Bayesian Sparse Fourier Representation of Off-Grid Targets with Application to Experimental Radar Data[☆,☆☆]

Marie Lasserre[a,*], Stéphanie Bidon[a], Olivier Besson[a], François Le Chevalier[b]

[a] *University of Toulouse, ISAE/DEOS, 10 Avenue Edouard Belin, 31055 Toulouse, France*
[b] *MS[3], TU Delft EEMCS Building 36, Mekelweg 4, 2628 CD Delft, The Netherlands*

Abstract

The problem considered is the estimation of a finite number of cisoids embedded in white noise, using a sparse signal representation (SSR) approach, a problem which is relevant in many radar applications. Many SSR algorithms have been developed in order to solve this problem, but they usually are sensitive to grid mismatch. In this paper, two Bayesian algorithms are presented, which are robust towards grid mismatch: a first method uses a Fourier dictionary directly parametrized by the grid mismatch while the second one employs a first-order Taylor approximation to relate linearly the grid mismatch and the sparse vector. The main strength of these algorithms lies in the use of a mixed-type distribution which decorrelates sparsity level and target power. Besides, both methods are implemented through a Monte-Carlo Markov chain algorithm. They are successfully evaluated on synthetic and experimental radar data, and compared to a benchmark algorithm.

Keywords:
sparse representation, grid mismatch, Bayesian inference, Monte-Carlo Markov Chain

1. Introduction

Usually in radar applications, the received signal consists of the signal of interest modeled by a sum of cisoids embedded in additive noise

$$\boldsymbol{y} = \sum_{n=1}^{N} \alpha_n \boldsymbol{a}_n + \boldsymbol{n} \quad \text{with} \quad [\boldsymbol{a}_n]_m = \exp\{j2\pi f_n m\} \tag{1}$$

where $\boldsymbol{y} \in \mathbb{C}^M$ is the observation vector and M is the size of the observation space; $\alpha_n, \boldsymbol{a}_n$ are respectively the complex amplitude and the steering vector with frequency f_n of the nth target signal and $\boldsymbol{n}$ is the noise vector. Several methods have been developed to estimate the target scene (α_n, f_n) using the observation vector $\boldsymbol{y}$; they can be divided into two classes. When no specific model about the noise covariance matrix is assumed, the technique is said to be non-parametric (e.g., Fourier transform, Capon algorithm [1], APES [2]). Otherwise, the technique is called parametric (e.g., subspace methods: MUSIC [3] and ESPRIT [4]). Recently, a new model for the estimation of the target scene has emerged, called sparse signal reconstruction (SSR) [5, chap.5]. In radar applications, the signal of interest is usually sparse, e.g., when a small number of targets are present in the target scene. A sparse representation is thus a natural choice and is deemed to "permit efficient fundamental signal processing" [6]. In what follows we adopt an SSR approach where the signal is described as a linear combination of a finite number of atoms from a dictionary. Using the Fourier basis as a sparsifying dictionary, the problem (1) can be recasted as

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} + \boldsymbol{n} \tag{2}$$

with

- $\boldsymbol{F} \in \mathbb{C}^{M \times \bar{M}}$ the Fourier dictionary of size $M \times \bar{M}$ where usually $\bar{M} \geq M$;
- $\boldsymbol{x} \in \mathbb{C}^{\bar{M}}$ the sparse vector having ideally exactly N nonzero components.

Several methods have been developed to estimate $\boldsymbol{x}$ from the noisy measurement $\boldsymbol{y}$. Most of them are formulated from a deterministic point of view and aim at finding the solution $\hat{\boldsymbol{x}}$ as

$$\hat{\boldsymbol{x}} = \arg\min \| \boldsymbol{x} \|_1 \quad \text{s.t.} \quad \| \boldsymbol{F}\boldsymbol{x} - \boldsymbol{y} \|_2 \leq \tau. \tag{3}$$

where $\boldsymbol{F}$ is a sparsifying dictionary and τ is the energy bound on the additive noise $\boldsymbol{n}$. The problem (3) can be

[☆] This paper has been presented in part at the IEEE Radar Conference, Lille, France in October 13-17, 2014.
[☆☆] The work of M. Lasserre, S. Bidon and O. Besson is supported by the Délégation Générale de l'Armement under grant 2012.60.0012.00.470.75.01.
[*] Corresponding author
Email addresses: marie.lasserre@isae.fr (Marie Lasserre), stephanie.bidon@isae.fr (Stéphanie Bidon), olivier.besson@isae.fr (Olivier Besson), F.LeChevalier@tudelft.nl (François Le Chevalier)

reformulated and solved with the Basis Pursuit [7] or Lasso [8] algorithms as well as thresholding-based techniques, e.g., [9].The so-called "greedy algorithms", e.g., Orthogonal Matching Pursuit (OMP) [10] or Subspace Pursuit [11], reduce the calculation load by iteratively identifying the support of vector $\boldsymbol{x}$.

Despite the prominence of techniques that solve (3), they all endure the so-called "grid mismatch problem" that arises when the grid frequencies do not match the true frequencies of the signal. This problem was first described in [12] and quantified several times in [13, 14, 15, 16]. They all conclude that it is essential to take into account the grid mismatch problem, at risk of dramatically deteriorate the sparse recovery performance. The most natural way to deal with grid mismatch is to refine the grid [14, 17, 12]. However, refining the grid increases the coherence in the sparsifying dictionary, which can damage the overall estimation performance. Thus, most of the robust SSR techniques aim at estimating jointly the grid mismatch and the sparse vector. A parametric model can be adopted where the dictionary directly depends on grid mismatch. For example, in [18], a Bayesian approach was adopted and a variational Expectation–Maximization (EM) algorithm implemented. Nonetheless, in general the dictionary is not linear with respect to (wrt) the grid mismatch, so this model induces quite heavy calculations [19]. Thus, most of the techniques choose alternatively to add a perturbation matrix to the dictionary [20], which is usually obtained from a first order Taylor approximation [19, 21, 22] so that the dictionary becomes linear wrt the grid mismatch. In this paper, such a model will be called a "first-derivative-based" model. Several estimation methods can be used with this model. In [20], the method used is a total least-squares optimization under special constraint for sparsity and error in the dictionary. In [21], a Bayesian framework is adopted and an EM algorithm is implemented. The Bayesian inference of [18] was taken up again in [19], as well as a variational Bayesian (VB) algorithm, but the calculations are simplified by the Taylor approximation. Finally, in [22] the classical algorithms OMP [10] and Lasso [8] are augmented to jointly estimate the grid mismatch and the sparse vector.

In this paper, we propose a robust hierarchical Bayesian scheme, which is an extension of a non-robust SSR method [23]. In particular, the analysis matrix is reduced to a Fourier dictionary and a step is added in the algorithm in order to sample the grid mismatch. The advantage of the proposed formulation over that of [18, 19, 21] is that it enforces more sparsity via the use of a hierarchical Bernoulli-complex Gaussian prior on $\boldsymbol{x}$. A mixed-type prior enables to decorrelate the sparsity level of the scene and the target power. Furthermore, it permits to estimate the grid mismatch only when a target is present, without previously setting a threshold or number of targets. No attempt has been made yet at robustifying a model using this prior. Besides, the sparse vector and grid mismatch are estimated for the first time using a Monte-Carlo Markov Chain algorithm, which provides an optimal solution. The motivation is above all to reach high estimation performance, and not to optimize the calculation load.

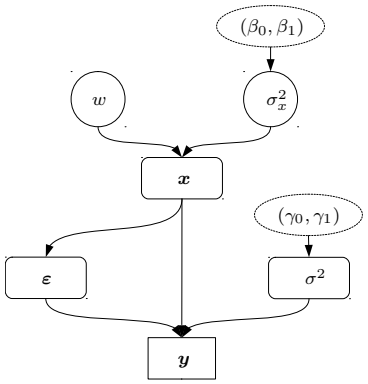


Figure 1: Directed acyclic graph associated with the hierarchical Bayesian parametric model.

The remaining of the paper is organized as follows. In section 2 a parametric model is presented together with the hierarchical Bayesian scheme used, and in particular the sampling of the grid mismatch. Then, in section 3 a first-derivative-based model is presented, which is widely used in the literature, as well as its associated hierarchical Bayesian scheme. This model simplifies the sampling of the grid mismatch. These two models are then compared to a state-of-the-art algorithm in section 4 on synthetic data and in section 5 on experimental data retrieved from the PARSAX radar system [24]. Section 6 concludes with the performance of the proposed algorithms and some future work considered.

2. Parametric signal model

In this section, we describe the so-called "parametric model". The hierarchical Bayesian model considered is represented graphically in Fig. 1.

2.1. Bayesian Model

2.1.1. Observation model

Modeling of grid mismatch. We consider the observation model $\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} + \boldsymbol{n}$ described in (2) with $\bar{M}$ fixed. When each true frequency of the signal matches a frequency in the analysis grid, an appropriate choice for the Fourier dictionary $\boldsymbol{F}$ is

$$[\boldsymbol{f}_{\bar{m}}]_m = 1/\sqrt{M} \exp\{j2\pi m(\bar{m}/\bar{M})\}$$

where $\boldsymbol{f}_{\bar{m}}$ is the $\bar{m}$th column of $\boldsymbol{F}$. In practice, the target frequencies do not match the ones in the analysis grid. As a consequence, as in [25] we propose to model the possible grid error by introducing a perturbation vector on the frequency axis denoted as $\boldsymbol{\varepsilon} = \begin{bmatrix}\varepsilon_0 & \dots & \varepsilon_{\bar{m}} & \dots & \varepsilon_{\bar{M}-1}\end{bmatrix}^T$ so that the Fourier dictionary is directly parameterized by $\boldsymbol{\varepsilon}$ as follows

$$\boldsymbol{F} \triangleq \boldsymbol{F}(\boldsymbol{\varepsilon}) = \begin{bmatrix}\boldsymbol{f}_0(\varepsilon_0) & \dots & \boldsymbol{f}_{\bar{m}}(\varepsilon_{\bar{m}}) & \dots & \boldsymbol{f}_{\bar{M}-1}(\varepsilon_{\bar{M}-1})\end{bmatrix}$$

where the $\bar{m}$th column of $\boldsymbol{F}(\boldsymbol{\varepsilon})$ is now expressed as

$$[\boldsymbol{f}_{\bar{m}}(\varepsilon_{\bar{m}})]_m = 1/\sqrt{M}\exp\{j2\pi m(\bar{m}+\varepsilon_{\bar{m}})/\bar{M}\}.$$

In order to avoid overlapping between the frequency bins of $\boldsymbol{F}$, the grid errors are assumed bounded such that for $\bar{m}=0,\dots,\bar{M}-1$, $\varepsilon_{\bar{m}}\in[-0.5,0.5)$. The observation model (2) can then be written as the "parametric model"

$$\boxed{\boldsymbol{y}=\boldsymbol{F}(\boldsymbol{\varepsilon})\boldsymbol{x}+\boldsymbol{n}} \tag{4}$$

where the Fourier matrix $\boldsymbol{F}$ is directly parameterized by the vector of frequency error $\boldsymbol{\varepsilon}$.

Likelihood. An additive white noise background is considered, and $\boldsymbol{n}$ is assumed to be centered Gaussian with power σ^2, which is denoted as

$$\boldsymbol{n}|\sigma^2 \sim \mathcal{CN}_M\left(\boldsymbol{0},\sigma^2\boldsymbol{I}\right) \tag{5}$$

where $\boldsymbol{I}$ is the identity matrix. Note that the clutter is supposedly represented by discretes with zero velocity. This representation may be sufficient in several radar applications and is relevant for the experimental data used in Section 5. The likelihood function is thus given by

$$f\left(\boldsymbol{y}|\boldsymbol{\varepsilon},\boldsymbol{x},\sigma^2\right)=\frac{1}{\pi^M\sigma^{2M}}\exp\left\{-\frac{\|\boldsymbol{y}-\boldsymbol{F}(\boldsymbol{\varepsilon})\boldsymbol{x}\|_2^2}{\sigma^2}\right\}. \tag{6}$$

A Bayesian framework is established in order to estimate the target scene $\boldsymbol{x},\boldsymbol{\varepsilon}$. Each unknown parameter is modeled by a random variable with a given prior probability density function (pdf). The choice of each prior density is aimed at facilitating the calculation of the estimation (mathematical tractability), yet preserving physical sense to the hierarchical model.

2.1.2. Prior pdfs of the parameters

Target amplitude vector. Ideally in SSR the vector $\boldsymbol{x}$ introduced in (2) has exactly N nonzero elements. As in [23] a Bernoulli-complex Gaussian prior is chosen to actually enforce sparsity in $\boldsymbol{x}$. The elements $x_{\bar{m}}\triangleq[\boldsymbol{x}]_{\bar{m}}$ of the amplitude vector are assumed independent and identically distributed (iid) according to the following mixed type pdf

$$f(x_{\bar{m}}|w,\sigma_x^2)=(1-w)\delta(|x_{\bar{m}}|)+w\frac{1}{\pi\sigma_x^2}\exp\left\{-\frac{|x_{\bar{m}}|^2}{\sigma_x^2}\right\}. \tag{7}$$

The prior (7), denoted as $x_{\bar{m}}|w,\sigma_x^2\sim\mathcal{BerCN}\left(w,0,\sigma_x^2\right)$, amounts to considering that a target with power σ_x^2 is present at the $\bar{m}$th frequency bin with probability w.

Grid errors. We propose to define the prior pdf of the grid error on the $\bar{m}$th frequency grid $\varepsilon_{\bar{m}}$ conditionally to the magnitude of $x_{\bar{m}}$: no grid error will be estimated if no target signal is present at the corresponding frequency bin. This notion is also used in [21, 18, 19] albeit either a threshold on the target amplitude has to be fixed, or the number of targets N known. Since we use a mixed-type prior on $\boldsymbol{x}$, there is no need of either setting up a target amplitude threshold or knowing the number of targets. Here, we consider that the $\varepsilon_{\bar{m}}|x_{\bar{m}}$ are iid with pdf

$$f(\varepsilon_{\bar{m}}|x_{\bar{m}}=0)=\delta(\varepsilon_{\bar{m}}) \tag{8a}$$

$$f(\varepsilon_{\bar{m}}|x_{\bar{m}}\neq 0)=\mathbb{I}_{[-0.5,0.5]}(\varepsilon_{\bar{m}}) \tag{8b}$$

where $\mathbb{I}_A(.)$ is the indicator function of the set A.

Noise power. An inverse-gamma prior is chosen for the white noise power σ^2 mostly since this distribution is conjugate to the likelihood (6). The prior pdf of σ^2 can therefore be expressed as

$$f(\sigma^2|\gamma_0,\gamma_1)\propto\frac{e^{-\gamma_1/\sigma^2}}{(\sigma^2)^{\gamma_0+1}}\mathbb{I}_{[0,+\infty)}(\sigma^2) \tag{9}$$

where γ_0,γ_1 are respectively the shape and scale parameters. The distribution (9) is denoted as $\sigma^2\,|\gamma_0,\gamma_1\sim\mathcal{IG}\left(\gamma_0,\gamma_1\right)$. The shape and scale parameters (γ_0,γ_1) allow to select a very informative, or on the contrary flat, prior. Nonetheless, they can be chosen in such a way that they keep hold of their physical sense. In radar applications, the thermal noise power is usually rather accurately known so that only a moderately informative prior is required in (9).

2.1.3. Prior pdfs of the hyperparameters

Since usually the probability w and target signal power σ_x^2 are both unknown, another level needs to be added to the hierarchical model.

Target signal power. Similarly to σ^2, an inverse-gamma prior is chosen for the target signal power σ_x^2 and is denoted as $\sigma_x^2|\beta_0,\beta_1\sim\mathcal{IG}\left(\beta_0,\beta_1\right)$. This time, the shape and scale parameters β_0,β_1 must be chosen carefully and in compliance with some prior knowledge about the target scene. Besides, the targets present in the signal may have different amplitudes from one to another, hence a not so informative prior might be necessary.

Level of occupancy. If no information is available about the sparsity level of the target scene, a convenient prior is a uniform pdf over the interval $[0,1]$, i.e., $w\sim\mathcal{U}_{[0,1]}$.

2.2. Bayesian estimation

Herein we propose an estimation scheme of the target scene $\boldsymbol{x},\boldsymbol{\varepsilon}$ based on the Bayesian hierarchical model described in Section 2.1. More precisely, our objective is to obtain the minimum mean square error (MMSE) estimators of $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$

$$\hat{\boldsymbol{x}}_{\text{MMSE}}=\int\boldsymbol{x}f(\boldsymbol{x}|\boldsymbol{y})d\boldsymbol{x}, \tag{10a}$$

$$\hat{\boldsymbol{\varepsilon}}_{\text{MMSE}}=\int\boldsymbol{\varepsilon}f(\boldsymbol{\varepsilon}|\boldsymbol{y})d\boldsymbol{\varepsilon}. \tag{10b}$$

In [23], the grid mismatch was not taken into account but the MMSE estimator of $\boldsymbol{x}$ was already intractable to derive analytically because of the complexity of the posterior distribution of $\boldsymbol{x}|\boldsymbol{y}$. When considering grid mismatch, the analytic calculation of the MMSE estimators (10a) and (10b) seems all the more demanding. As a consequence, a Monte-Carlo Markov Chain (MCMC) is implemented [26]. More specifically, an hybrid Gibbs sampler [26, chap.10] is used, which simulates iteratively samples $\sigma^{2^{(t)}}, \boldsymbol{\varepsilon}^{(t)}, \boldsymbol{x}^{(t)}, w^{(t)}, \sigma_x^{2^{(t)}}$ according to their conditional posterior distribution $f(\theta_i|\boldsymbol{y}, \boldsymbol{\theta}_{-i})$ where $\boldsymbol{\theta} = \left[\sigma^2, \boldsymbol{\varepsilon}^T, \boldsymbol{x}^T, w, \sigma_x^2\right]^T$ and $\boldsymbol{\theta}_{-i}$ is the vector $\boldsymbol{\theta}$ whose ith element has been removed. After a burn-in time N_{bi}, the samples are distributed according to their posterior distribution $f(\theta_i|\boldsymbol{y})$. When enough samples are acquired (N_r), the MMSE estimators can be built empirically as

$$\hat{\theta}_{i\text{MMSE}} = N_r^{-1} \sum_{t=1}^{N_r} {\theta_i}^{(t+N_{bi})}. \tag{11}$$

The conditional posterior distributions are obtained from the joint posterior pdf of $\sigma^2, \boldsymbol{\varepsilon}, \boldsymbol{x}, w, \sigma_x^2|\boldsymbol{y}$

$$\begin{aligned} f(\sigma^2, \boldsymbol{\varepsilon}, &\boldsymbol{x}, w, \sigma_x^2|\boldsymbol{y}) \propto \\ & f(\boldsymbol{y}|\boldsymbol{\varepsilon}, \boldsymbol{x}, \sigma^2) f(\boldsymbol{\varepsilon}|\boldsymbol{x}) f(\boldsymbol{x}|w, \sigma_x^2) f(w) f(\sigma_x^2) f(\sigma^2). \end{aligned} \tag{12}$$

In particular, both vectors $\boldsymbol{x}$ and $\boldsymbol{\varepsilon}$ are sampled element-wise with conditional posterior distributions of $\varepsilon_{\bar{m}}$ and $x_{\bar{m}}$ derived from their conditional joint posterior distribution

$$\begin{aligned} f(&\varepsilon_{\bar{m}}, x_{\bar{m}}|\boldsymbol{y}, \boldsymbol{\varepsilon}_{-\bar{m}}, \boldsymbol{x}_{-\bar{m}}, \sigma^2, w, \sigma_x^2) \\ &\propto \exp\left\{-\sigma^{-2}\left[|x_{\bar{m}}|^2 - x_{\bar{m}}^* \boldsymbol{f}_{\bar{m}}(\varepsilon_{\bar{m}})^H \boldsymbol{e}_{\bar{m}} - x_{\bar{m}} \boldsymbol{e}_{\bar{m}}^H \boldsymbol{f}_{\bar{m}}(\varepsilon_{\bar{m}})\right]\right\} \\ &\quad \times f(\varepsilon_{\bar{m}}|x_{\bar{m}}) f(x_{\bar{m}}|w, \sigma_x^2) \end{aligned} \tag{13}$$

In the preceding equation, we used (6) along with the fact that $\boldsymbol{y} - \boldsymbol{F}(\boldsymbol{\varepsilon})\boldsymbol{x} = \boldsymbol{e}_{\bar{m}} - \boldsymbol{f}_{\bar{m}} x_{\bar{m}}$ with $\boldsymbol{e}_{\bar{m}} = \boldsymbol{y} - \sum_{i \neq \bar{m}} \boldsymbol{f}_i(\varepsilon_i) x_i$.

2.2.1. Sampling of $\boldsymbol{x}$

Following [23], $\boldsymbol{x}$ is sampled element-wise. From inspection of (13), it is straightforward to see that the $\bar{m}$th element of $\boldsymbol{x}$ follows the distribution $\mathcal{B}er\mathcal{CN}\left(w_{\bar{m}}, \mu_{\bar{m}}, \eta_{\bar{m}}^2\right)$ with

$$\eta_{\bar{m}}^2 = \left(\frac{1}{\sigma^2} + \frac{1}{\sigma_x^2}\right)^{-1} \tag{14a}$$

$$\mu_{\bar{m}} = \frac{\eta_{\bar{m}}^2}{\sigma^2} \boldsymbol{f}_{\bar{m}}(\varepsilon_{\bar{m}})^H \boldsymbol{e}_{\bar{m}} \tag{14b}$$

$$w_{\bar{m}} = \frac{w \frac{\eta_{\bar{m}}^2}{\sigma_x^2} \exp\left\{\frac{|\mu_{\bar{m}}|^2}{\eta_{\bar{m}}^2}\right\}}{1 - w + w \frac{\eta_{\bar{m}}^2}{\sigma_x^2} \exp\left\{\frac{|\mu_{\bar{m}}|^2}{\eta_{\bar{m}}^2}\right\}}. \tag{14c}$$

2.2.2. Sampling of $\boldsymbol{\varepsilon}$

The parameter $\boldsymbol{\varepsilon}$ is sampled element-wise, in the same way as vector $\boldsymbol{x}$. Using (13), the conditional posterior distribution of $\varepsilon_{\bar{m}}$ is calculated

$$\begin{aligned} &f(\varepsilon_{\bar{m}}|\boldsymbol{y}, \boldsymbol{\varepsilon}_{-\bar{m}}, \boldsymbol{x}) \\ &\propto \exp\left\{-\sigma^{-2}\left[|x_{\bar{m}}|^2\right.\right. \\ &\quad \left.\left. - x_{\bar{m}}^* \boldsymbol{f}_{\bar{m}}(\varepsilon_{\bar{m}})^H \boldsymbol{e}_{\bar{m}} - x_{\bar{m}} \boldsymbol{e}_{\bar{m}}^H \boldsymbol{f}_{\bar{m}}(\varepsilon_{\bar{m}})\right]\right\} f(\varepsilon_{\bar{m}}|x_{\bar{m}}) \\ &\propto \exp\left\{2\sigma^{-2} \mathcal{R}e\left[x_{\bar{m}}^* \boldsymbol{f}_{\bar{m}}(\varepsilon_{\bar{m}})^H \boldsymbol{e}_{\bar{m}}\right]\right\} f(\varepsilon_{\bar{m}}|x_{\bar{m}}) \end{aligned} \tag{15}$$

$$\propto \exp\left\{\sum_{m=1}^{M-1} \kappa_m \cos\left(2\pi \frac{\varepsilon_{\bar{m}}}{\bar{M}} m - \phi_m\right)\right\} \times f(\varepsilon_{\bar{m}}|x_{\bar{m}}) \tag{16}$$

where $\kappa_m = \frac{2}{\sigma^2 \sqrt{\bar{M}}} \times |b_m|$ and $\phi_m = \angle b_m$. $\boldsymbol{b}$ has been defined as $\boldsymbol{b} = x_{\bar{m}}^* \boldsymbol{u}_{\bar{m}}^* \odot \boldsymbol{e}_{\bar{m}}$, where $\boldsymbol{u}_{\bar{m}} = \exp\{j2\pi m(\bar{m}/\bar{M})\}$.

Knowing that $f(\varepsilon_{\bar{m}}|x_{\bar{m}} \neq 0) = \mathbb{I}_{[-0.5,0.5]}(\varepsilon_{\bar{m}})$, we recognize from (16) a dilated and truncated generalized von Mises distribution [27]. Such a distribution can be troublesome to sample, so a Metropolis-Hastings (MH) algorithm is used [26]. This algorithm is based on a *proposal distribution* that should be easy to simulate from, and as close as possible to the *target distribution*. We represent in Fig.2 the conditional posterior distribution of $\varepsilon_{\bar{m}}$ for different values of mismatch and post-processing SNR defined as

$$M \times \mathcal{E}\{|\alpha_n|^2\}/\sigma^2.$$

We observed that when $\bar{M} \geq M$, in the case of low SNR, a flat proposal would be appropriate, while a Gaussian proposal would better fit the distribution in case of high SNR. Thus, in our MH algorithm we switch from a flat proposal to a Gaussian proposal (and vice-versa) depending on the estimated target power, namely $|x_{\bar{m}}^{(t)}|^2/\sigma^{2^{(t)}}$. The proposal scheme should be employed only during a burn-in period in order to preserve the convergence properties [26, chap.7], but in fact it does not damage the performance when used in the whole process.

2.2.3. Sampling of σ^2, w and σ_x^2

As in [23], the parameter σ^2 and hyperparameters w and σ_x^2 are sampled according to their conditional posterior distribution

$$\sigma^2|\boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\varepsilon} \sim \mathcal{IG}\left(\gamma_0 + M, \gamma_1 + \| \boldsymbol{y} - \boldsymbol{F}(\boldsymbol{\varepsilon})\boldsymbol{x} \|_2^2\right) \tag{17}$$

$$w|\boldsymbol{x} \sim \mathcal{B}e\left(1 + n_1, 1 + n_0\right) \tag{18}$$

$$\sigma_x^2|\boldsymbol{x} \sim \mathcal{IG}\left(\beta_0 + n_1, \beta_1 + \| \boldsymbol{x} \|_2^2\right) \tag{19}$$

where n_1 is the number of nonzero elements of $\boldsymbol{x}$ and $n_0 = \bar{M} - n_1$.

The hybrid-Gibbs sampler computed is summarized in Fig. 3 where the sampling of each parameter is detailed.

3. First-derivative-based signal model

3.1. Bayesian Model

The previous method relies on an exact modeling of the grid mismatch $\boldsymbol{\varepsilon}$, embedded in the exponential term.

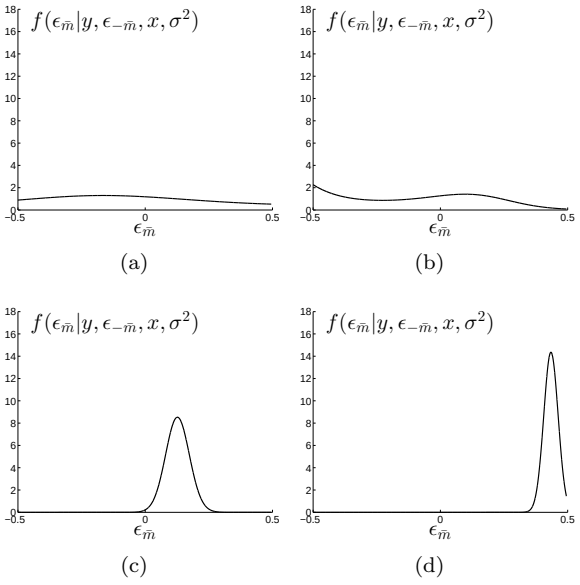


Figure 2: Shape of the conditional posterior distribution $f(\varepsilon_{\bar{m}}|\boldsymbol{y},\boldsymbol{\varepsilon}_{-\bar{m}},\boldsymbol{x},\sigma^2)$ for a single target at zero velocity with $\bar{M}=M$. (a) SNR=5dB, $\varepsilon_{\bar{m}}=.15$. (b) SNR=5dB, $\varepsilon_{\bar{m}}=.45$. (c) SNR=15dB, $\varepsilon_{\bar{m}}=.15$. (d) SNR=15dB, $\varepsilon_{\bar{m}}=.45$.

Require: $\boldsymbol{y}$, (γ_0,γ_1), (β_0,β_1)
Ensure: $\hat{\sigma}^2_{\text{MMSE}}$, $\hat{w}_{\text{MMSE}}$, $\hat{\sigma}^2_{x\text{MMSE}}$, $\hat{\boldsymbol{x}}_{\text{MMSE}}$, $\hat{\boldsymbol{\varepsilon}}_{\text{MMSE}}$
{Initialization}
$w^{(0)} \sim \mathcal{U}_{[0,1]}$
$\sigma_x^{2\,(0)} \sim \mathcal{IG}\left(\beta_0,\beta_1\right)$
$\boldsymbol{x}^{(0)} \sim \prod_{i=0}^{\bar{M}-1} \mathcal{B}er\mathcal{CN}\left(w^{(0)},0,\sigma_x^{2\,(0)}\right)$
$\boldsymbol{\varepsilon}^{(0)} \sim \mathcal{U}_{[-0.5,0.5]}$
{Iterations}
for $n=1$ to $N_{bi}+N_r$ **do**
  $\sigma^{2\,(n)}|\boldsymbol{y},\boldsymbol{x}^{(n-1)},\boldsymbol{\varepsilon}^{(n-1)} \sim$
    $\mathcal{IG}\left(\gamma_0+M,\gamma_1+\parallel \boldsymbol{y}-\boldsymbol{F}(\boldsymbol{\varepsilon}^{(n-1)})\boldsymbol{x}^{(n-1)} \parallel_2^2\right)$
  $w^{(n)}|\boldsymbol{x}^{(n-1)} \sim \mathcal{B}e\left(1+n_1,1+n_0\right)$
  $\sigma_x^{2\,(n)}|\boldsymbol{x}^{(n-1)} \sim \mathcal{IG}\left(\beta_0+n_1,\beta_1+\parallel \boldsymbol{x}^{(n-1)} \parallel_2^2\right)$
  for $\bar{m}=0$ to $\bar{M}-1$ **do**
    ${x_{\bar{m}}}^{(n)}|\boldsymbol{y},{\boldsymbol{x}_{-\bar{m}}}^{(n)},w^{(n)},\sigma_x^{2\,(n)},\sigma^{2\,(n)},\boldsymbol{\varepsilon}^{(n-1)} \sim$
      $\mathcal{B}er\mathcal{CN}\left(w_{\bar{m}},\mu_{\bar{m}},\eta_{\bar{m}}^2\right)$, as in (14)
    ${\varepsilon_{\bar{m}}}^{(n)}|\boldsymbol{y},{\boldsymbol{\varepsilon}_{-\bar{m}}}^{(n)},\boldsymbol{x}^{(n)},\sigma^{2\,(n)} \sim$
      $\text{dGvM}_{[-.5,+.5]}(\kappa_m,\phi_m)$
  end for
end for
{Estimators}
$\hat{\theta}_{\text{MMSE}} = \frac{1}{N_r}\sum_{n=1}^{N_r}\theta^{(n+N_{bi})}$

Figure 3: Hybrid-Gibbs sampler used with the parametric model. The acronym "$\text{dGvM}_{[-.5,+.5]}$" refers to a dilated truncated generalized von Mises distribution defined in (16). According to [26, p.268], this algorithm will converge whatever the initialization of the parameters.

However, this leads to a rather complicated posterior distribution (namely a truncated generalized von Mises distribution), which in turn induces additional complexity to the sampler. Therefore, we now turn to a possibly simpler model. Following an approach widely used in the literature [28, 19, 29, 22, 30], we also investigate a first-derivative-based signal model which will be later compared with the parametric model. Indeed, in the parametric model, the Fourier dictionary $\boldsymbol{F}$ is not linear wrt the grid mismatch vector $\boldsymbol{\varepsilon}$. A first-order Taylor approximation can be used to make $\boldsymbol{F}$ linear wrt $\boldsymbol{\varepsilon}$ and thus reduce the calculation load. The first-derivative-based model stems from a first-order Taylor approximation around $\boldsymbol{\varepsilon}=\boldsymbol{0}$

$$\begin{aligned}\boldsymbol{F}(\boldsymbol{\varepsilon}) &= \boldsymbol{F}(\boldsymbol{0}) + \left.\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{\varepsilon}}\right|_{\boldsymbol{\varepsilon}=\boldsymbol{0}} (\boldsymbol{\varepsilon}-\boldsymbol{0}) \\ &= \boldsymbol{F}(\boldsymbol{0}) + \boldsymbol{D}\boldsymbol{F}(\boldsymbol{0})\boldsymbol{D}_{\boldsymbol{\varepsilon}} \end{aligned} \tag{20}$$

with $\boldsymbol{D} = \text{diag}\left(j\frac{2\pi}{\bar{M}}\begin{bmatrix}0 & \dots & M-1\end{bmatrix}\right)$ and $\boldsymbol{D}_{\boldsymbol{\varepsilon}} = \text{diag}(\boldsymbol{\varepsilon})$.

The first-derivative-based observation model is then written as

$$\boxed{\boldsymbol{y} = (\boldsymbol{F}(\boldsymbol{0}) + \boldsymbol{D}\boldsymbol{F}(\boldsymbol{0})\boldsymbol{D}_{\boldsymbol{\varepsilon}})\boldsymbol{x} + \boldsymbol{n}} \tag{21}$$

where there is now a linear relationship between the grid mismatch vector $\boldsymbol{\varepsilon}$ and the sparse vector $\boldsymbol{x}$.

Remark 1. The accuracy of the approximation (20) highly depends on the number of points in the analysis grid $\bar{M}$: the larger its value, the better the approximation. As a consequence, the relative quadratic error of the reconstructed target scene $\boldsymbol{F}(\boldsymbol{\varepsilon})\boldsymbol{x}$, namely $\parallel (\boldsymbol{F}(0)+\boldsymbol{D}\boldsymbol{F}(\boldsymbol{0})\boldsymbol{D}_{\boldsymbol{\varepsilon}})\boldsymbol{x} - \boldsymbol{F}(\boldsymbol{\varepsilon})\boldsymbol{x} \parallel^2 / \parallel \boldsymbol{F}(\boldsymbol{\varepsilon})\boldsymbol{x} \parallel^2$, significantly decreases when $\bar{M}$ increases, especially in the case of high mismatch. However, in that case $\boldsymbol{F}(\boldsymbol{0})$ becomes highly coherent.

Likelihood. Just as with the parametric model, the additive white noise background $\boldsymbol{n}$ is assumed to be centered Gaussian with power σ^2. The likelihood function is now given by

$$f(\boldsymbol{y}|\boldsymbol{\varepsilon},\boldsymbol{x},\sigma^2) = \frac{1}{\pi^M\sigma^{2M}}\exp\left\{-\frac{\|\boldsymbol{y}-(\boldsymbol{F}(\boldsymbol{0})+\boldsymbol{D}\boldsymbol{F}(\boldsymbol{0})\boldsymbol{D}_{\boldsymbol{\varepsilon}})\boldsymbol{x}\|_2^2}{\sigma^2}\right\}. \tag{22}$$

A Bayesian framework is set up to estimate the target scene $\boldsymbol{x},\boldsymbol{\varepsilon}$. The prior pdfs assigned to the different parameters are the same as with the parametric model, as well as the hierarchical Bayesian model (Fig.1).

3.2. Bayesian estimation

The target scene is estimated through the calculation of the MMSE estimators (10) via an MCMC algorithm. The conditional posterior distributions of the parameters and hyperparameters are calculated using the joint posterior distribution (12), but they differ to a certain extent from the ones with the parametric model.

3.2.1. Sampling of $\boldsymbol{x}$

The sparse vector $\boldsymbol{x}$ is sampled element-wise, and $x_{\bar{m}}$ follows the distribution $\mathcal{B}er\mathcal{CN}\left(w_{\bar{m}}, \mu_{\bar{m}}, \eta_{\bar{m}}^2\right)$ with now

$$\eta_{\bar{m}}^2 = \left(\frac{\parallel \boldsymbol{g}_{\bar{m}} \parallel_2^2}{\sigma^2} + \frac{1}{\sigma_x^2} \right)^{-1} \tag{23a}$$

$$\mu_{\bar{m}} = \frac{\eta_{\bar{m}}^2}{\sigma^2} \boldsymbol{g}_{\bar{m}}^H \boldsymbol{e}_{\bar{m}} \tag{23b}$$

$$w_{\bar{m}} = \frac{w \frac{\eta_{\bar{m}}^2}{\sigma_x^2} \exp\left\{ \frac{|\mu_{\bar{m}}|^2}{\eta_{\bar{m}}^2} \right\}}{(1-w) + w \frac{\eta_{\bar{m}}^2}{\sigma_x^2} \exp\left\{ \frac{|\mu_{\bar{m}}|^2}{\eta_{\bar{m}}^2} \right\}}. \tag{23c}$$

where $\boldsymbol{g}_{\bar{m}}$ is the $\bar{m}$th column of $\boldsymbol{F}(\boldsymbol{0}) + \boldsymbol{DF}(\boldsymbol{0})\boldsymbol{D}_{\boldsymbol{\varepsilon}}$ and $\boldsymbol{e}_{\bar{m}} = \boldsymbol{y} - \sum_{i \neq \bar{m}} \boldsymbol{g}_i x_i$.

3.2.2. Sampling of $\boldsymbol{\varepsilon}$

The grid mismatch vector $\boldsymbol{\varepsilon}$ is sampled element-wise too. The conditional posterior distribution of $\varepsilon_{\bar{m}}$ is now written as

$$f(\varepsilon_{\bar{m}} | \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\varepsilon}_{-\bar{m}}, \sigma^2) \propto \exp\left\{ -\frac{\left(\varepsilon_{\bar{m}} - \mu_{\varepsilon_{\bar{m}}}\right)^2}{2\sigma_{\varepsilon_{\bar{m}}}^2} \right\} f(\varepsilon_{\bar{m}} | x_{\bar{m}}) \tag{24}$$

with

$$\sigma_{\varepsilon_{\bar{m}}}^2 = \frac{\sigma^2}{2 \parallel \dot{\boldsymbol{f}}_{\bar{m}} \parallel_2^2 |x_{\bar{m}}|^2}$$

$$\mu_{\varepsilon_{\bar{m}}} = 2\mathcal{R}e\left\{ x_{\bar{m}}^* \dot{\boldsymbol{f}}_{\bar{m}}^H \tilde{\boldsymbol{e}}_{\bar{m}} \right\} \frac{\sigma_{\varepsilon_{\bar{m}}}^2}{\sigma^2}$$

where $\tilde{\boldsymbol{e}}_{\bar{m}} = \boldsymbol{y} - \boldsymbol{F}(\boldsymbol{0})\boldsymbol{x} - \sum_{i \neq \bar{m}} \left[\boldsymbol{DF}(\boldsymbol{0}) \mathrm{diag}(\boldsymbol{x})\right]_i \varepsilon_i$ and $\dot{\boldsymbol{f}}_{\bar{m}}$ is the $\bar{m}$th column of $\boldsymbol{DF}(\boldsymbol{0})$.

Using $f(\varepsilon_{\bar{m}} | x_{\bar{m}} \neq 0) = \mathbb{I}_{[-0.5,0.5]}(\varepsilon_{\bar{m}})$, we see from (24) that the conditional posterior distribution of $\varepsilon_{\bar{m}} | x_{\bar{m}} \neq 0$ is a Gaussian distribution with mean $\mu_{\varepsilon_{\bar{m}}}$ and variance $\sigma_{\varepsilon_{\bar{m}}}^2$, truncated on the interval $[-0.5, +0.5)$.

An MH algorithm is used to draw samples following this distribution, with a Gaussian proposal with mean $\mu_{\varepsilon_{\bar{m}}}$ and variance $\sigma_{\varepsilon_{\bar{m}}}^2$. The sampling of $\boldsymbol{\varepsilon}$ in the first-derivative-based model is thus easier than with the parametric model. Note that other techniques can be used to draw samples following a truncated Gaussian distribution, e.g., [31].

3.2.3. Sampling of σ^2, w and σ_x^2

To finish, the conditional posterior distributions of σ^2, w and σ_x^2 are calculated

$$f(\sigma^2 | \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{\varepsilon}) \sim \mathcal{IG}\left(\gamma_0 + M, \gamma_1 + \parallel \boldsymbol{y} - (\boldsymbol{F}(\boldsymbol{0}) + \boldsymbol{DF}(\boldsymbol{0})\boldsymbol{D}_{\boldsymbol{\varepsilon}})\boldsymbol{x} \parallel_2^2\right) \tag{25}$$

$$f(w | \boldsymbol{x}) \sim \mathcal{B}e\left(1 + n_1, 1 + n_0\right) \tag{26}$$

$$f(\sigma_x^2 | \boldsymbol{x}) \sim \mathcal{IG}\left(\beta_0 + n_1, \beta_1 + \parallel \boldsymbol{x} \parallel_2^2\right) \tag{27}$$

where n_1 is the number of nonzero elements of $\boldsymbol{x}$ and $n_0 = \bar{M} - n_1$.

The hybrid-Gibbs sampler modified in the case of the first-derivative-based model is summarized in Fig. 4.

Require: $\boldsymbol{y}$, $\boldsymbol{F}(\boldsymbol{0})$, $\boldsymbol{D}$, (γ_0, γ_1), (β_0, β_1)
Ensure: $\hat{\sigma}^2_{\mathrm{MMSE}}$, $\hat{w}_{\mathrm{MMSE}}$, $\hat{\sigma}^2_{x\mathrm{MMSE}}$, $\hat{\boldsymbol{x}}_{\mathrm{MMSE}}$, $\hat{\boldsymbol{\varepsilon}}_{\mathrm{MMSE}}$
{Initialization}
$w^{(0)} \sim \mathcal{U}_{[0,1]}$
$\sigma_x^{2\,(0)} \sim \mathcal{IG}\left(\beta_0, \beta_1\right)$
$\boldsymbol{x}^{(0)} \sim \prod_{i=0}^{\bar{M}-1} \mathcal{B}er\mathcal{CN}\left(w^{(0)}, 0, \sigma_x^{2\,(0)}\right)$
$\boldsymbol{\varepsilon}^{(0)} \sim \mathcal{U}_{[-0.5,0.5]}$
{Iterations}
for $n = 1$ to $N_{bi} + N_r$ **do**
  $\sigma^{2\,(n)} | \boldsymbol{y}, \boldsymbol{x}^{(n-1)}, \boldsymbol{\varepsilon}^{(n-1)} \sim \mathcal{IG}\left(\gamma_0 + M, \gamma_1 + \parallel \boldsymbol{y} - (\boldsymbol{F}(\boldsymbol{0}) + \boldsymbol{DF}(\boldsymbol{0})\boldsymbol{D}_{\boldsymbol{\varepsilon}})\boldsymbol{x}^{(n-1)} \parallel_2^2\right)$
  $w^{(n)} | \boldsymbol{x}^{(n-1)} \sim \mathcal{B}e\left(1 + n_1, 1 + n_0\right)$
  $\sigma_x^{2\,(n)} | \boldsymbol{x}^{(n-1)} \sim \mathcal{IG}\left(\beta_0 + n_1, \beta_1 + \parallel \boldsymbol{x}^{(n-1)} \parallel_2^2\right)$
  for $\bar{m} = 0$ to $\bar{M} - 1$ **do**
    $x_{\bar{m}}{}^{(n)} | \boldsymbol{y}, \boldsymbol{x}_{-\bar{m}}{}^{(n)}, w^{(n)}, \sigma_x^{2\,(n)}, \sigma^{2\,(n)}, \boldsymbol{\varepsilon}^{(n-1)} \sim \mathcal{B}er\mathcal{CN}\left(w_{\bar{m}}, \mu_{\bar{m}}, \eta_{\bar{m}}^2\right)$, as in (23)
    $\varepsilon_{\bar{m}}{}^{(n)} | \boldsymbol{y}, \boldsymbol{\varepsilon}_{-\bar{m}}{}^{(n)}, \boldsymbol{x}^{(n)}, \sigma^{2\,(n)} \sim \mathcal{N}_{[-.5,+.5]}\left(\mu_{\varepsilon_{\bar{m}}}, \sigma_{\varepsilon_{\bar{m}}}^2\right)$
  end for
end for
{Estimators}
$\hat{\theta}_{\mathrm{MMSE}} = \frac{1}{N_r} \sum_{n=1}^{N_r} \theta^{(n+N_{bi})}$

Figure 4: Hybrid-Gibbs sampler used with the first-derivative-based model. The vector $\boldsymbol{\varepsilon}$ is now sampled following a truncated Gaussian distribution $\mathcal{N}_{[-.5,+.5]}\left(\mu_{\varepsilon_{\bar{m}}}, \sigma_{\varepsilon_{\bar{m}}}^2\right)$.

4. Results on synthetic data

First, the parametric and first-derivative-based models are evaluated through several numerical simulations on synthetic data. The synthetic data are generated according to (1) and (5). As underlined before, the radar operator must set the hyperparameters (β_0, β_1) and (γ_0, γ_1) with special care. They are chosen to give the desired mean and variance of the prior distributions of the thermal noise power σ^2 and target power σ_x^2. Indeed for an inverse gamma distribution $g \sim \mathcal{IG}\left(\nu_0, \nu_1\right)$ the mean and variance are respectively

$$\mathrm{m}_g = \frac{\nu_1}{\nu_0 - 1}, \quad \nu_0 > 1$$

$$\mathrm{var}_g = \frac{\nu_1^2}{(\nu_0 - 1)^2(\nu_0 - 2)}, \quad \nu_0 > 2.$$

These equations are used to tune the hyperparameters in the simulations.

4.1. Example after one realization

We begin with a first (non-quantitative) illustration of the algorithms behavior, from a single run. A simple scenario is considered with three targets with mismatch $\varepsilon_0 \in \{0, 0.15, 0.45\}$ and a post-processing SNR of 20 dB. Fig.5 shows the estimated target scene $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\varepsilon}})$. The results obtained with the proposed algorithms (parametric and first-derivative-based models) are compared with

the ones obtained with a non-robust method (proposed algorithm in Fig.3 without the sampling of $\boldsymbol{\varepsilon}$, i.e., $\boldsymbol{\varepsilon} = \mathbf{0}$). The target scene estimated with the Capon algorithm [1] is also represented for the sake of comparison with a classical high-resolution spectral method. We can see that when $|\varepsilon_{\bar{m}}| \to .5$ the target spreads over the surrounding frequency bins with the non-robust analysis and first-derivative-based model. However, the estimation performed with the parametric model is a lot more accurate in terms of both amplitude and mismatch estimation. Besides, the parametric model does not result in as much sidelobes as with the Capon algorithm and the non-robust analysis. Thus, in this example the practical interest of the non-robust analysis and first-derivative-based model is questionable when compared with a classical high-resolution spectral analysis such as Capon, but the parametric model clearly is advantageous.

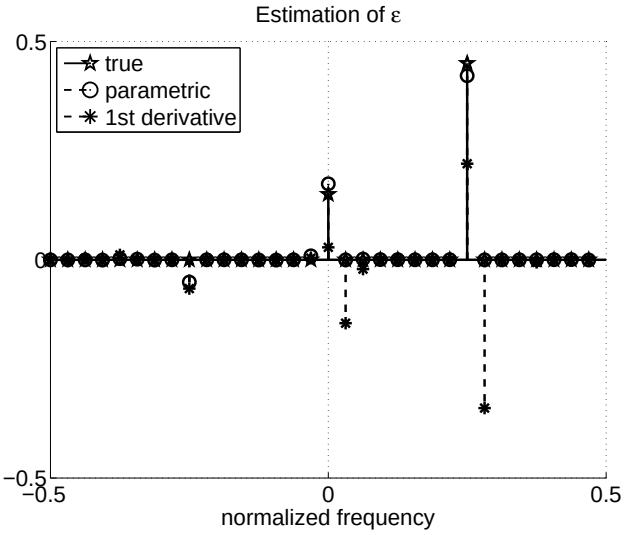
The histograms of the sampled parameters σ^2, w, σ_x^2, x_0 and ε_0 are represented in Fig.6 together with their prior distribution. They were obtained with the proposed algorithm and the parametric model. The estimation process results in a posterior distribution peaked around the estimated value, even when starting with a flat prior distribution (as with parameter w or ε_0). These estimated values seem quite reliable: $\sigma^2_{\text{MMSE}} = 1.02$ is close from its true value $\sigma^2 = 1$; $w_{\text{MMSE}} = 12.6\%$ while the true level of occupancy is $N/\bar{M} \approx 9.4\%$; $\sigma^2_{x\text{MMSE}} = 18.8$ dB while the average power of the scatterers in the scene is $1/N \sum_n \sigma^2 \text{SNR}_n = 20$ dB. If we focus on the target with zero velocity, the estimation is good too: $|x_{0\text{MMSE}}| = 9.53$ and $\varepsilon_{0\text{MMSE}} = 0.13$ whereas the true values are 10 (20 dB) and 0.15 respectively.

4.2. Results after several Monte-Carlo simulations

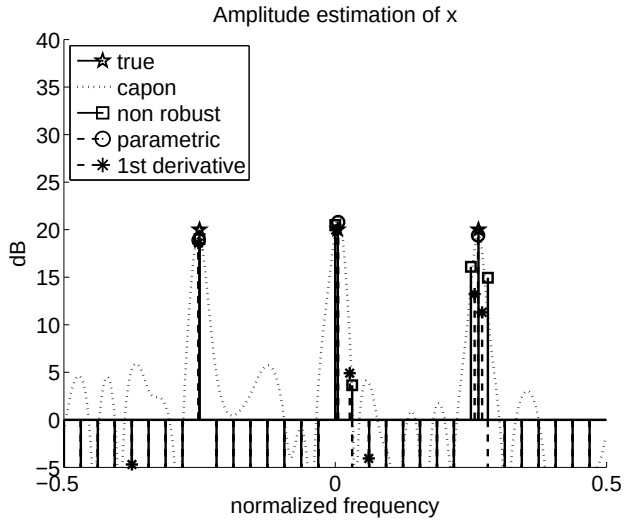
4.2.1. Discussion about an appropriate metric

In the case of SSR, finding an appropriate metric to evaluate the quality of the estimation is a delicate choice, e.g., [13]. Indeed, with a sparse representation there is an inherent ambiguity about the position of a target at the edge of a frequency bin. This ambiguity is represented in Fig.7: the target can be considered in the $\bar{m} - 1$th frequency bin with a mismatch of 0.5 ($\varepsilon_{\bar{m}-1} = +.5$) or in the $\bar{m}$th frequency bin with a mismatch of -0.5 ($\varepsilon_{\bar{m}} = -.5$).

If we consider a target such that $\varepsilon_{\bar{m}} = -.5$, the estimated target might be shifted on the previous frequency bin with a high positive mismatch (i.e., $\varepsilon_{\bar{m}-1} \to +.5$), or even split between the previous frequency bin and the true one but with acceptable mismatches (i.e., $\varepsilon_{\bar{m}-1} \to +.5$ and $\varepsilon_{\bar{m}} \to -.5$). In both cases, the estimation of the target frequency $(\bar{m} + \varepsilon_{\bar{m}})/\bar{M}$ is accurate and thus the reconstructed target scene can be correctly estimated. However, the MSEs of $\hat{\boldsymbol{x}}_{\text{MMSE}}$ and $\hat{\boldsymbol{\varepsilon}}_{\text{MMSE}}$ would be high since the target is not estimated in the correct bin. As a result, the MSE of $\boldsymbol{F}(\hat{\boldsymbol{\varepsilon}}_{\text{MMSE}})\hat{\boldsymbol{x}}_{\text{MMSE}}$ seems more representative of the estimation quality. More precisely, in what follows, the normalized MSE is calculated via $N_{mc} = 500$ Monte-Carlo runs as



(a)



(b)

Figure 5: SSR of synthetic target scene: $M = 32$, $N = 3$, $\sigma^2 = 1$, $\bar{M} = M$, $N_r = 1000$, $N_{bi} = 100$, $(\text{m}_{\sigma_x^2}, \sqrt{\text{var}_{\sigma_x^2}}) = M_{dB} + (0, 3.5)$ dB where $M_{dB} = 10\log_{10}(M)$, $(\text{m}_{\sigma^2}, \sqrt{\text{var}_{\sigma^2}}) = (0, 2.4)$ dB. (a) Grid mismatch $\boldsymbol{\varepsilon}$. (b) Target amplitude vector corrected by the estimated grid mismatch.

$$\text{nMSE}(\boldsymbol{F}(\hat{\boldsymbol{\varepsilon}}_{\text{MMSE}})\hat{\boldsymbol{x}}_{\text{MMSE}}) = \frac{1}{N_{mc}} \sum_{n=1}^{N_{mc}} \frac{\| \boldsymbol{F}(\hat{\boldsymbol{\varepsilon}}^{(n)}_{\text{MMSE}})\hat{\boldsymbol{x}}^{(n)}_{\text{MMSE}} - \boldsymbol{F}(\boldsymbol{\varepsilon})\boldsymbol{x} \|_2^2}{\| \boldsymbol{F}(\boldsymbol{\varepsilon})\boldsymbol{x} \|^2} \quad (28)$$

where $\hat{\boldsymbol{\varepsilon}}^{(n)}_{\text{MMSE}}$ refers to the value of $\hat{\boldsymbol{\varepsilon}}_{\text{MMSE}}$ at the nth run.(likewise for $\hat{\boldsymbol{x}}^{(n)}_{\text{MMSE}}$).

4.2.2. Performance of the proposed algorithms

The parametric and first-derivative-based models are studied in the case of a simple scenario: a single target with an SNR of 10 dB (Fig.8(a)) and 20 dB (Fig.8(b))

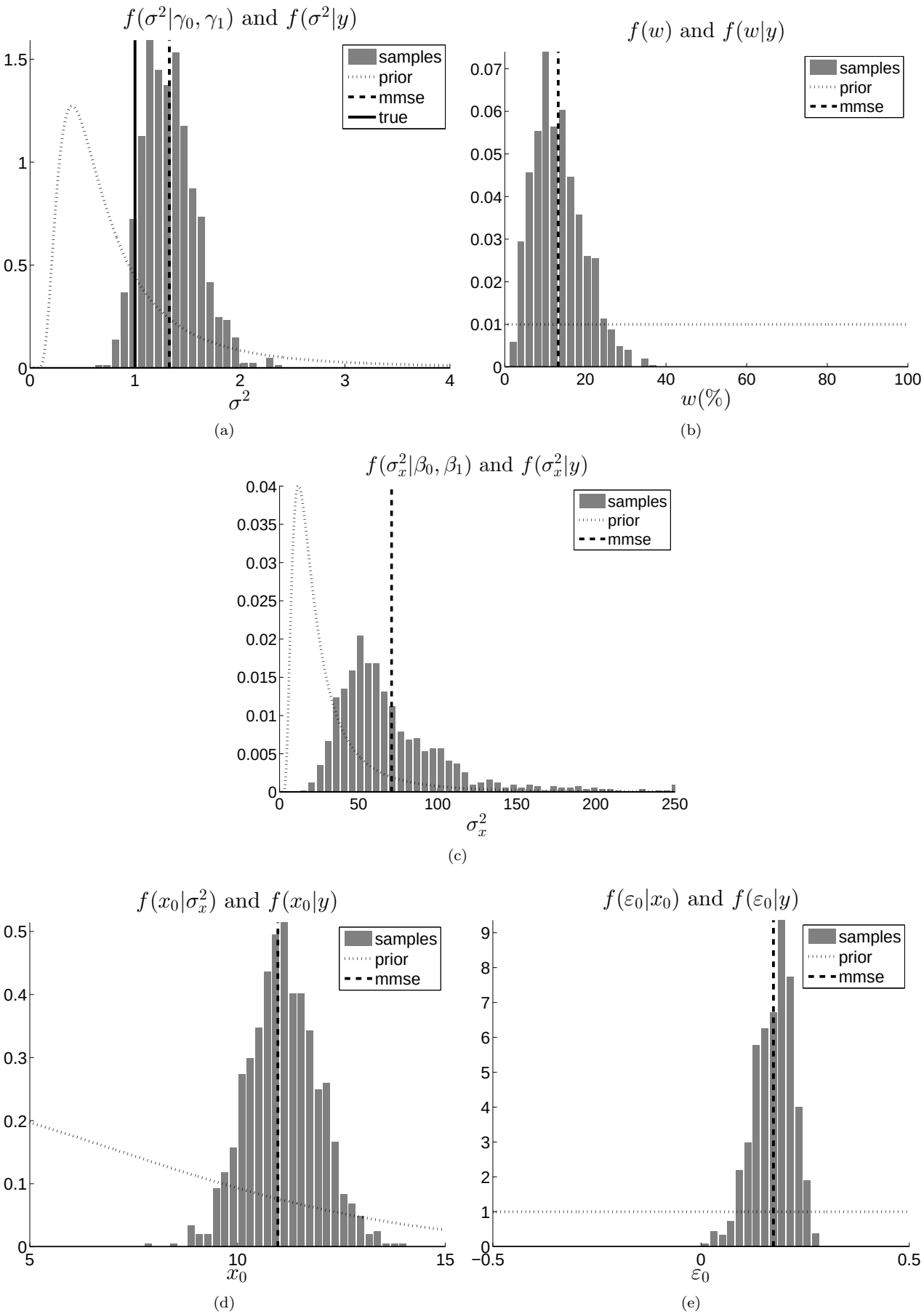


Figure 6: Prior and empirical posterior pdfs of the parameters: (a) Noise power σ^2. (b) Level of occupancy w. (c) Post-processing target power σ_x^2. (d) Amplitude of target x_0. (e) Associated grid mismatch ε_0. The dotted curve is the prior distribution. The dashed line is the estimated value of the parameter. The plain line is the true value in the case of σ^2.

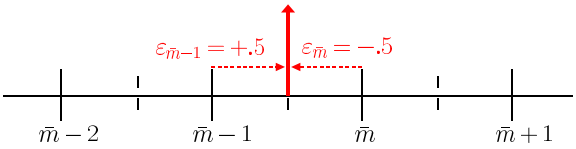


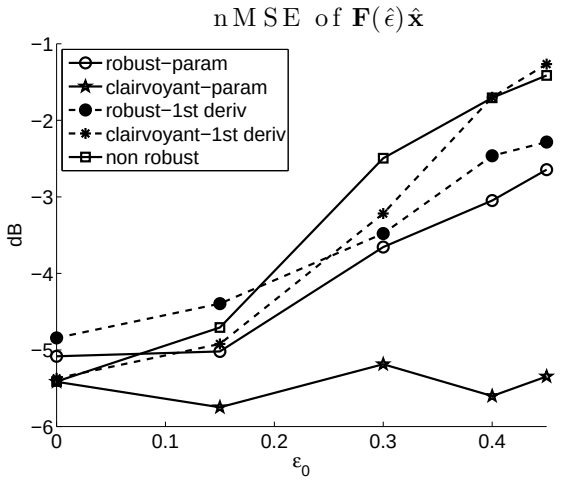
Figure 7: Illustration of the ambiguity about the representation of a target at the edge of a frequency bin. The plain lines represent the center of a frequency bin and the dashed lines the edge of the frequency bin. The target location is symbolized by an arrow.

and a varying mismatch ε_0. Their performance is assessed through the calculation of the MSE of the reconstructed target scene $\boldsymbol{F}(\hat{\boldsymbol{\varepsilon}}_{\text{MMSE}})\hat{\boldsymbol{x}}_{\text{MMSE}}$, and compared with the performance of the non-robust analysis, together with the performance of the clairvoyant analysis wrt $\boldsymbol{\varepsilon}$ (proposed algorithm in Fig.3 with the true value of $\boldsymbol{\varepsilon}$). First, the parametric model outperforms the first-derivative-based model and the non-robust analysis, except for very low values of mismatch ε_0 (in theory, it is better not to estimate the grid mismatch when it is null). The clairvoyant analysis based on the parametric model does not depend on the mismatch, contrary to the first-derivative-based model that highly depends on the mismatch *per se* because of the initial Taylor approximation. It is interesting to see that when SNR=20 dB, the performance of both models drops off when increasing the mismatch, but it occurs for a lower value with the first-derivative-based model and more dramatically.

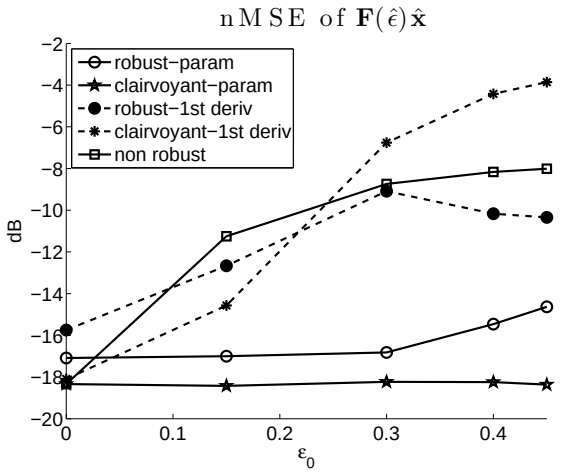
In the interest of fairness, our algorithm was then compared with another algorithm of the literature, the OGSBI algorithm [21]. This algorithm is an SSR technique that uses a first-derivative-based model. The sparse vector and grid mismatch are jointly estimated thanks to an Expectation–Maximization (EM) algorithm.

The comparison between our method of resolution and the OGSBI algorithm is represented in Fig.9 through the calculation of the normalized MSE of the reconstructed target scene $\boldsymbol{F}(\hat{\boldsymbol{\varepsilon}})\hat{\boldsymbol{x}}$. The proposed algorithms clearly outperform the OGSBI algorithm, whatever the SNR of the target, even if the OGSBI algorithm does not depend on grid mismatch as much as them. Then, we can assure that the parametric model and the proposed method of resolution are a lot more accurate than a first-derivative-based model used with an EM algorithm.

The high performance of the parametric model is rooted in the choice of prior for $\boldsymbol{x}$, as well as the use of an MCMC algorithm. It should be noted that in the OGSBI algorithm, as in several publications, the sparsity is only induced via a two-stage Laplacian prior on vector $\boldsymbol{x}$ or on the real and imaginary parts of vector $\boldsymbol{x}$. Then, only one hyperparameter can be adjusted in order to monitor both the sparsity level and the average target power. On the contrary, in the Bayesian scheme proposed the mixed-type prior on $\boldsymbol{x}$ allows to tune the sparsity level independently from the average target power, e.g., [32]. To finish, the algorithms mostly used in the literature provide sub-optimal



(a) SNR=10dB



(b) SNR=20dB

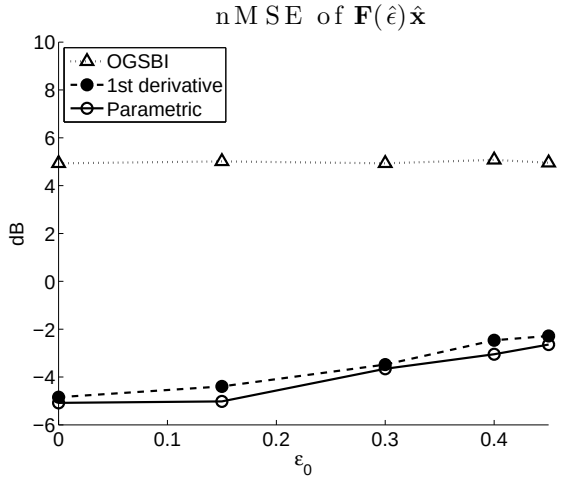
Figure 8: Performance on synthetic target scene: $M = 32$, $N = 1$, $\sigma^2 = 1$, $\bar{M} = M$, $(\mathrm{m}_{\sigma_x^2}, \sqrt{\mathrm{var}_{\sigma_x^2}}) = M_{dB} + (0, 3.5)$ dB, $(\mathrm{m}_{\sigma^2}, \sqrt{\mathrm{var}_{\sigma^2}}) = (0, 2.4)$ dB. Comparison with the clairvoyant wrt $\boldsymbol{\varepsilon}$ and non-robust estimation for both parametric and first-derivative-based model.

solutions compared with the MCMC algorithm, e.g., EM or VB algorithms. The EM algorithm can converge to a local, and not global, maximum [26, chap.5]. The VB algorithm is also sub-optimal because it is based on an approximation of the posterior distribution, and is dependent on the initialization [33].

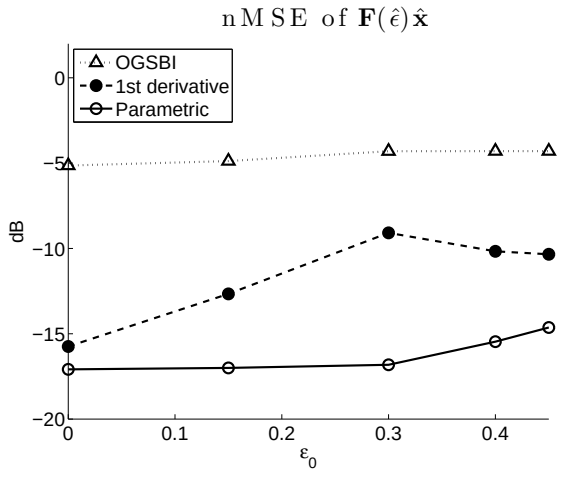
4.2.3. Discussion about the appropriate choice of grid length $\bar{M}$

As underlined in the Introduction, the most natural way to deal with grid mismatch is to refine the grid [12, 14, 17]. One can use this technique jointly with methods estimating grid mismatch, being based on different models, such as a parametric or first-derivative-based model. Here, we assess the influence of grid length on the proposed algorithms and on the OGSBI algorithm.

The different models are compared for different val-



(a) SNR=10dB



(b) SNR=20dB

Figure 9: Comparison between the proposed algorithms and the OGSBI algorithm. The scenario considered is a single target with grid mismatch ε_0. In the OGSBI algorithm, the noise precision was initialized to $1/\sigma^2$ in order to provide the best performance.

ues of grid length $\bar{M}$ in Fig. 10 via the calculation of the normalized MSE of the reconstructed target scene. The signal corresponds to a single target with a Doppler frequency uniformly distributed over the interval $[0, 1/M]$, and an SNR of 10, 20 and 40dB. In the case of low power (SNR=10dB), the grid length does not improve the normalized MSE of the reconstructed target scene, whatever the model considered. However, in the case of high power (SNR=20 and 40dB), the results obtained with a first-derivative-based model (the one presented here and OGSBI) improve when refining the grid, whereas they are unchanged with the parametric model. Then, we find the same results as in [18, 19]: there is no interest in refining the grid with a parametric model while this approach is successful with a first-derivative-based model. However, refining the grid is costly in the proposed algorithm because an MCMC is implemented.

5. Results on experimental data

Finally, the proposed model was tested on experimental data collected from the PARSAX radar [24] on November 2010. For the data set considered, the radar was pointing on a freeway during a heavy traffic time. The ground truth is unknown. The target amplitudes are very high comparing with the ones in the synthetic case, so the hyperparameters (β_0, β_1) are tuned accordingly (a similar tuning is made for the OGSBI algorithm). Fig.11 and Fig.12 show the results obtained on the data set with the parametric model, the first-derivative-based model and the OGSBI algorithm, as well as with the non-robust method. An initialization process was established in order to accelerate the convergence of the proposed algorithms. A Capon analysis was performed on each range bin and all the values under a predefined threshold (the assumed thermal noise level) were set to zero: the resulting vector is used as an initialization for vector $\boldsymbol{x}$ ($\boldsymbol{x}^{(0)}$), its number of non-zero elements is $w^{(0)}$ and its variance is $\sigma_x^{2\,(0)}$. The analysis was made for different values of grid length ($\bar{M} = M$ in Fig.11 and $\bar{M} = 2M$ in Fig.12). The number of realizations was observed as depending on the grid length: 1000 iterations are needed when $\bar{M} = M$ against 5000 iterations when $\bar{M} = 2M$. The amplitude estimated with the Capon analysis is displayed as a transparent background, and can be seen as an indicator of the ground truth.

First, we can see that the proposed algorithms are sparser than the OGSBI algorithm. As underlined before, this might be the consequence of the use of a Laplacian prior in the latter algorithm that induces a coupling between sparsity level and target power. In general, all the algorithms give a good estimation of the possible targets with high power. In particular, the possible target identified at range bins #3 and 4 and velocity $\approx -17\text{m.s}^{-1}$ is split with the proposed method as well as with the OGSBI algorithm so it is not an anomaly inherent to our method (e.g., [22]). However, this target is split only at range

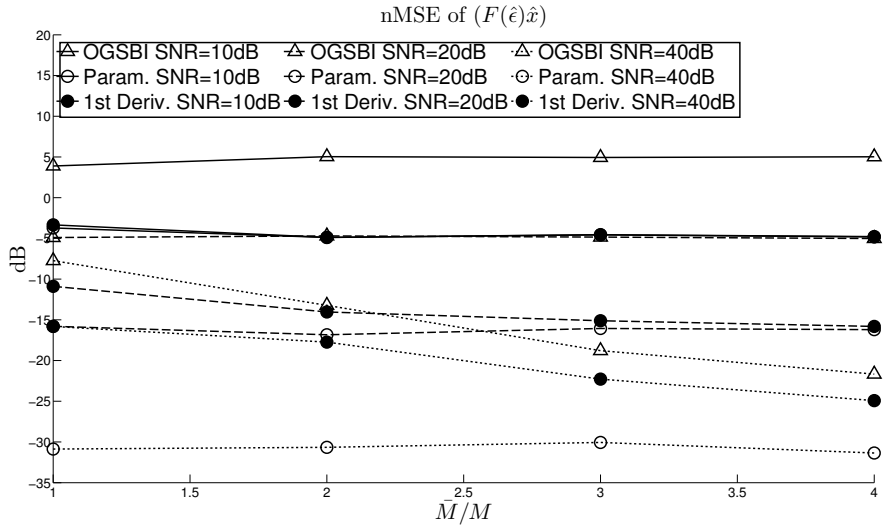


Figure 10: Influence of grid length $\bar{M}$ on the proposed algorithms and OGSBI technique in the case of a single target; $M = 32$, $N_r = 1000$, $N_{bi} = 100$.

bin #4 with the parametric model. Besides, some targets are only estimated with the parametric model: range bin #3, 4 and velocity ≈-5 m.s^{-1}; range bin #7 and velocity ≈-7 m.s^{-1}; range bins #7, 8 and velocity ≈-17, -19 m.s^{-1}. In the case when $\bar{M} = 2M$ the non-robust method better estimates the possible target at range bins #3, 4 and velocity ≈ -17m.s^{-1}: the high mismatch when $\bar{M} = M$ becomes almost null when $\bar{M} = 2M$. The first-derivative-based model improves, but neither the parametric model (some possible targets are split) nor the OGSBI (more possible false alarms and targets split) does. However, the parametric model is still the only one to estimate some possible targets.

6. Conclusions

In this paper, we presented two Bayesian algorithms for the sparse representation of off-grid targets in a Fourier basis and using a Monte-Carlo Markov chain. More specifically, an error vector representing grid mismatch was introduced: it directly parametrizes the Fourier basis in the parametric model, whereas it is used in an additive perturbation matrix in the first-derivative-based model. Both proposed algorithms were proved to be more robust to grid mismatch than a reference algorithm, at the cost of computational complexity. The sampling of the grid mismatch is more difficult with the parametric model, but this model is a lot more accurate in terms of estimation of the reconstructed target scene. Besides, the sparsity was well preserved thanks to the Bernoulli–Gaussian prior even in the case of high mismatch. This prior can also be tuned in a more flexible way, since it adjusts separately the sparsity and the target power.

Some points need to be further investigated. First, there seems to be an intrinsic limitation in the problem formulation since the case of high mismatch is limiting whatever the method used. Moreover, the parameters of the target power need to be set by the radar operator and they can significantly change the performance of the analysis.

Acknowledgement

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

7. References

[1] J. Capon, High-resolution frequency-wavenumber spectrum analysis, Proc. of the IEEE 57 (8) (1969) 1408–1418.

[2] J. Li, P. Stoica, An adaptive filtering approach to spectral estimation and SAR imaging, IEEE Trans. Signal Process. 44 (6) (1996) 1469–1484.

[3] R. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag. 34 (3) (1986) 276–280.

[4] R. Roy, T. Kailath, ESPRIT–estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust., Speech, Signal Process. 37 (7) (1989) 984–995.

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

[6] E. J. Candès, M. B. Wakin, An Introduction To Compressive Sampling, IEEE Signal Processing magazine (March 2008) 21–30.

[7] E. Van den Berg, M. P. Friedlander, Probing the Pareto Frontier for Basis Pursuit Solutions, SIAM Journal on Scientific Computing 31 (2) (2008) 890–912.

[8] R. Tibshirani, Regression Shrinkage and Selection via the Lasso, Journal of the Royal Statistical Society 58 (1) (1996) 267–288.

[9] T. Blumensath, M. E. Davies, Iterative hard thresholding for compressed sensing, Applied and Computational Harmonic Analysis 27 (3) (2009) 265–274.

[10] J. A. Tropp, A. C. Gilbert, Signal Recovery From Random Measurements Via Orthogonal Matching Pursuit, IEEE Transactions on Information Theory 53 (12) (2007) 4655–4666.

[11] W. Dai, O. Milenkovic, Subspace Pursuit for Compressive Sensing Signal Reconstruction, IEEE Transactions on Information Theory 55 (5) (2009) 2230–2249.

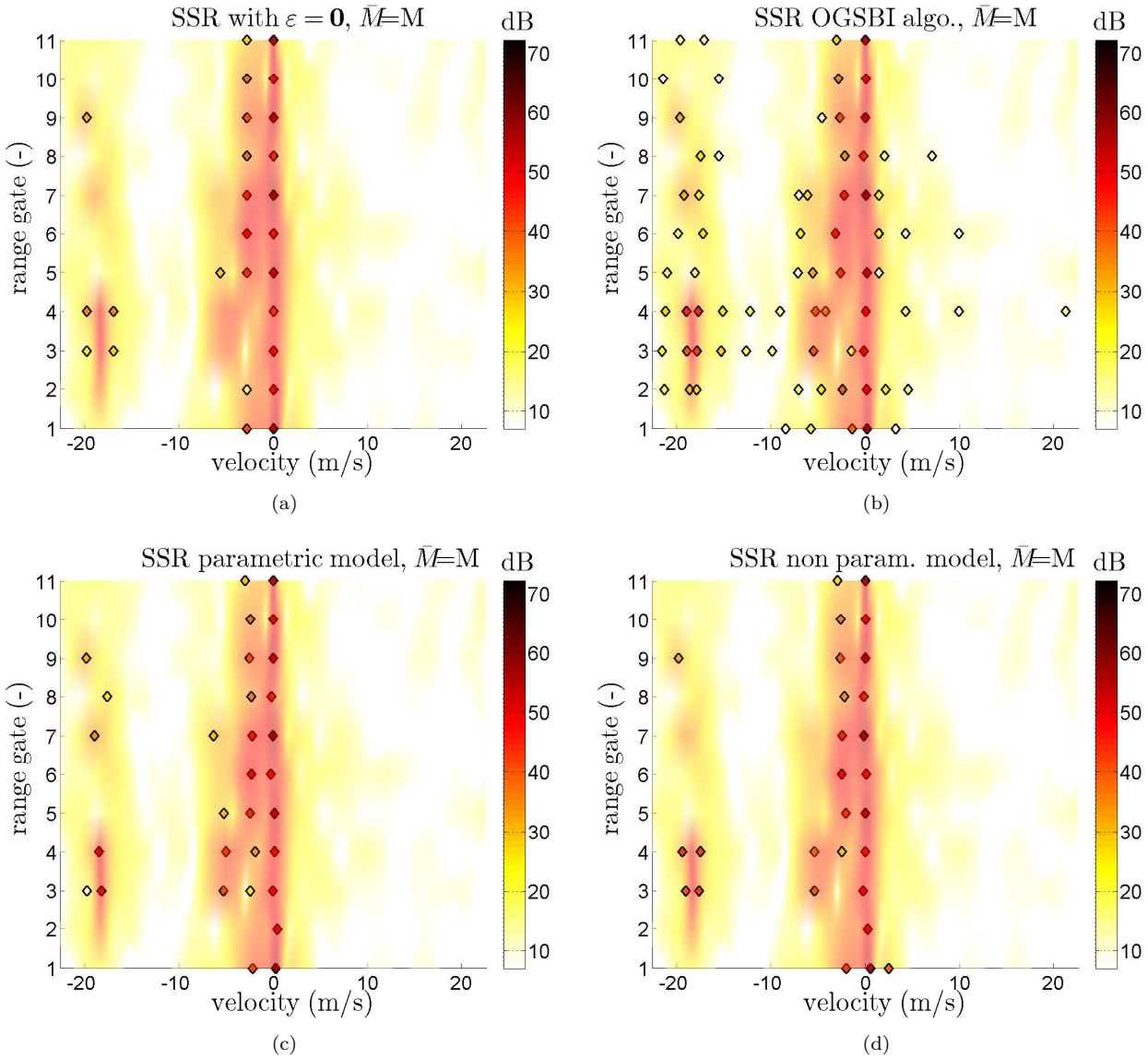


Figure 11: Comparison between parametric model, first-derivative-based model and OGSBI algorithm [21] on experimental PARSAX data. Range resolution of 1.5 m, $M = 16$, N unknown, $\bar{M} = M$, $\sigma^2 \approx 1$ (pre-normalization of the data), $(\mathrm{m}_{\sigma_x^2}, \sqrt{\mathrm{var}_{\sigma_x^2}}) = M_{dB} + (60, 3.5)$ dB, $(\mathrm{m}_{\sigma^2}, \sqrt{\mathrm{var}_{\sigma^2}}) = (0, 2.4)$ dB. The OGSBI algorithm is initialized as in [21]. (a) non-robust estimation, (b) OGSBI algorithm, (c) parametric model, (d) first-derivative-based model.

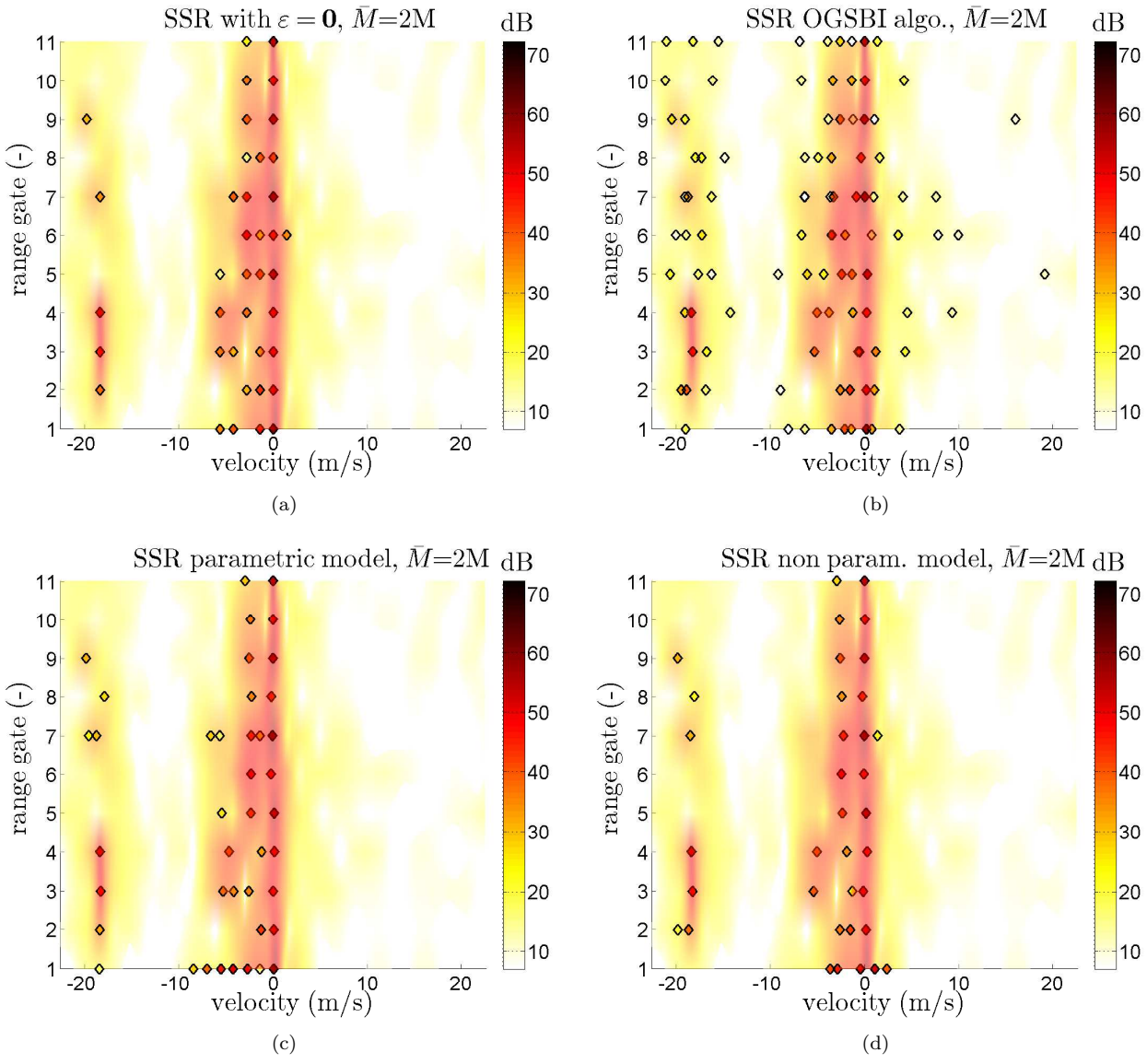


Figure 12: Comparison between parametric model, first-derivative-based model and OGSBI algorithm [21] on experimental PARSAX data. $\bar{M} = 2M$ (a) non-robust estimation, (b) OGSBI algorithm, (c) parametric model, (d) first-derivative-based model.

[12] D. Malioutov, M. Cetin, A. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Process. 53 (8) (2005) 3010–3022.
[13] Y. Chi, L. Scharf, A. Pezeshki, A. Calderbank, Sensitivity to basis mismatch in compressed sensing, IEEE Trans. Signal Process. (5) (2011) 2182–2195.
[14] M. Duarte, R. Baraniuk, Spectral compressive sensing, Appl. Comput. Harmon. Anal. 35 (2013) 111–129.
[15] M. Herman, T. Strohmer, General deviants: An analysis of perturbations in compressed sensing, IEEE Journal of Selected Topics in Signal Processing 4 (2) (2010) 342–349.
[16] M. Rosenbaum, A. Tsybakov, Sparse recovery under matrix uncertainty, The Annals of Statistics 38 (5) (2010) 2620–2651.
[17] A. C. Fannjiang, W. Liao, Coherence-pattern guided compressive sensing with unresolved grids, arXiv: 1106.5177 (June 2011).
[18] L. Hu, Z. Shi, J. Zhou, Q. Fu, Compressed sensing of complex sinusoids: An approach based on dictionary refinement, IEEE Trans. Signal Process. 60 (7) (2012) 3809–3822.
[19] L. Hu, J. Zhou, Z. Shi, Q. Fu, A fast and accurate reconstruction algorithm for compressed sensing of complex sinusoids, IEEE Trans. Signal Process. 61 (22) (2013) 5744–5754.
[20] H. Zhu, G. Leus, G. Giannakis, Sparsity-cognizant total least-squares for perturbed compressive sampling, IEEE Trans. Signal Process. 59 (5) (2011) 2002–2016.
[21] Z. Yang, L. Xie, C. Zhang, Off-grid direction of arrival estimation using sparse Bayesian inference, IEEE Trans. Signal Process. 61 (1) (2013) 38–43.
[22] Z. Tan, A. Nehorai, Sparse direction of arrival estimation using co-prime arrays with off-grid targets, IEEE Signal Process. Lett. 21 (1).
[23] S. Bidon, J.-Y. Tourneret, L. Savy, F. Le Chevalier, Bayesian sparse estimation of migrating targets for wideband radar, IEEE Trans. Aerosp. Electron. Syst. 50 (2) (2014) 871–886.
[24] O. A. Krasnov, G. P. Babur, Z. Wang, L. P. Ligthart, 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 2 (2010) 419–428.
[25] M. Lasserre, S. Bidon, O. Besson, F. Le Chevalier, Bayesian Sparse Fourier Representation of Off-Grid Targets, in: Proc. IEEE International Radar Conference, Lille, 2014.
[26] C. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, New York, NY, 2004.
[27] R. Gatto, S. Jammalamadaka, The generalized von Mises distribution, Statistical Methodology 4 (2007) 341–353.
[28] A. Gretsistas, M. Plumbley, An alternating descent algorithm for the off-grid DoA estimation problem with sparsity constraints, in: Proc. of the 20th European Signal Processing Conference (EUSIPCO 2012), 2012, pp. 874–878.
[29] R. Jagannath, K. Hari, Block sparse estimator for grid matching in single snapshot DoA estimation, IEEE Signal Process. Lett. 20 (11).
[30] Y. Zhang, Z. Ye, X. Xu, N. Hu, Off-grid DoA estimation using array covariance matrix and block-sparse Bayesian learning, Signal Process. 98 (2014) 197–201.
[31] N. Chopin, Fast simulation of truncated Gaussian distributions, Statistics and Computing 21 (2) (2011) 275–288.
[32] N. Dobigeon, A. O. Hero, J.-Y. Tourneret, Hierarchical bayesian sparse image reconstruction with application to MRFM, IEEE Trans. Image Process. 18 (9) (2009) 2059–2070.
[33] V. Smidl, The Variational Bayes Approach in Signal Processing, Ph.D. thesis, University of Dublin, Trinity College (2004).