ALGORITHMS FOR MARKOVIAN SOURCE SEPARATION BY ENTROPY RATE MINIMIZATION

Geng-Shen Fu**, Ronald Phlypo***, Matthew Anderson***, Xi-Lin Li*†*, and Tülay Adalı**

*University of Maryland, Baltimore County, Dept. of CSEE, Baltimore, MD 21250, USA
†Fortemedia, Sunnyvale, CA 94086, USA

ABSTRACT

Since in many blind source separation applications, latent sources are both non-Gaussian and have sample dependence, it is desirable to exploit both non-Gaussianity and sample dependency. In this paper, we use the Markov model to construct a general framework for the analysis and derivation of algorithms that take both properties into account. We also present two algorithms using two effective source priors. The first one is a multivariate generalized Gaussian distribution and the second is an autoregressive model driven by a generalized Gaussian distributed process. We derive the Cramér-Rao lower bound and demonstrate that the performance of the algorithms approach the lower bound especially when the underlying model matches the parametric model. We also demonstrate that a flexible semi-parametric approach exhibits very desirable performance.

***Index Terms*—** Blind source separation, Independent component analysis, Mutual information rate, Markov model.

1. INTRODUCTION

Independent component analysis (ICA) has been one of the most attractive solutions for the blind source separation (BSS) problem. ICA can estimate a demixing matrix and separate signals by assuming the source signals are mutually independent. It has been widely used in many applications such as biomedical signal processing, communications, and geophysical data analysis.

Most of the ICA algorithms exploit one of the following two properties: non-Gaussianity or sample dependency [1]. FastICA [2], the efficient variant of FastICA algorithm (EFICA) [3], Infomax [4], joint approximate diagonalization of eigenmatrices (JADE) [5], robust, accurate, direct independent components analysis algorithm (RADICAL) [6], ICA by entropy bound minimization (ICA-EBM) [7] only exploit non-Gaussianity property and assume samples are independently and identically distributed (i.i.d.). The second-order blind identification (SOBI) algorithm [8] and the weights-adjusted second-order blind identification (WA-SOBI) algorithm [9] use sample dependency, by only using the second order statistics. Other methods, such as the hybrid algorithm (MULTICOMBI) [10], autoregressive mixture of Gaussians (AR-MoG) [11], ERBM (introduced originally as FBSS) [12], and Markov source separation [13], aim at exploiting both non-Gaussianity and sample dependency.

In this paper, we discuss Markovian source separation since the Markov model is a very general model for many time series. We give the general cost function, update rule, and performance analysis for Markovian source separation. Also, we propose two algorithms, which exploit both sample dependency and non-Gaussianity by minimizing entropy rate under two different source prior assumptions: (1) a multivariate generalized Gaussian distribution (MGGD), and, (2) an autoregressive (AR) model driven by an i.i.d. process following a generalized Gaussian distribution (GGD). Under the first assumption, sources are separated by minimizing the difference of two joint entropies, which equals the entropy rate for a Markovian source. We call this entropy rate minimization via multivariate generalized Gaussian distribution (ERM-MG). The second algorithm relies on the GGD density function for the driving process of an AR model. We show that source separation can be achieved using entropy rate minimization via autoregression driven by generalized Gaussian process (ERM-ARG), and the algorithm approaches the Cramér-Rao lower bound (CRLB) effectively when the sources follow the model.

2. BACKGROUND

2.1. Linear Mixture Model

Let N statistically independent, zero mean sources $\mathbf{s}(t) = [s_1(t), \ldots, s_N(t)]^T$ be mixed through an $N \times N$ nonsingular mixing matrix $\mathbf{A}$ so that we obtain the mixtures $\mathbf{x}(t) = [x_1(t), \ldots, x_N(t)]^T$ as $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t), 1 \leq t \leq T$, where $(\cdot)^T$ denotes the transpose, and t is the sample (time) index. The mixtures are separated by $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$, where $\mathbf{y}(t) = [y_1(t), \ldots, y_N(t)]^T$, and $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]^T$ is the separation or demixing matrix to be estimated. These linear models can also be written in matrix form: $\mathbf{Y} = \mathbf{W}\mathbf{X} = \mathbf{W}\mathbf{A}\mathbf{S}$, where $\mathbf{S} = [\mathbf{s}_1, \ldots, \mathbf{s}_N]^T \in \mathbb{R}^{N\times T}$, $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]^T \in \mathbb{R}^{N\times T}$, $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_N]^T \in \mathbb{R}^{N\times T}$, $\mathbf{s}_i \in \mathbb{R}^T$, $\mathbf{x}_i \in \mathbb{R}^T$, and $\mathbf{y}_i \in \mathbb{R}^T$.

This work was supported by the NSF grant NSF-CIF 1117056.

2.2. Cost Function

A natural measure of dependence among N random processes y_i, $i = 1, \ldots, N$, is given by the mutual information rate:

$$\mathcal{I}_r(y_1; \ldots; y_N) = \sum_{i=1}^{N} H_r(y_i) - \log|\det(\mathbf{W})| - H_r(\mathbf{x}),$$

where $H_r(y_i) = \lim_{T\to\infty} H(y_i(1), ..., y_i(T))/T$ is the entropy rate of the process y_i, and $H_r(\mathbf{x})$ is the entropy rate of the observed vector-valued process $\mathbf{x}$, which is a constant with respect to $\mathbf{W}$. Hence, to achieve BSS, the cost function can be written as

$$\mathcal{J}_r(\mathbf{W}) = \sum_{i=1}^{N} H_r(y_i) - \log|\det(\mathbf{W})|. \tag{1}$$

If the samples are i.i.d., the cost function in (1) reduces to mutual information $\mathcal{I}(y_1; \ldots; y_N)$, and is equivalent to likelihood for i.i.d. samples [14,15], which has been widely used. For the general case where samples are not i.i.d., we can show that (1) is equivalent to likelihood given by $\mathcal{L}(\mathbf{W}) = -\sum_{i=1}^{N} \log p_{\mathbf{s}_i}(\mathbf{y}_i) - T\log|\det(\mathbf{W})|$.

3. MARKOVIAN SOURCE SEPARATION

To make the derivation tractable, we assume that the sources are stationary Kth-order Markov processes, where $K \ll T$. Hence, the entropy rate can be written as

$$\begin{aligned} H_r(s) &= \lim_{T\to\infty} H(s(T)|s(T-1), \ldots, s(1)) \\ &= H(s(T)|s(T-1), \ldots, s(T-K)) \\ &= H(\mathbf{s}_{(K+1)}) - H(\mathbf{s}_{(K)}) \end{aligned} \tag{2}$$

where $\mathbf{s}_{(P)} \in \mathbb{R}^P$ is a random vector, $\mathbf{s}_{(P)}(t) = [s(t), \ldots, s(t+P-1)]^T$ is the realization of $\mathbf{s}_{(P)}$ at index t, and $\mathbf{S}_{(P)} = [\mathbf{s}_{1,(P)}, \ldots, \mathbf{s}_{N,(P)}]^T \in \mathbb{R}^{N\times P}$. These definitions are also used for other variables.

3.1. Cost Function

For a Markovian source, the entropy rate equals to the difference of two joint entropies as in (2). Hence, under the Markov model assumption, ICA cost function in (1) can be written as

$$\begin{aligned} \mathcal{J}_r(\mathbf{W}) &= \sum_{i=1}^{N} (H(\mathbf{y}_{i,(K+1)}) - H(\mathbf{y}_{i,(K)})) - \log|\det(\mathbf{W})| \\ &= H(\mathbf{Y}_{(K+1)}) - H(\mathbf{Y}_{(K)}) - \log|\det(\mathbf{W})|. \end{aligned} \tag{3}$$

Comparing (3) with (1), we only need $K+1$ dimensional joint probability density function (PDF) $p_{\mathbf{s}_i}(\mathbf{y}_{i,(K+1)})$ in (3), instead of T dimensional joint PDF $p_{\mathbf{s}_i}(\mathbf{y}_i)$, for all sources with the Markovian assumption. By using the decoupling method [16, 17], we can divide the problem of minimizing $\mathcal{J}_r(\mathbf{W})$ with respect to the demixing matrix $\mathbf{W}$ into minimizing $\mathcal{J}_r(\mathbf{W})$ with respect to each of the row vectors $\mathbf{w}_i$, $i = 1, \ldots, N$. We can then write the cost as a function of only $\mathbf{w}_i$, which is

$$\mathcal{J}_{r_i}(\mathbf{w}_i) = H(\mathbf{y}_{i,(K+1)}) - H(\mathbf{y}_{i,(K)}) - \log|\mathbf{h}_i^T\mathbf{w}_i| + C,$$

where $\mathbf{h}_i$ is a unit Euclidian length vector which is perpendicular to all the row vectors of $\mathbf{W}$ except $\mathbf{w}_i$, and C is a constant term with respect to $\mathbf{w}_i$.

3.2. Update Rule

The gradient of $\mathcal{J}_r(\mathbf{W})$ with respect to $\mathbf{W}$ is given by

$$\frac{\partial \mathcal{J}_r(\mathbf{W})}{\partial \mathbf{W}} = E\left\{\boldsymbol{\varphi}(\mathbf{Y}_{(K+1)})\mathbf{X}_{(K+1)}^T\right\} - E\left\{\boldsymbol{\varphi}(\mathbf{Y}_{(K)})\mathbf{X}_{(K)}^T\right\} - \mathbf{W}^{-T},$$

where the score functions are given as

$$\begin{aligned} \boldsymbol{\varphi}(\mathbf{Y}_{(P)}) &= \left[\boldsymbol{\varphi}_1(\mathbf{y}_{1,(P)}), \ldots, \boldsymbol{\varphi}_N(\mathbf{y}_{N,(P)})\right]^T \in \mathbb{R}^{N\times P} \\ \boldsymbol{\varphi}_i(\mathbf{y}_{i,(P)}) &= \partial \log p_{\mathbf{y}_i}(\mathbf{y}_{i,(P)})/\partial \mathbf{y}_{i,(P)} \in \mathbb{R}^P. \end{aligned}$$

The natural gradient of $\mathcal{J}_r(\mathbf{W})$ with respect to $\mathbf{W}$ is given by

$$\begin{aligned} \Delta\mathbf{W} &= \frac{\partial \mathcal{J}_r(\mathbf{W})}{\partial \mathbf{W}}\mathbf{W}^T\mathbf{W} \\ &= \left(E\left\{\boldsymbol{\varphi}(\mathbf{Y}_{(K+1)})\mathbf{Y}_{(K+1)}^T\right\} - E\left\{\boldsymbol{\varphi}(\mathbf{Y}_{(K)})\mathbf{Y}_{(K)}^T\right\} - \mathbf{I}\right)\mathbf{W}. \end{aligned}$$

By using the decoupling method, the gradient update rule for each vector can be written as

$$\begin{aligned} \frac{\partial \mathcal{J}_{r_i}(\mathbf{W})}{\partial \mathbf{w}_i} &= E\left\{\mathbf{X}_{(K+1)}\boldsymbol{\varphi}(\mathbf{y}_{i,(K+1)})\right\} - E\left\{\mathbf{X}_{(K)}\boldsymbol{\varphi}(\mathbf{y}_{i,(K)})\right\} \\ &\quad -\mathbf{h}_i/\left(\mathbf{h}_i^T\mathbf{w}_i\right). \end{aligned}$$

3.3. Performance analysis

In [18, p.126], the general form of the Fisher information matrix (FIM) is given as

$$\mathbf{F}^{(ij)} = T\begin{pmatrix} f_{ij} & 1 \\ 1 & f_{ji} \end{pmatrix}, \tag{4}$$

where $f_{ij} \triangleq \frac{1}{T}\operatorname{Tr}(\boldsymbol{\Gamma}_i\mathbf{R}_j)$, $\boldsymbol{\Gamma}_i \triangleq E\{\boldsymbol{\varphi}_i(\mathbf{s}_i)\boldsymbol{\varphi}_i^T(\mathbf{s}_i)\} \in \mathbb{R}^{T\times T}$, and $\mathbf{R}_j \triangleq E\{\mathbf{s}_j\mathbf{s}_j^T\} \in \mathbb{R}^{T\times T}$. For $\mathbf{Y} = (\mathbf{I}+\boldsymbol{\mathcal{E}})\mathbf{S}$, the CRLB of $\mathcal{E}_{ij}$ is given by

$$E\left\{\mathcal{E}_{ij}^2\right\} \geq \frac{f_{ji}}{T(f_{ij}f_{ji}-1)}. \tag{5}$$

Since $\boldsymbol{\Gamma}_i$ is the autocorrelation matrix of $\boldsymbol{\varphi}_i(\mathbf{s}_i)$, then $\boldsymbol{\Gamma}_i$ is a symmetric Toeplitz matrix. In addition, since $\mathbf{s}_i$ is a Markovian source, $\boldsymbol{\Gamma}_i$ is a band diagonal matrix with left and right half-bandwidth equal to the Markov model order, K. To calculate f_{ij}, we only need to calculate a $(K+1)\times(K+1)$ submatrix of $\boldsymbol{\Gamma}_i$, denoted as $\boldsymbol{\Gamma}_{(K+1)}$, index i is suppressed for simplicity, which starts at (t,t) for $\forall t$.

$$\begin{aligned} \boldsymbol{\Gamma}_{(K+1)} &= E_{\mathbf{s}}\left\{\frac{\partial \log p(\mathbf{y})}{\partial \mathbf{y}_{(K+1)}(t)}\left(\frac{\partial \log p(\mathbf{y})}{\partial \mathbf{y}_{(K+1)}(t)}\right)^T\right\} \\ &= E_{\mathbf{s}_{(\tilde{K})}}\left\{\frac{\partial \log p(\mathbf{y}_{(\tilde{K})}(t-K))}{\partial \mathbf{y}_{(K+1)}(t)}\left(\frac{\partial \log p(\mathbf{y}_{(\tilde{K})}(t-K))}{\partial \mathbf{y}_{(K+1)}(t)}\right)^T\right\}, \end{aligned}$$

since $\mathbf{y}_{(\tilde{K})}(t-K)$ includes all entries of $\mathbf{y}$ that depend on $\mathbf{y}_{(K+1)}(t)$, where $\tilde{K} = 3K+1$. Hence, $\mathbf{\Gamma}_{(K+1)}$ is a submatrix of $\mathbf{\Gamma}_{(\tilde{K})} \triangleq E_{\mathbf{s}_{(\tilde{K})}}\{\boldsymbol{\varphi}(\mathbf{y}_{(\tilde{K})})\boldsymbol{\varphi}^T(\mathbf{y}_{(\tilde{K})})\}$, and FIM is determined.

To sum up, for the Markovian source, blind source separation and its performance analysis are achievable if we know the PDF of $\mathbf{s}_{(K+1)}$. Hence, the key problem in Markovian source separation is to determine the PDF of $\mathbf{s}_{(K+1)}$.

4. ALGORITHMS

4.1. Entropy rate minimization: Multivariate generalized Gaussian distribution model (ERM-MG)

So far, we have the general form for Markovian source separation, and modeling and estimation of the PDF of $\mathbf{s}_{(K+1)}$ is the key issue. The MGGD provides a flexible tool for data modeling and simulation. By assuming that $\mathbf{s}_{i,(P)}$ is a zero mean and P-dimensional MGGD, we can derive the score function and update rule. The zero mean MGGD density function is given by [19]

$$p_X(\mathbf{x}, \mathbf{\Sigma}, \beta, d) = \frac{\beta\Gamma(d/2)}{2^{d/2\beta}\pi^{d/2}\Gamma(d/2\beta)|\mathbf{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x})^\beta},$$

where $\mathbf{\Sigma}$ is the covariance parameter, β is the shape parameter, and d is the dimension of $\mathbf{x}$. MGGD will reduce to multivariate Gaussian if $\beta = 1$, and reduce to GGD if $d = 1$. By using the method of moments, $\mathbf{\Sigma}$ can be estimated as [19]

$$\widehat{\mathbf{\Sigma}} = \frac{d\Gamma(\frac{d}{2\beta})}{2^{1/\beta}\Gamma(\frac{d+2}{2\beta})}\mathbf{C}_{\mathbf{x}},$$

where $\mathbf{C}_{\mathbf{x}} = \sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t{}^T/T$ is the sample covariance matrix. MGGD parameters can also be estimated by using maximum likelihood (ML), but there are no closed form solutions. The entropy of an MGGD random vector is given by [20]

$$H(\mathbf{x}) = -\log\frac{\beta\Gamma(d/2)}{2^{d/2\beta}\pi^{d/2}\Gamma(d/2\beta)} + \frac{1}{2}\log|\mathbf{\Sigma}| + \frac{d}{2\beta}.$$

The MGGD score function can be written as

$$\boldsymbol{\varphi}(\mathbf{x}) = \beta(\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x})^{\beta-1}\mathbf{\Sigma}^{-1}\mathbf{x}. \quad (6)$$

Using the score function $\boldsymbol{\varphi}(\mathbf{x})$ given in (6), the matrix $\mathbf{\Gamma}_{(\tilde{K})}$ can be evaluated by generalized spherical coordinate transformation, as in [20, 21],

$$\begin{aligned}\mathbf{\Gamma}_{(\tilde{K})} &= E_{\mathbf{s}_{(\tilde{K})}}\left\{\boldsymbol{\varphi}_{\mathbf{y}_{(\tilde{K})}}\boldsymbol{\varphi}^T_{\mathbf{y}_{(\tilde{K})}}\right\} \\ &= \left(\frac{2\beta}{\tilde{K}\Gamma(\frac{\tilde{K}}{2\beta})}\right)^2 \Gamma\left(\frac{\tilde{K}+4\beta-2}{2\beta}\right)\Gamma\left(\frac{\tilde{K}+2}{2\beta}\right)\mathbf{\Sigma}^{-1}_{(\tilde{K})}.\end{aligned}$$

Hence, the FIM and CRLB can be evaluated by (4) and (5), respectively.

4.2. Entropy rate minimization: Autoregressive GGD source model (ERM-ARG)

Another effective Markovian source model is AR model driven by a GGD innovation process. For this model, we derive the cost function, update rule, and CRLB. We assume that the ith source is generated by the following AR model

$$s_i(t) = \sum_{k=1}^K q_{ik}s_i(t-k) + n_i(t),$$

where q_{ik} and K are the AR coefficients and order, respectively, and $n_i(t)$ is the white GGD process.

The mutual information rate cost in (3) can be derived as

$$\mathcal{J}_r(\mathbf{W}) = \sum_{i=1}^N H(n_i) - \log|\det(\mathbf{W})|. \quad (7)$$

From (7), we see that the entropy rate for source s_i is equivalent to the entropy of driving GGD process n_i. Hence, for this model, the cost function, update rule, and CRLB all have simpler forms than the general forms for Markovian source separation. By taking the derivative of the cost function in (7) with respect to $\mathbf{w}_i$, the update rule is given by

$$\begin{aligned}\frac{\partial\mathcal{J}_r(\mathbf{W})}{\partial\mathbf{w}_i} &= E\left\{\varphi_{n_i}(n_i)\frac{\partial n_i}{\partial\mathbf{w}_i}\right\} - \frac{\mathbf{h}_i}{\mathbf{h}_i^T\mathbf{w}_i} \\ &= \frac{1}{T}g_i(\mathbf{X})\boldsymbol{\varphi}_{\mathbf{n}_i} - \frac{\mathbf{h}_i}{\mathbf{h}_i^T\mathbf{w}_i},\end{aligned}$$

where $\boldsymbol{\varphi}_{\mathbf{n}_i} \triangleq [\varphi_{n_i}(n_i(1)), \ldots, \varphi_{n_i}(n_i(T))]^T$, $\varphi_{n_i}(n_i(t)) \triangleq -\partial\log p_{n_i}(n_i(t))/\partial n_i(t)$, $g_i(\mathbf{X}) \triangleq [g_i(\mathbf{x}_1), \ldots, g_i(\mathbf{x}_N)]^T \in \mathbb{R}^{N\times T}$, and $g_i(\mathbf{x}_j) \triangleq [x_j(1) - \sum_{k=1}^K q_{ik}x_j(1-k), \ldots, x_j(T) - \sum_{k=1}^K q_{ik}x_j(T-k)]^T \in \mathbb{R}^T$, which is the output of ith whitening filter driven by $\mathbf{x}_j$. The relative gradient of (7) is

$$\begin{aligned}\nabla\mathcal{J}_r(\mathbf{W}) &= \frac{\partial\mathcal{J}_r(\mathbf{W})}{\partial\mathbf{W}}\mathbf{W}^T \\ &= \frac{1}{T}\left(g_1(\mathbf{Y})\boldsymbol{\varphi}_{\mathbf{n}_1}, \ldots, g_N(\mathbf{Y})\boldsymbol{\varphi}_{\mathbf{n}_N}\right)^T - \mathbf{I}.\end{aligned}$$

Hence, the ijth entry of the likelihood relative gradient matrix is given by $\nabla\mathcal{L}(\mathbf{W})_{ij} = T\nabla\mathcal{J}_r(\mathbf{W})_{ij} = \boldsymbol{\varphi}^T_{\mathbf{n}_i}g_i(\mathbf{y}_j) - T\delta_{ij}$.

$$\begin{aligned}E\left\{(\nabla\mathcal{L}(\mathbf{W})_{ij})^2\right\} &= E\left\{(\boldsymbol{\varphi}^T_{\mathbf{n}_i}g_i(\mathbf{y}_j))^2\right\} \\ &= \mathrm{Tr}\left(E\left\{\boldsymbol{\varphi}_{\mathbf{n}_i}\boldsymbol{\varphi}^T_{\mathbf{n}_i}\right\}E\left\{g_i(\mathbf{y}_j)g_i^T(\mathbf{y}_j)\right\}\right) \\ &= E\left\{\varphi^2_{n_i}\right\}E\left\{\mathrm{Tr}\left(g_i(\mathbf{y}_j)g_i^T(\mathbf{y}_j)\right)\right\} \\ &= TE\left\{\varphi^2_{n_i}\right\}E\left\{g_i^2(y_j)\right\}.\end{aligned}$$

Hence, $f_{ij} = E\left\{\varphi^2_{n_i}\right\}E\left\{g_i^2(y_j)\right\}$, and the FIM and CRLB can be evaluated by (4) and (5), respectively.

5. EXPERIMENTAL RESULTS

In this section, we study the performances of the two proposed algorithms, ERM-MG and ERM-ARG, along with entropy rate bound minimization (ERBM), which provides a flexible and robust entropy rate estimator. We compare their performances with other BSS algorithms using simulated and real data. For simulated data, we compare performances in terms of interference to source ratio (ISR), which is given by $(1/N(N-1))\sum_{\{i,j=1,i\neq j\}}^{N} E\{g_{ij}^2\}$, where g_{ij} is the ijth entry of the the global demixing matrix $\mathbf{G} = \mathbf{WA}$, with CRLB to show that proposed algorithms exploit both sample dependency and non-Gaussianity. For real data, we compare performances in terms of inter-symbol-interference (ISI), which can be calculated as $(1/2N)(\sum_{i=1}^{N}\sum_{j=1}^{N}|g_{ij}|/\max_k|g_{ik}| + \sum_{j=1}^{N}\sum_{i=1}^{N}|g_{ij}|/\max_k|g_{kj}|) - 1$, to demonstrate the effectiveness of proposed algorithms if the data does not follow our model. All results are the average of 100 runs.

Experiment 1: We generate four sources using third-order AR models driven by GGD processes with shape parameters 0.5, 0.9, 1.5, and 5, respectively. From Fig.1, we see that ERM-ARG has the best performance among those algorithms because it is a ML method and the data fits the model exactly.

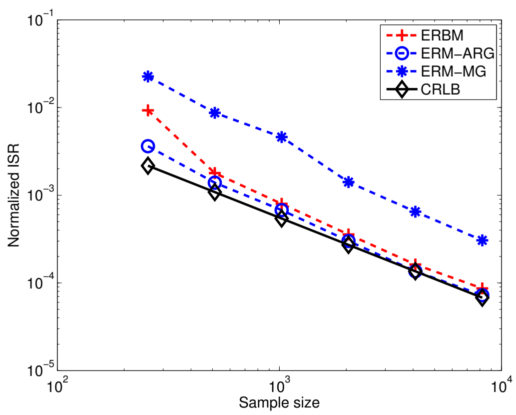

Fig. 1. Comparison of performances of ERBM, ERM-ARG, ERM-MG, and CRLB using simulated data generated by third order AR model driven by GGD processes.

Experiment 2: We compare the performances with the CRLB for the separation of two sources. One is an i.i.d. GGD source. The other one is generated by a first order AR model, with coefficients 0, 0.5, and 0.9, respectively, driven by GGD process. We introduce the diversity of both non-Gaussianity and sample dependency. As observed in the Fig.2, the CRLB decreases with increasing non-Gaussianity and/or diversity in terms of sample dependency, and sources are not separable when they are Gaussian with i.i.d. samples. Hence, as expected the performance improves when both types of diversity are used, and the model is identifiable for Gaussian sources as well, as long as their covariance matrices are not proportional.

Experiment 3: The performances of ten BSS algorithms are compared in the separation of artificial mixtures of ten image sources. JADE, EFICA, RADICAL and ICA-EBM only use the non-Gaussianity property for separation. WASOBI only uses the sample dependency property for separation. MULTICOMBI can exploit either non-Gaussianity or sample dependency for the estimation of a specific source.

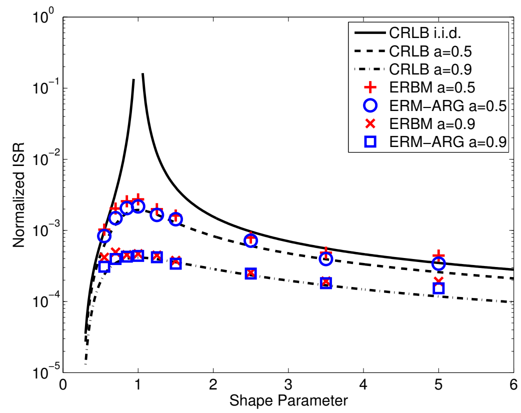

Fig. 2. Comparison of performances of ERBM and ERM-ARG, along with CRLB using simulated data generated by first order AR model driven by GGD process.

All others use both properties for source separation. As observed in Fig.3, both ERM-ARG and ERM-MG successfully separate image sources. But they cannot achieve the best performance among these algorithms because the real image data does not satisfy the assumptions for the source distribution exactly. ERBM has the best performance because it can model non-causal data and makes fewer assumptions on the distribution of the sources. Hence, its entropy rate estimator is more flexible and robust. We also see from Figs.1 and 2 that ERBM exhibits very desirable performance for simulated data.

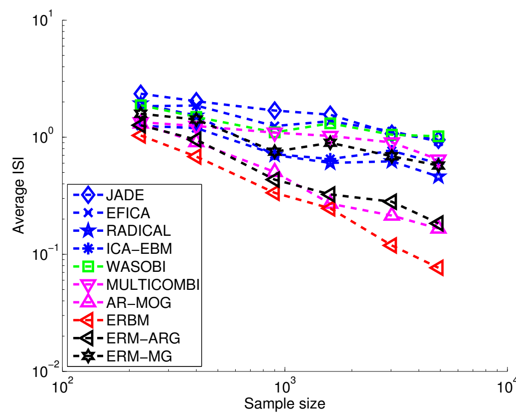

Fig. 3. Comparison of BSS algorithms in the separation of artificial mixtures of image sources.

6. CONCLUSION

We proposed a general cost function, update rule and performance analysis for Markovian source separation, and introduced two algorithms, ERM-MG and ERM-ARG. The algorithms assume that the sources come from MGGD, or are generated by AR models driven by GGD processes. We compared the performances of these two algorithms with the CRLB using simulated data, and demonstrated the effectiveness on real world data.

7. REFERENCES

[1] T. Adalı, M. Anderson, and G. Fu, "IVA and ICA: Use of diversity in independent decompositions," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, Aug. 2012, pp. 61–65.

[2] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *Neural Networks, IEEE Transactions on*, vol. 10, no. 3, pp. 626–634, May 1999.

[3] Z. Koldovský, P. Tichavský, and E. Oja, "Efficient variant of algorithm FastICA for independent component analysis attaining the Cramér-Rao lower bound," *Neural Networks, IEEE Transactions on*, no. 5, pp. 1265–1277.

[4] A. Bell and T. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[5] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Radar and Signal Processing, IEE Proceedings F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[6] E. G. Learned-Miller, J. W. Fisher III, T.-W. Lee, J.-F. Cardoso, E. Oja, and S.-I. Amari, "ICA using spacings estimates of entropy," *Journal of Machine Learning Research*, vol. 4, pp. 1271–1295, 2003.

[7] X.-L. Li and T. Adalı, "Independent component analysis by entropy bound minimization," *Signal Processing, IEEE Transactions on*, vol. 58, no. 10, pp. 5151–5164, Oct. 2010.

[8] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *Signal Processing, IEEE Transactions on*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[9] A. Yeredor, "Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting," *Signal Processing Letters, IEEE*, vol. 7, no. 7, pp. 197–200, Jul. 2000.

[10] P. Tichavsky, Z. Koldovsky, A. Yeredor, G. Gomez-Herrero, and E. Doron, "A hybrid technique for blind separation of non-Gaussian and time-correlated sources using a multicomponent approach," *Neural Networks, IEEE Transactions on*, vol. 19, no. 3, pp. 421–430, Mar. 2008.

[11] K. Hild, H. Attias, and S. Nagarajan, "An expectation-maximization method for spatio-temporal blind source separation using an AR-MOG source models," *Neural Networks, IEEE Transactions on*, vol. 19, no. 3, pp. 508–519, Mar. 2008.

[12] X.-L. Li and T. Adalı, "Blind spatiotemporal separation of second and/or higher-order correlated sources by entropy rate minimization," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, Mar. 2010, pp. 1934–1937.

[13] S. Hosseini, C. Jutten, and D. T. Pham, "Markovian source separation," *Signal Processing, IEEE Transactions on*, vol. 51, no. 12, pp. 3009–3019, Dec. 2003.

[14] J.-F. Cardoso, "High-order contrasts for independent component analysis," *Neural Comput.*, vol. 11, no. 1, pp. 157–192, Jan. 1999.

[15] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, Inc., 2002.

[16] X.-L. Li and X.-D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *Signal Processing, IEEE Transactions on*, vol. 55, no. 5, pp. 1803–1814, May 2007.

[17] M. Anderson, X.-L. Li, P. Rodriguez, and T. Adalı, "An effective decoupling method for matrix optimization and its application to the ICA problem," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, Mar. 2012, pp. 1885–1888.

[18] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, 1st ed. Academic Press, 2010.

[19] G. Verdoolaege and P. Scheunders, "Geodesics on the manifold of multivariate generalized Gaussian distributions with an application to multicomponent texture discrimination," *International Journal of Computer Vision*, vol. 95, pp. 265–286, 2011.

[20] G. Aulogiaris and K. Zografos, "A maximum entropy characterization of symmetric Kotz type and Burr multivariate distributions," *TEST*, vol. 13, pp. 65–83, 2004.

[21] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adalı, "Independent vector analysis, the Kotz distribution, and performance bounds," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2013, accepted.