

# Stochastic models for sparse and piecewise-smooth signals

Michael Unser, *Fellow, IEEE*, Pouya Dehghani Tafti *Student Member, IEEE*,

## Abstract

We introduce an extended family of continuous-domain stochastic models for sparse, piecewise-smooth signals. These are specified as solutions of stochastic differential equations, or, equivalently, in terms of a suitable innovation model; this is analogous conceptually to the classical interpretation of a Gaussian stationary process as filtered white noise. The two specific ingredients of our approach are: (1) signal generation is driven by a random stream of Dirac impulses (Poisson noise) instead of Gaussian white noise, and (2) the class of admissible whitening operators is considerably larger than what is allowed in the conventional theory of stationary processes. We provide a complete characterization of these finite-rate-of-innovation signals within Gelfand's framework of generalized stochastic processes. We then focus on the class of scale-invariant whitening operators which correspond to unstable systems. We show that these can be solved by introducing proper boundary conditions, which leads to the specification of random, spline-type signals that are piecewise-smooth. These processes are the Poisson counterpart of fractional Brownian motion; they are non-stationary and have the same  $1/\omega$ -type spectral signature. We prove that the generalized Poisson processes have a sparse representation in a wavelet-like basis subject to some mild matching condition. We also present a limit example of sparse process that yields a MAP signal estimator that is equivalent to the popular TV-denoising algorithm.

## I. INTRODUCTION

The hypotheses of Gaussianity and stationarity play a central role in the standard, textbook formulation of signal processing [1], [2]. They fully justify the use of the Fourier transform—as the optimal signal representation—and naturally lead to the derivation of optimal linear filtering algorithms for a large variety of statistical estimation tasks. The Gaussian world of signal processing and its related linear textbook material is elegant and reassuring, but it has reached its limits—it is not at the forefront of research anymore.

Starting with the discovery of the wavelet transform in the late 80s [3], [4], researchers in signal processing have progressively moved away from the Fourier transform and have uncovered powerful alternatives. Two examples of success are the wavelet-based JPEG-2000 standard for image compression [5], which outperforms the widely-used DCT-based JPEG method, and wavelet-domain image denoising which provides a good alternative to more traditional linear filtering [6]–[8]. The key property that makes these techniques work is that many naturally-occurring signals and images—in particular, the ones that are piecewise-smooth—have a sparse representation in the wavelet domain [9]. The concept of sparsity has been systematized and extended to other transforms, including redundant representations (a.k.a. frames); it is at the heart of recent developments in

The authors are with the Biomedical Imaging Group (BIG), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.

signal processing. Sparse signals are easy to compress; likewise, they can be denoised effectively by simple pointwise processing in the transform domain (the rationale being to discard small coefficients which are more likely to be noise). Sparsity provides an equally powerful framework for dealing with more difficult, ill-posed signal reconstruction problems [10], [11]. The strategy there is as follows: among the multitude of solutions that are consistent with the measurements, one should favor the "sparsest" one; that is, the one for which the  $\ell_0$ -norm of the expansion coefficients of the signal is minimum. In practice, one replaces the underlying  $\ell_0$ -norm minimization problem, which is NP hard, by a convex  $\ell_1$ -norm minimization which is computationally much more tractable. Remarkably, researchers have shown that the latter simplification of the problem does yield the correct solution to the problem under suitable conditions (e.g. restricted isometry or incoherent measurements) [12], [13]. This turns out to be one of the leading ideas behind the theory of compressed sensing that deals with the problem of the reconstruction of a signal from a minimal, but suitably chosen, set of measurements [10], [11], [14]. Another approach for breaking the traditional Nyquist's sampling barrier is to take advantage of specific knowledge of the form of the signal and to approach the signal reconstruction task as a parametric estimation problem. Vetterli and co-workers introduced the concept of signals with a finite rate of innovation (FRI) (the prototypical example being a stream of Dirac impulses with unknown locations and amplitudes) and demonstrated the possibility of recovering such signals from a set of uniform measurements at twice the "innovation rate", rather than twice the bandwidth [15]–[17].

The current formulations of compressed sensing and sparse signal recovery are based on solid variational principles, but they are fundamentally deterministic. By drawing on the analogy with the classical theory of signal processing, there are chances that further progress may be achieved via the investigation of stochastic processes that are the "sparse" counterparts of the stationary Gaussian ones. Ideally, the availability of such models would allow for: 1) the derivation of (near)-optimal signal representations for certain classes of signals, 2) the specification of signal-recovery procedures that are well-founded statistically, and 3) (near)-optimal sampling strategies and/or feature-extraction methods. Our goal in this paper is to set the theoretical foundation for such an approach by specifying an extended family of stochastic models that fulfills the following requirements:

- *Continuous-domain formulation.* The proper interpretation of qualifying terms such as "piecewise-smooth" and "scale-invariance", which is central to wavelet theory, calls for continuous-domain models of signals that are compatible with the notion of sparsity.
- *Beyond Gaussian statistics.* The statistical justification of nonlinear algorithms requires non-Gaussian models. However, moving in this direction is not trivial because: (1) decorrelation is no longer synonymous with independence, and (2) non-Gaussian distribution laws are generally not preserved under linear transformation.
- *Backward compatibility.* The formulation should be compatible with the classical theory of Gaussian stationary processes. In particular, the generation mechanism should provide a full control of the second-order statistics (autocorrelation/power spectrum) of the sparse signals so that the classical MMSE filtering, estimation and identification techniques remain applicable.

Our approach builds upon Vetterli et al.'s concept of signals with finite rate of innovation and provides a complete characterization of stochastic processes with the desired properties. While the rigorous specification

of these processes requires an appropriate mathematical formalism<sup>1</sup>, the payoff is a generative model that is simple conceptually and parallel to the classical white-noise filtering model for Gaussian stationary processes. The primary contributions of this work are as follows:

- An extended innovation model where the usual Gaussian white noise is substituted by impulsive Poisson noise with any prescribed amplitude distribution. The key descriptor of a process is its whitening operator  $L$  which is shift-invariant.
- A complete distributional characterization of such generalized Poisson processes by means of the *characteristic form* which condenses all statistical information [18]. The relevant theoretical framework, which is not standard in the field, is summarized in Appendix I.
- The extension of the traditional spectral shaping filters ( $L^{-1}$ ) to a larger class of inverse operators, including important ones that are unstable<sup>2</sup> in the classical sense. The prototypical example is the integrator which allows the generation of Brownian motion as well as piecewise-constant signals.
- The link with spline theory through a common operator formalism.
- The characterization of the sparsifying effect of the wavelet transform for a wide class of generalized Poisson processes.

The paper is organized as follows. In Section II, we show the relevance of the proposed class of random processes by contrasting the performance of the classical Wiener filter and sparsity-promoting restoration methods (total variation and wavelet denoising) in a denoising experiment that involves a matched pair of Gaussian vs. sparse processes. We then proceed in Section III with the definition of impulsive Poisson noise and the derivation of its characteristic form within the framework of Gelfand and Vilenkin's theory of generalized stochastic processes. In Section IV, we specify our generalized Poisson processes as the solutions of a stochastic differential equation driven by white impulsive noise, which is equivalent to the innovation model in Fig. 3 with whitening operator  $L$ . We then focus on the class of scale-invariant whitening operators and show how these can specify spline-type processes in one or several dimensions. In Section V, we consider the wavelet analysis of generalized Poisson processes, including those of mixed type, and prove that it generally yields a sparse signal decomposition. Finally, we illustrate the use of the proposed statistical formalism with the derivation of the likelihood function of a sparse, piecewise-constant process.

## II. MOTIVATION: BEYOND WIENER FILTERING

To motivate the stochastic models proposed in this paper, we consider the problem of the reconstruction of a continuously-defined signal  $s(x)$  given its noisy samples at the integers:  $\{g[k] = s(k) + n[k]\}_{k \in \mathbb{Z}}$  where  $n[k]$  is a discrete Gaussian white noise with zero mean and variance  $\sigma^2$ . When  $s(x)$  is a realization of a Gaussian stationary process, the minimum-mean-square-error (MMSE) solution to this problem is the well-known Wiener filter. The Wiener filter remains the best *linear* reconstruction algorithm when  $s(x)$  is non-Gaussian, but it is

<sup>1</sup>Impulse Poisson noise (random stream of Dirac impulses) can only be properly defined within the context of distribution theory. The other point is that many of the processes that will be considering here are non-stationary, meaning that they don't have a well-defined power spectrum; they also involve fractional derivative operators which are difficult to handle using conventional stochastic calculus.

<sup>2</sup>A convolution operator is BIBO-stable (bounded-input bounded-output) iff. its impulse response is in  $L_1(\mathbb{R})$ . The integrator is not BIBO-stable, but it is sometimes said to be marginally stable because its impulse response is bounded.

generally not globally optimal anymore. In particular, it has been observed that linear filtering is suboptimal for handling piecewise-smooth signals because it over-smoothes sharp transitions. For such signals, linear algorithms are typically outperformed by simple wavelet thresholding [8], [19], which is a non-linear type of processing.

To demonstrate this behavior, we selected two continuous-time test signals which are part of the class of stochastic processes considered in this paper. The first (Fig. 1a) is a Brownian motion (also known as the Wiener process), which is a classical example of Gaussian process. The second (Fig. 1b), which is piecewise-constant, is a compound Poisson process; the location of the singularities follow a spatial Poisson distribution with parameter  $\lambda$ , while the heights of the transitions are random and uniformly distributed. The important conceptual aspect is that these two signals share a common innovation model, as we shall prove in Section III. They are both whitened by the derivative operator  $D = \frac{d}{dx}$ , the distinction being that the innovation process is white Gaussian noise in the first case, and (sparse) impulsive Poisson noise in the second. Consequently, the two processes have identical second order statistics and they admit the same Wiener filter as the best linear signal estimator. In our earlier work, we have shown that the MMSE reconstruction of a Brownian motion signal corrupted by white Gaussian noise is provided by a piecewise-linear smoothing spline estimator [20]. Remarkably, this smoothing spline estimator  $\tilde{s}$  can also be defined as the solution of the variational problem

$$\begin{aligned} \tilde{s}_{\text{spline}}(x) &= \arg \min_{s(x)} \left\{ \sum_{k \in \mathbb{Z}} |g[k] - s(k)|^2 + \mu \int_{\mathbb{R}} |Ds(x)|^2 dx \right\} \\ &= \arg \min_{s(x)} \left\{ \|g - s\|_{\ell_2}^2 + \mu \|Ds\|_{L_2(\mathbb{R})}^2 \right\} \end{aligned} \quad (1)$$

with  $\mu = \frac{\sigma_0^2}{\sigma_g^2}$  with  $\sigma_0^2 = \text{Var}\{Ds(x)\}$ . Note that the above cost criterion includes a *discrete* data term (squared  $\ell_2$ -norm) and a *continuous* regularization functional (squared  $L_2$ -norm) that penalizes non-smooth solutions. In contrast to conventional digital signal processing, the solution of the minimization problem is continuously-defined: it corresponds to a hybrid form of Wiener filter (discrete input and continuous output).

A more satisfactory handling of the second piecewise-constant signal is based on the observation that it has a sparse decomposition in the Haar basis which is piecewise-constant as well. It therefore makes sense to seek a reconstruction that has few significant wavelet coefficients. This is achieved by introducing an  $\ell_p$ -norm penalty on the wavelet coefficients of  $s(x)$ :  $w_i[k] = \langle s, 2^{i/2} \psi(x/2^i - k) \rangle$  where  $i$  is the scale index and where  $\psi(x)$  is the Haar wavelet. This leads to the wavelet-based signal estimator

$$\tilde{s}_{\text{wave}}(x) = \arg \min_{s(x)} \left\{ \|g - s\|_{\ell_2}^2 + \sum_i \mu_i \|w_i\|_{\ell_p}^p \right\} \quad (2)$$

where  $\{\mu_i\}$  is an appropriate sequence of scale-dependant weights (typically,  $\lambda_0 \cdot 2^{i/2}$  to implement the Besov norm associated with the first derivative of the function [21]). By applying Parseval's identity to the data term and formulating the problem in the wavelet domain, one finds that the solution is obtained by applying a suitable pointwise non-linearity to the wavelet coefficients of the noisy signal [22]. For  $p = 1$ , the non-linearity corresponds to a standard soft-thresholding. Formally, we can also consider the case  $p = 0$ , which yields a sparse solution implemented by discarding all wavelet coefficients below a certain threshold.

Another popular reconstruction/denoising method is to penalize the total variation of the signal [23], which results in the TV estimator

$$\tilde{s}_{\text{TV}}(x) = \arg \min_{s(x)} \left\{ \|g - s\|_{\ell_2}^2 + \mu \text{TV}(s) \right\} \quad (3)$$

where  $\text{TV}(s)$  is the total variation of  $s(x)$ . Note that when  $s$  is differentiable,  $\text{TV}(s) = \int_{\mathbb{R}} |Ds(x)| dx$  so that criterion (3) is essentially the  $L_1$ -regularized counterpart of (1). A remarkable property of (3) is that the global optimum is achieved by a piecewise-constant function [24], which suggests that the TV criterion is ideally matched to our second class of signals.

We implemented the three proposed estimators and applied them to the reconstruction of our two test signals corrupted with various amounts of noise. In each case, we optimized the regularization parameter  $\mu \geq 0$  for maximum signal-to-noise ratio, taking the noise-free signal (oracle) as our reference. The results are summarized in Fig. 2. According to theoretical predictions, the smoothing spline estimator (Wiener filter) performs best in the Gaussian case. In the case of a piecewise-constant signal, the best linear estimator is outperformed by the TV estimator over the entire range of signal-to-noise ratios, and also by wavelet denoising at lower noise levels. The wavelet-based method performs adequately, but is sub-optimal—in fact, the results presented here were obtained by using cycle spinning which is a simple, effective way of boosting the performance of the basic threshold-based wavelet denoising algorithm [25].

This series of experiments confirms the well-documented observation that a change of regularization exponent (i.e.,  $p = 1$  vs.  $p = 2$ ) can have a significant effect on the restoration quality, especially for the second type of signal which is intrinsically sparse. Other than that, the regularization functionals used in the three estimators are qualitatively similar and perfectly matched to the spectral characteristics of the signals under considerations: the whitening operator  $D$  appears explicitly in (1) and (3), while it is also present implicitly in (2). The latter is seen by expressing the Haar wavelet as  $\psi(x) = D\phi(x)$  where the smoothing kernel  $\phi(x) = \frac{1}{2}\beta_+^1(2x)$  is a rescaled causal triangle function (or B-spline of degree 1). This implies that the wavelet coefficients, which can now be interpreted as smoothed derivatives, are statistically independent at any given scale  $i$ —indeed, by duality, we have that  $\langle s, \psi(\cdot - k) \rangle = -\langle Ds, \phi(\cdot - k) \rangle$  where  $Ds$  is a continuously-varying white noise process and where the integer-shifted versions of  $\phi(x)$  are non-overlapping.

While the Gaussian part of the story is well understood, it is much less so for the second class on non-gaussian signals which are more difficult to formalize statistically. Such Poisson processes, however, could be important conceptually because they yield prototypical signals for which some of the current popular signal processing methods (wavelets, compressed sensing,  $\ell_1$ -minimization, etc.) perform best. In the sequel, we will present a generalized distributional framework for the complete stochastic characterization of such processes and their extensions. We will come back to the above denoising problem at the end of the paper and show that the TV denoising solution (3) is compatible with the MAP signal estimator that can be derived for the second process.

### III. GENERALIZED POISSON NOISE

#### A. Classical Poisson processes

A Poisson process is a continuous-time process that is in direct correspondence with a series of independent point events randomly scattered over the real line according to a Poisson distribution with a rate parameter  $\lambda$ . Specifically, the probability of having a number  $N(x, x + T) = n$  of events in the time interval  $(x, x + T]$  with  $T \geq 0$  is

$$\text{Prob} \{N(x, x + T) = n\} = \frac{e^{-\lambda T} (\lambda T)^n}{n!}$$

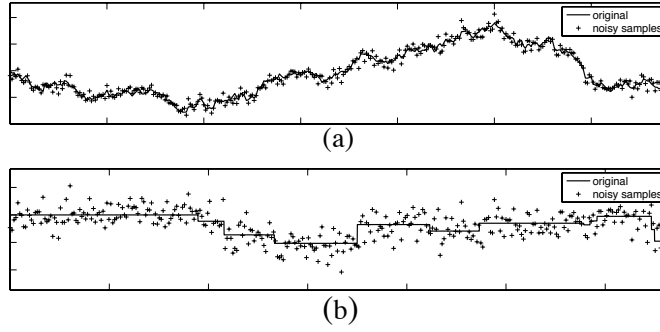


Fig. 1. Original test signals (solid lines) and their discrete noisy measurements. (a) Brownian motion. (b) Compound Poisson process (piecewise-constant signal).

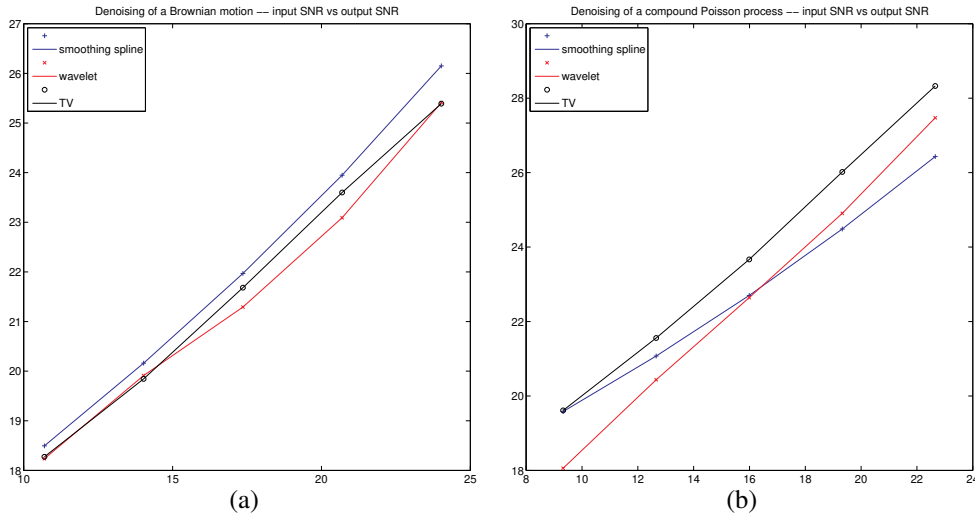


Fig. 2. Comparative evaluation (input SNR vs output SNR) of smoothing spline, wavelet and total variation reconstruction algorithms for a signal corrupted by white Gaussian noise. (a) Brownian motion. (b) Compound Poisson process.

with a mean value and variance given by  $\lambda T$ . The time locations of these events are ordered and denoted by  $x_k$  with  $k \in \mathbb{Z}$ . With this notation, the classical homogeneous Poisson process can be represented as

$$r(x) = \sum_{k \in \mathbb{Z}} 1_+(x - x_k)$$

where  $1_+(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$  is the unit step function. The statistical term “homogeneous” refers to the fact that the rate parameter  $\lambda$  is constant over time; the term will be dropped in the sequel. The realizations of such a Poisson process are piecewise-constant and monotonously increasing with unit increments; the parameter  $\lambda$  represents the expected number of discontinuities per unit interval.

An extended version of this process that is much better suited for modeling FRI signals is the so-called compound Poisson process

$$s(x) = \sum_{k \in \mathbb{Z}} a_k 1_+(x - x_k) \quad (4)$$

where the  $a_k$  are i.i.d. random variables associated with the probability measure  $dP(a)$ . This signal is piecewise-constant in each time interval  $[x_k, x_{k+1})$  and may be thought of as the stochastic version of a non-uniform

spline of degree 0 where both the knots  $(x_k)$  and heights of the piecewise-constant segments are random (cf. Fig. 1b). It is also the primary example of a concrete signal with a finite rate of innovation [15]; in average,  $s(x)$  has two degrees of freedom  $(x_k, a_k)$  per time interval of length  $\lambda^{-1}$ . While the compound Poisson process is clearly non-Gaussian, it has the interesting property of being indistinguishable from Brownian motion based on its second order statistics alone (covariances). It is part of the same class of “ $1/\omega$ ”-type processes, keeping in mind that the power spectrum of such signals is not defined in the conventional sense because of their non-stationary character. In the sequel, we will strengthen the connection between these two classes of processes by linking them to a common innovation model involving a spectral shaping operator and a specific white noise excitation which may or may not be Gaussian.

### B. White Poisson noise

By taking the distributional derivative of  $s(x)$ , we obtain a weighted stream of Dirac impulses  $\sum_{k \in \mathbb{Z}} a_k \delta(x - x_k)$  whose positions and amplitudes are random and independent of each other. In this paper, we will consider a more general multidimensional setting with signals and stochastic processes that are functions of the continuous-domain variable  $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ . We introduce the compound Poisson noise

$$w(\mathbf{x}) = \sum_{k \in \mathbb{Z}} a_k \delta(\mathbf{x} - \mathbf{x}_k) \quad (5)$$

where the  $\mathbf{x}_k$ 's are random point locations in  $\mathbb{R}^d$ , and where the  $a_k$  are i.i.d. random variables with cumulative probability distribution  $P(a)$ . The random events are indexed by  $k$  (using some arbitrary ordering); they are mutually independent and follow a spatial Poisson distribution. Specifically, let  $\Pi$  be any finite-measure subset of  $\mathbb{R}^d$ , then the probability of observing  $N(\Pi) = n$  events in  $\Pi$  is

$$\text{Prob}(N(\Pi) = n) = \frac{e^{-\lambda \text{Vol}(\Pi)} (\lambda \text{Vol}(\Pi))^n}{n!}$$

where  $\text{Vol}(\Pi)$  is the measure (or spatial volume) of  $\Pi$ . This is to say that the Poisson parameter  $\lambda$  represents the average number of random impulses per unit hyper-volume.

While the specification (5) of our compound Poisson noise is explicit and constructive, it is not directly suitable for deriving its stochastic properties. The presence of Dirac impulses makes it difficult to handle such entities using conventional stochastic calculus. Instead of trying to consider the point values of  $w(\mathbf{x})$  which are either zero or infinite, it makes more sense to investigate the (joint) statistical distribution of the scalar products (or linear functionals<sup>3</sup>)  $w(\varphi) := \langle w, \varphi \rangle$  between our Poisson noise and a collection of suitable test functions  $\varphi$ . The adequate mathematical formalism is provided by Gelfand's theory of generalized stochastic processes, whose main results are briefly reviewed in Appendix I. The conceptual foundation of this powerful framework is that a generalized stochastic process  $s$  is “indexed” by  $\varphi$  rather than by the spatial variable  $\mathbf{x}$ . It is thereby possible to fully characterize a real-valued process  $s$  by specifying its characteristic form

$$\mathcal{Z}_s(\varphi) = \mathcal{E}\{e^{-j\langle s, \varphi \rangle}\} \quad (6)$$

<sup>3</sup>The linear functional  $w(\varphi)$  is formally specified by the scalar-product integral  $\langle w, \varphi \rangle = \int_{\mathbb{R}^d} w(\mathbf{x})\varphi(\mathbf{x}) d\mathbf{x}$ . It is a well-defined linear, continuous mapping that associates a scalar to each  $\varphi$  within a suitable set of test functions.

where  $\mathcal{E}\{\cdot\}$  denotes the expectation operator and where  $s(\varphi) = \langle s, \varphi \rangle$  with  $\varphi$  fixed should be treated as a classical scalar random variable.  $\mathcal{Z}_s(\varphi)$  is a functional of the generic test function  $\varphi$  whose role is analogous to that of the index variable(s) used in the conventional definition of the characteristic function of a probability distribution. The powerful aspect of this generalization, which can be traced back to Kolmogoroff [26], is that  $\varphi$  has the ability to capture all the possible joint dependencies of the process. For instance, if we substitute  $\varphi = \omega_1 \varphi_1 + \dots + \omega_N \varphi_N$  in (6), then we obtain  $\mathcal{Z}_s(\varphi) = \mathcal{E}\{e^{-j(\omega_1 s_1 + \dots + \omega_N s_N)}\}$  with  $s_n = \langle s, \varphi_n \rangle$  and the  $\omega_n$ 's taking the role of frequency-domain variables; this is precisely the characteristic function of the  $N$ -vector random variable  $(s_1, \dots, s_N)$ , meaning that the joint probability distribution  $p(s_1, \dots, s_N)$  can be obtained, at least conceptually, by  $N$ -D inverse Fourier transformation. The corresponding distributional extension of the correlation function  $\mathcal{E}\{s(\mathbf{x}_1)s(\mathbf{x}_2)\}$  is the so-called correlation form

$$\mathcal{B}_s(\varphi_1, \varphi_2) = \mathcal{E}\{\langle s, \varphi_1 \rangle \cdot \langle s, \varphi_2 \rangle\} = \mathcal{E}\{s(\varphi_1) \cdot s(\varphi_2)\} \quad (7)$$

which can also be deduced from  $\mathcal{Z}_s(\varphi)$ . Clearly, (7) reverts to the classical correlation function if we formally substitute  $\varphi_1 = \delta(\cdot - \mathbf{x}_1)$  and  $\varphi_2 = \delta(\cdot - \mathbf{x}_2)$ .

In order to take advantage of this formalism, we need to obtain the characteristic form of the Poisson noise defined above. Before presenting these results, we introduce some notations and conventions:

- The integration element over  $\mathbb{R}^d$  is denoted by  $d\mathbf{x}$  with  $d\mathbf{x} = dx_1 \cdots dx_d$
- The amplitude statistics of the Poisson process are expressed in terms of the cumulative probability distribution  $P(a) = \int_{-\infty}^a dP(a)$  where  $dP(a)$  is the underlying probability measure. The  $p$ -order moment of the amplitude distribution is denoted by  $\mathcal{E}\{a^p\}$  with  $\mathcal{E}\{a^p\} = \int_{\mathbb{R}} a^p dP(a)$ .
- The Fourier transform of a "test" function  $\varphi \in \mathcal{S}(\mathbb{R}^d)$  is  $\hat{\varphi}$  with  $\hat{\varphi}(\boldsymbol{\omega}) = \int_{\mathbb{R}^d} \varphi(\mathbf{x}) e^{-j\langle \mathbf{x}, \boldsymbol{\omega} \rangle} d\mathbf{x}$ . A fundamental property is that the Fourier transform is a self-reversible mapping from  $\mathcal{S}(\mathbb{R}^d)$  (Schwartz's class of smooth and rapidly decreasing functions) into itself.
- $\|\varphi\|_{L_p(\mathbb{R}^d)}$  with  $1 \leq p \leq +\infty$  stands for the  $L_p$ -norm of  $\varphi$ ; it is bounded for all test functions.

*Theorem 1:* The characteristic form of the impulsive Poisson noise specified by (5) is

$$\mathcal{Z}_w(\varphi) = \mathcal{E}\{e^{-j\langle w, \varphi \rangle}\} = e^{u_{\lambda, P}(\varphi)} \quad (8)$$

with

$$u_{\lambda, P}(\varphi) = \lambda \int_{\mathbb{R}} \int_{\mathbb{R}^d} (e^{ja\varphi(\mathbf{x})} - 1) d\mathbf{x} dP(a); \quad (9)$$

where  $\lambda$  is the Poisson density parameter, and where  $P(a)$  is the cumulative amplitude probability distribution. The proof is given in Appendix II. Note that the above characteristic form is part of a more general family of functionals that is derived by Gelfand and Vilenkin starting from first principles (processes with independent values at every point, infinite divisibility) [18]. Here, we make the link between the abstract characterization of such processes and Eq. (5), which provides a concrete generation mechanism. A direct consequence of Theorem 1 is that the impulsive Poisson process is a bona fide white noise, albeit a not a Gaussian one.

*Corollary 1:* The correlation form of the generalized Poisson process defined by (8) is

$$B_w(\varphi_1, \varphi_2) = \lambda \mathcal{E}\{a^2\} \cdot \langle \varphi_1, \varphi_2 \rangle_{L_2(\mathbb{R}^d)} + \lambda^2 \mathcal{E}\{a\}^2 \hat{\varphi}_1(\mathbf{0}) \hat{\varphi}_2(\mathbf{0}).$$



Hence,  $w$  is a (non-Gaussian) white noise process with variance  $\sigma_w^2 = \lambda \mathcal{E}\{a^2\}$  provided that the random variable  $a$  has zero-mean.

*Proof:* We rely on (22) in Appendix I and partially differentiate (8) by applying the chain rule twice:

$$\frac{\partial^2 \mathcal{Z}_w(\omega_1 \varphi_1 + \omega_2 \varphi_2)}{\partial \omega_1 \partial \omega_2} = e^{u_{\lambda, P}} \left( \frac{\partial u_{\lambda, P}}{\partial \omega_1} \frac{\partial u_{\lambda, P}}{\partial \omega_2} + \frac{\partial^2 u_{\lambda, P}}{\partial \omega_1 \partial \omega_2} \right). \quad (10)$$

The required first derivative with respect to  $\omega_i, i = 1, 2$  is given by

$$\begin{aligned} \frac{\partial u_{\lambda, P}}{\partial \omega_i} &= \lambda \int_{\mathbb{R}} \int_{\mathbb{R}^d} \frac{\partial}{\partial \omega_i} (e^{ja(\omega_1 \varphi_1 + \omega_2 \varphi_2)} - 1) \, d\mathbf{x} \, dP(a) \\ &= \lambda \int_{\mathbb{R}} \int_{\mathbb{R}^d} (ja) \varphi_i(\mathbf{x}) e^{ja(\omega_1 \varphi_1 + \omega_2 \varphi_2)} \, d\mathbf{x} \, dP(a) \end{aligned}$$

which, when evaluated at the origin, simplifies to

$$\begin{aligned} \left. \frac{\partial u_{\lambda, P}}{\partial \omega_i} \right|_{\boldsymbol{\omega}=\mathbf{0}} &= \lambda j \int_{\mathbb{R}} a \, dP(a) \int_{\mathbb{R}^d} \varphi_i(\mathbf{x}) \, d\mathbf{x} \\ &= j \lambda \mathcal{E}\{a\} \hat{\varphi}_i(\mathbf{0}). \end{aligned}$$

Similarly, we get

$$\begin{aligned} \left. \frac{\partial^2 u_{\lambda, P}}{\partial \omega_1 \partial \omega_2} \right|_{\boldsymbol{\omega}=\mathbf{0}} &= \lambda (j)^2 \int_{\mathbb{R}} a^2 \, dP(a) \int_{\mathbb{R}^d} \varphi_1(\mathbf{x}) \varphi_2(\mathbf{x}) \, d\mathbf{x} \\ &= -\lambda \mathcal{E}\{a^2\} \langle \varphi_1, \varphi_2 \rangle_{L_2(\mathbb{R}^d)} \end{aligned}$$

The result then follows from (22) and the substitution of these expressions in (10).  $\blacksquare$

When the density distribution  $p(a)$  is symmetrical with respect to the origin, the Poisson functional  $u_{\lambda, P}$  takes the simplified form

$$u_{\lambda, P}(\varphi) = \lambda \int_{\mathbb{R}^d} \int_{\mathbb{R}} (\cos(a\varphi(\mathbf{x})) - 1) \, dP(a) \, d\mathbf{x}, \quad (11)$$

due to the cancellation<sup>4</sup> of imaginary terms. We will refer to this case as *symmetric Poisson noise*.

We conclude this section by presenting an expression for the characteristic form of symmetric Poisson noise that brings out the differences with the standard form of a Gaussian white noise and also gives us a better insight into the influence of the amplitude variable  $a$ . To this end, we write the Taylor series of (11) and manipulate it as follows

$$\begin{aligned} u_{\lambda, P}(\varphi) &= \lambda \int_{\mathbb{R}^d} \int_{\mathbb{R}} \sum_{n=1}^{+\infty} \frac{(-1)^n}{(2n)!} a^{2n} \varphi(\mathbf{x})^{2n} \, dP(a) \, d\mathbf{x} \\ &= \lambda \sum_{n=1}^{+\infty} \frac{(-1)^n}{(2n)!} \mathcal{E}\{a^{2n}\} \int_{\mathbb{R}^d} \varphi(\mathbf{x})^{2n} \, d\mathbf{x} \\ &= -\frac{\lambda}{2!} \mathcal{E}\{a^2\} \|\varphi\|_{L_2}^2 + \frac{\lambda}{4!} \mathcal{E}\{a^4\} \|\varphi\|_{L_4}^4 - \dots \end{aligned}$$

where we are assuming that the moments of  $P(a)$  are bounded in order to switch the order of summation. The final expression is enlightening since the first term, which is purely quadratic, precisely matches the standard Gaussian form  $u(\varphi) = \frac{1}{2} \|\varphi\|_{L_2(\mathbb{R}^d)}^2$  [cf. (20)]. This is consistent with the fact that the second-order properties of the process are indistinguishable from those of a Gaussian noise (cf. Corollary 1). The interesting aspect is

<sup>4</sup>This requires the interchange of the order of integration, which is justified by Lebesgue's dominated convergence theorem; specifically, we invoke the bounds  $|\cos(u) - 1| \leq \min(2, u^2)$  and  $|\sin(u)| \leq \min(1, |u|)$  together with the fact the test function  $\varphi(\mathbf{x})$  decays rapidly at infinity.

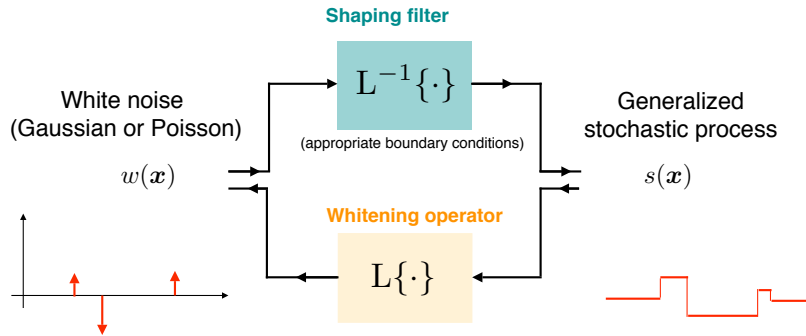


Fig. 3. Innovation model of a generalized stochastic process. The delicate mathematical issue is to make sure that the operator  $L$  and its inverse  $L^{-1}$  (resp., their duals  $L^*$  and  $L^{-1*}$ ) are well-defined over the space of tempered distributions (resp., Schwartz's class of infinitely differentiable and rapidly-decreasing test functions).

that the Poisson functional also includes higher-order terms involving the  $L_p$ -norms of  $\varphi$  for  $p$  even with the  $p$ th moments of the amplitude distribution acting as weighting factors. This last formula also shows that we have some freedom in shaping the Poisson noise via the control of the higher-order moments of  $a$ .

#### IV. GENERALIZED POISSON PROCESSES

Our quest in this paper is to specify stochastic models for the class of piecewise-smooth signals that are well represented by wavelets and that also conform to the notion of finite-rate of innovation. To maintain the backward compatibility with the classical Gaussian formulation, we are proposing the common innovation model in Fig. 3 driven by white noise  $w$ , which may be either Gaussian or impulsive Poisson. The remarkable feature is that the Poisson version of the model is capable of generating piecewise-smooth signals, in direct analogy with the method of constructing splines that is reviewed in Section IV-A. This naturally leads to the definition of generalized Poisson processes given in Section IV-B, with the catch that the underlying stochastic differential equations are typically unstable. In Section IV-C, we show how to bypass this difficulty via the specification of appropriate scale-invariant inverse operators. We then illustrate the approach by presenting concrete examples of sparse processes (Sections IV-D to IV-E).

##### A. The spline connection

Splines provide a convenient framework for modeling 1-D piecewise-smooth functions. They can be made quite general by allowing for non-uniform knots and different types of building blocks (e.g. piecewise polynomials or exponentials) [27]. An elegant, unifying formulation associates each brand of splines with a particular linear differential operator  $L$ . Here, we will assume that  $L$  is shift-invariant and that its null space is finite-dimensional and non-trivial. Its Green function (not unique) will be denoted by  $\rho_L(x)$  with the defining property that  $L\{\rho_L\} = \delta$ .

*Definition 1:* A function  $s(x)$  is a non-uniform  $L$ -spline with knot sequence  $x_1 < \dots < x_k < \dots < x_K$  iff.

$$L\{s\}(x) = \sum_{k=1}^K a_k \delta(x - x_k).$$

The knot points  $x_k$  correspond to the spline singularities. Interestingly, the Dirac impulse sequence on the right-hand side of this equation is essentially the same as the one used to define the Poisson noise in (5), with the important difference that it is now a deterministic entity.

We can formally integrate the above equation and obtain an explicit representation of the non-uniform spline as a linear combination of shifted Green functions plus a component  $p_0(x)$  that is in the null space of  $L$ :

$$s(x) = p_0(x) + \sum_{k=1}^K a_k \rho_L(x - x_k).$$

For the spline to be uniquely-defined, one also needs to specify some boundary conditions to fix the null-space component  $p_0(x)$  (typically,  $m$  linear constraints for a differential operator of order  $m$ ). The standard choice of a differential operator is  $L = D^m = \frac{d^m}{dx^m}$  which corresponds to the family of polynomial splines of degree  $m - 1$ . The generic form of such splines is

$$s_m(x) = \sum_{n=0}^{m-1} b_n x^n + \sum_{k=1}^K a_k \frac{(x - x_k)_+^{m-1}}{(m-1)!}$$

where the one-sided power function  $\frac{(x)_+^{m-1}}{(m-1)!}$  is the causal Green function of  $D^m$ , or, equivalently, the impulse response of the  $m$ -fold integrator  $D^{-m}$ . One can verify that  $s_m(x)$  coincides with a polynomial of degree  $m - 1$  in each interval  $[x_k, x_{k+1})$  and that it is differentiable up to order  $m - 1$  at the knot locations, implying that the polynomial segments are smoothly joined together.

An equivalent higher-level description of the above inversion process is to view our spline as the solution of the differential equation  $L\{s\} = r$  with driving term  $r(x) = \sum_k a_k \delta(x - x_k)$  and to express the solution as  $s = L^{-1}\{r\}$  where  $L^{-1}$  is an appropriate inverse operator that incorporates the desired boundary conditions. The mathematical difficulty in this formulation is that it requires a precise, unambiguous specification of  $L^{-1}$ . This is the approach that we will take here to define our generalized Poisson processes. Intuitively, these correspond to stochastic splines where both the weights and knot locations are random.

### B. Generalized processes with whitening operator $L$

Let us now return to the innovation model in Fig. 3. The idea is to define the generalized process  $s$  with whitening operator  $L$  as the solution of the stochastic partial differential equation (PDE)

$$L\{s\}(\mathbf{x}) = w(\mathbf{x}) \tag{12}$$

where the driving term  $w$  is a white noise process that is either Gaussian or Poisson (or possibly a combination of both). This definition is obviously only usable if we can specify a corresponding inverse operator  $T = L^{-1}$ ; in the case where the inverse is not unique, we will need to select one preferential operator, which is equivalent to imposing specific boundary conditions. Assuming that such an operator exists and that its adjoint  $L^{-1*}$  is mathematically well-defined on the chosen family of test functions, we are then able to formally solve the equation as

$$s(\mathbf{x}) = L^{-1}\{w\}(\mathbf{x}).$$

Moreover, based on the defining property  $\langle s, \varphi \rangle = \langle L^{-1}w, \varphi \rangle = \langle w, L^{-1*}\varphi \rangle$ , we can transfer the action of the operator onto the test function inside the characteristic form (cf. Appendix I.B) and obtain a complete statistical

characterization of the so-defined generalized stochastic process

$$\mathcal{Z}_s(\varphi) = \mathcal{Z}_{L^{-1}w}(\varphi) = \mathcal{Z}_w(L^{-1*}\varphi).$$

where  $\mathcal{Z}_w(\varphi)$  is specified by (20) or Theorem 1, depending on the type of driving noise. This simple manipulation yields the following explicit formulas for the characteristic forms of our two kinds of processes:

1) Generalized Gaussian

$$\mathcal{Z}_s(\varphi) = e^{-\frac{1}{2}\|L^{-1*}\varphi\|_{L_2(\mathbb{R}^d)}^2} \quad (13)$$

2) Generalized Poisson

$$\mathcal{Z}_s(\varphi) = \exp\left(\lambda \int_{\mathbb{R}} \int_{\mathbb{R}^d} (e^{jaL^{-1*}\{\varphi\}(\mathbf{x})} - 1) d\mathbf{x} dP(a)\right) \quad (14)$$

The correlation form, which is the same in both cases, is

$$\mathcal{B}_s(\varphi_1, \varphi_2) = \langle L^{-1*}\varphi_1, L^{-1*}\varphi_2 \rangle_{L_2} = \langle (L^{-1}L^{-1*})\varphi_1, \varphi_2 \rangle_{L_2}.$$

The latter implies that the Gaussian and sparse processes defined by (13) and (14) have identical second-order statistics, which is the matching condition emphasized in our introductory denoising experiment.

The above characterization is not only remarkably concise, but also quite general for it can handle a much larger class of linear operators than conventional stochastic calculus. This will prove to be very helpful for our investigation of spline-type processes in Section III.D.

In the special case where  $w$  is Gaussian and  $T = L^{-1}$  is a shift-invariant operator such that  $Ts = (h * w)$  where  $h \in L_2(\mathbb{R}^d)$  is a suitable, square-integrable convolution kernel, one recovers the classical family of Gaussian stationary processes with spectral power density  $\Phi_s(\omega) = |H(\omega)|^2$  where  $H(\omega)$  is the frequency response of the shaping filter  $h$  (cf. Appendix I-C). The corresponding autocorrelation function is given by  $\mathcal{B}_s(\delta, \delta(\cdot - \mathbf{x})) = (h^\vee * h)(\mathbf{x})$  with  $h^\vee(\mathbf{x}) = h(-\mathbf{x})$ , which is consistent with the Wiener-Kintchine theorem. If one switches to a Poisson excitation, one obtains a class of stationary random processes sometimes referred to as generalized shot noises [28], [29]. These signals are made up of shifted replicas of the impulse response of the shaping filter with some random amplitude factor:  $s(\mathbf{x}) = \sum_{k \in \mathbb{Z}} a_k h(\mathbf{x} - \mathbf{x}_k)$ . They are typically bumpy (depending on the localization properties of  $h$ ) and quite distinct from what one would commonly call a spline. Generating random splines is possible as well, but these will typically not be stationary.

### C. Scale-invariant operators and their inverse

Among the large variety of admissible operators  $L$ , we are especially interested in those that commute with the primary coordinate transformations: translation, dilation and rotation. This is because they are likely to generate processes with interesting properties. These operators, which are also tightly linked to splines and fractals, happen to be fractional derivatives.

We have shown in earlier work that the class of linear 1-D shift- and scale-invariant operators reduces to the  $(\gamma, \tau)$ -derivatives  $\partial_\tau^\gamma$  with  $\tau \in \mathbb{R}$  and  $\gamma \in \mathbb{R}+$  [30]. Their Fourier-domain definition is

$$\partial_\tau^\gamma f(x) \xleftrightarrow{\mathcal{F}} (-j\omega)^{\frac{\gamma}{2}-\tau} (j\omega)^{\frac{\gamma}{2}+\tau} \hat{f}(\omega)$$

where  $\hat{f}$  is the Fourier transform of  $f$  (in the sense of distributions). The parameter  $\tau$  is a phase factor that allows for a progressive transition between a causal operator ( $\tau = \frac{\gamma}{2}$ ) and an anti-causal one ( $\tau = -\frac{\gamma}{2}$ ), which is

the adjoint of the former (more generally, we have that  $(\partial_\tau^\gamma)^* = \partial_{-\tau}^\gamma$ ). Note that the causal fractional derivative  $\partial_{\frac{\gamma}{2}}^\gamma$ , whose frequency response is  $(j\omega)^\gamma$ , coincides with Liouville's fractional derivative of order  $\gamma$  which is often denoted by  $D^\gamma$ . When  $\gamma = m$  is integer, one recovers the traditional derivatives  $D^m = \frac{d^m}{dx^m}$ .

Adding the requirement of rotation invariance further narrows down the options. One is then left with the fractional Laplacians  $(-\Delta)^{\frac{\gamma}{2}}$  with  $\gamma \in \mathbb{R}^+$ , which are the only multidimensional linear operators that satisfy the requirement of simultaneous shift-, scale- and rotation invariance [31], [32]. The corresponding Fourier-domain definition is

$$(-\Delta)^{\frac{\gamma}{2}} f(\mathbf{x}) \xleftrightarrow{\mathcal{F}} \|\boldsymbol{\omega}\|^\gamma \hat{f}(\boldsymbol{\omega}).$$

Here too, there is a direct correspondence with the classical Laplace operator  $\Delta = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}$  when the order  $\gamma$  is even.

While the above differential operators are promising candidates for defining generalized stochastic processes, one is faced with a technical difficulty in defining the inverse operator  $L^{-1}$  because the frequency responses of  $\partial_\tau^\gamma$  and  $(-\Delta)^{\frac{\gamma}{2}}$  vanish at the origin. This means that the inversion problem, as such, is ill-posed. Fortunately, it is not insurmountable because the null space of our fractional derivative operator is finite-dimensional: it is made up of the polynomials of degree  $\lceil \gamma - 1 \rceil$ . Concretely, this means that we can uniquely specify the inverse operator (and solve our stochastic PDE) by imposing suitable boundary conditions. In previous work on fractional Brownian motion, we have shown that one can design an inverse operator  $L^{-1}$  that forces the process (and a proper number of derivatives) to vanish at the origin. Since the derivation of these inverse operators (which are fractional integrators with boundary conditions at the origin) and their duals is somewhat involved, we refer the reader to the corresponding publications for mathematical details [32], [33]. The basic results are summarized in Table 1.

To gain some insight into the type of transformation, let us have a closer look at the fractional integral operator  $L^{-1} = (-\Delta)_0^{-\frac{\gamma}{2}}$ , which is defined as follows

$$(-\Delta)_0^{-\frac{\gamma}{2}} \varphi(\mathbf{x}) = \int_{\mathbb{R}^d} \frac{e^{j\langle \mathbf{x}, \boldsymbol{\omega} \rangle} - \sum_{|\mathbf{k}|=0}^{\lfloor \gamma - \frac{d}{2} \rfloor} \frac{j^{|\mathbf{k}|} \mathbf{x}^{\mathbf{k}} \boldsymbol{\omega}^{\mathbf{k}}}{\mathbf{k}!}}{\|\boldsymbol{\omega}\|^\gamma} \hat{\varphi}(\boldsymbol{\omega}) \frac{d\boldsymbol{\omega}}{(2\pi)^d} \quad (15)$$

with the condensed multi-integer notations:  $|\mathbf{k}| = k_1 + \dots + k_d$ ,  $\mathbf{k}! = k_1! \dots k_d!$ , and  $\boldsymbol{\omega}^{\mathbf{k}} = \omega_1^{k_1} \dots \omega_d^{k_d}$ . We note that, except for the summation term within the integral, it corresponds to the inverse Fourier transform of  $\|\boldsymbol{\omega}\|^{-\gamma} \hat{\varphi}(\boldsymbol{\omega})$  which represents the filtering of  $\varphi$  with the (potentially unstable) inverse of the fractional Laplacian. The correction term amounts to a polynomial (in the form of a Taylor series at the origin) that ensures that the resulting function  $(-\Delta)_0^{-\frac{\gamma}{2}} \varphi(\mathbf{x})$  and its derivatives up to order  $\lfloor \gamma - \frac{d}{2} \rfloor$  are vanishing at  $\mathbf{x} = \mathbf{0}$ . This is justified formally by invoking the moment property of the Fourier transform:

$$\int_{\mathbb{R}^d} \sum_{|\mathbf{k}|=0}^N \frac{j^{|\mathbf{k}|} \mathbf{x}^{\mathbf{k}} \boldsymbol{\omega}^{\mathbf{k}}}{\mathbf{k}!} \hat{\varphi}(\boldsymbol{\omega}) \frac{d\boldsymbol{\omega}}{(2\pi)^d} = \sum_{|\mathbf{k}|=0}^N \frac{\mathbf{x}^{\mathbf{k}}}{\mathbf{k}!} \underbrace{\int_{\mathbb{R}^d} j^{|\mathbf{k}|} \boldsymbol{\omega}^{\mathbf{k}} \hat{\varphi}(\boldsymbol{\omega}) \frac{d\boldsymbol{\omega}}{(2\pi)^d}}_{\varphi^{(\mathbf{k})}(\mathbf{0})}.$$

where  $\varphi^{(\mathbf{k})}(\mathbf{x}) = \frac{\partial^{k_1 + \dots + k_d}}{\partial x_1^{k_1} \dots \partial x_d^{k_d}} \varphi(\mathbf{x})$ . Forcing the values of the (generalized) function and its derivatives to be zero at the origin is crucial for the specification of fractional Brownian motion as there, by definition, the process should equal zero at  $\mathbf{x} = \mathbf{0}$  [34]. Conveniently, these are precisely the boundary conditions that are imposed by all fractional integral operators  $L^{-1}$  in Table 1. Another important property is that  $LL^{-1}u = u$  in the distributional sense, which constitutes the foundation of the proposed innovation models.

The adjoint operator of  $(-\Delta)_0^{-\frac{\gamma}{2}}$  is specified by

$$(-\Delta)_0^{-\frac{\gamma}{2}*} \varphi(\mathbf{x}) = \int_{\mathbb{R}^d} e^{j\langle \mathbf{x}, \boldsymbol{\omega} \rangle} \frac{\hat{\varphi}(\boldsymbol{\omega}) - \sum_{|\mathbf{k}|=0}^{\lfloor \gamma - \frac{d}{2} \rfloor} \hat{\varphi}^{(\mathbf{k})}(\mathbf{0}) \frac{\boldsymbol{\omega}^{\mathbf{k}}}{\mathbf{k}!}}{\|\boldsymbol{\omega}\|^\gamma} \frac{d\boldsymbol{\omega}}{(2\pi)^d}$$

and has the same type of flavor. The difference is that it now includes a Taylor series correction in the frequency domain that sets the moments of the test function to zero up to order  $\lfloor \gamma - \frac{d}{2} \rfloor$ . Mathematically, this compensates the singularity of  $\|\boldsymbol{\omega}\|^{-\gamma}$  at  $\boldsymbol{\omega} = \mathbf{0}$  and ensures that this dual operator is well-behaved over the class of test functions  $\varphi \in \mathcal{S}$ . The domain of definition of the dual fractional integral operators can actually be extended to the weighted  $L_1$  space  $L_1(\mathbb{R}^d, w_\alpha) = \{f(\mathbf{x}) : \|w_\alpha f\|_{L_1(\mathbb{R}^d)} < +\infty\}$  where  $w_\alpha = (1 + \|\mathbf{x}\|)^\alpha$  and  $\alpha > \lfloor \gamma - d/2 \rfloor$ , which is considerably larger than  $\mathcal{S}$ . This is stated in the following theorem, whose proof is given in Appendix II.

*Theorem 2:* Let  $\varphi \in L_1(\mathbb{R}^d, w_N)$  with integer  $N > H = \gamma - d/2 \in \mathbb{R}^+ \setminus \mathbb{N}$ . Then,  $\|(-\Delta)_0^{-\frac{\gamma}{2}*} \varphi\|_{L_2(\mathbb{R}^d)} < +\infty$  and the corresponding generalized Gaussian and symmetric Poisson characteristic forms (13) and (14) are well-defined. The same applies in 1-D for the operators  $\partial^\gamma$  for any  $\tau \in \mathbb{R}$ .

Note that the cases where  $H$  is an integer are excluded because the corresponding  $L_2$ -norms are generally unbounded, including when  $\varphi \in \mathcal{S}$ . The proof that is given in Appendix II completely takes care of the Gaussian and symmetric Poisson cases. In more recent work, we have extended the results for the general, non-symmetric Poisson case by slightly modifying the inverse operators to make them stable in the  $L_1$ -sense [35].

The boundedness result in Theorem 2 ensures that the definition of the corresponding Gaussian and generalized Poisson processes is mathematically sound. By the same token, it also provides a constructive method for solving the stochastic differential equation (12), thanks to the inverse operator specified by (15). Indeed, we can show that  $(L^{-1*})L^*\varphi = \varphi$  and  $LL^{-1}\varphi = \varphi$  (which is the dual statement of the former) for all test functions  $\varphi \in \mathcal{S}$ . This means that  $L^{-1*}$  is the left inverse of  $L^*$ , while  $L^{-1}$  is the right inverse of  $L$ . In the first case, the operator  $L^*$  sets the moments of the intermediate function to zero so that the effect of  $L^{-1*}$  is equivalent to that of the unregularized inverse. In the second case,  $L$  sets to zero the polynomial component that was added by  $L^{-1}$  to fulfill the boundary conditions at  $\mathbf{x} = \mathbf{0}$ .

We conclude this technical discussion by mentioning that, unlike the fractional derivatives and Laplacian, the inverse operators specified in Table 1 are not shift-invariant (because of the boundary conditions). They are, however, invariant with respect to scaling as well as rotation in the case of the Laplacian. The implication is that the corresponding random processes will inherit some form of invariance, but that they will generally not be stationary.

#### D. Fractal and spline-type processes

By combining the scale-invariant operators of Table 1 with the general framework proposed in Section III.C, we obtain an interesting family of stochastic processes. Their key property is self-similarity in the sense that the spatially-rescaled versions of a process are also part of the same family. Our generation model is ultimately simple and boils down to a fractional integration of a white noise process subject to appropriate boundary conditions. The mathematical technicalities have been dealt with in the preceding section by specifying the proper fractional integration operators together with their domain of definition (cf. Theorem 2).

TABLE I  
INVERSION OF THE SCALE-INVARIANT OPERATORS IN ONE AND MULTIPLE DIMENSIONS.

L	$\rho = L^{-1}\delta$	$L^{-1}\varphi$	$L^{-1*}\varphi$
$D^\gamma$	$\frac{x_+^{\gamma-1}}{\Gamma(\gamma)}$	$\int_{\mathbb{R}} \frac{e^{j\omega x} - \sum_{k=0}^{\lfloor \gamma - \frac{1}{2} \rfloor} \frac{j^k x^k \omega^k}{k!} \hat{\varphi}(\omega)}{(j\omega)^\gamma} \frac{d\omega}{2\pi}$	$\int_{\mathbb{R}} e^{j\omega x} \frac{\hat{\varphi}(\omega) - \sum_{k=0}^{\lfloor \gamma - \frac{1}{2} \rfloor} \hat{\varphi}^{(k)}(0) \frac{\omega^k}{k!}}{(-j\omega)^\gamma} \frac{d\omega}{2\pi}$
$\partial_\tau^\gamma$	$A_\tau^\gamma  x ^{\gamma-1} + B_\tau^\gamma \text{sign}(x)  x ^{\gamma-1}$	$\int_{\mathbb{R}} \frac{e^{j\omega x} - \sum_{k=0}^{\lfloor \gamma - \frac{1}{2} \rfloor} \frac{j^k x^k \omega^k}{k!} \hat{\varphi}(\omega)}{(-j\omega)^{\frac{\gamma}{2}-\tau} (j\omega)^{\frac{\gamma}{2}+\tau}} \frac{d\omega}{2\pi}$	$\int_{\mathbb{R}} e^{j\omega x} \frac{\hat{\varphi}(\omega) - \sum_{k=0}^{\lfloor \gamma - \frac{1}{2} \rfloor} \hat{\varphi}^{(k)}(0) \frac{\omega^k}{k!}}{(j\omega)^{\frac{\gamma}{2}-\tau} (-j\omega)^{\frac{\gamma}{2}+\tau}} \frac{d\omega}{2\pi}$
$(-\Delta)^{\frac{\gamma}{2}}$	$C_\gamma \ \mathbf{x}\ ^{\gamma-d}, \gamma-d \notin 2\mathbb{N}$	$\int_{\mathbb{R}^d} \frac{e^{j\langle \mathbf{x}, \boldsymbol{\omega} \rangle} - \sum_{ k =0}^{\lfloor \gamma - \frac{d}{2} \rfloor} \frac{j^{ \mathbf{k} } x^{\mathbf{k}} \boldsymbol{\omega}^{\mathbf{k}}}{ \mathbf{k} !} \hat{\varphi}(\boldsymbol{\omega})}{\ \boldsymbol{\omega}\ ^\gamma} \frac{d\boldsymbol{\omega}}{(2\pi)^d}$	$\int_{\mathbb{R}^d} e^{j\langle \mathbf{x}, \boldsymbol{\omega} \rangle} \frac{\hat{\varphi}(\boldsymbol{\omega}) - \sum_{ k =0}^{\lfloor \gamma - \frac{d}{2} \rfloor} \hat{\varphi}^{(\mathbf{k})}(0) \frac{\boldsymbol{\omega}^{\mathbf{k}}}{ \mathbf{k} !}}{\ \boldsymbol{\omega}\ ^\gamma} \frac{d\boldsymbol{\omega}}{(2\pi)^d}$

$$\text{with } A_{\gamma,\tau} = -\frac{\cos \pi\tau}{2\Gamma(\gamma) \sin(\frac{\pi}{2}(\gamma-1))}, \quad B_{\gamma,\tau} = -\frac{\sin \pi\tau}{2\Gamma(\gamma) \cos(\frac{\pi}{2}(\gamma-1))}, \quad \text{and} \quad C_\gamma = \frac{2^\gamma \Gamma(\frac{d+\gamma}{2})}{\pi^{\frac{d}{2}} \Gamma(-\frac{\gamma}{2})}$$

We first consider the one-dimensional case. If the driving noise is Gaussian, the formulation is equivalent to that presented in [33]. Specifically, by taking  $L = \partial_\tau^\gamma$  with  $\frac{1}{2} < \gamma < \frac{3}{2}$  and any  $\tau \in \mathbb{R}$ , one obtains fractal processes that are equivalent to the fractional Brownian motion (fBm) introduced by Mandelbrot and Van Ness [36]. Rather than the order, fBMs are usually characterized by their Hurst exponent  $H = \gamma - \frac{d}{2}$  whose value is restricted to the open interval  $]0, 1[$ ; the fractal dimension is  $D = d + 1 - H$ . The case  $\gamma = 1$  (or  $H = \frac{1}{2}$ ) corresponds to the Brownian motion (or Wiener) process, which is illustrated in Fig. 1a. By definition, this process is whitened by all first-order derivative operators  $\partial_\tau^1$ ; in particular, by  $D = \partial_{\frac{1}{2}}^1$ , which corresponds to the optimal regularization functional for the Brownian motion denoising problem (1). The formalism is also applicable for  $H = \gamma - \frac{1}{2} > 1$  (but non-integer), in which case it yields the higher-order extensions of fBm introduced by Perrin et al. [37].

Alternatively, if we excite the system with impulse noise, we obtain a stochastic process that is a random spline of order  $\gamma$  or, equivalently, of degree  $\gamma - 1$ . The corresponding  $\partial_\tau^\gamma$ -Poisson processes are piecewise-smooth: they have pointwise discontinuities with a Hölder exponent<sup>5</sup>  $\gamma - 1$  at the spline knots and are infinitely differentiable in between (this follows from the properties of the Green functions in Table 1). In other words, they are infinitely differentiable almost everywhere, whereas the  $\gamma$ th order fBMs are uniformly rough everywhere (i.e., Hölder-continuous of order  $\gamma - 1$ ). Another distinction between the Gaussian and Poisson processes is the importance of the phase factor  $\tau$  for the latter. Specifically, we can use any fractional derivative operator  $\partial_\tau^\gamma$  to whiten an fBm of order  $\gamma$ , while in the case of a random spline, the value of  $\tau = \tau_0$  needs to be matched to the type of singularity (generation model) to recover Poisson noise. Some examples of extended fBMs and random splines with matching exponents are shown in Fig. 4. The important point is that the signals that are shown side by side share a common innovation model (same whitening operator  $L$ ) but yet are fundamentally different: the random splines on the right have a sparse representation (due to their finite rate of innovation), which is not the case for the Gaussian fBMs on the left. These signals were generated by inverse FFT using a discretized version of the Fourier representation of the operator  $L^{-1}$ . The Poisson noise was generated using uniform

<sup>5</sup>A function  $f(x)$  is  $r$ -Hölder-continuous iff  $\sup_{x,y} |f^{(N)}(y) - f^{(N)}(x)|/|y-x|^{r-N}$  is finite with  $N = \lceil r \rceil - 1$ .

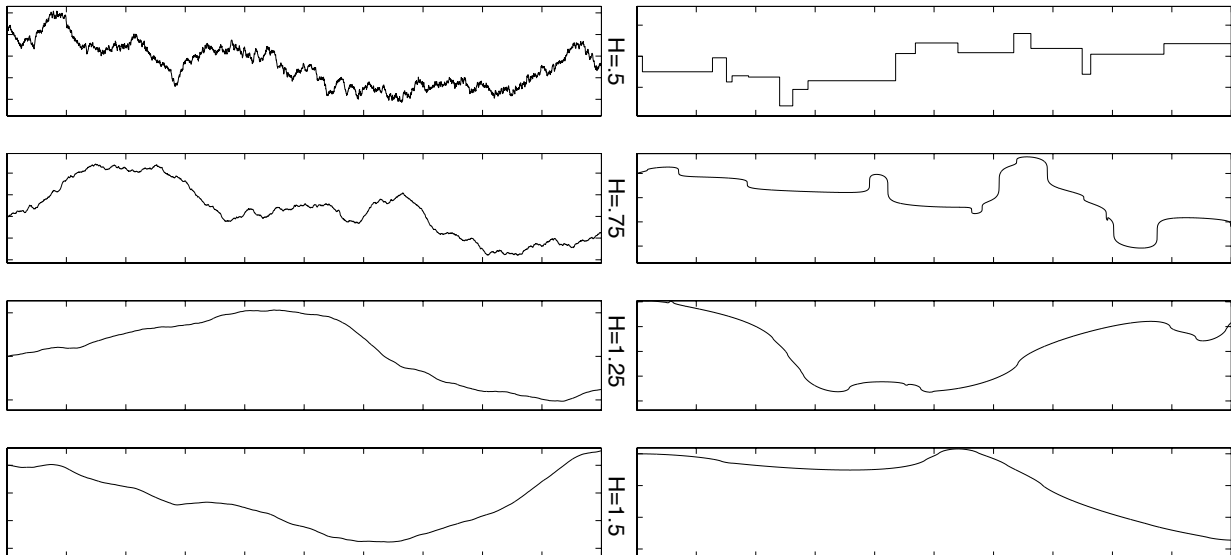


Fig. 4. Gaussian vs. sparse signals: Comparison of fractional Brownian motion (left column) and Poisson (right column) generalized stochastic processes as the order increases. The processes that are side-by-side have the same order  $\gamma = H + \frac{1}{2}$  and identical second order statistics.

random generators for the location and amplitude parameters. The underlying operators in this experiment are causal derivatives so that the generalized Poisson processes of integer order are polynomial splines (e.g., piecewise-constant for  $H = 0.5$  and piecewise-linear for  $H = 1.5$ ). As the order increases, the functions become smoother and the two types of processes are less and less distinguishable visually.

The generation mechanism naturally extends to multiple dimensions. In the case of a Gaussian excitation, we obtain fractional Brownian fields [38], whose detailed characterization as generalized random processes was presented in a recent paper [32]. If we switch to a Poisson noise, we generate random fields that are polyharmonic splines of order  $\gamma$ . For  $\frac{d}{2} < \gamma < d$ , these exhibit impulse-like singularities which are explained by the form of the Green function (cf. Table 1); the impulse-like behavior disappears as the order increases and as the process becomes smoother. A series of generalized fBms and random polyharmonic splines with matching exponents is presented in Fig. 5. Note the characteristic cloud-like appearance of low order fBms. Here too, the two types of processes become more and more similar visually as the order increases.

We did allude to the fact that the processes in Fig. 4 and 5 are self-similar. This property is well known for the fBms which are the prototypical examples of stochastic fractals [38], [39]. The random splines that we have defined here are self-similar as well, but in a weaker sense. Specifically, if one dilates such a generalized Poisson process by a factor of  $a$  one obtains a generalized Poisson process that is in the same family but with a rescaled Poisson parameter  $\lambda_a = \lambda_1/a^d$  where  $\lambda_1$  is the parameter of the primary process.

#### E. Mixed Poisson processes

Conveniently, the characteristic form of the sum of two independent processes is the product of their characteristic forms; i.e.,  $\mathcal{Z}_{s_1+s_2}(\varphi) = \mathcal{Z}_{s_1}(\varphi) \cdot \mathcal{Z}_{s_2}(\varphi)$  [18]. A direct implication is that a Poisson noise  $w(\mathbf{x})$  is infinitely divisible in the sense that it can always be broken into a sum of  $N$  i.i.d. Poisson processes.



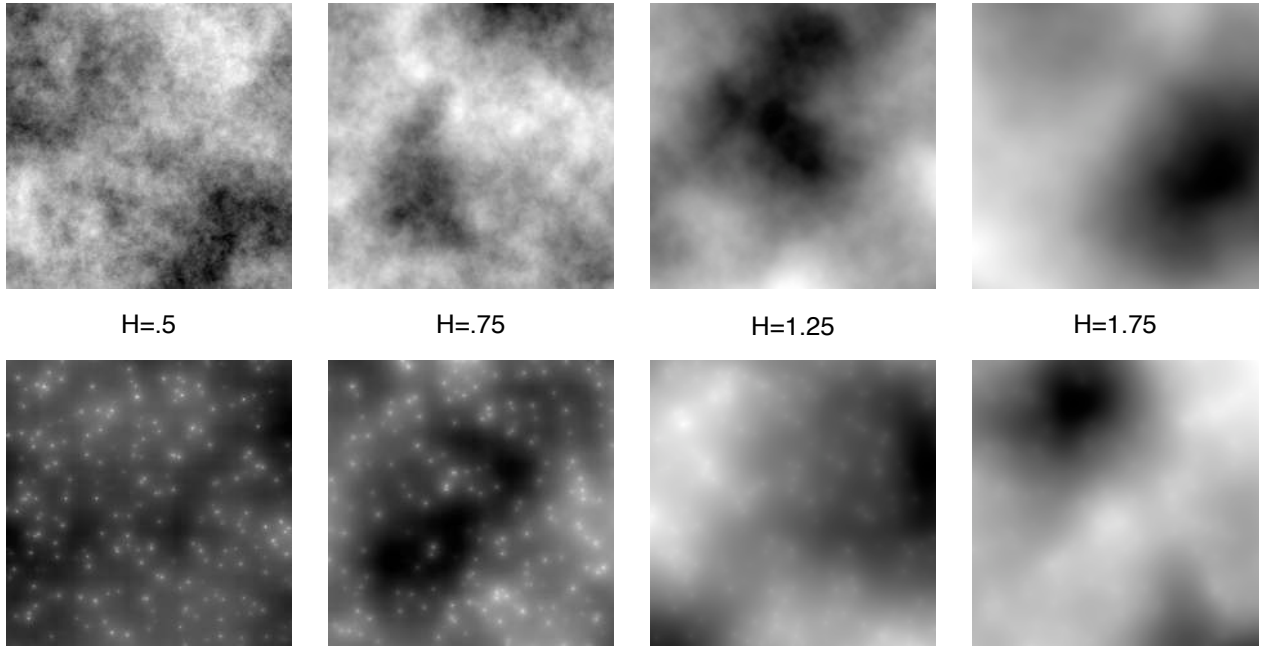


Fig. 5. Gaussian vs. sparse signals in 2-D: Comparison of fractional Brownian (upper row) and Poisson (lower row) generalized stochastic fields as the order increases. The processes in the same column have the same order  $\gamma = H + 1$  and identical second order statistics.

Following this line of thought, we propose to construct mixed Poisson processes by summing  $N$  independent generalized Poisson processes with parameters  $\{(L_n, \lambda_n, P_n(a))\}_{n=1}^N$ . The characteristic form of such a mixed process is

$$\mathcal{Z}_{s_1+\dots+s_N}(\varphi) = \prod_{n=1}^N \mathcal{Z}_{s_n}(\varphi) = \exp\left(\sum_{n=1}^N \lambda_n v_n(\varphi)\right)$$

where

$$v_n(\varphi) = \int_{\mathbb{R}} \int_{\mathbb{R}^d} (e^{jaL_n^{-1}\varphi(\mathbf{x})} - 1) d\mathbf{x} dP_n(a). \quad (16)$$

The corresponding random signals exhibit different types of singularities. They will generally be piecewise-smooth—sums of random splines—if the operators  $L_n$  are scale-invariant (cf. subsection IV.C). We can vary their structural complexity by including more or less independent Poisson components. Yet, their intrinsic sparsity (or rate of innovation) remains the same as long as  $\lambda = \sum_{n=1}^N \lambda_n = \text{Const}$ . The variations along this theme are countless, especially in 1-D, due to the large variety of available operators (e.g.,  $\partial_\tau^\gamma$  with  $\tau \in \mathbb{R}^+$  and  $\gamma > \frac{1}{2}$ ).

#### F. The Mondrian process

We became especially intrigued by the generalized Poisson process associated with the partial differential operator  $L = \frac{\partial^d}{\partial x_1 \dots \partial x_d}$  and decided to call it the “Mondrian process”. A colorful realization is shown in Fig. 6. The 2-D process  $u(x, y)$  corresponding to this illustration is the solution of the (marginally unstable) stochastic PDE:

$$\frac{\partial^2 u(x, y)}{\partial x \partial y} = \sum_{k=1}^K a_k \delta(x - x_k, y - y_k)$$

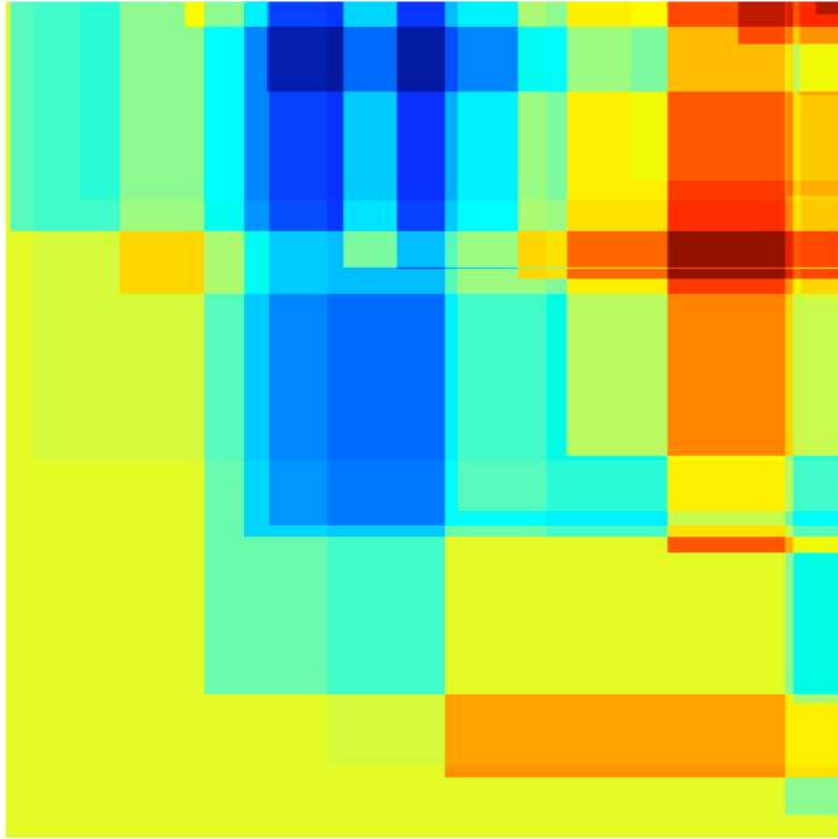


Fig. 6. Pseudo-color display of a realization of the Mondrian process with  $\lambda = 30$ .

where the  $a_k$ ,  $x_k$  and  $y_k$  are independent uniformly-distributed random variables. Since  $L$  is separable, its Green function is separable as well: it is the multidimensional Heaviside function  $1_+(\mathbf{x}) = \prod_{n=1}^d 1_+(x_n)$  (e.g.,  $L^{-1}$  is a quarter plane integrator in 2-D). This results in a signal that is the direct multi-D counterpart of (4).

The Mondrian process in Fig. 6 was constructed and first shown at a conference in the honor of Martin Vetterli in 2007. It is distinct from the Mondrian process of Roy and Teh [40] which is generated hierarchically via a random subdivision of rectangular regions. The characteristic feature of the present construct is the spatial, long-range dependance that is introduced by the (quarter plane) integration process. This type of pattern is more intriguing visually and intellectually than a mere random superposition of colored rectangles—remarkably, it is also shorter to describe mathematically (cf. equation above).

#### G. Poisson processes and system modeling

The proposed stochastic framework is general enough to handle many other types of operators; essentially, all those that lead to viable deterministic spline constructions since these require the resolution of the same type of operator equation [cf. (12)]. The choice of the “optimal” operator may be motivated by physical considerations. For instance, it is possible to model time-domain signals  $s(t)$  that are generated by a physical system driven by random impulse-like events. Specifically, the operator  $L$  associated with an  $N$ th order ordinary differential equation is characterized by the generic rational frequency response

$$\hat{L}(\omega) = \frac{(j\omega)^N + a_{N-1}(j\omega)^{N-1} + \cdots + a_1(j\omega) + a_0}{b_{N-1}(j\omega)^{N-1} + \cdots + b_1(j\omega) + b_0}. \quad (17)$$

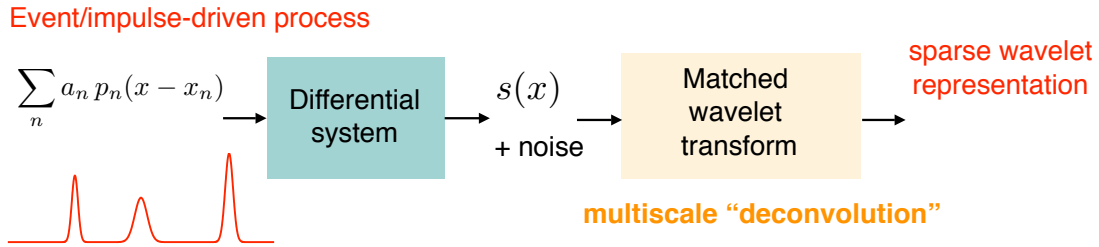


Fig. 7. A signal analysis paradigm where the operator  $L$  as well as the corresponding wavelet decomposition are matched to the characteristics of a physical, event-driven linear system.

It admits a stable causal inverse (i.e.,  $L^{-1}\delta(t) = h_+(t) \in L_1(\mathbb{R})$ , which is the impulse response of the system) iff. the poles of the system are in the left complex plane. Interestingly, there is no such restriction for defining the corresponding family of exponential splines, which are entirely specified by the poles and zeros of the system [20], nor for using this type of operator for specifying generalized stochastic processes. In general, the processes corresponding to BIBO stable inverse operators<sup>6</sup> are stationary, while the ones corresponding to unstable operators (poles on the imaginary axis) are non-stationary. The prototypical examples in this last category are the random polynomial splines; these are generated by  $N$ -fold integration which is basically an unstable operation ( $N$ th-order pole at 0).

We believe that the major benefit of adopting a system modeling point of view is that it provides us with a conceptual framework for developing new “designer wavelets” that are matched to particular classes of signals, in accordance with the scheme outlined in Fig. 7.

## V. ON THE WAVELET COMPRESSIBILITY OF GENERALIZED POISSON PROCESSES

### A. Operator-like wavelets

It is well known that conventional wavelets bases act like multi-scale derivatives [9]. We have exemplified this behavior for the Haar basis in our introductory discussion. More generally,  $N$ th order wavelets, which have  $N$  vanishing moments, behave like  $N$ th order derivatives; i.e., they can be represented as  $\psi(x) = D^N\phi(x)$  where  $\phi(x)$  is a suitable (lowpass) smoothing kernel. In the case where the wavelet is a polynomial spline (e.g. Battle-Lemarié or B-spline wavelet), the link with the differential operator can be made completely explicit. For instance, it is well-known that the cardinal spline wavelet

$$\psi_{D^N}(x) = D^N\phi_{\text{int}}(2x),$$

where  $\phi_{\text{int}}(x)$  is the unique cardinal spline interpolator of order  $2N$ , generates a semi-orthogonal Riesz basis of  $L_2(\mathbb{R})$  for any  $N \geq 1$  [41].

Remarkably, this concept carries over to more general classes of operators provided that there is a corresponding spline construction available. In particular, there exist wavelet bases that are perfectly matched to the complete range of fractional derivative operators in Table 1:

<sup>6</sup>The present mathematical framework can cope with poles that are in the right complex plane by allowing non-causal inverses, irrespective of any physical consideration.

- the fractional spline wavelets which are linked to the operator  $D^\gamma$  [42], and well as  $\partial_\tau^\gamma$  [43], [44];
- the multidimensional polyharmonic spline wavelets associated with the fractional Laplacian  $(-\Delta)^{\frac{\gamma}{2}}$  [45].

The latter are thin-plate spline functions that live in the span of the radial basis functions  $\|\mathbf{x} - \mathbf{k}\|^{\gamma-d}$ ,  $\mathbf{k} \in \mathbb{Z}^d$  (cf. Green's function in Table 1). The spline wavelets come in a variety of flavors (semi-orthogonal, B-spline, operator-like) and can also be constrained to be orthogonal.

When the operator  $L$  is not scale-invariant anymore, we can still build wavelets that form multiresolution hierarchies and yield Riesz bases of  $L_2(\mathbb{R}^d)$ , but that are no longer dilates of one another. Yet, as long as the operator is shift-invariant, the wavelets at a given resolution  $i$  remain translates of a single prototype whose generic form is

$$\psi_i(\mathbf{x}) = L^* \phi_i(\mathbf{x})$$

where  $\phi_i$  is a resolution-dependent smoothing kernel [46]. In the canonical construction,  $\phi_i$  is an  $L^*L$ -spline interpolator with respect to the grid at resolution  $i$  (in accordance with Definition 1), which ensures that  $\psi_i(\mathbf{x})$  is itself an  $L$ -spline. The corresponding transform can also be implemented using Mallat's fast filterbank algorithm but with resolution-dependent filters. In particular, we have shown how to construct 1-D wavelets that replicate the behavior of any given  $N$ th order differential operator, as specified by (17) [47]. While the basic wavelet prototypes are exponential splines, the scheme can be extended to obtain generalized, operator-like Daubechies wavelets that are both orthogonal and compactly supported [48].

### B. Wavelet analysis of generalized Poisson processes

We will now argue that the above operator-like wavelets are matched to the type of generalized Poisson processes introduced in this paper and that they will generally result in a sparse representation. By solving the operator equation (12), we obtain the following explicit representation of a generalized  $L_1$ -Poisson process:

$$s_1(\mathbf{x}) = p_{L_1}(\mathbf{x}) + \sum_n a_n \rho_{L_1}(\mathbf{x} - \mathbf{x}_n)$$

where  $p_{L_1}(\mathbf{x})$  is a finite-dimensional signal component that is in the null space of  $L_1$  and where  $\rho_{L_1}$  is the Green's function of  $L_1$  such that  $L_1 \rho_{L_1} = \delta$ ; the  $\mathbf{x}_n$ 's are random locations that follow a Poisson distribution with parameter  $\lambda$  while the  $a_n$ 's are i.i.d. random weights with probability distribution  $p(a)$ .

Let us consider the analysis of such a process with any higher-order  $L$ -wavelet transform with the property that  $L = L_0 L_1$  where the factor  $L_0$  is a proper<sup>7</sup> differential operator (the limit case being  $L_0 = \text{Identity}$ ). The wavelet coefficients at resolution  $i$  of  $s(\mathbf{x})$  are obtained as follows

$$\begin{aligned} \langle s_1, \psi_{i,\mathbf{k}} \rangle &= \langle s_1, L^* \phi_i(\cdot - 2^i \mathbf{k}) \rangle \\ &= \langle s_1, L_1^* L_0^* \phi_i(\cdot - 2^i \mathbf{k}) \rangle \\ &= \langle L_1 s_1, L_0^* \phi_i(\cdot - 2^i \mathbf{k}) \rangle \quad (\text{by duality}) \\ &= \left\langle \sum_n a_n \delta(\mathbf{x} - \mathbf{x}_n), L_0^* \phi_i(\cdot - 2^i \mathbf{k}) \right\rangle \\ &= \sum_n a_n L_0^* \phi_i(\mathbf{x}_n - 2^i \mathbf{k}). \end{aligned} \tag{18}$$

<sup>7</sup>The requirements are: 1) linear shift-invariance and 2)  $L_0^* \phi \in L_2(\mathbb{R}^d)$  with sufficient decay, the worst case being  $O(|\mathbf{x}|^{-\gamma_0-d})$  where  $\gamma_0$  is the (possibly fractional) order of the operator  $L_0$  [35, Proposition 2.4].

Recalling that the density of Poisson impulses is proportional to  $\lambda$ , this result suggests that the wavelet representation of such a process is intrinsically  $\lambda$ -sparse, provided that the  $L_0^*\phi_i$ 's are decaying sufficiently fast. Indeed, since the essential support of the smoothing kernel  $\phi_i$  is usually proportional to  $2^i$  (the wavelet scale), it implies that a singularity will affect the same number of wavelet basis functions at any resolution (cone-like region of influence). This last characteristic number, which is 1 for the Haar transform, can also be expected to increase with the order of  $L$  since higher-order wavelets are typically less localized. Another implicit effect is that the amplitude of the wavelet coefficients will exhibit a certain rate of decay (or growth) with  $i$  because of the unit-norm normalization of the basis functions. For instance, in the case of the Haar basis, we have that  $2^{-i/2}\psi(x/2^i) = D\phi_i(x)$  where  $\phi_i(x) = 2^{i/2}\phi_0(x/2^i)$  is a rescaled triangle function.

Interestingly, this sparsifying behavior subsists if we consider a mixed process of the type  $s(x) = s_1(x) + \dots + s_N(x)$  where the  $s_n$ 's are independent generalized  $L_n$ -Poisson processes (cf. subsection IV-E) with the  $L_n$ 's all being admissible factors of  $L$ . In the case of 1-D derivative-like wavelets, the argument applies to a broad class of piecewise-smooth signals because of the numerous ways of factorizing a derivative of order  $\gamma$

$$\partial_\tau^\gamma = \partial_{\tau_1}^{\gamma_1} \partial_{\tau - \tau_1}^{\gamma - \gamma_1};$$

where  $\tau_1 \in \mathbb{R}$  and  $\gamma_1 \in [0, \gamma]$  can be arbitrary. Concretely, this means that a conventional wavelet of order  $N$  will sparsify signals that contain any variety of  $\alpha$ -Hölder point singularities with  $\alpha \leq N - 1$  (cf. Green's function in Table 1). The spline operator-like wavelets discussed earlier turn out to be ideal for this task because the calculations of  $L_0^*\phi_i$  in (18) can be carried out analytically [44, Theorem 1]. Such a behavior of the wavelet transform of a piecewise-smooth signal is well documented in the literature, but it has not been made as explicit before, to the best of our knowledge.

The above analysis also suggests that wavelet compressibility is a robust property (i.e., it is not necessary to pick the "optimal" wavelet that precisely matches the whitening operator of the process). This is good practical news because it means that any wavelet will do, provided its order is sufficient.

Less standard is the extension of the argument for other types of operators that are not necessarily scale-invariant. In particular, if we have prior physical knowledge of the signal generation mechanism, we can invoke the above results to justify the signal processing paradigm in Fig. 8. The proposed wavelet-based approach is stable and much more robust to modeling errors and noise than a conventional deconvolution scheme. We have applied this strategy and demonstrated its benefits in two concrete applications: (1) the reconstruction of the dynamic positron emission tomograms with a constraint on the  $\ell_1$ -norm of the spatio-temporal wavelet coefficients [49], and (2) the detection of neuronal activity using functional magnetic resonance imaging [50]. In the first case, we have tuned the time-domain wavelets to the pharmacokinetic model that rules the time activity curve in response to the injection of a radioactive tracer. In the second case, we matched the wavelets to a linearized version of an established model of the hemodynamic response of the brain. In both instances, the results obtained with "designer" wavelets were superior to those obtained with conventional ones.

## VI. BACK TO THE COMPOUND POISSON PROCESS

To illustrate the suitability of the proposed formalism for deriving new signal estimators, we consider the compound Poisson process whose explicit form is given by (4). First, we note that the total variation of this signal

is  $\text{TV}(s) = \sum_{k \in \mathbb{Z}} |a_k| = \|a_k\|_{\ell_1}$ , which already makes an interesting connection between parametric (FRI) estimation and  $\ell_1$ -minimization (cf. Eq. (3)). To specify an optimal statistical estimator (MAP or minimum mean-square error) for the denoising problem in Section II, we need to know the  $N$ th-order joint probability density of the samples of the signal at the integers:  $p(s(1), s(2), \dots, s(N))$ . Instead of working with these samples which are strongly correlated, we propose to consider the finite-difference process  $w_d(x) = s(x) - s(x-1) = D_d s(x)$ , which is stationary with a very short correlation distance. To specify this latter process mathematically, we use the technique of Section IV-B to transfer the underlying operators onto the argument of the characteristic form of the Poisson noise  $w$  given by Theorem 1:

$$\mathcal{Z}_{w_d}(\varphi) = \mathcal{Z}_{D_d s}(\varphi) = \mathcal{Z}_{D_d D^{-1} w}(\varphi) = \mathcal{Z}_w(D^{-1*} D_d^* \varphi) = \mathcal{Z}_w(\beta_-^0 * \varphi)$$

where  $\beta_-^0(x)$  is the anti-causal B-spline of degree 0, which is piecewise-constant and compactly supported in  $x \in [-1, 0)$ . The critical step here is to show that  $D^{-1*} D_d^* \varphi(x) = (\beta_-^0 * \varphi)(x)$ , which is best achieved in the Fourier domain by using the relevant formula for  $D^{-1*}$  ( $\gamma = 1$ ) in Table 1:

$$D^{-1*} D_d^* \varphi(x) \xleftrightarrow{\mathcal{F}} \frac{(1 - e^{j\omega})\hat{\varphi}(\omega) - (1 - e^{j0})\hat{\varphi}(0)}{-j\omega} = \hat{\varphi}(\omega) \left( \frac{1 - e^{j\omega}}{-j\omega} \right)$$

where the right-hand side factor is precisely the Fourier transform of  $\beta_-^0$ . Note that the (forward) finite difference operator  $D_d^*$ , whose frequency response is  $(1 - e^{j\omega})$ , suppresses the zero-order correction term of  $D^{-1*}$  (integration constant), which is crucial for obtaining a stationary output. Next, we get the 2-D characteristic function  $\hat{p}(\omega_m, \omega_n)$  of the joint distribution  $p(w_d(m), w_d(n))$  with  $m, n \in \mathbb{Z}$  and  $m \neq n$  by evaluating the characteristic form of  $w_d$  for  $\varphi = \omega_m \delta(\cdot - m) + \omega_n \delta(\cdot - n)$ , which yields  $\hat{p}(\omega_m, \omega_n) = \mathcal{Z}_w(\omega_m \beta_-^0(\cdot - m) + \omega_n \beta_-^0(\cdot - n))$ . Moreover, since the B-splines  $\beta_-^0(\cdot - m)$  and  $\beta_-^0(\cdot - n)$  are non-overlapping and the Poisson characteristic form in Theorem 1 factorizes for functions with distinct support, we have that  $\hat{p}(\omega_m, \omega_n) = \hat{p}_w(\omega_m) \cdot \hat{p}_w(\omega_n)$  with

$$\hat{p}_w(\omega) = \mathcal{Z}_w(\omega \beta_-^0) = \exp\left(\lambda \int_{\mathbb{R}} \int_{\mathbb{R}} (e^{ja\omega \beta_-^0(x)} - 1) dx dP(a)\right) = \exp\left(\lambda \int_{\mathbb{R}} (e^{ja\omega} - 1) dP(a)\right)$$

where we have used the fact that  $\beta_-^0(x)$  is equal to one for  $x \in [-1, 0)$  and zero elsewhere to evaluate the inner integral over  $x$ . This factorization result proves independence and has the following implication:

*Proposition 1:* The integer samples of the finite-difference process  $w_d(x) = D_d s(x) = s(x) - s(x-1)$  where  $s$  is a generalized poisson process with parameters  $(D, \lambda, P(a))$  are i.i.d. random variables with probability distribution function  $p_w(w_d) = \mathcal{F}^{-1}\{\exp(\lambda \int_{\mathbb{R}} (e^{ja\omega} - 1) dP(a))\}(w_d)$ .

It follows that  $p_w$  provides the complete information for the statistical description of the sampled version of such signals. Proposition 1 allows us to express the regularization functional for the MAP estimator as a summation of independent log-likelihood terms, which results in a form that is compatible with the discretized version of the TV estimator described by (3). Interestingly, we can get an exact equivalence by making the formal substitution  $\frac{e^{-|a|}}{|a|} da = \lambda dP(a)$  in the Poisson functional. The relevant Fourier-domain identity is

$$\exp \int_{\mathbb{R}} (e^{ja\omega} - 1) \frac{e^{-|a|}}{|a|} da = \frac{1}{1 + \omega^2} = \mathcal{F}\left\{\frac{1}{2} e^{-|w_d|}\right\}(\omega).$$

where the integral on the left-hand side is convergent because  $e^{ja\omega} - 1 = ja\omega + O(a^2)$  as  $a \rightarrow 0$ . This translates into a pure  $\ell_1$ -norm log-likelihood term:  $\sum_k -\log(p(s(k) - s(k-1))) \propto \sum_k |s(k) - s(k-1)|$ , which may

explain the superiority of the TV algorithm in the denoising experiment in Section II. The existence of this limit<sup>8</sup> example is hard evidence of the relevance of the proposed stochastic framework for sparse signal recovery. We should keep in mind, however, that  $\ell_1$ -regularization is only one of the many possibilities, and that the proposed framework is rich enough to yield a board class of statistical estimators. The topic calls for a more detailed investigation/evaluation which is beyond the scope of the present paper.

Let us close by providing an intuitive justification for the digital prefiltering step that is implicit in Proposition 1: while the defining equation (12) would suggest applying the exact whitening/sparsifying operator  $L$  to the signal, this is not feasible conceptually nor practically because: 1) we cannot handle Dirac impulses directly, and 2) the measurements are discrete. The next best thing we can do is to apply a discrete approximation of the operator (e.g. finite difference instead of a derivative) to the samples of the signal to essentially replicate its whitening effect. Remarkably, this discretization does not result in any statistical approximation.

## VII. CONCLUSION

We introduced a unifying operator-based formulation of stochastic processes that encompasses the traditional Gaussian stationary processes, stochastic fractals which are Gaussian but non-stationary, as well as a whole new category of signals with finite rates of innovation. These signals are all specified as solutions of stochastic differential equations driven by white noise of the appropriate type. When the system is stable and the driving noise is Gaussian, the approach is equivalent to the traditional formulation of Gaussian stationary processes. Sparse or FRI signals are obtained in a completely analogous fashion by considering an impulsive Poisson noise excitation. It is important to note that these generalized Poisson processes are not Gaussian, irrespective of the choice of the amplitude distribution of the driving noise.

A particularly interesting situation occurs when the whitening operator is scale-invariant; while the corresponding system is unstable, we have shown that the operator can be inverted by introducing suitable boundary conditions. The corresponding Gaussian processes, which are self-similar, include Mandelbrot's famous fractional Brownian fields. The Poisson counterparts of these processes in one or multiple dimensions are random splines—unlike their fractal cousins, they are infinitely differentiable almost everywhere and piecewise-smooth by construction.

We believe that this latter class of signals constitutes a good test bed for the evaluation and comparison of sparsity-driven signal processing algorithms. The specification of a statistical model is obviously only a first step. A topic for future research is the investigation and extension of the type of result in Proposition 1 and the derivation and assessment of corresponding statistical estimators. While  $\ell_1$ -minimization and wavelet-based algorithms are attractive candidates, they are probably not the ultimate solution of the underlying statistical estimation problem.

<sup>8</sup>The proposed example does not correspond to a compound Poisson process in the strict sense of the term because the function  $|a|^{-1}e^{-|a|}$  is not integrable. It can be described as the limit of the Poisson process:  $\lambda \int_{\mathbb{R}} (e^{ja\omega} - 1) dP(a) = \int_{|a| > \frac{1}{n}} (e^{ja\omega} - 1) \frac{e^{-|a|}}{|a|} da$  with  $\lambda = \int_{|a| > \frac{1}{n}} \frac{e^{-|a|}}{|a|} da$ , as  $n$  tends to infinity. Taking the limit is acceptable and results in a well-defined stochastic process that is part of the extended Lévy family.

## APPENDIX I: RESULTS FROM THE THEORY OF GENERALIZED STOCHASTIC PROCESSES [18]

We recall that a multidimensional distribution  $u$  (or generalized function) is not defined through its point values (samples)  $u(\mathbf{x}), \mathbf{x} \in \mathbb{R}^d$ , but rather through its scalar products (linear functionals)  $u(\varphi) := \langle u, \varphi \rangle$  with all “test” functions  $\varphi \in \mathcal{S}(\mathbb{R}^d)$ . Here,  $\mathcal{S}(\mathbb{R}^d)$  denotes Schwartz’s class of indefinitely differentiable functions of rapid descent (i.e.,  $\varphi$  as well as all its higher order derivatives decay faster than  $O(\|\mathbf{x}\|^{-m}), \forall m \in \mathbb{N}^+$ ). In an analogous fashion, Gelfand defines a generalized stochastic process  $s$  via the probability law of its scalar products with arbitrary test functions  $\varphi \in \mathcal{S}(\mathbb{R}^d)$ , rather than by considering the probability law of its pointwise samples  $\{\dots, s(\mathbf{x}_1), s(\mathbf{x}_2), \dots, s(\mathbf{x}_N), \dots\}$ , as is customary in the conventional formulation.

## A. The characteristic form

Given a generalized process  $s$  and some test function  $\varphi \in \mathcal{S}(\mathbb{R}^d)$ ,  $y = \langle s, \varphi \rangle$  is a random variable characterized by a probability density  $p_\varphi(y) dy$ . The specification of this PDF for any  $\varphi$  allows one define the characteristic form of the process:

$$\mathcal{Z}_s(\varphi) = \mathcal{E}\{e^{-j\langle s, \varphi \rangle}\} = \int_{\mathbb{R}} e^{-jy} p_\varphi(y) dy \quad (19)$$

where  $\mathcal{E}\{\cdot\}$  is the expectation operator.  $\mathcal{Z}_s(\varphi)$  is a functional of  $\varphi$  that fully characterizes the generalized stochastic process  $s$ . In fact, Gelfand’s theory rests upon the principle that specifying  $\mathcal{Z}_s(\varphi)$  is equivalent to defining the underlying generalized stochastic process.

*Theorem 3 (Existence):* Let  $\mathcal{Z}(\varphi)$  be a positive-definite continuous functional on a test space  $\mathcal{S}$  such that  $\mathcal{Z}(0) = 1$ . Then there exists a generalized process  $s$  whose characteristic functional is  $\mathcal{Z}(\varphi)$ .

We will illustrate the concept with the (normalized) white Gaussian noise process  $g$ , which, in Gelfand’s framework, is succinctly defined by

$$\mathcal{Z}_g(\varphi) = e^{-\frac{1}{2}\|\varphi\|_{L_2(\mathbb{R}^d)}^2}. \quad (20)$$

If we now substitute  $\varphi = \omega_1 \varphi_1$  with the variable  $\omega_1 \in \mathbb{R}$  and define  $g_1 = \langle g, \varphi_1 \rangle$ , we get

$$\mathcal{Z}_g(\omega_1 \varphi_1) = \int_{\mathbb{R}} e^{-j\omega_1 g_1} p(g_1) dg_1 = e^{-\frac{1}{2}\omega_1^2 \|\varphi_1\|_{L_2(\mathbb{R}^d)}^2}$$

which is the characteristic function (in the classical sense) of the scalar random variable  $g_1$ . The PDF of  $g_1$  is obtained by inverse 1-D Fourier transformation, which yields

$$p(g_1) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{g_1^2}{2\sigma^2}}$$

with  $\sigma^2 = \|\varphi_1\|_{L_2(\mathbb{R}^d)}^2$ . This clearly shows that all first-order densities of the process are Gaussian. Similarly, to derive its second order statistics, we substitute  $\varphi = \omega_1 \varphi_1 + \omega_2 \varphi_2$  in (20) with  $g_1$  as above and  $g_2 = \langle g, \varphi_2 \rangle$ ; this produces the 2-D Gaussian-type characteristic function

$$\mathcal{Z}_g(\omega_1 \varphi_1 + \omega_2 \varphi_2) = \int_{\mathbb{R}^2} e^{-j(\omega_1 g_1 + \omega_2 g_2)} p(g_1, g_2) dg_1 dg_2 = e^{-\frac{1}{2}\boldsymbol{\omega}^T \mathbf{C} \boldsymbol{\omega}} \quad (21)$$

with  $\boldsymbol{\omega} = (\omega_1, \omega_2) \in \mathbb{R}^2$  and

$$\mathbf{C} = \begin{bmatrix} \langle \varphi_1, \varphi_1 \rangle_{L_2} & \langle \varphi_1, \varphi_2 \rangle_{L_2} \\ \langle \varphi_2, \varphi_1 \rangle_{L_2} & \langle \varphi_2, \varphi_2 \rangle_{L_2} \end{bmatrix}.$$



By taking the 2-D inverse Fourier transformation of (21), we readily derive the PDF of  $p(g_1, g_2) = \mathcal{N}(\mathbf{0}, \mathbf{C})$ , which is zero-mean Gaussian with covariance matrix  $\mathbf{C}$ . This result also yields the covariance form of the white Gaussian noise process

$$\mathcal{B}_g(\varphi_1, \varphi_2) = \mathcal{E}\{g_1 g_2\} = [\mathbf{C}]_{12} = \langle \varphi_1, \varphi_2 \rangle_{L_2(\mathbb{R}^d)}$$

More generally, based on the central equality in (21) which is valid for any process (including non-Gaussian ones), we invoke the general moment generating properties of the Fourier transform to relate the covariance form of a process  $s$  to a second-order derivative of its characteristic form:

$$\mathcal{B}_s(\varphi_1, \varphi_2) = (j)^2 \left. \frac{\partial^2 \mathcal{Z}_s(\omega_1 \varphi_1 + \omega_2 \varphi_2)}{\partial \omega_1 \partial \omega_2} \right|_{\omega_1=0, \omega_2=0}. \quad (22)$$

The generalized stochastic process  $w$  is called normalized white noise iff. its covariance properties are described by the simplest possible bilinear form  $\mathcal{B}_w(\varphi_1, \varphi_2) = \langle \varphi_1, \varphi_2 \rangle_{L_2}$ . Note that such a process need not be Gaussian and that it will essentially take independent values at every point<sup>9</sup>. To see this, we may select  $\varphi_2(\mathbf{x}) = \varphi_1(\mathbf{x} - \mathbf{x}_0)$  and consider a series of contracting functions  $\varphi_1$  converging to a Dirac impulse. In the limit, the correlation form will tend to  $\mathcal{B}_w(\delta(\cdot), \delta(\cdot - \mathbf{x}_0)) = \delta(\mathbf{x}_0)$  which is entirely localized at the origin and zero elsewhere.

### B. Linear transformation of a generalized process

While the characteristic form may look intimidating on first encounter, it is a powerful tool that greatly simplifies the characterization of derived processes that are obtained by linear transformation of the primary ones, including the cases where the operator is highly singular (e.g., derivatives). Specifically, let  $T$  be a linear operator whose action over Schwarz's space of tempered distributions ( $\mathcal{S}'(\mathbb{R}^d)$ ) is specified using a standard duality formulation

$$\forall u \in \mathcal{S}'(\mathbb{R}^d), \varphi \in \mathcal{S}(\mathbb{R}^d), \quad \langle Tu, \varphi \rangle = \langle u, T^* \varphi \rangle, \quad (23)$$

where  $\mathcal{S}'(\mathbb{R}^d)$  is the topological dual of  $\mathcal{S}(\mathbb{R}^d)$ . The key point in such a definition is to make sure that the adjoint operator  $T^*$  is such that it maps a test function  $\varphi \in \mathcal{S}$  into another test function  $T^* \varphi \in \mathcal{S}$ ; otherwise, the space  $\mathcal{S}$  of test functions needs to be modified accordingly. Then, it is a trivial matter to obtain the characteristic form of the transformed generalized stochastic process

$$\mathcal{Z}_{(Ts)}(\varphi) = \mathcal{Z}_s(T^* \varphi)$$

where we have used the adjoint's definition  $\langle Ts, \varphi \rangle = \langle s, T^* \varphi \rangle$  to move the operator onto the test function.

For instance, we can apply such an operator  $T$  to white Gaussian noise to obtain a generalized "colored" version of a noise process:

$$\mathcal{Z}_{Tw}(\varphi) = \mathcal{Z}_w(T^* \varphi) = e^{-\frac{1}{2} \|T^* \varphi\|_{L_2(\mathbb{R}^d)}^2}.$$

The noise will obviously remain white if and only if  $T$  (or, equivalently,  $T^*$ ) is norm preserving over  $L_2(\mathbb{R}^d)$ , which is equivalent to  $T$  being unitary. Note that this condition is fulfilled by the Fourier transform (up to some normalization factor), which proves that the Fourier transform of white noise is necessarily white as well.

<sup>9</sup>this is obviously a loose statement: white noise is discontinuous everywhere and there is no hope in trying to specify its samples in the traditional pointwise sense.

It is not hard to show that the correlation form of the linearly transformed noise process is

$$\mathcal{B}_{T^*w}(\varphi_1, \varphi_2) = \langle T^* \varphi_1, T^* \varphi_2 \rangle_{L_2(\mathbb{R}^d)} = \langle \varphi_1, (TT^*) \varphi_2 \rangle_{L_2(\mathbb{R}^d)},$$

where we observe that  $(TT^*) = \text{Id}$  iff.  $T$  is unitary.

### C. Generalized Power spectrum

By restricting ourselves to the class of linear, shift-invariant operations  $Tw = h * w$  where  $h$  is a suitable multidimensional convolution kernel and  $w$  is (non-necessarily Gaussian) white noise, we can use this transformation mechanism to generate an extended class of stationary processes. The corresponding correlation form is given by

$$\begin{aligned} \mathcal{B}_{(h*w)}(\varphi_1, \varphi_2) &= \langle h^\vee * \varphi_1, h^\vee * \varphi_2 \rangle_{L_2(\mathbb{R}^d)} \\ &= \int_{\mathbb{R}^d} \hat{\varphi}_1(\boldsymbol{\omega}) \hat{\varphi}_2^*(\boldsymbol{\omega}) |H(\boldsymbol{\omega})|^2 \frac{d\boldsymbol{\omega}}{(2\pi)^d} \\ &= \int_{\mathbb{R}^d} \hat{\varphi}_1(\boldsymbol{\omega}) \hat{\varphi}_2^*(\boldsymbol{\omega}) \Phi(\boldsymbol{\omega}) \frac{d\boldsymbol{\omega}}{(2\pi)^d} \end{aligned}$$

with  $h^\vee(\boldsymbol{x}) = h(-\boldsymbol{x})$  and  $\Phi(\boldsymbol{\omega}) = |H(\boldsymbol{\omega})|^2$ . Here,  $\Phi(\boldsymbol{\omega})$  is an extension of the classical power spectrum that remains valid when  $H(\boldsymbol{\omega})$  is not square-integrable. For instance,  $\Phi(\boldsymbol{\omega}) = 1$  corresponds to the case of white noise (e.g.  $h = \delta$ ). The filter  $h$  has the same spectral shaping role as in the classical theory of stochastic processes with the advantage that it is less constrained.

## APPENDIX II: PROOF OF THEOREM 1

The goal is to derive the characteristic form (8) starting from the explicit representation of the Poisson process (5). To that end, we select an arbitrary infinitely differentiable test function  $\varphi(\boldsymbol{x})$  of compact support, with its support included in, say, a centred cube  $\Pi_\varphi = [-c_\varphi, c_\varphi]^d$ . We denote by  $N_{w,\varphi}$  the number of Poisson points of  $w$  in  $\Pi_\varphi$ ; by definition, it is a Poisson random variable with parameter  $\lambda \text{Vol}(\Pi_\varphi)$ . The restriction of  $w$  to  $\Pi_\varphi$  corresponds to the random sum

$$\sum_{n=1}^{N_{w,\varphi}} a'_n \delta(\boldsymbol{x} - \boldsymbol{x}'_n),$$

using an appropriate relabeling of the variables  $\{(a_k, \boldsymbol{x}_k) | \boldsymbol{x}_k \in \Pi_\varphi\}$  in (5); correspondingly, we have  $\langle w, \varphi \rangle = \sum_{n=1}^{N_{w,\varphi}} a'_n \varphi(\boldsymbol{x}'_n)$ .

By the order statistics property of Poisson processes, the  $\boldsymbol{x}'_n$  are independent and all equivalent in distribution to a random variable  $\boldsymbol{x}'$  that is uniform on  $\Pi_\varphi$ .

Using the law of total expectation, we expand the characteristic functional of  $w$ ,  $\mathcal{Z}_w(\varphi) = \mathcal{E} \{ e^{j\langle w, \varphi \rangle} \}$ , as

$$\begin{aligned} \mathcal{Z}_w(\varphi) &= \mathcal{E} \left\{ \mathcal{E} \left\{ e^{j\langle w, \varphi \rangle} \middle| N_{w,\varphi} \right\} \right\} \\ &= \mathcal{E} \left\{ \mathcal{E} \left\{ \prod_{n=1}^{N_{w,\varphi}} e^{j a'_n \varphi(\boldsymbol{x}'_n)} \middle| N_{w,\varphi} \right\} \right\} \\ &= \mathcal{E} \left\{ \prod_{n=1}^{N_{w,\varphi}} \mathcal{E} \left\{ e^{j a' \varphi(\boldsymbol{x}')} \right\} \right\} \quad (\text{by independence}) \end{aligned}$$

$$\begin{aligned}
 &= \mathcal{E} \left\{ \prod_{n=1}^{N_{w,\varphi}} \mathcal{E} \left\{ \mathcal{E} \left\{ e^{ja\varphi(\mathbf{x}')} \middle| a \right\} \right\} \right\} && \text{(total expectation)} \\
 &= \mathcal{E} \left\{ \prod_{n=1}^{N_{w,\varphi}} \mathcal{E} \left\{ \frac{\int_{\Pi_\varphi} e^{ja\varphi(\mathbf{x}')} d\mathbf{x}'}{\text{Vol}(\Pi_\varphi)} \right\} \right\} && \text{(as } \mathbf{x}' \text{ is uniform in } \Pi_\varphi) \\
 &= \mathcal{E} \left\{ \prod_{n=1}^{N_{w,\varphi}} \frac{\int_{\mathbb{R}} \int_{\Pi_\varphi} e^{ja\varphi(\mathbf{x})} d\mathbf{x} dP(a)}{\text{Vol}(\Pi_\varphi)} \right\} && (24)
 \end{aligned}$$

The last expression has the inner expectation expanded in terms of the distribution  $P(a)$  of the random variable  $a$ . Defining the auxiliary functional

$$M(\varphi) = \int_{\mathbb{R}} \int_{\Pi_\varphi} e^{ja\varphi(\mathbf{x})} d\mathbf{x} dP(a),$$

we rewrite (24) as

$$\mathcal{E} \left\{ \prod_{n=1}^{N_{w,\varphi}} \frac{M(\varphi)}{\text{Vol}(\Pi_\varphi)} \right\} = \mathcal{E} \left\{ \left[ \frac{M(\varphi)}{\text{Vol}(\Pi_\varphi)} \right]^{N_{w,\varphi}} \right\}.$$

Next, we use the fact that  $N_{w,\varphi}$  is a Poisson random variable to compute the above expectation directly:

$$\begin{aligned}
 \mathcal{E} \left\{ \left[ \frac{M(\varphi)}{\text{Vol}(\Pi_\varphi)} \right]^{N_{w,\varphi}} \right\} &= \sum_{n \geq 0} \left[ \frac{M(\varphi)}{\text{Vol}(\Pi_\varphi)} \right]^n \frac{e^{-\lambda \text{Vol}(\Pi_\varphi)} [\lambda \text{Vol}(\Pi_\varphi)]^n}{n!} \\
 &= e^{-\lambda \text{Vol}(\Pi_\varphi)} \sum_{n \geq 0} \frac{[\lambda M(\varphi)]^n}{n!} \\
 &= e^{-\lambda \text{Vol}(\Pi_\varphi)} e^{\lambda M(\varphi)} && \text{(Taylor)} \\
 &= \exp[\lambda (M(\varphi) - \text{Vol}(\Pi_\varphi))].
 \end{aligned}$$

We now replace  $M(\varphi)$  by its integral equivalent, noting also that  $\text{Vol}(\Pi_\varphi) = \int_{\mathbb{R}} \int_{\Pi_\varphi} 1 d\mathbf{x} dP(a)$ , whereupon we obtain the expression

$$\mathcal{Z}_w(\varphi) = \exp\left[\lambda \int_{\mathbb{R}} \int_{\Pi_\varphi} (e^{ja\varphi(\mathbf{x})} - 1) d\mathbf{x} dP(a)\right].$$

As  $e^{ja\varphi(\mathbf{x})} - 1$  vanishes outside the support of  $\varphi(\mathbf{x})$  (and, therefore, outside  $\Pi_\varphi$ ), we may enlarge the domain of the inner integral to all of  $\mathbb{R}^d$ , yielding (8). Finally, we evoke a density/continuity argument to extend the result to the functions of the Schwartz class that are not compactly supported. ■

### APPENDIX III: PROOF OF THEOREM 2

First, we prove that  $(-\Delta)_0^{\frac{\gamma}{2}*} \varphi \in L_2(\mathbb{R}^d)$  for  $H = \gamma - \frac{d}{2} \in \mathbb{R}^+ \setminus \mathbb{N}$ .

The condition  $\varphi \in L_1(\mathbb{R}^d, w_N)$  ensures that the moments,  $\int_{\mathbb{R}^d} \varphi(\mathbf{x}) \mathbf{x}^{\mathbf{k}} d\mathbf{x} = j^{|\mathbf{k}|} \hat{\varphi}^{\mathbf{k}}(\mathbf{0})$ , are finite up to order  $N$ . More generally, it implies that  $\hat{\varphi}$  and its derivatives up to order  $N$  are uniformly bounded. The auxiliary function  $f(\boldsymbol{\omega}) = \hat{\varphi}(\boldsymbol{\omega}) - \sum_{|\mathbf{k}|=0}^{\lfloor H \rfloor} \hat{\varphi}^{\mathbf{k}}(\mathbf{0}) \frac{\boldsymbol{\omega}^{\mathbf{k}}}{\mathbf{k}!}$  is therefore well-defined, and the task reduces to showing that  $\left\| \hat{f}(\boldsymbol{\omega}) / \|\boldsymbol{\omega}\|^\gamma \right\|_{L_2(\mathbb{R}^d)} < +\infty$ .

To guarantee square integrability of the singularity of  $\hat{f}(\boldsymbol{\omega}) / \|\boldsymbol{\omega}\|^\gamma$  at the origin, we must make sure that  $\hat{f}(\boldsymbol{\omega}) = O(\|\boldsymbol{\omega}\|^{\gamma+\gamma'})$  with  $\gamma' > -d/2$  as  $\boldsymbol{\omega}$  tends to  $\mathbf{0}$ . Since  $\hat{\varphi}$  is sufficiently regular for its  $N$ th order Taylor series to be well-defined, we have that  $\hat{f}(\boldsymbol{\omega}) = O(\|\boldsymbol{\omega}\|^N)$  where  $N = \lfloor H \rfloor + 1$  so that the condition is automatically satisfied.

To establish  $L_2$ -integrability over the rest of the domain, we invoke the triangle inequality. The delicate aspect there is the decay at infinity of the elementary signals  $\frac{\omega^k}{\|\omega\|^\gamma} = O(\|\omega\|^{\gamma'})$  with  $\gamma' = |\mathbf{k}| - H - \frac{d}{2}$  for  $|\mathbf{k}| \leq \lfloor H \rfloor$ . The strict requirement is that  $\gamma' < -d/2$ , which is guaranteed for  $H$  non-integer, but not otherwise. ■

This boundedness result ensures that the Gaussian characteristic form (13) is well-defined. As for the Poisson functional, we can transfer the Gaussian bound to the real part of the argument in the exponential function in (14). Specifically, we have that

$$\operatorname{Re}(u_{\lambda, P}(L^{-1*}\varphi)) = \lambda \int_{\mathbb{R}} \int_{\mathbb{R}^d} [\cos(aL^{-1*}\varphi(\mathbf{x})) - 1] d\mathbf{x} dP(a),$$

which is bounded by

$$\int_{\mathbb{R}} \int_{\mathbb{R}^d} |\cos(aL^{-1*}\varphi(\mathbf{x})) - 1| d\mathbf{x} dP(a) < \int_{\mathbb{R}} \int_{\mathbb{R}^d} a^2 |L^{-1*}\varphi(\mathbf{x})|^2 dP(a) = \mathcal{E}(a^2) \|L^{-1*}\varphi\|_{L_2}^2,$$

based on the inequality  $|\cos(y) - 1| < y^2$ . This takes care entirely of the symmetric Poisson case.

We construct a similar bound for the imaginary part using the inequality  $|\sin(y)| \leq \min(|y|, 1)$ . It will be finite whenever  $g = L^{-1*}\varphi \in L_1(\mathbb{R}^d)$ , or more generally, if  $g(\mathbf{x})$  decays like  $\|\mathbf{x}\|^{-\gamma'}$  with  $\gamma' > d$  as  $\omega$  goes to infinity. In order to complete the proof for the non-even case, one needs to show that  $L^{-1*}\varphi$  meets the required conditions, which is presently left as an open issue. The problem is easily overcome when the moments of  $\varphi$  are zero up to order  $N + d$ , but this is probably too restrictive a condition.

Upon the completion of this work, we came up with an alternative approach where we further regularize the inverse operator by including higher-order correction terms in (15) to ensure that  $g = L^{-1*}\varphi \in L_1(\mathbb{R}^d)$  for all  $\varphi \in \mathcal{S}$  [35]. A remarkable finding is that the combination of scale-invariance and  $L_p$ -stability uniquely specifies the inverse, and that full  $L_1$ -stability is generally not compatible with  $L_2$ -stability.

#### ACKNOWLEDGEMENTS

The key ideas for this work originated during the preparation of a talk for Martin Vetterli's 50th birthday celebration, and the authors are pleased to dedicate this paper to him. The research was supported by the Swiss National Science Foundation under Grant 200020-109415.

#### REFERENCES

- [1] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw-Hill, 1991.
- [2] R. Gray and L. Davison, *An introduction to statistical signal processing*. Cambridge University Press, 2004.
- [3] I. Daubechies, "Orthogonal bases of compactly supported wavelets," *Comm. Pure Appl. Math.*, vol. 41, pp. 909–996, 1988.
- [4] S. G. Mallat, "A theory of multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [5] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: An overview," *IEEE Transactions on Consumer Electronics*, vol. 16, no. 4, pp. 1103–1127, 2000.
- [6] J. B. Weaver, X. Yansun, J. Healy, D. M., and L. D. Cromwell, "Filtering noise from images with wavelet transforms," *Magnetic Resonance in Medicine*, vol. 21, no. 2, pp. 288–95, 1991.
- [7] D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [8] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, 1994.
- [9] S. Mallat, *A wavelet tour of signal processing*. San Diego: Academic Press, 1998.
- [10] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

- [11] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Review*, vol. 51, no. 1, pp. 34–81, 2009.
- [12] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal  $\ell_1$ -norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.
- [13] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.
- [14] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [15] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Transactions on Signal Processing*, vol. 50, no. 6, pp. 1417–1428, June 2002.
- [16] T. Blu, P. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse Sampling of Signal Innovations," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 31–40, 2008.
- [17] P. Dragotti, M. Vetterli, and T. Blu, "Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang-Fix," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1741–1757, May 2007, part 1.
- [18] I. Gelfand and N. Y. Vilenkin, *Generalized Functions. Vol. 4. Applications of Harmonic Analysis*. New York, USA: Academic press, 1964.
- [19] R. A. Devore, "Nonlinear approximation," *Acta Numerica*, vol. 7, pp. 51–150, 1998.
- [20] M. Unser and T. Blu, "Generalized smoothing splines and the optimal discretization of the Wiener filter," *IEEE Transactions on Signal Processing*, vol. 53, no. 6, pp. 2146–2159, June 2005.
- [21] Y. Meyer, *Ondelettes et opérateurs I: Ondelettes*. Paris, France: Hermann, 1990.
- [22] A. Chambolle, R. DeVore, N.-Y. Lee, and B. Lucier, "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *IEEE Transactions on Image Processing*, vol. 7, no. 33, pp. 319–335, 1998.
- [23] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, no. 1-4, pp. 259–268, 1992.
- [24] E. Mammen and S. van de Geer, "Locally adaptive regression splines," *Annals of Statistics*, vol. 25, no. 1, pp. 387–413, 1997.
- [25] M. Raphan and E. P. Simoncelli, "Optimal denoising in redundant representations," *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1342–1352, 2008.
- [26] A. Kolmogoroff, "La transformation de Laplace dans les espaces linéaires," *C. R. Acad. Sci. Paris*, vol. 200, pp. 1717–1718, 1935, note of M. A. Kolmogoroff, presented by M. Jacques Hadamard.
- [27] L. Schumaker, *Spline functions: Basic theory*. New York: Wiley, 1981.
- [28] J. Rice, "Generalized shot noise," *Advances in Applied Probability*, vol. 9, no. 3, pp. 553–565, 1977.
- [29] P. Brémaud and L. Massoulié, "Power spectra of general shot noises and Hawkes point processes with a random excitation," *Advances in Applied Probability*, vol. 34, no. 1, pp. 205–222, 2002.
- [30] M. Unser and T. Blu, "Self-similarity: Part I—Splines and operators," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1352–1363, April 2007.
- [31] J. Duchon, "Splines minimizing rotation-invariant semi-norms in sobolev spaces," in *Constructive Theory of Functions of Several Variables*, W. Schempp and K. Zeller, Eds. Berlin: Springer-Verlag, 1977, pp. 85–100.
- [32] 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, April 2009.
- [33] T. Blu and M. Unser, "Self-similarity: Part II—Optimal estimation of fractal processes," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1364–1378, April 2007.
- [34] M. Taqqu, "Fractional Brownian motion and long-range dependence," in *Theory and applications of long-range dependence*, P. Doukhan, G. Oppenheim, and M. Taqqu, Eds. Boston, USA: Birkhauser, 2003, pp. 5–38.
- [35] Q. Sun and M. Unser, "Left inverses of fractional Laplacian and sparse stochastic processes," *Advances in Computational Mathematics*, in press.
- [36] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, 1968.
- [37] E. Perrin, R. Harba, C. Berzin-Joseph, I. Iribarren, and A. Bonami, " $n$ th-order fractional Brownian motion and fractional Gaussian noises," *IEEE Trans. Signal Processing*, vol. 49, no. 5, pp. 1049–1059, 2001.
- [38] B. B. Mandelbrot, *Gaussian Self-Affinity and Fractals*. Berlin: Springer, 2001.
- [39] ———, *The fractal geometry of nature*. San Francisco: Freeman, 1982.
- [40] D. M. Roy and Y. W. Teh, "The Mondrian process," in *Advances in Neural Information Processing Systems 21*, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Eds., 2009, pp. 1377–1384.

- [41] C. Chui and J. Wang, "A cardinal spline approach to wavelets," *Proc. Amer. Math. Soc.*, vol. 113, no. 3, pp. 785–793, 1991.
- [42] M. Unser and T. Blu, "Fractional splines and wavelets," *SIAM Review*, vol. 42, no. 1, pp. 43–67, March 2000.
- [43] T. Blu and M. Unser, "A complete family of scaling functions: The  $(\alpha, \tau)$ -fractional splines," in *Proceedings of the Twenty-Eighth IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, vol. VI, Hong Kong SAR, People's Republic of China, April 6-10, 2003, pp. 421–424.
- [44] D. Van De Ville, B. Forster-Heinlein, M. Unser, and T. Blu, "Analytical footprints: Compact representation of elementary singularities in wavelet bases," *IEEE Transactions on Image Processing*, submitted.
- [45] D. Van De Ville, T. Blu, and M. Unser, "Isotropic polyharmonic B-splines: Scaling functions and wavelets," *IEEE Trans. Image Processing*, vol. 14, no. 11, pp. 1798–1813, November 2005.
- [46] I. Khalidov, D. Van De Ville, T. Blu, and M. Unser, "Construction of wavelet bases that mimic the behaviour of some given operator," in *Proceedings of the SPIE Conference on Mathematical Imaging: Wavelet XII*, vol. 6701, San Diego CA, USA, August 26-29, 2007, pp. 67 010S–1–67 010S–7.
- [47] I. Khalidov and M. Unser, "From differential equations to the construction of new wavelet-like bases," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1256–1267, April 2006.
- [48] C. Vonesch, T. Blu, and M. Unser, "Generalized Daubechies wavelet families," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4415–4429, September 2007.
- [49] J. Verhaeghe, D. Van De Ville, I. Khalidov, Y. D'Asseler, I. Lemahieu, and M. Unser, "Dynamic PET reconstruction using wavelet regularization with adapted basis functions," *IEEE Transactions on Medical Imaging*, vol. 27, no. 7, pp. 943–959, 2008.
- [50] I. Khalidov, D. Van De Ville, J. Fadili, and M. Unser, "Activelets and sparsity: A new way to detect brain activation from fMRI data," in *Proceedings of the SPIE Conference on Mathematical Imaging: Wavelet XII*, vol. 6701, San Diego CA, USA, August 26-29, 2007, pp. 67 010Y–1–67 010Y–8.