Wavelet Statistics of Sparse and Self-Similar Images[*]

Julien Fageot[†], Emrah Bostan[‡], and Michael Unser[§]

Abstract. It is well documented that natural images are compressible in wavelet bases and tend to exhibit fractal properties. In this paper, we investigate statistical models that mimic these behaviors. We then use our models to make predictions on the statistics of the wavelet coefficients. Following an innovation modeling approach, we identify a general class of finite-variance self-similar sparse random processes. We first prove that spatially dilated versions of self-similar sparse processes are asymptotically Gaussian as the dilation factor increases. Based on this fundamental result, we show that the coarse-scale wavelet coefficients of these processes are also asymptotically Gaussian, provided the wavelet has enough vanishing moments. Moreover, we quantify the degree of Gaussianity by deriving the theoretical evolution of the kurtosis of the wavelet coefficients across scales. Finally, we apply our analysis to one- and two-dimensional signals, including natural images, and show that the wavelet coefficients tend to become Gaussian at coarse scales.

Key words. wavelet statistics, wide-sense self-similar processes, sparse processes, central-limit theorem, innovation model

AMS subject classifications. 60G18, 60G20, 60F05, 62M40

DOI. 10.1137/151003015

1. Introduction. The statistical analysis of natural images is a subject of great interest in image processing. The main motivation is that probabilistic models, generally combined with a Bayesian approach, are effectual in the derivation of imaging algorithms. From an observational point of view, the fundamental requirement for an acceptable model is its ability to replicate certain empirical properties. Among such observations, we shall put emphasis on *compressibility* and *scale invariance* [35, 36, 41, 46].

It is well-known that many natural images are compressible in some transform domain: by applying a suitable transform (e.g., wavelets), the energy of the signal gets concentrated in a small number of coefficients. In addition, the gray level intensities of images have finite variance while the corresponding histograms generally exhibit heavy tails. These observations have been validated practically for both first- and second-order marginals of the wavelet coefficients of natural images [9, 45].

Another remarkable characteristic of natural images is their scale invariance [13, 36]. Intuitively, it implies that images look roughly the same if viewed with different levels of magnification. Fractals are popular examples of scale-invariant objects since they can be well-described as the superposition of the same pattern repeated at different scales. The scale invariance

[†]STI, IMT, LIB, BM4139, EPFL, Lausanne 1006, Switzerland (julien.fageot@epfl.ch).
[‡]STI, IMT, LIB, BM4140, EPFL, Lausanne 1015, Switzerland (emrah.bostan@epfl.ch).
[§]Biomedical Imaging Group, EPFL, Lausanne CH-1015, Switzerland (michael.unser@epfl.ch).

property is also deeply related to the polynomial decay of the power spectrum of natural images [19, 25, 30]. It is thus not surprising that numerous models driven by these observations have been used in various problems in image reconstruction [1, 6, 8, 20, 22, 33].

1.1. From empirical observations to probabilistic models. The compressibility of natural images suggests that the use of Gaussian priors for deriving algorithms is not entirely adequate. Solutions beyond the Gaussian hypothesis were offered by wavelet methods [29] and by the compressed-sensing community [10, 15], which were initially deterministic. The concept of compressibility was also formalized in a probabilistic setting [3, 11, 24]. In Bayesian approaches, one traditionally models the marginal statistics of images—more generally, signals that are more compressible than Gaussian samples—with heavy-tailed distributions. Among the priors used, we mention generalized Gaussians [34], mixture of Gaussians, including Bernoulli–Gaussians [5, 51], and α-stable distributions [1, 42]. For us, the compressibility of images is modeled using infinitely divisible priors. Our motivation is that infinite divisible laws are in correspondence with random processes with independent increments [40, 43]. Hence, the family of infinitely divisible laws is the largest one we can use to model images as random processes [12, 36]. Non-Gaussian infinitely divisible laws are shown to have fatter tails than Gaussians [2].

The statistical scale invariance of natural images is captured by the concept of self-similar random processes. Starting from the fractional Brownian motion (fBm) [31], the class of self-similar processes has been considerably extended and applied to image processing [39]. The most popular self-similar stochastic models, however, are Gaussian, due to the convenient property that such processes are completely characterized by their second-order moments [32, 38]. Yet, there are also known examples of non-Gaussian self-similar models including symmetric α-stable models (SαS) [42, 44] and generalized Poisson processes [4, 25, 36].

1.2. Scope of the paper. Our goal in this paper is to introduce a unifying probabilistic framework that combines the hypotheses of sparsity (thus, non-Gaussianity) and self-similarity. We shall first establish a statistical model for the global image instead of its wavelet coefficients. Developing a model for the complete image is expectedly more demanding than the search for a local one. Other examples of global models include [12, 13, 23, 36]. The payoff is that it can provide us with deeper insights and predictions that can then be confronted with real data. We give three arguments in favor of our approach. First, the model contains the complete statistical information about the image; the same model can hence serve as the starting point for deriving different statistical methods. Second, the proposed model is inherently continuous: we define an image in the continuous domain and allow for any kind of geometric transformation without the artefacts of a discrete theory. Third, a complete model resolves the key aspects of first- or second-order statistics that can be missed by a local approach; for instance, in the context of a self-similar model, the probability distribution of the wavelet coefficients is expected to vary from one scale to the next unless the underlying model is stable [42].

In continuation of our previous works, we develop a global mechanism that we call the *innovation model* [53]. An image s is specified as the solution of the stochastic differential equation $\mathrm{L}s = w$, where L is a linear shift-invariant operator, typically a pseudodifferential operator [27, 28] called the whitening operator, and w is a white noise with finite variance but
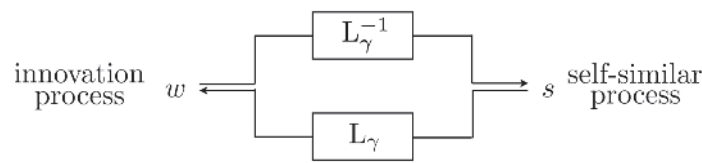
Figure 1. *Self-similar innovation model.*

possibly non-Gaussian. The statistics of any finite-dimensional marginal—especially the first- and second-order ones—are then deduced from the general theory. We reiterate that the main goal of our model is to be coherent with the observed compressibility and scale invariance of natural images. The efficiency of the innovation model in studying sparse signals is already well-understood [2, 54, 55] and lies beyond the scope of this paper. Therefore, we propose a continuous-domain innovation model that is particularly befitting self-similarity. We extend preliminary works that were restricted to Gaussian [48, 49] or compound-Poisson priors [52], or to specific whitening operators L [17, 47]. Our approach can be applied to any kind of signals, not only images, as we develop our theory in any number of dimensions $d \geq 1$. When $d \geq 2$, random processes are often called random fields [14, 16].

Our contributions are as follows:

(1) We specify a *self-similar* innovation model shown in Figure 1 and identify a particular class of linear operators L_γ that ensure the wide-sense self-similar properties of the process s (see section 3). The parameter $\gamma \geq 0$ measures the degree of self-similarity of s. The introduced class generalizes previously studied subclasses [17, 47, 48, 55].

(2) We provide a multiscale analysis of self-similar sparse processes by making the statistics of every rescaled version of the processes explicit and studying the limit behavior (see section 4). We show that any rescaled self-similar sparse process is asymptotically Gaussian.

(3) We deduce the complete statistical properties of the wavelet decomposition of self-similar sparse processes under the condition that the wavelet is admissible (see section 5). We show that the wavelet coefficients at coarse scales also become Gaussian. We further quantify the Gaussianity across scales by analyzing the cumulants of the process. In particular, we obtain the theoretical evolution of the kurtosis across scales.

(4) We provide experimental validations of our theoretical results on synthetic and natural images (see section 6). We show that synthetic data behave in accordance with our theoretical predictions. Moreover, the predicted Gaussian behavior of wavelet coefficients at coarse scales is qualitatively observed for different classes of natural images.

2. Preliminaries on the innovation model. We now introduce the mathematical notions required to define self-similar innovation models.

2.1. Generalized random processes. The innovation model is defined in the framework of generalized random processes [21], which is the extension of Schwartz' theory of generalized functions in a probabilistic setting. Thus, a random process s is not defined pointwise but by its duality products $\langle s, \varphi \rangle$ with test functions $\varphi \in \mathcal{S}(\mathbb{R}^d)$, the space of smooth and rapidly decreasing functions.

The space $\mathcal{S}(\mathbb{R}^d)$ is associated with its usual nuclear Fréchet topology [50]. Let $\mathcal{S}'(\mathbb{R}^d)$ be

the topological dual of $\mathcal{S}(\mathbb{R}^d)$, more commonly referred as the space of tempered generalized functions. We endow $\mathcal{S}'(\mathbb{R}^d)$ with the strong topology and denote by $\mathcal{B}(\mathcal{S}'(\mathbb{R}^d))$ the corresponding Borelian σ-field. Let $(\Omega, \mathcal{A}, \mathscr{P})$ be a probability space, $\mathcal{A}$ being a σ-field on Ω and $\mathscr{P}$ a probability measure on $\mathcal{A}$.

A *generalized random process* (over Ω) is a measurable map

$$(2.1) \qquad s : (\Omega, \mathcal{A}) \to (\mathcal{S}'(\mathbb{R}^d), \mathcal{B}(\mathcal{S}'(\mathbb{R}^d))).$$

As such, s is a random element of the space $\mathcal{S}'(\mathbb{R}^d)$. The *probability law* of s, denoted by $\mathscr{P}_s$, is the image measure of $\mathscr{P}$ by s; that is, the probability measure on $\mathcal{S}'(\mathbb{R}^d)$ defined by $\mathscr{P}_s(B) = \mathscr{P}(\{\omega \in \Omega, s(\omega) \in B\})$ for every Borelian set $B \in \mathcal{B}(\mathcal{S}'(\mathbb{R}^d))$. The *characteristic functional* of s is the infinite-dimensional Fourier transform of its probability law, defined as

$$(2.2) \qquad \widehat{\mathscr{P}_s}(\varphi) = \mathbb{E}[e^{i\langle s, \varphi \rangle}] = \int_{\mathcal{S}'(\mathbb{R}^d)} e^{i\langle u, \varphi \rangle} d\mathscr{P}_s(u) \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d).$$

The characteristic functional characterizes the law of s. The following fundamental result, known as the Minlos–Bochner theorem [21, section III-2.6, Theorem 5], fully describes the characteristic functionals of generalized random processes.

Theorem 2.1. *A functional $\widehat{\mathscr{P}} : \mathcal{S}(\mathbb{R}^d) \to \mathbb{C}$ is the characteristic functional of a generalized random process s if and only if $\widehat{\mathscr{P}}(0) = 1$ and $\widehat{\mathscr{P}}$ is continuous and positive-definite over $\mathcal{S}(\mathbb{R}^d)$.*

For more details on the definition of generalized random processes and the correspondence between s, its probability law $\mathscr{P}_s$ and its characteristic functional $\widehat{\mathscr{P}_s}$, we refer the reader to [26] and [21, Chapter III].

For fixed $\varphi_1, \ldots, \varphi_n \in \mathcal{S}(\mathbb{R}^d)$, the real random vector $\mathbf{X} = (\langle s, \varphi_1 \rangle, \ldots, \langle s, \varphi_n \rangle)$ is called a *finite-dimensional marginal* of s. Its characteristic function $\Phi_{\mathbf{X}}$ can be deduced from (2.2) by the relation

$$(2.3) \qquad \Phi_{\mathbf{X}}(\boldsymbol{\xi}) = \widehat{\mathscr{P}_s}(\xi_1 \varphi_1 + \cdots + \xi_n \varphi_n) \quad \forall \boldsymbol{\xi} = (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n.$$

A generalized random process s is said to have a *finite variance* if $\mathbb{E}[\langle s, \varphi \rangle^2] < \infty$ for any $\varphi \in \mathcal{S}(\mathbb{R}^d)$, which is the hypothesis that is made throughout this paper. A real random vector $\mathbf{X}$ is symmetric if its probability law satisfies $\mathscr{P}_{\mathbf{X}}(B) = \mathscr{P}_{\mathbf{X}}(-B)$ for every Borelian B. By extension, a generalized random process s is said to be *symmetric* if $(\langle s, \varphi_1 \rangle, \ldots, \langle s, \varphi_n \rangle)$ is symmetric for any $\varphi_1, \ldots, \varphi_n \in \mathcal{S}(\mathbb{R}^d)$.

Two processes s_1 and s_2 are *equal in law*, which we denote by $s_1 \overset{d}{=} s_2$ (d for distribution), if they have the same probability law, or equivalently the same characteristic functional. We deduce from (2.3) that this is equivalent with the equality in law of the finite-dimensional marginals of the two processes.

Definition 2.2. *Let s_n, $n \in \mathbb{N}$, and s be generalized random processes. We say that s_n converges in law to s, and denote it by*

$$s_n \xrightarrow[n \to \infty]{d} s,$$

if the underlying measures $\mathscr{P}_{s_n}$ are weakly converging to $\mathscr{P}_s$, which means that, for any $B \in \mathcal{B}(\mathcal{S}'(\mathbb{R}^d))$, $\mathscr{P}_{s_n}(B) \xrightarrow[n \to \infty]{} \mathscr{P}_s(B)$.

The space $\mathcal{S}(\mathbb{R}^d)$ being nuclear, the weak convergence of probability measures is equivalent to the pointwise convergence of the characteristic functionals [7, Theorem 4.5], meaning that

$$(2.4) \qquad \widehat{\mathscr{P}}_{s_n}(\varphi) \underset{n \to \infty}{\longrightarrow} \widehat{\mathscr{P}}_s(\varphi) \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d).$$

This result is the generalization of the Lévy continuity theorem to infinite dimensional real random variables (see [37] for more details).

2.2. Innovation processes. Continuous-domain white noises, also called innovation processes, cannot be defined pointwise. This is one of the main justifications for the introduction of the framework of generalized random processes. Innovation processes are in correspondence with infinite divisible random variables [53, Chapter IV]. A real random variable X is said to be infinitely divisible if, for every n, it can be decomposed as $X \overset{d}{=} X_1 + \cdots + X_n$, where the X_k are independently and identically distributed (i.i.d.) and $\overset{d}{=}$ denotes the equality in law (or in distribution). A function $f : \mathbb{R} \to \mathbb{C}$ is said to be a *Lévy exponent* if it is the log-characteristic function of a infinitely divisible random variable X, that is,

$$(2.5) \qquad \Phi_X(\xi) = \exp(f(\xi)) \quad \forall \xi \in \mathbb{R}.$$

If, moreover, X has a finite variance, we say by extension that f is a finite-variance Lévy exponent. The Lévy–Khintchine theorem [43, Theorem 8.1] ensures that we can decompose f as

$$(2.6) \qquad f(\xi) = i\mu\xi - \frac{\sigma^2}{2}\xi^2 + \int_{\mathbb{R}} (e^{i\xi t} - 1 - i\xi t 1_{|t| \le 1}) V(\mathrm{d}t) \quad \forall \xi \in \mathbb{R},$$

with $\mu \in \mathbb{R}$, $\sigma^2 \ge 0$, and V a *Lévy measure*; that is, a measure on $\mathbb{R}$ such that $\int_{\mathbb{R}} \min(1, t^2) V(\mathrm{d}t) < \infty$ and $V(\{0\}) = 0$. We call (μ, σ^2, V) the *Lévy triplet* associated with f. The infinite divisible random variable X with Lévy triplet (μ, σ^2, V) has a finite variance if and only if $m_2(V) = \int_{\mathbb{R}^2} t^2 V(\mathrm{d}t) < \infty$, in which case

$$(2.7) \qquad \mathrm{Var}(X) = \sigma^2 + m_2(V).$$

Furthermore, X is symmetric if and only if $\mu = 0$ and $V(B) = V(-B)$ for every Borelian B, in which case we say by extension that f and V are symmetric.

If f is a finite-variance Lévy exponent, then the functional

$$(2.8) \qquad \varphi \mapsto \exp\left(\int_{\mathbb{R}^d} f(\varphi(\mathbf{r})) \mathrm{d}\mathbf{r} \right)$$

was shown to be well-defined, continuous, and positive-definite over $\mathcal{S}(\mathbb{R}^d)$, taking value 1 at $\varphi = 0$ [17, Theorem 3]. Theorem 2.1 therefore ensures then that this functional is the characteristic functional of a generalized random process. This allows for the definition of finite-variance innovation processes on $\mathcal{S}'(\mathbb{R}^d)$.

Definition 2.3. A finite-variance innovation process w *on* $\mathcal{S}'(\mathbb{R}^d)$ *is a generalized random process with characteristic functional*

$$(2.9) \qquad \widehat{\mathscr{P}}_w(\varphi) = \exp\left(\int_{\mathbb{R}^d} f(\varphi(\mathbf{r})) \mathrm{d}\mathbf{r} \right) \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d),$$

Table 1
Examples of innovation processes with their associated Lévy exponent and pdf.

Innovation process	$f(\xi)$	$p_{\text{id}}(x)$
Gaussian	$-\frac{\sigma^2 \xi^2}{2}$	$\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-x^2/2\sigma^2}$
Laplace	$-\log(1 + b^2\xi^2)$	$\frac{1}{2b}\mathrm{e}^{-|x|/b}$
Student	no closed form	$\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}(1+x^2)^{-\frac{\nu+1}{2}}$

where f is a finite-variance Lévy exponent.

Since our goal in this paper is to develop a second-order theory, we shall only consider finite-variance innovation processes.

Innovation processes are stationary; that is, $w(\cdot - \mathbf{r}_0) \overset{d}{=} w$ for every $\mathbf{r}_0 \in \mathbb{R}^d$. It is also independent at every point, meaning that $\langle w, \varphi_1 \rangle$ and $\langle w, \varphi_2 \rangle$ are independent when φ_1 and φ_2 have disjoint supports. Consider an innovation $w \in \mathcal{S}'(\mathbb{R}^d)$ with Lévy triplet (μ, σ^2, V). The variance of $\langle w, \varphi \rangle$ is $(\sigma^2 + m_2(V))\|\varphi\|_2^2$ for every $\varphi \in \mathcal{S}(\mathbb{R}^d)$. By extension, we call the quantity $\sigma_w^2 = \sigma^2 + m_2(V)$ the *variance* of w. Moreover, w is symmetric as a generalized random process (see section 2.1) if and only if $\mu = 0$ and V is symmetric as a Lévy measure. We distinguish between two types of innovations w:

(i) *Gaussian innovations* (or Gaussian white noise) with variance σ^2 corresponding to the Lévy exponent $f(\xi) = -\frac{\sigma^2 \xi^2}{2}$. In this case, one has that $\widehat{\mathscr{P}_w}(\varphi) = \mathrm{e}^{-\frac{\sigma^2\|\varphi\|_2^2}{2}}$, where $\|\varphi\|_2 = \left(\int_{\mathbb{R}^d} \varphi(\mathbf{r})^2 \mathrm{d}\mathbf{r}\right)^{1/2}$. In particular, for every $\varphi \in \mathcal{S}(\mathbb{R}^d)$, the real random variable $\langle w, \varphi \rangle$ is Gaussian with variance $\sigma^2\|\varphi\|_2^2$.

(ii) The non-Gaussian innovations (or Lévy white noises) which are characterized by non-Gaussian infinitely divisible laws. Their tail is necessarily fatter than that of a Gaussian [2, Theorem 7]. These can therefore be used as a model for sparsity. For this reason, a non-Gaussian innovation is also called a *sparse innovation process* [53].

It can be convenient to characterize an innovation through its canonical distribution $p_{\text{id}}(x) = \mathcal{F}^{-1}\{\exp(f(\cdot))\}(x)$, which represents the probability density function, when it exists, of the infinitely divisible random variable $X = \langle w, \text{rect} \rangle$ (see section 3.3). Examples of innovation processes, including the ones that are used in section 6, are given in Table 1.

3. Self-similar innovation model. We first introduce some notation. Consider a linear operator L continuous from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{S}'(\mathbb{R}^d)$, for the topologies introduced in section 2.1. The adjoint L^* of L is the linear operator defined by

$$(3.1) \qquad \langle \mathrm{L}^*\varphi_1, \varphi_2 \rangle = \langle \mathrm{L}\varphi_2, \varphi_1 \rangle \quad \forall \varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}^d).$$

In the two duality products above, the first element is in $\mathcal{S}'(\mathbb{R}^d)$ and the second in $\mathcal{S}(\mathbb{R}^d)$, so the duality products are well-defined. The operator L^* is therefore also continuous from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{S}'(\mathbb{R}^d)$.

For $\mathbf{r}_0 \in \mathbb{R}^d$, the translation operator $\mathcal{T}_{\mathbf{r}_0}$ is $\mathcal{T}_{\mathbf{r}_0}\{\varphi\} = \varphi(\cdot - \mathbf{r}_0)$ with $\varphi \in \mathcal{S}(\mathbb{R}^d)$. For $a > 0$, the scaling operator S_a is $S_a\{\varphi\} = a^{-d/2}\varphi(\cdot/a)$ with $\varphi \in \mathcal{S}(\mathbb{R}^d)$. We have the relations $\mathcal{T}_{\mathbf{r}_0}^* = \mathcal{T}_{\mathbf{r}_0}^{-1} = \mathcal{T}_{-\mathbf{r}_0}$ and $S_a^* = S_a^{-1} = S_{a^{-1}}$. Translation and scaling operators are extended to $\mathcal{S}'(\mathbb{R}^d)$ by duality.

A linear operator L continuous from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{S}'(\mathbb{R}^d)$ is said to be *shift-invariant* if $\mathrm{L}\mathcal{T}_{\mathbf{r}_0} = \mathcal{T}_{\mathbf{r}_0}\mathrm{L}$ for all $\mathbf{r}_0 \in \mathbb{R}^d$, and *homogeneous of order γ* (or *γ-homogeneous*) with $\gamma \in \mathbb{R}$ if $\mathrm{L}S_a = a^{-\gamma}S_a\mathrm{L}$ for all $a > 0$.

3.1. Linear processes. In this section, we identify the class of processes s solutions of the general innovation model. We base our definition in the following result, which extends the domain of continuity of the characteristic functional of an innovation process.

Proposition 3.1. Let $w \in \mathcal{S}'(\mathbb{R}^d)$ be an innovation with Lévy triplet (μ, σ^2, V). Then, $\widehat{\mathscr{P}_w}$ is well-defined, continuous, and positive-definite in $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. If w is, moreover, symmetric, the same result holds in $L^2(\mathbb{R}^d)$.

The result is a corollary of [17, Proposition 4] for the case of finite-variance innovations. The positive definiteness comes from [17, Proposition 2].

Consider a linear shift-invariant operator L from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{S}'(\mathbb{R}^d)$ and an innovation w on $\mathcal{S}'(\mathbb{R}^d)$. We assume that there exists a linear operator L^{*-1} with the following properties:

(i) The operator L^{*-1} is a left-inverse of L^*, meaning that

$$(3.2) \qquad \mathrm{L}^{*-1}\mathrm{L}^*\{\varphi\} = \varphi \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d).$$

(ii) The operator L^{*-1} is continuous from $\mathcal{S}(\mathbb{R}^d)$ to $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. If w is symmetric, this second condition is relaxed and we only assume the continuity from $\mathcal{S}(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$.

Under condition (ii) and thanks to Proposition 3.1, the functional $\varphi \mapsto \widehat{\mathscr{P}_w}(\mathrm{L}^{*-1}\varphi)$ is well-defined, continuous, and positive-definite over $\mathcal{S}(\mathbb{R}^d)$, and takes value 1 for $\varphi = 0$. Then, according to Theorem 2.1, there exists a generalized random process s with characteristic functional

$$(3.3) \qquad \widehat{\mathscr{P}_s}(\varphi) = \exp\left(\int_{\mathbb{R}^d} f(\mathrm{L}^{*-1}\{\varphi\}(\mathbf{r}))\mathrm{d}\mathbf{r}\right).$$

We call s a *linear process* and summarize the above situation with the notation

$$(3.4) \qquad s = \mathrm{L}^{-1}w.$$

The crucial point for our analysis is that the "whitened" process Ls satisfies, for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$,

$$(3.5) \qquad \widehat{\mathscr{P}_{\mathrm{L}s}}(\varphi) = \mathbb{E}[\mathrm{e}^{\mathrm{i}\langle \mathrm{L}s, \varphi\rangle}] = \mathbb{E}[\mathrm{e}^{\mathrm{i}\langle s, \mathrm{L}^*\varphi\rangle}] = \widehat{\mathscr{P}_s}(\mathrm{L}^*\varphi) = \widehat{\mathscr{P}_w}(\mathrm{L}^{*-1}\mathrm{L}^*\varphi) = \widehat{\mathscr{P}_w}(\varphi),$$

which means that $\mathrm{L}s \overset{d}{=} w$.

The conditions under which we can define a left-inverse operator L^{*-1} such that (ii) is fulfilled is not the subject of this paper and was investigated in [17, 47, 53]. However, some examples will be detailed below for the case of self-similar processes.

3.2. Self-similar sparse processes. For the definition of the self-similar innovation model (see Figure 1), we restrict ourselves to the case of operators $\mathrm{L} = \mathrm{L}_\gamma$ that are shift-invariant and γ-homogeneous. The geometrical invariances of L are related to the statistical invariances of s. Thus, this leads us to the definition of sparse processes with self-similar properties.

Definition 3.2. A generalized random process s is said to be wide-sense self-similar of order $\gamma \geq 0$ *if the processes $a^\gamma S_a s$ and s have the same second-order statistics for every $a > 0$; that*

is, if for every $a > 0$,

$$(3.6) \qquad \mathbb{E}[\langle a^\gamma S_a s, \varphi_1 \rangle \langle a^\gamma S_a s, \varphi_2 \rangle] = \mathbb{E}[\langle s, \varphi_1 \rangle \langle s, \varphi_2 \rangle] \quad \forall \varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}^d).$$

The notion of self-similarity for generalized random processes was introduced in [14] in the Gaussian case. See also [53, section 7.5] for more details. We give sufficient conditions under which a linear process is wide-sense self-similar.

Proposition 3.3. Let $s = \mathrm{L}^{-1}w$ be a linear process with characteristic functional $\widehat{\mathscr{P}}_s(\varphi) = \widehat{\mathscr{P}}_w(\mathrm{L}^{*-1}\varphi)$. We assume that $\mathrm{L} = \mathrm{L}_\gamma$ is γ-homogeneous and $\mathrm{L}^{*-1} = \mathrm{L}_\gamma^{*-1}$ is $(-\gamma)$-homogeneous for some fixed $\gamma \geq 0$. Then the linear process $s = \mathrm{L}_\gamma^{-1}w$ is wide-sense self-similar.

Remark that, since L_γ^{*-1} is only a left-inverse, we cannot in general deduce the $(-\gamma)$-homogeneity of L_γ^{*-1} from the γ-homogeneity of L_γ.

Proof. For two functions $\varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}^d)$, we have

$$(3.7) \qquad \mathbb{E}[\langle s, \varphi_1 \rangle \langle s, \varphi_2 \rangle] = \mathbb{E}[\langle w, \mathrm{L}_\gamma^{*-1}\varphi_1 \rangle \langle w, \mathrm{L}_\gamma^{*-1}\varphi_2 \rangle] = \sigma_w^2 \langle \mathrm{L}_\gamma^{*-1}\varphi_1, \mathrm{L}_\gamma^{*-1}\varphi_2 \rangle$$

with σ_w^2 the variance of w. Therefore, for $a > 0$ and $\varphi_1, \varphi_2 \in \mathcal{S}(\mathbb{R}^d)$, we get

$$
\begin{aligned}
\mathbb{E}[\langle a^\gamma S_a s, \varphi_1 \rangle \langle a^\gamma S_a s, \varphi_2 \rangle] &= a^{2\gamma}\mathbb{E}[\langle s, S_{a^{-1}}\varphi_1 \rangle \langle s, S_{a^{-1}}\varphi_2 \rangle] \\
&= \sigma_w^2 a^{2\gamma} \langle \mathrm{L}_\gamma^{*-1}\{S_{a^{-1}}\varphi_1\}, \mathrm{L}_\gamma^{*-1}\{S_{a^{-1}}\varphi_2\} \rangle \\
&\overset{(i)}{=} \sigma_w^2 \langle S_{a^{-1}}\mathrm{L}_\gamma^{*-1}\varphi_1, S_{a^{-1}}\mathrm{L}_\gamma^{*-1}\varphi_2 \rangle \\
&\overset{(ii)}{=} \sigma_w^2 \langle \mathrm{L}_\gamma^{*-1}\varphi_1, \mathrm{L}_\gamma^{*-1}\varphi_2 \rangle \\
&= \mathbb{E}[\langle s, \varphi_1 \rangle \langle s, \varphi_2 \rangle],
\end{aligned}
$$

where (i) comes from the $(-\gamma)$-homogeneity of L_γ^{*-1} and (ii) from the fact that $S_{a^{-1}} = (S_a)^*$. In effect, we have proved (3.6) which means that s is wide-sense self-similar. ∎

Proposition 3.3 motivates the following definition.

Definition 3.4. Consider a linear process $s = \mathrm{L}^{-1}w$ with characteristic functional $\widehat{\mathscr{P}}_s(\varphi) = \widehat{\mathscr{P}}_w(\mathrm{L}^{*-1}\varphi)$. If, moreover, the operator $\mathrm{L} = \mathrm{L}_\gamma$ is γ-homogeneous and $\mathrm{L}^{*-1} = \mathrm{L}_\gamma^{*-1}$ is $(-\gamma)$-homogeneous for some $\gamma \geq 0$, we say that

(i) *$s = \mathrm{L}_\gamma^{-1}w$ is a self-similar Gaussian process of order γ if w is a Gaussian innovation;* or

(ii) *$s = \mathrm{L}_\gamma^{-1}w$ is a self-similar sparse process of order γ if w is a non-Gaussian—i.e., sparse—innovation.*

The good news is that the operators L_γ that are used in practice admit a left-inverse that satisfies the conditions of Definition 3.4, as illustrated in the following examples.

Example 1. Let $\gamma > 0$. The fractional Laplacian operator $(-\Delta)^{\gamma/2}$ is a pseudodifferential operator with symbol $\|\cdot\|^\gamma$. Hence, it is defined for $\varphi \in \mathcal{S}(\mathbb{R}^d)$ by

$$(3.8) \qquad (-\Delta)^{\gamma/2}\{\varphi\} = \mathcal{F}^{-1}\left\{\|\cdot\|^\gamma \mathcal{F}\{\varphi\}\right\},$$

where $\mathcal{F}$ ($\mathcal{F}^{-1}$, respectively) denotes the Fourier transform (the inverse Fourier transform, respectively). The fractional Laplacian is shift-invariant, γ-homogeneous, and self-adjoint. Its inverse operator is the Riesz potential I_γ, defined for $0 < \gamma < d$ by the relation

$$(3.9) \qquad \mathrm{I}_\gamma\{\varphi\} = \mathcal{F}^{-1}\left\{\|\cdot\|^{-\gamma}\mathcal{F}\{\varphi\}\right\} \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d)$$

and extended for noninteger $\gamma > d$ in [47]. However, the Riesz potential can be unstable, with the consequence that the functional $\varphi \mapsto \widehat{\mathscr{P}_w}(\mathrm{I}_\gamma \varphi)$ will generally not be well-defined on $\mathcal{S}(\mathbb{R}^d)$. The study of corrected versions of I_γ that are stable in given L^p-spaces ($1 \le p \le \infty$) is done in [47, Theorem 1.2]. More precisely, it is shown that, for fixed $p \ge 1$ and noninteger $\gamma > 0$ such that $(\gamma - d - \frac{d}{p}) \notin \mathbb{N}$, there exists a unique left inverse operator of $(-\Delta)^{\gamma/2}$, denoted by T_γ, that is linear, $(-\gamma)$-homogeneous, and continuous from $\mathcal{S}(\mathbb{R}^d)$ to $L^p(\mathbb{R}^d)$. This allows us to define self-similar sparse processes $s = ((-\Delta)^{\gamma/2})^{-1}w$ of order γ; see [17, section 4.1]. An additional property is that the fractional Laplacian is isotropic, so is the self-similar sparse process s.

Example 2. The partial-derivative operator D_{x_1} (or, more generally, D_{x_j} with $1 \le j \le d$) is linear, shift-invariant, and 1-homogeneous. Its adjoint is $(-\mathrm{D}_{x_1})$. The natural inverse operator of $(-\mathrm{D}_{x_1})$ is

$$(3.10) \qquad \mathrm{I}_{x_1}\{\varphi\}(\mathbf{r}) = \int_{r_1}^{+\infty} \varphi(\tau, r_2, \ldots, r_d) \mathrm{d}\tau$$

for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and $\mathbf{r} = (r_1, \ldots, r_d) \in \mathbb{R}^d$. Again, it is (-1)-homogeneous but unstable. As shown in [17, section 4.2], the operator T_{x_1} defined for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$ and $\mathbf{r} = (r_1, \ldots, r_d) \in \mathbb{R}^d$ by

$$(3.11) \qquad \mathrm{T}_{x_1}\{\varphi\}(\mathbf{r}) = \int_{\mathbb{R}} (1_{\tau \ge r_1} - 1_{r_1 \ge 0}) \varphi(\tau, r_2, \ldots, r_d) \mathrm{d}\tau$$

is a left inverse operator of $(-\mathrm{D}_{x_1})$ that is (-1)-homogeneous and continuous from $\mathcal{S}(\mathbb{R}^d)$ to $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. We can therefore define the self-similar sparse process $s = \mathrm{D}_{x_1}^{-1}w$ for any innovation w. Using the same principle, we can more generally define self-similar sparse processes $s = (\mathrm{D}^{\mathbf{n}})^{-1}w$ of order $|\mathbf{n}|$ for any innovation w, where $\mathrm{D}^{\mathbf{n}} = \mathrm{D}_{x_1}^{n_1} \cdots \mathrm{D}_{x_d}^{n_d}$ and $|\mathbf{n}| = n_1 + \cdots + n_d$. The operator $\mathrm{D}^{\mathbf{n}}$ is not isotropic for $d \ge 2$, and allows for the construction of nonisotropic self-similar sparse processes.

3.3. Extended domain of linear processes. One can always observe an innovation process w or a self-similar sparse process s through windows $\varphi \in \mathcal{S}(\mathbb{R}^d)$. Our goal here is to extend the definition of the duality products $\langle w, \varphi \rangle$ and $\langle s, \varphi \rangle$ to analysis functions φ that are not necessarily smooth neither rapidly decreasing.

We know with Proposition 3.1 that $\widehat{\mathscr{P}_w}$ is continuous over $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Thus, for φ a fixed function in $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, the function $\xi \mapsto \widehat{\mathscr{P}_w}(\xi\varphi)$ is continuous, positive-definite in $\mathbb{R}$, and takes values 1 at $\xi = 0$. From the Bochner theorem, it is therefore the characteristic function of a real random variable denoted by $\langle w, \varphi \rangle$.

In particular, the function $\mathrm{rect} = 1_{[0,1]^d}$ is in $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, so the real random variable $X = \langle w, \mathrm{rect} \rangle$ is always well-defined. Note that its characteristic function is

$$(3.12) \qquad \widehat{\mathscr{P}_X}(\xi) = \widehat{\mathscr{P}_w}(\xi \mathrm{rect}) = \exp\left(\int_{\mathbb{R}^d} f(\xi \mathrm{rect}(\boldsymbol{r})) \mathrm{d}\boldsymbol{r} \right) = \mathrm{e}^{f(\xi)} \quad \forall \xi \in \mathbb{R},$$

where f is the Lévy exponent of w. We can now extend the definition domain of linear processes from the one of innovation processes.

Definition 3.5. *Let $s = \mathrm{L}^{-1}w$ be a linear process. The function ψ is said to be* admissible *(relatively to s) if there exists $\phi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ with $\psi = \mathrm{L}^*\phi$. In this case, we can define*

the real random variable $\langle s, \psi \rangle$ by the relation

$$\langle s, \psi \rangle = \langle s, \mathrm{L}^* \phi \rangle = \langle \mathrm{L}s, \phi \rangle = \langle w, \phi \rangle. \tag{3.13}$$

More generally, we can define the real random vector $(\langle s, \psi_k \rangle)_{1 \le k \le n}$ using the same approach, assuming that the ψ_k are admissible. Note that the extension of the definition domain of w and s is done in a weak sense, meaning in law. For our purpose, this is sufficient because our main results concern convergence in law of generalized processes or of real random vectors. Extended definition domains allow, moreover, for the multiscale analysis of sparse processes with nonsmooth wavelets (see section 5).

4. Statistical evolution of sparse processes across scales. From now on, we assume that $s = \mathrm{L}_\gamma^{-1} w$ is a self-similar sparse process. Since s is a random element of $\mathcal{S}'(\mathbb{R}^d)$, we can consider its scaling transformation $S_{a^{-1}} s$ by the relation, for $\varphi \in \mathcal{S}(\mathbb{R}^d)$,

$$\langle S_{a^{-1}} s, \varphi \rangle = \langle s, S_{a^{-1}}^* \varphi \rangle = \langle s, S_a \varphi \rangle. \tag{4.1}$$

In this section, we study the evolution of $S_{a^{-1}} s$ with $a > 0$.

4.1. Characteristic functionals across scales. We first characterize the process $S_{a^{-1}} s$ by using characteristic functional.

Proposition 4.1. Let $s = \mathrm{L}_\gamma^{-1} w$ be a self-similar sparse process of order $\gamma \ge 0$. Then the characteristic functional of the process $S_{a^{-1}} s$ is

$$\widehat{\mathscr{P}}_{S_{a^{-1}} s}(\varphi) = \exp\left(a^d \int_{\mathbb{R}^d} f(a^{\gamma - \frac{d}{2}} \mathrm{L}_\gamma^{*-1} \{\varphi\}(\mathbf{r})) \mathrm{d}\mathbf{r} \right). \tag{4.2}$$

Proof. Using (4.1), we have $\widehat{\mathscr{P}}_{S_{a^{-1}} s}(\varphi) = \widehat{\mathscr{P}}_s(S_a \varphi)$. Moreover, because L_γ^{*-1} is $(-\gamma)$-homogeneous, we deduce that

$$\mathrm{L}_\gamma^{*-1} S_a \varphi = a^\gamma S_a \mathrm{L}_\gamma^{*-1} \varphi = a^{\gamma - \frac{d}{2}} \{\mathrm{L}_\gamma^{*-1} \varphi\}\left(\frac{\cdot}{a}\right). \tag{4.3}$$

Replacing $\mathrm{L}^{*-1} \varphi$ by $a^{\gamma - \frac{d}{2}} \{\mathrm{L}_\gamma^{*-1} \varphi\}(\frac{\cdot}{a})$ in (3.3), we obtain (4.2) after the change of variable $\mathbf{r} \leftarrow \frac{\mathbf{r}}{a}$. ∎

4.2. Coarse-scale behavior of sparse processes. We shall now show that the rescaled versions $S_{a^{-1}} s$ of a self-similar sparse process s converge to a Gaussian process when $a \to \infty$, up to a renormalization. The result is essentially a generalization of the central-limit theorem for real random vectors to the case of self-similar processes.

4.2.1. Self-similar Gaussian processes. The Gaussian innovation process of variance σ^2, denoted by w_{σ^2}, is the generalized random process with characteristic functional $\widehat{\mathscr{P}}_{w_{\sigma^2}}(\varphi) = \exp(-\frac{\sigma^2}{2} \|\varphi\|_2^2)$. Its Lévy exponent is $f(\xi) = -\frac{\sigma^2 \xi^2}{2}$ and its Lévy triplet is $(0, \sigma^2, 0)$. When $(\mathrm{L}, w_{\sigma^2})$ is compatible, we denote $s_{\mathrm{L}, \sigma^2} = \mathrm{L}^{-1} w_{\sigma^2}$. Recall that when $(\mathrm{L}_\gamma, w_{\sigma^2})$ is γ-compatible, $s_{\mathrm{L}_\gamma, \sigma^2}$ is called a self-similar Gaussian process.

4.2.2. Limit theorem for self-similar sparse processes. The process $S_{a^{-1}}s$ is a rescaled version of s. Increasing a is equivalent to observing s against dilated analysis functions. We obtain the limit behavior for $a \to +\infty$ in the sense of the convergence in law (see Definition 2.2), which is the main result of this paper.

Theorem 4.2. Let $s = L_\gamma^{-1}w$ be a self-similar sparse process of order $\gamma \geq 0$. We define the quantities

$$(4.4) \qquad \sigma_w^2 = \sigma^2 + m_2(V) = \sigma^2 + \int_{\mathbb{R}} t^2 V(\mathrm{d}t),$$

$$(4.5) \qquad \mu_w = \mu + \mu_V = \mu + \int_{|t|>1} t V(\mathrm{d}t).$$

We distinguish two cases.
- *If w is symmetric, then*

$$(4.6) \qquad a^{-\gamma} S_{a^{-1}} s \xrightarrow[a \to +\infty]{d} s_{L_\gamma, \sigma_w^2}.$$

- *If w is not symmetric, then $L_\gamma^{*-1}(\mathcal{S}(\mathbb{R}^d)) \subset L^1(\mathbb{R}^d)$ and we can define $v \in \mathcal{S}'(\mathbb{R}^d)$ by $\langle v, \varphi \rangle = \langle 1, L_\gamma^{*-1}\varphi \rangle$ for every $\varphi \in \mathcal{S}(\mathbb{R}^d)$. Then,*

$$(4.7) \qquad a^{-\gamma} S_{a^{-1}} s - \mu_w a^{d/2} v \xrightarrow[a \to +\infty]{d} s_{L_\gamma, \sigma_w^2}.$$

The proof is provided in Appendix A. The key ingredient is the equivalence between the convergence in law of generalized random processes and the pointwise convergence of the corresponding characteristic functionals, as expressed in (2.4).

The distinction between the nonsymmetric and the symmetric cases can be shortly explained. For a nonsymmetric innovation, the quantity $\mathbb{E}[\langle w, \varphi \rangle] = \mu_w \int_{\mathbb{R}^d} \varphi$ is in general nonzero and we have to compensate this drift, which is done in Theorem 4.2 by introducing the function v. On the other hand, for a symmetric innovation, $\mu_w = 0$ and it is therefore useless to introduce v.

The important consequence of Theorem 4.2 is that, under appropriate normalization, the statistics of wide-sense self-similar sparse processes are becoming Gaussian at coarse scales, despite the property that these processes are initially sparse and non-Gaussian.

5. Wavelet analysis of self-similar sparse processes. In this section, we use our previous results to deduce the statistics of the wavelet coefficients of self-similar sparse processes. We discretize the scale parameter a according to a geometric progression to fit with the traditional wavelet theory [29]. We denote by $\rho > 0$ the *dilation factor* and consider scale parameter of the form ρ^j with $j \in \mathbb{Z}$. (In practice, we chose $\rho = 2$ or $\sqrt{2}$.) Given a wavelet function ψ on $\mathbb{R}^d$, we define for $j \in \mathbb{Z}$ and $\mathbf{r}_0 \in \mathbb{R}^d$,

$$(5.1) \qquad \psi_{j,\mathbf{r}_0} = \frac{1}{\rho^{jd/2}} \psi\left(\frac{\cdot}{\rho^j} - \mathbf{r}_0\right) = S_{\rho^j} \mathcal{T}_{\mathbf{r}_0} \psi.$$

5.1. Self-similar sparse processes and admissible wavelets. Let $s = \mathrm{L}_\gamma^{-1} w$ be a self-similar sparse process of order $\gamma \geq 0$. For $\psi = \mathrm{L}^* \phi$ an admissible wavelet (see Definition 3.5), we can define the real random variables $\langle s, \psi_{j,\mathbf{r}_0} \rangle$ for every $(j, \mathbf{r}_0) \in \mathbb{Z} \times \mathbb{R}^d$, called the wavelet coefficients of s. Since

$$(5.2) \qquad \langle s, \psi_{j,\mathbf{r}_0} \rangle = \langle s, S_{\rho^j} \mathcal{T}_{\mathbf{r}_0} \psi \rangle = \langle S_{\rho^{-j}} s, \mathcal{T}_{\mathbf{r}_0} \psi \rangle,$$

the study of the processes $S_{a^{-1}} s$ can be translated into results on the wavelet coefficients of s at different scales j by setting $a = \rho^j$.

We say that a function ψ has vanishing moments until order $k \in \mathbb{N}$ if for every $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$ such that $|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_d \leq k$, we have $\int_{\mathbb{R}^d} \mathbf{r}^{\boldsymbol{\alpha}} \psi(\mathbf{r}) \mathrm{d}\mathbf{r} = 0$. For a self-similar sparse process s, the condition on the wavelet ψ to be admissible is closely related to the existence of vanishing moments. We have indeed the following proposition.

Proposition 5.1. Let $\gamma \geq 0$. Consider L_γ, a linear, shift-invariant, and γ-homogeneous operator, continuous from $\mathcal{S}(\mathbb{R}^d)$ to $\mathcal{S}'(\mathbb{R}^d)$, and ψ a wavelet such that $\psi = \mathrm{L}_\gamma^* \phi$ with $\phi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. We assume that the Fourier multiplier $\widehat{L_\gamma}$ of L_γ and the Fourier transform $\widehat{\phi}$ of ϕ are $(\lceil \gamma \rceil - 1)$ times differentiable, where $\lceil \gamma \rceil$ is the smallest integer larger than or equal to γ. Then, ψ has vanishing moments until order $(\lceil \gamma \rceil - 1)$.

Proof. The operator L_γ^* is shift-invariant and γ-homogeneous; hence, its Fourier multiplier $\widehat{L_\gamma^*}$ exists and satisfies for all $\boldsymbol{\omega} \in \mathbb{R}^d$ and $a > 0$,

$$(5.3) \qquad a^\gamma \widehat{L_\gamma^*}(\boldsymbol{\omega}) = \widehat{L_\gamma^*}(a\boldsymbol{\omega}).$$

Moreover, for φ a function with a sufficiently smooth Fourier transform, we have the relation

$$(5.4) \qquad \int_{\mathbb{R}^d} \mathbf{r}^{\boldsymbol{\alpha}} \varphi(\mathbf{r}) \mathrm{d}\mathbf{r} = (-\mathrm{i})^{|\boldsymbol{\alpha}|} \mathrm{D}^{\boldsymbol{\alpha}} \{\widehat{\varphi}\}(\mathbf{0}).$$

Because $\widehat{\psi}(\boldsymbol{\omega}) = \widehat{L_\gamma^*}(\boldsymbol{\omega}) \widehat{\phi}(\boldsymbol{\omega})$ and taking advantage of the Leibniz rule, showing that $\mathrm{D}^{\boldsymbol{\alpha}} \{\widehat{L_\gamma^*}\}(\mathbf{0}) = 0$ for $|\boldsymbol{\alpha}| \leq (\lceil \gamma \rceil - 1)$ is enough to deduce that ψ has vanishing moments until order $(\lceil \gamma \rceil - 1)$. Applying the operator $\mathrm{D}^{\boldsymbol{\alpha}}$ to (5.3), which is possible because $\widehat{L_\gamma^*}(\boldsymbol{\omega}) = \widehat{L_\gamma}(-\boldsymbol{\omega})$ and $\widehat{\phi}$ are regular enough by assumption, and setting $\boldsymbol{\omega} = 0$, we have $a^{\gamma - |\boldsymbol{\alpha}|} \mathrm{D}^{\boldsymbol{\alpha}} \{\widehat{L_\gamma^*}\}(\mathbf{0}) = \mathrm{D}^{\boldsymbol{\alpha}} \{\widehat{L_\gamma^*}\}(\mathbf{0})$ for all $a > 0$. Since $\gamma > |\boldsymbol{\alpha}|$, with $a \to 0$, we get $\mathrm{D}^{\boldsymbol{\alpha}} \{\widehat{L_\gamma^*}\}(\mathbf{0}) = 0$ and the result follows. ∎

The assumptions of Proposition 5.1 are satisfied by a vast majority of shift-invariant and homogeneous operators, including all the examples in section 3. In particular, the Fourier multipliers have the adequate regularity. Under some conditions of regularity, Proposition 5.1 says that the vanishing moments are a necessary condition to be admissible. Moreover, a large collection of wavelets with enough vanishing moments are admissible in the sense of Definition 3.5. For more details about the properties and the construction of admissible wavelets, see [53, section 6.5].

5.2. Statistics of wavelet coefficients. First, we focus on the wavelet coefficients of the innovation process w. Let ψ be in $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Then, because $\langle w, \psi_{j,\mathbf{r}_0} \rangle = \langle S_{\rho^{-j}} w, \mathcal{T}_{\mathbf{r}_0} \psi \rangle = \langle S_{\rho^{-j}} w, \psi \rangle$, the last equality being in law, the characteristic function $\Phi_{j,\mathbf{r}_0}$ of $\langle w, \psi_{j,\mathbf{r}_0} \rangle$ satisfies,

for all $\xi \in \mathbb{R}$,

$$(5.5) \qquad \Phi_{j,\mathbf{r}_0}(\xi) = \exp\left(\rho^{jd}\int_{\mathbb{R}^d} f(\rho^{-jd/2}\xi\varphi(\mathbf{r}))\mathrm{d}\mathbf{r}\right).$$

For a fixed j, $\Phi_{j,\mathbf{r}_0}$ does not depend on $\mathbf{r}_0$, which implies that the wavelet coefficients of an innovation process at a given scale are identically distributed. This result can be extended to self-similar sparse processes.

Proposition 5.2 (wavelet coefficients of sparse processes). *Let $s = \mathrm{L}_\gamma^{-1}w$ be a self-similar sparse process of order $\gamma \geq 0$ with admissible wavelet $\psi = \mathrm{L}_\gamma^*\phi$. The characteristic function of the wavelet coefficient $X = \langle s, \psi_{j,\mathbf{r}_0}\rangle$ is*

$$(5.6) \qquad \Phi_X(\xi) = \exp\left(\rho^{jd}\int_{\mathbb{R}^d} f(\rho^{j(\gamma-d/2)}\xi\phi(\mathbf{r}))\mathrm{d}\mathbf{r}\right) \qquad \forall \xi \in \mathbb{R}.$$

Proof. This result is a direct consequence of Proposition 4.1. From (2.3), the characteristic function of X is given by $\Phi_X(\xi) = \widehat{\mathscr{P}_s}(\xi\psi_{j,\mathbf{r}_0})$. We apply (4.2) with $a = \rho^j$ and $\varphi = \xi\psi(\cdot - \mathbf{r}_0)$ to obtain

$$(5.7) \qquad \Phi_X(\xi) = \exp\left(\rho^{jd}\int_{\mathbb{R}^d} f(\rho^{j(\gamma-d/2)}\xi\mathrm{L}_\gamma^{*-1}\{\psi(\cdot - \mathbf{r}_0)\}(\mathbf{r}))\mathrm{d}\mathbf{r}\right).$$

Moreover, because L_γ^* is stationary, $\psi(\cdot - \mathbf{r}_0) = \{\mathrm{L}_\gamma^*\phi\}(\cdot - \mathbf{r}_0) = \mathrm{L}_\gamma^*\{\phi(\cdot - \mathbf{r}_0)\}$. Hence,

$$(5.8) \qquad \mathrm{L}_\gamma^{*-1}\{\psi(\cdot - \mathbf{r}_0)\} = \mathrm{L}_\gamma^{*-1}\{\mathrm{L}_\gamma^*\{\phi(\cdot - \mathbf{r}_0)\}\} = \phi(\cdot - \mathbf{r}_0).$$

We substitute (5.8) in (5.7) and make the change of variable $\mathbf{r} \leftarrow \mathbf{r} - \mathbf{r}_0$ to obtain (5.6). ∎

From (5.6), we deduce that the wavelet coefficients at the scale j are identically distributed.

5.3. Wavelet coefficients at coarse scales. The convergence result in Theorem 4.2 implies that the wavelet coefficients of a self-similar sparse process are becoming Gaussian when the scale j goes to infinity.

Proposition 5.3 (central-limit theorem for wavelet coefficients). *Let $s = \mathrm{L}_\gamma^{-1}w$ be a symmetric—that is, with w a symmetric innovation—self-similar sparse process of order $\gamma \geq 0$ with admissible wavelet $\psi = \mathrm{L}_\gamma^*\phi$. Then, for every $(j_1,\ldots,j_n) \in \mathbb{Z}^n$ and $(\mathbf{r}_{0,1},\ldots,\mathbf{r}_{0,n}) \in (\mathbb{R}^d)^n$,*

$$(5.9) \qquad \frac{(\langle s, \psi_{j_1+j,\mathbf{r}_{0,1}}\rangle,\ldots,\langle s,\psi_{j_n+j,\mathbf{r}_{0,n}}\rangle)}{\rho^{j\gamma}} \xrightarrow[j\to+\infty]{d} (\langle s_{\mathrm{L}_\gamma,\sigma_w^2}, \psi_{j_1,\mathbf{r}_{0,1}}\rangle,\ldots,\langle s_{\mathrm{L}_\gamma,\sigma_w^2},\psi_{j_n,\mathbf{r}_{0,n}}\rangle),$$

with $\sigma_w^2 = \sigma^2 + m_2(V)$ and $s_{\mathrm{L}_\gamma,\sigma_w^2}$ the self-similar Gaussian process defined in section 4.2.1.

In particular, when $j \to +\infty$, $(\langle s, \psi_{j,\mathbf{r}_{0,1}}\rangle,\ldots,\langle s,\psi_{j,\mathbf{r}_{0,n}}\rangle)$ converges to a Gaussian vector, under the appropriate normalization.

Proof. We have already seen that $\langle s, \psi_{j_0+j,\mathbf{r}_0}\rangle = \langle S_{\rho^{-j}}s, S_{\rho^{j_0}}\mathcal{T}_{\mathbf{r}_0}\psi\rangle = \langle S_{\rho^{-j}}s, \psi_{j_0,\mathbf{r}_0}\rangle$. Moreover, from Theorem 4.2 we have $\frac{S_{\rho^{-j}}s}{\rho^{j\gamma}} \xrightarrow[j\to+\infty]{d} s_{\mathrm{L}_\gamma,\sigma_w^2}$. This implies the convergence of the finite-dimensional marginal

$$(5.10) \qquad \left(\left\langle \frac{S_{\rho^{-j}}s}{\rho^{j\gamma}}, \psi_{j_k,\mathbf{r}_{0,k}}\right\rangle\right)_{1\leq k\leq n} \xrightarrow[j\to+\infty]{d} \left(\langle s_{\mathrm{L}_\gamma,\sigma_w^2}, \psi_{j_k,\mathbf{r}_{0,k}}\rangle\right)_{1\leq k\leq n},$$

which is a rewriting of (5.9). ∎

5.4. Evolution of the cumulants. Let X be a real random variable with characteristic function Φ_X. When $\mathbb{E}[|X|^k] < \infty$ for all $k \in \mathbb{N}$, we have the series expansion $\log \Phi_X(\xi) = \sum_{n \in \mathbb{N}} \frac{\kappa_n(X)}{n!}(\mathrm{i}\xi)^n$. The coefficients $\kappa_n(X)$ are called the *cumulants* of X. If we only assume that $\mathbb{E}[|X|^N] < \infty$ for a fixed $N \geq 1$, then $\log \Phi_X$ admits the truncated Taylor expansion at 0

$$(5.11) \qquad \log \Phi_X(\xi) = \sum_{n=0}^{N} \frac{\kappa_n(X)}{n!}(\mathrm{i}\xi)^n + o(\xi^N),$$

and the cumulants up to order N are still well-defined. Especially, provided that they exist, we have $\kappa_1(X) = \mathbb{E}[X]$ and $\kappa_2(X) = \mathrm{Var}(X)$. Our goal in this section is to give conditions such that the cumulants of a sparse process exist, and to see how they evolve with scale.

Definition 5.4. A generalized random process s has finite Nth moments *if for all $\varphi \in \mathcal{S}(\mathbb{R}^d)$,* $\mathbb{E}[|\langle s, \varphi \rangle|^N] < \infty$.

We recall that for every innovation process w, the real random variable $\langle w, \mathrm{rect} \rangle$ is well-defined with characteristic function $\exp(f(\xi))$ where f is the Lévy exponent of w (see section 3.3).

Proposition 5.5. Let $N \geq 2$. The self-similar sparse process $s = \mathrm{L}_\gamma^{-1} w$ has finite Nth moments if

$$(5.12) \qquad \mathbb{E}[|\langle w, \mathrm{rect} \rangle|^N] < \infty.$$

Moreover, for any admissible function $\psi = \mathrm{L}_\gamma^* \phi$ with $\phi \in L^1(\mathbb{R}^d) \bigcap L^2(\mathbb{R}^d) \bigcap L^N(\mathbb{R}^d)$, we have if (5.12) holds, then $\mathbb{E}[|\langle s, \psi \rangle|^N] < \infty$.

Proof. The proof is based on the tools developed in [53, Chapter 9]. An infinitely divisible random variable X with Lévy measure V_X has a finite Nth moment if and only if $\int_{|t|>1} |t|^N V_X(\mathrm{d}t) < \infty$ [53, Theorem 9.9]. Hence, the condition $\mathbb{E}[|\langle w, \mathrm{rect} \rangle|^N] < \infty$ implies that $\int_{|t|>1} |t|^N V(\mathrm{d}t) < \infty$, where V is the Lévy measure of w.

For $\varphi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d) \cap L^N(\mathbb{R}^d)$, we know that the real random variable $\langle w, \varphi \rangle$ is well-defined (see section 3.3) and infinitely divisible with Lévy measure given by $V_\varphi(B) = \int_{\varphi(\mathbf{r}) \neq 0} V\left(\frac{B}{\varphi(\mathbf{r})}\right) \mathrm{d}\mathbf{r}$ for any Borelian set $B \subset \mathbb{R}$, where $B/a = \{x \in \mathbb{R}, ax \in B\}$ [53, Theorem 9.1]. With a slight abuse of notation, we denote $V_\varphi(\mathrm{d}t) = \int_{\varphi(\mathbf{r}) \neq 0} V\left(\frac{\mathrm{d}t}{\varphi(\mathbf{r})}\right) \mathrm{d}\mathbf{r}$. Then, we have

$$\int_{|t|>1} |t|^N V_\varphi(\mathrm{d}t) \leq \int_{\mathbb{R}} |t|^N V_\varphi(\mathrm{d}t)$$

$$= \int_{\mathbb{R}} \int_{\varphi(\mathbf{r}) \neq 0} |t|^N V\left(\frac{\mathrm{d}t}{\varphi(\mathbf{r})}\right) \mathrm{d}\mathbf{r}$$

$$= \int_{\mathbb{R}} \int_{\varphi(\mathbf{r}) \neq 0} |u\varphi(\mathbf{r})|^N V(\mathrm{d}u) \mathrm{d}\mathbf{r}$$

$$= \left(\int_{\mathbb{R}} |u|^N V(\mathrm{d}u) \right) \|\varphi\|_N^N$$

$$(5.13) \qquad\qquad\qquad < \infty,$$

where the finiteness comes from $\varphi \in L^N(\mathbb{R}^d)$ and

$$\int_{\mathbb{R}} |u|^N V(\mathrm{d}u) \leq \int_{|t|>1} |u|^N V(\mathrm{d}u) + \int_{|t| \leq 1} u^2 V(\mathrm{d}u) < \infty.$$

Hence, for any $\varphi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d) \cap L^N(\mathbb{R}^d)$, the real random variable $\langle w, \varphi \rangle$ has a finite Nth moment.

We can apply this result to the process s. For $\varphi \in \mathcal{S}(\mathbb{R}^d)$, the real random variable $\langle s, \varphi \rangle$ has Lévy measure $V_{\mathrm{L}_\gamma^{*-1}\varphi}$ with $\mathrm{L}_\gamma^{*-1}\varphi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d) \cap L^N(\mathbb{R}^d)$, we deduce that

$$\mathbb{E}[|\langle s, \varphi \rangle|^N] = \mathbb{E}[|\langle w, \mathrm{L}_\gamma^{*-1}\varphi \rangle|^N] < \infty.$$

This proves that the process s has finite Nth moments.

Similarly, if $\psi = \mathrm{L}_\gamma^*\phi$ is admissible with $\phi \in L^N(\mathbb{R}^d)$, the Lévy measure of $\langle s, \psi \rangle$ is V_ϕ and $\phi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d) \cap L^N(\mathbb{R}^d)$ by assumption. Again, we deduce that

$$\mathbb{E}[|\langle s, \psi \rangle|^N] = \mathbb{E}[|\langle w, \phi \rangle|^N] < \infty,$$

as expected. ∎

Proposition 5.6 (cumulants of sparse processes across scales). *Let $N \geq 2$. Consider $s = \mathrm{L}_\gamma^{-1}w$ to be a self-similar sparse process of order $\gamma \geq 0$ with admissible wavelet $\psi = \mathrm{L}_\gamma^*\phi$ and $\phi \in L^N(\mathbb{R}^d)$. Then, we have, for $1 \leq n \leq N$,*

$$(5.14) \qquad \kappa_n(j) = \kappa_n(\langle s, \psi_{j,\mathbf{r}_0} \rangle) = (-\mathrm{i})^n f_\phi^{(n)}(0)\rho^{j\left(n\left(\gamma-\frac{d}{2}\right)+d\right)},$$

where $f_\varphi(\xi) = \int_{\mathbb{R}^d} f(\xi\varphi(\mathbf{r}))\mathrm{d}\mathbf{r}$.

Proof. Using Proposition 5.2, we know that the characteristic function of $X = \langle s, \psi_{j,\mathbf{r}_0} \rangle$ satisfies $\log \Phi_X(\xi) = \rho^{jd} \int_{\mathbb{R}^d} f(\xi\rho^{j\left(\gamma-\frac{d}{2}\right)}\phi(\mathbf{r}))\mathrm{d}\mathbf{r} = \rho^{jd} f_\phi(\xi\rho^{j(\gamma-d/2)})$. By taking the derivative of order n and putting $\xi = 0$, we obtain (5.14). ∎

We deduce from Proposition 5.6 that the variance of $\langle s, \psi_{j,\mathbf{r}_0} \rangle$ is finite if s has finite 2nd-moments, in which case

$$(5.15) \qquad \mathrm{Var}(\langle s, \psi_{j,\mathbf{r}_0} \rangle) = \kappa_2(j) = -f_{\mathrm{L}_\gamma^{*-1}\psi}^{(2)}(0)\rho^{2j\gamma}.$$

In particular, the variance of the wavelet coefficient goes to ∞ at coarse scales when $\gamma > 0$. We rely this with the fact that, as is well-known, a self-similar sparse process s satisfies $\|s\|_2 = \infty$, which can be interpreted as an infinite energy. However, in most of the practical situations, typically for image processing, this effect disappears since the scale j is bounded below by some limit scale J_0.

Similarly, the kurtosis of $\langle s, \psi_{j,\mathbf{r}_0} \rangle$, denoted by $\mu(j)$ is well-defined if s has finite 4th-moments, in which case

$$(5.16) \qquad \mu(j) = \frac{\kappa_4(j)}{\kappa_2^2(j)} = \frac{f_{\mathrm{L}_\gamma^{*-1}\psi}^{(4)}(0)}{\left(f_{\mathrm{L}_\gamma^{*-1}\psi}^{(2)}(0)\right)^2}\rho^{-jd} = \mu(0)\rho^{-jd}.$$

Remarkably, $\mu(j)/\mu(0)$ does not depend on γ. We define the *decay rate of the kurtosis* τ by the relation $\log \mu(j)/\mu(0) = \tau j$. From (5.16), we have $\tau = -d$. Moreover, we see that the choice of a wavelet ψ, under the assumption that it is admissible, only affects the value of

$\kappa_n(0)$ in (5.14) in the sense that $\kappa_n(j)/\kappa_n(0)$ is independent of ψ. As such, the choice of a wavelet has only a small impact on the evolution of the cumulants of s across scales.

For the Gaussian case, we have $f_\phi(\xi) = -\frac{\|\phi\|_2^2 \xi^2}{2}$. Thus, $\kappa_n(j) = 0$ for all $n \geq 3$ and all $j \in \mathbb{Z}$, and the kurtosis $\mu_j = 0$ for all $j \in \mathbb{Z}$. Moreover, in the general case, when the kurtosis is well-defined, we have $\mu(j) \underset{j \to +\infty}{\longrightarrow} 0$. This is consistent with the results of Theorem 4.2: at coarse scales, the kurtosis of wavelet coefficients vanishes. We retain the kurtosis as an estimation of the degree of Gaussianity of the wavelet coefficients: the closer it is to 0, the more Gaussian it is.

The results of this section are presented for wavelet functions. However, our framework can be applied for any function ψ that is admissible. In practical terms, this means that one is allowed to use functions that are not necessarily wavelets, provided that $\psi = \mathrm{L}_\gamma^* \phi$ for some $\phi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$.

6. Experimental results. To corroborate our theoretical development, we have performed one- and two-dimensional experiments where all numerical schemes are implemented in MAT-LAB.

6.1. Simulations on synthetic data. First, we analyzed the wavelet coefficients of Lévy processes. Such processes are defined for $d = 1$ and generated by choosing the first-order derivative operator $\mathrm{L}_\gamma = \mathrm{D}$ [55, section VI]. In this case, all wavelet functions are admissible, since the said condition is equivalent to having vanishing moments until order 0—meaning that $\int_\mathbb{R} \psi(r)\mathrm{d}r = 0$.

Lévy processes of length 2^{15} were generated, where we considered a Laplace innovation (see Table 1). The processes were synthesized by integrating a discrete white noise, subject to the standard boundary condition $s(0) = 0$. To assess the effect of operating with different wavelets, we used two distinct wavelet transforms, namely, Haar and Daubechies 4. Both are implemented in nonredundant fashion and follow a dyadic scale progression—i.e., $\rho = 2$; see section 5.

We performed 2,000 realizations. For each realization, the variance and kurtosis of the wavelet coefficients at each scale were computed. The wavelet coefficients at the boundaries were excluded to avoid boundary effects. Empirical mean of the variance and kurtosis were obtained by simply averaging the variance and kurtosis values over all the realizations. It was observed that the variance values had insignificant fluctuation around the mean. By contrast, the kurtosis had a much larger variability as documented in Figure 2. We then regressed the parameters γ (self-similarity order) and τ (decay rate of the kurtosis) by using the empirical mean of the variance and kurtosis. Decomposition levels 2 to 8 were used for the regression.

Considering (5.15) in logarithmic scale, the computation of γ was performed through a simple linear regression. As for the τ parameter, we used (5.16) and carried out a nonlinear least-squares fitting provided by the standard Levenberg–Marquardt algorithm. Besides the issue of variability, a potential difficulty is that the kurtosis value gets closer to zero as the number of decomposition levels increases. Therefore, the approach for computing τ is observed to be more stable and robust than performing a linear regression (in a logarithmic scale). Finally, we compared the computed γ and τ values to the theoretical expectations.

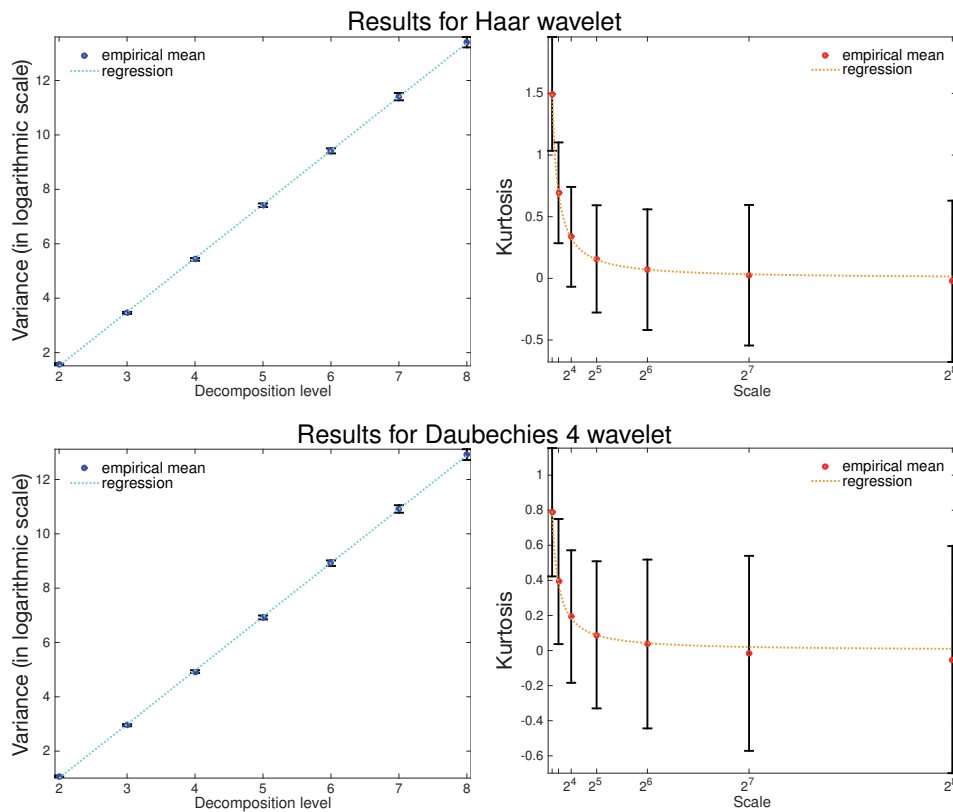Based on the results illustrated in Table 2, one sees that the regressed γ values are very

Figure 2. *Evolution of the variance and the kurtosis. Analysis of one-dimensional Lévy processes with Laplace innovation. Vertical bars represent one empirical standard deviation. Dashed curves illustrate appropriate polynomial fits to the empirical mean.*

Table 2
Regression results for parameters γ and τ based on the empirical mean of the variance and the kurtosis. Reported numbers correspond to the analysis of one-dimensional Lévy processes with Laplace innovation by using different wavelets.

		Haar wavelet		Daubechies 4 wavelet	
True γ	True τ	Regressed γ	Regressed τ	Regressed γ	Regressed τ
1	-1	0.99	-1.08	0.99	-1.05

close to the theoretical ones. This confirms the generality of our approach for non-Gaussian Lévy processes, extending in particular the results on the fractional Brownian motion of [49]. Further, we observe that the estimates of the decay rate of the kurtosis are also close to the expected value, which is (-1). By looking at the illustrations given in Figure 2, one sees that the kurtosis values are converging to 0 for the considered process. Notice that the choice of different wavelets affects only the constant factor $\mu(0)$ given in (5.16). As formalized by our theory, the decay rate does not depend on this factor.

In the second part of our experiments, we generated two-dimensional self-similar sparse

Table 3
Regression results for parameters γ and τ based on the empirical mean of the variance and the kurtosis. Reported numbers correspond to the analysis of two-dimensional self-similar sparse processes with Student's t-innovation for different orders of self-similarity.

True γ	True τ	Regressed γ	Regressed τ
0.8		0.79	-2.03
1.5	-2	1.49	-2.04
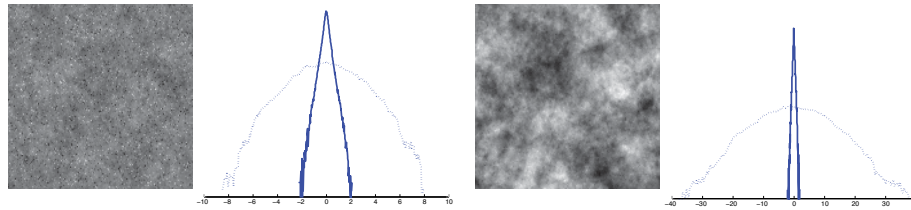
Figure 3. *Realizations of self-similar sparse processes driven by Student's t-innovation and their normalized histogram (in logarithmic scale) of the expansion coefficients at the second decomposition level (solid) and at the eighth decomposition level (dashed). The horizontal axis represent the value of the expansion coefficient, whereas the vertical one denotes the frequency. On left: $\gamma = 0.8$. On right: $\gamma = 1.5$.*

processes, each being realized on a $(1{,}024 \times 1{,}024)$ grid. The self-similar operator was chosen to be the fractional Laplacian operator specified in (3.8). For synthesizing the processes, a discrete version of the inverse fractional Laplacian operator was applied to a discrete white noise in the Fourier domain. We considered the self-similarity orders $\gamma = 0.8$ and 1.5 and a Student's t-innovation with a large enough ν parameter (see Table 1) to ensure the existence of the kurtosis. Realizations of such processes are given in Figure 3. As for the analysis, we operated with the Laplacian-of-Gaussian (LoG) kernel $\psi = \Delta G$, where G is a zero-mean Gaussian function with unit variance. The wavelet is isotropic and admissible for $\gamma < 2$, since $\psi = -(-\Delta)^{\gamma/2}\{(-\Delta)^{1-\gamma/2}G\}$. This is in fact related to ψ having vanishing moments until order 1 (see Proposition 5.1).

Similar to our previous experiments, computations of the mean variance and the mean kurtosis of the expansion coefficients were obtained though averaging over 2,000 realizations at every scale. The processes were analyzed using the LoG kernel with the dilation factor $\rho = \sqrt{2}$. The expansion coefficients at the boundaries were again discarded by applying a sufficiently large cropping. To have enough samples for statistical estimation (especially at coarser scales), subsampling operations were omitted.[1] We computed the parameters γ and τ as described in the one-dimensional experiments. We then compared the regressed γ and τ values and the theoretical ones (see Table 3). We also tested our framework on different natural images of size (512×512) (see Figure 4).

The results of our experiments reveal that the γ and τ values, estimated in two-dimensional contents, are again close to the true ones, further supporting over previous assertions. For the sake of completeness, the mean variance and the mean kurtosis of the expansion coefficients is illustrated in Figure 5. At this point, we stress that the decay rate of the kurtosis is

---

[1]This type of expansion is also known as the scale-space representation in the image-processing community.
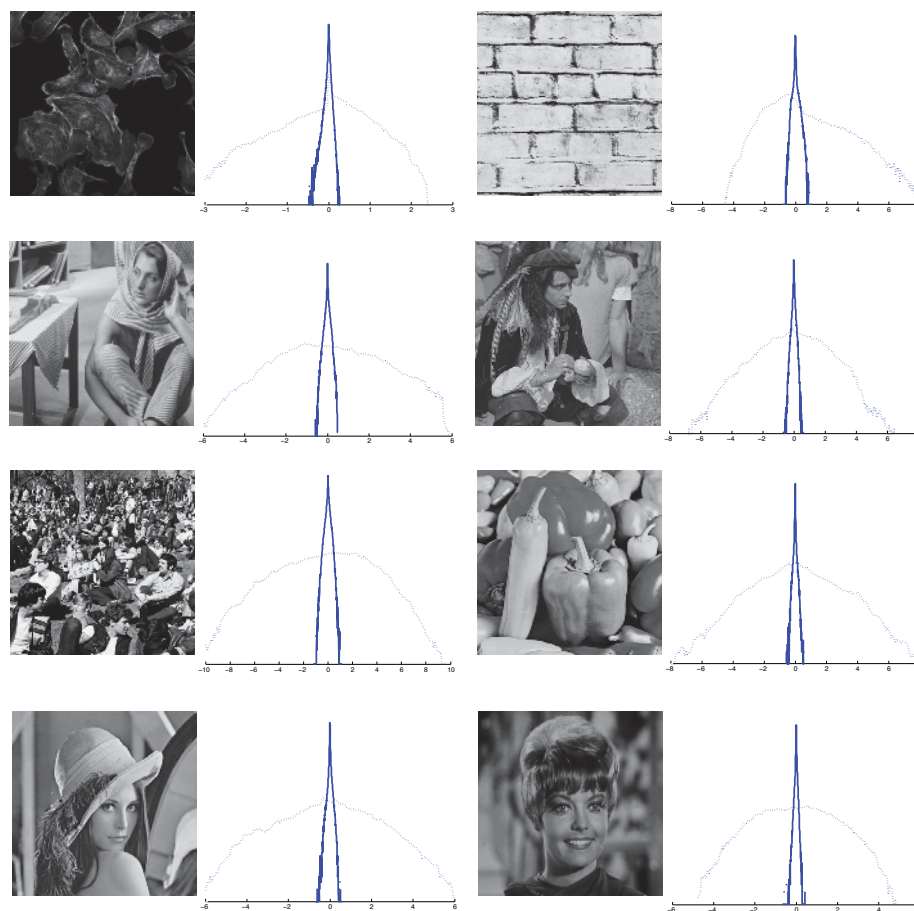
Figure 4. *Natural images used in the experiments and their normalized histograms (in logarithmic scale) of the expansion coefficients at the second decomposition level (solid) and at the eighth decomposition level (dashed). The horizontal axis represents the value of the expansion coefficient, whereas the vertical one denotes the frequency. On left, from top to bottom: Artery cells, Barbara, Crowd, Lena. On right, from top to bottom: Texture, Man, Peppers, Zelda.*

independent of γ.

6.2. Simulations on natural image. As for natural images, we confirm the Gaussianity of the coarse-scale coefficients in a qualitative manner by looking at the histograms (in logarithmic scale) in Figure 4. We observe that the given histograms fit a quadratic curve—what we expect for a Gaussian signal—more closely at coarse scales than at fine scales. This is corroborated by the plots in Figure 6: the decay of the kurtosis indicates an increasingly Gaussian behavior at coarse scales.

While the qualitative behavior is as expected, the observed rate of decay of the kurtosis is not quite as strong as predicted by the theory. First of all, the experiments on synthetic data have shown that the estimations based on the kurtosis are highly variable (see Figures 2 and 5). In Figure 6, the kurtosis is computed from a single estimate—the analyzed natural image, which makes the quantitative interpretation of the decay rate possibly problematic. In spite
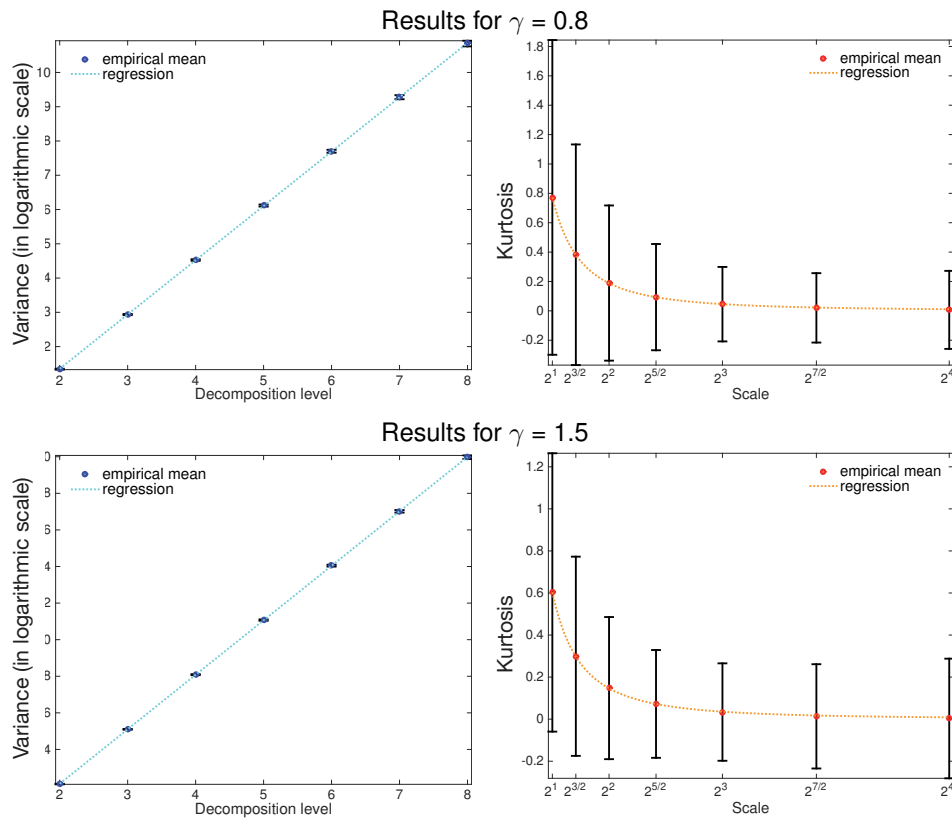
Figure 5. *Evolution of the variance and the kurtosis. Analysis of two-dimensional self-similar sparse processes with Student's t-innovation. Vertical bars represent one empirical standard deviation. Dashed curves illustrate appropriate polynomial fits to the empirical mean.*
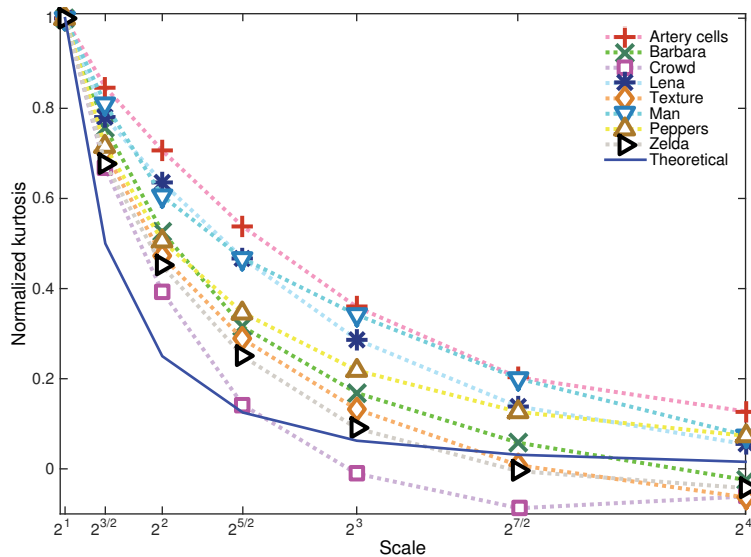

Figure 6. *Evolution of the kurtosis of the expansion coefficients of the natural images used in the experiments.*

of this uncertainty on the estimation, one sees in Figure 6 that the decay of the kurtosis of natural images deviate from the theoretical curve. This suggests that on top of self-similarity and sparsity, additional features need to be incorporated into the stochastic model to better capture the statistical properties of natural images.

7. Conclusion. This paper has been motivated by the search of stochastic models that are phenomenologically consistent with the compressibility and scale invariance properties of natural images. To that purpose, we modeled an image as a random process that can be transformed to a sparse white noise (i.e., an innovation process) by applying a homogeneous operator L_γ. The parameter γ denotes the self-similarity order of the process. We established and explained the mathematical framework that allows one to generate self-similar sparse processes. This development was followed by the analysis of such processes. In section 4, we studied the rescaled versions of self-similar sparse processes. We showed that the dilated versions of these processes are asymptotically Gaussian. The evolution of the process across scales had direct implications on the statistics of its wavelet coefficients. We further quantified in section 5 the degree of Gaussianity of the wavelet coefficients by providing the theoretical evolution of the kurtosis across scales. Finally, we tested our predictions in one- and two-dimensional settings. Our numerical simulations showed that the developed formalism is very well-matched to synthetic data. We also observed qualitatively a Gaussian behavior at coarse scales for natural images. Despite this good qualitative behavior, the evolution of the kurtosis of natural images were observed to deviate from the exact predicted decay rate. In this regard, the proposed framework can be seen as a first step towards a more precise stochastic model for natural images.

Appendix A. Proof of Theorem 4.2. In order to prove Theorem 4.2, we shall first control the Lévy exponent f with the following result.

Lemma A.1. *Let $w \in \mathcal{S}'(\mathbb{R}^d)$ be an innovation, f being its Lévy exponent, and let (μ, σ^2, V) be its Lévy triplet. We recall the notation $\mu_w = \mu + \int_{|t|>1} tV(\mathrm{d}t)$ and $\sigma_w^2 = \sigma^2 + \int_{\mathbb{R}} t^2 V(\mathrm{d}t)$. Then we have, for all $\xi \in \mathbb{R}$,*

$$(A.1) \qquad \left| f(\xi) - \mathrm{i}\mu_w\xi \right| \leq \sigma_w^2 \xi^2.$$

Proof. From (2.6), we know that

$$f(\xi) - \mathrm{i}\mu_w\xi = -\frac{\sigma^2\xi^2}{2} + \int_{\mathbb{R}} (\mathrm{e}^{\mathrm{i}t\xi} - 1 - \mathrm{i}\xi t 1_{|t|\leq 1})V(\mathrm{d}t) - \int_{\mathbb{R}} \mathrm{i}\xi t 1_{|t|>1} V(\mathrm{d}t)$$

$$(A.2) \qquad = -\frac{\sigma^2\xi^2}{2} + \int_{\mathbb{R}} (\mathrm{e}^{\mathrm{i}\xi t} - 1 - \mathrm{i}\xi t)V(\mathrm{d}t).$$

Using $|\mathrm{e}^{\mathrm{i}x} - 1 - \mathrm{i}x| \leq x^2$, we therefore deduce that

$$\left| f(\xi) - \mathrm{i}\mu_w\xi \right| \leq \frac{\sigma^2\xi^2}{2} + \int_{\mathbb{R}} t^2\xi^2 V(\mathrm{d}t) \leq \sigma_w^2\xi^2. \qquad \blacksquare$$

Proof of Theorem 4.2. We first treat the nonsymmetric case. The characteristic functional of $s = L_\gamma^{-1}w$ is $\widehat{\mathscr{P}_s}(\varphi) = \widehat{\mathscr{P}_w}(L_\gamma^{*-1}\varphi)$ with L_γ^{*-1} a $(-\gamma)$-homogeneous left-inverse of L_γ^*,

continuous from $\mathcal{S}(\mathbb{R}^d)$ to $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$ (see section 3.2). In particular, $\mathrm{L}_\gamma^{*-1}(\mathcal{S}(\mathbb{R}^d)) \subset L^1(\mathbb{R}^d)$ and we can define $v \in \mathcal{S}'(\mathbb{R}^d)$ as

$$\langle v, \varphi \rangle = \langle 1, \mathrm{L}_\gamma^{*-1} \varphi \rangle = \int_{\mathbb{R}^d} \mathrm{L}_\gamma^{*-1}\{\varphi\}(\mathbf{r}) \mathrm{d}\mathbf{r} \quad \forall \varphi \in \mathcal{S}(\mathbb{R}^d).$$

Theorem 4.2 can be seen as a generalization of the central-limit theorem for self-similar sparse processes. Let X be a finite-variance infinitely divisible random variable with Lévy exponent f and Lévy triplet (μ, σ^2, V). The mean of X is then $\mu_X = \mu + \int_{|t|>1} t V(\mathrm{d}t)$ and its variance $\sigma_X^2 = \sigma^2 + m_2(V)$. The central-limit theorem is equivalent to the pointwise convergence

$$(A.3) \qquad \tau f\left(\frac{\xi}{\sqrt{\tau}}\right) - \mathrm{i}\sqrt{\tau}\mu_X \xi \xrightarrow[\tau \to \infty]{} -\frac{\sigma_X^2 \xi^2}{2} \qquad \forall \xi \in \mathbb{R}.$$

Thanks to (2.4), we know that the convergence in law of generalized random processes is equivalent with the pointwise convergence of the corresponding characteristic functionals. Therefore, (4.7) is equivalent with

$$(A.4) \qquad \widehat{\mathscr{P}}_{a^{-\gamma} S_{a^{-1}} s - \mu_w a^{d/2} v}(\varphi) \xrightarrow[a \to \infty]{} \exp\left(-\frac{\sigma_w^2 \|\mathrm{L}_\gamma^{*-1}\varphi\|_2^2}{2}\right).$$

Let $\varphi \in \mathcal{S}(\mathbb{R}^d)$. We have

$$
\begin{aligned}
\log \widehat{\mathscr{P}}_{a^{-\gamma} S_{a^{-1}} s - \mu_w a^{d/2} v}(\varphi) &= \log \widehat{\mathscr{P}}_{a^{-\gamma} S_{a^{-1}} s}(\varphi) - \mathrm{i}\mu_w a^{d/2}\langle v, \varphi \rangle \\
&= \log \widehat{\mathscr{P}}_s(a^{-\gamma} S_a \varphi) - \mathrm{i}\mu_w a^{d/2}\langle v, \varphi \rangle \\
&\stackrel{(i)}{=} \log \widehat{\mathscr{P}}_w(\mathrm{L}_\gamma^{*-1}\{a^{-\gamma} S_a \varphi\}) - \mathrm{i}\mu_w a^{d/2}\langle 1, \mathrm{L}_\gamma^{*-1}\varphi \rangle \\
&\stackrel{(ii)}{=} \log \widehat{\mathscr{P}}_w(S_a \mathrm{L}_\gamma^{*-1}\varphi) - \mathrm{i}\langle \mu_w a^{d/2}, \mathrm{L}_\gamma^{*-1}\varphi \rangle \\
(A.5) \qquad &= \log \widehat{\mathscr{P}}_{S_{a^{-1}} w - \mu_w a^{d/2}}(\mathrm{L}_\gamma^{*-1}\varphi),
\end{aligned}
$$

where (i) is deduced from the relation $\widehat{\mathscr{P}}_s(\cdot) = \widehat{\mathscr{P}}_w(\mathrm{L}_\gamma^{*-1}\cdot)$ and (ii) by $(-\gamma)$-homogeneity. Based on (A.5) and since $\mathrm{L}_\gamma^{*-1}(\mathcal{S}(\mathbb{R}^d)) \subset L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$, we obtain (A.4) if we prove that, for all $\varphi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$,

$$(A.6) \qquad \log \widehat{\mathscr{P}}_{S_{a^{-1}} w - \mu_w a^{d/2}}(\varphi) \xrightarrow[a \to \infty]{} -\frac{\sigma_w^2 \|\varphi\|_2^2}{2}.$$

We have, for all $\varphi \in L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$,

$$
\begin{aligned}
\log \widehat{\mathscr{P}}_{S_{a^{-1}} w - \mu_w a^{d/2}}(\varphi) &= \int_{\mathbb{R}^d} a^d f(a^{-d/2}\varphi(\mathbf{r})) \mathrm{d}\mathbf{r} - \mathrm{i}\mu_w a^{d/2} \int_{\mathbb{R}^d} \varphi(\mathbf{r}) \mathrm{d}\mathbf{r} \\
&= \int_{\mathbb{R}^d} \left(a^d f(a^{-d/2}\varphi(\mathbf{r})) - \mathrm{i}\mu_w a^{d/2}\varphi(\mathbf{r})\right) \mathrm{d}\mathbf{r}.
\end{aligned}
$$

Applying (A.3) with $\tau = a^d$ and $\xi = \varphi(\mathbf{r})$ we obtained that, for all $\mathbf{r} \in \mathbb{R}^d$,

$$(A.7) \qquad a^d f(a^{-d/2}\varphi(\mathbf{r})) - \mathrm{i}\mu_w a^{d/2}\varphi(\mathbf{r}) \xrightarrow[a\to+\infty]{} -\frac{\sigma_w^2 \varphi(\mathbf{r})}{2}.$$

Moreover, applying (A.1) with $\xi = a^{-d/2}\varphi(\mathbf{r})$, we have, for all $\mathbf{r} \in \mathbb{R}^d$,

$$(A.8) \qquad \left| a^d f(a^{-d/2}\varphi(\mathbf{r})) - \mathrm{i}\mu_w a^{d/2}\varphi(\mathbf{r}) \right| \le \sigma_w^2 \varphi(\mathbf{r})^2,$$

which is integrable because $\varphi \in L^2(\mathbb{R}^d)$. Finally, we deduce (A.6) using the Lebesgue dominated-convergence theorem, which finishes the nonsymmetric case.

The symmetric case is almost identical and actually simpler. The first difference is that L_γ^{*-1} is not necessarily continuous in $L^1(\mathbb{R}^d)$, also we cannot introduce the generalized function v. However, since $\mu_w = 0$ for a symmetric innovation, there is no need to compensate the drift by introducing v. We can therefore follow the same proof with $\mu_w = 0$.

REFERENCES

[1] A. ACHIM, P. TSAKALIDES, AND A. BEZERIANOS, *SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling*, IEEE Trans. Geosci. Remote Sens., 41 (2003), pp. 1773–1784.

[2] A. AMINI AND M. UNSER, *Sparsity and infinite divisibility*, IEEE Trans. Inform. Theory, 60 (2014), pp. 2346–2358.

[3] A. AMINI, M. UNSER, AND F. MARVASTI, *Compressibility of deterministic and random infinite sequences*, IEEE Trans. Signal Process., 59 (2011), pp. 5193–5201.

[4] H. BIERMÉ, A. ESTRADE, AND I. KAJ, *Self-similar random fields and rescaled random balls models*, J. Theoret. Probab., 23 (2010), pp. 1110–1141.

[5] J. M. BIOUCAS-DIAS, *Bayesian wavelet-based image deconvolution: A GEM algorithm exploiting a class of heavy-tailed priors*, IEEE Trans. Image Process., 15 (2006), pp. 937–951.

[6] E. BOSTAN, U. KAMILOV, M. NILCHIAN, AND M. UNSER, *Sparse stochastic processes and discretization of linear inverse problems*, IEEE Trans. Image Process., 22 (2013), pp. 2699–2710.

[7] P. BOULICAUT, *Convergence cylindrique et convergence étroite d'une suite de probabilités de Radon*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 28 (1973/74), pp. 43–52.

[8] C. BOUMAN AND K. SAUER, *A generalized Gaussian image model for edge-preserving MAP estimation*, IEEE Trans. Image Process., 2 (1993), pp. 296–310.

[9] R. BUCCIGROSSI AND E. SIMONCELLI, *Image compression via joint statistical characterization in the wavelet domain*, IEEE Trans. Image Process., 8 (1999), pp. 1688–1701.

[10] E. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.

[11] V. CEVHER, *Learning with compressible priors*, Adv. Neural Inf. Process. Syst., 22 (2009), pp. 261–269.

[12] P. CHAINAIS, *Infinitely divisible cascades to model the statistics of natural images*, IEEE Trans. Pattern Anal. Mach. Intell., 29 (2007), pp. 2105–2119.

[13] Z. CHI, *Probability Models for Complex Systems*, Ph.D. thesis, Brown University, Providence, Rhode Island, 1998.

[14] R. L. DOBRUSHIN, *Gaussian and their subordinated self-similar random generalized fields*, Ann. Probab., (1979), pp. 1–28.

[15] D. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[16] A. DURAND AND S. JAFFARD, *Multifractal analysis of Lévy fields*, Probab. Theory Related Fields, 153 (2012), pp. 45–96.

[17] J. FAGEOT, A. AMINI, AND M. UNSER, *On the continuity of characteristic functionals and sparse stochastic modeling*, J. Fourier Anal. Appl., 20 (2014), pp. 1179–1211.

[18] J. FAGEOT, E. BOSTAN, AND M. UNSER, *Statistics of wavelet coefficients for sparse and self-similar images*, in Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP'14), Paris, France, 2014, pp. 6096–6100.

[19] P. FLANDRIN, *On the spectrum of fractional Brownian motions*, IEEE Trans. Inform. Theory, 35 (1989), pp. 197–199.

[20] G. FREEDMAN AND R. FATTAL, *Image and video upscaling from local self-examples*, ACM Trans. Graph., 30 (2011), article 12.

[21] I. M. GEL'FAND AND N. YA. VILENKIN, *Generalized Functions. Vol. 4: Applications of Harmonic Analysis*, translated by A. Feinstein, Academic Press, New York, London, 1964.

[22] D. GLASNER, S. BAGON, AND M. IRANI, *Super-resolution from a single image*, in 2009 IEEE 12th International Conference on Computer Vision, IEEE, 2009, pp. 349–356.

[23] U. GRENANDER AND A. SRIVASTAVA, *Probability models for clutter in natural images*, IEEE Trans. Pattern Anal. Mach. Intell., 23 (2001), pp. 424–429.

[24] R. GRIBONVAL, V. CEVHER, AND M. DAVIES, *Compressible distributions for high-dimensional statistics*, IEEE Trans. Inform. Theory, 58 (2012), pp. 5016–5034.

[25] J. HUANG AND D. MUMFORD, *Statistics of natural images and models*, in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, Vol. 1, IEEE, 1999.

[26] K. ITÔ, *Foundations of Stochastic Differential Equations in Infinite Dimensional Spaces*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 47, SIAM, Philadelphia, 1984.

[27] N. JACOB, *Pseudo Differential Operators & Markov Processes: Markov Processes and Applications*, Vol. 3, Imperial College Press, London, 2005.

[28] N. JACOB AND R. SCHILLING, *Lévy-type processes and pseudodifferential operators*, in Lévy Processes, Birkhäuser Boston, Boston, MA, 2001, pp. 139–168.

[29] S. MALLAT, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 1998.

[30] B. MANDELBROT, *The Fractal Geometry of Nature*, W. H. Freeman, San Francisco, 1982.

[31] B. B. MANDELBROT AND J. W. VAN NESS, *Fractional Brownian motions, fractional noises and applications*, SIAM Rev., 10 (1968), pp. 422–437.

[32] Y. MEYER, F. SELLAN, AND M. S. TAQQU, *Wavelets, generalized white noise and fractional integration: The synthesis of fractional Brownian motion*, J. Fourier Anal. Appl., 5 (1999), pp. 465–494.

[33] T. MICHAELI AND M. IRANI, *Blind deblurring using internal patch recurrence*, in Computer Vision–ECCV 2014, Springer, Cham, Switzerland, 2014, pp. 783–798.

[34] P. MOULIN AND J. LIU, *Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors*, IEEE Trans. Inform. Theory, 45 (1999), pp. 909–919.

[35] D. MUMFORD AND A. DESOLNEUX, *Pattern Theory. The Stochastic Analysis of Real-World Signals*, Applying Mathematics, A K Peters, Natick, MA, 2010.

[36] D. MUMFORD AND B. GIDAS, *Stochastic models for generic images*, Quart. Appl. Math., 59 (2001), pp. 85–111.

[37] D. H. MUSHTARI, *Probabilities and Topologies on Linear Spaces*, Kazan Tracts in Mathematics 1, Kazan Mathematics Foundation, Kazan, 1996.

[38] E. PERRIN, R. HARBA, C. BERZIN-JOSEPH, I. IRIBARREN, AND A. BONAMI, *Nth-order fractional Brownian motion and fractional Gaussian noises*, IEEE Trans. Signal Process., 49 (2001), pp. 1049–1059.

[39] B. PESQUET-POPESCU AND J. L. VÉHEL, *Stochastic fractal models for image processing*, IEEE Signal Processing Magazine, 19 (2002), pp. 48–62.

[40] B. S. RAJPUT AND J. ROSIŃSKI, *Spectral representations of infinitely divisible processes*, Probab. Theory Related Fields, 82 (1989), pp. 451–487.

[41] D. RUDERMAN AND W. BIALEK, *Statistics of natural images: Scaling in the woods*, Phys. Rev. Lett., 73 (1994), pp. 814–817.

[42] G. SAMORODNITSKY AND M. TAQQU, *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*, Chapman and Hall, London, 1994.

[43] K.-I. SATO, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge Stud. Adv. Math. 68, Cambridge University Press, Cambridge, 2013, translated from the 1990 Japanese original, revised edition of the 1999 English translation.

[44] M. Shao and C. Nikias, *Signal processing with fractional lower order moments: Stable processes and their applications*, Proceedings of the IEEE, 81 (1993), pp. 986–1010.

[45] E. P. Simoncelli, *Statistical models for images: Compression, restoration and synthesis*, in Conference Record of the Thirty-First Asilomar Conference on Signals, Systems & Computers, 1997, Vol. 1, IEEE, New York, 1997, pp. 673–678.

[46] A. Srivastava, A. B. Lee, E. P. Simoncelli, and S.-C. Zhu, *On advances in statistical modeling of natural images*, J. Math. Imaging Vision, 18 (2003), pp. 17–33.

[47] Q. Sun and M. Unser, *Left-inverses of fractional Laplacian and sparse stochastic processes*, Adv. Comput. Math., 36 (2012), pp. 399–441.

[48] P. D. Tafti and M. Unser, *Fractional Brownian vector fields*, Multiscale Model. Simul., 8 (2010), pp. 1645–1670.

[49] P. D. Tafti, D. V. D. Ville, and M. Unser, *Invariances, Laplacian-like wavelet bases, and the whitening of fractal processes*, IEEE Trans. Image Process., 18 (2009), pp. 689–702.

[50] F. Trèves, *Topological Vector Spaces, Distributions and Kernels*, Academic Press, New York, London, 1967.

[51] J. S. Turek, I. Yavneh, and M. Elad, *On MMSE and MAP denoising under sparse representation modeling over a unitary dictionary*, IEEE Trans. Signal Process., 59 (2011), pp. 3526–3535.

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

[53] M. Unser and P. Tafti, *An Introduction to Sparse Stochastic Processes*, Cambridge University Press, Cambridge, UK, 2014.

[54] M. Unser, P. D. Tafti, A. Amini, and H. Kirshner, *A unified formulation of Gaussian versus sparse stochastic processes—Part* II: *Discrete-domain theory*, IEEE Trans. Inform. Theory, 60 (2014), pp. 3036–3051.

[55] M. Unser, P. D. Tafti, and Q. Sun, *A unified formulation of Gaussian versus sparse stochastic processes—Part* I: *Continuous-domain theory*, IEEE Trans. Inform. Theory, 60 (2014), pp. 1945–1962.