A modified underdetermined blind source separation algorithm using competitive learning

Y. Luo and J.A. Chambers
Centre for Digital Signal Processing Research,
Division of Engineering, King's College London
Strand, London, WC2R 2LS. Email: Yuhui.Luo@kcl.ac.uk

Abstract

The problem of underdetermined blind source separation is addressed. An advanced classification method based upon competitive learning is proposed for automatically determining the number of active sources over the observation. Its introduction in underdetermined blind source separation successfully overcomes the drawback of an existing method, in which the goal of separating more sources than the number of available mixtures is achieved by exploiting the sparsity of the non-stationary sources in the time-frequency domain. Simulation studies are presented to support the proposed approach.

1 Introduction

The aim of blind source separation (BSS) is to separate a number, k, of unknown sources from m measurement signals. The word 'blind' in the statement refers to the lack of a priori information about the input sources and the channels between the sources and the measurement sensors. This blindness on one hand implies that the problem can be difficulty, but on the other hand makes BSS a powerful tool; as it could be applied in a wide range of applications where priori knowledge is difficult or impossible to achieve, for example where array calibration would be impractical.

Assuming that the input sources are statistically independent, many BSS algorithms resort to the approach of independent component analysis, e.g., the JADE algorithm, [4], and the natural gradient algorithm, [1]. By attempting to restore the mutual independence property of the sources, signal separation is achieved subject to a possible scaling and permutation ambiguity. Besides the mutual independence assumption, it is, moreover, a convention to assume that there exists at least as many sensors as sources, i.e., $k \leq m$. When this assumption is violated, the problem is termed underdetermined BSS in the literature. In [3], assuming the source distribution is sparse and the elements of the mixing matrix are uniformly distributed,

an algorithm which jointly estimates the mixing matrix and the sources is proposed. However, this sparsity assumption for the sources may not hold in many applications. In [5], an alternative approach is proposed by noticing that the received signals will be colinear with the corresponding steering vector (i.e., the column of the mixing matrix associated with that source) if only a single source is presented at a given time instant. Recently, an underdetermined BSS method is proposed by exploiting the information in the time-frequency (t-f) domain. Assuming that sources are orthogonal in the t-f domain, the t-f representations of different sources can be separated by clustering the t-f points which are associated with the same steering vectors, [6]. This algorithm requires, however, that the number of sources be known a priori and often leads to phantom sources being identified, because of the rough classification employed. In this paper, by applying an advanced classification method called self-splitting competitive learning (SSCL) algorithm, [7], not only is the performance of the time-frequency method greatly improved but also the newly proposed approach is able to separate different sources even when the number of active sources is not known. This advantage makes the proposed approach applicable in many practical applications.

2 Data Model

A multi-input and multi-output (MIMO) system is employed. The measurement signals are typically modelled as linear instantaneous mixtures for simplicity. Denote $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^*$ as the operations of transpose, conjugate transpose and complex conjugate respectively. For an array of m sensors, at discrete time instant t, the measurement signal $\mathbf{x}(t) = [x_1(t)\ x_2(t)\ \dots\ x_m(t)]^T$ is written as

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

where $\mathbf{s}(t) = [s_1(t)\ s_2(t)\ \dots\ s_k(t)]^T$ is the source vector contributed from k sources, A is an $m \times k$ mixing matrix and $\mathbf{n}(t)$ is the zero mean additive noise which maybe present in the measurement. As the mutual

independence property is lost after linear mixing, an unmixing matrix is introduced in independent component analysis to restore this property. Denote B as the unmixing matrix, which is of dimension k by m. The unmixing output $\mathbf{y}(t)$ is the estimator of the input sources subject to a scaling and permutational ambiguity.

$$\mathbf{y}(t) = B\mathbf{x}(t) \qquad (2)$$

In the approach of t-f analysis, as we will mention below, rather than directly looking for the unmixing matrix B, signal separation is achieved in two stages, i.e., the separation of t-f representations from different sources and the synthsis of the source signals from the corresponding separated t-f representations.

3 A t-f underdetermined BSS method and the problem of classification

Time-frequency analysis is a useful tool in processing non-stationary signals. With an appropriate t-f transform, a one-dimensional function of time is mapped into a two-dimensional function of time and frequency. By so doing, more information of the input sources can be revealed. In the method of [6], a quadratic t-f representation is employed. Denote $\phi(v, l)$ as the signal-independent kernel function. The discrete-time form of Cohen's class of t-f representation for signal $x_i(t)$, [2], which corresponds to the auto term in the quadratic t-f representation, is given by

$$D_{x_i x_i}(t, f) = \sum_{l=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} \phi_{ii}(v, l) \times \qquad (3)$$
$$x_i(t + v - l) x_i^*(t + v + l) e^{-j4\pi f l}$$

The cross t-f representation of two signals $x_i(t)$ and $x_j(t)$ is written as

$$D_{x_i x_j}(t, f) = \sum_{l=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} \phi_{ij}(v, l) \times \qquad (4)$$
$$x_i(t + v - l) x_j^*(t + v + l) e^{-j4\pi f l}$$

The extension of the t-f representation to accommodate vector signals leads to, [6] and [2],

$$D_{\mathbf{xx}}(t, f) = \sum_{l=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} \phi(v, l) \times \qquad (5)$$
$$\mathbf{x}(t + v - l) \mathbf{x}^H(t + v + l) e^{-j4\pi f l}$$

where $\phi(v, l)$ is a matrix whose $(i, j)^{th}$ entry $\phi_{ij}(v, l)$ is the kernel associated with the ith and the jth measurement sensor output. Notice that $D_{\mathbf{xx}}(t, f)$ is related to that of the input sources by the following relationship,

$$D_{\mathbf{xx}}(t, f) = A D_{\mathbf{ss}}(t, f) A^H \qquad (6)$$

where $D_{\mathbf{ss}}(t, f) = \sum_{l=-\infty}^{\infty} \sum_{v=-\infty}^{\infty} \phi(v, l) \mathbf{s}(t + v - l) \cdot \mathbf{s}^H(t + v + l) e^{-j4\pi f l}$ is the t-f matrix of the input sources. For underdetermined BSS, under the orthogonality assumption that the t-f distributions of different sources do not overlap, only one diagonal entry of $D_{\mathbf{xx}}(t, f)$ will be different from zero. Hence, if at the point (t_i, f_i), only source-j is present, the t-f representation $D_{\mathbf{xx}}(t_i, f_i)$ can be expressed as, [6],

$$D_{\mathbf{xx}}(t_i, f_i) = D_{s_j s_j}(t_i, f_i) \mathbf{a}_j \mathbf{a}_j^H \qquad (7)$$

where $\mathbf{a}_j$ is the jth column of the mixing matrix A, i.e., the steering vector of source-j. It can be observed that the scalar $D_{s_j s_j}(t, f)$ and the vector $\mathbf{a}_j$ are in fact the principal eigenvalue and eigenvector of the matrix $D_{\mathbf{xx}}(t_i, f_i)$. Moreover, if two t-f points are associated with the presence of the same source, in principle, the principle eigen vector of matrix $D_{\mathbf{xx}}(t_i, f_i)$ should remain the same. Classification according to the principal eigen vectors therefore provides a way to separate the signal spectra. The source signals can be synthesized with the estimator of $D_{s_j s_j}(t, f)$, for $j = 1 \ldots k$, [6].

A crucial problem in this approach is in the classification of the set of eigen vectors obtained at different t-f points. In [6], two normalized vectors $\widetilde{\mathbf{a}}_j$ and $\widetilde{\mathbf{a}}_{j'}$ are allocated into two different classes if the angle between them is larger than a certain threshold, i.e.,

$$\arccos\left(\widetilde{\mathbf{a}}_j^T \widetilde{\mathbf{a}}_{j'}\right) > \varepsilon \qquad (8)$$

where $\widetilde{\mathbf{a}}_j = \left[\mathrm{Re}(\mathbf{a}_j)^T \ \mathrm{Im}(\mathbf{a}_j)^T\right]^T$ and $\|\widetilde{\mathbf{a}}_j\| = 1$. However, this approach often gives more classes than the number of sources, as indicated by the authors, [6]. To avoid artificial sources, further thresholding is suggested and for its operation the knowledge of the number of active sources is necessary. As such information is likely to be unknown in many applications, to enable the algorithm to operate in a completely 'blind' fashion, better classification techniques are required. In the following section, the SSCL algorithm is introduced for this purpose.

4 Self-Splitting Competitive Learning in Underdetermined BSS

The objective of classification is to detect the hidden structure within some data set Z. If two members of the data set are judged to be similar (note that similarity can be measured in various ways), they will be assigned into one cluster, otherwise they will be allocated into different clusters. In our application of underdetermined BSS, the data set Z refers to the collection of eigenvectors $\widetilde{\mathbf{a}}_j$ and ideally it should be partitioned into k clusters, as there are k active sources present at the system input.

967   Proc. ISPA03

Assuming that the value of k is known, many clustering algorithms have been proposed, e.g., the well-known k-means algorithm. When the value of k is not known, however, only a few methods are available in the open literature. One of the widely used approaches determines the number of clusters by testing the error measure with respect to varying numbers of clusters. The idea is based on the fact that the error measure decreases monotonically as the number of clusters increases and the error reduction becomes much slower when a natural number of clusters has been reached. However, for a large data set, such an approach leads to heavy computational complexity due to the repeated computation of the error measure with respect to different numbers of clusters. In comparison, the SSCL algorithm partitions the data set in a completely different way. The algorithm is proposed from the view of competitive learning in neural networks, in which several prototypes $\mathbf{p}_i$ (A prototype corresponds to a neuron in neural networks.) are competing to 'win' as every member of Z is included in the examination. In the ideal situation, each cluster is associated with a prototype at its center. Therefore the problem of estimating the right number of clusters in classification is effectively the process of determination of the number of prototypes in competitive learing.

At the initialization stage of the SSCL algorithm, the whole data set is consider to be a single prototype, which will be split into further prototypes when a certain condition is met. The danger of this approach is recognized by the author in [7]. That is, one prototype might represent more than one cluster if the number of prototypes is less than the natural number of clusters, and as a result, none of the clusters could be correctly identified. To avoid this problem, an asymptotic property vector $\mathbf{f}_i$ is suggested to guide the learning of each prototype $\mathbf{p}_i$. By discriminating the data in Z, each prototype is able to represent only one cluster. Specifically, only the data points inside the neighborhood of $\mathbf{p}_i$ will contribute to the learning of $\mathbf{p}_i$ and the neighborhood is determined with the help of the asymptotic property vector $\mathbf{f}_i$. To ensure that enough members in Z to be included in the learning of $\mathbf{p}_i$, the asymptotic property vector $\mathbf{f}_i$ should be initialized far away from $\mathbf{p}_i$. In terms of the learning of $\mathbf{f}_i$ and $\mathbf{p}_i$, suppose at some moment during the learning of $\mathbf{p}_i$, the member selected from Z is $\widetilde{\mathbf{a}}$. The update of the asymptotic property vector $\mathbf{f}_i$ is written as

$$\mathbf{f}_i^{new} = \mathbf{f}_i + \frac{1}{n_{\mathbf{f}_i}} \delta_i \left(\widetilde{\mathbf{a}} - \mathbf{f}_i \right) \Psi(\mathbf{p}_i, \mathbf{f}_i, \widetilde{\mathbf{a}}) \qquad (9)$$

where $\delta_i = \left(\frac{\|\mathbf{p}_i - \mathbf{f}_i\|_2}{\|\mathbf{p}_i - \widetilde{\mathbf{a}}\|_2 + \|\mathbf{p}_i - \mathbf{f}_i\|_2} \right)^2$, the notation $\|\cdot\|_2$ stands for the Euclidean norm and the quantity $n_{\mathbf{f}_i}$ is the winning counter of $\mathbf{f}_i$ that is updated by $n_{\mathbf{f}_i}^{new} = n_{\mathbf{f}_i} + \delta_i \Psi(\mathbf{p}_i, \mathbf{f}_i, \widetilde{\mathbf{a}})$. The function $\Psi(\mathbf{p}_i, \mathbf{f}_i, \widetilde{\mathbf{a}})$ is given

by

$$\Psi(\mathbf{p}_i, \mathbf{f}_i, \widetilde{\mathbf{a}}) = \begin{cases} 1 & \text{if } \|\mathbf{p}_i - \mathbf{f}_i\|_2 \geq \|\mathbf{p}_i - \widetilde{\mathbf{a}}\|_2 \\ 0 & \text{otherwise} \end{cases} \qquad (10)$$

It can be observed that the above learning scheme guides the asymptotic property vector $\mathbf{f}_i$ to shift toward the $\widetilde{\mathbf{a}}$, if $\widetilde{\mathbf{a}}$ is in the neighbourhood of its associated prototype $\mathbf{p}_i$ and gives up those data outside this area. As more and more members are included in the learning, the asymptotic property vector $\mathbf{f}_i$ approaches $\mathbf{p}_i$, [7]. The update equation of $\mathbf{p}_i$ is given by

$$\mathbf{p}_i^{new} = \mathbf{p}_i + \alpha_i \left(\widetilde{\mathbf{a}} - \mathbf{p}_i \right) \qquad (11)$$

where α_i is some scalar whose value is related to $\mathbf{p}_i$, $\mathbf{f}_i$ and $\widetilde{\mathbf{a}}$, i.e.,

$$\alpha_i = \left(1 + \frac{\|\mathbf{p}_i - \widetilde{\mathbf{a}}\|_2}{\|\mathbf{p}_i - \mathbf{f}_i\|_2} \right)^{-2} \qquad (12)$$

According to eqn. (12) and (11), if the member $\widetilde{\mathbf{a}}$ is far away from the neighborhood of $\mathbf{p}_i$, the value α_i approaches to zero, which implies that $\widetilde{\mathbf{a}}$ will have little influence on the learning of $\mathbf{p}_i$. Upon convergence of $\mathbf{p}_i$, the distance between $\mathbf{p}_i$ and $\mathbf{f}_i$ is smaller than a certain threshold value, which implies no more members within the data set are eligible for the learning of $\mathbf{p}_i$.

To determine when the SSCL algorithm should split one of its prototypes, a center property vector $\mathbf{g}_i$ is employed for each cluster, which are updated with the k-means algorithm, as the exact arithmetic mean of the input data points for which a prototype $\mathbf{p}_i$ has so far been the winner. If $\|\mathbf{p}_i - \mathbf{g}_i\|_2$ is larger than a certain threshold, it suggests that an extra cluster exists which is trying to pull the cluster center from $\mathbf{p}_i$ to $\mathbf{g}_i$. For non-Gaussian distributed clusters, if $\widetilde{\mathbf{a}}_j$ is of dimension m, a simple way to determine the threshold is $\frac{1}{50} (\max(Sl_1, Sl_2 \ldots Sl_m))$, where Sl_i is the scale of the ith coordinate in the m dimension feature space, [7]. Once the algorithm decides there should be another new cluster, the current prototype is split into two. One stays at its current location, the other is initialized at some distant location. For more efficiency in the implementation, a distant property vector $\mathbf{r}_i$ is used and updated as the learning of $\mathbf{p}_i$ continues. But, in contrast to the asymptotic property vector $\mathbf{f}_i$, the distant property vector $\mathbf{r}_i$ will be updated to a distant location from $\mathbf{p}_i$. The algorithm continues until no further cluster is suggested by the splitting criterion.

The advantage of the SSCL algorithm is in its capability of automatically determining the number of clusters in the data set and moreover its computational simplicity when dealing with a large data set. Simulation results of the introduction of the SSCL algorithm in underdetermined BSS are shown below.
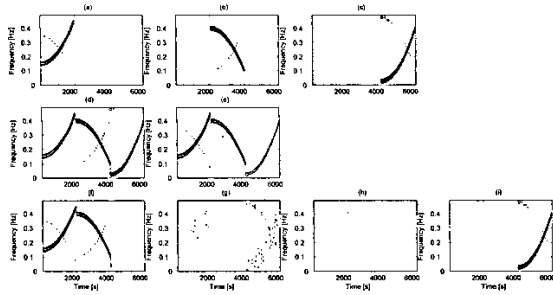
Figure 1. Separating three polynomial frequency modulated sources from two mixtures (a) Source-1 (b) Source-2 (c) Source-3 (d) Sensor-1 (e) Sensor-2 (f) Unmixing output-1 (fails) (g) Unmixing output-2 (fails) (h) Unmixing output-3 (fails) (i) Unmixing output-4 (successful)

5 Simulation

We assume a $k = 3$ sources and $m = 2$ sensors system. The source signals are assumed to be the frequency modulated signals as shown in Fig 1 (a)-(c). The kernel is selection as the Choi-Williams distribution in quadratic t-f representation. The mixing matrix is assumed to be $A =$
$$\begin{bmatrix} 0.4 + 0.7i & 0.9 + 0.2i & 0.6 + 0.5i \\ 0.6 + 0.6i & 0.7 + 0.2i & 0.4 + 0.3i \end{bmatrix}.$$ In the presence
of additive noise with signal to noise ratio of 20dB, the received signals are shown in 1 (d)-(e). With the rough classification method as [6], when the the threshold value ε in eqn (8) is set to 0.3, a number 4 of clusters have been found. The unmixing outputs in Fig 1 (f)-(i) show that although one phantom sources has been generated, source separation is not successful. Source separation is only successful when the threshold value is reduced to 0.1. However, 14 phantom sources have been generated, which is a rather large number compared with the number of actual sources. With the SSCL algorithm, we assume no knowledge of the number of sources. The algorithm finally splits into three prototypes (clusters), which is a desirable result. The t-f representations of different sources are successfully separated, as shown in Fig 2 (a)-(c). Simulations have also been carried out for linear chip sources and GMSK sources, but due to the limited space they are omitted here. But much improved performance of the SSCL algorithm is confirmed.

6 Conclusion

We are interested in the problem of underdetermined BSS by using t-f analysis. A major problem of the current t-f method in [6] is in its identification
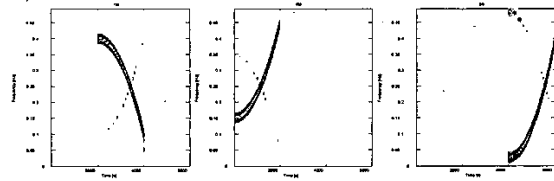

Figure 2. Successful source separation of three polynomial frequency modulated sources from two mixtures (a) Unmixing output-1 (b) Unmixing output-2 (c) Unmixing output-3

of the phantom sources and further treatment to the phantom sources requires knowledge of the number of active sources, which might not be available in pratice. By the introduction of the SSCL algorithm, not only has the problem of phantom sources been overcome but also the algorithm is able to operate in a complete blind sense. The requirement of knowing the number of active sources has been eliminated. Simulations results confirm these improvements.

References

[1] S.-I. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10:251–276, 1998.

[2] A. Belouchrani and M. G. Amin. Blind source separation based on time-frequency signal representation. *IEEE Trans. on Signal Processing*, 46(11):2888–2897, 1998.

[3] P. Bofill and M. Zibulevsky. Underdetermined blind source separation using sparse representation. *Signal Processing*, 81:2253–2362, 2001.

[4] J.R. Cardoso and A. Souloumiac. Blind beamforming for non-Gaussian signals. *IEE Proc. F on Radar and Signal Processing*, 140(6):362–370, 1993.

[5] L. Kelva, D. Erdokmus, C. Pantaleon, I. Santamaria, J. Pereda, and J. C. Principe. Underdetermined blind source separation in a time-varying environment. *Proc of IEEE Int. Conf. Acoust. Speech Signal Processing*, 3:3049–3052, 2002.

[6] L.-T. Nguyen, A. Belouchrain, K. Abed-Meraim, and B. Boashash. Separating more sources than sensors using time-frequency distribution. *Sixth International Symposium on Signal Processing and its Applications*, 2:583–586, 2001.

[7] Y.-J. Zhang and Z.-Q. Liu. Self-splitting competitive learning: A new on-line clustering paradigm. *IEEE Trans. on Neural Networks*, 13(2):369–380, 2002.