Optimized Wavelet Denoising for Self-Similar α-Stable Processes

Pedram Pad, *Student Member, IEEE,* Kasra Alishahi, and Michael Unser, *Fellow, IEEE*

***Abstract*— We investigate the performance of wavelet shrinkage methods for the denoising of symmetric-α-stable (SαS) self-similar stochastic processes corrupted by additive white Gaussian noise (AWGN), where α is tied to the sparsity of the process. The wavelet transform is assumed to be orthonormal and the shrinkage function minimizes the mean-square approximation error (MMSE estimator). We derive the corresponding formula for the expected value of the averaged estimation error. We show that the predicted MMSE is a monotone function of a simple criterion that depends on the wavelet and the statistical parameters of the process. Using the calculus of variations, we then optimize this criterion to find the best performing wavelet within the extended family of Meyer wavelets, which are bandlimited. These are compared with the Daubechies wavelets, which are compactly supported in time. We find that the wavelets that are shorter in time (in particular, the Haar basis) are better suited to denoise the sparser processes (say, $\alpha < 1.2$), while the bandlimited ones (including the Held and Shannon wavelets) offer the best performance for $\alpha > 1.6$, the limit corresponding to the Gaussian case (fBm) with $\alpha = 2$.**

***Index Terms*— Sparse signal processing, self-similar processes, discrete wavelet transforms, α-stable random variables, denoising.**

I. Introduction

A CLASSICAL example of a self-similar process is the fractional Brownian motion (fBm) [1]. It can be interpreted as the fractional integral of a continuous-domain white Gaussian noise (a.k.a. innovation) [2, Section 7.5.2]. The order of integration γ provides a direct control of the degree of fractality. This property makes the fBm a popular model of real-world signals. Modelling textures in images [3], [4], traffic in communication networks [5], and financial processes [6] are a few examples of applications of fBms. Higher-order generalizations of an fBm are characterized in [7].

The non-Gaussian counterpart of an fBm is the fractional stable motion [8], [9], which is generated by replacing the Gaussian innovation of the fBm by some α-stable white noise with $0 < \alpha \leq 2$ (the case $\alpha = 2$ corresponds to the Gaussian distribution and thus results in fBm). The fractional stable motion has also the properties of long-range dependency and self-similarity and enjoys a wide range of applications such as data traffic and network modelling [10]–[12], physics [13], [14], geophysical researches [15], [16] and water resource modelling [17].

For $\alpha < 2$, however, there is a special feature of α-stable processes that makes them fundamentally different from their Gaussian cousins: the fact that all their second-order moments (including the variance) are unbounded. This is equivalent to their statistical distributions being heavy tailed, which is the statistical transcription of the concept of sparsity [18], [19]. Again, this is very relevant to modern signal processing and to the development of algorithms for the recovery of sparse signals, including compressed sensing [20], [21].

To fully exploit the property of sparsity, it is still necessary to expand the signal in an appropriate basis. In the case of self-similar processes, the natural candidate is the wavelet transform whose decorrelation properties have been studied extensively in the Gaussian case [22]–[25]. Regarding the α-stable processes, there is some prior work on the determination of the statistical distribution of wavelet coefficients [26], as well as a recent demonstration of the ability of the Haar transform to provide an independent-component analysis of Lévy processes with $\gamma = 1$ and $\alpha < 1$ [27]. Our focus on wavelets is further motivated by the observation that they perform remarkably well in a variety of practical signal-processing tasks including coding [28], [29], signal reconstruction [30], [31], and denoising [32]–[34]. This is reinforced by fundamental results from approximation theory on the optimality of wavelets for the N-term approximation of functions in Besov spaces [35].

Our objective in this paper is to characterize the ability of wavelets to optimally denoise self-similar symmetric-α-stable (SαS) processes corrupted by additive white Gaussian noise (AWGN). We focus on the traditional architecture where the wavelet coefficients are processed independently of each other [36]–[39]. Our first objective is to predict the mean (the expected value) of the averaged estimation error (MAEE), in order to be able to compare the denoising performance of different wavelet bases. Interestingly, we can relate this quantity to a simple criterion: the L_α-norm of the γ th fractional integral of the mother wavelet. The availability of this criterion enables us to develop an infinite-dimensional optimization algorithm to find the optimal frequency profile for a Meyer wavelet [40].

The use of a component-wise minimum mean-square error (MMSE) estimation strategy ensures that the studied wavelet denoisers are the best solutions among the broad family of wavelet-shrinkage estimators. Moreover, it attains the global optimum (MMSE signal estimation) when the wavelet coefficients are perfectly decoupled (independent-component

Manuscript received April 4, 2016; revised October 24, 2016; accepted February 19, 2017. Date of publication March 22, 2017; date of current version August 16, 2017. This work was supported by the Hasler Foundation under Grant 16009. *(Corresponding author: Pedram Pad.)*

P. Pad and M. Unser are with the Biomedical Imaging Group, Ecole polytechnique fédérale de Lausanne, 1015 Lausanne, Switzerland.

K. Alishahi is with the Department of Mathematical Sciences, Sharif University of Technology, Tehran, Iran.

Communicated by A. H. Sayed, Associate Editor for Signal Processing.

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.2017.2686421

0018-9448 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

analysis). This suggests that the criterion can also be used as an indirect measure of the decoupling performance of a given wavelet basis for the underlying class of stochastic processes. Thus, a good wavelet according to this criterion should also be a good candidate for other applications such as coding.

To reach our goal, we have to prove two theorems that could be of interest on their own right. The first is a result that extends the use of the wavelet transform to signals that are not included in $L_2(\mathbb{R})$ but are only locally square-integrable. The second is a high-level characterization of the performance of the scalar MMSE estimator of an SαS random variable corrupted by Gaussian noise as a function of the dispersion parameter and the noise variance. Similar studies for finite-variance random variables have been conducted in [41]. Here, the fact that the variance of the signal is unbounded requires a more technical treatment.

The paper is organized as follows: In Section II, we present the notation and review the properties of self-similar SαS processes. Our signal-estimation problem is then formulated in Section III. In Section IV, we address the issue of the calculation of the average energy of a signal from its wavelet coefficients. In Section V, we specify the MMSE estimator of a scalar SαS random variable corrupted by AWGN and characterize its performance in terms of the dispersions of the signal and noise. The main result of the paper is presented in Section VI. In Section VII, using the calculus of variations, we propose an algorithm to find the optimal Meyer wavelet. The numerical results and the comparison between different Meyer and Daubechies wavelets are presented in Section VIII.

II. Notations and Preliminaries

A. Notations

Throughout the paper we use $\mathbb{R}$, $\mathbb{Z}$, $\mathbb{N}$ for the set of real, integer, and positive integer numbers, respectively. The Fourier transform of a function f is denoted by $\hat{f}$. Also, $\|f\|_\alpha$ represents the α-(pseudo)norm of f for any positive value α. $\langle f, \psi\rangle = \int f(t)\psi(t)\mathrm{d}t$ is the standard inner product between the functions f and ψ; when f is only locally integrable, ψ is assumed to have sufficient decay for it to be well-defined.

To denote the action of an operator L acting on a function f, we use the forms of $\mathrm{L}f$ or $\mathrm{L}\{f\}(x)$ depending on if we want to mention the whole function or its value at a specific point. $\mathbb{P}(A)$ and $\mathbb{E}[X]$ stand for the probability of the event A and the expected value of the random variable X.

B. Self-Similar SαS Processes

We now start with some preliminaries on self-similar SαS processes that are necessary for understanding the paper. We use the framework introduced in [2], [42], and [43] due to its convenience for the treatment of the wavelet coefficients of stochastic processes. However, we also make links to conventional stochastic calculus, which is more convenient for describing the behavior of the process in the time domain [8].

We start by the notion of white noise. Suppose that w is an SαS white noise with the Lévy exponent $-|a\omega|^\alpha$ for a given $a > 0$ [2, Chapter 4]. The quantity a^α is called the dispersion parameter. (It plays a role similar to the variance of the Gaussian case.) If we observe this noise through a window θ, the resultant random variable $W = \langle w, \theta\rangle$ has the characteristic function

$$\hat{p}_W(\omega) = \mathbb{E}\left[\mathrm{e}^{-\mathrm{j}\omega W}\right] = \exp\left(-\left|a\|\theta\|_\alpha\omega\right|^\alpha\right), \tag{1}$$

which is the characteristic function of the SαS random variable W with dispersion parameter $a^\alpha\|\theta\|_\alpha^\alpha$.

A process s is a self-similar SαS process of order $\gamma \geq 0$ if

$$\mathrm{D}^\gamma s = w, \tag{2}$$

where D^γ is the γth order derivative operator that is defined as [2, Chapter 7.5]

$$\mathrm{D}^\gamma\{f\}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} (\mathrm{j}\omega)^\gamma \hat{f}(\omega)\mathrm{e}^{\mathrm{j}\omega x}\mathrm{d}\omega. \tag{3}$$

For a test function θ, the random variable $\langle s, \theta\rangle$ can be identified with

$$\langle s, \theta\rangle = \langle \mathrm{D}^{-\gamma} w, \theta\rangle = \langle w, \mathrm{D}^{-\gamma *}\theta\rangle \tag{4}$$

in which $\mathrm{D}^{-\gamma *}$ is the L_α-stable adjoint of the inverse operator of D^γ that is defined by (see [2, Chapter 7.5])

$$\mathrm{D}^{-\gamma *}\{\theta\}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\hat{\theta}(\omega) - \sum_{k=0}^{\lfloor\gamma+\frac{1}{\alpha}\rfloor-1}\frac{\hat{\theta}^{(k)}(0)\omega^k}{k!}}{(\mathrm{j}\omega)^\gamma}\mathrm{e}^{\mathrm{j}\omega x}\mathrm{d}\omega \tag{5}$$

when $\gamma > 1 - \frac{1}{\alpha}$ and either $\gamma \in \mathbb{N}$ or $\gamma - \lfloor\gamma\rfloor > \frac{1}{\alpha} - 1$. Equivalently, in the space domain, we have that

$$\mathrm{D}^{-\gamma *}\{\theta\}(x) = \int_{\mathbb{R}} K(x,t)\theta(t)\mathrm{d}t, \tag{6}$$

where the kernel of the inverse operator is

$$K(x,t) = \frac{1}{\Gamma(\gamma)}(t-x)_+^{\gamma-1} - \sum_{k=0}^{\lfloor\gamma+\frac{1}{\alpha}\rfloor-1}\frac{t^k}{k!}\frac{(-x)_+^{\gamma-1-k}}{\Gamma(\gamma-k)} \tag{7}$$

in which $\Gamma(\cdot)$ is Euler's Gamma function where $(x)_+ = \max\{0, x\}$. Consequently, according to (1) and (4), $\langle s, \theta\rangle$ is an SαS random variable with the characteristic function

$$\hat{p}_{\langle s,\theta\rangle}(\omega) = \exp\left(-\left|a\|\mathrm{D}^{-\gamma *}\theta\|_\alpha\omega\right|^\alpha\right). \tag{8}$$

Despite the unifying aspect of these formulas in terms of α, the behavior of the sample paths of the process drastically change when we go from the finite-variance case ($\alpha = 2$) to the sparse case ($\alpha < 2$) [18]. For $\alpha = 2$, which is the Gaussian case, the sample paths are almost-surely continuous for any $\gamma > \frac{1}{2}$. By contrast, for $\alpha < 2$, the sample paths are almost-surely continuous only for $\gamma > 1$; otherwise, when $\gamma < 1$, the sample paths are unbounded on any interval with positive length [8, Chapter 10]. For the case $\gamma = 1$, which corresponds to Lévy processes, the sample paths are right-semicontinuous with left limit [44, Chapter 2]. On the asymptotic behavior of the sample paths, according to [44, Chapter 9] and [45, Theorem 1.3], we know that, if $\gamma \geq 1$, then there exists a constant $C_\eta \in \mathbb{R}$ for which

$$|s(x)| \leq C_\eta\ (1 + |x|)^\eta \tag{9}$$

for any $\eta > \gamma - 1 + \frac{1}{\alpha}$.

We mainly need two properties to describe the effect of the operator $\mathrm{D}^{-\gamma *}$. The first one is that the restriction of $\mathrm{D}^{-\gamma *}$

to the subspace of test functions with at least $\lfloor \gamma + \frac{1}{\alpha} \rfloor - 1$ vanishing moments is a shift- and scale-invariant operator (see (5)). More precisely, when θ is in this subspace, we have that

$$\mathrm{D}^{-\gamma *}\{\theta(c \cdot -b)\}(x) = c^{-\gamma}\, \mathrm{D}^{-\gamma *}\{\theta(\cdot)\}(cx - b) \tag{10}$$

for any $b, c \in \mathbb{R}$. If the test functions do not have enough vanishing moments, then $\mathrm{D}^{-\gamma *}$ loses its shift-invariance. In this situation, Lemma 1 applies.

Lemma 1: Suppose $\gamma > \max\{0, 1 - \frac{1}{\alpha}\}$ *and* θ *is a function for which there exists a constant* A *and* $n > 1$ *such that* $|\theta(x)| < A(1 + |x|)^{-n}$. *Also, assume that* $\theta(x)(1 + |x|)^{\gamma - 1} \in L_1(\mathbb{R})$ *and* $\int_{\mathbb{R}} \theta(x)\mathrm{d}x \neq 0$. *Under these assumptions, we have that*

$$\lim_{j \to \pm\infty} \|\mathrm{D}^{-\gamma *}\theta(\cdot - j)\|_\alpha = \infty. \tag{11}$$

The proof is given in Appendix A.

Lemma 1 means that the dispersion of $\langle s, \theta(\cdot - j)\rangle$ tends to infinity by letting j tend to infinity (see (8)). This fact results in significant simplifications in the derivation of our main result in Section VI.

III. Problem Formulation

Assume now that s is the self-similar SαS process defined by (2). Also, assume that z is an AWGN independent from s with variance σ^2. Then, the continuous-time stochastic process

$$\tilde{s} = s + z \tag{12}$$

is the noisy version of s.

We consider the following denoising problem: Given a realization of the process $\tilde{s}$, we want to estimate the corresponding realization of the process s. We denote the estimated version of s by $\hat{s}$. Notice that $\hat{s}$ is also a stochastic process which depends on s, z, and the method of estimation.

To quantify the performance of the estimation method, we use the mean of the averaged estimation error

$$\mathrm{MAEE}(s, \hat{s}) = \lim_{T \to \infty} \mathbb{E}_{s,\hat{s}} \left[\frac{1}{2T} \int_{-T}^{T} \big(s(x) - \hat{s}(x)\big)^2 \mathrm{d}x \right]. \tag{13}$$

Now, assume that φ and ψ are the father and mother wavelets of an orthonormal wavelet family, respectively. Then, for any $i_0 \in \mathbb{Z}$,

$$\{\varphi_{i_0 j}\}_{j \in \mathbb{Z}} \cup \left\{ \{\psi_{ij}\}_{j \in \mathbb{Z}} \right\}_{i = i_0}^{\infty} \tag{14}$$

is an orthonormal basis for $L_2(\mathbb{R})$ where

$$\varphi_{i_0 j}(x) = 2^{\frac{i_0}{2}} \varphi(2^{i_0} x - j) \tag{15}$$

and

$$\psi_{ij}(x) = 2^{\frac{i}{2}} \psi(2^i x - j). \tag{16}$$

Here, i_0 is the coarsest scale that we take into account. Under the conditions on s and on the wavelet basis functions discussed in Section IV, the equality

$$s(x) = \sum_{j \in \mathbb{Z}} \Pi_{i_0 j} \varphi_{i_0 j}(x) + \sum_{i \geq i_0} \sum_{j \in \mathbb{Z}} \Xi_{ij}\, \psi_{ij}(x) \tag{17}$$

holds almost everywhere for almost every realization of s when

$$\Pi_{i_0 j} = \langle s, \varphi_{i_0 j} \rangle, \tag{18}$$

$$\Xi_{ij} = \langle s, \psi_{ij} \rangle. \tag{19}$$

Now, the general wavelet-domain denoising method is that based on the wavelet coefficients of $\tilde{s}$, i.e.,

$$\widetilde{\Pi}_{i_0 j} = \langle \tilde{s}, \varphi_{i_0 j} \rangle, \tag{20}$$

$$\widetilde{\Xi}_{ij} = \langle \tilde{s}, \psi_{ij} \rangle, \tag{21}$$

we want to estimate the wavelet coefficients of s, i.e., $\Pi_{i_0 j}$ and Ξ_{ij}. We denote the corresponding estimated coefficients by $\widehat{\Pi}_{i_0 j}$ and $\widehat{\Xi}_{ij}$, respectively. Then, our estimation of the process s would be

$$\hat{s}(x) = \sum_{j \in \mathbb{Z}} \widehat{\Pi}_{i_0 j} \varphi_{i_0 j}(x) + \sum_{i \geq i_0} \sum_{j \in \mathbb{Z}} \widehat{\Xi}_{ij}\, \psi_{ij}(x). \tag{22}$$

To optimally calculate $\widehat{\Pi}_{i_0 j}$ and $\widehat{\Xi}_{ij}$, we would need to take into account all statistical dependencies among wavelet coefficients. But this is not computationally tractable except when the process is Gaussian. Therefore, most of the denoising algorithms calculate $\widehat{\Pi}_{i_0 j}$ and $\widehat{\Xi}_{ij}$ pointwise, based only on the corresponding wavelet coefficient $\widetilde{\Pi}_{i_0 j}$ and $\widetilde{\Xi}_{ij}$ of $\tilde{s}$. In other words, the dependencies to the other wavelet coefficients are neglected. Hence, these methods are called coefficient-wise denoising.

Our goal in this paper is to calculate $\mathrm{MAEE}(s, \hat{s})$ for a given wavelet basis using a coefficient-wise denoiser and to characterize the optimal solution. Since our criterion is in fact a function of φ, ψ, and i_0, we write it as $\mathrm{MAEE}_{i_0}(\varphi, \psi)$.

First, Theorem 1 tells us that the optimal coefficient-wise denoiser (the denoiser that minimizes $\mathrm{MAEE}_{i_0}(\varphi, \psi)$) is the coefficient-wise MMSE estimator. Then, we establish in Theorem 2 some properties for the MMSE function of denoising an SαS random variable. Based on these two theorems, we transform (13) into a concise formula. This formula allows us to easily compare different wavelets and to optimize a given design. We will also show in Section VI that for this denoiser the series on the right-hand side of (22) are convergent. This ensures that $\hat{s}$ is well-defined.

The studied estimator is globally suboptimal because the denoising is performed coefficient-wise. This suggests that the denoising result obtained with "more independent" wavelet coefficients should be closer to the global minimizer of (13). Therefore, the derived formula can also measure the ability of a given wavelet basis to decouple self-similar SαS stochastic processes.

IV. Calculating Average Energy of a Function Using Its Wavelet Coefficients

This section is devoted to the calculation of the average energy of a signal based on its wavelet coefficients. This is not a trivial task because the functions under consideration are not included in $L_2(\mathbb{R})$. Such a characterization is also required to lend meaning to (17).

Consider the wavelet family defined in (14)-(15). For a function $f \in L_2(\mathbb{R})$, we have that

$$f(x) = \sum_{j\in\mathbb{Z}} \langle f, \varphi_{i_0 j}\rangle \varphi_{i_0 j}(x) + \sum_{i\geq i_0}\sum_{j\in\mathbb{Z}} \langle f, \psi_{ij}\rangle \psi_{ij}(x) \quad (23)$$

and

$$\|f\|_2^2 = \sum_{j\in\mathbb{Z}} \langle f, \varphi_{i_0 j}\rangle^2 + \sum_{i\geq i_0}\sum_{j\in\mathbb{Z}} \langle f, \psi_{ij}\rangle^2. \quad (24)$$

This means that we can calculate the energy of a square-integrable function from its wavelet coefficients. In Theorem 1, we show that the concept generalizes to the determination of the average energy, even for signals that are not square-integrable, provided that the wavelets have a sufficient decay. Notice that the family of signals that have a finite average energy is much broader than $L_2(\mathbb{R})$ since L_2-functions have an average energy of zero.

Theorem 1: Suppose that the wavelets are such that

$$|\varphi(x)|, |\psi(x)| \leq \eta(x) = \frac{A}{(1+|x|)^{n+1+\epsilon}} \quad (25)$$

for some $A, n, \epsilon > 0$. Then, for any function f for which there exists $B > 0$ that satisfies

$$|f(x)| \leq g(x) = B(1+|x|)^n, \quad (26)$$

we have that

$$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f(x)^2 \mathrm{d}x = \lim_{N\to\infty} \frac{1}{2N}\left(\sum_{|j|\leq 2^{i_0}N} \langle f, \varphi_{i_0 j}\rangle^2 + \sum_{i\geq i_0}\sum_{|j|\leq 2^i N} \langle f, \psi_{ij}\rangle^2\right). \quad (27)$$

The proof is given in Appendix B.

This theorem helps us evaluate (13) for a wavelet-domain denoising method. We can also establish the complementary convergence result whose proof is given in Appendix C.

Proposition 1: Under the assumptions of Theorem 1,

$$f(x) = \sum_{j\in\mathbb{Z}} \langle f, \varphi_{i_0 j}\rangle \varphi_{i_0 j}(x) + \sum_{i\geq i_0}\sum_{j\in\mathbb{Z}} \langle f, \psi_{ij}\rangle \psi_{ij}(x) \quad (28)$$

holds almost everywhere.

V. MMSE Denoising of SαS Random Variables

In this section, we study the MMSE performance of a denoiser that is applied to a scalar SαS random variable contaminated by Gaussian noise. Suppose that

$$Y = X + Z, \quad (29)$$

where X is an SαS random variable with dispersion parameter a^α and Z is a Gaussian random variable, independent from X, with mean 0 and variance σ^2.

According to Stein's formula, the MMSE estimator of X given Y is [46]

$$\widehat{X} = \mathbb{E}[X|Y] = Y + \sigma^2 \frac{p_Y'(Y)}{p_Y(Y)} \quad (30)$$

in which p_Y and p_Y' are the probability density function of Y and its derivative, respectively. Notice that

$$p_Y = p_X * p_Z. \quad (31)$$

where p_X and p_Z are the probability density functions of X and Z, respectively, and $*$ denotes the convolution operator. Stein's theory also provides the MMSE that is achieved by the optimal denoiser as [27], [47]

$$\mathbb{E}\big[(X-\widehat{X})^2\big] = \sigma^2 - \sigma^4 \int \frac{\big(p_Y'(y)\big)^2}{p_Y(y)}\mathrm{d}y. \quad (32)$$

We call this function MMSE(a,σ), where a^α is the dispersion of the signal and σ^2 is the variance of the noise.

In Theorem 2, we prove some properties for MMSE(a,σ) that we use later in the derivation of our main result. A similar function is studied in [41] and [48], mostly in the scenario where X is a finite-variance random variable. The challenge here is that SαS random variables have infinite variance for $\alpha < 2$.

Theorem 2: Let $Y = X + Z$ where X is an SαS random variable with dispersion parameter a^α and Z is a Gaussian random variable, independent from X, with mean 0 and variance σ^2. Also, recall that MMSE(a,σ) is the minimum mean square error of estimating X from Y. Then, the following properties hold for MMSE(a,σ) as a function of its first argument a:

1) *Increasing over $[0,+\infty)$.*
2) *Towards infinity, we have*

$$\lim_{a\to\infty} MMSE(a,\sigma) = \sigma^2. \quad (33)$$

3) *Around zero, we have*

$$\lim_{a\to 0} \frac{MMSE(a,\sigma)}{a^{\alpha-\epsilon}} = 0 \quad (34)$$

for any $\epsilon > 0$.

For the proof, refer to Appendix D.

Also, notice that we have

$$\mathrm{MMSE}(a,\sigma) \leq \sigma^2 \quad (35)$$

for any a and σ since the identity estimator, i.e. $\widehat{X} = Y$, has the MSE of σ^2. This can also be seen from (32).

Additionally, by plotting the graph of MMSE(a,σ) around $a = 0$ in the computer, we propose Conjecture 1. This conjecture is not required for our analysis but helps us with a nice intuitive interpretation of the final result.

Conjecture 1: Around zero, we have

$$\lim_{a\to 0} \frac{MMSE(a,\sigma)}{a^{\alpha+\epsilon}} = \infty \quad (36)$$

for any $\epsilon > 0$.

VI. MAEE of Component-Wise Wavelet-Based Denoising

We now present our main result on the solution of the problem formulated in Section III. We concentrate on the case

where $\gamma \geq 1$ and the mother wavelet ψ has at least $\lfloor \gamma + \frac{1}{\alpha} \rfloor - 1$ vanishing moments. Also, we assume that

$$|\varphi(x)|, |\psi(x)| \leq \frac{A}{(1+|x|)^{\eta}} \tag{37}$$

for an $\eta > \gamma + \frac{1}{\alpha}$ and $A \in \mathbb{R}$. These assumptions, along with (9) and Proposition 1, yield

$$s(x) = \sum_{j \in \mathbb{Z}} \langle s, \varphi_{i_0 j} \rangle \varphi_{i_0 j}(x) + \sum_{i \geq i_0} \sum_{j \in \mathbb{Z}} \langle s, \psi_{ij} \rangle \psi_{ij}(x) \tag{38}$$

almost surely.

Also, we can straightforwardly characterize the wavelet coefficients of $\Xi_{ij} = \langle s, \psi_{ij} \rangle$. Assume that

$$\phi = \mathrm{D}^{-\gamma} * \psi. \tag{39}$$

As discussed in Section II-B and as a result of the vanishing moments of ψ, we can write

$$\phi_{ij}(x) = \mathrm{D}^{-\gamma} * \{\psi_{ij}\}(x) = 2^{i(\frac{1}{2}-\gamma)} \phi(2^i x - j). \tag{40}$$

Therefore, according to (4) and (8), the characteristic function of Ξ_{ij} (see also (19)) is

$$\hat{p}_{\Xi_{ij}}(\omega) = \exp\left(-\left|a \|\phi_{ij}\|_{\alpha} \omega\right|^{\alpha}\right). \tag{41}$$

A crucial point is that $\|\phi_{ij}\|_{\alpha}$ only depends on the scale index i. It is given by

$$\|\phi_{ij}\|_{\alpha} = 2^{i(\frac{1}{2}-\frac{1}{\alpha}-\gamma)} \|\phi\|_{\alpha}. \tag{42}$$

Similarly, we find that

$$\hat{p}_{\Pi_{i_0 j}}(\omega) = \exp\left(-\left|a \|\mathrm{D}^{-\gamma} * \varphi_{i_0 j}\|_{\alpha} \omega\right|^{\alpha}\right). \tag{43}$$

But, since φ is orthogonal to ψ, it does not have any vanishing moments. Thus, according to Section II-B, $\mathrm{D}^{-\gamma} * \varphi_{i_0 j}$ with j varying are not shifted versions of each other. However, Lemma 1 is all what we need about them.

The combination of (12) with (20)-(19) implies that

$$\widetilde{\Pi}_{i_0 j} = \Pi_{i_0 j} + Z'_{i_0 j}, \tag{44}$$

$$\widetilde{\Xi}_{ij} = \Xi_{ij} + Z_{ij}, \tag{45}$$

where

$$Z'_{i_0 j} = \langle z, \varphi_{i_0 j} \rangle, \tag{46}$$

$$Z_{ij} = \langle z, \psi_{ij} \rangle. \tag{47}$$

Thanks to the orthonormality of the wavelet family, the random variables $Z'_{i_0 j}$ and Z_{ij} are *iid* Gaussian with mean 0 and variance σ^2. This allows us to determine the coefficient-wise MMSE estimation of $\Pi_{i_0 j}$ and Ξ_{ij} by the direct application of the scalar estimators discussed in Section V.

In Section III, $\mathrm{MAEE}_{i_0}(\varphi, \psi)$ is the mean average error energy of the optimal coefficient-wise denoising. Thanks to Theorem 1, it is reformulated in the wavelet domain as

$$\begin{aligned}
&\mathrm{MAEE}_{i_0}(\varphi, \psi) \\
&\quad = \lim_{N \to \infty} \frac{1}{2N} \mathbb{E}\Bigg[\sum_{|j| \leq 2^{i_0} N} \left(\Pi_{i_0 j} - \widehat{\Pi}_{i_0 j}\right)^2 \\
&\qquad\qquad + \sum_{i \geq i_0} \sum_{|j| \leq 2^i N} \left(\Xi_{ij} - \widehat{\Xi}_{ij}\right)^2 \Bigg] \\
&\quad = \lim_{N \to \infty} \frac{1}{2N} \Bigg(\sum_{|j| \leq 2^{i_0} N} \mathbb{E}\left[\left(\Pi_{i_0 j} - \widehat{\Pi}_{i_0 j}\right)^2\right] \\
&\qquad\qquad + \sum_{i \geq i_0} \sum_{|j| \leq 2^i N} \mathbb{E}\left[\left(\Xi_{ij} - \widehat{\Xi}_{ij}\right)^2\right] \Bigg).
\end{aligned} \tag{48}$$

Based on this formula, we infer that the optimal coefficient-wise denoiser (the one that gives us the minimum MAEE) is the one that provides the MMSE for each coefficient. It is thus the coefficient-wise MMSE denoiser. In this case and according to Section V and (41)-(43), we have that

$$\mathbb{E}\left[\left(\Pi_{i_0 j} - \widehat{\Pi}_{i_0 j}\right)^2\right] = \mathrm{MMSE}\left(a \|\mathrm{D}^{-\gamma} * \varphi_{i_0 j}\}\|_{\alpha}, \sigma\right), \tag{49}$$

$$\mathbb{E}\left[\left(\Xi_{ij} - \widehat{\Xi}_{ij}\right)^2\right] = \mathrm{MMSE}\left(2^{i(\frac{1}{2}-\frac{1}{\alpha}-\gamma)} a \|\phi\|_{\alpha}, \sigma\right). \tag{50}$$

Replacing (49) and (50) in (48), we get

$$\begin{aligned}
&\mathrm{MAEE}_{i_0}(\varphi, \psi) \\
&\quad = \lim_{N \to \infty} \frac{1}{2N} \Bigg(\sum_{|j| \leq 2^{i_0} N} \mathrm{MMSE}\left(a \|\mathrm{D}^{-\gamma} * \varphi_{i_0 j}\}\|_{\alpha}, \sigma\right) \\
&\qquad + \sum_{i \geq i_0} (2^{i+1} N + 1) \mathrm{MMSE}\left(2^{i(\frac{1}{2}-\frac{1}{\alpha}-\gamma)} a \|\phi\|_{\alpha}, \sigma\right) \Bigg).
\end{aligned} \tag{51}$$

Now, (42) gives us

$$\begin{aligned}
&\mathrm{MAEE}_{i_0}(\varphi, \psi) \\
&\quad = \lim_{N \to \infty} \frac{1}{2N} \sum_{|j| \leq 2^{i_0} N} \mathrm{MMSE}\left(a \|\mathrm{D}^{-\gamma} * \varphi_{i_0 j}\}\|_{\alpha}, \sigma\right) \\
&\qquad + \sum_{i \geq i_0} 2^i \mathrm{MMSE}\left(2^{i(\frac{1}{2}-\frac{1}{\alpha}-\gamma)} a \|\phi\|_{\alpha}, \sigma\right).
\end{aligned} \tag{52}$$

Also, the existing limit in the right-hand side of (52) can be calculated by combining Lemma 1 and Part 1 of Theorem 2. Since $\varphi_{i_0 j}$ has no vanishing moments, $\|\mathrm{D}^{-\gamma} * \varphi_{i_0 j}\|_{\alpha}$ tends to infinity as j goes to infinity. Thus, we conclude that

$$\lim_{j \to \infty} \mathrm{MMSE}\left(a \|\mathrm{D}^{-\gamma} * \varphi_{i_0 j}\|_{\alpha}, \sigma\right) = \sigma^2. \tag{53}$$

This means that there is no asymptotic advantage to denoise the coefficients corresponding to the scaling functions (the low-pass basis functions) since this denoising performs no better than the identity estimator whose MSE is σ^2. Incorporating (53) in (52), we finally obtain

$$\begin{aligned}
\mathrm{MAEE}_{i_0}(\varphi, \psi) &= 2^{i_0} \sigma^2 + \sum_{i \geq i_0} 2^i \\
&\quad \times \mathrm{MMSE}\left(2^{i(\frac{1}{2}-\frac{1}{\alpha}-\gamma)} a \|\phi\|_{\alpha}, \sigma\right).
\end{aligned} \tag{54}$$

Since $\gamma > \frac{1}{2}$, there exists $\epsilon > 0$ for which

$$\big(\frac{1}{2} - \frac{1}{\alpha} - \gamma\big)(\alpha - \epsilon) + 1 < 0. \tag{55}$$

Thus, according to Part 3 of Theorem 2, the summation in (54) is finite, which is reassuring for our application. If it was otherwise, any coefficient-wise denoising in any wavelet domain would result in an unbounded mean average error energy, suggesting that there would be no advantage in trying to optimize the wavelet basis. Here are other interesting consequences of our analysis.

Remark 1: The finiteness of $\mathrm{MAEE}_{i_0}(\varphi, \psi)$ implies that the series in (22) are almost surely convergent when $\widehat{\Pi}_{i_0 j}$ and $\widehat{\Xi}_{ij}$ are the coefficient-wise MMSE estimations of $\Pi_{i_0 j}$ and Ξ_{ij} from $\widetilde{\Pi}_{i_0 j}$ and $\widetilde{\Xi}_{ij}$. Thus, almost every realization of $\hat{s}$ is a well-defined function that is locally square-integrable. Notice that the same statement does not hold for $\tilde{s}$ since almost every realization of z is not a locally L_2 function.

Remark 2: If we forget about the time-domain representation and consider (48) as our original definition of performance measure as motivated by Theorem 1, then the story can also be told with γ less than 1. The parameter γ can be seen as a measure of the spatial coupling of the process instants: $\gamma = 0$ corresponds to white noise with no coupling, and larger γ leads to more coupling. Then, Conjecture 1 reveals that it is possible to achieve a finite value for the estimation error by performing a *coefficient-wise* wavelet-domain denoising only if $\gamma \geq \frac{1}{2}$. This happens when the signal exhibits a sufficient amount of coupling. This is an interesting duality relation between the amount of dependency that exists among the spatial instants of the original process and the amount of dependency that we must utilize among its wavelet coefficients to denoise them.

Moreover, since $\mathrm{MMSE}(a, \sigma) \leq \sigma^2$ for any a, a decrease in i_0 results in a decrease of (54). This means that the use of more resolution levels in the denoising procedure gives better results. Thus, by letting i_0 tend to $-\infty$, we get the quantity

$$\mathrm{MAEE}(\psi) = \sum_{i \in \mathbb{Z}} 2^i \mathrm{MMSE}\left(2^{i(\frac{1}{2} - \frac{1}{\alpha} - \gamma)} a \|\phi\|_\alpha, \sigma\right) \tag{56}$$

which is the least achievable MAEE by performing a coefficient-wise wavelet-domain denoising for recovering SαS self-similar processes embedded in AWGN.

An interesting point about (54) and (56) is that, although the function $\mathrm{MMSE}(\cdot, \cdot)$ is not known analytically, a smaller $\|\phi\|_\alpha$ results in a smaller $\mathrm{MAEE}(\psi)$. This is because we know that $\mathrm{MMSE}(\cdot, \cdot)$ is an increasing function of its first argument (Part 2 of Theorem 2). This means that, to compare the denoising performance of two different families of wavelet for γ-order SαS processes, it is enough to compare the α-norm of the γ-order integration of their mother wavelets; i.e., $\left\|\mathrm{D}^{-\gamma} * \psi\right\|_\alpha$. This observation tremendously simplifies the design of the optimal wavelet. Also, it implies that the optimal wavelet depends neither on the dispersion a^α of the signal nor on the variance σ^2 of the noise. Notice that, although it is obvious that the multiplication of a and σ by a constant does not affect the optimal wavelet, the independence of the optimal wavelet from $\frac{a}{\sigma}$ is not obvious a priori.

VII. Optimal Meyer Wavelets for Denoising Self-Similar SαS Processes

Based on the result of Section VI and with the help of the calculus of variations, we now propose an algorithm to design the optimal wavelet for a given γ and α within the so-called Meyer family of bandlimited wavelets [40].

According to [40], [49], for a given function $\upsilon : [0, 1] \to \mathbb{R}$ such that

$$\upsilon(\rho) + \upsilon(1 - \rho) = 1, \tag{57}$$

the profile

$$\mathcal{W}\{\upsilon\}(\omega) = \begin{cases} \sin\left(\frac{\pi}{2}\upsilon(\frac{3}{2\pi}|\omega| - 1)\right), & \frac{2\pi}{3} < |\omega| \leq \frac{4\pi}{3} \\ \cos\left(\frac{\pi}{2}\upsilon(\frac{3}{4\pi}|\omega| - 1)\right), & \frac{4\pi}{3} < |\omega| \leq \frac{8\pi}{3} \\ 0, & \text{otherwise} \end{cases} \tag{58}$$

is the Fourier transform of the mother wavelet of an orthonormal wavelet basis. These wavelets are called Meyer wavelets.

We just saw that in order to rank the denoising performance of different wavelet bases, it is enough to compare the α-(pseudo)norm of the γ-order integration of their mother wavelets. Hence, for a Meyer wavelet, the criterion

$$\begin{aligned} Q_\alpha^\gamma(\upsilon) &= \int_{\mathbb{R}} \left| \frac{1}{2\pi} \int_{\mathbb{R}} \frac{\mathcal{W}\{\upsilon\}(\omega)}{(\mathrm{j}\omega)^\gamma} \mathrm{e}^{\mathrm{j}\omega x} \mathrm{d}\omega \right|^\alpha \mathrm{d}x \\ &= \frac{1}{\pi^\alpha} \int_{\mathbb{R}} \left| \int_{\frac{2\pi}{3}}^{\frac{8\pi}{3}} \frac{\mathcal{W}\{\upsilon\}(\omega)}{\omega^\gamma} \cos\big(\omega x - \frac{\pi}{2}\gamma\big) \mathrm{d}\omega \right|^\alpha \mathrm{d}x \end{aligned} \tag{59}$$

is a predictor of its denoising performance. Next, we apply a projected-gradient-descent algorithm with adaptive step size to find the function υ that minimizes $Q_\alpha^\gamma(\upsilon)$. An adaptive step size is specially important for $\alpha \leq 1$ for which the functional does not have a Lipschitz gradient. The pseudo-code of our optimization method is given in Algorithm 1. In the algorithm, ∇Q_α^γ is the infinite-dimensional gradient of the functional Q_α^γ in the Hilbert space of $L_2([0, 1])$. Also, $\mathcal{P}\{\tilde{\upsilon}\}$ is the projector that maps $\tilde{\upsilon}$ to the nearest function that satisfies (57).

According to Appendix E, ∇Q_α^γ is calculated as

$$\begin{aligned} \nabla Q_\alpha^\gamma\{\upsilon\}(\rho) = &\frac{1}{\pi^\alpha} \frac{\pi^2 \alpha}{3} \big(\frac{3}{2\pi}\big)^\gamma \frac{\cos\left(\frac{\pi}{2}\upsilon(\rho)\right)}{(\rho + 1)^\gamma} \\ &\times \int_{\mathbb{R}} \lambda\left(\int_{\frac{2\pi}{3}}^{\frac{8\pi}{3}} \frac{\mathcal{W}\{\upsilon\}(\omega)}{\omega^\gamma} \cos\big(\omega x - \frac{\pi}{2}\gamma\big)\, \mathrm{d}\omega \right) \\ &\times \cos\left(\frac{2\pi}{3}(\rho + 1)x - \frac{\pi}{2}\gamma\right) \mathrm{d}x \\ &- \frac{2}{\pi^\alpha} \frac{\pi^2 \alpha}{3} \big(\frac{3}{4\pi}\big)^\gamma \frac{\sin\left(\frac{\pi}{2}\upsilon(\rho)\right)}{(\rho + 1)^\gamma} \\ &\times \int_{\mathbb{R}} \lambda\left(\int_{\frac{2\pi}{3}}^{\frac{8\pi}{3}} \frac{\mathcal{W}\{\upsilon\}(\omega)}{\omega^\gamma} \cos\big(\omega x - \frac{\pi}{2}\gamma\big)\, \mathrm{d}\omega \right) \\ &\times \cos\left(\frac{4\pi}{3}(\rho + 1)x - \frac{\pi}{2}\gamma\right) \mathrm{d}x, \end{aligned} \tag{60}$$

where $\lambda(x) = \mathrm{sgn}(x)|x|^{\alpha - 1}$. Also, in Appendix F, we prove that

$$\mathcal{P}\{\tilde{\upsilon}\}(\rho) = \frac{\tilde{\upsilon}(\rho) - \tilde{\upsilon}(1 - \rho) + 1}{2}. \tag{61}$$

Algorithm 1 Optimal Meyer Wavelet

initialize: $\upsilon \in L_2([0,1])$
initialize: $\eta > 0$
initialize: $\kappa^+ \geq 1$ and $\kappa^- \leq 1$
$Q \leftarrow Q_\alpha^\gamma(\upsilon)$
repeat
  $\upsilon_{old} \leftarrow \upsilon$
  $Q_{old} \leftarrow Q$
  $\tilde{\upsilon} \leftarrow \upsilon - \eta \nabla Q_\alpha^\gamma\{\upsilon\}$
  $\upsilon \leftarrow \mathcal{P}\{\tilde{\upsilon}\}$
  $Q \leftarrow Q_\alpha^\gamma(\upsilon)$
  if $Q \leq Q_{old}$ **then**
    $\eta \leftarrow \kappa^+ \cdot \eta$
  else
    $\upsilon \leftarrow \upsilon_{old}$
    $Q \leftarrow Q_{old}$
    $\eta \leftarrow \kappa^- \cdot \eta$
  end if
until υ converges
return υ

Thus, we have all the ingredients to implement Algorithm 1.

It is worth mentioning that for $\alpha = 2$ as long as $\gamma > \frac{1}{2}$, which in fact includes all fBms, the minimizer of Q_2^γ can be derived analytically. It is indeed the Shannon wavelet, irrespective of the value of γ. This result is in accordance with the well-known result about the optimality of Shannon wavelets for the minimum approximation error of processes with non-increasing spectrum [50]. The formal statement of this result is given in Proposition 2.

Proposition 2: If $\alpha = 2$ and $\gamma > \frac{1}{2}$, then

$$\upsilon(\rho) = \mathbf{1}_{\left[\frac{1}{2},1\right]}(\rho), \tag{62}$$

which corresponds to the wavelet with the Fourier profile

$$\mathcal{W}\{\upsilon\}(\omega) = \mathbf{1}_{[-2\pi,-\pi]}(\omega) + \mathbf{1}_{[\pi,2\pi]}(\omega), \tag{63}$$

minimizes $Q_2^\gamma(\upsilon)$.

The proof is provided in Appendix G.

VIII. Numerical Optimization of Meyer Wavelets and Comparison of Different Wavelet Families

In this section, we give the wavelet for some values of γ and α optimized according to the derivation of Section VII. Additionally, we compare the performance of Meyer wavelets which are compactly supported in the Fourier domain with the Daubechies wavelets [51] that are compactly supported in the time domain.

To implement Algorithm 1, we take the samples of υ and $\mathcal{W}\{\upsilon\}$ uniformly in the Fourier and time domain, respectively. Since these wavelets are not compactly supported, we have to truncate them at some point in the time domain.

We give in Figure 1 the outcome of the algorithm for $\gamma = 1$ and $\alpha = 1.2, 1.8$, and 2. We show in Figure 2 the plots for the optimal wavelet when $\alpha = 1.2$ and $\gamma = 1, 2, 4$. An interesting phenomenon that is observed in these

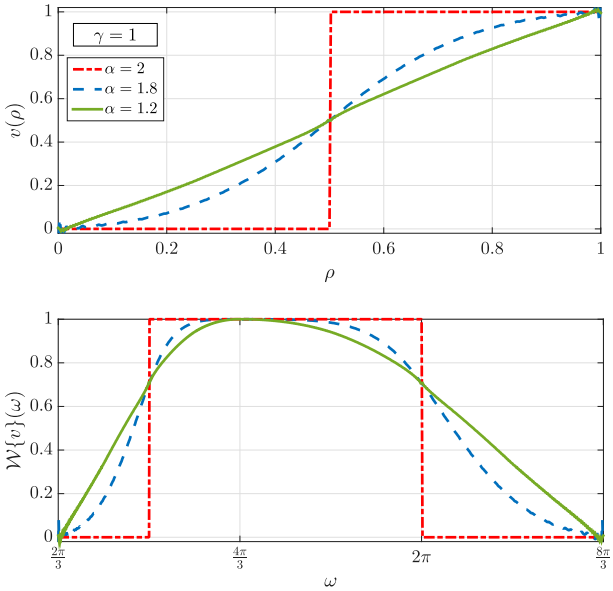

Fig. 1. Optimal $\upsilon(\rho)$ and corresponding wavelet profile $\mathcal{W}\{\upsilon\}(\omega)$ for $\gamma = 1$ and $\alpha = 1.2, 1.8$, and 2.

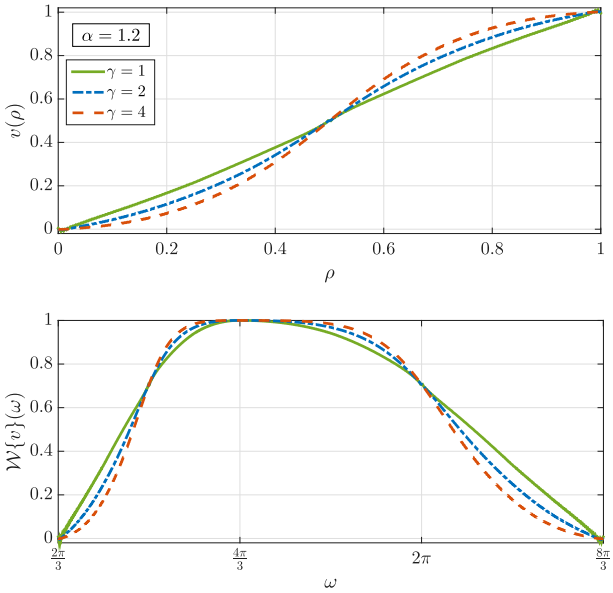

Fig. 2. Optimal $\upsilon(\rho)$ and corresponding wavelet profile $\mathcal{W}\{\upsilon\}(\omega)$ for $\alpha = 1.2$ and $\gamma = 1, 2$, and 4.

plots is that, by letting either α or γ increase, the wavelet approaches the Shannon wavelet ((62) and (63)). Regarding α, remember thet Proposition 2 states that the optimal wavelet for $\alpha = 2$ is exactly the Shannon wavelet. Regarding γ, we qualitatively shrink the high frequencies as γ increases. Thus, in order to have a smaller $Q_\alpha^\gamma(\upsilon)$, the frequency content tends toward higher frequencies and υ will have less weight on $[0, \frac{1}{2}]$. Therefore, the optimal wavelet approaches the Shannon wavelet which is vanishing on this interval (see (62)).

Conversely, the optimal wavelet tends to the Held wavelet ($\upsilon(\rho) = \rho$) when α decreases [52]. Thus, we can roughly say

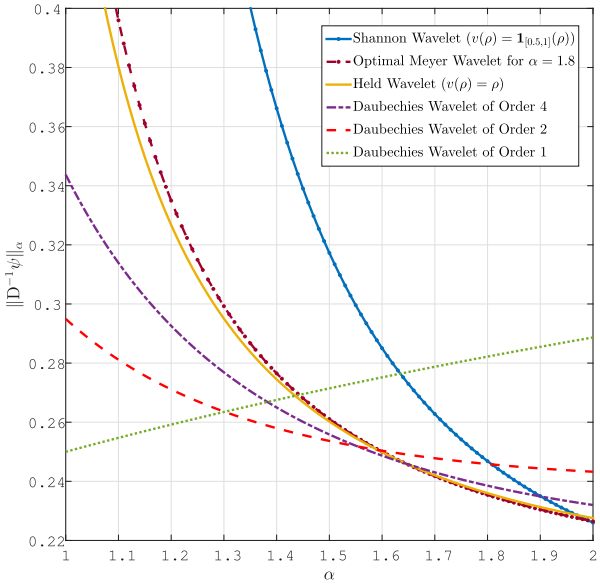

Fig. 3. The α-norm of the first-order integral of different mother wavelets versus α.

that, for very sparse cases (small α), it is better to use the Held wavelet; for less sparse cases (α close to 2), it is better to use the Shannon wavelet.

In the next step, we compare the MAEE performance of the optimal Meyer wavelets and Daubechies wavelets. Meyer wavelets lend themselves well to an FFT-based implementation due to their compact support in the Fourier domain (band-limitedness). Daubechies wavelets, on the other hand, are well-known for their minimal support in the time domain, which is valued in many applications.

According to the discussions in Section VI, comparing the MAEE obtained by different wavelets for denoising γ th-order SαS self-similar processes is equivalent to comparing the α-(pseudo)norm of the γ th-order integration of their mother wavelet. We plotted this quantity versus α in Figure 3 for the case of Lévy processes ($\gamma = 1$).

We observe that for very sparse signals (small α) it is better to use compactly supported wavelets (Daubechies wavelet). Moreover, we see that a smaller α favors a smaller wavelet support. Indeed, the Daubechies wavelet of order 1 (the Haar wavelet [53]) has the shortest support in the time domain and outperforms all the others for $\alpha \lesssim 1.3$. However, in less sparse cases (α close to 2), a compact support in the Fourier domain (band-limitedness) becomes more favorable. For special case $\alpha = 2$, the Shannon wavelet, which has the shortest support in the Fourier domain, outperforms the others, even if its superiority over the other Meyer wavelets is marginal.

IX. Summary and Conclusion

In this paper, we studied the performance of the wavelet-domain denoising of self-similar symetric-α-stable (SαS) processes corrupted by additive white Gaussian noise. We focused on the most classical denoising which proceeds coefficient-wise. We derived a simple formula for the mean average energy of estimation error (MAEE) for a given γ, α, and wavelet family, where γ is the order of the self-similar process under consideration. We showed that MAEE is an increasing function of the α-(pseudo)norm of the γ th-order integral of the mother wavelet ψ, i.e. $\|\mathrm{D}^{-\gamma} * \psi\|_\alpha$. This is an essential property that allows us to compare different wavelets based on this simple indicator of localization. The fact that the wavelet coefficients are treated independently is the only source of suboptimality of these denoisers. Thus, more correct is this assumption, MAEE is smaller. Therefore, the quantity $\|\mathrm{D}^{-\gamma} * \psi\|_\alpha$ can also be used to measure the decoupling performance of the wavelet. Moreover, the simplicity of the derived performance criterion allowed us to propose an optimization algorithm to find the optimal Meyer wavelet for a given γ and α. We could then compare the relative denoising performance of Meyer and Daubechies wavelets. For highly sparse signals (small α) we deduced that, it is better to use wavelets of compact support in the time domain; while for less sparse signals (α close to 2), it is better to use wavelets of compact support in the Fourier domain (Meyer wavelets).

To obtain these results, we proved two main theorems that are interesting on their own right. The first one enables us to calculate the average energy of a signal by using its wavelet coefficients (Theorem 1). The second one is about the minimum mean-square error (MMSE) function of estimating an SαS random variable given its summation with an independent Gaussian random variable (Theorem 2).

Appendix A
Proof of Lemma 1

First, notice that

$$\mathrm{D}^{-\gamma} * \{\theta(-\cdot)\}(x) = \mathrm{D}^{-\gamma} * \{\theta(\cdot)\}(-x). \tag{64}$$

Thus, it is enough to prove (11) when letting j tend to $+\infty$. According to (6) and (7), for $x > 0$, we have that

$$\mathrm{D}^{-\gamma} * \{\theta(\cdot - j)\}(x) = \frac{1}{\Gamma(\gamma)} \big((-\cdot)_+^{\gamma-1} * \theta(\cdot)\big)(x - j) \tag{65}$$

which implies that

$$\begin{aligned} &\|\mathrm{D}^{-\gamma} * \{\theta(\cdot - j)\}\|_\alpha \\ &\quad \geq \frac{1}{\Gamma(\gamma)} \left\| \big((-\cdot)_+^{\gamma-1} * \theta(\cdot)\big) \cdot \mathbf{1}_{[-j,0]} \right\|_\alpha, \end{aligned} \tag{66}$$

where $\mathbf{1}_{[-j,0]}$ is the indicator function of $[-j, 0]$. Now, since $\alpha(\gamma - 1) > -1$, it is sufficient to prove that $\big((-\cdot)_+^{\gamma-1} * \theta(\cdot)\big)(x)$ grows (decays) with the same rate as $(-x)_+^{\gamma-1}$ when x tends to $-\infty$; in other words

$$\lim_{x \to -\infty} \frac{\big((-\cdot)_+^{\gamma-1} * \theta(\cdot)\big)(x)}{(-x)_+^{\gamma-1}} = C, \tag{67}$$

where C is a nonzero real number. We first prove this statement when θ is a nonnegative function such that $\theta(x) \geq 0$ for all $x \in \mathbb{R}$, and then we generalize it to any θ.

Assuming that θ is a nonnegative function, $x < -1$, and $\frac{1}{n} < r < 1$, we write

$$\begin{aligned}\frac{\big((-\cdot)_+^{\gamma-1} * \theta(\cdot)\big)(x)}{(-x)_+^{\gamma-1}} &= \frac{\int_{\mathbb{R}}(t-x)_+^{\gamma-1}\theta(t)\mathrm{d}t}{(-x)^{\gamma-1}} \\ &= \int_{\mathbb{R}}\big(1-\frac{t}{x}\big)_+^{\gamma-1}\theta(t)\mathrm{d}t \\ &= \int_{|t|\le|x|^r}\big(1-\frac{t}{x}\big)_+^{\gamma-1}\theta(t)\mathrm{d}t \\ &\quad + \int_{|t|>|x|^r}\big(1-\frac{t}{x}\big)_+^{\gamma-1}\theta(t)\mathrm{d}t. \quad (68)\end{aligned}$$

If $\gamma \ge 1$, the first term of this summation is bounded by

$$\begin{aligned}&\big(1-\frac{1}{|x|^r}\big)^{\gamma-1}\int_{|t|\le|x|^r}\theta(t)\mathrm{d}t \\ &\le \int_{|t|\le|x|^r}\big(1-\frac{t}{x}\big)_+^{\gamma-1}\theta(t)\mathrm{d}t \\ &\le \big(1+\frac{1}{|x|^r}\big)^{\gamma-1}\int_{|t|\le|x|^r}\theta(t)\mathrm{d}t. \quad (69)\end{aligned}$$

Consequently, as x tends to $-\infty$, we obtain that

$$\lim_{x\to-\infty}\int_{|t|\le|x|^r}\big(1-\frac{t}{x}\big)_+^{\gamma-1}\theta(t)\mathrm{d}t = \int_{\mathbb{R}}\theta(t)\mathrm{d}t. \quad (70)$$

As for the second term in the right-hand side of (68), we write

$$\int_{-\infty}^{-|x|^r}\big(1-\frac{t}{x}\big)_+^{\gamma-1}\theta(t)\mathrm{d}t \le \big(1-\frac{1}{|x|^r}\big)^{\gamma-1}\int_{-\infty}^{-|x|^r}\theta(t)\mathrm{d}t \quad (71)$$

and, since $|x|^r < -x$, we have that

$$\int_{|x|^r}^{\infty}\big(1-\frac{t}{x}\big)_+^{\gamma-1}\theta(t)\mathrm{d}t \le \int_{|x|^r}^{\infty}\big(1+t\big)_+^{\gamma-1}\theta(t)\mathrm{d}t. \quad (72)$$

Since both integrals on the right-hand side of (71) and (72) are finite by assumption, by letting x tend to $-\infty$ we obtain that

$$\lim_{x\to-\infty}\int_{|t|>|x|^r}\big(1-\frac{t}{x}\big)_+^{\gamma-1}\theta(t)\mathrm{d}t = 0. \quad (73)$$

Up to now, we have proved that,

$$\lim_{x\to-\infty}\frac{\big((-\cdot)_+^{\gamma-1} * \theta(\cdot)\big)(x)}{(-x)_+^{\gamma-1}} = \int_{\mathbb{R}}\theta(t)\mathrm{d}t, \quad (74)$$

provided that θ is a nonnegative function. For a general function θ, let $\theta_+(x) = (\theta(x))_+$ and $\theta_-(x) = (-\theta(x))_+$. Thus, both of θ_+ and θ_- are nonnegative functions whilst $\theta = \theta_+ - \theta_-$. Incorporating this fact in (74) and using the bilinearity of the convolution operator, we can generalize (74) to any $\theta \in L_1(\mathbb{R})$. Therefore, since $\int_{\mathbb{R}}\theta(x)\mathrm{d}x \ne 0$, the argument (67) is proved for $\gamma > 1$.

For $\gamma < 1$, the argument is the same except that the upper- and lower-bounds need to be swapped in (69). Thus, (70) still holds. In (71), the inequality converts to

$$\begin{aligned}&\int_{-\infty}^{-|x|^r}\big(1-\frac{t}{x}\big)_+^{\gamma-1}\theta(t)\mathrm{d}t \\ &\quad\le \sup_{[x,-|x|^r]}|\theta(t)| \times \int_{x}^{-|x|^r}\big(1-\frac{t}{x}\big)_+^{\gamma-1}\mathrm{d}t \\ &\quad\le \frac{A}{(1+|x|^r)^n}\times\frac{-x}{\gamma}\big(1-\frac{1}{|x|^r}\big)^{\gamma}. \quad (75)\end{aligned}$$

Since $nr > 1$, the bound tends to zero as x tends to $-\infty$. The corresponding inequality for (72) with $\gamma < 1$ is

$$\int_{|x|^r}^{\infty}\big(1-\frac{t}{x}\big)_+^{\gamma-1}\theta(t)\mathrm{d}t \le \big(1+\frac{1}{|x|^r}\big)^{\gamma-1}\int_{|x|^r}^{\infty}\theta(t)\mathrm{d}t, \quad (76)$$

which again tends to zero as x tends to $-\infty$, and which completes the proof.

Appendix B
Proof of Theorem 1

First, notice that the conditions on f, φ, and ψ guarantee that the wavelet coefficients exist and are finite. We prove the result for $i_0 = 0$. It extends to other i_0 similarly. Define two projections

$$\mathrm{P}_T\{f\}(t) = f(t)\cdot\mathbf{1}_{[-T,T]} \quad (77)$$

in which $\mathbf{1}_{[-T,T]}$ is the indicator function of $[-T, T]$, and

$$\mathrm{Q}_N\{f\}(t) = \sum_{|j|\le N}\langle f, \varphi_{0j}\rangle\varphi_{0j}(t) + \sum_{i\ge0}\sum_{|j|\le 2^i N}\langle f, \psi_{ij}\rangle\psi_{ij}(t) \quad (78)$$

for $f : \mathbb{R}\to\mathbb{R}$ satisfying the conditions of the theorem. Later in the proof, we show that, for a fixed N, (78) is a converging series in L_2 and thus $\mathrm{Q}_N f$ is well-defined. In fact, P_T is the orthogonal projections onto the space of functions that are supported on $[-T, T]$, while Q_N is the orthogonal projection onto the space generated by $\{\varphi_{0j}\}_{|j|\le N}\cup\{\{\varphi_{ij}\}_{|j|\le 2^i N}\}_{i=0}^{\infty}$. Hence, we want to prove that

$$\lim_{T\to\infty}\frac{1}{2T}\|\mathrm{P}_T f\|_2^2 = \lim_{N\to\infty}\frac{1}{2N}\|\mathrm{Q}_N f\|_2^2. \quad (79)$$

Since P_T and Q_N are orthogonal projection, they do not increase the norm. Thus, we have that

$$\begin{aligned}\|\mathrm{P}_T f\|_2 &= \|\mathrm{P}_T\mathrm{Q}_N f + \mathrm{P}_T(1-\mathrm{Q}_N)f\|_2 \\ &\le \|\mathrm{P}_T\mathrm{Q}_N f\|_2 + \|\mathrm{P}_T(1-\mathrm{Q}_N)f\|_2 \\ &\le \|\mathrm{Q}_N f\|_2 + \|\mathrm{P}_T(1-\mathrm{Q}_N)f\|_2. \quad (80)\end{aligned}$$

Similarly, we have that

$$\|\mathrm{Q}_N f\|_2 \le \|\mathrm{P}_T f\|_2 + \|\mathrm{Q}_N(1-\mathrm{P}_T)f\|_2. \quad (81)$$

Therefore, we can write

$$\begin{aligned}&\sqrt{\frac{T}{N}}\Big(\frac{\|\mathrm{P}_T f\|_2}{\sqrt{T}} - \frac{\|\mathrm{P}_T(1-\mathrm{Q}_N)f\|_2}{\sqrt{T}}\Big) \\ &\quad\le \frac{\|\mathrm{Q}_N f\|_2}{\sqrt{N}} \\ &\quad\le \sqrt{\frac{T}{N}}\Big(\frac{\|\mathrm{P}_T f\|_2}{\sqrt{T}} + \frac{\|\mathrm{Q}_N(1-\mathrm{P}_T)f\|_2}{\sqrt{T}}\Big). \quad (82)\end{aligned}$$

Now, assume that $\delta > 0$ and let $N^+ = (1+\delta)^2 T$ and $N^- = (1-\delta)^2 T$. Using the inequalities in (82), by letting T tend to infinity we obtain that

$$(1-\delta)\limsup_{N\to\infty}\frac{\|\mathrm{Q}_N f\|_2}{\sqrt{N}} \le \lim_{T\to\infty}\frac{\|\mathrm{P}_T f\|_2}{\sqrt{T}} + \limsup_{T\to\infty}\frac{\|\mathrm{Q}_{N^-}(1-\mathrm{P}_T)f\|_2}{\sqrt{T}} \quad (83)$$

and

$$(1+\delta)\limsup_{N\to\infty}\frac{\|\mathrm{Q}_N f\|_2}{\sqrt{N}} \ge \lim_{T\to\infty}\frac{\|\mathrm{P}_T f\|_2}{\sqrt{T}} - \limsup_{T\to\infty}\frac{\|\mathrm{P}_T(1-\mathrm{Q}_{N^+})f\|_2}{\sqrt{T}}. \quad (84)$$

Hence, it is sufficient to prove that, for any $\delta > 0$, we have that

$$\limsup_{T\to\infty}\frac{\|\mathrm{Q}_N(1-\mathrm{P}_T)f\|_2^2}{T} = 0 \quad \text{if} \quad N \le (1-\delta)T \quad (85)$$

and

$$\limsup_{T\to\infty}\frac{\|\mathrm{P}_T(1-\mathrm{Q}_N)f\|_2^2}{T} = 0 \quad \text{if} \quad N \ge (1+\delta)T. \quad (86)$$

Notice that (85), at first hand, yields that (78) is a converging series in $L_2(\mathbb{R})$ and $\mathrm{Q}_N f$ is well-defiend. The reason is that

$$\begin{aligned}&\sum_{|j|\le N}\langle f,\varphi_{0j}\rangle^2 + \sum_{i\ge 0}\sum_{|j|\le 2^i N}\langle f,\psi_{ij}\rangle^2 \\ &\quad= \sum_{|j|\le N}\left(\langle \mathrm{P}_T f,\varphi_{0j}\rangle + \langle(1-\mathrm{P}_T)f,\varphi_{0j}\rangle\right)^2 \\ &\qquad+ \sum_{i\ge 0}\sum_{|j|\le 2^i N}\left(\langle \mathrm{P}_T f,\psi_{ij}\rangle + \langle(1-\mathrm{P}_T)f,\psi_{ij}\rangle\right)^2. \quad (87)\end{aligned}$$

Using the inequality $(a+b)^2 \le 2(a^2+b^2)$, we get

$$\begin{aligned}&\sum_{|j|\le N}\langle f,\varphi_{0j}\rangle^2 + \sum_{i\ge 0}\sum_{|j|\le 2^i N}\langle f,\psi_{ij}\rangle^2 \\ &\quad\le 2\Big(\sum_{|j|\le N}\langle \mathrm{P}_T f,\varphi_{0j}\rangle^2 + \sum_{i\ge 0}\sum_{|j|\le 2^i N}\langle \mathrm{P}_T f,\psi_{ij}\rangle^2\Big) \\ &\qquad+2\Big(\sum_{|j|\le N}\langle(1-\mathrm{P}_T)f,\varphi_{0j}\rangle^2 \\ &\qquad+\sum_{i\ge 0}\sum_{|j|\le 2^i N}\langle(1-\mathrm{P}_T)f,\psi_{ij}\rangle^2\Big) \\ &\quad= 2\|\mathrm{Q}_N\mathrm{P}_T f\|_2^2 + 2\|\mathrm{Q}_N(1-\mathrm{P}_T)f\|_2^2. \quad (88)\end{aligned}$$

Since $|P_T f(x)| \le P_T g(x)$ and $P_T g \in L_2(\mathbb{R})$, $P_T f$ also belongs to $L_2(\mathbb{R})$. This means that the right-hand side of (87) is finite. Consequently, the left-hand side of (87) is finite, too, and thus $\mathrm{Q}_N f$ belongs to $L_2(\mathbb{R})$.

To prove (85), we write

$$\begin{aligned}&\|\mathrm{Q}_N(1-\mathrm{P}_T)f\|_2^2 \\ &\quad= \sum_{|j|\le N}\Big(\int_{|x|>T} f(x)\varphi(x-j)\mathrm{d}x\Big)^2 \\ &\qquad+\sum_{i\ge 0}\sum_{|j|\le 2^i N}\Big(\int_{|x|>T} f(x)2^{\frac{i}{2}}\psi(2^i x-j)\mathrm{d}x\Big)^2 \\ &\quad= \sum_{|j|\le N}\Big(\int_{|x+j|>T} f(x+j)\varphi(x)\mathrm{d}x\Big)^2 \\ &\qquad+\sum_{i\ge 0}\sum_{|j|\le 2^i N}2^{-i}\Big(\int_{|\frac{x+j}{2^i}|>T} f\Big(\frac{x+j}{2^i}\Big)\psi(x)\mathrm{d}x\Big)^2. \quad (89)\end{aligned}$$

Using the facts that $|f(x)| \le g(x)$, $|\varphi(x)|, |\psi(x)| \le \eta(x)$, where g and η are even functions and g is increasing on the positive numbers, along with $T - N \ge \delta T$ and (89), we deduce that

$$\begin{aligned}&\|\mathrm{Q}_N(1-\mathrm{P}_T)f\|_2^2 \\ &\quad\le 4T\Big(\int_{x>\delta T} g(x+T)\eta(x)\mathrm{d}x\Big)^2 \\ &\qquad+\sum_{i\ge 0}2^{-i}\cdot 4\cdot 2^i T\Big(\int_{x>2^i\delta T} g(x+T)\eta(x)\mathrm{d}x\Big)^2 \\ &\quad\le 4T\Big[\Big(\int_{x>\delta T} g\big((1+\frac{1}{\delta})x\big)\eta(x)\mathrm{d}x\Big)^2 \\ &\qquad+\sum_{i\ge 0}\Big(\int_{x>2^i\delta T} g\big((1+\frac{1}{\delta})x\big)\eta(x)\mathrm{d}x\Big)^2\Big]. \quad (90)\end{aligned}$$

However, we know that there exists $c \in \mathbb{R}$ that satisfies $g\big((1+\frac{1}{\delta})x\big)\eta(x) < \frac{c}{x^{1+\epsilon}}$. Thus, we have that

$$\begin{aligned}&\|\mathrm{Q}_N(1-\mathrm{P}_T)f\|_2^2 \\ &\quad\le 4T\Big[\Big(\int_{\delta T}^{\infty}\frac{c}{x^{1+\epsilon}}\mathrm{d}x\Big)^2 + \sum_{i\ge 0}\Big(\int_{2^i\delta T}^{\infty}\frac{c}{x^{1+\epsilon}}\mathrm{d}x\Big)^2\Big] \\ &\quad= 4T\left(\frac{c^2}{\delta^2 T^{2\epsilon}} + \sum_{i\ge 0}\frac{c^2}{2^{2i}\delta^2 T^{2\epsilon}}\right) = c' T^{1-2\epsilon} \quad (91)\end{aligned}$$

with c' independent of T, which completes the proof of (85).

To prove (86), we write

$$\begin{aligned}\|\mathrm{P}_T(1-\mathrm{Q}_N)f\|_2^2 &= \int_{-T}^{T}|(1-\mathrm{Q}_N)\{f\}(x)|^2\mathrm{d}x \\ &\le 2T\Big[\sup_{|x|\le T}|(1-\mathrm{Q}_N)\{f\}(x)|\Big]^2. \quad (92)\end{aligned}$$

We have that

$$\begin{aligned}&(1-\mathrm{Q}_N)\{f\}(x) \\ &= \sum_{|j|>N}\Big(\int_{\mathbb{R}} f(u)\varphi(u-j)\mathrm{d}u\Big)\varphi(x-j) \\ &\quad+\sum_{i\ge 0}\sum_{|j|>2^i N}\Big(\int_{\mathbb{R}} f(u)2^{\frac{i}{2}}\psi(2^i u-j)\mathrm{d}u\Big)2^{\frac{i}{2}}\psi(2^i x-j). \quad (93)\end{aligned}$$

Using $|f(x)| \le g(x)$ and $|\varphi(x)|, |\psi(x)| \le \eta(x)$, and changing the variables of integrations, we get

$$\begin{aligned}|(1-\mathrm{Q}_N)\{f\}(x)| &\le \sum_{|j|>N}\Big(\int_{\mathbb{R}} g(u+j)\eta(u)\mathrm{d}u\Big)\eta(x-j)\\ &\quad+\sum_{i\ge 0}\sum_{|j|>2^iN}\Big(\int_{\mathbb{R}} g\big(\frac{u+j}{2^i}\big)\eta(u)\mathrm{d}u\Big)\eta(2^ix-j). \quad (94)\end{aligned}$$

For $|x| < T$, exploiting the fact that η is an even function and decreasing on positive numbers, along with $N \ge (1+\delta)T$, we write

$$\begin{aligned}|(1-\mathrm{Q}_N)\{f\}(x)| &\le \sum_{|j|>N}\Big(\int_{\mathbb{R}} g(u+j)\eta(u)\mathrm{d}u\Big)\eta(\delta' j)\\ &\quad+\sum_{i\ge 0}\sum_{|j|>2^iN}\Big(\int_{\mathbb{R}} g\big(\frac{u+j}{2^i}\big)\eta(u)\mathrm{d}u\Big)\eta(2^i\delta' j), \quad (95)\end{aligned}$$

where $\delta' = \frac{\delta}{1+\delta}$. We know that there exists $C \in \mathbb{R}$, independent of x and y, that satisfies

$$(x+y)^n \le C(|x|^n + |y|^n). \quad (96)$$

The same C satisfies

$$g(x+y) \le C(g(x)+g(y)). \quad (97)$$

Thus, we write

$$\begin{aligned}|(1-\mathrm{Q}_N)\{f\}(x)| &\le C\sum_{|j|>N}\Big(\int_{\mathbb{R}}(g(u)+g(j))\eta(u)\mathrm{d}u\Big)\eta(\delta' j)\\ &\quad+C\sum_{i\ge 0}\sum_{|j|>2^iN}\Big(\int_{\mathbb{R}}\big(g\big(\frac{u}{2^i}\big)+g\big(\frac{j}{2^i}\big)\big)\eta(u)\mathrm{d}u\Big)\eta(2^i\delta' j). \quad (98)\end{aligned}$$

By expanding the summations, we obtain

$$\begin{aligned}|(1-\mathrm{Q}_N)\{f\}(x)| &\le C\Big(\int_{\mathbb{R}} g(u)\eta(u)\mathrm{d}u\Big)\sum_{|j|>N}\eta(\delta' j)\\ &\quad+C\Big(\int_{\mathbb{R}}\eta(u)\mathrm{d}u\Big)\sum_{|j|>N} g(j)\eta(\delta' j)\\ &\quad+C\sum_{i\ge 0}\Big[\Big(\int_{\mathbb{R}} g\big(\frac{u}{2^i}\big)\eta(u)\mathrm{d}u\Big)\sum_{|j|>N}\eta(2^i\delta' j)\\ &\quad+\Big(\int_{\mathbb{R}}\eta(u)\mathrm{d}u\Big)\sum_{|j|>N} g\big(\frac{j}{2^i}\big)\eta(2^i\delta' j)\Big]. \quad (99)\end{aligned}$$

However, we have that

$$\int_{\mathbb{R}} g\big(\frac{u}{2^i}\big)\eta(u)\mathrm{d}u \le BC\Big(\int_{\mathbb{R}}\eta(u)\mathrm{d}u + \frac{1}{2^{in}}\int_{\mathbb{R}} u^n\eta(u)\mathrm{d}u\Big), \quad (100)$$

where B and C are as in (26) and (96), respectively. Likewise, we have that

$$\begin{aligned}&\int_{|u|>N} g\big(\frac{u}{2^i}\big)\eta(2^i\delta' u)\mathrm{d}u\\ &\le BC\Big(\int_{|u|>N}\eta(2^i\delta' u)\mathrm{d}u + \frac{1}{2^{in}}\int_{|u|>N} u^n\eta(2^i\delta' u)\mathrm{d}u\Big)\\ &\le \frac{BC}{2^i\delta'}\Big(\int_{|u|>2^i\delta' N}\eta(u)\mathrm{d}u + \frac{1}{(2^i\delta')^n}\int_{|u|>2^i\delta' N} u^n\eta(u)\mathrm{d}u\Big)\\ &\le \frac{BC}{2^i\delta'}\Big(\frac{\dot{c}}{(2^i\delta' N)^{n+\epsilon}} + \frac{1}{(2^i\delta')^n}\frac{\ddot{c}}{2^i\delta' N^\epsilon}\Big). \quad (101)\end{aligned}$$

Therefore, using the inequalities in (101) and bounding the summations in (99) by integrals, for a large enough N, we get

$$\begin{aligned}|(1-\mathrm{Q}_N)\{f\}(x)| &\le \frac{c_1}{N^{n+\epsilon}} + \frac{c_2}{N^\epsilon} + \sum_{i=0}^{\infty}\Big(\frac{c_3}{2^{(n+2)i}}\cdot\frac{1}{N^{n+\epsilon}} + \frac{c_4}{2^{(2n+1)i}N^\epsilon}\Big)\\ &\le \frac{c''}{T^\epsilon}, \quad (102)\end{aligned}$$

in which c_1 to c_4 and c'' are constants independent of N and T. This completes the proof of (86) and hence the proof of the theorem.

Appendix C
Proof of Proposition 1

Repeating the proof of (86), we show that

$$\limsup_{N\to\infty}\|\mathrm{P}_T(1-\mathrm{Q}_N)f\|_2^2 = 0 \quad (103)$$

for any fixed $T > 0$. This means that $\mathrm{Q}_N f$ converges to f on $[-T, T]$ almost surely for any T, which completes the proof.

Appendix D
Proof of Theorem 2

1) Assume that $a_1 \ge a_2 \ge 0$. We are going to show that $M(a_1, b) \ge M(a_2, b)$. Let X_1 and X_2 be two SαS random variables with dispersion parameters a_1^α and a_2^α, respectively. Due to the stability of the distribution, we can write $X_1 = X_2 + X_3$, where X_3 is another SαS random variable that is independent of X_2, with dispersion parameter $a_1^\alpha - a_2^\alpha$. If we denote the probability density function of X_i by p_{X_i}, then we have that

$$\begin{aligned}p_{X_1}(x) &= \int p_{X_2}(x-t)p_{X_3}(t)\mathrm{d}t\\ &= \mathbb{E}_{X_3}[p_{X_2}(x-X_3)]. \quad (104)\end{aligned}$$

The MMSE of estimating X_2 given $Y = X_2 + X_3 + Z$ and X_3, where Z is a Gaussian random variable with variance σ^2, is equal to $M(a_2, \sigma)$. Now, using the fact that the MMSE functional is a concave function of the input distribution p_X, we achieve the desired result [48].

2) This is a direct implication of [41, Theorem 11] since the distribution of an SαS random variable is absolutely continuous.

3) For $\alpha = 2$ (Gaussian distribution), we simply have that

$$M(a, \sigma) = \frac{2a^2\sigma^2}{2a^2+\sigma^2}, \quad (105)$$

which directly gives the result. Hence, we assume that $\alpha < 2$. The case $\epsilon \geq \alpha$ is trivial. Thus, we also assume that $\epsilon < \alpha$. The sketch of the proof is that we compute the mean-square error for the estimator

$$T(y) = \begin{cases} 0, & |y| < \Delta \\ y, & |y| \geq \Delta \end{cases} \tag{106}$$

and show that, upon an appropriate choice of Δ, the theorem holds for this estimator. Consequently, it would automatically hold for $M(a, \sigma)$.

Denote the pdf of X by

$$f_a(x) = \frac{1}{a} f\left(\frac{x}{a}\right), \tag{107}$$

where f is the standard SαS pdf with dispersion 1. Also, let

$$F_a(s,t) = \int_s^t f_a(x)\mathrm{d}x. \tag{108}$$

Moreover, let g be Gaussian pdf with mean 0 and variance σ^2 and define

$$G(s,t) = \int_s^t g(x)\mathrm{d}x. \tag{109}$$

Then, we have that

$$\begin{aligned} \mathrm{MSE}(T) = & \ \mathbb{E}_{X,Z}\left[(X - T(X+Z))^2\right] \\ = & \ \mathbb{P}\left(|X+Z| < \Delta\right) \mathbb{E}_{X,Z}\left[X^2 \mid |X+Z| < \Delta\right] \\ & + \mathbb{P}\left(|X+Z| \geq \Delta\right) \mathbb{E}_{X,Z}\left[Z^2 \mid |X+Z| \geq \Delta\right]. \end{aligned} \tag{110}$$

According to Bayes' rule, we can write

$$\begin{aligned} f_a\left(x \middle| |X+Z| < \Delta\right) &= \frac{\mathbb{P}\left(|x+Z| < \Delta\right) f_a(x)}{\mathbb{P}\left(|X+Z| < \Delta\right)} \\ &= \frac{G\left(-\Delta - x, \Delta - x\right) f_a(x)}{\mathbb{P}\left(|X+Z| < \Delta\right)} \end{aligned} \tag{111}$$

and, similarly,

$$g\left(z \middle| |X+Z| \geq \Delta\right) = \frac{\left(1 - F_a\left(-\Delta - z, \Delta - z\right)\right) g(z)}{\mathbb{P}\left(|X+Z| \geq \Delta\right)}. \tag{112}$$

Incorporating (111) and (112) in (110), we obtain

$$\begin{aligned} \mathrm{MSE}(T) = & \int_{\mathbb{R}} x^2 f_a(x) G(-\Delta - x, \Delta - x)\mathrm{d}x \\ & + \int_{\mathbb{R}} z^2 g(z) \left(1 - F_a\left(-\Delta - z, \Delta - z\right)\right) \mathrm{d}z. \end{aligned} \tag{113}$$

We investigate the behavior of the two terms on the right-hand side of (113) separately. For the first term, we know that, for $\alpha < 2$, we have that

$$\lim_{x \to \infty} \frac{f(x)}{x^{-(1+\alpha)}} = C, \tag{114}$$

where C is a positive finite value. Thus, we deduce that

$$x^{\frac{1}{2}+\alpha-\delta} f_a(x) \in L_2(\mathbb{R}) \tag{115}$$

for some positive $\delta < \epsilon$. The Cauchy-Schwartz inequality then yields

$$\begin{aligned} & \int x^2 f_a(x) G(-\Delta - x, \Delta - x)\mathrm{d}x \\ & \quad \leq \left\| x^{\frac{1}{2}+\alpha-\delta} f_a(x) \right\|_2 \cdot \left\| x^{\frac{3}{2}-\alpha+\delta} G(-\Delta - x, \Delta - x) \right\|_2. \end{aligned} \tag{116}$$

Notice that

$$G(-\Delta - x, \Delta - x) \leq \begin{cases} 1, & |x| \leq \Delta \\ \frac{\sigma^2}{\Delta} g(x), & |x| > \Delta \end{cases}. \tag{117}$$

Hence, since $\alpha < 2$, we have that

$$\left\| x^{\frac{3}{2}-\alpha+\delta} G(-\Delta - x, \Delta - x) \right\|_2^2 \leq \frac{\Delta^{4-2\alpha+2\delta}}{2 - \alpha + \delta} + \frac{\sigma}{2\Delta\sqrt{\pi}}. \tag{118}$$

Additionally, we can write

$$\begin{aligned} \left\| x^{\frac{1}{2}+\alpha-\delta} f_a(x) \right\|_2 &= \left(\int x^{1+2\alpha-2\delta} f_a^2(x)\mathrm{d}x \right)^{\frac{1}{2}} \\ &= \left(\int x^{1+2\alpha-2\delta} \frac{1}{a^2} f^2\left(\frac{x}{a}\right)\mathrm{d}x \right)^{\frac{1}{2}} \\ &= a^{\alpha-\delta} \left(\int x^{1+2\alpha-2\delta} f^2(x)\,\mathrm{d}x \right)^{\frac{1}{2}} \\ &= a^{\alpha-\delta} \left\| x^{\frac{1}{2}+\alpha-\delta} f(x) \right\|_2. \end{aligned} \tag{119}$$

Now, we investigate the second term of the right-hand side of (113). We specifically write

$$\begin{aligned} & \int_{\mathbb{R}} z^2 g(z) \left(1 - F_a\left(-\Delta - z, \Delta - z\right)\right) \mathrm{d}z \\ & \quad = \int_{\mathbb{R}} z^2 g(z) \int_{\mathbb{R}} \left(1 - \mathbf{1}_{[-\Delta,\Delta]}(x+z)\right) f_a(x)\mathrm{d}x\mathrm{d}z \\ & \quad = \int_{\mathbb{R}} \int_{\mathbb{R}} \left(1 - \mathbf{1}_{[-\Delta,\Delta]}(x+z)\right) z^2 g(z) f_a(x)\mathrm{d}z\mathrm{d}x \\ & \quad = 2\int_{\Delta}^{\infty} \int_{\mathbb{R}} z^2 g(z) f_a(t-z)\mathrm{d}z\mathrm{d}t, \end{aligned} \tag{120}$$

where we have used the change of variable $t = x + z$. Since both of $t^2 g(t)$ and $f_a(t)$ are symmetric functions that decrease on $t \geq \sigma\sqrt{2}$, we get

$$\begin{aligned} & \int_{\mathbb{R}} z^2 g(z) f_a(t-z)\mathrm{d}z \\ & \quad = \int_{|z| < \frac{t}{2}} z^2 g(z) f_a(t-z)\mathrm{d}z + \int_{|z| \geq \frac{t}{2}} z^2 g(z) f_a(t-z)\mathrm{d}z \\ & \quad \leq f_a(\tfrac{t}{2}) \int_{\mathbb{R}} z^2 g(z)\mathrm{d}z + \frac{t^2}{4} g(\tfrac{t}{2}) \int_{\mathbb{R}} f_a(z)\mathrm{d}z \\ & \quad = \sigma^2 f_a(\tfrac{t}{2}) + \frac{t^2}{4} g(\tfrac{t}{2}) \end{aligned} \tag{121}$$

for $t \geq 2\sigma\sqrt{2}$. Thus, we have that

$$\begin{aligned} & 2\int_{\Delta}^{\infty} \int_{\mathbb{R}} z^2 g(z) f_a(t-z)\mathrm{d}z\mathrm{d}t \\ & \quad \leq 2\sigma^2 \int_{\Delta}^{\infty} f_a(\tfrac{t}{2})\mathrm{d}t + 2\int_{\Delta}^{\infty} \frac{t^2}{4} g(\tfrac{t}{2})\mathrm{d}t \end{aligned} \tag{122}$$

for $\Delta \geq 2\sigma\sqrt{2}$. According to (114), there exists $C' \in \mathbb{R}$ for which

$$f(t) \leq \frac{C'}{t^{1+\alpha}}, \tag{123}$$

and thus

$$f_a(t) \leq a^\alpha \frac{C'}{t^{1+\alpha}}. \tag{124}$$

Also, there exists $A \in \mathbb{R}$ for which

$$t^2 \mathrm{e}^{-\frac{t^2}{2}} \leq A\mathrm{e}^{-t}. \tag{125}$$

Hence, relying on (122), we get

$$\begin{aligned} 2\int_\Delta^\infty \int_{\mathbb{R}} z^2 g(z) f_a(t-z)\mathrm{d}z\mathrm{d}t \\ \leq a^\alpha 2^{2+\alpha} C' \sigma^2 \frac{1+\alpha}{\Delta^\alpha} + 2A\sigma^2\sqrt{\frac{2}{\pi}}\mathrm{e}^{-\frac{\Delta}{2\sigma}}. \end{aligned} \tag{126}$$

Now, incorporating (116) and (126) in (113), we obtain

$$\begin{aligned} \mathrm{MSE}(T) \leq a^{\alpha-\delta} \left\| x^{\frac{1}{2}+\alpha-\delta} f(x) \right\|_2 \left(\frac{\Delta^{4-2\alpha+2\delta}}{4-2\alpha+2\delta} + 2\sigma\sqrt{\pi} \right) \\ + a^\alpha 2^{2+\alpha} C' \sigma^2 \frac{1+\alpha}{\Delta^\alpha} + 2A\sigma^2\sqrt{\frac{2}{\pi}}\mathrm{e}^{-\frac{\Delta}{2\sigma}}. \end{aligned} \tag{127}$$

Finally, setting

$$\Delta = 2\alpha\sigma\,|\log a| \tag{128}$$

completes the proof.

Appendix E
Calculation of the Gradient of Q_α^γ

According to the definition of the gradient, we have that

$$\begin{aligned} \mathrm{D}_u Q_\alpha^\gamma(v) &= \frac{\partial}{\partial\epsilon} Q_\alpha^\gamma(v+\epsilon u)\Big|_{\epsilon=0} \\ &= \int_0^\infty u(\rho)\ \nabla Q_\alpha^\gamma\{v\}(\rho)\mathrm{d}\rho, \end{aligned} \tag{129}$$

where $\mathrm{D}_u Q_\alpha^\gamma(v)$ is the Gâteaux derivative of Q_α^γ at point v in the direction of function u. Using the rules of differentiation of the calculus of variations, we write

$$\begin{aligned} \mathrm{D}_u Q_\alpha^\gamma(v) &= \frac{1}{(2\pi)^\alpha} \int_{\mathbb{R}} \frac{\partial}{\partial\epsilon} \left| \int_{\mathbb{R}} \frac{\mathcal{W}\{v+\epsilon u\}(\omega)}{(\mathrm{j}\omega)^\gamma} \mathrm{e}^{\mathrm{j}\omega x}\mathrm{d}\omega \right|^\alpha \Bigg|_{\epsilon=0} \mathrm{d}x \\ &= \frac{\alpha}{(2\pi)^\alpha} \int_{\mathbb{R}} \lambda\left(\int_{\mathbb{R}} \frac{\mathcal{W}\{v\}(\omega)}{(\mathrm{j}\omega)^\gamma} \mathrm{e}^{\mathrm{j}\omega x}\ \mathrm{d}\omega \right) \\ &\quad \times \int_{\mathbb{R}} \frac{\mathcal{W}'\{v\}(r)\bar{u}(\rho)}{(\mathrm{j}\rho)^\gamma} \mathrm{e}^{\mathrm{j}\rho x}\ \mathrm{d}\rho\ \mathrm{d}x \\ &= \frac{\alpha}{(2\pi)^\alpha} \int_{\mathbb{R}} \bar{u}(\rho) \frac{\mathcal{W}'\{v\}(\rho)}{(\mathrm{j}\rho)^\gamma} \\ &\quad \times \int_{\mathbb{R}} \lambda\left(\int_{\mathbb{R}} \frac{\mathcal{W}\{v\}(\omega)}{(\mathrm{j}\omega)^\gamma} \mathrm{e}^{\mathrm{j}\omega x}\ \mathrm{d}\omega \right) \mathrm{e}^{\mathrm{j}\rho x}\ \mathrm{d}x\ \mathrm{d}\rho \end{aligned} \tag{130}$$

in which $\lambda(x) = \mathrm{sgn}(x)|x|^{\alpha-1}$,

$$\mathcal{W}'\{v\}(\rho) = \frac{\pi}{2} \times \begin{cases} \cos(\frac{\pi}{2} v(\frac{3}{2\pi}|\rho| - 1)), & \frac{2\pi}{3} < |\rho| \leq \frac{4\pi}{3} \\ -\sin(\frac{\pi}{2} v(\frac{3}{4\pi}|\rho| - 1)), & \frac{4\pi}{3} < |\rho| \leq \frac{8\pi}{3} \\ 0, & \text{otherwise} \end{cases} \tag{131}$$

and

$$\bar{u}(\rho) = \begin{cases} u(\frac{3}{2\pi}\rho - 1), & \frac{2\pi}{3} < |\rho| \leq \frac{4\pi}{3} \\ u(\frac{3}{4\pi}\rho - 1), & \frac{4\pi}{3} < |\rho| \leq \frac{8\pi}{3} \end{cases}. \tag{132}$$

By breaking the outer integral of (130) into two integrals on $[\frac{2\pi}{3}, \frac{4\pi}{3}]$ and $[\frac{4\pi}{3}, \frac{8\pi}{3}]$, and by changing the variables of integrations, we get

$$\begin{aligned} \mathrm{D}_u Q_\alpha^\gamma(v) = \frac{(2\pi)^{1-\alpha}\alpha}{3} \int_0^1 u(\rho) \frac{\mathcal{W}'\{v\}(\frac{2\pi}{3}(\rho+1))}{(\mathrm{j}\frac{2\pi}{3}(\rho+1))^\gamma} \\ \times \int_{\mathbb{R}} \lambda\left(\int_{\mathbb{R}} \frac{\mathcal{W}\{v\}(\omega)}{(\mathrm{j}\omega)^\gamma} \mathrm{e}^{\mathrm{j}\omega x}\ \mathrm{d}\omega \right) \mathrm{e}^{\mathrm{j}\frac{2\pi}{3}(\rho+1)x}\ \mathrm{d}x\ \mathrm{d}\rho \\ + \frac{2(2\pi)^{1-\alpha}\alpha}{3} \int_0^1 u(\rho) \frac{\mathcal{W}'\{v\}(\frac{4\pi}{3}(\rho+1))}{(\mathrm{j}\frac{4\pi}{3}(\rho+1))^\gamma} \\ \times \int_{\mathbb{R}} \lambda\left(\int_{\mathbb{R}} \frac{\mathcal{W}\{v\}(\omega)}{(\mathrm{j}\omega)^\gamma} \mathrm{e}^{\mathrm{j}\omega x}\mathrm{d}\omega \right) \mathrm{e}^{\mathrm{j}\frac{4\pi}{3}(\rho+1)x}\ \mathrm{d}x\ \mathrm{d}\rho. \end{aligned} \tag{133}$$

Now, incorporating (58), (131) and some algebra, according to (129), we obtain (60).

Appendix F
Calculation of the Projection Operator $\mathcal{P}$

For a given function $\tilde{v}(\rho)$, define the two functionals

$$J(v) = \|v(\rho) - \tilde{v}(\rho)\|_2^2 \tag{134}$$

and

$$\bar{J}(v) = \frac{1}{2}\left(\|v(\rho) - \tilde{v}(\rho)\|_2^2 + \|(1 - v(1-\rho)) - \tilde{v}(\rho)\|_2^2 \right) \tag{135}$$

We are interested in

$$\mathcal{P}\{\tilde{v}\} = \mathrm{argmin}\ J(v) \tag{136}$$

subject to the constraint

$$v(x) + v(1-x) = 1. \tag{137}$$

Notice that $J(v) = \bar{J}(v)$ for any function v that satisfies (137). Also, notice that for any function v, $\bar{J}(v(\rho)) = \bar{J}(v(1-\rho))$. Since $\bar{J}(v)$ is strictly convex, it has a unique minimizer. Thus, the minimizer of $\bar{J}(v)$ satisfies (137). Hence, $\mathcal{P}\{\tilde{v}\}$ is the unconstrained minimizer of $\bar{J}(v)$.

To find the minimizer of $\bar{J}(v)$, we set its gradient to zero. According to the calculus of variation, the gradient of $\bar{J}(v)$ is

$$\nabla\bar{J}\{v\}(\rho) = (v(\rho) - \tilde{v}(\rho)) - (1 - v(\rho) - \tilde{v}(1-\rho)). \tag{138}$$

By solving $\nabla\bar{J}\{v\} = 0$, we obtain (61).

Appendix G
Proof of Proposition 2

For $\alpha = 2$, using Parseval, we can directly express Q_2^γ in the Fourier domain as

$$Q_2^\gamma(v) = \int_{\mathbb{R}} \left| \frac{\mathcal{W}\{v\}(\omega)}{(\mathrm{j}\omega)^\gamma} \right|^2 \mathrm{d}\omega. \tag{139}$$

Based on (58), we get

$$Q_2^{\gamma}(\upsilon) = 2\int_{\frac{2\pi}{3}}^{\frac{4\pi}{3}} \frac{\sin^2\left(\frac{\pi}{2}\upsilon(\frac{3}{2\pi}\omega-1)\right)}{\omega^{2\gamma}} + 2\int_{\frac{4\pi}{3}}^{\frac{8\pi}{3}} \frac{\cos^2\left(\frac{\pi}{2}\upsilon(\frac{3}{4\pi}\omega-1)\right)}{\omega^{2\gamma}}\mathrm{d}\omega. \quad (140)$$

Performing ordinary algebraic manipulations, (140) leads to

$$Q_2^{\gamma}(\upsilon) = = 2\Big((\frac{2\pi}{3})^{1-2\gamma} - (\frac{4\pi}{3})^{1-2\gamma}\Big)\int_0^1 \frac{\sin^2\left(\frac{\pi}{2}\upsilon(\rho)\right)}{(\rho+1)^{2\gamma}}\mathrm{d}\rho + \frac{2}{2\gamma-1}\Big((\frac{4\pi}{3})^{1-2\gamma} - (\frac{8\pi}{3})^{1-2\gamma}\Big). \quad (141)$$

Now, using (57), we get

$$Q_2^{\gamma}(\upsilon) = 2\Big((\frac{2\pi}{3})^{1-2\gamma} - (\frac{4\pi}{3})^{1-2\gamma}\Big) \times \int_0^{\frac{1}{2}} \sin^2(\frac{\pi}{2}\upsilon(\rho))\Big(\frac{1}{(\rho+1)^{2\gamma}} - \frac{1}{(2-\rho)^{2\gamma}}\Big)\mathrm{d}\rho + \frac{2}{2\gamma-1}\Big((\frac{\pi}{2})^{1-2\gamma} - \pi^{1-2\gamma}\Big). \quad (142)$$

However, since $(\rho+1)^{-2\gamma} - (2-\rho)^{-2\gamma}$ is positive on $[0, \frac{1}{2}]$, the function $\upsilon(\rho)$ that vanishes on this interval minimizes $Q_2^{\gamma}(\upsilon)$. Consequently, we obtain $\upsilon(\rho) = 0$ on $[0, \frac{1}{2}]$, and $\upsilon(\rho) = 1$ on $(\frac{1}{2}, 1]$ as the minimizer of $Q_2^{\gamma}(\upsilon)$.

REFERENCES

[1] B. B. Mandelbrot and J. W. Van Ness, "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, vol. 10, no. 4, pp. 422–437, Oct. 1968.

[2] M. Unser and P. D. Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge, U.K.: Cambridge Univ. Press, 2014.

[3] T. Lundahl, W. J. Ohley, S. M. Kay, and R. Siffert, "Fractional Brownian motion: A maximum likelihood estimator and its application to image texture," *IEEE Trans. Med. Imag.*, vol. 5, no. 3, pp. 152–161, Sep. 1986.

[4] B. Pesquet-Popescu and J. L. Véhel, "Stochastic fractal models for image processing," *IEEE Signal Process. Mag.*, vol. 19, no. 5, pp. 48–62, Sep. 2002.

[5] K. Park and W. Willinger, *Self-Similar Network Traffic and Performance Evaluation*. Hoboken, NJ, USA: Wiley, 2000.

[6] R. J. Elliott and J. Van Der Hoek, "Fractional Brownian motion and financial modelling," in *Mathematical Finance*. Basel, Switzerland: Birkhäuser, 2001, pp. 140–151.

[7] E. Perrin, R. Harba, C. Berzin-Joseph, I. Iribarren, and A. Bonami, "Nth-order fractional Brownian motion and fractional Gaussian noises," *IEEE Trans. Signal Process.*, vol. 49, no. 5, pp. 1049–1059, May 2001.

[8] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes*. London, U.K.: Chapman & Hall, 1994.

[9] P. Embrechts and M. Maejima, "An introduction to the theory of self-similar stochastic processes," *Int. J. Modern Phys. B*, vol. 14, nos. 12–13, pp. 1399–1420, May 2000.

[10] J. L. Véhel and R. Riedi, "Fractional Brownian motion and data traffic modeling: The other end of the spectrum," in *Fractals in Engineering*. London, U.K.: Springer, 1997, pp. 185–202.

[11] N. Laskin, I. Lambadaris, F. C. Harmantzis, and M. Devetsikiotis, "Fractional Lévy motion and its application to network traffic modeling," *Comput. Netw.*, vol. 40, no. 3, pp. 363–375, Oct. 2002.

[12] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*. New York, NY, USA: Wiley, 1995.

[13] S. Jespersen, R. Metzler, and H. C. Fogedby, "Lévy flights in external force fields: Langevin and fractional Fokker-Planck equations and their solutions," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 59, no. 3, p. 2736, 1999.

[14] M. M. Meerschaert, D. A. Benson, H.-P. Scheffler, and B. Baeumer, "Stochastic solution of space-time fractional diffusion equations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 65, no. 4, p. 041103, 2002.

[15] S. Painter and L. Paterson, "Fractional Lévy motion as a model for spatial variability in sedimentary rock," *Geophys. Res. Lett.*, vol. 21, no. 25, pp. 2857–2860, 1994.

[16] H.-H. Liu, J. Rutqvist, Q. Zhou, and G. S. Bodvarsson, 'Upscaling of normal stress-permeability relationships for fracture networks obeying fractional Levy motion," *Elsevier Geo-Eng. Book Ser.*, vol. 2, pp. 263–268, 2004.

[17] S. Painter, "Stochastic interpolation of aquifer properties using fractional Lévy motion," *Water Resour. Res.*, vol. 32, no. 5, pp. 1323–1332, 1996.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 910–917, Mar. 1992.

[23] G. W. Wornell, "Wavelet-based representations for the $1/f$ family of fractal processes," *Proc. IEEE*, vol. 81, no. 10, pp. 1428–1450, Oct. 1993.

[24] L. Delbeke and P. Abry, "Stochastic integral representation and properties of the wavelet coefficients of linear fractional stable motion," *Stochastic Process. Their Appl.*, vol. 86, no. 2, pp. 177–182, Apr. 2000.

[25] A. Ayache, F. Roueff, and Y. Xiao, "Linear fractional stable sheets: Wavelet expansion and sample path properties," *Stochastic Process. Their Appl.*, vol. 119, no. 4, pp. 1168–1197, Apr. 2009.

[26] B. Pesquet-Popescu, "Statistical properties of the wavelet decomposition of certain non-Gaussian self-similar processes," *Signal Process.*, vol. 75, no. 3, pp. 303–322, Jun. 1999.

[27] P. Pad and M. Unser, "Optimality of operator-like wavelets for representing sparse AR(1) processes," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4827–4837, Sep. 2015.

[28] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 205–220, Apr. 1992.

[29] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Trans. Image Process.*, vol. 1, no. 2, pp. 244–250, Apr. 1992.

[30] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 959–971, Oct. 1996.

[31] R. H. Chan, T. F. Chan, L. Shen, and Z. Shen, "Wavelet algorithms for high-resolution image reconstruction," *SIAM J. Sci. Comput.*, vol. 24, no. 4, pp. 1408–1432, Feb. 2003.

[32] B. K. Alsberg, A. M. Woodward, M. K. Winson, J. Rowland, and D. B. Kell, "Wavelet denoising of infrared spectra," *Analyst*, vol. 122, no. 7, pp. 645–652, Jul. 1997.

[33] M. K. Mihçak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 300–303, Dec. 1999.

[34] A. Achim, P. Tsakalides, and A. Bezerianos, "SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 8, pp. 1773–1784, Aug. 2003.

[35] R. A. DeVore, "Nonlinear approximation," *Acta Numer.*, vol. 7, pp. 51–150, Jan. 1998.

[36] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[37] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[38] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Trans. Med. Imag.*, vol. 20, no. 8, pp. 772–783, Aug. 2001.

[39] L. Boubchir and J. M. Fadili, "A closed-form nonparametric Bayesian estimator in the wavelet domain of images using an approximate α-stable prior," *Pattern Recognit. Lett.*, vol. 27, no. 12, pp. 1370–1382, Sep. 2006.

[40] Y. Meyer, "Principe d'incertitude, bases hilbertiennes et algebres d'operateurs," *Séminaire Bourbaki*, vol. 38, pp. 209–223, 1986.

[41] Y. Wu and S. Verdu, "MMSE dimension," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4857–4879, Aug. 2011.
[42] M. Unser, P. D. Tafti, and Q. Sun, "A unified formulation of Gaussian versus sparse stochastic processes—Part I: Continuous-domain theory," *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1945–1962, Mar. 2014.
[43] M. Unser, P. D. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian versus sparse stochastic processes—Part II: Discrete-domain theory," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 3036–3051, May 2014.
[44] K. Sato, *Lévy Processes and Infinite Divisibility*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
[45] A. Ayache, F. Roueff, and Y. Xiao, "Local and asymptotic properties of linear fractional stable sheets," *Comp. Rendus Math.*, vol. 344, no. 6, pp. 389–394, Mar. 2007.
[46] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 9, no. 6, pp. 1135–1151, 1981.
[47] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, Apr. 2005.
[48] Y. Wu and S. Verdú, "Functional properties of minimum mean-square error and mutual information," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1289–1301, Mar. 2012.
[49] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA, USA: SIAM, 1992.
[50] M. Unser, "On the optimality of ideal filters for pyramid and wavelet signal approximation," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3591–3596, Dec. 1993.
[51] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, no. 7, pp. 909–996, 1988.
[52] S. Held, M. Storath, P. Massopust, and B. Forster, "Steerable wavelet frames based on the riesz transform," *IEEE Trans. Image Process.*, vol. 19, no. 3, pp. 653–667, Mar. 2010.
[53] A. Haar, "Zur theorie der orthogonalen funktionensysteme," *Math. Ann.*, vol. 69, no. 3, pp. 331–371, Sep. 1910.

Pedram Pad (S'08) received the B.Sc. and M.Sc. degrees in electrical engineering (communications and signal processing) in 2009 and 2011, respectively, and the B.Sc. degree in mathematical sciences (pure mathematics) in 2009, all from Sharif University of Technology (SUT), Tehran, Iran. Since December 2011, he has been pursuing the Ph.D. degree with the Biomedical Imaging Group (BIG), Ecole Polytechnique Fedrale de Lausanne (EPFL), Lausanne, Switzerland. His research interests include different aspects of information theory and signal processing, specially applications of the former in the latter.

Kasra Alishahi received the B.Sc., M.Sc., and Ph.D. degrees all from the Department of Mathematical Sciences, Sharif University of Technology, Iran, in 2000, 2002, and 2008, respectively. He is currently an Assistant Professor of the Department of Mathematical Sciences at Sharif University of Technology. His research interests are stochastic processes and statistics.

Michael Unser (M'89–SM'94–F'99) is professor and director of EPFL's Biomedical Imaging Group, Lausanne, Switzerland. His primary area of investigation is biomedical image processing. He is internationally recognized for his research contributions to sampling theory, wavelets, the use of splines for image processing, stochastic processes, and computational bioimaging. He has published over 250 journal papers on those topics. He is the author with P. Tafti of the book "An introduction to sparse stochastic processes", Cambridge University Press 2014. From 1985 to 1997, he was with the Biomedical Engineering and Instrumentation Program, National Institutes of Health, Bethesda USA, conducting research on bioimaging. Dr. Unser has held the position of associate Editor-in-Chief (2003-2005) for the IEEE Transactions on Medical Imaging. He is currently member of the editorial boards of SIAM J. Imaging Sciences, IEEE J. Selected Topics in Signal Processing, and Foundations and Trends in Signal Processing. He is the founding chair of the technical committee on Bio Imaging and Signal Processing (BISP) of the IEEE Signal Processing Society. Prof. Unser is a fellow of the IEEE (1999), an EURASIP fellow (2009), and a member of the Swiss Academy of Engineering Sciences. He is the recipient of several international prizes including three IEEE-SPS Best Paper Awards and two Technical Achievement Awards from the IEEE (2008 SPS and EMBS 2010).