A Unified Formulation of Gaussian Versus Sparse Stochastic Processes—Part I: Continuous-Domain Theory

Michael Unser, *Fellow, IEEE*, Pouya D. Tafti, and Qiyu Sun

***Abstract*— We introduce a general distributional framework that results in a unifying description and characterization of a rich variety of continuous-time stochastic processes. The cornerstone of our approach is an innovation model that is driven by some generalized white noise process, which may be Gaussian or not (e.g., Laplace, impulsive Poisson, or alpha stable). This allows for a conceptual decoupling between the correlation properties of the process, which are imposed by the whitening operator L, and its sparsity pattern, which is determined by the type of noise excitation. The latter is fully specified by a Lévy measure. We show that the range of admissible innovation behavior varies between the purely Gaussian and super-sparse extremes. We prove that the corresponding generalized stochastic processes are well-defined mathematically provided that the (adjoint) inverse of the whitening operator satisfies some L_p bound for $p \geq 1$. We present a novel operator-based method that yields an explicit characterization of all Lévy-driven processes that are solutions of constant-coefficient stochastic differential equations. When the underlying system is stable, we recover the family of stationary continuous-time autoregressive moving average processes (CARMA), including the Gaussian ones. The approach remains valid when the system is unstable and leads to the identification of potentially useful generalizations of the Lévy processes, which are sparse and non-stationary. Finally, we show that these processes admit a sparse representation in some matched wavelet domain and provide a full characterization of their transform-domain statistics.**

***Index Terms*— Sparsity, non-Gaussian stochastic processes, innovation modeling, continuous-time signals, stochastic differential equations, wavelet expansion, Lévy process, infinite divisibility**

I. Introduction

In RECENT years, the research focus in signal processing has shifted away from the classical linear paradigm, which is intimately linked with the theory of stationary Gaussian processes [1], [2]. Instead of considering Fourier transforms and performing quadratic optimization, researchers are presently favoring wavelet-like representations and have adopted sparsity as design paradigm [3]–[8]. The property that a signal admits a sparse expansion can be exploited elegantly for compressive sensing, which is presently a very active area of research (cf. special issue of the Proceedings of the IEEE [9], [10]). The concept is equally helpful for solving inverse problems and has resulted in significant algorithmic advances for the efficient resolution of large scale ℓ_1-norm minimization problems [11]–[13].

The current formulations of compressed sensing and sparse signal recovery are fundamentally deterministic. Also, they are predominantly discrete and based on finite-dimensional mathematics, with the notable exception of the works of Eldar [14], Adcock and Hansen [15]. By drawing on the analogy with the classical theory of signal processing, it is likely that further progress may be achieved by adopting a statistical (or estimation theoretic) point of view for the description of sparse signals in the analog domain. This stands as our primary motivation for the investigation of the present class of continuous-time stochastic processes, the greater part of which is sparse by construction. These processes are specified as a superset of the Gaussian ones, which is essential for maintaining backward compatibility with traditional statistical signal processing.

The present construction is a generalization of a classical idea in communication theory and signal processing which is to view a stochastic process as filtered version of a white noise (a.k.a. innovation) [16]. The fundamental aspect here is that the modeling is done in the continuous domain, which, as we shall see, imposes strong constraints on the class of admissible innovations; that is, the generalized white noise that constitutes the input of the innovation model. The second ingredient is a powerful operational calculus (the generalization of the idea of filtering) for solving stochastic differential equations (SDE), including unstable ones, which is essential for inducing interesting (non-stationary) behaviors such as self-similarity. The combination of these ingredients results in the specification of an extended class of stochastic processes that are either Gaussian or sparse, at the exclusion of any other type. The proposed theory has a unifying character in that it connects a number of contemporary topics in signal processing, statistics and approximation theory:

- sparsity (in relation to compressed sensing) [3], [4]

Manuscript received September 21, 2012; revised October 7, 2013; accepted December 13, 2013. Date of publication January 9, 2014; date of current version February 12, 2014. This work was supported in part by the Swiss National Science Foundation under Grant 200020-144355, in part by the European Research Counsel under Grant ERC-2010-AdG-267439-FUNSP, and in part by the National Science Foundation under Grant DMS 1109063.

M. Unser and P. D. Tafti are with the Biomedical Imaging Group, École Polytechnique Fédérale de Lausanne, Lausanne CH-1015, Switzerland (e-mail: michael.unser@epfl.ch; pouya.tafti@epfl.ch).

Q. Sun is with the Department of Mathematics, University of Central Florida, Orlando, FL 32816 USA (e-mail: qiyu.sun@ucf.edu).

Communicated by V. Borkar, Associate Editor for Communication Networks.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2014.2298453

0018-9448 © 2014 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

- signals with a finite rate of innovation [17], [18]
- the classical theory of Gaussian stationary processes [1], [16]
- non-Gaussian continuous-domain modeling of signals [19], [20]
- stochastic differential equations [21], [22]
- splines, wavelets and linear system theory [5], [23].

Most importantly, it explains why certain classes of processes admit a sparse representation in a matched wavelet-like basis (see introductory example in Section II where the Haar transform outperfoms the classical Karhunen-Loève transform). Since these models are the natural functional extension of the Gaussian stationary processes, they may stimulate the development of novel algorithms for statistical signal processing. This has already been demonstrated in the context of biomedical image reconstruction [24], the derivation of statistical priors for discrete-domain signal representation [25], optimal signal denoising [26], and MMSE interpolation [27].

Because the proposed model is intrinsically linear, we have adopted a formulation that relies on generalized functions, rather than the traditional mathematical concepts (random measures and Itô integrals) from the theory of stochastic differential equations [21], [22], [28]. We are then taking advantage of the theory of generalized stochastic processes of Gelfand (arguably, the second most famous Soviet mathematician after Kolmogorov) and some powerful tools of functional analysis (Minlos-Bochner's theorem) [29] that are not widely known to engineers nor statisticians. While this may look like an unnecessary abstraction at first sight, it is very much in line with the intuition of an engineer who prefers to work with analog filters and convolution operators rather than with stochastic integrals. We are then able to use the whole machinery of linear system theory and the power of the characteristic functional to derive the statistics of the signal in any (linearly) transformed domain.

The paper is organized as follows. The basic flavor of the innovation model is conveyed in Section II by focusing on a first-order differential system which results in the generation of Gaussian and non-Gaussian AR(1) stochastic processes. We use of this model to illustrate that a properly-matched wavelet transform can outperform the classical Karhunen-Loève transform (or the DCT) for the compression of (non-Gaussian) signals. In Section III, we review the foundations of Gelfand's theory of generalized stochastic processes. In particular, we characterize the complete class of admissible continuous-time white noise processes (innovations) and give some argumentation as to why the non-Gaussian brands are inherently sparse. In Section IV, we give a high-level description of the general innovation model and provide a novel operator-based method for the solution of SDE. In Section V, we make use of Gelfand's formalism to fully characterize our extended class of (non-Gaussian) stochastic processes including the special cases of CARMA and Nth-order generalized Lévy processes. We also derive the statistics of the wavelet-domain representation of these signals, which allows for a common (stationary) treatment of the two latter classes of processes, irrespective of any stability consideration. Finally, in Section VI, we turn back to our introductory example by moving into the unstable regime (single pole at the origin) which yields a non-conventional system-theoretic interpretation of classical Lévy processes [28], [30], [31]. We also point out the structural similarity between the increments of Lévy processes and their Haar wavelet coefficients. For higher-order illustrations of sparse processes, we refer to our companion paper [32], which is specifically devoted to the study of the discrete-time implication of the theory and the way to best decouple (e.g. "sparsify") such processes. The notation, which is common to both papers, is summarized in [32, Table II].

II. Motivation: Gaussian vs. Non-Gaussian AR(1) Processes

A continuous-time Gaussian AR(1) (or Gauss-Markov) process can be formally generated by applying a first-order analog filter to a Gaussian white noise process w:

$$s_\alpha(t) = (\rho_\alpha * w)(t) \tag{1}$$

where $\rho_\alpha(t) = \mathbb{1}_+(t)e^{\alpha t}$ with $\mathrm{Re}(\alpha) < 0$ and $\mathbb{1}_+(t)$ is the unit-step function. Next, we observe that $\rho_\alpha = (\mathrm{D} - \alpha\mathrm{Id})^{-1}\delta$ where δ is the Dirac impulse and where $\mathrm{D} = \frac{\mathrm{d}}{\mathrm{d}t}$ and Id are the derivative and identity operators, respectively. These operators as well as the inverse are to be interpreted in the distributional sense (see Section III-A). This suggests that s_α satisfies the "innovation" model (cf. [1], [16])

$$(\mathrm{D} - \alpha\mathrm{Id})s_\alpha(t) = w(t), \tag{2}$$

or, equivalently, the stochastic differential equation (cf. [22])

$$\mathrm{d}s_\alpha(t) - \alpha s_\alpha(t)\mathrm{d}t = \mathrm{d}W(t),$$

where $W(t) = \int_0^t w(\tau)\mathrm{d}\tau$ is a standard Brownian motion (or Wiener process) excitation. In the statistical literature, the solution of the above first-order SDE is often called the Ornstein-Uhlenbeck process.

Let $(s_\alpha[k] = s_\alpha(t)|_{k=t})_{k\in\mathbb{Z}}$ denote the sampled version of the continuous-time process. Then, one can show that $s_\alpha[\cdot]$ is a discrete AR(1) autoregressive process that can be whitened by applying the first-order linear predictor:

$$s_\alpha[k] - e^\alpha s_\alpha[k-1] = u[k] \tag{3}$$

where $u[\cdot]$ (prediction error) is an i.i.d. Gaussian sequence. Alternatively, one can decorrelate the signal by computing its discrete cosine transform (DCT), which is known to be asymptotically equivalent to the Karhunen-Loève transform (KLT) of the process [33], [34]. Eq. (3) provides the basis for classical linear predictive coding (LPC), while the decorrelation property of the DCT is often invoked to justify the popular JPEG transform-domain coding scheme [35].

In this paper, we are concerned with the non-Gaussian counterpart of this story, which, as we shall see, will result in the identification of sparse processes. The idea is to retain the simplicity of the classical innovation model, while substituting the continuous-time Gaussian noise by some generalized Lévy innovation (to be properly defined in the sequel). This translates into Eqs. (1)–(3) remaining valid, except that the underlying random variates are no longer Gaussian. The more

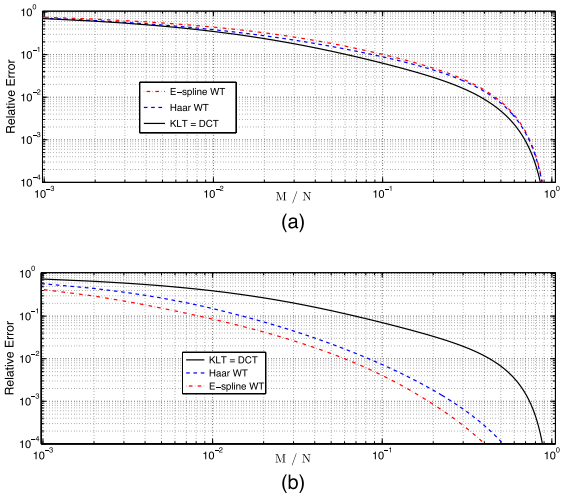

Fig. 1. Wavelets vs. KLT (or DCT) for the M-term approximation of Gaussian vs. sparse AR(1) processes with $\alpha = -0.1$: (a) classical Gaussian scenario, (b) sparse scenario with symmetric Cauchy innovations. The E-spline wavelets are matched to the innovation model. The displayed results (relative quadratic error as a function of M/N) are averages over 1000 realizations for AR(1) signals of length $N = 1024$; the performance of DCT and KLT is undistinguishable.

significant finding is that the KLT (or its discrete approximation by the DCT) is no longer optimal for producing the best M-term approximation of the signal. This is illustrated in Fig. 1, which compares the performance of various transforms for the compression of two kinds of AR(1) processes with correlation $e^{-0.1} \approx 0.90$: Gaussian vs. sparse where the latter innovation follows a Cauchy distribution. The key observation is that the E-spline wavelet transform, which is matched to the operator $\mathrm{L} = \mathrm{D} - \alpha\mathrm{Id}$, provides the best results in the non-Gaussian scenario over the whole range of experimentation [cf. Fig. 1(b)], while the outcome in the Gaussian case is as predicted by the classical theory with the KLT being superior. Examples of orthogonal E-spline wavelets at two successive scales are shown in Fig. 2 next to their Haar counterparts. We selected the E-spline wavelets because of their ability to decouple the process which follows from their operator-like behavior: $\psi_i = \mathrm{L}^*\phi_i$ where i is the scale index and ϕ_i a suitable smoothing kernel [36, Theorem 2]. Unlike their conventional cousins, they are not dilated versions of each other, but rather extrapolations in the sense that the slope of the exponential segments remains the same at all scales. They can, however, be computed efficiently using a perfect reconstruction filterbank with scale-dependent filters [36].

The equivalence with traditional wavelet analysis (Haar) and finite-differencing (as used in the computation of total variation) for signal "sparsification" is achieved by letting $\alpha \to 0$. The catch, however, is that the underlying system becomes unstable! Fortunately, the problem can be fixed, but it calls for an advanced mathematical treatment that is beyond the traditional formulation of stationary processes. The reminder of the paper is devoted to giving a proper sense to what has just been described informally, and to extending the approach to the whole class of ordinary differential operators,

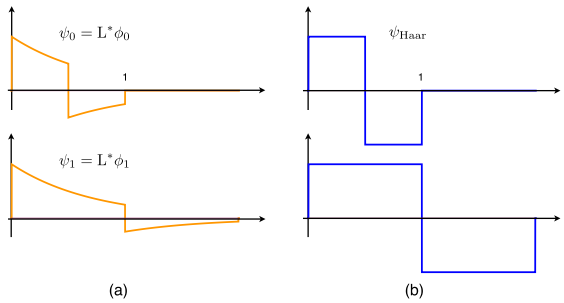

Fig. 2. Comparison of operator-like and conventional wavelet basis functions at two successive scales: (a) first-order E-spline wavelets with $\alpha = -0.5$. (b) Haar wavelets. The vertical axis is rescaled for full range display.

including the non-stable scenarios. The non-trivial outcome, as we shall see, is that many non-stable systems are linked with non-stationary stochastic processes. These, in turn, can be stationarized and "sparsified" by application of a suitable wavelet transformation. The companion paper [32] is focused on the discrete aspects of the theory, including the generalization of (3) for decoupling purposes and the full characterization of the underlying processes.

III. Mathematical Background

The purpose of this section is to introduce the distributional formalism that is required for the proper definition of continuous-time white noise that is the driving term of (1) and its generalization. We start with a brief summary of some required notions in functional analysis, which also serves us to set the notation. We then introduce the fundamental concept of characteristic functional which constitutes the foundation of Gelfand's theory of generalized stochastic processes. We proceed by giving the complete characterization of the possible types of continuous-domain white noises—not necessarily Gaussian—which will be used as universal input for our innovation models. We conclude the section by showing that the non-Gaussian brands of noises that are allowed by Gelfand's formulation are intrinsically sparse, a property that has not been emphasized before (to the best of our knowledge).

A. Functional and Distributional Context

The L_p-norm of a function $f = f(t)$ is $\|f\|_p = \left(\int_{\mathbb{R}} |f(t)|^p \mathrm{d}t\right)^{\frac{1}{p}}$ for $1 \leq p < \infty$ and $\|f\|_\infty = \operatorname{ess\,sup}_{t\in\mathbb{R}} |f(t)|$ for $p = +\infty$ with the corresponding Lebesgue space being denoted by $L_p = L_p(\mathbb{R})$. The concept is extendable for characterizing the rate of decay of functions. To that end, we introduce the weighted $L_{p,\alpha}$ spaces with $\alpha \in \mathbb{R}^+$

$$L_{p,\alpha} = \left\{f \in L_p : \|f\|_{p,\alpha} < +\infty\right\}$$

where the α-weighted L_p-norm of f is defined as

$$\|f\|_{p,\alpha} = \|(1 + |\cdot|^\alpha) f(\cdot)\|_p.$$

Hence, the statement $f \in L_{\infty,\alpha}$ implies that $f(t)$ decays at least as fast as $1/|t|^\alpha$ as t tends to $\pm\infty$; more precisely, that $|f(t)| \leq \frac{\|f\|_{\infty,\alpha}}{1+|t|^\alpha}$ almost everywhere. In particular, this

allows us to infer that $L_{\infty,\frac{1}{p}+\epsilon} \subset L_p$ for any $\epsilon > 0$ and $p \geq 1$. Another obvious inclusion is $L_{p,\alpha} \subseteq L_{p,\alpha_0}$ for any $\alpha \geq \alpha_0$. In the limit, we end up with the space of rapidly-decreasing functions $\mathcal{R} = \left\{f : \|f\|_{\infty,m} < +\infty,\ \forall m \in \mathbb{Z}^+\right\}$, which is included in all the others.[1]

We use $\varphi = \varphi(t)$ to denote a generic function in Schwartz's class $\mathcal{S}$ of rapidly-decaying and infinitely-differentiable test functions. Specifically, Schwartz's space is defined as:

$$\mathcal{S} = \left\{\varphi \in C^\infty : \|\mathrm{D}^n\varphi\|_{\infty,m} < +\infty,\ \forall m, n \in \mathbb{Z}^+\right\},$$

with the operator notation $\mathrm{D}^n = \frac{\mathrm{d}^n}{\mathrm{d}t^n}$ and the convention that $\mathrm{D}^0 = \mathrm{Id}$ (identity). $\mathcal{S}$ is a complete topological vector space with respect to the topology induced by the series of semi-norm $\|\mathrm{D}^n \cdot\|_{\infty,m}$ with $m, n \in \mathbb{Z}^+$. Its topological dual is the space of tempered distributions $\mathcal{S}'$; a distribution $\phi \in \mathcal{S}'$ is a continuous linear functional on $\mathcal{S}$ that is characterized by a duality product rule $\phi(\varphi) = \langle\phi, \varphi\rangle = \int_{\mathbb{R}} \phi(t)\varphi(t)\mathrm{d}t$ with $\varphi \in \mathcal{S}$ where the right-hand side expression has a literal interpretation as an integral only when $\phi(t)$ is true function of t. The prototypical example of a tempered distribution is the Dirac distribution δ, which is defined as $\delta(\varphi) = \langle\delta, \varphi\rangle = \varphi(0)$. In the sequel, we will drop the explicit dependence of the distribution on the generic test function $\varphi \in \mathcal{S}$ and simply write ϕ, $\phi(\cdot)$ or even $\phi(t)$ (with an abuse of notation) where t stands as our generic time index. For instance, we shall denote the shifted Dirac impulse[2] by $\delta(\cdot - t_0)$, or $\delta(t - t_0)$ which is the conventional notation used by engineers.

Let T be a continuous[3] linear operator that maps $\mathcal{S}$ into itself (or eventually some enlarged topological space such as L_p). It is then possible to extend the action of T over $\mathcal{S}'$ (or an appropriate subset of it) based on the definition $\langle\mathrm{T}\phi, \varphi\rangle = \langle\phi, \mathrm{T}^*\varphi\rangle$ for $\phi \in \mathcal{S}'$ if T^*, which is the adjoint of T, maps φ to another test function $\mathrm{T}^*\varphi \in \mathcal{S}$ continuously. An important example is the Fourier transform whose classical definition is $\mathcal{F}\{f\}(\omega) = \hat{f}(\omega) = \int_{\mathbb{R}} f(t)e^{-j\omega t}\mathrm{d}t$. Since $\mathcal{F}$ is a $\mathcal{S}$-continuous operator, it is extendable to $\mathcal{S}'$ based on the adjoint relation $\langle\mathcal{F}\phi, \varphi\rangle = \langle\phi, \mathcal{F}\varphi\rangle$ for all $\varphi \in \mathcal{S}$ (generalized Fourier transform).

A linear, shift-invariant (LSI) operator that is well-defined over $\mathcal{S}$ can always be written as a convolution product:

$$\mathrm{T}_{\mathrm{LSI}}\{\varphi\} = h * \varphi = \int_{\mathbb{R}} h(\tau)\varphi(\cdot - \tau)\mathrm{d}\tau$$

where $h = \mathrm{T}_{\mathrm{LSI}}\{\delta\}$ is the impulse response of the system. The adjoint operator is the convolution with the time-reversed version of h:

$$h^\vee(t) \equiv h(-t).$$

The better-known categories of LSI operators are the BIBO-stable (bounded input, bounded output) filters, and the ordinary differential operators. While the latter are not BIBO-stable, they do work well with test functions.

[1]The topology of $\mathcal{R}$ is defined by the family of semi-norms $\|\cdot\|_{\infty,m}$, $m = 1, 2, 3, \ldots$

[2]The precise definition is $\langle\delta(\cdot - t_0), \varphi\rangle = \varphi(t_0)$ for all $\varphi \in \mathcal{S}$.

[3]An operator T is continuous from a sequential topological vector space $\mathcal{V}$ into another one iff. $\varphi_k \to \varphi$ in the topology of $\mathcal{V}$ implies that $\mathrm{T}\varphi_k \to \mathrm{T}\varphi$ in the topology (or norm) of the second space. If the two spaces coincide, we say that T is $\mathcal{V}$-continuous.

1) L_p-Stable LSI Operators: The BIBO-stable filters correspond to the case where $h \in L_1$, or, more generally, when h corresponds to a complex-valued Borel measure of bounded variation. The latter extension allows for discrete filters of the form $h_d = \sum_{n\in\mathbb{Z}} d[n]\delta(\cdot - n)$ with $d[n] \in \ell_1$. We will refer to these filters as L_p-stable because they specify bounded operators in all the L_p spaces (by Young's inequality). L_p-stable convolution operators satisfy the properties of commutativity, associativity, and distributivity with respect to addition.

2) $\mathcal{S}$-Continuous LSI Operators: For an L_p-stable filter to yield a Schwartz function as output, it is necessary that its impulse response (continuous or discrete) be rapidly-decaying. In fact, the condition $h \in \mathcal{R}$ (which is much stronger than integrability) ensures that the filter is $\mathcal{S}$-continuous. The nth-order derivative D^n and its adjoint $\mathrm{D}^{n*} = (-1)^n\mathrm{D}^n$ are in the same category. The nth-order *weak* derivative of the tempered distribution ϕ is defined as $\mathrm{D}^n\phi(\varphi) = \langle\mathrm{D}^n\phi, \varphi\rangle = \langle\phi, \mathrm{D}^{n*}\varphi\rangle$ for any $\varphi \in \mathcal{S}$. The latter operator—or, by extension, any polynomial of distributional derivatives $P_N(\mathrm{D}) = \sum_{n=1}^N a_n\mathrm{D}^n$ with constant coefficients $a_n \in \mathbb{C}$—maps $\mathcal{S}'$ into itself. The class of these differential operators enjoys the same properties as its classical counterpart: shift-invariance, commutativity, associativity and distributivity.

B. Notion of Generalized Stochastic Process

Classically, a stochastic process is a random function $s(t), t \in \mathbb{R}$ whose statistical description is provided by the probability law of its point values $\{s(t_1), s(t_2), \ldots, s(t_n), \ldots\}$ for any finite sequence of time instants $\{t_n\}_{n=1}^N$. The implicit assumption there is that one has a mechanism for probing the value of the function s at any time $t \in \mathbb{R}$, which is only achievable approximately in the real physical world.

The leading idea in Gelfand and Vilenkin's theory of generalized stochastic processes is to replace the point measurements $\{s(t_n)\}$ by a series of scalar products $\{\langle s, \varphi_n\rangle\}$ with suitable "test" functions $\varphi_1, \ldots, \varphi_N \in \mathcal{S}$ [29]. The physical motivation that these authors give is that $X_n = \langle s, \varphi_n\rangle$ may represent the reading of a finite-resolution detector whose output is some "averaged" value $\int_{\mathbb{R}} s(t)\varphi_n(t)\mathrm{d}t$, which is a more plausible form of probing than ideal sampling. The additional hypothesis is that the linear measurement $X = \langle s, \varphi\rangle$ depends continuously on φ and that the quantities $X_n = \langle s, \varphi_n\rangle$ obtained for different test functions $\{\varphi_n\}$ are mutually compatible. Mathematically, this translates into defining a generalized stochastic process as a continuous linear random functional on some topological vector space such as $\mathcal{S}$.

Let s be such a generalized process. We first observe that the scalar product $X_1 = \langle s, \varphi_1\rangle$ with a given test function φ_1 is a conventional (scalar) random variable that is characterized by its probability density function (pdf) $p_{X_1}(x_1)$; the latter is in one-to-one correspondence (via the Fourier transform) with the characteristic function $\hat{p}_{X_1}(\omega_1) = \mathbb{E}\{e^{j\omega_1 X_1}\} = \int_{\mathbb{R}} e^{j\omega_1 x_1} p_{X_1}(x_1)\mathrm{d}x_1 = \mathbb{E}\{e^{j\langle s, \omega_1\varphi_1\rangle}\}$ where $\mathbb{E}\{\cdot\}$ is the expectation operator. The same applies for the 2nd-order pdf $p_{X_1,X_2}(x_1, x_2)$ associated with a pair of test functions φ_1 and φ_2 which is the inverse Fourier transform of the 2-D characteristic function $\hat{p}_{X_1,X_2}(\omega_1, \omega_2) = \mathbb{E}\{e^{j\langle s, \omega_1\varphi_1 + \omega_2\varphi_2\rangle}\}$, and so forth if one wants to specify higher-order dependencies.

The foundation for the theory of generalized stochastic processes is that one can deduce the complete statistical information about the process from the knowledge of its characteristic form

$$\widehat{\mathscr{P}}_s(\varphi) = \mathbb{E}\{e^{j\langle s,\varphi\rangle}\} \tag{4}$$

which is a continuous, positive-definite functional over $\mathcal{S}$ such that $\widehat{\mathscr{P}}_s(0) = 1$. Since the variable φ in $\widehat{\mathscr{P}}_s(\varphi)$ is completely generic, it provides the equivalent of an infinite-dimensional generalization of the characteristic function. Indeed, any finite dimensional version can be recovered by direct substitution of $\varphi = \omega_1\varphi_1 + \cdots + \omega_N\varphi_N$ in $\widehat{\mathscr{P}}_s(\varphi)$ where the φ_n are fixed and where $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_N)$ takes the role of the N-dimensional Fourier variable.

In fact, Gelfand's theory rests upon the principle that specifying an admissible functional $\widehat{\mathscr{P}}_s(\varphi)$ is equivalent to defining the underlying generalized stochastic process (Bochner-Minlos theorem). To explain this remarkable result, we start by recalling the fundamental notion of positive-definiteness for univariate functions [37].

Definition 1: A complex-valued function f of the real variable ω is said to be *positive-definite* iff.

$$\sum_{m=1}^{N}\sum_{n=1}^{N} f(\omega_m - \omega_n)\xi_m\overline{\xi}_n \geq 0$$

for every possible choice of $\omega_1, \ldots, \omega_N \in \mathbb{R}$, $\xi_1, \ldots, \xi_N \in \mathbb{C}$ and $N \in \mathbb{Z}^+$.

This is equivalent to the requirement that the $N \times N$ matrix $\mathbf{F}$ whose elements are given by $[\mathbf{F}]_{mn} = f(\omega_m - \omega_n)$ is positive semi-definite (that is, non-negative definite) for all N, no matter how the ω_n's are chosen.

Bochner's theorem states that a bounded, continuous function $\hat{p}$ is positive-definite if and only if it is the Fourier transform of a positive and finite Borel measure P:

$$\hat{p}(\omega) = \int_{\mathbb{R}} e^{j\omega x} P(\mathrm{d}x).$$

In particular, Bochner's theorem implies that a function $\hat{p}_X(\omega)$ is a valid characteristic function—that is, $\hat{p}_X(\omega) = \mathbb{E}\{e^{j\omega X}\} = \int_{\mathbb{R}} e^{j\omega x} P_X(\mathrm{d}x) = \int_{\mathbb{R}} e^{j\omega x} p_X(x)\mathrm{d}x$ where X is a random variable with probability measure P_X (or pdf p_X)—iff. $\hat{p}_X$ is continuous, positive-definite and such that $\hat{p}_X(0) = 1$.

The power of functional analysis is that these concepts carry over to functionals on some abstract nuclear space $\mathcal{X}$, the prime example being Schwartz's class $\mathcal{S}$ of smooth and rapidly-decreasing test functions [29].

Definition 2: A complex-valued functional $F(\varphi)$ defined over the function space $\mathcal{X}$ is said to be *positive-definite* iff.

$$\sum_{m=1}^{N}\sum_{n=1}^{N} F(\varphi_m - \varphi_n)\xi_m\overline{\xi}_n \geq 0$$

for every possible choice of $\varphi_1, \ldots, \varphi_N \in \mathcal{X}$, $\xi_1, \ldots, \xi_N \in \mathbb{C}$ and $N \in \mathbb{Z}^+$.

Definition 3: A functional $F : \mathcal{X} \to \mathbb{R}$ (or $\mathbb{C}$) is said to be continuous (with respect to the topology of the function space $\mathcal{X}$) if, for any convergent sequence (φ_i) in $\mathcal{X}$ with limit $\varphi \in \mathcal{X}$, the sequence $F(\varphi_i)$ converges to $F(\varphi)$; that is,

$$\lim_i F(\varphi_i) = F(\lim_i \varphi_i).$$

Theorem 1 (Minlos-Bochner): Given a functional $\widehat{\mathscr{P}}_s(\varphi)$ on a nuclear space $\mathcal{X}$ that is continuous, positive-definite and such that $\widehat{\mathscr{P}}_s(0) = 1$, there exists a unique probability measure $\mathscr{P}_s$ on the dual space $\mathcal{X}'$ such that

$$\widehat{\mathscr{P}}_s(\varphi) = \mathbb{E}\{e^{j\langle s,\varphi\rangle}\} = \int_{\mathcal{X}'} e^{j\langle s,\varphi\rangle}\mathscr{P}(\mathrm{d}s),$$

where $\langle s, \varphi\rangle$ is the dual pairing map. One further has the guarantee that all finite dimensional probabilities measures derived from $\widehat{\mathscr{P}}_s(\varphi)$ by setting $\varphi = \omega_1\varphi_1 + \cdots + \omega_N\varphi_N$ are mutually compatible.

The characteristic form therefore uniquely specifies the generalized stochastic process $s = s(\varphi)$ (via the infinite-dimensional probability measure $\mathscr{P}_s$) in essentially the same way as the characteristic function fully determines the probability measure of a scalar or multivariate random variable.

C. White Noise Processes (Innovations)

We define a white noise w as a generalized random process that is stationary and whose measurements for non-overlapping test functions are independent. A remarkable aspect of the theory of generalized stochastic processes is that it is possible to deduce the complete class of such noises based on functional considerations only [29]. To that end, Gelfand and Vilenkin consider the generic class of functionals of the form

$$\widehat{\mathscr{P}}_w(\varphi) = \exp\left(\int_{\mathbb{R}} f\big(\varphi(t)\big)\mathrm{d}t\right) \tag{5}$$

where f is a continuous function on the real line and φ is a test function from some suitable space. This functional specifies an independent noise process if $\widehat{\mathscr{P}}_w$ is continuous and positive-definite and $\widehat{\mathscr{P}}_w(\varphi_1 + \varphi_2) = \widehat{\mathscr{P}}_w(\varphi_1)\widehat{\mathscr{P}}_w(\varphi_2)$ whenever φ_1 and φ_2 have non-overlapping support. The latter property is equivalent to having $f(0) = 0$ in (5). Gelfand and Vilenkin then go on to prove that the complete class of functionals of the form (5) with the required mathematical properties (continuity, positive-definiteness and factorizability) is obtained by choosing f to be a *Lévy exponent,* as defined below.

Definition 4: A complex-valued continuous function $f(\omega)$ is a valid Lévy exponent if and only if $f(0) = 0$ and $g_\tau(\omega) = e^{\tau f(\omega)}$ is a positive-definite function of ω for all $\tau \in \mathbb{R}^+$.

In doing so, they actually establish a one-to-one correspondence between the characteristic form of an independent noise processes (5) and the family of infinite-divisible laws whose characteristic function takes the form $\hat{p}_X(\omega) = e^{f(\omega)} = \mathbb{E}\{e^{j\omega X}\}$ [38], [39]. While Definition 4 is hard to exploit directly, the good news is that there exists a complete constructive, characterization of Lévy exponents, which is a classical result in probability theory:

Theorem 2 (Lévy-Khintchine Formula): $f(\omega)$ is a valid Lévy exponent if and only if it can be written as

$$f(\omega) = jb_1'\omega - \frac{b_2\omega^2}{2} + \int_{\mathbb{R}\setminus\{0\}} [e^{ja\omega} - 1 - ja\omega \mathbb{1}_{\{|a|<1\}}(a)]\, V(\mathrm{d}a) \tag{6}$$

where $b_1' \in \mathbb{R}$ and $b_2 \in \mathbb{R}^+$ are some constants and V is a Lévy measure, that is, a (positive) Borel measure on $\mathbb{R}\setminus\{0\}$ such that

$$\int_{\mathbb{R}\setminus\{0\}} \min(1, a^2)\, V(\mathrm{d}a) < \infty. \tag{7}$$

The notation $\mathbb{1}_\Omega(a)$ refers to the indicator function that takes the value 1 if $a \in \Omega$ and zero otherwise. Theorem 2 is fundamental to the classical theories of infinite-divisible laws and Lévy processes [28], [31], [39]. To further our mathematical understanding of the Lévy-Khintchine formula (6), we note that $e^{ja\omega} - 1 - ja\omega\mathbb{1}_{\{|a|<1\}}(a) \sim -\frac{1}{2}a^2\omega^2$ as $a \to 0$. This ensures that the integral is convergent even when the Lévy measure V is singular at the origin to the extent allowed by the admissibility condition (7). If the Lévy measure is finite or symmetrical (i.e., $V(E) = V(-E)$ for any $E \subset \mathbb{R}$), it is then also possible to use the equivalent, simplified form of Lévy exponent

$$f(\omega) = jb_1\omega - \frac{b_2\omega^2}{2} + \int_{\mathbb{R}\setminus\{0\}} \left(e^{ja\omega} - 1\right) V(\mathrm{d}a) \tag{8}$$

with $b_1 = b_1' - \int_{0<|a|<1} a\, V(\mathrm{d}a)$. The bottomline is that a particular brand of independent noise process is thereby completely characterized by its Lévy exponent or, equivalently, its Lévy triplet (b_1, b_2, υ) where υ is the so-called *Lévy density* associated with V such that

$$V(E) = \int_E \upsilon(a)\mathrm{d}a$$

for any Borel set $E \subseteq \mathbb{R}$. With this latter convention, the three primary types of innovations encountered in the signal processing and statistics literature are specified as follows:

1) Gaussian: $b_1 = 0, b_2 = 1, \upsilon = 0$

$$f_{\text{Gauss}}(\omega) = -\frac{|\omega|^2}{2},$$
$$\widehat{\mathscr{P}}_w(\varphi) = e^{-\frac{1}{2}\|\varphi\|_{L_2}^2}. \tag{9}$$

2) Compound Poisson [18]: $b_1 = 0, b_2 = 0$, $\upsilon(a) = \lambda\ p_A(a)$ with $\int_{\mathbb{R}} p_A(a)\mathrm{d}a = \hat{p}_A(0) = 1$,

$$f_{\text{Poisson}}(\omega;\, \lambda,\, p_A) = \lambda \int_{\mathbb{R}} \left(e^{ja\omega} - 1\right) p_A(a)\mathrm{d}a,$$
$$\widehat{\mathscr{P}}_w(\varphi) = \exp\left(\lambda \int_{\mathbb{R}}\int_{\mathbb{R}} (e^{ja\varphi(t)} - 1)\ p_A(a)\mathrm{d}a\mathrm{d}t\right). \tag{10}$$

3) Symmetric alpha-stable (SαS) [40]: $b_1 = 0, b_2 = 0, \upsilon(a) = \frac{C_\alpha}{|a|^{\alpha+1}}$ with $0 < \alpha < 2$ and $C_\alpha = \frac{\sin(\frac{\pi\alpha}{2})}{\pi}$ a suitable normalization constant,

$$f_\alpha(\omega) = \frac{-|\omega|^\alpha}{\alpha!},$$
$$\widehat{\mathscr{P}}_w(\varphi) = e^{-\frac{1}{\alpha!}\|\varphi\|_{L_\alpha}^\alpha}. \tag{11}$$

The latter follows from the fact that $\frac{-|\omega|^\alpha}{\alpha!}$ is the generalized Fourier transform of $\frac{C_\alpha}{|t|^{\alpha+1}}$ with the convention that $\alpha! = \Gamma(\alpha + 1)$ where Γ is Euler's Gamma function [41].

While none of these innovations has a classical interpretation as a random function of t, we can at least provide an explicit description of the Poisson noise as an infinite random sequence of Dirac impulses (cf. [18, Theorem 1])

$$w_\lambda(t) = \sum_k A_k \delta(t - t_k)$$

where the t_k are random locations that are uniformly distributed over $\mathbb{R}$ with density λ, and where the weights A_k are i.i.d. random variables with pdf $p_A(a)$. Remarkably, this is the only innovation process in the family that has a finite rate of innovation [17]; however, it is, by far, not the only one that is sparse as explained next.

D. Gaussian Versus Sparse Categorization

To get a better understanding of the underlying class of white noises w, we propose to probe them through some localized analysis window φ, which will yield a conventional i.i.d. random variable $X = \langle w, \varphi\rangle$ with some pdf $p_\varphi(x)$. The most convenient choice is to pick the rectangular analysis window $\varphi(t) = \text{rect}(t) = \mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]}(t)$ when $\langle w, \text{rect}\rangle$ is well-defined. By using the fact that $e^{ja\omega\text{rect}(t)} - 1 = e^{ja\omega} - 1$ for $t \in [-\frac{1}{2}, \frac{1}{2}]$, and zero otherwise, we find that the characteristic function of X is simply given by

$$\hat{p}_{\text{rect}}(\omega) = \widehat{\mathscr{P}}_w\left(\omega \cdot \text{rect}(t)\right) = \exp\left(f(\omega)\right),$$

which corresponds to the generic (Lévy-Khinchine) form associated with an infinitely-divisible distribution [31], [39], [42]. The above result makes the mapping between generalized white noise processes and classical infinite-divisible (id) laws[4] explicit: The "canonical" id pdf of w, $p_{\text{id}}(x) = p_{\text{rect}}(x)$, is obtained by observing the noise through a rectangular window. Conversely, given the Lévy exponent of an id distribution, $f(\omega) = \log\left(\mathcal{F}\{p_{\text{id}}\}(\omega)\right)$, we can specify a corresponding innovation process w via the characteristic form $\widehat{\mathscr{P}}_w(\varphi)$ by merely substituting the frequency variable ω by the generic test function $\varphi(t)$, adding an integration over $\mathbb{R}$ and taking the exponential as in (5).

We note, in passing, that sparsity in signal processing may refer to two distinct notions. The first is that of a finite rate of innovation; i.e., a finite (but perhaps random) number of innovations per unit of time and/or space, which results in a mass at zero in the histogram of observations. The second possibility is to have a large, even infinite, number of innovations, but with the property that a few large innovations dominate the overall behavior. In this case the histogram of observations is distinguished by its 'heavy tails'. (A combination of the two is also possible, for instance in a compound Poisson process with a heavy-tailed amplitude distribution. For such a process one may observe a change of behavior in passing from one dominant type of sparsity to the other).

[4] A random variable X with pdf $p_X(x)$ is said to be infinitely divisible (id) if for any $n \in \mathbb{Z}^+$ there exist i.i.d. random variables $X_1, \ldots, X_n$ with pdf say $p_n(x)$ such that $X = X_1 + \cdots + X_n$ in law.

Our framework permits us to consider both types of sparsity, in the former case with compound Poisson models and in the latter with heavy-tailed infinitely-divisible innovations.

To make our point, we consider two distinct scenarios.

1) Finite Variance Case: We first assume that the second moment $m_2 = \int_{\mathbb{R}\setminus\{0\}} a^2\, V(\mathrm{d}a)$ of the Lévy measure V in (6) is finite. This allows us to rewrite the classical Lévy-Khinchine representation as

$$f(\omega) = jc_1\omega - \frac{b_2\omega^2}{2} + \int_{\mathbb{R}\setminus\{0\}} [e^{ja\omega} - 1 - ja\omega]\, V(\mathrm{d}a)$$

with $c_1 = b_1'' + \int_{|a|>1} aV(\mathrm{d}a)$ and where the Poisson part of the functional is now fully compensated. Indeed, we are guaranteed that the above integral is convergent because $|e^{ja\omega} - 1 - j\omega a| \lesssim |a\omega|^2$ as $a \to 0$ and $|e^{ja\omega} - 1 - j\omega a| \sim |a\omega|$ as $a \to \pm\infty$. An interesting non-Poisson example of infinitely-divisible probability laws that falls into this category (with non-finite V) is the Laplace distribution with Lévy triplet $(0, 0, v(a) = \frac{e^{-|a|}}{|a|})$ and $p_{\mathrm{id}}(x) = \frac{1}{2}e^{-|x|}$. This model is particularly relevant for sparse signal processing because it provides a tight connection between Lévy processes and total variation regularization [18, Section VI].

Now, if the Lévy measure is finite $\int_{\mathbb{R}} V(\mathrm{d}a) = \lambda < \infty$, the admissibility condition yields $\int_{\mathbb{R}\setminus\{0\}} a\, V(\mathrm{d}a) < \infty$, which allows us to pull the bias correction out of the integral. The representation then simplifies to (8). This implies that we can decompose X into the sum of two independent Gaussian and compound Poisson random variables. The variances of the Gaussian and Poisson components are $\sigma^2 = b_2$ and $\int_{\mathbb{R}} a^2 V(\mathrm{d}a)$, respectively. The Poisson component is sparse because its pdf exhibits a mass distribution $e^{-\lambda}\delta(x)$ at the origin, meaning that the chances for a continuous amplitude distribution of getting zero are overwhelmingly higher than any other value, especially for smaller values of $\lambda > 0$. It is therefore justifiable to use $0 \leq e^{-\lambda} < 1$ as our Poisson sparsity index.

2) Infinite Variance Case: We now turn our attention to the case where the second moment of the Lévy measure is unbounded, which we like to label as the "super-sparse" one. To substantiate this claim, we invoke the Ramachandran-Wolfe theorem which states that the pth moment $\mathbb{E}\{|X|^p\}$ with $p \in \mathbb{R}^+$ of an infinitely divisible distribution is finite iff. $\int_{|a|>1} |a|^p\ V(\mathrm{d}a) < \infty$ [43], [44]. For $p \geq 2$, the latter is equivalent to $\int_{\mathbb{R}\setminus\{0\}} |a|^p\ V(\mathrm{d}a) < \infty$ because of the admissibility condition (7). It follows that the cases that are not covered by the previous scenario (including the Gaussian + Poisson model) necessarily give rise to distributions whose moments of order p are unbounded for $p \geq 2$. The prototypical representatives of such heavy tail distributions are the alpha-stable ones or, by extension, the broad family of infinite divisible probability laws that are in their domain of attraction. Note that these distributions all fulfill the stringent conditions for ℓ_p compressibility [45], [46].

IV. Innovation Approach to Continuous-Time Stochastic Processes

Specifying a stochastic process through an innovation model (or an equivalent stochastic differential equation) is attractive conceptually, but it presupposes that we can provide an inverse operator (in the form of an integral transform) that transforms the innovation back into the initial stochastic process. This is the reason why, after laying out general conditions for existence, we shall spend the greater part of our effort investigating suitable inverse operators.

A. Stochastic Differential Equations

Our aim is to define the generalized process with whitening operator $\mathrm{L} : \mathcal{S}' \to \mathcal{S}'$ and Lévy exponent f as the solution of the stochastic linear differential equation

$$\mathrm{L}s = w, \tag{12}$$

where w is an innovation process, as described in Section III-C. This definition is obviously only usable if we can construct an inverse operator $\mathrm{T} = \mathrm{L}^{-1}$ that solves this equation. For the cases where the inverse is not unique, we will need to select one preferential operator, which is equivalent to imposing specific boundary conditions. We are then able to formally express the stochastic process as a transformed version of a white noise

$$s = \mathrm{L}^{-1}w. \tag{13}$$

The requirement for such a solution to be consistent with (12) is that the operator satisfies the right-inverse property $\mathrm{L}\mathrm{L}^{-1} = \mathrm{Id}$ over the underlying class of tempered distributions. By using the adjoint relation $\langle s, \varphi\rangle = \langle \mathrm{L}^{-1}w, \varphi\rangle = \langle w, \mathrm{L}^{-1*}\varphi\rangle$, we can then transfer the action of the operator onto the test function inside the characteristic form and obtain a complete statistical characterization of the so-defined generalized stochastic process

$$\widehat{\mathscr{P}}_s(\varphi) = \widehat{\mathscr{P}}_{\mathrm{L}^{-1}w}(\varphi) = \widehat{\mathscr{P}}_w(\mathrm{L}^{-1*}\varphi), \tag{14}$$

where $\widehat{\mathscr{P}}_w$ is given by (5) (or one of the specific forms in the list at the end of Section III-C) and where we are implicitly requiring that the adjoint L^{-1*} is mathematically well-defined (continuous) over $\mathcal{S}$, and that its composition with $\widehat{\mathscr{P}}_w$ is well-defined for all $\varphi \in \mathcal{S}$.

In order to realize the above idea mathematically, it is usually easier to proceed backwards: one specifies an operator T that satisfies the left-inverse property: $\forall \varphi \in \mathcal{S},\ \mathrm{T}\mathrm{L}^*\varphi = \varphi$, and that is continuous (i.e., bounded in the proper norm(s)) over the chosen class of test functions. One then characterizes the adjoint of T, which is the operator $\mathrm{T}^* : \mathcal{S}' \to \mathcal{S}'$ (or an appropriate subset thereof) such that, for a given $\phi \in \mathcal{S}'$,

$$\forall \varphi \in \mathcal{S}, \quad \langle \phi, \varphi\rangle = \langle \mathrm{L}\mathrm{T}^*\phi, \varphi\rangle = \langle \phi, \underbrace{\mathrm{T}\mathrm{L}^*}_{\mathrm{Id}}\varphi\rangle.$$

Finally, we set $\mathrm{L}^{-1} = \mathrm{T}^*$, which yields the proper distributional definition of the right inverse of L in (13).

B. General Conditions for Existence

To validate the proposed innovation model, we need to ensure that the solution $s = \mathrm{L}^{-1}w$ is a bona fide generalized stochastic process.

In order to simplify the analysis, we shall restrict our attention to an appropriate subclass of Lévy exponents.

Definition 5: A Lévy exponent f with derivative f' is *p-admissible* with $1 \leq p \leq 2$ if there exists a positive constant C such that $|f(\omega)| + |\omega| \cdot |f'(\omega)| \leq C|\omega|^p$ for all $\omega \in \mathbb{R}$.

Note that this p-admissibility condition is not very constraining and that it is satisfied by the great majority of members of the Lévy-Kintchine family (see Section III-C). For instance in the compound Poisson case, we can show that $|\omega| \cdot |f'(\omega)| \leq \lambda|\omega|\ \mathbb{E}\{|A|\}$ and $f(\omega) \leq \lambda|\omega|\ \mathbb{E}\{|A|\}$ by using the fact $|e^{jx} - 1| \leq |x|$; this implies that the bound in Definition 5 with $p = 1$ is always satisfied provided that the first (absolute) moment of the amplitude pdf $p_A(a)$ in (10) is finite. Similarly, all symmetric Lévy exponents with $-f''(0) < \infty$ (finite variance case) are p-admissible with $p = 2$, the prototypical example being the Gaussian. The only cases we are aware of that do not fulfill the condition are the alpha-stable noises with $0 < \alpha < 1$, which are notorious for their exotic behavior.

The first advantage of imposing p-admissibility is that it allows us to extend the set of acceptable analysis functions from $\mathcal{S}$ to L_p which is crucial if we intend to do conventional signal processing.

Theorem 3: If the Lévy exponent f is p-admissible, then the characteristic form $\widehat{\mathscr{P}}_w(\varphi) = \exp\left(\int_{\mathbb{R}} f\big(\varphi(t)\big)\mathrm{d}t\right)$ is a continuous, positive-definite functional over L_p.

Proof: Since the exponential function is continuous, it is sufficient to consider the functional

$$F(\varphi) = \log \widehat{\mathscr{P}}_w(\varphi) = \int_{\mathbb{R}} f(\varphi(t))\mathrm{d}t,$$

which is such that $F(0) = 0$. To show that $F(\varphi)$ $\big($and hence $\widehat{\mathscr{P}}_w(\varphi)\big)$ is well-defined over L_p, we note that

$$|F(\varphi)| \leq \int_{\mathbb{R}} |f(\varphi(t))|\,\mathrm{d}t \leq C\|\varphi\|_p^p,$$

which follows from the p-admissibility condition. The positive definiteness of $\widehat{\mathscr{P}}_w(\varphi)$ over $\mathcal{S}$ is a direct consequence of f being a Lévy exponent and is therefore also transferable to L_p. For the interested reader, this can be shown quite easily by proving that $F(\varphi)$ is conditionally positive-definite of order one (see [20]).

The only remaining work is to establish the L_p-continuity of $F(\varphi)$. To that end, we observe that

$$\begin{aligned}
|f(u) - f(v)| &= \left|\int_v^u f'(t)\mathrm{d}t\right| \\
&\leq C\left|\int_v^u t^{p-1}\mathrm{d}t\right| \\
&\qquad \text{(by the assumption on } f\text{)} \\
&\leq C\max(|u|^{p-1}, |v|^{p-1})|u - v| \\
&\leq C(|v|^{p-1} + |u - v|^{p-1})|u - v|. \\
&\qquad \text{(by the triangle inequality)}
\end{aligned}$$

Next, we pick a convergent sequence in L_p, $\{\varphi_n\}_{n=1}^{\infty}$, whose limit is denoted by φ. The convergence in L_p is expressed as

$$\lim_{n\to\infty} \|\varphi_n - \varphi\|_p = 0. \tag{15}$$

We then have

$$\begin{aligned}
&\left|\int_{\mathbb{R}} f(\varphi_n(t))\mathrm{d}t - \int_{\mathbb{R}} f(\varphi(t))\mathrm{d}t\right| \\
&\leq C\int_{\mathbb{R}} |\varphi(t)|^{p-1}|\varphi_n(t) - \varphi(t)| + |\varphi_n(t) - \varphi(t)|^p\mathrm{d}t \\
&\leq C\Big(\|\varphi\|_p^{p-1}\|\varphi_n - \varphi\|_p + \|\varphi_n - \varphi\|_p^p\Big) \\
&\qquad \text{(by Hölder's inequality)} \\
&\to 0 \text{ as } n \to \infty, \qquad \text{(by (15))}
\end{aligned}$$

which proves the continuity of the functional $\widehat{\mathscr{P}}_w$ on L_p. ■

Thanks to this result, we can then rely on the Minlos-Bochner theorem (Theorem 1) to state basic conditions on $\mathrm{T} = \mathrm{L}^{-1*}$ that ensure that $s = \mathrm{T}^* w$ is a well-defined generalized process over $\mathcal{S}'$.

Theorem 4 (Existence of Generalized Process): Let f be a valid Lévy exponent and T be an operator acting on $\varphi \in \mathcal{S}$ such that any one of the conditions below is met:

1) T is a continuous linear map from $\mathcal{S}$ into itself,
2) T is a continuous linear map from $\mathcal{S}$ into L_p and the Lévy exponent f is *p-admissible*.

Then, $\widehat{\mathscr{P}}_s(\varphi) = \exp\left(\int_{\mathbb{R}} f\big(\mathrm{T}\varphi(t)\big)\mathrm{d}t\right)$ is a continuous, positive-definite functional on $\mathcal{S}$ such that $\widehat{\mathscr{P}}_s(0) = 1$.

Proof: We already know that $\widehat{\mathscr{P}}_w$ is a continuous functional on $\mathcal{S}$ (resp., on L_p when f is p-admissible) by construction. This, together with the assumption that T is a continuous operator on $\mathcal{S}$ (resp., from $\mathcal{S}$ to L_p), implies that the composed functional $\widehat{\mathscr{P}}_s(\varphi) = \widehat{\mathscr{P}}_w(\mathrm{T}\varphi)$ is continuous on $\mathcal{S}$.

Given the functions $\varphi_1, \ldots, \varphi_N$ in $\mathcal{S}$ and some complex coefficients $\xi_1, \ldots, \xi_N$,

$$\begin{aligned}
&\sum_{1\leq m,n\leq N} \widehat{\mathscr{P}}_s(\varphi_m - \varphi_n)\xi_m\overline{\xi_n} \\
&\quad = \sum_{1\leq m,n\leq N} \widehat{\mathscr{P}}_w\big(\mathrm{T}(\varphi_m - \varphi_n)\big)\xi_m\overline{\xi_n} \\
&\quad = \sum_{1\leq m,n\leq N} \widehat{\mathscr{P}}_w(\mathrm{T}\varphi_m - \mathrm{T}\varphi_n)\xi_m\overline{\xi_n} \\
&\qquad \text{(by the linearity of the operator T)} \\
&\quad \geq 0. \qquad \text{(by the positivity of } \widehat{\mathscr{P}}_w \text{ over } \mathcal{S} \text{ or } L_p\text{)}
\end{aligned}$$

This proves the positive definiteness of the functional $\widehat{\mathscr{P}}_s$ on $\mathcal{S}$.

Lastly, $\widehat{\mathscr{P}}_s(0) = \widehat{\mathscr{P}}_w(\mathrm{T}0) = \widehat{\mathscr{P}}_w(0) = 1$. ■

The final fundamental issue relates to the interpretation of $s = \mathrm{L}^{-1}w$ as an ordinary stochastic process; that is, a random function $s(t)$ of the time variable t. This presupposes that the shaping operator L^{-1} performs a minimal amount of smoothing since the driving term of the model, w, is too rough to admit a pointwise representation.

Theorem 5 (Interpretation as Ordinary Stochastic Process): Let s be the generalized stochastic process whose characteristic function is given by (14) where f is a p-admissible Lévy exponent and L^{-1*} is a continuous operator from $\mathcal{S}$ to L_p (or a subset thereof). We also define the (generalized) impulse response

$$h(t, \tau) = \mathrm{L}^{-1}\{\delta(\cdot - \tau)\}(t), \tag{16}$$

with a slight abuse of notation since h is not necessarily an ordinary function. Then, $s = \mathrm{L}^{-1}w$ admits the pointwise representation for $t \in \mathbb{R}$

$$s(t) = \langle w, h(t, \cdot)\rangle \tag{17}$$

provided that $h(t, \cdot) \in L_p$ (with t fixed).

The form of $h(t, \tau)$ in (16) is the "time-domain" transcription of Schwartz's kernel theorem which gives the integral representation of a linear operator in terms of a (generalized) kernel $h \in \mathcal{S}' \times \mathcal{S}'$ (the infinite-dimensional generalization of a matrix multiplication). The more standard definition used in the theory of generalized functions is $\langle h(\cdot, \cdot), \varphi_1 \otimes \varphi_2\rangle = \langle \mathrm{L}^{-1*}\{\varphi_1\}, \varphi_2\rangle$, where $\varphi_1 \otimes \varphi_2(t, \tau) = \varphi_1(t)\varphi_2(\tau)$ for all $\varphi_1, \varphi_2 \in \mathcal{S}$.

Proof: The existence of the generalized stochastic process $s = \mathrm{L}^{-1}w$ is ensured by Theorem 4. We then consider the observation of the innovation $X_0 = \langle w, \varphi_0\rangle$ where $\varphi_0 = h(t_0, \cdot)$ with $\varphi_0 \in L_p$. Since $\widehat{\mathscr{P}}_w$ admits a continuous extension over L_p (by Theorem 3), we can specify the characteristic function of X_0 as

$$\hat{p}_{X_0}(\omega) = \mathbb{E}\{e^{j\omega X_0}\} = \widehat{\mathscr{P}}_w(\omega\varphi_0)$$

with φ_0 fixed. Thanks to the functional properties of $\widehat{\mathscr{P}}_w$, $\hat{p}_{X_0}(\omega)$ is a continuous, positive-definite function of ω such that $\hat{p}_{X_0}(0) = 1$ so that we can invoke Bochner's theorem to establish that X_0 is a well-defined conventional random variable with pdf p_{X_0} (the inverse Fourier transform of $\hat{p}_{X_0}$). ■

C. Inverse Operators

Before presenting our general method of solution, we need to identify a suitable set of elementary inverse operators that satisfy the continuity requirement in Theorem 4.

Our approach relies on the factorization of a differential operator into simple first-order components of the form $(\mathrm{D} - \alpha_n\mathrm{Id})$ with $\alpha_n \in \mathbb{C}$, which can then be treated separately. Three possible cases need to be considered.

1) Causal-Stable: $\mathrm{Re}(\alpha_n) < 0$. This is the classical textbook hypothesis which leads to a causal-stable convolution system. It is well known from the theory of distributions and linear systems (e.g., [47, Section 6.3], [48]) that the causal Green function of $(\mathrm{D} - \alpha_n\mathrm{Id})$ is the causal exponential function $\rho_{\alpha_n}(t)$ already encountered in the introductory example in Section II. Clearly, $\rho_{\alpha_n}(t)$ is absolutely integrable (and rapidly-decaying) iff. $\mathrm{Re}(\alpha_n) < 0$. It follows that $(\mathrm{D} - \alpha_n\mathrm{Id})^{-1} f = \rho_{\alpha_n} * f$ with $\rho_{\alpha_n} \in \mathcal{R} \subset L_1$. In particular, this implies that $\mathrm{T} = (\mathrm{D} - \alpha_n\mathrm{Id})^{-1}$ specifies a continuous LSI operator on $\mathcal{S}$. The same holds for $\mathrm{T}^* = (\mathrm{D} - \alpha_n\mathrm{Id})^{-1*}$, which is defined as $\mathrm{T}^* f = \rho^{\vee}_{\alpha_n} * f$.

2) Anti-Causal Stable: $\mathrm{Re}(\alpha_n) > 0$. This case is usually excluded because the standard Green function $\rho_{\alpha_n}(t) = \mathbb{1}_+(t)e^{\alpha_n t}$ grows exponentially, meaning that the system does not have a stable causal solution. Yet, it is possible to consider an alternative anti-causal Green function $\rho'_{\alpha_n}(t) = -\rho^{\vee}_{-\alpha_n}(t) = \rho_{\alpha_n}(t) - e^{\alpha_n t}$, which is unique in the sense that it is the only Green function[5] of $(\mathrm{D} - \alpha_n\mathrm{Id})$ that is Lebesgue-integrable and, by the same token, the proper inverse Fourier transform of $\frac{1}{j\omega - \alpha_n}$ for $\mathrm{Re}(\alpha_n) > 0$. In this way, we are able to specify an anti-causal inverse filter $(\mathrm{D} - \alpha_n\mathrm{Id})^{-1} f = \rho'_{\alpha_n} * f$ with $\rho'_{\alpha_n} \in \mathcal{R}$ that is L_p-stable and $\mathcal{S}$-continuous. In the sequel, we will drop the $'$ superscript with the convention that $\rho_\alpha(t)$ systematically refers to the unique Green function of $(\mathrm{D} - \alpha\mathrm{Id})$ that is rapidly-decaying when $\mathrm{Re}(\alpha) \neq 0$. For now on, we shall therefore use the definition

$$\rho_\alpha(t) = \begin{cases} \mathbb{1}_+(t)e^{\alpha t} & \text{if } \mathrm{Re}(\alpha) \leq 0 \\ -\mathbb{1}_+(-t)e^{\alpha t} & \text{otherwise.} \end{cases} \tag{18}$$

which also covers the next scenario.

3) Marginally Stable: $\mathrm{Re}(\alpha_n) = 0$ *or, equivalently,* $\alpha_n = j\omega_0$ *with* $\omega_0 \in \mathbb{R}$. This third case, which is incompatible with the conventional formulation of stationary processes, is most interesting theoretically because it opens the door to important extensions such as Lévy processes, as we shall see in Section V. Here, we will show that marginally-stable systems can be handled within our generalized framework as well, thanks to the introduction of appropriate inverse operators.

The first natural candidate for $(\mathrm{D} - j\omega_0\mathrm{Id})^{-1}$ is the inverse filter whose frequency response is

$$\hat{\rho}_{j\omega_0}(\omega) = \frac{1}{j(\omega - \omega_0)} + \pi\,\delta(\omega - \omega_0).$$

It is a convolution operator whose time-domain definition is

$$\begin{aligned} \mathrm{I}_{\omega_0}\varphi(t) &= (\rho_{j\omega_0} * \varphi)(t) \\ &= e^{j\omega_0 t}\int_{-\infty}^{t} e^{-j\omega_0\tau}\varphi(\tau)\mathrm{d}\tau. \end{aligned} \tag{19}$$

Its impulse response $\rho_{j\omega_0}(t)$ is causal and compatible with Definition (18), but not (rapidly) decaying. The adjoint of I_{ω_0} is given by

$$\begin{aligned} \mathrm{I}^*_{\omega_0}\varphi(t) &= (\rho^{\vee}_{j\omega_0} * \varphi)(t) \\ &= e^{-j\omega_0 t}\int_{t}^{+\infty} e^{j\omega_0\tau}\varphi(\tau)\mathrm{d}\tau. \end{aligned} \tag{20}$$

While $\mathrm{I}_{\omega_0}\varphi(t)$ and $\mathrm{I}^*_{\omega_0}\varphi(t)$ are both well-defined when $\varphi \in L_1$, the problem is that these inverse filters are not BIBO stable since their impulse responses, $\rho_{j\omega_0}(t)$ and $\rho^{\vee}_{j\omega_0}(t)$, are not in L_1. In particular, one can easily see that $\mathrm{I}_{\omega_0}\varphi$ (resp., $\mathrm{I}^*_{\omega_0}\varphi$) with $\varphi \in \mathcal{S}$ is generally not in L_p with $1 \leq p < +\infty$, unless $\hat{\varphi}(\omega_0) = 0$ (resp., $\hat{\varphi}(-\omega_0) = 0$). The conclusion is that $\mathrm{I}^*_{\omega_0}$ fails to be a bounded operator over the class of test functions $\mathcal{S}$.

This leads us to introduce some "corrected" version of the adjoint inverse operator $\mathrm{I}^*_{\omega_0}$,

$$\begin{aligned} \mathrm{I}^*_{\omega_0,t_0}\varphi(t) &= \mathrm{I}^*_{\omega_0}\left\{\varphi - \hat{\varphi}(-\omega_0)e^{-j\omega_0 t_0}\delta(\cdot - t_0)\right\}(t) \\ &= \mathrm{I}^*_{\omega_0}\varphi(t) - \hat{\varphi}(-\omega_0)e^{-j\omega_0 t_0}\rho^{\vee}_{j\omega_0}(t - t_0), \end{aligned} \tag{21}$$

where $t_0 \in \mathbb{R}$ is a fixed location parameter and where $\hat{\varphi}(-\omega_0) = \int_{\mathbb{R}} e^{j\omega_0 t}\varphi(t)\mathrm{d}t$ is the complex sinusoidal moment

[5]: ρ is a Green functions of $(\mathrm{D} - \alpha_n\mathrm{Id})$ iff. $(\mathrm{D} - \alpha_n\mathrm{Id})\rho = \delta$; the complete set of solutions is given $\rho(t) = \rho_{\alpha_n}(t) + Ce^{\alpha_n t}$ which is the sum of the causal Green function $\rho_{\alpha_n}(t)$ plus an arbitrary exponential component that is in the null space of the operator.

associated with the frequency ω_0. The idea is to correct for the lack of decay of $\mathrm{I}^*_{\omega_0}\varphi(t)$ as $t \to -\infty$ by subtracting a properly weighted version of the impulse response of the operator. An equivalent Fourier-based formulation is provided by the formula at the bottom of Table I; the main difference with the corresponding expression for $\mathrm{I}_{\omega_0}\varphi$ is the presence of a regularization term in the numerator that prevents the integrant from diverging at $\omega = \omega_0$. The next step is to identify the adjoint of $\mathrm{I}^*_{\omega_0,t_0}$, which is achieved via the following inner-product manipulation

$$\begin{aligned}\langle \varphi, \mathrm{I}^*_{\omega_0,t_0}\phi\rangle &= \langle \varphi, \mathrm{I}^*_{\omega_0}\phi\rangle - \hat{\phi}(-\omega_0)e^{-j\omega_0 t_0}\langle \varphi, \rho^{\vee}_{j\omega_0}(\cdot - t_0)\rangle \\ &= \langle \mathrm{I}_{\omega_0}\varphi, \phi\rangle - \langle e^{j\omega_0\cdot}, \phi\rangle\, e^{-j\omega_0 t_0}\ \mathrm{I}_{\omega_0}\varphi(t_0) \qquad \text{(using(19))} \\ &= \langle \mathrm{I}_{\omega_0}\varphi, \phi\rangle - \langle e^{j\omega_0(\cdot - t_0)}\mathrm{I}_{\omega_0}\varphi(t_0), \phi\rangle.\end{aligned}$$

Since the above is equal to $\langle \mathrm{I}_{\omega_0,t_0}\varphi, \phi\rangle$ by definition, we obtain that

$$\mathrm{I}_{\omega_0,t_0}\varphi(t) = \mathrm{I}_{\omega_0}\varphi(t) - e^{j\omega_0(t-t_0)}\ \mathrm{I}_{\omega_0}\varphi(t_0). \tag{22}$$

Interestingly, this operator imposes the boundary condition $\mathrm{I}_{\omega_0,t_0}\varphi(t_0) = 0$ via the subtraction of a sinusoidal component that is in the null space of the operator $(\mathrm{D} - j\omega_0\mathrm{Id})$, which gives a direct interpretation of the location parameter t_0. Observe that expressions (21) and (22) define linear operators, albeit not shift-invariant ones, in contrast with the classical inverse operators I_{ω_0} and $\mathrm{I}^*_{\omega_0}$.

For analysis purposes, it is convenient to relate the proposed inverse operators to the anti-derivatives corresponding to the case $\omega_0 = 0$. To that end, we introduce the modulation operator

$$\mathrm{M}_{\omega_0}\varphi(t) = e^{j\omega_0 t}\varphi(t)$$

which is a unitary map on L_2 with the property that $\mathrm{M}^{-1}_{\omega_0} = \mathrm{M}_{-\omega_0}$.

Proposition 1: The inverse operators defined by (19), (20), (22), and (21) satisfy the modulation relations

$$\begin{aligned}\mathrm{I}_{\omega_0}\varphi(t) &= \mathrm{M}_{\omega_0}\,\mathrm{I}_0\,\mathrm{M}^{-1}_{\omega_0}\varphi(t), \\ \mathrm{I}^*_{\omega_0}\varphi(t) &= \mathrm{M}^{-1}_{\omega_0}\,\mathrm{I}^*_0\,\mathrm{M}_{\omega_0}\varphi(t), \\ \mathrm{I}_{\omega_0,t_0}\varphi(t) &= \mathrm{M}_{\omega_0}\,\mathrm{I}_{0,t_0}\,\mathrm{M}^{-1}_{\omega_0}\varphi(t), \\ \mathrm{I}^*_{\omega_0,t_0}\varphi(t) &= \mathrm{M}^{-1}_{\omega_0}\,\mathrm{I}^*_{0,t_0}\,\mathrm{M}_{\omega_0}\varphi(t).\end{aligned}$$

Proof: These follow from the modulation property of the Fourier transform (i.e, $\mathcal{F}\{\mathrm{M}_{\omega_0}\varphi\}(\omega) = \mathcal{F}\{\varphi\}(\omega - \omega_0)$) and the observations that $\mathrm{I}_{\omega_0}\delta(t) = \rho_{j\omega_0}(t) = \mathrm{M}_{\omega_0}\rho_0(t)$ and $\mathrm{I}^*_{\omega_0}\delta(t) = \rho^{\vee}_{j\omega_0}(t) = \mathrm{M}_{-\omega_0}\rho^{\vee}_0(t)$ with $\rho_0(t) = \mathbb{1}_+(t)$ (the unit step function). ∎

The important functional property of $\mathrm{I}^*_{\omega_0,t_0}$ is that it essentially preserves decay and integrability, while $\mathrm{I}_{\omega_0,t_0}$ fully retains signal differentiability. Unfortunately, it is not possible to have the two simultaneously unless $\mathrm{I}_{\omega_0}\varphi(t_0)$ and $\hat{\varphi}(-\omega_0)$ are both zero.

Proposition 2: If $f \in L_{\infty,\alpha}$ with $\alpha > 1$, then there exists a constant C_{t_0} such that

$$|\mathrm{I}^*_{\omega_0,t_0}f(t)| \le C_{t_0}\frac{\|f\|_{\infty,\alpha}}{1+|t|^{\alpha-1}},$$

which implies that $\mathrm{I}^*_{\omega_0,t_0}f \in L_{\infty,\alpha-1}$.

Proof: Since modulation does not affect the decay properties of a function, we can invoke Proposition 1 and concentrate on the investigation of the anti-derivative operator I^*_{0,t_0}. Without loss of generality, we can also pick $t_0 = 0$ and transfer the bound to any other finite value of t_0 by adjusting the value of the constant C_{t_0}. Specifically, for $t < 0$, we write this inverse operator as

$$\begin{aligned}\mathrm{I}^*_{0,0}f(t) &= \mathrm{I}^*_0 f(t) - \hat{f}(0) \\ &= \int_t^{+\infty} f(\tau)\mathrm{d}\tau - \int_{-\infty}^{\infty} f(\tau)\mathrm{d}\tau \\ &= -\int_{-\infty}^{t} f(\tau)\mathrm{d}\tau.\end{aligned}$$

This implies that

$$\begin{aligned}|\mathrm{I}^*_{0,0}f(t)| = \left|\int_{-\infty}^{t} f(\tau)\mathrm{d}\tau\right| &\le \|f\|_{\infty,\alpha}\int_{-\infty}^{t}\frac{1}{1+|\tau|^{\alpha}}\mathrm{d}\tau \\ &\le \left(\frac{2\alpha}{\alpha-1}\right)\frac{\|f\|_{\infty,\alpha}}{1+|t|^{\alpha-1}}\end{aligned}$$

for all $t < 0$. For $t > 0$, $\mathrm{I}^*_{0,0}f(t) = \int_t^{\infty} f(\tau)\mathrm{d}\tau$ so that the above upper bounds remain valid. ∎

The interpretation of the above result is that the inverse operator $\mathrm{I}^*_{\omega_0,t_0}$ reduces inverse polynomial decay by one order. Proposition 2 actually implies that the operator will preserve the rapid decay of the Schwartz functions which are included in $L_{\infty,\alpha}$ for any $\alpha \in \mathbb{R}^+$. It also guarantees that $\mathrm{I}^*_{\omega_0,t_0}\varphi$ belongs to L_p for any Schwartz function φ. However, $\mathrm{I}^*_{\omega_0,t_0}$ will spoil the global smoothness properties of φ because it introduces a discontinuity at t_0, unless $\hat{\varphi}(-\omega_0)$ is zero in which case the output remains in the Schwartz class. This allows us to state the following theorem which summarizes the higher-level part of those results for further reference.

Theorem 6: The operator $\mathrm{I}^*_{\omega_0,t_0}$ defined by (22) is a continuous linear map from $\mathcal{R}$ into $\mathcal{R}$ (the space of bounded functions with rapid decay). Its adjoint $\mathrm{I}_{\omega_0,t_0}$ is given by (21) and has the property that $\mathrm{I}_{\omega_0,t_0}\varphi(t_0) = 0$. Together, these operators satisfy the complementary left- and right-inverse relations

$$\begin{cases}\mathrm{I}^*_{\omega_0,t_0}(\mathrm{D} - j\omega_0\mathrm{Id})^*\varphi = \varphi \\ (\mathrm{D} - j\omega_0\mathrm{Id})\mathrm{I}_{\omega_0,t_0}\varphi = \varphi\end{cases}$$

for all $\varphi \in \mathcal{S}$.

Having a tight control on the action of $\mathrm{I}^*_{\omega_0,t_0}$ over $\mathcal{S}$ allows us to extend the right-inverse operator $\mathrm{I}_{\omega_0,t_0}$ to an appropriate subset of tempered distributions $\phi \in \mathcal{S}'$ according to the rule $\langle \mathrm{I}_{\omega_0,t_0}\phi, \varphi\rangle = \langle \phi, \mathrm{I}^*_{\omega_0,t_0}\varphi\rangle$. Our complete set of inverse operators is summarized in Table I together with their equivalent Fourier-based definitions which are also interpretable in the generalized sense of distributions. The first three entries of the table are standard results from the theory of linear systems (e.g., [49, Table 4.1]), while the other operators are specific to this work.

D. Solution of Generic Stochastic Differential Equation

We now have all the elements to solve the generic stochastic linear differential equation

$$\sum_{n=0}^{N} a_n \mathrm{D}^n s = \sum_{m=0}^{M} b_m \mathrm{D}^m w \tag{23}$$

TABLE I
FIRST-ORDER DIFFERENTIAL OPERATORS AND THEIR INVERSES

L	$\mathrm{L}^{-1}f(t)$	Properties of inverse operator
Standard case: $\alpha_n \in \mathbb{C}, \mathrm{Re}(\alpha_n) \neq 0$		
$(\mathrm{D}-\alpha_n\mathrm{Id})$	$(\mathrm{D}-\alpha_n\mathrm{Id})^{-1}f(t) = \int_{\mathbb{R}} \hat{f}(\omega)\left(\frac{1}{j\omega-\alpha_n}\right)e^{j\omega t}\frac{\mathrm{d}\omega}{2\pi}$	L_p-stable, LSI, $\mathcal{S}$-continuous
$(\mathrm{D}-\alpha_n\mathrm{Id})^*$	$(\mathrm{D}^*-\alpha_n\mathrm{Id})^{-1}f(t) = \int_{\mathbb{R}} \hat{f}(\omega)\left(\frac{1}{-j\omega-\alpha_n}\right)e^{j\omega t}\frac{\mathrm{d}\omega}{2\pi}$	L_p-stable, LSI, $\mathcal{S}$-continuous
Critical case: $\alpha_n = j\omega_0, \omega_0 \in \mathbb{R}$		
$(\mathrm{D}-j\omega_0\mathrm{Id})$	$\mathrm{I}_{\omega_0}f(t) = \int_{\mathbb{R}} \hat{f}(\omega)\left(\frac{1}{j(\omega-\omega_0)}+\pi\delta(\omega-\omega_0)\right)e^{j\omega t}\frac{\mathrm{d}\omega}{2\pi}$	Causal, LSI
	$\mathrm{I}_{\omega_0,t_0}f(t) = \int_{\mathbb{R}} \hat{f}(\omega)\left(\frac{e^{j\omega t}-e^{j\omega_0(t-t_0)}e^{j\omega t_0}}{j(\omega-\omega_0)}\right)\frac{\mathrm{d}\omega}{2\pi}$	Output vanishes at $t=t_0$
$(\mathrm{D}-j\omega_0\mathrm{Id})^*$	$\mathrm{I}^*_{\omega_0}f(t) = \int_{\mathbb{R}} \left(\frac{\hat{f}(\omega)}{-j(\omega+\omega_0)}+\hat{f}(-\omega_0)\pi\delta(\omega+\omega_0)\right)e^{j\omega t}\frac{\mathrm{d}\omega}{2\pi}$	Anti-causal, LSI
	$\mathrm{I}^*_{\omega_0,t_0}f(t) = \int_{\mathbb{R}} \left(\frac{\hat{f}(\omega)-\hat{f}(-\omega_0)e^{-j(\omega+\omega_0)t_0}}{-j(\omega+\omega_0)}\right)e^{j\omega t}\frac{\mathrm{d}\omega}{2\pi}$	L_p-stable and decay preserving

where the a_n and b_m are arbitrary complex coefficients with the normalization constraint $a_N = 1$. While this reminds us of the textbook formula of an ordinary Nth-order differential system, the non-standard aspect in (23) is that the driving term is a innovation process w, which is generally not defined pointwise, and that we are not imposing any stability constraint. Eq. (23) thus covers the general case (12) where L is a shift-invariant operator with the rational transfer function

$$\hat{L}(\omega) = \frac{(j\omega)^N + a_{N-1}(j\omega)^{N-1} + \cdots + a_1(j\omega) + a_0}{b_M(j\omega)^M + \cdots + b_1(j\omega) + b_0} = \frac{P_N(j\omega)}{Q_M(j\omega)}. \tag{24}$$

The poles of the system, which are the roots of the characteristic polynomial $P_N(\zeta) = \zeta^N + a_{N-1}\zeta^{n-1} + \cdots + a_0$ with Laplace variable $\zeta \in \mathbb{C}$, are denoted by $\{\alpha_n\}_{n=1}^N$. While we are not imposing any restriction on their locus in the complex plane, we are adopting a special ordering where the purely imaginary roots (if present) are coming last. This allows us to factorize the numerator of (24) as

$$P_N(j\omega) = \prod_{n=1}^{N}(j\omega-\alpha_n) = \left(\prod_{n=1}^{N-n_0}(j\omega-\alpha_n)\right)\left(\prod_{m=1}^{n_0}(j\omega-j\omega_m)\right) \tag{25}$$

with $\alpha_{N-n_0+m} = j\omega_m$, $1 \leq m \leq n_0$, where n_0 is the number of purely-imaginary poles. The operator counterpart of this last equation is the decomposition

$$P_N(\mathrm{D}) = \underbrace{(\mathrm{D}-\alpha_1\mathrm{Id})\cdots(\mathrm{D}-\alpha_{N-n_0}\mathrm{Id})}_{\text{regular part}} \circ \underbrace{(\mathrm{D}-j\omega_1\mathrm{Id})\cdots(\mathrm{D}-j\omega_{n_0}\mathrm{Id})}_{\text{critical part}}$$

which involves a cascade of elementary first-order components. By applying the proper sequence of right-inverse operators from Table I, we can then formally solve the system as in (13). The resulting inverse operator is

$$\mathrm{L}^{-1} = \underbrace{\mathrm{I}_{\omega_{n_0},t_{n_0}}\cdots\mathrm{I}_{\omega_1,t_1}}_{\text{shift-variant}}\mathrm{T}_{\mathrm{LSI}} \tag{26}$$

with

$$\mathrm{T}_{\mathrm{LSI}} = (\mathrm{D}-\alpha_{N-n_0}\mathrm{Id})^{-1}\cdots(\mathrm{D}-\alpha_1\mathrm{Id})^{-1}Q_M(\mathrm{D}),$$

which imposes the n_0 boundary conditions

$$\begin{cases} s(t)|_{t=t_{n_0}} = 0 \\ (\mathrm{D}-j\omega_{n_0}\mathrm{Id})s(t)\big|_{t=t_{n_0-1}} = 0 \\ \quad\vdots \\ (\mathrm{D}-j\omega_2\mathrm{Id})\cdots(\mathrm{D}-j\omega_{n_0}\mathrm{Id})s(t)\big|_{t=t_1} = 0. \end{cases} \tag{27}$$

Implicit in the specification of these boundary conditions is the property that s and its derivatives up to order n_0-1 admit a pointwise interpretation in the neighborhood of $(t_1,\ldots,t_{n_0})$. This can be shown with the help of Theorem 5. For instance, if $n_0 = 1$ and $\omega_1 = 0$, then $s(t)$ with t fixed is given by (17) with $h(t,\cdot) = \mathrm{T}^*_{\mathrm{LSI}}\{\mathbb{1}_{[0,t)}\} \in \mathcal{R} \subset L_p$.

The adjoint of the operator specified by (26) is

$$\mathrm{L}^{-1*} = \mathrm{T}^*_{\mathrm{LSI}}\underbrace{\mathrm{I}^*_{\omega_1,t_1}\cdots\mathrm{I}^*_{\omega_{n_0},t_{n_0}}}_{\text{shift-variant}}, \tag{28}$$

and is guaranteed to be a continuous linear mapping from $\mathcal{S}$ into $\mathcal{R}$ by Theorem 6, the key point being that each of the component operators preserves the rapid decay of the test function to which it is applied. The last step is to substitute the explicit form (28) of L^{-1*} into (14) with a $\widehat{\mathscr{P}_w}$ that is well-defined on $\mathcal{R}$, which yields the characteristic form of the

stochastic process s defined by (23) subject to the boundary conditions (27).

We close this section with a comment about commutativity: while the order of application of the operators $Q_M(\mathrm{D})$ and $(\mathrm{D}-\alpha_n\mathrm{Id})^{-1}$ in the LSI part of (26) is immaterial (thanks to the commutativity of convolution), it is not so for the inverse operators $\mathrm{I}_{\omega_m,t_0}$ that appear in the "shift-variant" part of the decomposition. The latter do not commute and their order of application is tightly linked to the boundary conditions.

V. Sparse Stochastic Processes

This section is devoted to the characterization and investigation of the properties of the broad family of stochastic processes specified by the innovation model (12) where L is LSI. It covers the non-Gaussian stationary processes (V-A), which are generated by conventional analog filtering of a sparse innovation, as well as the whole class of processes that are solution of the (possibly unstable) differential equation (23) with a Lévy noise excitation (V-B). The latter category constitutes the higher-order generalization of the classical Lévy processes, which are non-stationary. The proposed method is constructive and essentially boils down to the specification of appropriate families of shaping operators L^{-1} and to making sure that the admissibility conditions in Theorem 4 are met.

A. Non-Gaussian Stationary Processes

The simplest scenario is when L^{-1} is LSI and can be decomposed into a cascade of BIBO-stable and ordinary differential operators. If the BIBO-stable part is rapidly-decreasing, then L^{-1} is guaranteed to be $\mathcal{S}$-continuous. In particular, this covers the case of an Nth-order differential system without any pole on the imaginary axis, as justified by our analysis in Section IV-D.

Proposition 3 (Generalized Stationary Processes): Let L^{-1} (the right-inverse of some operator L) be a $\mathcal{S}$-continuous convolution operator characterized by its impulse response $\rho_{\mathrm{L}}=\mathrm{L}^{-1}\delta$. Then, the generalized stochastic processes that are defined by $\widehat{\mathscr{P}_s}(\varphi)=\exp\left(\int_{\mathbb{R}} f\left(\rho_{\mathrm{L}}^{\vee}*\varphi(t)\right)\mathrm{d}t\right)$ where $f(\omega)$ is of the generic form (6) are *stationary* and well-defined solutions of the operator equation (12) driven by some corresponding innovation process w.

Proof: The fact that these generalized processes are well-defined is a direct consequence of the Minlos-Bochner Theorem since L^{-1*} (the convolution with $\rho_{\mathrm{L}}^{\vee}$) satisfies the first admissibility condition in Theorem 4. The stationarity property is equivalent to $\widehat{\mathscr{P}_s}(\varphi)=\widehat{\mathscr{P}_s}(\varphi(\cdot-t_0))$ for all $t_0\in\mathbb{R}$; it is established by simple change of variable in the inner integral using the basic shift-invariance property of convolution; i.e., $\left(\rho_{\mathrm{L}}^{\vee}*\varphi(\cdot-t_0)\right)(t)=(\rho_{\mathrm{L}}^{\vee}*\varphi)(t-t_0)$. ∎

The above characterization is not only remarkably concise, but also quite general. It extends the traditional theory of stationary Gaussian processes, which corresponds to the choice $f(\omega)=-\frac{\sigma_0^2}{2}\omega^2$. The Gaussian case results in the simplified form $\int_{\mathbb{R}} f(\mathrm{L}^{-1*}\varphi(t))\mathrm{d}t=-\frac{\sigma_0^2}{2}\|\rho_{\mathrm{L}}^{\vee}*\varphi\|_{L_2}^2=-\frac{1}{4\pi}\int_{\mathbb{R}}\Phi_s(\omega)|\hat{\varphi}(\omega)|^2\mathrm{d}\omega$ (using Parseval's identity) where $\Phi_s(\omega)=\frac{\sigma_0^2}{|\hat{L}(-\omega)|^2}$ is the spectral power density that is associated with the innovation model. The interest here is that we get access to a much broader family of non-Gaussian processes (e.g., generalized Poisson or alpha-stable) with matched spectral properties since they share the same whitening operator L.

The characteristic form condenses all the statistical information about the process. For instance, by setting $\varphi=\omega\delta(\cdot-t_0)$, we can explicitly determine $\widehat{\mathscr{P}_s}(\varphi)=\mathbb{E}\{e^{j\langle s,\varphi\rangle}\}=\mathbb{E}\{e^{j\omega s(t_0)}\}=\mathcal{F}\{p(s(t_0))\}(-\omega)$, which yields the characteristic function of the first-order probability density, $p(s(t_0))=p(s)$, of the sample values of the process. In the present stationary scenario, we find that $p(s)=\mathcal{F}^{-1}\{\exp\left(\int_{\mathbb{R}} f\left(-\omega\rho_{\mathrm{L}}(t)\right)\mathrm{d}t\right)\}(s)$, which requires the evaluation of an integral followed by an inverse Fourier transform. While this type of calculation is only tractable analytically in special cases, it may be performed numerically with the help of the FFT. Higher-order density functions are accessible as well as at the cost of some multi-dimensional inverse Fourier transforms. The same applies for moments which can be obtained through a simpler differentiation process, as exemplified in Section V-C.

B. Generalized Lévy Processes

The further reaching aspect of the present formulation is that it is also applicable to the characterization of non-stationary processes such as Brownian motion and Lévy processes, which are usually treated separately from the stationary ones, and that it naturally leads to the identification of a whole variety of higher-order extensions. The commonality is that these non-stationary processes can all be derived as solutions of an (unstable) Nth-order differential equation with some poles on the imaginary axis. This corresponds to the setting in Section IV-D with $n_0>0$.

Proposition 4 (Generalized Nth-order Lévy Processes): Let L^{-1} (the right-inverse of an Nth-order differential operator L) be specified by (26) with at least one non-shift-invariant factor $\mathrm{I}_{\omega_1,t_1}$. Then, the generalized stochastic processes that are defined by $\widehat{\mathscr{P}_s}(\varphi)=\exp\left(\int_{\mathbb{R}} f\left(\mathrm{L}^{-1*}\varphi(t)\right)\mathrm{d}t\right)$ where f is a p-admissible Lévy exponent are well-defined solutions of the stochastic differential equation (23) driven by some corresponding Lévy innovation w. These processes satisfy the boundary conditions (27) and are *non-stationary*.

Proof: The result is a direct consequence of the analysis in Section IV-D—in particular, Eqs. (26)–(28)—and Proposition 2. The latter implies that $\mathrm{L}^{-1*}\varphi$ is bounded in all $L_{\infty,m}$ norms with $m\geq 1$. Since $\mathcal{S}\subset L_{\infty,m}\subset L_p$ and the Schwartz topology is the strongest in this chain, we can infer that L^{-1*} is a continuous operator from $\mathcal{S}$ onto any of the L_p spaces with $p\geq 1$. The existence claim then follows from the combination of Theorem 4 and Minlos-Bochner. Since $\mathrm{L}^{-1*}\varphi$ is not shift-invariant, there is no chance for these processes to be stationary, not to mention the fact that they fulfill the boundary conditions (27). ∎

Conceptually, we like to view the generalized stochastic processes of Proposition 4 as "adjusted" versions of the stationary ones that include some additional sinusoidal (or

polynomial) trends. While the generation mechanism of these trends is random, there is a deterministic aspect to it because it imposes the boundary conditions (27) at $t_1, \cdots, t_{n_0}$. The class of such processes is actually quite rich and the formalism surprisingly powerful. We shall illustrate the use of Proposition 4 in Section V with the simplest possible operator $\mathrm{L} = \mathrm{D}$ which will gets us back to Brownian motion and the celebrated family of Lévy processes. We shall also show how the well-known properties of Lévy processes can be readily deduced from their characteristic form.

C. Moments and Correlation

The covariance form of a generalized (complex-valued) process s is defined as:

$$\mathcal{B}_s(\varphi_1, \varphi_2) = \mathbb{E}\{\langle s, \varphi_1\rangle \cdot \overline{\langle s, \varphi_2\rangle}\}.$$

where $\overline{\langle s, \varphi_2\rangle} = \langle s, \varphi_2\rangle$ when s is real-valued. Thanks to the moment generating properties of the Fourier transform, this functional can be calculated from the characteristic form $\widehat{\mathscr{P}}_s(\varphi)$ as

$$\mathcal{B}_s(\varphi_1, \varphi_2) = (-j)^2 \left.\frac{\partial^2 \widehat{\mathscr{P}}_s(\omega_1\varphi_1 + \omega_2\varphi_2)}{\partial\omega_1\partial\omega_2}\right|_{\omega_1=0,\omega_2=0}, \quad (29)$$

where we are implicitly assuming that the required partial derivative of the characteristic functional exists. The autocorrelation of the process is then obtained by making the formal substitution $\varphi_1 = \delta(\cdot - t_1)$ and $\varphi_2 = \delta(\cdot - t_2)$:

$$R_s(t_1, t_2) = \mathbb{E}\{s(t_1)s(t_2)\} = \mathcal{B}_s\left(\delta(\cdot - t_1), \delta(\cdot - t_2)\right).$$

Alternatively, we can also retrieve the autocorrelation function by invoking the kernel theorem: $\mathcal{B}_s(\varphi_1, \varphi_2) = \int_{\mathbb{R}^2} R_s(t_1, t_2)\varphi_1(t_1)\varphi(t_2)\mathrm{d}t_1\mathrm{d}t_2$.

The concept also generalizes for the calculation of the higher-order correlation form[6]

$$\mathbb{E}\{\langle s, \varphi_1\rangle \cdot \langle s, \varphi_2\rangle \cdots \langle s, \varphi_N\rangle\} = (-j)^N \left.\frac{\partial^N \widehat{\mathscr{P}}_s(\omega_1\varphi_1 + \cdots + \omega_N\varphi_N)}{\partial\omega_1 \cdots \partial\omega_N}\right|_{\omega_1=0,\cdots,\omega_N=0}$$

which provides the basis for the determination of higher-order moments and cumulants.

Here, we concentrate on the calculation of the second-order moments, which happen to be independent upon the specific type of noise. For the cases where the covariance is defined and finite, it is not hard to show that the generic covariance form of the innovation processes defined in Section III-C is

$$\mathcal{B}_w(\varphi_1, \varphi_2) = \sigma_0^2\ \langle\varphi_1, \varphi_2\rangle,$$

where σ_0^2 is a suitable normalization constant that depends on the noise parameters (b_1, b_2, v) in (7)–(10). We then perform the usual adjoint manipulation to transfer the above formula to the filtered version $s = \mathrm{L}^{-1}w$ of such a noise process.

Property 1 (Generalized Correlation): The covariance form of the generalized stochastic process whose characteristic form is $\widehat{\mathscr{P}}_s(\varphi) = \widehat{\mathscr{P}}_w(\mathrm{L}^{-1*}\varphi)$ where $\widehat{\mathscr{P}}_w$ is a white noise functional is given by

$$\mathcal{B}_s(\varphi_1, \varphi_2) = \sigma_0^2\ \langle \mathrm{L}^{-1*}\varphi_1, \overline{\mathrm{L}^{-1*}\varphi_2}\rangle = \sigma_0^2\ \langle \overline{\mathrm{L}^{-1}}\mathrm{L}^{-1*}\varphi_1, \varphi_2\rangle,$$

and corresponds to the correlation function

$$R_s(t_1, t_2) = \mathbb{E}\{s(t_1) \cdot \overline{s(t_2)}\} = \sigma_0^2\ \langle \overline{\mathrm{L}^{-1}}\mathrm{L}^{-1*}\delta(\cdot - t_1), \delta(\cdot - t_2)\rangle.$$

The latter characterization requires the determination of the impulse response of $\overline{\mathrm{L}^{-1}}\mathrm{L}^{-1*}$. In particular, when L^{-1} is LSI with convolution kernel $\rho_{\mathrm{L}} \in L_1$, we get that

$$\begin{aligned} R_s(t_1, t_2) &= \sigma_0^2\ \overline{\mathrm{L}^{-1}}\mathrm{L}^{-1*}\delta(t_2 - t_1) = r_s(t_2 - t_1) \\ &= \sigma_0^2\ (\overline{\rho}_{\mathrm{L}} * \rho_{\mathrm{L}}^{\vee})(t_2 - t_1), \end{aligned}$$

which confirms that the underlying process is wide-sense stationary. Since the autocorrelation function $r_s(\tau)$ is integrable, we also have a one-to-one correspondence with the traditional notion of power spectrum: $\Phi_s(\omega) = \mathcal{F}\{r_s\}(\omega) = \frac{\sigma_0^2}{|\hat{L}(-\omega)|^2}$, where $\hat{L}(\omega)$ is the frequency response of the whitening operator L.

The determination of the correlation function for the non-stationary processes associated with the unstable versions of (23) is more involved. We shall see in [32] that it can be bypassed if, instead of $s(t)$, we consider the generalized increment process $s_{\mathrm{d}}(t) = \mathrm{L}_{\mathrm{d}}s(t)$ where L_{d} is a discrete version (finite-difference type operator) of the whitening operator L.

D. Sparsification in a Wavelet-Like Basis

The implicit assumption for the next properties is that we have a wavelet-like basis $\{\psi_{i,k}\}_{i\in\mathbb{Z},k\in\mathbb{Z}}$ available that is matched to the operator L. Specifically, the basis functions $\psi_{i,k}(t) = \psi_i(t - 2^i k)$ with scale and location indices (i, k) are translated versions of some normalized reference wavelet $\psi_i = \mathrm{L}^*\phi_i$ where ϕ_i is an appropriate scale-dependent smoothing kernel. It turns out that such operator-like wavelets can be constructed for the whole class of ordinary differential operators considered in this paper [36]. They can be specified to be orthogonal and/or compactly supported (cf. examples in Fig. 2). In the case of the classical Haar wavelet, we have that $\psi_{\mathrm{Haar}} = \mathrm{D}\phi_i$ where the smoothing kernels $\phi_i \propto \phi_0(t/2^i)$ are rescaled versions of a triangle function (B-spline of degree 1). The latter dilation property follows from the fact that the derivative operator D commutes with scaling.

We note that the determination of the wavelet coefficients $v_i[k] = \langle s, \psi_{i,k}\rangle$ of the random signal s at a given scale i is equivalent to correlating the signal with the wavelet ψ_i (continuous wavelet transform) and sampling thereafter. The goods news is that this has a stationarizing and decoupling effect.

Property 2 (Wavelet-Domain Probability Laws): Let $v_i(t) = \langle s, \psi_i(\cdot - t)\rangle$ with $\psi_i = \mathrm{L}^*\phi_i$ be the ith channel of the continuous wavelet transform of a generalized (stationary or non-stationary) Lévy process s with whitening operator L and p-admissible Lévy exponent f. Then, $v_i(t)$ is a generalized stationary process with characteristic functional $\widehat{\mathscr{P}}_{v_i}(\varphi) = \widehat{\mathscr{P}}_w(\phi_i * \varphi)$ where $\widehat{\mathscr{P}}_w$ is defined by (5). Moreover, the characteristic function of the (discrete) wavelet coefficient $v_i[k] = v_i(2^i k)$—that is, the Fourier transform of the pdf

[6]For simplicity, we are only giving the formula for a real-valued process.

$p_{v_i}(v)$—is given by $\hat{p}_{v_i}(\omega) = \widehat{\mathscr{P}}_w(\omega\phi_i) = e^{f_i(\omega)}$ and is infinitely divisible with modified Lévy exponent

$$f_i(\omega) = \int_{\mathbb{R}} f\big(\omega\phi_i(t)\big)\mathrm{d}t.$$

Proof: Recalling that $s = \mathrm{L}^{-1}w$, we get

$$\begin{aligned} v_i(t) &= \langle s, \psi_i(\cdot - t)\rangle = \langle \mathrm{L}^{-1}w, \mathrm{L}^*\phi_i(\cdot - t)\rangle \\ &= \langle w, \mathrm{L}^{-1*}\mathrm{L}^*\phi_i(\cdot - t)\rangle = \big(\phi_i^\vee * w\big)(t) \end{aligned}$$

where we have used the fact that L^{-1*} is a valid (continuous) left-inverse of L^*. The wavelet smoothing kernel $\phi_i \in \mathcal{R}$ has rapid decay (e.g., compactly-support or, at worst, exponential decay); this allows us to invoke Proposition 3 to prove the first part.

As for the second part, we start from the definition of the characteristic function:

$$\begin{aligned} \hat{p}_{v_i}(\omega) &= \mathbb{E}\{e^{j\omega v_i}\} = \mathbb{E}\{e^{j\omega\langle s, \psi_{i,k}\rangle}\} = \mathbb{E}\{e^{j\langle s, \omega\psi_i\rangle}\} \\ &\qquad\qquad \text{(by stationarity)} \\ &= \widehat{\mathscr{P}}_s(\omega\psi_i) = \widehat{\mathscr{P}}_w(\mathrm{L}^{-1*}\mathrm{L}^*\phi_i\omega) \\ &= \widehat{\mathscr{P}}_w(\omega\phi_i) = \exp\left(\int_{\mathbb{R}} f\big(\omega\phi_i(t)\big)\mathrm{d}t\right) \end{aligned}$$

where we have used the left-inverse property of L^{-1*} and the expression of the Lévy noise functional. The result then follows by identification. [7] ∎

We determine the joint characteristic function of any two wavelet coefficients $Y_1 = \langle s, \psi_{i_1,k_1}\rangle$ and $Y_2 = \langle s, \psi_{i_2,k_2}\rangle$ with indices (i_1, k_1) and (i_2, k_2) using a similar technique.

Property 3 (Wavelet Dependencies): The joint characteristic function of the wavelet coefficients $Y_1 = v_{i_1}[k_1] = \langle s, \psi_{i_1,k_1}\rangle$ and $Y_2 = v_{i_2}[k_2] = \langle s, \psi_{i_2,k_2}\rangle$ of the generalized stochastic process s in Property 2 is given by

$$\begin{aligned} \hat{p}_{Y_1,Y_2}(\omega_1, \omega_2) = \exp\Bigg(\int_{\mathbb{R}} & f\big(\omega_1\phi_{i_1}(t - 2^{i_1}k_1) \\ & + \omega_2\phi_{i_2}(t - 2^{i_2}k_2)\big)\mathrm{d}t\Bigg) \end{aligned}$$

where f is the Lévy exponent of the innovation process w. The coefficients are independent if the kernels $\phi_{i_1}(t - 2^{i_1}k_1)$ and $\phi_{i_2}(t - 2^{i_2}k_2)$ have disjoint support; their correlation is given by

$$\mathbb{E}\{Y_1Y_2\} = \sigma_0^2\ \langle \phi_{i_1}(\cdot - 2^{i_1}k_1), \phi_{i_2}(\cdot - 2^{i_2}k_2)\rangle.$$

under the assumption that the variance σ_0^2 of w is finite.

Proof: The first formula is obtained by substitution of $\varphi = \omega_1\psi_{i_1,k_1} + \omega_2\psi_{i_2,k_2}$ in $\mathbb{E}\{e^{j\langle s,\varphi\rangle}\} = \widehat{\mathscr{P}}_w(\mathrm{L}^{-1*}\varphi)$, and simplification using the left-inverse property of L^{-1*}. The statement about independence follows from the exponential nature of the characteristic function and the property that $f(0) = 0$, which allows for the factorization of the characteristic function when the support of the kernels are distinct (independence of the noise at every point). The correlation formula is obtained by direct application of the first result in Property 1 with $\varphi_1 = \psi_{i_1,k_1} = \mathrm{L}^*\phi_{i_1}(\cdot - 2^{i_1}k_1)$ and $\varphi_2 = \psi_{i_2,k_2} = \mathrm{L}^*\phi_{i_2}(\cdot - 2^{i_2}k_2)$. ∎

These results provide a complete characterization of the statistical distribution of sparse stochastic processes in some matched wavelet domain. They also indicate that the representation is intrinsically sparse since the transformed-domain statistics are infinitely divisible. Practically, this translates into the wavelet domain pdfs being heavier tailed than a Gaussian (unless the process is Gaussian) (cf. argumentation in Section III-D).

To make matters more explicit, we consider the case where the innovation process is SαS. The application of Property 2 with $f(\omega) = -\frac{|\omega|^\alpha}{\alpha!}$ yields $f_i(\omega) = -\frac{|\sigma_i\omega|^\alpha}{\alpha!}$ with dispersion parameter $\sigma_i = \|\phi_i\|_{L_\alpha}$. This proves that the wavelet coefficients of a generalized SαS stochastic process follow SαS distributions with the spread of the pdf at scale i being determined by the L_α norm of the corresponding wavelet smoothing kernels. This strongly suggests that, for $\alpha < 2$, the process is compressible in the sense that the essential part of the "energy content" is carried by a tiny fraction of wavelet coefficients, as illustrated in Fig. 1.

It should be noted, however, that the quality of the decoupling is strongly dependent upon the spread of the wavelet smoothing kernels ϕ_i which should be chosen to be maximally localized for best performance. In the case of the first-order system (cf. example in Section II), the basis functions for i fixed are not overlapping which implies that the wavelet coefficients within a given scale are independent. This is not so across scale because of the cone-shaped region where the support of the kernels ϕ_{i_1} and ϕ_{i_2} overlap, which induces dependencies. Incidentally, the inter-scale correlation of wavelet coefficients is often exploited for improving coding performance [50] and signal reconstruction by imposing joint sparsity constraints [51].

VI. Lévy Processes Revisited

We now illustrate our method by specifying classical Lévy processes—denoted by $W(t)$—via the solution of the (marginally unstable) stochastic differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}W(t) = w(t) \tag{30}$$

where the driving term w is one of the independent noise processes defined earlier. It is important to keep in mind that Eq. (30), which is the limit of (2) as $\alpha \to 0$, is only a notation whose correct interpretation is $\langle \mathrm{D}W, \varphi\rangle = \langle w, \varphi\rangle$ for all $\varphi \in \mathcal{S}$. We shall consider the solution $W(t)$ for all $t \in \mathbb{R}$, but we shall impose the boundary condition $W(t_0) = 0$ with $t_0 = 0$ to make our construction compatible with the classical one which is defined for $t \geq 0$.

A. Distributional Characterization of Lévy Processes

The direct application of the operator formalism developed in Section III yields the solution of (30):

$$W(t) = \mathrm{I}_{0,0}w(t)$$

where $\mathrm{I}_{0,0}$ is the unique right inverse of D that imposes the required boundary condition at $t = 0$. The Fourier-based

[7] A technical remark is in order here: the substitution of a non-smooth function such as $\phi_i \in \mathcal{R}$ in the characteristic noise functional $\widehat{\mathscr{P}}_w$ is legitimate provided that the domain of continuity of the functional can be extended from $\mathcal{S}$ to $\mathcal{R}$, or, even less restrictively, to L_p when f is p-admissible (see Theorem 3).

expression of this anti-derivative operator is obtained from the 6th line of Table I by setting $(\omega_0, t_0) = (0, 0)$. By using the properties of the Fourier transform, we obtain the simplified expression

$$\mathrm{I}_{0,0}\varphi(t) = \begin{cases} \int_0^t \varphi(\tau)\mathrm{d}\tau, & t \geq 0 \\ -\int_t^0 \varphi(\tau)\mathrm{d}\tau, & t < 0, \end{cases} \tag{31}$$

which allows us to interpret $W(t)$ as the integrated version of w with the proper boundary conditions. Likewise, we derive the time-domain expression of the adjoint operator

$$\mathrm{I}_{0,0}^*\varphi(t) = \begin{cases} \int_t^\infty \varphi(\tau)\mathrm{d}\tau, & t \geq 0, \\ -\int_{-\infty}^t \varphi(\tau)\mathrm{d}\tau, & t < 0. \end{cases} \tag{32}$$

Next, we invoke Proposition 4 to obtain the characteristic form of the Lévy process

$$\widehat{\mathscr{P}}_W(\varphi) = \widehat{\mathscr{P}}_w(\mathrm{I}_{0,0}^*\varphi) \tag{33}$$

which is admissible provided that the Lévy exponent f fullfils the condition in Theorem 4.

We get the characteristic function of the sample values of the Lévy process $W(t_1) = \langle W, \delta(\cdot - t_1)\rangle$ by making the substitution $\varphi = \omega_1\delta(\cdot - t_1)$ in (33): $\widehat{\mathscr{P}}_W\big(\omega_1\delta(\cdot - t_1)\big) = \widehat{\mathscr{P}}_w\big(\omega_1\mathrm{I}_{0,0}^*\delta(\cdot - t_1)\big)$ with $t_1 > 0$. We then use (31) to evaluate $\mathrm{I}_{0,0}^*\delta(t - t_1) = \mathbb{1}_{(0,t_1]}(t)$. Since the latter indicator function is equal to one for $t \in (0, t_1]$ and zero elsewhere, it is easy to evaluate the integral over t in (5) with $f(0) = 0$, which yields

$$\begin{aligned} \mathbb{E}\{e^{j\omega_1 W(t_1)}\} &= \exp\left(\int_{\mathbb{R}} f\big(\omega_1 \mathbb{1}_{(0,t_1]}(t)\big)\mathrm{d}t\right) \\ &= e^{t_1 f(\omega_1)} \end{aligned}$$

This result is equivalent to the celebrated Lévy-Khinchine representation of the process [31].

B. Lévy Increments vs. Wavelet Coefficients

A fundamental property of Lévy processes is that their increments at equally-spaced intervals are i.i.d. [31]. To see how this fits into the present framework, we specify the increments on the integer grid as the special case of (3) with $\alpha = 0$:

$$\begin{aligned} u[k] = \Delta_0 W(k) &:= W(k) - W(k-1) \\ &= \int_{k-1}^{k} w(t)\mathrm{d}t = \langle w, \beta_0^\vee(\cdot - k)\rangle \end{aligned}$$

where $\beta_0(t) = \mathbb{1}_{[0,1)}(t) = \Delta_0\rho_0(t)$ is the causal B-spline of degree 0 (rectangular function). We are also introducing some new notation, which is consistent with the definitions given in [32, Table II], to set the stage for the generalizations to come. Δ_0 is the finite-difference operator, which is the discrete analog of the derivative operator D, while ρ_0 (unit step) is the Green function of the derivative operator D. The main point of the exercise is to show that determining increments is structurally equivalent to the computation of the wavelet coefficients in Property 2 with the smoothing kernel ϕ_i being substituted by $\beta_0^\vee$. It follow that the characteristic function of $u[\cdot]$ is given by

$$\hat{p}_u(\omega) = \exp\big(\textstyle\int_{\mathbb{R}} f(\omega\beta_0^\vee(t))\mathrm{d}t\big) = e^{f(\omega)} = \hat{p}_{\mathrm{id}}(\omega) \tag{34}$$

where the simplification of the integral results from the binary nature of β_0 which is either 1 (on a support of size 1) or zero. This implies that the increments of the Lévy process are independent (because the B-spline functions $\beta_0^\vee(\cdot - k)$ are non-overlapping) and that their pdf is given by the canonical id distribution of the innovation process $p_{\mathrm{id}}(x)$ (cf. discussion in Section III-D).

The alternative is to expand the Lévy process in the Haar basis which is ideally matched to it. Indeed, the Haar wavelet at scale $i = 1$ (lower-left function in Fig. 2) can be expressed as

$$\psi_{\mathrm{Haar}}(t/2) = \beta_0(t) - \beta_0(t-1) = \Delta_0\beta_0 = \mathrm{D}\beta_{(0,0)}(t) \tag{35}$$

where $\beta_{(0,0)} = \beta_0 * \beta_0$ is the causal B-spline of degree 1 (triangle function). Since $\mathrm{D}^* = -\mathrm{D}$, this confirms that the underlying smoothing kernels are dilated versions of a B-spline of degree 1. Moreover, since the wavelet-domain sampling is critical, there is no overlap of the basis functions within a given scale which implies that the wavelets coefficients are independent on a scale-by-scale basis (cf. Property 3). If we now compare the situation with that of the Lévy increments, we observe that the wavelet analysis involves one more layer of smoothing of the innovation with β_0 (due to the factorization property of $\beta_{(0,0)}$) which slightly complicates the statistical calculations.

While the smoothing effect on the innovation is qualitatively the same in both instances, there are fundamental differences, too. In the wavelet case, the underlying discrete transform is orthogonal, but the coefficients are not fully decoupled because of the inter-scale dependencies which are unavoidable, as explained in Section V-D. By contrast, the decoupling of the Lévy increments is perfect, but the underlying discrete transform (finite difference transform) is non-orthogonal. In our companion paper, we shall see how this latter strategy is extendable to the much broader family of sparse processes via the definition of the *generalized increment process*.

C. Examples of Lévy Processes

Realizations of four different Lévy processes are shown in Fig. 3 together with their Lévy triplets $\big(b_1, b_2, v(a)\big)$. The first signal is a Brownian motion (a.k.a. Wiener process) that is obtained by integration of a white Gaussian noise. This classical process is known to be nowhere differentiable in the classical sense, despite the fact that it is continuous everywhere (almost surely) as all the members of the Lévy family. While the sampled version of $\Delta_0 W$ is i.i.d. in all cases, it does not yield a sparse representation in this first instance because the underlying distribution remains Gaussian. The second process, which may be termed Lévy-Laplace motion, is specified by the Lévy density $v(a) = e^{-|a|}/|a|$ which is not in L_1. By taking the inverse Fourier transform of (34), we can show that its increment process has a Laplace distribution [18]; note that this type of generalized Gaussian model is often used to justify sparsity-promoting signal processing techniques based on ℓ_1 minimization [52]–[54]. The third piecewise-constant signal is a compound Poisson process. It is intrinsically sparse since a

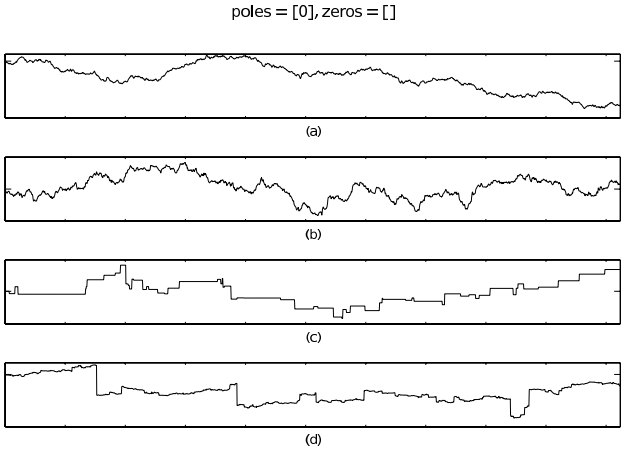

Fig. 3. Examples of Lévy motions $W(t)$ with increasing degrees of sparsity. (a) Brownian motion with Lévy triplet $(0, 1, 0)$. (b) Lévy-Laplace motion with $\left(0, 0, \frac{e^{-|a|}}{|a|}\right)$. (c) Compound Poisson process with $\left(0, 0, \lambda \frac{1}{\sqrt{2\pi}} e^{-a^2/2}\right)$ with $\lambda = \frac{1}{32}$. (d) Symmetric Lévy flight with $\left(0, 0, 1/|a|^{\alpha+1}\right)$ and $\alpha = 1.2$.

good proportion of its increments is zero by construction (with probability $e^{-\lambda}$). Interestingly, this is the only type of Lévy process that fulfills the finite rate of innovation property [17]. The fourth example is an alpha-stable Lévy motion (a.k.a. Lévy flight) with $\alpha = 1.2$. Here, the distribution of $\Delta_0 W$ is heavy-tailed (SαS) with unbounded moments for $p > \alpha$. Although this may not be obvious from the picture, this is the sparsest process of the lot because it is ℓ_α-compressible in the strongest sense [45]. Specifically, we can compress the sequence such as to preserve any prescribed portion $r < 1$ of its average ℓ_α energy by retaining an arbitrarily small fraction of samples as the length of the signal goes to infinity.

D. Link With Conventional Stochastic Calculus

Thanks to (30), we can view a white noise $w = \dot{W}$ as the weak derivative of some classical Lévy processes $W(t)$ which is well-defined pointwise (almost everywhere). This provides us with further insights on the range of admissible innovation processes of Section II.C which constitute the driving terms of the general stochastic differential equation (12). This fundamental observation also makes the connection with stochastic calculus[8] [55], [56], which avoids the notion of white noise by relying on the use of stochastic integrals of the form

$$s(t) = \int_{\mathbb{R}} h(t, t') \mathrm{d}W(t')$$

where W is a random (signed) measure associated to some canonical Brownian motion (or, by extension, a Lévy process) and where $h(t, t')$ is an integration kernel that formally corresponds to our inverse operator L^{-1} (see Theorem 5).

VII. Conclusion

We have set the foundations of a unifying framework that gives access to the broadest possible class of continuous-time stochastic processes specifiable by linear, shift-invariant equations, which is beneficial for signal processing purposes. We have shown that these processes admit a concise representation in a wavelet-like basis. We have applied our framework to the description of the classical Lévy processes, which, in our view, provide the simplest and most basic examples of sparse processes, despite the fact that they are non-stationary. We have also hinted at the link between Lévy increments and splines, which is the theme that we shall develop in full generality next [32].

We have demonstrated that the proposed class of stochastic models and the corresponding mathematical machinery (Fourier analysis, characteristic functional, and B-spline calculus) lends itself well to the derivation of transform-domain statistics. The formulation suggests a variety of new processes whose properties are compatible with the currently-dominant paradigm in the field which is focused on the notion of sparsity. In that respect, the sparse processes that are best matched to conventional wavelets[9] are those generated by N-fold integration (with proper boundary conditions) of a non-gaussian innovation. These processes, which are the solution of an unstable SDE (pole of multiplicity N at the origin), are intrinsically self-similar (fractal) and non-stationary. Last but not least, the formulation is backward compatible with the classical theory of Gaussian stationary processes.

Acknowledgment

The authors are thankful to Prof. Robert Dalang (EPFL Chair of Probabilities), Julien Fageot and Dr. Arash Amini for helpful discussions.

References

[1] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York, NY, USA: McGraw-Hill, 1991.

[2] R. Gray and L. Davisson, *An Introduction to Statistical Signal Processing*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[8]The Itô integral of conventional stochastic calculus is based on Brownian motion, but the concept can also be generalized to Lévy driving terms using the more advanced theory of semimartingales [55].

[9]A wavelet with N vanishing moments can always be rewritten as $\psi = \mathrm{D}^N \phi$ with $\phi \in L_2(\mathbb{R})$ where the operator $\mathrm{L} = \mathrm{D}^N$ is scale-invariant.

[3] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[4] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.

[5] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd ed. San Diego, CA, USA: Academic Press, 2009.

[6] J.-L. Starck, F. Murtagh, and J. M. Fadili, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge. U.K.: Cambridge Univ. Press, 2010.

[7] M. Elad, *Sparse and Redundant Representations. From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer-Verlag, 2010.

[8] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[9] R. Baraniuk, E. Candes, M. Elad, and Y. Ma, "Applications of sparse representation and compressive sensing," *Proc. IEEE*, vol. 98, no. 6, pp. 906–909, Jun. 2010.

[10] M. Elad, M. Figueiredo, and Y. Ma, "On the role of sparse and redundant representations in image processing," *Proc. IEEE*, vol. 98, no. 6, pp. 972–982, Jun. 2010.

[11] M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, Aug. 2003.

[12] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

[13] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[14] Y. C. Eldar, "Compressed sensing of analog signals in shift-invariant spaces," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 2986–2997, Aug. 2009.

[15] B. Adcock and A. Hansen, *Generalized Sampling and Infinite-Dimensional Compressed Sensing*. DAMTP Tech. Rep. 2011/NA13, Cambridge Univ., 2011.

[16] T. Kailath, "The innovations approach to detection and estimation theory," *Proc. IEEE*, vol. 58, no. 5, pp. 680–695, May 1970.

[17] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1417–1428, Jun. 2002.

[18] M. Unser and P. D. Tafti, "Stochastic models for sparse and piecewise-smooth signals," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 989–1005, Mar. 2011.

[19] A. Swami, G. B. Giannakis, and J. M. Mendel, "Linear modeling of multidimensional non-Gaussian processes using cumulants," *Multidimensional Syst. Signal Process.*, vol. 1, no. 1, pp. 11–37, 1990.

[20] P. Rao, D. Johnson, and D. Becker, "Generation and analysis of non-Gaussian Markov time series," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 845–856, Apr. 1992.

[21] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus*, 2nd ed. New York, NY, USA: Springer-Verlag, 1991.

[22] B. Økensdal, *Stochastic Differential Equations*, 6th ed. New York, NY, USA: Springer-Verlag, 2007.

[23] M. Unser, "Cardinal exponential splines: Part II—Think analog, act digital," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1439–1449, Apr. 2005.

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

[25] A. Amini, U. S. Kamilov, E. Bostan, and M. Unser, "Bayesian estimation for continuous-time sparse stochastic processes," *IEEE Trans. Signal Process.*, vol. 61, no. 4, pp. 907–920, Feb. 2013.

[26] U. S. Kamilov, P. Pad, A. Amini, and M. Unser, "MMSE estimation of sparse Lévy processes," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 137–147, Jan. 2013.

[27] A. Amini, P. Thévenaz, J. Ward, and M. Unser, "On the linearity of Bayesian interpolators for non-Gaussian continuous-time AR(1) processes," *IEEE Trans. Inf. Theory*, vol. 59, no. 8, pp. 5063–5074, Aug. 2013.

[28] D. Appelbaum, *Lèvy Processes and Stochastic Calculus*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[29] I. M. Gelfand and N. Y. Vilenkin, *Generalized Functions*, vol. 4. San Diego, CA, USA: Academic press, 1964.

[30] P. Lévy, *Le mouvement Brownien*. Paris, France: Gauthier–Villars, 1954.

[31] K.-I. Sato, *Lévy Processes and Infinitely Divisible Distributions*. Boston, MA, USA: Chapman & Hall, 1994.

[32] M. Unser, P. D. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian vs. sparse stochastic processes—Part II: Discrete-domain theory," *IEEE Trans. Inf. Theory*, in press.

[33] N. Ahmed, "Discrete cosine transform," *IEEE Trans. Commun.*, vol. 23, no. 1, pp. 90–93, Sep. 1974.

[34] M. Unser, "On the approximation of the discrete Karhunen-Loève transform for stationary processes," *Signal Process.*, vol. 7, no. 3, pp. 231–249, Dec. 1984.

[35] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Application to Speech and Video Coding*. Upper Saddle River, NJ, USA:Prentice-Hall, 1984.

[36] I. Khalidov and M. Unser, "From differential equations to the construction of new wavelet-like bases," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1256–1267, Apr. 2006.

[37] J. Stewart, "Positive definite functions and generalizations, an historical survey," *Rocky Mountain J. Math.*, vol. 6, no. 3, pp. 409–434, 1976.

[38] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 2. 2nd ed. New York, NY, USA: Wiley, 1971.

[39] F. W. Steutel and K. Van Harn, *Infinite Divisibility of Probability Distributions on the Real Line*. New York, NY, USA: Marcel Dekker, 2003.

[40] G. Samorodnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Boston, MA, USA: Chapman & Hall, 1994.

[41] I. M. Gelfand and G. Shilov, *Generalized Functions*, vol. 1. New York, NY, USA: Academic press, 1964.

[42] A. Bose, A. Dasgupta, and H. Rubin, "A contemporary review and bibliography of infinitely divisible distributions and processes," *Indian J. Statist., Ser. A*, vol. 64, no. 3, pp. 763–819, 2002.

[43] B. Ramachandran, "On characteristic functions and moments," *Indian J. Statist., Ser. A*, vol. 31, no. 1, pp. 1–12, 1969.

[44] S. J. Wolfe, "On moments of infinitely divisible distribution functions," *Ann. Math. Statist.*, vol. 42, no. 6, pp. 2036–2043, 1971.

[45] A. Amini, M. Unser, and F. Marvasti, "Compressibility of deterministic and random infinite sequences," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5193–5201, Nov. 2011.

[46] R. Gribonval, V. Cevher, and M. E. Davies, "Compressible distributions for high-dimensional statistics," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, pp. 5016–5034, Aug. 2012.

[47] A. H. Zemanian, *Distribution Theory and Transform Analysis: An Introduction to Generalized Functions, with Applications*. New York, NY, USA: Dover, 2010.

[48] W. Kaplan, *Operational Methods for Linear Systems*. Reading, MA, USA: Addison-Wesley, 1962.

[49] B. Lathi, *Signal Processing and Linear Systems*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[50] J. Shapiro, "Embedded image coding using zerotrees of wavelet coefficients," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 41, no. 12, pp. 3445–3462, Dec. 1993.

[51] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden Markov models," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 886–902, Apr. 1998.

[52] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, no. 3, pp. 296–310, Jul. 1993.

[53] M. W. Seeger and H. Nickisch, "Compressed sensing and Bayesian experimental design," in *Proc. 25th Int. Conf. Mach. Learn.*, 2008, pp. 912–919.

[54] S. Babacan, R. Molina, and A. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–64, Jan. 2010.

[55] P. Protter, *Stochastic Integration and Differential Equations*. New York, NY, USA: Springer-Verlag, 2004.

[56] P. J. Brockwell, "Lèvy-driven CARMA processes," *Ann. Inst. Statist. Math.*, vol. 53, no. 1, pp. 113–124, 2001.

Michael Unser (M'89–SM'94–F'99) received the M.S. (summa cum laude) and Ph.D. degrees in Electrical Engineering in 1981 and 1984, respectively, from the Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland. From 1985 to 1997, he worked as a scientist with the National Institutes of Health, Bethesda USA. He is now full professor and Director of the Biomedical Imaging Group at the EPFL.

His main research area is biomedical image processing. He has a strong interest in sampling theories, multiresolution algorithms, wavelets, the use of splines for image processing, and, more recently, stochastic processes. He has published about 250 journal papers on those topics.

Dr. Unser is currently member of the editorial boards of IEEE J. SELECTED TOPICS IN SIGNAL PROCESSING, *Foundations and Trends in Signal Processing*, SIAM J. Imaging Sciences, and the PROCEEDINGS OF THE IEEE. He co-organized the first IEEE International Symposium on Biomedical Imaging (ISBI 2002) and was the founding chair of the technical committee of the IEEE-SP Society on Bio Imaging and Signal Processing (BISP).

He received three Best Paper Awards (1995, 2000, 2003) from the IEEE Signal Processing Society, and two IEEE Technical Achievement Awards (2008 SPS and 2010 EMBS). He is an EURASIP Fellow and a member of the Swiss Academy of Engineering Sciences.

Pouya D. Tafti was born in Tehran in 1981. He received his BSc degree in Electrical Engineering from Sharif University of Technology, Tehran, in 2003, his MASc in Electrical and Computer Engineering from McMaster University, Hamilton, Ontario, in 2006, and his PhD in Computer, Information, and Communication Sciences from EPFL, Lausanne, in 2011. From 2006 to 2012 he was with the Biomedical Imaging Group at EPFL, where he worked on vector field imaging and statistical models for signal and image processing. He currently resides in Germany where he works as a data scientist.

Qiyu Sun received the BSc and PhD degree in mathematics from Hangzhou University, China in 1985 and 1990 respectively. He is a full professor with the Department of Mathematics, University of Central Florida. His prior position was with Zhejiang University (China), National University of Singapore (Singapore), Vanderbilt University, and University of Houston.

His research interests include sampling theory, Wiener's lemma, wavelet and frame theory, linear and nonlinear inverse problems, and Fourier analysis. He has published more than 100 papers on mathematics and signal processing, and written a book *An Introduction to Multiband Wavelets* (Zhejiang University Press, 2001) with Ning Bi and Daren Huang. He is on the editorial board of the journals *Advance in Computational Mathematics*, *Numerical Functional Analysis and Optimization* and *Sampling Theory in Signal and Imaging Processing*.