# Bayesian sparse representation in colored noise: prewhitening vs joint estimation

Stéphanie Bidon
*Department of Electronics Optronics and Signal*
*ISAE-SUPAERO, Université de Toulouse*
Toulouse, France
stephanie.bidon@isae-supaero.fr

***Abstract*—In this paper, we consider the problem of designing sparse signal representation (SSR) amid colored noise. Two processing architectures are examined under a Bayesian framework: i) a two-stage processing with a prewhitening operation followed by SSR assuming a perfect white noise ii) a joint approach estimating at the same time the sparse signal and the colored noise. Both approaches are compared; performance is numerically studied in case of conventional radar scenarios. Results show that the joint algorithm outperforms to some extent the prewhitened approach but at the expense of a higher complexity.**

## I. Introduction

These last few years a plethora of sparse signal representation techniques have been developed and advocated for to estimate radar target scene [1]. Quite often, they have been developed assuming a sparse signal of interest (i.e., target signatures) amid *white* noise. To adapt the algorithms to practical radar scenarios where diffuse clutter and hence *colored* noise arises, it is usually recommended to apply a prewhitening operation to filter the clutter component before performing SSR assuming a *perfect* white noise, e.g., [1], [2]. However, since the noise covariance matrix (CM) is not known, it is replaced in practice by an estimate thereby rendering the assumption of white noise incorrect.

Alternatively, a very few techniques are considering a joint estimation approach in which the colored noise is jointly estimated with the sparse signal [3]–[5]. This processing architecture bypasses *de facto* the white noise approximation encountered in the previous architecture.

One may thus wonder which approach yields better performance and, more particularly, if the perfect white noise assumption usually recommended is that deleterious compared to the joint approach. A similar question was addressed in [6] but investigation was conducted outside the framework of sparse representation. In this paper, we try to give first elements of response to this question considering *Bayesian SSR*. Particularly, an important task is first to clearly ask the question by i) defining the signal model considered; ii) describing both processing architectures; iii) defining outputs of interest to our processing as well as performance metrics. Then, since no closed-form estimation can be performed with our Bayesian model, numerical simulations need to be run to compare both architectures on specific scenarios. Hence, results obtained can determine only partially if, and how far, one architecture outperforms the other.

In the remaining of the paper, Section II describes the Bayesian signal model used to design both SSR architectures presented in Section III and detailed in Section IV. Numerical simulations are provided in Section V while the last Section includes some concluding remarks.

## II. Bayesian signal model

In this work, we consider a Gaussian and homogeneous environment for the data while prior information is injected about the sparse signal and the noise CM.

### A. Observation vectors

We assume to have primary and secondary data. Primary data consists of the sum of a sparse signal of interest plus colored Gaussian noise. The associated sparsifying dictionary is supposed to be known. Secondary data consists of colored Gaussian noise only with the same distribution as the primary noise. This is summarized by

$$\begin{aligned}&\text{for}\quad n \in \{0,\ldots,N_p-1\}\\ &\boldsymbol{y}_n^{(p)} = \boldsymbol{H}\boldsymbol{x}_n + \boldsymbol{n}_n^{(p)} \quad \text{with} \quad \boldsymbol{n}_n^{(p)} \sim \mathcal{CN}_M(\boldsymbol{0},\boldsymbol{R}) \qquad (1a)\\ &\text{for}\quad n \in \{0,\ldots,N_s-1\}\\ &\boldsymbol{y}_n^{(s)} = \boldsymbol{n}_n^{(s)} \quad \text{with} \quad \boldsymbol{n}_n^{(s)} \sim \mathcal{CN}_M(\boldsymbol{0},\boldsymbol{R}) \qquad (1b)\end{aligned}$$

with

| | |
|---|---|
| $N_p$, $N_s$ | the number of primary and secondary data; |
| $\boldsymbol{y}_n^{(p)}$, $\boldsymbol{y}_n^{(s)}$ | the primary and secondary observation vectors of length $M$; |
| $\boldsymbol{H}$ | the sparsifying dictionary of size $M \times \bar{M}$ where $\bar{M}$ is the length of the sparse signal to be reconstructed (usually $\bar{M} \gg M$); |
| $\boldsymbol{x}_n$'s | the $\bar{M}$-length sparse vectors to be estimated; |
| $\boldsymbol{R}$ | the unknown noise covariance matrix. |

We assume statistical independence between the $\boldsymbol{x}_n$, $\boldsymbol{n}_n^{(p)}$ and $\boldsymbol{n}_n^{(s)}$'s. In the following, matrix expressions are sometimes favored to describe the observation model (1), i.e.,

$$\boldsymbol{Y}^{(p)} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{N}^{(p)} \qquad (2a)$$

$$\boldsymbol{Y}^{(s)} = \boldsymbol{N}^{(s)} \qquad (2b)$$

where matrix notations are generically defined as $\boldsymbol{Y} \triangleq \left[\boldsymbol{y}_1 \quad \ldots \quad \boldsymbol{y}_{N_p}\right]$. The likelihood function is given by

$$f(\boldsymbol{Y}^{(p)}, \boldsymbol{Y}^{(s)}|\boldsymbol{X}, \boldsymbol{R}) \propto f(\boldsymbol{Y}^{(p)}|\boldsymbol{X}, \boldsymbol{R}) f(\boldsymbol{Y}^{(s)}|\boldsymbol{R})$$

with

$$f(\boldsymbol{Y}^{(p)}|\boldsymbol{X}, \boldsymbol{R}) \propto \frac{\operatorname{etr}\left\{-\boldsymbol{R}^{-1}(\boldsymbol{Y}^{(p)}-\boldsymbol{H}\boldsymbol{X})(\boldsymbol{Y}^{(p)}-\boldsymbol{H}\boldsymbol{X})^H\right\}}{|\boldsymbol{R}|^{N_p}} \quad (3)$$

$$f(\boldsymbol{Y}^{(s)}|\boldsymbol{R}) \propto |\boldsymbol{R}|^{-N_s} \operatorname{etr}\left\{-\boldsymbol{R}^{-1}\boldsymbol{Y}^{(s)}\boldsymbol{Y}^{(s)H}\right\} \quad (4)$$

where $\propto$ means proportional to, $|.|$ is the determinant of a matrix and $\operatorname{etr}\{.\}$ is the exponential of the trace.

In radar, the model (2) can represent the signal received from a narrowband waveform in a homogeneous environment with $\boldsymbol{y}_n^{(p)}$ the signal from a range gate where targets are likely present and $\boldsymbol{y}_n^{(s)}$ the signal from a range gate known to be target-free. For a pulse Doppler radar, $\boldsymbol{H}$ can typically be a(n) (oversampled) Fourier dictionary and $\boldsymbol{X}$ the target amplitude matrix in the (primary) range-Doppler map.

In any event, a Bayesian framework is further assumed, meaning that each unknown parameter (i.e., $\boldsymbol{X}$ and $\boldsymbol{R}$) is considered as random with a given prior as described next.

### B. Prior model

*1) Sparse signal:* A sparsity-inducing prior is assigned to the signal of interest $\boldsymbol{X}$[1]. As an *example* to answer to our central question, we assume the elements of $\boldsymbol{X}$ independent and identically distributed according to a Bernoulli-Student-*t* law, i.e.,

$$\pi(\boldsymbol{X}) = \prod_{n=0}^{N_p-1} \prod_{\bar{m}=0}^{\bar{M}-1} \pi([\boldsymbol{x}_n]_{\bar{m}}) \quad (5)$$

where each element $[\boldsymbol{x}_n]_{\bar{m}}$ has a two-stage hierarchical prior

$$\pi([\boldsymbol{x}_n]_{\bar{m}}) = \int\int \pi([\boldsymbol{x}_n]_{\bar{m}}|w, \sigma_x^2)\pi(w)\pi(\sigma_x^2)dwd\sigma_x^2 \quad (6)$$

and

$$\pi([\boldsymbol{x}_n]_{\bar{m}}|w, \sigma_x^2) = (1-w)\delta([\boldsymbol{x}_n]_{\bar{m}}) + w\mathrm{CN}\left([\boldsymbol{x}_n]_{\bar{m}}|0, \sigma_x^2\right) \quad (7a)$$

$$\pi(w) = \mathbb{I}_{[0,1]}(w) \quad (7b)$$

$$\pi(\sigma_x^2) = \mathrm{IG}\left(\sigma_x^2|\beta_0, \beta_1\right) \quad (7c)$$

with $\mathrm{CN}\left(|,\right)$ the complex Gaussian pdf probability density function (pdf) with given mean and variance, $\mathrm{IG}\left(|,\right)$ the inverse gamma pdf with given shape and scale parameters, $\delta()$ the Dirac delta function and $\mathbb{I}(.)$ the indicator function on a given set. The distributions described by the pdfs (7a), (7b) and (7c) are respectively Bernoulli-Gaussian, Beta and inverse Gamma; they are denoted respectively by $\mathcal{B}er\mathcal{CN}\left(w, 0, \sigma_x^2\right)$, $\mathcal{B}e\left(1, 1\right)$, $\mathcal{IG}\left(\beta_0, \beta_1\right)$.

Going back to our former pulse Doppler radar example, the prior (5)-(6)-(7) actually assumes that a target is likely present for each range gate $n$ and each Doppler bin $\bar{m}$ with an unknown probability $w$ and, if so, its amplitude is Gaussian distributed with an unknown power $\sigma_x^2$. This prior has proved to be adequate to recover sparse target scene, e.g., [5]. Note that in practice, the radar operator has only to set the values of $\beta_0, \beta_1$ (recommendations can be found e.g., in [5]).

*2) Noise covariance matrix:* As in [6] we restrict the study to a specific family of priors defined as

$$\pi(\boldsymbol{R}) \propto \frac{1}{|\boldsymbol{R}|^{\nu+M}} \operatorname{etr}\left\{-\boldsymbol{R}^{-1}\boldsymbol{A}\right\}. \quad (8)$$

We assume that $\nu \geq 0$ and $\boldsymbol{A}$ is a positive-semidefinite Hermitian matrix. The family of prior (8) enables not only mathematical tractability but also includes well known priors. In particular, it reduces to a so-called Jeffrey's prior when $\nu = 0$ and $\boldsymbol{A} = \boldsymbol{0}$ [7], i.e.,

$$\pi_J(\boldsymbol{R}) \propto |\boldsymbol{R}|^{-M} \quad (9)$$

and to an inverse complex Wishart prior when $\nu > M$ and $\boldsymbol{A}$ is positive definite [8], i.e.,

$$\pi_{IW}(\boldsymbol{R}) = \frac{|\boldsymbol{A}|^{\nu}}{|\boldsymbol{R}|^{\nu+M}} \frac{\operatorname{etr}\left\{-\boldsymbol{R}^{-1}\boldsymbol{A}\right\}}{\pi^{M(M-1)/2}\prod_{m=1}^{M}\Gamma(\nu-m+1)} \quad (10)$$

with $\Gamma()$ the Gamma function. The distribution described by (10) is denoted as $\boldsymbol{R} \sim \mathcal{CW}_M^{-1}\left(\boldsymbol{A}, \nu\right)$. The Jeffrey's prior is known as an noninformative prior whereas the inverse complex Wishart prior can be made very informative according to the degree of freedom $\nu$.

## III. Overview of two SSR architectures

In the following, we focus on two SSR architectures that can be designed from the Bayesian signal model described in Section II. They are represented in Figs. 1 and 2 and tagged by "preW-SSR" and "C-SSR", respectively.

The first considered architecture preW-SSR is often recommended in the literature. In this approach, secondary data $\boldsymbol{Y}^{(s)}$ are used only to estimate the noise CM $\boldsymbol{R}$ (or more precisely the inverse $\boldsymbol{R}^{-1}$) to further whiten the primary data $\boldsymbol{Y}^{(p)}$. Assuming then a perfect whitening, an SSR algorithm is applied to estimate the sparse signal $\boldsymbol{X}$ amid white noise. The second architecture C-SSR has not been intently investigated so far in the framework of sparse representation. In this approach, primary and secondary data are used jointly to estimate both the covariance matrix $\boldsymbol{R}$ and the sparse signal $\boldsymbol{X}$.

Some important differences appear between both architectures especially since we consider a Bayesian framework.

- In the preW-SSR architecture, primary data is whitened by an estimate of the noise CM. This means that one has to choose a specific estimator whereas in the C-SSR architecture no such hard decision is made within the processing. Using only a point estimate to whiten the data (instead of the full posterior) may lead to a loss of information about the noise CM. In this paper, we choose the MMSE (Minimum Mean Square Error) estimator as a point estimate.

[1] On the contrary, in [6] the prior on $\boldsymbol{X}$ is assumed noninformative and chosen uniform.

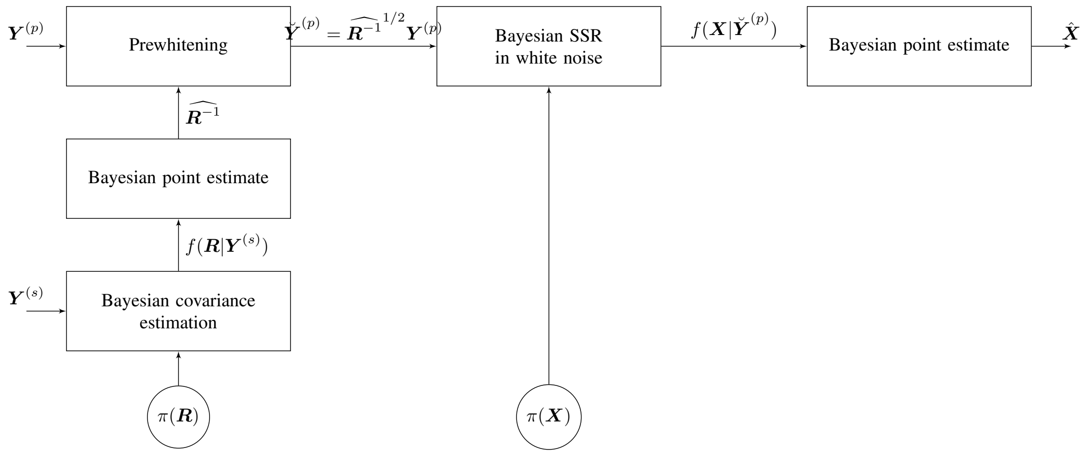


Fig. 1. Flowchart of Bayesian SSR architecture with prewhitening (preW-SSR).

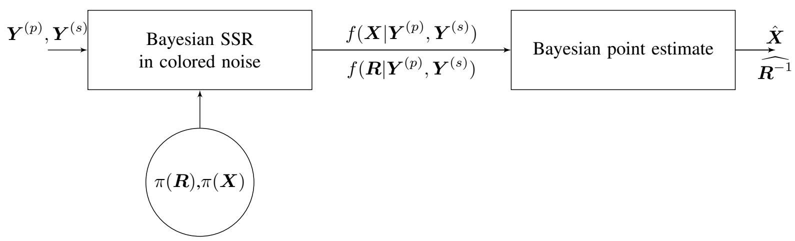


Fig. 2. Flowchart of Bayesian SSR architecture with joint estimation in colored noise (C-SSR).

- In the preW-SSR architecture only $N_s$ data are used to estimate the noise CM whereas $N_s+N_p$ are used in the C-SSR approach. If only a few secondary data is available, the preW-SSR architecture may thus be detrimental to the estimation of the noise CM and hence to prewhitening.
- In the preW-SSR architecture, the first-stage processing assumes perfect whitening. Though in practice, since an estimate is used rather than the true noise CM, this assumption is erroneous (for both the whitened noise and dictionary) and may lead to performance degradation in the subsequent SSR in white noise.
- Finally, each architecture leads to a posterior pdf of the sparse vector. A question arises then on how comparing the performance of both architectures. To that end, we choose in this work to focus on a specific Bayesian estimator of $\boldsymbol{X}$, namely the MMSE estimator, and use a specific performance metrics described later in Section V.

## IV. Bayesian SSR algorithms

In this Section, we detail the SSR algorithms implemented in both studied architectures.

### A. Principle of Bayesian SSR algorithm

To estimate the sparse signal $\boldsymbol{X}$, we need to derive first its posterior distribution. Given the Bayesian model of Section II, the latter has an intricate expression which prevents from obtaining conventional Bayesian estimators in closed form. We thus turn to a numerical method dubbed Gibbs sampling [9]. Let denote $\boldsymbol{\theta}$ the set of unknown random variables in the estimation problem. The Gibbs sampler is an iterative algorithm that generates for each parameter $\boldsymbol{\zeta} \in \boldsymbol{\theta}$ samples distributed according to its full conditional, i.e., $f(\boldsymbol{\zeta}|\boldsymbol{\theta}_{-\boldsymbol{\zeta}}, \boldsymbol{Y})$ where $\boldsymbol{\theta}_{-\boldsymbol{\zeta}}$ is $\boldsymbol{\theta}$ deprived of $\boldsymbol{\zeta}$ and $\boldsymbol{Y}$ represents the data available in the problem ($\boldsymbol{Y}$ is clarified for each architecture in what follows). After a burn in period, the iterative procedure is known to produce samples according to their posterior $f(\boldsymbol{\zeta}|\boldsymbol{Y})$. Particularly, the Gibbs sampling technique allows us to obtain the empirical posterior of $\boldsymbol{X}$ and consequently its MMSE estimator as an empirical mean, viz

$$\hat{\boldsymbol{X}}_{\text{mmse}} = \frac{1}{N_r} \sum_{t=1}^{N_r} \boldsymbol{X}^{(t+N_{bi})} \tag{11}$$

where

$\boldsymbol{X}^{(t)}$ is the sample of $\boldsymbol{X}$ generated at the $t$-th iteration;
$N_{bi}$ is the burn in period of the sampler;
$N_r$ is the number of useful samples.

The Gibbs sampling strategy is used in both SSR architectures.

### B. SSR with prewhitening (preW-SSR)

*1) Covariance estimation:* In the preW-SSR architecture, the noise CM is first estimated with the secondary data only. Using Bayes theorem with (4) and (8), the posterior of $\boldsymbol{R}$ can be expressed as

$$\begin{aligned} f(\boldsymbol{R}|\boldsymbol{Y}^{(s)}) &\propto f(\boldsymbol{Y}^{(s)}|\boldsymbol{R})\pi(\boldsymbol{R}) \\ &\propto \frac{1}{|\boldsymbol{R}|^{N_s+\nu}} \operatorname{etr}\left\{-\boldsymbol{R}^{-1}\left[\boldsymbol{Y}^{(s)}{\boldsymbol{Y}^{(s)}}^H + \boldsymbol{A}\right]\right\} \end{aligned}$$

where we recognize an inverse complex Wishart distribution $\boldsymbol{R}|\boldsymbol{Y}^{(s)} \sim \mathcal{CW}_M^{-1}\left(\boldsymbol{Y}^{(s)}\boldsymbol{Y}^{(s)H} + \boldsymbol{A}, N_s + \nu\right)$. As discussed in Section III, we choose to whiten the data with the MMSE estimator of $\boldsymbol{R}^{-1}$, i.e., [8]

$$\widehat{\boldsymbol{R}^{-1}} = \mathcal{E}\left\{\boldsymbol{R}^{-1}|\boldsymbol{Y}^{(s)}\right\} = \left[\frac{\boldsymbol{Y}^{(s)}\boldsymbol{Y}^{(s)H} + \boldsymbol{A}}{N_s + \nu}\right]^{-1}. \quad (12)$$

With this choice and in case of a Jeffrey's prior (9), we recognize in (12) the conventional sample covariance matrix defined by $N_s^{-1}\boldsymbol{Y}^{(s)}\boldsymbol{Y}^{(s)H}$ as in [6].

*2) Pre-whitening:* Once the (inverse) noise CM estimated, primary data can be filtered as $\breve{\boldsymbol{Y}}^{(p)} = \widehat{\boldsymbol{R}^{-1}}^{1/2}\boldsymbol{Y}^{(p)}$ and the observation model becomes

$$\breve{\boldsymbol{Y}}^{(p)} = \breve{\boldsymbol{H}}\boldsymbol{X} + \breve{\boldsymbol{N}}^{(p)} \quad (13)$$

where the breve mark $\breve{\cdot}$ designates the whitening operation via $\widehat{\boldsymbol{R}^{-1}}^{1/2}$. In practice, an important assumption is then usually made in the preW-SSR architecture. Namely, the whitened primary noise is supposed to be perfectly white, meaning here that for $n \in \{0, \ldots, N_p - 1\}$,

$$\breve{\boldsymbol{n}}_n^{(p)} \underset{\text{approx.}}{\sim} \mathcal{CN}_M(\boldsymbol{0}, \boldsymbol{I}). \quad (14)$$

*3) SSR in assumed white noise:* The sparse signal $\boldsymbol{X}$ is then estimated assuming the signal model (13)-(14). As explained in Section IV-A, a Gibbs sampling strategy is used. To determine the full conditionals required to implement each Gibbs move, we express the joint posterior of $\boldsymbol{\theta} \triangleq \left\{\boldsymbol{X}, w, \sigma_x^2\right\}$ as

$$f(\boldsymbol{X}, w, \sigma_x^2|\breve{\boldsymbol{Y}}^{(p)}) \propto f(\breve{\boldsymbol{Y}}^{(p)}|\boldsymbol{X})\pi(\boldsymbol{X}|w, \sigma_x^2)\pi(w)\pi(\sigma_x^2). \quad (15)$$

Full conditionals are then determined by fixing all but one parameter in (15). Similar calculations have been conducted in [5] and lead here to the following 3 Gibbs moves

$$w|\breve{\boldsymbol{Y}}^{(p)}, \boldsymbol{\theta}_{-w} \sim \mathcal{B}e\left(1 + \textstyle\sum_{n=0}^{N_p-1}\|\boldsymbol{x}_n\|_0, 1 + N_p\bar{M} - \sum_{n=0}^{N_p-1}\|\boldsymbol{x}\|_0\right)$$

$$\sigma_x^2|\breve{\boldsymbol{Y}}^{(p)}, \boldsymbol{\theta}_{-\boldsymbol{x}} \sim \mathcal{IG}\left(\textstyle\sum_{n=0}^{N_p-1}\|\boldsymbol{x}\|_0 + \beta_0, \sum_{n=0}^{N_p-1}\|\boldsymbol{x}\|_2^2 + \beta_1\right)$$

$$[\boldsymbol{x}_n]_{\bar{m}}|\breve{\boldsymbol{Y}}^{(p)}, \boldsymbol{\theta}_{-[\boldsymbol{x}_n]_{\bar{m}}} \sim \mathcal{B}er\mathcal{CN}\left(\breve{w}_{n,\bar{m}}, \breve{\mu}_{n,\bar{m}}, \breve{\eta}_{\bar{m}}^2\right)$$

where $\|\|_2$ is the Frobenius norm, $\|\|_0$ is the number of nonzero elements and

$$\begin{cases} \breve{w}_{n,\bar{m}} = \left[\dfrac{1-w}{w}\dfrac{\sigma_x^2}{\breve{\eta}_{\bar{m}}^2}\exp\left\{-\dfrac{|\breve{\mu}_{n,\bar{m}}|^2}{\breve{\eta}_{\bar{m}}^2}\right\} + 1\right]^{-1} \\ \breve{\eta}_{\bar{m}}^2 = \left\{\dfrac{1}{\sigma_x^2} + \|\breve{\boldsymbol{h}}_{\bar{m}}\|_2^2\right\}^{-1} \quad \text{and} \quad \breve{\mu}_{n,\bar{m}} = \breve{\eta}_{\bar{m}}^2\breve{\boldsymbol{h}}_{\bar{m}}^H\breve{\boldsymbol{e}}_{n,\bar{m}} \end{cases}$$

with $\breve{\boldsymbol{h}}_{\bar{m}}$ the $\bar{m}$th colon of $\breve{\boldsymbol{H}}$ and $\breve{\boldsymbol{e}}_{n,\bar{m}} = \breve{\boldsymbol{y}}_n - \sum_{\bar{m}' \neq \bar{m}}[\boldsymbol{x}_n]_{\bar{m}'}\breve{\boldsymbol{h}}_{\bar{m}'}$.

### C. SSR with joint estimation (C-SSR)

In the C-SSR architecture, full conditionals are determined using the joint posterior of $\boldsymbol{\theta} \triangleq \left\{\boldsymbol{X}, w, \sigma_x^2, \boldsymbol{R}\right\}$, i.e.,

$$f(\boldsymbol{X}, w, \sigma_x^2, \boldsymbol{R}|\boldsymbol{Y}^{(p)}, \boldsymbol{Y}^{(s)}) \propto f(\boldsymbol{Y}^{(p)}|\boldsymbol{X}, \boldsymbol{R})f(\boldsymbol{Y}^{(s)}|\boldsymbol{R}) \\ \pi(\boldsymbol{X}|w, \sigma_x^2)\pi(w)\pi(\sigma_x^2)\pi(\boldsymbol{R}). \quad (17)$$

One can show that the Gibbs sampler is the same as described previously for the preW-SSR architecture albeit an additional Gibbs move defined by

$$\boldsymbol{R}^{-1}|\boldsymbol{Y}^{(p)}, \boldsymbol{Y}^{(s)}, \boldsymbol{\theta}_{-\boldsymbol{R}} \sim \mathcal{CW}_M\left(\boldsymbol{M}^{-1}, N_s + N_p + \nu\right)$$

with $\boldsymbol{M} = \boldsymbol{Y}^{(s)}\boldsymbol{Y}^{(s)H} + (\boldsymbol{Y}^{(p)} - \boldsymbol{H}\boldsymbol{X})(\boldsymbol{Y}^{(p)} - \boldsymbol{H}\boldsymbol{X})^H + \boldsymbol{A}$ and $\mathcal{CW}(,)$ denotes the complex Wishart distribution. Note that this Gibbs move is highly computationally intensive.

## V. Numerical results

### A. Scenario

Herein, we assess performance of both architectures with synthetic data generated according to (1). The dictionary $\boldsymbol{H}$ is a Fourier matrix oversampled with zeropadding factor $n_{\text{zp}}$. The CM is built as $\boldsymbol{R} = \boldsymbol{R}_c + \boldsymbol{R}_n$ with $\boldsymbol{R}_n \propto \boldsymbol{I}$ the thermal noise CM and $\boldsymbol{R}_c$ the clutter CM such that $[\boldsymbol{R}_c]_{m,m'} \propto \exp\left\{-1/2[2\pi\sigma_v(m - m')]^2\right\}$ where $\sigma_v$ is the standard deviation of the velocity of internal clutter motion (expressed in ambiguous velocity) [10]. The clutter-to-noise ratio (CNR) is defined as $\text{CNR} = \text{Tr}\left\{\boldsymbol{R}_c\right\}/\text{Tr}\left\{\boldsymbol{R}_n\right\}$. A single target with constant amplitude is present in the primary data with signal-to-interference-plus-noise-ratio (SINR) defined as $\text{SINR} = |[\boldsymbol{x}_n]_{\bar{m}}|^2\boldsymbol{h}_{\bar{m}}^H\boldsymbol{R}^{-1}\boldsymbol{h}_{\bar{m}}$ with $n$ and $\bar{m}$ the range gate and Doppler bin of the target.

### B. Single run

Typical range-velocity maps obtained from the MMSE estimators of the sparse signal $\boldsymbol{X}$ are represented in Fig. 3. In this example, the number of secondary data is very low and a noninformative Jeffrey's prior (9) is chosen for the noise CM $\boldsymbol{R}$. Hence, the latter is badly estimated in the preW-SSR architecture and leads to poor prewhitening. This translates directly in many false estimations (cf. Fig. 3(b)). On the contrary, the joint C-SSR approach clearly benefits from the primary data as a source of information and/or the absence of approximation about the noise to remove clutter. Note also, that both architectures have difficulties to identify if the contribution around zero velocity comes from a discrete (conveyed in $\boldsymbol{X}$) or a diffuse clutter (conveyed in $\boldsymbol{R}$). The target is identified in both structures.

### C. Monte-Carlo simulations

To confirm and better apprehend trends observed in a single run, we proceed via Monte-Carlo simulations and estimate, as in [5], the power of the reconstructed MMSE scene *after* whitening, i.e.,

$$P_{\mathcal{Z}} \triangleq \mathcal{E}\left\{\left\|\widehat{\boldsymbol{R}^{-1}}^{1/2}\sum_{(n,\bar{m})\in\mathcal{Z}}\boldsymbol{h}_{\bar{m}}\widehat{[\boldsymbol{x}_n]}_{\bar{m}}\right\|_2^2\right\}$$

where $\mathcal{Z}$ is the set of indices $(n, \bar{m})$ defining a specific region in the range-Doppler map. In particular, we define a white noise zone as illustrated in Fig.3 (targets are excluded from this zone). The powers of the target $P_{\mathcal{T}}$ and the white noise zone $P_{\mathcal{W}}$ are depicted in Fig. 4 for the case $N_p = 5$ and $N_p = 1$. In both architectures and with an increased number of secondary data, $P_{\mathcal{T}}$ (which is always overestimated) converges to a value near to that of the true SINR, also the phenomenon of false estimation in the white noise zone diminishes. PreW-SSR and C-SSR techniques converge to the same performance. Overall, C-SSR architecture outperforms that of preW-SSR but only significantly with few secondary data and an increasing number of primary data.

## VI. Conclusion

In this work, we have specified two architectures to estimate a sparse radar scene in colored noise under a Bayesian framework. The first architecture is often prescribed in the literature and entails a prewhitened stage assumed to be perfect. The second architecture is a joint approach that estimates both the sparse signal and the noise covariance matrix. To conduct a performance analysis of both architectures, we have set the limitations of our study (choice of specific noise environment, priors, point estimators and performance metrics). Numerical results indicate that the joint approach outperforms the conventional prewhitening-based approach, particularly with few secondary data and an increasing amount of primary data. However this is obtained at the expense of a dramatically increased computational load. Further work should be devoted to the analysis of more scenarios.

## References


[1] J. T. Parker, M. A. Ferrara, and L. C. Potter, "Radar applications of sparse reconstruction and compressed sensing," in *Principle of Modern Radar. Advanced Techniques*, W. L. Melvin and J. A. Scheer, Eds. Edison, NJ: SciTech Publishing, 2013, ch. 5.

[2] M. A. Atassi and I. Abou-Faycal, "A reconstruction algorithm for noisy compressed sensing; the UWB channel estimation test case," in *2012 19th International Conference on Telecommunications (ICT)*, April 2012, pp. 1–6.

[3] J. J. Fuchs, "DOA estimation in the presence of unknown colored noise, the global matched filter approach," in *2010 18th European Signal Processing Conference*, Aug 2010, pp. 1369–1373.

[4] L. Ning, T. T. Georgiou, and A. Tannenbaum, "Separation of system dynamics and line spectra via sparse representation," in *49th IEEE Conference on Decision and Control (CDC)*, Dec 2010, pp. 473–478.

[5] S. Bidon, M. Lasserre, and F. Le Chevalier, "Unambiguous sparse recovery of migrating targets with a robustified bayesian model," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 1, pp. 108–123, Feb 2019.

[6] L. Svensson and M. Lundberg, "On posterior distributions for signals in Gaussian noise with unknown covariance matrix," *IEEE Transactions on Signal Processing*, vol. 53, no. 9, pp. 3554–3571, Sept 2005.

[7] H. Jeffreys, *Theory of probability*. Oxford, UK: Oxford University Press, 1961.

[8] J. A. Tague and C. I. Caldwell, "Expectations of useful complex wishart forms," *Multidimensional Systems and Signal Processing*, vol. 5, no. 3, pp. 263–279, Jul 1994.

[9] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York, NY: Springer Science, 2004.

[10] J. Ward, "Space-time adaptive processing for airborne radar," Lincoln Laboratory, MIT, Lexington, MA, Tech. Rep. 1015, Dec. 1994.


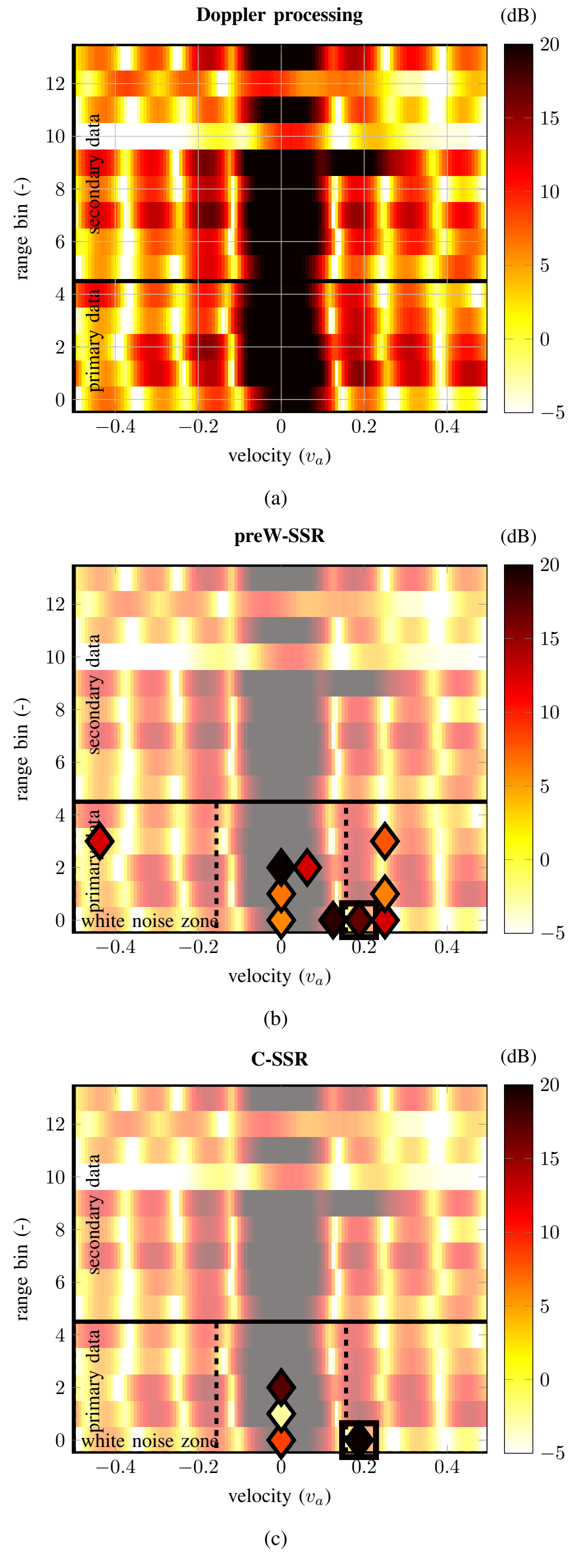


Fig. 3. Range-velocity map. Scenario: $M = 8$, $N_p = 5$, $N_s = M + 1$, CNR = 15 dB, $\sigma_v = 0.01 v_a$ ($v_a$ is the ambiguous velocity), target ($n = 1$, $\bar{m} = 3$, SINR=20 dB). Processing: $n_{\text{zp}} = 2$ (i.e., $\bar{M} = 16$), $\nu = 0$, $\boldsymbol{A} = \boldsymbol{0}$, $N_{bi} = 2E + 3$, $N_r = 500$, $\beta_0 \approx 1E3$, $\beta_1 =\approx 10E6$ (see [5]). (a) Doppler processing. (b) preW-SSR architecture with sampler initialization $\boldsymbol{x} = \boldsymbol{0}$. (c) C-SSR architecture with sampler initialization $\boldsymbol{x} = \boldsymbol{0}$ and $\boldsymbol{R}^{-1} = (N_s + \nu)(\boldsymbol{Y}^{(s)} \boldsymbol{Y}^{(s)H} + \boldsymbol{A})^{-1}$. Square markers indicate true target location. Diamond markers indicate estimated targets. Doppler processing depicted as a transparent background in (b) and (c).

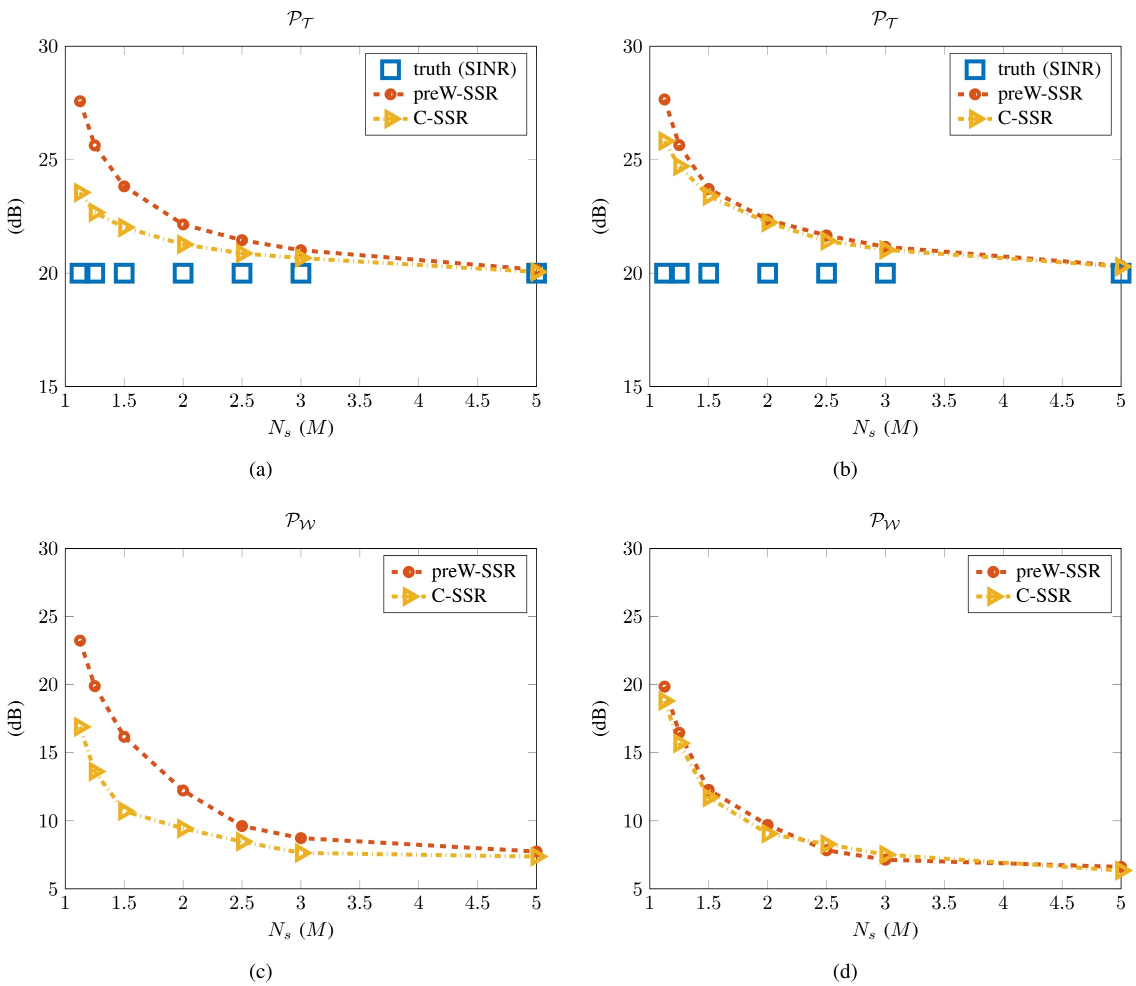


Fig. 4. Power of reconstructed scene *after* whitening. Scenario is that of Fig. 3 except from $N_p$ value and a varying number of secondary data $N_s$. $2E+3$ Monte-Carlo runs. Scenario with $N_p = 5$: (a) Power of reconstructed target and (c) Power of reconstructed white noise zone. Scenario with $N_p = 1$: (b) Power of reconstructed target. (d) Power of reconstructed white noise zone.