BLIND SEPARATION OF GENERALIZED HYPERBOLIC PROCESSES: UNIFYING APPROACH TO STATIONARY NON GAUSSIANITY AND GAUSSIAN NON STATIONARITY

Hichem Snoussi and Jérôme Idier

IRCCyN, UMR CNRS 6597, BP 92101, 1 rue de la Noë, 44321 Nantes Cedex 3, France
Email: name@irccyn.ec-nantes.fr

ABSTRACT

In this contribution, we propose a Bayesian sampling solution to the problem of noisy blind separation of generalized hyperbolic (GH) signals. GH 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 MCMC method applied to source separation. The incomplete data structure of the GH distribution is indeed compatible with the hidden variable nature of the source separation problem. Our algorithm involves hyperparameters estimation as well. Therefore, it can be used, independently, to fit the parameters of the GH distribution to real data.

1. INTRODUCTION

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

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

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. A challenging aspect of the BSS problem is the absence of information about the mixing matrix $\boldsymbol{A}$.

Many proposed algorithms are designed to linearly demixing the observations $\boldsymbol{x}_{1..T}$ on the basis of independent identically distributed (iid) source modeling. The separation principle in these methods relies on the statistical independence of the reconstructed sources. This is the case of Independent Component Analysis (ICA) [1]. However, ICA is designed in a noiseless framework. In addition, the separation necessarily relies also on higher order statistics, allowing at most one source to be Gaussian. In [2], the noisy case was tackled with the maximum likelihood approach using the EM algorithm, the sources being modeled by finite Gaussian mixture. However, exact implementation of the EM algorithm is computationally expensive. In addition, 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 [2]).

Our contribution is to efficiently implement a maximum likelihood solution in the noisy case, the sources being iid. The proposed method is based on the estimation of the mixing matrix, the source distribution parameters and the noise covariance matrix. Thus, the same algorithm can be applied to overdeterminate as well as underdeterminate cases without any prewhitening step. As the underdeterminate case can be solved by exploiting the sparsity of sources [3], the generalized hyperbolic (GH) distributions are well appropriate to model the sources and capture their heavy tails and also their skewness. The method implicitly incorporates a denoising procedure and it is consequently robust to high level noise. The key point is the use of Barndorff-Nielsen's **Generalized hyperbolic** distributions [4]. Their representation as normal mean-variance continuous mixture models is remarkably compatible with the hidden structure of the source separation problem: they can be interpreted either as stationary non Gaussian, or as Gaussian non stationary processes. This provides a new insight into the unification of the use of non stationary second order statistics and stationary higher order statistics to solve the problem of blind source separation. 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 [5] where sources are modeled by t-Student distributions, since the latter form a subclass of the GH family.

The paper is organized as follows. Section 2 is devoted to the GH law and to its properties. More specifically, we present an original Bayesian algorithm to fit the five parameters of the GH distribution from an observed finite sample. In Section 3, a Bayesian algorithm is introduced to solve the blind source separation problem in the noisy case. Finally, some simulation results corroborating the efficiency of the proposed algorithm are presented.

2. GENERALIZED HYPERBOLIC PROCESSES

2.1. Description and properties

In this paragraph, we briefly describe the GH distributions and their main properties. The GH law is mainly used to fit financial data. It corresponds to a five parametric family $H(\lambda, \alpha, \beta, \delta, \mu)$ introduced by Barndorff-Nielsen [4]. A random variable X belongs to $H(\lambda, \alpha, \beta, \delta, \mu)$ if its pdf reads:

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

where $\gamma^2 = \alpha^2 - \beta^2$ and K_λ 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 valid domain for the parameters is as follows:

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

GH distributions enjoy the property of being invariant under affine transformations:

$$X \sim H(\lambda, \alpha, \beta, \delta, \mu) \implies a\,X + b \sim H\left(\lambda, \frac{\alpha}{a}, \frac{\beta}{a}, a\delta, a\mu + b\right).$$

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. Figure 1 depicts examples of GH distributions. One can note that a wide range of tail behaviors is covered.
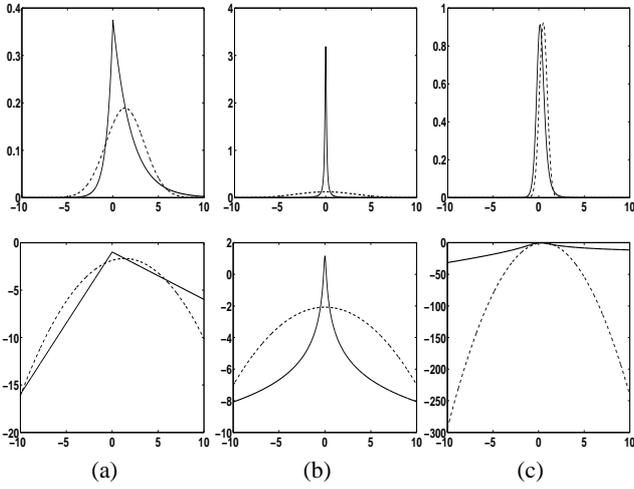

(a)   (b)   (c)

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:

$$H(x; \lambda, \alpha, \beta, \delta, \mu) = \int_0^\infty \mathcal{N}(x; \mu + \beta w, w)\, GIG(w; \lambda, \gamma, \delta)\, dw \tag{2}$$

where the variance W of each Gaussian component follows a Generalized Inverse Gaussian (GIG) distribution ($w > 0$):

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

In other words, the GH process can be seen as a doubly stochastic process:

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

Such a property will be a key point both in the estimation of the parameters and in the BSS problem.

2.2. Parameter estimation

The estimation of parameters $\boldsymbol{\eta} = (\lambda, \alpha, \beta, \delta, \mu)$ from an iid GH sample $\{x_i\}_{i=1..N}$ is a difficult task. As reported in [7], this difficulty is essentially due to the flatness of the likelihood with respect to the parameters and particularly with respect to the parameter λ. In the literature, several contributions are restricted to the estimation of parameters within particular subclasses (fixing the value of the parameter λ). Recently, Protassov [8] used the incomplete data structure of the problem (2) to propose an EM algorithm. The EM algorithm is however restricted to work within subclasses, that is for fixed λ. A Bayesian sampling solution is proposed by Lillestol [9] for the case of NIG distribution. However, the proposed algorithm is restricted to $\lambda = -1/2$. In this paper, we propose an original contribution to GH parameter estimation, without restrictions, exploiting the latent problem structure and based on Gibbs sampling. We propose a reparametrization of the GH distribution in order to efficiently sample the conditionals[2]. The Gibbs sampling algorithm for estimating the parameters $\boldsymbol{\eta}$ consists in alternating the sampling of the hidden variances $w_{1..N}$ (given the parameter $\boldsymbol{\eta}$) and the conditional sampling of the parameter of interest $\boldsymbol{\eta}$ (given the variances). The *a posteriori* distribution is, according to the Bayes rule,

$$p(w_{1..N} \mid x_{1..N}, \boldsymbol{\eta}) = \prod_{i=1}^{N} GIG\left(w_i; \lambda - \frac{1}{2}, \gamma^2 + \beta^2, \delta^2 + (x_i - \mu)^2\right) \tag{3}$$

where we note that the GIG density is a conjugate prior (the *a posteriori* density belongs to the same family). The sampling of the variances relies then on the efficient sampling of the GIG distribution. This is performed by the Ratio method [10] which is an exact rejection sampling method [6]. The second step of the Gibbs algorithm consists in sampling the parameter $\boldsymbol{\eta}$ according to its conditional *a posteriori* distribution $p(\boldsymbol{\eta} \mid x_{1..N}, w_{1..N})$ which is written as,

$$p(\boldsymbol{\eta} \mid x_{1..N}, w_{1..N}) \propto p(x_{1..N}, w_{1..N} \mid \boldsymbol{\eta})\, p(\boldsymbol{\eta}) \tag{4}$$

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

$$\begin{cases} \xi_1 = \lambda = \eta_1 \\ \xi_2 = \gamma/\delta = \sqrt{\eta_2^2 - \eta_3^2}/\eta_4 \\ \xi_3 = \gamma\,\delta = \sqrt{\eta_2^2 - \eta_3^2}\,\eta_4 \\ \xi_4 = \beta = \eta_3 \\ \xi_5 = \mu = \eta_5 \end{cases} \iff \begin{cases} \eta_1 = \xi_1 \\ \eta_2 = \sqrt{\xi_2\xi_3 + \xi_4} \\ \eta_3 = \xi_4 \\ \eta_4 = \sqrt{\xi_3/\xi_2} \\ \eta_5 = \xi_5 \end{cases}$$

---

[1]Among the Matlab files freely available from the first author, the program rGIG.m efficiently simulates a GIG random variable based on the ratio method [6].

[2]this reparametrization is different from that considered in [8].

3. BAYESIAN BLIND SEPARATION

In this section, we assume that n GH sources 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 following simple matrix form:

$$X = AS + N,$$

where the $(m \times T)$-matrix X contains the m observed rows, the $(n \times T)$-matrix S contains the n unobserved source rows and N is the noise. We assume that each source row $s_j = (s_j(1), .., s_j(T))$ follows a GH distribution $H(\lambda_j, \alpha_j, \beta_j, \delta_j, \mu_j)$ and that each noise row $n_j = (n_j(1), .., n_j(T))$ is white with a variance σ_j^2. The identification problem is very ill posed as the $(m \times n)$-mixing matrix, the sources S and their corresponding hyperparameters $\eta = (\lambda_j, \alpha_j, \beta_j, \delta_j, \mu_j)_{j=1}^n$ 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 explicitly modeled in the inference process and any additional prior information can be incorporated. Given the observations X, the *a posteriori* distribution of the unknowns $\theta = (A, R_n, S, \eta)$, according to the Bayesian rule is:

$$p(\theta \mid X, \mathcal{I}) \propto p(X \mid \theta, \mathcal{I}) \, p(\theta \mid \mathcal{I}), \tag{5}$$

where $\mathcal{I}$ contains the prior information such as the noisy model, the GH density of sources and the whiteness of the noise.

The posterior likelihood (5) incorporates our knowledge about the unknowns, but it does not provide a specific estimation procedure. In general, expression (5) corresponds to a complicated, multimodal function of the parameters. Bayesian sampling is an efficient tool to tackle this challenging inference problem. More specifically, Gibbs sampling is well suited to the separation problem. It produces a Markov chain $\tilde{\theta}^{(k)} = (\tilde{A}^{(k)}, \tilde{R}_n^{(k)}, \tilde{S}^{(k)}, \tilde{W}^{(k)}, \tilde{\eta}^{(k)})$ that converges, in distribution, to the *a posteriori* distribution (5).

The formulation of the GH 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.

3.1. Gibbs algorithm

The cyclic sampling steps are as follows:

1. Sample $S \sim p(S \mid X, \tilde{A}, \tilde{R}_n, \tilde{W}, \tilde{\eta})$
2. Sample $W \sim p(W \mid \tilde{S}, \tilde{\eta})$
3. Sample $\eta \sim p(\eta \mid \tilde{S}, \tilde{W})$
4. Sample $(A, R_n) \sim p(A, R_n \mid X, \tilde{S})$

$$\tag{6}$$

Given the data X and the remaining components of θ, the sources have temporally independent, *a posteriori* multivariate Gaussian distribution. This is obtained by applying the Bayes rule:

$$p(S \mid X, \tilde{\theta}) \propto \prod_{t=1}^{T} \mathcal{N}(s_t ; \boldsymbol{\mu}_s(t), \Gamma_s(t)). \tag{7}$$

The means and covariances of the sources at time t have the following expressions:

$$\begin{cases} \Gamma_s(t) = \left(A^* R_n^{-1} A + P_w^{-1}\right)^{-1} \\ \boldsymbol{\mu}_s(t) = \Gamma_s(t) \left[A^* R_n^{-1} x_t + P_w^{-1}(\mu + \beta \odot w_t)\right] \end{cases}$$

where $P_w = \text{diag}(w_t)$ is the *a priori* source covariance and $[\mu_j] + [\beta_j] \odot w_t$ is the *a priori* mean.

The conditioned sampling in the second and third steps of the Gibbs Algorithm (6), given the sampled sources $\tilde{S}$, are the same as in the previous Section 2. In fact, given the sources, the variances W and the hyperparameters η are independent of the data X as they are not related to the mixing process. They are spatially independent and, for each component $j = 1..n$, the variances row w_j is sampled according to a GIG distribution as in equation (3).

The sampling of the mixing matrix and the covariance matrix given the data and the sources is the same as in [2]. For a Jeffrey's prior (see [11] for details of Fisher matrix computation), the *a posteriori* distribution of (A, R_n^{-1}) is Normal-Wishart:

$$p(A, R_n^{-1}) = \mathcal{N}(A ; A_p, \Gamma_a) \, \mathcal{W}_m(R_n^{-1} ; \nu_p, \Sigma_p)$$

with parameters:

$$\begin{cases} A_p = R_{xs} R_{ss}^{-1}, \\ \Gamma_a = \frac{1}{T} R_{ss}^{-1} \otimes R_n, \end{cases} \quad \begin{cases} \nu_p = T - n, \\ \Sigma_p = \frac{T}{T-n}(R_{xx} - R_{xs} R_{ss}^{-1} R_{sx}) \end{cases}$$

where $\otimes$ is the Kronecker product and

$$R_{xx} = \frac{1}{T} \sum x_t x_t^*, \; R_{sx} = \frac{1}{T} \sum s_t x_t^*, \; R_{ss} = \frac{1}{T} \sum s_t s_t^*.$$

Remark 1 (over-relaxation) *The covariance Γ_p of the mixing matrix is inversely proportional to the signal to noise ratio. In the case of a high signal to noise ratio, the covariance is very small, which leads to a slow convergence of the Markov chain. In other words, the conditional distribution of the mixing matrix is very sharp around a mean value depending on the sampled sources due to a high correlation with this latter. The Markov chain is then unable to efficiently explore the parameter domain. To tackle this problem, a general solution proposed by Adler [12] in the Gaussian case consists in over-relaxing the chain by introducing a negative correlation between the updates. If the parameter to be updated θ has a Gaussian distribution $\mathcal{N}(m, LL^*)$, the retained value at iteration k is the following:*

$$\theta^{(k)} = m + \alpha (m - \theta^{(k-1)}) + \sqrt{1 - \alpha^2} \, Lu,$$

where u is a standard Gaussian vector and $\alpha \in [-1, 0]$ controls the degree of over-relaxation.

Remark 2 *Formally, the separation method matches the empirical data covariance R_{xx} to its theoretical expression $A P_w A^T + R_n$, where $P_w = \text{diag}(w_t)$ is the covariance of the non stationary sources simultaneously updated through the Gibbs iterations. This represents an unification between the use of higher order statistics and non stationary second order statistics.*

3.2. Simulation Results

In this paragraph, we illustrate the performance of the Gibbs separating algorithm on simulated data, in a very noisy context. Three sources are generated according to the GH model (1). They are artificially mixed by a mixing matrix $A^* = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & .5 \\ .9 & 1.8 & 4 \end{pmatrix}$ and corrupted by a white noise such that the SNR is respectively -0.4, 4 and 2 dB for each of the three detectors. Figure 2(a) shows the convergence of the empirical sums of the mixing matrix Markov

chains to their true values. Figure 2(b) shows the evolution of a performance index that evaluates the closeness of the matrix product $\mathcal{P} = \widehat{A}^{-1}A^*$ to the identity matrix. Following [13], it is defined by (when $\mathcal{P}$ approaches the identity matrix, the index converges to 0):

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

The convergence of the empirical mean of the performance index to -20 dB corroborates the effectiveness of the separating algorithm.

In Figure 3, the estimated source log-distributions 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 order to quantify the accuracy of the hyperparameters estimation, different measures of closeness between distributions are reported in Table 1: Kullback-Leibler divergence, Kolmogorov distance (maximum of the absolute difference between cumulative distributions), L1 and L2 distances.
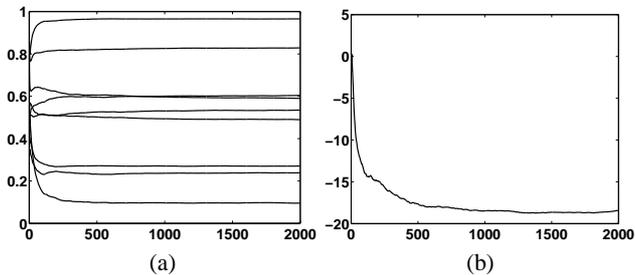

Fig. 2. (a) Convergence of the empirical means of the mixing coefficients Markov chains. (b) Convergence of the logarithm of the performance index to a satisfactory value of -20 dB.
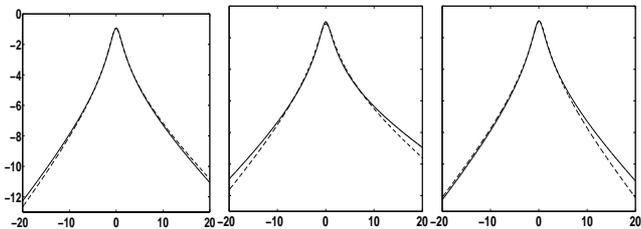

Fig. 3. Estimated log densities (in dashed lines) are almost identical to true log densities (in solid lines).

$\sqrt{D(\hat{p}\|p^*)}$	Source 1	Source 2	Source 3
$Kullback - Leibler$	0.03	0.07	0.04
$Kolmogorov$	0.01	0.02	0.02
$L1$	0.04	0.10	0.06
$L2$	0.02	0.03	0.02

Table 1. The estimated parameters $\hat{\eta}$ are close to the true parameters η^*. Different measures of distribution closeness corroborate the accuracy of the distribution estimation.

4. REFERENCES

[1] P. Comon, "Independent Component Analysis, a new concept ?", *Signal processing, Special issue on Higher-Order Statistics, Elsevier*, vol. 36 (3), pp. 287–314, April 1994.

[2] H. Snoussi and A. Mohammad-Djafari, "Fast joint separation and segmentation of mixed images", *Journal of Electronic Imaging*, vol. 13, no. 2, pp. 349–361, April 2004.

[3] M. Zibulevsky, B. Pearlmutter, P. Bofill, and P. Kisilev, "Blind source separation by sparce decomposition", *In Roberts, S.J., Everson, R.M., eds: Independent Component Analysis: Principles and Practice. Cambridge University Press*, 2001.

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

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

[6] J. Dagpunar, *Principles of random variate generation*, Clarendon Press, Oxford, 1988.

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

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

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

[10] H. Snoussi and J. Idier, "Bayesian Blind Separation of Generalized Hyperbolic Processes in Noisy and Underdeterminate Mixtures", *IEEE Trans. Signal Processing*, December 2004.

[11] H. Snoussi and A. Mohammad-Djafari, "Information Geometry and Prior Selection", in *Bayesian Inference and Maximum Entropy Methods*, C. Williams, Ed. MaxEnt Workshops, August 2002, pp. 307–327, Amer. Inst. Physics.

[12] S. L. Adler, "Over-relaxation method for Monte-Carlo evaluation of the partition function for multiquadratic actions", *Physical Review D*, vol. 23, pp. 2901–2904, June 1981.

[13] E. Moreau and O. Macchi, "High-order contrasts for self-adaptative source separation", in *Adaptative Control Signal Process. 10*, 1996, pp. 19–46.