中央研究院
資訊科學研究所
Institute of Information Science, Academia Sinica • Taipei, Taiwan, ROC

TR-IIS-13-001

Null Space Component Analysis for Noisy Blind Source Separation

Wen-Liang Hwang and Jinn Ho

Apr. 29, 2014(update) || Technical Report No. TR-IIS-13-001
http://www.iis.sinica.edu.tw/page/library/TechReport/tr2013/tr13.html

Null Space Component Analysis for Noisy Blind Source Separation

Wen-Liang Hwang and Jinn Ho
Institute of Information Science, Academia Sinica, Taiwan

Abstract

We propose an approach called the Null Space Component Analysis (NCA) algorithm to solve the noisy blind source separation (BSS) problem. In a set of m linearly independent source signals, each signal is associated with a separating operator that includes the signal in its null space and repels other signals from the space. The signal model induced by the m operators represents the space where each operator separates a single signal from the other signals. We show that the model can act as a constraint on the source signals in the noisy BSS problem. In contrast to the ICA-based and the sparsity-based approaches, NCA is a deterministic and data-adaptive algorithm that can solve both the under-determined and the over-determined BSS problem. To demonstrate the algorithm's efficiency, we process several signals, including real-life signals obtained from biomedical systems, and compare the results with those derived by other methods.

Keywords. Blind Source Separation, Null Space Operator

1 Introduction

The blind source separation (BSS) problem is a fundamental issue in applications of biomedical engineering, signal processing, and communications. Various topics related to the problem, including its history, are discussed in Common and Jutten's handbook [1]. The problem involves separating the source signals from observation signals under various source signal mixing models. Initially, linear instantaneous (memoryless) mixing models were used [2], followed by linear convolution mixing models [3]. More recently, nonlinear mixing models [4, 5, 6], bounded component analysis [7, 8], and the sparsity-based approach [9, 10] have been exploited.

In this paper, we consider the most basic BSS model, which assumes that the source signals are instantaneously mixed by an $n \times m$ matrix. Our objective is to estimate the m source signals and the mixing matrix from n observation signals for any m and $n \geq 2$. Darmois [11] observed that the problem is ill-posed because it does not provide a solution for Gaussian and temporally independent and identically distributed (i.i.d.) source signals. Usually, the problem can be well-posed by imposing assumptions on the source signals. For example, the second-order methods were developed for temporally correlated sources [12]. Otherwise, one can assume that the source signals are temporally i.i.d., but non-Gaussian. The assumption is the basis for the well known independent component analysis (ICA) algorithm [13, 14, 15]. If the number of observations is greater than or equal to the number of sources, i.e., $n \geq m$ (the over-deterministic case), the ICA approach can separate the statistically independent sources provided that at most one of the sources is Gaussian[1].

There have been a number of attempts to extend the ICA approach in order to address the problem in noisy environments. Several methods based on estimating the probability functions of the source signals or their high order statistics have been proposed to solve the noisy BSS problem. The methods can be classified according to which one of the following three key properties they exploit: 1. the noise variance is known [16, 17]; 2. the bias removal technique, which removes the bias introduced by noise in the mixing matrix, is used for pre-whitening [18, 19, 20, 21]; or 3. the source density is modeled [22]. An overview of some of the methods can be found in [23]. Each of the above methods has a limitation [24].For example, the higher order statistics methods are sensitive to outliers; the bias removal technique must estimate the noise variance; and the source modeling approach requires more data to estimate the parameters.

In contrast to the ICA approach, the sparsity-based approach can solve the under-determined BSS problem effectively [25, 26, 27, 28] when the number of observations n is less than the number of source signals m . The sparsity-based approach, or sparse component analysis (SCA), assumes that the source signals are

[1]For simplicity, we assume that the true order of the source signals is obtained. The ICA essentially computes the pseudo inverse W of the mixing matrix A from the observation X, to obtain the source signals S by

$$S =_p WX =_p WAS =_p (A^T A)^{-1} A^T AS, \tag{1}$$

where $=_p$ is the equality defined in the sense of probability. In order for the existence of $(A^T A)^{-1}$, A must be a full column rank over-deterministic matrix with $n \geq m$.

sparse with respect to some dictionaries. Let X be the observation signal matrix, A be the mixing matrix, and S be the source signal matrix. In addition, let $\mathcal{D}$ be a linear redundant transform (called the dictionary) such that the source signals, represented as row vectors in S, have sparse representations, i.e., $s_i^T = \mathcal{D}c_i^T$, and the coefficient (column) vector c_i^T is k-sparse if it contains k non-zero entries at most. After estimating the mixing matrix, e.g., with the clustering approach proposed in [29], the sparse assumption on the source signals can be represented as an L_1 norm optimization problem by solving the minimum of $\sum_{i=1}^{m} ||c_i^T||_1$ with the constraints $X = AS$ and $s_i = c_i D^T$. The SCA approach assumes the dictionary D that makes the source signals sparse is known as a priori information [30]. In fact, it is known that the Gabor dictionary is the sparse representation of oscillatory signals, and the wavelets comprise the sparse representation of piecewise-smooth signals. However, in general, a dictionary that can sparsely represent the source signals is dependent on the signals and cannot be determined easily without relying on a learning method [31].

Several adaptive techniques have been developed to solve the single channel source separation problem (a.k.a. the multiple inputs single output (MISO) problem), which attempts to separate superimposed source signals with a single observation signal. Like the BSS problem, well-posed solutions can be obtained by imposing assumptions on the source signals; for example, the empirical mode decomposition (EMD) algorithm assumes source signals are intrinsic mode functions [32]. Our approach is motivated by a recently proposed solution to the MISO problem that assumes a source signal is in the null space of an adaptive operator (so that the operator annihilates the signal) [33]. We extend the design concept of a signal-dependent operator from trying to annihilate a signal to separating one signal from other signals. Specifically, the separating operator simultaneously annihilates one signal and removes other signals from its null space. We define a set of signals $\{s_i | i = 1, \cdots, m\}$ as separable if there are $N \times N$ matrix operators $\{\mathcal{T}_i | i = 1, \cdots, m\}$ such that

$$\begin{cases} \mathcal{T}_i s_i^T = 0_N, \\ \mathcal{T}_i s_j^T \neq 0_N, \text{ for } j \neq i, \end{cases} \tag{2}$$

where 0_N is a zero vector; and we show that the signals that satisfy the above definition are linearly independent vectors. Then, we demonstrate that if the rank constraint is applied on the operators with

$$\text{Rank}(\mathcal{T}_i) = N - 1, \tag{3}$$

the rotation ambiguity that occurs in solving the BSS problem can be removed.

For convenience, the constraints on the source signals and operators in Equations (2) and (3) are referred to as the null space component analysis (NCA) constraints in the BSS problem. Based on the NCA constraints, we use a regularization approach to formulate the noisy BSS problem as a constrained optimization problem. The proposed NCA algorithm utilizes the alternating projection method, which iteratively derives the source signals, the rank-constrained matrix operators, and the mixing matrix from the observation signals. The algorithm is a deterministic and data-adaptive approach. We show that it can separate linearly independent source signals with an under-deterministic mixing matrix as well as an over-deterministic mixing matrix to solve the noisy BSS problem. To demonstrate the robustness and accuracy of the algorithm, we used it to separate noisy synthesized signals as well as a real-life signal, which involved removing electrocardiography (EOG) artifacts from electroencephalography (EEG) signals [34, 35].

The remainder of this paper is organized as follows. In Section 2, we introduce the NCA approach, including the concept of separating operators, the construction of the operators, the signal models induced by the operators, and the optimization methods used to derive the operators. In Section 3, we formulate the BSS problem with the proposed approach. We also show that the rotation ambiguity can be removed by applying the rank constraint on the operators. In Section 4, we present the NCA algorithm, which is based on the proposed models, to resolve the noisy BSS problem; and in Section 5, we discuss some implementation issues and the results of experiments on simulated and real-world signals. Section 6 contains our concluding remarks.

2 Null Space Component Analysis (NCA)

Given a set of signals, the NCA approach constructs an operator for each signal so that only the signal is in the operator's null space, and all other signals are excluded. First, we show that the operators exist for linearly independent signals. Then, we demonstrate that the optimal linear operators can be constructed based on the information about the signals.

2.1 Separating Operator

Let $s_1, \cdots, s_m$ be a set of m signals represented by row vectors in R^N, and let s_1 be linearly independent of s_i for all $i = 2, \cdots m$. We define the separating operator

$\mathcal{T}_1 : R^N \rightarrow R^N$ that separates s_1 from $\{s_i | i = 2, \cdots, m\}$ as follows:

$$\begin{cases} \mathcal{T}_1 s_1^T = 0_N, 0_N \text{ in } R^N \text{ whose elements are all equal to 0,} \\ \mathcal{T}_1 s_i^T \neq 0_N \text{ for } i = 2, \ldots, m. \end{cases} \tag{4}$$

The signal s_1 is called the separating signal of the operator $\mathcal{T}_1$; and the signals $\{s_i | i = 2, \cdots, m\}$ are called the supporting signals of $\mathcal{T}_1$ because they help $\mathcal{T}_1$ to remove the signal s_1. As $\mathcal{T}_1 s_1^T = 0_N$ and $s_1^T \neq 0_N$, the separating operator $\mathcal{T}_1$ is always a singular operator.

The operator $\mathcal{T}_1$ is well-defined because a linear operator can be constructed from the signals to satisfy the above definition. Let $s_i = u_i + v_i$ with u_i in the direction of s_1 and v_i in the direction perpendicular to s_1; then, it is straightforward to verify that the linear operator $\mathcal{T}_1 = \sum_{i=2}^{m} v_i^T v_i$ satisfies Equation (4). The separating operator $\mathcal{T}_1$ is the projection onto the subspace spanned by $\{v_2^T, \cdots v_m^T\}$. It is obvious that $\mathcal{T}_1$ is not unique. Let $||\mathcal{T}_1|| = 1$ where the operator is normalized with respect to a certain norm. Then, for numerical applications, we can define

$$\sum_{i=2}^{m} (||\mathcal{T}_1 s_i^T||_2^2 - ||\mathcal{T}_1 s_1^T||_2^2) \tag{5}$$

as the separation between $\{s_i | i = 2, \cdots, m\}$ and s_1 under the operator $\mathcal{T}_1$. Now, the optimal separating operator $\mathcal{T}_1^*$ that yields the maximum separation can be obtained by solving the following problem:

$$\begin{cases} \max_{||\mathcal{T}_1||=1} \sum_{i=2}^{m} (||\mathcal{T}_1 s_i^T||_2^2 - ||\mathcal{T}_1 s_1^T||_2^2), \\ \qquad ||\mathcal{T}_1 s_1^T||_2^2 = 0. \end{cases} \tag{6}$$

In practice, the equality constraint in Equation (6) can be relaxed to $||\mathcal{T}_1 s_1^T||_2^2 \leq \epsilon$; and the optimal separating operator can be derived by solving

$$\begin{cases} \max_{||\mathcal{T}_1||=1} \sum_{i=2}^{m} (||\mathcal{T}_1 s_i^T||_2^2 - ||\mathcal{T}_1 s_1^T||_2^2), \\ \qquad ||\mathcal{T}_1 s_1^T||_2^2 \leq \epsilon. \end{cases} \tag{7}$$

Intuitively, $\mathcal{T}_1^*$, the solution of Equation (7), yields the largest separation by mapping the signal s_1^T to within a small ball centered at 0_N, and simultaneously mapping the signals s_i, with $i \neq 1$, as far away as possible from 0_N. It can be shown that $\mathcal{T}_1^*$ can be derived by solving

$$\min_{||\mathcal{T}_1||=1} (||\mathcal{T}_1 s_1^T||_2^2 - \mu \sum_{i=2}^{m} (||\mathcal{T}_1 s_i^T||_2^2 - ||\mathcal{T}_1 s_1^T||_2^2)), \tag{8}$$

where $\mu \geq 0$ is the Lagrangian multiplier. Using $||\mathcal{T}_1 s_1^T||_2^2 \leq \epsilon$, Equation (8) has the

following upper-bound:

$$\begin{aligned} ||\mathcal{T}_1 s_1^T||_2^2 - \mu \sum_{i=2}^{m} (||\mathcal{T}_1 s_i^T||_2^2 - ||\mathcal{T}_1 s_1^T||_2^2) &\leq \epsilon + (m-1)\mu\epsilon - \mu \sum_{i=2}^{m} ||\mathcal{T}_1 s_i^T||_2^2 \\ &\leq \epsilon + (m-1)\mu\epsilon - \mu(\sum_{i=2}^{m} ||\mathcal{T}_1 s_i^T||_2^2 - ||\mathcal{T}_1 s_1^T||_2^2). \end{aligned}$$

Because $||\mathcal{T}_1 s_1^T||_2^2 - \mu \sum_{i=2}^{m} (||\mathcal{T}_1 s_i^T||_2^2 - ||\mathcal{T}_1 s_1^T||_2^2)$ must be a non-positive value, we can require that ϵ and μ satisfy

$$\epsilon + (m-1)\mu\epsilon - \mu(\sum_{i=2}^{m} ||\mathcal{T}_1 s_i^T||_2^2 - ||\mathcal{T}_1 s_1^T||_2^2) \leq 0. \tag{9}$$

By rearranging the terms in Equation (9) and letting $\mu > 0$, we obtain

$$\frac{\epsilon}{\mu} \leq \frac{\epsilon}{\mu} + (m-1)\epsilon \leq \sum_{i=2}^{m} (||\mathcal{T}_1 s_i^T||_2^2 - ||\mathcal{T}_1 s_1^T||_2^2). \tag{10}$$

To increase the separation metric $\sum_{i=2}^{m} (||\mathcal{T}_1 s_i^T||_2^2 - ||\mathcal{T}_1 s_1^T||_2^2)$, $\frac{\epsilon}{\mu}$ should be as large as possible. If ϵ is a small number, μ must also be a small number in order to yield a large $\frac{\epsilon}{\mu}$ value.

2.1.1 Local Linear Separating Operator

The solution of Equation (8) depends on the representation of an operator. A signal usually contains coherent local components, which suggests that we can derive the local operator for a segment of the signal, and then combine all the local operators to form the operator for the whole signal. The resulting operator is called the local linear separating operator.

Let $\mathcal{T}_1(j)$ be the j-th row of a matrix operator $\mathcal{T}_1$ of size $N \times N$. By the L_2 property, we have

$$||\mathcal{T}_1 s_1^T||_2^2 = \sum_{j=1}^{N} ||\mathcal{T}_1(j) s_1^T||_2^2. \tag{11}$$

Using Equation (11), and normalizing the 2-norm of each row, $||\mathcal{T}_1(j)||_2^2 = 1$, the objective function of Equation (8) can be re-arranged and re-written in terms of $\mathcal{T}_1(j)$ as

$$\min_{\{\mathcal{T}_1 \,|\, ||\mathcal{T}_1(j)||_2 = 1\}} \sum_{j=1}^{N} \mathcal{T}_1(j) \left[(1 - \mu + \mu m)\, s_1^T s_1 - \mu \sum_{i=2}^{m} s_i^T s_i \right] \mathcal{T}_1(j)^T. \tag{12}$$

Let $\mathcal{T}_1(j)$ operate on a local segment of length L_1 of a signal. We can impose the non-zero elements of $\mathcal{T}_1(j)$ on L_1 consecutive elements beginning at j, and use the

periodic boundary condition to extend the end point if necessary; for example,

$$\mathcal{T}_1(j) = [\ \underbrace{0\ldots0}_{j-1}\ \underbrace{*\ldots*}_{L_1}\ \underbrace{0\ldots0}_{N-L_1-j+1}\ ]. \tag{13}$$

If $\mathcal{T}_1(l)$ and $\mathcal{T}_1(k)$ with $l \neq k$ are derived independently, the summation and minimization in Equation (12) can be interchanged, and we have

$$\sum_{j=1}^{N} \min_{||\mathcal{T}_1(j)||_2=1} \mathcal{T}_1(j) \left[(1-\mu+\mu m)\, s_1^T s_1 - \mu \sum_{i=2}^{m} s_i^T s_i \right] \mathcal{T}_1(j)^T. \tag{14}$$

Obviously, the solution $\mathcal{T}_1(j)^*$ of Equation (14) is the eigenvector that corresponds to the smallest eigenvalue of the $L_1 \times L_1$ symmetric submatrix

$$\left[(1-\mu+\mu m)\, s_1^T s_1 - \mu \sum_{i=2}^{m} s_i^T s_i \right]_{j:((j+L_1-1)\ mod\ (N+1))+1, j:((j+L_1-1)\ mod\ (N+1))+1}. \tag{15}$$

The index range of the submatrix is the same as the support interval of $\mathcal{T}_1(j)$, and corresponds to the signal segment that contributes to the calculation of Equation (14) for row j. The optimal local separating operator $\mathcal{T}_1^*$ is the matrix formed by concatenating the N row vectors $\{\mathcal{T}_1^*(j)\}$

$$\mathcal{T}_1^* = [\mathcal{T}_1^{*T}(1)\ \ \cdots\ \ \mathcal{T}_1^{*T}(N)]^T. \tag{16}$$

2.1.2 Group of Separating Operators

For a collection of signals, we have defined the separating operator that separates one of the signals, say s_1, from the others. The sole requirement for the existence of the separating operator is that the signal s_1 is linearly independent of the other signals. We can extend this definition to cover an arbitrary collection of m linearly independent signals. Let $\mathcal{T}_i$ be the separating operator of s_i with the supporting signals $\{s_j | j \neq i\}$. Then, $\{\mathcal{T}_i\}$ can be defined as the group of separating operators for $\{s_i\}$ such that

$$\begin{cases} \mathcal{T}_i s_i^T = 0_N, \\ \mathcal{T}_i s_j^T \neq 0_N, \text{ for } j \neq i, \end{cases} \tag{17}$$

It is straightforward to extend the technique used to derive an optimal local linear separating operator (described in the previous section) to obtain the group of separating operators. Each operator $\mathcal{T}_i$ is associated with an interval L_i and is obtained by solving

$$\min_{\{||\mathcal{T}_i||=1\}} \left[\|\mathcal{T}_i s_i^T\|_2^2 - \mu \sum_{j=1}^{m} \left(\|\mathcal{T}_i s_j^T\|_2^2 - \|\mathcal{T}_i s_i^T\|_2^2 \right) \right], \tag{18}$$

where $||\mathcal{T}_i|| = 1$ means that each row of $\mathcal{T}_i$ is normalized to 1 with the L_2 norm.

2.1.3 Disjoint Subspaces Signal Model

We can induce the signal model for a group of m separating operators $\{\mathcal{T}_i\}$. The operator $\mathcal{T}_i$ partitions the space R^N into two subspaces M_i and N_i, where N_i is the null space and

$$R^N = M_i \oplus N_i. \tag{19}$$

Let Σ_i be the set of signals that satisfies

$$\begin{cases} \mathcal{T}_i \Sigma_i = 0_N, \\ \mathcal{T}_j \Sigma_i \neq 0_N \text{ for } j \neq i; \end{cases} \tag{20}$$

Σ_i can be characterized as follows:

$$\Sigma_i = N_i \cap (\cap_{j=1, j\neq i}^{j=m} M_j). \tag{21}$$

It is clear that Σ_i and Σ_j are disjoint because $\Sigma_i \cap \Sigma_j = \{0_N\}$ for $i \neq j$. If $\mathcal{T}_i$ is a linear operator, then Σ_i (the finite number of intersections of subspaces) is a subspace in R^N. Therefore, the signal model induced by the linear separating operators is the vector subspace $\Sigma_1 \oplus \cdots \oplus \Sigma_m$.

3 NCA for BSS

The BSS problem involves the retrieval of m unobserved sources, denoted as $\{s_i\}$, from n observed mixtures, $\{x_j\}$. Each signal, either s_i or x_j, is a row vector of size N (the number of sampling points). The problem can be formulated as follows:

$$x_i = \sum_{j=1}^{m} a_{ij} s_j, \quad \forall\, i \in \{1, \ldots, n\}. \tag{22}$$

Equation (22) can be re-written in matrix form with

$$X = AS, \tag{23}$$

where A is the mixing matrix, mapping from R^m to R^n; and $X = [x_1^T \cdots x_n^T]^T$ and $S = [s_1^T \cdots s_m^T]^T$ are matrices of size $n \times N$ and $m \times N$ respectively.

3.1 NCA Constraints for BSS

Based on the NCA approach, we introduce the NCA constraints for the BSS problem. The BSS model is affected by scaling, permutation, and rotation ambiguities[2].

[2]For any solution A and S of Equation (23), AC^{-1} and CS is also a solution. By using singular value decomposition (SVD), C can be factorized as a multiplication of the permutation, orthonormal, and scaling matrices.

Because of the scaling ambiguity, we usually assume that the variance of s_i is 1 or the 2-norm of each column of A is 1. The permutation ambiguity prevents us from determining the correct order of the source signals because

$$X = AP^T PS, \tag{24}$$

where P is a permutation matrix. The rotation ambiguity corresponds to

$$X = ABB^T S, \tag{25}$$

where B is an orthonormal matrix. The ICA approach exploits the i.i.d. but non-Gaussian source signal model to resolve the rotation ambiguity, but it still holds the scaling and permutation ambiguities.

The idea of the NCA approach is that the linearly independent sources and the separating operators can be used as constraints for the solution of the BSS problem. The source signals $s_1, \cdots, s_m$, which solve the BSS problem with the ordered operators $[\mathcal{T}_1, \cdots, \mathcal{T}_m]$, are defined as the signals that satisfy

$$\begin{cases} X = AS, \\ \mathcal{T}_i s_i^T = 0, \\ \mathcal{T}_i s_j^T \neq 0 \text{ for } i \neq j. \end{cases} \tag{26}$$

We can impose the additional constraint on the rank of the operators in Equation (26) and obtain

$$\begin{cases} X = AS, \\ \mathcal{T}_i s_i^T = 0, \\ \mathcal{T}_i s_j^T \neq 0 \text{ for } i \neq j \\ \text{Rank}(\mathcal{T}_i) = N - 1. \end{cases} \tag{27}$$

The signal model and operators that satisfy the last three equations in Equation (27) are called the NCA constraints for the BSS problem. The main reason for imposing the rank constraint on operators for solving the BSS problem is to remove the rotation ambiguity, as stated in the following lemma.

Lemma 1. Let $[\mathcal{T}_1, \cdots, \mathcal{T}_m]$ be a tuple of operators. Then, the BSS problem can be uniquely determined up to the scaling of signals by imposing the NCA constraints in Equation (27).

Proof.

Scaling ambiguity: Let D be a diagonal matrix with diagonal elements $\{d_i | d_i \neq 0\}$. The scaling ambiguity of the BSS problem still exists after replacing S with DS in Equation

(27) because

$$\begin{cases} X = AD^{-1}DS, \\ \mathcal{T}_i d_i s_i^T = d_i \mathcal{T}_i s_i^T = 0, \\ \mathcal{T}_i d_j s_j^T = d_j \mathcal{T}_i s_j^T \neq 0 \text{ for } i \neq j, \end{cases} \tag{28}$$

is satisfied.

Permutation ambiguity: The permutation ambiguity can be removed from the BSS problem because of the order of the operators in Equation (27). Specifically, we replace S with PS, where P is a permutation matrix, and let $(PS)_i$ represent the i-th signal (row) of PS. Then, $\mathcal{T}_i(PS)_i^T = 0$ if and only if $P = I$, the identity matrix; because $\mathcal{T}_i s_i^T = 0$ and $\mathcal{T}_i s_j^T \neq 0$ for $j \neq i$.

Rotation ambiguity: The rotation ambiguity can be removed from the BSS problem because of the rank constraint on the operators. We replace S in Equation (27) with $B^T S$, where B is an orthonormal matrix, and use $(B^T S)_i$ to denote the i-th signal of $B^T S$. The constraint on the operator rank in Equation (27) enforces that $\mathcal{T}_i(B^T S)_i^T = 0$ if and only if $B = I$ because $\mathcal{T}_i s_i^T = 0$ and $\text{Rank}(\mathcal{T}_i) = N - 1$. We conclude from the proof that with the NCA constraint and the order operators, the BSS problem can be uniquely determined up to the scaling of the signals. □

The rotation ambiguity is resolved by imposing the NCA constraint on the rank of the operators, not by imposing an additional assumption on the source signals. Note that the lemma does not imply that the true order of the source signals can be determined. It simply indicates that the order of the source signals can be derived after the order of the operators is provided.

4 NCA for Noisy BSS

Noisy BSS assumes that the observations in the BSS model are noisy and that they take the following form:

$$X = AS + V, \tag{29}$$

where $V = [v_1^T, \cdots, v_n^T]^T$ is an $n \times N$ matrix and v_i is sampled from i.i.d. white Gaussian noise with mean 0 and variance σ^2. By taking the log-likelihood of the Gaussian joint distribution of $X - AS$ and omitting the constant term in the log-

likelihood, we obtain

$$\begin{aligned}
\log\ \mathcal{L}(X|AS) &= \sum_{i=1}^{N} \frac{-1}{\sigma^2}(X[i]-AS[i])^T(X[i]-AS[i]) && (30)\\
&= \frac{-1}{\sigma^2}\text{Trace}(\sum_i (X[i]-AS[i])(X[i]-AS[i])^T) && (31)\\
&= \frac{-1}{\sigma^2}||X-AS||_F^2. && (32)
\end{aligned}$$

The maximum of the log-likelihood is the minimum of $||X-AS||_F^2$, the Frobenius norm of $X-AS$. The NCA approach proposes to solve the noisy BSS problem with the following formulation, which is more robust against the noisy environment:

$$\begin{cases}
\min\limits_{S,A,\{\mathcal{T}_i\}} \frac{1}{\lambda}\sum_{i=1}^{m}\left[\|\mathcal{T}_i s_i^T\|_2^2 - \mu\sum_{j=1}^{m}\left(\|\mathcal{T}_i s_j^T\|_2^2 - \|\mathcal{T}_i s_i^T\|_2^2\right)\right] + \|X-AS\|_F^2 \\
\text{Rank}(\mathcal{T}_i) = N-1, \\
||\mathcal{T}_i|| = 1, \ \ \textit{i.e.}\ \text{the 2-norm of each row is normalized to 1,}
\end{cases} \tag{33}$$

where each column of A is normalized to 1 and λ and μ are non-negative Lagrangian multipliers. Equation (33) is the final formulation we use to develop our algorithm.

4.1 The NCA Algorithm

If S in Equation (33) is fixed, the mixing matrix A and the rank-constrained operators can be derived separately; and if A and the operators are fixed, the source signals can be derived by quadratic optimization. This observation motivates us to begin with the initial estimated source signals and use an alternating projection method to derive the operators, the mixing matrix, and the source signals.

The NCA algorithm is comprised of four steps: 1. estimate the initial source signals; 2. estimate the rank-constrained operators with fixed source signals; 3. estimate the mixing matrix with fixed source signals; and 4. estimate the source signals from the fixed operators and the mixing matrix. We consider steps 2, 3, and 4 in the next three subsections; and analyze the computational complexity and convergence of the algorithm in Subsection 4.1.4. Various methods for estimating the initial source signals are discussed in Section 5.1.

4.1.1 Estimating the Rank-constrained Operators

Each operator in Equation (33) can be derived separately by using the method proposed in Section 2.1.1. We use $\{\hat{\mathcal{T}}_i\}$ to denote the derived operators according to Equations (11)-(16). If the operators do not meet the rank requirement, the operator $\hat{\mathcal{T}}_i$ can be modified based on the following optimization problem:

$$\min_{||\mathcal{T}_i||=1} \|\mathcal{T}_i - \widehat{\mathcal{T}}_i\|_F^2 \quad \text{satisfying} \begin{cases} \text{Rank}(\mathcal{T}_i) = N-1, \\ \mathcal{T}_i s_i^T = 0_N. \end{cases} \tag{34}$$

The condition where $\mathcal{T}_i s_i^T = 0$ occurs because $\mathcal{T}_i$ is the operator that separates the signal s_i from the other signals $\{s_j^T | j \neq i\}$. The solution of Equation (34) can be derived by first solving the following equation:

$$\min_{||\mathcal{T}_i||=1} \|\mathcal{T}_i - \hat{\mathcal{T}}_i\|_F^2 \quad \text{satisfying} \begin{cases} \text{Rank}(\mathcal{T}_i) \leq N-1, \\ \mathcal{T}_i s_i^T = 0_N. \end{cases} \tag{35}$$

Then, the rank of the obtained operator is modified to $N-1$.

Let $\hat{\mathcal{T}}_i = U_i \Sigma_i V_i^T$ be the SVD of $\hat{\mathcal{T}}_i$, where $U_i = [u_1, \cdots, u_N]$. First, we construct an orthonormal basis $\{\frac{s_i^T}{||s_i||^2}, w_1, \cdots, w_{N-1}\}$ in R^N by performing the Gram Schmidt procedure on $\{\frac{s_i^T}{||s_i||^2}, u_1, \cdots, u_{N-1}\}$. Then, we project each column of $\hat{\mathcal{T}}_i$ into the space of $\{w_1, \cdots, w_{N-1}\}$. This corresponds to finding the coefficients that minimize

$$\min_{\{b_{k,j}\}_{k=1,\cdots,N-1}} \sum_{j=1}^{N} \| \sum_{k=1}^{N-1} b_{k,j} w_k - \hat{\mathcal{T}}_i(j)^T \|^2, \tag{36}$$

where $\hat{\mathcal{T}}_i(j)^T$ denotes the j-th column of $\hat{\mathcal{T}}_i$. Let $W_i = [w_1, \cdots, w_{N-1}]$ and let $b_j = [b_{1,j}, \cdots, b_{N-1,j}]^T$. The coefficient vector b_j that solves Equation (36) is

$$b_j = \left({W_i}^T W_i\right)^{-1} {W_i}^T \hat{\mathcal{T}}_i(j)^T \tag{37}$$

for $j = 1, \cdots, N$.

Let $B_i = [b_1, \cdots, b_N]$ be an $N-1 \times N$ matrix. The operator that solves Equation (35) can be expressed as $(W_i B_i)^T$, where W_i is an $N \times N-1$ matrix. Note that $(W_i B_i)^T s_i^T = 0_N$ because ${W_i}^T s_i^T = 0_{N-1}$; and $\text{Rank}(B_i) \leq N-1$ because the columns of B_i are not necessarily independent of each other. To make the rank of $(W_i B_i)^T$ equal to $N-1$, we apply SVD to B_i and replace the zero singular values with small positive numbers. The rank of the resultant matrix $\tilde{B}_i$ is $N-1$; and that of the new operator $(W_i \tilde{B}_i)^T$ is also $N-1$ because of the following rank inequality:

$$(\text{Rank}(Y) + \text{Rank}(Z) - k) \leq \text{Rank}(YZ) \leq \min(\text{Rank}(Y), \text{Rank}(Z)), \tag{38}$$

where Y and Z are matrices of size $N \times k$ and $k \times N$ respectively.

Finally, we normalize the 2-norm of each row in the operator $(W_i \tilde{B}_i)^T$ by multiplying an $N \times N$ diagonal matrix D_i to the right of $(W_i \tilde{B}_i)^T$. The final operator is

$$\mathcal{T}_i = \tilde{B}_i^T W_i^T D_i, \tag{39}$$

which satisfies Equation (34).

4.1.2 Estimating the Mixing Matrix

If the source signals S are given, the mixing matrix A, with normalized column norm, for the following noisy BSS model

$$X = AS + V, \tag{40}$$

where V is an $n \times N$ matrix of white Gaussian noise, can be derived by solving the following optimization problem:

$$\min_A ||X - AS||_F^2, \tag{41}$$

whose solution is

$$\tilde{A} = XS^T(SS^T)^{-1}. \tag{42}$$

Because it is assumed that the source signals are linearly independent, SS^T is a full rank square matrix; therefore, $(SS^T)^{-1}$ exists. Then, we normalize the column norm of A by multiplying it by a diagonal matrix D_A to obtain

$$A = D_A\hat{A}. \tag{43}$$

4.1.3 Estimating the Source Signals

Next, we consider the estimation of the source signals with fixed A and $\{\mathcal{T}_i\}$, and the Lagrangian multipliers μ and λ. In the Appendix, we show that the necessary condition for the source signals to optimize the equation

$$\min_S \frac{1}{\lambda}\sum_{i=1}^{m}\left[\|\mathcal{T}_i s_i^T\|_2^2 - \mu\sum_{j=1}^{m}\left(\|\mathcal{T}_i s_j^T\|_2^2 - \|\mathcal{T}_i s_i^T\|_2^2\right)\right] + \|X - AS\|_F^2, \tag{44}$$

is the solution of the following Sylvester equation:

$$\begin{aligned} 0_m = {} & [(m-1)\mu + 1]\sum_{i=1}^{m} e_i e_i^T S(\mathcal{T}_i^T\mathcal{T}_i) + (-\mu)\left[\sum_{j=1}^{m} e_j e_j^T S\left(\sum_{k\neq j}^{m}\mathcal{T}_k^T\mathcal{T}_k\right)\right] \\ & + (-\lambda)\,A^TX + \lambda A^TAS, \end{aligned} \tag{45}$$

where $\{e_i\}$ is the standard (column) basis in R^m.

To calculate the source signals S, we use the Kronecker product, denoted as $\otimes$. Let the vec linear operator transform a matrix U into a column vector by stacking the columns of U on top of one another. A basic connection between the vec and the Kronecker product indicates that if U, V, and W are three matrices such that the matrix product UVW is defined, then

$$\text{vec}(UVW) = (W^T \otimes U)\,\text{vec}(V). \tag{46}$$

Applying the vec operator to Equation (45) and using Equation (46), we can derive that

$$\begin{aligned}\operatorname{vec} S = \Bigg\{ \frac{1}{\lambda} \sum_{i=1}^{m} \left[(m\mu + 1)\, \mathcal{T}_i^T \mathcal{T}_i + (-\mu) \sum_{k=1}^{m} \mathcal{T}_k^T \mathcal{T}_k \right] \otimes e_i e_i^T \\ + \left(\mathbf{I}_{N\times N} \otimes A^T A \right) \Bigg\}^{+} \operatorname{vec}\left(A^T X \right),\end{aligned} \tag{47}$$

where $+$ denotes the pseudo-inverse operation. The first term on the right-hand side of the open braces in Equation (47) is the bias term induced by the operators.

4.1.4 Convergence and Computational Complexity

We summarize Sections 4.1.1, 4.1.2, and 4.1.3 in the following algorithm.

Algorithm NCA:

INPUT: observation signals X (matrix of size $n \times N$), initial signals S^0 (matrix of size $m \times N$), stopping thresholds ϵ_s, Lagrangian $\mu > 0$, Lagrangian $\lambda > 0$, and the parameters $\{L_i | i = 1, \cdots, m\}$ for m local linear operators.

OUTPUT: m source signals S and the $n \times m$ mixing matrix A with the normalized column norm.

Step 1. Let $k = 0$.

Step 2. Estimate m separating operators $\mathcal{T}_i^k$ from S^k (Section 4.1.1).

Step 3. Estimate the mixing matrix A^k from S^k (Section 4.1.2).

Step 4. Estimate the source signals S^{k+1} with fixed $\mathcal{T}_i^k$ and A^k (Section 4.1.3).

Step 5. If $\text{RMSE}(S^k, S^{k+1}) \geq \epsilon_s$, where RMSE calculates the root mean square error of S^{k+1} against S^k, then set $k = k + 1$ and go to **Step 2**.

Step 6. Output: the source signals $S = S^{k+1}$ and the mixing matrix $A = A^k$.

The algorithm derives the operators, the mixing matrix, and the sources alternately after the initial source signals and the parameters are given. We discuss the estimation of the initial source signals and the selection of the parameter values in Sections 5.1 and 5.2 respectively. Next, we analyze the convergence and computational complexity of the NCA algorithm.

A. Convergence: The NCA method applies a sequence of orthogonal projections onto three sets of mixing matrices, source signals, and rank-constrained operators. Because the set of rank-constrained operators is not convex, the convergence of the alternating projection method to the intersection of sets is not guaranteed [36]. Depending on the initial point, there can be a situation where the sets do intersect, but the sequence of projections converges to a limited cycle, i.e., a periodic iteration

between points in the sets. However, in this case, the distance between the signals derived by two consecutive iterations can still converge and this is used by the NCA algorithm as the heuristic stopping condition.

B. Computational Complexity: The complexity of deriving the operators, the mixing matrix, and the source signals is analyzed as follows.

Operators: The operator $\hat{\mathcal{T}}_i$ in Equation (35) is obtained by concatenating N (column) eigenvectors, each of which corresponds to the smallest eigenvalue of an $L_i \times L_i$ matrix (See Equation (15)) and can be derived with time complexity $\mathcal{O}(L_i^3)$. Hence, the time complexity to obtain $\hat{\mathcal{T}}_i$ is $\mathcal{O}(NL_i^3)$. The SVD of $\hat{\mathcal{T}}_i$ takes $\mathcal{O}(N^3)$ and the matrix B_i (Equation (39)) can also be obtained in $\mathcal{O}(N^3)$ time. As a result, $\mathcal{O}(NL_i^3 + N^3)$ is used to derive the operator $\mathcal{T}_i$. Let $L = \max\{L_i | i = 1, \cdots, m\}$. Overall, the algorithm takes $\mathcal{O}(mNL^3 + mN^3)$ time to derive all the m operators.

Mixing Matrix: The mixing matrix is derived by Equation (42). Each element in the $m \times m$ matrix SS^T and $n \times m$ matrix XS^T is obtained by the inner product operation on length N vectors. Therefore, computing SS^T and its inverse takes $\mathcal{O}(m^3 + m^2N)$; computing XS^T takes $\mathcal{O}(mnN)$; and computing the matrix-matrix product $(XS^T)(SS^T)^{-1}$ takes $\mathcal{O}(m^2n)$. Overall, the algorithm takes $\mathcal{O}(m^3 + m^2N + mnN + m^2n)$ time to derive the mixing matrix A.

Source Signals: The source signals are derived by Equation (47) in which the computational cost is dominated by the pseudo-inverse operation, which can be obtained by computing the SVD of an $Nm \times Nm$ matrix. Therefore, the algorithm takes $\mathcal{O}(N^3m^3)$ time to estimate the source signals.

In practice, we can assume the size of the signal $N >> \max\{m, n, L\}$. Hence, the computational complexity of an iteration of the NCA algorithm is $\mathcal{O}(N^3)$.

5 Implementation Issues and Experiment Results

Next, we consider some implementation issues. We discuss the estimation of the initial source signals, the determination of the parameter values, and the performance measurements in Sections 5.1, 5.2, and 5.3 respectively. Then, in Section 5.4, we present the results of experiments performed on synthesized and real-life biomedical signals.

5.1 Estimation of Initial Source Signals

Factoring the matrix X into two matrices S and A is not a convex problem and the set of rank-constrained matrices is not convex; therefore, the final result of the

NCA algorithm depends on the initial source signals. Although good initial source signals can be derived by other BSS methods, in the following, we propose three initial source signal estimation methods, denote as **A**, **B**, and **C**, which we use in the experiments to derive the initial source signals.

Methods **A** and **B** use the prior information about the signals to determine if the sources are narrow-band or wide-band signals. Method **C**, which is particularly useful for processing biomedical signals, uses the reference signals and signals orthogonal to the references as initial source signals. Methods **A** and **B** are based on the NSP algorithm [33, 37], which is an MISO approach that decomposes an observation signal into the superimposition of narrow band signals. Let $X = [x_1^T, \cdots, x_n^T]^T$ represent the matrix of observation signals, where the row vector x_i is the i-th observation. The NSP algorithm decomposes x_i into the sum of r subcomponents u_i^j and a residual v_i as follows:

$$x_i = \sum_{j=1}^{r} u_i^j + v_i. \tag{48}$$

A. Narrow-band Source Signals: The objective is to divide all subcomponents $\{u_i^j\}$ into m groups, and choose a representative from each group as an initial source signal. Because the initial source signals should be linearly independent, signals in the same group should be as linearly dependent on each other as possible. This helps us avoid selecting subcomponents that are almost linearly dependent as initial signals. To realize the objective, we use the hierarchical clustering method [38], which clusters $\{u_i^j\}$ based on the similarity metric defined as follows:

$$\frac{|(u_i^j)(u_k^l)^T|}{||u_i^j||||u_k^l||}. \tag{49}$$

B. Wide-band Source Signals: The NSP subcomponents $\{u_i^j\}$ are narrow-band. To generate initial wide-band source signals, we subtract from an observation signal the subcomponent belonging to the other observation signals. For example, given two wide-band observations x_1 and x_2 , let V_1 and V_2 be the subspaces spanned by the narrow-band subcomponents $\{u_1^j\}$ and $\{u_2^j\}$ respectively. Then the initial source signals s_1^0 and s_2^0 are obtained by

$$s_1^0 = x_1 - P_{V_2}(x_1), \tag{50}$$

$$s_2^0 = x_2 - P_{V_1}(x_2) \tag{51}$$

respectively, where $P_{V_i}(x_j)$, with $i \neq j$, is the orthogonal projection of x_j onto the subspace V_i. Because $P_{V_i}(x_j)$ is a narrow-band signal in x_j, $x_j - P_{V_i}(x_j)$ is a wide-band signal.

C. Reference Signals are Provided: In certain applications, especially biomedical systems, some reference source signals are provided. The references can be used as the initial source signals. The other initial signals can be derived based on the orthogonal complement of the space of the reference signals. Let $r_1, \cdots r_k$ be the reference signals, and let the space of those signals be $\sum_{l=1}^{k} a_l r_l$. The observation signal x can be decomposed into two orthogonal components, one in the subspace of the references and the other in its complement subspace. The second component can be used as an initial source signal. The projection of x in the subspace of references is $\sum_{l=1}^{k} a_l^* r_l$, where the parameters $\{a_l^*\}$ minimize the objective

$$||x - \sum_{l=1}^{k} a_l r_l||^2. \tag{52}$$

By the orthogonality principle, the parameters a_l^* are the solution of

$$(x - \sum_{l} a_l r_l) r_i^T = 0, \text{for } i = 1, \cdots k. \tag{53}$$

The above equation can be re-written as the linear system $Ra^* = b$, with $R[i,j] = r_i r_j^T$, $b[i] = x r_i^T$, and $a[i]^* = a_i^*$. The component $x - \sum_{l=1}^{k} a_l^* r_l$ is in the complement subspace of the reference signals and can be used as an initial source signal.

5.2 Determining the Parameter Values

Finding good parameter values is more of an art than anything else. Here, we provide some guidelines for determining the parameter values of the NCA algorithm. There are $m + 2$ parameters in the algorithm: μ, λ, and L_i (the window size of each operator).

The parameter μ is related to an operator's signal separation power. According to the analysis in Equation (10), μ must be a small number in order to generate a large separation distance for signals inside and outside an operator's null space. In the NCA algorithm, this parameter is a constant number for all application examples.

The value of λ, which balances the BSS model and the NCA model, cannot be too small; otherwise, the minimization of the objective function in Equation (33) would be dominated by the minimization of the NCA model rather than by the BSS model. In the NCA algorithm, the value of λ is selected from the interval $[\lambda_1, \lambda_2]$ as a geometric series with a constant ratio, $\delta\lambda$.

The parameter L_i is the scale of the operator i and is also associated with the dimension of the null space of the operator. Let Y denote an $L_i \times L_i$ symmetric matrix of Equation (15). In addition, let us arrange the L_i real eigenvalues of the

symmetric matrix according to their absolute values in non-decreasing order. The spectral decomposition of Y is

$$Y = \sum_{i=1}^{L_i} a_i u_i^T u_i, \tag{54}$$

where the scalars $\{a_i\}$ and the column vectors $\{u_i^T\}$ are the eigenvalues and the eigenvectors of Y respectively, with $|a_i| \geq |a_{i+1}|$ for $1 \leq i \leq L_i - 1$. Because the eigenvector $u_{L_i}^T$ associated with the smallest eigenvalue is used as the operator, the null space of the local operator $u_{L_i}^T$ is spanned by the first $L_i - 1$ eigenvectors. Hence, the higher the value of L_i, the larger will be the dimension of the operator's null space. This analysis can be used as a guideline to determine the value of L_i for separating local narrow band signals. Because most of the local energy of a local narrow band signal is concentrated along the eigenvector associated with the largest eigenvalue, the signal can be annihilated by an operator derived with L_i set at 2 or 3.

Careful examination of the structure of a submatrix in Equation (15) reveals that an operator depends on the separating signal (the signal to be annihilated by a local operator) as well as on all the supporting signals (the signals to be expelled from the null space of the local operator). The nontrivial dependency between the signals inside and outside an operator's null space complicates the analysis when selecting of the scale of the operator. Therefore, we use a two-level coarse-to-fine approach to determine L_i values: initially, all L_i are assigned the same value; then, the value of each L_i is refined separately.

In the first level, we select an interval $[l_1, l_2]$; then, we sample the values in the interval with a geometric sequence. Let c be a value sampled in this way. We set $L_i = c$ for all i and take one value of λ (as previously described) as the input parameter of the NCA algorithm. After trying all combinations of L_i and λ values for the algorithm, we obtain a set of mixing matrix and source signal pairs. From the set, we choose the pair (A_1, S_1) that yields the best match to the BSS model against the observation X, i.e., $||X - A_1 S_1||_F^2$ is the minimum. Let $L_i = \hat{c}$ and $\lambda = \hat{\lambda}$ be the parameter values used to obtain the pair (A_1, S_1). In the second level, we refine the L_i value of each operator separately by searching for the solution that minimizes $||X - AS||_F^2$ in the neighborhood of $\hat{c}$.

5.3 Performance Measurement

The signal-to-interference ratio (SIR) metric is used to evaluate the performance of a BSS method [4]. The input SIR, SIR^{in}, measures the performance of mixed signals before they are processed by the BSS method. Let x_i be the observation of channel i with

$$x_i = \sum_{j=1}^{m} a_{i,j} s_j + n_i, \tag{55}$$

where $a_{i,j}$ denotes the mixing coefficient, s_j is the source j, n_i denotes the noise, and m is the number of sources. SIR^{in} calculates the signal-to-noise ratio (SNR) of the k-th scaled source signal $a_{i,k}s_k$ in observation channel i as follows:

$$\text{SIR}_k^{\text{in}}(i) = 10 \log_{10} \frac{a_{i,k}^2 ||s_k||^2}{||x_i - a_{i,k}s_k||^2}. \tag{56}$$

When all n observation channels are considered, the maximum SIR_k^{in}, is defined as

$$\text{SIR}_k^{\text{in}} = \max_{i=1,\cdots,n} \text{SIR}_k^{\text{in}}(i). \tag{57}$$

We use the same approach to define the output SIR, SIR^{out}. Let $\hat{A}_o$ and $\hat{S}_o$ denote, respectively, the mixing matrix and the source signals derived by the BSS method; and let A and S be the true mixing matrix and the source signals respectively. The order of the estimated source signals in $\hat{S}_o$ can be permuted to match that in S by solving the following optimization problem:

$$P_o = \arg\min_P ||S - P^T \hat{S}_o||_F^2, \tag{58}$$

where P is a permutation matrix. Then, we define $\hat{A} = \hat{A}_o P_o$ and $\hat{S} = P_o^T \hat{S}_o$ as the estimated mixing matrix and the source signals respectively. SIR^{out} calculates the SNR of the k-th scaled signal $a_{i,k}s_k$ from the BSS solution $(\hat{A}, \hat{S})$ as follows:

$$\text{SIR}_k^{\text{out}}(i) = 10 \log_{10} \frac{a_{i,k}^2 ||s_k||^2}{||a_{i,k}s_k - \hat{a}_{i,k}\hat{s}_k||^2}, \tag{59}$$

where $\hat{a}_{i,k}$ and $\hat{s}_k$ are the estimated $a_{i,j}$ and s_k in $\hat{A}$ and $\hat{S}$ respectively. We also define the maximum output SIR for the source signal k as

$$\text{SIR}_k^{\text{out}} = \max_{i=1,\cdots,n} \text{SIR}_k^{\text{out}}(i). \tag{60}$$

As defined in [4], the SIR gain for the source signal k is

$$\text{SIRI}_k = \text{SIR}_k^{\text{out}} - \text{SIR}_k^{\text{in}}, \tag{61}$$

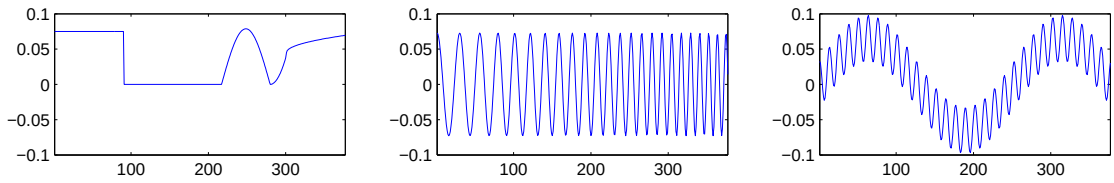


Figure 1: Left: the piecewise smooth signal; middle: the chirp signal; right: the superposition of two sinusoidal waves. The normalized inner-products of the source signals are: 0.0131 (the chirp and the superposition of sinusoidal waves), 0.5184 (the chirp and the piecewise smooth signal), and 0.1070 (the superposition of sinusoidal waves and the piecewise smooth signal).

and the average SIR gain for all source signals is

$$\text{SIRI} = \frac{1}{m}\sum_{k=1}^{m}\text{SIRI}_k. \tag{62}$$

The SIR and SSIR measurements are shown in dB.

Table 1 (SIRI for 3×3 BSS Systems)

$p = 0.2863$	$std(noise) = 0$	$std(noise) = 0.01$	$std(noise) = 0.02$
$\text{SIRI}(A, A_{\text{NCA}})$	**11.3754**	**3.9935**	**0.8269**
$\text{SIRI}(A, A_{\text{FastICA}})$	7.4330	0.3489	0.1603
$\text{SIRI}(A, A_{\text{WPICA}})$	7.6447	0.7890	1.2383
$\text{SIRI}(A, A_{\text{DSS}})$	7.2440	1.2050	-0.2048
$p = 0.4365$	$std(noise) = 0$	$std(noise) = 0.01$	$std(noise) = 0.02$
$\text{SIRI}(A, A_{\text{NCA}})$	**14.3899**	**7.2692**	**4.4084**
$\text{SIRI}(A, A_{\text{FastICA}})$	9.3325	3.0004	1.9854
$\text{SIRI}(A, A_{\text{WPICA}})$	9.4009	-1.1946	2.4601
$\text{SIRI}(A, A_{\text{DSS}})$	12.0154	4.0593	1.4415
$p = 0.5588$	$std(noise) = 0$	$std(noise) = 0.01$	$std(noise) = 0.02$
$\text{SIRI}(A, A_{\text{NCA}})$	9.9917	**7.3172**	**4.1820**
$\text{SIRI}(A, A_{\text{FastICA}})$	10.9433	4.0130	2.2885
$\text{SIRI}(A, A_{\text{WPICA}})$	14.1846	2.9122	2.9302
$\text{SIRI}(A, A_{\text{DSS}})$	**19.3100**	3.8601	3.1988

Table 2 (SIRI for 2×3 BSS Systems)

$p = 0.2863$	$std(noise) = 0$	$std(noise) = 0.01$	$std(noise) = 0.02$
$\mathrm{SIRI}(A, A_{\mathrm{NCA}})$	3.3239	3.0391	3.3626
$\mathrm{SIRI}(A, A_{\mathrm{WPICA}})$	1.8656	0.7667	0.7817
$\mathrm{SIRI}(A, A_{\mathrm{DSS}})$	**15.2357**	**6.0105**	**3.5799**
$p = 0.4365$	$std(noise) = 0$	$std(noise) = 0.01$	$std(noise) = 0.02$
$\mathrm{SIRI}(A, A_{\mathrm{NCA}})$	6.0557	4.3030	**5.4705**
$\mathrm{SIRI}(A, A_{\mathrm{WPICA}})$	2.4199	2.6492	1.7195
$\mathrm{SIRI}(A, A_{\mathrm{DSS}})$	**13.7027**	**5.0468**	5.1443
$p = 0.5588$	$std(noise) = 0$	$std(noise) = 0.01$	$std(noise) = 0.02$
$\mathrm{SIRI}(A, A_{\mathrm{NCA}})$	7.9840	**6.7063**	**4.3999**
$\mathrm{SIRI}(A, A_{\mathrm{WPICA}})$	4.7564	3.4830	3.4788
$\mathrm{SIRI}(A, A_{\mathrm{DSS}})$	**13.4106**	4.8929	4.2294

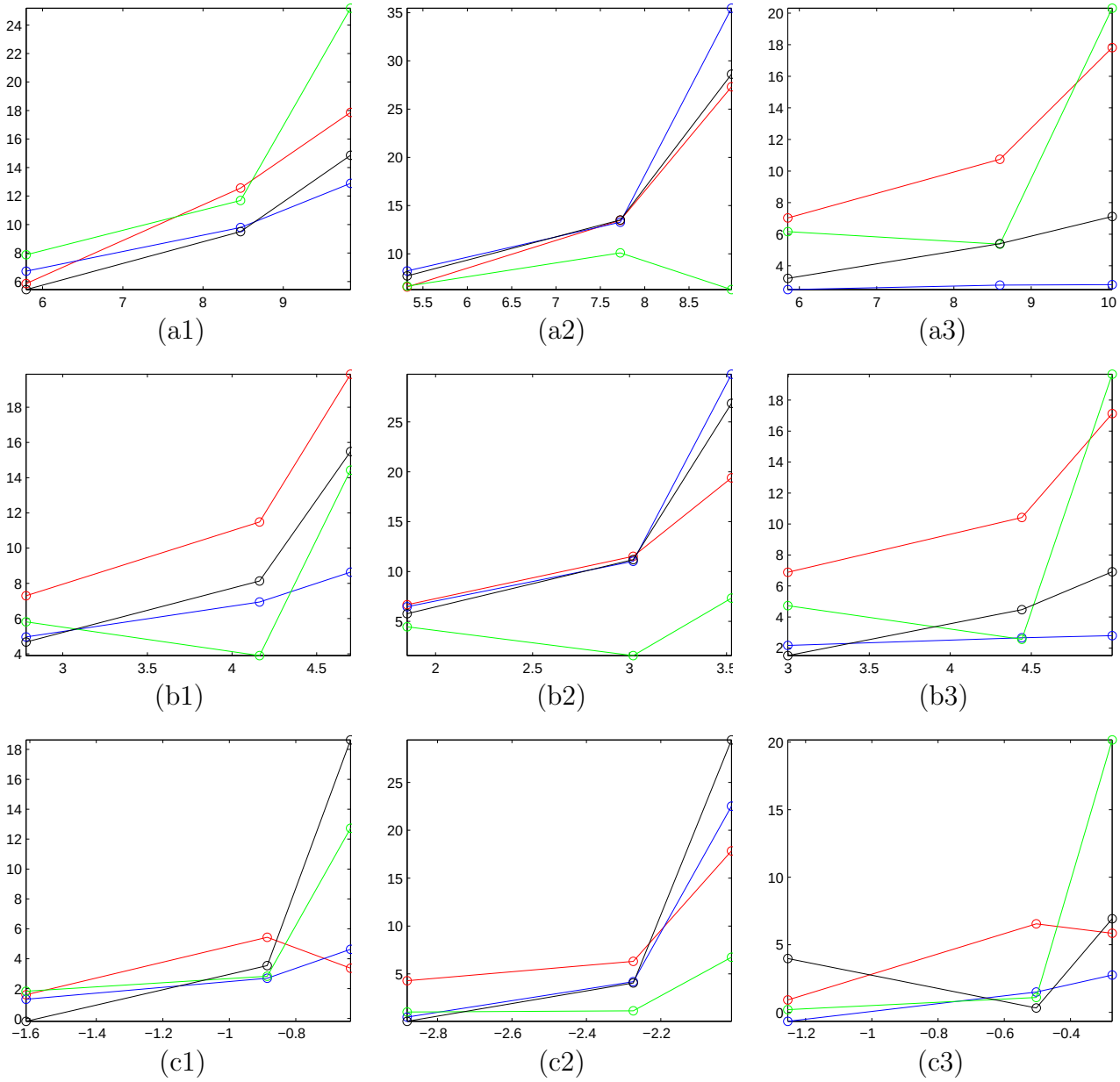


Figure 2: The $\mathrm{SIR}^{\mathrm{in}}$ (horizontal axis) vs $\mathrm{SIR}^{\mathrm{out}}$ (vertical axis) graph for the 3×3 BSS systems in data set 1. The performance curves of NCA, FastICA, WPICA, and DSS are shown in red, blue, green, and black respectively. Top row: subgraphs corresponding to $p = 0.2863$; middle row: subgraphs corresponding to $p = 0.4365$; bottom row: subgraphs corresponding to $p = 0.5588$. Left-hand column: subgraphs of the piecewise smooth signal; middle column: the chirp signal; right-hand column: the superposition of two sinusoids. The experiment setting for a subgraph can be derived from the location of the row and column of the subgraph. For example, (b2) is for the chirp and $p = 0.4365$. There are three points on each curve of a subgraph, namely: the leftmost point, the middle point, and the rightmost point, which are measured with the noise standard deviation set at 0.02, 0.01, and 0.0 (noiseless) respectively.

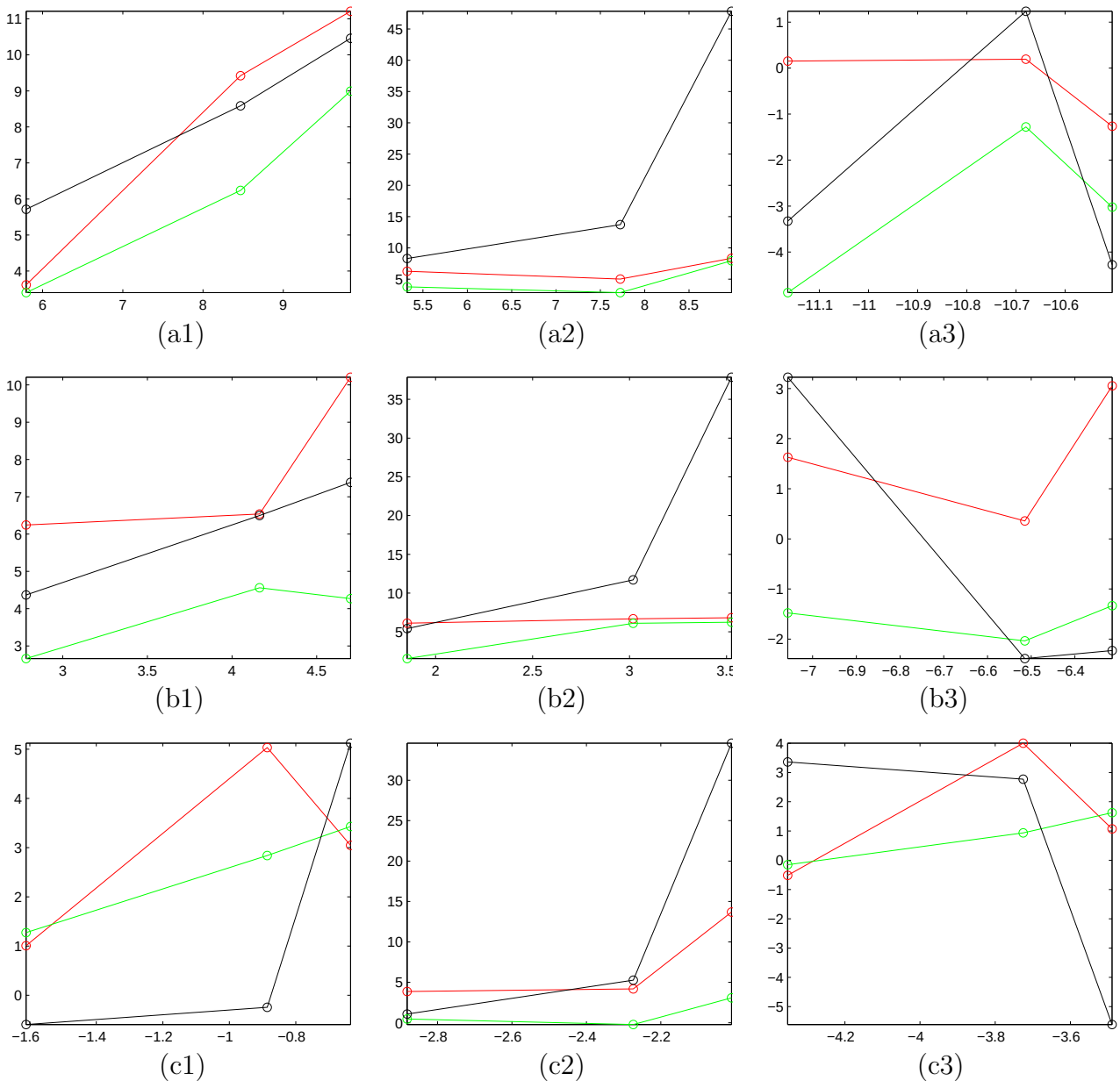


Figure 3: The $\mathrm{SIR}^{\mathrm{in}}$ (horizontal axis) vs $\mathrm{SIR}^{\mathrm{out}}$ (vertical axis) graph for the 2×3 BSS systems in data set 1. The performance curves of NCA, WPICA, and DSS are shown in red, green, and black respectively. Top row: subgraphs corresponding to $p = 0.2863$; middle row: subgraphs corresponding to $p = 0.4365$; bottom row: subgraphs corresponding to $p = 0.5588$. Left-hand column: subgraphs of the piecewise smooth signal; middle column: subgraphs of the chirp signal; right-hand column: subgraphs of the superposition of two sinusoids. The experiment setting for a subgraph can be derived from the location of the row and the column of the subgraph. For example, (a3) is for the piecewise signal and $p = 0.5588$. There are three points on each curve of a subgraph, namely: the leftmost point, the middle point, and the rightmost point which are measured with the noise standard deviation set at 0.02, 0.01, and 0.0 (noiseless) respectively.

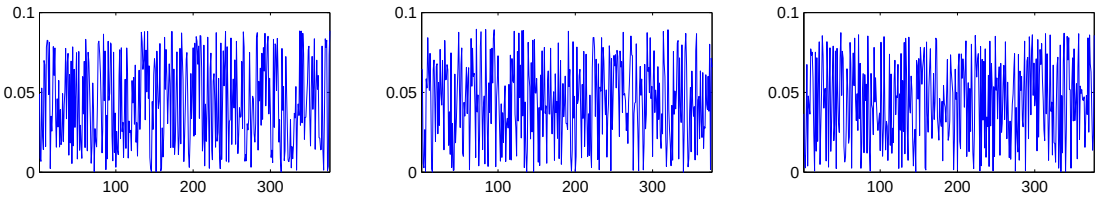


Figure 4: The wideband signals are generated by a t-copula distribution with 3 degrees of freedom.

Table 3 (SIRI for 3×3 BSS Systems)

$p = 0.2863$	$std(noise) = 0$	$std(noise) = 0.01$	$std(noise) = 0.02$
$\mathrm{SIRI}(A, A_{\mathrm{NCA}})$	**13.8268**	**5.4092**	**2.5503**
$\mathrm{SIRI}(A, A_{\mathrm{FastICA}})$	5.2779	0.8495	-0.4347
$\mathrm{SIRI}(A, A_{\mathrm{WPICA}})$	-2.1236	1.0433	0.5606
$\mathrm{SIRI}(A, A_{\mathrm{DSS}})$	2.2782	-1.5640	-2.4184
$p = 0.4365$	$std(noise) = 0$	$std(noise) = 0.01$	$std(noise) = 0.02$
$\mathrm{SIRI}(A, A_{\mathrm{NCA}})$	**18.0551**	**8.6005**	**6.3164**
$\mathrm{SIRI}(A, A_{\mathrm{FastICA}})$	8.9703	3.5462	2.1530
$\mathrm{SIRI}(A, A_{\mathrm{WPICA}})$	3.7952	5.2919	3.6666
$\mathrm{SIRI}(A, A_{\mathrm{DSS}})$	7.9019	2.0146	0.4421
$p = 0.5588$	$std(noise) = 0$	$std(noise) = 0.01$	$std(noise) = 0.02$
$\mathrm{SIRI}(A, A_{\mathrm{NCA}})$	**15.1815**	**10.3131**	**9.9406**
$\mathrm{SIRI}(A, A_{\mathrm{FastICA}})$	12.0891	5.3413	5.0667
$\mathrm{SIRI}(A, A_{\mathrm{WPICA}})$	9.8762	8.7152	7.9393
$\mathrm{SIRI}(A, A_{\mathrm{DSS}})$	15.1092	4.9441	4.5777

Table 4 (SIRI for 2×3 BSS Systems)

$p = 0.2863$	$std(noise) = 0$	$std(noise) = 0.01$	$std(noise) = 0.02$
$\mathrm{SIRI}(A, A_{\mathrm{NCA}})$	**8.5296**	**5.9812**	**5.6888**
$\mathrm{SIRI}(A, A_{\mathrm{WPICA}})$	2.0010	1.9346	3.0715
$\mathrm{SIRI}(A, A_{\mathrm{DSS}})$	7.9958	3.5221	2.6384
$p = 0.4365$	$std(noise) = 0$	$std(noise) = 0.01$	$std(noise) = 0.02$
$\mathrm{SIRI}(A, A_{\mathrm{NCA}})$	**8.7460**	**8.0618**	**6.6911**
$\mathrm{SIRI}(A, A_{\mathrm{WPICA}})$	5.6664	5.8222	5.1919
$\mathrm{SIRI}(A, A_{\mathrm{DSS}})$	8.4533	4.3198	3.4721
$p = 0.5588$	$std(noise) = 0$	$std(noise) = 0.01$	$std(noise) = 0.02$
$\mathrm{SIRI}(A, A_{\mathrm{NCA}})$	8.9486	**9.7891**	**8.9387**
$\mathrm{SIRI}(A, A_{\mathrm{WPICA}})$	**10.9817**	8.1801	8.5304
$\mathrm{SIRI}(A, A_{\mathrm{DSS}})$	10.4865	4.6377	4.5834

5.4 Experiment Results

Numerical experiments were performed to evaluate the source signal separation and mixing matrix estimation capabilities of the proposed algorithm. We considered three cases. In the first two cases, we experimented on synthesized data in noiseless and noisy environments, and measured the performance with the SIR and SIRI metrics. In the noiseless case, we compared our algorithm's performance with that of FastICA [41], Denoising Source Separation (DSS) [42, 43], a robust blind source separation method based on the E-M algorithm; and the Wavelet Packet Preprocessor for Independent Component Analysis (WPICA) [29, 44], which is based on adaptive selection of a sparse representation in a wavelet package tree. As the results of FastICA, DSS, and WPICA depend on randomly generated parameters, we took the average results after executing each method 20 times. The results of all executions of each method were aligned to ensure that the permutation order was consistent in all the executions. In the final case, we compared our approach's performance with that of FastICA on a set of real-life biomedical EEG signals.

Each source signal in the following two synthesized data sets is normalized to 1 with the L_2-norm. In a noisy environment, random elements in the matrix V of Equation (29) are sampled from i.i.d. Gaussian white noise. We set the standard deviation in the experiments at 0.01 or 0.02, corresponding to 1% or 2% of the source signal's energy respectively.

A. Synthesized Data Set 1

We used the proposed approach to separate the blind mixture of a chirp signal ($\cos(4t + 0.2\,t^2)$), a signal comprised of sinusoidal waves ($\sin(10(0.05\pi + t) + 2\sin(0.5t)$), and a piecewise smooth signal formed by concatenating the pieces from left to right: 2, 0, $2.1\sin(t)$, t^2, and $(0.05 + \frac{\pi}{3})^2 + \frac{1}{3}\sqrt{t}$. The source signals are shown in Figure 1. Note that wavelets are sparse representations of the source signals [45]. The initial source signals for the NCA method were derived by the procedure described in Section 5.1.**A**.

The source signals are mixed by unknown mixing matrices to obtain the respective 3×3 (over)-deterministic and 2×3 under-deterministic BSS system as follows:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} p & r & q \\ q & p & r \\ r & q & p \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} \quad \text{and} \tag{63}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \frac{p}{\sqrt{1-r^2}} & \frac{r}{\sqrt{1-q^2}} & \frac{q}{\sqrt{1-p^2}} \\ \frac{q}{\sqrt{1-r^2}} & \frac{p}{\sqrt{1-q^2}} & \frac{r}{\sqrt{1-p^2}} \end{bmatrix} \begin{bmatrix} \sqrt{1-r^2}s_1 \\ \sqrt{1-q^2}s_2 \\ \sqrt{1-p^2}s_3 \end{bmatrix}, \tag{64}$$

where $p^2 + q^2 + r^2 = 1$. For the sake of clarity, we set $q = p^2$ and $r = p^3$; therefore, the mixing matrices are determined by the parameter p. Under this setting, the dominating component of x_i in the 3×3 BSS system is the source signal s_i. The influence of the source signal s_i on the observation channels is also determined by the parameter p. The larger the value of p, the smaller will be the influence of the source signal s_i on the observation channel x_i and the greater will be the influence of s_i on the other observation channels. Note that the source signals in Equation (64) are scaled to ensure that the source signals 1 and 2 in the 3×3 and 2×3 BSS systems have the same SIR^{in} values.

Figures 2 and 3 plot, respectively, the SIR^{in} versus the SIR^{out} graphs of source signals obtained by different BSS methods for 3×3 and 2×3 BSS systems with various p values. In each subgraph, three sets of measurements are plotted for each BSS method: one in a noiseless environment and two in a noisy environment. The performances in terms of SIRI for the 3×3 and 2×3 systems are summarized in Tables 1 and 2 respectively. In the tables, the method that yields the highest score in each category is highlighted . For the 3×3 BSS systems, the NCA method achieves the best performance in all cases except when the source signal is a piecewise smooth signal, the environment is noiseless, and the mixing matrix parameter p is 0.5588.

For the 2×3 BSS systems, the DSS method outperforms the NCA method in all cases, except when the source signal is the superposition of two sinusoids, the noise standard deviation is 0.02, and the mixing matrix parameter p is set at 0.4365 or 0.5588. The superior performance of the DSS method is mainly due to the almost perfect recovery of the chirp signal in a low noise environment. In the noiseless case, the SIR^{out} of the DSS is more than 30 dB higher than that of other methods, as shown in (a2), (b2), and (c2) of Figure 3. Note that FastICA is not optimized for a noisy environment. Moreover, it cannot be implemented with under-deterministic systems, so we do not compare it with such systems.

B. Synthesized Data Set 2

In the second case, shown in Figure 4, the original source signals are three wide-band signals generated from a t-copula with a linear correlation matrix and 3 degrees of freedom. The t-copula is a multivariate probability distribution for which the marginal probability is the student t distribution; and the correlations between

the source signals are represented as the following 3×3 symmetry matrix:

$$\begin{bmatrix} 1 & 0.2 & 0.5 \\ 0.2 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}, \tag{65}$$

where the element (i, j) indicates the normalized correlation between the i-th and j-th source signals. The 3×3 and 3×2 mixing matrices used to obtain the observation signals are based on Equations (63) and (64) respectively.

The performances in terms of SIRI for the 3×3 and 2×3 systems with different p values are shown in Tables 3 and 4 respectively. In the tables, the method that yields the highest score in each category is highlighted. On this data set, except for the noiseless 2×3 system with $p = 0.5588$ shown in Table 4, the NCA method achieves the best performance in all cases irrespective of the source signals, mixing matrices, and parameter values used.

The parameter values for the two synthesized data sets of the NCA methods were derived by the method described in Section 5.2. Each point of the NCA method in Figures 2 and 3 represents a different parameter value. The parameter μ is a constant, and is set at 10^{-5} in all the experiments. The value of λ is in the range 0.05 to 50, and the geometric ratio $\delta\lambda$ between two successive items is set at 2 . The local scale size L of an operator is selected in the range 2 to 80, and the geometrical ratio δL between two successive items is set at 2.

Summarizing the results of the two data sets, we conclude that 1. the NCA method is more robust than the compared method for the noisy BSS problem; and 2. when p is large, the SIRI gain of NCA over the compared methods is reduced. This might be caused by the initial source signal estimation method because, in general, the performance of a signal channel decomposition method deteriorates when the energy of the source signals is almost equally mixed in the observation channel.

C. Real-life Data

Next, we compare the performances of NCA and FastICA on a set of real-life signals. The experiment involved separating electroocculographic (EOG) artifacts from electroencephalography (EEG) recordings. EEG signals record voltage fluctuations caused by ionic current flows in the neurons of the brain. To obtain EEG signals, multiple electrodes are placed on the subject's scalp (as shown in the right-hand subfigure of Figure 5(b)) to measure the brain's spontaneous electrical activity over a short period of time. In EEG signals, the major noise source is the electro-occulographic (EOG) artifacts caused by eye movements. The artifacts are difficult

to suppress, but the EOG signals can be measured during the EEG recording by placing the electrodes under the eyes, as shown in the left-hand subfigure of Figure 5(b). The recorded EOG signals are used as reference signals to solve the BSS problem.

Our subject was a healthy young man who was in resting state during the collection of EEG and EOG signals. The original size of each observation signal was $100,000$ samples. We took the sampling points from $42,000$ to $46,000$, and reduced the size of the segment through down-sampling by a factor of four. The resultant observations are shown in Figure 5(a). From the top to the bottom, the observation signals are read from $F5$, $F3$, $F1$, FZ, $F2$, $F4$, and $F6$ respectively. Based on those seven signals and the status of the subject, we assume the BSS problem estimates a 7×3 non-negative mixing matrix and three source signals; two of the signals are EOGs and the third is an EEG. The FastICA results are shown in Figure 6(b). The source signals are based on the average results after executing the FastICA algorithm 20 times, with the results of each execution well-aligned. The NCA results are shown in Figure 6(c). The two EOG signals in Figure 6(a) are used as initial sources; and the third initial source is the complement signal of the average of the seven observation signals in the space of the reference signals. The complement signal is derived by Method **C** described in Section 5.1.

The similarity of the source signals derived by various methods is measured by the absolute value of the normalized inner products of the signals. Let s_1 and s_2 be two (row) signals. The normalized inner product, which is defined as $\frac{(s_1)s_2^T}{||s_1||||s_2||}$, normalizes the norm of a signal so that the scaling uncertainty of the BSS problem can be ignored in the comparison. The similarity, denoted as C, is shown in Table 5, where $\hat{S}$ and S_{ICA} represent the source signals of NCA and FastICA respectively, and S_0 denotes the reference signals. In the $C(\hat{S}, S_0)$ column of Table 5, the second and third NCA source signals, $\hat{S}_2$ and $\hat{S}_3$, are similar to the two reference EOG signals, S_{01} and S_{02}. Moreover, in the same column, the first NCA source signal, $\hat{S}_1$, is not similar to any reference signal. Therefore, we can infer that NCA's first signal is an EEG signal whose EOG artifacts have been removed and are contained in the second and third source signals. In the $C(S_{ICA}, S_0)$ column of the table, FastICA's second source signal, S_{ICA2}, is similar to all the reference signals. However, FastICA's first and third source signals, S_{ICA1} and S_{ICA3}, are not similar to any reference signals because their C-values measured with the references are small. Thus, the FastICA method finds two EEG signals (S_{ICA1} and S_{ICA3}) and one EOG signal (S_{ICA2}). Finally, the $C(S_{ICA}, \hat{S})$ column shows that the conclusions derived by FastICA and

NCA are consistent because the EEG and EOG signals obtained by both methods have high C-values.

Table 5

$C(\hat{S}, S_0)$			$C(S_{ICA}, S_0)$			$C(S_{ICA}, \hat{S})$			
	S_{01}	S_{02}		S_{01}	S_{02}		$\hat{S}_1$	$\hat{S}_2$	$\hat{S}_3$
$\hat{S}_1$	0.2081	0.0383	S_{ICA1}	0.1302	0.2179	S_{ICA1}	**0.6681**	0.2989	0.0859
$\hat{S}_2$	**0.8471**	**0.6072**	S_{ICA2}	**0.6314**	**0.7562**	S_{ICA2}	0.1062	**0.7631**	**0.8518**
$\hat{S}_3$	**0.8006**	**0.7186**	S_{ICA3}	0.0860	0.0123	S_{ICA3}	**0.6887**	0.3208	0.1701

6 Concluding Remarks

We have proposed a deterministic method called the Null Space Component Analysis (NCA) algorithm to solve the noisy BSS problem. NCA is an adaptive approach because it derives signal-dependent operators to separate signals. Theoretically, we show that the NCA model can resolve the rotation ambiguity in the BSS problem and separate linearly independent source signals. It can also be applied to over-deterministic and under-deterministic BSS systems. Numerically, we demonstrate the efficiency of the NCA algorithm in a noisy environment and on real-life EEG signals by comparing its performance with that of other methods. The proposed method could be improved and extended in several directions, such as applying it to time-frequency representations of signals, and approximating the NCA constraint with convex functions.

A Appendix

We derive the necessary condition for the source signals in the optimization problem defined in Equation (44). First, we divide the following equation into three components;

$$\min_S \frac{1}{\lambda} \sum_{i=1}^{m} \left[\|\mathcal{T}_i s_i^T\|_2^2 - \mu \sum_{j=1}^{m} \left(\|\mathcal{T}_i s_j^T\|_2^2 - \|\mathcal{T}_i s_i^T\|_2^2 \right) \right] + \|X - AS\|_F^2, \qquad (66)$$

1. For the first component, we have

$$
\begin{aligned}
&[(m-1)\mu+1]\sum_{i=1}^{m}\|\mathcal{T}_i s_i^T\|_2^2 \\
=&\,[(m-1)\mu+1]\sum_{i=1}^{m} s_i\mathcal{T}_i^T\mathcal{T}_i s_i^T \\
=&\,[(m-1)\mu+1]\,[s_1,\ldots,s_m]\begin{bmatrix} \mathcal{T}_1^T\mathcal{T}_1 & 0 & \ldots & 0 \\ 0 & \mathcal{T}_2^T\mathcal{T}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathcal{T}_m^T\mathcal{T}_m \end{bmatrix}[s_1,\ldots,s_m]^T. \qquad (67)
\end{aligned}
$$

Let e_i be the standard basis of R^m. Equation (67) can be expressed as

$$
\begin{aligned}
&[(m-1)\mu+1]\sum_{i=1}^{m} e_i^T S\mathcal{T}_i^T\mathcal{T}_i S^T e_i \\
=&\,[(m-1)\mu+1]\sum_{i=1}^{m}\mathbf{Tr}\left[\left(e_i^T S\right)\left(\mathcal{T}_i^T\mathcal{T}_i S^T e_i\right)\right] \\
=&\,[(m-1)\mu+1]\sum_{i=1}^{m}\mathbf{Tr}\left(\mathcal{T}_i^T\mathcal{T}_i S^T e_i e_i^T S\right), \qquad (68)
\end{aligned}
$$

where $\mathbf{Tr}(\cdot)$ is the trace operation. If we take the derivative of Equation (68) with respect to S, we obtain

$$
2\,[(m-1)\mu+1]\sum_{i=1}^{m} e_i e_i^T S\mathcal{T}_i^T\mathcal{T}_i. \qquad (69)
$$

2. For the second component, we have

$$
\begin{aligned}
&(-\mu)\sum_{j=1}^{m}\sum_{k\neq j}^{m}\|\mathcal{T}_k s_j^T\|_2^2 \\
=&\,(-\mu)\sum_{j=1}^{m} e_j^T S\left(\sum_{k\neq j}^{m}\mathcal{T}_k^T\mathcal{T}_k\right)S^T e_j \\
=&\,(-\mu)\sum_{j=1}^{m}\mathbf{Tr}\left[\left(\sum_{k\neq j}^{m}\mathcal{T}_k^T\mathcal{T}_k\right)S^T e_j e_j^T S\right]. \qquad (70)
\end{aligned}
$$

Taking the derivative of Equation (70) with respect to S, we obtain

$$
2\,(-\mu)\sum_{j=1}^{m} e_j e_j^T S\left(\sum_{k\neq j}^{m}\mathcal{T}_k^T\mathcal{T}_k\right). \qquad (71)
$$

3. For the third component, we have

$$
\begin{aligned}
&\|X-AS\|_F^2 \\
=&\;\mathbf{Tr}[(X-AS)^T(X-AS)] \qquad (72)\\
=&\;\mathbf{Tr}(X^TX)-2\mathbf{Tr}(X^TAS)+\mathbf{Tr}(S^TA^TAS).
\end{aligned}
$$

Taking the derivative of Equation (72) with respect to S, we obtain

$$-2A^T X + 2A^T AS. \tag{73}$$

Finally, we combine the results derived from each component to obtain the necessary condition for S:

$$\begin{aligned} 0_m = [(m-1)\mu + 1] \sum_{i=1}^{m} e_i e_i^T S(\mathcal{T}_i^T \mathcal{T}_i) + (-\mu) \left[\sum_{j=1}^{m} e_j e_j^T S \left(\sum_{k \neq j}^{m} \mathcal{T}_k^T \mathcal{T}_k \right) \right] \\ + (-\lambda) A^T X + \lambda A^T AS. \end{aligned} \tag{74}$$

References

[1] P. Comon and C. Jutten, "Handbook of Blind Source Separation: Independent Component Analysis and Applications," *Academic Press*, Feb. 22, 2010.

[2] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. on Signal Processing*, vol.52, Issue.7, pp. 1830-1847, 2004.

[3] A. Aissa-el-Bey, K. Abed-Meraim, and Y. Grenier, "Blind separation of underdetermined convolutive mixtures using their time-frequency representation," *IEEE transactions on audio, speech and language processing*, July 2007, vol. 15, no. 5, pp. 1540-1550.

[4] Y. Deville and M. Puigt, "Temporal and time-frequency correlation-based blind source separation methods. Part I: Determined and underdetermined linear instantaneous mixtures," *Signal Processing*, vol. 87, Issue. 3 , pp.374-407, 2007.

[5] M. Puigt, A. Griffin, and A. Mouchtaris, "Post-nonlinear speech mixture identification using single-source temporal zones and curve clustering," *Proceedings of EUSIPCO*, pp. 1844-1848, Barcelona, Spain, August 29-September 2, 2011.

[6] L. T. Duarte, R. A. Ando, R. Attux, Y. Deville, and C. Jutten, "Separation of Sparse Signals in Overdetermined Linear-Quadratic Mixtures", *Springer Lecture Notes on Computer Science - 10th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2012)*, Tel-Aviv, Israel, 2012.

[7] S. Cruces, "Bounded component analysis of linear mixtures: A criterion of minimum convex perimeter," *IEEE Trans. on Signal Processing*, vol. 58, Issue. 4, pp.2141-2154, 2010.

[8] A. T. Erdogan, "*A family of bounded component analysis algorithm*," IEEE ICASSP, pp.1881-1884, March 2012.

[9] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, "Source estimation in noisy sparse component analysis, *Proceedings of IEEE 15th International Conference on Digital Signal Processing (DSP2007)*, Cardiff, UK, June 2007, pp. 219-222.

[10] H. Zayyani, M. Babaie-Zadeh, and C. Jutten, "On the Cramer-Rao bound for estimating the mixing matrix in noisy sparse component analysis," *IEEE Signal Processing Letters*, vol. 15, pp. 609-612, 2008.

[11] G. Darmois, "Analyse géngérale des liaisons stochastique," *Rev. Inst. Intern. stat.*, 21, pp. 2-8, 1953.

[12] A. Belouchrani, K. Aded-Meraim, J. F. Cardoso, and E. Moulines, "A Blind Source Separation Technique Using Second-Order Statistics," *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434-444, 1997.

[13] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 3, no. 36, pp.287-314, 1994.

[14] L. De Lathauwer, J. Castaing, and J.-F.Cardoso, "Fourth-order cumulant-based identification of underdetermined mixtures," *IEEE Trans. on Signal Processing*, vol. 55, pp. 2965-2973, 2007.

[15] A. Hyvarinen, and E. Oja, "Independent component analysis: algorithms and applications," *IMA Journal of Mathematics Applied in Medicine and Biology*, vol. 13(4-5), pp. 411-430, May-Jun 2000.

[16] S. Ikeda and K. Toyama, "Independent component analysis for noisy data X MEG data analysis," *Neural Networks*, vol. 13, Issue. 10, pp.1063-1074, Dec 2000.

[17] S. Narasimhan and S. Shan, "Model identification and error covariance matrix estimation from noisy data using PCA," *Control Eng. Prac.*, 16, 146-155, 2008.

[18] S. C. Douglas, A. Cichocki, and S. Amari, "Bias removal technique for blind source separation with noisy measurements," *Signal Processing*, vol.34, Issue. 14, pp. 1379-1380 , 1998.

[19] A. Hyvarinen, "Independent component analysis in the presence of Gaussian noise by maximizing joint likelihood," *Neurocomputing*, 22:49-67, 1998.

[20] A. Cichocki, S.C. Douglas and S. Amari, "Robust techniques for independent component analysis (ICA) with noisy data," *Neurocomputing*, vol. 22, pp.113-129, 1998.

[21] A. Hyvarinen, "Gaussian moments for noisy independent component analysis," *IEEE Signal Processing Letters*, vol. 6, Issue.6, pp.145-147, June 1999.

[22] H. Attias, "Independent factor analysis," *Neural Computation*, vol. 11, No. 4, pp. 803-851, May 1999.

[23] P. R. Babu and S. Narasimhan, "Multivariate techniques for preprocessing noisy data for source separation using ICA," *International Journal of Advances in Engineering Sciences and Applied Mathematics*, vol. 4, Issue.1-2, pp.32-40, June 2012.

[24] M. Davies, "*Identifiability, subspace selection and noisy ICA*," Independent component analysis and blind signal separation -Fifth International Conference, ICA 2004, pp.152-159, 2004.

[25] L. Vielva and D. Erdogmus, and J. C. Principe, "Underdetermined Blind Source Separation Using a Probabilistic Source Sparsity Model," *In 2nd International Workshop on Independent Component Analysis and Blind Source Separation*, 2001.

[26] P. Bofill and M. Zibulevsky, "Underdetermined Blind Source Separation Using Sparse Representation," *Signal Processing* 81, pp. 2353-2364, 2001.

[27] J. Bobin, J.-L. Starck, J. Fadili, and Y. Moudden, "*Sparsity and Morphological Diversity in Blind Source Separation*," IEEE Trans. on Image Processing, vol. 16, No. 11, pp. 2662-2674, 2007.

[28] S. Nam, M. Davies, M. Elad, and R. Gribonval, "The Cosparse Analysis Model and Algorithms," *Applied and Computational Harmonic Analysis*, Vol. 34, No. 1, pp. 30-56, 2013.

[29] M. Zibulevsky, B. A. Pearlmutter, P. Bofill, and P. Kisilev, "Blind Source Separation by Sparse Decomposition," chapter in the book:S. J. Roberts, and R. M. Everson eds. I*ndependent Component Analysis: Principles and Practice*, Cambridge, 2001.

[30] Y. Li, S. I. Amari, A. Cichocki, D. W. C. Ho, and S. Xie, "Underdetermined Blind Source Separation Based on Sparse Representation," *IEEE Trans. on Signal Processing*, vol.54, pp.423-437, 2006.

[31] M. Aharon, M. Elad, and A. M. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Trans. on Signal Processing*, vol.54, no. 11, 2006, pp.4311-4322.

[32] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, pp.903-995, 1998.

[33] S. L. Peng and W. L. Hwang, "Null Space Pursuit: An operator-based Approach to adaptive signal separation," *IEEE Trans. on Signal Processing*, vol.58, no.5, pp. 2475-2483, 2010.

[34] S. Vorobyov and A. Cichocki, "Blind noise reduction for multisensory signals using ICA and subspace filtering, with application to EEG analysis," *Biological Cybernetics*, vol. 86, Issue. 4, pp.293-303, 2002.

[35] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clinical Neurophusiology*, 118, pp. 98-104, 2007.

[36] M. Fazel, "Matrix Rank Minimization with Applications," Ph.D. Thesis, Stanford University, 2002.

[37] NSP codes in Matlab, Available: http://mda.ia.ac.cn/English/publications/publications_index.htm#2010

[38] V. G. Reju, S. N. Koh, and I. Y. Soon, "An algorithm for mixing matrix estimation in instantaneous blind source separation," Signal Processing, Vol. 89, Issue 9, September 2009, pp. 1762-1773.

[39] E. Vincent, S. Araki, and P. Bofill, "The 2008 Signal Separation Evaluation Campaign: A community-based approach to large-scale evaluation," Proc. of ICA, 2009, pp. 734-741.

[40] S. I. Amari, A. Cichocki, and H. H. Yang, "A New Learning Algorithm for Blind Source Separation," Avdances in Neural Information Processing Systems, 8, 757-763, 1996.

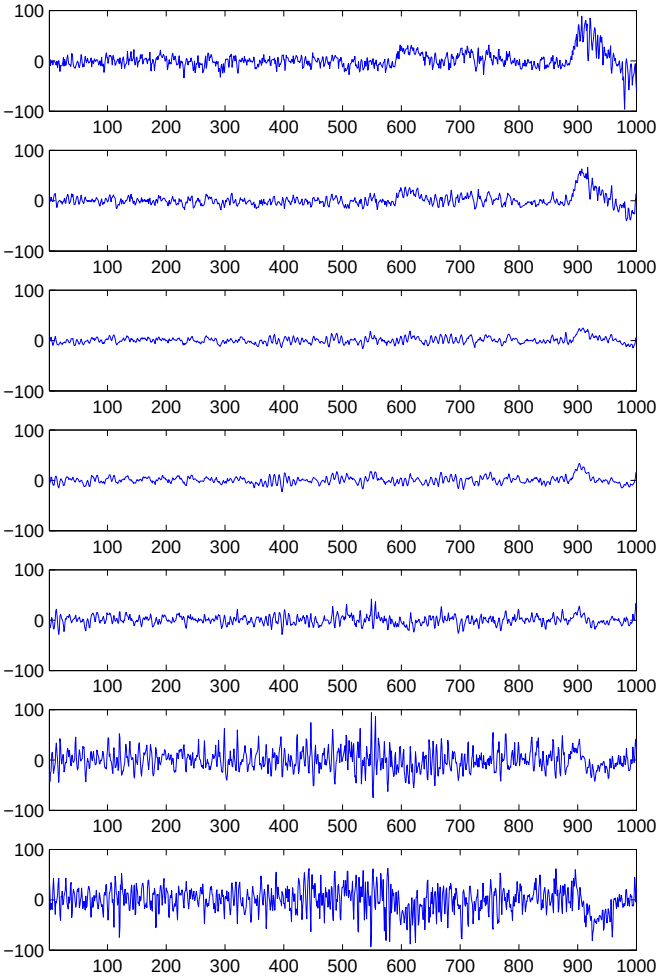
[41] FastICA package in Matlab, Available: http://research.ics.aalto.fi/ica/fastica/

[42] DSS package in Matlab, Available: http://www.cis.hut.fi/projects/dss/package/

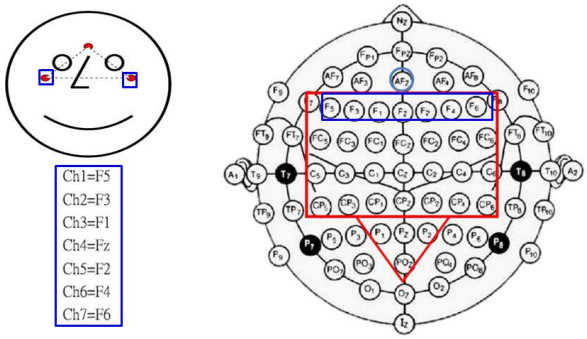
[43] J. Särelä , H. Valpola, "Denoising Source Separation," Journal of Machine Learning Research, vol.6, pp.233-272, 2005.

[44] Wavelet Packet Preprocessor for Independent Component Analysis in Matlab, Available: http://visl.technion.ac.il/bron/wpica

[45] R. Carmona, W. L. Hwang, and B. Torresani, "Characterization of Signals by the Ridges of Their Wavelet Transforms," IEEE Transactions on Signal Processing, vol. 45 (10), pp. 2586-2590, 1997.

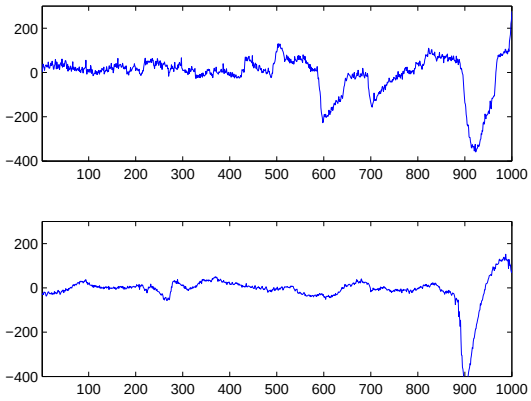


(a)

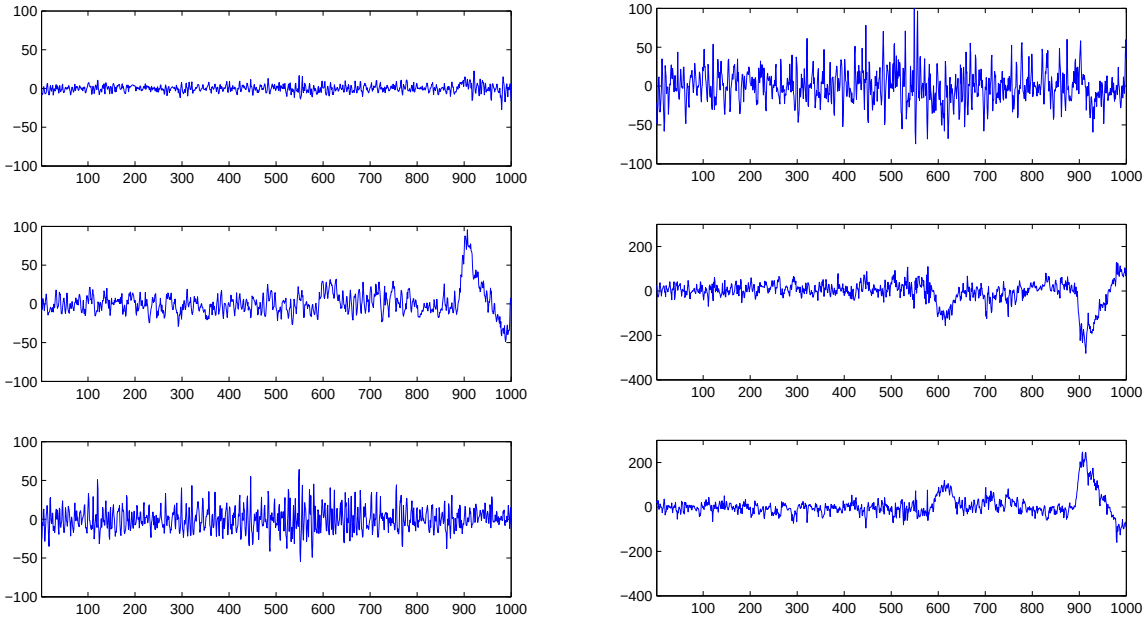


(b)

Figure 5: Recordings of the EOG and EEG signals of a healthy subject in a resting state: (a) the EEG observation signals (Ch1-Ch7) are read from electrodes, as shown in subfigure (b); (b) the positions of the EOG and EEG electrodes. The blue region in the right-hand subgraph indicates the positions for recording the EEG signals. The EOG reference signals are read from the blue boxes under the eyes in the left-hand subgraph.



(a)



(b) (c)

Figure 6: Comparison of the source separation results of the NCA and FastICA methods. (a) The EOG reference signals: S_{01} (top) and S_{02} (bottom). (b) The source signals derived by FastICA: from top to bottom, the signals are S_{ICA1}, S_{ICA2}, and S_{ICA3}. (c) The source signals derived by NCA: from top to bottom, the signals are $\hat{S}_1, \hat{S}_2$, and $\hat{S}_3$. The results of FastICA are the averages of running the algorithm 20 times. The parameter values of the NCA method are as follows: the λ is set at 0.1, the local scale size L of all seven operators is set at 3, and μ is set at 10^{-5}.