Available online at www.sciencedirect.com
ScienceDirect
Progress in Natural Science 19 (2009) 891–896
Progress in Natural Science
www.elsevier.com/locate/pnsc

A new geometric approach to blind source separation of bounded sources

Jinlong Zhang [a], Guoxu Zhou [a,*], Zuyuan Yang [a], Xiaoxin Liao [b]

[a] School of Electrics & Information Engineering, South China University of Technology, Guangzhou 510641, China
[b] Department of Control Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

Received 30 June 2008; received in revised form 21 July 2008; accepted 26 July 2008

Abstract

Based on the minimum-range approach, a new geometric approach is proposed to deal with blind source separation in this paper. The new approach is the batch mode of the original minimum-range approach. Compared with the original approach, the optimization algorithm of the proposed approach needs no parameters and is more efficient and reliable. In addition, the extension of minimum-range-based approaches is discussed. The simulations show the efficiency of the proposed approach.

Keywords: Blind source separation; Independent component analysis; Minimum-range approach

1. Introduction

Blind source separation (BSS) is a problem of recovering the underlying sources only from their mixtures [1,2]. Its linear instantaneous model is

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t) \tag{1}$$

where $\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_m(t)]^T$ and $\boldsymbol{s}(t) = [s_1(t), s_2(t), \ldots, s_n(t)]^T$ are the observation vector and source vector at time instant t, respectively. The matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is the mixing matrix, and $\boldsymbol{n}(t)$ is the additive noise. In this paper, the noise is neglected and we assume that $m = n$ (If $m > n$, the observations can be reduced by the principle component analysis (PCA), thus $m = n$ also holds finally). By this assumption, BSS can be realized by finding a matrix $\boldsymbol{B}$ (i.e. the so-called unmixing matrix) such that

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{x} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{s}(t) \tag{2}$$

where $\boldsymbol{P}$ is a permutation matrix and $\boldsymbol{D}$ is a diagonal matrix, thus the sources are recovered up to a permutation and scale [3].

To BSS, some prior information is used to separate the sources. For example, temporal structure, independence, and sparseness. The approaches based on the temporal structure are studied by Stone and Xie et al., and some significant results have been reported in [4,5]. The approaches based on the sparseness are generally studied in the context of sparse component analysis (SCA), and some valuable results can be found in [6–9]. In this paper, the sources are assumed to be bounded and independent, thus the approach proposed in this paper is also an alternative method for independent component analysis (ICA). BSS based on ICA has received many studies and new approaches still arise by using some other prior information [10–14]. These algorithms can be classified into the geometric approaches to BSS, since they prefer to the geometric information rather than the statistical characteristics. However, most of them cannot work efficiently in high dimensions (i.e. $m > 3$) [10,12,15].

In Ref. [11], the authors proposed a novel minimum-range approach. The approach was proved to be reliable

* Corresponding author. Tel./fax: +86 20 87114709.
E-mail address: zhou.guoxu@mail.scut.edu.cn (G. Zhou).

1002-0071/$ - see front matter © 2009 National Natural Science Foundation of China and Chinese Academy of Sciences. Published by Elsevier Limited and Science in China Press. All rights reserved.
doi:10.1016/j.pnsc.2008.07.022

in theorem. However, its optimization algorithm, named ICAForNDC, cannot guarantee the convergence of the algorithm. This algorithm updates the observations by a series of Givens rotations, and the rotation angle is determined by checking the values of $\pi\beta^t$ in the order of $t = 1, 2, \ldots$. Obviously a larger β results in a slower convergence speed, but a smaller one may terminate the algorithm by force even if the objective function has not converged to its stationary point yet. (If β is too small, $\pi\beta^t$ will approach zero after a few iterations, see Fig. 5 in [11] for details.) If the convergence of the objective function is not guaranteed, the correctness of the separation is not guaranteed. From the above analysis, the value of β is crucial to that approach, and the original approach has trouble in setting the value of β.

This paper proposes a new geometric algorithm for BSS. On the one hand, this algorithm is the batch separation mode of the minimum-range approach. On the other hand, the optimization algorithm proposed in this paper is more efficient and reliable than the original algorithm.

2. The proposed algorithm

The unmixing matrix $\boldsymbol{B}$ can be factored as $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{W}$, where $\boldsymbol{W}$ is the whitening matrix and consequently the matrix $\boldsymbol{U}$ is constrained to be orthogonal. In other words, $\boldsymbol{W}$ de-correlates the mixtures and then $\boldsymbol{U}$ rotates the mixtures. Assume that $s1$ and $s2$ are two uniformly distributed signals and they are mutually independent. We see that their scatter plot forms a square whose edges are parallel with the corresponding axis. After mixing via model (1), the scatter plot of their mixtures becomes a parallelogram. After whitening, a rotated version of the original square is obtained. Then the left task of BSS is to determine a rotation which will replace the square to its original position. This procedure can be done easily when there are only two sources, but rather complicated when more than two sources are involved. This paper is devoted to the geometric approach to BSS in high dimensions. In this paper, the observations are always assumed to be pre-whitened and are denoted by $\widetilde{\boldsymbol{x}}$.

A basic observation is, when the sources are independent, that the sum of their projection length on each axis is minimized. To see this, consider two independent uniformly distributed signals again. The sum of the projection length of their whitened mixtures is $l = \mathrm{AD} + \mathrm{DC}$ (see Fig. 1). This sum is minimized to be $\mathrm{A'D'} + \mathrm{D'C'}$ when $\mathrm{A'D'} // e_2$, which means that the sources are unmixed.

Now we consider this procedure in high dimensions. Let $e_i = [0, \ldots, 0, 1, 0, \ldots, 0]^T$, i.e. only the ith element is non-zero and it takes the value of 1. Thus the projection length of observations $\widetilde{\boldsymbol{x}}$ in the direction of e_i equals to

$$l_i = \left(e_i^T \widetilde{\boldsymbol{x}}\right)^{\max} - \left(e_i^T \widetilde{\boldsymbol{x}}\right)^{\min} = \widetilde{x}_i^{\max} - \widetilde{x}_i^{\min} \tag{3}$$

where l_i is often referred to as the range of $\widetilde{x}_i$. Thus the following objective function is used:

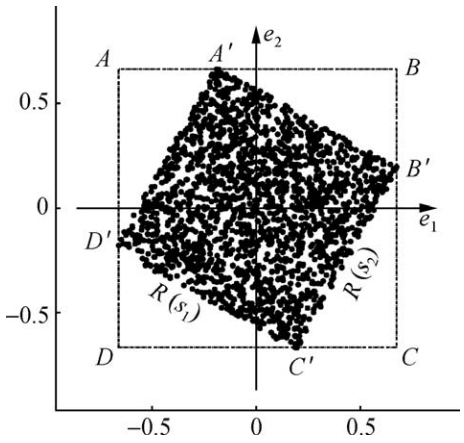

Fig. 1. The sources are unmixed when the sum of the projection length on each axis of the observations is minimized.

$$\min L(\boldsymbol{U}) = \sum_{i=1}^{m} l_i(\boldsymbol{U}\widetilde{\boldsymbol{x}}(t)) \quad s.t.\ \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I} \tag{4}$$

Since the unmixing matrix $\boldsymbol{U}$ must be orthogonal, it can be factorized as a product of a series of Givens (i.e. Jacobi) rotation matrices. In geometric, these matrices can be interpreted by they rotate the scatter plot of the observations till they replace it in its original position before mixing. Let $\boldsymbol{G}_{ij}^{\alpha}$ be a Givens matrix, i.e. $\boldsymbol{G}_{ij}^{\alpha}$ equals the identity matrix $\boldsymbol{I}$ except that $\left[\boldsymbol{G}_{ij}^{\alpha}\right]_{ii} = \left[\boldsymbol{G}_{ij}^{\alpha}\right]_{jj} = \cos(\alpha)$ and $\left[\boldsymbol{G}_{ij}^{\alpha}\right]_{ij} = -\left[\boldsymbol{G}_{ij}^{\alpha}\right]_{ji} = \sin(\alpha)$. Note that, after each rotation of $\widetilde{\boldsymbol{x}}^{(k+1)}(t) \leftarrow \boldsymbol{G}_{ij}^{\alpha}\widetilde{\boldsymbol{x}}^{(k)}(t)$, only the entries in the ith and jth rows of $\widetilde{\boldsymbol{x}}^{(k)}$ are changed, thus the minimization of (4) is equivalent to minimization of $l_i + l_j$.

Write $\widetilde{\boldsymbol{x}}_{ij}(t) = \left[\widetilde{x}_i(t), \widetilde{x}_j(t)\right]^T$ and let $\mathscr{C}_{ij}$ denote the convex polygon formed by $\widetilde{\boldsymbol{x}}_{ij}(t)$. It is obvious that $l_i + l_j$ equals half of the perimeter of the enclosing rectangle of $\mathscr{C}_{ij}$ whose edges are parallel with the axes (for convenience we say that this rectangle is parallel with the axes). Note that the perimeter of any enclosing rectangle of the convex polygon will not change in rotations. Therefore, to minimize $l_i + l_j$, first we would locate the minimum-perimeter enclosing rectangle of $\mathscr{C}_{ij}$, and then we rotate it till it is parallel with the axes. See Fig. 2(a), the vertexes of $\mathscr{C}_{ij}$ are marked by '⊙'. Before a rotation, the minimum-perimeter enclosing rectangle of $\mathscr{C}_{ij}$ is A′B′C′D′, thus $l_i + l_j = \mathrm{AD} + \mathrm{DC}$. When the rectangle A′B′C′D′ is rotated to be parallel with the axes, as shown in Fig. 2(b), $l_i + l_j$ is minimized to be $\mathrm{A'D'} + \mathrm{D'C'}$. Since A′B′C′D′ is the minimum-perimeter enclosing rectangle of $\mathscr{C}_{ij}$, $\mathrm{A'D'} + \mathrm{D'C'}$ is actually a global minimum of the current sub-problem.

There are $n(n-1)/2$ distinct Givens rotations (parameterized by α), and optimizing a set of these is known as a sweep. The algorithm often needs only a few sweeps to converge. Now we will consider how to locate the minimum-perimeter enclosing rectangle. From Theorem 3.2.1 in [16], at least one edge of the minimum-perimeter enclosing rectangle coincides with one edge of the polygon, thus we

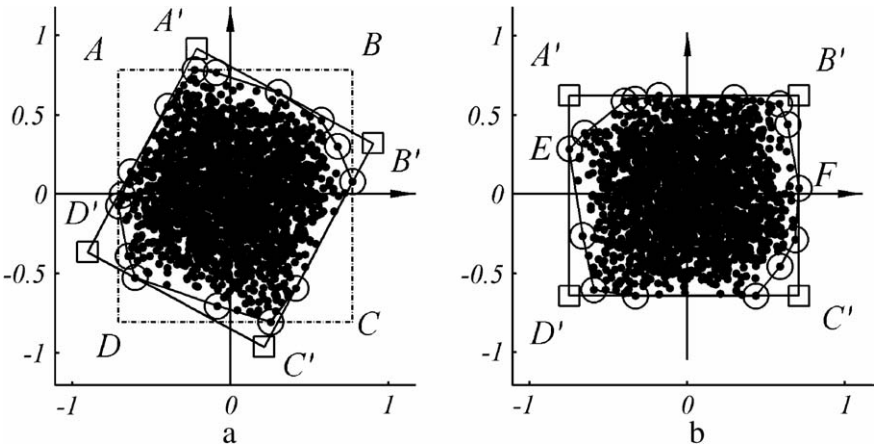

Fig. 2. Principle of the proposed algorithm.

may locate the minimum-perimeter enclosing rectangle by traversing the edges of $\mathscr{C}_{ij}$. The pseudo-code is displayed as follows:

For $k = 1{:}K$
  For $i = 1{:}m - 1$
    For $j = i{+}1{:}m$
      Generate $\mathscr{C}_{ij}$ from $\widetilde{\boldsymbol{x}}_{ij}(t)$, and locate the minimum-perimeter enclosing rectangle $\mathscr{R}$ of $\mathscr{C}_{ij}$ by traversing the edges of $\mathscr{C}_{ij}$
      $\widetilde{\boldsymbol{x}}^{(k+1)}(t) \leftarrow \boldsymbol{G}_{ij}^{\alpha}\widetilde{\boldsymbol{x}}^{(k)}(t)$ where α is the angle between $\mathscr{R}$ and the axis e_i
    End
  End
End

In the time complexity, the complexity of generating the convex hull $\mathscr{C}_{ij}$ is O(T log T) [17] and the complexity of locating the minimum-perimeter enclosing rectangle of $\mathscr{C}_{ij}$ is O(n) [16], where n is the number of vertexes with $n \ll T$. Thus the total time complexity is O(Km2T log T), where K is the number of iterations (sweeps). In our experiments, only 3–10 sweeps are needed, which is very efficient.

3. Discussion

3.1. Improvement of the robustness

In our algorithm the range of a signal is estimated from its maximum and minimum. To improve the robustness, after generating the convex polygon $\mathscr{C}_{ij}$, in practice, each vertex V of $\mathscr{C}_{ij}$ would be updated by

$$V^{\star} = \frac{1}{m}\sum_{k=1}^{m}\widetilde{\boldsymbol{x}}_k^{(V)} \tag{5}$$

where $V^{\star} = \left[\widetilde{x}_i^{V^{\star}}, \widetilde{x}_j^{V^{\star}}\right]^T$ and $\widetilde{\boldsymbol{x}}_k(V)$ is the kth nearest point of V in $\widetilde{\boldsymbol{x}}_{ij}$. Note that they need to be calculated only once because the neighborhood structure will be preserved in rotations. From Fig. 2(b), the minimum-range estimated by the proposed algorithm equals to

$$d^{\star} = \widetilde{x}_i^{F^{\star}} - \widetilde{x}_i^{E^{\star}} \tag{6}$$

where
$x_i^{F^{\star}} \approx \frac{1}{m}\sum_{k=1}^{m}\widetilde{x}_i(T-k+1)$, $x_i^{E^{\star}} \approx \frac{1}{m}\sum_{k=1}^{m}\widetilde{x}_i(k)$, $\widetilde{x}_i(k)$ is the kth smallest entry of $\widetilde{x}_i$. This method is actually equivalent to the averaged order statistics method [11]. The averaged order statistics uses the following estimator:

$$\langle R_m^{\star}\rangle = \frac{1}{m}\sum_{k=1}^{m}R_k^{\star}(\boldsymbol{x}) \tag{7}$$

where

$$R_k^{\star} \triangleq x_i(T-k+1) - x_i(k) \tag{8}$$

In fact, by substituting (8) into (7), we have

$$\langle R_m^{\star}\rangle = \frac{1}{m}\sum_{k=1}^{m}x_i(T-k+1) - \frac{1}{m}\sum_{k=1}^{m}x_i(k) \tag{9}$$

We see that the range estimated by $\langle R_m^{\star}\rangle$ equals the range estimated by (6) approximately. The selection of m has been discussed in [11] and we skip it.

3.2. Extension of the proposed approach

The proposed approach together with other geometric approaches often has poor performance when the finite sample sequence has few points near the scatter plot corners. Generally, they can separate the uniformly distributed sources or close to uniformly distributed sources, digital communication signals [14], and image signals. Some other types of available distributions were also discussed in [11]. These signals are often referred to as the sub-Gaussian.

One may want to know about the proposed approach to the other kinds of signals. A fact is, if the sample sequence has a few points near the scatter plot corners, the maximum-range criterion can be used to separate these kinds of signals, namely,

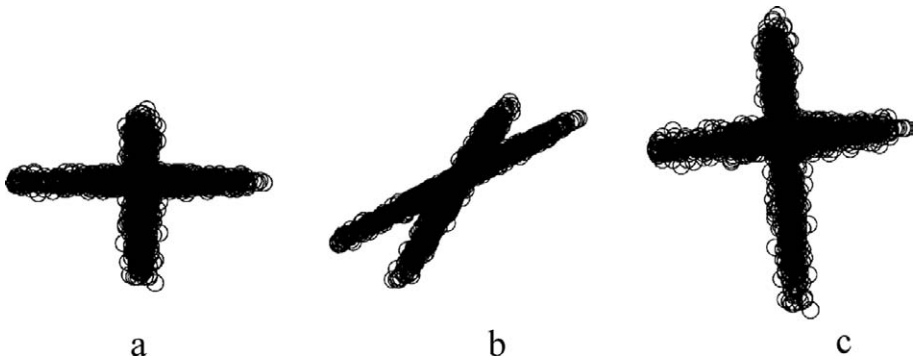

Fig. 3. Scatter plot of two speech signals, their mixtures and whitened mixtures.

$$\max L(\boldsymbol{U}) = \sum_{i=1}^{m} l_i(\boldsymbol{U}\widetilde{x}(t)) \quad s.t.\ \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I} \tag{10}$$

Fig. 3(a) is the scatter plot of two speech signals, and these two signals are sparse in a time domain. We see that the samples scatter in two orthogonal directions, and each direction is parallel with one axis. The scatter plots of their mixtures and the whitened mixtures are shown in Fig. 3(b) and (c), respectively. When a source is extracted successfully, its range is maximized. Recall that model (4) is based on the fact that $R(\boldsymbol{b}^T\boldsymbol{s}) = |\boldsymbol{b}|^T R(\boldsymbol{s})$ where $R(\boldsymbol{s}) = [R(s_1), R(s_2), \ldots, R(s_m)]^T$ are the ranges of independent components. However, this formula does not hold if the sources are sparse. For sparse signals, the following formula holds:

$$\max R(\boldsymbol{b}^T\boldsymbol{s}) = R(\boldsymbol{s}) \tag{11}$$

which can be interpreted as follows: since $\boldsymbol{s}$ are sparse, there is at most only one source being active or dominant at each time instant t, thus $x_i(t) = b_j s_j(t)$. If the sources are centered, $R(x_i) = R(\boldsymbol{b}^T\boldsymbol{s}) = R\{b_j s_j(t) | j = 1, 2, \ldots, m\}$, so (11) holds (note that $\boldsymbol{b}^T\boldsymbol{b} = 1$) and thus (10) yields, see Fig. 3(a) and (c). Also, the separation of these kinds of signals may be solved in the framework of classic sparse component analysis (SCA).

4. Simulations

The algorithm succeeds in many samples. Here the signals reported in [10,18] are adopted: $s1$ is the uniformly distributed noise, $s2 = 0.1\sin(400t)\cos(30t)$, $s3 = 0.01\text{sign}[\sin(500t + 9\cos(40t))]$, and the sample number is 1000. The sources $\boldsymbol{s}$, the mixtures $\boldsymbol{x}$ and the separated signals $\boldsymbol{y}$ are shown in Fig. 4.

We also compare our algorithm with ICAForNDC. For ICAForNDC, $\beta = 0.75$, the iteration number is 20, which are both recommended by the authors. The signal-to-noise ratio (SNR) is used as the performance index to evaluate the separation [1]:

$$\text{SNR}(s, y) = 10\log\frac{E[s^2]}{E[(y - s)^2]} \tag{12}$$

where s and y are normalized random variables with 0-mean and 1-variance. We run both algorithms 1000 times on an AMD Athlon 64 (1.8 GHz) processor with 448 MB

Table 1
Comparison between the proposed algorithm and ICA for NDC.

Algorithms	Time (s)		SNR (dB)	
ICAForNDC	0.48	35.65	39.87	47.78
Proposed algorithm	0.08	35.96	39.97	123.00

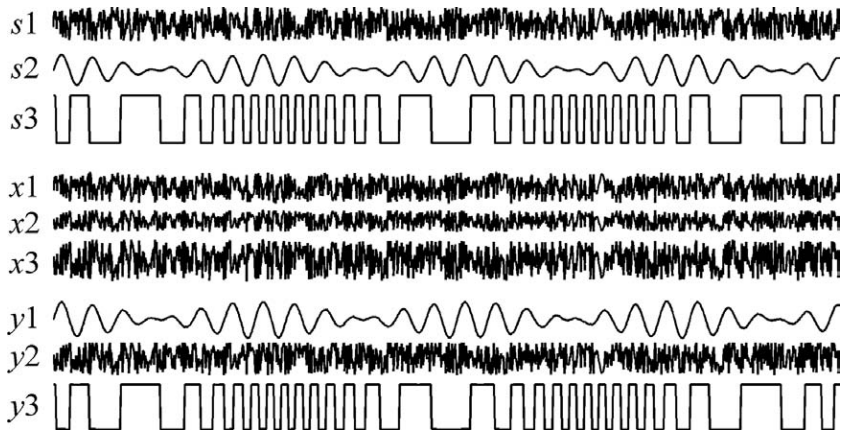

Fig. 4. The original, mixed and separated signals.

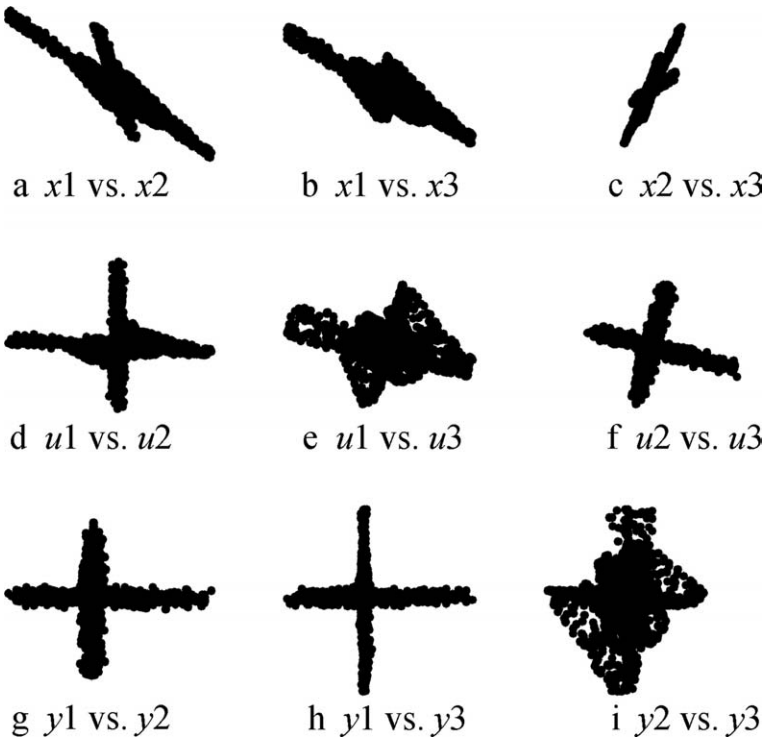

Fig. 5. (a–c) Scatter plot of the mixtures; (d–f) scatter plot of the whitened mixtures; (g–i) scatter plot of the separated signals.

of RAM. The sample number is 10,000 and the mixing matrix is regenerated randomly in each run. Their average performance is listed in Table 1.

From Table 1, the time consumption of the proposed algorithm is only 1/6 of that of ICAForNDC, while a better separation result is obtained.

We also performed experiments on a group of speech signals (they are sparse in time domain), and the maximum-range criterion is used. The scatter plot of their mixtures $\boldsymbol{x}$, whitened mixtures $\boldsymbol{u}$ and separated signals $\boldsymbol{y}$ are shown in Fig. 5(a)–(i), respectively. The corresponding SNRs are 35.28 dB, 53.45 dB, and 26.67 dB, respectively. However, it seems that this approach does not have obvious superiority to k-means-based approaches.

5. Conclusion

The minimum-range is a novel approach for ICA and BSS. This approach has reliable theoretical foundation. However, its optimization method, ICAForNDC, lacks efficiency and reliability.

The new approach is based on the minimum-range approach and it can be regarded as the batch mode of the original algorithm. The scatter plot of any two of the observations forms a convex polygon and its minimum-perimeter enclosing rectangle determines the rotation angle. Thus, our method is more efficient and reliable. To improve the robustness of the algorithm, its relation with averaged order statistics is also discussed.

The minimum-range approach is simple, but we must pay attention to the fact that it fails for some signals, for instance, sparse signals. Although for sparse signals we find that the maximum range is a solution, the applicability of the minimum/maximum-range approach needs more study, and this will be our future work.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. U0635001, 60674033 and 60774094).

References

[1] Cichocki A, Amari S. Adaptive blind signal and image processing: learning algorithms and applications. New York: John Wiley & Sons, Ltd.; 2002.
[2] Li YQ, Wang J, Zurada JM. Blind extraction of singularly mixed source signals. IEEE Trans Neural Netw 2000;11(6):1413–22.
[3] Common P. Independent component analysis, a new concept? Signal Process 1994;36(3):287–314.
[4] Stone JV. Blind deconvolution using temporal predictability. Neurocomputing 2002;49(1):79–86.
[5] Xie SL, He ZS, Fu YL. A note on Stone's conjecture of blind signal separation. Neural Comput 2005;17(2):321–30.
[6] Li YQ, Amari SI, Cichocki A, et al. Underdetermined blind source separation based on sparse representation. IEEE Trans Signal Process 2006;54(2):423–37.
[7] He ZS, Xie SL, Ding SX, et al. Convolutive blind source separation in the frequency domain based on sparse representation. IEEE Trans Audio Speech Lang Process 2007;15(5):1551–63.
[8] He ZS, Xie SL, Zhang LQ, et al. A note on Lewicki–Sejnowski gradient for learning overcomplete representations. Neural Comput 2008;20(3):636–43.
[9] Li YQ, Cichocki A, Amari S. Analysis of sparse representation and blind source separation. Neural Comput 2004;16(6):1193–234.

[10] Puntonet CG, Prieto A. Geometric approach for blind separation of signals. Electron Lett 1997;33(10):835–6.
[11] Vrins F, Lee JA, Verleysen M. A minimum-range approach to blind extraction of bounded sources. IEEE Trans Neural Netw 2007;18(3):809–22.
[12] Mansour A, Ohnishi N, Puntonet CG. Blind multiuser separation of instantaneous mixture algorithm based on geometrical concepts. Signal Process 2002;82(8):1155–75.
[13] Erdogan AT. A simple geometric blind source separation method for bounded magnitude sources. IEEE Trans Signal Process 2006;54(2):438–49.
[14] Erdogan AT. Globally convergent deflationary instantaneous blind source separation algorithm for digital communication signals. IEEE Trans Signal Process 2007;55(5):2182–92.
[15] Yamaguchi T, Hirokawa K, Itoh K. Independent component analysis by transforming a scatter diagram of mixtures of signals. Opt Commun 2000;173(1–6):107–14.
[16] Pirzadeh H. Computational geometry with the rotating calipers. Master thesis, McGill University, Montreal; 1999.
[17] Barber CB, Dobkin DP, Huhdanpaa HT. The quickhull algorithm for convex hulls. ACM Trans Math Software 1996;22(4): 469–83.
[18] Amari S, Cichocki A, Yang HH. A new learning algorithm for blind signal separation. Adv Neural Inf Process Syst 1996;8: 757–63.