MAP Estimators for Self-Similar Sparse Stochastic Models

Emrah Bostan, Julien Fageot, Ulugbek S. Kamilov and Michael Unser

Biomedical Imaging Group, EPFL, Lausanne, Switzerland

***Abstract*—We consider the reconstruction of multi-dimensional signals from noisy samples. The problem is formulated within the framework of the theory of continuous-domain sparse stochastic processes. In particular, we study the fractional Laplacian as the whitening operator specifying the correlation structure of the model. We then derive a class of MAP estimators where the priors are confined to the family of infinitely divisible distributions. Finally, we provide simulations where the derived estimators are compared against total-variation (TV) denoising.**

***Index Terms*—Innovation models, fractional Laplacian, fractals, invariance, self-similarity, sparse stochastic processes, MAP estimation.**

I. Introduction

Consider the signal denoising problem where the goal is to estimate the unknown signal $\mathbf{s} \in \mathbb{R}^K$ from the noisy measurement

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

where the vector $\mathbf{n} \in \mathbb{R}^K$ represents the noise that is assumed to be i.i.d. Gaussian with variance σ^2.

We consider the statistical formulation of the denoising problem based on the prior knowledge of the distribution of the signal and concentrate on MAP estimators. To that end, we first specify a *continuous-domain* signal model by using the theory of sparse stochastic processes [1]. The model has two fundamental elements: an *innovation process* governing the sparsity pattern and the *whitening operator* determining the correlation structure of the underlying signal.

The contribution of this work is to extend our previous line of work [2], [3] by using fractional-order Laplacians $(-\Delta)^{\gamma/2}$ with $\gamma > 0$ as our whitening operator. The unique feature of these operators is their invariance to translation, scaling, and rotation [4]. They also have been associated with $1/\|\omega\|^\gamma$-type power spectrum that appears in natural images [5], [6]. In this prospective, the derived estimators are suitable for removing noise from fractal-like images. We perform simulations and show that the derived estimators can improve upon TV denoising for particular images.

II. Mathematical foundations

We assume that the underlying signal $\mathbf{s}$ is the discretized version of a stochastic process $s(\mathbf{r})$ in $\mathbb{R}^d$ that is defined as the solution of the stochastic differential equation

$$\mathrm{L}s = w, \tag{2}$$

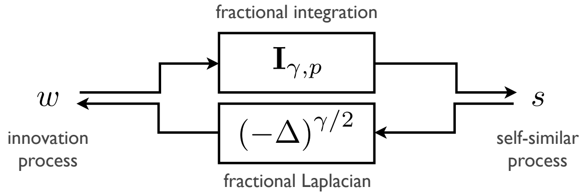

Fig. 1. Continuous-domain innovation model.

where w is a continuous-domain white noise that is *not necessarily Gaussian* and L is a suitable differential operator. What (2) implies is that the formal solution (if it exists) is given by $s = \mathrm{L}^{-1}w$. Therefore, the correlation structure of s is determined by the mixing operator L^{-1}, while its sparsity pattern is characterized by w that we shall call the *innovation process* (see Section II-A).

In the sequel, we restrict L to be in the subclass of *fractional Laplacians* (Section II-B). Our goal is to define a general class of *self-similar processes* (Section II-C) as illustrated in Figure 1.

A. Innovation processes

We define continuous-domain innovation processes in the framework of generalized functions of Schwartz [7]. In the one-dimensional setting, they are the weak derivative of the family of Lévy processes. As a member of the family of generalized stochastic processes, w is a random generalized function that is observed through scalar-products with test functions φ in the space $\mathcal{S}$ of smooth and rapidly decreasing functions. Hence, for fixed φ, the linear observation $\langle w, \varphi \rangle$ is a real random variable.

The innovation process w is a stationary stochastic process with independent value at every point. Its statistical properties are characterized by its characteristic functional (the infinite-dimensional generalization of the characteristic function of random variables) $\widehat{\mathscr{P}}_w(\varphi) = \mathbb{E}[e^{j\langle w,\varphi\rangle}]$. The characteristic functional of w has the general form

$$\widehat{\mathscr{P}}_w(\varphi) = \exp\left(\int_{\mathbb{R}^d} f(\varphi(\mathbf{r}))d\mathbf{r}\right), \tag{3}$$

where $f(\cdot)$ is called the Lévy exponent of w. The set of admissible Lévy exponents, and thus of innovation processes, is described in [7].

It is important to note that Lévy exponents are also in one-to-one correspondence with the so-called infinitely divisible (i.d.) distributions. Indeed, the characteristic functions of i.d. random variables are precisely of the form $e^{f(\omega)}$ [8]. Equivalently, an innovation process is characterized by its canonical pdf defined by $p_{\text{i.d.}} = \mathcal{F}^{-1}\left\{e^{f(\omega)}\right\}$, where $\mathcal{F}$ denotes the Fourier transform.

B. Fractional Laplacian operators and their inverses

As mentioned earlier, we choose L to be a member of the class of fractional Laplacian operators $(-\Delta)^{\gamma/2}$ for $\gamma > 0$. These isotropic differential operators defined in Fourier domain by

$$\mathcal{F}\left\{(-\Delta)^{\gamma/2}\varphi\right\}(\boldsymbol{\omega}) = |\boldsymbol{\omega}|^{\gamma}\mathcal{F}\{\varphi\}(\boldsymbol{\omega}),$$

where $\varphi \in \mathcal{S}$.

The fractional Laplacian is a linear, self-adjoint, and continuous operator with translation-, rotation-, and scaling-invariance properties. Its inverse operator is the Riesz potential I_{γ} for $\gamma < d$ and is extended for all non-integer $\gamma > d$ in Sun and Unser [4]. It has been also observed that the natural translation-invariant inverse I_{γ} of the fractional Laplacian operator can be unstable.

In accordance with this previous work, we define $\mathrm{I}_{\gamma,p}$ as the unique corrected version of the inverse mapping from $\mathcal{S}$ to the space L^p of functions with finite p-norm $\left(\int_{\mathbb{R}^d}|f(\mathbf{r})|^p d\mathbf{r}\right)^{1/p}$. The L^p stability comes with the cost of losing the translation-invariance for the inverse operator.

C. Self-similar processes

We now would like to define the process s. As we consider processes s such that $(-\Delta)^{\gamma/2}s = w$ is an innovation process, one formally writes

$$\langle s, \varphi\rangle = \langle \mathrm{I}_{\gamma,p}w, \varphi\rangle = \langle w, \mathrm{I}^*_{\gamma,p}\varphi\rangle,$$

where $\mathrm{I}_{\gamma,p}$ is the corrected inverse operator of $(-\Delta)^{\gamma/2}$ defined above.

To satisfy the admissibility conditions required between the Lévy exponent $f(\cdot)$ of w and the stability property of the inverse operator [1], one needs $p = 1$ for the Laplace innovations and $p = 2$ for the Gaussian ones. The characteristic functional of s is then given by

$$\widehat{\mathscr{P}}_s(\varphi) = \exp\left(\int_{\mathbb{R}^d} f\left(\mathrm{I}^*_{\gamma,p}\varphi(\mathbf{r})\right)d\mathbf{r}\right). \tag{4}$$

The resulting process s is called self-similar (in a stochastic sense) because an application of a similarity transformation such as scaling does not change its statistical behavior (up to some possible renormalization).

III. MAP ESTIMATION

After explaining that the process s is mathematically well-defined, we now concentrate on developing practical algorithms.

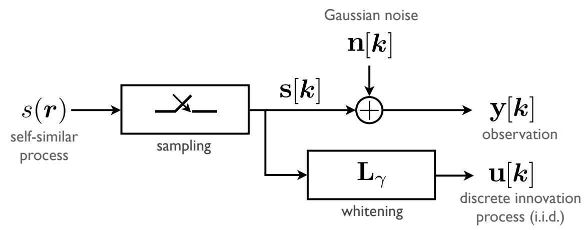

Fig. 2. Observation model.

A. Discrete innovation model

The discrete counterpart of innovation model (2) is obtained by introducing the discrete version $\mathbf{L}$ of the operator L [9]. Since we are only given the sampled version of s in real-world applications, one can think of formulating the discrete innovation model by applying $\mathbf{L}$ to the sampled process $\boldsymbol{s}[\boldsymbol{k}] = s(\mathbf{r} = \boldsymbol{k})$ for $\boldsymbol{k}$ being in a suitable discrete space Ω. In the case of fractional Laplacian operator, $\mathbf{L}_{\gamma}$ is efficiently implemented in Fourier domain via FFT operation. In effect, we define the discretized version of (2) as

$$\mathbf{u} = \mathbf{L}_{\gamma}\mathbf{s} \tag{5}$$

where $\mathbf{u}$ is called the discrete innovation process whose first-order pdf p_U is proven to be an infinitely divisible distribution [3].

As shown in [3], it is equivalent to define the discrete counterpart $(-\Delta)^{\gamma/2}_d$ of the operator $(-\Delta)^{\gamma/2}$ such that $\mathbf{u}[\boldsymbol{k}] = \left\{(-\Delta)^{\gamma/2}_d s\right\}(\mathbf{r} = \boldsymbol{k})$. In other terms, we have

$$\mathbf{u}[\boldsymbol{k}] = (\beta_{\gamma,p} * w)(\mathbf{r} = \boldsymbol{k}),$$

where $\beta_{\gamma,p} \in L^1$ is a polyharmonic B-spline and is the impulse response of the operator $(-\Delta)^{\gamma/2}_d \mathbf{I}_{\gamma,p}$. We note that the primary statistical features of $\mathbf{u}$ is related to the continuous-domain innovation process w.

Proposition 1. *If $p_{\text{i.d.}} = \mathcal{F}^{-1}\left\{e^{f(\omega)}\right\}$ is symmetric α-stable (in particular Gaussian case), then the same is true for p_U. If $p_{\text{i.d.}}$ is symmetric, unimodal with exponential decay, then the same is true for p_U.*

B. MAP estimation

We now formulate the MAP estimators for the denoising problem given in (1) under the assumption that the components $(\mathbf{u}[\boldsymbol{k}])_{\boldsymbol{k}\in\Omega}$ are i.i.d. random variables. We then get the posterior distribution $p_{S|Y}$ from the Bayes' rule

$$\begin{aligned} p_{S|Y}(\mathbf{s}|\mathbf{y}) &\propto p_N(\mathbf{y}-\mathbf{s})p_U(\mathbf{u}) \\ &\propto \exp\left(-\frac{\|\mathbf{y}-\mathbf{s}\|_2^2}{2\sigma^2}\right)\prod_{\boldsymbol{k}\in\Omega} p_U(\mathbf{L}_{\gamma}\mathbf{s}[\boldsymbol{k}]). \end{aligned}$$

We define the *potential function* $\Phi_U(x) = -\log p_U(x)$. Then, the MAP estimator $\mathbf{s}_{\text{MAP}} = \arg\max_{\mathbf{s}} p_{S|Y}(\mathbf{s}|\mathbf{y})$ is given by,

$$\begin{aligned} \mathbf{s}_{\text{MAP}} = \quad & \arg\min_{\mathbf{s}} \quad \frac{1}{2}\|\mathbf{y}-\mathbf{s}\|_2^2 + \sigma^2\sum_{\boldsymbol{k}\in\Omega}\Phi_U(\mathbf{u}[\boldsymbol{k}]) \\ & \text{subject to} \quad \mathbf{u} = \mathbf{L}_{\gamma}\mathbf{s}. \end{aligned} \tag{6}$$

By using the previous definitions and the inverse Fourier transform, we arrive at

$$\Phi_U(x) = -\log\left(\int_{\mathbb{R}} \exp\left(\int_{\mathbb{R}^d} f(\beta_{\gamma,p}(\mathbf{r})\omega)d\mathbf{r} + j\omega x\right)\frac{d\omega}{2\pi}\right). \quad (7)$$

We now characterize the asymptotical form of Φ_U using the Lévy exponent of the underlying innovation process.

Theorem 1. *There exist constants A_1, A_2, A_3, B_1, B_2 and B_3 depending on the parameter of the considered innovation such that*

- *If $f(\omega) = -\sigma_0^2\omega^2$ (Gaussian case),*

$$\Phi_U(x) = A_1x^2 + B_1$$

- *If $f(\omega) = -s|\omega|^\alpha$ (α-stable case),*

$$\Phi_U(x) \sim A_2\log(|x|) + B_2$$

- *If $f(\omega) = \log\left(\frac{\lambda^2}{\lambda^2+\omega^2}\right)$ (Laplace case),*

$$\Phi_U(x) \sim A_3|x| + B_3$$

where $f \sim g$ denotes that $f - g \to 0$.

Since the computation of the exact potential function (7) is challenging in the case of Laplacian innovation, we use its simplified asymptotic form $\Phi_U(x) = A_3|x| + B_3$. Note that the constants in Theorem 1 are irrelevant for the optimization task.

IV. Numerical Examples

We perform a simple simulation that compares the estimation performance of different estimators specified by our formalism. Particulary, we concentrate on denoising of a natural texture-type and a biological image that are shown in Figure 3. We consider two innovation processes (Gaussian and Laplacian). We also consider two different whitening operators: fractional Laplacian and the discrete gradient. For the latter case, we note that one obtains Tikhonov and TV denoising.

In the experiments, the noise-free images are degraded with various levels of AWGN where the noise variance σ^2 is specified to match some given input SNR. For denoising, we use FISTA [10] for 250 iterations without any stopping criteria. The multiplicative factors are optimized for all the estimators by using an oracle to obtain the highest-possible SNR. This optimization is done in a joint way for the γ parameter of the fractional Laplacian operator. The denoising results (output SNR in dB) are reported in Table I.

The results reported in Table I illustrate that the self-similarity assumption is well-suited for the particular images considered. For the clouds, it is coherent with the fact that the self-similar processes present a fractal-type statistical behavior. Moreover, the stem cells image is seemed to be appropriate for our model as corroborated by the results. For both images, it is observed that the performance of the self-similar models outperform TV denoising.

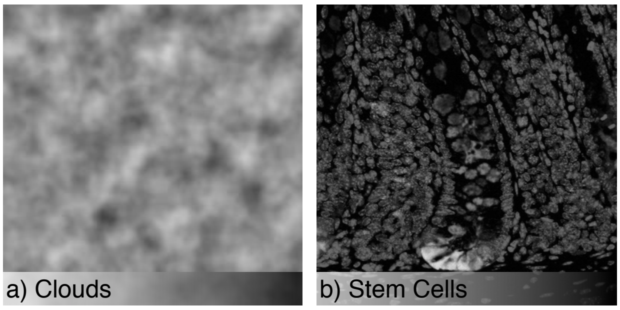

Fig. 3. Images used in the experiments.

TABLE I
DENOISING PERFORMANCE OF DIFFERENT MAP ESTIMATORS.

Input SNR (dB)	0	10	20	30
Estimator	Clouds			
Gaussian (discrete gradient)	20.44	24.68	29.93	35.31
Gaussian (fractional Laplacian)	**21.01**	**25.70**	**31.41**	36.45
Laplace (discrete gradient)	19.77	24.03	29.29	34.93
Laplace (fractional Laplacian)	20.22	25.29	31.16	**36.75**
Estimator	Stem cells			
Gaussian (discrete gradient)	11.16	15.85	22.13	30.40
Gaussian (fractional Laplacian)	**11.57**	**16.82**	**23.51**	**31.32**
Laplace (discrete gradient)	11.12	16.09	22.63	30.74
Laplace (fractional Laplacian)	11.20	16.70	**23.52**	**31.30**

V. Conclusion

The purpose of this work has been to drive MAP estimators that are suitable for reconstructing self-similar multidimensional signals from noisy samples. Our experiments showed that these estimators can outperform TV denoising for certain type of images.

VI. Acknowledgement

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

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] E. Bostan, U. Kamilov, and M. Unser, “Reconstruction of biomedical images and sparse stochastic modeling,” in *Proceedings of the 9^{th} IEEE International Symposium on Biomedical Imaging: From Nano to Macro (ISBI'12)*, Barcelona, Spain, May 2-5, 2012, pp. 880–883.

[3] E. Bostan, U. S. Kamilov, M. Nilchian, and M. Unser, “Sparse stochastic processes and discretization of linear inverse problems,” to appear in IEEE Transactions on Image Processing.

[4] Q. Sun and M. Unser, “Left-inverses of fractional Laplacian and sparse stochastic processes,” *Advances in Computational Mathematics*, vol. 36, no. 3, pp. 399–441, April 2012.

[5] B. B. Mandelbrot, *The Fractal Geometry of Nature*. W. H. Freeman, 1983.

[6] J. Huang and D. Mumford, “Statistics of natural images and models,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Collins, CO, 23-25 June 1999, pp. 637–663.

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

[8] K. Sato, *Lévy Processes and Infinitely Divisible Distributions.* Cambridge, 1994.

[9] M. Unser, P. D. Tafti, A. Amini, and H. Kirshner, “A unified formulation of Gaussian vs. sparse stochastic processes–Part II: Discrete-domain theory,” *arXiv:1108.6152v1*.

[10] A. Beck and M. Tebouelle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM*, vol. 2, no. 2, pp. 183–202, 2009.