

# A New Approach to Blind Separation of Overdetermined Mixtures

Jie Yang Jiashu Zhang

Sichuan Province Key Lab of Signal and Information Processing  
Southwest Jiaotong University, Chengdu 610031, P. R. China

## Abstract

Different from the general way to overdetermined blind source separation problem whose demixing matrix is a nonsquare one just as the mixing matrix, this paper proposed that using  $n$  (the number of sources) elements of the mixture signals as the new mixture signals to be separated, this means, the demixing matrix becomes a square one. First, the simulation shows the new approach works efficiently as the general approach. Second, the feasibility of the new approach is proved by comparing the entropy of the different mixtures. Finally, some interesting conclusions are obtained.

**Keywords:** Blind source separation, Overdetermined, Entropy

## 1. Introduction

The blind source separation (BSS) intends to extract statistically independent components from the sensor signals which are given as mixtures of unknown source signals, moreover, the mixing environment also isn't known a priori. Recently, this kind of blind technique has become an increasingly important research area due to its significant potential applications in various fields, such as telecommunication systems, image enhancement, sonar and radar systems, speech signal processing and biomedical signal processing [1].

Since the pioneering work of Jutten and Herault [2], various algorithms with different viewpoints have been proposed for BSS [3]-[6]. In addition, most of these algorithms in the literatures assume that the number of sensors is equal to the number of sources, thus, the mixing matrix is a square one, whereas, in most practical cases, this assumption is not always hold good. When sensors more (fewer) than sources, the BSS problem is referred to overdetermined or undercomplete [7] (underdetermined or overcomplete [8]) BSS. Therefore, in general opinion, the demixing matrix of the

overdetermined mixtures is not square just as the mixing matrix. But in practical experiments we found that it's not necessary to use a  $n \times m$  ( $n$ : the number of sources,  $m$ : the number of sensors) demixing matrix, instead, a  $n \times n$  demixing matrix also can works well. Further more, it's plausible to use overdetermined mixtures to improve upon the extracted signals of interest from mixtures. In this paper, we focus on the overdetermined mixtures, and prove the feasibility of the new approach with information geometry.

## 2. Problem statement

Assume that there exists an  $n$ -dimensional unobserved source vector  $\mathbf{s}_t = [s_1(t), \dots, s_n(t)]^T$  that are zero-mean and mutually statistically independent at each instant  $t$ . In the noise-free instantaneous case, the available sensor vector  $\mathbf{x}_t = [x_1(t), \dots, x_m(t)]^T$  is given by

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

where  $\mathbf{A} \in \mathbf{R}^{m \times n}$  is a constant and unknown mixing matrix of full rank. For the difficult case where there are fewer mixtures than sources,  $m < n$ , separation may be achievable only when the information about sources and mixing environment is known as a priori [8]-[9]. Therefore, in this paper we consider the overdetermined case,  $m > n$ . The objective of blind separation is to recover original signals  $\mathbf{s}_t$  from the observation  $\mathbf{x}_t$  without prior knowledge on the source signals and the mixing matrix  $\mathbf{A}$  except for the independence of source signals. For this purpose, the general solution is to adjust the so-called separating matrix  $\mathbf{W} \in \mathbf{R}^{n \times m}$  to optimize the criterions under certain cost function, then the output vector

$$\mathbf{y}_t = \mathbf{W}\mathbf{x}_t \quad (2)$$

is the estimation of the source signals.

In [7], Zhang *et al.* had proved that the natural gradient is available to blind separation of overdetermined mixtures in virtue of the Lie group and

Riemannian metric on the manifold of mixtures. Later, Ye *et al.* [10] showed that the mutual information of outputs of the separation network is a cost function for overdetermined BSS. Thus, the on-line learning algorithm is in the form

$$\Delta \mathbf{W}_t = \eta_t [\mathbf{I} - \Phi(\mathbf{y}_t) \mathbf{y}_t^T] \mathbf{W}_t \quad (3)$$

where  $\mathbf{W}_t$  is, in general, a nonsquare  $n \times m$  demixing matrix,  $\eta_t$  is a positive learning rate parameter and  $\Phi(\mathbf{y}_t) = [\varphi_1(y_1(t)), \dots, \varphi_n(y_n(t))]^T$  is a nonlinear-transformed vector of  $\mathbf{y}_t$ . This algorithm works efficiently and shows an "equivariant" [11] property. But, in our opinion, there exists redundant information in the overdetermined mixture signals  $\mathbf{x}_t$ , as when the sensor number  $m$  is equal to the source number  $n$ , the BSS problem can be solved well. Therefore, why don't we choose  $n$  elements from the  $m$ -dimensional sensor vector  $\mathbf{x}_t$  as the new mixtures to be separated? Thus, pursue a square  $n \times n$  demixing matrix instead of a nonsquare one.

Let the set of  $m$  elements of  $\mathbf{x}_t$  is  $X = \{x_1, x_2, \dots, x_m\}$ , and  $X_i$  ( $1 \leq i \leq P_m^n$ ) is the set which consists of the  $i$ th permutation of  $n$  elements of set  $X$ .  $\hat{\mathbf{x}}_t = [\hat{x}_1^i(t), \hat{x}_2^i(t), \dots, \hat{x}_n^i(t)]^T$ , where  $\hat{x}_j^i(t)$  denotes the  $j$ th element of the new vector  $\hat{\mathbf{x}}_t$ , and is equal to the  $j$ th element of set  $X_i$ , i.e.,  $\hat{\mathbf{x}}_t$  is the new  $n$ -dimensional vector whose elements are selected from  $\mathbf{x}_t$ . Therefore, the main objective of overdetermined BSS problem becomes to search an optimization for the square  $n \times n$  demixing matrix  $\hat{\mathbf{W}}$ , the output vector is in the form

$$\mathbf{y}_t = \hat{\mathbf{W}} \hat{\mathbf{x}}_t \quad (4)$$

In practical, we found that this new approach works as efficiently as the general way. For instance, we consider separation of the following source signals:

- Sign signal:  
 $s_1(t) = \text{sgn}(\cos(2\pi 155t));$
- Amplitude-modulated signal:  
 $s_2(t) = \sin(2\pi 90t) \sin(2\pi 300t);$
- Phase-modulated signal:  
 $s_3(t) = \sin(2\pi 300t + 6 \cos(2\pi 60t)).$

In simulations, there are five sensors, and the elements of  $5 \times 3$  mixing matrix  $\mathbf{A}$  are randomly assigned in the range  $[-1, +1]$ . For simplicity, the front 3 elements of the sensor vector  $\mathbf{x}_t$  are selected as  $\hat{\mathbf{x}}_t$  for the new approach. To measure the performance of the different ways, we introduce the

cross-talking error as the performance index [12],

$$\text{PI} = \frac{1}{m} \sum_{p=1}^m \left( \sum_{q=1}^n \frac{|c_{pq}|}{\max_l |c_{pl}|} - 1 \right) + \frac{1}{n} \sum_{q=1}^n \left( \sum_{p=1}^m \frac{|c_{pq}|}{\max_l |c_{lq}|} - 1 \right) \quad (5)$$

where  $\mathbf{C} = \mathbf{W}\mathbf{A} = \{c_{pq}\}$  is the combined mixing-separating matrix. Then, the performance of two approaches is showed in the Figure 1.

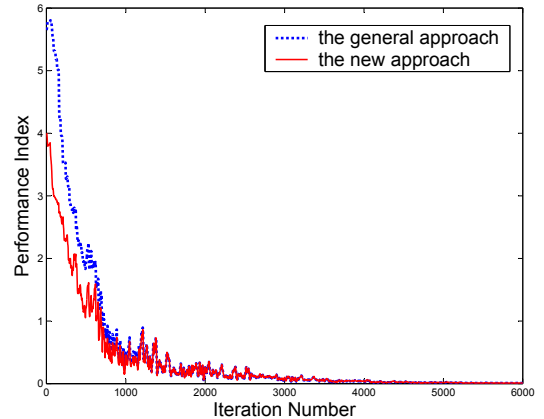


Fig. 1: Average performance indexes over 100 independent runs of the two approaches.

According to Figure 1, the new approach works as well as the general one, even more gets a better performance at the primary separating, and when the time goes, the different ways almost have equal performance at all as they converge in the same direction. Consequently, it's nonsignificant to use overdetermined mixtures to improve the separation algorithm, and actually, we can use the  $n$ -dimensional mixtures instead of the overdetermined mixtures with the same performance and less computation. In the next section, we study the validity of the new approach by comparing the (differential) entropy of mixtures.

### 3. Entropy of mixtures

Decomposing the mixing matrix in the following form (such as QR factorization or singular value decomposition):

$$\mathbf{A} = \mathbf{Q}^T \begin{bmatrix} \mathbf{A}_1 \\ 0 \end{bmatrix} \quad (6)$$

where matrix  $\mathbf{Q} \in \mathbf{R}^{m \times m}$  is an orthogonal normalized matrix,  $\mathbf{A}_1 \in \mathbf{R}^{n \times n}$  is a nonsingular matrix.

So the mixing model can be denoted as:

$$\mathbf{x} = \mathbf{A}\mathbf{s} = \mathbf{Q}^T \begin{bmatrix} \mathbf{A}_1\mathbf{s} \\ 0 \end{bmatrix} \quad (7)$$

Since the entropy is referred to a variable not a sample, the subscript  $t$  of the vectors is omitted in above equation. An important property of the entropy is that it is invariant for orthogonal normalized linear transformations. Thus, we have the entropy of mixtures of general approach:

$$H(\mathbf{x}) = H(\mathbf{A}_1\mathbf{s}) = H(\mathbf{s}) + \log |\det(\mathbf{A}_1)| \quad (8)$$

For simplicity, we consider the front  $n$  elements of the mixtures  $\mathbf{x}$  as the new mixtures  $\hat{\mathbf{x}}$ , because choosing another  $n$  elements is equivalent to add a permutation matrix to  $\mathbf{x}$ , and it is do nothing to the entropy of  $\hat{\mathbf{x}}$ . So the decomposition of  $\mathbf{x}$  with respect to  $\hat{\mathbf{x}}$  is given by:

$$\mathbf{x} = \begin{bmatrix} \hat{\mathbf{x}} \\ \tilde{\mathbf{x}} \end{bmatrix} \quad (9)$$

Correspondingly, we decompose the orthogonal normalized matrix:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_2 \\ \mathbf{Q}_3 & \mathbf{Q}_4 \end{bmatrix} \quad (10)$$

where  $\mathbf{Q}_1 \in \mathbf{R}^{n \times n}$  and  $\mathbf{Q}_2 \in \mathbf{R}^{n \times (m-n)}$ . According to equation (9) and (10), we obtain the entropy of mixtures of new approach:

$$\begin{aligned} H(\hat{\mathbf{x}}) &= H(\mathbf{Q}_1^T \mathbf{A}_1 \mathbf{s}) \\ &= H(\mathbf{s}) + \log |\det(\mathbf{A}_1)| + \log |\det(\mathbf{Q}_1^T)| \end{aligned} \quad (11)$$

Comparing the equation (8) and (11), we can find the relationship readily between the new mixtures  $\hat{\mathbf{x}}$  and the overdetermined mixtures  $\mathbf{x}$  in the viewpoint of information theory, that is,  $\hat{\mathbf{x}}$  is achieved when  $\mathbf{x}$  gets through another linear transformation  $\mathbf{Q}_1^T$  which is determined by the orthogonal normalized matrix  $\mathbf{Q}$ . Here, introduced another matrix  $\mathbf{B}_1 = \mathbf{A}_1 \mathbf{Q}_1^T$ , thus the relationship is illustrated in the Figure 2.

Obviously, the structure of the model of new approach is similar to the general one, and they are equivalent in the information-theoretic separation algorithms by introducing the matrix  $\mathbf{B}_1$  since there is not more information about  $\mathbf{x}$  than  $\hat{\mathbf{x}}$ . Therefore, we can get the following conclusion: it's available that using  $n$ -dimensional vector whose elements is selected from the overdetermined mixtures instead of the  $m$ -dimensional vector to be separated.

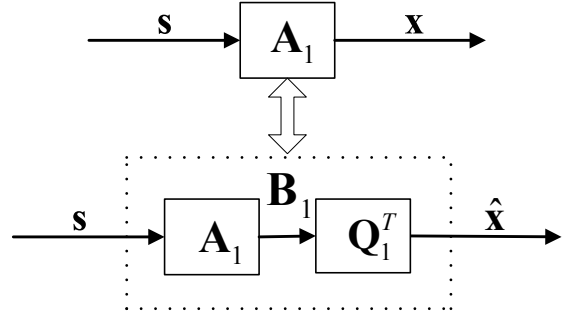


Fig. 2: the relationship of the two different approaches in information-theoretic sense.

## 4. Conclusion

In this paper, we study the overdetermined BSS problem where the sensor number  $m$  is not less than the source number  $n$ . Comparing with the general approach that using all sensor signals as the mixtures to be separated, a new approach that selecting  $n$  sensor signals as the mixture signals to be separated is proposed. Besides, the validity of this new approach is also be proved by means of comparing the entropy of the different mixtures. Obviously, the new approach needs less computation than the general one, because we search an optimal  $n \times n$  demixing matrix instead of a  $n \times m$  one, especially in array signal processing where the sensor number  $m$  may be large more than the source number  $n$ . Moreover, there is an additional interesting conclusion: now that the new approach works efficiently as the general one, that means, it's unuseful to use more sensors than sources to improve upon blind source separation algorithms in extracting the signals of interest from mixtures. A complete theoretical proof of this interesting conclusion will be given in the future work.

## References

- [1] A. Hyvarinen, J. Karhunen and E. Oja, editors. *Independent Component Analysis*, Wiley, New York, 2001.
- [2] C. Jutten and H. Herault, Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture, *Signal Processing*, 24:1–10, Elsevier, 1991.
- [3] A.J. Bell and T.J. Sejnowski, An information maximization approach to blind separation and blind deconvolution, *Neural Computation*, 7:1129–1159, MIT, 1995.

- [4] S. Amari, Natural gradient works efficiently in learning, *Neural Computation*, 10:251–276, MIT, 1998.
- [5] A. Hyvarinen, Fast and robust fixed-point algorithm for independent component analysis, *IEEE Trans. On Neural Network*, 10:626–634, IEEE, 1999.
- [6] X.-L. Zhu and X.-D. Zhang, Adaptive RLS algorithm for blind source separation using a natural gradient, *IEEE Signal Processing Letters*, 9:432–435, IEEE, 2002.
- [7] L.-Q. Zhang, A. Cichocki and S. Amari, Natural gradient algorithm for blind separation of overdetermined mixture with additive noise, *IEEE Signal Processing Letters*, 6:293–295, IEEE, 1999.
- [8] M.S. Lewick and T.J. Sejnowski, Learning overcomplete representations, *Neural Computation*, 12:337–365, MIT, 2000.
- [9] Y.-Q. Li, S. Amari, A. Cichocki, D.W.C. Ho and S.-L. Xie, Underdetermined blind source separation based on sparse representation, *IEEE Trans. on Signal Processing*, 54:423–437, IEEE, 2006.
- [10] J.-M. Ye, X.-L. Zhu and X.-D. Zhang, Adaptive blind separation with an unknown number of sources, *Neural Computation*, 16:1641–1660, MIT, 2004.
- [11] J. F. Cardoso and B. Laheld, Equivariant adaptive source separation, *IEEE Transactions on Signal Processing*, 44:3017–3030, IEEE, 1996.
- [12] H. H. Yang and S. Amari, Adaptive no-line learning algorithms for blind separation: Maximum entropy and minimum mutual information, *Neural Computation*, 9:1457–1482, MIT, 1997.