BAYESIAN ROBUST ADAPTIVE BEAMFORMING BASED ON RANDOM STEERING VECTOR WITH BINGHAM PRIOR DISTRIBUTION

Olivier Besson and Stéphanie Bidon

University of Toulouse-ISAE
10 Avenue Edouard Belin, 31055 Toulouse, France
olivier.besson@isae.fr, stephanie.bidon@isae.fr

ABSTRACT

We consider robust adaptive beamforming in the presence of steering vector uncertainties. A Bayesian approach is presented where the steering vector of interest is treated as a random vector with a Bingham prior distribution. Moreover, in order to also improve robustness against low sample support, the interference plus noise covariance matrix $\mathbf{R}$ is assigned a non informative prior distribution which enforces shrinkage to a scaled identity matrix, similarly to diagonal loading. The minimum mean square distance estimate of the steering vector as well as the minimum mean square error estimate of $\mathbf{R}$ are derived and implemented using a Gibbs sampling strategy. The new beamformer is shown to converge within a limited number of snapshots, despite the presence of steering vector errors.

Index Terms— Robust adaptive beamforming, Bayesian estimation, Bingham distribution, Gibbs sampling.

1. PROBLEM STATEMENT AND ASSUMPTIONS

Let us consider the classical problem of designing an adaptive beamformer $\mathbf{w}$ which maximizes the signal to interference and noise ratio (SINR) for an assumed signature of interest $\bar{\mathbf{v}}$, from observation of K measurements given by

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

In (1), $\mathbf{v}$ is the actual signal of interest signature and $\mathbf{n}_k$ stands for the disturbance (interference plus noise) contribution. The most intuitive method to solve this problem is to substitute the sample covariance matrix (SCM) $\hat{\mathbf{R}} = K^{-1}\mathbf{Z}\mathbf{Z}^H$ with $\mathbf{Z} = \begin{bmatrix}\mathbf{z}_1 & \cdots & \mathbf{z}_K\end{bmatrix}$ for the true covariance matrix $\mathbf{R} = \mathcal{E}\left\{\mathbf{n}_k\mathbf{n}_k^H\right\}$ in the so-called minimum power distortionless response beamformer (MPDR) [1], viz

$$\mathbf{w}_{\text{MPDR-SMI}} \propto \hat{\mathbf{R}}^{-1}\bar{\mathbf{v}}. \tag{2}$$

Unfortunately, the beamformer in (2) suffers from poor performance in low sample support [1,2]. Additionally, it degrades significantly in the presence of steering vector errors, i.e., as soon as $\mathbf{v} \neq \bar{\mathbf{v}}$. Indeed, since $\bar{\mathbf{v}}$ is the assumed signature of interest, the contribution $\alpha_k^*\mathbf{v}$ in $\mathbf{z}_k$ is perceived as an interference, which should be eliminated. It hence results in the so-called self nulling phenomenon, which leads to a dramatic SINR loss. The most celebrated approach to mitigating steering vector errors as well as small sample size is diagonal loading (DL) [3,4]:

$$\mathbf{w}_{\text{DL}-\bar{\mathbf{v}}} \propto \left(\hat{\mathbf{R}} + \mu\mathbf{I}\right)^{-1}\bar{\mathbf{v}}. \tag{3}$$

THIS WORK WAS PARTIALLY SUPPORTED BY DGA/MRIS UNDER GRANT NO. 2012.60.0012.00.470.75.01.

DL emerges as the solution of many different problems. It is a natural way to regularize the maximum likelihood covariance matrix estimate. It also corresponds to a MPDR beamformer for which a desired value of the white noise array gain is enforced [1, 5]. Accordingly, the robust Capon beamformers of [6–8] which, originally, attempt to minimize the output power subject to $\mathbf{v}$ belonging to a sphere centered around $\bar{\mathbf{v}}$, or to maintain a minimum gain within this sphere, boil down to diagonal loading. In the same vein, the doubly constrained Capon beamformer of [9] which imposes a norm constraint of the steering vector also results in a DL-type beamformer. In the latter references, the loading factor μ depends on the uncertainty sphere radius. Other researchers have also investigated some way to automatically choose the optimal loading level, see e.g., [10, 11].

In this paper, a Bayesian approach to robust adaptive beamforming in the presence of steering vector mismatches is proposed. More precisely, we assume that $\mathbf{v}$ is random and drawn from a prior distribution which depends on $\bar{\mathbf{v}}$. Additionally, we show how DL can be embedded in a Bayesian framework, by assigning a specific prior distribution to $\mathbf{R}$.

Let us thus describe our statistical model. First, the interference plus noise vectors $\mathbf{n}_k$ are assumed to be independent and identically distributed (i.i.d.) according to a zero-mean complex-valued Gaussian distribution with covariance matrix $\mathbf{R}$, i.e.,

$$p\left(\mathbf{n}_k|\mathbf{R}\right) = \pi^{-N}|\mathbf{R}|^{-1}\exp\left\{-\mathbf{n}_k^H\mathbf{R}^{-1}\mathbf{n}_k\right\} \tag{4}$$

where $|.|$ stands for the determinant of a matrix. Additionally, we assume that $\mathbf{R}$ is drawn from an inverse Wishart distribution [12, 13] with mean $\mu\mathbf{I}_N$ and ν degrees of freedom, viz

$$\pi\left(\mathbf{R}\right) \propto |\mathbf{R}|^{-(\nu+N)}\text{etr}\left\{-(\nu-N)\mu\mathbf{R}^{-1}\right\} \tag{5}$$

where $\propto$ means proportional to and $\text{etr}\left\{.\right\}$ stands for the exponential of the trace of the matrix between braces. Note that the distribution in (5) is *non informative* as it is a maximum entropy prior distribution subject to $\mathcal{E}\left\{\text{Tr}\left\{\mathbf{R}^{-1}\right\}\right\} = c_1$ and $\mathcal{E}\left\{\log|\mathbf{R}|\right\} = c_2$ [14], and it does not depend on any prior covariance matrix. As will be evidenced later, see (19), this choice results in a posterior mean of $\mathbf{R}$ which corresponds to diagonal loading. As for the amplitudes α_k, they are assumed to be i.i.d. and follow a complex Gaussian distribution with zero mean and variance σ_α^2, i.e.,

$$\pi\left(\boldsymbol{\alpha}|\sigma_\alpha^2\right) \propto \exp\left\{-\sigma_\alpha^{-2}\boldsymbol{\alpha}^H\boldsymbol{\alpha}\right\} \tag{6}$$

where $\boldsymbol{\alpha} = \begin{bmatrix}\alpha_1 & \cdots & \alpha_K\end{bmatrix}^T$. Furthermore, we assume that σ_α^2 follows an inverse-Gamma distribution, denoted as $\sigma_\alpha^2 \sim \mathsf{IG}\left(a, b\right)$, whose expression is

$$\pi(\sigma_\alpha^2) \propto \left(\sigma_\alpha^2\right)^{-(a+1)}\exp\left\{-b\sigma_\alpha^{-2}\right\}. \tag{7}$$

The above distribution is mainly chosen for mathematical tractability since it is a conjugate prior with respect to (6). Note however that, depending on the choice of a and b, this prior can be made rather non informative. Lastly, we consider the distribution of $\mathbf{v}$. Herein, we assume that $\|\mathbf{v}\| = 1$ and that it follows a complex Bingham distribution [15, 16]:

$$\pi(\mathbf{v}) \propto \exp\left\{\kappa|\mathbf{v}^H\bar{\mathbf{v}}|^2\right\} \tag{8}$$

where κ is a positive scalar and $\bar{\mathbf{v}}$ is the unit-norm nominal steering vector. The distribution in (8) depends on $\cos^2\theta$ where θ is for the angle between $\mathbf{v}$ and $\bar{\mathbf{v}}$: thus, all vectors $\mathbf{v}$ who lie on the frontier of a cone whose axis is $\bar{\mathbf{v}}$ and whose aperture is θ are equally likely. It should be mentioned that this model differs from the classical additive steering vector error. The scalar κ serves as a concentration parameter: the larger κ the closer $\mathbf{v}$ and $\bar{\mathbf{v}}$. More precisely, it is possible to show that

$$\mathcal{E}\left\{|\mathbf{v}^H\bar{\mathbf{v}}|^2\right\} = 1 - \frac{1}{\kappa}\frac{\gamma(N,\kappa)}{\gamma(N-1,\kappa)} \tag{9}$$

where $\mathcal{E}\left\{|\mathbf{v}^H\bar{\mathbf{v}}|^2\right\}$ is approximately the average square distance between $\mathbf{v}$ and $\bar{\mathbf{v}}$, and is seen to be inversely proportional to κ. This is illustrated in Figure 1 where we plot the distribution of θ for different values of κ.

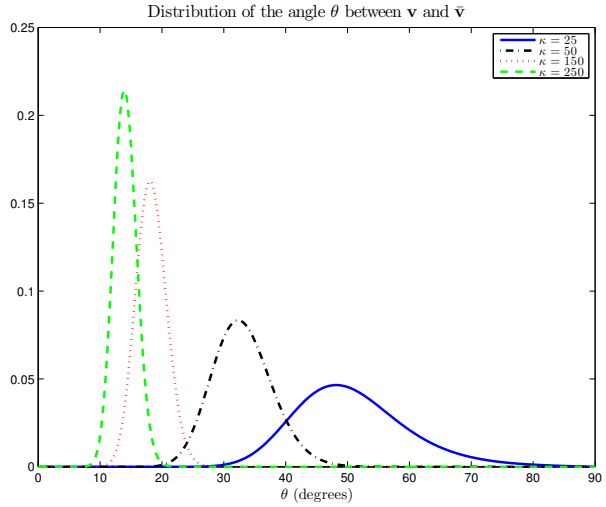

Fig. 1. Distribution of the angle between $\mathbf{v}$ and $\bar{\mathbf{v}}$, when $\mathbf{v}$ is drawn from (8), for different values of κ. $N = 16$.

Our objective is, from the statistical assumptions stated in (4)-(8), to obtain the minimum mean-square error (MMSE) estimate $\hat{\mathbf{R}}_{\text{mmse}}$ of $\mathbf{R}$

$$\hat{\mathbf{R}}_{\text{mmse}} = \mathcal{E}\left\{\mathbf{R}|\mathbf{Z}\right\} = \int \mathbf{R}\, p\left(\mathbf{R}|\mathbf{Z}\right) d\mathbf{R} \tag{10}$$

as well as the minimum mean-square distance (MMSD) estimate of $\mathbf{v}$. The latter is defined as [17]

$$\begin{aligned}\hat{\mathbf{v}}_{\text{mmsd}} &= \arg\min_{\hat{\mathbf{v}}} \mathcal{E}\left\{\sin^2\left(\hat{\mathbf{v}},\mathbf{v}\right)\right\} = \arg\max_{\hat{\mathbf{v}}} \mathcal{E}\left\{\left|\hat{\mathbf{v}}^H\mathbf{v}\right|^2\right\} \\ &= \mathcal{P}\left\{\int \mathbf{v}\mathbf{v}^H p\left(\mathbf{v}|\mathbf{Z}\right) d\mathbf{v}\right\}\end{aligned} \tag{11}$$

where $\mathcal{P}\{.\}$ stands for the principal eigenvector of the matrix between braces. These two quantities will then serve to obtain the beamformer $\mathbf{w} \propto \hat{\mathbf{R}}_{\text{mmse}}^{-1}\hat{\mathbf{v}}_{\text{mmsd}}$.

2. ESTIMATION OF $\mathbf{v}$ AND $\mathbf{R}$

In order to obtain $\hat{\mathbf{R}}_{\text{mmse}}$ in (10) and $\hat{\mathbf{v}}_{\text{mmsd}}$ in (11), one should theoretically obtain the posterior distribution $p\left(\mathbf{R}|\mathbf{Z}\right)$ and $p\left(\mathbf{v}|\mathbf{Z}\right)$ of $\mathbf{R}$ only and $\mathbf{v}$ only or at least their conditional joint posterior distribution $p\left(\mathbf{v},\mathbf{R}|\mathbf{Z}\right)$ by marginalizing with respect to the "nuisance" parameters $\boldsymbol{\alpha}$ and σ_α^2. It turns out that this is intractable [18]. To circumvent this problem, we turn to a solution where $\boldsymbol{\alpha}$ and σ_α^2 are estimated jointly with $\mathbf{v}$ and $\mathbf{R}$. More specifically, a Gibbs sampler [19] is now proposed which generates samples from $p\left(\mathbf{v}|\boldsymbol{\alpha},\mathbf{R},\sigma_\alpha^2,\mathbf{Z}\right)$, $p\left(\boldsymbol{\alpha}|\mathbf{v},\mathbf{R},\sigma_\alpha^2,\mathbf{Z}\right)$, $p\left(\mathbf{R}|\boldsymbol{\alpha},\mathbf{v},\sigma_\alpha^2,\mathbf{Z}\right)$ and $p\left(\sigma_\alpha^2|\boldsymbol{\alpha},\mathbf{v},\mathbf{R},\mathbf{Z}\right)$. As illustrated below, these conditional posterior distributions are easy to simulate.

Let us start with the joint posterior distribution of all variables:

$$\begin{aligned}&p\left(\boldsymbol{\alpha},\mathbf{v},\mathbf{R},\sigma_\alpha^2|\mathbf{Z}\right) \propto p\left(\mathbf{Z}|\boldsymbol{\alpha},\mathbf{v},\mathbf{R},\sigma_\alpha^2\right)\pi\left(\boldsymbol{\alpha}|\sigma_\alpha^2\right)\pi\left(\mathbf{v}\right)\pi\left(\mathbf{R}\right)\pi\left(\sigma_\alpha^2\right) \\ &\propto |\mathbf{R}|^{-K}\text{etr}\left\{-\left(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H\right)^H\mathbf{R}^{-1}\left(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H\right)\right\} \\ &\times\left(\sigma_\alpha^2\right)^{-(a+K+1)}\exp\left\{-\sigma_\alpha^{-2}\boldsymbol{\alpha}^H\boldsymbol{\alpha}\right\}\exp\left\{-b\sigma_\alpha^{-2}\right\} \\ &\times|\mathbf{R}|^{-(\nu+N)}\text{etr}\left\{-(\nu-N)\mathbf{R}^{-1}\right\}\exp\left\{\kappa|\mathbf{v}^H\bar{\mathbf{v}}|^2\right\}.\end{aligned} \tag{12}$$

Using the fact that

$$\begin{aligned}&\sigma_\alpha^{-2}\boldsymbol{\alpha}^H\boldsymbol{\alpha} + \text{Tr}\left\{\left(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H\right)^H\mathbf{R}^{-1}\left(\mathbf{Z}-\mathbf{v}\boldsymbol{\alpha}^H\right)\right\} \\ &= \sigma_\alpha^{-2}\boldsymbol{\alpha}^H\boldsymbol{\alpha} + \text{Tr}\left\{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{Z}\right\} - \mathbf{v}^H\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha} \\ &- \boldsymbol{\alpha}^H\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v} + \left(\boldsymbol{\alpha}^H\boldsymbol{\alpha}\right)\left(\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}\right) \\ &= \left(\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}\right)\left\|\boldsymbol{\alpha}-\frac{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}}{\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}\right\|^2 \\ &+ \text{Tr}\left\{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{Z}-\frac{\mathbf{Z}^H\mathbf{R}^{-1}\mathbf{v}\mathbf{v}^H\mathbf{R}^{-1}\mathbf{Z}}{\sigma_\alpha^{-2}+\mathbf{v}^H\mathbf{R}^{-1}\mathbf{v}}\right\}\end{aligned} \tag{13}$$

along with (12) it ensues that

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

Hence $\boldsymbol{\alpha}$, conditioned on $\mathbf{v},\mathbf{R},\sigma_\alpha^2,\mathbf{Z}$, is Gaussian distributed:

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

Accordingly,

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

which is recognized as a complex Bingham von Mises Fisher (BMF) distribution [20] with parameters $\kappa\bar{\mathbf{v}}\bar{\mathbf{v}}^H - \left(\boldsymbol{\alpha}^H\boldsymbol{\alpha}\right)\mathbf{R}^{-1}$ and $\mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}$, i.e.,

$$\mathbf{v}|\boldsymbol{\alpha},\mathbf{R},\sigma_\alpha^2,\mathbf{Z} \sim \mathsf{BMF}_c\left(\kappa\bar{\mathbf{v}}\bar{\mathbf{v}}^H - \left(\boldsymbol{\alpha}^H\boldsymbol{\alpha}\right)\mathbf{R}^{-1}, \mathbf{R}^{-1}\mathbf{Z}\boldsymbol{\alpha}\right). \tag{17}$$

An efficient sampling scheme for generating samples according to a real BMF distribution was proposed by Hoff [20]. Adaptation of this scheme to generate a complex BMF distributed vector is rather straightforward [18]. Next, the conditional posterior distribution of $\mathbf{R}$ is obtained as

$$p\left(\mathbf{R}|\boldsymbol{\alpha},\mathbf{v},\sigma_\alpha^2,\mathbf{Z}\right) \propto |\mathbf{R}|^{-(\nu+N+K)}\mathrm{etr}\left\{-\mathbf{R}^{-1}\mathbf{M}\left(\boldsymbol{\alpha},\mathbf{v},\mathbf{Z}\right)\right\} \tag{18}$$

with

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

This conditional posterior distribution is an inverse Wishart distribution with $\nu + K$ degrees of freedom and parameter matrix $\mathbf{M}\left(\boldsymbol{\alpha},\mathbf{v},\mathbf{Z}\right)$, which corresponds, up to a scaling factor, to the posterior mean of $\mathbf{R}|\boldsymbol{\alpha},\mathbf{v},\sigma_\alpha^2,\mathbf{Z}$. We would like to emphasize that the particular choice of the prior of $\mathbf{R}$ in (5) results in a posterior mean $\mathbf{M}\left(\boldsymbol{\alpha},\mathbf{v},\mathbf{Z}\right)$ which coincides with the usual diagonal loading. Therefore, the prior (5) with $(\nu - N)\mu\mathbf{I}$ as the parameter matrix is a way to embed diagonal loading in a Bayesian framework, and hence is a means to improve robustness to steering vector errors. It is also worth noticing that the loading level is $(\nu - N)\mu/K$, which provides a way to fix ν and μ. As a final remark, observe that the distributions $p\left(\boldsymbol{\alpha}|\mathbf{v},\mathbf{R},\sigma_\alpha^2,\mathbf{Z}\right)$ and $p\left(\mathbf{v}|\boldsymbol{\alpha},\mathbf{R},\sigma_\alpha^2,\mathbf{Z}\right)$ depend on $\mathbf{R}$ through its inverse $\mathbf{R}^{-1}$. Since our final objective is to derive a beamformer whose weight vector depends directly on $\mathbf{R}^{-1}$, the Gibbs sampler will generate directly the inverse of $\mathbf{R}$ from a Wishart distribution

$$\mathbf{R}^{-1}|\boldsymbol{\alpha},\mathbf{v},\sigma_\alpha^2,\mathbf{Z} \sim \mathsf{CW}\left(\nu+K,\left[\mathbf{M}\left(\boldsymbol{\alpha},\mathbf{v},\mathbf{Z}\right)\right]^{-1}\right). \tag{20}$$

Finally, the conditional posterior of σ_α^2 is given by

$$p\left(\sigma_\alpha^2|\boldsymbol{\alpha},\mathbf{v},\mathbf{R},\mathbf{Z}\right) \propto \left(\sigma_\alpha^2\right)^{-(a+K+1)}\exp\left\{-\sigma_\alpha^{-2}\left[b+\boldsymbol{\alpha}^H\boldsymbol{\alpha}\right]\right\} \tag{21}$$

and hence

$$\sigma_\alpha^2|\boldsymbol{\alpha},\mathbf{v},\mathbf{R},\mathbf{Z} \sim \mathsf{IG}\left(a+K, b+\boldsymbol{\alpha}^H\boldsymbol{\alpha}\right). \tag{22}$$

The Gibbs sampler will thus successively draw samples from (15), (17), (20) and (22), as described in Algorithm 1.

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

$$\hat{\mathbf{v}}_{\text{mmsd}} = \mathcal{P}\left\{\frac{1}{N_r}\sum_{n=N_{\text{bi}}+1}^{N_{\text{bi}}+N_r}\mathbf{v}(n)\mathbf{v}^H(n)\right\} \tag{23a}$$

$$\hat{\mathbf{R}}_{\text{mmse}}^{-1} = \frac{1}{N_r}\sum_{n=N_{\text{bi}}+1}^{N_{\text{bi}}+N_r}\mathbf{R}^{-1}(n) \tag{23b}$$

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

$$\mathbf{w} \propto \hat{\mathbf{R}}_{\text{mmse}}^{-1}\hat{\mathbf{v}}_{\text{mmsd}}. \tag{24}$$

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

Input: initial values $\mathbf{R}^{-1}(0)$, $\mathbf{v}(0)$, $\sigma_\alpha^2(0)$

1: **for** $n = 1, \cdots, N_{\text{bi}} + N_r$ **do**
2: sample $\boldsymbol{\alpha}(n)$ from $p\left(\boldsymbol{\alpha}|\mathbf{v}(n-1),\mathbf{R}(n-1),\sigma_\alpha^2(n-1)\mathbf{Z}\right)$ in (15).
3: sample $\mathbf{v}(n)$ from $p\left(\mathbf{v}|\boldsymbol{\alpha}(n),\mathbf{R}(n-1),\sigma_\alpha^2(n-1),\mathbf{Z}\right)$ in (17).
4: sample $\mathbf{R}^{-1}(n)$ from $p\left(\mathbf{R}^{-1}|\boldsymbol{\alpha}(n),\mathbf{v}(n),\sigma_\alpha^2(n-1),\mathbf{Z}\right)$ in (20).
5: sample $\sigma_\alpha^2(n)$ from $p\left(\sigma_\alpha^2|\boldsymbol{\alpha}(n),\mathbf{v}(n),\mathbf{R}(n),\mathbf{Z}\right)$ in (22).
6: **end for**

Output: sequence of random variables $\boldsymbol{\alpha}(n)$, $\mathbf{v}(n)$, $\mathbf{R}^{-1}(n)$, $\sigma_\alpha^2(n)$.

3. NUMERICAL SIMULATIONS

In this section, we assess the performance achieved with the beamformer in (24). We consider a uniform linear array of $N = 16$ elements spaced a half-wavelength apart. The receiver noise is assumed to be temporally and spatially white with power σ_n^2. The signal of interest (SOI) impinges from the broadside of the array so that $\bar{\mathbf{v}} = \mathbf{a}(0^\circ)$ where $\mathbf{a}(\varphi) = \begin{bmatrix}1 & e^{i\pi\sin\varphi} & \cdots & e^{i\pi(N-1)\sin\varphi}\end{bmatrix}^T/\sqrt{N}$ is the (normalized) steering vector of the array. The signal to noise ratio (SNR) is defined as

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

and is set to $SNR = 0$dB. In order to simulate steering vector errors, we consider pointing errors so that the SOI actually impinges from the direction of arrival (DOA) $\varphi_{\text{true}} = \delta \times HPBW$ where $HPBW$ stands for the half power beam width of the array [1]. It is noteworthy that the true steering vector is **not** generated according to the prior distribution. We also assume that two interference are present with directions of arrival -15° and 20°, and respective interference to noise ratio (INR) equal to 30dB and 20dB. In order to set the values for ν and μ, we use the expression of $\mathbf{M}\left(\boldsymbol{\alpha},\mathbf{v},\mathbf{Z}\right)$ in (19). We set $\nu = K + N$ so that the term due to the data and the term corresponding to diagonal loading have approximately the same weight, and hence μ corresponds to a diagonal loading level. We fix it to 5dB above the white noise level, a good rule of thumb in practice [1]. The Bayesian beamformer in (24) is compared with conventional diagonal loading using the presumed steering vector in (3) with a loading level 5dB above the white noise level. We also consider a "clairvoyant" diagonally loaded beamformer which would have knowledge of $\mathbf{v}$, i.e.,

$$\mathbf{w}_{\text{DL}-\mathbf{v}} \propto \left(K^{-1}\mathbf{Z}\mathbf{Z}^H + \mu\mathbf{I}_N\right)^{-1}\mathbf{v}. \tag{25}$$

The latter is hypothetical but it could serve as a benchmark since it is affected by finite-sample effects only but not steering vector errors. The performance metric for quality assessment of the adaptive beamformer will be the SINR loss with respect to the noise-only-environment, defined by [21]

$$SINR_{\text{loss}} = \frac{\left|\mathbf{w}^H\mathbf{v}\right|^2}{\mathbf{w}^H\mathbf{R}\mathbf{w}}\frac{1}{\sigma_n^{-2}\mathbf{v}^H\mathbf{v}}. \tag{26}$$

First, we study in Figure 2 the sensitivity of the Bayesian beamformer towards the parameter κ which is the only parameter in the Bingham prior distribution of $\mathbf{v}$. It can be observed that for moderate steering vector error ($\delta = 0.2$), the SINR loss of the Bayesian

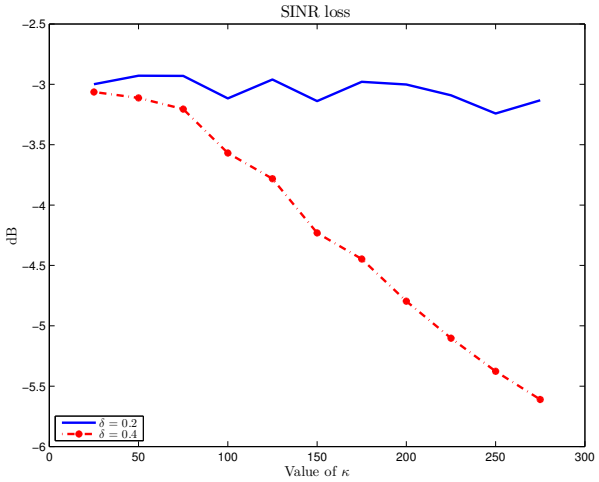

Fig. 2. SINR loss of the adaptive beamformer versus κ. $K = 32$.

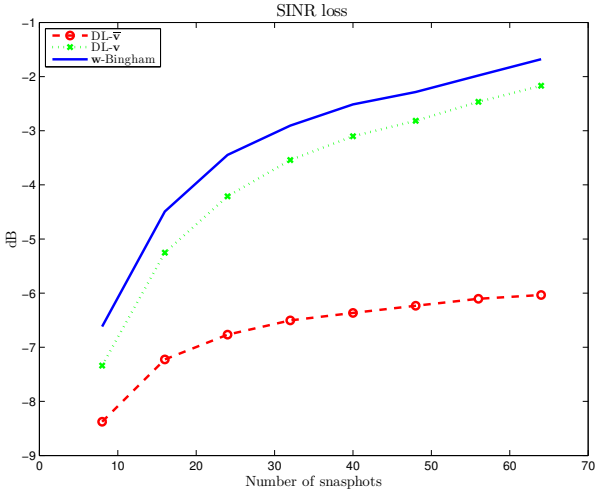

Fig. 3. SINR loss of the adaptive beamformers versus K. $\delta = 0.2$.

beamformer is nearly constant over a large range of values for κ. Therefore, in practice, one need not tune this parameter 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 ($\delta = 0.4$) the value of κ should be chosen appropriately (i.e., not too large) in order to accommodate the possibly large difference between $\mathbf{v}$ and $\bar{\mathbf{v}}$. If one has a rough knowledge of the distance between $\mathbf{v}$ and $\bar{\mathbf{v}}$ then equation (9) can be used to set κ. 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 κ.

We now investigate the performance versus K in Figure 3 and versus δ in Figure 4. There we set $\kappa = 50$. The Bayesian beamformer significantly improves upon conventional diagonal loading using the presumed steering vector. Remarkably enough it also outperforms the diagonally loaded beamformer constructed with the true steering vector. Moreover, it achieves a very high rate of convergence. Indeed, the SINR loss is lower than 3dB at about $K = 2N$, a rate of convergence commensurate with that of an MVDR beamformer, although the signal of interest is present in the measurements and despite the presence of steering vector errors.

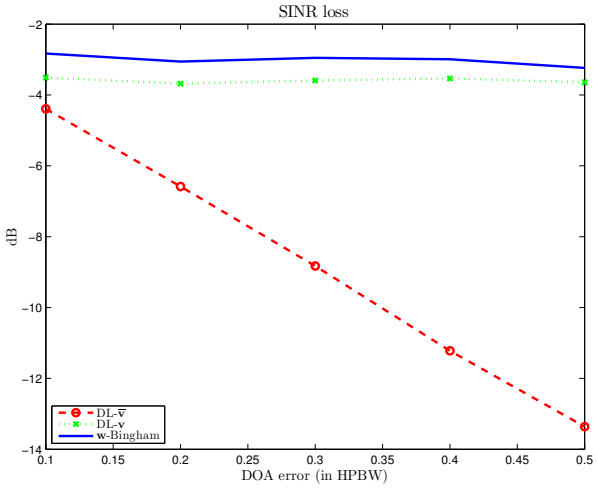

Fig. 4. SINR loss of the adaptive beamformers versus pointing error. $K = 32$.

4. CONCLUSIONS

A new Bayesian approach for robust adaptive beamforming in the presence of steering vector uncertainties was presented. It relies on a Bingham prior distribution for the steering vector of interest and an inverse Wishart prior distribution for the interference covariance matrix, with a parameter matrix proportional to $\mathbf{I}$ which amounts to introducing diagonal loading in a Bayesian framework. 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 algorithm was shown to provide very good performance in low sample support, despite steering vector errors.

5. REFERENCES

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

[2] D. M. Boroson, "Sample size considerations for adaptive arrays," *IEEE Transactions Aerospace Electronic Systems*, vol. 16, no. 4, pp. 446–451, July 1980.

[3] Y. I. Abramovich, "Controlled method for adaptive optimization of filters using the criterion of maximum SNR," *Radio Engineering and Electronic Physics*, vol. 26, pp. 87–95, March 1981.

[4] O. P. Cheremisin, "Efficiency of adaptive algorithms with regularised sample covariance matrix," *Radio Engineering and Electronic Physics*, vol. 27, no. 10, pp. 69–77, 1982.

[5] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Transactions Acoustics Speech Signal Processing*, vol. 35, no. 10, pp. 1365–1376, October 1987.

[6] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Transactions Signal Processing*, vol. 51, no. 7, pp. 1702–1715, July 2003.

[7] S. A. Vorobyov, A. B. Gershman, and Z. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Transactions Signal Processing*, vol. 51, no. 2, pp. 313–324, February 2003.

[8] R. Lorenz and S. P. Boyd, "Robust minimum variance beamforming," *IEEE Transactions Signal Processing*, vol. 53, no. 5, pp. 1684–1696, May 2005.

[9] J. Li, P. Stoica, and Z. Wang, "Doubly constrained robust Capon beamformer," *IEEE Transactions Signal Processing*, vol. 52, no. 9, pp. 2407–2423, September 2004.

[10] L. Du, T. Yardibi, J. Li, and P. Stoica, "Review of user parameter-free robust adaptive beamforming algorithms," *Digital Signal Processing-A Review Journal*, vol. 19, pp. 567–582, July 2009.

[11] L. Du, J. Li, and P. Stoica, "Fully automatic computation of diagonal loading levels for robust adaptive beamforming," *IEEE Transactions Aerospace Electronic Systems*, vol. 46, no. 1, pp. 449 –458, January 2010.

[12] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley, 1982.

[13] J. A. Tague and C. I. Caldwell, "Expectations of useful complex Wishart forms," *Multidimensional Systems and Signal Processing*, vol. 5, pp. 263–279, 1994.

[14] S. Sirianunpiboon, S. T. Howard, and D. Cochran, "A Bayesian derivation of generalized coherence detectors," in *Proceedings ICASSP*, Kyoto, Japan, March 26-30 2012, pp. 3253–3256.

[15] J.T. Kent, "The complex Bingham distribution and shape analysis," *Journal Royal Statistical Society Series B*, vol. 56, no. 2, pp. 285–299, 1994.

[16] K. V. Mardia and I. L. Dryden, "The complex Watson distribution and shape analysis," *Journal Royal Statistical Society Series B*, vol. 61, no. 4, pp. 913–926, 1999.

[17] O. Besson, N. Dobigeon, and J.-Y. Tourneret, "Minimum mean square distance estimation of a subspace," *IEEE Transactions Signal Processing*, vol. 59, no. 12, pp. 5709–5720, December 2011.

[18] O. Besson and S. Bidon, "Robust adaptive beamforming using a Bayesian steering vector error model," *Signal Processing*, 2013, to appear.

[19] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer Verlag, New York, 2nd edition, 2004.

[20] 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*, vol. 18, no. 2, pp. 438–456, June 2009.

[21] J. Ward, "Space-time adaptive processing for airborne radar," Tech. Rep. 1015, Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA, December 1994.