A SEMI-BLIND EM ALGORITHM FOR OVERCOMPLETE ICA

Qiuhua Lin[1], *Ning Xu*[1], *and Hualou Liang*[2]

[1] School of Electronic and Information Engineering,
Dalian University of Technology, Dalian 116023, China
[2] School of Health Information Sciences, University of Texas at Houston, Houston, TX 77030, USA

ABSTRACT

Overcomplete independent component analysis (ICA) is a challenge of ICA to estimate more sources from less mixtures. The statistical properties of the sources such as sparsity are often assumed to solve the problem. Other available information about the sources such as waveform, however, is scarcely used. Motivated by the fact that semi-blind ICA in complete case can improve the potential of ICA by incorporating source information, this paper proposes a semi-blind algorithm for overcomplete ICA by explicitly utilizing waveform information about some sources. An approximate expectation-maximization (EM) algorithm is explored to provide normal cost function of the semi-blind algorithm while the prior information is utilized to form an extended one. Computer simulations results demonstrate that the proposed algorithm has much improved performance in *SNR*, convergence speed, and elimination of order ambiguity compared to the original EM algorithm.

Index Terms—Independent component analysis, semi-blind ICA, overcomplete ICA, EM algorithm

1. INTRODUCTION

Independent component analysis (ICA) consists of recovering M maximally independent sources from their N observed mixtures without knowledge of the source signals and the mixing parameters (usually $M \le N$). One challenge of ICA is to recover more sources from less mixtures (i.e. $M > N$) since the estimates of sources are not unique even if the mixing matrix is known. This is called overcomplete ICA problem. Due to practical applications, overcomplete ICA has been gaining more and more attention, some algorithms have been proposed under certain assumptions [1]-[7]. For example, most of the algorithms assumed that the sources were sparse [2]-[5]. In addition, several algorithms utilized other statistical properties of the sources such as nonstationarity [6] and fourth-order cumulant [7].

In practice, other prior information about some sources is often available and has been utilized by complete ICA ($M = N$) as extra constraints. This actually leads to what is called semi-blind ICA [8]-[12]. For example, the temporal information about some sources was utilized to extract sources of interest in a constrained ICA algorithm [8]. The paradigm information was incorporated into the ICA analysis of event-related functional magnetic resonance imaging data [9]. The spatial topography of selected source sensor projection was used as spatial constraints for the fastICA algorithm [10], and geometric information exploited in beamforming was used to constrain the separation of convolutive speech [11]-[12]. As expected, semi-blind ICA demonstrated considerable promise in further improving ICA performance [8]-[12].

The semi-blind ICA, by far, has been developed mostly for complete ICA but scarcely for overcomplete ICA. As such, an approximate expectation-maximization (EM) algorithm, which is an efficient overcomplete ICA algorithm proposed by Zhong et al. in [3], was explored to incorporate waveform information about some sources. Comparison to the original EM algorithm was performed through simulations to demonstrate performance of the proposed semi-blind algorithm.

This paper is organized as follows. Section 2 briefly introduces the original EM algorithm. Section 3 presents our proposed semi-blind EM algorithm in detail. Section 4 consists of simulations and results which compare our algorithm with the original EM algorithm. In Section 5 we provide conclusion.

2. EM ALGORITHM

The approximate EM algorithm in [3] used the following noisy mixing model:

$$\mathbf{x} = \mathbf{A}\mathbf{s} + \boldsymbol{\varepsilon} \qquad (1)$$

where $\mathbf{s} = [s_1, s_2, ..., s_M]^T$ includes M independent sources, $\mathbf{x} = [x_1, x_2, ..., x_N]^T$ denotes N observed mixtures ($M > N$), $\mathbf{A}$ is an $N \times M$ mixing matrix, and $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N]^T$ includes N noise signals with zero mean and known covariance matrix $\mathbf{\Sigma}$. The sources are assumed sparse, the distribution of which can thus be denoted by the following Laplacian distribution:

$$p(\mathbf{s}) = \left(\sqrt{2}\right)^{-M} \prod_{i=1}^{M} \exp\left(-\sqrt{2}\left|s_i\right|\right) \tag{2}$$

By maximizing the log-likelihood below,

$$\begin{aligned} L(\mathbf{s}) &= \log\left\{p(\mathbf{s}|\mathbf{x},\mathbf{A})\right\} \\ &= \sum_{t=1}^{T}\left\{-\frac{1}{2}(\mathbf{x}-\mathbf{A}\mathbf{s})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{A}\mathbf{s})+\varphi(\mathbf{s})\right\}+C \end{aligned} \tag{3}$$

where $\varphi(\mathbf{s}) = \log\left\{p(\mathbf{s})\right\}$, C is a constant irrespective to $\mathbf{s}$, T is the data length, the learning rule for estimating the sources is obtained as follows [3]:

$$\hat{\mathbf{s}}^k = \hat{\mathbf{s}}^{k-1} + \eta\nabla_{\hat{\mathbf{s}}}L(\hat{\mathbf{s}}^{k-1}) \tag{4}$$

where $\hat{\mathbf{s}}$ is the estimate of $\mathbf{s}$, k is the number of iteration, η is the learning rate, $\nabla_{\hat{\mathbf{s}}}$ denotes the gradient with respect to $\hat{\mathbf{s}}$. Next, the mixing matrix $\mathbf{A}$ is learned by:

$$\mathbf{A}^{k+1} = \left\{\sum_{t=1}^{T}\mathbf{x}\hat{\mathbf{s}}^T\right\}\left\{\sum_{t=1}^{T}\left(H(\hat{\mathbf{s}})^{-1}+\hat{\mathbf{s}}\hat{\mathbf{s}}^T\right)\right\}^{-1} \tag{5}$$

where $H(\hat{\mathbf{s}}) = -\nabla_s\nabla_s L(\hat{\mathbf{s}})$ is the Hessian of $L(\hat{\mathbf{s}})$.

3. PROPOSED ALGORITHM

3.1. Block diagram of the algorithm

Fig. 1 shows the block diagram of the proposed algorithm, in which $x_1, x_2, ..., x_N$ denote N mixed signals, $\hat{s}_1, \ldots, \hat{s}_M$ are estimates of M sources $s_1, s_2, ..., s_M$. Assume that prior information about L ($1 \le L < M$) sources $s_1, \ldots, s_L$ is available, $r_1, \ldots, r_L$ are L reference signals for $s_1, \ldots, s_L$ constructed from the prior information, whereas $r_{L+1}, \ldots, r_M$ are randomly generated references for $s_{L+1}, \ldots, s_M$ without prior information. $g_i(\hat{s}_i, \mathbf{r})$, $i = 1, ..., M$, $\mathbf{r} = \left[r_1, \ldots, r_M\right]^T$ is a closeness measure between an estimate s_i and each reference in $\mathbf{r}$, and its maximal element is reached when a reference corresponds to its source. By constraining learning of $\mathbf{s}$ and $\mathbf{A}$ with $g_i(\hat{s}_i, \mathbf{r})$, the L sources $s_1, \ldots, s_L$ with prior information are recovered in the same order as $r_1, \ldots, r_L$, whereas the other sources $s_{L+1}, \ldots, s_M$ without prior information will be recovered in random order.

3.2. Basic algorithm

To incorporate prior information into the original EM algorithm, a new cost function is formulated for the proposed algorithm as follows:

$$J_{\hat{\mathbf{s}}} = L(\hat{\mathbf{s}}) + G(\hat{\mathbf{s}}) \tag{6}$$

where $L(\hat{\mathbf{s}})$ is the original cost function of the EM algorithm, $G(\hat{\mathbf{s}})$ is an extended part for incorporating prior information, and is defined from the closeness measure $g_i(\hat{s}_i, \mathbf{r})$.

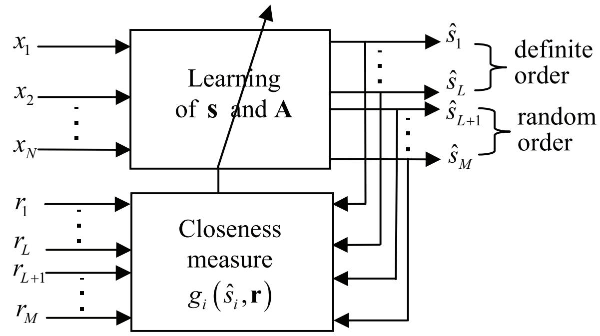

Fig. 1. Block diagram of the proposed algorithm

Indeed, $g_i(\hat{s}_i, \mathbf{r})$ can be defined in different ways. Here we use the correlation principle to utilize the waveform information about the sources:

$$g_i\left(\hat{s}_i, \mathbf{r}\right) = E(\hat{s}_i\mathbf{r}) = \left[E(\hat{s}_i r_1), ..., E(\hat{s}_i r_M)\right]^T \tag{7}$$

Obviously, $g_i(\hat{s}_i, \mathbf{r})$ has a maximal element when a reference corresponds to a source signal. Based upon this definition, we define a reasonable $G(\hat{\mathbf{s}})$ for estimating all of the sources as

$$G(\hat{\mathbf{s}}) = \sum_{i=1}^{M}\rho_i g_i\left(\hat{s}_i, \mathbf{r}\right)^T g_i\left(\hat{s}_i, \mathbf{r}\right) \tag{8}$$

where ρ_i is a positive correction factor. To fully utilize prior information, we learn ρ_i for each of sources differently:

$$\rho_i = \begin{cases} \lambda \max g_i\left(\hat{s}_i, \mathbf{r}\right), & \max g_i\left(\hat{s}_i, \mathbf{r}\right) \ge \xi. \\ \lambda \min g_i\left(\hat{s}_i, \mathbf{r}\right), & \max g_i\left(\hat{s}_i, \mathbf{r}\right) < \xi. \end{cases} \tag{9}$$

where λ is a positive constant, ξ is a threshold, these two parameters can be selected within (0, 1). As a result, the contribution of $G(\hat{\mathbf{s}})$ in (6) is enhanced by a large ρ_i when learning of a specific source s_i goes in a right direction.

By optimizing the new cost function with the steepest descent method, we obtain a new learning rule for $\mathbf{s}$ as follows:

$$\hat{\mathbf{s}}^{k+1} = \hat{\mathbf{s}}^k + \eta\left[\mathbf{A}^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mathbf{A}\hat{\mathbf{s}}^k) + \nabla_{\hat{s}}\varphi(\hat{\mathbf{s}}^k)\right] + \nabla_{\hat{\mathbf{s}}}G(\hat{\mathbf{s}}) \tag{10}$$

where

$$\begin{aligned} \nabla_{\hat{\mathbf{s}}}G(\hat{\mathbf{s}}) &= 2\rho_i\left[g_1\left(\hat{s}_1, \mathbf{r}\right), ..., g_M\left(\hat{s}_M, \mathbf{r}\right)\right]^T\left[E(r_1), ..., E(r_M)\right]^T \\ &\approx 2\rho_i\left[g_1\left(\hat{s}_1, \mathbf{r}\right), ..., g_M\left(\hat{s}_M, \mathbf{r}\right)\right]^T \mathbf{r} \end{aligned}$$

In the algorithm, no extra information about the mixing matrix is needed, thus $\mathbf{A}$ is still learned by (5).

3.3. Order correction

Since an element of $g_i(\hat{s}_i, \mathbf{r})$ will reach a maximum when a reference corresponds to a source, the M estimated sources can be readily divided into two groups according to the values of $g_i(\hat{s}_i, \mathbf{r})$, i.e., L sources with prior information (corresponding to L maxima of $g_i(\hat{s}_i, \mathbf{r})$) and the remainings. This result can then be employed to adjust the order of the M estimates, specifically, to recover the L sources $\hat{s}_1, \ldots, \hat{s}_L$ with prior information in the same order as $r_1, \ldots, r_L$, and to recover the other sources $s_{L+1}, \ldots, s_M$ without prior information in random order (refer to Fig. 1).

Based upon the original order of $\hat{s}_1, \ldots, \hat{s}_L$ (i.e., order of the L maxima), the order correction can be done by:

$$\hat{\mathbf{s}}^{k+1} = \mathbf{P}\hat{\mathbf{s}}^{k+1} \qquad (11)$$

where $\mathbf{P}$ is an $M \times M$ permuted identify matrix:

$$\mathbf{P} = \left[e_1, \ldots, e_L, e_{L+1}, \ldots, e_M\right]$$

The order of $e_1, \ldots, e_L$ is the same as that of $r_1, \ldots, r_L$, the position of element "1" in $e_1, \ldots, e_L$ represents the original order of $\hat{s}_1, \ldots, \hat{s}_L$. As a result, we have

$$\hat{\mathbf{s}}^{k+1} = \mathbf{P}\hat{\mathbf{s}}^{k+1} = [\hat{s}_1, \ldots, \hat{s}_L, \hat{s}_{L+1}, \ldots, \hat{s}_M]^T$$

Note that we need to do column permutation for $\mathbf{A}$ by $\mathbf{A}^{k+1} = \mathbf{A}^{k+1}\mathbf{P}$ as well after correcting order for $\hat{s}$.

3.4. Construction of the reference signals

The reference signals should be constructed based upon the prior information and the closeness measure $g_i(\hat{s}_i, \mathbf{r})$. Here we utilize waveform information about some sources, and define $g_i(\hat{s}_i, \mathbf{r})$ using correlation criterion between an estimate and a reference. Therefore, a reference can be any signal having bigger correlation with its corresponding source than with the other sources, i.e., the references can be roughly constructed from the waveform information in different ways. Typical references could be: (1) the rough envelope of a source, (2) a set of pulses the distribution of which corresponds to the main peaks of a source, (3) rectangular waveform the amplitudes of which approximate the rough polarity of a source, etc. Usually, the first kind of references has the biggest correlation with the sources while the third one has the smallest.

4. SIMULATIONS AND RESULTS

Some simulations are performed to evaluate the performance of the proposed algorithm. One example presented below used the three speech signals (sampled at 8kHz, 10000 intercepted samples) in [3] as the sources s_1 - s_3, as shown in Fig. 2. These signals are with silent segments sparsely distributed and can be approximated by a Laplacian model. Two mixed signals x_1 - x_2 are shown in Fig. 3.

Assume that the waveform information about s_2 and s_3 is available. We constructed two references r_1 and r_2 for s_2 and s_3 using the third kind of references (to represent more general cases of rough references), and generated the reference r_3 for s_1 with uniformly distributed pseudo-random values between 0 and 1. Fig. 4 shows the constructed r_1 - r_3 (only 50 samples are displayed for clarity).

By running the proposed semi-blind EM algorithm ($\lambda = 0.01$, $\xi = 0.4$) and the original EM algorithm, respectively, we obtained the three estimates by the two algorithms, as shown in Fig. 5(a) and Fig. 5(b). We see that the two sets of recovered signals are very close to the original source signals in Fig. 2, but the estimates by the proposed algorithm are much cleaner than those by the original EM. In addition, the order of the three estimates by the proposed algorithm is definite (2, 3, 1), i.e., the same as that of r_1 - r_3, but the estimates order by the original EM may change at different runs (here is 1, 2, 3).

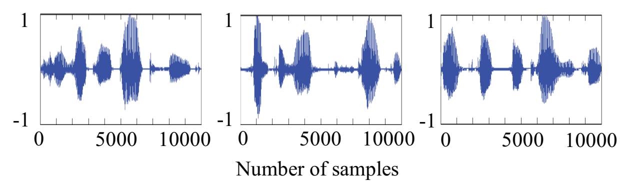

Fig. 2. Three speech signals used as sources s_1 - s_3

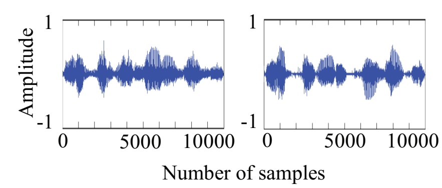

Fig. 3. Two mixed signals x_1 - x_2

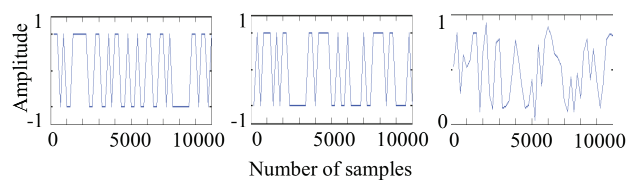

Fig. 4. Three reference signals r_1 - r_3

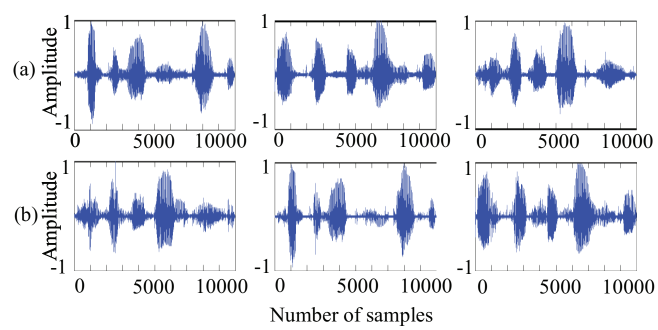

Fig. 5. Comparison of two algorithms. (a) Three estimates (with definite order 2, 3, 1) by the proposed algorithm utilizing waveform information about s_2 and s_3. (b) Three estimates by the original EM algorithm

To quantitatively compare performance of the two algorithms, we computed the signal-to-noise ratio (*SNR*) defined in [3] for the estimated signals. Table 1 shows the results. We see that the proposed algorithm has much higher

SNR than the original EM by use of prior information. We also compared the average *SNR* for all possible cases utilizing two references and utilizing one reference. Table 2 has the results which further confirm the advantages of incorporating prior information. In addition, when we compare the average *SNR* for the proposed algorithm by utilizing one reference with that by utilizing two references, we can find that the proposed algorithm has increased *SNR* when more prior information is used, e.g., the average *SNR* increases about 3dB when one more reference is used.

Table 1 Comparison of *SNR*(dB) for three estimates by the proposed algorithm utilizing waveform information about s_2 and s_3 and by the original EM algorithm

	s_1	s_2	s_3
Proposed	12.25	10.19	12.93
EM	5.28	9.08	9.77

Table 2 Comparison of average *SNR*(dB) for three estimates by the proposed algorithm (utilizing 1 reference and 2 references, respectively) and by the original EM algorithm

		s_1	s_2	s_3	average
Proposed	1 ref	5.42	9.72	11.13	8.76
	2 refs	12.81	10.15	12.96	11.97
EM		5.28	9.08	9.77	8.04

In addition, simulations indicate fast convergence of the proposed algorithm. Fig. 6 shows a comparison of *SNR* (for estimate of s_2) versus the number of iterations between the proposed algorithm and the original EM algorithm. We can see that the proposed algorithm converges at around 30 iterations whereas the EM algorithm converges at about 40 iterations. This may be achieved by the enhanced contribution of the extended cost function in the proposed algorithm when prior information is positively utilized.

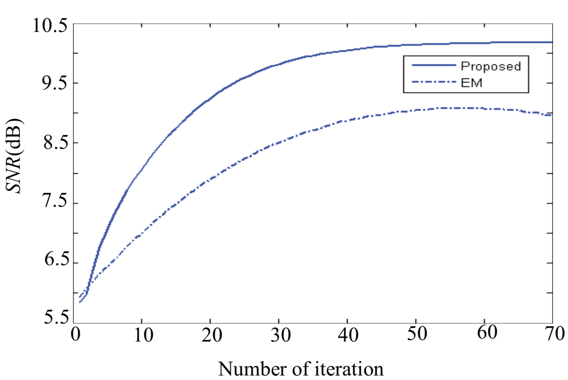

Fig. 6. Comparison of *SNR*(dB) (for estimate of s_2) versus the number of iterations between the proposed algorithm and the original EM algorithm

5. CONCLUSION

We proposed a semi-blind EM algorithm by incorporating waveform information about some sources into an approximate EM algorithm for overcomplete ICA. Simulation results demonstrate that the proposed semi-blind algorithm has much improved performance in *SNR*, convergence speed, and elimination of order ambiguity, compared to the original EM algorithm. Therefore, overcomplete ICA, similar to complete ICA, also benefits from use of prior information.

6. ACKNOWLEDGMENT

We thank Dr. Mingjun Zhong for many helpful comments on the manuscript. This work was supported by the National Natural Science Foundation of China under Grant No. 60402013, and the Liaoning Province Natural Science Foundation of China under Grant No. 20062174.

7. REFERENCES

[1] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Computation*, vol. 12, no. 2, pp. 337-365, 2000.

[2] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353-2362, 2001.

[3] M. Zhong, H. Tang, H. Chen, and Y. Tang, "An EM algorithm for learning sparse and overcomplete representations," *Neurocomputing*, vol. 57, pp. 469-476, 2004.

[4] C. Févotte and S. J. Godsill, "A bayesian approach for blind source separation of sparse sources," *IEEE Trans. Speech and Audio Processing*, vol. 14, no. 6, pp. 2174-2188, 2006.

[5] F. J. Theis, C. G. Puntonet, and E. W. Lang, "Median-based clustering for underdetermined blind signal processing," *IEEE Signal Processing Letters*, vol. 13, no. 2, pp. 96-99, 2006.

[6] J. Thomas, Y. Deville, and S. Hosseini, "Differential fast fixed-Point algorithms for underdetermined instantaneous and convolutive partial blind source separation," *IEEE Trans. Signal Processing*, vol. 55, no. 7, pp. 3717-3729, 2007.

[7] L. De Lathauwer, J. Castaing, and J. -F. Cardoso, "Fourth-order cumulant-based blind identification of underdetermined mixtures," *IEEE Trans. Signal Processing*, vol. 55, no. 6, pp. 2965-2973, 2007.

[8] W. Lu and J.C. Rajapakse, "Approach and applications of constrained ICA," *IEEE Trans. Neural Networks*, vol. 16, no. 1, pp. 203-212, 2005.

[9] V. D. Calhoun, T. Adali, M. C. Stevens, K. A. Kiehl, and J. J. Pekar, "Semi-blind ICA of fMRI: A method for utilizing hypothesis-derived time courses in a spatial ICA analysis," *NeuroImage*, vol. 25, no. 2, pp. 527-538, 2005.

[10] C. W. Hesse and C. J. James, "The FastICA algorithm with spatial constraints," *IEEE Signal Processing Letters*, vol. 12, no. 11, pp. 792-795, 2005.

[11] M. Knaak, S. Araki, and S. Makino, "Geometrically constrained independent components," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 15, No. 2, pp. 715-726, 2007.

[12] L. C. Parra and C. V. Alvino, "Geometric source separation: merging convolutive source separation with geometric beamforming," *IEEE Trans. Speech and Audio Processing*, vol. 10, no. 6, pp. 352-362, 2002.