A Statistical Framework to Investigate the Optimality of Signal-Reconstruction Methods

Pakshal Bohra, *Graduate Student Member, IEEE*, Pol del Aguila Pla, *Member, IEEE*, Jean-François Giovannelli, and Michael Unser, *Fellow, IEEE*

***Abstract*—We present a statistical framework to benchmark the performance of reconstruction algorithms for linear inverse problems, in particular, neural-network-based methods that require large quantities of training data. We generate synthetic signals as realizations of sparse stochastic processes, which makes them ideally matched to variational sparsity-promoting techniques. We derive Gibbs sampling schemes to compute the minimum mean-square error estimators for processes with Laplace, Student's t, and Bernoulli-Laplace innovations. These allow our framework to provide quantitative measures of the degree of optimality (in the mean-square-error sense) for any given reconstruction method. We showcase our framework by benchmarking the performance of some well-known variational methods and convolutional neural network architectures that perform direct nonlinear reconstructions in the context of deconvolution and Fourier sampling. Our experimental results support the understanding that, while these neural networks outperform the variational methods and achieve near-optimal results in many settings, their performance deteriorates severely for signals associated with heavy-tailed distributions.**

***Index Terms*—Inverse problems, minimum mean-square error, convolutional neural networks, sparse stochastic processes.**

I. Introduction

INVERSE problems are often encountered in biomedical imaging [1], particularly in modalities such as computed tomography (CT), magnetic resonance imaging (MRI), or deconvolution microscopy. Their goal is to reconstruct an unknown signal from its measurements. Often, these are hard to solve due to their ill-posedness, which implies that the underlying signal cannot be determined uniquely by the acquired measurements, unless one introduces some form of regularization. Therefore, prior knowledge about the signal of interest is required for the resolution of such problems.

A. Model-Based Methods

Model-based methods rely on the mathematical modeling of the signal of interest to counteract the ill-posedness of the inverse problem. We organize them in two categories.

The first category is composed of linear reconstruction methods (e.g., filtered back-projection), which are fast, well understood, and come with performance and stability guarantees [2], [3]. From a variational standpoint, they can be interpreted as minimizers of a cost functional that consists of a quadratic data term to ensure consistency with the measurements, along with an additive quadratic (Tikhonov) regularization term that imposes some smoothness on the solution. Interestingly, these methods can also be derived from a statistical perspective as optimal linear reconstructors under the Gaussian hypothesis [4].

The second category is composed of methods that exploit sparsity—the property that a signal admits a concise representation in some transform domain (e.g., wavelets) [5], [6], [7], [8]. This powerful concept supports the theory of compressed sensing, which gives conditions under which the reconstruction of an image from a limited set of measurements is feasible [9], [10], [11] and stable [12], [13]. To obtain a sparse reconstruction, one typically uses ℓ_1-norm regularization and solves the corresponding convex optimization problem using iterative algorithms such as the fast iterative shrinkage-thresholding algorithm (FISTA) [14] or the alternating direction method of multipliers (ADMM) [15]. In practice, sparsity-promoting regularizers such as total variation (TV) [16] generally improve the quality of the image. From a statistical point of view, many of these sparsity-based methods can be interpreted as maximum a posteriori (MAP) estimators for some specific choices of stochastic models for the signal of interest [17].

B. Learning-Based Methods

Neural-network-based methods that make use of prior information learned from a large collection of training data are now the focus of much of the current research in image reconstruction [18], [19]. They shine in extreme imaging scenarios where one wishes to achieve more with fewer data, for instance when operating with short integration times, which leads to an abundance of noise, or when collecting fewer measurements to reduce either the acquisition duration and/or the radiation exposure [20]. Here,

Manuscript received 13 September 2022; revised 20 January 2023; accepted 15 May 2023. Date of publication 1 June 2023; date of current version 14 June 2023. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Xiao Fu. This work was supported in part by the Swiss National Science Foundation under Grant 200020_184646 / 1 and in part by the European Research Council (ERC Project FunLearn) under Grant 101020573. *(Corresponding author: Pakshal Bohra.)*

Pakshal Bohra and Michael Unser are with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: pakshal.bohra@epfl.ch; michael.unser@epfl.ch).

Pol del Aguila Pla is with the Biomedical Imaging Group, École polytechnique fédérale de Lausanne, 1015 Lausanne, Switzerland, and also with the CIBM Center for Biomedical Imaging, 1015 Lausanne, Switzerland (e-mail: pol.delaguilapla@epfl.ch).

Jean-François Giovannelli is with the IMS (Univ. Bordeaux, CNRS, B-INP), UMR 5218, F-33400 Talence, France (e-mail: jean-francois.giovannelli@u-bordeaux.fr).

This article has supplementary downloadable material available at https://doi.org/10.1109/TSP.2023.3282062, provided by the authors.

Digital Object Identifier 10.1109/TSP.2023.3282062

1053-587X © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.

we focus on two classes of neural-network-based methods and classify them as the counterparts of the model-based ones.

The first successful applications of deep convolutional neural networks (CNNs) in imaging build upon the classical linear-reconstruction algorithms, training a CNN to correct for reconstruction artifacts in extreme imaging conditions [20], [21], [22], [23], [24]. Unrolling methods [25], [26], [27], [28], [29], [30] also fall into this class of direct, nonlinear reconstructions. Examples of successful applications include MRI, CT, optical imaging, and ultrasound. Their gain over the state-of-the-art is impressive and comparable in magnitude to the one afforded by a decade of refinement of the sparsity-promoting techniques.

The second class includes methods that attempt to reconstruct an image that is consistent with the measurements by replacing the proximal operator that is typically involved in the iterative sparsity-promoting methods by an appropriate denoising CNN, which then plays the role of the regularizer. They come in a variety of flavors, including plug-and-play (PnP) [31], [32], [33], [34], regularization-by-denoising (RED) [35], [36], [37], and projected-gradient-descent [38], [39] methods.

Despite their remarkable performance, CNN-based imaging methods have limitations that currently hinder their further development. Unlike the model-based methods, which are backed by sound mathematics, the development of CNN-based approaches is empirical. Expressivity is obtained through the composition of simple units, but the working of the whole is hard to comprehend and the architectural options are overwhelming (e.g., depth, number of channels, size of the filters). In practice, one usually proceeds by trial and error using the training, validation, and testing errors as quantitative criteria. Further, the training of CNNs is poorly understood and often difficult because of the underlying over-parameterization: getting a stochastic optimization algorithm to perform properly for a specific application typically requires a lot of adjustments and experimentation.

Beside the strain that this empirical approach exerts on developers, the performance greatly depends on the quality, cardinality, and representability of the training dataset, while the outcome is not necessarily transposable to other applications. The bottleneck with biomedical imaging is often a limited access to large, representative datasets. This is mostly because of legal issues in medical imaging and because of the lack of standardized protocols in biomicroscopy. Another issue is the chicken-and-egg nature of the training process because the desired image (the physical object that corresponds to the measurements) is not known precisely—in practice, the goldstandard is an image produced by a state-of-the-art model-based method with high-density/low-noise measurements. This is adequate for developing methods for compressed sensing, but not otherwise. This explains why the works that demonstrate the superiority of the CNN-based approaches over the more traditional model-based methods for image reconstruction have used limited benchmarks so far.

C. Contribution

In this work, we present an objective environment to benchmark the performance of reconstruction algorithms for linear inverse problems. Our proposed framework offers quantitative measures of the degree of optimality (in the mean-square-error sense) for any given reconstruction method. Further, it provides access to large amounts of training data, which enables the benchmarking of CNN-based approaches.

We synthesize ground-truth signals and then simulate the measurement process (e.g., convolution for deconvolution microscopy, Fourier sampling for MRI) in the presence of noise. Specifically, we consider a statistical framework where the underlying signals are realizations of 1D sparse stochastic processes (SSPs) [40]. The motivation there is that these processes are ideally matched to model-based methods, the most prominent of which can be interpreted as their MAP estimators [41]. Since the true statistical distribution of the signal is known exactly in our framework, the minimum-mean-square-error (MMSE) estimator is indeed optimal in the mean-square-error (MSE) sense. Therefore, we are able to provide statistical guarantees of optimality by specifying an upper limit on the reconstruction performance.

Our framework also provides training data for CNN-based approaches. Indeed, we can produce any desired number of training pairs for a given reconstruction task and some chosen stochastic signal model, which allows for an informed comparison of network architectures. Thus, the availability of the goldstandard (MMSE estimator) and training data make our benchmark a good ground for the tuning of CNN architectures and for the identification of the best designs in a tightly controlled environment.

The MAP estimates of SSPs are solutions of optimization problems that resemble the ones used in model-based methods, and can be computed efficiently. However, it has been observed that these MAP estimators are suboptimal in the MSE sense [41], [42], except in the Gaussian scenario where the MAP and MMSE estimators (generalized Wiener filter) coincide [4]. In this work, we focus on non-Gaussian signal models. In principle, the MMSE estimator involves the calculation of high-dimensional integrals, which are not numerically tractable in general. Thus, we develop efficient Gibbs-sampling-based algorithms to compute the MMSE estimator for specific classes of SSPs, with innovations following the Laplace, Student's t, and Bernoulli-Laplace distributions. To the best of our knowledge, no such working solution for generic linear inverse problems with SSPs has been presented in the literature.

Finally, we present experimental results that illustrate the usefulness of our framework. Specifically, we benchmark the performance of some well-known model-based methods and CNNs that perform direct nonlinear reconstructions, in the context of deconvolution and Fourier sampling for first-order SSPs. The CNNs that we consider are optimized by minimizing the MSE loss for training datasets. On one hand, when the innovations follow a Bernoulli-Laplace distribution, we observe that CNNs (with sufficient capacity and training data) outperform the sparsity-promoting methods, which are well-suited to these piecewise-constant signals. In fact, some of these CNNs achieve near-optimal MSE performance. On the other hand, our experiments with Student's t innovations indicate regimes where CNNs fail to reconstruct the signals well. More specifically, we observe that, when the tails of the Student's t distribution are made

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.

heavier (i.e., when we move towards a Cauchy distribution), CNNs perform rather poorly.

D. Roadmap

In Section II, we describe a continuous-domain model for the measurement process along with a way to discretize it. In Section III, we introduce Lévy processes as stochastic models for our signals and we derive the probability distribution for samples of such processes. We then discuss MAP and MMSE estimation in Section IV before we develop Gibbs samplers for Lévy processes associated with Laplace, Student's t, and Bernoulli-Laplace distributions in Section V. Finally, we present experimental results in Section VI.

II. Measurement Model

In the proposed framework, we consider the recovery of a continuous-domain signal $s : \mathbb{R} \to \mathbb{R}$ from a finite number M of measurements $\mathbf{y} = (y_m)_{m=1}^M$.

A. Continuous-Domain Measurement Model

We model the measurements $\mathbf{y} = (y_m)_{m=1}^M$ as

$$y_m = \int_{\mathbb{R}} s(t)\nu_m(t)\,\mathrm{d}t + n[m], \tag{1}$$

where $(\nu_m)_{m=1}^M$ are linear functionals that describe the physics of the acquisition process and $n[\cdot]$ is an additive white Gaussian noise (AWGN) with variance σ_{n}^2. By choosing appropriate functionals $(\nu_m)_{m=1}^M$, we can study a variety of linear inverse problems such as denoising, deconvolution, inpainting, and Fourier sampling.

B. Discrete Measurement Model

We need to discretize (1) to obtain a computationally feasible model for the measurements. To that end, we consider a finite region of interest $\Omega = (0, T)$ of the signal and approximate it with

$$s_h(t) = \sum_{k=1}^{K} s(kh)\mathrm{sinc}\left(\frac{t}{h} - k\right), \tag{2}$$

where h is the sampling step and $K = (\lfloor \frac{T}{h} \rfloor - 1)$. When h is small enough, s_h is a good approximation of s within the interval Ω [43]. On introducing (2) into (1), we get that

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n}, \tag{3}$$

where $\mathbf{s} = (s(kh))_{k=1}^K \in \mathbb{R}^K$ contains equidistant samples of the signal, $\mathbf{H} : \mathbb{R}^K \to \mathbb{R}^M$ is the discrete system matrix with

$$[\mathbf{H}]_{m,k} = \int_{\mathbb{R}} \mathrm{sinc}\left(\frac{t}{h} - k\right)\nu_m(t)\,\mathrm{d}t, \tag{4}$$

and $\mathbf{n} \in \mathbb{R}^M$ is the noise.

Thus, for any signal samples $\mathbf{s} \in \mathbb{R}^K$, we can simulate noisy measurements using (3). Next, we derive the discrete system matrices for deconvolution and Fourier sampling. Hereafter, we assume for simplicity that $h = 1$.

C. Deconvolution

In deconvolution, the measurements are acquired by sampling the result of the convolution between the signal and the point-spread function (PSF) ψ of the acquisition system, which we model by letting the measurement functionals be $\nu_m = \psi(m - \cdot)$. We assume that the cutoff frequency of ψ is $\omega_0 \leq \pi$, as this allows us to sample $(s * \psi)$ on an integer grid without aliasing effects. In this case, The entries of the resulting system matrix $\mathbf{H}$ are given by

$$\begin{aligned}[\mathbf{H}]_{m,k} &= \int_{\mathbb{R}} \mathrm{sinc}(t-k)\psi(m-t)\,\mathrm{d}t \\ &= \psi(m-k).\end{aligned} \tag{5}$$

Here, $\mathbf{H}$ is a discrete convolution matrix whose entries are samples of the bandlimited PSF ψ.

D. Fourier Sampling

In Fourier sampling, the measurements are acquired by sampling the Fourier transform of the signal at arbitrary frequencies $\{\omega_m\}_{m=1}^M$. Accordingly, the measurement functionals are the complex exponentials $\nu_m = \mathrm{e}^{-\mathrm{j}\omega_m \cdot}$. Assuming that $|\omega_m| \leq \pi$, we get that

$$\begin{aligned}[\mathbf{H}]_{m,k} &= \int_{\mathbb{R}} \mathrm{sinc}(t-k)\mathrm{e}^{-\mathrm{j}\omega_m t}\,\mathrm{d}t \\ &= \mathrm{e}^{-\mathrm{j}\omega_m k}.\end{aligned} \tag{6}$$

Here, $\mathbf{H}$ is a discrete Fourier-like matrix, except that the frequencies ω_m do not necessarily lie on an uniform grid.

III. Stochastic Signal Model

In this section, we describe a continuous-domain stochastic model for the signal. We also derive the probability distribution for the discrete signal vector $\mathbf{s} = (s(k))_{k=1}^K$.

A. Lévy Processes

In our framework, the underlying signals are realizations of a well-known class of first-order sparse stochastic processes: the Lévy processes [40], [44].

Definition 1 (Lévy process): A stochastic process $s = \{s(t) : t \in \mathbb{R}^+\}$ is a Lévy process if

1) $s(0) = 0$ almost surely;
2) (independent increments) for any $N \in \mathbb{N} \setminus \{0, 1\}$ and $0 \leq t_1 < t_2 \cdots < t_N < \infty$, the increments $(s(t_2) - s(t_1)), (s(t_3) - s(t_2)), \ldots, (s(t_N) - s(t_{N-1}))$ are mutually independent;
3) (stationary increments) for any given step h, the increment process $u_h = \{s(t) - s(t-h) : t \in \mathbb{R}^+\}$ is stationary;
4) (stochastic continuity) for any $\epsilon > 0$ and $t \geq 0$

$$\lim_{h \to 0} \Pr\{|s(t+h) - s(t)| > \epsilon\} = 0.$$

Lévy processes are closely linked to infinitely divisible (id) distributions.

Definition 2 (Infinite divisibility): A random variable X is infinitely divisible if, for any $N \in \mathbb{N} \setminus \{0\}$, there exist independent

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.

and identically distributed (i.i.d.) random variables $X_1, \ldots, X_N$ such that $X = X_1 + \cdots + X_N$.

For any Lévy process s, the random variable $s(t)$ for some $t > 0$ is infinitely divisible. Moreover, its probability density function (pdf) is given by

$$p_{s(t)}(x) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} p_{s(1)}(y) \mathrm{e}^{\mathrm{j}\omega y} \, \mathrm{d}y \right)^t \mathrm{e}^{-\mathrm{j}\omega x} \frac{\mathrm{d}\omega}{2\pi}. \tag{7}$$

Conversely, for any id distribution with pdf p_{id}, it is possible to construct a Lévy process s such that $p_{s(1)} = p_{\text{id}}$. Thus, there is a one-to-one correspondence between Lévy processes and id distributions [44].

Among all id distributions, the pdf of the Gaussian distribution exhibits the fastest rate of decay at infinity. In this sense, we refer to the non-Gaussian, heavier-tailed members (e.g., Laplace, Bernoulli-Laplace, Student's t, symmetric-alpha-stable) of the class of id distributions as sparse [45]. Indeed, some of these sparse distributions have a mass at the origin in their probability distribution (e.g., Bernoulli-Laplace) and some of them are strongly compressible (e.g., Student's t, symmetric-alpha-stable) [46].

The stochastic model of Lévy processes allows us to consider a variety of signals with different types of sparsity. In our framework, we focus on the subclass of Lévy processes associated with the Gaussian, Laplace, Bernoulli-Laplace and Student's t distributions. Some realizations of these processes are shown in Fig. 1.

B. Discrete Stochastic Model

Now, we derive the pdf of the random vector $\mathbf{s} = (s(k))_{k=1}^K$, which contains uniform samples of a Lévy process. Consider the stationary increment process $u(t) = \{s(t) - s(t-1) : t \in \mathbb{R}^+\}$ whose first-order pdf p_u is the same as $p_{s(1)}$ and so is infinitely divisible. Its samples $\mathbf{u} = (u(k))_{k=1}^K$ can be expressed as

$$\mathbf{u} = \mathbf{D}\mathbf{s}, \tag{8}$$

where $\mathbf{D}$ is a finite-difference matrix of the form

$$\mathbf{D} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ & & \ddots & \ddots & \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix}. \tag{9}$$

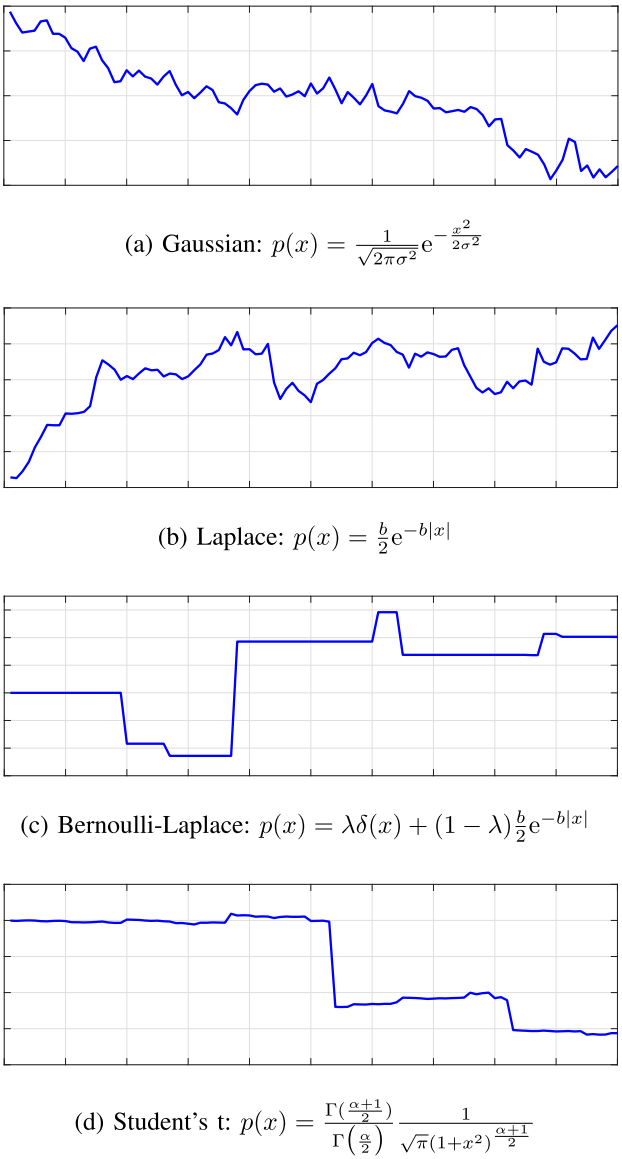
Using (8) and the fact that the increments are independent, we obtain the pdf of the discrete signal as

$$p_{\mathbf{s}}(\mathbf{s}) = \prod_{k=1}^{K} p_u\left([\mathbf{D}\mathbf{s}]_k\right). \tag{10}$$

Note that (8) can also be written as

$$[\mathbf{s}]_k = \sum_{n=1}^{k} [\mathbf{u}]_n, \quad k = 1, \ldots, K, \tag{11}$$

which gives us a direct way to generate samples of Lévy processes.



(a) Gaussian: $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{x^2}{2\sigma^2}}$

(b) Laplace: $p(x) = \frac{b}{2} \mathrm{e}^{-b|x|}$

(c) Bernoulli-Laplace: $p(x) = \lambda\delta(x) + (1-\lambda)\frac{b}{2}\mathrm{e}^{-b|x|}$

(d) Student's t: $p(x) = \frac{\Gamma(\frac{\alpha+1}{2})}{\Gamma\left(\frac{\alpha}{2}\right)} \frac{1}{\sqrt{\pi}(1+x^2)^{\frac{\alpha+1}{2}}}$

Fig. 1. Realizations of different Lévy processes as characterized by the corresponding infinitely divisible pdfs.

C. Extensions

In this work, we have considered inverse problems involving 1D signals that are modelled as realizations of Lévy processes with increments that follow the Gaussian, Laplace, Bernoulli-Laplace and Student's t distributions. Our framework can further be extended in a straightforward manner to include the more general signal model of continuous-domain first-order autoregressive processes [40, Chapter 7] driven by white noises associated with the aforementioned distributions. These AR(1) processes yield a discrete stochastic model that is similar to the

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.

one described in (10). There, the application of a suitable transformation matrix to the discrete signal vector, which contains equidistant samples of the process, decouples it and generates a random vector (called the innovation or generalized increments) with i.i.d. entries. Thus, the MMSE estimation methods presented in Section V can be readily adapted for such AR(1) processes.

We can also directly extend the proposed framework to handle multidimensional signals for the particular stochastic model of continuous-domain AR Lévy sheets [47, Chapter 3], [40] associated with the Gaussian, Laplace, Bernoulli-Laplace and Student's t distributions. These are higher-dimensional generalizations (based on separable whitening operators) of the corresponding AR(1) processes and they result in desirable discrete models of the form (10). Unfortunately, the vectorized discrete signal for other ("non-separable") higher-dimensional stochastic processes described in [40] cannot be fully decoupled by applying a linear transformation. This makes the task of designing schemes to compute their MMSE estimators very challenging. An alternate way of extending our framework could be to define a new class of continuous-domain multidimensional stochastic models using the spline-operator-based framework of [48]. However, this approach would require substantial development of novel mathematical ideas and is thus not discussed further in this article.

IV. Bayesian Inference

So far, we have introduced the signal and measurement models that allow us to generate our ground-truth signals and simulate their noisy measurements for a certain acquisition setup. Next, we focus on statistical estimators for the reconstruction problem at hand, which is to recover the signal $\mathbf{s}$ from the measurements $\mathbf{y}$.

In Bayesian inference, the goal is to characterize the posterior distribution $p_{\mathbf{s}|\mathbf{y}}$ and derive estimators based on it. Using Bayes' rule and (10), we get

$$p_{\mathbf{s}|\mathbf{y}}(\mathbf{s}|\mathbf{y}) = \frac{p_{\mathbf{y}|\mathbf{s}}(\mathbf{y}|\mathbf{s})p_{\mathbf{s}}(\mathbf{s})}{\int_{\mathbb{R}^K} p_{\mathbf{y}|\mathbf{s}}(\mathbf{y}|\mathbf{s})p_{\mathbf{s}}(\mathbf{s})\,\mathrm{d}\mathbf{s}} \propto \exp\left(-\frac{\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2}{2\sigma_{\mathrm{n}}^2}\right)\prod_{k=1}^{K} p_u\left([\mathbf{D}\mathbf{s}]_k\right). \quad (12)$$

A. Maximum a Posteriori Estimator

The MAP estimator calculates the mode of the posterior distribution $p_{\mathbf{s}|\mathbf{y}}$ and is given by

$$\widehat{\mathbf{s}}_{\mathrm{MAP}}(\mathbf{y}) = \arg\max_{\mathbf{s}\in\mathbb{R}^K} p_{\mathbf{s}|\mathbf{y}}(\mathbf{s}|\mathbf{y}) = \arg\min_{\mathbf{s}\in\mathbb{R}^K}\left(\frac{1}{2\sigma_{\mathrm{n}}^2}\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \sum_{k=1}^{K}\Phi_u\left([\mathbf{D}\mathbf{s}]_k\right)\right), \quad (13)$$

where $\Phi_u(x) = -\log(p_u(x))$. The cost functional in (13) consists of a quadratic data-fidelity term and a penalty term that encodes the prior signal model. The optimization task in (13) resembles the one formulated in the variational model-based methods. For instance, if p_u is a Gaussian pdf, then the penalty term is proportional to $\|\mathbf{D}\mathbf{s}\|_2^2$, which is a classical Tikhonov regularizer [2]. However, if p_u is a Laplace pdf, then we have a sparsity-promoting ℓ_1-norm penalty term $\|\mathbf{D}\mathbf{s}\|_1$, which corresponds to the popular TV regularizer [16].

These MAP estimators can be computed efficiently with the help of iterative algorithms such as gradient descent, FISTA [14], and ADMM [15].

B. Minimum Mean-Square Error Estimator

The MMSE estimator is given by

$$\widehat{\mathbf{s}}_{\mathrm{MMSE}}(\mathbf{y}) = \arg\min_{\widehat{\mathbf{s}}\in\mathbb{R}^K}\left(\int_{\mathbb{R}^K}\|\mathbf{s}-\widehat{\mathbf{s}}\|_2^2\, p_{\mathbf{s}|\mathbf{y}}(\mathbf{s}|\mathbf{y})\,\mathrm{d}\mathbf{s}\right) = \int_{\mathbb{R}^K}\mathbf{s}\, p_{\mathbf{s}|\mathbf{y}}(\mathbf{s}|\mathbf{y})\,\mathrm{d}\mathbf{s}, \quad (14)$$

which is the mean of the posterior distribution $p_{\mathbf{s}|\mathbf{y}}$. For a fixed stochastic model, the MMSE estimator is the optimal reconstructor in the MSE sense and thus serves as the goldstandard in our benchmarking framework.

In the Gaussian case, the MMSE estimator is known to coincide with the MAP estimator and is straightforward to calculate [1], [4]. However, in the non-Gaussian case, we need to numerically evaluate the high-dimensional integral in (14), which is computationally challenging.

V. MMSE Estimators for Sparse Lévy Processes

In this section, we present efficient methods to compute the MMSE estimator for sparse Lévy processes with increments that follow the Laplace, Student's t, and Bernoulli-Laplace distributions, which constitutes a key contribution of this article.

A. Markov Chain Monte Carlo Methods

The MMSE estimator $\widehat{\mathbf{s}}_{\mathrm{MMSE}}$ involves the calculation of the integral (14). The high dimensionality of this integral makes its approximation by simple techniques such as uniform-grid-based Riemann sums infeasible. Instead, one can use Markov Chain Monte Carlo (MCMC) methods [49], [50], [51], [52] for the numerical approximation of (14) in a tractable manner.

MCMC methods are designed for generating random samples from nontrivial high-dimensional probability distributions. Broadly speaking, the idea in MCMC is to design a Markov chain such that the distribution that one wishes to draw samples from is its stationary distribution. The desired samples can be obtained by simulating the Markov chain and recording its states after convergence.

In order to compute the integral in (14), we first generate samples $\{\mathbf{s}^{(q)}\}_{q=1}^{Q}$ from $p_{\mathbf{s}|\mathbf{y}}$ using an MCMC method. We then approximate $\widehat{\mathbf{s}}_{\mathrm{MMSE}}$ by the empirical mean $\widehat{\mathbf{s}}_Q = \frac{1}{Q}\sum_{q=1}^{Q}\mathbf{s}^{(q)}$. Although the collected samples are correlated, the Markov chain central limit theorem [53] guarantees that $\widehat{\mathbf{s}}_Q$ is a good approximation of $\widehat{\mathbf{s}}_{\mathrm{MMSE}}$ for a large-enough Q.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Gibbs Sampling.

Input: Q (number of samples), B (burn-in period)
Initialization: $(\widetilde{x}^{(0)}, \widetilde{y}^{(0)})$
for $q = 1, \ldots, B+Q$ **do**
  Generate $\widetilde{x}^{(q)} \sim p_{x|y}(x|\widetilde{y}^{(q-1)})$
  Generate $\widetilde{y}^{(q)} \sim p_{y|x}(y|\widetilde{x}^{(q)})$
end for
Output: $\{(x^{(q)}, y^{(q)})\}_{q=1}^{Q} = \{(\widetilde{x}^{(q+B)}, \widetilde{y}^{(q+B)})\}_{q=1}^{Q}$

B. Gibbs Sampling

In this work, we propose to use the MCMC method called Gibbs sampling [54], [55] to generate samples $\{\mathbf{u}^{(q)}\}_{q=1}^{Q}$ from the posterior distribution $p_{\mathbf{u}|\mathbf{y}}$. These can then be transformed in accordance with (11) to obtain samples $\{\mathbf{D}^{-1}\mathbf{u}^{(q)}\}_{q=1}^{Q}$ from $p_{\mathbf{s}|\mathbf{y}}$. We now give the gist of this algorithm.

Let x and y be two random variables. Consider the task of generating samples from their joint distribution $p_{x,y}$. Gibbs sampling is advantageous whenever it is computationally difficult to sample from the joint distribution directly but the conditional distributions $p_{x|y}$ and $p_{y|x}$ are easy to sample from. The steps involved in this method are presented in Algorithm 1. They yield a Markov chain whose stationary distribution is indeed $p_{x,y}$ [55]. In practice, one discards some of the initial samples (burn-in period) to allow the chain to converge. Moreover, quantities (expectation integrals) based on the marginal distributions p_x and p_y can be computed from the individual samples $\{x^{(q)}\}_{q=1}^{Q}$ and $\{y^{(q)}\}_{q=1}^{Q}$, respectively.

Next, we present Gibbs sampling schemes for Lévy processes with Laplace, Student's t, and Bernoulli-Laplace increments. Our strategy is to introduce an auxiliary vector $\mathbf{w}$ and perform Gibbs sampling for the joint distribution $p_{\mathbf{u},\mathbf{w}|\mathbf{y}}$ [56], [57]. The key is to choose $\mathbf{w}$ such that the conditional distributions $p_{\mathbf{u}|\mathbf{w},\mathbf{y}}$ and $p_{\mathbf{w}|\mathbf{u},\mathbf{y}}$ can be sampled from in an efficient manner.

Hereafter, we assume that the noise variance σ_{n}^2 and the parameters of the signal model are known.

C. Laplace Increments

For Lévy processes with Laplace increments, we adapt the approach that was developed in [58].

The pdf for the Laplace distribution is

$$p_u(u) = \frac{b}{2}\exp\left(-b|u|\right), \tag{15}$$

where b is the scale parameter. The density in (15) can be expressed as a scale mixture of normal distributions [59], as

$$p_u(u) = \int_{\mathbb{R}} p_{u|w}(u|w)p_w(w)\,\mathrm{d}w, \tag{16}$$

where

$$p_{u|w}(u|w) = \frac{1}{\sqrt{2\pi w}}\exp\left(-\frac{u^2}{2w}\right) \tag{17}$$

is the Gaussian pdf and

$$p_w(w) = \frac{b^2}{2}\exp\left(-\frac{b^2 w}{2}\right)\mathbb{1}_+(w) \tag{18}$$

is a mixing exponential pdf[1] with $\lambda = 2/b^2$. This property allows us to define an auxiliary random vector $\mathbf{w} \in \mathbb{R}^K$ with i.i.d. entries following the distribution p_w in (18), such that

$$p_{\mathbf{u}|\mathbf{w}}(\mathbf{u}|\mathbf{w}) = \prod_{k=1}^{K} p_{u|w}\left([\mathbf{u}]_k|[\mathbf{w}]_k\right), \tag{19}$$

where $p_{u|w}$ is shown in (17).

Due to the chain rule of probability (or the general product rule), the full joint distribution $p_{\mathbf{y},\mathbf{u},\mathbf{w}}$ can be written as

$$\begin{aligned} p_{\mathbf{y},\mathbf{u},\mathbf{w}}(\mathbf{y},\mathbf{u},\mathbf{w}) &= p_{\mathbf{y}|\mathbf{u},\mathbf{w}}(\mathbf{y}|\mathbf{u},\mathbf{w})p_{\mathbf{u},\mathbf{w}}(\mathbf{u},\mathbf{w}) \\ &= p_{\mathbf{y}|\mathbf{u}}(\mathbf{y}|\mathbf{u})p_{\mathbf{u}|\mathbf{w}}(\mathbf{u}|\mathbf{w})p_{\mathbf{w}}(\mathbf{w}). \end{aligned} \tag{20}$$

Consequently, the distribution $p_{\mathbf{u},\mathbf{w}|\mathbf{y}}$ takes the form

$$\begin{aligned} p_{\mathbf{u},\mathbf{w}|\mathbf{y}}(\mathbf{u},\mathbf{w}|\mathbf{y}) \propto\ & \exp\left(-\frac{1}{2\sigma_{\mathrm{n}}^2}\|\mathbf{y}-\mathbf{A}\mathbf{u}\|_2^2\right) \\ & \times \prod_{k=1}^{K}[\mathbf{w}]_k^{-\frac{1}{2}}\exp\left(-\frac{[\mathbf{u}]_k^2}{2[\mathbf{w}]_k}\right) \\ & \times \prod_{k=1}^{K}\frac{b^2}{2}\exp\left(-\frac{b^2[\mathbf{w}]_k}{2}\right)\mathbb{1}_+\left([\mathbf{w}]_k\right), \end{aligned} \tag{21}$$

where $\mathbf{A} := \mathbf{H}\mathbf{D}^{-1}$.

Based on (21), the conditional distribution $p_{\mathbf{u}|\mathbf{w},\mathbf{y}}$ is then obtained as

$$\begin{aligned} p_{\mathbf{u}|\mathbf{w},\mathbf{y}}(\mathbf{u}|\mathbf{w},\mathbf{y}) \propto\ & \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_{\mathrm{n}}^2}\|\mathbf{y}-\mathbf{A}\mathbf{u}\|_2^2\right.\right. \\ & \left.\left.+\,\mathbf{u}^T\mathbf{C}_{\mathrm{L}}(\mathbf{w})\mathbf{u}\right)\right), \end{aligned} \tag{22}$$

where $\mathbf{C}_{\mathrm{L}}(\mathbf{w})$ is a diagonal matrix with elements $([\mathbf{w}]_k^{-1})_{k=1}^{K}$. Specifically, $p_{\mathbf{u}|\mathbf{w},\mathbf{y}}$ is a multivariate Gaussian pdf with mean $\overline{\mathbf{u}} = \sigma_{\mathrm{n}}^{-2}(\sigma_{\mathrm{n}}^{-2}\mathbf{A}^T\mathbf{A} + \mathbf{C}_{\mathrm{L}}(\mathbf{w}))^{-1}\mathbf{A}^T\mathbf{y}$ and covariance matrix $\overline{\mathbf{R}} = (\sigma_{\mathrm{n}}^{-2}\mathbf{A}^T\mathbf{A} + \mathbf{C}_{\mathrm{L}}(\mathbf{w}))^{-1}$. There exist several methods for the efficient generation of samples from a multivariate Gaussian density [60], [61], [62], [63].

The conditional distribution $p_{\mathbf{w}|\mathbf{u},\mathbf{y}}$ is

$$p_{\mathbf{w}|\mathbf{u},\mathbf{y}}(\mathbf{w}|\mathbf{u},\mathbf{y}) \propto \prod_{k=1}^{K} p_{w|u,\mathbf{y}}\left([\mathbf{w}]_k|[\mathbf{u}]_k,\mathbf{y}\right), \tag{23}$$

where

$$\begin{aligned} p_{w|u,\mathbf{y}}\left(w|u,\mathbf{y}\right) \propto\ & \exp\left(-\frac{1}{2}\left(\frac{u^2}{w}+b^2 w\right)\right) \\ & \times\ w^{-\frac{1}{2}}\mathbb{1}_+(w) \end{aligned} \tag{24}$$

[1] The pdf of the exponential distribution is

$$p_{\exp}(x) = (1/\lambda)\mathrm{e}^{-x/\lambda}\mathbb{1}_+(x),$$

where $\lambda > 0$ is the scale parameter.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.

belongs to the family of generalized inverse Gaussian distributions[2] with $\lambda_1 = b^2$, $\lambda_2 = u^2$ and $a = 0.5$. We use the method proposed in [64] to draw samples from the pdf in (24).

To summarize, at each iteration q of the constructed blocked Gibbs sampler, we generate $\mathbf{u}^{(q)} \sim p_{\mathbf{u}|\mathbf{w},\mathbf{y}}(\mathbf{u}|\mathbf{w}^{(q-1)}, \mathbf{y})$ and $[\mathbf{w}^{(q)}]_k \sim p_{w|u,\mathbf{y}}(w|[\mathbf{u}^{(q)}]_k, \mathbf{y})$ for all $k \in \{1, \dots, K\}$. The collected samples $\{\mathbf{u}^{(q)}\}_q$ follow the desired distribution $p_{\mathbf{u}|\mathbf{y}}$.

D. Student's t Increments

The case of Student's t increments can be handled by adapting the method shown in [65], which is in fact similar to the one we described for Laplace increments.

The Student's t pdf is given by

$$p_u(u) = \frac{\Gamma(\frac{\alpha+1}{2})}{\Gamma\left(\frac{\alpha}{2}\right)} \frac{1}{\sqrt{\pi}(1+u^2)^{\frac{\alpha+1}{2}}}, \tag{25}$$

where α is the number of degrees of freedom and controls the tail of the distribution, and where Γ denotes the gamma function. It can also be expressed as

$$p_u(u) = \int_{\mathbb{R}} p_{u|w}(u|w) p_w(w)\, \mathrm{d}w, \tag{26}$$

where

$$p_{u|w}(u|w) = \sqrt{\frac{w}{2\pi}} \exp\left(-\frac{wu^2}{2}\right) \tag{27}$$

is a Gaussian pdf and

$$p_w(w) = \frac{(0.5)^{\frac{\alpha}{2}}}{\Gamma(\frac{\alpha}{2})} w^{\frac{\alpha}{2}-1} \exp\left(-\frac{w}{2}\right) \mathbb{1}_+(w) \tag{28}$$

is the pdf of a gamma[3] distribution. Again, we introduce an auxiliary vector $\mathbf{w} \in \mathbb{R}^K$ whose i.i.d. entries follow p_w defined in (28). It is such that

$$p_{\mathbf{u}|\mathbf{w}}(\mathbf{u}|\mathbf{w}) = \prod_{k=1}^{K} p_{u|w}\left([\mathbf{u}]_k|[\mathbf{w}]_k\right), \tag{29}$$

where $p_{u|w}$ is defined in (27).

Here, the distribution $p_{\mathbf{u},\mathbf{w}|\mathbf{y}}$ is given by

$$p_{\mathbf{u},\mathbf{w}|\mathbf{y}}(\mathbf{u},\mathbf{w}|\mathbf{y}) \propto \exp\left(-\frac{1}{2\sigma_{\mathrm{n}}^2}\|\mathbf{y}-\mathbf{A}\mathbf{u}\|_2^2\right) \times \prod_{k=1}^{K} [\mathbf{w}]_k^{\frac{1}{2}} \exp\left(-\frac{[\mathbf{w}]_k [\mathbf{u}]_k^2}{2}\right) \times \prod_{k=1}^{K} [\mathbf{w}]_k^{\frac{\alpha}{2}-1} \exp\left(-\frac{[\mathbf{w}]_k}{2}\right) \mathbb{1}_+([\mathbf{w}]_k), \tag{30}$$

where $\mathbf{A} := \mathbf{H}\mathbf{D}^{-1}$.

Now, the conditional distribution $p_{\mathbf{u}|\mathbf{w},\mathbf{y}}(\mathbf{u}|\mathbf{w},\mathbf{y})$ turns out to be

$$p_{\mathbf{u}|\mathbf{w},\mathbf{y}}(\mathbf{u}|\mathbf{w},\mathbf{y}) \propto \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_{\mathrm{n}}^2}\|\mathbf{y}-\mathbf{A}\mathbf{u}\|_2^2 + \mathbf{u}^T \mathbf{C}_{\mathrm{T}}(\mathbf{w})\mathbf{u}\right)\right), \tag{31}$$

where $\mathbf{C}_{\mathrm{T}}(\mathbf{w})$ is a diagonal matrix with entries $([\mathbf{w}]_k)_{k=1}^K$. Similar to the Laplace case, $p_{\mathbf{u}|\mathbf{w},\mathbf{y}}$ is a multivariate Gaussian density with mean $\overline{\mathbf{u}} = \sigma_{\mathrm{n}}^{-2}(\sigma_{\mathrm{n}}^{-2}\mathbf{A}^T\mathbf{A} + \mathbf{C}_{\mathrm{T}}(\mathbf{w}))^{-1}\mathbf{A}^T\mathbf{y}$ and covariance matrix $\overline{\mathbf{R}} = (\sigma_{\mathrm{n}}^{-2}\mathbf{A}^T\mathbf{A} + \mathbf{C}_{\mathrm{T}}(\mathbf{w}))^{-1}$.

The distribution $p_{\mathbf{w}|\mathbf{u},\mathbf{y}}$ is again separable and takes the form

$$p_{\mathbf{w}|\mathbf{u},\mathbf{y}}(\mathbf{w}|\mathbf{u},\mathbf{y}) \propto \prod_{k=1}^{K} p_{w|u,\mathbf{y}}\left([\mathbf{w}]_k|[\mathbf{u}]_k,\mathbf{y}\right), \tag{32}$$

where

$$p_{w|u,\mathbf{y}}(w|u,\mathbf{y}) \propto \exp\left(-\frac{(1+u)^2 w}{2}\right) \times w^{\frac{\alpha-1}{2}} \mathbb{1}_+(w). \tag{33}$$

is a gamma distribution with $\lambda_1 = \frac{\alpha+1}{2}$ and $\lambda_2 = \frac{2}{(1+u)^2}$, which can easily be sampled from.

E. Bernoulli-Laplace Increments

In [66], Gibbs sampling schemes have been designed for a deconvolution problem where the underlying signal is an i.i.d. spike train that follows the Bernoulli-Gaussian distribution. Unfortunately, the Bernoulli-Gaussian distribution is not infinitely divisible and so is not compatible with our framework of Lévy processes. While there exists some work [67] on Bernoulli-Laplace priors, according to the analysis presented in [66], their proposed sampler would have a tendency to get stuck in certain configurations. Thus, we build upon the method in [66] and develop a novel Gibbs sampler for Lévy processes with Bernoulli-Laplace increments.

The Bernoulli-Laplace pdf is

$$p_u(u) = \lambda\delta(u) + (1-\lambda)\frac{b}{2}\exp\left(-b|u|\right), \tag{34}$$

where $\lambda \in (0,1)$ denotes the mass probability at the origin and b is a scale parameter. We can represent this same density as

$$p_u(u) = \int_{\mathbb{R}} \left(\sum_{v=0}^{1} p_{u|v,w}(u|v,w)p(v)\right) p(w)\, \mathrm{d}w, \tag{35}$$

where

$$p_v(v) = (\lambda)^{1-v}(1-\lambda)^v \quad \text{for } v \in \{0,1\} \tag{36}$$

[2]The pdf of the generalized inverse Gaussian distribution is

$$p_{\mathrm{gig}}(x) = \frac{(\lambda_1/\lambda_2)^{a/2}}{2K_a(\sqrt{\lambda_1\lambda_2})} x^{a-1} \mathrm{e}^{-(\lambda_1 x + \lambda_2/x)/2} \mathbb{1}_+(x),$$

where K_a is the modified Bessel function of the second kind, $\lambda_1 > 0$, $\lambda_2 > 0$, and $a \in \mathbb{R}$.

[3]The pdf of the gamma distribution is

$$p_{\mathrm{gam}}(x) = \frac{1}{\lambda_2^{\lambda_1}\Gamma(\lambda_1)} x^{\lambda_1-1} \mathrm{e}^{-x/\lambda_2} \mathbb{1}_+(x),$$

where $\lambda_1 > 0$ and $\lambda_2 > 0$ are the shape and scale parameters, respectively.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.

is a Bernoulli distribution,

$$p_w(w) = \frac{b^2}{2} \exp\left(- \frac{b^2 w}{2} \right) \mathbb{1}_+(w) \tag{37}$$

is an exponential pdf, and $p_{u|v,w}$ is defined such that

$$p_{u|v,w}(u|v=0,w) = \delta(u) \tag{38}$$

$$p_{u|v,w}(u|v=1,w) = \frac{1}{\sqrt{2\pi w}} \exp\left(- \frac{u^2}{2w} \right). \tag{39}$$

Based on this representation, we introduce two independent auxiliary vectors $\mathbf{v} \in \mathbb{R}^K$ and $\mathbf{w} \in \mathbb{R}^K$. Their elements are i.i.d. and follow the distributions p_v and p_w, as defined in (36) and (37), respectively. Further, these vectors satisfy

$$p_{\mathbf{u}|\mathbf{v},\mathbf{w}}(\mathbf{u}|\mathbf{v},\mathbf{w}) = \prod_{k=1}^{K} p_{u|v,w}\left([\mathbf{u}]_k|[\mathbf{v}]_k,[\mathbf{w}]_k\right), \tag{40}$$

where $p_{u|v,w}$ is defined in (38) and (39).

Here, the full joint distribution $p_{\mathbf{y},\mathbf{u},\mathbf{v},\mathbf{w}}$ is given by

$$\begin{aligned} p_{\mathbf{y},\mathbf{u},\mathbf{v},\mathbf{w}}(\mathbf{y},\mathbf{u},\mathbf{v},\mathbf{w}) &= p_{\mathbf{y}|\mathbf{u},\mathbf{v},\mathbf{w}}(\mathbf{y}|\mathbf{u},\mathbf{v},\mathbf{w}) p_{\mathbf{u},\mathbf{v},\mathbf{w}}(\mathbf{u},\mathbf{v},\mathbf{w}) \\ &= p_{\mathbf{y}|\mathbf{u}}(\mathbf{y}|\mathbf{u}) p_{\mathbf{u}|\mathbf{v},\mathbf{w}}(\mathbf{u}|\mathbf{v},\mathbf{w}) \\ &\quad \times p_{\mathbf{v}}(\mathbf{v}) p_{\mathbf{w}}(\mathbf{w}). \end{aligned} \tag{41}$$

As a result, the distribution $p_{\mathbf{u},\mathbf{v},\mathbf{w}|\mathbf{y}}$ takes the form

$$\begin{aligned} p_{\mathbf{u},\mathbf{v},\mathbf{w}|\mathbf{y}}(\mathbf{u},\mathbf{v},\mathbf{w}|\mathbf{y}) \propto\ & \exp\left(- \frac{1}{2\sigma_{\mathrm{n}}^2} \|\mathbf{y} - \mathbf{A}\mathbf{u}\|_2^2 \right) \\ & \times \prod_{k=1}^{K} p_{u|v,w}\left([\mathbf{u}]_k|[\mathbf{v}]_k,[\mathbf{w}]_k\right) \\ & \times \prod_{k=1}^{K} \lambda^{1-[\mathbf{v}]_k} (1-\lambda)^{[\mathbf{v}]_k} \\ & \times \prod_{k=1}^{K} \frac{b^2}{2} \exp\left(- \frac{b^2 [\mathbf{w}]_k}{2} \right) \mathbb{1}_+([\mathbf{w}]_k), \end{aligned} \tag{42}$$

where $\mathbf{A} = \mathbf{H}\mathbf{D}^{-1}$.

Let us now introduce some notations. For any binary vector $\mathbf{q} \in \mathbb{R}^K$, let $\mathcal{I}_{\mathbf{q},0}$ and $\mathcal{I}_{\mathbf{q},1}$ denote sets of indices such that $[\mathbf{q}]_k = 0$ for $k \in \mathcal{I}_{\mathbf{q},0}$ and $[\mathbf{q}]_k = 1$ for $k \in \mathcal{I}_{\mathbf{q},1}$. Further, let $\mathbf{A}(\mathbf{q})$ be the matrix constructed by taking the columns of $\mathbf{A}$ corresponding to the indices in $\mathcal{I}_{\mathbf{q},1}$. We then define the matrix $\mathbf{B}(\mathbf{q},\mathbf{r}) = \sigma_{\mathrm{n}}^2 \mathbf{I} + \mathbf{A}(\mathbf{q}) \mathbf{C}_{\mathrm{BL}}(\mathbf{q},\mathbf{r}) \mathbf{A}(\mathbf{q})^T$, where $\mathbf{r} \in \mathbb{R}^K$ is a vector with positive entries and $\mathbf{C}_{\mathrm{BL}}(\mathbf{q},\mathbf{r})$ is a diagonal matrix with entries $([\mathbf{r}]_k)_{k \in \mathcal{I}_{\mathbf{q},1}}$. Here, we also introduce the vector $\mathbf{q}_{(-k)} \in \mathbb{R}^{K-1}$ that contains all the entries of $\mathbf{q}$ except the kth one, so that $\mathbf{q}_{(-k)} = ([\mathbf{q}]_1, \ldots, [\mathbf{q}]_{k-1}, [\mathbf{q}]_{k+1}, \ldots, [\mathbf{q}]_K)^T$. Lastly, for $q \in \{0,1\}$, we define the vector $\mathbf{q}^q_{(-k)} \in \mathbb{R}^K$ such that $\mathbf{q}^q_{(-k)} = ([\mathbf{q}]_1, \ldots, [\mathbf{q}]_{k-1}, q, [\mathbf{q}]_{k+1}, \ldots, [\mathbf{q}]_K)^T$.

First, we look at the conditional distribution $p_{\mathbf{u}|\mathbf{v},\mathbf{w},\mathbf{y}}$. From (38) and (42), we deduce that any sample from $p_{\mathbf{u}|\mathbf{v},\mathbf{w},\mathbf{y}}$ takes the value of zero at the indices in $\mathcal{I}_{\mathbf{v},0}$. If we define $\mathbf{u}_1 = ([\mathbf{u}]_k)_{k \in \mathcal{I}_{\mathbf{v},1}}$, then we get

$$\begin{aligned} p_{\mathbf{u}_1|\mathbf{v},\mathbf{w},\mathbf{y}}(\mathbf{u}_1|\mathbf{v},\mathbf{w},\mathbf{y}) \propto \exp\Bigg(- \frac{1}{2} \Bigg(& \frac{1}{\sigma_{\mathrm{n}}^2} \|\mathbf{y} - \mathbf{A}(\mathbf{v})\mathbf{u}_1\|_2^2 \\ & + \mathbf{u}_1^T \mathbf{C}_{\mathrm{BL}}(\mathbf{v},\mathbf{w}) \mathbf{u}_1 \Bigg) \Bigg). \end{aligned} \tag{43}$$

Thus, $p_{\mathbf{u}_1|\mathbf{v},\mathbf{w},\mathbf{y}}$ is a multivariate Gaussian density with mean $\overline{\mathbf{u}_1} = \sigma_{\mathrm{n}}^{-2} (\sigma_{\mathrm{n}}^{-2} \mathbf{A}(\mathbf{v})^T \mathbf{A}(\mathbf{v}) + \mathbf{C}_{\mathrm{BL}}(\mathbf{v},\mathbf{w}))^{-1} \mathbf{A}(\mathbf{v})^T \mathbf{y}$ and covariance matrix $\overline{\mathbf{R}} = (\sigma_{\mathrm{n}}^{-2} \mathbf{A}(\mathbf{v})^T \mathbf{A}(\mathbf{v}) + \mathbf{C}_{\mathrm{BL}}(\mathbf{v},\mathbf{w}))^{-1}$.

The conditional distribution $p_{\mathbf{w}|\mathbf{u},\mathbf{v},\mathbf{y}}$ takes the form

$$p_{\mathbf{w}|\mathbf{u},\mathbf{v},\mathbf{y}}(\mathbf{w}|\mathbf{u},\mathbf{v},\mathbf{y}) \propto \prod_{k=1}^{K} p_{w|u,v,\mathbf{y}}\left([\mathbf{w}]_k|[\mathbf{u}]_k,[\mathbf{v}]_k,\mathbf{y}\right), \tag{44}$$

where $p_{w|u,v,\mathbf{y}}$ is given by

$$p_{w|u,v,\mathbf{y}}(w|u,v=0,\mathbf{y}) \propto \frac{b^2}{2} \exp\left(- \frac{b^2 w}{2} \right) \mathbb{1}_+(w) \tag{45}$$

$$\begin{aligned} p_{w|u,v,\mathbf{y}}(w|u,v=1,\mathbf{y}) \propto\ & \exp\left(- \frac{1}{2} \left(\frac{u^2}{w} + b^2 w \right) \right) \\ & \times\ w^{-\frac{1}{2}} \mathbb{1}_+(w). \end{aligned} \tag{46}$$

The densities in (45) and (46) correspond to the exponential distribution with $\lambda = 2/b^2$ and the generalized inverse Gaussian distribution with $\lambda_1 = b^2$, $\lambda_2 = u^2$, and $a = 0.5$.

Next, inspired by the work in [66], we consider sampling from the marginalized conditional distribution of $[\mathbf{v}]_k$ in a sequential manner as this can allow for a more efficient exploration of configurations of $\mathbf{v}$. More specifically, at each iteration q, we draw $[\mathbf{v}^{(q)}]_k$ from the distribution $p_{[\mathbf{v}]_k|\mathbf{v}_{(-k)},\mathbf{w},\mathbf{y}}(v|\mathbf{v}^{(q)}_{(-k)}, \mathbf{w}^{(q)}, \mathbf{u}^{(q-1)})$, where $\mathbf{v}^{(q)}_{(-k)} = ([\mathbf{v}^{(q)}]_1, \ldots, [\mathbf{v}^{(q)}]_{k-1}, [\mathbf{v}^{(q-1)}]_{k+1}, \ldots, [\mathbf{v}^{(q-1)}]_K)$ and $k \in \{1, \ldots, K\}$.

The marginalized posterior distribution $p_{\mathbf{v},\mathbf{w}|\mathbf{y}}$ is given by

$$p_{\mathbf{v},\mathbf{w}|\mathbf{y}}(\mathbf{v},\mathbf{w}|\mathbf{y}) \propto p_{\mathbf{y}|\mathbf{v},\mathbf{w}}(\mathbf{y}|\mathbf{v},\mathbf{w}) p_{\mathbf{v}}(\mathbf{v}) p_{\mathbf{w}}(\mathbf{w}), \tag{47}$$

where

$$p_{\mathbf{y}|\mathbf{v},\mathbf{w}}(\mathbf{y}|\mathbf{v},\mathbf{w}) = \int_{\mathbb{R}^K} p_{\mathbf{y}|\mathbf{u},\mathbf{v},\mathbf{w}}(\mathbf{y}|\mathbf{u},\mathbf{v},\mathbf{w}) p_{\mathbf{u}|\mathbf{v},\mathbf{w}}(\mathbf{u}|\mathbf{v},\mathbf{w}) \, \mathrm{d}\mathbf{u}. \tag{48}$$

It can be shown that (47) and (48) lead to

$$\begin{aligned} p_{\mathbf{v},\mathbf{w}|\mathbf{y}}(\mathbf{v},\mathbf{w}|\mathbf{y}) \propto\ & |\mathbf{B}(\mathbf{v},\mathbf{w})|^{-\frac{1}{2}} \exp\left(- \frac{1}{2} \mathbf{y}^T \mathbf{B}(\mathbf{v},\mathbf{w})^{-1} \mathbf{y} \right) \\ & \times \prod_{k=1}^{K} \lambda^{1-[\mathbf{v}]_k} (1-\lambda)^{[\mathbf{v}]_k} \\ & \times \prod_{k=1}^{K} \frac{b^2}{2} \exp\left(- \frac{b^2 [\mathbf{w}]_k}{2} \right) \mathbb{1}_+([\mathbf{w}]_k). \end{aligned} \tag{49}$$

From (49), we see that $p_{[\mathbf{v}]_k|\mathbf{v}_{(-k)},\mathbf{w},\mathbf{y}}$ is a Bernoulli distribution with

$$p_{[\mathbf{v}]_k|\mathbf{v}_{(-k)},\mathbf{w},\mathbf{y}}(v|\mathbf{v}_{(-k)},\mathbf{w},\mathbf{y})$$

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.

$$= \left(1+\exp\left(-\frac{1}{2}\left(h\left(1-v;\mathbf{v}_{(-k)},\mathbf{w},\mathbf{y}\right)\right.\right.\right.$$
$$\left.\left.\left.-h\left(v;\mathbf{v}_{(-k)},\mathbf{w},\mathbf{y}\right)\right)\right)\right)^{-1}, \tag{50}$$

where

$$h\left(v;\mathbf{v}_{(-k)},\mathbf{w},\mathbf{y}\right) = \mathbf{y}^T\mathbf{B}\left(\mathbf{v}^v_{(-k)},\mathbf{w}\right)^{-1}\mathbf{y} + \log\left(\left|\mathbf{B}(\mathbf{v}^v_{(-k)},\mathbf{w})\right|\right) + 2v\log\left(\frac{\lambda}{1-\lambda}\right). \tag{51}$$

To summarize, in each iteration q of the above-described sampler, we generate $\mathbf{w}^{(q)} \sim p_{\mathbf{w}|\mathbf{u},\mathbf{v},\mathbf{y}}(\mathbf{w}|\mathbf{u}^{(q-1)},\mathbf{v}^{(q-1)},\mathbf{y})$, $[\mathbf{v}^{(q)}]_k \sim p_{[\mathbf{v}]_k|\mathbf{v}_{(-k)},\mathbf{w},\mathbf{y}}(v|\mathbf{v}^{(q)}_{(-k)},\mathbf{w}^{(q)},\mathbf{u}^{(q-1)})$ for all k and $\mathbf{u}^{(q)} \sim p_{\mathbf{u}|\mathbf{v},\mathbf{w},\mathbf{y}}(\mathbf{u}|\mathbf{v}^{(q)},\mathbf{w}^{(q)},\mathbf{y})$. This particular order of updates is important as it yields a partially collapsed Gibbs sampler [68] where the stationary distribution is still $p_{\mathbf{u},\mathbf{v},\mathbf{w}|\mathbf{y}}$.

VI. Experimental Results

In our experiments, we benchmark the performance of some popular signal reconstruction schemes, including a CNN-based method, on deconvolution and Fourier sampling problems with Lévy processes associated with the Bernoulli-Laplace and Student's t distributions.

A. Signal Models

We consider a signal vector $\mathbf{s} \in \mathbb{R}^{100}$ that contains samples of a Lévy process whose increments follow the Bernoulli-Laplace or Student's t distribution.

1) Bernoulli-Laplace Increments: The Bernoulli-Laplace pdf (34) is characterized by the parameters λ and b, where λ determines the mass probability at the origin and b represents the scale of the Laplace component. We perform experiments for models corresponding to $\lambda \in \{0.6, 0.7, 0.8, 0.9\}$. The scale parameter is set to $b = 1$ for each case.

2) Student't t Increments: The Student's t pdf (25) is parameterized by α, which controls the tails of the distribution. We conduct experiments for $\alpha \in \{1, 3, 5, 39\}$.

B. Measurement Models

We consider both deconvolution and Fourier sampling problems for each of the above-described signal models.

1) Deconvolution: As shown in Section II-C, the system matrix $\mathbf{H}$ for deconvolution is a discrete convolution matrix. Accordingly, we construct $\mathbf{H} : \mathbb{R}^{100} \to \mathbb{R}^{88}$ such that

$$\mathbf{H} = \begin{bmatrix} [\mathbf{h}]_{13} & \cdots & [\mathbf{h}]_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & [\mathbf{h}]_{13} & \cdots & [\mathbf{h}]_1 \end{bmatrix}, \tag{52}$$

where $\mathbf{h} \in \mathbb{R}^{13}$ consists of the central samples of a truncated Gaussian PSF with variance $\sigma_0^2 = 4$.

2) Fourier Sampling: For Fourier sampling in 1D, which is reminiscent of MRI, the forward model $\mathbf{H}$ resembles a discrete Fourier matrix (see Section II-D). Thus, in order to construct $\mathbf{H}$, we first sample $M' = 16$ rows of the DFT matrix. The first row of the DFT matrix (DC component) is always kept, while the remaining ones are selected in a quasi-random fashion with a denser sampling at low frequencies. We then create the real system matrix $\mathbf{H} : \mathbb{R}^{100} \to \mathbb{R}^M$, where $M = 2M' - 1$, by separating the real and imaginary parts.

In both measurement models, the AWGN variance σ_{n}^2 is chosen such that the (average) measurement SNR is around 30 dB.

C. Reconstruction Algorithms

For each combination of the signal and measurement models, we compare the performance of some (variational) model-based techniques, a CNN-based scheme and the MMSE estimator. We generate validation and test datasets, each consisting of 1,000 pairs of ground-truth signals and their noisy measurements. Further, in order to train the CNNs, we also synthesize a repository $\mathcal{R}$ containing a large number of training examples.

1) Model-Based methods: We consider the model-based methods

$$\widehat{\mathbf{s}}_{\ell_2} = \arg\min_{\mathbf{s}\in\mathbb{R}^K}\left(\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \tau\|\mathbf{D}\mathbf{s}\|_2^2\right), \tag{53}$$

$$\widehat{\mathbf{s}}_{\ell_1} = \arg\min_{\mathbf{s}\in\mathbb{R}^K}\left(\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \tau\|\mathbf{D}\mathbf{s}\|_1\right), \tag{54}$$

and

$$\widehat{\mathbf{s}}_{\text{log}} = \arg\min_{\mathbf{s}\in\mathbb{R}^K}\left(\|\mathbf{y}-\mathbf{H}\mathbf{s}\|_2^2 + \tau\sum_{k=1}^{K}\log\left(1+([\mathbf{D}\mathbf{s}]_k)^2\right)\right), \tag{55}$$

where $\tau \in \mathbb{R}_+$. (53), (54) and (55) resemble the MAP estimators of Lévy processes associated with Gaussian, Laplace, and Student's t distributions, respectively. However, unlike the MAP estimators, these include an adjustable hyperparameter τ. For each of these methods, the same regularization parameter τ is used for the entire test dataset. This particular value of τ is the one that yields the lowest MSE for the validation dataset.

These estimators are implemented in MATLAB using GlobalBioIm [69]—a library for solving inverse problems. Specifically, the ℓ_2 estimator is expressed in closed-form as

$$\widehat{\mathbf{s}}_{\ell_2} = \left(\mathbf{H}^T\mathbf{H} + \tau\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{H}^T\mathbf{y}. \tag{56}$$

The ℓ_1 and log estimators are computed iteratively using ADMM. Since the cost functional in (55) is non-convex, we initialize ADMM for $\widehat{\mathbf{s}}_{\text{log}}$ with $\widehat{\mathbf{s}}_{\ell_1}$ so that it can reach a better local minimum.

2) CNN-Based Method: The concept here is to train a CNN as a regressor that maps an initial low-quality reconstruction $\widehat{\mathbf{s}}_0$ to a high-quality one $\widehat{\mathbf{s}}_{\text{CNN}}$ [20], [21], [22], [23], [24]. The architecture of the CNN used in our experiments is based on the well-known denoising network DnCNN [70] and is described in Fig. 2 and Table I.

First, we build a training dataset of cardinality M_T by taking the first M_T examples $\{\mathbf{s}_m, \mathbf{y}_m\}_{m=1}^{M_T}$ from the repository $\mathcal{R}$. We

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.

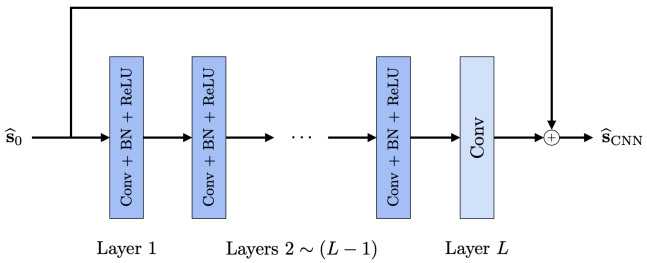


Fig. 2. Architecture of the CNN, where BN denotes the operation of batch normalization.

TABLE I
CONVOLUTION LAYERS

Layer	Filter size	Input channels	Output channels
1	$(F \times 1)$	1	C
$2 \sim (L-1)$	$(F \times 1)$	C	C
L	$(F \times 1)$	C	1

then train the model by minimizing the MSE loss function

$$\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{M_T} \sum_{m=1}^{M_T} \| \mathbf{s}_m - \widehat{\mathbf{s}}_{\text{CNN}} \left(\boldsymbol{\theta}; \widehat{\mathbf{s}}_0(\mathbf{y}_m) \right) \|_2^2, \tag{57}$$

where $\boldsymbol{\theta}$ represents the learnable parameters of the network, with the help of the ADAM optimizer [71]. The CNN is trained for 1,000 epochs with a batch size of 256 and a weight decay of γ. The initial learning rate is set as 10^{-2}. For some duration of the training (first 600 epochs for deconvolution and first 750 epochs for Fourier sampling), it is decreased by a factor of 0.5 every 50 epochs. We choose the initial low-quality reconstruction to be $\widehat{\mathbf{s}}_0(\mathbf{y}) = \mathbf{H}^T \mathbf{y}$ for the deconvolution problems. In the case of Fourier sampling, $\widehat{\mathbf{s}}_0(\mathbf{y})$ is the zero-filled reconstruction. All the CNN-based reconstruction schemes are implemented in PyTorch.

3) Goldstandard (MMSE Estimator): Our MMSE estimators are implemented in MATLAB, according to the methods detailed in Section V. There, we set the number of samples as $Q = 8{,}000$ and the burn-in period as $B = 3{,}000$ for signals with Bernoulli-Laplace increments. We use $Q = 15{,}000$ and $B = 5{,}000$ for signals associated with the Student's t distribution.

D. Results

We present our results for all the test datasets in Figs. 3, 4, 5 and 6. For the sake of clarity, instead of the MSE, we display the "MSE optimality gap" which is the difference between the MSE obtained by a specific method and the MSE attained by the MMSE estimator. In these figures, the CNNs are labelled using the tuple (F, C, L, M_T, γ), where F is the filter size, C is the number of channels, L is the number of layers, M_T is the cardinality of the training dataset and γ is the weight decay. For the interested reader, we also provide information about the computation times required by all the methods in the supplementary material.

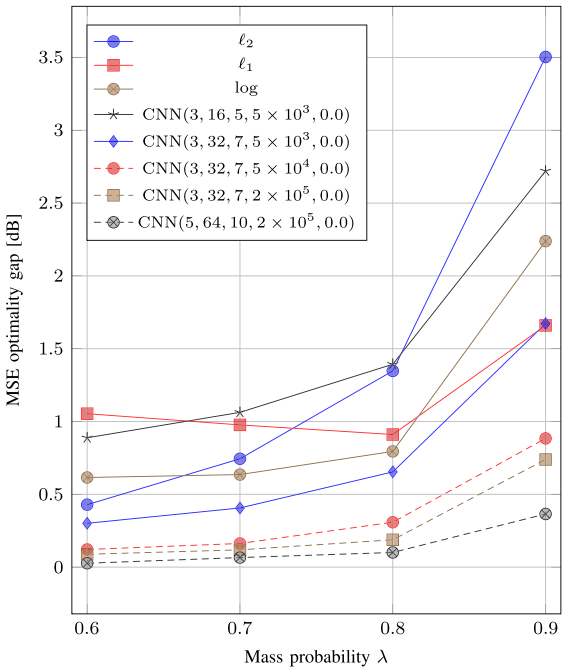


Fig. 3. Deconvolution of Lévy processes with Bernoulli-Laplace increments.

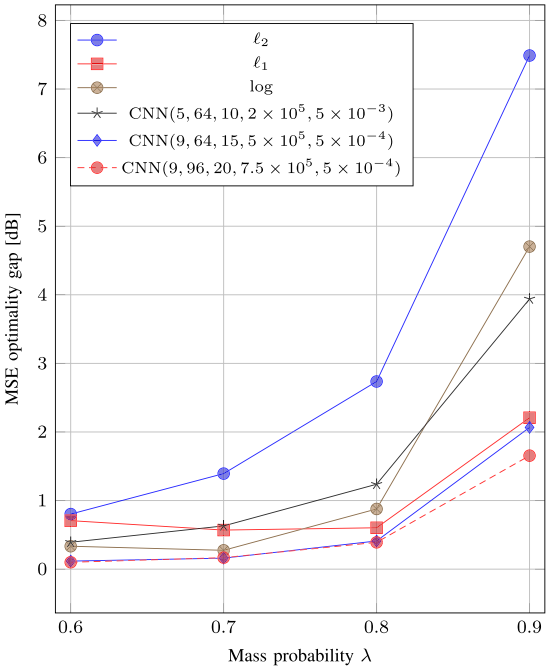


Fig. 4. Fourier sampling of Lévy processes with Bernoulli-Laplace increments.

1) Lévy Processes with Bernoulli-Laplace Increments: Here, we summarize our observations for both the deconvolution and Fourier sampling experiments (Figs. 3 and 4). The sparsity-promoting ℓ_1 estimator, which corresponds to the popular TV

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.

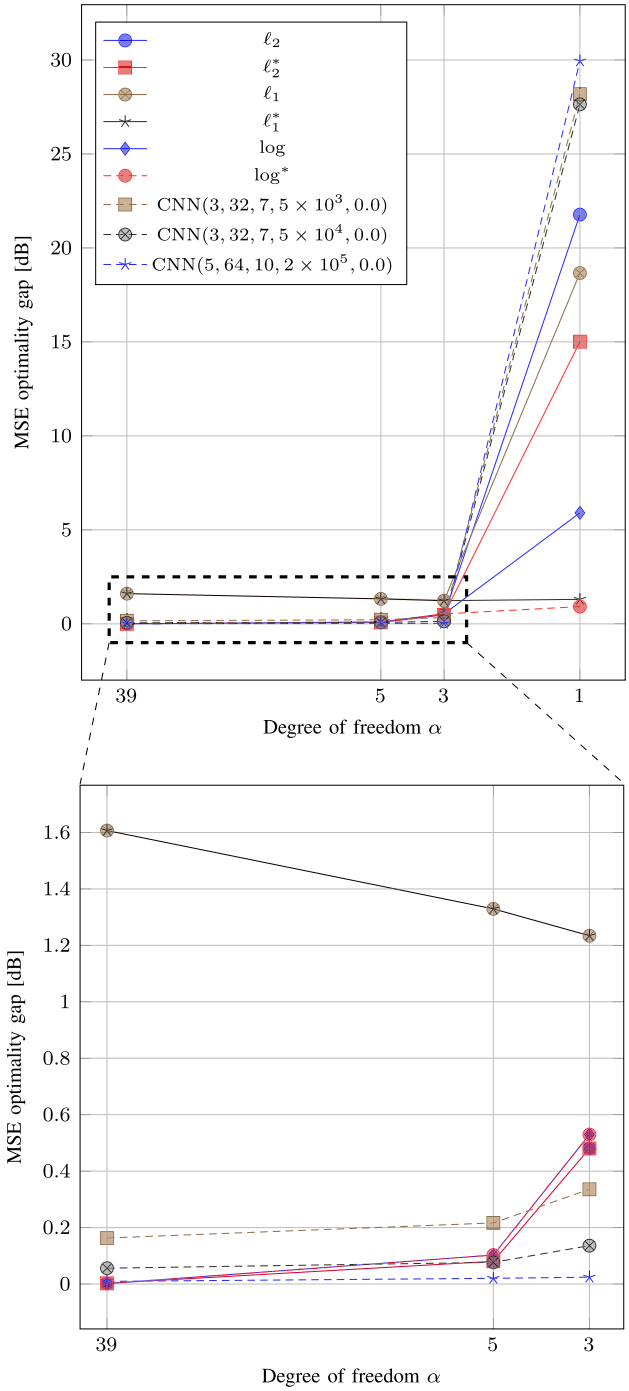


Fig. 5. Deconvolution of Lévy processes with Student's t increments. The figure at the bottom is a zoomed-in version of the dotted rectangular box shown in the figure at the top.

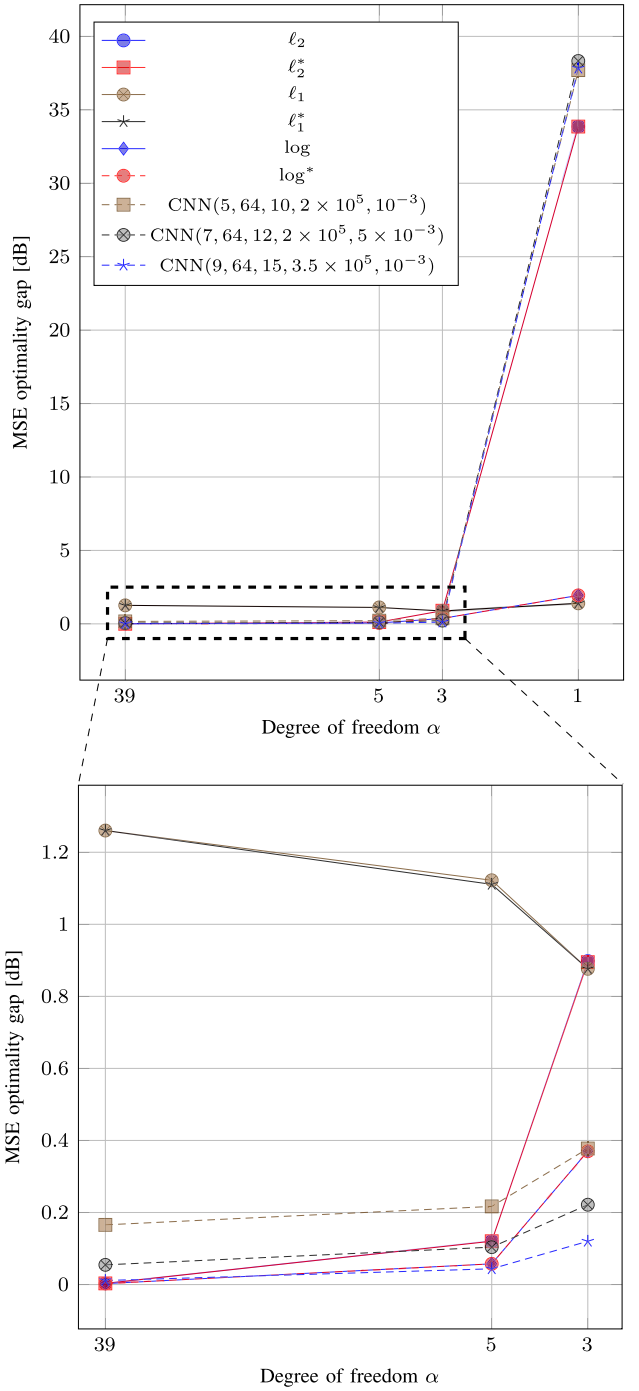


Fig. 6. Fourier sampling of Lévy processes with Student's t increments. The figure at the bottom is a zoomed-in version of the dotted rectangular box shown in the figure at the top.

regularization, is known to be well-suited to piecewise-constant Lévy processes with Bernoulli-Laplace increments. As the value of λ increases, these signals become sparser and exhibit fewer jumps. Consequently, we observe that the ℓ_1 estimator performs better than the ℓ_2 estimator. The log estimator also promotes sparse solutions [72] and we see that it performs well for these piecewise-constant signals. However, despite the good fit, there is still some gap between the MSE attained by the ℓ_1 and log, and MMSE estimators.

The performance of the CNN-based method improves as we increase the capacity of the CNN and the amount of training data. In fact, with sufficient capacity and training data, they

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.

outperform the ℓ_1 and log estimators. Remarkably, some of the CNNs achieve a near-optimal MSE.

2) Lévy Processes With Student's t Increments: The parameter α allows us to consider a wide range of signals. As $\alpha \to \infty$, we approach the Gaussian regime. The other extreme is $\alpha = 1$, which corresponds to the super heavy-tailed (sparse) Cauchy distribution. This scenario can be particularly challenging for the correct setting of algorithm parameters. Due to the heavy tails of the Cauchy distribution, the validation and test datasets may contain signals with a vastly different range of values. Consequently, for a given model-based method, the regularization parameter τ that is chosen to yield the lowest MSE for the validation dataset may differ significantly from the value τ^* that achieves the lowest MSE on the test dataset. Thus, in Figs. 5 and 6, we also include the performance of model-based methods when their regularization parameter is tuned for optimal MSE performance on the test dataset directly. These "boosted" model-based methods are labelled as ℓ_2^*, ℓ_1^* and $\log^*$.

In Figs. 5 and 6, we can see that the ℓ_2 estimator is optimal for a large value of α. As the value of α decreases, the performance of the ℓ_2 estimator deteriorates and becomes worse than that of the ℓ_1 estimator. For all the cases, the log estimator, which corresponds to a tunable MAP estimator for the Student's t distribution, attains reasonable MSE values. Note that for the deconvolution experiment involving Cauchy signals, there is a significant gap between the MSE values obtained by the ℓ_2 and ℓ_1 and ℓ_2^* and ℓ_1^* estimators, respectively. Interestingly, the log estimator is less affected by this issue.

Finally, for both deconvolution and Fourier sampling problems, CNNs with sufficient capacity and training data perform well up to $\alpha = 3$, after which there seems to be a steep transition and their performance drops sharply. In fact, for Cauchy signals, we observe that the training process for these CNNs is quite unstable—the training loss marginally decreases and seems to converge, and the networks do not generalize to the validation (or test) datasets. We believe that this last example poses an open challenge for designing robust neural-network-based schemes that can handle signals following (super) heavy-tailed distributions.

VII. Conclusion

We have introduced a controlled environment, based on sparse stochastic processes (SSPs), for the objective benchmarking of reconstruction algorithms, including CNN-based methods that require lots of training data, in the context of linear inverse problems. We have developed efficient posterior sampling schemes to compute the minimum-mean-square-error estimators for specific classes of SSPs. These yield the upper limit on reconstruction performance and allow us to provide a measure of statistical optimality. We have highlighted the abilities of our framework by benchmarking some popular variational methods and convolutional neural-network (CNN) architectures for deconvolution and Fourier-sampling problems. In particular, we have observed that, while CNNs outperform the variational methods and achieve a near-optimal performance in terms of mean-square error for a wide range of conditions, they can sometimes fail too, especially for signals with heavy-tailed innovations.

Acknowledgement

We acknowledge access to the facilities and expertise of the CIBM Center for Biomedical Imaging, a Swiss research center of excellence founded and supported by Lausanne University Hospital (CHUV), University of Lausanne (UNIL), École polytechnique fédérale de Lausanne (EPFL), University of Geneva (UNIGE), and Geneva University Hospitals (HUG).

References

[1] M. Unser and M. T. McCann, "Biomedical image reconstruction: From the foundations to deep neural networks," *Found. Trends Signal Process.*, vol. 13, pp. 280–359, 2019.
[2] A. N. Tikhonov, "Solution of incorrectly formulated problems and the regularization method," *Sov. Math.*, vol. 4, pp. 1035–1038, 1963.
[3] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*. Boca Raton, FL, USA: CRC Press, 1998.
[4] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall,1993.
[5] S. Mallat, *A Wavelet Tour of Signal Processing*. Amsterdam, The Netherlands: Elsevier, 1999.
[6] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.
[7] R. G. Baraniuk, E. Candès, M. Elad, and Y. Ma, "Applications of sparse representation and compressive sensing [scanning the issue]," *Proc. IEEE*, vol. 98, no. 6, pp. 906–909, Jun. 2010.
[8] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, vol. 2. Berlin, Germany: Springer, 2010.
[9] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[10] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[11] S. Foucart and H. Rauhut, "An invitation to compressive sensing," in *A Mathematical Introduction to Compressive Sensing*. Berlin, Germany: Springer, 2013, pp. 1–39.
[12] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.
[13] P. del Aguila Pla, S. Neumayer, and M. Unser, "Stability of image-reconstruction algorithms," *IEEE Trans. Comput. Imag.*, vol. 9, pp. 1–12, 2023.
[14] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
[15] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.
[16] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1/4, pp. 259–268, 1992.
[17] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, Jan. 2010.
[18] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 85–95, Nov. 2017.
[19] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, "Deep learning techniques for inverse problems in imaging," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 39–56, May 2020.
[20] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.
[21] H. Chen et al., "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.

[22] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo, "Deep learning for undersampled MRI reconstruction," *Phys. Med. Biol.*, vol. 63, no. 13, 2018, Art. no. 135007.
[23] K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," *Opt. Exp.*, vol. 27, no. 20, pp. 28075–28090, 2019.
[24] D. Perdios, M. Vonlanthen, F. Martinez, M. Arditi, and J.-P. Thiran, "CNN-based image reconstruction method for ultrafast ultrasound imaging," *IEEE Trans. Ultrason., Ferroelectr, Freq. Control*, vol. 69, no. 4, pp. 1154–1168, Apr. 2022.
[25] K. Gregor and Y. LeCun, "Learning fast approximations of sparse coding," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 399–406.
[26] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1256–1272, Jun. 2016.
[27] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-Net for compressive sensing MRI," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 10–18.
[28] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 394–405, Feb. 2019.
[29] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1322–1332, Jun. 2018.
[30] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, Mar. 2021.
[31] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2013, pp. 945–948.
[32] E. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 5546–5557.
[33] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6360–6376, Oct. 2022.
[34] Y. Sun, Z. Wu, X. Xu, B. Wohlberg, and U. S. Kamilov, "Scalable plug-and-play ADMM with convergence guarantees," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 849–863, 2021.
[35] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (RED)," *SIAM J. Imag. Sci.*, vol. 10, no. 4, pp. 1804–1844, 2017.
[36] Y. Sun, J. Liu, and U. S. Kamilov, "Block coordinate regularization by denoising," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 382–392.
[37] Z. Wu, Y. Sun, A. Matlock, J. Liu, L. Tian, and U. S. Kamilov, "SIMBA: Scalable inversion in optical tomography using deep denoising priors," *IEEE J. Sel. Topics Signal Process.*, vol. 14, no. 6, pp. 1163–1175, Oct. 2020.
[38] J. H. R. Chang, C.-L. Li, B. Poczos, B. V. K. Vijaya Kumar, and A. C. Sankaranarayanan, "One network to solve them all–Solving linear inverse problems using deep projection models," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 5888–5897.
[39] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser, "CNN-based projected gradient descent for consistent CT image reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1440–1453, Jun. 2018.
[40] M. Unser and P. D. Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge, U.K.: Cambridge Univ. Press, 2014 .
[41] U. S. Kamilov, P. Pad, A. Amini, and M. Unser, "MMSE estimation of sparse Lévy processes," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 137–147, Jan. 2013.
[42] A. Amini, U. S. Kamilov, E. Bostan, and M. Unser, "Bayesian estimation for continuous-time sparse stochastic processes," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 907–920, Feb. 2013.
[43] M. Unser, "Sampling–50 Years after Shannon," *Proc. IEEE*, vol. 88, no. 4, pp. 569–587, Apr. 2000.
[44] S. Ken-Iti, *Lévy Processes and Infinitely Divisible Distributions*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
[45] A. Amini and M. Unser, "Sparsity and infinite divisibility," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2346–2358, Apr. 2014.
[46] 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.
[47] J. R. Fageot, "Gaussian versus sparse stochastic processes: Construction, regularity, compressibility," Ph.D. dissertation, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2017.
[48] M. Unser, "Ridges, neural networks, and the Radon transform," *J. Mach. Learn. Res.*, vol. 24, no. 37, pp. 1–33, 2023.
[49] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
[50] C. J. Geyer, "Practical Markov chain Monte Carlo," *Statist. Sci.*, vol. 7, no. 4, pp. 473–483, 1992.
[51] W. R. Gilks, S. Richardson, and D. Spiegelhalter, *Markov chain Monte Carlo in Practice*. Boca Raton, FL, USA: CRC Press, 1995.
[52] D. Gamerman and H. F. Lopes, *Markov chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Boca Raton, FL, USA: CRC Press, 2006.
[53] M. I. Gordin and B. A. Lifšic, "The central limit theorem for stationary Markov processes," *Doklady Akademii Nauk*, vol. 239, no. 4, pp. 766–767, 1978.
[54] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.
[55] G. Casella and E. I. George, "Explaining the Gibbs sampler," *Amer. Statist.*, vol. 46, no. 3, pp. 167–174, 1992.
[56] M. A. Tanner and W. H. Wong, "The calculation of posterior distributions by data augmentation," *J. Amer. Stat. Assoc.*, vol. 82, no. 398, pp. 528–540, 1987.
[57] A. Mira and L. Tierney, "On the use of auxiliary variables in Markov chain Monte Carlo sampling," *Scand. J. Statist.*, 1997.
[58] T. Park and G. Casella, "The Bayesian LASSO," *J. Amer. Stat. Assoc.*, vol. 103, no. 482, pp. 681–686, 2008.
[59] D. F. Andrews and C. L. Mallows, "Scale mixtures of normal distributions," *J. Roy. Stat. Soc.: Ser. B.*, vol. 36, no. 1, pp. 99–102, 1974.
[60] H. Rue, "Fast sampling of Gaussian Markov random fields," *J. Roy. Stat. Soc.: Ser. B. (Stat. Methodol.)*, vol. 63, no. 2, pp. 325–338, 2001.
[61] F. Orieux, O. Féron, and J.-F. Giovannelli, "Sampling high-dimensional Gaussian distributions for general linear inverse problems," *IEEE Signal Process. Lett.*, vol. 19, no. 5, pp. 251–254, May 2012.
[62] C. Gilavert, S. Moussaoui, and J. Idier, "Efficient Gaussian sampling for solving large-scale inverse problems using MCMC," *IEEE Trans. Signal Process.*, vol. 63, no. 1, pp. 70–80, Jan. 2015.
[63] M. Vono, N. Dobigeon, and P. Chainais, "High-dimensional Gaussian sampling: A review and a unifying approach based on a stochastic proximal point algorithm," *SIAM Rev.*, vol. 64, no. 1, pp. 3–56, 2022.
[64] L. Devroye, "Random variate generation for the generalized inverse Gaussian distribution," *Statist. Comput.*, vol. 24, no. 2, pp. 239–246, 2014.
[65] C. Fevotte and S. J. Godsill, "A Bayesian approach for blind separation of sparse sources," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 6, pp. 2174–2188, Nov. 2006.
[66] D. Ge, J. Idier, and E. Le Carpentier, "Enhanced sampling schemes for MCMC based blind Bernoulli–Gaussian deconvolution," *Signal Process.*, vol. 91, no. 4, pp. 759–772, 2011.
[67] L. Chaari, J.-Y. Tourneret, and H. Batatia, "Sparse Bayesian regularization using Bernoulli-Laplacian priors," in *Proc. IEEE 21st Eur. Signal Process. Conf.*, 2013, pp. 1–5.
[68] D. A. Van Dyk and T. Park, "Partially collapsed Gibbs samplers: Theory and methods," *J. Amer. Stat. Assoc.*, vol. 103, no. 482, pp. 790–796, 2008.
[69] E. Soubies et al., "Pocket guide to solve inverse problems with GlobalBioIm," *Inverse Problems*, vol. 35, no. 10, 2019, Art no. 104006.
[70] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[71] D. Kingma and J. Ba, "ADAM: A. method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015.
[72] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 317–329, Apr. 2010.

Authorized licensed use limited to: EPFL LAUSANNE. Downloaded on June 27,2023 at 12:14:53 UTC from IEEE Xplore. Restrictions apply.