Bayesian Blind Separation of Generalized Hyperbolic Processes in Noisy and Underdeterminate Mixtures

Hichem Snoussi and Jérôme Idier

***Abstract*—In this paper, we propose a Bayesian sampling solution to the noisy blind separation of generalized hyperbolic signals. Generalized hyperbolic models, introduced by Barndorff–Nielsen in 1977, represent a parametric family able to cover a wide range of real signal distributions. The alternative construction of these distributions as a normal mean variance (continuous) mixture leads to an efficient implementation of the Markov chain Monte Carlo method applied to source separation. The incomplete data structure of the generalized hyperbolic distribution is indeed compatible with the hidden variable nature of the source separation problem. Both overdeterminate and underdeterminate noisy mixtures are solved by the same algorithm without a prewhitening step. Our algorithm involves hyperparameters estimation as well. Therefore, it can be used, independently, to fitting the parameters of the generalized hyperbolic distribution to real data.**

***Index Terms*—Blind source separation, generalized hyperbolic distributions, Gibbs sampling, noisy mixture, underdeterminate mixture.**

I. Introduction

In this paper, we consider the blind source separation problem as the reconstruction of the sources from the noisy linear instantaneous mixture

$$\boldsymbol{x}_t = \boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{n}_t, \quad t = 1, \ldots, T \tag{1}$$

where $\boldsymbol{x}_t, \boldsymbol{s}_t$ and $\boldsymbol{n}_t$ are, respectively, the $(m\times 1)$ observation vector, the $(n\times 1)$ unknown source vector, and the $(m\times 1)$ unknown noise vector at instant t. $\boldsymbol{A}$ is the $(m \times n)$ unknown mixing matrix. m can be lower or greater than n. The challenging aspect of the blind source separation (BSS) problem is the absence of any exact information about the mixing matrix $\boldsymbol{A}$.

Based on independent identically distributed (i.i.d.) source modeling, many proposed algorithms are designed to linearly demixing the observations $\boldsymbol{x}_{1\ldots T}$. The separation principle in these methods is based on the statistical independence of the reconstructed sources [independent component analysis (ICA)] [1]–[3]. However, ICA is designed to efficiently work in the noiseless case. In addition, with the i.i.d. assumption, the separation capability necessarily relies on high-order statistics allowing at most one source to be Gaussian. The noisy case was treated with the maximum likelihood approach using the expectation maximization (EM) algorithm [4]–[6], the sources being modeled by finite Gaussian mixture. The exact implementation of the EM algorithm leads to a high computational cost. Other stochastic EM variants are used in [6] in order to accelerate the algorithm convergence. However, the choice of the number of Gaussian components remains a difficult task and limits the use of the separation method to some particular types of real signals (e.g., audio signals, piecewise homogeneous images [7]).

Putting the i.i.d. assumption aside, source separation can be achieved with second-order statistics. For instance, second-order correlation diversity in the time domain [8], frequency domain [9], or time-frequency domain [10] are successfully used to blindly separate the sources. Nonstationary second-order-based methods are also proposed in [11]–[14]. Stationarity and decorrelated nonstationarity can approximately be seen as dual under Fourier transformation. For instance, based on the circular approximation, it can be shown that a finite sample correlated temporal stationary signal has a Fourier transform with nonstationary decorrelated samples. Recently, a maximum likelihood method has been proposed to separate noisy mixture of Gaussian stationary sources exploiting this temporal/spectral duality [15], [16]. The Gaussian model of sources allows an efficient implementation of the EM algorithm [17].

The original contribution of this work is to efficiently implement a sampling Bayesian solution in the noisy case, the sources being i.i.d. modeled. The Markov chain Monte Carlo (MCMC) sampling algorithm yields an ergodic Markov chain that has the target posterior distribution as its equilibrium distribution in the stationary regime. From this chain, one can build an estimator based on a selected cost function, without being constrained to the maximum a posteriori (or maximum likelihood) estimator as is the case with the EM algorithm. The proposed separating algorithm yields an estimation of the mixing matrix, the parameters of the source distributions, and the noise covariance matrix. The key point is the use of *generalized hyperbolic* (GH) distributions of Barndorff–Nielsen [18]. Their normal mean-variance continuous mixture representation is remarkably compatible with the hidden structure of the source separation problem. Moreover, the same algorithm can be applied to overdeterminate and underdeterminate cases without any prewhitening step. As the underdeterminate case can be solved exploiting the sparsity of sources [19], the GH distributions represent a well appropriate parametric model for sources able to capture their heavy tails and also their skewness. In addition, a different tail behavior or skewness between sources will enhance their statistical diversity and thus ameliorate the separation performance. The method implicitly

Manuscript received December 18, 2004; revised September 6, 2005. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Vikram Krishnamurthy.

H. Snoussi is with ISTIT/M2S, University of Technology of Troyes, 10000 Troyes, France (e-mail: snoussi@utt.fr).

J. Idier is with IRCCyN, Institut de Recherche en Communications et Cybernétiques de Nantes, Ecole Centrale de Nantes, 44321 Nantes, France (e-mail: idier@irccyn.ec-nantes.fr).

Digital Object Identifier 10.1109/TSP.2006.877660

1053-587X/$20.00 © 2006 IEEE

incorporates a denoising procedure, and it is consequently robust to high-level noise. The double interpretation of this statistical modeling as stationary non-Gaussian and Gaussian nonstationary gives a new insight into the unification of the use of nonstationary second-order statistics and stationary higher order statistics to solve the blind source separation problem. In addition, this leads to an efficient Bayesian Gibbs sampling implementation as the conditionals of the sources and the mixing matrix are Gaussian. To this extent, we obtain a generalization of the finite Gaussian mixture modeling while preserving the benefit of normal conditioning in the Gibbs sampling solution. This work also generalizes the Gibbs separating algorithm in [20], where sources are modeled by t-Student distributions as they are a particular class of the generalized hyperbolic modeling.

This paper is organized as follows. Section II is devoted to generalized hyperbolic processes and their properties. In this section, we present an original Bayesian algorithm to fit the parameters of the generalized hyperbolic distribution to an observed finite time series sample. In Section III, we present the Bayesian blind source separation algorithm in the noisy case. In Section IV, some simulation results corroborating the efficiency of the proposed algorithm are presented.

II. Generalized Hyperbolic Processes

A. Description and Properties

In this paragraph, we briefly describe the generalized hyperbolic distributions and their main properties (for more details refer to Barndorff–Nielsen's original work [18] or Bibby and Sorensen [21]). Generalized hyperbolic distributions form a five-parameter family $\mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu)$ introduced by Barndorff–Nielsen. If the random variable X follows the distribution $\mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu)$, then its probability density function reads

$$\frac{(\gamma/\delta)^{\lambda}}{\sqrt{2\pi}K_{\lambda}(\delta\gamma)} \cdot \frac{K_{\lambda-\frac{1}{2}}(\alpha\sqrt{\delta^2+(x-\mu)^2})}{(\sqrt{\delta^2+(x-\mu)^2}/\alpha)^{\frac{1}{2}-\lambda}} \cdot e^{\beta(x-\mu)}, \quad x \in \mathbb{R} \tag{2}$$

where $\gamma^2 = \alpha^2 - \beta^2$ and $K_{\lambda}(\cdot)$ is the modified Bessel function of third kind

$$K_{\lambda}(y) = \frac{1}{2}\int_0^{\infty} u^{\lambda-1} e^{-\frac{1}{2}y(u+u^{-1})}\, du.$$

The validity parameter domain is as follows:

$$\lambda \in \mathbb{R}, \quad \begin{cases} \delta \geq 0, \alpha > 0, \alpha^2 > \beta^2, & \text{for } \lambda > 0 \\ \delta > 0, \alpha > 0, \alpha^2 > \beta^2, & \text{for } \lambda = 0 \\ \delta > 0, \alpha \geq 0, \alpha^2 \geq \beta^2, & \text{for } \lambda < 0 \end{cases}.$$

GH distributions enjoy the property of being invariant under affine transformations. If $X \sim \mathrm{GH}(\lambda, \alpha, \beta, \delta, \mu)$, then the random variable $aX + b$ follows the distribution $\mathrm{GH}(\lambda, \alpha/\mathrm{a}, \beta/\mathrm{a}, \mathrm{a}\delta, \mathrm{a}\mu + b)$. Many known subclasses can be obtained, either by fixing some parameters or by considering limiting cases: $\lambda = 1$ and $\lambda = -1/2$, respectively, yield the hyperbolic and the NIG distributions (the latter being closed under convolution); $\lambda = 1$ with $\delta \to 0$ provides the asymmetric Laplace distribution; $\lambda = -1/2$ with $\alpha \to 0$ corresponds to the Cauchy distribution; the asymmetric scaled t-distribution is obtained for $\alpha = |\beta|$, etc. Thus, varying the parameters of the GH distributions yields a wide range of tail behaviors, from Gaussian tails to the heavy tails of the Student t-distributions. Fig. 1 depicts examples of GH distributions. One can note that a wide range of tail behaviors is covered and the possibility of modeling the distribution asymmetry (via the parameter β).

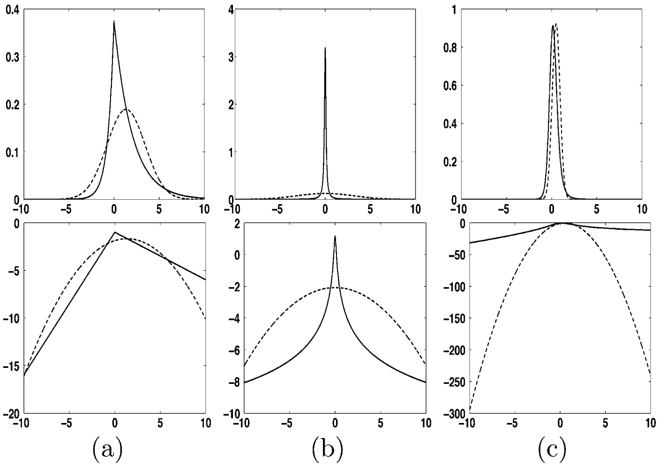

Fig. 1. Examples of the GH distributions. (a) Hyperbolic case: $\lambda = 1, \alpha = 1, \beta = .5, \delta = .001, \mu = 0$. (b) Cauchy case: $\lambda = -.5, \alpha = .01, \beta = .001, \delta = .01, \mu = 0$. (c) Student case: $\lambda = 3, \alpha = 1, \beta = 1, \delta = 1, \mu = 0$. PDFs appear on top row, log densities on bottom row. The dashed line corresponds to the Gaussian distribution with the same mean and variance.

An important feature of the GH distribution is its expression as a continuous normal mean-variance mixture

$$\mathrm{GH}(\mathrm{x}; \lambda, \alpha, \beta, \delta, \mu) = \int_0^{\infty} \mathcal{N}(\mathrm{x}; \mu + \beta\mathrm{w}, \mathrm{w})\, \mathrm{GIG}(\mathrm{w}; \lambda, \gamma, \delta)\, \mathrm{dw} \tag{3}$$

where the variance W of each Gaussian component follows a generalized inverse Gaussian (GIG) distribution

$$\begin{cases} \mathrm{GIG}(w; \lambda, \gamma, \delta) = \frac{(\gamma/\delta)^{\lambda}}{2K_{\lambda}(\delta\gamma)} \cdot w^{\lambda-1} \cdot \exp\left[-\frac{1}{2}(\delta^2 w^{-1} + \gamma^2 w)\right]. \\ w > 0 \end{cases}$$

In other words, the generalized hyperbolic process can be seen as a double stochastic process.

1) First generate[1] $W \sim \mathrm{GIG}(\lambda, \gamma, \delta)$.
2) Then generate $X \sim \mathcal{N}(\mu + \beta W, W)$.

The normal mean-variance expression of the GH distribution will be a key point both in the estimation of the parameters and when incorporating this modeling in the blind source separation problem.

B. Parameter Estimation

Based on an i.i.d. GH sample $\{x_i\}_{i=1\dots N}$, the estimation of the parameters $(\lambda, \alpha, \beta, \delta, \mu)$ is a difficult task. As reported in [22] and [23], this difficulty is essentially due to the flatness of the likelihood with respect to the parameters and particularly

[1] Among the Matlab files freely available from the first author, the program rGIG.m efficiently simulates a GIG random variable.

with respect to the parameter λ. For instance, Barndorff–Nielsen gives an example in [24] where an hyperbolic distribution ($\lambda = 1$) is almost identical to an NIG distribution ($\lambda = -1/2$). Consequently, using standard optimization methods as gradient-like algorithms fails to solve the inference problem as the gradient is too small. Therefore, most of the previous contributions are restricted to a given subclass (typically, λ is set constant). Blaesild and Sorensen propose in [25] the algorithm "hyp" to estimate the parameters of hyperbolic distributions. However, in addition to the restriction $\lambda = 1$, this method suffers from a high computational cost in the multivariate case as was reported by Prause [23], who proposed another restriction to symmetric distributions ($\beta = 0$) able to handle the multivariate case. Recently, Protassov [26] exploits the incomplete data structure of the problem (3) to propose an EM algorithm [17] allowing an efficient implementation in the multivariate case. The EM algorithm is, however, restricted to work within the subclass of NIG distribution ($\lambda = -1/2$). A Bayesian sampling solution is proposed in [27] and [28] for the case of NIG distribution. However, the proposed algorithm cannot be applied for other values of λ. In this paper, we propose an original contribution to Generalized Hyperbolic parameter estimation, without restrictions, exploiting the latent problem structure and based on Gibbs sampling. We propose a new reparametrization in order to sample efficiently the conditionals. The proposed algorithm outperforms the EM algorithm of Protassov [26] in several respects.

1) The conditional sampling steps can be incorporated in more general problems in a hierarchical way. For instance, when solving the blind source separation problem in the next section, the parameters of the sources models are updated, through iterations, according to the same conditionals as in this paragraph.
2) It has the possibility to avoid local maxima.
3) At convergence, one can be aware of the inference problem difficulty by plotting the marginal posteriors.
4) Although, in practice, it may still be more convenient to set the value of λ constant, due to the flatness of the likelihood with respect to this parameter, the Bayesian algorithm includes an optional step for the estimation of λ.

In the following, we outline the Gibbs sampling algorithm for estimating the parameters $\boldsymbol{\eta} = (\lambda, \alpha, \beta, \delta, \mu)$ based on an i.i.d. samples $\{x_i\}_{i=1\ldots N}$. The Bayesian solution consists in sampling the a posteriori distribution of the parameter $\boldsymbol{\eta}$ ($\boldsymbol{\eta}^{(k)} \sim p(\boldsymbol{\eta} \,|\, x_{1\ldots N}), k = 1 \ldots K$). Then, based on the parameter samples $\{\boldsymbol{\eta}^{(k)}\}_{k=1}^{K}$, each estimator $\mathrm{E}[h(\boldsymbol{\eta}) \,|\, x_{1.N}]$ can be approximated by its empirical sum

$$\mathrm{E}[h(\boldsymbol{\eta}) \,|\, x_{1.N}] \approx \frac{1}{K} \sum_{k=1}^{K} h\left(\boldsymbol{\eta}^{(k)}\right). \tag{4}$$

From the Bayes rule

$$p(\boldsymbol{\eta} \,|\, x_{1\ldots N}) \propto p(x_{1\ldots N} \,|\, \boldsymbol{\eta})\, p(\boldsymbol{\eta})$$

one can easily note the difficulty of sampling the posterior because of the complicated form (2) of the likelihood function. However, using the hidden structure of the generalized hyperbolic distribution (3), we can take advantage of the powerful tool of MCMC methods [29]. This can be obtained by implementing a Gibbs sampling algorithm which consists in alternating the sampling of the hidden variances $w_{1\ldots N}$ (given the parameters $\boldsymbol{\eta}$) and the conditional sampling of the parameter of interest $\boldsymbol{\eta}$ (given the variances). The proposed algorithm, called GibbsgHyp, has the following scheme:

$$\left[\begin{array}{l} \text{Set initial values } \boldsymbol{\eta}^{(0)} \text{ and } w_{1\ldots N}^{(0)} \\ \text{Repeat until convergence} \\ \text{1. Sample the variances: } w_{1\ldots N}^{(k)} \sim p\left(w_{1\ldots N} \,\middle|\, x_{1\ldots N}, \boldsymbol{\eta}^{(k-1)}\right) \\ \text{2. Sample the variances: } \boldsymbol{\eta}^{(k)} \sim p\left(\boldsymbol{\eta} \,\middle|\, x_{1\ldots N}, w_{1\ldots N}^{(k)}\right). \end{array}\right. \tag{5}$$

Hence, under weak conditions, the empirical sum $\sum_{k=1}^{K} h(\boldsymbol{\eta}^{(k)})/K$ converges to the expectation $\mathrm{E}[h(\boldsymbol{\eta}) \,|\, x_{1\ldots N}$ as K goes to ∞ (almost sure convergence). The convergence of the empirical sums needs weaker conditions than the convergence of the underlying Markov chain. In fact, the invariance and the irreducibility of the transition kernel are enough to ensure this convergence.

1) Sampling the Variances According to $p(w_{1\ldots N} \,|\, x_{1\ldots N}, \boldsymbol{\eta})$: The first step of the Gibbs algorithm (5) consists in a posteriori sampling the hidden variances. The a posteriori distribution is, according to the Bayes rule

$$\begin{aligned} &p(w_{1\ldots N} \,|\, x_{1\ldots N}, \boldsymbol{\eta}) \\ &\quad \propto p(x_{1\ldots N} \,|\, w_{1\ldots N}, \boldsymbol{\eta})\, p(w_{1\ldots N} \,|\, \boldsymbol{\eta}) \\ &\quad \propto \prod_{i=1}^{N} \mathcal{N}(x_i; \mu + \beta w_i, w_i)\, \mathrm{GIG}(w_i; \lambda, \gamma, \delta) \\ &\quad \propto \prod_{i=1}^{N} \mathrm{GIG}(w_i; \lambda - 1/2, \gamma^2 + \beta^2, \delta^2 + (x_i - \mu)^2) \end{aligned} \tag{6}$$

where we note that the GIG density plays the role of a conjugate prior (the a posteriori density remains in the family of the a priori distribution). The sampling of the variances relies then on the efficient sampling of the GIG distribution. This is performed by the ratio method, which is an exact rejection sampling method based on the calculation of $\sup_z \sqrt{f(z)}$ and $\sup_z |z|\sqrt{f(z)}$, where $f(\,\cdot\,)$ is a function proportional to the density to sample from. For details of this method and its application to sampling the GIG distribution, see Appendix VI.

Remark 1: In the particular case of Student t-distributions ($\gamma = \beta = 0$), the variances are also a posteriori distributed according to an inverse gamma distribution

$$\forall n = 1 \ldots N, \quad \left\{\begin{array}{l} \text{1. Sample } u_n \sim \mathrm{Gamma}(1/2 - \lambda, 1) \\ \text{2. } v_n = u_n \frac{2}{\delta^2 + (x_n - \mu)^2} \\ \text{3. } w_n = 1/v_n \end{array}\right. .$$

2) Sampling the Parameters According to $p(\boldsymbol{\eta} \,|\, x_{1\ldots N}, w_{1\ldots N})$: The second step of the Gibbs algorithm (5) consists in sampling the parameter $\boldsymbol{\eta}$ according to its conditional a posteriori distribution $p(\boldsymbol{\eta} \,|\, x_{1\ldots N}, w_{1\ldots N})$, which reads

$$\begin{aligned} & p(\boldsymbol{\eta} \,|\, x_{1\ldots N}, w_{1\ldots N}) \\ & \quad \propto p(x_{1\ldots N}, w_{1\ldots N} \,|\, \boldsymbol{\eta})\, p(\boldsymbol{\eta}) \\ & \quad \propto p(x_{1\ldots N} \,|\, w_{1\ldots N}, \mu, \beta)\, p(w_{1\ldots N} \,|\, \lambda, \gamma, \delta) p(\boldsymbol{\eta}) \end{aligned} \tag{7}$$

where $p(\boldsymbol{\eta})$ is the a priori distribution of the parameter $\boldsymbol{\eta}$. A key point in the proposed Gibbs sampling is the reparametrization of the hyperbolic distribution $\boldsymbol{\xi} = \phi(\boldsymbol{\eta}) = (\lambda, a, b, \beta, \mu)$:

$$\begin{bmatrix} \xi_1 = \lambda = \eta_1 \\ \xi_2 = a = \gamma/\delta = \sqrt{\eta_2^2 - \eta_3^2} \Big/ \eta_4 \\ \xi_3 = b = \gamma\delta = \sqrt{\eta_2^2 - \eta_3^2}\, \eta_4 \\ \xi_4 = \beta = \eta_3 \\ \xi_5 = \mu = \eta_5 \end{bmatrix} \Leftrightarrow \begin{bmatrix} \eta_1 = \xi_1 \\ \eta_2 = \sqrt{\xi_2 \xi_3 + \xi_4^2} \\ \eta_3 = \xi_4 \\ \eta_4 = \sqrt{\xi_3/\xi_2} \\ \eta_5 = \xi_5 \end{bmatrix}.$$

With the new parametrization, we assume a Gaussian prior for the couple (μ, β), a gamma prior for b and a GIG prior for a given b [28]

$$\begin{cases} \begin{pmatrix} \mu \\ \beta \end{pmatrix} \sim \mathcal{N}(\boldsymbol{m}_p, \boldsymbol{R}_p) \\ b \sim \text{Gamma}(\zeta, \chi) \\ a \,|\, b \sim \text{GIG}(-1/2, \sqrt{b\omega\psi}, \sqrt{b\omega/\psi}) \end{cases}$$

where $(\boldsymbol{m}_p, \boldsymbol{R}_p, \zeta, \chi, \omega, \psi)$ are additional fixed hyperparameters. The gamma and GIG priors ensure the positivity constraint on the parameters a and b. The a posteriori distribution of the parameter $\boldsymbol{\xi}$ becomes

$$\begin{aligned} & p(\boldsymbol{\xi} \,|\, x_{1\ldots N}, w_{1\ldots N}) \\ & \quad \propto p(x_{1\ldots N} \,|\, w_{1\ldots N}, \mu, \beta)\, p(w_{1\ldots N} \,|\, \lambda, a, b) \\ & \quad \propto \mathcal{N}\left(\begin{pmatrix} \mu \\ \beta \end{pmatrix}; \boldsymbol{m}_g, \boldsymbol{R}_g \right) f(\lambda, a, b) \end{aligned} \tag{8}$$

where we note that the posterior is separable into two subvectors (μ, β) and (λ, a, b). The first subvector has a Gaussian distribution with the following mean and covariance:

$$\begin{cases} \boldsymbol{m}_g = \boldsymbol{R}_g(\boldsymbol{R}_d^{-1} \boldsymbol{m}_d + \boldsymbol{R}_p^{-1} \boldsymbol{m}_p) \\ \boldsymbol{R}_g = \left(\boldsymbol{R}_d^{-1} + \boldsymbol{R}_p^{-1} \right)^{-1} \end{cases}$$

where the data-dependent quantities $\boldsymbol{m}_d$ and $\boldsymbol{R}_d$ have the following expressions:

$$\boldsymbol{m}_d = \boldsymbol{R}_d \begin{pmatrix} \sum x_i / w_i \\ \sum x_i \end{pmatrix}, \quad \boldsymbol{R}_d^{-1} = N \begin{pmatrix} \frac{\sum w_i^{-1}}{N} & 1 \\ 1 & \frac{\sum w_i}{N} \end{pmatrix}$$

and hence the couple (μ, β) can be exactly sampled. However, to sample the distribution $f(\,\cdot\,)$ of the second subvector, we need another Gibbs cycle of the three parameters (λ, a, b). In other words, we adopt the following scheme:

$$\begin{aligned} & 1.\ \lambda \,|\, a, b \sim f(\lambda) \propto \frac{\left(a^N \prod w_i\right)^{\lambda}}{K_\lambda(b)^N} \\ & 2.\ a \,|\, \lambda, b \sim \text{GIG}\left(N\lambda - 1/2, \sqrt{b\left(\sum w_i + \omega\psi\right)} \right. \\ & \qquad \left. \sqrt{b\left(\sum w_i^{-1} + \omega/\psi\right)} \right) \\ & 3.\ b \,|\, \lambda, a \sim f(b) \propto \frac{b^{\zeta-1} e^{-\frac{1}{2} b \left(a \sum w_i + \frac{1}{a} \sum w_i^{-1} + 2\chi\right)}}{K_\lambda(b)^N} \end{aligned} \tag{9}$$

where the sampling of the parameters λ and b is performed by the ratio method as in the GIG case.

III. Bayesian Blind Separation

In this section, we assume that n sources modeled by the generalized hyperbolic distribution (2) are indirectly observed. The collected data are a noisy linear mixture of the sources. The forward model of the observation process can be cast in the simple matrix form (1) or, equivalently

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} + \boldsymbol{N}$$

where the $(m \times T)$-matrix $\boldsymbol{X}$ contains the m observed rows, the $(n \times T)$-matrix $\boldsymbol{S}$ contains the n unobserved source rows, and $\boldsymbol{N}$ is the noise corrupting the observations. We assume that each source row $\boldsymbol{s}_j = (s_j(1), \ldots, s_j(T))$ follows a generalized hyperbolic distribution $H(\lambda_j, \alpha_j, \beta_j, \delta_j, \mu_j)$ and that each noise row $\boldsymbol{n}_j = (n_j(1), \ldots, n_j(T))$ is white and Gaussian with a variance σ_j^2 (i.e., at time t, the noise covariance is $\boldsymbol{R}_n = \text{diag}(\sigma_1^2, \ldots, \sigma_m^2)$). This source model, in the context of the blind source separation problem, is mainly motivated by the flexibility and the normal mean-variance mixture form of the GH distribution. In fact, as will be seen later in this section, the hidden structure of the normal mixture is compatible with the BSS structure, yielding an efficient implementation of the Gibbs sampling algorithm. In addition, the GH model is able to capture both the heavy tails and the asymmetry of the sources. Thus, it provides an implicit way to exploit the sparsity and skewness of the sources and enhance their statistical diversity. The identification problem is very ill posed as the $(m \times n)$-mixing matrix, the sources $\boldsymbol{S}$, and their corresponding hyperparameters $\boldsymbol{\eta} = (\lambda_j, \alpha_j, \beta_j, \delta_j, \mu_j)_{j=1}^n$ are unknown.

The Bayesian formulation is adapted to this ill-posed problem as it consistently takes the structure of the observation process into account. The noise is indeed modeled in the inference process and any additional prior information can be incorporated. Given the observations $\boldsymbol{X}$, the a posteriori distribution of the unknowns $\boldsymbol{\theta} = (\boldsymbol{A}, \boldsymbol{R}_n, \boldsymbol{S}, \boldsymbol{\eta})$, according to the Bayesian rule, is

$$(\boldsymbol{\theta} \,|\, \boldsymbol{X}, \mathcal{I}) \propto p(\boldsymbol{X} \,|\, \boldsymbol{\theta}, \mathcal{I}) p(\boldsymbol{\theta} \,|\, \mathcal{I}) \tag{10}$$

where $\mathcal{I}$ contains the prior information such as the noisy mixture, the generalized hyperbolic density of sources, the whiteness of the noise, and so forth.

In general, (10) yields a complicated non linear function of the parameters to estimate. However, the Bayesian sampling tool is efficient to deal with the challenging inferential task. For instance, the Gibbs sampling is appropriate for the separation problem. It produces a Markov chain $\tilde{\theta}^{(k)} = (\tilde{\boldsymbol{A}}^{(k)}, \tilde{\boldsymbol{R}}_n{}^{(k)}, \tilde{\boldsymbol{S}}^{(k)}, \tilde{\boldsymbol{W}}^{(k)}, \tilde{\boldsymbol{\eta}}^{(k)})$, which converges, in distribution, to the target a posteriori (10). This can be seen when considering the sources $\boldsymbol{S}$ as the missing variables to estimate the parameters $(\boldsymbol{A}, \boldsymbol{R}_n)$ and that both the sources $\boldsymbol{S}$ and the hidden variances $\boldsymbol{W}$ are the missing data for the estimation of the hyperparameters $\boldsymbol{\eta}$. Thus, we have a data augmentation problem with two missing variables shells. The formulation of the generalized hyperbolic density as a continuous mean-variance normal mixture leads to an efficient implementation of the Gibbs sampling as the conditioning of the sources is Gaussian and that of the hyperparameters is implementable with the ratio method. In the following, we outline the Gibbs sampling scheme for the source separation problem.

A. Gibbs Algorithm

The cyclic sampling steps are as follows:

$$\begin{aligned}
&1.\ \text{Sample } \tilde{\boldsymbol{S}} \sim p(\boldsymbol{S} \mid \boldsymbol{X}, \tilde{\boldsymbol{A}}, \tilde{\boldsymbol{R}}_n, \tilde{\boldsymbol{W}}, \tilde{\boldsymbol{\eta}})\\
&2.\ \text{Sample } \tilde{\boldsymbol{W}} \sim p(\boldsymbol{W} \mid \tilde{\boldsymbol{S}}, \tilde{\boldsymbol{\eta}})\\
&3.\ \text{Sample } \tilde{\boldsymbol{\eta}} \sim p(\boldsymbol{\eta} \mid \tilde{\boldsymbol{S}}, \tilde{\boldsymbol{W}})\\
&4.\ \text{Sample } (\tilde{\boldsymbol{A}}, \tilde{\boldsymbol{R}}_n) \sim p(\boldsymbol{A}, \boldsymbol{R}_n \mid \boldsymbol{X}, \tilde{\boldsymbol{S}}).
\end{aligned} \tag{11}$$

1) Sampling the Sources: Given the data $\boldsymbol{X}$ and the remaining components of $\boldsymbol{\theta}$, the sources are temporally independent and their a posteriori distribution is multivariate Gaussian, obtained by applying the Bayes rule

$$\begin{aligned}
&p(\boldsymbol{S} \mid \boldsymbol{X}, \tilde{\boldsymbol{\theta}})\\
&\quad \propto \prod_{t=1}^{T} \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{A}\boldsymbol{s}_t, \boldsymbol{R}_n) \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{\mu} + \boldsymbol{\beta} \odot \boldsymbol{w}_t, \operatorname{diag}(\boldsymbol{w}_t))\\
&\quad \propto \prod_{t=1}^{T} \mathcal{N}(\boldsymbol{s}_t; \boldsymbol{\mu}_s(t), \boldsymbol{\Gamma}_s(t))
\end{aligned} \tag{12}$$

where $\odot$ is the element by element multiplication operator. The means and the covariances of the sources at time t have the following expressions:

$$\begin{cases}
\boldsymbol{\Gamma}_s(t) = \left[\boldsymbol{A}^* \boldsymbol{R}_n^{-1} \boldsymbol{A} + \boldsymbol{P}_w^{-1}\right]^{-1}\\
\boldsymbol{\mu}_s(t) = \boldsymbol{\Gamma}_s(t) \left[\boldsymbol{A}^* \boldsymbol{R}_n^{-1} \boldsymbol{x}_t + \boldsymbol{P}_w^{-1}(\boldsymbol{\mu} + \boldsymbol{\beta} \odot \boldsymbol{w}_t)\right]
\end{cases} \tag{13}$$

where $\boldsymbol{P}_w = \operatorname{diag}(\boldsymbol{w}_t)$ is the a priori source covariance and $\boldsymbol{\mu} + \boldsymbol{\beta} \odot \boldsymbol{w}_t$ is the a priori mean.

2) Sampling the Variances and the Hyperparameters: The conditioned sampling in the second and third steps of the Gibbs algorithm (11), given the sampled sources $\tilde{\boldsymbol{S}}$, are the same as in the previous Section II. In fact, given the sources, the variances $\boldsymbol{W}$ and the hyperparameters $\boldsymbol{\eta}$ are independent of the data $\boldsymbol{X}$, as they are not related to the mixing process. They are spatially independent and for each component $j = 1 \ldots n$, the variances row $\boldsymbol{w}_j$ is sampled according to a GIG distribution as in (6) and the hyperparameters $\eta_j = (\lambda_j, \alpha_j, \beta_j, \delta_j, \mu_j)$ are sampled according to the distributions (8) and (9).

Remark 2: The Gibbs sampling scheme (11) can be improved by integrating with respect to the sources $\boldsymbol{S}$ when sampling the variances $\boldsymbol{W}$. The a posteriori distribution of the variances is then

$$\begin{aligned}
f(\boldsymbol{W}) &\propto p(\boldsymbol{W} \mid \boldsymbol{X}, \boldsymbol{\eta}, \boldsymbol{A}, \boldsymbol{R}_n)\\
&\propto p(\boldsymbol{X} \mid \boldsymbol{W}, \boldsymbol{A}, \boldsymbol{R}_n) p(\boldsymbol{W} \mid \boldsymbol{\eta})\\
&\propto \int p(\boldsymbol{X}, \boldsymbol{S} \mid \boldsymbol{W}, \boldsymbol{A}, \boldsymbol{R}_n)\, d\boldsymbol{S}\, p(\boldsymbol{W} \mid \boldsymbol{\eta})\\
&\propto \prod_{t=1}^{T} \mathcal{N}(\boldsymbol{x}_t; \boldsymbol{A}(\boldsymbol{\mu} + \boldsymbol{\beta} \odot \boldsymbol{w}_t), \boldsymbol{A}\boldsymbol{P}_w\boldsymbol{A}^* + \boldsymbol{R}_n)\\
&\quad \times p(\boldsymbol{W} \mid \boldsymbol{\eta})
\end{aligned}$$

where $\boldsymbol{P}_w$ is the diagonal covariance $\operatorname{diag}(\boldsymbol{w}_t)$. As an exact sampling procedure for this distribution is not available, we can implement a hybrid version of Gibbs/Hasting-Metropolis version where the instrumental distribution $g(\,\cdot\,)$ is the first one proposed in the Gibbs algorithm (11), that is

$$\begin{aligned}
g(w_j(t)) = \mathrm{GIG}(w_j(t); \lambda_j - 1/2, \gamma_j^2 + \beta_j^2, \delta_j^2\\
+ (s_j(t) - \mu_j)^2).
\end{aligned}$$

The new hybrid version has the following scheme:

$$\begin{aligned}
&1.\ \text{Sample } \tilde{\boldsymbol{W}} \sim g(\boldsymbol{W})\\
&\qquad \text{accept } \tilde{\boldsymbol{W}} \text{ with probability } \rho\\
&\qquad\qquad = \min\left(1, \frac{\mathrm{f}(\tilde{\boldsymbol{W}})\mathrm{g}\left(\boldsymbol{W}^{(\mathrm{k}-1)}\right)}{\mathrm{g}(\tilde{\boldsymbol{W}})\mathrm{f}\left(\boldsymbol{W}^{(\mathrm{k}-1)}\right)}\right)\\
&2.\ \text{Sample } \boldsymbol{S} \sim p(\boldsymbol{S} \mid \boldsymbol{X}, \tilde{\boldsymbol{A}}, \tilde{\boldsymbol{R}}_n, \tilde{\boldsymbol{W}}, \tilde{\boldsymbol{\eta}})\\
&3.\ \text{Sample } \boldsymbol{\eta} \sim p(\boldsymbol{\eta} \mid \tilde{\boldsymbol{S}}, \tilde{\boldsymbol{W}})\\
&4.\ \text{Sample } (\boldsymbol{A}, \boldsymbol{R}_n) \sim p(\boldsymbol{A}, \boldsymbol{R}_n \mid \boldsymbol{X}, \tilde{\boldsymbol{S}}).
\end{aligned} \tag{14}$$

The gain of performance of the hybrid version (14) with respect to the first Gibbs scheme (11) can be shown when considering the effective number of subvectors in the Gibbs sampling cycle. For instance, in the scheme (11), the number of subvectors is three: 1) sample $\boldsymbol{S}$, 2) sample $\boldsymbol{\eta}$, 3) sample $(\boldsymbol{W}, \boldsymbol{A}, \boldsymbol{R}_n)$ [steps 2 and 4 are independent]. However, in the hybrid version, assuming the first step is exact, the number of effective subvectors is onlytwo: 1) sample $(\boldsymbol{W}, \boldsymbol{S})$ and 2) sample $(\boldsymbol{\eta}, \boldsymbol{A}, \boldsymbol{R}_n)$.

3) Sampling the Parameters $\boldsymbol{A}$ and $\boldsymbol{R}_n$: The sampling of the mixing matrix and the covariance matrix given the data and the sources is the same as in [7]. For sake of completeness, we report hereafter the sampling distributions. For a Jeffrey prior (see [30] for details of Fisher matrix computation)

$$p\left(\boldsymbol{A}, \boldsymbol{R}_n^{-1}\right) \propto \left|\boldsymbol{R}_n^{-1}\right|^{\frac{n-(m+1)}{2}}$$

the a posteriori distribution of $(\boldsymbol{A}, \boldsymbol{R}_n^{-1})$ is Normal–Wishart

$$p\left(\boldsymbol{A}, \boldsymbol{R}_n^{-1}\right) = \mathcal{N}(\boldsymbol{A}; \boldsymbol{A}_p, \boldsymbol{\Gamma}_a)\mathcal{W}_m\left(\boldsymbol{R}_n^{-1}; \nu_p, \boldsymbol{\Sigma}_p\right)$$

with parameters

$$\begin{cases} \boldsymbol{A}_p = \boldsymbol{R}_{xs}\boldsymbol{R}_{ss}^{-1} \\ \boldsymbol{\Gamma}_a = \frac{1}{T}\boldsymbol{R}_{ss}^{-1} \otimes \boldsymbol{R}_n \end{cases}$$
$$\begin{cases} \nu_p = T - n \\ \boldsymbol{\Sigma}_p = \frac{T}{T-n}\left(\boldsymbol{R}_{xx} - \boldsymbol{R}_{xs}\boldsymbol{R}_{ss}^{-1}\boldsymbol{R}_{sx}\right) \end{cases}$$

where $\otimes$ is the Kronecker product [31] and $\boldsymbol{R}_{ab}$ denotes the empirical covariance between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ for T samples

$$\boldsymbol{R}_{ab} = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{a}_t \boldsymbol{b}_t^*.$$

Remark 3 (Overrelaxation): The covariance $\boldsymbol{\Gamma}_p$ of the mixing matrix is inversely proportional to the signal-to-noise ratio (SNR). Thus, in the case of high SNR, the covariance is very small leading to a slow convergence of the Markov chain. In other words, the conditional distribution of the mixing matrix is very picky around a mean value depending on the sampled sources due to a high correlation with the latter. The Markov chain is then unable to efficiently explore the parameter domain. To alleviate this problem, a general solution proposed by Adler [32] for the Gaussian case (and reconsidered by Neal [33] for more general conditional distributions) consists in overrelaxing the chain by introducing a negative correlation between the updates. If the parameter to be updated $\boldsymbol{\theta}$ has a Gaussian distribution $\mathcal{N}(\boldsymbol{m}, \boldsymbol{L}\boldsymbol{L}^*)$, the retained value at iteration k is the following:

$$\boldsymbol{\theta}^{(k)} = \boldsymbol{m} + \alpha(\boldsymbol{m} - \boldsymbol{\theta}^{(k-1)}) + \sqrt{1-\alpha^2}\boldsymbol{L}\boldsymbol{u}$$

where $\boldsymbol{u}$ is a standard Gaussian vector and α has a value in the interval [0, 1[controlling the overrelaxation degree. The modified Markov chain is still ergodic and has the posterior as its equilibrium distribution [33]. In the BSS algorithm (11), we only modify the update steps 1 and 4 corresponding to the conditional Gaussian sampling of the sources and the mixing matrix. The remaining non-Gaussian steps are kept unchanged.

Remark 4: In its algorithmic aspect, the separating method relies on matching the empirical data covariance $\boldsymbol{R}_{xx}$ to its theoretical expression $\boldsymbol{A}\boldsymbol{P}_w\boldsymbol{A}^T + \boldsymbol{R}_n$, where $\boldsymbol{P}_w = \mathrm{diag}(\boldsymbol{w}_t)$ is the nonstationary source covariance simultaneously updated through the Gibbs iterations. This represents an unification between the use of higher order statistics and nonstationary second-order statistics. In fact, virtually introducing the hidden variances $\boldsymbol{w}_{1\ldots T}$ makes the sources Gaussian but not stationary. The log-likelihood is then the Kullback–Leibler divergence between the matrices $\boldsymbol{R}_{xx}$ and $\boldsymbol{A}\boldsymbol{P}_w\boldsymbol{A}^T + \boldsymbol{R}_n$. As the variances are sampled through the Gibbs iterations, the effective distribution of the sources, at convergence, is temporally stationary but not Gaussian.

B. Source Estimation

In the noisy mixture case, the estimation of the mixing matrix is not equivalent to the estimation of the sources. The matrix $\hat{\boldsymbol{A}}^{-1}$ is not a separating matrix. In other words, the signals $\hat{\boldsymbol{A}}^{-1}\boldsymbol{X}$ are not a consistent estimate of the sources. However, the Bayesian framework allows an efficient consistent joint estimation of the sources. After convergence in distribution (after a burn in period k_0), the Gibbs algorithm (11) yields a sequence $(\tilde{\boldsymbol{A}}^{(k)}, \tilde{\boldsymbol{R}}_n^{(k)}, \tilde{\boldsymbol{S}}^{(k)}, \tilde{\boldsymbol{W}}^{(k)}, \tilde{\boldsymbol{\eta}}^{(k)})_{k \geq k_0}$ distributed according to its a posteriori $p(\boldsymbol{A}, \boldsymbol{R}_n, \boldsymbol{S}, \boldsymbol{W}, \boldsymbol{\eta} \,|\, \boldsymbol{X}, \mathcal{I})$. Based on this sequence, two estimates can be simply obtained: the posterior mean (PM) minimizing the expected quadratic loss and the conditional posterior mean minimizing the conditional quadratic loss.

1) Posterior Mean: The expected quadratic loss is the posterior mean of the quadratic error and is defined as follows:

$$\begin{aligned}\mathcal{C}(\boldsymbol{S}) &= \mathrm{E}[(\boldsymbol{S} - \boldsymbol{S}^*)^2 \,|\, \boldsymbol{X}, \mathcal{I}] \\ &= \int_{\boldsymbol{S}^*} (\boldsymbol{S} - \boldsymbol{S}^*)^2 p(\boldsymbol{S}^* \,|\, \boldsymbol{X}, \mathcal{I})\, \mathrm{d}\boldsymbol{S}^*.\end{aligned}$$

Minimization of $\mathcal{C}(\boldsymbol{S})$ yields the posterior mean estimate $\hat{\boldsymbol{S}}_{\mathrm{PM}}$ approximated by the empirical mean of the sequence $(\tilde{\boldsymbol{S}}^{(k)})_{k \geq k_0}$

$$\begin{aligned}\hat{\boldsymbol{S}}_{\mathrm{PM}} &= \mathrm{E}[\boldsymbol{S} \,|\, \boldsymbol{X}, \mathcal{I}] \\ &\approx \frac{1}{K}\sum_{k=k_0}^{K} \tilde{\boldsymbol{S}}^{(k)}.\end{aligned}$$

Remark 5: The sequence $(\tilde{\boldsymbol{S}}^{(k)})_{k \geq k_0}$ is obtained through the Gibbs sampling in an augmented variable procedure. The augmented variables are the remaining components of $\boldsymbol{\theta}$ ($\boldsymbol{\theta}_{\backslash \boldsymbol{S}} = (\boldsymbol{A}, \boldsymbol{R}_n, \boldsymbol{W}, \boldsymbol{\eta})$). The variance of the empirical mean estimate can be further reduced by using instead the sequence of conditional expectations of the sources given the sampled parameters $\tilde{\boldsymbol{\theta}}_{\backslash \boldsymbol{S}}^{(k)}$. This is known as the Rao–Blackwell scheme [29]

$$\hat{\boldsymbol{S}}_{\mathrm{RB-PM}} = \frac{1}{K}\sum_{k=k_0}^{K} \mathrm{E}\left[\boldsymbol{S} \,|\, \tilde{\boldsymbol{\theta}}_{\backslash \boldsymbol{S}}^{(k)}, \boldsymbol{X}\right]\Big)_{k \geq k_0}$$

where the conditional expectations are simply obtained from (13).

2) Conditional Posterior Mean: Given an estimate of the mixing matrix, the noise covariance, the hidden variance, and the GH hyperparameters

$$\hat{\boldsymbol{A}} = \frac{1}{K}\sum_{k=k_0}^{K} \tilde{\boldsymbol{A}}^{(k)}; \quad \hat{\boldsymbol{R}}_n = \frac{1}{K}\sum_{k=k_0}^{K} \tilde{\boldsymbol{R}}_n^{(k)};$$
$$\hat{\boldsymbol{W}} = \frac{1}{K}\sum_{k=k_0}^{K} \tilde{\boldsymbol{W}}^{(k)}$$
$$\hat{\boldsymbol{\eta}} = \frac{1}{K}\sum_{k=k_0}^{K} \tilde{\boldsymbol{\eta}}^{(k)}$$

the conditional distribution of the sources is a multivariate Gaussian at each instant t. Therefore, minimizing the expected quadratic loss $\mathcal{C}(\boldsymbol{S}\,|\,\hat{\boldsymbol{\theta}}_{\setminus \boldsymbol{S}})$ is equivalent to maximizing the a posteriori distribution and yields the following linear (with respect to data) estimate:

$$\hat{\boldsymbol{s}}_{\mathrm{CPM}}(t) = \left[\hat{\boldsymbol{A}}^* \hat{\boldsymbol{R}}_n^{-1} \hat{\boldsymbol{A}} + \hat{\boldsymbol{P}}_w^{-1}\right]^{-1} \left[\hat{\boldsymbol{A}}^* \hat{\boldsymbol{R}}_n^{-1} \boldsymbol{x}_t + \hat{\boldsymbol{P}}_w^{-1}(\hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\beta}} \odot \hat{\boldsymbol{w}}_t)\right]$$

where $\hat{\boldsymbol{P}}_w$ is the prior diagonal covariance of the sources varying in time

$$\hat{\boldsymbol{P}}_w = \mathrm{diag}[\hat{\boldsymbol{w}}_t].$$

IV. Simulation Results

A. Parameter Estimation

In this paragraph, we illustrate the performance of the Bayesian sampling algorithm GibbsgHyp (5) proposed in Section II-B. Raw data are observed and our purpose is the estimation of the generalized hyperbolic parameters. Following the example in [27], we consider the estimation of the parameters of an NIG distribution, in which case λ is set to $-1/2$ rather than sampled. A time series of 5000 i.i.d. samples is considered. The true values are $\alpha = 2, \beta = 1, \delta = 1, \mu = 2$. The hyperparameters are fixed as follows: $\boldsymbol{m}_p = (0,0), \boldsymbol{R}_p = \boldsymbol{I}, \zeta = 0.1, \chi = 0.1, \omega = 0.01, \psi = 0.2$. In Fig. 2, the first column illustrates, from top to bottom, the NIG time series, its empirical histogram, and the estimated log-distribution superimposed to the true log-distribution. One can note the heavy tails, the asymmetry of the distribution, and the accuracy of its identification. In the second column, we have plotted the evolution of the parameters Markov chains $\boldsymbol{\eta}^{(k)}$ and in the third column the evolution of their corresponding empirical sums. We note the convergence of the empirical sums near the true values. Table I contains the posterior mean estimates of the GH parameters. In order to quantify this accuracy, it is more meaningful to evaluate the closeness of the estimated distribution to the true distribution in a parametric-free way. In Table I, several divergence measures between densities are computed: Kullback–Leibler divergence, Kolmogorov distance (maximum of the absolute difference between cumulative distributions), and the L1 and L2 distances. We note the performance of the proposed algorithm in terms of the free parametric accuracy evaluation.

In order to illustrate the difficulty of estimating the parameter λ, the likelihood (for the same sample $x_{1\ldots N}$) with respect to λ when the other parameters are fixed

$$\begin{aligned}\mathcal{L}(\lambda) &= p(x_{1\ldots N}\,|\,\lambda, \alpha, \beta, \delta, \mu)\\ &= \prod_{i=1}^{N} \frac{(\gamma/\delta)^{\lambda}}{\sqrt{2\pi}K_{\lambda}(\delta\gamma)} \cdot \frac{K_{\lambda-\frac{1}{2}}(\alpha\sqrt{\delta^2+(x_i-\mu)^2})}{(\sqrt{\delta^2+(x_i-\mu)^2}/\alpha)^{\frac{1}{2}-\lambda}}\\ &\quad \cdot e^{\beta(x_i-\mu)}\end{aligned}$$

is plotted in Fig. 3 for different values of the remaining parameters in $\boldsymbol{\eta}$. As clearly appears in Fig. 3, the likelihood is peaked at a position which is strongly dependent on the values of the other parameters. When the parameters are set to their true values, the likelihood profile is concentrated around its value $(\lambda^* = -0.5)$. However, when the parameters are not set to their true values, the position of the peak is not close to the true parameter.

B. BSS: Low SNR

In this paragraph, we illustrate the performance of the Gibbs separating algorithm on simulated data in a very noisy context. The number of sources is equal to the number of detectors. Three sources are generated according to the GH model (2). The number of samples is 5000. They are artificially mixed by a mixing matrix

$$\boldsymbol{A}^* = \begin{pmatrix} 0.59 & 0.27 & 0.24 \\ 0.59 & 0.82 & 0.12 \\ 0.53 & 0.49 & 0.96 \end{pmatrix}$$

and corrupted by a white noise $(\boldsymbol{R}_n^* = 2\boldsymbol{I})$ such that the SNRs are 0, 4, and 4 dB for the three detectors, respectively. Given only the observed data, the Gibbs algorithm yields a Markov chain $(\tilde{\boldsymbol{A}}^{(k)}, \tilde{\boldsymbol{R}_n}^{(k)}, \tilde{\boldsymbol{S}}^{(k)}, \tilde{\boldsymbol{W}}^{(k)}, \tilde{\boldsymbol{\eta}}^{(k)})_{k\in\mathbb{N}}$ based on which one can obtain the estimation of the mixing matrix, the noise covariance, the sources, and their probability densities. Fig. 4 shows the convergence of the empirical sums of the mixing matrix Markov chains near their true values. The norm of the columns of $\tilde{\boldsymbol{A}}^{(k)}$ is set fixed to one in order to fix the scale indeterminacy of BSS. The empirical mixing matrix mean is

$$\hat{\boldsymbol{A}} = \begin{pmatrix} 0.5761 & 0.2788 & 0.2485 \\ 0.5990 & 0.8223 & 0.1310 \\ 0.5553 & 0.4954 & 0.9591 \end{pmatrix}.$$

Moreover, Fig. 5 shows the evolution of a performance index that evaluates the closeness of the matrix product $\mathcal{P} = \hat{\boldsymbol{A}}^{-1}\boldsymbol{A}^*$ to the identity matrix. Following [34], it is defined by (when $\mathcal{P}$ approaches the identity matrix, the index converges to zero)

$$\frac{1}{2}\left[\sum_i \left(\sum_j \frac{|\mathcal{P}_{ij}|^2}{\max_l |\mathcal{P}_{il}|^2} - 1\right) + \sum_j \left(\sum_i \frac{|\mathcal{P}_{ij}|^2}{\max_l |\mathcal{P}_{lj}|^2} - 1\right)\right].$$

The convergence of the empirical mean of the performance index to -18 dB corroborates the effectiveness of the separating algorithm in estimating the mixing matrix. At convergence, the product $\mathcal{P}$ is

$$\mathcal{P} = \hat{\boldsymbol{A}}^{-1}\boldsymbol{A}^* = \begin{pmatrix} 1.0682 & -0.0096 & -0.0175 \\ -0.0464 & 1.0094 & -0.0007 \\ -0.0331 & -0.0002 & 1.0141 \end{pmatrix}.$$

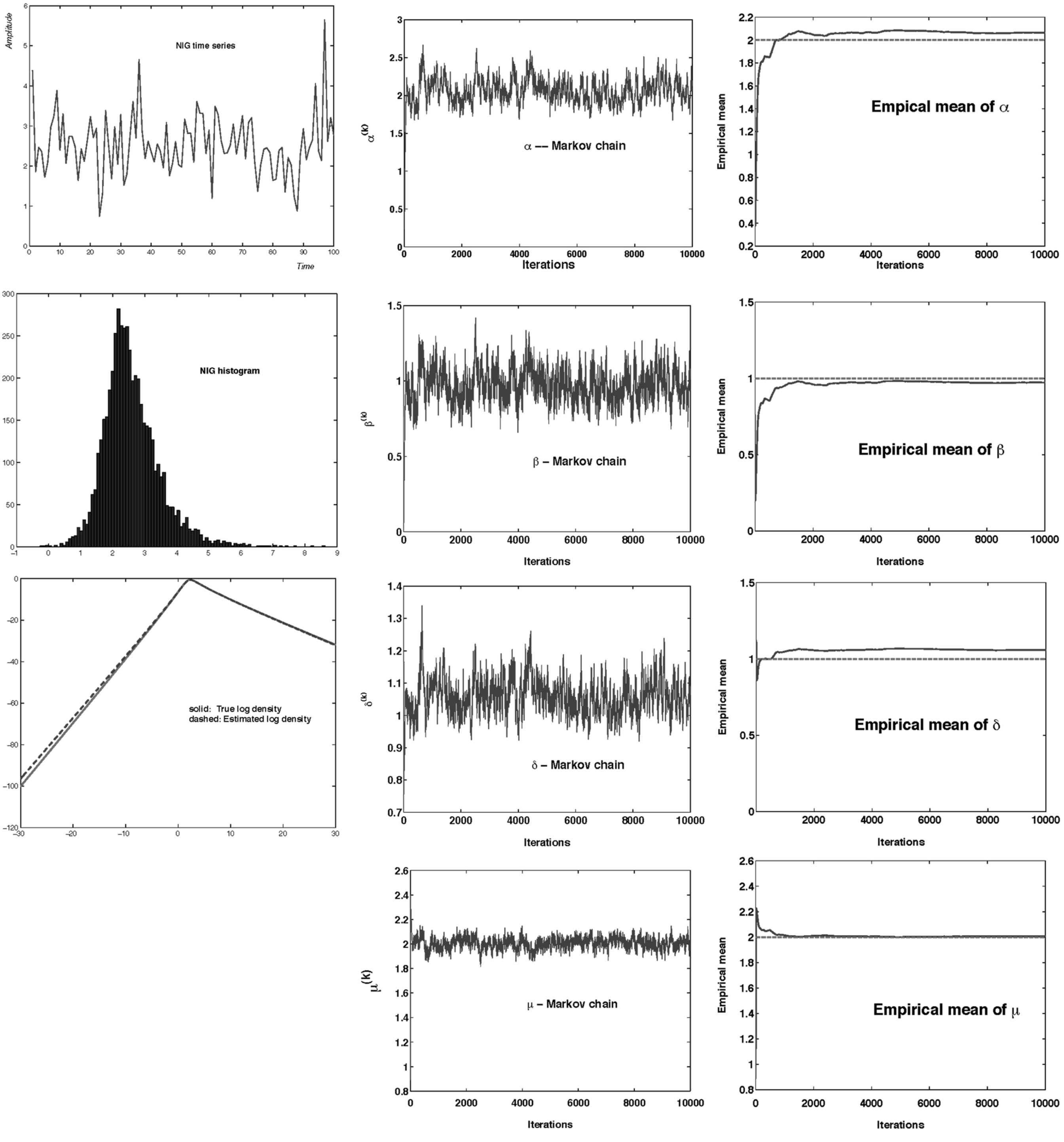

Fig. 2. From top to bottom, the first column, respectively, illustrate a portion of a simulated time series of NIG distribution, the empirical histogram, and the estimated distribution superimposed with the true sampling distribution. The second column shows the Markov chains of the parameters $(\alpha, \beta, \delta, \mu)$, respectively. The third column contains the corresponding empirical sums, which converge near the true values.

Fig. 6 illustrates the convergence of the empirical mean of the noise variance Markov chain $\tilde{\sigma}_n^{(k)}$ close the true value σ_n^*. Therefore, the Gibbs separating algorithm does not need a previous knowledge of the noise covariance, unlike the proposed methods in literature (FastICA [35] and SOBI [8]) dealing with the noisy case. In Fig. 7, the estimated source log-distributions (corresponding to $\hat{\boldsymbol{\eta}} = (1/K)\sum\tilde{\boldsymbol{\eta}}^{(k)}$) are superimposed to the true sampling distributions. We note the heavy tails and the asymmetry of the distributions and the accuracy of their estimation. In Table II, we have reported the estimated GH parameters for the three sources, within the true values. In order to better quantify the accuracy of the hyperparameter estimation, different measures of closeness between distributions are reported in Table III: Kullback–Leibler divergence, Kolmogorov distance (maximum of the absolute difference between cumulative distributions), L1 and L2 distances. The

TABLE I
THE ESTIMATED PARAMETERS $\hat{\boldsymbol{\eta}}$ ARE CLOSE TO THE TRUE PARAMETERS $\boldsymbol{\eta}^*$. DIFFERENT MEASURES OF DISTRIBUTION CLOSENESS CORROBORATE THE ACCURACY OF THE DISTRIBUTION ESTIMATION

$D(\hat{p}\|\|p^*)$			$\boldsymbol{\eta}^*$	$\hat{\boldsymbol{\eta}}$
$Kullback-Leibler$	0.0002	α	2	2.06
$Kolmogorov$	0.0034	β	1	0.97
$L1$	0.0104	δ	1	1.05
$L2$	0.0045	μ	2	2.00

sources can be estimated by either the posterior mean $\hat{\boldsymbol{S}}_{\mathrm{PM}}$ (approximated by the empirical mean of the sequence $\tilde{\boldsymbol{S}}^{(k)}$), its Rao–Blackwellized version $\hat{\boldsymbol{S}}_{\mathrm{RB-PM}}$, or the conditional posterior mean $\hat{\boldsymbol{S}}_{\mathrm{CPM}}$ (see expressions in Section III-B). The mean quadratic error between the estimate and the true sources is -3.66 dB, -3.66 dB, and 0 dB for $\hat{\boldsymbol{S}}_{\mathrm{PM}}, \hat{\boldsymbol{S}}_{\mathrm{RB-PM}}$, and $\hat{\boldsymbol{S}}_{\mathrm{CPM}}$, respectively. The accuracy of the source estimation is less impressive than the estimation of their probability distributions. This can be expected, as the noise level is high and the size of the estimated matrix $\hat{\boldsymbol{S}}$ is equal to the size of the observed matrix $\boldsymbol{X}$. In other words, the number of unknowns is equal to the number of observed data. However, one can note that the performance of the posterior mean estimate $\hat{\boldsymbol{S}}_{\mathrm{PM}}$ is better than the conditional posterior mean $\hat{\boldsymbol{S}}_{\mathrm{CPM}}$ as all the remaining parameters are integrated over within the former estimate.

C. BSS: Underdeterminate Case

The proposed Gibbs algorithm (11) represents an efficient solution to the difficult case of noisy underdetermined mixture with a high SNR. In fact, the generalized hyperbolic distribution provides a flexible parametric framework to exploit the sparsity of the sources. Three GH sources are generated and artificially mixed by a (2 × 3) mixing matrix $\boldsymbol{A}^* = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 0.5 \end{pmatrix}$. A white Gaussian noise, such that the SNR is 26 dB, is added to the mixture. Five thousand samples are considered. In order to fix the scale indeterminacy, the first raw of the mixing matrix is fixed to [1, 1, 1].

The convergence of the empirical means of the Markov chains produced by the Gibbs algorithm is shown in Fig. 8. One can note the convergence of the empirical posterior mean near the true value of the mixing matrix. However, as was shown in [7], the convergence (in distribution) of the Markov chain is very slow in the high SNR case. In fact, the mixing matrix covariance is proportional to the inverse of the signal to noise ratio. Therefore, the Markov chain does not explore the parameter space efficiently. Fortunately, the Markov chain is often stacked near the global mode of the posterior distribution. At convergence, the posterior mean estimator is

$$\hat{\boldsymbol{A}} = \begin{pmatrix} 1 & 1 & 1 \\ 1.0014 & 3.0519 & 0.4872 \end{pmatrix}.$$

The Gibbs algorithms yields an accurate estimation of the source distributions as well. In Fig. 9, the estimated log-densities are superimposed to their true shapes. Quantitative evaluation is illustrated in Tables IV and V, where the estimated GH parameters $\hat{\boldsymbol{\eta}}$ and several density divergences are reported. The proposed Gibbs algorithm is thus able to blindly estimate

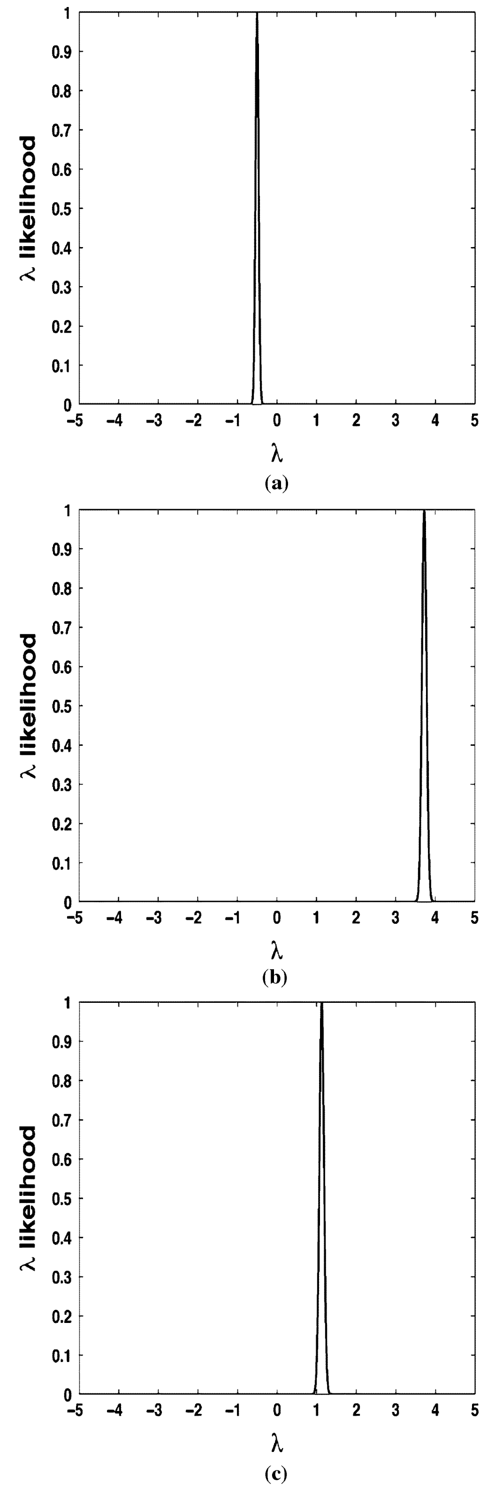

Fig. 3. The profile of the likelihood (incomplete likelihood) $p(\lambda \mid x_{1\ldots N}, \boldsymbol{\eta})$ is peaked around a value which is highly dependent on the current values of the remaining GH parameters. (a) λ-likelihood $p(\lambda \mid x_{1\ldots N}, \boldsymbol{\eta}^*)$ with true parameters ($\alpha^* = 2, \beta^* = 1, \delta^* = 1, \mu^* = 2$); (b) λ-likelihood with parameters ($\alpha = 1, \beta = 0.2, \delta = 0.1, \mu = 0$); (c) λ-likelihood with parameters ($\alpha = 1.5, \beta = 0.5, \delta = 2, \mu = 1$).

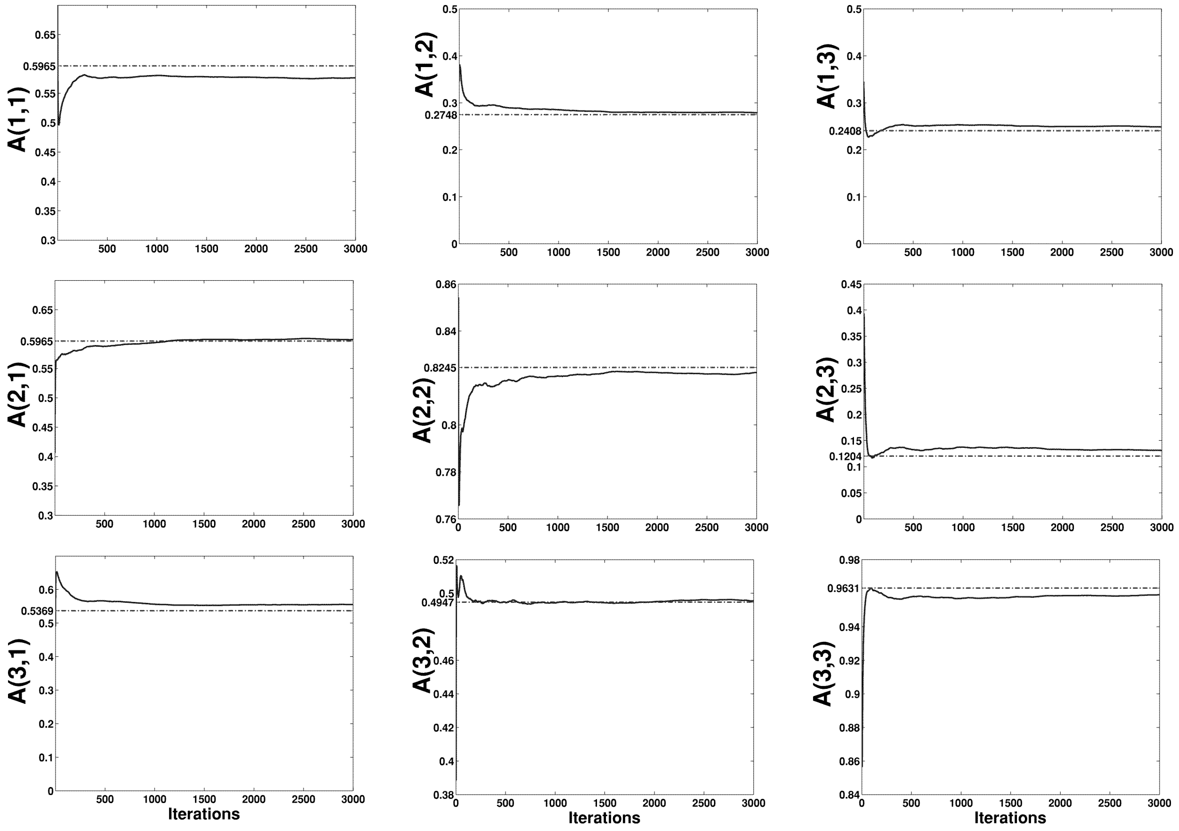

Fig. 4. Convergence of the empirical means of the mixing coefficient Markov chains.

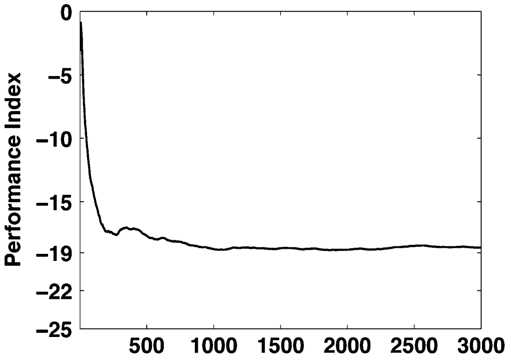

Fig. 5. Convergence of the logarithm of the performance index to a satisfactory value of -18 dB.

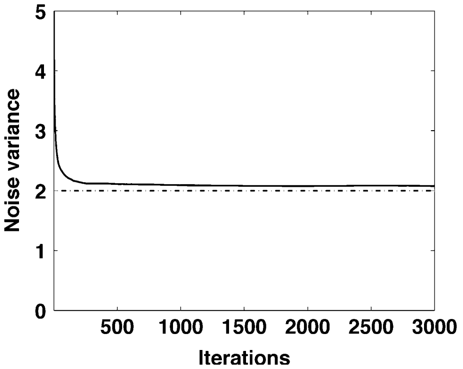

Fig. 6. Convergence of the empirical mean of the noise variance Markov chain near the true value $\sigma_n^* = 2$.

the heavy tails and the asymmetry (within its sign) of the sources.

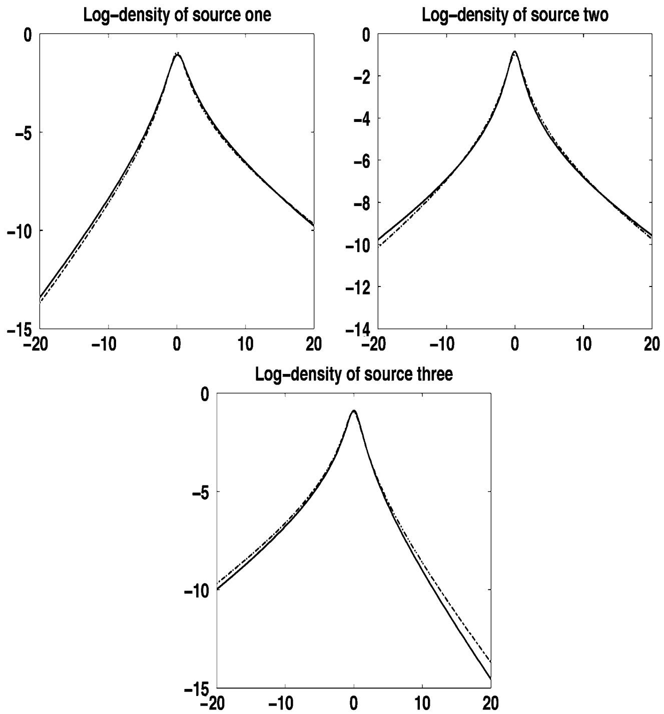

Fig. 7. Estimated log densities (in dashed lines) are almost identical to true log densities (in solid lines).

V. Conclusion and Future Research

To conclude, we have proposed a Bayesian sampling solution to the source separation problem. The proposed algorithm has shown promising results in two difficult cases: low SNR and underdeterminate mixture. The key point of this contribution

TABLE II
EMPIRICAL MEANS $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\mu})$ OF THE GH MARKOV CHAINS AND THEIR TRUE VALUES $(\alpha^*, \beta^*, \delta^*, \mu^*)$

	$Source\,1$		$Source\,2$		$Source\,3$	
	$\boldsymbol{\eta}^*$	$\hat{\boldsymbol{\eta}}$	$\boldsymbol{\eta}^*$	$\hat{\boldsymbol{\eta}}$	$\boldsymbol{\eta}^*$	$\hat{\boldsymbol{\eta}}$
α	0.3	0.3056	0.2	0.1715	0.2	0.3291
β	0.1	0.0903	0.01	0.0055	-0.1	-0.1154
δ	1	1.1634	1	0.8328	1	0.9473
μ	0	0.0385	0	-0.0418	0	0.0776

TABLE III
DIFFERENT DENSITY DIVERGENCE MEASURES BETWEEN THE TRUE DISTRIBUTION (WITH PARAMETER $\boldsymbol{\eta}^*$) AND THE ESTIMATED ONE (WITH PARAMETER $\hat{\boldsymbol{\eta}}$)

$\sqrt{D(\hat{p}\|\|p^*)}$	$Source$ 1	$Source$ 2	$Source$ 3
$Kullback-Leibler$	0.0048	0.0062	0.0033
$Kolmogorov$	0.0304	0.0372	0.0310
$L1$	0.1106	0.1126	0.1048
$L2$	0.0418	0.0462	0.0396

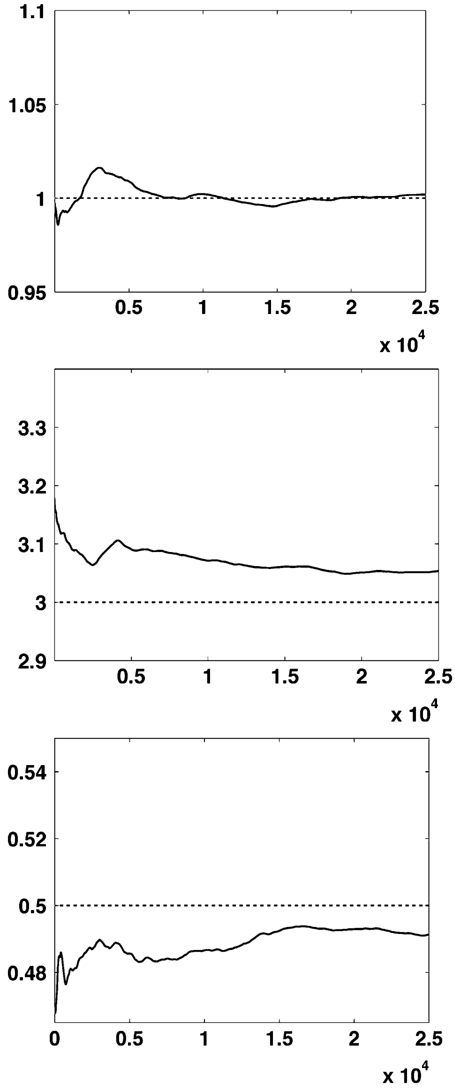

Fig. 8. Convergence of the empirical means of the mixing coefficient Markov chains in the underdetermined case. The first raw of $\tilde{\boldsymbol{A}}^{(k)}$ is set fixed to [1, 1, 1].

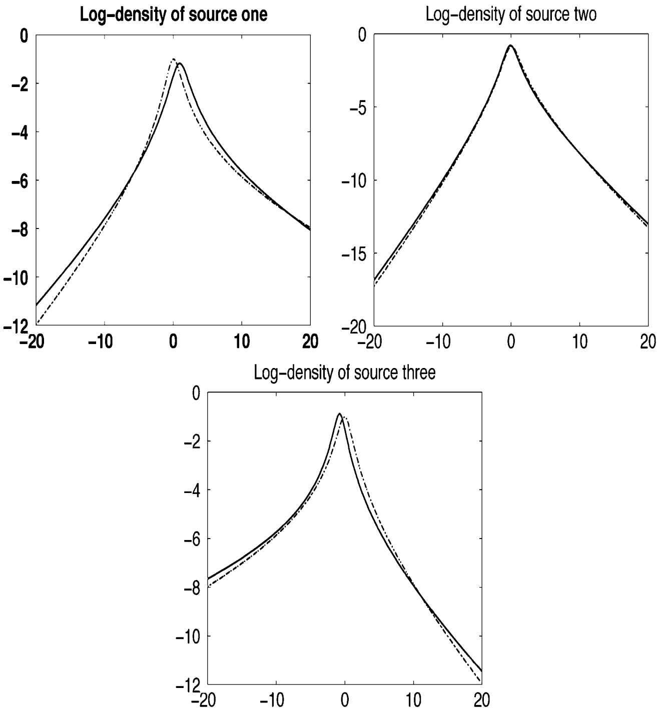

Fig. 9. True log-densities (in dashed lines) superimposed to the estimated log-densities (in solid lines) in the underdetermined case.

TABLE IV
EMPIRICAL MEANS $(\hat{\alpha}, \hat{\beta}, \hat{\delta}, \hat{\mu})$ OF THE GH MARKOV CHAINS AND THEIR TRUE VALUES $(\alpha^*, \beta^*, \delta^*, \mu^*)$ IN THE UNDERDETERMINED CASE

	$Source\,1$		$Source\,2$		$Source\,3$	
	$\boldsymbol{\eta}^*$	$\hat{\boldsymbol{\eta}}$	$\boldsymbol{\eta}^*$	$\hat{\boldsymbol{\eta}}$	$\boldsymbol{\eta}^*$	$\hat{\boldsymbol{\eta}}$
α	0.2	0.1901	0.5	0.4767	0.2	0.1603
β	0.1	0.0654	0.1	0.0995	-0.1	-0.0862
δ	1	1.2196	1	0.92528	1	0.8433
μ	0	0.9076	0	-0.1354	0	-0.7417

TABLE V
DIFFERENT DENSITY DIVERGENCE MEASURES BETWEEN THE TRUE DISTRIBUTION (WITH PARAMETER $\boldsymbol{\eta}^*$) AND THE ESTIMATED ONE (WITH PARAMETER $\hat{\boldsymbol{\eta}}$) IN THE UNDERDETERMINED CASE

$\sqrt{D(\hat{p}\|\|p^*)}$	$Source$ 1	$Source$ 2	$Source$ 3
$Kullback-Leibler$	0.1851	0.0096	0.1644
$Kolmogorov$	0.2740	0.0638	0.2594
$L1$	0.8023	0.1423	0.6713
$L2$	0.3631	0.0740	0.3119

is the modeling of sources by the generalized hyperbolic processes. The GH process is a five-parameter distribution able to cover a wide range of tail behaviors as well as the asymmetry characteristics. Its normal mean-variance continuous mixture is compatible with the hidden variable structure of the source separation problem. Therefore, the Gibbs sampling is efficiently implemented. Morover, this provides us with an original unification of the exploitation of the high-order statistics and the nonstationary second-order statistics to solve the BSS problem. Taking into account the noise in the model and the joint estimation of its noise covariance are the main reasons of the robustness of the proposed method in a high noisy environment. Although we have proposed a sampling step for the parameter λ of the generalized hyperbolic distribution, the Gibbs algorithm yields poor results when this parameter is not set to a fixed value.

We have noted this behavior also in the estimation of the GH parameters of an observed raw sample as shown in Fig. 3. Further research should be done to clarify the role of the parameter λ and the possibility of incorporating its sampling in the separating Gibbs algorithm. Extending the generalized hyperbolic model to incorporate a temporal correlation is also an interesting direction to investigate in order to improve the separating algorithm performance.

APPENDIX
RATIO METHOD

The ratio method is an acceptation/rejection sampling procedure [36]. Let $f(x)$ the density to sample from and assume we have the expression of $h(x) \propto f(x)$. Consider the set S_h

$$S_h = \{(u,v)\,|\,0 < u \leq \sqrt{h(v/u)}\}.$$

Then, we have the following theorem.

Theorem 1: If $k_u = \sup_x \sqrt{h(x)}$ and $k_v = \sup_x |x|\sqrt{h(x)}$, then we have the following statements.

- The rectangle $(0, k_u; -k_v, k_v)$ encloses the set S_h.
- If $h(x) = 0$ for $x < 0$, the enclosing rectangle is $(0, k_u; 0, k_v)$.
- Let the point (U, V) be sampled uniformly within the rectangle $(0, k_u; -k_v, k_v)$ (or $(0, k_u; 0, k_v)$). If $(U, V) \in S_h$, then $X = V/U$ is distributed according to $f(x)$.

The ratio rejection algorithm is then the following:

$$\begin{aligned}
&1.\ \text{Compute } k_u = \sup_x \sqrt{h(x)} \text{ and } k_v = \sup_x |x|\sqrt{h(x)}.\\
&2.\ \text{Sample uniformly } U \text{ and } V \text{ in } [0, k_u] \text{ and } [0, k_v].\\
&3.\ \text{If } U < \sqrt{h(X)}, \text{accept } X = V/U, \text{else return to 2.}
\end{aligned} \tag{15}$$

Example 1 (GIG Sampling): Let $f(x) \propto x^{\lambda-1}\exp(-0.5(\gamma^2 x + \delta^2 x^{-1}))$ be a GIG distribution to sample from. First, assume $\lambda \geq 0$; the case $\lambda < 0$ will be studied later. Then, let the following transformation:

$$\begin{cases} b = \gamma\delta \\ \eta = \gamma/\delta \end{cases}$$

which leads to the following function $h(x)$

$$h(x) = x^{\lambda-1}\exp(-0.5b(\eta x + \eta^{-1}x^{-1}))$$

where we note that we can sample x according to $p(x) \propto h(x) = x^{\lambda-1}\exp(-0.5b(x + x^{-1}))$ and then transform x to x/η.

Now, the key point of the sampling procedure is the fact that we have explicit formula for the maximizers of $\sqrt{h(x)}$ and $|x|\sqrt{h(x)}$

$$\begin{cases} x_m = \arg\max_x \sqrt{h(x)} = \frac{\lambda-1+\sqrt{(\lambda-1)^2+b^2}}{b} \\ y_m = \arg\max_x |x|\sqrt{h(x)} = \frac{\lambda+1+\sqrt{(\lambda+1)^2+b^2}}{b} \end{cases}. \tag{16}$$

Then, compute k_u and k_v according to

$$\begin{cases} k_u = \sqrt{h(x_m)} \\ k_v = |y_m|\sqrt{h(y_m)} \end{cases}$$

and perform the two remaining steps of the algorithm (15). The obtained x is divided by the parameter η.

The case of $\lambda < 0$ is simply obtained as follows.

1. Sample $y \sim \text{GIG}(-\lambda, b, 1)$.
2. Put $x = 1/y$.
3. Replace x by x/η.

REFERENCES

[1] C. Jutten and J. Herault, "Blind separation of sources, I: An adaptive algorithm based on neuromimetic architecture," *Signal Process.*, vol. 24, no. 1, pp. 1–10, 1991.

[2] P. Comon, "Independent component analysis, a new concept?," *Signal Process. (Special Issue on Higher-Order Statistics)*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[3] A. J. Bell and T. J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neur. Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.

[4] E. Moulines, J. Cardoso, and E. Gassiat, "Maximum likelihood for blind separation and deconvolution of noisy signals using mixture models," presented at the Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Munich, Germany, Apr. 1997.

[5] H. Attias, "Independent factor analysis," *Neur. Comput.*, vol. 11, pp. 803–851, 1999.

[6] H. Snoussi and A. Mohammad-Djafari, "Bayesian unsupervised learning for source separation with mixture of Gaussians prior," *Int. J. VLSI Signal Process. Syst. Signal, Image, Video Technol.*, vol. 37, no. 2–3, pp. 263–279, Jun.–Jul. 2004.

[7] ——, "Fast joint separation and segmentation of mixed images," *J. Electron. Imag.*, vol. 13, no. 2, pp. 349–361, Apr. 2004.

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

[9] K. Rahbar and J. Reilly, "Blind source separation of convolved sources by joint approximate diagonalization of cross-spectral density matrices," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Salt Lake City, UT, May 2001, vol. 5, pp. 2745–2748.

[10] A. Belouchrani and M. Amin, "Blind source separation using time-frequency distributions: Algorithm and asymptotic performance," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Munich, Germany, 1997, pp. 3469–3472.

[11] K. Matsuoka, M. Ohya, and M. Kawamoto, "A neural net for blind separation of nonstationary sources," *Neur. Netw.*, vol. 8, no. 3, pp. 411–419, 1995.

[12] S. Choi and A. Cichocki, "Blind separation of nonstationary sources in noisy mixtures," *Electron. Lett.*, vol. 36, no. 9, pp. 848–849, Apr. 2000.

[13] A. Souloumiac, "Blind source detection and separation using second order nonstationarity," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1995, pp. 1912–1915.

[14] D.-T. Pham and J. Cardoso, "Blind separation of instantaneous mixtures of non stationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 11, pp. 1837–1848, Nov. 2001.

[15] H. Snoussi, G. Patanchon, J. F. Macías-Pérez, A. Mohammad-Djafari, and J. Delabrouille, "Bayesian blind component separation for cosmic microwave background observation," in *Bayesian Inference and Maximum Entropy Methods*, R. L. Fry, Ed., 2001, pp. 125–140, American Institute of Physics.

[16] J. Cardoso, H. Snoussi, J. Delabrouille, and G. Patanchon, "Blind separation of noisy Gaussian stationary sources. Application to cosmic microwave background imaging," presented at the EUSIPCO, Toulouse, France, Sep. 2002.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Statist. Soc. B*, vol. 39, pp. 1–38, 1977.

[18] O. Barndorff-Nielsen, "Exponentially decreasing distributions for the logarithm of particle size," in *Proc. Roy. Soc.*, London, U.K., 1977, vol. 353, pp. 401–419.

[19] M. Zibulevsky, B. Pearlmutter, P. Bofill, and P. Kisilev, , S. J. Roberts and R. M. Everson, Eds., "Blind source separation by sparce decomposition," in *Independent Component Analysis: Principles and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2001.

[20] C. Févotte, S. Godsill, and P. Wolfe, "Bayesian approach for blind separation of underdetermined mixtures of sparse sources," presented at the 5th Int. Conf. Independent Component Analysis Blind Source Separation (ICA 2004), Granada, Spain, 2004.

[21] B. Bibby and M. Sorensen, "Hyperbolic processes in finance," in *Handbook of Heavy Tailed Distributions in Finance*. Amsterdam, The Netherlands: Elsevier Science, 2003, pp. 211–248.

[22] O. Barndorff-Nielsen and P. Blaesild, , C. Taillie, G. Patil, and B. Baldessari, Eds., "Hyperbolic distributions and ramifications: Contribution to theory and application," in *Statistical Distributions in Scientific Work*. Dordrecht, The Netherlands: Reidel, 1981, vol. 4, pp. 19–44.

[23] K. Prause, "The generalized hyperbolic model: Estimation, financial derivatives and risk measurement," Ph.D. dissertation, Mathematics Faculty, Freiburg Univ., Freiburg, Germany, 1999.

[24] O. Barndorff-Nielsen, Normal inverse Gaussian processes and the modeling of stock returns Dept. of Theoretical Statistics, Aarhus Univ., 1995, Tech. Rep. 300.

[25] P. Blaesild and M. Sorensen, "'Hyp'—A computer program for analyzing data by means of the hyperbolic distribution," Dept. of Theoretical Statistics, Aarhus Univ., 2002, Tech. Rep. 248.

[26] R. Protassov, "EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ," *Stat. Comput.*, vol. 14, pp. 67–77, 2004.

[27] J. Lillestol, "Bayesian estimation of NIG-parameters by Markov chain Monte Carlo methods," Norwegian School of Economics and Business Administration, 2001, Tech. Rep. 2001/3.

[28] D. Karlis and J. Lillestol, "Bayesian estimation of NIG models via Markov chain Monte Carlo methods," *Appl. Stochastic Models Bus. Ind.*, to be published.

[29] C. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.

[30] H. Snoussi and A. Mohammad-Djafari, , C. Williams, Ed., "Information geometry and prior selection," in *Bayesian Inference and Maximum Entropy Methods*. College Park, MD: American Institute of Physics, Aug. 2002, pp. 307–327.

[31] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Trans. Circuits. Syst.*, vol. CS-25, no. 9, pp. 772–781, 1978.

[32] S. L. Adler, "Over-relaxation method for Monte-Carlo evaluation of the partition function for multiquadratic actions," *Phys. Rev. D*, vol. 23, pp. 2901–2904, Jun. 1981.

[33] R. M. Neal, , M. I. Jordan, Ed., "Suppressing random walks in Markov chain Monte Carlo using ordered overrelaxation," in *Learning in Graphical Models*. Dordrecht, The Netherlands: Kluwer Academic, 1998, pp. 205–225.

[34] E. Moreau and O. Macchi, "High-order contrasts for self-adaptative source separation," *Adapt. Contr. Signal Process.*, vol. 10, pp. 19–46, 1996.

[35] A. Hyvärinen, "Noisy independent component analysis by Gaussian moments," in *Proc. 1st Int. Conf. Indep. Comp. Anal. Blind Source Separation (ICA)*, Aussois, France, Jan. 11–15, 1999, pp. 473–478.

[36] J. Dagpunar, *Principles of Random Variate Generation*. Oxford, U.K.: Clarendon Press, 1988.

Hichem Snoussi was born in Bizerta, Tunisia, in 1976. He received the diploma degree in electrical engineering from the Ecole Supérieure d'Electricité (Supélec), Gif-sur-Yvette, France, in 2000. He received the DEA degree and the Ph.D. degree in signal processing from the University of Paris-Sud, Orsay, France, in 2000 and 2003, respectively.

Between 2003 and 2004, he was a Postdoctoral Researcher with IRCCyN, Institut de Recherches en Communications et Cybernétiques de Nantes, France. Since 2005, he has been an Associate Professor at the Institute of Technology of Troyes, France. His research interests include Bayesian technics for source separation, information geometry, differential geometry, machine learning, and robust statistics with application to brain signal processing, astrophysics, and so forth.

Jérôme Idier was born in France in 1966. He received the diploma degree in electrical engineering from École Supérieure d'Électricité, Gif-sur-Yvette, France, in 1988 and the Ph.D. degree in physics from University of Paris-Sud, Orsay, France, in 1991.

In 1991, he joined the Centre National de la Recherche Scientifique. He is currently with the Institut de Recherche en Communications et Cybernétique de Nantes, France. His major scientific interest is in probabilistic approaches to inverse problems for signal and image processing.