STATISTICS OF WAVELET COEFFICIENTS FOR SPARSE SELF-SIMILAR IMAGES

Julien Fageot, Emrah Bostan, and Michael Unser

Biomedical Imaging Group, EPFL, Lausanne, Switzerland

ABSTRACT

We study the statistics of wavelet coefficients of non-Gaussian images, focusing mainly on the behaviour at coarse scales. We assume that an image can be whitened by a fractional Laplacian operator, which is consistent with an $\|\omega\|^{-\gamma}$ spectral decay. In other words, we model images as sparse and self-similar stochastic processes within the framework of generalised innovation models. We show that the wavelet coefficients at coarse scales are asymptotically Gaussian even if the prior model for fine scales is sparse. We further refine our analysis by deriving the theoretical evolution of the cumulants of wavelet coefficients across scales. Especially, the evolution of the kurtosis supplies a theoretical prediction for the Gaussianity level at each scale. Finally, we provide simulations and experiments that support our theoretical predictions.

***Index Terms*—**Wavelet statistics, innovation modelling, sparse stochastic processes, self-similarity.

1. INTRODUCTION

Natural and biomedical images are known to be sparse in the sense that their energy is concentrated in few coefficients in a well-chosen transform domain [1]. This concentration effect is especially noticeable in the wavelet domain and is often used as justification for the deployment of ℓ_1-regularization methods.

The empirical concept of sparsity can be formalised as follows. A signal is statistically sparse if its density has heavier tails than the Gaussian distribution [2]. It therefore implies that a small fraction of all coefficients of a sparse signal carries most of its energy. This definition highlights the limitations of Gaussian priors and suggests using more refined statistical models to tackle practical image-processing tasks [3].

In this paper, we investigate the statistics of wavelet coefficients of images within a non-Gaussian setting. At any scale of a wavelet decomposition, wavelet coefficients are known to be empirically sparse. However, at coarser scales, wavelet coefficients correspond to a renormalized average of a large number of image pixels. This fact is reminiscent of the settings of the central limit theorem. It therefore suggests that the statistics of the wavelet coefficients are asymptotically Gaussian at coarse scales, a conclusion that stands in contradiction with the traditional assumption of sparsity.

In our framework, we consider a continuous-domain probabilistic model of images using the theory of sparse stochastic processes [4]. This is consistent with the observed sparsity of natural images. Because of its fractal properties, our model is also well-adapted to signals with an $\|\omega\|^{-\gamma}$ spectral decay [5]. It is therefore appropriate for the study of natural images [6, 7].

Our contributions in this paper are threefold. We first give theoretical predictions of the statistical behaviour of wavelet coefficients at coarse scales. Our main theoretical result shows that the asymptotic behaviour is indeed Gaussian. Then, we quantify the Gaussianity by deriving the evolution of the cumulants across scales. This analysis allows us to directly obtain variance and kurtosis for a general (possibly non-Gaussian) process. We thus generalise the results of Tafti *et al.* [8] who focus on the variance of Gaussian self-similar processes. Finally, we perform simulations to test the stability of our theoretical predictions. We also provide experimental evidence that the wavelet coefficients of natural images become Gaussian at coarser scales.

2. SELF-SIMILAR MODEL

We model an image by the discretized version of the *continuous-domain* stochastic process $s(\mathbf{r})$ indexed over $\mathbf{r} \in \mathbb{R}^2$. Further, we define the fractional Laplacian operator $(-\Delta)^{\gamma/2}$ for functions φ in the space of smooth and rapidly decreasing functions $\mathcal{S}(\mathbb{R}^2)$ as

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

where $\mathcal{F}$ denotes the Fourier transform. Furthermore, we assume that $\widehat{\mathscr{P}}$ whitens s. Mathematically speaking, s satisfies the self-similar *innovation model* expressed as:

$$(-\Delta)^{\gamma/2}s = w, \tag{2}$$

where w is an innovation process (continuous-domain white noise) and $\gamma \geq 0$ [4, 9]. Essentially, the innovation process w determines the underlying sparsity pattern of s whereas the operator $(-\Delta)^{\gamma/2}$ governs the correlation structure and the level of self-similarity.

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

The crucial point of (2) is that it allows for sparse probabilistic models that are potentially a better fit than the Gaussian ones for real-world images [7].

2.1. The Whitening Operator: $(-\Delta)^{\gamma/2}$

The operator $(-\Delta)^{\gamma/2}$ defined in (1) is linear, shift-invariant, and isotropic. Moreover, it is homogeneous of order γ, implying that

$$\left\{(-\Delta)^{\gamma/2}\{\varphi\}\right\}\left(\frac{\cdot}{\sigma}\right) = |\sigma|^{\gamma}(-\Delta)^{\gamma/2}\left\{\varphi\left(\frac{\cdot}{\sigma}\right)\right\} \quad (3)$$

for all $\varphi \in \mathcal{S}(\mathbb{R}^2)$ and $\sigma \neq 0$. Note that (3) qualifies $(-\Delta)^{\gamma/2}$ as the suitable operator for analysis since scale-invariant probability models are commonly used for image processing [6].

To formally define the process s such that $(-\Delta)^{\gamma/2}s = w$, one needs to specify the inverse I_γ of the fractional Laplacian. This operator, named as Riesz potential, is well-defined for $\gamma < 2$ (classical Riesz potential) and is extended for non-integer $\gamma > 2$ in [10]. It is shift-invariant and homogeneous of order $(-\gamma)$, though I_γ can be unstable. According to [10], we have that for a fixed $\varphi \in \mathcal{S}(\mathbb{R}^2)$

$$\forall \mathbf{k} \in \mathbb{N}^2, |\mathbf{k}| \leq \lfloor\gamma\rfloor, \int_{\mathbb{R}^2} \mathbf{r}^{\mathbf{k}}\varphi(\mathbf{r})\mathrm{d}\mathbf{r} = 0 \Rightarrow \mathrm{I}_\gamma\{\varphi\} \in L^2(\mathbb{R}^2), \quad (4)$$

where $|\mathbf{k}| = k_1 + k_2$. This condition will be sufficient for our analysis and encourages us to use functions φ with enough vanishing moments.

2.2. The Innovation Process: w

Continuous-domain white noises, also called innovation processes, are defined as generalised stochastic processes since they cannot be defined pointwise [11]. This framework is the stochastic counterpart of Schwartz' theory of generalised functions. The process w is thus observed through scalar products $\langle w, \varphi\rangle$ with test functions $\varphi \in \mathcal{S}(\mathbb{R}^2)$. Note that the scalar products yield conventional random variables.

Innovation processes are stationary processes with independent values at every point. In this sense, they are the proper generalisation of the discrete white noises that are collection of independent and identically distributed (iid) random variables. Throughout this work, we shall focus on the finite-variance and symmetric innovation processes (*i.e.*, the random variables $\langle w, \varphi\rangle$ have a finite variance and a symmetric pdf for all φ) by arguing that these properties are empirically fulfilled for natural images [7].

At this point, some mathematical concepts on which we shall base our derivation need to be explained. We first note that w is completely specified by its characteristic functional, defined for $\varphi \in \mathcal{S}(\mathbb{R}^2)$ as

$$\widehat{\mathscr{P}}_w(\varphi) = \mathbb{E}\left[\mathrm{e}^{\mathrm{j}\langle w,\varphi\rangle}\right] = \exp\left(\int_{\mathbb{R}^2} f(\varphi(\mathbf{r}))\mathrm{d}\mathbf{r}\right), \quad (5)$$

where $f(\cdot)$ is the corresponding Lévy exponent of w. For instance, in the Gaussian case, $f(\omega) = -\sigma^2\omega^2/2$ and hence $\widehat{\mathscr{P}}_w(\varphi) = \mathrm{e}^{-\frac{1}{2}\sigma^2\|\varphi\|_2^2}$. The complete characterisation of admissible Lévy exponents is described in [4].

Because of the finite variance and symmetry of w, one can show that the functional $\widehat{\mathscr{P}}_w$ can be extended continuously over the function space $L^2(\mathbb{R}^2)$ [4]. This means that we can give a meaning to random variables $\langle w, \varphi\rangle$ for $\varphi \in L^2(\mathbb{R}^2)$. Moreover, the characteristic function of $\langle w, \varphi\rangle$ for a fixed φ can be directly deduced from (5) as being equal to

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{j}\omega\langle w,\varphi\rangle}\right] = \widehat{\mathscr{P}}_w(\omega\varphi) = \exp\left(\int_{\mathbb{R}^2} f(\omega\varphi(\mathbf{r}))\mathrm{d}\mathbf{r}\right). \quad (6)$$

We also recall the basic fact that there is a one-to-one correspondence between the Lévy exponents and the infinitely divisible (id) probability densities p_{id} [4]. This relation is given by $\mathrm{e}^{f(\omega)} = \mathcal{F}\{p_{\text{id}}\}(\omega)$. By identifying the Taylor expansions in the previous equation, we obtain that the nth cumulant of p_{id}, if it exists, is equal to $\kappa_n = (-\mathrm{j})^n f^{(n)}(0)$.

2.3. Self-Similar Processes: $(-\Delta)^{\gamma/2}s = w$

We fix here a finite-variance and symmetric innovation process w and $\gamma \geq 0$ with $(\gamma - 2) \notin \mathbb{N}$. Under these conditions, we can define the process s as the solution of (2) (see [9], Section 4). Moreover, if ψ is a smooth function such that $\psi = (-\Delta)^{\gamma/2}\phi$ with $\phi \in L^2(\mathbb{R}^2)$, then ψ satisfies (4). Consequently, (2) is formally equivalent to

$$\langle s, \psi\rangle = \langle s, (-\Delta)^{\gamma/2}\phi\rangle = \langle(-\Delta)^{\gamma/2}s, \phi\rangle = \langle w, \phi\rangle \quad (7)$$

using the self-adjoint property of $(-\Delta)^{\gamma/2}$ and the extension of $\langle w, \cdot\rangle$ for functions in $L^2(\mathbb{R}^2)$. Simply put, the statistics of the random variable $\langle s, \psi\rangle$ are deduced from the ones of $\langle w, \phi\rangle$ using (6) with $\varphi = \phi$.

3. MULTISCALE ANALYSIS

Now that the self-similar process s is well-defined, we characterise the statistics of its wavelet coefficients. Specifically, we prove that the wavelet coefficients are becoming Gaussian as the scale gets coarser. To obtain this result, we start by introducing the wavelet decomposition of self-similar processes.

3.1. Wavelet Decomposition of Self-Similar Processes

Let us assume that $(-\Delta)^{\gamma/2}s = w$ with $\gamma \geq 0$, $(\gamma - 2) \notin \mathbb{N}$, and w a finite-variance and symmetric innovation process. We choose a two-dimensional non-separable wavelet ψ with vanishing moments up to order $\lfloor\gamma\rfloor$. Then, Condition (4) is fulfilled and $\phi = \mathrm{I}_\gamma\psi \in L^2(\mathbb{R}^2)$ is such that $\psi = (-\Delta)^{\gamma/2}\phi$. We define

$$\psi_{a,\mathbf{r}_0}(\mathbf{r}) = \frac{1}{a}\psi\left(\frac{\mathbf{r}}{a} - \mathbf{r}_0\right) \quad (8)$$

for all $a > 0$ and $\mathbf{r}_0 \in \mathbb{R}^2$ as the shifted and dilated version of ψ. Note that $\|\psi_{a,\mathbf{r}_0}\|_2 = \|\psi\|_2$. The wavelet coefficients $\langle s, \psi_{a,\mathbf{r}_0}\rangle$ of the process s are random variables.

We now specify the characteristic function and the pdf of the wavelet coefficients.

Theorem 1. *Under the previous assumptions on s and ψ with $\psi = (-\Delta)^{\gamma/2}\phi$, we have that*

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{j}\omega\langle s,\psi_{a,\mathbf{r}_0}\rangle}\right] = \exp\left(a^2 \int_{\mathbb{R}^2} f(\omega a^{\gamma-1}\phi(\mathbf{r}))\mathrm{d}\mathbf{r}\right). \quad (9)$$

Moreover, the pdf of $\langle s, \psi_{a,\mathbf{r}_0}\rangle$ does not depend on $\mathbf{r}_0$ and can be computed via

$$p_a(x) = \mathcal{F}^{-1}\left\{\mathrm{e}^{a^2 f_\phi(a^{\gamma-1}\omega)}\right\}(x), \quad (10)$$

where $f_\phi(\omega) = \int_{\mathbb{R}^2} f\left(\omega\phi(\mathbf{r})\right)\mathrm{d}\mathbf{r}$ is a valid Lévy exponent.

Proof. Using the shift-invariance and the $(-\gamma)$-homogeneity of I_γ we have $\mathrm{I}_\gamma\{\psi_{a,\mathbf{r}_0}\} = a^{\gamma-1}\mathrm{I}_\gamma\{\psi\}\left(\frac{\cdot}{a} - \mathbf{r}_0\right)$ or, equivalently

$$\psi_{a,\mathbf{r}_0} = (-\Delta)^{\gamma/2}\left\{a^{\gamma-1}\phi\left(\frac{\cdot}{a} - \mathbf{r}_0\right)\right\}. \quad (11)$$

The same calculation as for equation (7) gives $\langle s, \psi_{a,\mathbf{r}_0}\rangle = \langle w, a^{\gamma-1}\phi\left(\frac{\cdot}{a} - \mathbf{r}_0\right)\rangle$. Injecting this in (6), we obtain the desired result after the substitution $\mathbf{r} \leftarrow \left(\frac{\mathbf{r}}{a} - \mathbf{r}_0\right)$ in the integral. The pdf can be directly derived from here. The fact that f_ϕ is an admissible Lévy exponent was shown in [4]. □

One fundamental implication of Theorem 1 is that the statistics of $\langle s, \psi_{a,\mathbf{r}_0}\rangle$ do not depend on the translation factor $\mathbf{r}_0$. Since we limit ourselves to first-order statistics within this work, we shall drop the shift parameter $\mathbf{r}_0$ in the remaining sections.

3.2. Statistics of Wavelet Coefficients at Coarse Scales

We are now interested in the statistical behaviour of the wavelet coefficients when $a \to +\infty$, which corresponds to coarser scales. The next theorem states that these are asymptotically Gaussian even with a sparse innovation process.

Theorem 2. *Under the conditions of Theorem 1, we have*

$$\frac{\langle s, \psi_a\rangle}{a^\gamma} \xrightarrow[a\to+\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2\|\phi\|_2^2), \quad (12)$$

where $\mathcal{L}$ denotes the convergence in law.

The proof requires technical developments and is omitted in the interest of space. It essentially consists in proving the pointwise convergence $\log\left(\mathbb{E}\left[\mathrm{e}^{\mathrm{j}\omega\langle s,\psi_a\rangle/a^\gamma}\right]\right) \xrightarrow[a\to+\infty]{} -\frac{\sigma^2\|\phi\|_2^2\omega^2}{2}$.

Fundamentally, Theorem 2 formalizes the intuitive expectation of observing a Gaussian behaviour (up to an adapted normalization) at coarser scales, since there the wavelet function gets wider and combines more data additively.

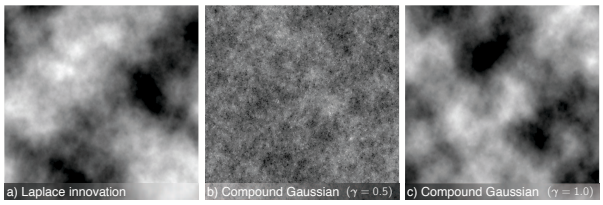

Fig. 1. Illustration of a single realisation of self-similar processes used in the experiments.

3.3. Cumulants of Wavelet Coefficients

Theorem 2 ensures that the coarser wavelet coefficients are becoming Gaussian. It is also interesting to quantify the level of Gaussianity across scales. To do so, we retain the cumulants as our indicator of Gaussianity. We provide here the theoretical prediction for the evolution of the cumulants across scales, with a special emphasis on the variance and the kurtosis.

Proposition 1. *Under the assumptions of Theorem 2, the following results hold.*

- *The nth cumulant $\kappa_n(a)$ of $\langle s, \psi_a\rangle$ exists if $\mathbb{E}\left[|\langle s, \varphi\rangle|^n\right]$ is finite for all $\varphi \in \mathcal{S}(\mathbb{R}^2)$.*
- *Assuming the existence of $\kappa_n(a)$, we have that*

$$\kappa_n(a) = (-\mathrm{j})^n f_\phi^{(n)}(0) a^{n(\gamma-1)+2}, \quad (13)$$

  where f_ϕ is defined in Theorem 1.
- *Assuming the existence of $\kappa_2(a)$ and $\kappa_4(a)$, the variance $\sigma^2(a) = \kappa_2(a)$ and the kurtosis $\eta(a) = \frac{\kappa_4(a)}{\kappa_2(a)^2}$ of the wavelet coefficients are given by*

$$\sigma^2(a) = -f_\phi^{(2)}(0)a^{2\gamma}, \quad (14)$$

$$\eta(a) = \frac{f_\phi^{(4)}(0)}{f_\phi^{(2)}(0)}a^{-2}. \quad (15)$$

Proof. The condition of existence of the cumulants is demonstrated in [4] and is related to the differentiability of $f(\cdot)$ at 0. We rewrite (9) as $\log(\Phi_a(\omega)) = a^2 f_\phi(a^{\gamma-1}\omega)$, where Φ_a denotes the characteristic function of $\langle s, \psi_a\rangle$. The cumulants, if they exist, give the Taylor expansion of $\log(\Phi_a)$ at 0, that is, $\log(\Phi_a(\omega)) = \sum_{n\geq 1}\kappa_n(a)\frac{(\mathrm{j}\omega)^n}{n!}$. Then, (13) is derived from the identification of the two Taylor expansions. Finally, (14) and (15) are clear consequences of (13). □

In the Gaussian case, we have that $f_\phi(\omega) = -\sigma^2\|\phi\|_2^2\omega^2$ and then $f_\phi(0) = 0$ for $n \geq 3$. Proposition 1 implies that the nth cumulant (if it exists) of the random variable $\frac{\langle s,\psi_a\rangle}{a^\gamma}$ is $\frac{\kappa_n(a)}{a^{n\gamma}} \propto a^{2-n}$. In the limit, we have that $a^{2-n} \xrightarrow[a\to\infty]{} 0$ for $n \geq 3$. This is consistent with Theorem 2 in the sense that the asymptotic cumulants are equal to the Gaussian ones, so that, $\kappa_n = 0$ for $n > 2$.

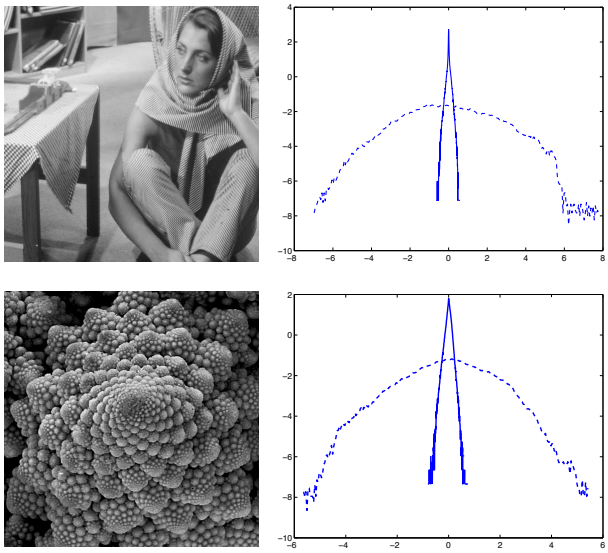

Fig. 2. Normalised histograms (in log scale) of the wavelet coefficients at fine (solid) and coarse scales (dashed) for Barbara and Broccoli.

Table 1. Estimation of the parameter γ.

Innovation	γ	γ_{est}
Laplace	1	0.995
Compound Poisson	1	0.994
Compound Poisson	0.5	0.494

4. NUMERICAL RESULTS

We now let experimental results corroborate our theoretical findings. In order to satisfy the requirements of our analysis, we use the isotropic polyharmonic wavelets developed in [12], with a sufficiently large order. These wavelets enjoy a fast FFT-based implementation. We use a quincunx subsampling scheme that involves a single quasi-isotropic mother wavelet.

To test our predictions, we generate three different self-similar processes, each being realised on a (512×512) grid. For the first process, we consider a Laplace innovation whereas a compound Poisson innovation is used for the two others. We compute the variance of the wavelet coefficients at each scale (averaged over 100 realisations), from which we regress the slope. The empirical results are then compared against the theoretical values (see (14)). To validate the Gaussian behaviour, we compute the normalised kurtosis at each scale. We also test our predictive framework on natural images of size (512×512) (see Figure 2). To have enough samples for conducting statistical estimation, we dilate the wavelet function at every scale as the image size is kept fixed.

In Table 1, we provide the estimated γ values. The fact that are very close to the theoretical ones confirms the validity of our generalised result for evolution of the variance of the non-Gaussian innovations. Also, we see from Figure 3 that the kurtosis curves converge to 0, as predicted by (15), including the ones of the natural images used in the experiments. We note that the behaviour of the normalised kurtosis $\eta(a)/\eta(1)$ does not depend on the innovation process and the γ parameter, as formalised by our theory. By looking at the histograms given in Figure 2, we confirm the Gaussiannity of the coarse-scale coefficients in a qualitative manner. For further validation, we fit a generalised Gaussian distribution to the wavelet coefficients and obtain the shape parameter (exponent p) via a maximum-likelihood estimation. We notice that the shape parameter goes to 2, which is consistent with the convergence of the kurtosis to 0.

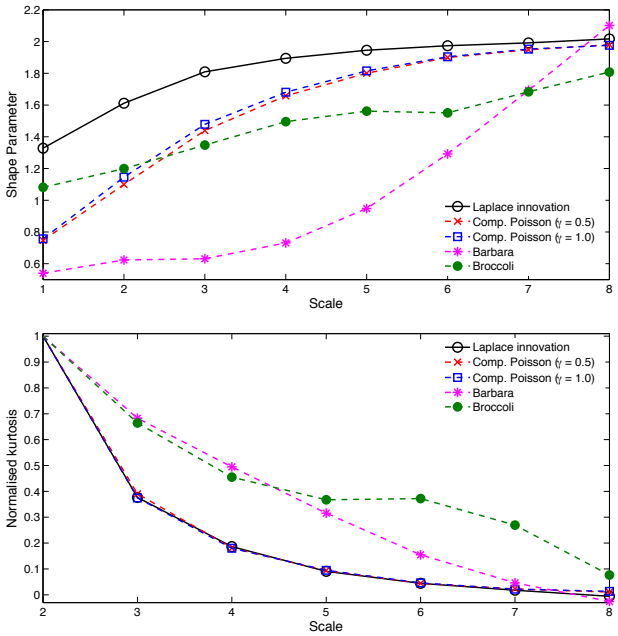

Fig. 3. Evolution of the shape parameter and the kurtosis.

In other words, a Gaussian behaviour is confirmed both for the simulated processes and for the natural images. This suggests that our stochastic self-similar models can provide predictive insights for particular types of natural images.

5. CONCLUSION

Our purpose in this work has been to derive theoretical predictions for the evolution of the statistics of wavelet coefficients. We proved an asymptotic Gaussian behaviour at coarser scales. We based our investigation on a continuous-domain and stochastic model adapted for sparse and self-similar signals. Our experiments showed that our framework is in agreement with observations for certain images.

6. REFERENCES

[1] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 2008.

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

[3] E. Bostan, U. Kamilov, M. Nilchian, and M. Unser, "Sparse stochastic processes and discretization of linear inverse problems," *IEEE Transactions on Image Processing*, vol. 22, no. 7, pp. 2699–2710, 2013.

[4] M. Unser and P. Tafti, *An introduction to sparse stochastic processes*. Cambridge Univerity Press, in press.

[5] E. Bostan, J. Fageot, U. Kamilov, and M. Unser, "MAP estimators for self-similar sparse stochastic models," in *Proceedings of the Tenth International Workshop on Sampling Theory and Applications (SampTA'13)*, Bremen, Federal Republic of Germany, July 1-5, 2013, pp. 197–199.

[6] B.P.-Popescu and J. Vehel, "Stochastic fractal models for image processing," *IEEE Signal Processing Magazine*, vol. 19, no. 5, pp. 48–62, 2002.

[7] D. Mumford and B. Gidas, "Stochastic models for generic images," *Quarterly of Applied Mathematics*, vol. 59, no. 1, pp. 85–112, 2001.

[8] P. Tafti, D. Van De Ville, and M. Unser, "Invariances, Laplacian-like wavelet bases, and the whitening of fractal processes," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 689–702, 2009.

[9] J. Fageot, A. Amini, and M. Unser, "On the continuity of characteristic functionals and sparse stochastic modeling," *arXiv:1401.6850v2*, 2014.

[10] 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, 2012.

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

[12] D. Van De Ville, T. Blu, and M. Unser, "Isotropic polyharmonic B-splines: Scaling functions and wavelets," *IEEE Transactions on Image Processing*, vol. 14, no. 11, pp. 1798–1813, 2005.