Sparse Stochastic Processes and Discretization of Linear Inverse Problems

Emrah Bostan, *Student Member, IEEE*, Ulugbek S. Kamilov, *Student Member, IEEE*, Masih Nilchian, *Student Member, IEEE*, and Michael Unser, *Fellow, IEEE*

***Abstract*— We present a novel statistically-based discretization paradigm and derive a class of maximum a posteriori (MAP) estimators for solving ill-conditioned linear inverse problems. We are guided by the theory of sparse stochastic processes, which specifies continuous-domain signals as solutions of linear stochastic differential equations. Accordingly, we show that the class of admissible priors for the discretized version of the signal is confined to the family of infinitely divisible distributions. Our estimators not only cover the well-studied methods of Tikhonov and ℓ_1-type regularizations as particular cases, but also open the door to a broader class of sparsity-promoting regularization schemes that are typically nonconvex. We provide an algorithm that handles the corresponding nonconvex problems and illustrate the use of our formalism by applying it to deconvolution, magnetic resonance imaging, and X-ray tomographic reconstruction problems. Finally, we compare the performance of estimators associated with models of increasing sparsity.**

***Index Terms*— Innovation models, maximum a posteriori (MAP) estimation, nonconvex optimization, non-Gaussian statistics, sparse stochastic processes, sparsity-promoting regularization.**

I. Introduction

We consider linear inverse problems that occur in a variety of biomedical imaging applications [1]–[3]. In this class of problems, the measurements $\mathbf{y}$ are obtained through the forward model

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

where $\mathbf{s}$ represents the true signal/image. The linear operator $\mathbf{H}$ models the physical response of the acquisition/imaging device and $\mathbf{n}$ is some additive noise. A conventional approach for reconstructing $\mathbf{s}$ is to formulate the reconstructed signal $\mathbf{s}^\star$ as the solution of the optimization problem

$$\mathbf{s}^\star = \arg\min_{\mathbf{s}} \mathcal{D}(\mathbf{s}; \mathbf{y}) + \lambda \mathcal{R}(\mathbf{s}), \tag{2}$$

where $\mathcal{D}(\mathbf{s}; \mathbf{y})$ quantifies the distance separating the reconstruction from the observed measurements, $\mathcal{R}(\mathbf{s})$ measures the regularity of the reconstruction, and $\lambda > 0$ is the regularization parameter.

In the classical quadratic (Tikhonov-type) reconstruction schemes, one utilizes ℓ_2-norms for measuring both the data consistency and the reconstruction regularity [3]. In a general setting, this leads to a smooth optimization problem of the form

$$\mathbf{s}^\star = \arg\min_{\mathbf{s}} \|\mathbf{y} - \mathbf{H}\mathbf{s}\|_2^2 + \lambda \|\mathbf{R}\mathbf{s}\|_2^2 \tag{3}$$

where $\mathbf{R}$ is a linear operator and the formal solution is given by

$$\mathbf{s}^\star = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H} + \lambda \mathbf{R}^{\mathrm{T}}\mathbf{R}\right)^{-1} \mathbf{H}^{\mathrm{T}}\mathbf{y}. \tag{4}$$

The linear reconstruction framework expressed in (3) and (4) can also be derived from a statistical perspective. Under the hypothesis that $\mathbf{s}$ follows a multivariate zero-mean Gaussian distribution with covariance matrix $\mathbf{C_{ss}} = \mathbb{E}\{\mathbf{s}\mathbf{s}^{\mathrm{T}}\}$, the operator $\mathbf{C_{ss}}^{-1/2}$ *whitens* $\mathbf{s}$ (i.e., renders its components independent). Moreover, if $\mathbf{n}$ is additive white Gaussian noise (AWGN) of variance σ^2, the maximum a posteriori (MAP) formulation of the reconstruction problem yields

$$\mathbf{s}_{\mathrm{MAP}} = (\mathbf{H}^{\mathrm{T}}\mathbf{H} + \sigma^2 \mathbf{C_{ss}}^{-1})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{y} \tag{5}$$

which is equal to (4) when $\mathbf{C_{ss}}^{-1/2} = \mathbf{R}$ and $\sigma^2 = \lambda$. In the Gaussian scenario, the MAP estimator is known to yield the minimum mean square error (MMSE) solution. The equivalent Wiener solution (5) is also applicable for non-Gaussian models with known covariance $\mathbf{C_{ss}}$ and is commonly referred to as the linear minimum mean square error (LMMSE) [4].

In recent years, the paradigm in variational formulations for signal reconstruction has shifted from the classical linear schemes to the *sparsity-promoting* methods motivated by the observation that many signals that occur naturally have sparse or nearly-sparse representations in some transform domain [5]. The promotion of sparsity is achieved by specifying well-chosen non-quadratic regularization functionals and results in nonlinear reconstruction. One common choice for the regularization functional is $\mathcal{R}(\mathbf{v}) = \|\mathbf{v}\|_1$, where $\mathbf{v} = \mathbf{W}^{-1}\mathbf{s}$ represents the coefficients of a wavelet (or a wavelet-like multiscale) transform [6]. An alternative choice is $\mathcal{R}(\mathbf{s}) = \|\mathbf{L}\mathbf{s}\|_1$, where $\mathbf{L}$ is the discrete version of the gradient or Laplacian operator, with the gradient one being known as total-variation (TV) regularization [7]. Although using the ℓ_1 norm as regularization functional has been around for

Manuscript received August 14, 2012; revised December 11, 2012; accepted March 11, 2013. Date of publication March 28, 2013; date of current version May 16, 2013. This work was supported in part by the Center for Biomedical Imaging of the Geneva-Lausanne Universities and EPFL, the Foundations Leenaards and Louis-Jeannet, and the European Commission under Grant ERC-2010-AdG 267439-FUN-SP. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jean-Philippe Thiran.

The authors are with the Biomedical Imaging Group, École Polytechnique fédérale de Lausanne, Lausanne CH–1015, Switzerland (e-mail: emrah.bostan@gmail.com; kamilov@ieee.org; masih.nilchian@epfl.ch; michael.unser@epfl.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2255305

1057-7149/$31.00 © 2013 IEEE

some time (for instance, see [8], [9]), it is currently at the heart of sparse signal reconstruction problems. Consequently, a significant amount of research is dedicated to the design of efficient algorithms for nonlinear reconstruction methods [10].

The current formulations of sparsity-promoting regularization are based on solid variational principles and are predominantly deterministic. They can also be interpreted in statistical terms as MAP estimators by considering generalized Gaussian or Laplace priors [11]–[13]. These models, however, are tightly linked to the choice of a given sparsifying transform, with the downside that they do not provide further insights on the true nature of the signal.

A. Contributions

In this paper, we revisit the signal reconstruction problem by specifying upfront a *continuous-domain model* for the signal that is independent from the subsequent reconstruction task/algorithm and apply a *proper discretization scheme* to derive the corresponding MAP estimator. Our approach builds upon the theory of continuous-domain sparse stochastic processes [14]. In this framework, the stochastic process is defined through an innovation model that can be driven by a *non-Gaussian* excitation.[1] The primary advantage of our continuous-domain formulation is that it lends itself to an analytical treatment. In particular, it allows for the derivation of the probability density function (pdf) of the signal in any transform domain, which is typically much more difficult in a purely discrete framework. Remarkably, the underlying class of models also provides us with a strict derivation of the class of admissible regularization functionals which happen to be confined to two categories: Gaussian or sparse.

The main contributions of the present work are as follows:

- The introduction of continuous-domain stochastic models in the formulation of inverse problems. This leads to the use of non-quadratic reconstruction schemes.
- A general scheme for the proper discretization of these problems. This scheme specifies feasible statistical estimators.
- The characterization of the complete class of admissible potential functions (prior log-likelihoods) and the derivation of the corresponding MAP estimators. The connections between these estimators and the existing deterministic methods such as TV and ℓ_1 regularizations are also explained.
- A general reconstruction algorithm, based on variable-splitting techniques, that handles different estimators, including the nonconvex ones. The algorithm is applied to deconvolution and to the reconstruction of MR and X-ray images.

B. Outline

The paper is organized as follows: In Section II, we explain the acquisition model and obtain the corresponding representation of the signal $\mathbf{s}$ and the system matrix $\mathbf{H}$. In Section III, we introduce the continuous-domain innovation model that defines a generalized stochastic process. We then statistically specify the discrete-domain counterpart of the innovation model and characterize admissible prior distributions. Based on this characterization, we derive the MAP estimation as an optimization problem in Section IV. In Section V, we provide an efficient algorithm to solve the optimization problem for a variety of admissible priors. Finally, in Section VI, we illustrate our discretization procedure by applying it to a series of deconvolution and of MR and X-ray image-reconstruction problems. This allows us to compare the effect of different sparsity priors on the solution.

[1] It is noteworthy that the theory includes the stationary Gaussian processes.

C. Notations

Throughout the paper, we assume that the measurement noise is AWGN of variance σ^2. The input argument $\boldsymbol{x} \in \mathbb{R}^d$ of the continuous-domain signals is written inside parenthesis (e.g., $s(\boldsymbol{x})$) whereas, for discrete-domain signals, we employ $\boldsymbol{k} \in \mathbb{Z}^d$ and use brackets (e.g., $s[\boldsymbol{k}]$). The scalar product is represented by $\langle \cdot, \cdot \rangle$ and $\delta(\cdot)$ denotes the Dirac impulse.

II. Measurement Model

In this section, we develop a discretization scheme that allows us to obtain a tractable representation of continuously-defined measurement problem, with minimal loss of information. Such discrete representation is crucial since the resulting reconstruction algorithms are implemented numerically.

A. Discretization of the Signal

To obtain a clean analytical discretization of the problem, we consider the generalized sampling approach using "shift-invariant" reconstruction spaces [15]. The advantage of such a representation is that it offers the same type of error control as finite-element methods (i.e., the approximation error between the original signal and its representation in the reconstruction space can be made arbitrarily small by choosing a sufficiently fine reconstruction grid, see [15]).

The idea is to represent the signal s by projecting it onto a reconstruction space. We define our reconstruction space at resolution T as

$$V_T(\varphi_{\text{int}}) = \left\{ s_T(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} s[\boldsymbol{k}]\, \varphi_{\text{int}}\left(\frac{\boldsymbol{x}}{T} - \boldsymbol{k}\right) : s[\boldsymbol{k}] \in \ell_\infty(\mathbb{Z}^d) \right\}, \quad (6)$$

where $s[\boldsymbol{k}] = s(\boldsymbol{x})|_{\boldsymbol{x}=T\boldsymbol{k}}$, and φ_{int} is an interpolating basis function positioned on the reconstruction grid $T\mathbb{Z}^d$. The interpolation property is $\varphi_{\text{int}}(\boldsymbol{k}) = \delta[\boldsymbol{k}]$. For the representation of s in terms of its samples $s[\boldsymbol{k}]$ to be stable and unambiguous, φ_{int} has to be a valid Riesz basis for $V_T(\varphi_{\text{int}})$. Moreover, to guarantee that the approximation error decays as a function of T, the basis function should satisfy the partition of unity property [15]

$$\sum_{\boldsymbol{k} \in \mathbb{Z}^d} \varphi_{\text{int}}(\boldsymbol{x} - \boldsymbol{k}) = 1, \ \forall \boldsymbol{x} \in \mathbb{R}^d. \quad (7)$$

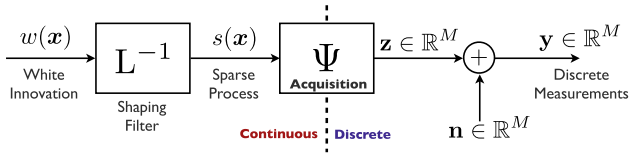


Fig. 1. General form of the linear, continuous-domain measurement model considered in this paper. The signal $s(\boldsymbol{x})$ is acquired through linear measurements of the form $z_m = [\Psi s]_m = \langle s, \psi_m \rangle$. The resulting vector $\mathbf{z} \in \mathbb{R}^M$ is corrupted with AWGN. Our goal is to estimate the original signal s from noisy measurements $\mathbf{y}$ by exploiting the knowledge that s is a realization of a sparse stochastic process that satisfies the innovation model $\mathrm{L}s = w$, where w is a non-Gaussian white innovation process.

The projection of the signal onto the reconstruction space $V_T(\varphi_{\text{int}})$ is given by

$$P_{V_T} s(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} s(T\boldsymbol{k}) \varphi_{\text{int}} \left(\frac{\boldsymbol{x}}{T} - \boldsymbol{k} \right), \tag{8}$$

with the property that $P_{V_T} P_{V_T} s(\boldsymbol{x}) = P_{V_T} s(\boldsymbol{x})$ (since P_{V_T} is a projection operator). To simplify the notation, we shall use a unit sampling $T = 1$ with the implicit assumption that the sampling error is negligible. (If the sampling error is large, one can use a finer sampling and rescale the reconstruction grid appropriately.) Thus, the resulting discretization is

$$s_1(\boldsymbol{x}) = P_{V_1} s(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} s[\boldsymbol{k}] \varphi_{\text{int}}(\boldsymbol{x} - \boldsymbol{k}). \tag{9}$$

To summarize, $s_1(\boldsymbol{x})$ is the discretized version of the original signal $s(\boldsymbol{x})$ and it is uniquely described by the samples $s[\boldsymbol{k}] = s(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{k}}$ for $\boldsymbol{k} \in \mathbb{Z}^d$. The main point is that the reconstructed signal is represented in terms of samples even though the problem is still formulated in the continuous-domain.

B. Discrete Measurement Model

By using the discretization scheme in (9), we are now ready to formally link the continuous model in Figure 1 and the corresponding discrete linear inverse problem. Although the signal representation (9) is an infinite sum, in practice we restrict ourselves to a subset of N basis functions with $\boldsymbol{k} \in \Omega$, where Ω is a discrete set of integer coordinates in a region of interest (ROI). Hence, we rewrite (9) as

$$s_1(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \Omega} s[\boldsymbol{k}] \varphi_{\boldsymbol{k}}(\boldsymbol{x}) \tag{10}$$

where $\varphi_{\boldsymbol{k}}(\boldsymbol{x})$ corresponds to $\varphi_{\text{int}}(\boldsymbol{x} - \boldsymbol{k})$ up to modifications at the boundaries (periodization or Neumann boundary condition).

We first consider a noise-free signal acquisition. The general form of a linear, continuous-domain noise-free measurement system is

$$z_m = \int_{\mathbb{R}^d} s(\boldsymbol{x}) \psi_m(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}, \quad (m = 1, \ldots, M) \tag{11}$$

where $s(\boldsymbol{x})$ is the original signal, and the measurement function $\psi_m(\boldsymbol{x})$ represents the spatial response of the mth detector which is application dependent as we shall explain in Section VI.

By substituting the signal representation (9) into (11), we discretize the measurement model and write it in matrix-vector form as

$$\mathbf{y} = \mathbf{z} + \mathbf{n} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{12}$$

where $\mathbf{y}$ is the M-dimensional measurement vector, $\mathbf{s} = (s[\boldsymbol{k}])_{\boldsymbol{k} \in \Omega}$ is the N-dimensional signal vector, $\mathbf{n}$ is the M-dimensional noise vector, and $\mathbf{H}$ is the $M \times N$ system matrix whose entry $(m, \boldsymbol{k})$ is given by

$$[\mathbf{H}]_{m,\boldsymbol{k}} = \langle \psi_m, \varphi_{\boldsymbol{k}} \rangle = \int_{\mathbb{R}^d} \psi_m(\boldsymbol{x}) \varphi_{\boldsymbol{k}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}. \tag{13}$$

This allows us to specify the discrete linear forward model given in (1) which is compatible with the continuous-domain formulation. The solution of this problem yields the representation $s_1(\boldsymbol{x})$ of $s(\boldsymbol{x})$ which is parameterized in terms of the signal samples $\mathbf{s}$. Having the forward model explained, our next aim is to obtain the statistical distribution of $\mathbf{s}$.

III. Sparse Stochastic Models

We now proceed by introducing our stochastic framework which will provide us with a signal prior. For that purpose, we assume that $s(\boldsymbol{x})$ is a realization of a stochastic process that is defined as the solution of a linear stochastic differential equation (SDE) with a driving term that is not necessarily Gaussian. Starting from such a continuous-domain model, we aim at obtaining the statistical distribution of the sampled version of the process (discrete signal) that will be needed to formulate estimators for the reconstruction problem.

A. Continuous-Domain Innovation Model

As mentioned in Section I, we specify our relevant class of signals as the solution of an SDE in which the process s is assumed to be whitened by a linear operator. This model takes the form

$$\mathrm{L}s = w, \tag{14}$$

where w is a continuous-domain white innovation process (the driving term), and L is a (multidimensional) differential operator. The right-hand side of (14) represents the unpredictable part of the process, while L is called the whitening operator. Such models are standard in the classical theory of stationary Gaussian processes [16]. The twist here is that the driving term w is not necessarily Gaussian. Moreover, the underlying differential system is potentially unstable to allow for self-similar models.

In the present model, the process s is characterized by the formal solution $s = \mathrm{L}^{-1} w$, where L^{-1} is an appropriate right inverse of L. The operator L^{-1} 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 shaping operator L^{-1}, whereas its statistical properties and sparsity structure is determined by the driving term w. As an example in the one-dimensional setting, the operator L can be chosen as the first-order continuous-domain derivative operator $\mathrm{L} = \mathrm{D}$. For multidimensional signals, an attractive class of operators is the fractional Laplacian $(-\Delta)^{\frac{\gamma}{2}}$ which is invariant to translation, dilation, and rotation

in $\mathbb{R}^d$ [17]. This operator gives rise to "$1/\|\boldsymbol{\omega}\|^{\gamma}$"-type power spectrum and is frequently used to model certain types of images [18], [19].

The mathematical difficulty is that the innovation w cannot be interpreted as an ordinary function because it is highly singular. The proper framework for handling such singular objects is Gelfand and Vilenkin's theory of generalized stochastic processes [20]. In this framework, the stochastic process s is observed by means of scalar-products $\langle s, \varphi \rangle$ with $\varphi \in \mathcal{S}(\mathbb{R}^d)$, where $\mathcal{S}(\mathbb{R}^d)$ denotes the Schwartz class of rapidly decreasing test functions. Intuitively, this is analogous to measuring an intensity value at a pixel after integration through a CCD detector.

A fundamental aspect of the theory is that the driving term w of the innovation model (14) is uniquely specified in terms of its Lévy exponent $f(\cdot)$.

Definition 1: A complex-valued function $f\colon \mathbb{R} \to \mathbb{C}$ is a valid Lévy exponent iff. it satisfies the three following conditions:

1) it is continuous;
2) it vanishes at the origin;
3) it is conditionally positive-definite of order one in the sense that
$$\sum_{m=1}^{N}\sum_{n=1}^{N} f(\omega_m - \omega_n)\xi_m\overline{\xi}_n \geq 0$$
under the condition $\sum_{m=1}^{N}\xi_m = 0$ for every possible choice of $\omega_1, \ldots, \omega_N \in \mathbb{R}$, $\xi_1, \ldots, \xi_N \in \mathbb{C}$, and $N \in \mathbb{N} \setminus \{0\}$.

An important subset of Lévy exponents are the p-admissible ones, which are central to our formulation.

Definition 2: A Lévy exponent f with derivative f' is called p-admissible if it satisfies the inequality
$$|f(\omega)| + |\omega||f'(\omega)| \leq C|\omega|^p$$
for some constant $C > 0$ and $0 < p \leq 2$.

A typical example of a p-admissible Lévy exponent is $f(\omega) = -s_0|\omega|^{\alpha}$ with $s_0 > 0$. The simplest case is $f_{\text{Gauss}}(\omega) = -\frac{1}{2}|\omega|^2$; it will be used to specify Gaussian processes.

Gelfand and Vilenkin have characterized the whole class of continuous-domain white innovation and have shown that they are fully specified by the generic characteristic form
$$\widehat{\mathscr{P}}_w(\varphi) = \mathbb{E}\left\{\mathrm{e}^{\mathrm{j}\langle w,\varphi\rangle}\right\} = \exp\left(\int_{\mathbb{R}^d} f(\varphi(\boldsymbol{x}))\mathrm{d}\boldsymbol{x}\right), \quad (15)$$
where f is the corresponding Lévy exponent of the innovation process w. The powerful aspect of this characterization is that $\widehat{\mathscr{P}}_w$ is indexed by a test 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.

Recently, Unser *et al.* characterized the class of stochastic processes that are solutions of (14) where L is a linear shift-invariant (LSI) operator and w is a member of the class of so-called Lévy noises [14, Theorem 3].

Theorem 1: Let w be a Lévy noise as specified by (15) and L^{-1*} be a left inverse of the adjoint operator L^* such that either one of the conditions below is met:

1) L^{-1*} is a continuous linear map from $\mathcal{S}(\mathbb{R}^d)$ into itself;
2) f is p-admissible and L^{-1*} is a continuous linear map from $\mathcal{S}(\mathbb{R}^d)$ into $L_p(\mathbb{R}^d)$; that is,
$$\|\mathrm{L}^{-1*}\varphi\|_{L_p} < C\|\varphi\|_{L_p}, \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d)$$
for some constant C and some $p \geq 1$.

Then, $s = \mathrm{L}^{-1}w$ is a well-defined generalized stochastic process over the space of tempered distributions $\mathcal{S}'(\mathbb{R}^d)$ and is uniquely characterized by its characteristic form
$$\widehat{\mathscr{P}}_s(\varphi) = \mathbb{E}\left\{\mathrm{e}^{\mathrm{j}\langle s,\varphi\rangle}\right\} = \exp\left(\int_{\mathbb{R}^d} f\big(\mathrm{L}^{-1*}\varphi(\boldsymbol{x})\big)\mathrm{d}\boldsymbol{x}\right). \quad (16)$$
It is a (weak) solution of the stochastic differential equation $\mathrm{L}s = w$ in the sense that $\langle \mathrm{L}s, \varphi\rangle = \langle w, \varphi\rangle$ for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$.

Before we move on, it is important to emphasize that Lévy exponents are in one-to-one correspondence with the so-called infinitely divisible (i.d.) distributions [21].

Definition 3: A generic pdf p_X is infinitely divisible if, for any positive integer n, it can be represented as the n-fold convolution $(p * \cdots * p)$ where p is a valid pdf.

Theorem 2 (Lévy-Schoenberg): Let $\hat{p}_X(\omega) = \mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega X}\} = \int_{\mathbb{R}} \mathrm{e}^{\mathrm{j}\omega x} p_X(x)\,\mathrm{d}x$ be the characteristic function of an infinitely divisible random variable X. Then,
$$f(\omega) = \log \hat{p}_X(\omega)$$
is a Lévy exponent in the sense of Definition 1. Conversely, if $f(\omega)$ is a valid Lévy exponent, then the inverse Fourier integral
$$p_X(x) = \int_{\mathbb{R}} \mathrm{e}^{f(\omega)}\mathrm{e}^{-\mathrm{j}\omega x}\frac{\mathrm{d}\omega}{2\pi}$$
yields the pdf of an i.d. random variable.

Another important theoretical result is that it is possible to specify the complete family of i.d. distributions thanks to the celebrated Lévy-Khintchine representation [22] which provides a constructive method for defining Lévy exponents. This tight connection will be essential for our formulation and limits us to a certain family of prior distributions.

B. Statistical Distribution of Discrete Signal Model

The interest is now to statistically characterize the discretized signal described in Section II-A. To that end, the first step is to formulate a discrete version of the continuous-domain innovation model (14). Since, in practical applications, we are only given the samples $s[\boldsymbol{k}]_{\boldsymbol{k}\in\Omega}$ of the signal, we obtain the discrete-domain innovation model by applying to them the discrete counterpart $\mathrm{L_d}$ of the whitening operator L. The fundamental requirement for our formulation is that the composition of $\mathrm{L_d}$ and L^{-1} results in a stable, shift-invariant operator whose impulse response is well localized [23]
$$\left(\mathrm{L_d}\mathrm{L}^{-1}\delta\right)(\boldsymbol{x}) = \beta_{\mathrm{L}}(\boldsymbol{x}) \in L_1(\mathbb{R}^d). \quad (17)$$
The function β_{L} is the generalized B-spline associated with the operator L. Ideally, we would like it to be maximally localized.

A necessary requirement for $\mathrm{L_d}$ is that its null space includes the one of L (see [23]).

To give more insight, let us consider $\mathrm{L} = \mathrm{D}$ and $\mathrm{L_d} = \mathrm{D_d}$ (the finite-difference operator associated to D). Then, the associated B-spline is $\beta_{\mathrm{D}}(x) = \mathrm{D_d}\mathbb{1}_+(x) = \mathbb{1}_+(x) - \mathbb{1}_+(x-1)$ where $\mathbb{1}_+(x)$ is the unit step (Heaviside) function. Hence, $\beta_{\mathrm{D}}(x) = \mathrm{rect}(x - \frac{1}{2})$ is a causal rectangle function (polynomial B-spline of degree 0).

The practical consequence of (17) is

$$u = \mathrm{L_d}s = \mathrm{L_d}\mathrm{L}^{-1}w = \beta_{\mathrm{L}} * w. \tag{18}$$

Since $(\beta_{\mathrm{L}} * w)(\boldsymbol{x}) = \langle w, \beta_{\mathrm{L}}^{\vee}(\cdot - \boldsymbol{x})\rangle$ where $\beta_{\mathrm{L}}^{\vee}(\boldsymbol{x}) = \beta_{\mathrm{L}}(-\boldsymbol{x})$ is the space-reversed version of β_{L}, it can be inferred from (18) that the evaluation of the samples of $\mathrm{L_d}s$ is equivalent to the observation of the innovation through a B-spline window.

From a system-theoretic point of view, $\mathrm{L_d}$ is understood as a finite impulse response (FIR) filter. This impulse response is of the form $\sum_{\boldsymbol{k}\in\Omega} d[\boldsymbol{k}]\delta(\cdot - \boldsymbol{k})$ with some appropriate weights $d[\boldsymbol{k}]$. Therefore, we write the discrete counterpart of the continuous-domain innovation variable as

$$u[\boldsymbol{k}] = \mathrm{L_d}s(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{k}} = \sum_{\boldsymbol{k}'\in\Omega} d[\boldsymbol{k}']s(\boldsymbol{k} - \boldsymbol{k}').$$

This allows us to write in matrix-vector notation the discrete-domain version of the innovation model (14) as

$$\mathbf{u} = \mathbf{L}\mathbf{s}, \tag{19}$$

where $\mathbf{s} = (s[\boldsymbol{k}])_{\boldsymbol{k}\in\Omega}$ represents the discretization of the stochastic model with $s[\boldsymbol{k}] = s(\boldsymbol{x})|_{\boldsymbol{x}=\boldsymbol{k}}$ for $\boldsymbol{k} \in \boldsymbol{\Omega}$, $\mathbf{L}$:$\mathbb{R}^N \to \mathbb{R}^N$ is the matrix representation of $\mathrm{L_d}$, and $\mathbf{u} = (u[\boldsymbol{k}])_{\boldsymbol{k}\in\Omega}$ is the discrete innovation vector.

We shall now rely on (16) to derive the pdf of the discrete innovation variable, which is one of the key results of this paper.

Theorem 3: Let s be a stochastic process whose characteristic form is given by (16) where f is a p-admissible Lévy exponent, and $\beta_{\mathrm{L}} = \mathrm{L_d}\mathrm{L}^{-1}\delta \in L_p(\mathbb{R}^d)$ for some $p \in (0, 2]$. Then, $u = \mathrm{L_d}s$ is stationary and infinitely divisible. Its first-order pdf is given by

$$p_U(u) = \int_{\mathbb{R}} \exp\left(f_{\beta_{\mathrm{L}}^{\vee}}(\omega)\right) \mathrm{e}^{\mathrm{j}\omega u} \frac{\mathrm{d}\omega}{2\pi}, \tag{20}$$

with Lévy exponent

$$f_{\beta_{\mathrm{L}}^{\vee}}(\omega) = \log \hat{p}_U(\omega) = \int_{\mathbb{R}^d} f\left(\omega\beta_{\mathrm{L}}^{\vee}(\boldsymbol{x})\right) \mathrm{d}\boldsymbol{x}, \tag{21}$$

which is p-admissible as well.

Proof: Taking (18) into account, we derive the characteristic form of u which is given by

$$\begin{aligned}\widehat{\mathscr{P}}_u(\varphi) &= \mathbb{E}\left\{\mathrm{e}^{\mathrm{j}\langle u,\varphi\rangle}\right\} = \mathbb{E}\left\{\mathrm{e}^{\mathrm{j}\langle \beta_{\mathrm{L}} * w,\varphi\rangle}\right\} = \mathbb{E}\left\{\mathrm{e}^{\mathrm{j}\langle w,\beta_{\mathrm{L}}^{\vee} * \varphi\rangle}\right\} \\ &= \widehat{\mathscr{P}}_w(\beta_{\mathrm{L}}^{\vee} * \varphi) \\ &= \exp\left(\int_{\mathbb{R}^d} f\left(\beta_{\mathrm{L}}^{\vee} * \varphi(\boldsymbol{x})\right) \mathrm{d}\boldsymbol{x}\right).\end{aligned} \tag{22}$$

The fact that u is stationary is equivalent to $\widehat{\mathscr{P}}_u(\varphi) = \widehat{\mathscr{P}}_u\left(\varphi(\cdot - \boldsymbol{x}_0)\right)$ for any $\boldsymbol{x}_0 \in \mathbb{R}^d$, which is established by a simple change of variable in (22). We now consider the random variable $U = \langle u, \delta\rangle = \langle w, \beta_{\mathrm{L}}^{\vee}\rangle$. Its characteristic function is obtained as

$$\begin{aligned}\hat{p}_U(\omega) &= \mathbb{E}\left\{\mathrm{e}^{\mathrm{j}\omega U}\right\} = \mathbb{E}\left\{\mathrm{e}^{\mathrm{j}\langle w,\omega\beta_{\mathrm{L}}^{\vee}\rangle}\right\} \\ &= \widehat{\mathscr{P}}_w(\omega\beta_{\mathrm{L}}^{\vee}) \\ &= \exp\left(f_{\beta_{\mathrm{L}}^{\vee}}(\omega)\right)\end{aligned}$$

where the substitution $\varphi = \omega\beta_{\mathrm{L}}^{\vee}$ in $\widehat{\mathscr{P}}_w(\varphi)$ is valid since $\widehat{\mathscr{P}}_w$ is a continuous functional on $L_p(\mathbb{R}^d)$ as a consequence of the p-admissibility condition. To prove that $f_{\beta_{\mathrm{L}}^{\vee}}(\omega)$ is a p-admissible Lévy exponent, we start by establishing the bound

$$\begin{aligned}C\|\varphi\|_{L_p}^p|\omega|^p &\geq \int_{\mathbb{R}^d} \left|f\left(\omega\beta_{\mathrm{L}}^{\vee}(\boldsymbol{x})\right)\right| \mathrm{d}\boldsymbol{x} \\ &\quad + |\omega| \int_{\mathbb{R}^d} \left|f'\left(\omega\beta_{\mathrm{L}}^{\vee}(\boldsymbol{x})\right)\varphi(\boldsymbol{x})\right| \mathrm{d}\boldsymbol{x} \\ &\geq \left|f_{\beta_{\mathrm{L}}^{\vee}}(\omega)\right| + |\omega| \left|f'_{\beta_{\mathrm{L}}^{\vee}}(\omega)\right|,\end{aligned} \tag{23}$$

which follows from the p-admissibility of f. We are also relying on Lebesgue's dominated convergence theorem to move the derivative with respect to ω inside the integral that defines $f_{\beta_{\mathrm{L}}^{\vee}}(\omega)$. In particular, (23) implies that $f_{\beta_{\mathrm{L}}^{\vee}}$ is continuous and vanishes at the origin. The last step is to establish its conditional positive definiteness which is achieved by interchanging the order of summation. We write

$$\begin{aligned}&\sum_{m=1}^{N}\sum_{n=1}^{N} f_{\beta_{\mathrm{L}}^{\vee}}(\omega_m - \omega_n)\xi_m\overline{\xi}_n = \\ &\int_{\mathbb{R}^d} \underbrace{\sum_{m=1}^{N}\sum_{n=1}^{N} f\left(\omega_m\beta_{\mathrm{L}}^{\vee}(\boldsymbol{x}) - \omega_n\beta_{\mathrm{L}}^{\vee}(\boldsymbol{x})\right)\xi_m\overline{\xi}_n}_{\geq 0} \mathrm{d}\boldsymbol{x} \geq 0\end{aligned} \tag{24}$$

under the condition $\sum_{m=1}^{N}\xi_m = 0$ for every possible choice of $\omega_1, \ldots, \omega_N \in \mathbb{R}$, $\xi_1, \ldots, \xi_N \in \mathbb{C}$, and $N \in \mathbb{N} \setminus \{0\}$. ■

The direct consequence of Theorem 3 is that the primary statistical features of $\mathbf{u}$ is directly related to the continuous-domain innovation process w via the Lévy exponent. This implies that the sparsity structure (tail behavior of the pdf and/or presence of a mass distribution at the origin) is primarily dependent upon f. The important conceptual aspect, which follows from the Lévy-Schoenberg theorem, is that the class of admissible pdfs is restricted to the family of i.d. laws since $f_{\beta_{\mathrm{L}}^{\vee}}(\omega)$, as given by (21), is a valid Lévy exponent. We emphasize that this result is attained by taking advantage of considerations in the continuous-domain.

C. Specific Examples

We now would like to illustrate our formalism by presenting some examples. If we choose $\mathrm{L} = \mathrm{D}$, then the solution of (14) with the boundary condition $s(0) = 0$ is given by

$$s(x) = \int_0^x w(x')\mathrm{d}x'$$

and is a Lévy process. It is noteworthy that the Lévy processes—a fundamental and well-studied family of stochastic processes—include Brownian motion and Poisson

TABLE I
FOUR MEMBERS OF THE FAMILY OF INFINITELY DIVISIBLE DISTRIBUTIONS AND THE CORRESPONDING POTENTIAL FUNCTIONS

	$p_U(x)$	$\Phi_U(x)$	Property
Gaussian	$\frac{1}{\sigma_0\sqrt{2\pi}}e^{-x^2/2\sigma_0^2}$	$M_1x^2+C_1$	smooth, convex
Laplace	$\frac{\tau}{2}e^{-\tau\|x\|}$	$M_2\|x\|+C_2$	nonsmooth, convex
Student's	$\frac{1}{\epsilon B(r,\frac{1}{2})}\left(\frac{1}{(x/\epsilon)^2+1}\right)^{r+\frac{1}{2}}$	$M_3\log\left(\frac{x^2+\epsilon^2}{\epsilon^2}\right)+C_3$	smooth, nonconvex
Cauchy	$\frac{1}{\pi s_0}\frac{1}{(x/s_0)^2+1}$	$\log\left(\frac{x^2+s_0^2}{s_0^2}\right)+C_4$	smooth, nonconvex

processes which are commonly used to model random physical phenomena [21]. In this case, $\beta_{\mathrm{D}}(x)=\mathrm{rect}(x-\frac{1}{2})$ and the discrete innovation vector $\mathbf{u}$ is obtained by

$$\begin{aligned} u[k] &= \langle w, \mathrm{rect}(\cdot+\tfrac{1}{2}-k)\rangle \\ &= s(k)-s(k-1). \end{aligned}$$

Since the B-splines are non-overlapping, we can deduce that the increments $(u[k])_{k\in\Omega}$ are i.i.d., which is the defining property of a Lévy process [21]. Evaluating (21) together with $f(0)=0$ (see Definition 1), we obtain

$$f_{\beta_{\mathrm{D}}^\vee}(\omega)=\int_{-1}^{0} f(\omega)\mathrm{d}x = f(\omega).$$

In particular, we generate Lévy processes with Laplace-distributed increments by choosing $f(\omega)=\log\left(\frac{\tau^2}{\tau^2+\omega^2}\right)$ with the scale parameter $\tau>0$. To see that, we write $\exp(f_{\beta_{\mathrm{D}}^\vee}(\omega))=\hat{p}_U(\omega)=\frac{\tau^2}{\tau^2+\omega^2}$ via (21). The inverse Fourier transform of this rational function is known to be

$$p_U(u)=\frac{\tau}{2}\mathrm{e}^{-\tau|u|}.$$

Also, we rely on Theorem 3 in a more general aspect. For instance, a special case of interest is the Gaussian (nonsparse) scenario where $f_{\mathrm{Gauss}}(\omega)=-\frac{1}{2}|\omega|^2$. Therefore, one gets $f_{\beta_{\mathrm{L}}^\vee}(\omega)=\log\hat{p}_U(\omega)=-\frac{1}{2}\omega^2\|\beta_{\mathrm{L}}\|_2^2$ from (21). Plugging this into (20), we deduce that the discrete innovation vector is zero-mean Gaussian with variance $\|\beta_{\mathrm{L}}\|_2^2$ (i.e., $p_U(u)=\mathcal{N}(0,\|\beta_{\mathrm{L}}\|_2^2)$).

Additionally, when $f(\omega)=\frac{-|\omega|^\alpha}{2}$ with $\alpha\in[1,2]$, one finds that $f_{\beta_{\mathrm{L}}^\vee}(\omega)=\log\hat{p}_U(\omega)=-\frac{|\omega|^\alpha}{2}\|\beta_{\mathrm{L}}\|_{L_\alpha}^\alpha$. This indicates that $\mathbf{u}$ is a symmetric α-stable ($S\alpha S$) distribution with scale parameter $s_0=\|\beta_{\mathrm{L}}\|_{L_\alpha}^\alpha$. For $\alpha=1$, we have the Cauchy distribution (or Student's with $r=1/2$). For other i.d. laws, the inverse Fourier transformation (20) is often harder to compute analytically, but it can still be performed numerically to determine $p_U(u)$ (or its corresponding potential function $\Phi_U=-\log p_U$). In general, p_U will be i.d. and will typically imply heavy tails. Note that heavy-tailed distributions are compressible [24].

IV. BAYESIAN ESTIMATION

We now use the results of Section III to derive solutions to the reconstruction problem in some well-defined statistical sense. To that end, we concentrate on the MAP solutions that are presently derived under the decoupling assumption that the components of $\mathbf{u}$ are independent and identically distributed (i.i.d.). This assumption is exact when L is a first-order differential operator (such as the derivative) in which case the B-spline is of unit support. For higher-order operators, the decoupling has local dependencies over the support of β_{L} that can be worked out explicitly [25]. However, taking these into account results in more complicated estimation algorithms.

In order to reconstruct the signal, 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 (since $\mathbf{u}=\mathbf{Ls}$). The direct application of Bayes' rule is

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

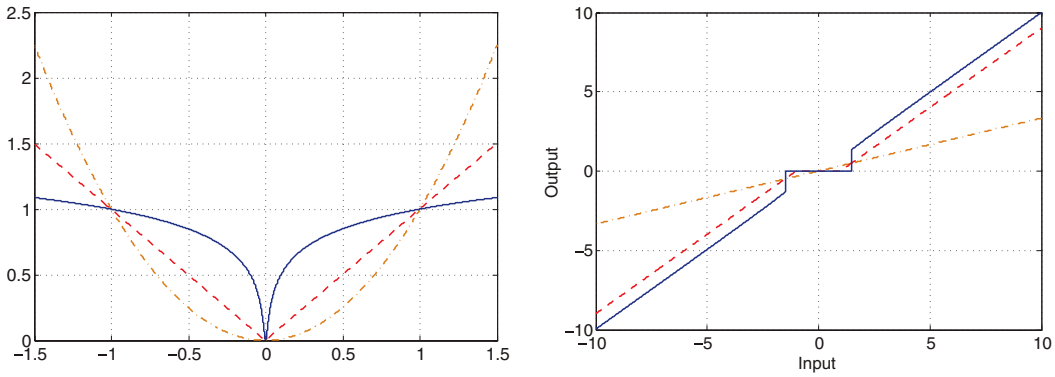
Then, we write the MAP estimation for $\mathbf{s}$ as

$$\begin{aligned} \mathbf{s}_{\mathrm{MAP}} &= \arg\max_{\mathbf{s}} p_{S|Y}(\mathbf{s}\mid\mathbf{y}) \\ &= \arg\min_{\mathbf{s}}\left(\tfrac{1}{2}\|\mathbf{Hs}-\mathbf{y}\|_2^2+\sigma^2\sum_{\boldsymbol{k}\in\Omega}\Phi_U\big([\mathbf{Ls}]_{\boldsymbol{k}}\big)\right), \quad (25) \end{aligned}$$

where $\Phi_U(x)=-\log p_U(x)$ is called the *potential function* corresponding to p_U. Note that (25) is compatible with the standard form of the variational reconstruction formulation given in (2). In the next section, we focus on the potential functions.

A. 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 $f_{\beta_{\mathrm{L}}^\vee}$ which is itself in direct relation with the continuous-domain innovation w via (21). For illustration purposes, we consider three members of the i.d. family: Gaussian, Laplace, and Student's (or, equivalently, Cauchy) distributions. We provide the potential functions for these priors in Table I. The exact values of the constants C_1, C_2, and C_3 and the positive scaling factors M_1, M_2, and M_3 have been omitted since they are irrelevant to the optimization problem. On one hand, we already know that the Gaussian prior does not correspond to a sparse reconstruction. On the other hand, the Student's prior has a slower tail decay and promotes sparser solutions than the Laplace prior. Also, to provide a geometrical intuition of how the Student's prior increases the sparsity of the solution, we plot the potential functions for Gaussian, Laplacian, and Student's (with $\epsilon=10^{-2}$) estimators



(a) (b)

Fig. 2. (a) Potential functions and (b) the corresponding proximity operators of different estimators. Dash-dotted line: Gaussian estimator. Dashed line: Laplacian estimator. Solid line: Student's estimator with $\epsilon = 10^{-2}$. The multiplication factors are set such that $\Phi_U(1) = 1$ for all potential functions.

in Figure 2. By looking at Figure 2, we see that the Student's estimator penalizes small values more than the Laplacian or Gaussian counterparts do. Conversely, it penalizes the large values less.

Let us point out some connections between the general estimator (25) and the standard variational methods. The first quadratic potential function (Gaussian estimator) yields the classical Tikhonov-type regularizer and produces a stabilized linear solution, as explained in Section I. The second potential function (Laplace estimator) provides the ℓ_1-type regularizer. Moreover, the well-known TV regularizer [7] is obtained if the operator L is a first-order derivative operator. Interestingly, the third log-based potential (Student's estimator) is linked to the limit case of the ℓ_p relaxation scheme as $p \to 0$ [26]. To see the relation, we note that minimizing $\lim_{p\to 0} \sum_i |x_i|^p$ is equivalent to minimizing $\lim_{p\to 0} \sum_i \frac{|x_i|^p - 1}{p}$. As pointed out in [26], it holds that

$$\lim_{p\to 0} \sum_i \frac{|x_i|^p - 1}{p} = \sum_i \log|x_i| = \sum_i \tfrac{1}{2}\log|x_i|^2 \leq \frac{1}{2}\sum_i \log(x_i^2 + \kappa) \tag{26}$$

for any $\kappa \geq 0$. The key observation is that the upper-bounding log-based potential function in (26) is interpretable as a Student's prior. This kind of regularization has been considered by different authors (see [27]–[29] and also [30] where the authors consider a similar log-based potential).

V. Reconstruction Algorithm

We have now the necessary elements to derive the general MAP solution of our reconstruction problem. By using the discrete innovation vector $\mathbf{u}$ as an auxiliary variable, we recast the MAP estimation as the constrained optimization problem

$$\mathbf{s}_{\text{MAP}} = \arg\min_{\mathbf{s}\in\mathbb{R}^K} \left(\frac{1}{2}\|\mathbf{H}\mathbf{s} - \mathbf{y}\|_2^2 + \sigma^2 \sum_{\boldsymbol{k}\in\Omega} \Phi_U\left(u[\boldsymbol{k}]\right) \right) \quad \text{subject to } \mathbf{u} = \mathbf{L}\mathbf{s}. \tag{27}$$

This representation of the solution naturally suggests using the type of splitting-based techniques that have been employed by various authors for solving similar optimization problems [31]–[33]. Rather than dealing with a constrained optimization problem directly, we prefer to formulate an equivalent unconstrained problem. To that purpose, we rely on the augmented-Lagrangian method [34] and introduce the corresponding *augmented Lagrangian* (AL) functional of (27) given by

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

where $\boldsymbol{\alpha} \in \mathbb{R}^K$ denotes the *Lagrange-multiplier* vector and $\mu \in \mathbb{R}$ is called the *penalty parameter*. The resulting optimization problem takes of the form

$$\min_{(\mathbf{s}\in\mathbb{R}^K,\ \mathbf{u}\in\mathbb{R}^K)} \mathcal{L}_{\mathcal{A}}(\mathbf{s},\mathbf{u},\boldsymbol{\alpha}). \tag{28}$$

Unlike the quadratic penalty approach [31], (28) does not require μ to be taken to infinity to guarantee convergence to the solution of (27). To obtain the solution, we apply the *alternating-direction method of multipliers* (ADMM) [35] that replaces the joint minimization of the AL functional over $(\mathbf{s},\mathbf{u})$ by the partial minimization of $\mathcal{L}_{\mathcal{A}}(\mathbf{s},\mathbf{u},\boldsymbol{\alpha})$ with respect to each independent variable in turn, while keeping the other variable fixed. These independent minimizations are followed by an update of the Lagrange multiplier. In summary, the algorithm results in the following scheme at iteration t:

$$\mathbf{u}^{t+1} \leftarrow \arg\min_{\mathbf{u}} \ \mathcal{L}_{\mathcal{A}}(\mathbf{s}^t,\mathbf{u},\boldsymbol{\alpha}^t) \tag{29a}$$

$$\mathbf{s}^{t+1} \leftarrow \arg\min_{\mathbf{s}} \ \mathcal{L}_{\mathcal{A}}(\mathbf{s},\mathbf{u}^{t+1},\boldsymbol{\alpha}^t) \tag{29b}$$

$$\boldsymbol{\alpha}^{t+1} = \boldsymbol{\alpha}^t + \mu(\mathbf{L}\mathbf{s}^{t+1} - \mathbf{u}^{t+1}). \tag{29c}$$

From the Lagrangian duality point of view, (29c) can be interpreted as the maximization of the dual functional so that, as the above scheme proceeds, feasibility is imposed [35].

Now, we focus on the sub–problem (29a). In effect, we see that the minimization is separable, which implies that (29a) reduces to performing K scalar minimizations of the form

$$\min_{u[\boldsymbol{k}]\in\mathbb{R}} \left(\sigma^2\Phi_U(u[\boldsymbol{k}]) + \frac{\mu}{2}\left(u[\boldsymbol{k}] - z[\boldsymbol{k}]\right)^2\right), \ \forall \boldsymbol{k} \in \Omega, \tag{30}$$

where $\mathbf{z} = \mathbf{L}\mathbf{s} + \boldsymbol{\alpha}/\mu$. One sees that (30) is nothing but the proximity operator of $\Phi_U(\cdot)$ that is defined below.

Definition 4: The proximity operator associated to the function $\lambda\Phi_U(\cdot)$ with $\lambda \in \mathbb{R}_+$ is defined as

$$\text{prox}_{\Phi_U}(y;\lambda) = \arg\min_{x\in\mathbb{R}} \frac{1}{2}(y-x)^2 + \lambda\Phi_U(x). \tag{31}$$

Consequently, (29a) is obtained by applying $\text{prox}_{\Phi_U}(z;\ \frac{\sigma^2}{\mu})$ in a *component-wise* fashion to $\mathbf{z} = \mathbf{L}\mathbf{s}^t + \boldsymbol{\alpha}^t/\mu$. The closed-form solutions for the proximity operator are well-known for the Gaussian and Laplace priors. They are given by

$$\text{prox}_{(\cdot)^2}(z;\lambda) = z(1+2\lambda)^{-1}, \tag{32a}$$

$$\text{prox}_{|\cdot|}(z;\lambda) = \max(|z|-\lambda, 0)\text{sgn}(z), \tag{32b}$$

respectively. The proximity operator has no closed-form solution for the Student's potential. For this case, we propose to precompute and store it in a lookup table (LUT) (see Figure 2(b)). This idea suggests a very fast implementation of the proximal step which is applicable to the entire class of i.d. potentials.

We now consider the second minimization problem (29b), which amounts to the minimization of a quadratic problem for which the solution is given by

$$\mathbf{s}^{t+1} = (\mathbf{H}^{\mathrm{T}}\mathbf{H} + \mu\mathbf{L}^{\mathrm{T}}\mathbf{L})^{-1}\left(\mathbf{H}^{\mathrm{T}}\mathbf{y} + \mu\mathbf{L}^{\mathrm{T}}\left(\mathbf{u}^{t+1} - \frac{\boldsymbol{\alpha}^t}{\mu}\right)\right). \tag{33}$$

Interestingly, this part of the reconstruction algorithm is equivalent to the Gaussian solution given in (4) and (5). In general, this problem can be solved iteratively using a linear solver such as the conjugate-gradient (CG) method. Also in some cases, the direct inversion is possible. For instance, when $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ has a convolution structure, as in some of our series of experiments, the direct solution can be obtained by using the FFT [36].

We conclude this section with some remarks regarding the optimization algorithm. We note that the computational complexity of the shrinkage step is $\mathcal{O}(N)$. Therefore, the dominating part of the computation is applying $\mathbf{H}$ and $\mathbf{H}^T$.

We also note that the method remains applicable when $\Phi_U(x)$ is nonconvex, with the following caveat: The convergence is not guaranteed. However, when the ADMM converges and Φ_U is nonconvex, it converges to a local minimum, including the case where the sub-minimization problems are solved exactly [35]. As the potential functions considered in the present context are closed and proper, we stress the fact that if $\Phi_U{:}\mathbb{R} \to \mathbb{R}_+$ is convex and the unaugmented Lagrangian functional has a saddle point, then the constraint in (27) is satisfied and the objective functional reaches the optimal value as $t \to \infty$ [35]. Meanwhile, in the case of nonconvex problems, the algorithm can converge to different local minimums depending on the initial point. It is therefore recommended to apply a deterministic continuation method or to consider a *warm start* that can be obtained by solving the problem first with Gaussian or Laplace priors. We have opted for the latter solution as an effective remedy for obtaining sparser solutions.

VI. Numerical Results

In the sequel, we illustrate our method with some concrete examples. We concentrate on three different imaging modalities and consider the problems of deconvolution, MR image reconstruction from partial Fourier coefficients, and image reconstruction from X-ray tomograms. For each of these problems, we present how the discretization paradigm is applied. In addition, our aim is to show that the adequacy of a given potential function is dependent upon the type of image being considered. Thus, for a fixed imaging modality, we perform model-based image reconstruction, where we highlight images that suit well to a particular estimator. For all the experiments, we choose $\mathbf{L}$ to be the discrete-gradient operator that is implemented by using forward finite-differences. In all the experiments, the signals extended using periodic boundary conditions. As a result, we update the proximal operators in Section V to their vectorial counterparts. The regularization parameters are optimized via an oracle to obtain the highest-possible SNR. The reconstruction is initialized in a systematic fashion: The solution of the Gaussian estimator is used as initial solution for the Laplace estimator and the solution of the Laplace estimator is used as initial solution for Student's estimator. The ϵ parameter for Student's estimator is set to 10^{-2}. The penalty parameter μ is set to 10λ for convex problems and $5\times10^3\lambda$ for the nonconvex ones. These heuristics are found to provide a good convergence speed.

A. Image Deconvolution

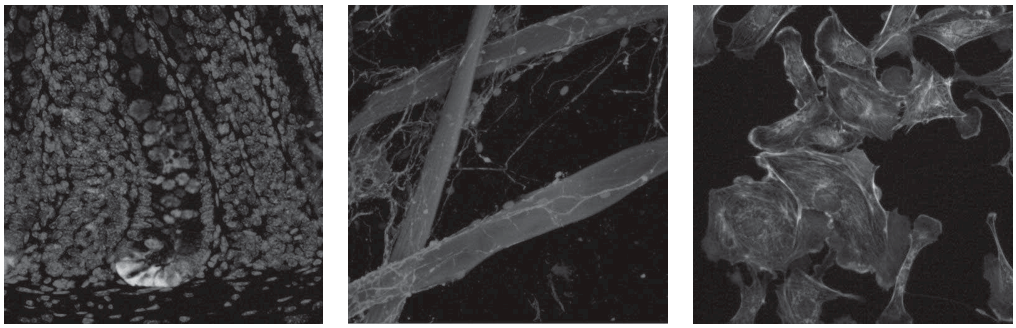
The first problem we consider is the deconvolution of microscopy images. In deconvolution, the measurement function in (11) corresponds to the shifted version of the point-spread function (PSF) of the microscope on the sampling grid: $\psi_m^{\mathrm{D}}(\boldsymbol{x}) = \psi^{\mathrm{D}}(\boldsymbol{x} - \boldsymbol{x}_m)$ where ψ^{D} represents the PSF. We discretize the model by choosing $\varphi_{\text{int}}(\boldsymbol{x}) = \text{sinc}(\boldsymbol{x})$ with $\varphi_{\boldsymbol{k}}(\boldsymbol{x}) = \varphi_{\text{int}}(\boldsymbol{x} - \boldsymbol{x}_{\boldsymbol{k}})$. The entries of the resulting system matrix $\mathbf{H}$ are given by

$$[\mathbf{H}]_{m,\boldsymbol{k}} = \langle \psi_m^{\mathrm{D}}(\cdot), \text{sinc}(\cdot - \boldsymbol{x}_{\boldsymbol{k}})\rangle, \tag{34}$$

In effect, (34) corresponds to the samples of the band-limited version of the PSF.

We perform controlled experiments, where the blurring of the microscope is simulated by a Gaussian PSF kernel of support 9×9 and standard deviation $\sigma_{\mathrm{b}} = 4$, on three microscopic images of size 512×512 that are displayed in Figure 3. In Figure 3(a), we show stem cells surrounded by numerous goblet cells. In Figure 3(b), we illustrate nerve cells growing along fibers, and we show in Figure 3(c) bovine pulmonary artery cells.

For deconvolution, the algorithm is run for a maximum of 500 iterations, or until the relative error between the successive iterates is less than 5×10^{-6}. Since $\mathbf{H}$ is block-Toeplitz, it can be diagonalized by a Fourier transform. Therefore, we use a direct FFT-based solver for (33). The measurements are degraded with different levels of AWGN. The noise variance σ^2 is determined through blurred SNR (BSNR) that is defined as $\text{BSNR} = \text{var}(\mathbf{H}\mathbf{s})/\sigma^2$. The results are summarized in Table II.



(a) (b) (c)

Fig. 3. Images used in deconvolution experiments. (a) Stem cells surrounded by goblet cells. (b) Nerve cells growing around fibers. (c) Artery cells.

TABLE II
DECONVOLUTION PERFORMANCE OF MAP ESTIMATORS BASED ON DIFFERENT PRIOR DISTRIBUTIONS

		Estimation Performance (SNR in dB)		
	BSNR (dB)	Gaussian	Laplace	Student's
Stem cells	20	**14.43**	13.76	11.86
Stem cells	30	**15.92**	15.77	13.15
Stem cells	40	**18.11**	**18.11**	13.83
Nerve cells	20	13.86	**15.31**	14.01
Nerve cells	30	15.89	**18.18**	15.81
Nerve cells	40	18.58	**20.57**	16.92
Artery cells	20	14.86	**15.23**	13.48
Artery cells	30	16.59	**17.21**	14.92
Artery cells	40	18.68	**19.61**	15.94

We conclude from the results of Table II that the MAP estimator based on a Laplace prior yields the best performance for images having sharp edges with a moderate amount of texture, such as those in Figures 3(b)–3(c). This confirms the observation that, by promoting solutions with sparse gradient, it is possible to improve the deconvolution performance. However, enforcing sparsity too heavily, as is the case for Student's priors, results in a degradation of the deconvolution performance for the biological images considered. Finally, for a heavily textured image like the one found in Figure 3(a), image deconvolution based on Gaussian priors yields the best performance. We note that the derived algorithms are compatible with the methods commonly used in the field (e.g., Tikhonov regularization [37] and TV regularization [38]).

B. MRI Reconstruction

We consider the problem of reconstructing MR images from undersampled $\boldsymbol{k}$-space trajectories. The measurement function represents a complex exponential at some fixed frequencies and is defined as $\psi_m^{\mathrm{M}}(\boldsymbol{x}) = \mathrm{e}^{2\pi \mathrm{j}\langle \boldsymbol{k}_m, \boldsymbol{x}\rangle}$ where $\boldsymbol{k}_m$ represents the sample point in $\boldsymbol{k}$-space. As in Section VI-A, we choose $\varphi_{\mathrm{int}}(\boldsymbol{x}) = \mathrm{sinc}(\boldsymbol{x})$ for the discretization of the forward model, which results in a system matrix with the entries

$$[\mathbf{H}]_{m,\boldsymbol{k}} = \langle \psi_m^{\mathrm{M}}(\boldsymbol{x}), \mathrm{sinc}(\cdot - \boldsymbol{x}_{\boldsymbol{k}})\rangle = \mathrm{e}^{-\mathrm{j}2\pi\langle \boldsymbol{k}_m, \boldsymbol{x}_{\boldsymbol{k}}\rangle} \quad \text{if} \quad |\boldsymbol{k}_m|_\infty \leq \tfrac{1}{2}. \tag{35}$$

The effect of choosing a sinc function is that the system matrix reduces to the discrete version of complex Fourier exponentials.

We study the reconstruction of the two MR images of size 256×256 illustrated Figure 4 – a cross-section of a wrist is displayed in the first image, followed by an MR angiography image—and consider a radial sampling pattern in $\boldsymbol{k}$-space (see Figure 4(c)).

The reconstruction algorithm is run with the stopping criteria set as in Section VI-A and an FFT-based solver is used for (33). We show in Table III the reconstruction performance of our estimators for different number of radial lines.

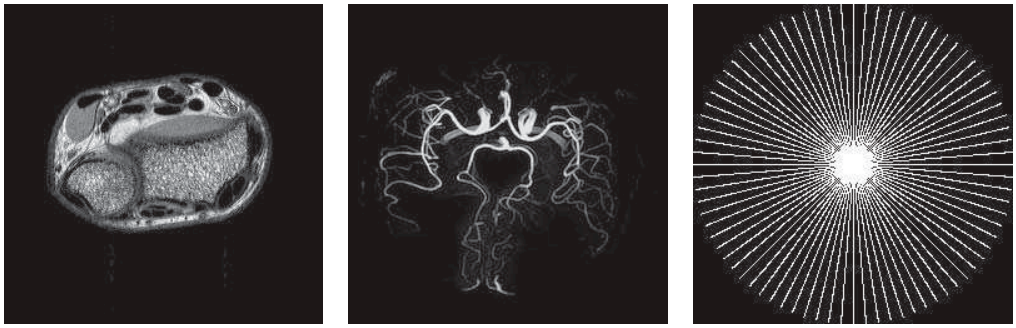
On one hand, the estimator based on Laplace priors yield the best solution in the case of the wrist image, which has sharp edges and some amount of texture. Meanwhile, the reconstructions using Student's priors are suboptimal because they are too sparse. This is similar to what was observed with microscopic images. On the other hand, Student's priors are quite suitable for reconstructing the angiogram, which is mostly composed of piecewise-smooth components. We also observe that the performance of Gaussian estimators is not competitive for the images considered. Our reconstruction algorithms are tightly linked with the deterministic approaches used for MRI reconstruction including TV [39] and log-based reconstructions [40].

C. X-Ray Tomographic Reconstruction

X-ray computed tomography (CT) aims at reconstructing an object from its projections taken along different directions. The mathematical model of a conventional CT is based on the Radon transform

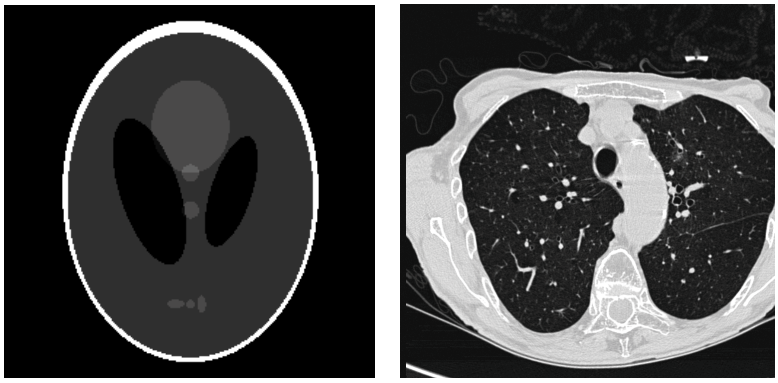
$$g_{\theta_m}(t_m) = \mathcal{R}_{\theta_m}\{s(\boldsymbol{x})\}(t_m) = \int_{\mathbb{R}^2} s(\boldsymbol{x})\delta(t_m - \langle \boldsymbol{x}, \boldsymbol{\theta_m}\rangle)\mathrm{d}\boldsymbol{x},$$

where $s(\boldsymbol{x})$ is the absorption coefficient distribution of the underlying object, t_m is the sampling point and $\boldsymbol{\theta}_m = (\cos(\theta_m), \sin(\theta_m))$ is the angular parameter. Therefore, the measurement function $\psi_m^{\mathrm{X}}(\boldsymbol{x}) = \delta(t_m - \langle \boldsymbol{x}, \boldsymbol{\theta}_m\rangle)$ denotes an idealized line in $\mathbb{R}^2$ perpendicular to $\boldsymbol{\theta}_m$. In our formulation, we represent the absorption distribution in the space spanned



(a) (b) (c)

Fig. 4. Data used in MR reconstruction experiments. (a) Cross section of a wrist. (b) Angiography image. (c) K-space sampling pattern along 40 radial lines.



(a) (b)

Fig. 5. Images used in X-ray tomographic reconstruction experiments. (a) SL phantom. (b) Cross section of the lung.

by the tensor product of two B-splines

$$s(\mathbf{x}) = \sum_{\mathbf{k}} s[\mathbf{k}]\varphi_{\text{int}}(\boldsymbol{x} - \boldsymbol{k}),$$

where $\varphi_{\text{int}}(\boldsymbol{x}) = \text{tri}(x_1)\text{tri}(x_2)$, with $\text{tri}(x) = (1 - |x|)$ denoting the linear B-spline function. The entries of the system matrix are then determined explicitly using the B-spline calculus described in [41], which leads to

$$\begin{aligned}[\mathbf{H}]_{m,\boldsymbol{k}} &= \langle \delta(t_m - \langle \boldsymbol{x}, \boldsymbol{\theta}_m \rangle), \varphi_{\text{int}}(\boldsymbol{x} - \boldsymbol{k}) \rangle \\ &= \frac{\triangle^2_{|\cos\theta_m|} \triangle^2_{|\sin\theta_m|}}{3!} (t_m - \langle \boldsymbol{k}, \boldsymbol{\theta}_m \rangle)^3_+,\end{aligned}$$

where $\triangle_h f(t) = \frac{f(t) - f(t-h)}{h}$ is the finite-difference operator, $\triangle^n_h f(t)$ is its n-fold iteration, and $t_+ = \max(0, t)$. This approach provides an accurate modeling, as demonstrated in [41] where further details regarding the system matrix, its implementation and execution time are provided.

We consider the two images shown in Figure 5. The Shepp-Logan (SL) phantom has size 256×256, while the cross section of the lung has size 750×750. In the simulations of the forward model, the Radon transform is computed along 180 and 360 directions for the lung image and along 120 and 180 directions for the SL phantom. The measurements are degraded with the Gaussian noise such that the signal-to-noise ratio is 20 dB.

TABLE III

MR IMAGE RECONSTRUCTION PERFORMANCE OF MAP ESTIMATORS BASED ON DIFFERENT PRIOR DISTRIBUTIONS

	Estimation Performance (SNR in dB)		
	Gaussian	Laplace	Student's
Wrist (20 radial lines)	8.82	**11.8**	5.97
Wrist (40 radial lines)	11.30	**14.69**	13.81
Angiogram (20 radial lines)	4.30	9.01	**9.40**
Angiogram (40 radial lines)	6.31	14.48	**14.97**

TABLE IV

RECONSTRUCTION RESULTS OF X-RAY COMPUTED TOMOGRAPHY USING DIFFERENT ESTIMATORS

	Estimation Performance (SNR in dB)		
	Gaussian	Laplace	Student's
SL Phantom (120 direction)	16.8	17.53	**18.76**
SL Phantom (180 direction)	18.13	18.75	**20.34**
Lung (180 direction)	**22.49**	21.52	21.45
Lung (360 direction)	**24.38**	22.47	22.37

For the reconstruction, we solve the quadratic minimization problem (33) iteratively by using 50 CG iterations. The reconstruction results are reported in Table IV.

The SL phantom is a piecewise-smooth image with sparse gradient. We observe that the imposition of more sparsity brought by Student's priors significantly improves the reconstruction quality for this particular image. On the other hand, we find that the Gaussian priors for the lung image outperform the other priors. Like the deconvolution and MRI problems, our algorithms are in line with Tikhonov-type [42] and TV [43] reconstructions used for X-ray CT.

D. Discussion

As our experiments on different types of imaging modalities have revealed, sparsity-promoting reconstructions are powerful methods for solving biomedical image reconstruction problems. However, encouraging sparser solutions does not always give the best reconstruction performance and nonsparse solutions provided by Gaussian priors still yield better reconstructions for certain images. The efficiency of a potential function is primarily dependent upon the type of image being considered. In our model, this is related to the Lévy exponent of the underlying continuous-domain innovation process w which is in direct relationship with the signal prior. The implicit assumption in this type of formulation is that the gradient of the image is independent from one location to the next, which is quite likely to be an oversimplification [44]. While the results obtained with this assumption are very promising, there is interest in investigating higher-order levels of dependencies.

VII. Conclusion

The purpose of this paper has been to develop a practical scheme for linear inverse problems by combining a proper discretization method and the theory of continuous-domain sparse stochastic processes. On the theoretical side, an important implication of our approach is that the potential functions cannot be selected arbitrarily as they are necessarily linked to infinitely divisible distributions. The latter puts restrictions but also constitutes the largest family of distributions that is closed under linear combinations of random variables. On the practical side, we have shown that the MAP estimators based on these prior distributions cover the current state-of-the-art methods in the field including ℓ_1-type regularizers. The class of said estimators is sufficiently large to reconstruct different types of images.

Another interesting observation is that we face an optimization problem for MAP estimation that is generally nonconvex, with the exception of the Gaussian and the Laplacian priors. We have proposed a computational solution, based on alternating-direction method of multipliers, that applies to arbitrary potential functions by suitable adaptation of the proximity operator.

In particular, we have applied our framework to deconvolution, MRI, and X-ray tomographic reconstruction problems and have compared the reconstruction performance of different estimators corresponding to models of increasing sparsity.

In basic terms, our model is composed of two fundamental concepts: the whitening operator L, which is in connection with the regularization operator, and the Lévy exponent f, which is related to the prior distribution. A further advantage of continuous-domain stochastic modeling is that it enables us to investigate the statistical characterization of the signal in any transform domain. This observation designates key research directions: (1) the identification of the optimal whitening operators and (2) the proper fitting of the Lévy exponent of the continuous-domain innovation process w to the class of images of interest.

References

[1] C. Vonesch, F. Aguet, J.-L. Vonesch, and M. Unser, "The colored revolution of bioimaging," *IEEE Signal Process. Mag.*, vol. 23, no. 3, pp. 20–31, May 2006.

[2] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. New York, NY, USA: Taylor & Francis, 1998.

[3] A. Ribes and F. Schmitt, "Linear inverse problems in imaging," *IEEE Signal Process. Mag.*, vol. 25, no. 4, pp. 84–99, Jul. 2008.

[4] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[5] S. Mallat, *A Wavelet Tour of Signal Processing*. San Francisco, CA, USA: Academic, 2008.

[6] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Resonance Med.*, vol. 58, no. 6, pp. 1182–95, Dec. 2007.

[7] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.

[8] J. F. Claerbout and F. Muir, "Robust modeling with erratic data," *Geophysics*, vol. 38, no. 5, pp. 826–844, Oct. 1973.

[9] H. L. Taylor, S. C. Banks, and J. F. McCoy, "Deconvolution with the ℓ_1 norm," *Geophysics*, vol. 44, no. 1, pp. 39–52, Jan. 1979.

[10] M. Zibulevsky and M. Elad, "L1-L2 optimization in signal and image processing," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 76–88, May 2010.

[11] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 296–310, Jul. 1993.

[12] H. Choi and R. Baraniuk, "Wavelet statistical models and Besov spaces," *Proc. SPIE*, vol. 3813, pp. 489–501, Jul. 1999.

[13] S. D. Babacan, R. Molina, and A. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–64, Jan. 2010.

[14] M. Unser, P. D. Tafti, and Q. Sun, "A unified formulation of Gaussian vs. sparse stochastic processes-part I: Continuous-domain theory," 2011, arXiv:1108.6150v1.

[15] M. Unser, "Sampling-50 years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.

[16] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York, NY, USA: McGraw-Hill, 1991.

[17] Q. Sun and M. Unser, "Left-inverses of fractional Laplacian and sparse stochastic processes," *Adv. Comput. Math.*, vol. 36, no. 3, pp. 399–441, Apr. 2012.

[18] B. B. Mandelbrot, *The Fractal Geometry of Nature*. San Francisco, CA, USA: Freeman, 1983.

[19] J. Huang and D. Mumford, "Statistics of natural images and models," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1999, pp. 637–663.

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

[21] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[22] F. W. Steutel and K. V. Harn, *Infinite Divisibility of Probability Distributions on the Real Line*. New York, NY, USA: Marcel Dekker, 2004.

[23] M. Unser, P. D. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian vs. sparse stochastic processes-part II: Discrete-domain theory," 2012, arXiv:1108.6152v1.

[24] A. Amini, M. Unser, and F. Marvasti, "Compressibility of deterministic and random infinite sequences," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5193–5201, Nov. 2011.

[25] A. Amini, U. S. Kamilov, E. Bostan, and M. Unser, "Bayesian estimation for continuous-time sparse stochastic processes," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 907–920, Feb. 2013.

[26] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 317–329, Apr. 2010.
[27] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar.–Apr. 2008, pp. 3869–3872.
[28] Y. Zhang and N. Kingsbury, "Fast L0-based sparse signal recovery," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Aug.–Sep. 2010, pp. 403–408.
[29] U. Kamilov, E. Bostan, and M. Unser, "Wavelet shrinkage with consistent cycle spinning generalizes total variation denoising," *IEEE Signal Process. Lett.*, vol. 19, no. 4, pp. 187–190, Apr. 2012.
[30] E. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, Dec. 2008.
[31] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM J. Imag. Sci.*, vol. 1, no. 3, pp. 248–272, Jul. 2008.
[32] S. Ramani and J. A. Fessler, "Regularized parallel MRI reconstruction using an alternating direction method of multipliers," in *Proc. IEEE Int. Symp. Biomed. Imag., From Nano Macro*, Mar.–Apr. 2011, pp. 385–388.
[33] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011.
[34] J. Nocedal and S. J. Wright, *Numerical Optimization*. New York, NY, USA: Springer-Verlag, 2006.
[35] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
[36] C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering*. Philadelphia, PA, USA: SIAM, 2006.
[37] C. Preza, M. I. Miller, and J.-A. Conchello, "Image reconstruction for 3D light microscopy with a regularized linear method incorporating a smoothness prior," *Proc. SPIE*, vol. 1905, pp. 129–139, Jul. 1993.
[38] N. Dey, L. Blanc-Féraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia, "Richardson–Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution," *Microscopy Res. Tech.*, vol. 69, no. 4, pp. 260–266, Apr. 2006.
[39] K. T. Block, M. Uecker, and J. Frahm, "Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraints," *Magn. Resonance Med.*, vol. 57, no. 6, pp. 1086–1098, Jun. 2007.
[40] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization," *IEEE Trans. Med. Imag.*, vol. 28, no. 1, pp. 106–121, Jan. 2009.
[41] A. Entezari, M. Nilchian, and M. Unser, "A box spline calculus for the discretization of computed tomography reconstruction problems," *IEEE Trans. Med. Imag.*, vol. 31, no. 7, pp. 1289–1306, Aug. 2012.
[42] J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography," *IEEE Trans. Med. Imag.*, vol. 25, no. 10, pp. 1272–1283, Oct. 2006.
[43] Z. Xiao-Qun and F. Jacques, "Constrained total variation minimization and application in computerized tomography," in *Proc. 5th Int. Conf. Energy Minimizat. Methods Comput. Vis. Pattern Recognit.*, 2005, pp. 456–472.
[44] A. J. Nasrallah and L. D. Griffin, "Gradient direction dependencies in natural images," *Spatial Vis.*, vol. 20, no. 3, pp. 277–299, 2007.

Emrah Bostan (S'11) received the B.Sc. degree in telecommunication engineering from Istanbul Technical University, Istanbul, Turkey, in 2009, and the M.Sc. degree in electrical engineering from École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2011. He is currently pursuing the Ph.D. degree with the Biomedical Imaging Group, EPFL.

His current research interests include inverse problems, variational models, and computational imaging.

Ulugbek S. Kamilov (S'11) received the M.Sc. degree in communications systems from the École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 2011, where he is currently pursuing the Ph.D. degree.

He was an Exchange Student in electrical and computer engineering with Carnegie Mellon University, Pittsburgh, PA, USA, from 2007 to 2008. He was a Research Intern with the Telecommunications Research Center, Vienna, Austria, in 2008. He was a Software Engineering Intern with Microsoft in 2009. He was a Visiting Student with the Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA, from 2010 to 2011. He joined the Biomedical Imaging Group, EPFL, in 2011. His current research interests include message-passing algorithms and the application of signal processing techniques to various biomedical problems.

Masih Nilchian (S'10) received the B.Sc. and M.Sc. degrees in electrical engineering (communication and information theory) from Sharif University of Technology, Tehran, Iran, in 2006 and 2008, respectively. He is currently pursuing the Ph.D. degree with the Biomedical Imaging Group, EPFL.

His current research interests include information theory, optimization algorithms, variational models, and their applications in biomedical imaging.

Michael Unser (M'89–SM'94–F'99) received the M.S. (*summa cum laude*) and Ph.D. degrees in electrical engineering from the École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland, in 1981 and 1984, respectively.

He was a Scientist with the National Institutes of Health, Bethesda, MD, USA, from 1985 to 1997. He is currently a full Professor and Director with the Biomedical Imaging Group, EPFL. His current research interests include biomedical image processing. He has a strong interest in sampling theories, multiresolution algorithms, wavelets, and the use of splines for image processing. He has authored over 200 journal papers on those topics.

Dr. Unser was an Associate Editor-in-Chief for the IEEE TRANSACTIONS ON MEDICAL IMAGING from 2003 to 2005, and has served as an Associate Editor for the IEEE TRANSACTIONS ON MEDICAL IMAGING from 1999 to 2002 and from 2006 to 2007, the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1992 to 1995, and the IEEE Signal Processing Letters from 1994 to 1998. He is currently a member of the editorial boards of the IEEE J. Selected Topics in Signal Processing, the Foundations and Trends in Signal Processing, and the Sampling Theory in Signal and Image Processing. He co-organized the first IEEE International Symposium on Biomedical Imaging (ISBI2002) and was the founding Chair of the Technical Committee of the IEEE-SP Society on Bio Imaging and Signal Processing. He was the recipient of three Best Paper Awards (1995, 2000, 2003) from the IEEE Signal Processing Society, and two IEEE Technical Achievement Awards (2008 SPS and 2010 EMBS). He is an EURASIP Fellow and a member of the Swiss Academy of Engineering Sciences.