RECONSTRUCTION OF BIOMEDICAL IMAGES AND SPARSE STOCHASTIC MODELING

Emrah Bostan, Ulugbek Kamilov, and Michael Unser

Biomedical Imaging Group, EPFL, Switzerland

ABSTRACT

We propose a novel statistical formulation of the image-reconstruction problem from noisy linear measurements. We derive an extended family of MAP estimators based on the theory of continuous-domain sparse stochastic processes. We highlight the crucial roles of the whitening operator and of the Lévy exponent of the innovations which controls the sparsity of the model. While our family of estimators includes the traditional methods of Tikhonov and total-variation (TV) regularization as particular cases, it opens the door to a much broader class of potential functions (associated with infinitely divisible priors) that are inherently sparse and typically nonconvex. We also provide an algorithmic scheme—naturally suggested by our framework—that can handle arbitrary potential functions. Further, we consider the reconstruction of simulated MRI data and illustrate that the designed estimators can bring significant improvement in reconstruction performance.

Index Terms— sparse stochastic processes, statistical estimation, sparsity-promoting regularization, nonconvex optimization.

1. INTRODUCTION

The problem of image reconstruction from integrated (or blurred) measurements is central to modern imaging, for instance in parallel MRI, X-ray tomography, and electron and fluorescence microscopy. The image-reconstruction task is usually formulated as the linear inverse problem

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{1}$$

where the goal is to reconstruct the unknown discrete signal $\mathbf{s} \in \mathbb{R}^K$ from the noisy measurements $\mathbf{y} \in \mathbb{R}^M$. The matrix $\mathbf{H} \in \mathbb{R}^{M \times K}$, with $M \leq K$, models the spatial response of the imaging device. The vector $\mathbf{n}$ represents the measurement noise which, from here on, is assumed to be i.i.d. Gaussian with variance σ^2.

In this work, we are concerned with the statistical formulation of the reconstruction task based on the prior knowledge of the distribution of the signal. For instance, when the signal is zero-mean Gaussian with covariance $\mathbf{C_s}$, it is well known that the MMSE solution is given by the Wiener filter. This can also be recast in terms of the quadratic minimization problem

$$\mathbf{s}^* = \arg\min_{\mathbf{s}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \sigma^2 \|\mathbf{C}_{\mathbf{s}}^{-1/2}\mathbf{s}\|_2^2, \tag{2}$$

which provides a direct link between the Gaussian MMSE/MAP estimation and the classical method of Tikhonov regularization. Observe that (2) imposes a constraint on the ℓ_2 norm of the auxiliary "innovation" or whitened signal $\mathbf{u} = \mathbf{C}_s^{-1/2}\mathbf{s}$, which is i.i.d. Gaussian with a unit variance.

The recent trend in the field has been to move away from quadratic regularization and to replace the second term in (2) by some suitable ℓ_1 norm in order to promote sparse solutions [1]. A preferred choice of regularization is $\mathfrak{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_1$ where $\mathbf{L}$ is the discrete counterpart of gradient or Laplacian operators, with the former corresponding to total-variation regularization. A considerable amount of effort has been put in designing efficient algorithms for the reconstruction of signals under such sparsity constraints.

While the current formulations of sparse signal recovery are based on solid variational principles, they are fundamentally deterministic. The purpose of this paper is to introduce a stochastic framework that justifies the use of nonquadratic regularization functionals by tying them to some generative, continuous-domain signal model. The approach builds upon the recent theory of sparse stochastic processes which are specified as solutions of stochastic differential equations (SDE) driven by non-Gaussian noise [2].

The main contributions of this work are as follows:

- The derivation of the prior distribution of some non-Gaussian discrete-domain innovation sequence $\mathbf{u} = \mathbf{L}\mathbf{s}$ based on the modeling of the signal $s = \mathrm{L}^{-1}w$ in the *continuous-domain* as a sparse stochastic process. Our key finding is that the probability density function (pdf) must be part of the extended family of *infinitely divisible* distributions.
- The characterization of admissible potential functions (prior log-likelihoods) and the specification of the corresponding MAP estimators. We also illustrate the connection between these statistical estimators and existing deterministic methods.
- The proposal of a generic reconstruction algorithm, based on variable splitting, that efficiently handles the different estimators, including the nonconvex ones. The technique is applied to the reconstruction of MRI from non-uniform k-space samples.

It is noteworthy that the proposed formulation lends itself to an analytical treatment using Fourier-domain techniques. Moreover, it is backward compatible with the Gaussian formulation and the traditional theory of stationary processes.

2. SPARSE STOCHASTIC PROCESSES

Our formulation is based on the assumption that $\mathbf{s}$ is the discretized version of the continuous-domain stochastic process $s(\boldsymbol{x})$, with $\boldsymbol{x} \in \mathbb{R}^d$, which is specified as the solution of a SDE.

2.1. Continuous-domain innovation model

The generic form of a linear stochastic differential equation is

$$\mathrm{L}s = w, \tag{3}$$

This work was partially supported by the Center for Biomedical Imaging of the Geneva-Lausanne Universities and EPFL, as well as by the Foundations Leenaards and Louis-Jeannet and by the European Commission under Grant ERC-2010-AdG 267439-FUN-SP.

where w is a continuous-domain white noise (not necessarily Gaussian) and L is a (multidimensional) differential operator. In the classical theory of stationary processes, the right-hand side of (3) is often interpreted as the innovation (unpredictable) part of the process, while L is referred to as the whitening operator. The formal solution (if it exists) can be written as $s = \mathrm{L}^{-1}w$, where L^{-1} is an appropriate right inverse of L, which amounts to some generalized "integration" of the innovation w. The implication is that the correlation structure of the stochastic process s is determined by the mixing operator L^{-1}, while its randomness and sparsity pattern is governed by the excitation noise w.

In Gelfand's theory of generalized stochastic processes [3], s is observed indirectly by taking inner-products $\langle s, \varphi \rangle$ with rapidly decreasing test functions φ in Schwartz class $\mathcal{S}$. In that framework, the stochastic process s is completely characterized by its characteristic form $\widehat{\mathscr{P}_s}(\varphi) = \mathbb{E}\{e^{j\langle s,\varphi\rangle}\}$, where $\mathbb{E}\{\cdot\}$ denotes the expectation. The powerful aspect of this characterization is that $\widehat{\mathscr{P}_s}$ is indexed by a generic function $\varphi \in \mathcal{S}$ rather than by a scalar (or vector) Fourier variable ω. As such, it constitutes the *infinite-dimensional* generalization of the characteristic function of a conventional random variable.

In a recent work, we characterized the complete class of stochastic processes that are specified by (3) where L is shift-invariant and w is part of the class of so-called Lévy noises [4]. In that context, we have shown that $\widehat{\mathscr{P}_s}(\varphi)$ is given by

$$\widehat{\mathscr{P}_s}(\varphi) = \exp\left(\int_{\mathbb{R}^d} f\big(\mathrm{L}^{-1*}\varphi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}\right), \tag{4}$$

where the univariate function $f(\cdot)$ is the Lévy exponent of the white noise w and L^{-1*} is the adjoint of L^{-1}. We have also proven that a sufficient condition for the existence of the process is that $\|\mathrm{L}^{-1*}\varphi\|_{L_p} < C\|\varphi\|_{L_p}$ for all $\varphi \in L_p(\mathbb{R}^d)$ and some $p \geq 1$.

The particular case of Gaussian processes is recovered by taking $f(\omega) = -|\omega|^2$ and $p = 2$. More generally, the Lévy exponent $f(\omega)$ can be chosen to be any conditionally positive-definite function of order one.

2.2. Discrete-Domain Innovations

We now would like to obtain a discrete version of the innovation model (3) by applying the discrete counterpart L_d of the whitening operator L to the samples $s[\boldsymbol{k}] = s(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{k}}$ of the process. The implicit requirement is that the composition of L_d and L^{-1} results in a stable, shift-invariant operator whose impulse response

$$\mathrm{L}_d\mathrm{L}^{-1}\delta(x) = \beta_\mathrm{L}(x) \in L_1(\mathbb{R}^d) \tag{5}$$

is maximally localized [5]. We call β_L the *generalized B-spline* associated with the operator L. For instance, when $\mathrm{L} = \mathrm{D}$ (derivative operator) and $\mathrm{L}_\mathrm{d} = \mathrm{D}_\mathrm{d}$ (finite-difference operator), then $\beta_\mathrm{D}(x) = \mathrm{rect}(x - \frac{1}{2})$ is a causal rectangle (polynomial B-spline of degree 0). In matrix-vector notations, we write the discrete-domain counterpart of the innovation model (3) as

$$\mathbf{u} = \mathbf{L}\mathbf{s}, \tag{6}$$

where $\mathbf{u} = (u[\boldsymbol{k}])_{\boldsymbol{k}\in\Omega}$ is the discrete *innovation variable* (with a slight abuse of language). A key point of our formulation is that we can rely on (4) to derive the pdf of u.

Proposition 1. *Let s be a stochastic process whose characteristic form is given by (4). Then, $u[\boldsymbol{k}] = \mathrm{L}_\mathrm{d}s(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{k}}$ is stationary and such that $u[\boldsymbol{k}] = (\beta_\mathrm{L} * w)(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{k}}$ where β_L is defined by (5). Its first-order pdf is given by the inverse Fourier integral*

$$p_U(u) = \int_{\mathbb{R}} \exp\left(\int_{\mathbb{R}^d} f\big(\omega\beta_\mathrm{L}(-\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}\right) e^{-j\omega u}\frac{\mathrm{d}\omega}{2\pi}, \tag{7}$$

and is part of the family of infinitely divisible distributions.

	$p_U(x)$	$\Phi_U(x)$
Gaussian	$\frac{1}{\sigma_0\sqrt{2\pi}}e^{-x^2/2\sigma_0^2}$	$M_1x^2 + C_1$
Laplace	$\frac{\lambda}{2}e^{-\lambda\lvert x\rvert}$	$M_2\lvert x\rvert + C_2$
Student's	$\frac{1}{\epsilon B(r,\frac{1}{2})}\left(\frac{1}{(x/\epsilon)^2+1}\right)^{r+\frac{1}{2}}$	$M_3\log(\frac{x^2+\epsilon^2}{\epsilon^2}) + C_3$

Table 1. Some infinitely divisible distributions and the corresponding potential functions.

Proposition 1 indicates that the pdf of $\mathbf{u}$ is completely determined by the continuous-domain innovation model via the generalized B-spline β_L and the Lévy exponent of the excitation noise w. Its sparsity pattern (tail behavior of the pdf and/or presence of a mass distribution at the origin) is primarily dependent upon $f(\cdot)$. The important conceptual aspect is that the class of admissible pdfs is restricted to the (classical) family of *infinitely divisible* (id) laws.

A case of special interest is $f(\omega) = -|\omega|^\alpha$ with $\alpha \in (1, 2]$. It is not difficult to show that the corresponding characteristic function is $\hat{p}_U(\omega) = e^{-|s_0\omega|^\alpha}$ with $s_0 = \|\beta_\mathrm{L}\|_{L_\alpha}$. The implication is that u has a symmetric α-stable (SαS) distribution, irrespective of L. The case $\alpha = 2$ corresponds to the only nonsparse scenario (Gaussian), while smaller values of α give rise to much heavier-tailed distributions. For $\alpha = 1$, we obtain the Cauchy distribution (or Student with $r = 1/2$) whose standardized version is $p_U(x) = \frac{1}{\pi(1+x^2)}$. Note that all the members of the SαS family with $\alpha \neq 2$ satisfy some stringent criterion for compressibility [6].

2.3. MAP Estimation

The present MAP estimator is derived under the decoupling assumption that $\mathbf{u}$ is i.i.d. In order to reconstruct the image, we seek an estimate of $\mathbf{s}$ that maximizes the posterior distribution $p_{S|Y}$ which depends upon the prior distribution p_S, assumed to be proportional to p_U, and the Gaussian noise distribution p_N. The direct application of Bayes' rule yields

$$\begin{aligned} p_{S|Y}(\mathbf{s}|\mathbf{y}) &\propto p_N(\mathbf{y} - \mathbf{H}\mathbf{s})p_U(\mathbf{u}) \\ &\propto \exp\left(-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{s}\|^2}{2\sigma^2}\right)\prod_{\boldsymbol{k}\in\Omega} p_U\big([\mathbf{L}\mathbf{s}]_{\boldsymbol{k}}\big), \end{aligned}$$

where Ω is the region-of-interest (ROI) and σ^2 is the variance of the noise. We can then specify the MAP estimator of $\mathbf{s}$ as

$$\begin{aligned} \mathbf{s}_\mathrm{MAP} &= \arg\max_{\mathbf{s}} p_{S|Y}(\mathbf{s}|\mathbf{y}) \\ &= \arg\min_{\mathbf{s}} \tfrac{1}{2}||\mathbf{H}\mathbf{s} - \mathbf{y}||_2^2 + \sigma^2\sum_{k\in\Omega}\Phi_U\big([\mathbf{L}\mathbf{s}]_{\boldsymbol{k}}\big), \end{aligned} \tag{8}$$

where $\Phi_U(x) = -\log p_U(x)$ is the potential function.

2.4. Potential Functions

Recall that, in the current Bayesian formulation, the potential function $\Phi_U(x) = -\log p_U(x)$ is specified by the Lévy exponent $f(\omega)$ of the

continuous-domain innovation w. We now point out the connections between the general estimator (8) and the standard variational methods. The three members of the id family that are of interest to the discussion are listed in Table 1. The exact values of the constants C_1, C_2, and C_3 and scaling factors M_1, M_2, and M_3 have been omitted as they are not relevant to the optimization task.

The first quadratic potential function yields the classical Tikhonov regularizer that produces a stabilized linear solution $\big($set $\mathbf{L} = \mathbf{C_s}^{-1/2}$ in (2)$\big)$. The second potential function provides the ℓ_1-type regularizer. Moreover, the well-known TV regularizer is obtained if the operator L is a first-order derivative operator. Interestingly, the third log-based potential is linked to the limit case of the ℓ_p relaxation scheme as p tends to zero [7].

3. MINIMIZATION ALGORITHM

We now formulate the MAP estimation as the constrained optimization

$$\begin{aligned} \mathbf{s}_{\text{MAP}} = \quad & \underset{\mathbf{s}\in\mathbb{R}^K}{\arg\min} \quad \left(\frac{1}{2}||\mathbf{Hs}-\mathbf{y}||_2^2 + \sigma^2 \sum_{\boldsymbol{k}\in\Omega} \Phi_U\left(u[\boldsymbol{k}]\right)\right) \\ & \text{subject to} \qquad \mathbf{u} = \mathbf{Ls}, \end{aligned} \tag{9}$$

where Ω is the ROI of size K and $\mathbf{L}:\mathbb{R}^K \to \mathbb{R}^K$ is a linear operator. Note that this reformulation is very much in line with the *variable-splitting* technique that has been promoted by various authors [8, 9]. Instead of considering the constrained problem directly, we introduce the corresponding *augmented Lagrangian* (AL) functional

$$\begin{aligned} \mathcal{L}_{\mathcal{A}}(\mathbf{s},\mathbf{u},\boldsymbol{\alpha}) \quad &= \quad \frac{1}{2}||\mathbf{Hs}-\mathbf{y}||_2^2 + \sigma^2 \sum_{\boldsymbol{k}\in\Omega} \Phi_U\left(u[\boldsymbol{k}]\right) \\ &+ \quad \boldsymbol{\alpha}^T(\mathbf{Ls}-\mathbf{u}) + \frac{\mu}{2}||\mathbf{Ls}-\mathbf{u}||_2^2, \end{aligned}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^K$ denotes the *Lagrange-multiplier* vector and $\mu \in \mathbb{R}$ is called the *penalty parameter*. Then, we face the equivalent unconstrained problem

$$\underset{\left(\mathbf{s}\in\mathbb{R}^K,\ \mathbf{u}\in\mathbb{R}^K\right)}{\text{minimize}} \ \mathcal{L}_{\mathcal{A}}\ . \tag{10}$$

As a replacement for the joint minimization of the AL functional over $(\mathbf{s},\mathbf{u})$, we apply the *alternating-direction method of multipliers* (ADMM) [10] and consider minimizing $\mathcal{L}_{\mathcal{A}}(\mathbf{s},\mathbf{u},\boldsymbol{\alpha})$ with respect to each variable while keeping the others fixed. This is followed by an update of the Lagrange multiplier which results in the following scheme at iteration t:

$$\mathbf{u}^{t+1} \leftarrow \underset{\mathbf{u}}{\arg\min} \quad \mathcal{L}_{\mathcal{A}}(\mathbf{s}^t,\mathbf{u},\boldsymbol{\alpha}^t) \tag{11a}$$

$$\mathbf{s}^{t+1} \leftarrow \underset{\mathbf{s}}{\arg\min} \quad \mathcal{L}_{\mathcal{A}}(\mathbf{s},\mathbf{u}^{t+1},\boldsymbol{\alpha}^t) \tag{11b}$$

$$\boldsymbol{\alpha}^{t+1} = \boldsymbol{\alpha}^t + \mu(\mathbf{Ls}^{t+1}-\mathbf{u}^{t+1}). \tag{11c}$$

From the Lagrangian duality point of view, (11c) can be interpreted as a maximization of the dual functional so that, as the above scheme proceeds, feasibility is imposed [10]. Seeing that our potential functions are closed and proper, we deduce Remark 1 from [10].

Remark 1. *If $\Phi_U : \mathbb{R} \to \mathbb{R}_+$ is convex and the unaugmented Lagrangian functional has a saddle point, then the constraint in* (9) *is satisfied and the objective functional reaches the optimal value as $t \to \infty$.*

3.1. Minimization (11a)

The minimization is separable. Hence, (11a) reduces to performing K scalar minimizations:

$$\min_{u[\boldsymbol{k}]\in\mathbb{R}} \left(\sigma^2\Phi_U(u[\boldsymbol{k}]) + \frac{\mu}{2}\left(u[\boldsymbol{k}]-z[\boldsymbol{k}]\right)^2\right), \forall \boldsymbol{k}\in\Omega, \tag{12}$$

where $\mathbf{z} = \mathbf{Ls} + \boldsymbol{\alpha}/\mu$.

Definition 1. *The proximity operator with weighting factor λ associated to the function Φ_U is defined as*

$$\text{prox}_{\Phi_U}(y;\lambda) = \arg\min_{x\in\mathbb{R}} \frac{1}{2}(y-x)^2 + \lambda\Phi_U(x). \tag{13}$$

Therefore, (11a) is obtained by applying $\text{prox}_{\Phi_U}(z;\ \frac{\sigma^2}{\mu})$ in a *component-wise* fashion to $\mathbf{z} = \mathbf{Ls}^t + \boldsymbol{\alpha}^t/\mu$. The closed-form solutions for the proximity operator are well-known for the Gaussian and Laplace priors. They are given by

$$\text{prox}_{(\cdot)^2}(z;\lambda) = z(1+2\lambda)^{-1} \tag{14a}$$

$$\text{prox}_{|\cdot|}(z;\lambda) = \max(|z|-\lambda,0)\text{sgn}(z). \tag{14b}$$

While the Student proximity operator has no closed-form solution, it can still be precomputed and stored in a *look-up table* (LUT). This suggests a very fast implementation of the proximal step that is applicable for any potential function $\Phi_U(x)$. We also note that the method remains applicable when $\Phi_U(x)$ is nonconvex, with the following caveat:

Remark 2. *When ADMM converges and Φ_U is nonconvex, it converges to a local minimum, including the case where the sub-minimization problems are solved exactly [10].*

In the case of a nonconvex problem, the algorithm can potentially get trapped in local minima in the early stages of the optimization. It is therefore recommended to apply a deterministic *continuation* method or to consider a *warm start* that can be obtained via solving the problem first with Gaussian or Laplace priors. We have opted for the latter solution and have not encountered any convergence problem in practice.

3.2. Minimization (11b)

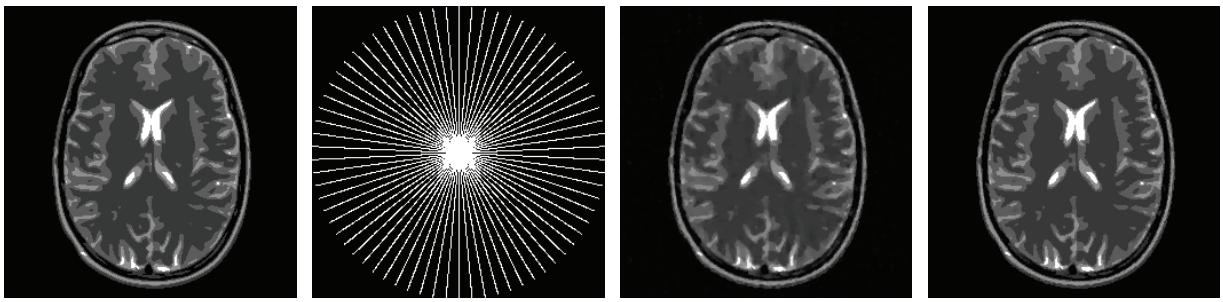
This step of the optimization amounts to solving a quadratic problem. It yields a linear estimate of the form

$$\mathbf{s}^{t+1} = (\mathbf{H}^H\mathbf{H} + \mu\mathbf{L}^H\mathbf{L})^{-1}\left(\mathbf{H}^H\mathbf{y} + \mu\mathbf{L}^H\left(\mathbf{u}^{t+1} - \frac{\boldsymbol{\alpha}^t}{\mu}\right)\right). \tag{15}$$

Interestingly, this part of the problem is equivalent to the Gaussian solution. It can either be solved iteratively using the conjugate-gradient method, or directly, for instance by means of FFTs when $\mathbf{H}^H\mathbf{H}$ has a convolution structure, as in our series of experiments.

4. NUMERICAL RESULTS

To assess the numerical performance of the estimators, we simulated the reconstruction of a realistic analytical brain phantom [11] from its undersampled discrete Fourier-domain coefficients. The corresponding system matrix is $\mathbf{H} = \boldsymbol{\Gamma}\mathbf{F}$, where $\mathbf{F}$ is the DFT matrix and $\boldsymbol{\Gamma}$ is the underlying downsampling operator. For the experiment, we



(a) (b) (c) (d)

Fig. 1. Reconstruction of the realistic analytical brain phantom: (a) original phantom; (b) Fourier sampling pattern; (c) reconstruction using Laplace prior, SNR = 17.83 dB; (d) reconstruction using Student's prior, SNR = 21.33 dB.

choose $\mathbf{L}$ to be the gradient operator and adopt the vectorial counterparts of the proximity operators. For this MRI model, we have that

$$\mathbf{H}^H\mathbf{H} = \mathbf{F}^H\mathbf{\Psi}\mathbf{F},$$

where $\mathbf{\Psi} = \mathbf{\Gamma}^T\mathbf{\Gamma}$ is a diagonal matrix. Moreover, $\mathbf{L}$ is also diagonalized by the operator $\mathbf{F}$, which suggests an efficient Fourier-domain solution of (15).

As illustrated in Figure 1, we consider a phantom of size 256×256 and a radial sampling pattern composed of 35 lines, which corresponds to an undersampling factor of 6.5. In the experiments, the measurements are corrupted with a Gaussian noise with $\sigma = 10^{-4}$. For the Laplace estimator, we initialize the algorithm with the zero-filled back-projected solution $\mathbf{s}^0 = \mathbf{H}^H\mathbf{y}$. For the Student's estimator, we use the solution of the Laplace estimator as the initial solution. The ϵ parameter of Student's prior is set to 10^{-2}. The regularization weights, for both estimators, are optimized by using an oracle in order to obtain the highest-possible SNR.

We illustrate the results in Figure 1. Observe that the image reconstruction with the Laplace prior exhibits some artifacts and lack of sharpness in the transition zones. The artifacts are substantially reduced in the image estimated with the Student prior. The sparsity constraint helps restoring the sharpness of the edges and turns out to be particularly well-suited for this kind of piecewise-smooth image.

5. CONCLUSION

We have applied the recent theory of sparse stochastic processes to derive a general family of MAP estimators for the reconstruction of biomedical images. An important conceptual point is that the potential function cannot be selected arbitrarily; moreover, the resulting optimization problem is generally nonconvex with the notable exception of the Gaussian and the Laplacian priors. We have emphasized the advantage of using the discrete innovation as auxiliary variable both when formulating the problem and when implementing the algorithm. By focusing on the unconstrained equivalent problem via an augmented Lagrangian functional, we have proposed an ADMM-based minimization scheme that can handle arbitrary potential functions. Finally, we have illustrated the suitability of this type of sparse modeling for the reconstruction of simulated MRI data. Our results demonstrate that the proper use of (nonconvex) sparse priors can outperform ℓ_1-type regularizations.

6. REFERENCES

[1] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, pp. 1182–95, 2007.

[2] M. Unser and P. D. Tafti, "Stochastic models for sparse and piecewise-smooth signals," *IEEE Transactions on Signal Processing*, vol. 59, pp. 989–1006, 2011.

[3] I. Gelfand and N. Y. Vilenkin, *Generalized Functions. Vol. 4. Applications of Harmonic Analysis.* New York, USA: Academic Press, 1964.

[4] M. Unser, P. D. Tafti, and Q. Sun, "A unified formulation of Gaussian vs. sparse stochastic processes–Part I: Continuous-domain theory." *arXiv:1108.6150v1*.

[5] M. Unser, P. D. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian vs. sparse stochastic processes–Part II: Discrete-domain theory." *arXiv:1108.6152v1*.

[6] A. Amini, M. Unser, and F. Marvasti, "Compressibility of deterministic and random infinite sequences," *IEEE Transactions on Signal Processing*, vol. 59, pp. 5193–5201, 2011.

[7] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, pp. 317 –329, 2010.

[8] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, pp. 248–272, 2008.

[9] S. Ramani and J. A. Fessler, "Regularized parallel MRI reconstruction using an alternating direction method of multipliers," in *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 385–388, 2011.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, pp. 1–122, 2010.

[11] M. Guerquin-Kern, L. Lejeune, K. P. Pruessmann, and M. Unser, "Realistic analytical phantoms for parallel magnetic resonance imaging," *to appear in IEEE Transactions on Medical Imaging*.