

Article

Semi-Blind Signal Extraction for Communication Signals by Combining Independent Component Analysis and Spatial Constraints

Xiang Wang *, Zhitao Huang and Yiyu Zhou

College of Electronic Science and Engineering, National University of Defense Technology (NUDT), Changsha 410073, China; E-Mails: taldcn@sohu.com (Z.H.); zhouyiyu@sohu.com (Y.Z.)

* Author to whom correspondence should be addressed; E-Mail: christopherwx@163.com; Tel.: +86-1375-5118-941; Fax: +86-731-572-518.

Received: 28 April 2012; in revised form: 15 June 2012 / Accepted: 28 June 2012 /

Published: 2 July 2012

Abstract: Signal of interest (SOI) extraction is a vital issue in communication signal processing. In this paper, we propose two novel iterative algorithms for extracting SOIs from instantaneous mixtures, which explores the spatial constraint corresponding to the Directions of Arrival (DOAs) of the SOIs as a priori information into the constrained Independent Component Analysis (cICA) framework. The first algorithm utilizes the spatial constraint to form a new constrained optimization problem under the previous cICA framework which requires various user parameters, *i.e.*, Lagrange parameter and threshold measuring the accuracy degree of the spatial constraint, while the second algorithm incorporates the spatial constraints to select specific initialization of extracting vectors. The major difference between the two novel algorithms is that the former incorporates the prior information into the learning process of the iterative algorithm and the latter utilizes the prior information to select the specific initialization vector. Therefore, no extra parameters are necessary in the learning process, which makes the algorithm simpler and more reliable and helps to improve the speed of extraction. Meanwhile, the convergence condition for the spatial constraints is analyzed. Compared with the conventional techniques, *i.e.*, MVDR, numerical simulation results demonstrate the effectiveness, robustness and higher performance of the proposed algorithms.

Keywords: independent component analysis (ICA); constrained ICA; signal extraction; spatial constraint; initialization

1. Introduction

The problem of blind source separation arises in a wide range of application fields, such as speech processing [1], image analysis [2], medical diagnosis [3] and wireless communication [4], *etc.* Suppose that there exist M independent source signals $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_M(n)]^T$ and N observed mixtures of the source signals $\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_N(n)]^T$. The linear instantaneous model of Blind Source Separation (BSS) is as follows:

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{v}(n) \quad (1)$$

where \mathbf{A} is an $N \times M$ mixing matrix and $\mathbf{v}(n) = [v_1(n), v_2(n), \dots, v_N(n)]^T$ denotes the noise vector. The goal of BSS is to determine the original sources $\mathbf{s}(n)$ from their mixtures $\mathbf{x}(n)$ without any other a priori knowledge being necessary.

However, some applications of BSS often wish to extract only one signal of interest (SOI) or a desired subset of sources and automatically discard uninteresting sources: for example, the extraction of “interesting” signal from the interference signals in communication application. In such cases, the BSS problem reduces to a blind signal extraction (BSE) problem.

Generally speaking, two kinds of promising techniques have been proposed to address the BSE problem. One is beamforming, which attempts to cover specific cell sectors so that the signal of interest (SOI) can be extracted while suppressing other signals. Traditional beamforming techniques such as MVDR, LCMV [5] are based on the accurate knowledge of the direction vector associated to the SOI and the perfect array calibration, both of which are not often available in practice. Another technique is independent component analysis (ICA), which is perhaps most widely used for performing BSS. ICA attempts to exploit the assumed mutual statistical independence of the source components to estimate the mixing matrix and/or the source signals which can have an arbitrary permutation of the original sources [6]. Many existing ICA algorithms recover the source signals simultaneously whose number is same as that of the observed mixtures, while in many practical applications, it is not necessary for us to extract all the source signals since the number of desired signals needed to be recovered is less than that of the mixtures. Therefore, classical ICA algorithms involve redundant computation, require large memory for estimating uninteresting signals, degrade the quality of the signals recovered and need complex post-processing to detect and identify sources of interest. Furthermore, Hiroshi *et al.* proposed a new algorithm [7,8] combining subband ICA and beamforming to solve convolutive blind source separation problem in frequency domain, which not only helps to solve arbitrariness of permutation and gain problem at each frequency bin, but also improve the separation performance. In [9], the geometric source separation (GSS) algorithm, which combines the optimization criteria of source separation, while constraining the responses of multiple beams based on readily available geometric information, can be used to extract the SOIs while reducing undesired interferences. However, it is a symmetric algorithm which recovers the source signals simultaneously whose number is same as that of the observed mixtures, though often in the

BSE problem, especially when the number of components needed to be recovered is much less, the one-unit (deflation) scheme is recommended. Although some one-unit algorithms were proposed to extract all source signals one by one with a deflation process, the inefficiency of such techniques and the arbitrary order of extraction remain as major drawbacks. Elimination of indeterminacy in the ICA not only resolves the aforementioned problems but facilitates more applications by providing stable and unique solutions.

Recently, Lu and Rajapakse proposed a new technique of constrained independent component analysis (cICA) [10], which incorporates the *a priori* information as additional constraints into the conventional ICA learning process and means that only a single statistically independent component will be extracted for each given constraint, which can help to reduce the dimensionality of the output of the ICA method. The *a priori* information in [11,12] are rough templates or reference signals of the desired signals, and the constraints are denoted by the correlation measure between the recovered signals and their corresponding reference signals. James *et al.* have applied with great success the cICA method to artifact rejection in EEG/MEG signal analysis [13]. Lee *et al.* used the cICA method for the extraction of fetal abdominal ECGs [3]. In practice, it is difficult for us to accurately compensate the time delay between the recovered signal and the reference signal in order to make the phase between them be closely matched, let alone that the reference signals are not available in communication applications Mitianoudis *et al.* [14] proposed a new cICA method exploiting the smoothness constraint to extract the smooth source signals with slowly varying temporal structures. Furthermore, the previous cICA methods generally view the *a priori* information as inequality constraints and transform the BSE problem into a constrained optimization problem. It means that the *a priori* information has been incorporated into the learning process to guarantee the algorithms converge to the desired solution. Yet, all of these algorithms need a very important parameter, e.g., closeness tolerance in [11–13] and smoothness degree in [14], to measure the corresponding constraint. Unfortunately, it is not easy for us to obtain a suitable candidate, which needs a great deal of trial.

We notice that, in communication applications with antenna arrays, the mixing matrix is closely related to the Direction of Arrivals (DOAs) of the narrowband source signals and the advantage of the beamforming technique over the source separation method lies in its use of geometric information. The *a priori* information about the structure of selected source sensor projections is often readily available and can similarly be used as a reference or spatial constraint, which is associated with the DOA of the SOI as well as in the beamforming theory. This paper is concerned with the use of spatial constraints in the cICA method and aims to introduce a novel method by incorporating the rough spatial knowledge into the initialization of the ICA method, instead of its learning process. Our purpose is to describe how the computationally fast and efficient, the ICA algorithm can be adapted to accommodate spatial constraints and thus provide an improved algorithm. This manuscript is organized as follows: in Section 2, the attenuation delay mixing model is given, along with the assumptions and the notations. Section 3 briefly reviews the ICA concept and the cICA framework. Section 4 introduces how the spatial constraint can be incorporated into the ICA process and demonstrates its efficacy in extracting the SOI. Section 5 presents the results of a computer simulation compared with the beamforming method and finally Section 6 provides the conclusions and discussion.

2. Problem Formulation, Definition, Assumption and Notation

2.1. Notation

Conventional notation is used in this paper. Scalars, matrices and vectors are represented by lower case, upper case and boldface lower case letters, respectively. The i^{th} component of vector \mathbf{x} is denoted by x_i . The expectation operator is $E\{\cdot\}$, \mathbf{A}^T , \mathbf{A}^* and \mathbf{A}^H denote transpose, complex conjugate, and Hermitian transpose of the matrix \mathbf{A} , respectively. The identity matrix is denoted by \mathbf{I} . Furthermore, $kurt(\cdot)$ denotes the kurtosis operator and $\|\cdot\|$ represents the L2 norm of a vector.

2.2. Problem Formulation

In narrowband (NB) array signal processing, the attenuation delay mixing model is more suitable than the instantaneous mixing model. Suppose N narrowband source signals impinge on a linear array of M sensors, the i^{th} mixture $x_i(n)$ can be formulated as:

$$x_i(n) = \sum_{k=1}^N b_{ik} s_k(n - \tau_{ik}) + v_i(n) \quad (2)$$

where b_{ik} is the attenuation coefficients, and τ_{ik} denotes the propagation time delays associated with the path from the k^{th} source signal to the i^{th} sensor which can be represented by $c^{-1}d_i \sin \theta_k$ where d_i , c and θ_k denote the position of the i^{th} sensor, the propagation velocity and the DOA of the k^{th} source signal, respectively. According to the NB assumption, Equation (2) can be formulated as complex-valued form:

$$x_i(n) = \sum_{k=1}^N b_{ik} s_k(n) e^{-j\pi f_k c^{-1} d_i \sin \theta_k} + v_i(n) \quad (3)$$

Equation (3) can be rewritten as the matrix form (1), where the mixing matrix can be represented by the DOAs of the source signals $\mathbf{A} = [\mathbf{a}_1(\theta_1), \dots, \mathbf{a}_N(\theta_N)]^T \in \mathbf{C}^{M \times N}$. For instance, in a uniform linear array (ULA) system where the inter-element equals to half of the wavelength of the source signal, the steering vectors in the mixing matrix can be denoted by $\mathbf{a}_k(\theta_k) = [1, e^{-j\pi \sin \theta_k}, \dots, e^{-j(M-1)\pi \sin \theta_k}]^T \in \mathbf{C}^{M \times 1}$. The model throughout this paper is a uniform linear array.

2.3. Spatial Constraints on the Mixing Matrix

Given the *a priori* knowledge on the rough estimate of DOAs $\hat{\boldsymbol{\theta}} = [\theta_1, \theta_2, \dots, \theta_l]^T$ of the SOIs, we can define a spatial constraint on the mixing matrix, *i.e.*, the spatial constraint regarding the sensor projections of the SOIs operates on selected columns of \mathbf{A} and are enforced with reference to a set of predetermined constraint sensor projections, denoted by \mathbf{A}_c . Thus, the spatially constrained mixing matrix comprises two types of columns [15]:

$$\mathbf{A} = [\hat{\mathbf{A}}_c \ \mathbf{A}_u] \quad (4)$$

where $\hat{\mathbf{A}}_c \approx \mathbf{A}_c$ are columns subject to the constrained and can be written as follows:

$$\hat{\mathbf{a}}_i = [1 \ e^{-j\pi \hat{f}_i c^{-1} d_2 \sin \hat{\theta}_i} \ \dots \ e^{-j\pi \hat{f}_i c^{-1} d_N \sin \hat{\theta}_i}] \quad i = 1, 2, \dots, l \quad (5)$$

and \mathbf{A}_u are otherwise unconstrained columns.

To reflect the degree of certainty about the accuracy of the constrained topographies \mathbf{A}_c and the extent to which $\widehat{\mathbf{A}}_c$ may diverge from \mathbf{A}_c , we make the definition of the inverse of the mixing matrix as follows:

$$\mathbf{A}^{-1} = [\tilde{\mathbf{a}}_1^H \tilde{\mathbf{a}}_2^H \cdots \tilde{\mathbf{a}}_N^H] \quad (6)$$

In this manuscript, the accurate degree of the constrained column $\hat{\mathbf{a}}_i$ for the desired signal $s_i(n)$ contrast with sensor projections for other signal $s_j(n)$ is defined as follows:

$$AD(\hat{\mathbf{a}}_i | \tilde{\mathbf{a}}_j) = \left| \frac{\tilde{\mathbf{a}}_j^H \hat{\mathbf{a}}_i}{\tilde{\mathbf{a}}_j^H \tilde{\mathbf{a}}_i} \right| \quad j = 1, 2, \dots, N \quad j \neq i \quad (7)$$

2.4. Assumption

In our implementation of cICA with the spatial knowledge, we make assumptions that are in keeping with the general assumptions governing the application of ICA. In particular, we assume the following:

AS1: All the source signals are independent from each other.

In practice, this assumption is not strict and easy to satisfy.

AS2: The SOI is not Gaussian.

Most digital communication signals can be considered as sub-Gaussian and therefore this assumption is also within reason.

AS3: The number of sensors is assumed to be identical to that of source signals for simplicity and the mixing matrix \mathbf{A} is of full-rank.

This assumption is necessary for preventing the BSS problem from becoming an underdetermined case which requires other separation methods. As long as there are no two signals whose frequencies and DOAs equal to each other completely at the same time, the mixing matrix can be considered as full-rank.

AS4: The accuracy degree of each spatial reference or constraint satisfies the following condition:

$$AD(\hat{\mathbf{a}}_i | \tilde{\mathbf{a}}_j) > \max \left(1, \left\| \frac{kurt(s_j)}{kurt(s_i)} \right\|^{1/2} \right) \quad j = 1, 2, \dots, N \quad i \neq j \quad (8)$$

Remark: The accuracy degree required in the following analysis is related to the “non-Gaussianity” of the source signals and the mixing vectors corresponding to other “uninteresting” signals. Here, we use kurtosis as the measurement of non-Gaussianity of the source signals for simplicity.

3. Constrained Independent Component Analysis

3.1. Independent Component Analysis

As ICA is a building-block in the cICA algorithm, we start with a short description. ICA is a statistical method for transforming an observed multidimensional random vector into components that

are statistically as independent from each other as possible. Thus the starting point for ICA is the very simple assumption that the components (called source signals in BSS) are statistically independent. ICA attempts to find an $N \times M$ demixing matrix \mathbf{W} to recover source signals as follows:

$$\hat{\mathbf{s}}(n) = \mathbf{W}\mathbf{x}(n) = \mathbf{W}\mathbf{A}\mathbf{s}(n) = \mathbf{P}\mathbf{D}\mathbf{s}(n) \quad (9)$$

where $\mathbf{P} \in \mathbf{C}^{N \times N}$ is a permutation matrix, and $\mathbf{D} \in \mathbf{C}^{N \times N}$ is a diagonal scaling matrix. Consequently, the source signals are recovered up to scaling and permutation ambiguities. In the past decade, a variety of ICA algorithms based on different methodologies and theories have been widely studied, like Bell and Sejnowski's infomax algorithm [16], Cardoso's joint approximate diagonalization of eigen matrices (JADE) [17], Hyvärinen's fixed point algorithm [18,19], *etc.* (for more recent work see [20–22]). We will restrict ourselves to the exploitation of the FastICA algorithm, mainly because of its ease of implementation and speed of separation. The FastICA algorithm is a two step procedure where the mixture data are whitened in the first step, and in the next step, unitary separate vector is updated in order to produce independent output components. Whitening means that the original observed data \mathbf{x} is linearly transformed to vectors $\mathbf{z} = \mathbf{V}\mathbf{x}$, such that the correlation matrix of \mathbf{z} equals unity: $E\{\mathbf{z}\mathbf{z}^H\} = \mathbf{I}$, which can reduce the search for an extracting vector to the group of vectors with unity norm and orthogonal to each other. This transformation can be accomplished by a principle component analysis (PCA) step. The FastICA algorithm separates the source signals based on the “non-Gaussianity” of the recovered source signals, which is measured by kurtosis or negentropy. In the FastICA algorithm for complex valued signals, the flexible and reliable approximation of negentropy was introduced in [19] as follows:

$$J_1(\mathbf{w}) = E\left\{G\left(|\mathbf{w}^H \mathbf{z}|^2\right)\right\} \quad (10)$$

where $G(\cdot)$ is a non-quadratic function.

3.2. cICA Framework

cICA can eliminate the indeterminacy of classical ICA on permutation and consequently get the unique result by incorporating the additional requirements and available *a priori* information [10]. The general framework of cICA can be expressed as follows:

$$\begin{aligned} & \max_{\mathbf{w}} J_1(\mathbf{w}) \\ & s.t. \quad \mathbf{J}_2(\mathbf{w}) \leq 0 \\ & \quad \mathbf{J}_3(\mathbf{w}) = 0 \\ & \quad J_4(\mathbf{w}) = \|\mathbf{w}\|^2 - 1 = 0 \end{aligned} \quad (11)$$

where $J_1(\mathbf{w})$ is the contrast function which can be defined by the independence between estimated signals, while $J_2(\mathbf{w}) = [J_{21}(\mathbf{w}), \dots, J_{2m}(\mathbf{w})]^T$ and $J_3(\mathbf{w}) = [J_{31}(\mathbf{w}), \dots, J_{3n}(\mathbf{w})]^T$ are inequality and equality constraints denoting the available *a priori* information which of course can be combined into the contrast function or exploited in the initialization of the solution vector [23] as described in Section 4.2. Therefore, $J_2(\mathbf{w})$ and $J_3(\mathbf{w})$ are not often necessary according to different application. $J_4(\mathbf{w})$ indicates constrains the norm of extracting vector \mathbf{w} to be unity due to the fact that the variance of recovered sources must be constrained to unity for whitened data.

4. The Proposed Algorithm with Spatial Constraint

Firstly, we derive a new cICA algorithm with the spatial constraint by using the gradient ascent method and then we will propose a novel method by incorporating the spatial knowledge into the initialization of the extracting vector instead of the learning process. Therefore, no extra parameters are involved in the algorithm, which is superior to the previous algorithms.

4.1. Conventional Approach with Spatial Constraint

If the minimum of half the wavelength of all source signals is longer than the sensor spacing, there is no spatial aliasing. In most such cases, the desired solution \mathbf{w}_i forms spatial nulls in the directions of jammer signals and extracts the SOI in another direction [5]. Thus, the extraction system for the SOI $s_i(n)$ can be viewed as an impulse response from the mixtures to its estimate $\hat{s}_i(n)$, which can be written as follows:

$$u_i(\theta_i) = \mathbf{w}_i^H \mathbf{h}_i = \mathbf{w}_i^H (\mathbf{V} \mathbf{a}_i) = \sum_{j=1}^N w_{ij}^* \sum_{k=1}^N v_{ik} e^{-j\pi(k-1)\sin\theta_i} \quad (12)$$

In beamforming theory, the impulse response is thus called a directivity pattern, which can measure the spatial constraint. If the i^{th} row in the separating matrix produces the “interesting” source signal originating from the direction θ_i , it should maximize the gain of $|u_i(\theta_i)|$, that is:

$$|\mathbf{w}_i^H \mathbf{V} \hat{\mathbf{a}}_i| > \gamma > |\mathbf{w}_k^H \mathbf{V} \hat{\mathbf{a}}_i| \quad k = 1, 2, \dots, N \quad k \neq i \quad (13)$$

where γ is a threshold measuring accuracy degree of the spatial constraint. Thus, an inequality constraint can be defined for the desired output component with the directivity pattern at θ_i more than or equal to a threshold γ , that is:

$$J_2(\mathbf{w}) = \gamma - |\mathbf{w}^H \mathbf{V} \hat{\mathbf{a}}_i| \leq 0 \quad (14)$$

Similarly, the problem of spatial constrained ICA can be modeled in the cICA framework as a constrained optimization problem:

$$\begin{aligned} \max_{\mathbf{w}} J_1(\mathbf{w}) \\ \text{s.t. } J_2(\mathbf{w}) = \gamma - |\mathbf{w}^H \mathbf{V} \hat{\mathbf{a}}_i| \leq 0 \\ J_4(\mathbf{w}) = \|\mathbf{w}\|^2 - 1 = 0 \end{aligned} \quad (15)$$

Firstly, we replace the inequality constraint by the equality constraint $\max(J_2(\mathbf{w}), 0) = 0$ for simplicity and then a neural algorithm using the augmented Lagrange multipliers method and the gradient ascent learning approach can be derived to obtain the desired optimal solution. The corresponding Lagrangian function $L(\mathbf{w}, \lambda)$ is given by:

$$L(\mathbf{w}, \lambda) = J_1(\mathbf{w}) - \lambda \max(J_2(\mathbf{w}), 0) \quad (16)$$

The unit-norm constraint in Equation (15) is enforced by the projection of the estimated \mathbf{w} on the unit-sphere in each iteration, that is:

$$\mathbf{w} = \mathbf{w} / \|\mathbf{w}\| \quad (17)$$

Following the strategy proposed in [24], the above optimization problem is addressed using alternative optimization. That is, given the current estimate λ , a new estimate for \mathbf{w} is searched, and then given the estimated for \mathbf{w} , we update λ . The cICA algorithm reported here search for the optimal solution of \mathbf{w} and the Lagrangian parameter λ by using conventional gradient descent for complex variable [24]:

$$\begin{aligned}\mathbf{w} &\leftarrow \mathbf{w} + \eta_1 \nabla_{\mathbf{w}} L \\ \lambda &\leftarrow \lambda - \eta_2 \max(J_2(\mathbf{w}), 0)\end{aligned}\quad (18)$$

where η_1, η_2 are the corresponding learning rate and $\nabla_{\mathbf{w}} L$ denotes the gradient vector of $L(\mathbf{w}, \lambda)$ (See the Appendix I):

$$\nabla_{\mathbf{w}} L = \mathbb{E} \left\{ \mathbf{z} (\mathbf{w}^H \mathbf{z})^* g(|\mathbf{w}^H \mathbf{z}|^2) \right\} + \frac{\lambda}{2} (\text{sign}(J_2) + 1) \left[\mathbf{V} \hat{\mathbf{a}}_i (\mathbf{w}^H \mathbf{V} \hat{\mathbf{a}}_i)^* \right] \quad (19)$$

where $g(\cdot)$ denotes the first-order derivative of $G(\cdot)$. In fact the selection of the learning rate (step size) is a crucial point and it has been a research focus for several decades where various step size selection schemes have been developed (see for example [25,26]).

If there is a subset of SOIs with the same spatial constraint, we need to run the same procedure by re-initializing the extracting vector \mathbf{w}_i in order to identify the whole subset of SOIs. To prevent different vectors from converging to the same independent component, we must decorrelate the outputs $\mathbf{w}_1^H \mathbf{z}, \mathbf{w}_2^H \mathbf{z}, \dots$. As introduced in [6], there are two varieties of the FastICA algorithm: the one-unit, or deflation algorithm and the symmetric algorithm. So does our algorithm for extracting a desired subset of SOIs with the same cyclic frequencies. The one-unit approach estimates the source signals successively under orthogonality condition, while the symmetric algorithm estimates all the source signals in parallel and each step is completed by a symmetric orthogonalization of the extracting matrix. In the one-unit approach, the i^{th} extracting vector \mathbf{w}_i can be orthogonal to the space spanned by the vector $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{i-1}$, by Gram-Schmidt method, that is $\mathbf{w}_i = \mathbf{w}_i - \sum_{j=1}^{i-1} (\mathbf{w}_j^H \mathbf{w}_i) \mathbf{w}_j$.

In the symmetric algorithm the symmetric orthogonalization procedure can be approximately finished by $(\mathbf{W}\mathbf{W}^H)^{-1/2}\mathbf{W}$. Therefore, the one-unit and symmetric version of the proposed algorithm with spatial constraint are summarized in Algorithms 1 and 2 respectively. We refer them to as Alg 1 and Alg 2 in the later analysis for simplicity.

Algorithm 1. The one-unit extracting algorithm with spatial constraint.

Initialization

Whitened the observation data \mathbf{x} to give $\mathbf{Z} = \mathbf{V}\mathbf{x}$;

for $P = 1, \dots, l$

Set $\lambda^{(0)}, \eta_1, \eta_2$ and choose a random initial weight vector $\mathbf{w}^{(0)}$ with unity norm

Iteration

At the i^{th} iteration for obtaining \mathbf{W}_p ,

Calculate $\Delta_{\mathbf{w}} \zeta$ according to Equation (19) by utilizing $\mathbf{W}_p^{(i-1)}$ respectively

$$\mathbf{w}_p^{(i)} \leftarrow \mathbf{w}_p^{(i-1)} + \eta_1 \nabla_{\mathbf{w}} \zeta \left(\mathbf{w}_p^{(i-1)} \right)$$

Algorithm 1. Cont.

$$\mathbf{w}_p^{(i)} \leftarrow \mathbf{w}_p^{(i)} - \sum_{j=1}^{p-1} \mathbf{w}_j^H \mathbf{w}_p^{(i)} \mathbf{w}_j$$

$$\mathbf{w}_p^{(i)} \leftarrow \mathbf{w}_p^{(i)} / \|\mathbf{w}_p^{(i)}\|$$

$$\lambda^{(i)} \leftarrow \lambda^{(i-1)} - \eta_2 \max\left(J_2(\mathbf{w}_p^{(i)}), 0\right)$$

Termination

The iteration is terminated when the relative change $\|\mathbf{w}_p^{(i)} - \mathbf{w}_p^{(i-1)}\|$ is less than a specified tolerance.

end for

Algorithm 2. The symmetric extracting algorithm with spatial constraint.

Initialization

Whitened the observation data \mathbf{x} to give $\mathbf{z} = \mathbf{V}\mathbf{x}$;

Set $\lambda^{(0)}$, η_1 , η_2 and choose a random initial weight matrix $\mathbf{W}^{(0)} = [\mathbf{w}_1^{(0)}, \dots, \mathbf{w}_l^{(0)}]$ with $\mathbf{w}_l^{(0)}$ having unity norm

Iteration

At the i^{th} iteration for obtaining \mathbf{W} ,

for $P = 1, \dots, l$

Calculate $\nabla_{\mathbf{w}} \zeta$ according to Equation (19) by utilizing $\mathbf{w}_p^{(i-1)}$ respectively

$$\mathbf{w}_p^{(i)} \leftarrow \mathbf{w}_p^{(i-1)} + \eta_1 \nabla_{\mathbf{w}} \zeta(\mathbf{w}_p^{(i-1)})$$

end for

$$\mathbf{W}^{(i)} \leftarrow \left(\mathbf{W}^{(i)} \left(\mathbf{W}^{(i)} \right)^H \right)^{-1/2} \mathbf{W}^{(i)}$$

$$\lambda^{(i)} \leftarrow \lambda^{(i-1)} - \eta_2 \max\left(\min\left(J_2(\mathbf{w}_p^{(i)})\right), 0\right) \quad p = 1, \dots, l$$

Termination

The iteration is terminated when the relative change $\|\mathbf{W}^{(i)} - \mathbf{W}^{(i-1)}\|$ is less than a specified tolerance.

Yet Alg1 and Alg2 require a user parameter which may affect the final results significantly. The selection of the threshold in the algorithm is of vital importance for extracting the desired signal successfully, which can be found in Section 5 by simulation. Furthermore, the update step of the Lagrangian parameter in each iteration will increase the computational load of the algorithm. Therefore, it is important to develop user parameters free methods.

4.2. A Novel Method

If the number of source signals is N , there will be $2N$ local maxima of negentropy, each one of which corresponds to $\pm s_i(n)$. The FastICA algorithm cannot theoretically obtain particular desired

independent sources other than those having the maximum negentropy among the sources. Furthermore, as we know, the FastICA algorithm is a local optimization algorithm which may arbitrarily converge to different local maxima from time to time because the local convergence depends on a number of factors such as the initial weight vector and the learning rate. When one desires a specific solution, the FastICA algorithm is of little use, unless the “interesting” independent source lies in the neighborhood of the initialization. Therefore, as long as we predispose the initial \mathbf{w}_0 in the neighborhood of the SOI by utilizing the spatial constraint, the algorithm will automatically converge to it. In this case, the specific initial \mathbf{w}_0 for the SOI $s_i(n)$ is obtained based on the maximization of the directivity pattern corresponding to the spatial information $\hat{\mathbf{a}}_i$ as follows:

$$\begin{aligned} \max & \left| \mathbf{w}_0^H \mathbf{V} \hat{\mathbf{a}}_i \right| \\ \text{s.t.} & \left\| \mathbf{w}_0 \right\|^2 - 1 = 0 \end{aligned} \quad (20)$$

The Lagrange multiplier method is adopted to obtain the optimal solution of Equation (20). The corresponding Lagrangian function is given by:

$$L(\mathbf{w}_0, \mu) = \left| \mathbf{w}_0^H \mathbf{V} \hat{\mathbf{a}}_i \right| - \mu \left(\left\| \mathbf{w}_0 \right\|^2 - 1 \right) \quad (21)$$

where μ is Lagrangian parameter. Let $\nabla_{\mathbf{w}_0} L(\mathbf{w}_0, \mu) = 0$, we have:

$$\mathbf{w}_0 = \mathbf{V} \hat{\mathbf{a}}_i \quad (22)$$

Since \mathbf{w}_0 is on the unit sphere, the result is:

$$\mathbf{w}_0 = \mathbf{V} \hat{\mathbf{a}}_i / \left\| \mathbf{V} \hat{\mathbf{a}}_i \right\| \quad (23)$$

The whole estimation of initial \mathbf{w}_0 for $s_i(n)$ via maximizing the corresponding directivity pattern does not need learning, so it is easy to obtain. In [6], Hyvarinen and Oja have shown that if the initial \mathbf{w}_0 is located in the neighborhood of \mathbf{w}_i which is the desired projection direction to extract $s_i(n)$, the learning process will automatically converge to $s_i(n)$. Therefore, the one-unit and symmetric algorithm with purpose-designed initialization under the spatial constraint are summarized in Algorithms 3 and 4 respectively. We refer them to as Alg 3 and Alg 4 respectively in the later analysis for simplicity.

Algorithm 3. The one-unit version of the cICA algorithm with the purpose-designed initialization.

Initialization

Whitened the observation data \mathbf{x} to give $\mathbf{z} = \mathbf{V}\mathbf{x}$;

for $p = 1, \dots, l$

Compute the specific initial weight vector $\mathbf{w}_p^{(0)}$ corresponding the spatial constraint, that is $\mathbf{w}_p^{(0)} = \mathbf{V} \hat{\mathbf{a}}_p / \left\| \mathbf{V} \hat{\mathbf{a}}_p \right\|$.

Iteration

At the i^{th} iteration for obtaining \mathbf{w}_p ,

$$\begin{aligned} \mathbf{w}_p^{(i)} \leftarrow & \mathbb{E} \left\{ \mathbf{z} \left(\left(\mathbf{w}_p^{(i-1)} \right)^H \mathbf{z} \right)^* \mathcal{G} \left(\left| \left(\mathbf{w}_p^{(i-1)} \right)^H \mathbf{z} \right|^2 \right) \right\} - \\ & \left[\mathbb{E} \left\{ \mathcal{G} \left(\left| \left(\mathbf{w}_p^{(i-1)} \right)^H \mathbf{z} \right|^2 \right) \right\} + \left| \left(\mathbf{w}_p^{(i-1)} \right)^H \mathbf{z} \right|^2 \mathcal{G}' \left(\left| \left(\mathbf{w}_p^{(i-1)} \right)^H \mathbf{z} \right|^2 \right) \right] \mathbf{w}_p^{(i-1)} \end{aligned}$$

Algorithm 3. Cont.

$$\mathbf{w}_p^{(i)} \leftarrow \mathbf{w}_p^{(i)} - \sum_{j=1}^{p-1} \mathbf{w}_j^H \mathbf{w}_p^{(i)} \mathbf{w}_j$$

$$\mathbf{w}_p^{(i)} \leftarrow \mathbf{w}_p^{(i)} / \|\mathbf{w}_p^{(i)}\|$$

Termination

The iteration is terminated when the relative change $\|\mathbf{w}_p^{(i)} - \mathbf{w}_p^{(i-1)}\|$ is less than a specified tolerance.

end for

Algorithm 4. The symmetric version of the cICA algorithm with the purpose-designed initialization.

Initialization

Whitened the observation data \mathbf{x} to give $\mathbf{z} = \mathbf{V}\mathbf{x}$;

Compute the specific initial matrix $\mathbf{W}^{(0)} = [\mathbf{w}_1^{(0)}, \dots, \mathbf{w}_l^{(0)}]$ corresponding the spatial constraints with $\mathbf{w}_p^{(0)} = \mathbf{V}\hat{\mathbf{a}}_p / \|\mathbf{V}\hat{\mathbf{a}}_p\|$, $p = 1, \dots, l$

Iteration

At the i^{th} iteration for obtaining \mathbf{W} ,

for $p = 1, \dots, l$

$$\mathbf{w}_p^{(i)} \leftarrow \mathbb{E} \left\{ \mathbf{z} \left(\left(\mathbf{w}_p^{(i-1)} \right)^H \mathbf{z} \right)^* \mathcal{G} \left(\left| \left(\mathbf{w}_p^{(i-1)} \right)^H \mathbf{z} \right|^2 \right) \right\} - \left[\mathbb{E} \left\{ \mathcal{G} \left(\left| \left(\mathbf{w}_p^{(i-1)} \right)^H \mathbf{z} \right|^2 \right) \right\} + \left| \left(\mathbf{w}_p^{(i-1)} \right)^H \mathbf{z} \right|^2 \mathcal{G}' \left(\left| \left(\mathbf{w}_p^{(i-1)} \right)^H \mathbf{z} \right|^2 \right) \right] \mathbf{w}_p^{(i-1)}$$

end for

$$\mathbf{W}^{(i)} \leftarrow \left(\mathbf{W}^{(i)} \left(\mathbf{W}^{(i)} \right)^H \right)^{-1/2} \mathbf{W}^{(i)}$$

Termination

The iteration is terminated when the relative change $\|\mathbf{W}^{(i)} - \mathbf{W}^{(i-1)}\|$ is less than a specified tolerance.

In summary, our new method is similar to the FastICA algorithm, including whitening, choosing an appropriate non-quadratic for contrast function, as well as the learning process for the optimization problem except for the initialization of the extracting vectors. The following theorem indicates the fact that the specific initialization obtained by Equation (23) can guarantee the learning process to the desired local maxima under certain condition.

Theorem 1: If the spatial constraint satisfies the AS4, that is to say Equation (8) comes into existence, the Alg 3 and Alg 4 must converge to the desired signal corresponding to the spatial constraint automatically.

Proof: See Appendix II.

The theorem shows that higher accuracy of the information about the DOA is required to dispose the initial vector in the neighborhood of the solution when the non-gaussianity of the SOI is weak. In general, most communication signals are sub-Gaussian signals. Therefore, the requirement of the prior information about the DOA of the SOI is not very strict, which can be found in Section 5 by simulation.

As we know, the conventional beamforming (CBF) technique can extract the signal in the desired direction and reject all other signals in other directions, however, there are two major drawbacks: (1) the “uninteresting” source signals out of the beam can only be suppressed to some extent by the sidelobes while some “uninteresting” source signals in the beam cannot be removed, both of which are influenced by the spatial resolution; (2) The *a priori* information about the DOAs for the desired signals should be as accurate as possible. While the proposed algorithms (Alg 1–4) relax the accuracy requirement for the spatial information and may achieve better performance since they exploit the independence property of the source signals.

On the other hand, the Alg 3 and Alg 4 have obvious advantages over the conventional cICA method (*i.e.*, Alg 1 and Alg 2) in the computation load since the update of Lagrangian parameter is not involved in the learning process. Furthermore, the Alg 3 and Alg 4 are more reliable and stable due to the fact that they are user parameters free algorithms without choosing an approximate parameter measuring the corresponding constraint. We can easily see that the Alg 2 reduces to the Alg 1 and the Alg 4 reduces to the Alg 3 when there is only one SOI. The comparison results will be demonstrated in Section 5 by simulation.

5. Simulation Experiments

To evaluate the performance of the proposed algorithm, we adopt the average signal to interference ratio (SIR). SIR in decibels can be defined as follows:

$$\text{SIR} (dB) = 10 \log_{10} \left(\frac{\text{E}\{s^2(n)\}}{\text{E}\{(s(n) - \hat{s}(n))^2\}} \right) \quad (24)$$

Experiment 1: The performance of the proposed algorithm and comparison with the FastICA algorithm and the MVDR algorithm.

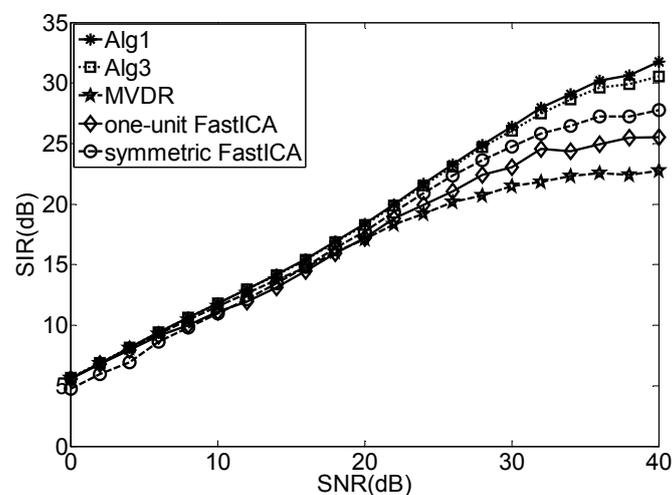
In this experiment, we used four QPSK modulated signals with the carrier frequencies 12, 12.01, 12.02, 12.03 MHz, and the symbol rates 1, 2, 3, 4 Mbps, respectively. The Directions of Arrival (DOAs) of the sources were set 10^0 , 30^0 , 50^0 , and 70^0 . The signals are received by a uniform linear array with four sensors without considering the imperfect array calibration. Without loss of generality, we assume the signal with the carrier frequency 12 MHz and DOA 10^0 to be the SOI. We chose a random initialization of \mathbf{w} , $\eta_1 = 1$, $\eta_2 = 0.01$ and the initialization of Lagrange parameter $\lambda = 10$ in the Alg 1, while the threshold γ was set 2. In addition, the non-quadratic function $G(u) = \sqrt{0.1 + u}$ was used in the adaptation of \mathbf{w} in both of the Alg 1 and Alg 3. More selection for $G(u)$ can be found in [6]. The experiment was repeated 100 times with fixed length of data samples 12,500. The following experiments will exploit the same scenario and the parameter settings without extra statements. Moreover, the FastICA algorithm identified the SOI after separating all the source signals and the extracting vector corresponding to the desired signal can be obtained by:

$$\hat{i} = \arg \max_k |\hat{\mathbf{w}}_k \mathbf{V} \hat{\mathbf{a}}_i| \quad k = 1, \dots, N \quad (25)$$

where $\hat{\mathbf{w}}_k$ ($k = 1, \dots, N$) is the columns of the separating matrix estimated by the FastICA algorithm.

Figure 1 shows the SIRs for different input SNRs by using Alg 1, Alg3, MVDR and the one-unit and symmetric FastICA algorithm. The SIRs of the “interesting” signals of the FastICA algorithm are worse than those of Alg 1 and Alg 3 since the FastICA algorithm attempts to separate all the other “uninteresting” components before identifying the desired signal. Again, the proposed algorithms outperform the conventional beamforming technique as our technique utilizes the independence between the source signals while the beamforming technique suffers from the leakage or cross-talk problem.

Figure 1. Comparison of the performance by using the algorithm of the Alg 1, Alg 3, MVDR, one-unit FastICA and symmetric FastICA.

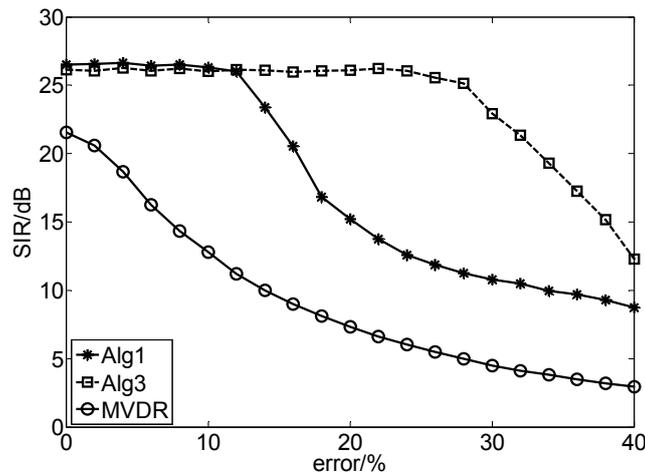


Experiment 2: The performance comparisons among the algorithm of Alg 1, Alg 3 and MVDR with the error of the spatial constraint.

There are various factors influencing the accuracy of the spatial constraint and in this experiment we consider the following two types: (1) error of the DOA of the desired signal; (2) amplitude and phase mismatch of the array element.

The distorted DOA of the desired signal covered from 10^0 to 12^0 , spaced with 0.1^0 and the error of the DOA can be expressed by $|\theta - \theta_{true}| \times 100/\theta_{true}$ (%). We ran this experiment 100 times for each value of error with fixed length of samples 12,500 when the SNR is 30 dB. The results are depicted in Figure 2. The results show that the proposed method is not sensitive to the error of DOA while the performance of the MVDR algorithm attenuates quickly as the increase of the error due to the fact that our algorithms combine the independence between the source signals and the spatial information of the SOI. Moreover, Alg 3 performs better than Alg 1 due to the different usage of the spatial information. In Alg 1, the spatial constraint has been incorporated into the learning process to guarantee the algorithm to converge to the desired solution, while in Alg 3, the spatial information is utilized to select the specific initial vector of the extracting vector. Therefore, the requirement of the accuracy of the spatial information for Alg 3 is less strict than that for Alg 1.

Figure 2. Comparison of the performance *versus* the error of DOA by using different algorithms.

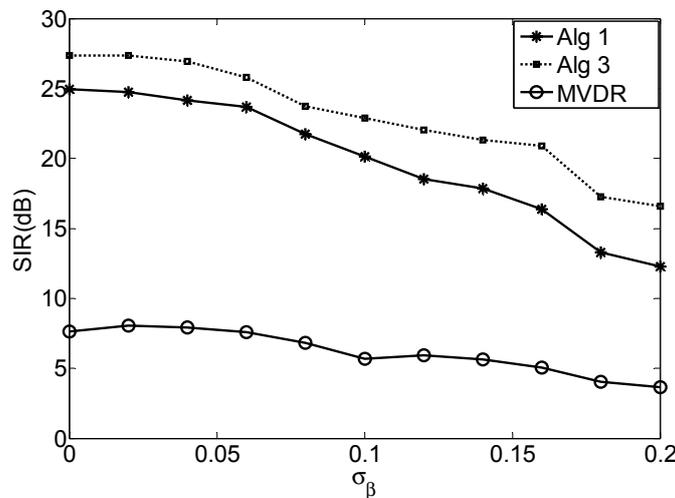


Let Γ be an $M \times M$ diagonal array error matrix defined as $\Gamma = \text{diag}\{\alpha_1 e^{-j\beta_1}, \dots, \alpha_M e^{-j\beta_M}\}$ where α_m and β_m are the gain and the phase of the m^{th} element. We assume that α_m and β_m ($m = 1, \dots, M$) are independent and Gaussian random variables with $\mu_\alpha = 1, \mu_\beta = 0, \sigma_\alpha$ and σ_β , where $\mu_\alpha, \mu_\beta, \sigma_\alpha$ and σ_β are the mean value and the standard deviation of α_m and β_m ($m = 1, \dots, M$), respective. Thus, the mixing matrix defined in Equation (1) can be rewritten as:

$$\mathbf{x}(n) = \Gamma \mathbf{A} \mathbf{s}(n) + \mathbf{v}(n) \tag{26}$$

In this experiment, σ_α is assumed to be 0.01 and σ_β varies from 0 to 0.2, spaced with 0.02. Figure 3 shows the SIR of the algorithms of Alg 1, Alg 3 and MVDR with respect to the standard deviation of the phase errors (σ_β). The results indicate that the proposed algorithms perform more robustly than the MVDR algorithm for a wide range of phase error. Again, Alg 3 outperforms Alg 1.

Figure 3. Comparison of the performance *versus* the variance of the amplitude and phase by using different algorithms.



The requirement of the accuracy of the spatial information for the proposed algorithm is related to various factors including the “non-Gaussianity” of the source signals and the mixing vectors corresponding to other “uninteresting” signals, as stated in Theorem 1.

Experiment 3: The selection principle of the threshold γ in spatial constraint in the algorithm of Alg 1 and Alg 2.

In this experiment, we investigated the selection principle and established the suitable range for the threshold by simulations. We changed the value of the threshold from 0.01 to 50 for fixed Lagrange parameter λ covering from 10 to 20. The SIRs are displayed in Figure 4, where the x-axis and y-axis represent the varying threshold and Lagrange parameter, respectively.

Figure 4. SIRs *versus* varying thresh γ and varying Lagrange parameter λ .

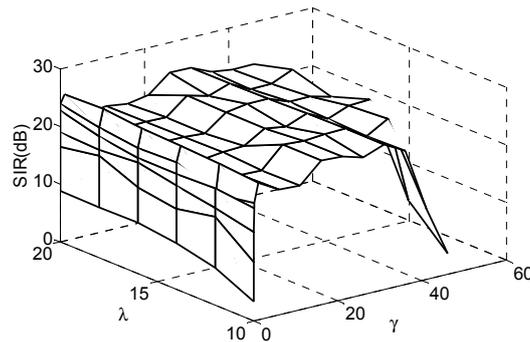


Figure 5 is the slice results of Figure 4 when $\lambda = 10$, and Figure 6 gives the rate of “correct” extraction by setting different thresholds. The results indicate that the reasonable domain for the threshold γ in $J_2(\mathbf{w})$ should be $\gamma \in (1, 36)$. The results demonstrate the key role of the threshold in Alg 1 and Alg 2. If it is beyond some limit, the output component may be unstable to produce any desired signal because the corresponding constraint $J_2(\mathbf{w})$ causes the learning process of that neuron to become unpredictable. If it is too small, the constraint will fail to guide the learning process to converge to the desired solution, since the spatial information of other source signals besides the SOIs may satisfy the constraint $J_2(\mathbf{w})$. Consequently, the algorithm of Alg 1 and Alg 2 cannot produce the desired solution due to the improper parameter selection, which perhaps needs a great deal of trial in different application and will not be a problem by utilizing the algorithm of Alg 3 and Alg 4.

Figure 5. SIRs *versus* varying threshold γ when $\lambda = 10$.

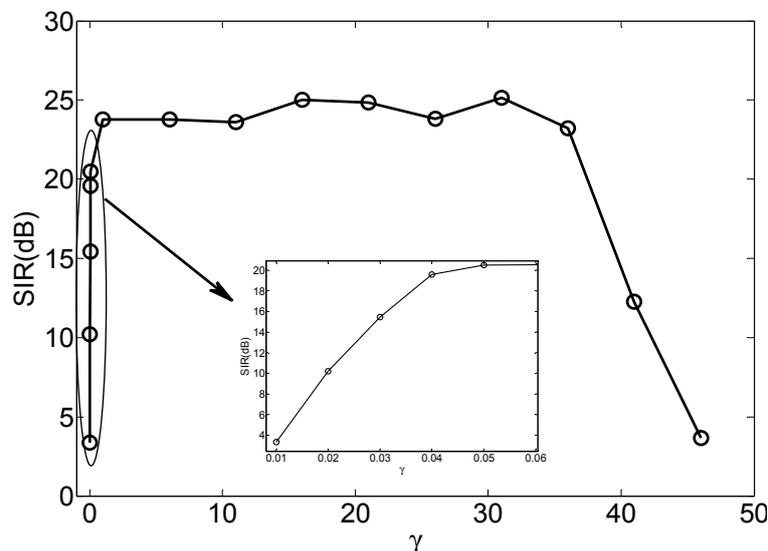
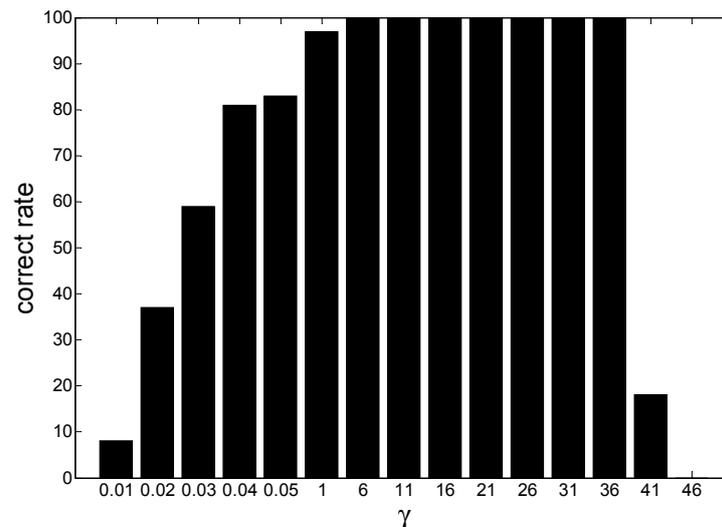
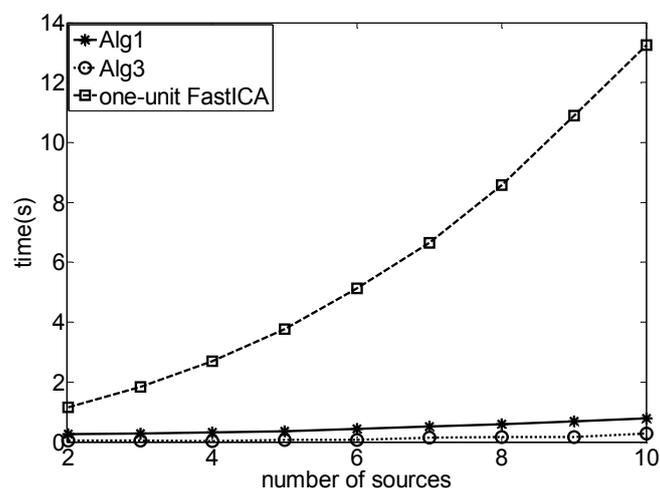


Figure 6. The rate of “correct” extractions *versus* varying threshold γ .

Experiment 4: The comparison of the consuming time between the proposed algorithm and the FastICA algorithm.

In this experiment, we added six QPSK signals with the carrier frequencies from 12.04 to 12.09 MHz spaced with 0.01 MHz. The symbol rates are all 5 Mbps and the DOAs of all the source signals were set -80° , -60° , -40° , -25° , -10° , 0° , 15° , 35° , 55° and 75° respectively. Figure 7 gives the comparison of the time consumed by the different algorithms with different number of source signals through the Matlab execution. The results show that the proposed algorithm runs much faster than the FastICA algorithm in extracting the SOI, because the FastICA algorithm needed redundant computation in estimating unnecessary signals. For the FastICA algorithm, the larger the number of the source signals is, the larger the computation load will be. Consequently, much more time is necessary for extracting the SOI, while the time consumed by the proposed algorithm still maintained. Furthermore, the result also indicates that Alg 1 consumes more time than Alg 3 due to the fact that the update of the Lagrangian parameter and the calculation of the inequality constraint are involved in the learning process of Alg 1, both of which makes the algorithm more complex and slows down the convergence of \mathbf{w} .

Figure 7. Comparison of the performance in the case of different number of the sources.

Experiment 5: Extraction of a subset of SOIs by the algorithm of Alg 1, Alg 2, Alg 3 and Alg 4.

This experiment exploited the same scenario and the same parameters as in Experiment 1 except that there were two desired signals with the DOA 10^0 and 30^0 to be extracted. The measurements indicating the quality of the extraction of the SOIs by using different algorithms are given in Table 1. We should note that the proposed deflationary approach, *i.e.*, Alg 1 or Alg 3, has an inherent error accumulation problem and lower convergence speed. Since we proceed from one phase to the other without complete convergence due to the finite number of iteration existence of noise and nonlinear factors, the assumption about the elimination of previous sources will fail and this will create extra noise effect for the following stage, which will accumulate throughout the multiple stages of the algorithm.

Table 1. Comparison of the performance of extracting two sources of interest by using the algorithm of Alg 1, Alg 3, Alg 3 and Alg 4.

Algorithm	Output	SIR(dB)
Alg 1	S1	28.9574
	S2	26.3267
Alg 2	S1	27.5611
	S2	27.8531
Alg 3	S1	27.6955
	S2	25.4508
Alg 4	S1	27.5901
	S2	27.9003

6. Discussion and Conclusions

In this manuscript, the authors have extended current work in the cICA framework. Spatial constraints/references are associated with sensor projections, whose location in the source mixture model is specified a priori. We incorporated the spatial constraint into the ICA model in different ways and derived two types of iterative algorithms in order to extract the “interesting” signal from the instantaneous mixture while discard other “uninteresting” sources automatically. The first class of iterative algorithms, *i.e.*, Alg 1 and Alg 2, combines the spatial constraints and the gradient descent method, and the second class of iterative algorithms, *i.e.*, Alg 3 and Alg 4, incorporates the spatial constraints into the initialization of the extracting vectors by maximizing the corresponding directivity pattern. The experimental results showed the advantages and superiority of our method compared to the previous methods. Compared with the ICA method, our algorithms perform separation and selection of the SOI simultaneously and eliminate the need for complex post-processing to detect and identify the SOI. Compared with the beamforming techniques, our methods are less sensitive to the considerable error in the DOAs of the desired signals, as well as other factors influencing the accuracy of the sensor projections, such as amplitude and phase mismatch, *etc.* Furthermore, Alg 1 and Alg 2 outperform Alg 3 and Alg 4 in both computation complexity and robustness since no extra parameters such as Lagrangian parameter and threshold are involved in the learning process, which makes the algorithm simpler and more reliable and helps to improve the speed of extraction.

Overall, the application of the cICA method has met with considerable success, especially in EM brain signal analysis, ECG/EMG signal extraction, *etc.* and it demonstrates great potential in the field

of cooperative/non-cooperative communication signal processing. In fact, the cICA method can be considered as very closely related to the method called semi-blind source separation/extraction. There are numerous advantages of the cICA technique over the previous approaches: it produces only the desired independent sources and facilitates subsequent applications; the computation time and storage requirements are reduced; and the incorporation of the *a priori* information improves the quality and accuracy of the separation of the interested components or convergence speed. cICA as applied in the previous literature with temporal constraint, *i.e.*, reference signals or smoothness property, results in a useful technique for the fast and efficient extraction of the desired signals with the corresponding constraint from multichannel recordings, while the spatial constraint is exploited in our manuscript. Furthermore, when the *a priori* information mentioned in the previous literature is not available, new constraints should be explored. Moreover, different constraints can be incorporated in the cICA framework in different ways, including inequality constraints as well as specific initialization described in this manuscript. In conclusion, we need to answer the following two questions when dealing with the cICA problem.

- (i) Which types of the *a priori* information can be exploited in the cICA problem?
- (ii) How can the *a priori* information be exploited in the cICA problem?

Additional work is being carried out to explore new features to validate the algorithm's output automatically and a new manuscript concerned about the overall description on the cICA framework answering the above two questions is being prepared by the authors.

Acknowledgments

This work was supported by the National Science Foundation (No. 61072120), the program for New Century Excellent Talents in University (NCET).

References

1. Smith, D.; Lukasiak, J.; Burnett, I. An analysis of the limitations of blind signal separation application with speech. *Signal Proc.* **2006**, *86*, 353–359.
2. Jiménez-Hernández, H. Background subtraction approach based on independent component analysis. *Sensors* **2010**, *10*, 6092–6114.
3. Lee, J.; Park, K.L.; Lee, K.J. Temporally constrained ICA-based fetal ECG separation. *Electron. Lett.* **2005**, *41*, 1158–1160.
4. Yang, X.N.; Yao, J.L.; Jiang, D.; Zhao, L. Performance analysis of the FastICA algorithm in ICA-Based Co-Channel Communication System. In *Proceedings of the 5th International Conference on Wireless Communication, Network and Mobile Computing*, Beijing, China, 24–26 September 2009; pp. 2055–2058.
5. Johnson, D.H.; Dudgeon, D.E. *Array Signal Processing: Concepts and Techniques*; Simon & Schuster: Parsippany, NJ, USA, 1992.
6. Hyvarinen, A.; Karhunen, J.; Oja, E. *Independent Component Analysis*; Wiley-Interscience: New York, NY, USA, 2001.

7. Saruwatari, H.; Takeda, K.; Itakura, F.; Nishikawa, T.; Shikano, K. Blind source separation combining independent component analysis and beamforming. *EURASIP J. Appl. Signal Proc.* **2003**, *11*, 1135–1145.
8. Saruwatari, H.; Takeda, K.; Nishikawa, T.; Itakura, F.; Shikano, K. Blind source separation based on a fast-convergence algorithm combining ica and beamforming. *IEEE Trans. Audio Speech Lang. Proc.* **2006**, *14*, 666–678.
9. Parra, L.C.; Alvino, C.V. Geometric source separation: Merging convolutive source separation with geometric beamforming. *IEEE Trans. Speech Audio Proc.* **2002**, *10*, 352–362.
10. Lu, W.; Rajapakse, J.C. Approach and applications of constrained ICA. *IEEE Trans. Neural Networks* **2005**, *16*, 203–212.
11. Lu, W.; Rajapakse, J.C. ICA with reference. *Neurocomputing* **2006**, *69*, 2244–2257.
12. Lu, W.; Rajapakse, J.C. Eliminating indeterminacy in ICA. *Neurocomputing* **2003**, *50*, 271–290.
13. James, C.J.; Gibson, O.J. Temporally constrained ICA: An application to artifact rejection in electromagnetic brain signal analysis. *IEEE Trans. Biomed. Eng.* **2003**, *50*, 1108–1116.
14. Mitianoudis, N.; Stathaki, T.; Constantinides, A.G. Smooth signal extraction from instantaneous mixtures. *IEEE Signal Proc. Lett.* **2007**, *14*, 271–274.
15. Hesse, C.W.; James, C.J. The FastICA algorithm with spatial constraints. *IEEE Signal Proc. Lett.* **2005**, *12*, 792–795.
16. Bell, A.J.; Sejnowski, T.J. An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.* **1995**, *7*, 1129–1159.
17. Cardoso, J.F.; Souloumiac, A. Blind beamforming for non-Gaussian signals. *IEE Proc.* **1993**, *140*, 362–370.
18. Hyvarinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Networks* **1999**, *10*, 626–634.
19. Bingham, E.; Hyvarinen, A. A fast fixed-point algorithm for independent component analysis of complex valued signals. *Int. J. Neural Syst.* **2000**, *10*, 1–8.
20. Douglas, S.C. Fixed-Point algorithms for the blind separation of arbitrary complex-valued non-gaussian signal mixtures. *EURASIP J. Adv. Signal Proc.* **2007**, *2007*, 83–98.
21. Li, X.L.; Adali, T.L. Independent component analysis by entropy bound minimization. *IEEE Trans. Signal Proc.* **2010**, *58*, 5151–5164.
22. Javidi, S.; Took, C.C.; Mandic, D.P. Fast independent component analysis algorithm for quaternion valued signals. *IEEE Trans. Neural Networks* **2011**, *22*, 1967–1978.
23. Mi, J.X.; Gui, J. A method for ICA with reference signals. *Adv. Intell. Comput. Theor. Appl. Asp. Artif. Intell.* **2010**, *6216*, 156–162.
24. Nocedal, J.; Wright, S.J. *Numerical Optimization*; Springer Verlag: New York, NY, USA, 1999.
25. Zarzoso, V.; Comon, P. Comparative speed analysis of fastica. *Lect. Notes Comput. Sci.* **2007**, *4666*, 293–300.
26. Zarzoso, V.; Comon, P. Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size. *IEEE Trans. Neural Networks* **2010**, *21*, 248–261.
27. Shen, H.; Kleinstueber, M.; Hüper, K. Local convergence analysis of fastica and related algorithms. *IEEE Trans. Neural Networks* **2008**, *19*, 1022–1032.

Appendix I—Derivation of Equation (19)

The first-order derivative of $J_1(\mathbf{w})$ is calculated, as follows:

$$\nabla_{\mathbf{w}} J_1(\mathbf{w}) = \frac{\partial J_1(\mathbf{w})}{\partial \mathbf{w}^*} = \mathbb{E} \left\{ \mathbf{z}(\mathbf{w}^H \mathbf{z})^* g(|\mathbf{w}^H \mathbf{z}|^2) \right\} \quad (27)$$

And the first-order derivative of $J_2(\mathbf{w})$ are calculated, as follows:

$$\nabla_{\mathbf{w}} J_2(\mathbf{w}) = \frac{\partial J_2(\mathbf{w})}{\partial \mathbf{w}^*} = -\mathbf{V} \hat{\mathbf{a}}_i (\mathbf{w}^H \mathbf{V} \hat{\mathbf{a}}_i)^* \quad (28)$$

Also, $\max(J_2(\mathbf{w}), 0)$ can be expressed as follows:

$$\max(J_2(\mathbf{w}), 0) = \begin{cases} J_2(\mathbf{w}) & \text{if } J_2(\mathbf{w}) \geq 0 \\ 0 & \text{if } J_2(\mathbf{w}) < 0 \end{cases} \quad (29)$$

Therefore, we can derive the following derivative

$$\begin{aligned} \nabla_{\mathbf{w}} \max(J_2(\mathbf{w}), 0) &= \frac{\partial}{\partial \mathbf{w}^*} \max(J_2(\mathbf{w}), 0) \\ &= \frac{\text{sign}(J_2(\mathbf{w})) + 1}{2} \frac{\partial J_2(\mathbf{w})}{\partial \mathbf{w}^*} \end{aligned} \quad (30)$$

Therefore, the gradient of $L(\mathbf{w}, \lambda)$ is calculated as follows:

$$\begin{aligned} \nabla_{\mathbf{w}} L(\mathbf{w}, \lambda) &= \nabla_{\mathbf{w}} J_1(\mathbf{w}) - \lambda \nabla_{\mathbf{w}} \max(J_2(\mathbf{w}), 0) \\ &= \frac{\partial J_1(\mathbf{w})}{\partial \mathbf{w}^*} - \frac{\lambda (\text{sign}(J_2(\mathbf{w})) + 1)}{2} \frac{\partial J_2(\mathbf{w})}{\partial \mathbf{w}^*} \end{aligned} \quad (31)$$

Substituting Equations (27) and (28) in Equation (31), we can easily obtain

$$\nabla_{\mathbf{w}} L = \mathbb{E} \left\{ \mathbf{z}(\mathbf{w}^H \mathbf{z})^* g(|\mathbf{w}^H \mathbf{z}|^2) \right\} + \frac{\lambda}{2} (\text{sign}(J_2) + 1) \left[\mathbf{V} \hat{\mathbf{a}}_i (\mathbf{w}^H \mathbf{V} \hat{\mathbf{a}}_i)^* \right] \quad (32)$$

Appendix II—Proof of Theorem 1

Without loss of generality, we assume s_p is the only one SOI, \mathbf{w}_p is the solution of extracting vector, and the initial weight of \mathbf{w}_p is $\mathbf{w}_p^{(0)} = \mathbf{V} \hat{\mathbf{a}}_p / \|\mathbf{V} \hat{\mathbf{a}}_p\|$. So we have

$$s_p = \mathbf{w}_p^H \mathbf{z} = \mathbf{w}_p^H \mathbf{V} \mathbf{A} \mathbf{s} = \mathbf{e}_p^H \mathbf{s} \quad (33)$$

$$\hat{s}_p = \left(\mathbf{w}_p^{(0)} \right)^H \mathbf{z} = \left(\mathbf{w}_p^{(0)} \right)^H \mathbf{V} \mathbf{A} \mathbf{s} = \mathbf{q}_p^H(0) \mathbf{s} \quad (34)$$

where $\mathbf{e}_p = [0, \dots, c, \dots, 0]^H$ with $|c| = 1$ and $\mathbf{q}_p(0) = (\mathbf{V} \mathbf{A})^H (\mathbf{V} \mathbf{A}) \mathbf{A}^{-1} \hat{\mathbf{a}}_p = \mathbf{A}^{-1} \hat{\mathbf{a}}_p = [q_{p,1}, \dots, q_{p,N}]^H$ with the unit-norm constraint $\|\mathbf{q}_p(0)\| = 1$. Now, we just need to proof \mathbf{q}_p must converge to \mathbf{e}_p . Here, we use kurtosis as the measurement of non-Gaussianity instead of negentropy and gradient learning algorithm as the optimization method for simplicity, since Newton-like method has the same extreme value which has been shown in [6].

According to 0, kurtosis has linearity property that

$$\text{kurt}(\mathbf{q}_p^H \mathbf{s}) = \sum_{i=1}^N |q_{p,i}|^4 \text{kurt}(s_i) \quad (35)$$

And the gradient can be calculated as

$$\frac{\partial \text{kurt}(\mathbf{q}_p^H \mathbf{s})}{\partial \mathbf{q}_p^*} = \left[2q_{p,1} |q_{p,1}|^2 \text{kurt}(s_1), 2q_{p,2} |q_{p,2}|^2 \text{kurt}(s_2), \dots, 2q_{p,N} |q_{p,N}|^2 \text{kurt}(s_N) \right]^T \quad (36)$$

Equating the direction of gradient with new \mathbf{q}_p , we have

$$\mathbf{q}_p \propto \frac{\partial \text{kurt}(\mathbf{q}_p^H \mathbf{s})}{\partial \mathbf{q}_p^*} \quad \text{and} \quad \mathbf{q}_p = \mathbf{q}_p / \|\mathbf{q}_p\| \quad (37)$$

Therefore, the learning process of the elements of \mathbf{q}_p can be written as

$$q_{p,i}(k+1) = c(k) \text{kurt}(s_p) q_{p,i}(k) |q_{p,i}(k)|^2 \quad i=1, \dots, N \quad (38)$$

where $c(k)$ is a constant which makes the norm of $\mathbf{q}_p(k+1)$ equal unity. Then we have

$$\left| \frac{q_{p,p}(k+1)}{q_{p,i}(k+1)} \right| = \left| \frac{c(k) c^3(k-1) \dots c^{3^k}(0) \text{kurt}^{\frac{3^k-1}{2}}(s_p) q_{p,p}^{3^k}(0)}{c(k) c^3(k-1) \dots c^{3^k}(0) \text{kurt}^{\frac{3^k-1}{2}}(s_i) q_{p,i}^{3^k}(0)} \right| = \left| \frac{\text{kurt}(s_p)}{\text{kurt}(s_i)} \right|^{\frac{3^k-1}{2}} \left| \frac{\tilde{\mathbf{a}}_p^H \hat{\mathbf{a}}_p}{\tilde{\mathbf{a}}_i^H \hat{\mathbf{a}}_p} \right| \quad p \neq i \quad (39)$$

Case 1: $|\text{kurt}(s_p)| \geq |\text{kurt}(s_i)|$ and $\left| \frac{\tilde{\mathbf{a}}_p^H \hat{\mathbf{a}}_p}{\tilde{\mathbf{a}}_i^H \hat{\mathbf{a}}_p} \right| > 1$

$$\left| \frac{q_{p,p}(k+1)}{q_{p,i}(k+1)} \right| > \left| \frac{\text{kurt}(s_p)}{\text{kurt}(s_i)} \right|^{\frac{3^k-1}{2}} > 1 \quad (40)$$

Case 2: $|\text{kurt}(s_p)| < |\text{kurt}(s_i)|$

$$\begin{aligned} \left| \frac{q_{p,p}(k+1)}{q_{p,i}(k+1)} \right| &= \left| \frac{\text{kurt}(s_p)}{\text{kurt}(s_i)} \right|^{\frac{3^k}{2}} \left| \frac{\text{kurt}(s_i)}{\text{kurt}(s_p)} \right|^{\frac{1}{2}} \left| \frac{\tilde{\mathbf{a}}_p^H \hat{\mathbf{a}}_p}{\tilde{\mathbf{a}}_i^H \hat{\mathbf{a}}_p} \right| \\ &> \left| \frac{\text{kurt}(s_p)}{\text{kurt}(s_i)} \right|^{\frac{3^k}{2}} \left| \frac{\text{kurt}(s_i)}{\text{kurt}(s_p)} \right|^{\frac{3^k}{2}} \left| \frac{\text{kurt}(s_i)}{\text{kurt}(s_p)} \right|^{\frac{1}{2}} \\ &= \left| \frac{\text{kurt}(s_i)}{\text{kurt}(s_p)} \right|^{\frac{1}{2}} > 1 \end{aligned} \quad (41)$$

Therefore, for any k in the learning process, the following inequality always holds

$$\left| \frac{q_{p,p}(k+1)}{q_{p,i}(k+1)} \right| > 1 \quad p \neq i \quad (42)$$

Since [6,27] has proved that the FastICA algorithm will converge and once the algorithm converges, only one element of \mathbf{q}_p equals c with $|c| = 1$, and according to Equation (42), \mathbf{q}_p must converge to \mathbf{e}_p . This completes the proof.

© 2012 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).