

A HIERARCHICAL SPARSITY-SMOOTHNESS BAYESIAN MODEL FOR $\ell_0 + \ell_1 + \ell_2$ REGULARIZATION

Lotfi Chaari, Hadj Batatia, Nicolas Dobigeon and Jean-Yves Tournier

University of Toulouse, IRIT - INP-ENSEEIH, France

firstname.lastname@enseeiht.fr

ABSTRACT

Sparse signal/image recovery is a challenging topic that has captured a great interest during the last decades. To address the ill-posedness of the related inverse problem, regularization is often essential by using appropriate priors that promote the sparsity of the target signal/image. In this context, $\ell_0 + \ell_1$ regularization has been widely investigated. In this paper, we introduce a new prior accounting simultaneously for both sparsity and smoothness of restored signals. We use a Bernoulli-generalized Gauss-Laplace distribution to perform $\ell_0 + \ell_1 + \ell_2$ regularization in a Bayesian framework. Our results show the potential of the proposed approach especially in restoring the non-zero coefficients of the signal/image of interest.

Index Terms— MCMC, sparsity, smoothness, hierarchical Bayesian models, restoration

1. INTRODUCTION

Sparse signal and image restoration is an open issue and has been the focus of numerous works during the last decades. More recently, and due to the emergence of the compressed sensing theory [1], sparse models have gained more interest. Indeed, recent applications generally produce large data sets that have the particularity to be highly sparse in a transformed domain. Since these data are generally modeled using ill-posed observation systems, regularization is usually required to improve the quality of the reconstructed signals/images through the use of appropriate prior information. A natural way to promote sparsity is to penalize or constrain the ℓ_0 pseudo-norm of the reconstructed signal. Unfortunately, optimizing the resulting criterion is a combinatorial problem. Suboptimal greedy algorithms, such as matching pursuit [2] or its orthogonal counterpart [3] may provide reasonable solutions to this NP-hard problem. However, despite recent advances which made the ℓ_0 -penalized problem feasible in a variational framework [4], fixing the regularization hyperparameters is still an open issue. Conversely, the solutions of the ℓ_0 -penalized problem can coincide with those of a ℓ_1 -penalized problem [5] provided that appropriate sufficient conditions are fulfilled. Based on this convex relaxation of the problem, an amount of works has been conducted to propose efficient algorithms to solve ℓ_1 -penalized problems (see for instance [6, 7]). Again, choosing appropriate values for the hyperparameters associated with the ℓ_1 -penalized (or the ℓ_1 -constrained) problems remains a difficult task [8]. These hyperparameters can for instance be estimated using empirical assessments, cross-validation or some

external empirical Bayes approaches such as [9, 10]. In this context, fully Bayesian approaches have demonstrated their flexibility to overcome these issues. More specifically, Bernoulli-based models [11–14] have been proven to be efficient tools to build sparsity promoting priors. Moreover, these Bayesian approaches allow the target signal and the regularization hyperparameters to be jointly estimated directly from the data, avoiding a difficult and painful tuning of these regularization hyperparameters.

In this paper, a hierarchical Bayesian model is proposed to enforce a smoothness-sparsity constraint by using an $\ell_0 + \ell_1 + \ell_2$ regularization. At the first level of the model, a sparsity constraint is guaranteed by using a Bernoulli process, equivalent to an ℓ_0 -penalization which favors zeroes in the reconstructed signal. At the second level of the model, the non-zero signal values are subject to a $\ell_1 + \ell_2$ penalization which allows both sparse (ℓ_1) and smooth (ℓ_2) parts of the target signal to be recovered. The use of this twofold penalization has been for instance previously advocated in [15, 16] and the resulting so-called “elastic net” model has demonstrated its efficiency to perform smooth regularization and variable selection jointly. In this paper, the $\ell_1 + \ell_2$ penalization is modeled within a Bayesian framework using a generalized-Gauss-Laplace (GGL) distribution [15]. The resulting sparsity promoting prior consists of a distribution mixture leading to a Bernoulli-GGL (BGGL) prior. To the best of our knowledge, this is the first time that an $\ell_0 + \ell_1 + \ell_2$ regularization model is fully developed. Such a regularization is still an open issue in the variational regularization literature since the inherent cost function is not convex (see Section 2), and is thus not easy to optimize with standard algorithms. Note that recent works have addressed the $\ell_0 + \ell_2$ regularization [4, 17, 18] in a variational framework. This variational regularization would be equivalent to its Bayesian counterpart in which Bernoulli-Gaussian models are used as priors [19].

Moreover, as for the variational formulation of the sparsity regularized problems, the quality of the Bayesian reconstruction drastically depends on the values of the three hyperparameters associated with the penalizing terms ℓ_0 , ℓ_1 and ℓ_2 . In this paper, following the unsupervised approaches of [12, 14], these hyperparameters are included within the Bayesian model by assigning them non-informative prior distributions. Finally, these hyperparameters and the signal of interest are jointly estimated from the data in a fully unsupervised framework.

This paper is organized as follows. Section 2 introduces the $\ell_0 + \ell_1 + \ell_2$ regularized problem we intend to solve. This problem is reformulated within a hierarchical Bayesian model detailed in Section 3. Section 4 presents a Gibbs sampler which can be used to generate samples asymptotically distributed according to the posterior of this Bayesian model and thus to compute Bayesian estimators of the unknown model parameters. Finally, we validate the proposed method in Section 5 before concluding in Section 6.

Part of this work has been conducted within the image processing trimester of the CIMI Labex. This work has also been performed in the frame of the Hypanema ANR Project (ANR-12-BS03-003) and the STIC AmSud project DynBrain.

2. HIERARCHICAL SPARSE REGULARIZATION

2.1. Problem formulation

In this paper we focus on real-valued digital signals of length M as elements of the Euclidean space \mathbb{R}^M endowed with the usual scalar product and norm denoted as $\langle \cdot | \cdot \rangle$ and $\| \cdot \|$, respectively. Let $\mathbf{x} \in \mathbb{R}^M$ be our target signal, which is measured by $\mathbf{y} \in \mathbb{R}^N$ through a distortion linear operator \mathcal{H} . The resulting observation model can be written

$$\mathbf{y} = \mathcal{H}\mathbf{x} + \mathbf{n} \quad (1)$$

where \mathbf{n} is an additive noise often considered as white Gaussian with covariance matrix $\sigma_n^2 \mathbf{I}_N$. Since we generally have $M \gg N$, the inverse problem in Eq. (1) is ill-posed. In this situation, its direct inversion yields distorted solutions presenting reconstruction artifacts that possibly interfere with the useful signal. This is the case in a number of recent applications in the field of signal and image processing, such as in parallel MRI [15, 20] and positron emission tomography (PET) [21]. This paper focuses on such kind of problems where the target signal/image \mathbf{x} is sparse. Consequently, we propose here to adopt a sparse regularization strategy for estimating the unknown signal/image \mathbf{x} . More precisely, the signal of interest \mathbf{x} is assumed to contain both zero and non-zero coefficients. Moreover, the non-zero coefficients are decomposed into sparse and smooth groups. Under these assumptions, we propose to investigate an $\ell_0 + \ell_1 + \ell_2$ regularization to tackle a hierarchical sparsity model.

2.2. Variational formulation

Performing an $\ell_0 + \ell_1 + \ell_2$ regularization consists of solving the following minimization problem

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{R}^M} \frac{1}{\sigma_n^2} \|\mathbf{y} - \mathcal{H}\mathbf{x}\|_2^2 + \lambda_0 \|\mathbf{x}\|_0 + \lambda_1 \|\mathbf{x}\|_1 + \lambda_2 \|\mathbf{x}\|_2^2 \quad (2)$$

where λ_0 , λ_1 and λ_2 are regularization parameters that have to be estimated. In Eq. (2) $\|\cdot\|_0$, $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the ℓ_0 pseudo-norm and the ℓ_1 and ℓ_2 norms, respectively. To the best of our knowledge, mainly because the problem in Eq. (2) is not convex, it cannot be solved using standard optimization algorithms. For this reason, we propose to define a new hierarchical Bayesian model with appropriate prior distributions allowing Eq. (2) to be solved in a fully Bayesian framework.

3. BAYESIAN MODEL FOR HIERARCHICAL SPARSE REGULARIZATION

In a Bayesian framework, \mathbf{y} and \mathbf{x} are assumed to be realizations of random vectors \mathbf{Y} and \mathbf{X} . We then aim at characterizing the probability distribution of $\mathbf{X}|\mathbf{Y}$, by considering some parametric probabilistic model and by estimating the associated parameters and hyperparameters. In the following, we derive the hierarchical Bayesian model proposed for the sparse regularization problem of Eq. (2).

3.1. Likelihood

Under the assumption of additive white Gaussian noise of variance σ_n^2 , the likelihood can be expressed as follows

$$f(\mathbf{y}|\mathbf{x}, \sigma_n^2) = \left(\frac{1}{2\pi\sigma_n^2} \right)^{N/2} \exp \left(-\frac{\|\mathbf{y} - \mathcal{H}\mathbf{x}\|_2^2}{2\sigma_n^2} \right). \quad (3)$$

3.2. Priors

Let us denote by $\boldsymbol{\theta} = (\mathbf{x}, \sigma_n^2)^\top$ the unknown parameter vector to be estimated. For the noise variance σ_n^2 , we use a non-informative prior that guarantees the positivity of this parameter. More precisely, σ_n^2 is assigned a Jeffreys' prior distribution defined as (see [22] for motivations)

$$f(\sigma_n^2) \propto \frac{1}{\sigma_n^2} \mathbf{1}_{\mathbb{R}^+}(\sigma_n^2) \quad (4)$$

where $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ is the indicator function on \mathbb{R}^+ , i.e., $\mathbf{1}_{\mathbb{R}^+}(\xi) = 1$ if $\xi \in \mathbb{R}^+$ and 0 otherwise.

In order to promote the sparsity of the target signal, one can choose a Bernoulli-Gaussian (BG) [11, 23], a Bernoulli-exponential [12] (for positive real-valued signals), or a Bernoulli-Laplace (BL) [14] prior for every x_i ($i = 1, \dots, M$). To promote hierarchical sparsity and further distinguish smooth and sparse coefficients for the non-zero part of the target signal, we use here a Bernoulli-Generalized Gauss-Laplace (BGGL) distribution for every x_i

$$f(x_i|\boldsymbol{\Phi}) = (1 - \omega)\delta(x_i) + \omega \mathcal{GGL}(x_i|\alpha, \beta) \quad (5)$$

with $\boldsymbol{\Phi} = (\omega, \alpha, \beta)^\top$ is the vector of unknown hyperparameters and

$$\mathcal{GGL}(x_i|\alpha, \beta) = \frac{\sqrt{\frac{\beta}{2\pi}}}{\text{erfc}(\frac{\alpha}{\sqrt{2\beta}})} \exp \left[-\left(\alpha|x_i| + \frac{\beta}{2}x_i^2 + \frac{\alpha^2}{2\beta} \right) \right] \quad (6)$$

where $\text{erfc}(\cdot)$ denotes the complementary error function

$$\text{erfc}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

In (5), $\delta(\cdot)$ is the Dirac delta function and $\omega \in [0, 1]$ represents the prior probability of having a non-zero signal component. We use a generalized Gauss-Laplace (GGL) model as a prior for the non-zero coefficients x_i in order to account for both smoothness and sparsity constraints for the x_i 's. Using the BGGL model for x_1, \dots, x_m and assuming these variables are a priori independent, the joint prior distribution for the full signal vector \mathbf{x} is

$$\begin{aligned} f(\mathbf{x}|\boldsymbol{\Phi}) &= \prod_{i=1}^M f(x_i|\boldsymbol{\Phi}) \\ &= \prod_{i=1}^M \left\{ (1 - \omega)\delta(x_i) + \omega \mathcal{GGL}(x_i|\alpha, \beta) \right\}. \end{aligned} \quad (7)$$

The resulting BGGL model consists of a two-level sparsity promoting prior, and also accounts for possible smoothness properties of the target signal. The first level of sparsity is guaranteed thanks to the Bernoulli model and the Dirac delta function. The second level of sparsity is ensured by the GGL distribution. This prior distribution generalizes several standard regularizations used in the statistics and signal/image processing literatures. Indeed, for $\omega = 1$, the BGGL model is reduced to a GGL, which can be interpreted as the Bayesian counterpart of the elastic net model introduced in [16] and successfully used in [15] for parallel MRI reconstruction. Moreover, for $\alpha = 0$, the GGL distribution reduces to a Gaussian distribution, inducing a standard smoothing ℓ_2 -regularization, which results in a Bernoulli-Gaussian prior for x_i , for instance used in [11, 19, 23, 24]. Finally, for $\beta = 0$, the GGL distribution boils down to a Laplace prior distribution, i.e., a sparsity inducing ℓ_1 -regularization advocated in [25] within a Bayesian framework. In this later case, the prior for the signal component x_i is a Bernoulli-Laplace process introduced in [26] and successfully used in [14].

3.3. Hyperparameter priors

In the variational formulation of the considered $\ell_0 + \ell_1 + \ell_2$ regularization in (2), the levels of the various penalizations are adjusted via the hyperparameters λ_0 , λ_1 and λ_2 for a given noise variance σ_n^2 . Choosing appropriate values for these regularization hyperparameters is a challenging issue that is usually addressed using empirical approaches, e.g., cross-validation or subjective inspections of multiple results. In the Bayesian formulation of the $\ell_0 + \ell_1 + \ell_2$ regularization, similar roles are played by the hyperparameters ω , α and β . It can be easily observed that the quality of the Bayesian reconstruction also drastically depends on these hyperparameters that need to be properly chosen. In absence of additional prior knowledge regarding the signal to be reconstructed (e.g., proportion and mean of non-zero signal components), these hyperparameters can be included within the Bayesian model by assigning them prior distributions. Consequently, these hyperparameters can be directly estimated from the data, in a fully unsupervised framework. It is the strategy considered in this paper and the hyperparameter prior distributions are detailed below.

Individual non-informative priors are used for the hyperparameters ω , α and β which are assumed to be a priori independent. First, to reflect the absence of prior knowledge regarding the proportion of non-zero signal components, a uniform distribution on the simplex $[0, 1]$ can be used for ω , i.e., $\omega \sim \mathcal{U}_{[0,1]}$. Since the parameters α and β are real-positive, a commonly used prior in this situation is a conjugate inverse-gamma (IG) distribution $\mathcal{IG}(\alpha|a, b)$ defined as

$$\mathcal{IG}(\alpha|a, b) = \frac{b^a}{\Gamma(a)} \alpha^{-a-1} \exp\left(-\frac{b}{\alpha}\right) \quad (8)$$

where $\Gamma(\cdot)$ is the gamma function, and a and b are hyperparameters to be fixed to obtain vague hyper-priors (in the experiments reported in Section 5, these hyperparameters have been set to $a = b = 10^{-3}$ both for α and β).

4. RESOLUTION SCHEME

Using a maximum a posteriori (MAP) strategy, the model parameter vector $\theta = (\mathbf{x}, \sigma_n^2)^\top$ is estimated based on the likelihood $f(\mathbf{y}|\theta)$, the priors $f(\theta|\Phi)$ and hyperpriors $f(\Phi)$ introduced in the previous section. According to the Bayes' paradigm, the joint posterior distribution of $\{\theta, \Phi\}$ can be expressed as

$$f(\theta, \Phi|\mathbf{y}) \propto f(\mathbf{y}|\theta)f(\theta|\Phi)f(\Phi) \quad (9)$$

$$\propto f(\mathbf{y}|\mathbf{x}, \sigma_n^2)f(\mathbf{x}|\omega, \alpha, \beta)f(\sigma_n^2)f(\omega|\mathbf{x})f(\beta|\mathbf{x})f(\alpha|\mathbf{x}).$$

We propose here to resort to a Gibbs sampler [22] that iteratively samples according to the conditional posteriors $f(\mathbf{x}|\mathbf{y}, \omega, \alpha, \beta, \sigma_n^2)$, $f(\sigma_n^2|\mathbf{y}, \mathbf{x})$, $f(\omega|\mathbf{x})$, $f(\alpha|\mathbf{x})$ and $f(\beta|\mathbf{x})$. Calculations similar to [12, 14] show that the posteriors for σ_n^2 and ω are simply inverse gamma and beta distributions, respectively

$$\sigma_n^2|\mathbf{x}, \mathbf{y} \sim \mathcal{IG}\left(\sigma_n^2|N/2, \|\mathbf{y} - \mathcal{H}\mathbf{x}\|^2/2\right)$$

$$\omega \sim \mathcal{B}(1 + \|\mathbf{x}\|_0, 1 + M - \|\mathbf{x}\|_0). \quad (10)$$

Unfortunately, no closed-form expression can be obtained for the conditional distributions of $\alpha|\mathbf{x}, \omega, \beta$ and $\beta|\mathbf{x}, \omega, \alpha$. Metropolis-Hastings moves with positively truncated Gaussian proposals are therefore used to sample according to $f(\beta|\mathbf{x}, \omega, \alpha) = f(\beta|\mathbf{x})$ and $f(\alpha|\mathbf{x}, \omega, \beta) = f(\alpha|\mathbf{x})$.

The distribution of x_i conditionally to the rest of the signal \mathbf{x}_{-i} and the other model parameters is easy to be derived. Straightforward computations lead to the following result

$$f(x_i|\mathbf{y}, \mathbf{x}_{-i}, \omega, \alpha, \beta) = \omega_{1,i}\delta(x_i) + \omega_{2,i}\mathcal{N}^+(\mu_{i+}, \sigma_i^2) \quad (11)$$

$$+ \omega_{3,i}\mathcal{N}^-(\mu_{i-}, \sigma_i^2)$$

where \mathcal{N}^+ and \mathcal{N}^- denote the truncated Gaussian distribution on \mathbb{R}^+ and \mathbb{R}^- , respectively. Akin to [12, 14], we first decompose \mathbf{x} on the orthonormal basis $B = \{\mathbf{e}_1, \dots, \mathbf{e}_M\}$ such that $\mathbf{x} = \tilde{\mathbf{x}}_{-i} + x_i\mathbf{e}_i$, where $\tilde{\mathbf{x}}_{-i}$ is the signal vector \mathbf{x} whose i^{th} element is set to 0. Denoting $\mathbf{v}_i = \mathbf{y} - \mathcal{H}\mathbf{x}_{-i}$ and $\mathbf{h}_i = \mathcal{H}\mathbf{e}_i$, the weights $(\omega_{l,i})_{1 \leq l \leq 3}$ are given by

$$\omega_{l,i} = \frac{u_{l,i}}{\sum_{l=1}^3 u_{l,i}} \quad (12)$$

where $u_{1,i} = 1 - \omega$

$$u_{2,i} = \omega \frac{\sqrt{\frac{\beta}{2\pi}} e^{-\frac{\alpha^2}{2\beta} + \frac{\mu_{i+}^2}{2\sigma_i^2}}}{\text{erfc}\left(\frac{\alpha}{\sqrt{2\beta}}\right)} \sqrt{2\pi\sigma_i^2} C(\mu_{i+}, \sigma_i^2)$$

$$u_{3,i} = \omega \frac{\sqrt{\frac{\beta}{2\pi}} e^{-\frac{\alpha^2}{2\beta} + \frac{\mu_{i-}^2}{2\sigma_i^2}}}{\text{erfc}\left(\frac{\alpha}{\sqrt{2\beta}}\right)} \sqrt{2\pi\sigma_i^2} C(\mu_{i-}, \sigma_i^2) \quad (13)$$

and

$$\sigma_i^2 = \frac{\sigma_n^2}{\|\mathbf{h}_i\|^2 + \beta\sigma_n^2}$$

$$\mu_{i+} = \sigma_i^2 \left(\frac{\mathbf{h}_i^\top \mathbf{v}_i}{\sigma_n^2} - \alpha \right),$$

$$\mu_{i-} = \sigma_i^2 \left(\frac{\mathbf{h}_i^\top \mathbf{v}_i}{\sigma_n^2} + \alpha \right)$$

$$C(\mu, \sigma^2) = \sqrt{\frac{\sigma^2\pi}{2}} \left[1 + \text{erf}\left(\frac{\mu}{\sigma^2}\right) \right]. \quad (14)$$

The resulting sampler is summarized in Algorithm 1. After convergence, Algorithm 1 provides samples that are asymptotically distributed according to the full posterior of interest. These samples can be used to compute a MAP estimator in order to get $\hat{\mathbf{x}}$, as in [24]. Moreover, the proposed algorithm also allows $\hat{\sigma}_n^2$, $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\omega}$ to be computed.

Algorithm 1 Gibbs sampler.

Initialize with some $\mathbf{x}^{(0)}$.
repeat
 Sample σ_n^2 according to $f(\sigma_n^2|\mathbf{y}, \mathbf{x})$.
 Sample α according to $f(\alpha|\mathbf{x}, a, b)$.
 Sample β according to $f(\beta|\mathbf{x}, a, b)$.
 Sample ω according to $f(\omega|\mathbf{x})$.
 for $i = 1$ to M **do**
 Sample x_i according to Eq. (11).
 end for
until convergence

5. EXPERIMENTAL VALIDATION

The conducted experiment addresses a 1D signal recovery problem based on realistic simulated data. A sparse signal \mathbf{x} of size 100 is

recovered from its distorted observation \mathbf{y} according to the observation model in Eq. (1). Distortion is due to the application of the second order difference operator (\mathcal{H}) in addition to a white Gaussian noise of variance $\sigma_n^2 = 0.5$. The results are compared to other regularization techniques based on visual inspections as well as output signal-to-noise ratios given by

$$\text{SNR} = 20 \log_{10} \frac{\|\mathbf{x}^0\|}{\|\mathbf{x}^0 - \hat{\mathbf{x}}\|}$$

where \mathbf{x}^0 and $\hat{\mathbf{x}}$ are the reference and estimated signals, respectively. For the sake of comparison, the sparse regularization scheme (BL) of [14] is applied in addition to the proposed method (BGGL). Moreover, results using the orthogonal matching pursuit (OMP) algorithm are also provided. Fig. 1 illustrates the ground truth and the reconstructed signals using the BL and BGGL models, in addition to the OMP algorithm. Visual inspection of restored signals show very similar performance for the BL and BGGL models. Indeed, these two methods recover an accurate sparsity support (non-zero coefficients): $\|\hat{\mathbf{x}}_{\text{BL}}\|_0 = \|\hat{\mathbf{x}}_{\text{BGGL}}\|_0 = 29$ and $\|\mathbf{x}^0\|_0 = 28$. As regards OMP restoration, visual inspection show that non-zero coefficients are not well recovered even if the sparsity support is quite accurately recovered ($\|\hat{\mathbf{x}}_{\text{OMP}}\|_0 = 24$).

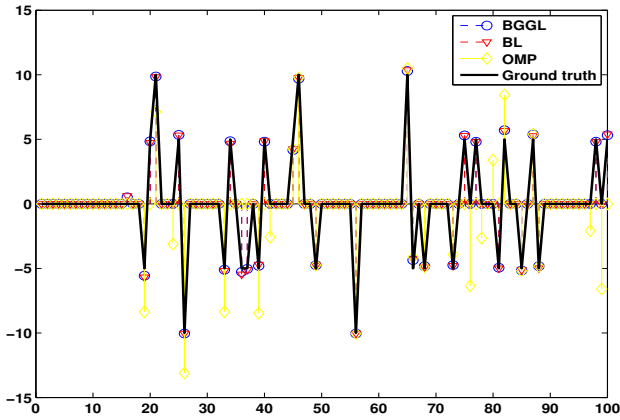


Fig. 1. Original and restored signals using the proposed method (BGGL) and BL regularization in [14].

To quantitatively assess the reconstruction quality, output SNR are computed for the ℓ_1 - and $\ell_0 + \ell_1 + \ell_2$ -regularized restoration methods: $\text{SNR}_{\text{BGGL}} = 25.61$ dB and $\text{SNR}_{\text{BL}} = 24.56$ dB. Since the two methods recover the same sparsity support, this performance gain is due to a better estimation of the non-zero coefficients with the proposed method. Indeed, the flexibility of the GGL distribution allows us to better model both sparsity and smoothness of non-zero coefficients, leading to better restoration results.

Moreover, since the proposed Gibbs algorithm generates samples asymptotically distributed according to the joint posterior distribution (9), the conditional posterior distributions for the noise variance σ_n^2 and the regularization hyperparameters ω , α and β can also be estimated. These estimated posteriors are depicted in Fig. 2 and the estimated parameters are reported below each plot.

To further assess the restoration performance of the proposed method, 50 Monte Carlo simulations have been conducted with different acquisition noise levels ($\sigma_n^2 \in \{0.5, 1, 1.5, 2, 2.5, 3\}$). The average SNR values computed using the 50 Monte Carlo runs are

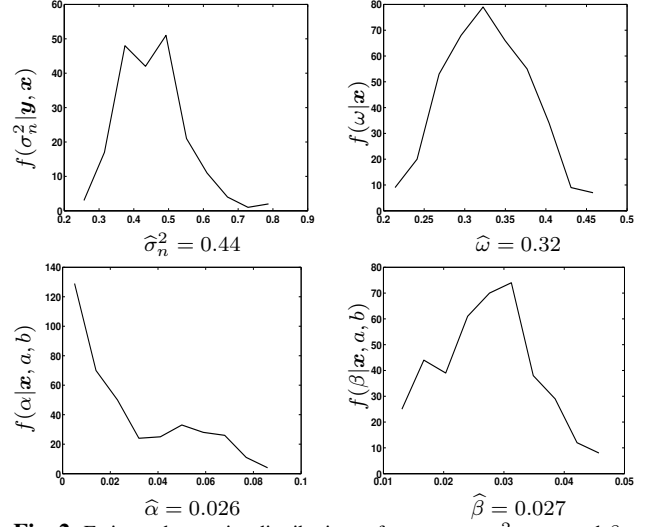


Fig. 2. Estimated posterior distributions of parameters σ_n^2 , ω , α and β .

depicted in Fig. 3 for all noise levels. The observed SNR values confirm the ability of the proposed BGGL model to better restore non-zero signal coefficients. Fig. 3 also shows that the proposed method may be more efficient at high noise levels. As expected, OMP gives lower performance compared to the two other methods.

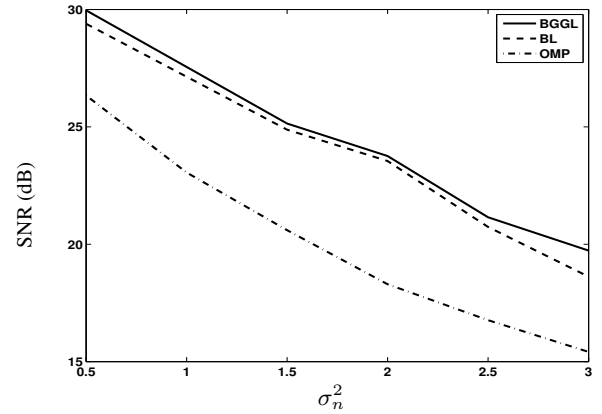


Fig. 3. Output SNR w.r.t. input noise variance σ_n^2 . Mean values are calculated based on 50 Monte Carlo simulations for every noise level.

6. CONCLUSION

In this contribution, we proposed a new method for hierarchical sparse-smooth regularization involving $\ell_0 + \ell_1 + \ell_2$ penalization. The proposed method relied on a hierarchical Bayesian model with appropriate priors for the model parameters and hyperparameters, the latter being automatically estimated from the data. Promising results showed the potential of the proposed approach. Future work will investigate the application of this method to real magnetic resonance imaging (MRI) and electroencephalography (EEG) signal recovery.

7. REFERENCES

- [1] D.L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [2] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
- [3] Y. Pati, R. Rezaifar, and P. Krishnaprasad, "Orthogonal matching pursuit : recursive function approximation with application to wavelet decomposition," in *Proc. IEEE Asilomar Conf. Signals, Systems, Computers*, 1993, pp. 40–44.
- [4] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot., "A majorize-minimize subspace approach for l2-l0 image regularization," *SIAM Journal on Imaging Science*, vol. 6, no. 1, pp. 563–591, 2013.
- [5] D. Donoho and X. Huo, "Uncertainty principles and ideal atomic decompositions," *IEEE Trans. Inf. Theory*, vol. 47, no. 10, pp. 2845–2862, Nov. 2001.
- [6] Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
- [7] E. Candes and J. Romberg, " ℓ_1 -magic: Recovery of sparse signals via convex programming," Tech. Rep., California Inst. Technol., Pasadena, CA, Oct. 2005.
- [8] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, June 2010.
- [9] M. Ting, R. Raich, and A. O. Hero, "Sparse image reconstruction for molecular imaging," *IEEE Trans. Image Process.*, vol. 18, no. 6, pp. 1215–1227, Jan. 2009.
- [10] L. Chaari, J.-C. Pesquet, J.-Y. Tournet, 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.
- [11] J. Idier and Y. Goussard, "Stack algorithm for recursive deconvolution of Bernoulli-Gaussian processes," *IEEE Trans. Signal Process.*, vol. 28, no. 9, pp. 67–79, Sep. 1990.
- [12] N. Dobigeon, A. O. Hero, and J.-Y. Tournet, "Hierarchical Bayesian sparse image reconstruction with application to MRFM," *IEEE Trans. Image Process.*, vol. 18, no. 9, pp. 2059–2070, Sept. 2009.
- [13] C. Soussen, J. Idier, D. Brie, and J. Duan, "From Bernoulli-Gaussian deconvolution to sparse signal restoration," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4572–4584, Oct. 2011.
- [14] L. Chaari, J.-Y. Tournet, and H. Batatia, "Sparse Bayesian regularization using Bernoulli-Laplacian priors," in *Proc. European Signal Processing Conference (EUSIPCO)*, Marrakech, Morocco, September, 9-13 2013.
- [15] L. Chaari, J.-C. Pesquet, A. Benazza-Benyahia, and P. Ciuciu, "A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging," *Med. Image Anal.*, vol. 15, no. 2, pp. 185–201, Nov. 2011.
- [16] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. Roy. Stat. Soc. Ser. B*, vol. 67, pp. 301–320, 2005.
- [17] J. Duan, C. Soussen, D. Brie, and J. Idier, "A continuation approach to estimate a solution path of mixed l2-l0 minimization problems," in *Proc. Signal Processing with Adaptive Sparse Structured Representations Workshop (SPARS)*, Saint-Malo, France, April 6-9 2009.
- [18] 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 *Proc. IEEE International Conference on Image Processing (ICIP)*, Brussels, Belgium, Sep. 11-14 2011, pp. 2717 – 2720.
- [19] M. Lavielle, "Bayesian deconvolution of Bernoulli-Gaussian processes," *Signal Process.*, vol. 33, no. 1, pp. 67–79, Jul. 2009.
- [20] L. Chaari, P. Ciuciu, S. Mériaux, and J.-C. Pesquet, "Spatio-temporal wavelet regularization for parallel mri reconstruction: application to functional mri," *Magnetic Resonance Materials in Physics, Biology and Medicine (MAGMA)*, 2014, DOI: 10.1007/s10334-014-0436-5.
- [21] N. Pustelnik, C. Chaux, J.-C. Pesquet, and C. Comtat, "Parallel algorithm and hybrid regularization for dynamic PET reconstruction," in *Proc. IEEE Medical Imaging Conference*, Knoxville, Tennessee, 30, Oct - 6, Nov. 2010.
- [22] C. Robert and G. Casella, *Monte Carlo statistical methods*, Springer, New York, 2004.
- [23] A. Doucet and P. Duvaut, "Bayesian estimation of state-space models applied to deconvolution of Bernoulli-Gaussian processes," *Signal Processing*, vol. 57, no. 2, pp. 147–161, Mar. 1997.
- [24] N. Dobigeon and J.-Y. Tournet, "Bayesian orthogonal component analysis for sparse representation," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2675–2685, May 2010.
- [25] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, Jan. 2010.
- [26] I. M. Johnstone and B. W. Silverman, "Needles and straw in haystacks: empirical Bayes estimates of possibly sparse sequences," *Ann. Stat.*, vol. 32, no. 4, pp. 1594–1649, 2004.