MMSE Estimation of Sparse Lévy Processes

Ulugbek S. Kamilov, *Student Member, IEEE,* Pedram Pad, *Student Member, IEEE,* Arash Amini, and Michael Unser, *Fellow, IEEE*

***Abstract*—We investigate a stochastic signal-processing framework for signals with sparse derivatives, where the samples of a Lévy process are corrupted by noise. The proposed signal model covers the well-known Brownian motion and piecewise-constant Poisson process; moreover, the Lévy family also contains other interesting members exhibiting heavy-tail statistics that fulfill the requirements of compressibility. We characterize the maximum-a-posteriori probability (MAP) and minimum mean-square error (MMSE) estimators for such signals. Interestingly, some of the MAP estimators for the Lévy model coincide with popular signal-denoising algorithms (e.g., total-variation (TV) regularization). We propose a novel non-iterative implementation of the MMSE estimator based on the belief-propagation algorithm performed in the Fourier domain. Our algorithm takes advantage of the fact that the joint statistics of general Lévy processes are much easier to describe by their characteristic function, as the probability densities do not always admit closed-form expressions. We then use our new estimator as a benchmark to compare the performance of existing algorithms for the optimal recovery of gradient-sparse signals.**

***Index Terms*—Lévy process, stochastic modeling, sparse-signal estimation, non linear reconstruction, total-variation estimation, belief propagation (BP), message passing.**

I. INTRODUCTION

ESTIMATION of signals from incomplete or distorted measurements is a fundamental problem in signal processing. It inevitably arises during any realistic measurement process relying on some physical acquisition device.

Consider the problem of estimating a signal $\mathbf{x} \in \mathbb{R}^n$ from a noisy vector $\mathbf{y} = \mathbf{x} + \mathbf{n} \in \mathbb{R}^n$ where the components of $\mathbf{n}$ are independent and distributed with a known probability distribution. If we suppose that the components of the vector $\mathbf{x}$ are also independent, then the estimation problem becomes separable and reduces to n scalar estimation problems. In practice, however, due to correlations between the components of $\mathbf{x}$, simple pointwise techniques are suboptimal and more refined methods often perform significantly better. In this paper, we consider the problem of estimating signals with sparse derivatives. We take continuous-domain perspective and propose Lévy processes [1]–[4] as a natural approach to model such signals. The fundamental defining property of Lévy process is that it has independent and stationary increments. Therefore, the application of a finite-difference operator on samples of a Lévy process decouples it into a sequence of independent random variables. Interestingly, the class of Lévy processes is in one-to-one correspondence with the class of infinitely divisible distributions. Such distributions typically exhibit a heavy-tail behavior that has recently been proven to fulfill the requirements of compressibility [5]. Therefore, Lévy processes can be considered as the archetype of sparse stochastic signals [3].

The authors are with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne (EPFL), Station 17, CH-1015 Lausanne, Switzerland (e-mail: ulugbek.kamilov@epfl.ch; pedram.pad@epfl.ch; arash.amini@epfl.ch; michael.unser@epfl.ch)

A. Contributions

Many recent algorithms for the recovery of sparse signals can be interpreted as maximum-a-posteriori (MAP) estimators relying on some specific priors. From this Bayesian perspective, state-of-the-art methods based on gradient regularizers, such as total-variation (TV) [6] minimization, implicitly assume the signals to be sampled instances of Lévy processes [7, Section II]. In this paper, we investigate the minimum-mean-squared error (MMSE) estimator for Lévy processes. The performance of the estimator can be interpreted as a lower-bound on the MSE for the problem of recovery of gradient-sparse signals. Unfortunately, due to high-dimensional integration, MMSE estimators are computationally intractable for general signals. By considering the Lévy signal model, we propose a novel method for computing MMSE estimator based on the *belief-propagation* (BP) algorithm on cycle-free factor graphs [8]–[10].

The main contributions of this work are as follows:

- Bayesian formulation of the signal recovery problem under the Lévy hypothesis for a general "signal+noise" measurement model. With this formulation, we are able to derive an equivalence between MAP estimators for Lévy processes and some existing algorithms for the recovery of sparse signals.
- Characterization of the MSE optimal solution and the determination of performance bounds. We show that the MMSE estimator can be computed directly with the BP algorithm. The algorithm also obtains the marginals of the posterior distribution, which allows us to estimate the MSE of the reconstruction and provide confidence intervals.
- Development of a novel frequency-domain message-passing algorithm specifically tailored to the MMSE estimation of Lévy processes. Some of the sparsest priors considered here do not have closed-form probability density functions. Indeed, they are represented in terms of their characteristic function obtained by the Lévy-Khintchine theorem [1], [2]. The frequency-domain algorithm allows us to use the characteristic function directly without any numerical inversion.
- Experimental evaluation and comparison with standard solutions such as LMMSE, ℓ_1-minimization, and ℓ_p-relaxation [11]. In particular, the availability of MMSE allows us to benchmark these estimators on signals with desired properties such as sparsity.

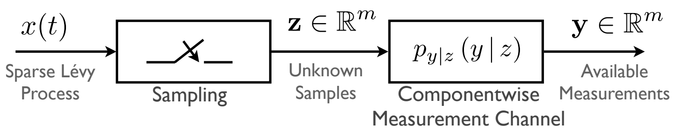

Fig. 1. Signal model considered in this work. The continuous-domain Lévy process $x(t)$ is sampled, and the resulting vector $\mathbf{z} \in \mathbb{R}^m$ is passed through a separable measurement channel $p_{y|z}(y|z)$ to yield $\mathbf{y} \in \mathbb{R}^m$. We investigate the estimation of interpolated vectors $\mathbf{x} \in \mathbb{R}^n, n \geq m$, from noisy measurements $\mathbf{y}$.

B. Outline

The paper is organized as follows: In Section II, we introduce our signal and measurement model. In particular, we review the theory of Lévy processes and their connection to sparse estimation. In Section III, we characterize the Bayesian MAP and MMSE estimators. In Section IV, we illustrate the connections between the MAP estimator for Lévy processes and standard variational approaches. In Section V, we provide closed-form formulae for evaluating the MMSE under the assumption of AWGN. In Section VI, we present the belief propagation algorithm as an efficient way to compute the MMSE estimator. We then introduce a new frequency-domain algorithm particularly well suited for Lévy processes. In Section VII, we provide numerical experiments demonstrating the applications of the method.

C. Notations

Throughout the paper, we typeset matrices in an uppercase boldface, vectors in a lowercase boldface, and scalars in italic typeface. Random and deterministic quantities are not distinguished typographically. We use p_x and $p_{y|x}$ to indicate probability distribution functions (pdf), and $\hat{p}_x$ and $\hat{p}_{y|x}$ to denote the corresponding characteristic functions. The pdf of a Gaussian random variable $x \sim \mathcal{N}(\mu, \sigma^2)$ will often be denoted as $\mathcal{G}(x-\mu;\sigma^2)$. The symbol $\stackrel{d}{=}$ indicates equality in distribution so that, for any two random variables, we have that $x \stackrel{d}{=} y$ if $\mathrm{Prob}\,(x \leq a) = \mathrm{Prob}\,(y \leq a)$ for all $a \in \mathbb{R}$.

II. Signal and Measurement Model

In this section, we describe the signal model summarized in Figure 1. We first give a powerful, yet simple continuous-domain stochastic formulation of the signal. The one-to-one mapping between our model and the extended family of infinitely divisible distributions is discussed. We finally describe the measurement model and provide examples of typical measurement channels.

A. Lévy Processes

Stochastic processes are often used to model random signals with the Brownian motion and the Poisson process being two most common examples. Lévy processes—often seen as analogues of random walks in continuous time—extend those two processes to a larger family of distributions. They represent a fundamental and well-studied class of stochastic processes [1], [2]. Let $\{x(t) : t \geq 0\}$ be a continuous-time stochastic process. It is called a Lévy process if

1) $x(0) = 0$ almost surely;
2) for each $n \in \mathbb{N}$ and $0 \leq t_1 < t_2 < \cdots < t_n < \infty$ the random variables $\{x(t_{k+1}) - x(t_k) : 1 \leq k \leq n-1\}$ are independent;
3) for each $0 \leq t_1 < t_2 < \infty$, the random variable $x(t_2) - x(t_1)$ is equal in distribution to $x(t_2 - t_1)$;
4) for all $\epsilon > 0$ and for all $t_1 \geq 0$

$$\lim_{t_2 \to t_1} \mathrm{Prob}\,(|x(t_2) - x(t_1)| > \epsilon) = 0.$$

Together, Properties 2) and 3) are commonly referred to as the *stationary-independent-increments* property, while Property 4) is called the *stochastic continuity*.

One of the most powerful results concerning Lévy processes is that they are in one-to-one correspondence with the class of infinitely divisible probability distributions. The random variable x is said to be *infinitely divisible* if, for any positive $n \in \mathbb{N}$, there exist i.i.d. random variables $y^{(1)}, \ldots, y^{(n)}$ such that

$$x \stackrel{d}{=} y^{(1)} + \cdots + y^{(n)}.$$

In other words, it must be possible to express the pdf p_x as the n-th convolution power of p_y. In fact, it is easy to show that the pdf of the increment $u_t = x(t+s) - x(s)$ of length t of any Lévy process is infinitely divisible

$$u_t \stackrel{d}{=} x(t) \stackrel{d}{=} u_{t/n}^{(1)} + \cdots + u_{t/n}^{(n)},$$

where each

$$u_{t/n}^{(k)} = x\left(\frac{kt}{n}\right) - x\left(\frac{(k-1)t}{n}\right).$$

The increments $u_{t/n}^{(k)}$ are of length t/n and are i.i.d. by the stationary-independent-increments property. Conversely, it has also been proved that there is a Lévy process for each infinitely divisible probability distribution [1].

The fundamental *Lévy-Khintchine formula* provides the characteristic function of *all* infinitely divisible distributions: p_u is an infinitely divisible probability distribution if and only if its characteristic function can be written as

$$\begin{aligned}\hat{p}_u(\omega) =& \mathbb{E}\left[\mathrm{e}^{\mathrm{j}\omega u}\right] \\ =& \exp\Big(\mathrm{j}a\omega - \tfrac{1}{2}b\omega^2 \\ &+ \int_{\mathbb{R}\setminus\{0\}} \left(\mathrm{e}^{\mathrm{j}\omega z} - 1 - \mathrm{j}z\omega\mathbb{1}_{|z|<1}(z)\right) v(z)\mathrm{d}z\Big),\end{aligned}$$

where $a \in \mathbb{R}, b \geq 0$, and where $\mathbb{1}_{|z|<1}$ is an indicator function. The function $v \geq 0$ is the *Lévy density* satisfying

$$\int_{\mathbb{R}\setminus\{0\}} \min\left(1, z^2\right) v(z)\mathrm{d}z < \infty.$$

The representation of $\hat{p}_u$ by a triplet $(a, b, v(\cdot))$ is unique. In this paper, we limit our attention to even-symmetric Lévy densities $v(z) = v(-z)$, which results in the simplified Lévy-Khintchine formula

$$\hat{p}_u(\omega) = \exp\Big(\mathrm{j}a\omega - \tfrac{1}{2}b\omega^2 + \int_{\mathbb{R}\setminus\{0\}} \left(\mathrm{e}^{\mathrm{j}\omega z} - 1\right) v(z)\mathrm{d}z\Big). \quad (1)$$

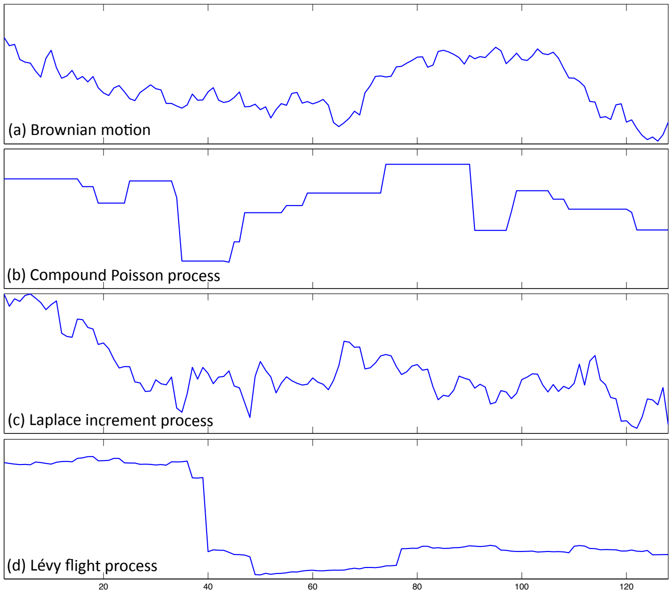

Fig. 2. Sample paths of Lévy processes discussed in this paper.

B. Examples of Lévy Processes

We now give examples of a few Lévy processes that are particularly interesting for us. Sample paths of these processes are summarized in Figure 2. Without loss of generality, we assume an increment $u = x(s) - x(s-1)$ for some fixed $s \geq 0$.

1) Brownian Motion: By setting $a = 0$ and choosing the Lévy density $v(z) = 0$, we obtain the familiar Brownian motion that has stationary independent increments characterized by

$$\hat{p}_u(\omega) = \mathrm{e}^{-\frac{1}{2}b\omega^2}, \tag{2}$$

with $b \geq 0$. This implies that the increments of the resulting Lévy process are Gaussian random variables with mean 0 and variance b, which corresponds to $u \sim \mathcal{N}(0, b)$. We illustrate in Figure 2(a) a single realization of a Brownian motion.

2) Compound Poisson Process: Let $\{z_k : k \in \mathbb{N}\}$ be a sequence of i.i.d. random variables with distribution p_z and let $n(t) \sim \mathrm{Poisson}(\lambda)$ be a Poisson process of intensity $\lambda > 0$ that does not depend on any z_k. The compound Poisson process y is then defined as

$$y(t) = \sum_{k=1}^{n(t)} z_k,$$

for each $t \geq 0$. This is a Lévy process obtained by setting the parameter triplet to $(0, 0, v(z) = \lambda p_z(z))$, which results in the characterization of increments

$$\hat{p}_u(\omega) = \mathrm{e}^{\lambda(\hat{p}_z(\omega)-1)}, \tag{3}$$

where $\hat{p}_z$ is the Fourier transform of p_z. On finite intervals, the sample paths of the process are piecewise-constant (Figure 2(b)), while the size of the jumps is determined by p_z [2]. Compound Poisson processes are piecewise-constant signals for which TV-like estimation algorithms are well suited [12]. The parameter λ controls the sparsity of the signal; it represents the rate of discontinuities. Compound Poisson processes are of special importance in Lévy-Itō decomposition of Lévy processes. The latter decomposition expresses any Lévy process as the sum of three processes, two of which are Brownian motion and Compound Poisson. More details are provided in Appendix A.

3) Laplace Increment Process: The Lévy process with Laplace-distributed increment u is obtained by setting the parameter triplet to $\left(0, 0, v(z) = \mathrm{e}^{-\gamma|z|}/|z|\right)$, which results in

$$\hat{p}_u(\omega) = \frac{\gamma^2}{\gamma^2 + \omega^2}, \tag{4}$$

where $\gamma > 0$ is the scale parameter of the Laplace distribution. To obtain the characteristic function (4), we remark that

$$\begin{aligned}\log\left(\hat{p}_u(\omega)\right) &= \int_{\mathbb{R}\setminus\{0\}} \left(\mathrm{e}^{\mathrm{j}\omega z} - 1\right) \frac{\mathrm{e}^{-\gamma|z|}}{|z|}\mathrm{d}z \\ &= 2\int_0^{\infty} \left(\cos\left(\omega z\right) - 1\right) \frac{\mathrm{e}^{-\gamma z}}{z}\mathrm{d}z.\end{aligned}$$

Then, by differentiation with respect to ω and integrating back using the condition $\hat{p}_u(0) = 1$, we obtain (4). The corresponding pdf is

$$p_u(u) = \frac{\gamma}{2}\mathrm{e}^{-\gamma|u|}. \tag{5}$$

An interesting observation is that the Bayesian MAP interpretation of the TV regularization method with a first-order finite-differences operator inherently assumes the underlying signal to be a Lévy process with Laplace increments. We give in Figure 2(c) an illustration of such process.

4) Lévy Flight Process: Stable random variables are such that a linear combination of two independent such random variables results in a third stable random variable [1]. In the symmetric case, they are often referred to as symmetric α-stable random variables and written as $u \sim S\alpha S$, where $0 < \alpha < 2$ is the stability parameter. It is possible to generate a Lévy process with α-stable increments by setting $\left(0, 0, v(z) = c_\alpha/|z|^{1+\alpha}\right)$, which results in

$$\hat{p}_u(\omega) = \mathrm{e}^{-\rho|\omega|^\alpha}, \tag{6}$$

with $\rho > 0$ and $0 < \alpha < 2$. It has been recently shown that such heavy-tail distributions result in highly compressible sequences [5]. A sample signal generated from a Cauchy increment Lévy flight, which corresponds to the α-stable process with $\alpha = 1$, is illustrated in Figure 2(d).

C. Innovation Modeling

Recently, an alternative system-theoretic formulation of Lévy processes was proposed in the context of the general theory of sparse stochastic processes [3], [4]. The authors specify the Lévy process $\{x(t) : t \geq 0\}$ as the solution of the stochastic differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}x(t) = w(t), \tag{7}$$

where the differentiation is interpreted in the weak sense of distributions. The process w is a non-Gaussian white noise referred to as *continuous-time innovation process*. According to the formalism developed in [3], the Lévy process is then generated by integrating the white noise according to

$$x(t) = \int_0^t w(t')\mathrm{d}t', \tag{8}$$

which provides a convenient linear-system interpretation. The delicate aspect is that the integrator is not BIBO stable and the white noise does not admit a classical interpretation as a function. The result confirms that, for all positive $k \in \mathbb{N}$, the quantities

$$\begin{aligned} u_k &= x(k) - x(k-1) = \mathrm{D_d} x(k) \\ &= \int_{k-1}^{k} w(t)\mathrm{d}t = \left\langle \mathrm{rect}\left(\cdot - k + \tfrac{1}{2}\right), w(\cdot)\right\rangle \end{aligned} \tag{9}$$

are i.i.d. random variables that can be seen as discrete innovations. The symbol $\langle\cdot,\cdot\rangle$ denotes an inner product between two functions, $\mathrm{D_d}$ is the finite-difference operator, and rect is the rectangular function. The fundamental observation is that the increment is obtained by applying the discrete version of the derivative to $x(t)$, in an attempt to emulate (7) using discrete means only.

D. Measurement Model

Consider the measurement model illustrated in Figure 1. The vector $\mathbf{z} \in \mathbb{R}^m$ contains uniformly sampled values of $x(t)$

$$z_i = x(iT_s), \quad i \in [1 \dots m], \tag{10}$$

where $T_s > 0$ is the sampling interval. The components of $\mathbf{y}$ are generated by a separable measurement channel given by the conditional probability distribution

$$p_{\mathbf{y}|\mathbf{z}}(\mathbf{y} \mid \mathbf{z}) = \prod_{i=1}^{m} p_{y|z}(y_i \mid z_i). \tag{11}$$

The measurement channel models distortions affecting the signal during the acquisition process. This paper addresses the computation of the estimator $\widehat{\mathbf{x}}$ of the vector $\mathbf{x} \in \mathbb{R}^n$ containing the samples of the original signal x on some uniform grid

$$x_k = x(kT_e), \quad k \in [1 \dots n], \tag{12}$$

where $T_e > 0$ is the interpolation interval. We wish to minimize the squared-error of the reconstruction in the situations when $T_s = m_s T_e$ for some positive $m_s \in \mathbb{N}$. This implies that in general $n \geq m$. The special case $n = m$ reduces the problem to signal denoising. In the sequel, we assume $T_s(m-1) = T_e(n-1)$ and set $T_e = 1$ to simplify the expressions. In particular, this implies that for any $m_s = T_s/T_e$ we have $z_i = x_{m_s(i-1)+1}$ for all $i \in [1 \dots m]$.

The generality of measurement channel allows us to handle both signal-dependent and independent distortions. Some common noise models encountered in practice are

1) Additive White Gaussian Noise (AWGN): The measurements in the popular AWGN noise model are given by $\mathbf{y} = \mathbf{z} + \mathbf{n}$, where $\mathbf{n} \in \mathbb{R}^m$ is a signal independent-Gaussian vector with i.i.d components $n_i = y_i - z_i \sim \mathcal{N}(0, \sigma^2)$. The transitional probability distribution then reduces to

$$p_{y|z}(y \mid z) = \mathcal{G}(y - z; \sigma^2). \tag{13}$$

2) Scalar Quantization: Another common source of signal distortion is the analog-to-digital converter (ADC). When the conversion corresponds to a simple mapping of the analog voltage input to some uncoded digital output, it can be modeled as standard AWGN followed by a lossy mapping $\mathrm{Q} : \mathbb{R} \to \mathcal{C}$. The nonlinear function Q is often called a K-level scalar quantizer [13]. It maps the K-partitions of the real line $\left\{\mathrm{Q}^{-1}(c_i) : i = 1, \dots, K\right\} \subseteq \mathbb{R}$ into the set of discrete output levels $\mathcal{C} = \{c_i : i = 1, \dots, K\}$. This channel is signal-dependent. It is described in terms of the transitional probability distribution

$$p_{y|z}(y \mid z) = \int_{\mathrm{Q}^{-1}(y)} \mathcal{G}(z' - z; \sigma^2)\mathrm{d}z', \tag{14}$$

where $\mathrm{Q}^{-1}(y) = \{z \in \mathbb{R} : \mathrm{Q}(z) = y\}$ denotes a single partition.

III. Bayesian Formulation

We now specify explicitly the class of problems we wish to solve and identify corresponding statistical estimators. Consider the vector $\mathbf{u} \in \mathbb{R}^n$ obtained by applying the finite-difference matrix $\mathbf{D}$ to $\mathbf{x}$ in (12). Then, from the stationary independent increments property of Lévy processes the components

$$u_k = [\mathbf{D}\mathbf{x}]_k = x_k - x_{k-1}, \tag{15}$$

of the vector $\mathbf{u}$ are realizations of i.i.d. random variables characterized by the simplified Lévy-Khintchine formula (1). Note that, from the definition of the Lévy process we have $x_0 = 0$. We construct the conditional probability distribution for the signal $\mathbf{x}$ given the measurements $\mathbf{y}$ as

$$\begin{aligned} p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y}) &\propto p_{\mathbf{y}|\mathbf{x}}(\mathbf{y} \mid \mathbf{x})\, p_{\mathbf{x}}(\mathbf{x}) \\ &\propto \prod_{i=1}^{m} p_{y|z}(y_i \mid z_i) \prod_{k=1}^{n} p_u\left([\mathbf{D}\mathbf{x}]_k\right), \end{aligned} \tag{16}$$

where we use $\propto$ to denote identity after normalization to unity. The distribution of the whitened elements p_u is, in principle, obtained by taking the inverse Fourier transform $p_u(u) = \mathcal{F}^{-1}\{\hat{p}_u\}(u)$; however, it does not necessarily admit a closed-form formula. The posterior distribution (16) of the signal provides a complete statistical characterization of the problem. In particular, the MAP and MMSE estimators of $\mathbf{x}$ are specified by

$$\widehat{\mathbf{x}}_{\mathrm{MAP}} = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmax}} \left\{p_{\mathbf{x}|\mathbf{y}}(\mathbf{x} \mid \mathbf{y})\right\} \tag{17}$$

$$\widehat{\mathbf{x}}_{\mathrm{MMSE}} = \mathbb{E}[\mathbf{x} \mid \mathbf{y}]. \tag{18}$$

Finding efficient methods to evaluate (17) and (18) is a common challenge encountered in signal processing.

IV. MAP Estimation

An estimation based on the minimization of some cost functional is a popular way of obtaining the MAP estimator $\widehat{\mathbf{x}}_{\mathrm{MAP}}$. The availability of efficient numerical methods for convex and nonconvex optimization partially explain the success of such methods [12], [14]–[16]. The MAP estimator in (17) can be reformulated as

$$\begin{aligned}\widehat{\mathbf{x}}_{\mathrm{MAP}} &= \operatorname*{argmax}_{\mathbf{x}\in\mathbb{R}^n}\left\{p_{\mathbf{x}|\mathbf{y}}\left(\mathbf{x}\,|\,\mathbf{y}\right)\right\} \\ &= \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^n}\left\{-\log\left(p_{\mathbf{x}|\mathbf{y}}\left(\mathbf{x}\,|\,\mathbf{y}\right)\right)\right\} \\ &= \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^n}\left\{\mathcal{D}\left(\mathbf{x},\mathbf{y}\right)+\mathcal{R}\left(\mathbf{x}\right)\right\},\end{aligned} \tag{19a}$$

where

$$\mathcal{D}\left(\mathbf{x},\mathbf{y}\right) = -\sum_{i=1}^{m}\log\left(p_{y|z}(y_i\,|\,z_i)\right), \tag{19b}$$

$$\mathcal{R}\left(\mathbf{x}\right) = -\sum_{k=1}^{n}\log\left(p_u\left([\mathbf{D}\mathbf{x}]_k\right)\right). \tag{19c}$$

The term $\mathcal{D}(\cdot)$ is the *data term* and $\mathcal{R}(\cdot)$ the *regularization term*.

In the AWGN, model the MAP estimation reduces to the popular regularized least-squares minimization problem

$$\widehat{\mathbf{x}}_{\mathrm{MAP}} = \operatorname*{argmin}_{\mathbf{x}\in\mathbb{R}^n} \tfrac{1}{2}\left\|\mathbf{y}-\mathbf{z}\right\|^2 + \sigma^2\sum_{k=1}^{n}\phi_u([\mathbf{D}\mathbf{x}]_k), \tag{20}$$

where $\mathbf{z}\in\mathbb{R}^m$ is given in (10) and $\phi_u(x) = -\log\left(p_u(x)\right)$ is the *potential function*.

The estimator in (20) clearly illustrates the connections between the standard variational methods and our stochastic model. In particular, in the framework of the Lévy process, the Brownian motion yields the classical Tikhonov regularizer. The Lévy process with Laplace increments provides the ℓ_1-based TV regularizer. Finally, the Lévy flight process results in a log-based regularizer that is linked to the limit case of the ℓ_p relaxation as p tends to zero [11]. Such regularizers have been shown to be effective in several problems of the recovery of sparse signals [12], [15]. In [17] the authors have proposed an efficient method for solving the regularized-least-squares-based MAP denoising of Lévy processes. We also point out that the MAP estimation of compound-Poisson processes yields a trivial solution due to a point mass at zero.

V. MMSE Estimation in AWGN

In this section, we present some theoretical results related to MMSE denoising. For detailed derivation of the results, we refer the reader to Appendix B. Consider AWGN denoising problem

$$\mathbf{y} = \mathbf{z}+\mathbf{n} \qquad \text{with} \qquad \mathbf{z} = \mathbf{x}, \tag{21}$$

where each noise component $n_i \sim \mathcal{N}(0,\sigma^2)$. Then, for any distribution on $\mathbf{x}$, it is possible to characterize the MMSE estimator as

$$\widehat{\mathbf{x}}_{\mathrm{MMSE}} = \mathbf{y} + \sigma^2\;\boldsymbol{\nabla}\log p_{\mathbf{y}}(\mathbf{y}), \tag{22}$$

where $\boldsymbol{\nabla}$ denotes the gradient and $p_{\mathbf{y}}$ is the pdf of the noisy vector $\mathbf{y}$ [18], [19]. Note that $p_{\mathbf{y}}$ is given by the convolution $p_{\mathbf{y}} = p_{\mathbf{x}} * p_{\mathbf{n}}$, where $p_{\mathbf{x}}$ is the prior and $p_{\mathbf{n}}$ is the pdf of the AWGN. Then, MMSE of the estimation problem is given by

$$\begin{aligned}\mathrm{MMSE}(n) &= \frac{1}{n}\mathbb{E}\left[\left\|\mathbf{x}-\widehat{\mathbf{x}}_{\mathrm{MMSE}}\right\|^2\right] \\ &= \sigma^2 + \frac{\sigma^4}{n}\int p_{\mathbf{y}}(\mathbf{y})\Delta\log p_{\mathbf{y}}(\mathbf{y})\mathrm{d}\mathbf{y},\end{aligned} \tag{23}$$

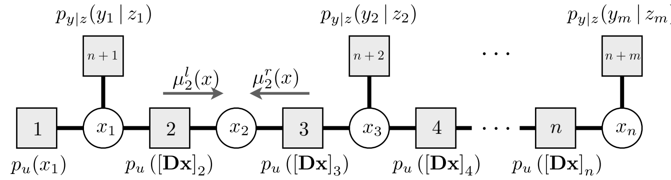

Fig. 3. Factor-graph representation of the posterior distribution (16) with $m_s = 2$. In the graph, square factor nodes represent the probability densities and circled variable nodes represent the unknowns. Functions μ_2^l and μ_2^r represent beliefs at the variable node 2.

where Δ is the Laplacian with respect to $\mathbf{y}$.

Although elegant, Equations (22) and (23) are not tractable for arbitrary distributions on $\mathbf{x}$. In the special case of Brownian motion, where the increments are Gaussian random variables, the MMSE estimator reduces to the well-known Wiener filter, which is commonly referred to as linear minimum mean-square error (LMMSE) estimator. As described in Appendix B, by using the Central Limit theorem argument in the Karhunen-Loève Transform (KLT) domain, it is possible to obtain the following asymptotic description of the MMSE

$$\lim_{n\rightarrow\infty}\mathrm{MMSE}(n) = \frac{\sigma^2}{\sqrt{1+4\frac{\sigma^2}{\sigma_u^2}}}. \tag{24}$$

In general, (24) is not equivalent to the MMSE for non-Gaussian increments; however it still corresponds to the performance of LMMSE estimator.

VI. Message Passing Estimation

A. Exact Formulation

In this section, we specify the MMSE estimator $\widehat{\mathbf{x}}_{\mathrm{MMSE}}$ in (18) for the signals under the Lévy-process model. Unfortunately, due to the high-dimensionality of the integral, this estimation is intractable in the direct form. However, several computational methods exist for computing this integral. We adopt the *sum-product belief-propagation* (BP) [8] method, which efficiently approximates the computationally intractable direct marginalization of the posterior (16). The BP-based message-passing methods have successfully been used in numerous inference problems in statistical physics, computer vision, channel coding, and signal processing [8]–[10], [20]–[25].

In order to apply the BP, we construct the bipartite factor-graph $G = (V,F,E)$, structured according to the posterior distribution in (16). We illustrate in Figure 3 an example of a factor-graph for $m_s = 2$. The graph consists of two sets of nodes, the *variable nodes* $V = \{1,\dots,n\}$ (circles), the *factor nodes* $F = \{1,\dots,n+m\}$ (squares), and a set of edges E linking variables to the factors they participate in. To introduce the BP algorithm, we define the functions μ_i^l and μ_i^r, which denote the messages exchanged along the edges of the graph. These messages—often referred to as *beliefs*—are in fact pdfs representing the desirable state of the variable node i. We also define for all $i\in[1\dots n]$ and $j = 1+(i-1)/m_s$ the function

$$\eta_i(x) = \begin{cases} p_{y|z}(y_j\,|\,x), & \text{when } j\in\mathbb{N} \\ 1, & \text{otherwise.}\end{cases} \tag{25}$$

Whenever the component x_i has a corresponding measurement, the function η_i is equivalent to the channel pdf. Otherwise, η_i is equivalent to the constant function.

Given the measurements $\mathbf{y} \in \mathbb{R}^m$ and the functions η_i and p_u, the steps of the BP estimation are:

1) *Initialization:* Set

$$\mu_1^l(x) = p_u(x), \tag{26a}$$
$$\mu_n^r(x) = 1. \tag{26b}$$

2) *Message Updates:* For $i = 1, \ldots, n-1$, compute

$$\mu_{i+1}^l(x) \propto \int_{\mathbb{R}} p_u(x-z)\eta_i(z)\mu_i^l(z)\mathrm{d}z, \tag{27a}$$
$$\mu_{n-i}^r(x) \propto \int_{\mathbb{R}} p_u(z-x)\eta_j(z)\mu_j^r(z)\mathrm{d}z, \tag{27b}$$

where $j = n-i+1$. The symbol $\propto$ denotes identity after normalization to unity. Since the pdf p_u is symmetric, the expressions can be rewritten in terms of the convolutions $\mu_{i+1}^l \propto p_u * \eta_i\mu_i^l$ and $\mu_{n-i}^r \propto p_u * \eta_j\mu_j^r$.

3) *Result:* For $i = 1, \ldots, n$, compute

$$[\widehat{\mathbf{x}}_{\mathrm{MMSE}}]_i = \int_{\mathbb{R}} x p_{x_i|\mathbf{y}}(x\,|\,\mathbf{y})\,\mathrm{d}x, \tag{28a}$$

where the marginal pdf is obtained by

$$p_{x_i|\mathbf{y}}(x\,|\,\mathbf{y}) \propto \mu_i^l(x)\mu_i^r(x)\eta_i(x). \tag{28b}$$

The proposed update rules recursively marginalize the posterior distribution, reducing intractable high-dimensional integration into $2n$ convolutions. It is well-known that BP gives exact marginal probabilities for all the nodes in any singly connected graph. Consequently, for our problem the solution of the algorithm coincides with $\widehat{\mathbf{x}}_{\mathrm{MMSE}}$.

B. Fourier-Domain Alternative

The BP algorithm presented in Section VI-A assumes availability of a closed-form expression for the pdf p_u. Unfortunately this form is often unavailable, since the distribution is defined by its characteristic function $\hat{p}_u$ obtained by the Lévy-Khintchine formula (1). When the general shape of the pdf is unknown, a naïve numerical evaluation of the inverse Fourier-transform of the characteristic function can lead to unexpected results. As an example, consider the compound Poisson process. The characteristic function (3), describing the distribution of the increments, does not generally admit a closed-form expression of its inverse Fourier transform. Moreover, it results in a pdf containing a probability mass (a Dirac delta function) at zero, which needs to be taken into account explicitly for a correct numerical inversion.

Fortunately, the BP algorithm presented above can readily be performed in the frequency domain. The message-update equations are obtained by the convolution property of the Fourier transform, which amounts to switching the role of multiplications and convolutions in (27) and (28b). The final estimation step is also simplified by applying the moment property

$$\int_{\mathbb{R}} x^n f(x)\mathrm{d}x = \mathrm{j}^n \left.\frac{\mathrm{d}^n}{\mathrm{d}\omega^n}\hat{f}(\omega)\right|_{\omega=0}, \tag{29}$$

where $\hat{f}(\omega) = \int_{\mathbb{R}} f(x)\mathrm{e}^{-\mathrm{j}\omega x}\mathrm{d}x$ is the Fourier transform of f.

1) *Initialization:* Set

$$\hat{\mu}_1^l(\omega) = \hat{p}_u(\omega), \tag{30a}$$
$$\hat{\mu}_n^r(\omega) = \delta(\omega), \tag{30b}$$

where δ is the Dirac delta function.

2) *Message updates:* For $i = 1, \ldots, n-1$, compute

$$\hat{\mu}_{i+1}^l(\omega) \propto \hat{p}_u(\omega)\cdot(\hat{\eta}_i * \hat{\mu}_i^l)(\omega), \tag{31a}$$
$$\hat{\mu}_{n-i}^r(\omega) \propto \hat{p}_u(\omega)\cdot(\hat{\eta}_j * \hat{\mu}_j^r)(\omega), \tag{31b}$$

where $j = n-i+1$. The symbol $\propto$ denotes identity after normalization by the zero frequency component. Note that, functions $\hat{\eta}_i$ represent the Fourier transform of (25).

3) *Result:* For $i = 1, \ldots, n$, compute

$$[\widehat{\mathbf{x}}_{\mathrm{MMSE}}]_i = \mathrm{j}\left.\frac{\mathrm{d}}{\mathrm{d}\omega}\hat{p}_{x_i|\mathbf{y}}(\omega\,|\,\mathbf{y})\right|_{\omega=0}, \tag{32a}$$

where the characteristic function $\hat{p}_{x_i|\mathbf{y}}(\omega\,|\,\mathbf{y})$ of the marginalized posterior is obtained by

$$\hat{p}_{x_i|\mathbf{y}}(\omega\,|\,\mathbf{y}) \propto \left(\hat{\mu}_i^l * \hat{\mu}_i^r * \hat{\eta}_i\right)(\omega). \tag{32b}$$

Note that (32a) and (32b) can be evaluated with a single integral. This is achieved by reusing convolutions in (31) and evaluating the derivative only at zero.

C. Implementation

In principle, the BP equations presented above yield the exact MMSE estimator for our problem. However, due to the existence of continuous-time integrals in the updates, they cannot be implemented in the given form. To obtain a realizable solution, we need to choose some practical discrete parameterization for the messages exchanged in the algorithm. The simplest and the most generic approach is to sample the functions and represent them on a uniform grid with finitely many samples. In our implementation, we fix the support set of the functions to $[-N_{\delta,\epsilon}\delta, N_{\delta,\epsilon}\delta]_{\mathbb{Z}}$. $\epsilon > 0$ controls the truncation. We keep only samples such that $f(x) \geq \epsilon$, thus, the total number of samples for representing the function depends on both the truncation parameter ϵ and on the sampling step $\delta > 0$. It is given by $M_{\delta,\epsilon} = 2N_{\delta,\epsilon} + 1$. The proper parameter values depend on the distribution to represent and on the measurements $\mathbf{y}$. Then, both time- and frequency-domain versions can be obtained by approximating continuous integrals by standard quadrature rules. In our implementation, we use Riemann sum to approximate the integrals.

VII. Experimental Results

In this section, we present several experiments with the goal of comparing various signal-estimation methods. The performance of the estimator is judged based on MSE reduction given by

$$\mathrm{MSE} = 10\log_{10}\left(\frac{1}{n}\|\mathbf{x} - \widehat{\mathbf{x}}\|^2\right), \tag{33}$$

where $\mathbf{x}, \widehat{\mathbf{x}} \in \mathbb{R}^n$.

We concentrate on the four Lévy processes discussed in Section II-B and set the parameters of these processes as

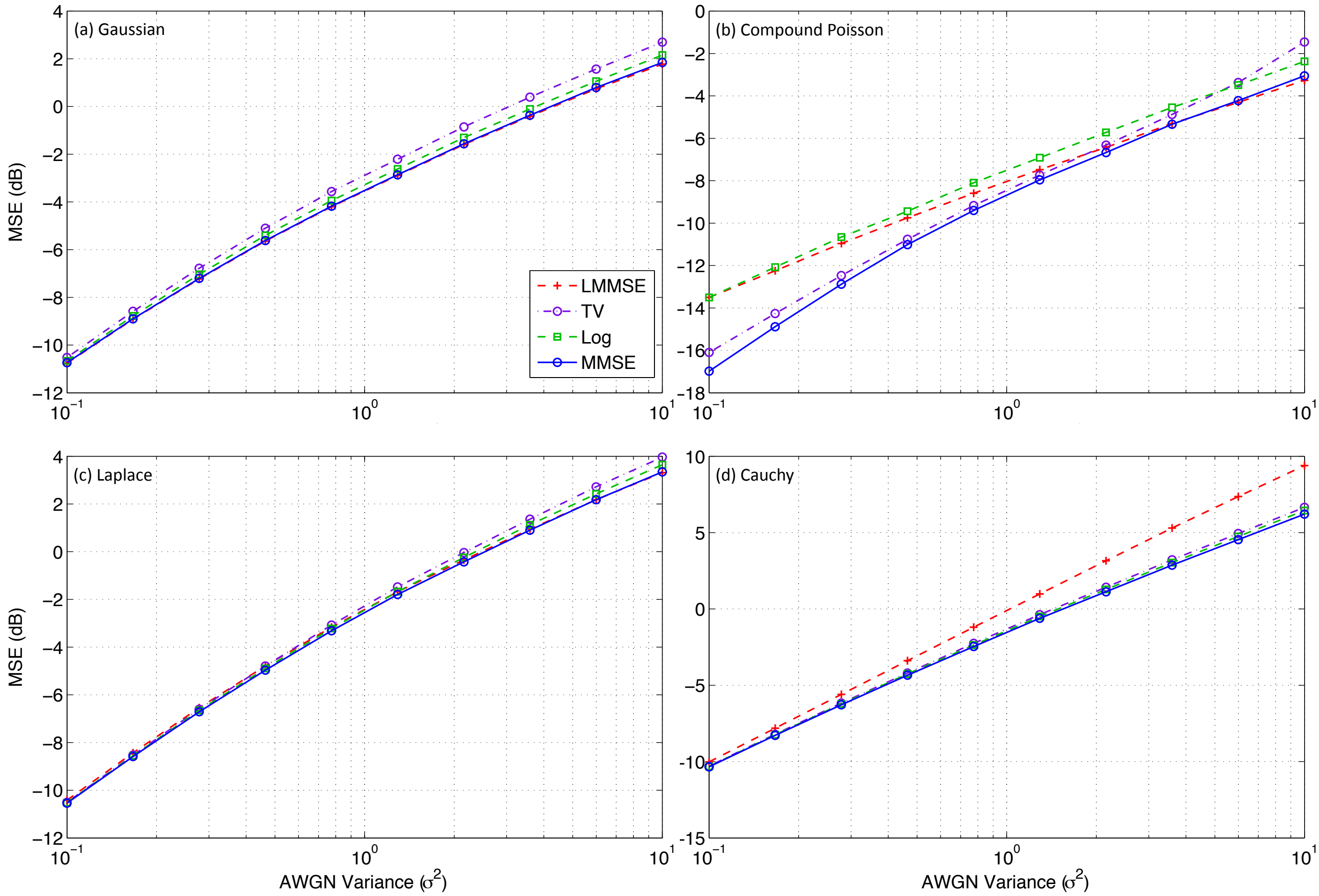

Fig. 4. AWGN denoising of (a) Brownian motion, (b) compound Poisson, (c) Lévy process with Laplace increments, (d) Lévy flight processes. We compare LMMSE, TV regularization, Log regularization, and MMSE. (a) Since linear estimation is optimal for Gaussian signals, both LMMSE and MMSE yield identical results. (b) The excellent performance of TV regularization is confirmed for piecewise-constant signals. (c) TV regularization results in a MAP estimator, which leads to a suboptimal MSE performance. (d) Log regularization amounts to the MAP estimator. For such processes, linear estimators over-smooth the edges resulting in poor performance.

- *Brownian Motion:* The increments are generated from a standard Gaussian distribution with $u_k = [\mathbf{D}\mathbf{x}]_k \sim \mathcal{N}(0,1)$.
- *Compound-Poisson Process:* We concentrate on sparse signals and set the mass probability to $P(u_k = 0) = \mathrm{e}^{-\lambda} = 0.9$. The size of the jumps follow the standard Gaussian distribution.
- *Laplace Increment Process:* The increments are generated from the Laplace distribution of scale $\gamma = 1$.
- *Lévy Flight:* We set the distribution of the increments to be Cauchy ($\alpha = 1$) with scale parameter $\rho = 1$.

A. AWGN Denoising

In the first set of experiments, we consider the denoising of Lévy processes in AWGN. We compare the performance of several popular estimation methods over a range of noise levels σ^2. In Figures 4(a)–(d), we perform 1000 random realization of the denoising problem for each value of σ^2 and plot the average MSE reduction after estimation. The signal length is set to $n = m = 200$. The proposed message-passing estimator is compared with the regularized least-squares estimators

$$\widehat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}} \ \tfrac{1}{2}\left\|\mathbf{y}-\mathbf{x}\right\|^2 + \tau \sum_{k=1}^{n} \phi_u([\mathbf{D}\mathbf{x}]_k), \qquad (34)$$

where $\mathbf{D}$ is the finite-difference matrix and $\tau > 0$ is the regularization parameter optimized for the best MSE performance.

The curve labeled LMMSE corresponds to the MSE optimal linear estimator, which can be obtained by setting the potential function $\phi_u(x) = x^2$ [26]. The TV method corresponds to the potential function $\phi_u(x) = |x|$ and can be efficiently implemented by using the FISTA algorithm described in [16]. The Log estimator corresponds to the potential function $\phi_u(x) = \log\left(x^2+\epsilon\right)$, where the parameter $\epsilon > 0$ controls the sparsity of the signal. Log-based regularizers have been shown to outperform traditional ℓ_1-based regularizers in various applications [12], [15]. In our experiments, we fix $\epsilon = 1$, which corresponds to the MAP estimator for the Lévy flight process with Cauchy increments. Efficient implementation of the Log-based denoising was obtained by using the algorithm [17].

It is well known that the LMMSE estimator is linear and optimal for Brownian motion. In Figure 4(a), it is precisely matched by the message-passing MMSE estimator. Moreover,

TABLE I
MMSE PREDICTION.

Prior	Noise (σ^2)	Oracle MSE	Predicted MSE
Gaussian	0.1	-10.74 dB	$-10.73 \pm 5.4 \times 10^{-5}$ dB
	1	-3.54 dB	$-3.49 \pm 5.9 \times 10^{-5}$ dB
	10	1.85 dB	$1.95 \pm 6.5 \times 10^{-5}$ dB
Cauchy	0.1	-10.37 dB	-10.34 ± 0.03 dB
	1	-1.54 dB	-1.53 ± 0.11 dB
	10	6.15 dB	6.22 ± 0.21 dB

we have observed that—even for $n = 200$—the asymptotic prediction (24) closely matches the simulation results (within 0.06 dB). Since the curve for the asymptotic prediction is hidden under LMMSE and MMSE, we have omitted it from Figure 4(a). The worst performance is observed for TV regularization, which yields piecewise-constant solutions by removing small variations of the signal. The performance of the Log-based method is significantly better; it preserves important details by allowing small variations of the signal.

In Figure 4(b), we observe excellent MSE performance of TV for compound Poisson processes over many noise levels. It is well known that TV estimators yield a piecewise-constant solution, which makes it ideally matched for such signals. In this experiment, we have also measured the average running times for all the algorithms. For example, for $\sigma^2 = 1$ the average estimation times for LMMSE, TV, Log, and MMSE were 0.03, 0.05, 0.01, and 0.29 seconds, respectively. The theoretical implications of the compound Poisson process is extensively discussed in [7].

In Figure 4(c), we observe a surprisingly poor performance of TV, which corresponds to the MAP estimator for Lévy processes with Laplace increments. This highlights the fact that, in some situations, a MAP estimator can result in suboptimal MSE performance.

In Figure 4(d), we observe that LMMSE performs poorly for a Lévy flight process. It fails to preserve signal edges, which results in a suboptimal MSE performance for all noise levels. Both TV and Log methods are known to be edge-preserving. In fact, they obtain solutions close to the MMSE estimator (within 0.2 dB for Log). For such signals, Log-based regularizers yield the MAP estimator.

The message passing algorithm considered in this paper computes the marginals of the posterior distribution. The algorithm yields the MMSE estimator by finding the mean of the marginalized distribution. But the posterior distribution actually provides much more information. For example, the algorithm can predict the MSE of the reconstruction by computing the variance of the posterior

$$\mathrm{Var}\left[x_k \,|\, \mathbf{y}\right] = \mathbb{E}\left[x_k^2 \,|\, \mathbf{y}\right] - \left(\left[\widehat{\mathbf{x}}_{\mathrm{MMSE}}\right]_k\right)^2,$$

where $\left[\widehat{\mathbf{x}}_{\mathrm{MMSE}}\right]_k$ is given in (32). The second moment can be evaluated by using the moment property (29).

The capability to predict the MSE of the reconstruction is useful to complement the solution of the estimator with a confidence interval. In Table I, the MSE predicted by the algorithm is presented for Gaussian and Cauchy increment processes. For comparison, we also provide the oracle MSE obtained by comparing the true signal $\mathbf{x}$ with $\widehat{\mathbf{x}}$. The average predicted MSE is obtained from 1000 random realizations of the problem. The table also provides the standard deviation of the predicted MSE values around the mean. This illustrates the accuracy of the predicted MSE values across noise levels.

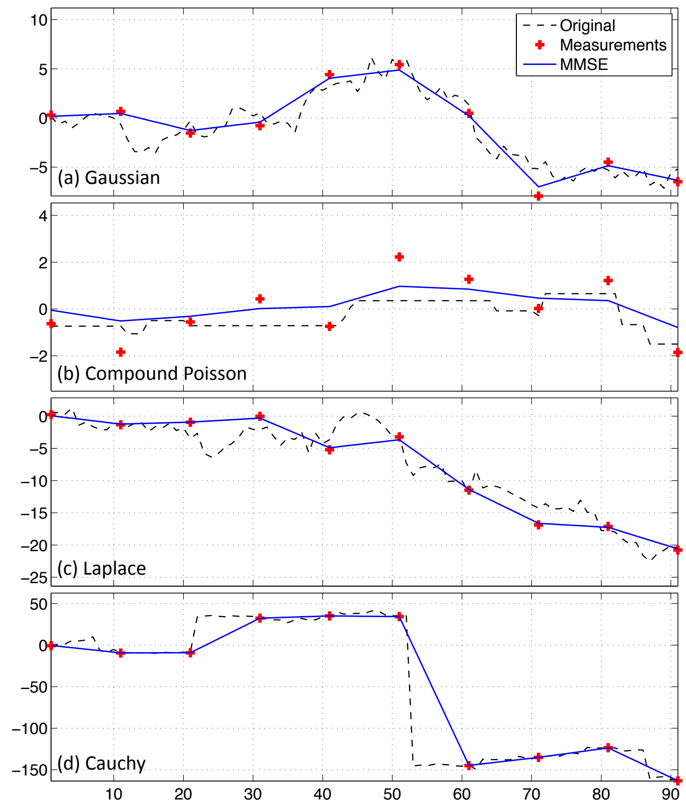

Fig. 5. Ten-fold interpolation of Lévy processes from AWGN measurements. From top to bottom: (a) Brownian motion; (b) compound Poisson process; (c) Lévy process with Laplace increments; (d) Lévy flight process. Surprisingly, for all priors the optimal estimator appears to be a piecewise linear function.

B. Signal Interpolation

In Figure 5, we illustrate the interpolation of Lévy processes from noisy measurements. We assume AWGN of variance $\sigma^2 = 1$ and set the interpolation rate to $m_s = T_s/T_e = 10$. Given 10 noisy measurements, this results in 91 estimated values. In the topmost graph, the signal is a Brownian motion with increments of unit variance. The second graph illustrates the interpolation of a compound Poisson process. The third graph illustrates the interpolation of a Lévy process with Laplace increments. Finally, the bottom-most graph illustrates the interpolation of a Lévy flight. An interesting observation is that the MSE optimal interpolator seems to yield piecewise linear results independently of the process considered. In fact, it is known that, for the Brownian motion, piecewise-linear interpolation is optimal [27]. Note that this does not imply that the estimator is itself linear—in general, it is not.

In Table II, we compare the MSE performance of message-passing estimators with linear estimators for the interpolation problem with $m_s = 2$. Each value in the table is obtained by averaging over 1000 problem instances. Note that the Lévy

TABLE II
INTERPOLATION OF LÉVY PROCESSES: MSE FOR DIFFERENT NOISE LEVELS.

Prior	**Noise (σ^2)**	**LMMSE**	**MMSE**
Gaussian	0.1 dB	−4.9315 dB	−4.9315 dB
	1 dB	−1.3866 dB	−1.3866 dB
	10 dB	3.4221 dB	3.4221 dB
Compound Poisson	0.1 dB	−11.3233 dB	−12.7016 dB
	1 dB	−6.3651 dB	−6.8164 dB
	10 dB	−1.5267 dB	−1.6012 dB
Laplace	0.1 dB	−2.4691 dB	−2.4724 dB
	1 dB	0.2644 dB	0.2279 dB
	10 dB	4.9509 dB	4.9406 dB

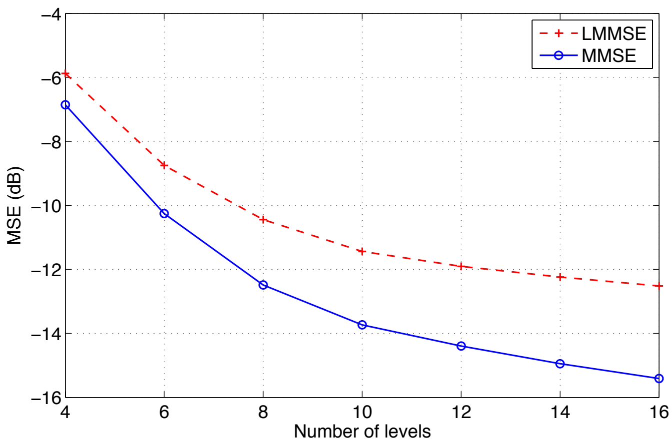

Fig. 6. Estimation of the compound Poisson process from quantized measurements. We compare the standard LMMSE against MMSE, thereby illustrating the suboptimality of standard linear reconstructions.

flight process was omitted from the table. For the interpolation problem, the average estimation MSE for this process is not defined and can only be characterized conditioned on a given $\mathbf{y}$.

C. Estimation from Quantized Samples

We next consider the highly non linear problem of estimating Lévy processes from quantized measurements (14). We generate the compound Poisson process of length $n = 200$. An AWGN of variance 0.1 is added to the signal prior to quantization. The quantizer is uniform with granular region of length $2\|\mathbf{y}\|_\infty$. It is centered at the origin.

In Figure 6, we compare the MSE performance of the message-passing estimator with the standard LMMSE estimator. The parameter τ of the linear estimator was optimized for the best MSE performance. In this figure, we plot the mean of the MSE from 1000 problem instances for several quantization levels K. For such non linear measurement channels, the message-passing estimator yields significant improvements in the reconstruction performance over the standard linear estimator.

VIII. CONCLUSION

We have presented an in-depth investigation of the Lévy-process framework for modeling signals with sparse derivatives. We have also characterized the corresponding statistical estimators. Lévy processes are fundamental members of a recently proposed continuous-domain stochastic framework for modeling sparse signals. We have presented a simple message-passing algorithm for the MMSE estimation of Lévy processes from noisy measurements. The proposed algorithm can handle a large class of priors, including those that do not have closed-form pdfs. Moreover, it can incorporate a large class of noise distributions, provided that the noise components are independent among themselves. The algorithm has also the ability to handle signal-dependent noise. Due to the tree-like structure of the underlying factor graph, when the messages are continuous-time functions the message-passing algorithm obtains the MMSE estimator of the signal. This motivates its application as a benchmark to judge the optimality of various existing gradient-based estimators including TV- and Log-regularization algorithms.

APPENDIX A
LÉVY-ITŌ DECOMPOSITION

For a Lévy process $x(t)$, let $\hat{p}_{x(t)}(\omega)$ denote the characteristic function of the random variable $x(t)$. Furthermore, let $(a, b, v(\cdot))$ be the Lévy-Khintchine triplet associated with the random variable $x(1)$, which has the same distribution as $x(t+1) - x(t)$, for all t.

For arbitrary integers m, n, the two representations of the random variable $x(m)$ written as

$$\begin{aligned} x(m) &= \sum_{i=0}^{n-1} x\Big((i+1)\frac{m}{n}\Big) - x\Big(i\frac{m}{n}\Big) \\ &= \sum_{i=0}^{m-1} x(i+1) - x(i) \end{aligned} \tag{35}$$

show that

$$\Big(\hat{p}_{x(\frac{m}{n})}(\omega)\Big)^n = \Big(\hat{p}_{x(1)}(\omega)\Big)^m. \tag{36}$$

By using the continuity property in the definition of Lévy processes, we can further generalize (36) to

$$\hat{p}_{x(t)}(\omega) = \Big(\hat{p}_{x(1)}(\omega)\Big)^t. \tag{37}$$

This suggests the Lévy-Khintchine triplet $\big(t\,a, t\,b, t\,v(\cdot)\big)$ for the random variable $x(t)$. The triplet can be decomposed as

$$\underbrace{(ta, tb, 0)}_{\text{BM}} + \underbrace{\big(0, 0, tv_1(\cdot)\big)}_{\text{CP}} + \underbrace{\big(0, 0, tv_2(\cdot)\big)}_{\text{PJ}}, \tag{38}$$

where v_1 is an absolutely integrable function, v_2 is a pure singular distribution, and $v = v_1 + v_2$. The latter decomposition is achieved by adapting Lebesgue's decomposition theorem for distributions corresponding to measures. In (38), the term BM reveals the Lévy-Khintchine triplet of a Brownian motion with non-zero mean. Similarly, since v_1 is integrable, the term CP reflects a compound Poisson process. The last term PJ, due to singular nature of v_2, is referred to as the pure jump component. Note that the decomposition (38) is equivalent to decomposing the process itself to three independent processes as

$$x(t) \overset{d}{=} x_{BM}(t) + x_{CP}(t) + x_{PJ}(t), \tag{39}$$

which is known as the Lévy-Itō decomposition.

APPENDIX B
MMSE ESTIMATION OF LÉVY PROCESSES

A. Derivation of MMSE formula (23)

To prove (23), we start by the definition of MMSE and we apply the explicit form of $\widehat{\mathbf{x}}_{\text{MMSE}} = \mathbb{E}_{\mathbf{x}|\mathbf{y}}\{\mathbf{x}\}$ to simplify the equations.

$$\begin{aligned} n\ \text{MMSE}(n) &= \mathbb{E}\left\{\left\|\mathbf{x}-\mathbb{E}_{\mathbf{x}|\mathbf{y}}\{\mathbf{x}\}\right\|^2\right\} \\ &= \mathbb{E}\left\{\left\|\mathbf{x}-\mathbf{y}-\mathbb{E}_{\mathbf{x}|\mathbf{y}}\{\mathbf{x}-\mathbf{y}\}\right\|^2\right\} \\ &= \mathbb{E}\left\{\|\mathbf{x}-\mathbf{y}\|^2\right\}-\mathbb{E}_{\mathbf{y}}\left\{\left\|\mathbb{E}_{\mathbf{x}|\mathbf{y}}\{\mathbf{x}-\mathbf{y}\}\right\|^2\right\} \\ &\overset{(22)}{=} n\sigma^2-\sigma^4\mathbb{E}_{\mathbf{y}}\left\{\|\nabla\log p_{\mathbf{y}}(\mathbf{y})\|^2\right\} \\ &= n\sigma^2-\sigma^4\int\nabla^{\mathrm{T}}p_{\mathbf{y}}(\mathbf{y})\nabla\log p_{\mathbf{y}}(\mathbf{y})\mathrm{d}\mathbf{y} \\ &= n\sigma^2+\sigma^4\Bigg(\int p_{\mathbf{y}}(\mathbf{y})\Delta\log p_{\mathbf{y}}(\mathbf{y})\mathrm{d}\mathbf{y} \\ &\quad -\sum_{i=1}^{n}\int_{\mathbb{R}^{n-1}}\frac{\partial}{\partial y_i}p_{\mathbf{y}}(\mathbf{y})\Big|_{y_i=-\infty}^{\infty}\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}y_i}\Bigg) \\ &= n\sigma^2+\sigma^4\int p_{\mathbf{y}}(\mathbf{y})\Delta\log p_{\mathbf{y}}(\mathbf{y})\mathrm{d}\mathbf{y}. \end{aligned} \tag{40}$$

B. Derivation of MMSE Estimator for Brownian Motion

If $\mathbf{x}$ is an n-dimensional Brownian motion, we can write it as $\mathbf{x} = \mathbf{D}^{-1}\mathbf{u}$, where $\mathbf{u}$ is an n-dimentional random vector with distribution $\mathcal{N}(0,\sigma_u^2\mathbf{I}_n)$. Thus, the distribution of $\mathbf{y}$ would be $\mathcal{N}(0,\sigma^2\mathbf{I}_n+\sigma_u^2\mathbf{D}^{-1}\mathbf{D}^{-\mathrm{T}})$. Now, according to (22) we have,

$$\begin{aligned} \widehat{\mathbf{x}}_{\text{MMSE}} &= \mathbf{y}+\sigma^2\nabla\log p_{\mathbf{y}}(\mathbf{y}) = \mathbf{y}+\sigma^2\frac{\nabla p_{\mathbf{y}}(\mathbf{y})}{p_{\mathbf{y}}(\mathbf{y})} \\ &= \mathbf{y}-\sigma^2\left(\sigma^2\mathbf{I}+\sigma_u^2(\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1}\right)^{-1}\mathbf{y} \\ &= \left(\mathbf{I}_n-\sigma^2\left(\sigma^2\mathbf{I}+\sigma_u^2(\mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1}\right)^{-1}\right)\mathbf{y}. \end{aligned} \tag{41}$$

This is the so called Winner filter.

Notice that for any Levy process for which the variance of the steps is σ_u^2, this estimator gives the minimum MSE among all linear estimators (LMMSE estimator).

C. KLT of Finite-Variance Lévy Processes

While the increments are stationary, the actual Lévy processes are not, which complicates their analysis. The autocorrelation function in the finite-variance case is

$$\begin{aligned} &\mathbb{E}\left[(x(t_1)-\mathbb{E}[x(t_1)])(x(t_2)-\mathbb{E}[x(t_2)])\right] \\ &= \frac{c}{2}\left(|t_1|+|t_2|-|t_2-t_1|\right), \end{aligned} \tag{42}$$

where c is a constant [27]. Thus, if we consider the vector $\mathbf{x}$ containing samples of a normalized Lévy process with increments of unit variance, the covariance matrix of $\mathbf{x}$ would be

$$\mathbb{E}\left[\mathbf{x}\mathbf{x}^{\mathrm{T}}\right] = \mathbf{C} = \left(\mathbf{D}^{\mathrm{T}}\mathbf{D}\right)^{-1}, \tag{43}$$

where, for all $i,j\in[1\dots n]$ the components $c_{ij} = [\mathbf{C}]_{ij}$ are given by

$$c_{ij} = \left[\left(\mathbf{D}^{\mathrm{T}}\mathbf{D}\right)^{-1}\right]_{ij} = \min\{i,j\}. \tag{44}$$

We are interested in finding the eigenvalues of $\mathbf{C}^{-1}$, where

$$\mathbf{C}^{-1} = \mathbf{D}^{\mathrm{T}}\mathbf{D} = \begin{cases} 2, & i=j\neq n \\ 1, & i=j=n \\ -1, & |i-j|=1 \\ 0, & \text{otherwise.} \end{cases} \tag{45}$$

Then, by writing the eigenvalue equation of the matrix $\mathbf{C}$ as

$$\mathbf{C}^{-1}\mathbf{v} = \lambda\mathbf{v}, \tag{46}$$

where $\mathbf{v} = [v_1\cdots v_n]^{\mathrm{T}}$, we obtain the recursive set of equations

$$\begin{cases} 2v_1-v_2=\lambda v_1 \\ -v_{i-1}+2v_i-v_{i+1}=\lambda v_i, \quad \text{for } i=2,\dots,n-1 \\ -v_{n-1}+v_n=\lambda v_n. \end{cases} \tag{47}$$

The solution of these equations is given by

$$v_i = \frac{1}{\sqrt{\lambda(\lambda-4)}}\left(\left(\frac{2-\lambda+\sqrt{\lambda(\lambda-4)}}{2}\right)^i - \left(\frac{2-\lambda-\sqrt{\lambda(\lambda-4)}}{2}\right)^i\right) \tag{48}$$

for $i=1,\dots,n$. Finally, by plugging (48) into (47) and performing some algebraic manipulations, we obtain

$$\lambda = 4\sin^2\left(\frac{\pi}{2}\frac{2k-1}{2n+1}\right) \tag{49}$$

for $k=1,\dots,n$. The entries of the corresponding eigenvectors $\mathbf{v}^k = \left[v_1^k\dots v_n^k\right]^{\mathrm{T}}$ are given by

$$v_i^k = \frac{2}{\sqrt{2n+1}}\sin\left(\frac{2k-1}{2n+1}i\pi\right), \tag{50}$$

for $k=1,\dots,n$.

D. MMSE of Estimation of Brownian Motion or Equivalently Performance of LMMSE

MMSE estimator for Brownian motion or the LMMSE estimator for any finite variance Levy process is equivalent to the entry-wise MMSE in the KLT domain. In KLT domain we have

$$\tilde{\mathbf{y}} = \mathbf{V}^{\mathrm{T}}\mathbf{y} = \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{V}^{\mathrm{T}}\mathbf{u}+\mathbf{V}^{\mathrm{T}}\mathbf{n} = \mathbf{\Lambda}^{-\frac{1}{2}}\tilde{\mathbf{u}}+\tilde{\mathbf{n}}. \tag{51}$$

where $\mathbf{V} = \left[\mathbf{v}^1|\cdots|\mathbf{v}^n\right]$ and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1,\dots,\lambda_n)$. Notice that since $\mathbf{u}$ and $\mathbf{n}$ have distributions $\mathcal{N}(0,\sigma_u^2\mathbf{I}_n)$ and $\mathcal{N}(0,\sigma^2\mathbf{I}_n)$, respectively, and $\mathbf{V}$ is a unitary matrix, $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{n}}$ are also distributed as $\mathcal{N}(0,\sigma_u^2\mathbf{I}_n)$ and $\mathcal{N}(0,\sigma^2\mathbf{I}_n)$.

Now, the MSE of estimating the i^{th}-entry of $\mathbf{\Lambda}^{-\frac{1}{2}}\tilde{\mathbf{u}}$ from the i^{th}-entry of $\tilde{\mathbf{y}}$, simply is $\sigma_u^2/\left(\frac{\sigma_u^2}{\sigma^2}+\lambda_i\right)$. Thus, we have

$$\text{MMSE}(n) = \frac{1}{n}\sum_{i=1}^{n}\frac{\sigma_u^2}{\frac{\sigma_u^2}{\sigma^2}+4\sin^2\left(\frac{\pi}{2}\frac{2i-1}{2n+1}\right)}. \tag{52}$$

If we tend n to infinity, we get

$$\lim_{n\to\infty} \mathrm{MMSE}(n) = \int_0^1 \frac{\sigma_u^2}{\frac{\sigma_u^2}{\sigma^2} + 4\sin^2\left(\frac{\pi}{2}t\right)} \mathrm{d}t = \frac{\sigma^2}{\sqrt{1 + 4\frac{\sigma^2}{\sigma_u^2}}}. \tag{53}$$

REFERENCES

[1] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*. Chapman & Hall, 1994.

[2] D. Applebaum, *Lévy Processes and Stochastic Calculus*. Cambridge University Press, 2009.

[3] 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.

[4] M. Unser, P. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian vs. sparse stochastic processes—Part II: Discrete-domain theory," 2011, arXiv:1108.6152v1.

[5] 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, November 2011.

[6] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1–4, pp. 259–268, November 1992.

[7] M. Unser and P. D. Tafti, "Stochastic models for sparse and piecewise-smooth signals," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 989–1006, March 2011.

[8] J. S. Yedidia, W. T. Freeman, and Y. Weiss, *Understanding Belief Propagation and Its Generalizations*. San Francisco, CA: Morgan Kaufmann Publishers, 2003, ch. Exploring Artificial Intelligence in the New Millennium, pp. 239–269.

[9] H.-A. Loeliger, J. Dauwels, J. Hu, S. Korl, L. Ping, and F. R. Kschischang, "The factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295–1322, June 2007.

[10] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.

[11] 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, April 2010.

[12] M. Nikolova, "Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares," *SIAM Multiscale Model. Simul.*, vol. 4, no. 3, pp. 960–991, September 2005.

[13] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, October 1998.

[14] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, December 2007.

[15] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. of Fourier Anal. Appl.*, vol. 14, no. 5–6, pp. 877–905, October 2008.

[16] A. Beck and M. Teboulle, "Fast gradient-based algorithm for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, November 2009.

[17] U. S. 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, April 2012.

[18] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, 1981.

[19] F. Luisier, "The SURE-LET approach to image denoising," Ph.D. dissertation, Swiss Federal Institute of Technology Lausanne, January 2010, EPFL Thesis no. 4566.

[20] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Trans. Inf. Theory*, vol. 47, no. 2, pp. 599–618, February 2001.

[21] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, November 2009.

[22] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 269–280, January 2010.

[23] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," in *Proc. Conf. on Inform. Sci. & Sys.*, Princeton, NJ, March 2010, pp. 1–6.

[24] ——, "Generalized approximate message passing for estimation with random linear mixing," in *Proc. IEEE Int. Symp. Information Theory*, St. Petersburg, Russia, July-August 2011, pp. 2168–2172.

[25] U. S. Kamilov, V. K. Goyal, and S. Rangan, "Message-passing estimation from quantized samples," November 2011, arXiv:1105.4652 [cs.IT].

[26] M. Unser and T. Blu, "Generalized smoothing splines and optimal discretization of the Wiener filter," *IEEE Trans. Signal Process.*, vol. 53, no. 6, pp. 2146–2159, June 2005.

[27] T. Blu and M. Unser, "Self-similarity: Part II—Optimal estimation of fractal processes," *IEEE Trans. Signal Process.*, vol. 55, no. 4, pp. 1364–1378, April 2007.