Bayesian Separation of Non-Stationary Mixtures of Dependent Gaussian Sources†

Deniz Gençağa[1,‡], Ercan E. Kuruoğlu[2], Ayşın Ertüzün[1]

(1) Boğaziçi University, Electrical and Electronic Engineering Department,Bebek,34342, İstanbul, Turkey
(2) ISTI, Consiglio Nazionale delle Ricerche, via G. Moruzzi 1, 56124, Pisa, Italy

Abstract. In this work, we propose a novel approach to perform Dependent Component Analysis (DCA). DCA can be thought as the separation of latent, *dependent* sources from their observed mixtures which is a more realistic model than Independent Component Analysis (ICA) where the sources are assumed to be independent. In general, the sources can be spatio-temporally dependent and the mixing system may be non-stationary. Here, we propose a DCA algorithm, that combines concepts of particle filters and Markov Chain Monte Carlo (MCMC) methods in order to separate *non-stationary mixtures* of spatially *dependent* Gaussian sources.

Keywords: Bayesian Source Separation, Particle Filters, Markov Chain Monte Carlo.

INTRODUCTION

Source separation problem has always attracted many researchers from different disciplines, such as telecommunications, biomedicine, audio, speech and astrophysics. This arises as a result of the need to investigate different and relevant properties of the desired signals, which are generally hidden at the mixed and noisy observations. In the last decade, research on source separation was mainly focused on the independence assumption of the sources that are mixed, hence called as the ICA [1]. However, in the physical world, the independence assumption cannot always hold and the dependencies should be taken into account. In literature, there is a limited number of references considering the problem of dependent sources [2-3].

Moreover, in the classical ICA approaches, the separation problem is generally handled as a blind method. However, in physical world, the researchers do have some *a priori* knowledge about their specific problems and can also make use of this information. Therefore, instead of blind processing, this *a priori* information can be exploited [4-5].

In source separation problems, another investigation topic is the stationarity of the signals. If the mixing system is constant over time and the sources are stationary, then MCMC methods, can be applied. Rowe [3] proposes MCMC methods for separating both dependent and independent sources, for such cases.

On the other hand, for non-stationary applications, another Bayesian approach, known as particle filters [6-7], have been widely used for processing the data

† This work is supported partially by TÜBİTAK-CNR project number: 102E027 and partially by Boğaziçi University Scientific Research Fund project number: 04A201.

‡ The researcher was supported by NATO-TÜBİTAK A2 fellowship, throughout his research at ISTI-CNR, Italy.

sequentially, as opposed to the batch nature of the MCMC methods. In order to model the non-stationarities, particle filters have been applied to the separation of *autoregressive* (AR) *sources*, whose AR coefficients are time-varying [8] and the sources are independent at a given time instant. In [9-11], other non-stationary ICA problems are solved by particle filters for *non-Gaussian sources*, where the mixing systems are time-varying.

In this work, we propose a novel method in order to separate two *dependent* Gaussian sources from their *non-stationary* mixtures. In this approach, we used particle filters to estimate the coefficients of mixing matrices and MCMC to separate the sources. In this approach, particle filters can be used for estimating *any* non-stationary mixing matrix. Thus, it is different from the method in [3, Ch.13] where special forms of non-stationary mixing matrices are estimated by MCMC.

PARTICLE FILTERS

Particle filters are used in order to sequentially update *a priori* knowledge about some predetermined state variables by using the observation data. In general, these state variables are the hidden variables in a non-Gaussian and nonlinear state-space modelling system. Such a system can be given by the following equations:

$$\begin{aligned} \boldsymbol{\theta}_t &= f_t(\boldsymbol{\theta}_{t-1}, \mathbf{v}_t) \\ \mathbf{x}_t &= h_t(\boldsymbol{\theta}_t, \mathbf{n}_t) \end{aligned} \tag{1}$$

where $\boldsymbol{\theta}_t$, $\mathbf{x}_t$, $\mathbf{v}_t$ and $\mathbf{n}_t$ represent states, observation, process and observation noises, respectively. Here, the objective is to sequentially obtain the *a posteriori* distribution of the state variables obtained via the observation data gathered up to that time, i.e. $p(\boldsymbol{\theta}_{0:t} \mid \mathbf{x}_{1:t})$. Distributions are approximated in terms of particles as follows [6-7]:

$$p(\boldsymbol{\theta}_{0:t} \mid \mathbf{x}_{1:t}) \approx \sum_{i=1}^{K} w_t^i \delta(\boldsymbol{\theta}_{0:t} - \boldsymbol{\theta}_{0:t}^i) \tag{2}$$

where w_t^i, $\boldsymbol{\theta}_{0:t}^i$, $\delta(.)$ denote the weight, the i^{th} particle and the Kronecker delta operator, respectively. The particles that take place in equation (2) are drawn by a method known as the "Sequential Importance Sampling" [6-7] and the corresponding "Importance Weight" for each of them is denoted by w_t^i , which is defined as follows:

$$w_t^i \propto w_{t-1}^i \frac{p(\mathbf{x}_t \mid \boldsymbol{\theta}_t^i) p(\boldsymbol{\theta}_t^i \mid \boldsymbol{\theta}_{t-1}^i)}{q(\boldsymbol{\theta}_t^i \mid \boldsymbol{\theta}_{0:t-1}^i, \mathbf{x}_{1:t})} \tag{3}$$

where q(.) function is called as the "Importance Function" and drawing samples from this probability density function (pdf) is easier than that of original distribution [6,7].

MCMC METHOD FOR SEPARATING STATIONARY MIXTURES

In this section, brief background information is presented about the MCMC method that is utilized for the separation of *stationary* mixtures of sources in [3]. This method assumes that the mixing system and the sources are stationary. The following model is used for the mixtures:

$$\mathbf{x}_t = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{s}_t + \mathbf{n}_t \tag{4}$$

where $\mathbf{x}_t, \boldsymbol{\mu}, \boldsymbol{\Lambda}, \mathbf{s}_t, \mathbf{n}_t$ denote the mixture, overall constant mean, mixing matrix, source and the noise matrices, respectively. These matrices are represented as follows: $\mathbf{x}_t = (x_1(t),...,x_p(t))'$, $\boldsymbol{\mu} = (\mu_1,...,\mu_p)'$, $\mathbf{s}_t = (s_1(t),...,s_m(t))'$, $\mathbf{n}_t = (n_1(t),...,n_p(t))'$, $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1', \boldsymbol{\lambda}_2',...,\boldsymbol{\lambda}_p')'$ where (.)' denotes the transposition operator. In this work, the sources are *spatially dependent* Gaussian processes and each source is temporally uncorrelated. Unlike the sources, the additive noise components are both spatially and temporally independent. Because (4) represents a mixture model at a given time instant *t*, a number of observations, say *n*, form the following matrix notation for the estimation from this *n* sampled batch [3]:

$$\mathbf{X} = \mathbf{e}_n\boldsymbol{\mu}' + \mathbf{S}\boldsymbol{\Lambda}' + \mathbf{N} \tag{5}$$

where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2,...,\mathbf{x}_n)'$, $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2,...,\mathbf{s}_n)'$, $\mathbf{N} = (\mathbf{n}_1, \mathbf{n}_2,...,\mathbf{n}_n)'$ and $\mathbf{e}_n$ denote an n-dimensional vector of ones [3]. By augmenting the $\boldsymbol{\mu}$ and the $\mathbf{e}_n$ vectors, (5) is put into a more compact form as follows:

$$\mathbf{X} = \mathbf{Z}\mathbf{C}' + \mathbf{N} \tag{6}$$

where $\mathbf{C} = (\boldsymbol{\mu}, \boldsymbol{\Lambda})$ and $\mathbf{Z} = (\mathbf{e}_n, \mathbf{S})$. Then, the likelihood function is given as follows:

$$p(\mathbf{X} \mid \mathbf{C}, \mathbf{Z}, \boldsymbol{\Psi}) \propto |\boldsymbol{\Psi}|^{-n/2} \exp\left(-\frac{1}{2} tr(\mathbf{X} - \mathbf{Z}\mathbf{C}')\boldsymbol{\Psi}^{-1}(\mathbf{X} - \mathbf{Z}\mathbf{C}')'\right) \tag{7}$$

where $\boldsymbol{\Psi}$ denotes the diagonal covariance matrix of the noise vector and *tr*(.) represents the trace operator. Distribution in (7) is known as the Matrix-Normal distribution [3]. For the model of (6), the following conjugate prior distributions can be used for the model parameters [3], where MN and IW stand for Matrix Normal and Inverted Wishart distributions, respectively. In the following equations, $\mathbf{S}_0, \mathbf{C}_0, \eta, \nu, \mathbf{V}, \mathbf{B}, \mathbf{H}$ denote the hyperparameters, through which, the *a priori* information is exploited.

$$
\begin{aligned}
&p(\mathbf{S},\mathbf{R},\mathbf{C},\mathbf{\Psi}) = p(\mathbf{S}\,|\,\mathbf{R})\,p(\mathbf{R})\,p(\mathbf{C}\,|\,\mathbf{\Psi})\,p(\mathbf{\Psi})\\
&p(\mathbf{S}\,|\,\mathbf{R}) \propto |\mathbf{R}|^{-n/2}\exp\left(-\frac{1}{2}tr(\mathbf{S}-\mathbf{S_0})\mathbf{R}^{-1}(\mathbf{S}-\mathbf{S_0})'\right) \rightarrow MN\\
&p(\mathbf{R}) \propto |\mathbf{R}|^{-\eta/2}\exp\left(-\frac{1}{2}tr\mathbf{R}^{-1}\mathbf{V}\right) \rightarrow IW(\mathbf{V},p,\eta),\\
&p(\mathbf{\Psi}) \propto |\mathbf{\Psi}|^{-\nu/2}\exp\left(-\frac{1}{2}tr\mathbf{\Psi}^{-1}\mathbf{B}\right) \rightarrow IW(\mathbf{B},m,\nu),\\
&p(\mathbf{C}\,|\,\mathbf{\Psi}) \propto |\mathbf{H}|^{-p/2}|\mathbf{\Psi}|^{-(m+1)/2}\exp\left(-\frac{1}{2}tr\mathbf{\Psi}^{-1}(\mathbf{C}-\mathbf{C_0})\mathbf{H}^{-1}(\mathbf{C}-\mathbf{C_0})'\right) \rightarrow MN
\end{aligned}
\tag{8}
$$

Above, **R** denotes the covariance matrix of the sources and it is not constrained to be diagonal and all of its elements are left free, to model the dependencies between the sources. After some algebra, the following posterior conditionals are obtained to be utilized in the Gibbs sampling, which is the preferred MCMC method here [3]:

$$
\begin{aligned}
&p(\mathbf{C}\,|\,\mathbf{S},\mathbf{R},\mathbf{\Psi},\mathbf{X}) \propto \exp\left(-\frac{1}{2}tr\mathbf{\Psi}^{-1}(\mathbf{C}-\tilde{\mathbf{C}})(\mathbf{H}^{-1}+\mathbf{Z}'\mathbf{Z})(\mathbf{C}-\tilde{\mathbf{C}})'\right),\\
&\tilde{\mathbf{C}} = (\mathbf{C_0}\mathbf{H}^{-1}+\mathbf{X}'\mathbf{Z})(\mathbf{H}^{-1}+\mathbf{Z}'\mathbf{Z})^{-1}\\
&p(\mathbf{\Psi}\,|\,\mathbf{S},\mathbf{R},\mathbf{C},\mathbf{X}) \propto |\mathbf{\Psi}|^{-(n+\nu+m+1)/2}\exp\left(-\frac{1}{2}tr\mathbf{\Psi}^{-1}\mathbf{G}\right),\\
&\mathbf{G} = (\mathbf{X}-\mathbf{Z}\mathbf{C}')'(\mathbf{X}-\mathbf{Z}\mathbf{C}')+(\mathbf{C}-\mathbf{C_0})\mathbf{H}^{-1}(\mathbf{C}-\mathbf{C_0})'+\mathbf{B}\\
&p(\mathbf{R}\,|\,\mathbf{C},\mathbf{S},\mathbf{\Psi},\mathbf{X}) \propto |\mathbf{R}|^{-(n+\eta)/2}\exp\left(-\frac{1}{2}tr\mathbf{R}^{-1}\left[(\mathbf{S}-\mathbf{S_0})'(\mathbf{S}-\mathbf{S_0})+\mathbf{V}\right]\right),\\
&p(\mathbf{S}\,|\,\mathbf{C},\mathbf{R},\mathbf{\Psi},\mathbf{X}) \propto \exp\left(-\frac{1}{2}tr(\mathbf{S}-\tilde{\mathbf{S}})(\mathbf{R}^{-1}+\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda})(\mathbf{S}-\tilde{\mathbf{S}})'\right),\\
&\tilde{\mathbf{S}} = \left[\mathbf{S_0}\mathbf{R}^{-1}+(\mathbf{X}-\mathbf{e_n}\boldsymbol{\mu}')\mathbf{\Psi}^{-1}\mathbf{\Lambda}\right](\mathbf{R}^{-1}+\mathbf{\Lambda}'\mathbf{\Psi}^{-1}\mathbf{\Lambda})^{-1}
\end{aligned}
\tag{9}
$$

By cycling through the posterior conditionals given above, Gibbs sampling can be performed in order to estimate the model parameters [3].

THE PROPOSED METHOD

In literature, MCMC techniques are generally used for batch processing, where the model parameters are assumed not to change within the observed data block [12]. That is why the particle filtering methods have been developed in order to make the estimations sequentially in case of non-stationarities. In this work, the objective is to separate dependent Gaussian sources from their mixtures, where the mixing system is *time-varying* unlike the scenario given in the previous section. That is, the elements of the mixing matrix in (5) change over time. So, (5) can be written as follows:

$$\mathbf{X} = \mathbf{e_n}\boldsymbol{\mu}' + \mathbf{S}\boldsymbol{\Lambda_t}' + \mathbf{N} \tag{10}$$

where $\boldsymbol{\Lambda}_t = \left(\boldsymbol{\lambda}_1'(t), \boldsymbol{\lambda}_2'(t), \ldots, \boldsymbol{\lambda}_p'(t)\right)'$. So, we propose to use particle filtering to estimate the time-varying elements of the mixing matrix. Here, it is assumed that there is no constant mean in the mixture, i.e. $\boldsymbol{\mu} = \mathbf{0}$, and we have *a priori* information about the statistics of the sources and the noise, so that we can form *informative priors* for these. Even if all the statistics of the sources and the noise are known *a priori*, separating the time evolution of the mixing matrix and the sources needs two sets of states in the particle filtering model, i.e. $\boldsymbol{\theta}_t^i = \left\{\mathbf{s}_t^i, \tilde{\boldsymbol{\Lambda}}_t^i\right\}$ of (1), at any time instant *t*. Here, $\tilde{\boldsymbol{\Lambda}}_t^i$ is the vector that is formed by concatenating the rows of the matrix $\boldsymbol{\Lambda}_t^i$ one by one. Since particle filter estimates the *joint* pdf given by (2), sources, must be integrated out from this *joint* pdf in order to obtain the *marginal* pdf estimate of the mixing matrix. From (3), it is seen that the importance weights of each particle is expressed in terms of the likelihood function, if *a priori* transitions are used for the states [6-7]. Then, importance weight estimation (3) takes the following form:

$$w_t^i \propto w_{t-1}^i p(\mathbf{x}_t \mid \boldsymbol{\theta}_t^i) \tag{11}$$

To obtain the marginal importance weight of $\tilde{\boldsymbol{\Lambda}}_t^i$, the likelihood function is integrated:

$$p(\mathbf{x}_t \mid \tilde{\boldsymbol{\Lambda}}_t^i) \propto \int p(\mathbf{x}_t \mid \tilde{\boldsymbol{\Lambda}}_t^i, \mathbf{s}_t^i)\, p(\mathbf{s}_t^i)\, d\mathbf{s}_t^i \tag{12}$$

From (10), it is seen that $p(\mathbf{x}_t \mid \tilde{\boldsymbol{\Lambda}}_t^i, \mathbf{s}_t^i)$ conditional density has a Gaussian distribution. Since the sources are also Gaussian distributed, the integration (12) has also a Gaussian distribution. Thus, instead of estimating (12) for each particle, only mean and variance estimations can be found as follows, using Monte Carlo integration:

$$\begin{aligned}
\hat{\boldsymbol{\mu}}_x &= E\left\{\mathbf{x}_t^i \mid \tilde{\boldsymbol{\Lambda}}_t^i\right\} = \int \mathbf{x}_t^i p(\mathbf{x}_t^i \mid \tilde{\boldsymbol{\Lambda}}_t^i) d\mathbf{x}_t^i \approx \frac{1}{K}\sum_{j=1}^{K} (\mathbf{x}_t^i)_j \\
\hat{\boldsymbol{\Sigma}}_x &= E\left\{\left(\mathbf{x}_t^i - E\left\{\mathbf{x}_t^i \mid \tilde{\boldsymbol{\Lambda}}_t^i\right\}\right)^2\right\} = \int \left\{\mathbf{x}_t^i - E\left\{\mathbf{x}_t^i \mid \tilde{\boldsymbol{\Lambda}}_t^i\right\}\right\}^2 p(\mathbf{x}_t^i \mid \tilde{\boldsymbol{\Lambda}}_t^i) d\mathbf{x}_t^i \\
&\approx \frac{1}{K}\sum_{j=1}^{K} \left\{\left(\mathbf{x}_t^i\right)_j - E\left\{\mathbf{x}_t^i \mid \tilde{\boldsymbol{\Lambda}}_t^i\right\}\right\}^2
\end{aligned} \tag{13}$$

where samples $\left(\mathbf{x}_t^i\right)_j$, can be easily drawn from the Gaussian pdf, given by (12) and K is the total number of particles. By these operations, (11) becomes as follows:

$$w_t^i \propto w_{t-1}^i p(\mathbf{x}_t \mid \tilde{\boldsymbol{\Lambda}}_t^i) \tag{14}$$

where $p(\mathbf{x}_t \mid \tilde{\mathbf{\Lambda}}_t^i)$ is Gaussian, whose mean and covariance matrices can be found by (13). In addition, the *a priori* state transition, which is discussed above, can be written as follows:

$$\tilde{\mathbf{\Lambda}}_t^i = \tilde{\mathbf{\Lambda}}_{t-1}^i + \mathbf{v}_t^i \tag{15}$$

where $\mathbf{v}_t \sim N(\mathbf{0},\mathbf{Q})$ and $\mathbf{Q}$ is diagonal. (10) and (15) fits into the general state-space formulation of the particle filtering, shown in (1). Having found the time evolution of the mixing matrix elements by integrating the sources as outlined above, sources also need to be extracted from the mixtures. Had the mixing matrix coefficients been constant over time, the MCMC method given above could have been used for estimating the sources. However, this is not the case in our scenario. Thus, we propose to assume that the elements of the mixing matrix do not change considerably over small blocks of data. In this case, one may use MCMC in these blocks, however the choice of the appropriate prior distributions of the mixing matrix elements arises as a problem. This is due to the requirement of using *informative prior distributions* for the parameters given in (8) [3, pp. 56].

Since we can obtain estimates of these mixing matrix elements by using the particle filtering scheme, which is explained above, we propose to use these estimates in order to form some approximate informative priors for the mixing matrix elements, which are denoted by $\mathbf{C_0}$ in (8) and (9).

Table 1. Pseudo-code

PARTICLE FILTER PART

I. Initialize the particles $\tilde{\mathbf{\Lambda}}_t^i \sim N(\boldsymbol{\mu}_\lambda, \mathbf{P}_\lambda)$, $w_0^i = 1$ for $i = 1,2,...,K$ where $\mathbf{P}_\lambda$ is a diagonal covariance matrix.

II. Draw new samples $\{\boldsymbol{\theta}_t^i\}_{i=1}^K$: $\tilde{\mathbf{\Lambda}}_t^i = \tilde{\mathbf{\Lambda}}_{t-1}^i + \mathbf{v}_t^i$ and $\mathbf{s}_t^i \sim N(\boldsymbol{\mu}_\mathbf{s}, \mathbf{R})$
Hyperparameters: $\mathbf{v}_t^i \sim N(\mathbf{0},\mathbf{Q}), \boldsymbol{\mu}_\mathbf{s} \sim N(\boldsymbol{\mu}_\mathbf{1}, \mathbf{P}_\mathbf{1}), \mathbf{R} \sim IW(\mathbf{V}, p, \eta)$

III. Calculate the importance weights by integrating out the sources:
$w_t^i \propto w_{t-1}^i p(\mathbf{x}_t \mid \tilde{\mathbf{\Lambda}}_t^i)$ where $p(\mathbf{x}_t \mid \tilde{\mathbf{\Lambda}}_t^i) = N(\mathbf{x}_t; \hat{\boldsymbol{\mu}}_x, \hat{\mathbf{\Sigma}}_x)$

$N(\mathbf{x}_t; \hat{\boldsymbol{\mu}}_x, \hat{\mathbf{\Sigma}}_x)$ denotes the evaluation of the observation data $\mathbf{x_t}$ at the Gaussian pdf, whose mean and covariance matrices are given by (13)

IV. Normalize the importance weights: $\tilde{w}_t^i = w_t^i \Big/ \sum_{i=1}^K w_t^i$, $i = 1,...,K$

V. Resample at each iteration from $\{\mathbf{\Lambda}_t^i, \tilde{w}_t^i\}_{i=1}^K$ and make the unnormalized importance weights equal to each other.

VI. Go to Step (II) and repeat.

MCMC PART

I. Estimate the Minimum Mean Square Error (MMSE) estimate of $\tilde{\mathbf{\Lambda}}$ as follows: $\hat{\mathbf{\Lambda}}_t = \int \mathbf{\Lambda}_{0:t}\, p(\mathbf{\Lambda}_{0:t} \mid \mathbf{x}_{1:t})\, d\mathbf{\Lambda}_{0:t} \approx \sum_{i=1}^{K} \mathbf{\Lambda}_{0:t}^{i} \tilde{w}_t^i$

II. For data blocks of size M, calculate the mean of each mixing matrix element (found above) and use that as the mixing matrix prior in the first equation of (9): $\mathbf{\Lambda}_0 = \frac{1}{M}\sum_{i=1}^{M} \hat{\mathbf{\Lambda}}_i \rightarrow \mathbf{C}_0 = (\boldsymbol{\mu}_0, \mathbf{\Lambda}_0)$

For other parameters, use the priors given in (8).

III. Then cycle through the Gibbs iterations by using the posterior conditionals in (9): i) Start with the initial $\mathbf{S}_0, \mathbf{\Psi}_0$, ii) Cycle: $\mathbf{C}_{(l+1)} \sim p(\mathbf{C} \mid \mathbf{S}_{(l)}, \mathbf{R}_{(l)}, \mathbf{\Psi}_{(l)}, \mathbf{X}), \mathbf{\Psi}_{(l+1)} \sim p(\mathbf{\Psi} \mid \mathbf{S}_{(l)}, \mathbf{R}_{(l)}, \mathbf{C}_{(l+1)}, \mathbf{X})$

$\mathbf{R}_{(l+1)} \sim p(\mathbf{R} \mid \mathbf{S}_{(l)}, \mathbf{C}_{(l+1)}, \mathbf{\Psi}_{(l+1)}, \mathbf{X}), \mathbf{S}_{(l+1)} \sim p(\mathbf{S} \mid \mathbf{R}_{(l+1)}, \mathbf{C}_{(l+1)}, \mathbf{\Psi}_{(l+1)}, \mathbf{X})$

iii) Discard the variates of burn in period and estimate the parameters as follows:

$$\mathbf{S} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{S}_l, \mathbf{R} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{R}_l, \mathbf{C} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{C}_l, \mathbf{\Psi} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{\Psi}_l$$

EXPERIMENT

To verify the performance, the following scenario is simulated on the computer:

$$\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 1 & a_1(t) \\ a_2(t) & 1 \end{bmatrix}\begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix} + \begin{bmatrix} n_1(t) \\ n_2(t) \end{bmatrix}, a_1(t) = \cos\left(\frac{\pi t}{64}\right), a_2(t) = \sin\left(\frac{\pi t}{64}\right),$$

$$\begin{bmatrix} s_1(t) \\ s_2(t) \end{bmatrix} \sim N\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 & -0.5 \\ -0.5 & 3 \end{bmatrix}\right), \begin{bmatrix} n_1(t) \\ n_2(t) \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix}\right)$$

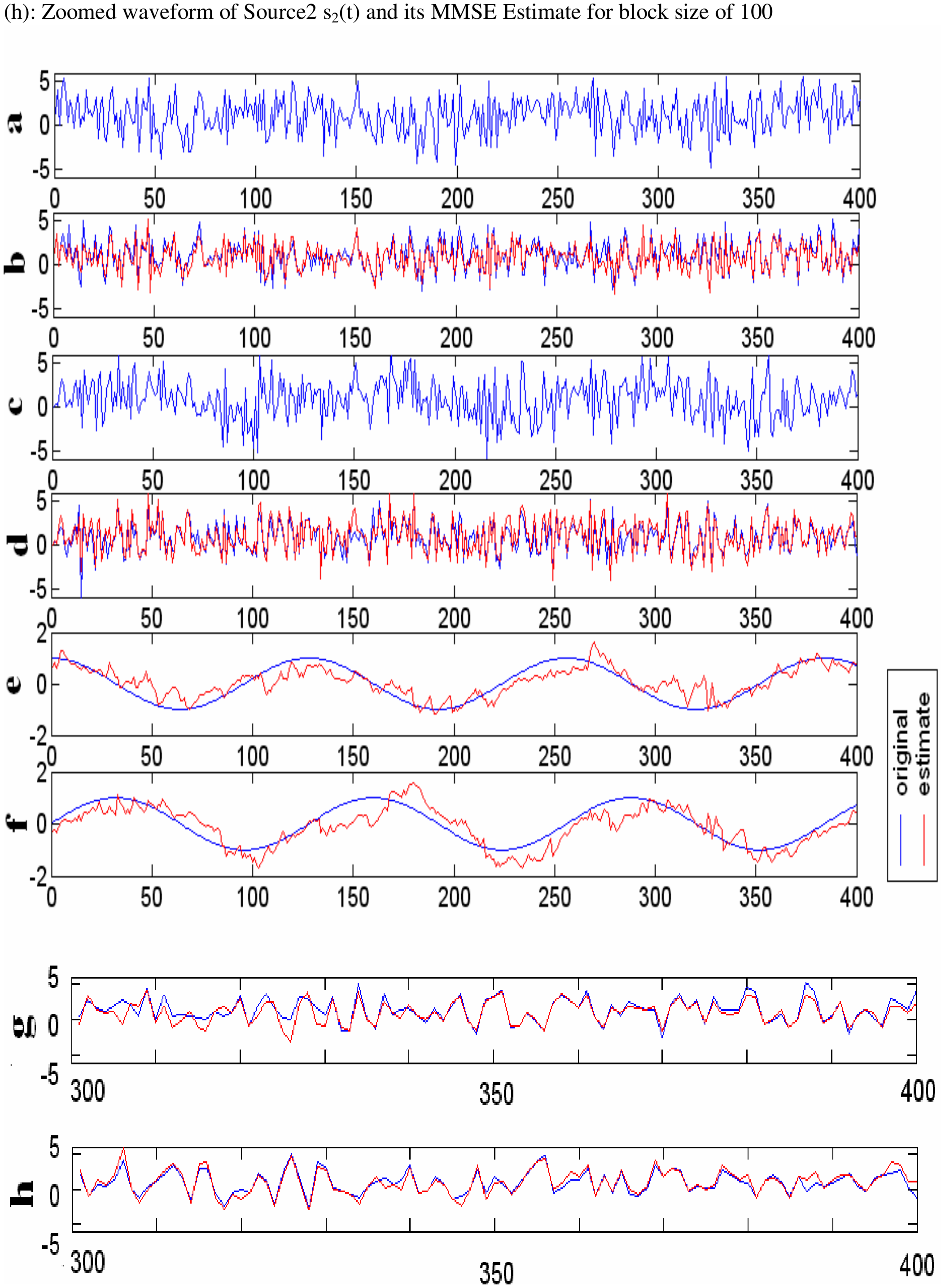
$$hyperparameters: \eta = 1.8\times10^7, \nu = 58, p = 2, m = 2,$$

$$\boldsymbol{\mu}_\lambda = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{P}_\lambda = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, \mathbf{Q} = \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix}, \boldsymbol{\mu}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{P}_1 = \begin{bmatrix} 10^{-5} & 0 \\ 0 & 10^{-5} \end{bmatrix}$$

$$\mathbf{V} = \begin{bmatrix} 5.4\times10^7 & -9\times10^6 \\ -9\times10^6 & 5.4\times10^7 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 2.6 & 0 \\ 0 & 2.6 \end{bmatrix}, \mathbf{H} = \begin{bmatrix} 0.0500 & 0 \\ 0 & 0.0500 \end{bmatrix}$$

By using these informative priors, both the mixing matrix elements and the sources are estimated satisfactorily as shown in Figure 1.

FIGURE 1. (a) Observation1: $x_1(t)$, (b) Source1 $s_1(t)$ and its MMSE Estimate, (c) Observation2: $x_2(t)$, (d) Source2: $s_2(t)$ and its MMSE Estimate, (e) First mixing matrix element and its estimate: $a_1(t)$, (f) Second mixing matrix element and its estimate: $a_2(t)$
(g): Zoomed waveform of Source1 $s_1(t)$ and its MMSE Estimate for block size of 100
(h): Zoomed waveform of Source2 $s_2(t)$ and its MMSE Estimate for block size of 100



CONCLUSIONS AND FUTURE WORK

In this work, we propose a novel method to separate *non-stationary* mixtures of *spatially dependent* Gaussian sources. Here, particle filter is utilized to estimate the non-stationary mixing matrix, while Gibbs sampling is used to extract the sources. By simulations, it is observed that the Signal to Interference Ratio is raised approximately to 6 dB from 1 dB as a result of the separation. The success of the proposed DCA algorithm is a promising result, which can be used in future applications, where the independence assumption of ICA is avoided for more realistic physical problem modelings.

REREFENCES

1. A. Hyvarinen, J. Karhunen, E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.
2. A. K. Barros, "The Independence Assumption: Dependent Component Analysis," in *Advances in Independent Component Analysis,* edited by M. Girolami, Springer, 2000.
3. D. B. Rowe, *Multivariate Bayesian Statistics*, *Models for Source Separation and Signal Unmixing*, Chapman & Hall/CRC, 2003.
4. K.H. Knuth, "A Bayesian approach to source separation", *Proceedings of the First International Workshop on Independent Component Analysis and Signal Separation: ICA'99*, Aussois, France, Jan. 1999, pp. 283-288.
5. A. Mohammad-Djafari, "A Bayesian approach to source separation", *19th International Workshop on Maximum Entropy and Bayesian Methods*, Idaho, USA, Aug. 4-8, 1999.
6. A. Doucet, N. de Freitas, N. J. Gordon, *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, 2001.
7. A. Doucet, S. Godsill, C. Andrieu, "On Sequential Monte Carlo sampling methods for Bayesian filtering", *Statistics and Computing* (2000) 10, pp. 197-208.
8. C. Andrieu, S. Godsill, "A particle filter for model based audio source separation", *Proceedings of the International Workshop on Independent Component Analysis and Signal Separation: ICA'2000*, Helsinki, Finland, June. 2000.
9. A. Ahmed, "Signal Separation", Ph.D. Thesis, Cambridge University, 2000.
10. A. Ahmed, C. Andrieu, A. Doucet, P. J. W. Rayner, "On-line non-stationary ICA using mixture models", *ICASSP 2000*, pp. 3148-3151.
11. R. Everson, S. Roberts, "Particle Filters for Non-Stationary ICA" in *Advances in Independent Component Analysis*, edited by., M. Girolami, Springer, 2000.
12. C. P. Robert, G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, 1999.
13. D. B. Rowe, "A Bayesian Approach to Blind Source Separation", *Journal of Interdisciplinary Mathematics*, Vol. 5, No. 1, pp. 49-76.