New Sparse-Promoting Prior for the Estimation of a Radar Scene with Weak and Strong Targets

Marie Lasserre and Stéphanie Bidon
ISAE / Université de Toulouse, Toulouse, France
Email: firstname.lastname@isae.fr
François Le Chevalier
MS^3,Delft University of Technology, Delft, The Netherlands
Email: F.LeChevalier@tudelft.nl

***Abstract*—In this paper, we consider the problem of estimating a signal of interest embedded in noise using a sparse signal representation (SSR) approach. This problem is relevant in many radar applications. In particular, estimating a radar scene consisting of targets with wide amplitude range can be challenging since the sidelobes of a strong target can disrupt the estimation of a weak one. Within a Bayesian framework, we present a new sparse-promoting prior designed to estimate this specific type of radar scene. The main strength of this new prior lies in its mixed-type structure which decorrelates sparsity level and target power, as well as in its subdivided support which enables the estimation process to span the whole target power range. This algorithm is implemented through a Monte-Carlo Markov chain. It is successfully evaluated on synthetic and semiexperimental radar data and compared to state-of-the-art algorithms.**

***Index Terms*—sparse representation, Bayesian estimation, Monte Carlo Markov Chain.**

I. Introduction

A recurrent problem in radar applications is the estimation of weak targets in radar scenes with high dynamic range, for example a drone masked by a liner or a vehicle. The CLEAN algorithm was first described in [1] as a means to address this issue. This algorithm "successively removes large targets and their sidelobe responses by subtracting the point spread function of the receiving system centered at the locations of the bright targets" [2]. In other words, when using the CLEAN algorithm on a radar scene with a weak target and a strong one, the latter will first be removed, and the weak target will then appear as the brightest and be estimated. Later on, the so-called "greedy methods", such as matching pursuit (MP) [3] and orthogonal matching pursuit (OMP) [4], [5], were developed in order to solve the same problem and were based on the same principle as the CLEAN algorithm. In fact, these algorithms rely on a sparse representation of the target scene, which allows the estimation of a sidelobe-less signal of interest (SOI), leading to an increased target dynamic range. Consequently, such sparse signal representation (SSR) methods seem to be of particular interest when dealing with target scenes with both weak and strong targets. In SSR, the signal is described as a linear combination of a finite number of atoms from a dictionary, and the problem is written as

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

with

- $\boldsymbol{F} \in \mathbb{C}^{M\times\bar{M}}$ a sparsifying dictionary of size $M \times \bar{M}$ where usually $\bar{M} \geq M$, M being for instance the number of pulses;
- $\boldsymbol{x} \in \mathbb{C}^{\bar{M}}$ the sparse vector having ideally exactly N nonzero components, N being the number of scatterers in the target scene;
- $\boldsymbol{n} \in \mathbb{C}^{M}$ the noise vector of size M.

Within the SSR scope, the greedy methods are very effective in terms of calculation load but several algorithms having higher complexity give better performance, such as the ℓ_1-penalized least-squares algorithm (e.g., basis pursuit de-noising, i.e., BPDN [6], and LASSO [7]). In fact, these methods are equivalent to a Bayesian estimation using a Laplacian sparse-promoting prior distribution [8, p.160].

The Laplacian prior distribution is part of the general class of "normal variance mixture" [9] (also called "scale mixture of Gaussians") that consists of a hierarchical prior distribution $\pi(\boldsymbol{x})$ subject to (s.t.)

$$\pi(\boldsymbol{x}) = \int \pi(\boldsymbol{x}|\boldsymbol{\sigma_x^2})\pi(\boldsymbol{\sigma_x^2})d\boldsymbol{\sigma_x^2} \tag{2}$$

where

$$\pi(\boldsymbol{x}|\boldsymbol{\sigma_x^2}) = \prod_{\bar{m}=0}^{\bar{M}-1} \pi(x_{\bar{m}}|0,\sigma^2_{x\bar{m}}) = \prod_{\bar{m}=0}^{\bar{M}-1} \mathcal{N}\left(x_{\bar{m}}|0,\sigma^2_{x\bar{m}}\right) \tag{3}$$

is the conditional prior assigned to the sparse vector $\boldsymbol{x}$, and $\pi(\boldsymbol{\sigma_x^2})$ is the "mixing distribution"; $\mathcal{N}\left(x_{\bar{m}}|0,\sigma^2_{x\bar{m}}\right)$ designates the normal distribution assigned to $x_{\bar{m}}$ with mean 0 and variance $\sigma^2_{x\bar{m}}$.

Extensive work has been conducted on the normal variance mixtures, most of it consisting in changing the mixing distribution $\pi(\boldsymbol{\sigma_x^2})$, leading to new priors $\pi(\boldsymbol{x})$. In particular, a Jeffreys mixing distribution leads to a Jeffreys prior [10], a Gamma mixing distribution to a Laplace prior [11] and an inverse-Gamma mixing distribution to a Student's *t*-prior [12]. In general, the hierarchical nature of the normal variance mixture model (2) is used to derive efficient inference procedures.

The work of M. Lasserre and S. Bidon is supported by the Délégation Générale de l'Armement under grant 2014.60.0045 ISAE 2014-CIF-R90 (M. Lasserre) and grant 2015.60.0090.00.470.75.01 (S. Bidon).

Part of this work has been presented at the CoSeRa International Workshop in Pisa, Italy in June 2015.

A distinct sparse-promoting prior was presented in [13] that differs from the normal variance mixtures thanks to its mixed type structure. It consists of an atom at 0 and a hierarchical Student's *t*-prior (denoted "HBerStu" in the rest of the paper), and hence is part of the class of "spike and slab" priors [14]. It is formulated as

$$\pi(x_{\bar{m}}|w,\sigma_x^2) = (1-w)\delta(|x_{\bar{m}}|) + w\frac{1}{\pi\sigma_x^2}\exp\left\{-\frac{|x_{\bar{m}}|^2}{\sigma_x^2}\right\} \quad (4a)$$

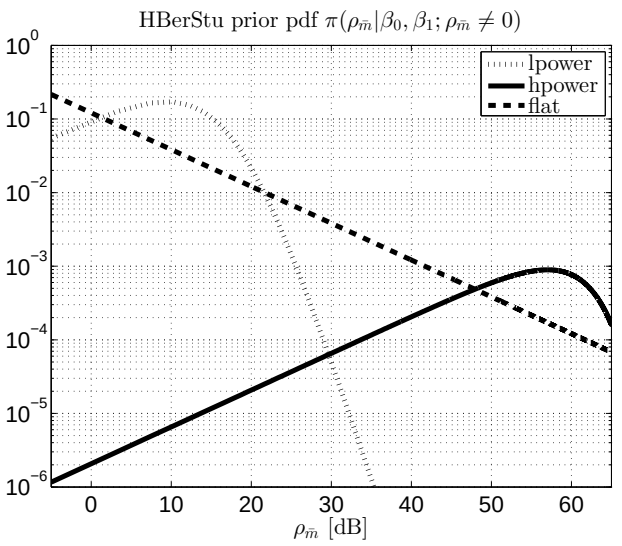
$$\pi(\sigma_x^2|\beta_0,\beta_1) \propto \frac{e^{-\beta_1/\sigma_x^2}}{(\sigma_x^2)^{\beta_0+1}}\mathbb{I}_{[0,+\infty)}(\sigma_x^2). \quad (4b)$$

The atom at zero is added to the normal variance mixture in order to force the sparsity. Then, the continuous part of the prior only deals with the estimation of non-zero targets and does not address the sparsity constraint. It can be seen in (4b) that the mixing distribution is an inverse-gamma distribution with shape and scale parameters (β_0,β_1); its mean and variance, if they exist, are denoted $\left(\mathrm{m}_{\sigma_x^2} = \frac{\beta_1}{\beta_0-1}, \mathrm{var}_{\sigma_x^2} = \frac{\beta_1^2}{(\beta_0-1)^2(\beta_0-2)}\right)$ respectively (resp.). In order to facilitate the upcoming comparison with the new prior proposed, in the following the prior from (4) is reformulated in polar coordinates. In particular, the prior distribution of the modulus $\rho_{\bar{m}} = |x_{\bar{m}}|$ is then expressed as

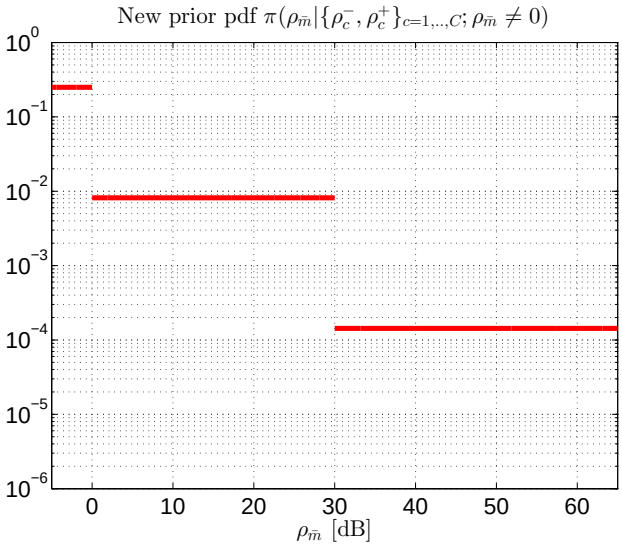
$$\begin{aligned}\pi(\rho_{\bar{m}}|\beta_0,\beta_1) &= \int\int \pi(\rho_{\bar{m}}|\sigma_x^2)\pi(\sigma_x^2)\pi(w)d\sigma_x^2 dw \\ &= \frac{1}{2}\delta(\rho_{\bar{m}}) + \rho_{\bar{m}}\frac{\beta_0}{\beta_1}\left[1+\frac{\rho_{\bar{m}}^2}{\beta_1}\right]^{-(\beta_0+1)} \end{aligned} \quad (5)$$

where we considered that the level of occupancy w is a priori uniformly distributed over $[0,1]$. When using such a prior the operator needs to adjust the hyperparameters of the mixing distribution (β_0,β_1). In Fig.1(a), the continuous part of the prior $\pi(\rho_{\bar{m}})$ from (5) is represented for different tuning of these hyperparameters corresponding to an informative prior centered on either low-power, or high-power targets, or corresponding to a flat non-informative prior. Note that the latter is a priori the most appropriate when no information is available about the power of the targets present in the scene. Nonetheless, it can be seen that even in the case of a flat/non-informative mixing distribution, it is not possible to efficiently scan the whole power range. In practice, and as shown later in Section IV, we have observed that when using such prior, weak targets are not properly estimated in presence of strong ones.

In this paper, we focus on the estimation of radar scenes with high dynamic range, and more precisely on the estimation of weak targets that are possibly hidden by strong ones. Our approach is to consider different classes of power amongst the targets present in the scene (with at least two classes to discriminate weak and strong targets) in order to better estimate the weak targets. The proposed prior is detailed later in Section II; note that the new prior proposed also includes an atom at zero that corresponds to the case when no target is present at the $\bar{m}$-th bin. Its continuous part is represented in Fig.1(b); it can be seen that the new prior is



(a) Former prior



(b) New prior

Fig. 1. Prior pdf of $\rho_{\bar{m}} = |x_{\bar{m}}|$. 1(a) Hierarchical Student's *t*-prior on $x_{\bar{m}}$ for different tuning of the hyperparameters of the mixing distribution: "flat" Non-informative prior, "lpower" Informative prior centered on low-power targets: $(\mathrm{m}_{\sigma_x^2}, \sqrt{\mathrm{var}_{\sigma_x^2}}) = (15,15)$ dB, "hpower" Informative prior centered on high-power targets: $(\mathrm{m}_{\sigma_x^2}, \sqrt{\mathrm{var}_{\sigma_x^2}}) = (60,15)$ dB. 1(b) New approach with three target power classes.

more flexible than a normal variance mixture so far as setting up the hyperparameters in order to efficiently scan a predefined choice of power range scanning is more of an automatic process. To support this point, an automatic setup of the power class boundaries $\{\rho_c^-,\rho_c^+\}_{c=1,..,C}$ is proposed for high dynamic range scenarii. This new prior and the corresponding estimation algorithm lead to better performance on a wide variety of scenarii of interest with weak and strong targets, as shown later in Section IV.

The rest of the paper is organized as follows. First, the new Bayesian model adopted is presented in Section II. Then, the estimation process is detailed in Section III. To conclude, the proposed algorithm is successfully evaluated and compared

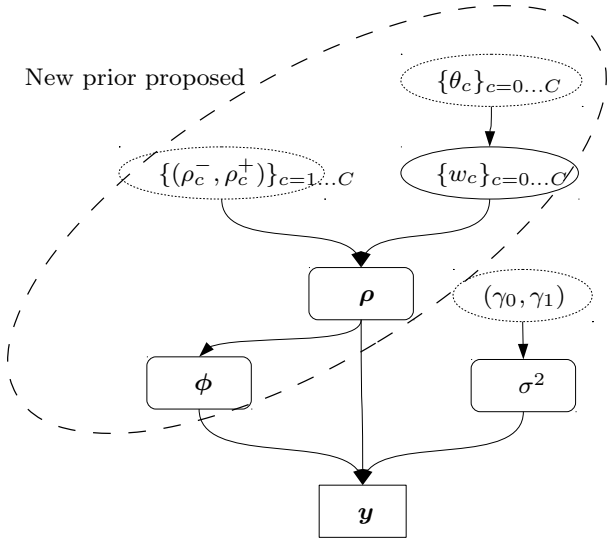


Fig. 2. Directed acyclic graph associated with the hierarchical Bayesian model proposed. The parameters in the dotted circles need to be adjusted by the operator.

to state-of-the-art algorithms in Section IV via numerical simulations on both synthetic and semiexperimental data.

II. Bayesian Model

In this section, we describe the hierarchical Bayesian model adopted, represented graphically in Fig. 2. In particular, a new sparse-promoting prior distribution is presented, that has been designed to estimate radar scenes consisting of targets with very different amplitude.

A. Likelihood

The observation model (1) is adopted. The additive noise $\boldsymbol{n}$ is considered white and distributed following a centered Gaussian distribution with power σ^2

$$\boldsymbol{n}|\sigma^2 \sim \mathcal{CN}_M\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}\right) \tag{6}$$

where $\boldsymbol{I}$ is the identity matrix. The likelihood function is thus given by

$$f(\boldsymbol{y}|\boldsymbol{x},\sigma^2) = \frac{1}{\pi^M\sigma^{2M}}\exp\left\{-\frac{\|\boldsymbol{y}-\boldsymbol{F}\boldsymbol{x}\|_2^2}{\sigma^2}\right\}. \tag{7}$$

B. Prior pdfs of the parameters

1) Sparse vector $\boldsymbol{x}$: As mentioned in Section I, the aim is to estimate a radar scene with targets having wide amplitude range. A new approach is considered and the prior is designed knowing first that the target amplitude is limited by the receiver dynamic range [8, chap.11]. *Then, the power range is divided into several classes, within the dynamic range limitation.*

In this paper, instead of considering $\boldsymbol{x} \in \mathbb{C}^{\bar{M}}$ we parametrize $\boldsymbol{x}$ in terms of modulus $\boldsymbol{\rho}$ and angle $\boldsymbol{\phi}$ subject to (s.t.)

$$\boldsymbol{x} \triangleq \boldsymbol{\rho} \odot e^{i\boldsymbol{\phi}}, \tag{8}$$

where $\odot$ refers to the Hadamard product. (8) is equivalent to the formulation $\forall \bar{m} \in \left\{0, \ \ldots, \ \bar{M}-1\right\}, x_{\bar{m}} = \rho_{\bar{m}}e^{i\phi_{\bar{m}}}$. The elements in $\boldsymbol{\rho}$ and $\boldsymbol{\phi}$ are independent and identically distributed (iid) a priori. Then, we design the new sparse-promoting prior of the modulus vector $\boldsymbol{\rho}$ that consists of an atom at zero and a mixture of uniform distributions

$$\boxed{\pi(\rho_{\bar{m}}|\boldsymbol{w}) = w_0\ \delta(\rho_{\bar{m}}) + \sum_{c=1}^{C} w_c k_c \mathbb{I}_{[\rho_c^-,\rho_c^+)}(\rho_{\bar{m}})} \tag{9}$$

where $\boldsymbol{w} = [w_c]_{c=0,\ldots,C}$ is the vector of class probabilities with $w_0 = 1 - \sum_{c=1}^{C} w_c$. k_c is a scaling constant easily calculated as $k_c = 1/(\rho_c^+ - \rho_c^-)$. Using the prior (9) means that $\rho_{\bar{m}}$ is null with probability w_0; it belongs to class c with probability w_c and is uniformly distributed within this class. Hence, a new approach is considered and the prior is designed knowing first that the target amplitude is limited by the receiver dynamic range, represented by the upper limit ρ_C^+, and then several classes with desired length $\rho_c^+ - \rho_c^-$ and weight w_c are created in order to scan the power range as desired.

The prior distribution assigned to the phase $\boldsymbol{\phi}$ is

$$\pi(\phi_{\bar{m}}|\rho_{\bar{m}} = 0) = \delta(\phi_{\bar{m}}) \tag{10a}$$

$$\pi(\phi_{\bar{m}}|\rho_{\bar{m}} \neq 0) = \frac{1}{2\pi}\mathbb{I}_{[0,2\pi]}(\phi_{\bar{m}}). \tag{10b}$$

Note that the difference between the new prior proposed and the one used by the authors in [13] does not lie in a simple change of mixing distribution as usually done in normal variance mixture models (2).

2) Noise power σ^2: An inverse gamma prior distribution is classically assigned to the white noise power σ^2, mainly for mathematical tractability since it is conjugate to the likelihood (7). The prior pdf of σ^2 is denoted $\sigma^2|\gamma_0,\gamma_1 \sim \mathcal{IG}\left(\gamma_0,\gamma_1\right)$ and expressed as

$$f(\sigma^2|\gamma_0,\gamma_1) \propto \frac{e^{-\gamma_1/\sigma^2}}{(\sigma^2)^{\gamma_0+1}}\mathbb{I}_{[0,+\infty)}(\sigma^2) \tag{11}$$

where γ_0, γ_1 are the shape and scale parameters respectively (resp.). Adequate choice of the hyperparameters (γ_0,γ_1) allows to select an informative, or on the contrary non-informative flat prior distribution. The white noise model can be used in several radar applications (e.g., look-up mode in airborne radars).

C. Prior pdfs of the hyperparameters

1) Vector of class probabilities $\boldsymbol{w}$: Using a conventional solution for class probability vectors, a multivariate Dirichlet distribution is assigned to $\boldsymbol{w}$ [15], denoted $\boldsymbol{w} \sim \mathcal{D}ir\left(\theta_0,...,\theta_C\right)$

$$\pi(\boldsymbol{w}|\theta_0,\ldots,\theta_C) \propto w_0^{\theta_0-1}\mathbb{I}_{[0,1]}(w_0) \times \prod_{c=1}^{C} w_c^{\theta_c-1}\mathbb{I}_{[0,1]}(w_c) \tag{12}$$

where $\theta_0,\ldots,\theta_C > 0$ are the concentration parameters. They can be adjusted in order to favor some classes over the other ones, or on the contrary be equal to 1 to form a non-informative symmetric Dirichlet distribution.

Remark. *The new hierarchical prior proposed requires adjusting more hyperparameters (number of classes C, power class boundaries $\{\rho_c^-, \rho_c^+\}_{c=1,\dots,C}$) than the hierarchical Student's* t*-prior with an atom at 0 from [13] which needs only the two hyperparameters of the mixing distribution (β_0, β_1). Despite this apparent difficulty, it is easy to set up the hyperparameters of the new prior and this will lead to better estimation performance. Both models require approximate information about the thermal noise level σ^2, which is relevant in many radar applications.*

III. Bayesian Estimation

Herein we propose an estimation scheme of the target scene $\boldsymbol{x}$ based on the Bayesian hierarchical model described in (7), (9), (10), (11), (12). In general, the minimum mean square error (MMSE) estimator is calculated

$$\hat{\boldsymbol{x}}_{\text{MMSE}} = \int \boldsymbol{x} f(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x}. \tag{13}$$

If this integral is intractable to derive analytically, a Monte-Carlo Markov chain method can be adopted, and more precisely a Gibbs sampler can be implemented [13]. Generally speaking, considering a target distribution $f(\boldsymbol{\zeta}|\boldsymbol{y})$ where $\boldsymbol{\zeta}$ is a parameter vector, a Gibbs sampler consists in drawing iteratively samples ${\zeta_i}^{(t)}$ according to their conditional posterior distribution $f(\zeta_i|\boldsymbol{y}, \boldsymbol{\zeta}_{-i})$, where $\boldsymbol{\zeta}_{-i}$ is the vector $\boldsymbol{\zeta}$ whose ith element has been removed. Convergence of the sampling chain ${\boldsymbol{\zeta}}^{(t)}$ and of each sub-chain ${\boldsymbol{\zeta}_i}^{(t)}$ is assured, and the sampling chain (resp. each sub-chain) is distributed following the posterior distribution $f(\boldsymbol{\zeta}|\boldsymbol{y})$ (resp. $f(\zeta_i|\boldsymbol{y})$) [16].

In this paper, we use the decomposition $\boldsymbol{x} = \boldsymbol{\rho} \odot e^{i\boldsymbol{\phi}}$ so we calculate a different estimator of the target scene $\boldsymbol{x}$

$$\hat{\boldsymbol{x}}_{\text{class}} = \mathcal{E}\left\{\boldsymbol{\rho} \odot e^{i\boldsymbol{\phi}} | \boldsymbol{y}\right\} \tag{14a}$$

$$= \int_{\boldsymbol{\rho},\boldsymbol{\phi}} \{\boldsymbol{\rho} \odot e^{i\boldsymbol{\phi}}\} f(\boldsymbol{\rho}, \boldsymbol{\phi}|\boldsymbol{y}) d\boldsymbol{\rho} d\boldsymbol{\phi}. \tag{14b}$$

Since this last integral is intractable to derive analytically, we demarginalize it as

$$\hat{\boldsymbol{x}}_{\text{class}} = \int_{\sigma^2,\boldsymbol{\rho},\boldsymbol{\phi},\boldsymbol{w}} \{\boldsymbol{\rho} \odot e^{i\boldsymbol{\phi}}\} f(\sigma^2, \boldsymbol{\rho}, \boldsymbol{\phi}, \boldsymbol{w}|\boldsymbol{y}) d\sigma^2 d\boldsymbol{\rho} d\boldsymbol{\phi} d\boldsymbol{w}. \tag{15}$$

Then, a Gibbs sampler is implemented to obtain samples following the joint posterior distribution $f(\sigma^2, \boldsymbol{\rho}, \boldsymbol{\phi}, \boldsymbol{w}|\boldsymbol{y})$ and compute $\hat{\boldsymbol{x}}_{\text{class}}$. More specifically, samples $\boldsymbol{\zeta}^{(t)} = [{\sigma^2}^{(t)}, \boldsymbol{\rho}^{(t)}, \boldsymbol{\phi}^{(t)}, \boldsymbol{w}^{(t)}]$ are generated iteratively according to their conditional posterior distribution. The estimator $\hat{\boldsymbol{x}}_{\text{class}}$ is built empirically using N_r samples drawn after a burn-in period N_{bi} as

$$\hat{\boldsymbol{x}}_{\text{class}} = N_r^{-1} \sum_{t=1}^{N_r} \boldsymbol{\rho}^{(t+N_{bi})} \odot e^{i\boldsymbol{\phi}^{(t+N_{bi})}} \tag{16}$$

which is the empirical mean of all the samples $\boldsymbol{x}^{(t)} = \boldsymbol{\rho}^{(t)} \odot e^{i\boldsymbol{\phi}^{(t)}}$. The conditional posterior distributions used in the Gibbs sampler are obtained from the joint posterior pdf of $\boldsymbol{\rho}, \boldsymbol{\phi}, \sigma^2, \boldsymbol{w}|\boldsymbol{y}$

$$f(\boldsymbol{\rho}, \boldsymbol{\phi}, \sigma^2, \boldsymbol{w}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\rho}, \boldsymbol{\phi}, \sigma^2) \pi(\boldsymbol{\rho}|\boldsymbol{w}) \pi(\boldsymbol{\phi}|\boldsymbol{\rho}) \pi(\sigma^2) \pi(\boldsymbol{w}). \tag{17}$$

They are expressed in the remaining of this Section.

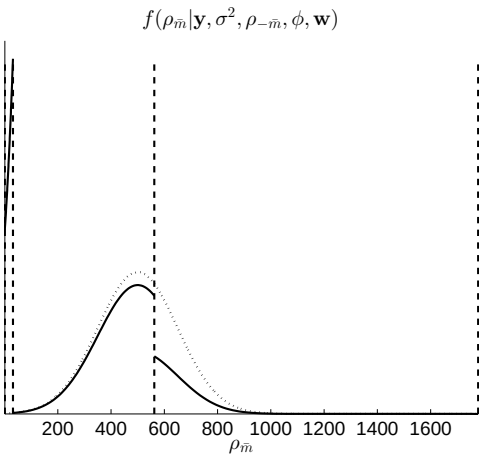


Fig. 3. Conditional posterior distribution (18) of $\rho_{\bar{m}}$ with $\mu_{\bar{m}} = 500$, $\eta_{\bar{m}} = 150$ and $\boldsymbol{w}_{\bar{m}} = [10, 15, 60, 15]\%$. The dashed lines represent the limits of the three non-zero power classes. The dotted curve is the Gaussian $\mathcal{N}\left(\mu_{\bar{m}}, \eta_{\bar{m}}^2\right)$.

A. Sampling of $\boldsymbol{\rho}$

The elements in $\boldsymbol{\rho}$ are a priori iid so the vector can be easily sampled element-wise. (9) and (17) are used to calculate the conditional posterior pdf of $\rho_{\bar{m}}$

$$\begin{aligned} f(\rho_{\bar{m}}|\boldsymbol{y}, \sigma^2, \boldsymbol{\rho}_{-\bar{m}}, \boldsymbol{\phi}, \boldsymbol{w}) &\propto f(\boldsymbol{y}|\boldsymbol{\rho}, \boldsymbol{\phi}, \sigma^2) \pi(\rho_{\bar{m}}|\boldsymbol{w}) \\ &\propto \exp\left\{-\sigma^{-2}\left[\rho_{\bar{m}}^2 \parallel \boldsymbol{f}_{\bar{m}} \parallel^2 - 2\rho_{\bar{m}} \mathcal{R}\text{e}\left\{\boldsymbol{e}_{\bar{m}}^H \boldsymbol{f}_{\bar{m}} e^{i\phi_{\bar{m}}}\right\}\right]\right\} \\ &\quad \times \left[w_0\, \delta(\rho_{\bar{m}}) + \sum_{c=1}^{C} w_c k_c \mathbb{I}_{[\rho_c^-, \rho_c^+)}(\rho_{\bar{m}})\right] \\ &\propto w_{\bar{m},0}\, \delta(\rho_{\bar{m}}) \\ &\quad + \sum_{c=1}^{C} \frac{w_{\bar{m},c} k_{\bar{m},c}}{\sqrt{2\pi\eta_{\bar{m}}^2}} \exp\left\{-\frac{\eta_{\bar{m}}^{-2}}{2}(\rho_{\bar{m}} - \mu_{\bar{m}})^2\right\} \mathbb{I}_{[\rho_c^-, \rho_c^+)}(\rho_{\bar{m}}) \end{aligned} \tag{18}$$

where for $c = 1, \dots, C$

$$\eta_{\bar{m}}^2 = \frac{1}{2}\sigma^2 \parallel \boldsymbol{f}_{\bar{m}} \parallel^{-2} \tag{18a}$$

$$\mu_{\bar{m}} = \parallel \boldsymbol{f}_{\bar{m}} \parallel^{-2} \mathcal{R}\text{e}\left\{\boldsymbol{e}_{\bar{m}}^H \boldsymbol{f}_{\bar{m}} e^{i\phi_{\bar{m}}}\right\} \tag{18b}$$

$$w_{\bar{m},c} = \frac{w_c \times \dfrac{k_c}{k_{\bar{m},c}} (2\pi\eta_{\bar{m}}^2)^{1/2} \exp\left\{\dfrac{1}{2}\eta_{\bar{m}}^{-2}\mu_{\bar{m}}^2\right\}}{w_0 + \sum_{c=1}^{C} w_c \dfrac{k_c}{k_{\bar{m},c}} (2\pi\eta_{\bar{m}}^2)^{1/2} \exp\left\{\dfrac{1}{2}\eta_{\bar{m}}^{-2}\mu_{\bar{m}}^2\right\}} \tag{18c}$$

and $w_{\bar{m},0} = 1 - \sum_{c=1}^{C} w_{\bar{m},c}$. $k_{\bar{m},c}$ is a scaling constant consecutive to the truncation of the Gaussian with mean $\mu_{\bar{m}}$ and variance $\eta_{\bar{m}}^2$ on the interval $[\rho_c^-, \rho_c^+)$; this truncated normal distribution is denoted $\mathcal{N}_{[\rho_c^-, \rho_c^+)}\left(\mu_{\bar{m}}, \eta_{\bar{m}}^2\right)$. The conditional posterior distribution of $\rho_{\bar{m}}$ (18) is a mixed-type distribution: it has an atom at 0 and a continuous component which is a mixture of truncated Gaussian distributions. Note that the mean and variance $(\mu_{\bar{m}}, \eta_{\bar{m}}^2)$ of each Gaussian do not depend on the class c, but that the distribution $\mathcal{N}\left(\mu_{\bar{m}}, \eta_{\bar{m}}^2\right)$ is weighted differently on each class by $w_{\bar{m},c}$. This is illustrated in Fig. 3 where we can see an example of the continuous part of the conditional posterior pdf of $\rho_{\bar{m}}$.

Remark. *When using a Fourier dictionary as in the numerical*

simulations, $\| \boldsymbol{f}_{\bar{m}} \|^2 = 1$ *so that* $\eta_{\bar{m}}^2 = \eta^2$ *does not depend on the grid index* $\bar{m}$.

B. Sampling of $\boldsymbol{\phi}$

Vector $\boldsymbol{\phi}$ can also be sampled element-wise easily since its elements are supposed a priori iid. (10) and (17) are used to calculate the conditional posterior pdf of $\phi_{\bar{m}}$

$$\begin{aligned} f(\phi_{\bar{m}}|\boldsymbol{y},\sigma^2,\boldsymbol{\phi}_{-\bar{m}},\boldsymbol{\rho}) &\propto f(\boldsymbol{y}|\boldsymbol{\rho},\boldsymbol{\phi},\sigma^2)\pi(\phi_{\bar{m}}|\rho_{\bar{m}}) \\ &\propto \exp\left\{-\sigma^{-2}\left[-2\rho_{\bar{m}}\,\mathcal{R}\mathrm{e}\left\{\boldsymbol{e}_{\bar{m}}^H\boldsymbol{f}_{\bar{m}}e^{i\phi_{\bar{m}}}\right\}\right]\right\}\pi(\phi_{\bar{m}}|\rho_{\bar{m}}) \\ &\propto \exp\left\{2\sigma^{-2}\rho_{\bar{m}}|\boldsymbol{f}_{\bar{m}}^H\boldsymbol{e}_{\bar{m}}|\cos\left(\phi_{\bar{m}}-\psi_{\bar{m}}\right)\right\}\pi(\phi_{\bar{m}}|\rho_{\bar{m}}). \end{aligned}$$

According to (10a), we have

$$f(\phi_{\bar{m}}|\boldsymbol{y},\sigma^2,\boldsymbol{\phi}_{-\bar{m}},\boldsymbol{\rho};\rho_{\bar{m}}=0) = \delta(\phi_{\bar{m}}) \tag{19}$$

and

$$\begin{aligned} f(\phi_{\bar{m}}|\boldsymbol{y},\sigma^2,\boldsymbol{\phi}_{-\bar{m}},\boldsymbol{\rho};\rho_{\bar{m}}\neq 0) \propto \\ \exp\left\{\kappa_{\bar{m}}\cos\left(\phi_{\bar{m}}-\psi_{\bar{m}}\right)\right\}\mathbb{I}_{[0,2\pi)}(\phi_{\bar{m}}). \end{aligned} \tag{20}$$

We recognize from (20) a von Mises-Fisher distribution, denoted as $\phi_{\bar{m}}|\boldsymbol{y},\sigma^2,\boldsymbol{\phi}_{-\bar{m}},\boldsymbol{\rho};\rho_{\bar{m}}\neq 0 \sim \mathcal{VM}(\kappa_{\bar{m}},\psi_{\bar{m}})$, with concentration parameter and mean direction resp. [17]

$$\kappa_{\bar{m}} = 2\rho_{\bar{m}}\sigma^{-2}|\boldsymbol{f}_{\bar{m}}^H\boldsymbol{e}_{\bar{m}}| \tag{20a}$$

$$\psi_{\bar{m}} = \angle\boldsymbol{f}_{\bar{m}}^H\boldsymbol{e}_{\bar{m}}, \tag{20b}$$

where $\angle$ represents the angle in $[0,2\pi)$.

C. Sampling of σ^2

The conditional posterior distribution of σ^2 is simply calculated as

$$\sigma^2|\boldsymbol{y},\boldsymbol{\rho},\boldsymbol{\phi} \sim \mathcal{IG}\left(\gamma_0+M,\gamma_1+\| \boldsymbol{y}-\boldsymbol{F}(\boldsymbol{\rho}\odot e^{i\boldsymbol{\phi}})\|_2^2\right). \tag{21}$$

D. Sampling of $\boldsymbol{w}$

The conditional posterior distribution of $\boldsymbol{w}$ is calculated using (12) and (17)

$$\begin{aligned} f(\boldsymbol{w}|\boldsymbol{y},\boldsymbol{\rho}) &\propto \pi(\boldsymbol{\rho}|\boldsymbol{w})\pi(\boldsymbol{w}) \\ &\propto \prod_{\bar{m}=0}^{\bar{M}-1}\left\{w_0\;\delta(\rho_{\bar{m}}) + \sum_{c=1}^{C} w_c k_c \mathbb{I}_{[\rho_c^-,\rho_c^+)}(\rho_{\bar{m}})\right\} \times \prod_{c=0}^{C} w_c^{\theta_c-1} \\ &\propto w_0^{n_0+\theta_0-1} \times \prod_{c=1}^{C} w_c^{n_c+\theta_c-1} \end{aligned} \tag{22}$$

where $n_c = \#\{\bar{m}|\rho_{\bar{m}} \in [\rho_c^-,\rho_c^+)\}$, i.e., the number of scatterers in class c, and $n_0 = \bar{M} - \sum_{c=1}^{C} n_c$. We recognize from (22) a multivariate Dirichlet distribution with concentration parameters $(n_0+\theta_0, n_1+\theta_1, \ldots, n_C+\theta_C)$. $\boldsymbol{w}$ is sampled after the method proposed in [18, chap.11], which uses a gamma distribution with parameters $(n_c+\theta_c, 1)$.

The Gibbs sampler computed is summarized in Fig. 4 where the sampling of each parameter is detailed. Note that the algorithm does not require the number of targets in the scene as an input parameter.

Require: $\boldsymbol{y}$, (γ_0,γ_1), $(\rho_c^-,\rho_c^+)_{c=1,\ldots,C}$, $[\theta_c]_{c=0,\ldots,C}$
Ensure: $\hat{\boldsymbol{x}}_{\text{class}}$
  {Initialization}
  $\boldsymbol{\rho}^{(0)}$, $\boldsymbol{\phi}^{(0)}$, $\boldsymbol{\alpha}^{(0)}$
  {Iterations}
  for $n = 1$ to $N_{bi}+N_r$ **do**
    $\sigma^{2(n)}|\boldsymbol{y},\boldsymbol{x}^{(n-1)} \sim \mathcal{IG}\left(\gamma_0+M,\gamma_1+\| \boldsymbol{y}-\boldsymbol{F}\boldsymbol{x}\|_2^2\right)$
    $\boldsymbol{w}^{(n)}|\boldsymbol{\rho}^{(n-1)} \sim \mathcal{D}ir\left(\theta_0+n_0,\ldots,\theta_C+n_C\right)$
    for $\bar{m} = 0$ to $\bar{M}-1$ **do**
      ${\alpha_{\bar{m}}}^{(n)} \sim \text{discrete}\left(\{0,\ldots,C\},\boldsymbol{w}_{\bar{m}}\right)$
      if ${\alpha_{\bar{m}}}^{(n)} = c$ with $c \neq 0$ **then**
        ${\rho_{\bar{m}}}^{(n)}|\boldsymbol{y},{\sigma^{2}}^{(n)},{\boldsymbol{\rho}_{-\bar{m}}}^{(n)},{\boldsymbol{\phi}}^{(n)} \sim \mathcal{N}_{[\rho_c^-,\rho_c^+)}\left(\mu_{\bar{m}},\eta_{\bar{m}}^2\right)$
        ${\phi_{\bar{m}}}^{(n)}|\boldsymbol{y},{\sigma^{2}}^{(n)},{\boldsymbol{\phi}_{-\bar{m}}}^{(n)},{\boldsymbol{\rho}}^{(n)} \sim \mathcal{VM}(\kappa_{\bar{m}},\psi_{\bar{m}})$
      else
        ${\rho_{\bar{m}}}^{(n)} = 0$
        ${\phi_{\bar{m}}}^{(n)} = 0$
      end if
      ${x_{\bar{m}}}^{(n)} = {\rho_{\bar{m}}}^{(n)}e^{i{\phi_{\bar{m}}}^{(n)}}$
    end for
  end for
  {Estimator}
  $\hat{\boldsymbol{x}}_{\text{class}} = \frac{1}{N_r}\sum_{t=1}^{N_r}\boldsymbol{\rho}^{(t+N_{bi})} \odot e^{i\boldsymbol{\phi}^{(t+N_{bi})}}$

Fig. 4. Gibbs sampler used in the proposed algorithm. $\alpha_{\bar{m}}$ is the label associated with the $\bar{m}$th element of vectors $\boldsymbol{\rho}$ and $\boldsymbol{\phi}$; it is described in Appendix A. It follows a discrete distribution within $\{0,\ldots,C\}$ with corresponding probabilities $\{w_{\bar{m},0},\ldots,w_{\bar{m},C}\}$.

IV. Numerical simulations

The proposed algorithm is evaluated via numerical simulations on both synthetic and semiexperimental radar data. It is compared with the algorithm based on the hierarchical Student's *t*-prior with an atom at 0 from [13] for different tuning of the mixing distribution: flat prior, prior adjusted to low-power targets or high-power targets (numerical values given in Fig. 1). It is also compared to state-of-the-art algorithm CLEAN [1], [2] and displayed with a classical spectral estimate, namely APES [19].

Since we are considering conventional radar data, a simple Fourier matrix is used as a sparsifying dictionary $\boldsymbol{F}$. The proposed algorithm is adjusted to the following values. Three non-zero target power classes are used, and the class boundaries $\{\rho_c^-,\rho_c^+\}_{c=1,\ldots,C}$ are:

- $(-\infty,\sigma_{dB}^2)$ dB,
- $[\sigma_{dB}^2, 30+\sigma_{dB}^2)$ dB,
- $[30+\sigma_{dB}^2, 65+\sigma_{dB}^2)$ dB

where $\sigma_{dB}^2 = 10\log 10(\sigma^2)$. The hyperparameters $\{\theta_c\}_{c=0\ldots C}$ adjusting the vector of class probabilities $\boldsymbol{w}$ are set to 1 (non-informative prior). A non-informative Jeffreys prior is chosen for the noise power σ^2, so that $(\gamma_0,\gamma_1) = (0,0)$.

A. Synthetic data

The performance of the proposed algorithm is first evaluated on synthetic radar data via the calculation of the normalized mean square error (nMSE) of the reconstructed target scene $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$ (see discussion in [20], [21]) and the nMSEs of the elements of $\hat{\boldsymbol{x}}_{\text{class}}$ where a target is present in order to evaluate the estimation performance of each target individually. These performance metrics are calculated after $N_{mc} = 200$ Monte-Carlo simulations as

$$\text{nMSE}(\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}) = \frac{1}{N_{mc}} \sum_{n=1}^{N_{mc}} \frac{\| \boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}{}^{(n)} - \boldsymbol{F}\boldsymbol{x} \|_2^2}{\| \boldsymbol{F}\boldsymbol{x} \|_2^2} \quad (23a)$$

$$\text{nMSE}(\hat{\boldsymbol{x}}_{\text{class}_i}) = \frac{1}{N_{mc}} \sum_{n=1}^{N_{mc}} \frac{|\hat{\boldsymbol{x}}^{(n)}_{\text{class}_i} - x_i|^2}{|x_i|^2}. \quad (23b)$$

The synthetic data are generated following a model usually used in radar applications where the SOI is modeled by a sum of cisoids embedded in additive white noise $\boldsymbol{n}$ generated using (6) [22]; the noise power σ^2 is set to 1. The sparsifying dictionary $\boldsymbol{F}$ is a simple 1D-Fourier matrix.

1) Performance of the proposed algorithm: In the first scenario, the target scene consists of a strong target with a post-processing signal-to-noise ratio (SNR) of 60 dB, defined as (the columns of $\boldsymbol{F}$ being unitary)

$$\text{SNR} = |x_{\bar{m}}|^2/\sigma^2, \quad (24)$$

surrounded by two weak targets with varying SNR located on the previous and next frequency bins of the analysis grid, i.e., $f_{d_{l_1}} = 0$, $f_{d_s} = f_{d_{l_1}} + 1/\bar{M} = f_{d_{l_2}} - 1/\bar{M}$ where $f_{d_{l_1}}$ (resp. $f_{d_{l_2}}$, f_{d_s}) is the normalized Doppler frequency of the left-hand side low-power target (resp. right-hand side low-power target, strong target). Note that with definition (24), the integration gain $10\log_{10}(M)$ is already comprised in the SNR value.

a) Result after one realization: Fig. 5 shows an example of the target scene estimated by the proposed algorithm and the one based on the hierarchical Student's *t*-prior with an atom at 0 from [13] for different tuning of the mixing distribution. In this scenario, the weak targets have a post-processing SNR of 14 dB. It can be seen that both algorithms estimate the strong target, but that the proposed algorithm is more accurate in terms of amplitude estimation of the weak targets (the latter are even missed in this specific realization when using a flat mixing distribution). Note that the weak targets are not resolved by the APES algorithm. This first realization highlights tendencies that are confirmed in the following Monte-Carlo simulations.

b) Monte Carlo results for different SNRs of weak targets: The performance of the algorithm with respect to (wrt) the SNR of the weak targets is evaluated through the calculation of the nMSE of the target scene $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$, the nMSE of $\hat{x}_{l_1}$ (which corresponds to the left-hand side weak target) and the nMSE of $\hat{x}_s$ (which corresponds to the strong target). The results are presented in Fig. 6. It can be seen that the proposed algorithm outperforms the algorithm based on the hierarchical Student's *t*-prior with an atom at 0 from [13] in terms of reconstruction of the target scene $\boldsymbol{F}\boldsymbol{x}$, regardless of the tuning of the mixing distribution. In fact, both algorithms give the same performance in terms of estimation of the strong target (x_s, Fig. 6(c)) but the proposed algorithm better estimates the weak targets (e.g., x_{l_1}, Fig. 6(b)) (the corresponding figure for x_{l_2} is not depicted but similar). More precisely, it should be noted that the nMSE of $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$ is not monotonic in Fig. 6(a):

- in the first part, the normalized MSE increases with the SNR of the weak targets. The nMSE obtained with the hierarchical Student's *t*-prior with an atom at 0 from [13] (square-markers) matches the theoretical normalized MSE when both weak targets are missed (green line), namely $10\log_{10}\left(2|x_l|^2/\left(|x_s|^2+2|x_l|^2\right)\right)$. Thus, we can affirm that the weak targets are missed by the estimation algorithm based on the hierarchical Student's *t*-prior with an atom at 0 when their SNR is below 14 dB.
- in the second part, the nMSE decreases, meaning that the weak targets are more often seen and correctly estimated by the algorithm.

Thus, misdetection of the very low targets is more recurrent with the algorithm based on the hierarchical Student's *t*-prior with an atom at 0 from [13] than with the proposed algorithm.

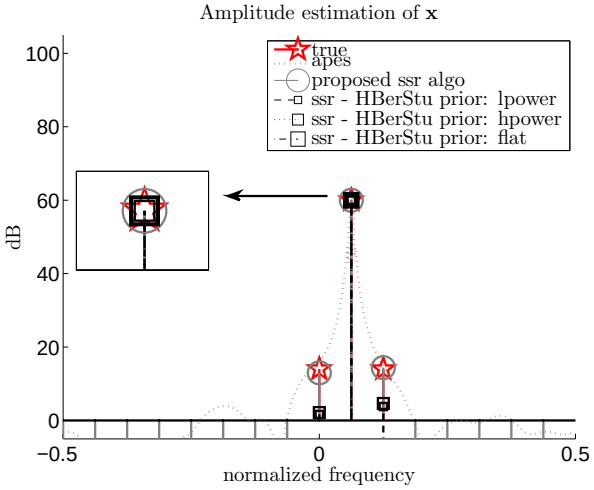


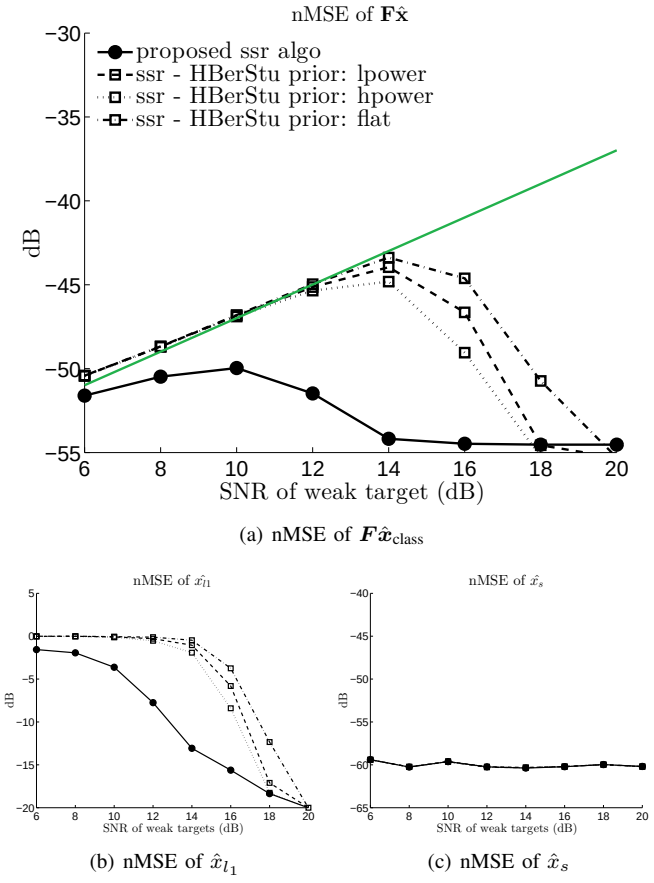
Fig. 5. Example after one realization of the proposed SSR algorithm and the one based on the hierarchical Student's *t*-prior with an atom at 0 from [13] ("ssr - HBerStu prior") for different shape of the mixing distribution: flat prior ("flat prior"), prior adjusted to low-power targets ("lpower") or high-power targets ("hpower"). The scenario consists in a strong target (SNR=60 dB) surrounded by two weak targets with SNR=14 dB. $\bar{M} = M = 16$, $N_r = 1000$, $N_{bi} = 200$.

2) Discussion of the choice of power classes: As mentioned in Section II, the radar operator needs to adjust the hyperparameters controlling the target power classes. Thus, the influence of these classes on the performance of the proposed algorithm is investigated. Using the scenario of Section IV-A1 and considering $\sigma^2 = 1$, five different choices of power classes are studied:

1) $(-\infty \quad 65)$ dB
2) $(-\infty \quad 0), [0 \quad 30), [30 \quad 65)$ dB
3) $(-\infty \quad 0), [0 \quad 30), [30 \quad 55), [55 \quad 65)$ dB
4) $[0 \quad 30), [30 \quad 65)$ dB
5) $(-\infty \quad 30), [30 \quad 65)$ dB

These choices will first indicate if it really is necessary to subdivide the truncated power range into several classes. They will also show the importance of the first non-zero class $(-\infty \quad 0)$ dB, and the nonnecessity of a fine subdivision within the high-power range.

The performance of the proposed algorithm with the different choices of target power classes is assessed on this scenario via the calculation of the nMSE of the reconstructed target scene $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$ as represented in Fig. 7. First, it can be seen that using only one class $(-\infty \quad 65)$ dB gives the worst



(a) nMSE of $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$

(b) nMSE of $\hat{x}_{l_1}$ (c) nMSE of $\hat{x}_s$

Fig. 6. Comparison between the performance of the proposed algorithm (black plain line and circles) and the one based on the hierarchical Student's *t*-prior with an atom at 0 from [13] for different shape of the mixing distribution: flat prior (dash-dotted line and squares), prior centered on low-power targets (dashed line and squares) and high-power targets (dotted line and squares). The scenario consists of a strong target (SNR=60 dB) surrounded by two weak targets with varying SNR s.t. $f_{d_{l_1}} = 0$, $f_{d_s} = f_{d_{l_1}} + 1/\bar{M} = f_{d_{l_2}} - 1/\bar{M}$. $\bar{M} = M = 16$, $N_r = 1000$, $N_{bi} = 200$.

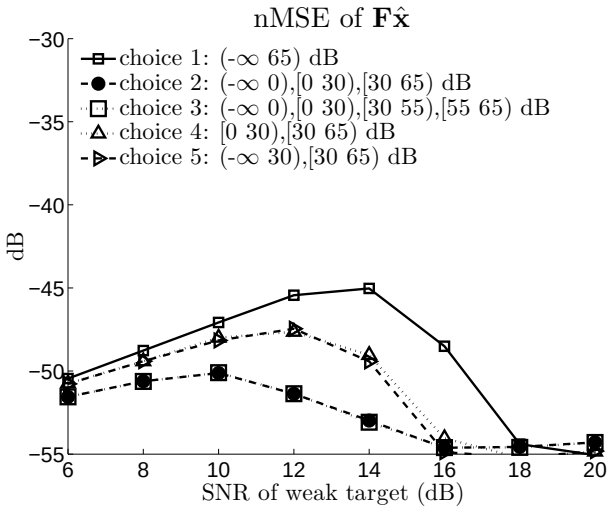


Fig. 7. Performance of the proposed algorithm for different choice of power classes (nMSE of $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$). The scenario is the same as in Section IV-A1. $\bar{M} = M = 16$, $N_r = 1000$, $N_{bi} = 200$.

performance: the weak targets are not seen by the algorithm. Then, comparing choices 2 and 3 shows that it does not seem necessary to finely subdivide the high-power range: the strong targets belonging to classes $[30 \quad 55), [55 \quad 65)$ dB are way above noise level so the algorithm easily estimates their amplitude, whatever the power classes used. To finish, by focusing on choices 2 and 4, it can be seen that adding the class $(0 \quad 1)$, i.e., $(-\infty \quad 0)$ dB improves the performance. Indeed, it gives the algorithm an alternative when it hesitates between the two choices "no target" (zero class) and "a target above noise level" (class $[0 \quad 30)$ dB). Comparing choices 2 and 5 supports the affirmation that a subdivision of the low-power range via the class $(-\infty \quad 0)$ dB is necessary.

Thus, an adequate setup of the class boundaries would be (in dB):

1) a first class $(-\infty \quad \sigma^2_{dB})$,
2) a second class $[\sigma^2_{dB} \quad \sigma^2_{dB} + 30)$,
3) a last class $[\sigma^2_{dB} + 30 \quad \mathcal{A}_{\max})$,

where $\mathcal{A}_{\max}$ designates the maximum detectable amplitude (corresponding to the receiver limitation). Note that the operator needs to first estimate the noise level σ^2, which is quite common in radar applications. Robustness of the algorithm towards the σ^2-estimate used when setting up the power classes is discussed in Section IV-A3.

The selected setup of the target power classes has been tested on many different scenarii and shown to be appropriate compared to other settings. In order to illustrate this point, we have evaluated the proposed algorithm on a new scenario consisting of five targets randomly situated on the frequency axis. The post-processing amplitude of three out of five targets is uniformly distributed over the interval [10 14] dB, the amplitude of the other two over the interval [30 65] dB. The performance of the algorithm is assessed via the calculation of the MSE of the reconstructed target scene $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$, the MSE of the lowest target in the scene x_l and of the strongest target x_s; the results are presented in Table I. It can be seen that the proposed algorithm is always more accurate than the one based on the hierarchical Student's *t*-prior with an atom at 0 from [13]: the MSE of the reconstructed target scene $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$ is lower by more than 4 dB. More precisely, the weak targets are correctly estimated by the proposed algorithm, while this is not always the case with the one based on the hierarchical Student's *t*-prior with an atom at 0. This confirms the adequacy of the proposed target power classes setup.

3) Robustness towards thermal noise approximation: As explained earlier, an approximate value of the thermal noise σ^2 is required to setup the power classes used in the algorithm, and usually in radar applications it is rather accurately known. Nonetheless, we evaluate the algorithm's robustness towards the thermal noise approximation via the calculation of the nMSE of $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$ for different choices of power classes corresponding to an error of $\delta_{\sigma^2} = \pm 5$ dB on the thermal noise approximate. The target power classes used are

- $(-\infty \quad 5), [5 \quad 35), [35 \quad 65)$ dB corresponding to an error of $\delta_{\sigma^2} = +5$ dB,
- $(-\infty \quad -5), [-5 \quad 25), [25 \quad 65)$ dB corresponding to an error of $\delta_{\sigma^2} = -5$ dB.

The scenario considered is the same as that of Section IV-A1,

TABLE I
COMPARISON BETWEEN THE ALGORITHM WITH OR WITHOUT POWER CLASSES VIA DIFFERENT METRICS ON A SCENARIO WITH 5 RANDOMLY-LOCATED TARGETS (THREE WEAK AND TWO STRONG). THE MSE OF x_s (RESP. x_l) CORRESPONDS TO THE MSE ON THE STRONGEST ELEMENT OF $\boldsymbol{x}$ (RESP. THE LOWEST).

	pw classes	npwclasses: lpower	npwclasses: hpower	npwclasses: flat
MSE of $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$	12.14	16.50	16.95	16.75
MSE of x_s	-0.34	-0.33	-0.34	-0.33
MSE of x_l	6.71	10.54	10.93	10.73

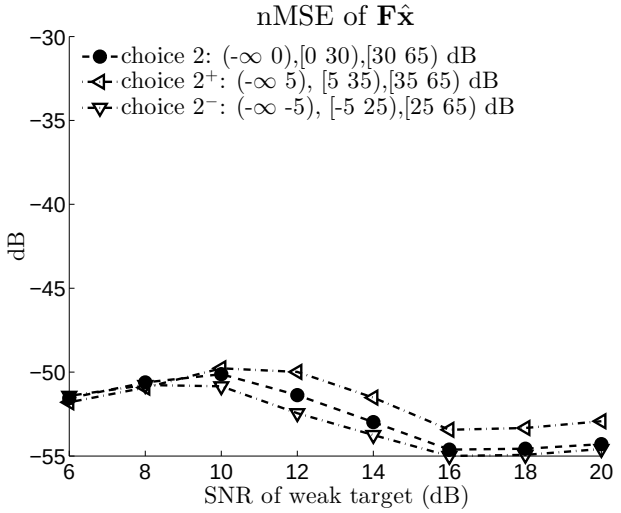


Fig. 8. Robustness of the proposed algorithm wrt the thermal noise approximate (nMSE of $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$). The scenario is the same as in Section IV-A1. $\bar{M} = M = 16$, $N_r = 1000$, $N_{bi} = 200$.

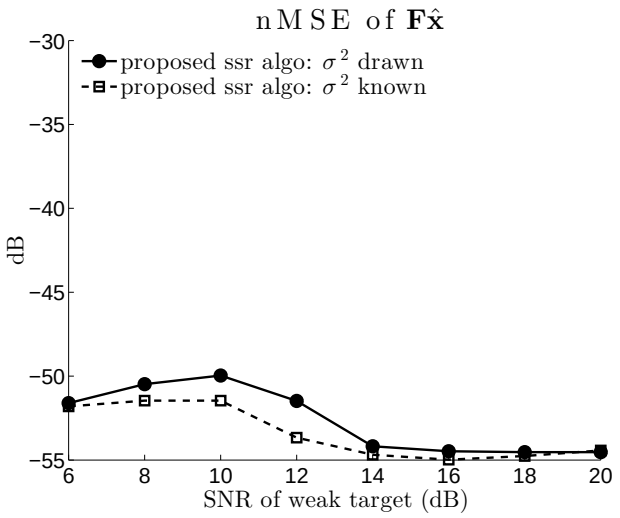


Fig. 9. Performance of the proposed algorithm when the thermal noise parameter σ^2 is known compared with the ones when it is drawn and a non-informative prior is used (nMSE of $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$). The scenario is the same as in Section IV-A1. $\bar{M} = M = 16$, $N_r = 1000$, $N_{bi} = 200$.

and the results are presented in Fig. 8. It can be seen that the results do not dramatically change with an error of ± 5 dB on the thermal noise approximate. Nonetheless, underestimating σ^2 is more advantageous.

On the other hand, we explained in Section IV-A that a non-informative prior is chosen for σ^2. This choice aims at facilitating the σ^2-hyperparameters setup, and has proven to give accurate results. For the sake of transparency, we represent in Fig. 9 the nMSE of the reconstructed target scene $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$ estimated by the proposed algorithm when σ^2 is known, which corresponds to a clairvoyant estimation wrt σ^2. The scenario is also the same as in Section IV-A1. It can be seen that using a non-informative prior in the proposed algorithm does not dramatically decrease the performance when compared to the σ^2-clairvoyant case.

4) Influence of the level of occupation: To finish, we study the influence of sparsity level, and thus test the proposed algorithm on a scenario with a strong target and an increasing number of weak targets. The post-processing amplitude of the strong target is uniformly distributed within [30 65] dB and that of the weak targets within [10 14] dB. The MSE of $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$ is calculated and displayed in Fig.10. The results are consistent with the well-known assumption that lower sparsity degrades the performance of SSR algorithms. However, it can be seen that the proposed sparse-promoting prior gives better performance than the hierarchical Student's *t*-prior with an atom at 0 from [13].

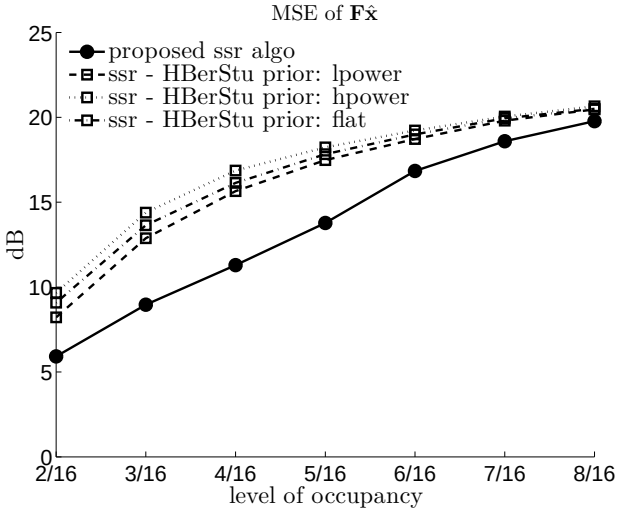


Fig. 10. MSE of $\boldsymbol{F}\hat{\boldsymbol{x}}_{\text{class}}$ evaluated after 200 Monte-Carlo simulations on a scenario with one strong target and an increasing number of weak targets (from 1 to 7). $M = \bar{M} = 16$, $N_r = 1000$, $N_{bi} = 200$.

B. Semiexperimental data

To finish, the proposed algorithm is evaluated on semiexperimental data recorded in November 2014 using the PARSAX radar [23] installed at TU Delft, The Netherlands. The semiexperimental data were built adding synthetic two-dimensional targets to a deramped thermal noise signal (the target signature is a two-dimensional cisoid with given range and Doppler frequency). Thus, the proposed algorithm is tested on realistic noise while having cooperative targets. The target scene is represented in Fig. 11(a); the strong targets present in the scene have a post-processing SNR of 60 dB, the weak ones of 12 dB and the medium one of 30 dB. The semiexperimental data is normalized beforehand, so that $\sigma^2 \approx 1$. In order to process this type of target scene, a two-dimensional Fourier matrix is used as a sparsifying dictionary $\boldsymbol{F}$.

Fig. 11(b) shows that the proposed algorithm accurately estimates the target scene. In particular, the weak targets are accurately estimated whereas the algorithm based on the hierarchical Student's t-prior with an atom at zero from [13] cannot see them if they are competing with a strong one (Fig. 11(d)). It is also interesting to note that both algorithms estimate zero-velocity components that most probably correspond to offsets of the coders. For the sake of comparison, the target scene estimated by the CLEAN algorithm is represented in Fig. 11(c). Following the approach described in [24], the stopping criterion adopted is reached when the standard deviation of the signal after removing the targets is lower than the noise standard deviation $\sigma \approx 1$. It can be seen that the CLEAN algorithm does estimate the weak targets at 12 dB, but more importantly it estimates several false targets.

V. Conclusion

In this paper, a new sparse-promoting prior and the corresponding MCMC estimation algorithm have been presented, and were proved to give accurate estimation of a radar scene consisting of weak and strong targets. This new prior enforces sparsity and decorrelates it from the target power level via its mixed-type structure. The novelty resides in the subdivision of the prior's support into several classes which corresponds to target power. In addition, an automatic setup of the class boundaries is proposed for high dynamic range scenarii. The proposed algorithm was tested on both synthetic and semiexperimental data, and showed that it outperforms state-of-the-art algorithms. More precisely, it is able to recover weak targets the estimation of which can be disrupted by strong ones. Hence, it could lead to a more recurrent detection of these weak targets. Nonetheless, the proposed algorithm lacks computational efficiency, especially when compared to the CLEAN algorithm.

In the future, the problem of finding a bound for the MSE of the reconstructed target scene within the scope of SSR could be investigated. Furthermore, the new sparse-promoting prior could be used with a non-optimal estimation algorithm in order to decrease the calculation load, for example a variational Bayes approximation.

Acknowledgement

The authors would like to thank O. Krasnov at TU Delft for kindly providing the PARSAX experimental data.

Appendix A
Implementation

The aim of this section is to provide the reader with details about the implementation of the algorithm. More precisely, an equivalent formulation of (9) and (10) consists in using labels $\alpha_{\bar{m}} = [\boldsymbol{\alpha}]_{\bar{m}}$ s.t. $\alpha_{\bar{m}} \in \{0, 1, \ldots, C\}$ that indicate in which class lands the $\bar{m}$th element of $\boldsymbol{\rho}$

$$
\begin{aligned}
&\Pr\left[\alpha_{\bar{m}} = c\right] = w_c, \quad c = 0, \ldots, C \\
&\begin{cases} \pi(\rho_{\bar{m}}|\alpha_{\bar{m}} = 0) & = \delta(\rho_{\bar{m}}) \\ \pi(\rho_{\bar{m}}|\alpha_{\bar{m}} = c) & = k_c \mathbb{I}_{[\rho_c^-, \rho_c^+)}(\rho_{\bar{m}}), \quad c = 1, \ldots, C \end{cases} \\
&\begin{cases} \pi(\phi_{\bar{m}}|\alpha_{\bar{m}} = 0) & = \delta(\phi_{\bar{m}}) \\ \pi(\phi_{\bar{m}}|\alpha_{\bar{m}} = c) & = \frac{1}{2\pi} \mathbb{I}_{[0,2\pi]}(\phi_{\bar{m}}), \quad c = 1, \ldots, C. \end{cases}
\end{aligned}
$$

In order to facilitate the implementation of the algorithm, (18) and (20) are reformulated using the label vector $\boldsymbol{\alpha}$ as

$$
\begin{aligned}
&\Pr\left[\alpha_{\bar{m}} = c\right] = w_{\bar{m},c}, \quad c = 0, \ldots, C \\
&\begin{cases} f(\rho_{\bar{m}}|\alpha_{\bar{m}} = 0) & = \delta(\rho_{\bar{m}}) \\ f(\rho_{\bar{m}}|\alpha_{\bar{m}} = c) & = \mathcal{N}_{[\rho_c^-, \rho_c^+)}\left(\mu_{\bar{m}}, \eta_{\bar{m}}^2\right), \quad c = 1, \ldots, C. \end{cases} \\
&\begin{cases} f(\phi_{\bar{m}}|\alpha_{\bar{m}} = 0) & = \delta(\phi_{\bar{m}}) \\ f(\phi_{\bar{m}}|\alpha_{\bar{m}} = c) & = \mathcal{VM}(\kappa_{\bar{m}}, \psi_{\bar{m}}). \end{cases}
\end{aligned}
$$

Thus, $\alpha_{\bar{m}}$ is sampled following a discrete distribution within $\{0, ..., C\}$ with corresponding probabilities $\{w_{\bar{m},0}, ..., w_{\bar{m},C}\}$; it is denoted as $\alpha_{\bar{m}} \sim \text{discrete}\left(\{0, ..., C\}, \boldsymbol{w}_{\bar{m}}\right)$. Then, depending on the value of $\alpha_{\bar{m}}$, the method proposed in [25] can be used to draw samples of $\rho_{\bar{m}}$ that follow a truncated Gaussian distribution; the samples $\phi_{\bar{m}}$ follow a von Mises-Fisher distribution and can be drawn following the method described in [17].

References

[1] J. Hogbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astronomy and Astrophysics Supplement*, vol. 15, pp. 417–426, 1974.

[2] J. Tsao and B. D. Steinberg, "Reduction of sidelobe and speckle artifacts in microwave imaging: the CLEAN technique." *IEEE Transactions on Antennas and Propagation*, vol. 36, no. 4, pp. 543–557, 1988.

[3] S. G.Mallat and Z. Zhang, "Matching Pursuits with time-frequence dictionaries," *IEEE transactions on signal processing*, vol. 41, no. 12, pp. 3397–3415, 1993.

[4] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition," *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*, pp. 40–44, 1993.

[5] G. Davis, "Adaptive greedy approximations," *Constructive Approximation*, vol. 13, pp. 57–98, 1997.

[6] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

[7] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 267–288, 1996.

[8] W. Melvin and J. Scheer, Eds., *Principles of Modern Radar: Advanced Techniques*. Edison, NJ: Institution of Engineering and Technology, 2012.

[9] S. D. Babacan, S. Nakajima, and M. N. Do, "Bayesian group-sparse modeling and variational inference," *IEEE Transactions on Signal Processing*, vol. 62, no. 11, pp. 2906–2921, 2014.

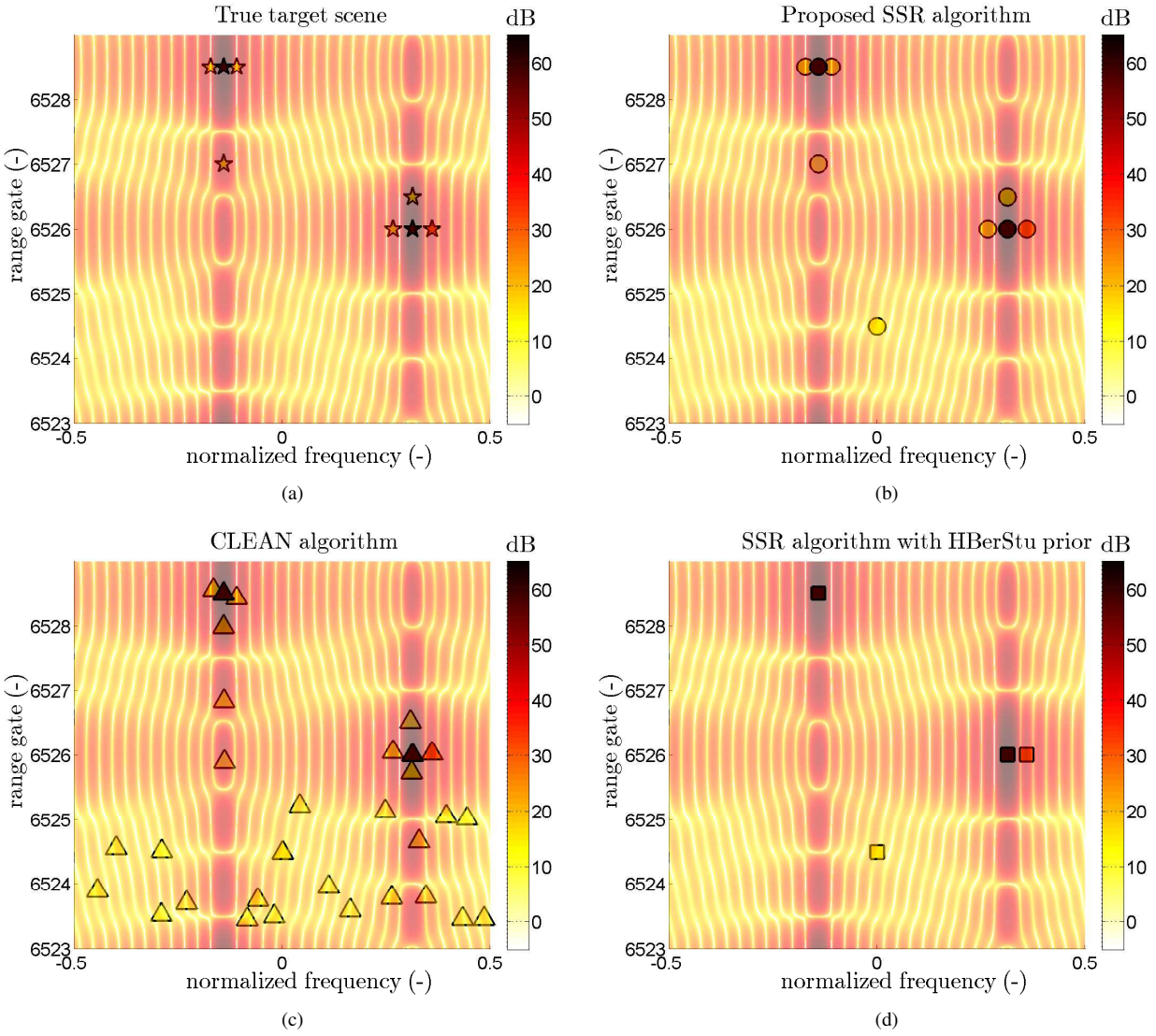


Fig. 11. Evaluation on semiexperimental data: 11(a) true target scene (stars) 11(b) target scene estimated by the proposed algorithm (circles) 11(d) target scene estimated by the algorithm without power classes of [13] (squares). For these two algorithms, $M = 32$, $K = 6$, $\bar{M} = 2M$, $\bar{K} = 2K$, $N_r = 1000$, $N_{bi} = 200$. 11(c) target scene estimated by the CLEAN algorithm (triangles). $M_{\text{fft}} = 1024$, $K_{\text{fft}} = 512$. The background corresponds to the coherent integration of the target scene.

[10] M. A. T. Figueiredo, "Adaptative sparseness for supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1150–1159, 2003.

[11] S. D. Babacan, R. Moline, and A. K. Katsaggelos, "Bayesian Compressive Sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, 2010.

[12] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211 – 244, 2001.

[13] S. Bidon, J.-Y. Tourneret, L. Savy, and F. Le Chevalier, "Bayesian sparse estimation of migrating targets for wideband radar," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 2, pp. 871–886, Apr. 2014.

[14] T. J. Mitchell and J. J. Beauchamp, "Bayesian Variable Selection in Linear Regression," *Journal of the American Statistical Association*, vol. 83, no. 404, pp. 1023–1032, 1988.

[15] S. Kotz, N. Balakrishnan, and N. L. Johnson, *Continuous Multivariate Distributions*, 2nd ed. Wiley Series in Probability and Statistics, 2000, vol. 1.

[16] C. Robert and G. Casella, *Monte Carlo Statistical Methods*. Springer, 2004.

[17] D. J. Best and N. I. Fisher, "Efficient simulation of the von Mises distribution," *Journal of the Royal Statistical Society*, no. 2, pp. 152–157.

[18] L. Devroye, *Non-Uniform Random Variate Generation*. Springer-Verlag, New York, 1986.

[19] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Process.*, vol. 44, no. 6, pp. 1469–1484, Jun. 1996.

[20] M. Lasserre, S. Bidon, O. Besson, and F. Le Chevalier, "Bayesian sparse Fourier representation of off-grid targets with application to experimental radar data," *Signal Processing*, vol. 111, pp. 261–273, 2015.

[21] Y. Chi, L. Scharf, A. Pezeshki, and A. Calderbank, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, no. 5, pp. 2182–2195, May 2011.

[22] M. Richards, *Fundamentals of Radar Signal Processing*. McGraw-Hill Electronic Engineering, 2005.

[23] O. A. Krasnov, G. P. Babur, Z. Wang, L. P. Ligthart, and F. van der Zwan, "Basics and first experiments demonstrating isolation improvements in the agile polarimetric FM-CW radar – PARSAX," *International Journal of Microwave and Wireless Technologies*, vol. 2, pp. 419–428, 8 2010.

[24] H. Deng, "Effective CLEAN algorithms for performance-enhanced detection of binary coding radar signals," *IEEE Trans. Signal Process.*, vol. 52, no. 1, pp. 72–78, January 2004.

[25] N. Chopin, "Fast simulation of truncated Gaussian distributions," *Statistics and Computing*, vol. 21, no. 2, pp. 275–288, 2011.