SPARSE SIGNAL RECOVERY USING A BERNOULLI GENERALIZED GAUSSIAN PRIOR

Lotfi Chaari[1,2]*, Jean-Yves Tourneret*[1] *and Caroline Chaux*[3]

[1] University of Toulouse, IRIT - INP-ENSEEIHT, France
firstname.lastname@enseeiht.fr

[2] University of Sfax, MIRACL laboratory, Tunisia

[3] Aix-Marseille University, CNRS, Centrale Marseille, I2M, UMR 7373, France
caroline.chaux@univ-amu.fr

ABSTRACT

Bayesian sparse signal recovery has been widely investigated during the last decade due to its ability to automatically estimate regularization parameters. Prior based on mixtures of Bernoulli and continuous distributions have recently been used in a number of recent works to model the target signals, often leading to complicated posteriors. Inference is therefore usually performed using Markov chain Monte Carlo algorithms. In this paper, a Bernoulli-generalized Gaussian distribution is used in a sparse Bayesian regularization framework to promote a two-level flexible sparsity. Since the resulting conditional posterior has a non-differentiable energy function, the inference is conducted using the recently proposed non-smooth Hamiltonian Monte Carlo algorithm. Promising results obtained with synthetic data show the efficiency of the proposed regularization scheme.

***Index Terms*—** Sparse Bayesian regularization, MCMC, ns-HMC, restoration

1. INTRODUCTION

Sparse signal and image restoration has been of increasing interest during the last decades. This issue arises especially in large data applications where regularization is essential to recover a stable solution of the related inverse problem. These applications include remote sensing and medical image reconstruction. Since observation systems are mostly ill-posed, regularization is usually needed to improve the quality of the reconstructed signals. The underlying idea of regularization is to constrain the search space through some prior information added to the model in order to stabilize the inverse problem. This prior information usually involves additional parameters that have to be tuned. Fixing these parameters can be a problem since they can have a deep impact on the quality of the target solution. However, they can be estimated using Bayesian methods that have already shown their efficiency in a number of recent works [1, 2]. As regards regularization, Bayesian techniques have also been widely investigated in a number of recent works using hierarchical Bayesian models, e.g., for medical [3] and hyperspectral imaging [4]. In the recent literature, priors constructed from mixtures of Bernoulli and continuous distributions have gained a significant interest due to their ability to approach the ℓ_0 penalization [3, 5]. These priors are generally combined with the likelihood to derive the posterior distribution of the parameters of interest. The resulting posterior is then used to derive estimators for the model parameters such as the maximum a posteriori (MAP) estimator. Since the target posterior is difficult to handle in most cases, Markov chain Monte Carlo (MCMC) sampling schemes can be used to draw samples according to the posterior. The Metropolis-Hastings (MH) algorithm [6] is one of the most used sampling techniques, even if it requires to design efficient proposal distributions especially for high-dimensional signals or images. This task is not always easy to perform, which has led to the emergence of the Hamiltonian Monte Carlo (HMC) technique [7].

In this paper, we propose a sparse Bayesian regularization scheme involving a Bernoulli-generalized Gaussian (BGG) prior. This prior has the advantage to promote a two-level structured sparsity. The first level is due to the Bernoulli random variable, while the second one is due to the ℓ_p-like energy function of the GG distribution. For the same reasons aforementioned, sampling techniques are required to build the inference and derive the MAP estimator. However, the target distribution involves a non-differentiable energy function, which makes the use of the HMC technique impossible. We therefore resort to the recently proposed non-smooth HMC technique (ns-HMC) which handles such kind of posterior distributions [8] (see also [9] for similar techniques). This technique presents the advantage of fast convergence speed and reduced correlation between the generated samples.

The rest of this paper is organized as follows. The addressed problem is formulated in Section 2. The proposed hierarchical Bayesian model is detailed in Section 3 before introducing the inference scheme in Section 4. Experimental results are presented in Section 5. Finally, some conclusions and perspectives are drawn in Section 6.

2. PROBLEM FORMULATION

Let $\boldsymbol{x} \in \mathbb{R}^M$ be the target signal measured by $\boldsymbol{y} \in \mathbb{R}^N$ through a linear distortion operator $\mathcal{K}$. Accounting for a pos-

This work was supported by the CNRS Imag'In project under grant OPTIMISME

sible additive noise $\boldsymbol{n}$, the observation model writes

$$\boldsymbol{y} = \mathcal{K}\boldsymbol{x} + \boldsymbol{n}. \tag{1}$$

When the above inverse problem is ill-posed, its direct inversion yields distorted solutions. We propose here a framework for sparse Bayesian regularization involving suitable priors for estimating high-dimensional signals $\boldsymbol{x}$ from the observed vector $\boldsymbol{y}$. Without loss of generality, we focus in this paper on linear observation problems with an additive Gaussian noise of covariance matrix $\sigma_n^2 \boldsymbol{I}_N$, where $\boldsymbol{I}_N$ is the identity matrix.

3. HIERARCHICAL BAYESIAN MODEL

By adopting a probabilistic approach, $\boldsymbol{y}$ and $\boldsymbol{x}$ in (1) are assumed to be realizations of random vectors $\boldsymbol{Y}$ and $\boldsymbol{X}$. In this context, our goal is to characterize the probability distribution of $\boldsymbol{X}|\boldsymbol{Y}$, by considering a parametric probabilistic model and by estimating the associated hyperparameters.

3.1. Likelihood

Assuming that the observation noise $\boldsymbol{n}$ is additive and Gaussian with variance σ_n^2, the likelihood can be expressed as

$$f(\boldsymbol{y}|\boldsymbol{x}, \sigma_n^2) = \left(\frac{1}{2\pi\sigma_n^2}\right)^{N/2} \exp\Big(-\frac{\|\boldsymbol{y} - \mathcal{K}\boldsymbol{x}\|_2^2}{2\sigma_n^2}\Big) \tag{2}$$

where $\|.\|_2$ denotes the Euclidean norm.

3.2. Priors

In our model, the vector of unknown parameters is denoted by $\boldsymbol{\theta} = \{\boldsymbol{x}, \sigma_n^2\}$. As regards the noise variance, the only available knowledge is the positivity of this parameter. Therefore, and as motivated in [10], we use a Jeffrey's prior defined as

$$f(\sigma_n^2) \propto \frac{1}{\sigma_n^2} 1_{\mathbb{R}^+}(\sigma_n^2) \tag{3}$$

where $1_{\mathbb{R}^+}$ is the indicator function on $\mathbb{R}^+$, i.e., $1_{\mathbb{R}^+}(\xi) = 1$ if $\xi \in \mathbb{R}^+$ and 0 otherwise. For the signal $\boldsymbol{x}$, we define a prior denoted as $f(\boldsymbol{x}|\boldsymbol{\Phi})$, where $\boldsymbol{\Phi}$ is a hyperparameter vector. Assuming that the signal components x_i are *a priori* independent, the resulting prior distribution for $\boldsymbol{x}$ writes

$$f(\boldsymbol{x}|\boldsymbol{\Phi}) = \prod_{i=1}^{M} f(x_i|\boldsymbol{\Phi}). \tag{4}$$

Promoting the sparsity of the target signal, one can use a Bernoulli-Gaussian (BG) prior for every x_i ($i = 1, \ldots, M$). To better capture the zero signal coefficients and well regularize the non-zero part, we use here a Bernoulli-Generalized Gaussian prior (BGG) for x_i which is defined as

$$f(x_i|\boldsymbol{\Phi}) = (1-\omega)\delta(x_i) + \omega \mathrm{GG}(x_i|\lambda, p) \tag{5}$$

with $\boldsymbol{\Phi} = \{\omega, \lambda, p\}$ and

$$\mathrm{GG}(x_i|\lambda, p) = \frac{p}{2\lambda\Gamma(1/p)} \exp\left(-\frac{|x_i|^p}{\lambda^p}\right) \tag{6}$$

where $\lambda > 0$ and $p > 0$ are the scale and shape parameters, and $\Gamma(.)$ denotes the gamma function. In (5), $\delta(.)$ is the Dirac delta function and ω is a weight belonging to $[0, 1]$. The adopted BGG model promotes a *two-level structured sparsity*. The first level is guaranteed due to the Bernoulli model and the Dirac delta function. The second level is ensured by the GG distribution, especially when $p \leq 1$. This distribution has been widely used in sparse signal recovery [11–13]. When p varies between 0 and 2, the GG distribution gives more flexibility to represent signals with different sparsity levels. BGG regularization therefore approaches the $\ell_0 + \ell_p$ one.
Using a BGG model for every x_i, and assuming independence between the different priors, the prior in (4) reduces to

$$f(\boldsymbol{x}|\boldsymbol{\Phi}) = \prod_{i=1}^{M} \left[(1-\omega)\delta(x_i) + \omega \frac{p}{2\lambda\Gamma(1/p)} \exp\left(-\frac{|x_i|^p}{\lambda^p}\right)\right]. \tag{7}$$

3.3. Hyperparameter priors

Separable non-informative priors are used for each variable of the hyperparameter vector $\boldsymbol{\Phi} = \{\omega, \lambda, p\}$. A uniform distribution on the simplex $[0, 1]$ may be used for ω, i.e., $\omega \sim \mathcal{U}_{[0,1]}$. As regards $\lambda \in \mathbb{R}_+^*$, a suitable prior is the conjugate inverse-gamma (IG) distribution denoted as $\mathcal{IG}(\lambda|\alpha, \beta)$

$$f(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-\alpha-1} \exp\left(-\frac{\beta}{\lambda}\right) 1_{\mathbb{R}^+}(\lambda) \tag{8}$$

where α and β are fixed hyperparameters (in our experiments these hyperparameters were empirically set to $\alpha = \beta = 10^{-3}$). For the shape parameter p, a uniform prior on the simplex $[0, 2]$ is considered. Setting $p \sim \mathcal{U}_{[0,2]}$ guarantees a high level of sparsity for the target solution while using a non-informative prior at the same time.

4. INFERENCE SCHEME

Adopting a MAP strategy, the joint posterior distribution of $\{\boldsymbol{\theta}, \boldsymbol{\Phi}\}$ can be expressed as

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{\Phi}) f(\boldsymbol{\Phi}) \\ &\propto f(\boldsymbol{y}|\boldsymbol{x}, \sigma_n^2) f(\boldsymbol{x}|\omega, p, \lambda) f(\sigma_n^2) f(\omega) f(p) f(\lambda|\alpha, \beta). \end{aligned} \tag{9}$$

Based on the hierarchical Bayesian model described in Section 3, the joint posterior can be written

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\boldsymbol{y}) \propto & \left(\frac{1}{2\pi\sigma_n^2}\right)^{N/2} \exp\Big(-\frac{\|\boldsymbol{y} - \mathcal{K}\boldsymbol{x}\|_2^2}{2\sigma_n^2}\Big) \times \\ & \prod_{i=1}^{M} \left[(1-\omega)\delta(x_i) + \omega \frac{p}{2\lambda\Gamma(1/p)} \exp\left(-\frac{|x_i|^p}{\lambda^p}\right)\right] \times \\ & \frac{1}{\sigma_n^2} 1_{\mathbb{R}^+}(\sigma_n^2) \mathcal{U}_{[0,1]}(\omega) \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-\alpha-1} \exp\left(-\frac{\beta}{\lambda}\right) \mathcal{U}_{[0,2]}(p). \end{aligned} \tag{10}$$

Unfortunately, no closed-form expression for the Bayesian estimators associated with (10) can be obtained. Thus, we propose to design a Gibbs sampler (GS) that generates samples asymptotically distributed according to (10). The GS iteratively generates samples distributed according to the conditional distributions associated with the target distribution. More precisely, the GS iteratively samples according to $f(\sigma_n^2|\boldsymbol{y},\boldsymbol{x})$, $f(\omega|\boldsymbol{x})$, $f(p|\boldsymbol{x},\omega,\lambda)$, $f(\lambda|\boldsymbol{x},p,\alpha,\beta)$ and $f(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\Phi})$ as detailed below.

4.1. Gibbs sampler

Following the strategy detailed above, the main steps of the proposed sampling algorithm are summarized in Algorithm 1. The adopted sampling strategy generates samples

Algorithm 1: Gibbs sampler for Sparse Bayesian regularization.

- Initialize with some $\boldsymbol{\theta}^{(0)}$ and $\boldsymbol{\Phi}^{(0)}$ and set $r=0$;
for $r=1\dots S$ **do**
  ① Sample ${\sigma_n^2}^{(r)}$ according to its posterior $f(\sigma_n^2|\boldsymbol{x},\boldsymbol{y})$;
  ② Sample ω according to its posterior $f(\omega|\boldsymbol{x})$;
  ③ Sample p according to its posterior $f(p|\boldsymbol{x})$;
  ④ Sample λ according to its posterior $f(\lambda|\boldsymbol{x},\alpha,\beta)$;
  ⑤ Sample $\boldsymbol{x}^{(r)}$ according to its posterior $f(\boldsymbol{x}|\boldsymbol{y},\omega,\lambda,\sigma_n^2,p)$;
  ⑥ Set $r\leftarrow r+1$;
end

$\{\boldsymbol{x}^{(r)}\}_{r=1,\dots,S}$ that will be used to estimate the target signal/image, where S is the total number of generated samples. Using the sampled chain $\{\boldsymbol{x}^{(r)}\}_{r=1,\dots,S}$, $\widehat{\boldsymbol{x}}$ can therefore be approximated by calculating the MAP or the MMSE estimate not over $\mathbb{R}^M$, but only over a part of the sampled values which theoretically form a representative sample of the space in which $\boldsymbol{x}$ is living. The retained part must exclude, from the sampled chain, vectors generated during the burn-in period. Results below are given using the MAP estimator since it helps retreiving more precise estimates for the zero coefficients with our model.

In addition to $\boldsymbol{x}$, and using the same strategy, the proposed algorithm will also allow the estimation of σ_n^2, λ and ω based on the sampled chains $\{{\sigma_n^2}^{(r)}\}_{r=1,\dots,S}$, $\{\lambda^{(r)}\}_{r=1,\dots,S}$ and $\{\omega^{(r)}\}_{r=1,\dots,S}$, respectively. The conditional distributions used in the GS are detailed below.

4.2. Conditional distributions of $f(\boldsymbol{\theta},\boldsymbol{\Phi}|\boldsymbol{y})$

4.2.1. Sampling according to $f(\sigma_n^2|\boldsymbol{y},\boldsymbol{x})$

From the posterior in (10), straightforward calculations lead to the following conditional distribution for the noise variance

$$\sigma_n^2|\boldsymbol{x},\boldsymbol{y}\sim\mathcal{IG}\Big(\sigma_n^2|N/2,\|\boldsymbol{y}-\mathcal{K}\boldsymbol{x}\|_2^2/2\Big) \tag{11}$$

which is easy to sample.

4.2.2. Sampling according to $f(\omega|\boldsymbol{x})$

Starting from (10), we can show that the posterior of ω is the following beta distribution

$$\omega\sim\mathcal{B}(1+\|\boldsymbol{x}\|_0,1+M-\|\boldsymbol{x}\|_0) \tag{12}$$

which is also easy to sample.

4.2.3. Sampling according to $f(p|\boldsymbol{x},\omega,\lambda)$

The conditional distribution associated with the hyperparameter p can be written as

$$f(p|\boldsymbol{x})\propto\prod_{i=1}^{M}\left[(1-\omega)\delta(x_i)+\frac{\omega p}{2\lambda\Gamma(1/p)}\exp\left(-\frac{|x_i|^p}{\lambda^p}\right)\right]\times\mathcal{U}_{[0,2]}(p). \tag{13}$$

An MH step with a positive truncated Gaussian proposal can be used to sample according to (13).

4.2.4. Sampling according to $f(\lambda|\boldsymbol{x},p,\alpha,\beta)$

The conditional distribution of λ is

$$f(\lambda|\boldsymbol{x},\alpha,\beta)\propto\prod_{i=1}^{M}\left[(1-\omega)\delta(x_i)+\frac{\omega p}{2\lambda\Gamma(1/p)}\exp\left(-\frac{|x_i|^p}{\lambda^p}\right)\right]\times\frac{\beta^\alpha}{\Gamma(\alpha)}\lambda^{-\alpha-1}\exp\left(-\frac{\beta}{\lambda}\right). \tag{14}$$

We propose to use a standard MH move with a positive truncated Gaussian proposal to sample according to $f(\lambda|\boldsymbol{x},\alpha,\beta)$.

4.2.5. Sampling according to $f(\boldsymbol{x}|\boldsymbol{y},\omega,\lambda,\sigma_n^2,p)$

Because of the sophisticated prior used for $\boldsymbol{x}$, sampling according to $f(\boldsymbol{x}|\boldsymbol{y},\omega,\lambda,\sigma_n^2,p)$ is not straightforward. However, we can derive the conditional distribution of each element x_i given all the other parameters. Precisely, we decompose $\boldsymbol{x}$ onto the orthonormal basis $B=\{\boldsymbol{e}_1,\dots,\boldsymbol{e}_M\}$ such that $\boldsymbol{x}=\widetilde{\boldsymbol{x}}_{-i}+x_i\boldsymbol{e}_i$. We denote by $\widetilde{\boldsymbol{x}}_{-i}$ the vector $\boldsymbol{x}$ whose i^{th} element is set to 0 denoted as $\boldsymbol{v}_i=\boldsymbol{y}-\mathcal{K}\boldsymbol{x}_{-i}$, and $\boldsymbol{k}_i=\mathcal{K}\boldsymbol{e}_i$. Straightforward calculations lead to

$$\begin{aligned}
&f(x_i|\boldsymbol{y},\boldsymbol{x}_{-i},\omega,\lambda,p)\propto(1-\omega)\exp\left(-\frac{\|\boldsymbol{v}_i\|_2^2}{2\sigma_n^2}\right)\delta(x_i)+\\
&\omega\frac{p}{2\lambda\Gamma(1/p)}\exp\left(-\frac{|x_i|^p}{\lambda^p}-\frac{\|\boldsymbol{v}_i-\boldsymbol{k}_ix_i\|_2^2}{2\sigma_n^2}\right)\\
&\propto(1-\omega)\exp\left(-\frac{\|\boldsymbol{v}_i\|_2^2}{2\sigma_n^2}\right)\delta(x_i)+\\
&\frac{\omega p}{2\lambda\Gamma(1/p)}\exp\left(-\frac{|x_i|^p}{\lambda^p}-\frac{\boldsymbol{v}_i^\intercal\boldsymbol{v}_i+\boldsymbol{k}_i^\intercal\boldsymbol{k}_ix_i^2-2\boldsymbol{v}_i^\intercal\boldsymbol{k}_ix_i}{2\sigma_n^2}\right)\\
&\propto(1-\omega_i)\delta(x_i)+\frac{\omega_i}{Z_i(\lambda,p)}\exp\left[-U_i(x_i)\right]
\end{aligned} \tag{15}$$

where $$U_i(x_i) = \frac{|x_i|^p}{\lambda^p} + \frac{\boldsymbol{k}_i^\intercal \boldsymbol{k}_i x_i^2 - 2\boldsymbol{v}_i^\intercal \boldsymbol{k}_i x_i}{2\sigma_n^2} \quad (16)$$

and $$\omega_i = \frac{u_i}{u_i + (1-\omega)}, \; u_i = \omega Z_i(\lambda, p) \frac{p}{2\lambda\Gamma(1/p)}. \quad (17)$$

In (15) and (17), $Z_i(\lambda, p)$ is a normalizing constant such that $D(x_i|\lambda, p) = Z_i(\lambda, p) \exp\left[-U_i(x_i)\right]$ defines a probability density function. Sampling according to the conditional distribution in (15) can be performed in two steps

1- Generate z_i according to a Bernoulli distribution, i.e., set $z_i = 0$ with probability $1 - \omega_i$, and $z_i = 1$ with probability ω_i.

2- set $$\begin{cases} x_i = 0 & \text{if } z_i = 0 \\ x_i \sim D(x_i|\lambda, p) & \text{if } z_i = 1. \end{cases}$$

For efficiency reasons, we resort to ns-HMC to sample from $D(x_i|\lambda, p)$. Indeed, ns-HMC has recently been proposed [8] to make feasible the use Hamiltonian dynamics for sampling from non-differentiable log-concave distributions such as $D(x_i|\lambda, p)$. The resulting schemes are therefore more efficient than the standard MH algorithm both in terms of convergence speed and decorrelation properties [8]. As detailed in [8], to sample according to $D(x_i|\lambda, p)$, we need to calculate the proximity operator [14] of $U(x_i)$ in (16). If $p \geq 1$ so that U remains convex, following [15, Lemma 2.5(i)], the proximity operator of U can be obtained as follows

$$\forall s \in \mathbb{R}, \; \text{prox}_U(s) = \text{prox}_{\varphi/(\kappa+1)}\left[(s-u)/(\kappa+1)\right] \quad (18)$$

where $\kappa = \|\boldsymbol{k}_i\|_2^2/\sigma_n^2$, $u = -\boldsymbol{v}_i^\intercal \boldsymbol{k}_i/\sigma_n^2$ and $\varphi(\cdot) = |\cdot|^p/\lambda^p$. The proximity operator for the function φ is given by

$$\forall x \in \mathbb{R}, \; \text{prox}_\varphi(x) = \text{sign}(x)\rho \quad (19)$$

where ρ is the unique minimizer of $\rho + \frac{p}{\lambda^p}\rho^{p-1}$. Note that prox_φ reduces to the soft thresholding operator for $p = 1$, i.e.

$$\text{prox}_\varphi(x) = \text{sign}(x) \max\left\{|x| - \frac{1}{\lambda}, 0\right\}. \quad (20)$$

Other forms for different values of p can be found in [15].

5. EXPERIMENTAL VALIDATION

This section validates the proposed Bayesian sparse regularization model both for 1D and 2D signal recovery problems.

5.1. 1D sparse signal recovery

In this experiment, a 1D synthetic sparse signal $\boldsymbol{x}$ of size 100 is recovered from its distorted version $\boldsymbol{y}$ observed according to (1), where the observation operator $\mathcal{K}$ is the second order difference operator[1] and the additive noise $\boldsymbol{n}$ is Gaussian with diagonal covariance matrix $\boldsymbol{\Sigma} = \sigma_n^2 \boldsymbol{I}_M$, with $\sigma_n^2 = 1$. In addition to the reference signal, Fig. 1 shows the recovered signal obtained with the proposed algorithm (ns-HMC). For the sake of comparison, a Bernoulli-Gaussian model has been used in a Bayesian approach similar to the one in [5]. This model approaches the variational $\ell_0 + \ell_2$ regularization recently investigated in [16].

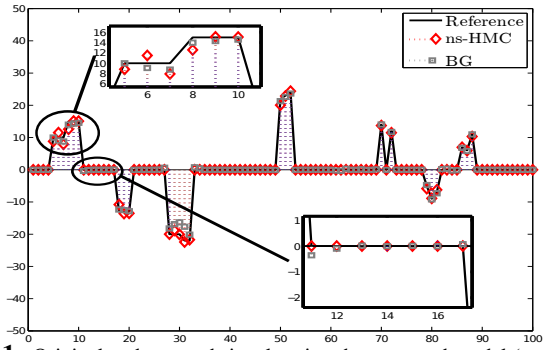

Fig. 1. Original and restored signals using the proposed model (ns-HMC) and the BG regularization.

Fig. 1 clearly shows that the proposed model gives very close results to the reference, especially where the original signal is exactly zero. From a quantitative point of view, results are evaluated in terms of signal to noise ratio (SNR). SNR values are given in Table 1 and confirm the good performance of the proposed method. To further evaluate the robustness of the proposed method, sparsity cardinalities are also given in Table 1 through the pseudo-norm $\|\widehat{\boldsymbol{x}}\|_0$. The reported values indicate that the prox-MC method allows the support of the reference to be recovered very accurately. Note that the shape parameter p has been fixed to 1 in the above simulations.

Table 1. SNR values and sparsity measures for the reference 1D signal, ns-HMC and BG.

	SNR (dB)	$\|\widehat{\boldsymbol{x}}\|_0$
Reference	-	25
Observation	-5.59	100
ns-HMC	23.41	26
BG	20.88	26

Regarding the computational cost, using a Matlab implementation on a 64-bit 3.2GHz Xeon architecture, the ns-HMC results have been obtained in 17 seconds, while the BG algorithm required only 4 seconds. The same burn-in period of 300 iterations has been considered for both algorithms. After the burn-in, 300 more iterations are collected to calculate the MAP estimators ($S = 600$).

5.2. 2D sparse image recovery

In this experiment, a 2D sparse image of size 26×26 is recovered from its distorted version $\boldsymbol{y}$ observed according to the model (1), with the same linear operator as in the previous ex-

[1] One can also cosider other linear operators such as uniform blur

Table 2. SNR values and sparsity measures for the reference image, ns-HMC and BG.

	SNR (dB)	$\|\widehat{\boldsymbol{x}}\|_0$
Reference	-	156
Observation	-8.68	676
ns-HMC	15.63	126
BG	11.89	138

periment and $\sigma_n^2 = 10^{-2}$. Under these assumptions, the original and distorted images obtained with SNR=-8.68 dB are displayed in Fig. 2. This figure also shows restored images using ns-HMC and the BG regularization. From a visual viewpoint, both methods provide similar performance except for some non-zero pixels where the dynamic is better recovered with the proposed model. This is better explained through the quantitative evaluation in Table 2 which provides the SNR and support cardinalities of the recovered signals.

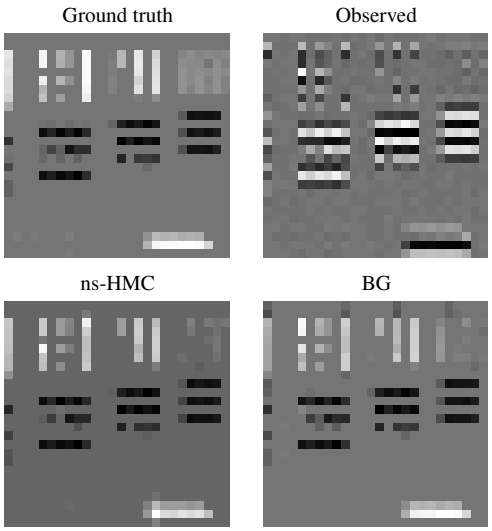

Fig. 2. Ground truth, observed and restored images using the proposed method (ns-HMC) and the BG regularization.

Indeed, our model and the BG regularization recover the same number of non-zero coefficients, while the SNR is higher using our approach. Regarding computational burden, and keeping the same burn-in period and physical architecture, the proposed algorithm takes a reasonable time with 115 seconds while the BG regularization required 48 seconds.

6. CONCLUSION

This paper proposed a new method for Bayesian sparse regularization using a Bernoulli generalized Gaussian prior promoting a two-level structured sparsity. The proposed method is based on a hierarchical Bayesian model and the recent ns-HMC technique is used for sampling the posterior of this model and to build estimators for its unknown parameters and hyperparameters. Promising results obtained with signal and image restoration experiments showed the good performance of the proposed technique. Future work will be focused on the application of this technique to large-scale real data for medical image reconstruction.

7. REFERENCES

[1] M. Vega, J. Mateos, R. Molina, and A. K. Katsaggelos, "Bayesian parameter estimation in image reconstruction from subsampled blurred observations," in *IEEE Int. Conf. on Image Process. (ICIP)*, Brussels, Belgium, Sep. 14-17 2003, pp. 969–972.

[2] L. Chaari, J.-C. Pesquet, J.-Y. Tourneret, Ph. Ciuciu, and A. Benazza-Benyahia, "A hierarchical Bayesian model for frame representation," *IEEE Trans. on Signal Process.*, vol. 18, no. 11, pp. 5560–5571, Nov. 2010.

[3] N. Dobigeon, A. O. Hero, and J.-Y. Tourneret, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2059–2070, Sept. 2009.

[4] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret, "Nonlinear spectral unmixing of hyperspectral images using Gaussian processes," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2442–2453, May 2013.

[5] L. Chaari, J.-Y. Tourneret, and H. Batatia, "Sparse Bayesian regularization using Bernoulli-Laplacian priors," in *Proc. Europ. Signal Process. Conf. (EUSIPCO)*, Marrakech, Morocco, September, 9-13 2013, pp. 1–5.

[6] W. K. Hastings, "Monte Carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[7] R. M. Neal, "MCMC using Hamiltonian dynamics," in *Handbook of Markov Chain Monte Carlo*, G. Jones X. L. Meng S. Brooks, A. Gelman, Ed., chapter 5. Chapman and Hall/CRC Press, 2010.

[8] L. Chaari, J.-Y. tourneret, C. Chaux, and H. Batatia, "A Hamiltonian Monte Carlo method for non-smooth energy sampling," 2015, http://arxiv.org/abs/1401.3988.

[9] M. Girolami and B. Calderhead, "Riemann manifold Langevin and Hamiltonian Monte Carlo methods," *J. R. Statist. Soc. B*, vol. 73, pp. 123–214, 2011.

[10] C. Robert and G. Castella, *Monte Carlo statistical methods*, Springer, New York, 2004.

[11] M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaussian density and Kullback-Leibler distance," *IEEE Trans.on Image Process.*, vol. 11, no. 2, pp. 146–158, Feb. 2002.

[12] M. W. Seeger and H. Nickisch, "Compressed sensing and Bayesian experimental design," in *Int. Conf. on Machine Learn.*, Helsinki, Finland, July 5-9 2008, pp. 912–919.

[13] L. Chaari, P. Ciuciu, S. Mériaux, and J.-Ch. Pesquet, "Spatio-temporal wavelet regularization for parallel MRI reconstruction: application to functional MRI," *Magn. Reson. Mater. in Phys., Bio. Med.*, vol. 27, no. 6, pp. 509–529, 2014.

[14] J.-J. Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société Mathématique de France*, vol. 93, pp. 273–299, 1965.

[15] C. Chaux, P. Combettes, J.-C. Pesquet, and V.R Wajs, "A variational formulation for frame-based inverse problems," *Inv. Prob.*, vol. 23, no. 4, pp. 1495–1518, Aug. 2007.

[16] E. Chouzenoux, J.-C. Pesquet, H. Talbot, and A. Jezierska, "A memory gradient algorithm for $\ell_2 - \ell_0$ regularization with application to image restoration," in *IEEE Int. Conf. on Image Process. (ICIP)*, Brussels, Belgium, Sep. 11-14 2011, pp. 2717 – 2720.