Improved Techniques for Blind Source Separation

Yongjian Zhao

Information Engineering Institute, Shandong University (Weihai), China

jian123cn@sdu.edu.cn

Keywords: Separation, Factor, Matrix, Distribution, Recovery, Independence.

Abstract. In many practical applications such as biomedical signal processing, it is often desirable to extract one or a few source signals instead of all signals. The classical FASTICA algorithm can extract a source signal which has the maximum negentropy of all signals. However, the extracted signal is not necessarily the desired one. To address these problems above, a constraint is introduced to a negentropy based cost function. As a result, a constrained method is proposed which can extract a desired signal from its mixture exclusively.

Introduction

Blind source separation (BSS) is a statistical and computational technique for revealing hidden factors that underlie sets of measurements [1]. The BSS problem arises when linear, instantaneous mixtures or observations, generated as a set of signals are mixed by traversing an unknown medium, need to be processed to estimate or recover a number of the original sources. In last decades, the BSS technique has shown its efficiency for data analysis originate from digital images, document databases and economic indicators [2,3].

Until now, a few source extraction algorithms have been developed for extraction of a specific signal by using some priori information, such as non-Gaussianity, smoothness or linear predictability, coding complexity, etc [5-7]. For example, Barros and Cichocki formally defined the prerequisites to extract a desired signal with a linear filter using only information about its autocorrelation function [7]. This algorithm only needs to estimate the period of the desired source signal in advance. It is simple and runs quickly, but it often cannot recover a clean desired signal. In other words, the extracted desired source signal is often mixed with other undesired sources. What is more, it is sensitive to the estimation error of time delay.

The famous FASTICA algorithm [1,5], proposed by Hyvärinen, attempts to separate underlying components from a given set of mixed measurement channels on the basis of their non-Gaussianity. After a deflated stage, the extracted signal can be removed from its mixture. The new deflated mixture of the remaining sources then undergoes the next extraction process to recover the second source. Repeating this iterative process, source signals can be extracted from a set of mixed measurements one by one. In many practical applications such as biomedical signal processing, it is often desirable to extract one or a few source signals instead of all signals. In such applications, the one-unit FASTICA algorithm is computationally expensive.

To address these problems above, a constraint about the desired signal is introduced to guide the signal separation process. Then the ill-posed BSS is converted to a better-posed problem, facilitating more applications. Solving this BSS problem, an improved BSS algorithm is presented for extraction of one or a few desired source signals exclusively. Extensive computer simulations demonstrate its validity and usefulness.

Improved BSS Techniques

Generally speaking, the basic noise-free mixture model [1,2] for BSS problem can be described as follows:

$$x = As, \tag{1}$$

where $s = (s_1, \cdots, s_m)^T$ denotes the original independent sources, $x = (x_1, \cdots, x_n)^T$ denotes mixtures of original sources and A is an unknown mixing matrix.

The goal of source extraction is to find a vector w so that
$$y = w^T x, \tag{2}$$
is an estimated source signal up to a scalar.

To cope with ill-conditioned case and to make algorithms simpler and faster [5,6], whitening is often used to transform the observed signals x to $\tilde{x} = Vx$ so that $E[\tilde{x}\tilde{x}^T] = I$, where V is a whitening matrix. For convenience, in the following we assume that x are the whitened observed signals and $n = m$.

To formulate the BSS problem, one must introduce a reliable cost function, which is greatly dependent on the parameters of the specified neural network model in (1). Maximization/Minimization of such a cost function should cause the output satisfy the statistical conditions of stochastic independence and other inherent properties. Generally speaking, non-Gaussianity is a simple and intuitive principle for solving BSS problem. Kurtosis, the fourth-order cumulant, is the first practical measure of non-Gaussianity [1-3]. However, such principle gives estimators that are very sensitive to outliers and have large mean square errors. As an alternative measure of non-Gaussian, Negentropy has a desirable property that it is invariant for invertible linear transformations. As far as the statistical performance is concerned, negentropy has become the optimal estimator of non-Gaussianity

A classical negentropy based cost function introduced by Hyvärinen is constructed as the cost function of the FASTICA algorithm [1,5] :
$$J(y) = \rho[E\{G(y)\} - E\{G(v)\}]^2, \tag{3}$$
where v is a Gaussian variable with zero mean and unity variance, ρ is a positive constant, and $G(*)$ can be any non-quadratic function, such as
$$G_1(y) = \log coash(a_1 y)/a_1, \tag{4}$$
$$G_2(y) = \exp(-\frac{1}{2}a_2 y^2)/a_2, \tag{5}$$
$$G_3(y) = y^4/4, \tag{6}$$
where $1 \le a_1 \le 2$ and $a_2 \approx 1$. G_1 is a good general purpose function; G_2 may be better when the independent components are highly super-Gaussian, or when robustness is very important; G_3 is justified on statistical grounds only for estimating sub-Gaussian signals when there are no outliers.

The gradient of $J(y)$ with respect to w can be deduced as
$$\nabla_w J(y) \propto 2\gamma E\{xg(w^T x)\} \tag{7}$$
where $\gamma = E\{G(w^T x)\} - E\{G(v)\}$. Thus one can deduce the famous FastICA algorithm [2,4]:
$$w \leftarrow 2\gamma E\{xg(w^T x)\}. \tag{8}$$

As a classical one-unit BSS algorithm, FastICA has many desirable properties [4-6]. For each process, it converges to an independent component (source signal) which has the maximum negentropy of all underlying components. Through a deflated stage, the source signals can be estimated from their mixtures one by one. In many practical applications such as biomedical signal processing, the number of source signals is large while only one or a few is desired. FastICA has to estimate source signals one by one unless the desired signal has the maximum negentropy of all underlying components. That is to say, the famous one-unit algorithm is time-consuming and not convenient to obtain the desired source by the deflation scheme.

In most cases of biomedical signal processing, one may know specific prior information about a desired source signal in advance [5,6]. If such prior information carries enough knowledge to distinguish the desired signal from others, it is called reference [3,4]. In this paper, we introduce such reference to guide the one-unit algorithm (8), thus ensuring the output is necessarily the desired one.

The closeness measure between an estimated output y and its corresponding reference r is denoted by $\xi(y,r)$. Further suppose that w_m or w^* is the demixing vector for the desired signal and w_i $(i=1,\cdots,m-1)$ are the demixing vectors for undesired components. Note that the desired source signal is the only one closest to the reference signal r, satisfying the following inequality relationship:

$$\varepsilon(w^{*T}\tilde{x}, r) \le \varepsilon(w_1^T \tilde{x}, r) \le \cdots \le \varepsilon(w_{m-1}^T \tilde{x}, r). \tag{9}$$

In fact, the extracted signal $w^{*T}\tilde{x}$ is equal to y^* if and only if the following inequality relationship is satisfied:

$$g(w) = \varepsilon(w^T\tilde{x}, r) - \xi \le 0, \tag{10}$$

where $\xi \in [\varepsilon(w^{*T}\tilde{x}, r), \varepsilon(w_1^T \tilde{x}, r))$ is a threshold parameter.

Through introducing (10) into the negentropy based cost function (3) as a constrained term, one can deduce the following constrained optimization problem:

$$\begin{cases} \max \quad J(y) = \rho[E\{G(y)\} - E\{G(v)\}]^2 \\ s.t. \quad g(w) \le 0, \mathrm{h}(w) = \mathrm{E}(y^2) - 1 = 0 \end{cases}, \tag{11}$$

where the equality constraint aims at bounding $J(y)$ and w.

The inequality constraints in (10) can be converted into equality ones by introducing a vector of slack variables, so the optimal solution can be obtained by the Lagrange multipliers method [1,6]. By seeking for the maximum of the augmented Lagrange function, one can deduce the following learning rule for w:

$$w_{t+1} = w_t - \eta R_z^{-1}\Psi_1 / \psi_2, \tag{12}$$

$$\psi_1 = \overline{\rho}E\{zG_y^{'}(y)\} - 0.5\mu E\{zg_y^{'}(y)\} - \lambda E\{zy\}, \tag{13}$$

$$\psi_2 = \overline{\rho}E\{zG_{y^2}^{''}(y)\} - 0.5\mu E\{zg_{y^2}^{''}(y)\} - \lambda, \tag{14}$$

where t is the iteration count, $R_z = E\{zz^T\}$, $\overline{\rho} = \rho \cdot sign(EG(y) - EG(v))$, $G_y^{'}(y)$ and $g_y^{'}(y)$ are the first derivatives of $G(y)$ and $g(y)$ with respect to y, and $G_{y^2}^{''}(y)$ and $g_{y^2}^{''}(y)$ are the corresponding second derivatives. The optimum multipliers μ^* and λ^* are found by iteratively updating them based on the following gradient-ascent method:

$$\mu_t = \max\{0, \mu_{t-1} + \gamma g(w_{t-1})\}, \tag{15}$$

$$\lambda_t = \lambda_{t-1} + \gamma h(w_{t-1}). \tag{16}$$

Simulation Results

We tested the proposed BSS algorithm with a synthetic data set of four known source signals, which was also used for testing the previous BSS methods in [5-7]. To make comparison, we also ran the ICA-R algorithm [5,6] and the BCBSE algorithm [7], respectively. We tried our best to adjust the parameters of these algorithms in advance to ensure the best performance. To quantitatively compare their extraction quality, we selected the following performance index (*PI*):

$$PI = \sum_{j=1}^{M} \frac{|p_j|}{\max_k |p_k|} - 1, \quad k = 1,\cdots M \tag{17}$$

where p_j denotes the j elements of the global vector $p = w^T B$. In practice, *PI* is a very small number. *PI* is zero if and only if the desired signal is perfectly extracted.

The performance was estimated as the average *PI* values of 100 independent trials. The averaged performance was depicted in Fig.1. The convergence results were also depicted in Fig.5 for the four BSE algorithms. Here GRBSE1 is the proposed algorithm where $\varepsilon(y,r) = -E\{(y-r)^2\}$ while GRBS2 is the proposed algorithm where $\varepsilon(y,r) = (E\{yr\})^2$. Obviously, the BCBSE algorithm could

not extract the desired signal efficiently since it failed to converge. It was worthy of mention that the two new algorithms only needed about 30 iterations to achieve convergence and had the best extraction performance with small *PI*.
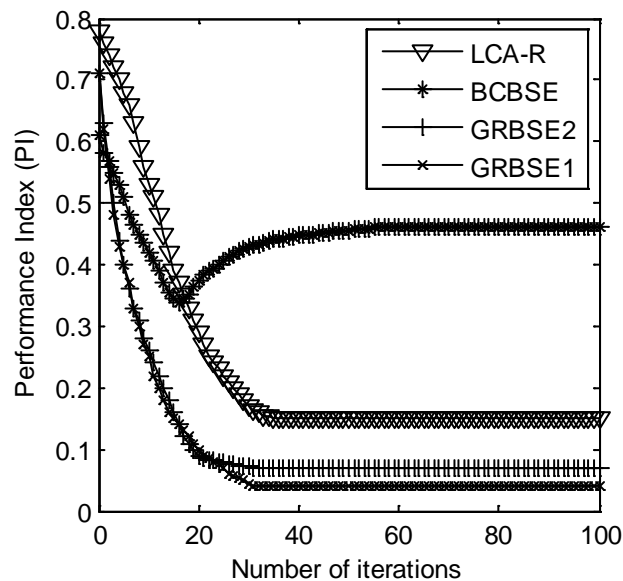

Fg.1 Comparisons of average *PI*s over 100 independent trials for four algorithms

Conclusion

As an increasingly popular data analysis tool, BSS technique has received wide attention in various fields such as biomedical signal processing, wireless telecommunication system, data mining, etc. Through incorporating a priori information as the reference into the negentropy cost function for FastICA, the model of constrained optimization problem is constructed. In such BSS problem, inequality constraint term is a closeness measure between the desired independent component and its corresponding reference. Solving this constrained optimization problem with the Lagrange multipliers method, an improved method is introduced, which can estimate a specific source signal from its mixture.

The proposed BSS method may have several advantages over simultaneous BSS. For example, only interesting source signals can be extracted, thus achieving good flexibility and relatively low computation load compared with the earlier methods; signals can be extracted in a specific order according to some features of source signals.

References

[1] Hyvärinen A, Karhunen J, Oja E. Independent Component Analysis. New York: Wiley, 2001

[2] Leong W Y, Mandic D P. Noisy component extraction (NoiCE). IEEE Transactions on Circuits and Systems, 2010, 57 (3) :664-671

[3] R.J. Ong, J.T. Dawley and P.G. Clem: submitted to Journal of Materials Research (2003)

[4] P.G. Clem, M. Rodriguez, J.A. Voigt and C.S. Ashley, U.S. Patent 6,231,666. (2001)

[5] Lu W, Rajapakse J C. ICA with reference. Neurocomputing, 2006, 69 (16-18) :2244-2257

[6] Lu W, Rajapakse J C. Approach and applications of constrained ICA. IEEE Transactions on Neural Networks, 2005, 16 (1) :203-212

[7] Barros A K, Cichocki A. Extraction of specific signals with temporal structure. Neural Computation, 2001, 13 (9) :1995-2003