Sparsity and Infinite Divisibility

Arash Amini and Michael Unser, *Fellow, IEEE*.

***Abstract*—We adopt an innovation-driven framework and investigate the sparse/compressible distributions obtained by linearly measuring or expanding continuous-domain stochastic models. Starting from first principles, we show that all such distributions are necessarily infinitely divisible. This property is satisfied by many distributions used in statistical learning such as Gaussian, Laplace, and a wide range of fat-tailed distributions such as Student's-t and α-stable laws. However, it excludes some popular distributions used in compressed sensing such as the Bernoulli-Gaussian distribution and distributions that decay like $\exp\big(-\mathcal{O}(|x|^p)\big)$ for $1 < p < 2$. We further explore the implications of infinite divisibility on distributions and conclude that tail decay and unimodality are preserved by all linear functionals of the same continuous-domain process. We explain how these results help in distinguishing suitable variational techniques for statistically solving inverse problems such as denoising.**

***Index Terms*—Decay gap, infinite-divisibility, Lévy-Khinchine representation, Lévy process, sparse stochastic process.**

I. Introduction

GAUSSIAN processes are by far the most studied stochastic models. There are many advantages in favor of Gaussian models, such as simplicity of statistical analysis (*e.g.*, in inference problems), stability of the Gaussian distribution (*i.e.*, closedness under linear combinations), and unified parameterization of all marginal distributions. Yet, one of the downsides of Gaussian distributions is that they fail to properly represent sparse or compressible data, which is a good incentive for the study of alternative models.

The identification of compressible distributions (whose high-dimensional i.i.d. realizations likely consist of a small number of large elements that capture most of the energy of the sequence) is an active field of research where established findings indicate that rapidly decaying distributions such as Gaussian and Laplace are not compressible. Meanwhile, fat-tailed distributions are potential candidates for distributions with compressibility [1], [2]. For instance, mixture models are common representatives of the compressible distributions in the literature: one might think of a mixture of two zero-mean Gaussian laws with considerably different variances, where the outcome of the one with the smaller variance forms the insignificant or compressible part. The so-called Bernoulli-Gaussian distribution is an extreme case where one of the variances is zero.

Manuscript received on December 4, 2012

A. Amini is with EE department, Sharif university of technology, Tehran, Iran. Email: aamini@sharif.edu.

M. Unser is with the Biomedical Imaging Group (BIG), École polytechnique fédérale de Lausanne (EPFL), Lausanne, Switzerland. Email: michael.unser@epfl.ch

The research was supported by the European Research Council under Grant ERC-2010-AdG 267439-FUN-SP.

To understand physical phenomena, we need to establish mathematical models that are calibrated with a finite number of measurements. These measurements are usually described by discrete stochastic models. Nevertheless, there are two fundamentally different modeling approaches.

1) Assume a discrete-domain model right from the start.
2) Initially adopt a continuous-domain model and discretize it later to describe the measurements.

We refer to the two as the discrete and the discretized models, respectively. One way to formalize stochastic processes is through an innovation model. In words, we assume that the stochastic objects are linked to discrete/continuous-domain innovation processes by means of linear operators. In this paper, we study innovation-driven *discretized* models. The framework was recently introduced in [3], [4] under the name "Sparse Stochastic Processes".

A. Motivation

In many applications, the signals of interest possess sparse/compressible representations in some transform domains, although the observations are rarely sparse themselves. For analyzing such signals, it is befitting to establish sparse/compressible signal models. One way to incorporate sparsity into the model is to assume a sparsity-inducing probability distribution for the signal. Although the statistics of the signal are often known in the observation domain, the common trend is to assume a sparse/compressible distribution on the coefficients of a sparse representation. The typical example is to assume independent and identically distributed (i.i.d.) coefficients in a transform domain with Bernoulli-Gaussian law [5].

Our approach in this paper is to assume a continuous-domain innovation-driven model for the signal, where the statistics are imposed on the innovation process. The continuous-domain models are known to explain physical phenomena more accurately. In addition, as we will show in this paper, the innovation-driven models allow for specifying the statistics in any transform domain. This advantage is better understood when compared to the conventional Bernoulli-Gaussian discrete model. In the latter case, any transformation of the signal involves linear combinations of Bernoulli-Gaussian random variables. In general, such distributions are found by n-fold convolution of the constituent density functions. However, in the case of innovation-driven models there is a direct way of expressing all such statistics based on the distribution of the innovation process. Particularly, it turns out that the statistics of the innovation process determine whether the process of interest has a sparse/compressible representation.

Let us assume that the structure of the process is such that it has a sparse/compressible representation in a given domain (*e.g.*, a continuous-domain wavelet transform). In order to represent a realization of the process based on a finite number of measurements, it is favorable to estimate the sparse wavelet coefficients. However, we need the probability laws of the coefficients in order to estimate them. In other words, unlike the previous scenario in which we assume the statistics of the sparse representation, we need now to *derive* them.

The knowledge about the distribution of the coefficients in the sparse representation can also be exploited in signal-recovery problems (*e.g.*, inverse problems) by devising statistical techniques such as maximum a posteriori (MAP) and minimum mean-square error (MMSE) methods. The common implementation of such statistical techniques is to reformulate them as variational problems in which we minimize a cost that involves the prior or posterior probability density functions (pdf). Thus, the shape of the pdf plays a significant role in the minimization procedure and, therefore, the recovery method.

In this paper, we maintain awareness of the analog perspective of the model while studying the sparse/compressible distributions that arise from innovation-driven models. Such models serve as common ground for the conventional continuous-domain and the modern sparsity-based models. In particular, we investigate the implications of these models in minimization problems linked with statistical recovery methods.

B. Contribution

To study innovation-driven processes, we introduce innovation processes formally. The approach towards these processes is based on observations through analysis functions rather than through conventional pointwise samples. It allows us to deduce the statistics of any linear functional of the innovation process (or observations through some arbitrary test functions). As starting point, we show in Proposition 1 that the observation of an innovation process through a rectangular window characterizes the whole process. The result can also be used to characterize innovation-driven processes, by mapping the observations onto the innovation process itself. The practical advantage is that this formulation lends itself to the derivation of statistics in any linear transform domain.

At first glance, it would seem that there is no obvious distinction between the discrete and discretized versions of innovation models. Nevertheless, we shall show in Theorems 2 and 3 that the discretized models are strictly embedded in the discrete family. The reason for this is that every probability distribution associated with linear measurements of a continuous-domain innovation model is necessarily *infinitely divisible* (id), while there is no such restriction on discrete-domain models. The key property is that infinite divisibility, which is classically associated with Lévy processes [6], is preserved by linear transformations.

To highlight the implications of the discretized model, we focus on the tail behavior of probability density functions in Theorem 5. It is well-known that, among the family of id laws, Gaussian distributions have the fastest decay. Since the degree of sparsity/compressibility of a distribution is in inverse relation with its decay rate [1], all non-Gaussian members of the id family are sparser than the Gaussians. Therefore, distributions with super-Gaussian decay (*e.g.*, distributions with finite support such as the uniform distribution) cannot be id. As we shall show in Theorem 7, there is even a gap between the Gaussian rate of decay and the rest of the id family, in the sense that the non-Gaussian id pdfs cannot decay faster than $\mathrm{e}^{-\mathcal{O}(|x|\log|x|)}$.

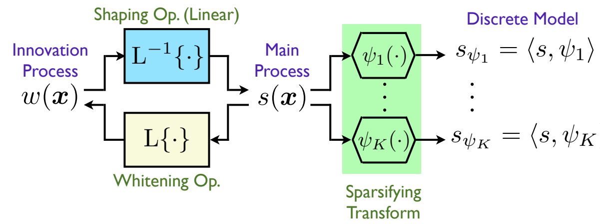

Fig. 1. Schematic of the stochastic framework.

Besides the tail behavior, we study the unimodality and moment indeterminacy of id laws in Theorems 9 and 11. We shall show that linear transformations preserve these properties along with infinite divisibility.

C. Outline

We address continuous-domain models in Section II. This includes the definition and characterization of innovation processes. In Section III, we deduce the infinite-divisibility property as a major consequence of adopting an innovation-based model. This property helps us characterize the set of admissible probability distributions. In Section IV, we cover some key properties that are shared among infinitely divisible distributions obtained from the model, such as unimodality and the state of decay of the tail.

II. Stochastic Framework

The notations in this paper are consistent with the previous works [3], [4]. We denote the continuous-domain stochastic process which models a real/complex-valued physical phenomenon by $s(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathbb{R}^d$. We assume that the process is the result of applying a linear operator on a continuous-domain innovation process. The innovation process is represented by w. The model presupposes the existence of a pair of operators L^{-1} and L with $\mathrm{L}\mathrm{L}^{-1} = \mathrm{I}$ (identity operator) that transform w to s and vice versa, respectively. The former is known as the *shaping* operator L^{-1} and the latter as the *whitening* operator L.

To discretize the continuous-domain process or to project it onto a Riesz basis (transform domain), we consider generalized sampling through sampling kernels (or dual frame) $\psi_1, \dots, \psi_K$. The constraint on the kernels is that $\phi_i = \mathrm{L}^{-1^*}\psi_i$ should have a finite L_p norm for certain values of p, where L^{-1^*} refers to the adjoint of L^{-1}. We show in Figure 1 the schematic of our stochastic model.

A. Characteristic Functionals

Since the considered processes are not necessarily Gaussian, the first and second-order statistics (such as mean and covariance) are not sufficient to fully characterize them. The alternative used here is the so-called characteristic functional. It is conceptually the same as the characteristic function but it is associated with random processes rather than random variables [7]. For an arbitrary random process z, the characteristic functional $\widehat{\mathscr{P}}_z$ is defined as

$$\widehat{\mathscr{P}}_z(\psi) = \mathbb{E}\big\{\mathrm{e}^{\mathrm{j}\langle z,\psi\rangle}\big\}, \tag{1}$$

where ψ is a suitable test function and

$$Z_\psi = \langle z, \psi\rangle = \int_{\mathbb{R}^d} z(\boldsymbol{x})\psi(\boldsymbol{x})\mathrm{d}\boldsymbol{x} \tag{2}$$

is a real/complex-valued random variable which is a linear functional of z. Note that the characteristic functional is indexed by test functions rather than scalars.

The specification of the input domain of a characteristic functional is part of its definition. For a process z, we denote the set of all valid test functions by Ξ_z. It is required that this set is a function space. This implies that linear combinations of elements in Ξ_z belong to Ξ_z. Hence, if $\psi_1,\dots,\psi_K \in \Xi_z$ and $\omega_1,\dots,\omega_K$ is a set of real variables, then $\sum_{k=1}^K \omega_k\psi_k \in \Xi_z$ and

$$\begin{aligned}\widehat{\mathscr{P}}_z\Big(\sum_{k=1}^{K}\omega_k\psi_k\Big) &= \mathbb{E}\Big\{\mathrm{e}^{\mathrm{j}\sum_{k=1}^K \omega_k Z_{\psi_k}}\Big\}\\ &= \int_{\mathbb{R}^K} p_{Z_{\psi_1}:Z_{\psi_K}}(x_1,\dots,x_K)\mathrm{e}^{\mathrm{j}\sum_{k=1}^K\omega_k x_k}\prod_{k=1}^K \mathrm{d}x_k\\ &= \mathcal{F}\Big\{p_{Z_{\psi_1}:Z_{\psi_K}}\Big\}(\omega_1,\dots,\omega_K),\end{aligned} \tag{3}$$

where $\mathcal{F}\{\cdot\}$ represents the Fourier transform operator and $p_{Z_{\psi_1}:Z_{\psi_K}}$ is the joint pdf of the linear observations of the process z: $Z_{\psi_1},\dots,Z_{\psi_K}$. Therefore, all finite-dimensional pdfs can be derived from the characteristic functional. The function spaces Ξ used in this paper are the intersection of two L_p spaces (functions with finite p-norm).

The main interest of characteristic functionals is that they provide a concise and rigorous way of defining stochastic processes. In brief, a functional $\widehat{\mathscr{P}}_z$ for which $\widehat{\mathscr{P}}_z(0) = 1$ and $\widehat{\mathscr{P}}_z\big(\sum_{k=1}^K\omega_k\psi_k\big)$ is a valid characteristic function for all $\psi_k \in \Xi_z$ corresponds to a unique random process z. More details about this fact are provided in Appendix A.

B. Innovation Process

The innovation process (a.k.a. white noise) is a random object composed of i.i.d. constituents. A discrete-domain innovation process is simply a sequence of i.i.d. random variables that is completely characterized by its pdf, which can be arbitrary. By contrast, a continuous-domain innovation process is defined by its observations through test functions.

Definition 1. *The process w is a continuous-domain innovation process if*

1) *(Stationarity) for any test function $\varphi \in \Xi_w$ and arbitrary $\boldsymbol{\tau}_1,\boldsymbol{\tau}_2 \in \mathbb{R}^d$, the two random variables $\langle w, \varphi(\cdot-\boldsymbol{\tau}_1)\rangle$ and $\langle w, \varphi(\cdot-\boldsymbol{\tau}_2)\rangle$ have identical distributions.*
2) *(Independent atoms) for test functions $\varphi_1,\varphi_2 \in \Xi_w$ with disjoint supports such that $\varphi_1(\boldsymbol{x})\varphi_2(\boldsymbol{x}) \equiv 0$, the random variables $\langle w,\varphi_1\rangle$ and $\langle w,\varphi_1\rangle$ are independent.*

The independent-atom property implies that, for $\varphi_1,\varphi_2 \in \Xi_w$ with disjoint supports, we should have that

$$\widehat{\mathscr{P}}_w(\varphi_1+\varphi_2) = \widehat{\mathscr{P}}_w(\varphi_1)\widehat{\mathscr{P}}_w(\varphi_2). \tag{4}$$

In Section III, we give a full characterization of continuous-domain innovation processes using the Gelfand-Vilenkin approach.

It is known that two Gaussian random variables are independent if and only if they are uncorrelated. Thus, for Gaussian innovation processes, it is common to express the independent-atom property based on two orthogonal test functions ($\int\varphi_1(\boldsymbol{x})\varphi_2(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 0$) without considering their supports. The characteristic functional of a Gaussian innovation process with symmetric distribution is given by [8]

$$\widehat{\mathscr{P}}_{w_g}(\varphi) = \mathrm{e}^{-\frac{1}{2}\sigma_g^2\|\varphi\|_2^2}. \tag{5}$$

This functional is well-defined for $\varphi \in L_2$. It satisfies the requirements of Definition 1 owing to the fact that $\|\varphi(\cdot-\boldsymbol{\tau})\|_2 = \|\varphi\|_2$ and $\|\varphi_1+\varphi_2\|_2^2 = \|\varphi_1\|_2^2 + \|\varphi_2\|_2^2 + 2\langle\varphi_1,\varphi_2\rangle$.

C. Linear Operators

The linear operator L in Figure 1 is the continuous-domain analog of the sparsifying matrix used in compressed sensing. Conversely, the inverse operator L^{-1} mixes the independent components of the innovation process to form specific correlation patterns. In order to formally define the application of L^{-1} on w, one might think of studying the effect of the operator on the realizations. Here, we concentrate on the characteristic functionals. The key to our study is the concept of adjoint operator which enables us to write

$$S_\psi = \langle s,\psi\rangle = \langle \mathrm{L}^{-1}w,\psi\rangle = \langle w, \mathrm{L}^{-1*}\psi\rangle = W_\phi, \tag{6}$$

where L^{-1*} is the adjoint operator of L^{-1} and $\phi = \mathrm{L}^{-1*}\psi$. Hence, the characteristic functional of the process s can be expressed as

$$\widehat{\mathscr{P}}_s(\psi) = \widehat{\mathscr{P}}_w(\mathrm{L}^{-1*}\psi) = \widehat{\mathscr{P}}_w(\phi). \tag{7}$$

This definition implies that the domain of $\widehat{\mathscr{P}}_s$ is made up of those test functions ψ for which $\phi = \mathrm{L}^{-1*}\psi \in \Xi_w$. For existence considerations and the proper interpretation of s as a generalized process over tempered distributions ($\mathcal{S}'$), it is important that the domains of both $\widehat{\mathscr{P}}_w$ and $\widehat{\mathscr{P}}_s$ include the Schwartz function space $\mathcal{S}$ (see Appendix A). This constrains L^{-1*} to form a mapping from the Schwartz space to a subset of Ξ_w. In this paper we assume that the operator L^{-1} satisfies the required admissibility properties. More details regarding the specification of suitable inverse operators can be found in [3], [9].

It is worth mentioning that L needs not be uniquely invertible to have an acceptable shaping operator L^{-1}. The real requirement is that L has a finite-dimensional null space and L^{-1} is a right inverse of L that continuously maps $\mathcal{S}$ into Ξ_w [9]. Ordinary differential operators with constant coefficients are among examples of L that admit suitable right-inverses [10], [11]. For such operators, L^{-1} is an integral operator. The formalism also extends to the cases where the underlying system is unstable, which requires the imposition of suitable linear boundary conditions in order to enforce unicity. Linearity is of fundamental importance to the formulation and is embedded in the definition of s via the use of the adjoint operator L^{-1*}.

D. Discretization

We use the term *discretization* for observations of the continuous-time process s through test functions (sampling kernels). We now explain two objectives of discretization.

1) Let us consider the expansion of the realizations of s in a Riesz basis, namely, $\{\tilde{\psi}_i\}$. The coefficients in the expansion are found as the inner products of s with the dual basis $\{\psi_i\}$ as in
$$s_{\tilde{\psi}} = \sum_i \langle s, \psi_i \rangle \tilde{\psi}_i. \quad (8)$$
This suggests that the statistics of s would be encapsulated in the coefficients $\langle s, \psi_i \rangle$. In this scenario, the task of finding the coefficients and their statistics in a transform domain is referred to as discretization.
2) We are practically limited to sense physical phenomena through a finite number of measurements. The stochastic models that describe such phenomena are usually characterized by a set of parameters, and the measurements can be used to estimate these parameters. For instance, one might think of the optimal set of parameters as the one that best explains the statistics of the measurements. In many applications, the measurements are point samples or, more generally, linear samples of the physical phenomena by means of sampling kernels $\psi_1, \dots, \psi_K$. In this context, the discretization procedure translates into linearly measuring the process. It might also involve an additive noise term.

In spite of different objectives, all discretizations are centered on the random variables $s_{\psi_i} = \langle s, \psi_i \rangle$. The definition of s restricts the kernels ψ_i to satisfy $\phi_i = \mathrm{L}^{-1*}\psi_i \in \Xi_w$. This guarantees the inclusion of the random variables s_{ψ_i} in the established framework by the way of the characteristic functional.

Our results in this paper do not depend on the choice of sampling kernels used in discretization, as long as they are admissible. However, certain kernels are preferable for the purpose of sparse representation. For the sake of simplicity, let us consider a hypothetical setting in which the s_{ψ_i} are i.i.d. If the distribution is also compressible, then we are dealing with a linear transform domain with compressible i.i.d. coefficients, which is an ideal scenario for compressed sensing. In most cases, however, such a linear transformation does not exist. Instead, one may think of a linear transformation which best uncouples the coefficients. For instance, it is shown in [10] that, if L is a differential operator, then the generalized differences of uniform point samples of s have finite-length dependencies. Note that the generalized differences are linear functionals of the process and can be written in the form $\langle s, \psi_i \rangle$. In turn, the functions $\phi_i = \mathrm{L}^{-1*}\psi_i$ are exponential splines associated with the differential operator L and are of finite support [12]. The relaxation of the i.i.d. property to finite-length dependencies is still useful because the sequence can be written as the union of a finite (but more than one) number of i.i.d. subsequences.

III. Infinite divisibility

The main property of the measurements that we are going to explore is the infinite divisibility stated in Definition 2.

Definition 2. *A random variable X (or its distribution) is said to be infinitely divisible if, for all positive integers n, we can write X as the sum of n independent and identically distributed (i.i.d.) random variables.*

It is easy to check that the sum of n independent Gaussian random variables with mean $\frac{\mu}{n}$ and variance $\frac{\sigma^2}{n}$ is a Gaussian random variable with mean μ and variance σ^2. Thus, all Gaussian random variables are infinitely divisible. The same argument can be extended to other stable distributions. Nevertheless, the stable distributions are only a small part of the id family. The complete family is characterized by the celebrated Lévy-Khinchine representation theorem in Section III-B.

In the following, we first show that the discretizations of an innovation process through rectangular test functions are infinitely divisible. Then, we characterize all id laws. This in turn characterizes the innovation processes. The final result of this section is that infinite divisibility is a general property that is shared by all linear measurements and is not restricted to rectangular test functions.

A. Observations with Rectangular Windows

To measure a process through a given test function, we should first make sure that the test function belongs to the associated function space. For the innovation process w, this requires the knowledge of the space Ξ_w. Although we postpone the exact identification of Ξ_w to Section III-D, we can already use the result that Ξ_w is the intersection of some L_p spaces, which certainly contains the intersection of all L_p spaces.

We first examine the unit rectangular test function $\mathrm{rect}(\boldsymbol{x})$ that takes the value 1 for $\boldsymbol{x} \in [0,1[^d$ and 0 otherwise. Since this function is bounded and has finite support, it belongs to $\bigcap_{p\geq 0} L_p(\mathbb{R}^d)$.

Lemma 1. *The random variable $\langle w, \mathrm{rect}\rangle$, where w is an innovation process as specified in Definition 1, is infinitely divisible.*

Proof. A distinguishing property of the rectangular function is its refinablity given by
$$\mathrm{rect}(\boldsymbol{x}) = \sum_{i=0}^{n-1} \mathrm{rect}(nx_1 - i,\, x_2,\, \cdots,\, x_d), \quad (9)$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)$ and n is any positive integer. This implies that

$$\begin{aligned} X &= \langle w, \mathrm{rect}\rangle \\ &= \sum_{i=0}^{n-1} \underbrace{\langle w(\boldsymbol{x})\,,\, \mathrm{rect}(nx_1 - i,\, x_2,\, \cdots,\, x_d)\rangle}_{X_i}. \end{aligned} \tag{10}$$

Note that the test functions $\mathrm{rect}(nx_1 - i,\, x_2,\, \ldots,\, x_d)$ for $i = 0, \ldots, (n-1)$ differ only by the shift parameter i. Moreover, they have disjoint supports. Hence, due to the stationarity and independent-atom property of the innovation process w, the random variables X_i are i.i.d. Consequently, for arbitrary n, we have a representation of X as the sum of n i.i.d. random variables, thus X is infinitely divisible. $\square$

B. Characterization of id Distributions

The concept of infinite divisibility was introduced and studied in the late 1920's and early 1930's by Finetti, Kolmogorov, and Lévy. The complete characterization of id distributions is given by the Lévy-Khinchine representation theorem.

Theorem 1 (Lévy-Khinchine [6]). *The random variable X is infinitely divisible if and only if its characteristic function has the form $\hat{p}_X(\omega) = \exp\big(f(\omega)\big)$ with*

$$\begin{aligned} f(\omega) =& \mathrm{j}\theta\omega - \frac{\sigma^2}{2}\omega^2 \\ &+ \int_{\mathbb{R}\setminus\{0\}} \big(\mathrm{e}^{\mathrm{j}a\omega} - 1 - \mathrm{j}a\omega \mathbb{1}_{|a|<1}(a)\big)\, \mathrm{d}V(a), \end{aligned} \tag{11}$$

where $\mathbb{1}_{h(a)<1}(a) = 1$ for $\{a \,|\, h(a) < 1\}$ and 0 otherwise, θ, σ are constants, and V (the Lévy measure) is a positive measure that satisfies

$$\int_{\mathbb{R}\setminus\{0\}} \min(1, a^2)\mathrm{d}V(a) < \infty. \tag{12}$$

The function f in (11) is usually referred to as the *Lévy exponent*. Its finiteness implies that the characteristic function of an id random variable does not vanish. It is interesting to point out how three important properties of characteristic functions impact the Lévy exponent.

1) Normalization: $\hat{p}_X(0) = \int_{\mathbb{R}} p_X(x)\mathrm{d}x = 1$. This implies that $f(0) = 0$, which is consistent with (11).
2) Since the characteristic function is the Fourier transform of a nonnegative distribution, we have that $|\hat{p}_X(\omega)| \leq \hat{p}_X(0)$. This is equivalent to $\Re\{f(\omega)\} \leq 0$, with equality at $\omega = 0$.
3) Continuity: $\hat{p}$ is the Fourier transform of a non-negative integrable distribution. Thus, it is continuous. This translates into $f = \log \hat{p}$ being continuous as well.

Theorem 1 indicates that id distributions are uniquely characterized by the triplet (θ, σ, V). For instance, a Gaussian distribution with mean μ_g and variance σ_g^2 corresponds to the triplet $(\mu_g, \sigma_g, V \equiv 0)$. In fact, the term associated with the constant σ is usually regarded as the Gaussian term; this becomes even more evident in the Lévy-Itō decomposition of Lévy processes.

When the Lévy measure is symmetric, with $V(I) = V(-I)$ for all measurable sets I, then the Lévy exponent admits the simplified form

$$f(\omega) = \mathrm{j}\theta\omega - \frac{\sigma^2}{2}\omega^2 - \int_{\mathbb{R}\setminus\{0\}} \big(1 - \cos(a\omega)\big)\mathrm{d}V(a). \tag{13}$$

By Theorem 1 and Lemma 1, it follows that the characteristic function of $\langle w, \mathrm{rect}\rangle$ takes the generic form

$$\hat{p}_{\langle w, \mathrm{rect}\rangle}(\omega) = \widehat{\mathscr{P}}_w(\omega\, \mathrm{rect}) = \mathrm{e}^{f(\omega)}, \tag{14}$$

where f is a valid Lévy exponent.

C. Discretizations with general Windows

So far, we have considered rectangular windows. Next, we study the implications of rectangular windows on more general test functions.

By employing the independent-atom and stationarity properties of the innovation process and the refinement formula

$$\mathrm{rect}(\boldsymbol{x}) = \sum_{\boldsymbol{k}\in\{0,\ldots,n-1\}^d} \mathrm{rect}(n\boldsymbol{x} - \boldsymbol{k}), \tag{15}$$

we conclude that

$$\widehat{\mathscr{P}}_w\big(\omega\, \mathrm{rect}(n\boldsymbol{x} - \boldsymbol{k})\big) = \big(\widehat{\mathscr{P}}_w(\omega\, \mathrm{rect})\big)^{\frac{1}{n^d}} = \mathrm{e}^{\frac{f(\omega)}{n^d}}. \tag{16}$$

This allows us to further identify the value of $\widehat{\mathscr{P}}_w$ for the general class of piecewise-constant functions of the form $\varphi = \sum a_{\boldsymbol{k}} \mathrm{rect}(n\boldsymbol{x} - \boldsymbol{k})$ such as

$$\begin{aligned} \widehat{\mathscr{P}}_w\Big(\omega \sum_{\boldsymbol{k}\in\mathcal{A}_K} a_{\boldsymbol{k}}\, \mathrm{rect}(n\boldsymbol{x} - \boldsymbol{k})\Big) &= \prod_{\boldsymbol{k}\in\mathcal{A}_K} \widehat{\mathscr{P}}_w\big(\omega a_{\boldsymbol{k}}\, \mathrm{rect}(n\boldsymbol{x} - \boldsymbol{k})\big) \\ &= \mathrm{e}^{\frac{1}{n^d}\sum_{\boldsymbol{k}\in\mathcal{A}_K} f(\omega a_{\boldsymbol{k}})}, \end{aligned} \tag{17}$$

where $\mathcal{A}_K = \{-K, \ldots, K\}^d$ and $a_{\boldsymbol{k}} \in \mathbb{R}$ are arbitrary coefficients. In other words, the characterization of $\langle w, \mathrm{rect}\rangle$ results in the identification of $\widehat{\mathscr{P}}_w$ over the set of d-dimensional piecewise-constant signals of finite support with corners at rational grid points.

These step functions can also be used to approximate other test functions; by increasing n and K, we make the step functions finer and wider in support, respectively.

Proposition 1. *For a given test function $\varphi \in \Xi_w$, where $\Xi_w = L_{p_1} \cap L_{p_2}$ and $|\varphi|^{p_i}$ is Riemann-integrable, we have that*

$$\forall\, \omega : \quad \widehat{\mathscr{P}}_w(\omega\varphi) = \exp\Big(\int_{\mathbb{R}^d} f\big(\omega\varphi(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau}\Big). \tag{18}$$

Proof. The key idea is that, due to Riemann integrability of $|\varphi|^{p_i}$, it is possible to find a sequence of step functions $\{\varphi_n\}_{n\in\mathbb{N}}$ such that $|\varphi_n| \leq |\varphi|$ and $\lim_{n\to\infty} \varphi_n = \varphi$. For each realization of w like w_r we have that

$$\langle w_r, \varphi\rangle = \lim_{n\to\infty} \langle w_r, \varphi_n\rangle. \tag{19}$$

Therefore, the random variables $W_{\varphi_n} = \langle w, \varphi_n\rangle$ converge to the random variable $W_\varphi = \langle w, \varphi\rangle$, almost surely. According to

Lévy's continuity theorem, a similar convergence result holds for the characteristic functions

$$\begin{aligned}\hat{p}_{W_\varphi}(\omega) &= \mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega W_\varphi}\} = \lim_{n\to\infty} \mathbb{E}\{\mathrm{e}^{\mathrm{j}\omega W_{\varphi_n}}\} \\ &= \lim_{n\to\infty} \hat{p}_{W_{\varphi_n}}(\omega). \end{aligned} \tag{20}$$

Note that $\hat{p}_{W_\phi}(\omega) = \widehat{\mathscr{P}}_w(\omega\phi)$ and, for step functions φ_n, we already know the validity of (18) from (17), so that

$$\widehat{\mathscr{P}}_w(\omega\varphi_n) = \exp\Big(\int_{\mathbb{R}^d} f\big(\omega\varphi_n(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau}\Big). \tag{21}$$

Hence,

$$\begin{aligned}\widehat{\mathscr{P}}_w(\omega\varphi) &= \lim_{n\to\infty} \exp\Big(\int_{\mathbb{R}^d} f\big(\omega\varphi_n(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau}\Big) \\ &= \exp\Big(\lim_{n\to\infty}\int_{\mathbb{R}^d} f\big(\omega\varphi_n(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau}\Big). \end{aligned} \tag{22}$$

The proof is completed by

$$\lim_{n\to\infty}\int_{\mathbb{R}^d} f\big(\omega\varphi_n(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau} = \int_{\mathbb{R}^d} f\big(\omega \underbrace{\lim_{n\to\infty}\varphi_n(\boldsymbol{\tau})}_{\varphi(\boldsymbol{\tau})}\big)\mathrm{d}\boldsymbol{\tau}, \tag{23}$$

where we invoke the continuity of f and Lebesgue's dominated convergence theorem to justify the interchange of limits. This requires the upperbound of Lemma 3 (Section III-D) on $|f(\omega)|$ which implies an upperbound on $\int_{\mathbb{R}^d} f\big(\omega\varphi_n(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau}$ in terms of some L_p norms of φ_n, and consequently, of φ. □

One can check that (18) is consistent with previous assumptions regarding the rectangular window. In particular, the characteristic function of $X = \langle w, \mathrm{rect}\rangle$ predicted by (18) matches $\hat{p}_X(\omega) = \exp\big(f(\omega)\big)$. Moreover, by setting $\omega = 1$ in (18), we can interpret the result of Proposition 1 in terms of the characteristic functional by

$$\widehat{\mathscr{P}}_w(\varphi) = \exp\Big(\int_{\mathbb{R}^d} f\big(\varphi(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau}\Big). \tag{24}$$

This form is multiplicative for disjointly supported test functions φ_1, φ_2, which guarantees the independent-atom property of the process. Conversely, Gelfand and Vilenkin proved in [8] that (24) is a valid characteristic functional over the space of smooth and compactly supported functions if and only if f is a valid Lévy exponent. In this work, we shall investigate the extent to which we can expand the class of test functions.

D. Characteristic Functional over Ξ_w

Bydefining a characteristic functional over some function space Ξ, we imply that the probability measure of the process is supported on the dual of Ξ (Appendix A). To highlight this point, let $\hat{\Xi}$ be a strict subspace of Ξ. The definition of the characteristic functional over Ξ induces a definition over $\hat{\Xi}$. The latter definition results in an extension of the probability space to the algebraic dual of $\hat{\Xi}$, typically via the inclusion of new sets with probability measure zero. Therefore, it is desirable to base the definition of the characteristic functional on the largest-possible space, so as to maximally constrain the support of the probability measure.

Definition 3. *The Lévy measure V is said to be (p_1, p_2)-bounded for $0 \le p_1 \le p_2 \le 2$ if*

$$\int_{\mathbb{R}\setminus\{0\}} \min(|a|^{p_1}, |a|^{p_2})\mathrm{d}V(a) < \infty. \tag{25}$$

The concept of (p_1, p_2)-boundedness is to refine the $(0, 2)$-boundedness imposed by (12) in order to better represent the properties of a given Lévy measure. As Lemma 2 indicates, a (p_1, p_2)-bounded measure is automatically $(0, 2)$-bounded.

Lemma 2. *If $0 \le q_1 \le p_1 \le p_2 \le q_2 \le 2$, then, (p_1, p_2)-boundedness of a measure V implies its (q_1, q_2)-boundedness.*

Proof. By separately studying the cases of $|a| \le 1$ and $|a| > 1$, we can check that

$$\min(|a|^{q_1}, |a|^{q_2}) \le \min(|a|^{p_1}, |a|^{p_2}). \tag{26}$$

This yields

$$\begin{aligned}&\int_{\mathbb{R}\setminus\{0\}} \min(|a|^{p_1}, |a|^{p_2})\mathrm{d}V(a) \\ \le &\int_{\mathbb{R}\setminus\{0\}} \min(|a|^{q_1}, |a|^{q_2})\mathrm{d}V(a) < \infty. \end{aligned} \tag{27}$$

□

The particular instance of Lemma 2 for $q_1 = 0$ and $q_2 = 2$ suggests that (p_1, p_2)-boundedness is a more restrictive property than the classical constraint (12).

In order to specify Ξ_w, we need to take into account the properties of the Lévy triplet (θ, σ, V). The concept of (p_1, p_2)-boundedness describes some properties of the Lévy measure V, such as the decay of its tail. We further refine this concept in Definition 4 by including the two other elements of the triplet.

Definition 4. *We say that the pair $(p_{\min}, p_{\max})$, where $0 \le p_{\min} \le p_{\max} \le 2$, bounds the Lévy triplet (θ, σ, V) if*

1) *V is $(p_{\min}, p_{\max})$-bounded (see Definition 3),*
2) *$1 \in [p_{\min}, p_{\max}]$ in case $\theta \ne 0$ or V is asymmetric (no constraint when $\theta = 0$ and V is symmetric), and*
3) *$p_{\max} = 2$ for $\sigma \ne 0$ (no constraint for $\sigma = 0$).*

Similar to (p_1, p_2)-boundedness, it is easy to check that $(0, 2)$ bounds all Lévy triplets. Furthermore, if $(p_{\min}, p_{\max})$ bounds a given triplet, then all pairs of $(q_{\min}, q_{\max})$ such that $0 \le q_{\min} \le p_{\min} \le p_{\max} \le q_{\max} \le 2$ bound the triplet as well.

The significance of Definition 4 is in identifying the L_p spaces whose intersection results in a valid function space Ξ_w for the domain of the characteristic functional. We show in Theorem 2 that, if $(p_{\min}, p_{\max})$ bounds the Lévy triplet, then

$$\Xi_w = L_{p_{\min}}(\mathbb{R}^d) \cap L_{p_{\max}}(\mathbb{R}^d) \tag{28}$$

is a suitable function space as the input domain of the characteristic functional of the innovation process. By convention, the limit case L_0 denotes the space of bounded and compactly supported functions.

For $0 \le q_{\min} \le p_{\min} \le p_{\max} \le q_{\max} \le 2$, we have that $L_{q_{\min}} \cap L_{q_{\max}} \subseteq L_{p_{\min}} \cap L_{p_{\max}}$. Thus, the tighter bounding pair $(p_{\min}, p_{\max})$ on the Lévy triplet allows for a larger function space Ξ_w. Concretely, this means that by finding tight bounding pairs for the Lévy triplet we get a better description of the properties of the innovation process.

In Theorem 2, we prove that the function space Ξ_w defined in (28) is a suitable domain for the characteristic functional of (24). Lemma 3, whose proof is given in Appendix B, is our main tool for establishing this fact.

Lemma 3. *Let V be a (p_1, p_2)-bounded Lévy measure and define*

$$g(\omega) = \int_{\mathbb{R}\setminus\{0\}} \big(\mathrm{e}^{\mathrm{j}a\omega} - 1 - \mathrm{j}a\omega \mathbb{1}_{|a|<1}(a)\big)\,\mathrm{d}V(a). \tag{29}$$

We have that

$$\big|g(\omega)\big| \leq \kappa_1|\omega|^m + \kappa_2|\omega|^M, \tag{30}$$

where κ_1 and κ_2 are nonnegative constants. Here, $(m, M) = (p_1, p_2)$ if V is symmetric, and $m = \min(1, p_1)$ and $M = \max(1, p_2)$ otherwise.

Theorem 2. *Let f be a Lévy exponent characterized by the triplet (b, σ, V) and let $(p_{\min}, p_{\max})$ be a pair that bounds the triplet. Then, the characteristic functional*

$$\widehat{\mathscr{P}}_w(\varphi) = \exp\Big(\int_{\mathbb{R}^d} f\big(\varphi(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau}\Big)$$

is finite (well-defined) over $\Xi_w = L_{p_{\min}}(\mathbb{R}^d) \cap L_{p_{\max}}(\mathbb{R}^d)$.

Proof. Using the Lévy-Khintchine representation (11), we rewrite the exponent of the characteristic functional as

$$\int_{\mathbb{R}^d} f\big(\varphi(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau} = \underbrace{\mathrm{j}\,\theta\int_{\mathbb{R}^d}\varphi(\boldsymbol{\tau})\mathrm{d}\boldsymbol{\tau}}_{T_1} - \underbrace{\frac{\sigma^2}{2}\int_{\mathbb{R}^d}\varphi^2(\boldsymbol{\tau})\mathrm{d}\boldsymbol{\tau}}_{T_2} + \underbrace{\int_{\mathbb{R}^d} g\big(\varphi(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau}}_{T_3}. \tag{31}$$

We establish finiteness for each of the terms contributing in (31).

- If $\theta = 0$, then $T_1 = 0$. For $\theta \neq 0$, Definition 4 implies that $1 \in [p_{\min}, p_{\max}]$ or $\varphi \in L_1(\mathbb{R}^d)$. The inequality $|T_1| \leq |\theta| \cdot \|\varphi\|_1$ confirms that T_1 is finite.
- Similarly, $\sigma = 0$ yields $T_2 = 0$; thus, we assume $\sigma \neq 0$. Under this assumption, Definition 4 necessitates that $p_{\max} = 2$ or $\varphi \in L_2(\mathbb{R}^d)$. The finiteness of T_2 is obtained by the inequality $|T_2| \leq \frac{\sigma^2}{2}\|\varphi\|_2^2$.
- We prove the finiteness of T_3 by applying Lemma 3. Note that the pair $(p_{\min}, p_{\max})$ also satisfies the requirements of Lemma 3. This provides us with $\varphi \in L_{p_{\min}}(\mathbb{R}^d) \cap L_{p_{\max}}(\mathbb{R}^d)$ and

$$|T_3| \leq \kappa_1\|\varphi\|_{p_{\min}}^{p_{\min}} + \kappa_2\|\varphi\|_{p_{\max}}^{p_{\max}}. \tag{32}$$

□

E. Infinite Divisibility of All Discretizations

Our last contribution in this section is to show that all the measurements $W_\varphi = \langle w, \varphi\rangle$ are infinitely divisible.

Theorem 3. *Let (θ, σ, V) be a Lévy triplet representing the Lévy exponent f and let Ξ_w and w be the corresponding function space and innovation process as defined in Theorem 2, respectively. For a given $\varphi \in \Xi_w$, define μ_φ to be the measure describing the amplitude distribution of φ. Then, the random variable $X_\varphi = \langle w, \varphi\rangle$ is infinitely divisible with the Lévy exponent*

$$f_\varphi(\omega) = \int_{\mathbb{R}^d} f\big(\omega\varphi(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau}, \tag{33}$$

which can be represented by the triplet $(\theta_\varphi, \sigma_\varphi, V_\varphi)$, where

$$\begin{cases} \sigma_\varphi = \sigma\|\varphi\|_2 & (0, \text{ if } \sigma = 0), \\ V_\varphi(I) = \int_{a\tau\in I}\mathrm{d}V(a)\mathrm{d}\mu_\varphi(\tau) & (0 \notin I \subset \mathbb{R}) \,. \end{cases} \tag{34}$$

Furthermore, except for $\varphi \equiv 0$, V_φ is (p_1, p_2)-bounded if and only if V is (p_1, p_2)-bounded.

To facilitate reading of the paper, the proof is postponed to Appendix C. The main message in Theorem 3 is that all linear observations of an innovation-driven process (subject to the admissibility condition $\varphi = \mathrm{L}^{-1*}\psi \in \Xi_w$) are infinitely divisible with roughly similar Lévy measures.

IV. Properties of Infinitely Divisible Distributions

In this section, we study properties of the id family such as decay and unimodality of the probability density functions. We also investigate their consequence on transform-domain statistics. Our approach is based on expressing various properties of the pdf in terms of the associated Lévy measure. As discussed in Section III, certain high-level properties of the Lévy measures are shared among different linear measurements of an innovation-driven process. The links between the Lévy measures and pdfs help us in establishing the implications that this has on the probability laws.

Remark 1. *Let X be an infinitely divisible random variable with Lévy triplet (θ, σ, V), and N be a Gaussian random variable independent of X with mean μ_g and variance σ_g^2. Then, the random variable $X + N$ is also infinitely divisible with the Lévy triplet $\big(\theta+\mu_g, \sqrt{\sigma^2 + \sigma_g^2}, V\big)$. This can be easily verified by stating the independence of X and N in the form $\hat{p}_{X+N}(\omega) = \hat{p}_X(\omega)\hat{p}_N(\omega)$. The main consequence is that the existence of an additive Gaussian noise does not change those properties of X that are related to its Lévy measure.*

A. Decay Rate

The id property is typically associated with slowly decaying pdfs. More specifically, it will be proved that Gaussian laws have the fastest rate of decay among id distributions. Thus, all distributions with super-Gaussian decay are necessarily non-id. However, a sub-Gaussian decay does not necessarily imply infinite divisibility. As we shall demonstrate, there is a decay-gap between the Gaussians and the rest of the id family.

We start our investigation by recalling a standard result in the theory of id laws.

Theorem 4 (25.3 in [6]). *Let V be the Lévy measure of an infinitely divisible random variable X. Then, for all locally bounded functions $g : \mathbb{R} \mapsto \mathbb{R}$ such that*

$$g(x+y) \leq g(x)g(y), \quad \forall\, x, y \in \mathbb{R}, \tag{35}$$

we have that

$$\mathbb{E}_X\{g(x)\} < \infty \iff \int_{|a|\geq 1} g(a)\mathrm{d}V(a) < \infty. \tag{36}$$

The functions g satisfying (35) are called submultiplicative. They include all functions of the form $g(x) = (1+|x|)^{\delta_1}\big(1+\ln(1+|x|)\big)^{\delta_2} e^{\delta_3 |x|^{\delta_4}}$, where $\delta_{1,2,3} \geq 0$ and $0 < \delta_4 \leq 1$. In particular, Theorem 4 can be applied to investigate the existence of moments.

Lemma 4. *Let X be an id random variable with Lévy measure V. Then, for all $p \geq 0$,*

$$\mathbb{E}\{|X|^p\} < \infty \iff \int_{|a|\geq 1} |a|^p \mathrm{d}V(a) < \infty. \tag{37}$$

Proof. The function $g(x) = |x|^p$ does not satisfy the requirement of Theorem 4. Therefore, we continue with

$$\begin{aligned}\mathbb{E}\{|X|^p\} < \infty &\iff \mathbb{E}\{1+|X|^p\} < \infty\\ &\iff \mathbb{E}\{(1+|X|)^p\} < \infty,\end{aligned} \tag{38}$$

where we used the inequalities

$$1+|x|^p \leq (1+|x|)^p \leq 2^{p-1}(1+|x|^p).$$

Now, the function $g(x) = (1+|x|)^p$ fulfills the requirement of Theorem 4. Thus,

$$\begin{aligned}\mathbb{E}\{|X|^p\} < \infty &\iff \int_{|a|\geq 1} (1+|a|)^p \mathrm{d}V(a) < \infty\\ &\iff \int_{|a|\geq 1} |a|^p \mathrm{d}V(a) < \infty.\end{aligned} \tag{39}$$

□

Theorem 5. *Let w be an innovation process and let X and X_φ be $\langle w, \mathrm{rect}\rangle$ and $\langle w, \varphi\rangle$, respectively, where $0 \neq \varphi \in \Xi_w \cap L_p \cap L_{\max(2,p)}$. Then, we have that*

$$\mathbb{E}\{|X|^p\} < \infty \iff \mathbb{E}\{|X_\varphi|^p\} < \infty. \tag{40}$$

In a nutshell, Theorem 4 and Lemma 4 imply that the pdf and the Lévy measure of an id distribution have the same rate of decay of their tails. Theorem 5 states that X and X_φ are equivalent random variables in the sense of existence of moments. The additional restriction $\varphi \in L_p \cap L_{\max(2,p)}$ is to ensure that the amplitude distribution measure μ_φ has finite pth or both pth and second-order moments. The proof of Theorem 5 is postponed to appendix D.

Our next step is to show that the linear observations of an innovation process through test functions in $\bigcap_p L_p$ all have the same fat-tail behavior.

Lemma 5. *Let Ξ_w be the domain of the characteristic functional of an innovation process w. If the distribution of $X_{\bar\varphi} = \langle w, \bar\varphi\rangle$ for a given $\bar\varphi \in \bigcap_p L_p \setminus \{0\}$ is fat-tailed with $\lim_{|x|\to\infty} |x|^p\, \mathbb{P}(|X_{\bar\varphi}| > |x|) \in]0, \infty[$ for some $p > 0$, then, the distribution of $\langle w, \varphi\rangle$ for all $\varphi \in \bigcap_p L_p \setminus \{0\}$ is fat-tailed with the same decay rate $|x|^{-p}$. Moreover, the addition of Gaussian noise to the measurement does not change the fat-tail property.*

Proof. First note that, due to Theorem 3, all the random variables $\langle w, \varphi\rangle$ are infinitely divisible. Let X, $X_{\bar\varphi}$, and X_φ denote the random variables $\langle w, \mathrm{rect}\rangle$, $\langle w, \bar\varphi\rangle$, and $\langle w, \varphi\rangle$, respectively. Then, the condition $0 < \lim_{|x|\to\infty} |x|^p\, \mathbb{P}(|X_{\bar\varphi}| > |x|) < \infty$ (fat-tail property of $X_{\bar\varphi}$) indicates that $\mathbb{E}\{|X_{\bar\varphi}|^r\}$ is finite for all $0 < r < p$ and is infinite for $r \geq p$. Recalling Theorem 5, we conclude that $\mathbb{E}\{|X|^r\}$ and, therefore, $\mathbb{E}\{|X_\varphi|^r\}$, are finite for $0 < r < p$ and infinite for $r \geq p$. Thus, X_φ is also fat-tailed with the same decay rate $|x|^{-p}$.

The effect of an additive Gaussian noise is cast in the Gaussian parameter σ of the Lévy triplet. Lemma 4 shows that the fat-tail property is solely determined by the Lévy measure. □

Compressible distributions are closely related to fat-tailed distributions [1], [2]. In fact, Lemma 5 states that the compressibility of a linear observation is a property that is inherited from the innovation process and is independent of how it is measured or expanded.

Illustration 1. *Let us consider the recovery of compressible vectors from noisy linear measurements. For this purpose, let $\mathbf{x}$ be an i.i.d. random vector with a fat-tailed distribution and let $\mathbf{y} = \mathbf{A}\mathbf{x}+\mathbf{n}$ be the measurements, where $\mathbf{A}$ is a known sensing matrix and $\mathbf{n}$ stands for a vector of white Gaussian noise with variance σ_n^2. In our framework, this problem can reflect the discretization of a continuous-domain process where $\mathbf{A}$ and $\mathbf{x}$ correspond to the discretizations of L^{-1} and the innovation process, respectively. An example of such discretization can be found in [13]. For the sake of simplicity, we focus on the MAP estimator which is known to take the form*

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \ \frac{1}{2\sigma_n^2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + J(\mathbf{x}), \tag{41}$$

where $J(\mathbf{x}) = -\log p_X(\mathbf{x}) = -\sum_i \log p_X(x_i)$. The common sparsifying penalty term used in compressed sensing is $J(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_i |x_i|$ which is obtained for $\mathbf{x}$ vectors following a Laplace distribution. Several authors have pointed out that the Laplace distribution is by no means sparse or compressible. Furthermore, the classical least-square estimator outperforms the MAP estimator under Laplace distributions [10].

For fat-tailed distributions of $\mathbf{x}$, the penalty term $J(\mathbf{x})$ is of the form $\sum_i \Psi(x_i)$ with $\Psi(x) = \mathcal{O}\big(\log|x|\big)$, which is fundamentally different from $|x|$. Nevertheless, the penalty term $\log(\cdot)$ can be regarded as an ℓ_1-ℓ_0 relaxation [14] and is useful in image recovery [15]. Moreover, for fat-tailed distributions, the MAP estimator is a biased but still fair approximation of the Bayesian (posterior mean) estimator [10].

As the rate of decay of the tail of a distribution increases (faster decay), it becomes less compressible. One of the properties of the id family is that the Gaussian distributions are the least-compressible members. In fact, Gaussian distributions are somewhat isolated members, not only because of their

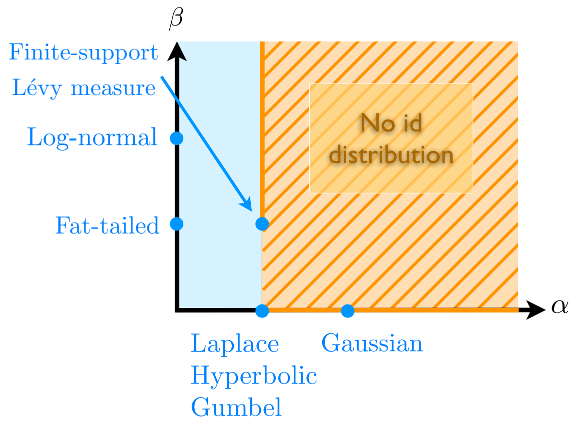

Fig. 2. Identification of id distributions with respect to their tail probabilities in the form of $\exp\big(-\mathcal{O}(|x|^{\alpha}(\log|x|)^{\beta})\big)$. Examples include $(\alpha=2,\beta=0)$ for Gaussians; $(\alpha=1,\beta=0)$ for Laplace, hyperbolic and Gumbel distributions; $(\alpha=0,\beta=1)$ for all the fat-tailed laws; $(\alpha=0,\beta=2)$ for log-normal distributions; and $(\alpha=1,\beta=1)$ for all id laws with non-zero but finitely supported Lévy measures. The only id distributions in the shaded area are Gaussians.

extreme rate of decay, but also due to a gap between their rate of decay and that of the rest of the family. Theorem 6 paves the road for specifying this gap.

Theorem 6 (26.1 in [6]). *Let X be an id random variable corresponding to a Lévy measure V. Define*

$$c=\inf\Big\{a>0:\ S_V\subseteq\{x:\ |x|\le a\}\Big\},\tag{42}$$

where S_V denotes the support set of the Lévy measure V. We also allow c to take the values 0 and ∞. Then, for the super-exponential moments, we have that

$$\begin{cases} 0<\alpha<\frac{1}{c}: & \mathbb{E}_X\Big\{\mathrm{e}^{\alpha|x|\log|x|}\Big\} < \infty,\\ \frac{1}{c}<\alpha: & \mathbb{E}_X\Big\{\mathrm{e}^{\alpha|x|\log|x|}\Big\} = \infty.\end{cases}\tag{43}$$

Theorem 7. *The only id distributions that decay faster than $\mathrm{e}^{-\mathcal{O}(|x|\log|x|)}$ are the Gaussians.*

Proof. A tail decaying faster than $\exp\big(-\mathcal{O}(|x|\log|x|)\big)$ implies that all super-exponential moments $\mathbb{E}_X\big\{\mathrm{e}^{\alpha|x|\log|x|}\big\}$ are finite. By using the result of Theorem 6, this implies that $\frac{1}{c}=\infty$, where c is defined in (42). Thus, we shall have $c=0$, which confirms that V is supported only at $\{0\}$. Besides, note that $\{0\}$ is excluded in all the integrals involving V. Hence, such a V is effectively equivalent to the zero measure. Evidently, an id distribution with zero Lévy measure is a Gaussian distribution (see Section III-B). □

Illustration 2. *Let us consider the pdfs that have a rate of decay of the form $\exp\big(-\mathcal{O}(|x|^{\kappa})\big)$. The Gaussian and Laplace distributions are id examples that correspond to $\kappa=2$ and $\kappa=1$, respectively. However, Theorem 7 states that the pdfs corresponding to $1<\kappa<2$ are not infinitely divisible (the gap). For a better understanding of this result, we revisit the MAP estimator of Illustration 1. It is well-known that, for Gaussian and Laplace distributions of $\mathbf{x}$, the penalty term $J(\mathbf{x})$ in (41) transforms into $\mathcal{O}\big(\|\mathbf{x}\|_2^2\big)$ and $\mathcal{O}\big(\|\mathbf{x}\|_1\big)$, respectively. A simple consequence of the gap in the decay of the tail of id distributions is that penalty terms of the form $\|\mathbf{x}\|_p^p$ for $1<p<2$ are not allowed. We illustrate this gap in Figure 2.*

B. Unimodality

The modes of a real-valued function are the points at which the function attains its local maxima or minima. A pdf is unimodal if it has a unique local maximum and no local minima. In words, a unimodal pdf is decreasing on the right side of its mode and increasing on its left side. Unimodality is useful in optimization problems such as MAP.

Here, we want to show that the unimodality of the pdf is a property that is inherited from the innovation process. Similar to the decay of the tail, we investigate the implications of the Lévy measure on the pdf in terms of unimodality.

Definition 5 ([16]). *A measure V is said to be unimodal with mode a_0 if it can be expressed as*

$$V(\mathrm{d}a)=c\delta_{a_0}(\mathrm{d}a)+v(a)\mathrm{d}a,\tag{44}$$

where c is a nonnegative real number, δ_{a_0} is Dirac's delta function supported at a_0, and v is an increasing function on $]-\infty,a_0[$ and decreasing on $]a_0,\infty[$.

We use Theorem 8 proved in [17] as the main tool for connecting the unimodality of the pdf to that of the Lévy measure.

Theorem 8 ([17]). *If a Lévy measure V is symmetric and unimodal with mode 0, all the id random variables identified by the Lévy triplet (θ,σ,V) have unimodal pdfs.*

Theorem 9. *Let w be an innovation process for which the random variable $\langle w, rect\rangle$ admits the Lévy triplet (θ,σ,V). If V is symmetric and unimodal with mode 0, then the pdf of $\langle w,\varphi\rangle$ for all $\varphi\in\Xi_w$ is unimodal.*

Proof. By using Theorem 8, it is sufficient to show that the Lévy measure V_φ of $\langle w,\varphi\rangle$ is also symmetric and unimodal with mode 0. To show its symmetry, we recall Theorem 3 and write $V_\varphi(-I)$ for $0\notin I\subset\mathbb{R}$ as

$$\begin{aligned}V_\varphi(-I)&=\int_{a\tau\in -I}\mathrm{d}V(a)\mathrm{d}\mu_\varphi(\tau)=\int_{a\tau\in I}\mathrm{d}V(-a)\mathrm{d}\mu_\varphi(\tau)\\ &=\int_{a\tau\in I}\mathrm{d}V(a)\mathrm{d}\mu_\varphi(\tau)=V_\varphi(I).\end{aligned}\tag{45}$$

Unimodality of V with mode 0 requires the corresponding delta term of $V(\mathrm{d}x)$ to be placed at zero. However, as pointed out earlier, zero is excluded in all the integrals over V. This fact, in conjunction with the symmetry of V, suggests that $V(\mathrm{d}x)$ can be effectively written as $v(|x|)\mathrm{d}x$ where v is a decreasing function. Hence, for $|\bar{a}|\ge|a|>0$ we can write that

$$\begin{aligned}V_\varphi(\mathrm{d}a)&=\int_{\tau\neq 0}v\Big(\frac{a}{\tau}\Big)\mathrm{d}\mu_\varphi(\tau)=\int_{\tau\neq 0}v\Big(\Big|\frac{a}{\tau}\Big|\Big)\mathrm{d}\mu_\varphi(\tau)\\ &\ge\int_{\tau\neq 0}v\Big(\Big|\frac{\bar{a}}{\tau}\Big|\Big)\mathrm{d}\mu_\varphi(\tau)=V_\varphi(\mathrm{d}\bar{a}),\end{aligned}\tag{46}$$

which proves the unimodality of V_φ. □

C. Moment Indeterminacy

The problem of moments, or Hamburger-moment problem, is to answer whether the set of moments

$$m_n^{(U)} = \int_{\mathbb{R}} a^n \mathrm{d}U(a), \quad n = 0, 1, 2, \dots$$

uniquely determines the measure U. In case the answer is negative, the measure is called moment-indeterminate or, briefly, indeterminate. There are simple necessary or sufficient conditions (namely, Krein's and Carleman's conditions, respectively) for indeterminacy of a measure, while necessary and sufficient conditions are more complicated to formulate.

If at least one of moment of a distribution is infinite, then the distribution is automatically considered as indeterminate. Theorem 5 states that $\langle w, \mathrm{rect}\rangle$ and $\langle w, \varphi\rangle$ for $\varphi \in \bigcap_p L_p$ are equivalent in the sense of existing moments. Thus, if $\langle w, \mathrm{rect}\rangle$ is indeterminate by means of having infinite moments, the same applies to all $\langle w, \varphi\rangle$. However, when all the moments are finite, Theorem 5 does not settle the issue of in/determinacy.

Theorem 10 ([18]). *An infinitely divisible distribution that corresponds to an indeterminate Lévy measure is itself indeterminate.*

Next, we show that the indeterminacy of a Lévy measure and, consequently, of its associated probability distribution, is a property that is shared by all linear measurements of an innovation process.

Theorem 11. *If the Lévy measure V of $\langle w, \mathrm{rect}\rangle$ is indeterminate, where w is an innovation process, then the distribution of $\langle w, \varphi\rangle$ for all $0 \neq \varphi \in \bigcap_p L_p$ is indeterminate.*

Proof. By applying Theorem 3, we know that

$$\begin{aligned}\int_{\mathbb{R}\setminus\{0\}} \bar{a}^n \mathrm{d}V_\varphi(\bar{a}) &= \int_{\mathbb{R}\setminus\{0\}}\int_{\mathbb{R}\setminus\{0\}} (a\tau)^n \mathrm{d}V(a)\mathrm{d}\mu_\varphi(\tau) \\ &= \underbrace{\Big(\int_{\mathbb{R}\setminus\{0\}} a^n \mathrm{d}V(a)\Big)}_{m_n^{(V)}}\underbrace{\Big(\int_{\mathbb{R}\setminus\{0\}} \tau^n \mathrm{d}\mu_\varphi(\tau)\Big)}_{m_n^{(\mu_\varphi)}}. \end{aligned} \quad (47)$$

Since $\varphi \in \bigcap_p L_p$, the moments $\{m_n^{(\mu_\varphi)}\}_{n=0}^\infty$ are all finite. This implies a one-to-one mapping between the moments $\{m_n^{(V)}\}_{n=0}^\infty$ and $\{m_n^{(V_\varphi)}\}_{n=0}^\infty$ for a given $\varphi \in \bigcap_p L_p$. In other words, $\{m_n^{(V)}\}_n$ uniquely determines V if and only if $\{m_n^{(V_\varphi)}\}_{n=0}^\infty$ uniquely determines V_φ. Hence, indeterminacy of V translates into indeterminacy of V_φ, which in turn establishes the indeterminacy of the distribution of $\langle w, \varphi\rangle$ through Theorem 10. □

V. Conclusion

We considered an innovation-driven continuous-domain model from which we obtain linear measurements. Our goal was to identify the sparse/compressible distributions that can describe the distribution of such measurements. We showed that a common property of such distributions is infinite divisibility. One of the important implications of this property is the exclusion of all distributions that decay faster than Gaussians. Furthermore, we revealed a gap between the decay rate of Gaussian distributions and other id distributions.

The Lévy-Khinchine representation theorem characterizes all infinitely divisible distributions by means of a measure known as the Lévy measure. It was already known that many properties of infinitely divisible distributions can be expressed in terms of their Lévy measure. The contribution of this paper is to show that most of the higher-level properties of pdfs (finiteness of moments, rate of decay, and unimodality) are also preserved through linear measurements. For instance, if a model generates a compressible distribution in a particular measurement scheme, the distribution of all possible measurements would be compressible. Furthermore, this compressibility can be identified a priori through the Lévy measure associated with the innovation process.

Acknowledgment

The authors would like to thank J. Ward, C. Vonesch, and P. Thévenaz for their constructive comments.

Appendix A
Random Processes via Characteristic Functionals

A given functional over the space Ξ defines a probability measure on the algebraic or topological (continuous) dual of Ξ (the set of realizations) if the functional satisfies certain conditions. The two main results are the Kolmogorov extension theorem [19] and the Bochner-Minlos theorem [20]. For suitable characteristic functionals, the Kolmogorov extension theorem demonstrates the existence of a random process supported over the algebraic dual of Ξ, while the Bochner-Minlos theorem narrows down the support to the continuous dual of Ξ, provided that the latter space is nuclear.

Theorem 12 (Bochner-Minlos [20]). *Let Ξ be a nuclear space over $\mathbb{R}$ and $C : \Xi \mapsto \mathbb{C}$ be a continuous functional. If $C(\mathbf{0}) = 1$ and C is semipositive-definite, then C is the characteristic functional of a unique random process supported on the continuous dual of Ξ denoted by Ξ'. Semipositive definiteness of C means that, for any positive integer k and for all $z_1, \dots, z_K \in \mathbb{C}$ and $\varphi_1, \dots, \varphi_K \in \Xi$, the value*

$$\sum_{k_1,k_2=1}^{K} z_{k_1}\bar{z}_{k_2} C(\varphi_{k_1} - \varphi_{k_2})$$

is real and nonnegative.

The function space Ξ of the characteristic functionals considered in this paper is the intersection of L_p spaces. Such spaces are not nuclear, thus, the Bochner-Minlos theorem does not apply. However, the Kolmogorov extension theorem [19] implies that the functional C defines a random process over the algebraic dual of Ξ represented as Ξ^*. Additionally, C gives rise to a cylinder set measure (a quasi-measure) over Ξ' (continuous dual). This quasi-measure is equivalent to a random process as long as finite-dimensional pdfs are considered. The key observation for our purpose is that the Schwartz space $\mathcal{S}$ of rapidly decreasing functions is a nuclear space which is included in all L_p spaces. In terms of duals, this translates into $\Xi^{'} \subset \mathcal{S}^{'} \subset \Xi^{*}$. The Bochner-Minlos theorem therefore guarantees that a random process over Ξ^{*} is indeed supported over $\mathcal{S}^{'}$ (the space of tempered distributions).

Appendix B
Proof of Lemma 3

We first decompose g into three components as

$$g(\omega) = \mathrm{j}\underbrace{\int_{|a|<1}\big(\sin(a\omega) - a\omega\big)\mathrm{d}V(a)}_{\mathring{g}_{\Im}(\omega)}$$
$$+\mathrm{j}\underbrace{\int_{|a|\geq 1}\sin(a\omega)\mathrm{d}V(a)}_{\breve{g}_{\Im}(\omega)} \underbrace{-2\int_{\mathbb{R}\setminus\{0\}}\sin^2\Big(\frac{a\omega}{2}\Big)\mathrm{d}V(a)}_{g_{\Re}(\omega)}. \quad (48)$$

Our next step is to upperbound each term separately. For this purpose, note that $|\sin(x)| \leq |x|$ and $|\sin(x)| \leq 1$. Hence, for all $a \in [0,1]$, we conclude that $|\sin(x)| \leq \min(1, |x|) \leq |x|^a$. In addition, since V is (m, M)-bounded, we have that

$$\int_{|a|<1}|a|^M \mathrm{d}V(a) < \infty \quad , \quad \int_{|a|\geq 1}|a|^m \mathrm{d}V(a) < \infty.$$

1) In case V is symmetric, due to the odd symmetry of the integrand, we conclude that $\mathring{g}_{\Im} \equiv 0$. For asymmetric measures, we continue as

$$\mathring{g}_{\Im}(\omega) = -\int_{|a|<1} a\omega\big(1 - \mathrm{sinc}(a\omega)\big)\mathrm{d}V(a), \quad (49)$$

where $\mathrm{sinc}(x) = \frac{\sin(x)}{x}$. We further know that $|\mathrm{sinc}(x)| \leq 1$ and $0 \leq 1 - \mathrm{sinc}(x) \leq |x|$. The latter is obtained by observing that $|x| \geq \sin^2(x/2)$, which confirms that the function $|x|^2 + \sin|x| - |x|$ is increasing with respect to $|x|$. The two inequalities for sinc lead to $|1 - \mathrm{sinc}(x)| \leq \min(2, |x|)$. Thus, we can bound $\mathring{g}_{\Im}(\omega)$ as

$$\begin{aligned} |\mathring{g}_{\Im}(\omega)| &\leq \int_{|a|<1}|a\omega|\cdot|1 - \mathrm{sinc}(a\omega)|\,\mathrm{d}V(a) \\ &\leq \int_{|a|<1}|a\omega|\min(2, |a\omega|)\mathrm{d}V(a) \\ &\leq \int_{|a|<1}|a\omega|^M \mathrm{d}V(a) \\ &= |\omega|^M\underbrace{\int_{|a|<1}|a|^M\mathrm{d}V(a)}_{<\infty}, \end{aligned} \quad (50)$$

where we used $\min(2, |x|) \leq |x|^{M-1}$, which is justified by $1 \leq M \leq 2$ for asymmetric Lévy measures.

2) Similar to $\mathring{g}_{\Im}$, we have that $\breve{g}_{\Im} \equiv 0$ for symmetric Lévy measures V. We can write that

$$\begin{aligned} |\breve{g}_{\Im}(\omega)| &\leq \int_{|a|\geq 1}|\sin(a\omega)|\,\mathrm{d}V(a) \\ &\leq \int_{|a|\geq 1}|a\omega|^m\mathrm{d}V(a) \\ &\leq |\omega|^m\underbrace{\int_{|a|\geq 1}|a|^m\mathrm{d}V(a)}_{<\infty}, \end{aligned} \quad (51)$$

where we used $0 \leq m \leq 1$ for asymmetric Lévy measures.

3) We employ a trigonometric rule to simplify $g_{\Re}$ as

$$\begin{aligned} |g_{\Re}(\omega)| &= 2\int_{\mathbb{R}\setminus\{0\}}\sin^2\Big(\frac{a\omega}{2}\Big)\mathrm{d}V(a) \\ &\leq 2\int_{\mathbb{R}\setminus\{0\}}\min\Big(1, \Big|\frac{a\omega}{2}\Big|^2\Big)\mathrm{d}V(a) \\ &\leq 2\int_{\mathbb{R}\setminus\{0\}}\min\Big(\Big|\frac{a\omega}{2}\Big|^m, \Big|\frac{a\omega}{2}\Big|^M\Big)\mathrm{d}V(a) \\ &\leq 2\Big(\frac{|\omega|^m}{2^m} + \frac{|\omega|^M}{2^M}\Big)\times \\ &\quad \underbrace{\int_{\mathbb{R}\setminus\{0\}}\min(|a|^m, |a|^M)\mathrm{d}V(a)}_{<\infty}. \end{aligned} \quad (52)$$

Finally, we combine the individual upperbounds using the triangular inequality, which completes the proof.

Appendix C
Proof of Theorem 3

Recalling (3) and Theorem 2, we obtain the characteristic function of X_φ as

$$\hat{p}_{X_\varphi}(\omega) = \widehat{\mathscr{P}}_w(\omega\varphi) = \exp\Big(\int_{\mathbb{R}^d} f\big(\omega\varphi(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau}\Big). \quad (53)$$

Similar to (31) we can write that

$$\int_{\mathbb{R}^d} f\big(\omega\varphi(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau} = \mathrm{j}\theta_{1,\varphi}\,\omega - \frac{\sigma_\varphi^2}{2}\omega^2 + g_\varphi(\omega), \quad (54)$$

where

$$\begin{cases} \theta_{1,\varphi} &= \theta\int_{\mathbb{R}^d}\varphi(\boldsymbol{\tau})\mathrm{d}\boldsymbol{\tau}, \\ \sigma_\varphi^2 &= \sigma^2\int_{\mathbb{R}^d}\varphi^2(\boldsymbol{\tau})\mathrm{d}\boldsymbol{\tau}, \\ g_\varphi(\omega) &= \int_{\mathbb{R}^d} g\big(\omega\varphi(\boldsymbol{\tau})\big)\mathrm{d}\boldsymbol{\tau}. \end{cases} \quad (55)$$

The upperbound on g imposed by Lemma 3 indicates that $g_\varphi(\omega)$ is finite for all ω. We simplify g_φ by rewriting it as

$$\begin{aligned} g_\varphi(\omega) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}\setminus\{0\}}\Big(\mathrm{e}^{\mathrm{j}a\omega\varphi(\boldsymbol{\tau})} - 1 \\ &\quad -\mathrm{j}a\omega\varphi(\boldsymbol{\tau})\mathbb{1}_{|a|<1}(a)\Big)\mathrm{d}V(a)\mathrm{d}\boldsymbol{\tau} \\ &= \bar{g}_\varphi(\omega) + \mathrm{j}\omega\theta_{2,\varphi}, \end{aligned} \quad (56)$$

where

$$\begin{aligned} \bar{g}_\varphi(\omega) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}\setminus\{0\}}\Big(\mathrm{e}^{\mathrm{j}a\omega\varphi(\boldsymbol{\tau})} - 1 \\ &\quad -\mathrm{j}a\omega\varphi(\boldsymbol{\tau})\mathbb{1}_{|a\varphi(\boldsymbol{\tau})|<1}(a)\Big)\mathrm{d}V(a)\mathrm{d}\boldsymbol{\tau} \end{aligned} \quad (57)$$

and

$$\begin{aligned} \theta_{2,\varphi} &= \int_{\mathbb{R}^d}\int_{\mathbb{R}\setminus\{0\}} a\varphi(\boldsymbol{\tau})\Big(\mathbb{1}_{|a\varphi(\boldsymbol{\tau})|<1}(a) \\ &\quad -\mathbb{1}_{|a|<1}(a)\Big)\mathrm{d}V(a)\mathrm{d}\boldsymbol{\tau}. \end{aligned} \quad (58)$$

In (57), the integration parameter $\boldsymbol{\tau}$ is used only as the input argument of φ. Consequently, $\varphi(\boldsymbol{\tau})$ can be replaced with its amplitude distribution measure μ_φ. On the other hand, whenever $\varphi(\boldsymbol{\tau}) = 0$ the integrand in (57) is also zero. Thus, those values of $\boldsymbol{\tau}$ for which $\varphi(\boldsymbol{\tau}) = 0$ do not contribute in the integral. In summary, we can rewrite (57) as

$$\begin{aligned}\bar{g}_\varphi(\omega) &= \int_{\mathbb{R}\setminus\{0\}}\int_{\mathbb{R}\setminus\{0\}} \Big(\mathrm{e}^{\mathrm{j}a\omega\bar{\tau}} - 1 \\ &\qquad - \mathrm{j}a\omega\bar{\tau}\mathbb{1}_{|a\bar{\tau}|<1}(a)\Big)\mathrm{d}V(a)\mathrm{d}\mu_\varphi(\bar{\tau}) \\ &= \int_{\mathbb{R}\setminus\{0\}} \left(\mathrm{e}^{\mathrm{j}\bar{a}\omega} - 1 - \mathrm{j}\bar{a}\omega\mathbb{1}_{|\bar{a}|<1}(\bar{a})\right)\mathrm{d}V_\varphi(\bar{a}), \quad (59)\end{aligned}$$

where we used the change of variables $\bar{a} = a\bar{\tau}$. Equations (54)-(59) suggest $(\theta_{1,\varphi} + \theta_{2,\varphi}, \sigma_\varphi, V_\varphi)$ as the Lévy triplet of X_φ, provided that $\theta_{2,\varphi}$ is finite and V_φ satisfies the requirement (12). Note that the $(0, 2)$-boundedness of V_φ (Requirement (12)) implies the finiteness of $\bar{g}_\varphi$ through Lemma 3. This establishes the finiteness of $\theta_{2,\varphi}$, since the finiteness of g_φ is guaranteed by Theorem 2. Thus, to prove that X_φ is infinitely divisible with the suggested Lévy triplet, it suffices to show that V_φ is $(0, 2)$-bounded. Instead, we prove a stronger statement: if V is (p_1, p_2)-bounded, then V_φ is also (p_1, p_2)-bounded. Specifically,

$$\begin{aligned}&\int_{\mathbb{R}\setminus\{0\}} \min(|\bar{a}|^{p_1}, |\bar{a}|^{p_2})\mathrm{d}V_\varphi(\bar{a}) \\ &= \int_{\mathbb{R}\setminus\{0\}}\int_{\mathbb{R}\setminus\{0\}} \min(|a\bar{\tau}|^{p_1}, |a\bar{\tau}|^{p_2})\mathrm{d}V(a)\mathrm{d}\mu_\varphi(\bar{\tau}) \\ &\leq \int_{\mathbb{R}\setminus\{0\}} (|\bar{\tau}|^{p_1} + |\bar{\tau}|^{p_2})\, \min(|a|^{p_1}, |a|^{p_2})\mathrm{d}V(a)\mathrm{d}\mu_\varphi(\bar{\tau}) \\ &= \left(\|\varphi\|_{p_1}^{p_1} + \|\varphi\|_{p_2}^{p_2}\right) \underbrace{\int_{\mathbb{R}\setminus\{0\}} \min(|a|^{p_1}, |a|^{p_2})\mathrm{d}V(a)}_{<\infty}. \quad (60)\end{aligned}$$

To complete the proof of Theorem 3, we establish the converse statement: if V_φ is (p_1, p_2)-bounded and $\varphi \in \Xi \setminus \{0\}$, then V is also (p_1, p_2)-bounded, since

$$\begin{aligned}&\int_{\mathbb{R}\setminus\{0\}} \min(|a|^{p_1}, |a|^{p_2})\mathrm{d}V(a) \\ &\leq \frac{\int_{\mathbb{R}\setminus\{0\}}\int_{\mathbb{R}\setminus\{0\}} \min(|a\bar{\tau}|^{p_1}, |a\bar{\tau}|^{p_2})\mathrm{d}V(a)\mathrm{d}\mu_\varphi(\bar{\tau})}{\int_{\mathbb{R}\setminus\{0\}} \min(|\bar{\tau}|^{p_1}, |\bar{\tau}|^{p_2})\mathrm{d}\mu_\varphi(\bar{\tau})} \\ &= \frac{\int_{\mathbb{R}\setminus\{0\}} \min(|\bar{a}|^{p_1}, |\bar{a}|^{p_2})\mathrm{d}V_\varphi(\bar{a})}{\int_{\mathbb{R}\setminus\{0\}} \min(|\bar{\tau}|^{p_1}, |\bar{\tau}|^{p_2})\mathrm{d}\mu_\varphi(\bar{\tau})}. \quad (61)\end{aligned}$$

The numerator in (61) is finite since V_φ is assumed to be (p_1, p_2)-bounded. The integrand in the denominator is also positive and the integral is nonzero because of $\varphi \not\equiv 0$. The boundedness of the denominator is readily confirmed by

$$\int_{\mathbb{R}\setminus\{0\}} \min(|\bar{\tau}|^{p_1}, |\bar{\tau}|^{p_2})\mathrm{d}\mu_\varphi(\bar{\tau}) \leq \min(\|\varphi\|_{p_1}^{p_1}, \|\varphi\|_{p_2}^{p_2}). \quad (62)$$

Appendix D
Proof of Theorem 5

According to Theorem 3, for all $\varphi \in \Xi_w$ the random variable $\langle w, \varphi\rangle$ is infinitely divisible. Let V and V_φ denote the Lévy measures of $\langle w, \mathrm{rect}\rangle$ and $\langle w, \varphi\rangle$, respectively. By using Lemma 4, we reformulate the claim in Theorem 5 as

$$\int_{|a|\geq 1} |a|^p\mathrm{d}V(a) < \infty \iff \int_{|\bar{a}|\geq 1} |\bar{a}|^p\mathrm{d}V_\varphi(\bar{a}) < \infty. \quad (63)$$

We apply Theorem 3 to rewrite the integral against the measure V_φ in the form

$$\begin{aligned}\int_{|\bar{a}|\geq 1} |\bar{a}|^p\mathrm{d}V_\varphi(\bar{a}) &= \int_{|a\tau|\geq 1} |a\tau|^p\mathrm{d}V(a)\mathrm{d}\mu_\varphi(\tau) \\ &= \int_{\mathbb{R}\setminus\{0\}} |\tau|^p\bigg(\int_{|a|>\frac{1}{|\tau|}} |a|^p\mathrm{d}V(a)\bigg)\mathrm{d}\mu_\varphi(\tau). \quad (64)\end{aligned}$$

This yields

$$\begin{aligned}\int_{|\bar{a}|\geq 1} |\bar{a}|^p\mathrm{d}V_\varphi(\bar{a}) &\leq \int_{\mathbb{R}\setminus\{0\}} |\tau|^p\mathrm{d}\mu_\varphi(\tau)\cdot\int_{|a|\geq 1} |a|^p\mathrm{d}V(a) \\ &\quad + \int_{|\tau|\geq 1} |\tau|^p\int_{\frac{1}{|\tau|}\leq|a|\leq 1} |a|^p\mathrm{d}V(a)\mathrm{d}\mu_\varphi(\tau). \quad (65)\end{aligned}$$

Note that $\int_{\mathbb{R}\setminus\{0\}} |\tau|^p\mathrm{d}\mu_\varphi(\tau) = \|\varphi\|_p^p$ and

$$\begin{aligned}\int_{\frac{1}{|\tau|}\leq|a|\leq 1} |a|^p\mathrm{d}V(a) &\leq \frac{1}{|\tau|^{p_2}}\int_{\frac{1}{|\tau|}\leq|a|\leq 1} |a|^2\mathrm{d}V(a) \\ &\leq \frac{1}{|\tau|^{p_2}}\underbrace{\int_{\mathbb{R}\setminus\{0\}} \min(1, |a|^2)\mathrm{d}V(a)}_{c_V} \\ &= \frac{c_V}{|\tau|^{p_2}}, \quad (66)\end{aligned}$$

where $p_2 = \min(0, p-2)$. Hence,

$$\begin{aligned}&\int_{|\bar{a}|\geq 1} |\bar{a}|^p\mathrm{d}V_\varphi(\bar{a}) \leq \\ &\|\varphi\|_p^p\int_{|a|\geq 1} |a|^p\mathrm{d}V(a) + c_V\|\varphi\|_{\max(2,p)}^{\max(2,p)}. \quad (67)\end{aligned}$$

This proves that

$$\int_{|a|\geq 1} |a|^p\mathrm{d}V(a) < \infty \quad\Rightarrow\quad \int_{|\bar{a}|\geq 1} |\bar{a}|^p\mathrm{d}V_\varphi(\bar{a}) < \infty. \quad (68)$$

Next, we prove the converse statement. The assumption $\varphi \neq 0$ necessitates the existence of $0 < T \leq 1$ such that $\|\varphi\|_{p,(T)}^p = \int_{|\tau|\geq T} |\tau|^p\mathrm{d}\mu_\varphi(\tau)$ is strictly positive. This helps us bound (64) as

$$\begin{aligned}\int_{|\bar{a}|\geq 1} |\bar{a}|^p\mathrm{d}V_\varphi(\bar{a}) &\geq \int_{|\tau|\geq T} |\tau|^p\mathrm{d}\mu_\varphi(\tau)\cdot\int_{|a|\geq\frac{1}{T}} |a|^p\mathrm{d}V(a) \\ &= \|\varphi\|_{p,(T)}^p\bigg(\int_{|a|\geq 1} |a|^p\mathrm{d}V(a) - \int_{1\leq|a|\leq\frac{1}{T}} |a|^p\mathrm{d}V(a)\bigg) \\ &\geq \|\varphi\|_{p,(T)}^p\bigg(\int_{|a|\geq 1} |a|^p\mathrm{d}V(a) - \frac{c_V}{|T|^p}\bigg), \quad (69)\end{aligned}$$

which completes the proof.

Arash Amini received B.Sc., M.Sc., and Ph.D. degrees in electrical engineering (Communications and Signal Processing) in 2005, 2007, and 2011, respectively, and the B.Sc. degree in petroleum engineering (Reservoir) in 2005, all from Sharif University of Technology (SUT), Tehran, Iran. He was a researcher at École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland from 2011 to 2013 working on statistical approaches towards modeling sparsity in continuous-domain.

Since September 2013, he is an assistant professor at Sharif University of Technology, Tehran, Iran. His research interests include various topics in statistical signal processing, specially compressed sensing.

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 (ISBI2002) 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.

REFERENCES

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

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

[3] M. Unser, P. Tafti, and Q. Sun, "A unified formulation of Gaussian *vs.* sparse stochastic processes: Part I—Continuous-domain theory," *arXiv:1108.6150v1 [cs.IT]*.

[4] M. Unser, P. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian vs. sparse stochastic processes: Part II—Discrete-domain theory," *arXiv:1108.6152v1 [cs.IT]*.

[5] J. S. Turek, I. Yavneh, M. Protter, and M. Elad, "On MMSE and MAP denoising under sparse representation modeling over a unitary dictionary," *IEEE Trans. Sig. Proc.*, vol. 59, no. 8, pp. 3526–3535, Aug. 2011.

[6] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*. Chapman & Hall, 1994.

[7] A. N. Kolmogorov, "La transformation de Laplace dans les espaces linéaires," *C. R. Acad. Sci. Paris*, vol. T200, pp. 1717–1718, Jan.-Jun. 1935.

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

[9] Q. Sun and M. Unser, "Left-inverses of fractional Laplacian and sparse stochastic processes," *Adv. Comp. Math.*, vol. 36, no. 3, pp. 399–441, Apr. 2012.

[10] A. Amini, U. Kamilov, E. Bostan, and M. Unser, "Bayesian estimation for continuous-time sparse stochastic processes," *to appear in IEEE Trans. Sig. Proc.*

[11] A. Amini, U. Kamilov, and M. Unser, "Bayesian denoising of generalized Poisson processes with finite rate of innovation," in *Intern. Conf. on Acous., Speech, and Sig. Proc. (ICASSP)*, Kyoto, Japan, Mar. 25–30, 2012, pp. 3629–3632.

[12] M. Unser and T. Blu, "Cardinal exponential splines: Part I—Theory and filtering algorithms," *IEEE Trans. Sig. Proc.*, vol. 53, no. 4, pp. 1425–1438, Apr. 2005.

[13] U. S. Kamilov, P. Pad, A. Amini, and M. Unser, "MMSE estimation of sparse Lévy processes," *to appear in IEEE Trans. Sig. Proc.*

[14] E. J. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 877–905, Dec. 2008.

[15] E. Bostan, U. Kamilov, and M. Unser, "Reconstruction of biomedical images and sparse stochastic modeling," in *Intern. Symp. Biomed. Imag. (ISBI)*, Barcelona, Spain, May 2–5, 2012, pp. 880–883.

[16] O. E. Barndorff-Nielsen, T. Mikosch, and S. I. Resnick, *Lévy Processes: Theory and Applications*. Birkhäuser, 2001.

[17] S. J. Wolfe, "On the unimodality of infinitely divisible distribution functions," *Prob. Theo. & Rel. Fields*, vol. 45, no. 4, pp. 329–335, Dec. 1978.

[18] C. C. Heyde, "Some remarks on the moment problem (ii)," *Quart. J. Math. Oxford Ser.*, vol. 14, no. 1, pp. 97–105, 1963.

[19] B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*. Springer, Berlin, 2003.

[20] T. R. Johansen, "The Bochner-Minlos theorem for nuclear spaces and an abstract white noise space," Louisiana State University, Tech. Rep., 2003.