

New Negentropy Optimization Schemes for Blind Signal Extraction of Complex Valued Sources

Peng-cheng XU¹, Yue-hong SHEN¹, Hui LI²

¹College of Communications Engineering, PLA University of Science and Technology, 210007, Nanjing, P. R. China

²PLA Troop 78020, 650223, Kunming, P. R. China

pchxu2005@gmail.com, chunfeng22259@126.com, leehoo86@163.com

Abstract. *Blind signal extraction, a hot issue in the field of communication signal processing, aims to retrieve the sources through the optimization of contrast functions. Many contrasts based on higher-order statistics such as kurtosis, usually behave sensitive to outliers. Thus, to achieve robust results, nonlinear functions are utilized as contrasts to approximate the negentropy criterion, which is also a classical metric for non-Gaussianity. However, existing methods generally have a high computational cost, hence leading us to address the problem of efficient optimization of contrast function. More precisely, we design a novel "reference-based" contrast function based on negentropy approximations, and then propose a new family of algorithms (Alg. 1 and Alg. 2) to maximize it. Simulations confirm the convergence of our method to a separating solution, which is also analyzed in theory. We also validate the theoretic complexity analysis that Alg. 2 has a much lower computational cost than Alg. 1 and existing optimization methods based on negentropy criterion. Finally, experiments for the separation of single sideband signals illustrate that our method has good prospects in real-world applications.*

Keywords

Negentropy criterion, blind signal extraction, quadratic optimization, global convergence, reference system

1. Introduction

Source separation, an active research area, has found wide applicability. Historically, the source separation problem has been posed with flexible and general assumptions as well as minimal priors, hence leading to the designation blind source separation. In particular, blind extraction technique for complex valued sources has found utility in many applications such as communications [1–3], face recognition [4], analysis of functional magnetic resonance imaging [5], electroencephalograph [6], [7], and radar data [8], [9]. Depending on the applications, the sources may be both sub-Gaussian (e.g. digital communication signals) and super-Gaussian distributed (e.g. radio broadcast signals).

In a multi-input/multi-output (MIMO) context, the problem of blind source separation has found interesting solutions through the optimization of so-called contrast functions (separation criteria) [10]. Many contrasts relying on higher-order statistics (e.g. the kurtosis contrast) have been easily extended from the real to the complex domain [11–13]. However, using these contrast functions generally results in an estimate sensitive to outliers [14]. An alternative approach to achieve robust results is to use nonlinear functions as contrast functions to approximate the probability density functions (pdf) of sources (like in the maximum likelihood and infomax criteria) or the outputs' negentropy [15]. Several contrast functions and algorithms have been proposed based on the nonlinearities of the outputs. In [16], an extension of the well known negentropy-based FastICA to the complex case by Bingham and Hyvarinen was proposed using the modulus information. Novey and Adali proposed in [17] a complex-FastICA algorithm for non-circular sources using the information of pseudo-covariance matrix. In [15], Novey and Adali derived both a gradient-descent and a quasi-Newton algorithm that use the full second-order statistics for complex sources, generalizing their work in [17]. Duran-Diaz et al. presented an algorithm by decoupling the arguments of the criterion so that the algorithm maximizes it cyclically with respect to each argument [18]. But these methods generally have a high computational cost.

On the other hand, contrast functions referred to as "reference-based" have been recently proposed [19], [20]. The output variables in the traditional contrast functions are decoupled into several relatively independent ones, so it is the same with extraction vectors. Then some extraction vectors are artificially specified as primary extraction vectors, while others as reference extraction vectors. Practically, "reference" implies that updating extraction vector should refer to the fixed reference vector so as to determine the optimization direction. Thus, it is worth noting that these "reference-based" approaches are different from the classical concept like "data-aided" or semi-blind method, but in fact any signal or vector fixed prior in the separation can be called "reference" signal or "reference" vector. They are particularly appealing because the corresponding maximization problem can be evolved into a quadratic optimization problem with respect to the searched parameters. Tak-

ing advantage of this quadratic feature, a few maximization algorithms based on kurtosis contrast function have been proposed and significantly quicker than traditional kurtosis-based algorithms [21–23]. However, these methods generally require searching for the optimal step-size and behave sensitive to outliers. In these contexts, we consider the optimization of “reference-based” negentropy approximations.

In this paper, we remodel the contrast function based on negentropy approximations and propose a new family of algorithms (Alg. 1 and Alg. 2) to search for the solution. Starting from a basic “reference-based” algorithm (i.e. Alg. 1), the method is extended by introducing two different number of iteration parameters, and we call the algorithm Alg. 2. Depending on these parameters, a tradeoff can be adjusted between separation performance and speed of the optimization method. The main novelties are the following ones:

- the possibility in the new method to adjust two parameters for improved performance;
- a detailed theoretic analysis of complexity, which well explains the computational cost of all the algorithms in simulations;
- a link with Expectation-Maximization and Minimization-Maximization algorithms, which justifies the convergence of our “fixed-point” like algorithm (i.e. Alg. 2).

The remainder of the paper is organized as follows. Section 2 introduces the data model and assumptions of the paper. In Sec. 3, the remodeled contrast function and the separating method are given. The optimization algorithm (Alg. 1) is improved, and a proof of the convergence for Alg. 2 is given in Sec. 4. Section 5 presents simulation results and discussion. Finally, Section 6 concludes the paper.

2. Model and Assumptions

In the noise-free instantaneous case, we assume that n unknown statistically independent source signals of zero mean and unit variance, contained within $\mathbf{s}(t) = [s_1(t), \dots, s_n(t)]^T \in \mathbb{C}^n$ pass through an unknown full-column rank mixing matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ ($m \geq n$), therefore m mixed signals $\mathbf{x}(t) = [x_1(t), \dots, x_m(t)]^T$ can be modeled as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \quad (1)$$

where t is the sample index. To simplify the problem, we further assume that the number of sources matches the number of mixtures i.e. $m = n$, an exactly determined problem. The prewhitening of the observations by the matrix \mathbf{U} yields the new observations vector, $\mathbf{z}(t) = \mathbf{U}\mathbf{x}(t)$, whose covariance $E[\mathbf{z}\mathbf{z}^H]$ is the identity matrix of dimension $N \times N$.

Then the output signal $y(t)$ that estimates one of the sources is obtained as $y(t) = \mathbf{w}^H \mathbf{z}(t)$, where \mathbf{w} is the unit norm extraction vector (enforcing the output to have unit variance). More specifically, our approach is an iterative one and the extraction of all sources can then be performed through a deflation or symmetric orthogonalization procedure.

In this article, the superscripts $()^*$, $()^T$, $()^H$ denote conjugate, transpose and conjugate transpose, respectively. Bold upper (respectively, lower) case letters are used for matrices (respectively, vectors).

3. Separation Method

3.1 Contrast Function

According to the central limit theorem, which states that the linear combination of independent random variables with finite support pdfs tends toward the Gaussian distribution, one should choose \mathbf{w} such that y has the most non-Gaussian distribution. In other words, if \mathbf{w} is a column of the inverse matrix of $\mathbf{U}\mathbf{A}$, $\mathbf{w}^H \mathbf{U}\mathbf{A}$ has only one nonzero element. However as $\mathbf{w}^H \mathbf{U}\mathbf{A}$ deviates from this solution, this implies a linear mixture of the sources, causing the output signal y to become more Gaussian. This justifies the maximization of non-Gaussianity, which can be measured by kurtosis or negentropy.

Kurtosis (or called the fourth-order cumulant) of y is classically defined by [24]

$$\text{Kurt}(y) = E\{y^4\} - 3\left(E\{y^2\}\right)^2. \quad (2)$$

Actually, since we assumed that y is of unit variance, the right-hand side simplifies to $E\{y^4\} - 3$, which shows that kurtosis is simply a normalized version of the fourth moment. Kurtosis is zero for a Gaussian random variable, while for most non-Gaussian random variables, kurtosis is nonzero. Random variables that have a negative kurtosis are called sub-Gaussian, and those with positive kurtosis are called super-Gaussian. From the view of pdf, super-Gaussian random variables have typically a “spiky” pdf with heavy tails, i.e. the pdf is relatively large at zero while being small at large values of the variable, on the other hand, sub-Gaussian random variables have typically a “flat” pdf, which is rather constant near zero, and also very small for values larger than a certain threshold of the variable.

Using kurtosis contrast function generally results in an estimate sensitive to outliers. So we consider utilizing negentropy contrast instead. Negentropy is defined as

$$J(y) = H(y_{\text{gauss}}) - H(y) \quad (3)$$

where $H(y) = -E\{\log(p(y))\}$ is the differential entropy, $p(y)$ is the pdf, and y_{gauss} is a complex Gaussian random

variable with the same variance as y . Negentropy is always positive but equals zero if and only if y is Gaussian. Thus the maximization of negentropy is the same as the maximization of non-Gaussianity, which leads to the estimation of one source.

Using negentropy as the contrast function requires prior knowledge of the pdf which is generally unknown, or online estimation of the pdf costs too much computational time. To overcome these difficulties, nonlinear functions are utilized to approximate the negentropy and hence (3) is rewritten as [18]

$$J(\mathbf{w}) = h_N - \nu E \left[\left| G(\mathbf{w}^H \mathbf{z}) \right|^2 \right] \quad (4)$$

where h_N represents the entropy of a complex Gaussian random variable with unit variance, and $G(y)$ is a selected nonlinearity. Any value selected for h_N larger than a certain threshold, usually requiring $h_N > 2$, guarantees the ascent of $J(\mathbf{w})$. Besides, the parameter ν is introduced to practically guarantee the stability of the algorithm. It takes values within the set $\{-1, +1\}$, depending on whether the sources are super-Gaussian or sub-Gaussian distributed.

It has been proved in [15] and [18] that (4) is a contrast function: this means by definition that when maximized with respect to the separating vector, this criterion leads to the separation of one source signal. In this paper, we remodel this negentropy-based approximation and propose a new method to optimize it.

Generally, $G(y)$ can be expressed as the form of power series expansion $G(y) = \sum_{i=1}^{\infty} c_i y^i$ where $c_i \in \mathbb{R}$.

To begin with, we can define a function $F(y)$ such that $G(y) = y \sum_{i=1}^{\infty} c_i y^{i-1} = y F(y)$, which will be useful for

decoupling the problem, as it will be shown below. Since the output signal y has unit variance, we can rewrite (4) as

$$J(\mathbf{w}) = h_N E [y y^*] - \nu E \left[y |F(y)|^2 y^* \right]. \quad (5)$$

When function $G(y)$ is properly selected we have $h_N \geq E \left[y |F(y)|^2 y^* \right]$, so expression (5) should be rewritten as

$$J(\mathbf{w}) = \left| h_N E [y y^*] - \nu E \left[y |F(y)|^2 y^* \right] \right| \quad (6)$$

$$= \left| h_N \mathbf{w}^H \mathbf{w} - \nu \mathbf{w}^H E \left[\mathbf{z} |F(\mathbf{w}^H \mathbf{z})|^2 \mathbf{z}^H \right] \mathbf{w} \right|$$

where the second equality follows from the fact that $y(t) = \mathbf{w}^H \mathbf{z}(t)$ and $E[\mathbf{z}\mathbf{z}^H] = \mathbf{I}$.

In order to optimize the criterion in “reference-based” way, we mark the involved output variable y and rewrite (6) as follows

$$J(\mathbf{w}) = \left| h_N E \left[y^{[1]} (y^{[1]})^* \right] - \nu E \left[y^{[1]} |F(y^{[2]})|^2 (y^{[1]})^* \right] \right|$$

$$= \left| h_N (\mathbf{w}^{[1]})^H \mathbf{w}^{[1]} - \nu (\mathbf{w}^{[1]})^H E \left[\mathbf{z} |F((\mathbf{w}^{[2]})^H \mathbf{z})|^2 \mathbf{z}^H \right] \mathbf{w}^{[1]} \right| \quad (7)$$

Then, we remodel the contrast function by decoupling two variables $y^{[1]}$ and $y^{[2]}$, which allows us to isolate the term $|F((\mathbf{w}^{[2]})^H \mathbf{z})|^2$, the only one for which we cannot obtain a quadratic dependence. This choice yields a quadratic dependence with respect to the optimized parameter $\mathbf{w}^{[1]}$. More precisely, after decoupling the variables, the resulting function is

$$I(\mathbf{w}, \mathbf{v}) = \left| h_N \mathbf{w}^H \mathbf{w} - \nu \mathbf{w}^H E \left[\mathbf{z} |F(\mathbf{v}^H \mathbf{z})|^2 \mathbf{z}^H \right] \mathbf{w} \right| \quad (8)$$

where \mathbf{w} and \mathbf{v} are two independent candidates for the extraction vector. We introduce a “reference signal” $r(t)$ which is the output of “reference extraction vector” denoted by \mathbf{v} , i.e. $r(t) = \mathbf{v}^H \mathbf{z}(t)$. In the work of previous researchers, the reference signal was always interpreted as a prior knowledge on the source to be extracted [19]. Another idea, which is in contrast to this concept, indicates that the reference signal can be artificially introduced in an algorithm for the purpose of facilitating the optimization of the contrasts. Thus, any signal or vector fixed antecedently in the separation can be called “reference” signal or “reference” vector. It hence appears as an efficient way of exploiting the properties of the criterion to be maximized.

Obviously, the contrast function (8) is linked with (6) by $J(\mathbf{w}) = I(\mathbf{w}, \mathbf{w})$. Our separation system and most notations are summed up in Fig. 1.

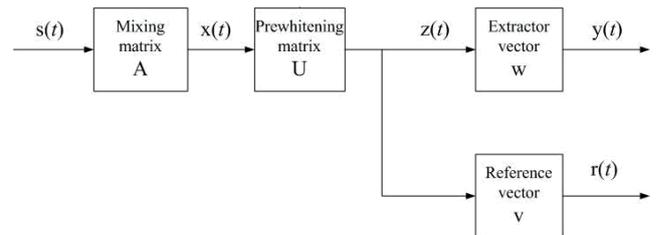


Fig. 1. System and notations summary.

3.2 Optimization Algorithm

Basically, in order to maximize the contrast function (8), we can introduce a steepest ascent algorithm which moves from one point to another following the gradient direction of the criterion. However, improper choosing of step-size in gradient algorithm usually leads to slow convergence speed and poor stability performance [25]. To derive a more efficient algorithm, we note that at a stable point of the gradient algorithm, the gradient of the contrasts must point in the direction of extraction vector, that is, the

gradient must be equal to extraction vector, and the algorithm can converge [26]. This can be proven more rigorously using the technique of Lagrange multipliers. More precisely, we give the basic and simplest version of the algorithm. Further refinements will be addressed in Sec. 4.

Algorithm 1

- Initialization: \mathbf{v}_0 is created randomly, and the corresponding reference signal $r_0(t) = \mathbf{v}_0^H \mathbf{z}(t)$.
Set the initial extraction vector $\mathbf{w}_0 = \mathbf{v}_0$.
- For $k = 0, 1, 2, \dots, k_{\max} - 1$, repeat (M0-U):
(M0) Set $\mathbf{d}_k = \partial I(\mathbf{w}_k, \mathbf{v}_k) / \partial \mathbf{w}_k^*$
 $\mathbf{w}_{k+1} \leftarrow \mathbf{d}_k$ and renormalize:
 $\mathbf{w}_{k+1} \leftarrow \mathbf{w}_{k+1} / \|\mathbf{w}_{k+1}\|$
(U) Update $\mathbf{v}_{k+1} \leftarrow \mathbf{w}_{k+1}$

As is seen from Alg. 1, at each iterative, the direction of extraction vector is modified by the partial gradient of the first argument \mathbf{w} , while the second argument \mathbf{v} is fixed to a constant vector. Then the extraction vector \mathbf{w} is modified, and also the reference extraction vector \mathbf{v} is updated to fabricate a new “reference signal” $r(t)$. Furthermore, the iteration formula $\mathbf{d}_k = \partial I(\mathbf{w}_k, \mathbf{v}_k) / \partial \mathbf{w}_k^*$ describes the “reference-based” concept specifically from the view of implementation.

The two arguments \mathbf{w} and \mathbf{v} like two persons climbing a mountain, one of who (\mathbf{v}) does guard duty while another person (\mathbf{w}) starts from the first person (\mathbf{v}) to look for the way as easy as possible to the top. Once the person \mathbf{w} finds the way, he motions to \mathbf{v} to reach his position. They repeat this action until they reach the mountaintop. This vividly explains why \mathbf{v} is called “reference vector”.

Now we derive the gradient direction in Alg. 1. Since the formula (5) is real valued, we can rewrite (8) as

$$I(\mathbf{w}, \mathbf{v}) = \left[\left(h_N \mathbf{w}^H \mathbf{w} - \nu \mathbf{w}^H \mathbf{E} \left[\mathbf{z} \left| F(\mathbf{v}^H \mathbf{z}) \right|^2 \mathbf{z}^H \right] \mathbf{w} \right)^2 \right]^{\frac{1}{2}} \quad (9)$$

Therefore, we have

$$\begin{aligned} \mathbf{d} &= \partial I(\mathbf{w}, \mathbf{v}) / \partial \mathbf{w}^* \\ &= \frac{1}{2} \left[\left(h_N \mathbf{w}^H \mathbf{w} - \nu \mathbf{w}^H \mathbf{E} \left[\mathbf{z} \left| F(\mathbf{v}^H \mathbf{z}) \right|^2 \mathbf{z}^H \right] \mathbf{w} \right)^2 \right]^{\frac{1}{2}} \\ &\quad \times 2 \left(h_N \mathbf{w}^H \mathbf{w} - \nu \mathbf{w}^H \mathbf{E} \left[\mathbf{z} \left| F(\mathbf{v}^H \mathbf{z}) \right|^2 \mathbf{z}^H \right] \mathbf{w} \right) \\ &\quad \times \left(h_N \mathbf{w} - \nu \mathbf{E} \left[\mathbf{z} \left| F(\mathbf{v}^H \mathbf{z}) \right|^2 \mathbf{z}^H \right] \mathbf{w} \right) \\ &= h_N \mathbf{w} - \nu \mathbf{E} \left[\mathbf{z} \left| F(\mathbf{v}^H \mathbf{z}) \right|^2 \mathbf{z}^H \right] \mathbf{w} \end{aligned} \quad (10)$$

By the Modification and Update (M0-U) rule in Alg. 1, one can then verify the following more practical expression as iteration formula.

$$\begin{aligned} \mathbf{w}_{k+1} &= h_N \mathbf{w}_k - \nu \mathbf{E} \left[\mathbf{z} \left| F(\mathbf{w}_k^H \mathbf{z}) \right|^2 \mathbf{z}^H \right] \mathbf{w}_k \\ \mathbf{w}_{k+1} &\leftarrow \mathbf{w}_{k+1} / \|\mathbf{w}_{k+1}\| \end{aligned} \quad (11)$$

Although the formula (11) is simple and effective for real-world application, it obscures the crucial concept of the “reference-based” method to some extent.

4. Improvements and Convergence Analysis

In order to improve Alg. 1, it is natural to replace the update step (M0) by a more efficient operation. Here, we consider replacing the optimization along a line in (M0) by a global optimization, i.e., the step (M0) should achieve the goal that

$$\mathbf{v}_{k+1} \leftarrow \arg \max_{\mathbf{w}} I(\mathbf{w}, \mathbf{v}_k). \quad (12)$$

Therefore, the metaphor is changed here: Two persons (\mathbf{w} and \mathbf{v}) climb a mountain. One of them (\mathbf{v}) still does guard duty, but another person (\mathbf{w}) starts from the first person (\mathbf{v}) to look for the highest position around \mathbf{v} . Once the person \mathbf{w} finds the position, he motions to \mathbf{v} to reach his position. They repeat this action until they reach the mountaintop. This yields the following algorithm.

Algorithm 2

- Initialization: \mathbf{v}_0 is created randomly, and the corresponding reference signal $r_0(t) = \mathbf{v}_0^H \mathbf{z}(t)$.
- For $k = 0, 1, 2, \dots, k_{\max} - 1$, set $\mathbf{w}_0 = \mathbf{v}_k$ and repeat (M1-U):
(M1) For $l = 0, 1, 2, \dots, l_{\max} - 1$ repeat (M0):
(M0) Set $\mathbf{d}_l = \partial I(\mathbf{w}_l, \mathbf{v}_k) / \partial \mathbf{w}_l^*$
Set $\mathbf{w}_{l+1} \leftarrow \mathbf{d}_l$ and renormalize:
 $\mathbf{w}_{l+1} \leftarrow \mathbf{w}_{l+1} / \|\mathbf{w}_{l+1}\|$
(U) Update $\mathbf{v}_{k+1} \leftarrow \mathbf{w}_{l_{\max}}$.

First note that a new parameter l_{\max} has been introduced in Alg. 2 in addition to k_{\max} . Depending on these two parameter values, Alg. 2 is provided with both “reference” and “fixed-point” properties. Indeed, Alg. 2 is intermediate between Alg. 1 ($l_{\max} = 1$) and a “fixed-point” like algorithm ($k_{\max} = 1$). In other words, at iteration steps reference vector \mathbf{v}_k in Alg. 2 is no longer updated every iterative as Alg. 1, either remains unchanged all the time as a “fixed-point” \mathbf{v}_0 , but is fixed in several steps l_{\max} to search for the optimal solution \mathbf{w} as the next reference vector.

Additionally, starting from an initial “reference” \mathbf{v}_0 , Alg. 2 iteratively produces a sequence of better estimate \mathbf{v}_k for the separating system. This point of view exhibits the similarity with Expectation-Maximization (EM) [27] or Maximization-Minimization (MM) methods [28].

It is worth noting that our approach is related to other works. First, we have noticed that, in [29–33] the timing synchronization for turbo receivers in environments like AWGN channel and fast time-variant channel has been addressed. Although they are not directly linked to our approach, it is shown in these works how maximum-likelihood estimation of the synchronization parameters like frequency offset, propagation delays, carrier phase offsets and received amplitudes can be implemented by means of the iterative expectation-maximization (EM) algorithm. Moreover [34] has generalized the synchronizer based on EM-algorithm to the multi-user case.

In these previous works, the EM algorithm maximizes a likelihood function $L(\mathbf{v})$ by producing a set of successive estimates \mathbf{v}_k obtained by iteratively maximizing a Q-function

$$\mathbf{v}_{k+1} = \arg \max_{\mathbf{v}} Q(\mathbf{v}, \mathbf{v}_k). \quad (13)$$

In EM methods, the Q-function is obtained by introducing some hidden data in the model and taking the expectation of the complete log-likelihood. At each iterative, the EM algorithm breaks up into two steps: in the Expectation step (E-step), the average log-likelihood of the complete data is computed, while in the Maximization step (M-step), the average log-likelihood is maximized. In contrast to EM methods, MM methods do not require to identify a complete data set and take advantage of convexity and inequalities satisfied by the objective function [35].

The MM methods, which are based on (13) show convergence properties under the condition that at each step, the L and Q-function satisfies

$$L(\mathbf{v}) \geq Q(\mathbf{v}, \mathbf{v}_k) \text{ and } L(\mathbf{v}_k) = Q(\mathbf{v}_k, \mathbf{v}_k). \quad (14)$$

Such condition has been proved in the EM context using Jensen's inequality, and then

$$\begin{aligned} L(\mathbf{v}_{k+1}) &= \underbrace{L(\mathbf{v}_{k+1}) - Q(\mathbf{v}_{k+1}, \mathbf{v}_k)}_{\geq 0} + Q(\mathbf{v}_{k+1}, \mathbf{v}_k) \\ &\geq Q(\mathbf{v}_{k+1}, \mathbf{v}_k) \\ &\geq Q(\mathbf{v}_k, \mathbf{v}_k) = L(\mathbf{v}_k) \end{aligned} \quad (15)$$

where the last inequality follows from (13). Consequently, the objective criterion monotonically increases, which results in the convergence of the algorithm.

It can be seen that Alg. 2 satisfies all properties of the MM maximization methods with L corresponding to J and Q corresponding to I . Meanwhile, the convergence conditions (13) and (14) can also be satisfied by (12) and the convexity of criterion. Thereby, one can see that Alg. 2 belongs to the class of EM or MM methods, and monotonically increases the criterion, which justifies its convergence.

In practice, Alg. 2 can be implemented in a more effective way. Now we define the column vector

$$\mathbf{z}\mathbf{f}(\mathbf{v}_k) = \mathbf{z} \left| \mathbf{F}(\mathbf{v}_k^H \mathbf{z}) \right|^2. \quad (16)$$

Then, the expression of gradient direction $\partial I(\mathbf{w}_l, \mathbf{v}_k) / \partial \mathbf{w}_l^*$ given by (10) can be rewritten as

$$\partial I(\mathbf{w}_l, \mathbf{v}_k) / \partial \mathbf{w}_l^* = h_N \mathbf{w}_l - \nu \mathbb{E} \left[\mathbf{z}\mathbf{f}(\mathbf{v}_k) y_l^* \right] \quad (17)$$

where $y_l = \mathbf{w}_l^H \mathbf{z}$. More precisely, since the values of column vector $\mathbf{z}\mathbf{f}(\mathbf{v}_k)$ are constant during the iteration step (M0), we can complete the computation of this vector earlier in the step (M1), which hence saves considerable execution time for the step (M0).

As is seen from Sec. 3 and Sec. 4, we can insight into the close relationship between contrast function and corresponding algorithms. Since the contrast function $J(\mathbf{w})$ is remodeled into $I(\mathbf{w}, \mathbf{v})$ depending quadratically on \mathbf{w} , it is no doubt that the algorithms (Alg. 1 and Alg. 2) have the potentiality of two-dimensional optimization. However, for classical methods such as complex fast independent component analysis (FastICA) algorithm [16] and complex maximization of non-Gaussianity (CMN) algorithm [15], with contrast function of only one variable, the optimization algorithms have to be one-dimensional.

5. Simulation Results and Discussion

5.1 Validity and Separation Quality

a) Simulation settings: In these simulations, we mixed three complex valued 4QAM sources, which have been generated taking values in $\{e^{j\pi/4}, e^{-j\pi/4}, e^{j3\pi/4}, e^{-j3\pi/4}\}$

with equal probability 1/4. The mixing matrix \mathbf{A} in data model of (1) and the initial extraction vector \mathbf{v}_0 were created randomly. The data length was varied from 100 to 5000 samples. We compared the proposed algorithms with classical algorithms based on negentropy criterion, i.e., FastICA with the contrast function

$J(\mathbf{w}) = \mathbb{E} \left[G \left(\left| \mathbf{w}^H \mathbf{z} \right|^2 \right) \right]$ and CMN algorithm with the con-

trast function $J(\mathbf{w}) = \mathbb{E} \left[\left| G(\mathbf{w}^H \mathbf{z}) \right|^2 \right]$. The exponent of

1.25 used in the nonlinear function $y^{1.25}$ was chosen in that it makes a contrast function grow slower than classical kurtosis contrast function, and the latter can be treated as a special case of the complex FastICA with the nonlinear function y^2 . This choice provides a more robust performance, and it has been also proven in [15] that the CMN algorithm with $G(y) = y^{1.25}$ for separating many kinds of sub-Gaussian and super-Gaussian sources performs optimal. Considering the above reasons, also allowing fair comparison, for all the algorithms we chose the same nonlinearity $G(y) = y^{1.25}$ (i.e., $F(y) = y^{0.25}$).

The total number of iterations was set to the same value 1000. We set $\nu = 1$ for 4QAM signals, the pdf of which is ‘‘flat’’ with sub-Gaussian distribution. All the results presented below result from Monte Carlo Simulations

Separation methods			Average MSE					Standard Deviation of MSE				
			Number of samples					Number of samples				
parameters	k_{max}	l_{max}	100	500	1000	2000	5000	100	500	1000	2000	5000
CMN	1000	-	3.68E-4	7.14E-5	3.39E-5	1.68E-5	7.38E-6	4.33E-4	9.27E-5	4.01E-5	2.32E-5	9.76E-6
FastICA	1000	-	3.89E-4	7.41E-5	3.63E-5	1.85E-5	7.61E-6	4.24E-4	9.41E-5	4.72E-5	2.66E-5	9.89E-6
Alg.1	1000	-	4.14E-4	6.73E-5	3.63E-5	1.73E-5	6.92E-6	4.51E-4	9E-5	4.73E-5	2.21E-5	9.02E-6
Alg.2	200	5	3.78E-4	6.87E-5	3.45E-5	1.78E-5	6.96E-6	4.32E-4	9.01E-5	4.59E-5	1.93E-5	9.16E-6
Alg.2	100	10	3.76E-4	6.91E-5	3.58E-5	1.75E-5	6.91E-6	4.39E-4	9.1E-5	4.61E-5	2.24E-5	9E-6
Alg.2	50	20	3.6E-4	7.22E-5	3.48E-5	1.79E-5	6.84E-6	4.34E-4	9.34E-5	4.62E-5	2.3E-5	9.3E-6
Alg.2	25	40	3.38E-4	7.06E-5	3.54E-5	1.68E-5	7.16E-6	4.46E-4	9.12E-5	4.69E-5	2.31E-5	9.43E-6
Alg.2	10	100	3.64E-4	7.21E-5	3.53E-5	1.65E-5	7.09E-6	4.61E-4	9.17E-5	4.77E-5	2.32E-5	9.35E-6
Alg.2	8	125	4.99E-4	7.25E-5	3.71E-5	1.82E-5	7.5E-6	5.04E-4	1.08E-4	5.18E-5	2.8E-5	1.02E-5
Alg.2	5	200	3.54E-3	6.15E-4	2.91E-4	1.4E-4	5.49E-5	8.27E-4	1.61E-4	7.81E-5	3.85E-5	1.54E-5

Tab. 1. MSE for different separation methods. (Three 4QAM Sources, Random Mixing Matrix. 1000 Monte Carlo Realization).

involving 1000 realizations. At each run, the mixing system, the sources and the initial coefficients of the reference system have been drawn randomly.

b) Extraction of one source: Results are shown in Tab. 1 where the average and standard deviation values of mean square error (MSE)

$$MSE = \frac{1}{T} \sum_{t=1}^T \|y(t) - s(t)\|^2 \quad (18)$$

are adopted as our separation performance measurement, where s is the source signal and y is the recovered corresponding signal, both of which are normalized.

The top row collects the results provided by CMN, which maximizes the contrast function by Newton algorithm (or called second-order learning algorithm) [15]. The second row corresponds to the complex FastICA algorithm which also follows Newton iteration [16]. The third row shows the results provided by Alg. 1, which is described in detail in Sec. 3. Finally, the successive rows present the results given by Alg. 2 with different values of the parameters k_{max} , l_{max} .

From the first three rows of Tab. 1, we can observe that FastICA, CMN and Alg. 1 yield similar values of MSE, which confirms the validity of Alg. 1. In addition, an equally good value of MSE can be obtained by Alg. 2, which proves its validity and convergence to a separating point.

c) Extraction of all sources: It has been illustrated that reference-based contrasts may be unsatisfying when extracting all sources [23]. We hence tested here the behavior of our optimization algorithms when extracting all sources. Among the methods that prevent the outputs from converging to the same source, we considered the symmetric orthogonalization and the deflation schemes. The symmetric orthogonalization consists in the execution of the algorithm N times simultaneously and the symmetric decorrelation of the extraction vectors [16]. Taking Alg. 2 for example, this decorrelation is carried out after each iteration ($M0$) by means of

$$\mathbf{W}_l \leftarrow \mathbf{W}_l (\mathbf{W}_l^H \mathbf{W}_l)^{-1/2} \quad (19)$$

where \mathbf{W}_l is the extraction matrix at the l th iteration. In contrast, the deflation method estimates the sources sequentially, which probably has the advantage that convergence performance remains valid, due to the fact that the iterations of the extraction algorithm are not affected [36]. In this case the deflation method removes each source from the data after each estimate, i.e., once the $j-1$ previous extracted sources have been estimated, the observations vector for the j th extraction becomes

$$\mathbf{x}_j(t) = \left(\mathbf{I} - \sum_{q=1}^{j-1} \mathbf{v}_q \mathbf{v}_q^H \right) \mathbf{z}(t) \quad (20)$$

where \mathbf{v}_q is the extraction vector for the q th extracted sources. Then the extraction algorithms are carried on to achieve the j th extraction vector and extract the j th source. These processes would be executed several times until all sources have been extracted. The results are provided in Tab. 2 and show that the proposed algorithms behave equally well as (or even better than) complex FastICA algorithm.

Extracted source number	Deflation method			Symmetric orthogonalization	
	1st	2nd	3rd		
Average MSE	FastICA	6.57E-6	1.84E-5	3.10E-5	1.83E-5
	Alg. 1	2.98E-6	1.24E-5	3.66E-5	1.86E-5
	Alg. 2	3.02E-6	1.31E-5	3.77E-5	1.80E-5
Standard Deviation of MSE	FastICA	9.68E-6	2.38E-5	2.61E-5	1.64E-5
	Alg. 1	5.53E-6	1.57E-5	2.52E-5	1.52E-5
	Alg. 2	6.54E-6	1.53E-5	2.49E-5	1.53E-5

Tab. 2. Average and Standard Deviation of MSE for extraction of three 4QAM sources. ($T=2000$ samples, $k_{max}=50$, $l_{max}=20$. Number of Monte Carlo runs = 1000)

d) Experiments on the separation performance in Additive White Gaussian Noise (AWGN) channel: We also investigated the effect of noise on the extraction performance when received signals are contaminated with complex white Gaussian noise, i.e., $\mathbf{x}(t) = \mathbf{A}s(t) + \mathbf{b}(t)$, where $\mathbf{b}(t) = [b_1(t), b_2(t), \dots, b_n(t)]^T$ is the noise vector. A detailed performance of MSE with respect to signal to noise ratios (10 dB, 20 dB, 30 dB, 40 dB, 50 dB) was obtained among complex FastICA, Alg. 1 and Alg. 2. Here we implemented the extraction of all sources by symmetric orthogonalization schemes. The results are provided in Tab. 3.

Separation Methods	Average MSE					Standard Deviation of MSE				
	Signal to Noise Ratio (dB)					Signal to Noise Ratio (dB)				
	10	20	30	40	50	10	20	30	40	50
FastICA	4.99E-5	2.2E-5	2.04E-5	1.85E-5	1.81E-5	3.58E-5	1.92E-5	1.74E-5	1.68E-5	1.55E-5
Alg.1	4.43E-5	2.38E-5	1.91E-5	1.87E-5	1.85E-5	2.99E-5	2.05E-5	1.7E-5	1.69E-5	1.54E-5
Alg.2	4.79E-5	2.54E-5	2.01E-5	1.82E-5	1.8E-5	2.99E-5	2.08E-5	1.66E-5	1.65E-5	1.53E-5

Tab. 3. Average and Standard Deviation of MSE for extraction of three 4QAM sources in AWGN channel. ($T=2000$ samples, $k_{\max}=50$, $l_{\max}=20$. Number of Monte Carlo runs = 1000).

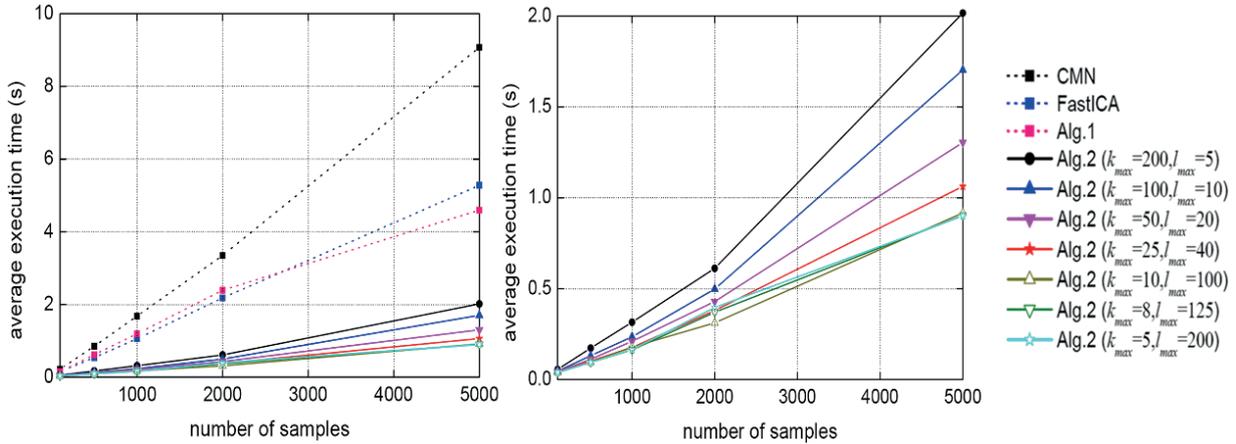


Fig. 2. Execution time for different separation methods in the case of extracting one source.

From Tab. 3 we can observe that the proposed method seems still valid and equally good as complex FastICA, even when the signal to noise ratio decreases by 10 dB. This demonstrates the robustness of the proposed algorithms. What's more, as the signal to noise ratio increases, these methods perform steadier in terms of standard deviation of MSE.

5.2 Computational Load Analysis

It has been observed in the previous subsection that our methods yield a separation performance which is as good as the one provided by some existing negentropy-based algorithms.

Here, we checked the computational time for different separation methods in the case of extraction of one source. The data length was still varied from 100 to 5000 samples, and the total number of iterations was set to the same value 1000. Fig. 2 (left) presents the results for all separation methods, which obviously demonstrates that a good choice of k_{\max} and l_{\max} for Alg. 2 leads to a significantly lower computational time than the others. This is particularly appealing for large number of samples, where the execution time is significantly reduced. The clearer results provided by Alg. 2 are scaled up in Fig. 2 (right). Taking the advantages of reference-based contrasts and ‘‘fixed-point’’ iteration, it can be noticed that Alg. 2 leaves a degree of freedom through the choice of k_{\max} and l_{\max} , which is further investigated next.

A careful inspection indeed reveals that the most time consuming steps in Alg. 2 are:

- (i) the computation of $\mathbf{zf}(\mathbf{v}_k)$ defined in (16),

- (ii) the computation of the gradient of criterion value $\partial I(\mathbf{w}_l, \mathbf{v}_k) / \partial \mathbf{w}_l^*$ given in (17).

As (16) indicates, the complexity of step (i) is an increasing function of both the number of sample size T and the number of sources N . Likewise, step (ii) involves the computation from the set of sample data and is in proportion to N . Let $\beta(T, N)$ and $\gamma(T, N)$ respectively denote the complexity of step (i) and (ii). Now, it can be seen from the description of Alg. 2 that step (i) is repeated k_{\max} times through the algorithm because the column vector $\mathbf{zf}(\mathbf{v}_k)$ changes only after each update (U) step. The overall complexity related to step (i) is thus $O(k_{\max}\beta(T, N))$. Meanwhile, step (ii) are repeated $k_{\max}l_{\max}$ times through the algorithm, which yields a complexity of $O(k_{\max}l_{\max}\gamma(T, N))$. It follows that the order of the complexity of Alg. 2 is:

$$O(k_{\max}\beta(T, N) + k_{\max}l_{\max}\gamma(T, N)) \quad (21)$$

where β and γ are increasing functions of their parameters. It is out of the scope of the paper to specify further β and γ .

The complexity of Alg. 1 is given by the above results with $l_{\max} = 1$, that is:

$$O(k'_{\max}(\beta(T, N) + \gamma(T, N))) \quad (22)$$

which implies that the overall complexity of Alg. 2 must be lower than the one of Alg. 1, due to the fact that these algorithms are equal in total number of iterations, i.e., $k'_{\max} = k_{\max}l_{\max}$. Moreover, note that complex FastICA and CMN algorithms have a complexity which is of the same order of magnitude as Alg. 1. In fact such algorithms perform steps whose complexity is roughly described above. It demonstrates that the execution time of complex FastICA,

CMN and Alg. 1 is in proportion to k'_{\max} , whereas the one of Alg. 2 is in proportion to k_{\max} . Therefore, it is not difficult for us to understand the slope properties of the curves in Fig. 2, that is, in general the slope of Alg. 2's curves declines as the parameter k_{\max} decreases, and it can be much smaller than the other three algorithms since $k_{\max} < k'_{\max}$.

The above analysis confirms the simulation results and the fact that the computational load of Alg. 2 increases swiftly with the increasing of sample size T for large value of k_{\max} , while the opposite for small value of k_{\max} . Finally, given the number of samples, the complexity of Alg. 2 is almost directly proportional to k_{\max} . The analysis can be summed up by a few empirical rules to indicate how k_{\max} and l_{\max} should be chosen:

- for large sample size T (say $T > 5000$), increasing k_{\max} generally ensures a better convergence performance.
- increasing k_{\max} leads to a considerable computational load, especially for large number of samples T . Hence, k_{\max} should be as small as possible (say $k_{\max} < 25$) to obtain lower computational load, and meanwhile it is possible to increase l_{\max} for adequate separation performance.
- in general, k_{\max} should be great enough (approximately $k_{\max} > 10$) to avoid poor separation quality.

5.3 Experiments on the Separation of Single Sideband (SSB) Signals

Finally we show the separation validity and efficiency of our proposed algorithms for practical single sideband (SSB) signals carrying the music information.

To begin with, three baseband sources with three types were utilized: (1) bass; (2) tenor; and (3) jazz music. The sources are all with sampling frequency 8 kHz and time span 2.5 s. Then all the baseband sources were represented as equivalent low-pass signals with single sideband

$$s_i(t) = s(t) + j\hat{s}(t) \quad (23)$$

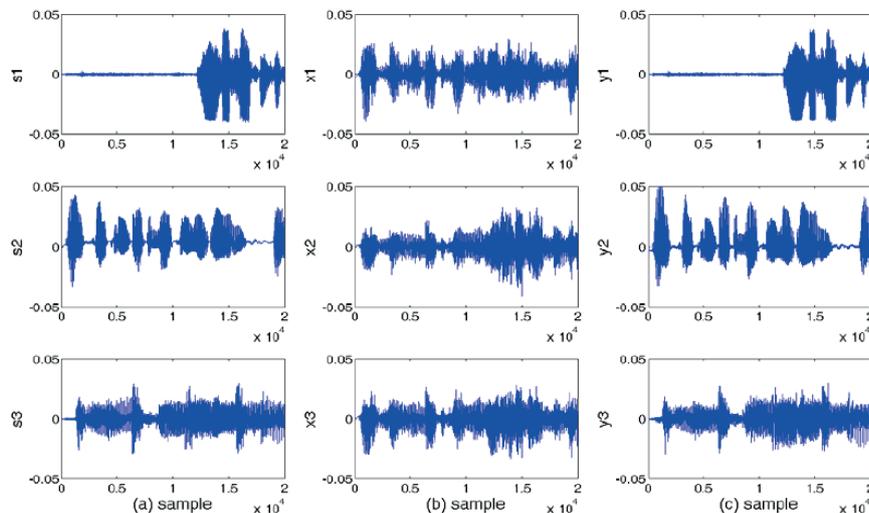


Fig. 3. Separation result of three SSB signals with bass, tenor and jazz music information respectively. (a) Original sources, (b) Observations, (c) Extracted sources.

where $\hat{s}(t)$ denotes the Hilbert transform of $s(t)$. As we know, in wireless communications, the equivalent low-pass representation is meaningful in that the signals which are single-side band modulated save considerable spectrum and power resources comparing to the double-side band signals [37]. Furthermore, we assume that the communication channel is linear without complicated factors like IQ unbalance, multi-path effect and non-linear distortion. Thus in the simulations, the MIMO channel can be characterized by the time-invariant mixing matrix \mathbf{A} that was created randomly in data model of (1). At last, we chose the nonlinearity $G(y) = y^{1.25}$, and set $\nu = -1$ for these super-Gaussian sources.

Fig. 3 gives the separation results provided by Alg. 2 in the waveforms of the real parts of signals, whose imaginary parts need not be redundantly presented here since the imaginary parts are the Hilbert transform of the real parts. It illustrates that our method seems to have good prospects in real-world applications.

A comparison on the performance of MSE and average execution time was also made among complex FastICA, Alg. 1 and Alg. 2. The experiments were carried out in the deflation schemes, and the results are provided in Tab. 4, which demonstrate that Alg. 2 behaves best among the algorithms for its optimal separation quality and lowest computational time.

Extracted source number		Deflation method		
		1st	2nd	3rd
Average MSE	FastICA	1.54E-6	6.32E-6	6.47E-6
	Alg. 1	9.34E-7	2.30E-6	7.38E-6
	Alg. 2	6.92E-7	1.64E-6	2.46E-6
Standard Deviation of MSE	FastICA	9.97E-7	1.86E-6	1.81E-6
	Alg. 1	9.77E-7	1.17E-6	8.55E-6
	Alg. 2	6.56E-7	8.44E-7	5.26E-7
Average execution time (s)	FastICA	47.5		
	Alg. 1	45.8		
	Alg. 2	7.2		

Tab. 4. Comparison among three algorithms for the separation of SSB signals. ($T = 2 \times 10^4$ samples, $k_{\max} = 25$, $l_{\max} = 40$. Number of Monte Carlo runs = 200).

6. Conclusion

In this paper, we have proposed a new method for the extraction of complex sources through the maximization of “reference-based” negentropy approximations. Moreover, our methods perform several optimizations (Alg. 1 and Alg. 2) of this “reference-based” contrast. Simulations confirm the theoretical analysis of the convergence to a separation point. Additionally, the method Alg. 2, which introduces the “fixed-point” like iterations in the context of reference contrasts, allows one to adjust two iteration number parameters to improve the performance. Indeed an appropriate tuning of the number of iterations significantly reduces the computational load and even provides a better separation quality than existing methods. Simulations have shown that Alg. 2 is particularly striking, as it yields an impressive improvement in terms of computational speed for large number of samples. However, the channel model in this paper is simply represented by instantaneous mixtures, such that further research is still needed to determine if our method can be extended to more practical and complicated scenarios like convolutive mixing case or fast time-variant channel.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 61172061 and the Natural Science Foundation of JiangSu Province in China under Grant BK2011117.

References

- [1] VAN DER VEEN, A.-J. Algebraic methods for deterministic blind beamforming. *Proceedings of IEEE*, Oct. 1998, vol. 86, no. 10, p. 1987–2008. DOI: 10.1109/5.720249
- [2] RAJU, K., RISTANIEMI, T., KARHUNEN, J., OJA, E. Suppression of bitpulsed jammer signals in DS-CDMA array system using independent component analysis. In *Proceedings of the Int. Symp. on Circuits and Systems ISCAS 2002*. Phoenix-Scottsdale (Arizona, USA), May 2002, vol. 1, p. 189–192. DOI: 10.1109/ISCAS.2002.1009809
- [3] WAHEED, K. SALEM, F. M. Blind information-theoretic multiuser detection algorithms for DS-CDMA and WCDMA downlink systems. *IEEE Transactions on Neural Networks*, July 2005, vol. 16, no. 4, p. 937–948. DOI: 10.1109/TNN.2005.849848
- [4] KWAK, K. C., PEDRYCZ, W. Face recognition using an enhanced independent component analysis approach. *IEEE Transactions on Neural Networks*, March 2007, vol. 18, no. 2, p. 530–541. DOI: 10.1109/TNN.2006.885436
- [5] CALHOUN, V. D., ADALI, T. Unmixing fMRI with independent component analysis. *IEEE Engineering in Medicine and Biology Magazine*, Apr. 2006, vol. 25, no. 2, p. 79–90.
- [6] ANNEMULLER, J., SEJNOWSKI, T. J., MAKEIG, S. Complex spectral domain independent component analysis of electroencephalographic data. In *4th international Symposium on Independent Component Analysis and Blind Signal Separation ICA2003*. Nara (Japan), March 2003, p. 47–52.
- [7] HUANG, D. S., MI, J. X. A new constrained independent component analysis method. *IEEE Transactions on Neural Networks*, Sep. 2007, vol. 18, no. 5, p. 1532–1535. DOI: 10.1109/TNN.2007.895910
- [8] CHAUMETTE, E., COMON, P., MULLER, D. ICA-based technique for radiating sources estimation: Application to airport surveillance. In *Inst. Electr. Eng. Proc.*, 1993, vol. 140, no. 6, p. 395–401.
- [9] KARLSEN, B., SORENSEN, H. B., LARSEN, J., JACKOBSEN, K. B. Independent component analysis for clutter reduction in ground penetrating radar data. In *Proc. SPIE, Detection and Remediation Technologies for Mines and Minelike Targets VII*, 2002, vol. 4742, p. 378–389. DOI:10.1117/12.479110
- [10] COMON, P., JUTTEN, C. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. New York: Elsevier, 2010.
- [11] MOREUA, E., THIRION-MOREUA, N. Nonsymmetrical contrasts for sources separation. *IEEE Transactions on Signal Processing*, 1999, vol. 47, no. 8, p. 2241–2252. DOI: 10.1109/78.774767
- [12] DOUGLAS, S. C. Fixed-point algorithms for the blind separation of arbitrary complex-valued non-Gaussian signal mixtures. *EURASIP Journal on Advances in Signal Processing*, 2007, p. 1–15. DOI: 10.1155/2007/36525
- [13] LI, Y., WANG, J., CICHOCKI, A. Blind source extraction from convolutive mixtures in ill-conditioned multi-input multi-output channels. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2004, vol. 51, no. 9, p. 1814–1822. DOI: 10.1109/TCSI.2004.832723
- [14] HYVARINEN, A. One-unit contrast functions for independent component analysis: A statistical analysis. In *Proceedings of the 1997 IEEE Workshop on Neural Networks for Signal Processing*. Amelia Island (FL, USA), Sep. 1997, p. 388–397. DOI: 10.1109/NNSP.1997.622420
- [15] NOVEY, M., ADALI, T. Complex ICA by negentropy maximization. *IEEE Transactions on Neural Networks*, 2008, vol. 19, no. 4, p. 596–609. DOI: 10.1109/TNN.2007.911747
- [16] BINGHAM, E., HYVARINEN, A. A fast fixed-point algorithm for independent component analysis of complex valued signals. *International Journal of Neural Systems*, 2000, vol. 10, p. 1–8. DOI: 10.1142/S0129065700000028
- [17] NOVEY, M., ADALI, T. On extending the complex fastICA algorithm to noncircular sources. *IEEE Transactions on Signal Processing*, 2008, vol. 56, no. 5, p. 2148–2154. DOI: 10.1109/TSP.2007.911278
- [18] DURAN-DIAZ, I., CRUCES, S., SARMIENTO-VEGA, M. A., AGUILERA-BONET, P. Cyclic maximization of non-Gaussianity for blind signal extraction of complex-valued sources. *Neurocomputing*, 2011, vol. 74, p. 2867–2873. DOI: 10.1016/j.neucom.2011.03.031
- [19] CASTELLA, M., RHIOUI, S., MOREAU, E., PESQUET, J.-C. Quadratic higher-order criteria for iterative blind separation of a MIMO convolutive mixture of sources. *IEEE Transactions on Signal Processing*, Jan. 2007, vol. 55, no. 1, p. 218–232. DOI: 10.1109/TSP.2006.882113
- [20] DUBROCA, R., DE LUIGI, C., CASTELLA, M., MOREAU, E. A general algebraic algorithm for blind extraction of one source in a MIMO convolutive mixture. *IEEE Transactions on Signal Processing*, May 2010, vol. 58, no. 5, p. 2484–2493. DOI: 10.1109/TSP.2010.2042487
- [21] CASTELLA, M., MOREAU, E. New kurtosis optimization schemes for MISO equalization. *IEEE Transactions on Signal Processing*, March 2012, vol. 60, no. 3, p. 1319–1330. DOI: 10.1109/TSP.2011.2177828
- [22] KAWAMOTO, M., KOHNO, K., INOUE, Y. Eigenvector algorithms incorporated with reference systems for solving blind deconvolution of MIMO-IIR linear systems. *IEEE Signal*

- Processing Letters*, Dec. 2007, vol. 14, no. 12, p. 996–999. DOI: 10.1109/LSP.2007.906225
- [23] CASTELLA, M., MOREAU, E. A new optimization method for reference-based quadratic contrast functions in a deflation scenario. In *Proc. of IEEE Int. Conf. Acoustics, Speech and Signal Processing ICASSP 2009*. Taipei, (Taiwan), Apr. 2009, p. 3161–3164. DOI: 10.1109/ICASSP.2009.4960295
- [24] HYVARINEN, A., OJA, E. Independent component analysis: algorithms and applications. *Neural Networks*, June 2000, vol. 13, no. 4-5, p. 411–430.
- [25] AMARI, S. Natural gradient works efficiently in learning. *Neural Computation*, 1998, vol. 10, p. 251–276.
- [26] HYVARINEN, A., KARHUNEN, J., OJA, E. *Independent Component Analysis*. New York: Wiley press, 2001.
- [27] DEMPSTER, A. P., LAIRD, N. M., RUBIN, D. B. Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Ser. B*, 1977, vol. 39, no. 1, p. 1–38.
- [28] HUNTER, D. R., LANGE, K. A tutorial on MM algorithms. *Amer. Statist.*, Feb. 2004, vol. 58, no. 1, p. 30–37.
- [29] HERZET, C., RAMON, V., VANDENDORPE, L., MOENECLAEY, M. EM algorithm-based timing synchronization in turbo receivers. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2003*. Hong-Kong (China), 2003, vol. 4, p. 612–615. DOI: 10.1109/ICASSP.2003.1202717
- [30] SYKORA, J., VCELAK, J. Iterative EM based IMD synchronization for fast time-variant channel with subspace order recursive LS iterator. In *Asia-Pacific Conference on Communications*. Perth (Australia), 2005, vol. 2005, p. 921–925. DOI: 10.1109/APCC.2005.1554197
- [31] GUENACH, M., WYMEERSCH, H., MOENECLAEY, M. Joint estimation of path delay and complex gain for coded systems using the EM algorithm. In *International Zurich Seminar on Digital Communications*. 2004, p. 216–219. DOI: 10.1109/IZS.2004.1287428
- [32] SYKORA, J., BURR, A. G. Iterative decoding networks with iteratively data eliminating SDD and EM based channel state estimator. In *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. 2004, vol. 2, p. 785–790. DOI: 10.1109/PIMRC.2004.1373807
- [33] NOELS, N., HERZET, C., DEJONGHE, A., LOTTICI, V., STEENDAM, H., MOENECLAEY, M., LUISE, M., VANDENDORPE, L. Turbo synchronization: An EM algorithm interpretation. In *IEEE International Conference on Communications ICC 2003*. Anchorage (Alaska, USA), 2003, vol. 4, p. 2933–2937. DOI: 10.1109/ICC.2003.1204575
- [34] RAMON, V., HERZET, C., VANDENDORPE, L., MOENECLAEY, M. EM algorithm-based multiuser synchronization in turbo receivers. In *IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP 2004*. Montreal (Canada), 2004, vol. 4, p. 849–852. DOI: 10.1109/ICASSP.2004.1326960
- [35] FIGUEIREDO, M. A. T., BIOUCAS-DIAS, J. M., NOWAK, R. D. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Transactions on Image Processing*, Dec. 2007, vol. 16, no. 12, p. 2980–2991. DOI: 10.1109/TIP.2007.909318
- [36] DELFOSSE, N., LOUBATON, P. Adaptive blind separation of independent sources: a deflation approach. *Signal Processing*, 1995, vol. 45, p. 59–83. DOI:10.1016/0165-1684(95)00042-C
- [37] PROAKIS, J. G. *Digital Communications*. 5th ed. Asia: McGraw-Hill Education Co., 2007.

About the Authors...

Pengcheng XU was born in 1987. He received his M.S. degree from the College of Communications Engineering, PLA University of Science and Technology in 2012. He is currently pursuing the Ph.D. degree in the College of Communications Engineering, PLA University of Science and Technology. His research interests include blind signal processing and wireless communication systems.

Yuehong SHEN was born in 1959. He received his Ph.D. degree in Communication Engineering from Nanjing University of Science and Technology in 1999. And now, he is a Professor and doctor advisor of the College of Communications Engineering, PLA University of Science and Technology. His current research interests include wireless communication systems, digital communications and signal processing.

Hui LI was born in 1986. She received her Bachelor degree and Ph.D. degree from the College of Communications Engineering, PLA University of Science and Technology in 2007 and 2013, respectively. She is currently an engineer in PLA Troop 78020. Her research interests include blind signal processing and wireless communications.