Robust adaptive beamforming using a Bayesian steering vector error model

Olivier Besson *, Stéphanie Bidon

University of Toulouse-ISAE, Department of Electronics, Optronics and Signal, 10 Avenue Edouard Belin, 31055 Toulouse, France

ARTICLE INFO

Keywords:
Robust adaptive beamforming
Bayesian estimation
Bingham distribution
Gibbs sampling

ABSTRACT

We propose a Bayesian approach to robust adaptive beamforming which entails considering the steering vector of interest as a random variable with some prior distribution. The latter can be tuned in a simple way to reflect how far is the actual steering vector from its presumed value. Two different priors are proposed, namely a Bingham prior distribution and a distribution that directly reveals and depends upon the angle between the true and presumed steering vector. Accordingly, a non-informative prior is assigned to the interference plus noise covariance matrix **R**, which can be viewed as a means to introduce diagonal loading in a Bayesian framework. The minimum mean square distance estimate of the steering vector as well as the minimum mean square error estimate of **R** are derived and implemented using a Gibbs sampling strategy. Numerical simulations show that the new beamformers possess a very good rate of convergence even in the presence of steering vector errors.

1. Introduction and problem statement

Designing robust adaptive beamformers is a major requirement of most practical systems where one is most likely faced with partially unknown array characteristics [1,2]. Steering vectors errors due e.g., to partially uncalibrated arrays, uncertainties about the direction of arrival, pointing errors are known to be very detrimental to conventional adaptive beamformers unless some proper action is taken. This is especially the case when the signal of interest (SOI) is present in the measurements [3–6]. Indeed, any input signal whose steering vector differs from the presumed SOI steering vector is deemed an interference, and hence should be suppressed. This results in the self-nulling phenomenon where the adaptive beamformer tends to place nulls towards the SOI. Accordingly, limited observation time is another limitation of minimum power distortionless response (MPDR) beamformers since errors in the estimation of the interference plus noise covariance matrix result in a significant increase of the number of snapshots required to come close to the optimal beamformer. Thus it comes at no surprise that the literature about robust adaptive beamforming (RAB) is abundant, see e.g., [1,2,6]. Our aim is not here to provide an exhaustive review of the literature, rather give a short overview of the main approaches available so far.

The first approach that undoubtedly comes to mind when dealing with steering vector errors or a limited number of snapshots is diagonal loading [7–9]. Notwithstanding the issue of setting the loading level, diagonal loading (DL) is a simple yet powerful method to mitigate steering vector errors and/or small sample size. In addition to be a natural way to regularize the covariance matrix estimation or to enforce a desired value for the white noise array gain [10,6], diagonal loading also emerges as the solution to RAB approaches which, from their very principle, do not enforce diagonal loading. For instance, the robust Capon beamformers of [11–13] which

* Corresponding author. Tel.: +33 561339125.
E-mail addresses: olivier.besson@isae.fr (O. Besson), stephanie.bidon@isae.fr (S. Bidon).

either entail minimizing the output power subject to the SOI steering vector belonging to a sphere centered around its presumed value, or maintaining a minimum gain within this sphere, boil down to diagonal loading. The doubly constrained Capon beamformer of [14] which imposes a norm constraint of the steering vector also results in a DL-type beamformer. The main difference lies in the fact that the loading level is now computed as a function of the sphere radius. This issue of determining an optimal loading level has attracted a great deal of attention and, recently, some references have addressed the problem of computing automatically the loading level, see e.g., [15,16]. In a similar vein albeit with a different methodology, the idea of ensuring a given gain around the presumed steering vector using some constraints forms the basis of many methods, see e.g., [17–19]. One difference between the aforementioned methods [11–13] and the original DL is that an estimate of the SOI steering vector is made available. Indeed, SOI steering vector estimation is also a viable approach to combat steering vector errors, capitalizing on a better estimate of the SOI steering vector. Refs. [20–22] are a few examples of such an approach.

The references presented so far mostly adopt a deterministic view of steering vector errors. However, stochastic modeling of steering vector errors has been addressed in some references. For instance, [23] assumes that the difference between the actual and presumed steering vector is a random quantity with a given distribution. Then, [23] considers minimizing the output power under the constraint that the gain towards the actual steering vector be larger than one with a given probability. Ref. [24] also addresses the design of an adaptive beamformer under a Gaussian hypothesis for the steering vector errors. A penalizing term is introduced in the output power minimization that takes into account the distribution of the steering vector. In [25] Kristine Bell considers a Bayesian approach to RAB with uncertainties in direction of arrival, modeling the latter as a random variable with a given prior distribution. A Bayesian beamformer is derived, which amounts to a weighted sum of Wiener filters where the weights depend on the posterior distribution of each pointing direction. A Bayesian beamformer is also derived in [26] based on maximum a posteriori or minimum mean square error estimation of the SOI waveform in the case of a Gaussian distributed SOI steering vector.

In this paper, we also address the problem of designing an adaptive beamformer in the case of steering vector uncertainties. The latter are considered as random with some prior distribution. A Bayesian framework is thus formulated and the minimum mean square distance estimation of the steering vector as well as the minimum mean square error estimate of the interference plus noise covariance matrix are derived. They are then used to compute a beamformer which hopefully could perform well in spite of the presence of steering vector errors.

2. Data model and assumptions

In this section, we state the assumptions regarding our data model. We assume that K snapshots are received on the array, which can be written as

$$\mathbf{z}_k = \alpha_k^* \mathbf{v} + \mathbf{n}_k; \quad k = 1, \ldots, K \tag{1}$$

where

- $\mathbf{z}_k \in \mathbb{C}^{N\times 1}$ is the output of the array (snapshot) at time k. In the sequel we let $\mathbf{Z} = [\mathbf{z}_1 \ \cdots \ \mathbf{z}_K]$ denote the data matrix.
- $\mathbf{v}$ is the signal of interest signature and is assumed to be a random vector with some prior distribution $\pi(\mathbf{v})$. Two different priors will be considered in the sequel, namely a Bingham distribution and a distribution which depends directly on the angle between $\mathbf{v}$ and its presumed value $\overline{\mathbf{v}}$.
- the interference plus noise vectors $\mathbf{n}_k$ are assumed to be independent, complex-valued Gaussian distributed, with zero-mean and covariance matrix $\mathbf{R}$, i.e.,

$$p(\mathbf{n}_k|\mathbf{R}) = \pi^{-N}|\mathbf{R}|^{-1}\exp\{-\mathbf{n}_k^H\mathbf{R}^{-1}\mathbf{n}_k\} \tag{2}$$

  where $|\cdot|$ stands for the determinant of a matrix. Herein, we assume that $\mathbf{R}$ is a random matrix, drawn from an inverse Wishart distribution [27,28] with mean $\mu\mathbf{I}_N$ and ν degrees of freedom, viz

$$\pi(\mathbf{R}|\nu,\mu) \propto |\mathbf{R}|^{-(\nu+N)}\mathrm{etr}\{-(\nu-N)\mu\mathbf{R}^{-1}\} \tag{3}$$

  where $\propto$ means proportional to and $\mathrm{etr}\{\cdot\}$ stands for the exponential of the trace of the matrix between braces. The choice of an inverse Wishart distribution for $\mathbf{R}$ is mostly due to the fact that it is conjugate with respect to the Gaussian distribution of the snapshots in (2). This is a usual choice in Bayesian estimation and it facilitates mathematical derivation of the posterior distributions. Note however that the distribution in (3) is *non-informative* as it is a maximum entropy prior distribution subject to $\mathcal{E}\{\mathrm{Tr}\{\mathbf{R}^{-1}\}\} = c_1$ and $\mathcal{E}\{\log|\mathbf{R}|\} = c_2$ [29], and it does not depend on any prior covariance matrix. This prior is mainly aimed at increasing robustness of the adaptive beamformer. Indeed, (3) means that $\mathbf{R}$ should be close to a scaled identity matrix and is tantamount to introducing diagonal loading in the beamformer calculation, see below for further details.
- the amplitudes α_k are assumed to be independent and identically distributed according to a complex Gaussian distribution with zero mean and (known) variance σ_α^2, i.e.,

$$\pi(\boldsymbol{\alpha}|\sigma_\alpha^2) = \prod_{k=1}^{K} \pi^{-1}\sigma_\alpha^{-2}\exp\{-\sigma_\alpha^{-2}|\alpha_k|^2\} \propto \exp\{-\sigma_\alpha^{-2}\boldsymbol{\alpha}^H\boldsymbol{\alpha}\} \tag{4}$$

  where $\boldsymbol{\alpha} = [\alpha_1 \ \cdots \ \alpha_K]^T$. We denote this distribution as $\boldsymbol{\alpha} \sim \mathrm{CN}(\mathbf{0}, \sigma_\alpha^2\mathbf{I}_K)$ and note that it is conjugate with respect to the conditional Gaussian distribution of $\mathbf{z}_k$. The assumption that σ_α^2 is known means that we have a rough idea of the signal power level, which is reasonable in numerous applications. In Appendix B we extend the algorithms developed below to the case of an unknown random σ_α^2. Additionally, in the numerical section, we study the robustness of our algorithms to a non-perfect knowledge of σ_α^2.

The statistical model is thus summarized by the likelihood function

$$p(\mathbf{Z}|\boldsymbol{\alpha},\mathbf{v},\mathbf{R}) = \pi^{-NK}|\mathbf{R}|^{-K}\text{etr}\{-(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H)^H\mathbf{R}^{-1}(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H)\} \quad (5)$$

and the prior distributions $\pi(\mathbf{v})$, $\pi(\mathbf{R}|\nu,\mu)$ and $\pi(\boldsymbol{\alpha}|\sigma_\alpha^2)$. Our objective is, from the data measurements $\mathbf{Z}$ and given the statistical assumptions described above, to obtain a beamformer aimed at recovering the SOI with maximum signal to interference and noise ratio (SINR). Since the optimal filter is $\mathbf{w}_{\text{opt}} \propto \mathbf{R}^{-1}\mathbf{v}$, we will obtain estimates of both $\mathbf{v}$ and $\mathbf{R}$, so as to approach $\mathbf{w}_{\text{opt}}$ as closely as possible.

3. Bingham model

As stated previously, we assume that $\mathbf{v}$ is a random vector which is close to a nominal steering vector $\overline{\mathbf{v}}$. First note that any scaling factor affecting $\mathbf{v}$ can be reported in σ_α^2. Herein, we assume that $\|\mathbf{v}\| = 1$, i.e., $\mathbf{v} \in \mathcal{S}_N$, where $\mathcal{S}_N$ is the unit sphere in $\mathbb{C}^N$. We further assume that $\mathbf{v}$ follows a complex Bingham distribution (see [30,31] for definition and properties of real Bingham distributions, [32,33] for complex Bingham distributions and Appendix A for a brief overview and an extension to complex Bingham von Mises Fisher distributions), i.e.,

$$\pi(\mathbf{v}) \propto \exp\{\kappa|\mathbf{v}^H\overline{\mathbf{v}}|^2\} \quad (6)$$

where κ is a positive scalar and $\overline{\mathbf{v}} \in \mathcal{S}_N$ is the nominal steering vector. The distribution in (6) is often referred to as the complex Watson distribution [33] which is a special case of the complex Bingham distribution. Note that the distribution in (6) depends on $\cos^2\theta$ where θ stands for the angle between $\mathbf{v}$ and $\overline{\mathbf{v}}$: $\pi(\mathbf{v})$ is thus constant for any vector lying on a cone whose axis is $\overline{\mathbf{v}}$ and whose aperture is θ. The scalar κ serves as a concentration parameter: the larger κ the closer $\mathbf{v}$ and $\overline{\mathbf{v}}$. Therefore, κ reflects our knowledge of the amplitude of steering vector uncertainties. More precisely, using changes of variables similar to those in [32,33], it is possible to show that the probability density function (pdf) of θ is given by

$$p(\theta) = \frac{\kappa^{N-1}\exp\{-\kappa\}}{\gamma(N-1,\kappa)}\sin(2\theta)(\sin^2\theta)^{N-2}\exp\{\kappa\cos^2\theta\} \quad (7)$$

where $\gamma(a,x) = \int_0^x t^{a-1}\exp\{-t\}\,dt$ is the incomplete Gamma function [34]. Therefore, the choice of the complex Watson distribution in (6) and a given value of κ lead to a given distribution for θ. Moreover, it is straightforward to show that

$$\mathcal{E}\{|\mathbf{v}^H\overline{\mathbf{v}}|^2\} = 1 - \frac{1}{\kappa}\frac{\gamma(N,\kappa)}{\gamma(N-1,\kappa)} \quad (8)$$

Although $\mathcal{E}\{|\mathbf{v}^H\overline{\mathbf{v}}|^2\}$ is not exactly the average square distance between $\mathbf{v}$ and $\overline{\mathbf{v}}$, it is close to, see the discussion below. Hence, the previous formula indicates that $\mathcal{E}\{|\mathbf{v}^H\overline{\mathbf{v}}|^2\}$ is roughly proportional to κ^{-1}, and hence that κ is indeed a concentration parameter. This is illustrated in Fig. 1 where we plot $p(\theta)$ for different values of κ.

The statistical description of our model is now complete and we turn to the estimation of $\mathbf{v}$ and $\mathbf{R}$, in order to compute a filter $\mathbf{w} \propto \hat{\mathbf{R}}^{-1}\hat{\mathbf{v}}$ which is hopefully close to the optimal filter $\mathbf{w}_{\text{opt}} \propto \mathbf{R}^{-1}\mathbf{v}$. Prior to that, a few observations are in order regarding the estimation of $\mathbf{v}$. Note that $\mathbf{v}$ is not an arbitrary vector in $\mathbb{C}^N$ but lies on the unit sphere, and hence its estimate should also inherit this property. However, the natural metric on the unit sphere is not the mean square error between $\hat{\mathbf{v}}$ and $\mathbf{v}$: hence the usual minimum mean square error (MMSE) approach should be revisited [35,36]. Indeed, the natural distance on the sphere is given by the angle between $\hat{\mathbf{v}}$ and $\mathbf{v}$ and, therefore, it seems logical to estimate $\mathbf{v}$ by minimizing this angle. However the latter is given by $\arccos|\hat{\mathbf{v}}^H\mathbf{v}|$ and it appears intractable to obtain an expression for the estimator that minimizes this distance. In contrast, minimizing the average square sine angle between $\hat{\mathbf{v}}$ and $\mathbf{v}$ leads in a straightforward way to

$$\begin{aligned}\mathbf{v}_{\text{mmsd}} &= \arg\min_{\hat{\mathbf{v}}} \mathcal{E}\{\sin^2(\hat{\mathbf{v}},\mathbf{v})\} = \arg\max_{\hat{\mathbf{v}}} \mathcal{E}\{|\hat{\mathbf{v}}^H\mathbf{v}|^2\} \\ &= \arg\max_{\hat{\mathbf{v}}} \int\left[\int |\hat{\mathbf{v}}^H\mathbf{v}|^2 p(\mathbf{v}|\mathbf{Z})\,d\mathbf{v}\right]p(\mathbf{Z})\,d\mathbf{Z} \\ &= \arg\max_{\hat{\mathbf{v}}} \hat{\mathbf{v}}^H\left[\int \mathbf{v}\mathbf{v}^H p(\mathbf{v}|\mathbf{Z})\,d\mathbf{v}\right]\hat{\mathbf{v}} \\ &= \mathcal{P}\left\{\int \mathbf{v}\mathbf{v}^H p(\mathbf{v}|\mathbf{Z})\,d\mathbf{v}\right\} \end{aligned} \quad (9)$$

where $\mathcal{P}\{\cdot\}$ stands for the principal eigenvector of the matrix between braces. With a slight abuse of language, we refer to (9) as the minimum mean square distance (MMSD) estimate of $\mathbf{v}$: it does not minimize the true distance but, for small θ, $\sin\theta \simeq \theta$, and hence (9) is meaningful. The MMSD estimator of $\mathbf{v}$ thus amounts to computing the principal eigenvector of the posterior mean of the projection matrix $\mathbf{v}\mathbf{v}^H$. The problem associated with this approach is that we do not know how to obtain an expression for $p(\mathbf{v}|\mathbf{Z})$: marginalizing $p(\mathbf{Z}|\boldsymbol{\alpha},\mathbf{v},\mathbf{R})$ with respect to (w.r.t.) $\boldsymbol{\alpha}$ is straightforward – see (12) below – but further marginalizing w.r.t. $\mathbf{R}$ is intractable. Furthermore, $\mathbf{v}$ is not the only parameter of interest as we also need to estimate $\mathbf{R}$. In contrast, $\boldsymbol{\alpha}$ can be considered as a nuisance parameter, which is not necessarily to be estimated. Therefore, the first approach that crosses one's

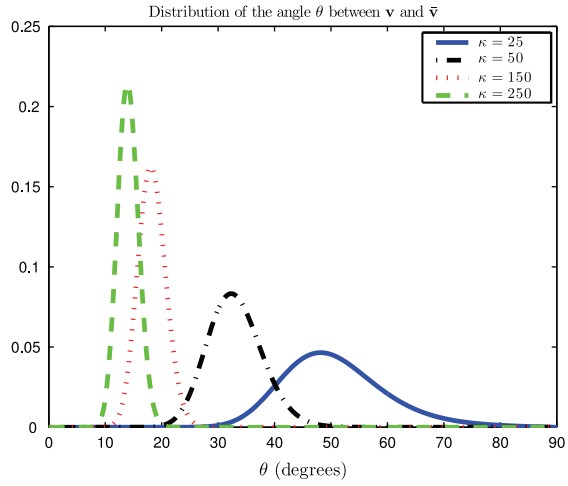

Fig. 1. Distribution of the angle between $\mathbf{v}$ and $\overline{\mathbf{v}}$, when $\mathbf{v}$ is drawn from (6), for different values of κ. N=16.

mind consists in computing the MMSD estimator of $\mathbf{v}$ and the MMSE estimator of $\mathbf{R}$ which is given by

$$\mathcal{E}\{\mathbf{R}|\mathbf{Z}\} = \int \mathbf{R}p(\mathbf{R}|\mathbf{Z})\, d\mathbf{R} \tag{10}$$

The first step towards obtaining these estimates is to marginalize w.r.t. $\boldsymbol{\alpha}$ in order to derive the joint posterior distribution $p(\mathbf{v},\mathbf{R}|\mathbf{Z})$ of $\mathbf{v}$ and $\mathbf{R}$ only. Observing that

$$\begin{aligned}
&\sigma_\alpha^{-2}\boldsymbol{\alpha}^H\boldsymbol{\alpha}+\mathrm{Tr}\{(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H)^H\mathbf{R}^{-1}(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H)\}\\
&\quad=\sigma_\alpha^{-2}\boldsymbol{\alpha}^H\boldsymbol{\alpha}+\mathrm{Tr}\{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{Z}\}-\mathbf{v}^H\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}\\
&\qquad-\boldsymbol{\alpha}^H\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}+(\boldsymbol{\alpha}^H\boldsymbol{\alpha})(\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})\\
&\quad=(\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})\left\|\boldsymbol{\alpha}-\frac{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}}{\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}\right\|^2\\
&\qquad+\mathrm{Tr}\left\{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{Z}-\frac{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}\mathbf{v}^H\mathbf{R}^{-1}\mathbf{Z}}{\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}\right\}
\end{aligned} \tag{11}$$

it follows that:

$$\begin{aligned}
p(\mathbf{Z}|\mathbf{v},\mathbf{R}) &= \int p(\mathbf{Z}|\boldsymbol{\alpha},\mathbf{v},\mathbf{R})\pi(\boldsymbol{\alpha}|\sigma_\alpha^2)\, d\boldsymbol{\alpha}\\
&\propto(\sigma_\alpha^2)^{-K}\int|\mathbf{R}|^{-K}\mathrm{etr}\{-(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H)^H\mathbf{R}^{-1}(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H)\}\\
&\quad\exp\{-\sigma_\alpha^{-2}\boldsymbol{\alpha}^H\boldsymbol{\alpha}\}\, d\boldsymbol{\alpha}\\
&\propto(\sigma_\alpha^2)^{-K}|\mathbf{R}|^{-K}\mathrm{etr}\left\{-\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{Z}+\frac{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}\mathbf{v}^H\mathbf{R}^{-1}\mathbf{Z}}{\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}\right\}\\
&\quad\times\int\exp\left\{(\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})\left\|\boldsymbol{\alpha}-\frac{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}}{\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}\right\|^2\right\}d\boldsymbol{\alpha}\\
&\propto(\sigma_\alpha^2)^{-K}|\mathbf{R}|^{-K}(\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})^{-K}\\
&\quad\mathrm{etr}\left\{-\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{Z}+\frac{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}\mathbf{v}^H\mathbf{R}^{-1}\mathbf{Z}}{\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}\right\}\\
&\propto|\mathbf{R}+\sigma_\alpha^2\mathbf{v}\mathbf{v}^H|^{-K}\mathrm{etr}\{-\mathbf{Z}^H(\mathbf{R}+\sigma_\alpha^2\mathbf{v}\mathbf{v}^H)^{-1}\mathbf{Z}\}.
\end{aligned} \tag{12}$$

At this stage, it appears intractable to derive an analytic expression for $p(\mathbf{Z}|\mathbf{v})$ and $p(\mathbf{Z}|\mathbf{R})$ which are required to compute the MMSD estimator of $\mathbf{v}$ and the MMSE estimator of $\mathbf{R}$. To circumvent this problem, one may think of resorting to Markov chain Monte-Carlo (MCMC) simulation methods [37], more precisely to a Gibbs sampler that would draw samples from the posterior distributions $p(\mathbf{v}|\mathbf{R},\mathbf{Z})$ or $p(\mathbf{R}|\mathbf{v},\mathbf{Z})$ and approximate the integrals in (9)–(10) by arithmetic means. However, using (12), the conditional posterior distributions $p(\mathbf{v}|\mathbf{R},\mathbf{Z})$ and $p(\mathbf{R}|\mathbf{v},\mathbf{Z})$ are given by

$$\begin{aligned}
p(\mathbf{v}|\mathbf{R},\mathbf{Z}) &\propto p(\mathbf{Z}|\mathbf{v},\mathbf{R})\pi(\mathbf{v}|\sigma_\alpha^2)\\
&\propto(\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})^{-K}\exp\left\{\kappa|\mathbf{v}^H\overline{\mathbf{v}}|^2+\frac{\mathbf{v}^H\mathbf{R}^{-1}\mathbf{Z}\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}}{\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}\right\}
\end{aligned} \tag{13a}$$

$$\begin{aligned}
p(\mathbf{R}|\mathbf{v},\mathbf{Z}) &\propto p(\mathbf{Z}|\mathbf{v},\mathbf{R})\pi(\mathbf{R}|\nu,\mu)\\
&\propto|\mathbf{R}|^{-(\nu+N)}|\mathbf{R}+\sigma_\alpha^2\mathbf{v}\mathbf{v}^H|^{-K}\mathrm{etr}\{-(\nu-N)\mu\mathbf{R}^{-1}\}\\
&\quad\mathrm{etr}\{-\mathbf{Z}^H(\mathbf{R}+\sigma_\alpha^2\mathbf{v}\mathbf{v}^H)^{-1}\mathbf{Z}\}
\end{aligned} \tag{13b}$$

These distributions do not belong to a familiar class, rendering generation of samples according to $p(\mathbf{v}|\mathbf{R},\mathbf{Z})$ or $p(\mathbf{R}|\mathbf{v},\mathbf{Z})$ a difficult issue. To conclude, there is no way but to estimate $\boldsymbol{\alpha}$ jointly with $\mathbf{v}$ and $\mathbf{R}$. More specifically, a Gibbs sampler is now proposed which generates samples from $p(\mathbf{v}|\boldsymbol{\alpha},\mathbf{R},\mathbf{Z})$, $p(\boldsymbol{\alpha}|\mathbf{v},\mathbf{R},\mathbf{Z})$ and $p(\mathbf{R}|\boldsymbol{\alpha},\mathbf{v},\mathbf{Z})$. As illustrated below, these conditional posterior distributions are easy to simulate.

Indeed, using (5) along with (11), one can write

$$p(\boldsymbol{\alpha}|\mathbf{v},\mathbf{R},\mathbf{Z})\propto\exp\left\{-(\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})\left\|\boldsymbol{\alpha}-\frac{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}}{\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}\right\|^2\right\} \tag{14}$$

and hence $\boldsymbol{\alpha}$, conditioned on $\mathbf{v}$, $\mathbf{R}$, $\mathbf{Z}$, is Gaussian distributed, i.e.,

$$\boldsymbol{\alpha}|\mathbf{v},\mathbf{R},\mathbf{Z}\sim\mathrm{CN}\left(\frac{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}}{\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}},(\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})^{-1}\mathbf{I}_K\right) \tag{15}$$

Accordingly, we deduce from (6), (5) and (11) that

$$\begin{aligned}
p(\mathbf{v}|\boldsymbol{\alpha},\mathbf{R},\mathbf{Z})\propto\exp\{&\kappa|\mathbf{v}^H\overline{\mathbf{v}}|^2-(\boldsymbol{\alpha}^H\boldsymbol{\alpha})(\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})\\
&+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}+\boldsymbol{\alpha}^H\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}\}
\end{aligned} \tag{16}$$

which is recognized as a complex Bingham von Mises Fisher (BMF) distribution with parameters $\kappa\overline{\mathbf{v}}\mathbf{v}^H-(\boldsymbol{\alpha}^H\boldsymbol{\alpha})\mathbf{R}^{-1}$ and $\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}$, i.e.,

$$\mathbf{v}|\boldsymbol{\alpha},\mathbf{R},\mathbf{Z}\sim\mathrm{BMF}_c(\kappa\overline{\mathbf{v}}\mathbf{v}^H-(\boldsymbol{\alpha}^H\boldsymbol{\alpha})\mathbf{R}^{-1},\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}) \tag{17}$$

An efficient sampling scheme for generating samples according to a real BMF distribution was proposed by Hoff [38]. This scheme can be adapted to generate a complex BMF distributed vector, utilizing the relation between real and complex BMF distributions, see Appendix A. Finally, the conditional posterior distribution of $\mathbf{R}$ is obtained as

$$p(\mathbf{R}|\boldsymbol{\alpha},\mathbf{v},\mathbf{Z})\propto|\mathbf{R}|^{-(\nu+N+K)}\mathrm{etr}\{-\mathbf{R}^{-1}\mathbf{M}(\boldsymbol{\alpha},\mathbf{v},\mathbf{Z})\} \tag{18}$$

with

$$\mathbf{M}(\boldsymbol{\alpha},\mathbf{v},\mathbf{Z})=(\nu-N)\mu\mathbf{I}_N+(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H)(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H)^H \tag{19}$$

This conditional posterior distribution is an inverse Wishart distribution with $\nu+K$ degrees of freedom and parameter matrix $\mathbf{M}(\boldsymbol{\alpha},\mathbf{v},\mathbf{Z})$. The latter is, up to a scaling factor, the posterior mean of $\mathbf{R}|\boldsymbol{\alpha},\mathbf{v},\mathbf{Z}$. It is instructive to observe that, due to the choice of the prior of $\mathbf{R}$ in (3), the form of $\mathbf{M}(\boldsymbol{\alpha},\mathbf{v},\mathbf{Z})$ bears strong resemblance with the usual diagonal loading, as we hinted at in the introduction. Since diagonal loading is known to be efficient to mitigate steering vector errors, the introduction of the non-informative prior $\pi(\mathbf{R})$ in (3) can be viewed as a means to improve robustness. It is also worth noticing that $(\nu-N)\mu/K$ corresponds to the loading level, which provides a way to fix ν and μ.

Before proceeding, we note that the distributions $p(\boldsymbol{\alpha}|\mathbf{v},\mathbf{R},\mathbf{Z})$ and $p(\mathbf{v}|\boldsymbol{\alpha},\mathbf{R},\mathbf{Z})$ depend on $\mathbf{R}$ through its inverse $\mathbf{R}^{-1}$. Moreover, our final objective is to derive a beamformer whose weight vector depends directly on $\mathbf{R}^{-1}$. Therefore, since we look for an estimate of $\mathbf{R}^{-1}$ rather than an estimate of $\mathbf{R}$, the Gibbs sampler will generate directly the inverse of $\mathbf{R}$ from a Wishart distribution

$$\mathbf{R}^{-1}|\boldsymbol{\alpha},\mathbf{v},\mathbf{Z}\sim\mathrm{CW}(\nu+K,[\mathbf{M}(\boldsymbol{\alpha},\mathbf{v},\mathbf{Z})]^{-1}) \tag{20}$$

The Gibbs sampler will thus successively draw samples from (15), (17) and (20), as described in Algorithm 1.

Algorithm 1. Gibbs sampler for estimation of $\mathbf{v}$ and $\mathbf{R}^{-1}$.

Require: initial values $\mathbf{R}^{-1}(0)$, $\mathbf{v}(0)$
for $n=1,\ldots,N_{\text{bi}}+N_r$ **do**
  sample $\boldsymbol{\alpha}(n)$ from $p(\boldsymbol{\alpha}|\mathbf{v}(n-1),\mathbf{R}(n-1),\mathbf{Z})$ in (15).
  sample $\mathbf{v}(n)$ from $p(\mathbf{v}|\boldsymbol{\alpha}(n),\mathbf{R}(n-1),\mathbf{Z})$ in (17).
  sample $\mathbf{R}^{-1}(n)$ from $p(\mathbf{R}^{-1}|\boldsymbol{\alpha}(n),\mathbf{v}(n),\mathbf{Z})$ in (20).
end for
Ensure: sequence of random variables $\boldsymbol{\alpha}(n)$, $\mathbf{v}(n)$, $\mathbf{R}^{-1}(n)$.

Once these samples are available, the MMSD estimator of $\mathbf{v}$ and the MMSE estimator of $\mathbf{R}^{-1}$ can be approximated by

$$\hat{\mathbf{v}}_{\text{mmsd}}=\mathcal{P}\left\{\frac{1}{N_r}\sum_{n=N_{\text{bi}}+1}^{N_{\text{bi}}+N_r}\mathbf{v}(n)\mathbf{v}^H(n)\right\} \tag{21a}$$

$$\hat{\mathbf{R}}^{-1}_{\text{mmse}}=\frac{1}{N_r}\sum_{n=N_{\text{bi}}+1}^{N_{\text{bi}}+N_r}\mathbf{R}^{-1}(n) \tag{21b}$$

where N_{bi} stands for the number of burn-in iterations and N_r is the effective number of iterations. Finally, with the above estimates available, a beamformer can be designed whose weight vector is given by

$$\mathbf{w}\propto\hat{\mathbf{R}}^{-1}_{\text{mmse}}\hat{\mathbf{v}}_{\text{mmsd}} \tag{22}$$

4. A model based on the angle between v and $\overline{\mathbf{v}}$

In this section, we consider a slightly different model for $\mathbf{v}$ which stems from the following observation. One drawback of the Bingham distribution is that it is not easy to set a value for κ, even if a rough knowledge of the average value of $|\mathbf{v}^H\overline{\mathbf{v}}|^2$ along with (8) can serve as a guide to select a value for κ. Additionally, the choice of a Bingham distribution for $\mathbf{v}$ results in a given distribution for θ and hence the user cannot choose the latter. In contrast, we would be interested in a model that directly depends on θ and where the prior distribution of θ could be set by the user. This is the approach we take in this section. With no loss of generality we assume now that $\overline{\mathbf{v}}=[1\ 0\ \cdots\ 0]^T$: if this is not the case, the measurements $\mathbf{Z}$ can be pre-multiplied by the unitary matrix $\mathbf{Q}$ – without it modifying the distribution of $\mathbf{Z}$ in (5) – such that $\mathbf{Q}^H\overline{\mathbf{v}}=[1\ 0\ \cdots\ 0]^T$. Then we use a model for $\mathbf{v}$ that directly involves the angle θ between $\mathbf{v}$ and $\overline{\mathbf{v}}$, namely

$$\mathbf{v}=\begin{bmatrix}\cos\theta\\ \mathbf{v}_2\sin\theta\end{bmatrix}e^{i\phi} \tag{23}$$

In (23), $\mathbf{v}_2$ is an arbitrary vector in $\mathcal{S}_{N-1}$ and we assume that it is uniformly distributed on the sphere. As for ϕ, we note that since $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}e^{i\phi}$ have the same distribution, $e^{i\phi}$ can be absorbed in $\boldsymbol{\alpha}$ which amounts to set $\phi=0$ in (23). Regarding θ we assume that it is uniformly distributed on $[0,\theta_{\max}]$, i.e., $\theta\sim\mathcal{U}([0,\theta_{\max}])$. $\theta_{\max}$ sets the maximum angle between $\mathbf{v}$ and $\overline{\mathbf{v}}$ and thus indicates how confident we are in $\overline{\mathbf{v}}$. This model is more intuitive than the Bingham model as $\theta_{\max}$ is an intelligible parameter which is easier to set than κ. Moreover, the pdf of the angle between $\mathbf{v}$ and $\overline{\mathbf{v}}$ is uniform and thus different from that in a Bingham distribution, see (7). Finally, the average distance between $\mathbf{v}$ and $\overline{\mathbf{v}}$ as well as the average value of $\cos^2\theta$ can be evaluated in a straightforward manner as

$$\mathcal{E}\{|\mathbf{v}^H\overline{\mathbf{v}}|^2\}=\frac{1}{2}+\frac{1}{4}\frac{\sin 2\theta_{\max}}{\theta_{\max}} \tag{24a}$$

$$\mathcal{E}\{d(\mathbf{v},\overline{\mathbf{v}})\}=\mathcal{E}\{\theta\}=\frac{\theta_{\max}}{2} \tag{24b}$$

Note that (23) is the vector version of the CS (cosine–sine) decomposition for unitary matrices [39], and a similar model for subspaces of rank greater than one has been used in [40].

We use the same approach as before, except that now we need to derive the conditional posterior distributions of $\mathbf{v}_2$ and θ: indeed, the conditional posterior distributions of $\boldsymbol{\alpha}$ and $\mathbf{R}$ remain the same as in (15) and (20). Let us begin by observing that, for an arbitrary Hermitian matrix $\mathbf{A}$ and an arbitrary vector $\mathbf{c}$

$$\begin{aligned}\mathbf{v}^H\mathbf{A}\mathbf{v}&=[\cos\theta\ \mathbf{v}_2^H\sin\theta]\begin{bmatrix}A_{11}&\mathbf{A}_{12}\\ \mathbf{A}_{21}&\mathbf{A}_{22}\end{bmatrix}\begin{bmatrix}\cos\theta\\ \mathbf{v}_2\sin\theta\end{bmatrix}\\&=A_{11}\cos^2\theta+(\mathbf{v}_2^H\mathbf{A}_{22}\mathbf{v}_2)\sin^2\theta+(\mathbf{v}_2^H\mathbf{A}_{21}+\mathbf{A}_{21}^H\mathbf{v}_2)\cos\theta\sin\theta\end{aligned} \tag{25}$$

$$\begin{aligned}\mathbf{v}^H\mathbf{c}&=[\cos\theta\ \mathbf{v}_2^H\sin\theta]\begin{bmatrix}c_1\\ \mathbf{c}_2\end{bmatrix}\\&=c_1\cos\theta+(\mathbf{v}_2^H\mathbf{c}_2)\sin\theta\end{aligned} \tag{26}$$

Therefore, from (5), we have

$$\begin{aligned}p(\theta,\mathbf{v}_2|\boldsymbol{\alpha},\mathbf{R},\mathbf{Z})&\propto\exp\{-(\boldsymbol{\alpha}^H\boldsymbol{\alpha})(\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v})+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}+\boldsymbol{\alpha}^H\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}\}\\&\propto\exp\{-(\boldsymbol{\alpha}^H\boldsymbol{\alpha})([\mathbf{R}^{-1}]_{11}\cos^2\theta+(\mathbf{v}_2^H[\mathbf{R}^{-1}]_{22}\mathbf{v}_2)\sin^2\theta]\}\\&\times\exp\{-(\boldsymbol{\alpha}^H\boldsymbol{\alpha})[\mathbf{v}_2^H[\mathbf{R}^{-1}]_{21}+[\mathbf{R}^{-1}]_{21}^H\mathbf{v}_2]\cos\theta\sin\theta\}\\&\times\exp\{[\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}]_1\cos\theta+(\mathbf{v}_2^H[\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}]_2)\sin\theta\}\\&\times\exp\{[\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}]_1^*\cos\theta+([\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}]_2^H\mathbf{v}_2)\sin\theta\}\end{aligned} \tag{27}$$

The conditional posterior distribution of θ only is thus given by

$$\begin{aligned}p(\theta|\mathbf{v}_2,\boldsymbol{\alpha},\mathbf{R},\mathbf{Z})&\propto\exp\{-(\boldsymbol{\alpha}^H\boldsymbol{\alpha})[[\mathbf{R}^{-1}]_{11}-\mathbf{v}_2^H[\mathbf{R}^{-1}]_{22}\mathbf{v}_2]\cos^2\theta\}\\&\times\exp\{-(\boldsymbol{\alpha}^H\boldsymbol{\alpha})\mathrm{Re}(\mathbf{v}_2^H[\mathbf{R}^{-1}]_{21})\sin 2\theta\}\\&\times\exp\{2\mathrm{Re}([\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}]_1)\cos\theta+2\mathrm{Re}(\mathbf{v}_2^H[\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}]_2)\sin\theta\}\end{aligned} \tag{28}$$

This distribution does not belong to a known family. Since $\theta\in[0,\pi/2]$, a simple way to draw samples from (28) consists in using an inverse cumulative density function approach. The conditional posterior distribution of $\mathbf{v}_2$ can be written as

$$p(\mathbf{v}_2|\theta,\boldsymbol{\alpha},\mathbf{R},\mathbf{Z})\propto\exp\{\mathbf{v}_2^H\mathbf{b}+\mathbf{b}^H\mathbf{v}_2+\mathbf{v}_2^H\mathbf{B}\mathbf{v}_2\} \tag{29}$$

with

$$\mathbf{b}=-(\boldsymbol{\alpha}^H\boldsymbol{\alpha})[\mathbf{R}^{-1}]_{21}\cos\theta\sin\theta+[\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}]_2\sin\theta \tag{30}$$

$$\mathbf{B}=-(\boldsymbol{\alpha}^H\boldsymbol{\alpha})[\mathbf{R}^{-1}]_{22}\sin^2\theta \tag{31}$$

It ensues that

$$\mathbf{v}_2|\theta,\boldsymbol{\alpha},\mathbf{R},\mathbf{Z}\sim\mathrm{BMF}_c(\mathbf{B},\mathbf{b}) \tag{32}$$

The Gibbs sampler corresponding to the CS model will thus work similarly to that of Algorithm 1 except that, on

line 3, we need to draw θ and $\mathbf{v}_2$ according to their conditional posterior distributions in (28) and (29), and then reconstruct $\mathbf{v}$ using (23). The MMSD estimate of $\mathbf{v}$ and MMSE estimate of $\mathbf{R}^{-1}$ will still be obtained as in (21a)–(21b), and the final beamformer constructed as in (22).

5. Numerical simulations

In this section, we assess the performance achieved with the beamformer $\mathbf{w} \propto \hat{\mathbf{R}}_{\text{mmse}}^{-1}\hat{\mathbf{v}}_{\text{mmsd}}$. We consider a uniform linear array of $N=16$ elements spaced a half-wavelength apart. The signal received on the array is the superposition of the signal of interest, the interferences and the receiver noise, which is assumed to be temporally and spatially white with power σ_n^2. As for the interferences, we assume that there are two of them, impinging from directions -15° and 20°, with respective interference to noise ratio (INR) equal to 30 dB and 20 dB. The signal of interest (SOI) is assumed to propagate from the broadside of the array so that $\bar{\mathbf{v}} = \mathbf{a}(0^\circ)$ where $\mathbf{a}(\varphi) = [1 \; e^{i\pi \sin\varphi} \; \ldots \; e^{i\pi(N-1)\sin\varphi}]^T/\sqrt{N}$ is the (normalized) steering vector of the array. We consider pointing errors so that the SOI actually impinges from the direction of arrival (DOA) φ_{true}. The latter is expressed in fraction of the half power beam width (HPBW) [6] as $\varphi_{\text{true}} = \delta \times HPBW$. We stress the fact that the *true steering vector is not generated according to the prior distribution* assumed by each Bayesian beamformer. The signal to noise ratio (SNR) defined as

$$SNR = 10\log_{10}\frac{\sigma_\alpha^2 \mathbf{v}^H\mathbf{v}}{N\sigma_n^2}$$

is set to $SNR=0$ dB. In order to set the values for ν and μ, we use the expression of $\mathbf{M}(\boldsymbol{\alpha},\mathbf{v},\mathbf{Z})$ in (19), which corresponds, up to a scaling factor, to the posterior mean of $\mathbf{R}$ conditioned on $\boldsymbol{\alpha}$, $\mathbf{v}$ and $\mathbf{Z}$. In order for the term due to the data and the term corresponding to diagonal loading to have approximately the same weight, we set $\nu = K+N$ so that μ is tantamount to a diagonal loading level. We fix it to 5 dB above the white noise level, a good rule of thumb in practice [6].

The performance metric of the adaptive beamformer will be the SINR loss with respect to the noise-only-environment defined by [41]

$$SINR_{\text{loss}} = \frac{|\mathbf{w}^H\mathbf{v}|^2}{\mathbf{w}^H\mathbf{R}\mathbf{w}}\frac{1}{\sigma_n^{-2}\mathbf{v}^H\mathbf{v}} \tag{33}$$

First, we investigate the sensitivity of the Bayesian beamformers towards the parameters one has to set, namely κ for the Bingham-based model and $\theta_{\max}$ for the CS-based model. The results are shown in Figs. 2 and 3. For $\delta = 0.2$ [respectively $\delta = 0.4$] the angle between $\bar{\mathbf{v}}$ and $\mathbf{v}$, i.e., $\arccos|\bar{\mathbf{v}}^H\mathbf{v}|$, is equal to 18.43° [respectively 36.45°]. The following comments about these figures can be made. For moderate steering vector error ($\delta = 0.2$), the Bayesian beamformers are seen to achieve a quasi-constant SINR loss over a large range of values for κ and $\theta_{\max}$. This is very important from a practical point of view as it means that the user does not have to tune these

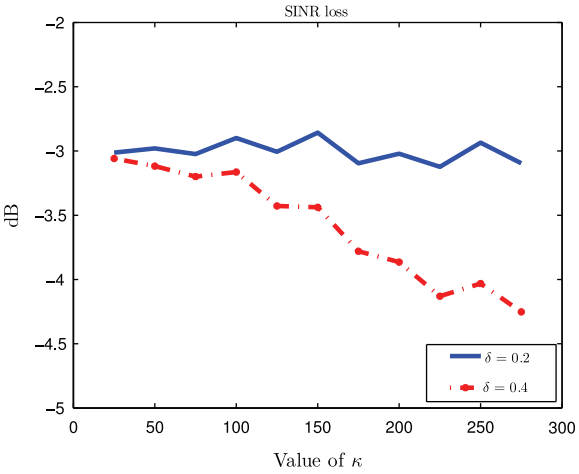

Fig. 2. SINR loss of the adaptive beamformer in the Bingham model versus κ. $K=32$.

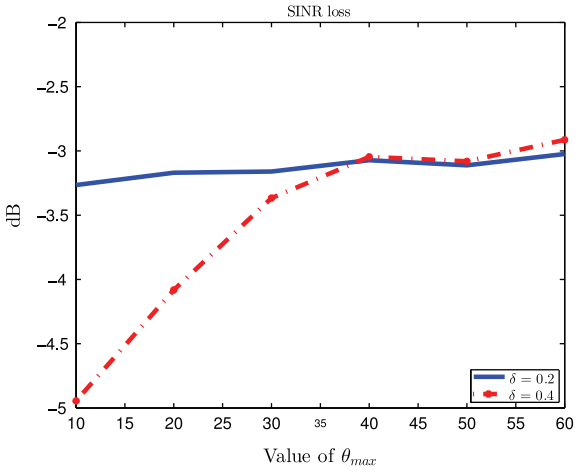

Fig. 3. SINR loss of the adaptive beamformer in the CS model versus $\theta_{\max}$. $K=32$.

parameters very accurately: the performance is guaranteed to be almost the same over a rather large interval. On the other hand, for large steering vector errors ($\delta = 0.4$) it is mandatory to adapt the values of κ and $\theta_{\max}$. The former should not be taken too large while $\theta_{\max}$ should not be chosen too small in order to accommodate the possibly large difference between $\mathbf{v}$ and $\bar{\mathbf{v}}$: recall that a large κ or a small $\theta_{\max}$ implies that $\mathbf{v}$ should be close to $\bar{\mathbf{v}}$. Note however that the case $\delta = 0.4$ corresponds to a rather large error, the case $\delta = 0.2$ may be more representative. In the latter situation, hopefully there is no need to select very accurately the values of κ and $\theta_{\max}$. To summarize this sensitivity analysis, we can conclude that the Bayesian beamformers are rather robust to the choice of the parameters κ and $\theta_{\max}$: the latter need not be set very accurately, a very interesting property in practice.

We now study the influence of the number of snapshots K and the pointing error δ in Figs. 4–7. The beamformers are compared with conventional diagonal

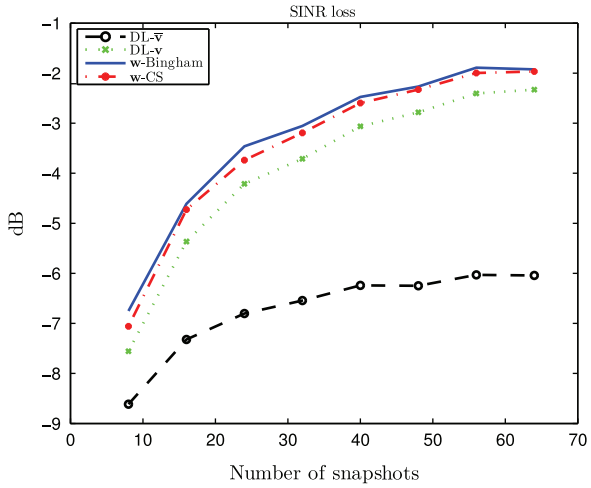

Fig. 4. SINR loss of the adaptive beamformers versus number of snapshots K. $\kappa = 50$, $\theta_{\max} = 45^\circ$ and $\delta = 0.2$.

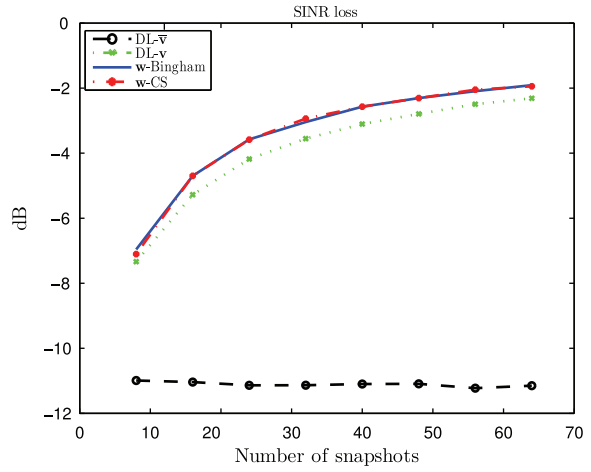

Fig. 5. SINR loss of the adaptive beamformers versus number of snapshots K. $\kappa = 50$, $\theta_{\max} = 45^\circ$ and $\delta = 0.4$.

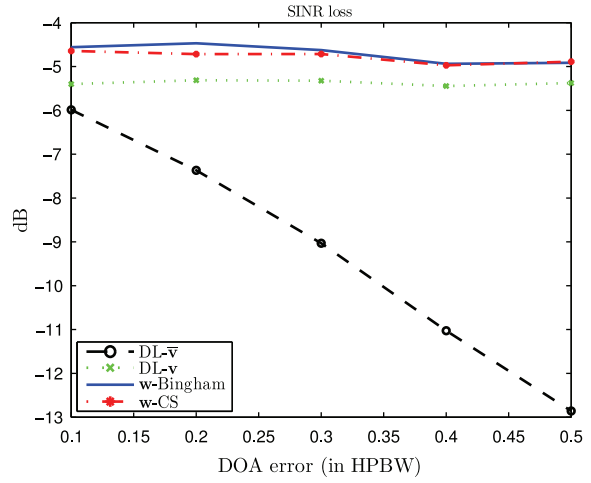

Fig. 6. SINR loss of the adaptive beamformers versus pointing error. $K = 16$, $\kappa = 50$ and $\theta_{\max} = 45^\circ$.

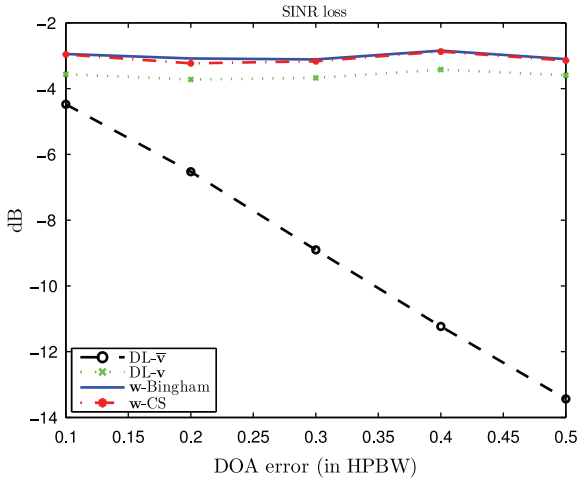

Fig. 7. SINR loss of the adaptive beamformers versus pointing error. $K = 32$, $\kappa = 50$ and $\theta_{\max} = 45^\circ$.

loading using the presumed steering vector, i.e.,

$$\mathbf{w}_{\mathrm{DL}-\bar{\mathbf{v}}} \propto (K^{-1}\mathbf{Z}\mathbf{Z}^H + \mu\mathbf{I}_N)^{-1}\bar{\mathbf{v}} \tag{34}$$

with a loading level 5 dB above the white noise level. We also consider a "clairvoyant " diagonally loaded beamformer which would have knowledge of $\mathbf{v}$, i.e.,

$$\mathbf{w}_{\mathrm{DL}-\mathbf{v}} \propto (K^{-1}\mathbf{Z}\mathbf{Z}^H + \mu\mathbf{I}_N)^{-1}\mathbf{v} \tag{35}$$

Of course, the latter is hypothetical but it enables to consider only finite-sample effects without any steering vector error.

These figures call for the following observations:

- The Bayesian beamformers significantly improve upon conventional diagonal loading using the presumed steering vector but they also outperform the diagonally loaded beamformer constructed with the true steering vector. This is a rather remarkable result, especially with large steering vector errors.
- The CS-based beamformer and its Bingham counterpart result in approximately the same output SINR.
- For large steering vector errors, diagonal loading with $\bar{\mathbf{v}}$ performs very poorly and the performance is mainly dominated by the steering vector error: for instance the output SINR does not improve when K increases, see Fig. 5. On the contrary for moderate steering vector errors ($\delta = 0.2$) the output SINR increases when K is increased. The diagonally loaded beamformer which knows $\mathbf{v}$ does not suffer from this phenomenon of course: its output SINR is independent of δ, see Figs. 6 and 7, and increases when K increases.
- Remarkably enough, the Bayesian beamformers are seen to be quite insensitive to the magnitude of the steering vector errors, see Figs. 6 and 7. In contrast, their performance depends on the number of snapshots, cf. Figs. 4 and 5.
- The Bayesian beamformers have a very high rate of convergence. Indeed, the SINR loss is inferior to 3 dB at about $K = 2N$, a rate of convergence commensurate with that of an MVDR beamformer (where the signal of

interest is not contained in the data) and much better than that of a conventional MPDR beamformer. This high rate of convergence is achieved despite steering vector errors being present.

To summarize, the new Bayesian beamformers presented herein enable one to achieve a close to optimal performance very rapidly, despite the presence of steering vector errors which they are not very sensitive to.

As a final simulation, we investigate the robustness of the beamformers to an imprecise knowledge of σ_α^2, which was assumed to be known in the previous simulations. First, we study the SINR loss obtained when the assumed value of σ_α^2 – or equivalently the assumed value of the SNR- differs from its true value. More specifically, we fix the value of the SNR to 0 dB – which sets σ_α^2 – and we run the beamformers with an assumed value of the SNR that might differ from the true one. The SINR loss, relative to the SINR loss obtained when the SNR is known, is plotted in Fig. 8. As can be observed, the degradation is not very important (about -1 dB) for differences between true and assumed SNR up to ± 6 dB. The beamformer based on the Bingham prior distribution appears to be slightly more robust than the CS-based beamformer. Therefore, a wrong guess of the SNR does not affect too much the beamformers provided that the assumed SNR stays within a few decibels from the actual SNR.

If our knowledge about the SOI power is not that accurate, then the extended Gibbs sampler presented in Appendix B can be used. The latter assumes an inverse Gamma prior distribution for σ_α^2 and the Gibbs sampler is extended to draw samples from the posterior distribution of σ_α^2, see the Appendix for more details. In Fig. 9 we compare the SINR loss achieved with these extended beamformers to the SINR loss obtained when σ_α^2 is known. As can be observed (similar results were obtained in other simulations not reported here), the difference is marginal, at most 0.2 dB. Therefore, in the case where σ_α^2 is not known precisely, the extended beamformers can be used without any performance degradation.

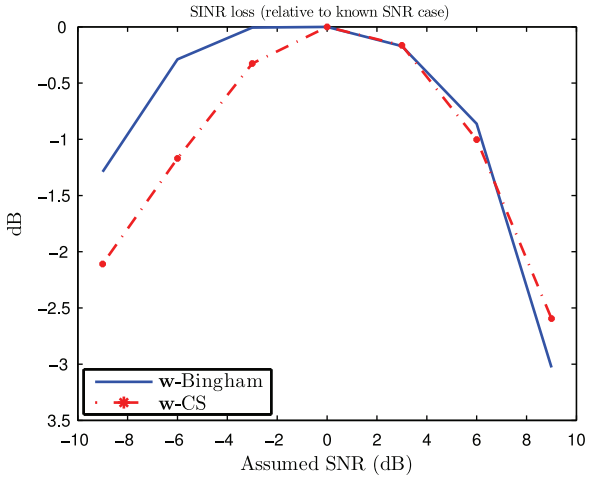

Fig. 8. SINR loss of the Bayesian beamformers versus assumed SNR (true SNR is 0 dB). $K=32$, $\kappa=50$, $\theta_{\max}=45^\circ$ and $\delta=0.2$.

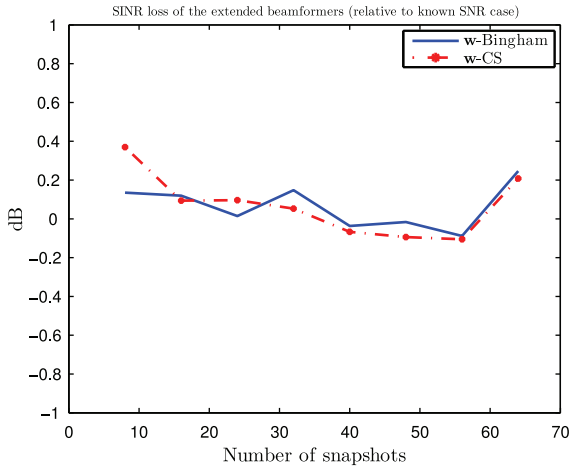

Fig. 9. SINR loss of the extended Bayesian beamformers versus number of snapshots K. $\kappa=50$, $\theta_{\max}=45^\circ$ and $\delta=0.2$.

6. Conclusions

We presented a new Bayesian approach of robust adaptive beamforming based on two novel prior distributions for the steering vector of interest, namely a Bingham (or Watson) distribution or a CS-based distribution which depends only on the angle between the actual steering vector and its presumed value. Additionally, an inverse Wishart distribution with a parameter matrix proportional to $\mathbf{I}$ was assumed for the interference covariance matrix, which amounts to introducing diagonal loading and hence improving robustness. For both models, the MMSE estimator of the interference covariance matrix as well as the MMSD estimator of the steering vector were derived and implemented through a Gibbs sampling procedure. The new algorithms were shown to significantly improve over conventional diagonal loading especially in low sample support and under steering vector errors.

Acknowledgments

This work was partially supported by DGA/MRIS under Grant no. 2009.60.033.00.470.75.01.

Appendix A. Complex Bingham and Bingham von Mises Fisher distributions

In this appendix we briefly review the complex Bingham distribution [32,33] and introduce the complex Bingham von Mises Fisher distribution from its real-valued counterpart. For any vector $\mathbf{v}$, we let $\tilde{\mathbf{v}}=[\mathbf{v}_R^T\ \mathbf{v}_I^T]^T$ denote the $2N$-length real-valued vector obtained by concatenating the real ($\mathbf{v}_R$) and imaginary ($\mathbf{v}_I$) parts of $\mathbf{v}$. Accordingly, for any Hermitian matrix $\mathbf{A}$ we let $\tilde{\mathbf{A}}=\left[\begin{smallmatrix}\mathbf{A}_R & -\mathbf{A}_I\\ \mathbf{A}_I & \mathbf{A}_R\end{smallmatrix}\right]$ which is a $2N\times 2N$ symmetric matrix.

A complex-valued vector $\mathbf{v}\in\mathcal{S}_N$ is said to follow a (complex) Bingham distribution with parameter matrix $\mathbf{A}$ if its probability density function can be written as $p(\mathbf{v})\propto\exp\{\mathbf{v}^H\mathbf{A}\mathbf{v}\}$. We denote this distribution as

$\mathbf{v} \sim \mathrm{B}_c(\mathbf{A})$. Since

$$\mathbf{v}^H\mathbf{A}\mathbf{v} = [\mathbf{v}_R^T\ \mathbf{v}_I^T]\begin{bmatrix}\mathbf{A}_R & -\mathbf{A}_I\\ \mathbf{A}_I & \mathbf{A}_R\end{bmatrix}\begin{bmatrix}\mathbf{v}_R\\ \mathbf{v}_I\end{bmatrix} = \tilde{\mathbf{v}}^T\tilde{\mathbf{A}}\tilde{\mathbf{v}}$$

we have the equivalence $\mathbf{v} \sim \mathrm{B}_c(\mathbf{A}) \equiv \tilde{\mathbf{v}} \sim \mathrm{B}(\tilde{\mathbf{A}})$ where $\mathrm{B}(-)$ denotes the real-valued Bingham distribution. A special case of interest is when $\mathbf{A} = \kappa\overline{\mathbf{v}}\overline{\mathbf{v}}^H$ where $\overline{\mathbf{v}} \in \mathcal{S}_N$: the distribution is then referred to as the complex Watson distribution [33]. In this case, one has $p(\mathbf{v}) \propto \exp\{\kappa|\mathbf{v}^H\overline{\mathbf{v}}|^2\}$ and the term $|\mathbf{v}^H\overline{\mathbf{v}}|^2$ within the exponential corresponds to the square cosine angle between $\mathbf{v}$ and $\overline{\mathbf{v}}$. We emphasize that

$$\begin{aligned}|\mathbf{v}^H\overline{\mathbf{v}}|^2 &= (\mathbf{v}_R^T\overline{\mathbf{v}}_R + \mathbf{v}_I^T\overline{\mathbf{v}}_I)^2 + (\mathbf{v}_R^T\overline{\mathbf{v}}_I - \mathbf{v}_I^T\overline{\mathbf{v}}_R)^2\\ &= \left([\mathbf{v}_R^T\ \mathbf{v}_I^T]\begin{bmatrix}\overline{\mathbf{v}}_R\\ \overline{\mathbf{v}}_I\end{bmatrix}\right)^2 + \left([\mathbf{v}_R^T\ \mathbf{v}_I^T]\begin{bmatrix}\overline{\mathbf{v}}_I\\ -\overline{\mathbf{v}}_R\end{bmatrix}\right)^2\end{aligned}$$

Therefore, $\mathbf{v} \sim \mathrm{B}_c(\kappa\overline{\mathbf{v}}\overline{\mathbf{v}}^H)$ is *not* equivalent to $\tilde{\mathbf{v}} \sim \mathrm{B}(\kappa[\overline{\mathbf{v}}_R^T\ \overline{\mathbf{v}}_I^T]^T[\overline{\mathbf{v}}_R^T\ \overline{\mathbf{v}}_I^T])$, where the last distribution depends on the square cosine angle between $\tilde{\mathbf{v}}$ and $[\overline{\mathbf{v}}_R^T\ \overline{\mathbf{v}}_I^T]^T$.

A complex-valued vector $\mathbf{v} \in \mathcal{S}_N$ is said to follow a (complex) Bingham von Mises Fisher distribution with parameter matrix $\mathbf{A}$ and parameter vector $\mathbf{c}$ if its probability density function can be written as $p(\mathbf{v}) \propto \exp\{\mathbf{v}^H\mathbf{c} + \mathbf{c}^H\mathbf{v} + \mathbf{v}^H\mathbf{A}\mathbf{v}\}$. We denote this distribution as $\mathbf{v} \sim \mathrm{BMF}_c(\mathbf{A},\mathbf{c})$. Since

$$\mathbf{v}^H\mathbf{c} + \mathbf{c}^H\mathbf{v} = 2(\mathbf{v}_R^T\mathbf{c}_R + \mathbf{v}_I^T\mathbf{c}_I) = 2\tilde{\mathbf{v}}^T\tilde{\mathbf{c}}$$

we have the equivalence $\mathbf{v} \sim \mathrm{BMF}_c(\mathbf{A},\mathbf{c}) \equiv \tilde{\mathbf{v}} \sim \mathrm{BMF}(\tilde{\mathbf{A}},2\tilde{\mathbf{c}})$ where $\mathrm{BMF}(-,-)$ denotes the real-valued BMF distribution.

Appendix B. Extension to random σ_α^2

In this appendix, we relax the assumption that σ_α^2 is known and consider it as a random variable with a possibly non-informative prior. More precisely, we assume that σ_α^2 follows an inverse-Gamma distribution, denoted as $\sigma_\alpha^2 \sim \mathrm{IG}(a,b)$, whose expression is

$$\pi(\sigma_\alpha^2) \propto (\sigma_\alpha^2)^{-(a+1)}\exp\{-b\sigma_\alpha^{-2}\} \tag{36}$$

The above distribution is mainly chosen for mathematical tractability since it is a conjugate prior with respect to (4). Note however that, depending on the choice of a and b, this prior can be made rather non-informative. We now proceed to the derivation of the new conditional posterior distributions for this case: the latter should be used accordingly in a modification of the Gibbs sampler. The joint posterior distribution of all variables becomes

$$\begin{aligned}p(\boldsymbol{\alpha},\mathbf{v},\mathbf{R},\sigma_\alpha^2|\mathbf{Z}) &\propto p(\mathbf{Z}|\boldsymbol{\alpha},\mathbf{v},\mathbf{R},\sigma_\alpha^2)\pi(\boldsymbol{\alpha}|\sigma_\alpha^2)\pi(\mathbf{v})\pi(\mathbf{R})\pi(\sigma_\alpha^2)\\ &\propto |\mathbf{R}|^{-K}\mathrm{etr}\{-(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H)^H\mathbf{R}^{-1}(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H)\}\\ &\times(\sigma_\alpha^2)^{-(a+K+1)}\exp\{-\sigma_\alpha^{-2}\boldsymbol{\alpha}^H\boldsymbol{\alpha}\}\exp\{-b\sigma_\alpha^{-2}\}\\ &\times|\mathbf{R}|^{-(v+N)}\mathrm{etr}\{-(v-N)\mathbf{R}^{-1}\}\pi(\mathbf{v})\end{aligned} \tag{37}$$

It is straightforward to show that $p(\boldsymbol{\alpha}|\mathbf{v},\mathbf{R},\sigma_\alpha^2,\mathbf{Z})$, $p(\mathbf{v}|\boldsymbol{\alpha},\mathbf{R},\sigma_\alpha^2,\mathbf{Z})$ and $p(\mathbf{R}|\boldsymbol{\alpha},\mathbf{v},\sigma_\alpha^2,\mathbf{Z})$ are the same as in (14), (16) and (18). The conditional posterior of σ_α^2 is given by

$$p(\sigma_\alpha^2|\boldsymbol{\alpha},\mathbf{v},\mathbf{R},\mathbf{Z}) \propto (\sigma_\alpha^2)^{-(a+K+1)}\exp\{-\sigma_\alpha^{-2}[b+\boldsymbol{\alpha}^H\boldsymbol{\alpha}]\} \tag{38}$$

and hence $\sigma_\alpha^2|\boldsymbol{\alpha},\mathbf{v},\mathbf{R},\mathbf{Z} \sim \mathrm{IG}(a+K,b+\boldsymbol{\alpha}^H\boldsymbol{\alpha})$. The extended Gibbs sampler will work similarly to that of Algorithm 1, except that σ_α^2 needs now to be generated according to $p(\sigma_\alpha^2|\boldsymbol{\alpha},\mathbf{v},\mathbf{R},\mathbf{Z})$ in (38). Observe that the Gibbs sampler of Algorithm 1 requires σ_α^{-2} and hence the latter can be generated according to a Gamma distribution with parameters $a+K$ and $b+\boldsymbol{\alpha}^H\boldsymbol{\alpha}$.

References

[1] A.B. Gershman, Robust adaptive beamforming in sensor arrays, International Journal of Electronics and Communications 53 (December (6)) (1999) 305–314.

[2] A.B. Gershman, Robustness issues in adaptive beamforming and high-resolution direction finding, in: Y. Hua, A. Gershman, Q. Chen (Eds.), High Resolution and Robust Signal Processing, Marcel Dekker, 2003, pp. 63–110. (Chapter 2).

[3] D.M. Boroson, Sample size considerations for adaptive arrays, IEEE Transactions on Aerospace and Electronic Systems 16 (July (4)) (1980) 446–451.

[4] R.T. Compton, The effect of random steering vectors in the Applebaum adaptive array, IEEE Transactions on Aerospace and Electronic Systems 18 (September (5)) (1982) 392–400.

[5] M. Wax, Y. Anu, Performance analysis of the minimum variance beamformer in the presence of steering vector errors, IEEE Transactions on Signal Processing 44 (April (4)) (1996) 938–947.

[6] H.L. Van Trees, Optimum Array Processing, John Wiley, New York, 2002.

[7] Y.I. Abramovich, Controlled method for adaptive optimization of filters using the criterion of maximum SNR, Radio Engineering and Electronic Physics 26 (March) (1981) 87–95.

[8] Y.I. Abramovich, A.I. Nevrev, An analysis of effectiveness of adaptive maximization of the signal to noise ratio which utilizes the inversion of the estimated covariance matrix, Radio Engineering and Electronic Physics 26 (December) (1981) 67–74.

[9] O.P. Cheremisin, Efficiency of adaptive algorithms with regularised sample covariance matrix, Radio Engineering and Electronic Physics 27 (10) (1982) 69–77.

[10] H. Cox, R.M. Zeskind, M.M. Owen, Robust adaptive beamforming, IEEE Transactions on Acoustics Speech and Signal Processing 35 (October (10)) (1987) 1365–1376.

[11] J. Li, P. Stoica, Z. Wang, On robust Capon beamforming and diagonal loading, IEEE Transactions on Signal Processing 51 (July (7)) (2003) 1702–1715.

[12] S.A. Vorobyov, A.B. Gershman, Z. Luo, Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem, IEEE Transactions on Signal Processing 51 (February (2)) (2003) 313–324.

[13] R. Lorenz, S.P. Boyd, Robust minimum variance beamforming, IEEE Transactions on Signal Processing 53 (May (5)) (2005) 1684–1696.

[14] J. Li, P. Stoica, Z. Wang, Doubly constrained robust Capon beamformer, IEEE Transactions on Signal Processing 52 (September (9)) (2004) 2407–2423.

[15] L. Du, T. Yardibi, J. Li, P. Stoica, Review of user parameter-free robust adaptive beamforming algorithms, Digital Signal Processing—A Review Journal 19 (July) (2009) 567–582.

[16] L. Du, J. Li, P. Stoica, Fully automatic computation of diagonal loading levels for robust adaptive beamforming, IEEE Transactions on Aerospace and Electronic Systems 46 (January (1)) (2010) 449–458.

[17] M.H. Er, A. Cantoni, An alternative formulation for an optimum beamformer with robustness capability, IEE Proceedings Radar Sonar and Navigation 132 (October (6)) (1985) 447–460.

[18] K.-C. Huarng, C.-C. Yeh, Performance analysis of derivative constraint adaptive arrays with pointing errors, IEEE Transactions on Antennas and Propagation 40 (August (8)) (1986) 975–981.

[19] Z.L. Yu, W. Ser, M.H. Er, Z. Gu, Y. Li, Robust adaptive beamformers based on worst-case optimization and constraints on magnitude response, IEEE Transactions on Signal Processing 57 (July (7)) (2009) 2615–2628.

[20] D.D. Feldman, L.J. Griffiths, A projection approach for robust adaptive beamforming, IEEE Transactions on Signal Processing 42 (April (4)) (1994) 867–876.

[21] A. Khabbazibasmenj, S.A. Vorobyov, A. Hassanien, Robust adaptive beamforming based on steering vector estimation with as little as possible prior information, IEEE Transactions on Signal Processing 60 (June (6)) (2012) 2974–2987.

[22] Y. Gu, A. Leshem, Robust adaptive beamforming based on interference covariance matrix reconstruction and steering vector estimation, IEEE Transactions on Signal Processing 60 (July (7)) (2012) 3881–3885.
[23] S. Vorobyov, H. Chen, A. Gershman, On the relationship between robust minimum variance beamformers with probabilistic and worst-case distortionless response constraints, IEEE Transactions on Signal Processing 56 (November (11)) (2008) 5719–5724.
[24] C.-C. Lee, J.-H. Lee, Robust adaptive beamforming under steering vector errors, IEEE Transactions on Antennas and Propagation 45 (January (1)) (1997) 168–175.
[25] K.L. Bell, Y. Ephraim, H.L.V. Trees, A Bayesian approach to robust adaptive beamforming, IEEE Transactions on Signal Processing 48 (February (2)) (2000) 386–398.
[26] O. Besson, A. Monakov, C. Chalus, Signal waveform estimation in the presence of uncertainties about the steering vector, IEEE Transactions on Signal Processing 52 (September (9)) (2004) 2432–2440.
[27] R.J. Muirhead, Aspects of Multivariate Statistical Theory, John Wiley, 1982.
[28] J.A. Tague, C.I. Caldwell, Expectations of useful complex Wishart forms, Multidimensional Systems and Signal Processing 5 (1994) 263–279.
[29] S. Sirianunpiboon, S.T. Howard, D. Cochran, A Bayesian derivation of generalized coherence detectors, in: Proceedings ICASSP, Kyoto, Japan, March 26–30, 2012, pp. 3253–3256.
[30] K.V. Mardia, P.E. Jupp, Directional Statistics, John Wiley & Sons, 1999.
[31] Y. Chikuse, Statistics on Special Manifolds, Springer Verlag, New York, 2003.
[32] J. Kent, The complex Bingham distribution and shape analysis, Journal Royal Statistical Society Series B 56 (2) (1994) 285–299.
[33] K.V. Mardia, I.L. Dryden, The complex Watson distribution and shape analysis, Journal of the Royal Statistical Society Series B 61 (4) (1999) 913–926.
[34] M. Abramowitz, I.A. Stegun (Eds.), Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Dover Publications, New York, 1970 (November).
[35] A. Srivastava, A Bayesian approach to geometric subspace estimation, IEEE Transactions on Signal Processing 48 (May (5)) (2000) 1390–1400.
[36] O. Besson, N. Dobigeon, J.-Y. Tourneret, Minimum mean square distance estimation of a subspace, IEEE Transactions on Signal Processing 59 (December (12)) (2011) 5709–5720.
[37] C.P. Robert, G. Casella, Monte Carlo Statistical Methods, 2nd ed. Springer Verlag, New York, 2004.
[38] P.D. Hoff, Simulation of the matrix Bingham–von Mises–Fisher distribution, with applications to multivariate and relational data, Journal of Computational and Graphical Statistics 18 (June (2)) (2009) 438–456.
[39] G. Golub, C.V. Loan, Matrix Computations, 3rd ed. John Hopkins University Press, Baltimore, 1996.
[40] O. Besson, N. Dobigeon, J.-Y. Tourneret, CS decomposition based Bayesian subspace estimation, IEEE Transactions on Signal Processing 60 (August (8)) (2012) 4210–4218.
[41] J. Ward, Space-time adaptive processing for airborne radar, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA, Technical Report 1015, December 1994.