BAYESIAN NOISY ICA FOR SOURCE SWITCHING ENVIRONMENTS

Jun-ichiro HIRAYAMA, Shin-ichi MAEDA and Shin ISHII

Graduate School of Information Science,
Nara Institute of Science and Technology
8916–5 Takayama, Ikoma, Nara, 630–0192 Japan

ABSTRACT

Most of the existing algorithms for blind source separation (BSS) assume that the number of sources is known and constant for all samples. Real situations, however, often have difficult non-stationarity such that each source signal abruptly switches to appear or disappear and hence the number of sources varies with time. In this article, we propose a noisy independent component analysis (ICA) algorithm that assumes unknown and varying number of sources. We employ Bayesian variable selection in combination with the hidden Markov model to automatically select and switch the set of sources which are temporally active in a certain period. We formulate our algorithm based on the Bayesian inference using the variational Bayes method. A simulation study using artificial data showed that our approach successfully recovered source signals even when the number of sources varied with time.

1. INTRODUCTION

Blind source separation (BSS) is the problem of recovering unknown source signals when only their linear mixtures are observed, without having knowledge about the mixing process. The problem has many potential applications, such as signal processing of speech, image or biomedical signals. A popular approach to address BSS is the independent component analysis (ICA) which utilizes statistical independence of source signals to recover the sources and mixing coefficients.

Most of the existing algorithms for BSS/ICA assume that the number of sources is known and constant for all samples. In practice, however, this assumption does not necessarily hold. The number of sources is rarely known a priori, and real situations often have difficult non-stationarity such that each source signal abruptly switches to appear or disappear and hence the number of sources varies with time. Recently some literatures have addressed the problem with unknown (but fixed) number of sources [1, 2, 3, 4], while the cases with unknown and varying number of sources were not explicitly addressed so far. In this study, particularly focussing on the latter *source switching* situations in BSS, we propose an ICA method that assumes unknown and varying number of sources.

We employ the main idea of Bayesian variable selection [5, 6, 7, 8] combined with the hidden Markov model (HMM) [9] to select a set of sources which are temporally active in a certain period. In the context of variable selection in regression problems, independent variables necessary for explaining observations can be selected from many potential variables. A practical scheme to do this is the Bayesian variable selection. The basic idea is to introduce a hierarchical prior to the regression coefficient for each variable, which makes the coefficient zero and then vanished when it is irrelevant. Conventionally, however, this approach is not regarded as suitable for the case where necessary variables may switch with time. In this article, we extend the idea with utilizing HMM in order to introduce temporal structure into the prior, and the prior is applied to latent sources instead of mixing vectors (i.e., regression coefficients). Hence, our approach provides a "dynamic" variable selection scheme, which is different from the conventional "static" one.

Although most ICA algorithms consider noise-free situations, we focus on noisy cases for its practical significance. Several noisy ICA algorithms have already been proposed, which were mainly based on the maximum likelihood (ML) with assuming parametric models for source distribution [10, 11, 12, 13, 14, 15], whereas some semiparametric methods which have no assumption for the form of source distribution were also proposed [16, 17]. Some authors have recently developed Bayesian approaches to noisy ICA [2, 18, 19] as natural extensions of ML-based ones, using the variational Bayes (VB) [20] method. For ML- and VB-ICA, a common way to model source distribution is to use a mixture of Gaussian (MoG) distribution [10, 13, 15]. In this article, we employ a special variation of MoG, a scale mixture of Gaussian [21, 22], to represent the super-Gaussian nature of source signals, and propose a Bayesian noisy ICA algorithm using the VB method, with introducing a dynamic variable selection mechanism.

2. PROPOSED APPROACH

2.1. Linear mixture with isotropic Gaussian noise

The situation is assumed as follows. d observations at time step t, $x_{j,t}\,(j=1,\ldots,d)$, are generated as a linear mixture of n independent source signals, $s_{i,t}\,(i=1,\ldots,n)$, plus Gaussian noise. The probabilistic generative model for the observation vector $\boldsymbol{x}_t=(x_{1,t},x_{2,t},\ldots,x_{d,t})'$ is given by

$$\boldsymbol{x}_t=\boldsymbol{A}\boldsymbol{s}_t+\boldsymbol{\nu}_t, \tag{1}$$

where $\boldsymbol{s}_t=(s_{1,t},s_{2,t},\ldots,s_{n,t})'$ is the source vector. $d\times n$ matrix $\boldsymbol{A}$ is the mixing matrix, and each component of the d-dimensional vector $\boldsymbol{\nu}_t$ is independently and identically distributed as Gaussian with the mean zero and the variance τ^{-1}. Conditional density of $\boldsymbol{x}_t$ is then

$$p(\boldsymbol{x}_t\mid\boldsymbol{s}_t,\boldsymbol{A},\tau)=\mathrm{N}_d(\boldsymbol{x}_t\mid\boldsymbol{A}\boldsymbol{s}_t,\tau^{-1}\boldsymbol{I}_d), \tag{2}$$

where $\mathrm{N}_d(\cdot\mid\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes a d-dimensional normal distribution with a mean $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{\Sigma}$.

2.2. Markov-switching sources

We introduce binary variables $z_{i,t}\in\{0,1\}\,(i=1,\ldots,n)$ to indicate the existence of the i-th source at time step t; $z_{i,t}$ equals one if the i-th source signal $s_{i,t}$ exists (is active) at that time, while zero if the source signal does not exist (is inactive). Using this hidden indicator variables, the prior distribution of $s_{i,t}$ is defined as a two-component mixture model as in the usual setting of Bayesian variable selection [7]. In this article, we model the source distribution as a scale mixture of two Gaussian distributions, conditional on $z_{i,t}=1$. The conditional density is given as follows, by introducing the corresponding latent variable $y_{i,t}\in\{0,1\}$ for the two Gaussian components:

$$\begin{aligned}&p(s_{i,t}\,|y_{i,t},z_{i,t}=1)\\&\qquad=\mathrm{N}(s_{i,t}\mid 0,\phi_1^2)^{y_{i,t}}\mathrm{N}(s_{i,t}\mid 0,\phi_0^2)^{1-y_{i,t}},\end{aligned} \tag{3a}$$

$$p(y_{i,t})=\rho^{y_{i,t}}(1-\rho)^{1-y_{i,t}}, \tag{3b}$$

where ρ is the mixing rate for the two Gaussian components, and ϕ_0^2 and ϕ_1^2 are their variances. Although these are fixed in this study, they can also be adjusted from data [13]. A scale MoG allows us to obtain an analytical solution of the posterior distribution for super-Gaussian source signals. We note that other kinds of source distribution can be employed, such as Laplace [14], Student-t [3], or more general mixture of Gaussian distribution [10, 13], possibly under some approximation. When $z_{i,t}=0$, in contrast, each source distribution is modeled as a Dirac's delta distribution:

$$p(s_{i,t}\mid z_{i,t}=0)=\delta(s_{i,t}), \tag{4}$$

where $\delta(\cdot)$ is the Dirac's delta function. This setting allows our model to remove unnecessary sources effectively. The conditional density of the source vector $\boldsymbol{s}_t$ is then given by

$$p(\boldsymbol{s}_t\mid\boldsymbol{y}_t,\boldsymbol{z}_t)=\mathrm{N}_n(\boldsymbol{s}_t\mid\boldsymbol{0}_n,\boldsymbol{V}_t(\boldsymbol{y}_t,\boldsymbol{z}_t)), \tag{5}$$

where $\boldsymbol{V}_t(\boldsymbol{y}_t,\boldsymbol{z}_t)$ is a diagonal covariance matrix whose i-th diagonal element is $\sigma_{i,t}^2(y_{i,t},z_{i,t})=z_{i,t}\phi_{i,t}^2(y_{i,t})$ where $\phi_{i,t}^2(y_{i,t})=y_{i,t}\phi_1^2+(1-y_{i,t})\phi_0^2$.

If the indicator variable $\boldsymbol{z}_t$ is an i.i.d random vector having no time-dependency, the mixture source model is equivalent to those used in sparse coding methods [23, 24]. On the other hand, if the indicator has the same value at any time, $\boldsymbol{z}_1=\boldsymbol{z}_2=\cdots=\boldsymbol{z}_T$, this case is equivalent to that of Bayesian variable selection in its usual sense. In this article, to incorporate time-structure into the source model, we model the indicator variable $\boldsymbol{z}_t$ as a Markov process, which is, for simplicity, assumed to be independent of the other sources:

$$p(z_{i,1}) = \begin{cases} 1-\pi_i & \text{if } z_{i,1}=0 \\ \pi_i & \text{if } z_{i,1}=1 \end{cases}, \tag{6a}$$

$$p(z_{i,t} \mid z_{i,t-1}) = \begin{cases} 1-r_{1i} & \text{if } (z_{i,t}, z_{i,t-1}) = (0,0) \\ r_{1i} & \text{if } (z_{i,t}, z_{i,t-1}) = (1,0) \\ r_{0i} & \text{if } (z_{i,t}, z_{i,t-1}) = (0,1) \\ 1-r_{0i} & \text{if } (z_{i,t}, z_{i,t-1}) = (1,1) \end{cases} \tag{6b}$$

where π_i is the probability for initial existence of the i-th source, and $r_{1i}, r_{0i} \in [0,1]$ are probabilities that the i-th source switches to active and inactive, respectively. The source model is thus formulated as an HMM. Although the time-structure of the source model is the key of our dynamic variable selection mechanism, our model includes the one without any dynamics as its special case, because the special setting: $r_{0i}+r_{1i}=1$ for any i implies that $p(z_{i,t}=1)=r_{1i}$ and $p(z_{i,t}=0)=r_{0i}$ for every time step.

2.3. Variational Bayes inference

A Bayesian inference for the model described above is performed by the variational Bayes (VB) method [20]. Let $\boldsymbol{\xi} = \{\boldsymbol{s}_t, \boldsymbol{y}_t, \boldsymbol{z}_t | t=1,\dots,T\}$ denote the set of latent variables, $\boldsymbol{\theta} = \{\boldsymbol{A}, \tau\}$ the set of observation model parameters, and $\boldsymbol{\omega} = \{\pi_i, r_{1i}, r_{0i} \mid i=1,\dots,n\}$ the set of dynamics model parameters. The VB method solves a functional optimization problem to obtain the probability distribution $q(\boldsymbol{\xi},\boldsymbol{\theta},\boldsymbol{\omega})$ that well approximates the true posterior $p(\boldsymbol{\xi},\boldsymbol{\theta},\boldsymbol{\omega}|\boldsymbol{X})$, where $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_T)$. We introduce the factorization assumption: $q(\boldsymbol{\xi},\boldsymbol{\theta},\boldsymbol{\omega}) \approx q(\boldsymbol{\xi})q(\boldsymbol{\theta})q(\boldsymbol{\omega})$ to allow the solution to have closed forms with respect to $q(\boldsymbol{\xi})$, $q(\boldsymbol{\theta})$ and $q(\boldsymbol{\omega})$. For the parameters $\boldsymbol{\theta}$, we use a conjugate prior:

$$p_0(\boldsymbol{A},\tau) = \mathrm{N}_{d\times n}\left(\boldsymbol{A}|\boldsymbol{M}_0, \tau^{-1}\boldsymbol{I}_d, \boldsymbol{G}_0^{-1}\right)\mathrm{Ga}(\tau|\kappa_0,\lambda_0), \tag{7}$$

where $\mathrm{N}_{d\times n}(\cdot \mid \cdot,\cdot,\cdot)$ denotes a matrix normal distribution (Appendix A), and $\mathrm{Ga}(\cdot \mid \cdot,\cdot)$ is a Gamma distribution. For $\boldsymbol{\omega}$, we also use the following conjugate prior:

$$p_0(\pi_i) = \mathrm{Be}(\pi_i \mid u_{\pi_i}, w_{\pi_i}), \tag{8a}$$

$$p_0(r_{0i}) = \mathrm{Be}(r_{0i} \mid u_{r_{0i}}, w_{r_{0i}}), \tag{8b}$$

$$p_0(r_{1i}) = \mathrm{Be}(r_{1i} \mid u_{r_{1i}}, w_{r_{1i}}), \tag{8c}$$

for $i=1,\dots,n$, where $\mathrm{Be}(\cdot \mid \cdot,\cdot)$ is the Beta distribution.

According to the VB method, the approximate posterior for $\boldsymbol{s}_t$, conditional on $\boldsymbol{y}_t$ and $\boldsymbol{z}_t$, is derived as

$$q(\boldsymbol{s}_t \mid \boldsymbol{y}_t, \boldsymbol{z}_t) = \frac{\psi(\boldsymbol{x}_t, \boldsymbol{s}_t)\, p(\boldsymbol{s}_t \mid \boldsymbol{y}_t, \boldsymbol{z}_t)}{\int d\boldsymbol{s}_t \psi(\boldsymbol{x}_t, \boldsymbol{s}_t)\, p(\boldsymbol{s}_t \mid \boldsymbol{y}_t, \boldsymbol{z}_t)} \tag{9a}$$

$$= \mathrm{N}_n(\boldsymbol{s}_t \mid \boldsymbol{\mu}_t(\boldsymbol{y}_t,\boldsymbol{z}_t), \hat{\boldsymbol{V}}_t(\boldsymbol{y}_t,\boldsymbol{z}_t)), \tag{9b}$$

$$\hat{\boldsymbol{V}}_t(\boldsymbol{y}_t,\boldsymbol{z}_t) = \left(\langle \tau \boldsymbol{A}'\boldsymbol{A} \rangle + \boldsymbol{V}_t(\boldsymbol{y}_t,\boldsymbol{z}_t)^{-1}\right)^{-1}, \tag{9c}$$

$$\boldsymbol{\mu}_t(\boldsymbol{y}_t,\boldsymbol{z}_t) = \hat{\boldsymbol{V}}_t(\boldsymbol{y}_t,\boldsymbol{z}_t)\, \langle \tau \boldsymbol{A} \rangle' \boldsymbol{x}_t, \tag{9d}$$

where $\psi(\boldsymbol{x}_t, \boldsymbol{s}_t) \equiv \exp\left(\langle \log p(\boldsymbol{x}_t \mid \boldsymbol{s}_t, \boldsymbol{A}, \tau)\rangle_{\boldsymbol{A},\tau}\right)$. The expectations $\langle \tau\boldsymbol{A}\rangle$ and $\langle \tau \boldsymbol{A}'\boldsymbol{A}\rangle$ are given by Eq. (13) below. Let $l(\boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{z}_t)$ denote the denominator in Eq. (9a). The marginal likelihood for $\boldsymbol{z}_t$ is given as

$$e(\boldsymbol{x}_t, \boldsymbol{z}_t) = \sum_{\boldsymbol{y}_t} l(\boldsymbol{x}_t, \boldsymbol{y}_t, \boldsymbol{z}_t)\, p(\boldsymbol{y}_t), \tag{10}$$

which is the summation over 2^n possible values of the n-dimensional binary vector $\boldsymbol{y}_t$. Using the marginal likelihood, the approximate joint posterior of $\boldsymbol{z}_t$ for all t is

$$q(\boldsymbol{z}_1, \boldsymbol{z}_2, \dots, \boldsymbol{z}_T) = \frac{1}{C_{\boldsymbol{\xi}}}\, e(\boldsymbol{x}_1, \boldsymbol{z}_1)\psi_\pi(\boldsymbol{z}_1) \prod_{t=2}^{T} e(\boldsymbol{x}_t, \boldsymbol{z}_t)\psi_r(\boldsymbol{z}_t, \boldsymbol{z}_{t-1}), \tag{11}$$

where $\psi_\pi(\boldsymbol{z}_1) \equiv \exp(\langle \log p(\boldsymbol{z}_1 \mid \boldsymbol{\pi})\rangle_{\boldsymbol{\pi}})$ and $\psi_r(\boldsymbol{z}_t, \boldsymbol{z}_{t-1}) \equiv \exp(\langle \log p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1})\rangle_{\boldsymbol{r}_0, \boldsymbol{r}_1})$. $C_{\boldsymbol{\xi}}$ is the normalization term. Because Eq. (11) is a simple HMM whose transition and observation processes are given by $\psi_r(\boldsymbol{z}_t, \boldsymbol{z}_{t-1})$ and $e(\boldsymbol{x}_t, \boldsymbol{z}_t)$, respectively, the marginal distributions, $q(\boldsymbol{z}_1), \dots, q(\boldsymbol{z}_T)$, are exactly calculated by the standard Forward-Backward algorithm [9]. The pairwise posteriors $q(\boldsymbol{z}_t, \boldsymbol{z}_{t-1})$ can also be exactly calculated for updating the approximate posterior for π_i, r_{1i}, r_{0i} which are given as Beta distributions. The detail of Bayesian learning of HMM can be found in [25]. Taking expectations with respect to the posteriors, the following statistics are calculated:

$$\boldsymbol{R}_{ss} = \langle \boldsymbol{S}\boldsymbol{S}' \rangle + \boldsymbol{G}_0, \tag{12a}$$

$$\boldsymbol{R}_{xs} = \boldsymbol{X}\langle \boldsymbol{S} \rangle' + \boldsymbol{M}_0\boldsymbol{G}_0, \tag{12b}$$

$$\boldsymbol{R}_{xx} = \boldsymbol{X}\boldsymbol{X}' + \boldsymbol{M}_0\boldsymbol{G}_0\boldsymbol{M}_0', \tag{12c}$$

$$\boldsymbol{R}_{x|s} = \boldsymbol{R}_{xx} - \boldsymbol{R}_{xs}\boldsymbol{R}_{ss}^{-1}\boldsymbol{R}_{xs}', \tag{12d}$$

where $\boldsymbol{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \dots, \boldsymbol{s}_T)$. Then the approximate posterior for $\boldsymbol{A}$ and τ is given as

$$q(\boldsymbol{A},\tau) = \mathrm{N}_{d\times n}(\boldsymbol{A} \mid \hat{\boldsymbol{M}}, \tau^{-1}\boldsymbol{I}_d, \hat{\boldsymbol{G}}^{-1})\mathrm{Ga}(\tau \mid \hat{\kappa}, \hat{\lambda}), \tag{13}$$

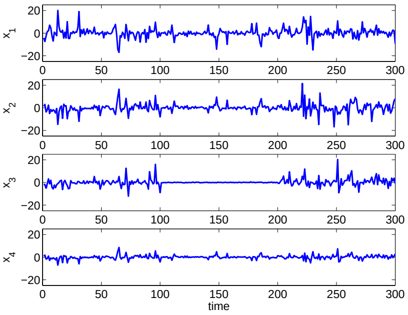

Fig. 1. Test time-series of four observation signals.

where

$$\hat{\boldsymbol{M}} = \boldsymbol{R}_{xs}\boldsymbol{R}_{ss}^{-1}; \quad \hat{\boldsymbol{G}} = \boldsymbol{R}_{ss}, \tag{14a}$$

$$\hat{\kappa} = \kappa_0 + \frac{d(n+T)}{2}; \quad \hat{\lambda} = \lambda_0 + \frac{1}{2}\mathrm{tr}\left[\boldsymbol{R}_{x|s}\right]. \tag{14b}$$

The VB inference is performed by calculating the expectations with respect to $\boldsymbol{\xi}$, $\boldsymbol{\theta}$ and $\boldsymbol{\omega}$ alternately, with supplementary calculation of the marginal posterior distributions of $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T$ by the Forward-Backward procedure at each iteration. After convergence, all unknown variables are conveniently estimated as their expected values under the LMS (least mean square) loss criterion, while other estimators, such as the MAP (maximum a posteriori) estimator, can also be employed.

3. RESULTS

Figure 1 shows the test time-series we used. The number of observation signals was $d = 4$ and the time-series length was $T = 300$. The time-series were synthesized according to the generative model given by Eq. (1) with three artificial source signals which are shown in the left column of Figure 2 (with one always-inactive source at the bottom). The bar graph at the top of each panel shows the source existence, where the brightness indicates the value from zero to one as depicted in the right side of the figure. The number of active source signals was varied with time: two, one and three for each 100 time steps. The source signals for active periods were artificially generated from the scale mixture of Gaussian, Eq.(3), with $\phi_0 = 0.5, \phi_1 = 2.5$, and $\rho = 0.25$. The Hinton diagram [26] of the true mixing matrix is shown in the left panel of Figure 3. The noise inverse variance was set at $\tau = 26.2$ (average SNR ≈ 25 (dB)). The VB steps were iterated 400 times. Prior hyperparameters were set at: $\boldsymbol{M}_0 = \boldsymbol{0}, \boldsymbol{G}_0 = 0.05\boldsymbol{I}_n, v_0 = 1, w_0 = 1 \times 10^{-3}, u_\cdot = w_\cdot = 0.5$. The mean parameter of trial distribution for mixing matrix, $\boldsymbol{M}$, was initialized by JADE [27]. The other parameters of trial distributions were randomly chosen.

Figure 2 (right) shows estimated sources and the existence by our approach, which were obtained as LMS estimates. The bar graphs shows the estimate (possibly analog) of existence indicator. Our approach recovered the three source signals and their existence almost correctly, so that the estimated number of sources varies with time like in the reality. The mixing matrix was also well estimated as shown in the right panel of Figure 3, where the right three columns correspond to the true ones. The fourth column was automatically pruned in the learning process, because the fourth source was estimated as consistently inactive so that no information was offered to learn the fourth column vector and then the prior estimate was filled in. It should be noted that the order and signs of estimated sources and mixing matrix in these figures were manually adjusted to clarify the correspondence to the true ones.

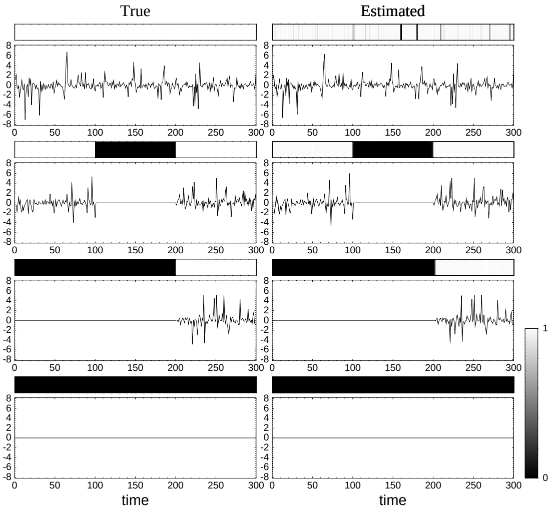

Fig. 2. True (left column) and estimated (right column) source signals and their existence.

4. DISCUSSION

We proposed a novel method to address the source switching situations in BSS, in which the number of sources is

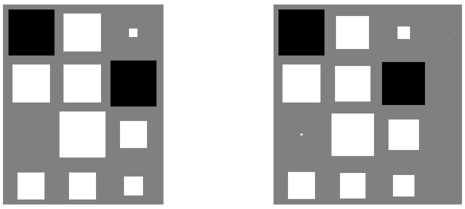

Fig. 3. Hinton diagrams [26] of the true (left) and estimated (right) mixing matrices.

not only unknown but also varying with time. A dynamical model was introduced as the source model to achieve automatic selection of active sources, which was based on the combination of the hierarchical prior which has been used in Bayesian variable selection and the HMM. We formulated a generative model including the dynamic source model and noisy linear mixing, and derived the Bayesian inference algorithm using the VB method. A simulation study using artificial data showed that our approach successfully recovered source signals in a non-stationary situation such that each source occasionally switched to appear or disappear and then the true (unknown) number of sources varied with time.

In this article, we showed that our approach had a potential to overcome the source switching problem in BSS at least in principle. The quantitative performance of our method and its comparison to alternative methods are currently under investigation. In our formulation, however, the computational cost is too high to directly apply to problems with many potential sources. Actually, the cost grows in an exponential order of the number of potential sources, n; then, it is intractable when n is large. The reason was twofold. One reason is to use a MoG as an active source model so that the posterior for all sources becomes an n-dimensional MoG with an intractable number of components if n is large. This may be solved in part by using the factorization of trial distribution of the sources as introduced in [13]. Another reason, which is more crucial, is the load to calculate exactly the forward/backward probabilities for the whole 2^n states of $\boldsymbol{z}_t$. In the Bayesian variable selection literatures [5, 6, 7, 8], such a problem was overcome using the Markov Chain Monte Carlo (MCMC) technique, while the time dependency assumed in our model refuses a direct application of it. A potential way to address this issue is the use of sequential Monte Carlo [28]. We are now seeking on this way to achieve a computationally efficient formulation of our algorithm.

A. APPENDIX

A matrix normal distribution is defined as

$$
\begin{aligned}
&\mathrm{N}_{d\times n}\left(\boldsymbol{A} \mid \boldsymbol{M}, \boldsymbol{V}, \boldsymbol{K}\right) \\
&= (2\pi)^{-\frac{dn}{2}} |\boldsymbol{K}|^{-\frac{d}{2}} |\boldsymbol{V}|^{-\frac{n}{2}} \\
&\quad \times \exp\left(-\frac{1}{2}\mathrm{tr}\left[\left(\boldsymbol{A}-\boldsymbol{M}\right)' \boldsymbol{V}^{-1}\left(\boldsymbol{A}-\boldsymbol{M}\right)\boldsymbol{K}^{-1}\right]\right),
\end{aligned}
$$

where $\boldsymbol{A} \in \Re^{d\times n}$, $\boldsymbol{M} \in \Re^{d\times n}$, $\boldsymbol{K} \in \Re^{n\times n}$, and $\boldsymbol{V} \in \Re^{d\times d}$. $\boldsymbol{M}$ denotes the mean of $\boldsymbol{A}$; $\boldsymbol{K}$ and $\boldsymbol{V}$ are two covariance matrices of $\boldsymbol{A}$ [29].

B. REFERENCES

[1] A. Cichocki, J. Karhunen, W. Kasprzak, and R. Vigario, "Neural networks for blind separation with unknown number of sources," *Neurocomputing*, vol. 24, no. 1, pp. 55–93, 1999.

[2] R. Choudrey and S. Roberts, "Flexible Bayesian independent component analysis for blind source separation," in *Proc. Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, 2001, pp. 90–95.

[3] J. T. Cao, N. Murata, S. Amari, A. Cichocki, and T. Takeda, "A robust approach to independent component anslysis of signals with high-level noise measurements," *IEEE Trans. on Neural Networks*, vol. 14, no. 3, pp. 631–645, 2003.

[4] J. Ye, X. Zhu, and X. Zhang, "Adaptive blind separation with an unknown number of sources," *Neural Computation*, vol. 16, pp. 1641–1660, 2004.

[5] E. I. George and R. E. McCulloch, "Variable selection via gibbs sampling," *J. Am. Statist. Ass.*, vol. 88, pp. 881–889, 1993.

[6] J. Geweke, "Variable selection and model comparison in regression," in *Bayesian Statistics 5*, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Eds., pp. 609–620. Clarendon Press, Oxford, UK, 1996.

[7] E. I. George and R. E. McCulloch, "Approaches for Bayesian variable selection," *Statistica Sinica*, vol. 7, no. 2, pp. 339–373, 1997.

[8] P. J. Brown, M. Vannucci, and T. Fearn, "Multivariate Bayesian variable selection and prediction," *Journal of the Royal Statistical Society, Series B*, vol. 60, no. 3, pp. 627–641, 1998.

[9] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[10] E. Moulines, J.-F. Cardoso, and E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models," in *Proc. ICASSP*, 1997, vol. 5, pp. 3617–3620.

[11] A. Cichocki, S. C. Douglas, and S. Amari, "Robust techniques for independent component analysis (ICA) with noisy data," *Neurocomputing*, vol. 22, no. 1–3, pp. 113–129, 1998.

[12] A. Hyvärinen, "Independent component analysis in the presence of gaussian noise by maximizing joint likelihood," *Neurocomputing*, vol. 22, pp. 49–67, 1998.

[13] H. Attias, "Independent factor analysis," *Neural Computation*, vol. 11, no. 4, pp. 803–851, 1999.

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

[15] M. Welling and M. Weber, "A constrainted EM algorithms for independent component analysis," *Neural Computation*, vol. 13, pp. 677–689, 2001.

[16] A. Hyvärinen, "Gaussian moments for noisy independent component analysis," *IEEE Signal Processing Letters*, vol. 6, pp. 145–147, 1999.

[17] M. Kawanabe and N. Murata, "Independent component analysis in the presence of gaussian noise based on estimating functions," in *Proc. International Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, 2000, pp. 39–44.

[18] J. Miskin and D. MacKay, "Ensemble learning for blind source separation," in *Independent Component Analysis: Principles and Practice*, S. Roberts and R. Everson, Eds., pp. 209–233. Cambridge University Press, 2001.

[19] K. Chan, T.-W. Lee, and T. J. Sejnowski, "Variational Bayesian learning of ica with missing data," *Neural Computation*, vol. 15, pp. 1991–2011, 2003.

[20] H. Attias, "Inferring parameters and structure of latent variable models by variational Bayes," in *Proc. 15th Conf. on Uncertainty in AI*, 1999, pp. 21–30.

[21] D. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *J. Royal Stat. Soc. B*, vol. 36, pp. 99–102, 1974.

[22] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of gaussians and the statistics of natural images," in *Adv. Neural Information Processing Systems*, Cambridge, MA, 2000, vol. 12, pp. 855–861, MIT Press.

[23] B. A. Olshausen and K. J. Millman, "Learning sparse codes with a mixture-of-gaussians prior," in *Advances in Neural Information Processing Systems*. 2000, vol. 12, pp. 841–884, MIT Press.

[24] P. Sallee and B. A. Olshausen, "Learning sparse multiscale image representations," in *Advances in Neural Information Processing Systems*, 2003, vol. 15, pp. 1327–1334.

[25] D. J. C. MacKay, "Ensemble learning for hidden Markov models," Tech. Rep., Cavendish Laboratory, University of Cambridge, 1997.

[26] C. M. Bishop, *Neural networks for pattern recognition*, Oxford University Press, Oxford, UK, 1995.

[27] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEEE Proc. -F*, vol. 140, no. 6, pp. 362–370, 1993.

[28] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo in Practice*, Springer-Verlag, 2001.

[29] T. Minka, "Bayesian linear regression," Tech. Rep., MIT, 2000.