A von Mises–Fisher prior to Remove Scale Ambiguity in Blind Deconvolution

Benjamin Traullé, Stéphanie Bidon, Damien Roque
ISAE-SUPAERO, Université de Toulouse, France
{firstname.lastname}@isae-supaero.fr

***Abstract*—We propose a von Mises–Fisher prior to remove scale ambiguity arising in blind deconvolution (BD). Indeed, traditional Bayesian BD methods rely on Gaussian priors that address only partially this ambiguity. We first derive the posteriors in closed-form to underline the benefit of a von Mises–Fisher prior compared with a conventional Gaussian prior. We also showcase its applicability within an augmented Gibbs sampler that includes a state-of-the-art re-scaling step. However, state-space exploration issues may still occur owing to the multimodal nature of the posteriors. These preliminary results encourage the design of BD-specific sphere-constrained sampling techniques.**

***Index Terms*—Blind deconvolution, scale ambiguity, Gibbs sampler, von Mises–Fisher prior.**

I. Introduction

Blind deconvolution (BD) problems arise in many areas such as physiological signal analysis [1], [2], seismic exploration [3], multi-channel imaging restoration [4]. The general principle of BD is to recover two vectors $\mathbf{x}$ and $\mathbf{g}$ from a convolutional noisy observation $\mathbf{y}$ as defined in (1). However, without any further information about $\mathbf{x}$ and $\mathbf{g}$, the problem is ambiguous and thus ill-posed. For instance, the so-called scale ambiguity—which is of interest in this work—reflects that if $(\mathbf{x}, \mathbf{g})$ is a solution to (1), then $(\alpha\mathbf{x}, \mathbf{g}/\alpha)$ is also a solution for any scaling parameter $\alpha \neq 0$. In any event, additional assumptions about $\mathbf{x}$ and $\mathbf{g}$ are required to make BD feasible.

In a Bayesian framework, this is naturally attempted by assigning prior distributions to $\mathbf{x}$ and $\mathbf{g}$ (*e.g.*, [5]–[7]). In particular, Gaussian priors (GP)—along with Bernoulli GP—are extensively used in BD [1], [3], [8]–[10]. However, traditional Gibbs sampling exhibit then state-space exploration issues that have motivated various enhancement techniques [8]–[11]. In [8], a Metropolis-Hastings move is introduced to favor sampling of time-shifted versions of $(\mathbf{x}, \mathbf{g})$. In [9], a K-tuple and a collapsed Gibbs samplers are designed to avoid the support of $\mathbf{g}$ to be stuck in a given configuration. In [11], a scale-sampling step is added in the Gibbs sampler to stimulate the exploration of the scale of $(\mathbf{x}, \mathbf{g})$ while preserving the target distribution. Nonetheless, in absence of theoretical performance guarantees, simulations still needs to be run to assess the efficiency of these methods.

Interestingly, several authors have more recently proposed to constrain $\mathbf{x}$ to have unit ℓ_2-norm [7]. While this sphere-constraint removes scale ambiguity of BD (up to a sign), some results have shown that it actually can facilitate to solve nonconvex optimization problems [4], [12]–[16].

In light of these promising results, we propose in this paper to study the benefit of sphere-constrained prior in Bayesian model for BD problems. In a first attempt, a simple model is considered in what follows. We set $\mathbf{g}$ as *a priori* centered and white Gaussian; the mean direction of $\mathbf{x}$ is *a priori* known, as in many realistic applications (*e.g.*, radar, sonar). We propose a von Mises–Fisher prior (VMP) for $\mathbf{x}$ that narrows the scale ambiguity to a sign ambiguity [17]; we alternatively consider a conventional GP for $\mathbf{x}$ as a baseline scenario. We first derive the theoretical posteriors of $\mathbf{x}$ and $\mathbf{g}$ to discuss the benefits of VMP. We also illustrate state-space exploration issues of the Gibbs sampler with both VMP and GP, even if the re-scaling technique in [11] is used[1].

II. Bayesian Model

A. Observation Model

We consider a linear convolution in presence of noise:

$$\mathbf{y} = \mathbf{x} * \mathbf{g} + \mathbf{n}. \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^L$, $\mathbf{g} \in \mathbb{R}^M$ and $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2\mathbf{I})$ is an additive white Gaussian noise vector with zero mean and known power σ_n^2. The convolution between $\mathbf{x}$ and $\mathbf{g}$ in (1) can be rewritten such that $\mathbf{x} * \mathbf{g} = \mathbf{X}\mathbf{g} = \mathbf{G}\mathbf{x}$ with $\mathbf{X}$ the Toeplitz matrix with first column $[\mathbf{x}^\top, \mathbf{0}_{1,M-1}]^\top$ and first line $[x_1, \mathbf{0}_{1,M-1}]$. $\mathbf{G}$ is built similarly with appropriate dimension. From the observation (1), the likelihood of $\mathbf{x}$ and $\mathbf{g}$ is

$$\mathbf{y}|\mathbf{x}, \mathbf{g} \sim \mathcal{N}\left(\mathbf{X}\mathbf{g}, \sigma_n^2\mathbf{I}\right). \tag{2}$$

BD aims at jointly estimating $\mathbf{x}$ and $\mathbf{g}$ using the likelihood (2). Nonetheless, as previously discussed, (2) is an ill-posed problem due to ambiguities. Thus, we regularize (1) by assigning prior distributions to $\mathbf{x}$ and $\mathbf{g}$.

B. A Priori Model

Herein, $\mathbf{x}$ and $\mathbf{g}$ are supposed to be *a priori* independent. Throughout the paper, we will compare two priors for $\mathbf{x}$:

VMP is the proposed von Mises–Fisher prior [17], characterized by its concentration factor κ and its mean direction

The work of B. Traullé is funded by *École normale supérieure de Rennes*. The work of S. Bidon and D. Roque is supported by DGA/AID under grant 019.65.0068.00.470.75.01 (NEWAVE project).

[1] In our simple model, $\mathbf{g}$ is not sparse so that the technique introduced in [9] is not needed. In addition, in the simulations presented later in Section V, the lengths of $\mathbf{x}$ and $\mathbf{g}$ are small so that the time-shift problem addressed in [8] is not noticeable and does need to be addressed.

	Priors		Joint posteriors
VMP	$\pi(\mathbf{x}) \propto \exp\left\{\kappa\tilde{\mathbf{x}}^\top\mathbf{x}\right\}\mathbb{1}(\mathbf{x})$ (3)	$\pi(\mathbf{g}) \propto \exp\left\{-\frac{\|\mathbf{g}\|^2}{2\sigma_g^2}\right\}$ (5)	$f(\mathbf{x},\mathbf{g}\|\mathbf{y}) \propto \exp\left\{-\frac{\|\mathbf{y}-\mathbf{X}\mathbf{g}\|^2}{2\sigma_n^2}+\kappa\tilde{\mathbf{x}}^\top\mathbf{x}-\frac{\|\mathbf{g}\|^2}{2\sigma_g^2}\right\}\mathbb{1}(\mathbf{x})$ (6)
GP	$\pi(\mathbf{x}) \propto \exp\left\{-\frac{\kappa}{2}\|\mathbf{x}-\tilde{\mathbf{x}}\|^2\right\}$ (4)		$f(\mathbf{x},\mathbf{g}\|\mathbf{y}) \propto \exp\left\{-\frac{\|\mathbf{y}-\mathbf{X}\mathbf{g}\|^2}{2\sigma_n^2}+\frac{\kappa}{2}\|\mathbf{x}-\tilde{\mathbf{x}}\|^2-\frac{\|\mathbf{g}\|^2}{2\sigma_g^2}\right\}$ (7)

	Closed-form posteriors	
	$\mathbf{x}$	$\mathbf{g}$
VMP	$f(\mathbf{x}\|\mathbf{y}) \propto \frac{1}{\sqrt{\|\mathbf{\Sigma}_{\mathbf{y}\|\mathbf{x}}\|}}\exp\left\{-\frac{1}{2}\mathbf{y}^\top\mathbf{\Sigma}_{\mathbf{y}\|\mathbf{x}}^{-1}\mathbf{y}+\kappa\tilde{\mathbf{x}}^\top\mathbf{x}\right\}\mathbb{1}(\mathbf{x})$ (8)	$f(\mathbf{g}\|\mathbf{y}) \propto C_{\text{FB}}(\mathbf{v},\mathbf{A}) \times \exp\left\{-\frac{\|\mathbf{g}\|^2}{2\sigma_g^2}\right\}$ (9)
GP	$f(\mathbf{x}\|\mathbf{y}) \propto \frac{1}{\sqrt{\|\mathbf{\Sigma}_{\mathbf{y}\|\mathbf{x}}\|}}\exp\left\{-\frac{1}{2}\left[\mathbf{y}^\top\mathbf{\Sigma}_{\mathbf{y}\|\mathbf{x}}^{-1}\mathbf{y}+\kappa\|\mathbf{x}\|^2\right]\right\} \times \exp\left\{\kappa\tilde{\mathbf{x}}^\top\mathbf{x}\right\}$ (10)	$f(\mathbf{g}\|\mathbf{y}) \propto \frac{1}{\sqrt{\|\mathbf{\Sigma}_{\mathbf{y}\|\mathbf{g}}\|}}\exp\left\{-\frac{1}{2}\left[\mathbf{y}^\top\mathbf{\Sigma}_{\mathbf{y}\|\mathbf{g}}^{-1}\mathbf{y}+\tilde{\mathbf{x}}^\top\mathbf{G}^\top\mathbf{\Sigma}_{\mathbf{y}\|\mathbf{g}}^{-1}\mathbf{G}\tilde{\mathbf{x}}+\frac{\|\mathbf{g}\|^2}{\sigma_g^2}\right]\right\} \times \exp\left\{\left(\mathbf{\Sigma}_{\mathbf{y}\|\mathbf{g}}^{-1}\mathbf{G}\tilde{\mathbf{x}}\right)^\top\mathbf{y}\right\}$ (11)

	Full conditionals	
	$\mathbf{x}$	$\mathbf{g}$
VMP	$f(\mathbf{x}\|\mathbf{y},\mathbf{g}) \propto \exp\left\{\mathbf{v}^\top\mathbf{x}-\mathbf{x}^\top\mathbf{A}\mathbf{x}\right\}\mathbb{1}(\mathbf{x})$ (12)	$f(\mathbf{g}\|\mathbf{y},\mathbf{x}) \propto \exp\left\{-\frac{(\mathbf{g}-\bar{\mathbf{g}})^\top\mathbf{\Sigma}_{\bar{\mathbf{g}}}^{-1}(\mathbf{g}-\bar{\mathbf{g}})}{2}\right\}$ (14)
GP	$f(\mathbf{x}\|\mathbf{y},\mathbf{g}) \propto \exp\left\{-\frac{(\mathbf{x}-\bar{\mathbf{x}})^\top\mathbf{\Sigma}_{\bar{\mathbf{x}}}^{-1}(\mathbf{x}-\bar{\mathbf{x}})}{2}\right\}$ (13)	

$\tilde{\mathbf{x}}$, *i.e.*, $\mathbf{x} \sim \mathcal{VM}(\kappa,\tilde{\mathbf{x}})$ as detailed in (3) where $\mathbb{1}$ is the indicator function on the L-dimension sphere $\mathbb{S}_{\mathbb{R}}^{L-1}$;

GP is the conventional Gaussian prior; for a fair comparison, we choose $\mathbf{x} \sim \mathcal{N}(\tilde{\mathbf{x}},\mathbf{I}/\kappa)$ — as detailed in (4) — so that the conditional density of $\mathbf{x}$ given $\|\mathbf{x}\| = 1$ boils down to the von Mises–Fisher prior (3) [17, p. 175].

A conventional centered Gaussian prior is chosen for $\mathbf{g}$, *i.e.*, $\mathbf{g} \sim \mathcal{N}(\mathbf{0},\sigma_g^2\mathbf{I})$ with known variance σ_g^2, as recalled in (5).

III. Theoretical Closed-form Posteriors

In this work, as previously discussed, we aim at firstly assessing the impact of VMP on the posteriors of $\mathbf{x}$ and $\mathbf{g}$ compared to that of GP. Given our Bayesian models, posteriors can actually be obtained in closed-forms (up to a normalizing constant). They will be also useful to assess the convergence of the Gibbs sampler described in the next Section.

The posterior of $\mathbf{x}$ (that of $\mathbf{g}$ is obtained similarly) is derived using Bayes theorem as $f(\mathbf{x}|\mathbf{y}) \propto f(\mathbf{y}|\mathbf{x})\pi(\mathbf{x})$ with $f(\mathbf{y}|\mathbf{x}) = \int_{\mathbf{g}} f(\mathbf{y}|\mathbf{x},\mathbf{g})\pi(\mathbf{g})d\mathbf{g}$. Details of calculations are omitted in what follows.

In case of VMP, using (2), (3) and (5), the posteriors of $\mathbf{x}$ and $\mathbf{g}$ respectively result in (8) and (9), with $\mathbf{\Sigma}_{\mathbf{y}|\mathbf{x}} = \sigma_g^2\mathbf{X}\mathbf{X}^\top + \sigma_n^2\mathbf{I}$, $\mathbf{v} = \frac{\mathbf{G}^\top\mathbf{y}}{\sigma_n^2} + \kappa\tilde{\mathbf{x}}$, $\mathbf{A} = \frac{\mathbf{G}^\top\mathbf{G}}{2\sigma_n^2}$. C_{FB} is the normalization constant of the Fisher–Bingham distribution[2] *viz.* $C_{\text{FB}}(\gamma\mathbf{w},\mathbf{B}) = \int_{\mathbb{S}_{\mathbb{R}}^{L-1}} \exp\{\gamma\mathbf{w}^\top\mathbf{x}-\mathbf{x}^\top\mathbf{B}\mathbf{x}\}d\mathbf{x}$ with $\gamma > 0$, $\|\mathbf{w}\| = 1$ and without loss of generality $\mathbf{B}^\top = \mathbf{B}$.

In case of GP, using (2), (4) and (5), the posteriors of $\mathbf{x}$ and $\mathbf{g}$ result in (10) and (11), respectively, with $\mathbf{\Sigma}_{\mathbf{y}|\mathbf{g}} = \kappa^{-1}\mathbf{G}\mathbf{G}^\top + \sigma_n^2\mathbf{I}$.

[2]In Section V, C_{FB} is numerically calculated for $L = 2$ with cylindrical formulation of the integrand.

IV. Monte Carlo Markov Chain Estimation Method

Even though closed-form posteriors are obtained in Section III, it may be arduous to build Bayesian estimators from them. The situation could worsen with more advanced hierarchical Bayesian models. For this reason, a numerical approach is now considered to sample the posterior distributions.

A. Traditional Gibbs Sampler

Gibbs sampling is a Monte Carlo Markov Chain algorithm that consists, in our case, in drawing alternatively $\mathbf{x}$ and $\mathbf{g}$ from their full conditional. After a certain burn-in time (N_{bi} iterations), samples of $\mathbf{x}$ and $\mathbf{g}$ are in theory distributed according to their respective posterior [18]. N_{r} samples are then retained to build histograms and/or compute estimates of $\mathbf{x}$ and $\mathbf{g}$.

Using Bayes theorem, the likelihood (2) and the *a priori* independence between $\mathbf{x}$ and $\mathbf{g}$, we obtain the joint posteriors up to a constant factor for VMP in (6) and for GP in (7). From them, we can directly express the full conditionals of $\mathbf{x}$ and $\mathbf{g}$ as reported in (12)-(14) for VMP and in (13)-(14) for GP. Note that, while (13) and (14) are simply simulated as Gaussian distributions, (12) is recognized as a Fisher–Bingham distribution sampled according to [19]. Parameters of these distributions are $\mathbf{v}$ and $\mathbf{A}$ from (9), $\mathbf{\Sigma}_{\bar{\mathbf{x}}} = (2\mathbf{A}+\kappa\mathbf{I})^{-1}$, $\bar{\mathbf{x}} = \mathbf{\Sigma}_{\bar{\mathbf{x}}}\mathbf{v}$, $\mathbf{\Sigma}_{\bar{\mathbf{g}}} = \left(\frac{\mathbf{X}^\top\mathbf{X}}{\sigma_n^2}+\frac{1}{\sigma_g^2}\mathbf{I}\right)^{-1}$, $\bar{\mathbf{g}} = \mathbf{\Sigma}_{\bar{\mathbf{g}}}\frac{\mathbf{X}^\top\mathbf{y}}{\sigma_n^2}$.

B. Poor Mixing Property and Workaround

In BD, the traditional Gibbs sampler suffers from state-space exploration issues [8]–[11]. We notably observe correlated chains with local modes that can be difficult to escape, leading to a very slow convergence towards the target distribution. In a first attempt to deal with this problem (still

persistent with VMP and GP), we implement in both Bayesian models the scale-sampling method described in [11]. The latter is actually tantamount to a Reversible Jump Monte Carlo Markov Chain (RJMCMC) step [20], [21]—that preserves the target distribution—with a unitary acceptance probability. In particular, after sampling $\mathbf{x}$ and $\mathbf{g}$, a scaling factor is sampled from the pdf

$$f(s|\mathbf{x},\mathbf{g}) \propto |s|^{L-M-1}\pi(s\mathbf{x})\pi(\mathbf{g}/s), \tag{16}$$

and then applied as $\{\mathbf{x},\mathbf{g}\} \leftarrow \{s\mathbf{x},\mathbf{g}/s\}$.

1) Scaling factor with VMP: Using (3) and (5), the scaling factor pdf (16) is shown to be non-zero iff $|s| = 1$ and depends only on $\mathbf{x}$. Its probability mass function (pmf) is $P(s=1|\mathbf{x}) = 1/(1+\exp\{-2\kappa\tilde{\mathbf{x}}^\top\mathbf{x}\})$. In other words, the scale-sampling reduces to a sign-sampling.

2) Scaling factor with GP: Using (4) and (5), the scaling factor pdf (16) reduces to

$$f(s|\mathbf{x},\mathbf{g}) \propto |s|^{L-M-1}\exp\left\{\kappa\tilde{\mathbf{x}}^\top\mathbf{x}s - \kappa\frac{\|\mathbf{x}\|^2}{2}s^2 - \frac{\|\mathbf{g}\|^2}{2\sigma_g^2}s^{-2}\right\}.$$

We sample it as $s \triangleq \lambda r$ with joint pdf

$$f(r,\lambda|\mathbf{x},\mathbf{g}) \propto r^{L-M-1}\exp\left\{\kappa\tilde{\mathbf{x}}^\top\mathbf{x}\lambda r - \kappa\frac{\|\mathbf{x}\|^2}{2}r^2 - \frac{\|\mathbf{g}\|^2}{2\sigma_g^2}r^{-2}\right\},$$

where $r > 0$ and $\lambda \in \{-1,1\}$. To that end, we implement a hybrid Gibbs sampler that iterates between $r|\lambda,\mathbf{x},\mathbf{g}$ and $\lambda|r,\mathbf{x},\mathbf{g}$. The latter follows a Bernoulli distribution with pmf $P(\lambda=1|r,\mathbf{x}) = 1/(1+\exp\{-2\kappa\tilde{\mathbf{x}}^\top\mathbf{x}r\})$. The former does not follow a common distribution from which standard sampling method exists. Nevertheless, a polynomial analysis of this distribution shows it can be strictly log-concave (thus unimodal), otherwise either unimodal or bimodal. For the former case, we choose a Metropolis-adjusted Langevin algorithm (MALA) [22], [23] that is adequate for strictly log-concave distribution. Otherwise, alternate sampling techniques can be used such as slice-sampling [24]. Note that the proposed hybrid Gibbs sampler allows the scaling factor s to be updated both conditionally or independently of its previous values [25].

C. Augmented Gibbs Sampler Algorithm

Finally, the augmented Gibbs sampler implemented in Section V for both VMP and GP is summarized in Algorithm 1.

Algorithm 1 (Gibbs sampler with scaling step)**.**
given $\mathbf{x}^{(k-1)}$ and $\mathbf{g}^{(k-1)}$, sample:
1: $\mathbf{x}^{(k)}$ from $f\left(\mathbf{x}|\mathbf{y},\mathbf{g}^{(k-1)}\right)$
2: $\mathbf{g}^{(k)}$ from $f\left(\mathbf{g}|\mathbf{y},\mathbf{x}^{(k)}\right)$
3: s from (16)
4: Update $\{\mathbf{x}^{(k)},\mathbf{g}^{(k)}\} \leftarrow \{s\mathbf{x}^{(k)},\mathbf{g}^{(k)}/s\}$

V. Simulation and Results

For illustrative purposes, we adopt a 2-dimensional space for $\mathbf{x}$ end $\mathbf{g}$, *i.e.*, $L = M = 2$, as it makes more convenient representations of posteriors. Nevertheless, same observations could be made for higher dimensions and more realistic scenarios, especially for the state-space exploration issue.

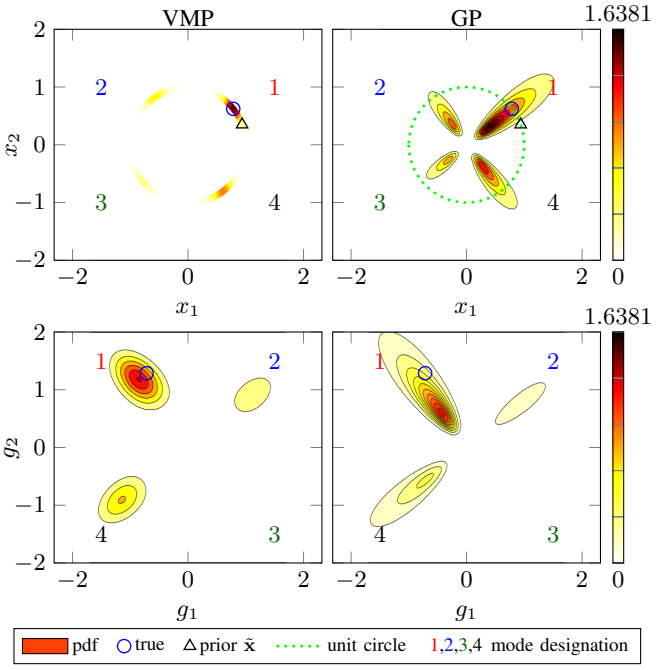

Fig. 1. Theoretical posteriors of $\mathbf{x}$ (top) and $\mathbf{g}$ (bottom) with VMP (left) and GP (right). #1 is the major mode. Counterclockwise direction modes numbering of $f(\mathbf{x}|\mathbf{y})$. Clockwise direction for $f(\mathbf{g}|\mathbf{y})$.

To generate the signal $\mathbf{y}$ in (1), we set $\mathbf{x} = [0.7826, 0.6226]^\top \in \mathbb{S}_{\mathbb{R}}^{L-1}$ and $\mathbf{g} = [-0.7182, 1.2861]^\top$ is drawn from its prior (5). We define the signal-to-noise ratio as $\text{SNR} = \mathbb{E}_{|\mathbf{x}}\{\|\mathbf{x}*\mathbf{g}\|^2\}/\mathbb{E}\{\|\mathbf{n}\|^2\} = M\sigma_g^2/((L+M-1)\sigma_n^2)$ to generate a noise vector at 18 dB. A realization gives $\mathbf{n} = -[0.1423, 0.1656, 0.0164]^\top$. Prior parameters are set as $\tilde{\mathbf{x}} = \mathbf{\Omega}\mathbf{x}$, $\kappa = 1$ and $\sigma_g^2 = 5$, where $\mathbf{\Omega}$ is the rotation matrix of angle $-\pi/10$.

A. Theoretical Posteriors

Theoretical posteriors (8)-(9) and (10)-(11) are depicted in Fig. 1. The posteriors of $\mathbf{x}$ have 4 apparent modes both with VMP and GP. It is also the case for $\mathbf{g}$ even if the color-scale makes mode #3 not clearly visible. Mode #1 is the major mode, mode #4 corresponds to a sign ambiguity and modes #2 and #3 are induced by a rotational ambiguity. As expected, we have observed that these modes tend to be more peaked (resp. spread) as the SNR increases (resp. decreases). In addition the higher κ, the more predominant becomes mode #1 that, at the same time, gets closer to $\tilde{\mathbf{x}}$. Interestingly, the VMP posterior of $\mathbf{x}$ in (8) stems directly from the conditioning to $\|\mathbf{x}\| = 1$ of the GP posterior (10)[3]. Using a prior on the sphere alleviates as expected the scale ambiguity inherent to (1) for $\mathbf{x}$, but also decreases the radial uncertainty for $\mathbf{g}$. Thus, using VMP instead of GP could be relevant to BD problems.

[3]It is in fact the case for all the pdf of $\mathbf{x}$: (3)-(4) and (6)-(7).

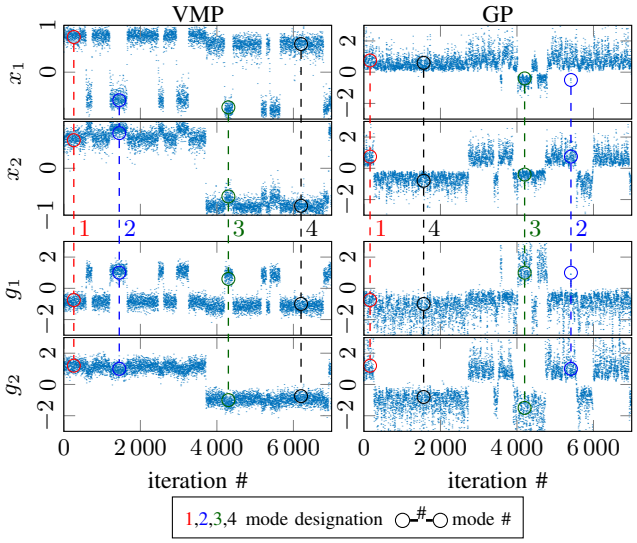

Fig. 2. Chain of samples of $f(\mathbf{x}|\mathbf{y})$ (top) and $f(\mathbf{g}|\mathbf{y})$ (bottom), with VMP (left) and GP (right). Same modes numbering as in Fig. 1 – Traditional Gibbs sampler.

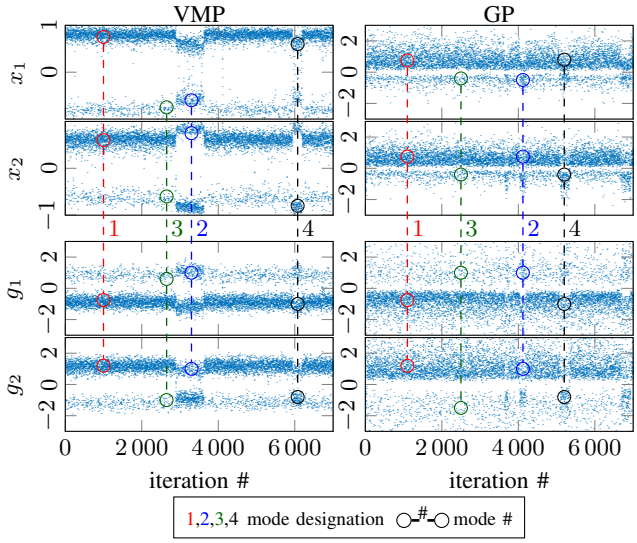

Fig. 3. Chain of samples of $f(\mathbf{x}|\mathbf{y})$ (top) and $f(\mathbf{g}|\mathbf{y})$ (bottom), with VMP (left) and GP (right). Same modes numbering as in Fig. 1 – Augmented Gibbs sampler.

B. Empirical Posteriors and Gibbs Sampler Behavior

We run both the traditional and the augmented Gibbs sampler with VMP and GP. Samples from the first 7000 iterations are represented in Fig. 2 for the traditional Gibbs sampler, and in Fig. 3 for the augmented. Both samplers are initialized with $\mathbf{x}^{(0)} \leftarrow \tilde{\mathbf{x}}$ to force the exploration of the main mode #1 at the first iteration, and then to observe the capability to explore other modes during sampling. As shown in Fig. 3 compared with Fig. 2, the augmented sampler for GP does avoid correlated chains "conditionally to a mode", but does not prevent from lingering in a given mode whether a von Mises–Fisher or a Gaussian prior for $\mathbf{x}$ is proposed. For instance, the sampling chains here explore modes #1 and #3 for approximately more than 3000 iterations within the augmented Gibbs sampler. We also observe that the sampling of $(\mathbf{x}, \mathbf{g})$ occurs by pair of "opposite" modes: modes $\#(1,3)$ and modes $\#(2,4)$ when the re-scaling technique is done. As a matter of fact, in both VMP and GP the scaling factor sampled from (16) has a signed value, namely $s \in \{-1, 1\}$ for VMP and $\lambda \triangleq s/|s| \in \{-1, 1\}$ for GP. Hence, if $(\mathbf{x}, \mathbf{g})$ is near mode #1, then its scaled version $(s\mathbf{x}, \mathbf{g}/s)$ is either near mode #1 or #3. In any event, being trapped in modes makes the empirical posteriors of $\mathbf{x}$ and $\mathbf{g}$ highly dependent on the chain length and the initialization. As exemplified in Fig. 4 for the traditional Gibbs sampler where modes #3 and #4 with VMP, modes #2 and #3 with GP, are not sampled yet (cf. Fig. 2), and in Fig. 5 for the augmented where modes #2 and #4 are not sampled yet (cf. Fig. 3) for both VMP and GP — with $N_{\text{bi}} = 100$ and $N_{\text{r}} = 3000$. It can thus lead to poor estimates of the theoretical posteriors, even if the re-scaling technique is used.

VI. Conclusion

In this paper, we studied the BD problem in a Bayesian context. We assigned a von Mises–Fisher prior to one of the two signals of interest; the other being *a priori* Gaussian. The scale ambiguity is thus removed (up to a sign); thereby greatly contributing to BD's regularization compared to traditional Gaussian priors. This is supported *via* the derivation of the posteriors in closed-form as well as their graphical interpretation. Nevertheless, we showed that a Gibbs sampler-based BD still suffers from state-space exploration issues, even if the resampling technique in [11] is used. These results are particularly exacerbated at high SNR and with low concentration parameter for the von Mises–Fisher prior; they suggest the design of sampling methods tailored to the sphere-constrained BD problem.

References

[1] C. Lin, C. Mailhes, and J.-Y. Tourneret, "P- and T-wave delineation in ECG signals using a Bayesian approach and a partially collapsed Gibbs sampler," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2840–2849, Dec. 2010.

[2] D. Ge, E. Le Carpentier, J. Idier, and D. Farina, "Spike sorting by stochastic simulation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 19, no. 3, pp. 249–259, Feb. 2011.

[3] Q. Cheng, R. Chen, and T.-H. Li, "Simultaneous wavelet estimation and deconvolution of reflection seismic signals," *IEEE Trans. Geosci. Remote Sens.*, vol. 34, no. 2, pp. 377–384, Mar. 1996.

[4] Y. Li and Y. Bresler, "Multichannel sparse blind deconvolution on the sphere," *IEEE Trans. Inf. Theory*, vol. 65, no. 11, pp. 7415–7436, July 2019.

[5] A. Doucet and P. Duvaut, "Bayesian estimation of state-space models applied to deconvolution of Bernoulli–Gaussian processes," *Signal Process.*, vol. 57, no. 2, pp. 147–161, Mar. 1997.

[6] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding blind deconvolution algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2354–2367, July 2011.

[7] D. Wipf and H. Zhang, "Revisiting Bayesian blind deconvolution," *J. Mach. Learn. Res.*, vol. 15, pp. 3595–3634, 2014.

[8] C. Labat and J. Idier, "Sparse blind deconvolution accounting for time-shift ambiguity," in *IEEE Int. Conf. on Acoust. Speech and Signal Process.*, May 2006, vol. 3, pp. 616–619.

[9] D. Ge, J. Idier, and E. Le Carpentier, "Enhanced sampling schemes for MCMC based blind Bernoulli–Gaussian deconvolution," *Signal Process.*, vol. 91, no. 4, pp. 759–772, Apr. 2011.

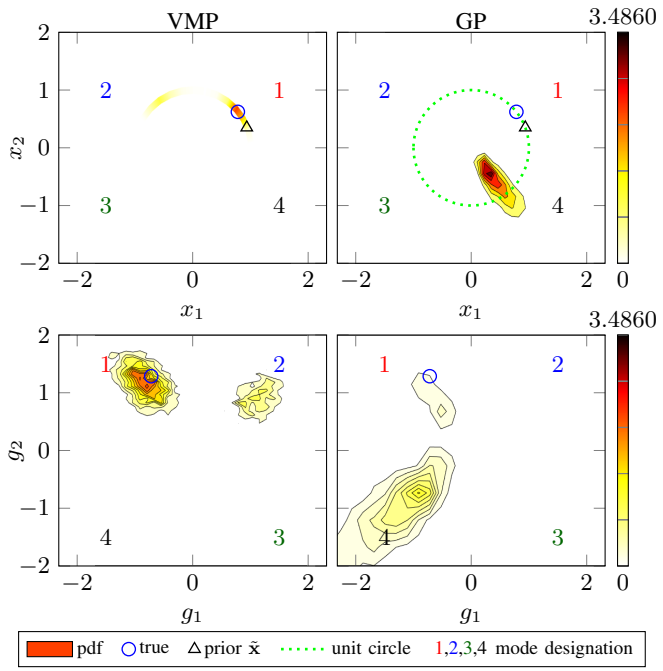

Fig. 4. Histograms of $\{\mathbf{x}^{(k)}\}$ (top) and $\{\mathbf{g}^{(k)}\}$ (bottom) with VMP (left) and GP (right) ($N_{\text{bi}} = 100$, $N_{\text{r}} = 3000$). Same modes numbering as in Fig. 1 – Traditional Gibbs sampler.

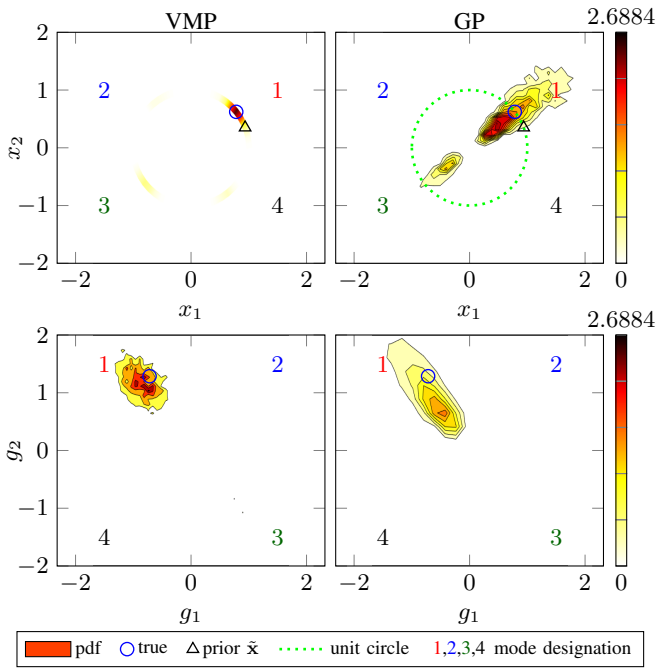

Fig. 5. Histograms of $\{\mathbf{x}^{(k)}\}$ (top) and $\{\mathbf{g}^{(k)}\}$ (bottom) with VMP (left) and GP (right) ($N_{\text{bi}} = 100$, $N_{\text{r}} = 3000$). Same modes numbering as in Fig. 1 – Augmented Gibbs sampler.

[10] G. Kail, J.-Y. Tourneret, F. Hlawatsch, and N. Dobigeon, "Blind deconvolution of sparse pulse sequences under a minimum distance constraint: A partially collapsed Gibbs sampler method," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2727–2743, June 2012.

[11] T. Veit and J Idier, "Rééchantillonnage de l'échelle dans les algorithmes MCMC pour les problèmes inverses bilinéaires," *Traitement du Signal*, vol. 25, no. 4, pp. 329–344, 2008.

[12] H.-W. Kuo, Y. Zhang, Y. Lau, and J. Wright, "Geometry and symmetry in short-and-sparse deconvolution," *SIAM J. Math. Data Sci.*, vol. 2, no. 1, pp. 216–245, Feb. 2020.

[13] Y. Lau, Q. Qu, H.-W. Kuo, P. Zhou, Y. Zhang, and J. Wright, "Short-and-sparse deconvolution – a geometric approach," arXiv: 1908.10959v2, 2019.

[14] Q. Qu, X. Li, and Z. Zhu, "Exact and efficient multi-channel sparse blind deconvolution – a nonconvex approach," in *IEEE Asilomar Conf. on Signals, Syst., and Comput.*, Nov. 2019, pp. 640–644.

[15] Y. Zhang, H.-W. Kuo, and J. Wright, "Structured local optima in sparse blind deconvolution," *IEEE Trans. Inf. Theory*, vol. 66, no. 1, pp. 419–452, Sept. 2020.

[16] Y. Zhang, Y. Lau, H.-W. Kuo, S. Cheung, A. Pasupathy, and J. Wright, "On the global geometry of sphere-constrained sparse blind deconvolution," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 3, pp. 1–1, Sept. 2019.

[17] K. V. Mardia and P. E. Jupp, *Directional Statistics*, Wiley Series in Probability and Statistics. Wiley, 2008.

[18] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer–Verlag, 2004.

[19] P. D. Hoff, "Simulation of the matrix Bingham–von Mises–Fisher distribution, with applications to multivariate and relational data," *J. Comput. Graph. Stat.*, vol. 18, no. 2, pp. 438–456, Jan. 2012.

[20] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, Dec. 1995.

[21] R. Waagepetersen and D. Sorensen, "A tutorial on reversible jump MCMC with a view toward applications in QTL-mapping," *Int. Stat. Rev.*, vol. 69, no. 1, pp. 49–61, Dec. 2001.

[22] G. O. Roberts and R. L. Tweedie, "Exponential convergence of Langevin distributions and their discrete approximations," *Bernoulli*, vol. 2, no. 4, pp. 341–363, Dec. 1996.

[23] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. O. Hero, and S. McLaughlin, "A survey of stochastic simulation and optimization methods in signal processing," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 2, pp. 224–241, Mar. 2016.

[24] Radford M. Neal, "Slice sampling," *The Annals of Statistics*, vol. 31, no. 3, jun 2003.

[25] S. P. Brooks, P. Giudici, and G. O. Roberts, "Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions," *J. R. Stat. Soc. Series B: Statistical Methodology*, vol. 65, no. 1, pp. 3–39, Jan. 2003.