HIGHER-ORDER BLIND ESTIMATION OF GENERALIZED EIGENFILTERS USING INDEPENDENT COMPONENT ANALYSIS

Toni Huovinen, Ali Shahed hagh ghadam, Mikko Valkama

Tampere University of Technology
Institute of Communications Engineering
P.O.Box 553, FIN-33101, Tampere Finland,
Email: {toni.huovinen,ali.shahed,mikko.e.valkama}@tut.fi

ABSTRACT

Assuming a noisy linear mixing model of source random variables or signals, maximizing the output signal-to-interference-and-noise-ratio (SINR) among linear transformations of observed data leads to solving the generalized eigenvalue problem. The explicit solution of the problem assumes the knowledge of the mixing coefficients and noise variance and, for this reason, is not a blind method as such. However, we show in this paper that the solution can be estimated *blindly* and *directly* using basic independent component analysis (ICA) designed for noise-free linear models. In addition, the theoretical and numerical results of the paper show that one of the most widely applied ICA algorithms, the equivariant adaptive source identification (EASI) algorithm, is, in practice, identical with SINR-maximizing generalized eigenfiltering, even though it does not use explicit knowledge of the mixing coefficients nor source and noise statistics.

***Index Terms*—** Independent Component Analysis, Noisy Linear Models, Generalized Eigenvalue Problem, Adaptive Source Identification

1. INTRODUCTION

Independent component analysis (ICA) [1] is a statistical signal processing technique which has attracted a lot of attention recently. Especially, it has been applied successfully to solving blind source separation (BSS) problem. Typically, a *noise-free* linear mixing model is assumed in derivation of ICA algorithms in the literature [1]. Needless to say, the noise-free model is unrealistic in most of the practical applications. Consequently, applications of ICA often assume a noisy linear model, but exploit one of the ICA algorithms developed for noise-free models. Thus, the presence of reasonable level of additive noise is thought to cause "only" some feasible distortion due to the model mismatch. However, numerical experiments reported in [2] indicate that, although noise can never be suppressed completely by any linear technique, the performance gain in terms of input-output signal-to-interference-and-noise-ratio (SINR) obtained using ICA is practically identical to that of the optimum (i.e., SINR-maximizing) linear transformation utilizing known channel and noise statistics. That paper is, nevertheless, restricted to a certain telecommunications related interference suppression application in which it is enough to assume (and is assumed) rather simple $2 \times N$ ($N \geq 2$) noisy mixing model. Also some other earlier ICA works (see, e.g., [3]) have found out (by numerical experiments) similar results.

In this paper, our main objective is to generalize the experimental findings of [2] to $M \times N$ ($M \geq N$) models *in the mathematically rigorous way*. In particular, the forthcoming analytical study shows that conceptually ICA can indeed identify a linear input-output SINR maximizing transformation *directly*. Recall, that identifying the mixing matrix and inverting it do not lead to the linear SINR maximizing transformation as such due to noise enhancement. We also give conditions under which one of the most widely applied ICA algorithms, the equivariant adaptive source identification (EASI) algorithm [4], is, in theory, identical with SINR-maximizing generalized eigenfiltering [5]. More specifically, we give the necessary and sufficient conditions under which the matrix of the generalized eigenfilters is a stationary point of the EASI algorithm. We also show some numerical results verifying convergence of the EASI algorithm to the SINR-maximizing linear solution and showing that the performance of the EASI algorithm is remarkably close to the optimal (i.e., the maximal output SINR among all linear transforms of observed data) also in cases in which the above mentioned theoretical optimality condition is not met. In addition, we prove that, in the special case of interference-free (that is, noise only) system, the EASI algorithm can attain exactly the maximum diversity gain blindly, i.e., perform as a *blind* maximal ratio combiner (MRC).

2. NOISY ICA MODEL AND OPTIMAL LINEAR FILTERING

A complex valued, linear, $M \times N$ ($M \geq N$) ICA model [1] with additive white Gaussian noise (AWGN) is assumed throughout this paper. Thus, the M-dimensional random observation vector, $\mathbf{x}$, is given as

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \boldsymbol{\eta}, \tag{1}$$

in which $\mathbf{s} = [s_1\ s_2 \ldots s_N]^T$ is a random source vector with complex valued, *mutually independent components* and $\boldsymbol{\eta}$ is a zero mean complex valued Gaussian noise vector (independent from $\mathbf{s}$) with a strictly positive variance σ^2 and the covariance $\mathbb{E}\left\{\boldsymbol{\eta}\boldsymbol{\eta}^H\right\} = \sigma^2\mathbf{I}$. Further, $\mathbf{A} = [\mathbf{h}_1\ \mathbf{h}_2 \ldots \mathbf{h}_N] \in \mathbb{C}^{M \times N}$ is a full rank mixing (or channel) matrix. Without loss of generality (see, [1]), it is assumed that the source components, $s_1, s_2, \ldots$ and s_N, have zero mean and unit variance. Together with the independence assumption, this implies that $\mathbb{E}\left\{\mathbf{s}\mathbf{s}^H\right\} = \mathbf{I}$. In ICA literature, the complex valued sources are usually assumed also to be circularly symmetric at least up to second-order. A complex valued random variable, say z, is said to be second-order circularly symmetric, if $\mathbb{E}\left\{z^2\right\} = 0$ [6]. Circular symmetry of the source components is assumed also here. This assumption yields (again together with independency and zero mean assumptions) that $\mathbb{E}\left\{\mathbf{s}\mathbf{s}^T\right\} = \mathbf{0}$. In the analysis, the second-order circular symmetry property is assumed also for the noise vector $\boldsymbol{\eta}$.

Definition 1 *Let $\mathbf{w} \in \mathbb{C}^M \backslash \{\mathbf{0}\}$ be an arbitrary linear filter and $y_\mathbf{w} = \mathbf{w}^H\mathbf{x}$ the corresponding linear output. Signal-to-interference-and-noise ratio (SINR) wrt. the n-th source component, s_n, at the output $y_\mathbf{w}$ is then defined as*

$$\rho_n(\mathbf{w}) = \frac{\mathbf{w}^H\mathbf{R}_n\mathbf{w}}{\mathbf{w}^H\mathbf{R'}_n\mathbf{w}}, \tag{2}$$

in which

$$\mathbf{R}_n := \mathbb{E}\left\{\mathbf{h}_n s_n s_n^* \mathbf{h}_n^H\right\} = \mathbf{h}_n\mathbf{h}_n^H \tag{3}$$

and

$$\begin{aligned}\mathbf{R'}_n &:= \mathbb{E}\left\{\left(\sum_{k\neq n}\mathbf{h}_k s_k + \boldsymbol{\eta}\right)\left(\sum_{k\neq n}\mathbf{h}_k s_k + \boldsymbol{\eta}\right)^H\right\} \\ &= \sum_{k\neq n}\mathbf{h}_k\mathbf{h}_k^H + \sigma^2\mathbf{I}.\end{aligned} \tag{4}$$

Remark 1 *Since all eigenvalues of the Hermitian matrix $\mathbf{R'}_n$ are greater than or equal to σ^2, the Hermitian form in the denominator of (2) is positive definite, i.e., strictly positive for all $\mathbf{w} \in \mathbb{C}^M\backslash\{\mathbf{0}\}$, provided that $\sigma^2 > 0$. Consequently, (2) is well-defined for all $\mathbf{w} \in \mathbb{C}^M\backslash\{\mathbf{0}\}$.*

Now, as seen in (2), maximizing SINR among all linear transformations of observed data, i.e., maximizing $\rho_n(\mathbf{w})$, equals to solving the generalized eigenvalue problem [5] associated with matrix pair $(\mathbf{R}_n, \mathbf{R'}_n)$. Hence,

$$\max_{\mathbf{w}\in\mathbb{C}^M\backslash\{\mathbf{0}\}} \rho_n(\mathbf{w}) = \lambda_n \tag{5}$$

and

$$\arg\max_{\mathbf{w}\in\mathbb{C}^M\backslash\{\mathbf{0}\}} \rho_n(\mathbf{w}) = \mathbf{e}_n, \tag{6}$$

in which λ_n stands for the greatest eigenvalue of the Hermitian matrix $(\mathbf{R'}_n)^{-1}\mathbf{R}_n$ (the matrix inverse exists, see Remark 1) and $\mathbf{e}_n$ for the corresponding eigenvector. To be specific, since SINR $\rho_n(\mathbf{w})$ is scale invariant, $\mathbf{e}_n$ can be any vector in one-dimensional eigensubspace corresponding to the eigenvalue λ_n.

Also the linear minimum mean square error (LMMSE) estimator of a source can be shown to yield the maximum SINR among linear transformations. This is basically stated in [7] and in references therein. The LMMSE transformation for n-th source in model (1) is essentially given as

$$\mathbf{w}'_n = \mathbf{C_x}^{-1}\mathbf{h}_n, \tag{7}$$

in which $\mathbf{C_x} = \mathbf{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{A}\mathbf{A}^H + \sigma^2\mathbf{I}$ is the observation covariance. This linear transformation, thus, gives an explicit solution to the generalized eigenvalue problem above, i.e., $\mathbf{e}_n = \mathbf{w}'_m$. Nevertheless, the solution assumes the knowledge of the mixing coefficients and noise variance and, for this reason, is not a blind method as such. However, it will be shown that the solution can be estimated blindly using ICA.

Next some further notations used throughout this paper are defined and the most important properties of filters maximizing the linear SINR are given.

Definition 2 (i) *The vector $\mathbf{e}_n$ is called SINR-maximizing generalized eigenfilter (M-GEF) wrt. the source component s_n.* (ii) *The matrix $\mathbf{E} := [\mathbf{e}_1\ \mathbf{e}_2 \ldots \mathbf{e}_N] \in \mathbb{C}^{N\times M}$ is called M-GEF transformation. Recall, that it also has the LMMSE characterization as*

$$\mathbf{E} = \mathbf{C_x}^{-1}\mathbf{A}. \tag{8}$$

(iii) *The filtered output $\mathbf{y} := \mathbf{E}^H\mathbf{x}$ is called M-GEF output.*

Lemma 1 *If the columns of mixing matrix $\mathbf{A}$ are mutually orthogonal, i.e., $\mathbf{A}^H\mathbf{A} = \mathbf{D_A}$ for some diagonal matrix $\mathbf{D_A} \in \mathbb{C}^{N\times N}$, then the M-GEF transformation $\mathbf{E}$ has mutually orthogonal columns and, moreover, $\mathbf{E} = \mathbf{A}\mathbf{D}_E$ for some diagonal matrix $\mathbf{D_E} \in \mathbb{C}^{N\times N}$.*

Proof. Let the columns of $\mathbf{A} = [\mathbf{h}_1\ \mathbf{h}_2 \ldots \mathbf{h}_N]$ be orthogonal. Now, it is enough to prove that the mixing coefficient vector $\mathbf{h}_n$ is the M-GEF vector wrt. the source component $\mathbf{s}_n$ for all $n \in \{1, 2, \ldots, N\}$. First, due to orthogonality of the columns of $\mathbf{A}$, $\mathbf{R}_k\mathbf{h}_n = 0$ for all $k \neq n$, which implies that $\mathbf{R'}_n\mathbf{h}_n = \sigma^2\mathbf{h}_n$. Consequently,

$$\mathbf{R}_n\mathbf{h}_n = \|\mathbf{h}_n\|^2\mathbf{h}_n = \frac{\|\mathbf{h}_n\|^2}{\sigma^2}\mathbf{R'}_n\mathbf{h}_n, \tag{9}$$

i.e., $\mathbf{h}_n$ is *an* eigenvector of $(\mathbf{R'}_n)^{-1}\mathbf{R}_n =: \mathbf{M}_n$ and the corresponding eigenvalue is $\|\mathbf{h}_n\|^2/\sigma^2 =: \tilde{\lambda}_n$.

Second, it is still needed to show that $\tilde{\lambda}_n$ is *the greatest* eigenvalue of $\mathbf{M}_n$, or equally, that $\tilde{\lambda}_n$ really is the maximum value of the linear output SINR $\rho_n(\mathbf{w})$ in $\mathbb{C}^M\backslash\{\mathbf{0}\}$. But, using Schwarz's inequality, it is seen that, for all $\mathbf{w} \in \mathbb{C}^M\backslash\{\mathbf{0}\}$,

$$\rho_n(\mathbf{w}) \leq \frac{\|\mathbf{h}_n\|^2}{\sigma^2}. \tag{10}$$

Hence, this concludes the proof. ∎

Lemma 2 *The following assertions are equivalent:*

1. *The matrix $\mathbf{A} = [\mathbf{h}_1\ \mathbf{h}_2 \ldots \mathbf{h}_N]$ has orthogonal columns.*
2. *$\mathbf{E}^H\mathbf{A} = \mathbf{D} := \mathrm{diag}(d_1, d_2, \ldots, d_N)$ for some complex numbers $d_1, d_2, \ldots, d_N$.*
3. *The M-GEF output $\mathbf{y} = [y_1\ y_2 \ldots y_N]$ has uncorrelated components, i.e., $\mathbb{E}\left\{y_i y_j^*\right\} = 0$ for all $i \neq j$.*

Proof. First of all, notice that the complex number d_n in the second assertion can be written as $d_n = \mathbf{e}_n^H\mathbf{h}_n$ for any $n \in \{1, 2, \ldots, N\}$. For this reason, d_n is necessarily non-zero for any n, since otherwise λ_n would be zero. Now, (i) The first assertion implies the second one according to Lemma 1. (ii) Also the inverse is correct, since assuming that the second assertion is true and using the definition of M-GEF filter $\mathbf{e}_n$ we have for all n, that

$$d_n^*\mathbf{h}_n = \mathbf{R}_n\mathbf{e}_n = \lambda_n\mathbf{R'}_n\mathbf{e}_n = \lambda_n\sigma^2\mathbf{e}_n. \tag{11}$$

Hence, $\mathbf{A} = \mathbf{E}\widetilde{\mathbf{D}}$ with appropriate diagonal matrix $\widetilde{\mathbf{D}}$ and, further, $\mathbf{A}^H\mathbf{A} = \widetilde{\mathbf{D}}^H\mathbf{E}^H\mathbf{A} = \widetilde{\mathbf{D}}^H\mathbf{D}$. Finally, (iii) the second and the third assertions are equivalent, since the covariance of the observation is

$$\mathbb{E}\left\{\mathbf{x}\mathbf{x}^H\right\} = \mathbf{A}\mathbf{A}^H + \sigma^2\mathbf{I} = \sum_{k=1}^{M}\mathbf{R}_k + \sigma^2\mathbf{I} = \mathbf{R}_n + \mathbf{R'}_n \tag{12}$$

for any n. Hence, the correlation between the i-th and j-th $(i \neq j)$ output components, $\mathbb{E}\left\{y_i y_j^*\right\}$, reduces now as follows:

$$\mathbb{E}\left\{y_i y_j^*\right\} = \mathbf{e}_i^H\left(\mathbf{R}_j + \mathbf{R'}_j\right)\mathbf{e}_j = (1 + \frac{1}{\lambda_j})\mathbf{e}_i^H\mathbf{R}_j\mathbf{e}_j. \tag{13}$$

Here, λ_j is *strictly* positive by definition, thus, the correlation is zero if and only if

$$0 = \mathbf{e}_i^H \mathbf{R}_j \mathbf{e}_j = \mathbf{e}_i^H \mathbf{h}_j \mathbf{h}_j^H \mathbf{e}_j \quad (14)$$

Therefore and because the mixing matrix $\mathbf{A} = [\mathbf{h}_1\ \mathbf{h}_2 \ldots \mathbf{h}_N]$ is assumed to be full rank, all the output components are mutually uncorrelated if and only if $\mathbf{E}^H \mathbf{A} = \mathbf{D}$. ■

A conventional diversity combining, or maximal ratio combining (MRC) as often referred to, is closely related to M-GEF. Nevertheless, in MRC, the main objective is to maximize the signal-to-(AWG)-noise ratio (SNR) and, consequently, MRC ignores possible interfering source components which results in a suboptimum SINR performance, in general. However, M-GEF wrt., say, the n-th source component is consistent with MRC (wrt. the same component) in the sense that, without interfering source components, $s_k,\ k \neq n$, these two methods coincide. This is easy to see by setting interfering sources to zero and constraining $\mathbf{w}$ to have, say, a unit norm in which case (2) reduces to

$$\tilde{\rho}_n(\mathbf{w}) = \frac{\mathbf{w}^H \mathbf{R}_n \mathbf{w}}{\sigma^2}. \quad (15)$$

Thus now, maximizing (15) is an ordinary eigenvalue problem, which, for one, is well known to yield the MRC solution [8], i.e.,

$$\arg\max_{\mathbf{w}\in\mathbb{C}^M} \tilde{\rho}_n(\mathbf{w}) = c\mathbf{h}_n \text{ for any } c \in \mathbb{C}. \quad (16)$$

An important difference between MRC and M-GEF is, nevertheless, the order of observation statistics needed to solve the problems blindly. Estimation of the covariance of the observation vector $\mathbf{x}$, $\mathbf{C}_\mathbf{x} = \mathbb{E}\left\{\mathbf{x}\mathbf{x}^H\right\} = \mathbf{R}_n + \sigma^T \mathbf{I}$ (in noise-only system), is sufficient to solve the MRC problem, since

$$\arg\max_{\mathbf{w}\in\mathbb{C}^M} \tilde{\rho}_n(\mathbf{w}) = \arg\max_{\mathbf{w}\in\mathbb{C}^\mathbf{M}} \mathbf{w}^\mathbf{H} \mathbf{C}_\mathbf{x} \mathbf{w}. \quad (17)$$

Hence, the MRC solution is obtainable blindly using only second-order statistics of $\mathbf{x}$. Also SINR is a second-order measure of output's "goodness" under general model (1), or more precisely, M-GEF depends only on second-order statistics of the contribution of the desired source, $\mathbf{R}_n$, and second-order statistics of the contribution of interference and noise, $\mathbf{R}'_n$, in the observation $\mathbf{x}$. However, $\mathbf{R}_n$ and $\mathbf{R}'_n$ can not be separated blindly from second-order statistics of the observation ($\mathbf{C}_\mathbf{x} = \mathbf{R}_n + \mathbf{R}'_n$ for all n), but both are, anyway, needed separately when solving M-GEF problem. This suggests that blind linear output SINR maximization is rather a generalization of higher-order (noise-free) blind source separation than of MRC.

3. INDEPENDENT COMPONENT ANALYSIS

In basic ICA, the goal is essentially to invert the model (1) blindly, that is, to find a demixing matrix $\mathbf{B} \in \mathbb{C}^{N\times M}$ such that $\mathbf{BA}$ is as close to identity as possible by using only the observations $\mathbf{x}$. Because of the blindness, a solution of the ICA problem, $\mathbf{B}$, can be unique only up to left multiplication by an arbitrary permutation and diagonal matrices. Blind identifiability of such a $\mathbf{B}$ is proved for the noise-free ICA model in [9] and for the noisy model in [10]. Nevertheless, in general, transforming the observation by inverse of $\mathbf{A}$ does not lead, as such, to the best linear SINR gain (i.e., to M-GEF solution) in a noisy system due to arbitrary noise amplification. (This is also seen in Lemma 2.) Naturally, a linear SINR maximizing transformation can be constructed as the LMMSE matrix (8) after the identification of $\mathbf{A}$ given that the observation covariance, $\mathbf{C}_\mathbf{x}$, is also estimated. However, experimental results in [2] suggest that some ICA algorithms developed for noise-free models are able to provide *directly* input-output SINR gains very close to the best linear gain possible, in particular, the gains clearly better than with inverse transform of $\mathbf{A}$.

Good performance under noise can be basically explained by whitening which is accomplished in typical ICA algorithms, that is, the observed signal, $\mathbf{x}$, is first transformed linearly to vector $\mathbf{z} = \mathbf{Vx}$ such that $\mathbb{E}\left\{\mathbf{z}\mathbf{z}^H\right\} = \mathbf{I}$ [1]. In some algorithms (in EASI algorithm, for instance), linear whitening is performed implicitly during the actual ICA separation procedure [4]. The whitening transformation, $\mathbf{V}$, is not unique, however, one popular way is to use $\mathbf{V} = \mathbf{C}_\mathbf{x}^{-\frac{1}{2}}$. The whitening changes the ICA model (1) into

$$\mathbf{z}(t) = \mathbf{Vx} = \mathbf{VAs}(t) + \mathbf{V}\boldsymbol{\eta}(t).$$

Hence, the new mixing matrix is $\hat{\mathbf{A}} := \mathbf{VA}$. Now, since covariance of $\mathbf{z}$ equals to the identity matrix ($\mathbf{C}_\mathbf{z} = \mathbf{I}$), $\hat{\mathbf{A}}$ is exactly the LMMSE matrix, $\mathbf{C}_\mathbf{z}^{-1}\hat{\mathbf{A}}$, for $\mathbf{z}$. Consequently, after whitening, ICA algorithms using an optimization criterion that is invariant to additive Gaussian noise [1] actually estimate the M-GEF transformation (or inverse of it) directly. This is not well-understood in the literature.

Majority of the well known ICA algorithms assume, that the mixing matrix after whitening is orthogonal (or unitary in the complex valued case), which is valid, in general, only if the linear model is noise-free. In the noisy model (1), this assumption is not true, or in other words, the matrix $\hat{\mathbf{A}}$ is not orthogonal (unitary) in general. For this reason, the algorithms using the orthogonality constraint can not attain the M-GEF solution exactly in theory, but one can think that they tend to produce an orthogonalized estimate of the M-GEF (or LMMSE) transformation. We demonstrate this in the following, by considering the widely applied EASI algorithm as an example of the algorithms that are originally intended to perform noise-free ICA and have the orthogonality constraint. Notice that there exists also ICA algorithms that do not use the orthogonality constraint (see, e.g., [11]).

4. EASI ALGORITHM AND NOISY MODEL

EASI algorithm is a recursive online algorithm which operates on individual samples of observed data. One recursion step of the EASI algorithm, i.e., of searching the demixing matrix $\mathbf{B} \in \mathbb{C}^{N\times M}$, is given as

$$\mathbf{B}_{t+1} = \mathbf{B}_t - \mu \mathbf{U}_t(\mathbf{y}_t)\mathbf{B}_t, \quad (18)$$

in which μ is a scalar step size and the update matrix, $\mathbf{U}_t(\mathbf{y}_t) \in \mathbb{C}^{N\times N}$, is defined as

$$\mathbf{U}_t(\mathbf{y}_t) = \mathbf{y}_t\mathbf{y}_t^H - \mathbf{I} + \mathbf{g}(\mathbf{y}_t)\mathbf{y}_t^H - \mathbf{y}_t\mathbf{g}(\mathbf{y}_t)^H. \quad (19)$$

Here $\mathbf{y}_t = \mathbf{B}_t\mathbf{x}_t$, $\mathbf{I}$ stands for identity matrix and $\mathbf{g} : \mathbb{C}^N \to \mathbb{C}^N$ is an arbitrary nonlinear function. On right-hand side of (19), two first terms tends to whiten the output $\mathbf{y}_t$, thus, the algorithm uses implicitly the orthogonality constraint discussed in the previous section. Since only the current sample is used in each step of the algorithm, the update matrix (19) does not vanish asymptotically. Instead, a stationary point of the algorithm is defined stochastically as follows: [4]

Definition 3 *An $N \times M$ matrix $\mathbf{B}'$ is a stationary point of the EASI algorithm if the expected value of the update term, $\mathbf{U}(\mathbf{y})$, is zero, i.e.,*

$$\mathbb{E}\left\{\mathbf{y}\mathbf{y}^H - \mathbf{I} + \mathbf{g}(\mathbf{y})\mathbf{y}^H - \mathbf{y}\mathbf{g}(\mathbf{y})^H\right\} = \mathbf{0}, \quad (20)$$

for $\mathbf{y} = \mathbf{B}'\mathbf{x}$.

The convergence of the EASI algorithm is given as stability of the stationary point in noise-free model [4]. In other words, at least local convergence is guaranteed in theory. In the following theoretical analysis, we do not pay attention to the issue of convergence but solely show that the EASI stationary point is very closely related to the M-GEF solution. The convergence is verified with numerical experiments in the section 5.

4.1. Blind Maximal Ratio Combining

It was seen in the section 2 that, in the case of an *interference-free* linear model (i.e., the model with one source and noise only), *blind* maximization of the linear output SNR (or equally SINR) can be carried out using second-order observation statistics. In this section, we show that also the EASI algorithm can maximize SNR in interference-free system. Of course, using a higher-order statistical method to solve the second-order statistical problem is not advisable in practise, if one knows beforehand that the system is interference-free. Sometimes the possible absence of interfering source components is not known in which case using the higher-order ICA method can be reasonable. The following result is also a natural starting point to the more general analysis in the next section.

Strictly speaking, we prove here that, in the particular case of the interference-free model,

$$\tilde{\mathbf{x}} = \mathbf{h}s + \boldsymbol{\eta}, \tag{21}$$

the MRC filter, which thus maximizes linear output SNR, is related to stationary point of EASI algorithm provided that a nonlinearity $\mathbf{g} = [g_1\ g_2 \dots g_N]^T : \mathbb{C}^N \to \mathbb{C}^N$;

$$g_n(\mathbf{z}) = |z_n|^2 z_n, n = 1, \dots, N, \tag{22}$$

is used. This is formulated rigorously in the following proposition.

Proposition 1 *Let $\widetilde{N} \in \{1, 2, \dots, M\}$ be arbitrary and assume the model (21). In addition, let $\{\tilde{\mathbf{e}}_1, \tilde{\mathbf{e}}_2, \dots, \tilde{\mathbf{e}}_{\widetilde{N}}\}$ be a set of orthonormal vectors in $\mathbb{C}^M \backslash \{\mathbf{0}\}$ such that $\tilde{\mathbf{e}}_1$ equals to normalized MRC filter, i.e., $\tilde{\mathbf{e}}_1 = \mathbf{h}/\|\mathbf{h}\|$. Further, let $\mathbf{g} : \mathbb{C}^{\widetilde{N}} \to \mathbb{C}^{\widetilde{N}}$ be defined as in (22) and $\widetilde{\mathbf{E}} := [\tilde{\mathbf{e}}_1\ \tilde{\mathbf{e}}_2\ \dots \tilde{\mathbf{e}}_{\widetilde{N}}]$. Now, $\exists\ \boldsymbol{\alpha} = [\alpha_1\ \alpha_2\ \dots\ \alpha_{\widetilde{N}}]^T \in (\mathbb{C}\backslash\{0\})^{\widetilde{N}}$ such that*

$$\widetilde{\mathbf{B}}_{\boldsymbol{\alpha}} := \operatorname{diag}(\alpha_1, \alpha_2, \dots, \alpha_{\widetilde{N}})\widetilde{\mathbf{E}}^H \tag{23}$$

is a stationary point of the EASI algorithm, i.e.,

$$\mathbb{E}\left\{\tilde{\mathbf{y}}_{\boldsymbol{\alpha}}\tilde{\mathbf{y}}_{\boldsymbol{\alpha}}^H - \mathbf{I} + \mathbf{g}(\tilde{\mathbf{y}}_{\boldsymbol{\alpha}})\tilde{\mathbf{y}}_{\boldsymbol{\alpha}}^H - \tilde{\mathbf{y}}_{\boldsymbol{\alpha}}\mathbf{g}(\mathbf{y}_{\boldsymbol{\alpha}})^H\right\} = \mathbf{0}, \tag{24}$$

in which $\tilde{\mathbf{y}}_{\boldsymbol{\alpha}} := \widetilde{\mathbf{B}}_{\boldsymbol{\alpha}}\tilde{\mathbf{x}}$.

Proof. To begin with, let us prove that the higher-order term of (24),

$$\widetilde{\mathbf{H}}_{\boldsymbol{\alpha}} := \mathbb{E}\left\{\mathbf{g}(\tilde{\mathbf{y}}_{\boldsymbol{\alpha}})\tilde{\mathbf{y}}_{\boldsymbol{\alpha}}^H - \tilde{\mathbf{y}}_{\boldsymbol{\alpha}}\mathbf{g}(\tilde{\mathbf{y}}_{\boldsymbol{\alpha}})^H\right\} \in \mathbb{C}^{\widetilde{N}\times\widetilde{N}}, \tag{25}$$

vanishes for all $\boldsymbol{\alpha} \in (\mathbb{C}\backslash\{0\})^{\widetilde{N}}$. First, diagonal elements of $\widetilde{\mathbf{H}}_{\boldsymbol{\alpha}}$ are zeros due to skew-symmetry. Second, the (i,j)-th ($i \neq j$) off-diagonal element of $\widetilde{\mathbf{H}}_{\boldsymbol{\alpha}}$ is given as

$$\left(\widetilde{\mathbf{H}}_{\boldsymbol{\alpha}}\right)_{i,j} = \begin{cases} \mathbb{E}\left\{\tilde{y}_i\tilde{y}_j^*(|\tilde{y}_i|^2 - |\tilde{y}_j|^2)\right\} & \text{if } i < j \\ \mathbb{E}\left\{\tilde{y}_i^*\tilde{y}_j(|y_j|^2 - |\tilde{y}_i|^2)\right\} & \text{if } i > j \end{cases} \tag{26}$$

in which $\tilde{y}_n$ is the n-th component of $\tilde{\mathbf{y}}_{\boldsymbol{\alpha}}$. Hence, it is enough to show that the off-diagonal elements are zeros for all $i < j$. Now, a straightforward calculation shows that $\mathbb{E}\left\{\tilde{y}_i\tilde{y}_j^*|\tilde{y}_i|^2\right\} = 0$ and $\mathbb{E}\left\{\tilde{y}_i\tilde{y}_j^*|\tilde{y}_j|^2\right\} = 0$ for all $\boldsymbol{\alpha} \in (\mathbb{C}\backslash\{0\})^N$. Therefore, the off-diagonal terms and, consequently, the entire higher-order term $\widetilde{\mathbf{H}}_{\boldsymbol{\alpha}}$ vanishes for all $\boldsymbol{\alpha} \in (\mathbb{C}\backslash\{0\})^{\widetilde{N}}$.

To finish the proof, we need that also the second-order term of (24), $\mathbb{E}\left\{\tilde{\mathbf{y}}_{\boldsymbol{\alpha}}\tilde{\mathbf{y}}_{\boldsymbol{\alpha}}^H - \mathbf{I}\right\}$, vanishes for appropriate selection of the vector $\boldsymbol{\alpha} \in (\mathbb{C}\backslash\{0\})^{\widetilde{N}}$. Now, denoting $\mathbf{D}_{\boldsymbol{\alpha}} = \operatorname{diag}(\alpha_1,\ \alpha_2, \dots, \alpha_{\widetilde{N}})$, the covariance term is

$$\begin{aligned} \mathbb{E}\left\{\tilde{\mathbf{y}}_{\boldsymbol{\alpha}}\tilde{\mathbf{y}}_{\boldsymbol{\alpha}}^H\right\} &= \widetilde{\mathbf{B}}_{\boldsymbol{\alpha}}\mathbf{h}\mathbf{h}^H\widetilde{\mathbf{B}}_{\boldsymbol{\alpha}}^H + \sigma^2\widetilde{\mathbf{B}}_{\boldsymbol{\alpha}}\widetilde{\mathbf{B}}_{\boldsymbol{\alpha}}^H \\ &= \mathbf{D}_{\boldsymbol{\alpha}}\widetilde{\mathbf{E}}^H\mathbf{h}\mathbf{h}^H\widetilde{\mathbf{E}}\mathbf{D}_{\boldsymbol{\alpha}}^H + \sigma^2\mathbf{D}_{\boldsymbol{\alpha}}\widetilde{\mathbf{E}}^H\widetilde{\mathbf{E}}\mathbf{D}_{\boldsymbol{\alpha}}^H \\ &= \operatorname{diag}\left(|\alpha_1|^2(\|\mathbf{h}\|^2 + \sigma^2),\ \sigma^2|\alpha_2|^2, \dots, \sigma^2|\alpha_N|^2\right). \end{aligned} \tag{27}$$

From which we see that the components of $\tilde{\mathbf{y}}_{\boldsymbol{\alpha}}$ are uncorrelated for all $\boldsymbol{\alpha} \in (\mathbb{C}\backslash\{0\})^{\widetilde{N}}$ and, surely, it is possible to choose $\alpha_1, \alpha_2, \dots$ and $\alpha_{\widetilde{N}} \in \mathbb{C}$ such that the covariance matrix equals unity. ■

Remark 2 (i) *In Prop. 1, the selection that the vector $\tilde{\mathbf{e}}_1$ is the MRC filter, is fully technical. Selecting any vector $\tilde{\mathbf{e}}_n,\ n \in \{1, 2, \dots, \widetilde{N}\}$, to be the MRC filter leads to the same conclusion. This is, of course, consistent with permutation unambiguity of the EASI algorithm.*

(ii) *The output dimension, $\widetilde{N}$, of transformation $\widetilde{\mathbf{E}}$ can be chosen freely (as far as $\widetilde{N} \leq M$). Consequently, the use of the higher-order EASI algorithm to solve the second-order MRC problem is sensible also in practise, if it is not known beforehand wether the system is interference-free.*

4.2. Blind SINR Maximization

The numerical experiments in the next section and also in [2] show that the EASI algorithm provides almost identical SINR performance with M-GEF in practice also under the general model (1). Unfortunately, in theory, the performances are not exactly equal in general, which is seen as follows. Since the higher-order term,

$$\mathbf{g}(\mathbf{y})\mathbf{y}^H - \mathbf{y}\mathbf{g}(\mathbf{y})^H, \tag{28}$$

in the EASI algorithm (19) is skew-Hermitian, the output, $\hat{\mathbf{y}}$, corresponding to any stationary point of the algorithm is necessarily white, i.e., $\mathbb{E}\left\{\hat{\mathbf{y}}\hat{\mathbf{y}}^H\right\} = \mathbf{I}$, due to the structure of the second-order term, $\mathbf{y}\mathbf{y}^H - \mathbf{I}$. On the other hand, according to Lemma 2, the M-GEF output, $\mathbf{y}$, has correlated components unless the mixing matrix $\mathbf{A}$ has orthogonal columns. Therefore, the M-GEF transformation, $\mathbf{E}$, can not be, in general, exactly a stationary point of the EASI algorithm.

It is, however, interesting to notice that the M-GEF transformation is, indeed, a stationary point of the EASI algorithm, if the mixing matrix has orthogonal columns. This is stated in the following proposition.

Proposition 2 *Assume the model (1) with a mixing matrix $\mathbf{A}$ for which $\mathbf{A}^H\mathbf{A} = \mathbf{D}_{\mathbf{A}}$ for some diagonal matrix $\mathbf{D}_{\mathbf{A}} \in \mathbb{C}^{N\times N}$ and let the nonlinearity $\mathbf{g}$ be defined again as in (22). Now, $\exists\ \boldsymbol{\alpha} = [\alpha_1\ \alpha_2 \dots \alpha_N]^T \in (\mathbb{C}\backslash\{0\})^N$ such that*

$$\mathbf{B}_{\boldsymbol{\alpha}} := \operatorname{diag}(\alpha_1, \alpha_2, \dots, \alpha_N)\mathbf{E}^H \tag{29}$$

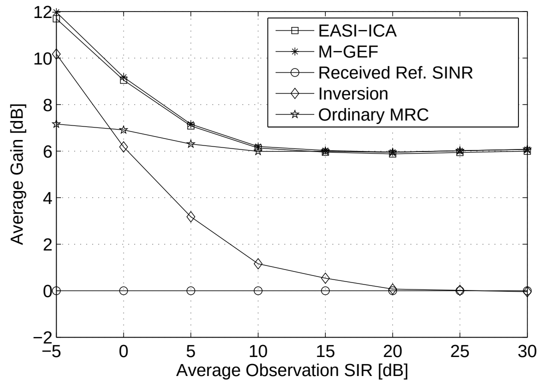

Fig. 1. Output SINR gains as a function of SIR wrt. the strongest component. $N = 4$ source components as well as $M = 4$ observations are used. Observed SNR is fixed to 5 dB.

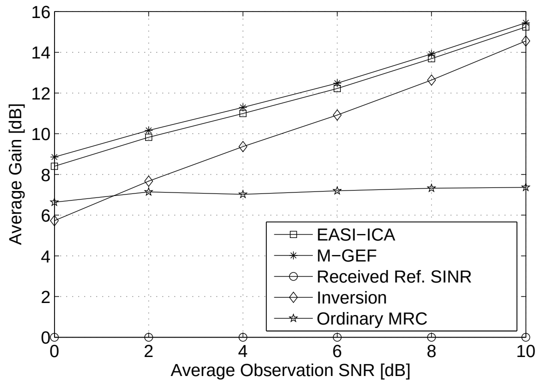

Fig. 2. Output SINR gains as a function of observed SNR. All $N = 4$ sources have equal observed power (i.e., SIR ≈ -4.8 dB). Number of observations is $M = 4$.

is a stationary point of the EASI algorithm, i.e.,

$$\mathbb{E}\left\{\mathbf{y}_{\boldsymbol{\alpha}}\mathbf{y}_{\boldsymbol{\alpha}}^H - \mathbf{I} + \mathbf{g}(\mathbf{y}_{\boldsymbol{\alpha}})\mathbf{y}_{\boldsymbol{\alpha}}^H - \mathbf{y}_{\boldsymbol{\alpha}}\mathbf{g}(\mathbf{y}_{\boldsymbol{\alpha}})^H\right\} = \mathbf{0}, \quad (30)$$

in which $\mathbf{y}_{\boldsymbol{\alpha}} := \mathbf{B}_{\boldsymbol{\alpha}}\mathbf{x}$.

Proof. In essence, the proof is similar to the proof of Prop. 1. Namely, Lemma 1 implies that, for all $n \in \{1, 2, \ldots, N\}$, $\mathbf{e}_n^H\mathbf{A}\mathbf{s} = c_n s_n$ for some complex constant c_n, from which it follows easily that the higher-order term in (30) vanishes for all selections of coefficient vector $\boldsymbol{\alpha} \in (\mathbb{C}\backslash\{0\})^N$. In addition, it is also straightforward to see that it is possible to choose such a vector $\boldsymbol{\alpha} \in (\mathbb{C}\backslash\{0\})^N$ that the covariance of $\mathbf{y}_{\boldsymbol{\alpha}}$ equals to unity. (The covariance is diagonal for all the selections according to Lemma 2.) ■

Remark 3 *The Hermitian transpose of the M-GEF transformation with appropriate scale of columns is* exactly *a stationary point of the EASI algorithm* if and only if *the mixing matrix* $\mathbf{A}$ *has orthogonal columns (i.e., orthogonality is the* necessary and sufficient *condition). This is a direct consequence of Prop. 2 and Lemma 2.*

5. NUMERICAL EXPERIMENTS

Numerical results in this section set against the performance of EASI algorithm and SINR-maximizing M-GEF approach under noisy environment. Also SINR performances of ordinary maximal ratio combining (MRC) and inversion of the mixing matrix, the matrix $\mathbf{A}$ in model (1), are simulated in the experiments. Both of the latter methods are, thus, suboptimum since both interfering source components and additive noise are present. Here, signal-to-noise ratio (SNR) wrt. n-th source signal is defined as the average ratio of the power of n-th source signal's contribution and additive noise power in the *observed* signals. The signal-to-interference ratio (SIR) wrt. n-th source signal, in turn, is the average ratio of the power of n-th source signal's contribution and the powers of other components' contribution in *observed* signals. Given the power normalization of the formal sources stated below (1), the SIR values other than 0 dB are implemented by appropriate scaling of the mixing coefficients.

In the experiments, four QPSK sources ($N = 4$) and four mixtures ($M = 4$) of them are used. The selection of source constellations is more or less arbitrary, and it should not affect the general validity of the results. Mixing coefficients, i.e., elements of the matrix $\mathbf{A}$, are drawn randomly from zero mean Gaussian distributions (one distribution for each source component) for each processing block of $N = 50000$ symbols of data. Variances of these distributions are selected such that observed SNR and SIR values correspond to given values on average. M-GEF bound is evaluated directly from the data model for each block. Hence, the bounds are not affected by finite sample statistics and, more importantly, they are the absolute upper bounds among all linear transformations of received data in case of each realization. Also output SINR's of the MRC and inversion of $\mathbf{A}$ are evaluated from the model.

The third-order nonlinearity (22) is used in the EASI algorithm in all the following experiments. A permutation ambiguity of EASI outputs is circumvented by, first, evaluating the output SINR wrt. all the source components for all EASI outputs and, then, selecting the maximum ones. Practical ways to identify the output components are not considered in this paper. All the gains plotted are average gains wrt. the received SINR over one thousand mixture realization.

In Fig. 1, it is assumed that one of the sources is the desired one and the other three are interfering ones. The figure shows SINR gains for the desired source as a function of SIR with fixed observed SNR. The figure illustrates that the gain of EASI algorithm is nearly identical with M-GEF bound. A difference is roughly 0.25 dB-unit for low SIR values and tends to zero as SIR increases. This is well-consistent with the theoretic results, and especially with Prop. 1, since asymptotically (i.e., as SIR $\to \infty$) the system tends to noise-only system. Notice also that asymptotic SINR gain for the EASI algorithm is roughly 6 dB ($\approx 10\log_{10}(4)$ dB) which equals to theoretic MRC gain with four observations.

Figs. 2 and 3 give two examples of SINR gain vs. observed SNR with fixed received SIR. In the former figure, all the source components have equal observed power on average, which results in approximately SIR of -4.8 dB. Performance is plotted only wrt. one source component, since SINR gains wrt. the other components are, naturally, similar in this case. The latter figure gives an example of system with unequal observed average powers between the sources. Observed average powers of the three weakest sources are -15, -10, and -5 dB wrt. the strongest source. Both figures indicate that the performance behavior of EASI is almost identical with the M-GEF bound in practice, although, in theory, EASI is not able to attain exactly the bound in general.

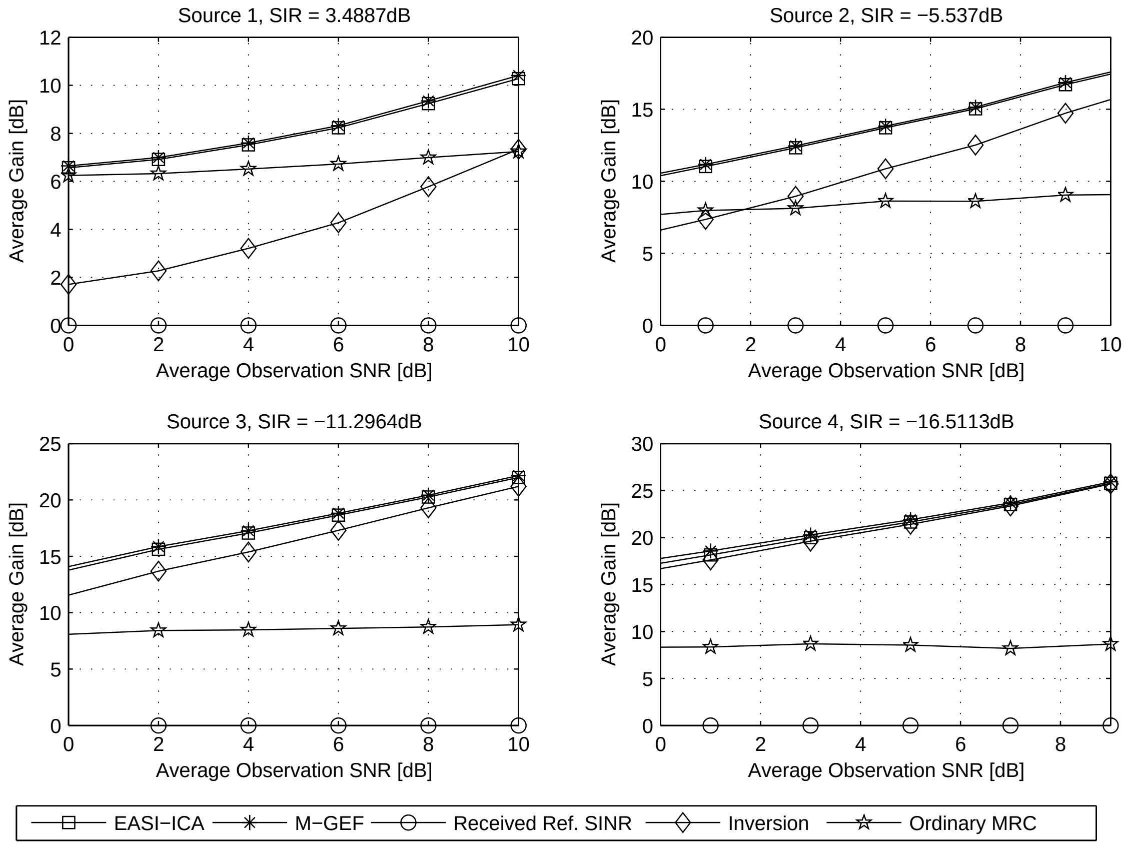

Fig. 3. Output SINR gains as a function of SNR. The $N = 4$ sources have unequal observed power. Observation SIR's wrt. each source are given above the plots. Number of observations is $M = 4$.

6. CONCLUSIONS

In this paper, we illustrated that basic independent component analysis (ICA) designed for noise-free linear models is able to solve, blindly and directly, the generalized eigenvalue problem, i.e., to provide essentially the best possible output SINR among all linear transformations of observed data, in the challenging case of having both additive noise and interference disturbing the desired signal observation. However, ICA algorithms constraining the estimated demixing matrix to be orthogonal (or unitary) can not exactly attain the optimal solution in general, but in a sense they produce an orthogonalized version of the solution. In addition, the theoretical and numerical results of the paper showed that one of the most widely applied ICA algorithms, the equivariant adaptive source identification (EASI) algorithm, is, in practice, identical with SINR-maximizing generalized eigenfiltering. Strictly speaking, we gave the necessary and sufficient condition under which the stationary point of the EASI algorithm maximizes the linear output SINR. We also proved that, in the special case of interference-free (that is, noise only) system, the EASI algorithm attains exactly the greatest diversity gain blindly, i.e., performs as a blind maximal ratio combiner (MRC). In addition, the numerical results were given to show that the performance of the EASI algorithm is remarkably close to the optimal (i.e., the maximal output SINR among all linear transforms of observed data) also in cases in which the above mentioned theoretical optimality condition is not met.

7. REFERENCES

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, vol. 2, Wiley, New York, 2001.

[2] T. Huovinen, A. Shahed hagh ghadam, and M. Valkama, "Blind diversity reception and interference cancellation using ICA," in *Proc. IEEE ICASSP2007*, April 2007, pp. 685–688.

[3] Z. Koldovsk and P. Tichavsk, "Methods of fair comparison of performance of linear ICA techniques in presence of additive noise," in *Proc. IEEE ICASSP2006*, May 2006, pp. 873–876.

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

[5] G. Strang, *Linear Algebra and Its Applications*, Thompson Learning, USA, 3rd edition, 1988.

[6] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2637–2640, 1996.

[7] J. H. Winters, "Optimum combining in digital mobile radio with cochannel interference," *IEEE Trans. Veh. Technol.*, vol. 33, no. 3, pp. 144–155, 1984.

[8] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Upper Saddle River, NJ, 3rd edition, 1996.

[9] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1992.

[10] M. Davies, "Identifiability issues is noisy ICA," *IEEE Signal Processing Letters*, vol. 11, no. 5, pp. 470–473, 2004.

[11] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A technique for higher-order-only blind source separation," in *Proc. ICONIP*, 1996.