MMSE DENOISING OF SPARSE LÉVY PROCESSES VIA MESSAGE PASSING

Ulugbek Kamilov, Arash Amini, and Michael Unser

Biomedical Imaging Group, EPFL, Lausanne, Switzerland.

ABSTRACT

Many recent algorithms for sparse signal recovery 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 discrete-gradient regularizers, such as total-variation (TV) minimization, implicitly assume the signals to be sampled instances of Lévy processes with independent Laplace-distributed increments. By extending the concept to more general Lévy processes, we propose an efficient minimum-mean-squared error (MMSE) estimation method based on message-passing algorithms on factor graphs. The resulting algorithm can be used to benchmark the performance of the existing or design new algorithms for the recovery of sparse signals.

***Index Terms*—** signal denoising, sparse estimation, TV denoising

1. INTRODUCTION

Lévy processes constitute the archetype of sparse stochastic signals [1]. Except for Brownian motion, which is the only Gaussian member of the family, these processes exhibit fat-tail statistics shown to fulfill the requirements of compressibility [2]. The simplest example is the compound Poisson process, which is a piecewise-constant signal with a finite rate of innovation. By a limiting argument, one can show that such signals can result in a MAP estimator that is equivalent to the popular TV [3,4]. Due to the availability of a complete statistical description, Lévy processes are particularly interesting for designing or testing algorithms aimed at the recovery of sparse signals.

Our present contribution is the derivation of an MMSE estimator for general Lévy processes. While the joint probability densities of the processes do not always admit closed-form expressions, they are much easier to describe by their characteristic function. For that reason, we perform the estimation in the Fourier domain by representing the characteristic function with its samples. The tree-like structure of the underlying graphical model allows us to perform MMSE estimation of the sampled signal from its noisy observations. Having access to this estimator, we can compare its MSE performance against that of the TV for various stochastic signals. In order to keep things simple, we concentrate our efforts on the 1-D case, while generalizations will be considered in forthcoming publications.

This work was supported by the European Commission under Grant ERC-2010-AdG 267439-FUN-SP.

2. ESTIMATION PROBLEM

2.1. Lévy Denoising Problem

Consider the standard signal-denoising problem

$$\mathbf{y} = \mathbf{x} + \mathbf{e}, \tag{1}$$

where we would like to estimate the unknown discrete signal $\mathbf{x} \in \mathbb{R}^n$ from its noisy observations $\mathbf{y}$. We assume the noise to be white Gaussian of variance σ^2 and our underlying signal model is that $\mathbf{x}$ are integer samples of a Lévy process $X = (X(t) : t \geq 0)$

$$x_i = X(i),\ i = 1, \dots, n. \tag{2}$$

The fundamental defining property of Lévy processes is that they have stationary and independent increments [5], which means that, by taking the finite differences of the discrete signal $\mathbf{x}$, we can decouple it into a sequence of independent identically distributed (i.i.d.) random variables. Consider the vector $\mathbf{w} \in \mathbb{R}^n$ obtained by applying the finite-difference operator L to $\mathbf{x}$. Then, we have

$$w_i = (\mathrm{L}\mathbf{x})_i = x_i - x_{i-1} \sim p_W(\cdot), \tag{3}$$

where $i \in \{1, \dots, n\}$. Note that, by definition of Lévy processes, we must have $x_0 = 0$ [5]. The probability distribution p_W of the Lévy process increments can be uniquely described in its characteristic form by the well-known *Lévy-Khintchine theorem* [5]

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

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

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

Although, the formulation above might look mysterious at first, it allows us to completely characterize a Lévy process with a triplet $(a, b, v(z))$. In principle, the distribution p_W is obtained by taking the inverse Fourier transform $p_W(x) = \mathcal{F}^{-1}\{\hat{p}_W(\omega)\}(x)$. However, it does not necessarily admit a closed form.

What makes the Lévy processes particularly interesting for the sparse estimation problem is the fact that the sparsity of the vector $\mathbf{w}$ is completely determined by the parameters of the Lévy-Khintchine formula. Compound-Poisson and α-stable vectors are two interesting members of the family and have been demonstrated to be sparse [1, 2]. This allows us to test different estimation algorithms on signals of varying degrees of sparsity and to objectively compare their performance. Moreover, having access to the characteristic function $\hat{p}_W$ allows us to derive regularized least-squares functionals for performing the MAP estimation of the signal $\mathbf{x}$. Hence, the estimation can be performed by solving

$$\hat{\mathbf{x}} = \underset{\mathbf{x}\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ \tfrac{1}{2}\|\mathbf{x}-\mathbf{y}\|_{\ell_2}^2 + \sum_{i=1}^{n} \phi(x_i - x_{i-1}) \right\}, \tag{4}$$

where $x_0 = 0$ and the potential function ϕ is given by

$$\phi(x) = -\log\left(\mathcal{F}^{-1}\left\{\hat{p}_W(\cdot)\right\}(x)\right), \tag{5}$$

where $\mathcal{F}^{-1}$ represents the inverse Fourier transform of the characteristic function $\hat{p}_W$. Note that the popular TV-denoising algorithm can be derived as the MAP estimator for the Lévy process with the parameter triplet $(0, 0, v(z) = \frac{1}{|z|}\mathrm{e}^{-|z|})$, which results in increments distributed according to the Laplace distribution $p_W(x) = \frac{\lambda}{2}\mathrm{e}^{-\lambda|x|}$.

3. MESSAGE PASSING

3.1. Exact Formulation

In this section, we specify the MMSE estimator for the signal $\mathbf{x}$ under the Lévy-process model described in the previous section. We begin by constructing the following conditional probability distribution for the variable $\mathbf{x}$ given the measurements $\mathbf{y}$:

$$p_{\mathbf{X}|\mathbf{Y}}(\mathbf{x}|\mathbf{y}) = \frac{1}{Z(\mathbf{y})}\prod_{i=1}^{n}\mathcal{G}\left(y_i - x_i, \sigma^2\right)\prod_{i=1}^{n} p_W\left(x_i - x_{i-1}\right), \tag{6}$$

where $Z(\mathbf{y})$ is a normalization constant, $x_0 = 0$, and $\mathcal{G}$ is the normal probability density function

$$\mathcal{G}\left(x-\mu; \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{7}$$

The posterior distribution (6) of the signal provides a complete statistical characterization of the problem. For instance, the MMSE estimator is given by the conditional expectation

$$\hat{\mathbf{x}}^{\text{MMSE}}(\mathbf{y}) = \mathbb{E}\left[\mathbf{X} \mid \mathbf{Y} = \mathbf{y}\right]. \tag{8}$$

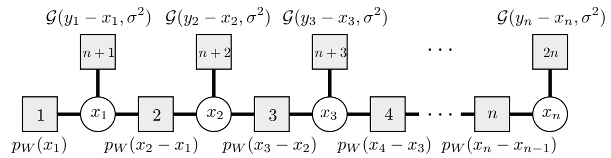

Fig. 1: Factor-graph representation of the posterior distribution.

Algorithm 1 Message-Passing in time-domain

input : $\mathbf{y} \in \mathbb{R}^n$, $\sigma^2 > 0$, and $p_W(\cdot)$
initialize : $\mu_1^l(x) = p_W(x)$, $\mu_n^r(x) = 1$
for $i = 1, \ldots, n-1$ **do**
  $\mu_{i+1}^l(x) \propto \int_{\mathbb{R}} p_W(x-z)\mathcal{G}(y_i - z; \sigma^2)\mu_i^l(z)\mathrm{d}z$
  $\mu_{n-i}^r(x) \propto \int_{\mathbb{R}} p_W(z-x)\mathcal{G}(y_{n-i+1} - z; \sigma^2)\mu_{n-i+1}^r(z)\mathrm{d}z$
end for
set : $p_{X_i|\mathbf{Y}}(x \mid \mathbf{y}) \propto \mu_i^l(x)\mu_i^r(x)\mathcal{G}(y_i - x; \sigma^2)$
return $\hat{x}_i = \int_{\mathbb{R}} x p_{X_i|\mathbf{Y}}(x \mid \mathbf{y})\mathrm{d}x$, for all $i \in \{1, \ldots, n\}$

Remark : The $\propto$ symbol means that the expression on the right-hand side should be normalized to unity.

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 iteratively. One natural approach is to use *sum-product message passing*, which iteratively updates the estimates by passing messages along a graph [6, 7]. In communications and coding communities, the algorithm is commonly known as *belief propagation* and was successfully applied for iterative decoding of LDPC codes [8].

In order to introduce the method, we consider the *factor graph* $G = (V, F, E)$ shown in Figure 1. The graph consists of two sets of nodes, the *variable nodes* $V = \{1, \ldots, n\}$ (circles), the *factor nodes* $F = \{1, \ldots, 2n\}$ (squares), and a set of edges E linking variables to the factors they participate in. The graph is structured according to the factorization of the posterior distribution in (6). Hence, G is a *cycle-free* bipartite graph with n variable nodes and $2n$ factor nodes.

The simple tree-like structure of the factor graph gives rise to the efficient message-passing Algorithm 1. This algorithm computes the marginals of the posterior distribution by sending messages along the edges of the tree. The messages are functions on $\mathbb{R}$ that represent marginals of parts of the posterior, and, combined together, result in the marginals $p_{X_i|\mathbf{Y}}$ for all $i \in \{1, \ldots, n\}$. The algorithm is initialized at the left and right edges of the tree and builds the MMSE estimate by making a single pass along the tree.

The message-passing Algorithm 1 reduces intractable high-dimensional integration into $2n$ convolutions. However, it requires a closed-form solution for the probability distribution p_W. As indicated before, this form is not always available, since the distribution is obtained via the Lévy-Khintchine theorem and is defined by its characteristic

Algorithm 2 Message-Passing in frequency-domain

input : $\mathbf{y} \in \mathbb{R}^n$, $\sigma^2 > 0$, and $\hat{p}_W(\cdot)$
initialize : $\mu_1^l(\omega) = \hat{p}_W(\omega)$, $\mu_n^r(\omega) = \delta(\omega)$
for $i = 1, \ldots, n-1$ **do**
  $\mu_{i+1}^l(\omega) \propto \hat{p}_W(\omega) \int_{\mathbb{R}} \hat{\mathcal{G}}(\omega - \nu; y_i, \sigma^2) \mu_i^l(\nu) d\nu$
  $\mu_{n-i}^r(\omega) \propto \hat{p}_W(-\omega) \int_{\mathbb{R}} \hat{\mathcal{G}}(\omega - \nu; y_{n-i+1}, \sigma^2) \mu_{n-i+1}^r(\nu) d\nu$
end for
set : $\hat{p}(\omega \,|\, \mathbf{y}) \propto \int_{\mathbb{R}} \hat{\mathcal{G}}(\omega - \nu; y_i, \sigma^2) \left(\mu_i^l(\omega) * \mu_i^r(\omega)\right) d\nu$
return $\hat{x}_i = j \frac{d}{d\omega} \hat{p}(\omega \,|\, \mathbf{y})_{|\omega=0}$ for all $i \in \{1, \ldots, n\}$

Remark 1 : The $\propto$ symbol means that the expression on the right-hand side should be normalized by its zero frequency component.
Remark 2 : The $*$ symbol denotes a convolution of two functions.

function. Hence, we propose to perform the estimation in the frequency domain. Denote with $\hat{\mathcal{G}}\left(\omega; y, \sigma^2\right)$ the Fourier transform of the Gaussian pdf in (7). Then, we can proceed with Algorithm 2 in the frequency domain. The algorithm has been obtained by using the convolution property of the Fourier transform. Note that the last estimation step is obtained by applying the moment property of the Fourier transform

$$\int_{\mathbb{R}} x f(x) \mathrm{d}x = \mathrm{j} \frac{\mathrm{d}}{\mathrm{d}w} \hat{f}(\omega)_{|\omega=0},$$

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

3.2. Computational implementation

In principle, our message-passing algorithms result in the exact MMSE estimation of the signal $\mathbf{x}$ from its noisy measurements $\mathbf{y}$. However, the algorithms cannot be implemented in direct form, as the computations involve continuous-time integrals. To obtain a realizable solution, we need to choose some practical discrete parametrization 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 $\delta\,[-N, N]_{\mathbb{Z}}$. The parameters $\delta \in \mathbb{R}_+$ and $N \in \mathbb{N}$ depend on the distribution to represent and on the measurements $\mathbf{y}$. Then, both time- and frequency-domain versions can be obtained by implementing continuous integrals via some quadrature rules.

Our frequency-domain implementation has the advantage of directly using the characteristic function $\hat{p}_W$. The numerical tabulation of the probability density p_W, which would be difficult for some (heavy-tail) distributions, is avoided Moreover, as the estimation converges, the frequency-domain messages become broader, which allows us to adapt the grid accordingly.

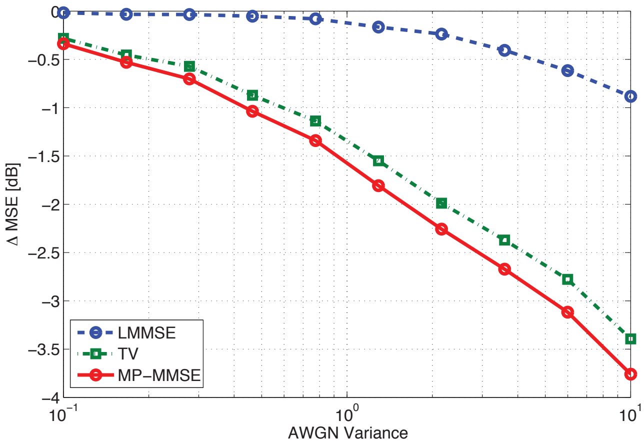

Fig. 2: Comparison of the performance of the LMMSE, TV estimator, and MMSE on the Lévy processes with Cauchy prior as function of the noise power.

4. EXPERIMENTS

We implemented our method in MATLAB and conducted three simple numerical experiments. In the experiments, we compare three estimators: LMMSE, TV, and MMSE obtained via frequency-domain message-passing method described in Algorithm 2. The TV algorithm was implemented using FISTA [9] and was allowed to run for 500 iterations. For each realization of the problem, the regularization parameter $\lambda > 0$ of TV was optimized for the best MSE performance.

The simulations in Figures 2 and 3 were conducted with 50 random realizations of the problem, while the one in Figure 4 with 500. Signals of length $n = 500$ were considered. On the horizontal axis, we plot the input noise level, while on the vertical we show the average MSE reduction after estimation given by

$$\Delta\mathrm{MSE} = 10 \log_{10}\left(\frac{\|\mathbf{x} - \hat{\mathbf{x}}\|_{\ell_2}^2}{\|\mathbf{x} - \mathbf{y}\|_{\ell_2}^2}\right),$$

where $\hat{\mathbf{x}}$ is the estimate of $\mathbf{x}$. We approximated integrals in the message-passing rules by simple Riemann sums. As can be seen, the MP-MMSE estimator outperforms TV and linear MMSE reconstructions over the whole range of experiments and noise variances.

In Figure 2, the signal increments are distributed according to a Cauchy distribution, which has been shown to be highly compressible due to its heavy-tail nature [2]. We observe that TV outperforms LMMSE for all noise levels. This is due to the fact that the TV method is better suited for sparse estimation, preserving the tails of the prior.

In Figure 3, we denoise the sample values of a Lévy process with Laplace prior p_W for which the MAP estimator is given by TV. Although TV performs well at lower noise levels, it is outperformed by LMMSE for high noise powers. At

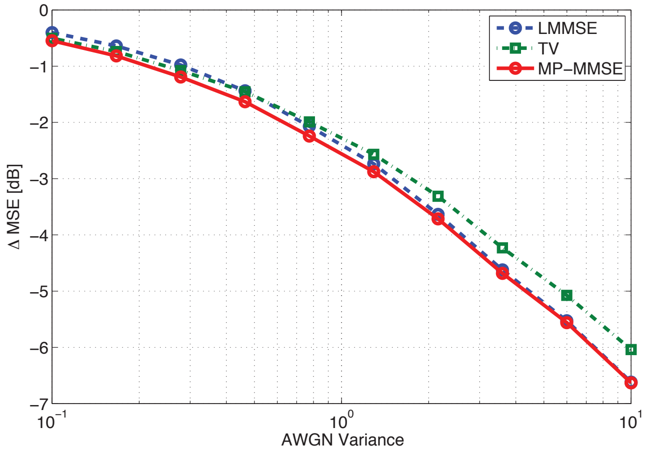

Fig. 3: Comparison of the performance of the LMMSE, TV estimator, and MMSE on the Lévy processes with Laplace prior as function of the noise power.

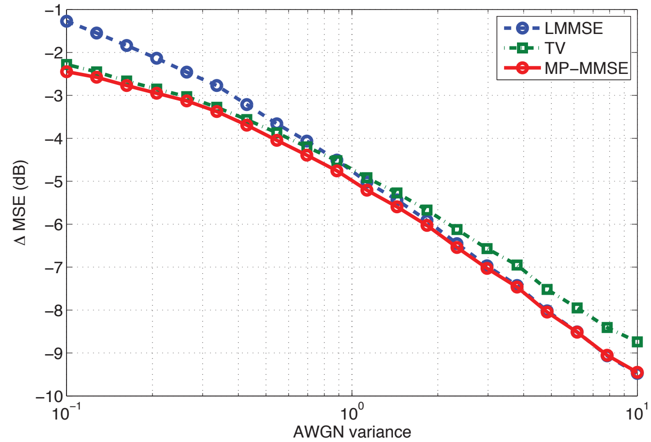

Fig. 4: Comparison of the performance of the LMMSE, TV estimator, and MMSE on the Lévy processes with compound Poisson prior as function of the noise power.

high noise levels, the performance of the LMMSE is identical to the MP-MMSE, as the statistics of the measurements $\mathbf{y}$ are dominated by Gaussian noise.

In the final experiment (Figure 4), we generate the compound Poisson process (piecewise-constant signal) considered in [3]. The distribution of the increments of these signals is sparse, in the sense that it contains a probability mass at 0. In our experiment, we set the mass probability to $\text{Prob}(w_i = 0) = \mathrm{e}^{-0.5} \approx 0.6$ with Gaussian distributed amplitudes. We observe that, for lower noise levels, TV almost achieves the MP-MMSE performance, which may motivate it as an approximate MMSE for compound Poisson signals at low noise levels.

5. CONCLUSION

We have derived and implemented a minimum mean-squared error estimator for denoising samples of sparse Lévy processes. The availability of a complete statistical characterization of such signals makes them attractive for designing new and testing standard sparse estimation algorithms. Our method is based on the message-passing algorithms over bipartite graphs and is implemented entirely in the frequency domain. This is consistent with the fact that statistics of the Lévy processes are described by their characteristic function. Finally, we have demonstrated the superior performance of our method by an empirical comparison with the standard LMMSE and TV estimation methods.

6. REFERENCES

[1] 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*.

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

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

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

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

[6] 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, pp. 1295–1322, 2007.

[7] 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.

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

[9] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.