

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 (BP) 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—Belief propagation, Lévy process, message passing, nonlinear reconstruction, sparse-signal estimation, stochastic modeling, total-variation estimation.

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 that have sparse derivatives. We

take a continuous-domain perspective and propose Lévy processes [1]–[4] as a natural approach to model such signals. The fundamental defining property of a 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], [6]. Therefore, Lévy processes can be considered as the archetype of sparse stochastic signals [3].

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) [7] minimization, implicitly assume the signals to be sampled instances of Lévy processes [8]. In this paper, we investigate a minimum-mean-squared-error (MMSE) estimator for Lévy processes. The estimator provides a lower-bound on the mean-squared error (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 the MMSE estimator based on the *belief-propagation* (BP) algorithm on cycle-free factor graphs [9]–[11].

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 to 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

Manuscript received March 08, 2012; revised June 21, 2012 and September 07, 2012; accepted September 17, 2012. Date of publication October 03, 2012; date of current version December 12, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Benoit Champagne. This work was supported by the European Commission by Grant ERC-2010-AdG 267439-FUN-SP.

The authors are with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland (e-mail: ulugbek.kamilov@epfl.ch; pedram.pad@epfl.ch; arash.amini@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/TSP.2012.2222394

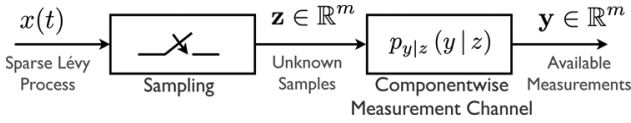


Fig. 1. Signal model considered in this work. The continuous-domain Lévy process $x(t)$ is sampled, and the resulting vector $\mathbf{z} \in \mathbb{R}^m$ is passed through a separable measurement channel $p_{y|z}(y|z)$ to yield $\mathbf{y} \in \mathbb{R}^m$. We investigate the estimation of interpolated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

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 [12]. In particular, the availability of MMSE allows us to benchmark these estimators on signals with desired properties such as sparsity.

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 additive white Gaussian noise (AWGN). In Section VI, we present the BP 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.

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 $\text{Prob}(x \leq a) = \text{Prob}(y \leq a)$ for all $a \in \mathbb{R}$.

II. SIGNAL AND MEASUREMENT MODEL

In this section, we describe the signal model summarized in Fig. 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, the Brownian motion and the Poisson process being the 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 $0 \leq t_1 < t_2 < \dots < t_n < \infty$ 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 $t_1 \geq 0$

$$\lim_{t_2 \rightarrow t_1} \text{Prob}(|x(t_2) - x(t_1)| > \epsilon) = 0.$$

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 $y^{(1)}, \dots, y^{(n)}$ such that

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

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

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

where each

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

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

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

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

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

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 pdfs

$p_u(x) = p_u(-x)$ which results in the simplified Lévy-Khintchine formula

$$\hat{p}_u(\omega) = \exp\left(-\frac{1}{2}b\omega^2 - \int_{\mathbb{R}\setminus\{0\}} (1 - \cos(\omega z))v(z)dz\right). \quad (1)$$

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 Fig. 2. Without loss of generality, we assume an increment $u = x(s) - x(s-1)$ for some fixed $s \geq 1$.

- 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

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

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 Fig. 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 $n(t) \sim \text{Poisson}(\lambda)$ 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

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

where \hat{p}_z is the Fourier transform of p_z . On finite intervals, the sample paths of the process are piecewise-constant (Fig. 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 [13]. The parameter λ controls the sparsity of the signal; it represents the rate of discontinuities. Compound Poisson processes are of special importance in the Lévy-Ito decomposition of Lévy processes. This decomposition expresses any Lévy process as the sum of three processes, one being a Brownian motion and another being Compound Poisson. More details are provided in Appendix I.

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

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

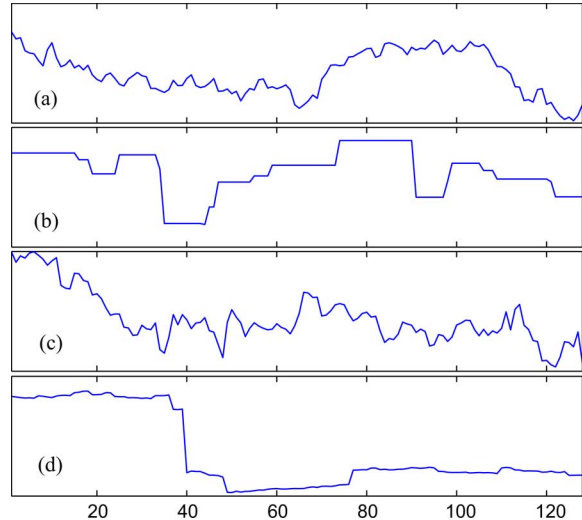


Fig. 2. Sample paths of Lévy processes discussed in this paper.

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

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

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

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

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 Fig. 2(c) an illustration of such a process.

- 4) *Lévy-Flight Process*: Stable random variables are such that a linear combination of two independent ones 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 $(0, 0, v(z) = c_\alpha/|z|^{1+\alpha})$, which results in

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

with $\rho > 0$ and $0 < \alpha < 2$. Such distributions are heavy-tailed and are known to result in highly compressible sequences [6]. A sample signal generated from a Cauchy-increment Lévy flight, which corresponds to the α -stable process with $\alpha = 1$, is illustrated in Fig. 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

$$\frac{d}{dt}x(t) = w(t), \quad (7)$$

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

$$x(t) = \int_0^t w(t')dt', \quad (8)$$

which provides a convenient linear-system interpretation. The only delicate aspect of this interpretation is that the white noise must be considered as a tempered distribution, since it is too rough to admit a classical interpretation as a function of t . The result confirms that, for all positive $k \in \mathbb{N}$, the quantities

$$\begin{aligned} u_k &= x(k) - x(k-1) = D_d x(k) \\ &= \int_{k-1}^k w(t)dt = \langle \text{rect}(\cdot - k + \frac{1}{2}), w(\cdot) \rangle \end{aligned} \quad (9)$$

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, D_d is the finite-difference operator, and rect is the rectangular function, which is 1 inside the interval $[-1/2, 1/2]$ and zero outside. 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 Fig. 1. The vector $\mathbf{z} \in \mathbb{R}^m$ contains the uniformly sampled values of x

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

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

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

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

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

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 $n \geq m$. In other words, we seek more estimates than there are samples. 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 $m_s = \frac{T_s}{T_e}$ we have $z_i = x_{m_s(i-1)+1}$ for all $i \in [1 \dots m]$.

The generality of the 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*: 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

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

- 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 $Q : \mathbb{R} \rightarrow \mathcal{C}$. The nonlinear function Q is often called a K -level scalar quantizer [14]. It maps the K -partitions of the real line $\{Q^{-1}(c_i) : i = 1, \dots, K\} \subseteq \mathbb{R}$ 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

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

where $Q^{-1}(y) = \{z \in \mathbb{R} : Q(z) = y\}$ 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} , whose components are given in (12). Then, from the stationary independent increments property of Lévy processes the components

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

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

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

where we use \propto to denote identity after normalization to unity. The distribution of the whitened elements p_{u_k} is, in principle, obtained by taking the inverse Fourier transform $p_{u_k}(u) = \mathcal{F}^{-1}\{\hat{p}_{u_k}\}(u)$; 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

$$\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmax}} \{p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})\} \quad (17)$$

$$\hat{\mathbf{x}}_{\text{MMSE}} = \mathbb{E}[\mathbf{x}|\mathbf{y}]. \quad (18)$$

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

IV. MAP ESTIMATION

An estimation based on the minimization of some cost functional is a popular way of obtaining the MAP estimator $\hat{\mathbf{x}}_{\text{MAP}}$. The availability of efficient numerical methods for convex and nonconvex optimization partially explain the success of such methods [13], [15]–[17]. The MAP estimator in (17) can be reformulated as

$$\begin{aligned} \hat{\mathbf{x}}_{\text{MAP}} &= \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmax}} \{p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})\} \\ &= \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \{-\log(p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}))\} \\ &= \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \{\mathcal{D}(\mathbf{x}, \mathbf{y}) + \mathcal{R}(\mathbf{x})\}, \end{aligned} \quad (19a)$$

where

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

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

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

$$\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x} \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{z}\|^2 + \sigma^2 \sum_{k=1}^n \phi_u([\mathbf{D}\mathbf{x}]_k), \quad (20)$$

where $\mathbf{z} \in \mathbb{R}^m$ is given in (10) and $\phi_u(x) = -\log(p_u(x))$ 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 [12]. Such regularizers have been shown to be effective in several problems of the recovery of sparse signals [13], [16]. In [18] 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 focus on theoretical results related to MMSE denoising. Consider the AWGN denoising problem

$$\mathbf{y} = \mathbf{z} + \mathbf{n} \quad \text{with} \quad \mathbf{z} = \mathbf{x}, \quad (21)$$

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

$$\hat{\mathbf{x}}_{\text{MMSE}} = \mathbf{y} + \sigma^2 \nabla \log p_{\mathbf{y}}(\mathbf{y}), \quad (22)$$

where ∇ denotes the gradient with respect to \mathbf{y} and $p_{\mathbf{y}} = p_{\mathbf{x}} * p_{\mathbf{n}}$ is the pdf of the noisy vector \mathbf{y} [19], [20]. The functions $p_{\mathbf{x}}$ and $p_{\mathbf{n}}$ are the pdfs of the prior and AWGN, respectively. Then, the MMSE of the estimation problem is given by

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

where Δ is the Laplacian with respect to \mathbf{y} (cf. Appendix II-A).

Although elegant, (22) and (23) are not tractable for arbitrary distributions on \mathbf{x} . However, 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 the linear minimum-mean-squared-error (LMMSE) estimator (cf. Appendix II-B). In the Karhunen-Loève transform (KLT) domain, the Wiener filter reduces to a simple pointwise linear estimator. We finally obtain the asymptotic description of the MMSE (cf. Appendix II-C and Appendix II-D) as

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

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

VI. MESSAGE-PASSING ESTIMATION

A. Exact Formulation

In this section, we specify the MMSE estimator $\hat{\mathbf{x}}_{\text{MMSE}}$ 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. We propose to use the *sum-product belief-propagation* [9] method to efficiently approximate the marginalization of the posterior (16), whose direct computation is intractable otherwise. The BP-based message-passing methods have successfully been used in numerous inference problems in statistical physics, computer vision, channel coding, and signal processing [9]–[11], [21]–[26].

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 Fig. 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

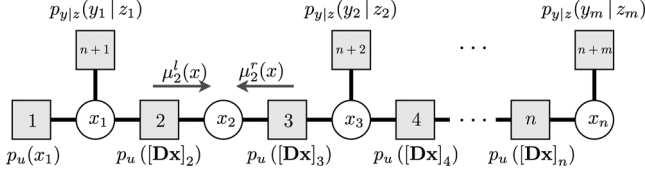


Fig. 3. Factor-graph representation of the posterior distribution (16) with $m_s = 2$. In the graph, square factor nodes represent the probability densities and circled variable nodes represent the unknowns. The functions μ_2^l and μ_2^r represent beliefs at the variable node 2.

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

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

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

$$\mu_1^l(x) = p_u(x), \quad (26a)$$

$$\mu_n^r(x) = 1. \quad (26b)$$

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

$$\mu_{i+1}^l(x) \propto \int_{\mathbb{R}} p_u(x - z) \eta_i(z) \mu_i^l(z) dz, \quad (27a)$$

$$\mu_{n-i}^r(x) \propto \int_{\mathbb{R}} p_u(z - x) \eta_j(z) \mu_j^r(z) dz, \quad (27b)$$

where $j = n - i + 1$. As in (16), 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 $i = 1, \dots, n$, compute

$$[\hat{\mathbf{x}}_{\text{MMSE}}]_i = \int_{\mathbb{R}} x p_{x_i|\mathbf{y}}(x | \mathbf{y}) dx, \quad (28a)$$

where the marginal pdf is obtained by

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

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 $\hat{\mathbf{x}}_{\text{MMSE}}$.

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) describes the distribution of the increments, but 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

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

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

1) *Initialization*: Set

$$\hat{\mu}_1^l(\omega) = \hat{p}_u(\omega), \quad (30a)$$

$$\hat{\mu}_n^r(\omega) = \delta(\omega), \quad (30b)$$

where δ is the Dirac delta function.

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

$$\hat{\mu}_{i+1}^l(\omega) \propto \hat{p}_u(\omega) \cdot (\hat{\eta}_i * \hat{\mu}_i^l)(\omega), \quad (31a)$$

$$\hat{\mu}_{n-i}^r(\omega) \propto \hat{p}_u(\omega) \cdot (\hat{\eta}_j * \hat{\mu}_j^r)(\omega), \quad (31b)$$

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

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

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

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

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

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 $[-N_{\Omega, \epsilon} \Omega, N_{\Omega, \epsilon} \Omega]_{\mathbb{Z}}$. We retain only these samples for which

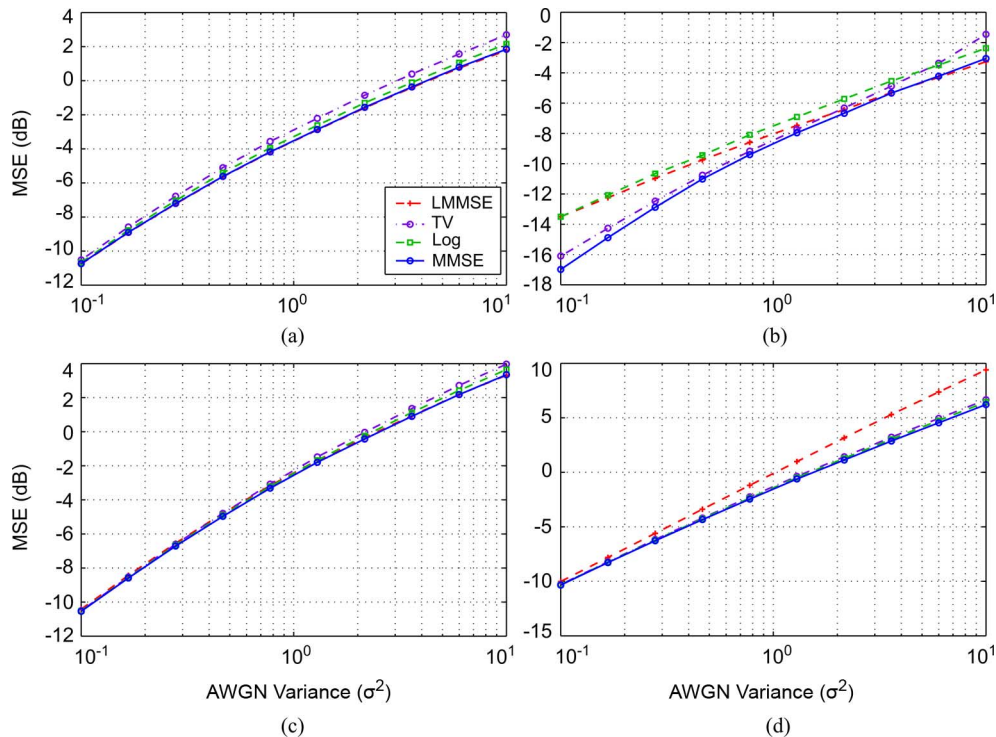


Fig. 4. AWGN denoising of: (a) Brownian motion. (b) Compound Poisson process. (c) Lévy process with Laplace increments. (d) Lévy-flight process.

$f(x) \geq \epsilon$. Thus, the total number of samples for representing the function depends on both the truncation parameter $\epsilon > 0$ and on the sampling step $\Omega > 0$. It is given by $M_{\Omega, \epsilon} = 2N_{\Omega, \epsilon} + 1$. 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 sums to approximate the integrals.

VII. EXPERIMENTAL RESULTS

We now present several experiments with the goal of comparing various signal-estimation methods. The performance of the estimator is judged based on the MSE given by

$$\text{MSE} = 10 \log_{10} \left(\frac{1}{n} \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \right), \quad (33)$$

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

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

- *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 $P(u_k = 0) = e^{-\lambda} = 0.9$. 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 . We perform 1000 random realization of the denoising problem for each value of σ^2 and plot the average MSE in Figs. 4(a)–4(d). The signal length is set to $n = m = 200$. The proposed message-passing estimator is compared with the regularized least-squares estimators

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^n}{\text{argmin}} \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 + \tau \sum_{k=1}^n \phi_u([\mathbf{D}\mathbf{x}]_k), \quad (34)$$

where \mathbf{D} is a 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 $\phi_u(x) = x^2$ [27]. The TV method corresponds to the potential function $\phi_u(x) = |x|$ and can be efficiently implemented by the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) described in [17]. The Log estimator corresponds to the potential function $\phi_u(x) = \log(x^2 + \epsilon)$, 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 [13], [16]. In our experiments, we fix $\epsilon = 1$, which corresponds to the MAP estimator for the Lévy-flight process with Cauchy increments. The Log-based denoising was implemented efficiently by the algorithm described in [18].

It is well known that the LMMSE estimator is optimal for Brownian motion. In Fig. 4(a), it is precisely matched by the message-passing MMSE estimator. Moreover, 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 Fig. 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 Fig. 4(b), we observe excellent MSE performance of TV for compound Poisson processes over many noise levels. This happens because the piecewise-constant solution of TV estimator is well matched to 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 relevance of the compound Poisson process is extensively discussed in [8].

In Fig. 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.

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

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

where $[\hat{\mathbf{x}}_{\text{MMSE}}]_k$ 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 $\hat{\mathbf{x}}$. The average predicted MSE is obtained from 1000 random realizations of each process. Table I also provides the standard deviation of the predicted MSE values across noise levels.

B. Signal Interpolation

In Fig. 5, we illustrate the interpolation of four types of Lévy processes from noisy measurements. We assume AWGN of variance $\sigma^2 = 1$ and set the interpolation rate to $m_s = \frac{T_s}{T_e} = 10$. Given 10 noisy measurements, this results in

TABLE I
MMSE PREDICTION

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

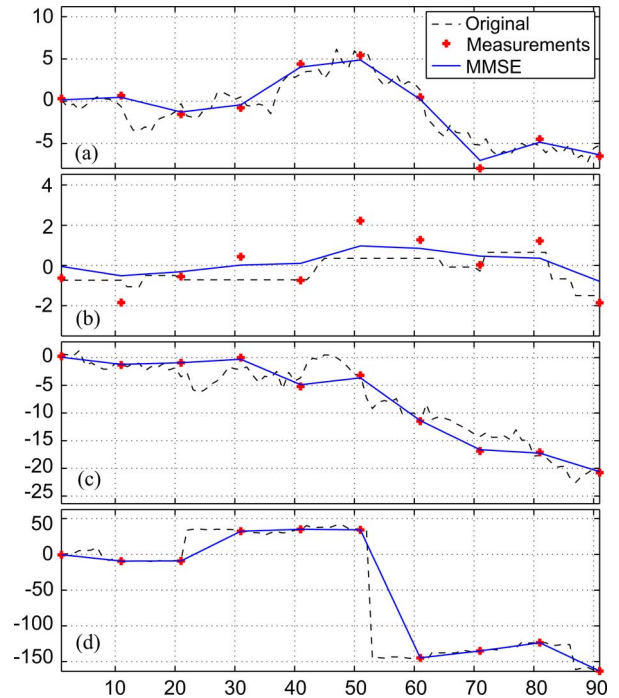


Fig. 5. Tenfold interpolation of Lévy processes from AWGN measurements. From top to bottom: (a) Brownian motion. (b) Compound Poisson process. (c) Lévy process with Laplace increments. (d) Lévy flight process. Surprisingly, for all priors the optimal estimator appears to be a piecewise linear function.

91 estimated values. 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 Brownian motion, piecewise-linear interpolation is optimal [28]. Note that this does not imply that the estimator is itself linear, in the sense of being homogeneous and additive—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. For the interpolation problem, the average estimation MSE for the Lévy-flight process is not defined and can only be characterized conditioned on a given \mathbf{y} . Therefore, this process was omitted from the table

C. Estimation From Quantized Samples

We next consider the highly nonlinear problem of estimating Lévy processes from quantized measurements (14). To do so, we generate a 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.

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

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

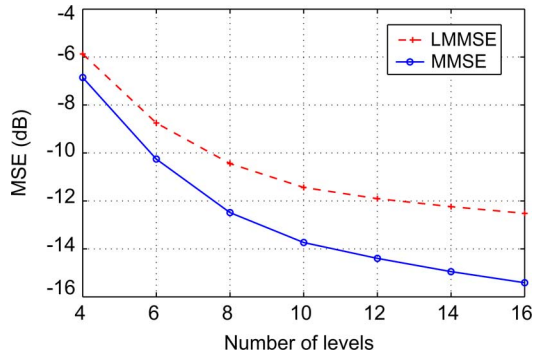


Fig. 6. Estimation of the compound Poisson process from quantized measurements. We compare the standard LMMSE against MMSE, thereby illustrating the suboptimality of standard linear reconstructions.

In Fig. 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 nonlinear 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 family of stochastic processes for the continuous-domain modeling of sparse signals. The key contribution of this paper is 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 I LÉVY-ITO 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

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

show that

$$\left(\hat{p}_{x\left(\frac{m}{n}\right)}(\omega)\right)^n = \left(\hat{p}_{x(1)}(\omega)\right)^m. \quad (36)$$

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

$$\hat{p}_{x(t)}(\omega) = \left(\hat{p}_{x(1)}(\omega)\right)^t. \quad (37)$$

This suggests the Lévy-Khintchine triplet $(ta, tb, tv(\cdot))$ for the random variable $x(t)$. The triplet can be decomposed as

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

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 the singular nature of v_2 , is referred to as the pure-jump component. The decomposition (38) is equivalent to decomposing the process itself in three independent processes as

$$x(t) \stackrel{d}{=} x_{\text{BM}}(t) + x_{\text{CP}}(t) + x_{\text{PJ}}(t), \quad (39)$$

which is known as the Lévy-Ito decomposition.

APPENDIX II MMSE ESTIMATION OF LÉVY PROCESSES

A. Derivation of the MMSE Formula (23)

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

$$\begin{aligned} n \text{ MMSE}(n) &= \mathbb{E} \left\{ \|\mathbf{x} - \mathbb{E}_{\mathbf{x}|\mathbf{y}}\{\mathbf{x}\}\|^2 \right\} \\ &= \mathbb{E} \left\{ \|\mathbf{x} - \mathbf{y} - \mathbb{E}_{\mathbf{x}|\mathbf{y}}\{\mathbf{x} - \mathbf{y}\}\|^2 \right\} \\ &= \mathbb{E} \left\{ \|\mathbf{x} - \mathbf{y}\|^2 \right\} - \mathbb{E}_{\mathbf{y}} \left\{ \|\mathbb{E}_{\mathbf{x}|\mathbf{y}}\{\mathbf{x} - \mathbf{y}\}\|^2 \right\}. \end{aligned}$$

By introducing (22), we obtain

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

B. Derivation of the MMSE Denoising Estimator for Brownian Motion

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

$$\begin{aligned}
\hat{\mathbf{x}}_{\text{MMSE}} &= \mathbf{y} + \sigma^2 \nabla \log p_{\mathbf{y}}(\mathbf{y}) = \mathbf{y} + \sigma^2 \frac{\nabla p_{\mathbf{y}}(\mathbf{y})}{p_{\mathbf{y}}(\mathbf{y})} \\
&= \mathbf{y} - \sigma^2 (\sigma^2 \mathbf{I} + \sigma_u^2 (\mathbf{D}^T \mathbf{D})^{-1})^{-1} \mathbf{y} \\
&= (\mathbf{I}_n - \sigma^2 (\sigma^2 \mathbf{I} + \sigma_u^2 (\mathbf{D}^T \mathbf{D})^{-1})^{-1}) \mathbf{y}. \quad (40)
\end{aligned}$$

This estimator is the so-called Wiener filter, which gives the minimum MSE among all linear estimators (LMMSE estimator) for any Lévy process with increments of variance σ_u^2 .

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

$$\begin{aligned}
\mathbb{E} [(x(t_1) - \mathbb{E}[x(t_1)])(x(t_2) - \mathbb{E}[x(t_2)])] \\
= \frac{c}{2} (|t_1| + |t_2| - |t_2 - t_1|), \quad (41)
\end{aligned}$$

where c is a constant [28]. 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

$$\mathbb{E} [\mathbf{x}\mathbf{x}^T] = \mathbf{C} = (\mathbf{D}^T \mathbf{D})^{-1}, \quad (42)$$

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

$$c_{ij} = [(\mathbf{D}^T \mathbf{D})^{-1}]_{ij} = \min\{i, j\}. \quad (43)$$

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

$$\mathbf{C}^{-1} = \mathbf{D}^T \mathbf{D} = \begin{cases} 2, & i = j \neq n \\ 1, & i = j = n \\ -1, & |i - j| = 1 \\ 0, & \text{otherwise.} \end{cases} \quad (44)$$

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

$$\mathbf{C}^{-1} \mathbf{v} = \lambda \mathbf{v}, \quad (45)$$

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

$$\begin{cases} 2v_1 - v_2 = \lambda v_1 \\ -v_{i-1} + 2v_i - v_{i+1} = \lambda v_i, \text{ for } i = 2, \dots, n-1 \\ -v_{n-1} + v_n = \lambda v_n. \end{cases} \quad (46)$$

The solution of these equations is given by

$$v_i = \frac{1}{\sqrt{\lambda(\lambda-4)}} \left(\left(\frac{2-\lambda+\sqrt{\lambda(\lambda-4)}}{2} \right)^i - \left(\frac{2-\lambda-\sqrt{\lambda(\lambda-4)}}{2} \right)^i \right) \quad (47)$$

for $i = 1, \dots, n$. Finally, by plugging (47) into (46) and performing some algebraic manipulations, we obtain

$$\lambda = 4 \sin^2 \left(\frac{\pi}{2} \frac{2k-1}{2n+1} \right) \quad (48)$$

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

$$v_i^k = \frac{2}{\sqrt{2n+1}} \sin \left(\frac{2k-1}{2n+1} i \pi \right), \quad (49)$$

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

D. MMSE of Estimation of Brownian Motion

The MMSE estimator for Brownian motion, or, equivalently the LMMSE estimator for any finite-variance Lévy process, is also equivalent to the entry-wise MMSE in the KLT domain. The KLT of \mathbf{y} is given by

$$\tilde{\mathbf{y}} = \mathbf{V}^T \mathbf{y} = \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{V}^T \mathbf{u} + \mathbf{V}^T \mathbf{n} = \mathbf{\Lambda}^{-\frac{1}{2}} \tilde{\mathbf{u}} + \tilde{\mathbf{n}}, \quad (50)$$

where $\mathbf{V} = [\mathbf{v}_1 \dots \mathbf{v}_n]$ and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$. 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 since \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 $\tilde{\mathbf{y}}$ is $\sigma_u^2 / \left(\frac{\sigma_u^2}{\sigma^2} + \lambda_i \right)$ (pointwise Wiener filter). Thus, we have

$$\text{MMSE}(n) = \frac{1}{n} \sum_{i=1}^n \frac{\sigma_u^2}{\frac{\sigma_u^2}{\sigma^2} + 4 \sin^2 \left(\frac{\pi}{2} \frac{2i-1}{2n+1} \right)}. \quad (51)$$

When n tends to infinity, we get

$$\begin{aligned}
\lim_{n \rightarrow \infty} \text{MMSE}(n) &= \int_0^1 \frac{\sigma_u^2}{\frac{\sigma_u^2}{\sigma^2} + 4 \sin^2 \left(\frac{\pi}{2} t \right)} dt \\
&= \frac{\sigma^2}{\sqrt{1 + 4 \frac{\sigma_u^2}{\sigma^2}}}. \quad (52)
\end{aligned}$$

REFERENCES

- [1] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [2] D. Applebaum, *Lévy Processes and Stochastic Calculus*. Cambridge, U.K.: Cambridge Univ. 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,” *arXiv:1108.6150v1 [cs.IT]*, 2011 [Online]. Available: <http://arxiv.org/abs/1108.6150>
- [4] M. Unser, P. Tafti, A. Amini, and H. Kirshner, “A unified formulation of Gaussian vs. sparse Stochastic processes—Part II: Discrete-domain theory,” *arXiv:1108.6152v1 [cs.IT]*, 2011 [Online]. Available: <http://arxiv.org/abs/1108.6152>
- [5] V. Cevher, “Learning with compressible priors,” in *Proc. NIPS*, Vancouver, BC, Canada, Dec. 2009.
- [6] 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.
- [7] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Phys. D*, vol. 60, no. 1–4, pp. 259–268, Nov. 1992.
- [8] 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, Mar. 2011.
- [9] J. S. Yedidia, W. T. Freeman, and Y. Weiss, *Understanding Belief Propagation and its Generalizations ch. Exploring Artificial Intelligence in the New Millennium*. San Francisco, CA: Morgan Kaufmann, 2003, pp. 239–269.
- [10] 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, Jun. 2007.
- [11] M. J. Wainwright and M. I. Jordan, “Graphical models, exponential families, and variational inference,” *Found. Trends in Mach. Learn.*, vol. 1, no. 1–2, pp. 1–305, 2008.
- [12] D. Wipf and S. Nagarajan, “Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions,” *IEEE J. Sel. Top. Signal Process.*, vol. 4, no. 2, pp. 317–329, Apr. 2010.
- [13] 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, Sep. 2005.
- [14] R. M. Gray and D. L. Neuhoff, “Quantization,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.
- [15] 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, Dec. 2007.
- [16] E. J. Candès, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *J. Fourier Anal. Appl.*, vol. 14, no. 5–6, pp. 877–905, Oct. 2008.
- [17] 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, Nov. 2009.
- [18] 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, Apr. 2012.
- [19] C. M. Stein, “Estimation of the mean of a multivariate normal distribution,” *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, 1981.
- [20] F. Luisier, “The SURE-LET approach to image denoising,” Ph.D. dissertation, Swiss Fed. Inst. Technol., Lausanne, 2010.
- [21] 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, Feb. 2001.
- [22] D. L. Donoho, A. Maleki, and A. Montanari, “Message-passing algorithms for compressed sensing,” *Proc. Nat. Acad. Sci.*, vol. 106, no. 45, pp. 18914–18919, Nov. 2009.
- [23] 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, Jan. 2010.
- [24] S. Rangan, “Estimation with random linear mixing, belief propagation and compressed sensing,” in *Proc. Conf. Inf. Sci. Syst.*, Princeton, NJ, Mar. 2010, pp. 1–6.
- [25] S. Rangan, “Generalized approximate message passing for estimation with random linear mixing,” in *Proc. IEEE Int. Symp. Inf. Theory*, St. Petersburg, Russia, Jul.–Aug. 2011, pp. 2168–2172.
- [26] U. S. Kamilov, V. K. Goyal, and S. Rangan, “Message-passing de-quantization with applications to compressed sensing,” *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6270–6281, 2011.
- [27] 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, Jun. 2005.
- [28] 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, Apr. 2007.



Ulugbek S. Kamilov (S'11) received the M.Sc. degree in communications systems from the École Polytechnique Fédérale de Lausanne (EPFL), Switzerland, in 2011.

During 2007–2008, he was an exchange student with the Department of Electrical and Computer Engineering, Carnegie-Mellon University, Pittsburgh, PA. In 2008, he worked as a research intern with the Telecommunications Research Center, Vienna, Austria. In 2009, he worked as a software engineering intern at Microsoft. During 2010–2011, he was a visiting student at the Massachusetts Institute of Technology, Cambridge. In 2011, he joined the Biomedical Imaging Group at EPFL where he is currently working toward his Ph.D. His research interests include message-passing algorithms and the application of signal processing techniques to various biomedical problems.



Pedram Pad (S'08) received the B.Sc. and M.Sc. degrees in electrical engineering (communications and signal processing) in 2009 and 2011, respectively, and the B.Sc. degree in mathematical sciences (pure mathematics) in 2009, all from Sharif University of Technology (SUT), Tehran, Iran.

Since December 2011, he has been pursuing the Ph.D. degree with the Biomedical Imaging Group (BIG), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. His research interests include different aspects of information theory and signal processing, specially applications of the former in the latter.



Arash Amini received the B.Sc., M.Sc. and Ph.D. degrees in electrical engineering (communications and signal processing) in 2005, 2007, and 2011, respectively, and the B.Sc. degree in petroleum engineering (reservoir) in 2005, all from Sharif University of Technology (SUT), Tehran, Iran.

Since April 2011, he is a postdoctoral researcher in Biomedical Imaging Group (BIG), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. His research interests include different aspects of sampling, specially compressed sensing.



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

From 1985 to 1997, he worked as a scientist with the National Institutes of Health, Bethesda, MD. He is now Full Professor and Director of the Biomedical Imaging Group, EPFL. His main research area is biomedical image processing. He has a strong interest in sampling theories, multiresolution algorithms, wavelets, and the use of splines for image processing. He has published 200 journal papers on those topics.

Dr. Unser has held the position of Associate Editor-in-Chief (2003–2005) for the IEEE TRANSACTIONS ON MEDICAL IMAGING and has served as an Associate Editor for the same JOURNAL (1999–2002; 2006–2007), the IEEE TRANSACTIONS ON IMAGE PROCESSING (1992–1995), and the IEEE SIGNAL PROCESSING LETTERS (1994–1998). He is currently a member of the editorial boards of *Foundations and Trends in Signal Processing*, and *Sampling Theory in Signal and Image Processing*. He coorganized 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 (BISP). He received the 1995 and 2003 Best Paper Awards, the 2000 Magazine Award, 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.