du}$$
$$> \int p_{s_j}(u)\gamma(u)du = \mathbb{E}\{\gamma(s_j)\}.$$

These results hold if we impose the condition of (5), since we have from Hölders inequality that $\int_{\overline{\mathbb{B}}} p_{s_j}(u)du \int_{\overline{\mathbb{B}}} \gamma(u)du \geq \int_{\overline{\mathbb{B}}} p_{s_j}(u)\gamma(u)du$ and thus $\int_{\overline{\mathbb{B}}} p_{s_j}(u)du \rightarrow 0 \Rightarrow \int_{\overline{\mathbb{B}}} p_{s_j}(u)\gamma(u)du \rightarrow 0$ for all positive valued functions γ. As a consequence, we have $\mathbb{E}_{s_j}\{\gamma(s_j)\} \geq \mathbb{E}\{\gamma(s_j)\}$ with equality if and only if $\int_{\mathbb{B}} p_{s_j}(u)du = 1$, i.e., $C < \min |s_j|$. Since u^2 is a nonnegative valued function, $C > 0$ and we generally have $\min |s_j[k]| = 0$

(continuous distributions defined on the whole real line), we have $\mathbb{E}_{s_j}\{s_j^2\} > \mathbb{E}\{s_j^2\}$. ■

REFERENCES

[1] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture," *Signal Process.*, vol. 24, no. 1, pp. 1–10, 1991.

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

[3] L. Tong, R.-W. Liu, V. C. Soon, and Y.-F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits Syst.*, vol. 38, no. 5, pp. 499–509, 1991.

[4] S. A. Cruces-Alvarez, A. Cichocki, and S.-i. Amari, "From blind signal extraction to blind instantaneous signal separation: Criteria, algorithms and stability," *IEEE Trans. Neural Netw.*, vol. 15, no. 4, pp. 859–873, 2004.

[5] E. Moreau and P. Comon, "Fonctions de Contraste," in *Séparation de Sources* (in French). Paris, France: Hermès-Lavoisier, 2007, vol. 1, pp. 75–115.

[6] A. Cichocki, R. Thawonmas, and S.-I. Amari, "Sequential blind signal extraction in order specified by stochastic properties," *Electron. Lett.*, vol. 33, no. 1, pp. 64–65, 1997.

[7] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Process.*, vol. 45, pp. 59–83, 1995.

[8] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, no. 4–5, pp. 411–430, 2000.

[9] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 3017–3030, 1996.

[10] W. Lu and J. C. Rajapakse, "ICA with reference," in *Proc. Int. Conf. ICA BSS*, 2001, pp. 120–125.

[11] W. Lu and J. C. Rajapakse, "Approach and applications of constrained ica," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 203–212, 2005.

[12] B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, and R. C. Goodlin, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, no. 12, pp. 1692–1716, 1975.

[13] M. Sato, Y. Kimura, S. Chida, T. Ito, N. Katayama, K. Okamura, and M. Nakao, "A novel extraction method of fetal electrocardiogram from the composite abdominal signal," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 1, pp. 49–58, 2007.

[14] A. Adib, E. Moreau, and D. Aboutajdine, "Source separation contrasts using a reference signal," *IEEE Signal Process. Lett.*, vol. 11, no. 3, pp. 312–315, 2004.

[15] M. Castella, S. Rhioui, E. Moreau, and J. C. Pesquet, "Quadratic higher order criteria for iterative blind separation of a mimo convolutive mixture of sources," *IEEE Trans. Signal Process.*, vol. 55, pp. 218–232, Jan. 2007.

[16] B. Xerri and B. Borloz, "An iterative method using conditional second-order statistics applied to the blind source separation problem," *IEEE Trans. Signal Process.*, vol. 52, pp. 313–328, Feb. 2004.

[17] D.-T. Pham and P. Garat, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1712–1725, 1997.

[18] J.-F. Cardoso, "Infomax and maximum likelihood for blind source separation," *IEEE Signal Process. Lett.*, vol. 4, no. 4, pp. 112–114, 1997.

[19] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[20] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley-Interscience, 2001.

[21] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, 5th ed. New York: Wiley, 2002.

[22] *Independent Component Analysis: Principles and Practice*, S. Roberts and R. Everson, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[23] J.-F. Cardoso, "On the performance of orthogonal source separation algorithms," in *Proc. EUSIPCO*, Edinburgh, 1994, pp. 776–779.

[24] C. J. James and O. Gibson, "ICA with a reference: Extracting desired electromagnetic brain signals," *Med. Applicat. Signal Process.*, 2002.

[25] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[26] R. Gribonval, "Piecewise linear source separation," in *Proc. SPIE'03*, San Diego, CA, 2003, vol. 5207, ser. Wavelets: Applications in Signal and Image Processing.

[27] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, pp. 4311–4322, Nov. 2006.

[28] C. J. James and O. J. Gibson, "Temporally constrained ICA: An application to artifact rejection in electromagnetic brain signal analysis," *IEEE Trans. Biomed. Eng.*, vol. 30, no. 9, pp. 1108–1115, 2003.

[29] T. Netabayashi, Y. Kimura, S. Chida, T. Ito, K. Ohwada, N. Katayama, K. Okamura, and M. Nakao, "Robustness of the blind source separation with reference against uncertainties of the reference signals," in *Proc. 30th Annu. Int. IEEE EMBS Conf.*, Vancouver, BC, Canada, 2008, vol. 30, pp. 1875–1878.

[30] R. Phlypo, V. Zarzoso, and I. Lemahieu, "Eigenvector analysis for separation of a spectrally concentrated source from a mixture," in *Proc. 30th Annu. Int. IEEE EMBS Conf.*, Vancouver, BC, Canada, 2008, pp. 1863–1866.

Vicente Zarzoso (S'94–M'03) was born in Valencia, Spain, in 1973. He graduated (with highest distinction) in telecommunications engineering from the Universidad Politécnica de Valencia, Spain, in 1996. He received the Ph.D. degree from the University of Liverpool, U.K., in 1999.

He received a scholarship from the University of Strathclyde, Glasgow, U.K., and the Defence Evaluation and Research Agency, U.K. He spent five years with the University of Liverpool under a Research Fellowship from the Royal Academy of Engineering, U.K. Since September 2005, he has been a Lecturer with the Université de Nice-Sophia Antipolis and a Researcher with the Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis, France. His research interests include blind statistical signal and array processing and its application to biomedical problems and communications.

Ronald Phlypo (S'06–M'10) was born in Ostend, Belgium, in 1981. He graduated in industrial engineering from KHBO, Ostend, in 2003 and received the master's degree in artificial intelligence from KULeuven, Leuven, Belgium, in 2004. He is currently pursuing the Ph.D. degree in engineering sciences in the Medical Image and Signal Processing Research Group, University of Ghent, Belgium.

He has been a Visitor with the Laboratoire d'Informatique, Signaux et Systèmes de Sophia Antipolis, France, since 2006. His research interests include blind statistical signal and array processing and its application to biomedical problems.

Ignace Lemahieu (M'92–SM'00) was born in Belgium in 1961. He graduated and received the doctoral degree from Ghent University, Ghent, Belgium, in 1983 and 1988, respectively, both in physics.

He joined the Department of Electronics and Information Systems (ELIS), Ghent University, in 1989 as a Research Associate with the Fund for Scientific Research. He is now a Professor of medical image and signal processing and head of the Medical Image and Signal Processing Research Group. His research interests comprise all aspects of image processing and biomedical signal processing, including image reconstruction from projections, pattern recognition, image fusion, and compression. He is a coauthor of more than 200 papers.

Dr. Lemahieu is a member of SPIE, the European Society for Engineering and Medicine, and the European Association of Nuclear Medicine.