SPARSE BAYESIAN REGULARIZATION USING BERNOULLI-LAPLACIAN PRIORS

Lotfi Chaari, Jean-Yves Tourneret and Hadj Batatia

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

ABSTRACT

Sparse regularization has been receiving an increasing interest in the literature. Two main difficulties are encountered when performing sparse regularization. The first one is how to fix the parameters involved in the regularization algorithm. The second one is to optimize the inherent cost function that is generally non differentiable, and may also be non-convex if one uses for instance an ℓ_0 penalization. In this paper, we handle these two problems jointly and propose a novel algorithm for sparse Bayesian regularization. An interesting property of this algorithm is the possibility of estimating the regularization parameters from the data. Simulation performed with 1D and 2D restoration problems show the very promising potential of the proposed approach. An application to the reconstruction of electroencephalographic signals is finally investigated.

Index Terms**—** Sparse Bayesian restoration, MCMC methods, parameter estimation, $\ell_0 + \ell_1$ regularization

1. INTRODUCTION

Sparse signal and image restoration has been of increasing interest during the last decades. It has found several fields of applications such as remote sensing [1] and medical image reconstruction [2], especially after the emergence of the compressed sensing theory [3]. Data volumes are continuously increasing, e.g., in recent applications where imaging systems may deliver multidimensional signals up to 4D (3D + time for instance) [4]. Accounting for sparsity properties while reconstructing such signals is therefore of great interest. Since observation systems are generally ill-posed, regularization is usually requested to improve quality of reconstructed signals. Regularization consists of constraining the search space through some prior information that we inject in the model to stabilize the inverse problem. Such prior information generally involves additional parameters that have to be tuned. Fixing these parameters is actually an open issue, since they deeply impact the target solution quality. One can fix them by cross-validation or by using some Bayesian methods such as [5]. However, the estimation of these parameters still relies on an external algorithm, which makes the regularization problem not fully automatic. After estimating these parameters, the cost function associated with the restoration problem has to be optimized. In the recent literature, variational approaches have been widely used to solve the corresponding optimization problem. They generally rely on some iterative optimization algorithm since most of the edge-preserving penalizations yield to non-differentiable cost functions whose extrema have no closed-form expression. In the recent image processing literature [6, 4, 7], proximal algorithms have been notably investigated such as forward-backward (FB) [7] and parallel proximal algorithms (PPXA) [8]. However, these algorithms can only handle convex cost functions such as those involving ℓ_1 or $\ell_1 + \ell_2$ penalizations [9]. In other words, regularization problems involving an ℓ_0 pseudo-norm penalization cannot be solved using these algorithms.

In this paper, we propose a novel approach to handle this ℓ_0 regularization problem in a Bayesian framework. The main advantage of our Bayesian restoration method is that the regularization parameters can be estimated from the data, allowing the sparsity level of the target signal/image to be determined. The proposed method is fully automatic and does not need any user interaction. Indeed, Bayesian method have been widely promoted during the last decades due to their flexibility in handling complicated models, especially when model hyperparameters are not easy to set. For instance, such investigation is clearly stated in the biomedical [10] and hyperspectral imaging [11] fields.

In recent literature, different priors have been used for sparse regularization such as Bernoulli-Gaussian [12] or Bernoulli-exponential priors [13]. The Bernoulli-Laplacian prior investigated in this work provides sparser solutions when compared to the Bernoulli-Gaussian one. This increased sparsity is due to the Laplacian term, while accounting for both positive and negative signal values in contrast to the Bernoulli-exponential prior.

The paper is organized as follows. In Section 2, the proposed model and inference scheme are introduced. Experimental validations are presented in Section 3. Conclusions and future work are reported in Section 4.

The authors would like to thank Yoann Altmann for helping them in the Matlab implementation, as well as Prof. Alexandre Gramfort for his help in EEG experiments.

2. BAYESIAN RESTORATION METHOD

2.1. Problem formulation

Let $\boldsymbol{x} \in \mathbb{R}^M$ be our target signal, which is measured by $\boldsymbol{y} \in \mathbb{R}^P$ through a linear observation operator $\mathcal{H}$. Accounting for the additive acquisition noise, the observation model we are interested in can be written as

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

Without loss of generality, we only focus here on multiplicative linear operators and additive noise. When $\mathcal{H}$ is ill-conditioned, the above inverse problem may be ill-posed. We propose here to adopt a sparse regularization strategy for estimating the unknown signal/image $\boldsymbol{x}$ via a Bayesian framework. In the following section, the hierarchical Bayesian model used for regularization is detailed.

2.2. Hierarchical Bayesian model

2.2.1. Likelihood

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

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

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

2.2.2. Priors

In our model, the unknown parameter vector to be estimated is denoted by $\boldsymbol{\theta} = \{\boldsymbol{x}, \sigma_n^2\}$. In what follows, we introduce the prior distributions to be used for these two parameters.

Prior for $\boldsymbol{x}$

In order to promote the sparsity of the target signal, we adopt here a Bernoulli-Laplace prior for every x_i ($i = 1, \dots, M$), given by:

$$f(x_i|\omega, \lambda) = (1 - \omega)\delta(x_i) + \frac{\omega}{2\lambda} \exp\left(-\frac{|x_i|}{\lambda}\right) \tag{3}$$

where $\delta(.)$ is the Dirac delta function, $\lambda > 0$ is the parameter of the Laplace distribution, and w is a weight belonging to $[0, 1]$. This prior is similar to the one used in [13]. However, since we are considering signals with both positive and negative coefficients, the exponential distribution is replaced here by the Laplace one. Assuming the coefficients x_i *a priori* independent, the prior distribution for $\boldsymbol{x}$ writes:

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

Prior for σ_n^2

To guarantee the positivity of σ_n^2 and keep this prior non-informative, we use here a Jeffrey's prior defined as:

$$f(\sigma_n^2) \propto \frac{1}{\sigma_n^2} 1_{\mathbb{R}^+}(\sigma_n^2) \tag{5}$$

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. Motivations for using this kind of prior for the noise variance can be found in standard textbooks on Bayesian inference such as [14].

2.2.3. Hyperparameter priors

Hyperprior for ω

For simplicity reasons, and to use non-informative priors, we use here a uniform distribution on the simplex $[0, 1]$ for ω, i.e., $\omega \sim \mathcal{U}_{[0,1]}$.

Hyperprior for λ

Since λ is real positive, a conjugate inverse-gamma (IG) distribution has been used as a hyper-prior:

$$f(\lambda|\alpha, \beta) = \mathcal{IG}(\lambda|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{-\alpha-1} \exp\left(-\frac{\beta}{\lambda}\right) \tag{6}$$

where $\Gamma(.)$ is the gamma function, and α and β are hyperparameters to be fixed (in our experiments these hyperparameters have been set to $\alpha = \beta = 10^{-3}$).

2.3. Bayesian inference scheme

We adopt here a maximum a posteriori (MAP) strategy in order to estimate the model parameter vector $\boldsymbol{\theta}$ based on the likelihood, the priors and hyperpriors introduced hereabove. If we denote by $\boldsymbol{\Phi} = \{\lambda, \omega\}$ the model hyperparameters, the joint posterior distribution of $\{\boldsymbol{\theta}, \boldsymbol{\Phi}\}$ can be expressed as

$$f(\boldsymbol{\theta}, \boldsymbol{\Phi}|\boldsymbol{y}, \alpha, \beta) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\boldsymbol{\Phi}) f(\boldsymbol{\Phi}|\alpha, \beta). \tag{7}$$

Akin to [13], we propose here to use a Gibbs algorithm [14] that iteratively samples according to the conditional posteriors $f(\boldsymbol{x}|\boldsymbol{y}, \omega, \lambda, \sigma_n^2)$, $f(\sigma_n^2|\boldsymbol{y}, \boldsymbol{x})$, $f(\lambda|\boldsymbol{x}, \alpha, \beta)$ and $f(\omega|\boldsymbol{x})$.

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

Straightforward calculations combining the likelihood and the prior distribution of σ_n^2 lead to the following posterior:

$$\sigma_n^2|\boldsymbol{x}, \boldsymbol{y} \sim \mathcal{IG}\Big(\sigma_n^2|P/2, ||\boldsymbol{y} - \mathcal{H}\boldsymbol{x}||^2/2\Big) \tag{8}$$

which is easy to sample.

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

Calculations similar to [13] lead to the following posterior which is also simple to sample:

$$\lambda|\boldsymbol{x},\alpha,\beta \sim \mathcal{IG}\Big(\lambda|\alpha+||\boldsymbol{x}||_0,\beta+||\boldsymbol{x}||_1\Big) \tag{9}$$

where $||.||_0$ denotes the ℓ_0 pseudo-norm calculating the number of non-zero coefficients, and $||.||_0$ is the ℓ_1 norm defined as $\|\boldsymbol{x}\|_1 = \sum_{i=1}^{M} |x_i|$

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

Straightforward calculations show that the posterior of ω is a beta distribution:

$$\omega \sim \mathcal{B}(1+||\boldsymbol{x}||_0, 1+M-||\boldsymbol{x}||_0) \tag{10}$$

according to which it is easy to sample.

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

We can easily derive the posterior distribution of each signal element x_i conditionally to the rest of the signal. Straightforward computations lead to lead to the following form of this posterior

$$\begin{aligned} f(x_i|\boldsymbol{y},\boldsymbol{x}_{-i},\omega,\lambda) =& \omega_{1,i}\delta(x_i) \\ &+ \omega_{2,i}\mathcal{N}^+(\mu_i^+,\sigma_i^2) + \omega_{3,i}\mathcal{N}^-(\mu_i^-,\sigma_i^2) \end{aligned} \tag{11}$$

where $\mathcal{N}^+$ and $\mathcal{N}^-$ denote the truncated Gaussian distribution on $\mathbb{R}^+$ and $\mathbb{R}^-$, respectively. By decomposing $\boldsymbol{x}$ on the orthonormal basis $B = \{\boldsymbol{e}_1,\ldots,\boldsymbol{e}_M\}$ such that $\boldsymbol{x} = \widetilde{\boldsymbol{x}}_{-i} + x_i\boldsymbol{e}_i$ where $\widetilde{\boldsymbol{x}}_{-i}$ is nothing but $\boldsymbol{x}$ whose i^{th} element is set to 0, and denoting $\boldsymbol{v}_i = \boldsymbol{y} - \mathcal{H}\boldsymbol{x}_{-i}$ and $\boldsymbol{h}_i = \mathcal{H}\boldsymbol{e}_i$, the weights $(\omega_{l,i})_{1\le l\le 3}$ are given by

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

where

$$\begin{aligned} u_{1,i} =& 1-\omega \\ u_{2,i} =& \frac{\omega}{2\lambda}\exp\left(\frac{\mu_{i+}^2}{2\sigma_i^2}\right)\sqrt{2\pi\sigma_i^2}C(\mu_{i+},\sigma_i^2) \\ u_{3,i} =& \frac{\omega}{2\lambda}\exp\left(\frac{\mu_{i-}^2}{2\sigma_i^2}\right)\sqrt{2\pi\sigma_i^2}C(\mu_{i-},\sigma_i^2) \end{aligned} \tag{13}$$

and

$$\begin{aligned} \sigma_i^2 =& \frac{\sigma_n^2}{||\boldsymbol{h}_i||^2} \\ \mu_{i+} =& \sigma_i^2\Big(\frac{\boldsymbol{h}_i^\intercal \boldsymbol{v}_i}{\sigma_n^2} - \frac{1}{\lambda}\Big), \\ \mu_{i-} =& \sigma_i^2\Big(\frac{\boldsymbol{h}_i^\intercal \boldsymbol{v}_i}{\sigma_n^2} + \frac{1}{\lambda}\Big) \\ C(\mu,\sigma^2) =& \sqrt{\frac{\sigma^2\pi}{2}}\Big(1+\operatorname{erf}\big(\frac{\mu}{2\sigma^2}\big)\Big). \end{aligned} \tag{14}$$

The main steps of the proposed sampling algorithm are summarized in Algorithm 1.

Algorithm 1 Gibbs sampler.

Initialize with some $\boldsymbol{x}^{(0)}$
repeat
  Sample σ_n^2 according to Eq. (8).
  Sample λ according to Eq. (9).
  Sample ω according to Eq. (10).
  for $i = 1$ to M **do**
    Sample x_i according to Eq. (11).
  end for
until convergence

After convergence, the proposed algorithm ends up with sampled sequences that will be used to compute the minimum mean square error (MMSE) estimator of the unknown parameter vector, allowing us to compute the estimated signal $\widehat{\boldsymbol{x}}$, in addition to $\widehat{\sigma_n^2}$, $\widehat{\lambda}$ and $\widehat{\omega}$.

3. EXPERIMENTAL VALIDATION

In order to validate the proposed method for sparse signal and image restoration, three experiments have been conducted. The first two experiments correspond to 1D and 2D signal restorations. The third experiment handles an electroencephalography (EEG) reconstruction problem. Since we are in a simulation context, results are evaluated in terms of signal to noise ratio (SNR) given by $20\log_{10}\frac{||\boldsymbol{x}^0||}{||\boldsymbol{x}^0-\widehat{\boldsymbol{x}}||}$, where $\boldsymbol{x}^0$ and $\widehat{\boldsymbol{x}}$ are the reference and estimated signals, respectively.

3.1. 1D signal restoration

In this experiment, a 1D sparse signal $\boldsymbol{x}$ of size 100 is recovered from its distorted version $\boldsymbol{y}$ observed according to the model in Eq. (1), where the observation operator $\mathcal{H}$ is the second order difference operator. The observation $\boldsymbol{y}$ has been simulated by adding a Gaussian noise $\boldsymbol{n}$ of variance $\sigma_n^2 = 1$. The regularization scheme detailed in Section 2 is used with the same sparsity promoting priors and hyperparameter setting. Fig. 1 shows the original signal (Reference), the restored signal obtained with the proposed algorithm (MCMC) and the restored signal obtained with a PPXA algorithm (referred to as ℓ_1) whose parameters have been initialized by the output of the proposed algorithm.

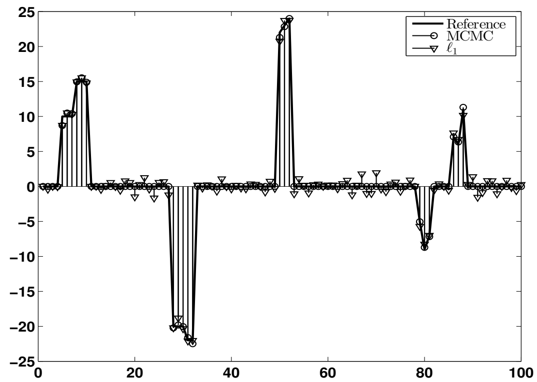

Fig. 1. Original and restored signals using the proposed method and ℓ_1 regularization.

The proposed method provides the smallest estimation error since the corresponding markers are always closer to the ground truth signal. Quantitatively speaking, with an initial SNR of 5.71 dB (observed signal), we achieve an SNR = 31.33 dB with our method, while the ℓ_1 regularization can only reach SNR = 20.05 dB. In addition to the automatic estimation of the regularization parameter, the proposed method ensures a better sparsity of the target signal due to the Bernoulli-Laplace prior, which is equivalent to an $\ell_0 + \ell_1$ regularization.
Regarding parameter estimation, Fig. 2 illustrates the posterior distributions of parameters σ_n^2, λ and ω, as well as their estimated values. We can easily notice that the estimation of σ_n^2 is very accurate since the estimated value (0.98) is very close to the reference one which is indicated by a vertical line in Fig. 2[top-right]. Moreover, the restored signal estimated with our method shows higher sparsity level since $\|\widehat{\boldsymbol{x}}_{\mathrm{MCMC}}\|_0 = 20$ and $\|\widehat{\boldsymbol{x}}_{\ell_1}\|_0 = 100$.

Finally, it is interesting to note that these results have been obtained after 600 iterations including a burn-in period of 200 iterations. The requested computation time using a Matlab implementation on a 64-bit 2.00GHz i7-3667U architecture was about 22 seconds.

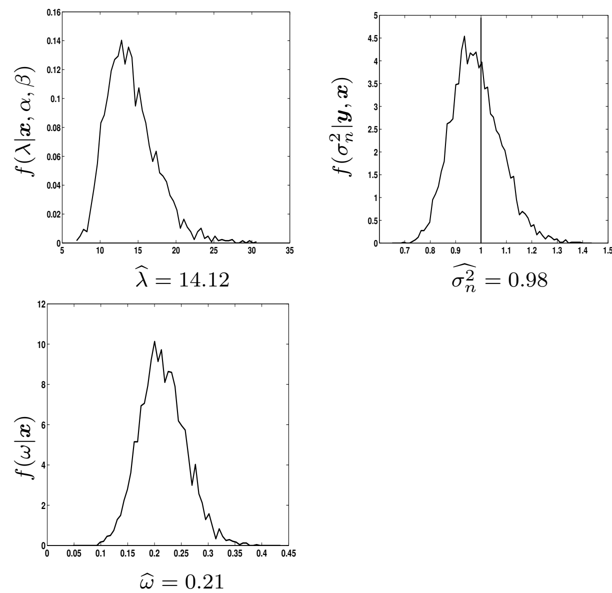

Fig. 2. Estimated posterior distributions of parameters σ_n^2, λ and ω.

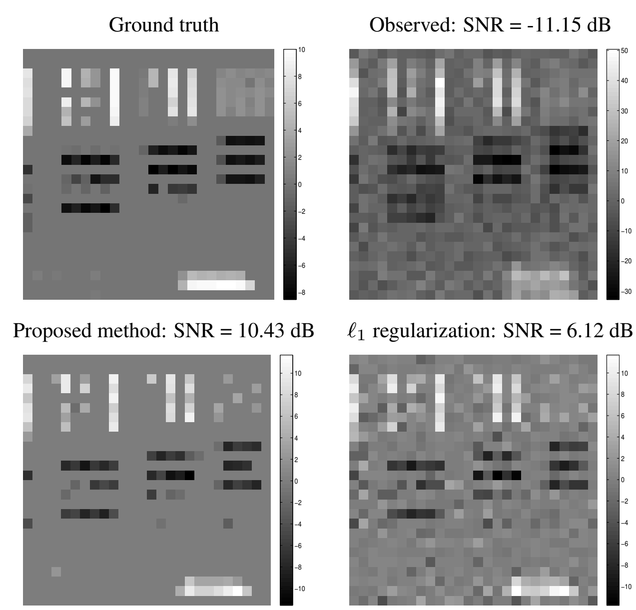

Fig. 3. Original, observed and restored images using the proposed method and ℓ_1 regularization.

3.2. 2D image restoration

In this experiment, a 2D sparse image $\boldsymbol{x}$ of size 26×26 is recovered from its distorted version $\boldsymbol{y}$ observed according to the model in Eq. (1) (with the same observation operator and noise level as in the first experiment). Fig. 3 illustrates the original, observed, and reconstructed images using our method and an ℓ_1 regularization.

The same conclusions apply to this example. The reported SNR values corroborate the clear advantage of the autocalibrated $\ell_0 + \ell_1$ Bayesian regularization made possible by our algorithm, compared to the variational ℓ_1 regularization (performed using parameters estimated by our algorithm). Sparsity levels of the estimated images show also the improvement obtained with our method since $\|\widehat{\boldsymbol{x}}_{\mathrm{MCMC}}\|_0 = 103$ and $\|\widehat{\boldsymbol{x}}_{\ell_1}\|_0 = 660$. For this experiment, the computational time with the same implementation and number of iterations as in the previous experiment was about 94 seconds.

3.3. EEG reconstruction

The last experiment addresses an EEG reconstruction problem where the observed signals correspond to the activity measured by each electrode during the acquisition time. Using the MNE software[1], we simulated a small EEG dataset using 20 electrodes with 35 sources (35 voxels on the brain surface), where only 4 of them have been chosen to be active. The simulation involved 21 time points, which means that we have to recover a 35×21 image $\boldsymbol{x}$ from an observation $\boldsymbol{y}$ of size 20×21. The linear operator here represents simply the brain model geometry. Fig. 4 displays the ground truth and reconstructed images using our method and an ℓ_1 regularization. Visually speaking, our method gives sparser recovered signal while preserving most of the activated sources. The obtained SNR values show also that the proposed algorithm outperforms the ℓ_1 regularization from a quantitative viewpoint. As regards sparsity levels, the same conclusion as for the first two experiments applies, since we have $\|\widehat{\boldsymbol{x}}_{\mathrm{MCMC}}\|_0 = 51$ and $\|\widehat{\boldsymbol{x}}_{\ell_1}\|_0 = 730$.
Computational time with the same implementation and number of iterations as in the previous two experiments was about 95 seconds.

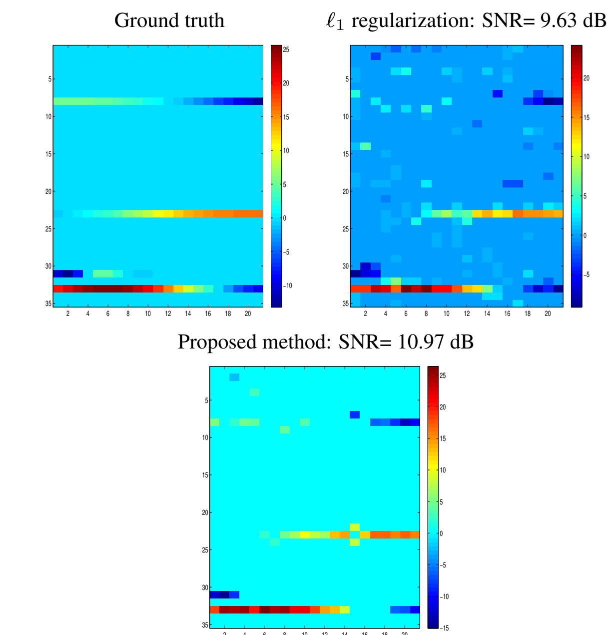

Fig. 4. Original, observed and restored images using the proposed method and ℓ_1 regularization.

4. CONCLUSION

We proposed in this paper a method for Bayesian sparse regularization using a Bernoulli-Laplace prior. This prior makes the $\ell_0 + \ell_1$ non-convex regularization problem feasible in a Bayesian framework, in contrast to variational methods which require the convexity of the cost function. In addition, the proposed method estimates the regularization parameters directly from the data. Validation on 1D and 2D restoration problems, as well as in EEG reconstruction, show the efficiency of the proposed method for recovering sparse signals. As a perspective, we will handle the problem of correlation between signal coefficients. Future work will also focus on the validation of our method on real EEG signals, as well as for other real world applications such as image deconvolution.

[1] http://www.martinos.org/mne/

5. REFERENCES

[1] C. Chaux, P. L. Combettes, J.-C Pesquet, and V. Wajs, "Iterative image deconvolution using overcomplete representations," in *European Signal Processing Conference*, Florence, Italy, Sep. 4-8 2006, 5 pp.

[2] 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.

[3] D.L Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] L. Chaari, S. Mériaux, S. Badillo, J.-Ch. Pesquet, and P. Ciuciu, "Multidimensional wavelet-based regularized reconstruction for parallel acquisition in neuroimaging," *EURASIP Journal on Advances in Signal Processing*, 2013, under revision: *http://arxiv.org/abs/1201.0022*.

[5] 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.

[6] L. Chaari, J.-C. Pesquet, A. Benazza-Benyahia, and P. Ciuciu, "Minimization of a sparsity promoting criterion for the recovery of complex-valued signals," in *Signal Processing with Adaptive Sparse Structured Representations Workshop (SPARS)*, Saint-Malo, France, April 6-9 2009.

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

[8] P. L. Combettes and J.-C. Pesquet, "A proximal decomposition method for solving convex variational inverse problems," *Inverse Problems*, vol. 24, no. 6, Dec. 2008, 27 p.

[9] A. Gramfort, D. Strohmeier, J. Haueisen, M.S. Hmlinen, and M. Kowalski, "Time-frequency mixed-norm estimates: Sparse M/EEG imaging with non-stationary source activations," *NeuroImage*, vol. 70, pp. 410–422, 2013.

[10] L. Chaari, T. Vincent, F. Forbes, M. Dojat, and P. Ciuciu, "Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach," *IEEE Trans. on Medical Imaging*, vol. 32, no. 5, pp. 821–837, May 2013.

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

[12] M. Lavielle, "Bayesian deconvolution of Bernoulli-Gaussian processes," *Signal Process.*, vol. 33, no. 1, pp. 67–79, Jul. 2009.

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

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