Optimal Representations of Sparse Stochastic Processes with Applications in Image Processing

Pedram Pad

Abstract

We establish in the world of stochastic processes a theoretical relation between sparsity and wavelets. The underlying principle is to treat stochastic processes as generalized functions, which facilitates the study of their properties in a transform domain. We focus on symmetric-α-stable (SαS) processes, with $\alpha \in (0, 2]$. They are central to a recently proposed framework for sparse stochastic processes. The case $0 < \alpha < 2$ corresponds to heavy-tail processes and, thus, to sparse signals with the level of sparsity being more pronounced for smaller values of α. The limit case $\alpha = 2$ yields the classical Gaussian processes, which are not considered to be sparse.

In the first part of the thesis, we identify a particular class of wavelets and show that they provide an independent-component analysis for SαS AR(1) signals, which form an important subclass of sparse processes. This contribution can be considered as the first theoretical result ever found about the optimality of wavelets in a stochastic framework. It is remarkable that, by decreasing the parameter α, we continuously run the gamut from Fourier (for $\alpha = 2$) to wavelets (for $\alpha \leq 1$), thus providing a unifying view that sees Fourier and wavelet transforms as two extremes of a single spectrum.

In the second part of the thesis, we study the wavelet decomposition of self-similar processes—which form a second important family of sparse SαS processes. In particular, in the context of the wavelet-based denoising of a stochastic process, we are able to theoretically predict the exact performance of an arbitrary orthonormal wavelet basis. As it turns out, our prediction takes a tractable simple form, which allows us to design the optimal Meyer wavelet as an example of application.

Our theoretical results suggest that the performance depends crucially on the localization of wavelets. Consequently, in an image-processing context, we introduce a moment-based measure of localization and propose an optimization framework to design better wavelets. We end the thesis with the experimental confirmation that the proposed wavelets

outperform the previously existing ones.

Résumé

Dans notre travail, nous établissons une relation entre la parcimonie et les ondelettes dans le monde des processus stochastiques. Nous considérons les processus stochastiques comme des fonctions généralisées, ce qui facilite leur étude dans un domain de transformation. Nous nous concentrons sur des processus symétriques et α-stable (SαS), avec $0 < \alpha \leq 2$, qui sont au coeur de la récente théorie des processus stochastiques parcimonieux. Pour $0 < \alpha < 2$, les processus étudiés sont à queue épaisse, et correspondent donc à des signaux parcimonieux dont le niveau de parcimonie s'accroît lorsque α diminue. Dans le cas limite $\alpha = 2$, on retrouve les processus Gaussien, qui ne sont traditionnellement pas considérés comme parcimonieux.

Dans la première partie de notre thèse, nous identifions une classe d'ondelettes et montrons qu'elles permettent une analyse en composantes indépandantes d'un signal SαS correspondant au modèle AR(1). Cette contribution est le premier résultat théorique jamais obtenu sur l'optimalité des ondelettes pour l'analyse de modèles stochastiques. En réduisant le paramètre α, il est remarquable de constater le passage continu de la transformée de Fourier, optimale pour $\alpha = 2$, aux ondelettes, optimales pour $\alpha \leq 1$. Nous offrons ainsi une vue unifée où les transformées de Fourier et en ondelettes sont les deux extrèmes d'un spectre unique.

Dans la seconde partie de la thèse, nous étudions la décomposition en ondelettes des processes auto-similaires, qui sont une seconde famille importante des processus SαS parcimonieux. En particulier, nous sommes capable de prédire la performance exacte d'une base arbitraire d'ondelettes orthonormales pour le débruitage d'un tel processus stochastique. Notre prediction prend une forme particulièrement maniable. Ceci nous permet de construire l'ondelette de Meyer optimale comme application de notre résultat.

Nos résultats théoriques suggèrent que la performance d'une ondelette dépend fortement de ses propriétés de localisation. Dans le cadre du traitement de l'image, nous intro-

duisons de ce fait une mesure de cette localisation basée sur les moments de l'ondelette et proposons une méthode d'optimisation adaptée à la construction de meilleures ondelettes. La thèse s'achève sur des expériences confirmant que nos nouvelles ondelettes supplantent celles existantes.

Mots-clefs : Processsus stochastique parcimonieux, SαS, processus auto-régressifs, processus auto-similaires, analyse en composantes indépendantes, divergence de Kullback-Leibler, optimisation en dimension infinie, opérateurs adapté à une transformation en ondelettes, transformée en ondelette isotropique, pyramide orientable, calcul des variations, traitement de l'image

To my mother, father and wife
for all their supports and sacrifices for me

Acknowledgement

This thesis would not have been written without the help and encouragement of many people. I take this opportunity to highlight some of them.

First and foremost, I would like to express my gratitude to my PhD advisor, Prof. Michaël Unser. His true researcher attitude, which is illustrated in being very receptive to new ideas, his passion for science, which results in being very knowledgeable and hard-working scientist, and his sense of responsibility, which results in easy reachability make him an ideal supervisor. Also, I would like to thank my supervisor during the bachelor and master periods at Sharif University of Technology, Prof. Farokh Marvasti, to whom I'm especially in debt for the very first encouragement and support that started my path through the world of research.

I express my sincere thanks to the president of the thesis committee, Prof. Pascal Frossard, and the official members, Prof. Helmut Bölcskei, Prof. Patrick Flandrin and Porf. Dimitri Van de Ville, for accepting to review my manuscript and coming to Lausanne for the defence.

I thank all my friends and collaborators during all these years. To name a few, I thank Dr. Philippe Thevenaz for his careful proofreading of my papers and other manuscripts, and Dr. Daniel Sage for his flawless management of computer related issues in the lab. I thank all of my BIG (Biomedical Imaging Group) fellows especially Dr. Emrah Bostan, Dr. John Paul Ward, Dr. Masih Nilchian, Dr. Ulugbek Kamilov, Dr. Michael McCann, Virginie Uhlmann, Dr. Zsuzsanna Puspoki, Julien Fageot, Prof. Adrien Depeursinge, Daniel Schmitter, Laurene Donati and Anais Badoual. I would also like to thank some of my Sharif friends, especially Prof. Kasra Alishahi, Prof. Arash Amini, Prof. Adel Javanmard, Prof. Masoud Babaeizadeh, Pouria Kabir, Seyyed Mohammad Mansouri, Shayan

Contents

Chapter 1

Introduction

Stochastic signal modeling and *transform-domain signal processing* are tools that pervade all branches of signal processing. On one hand, stochastic processes are the prime candidate to model practical phenomena since they view real signals or operations only through their statistics. This offers some slack over the actual observed data, whereas the deterministic view of data invariably fails to fulfill the idealized assumptions placed upon them [1]. On the other hand, the analysis of most signals is best performed in a transform domain, which reveals more about their intimate structure than could ever be found in the original domain. In the intersection of these two fields, the investigation of transform-domain properties of stochastic processes has naturally emerged in the literature as a fecund approach.

Classical statistical signal processing is predicated upon the notion that processes are Gaussian. This goes hand-in-hand with a Fourier-transform representation of data, which is widely used in classical signal processing, the Fourier transform being often the most appropriate choice to represent Gaussian processes. (Similar observations hold for general finite-variance processes up to second-order statistics [2, 3].) Unfortunately, practical signals often do not exhibit the *regular* behavior of Gaussian processes; moreover, *global* transformations like Fourier are not necessarily the most appropriate match for them. Therefore, modern signal processing focuses instead on the concepts of sparsity for signal modeling [4] and on localized transformations like wavelets as the representation domain [5].

In the late 1980s, the advent of the *wavelet transform* has opened new avenues of research. It is a prominent component of modern signal processing that has replaced the

Fourier transform in many different applications such as signal reconstruction and compression [6, 7, 8]. As further development, new redundant and non-redundant representations of signals (local cosines, curvelets) have emerged during the past two decades and have led to better algorithms for data compression, processing, and feature extraction. An important example is the joint photographic experts group (JPEG) 2000 standard for image compression which gives almost 30% better performance over the older discrete cosine transform (DCT)-based JPEG method. However, because they were abandoning Fourier, researchers had also to progressively abandon the modeling of signals as stationary Gaussian stochastic processes, adopting instead a more deterministic approximation-theoretic point of view. As it turns out, many naturally occurring signals and images happen to be experimentally sparse in some wavelet basis [9], while a Gaussian process never is.

Since the early 2000s, the different fields of signal processing have seen a tremendous amount of research been conducted around the concept of sparsity [10]. Loosely said, a signal is said to be sparse whenever its representation in an appropriate domain has only a small number of significant values. Signals that have a sparse representation in some transform domain are easy to compress; similarly, a simple pointwise processing will be highly successful at attenuating their noise. The concept of sparsity has penetrated deep into both the stochastic and the deterministic signal-processing frameworks. In the stochastic framework, studies such as [11] and [12] give mathematically rigorous meaning to sparsity. In the deterministic framework, sparsity is also a proven property of a particular class of signals [13].

Attempts have been made to connect wavelets with sparsity. Following an empirical approach, the authors of [9] and [14] have suggested that some wavelet-like bases tend to be optimal to control sparsity and for decoupling purposes. Following an approximation-theoretic approach (without combining it with a stochastic one), the authors of [13] have demonstrated the theoretical optimality of wavelets. More generally, many other researchers have actively sought to propose criteria to design and optimize wavelets during the last two decades, but most of these criteria are driven from approximation-theoretic considerations rather than statistics. In fact, despite all these efforts, the connection between the concept of wavelets and that of sparsity has not been well established, at least not from the *statistical* signal-processing point of view—even if the statistical behavior of wavelets was nevertheless studied in a few rare cases [15].

The main reason for this shortage of statistics-related results could be stated to be the lack of a good stochastic framework for modeling sparse signals. Fortunately, a timely innovation model has been introduced recently in [16, 17, 18]. It is capable of modeling a wide range of stochastic processes, from Gaussian to sparse, and brings to the signal-

processing literature the overly abstract stochastic model originally proposed by Russian mathematicians [19]. Indeed, the new framework is particularly convenient for investigating the transform-domain representations of stochastic processes.

Relying on this framework, the general aim of the thesis is to investigate the link between wavelets and sparsity—from a stochastic point of view. In particular, we harness the information-theoretic properties of the wavelet coefficients of sparse stochastic processes to build, for the very first time, a mathematical model in which we can demonstrate the fitness of wavelets to sparse stochastic processes. Then, we turn our attention to the characterization of different wavelet bases for working with sparse stochastic processes. Finally, we put out theoretical musings to fruition by designing new classes of wavelets that outperform the ones previously known in the literature, a claim that we are able to validate with practical experiments. The summary of our thesis is told in more details in Section 1.1.

1.1   Outline and Contributions of the Thesis

In this section, we give a brief hierarchical outline of the contributions of the thesis. We put forward the main result of each chapter and highlight a few select intermediate results as well.

Chapter 2 contains preliminaries on the fundamental concepts of the thesis. There, we describe the innovation model of sparse stochastic processes and the symmetric-α-stable processes. Throughout the thesis, these concepts form the basis of our stochastic analysis. Then, we present the theoretical and practical definitions and properties that apply to conventional and operator-like wavelet bases and frames.

The main body of the thesis starts in earnest after the preliminaries of Chapter 2. It consists of two major parts, namely, *theory* and *applications*. The part on theory is contained in Chapters 3 and 4; the part on applications is contained in Chapter 5, where we use our theoretical results to design wavelets that perform optimally in an image-processing context.

We start Chapter 3 by investigating the optimality of wavelets for decoupling stochastic processes. We show, in particular, that operator-like wavelets are optimal to represent a certain class of sparse stochastic processes. When we switch to other classes of sparse stochastic processes, however, optimality may be lost. But, in Chapter 4, we are still able to distinguish between wavelet bases on the ground of their performance to attenuate noise for some particular class of sparse stochastic processes.

In the part on theory, we essentially consider two classes of stochastic processes: 1) autoregressive processes; and 2) self-similar processes. They are of special importance in statistical signal processing, both from a theoretical and a practical point of view. They can be considered as two separate extensions of the Lévy processes, which are at the heart of the theory of continuous-domain stochastic processes [20].

Autoregressive processes are related to Lévy processes through Markov processes. Indeed, the class of autoregressive processes of first order (AR(1)) identifies with the class of Markov processes, which are such that the knowledge of a sample of the process is enough to decouple past and future—in the sense that it makes them independent. As it turns out, the Lévy processes *are* Markov processes. Likewise, self-similar processes are related to Lévy processes, too. Indeed, an important property of Lévy processes is their self-similarity. More precisely, they have a fractal property of first order.

All excitation noises that we consider for these processes have an α-stable distribution. This is a rich family of distributions that runs the gamut from Gaussian distributions to heavy-tail (and thus, sparse) distributions. This enables us to roam between the classical and modern schools of signal processing simply by sweeping over α.

Our theoretical results are as follows:

- **Proof of the Optimality of Operator-Like Wavelets for Decoupling α-Stable AR(1) Processes**

  - Our result on wavelets must be contrasted with the situation where the transform domain is Fourier, in which case finite-variance processes must be considered instead of α-stable AR(1) ones.

  - We derive a closed-form formula for the mutual information of entries of a vast class of α-stable random vectors. Interestingly, our formula involves ℓ_α-norm calculations that are in common use in the context of sparse-recovery algorithms.

  - We relate the sparse-component analysis (SCA) to the independent-component analysis (ICA). This is of utmost relevance from an information-theoretic point of view.

- **Gauge of the Total Performance of a Wavelet when Attenuating the Noise of α-Stable Self-Similar Processes**

  - We show how to determine the time-averaged energy of a signal through its wavelet coefficients. It is noteworthy to observe that we deal with the wavelet

decomposition of signals that have a finite *average energy*, rather that finite-energy signals.

– We study the behavior of minimum mean-square error (MMSE) functions for infinite-variance random variables when the width parameter tends to zero. In particular, we extend to the infinite-variance case some existing results in the literature about finite-variance random variables.

– We derive the optimal Meyer wavelet for attenuating the noise found in α-stable self-similar processes [21].

– We obtain closed-form formulæ for the projector of a profile onto the infinite-dimensional manifold of Meyer wavelets.

The part on applications revolves around solving practical problems found in the context of image processing. More precisely, we consider the two wavelet-based tasks of 1) attenuating the noise in images and 2) reconstructing images. Inspired by our theoretical results, we design new wavelets tuned to each one of the considered applications. Our experiments confirm that the wavelets we designed outperform previously existing ones.

Our practical results are as follows:

- **Numerical Design of Maximally Localized Isotropic Tight Wavelet Frames**

  – We derive a novel closed-form formula for the projector of isotropic wavelets onto the infinite-dimensional manifold of tight frames.

- **Illustration of the Superiority of the Proposed Localized Wavelets Through Practical Experiments**

We conclude the thesis by summarizing our results and suggesting topics for future studies.

Chapter 2

Mathematical Preliminaries

There are two major theories that constitute the formulation of this thesis. The first one is the theory of sparse stochastic processes and the second one is the theory of wavelets. To have a self-contained text, in this section we give a brief introduction about these two major topics. Before that, we present the notations that we will use throughout the text.

2.1 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 from $\mathbb{R}^n$ to $\mathbb{R}$ is denoted by $\hat{f}$. Thus, we have

$$\hat{f}(\boldsymbol{\omega}) = \int_{\mathbb{R}} f(\mathbf{x}) \mathrm{e}^{-\mathrm{j}\langle \boldsymbol{\omega}, \mathbf{x} \rangle} \mathrm{d}\mathbf{x} \tag{2.1}$$

and

$$f(\mathbf{x}) = \frac{1}{(2\pi)^n} \int_{\mathbb{R}} \hat{f}(\boldsymbol{\omega}) \mathrm{e}^{\mathrm{j}\langle \boldsymbol{\omega}, \mathbf{x} \rangle} \mathrm{d}\boldsymbol{\omega} \tag{2.2}$$

with $\mathbf{x} = (x_1, \cdots, x_n) \in \mathbb{R}^n$ and $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_n) \in \mathbb{R}^n$ being the space and the frequency domain variables, respectively. Also, $\|f\|_{\alpha}$ represents the α-(pseudo)norm of f for any positive value α. The notation $\langle \cdot, \cdot \rangle$ is the standard inner product of two vectors, functions or a linear functional with a test function.

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.

2.2 Sparse Stochastic Processes

To define a stochastic process, several different approaches have been proposed up to now. Each approach has its own advantages and limitations. Most of the conventional approaches are based on constructing the sample paths of the processes. Although this constructions give us intuitions about how a stochastic process looks like, they have technical difficulties for certain signal processing procedures. For examples, taking a signal into a transformation domain is very common in signal processing. But, with these constructions of stochastic processes, it is usually complicated to characterize the transform domain properties of the stochastic processes.

In [19], the authors proposed a framework for defining stochastic processes that are very convenient for studying transform domain operations. Although it is a relatively old framework, it did not attract much attention within the signal processing community due to its mathematical technicalities. Recently, the authors of [17, 18, 22], reinterpreted this theory into the conventional language of signal processing which is now very handy to work with. In the following subsections, we give a brief preliminary about this framework.

2.2.1 Innovation Model

In [17], the authors define a sparse stochastic process s as the solution of the linear differential equation

$$\mathrm{L}s = w \tag{2.3}$$

where L is a suitable differential operator and w a Gaussian/non-Gaussian continuous-domain white noise (or innovation process). However, the more interesting noises are the non-Gaussian ones which are actually the source of the adjective "sparse" in the name of these processes. Formally, (2.3) results into the solution

$$s = \mathrm{L}^{-1}w \tag{2.4}$$

where the linear operator L^{-1} is the inverse of the whitening operator L, which also indicates that a sparse stochastic process is a filtered version of a non-Gaussian white noise.

The delicate aspect with this simple operational description is that w is a highly singular entity that does not admit an interpretation as a conventional function of the time variable t (think of w as the stochastic counterpart of the Dirac distribution δ whose explicit definition as a tempered distribution is $\langle \delta, \varphi \rangle = \varphi(0)$ for all "test" functions φ). In fact, specifying a stochastic process s in this framework is equivalent to prescribing the probability law of the random vector $[\langle s, \varphi_1 \rangle, \cdots, \langle s, \varphi_N \rangle]^\top$ for any N test functions $\{\varphi_1, \cdots, \varphi_N\}$. The mathematical framework for the correct interpretation of (2.4) is Gelfand's theory of generalized stochastic processes [19], which is briefly summarized as follows:

White Noise

The theory starts with the definition of the white noise. A generalized white noise is a probability measure on the dual space of a set of test functions $\varphi : \mathbb{R} \to \mathbb{R}$ (typically, Schwartz space of tempered distributions) that has the following properties:

- For a given test function φ, the statistics of the random variable $\langle w, \varphi \rangle$ do not change upon shifting φ, where w denotes a generic random element in the dual space of test functions .

- If the test functions in the collection $\{\varphi_\beta\}_{\beta \in B}$ (B is an index set) have disjoint supports, then the random variables in $\{\langle w, \varphi_\beta \rangle\}_{\beta \in B}$ are independent.

Under some mild regularity conditions, there is a one-to-one correspondence between the infinitely divisible random variables and the white noises specified above. A random variable X is called infinite divisible if for any $n \in \mathbb{N}$, there exists n iid random variables X_1 to X_n such that $X_1 + \cdots + X_n$ has the same distribution as X. Thus, if the characteristic function of X is

$$\mathbb{E}\left[e^{j\omega X}\right] = e^{f(\omega)}, \tag{2.5}$$

then for any $n \in \mathbb{N}$, $\exp\left(\frac{1}{n}f(\omega)\right)$ is also a valid characteristic function of a random variable. The function f is called the Lévy exponent of the infinite-divisible random variable X. Now, the white noise corresponding to the random variable X is the random process w for which we have

$$\mathbb{E}\left[\exp\left(j\sum_{i=1}^{N}\omega_i\langle w, \varphi_i\rangle\right)\right] = \exp\left(\int_{\mathbb{R}} f\left(\sum_{i=1}^{N}\omega_i\varphi_i(x)\right)dx\right) \tag{2.6}$$

where φ_1 to φ_N are N arbitrary test functions. This completely characterizes the white noise w.

Stochastic Processes

To define the stochastic process s in (2.4), we take advantage of the formula

$$\langle s, \varphi \rangle = \langle \mathrm{L}^{-1}w, \varphi \rangle = \langle w, \mathrm{L}^{-1*}\varphi \rangle \tag{2.7}$$

where L^{-1*} is the adjoint operator of L^{-1}. It means that one can readily deduce the statistical distribution of $\langle s, \varphi \rangle$ given the white noise w and the operator L^{-1*}. The combination of (2.6) and (2.7) gives the complete characterization of the process s, i.e.

$$\mathbb{E}\left[\exp\left(\mathrm{j}\sum_{i=1}^{N}\omega_i\langle s, \varphi_i\rangle\right)\right] = \exp\left(\int_{\mathbb{R}} f\left(\sum_{i=1}^{N}\omega_i\mathrm{L}^{-1*}\{\varphi_i\}(x)\right)\mathrm{d}x\right) \tag{2.8}$$

for any N test functions φ_1 to φ_N.

2.2.2 Symmetric-α-Stable (SαS) White Noise

An important family of white noises is the α-stable ones. They are an extension of the Gaussian white noise that play a central role in the theory of stochastic processes. In this thesis, we focus on this class of white noises; specifically the symmetric-α-stable (SαS) white noises. It is a rich family that exhibit a wide range of desirable properties to model practical signals which will be discussed later. Additionally, such stable models are attractive for statistical signal processing because they lend themselves well to analytic calculations [23]. Areas of applications include detection theory [24], communications [25], and signal denoising [26].

To describe the SαS white noises, we first describe the SαS random variables.

Definition 1. *The random variable X is SαS if its characteristic function is of the form*

$$\mathbb{E}\left[\mathrm{e}^{\mathrm{j}\omega X}\right] = \mathrm{e}^{-\|a\omega\|_\alpha^\alpha}, \tag{2.9}$$

for some $a \in \mathbb{R}^+$. Notice that in order for (2.9) to be a valid characteristic function α needs to be in $(0, 2]$, where $\alpha = 2$ corresponds to the Gaussian case. The quantity a^α is known as the dispersion parameter which plays a role similar to the variance of Gaussian random variables.

We also recall a fundamental property of stable distributions that can also serve as their definition.

Property 1 (Linear combination of SαS random variables)**.** *Let* $\bar{r} = \sum_{n=1}^{N} a_n r_n$ *where the* r_m *are iid SαS random variables that have the same distribution as an SαS random variable* r*. Then,* $\bar{r}$ *is an SαS random variable as well with the same distribution as* $\|[a_1, \cdots, a_N]\|_\alpha \, r$ *[23].*

To establish this property, we consider N iid SαS random variables $r_1, \ldots, r_N$ with common characteristic function $\exp(-|a\omega|^\alpha)$, and a corresponding sequence of real-valued weights $a_1, \ldots, a_N$. The characteristic function of the random variable $\bar{r} = \sum_{n=1}^{N} a_n r_n$ is given by

$$\hat{p}_{\bar{r}}(\omega) = \prod_{n=1}^{N} e^{-|a\, a_n \omega|^\alpha} = e^{-\left|a\left(\sum_{n=1}^{N} |a_n|^\alpha\right)^{1/\alpha}\omega\right|^\alpha}. \tag{2.10}$$

Thus, $\bar{r}$, which is a linear combination of iid SαS random variables, is an SαS random variable with the same distribution as one of them multiplied by the factor $(\sum_{n=1}^{N} |a_n|^\alpha)^{1/\alpha}$; i.e.

$$\bar{r} \overset{d}{=} \left(\sum_{n=1}^{N} |a_n|^\alpha \right)^{1/\alpha} r_1. \tag{2.11}$$

It is easy to see that stable distributions belong to the class of infinite divisible distributions. Thus, there exist corresponding white noises that are called SαS white noises. According to (2.5), the Lévy exponent corresponding to r is $f(\omega) = -|a\omega|^\alpha$. Therefore, according to (2.6), the corresponding SαS white noise w is characterized as follows

$$\mathbb{E}\left[\exp\left(j \sum_{i=1}^{N} \omega_i \langle w, \varphi_i \rangle \right)\right] = \exp\left(\int_{\mathbb{R}} \left| \sum_{i=1}^{N} \omega_i \varphi_i(x) \right|^\alpha dx \right)$$

$$= \exp\left(\left\| a \sum_{i=1}^{N} \omega_i \varphi_i(x) \right\|_\alpha^\alpha \right) \tag{2.12}$$

where φ_1 to φ_N are any N number of test functions. Specifically, for a random variable $\langle w, \varphi \rangle$, we have

$$\hat{p}_{\langle w, \varphi \rangle}(\omega) = \mathbb{E}\left[e^{j\omega\langle w, \varphi \rangle}\right] = e^{-|a\|\varphi\|_\alpha \omega|^\alpha} \tag{2.13}$$

which is the characteristic function of the SαS random variable with dispersion parameter $a^\alpha \|\varphi\|_\alpha^\alpha$. This means that the observation of an SαS white noise through any window is an SαS random variable.

The case of $\alpha = 2$ is the well studied Gaussian distribution. Except that, the probability density function (pdf) of an SαS random variable is heavy tailed for $\alpha < 2$. Precisely, it decays with the order of $x^{-(\alpha+1)}$ asymptotically. Therefore, according to [11] and [12], these white noises are good candidates for modelling sparse processes.

2.3   Discrete Wavelet Transform

Another main ingredient of this thesis are the wavelet bases and frames. In this subsection, we provide a brief preliminaries about them which is enough to comprehend the next chapters.

The construction of wavelets start with the notion of a multi-resolution analysis for $L_2(\mathbb{R})$. Assume that $\{V_i\}_{i\in\mathbb{Z}}$ is a sequence of linear subspaces of $L_2(\mathbb{R})$ that satisfy the following three conditions:

1. V_i are nested, i.e. for all $i \in \mathbb{Z}$, we have

$$V_i \subseteq V_{i+1}. \tag{2.14}$$

2. They generate $L_2(\mathbb{R})$, i.e.

$$\overline{\bigcup_{i\in\mathbb{Z}} V_i} = L_2(\mathbb{R}) \tag{2.15}$$

   where bar means the set closure operation.

3. Their intersection is empty, i.e.

$$\bigcap_{i\in\mathbb{Z}} V_i = \emptyset. \tag{2.16}$$

Also, assume that V_i is generated by the shifted versions of a single function φ_i over a uniform grid, i.e.

$$V_i = \mathrm{span}\Big\{ \varphi_{ij}(x) \big| \varphi_{ij}(x) = \varphi_i(x - jT_i) \Big\} \tag{2.17}$$

where T_i is the step size of grid corresponds to V_i. Now, this sequence of nested spaces gives us a multi-resolution basis for $L_2(\mathbb{R})$ since by increasing i, projection of a function

$f \in L_2(\mathbb{R})$ on V_i results in an approximation of f with an intuitively higher resolution. This structure implies that there exist $h_{ij} \in \mathbb{R}$ for $i, j \in \mathbb{Z}$ such that

$$\varphi_i(x) = \sum_{j \in \mathbb{Z}} h_{ij} \varphi_{i+1,j}(x). \tag{2.18}$$

Now, let W_i be the orthogonal complement of V_i inside V_{i+1} (see (2.14)). Thus, we can write

$$V_{i+1} = V_i \oplus W_i \tag{2.19}$$

where $\oplus$ denotes the direct sum of vector spaces. Using this expression recursively, according to (2.15), we deduce that, for any fixed $i_0 \in \mathbb{Z}$,

$$\overline{V_{i_0} \oplus \bigoplus_{i=i_0}^{\infty} W_i} = \overline{\bigcup_{i=i_0}^{\infty} V_i} = L_2(\mathbb{R}). \tag{2.20}$$

Incorporating (2.16), we obtain

$$\overline{\bigoplus_{i \in \mathbb{Z}} W_i} = L_2(\mathbb{R}). \tag{2.21}$$

Since W_i is orthogonal to V_i and $V_j \subseteq V_i$ for all $j < i$, W_i is orthogonal to W_j when $i \neq j$. Hence, $\{W_i\}_{i \in \mathbb{Z}}$ is a sequence of orthogonal spaces. Moreover, under certain conditions W_i is also generated by shifted versions of a function ψ_i on the grid with step size T_i, i.e.

$$W_i = \text{span}\Big\{ \psi_{ij}(x) \big| \psi_{ij}(x) = \psi_i(x - jT_i) \Big\}. \tag{2.22}$$

Thus, relying on (2.20) for any $i_0 \in \mathbb{Z}$, $\{\varphi_{i_0 j}\}_{j \in \mathbb{Z}} \cup \Big\{ \{\psi_{ij}\}_{i=i_0}^{\infty} \Big\}_{j \in \mathbb{Z}}$ generates $L_2(\mathbb{R})$. Also, relying on (2.21), $\{\psi_{ij}\}_{i,j \in \mathbb{Z}}$ generates $L_2(\mathbb{R})$ as well. These two basis for $L_2(\mathbb{R})$ are called the *wavelet* bases.

Similar to (2.18), we can also write

$$\psi_i(x) = \sum_{j \in \mathbb{Z}} g_{ij} \varphi_{i+1}(x - jT_{i+1}) \tag{2.23}$$

with $g_{ij} \in \mathbb{R}$ for all $i, j \in \mathbb{Z}$. These coefficients h_{ij} and g_{ij} allow us to perform a fast algorithm for calculating the representation coefficients of a function in the wavelet bases described above [27].

2.3.1   Conventional Wavelets

In a conventional wavelet basis, it is additionally assumed that φ_i are dilated versions of a fixed function φ for any i. Precisely, we have

$$\varphi_i(x) = 2^{\frac{i}{2}} \varphi(2^i x) \tag{2.24}$$

(the multiplicative constant is to keep the L_2-norm of all φ_i fixed). It is also assumed that $T_i = 2^{-i}$. Consequently, we have

$$\varphi_{ij}(x) = \varphi_i(x - jT_i) = 2^{\frac{i}{2}} \varphi(2^i x - j). \tag{2.25}$$

Considering the case of $i = 0$, by substituting (2.24) in (2.18), we obtain

$$
\begin{aligned}
\varphi(x) = \varphi_0(x) &= \sum_{j \in \mathbb{Z}} h_{0j} \varphi_1\big(x - 2^{-1}j\big) \\
&= \sqrt{2} \sum_{j \in \mathbb{Z}} h_{0j} \varphi_0\big(2x - j\big) \\
&= \sqrt{2} \sum_{j \in \mathbb{Z}} h_{0j} \varphi\big(2x - j\big).
\end{aligned} \tag{2.26}
$$

This equation means that φ has a kind of fractal property since it is a linear combination of shifts of its contracted version. According to (2.26), for any $i \in \mathbb{Z}$, we can write

$$
\begin{aligned}
\varphi_i(x) &= 2^{\frac{i}{2}} \varphi(2^i x) \\
&= 2^{\frac{i+1}{2}} \sum_{j \in \mathbb{Z}} h_{0j} \varphi\big(2^{i+1} x - j\big) \\
&= \sum_{j \in \mathbb{Z}} h_{0j} \varphi_{i+1}\big(x - jT_{i+1}\big).
\end{aligned} \tag{2.27}
$$

Therefore, when (2.24) holds, the h_{ij} in (2.18) are equal to h_{0j} irrelevant to i. Thus, we drop the index 0 in h_{0j} from now on (so we just write h_j).

Also, we can deduce that there exist a function $\psi(x)$ such that for any i, we have

$$\psi_i(x) = 2^{\frac{i}{2}} \psi(2^i x) \tag{2.28}$$

and g_{ij} in (2.23) equals g_{0j} for all i. We drop the index i for g_{ij}, too.

Therefore, in brief, a conventional wavelet basis is $\{\varphi_{i_0 j}\}_{j \in \mathbb{Z}} \cup \left\{ \{\psi_{ij}\}_{i=i_0}^{\infty} \right\}_{j \in \mathbb{Z}}$ or $\{\psi_{ij}\}_{i,j \in \mathbb{Z}}$ that generates $L_2(\mathbb{R})$ and in which we have

$$\varphi_{ij}(x) = 2^{\frac{i}{2}} \varphi(2^i x - j), \tag{2.29}$$

$$\psi_{ij}(x) = 2^{\frac{i}{2}} \psi(2^i x - j). \tag{2.30}$$

Also, the function φ and ψ satisfy the two-scale relation

$$\varphi(x) = \sqrt{2} \sum_{j \in \mathbb{Z}} h_j \varphi(2x - j), \tag{2.31}$$

$$\psi(x) = \sqrt{2} \sum_{j \in \mathbb{Z}} g_j \varphi(2x - j). \tag{2.32}$$

They are called the father wavelet (or scaling function) and the mother wavelet, respectively. Moreover, the sequences $\{h_j\}_{j \in \mathbb{Z}}$ and $\{g_j\}_{j \in \mathbb{Z}}$ are respectively the impulse response of a low-pass and a high-pass filter. These filters play a fundamental role in Mallat's efficient filter-bank implementation of the wavelet decomposition [27].

The most famous wavelets are Daubechies wavelets [28] that are compactly supported in the time domain and Meyer wavelets [29] that are compactly supported in the Fourier domain. We explain the latter later on due to their importance in this thesis.

2.3.2 Operator-like Wavelets

Conventional wavelet bases act as smoothed versions of the derivative operator [30]. However, there exist multi-resolution wavelet-like bases which essentially behave like a given differential operator L. These bases are called operator-like wavelets [31]. The operator-like wavelet at scale i and location j is given by

$$\psi_{ij}(x) = L^* \{\phi_i\}(x - j2^{-i}T), \tag{2.33}$$

where ϕ_i is a scale-dependent smoothing kernel.

In our study, we work only with operator-like wavelets tailored to first order differential operators $L = D + \kappa I$ where D is the differentiator and κ is a non-negative real number (they are also called first order autoregressive (AR(1)) systems). In this case, the

operator-like wavelets proposed in [31] are very similar to Haar wavelets [32] (or first or-
der Daubechies wavelets). The only difference is that Haar wavelets are piecewise constant
while these operator-like wavelets are piecewise exponential. Indeed, we have

$$\psi_{ij}(t) \propto e^{-2^{-i}\kappa T} \beta_{\kappa,2^{-i}T}(t - j2^{-i}T) - \beta_{\kappa,2^{-i}T}(t - (j+1)2^{-i}T)$$

$$= \begin{cases} 0 & t < j2^{-i}T \\ e^{-\kappa(t-(j-1)2^{-i}T)} & j2^{-i}T \le t < (j+1)2^{-i}T \\ -e^{-\kappa(t-(j+1)2^{-i}T)} & (j+1)2^{-i}T \le t < (j+2)2^{-i}T \\ 0 & (j+2)2^{-i}T \le t \end{cases}. \tag{2.34}$$

where $\beta_{\kappa,2^{-i}T}$ is the exponential B-spline with parameters κ and $2^{-i}T$ [18]. In essence,
this amounts to replacing the finite-difference operations of the conventional wavelet trans-
form algorithm by a suitable series of linear prediction errors where the coefficients are de-
termined by the pole of the AR(1) system. In contrast with conventional wavelets, operator-
like wavelets have scale dependent h_{ij} and g_{ij} coefficients. However, it does not affect the
efficiency of the filter-bank implementation of them.

2.3.3   One-Dimensional Wavelet Frames

The story of conventional wavelets can be told from another point of view. Assume that
for a given function ψ, we define

$$\psi_{ij}(x) = 2^{\frac{i}{2}} \psi(2^i x - j) \tag{2.35}$$

for $i, j \in \mathbb{Z}$, the same as what we had in the previous subsection. For $\{\psi_{ij}\}_{i,j \in \mathbb{Z}}$ to be
a frame (not necessarily a basis) for $L_2(\mathbb{R})$ with robust reconstruction, there need to exist
$A, B > 0$ such that for any $f \in L_2(\mathbb{R})$, we have [33]

$$A\|f\|_2^2 \le \sum_{i,j \in \mathbb{Z}} |\langle f, \psi_{ij} \rangle|^2 \le B\|f\|_2^2. \tag{2.36}$$

If $A = B$, then the frame is called a tight frame. Also, there exist a dual frame $\{\tilde{\psi}_{ij}\}_{i,j \in \mathbb{Z}}$,
which gives us the representation coefficients, i.e.

$$f(x) = \sum_{i,j \in \mathbb{Z}} \langle f, \tilde{\psi}_{ij} \rangle \psi_{ij}(x) \tag{2.37}$$

for any $f \in L_2(\mathbb{R})$.

In [33], it is shown that for a given ψ, A and B which satisfy (2.36) can be calculated as

$$A = \inf_{\rho} \left\{ \sum_{i \in \mathbb{Z}} \left| \hat{\psi}(2^i \rho) \right|^2 - \sum_{k \in \mathbb{Z} \backslash \{0\}} \left(\theta(2\pi k)\theta(-2\pi k) \right)^{\frac{1}{2}} \right\} \tag{2.38}$$

and

$$B = \sup_{\rho} \left\{ \sum_{i \in \mathbb{Z}} \left| \hat{\psi}(2^i \rho) \right|^2 + \sum_{k \in \mathbb{Z} \backslash \{0\}} \left(\theta(2\pi k)\theta(-2\pi k) \right)^{\frac{1}{2}} \right\} \tag{2.39}$$

where $\hat{\psi}$ is the Fourier transform of ψ and

$$\theta(s) = \sup_{\rho} \sum_{\ell \in \mathbb{Z}} \left| \hat{\psi}(2^\ell \rho) \right| \left| \hat{\psi}(2^\ell \rho + s) \right|. \tag{2.40}$$

Therefore, if we assume that ψ is band-limited over $[-\pi, \pi]$ and

$$\sum_{i \in \mathbb{Z}} \left| \hat{\psi}(2^i \rho) \right|^2 = 1, \tag{2.41}$$

then we get a frame which is tight with $A = B = 1$. These wavelets have two favorable properties. The first one is norm preservation due to their tight frame property with $A = B = 1$. This property implies that the dual frame is the same as the original frame, i.e.

$$f(x) = \sum_{i,j \in \mathbb{Z}} \langle f, \psi_{ij} \rangle \psi_{ij}(x). \tag{2.42}$$

for any $f \in L_2(\mathbb{R})$, which is important in reconstruction applications. The second one is the band-limitedness that can be translated into an efficient wavelet transform algorithm in the Fourier domain.

Meyer Wavelet Bases

An interesting subfamily of wavelets that satisfy (2.41) are the Meyer wavelets. Meyer wavelets are not only tight frames but also orthonormal bases [29, 33]. This subfamily is characterized by a function $v : [0, 1] \to \mathbb{R}$ such that

$$v(\rho) + v(1 - \rho) = 1. \tag{2.43}$$

Having this function, the mother wavelet profile in Fourier domain is

$$\hat{\psi}(\rho) = e^{j\rho/2} \times \begin{cases} \sin\left(\frac{\pi}{2} v(\frac{3}{2\pi}|\rho| - 1)\right), & \frac{2\pi}{3} < |\rho| \le \frac{4\pi}{3} \\ \cos\left(\frac{\pi}{2} v(\frac{3}{4\pi}|\rho| - 1)\right), & \frac{4\pi}{3} < |\rho| \le \frac{8\pi}{3} \\ 0, & \text{otherwise} \end{cases} . \tag{2.44}$$

Notice that the dilation that expands the support of $\hat{\psi}(\rho)$ to $\left[-\frac{8\pi}{3}, -\frac{2\pi}{3}\right] \cup \left[\frac{2\pi}{3}, \frac{8\pi}{3}\right]$, rather than a subset of $[-\pi, \pi]$, is crucial to obtain an orthonormal basis and not only a tight frame. In fact, this specific construction results in miraculous cancelations (stated as such by Daubechies [33]) that result in the emergence of a basis. Moreover, this charaterization helps us to easily change different properties of the wavelet such as localization while keeping it orthonormal.

2.3.4   Two-Dimensional Isotropic Wavelet Frames

Similar to the 1-dimensional case, we can construct a wavelet frame for $L_2(\mathbb{R}^2)$. Let the mother wavelet ψ be a function from $\mathbb{R}^2$ to $\mathbb{R}$. The complete wavelet frame is given by

$$\psi_{i,\mathbf{k}}(\mathbf{x}) = 2^i \, \psi\left(2^i \mathbf{x} - \mathbf{k}\right), \tag{2.45}$$

in which $i \in \mathbb{Z}$ and $\mathbf{k} \in \mathbb{Z}^2$.

In particular, we are interested in isotropic wavelets. These wavelets are circular symmetric, thus, enabling us to obtain orientation-free analysis of images. This means that the value of a wavelet coefficient at a specific position does not change if one rotates the underlying image with respect to the center of the wavelet. Mathematically, we can write

$$\psi(\mathbf{x}) = \psi(r), \tag{2.46}$$

where $r = \|\mathbf{x}\|$. Here, for simplicity, we use the notation ψ both for the wavelet and for the radial profile of the wavelet. Let $h(\boldsymbol{\omega})$ be the 2-dimensional Fourier transform of $\psi(\mathbf{x})$. According to the properties of the Fourier transform, h is also isotropic. Hence, like (2.46), we write

$$h(\boldsymbol{\omega}) = h(\rho), \tag{2.47}$$

where $\rho = \|\boldsymbol{\omega}\|$. According to [34], ψ and h are related through the Hankel transform as

$$\psi(r) = \mathcal{H}\{h\}(r) = \int_0^\infty h(t) \, J_0(rt) \, t \, dt. \tag{2.48}$$

for $r \geq 0$. This formula allows us to calculate the radial profile of an isotropic wavelet in space domain from its radial profile in Fourier domain through a 1-dimensional integral, rather than calculating 2-dimensional Fourier transforms.

The second assumption is that ψ creates a tight-frame family. According to [35, 36], this condition is satisfied if

$$\sum_{i \in \mathbb{Z}} |h(2^i \rho)|^2 = 1 \quad \text{for} \quad \forall \rho \in \mathbb{R}^+ \backslash \{0\}. \tag{2.49}$$

Comparing (2.49) with (2.41), we conclude that by using the Fourier transform of a 1-dimensional mother wavelet of a tight wavelet family as radial frequency profile, we obtain a mother wavelet of a 2-dimensional mother wavelet. Also, there is an equivalence between stating (2.49) for all positive ρ and stating (2.49) for $\rho \in [\frac{\pi}{2}, \pi]$ because $2^i \rho$, for $i \in \mathbb{Z}$, can reach any arbitrary positive value.

Filter-Bank Implementation of the Isotropic Wavelets

Since in practice we are working with discrete-domain signals, we are interested in obtaining the wavelet coefficients of a discrete signal. To be precise, assume that $f[\mathbf{k}]$ for $\mathbf{k} \in \mathbb{Z}^2$ is a discrete-domain signal. Based on it, we construct a continuous-domain signal

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^2} f[\mathbf{k}] \, \text{sinc}(x_1 - k_1) \text{sinc}(x_2 - k_2) \tag{2.50}$$

where $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ and $\mathbf{k} = (k_1, k_2)$. Now, we are interested in calculating the wavelet coefficients of $f(\mathbf{x})$.

To have a straightforward exact implementation of the wavelet transform, we assume that the wavelet profile $h(\rho)$ is supported on $[\frac{\pi}{4}, \pi]$. We see the advantage of this assumption during the derivation of the method. First, define the highpass h_H and lowpass h_L filters as

$$h_H(\boldsymbol{\omega}) = \begin{cases} 0 & \|\boldsymbol{\omega}\| \leq \frac{\pi}{4} \\ h(\|\boldsymbol{\omega}\|) & \frac{\pi}{4} < \|\boldsymbol{\omega}\| < \frac{\pi}{2} \\ 1 & \frac{\pi}{2} \leq \|\boldsymbol{\omega}\| \end{cases}, \tag{2.51}$$

$$h_L(\boldsymbol{\omega}) = \begin{cases} 1 & \|\boldsymbol{\omega}\| \leq \frac{\pi}{4} \\ h(2\|\boldsymbol{\omega}\|) & \frac{\pi}{4} < \|\boldsymbol{\omega}\| < \frac{\pi}{2} \\ 0 & \frac{\pi}{2} < \|\boldsymbol{\omega}\| \end{cases}, \tag{2.52}$$

respectively. Notice that for these h_H and h_L, we have

$$h(\boldsymbol{\omega}) = h_L\left(\frac{\boldsymbol{\omega}}{2}\right) h_H(\boldsymbol{\omega}), \tag{2.53}$$

and also according to the tight-frame constraint (2.49), we automatically have that

$$h_L^2(\boldsymbol{\omega}) + h_H^2(\boldsymbol{\omega}) = 1. \tag{2.54}$$

For $i \in \mathbb{Z}$ and $\mathbf{k} \in \mathbb{Z}^2$, we have

$$\begin{aligned} w_{i,\mathbf{k}} = \langle f(\mathbf{x}), \psi_{i,\mathbf{k}}(\mathbf{x}) \rangle &= (f(\mathbf{x}) * \psi_{i,\mathbf{0}}(\mathbf{x})) \left(2^{-i}\mathbf{k}\right) \\ &= \mathcal{F}^{-1}\left\{\hat{f}(\boldsymbol{\omega})2^{-i}h\left(2^{-i}\boldsymbol{\omega}\right)\right\}\left(2^{-i}\mathbf{k}\right). \end{aligned} \tag{2.55}$$

Since the support of $\hat{f}(\boldsymbol{\omega})$ is a subset of $[-\pi, \pi] \times [-\pi, \pi]$, for all $i > 1$, $w_{i,\mathbf{k}} = 0$. Thus, we let

$$w_{0,\mathbf{k}} = \mathcal{F}^{-1}\left\{\hat{f}(\boldsymbol{\omega})h_H\left(\boldsymbol{\omega}\right)\right\}(\mathbf{k}) \tag{2.56}$$

which actually accumulates the effect of wavelet coefficients for $i \geq 0$. For $i < 0$, according to the properties of the Fourier transform, we have

$$\mathcal{F}^{-1}\left\{\hat{f}(\boldsymbol{\omega})2^{-i}h\left(2^{-i}\boldsymbol{\omega}\right)\right\}\left(2^{-i}\mathbf{k}\right) = 2^i\,\mathcal{F}^{-1}\left\{\hat{f}\left(2^i\boldsymbol{\omega}\right)h\left(\boldsymbol{\omega}\right)\right\}(\mathbf{k}). \tag{2.57}$$

Using (2.53), we can write

$$2^i\,\mathcal{F}^{-1}\left\{\hat{f}\left(2^i\boldsymbol{\omega}\right)h\left(\boldsymbol{\omega}\right)\right\}(\mathbf{k}) = 2^i\,\mathcal{F}^{-1}\left\{\hat{f}\left(2^i\boldsymbol{\omega}\right)h_L\left(\frac{\boldsymbol{\omega}}{2}\right)h_H(\boldsymbol{\omega})\right\}(\mathbf{k}). \tag{2.58}$$

Notice that $h_L(\boldsymbol{\omega}) = 0$ for $\|\boldsymbol{\omega}\| > \frac{\pi}{2}$. Thus, downsampling $\mathcal{F}^{-1}\left\{\hat{f}\left(2^{i+1}\boldsymbol{\omega}\right)h_L\left(\boldsymbol{\omega}\right)\right\}(\mathbf{k})$ by factor of 2, we obtain $\mathcal{F}^{-1}\left\{\hat{f}\left(2^i\boldsymbol{\omega}\right)h_L\left(\frac{\boldsymbol{\omega}}{2}\right)\right\}(\mathbf{k})$ without any frequency aliasing. Also, notice that

$$h_L\left(\frac{\boldsymbol{\omega}}{2}\right)h_L(\boldsymbol{\omega}) = h_L(\boldsymbol{\omega}). \tag{2.59}$$

These yield that by filtering $\mathcal{F}^{-1}\left\{\hat{f}\left(2^i\boldsymbol{\omega}\right)h_L\left(\frac{\boldsymbol{\omega}}{2}\right)\right\}(\mathbf{k})$ with $h_L(\boldsymbol{\omega})$ and then downsampling by the factor of 2, we get $\mathcal{F}^{-1}\left\{\hat{f}\left(2^{i-1}\boldsymbol{\omega}\right)h_L\left(\frac{\boldsymbol{\omega}}{2}\right)\right\}(\mathbf{k})$. Combining this fact with

---

Algorithm 1: Efficient Recursive Method for Calculating the Wavelet Coefficient of a Discrete Signal

---

1: **input:** $f[\mathbf{k}]$ for $\mathbf{k} \in \mathbb{Z}^2$
2: $f_0 \leftarrow f$
3: **for** $j = 0$ **to** ∞ **do**
4: $\quad w_{i,\mathbf{k}} \leftarrow \mathcal{F}^{-1}\left\{\hat{f}_i(\boldsymbol{\omega}) h_H(\boldsymbol{\omega})\right\}(\mathbf{k})$
5: $\quad f_{i-1}[\mathbf{k}] \leftarrow \left(\frac{1}{2}\mathcal{F}^{-1}\left\{\hat{f}_i(\boldsymbol{\omega}) h_L(\boldsymbol{\omega})\right\}(\mathbf{k})\right) \downarrow 2$
6: $\quad i \leftarrow i - 1$
7: **end for**
8: **return** $\left\{\{w_{i,\mathbf{k}}\}_{\mathbf{k}\in\mathbb{Z}^2}\right\}_{i\in\mathbb{Z}\backslash\mathbb{N}}$

(2.58), one can use the recursive Algorithm 1 for calculating the wavelet coefficients of a discrete signal f:

Therefore, Algorithm 1 gives us an efficient implementation of the wavelet transform for discrete signals which is equivalent to the standard decimated filter-bank implementation of wavelets. The only difference is that we do not down-sample the wavelet coefficients in the high-pass channel of highest resolution ($w_{0,\mathbf{k}}$) which means that the redundancy factor of these frames are $4/3$.

Chapter 3

Optimal Representation of Sparse AR(1) Processes

The discrete cosine transform (DCT) is known to be asymptotically equivalent to the Karhunen-Loève transform (KLT) of *Gaussian* stationary processes including first-order auto-regressive (AR(1)) ones. Since being uncorrelated under the Gaussian hypothesis is synonymous with independence, it also yields an independent-component analysis (ICA) of such signals. In this chapter, we present a constructive non-Gaussian generalization of this result: the characterization of the optimal orthogonal transform (ICA) for the family of symmetric-α-stable AR(1) processes. The degree of sparsity of these processes is controlled by the stability parameter $0 < \alpha \le 2$ with the only non-sparse member of the family being the classical Gaussian AR(1) process with $\alpha = 2$. Specifically, we prove that, for $\alpha < 2$, a fixed family of operator-like wavelet bases systematically outperforms the DCT in terms of compression and denoising ability. The effect is quantified with the help of two performance criteria (one based on the Kullback-Leibler divergence, and the other on Stein's formula for the minimum estimation error) that can also be viewed as statistical measures of independence. Finally, we observe that, for the sparser kind of processes with $0 < \alpha \le 1$, the operator-like wavelet basis, as dictated by linear system theory, is undistinguishable from the ICA solution obtained through numerical optimization. Our framework offers a unified view that encompasses sinusoidal transforms such as the DCT and a family of orthogonal Haar-like wavelets that is linked analytically to the underlying signal model.

23

3.1   Introduction

Transform-domain processing is a classical approach to compress signals, model data, and extract features. The guiding principle is to produce transform-domain coefficients that are decoupled statistically so that a simple component-wise processing can be applied; i.e, each coefficient is processed independently of the others. The reference solution in the field is the Karhunen-Loève transform (KLT) which yields transform-coefficients that are uncorrelated and therefore also independent, provided the process is Gaussian. Also, if the process is stationary with finite variance and infinite length, then the KLT is a Fourier-like transform [2]. Moreover, it has been shown that the discrete cosine transform (DCT) [37] is asymptotically equivalent to the KLT for the whole class of stationary processes [3], including the AR(1) model [38]; thus, for a Gaussian input, all these transforms result in a fully decoupled (independent) representation. However, this favorable independence-related property is extinguished for non-Gaussian processes. In this case, the coefficients are only partially decoupled and the representation of the signal afforded by the KLT is no longer optimal.

In recent years, wavelets have emerged as an alternative representation of signals and images. Typical examples of successful applications are JPEG2000 for image compression [39] and shrinkage methods for attenuating noise [6, 40]. The fact that wavelets are so effective in transform-domain applications suggests that they are naturally suited to represent practical processes. This empirical observation was established by early studies that include [9], where many natural images were subjected to an independent-component analysis (ICA). It was found that the resulting components have properties that are reminiscent of 2D wavelets and/or Gabor functions. Additional ICA experiments were performed in [14] on realizations of the stationary sawtooth process and of Meyer's ramp process [41]; for both processes, the basis vectors of ICA exhibit a wavelet-like multiresolution structure.

Despite their empirical usefulness, the optimality of wavelets for the representation of non-Gaussian stochastic processes remains poorly understood from a theoretical point of view. An early study can be traced back to [42], where the decomposition of fractional Brownian motions over a wavelet basis was shown to result in almost uncorrelated coefficients, under some conditions. By contrast, in the deterministic framework, it is well known that wavelets are optimal (up to some constant) for the N-term approximation of functions in Besov spaces [13]; the extension of this result to a statistical setting could be achieved only experimentally.

The general distributional framework for the specification of sparse stochastic processes presented in Section 2.2 is particularly well suited to the specification of symmetric-

α-stable (SαS) white noises. These noises can be used to drive first-order stochastic differential equations (SDE) to synthesize AR(1) processes. As it turns out, AR(1) systems and α-stable distributions are at the core of signal modeling and probability theory. As discussed in Section 2.2.2, the classical Gaussian processes correspond to $\alpha = 2$, while $0 < \alpha < 2$ yields stable processes that have heavy-tailed statistics and that are prototypical representatives for sparse signals. Also, specifically heavy-tailed AR have been used to model phenomena in network [43], sea surface [44], economy and finance [45].

In this chapter, we take advantage of the framework presented in Section 2.2 to establish the optimality of a certain class of wavelets in a stochastic sense. We start by characterizing the amount of dependency between the coefficients of stochastic processes represented in an arbitrary transform domain. To that end, we introduce two performance criteria. The first assesses the coding performance of the transform: it is given by the Kullback-Leibler divergence between the joint probability density function (pdf) of the original signal and the product of the marginals in the transformed domain. The second is a theoretical prediction of denoising performance under the hypothesis of additive white Gaussian noise (AWGN). It is based on Stein's formula for the mean-square estimation error and also takes the form of a divergence between the joint pdf of the original signal and the product of the marginals in the transformed domain. Then, we seek the orthogonal transformation that minimizes these statistical criteria. We confirm the loss of optimality of the DCT for $0 < \alpha < 2$ and validate the superiority of a special brand of operator-like wavelet transform that is matched to the underlying signal model (see Section 2.3.2). Our reference method in this comparison is the ICA solution that is determined by numerical means for different values of α. The remarkable empirical finding of this chapter is that the ICA solution converges to the operator-like wavelets for values of α below one. Moreover, the practical relevance of these results is that, unlike ICA, the operator-like wavelets are known in analytical form in terms of the pole of the underlying system $\big($see Eq. (2.34)$\big)$.

This chapter is organized as follows: In Section 3.2, we introduce two measures of divergence between distributions that are suitable for either noise attenuation or compression applications. The signal model fundamental to this chapter is discussed in Section 3.3. In Section 3.4, we derive the explicit form of our performance criteria for the SαS model in the context of transform-domain compression and noise attenuation. In addition, we provide an iterative algorithm to find the optimal basis (Section 3.5). Results for different AR(1) processes and different transform domains are discussed in Section 3.6. The last section is dedicated to the recapitulation of the main results, the relation to prior works, and topics for future studies.

3.2   Information Theoretic Performance Measures

In statistical signal processing, it is of interest to precisely quantify the best-achievable performance when the model is not perfectly matched to the signal under investigation, or when certain simplifying hypothesis, such as independence, are being made. In the following, we address this issue for the two problems of compression and denoising when the assumed distribution and the real one may differ.

3.2.1   Compression Based on Non-Exact Distribution

It is well-known that, if we have a source $\mathbf{s}$ of iid random vectors with common pdf $p_{\mathbf{s}}$, then the logarithm of measure of the coding set per sample can be at least

$$\mathbb{H}(p_{\mathbf{s}}) = -\int p_{\mathbf{s}}(\mathbf{s}) \log p_{\mathbf{s}}(\mathbf{s}) \mathrm{d}\mathbf{s} \tag{3.1}$$

which is the entropy[1] of the source [46]. However, if we compress $\mathbf{s}$ assuming that it is distributed according to $q_{\mathbf{s}}$ (rather than $p_{\mathbf{s}}$), then

$$\begin{aligned}\mathbb{H}(q_{\mathbf{s}}) &= \mathbb{H}(p_{\mathbf{s}}) + \mathbb{D}(p_{\mathbf{s}} \| q_{\mathbf{s}}) \\ &= \mathbb{H}(p_{\mathbf{s}}) + \int p_{\mathbf{s}}(\mathbf{s}) \log \frac{p_{\mathbf{s}}(\mathbf{s})}{q_{\mathbf{s}}(\mathbf{s})} \mathrm{d}\mathbf{s}\end{aligned} \tag{3.2}$$

in which $\mathbb{D}(\cdot \| \cdot)$ is the Kullback-Leibler divergence.

Typically, when there is a statistical dependency between the entries of $\mathbf{s}$, compressing the vector based on the exact distribution is often intractable. Thus, the common strategy is to expand the vector in some other basis and to then do the compression entry-wise (neglecting the dependency between entries of the transformed vector). This is equivalent to doing the compression assuming that the signal distribution is the product of the marginal distributions. Thus, if the transformed vector is $\mathbf{y} = \mathbf{H}\mathbf{s}$, then the normalized redundant information remaining in the compressed signal is

$$\begin{aligned}\mathrm{R}(\mathbf{H}) &= \frac{1}{N} \left(\mathbb{H}(p_{y_1}(y_1) \cdots p_{y_N}(y_N)) - \mathbb{H}(p_{\mathbf{s}}) \right) \\ &= \frac{1}{N} \mathbb{D}(p_{\mathbf{y}}(\mathbf{y}) \| p_{y_1}(y_1) \cdots p_{y_N}(y_N)),\end{aligned} \tag{3.3}$$

---

[1]$\mathbb{H}(\cdot)$ is used for the random variable or its pdf interchangeably.

where N is the number of entries in $\mathbf{s}$. This is the first measure of performance of the transform $\mathbf{H}$ that we use in this chapter. Also, this criterion is commonly used in ICA to find the "most-independent" representation [47].

3.2.2 Denoising Based on Non-Exact Distribution

Although the Kullback-Leibler divergence is widely used to measure the distance between two distributions, it is inherently tied to the application of compression. Here, we introduce a novel measure of divergence between distributions that is more specifically targeted to the classical denoising task. Consider the problem of estimating $\mathbf{s}$ from the noisy measurement

$$\mathbf{z} = \mathbf{s} + \mathbf{n} \tag{3.4}$$

where $\mathbf{n}$ is an N-dimensional Gaussian random vector with iid entries with variance σ^2 that is also independent from $\mathbf{s}$. Our prior knowledge is the Nth order pdf $p_{\mathbf{s}}(\cdot)$ of the signal. Under these assumptions and according to Stein [48], the optimal signal estimator that obtains minimum mean-square error (MMSE) is

$$\mathbb{E}\left[\mathbf{s}|\mathbf{z}\right] = \mathbf{z} + \sigma^2 \boldsymbol{\nabla} \log p_{\mathbf{z}}(\mathbf{z}) \tag{3.5}$$

where $\mathbb{E}\left[\mathbf{s}|\mathbf{z}\right]$ is the expected value of $\mathbf{s}$ given $\mathbf{z}$, $p_{\mathbf{z}}(\mathbf{z}) = (p_{\mathbf{s}} * p_{\mathbf{n}})(\mathbf{z})$ is the Nth order pdf of the noisy measurements, and $\boldsymbol{\nabla}$ represents the gradient operator. Thus, the MSE given $\mathbf{z}$ is

$$\mathbb{E}\left[(\mathbf{s} - \mathbb{E}\left[\mathbf{s}|\mathbf{z}\right])^2\big|\mathbf{z}\right] = \int \|\mathbf{s} - \mathbf{z}\|^2 p(\mathbf{s}|\mathbf{z})\mathrm{d}\mathbf{s} - \sigma^4 \|\boldsymbol{\nabla} \log p_{\mathbf{z}}(\mathbf{z})\|^2$$
$$= N\sigma^2 + \sigma^4 \Delta \log p_{\mathbf{z}}(\mathbf{z}). \tag{3.6}$$

where Δ is the Laplacian operator. Averaging over $\mathbf{z}$, we have

$$\begin{aligned} \mathrm{MMSE} &= N\sigma^2 - \sigma^4 \int p_{\mathbf{z}}(\mathbf{z}) \|\boldsymbol{\nabla} \log p_{\mathbf{z}}(\mathbf{z})\|^2 \, \mathrm{d}\mathbf{z} \\ &= N\sigma^2 + \sigma^4 \int p_{\mathbf{z}}(\mathbf{z}) \Delta \log p_{\mathbf{z}}(\mathbf{z})\mathrm{d}\mathbf{z}, \end{aligned} \tag{3.7}$$

However, if we apply this signal estimator based on an incorrect prior $q_{\mathbf{s}}$ (instead of the true distribution $p_{\mathbf{s}}$) as the distribution of $\mathbf{s}$, then by using (3.5)-(3.7), the MSE of

estimation becomes

$$\mathrm{MSE}(q_{\mathbf{s}}) = \mathrm{MMSE} + \sigma^4 \int p_{\mathbf{z}}(\mathbf{z}) \left\| \boldsymbol{\nabla} \log \frac{p_{\mathbf{z}}(\mathbf{z})}{q_{\mathbf{z}}(\mathbf{z})} \right\|^2 \mathrm{d}\mathbf{z} \tag{3.8}$$

where $q_{\mathbf{z}}(\mathbf{z})$ is the distribution induced on $\mathbf{z}$ in (3.4) when the distribution on $\mathbf{s}$ is $q_{\mathbf{s}}(\mathbf{s})$. Here, notice the pleasing similarity between (3.1)-(3.2) and (3.7)-(3.8).

If the entries of $\mathbf{s}$ are dependent, then the entries of $\mathbf{z}$ are dependent, too. Then, performing the exact MMSE estimator is once again often infeasible. The common scheme is then to take $\mathbf{z}$ into a transform domain, perform an entry-wise denoising (regardless of the dependency between coefficients), and map the result back into the original domain. This is justifiable when the transformation $\mathbf{H}$ is unitary because the transform-domain noise remains Gaussian iid while the ℓ_2-norm of the signal is preserved. Hence, the expected performance of this scalar denoising scheme is $\mathrm{MSE}(p_{\tilde{y}_1}(\tilde{y}_1) \cdots p_{\tilde{y}_N}(\tilde{y}_N))$ where $p_{\tilde{y}_n}(\tilde{y}_n)$ is the marginal distribution of the nth entry of $\tilde{\mathbf{y}} = \mathbf{H}\mathbf{z}$. We write this as a function of $\mathbf{H}$ normalized by the dimensionality of $\mathbf{s}$, with

$$\mathrm{MSE}(\mathbf{H}) = \frac{1}{N} \mathrm{MSE}(p_{\tilde{y}_1}(\tilde{y}_1) \cdots p_{\tilde{y}_N}(\tilde{y}_N)) \tag{3.9}$$

which is the second measure of performance that we consider in this chapter.

3.3   Continuous-Time SαS AR(1) Processes

In this section, we present a continuous-domain description of a SαS AR(1) process as the solution of a first-order stochastic differential equation. This differential formulation is central to our argumentation since it results in the identification of the operator-like wavelets, as discussed in Section 2.3.2. We also show that the continuous-domain representation is consistent with the more standard discrete AR(1) model in the sense that the latter is the sampled version of the former.

In the case of an AR(1) process, we have that

$$\mathrm{L} = \mathrm{D} + \kappa \mathrm{I} \tag{3.10}$$

where D and I are respectively the differentiator and the identity operator; then, s in (2.4) is a continuous-domain SαS AR(1) process. It follows from the theory of linear systems that the impulse response of L^{-1} is the causal exponential

$$\rho_\kappa(t) = \mathrm{e}^{-\kappa t} \mathbf{1}_+(t) \tag{3.11}$$
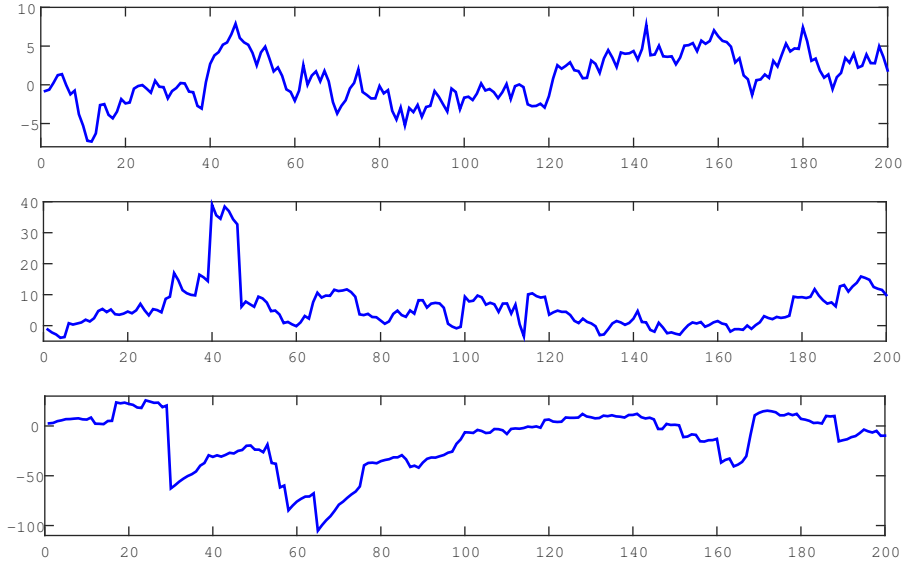
Figure 3.1: Examples of AR(1) processes for different α.

where $\mathbf{1}_+(t)$ is the unit step. Thus, as a function of t, we can write

$$s(t) = (\rho_\kappa * w)(t) \tag{3.12}$$

where $*$ denotes the continuous-domain convolution operation. The AR(1) process is well-defined for $\kappa > 0$. The limit case $\kappa = 0$ can also be handled by setting the boundary condition $s(0) = 0$, which results in a Lévy process that is non-stationary. Realizations of AR(1) processes for $\kappa = 0.05$ and for different values of α are depicted in Figure 3.1. When α decreases, the process becomes sparser in the sense that its innovation becomes more and more heavy-tailed.

3.3.1   Discretization of AR(1) Processes

Now, for a given integer k and time period T, we set

$$\varphi_k(t) = \delta(t - kT) - e^{-\kappa T}\delta\left(t - (k-1)T\right) \tag{3.13}$$

where δ is the Dirac impulse, and define w_k as

$$w_k = \langle s, \varphi_k(t)\rangle = s(kT) - e^{-\kappa T}s((k-1)T). \tag{3.14}$$

This means that the sampled version $\{s_k = s\left((k-1)T\right)\}_{k\in\mathbb{Z}}$ of $s(t)$ satisfies the first-order difference equation

$$s_k = e^{-\kappa T}s_{k-1} + w_k. \tag{3.15}$$

Also, we have that

$$w_k = \langle s, \varphi_k(t)\rangle = \langle w, (\tilde{\rho}_\kappa * \varphi_k)(t)\rangle \tag{3.16}$$

where $\tilde{\rho}_\kappa(t) = \rho_\kappa(-t)$ is the impulse response of L^{-1*} in (2.7). Also,

$$(\tilde{\rho}_\kappa * \varphi_k)(t) = \beta_{\kappa,T}(t - kT) = \mathbf{1}_{\left[kT, (k+1)T\right)} e^{-\kappa(t-kT)} \tag{3.17}$$

where $\mathbf{1}_{\left[kT, (k+1)T\right)}$ is the indicator function of the set $\left[kT, (k+1)T\right)$. The fundamental property here is that the kernels $\{\beta_{\kappa,T}(\cdot - kT)\}_{k\in\mathbb{Z}}$ are shifted replicates of each other and have compact and disjoint supports. Thus, according to the definition of a white noise in Section 2.2.2, $\{w_k\}_{k\in\mathbb{Z}}$ is an iid sequence of SαS random variables with the common characteristic function

$$\hat{p}_w(\omega) = \mathbb{E}\left[e^{j\omega\langle w, \beta_{\kappa,T}\rangle}\right] = e^{-\|\|\beta_{\kappa,T}\|_\alpha\omega\|^\alpha}. \tag{3.18}$$

The conclusion is that a continuous-domain AR(1) process maps into the discrete AR(1) process $\{s_k\}_{k\in\mathbb{Z}}$ that is uniquely specified by (3.15) and (3.18).

We now consider N consecutive samples of the process and define the random vectors $\mathbf{s} = [s_1 \ \cdots \ s_N]^\top$ and $\mathbf{w} = [w_1 \ \cdots \ w_N]^\top$.

This allows us to rewrite (3.15) as

$$\mathbf{s} = \mathbf{L}^{-1}\mathbf{w} \tag{3.19}$$

where $\mathbf{L}^{-1} = \left[\bar{l}_{ij} \right]_{N \times N}$ and

$$\bar{l}_{ij} = \mathrm{e}^{-\kappa T (j-i)} \cdot \mathbf{1}_{\{j \geq i\}} \tag{3.20}$$

which is the discrete-domain counterpart of (3.11).

In the next sections, we are going to study linear transforms applied to the signal s (or $\mathbf{s}$).

3.4 Performance Measures for AR(1) Processes

3.4.1 Mutual Information of Representation Coefficients

From now on, we assume that the signal vector $\mathbf{s} = [s_1 \; \cdots \; s_N]^\top$ with $s_k = s\left((k-1)T\right)$ is obtained from the samples of an SαS AR(1) process and satisfies the discrete innovation model (3.15). The representation of the signal $\mathbf{s}$ in (3.19) in the transform domain is denoted by $\mathbf{y} = [y_1 \; \cdots \; y_N]^\top = \mathbf{Hs}$, where $\mathbf{H} = [h_{ij}]_{N \times N}$ is the underlying orthogonal transformation matrix (e.g. DCT, wavelet transform). The idea is now to rely on Property 1 to derive the explicit form of the proposed performance criteria under the SαS hypothesis. This, in turn, will allow us to determine the optimal transform (ICA solution) based on numerical optimization.

Let us now use (3.3) to characterize the performance of a given transformation matrix $\mathbf{H}$. First, we simplify (3.3) to

$$
\begin{aligned}
\mathrm{R}(\mathbf{H}) \;&=\; \frac{1}{N} \sum_{n=1}^{N} \mathbb{H}(y_n) - \frac{1}{N} \mathbb{H}(\mathbf{y}) \\
&=\; \frac{1}{N} \sum_{n=1}^{N} \mathbb{H}(y_n) - \mathbb{H}(w_1) - \frac{1}{N} \log \det \mathbf{HL}^{-1},
\end{aligned}
\tag{3.21}
$$

where $\mathbb{H}(\cdot)$ is the differential entropy defined in (3.1). Also, we observe that $\log \det \mathbf{HL}^{-1} = 0$. In addition, since the w_m is α-stable, according to Property 1, we can write

$$y_n \overset{d}{=} \bar{h}_n w_1, \tag{3.22}$$

where $\bar{h}_n$ is the α-(pseudo)norm of the nth row of $\mathbf{HL}^{-1}$ given by

$$\bar{h}_n = \left(\sum_{r=1}^{N} \left| \sum_{m=1}^{N} h_{nm} \bar{l}_{mr} \right|^{\alpha} \right)^{\frac{1}{\alpha}}. \tag{3.23}$$

It follows that

$$\mathrm{R}(\mathbf{H}) = \frac{1}{N} \sum_{n=1}^{N} \log \bar{h}_n, \tag{3.24}$$

which can be readily calculated for any given $\mathbf{H}$.

Note 1. *This criterion is reminiscent of the sum-of-dispersion criterion $\sum_{n=1}^{N} \bar{h}_n$ which is frequently used in the study of α-stable stochastic processes [49, 50]. However, unlike (3.24), the latter dispersion criterion does not have a direct information-theoretic interpretation.*

3.4.2   Denoising-Oriented Decoupling Performance

As a second option, we use the criterion (3.9) to measure the performance of a given transform matrix $\mathbf{H}$ for the denoising task. Similar to the case in (3.21), it can be simplified to

$$\mathrm{MSE}(\mathbf{H}) = \sigma^2 - \frac{\sigma^4}{N} \sum_{n=1}^{N} \int \frac{\left(p'_{\tilde{y}_n}(\tilde{y}_n)\right)^2}{p_{\tilde{y}_n}(\tilde{y}_n)} \mathrm{d}\tilde{y}_n, \tag{3.25}$$

in which σ^2 is the noise variance and $\tilde{y}_n$ is the nth entry of

$$\tilde{\mathbf{y}} = \mathbf{Hz} = \mathbf{Hs} + \mathbf{Hn} = \mathbf{y} + \tilde{\mathbf{n}}. \tag{3.26}$$

Since $\mathbf{H}$ is a unitary matrix, $\tilde{\mathbf{n}}$ has the same distribution as $\mathbf{n}$. Also, according to (3.22),

$$\tilde{y}_n \stackrel{d}{=} \bar{h}_n w_1 + n_1, \tag{3.27}$$

where n_1 is a standard Gaussian random variable. This allows us to deduce the pdf expression

$$p_{\tilde{y}_n}(y) = \frac{1}{\bar{h}_n} p_{w_1}\left(\frac{y}{\bar{h}_n}\right) * p_{n_1}(y) \tag{3.28}$$

which involves the convolution of a rescaled SαS law with a Gaussian of standard deviation σ. Thus, (3.25) is calculable through one-dimensional integrals.

3.5 Optimization of the Representation Basis

3.5.1 Gradient of the Measures

Based on equations (3.24) and (3.25), we can now attempt to find the optimal transformation $\mathbf{H}_{\text{ICA}}$ by minimizing these expressions over the space of all orthonormal matrices of size N.

To guide this optimization process, we first derive the gradient of the cost functions R and MSE with respect to $\mathbf{H}$. Specifically, according to (3.23) and (3.24), the partial derivative of R($\mathbf{H}$) is

$$\frac{\partial \text{R}}{\partial h_{ij}} = \frac{1}{N\alpha \bar{h}_i^\alpha} \frac{\partial \bar{h}_i^\alpha}{\partial h_{ij}} \tag{3.29}$$

where

$$\frac{\partial \bar{h}_i^\alpha}{\partial h_{ij}} = \alpha \sum_{r=1}^{N} l_{jr} \text{sgn}\left(\sum_{n=1}^{N} h_{ik}l_{kr}\right) \left|\sum_{n=1}^{N} h_{ik}l_{kr}\right|^{\alpha-1}. \tag{3.30}$$

Also, the partial derivative of MSE($\mathbf{H}$) in (3.25) is

$$\frac{\partial \text{MSE}}{\partial h_{ij}} = -\frac{\sigma^4}{N} \frac{\partial}{\partial \bar{h}_i} \int \frac{\left(p_{\tilde{y}_i}^{(1)}(u)\right)^2}{p_{\tilde{y}_i}(u)} \mathrm{d}u \times \frac{\bar{h}_i^{1-\alpha}}{\alpha} \frac{\partial \bar{h}_i^\alpha}{\partial h_{ij}} \tag{3.31}$$

$$= -\frac{\sigma^4}{N} \left(2 \int \frac{\partial}{\partial \bar{h}_i} p_{\tilde{y}_i}^{(1)}(u) \frac{p_{\tilde{y}_i}^{(1)}(u)}{p_{\tilde{y}_i}(u)} \mathrm{d}u - \int \frac{\partial}{\partial \bar{h}_i} p_{\tilde{y}_i}(u) \left(\frac{p_{\tilde{y}_i}^{(1)}(u)}{p_{\tilde{y}_i}(u)}\right)^2 \mathrm{d}u\right)$$
$$\times \frac{\bar{h}_i^{1-\alpha}}{\alpha} \frac{\partial \bar{h}_i^\alpha}{\partial h_{ij}}$$

in which $p_{\tilde{y}_i}^{(k)}(y)$ is the kth derivative of $p_{\tilde{y}_i}(y)$ which, according to (3.28), can be written as

$$p_{\tilde{y}_i}^{(k)}(y) = p_{y_i}(y) * \frac{\mathrm{d}^k}{\mathrm{d}y^k}\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\right). \tag{3.32}$$

Also, we have that

$$\frac{\partial}{\partial \bar{h}_i} p_{\tilde{y}_i}(y) = -\frac{1}{\bar{h}_i} p_{\tilde{y}_i}(y) - \frac{y}{\bar{h}_i} p_{\tilde{y}_i}^{(1)}(y) - \frac{1}{\bar{h}_i} p_{\tilde{y}_i}^{(2)}(y) \tag{3.33}$$

and

$$\frac{\partial}{\partial \bar{h}_i} p_{\tilde{y}_i}^{(1)}(y) = -\frac{2}{\bar{h}_i} p_{\tilde{y}_i}^{(1)}(y) - \frac{y}{\bar{h}_i} p_{\tilde{y}_i}^{(2)}(y) - \frac{1}{\bar{h}_i} p_{\tilde{y}_i}^{(3)}(y). \tag{3.34}$$

Now, since the y_i have nice characteristic functions, we can calculate (3.32) efficiently through the inverse Fourier transform

$$p_{y_i}^{(k)}(y) = \mathcal{F}_\omega^{-1} \left\{ (j\omega)^k e^{-|\bar{h}_i \omega|^\alpha - \frac{\sigma^2}{2} \omega^2} \right\}(y) \tag{3.35}$$

using the FFT algorithm.

Thus, we can use gradient-based optimization to obtain the optimal transformations for different values of κ, α, and N. For our experiments, we implemented a gradient-descent algorithm with adaptive step size to efficiently find the optimal transform matrix. Since the transform matrix may deviate from the space of unitary matrices, after each step, we project it on that space using the method explained in the next subsection.

3.5.2   Projection on the Space of Unitary Matrices

Suppose that $\mathbf{A}$ is an $N \times N$ matrix. Our goal is to find the unitary matrix $\mathbf{H}^*$ that is the closest to $\mathbf{A}$ in Frobenius norm, in the sense that

$$\mathbf{H}^* = \arg \min_{\mathbf{H}} \|\mathbf{A} - \mathbf{H}\|_F. \tag{3.36}$$

According to singular-value decomposition (SVD), we can write $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top$ where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{\Lambda}$ is a diagonal matrix with nonnegative diagonal entries.

Since the Frobenius norm is unitarily invariant, we have that

$$\|\mathbf{A} - \mathbf{H}\|_F = \|\mathbf{\Lambda} - \mathbf{U}^\top \mathbf{H} \mathbf{V}\|_F \tag{3.37}$$

in which $\mathbf{U}^\top \mathbf{H} \mathbf{V}$ is a unitary matrix that we call $\mathbf{K}$. The expansion of the right-hand side

of (3.37) gives

$$
\begin{aligned}
\|\mathbf{\Lambda} - \mathbf{K}\|_F^2 &= \sum_{1 \leq i,j \leq N} k_{ij}^2 + \sum_{i=1}^{N} \lambda_{ii}^2 - 2 \sum_{i=1}^{N} \lambda_{ii} k_{ii} \\
&= N + \sum_{i=1}^{N} \lambda_{ii}^2 - 2 \sum_{i=1}^{N} \lambda_{ii} k_{ii}.
\end{aligned}
\tag{3.38}
$$

Since $\mathbf{K}$ is unitary, $|k_{ii}| \leq 1$ for $i = 1, \ldots, N$. Thus, setting $k_{ii} = 1$, which means setting $\mathbf{K} = \mathbf{I}$, minimizes (3.38). Consequently, the projection of $\mathbf{A}$ on the space of unitary matrices is $\mathbf{H}^* = \mathbf{U}\mathbf{V}^\top$.

3.5.3 Optimization Algorithm

Given the measure of independence C (i.e., R or MSE), the algorithm is as follows:

Algorithm 2: Steepest-Descent Algorithm with Adaptive Step-Size to Apply ICA to Discrete SαS AR(1) Processes

---
1: **input:** N, α, κ
2: **initialize:** $\mathbf{H}_{\text{old}}, \mu, a \in [1, +\infty)$ and $b \in [0, 1]$
3: **repeat**
4:     $\widetilde{\mathbf{H}}_{\text{new}} = \mathbf{H}_{\text{old}} - \mu \nabla C|_{\mathbf{H}_{\text{old}}}$
5:     Set $\mathbf{H}_{\text{new}}$ to the projection of $\widetilde{\mathbf{H}}_{\text{new}}$ onto the space of unitary matrices
6:     **if** $C(\mathbf{H}_{\text{new}}) < C(\mathbf{H}_{\text{old}})$ **then**
7:        $\mathbf{H}_{\text{old}} \leftarrow \mathbf{H}_{\text{new}}$
8:        $\mu \leftarrow a \cdot \mu$
9:     **else**
10:       $\mathbf{H}_{\text{new}} \leftarrow \mathbf{H}_{\text{old}}$
11:       $\mu \leftarrow b \cdot \mu$
12:     **end if**
13: **until** convergence
14: **return** $\mathbf{H}_{\text{new}}$

Algorithm 3 can be viewed as a model-based version of ICA. We take advantage of the underlying stochastic model to derive an optimal solution based on the minimization

of (3.24) and (3.25), which involves the computation of ℓ_α-norms of the transformation matrix. By contrast, the classical version of ICA is usually determined empirically based on the observations of a process, but the ultimate aim is similar; namely, the decoupling of the data vector.

3.6 Numerical Results and Discussion

The majority of experiments on ICA published in the literature are data-driven. The present formulation, by contrast, is model-based so that it does not require the generation of signal samples. To make an analogy, it is to ICA what the Karhunen-Loève transform is to principal components (PCA). We can therefore rely on (3.24)-(3.25) to compute the performance of a transform analytically. Also, the optimal transform (referred to as ICA) is found numerically by running Algorithm 2. We recall that our theoretical figures of merit are relevant to practical signal processing: the first (mutual information) gives in a direct measure of the coding gain in a compression experiment, while the second measures the signal-to-noise ratio (SNR) improvement for signal denoising, as justified in Section 3.2.2.

Also, notice that the wavelet coefficients of the signal s are

$$
\begin{aligned}
v_{ij} = \langle s, \psi_{ij} \rangle &= \langle \mathrm{L}^{-1}w, \psi_{ij} \rangle \\
&= \langle w, \mathrm{L}^{-1*}\mathrm{L}^*\phi_i(\cdot - 2^i jT) \rangle = \langle w, \phi_i(\cdot - 2^i jT) \rangle.
\end{aligned}
\tag{3.39}
$$

where we have $\psi_{ij} = \mathrm{L}^*\phi_i(\cdot - 2^i jT)$. Based on this equality, according to (3.18), we understand that, for any given i and for all j, the v_{ij} follows an SαS distribution with dispersion parameter $\|\phi_i\|_\alpha^\alpha$. Also, since w is independent at every point, intuitively, the level of decoupling has a direct relation to the overlap of the smoothing kernels $\phi_i(\cdot - 2^i jT)$. For the operator-like wavelets decdibed in Section 2.3.2, the supports of $\phi_i(\cdot - 2^i jT)$ do not overlap within the given scale i. Thus, the wavelet coefficients at scale i are independent and identically distributed. This property suggests that this type of transform is an excellent candidate for decoupling AR(1) processes. The illustration of plugging these wavelets into (3.39) is given in Figure 3.2.

Initially, we investigate the effect of the signal length N on the value of R and MSE. We consider the case of a Lévy process (i.e., $\kappa = 0$) and numerically optimize the criteria for different α and plot it as a function of N. Results are depicted in Figure 3.3. As we see, the criteria values converge quickly to their asymptotic values. Thus, for the remainder of the experiments, we choose $N = 64$. This is a block size that is reasonable computationally and large enough to be representative of the asymptotic regime.
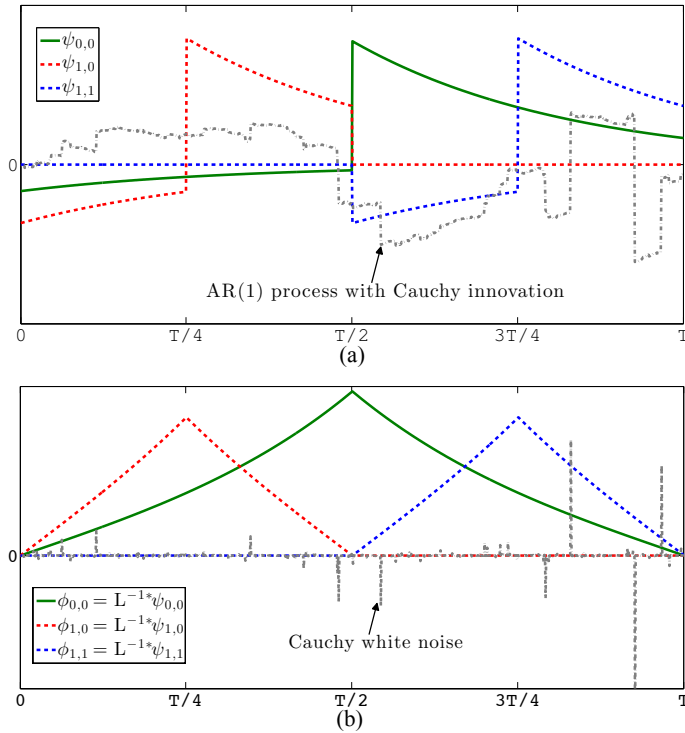
Figure 3.2: Two equivalent interpretations of the wavelet analysis of a sparse process. (a) Operator-like wavelets at two consecutive scales acting on an Cauchy AR(1) process. (b) The equivalent windows (smoothing kernels) acting on the underlying Cauchy white noise. Note that $\psi_{1,0}$ and $\psi_{1,1}$ ($\phi_{1,0}$ and $\phi_{1,1}$, respectively) are non-overlapping.

Then, we investigate the performance of different transforms for various processes. First, we focus on the Lévy processes. In this case, the operator-like wavelet transform is the classical Haar wavelet transform (HWT). The performance criteria R and MSE as a function of α for various transforms are plotted in Figures 3.4 and 3.5, respectively. The considered transformations are as follows: identity as the baseline, discrete cosine transform (DCT), Haar wavelet transform (HWT), and optimal solution (ICA) provided by the
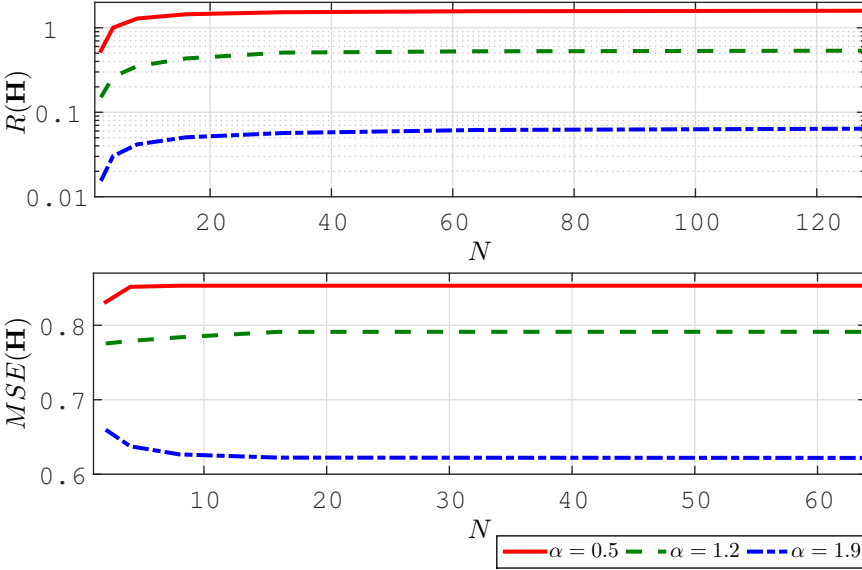
Figure 3.3: Minimum value of R($\mathbf{H}$) and MSE($\mathbf{H}$) for Lévy processes as a function of N for different values of α. In the second plot $\sigma^2 = 1$.

proposed algorithm. In the case of $\alpha = 2$ (Gaussian scenario), the process s is a Brownian motion whose KLT is a sinusoidal transform that is known analytically [51]. In this case, the DCT and the optimal transform converge to the KLT since being decorrelated is equivalent to being independent. We see this coincidence in both Figures 3.4 and 3.5. The vanishing of R at $\alpha = 2$ indicates perfect decoupling. By contrast, as α decreases, neither the DCT nor the optimal transform decouples the signal completely. The latter means that there is no unitary transform that completely decouples stable non-Gaussian Lévy processes. However, we see that, based on both criteria R and MSE, and as α decreases, the DCT becomes less favorable while the performance of the HWT gets closer to the optimal one. Moreover, Figures 3.4 and 3.5 even suggest that the Haar wavelet transform is equivalent to the ICA solution for $\alpha \leq 1$.
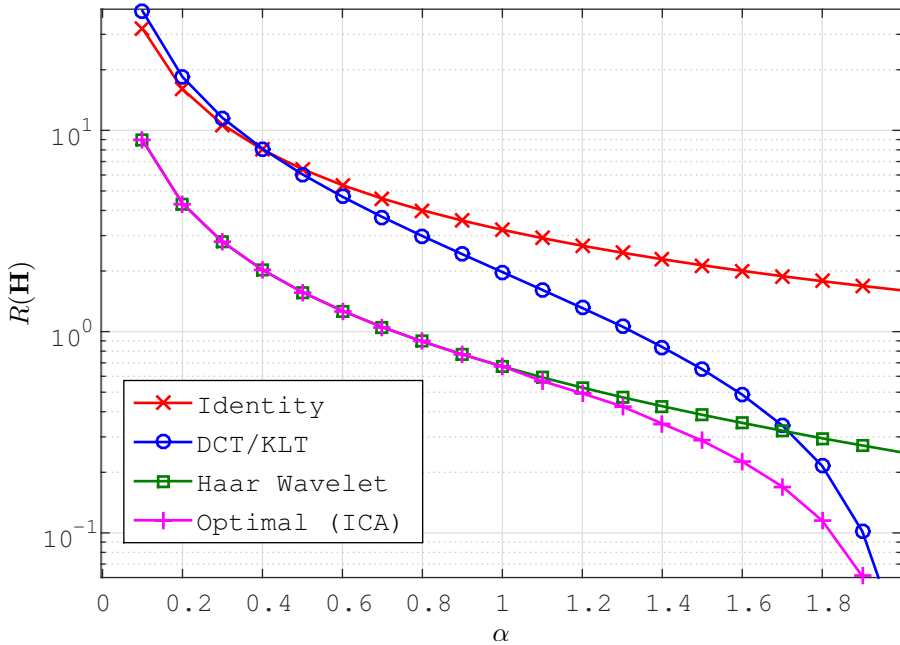
Figure 3.4: R($\mathbf{H}$) of Lévy processes versus α when $N = 64$ for different $\mathbf{H}$.

Also, to see the transition from sinusoidal bases to Haar wavelet bases, we plot the optimal basis which is obtained by the proposed algorithm at two consequent scales. In Figure 3.6, we see the progressive evolution of the ICA solution from the sinusoidal basis to the Haar basis while changing the parameter α of the model.

Next, we consider a stationary AR(1) process with $e^{-\kappa T} = 0.9$ and $n = 64$. For $\alpha = 2$, we get the well-known classical Gaussian AR(1) process for which the DCT is known to be asymptotically optimal [2, 3]. For such a process, the operator-like wavelet is known before hand and given by (2.34). The performance criterion R versus α for the DCT, the HWT, the operator-like wavelet matched to the process, and the optimal ICA solution are plotted in Figure 3.7. Here too we see that, for $\alpha = 2$, ICA is equivalent the DCT. But, as α decreases, the DCT loses its optimality and the matched operator-like wavelet becomes closer to optimum. Again, we observe that, for $\alpha \leq 1$, the ICA
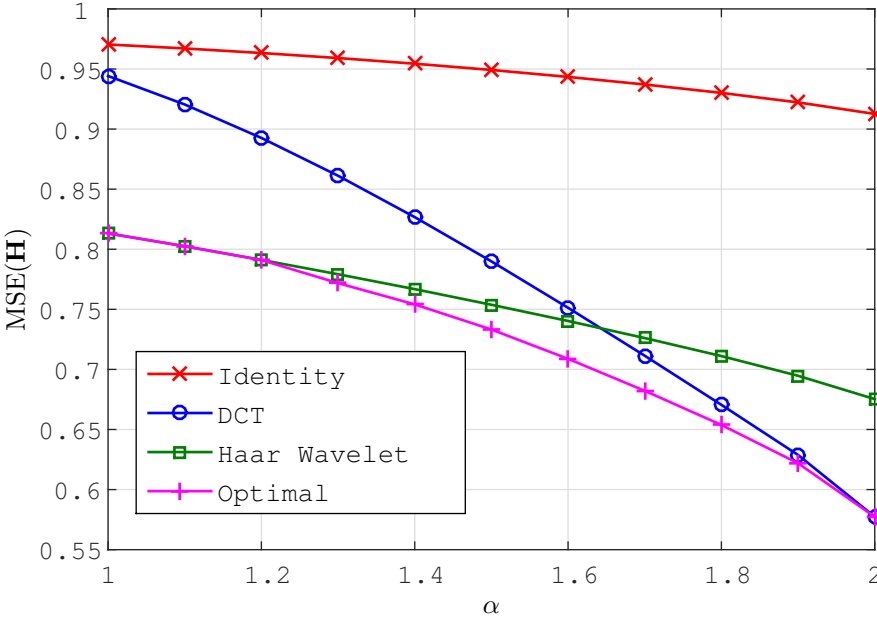
Figure 3.5: MSE($\mathbf{H}$) of Lévy processes versus α when $N = 64$ for different $\mathbf{H}$ when $\sigma^2 = 1$.

solution is the matched operator-like wavelet described in Section 2.3.2. The fact that the matched operator-like wavelet outperforms the HWT shows the benefit of the tuning of the wavelet to the differential characteristics of the process. Also, as shown in Figure 3.8, experimentally determined ICA basis functions for $\alpha = 1$ are indistinguishable from the wavelets in Figure 3.2.

To substantiate those findings, we present a theorem that states that, based on the above mentioned criteria and for any $\alpha < 2$, the operator-like wavelet transform outperforms the DCT (or, equivalently, the KLT associated with the Gaussian member of the family) as the block-size N tends to infinity.

α
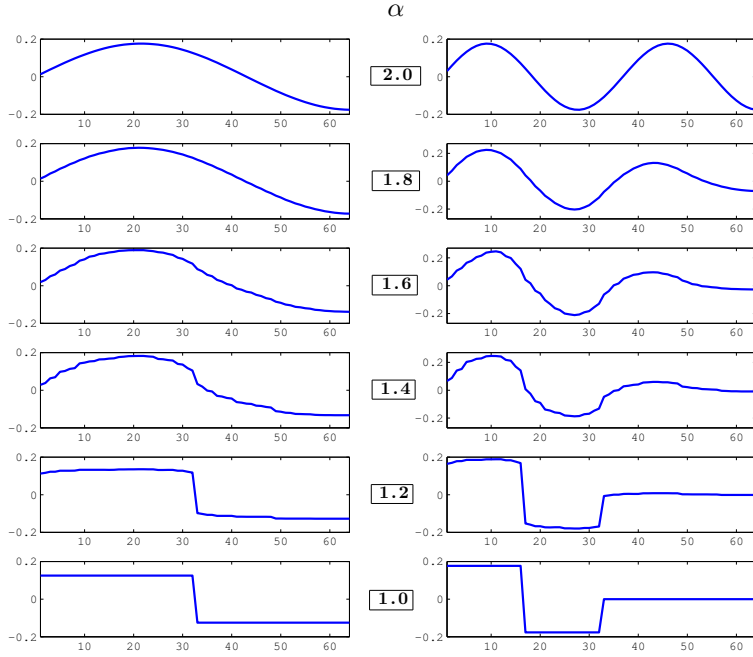
Figure 3.6: Two rows of the optimal **H** (ICA) for $\alpha = 2$ down to 1 when $N = 64$. In each row, we see the evolution from sinusoidal waves to Haar wavelets by increasing the sparsity of the underlying innovation process.

Theorem 1. *If $\alpha < 2$ and $\kappa \geq 0$, we have that*

$$\lim_{N \to \infty} R(OpWT) < \lim_{N \to \infty} R(DCT) = \infty \tag{3.40}$$

and

$$\lim_{N \to \infty} MSE(OpWT) < \lim_{N \to \infty} MSE(DCT) = \sigma^2, \tag{3.41}$$

where OpWT stands for the operator-like wavelet transform.
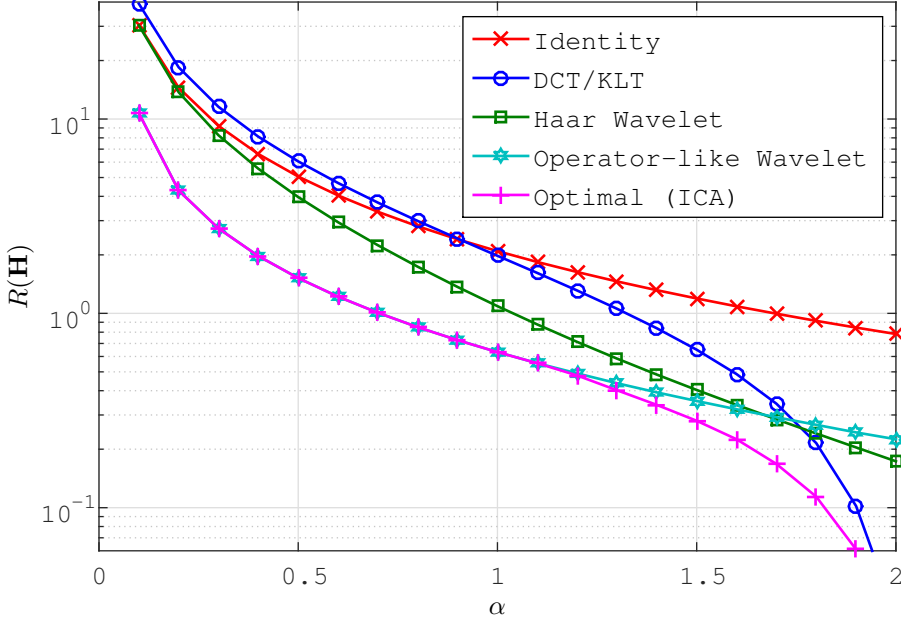
Figure 3.7: R(**H**) versus α when $e^{-\kappa T} = 0.9$ and $N = 64$ for different **H**.

The proof is given in Appendix 3.A.

In addition, this theorem states that, for $\alpha < 2$ and as N tends to ∞, the performance of the DCT is equivalent to the trivial identity operator. This is surprising because, since the DCT is optimal for the Gaussian case ($\alpha = 2$), one may expect that it has a good result for other AR(1) processes. However, although this theorem does not assert that operator-like wavelets are the optimal basis, it still shows that, by applying them, we obtain better performance than trivial transformations. Also, through simulations, we observed that operator-like wavelets are close to optimal transform as α gets smaller. In such extreme scenari, the probabilities densities of the signal and of its transformed-domain coefficients are extremely heavy-tailed which conforms with a statistical notion of sparsity [11, 12].
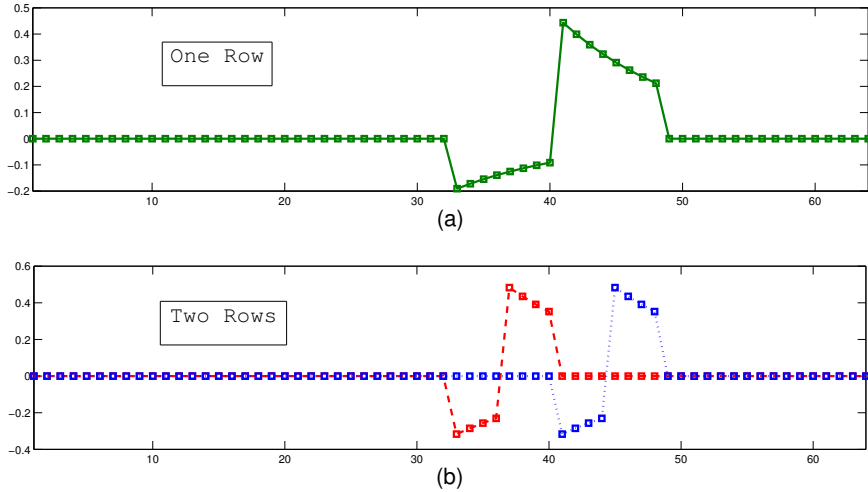
Figure 3.8: Three rows of the optimal **H** for $\alpha = 1$ and $N = 64$. Parts (a) and (b) show the dyadic structure of the wavelets.

It is worth mentioning that, in addition to the gain in performance, operator-like wavelets are cheaper to compute than the DCT. They can be implemented with the same type of filter-bank algorithm as the Haar transform, the only difference being that the filters are scale-dependent. The resulting cost is of $\mathcal{O}(N)$ (two operations per coefficient) which compares favorably with the $\mathcal{O}(N \log N)$ of the DCT. Using operator-like wavelets is also immensely more efficient than deploying the full ICA machinery. The latter requires the estimation of the transform and then its full matrix computation ($\mathcal{O}(N^2)$) which cannot benefit from any acceleration due to lack of structure.

3.7 Summary

In this chapter, we focused on the simplest version (first-order differential system with an SαS excitation) of the sparse stochastic processes which have been proposed by Unser et

al [17, 18]. Because of the underlying innovation model and the properties of SαS random variables, we could obtain a closed-form formula for the performance of different transform-domain representations and characterize the optimal transform. This is a novel model-based point of view for ICA. We proved that operator-like wavelets are better than sinusoidal transforms for decoupling the sparse AR(1) processes ($\alpha < 2$). This result is remarkable since sinusoidal bases are known to be asymptotically optimal for the classical case of $\alpha = 2$ [2, 3]. Moreover, we showed that, for very sparse excitations ($\alpha \lesssim 1$), operator-like wavelets are equivalent to the ICA. As far as we know, this is the first theoretical results on the optimality of wavelet-like bases for a given class of stochastic processes.

Another interesting aspect of this study is that it gives a unified framework for Fourier-type transforms and a class of wavelet transforms. Now, the Fourier transform and the wavelet transforms were based on two different intuitions and philosophies. However, here we have a model in which we obtain both transform families just by changing the underlying parameters.

The next step in this line of research is to investigate the extent to which these findings can be generalized to other white noises or higher-order differential operators. Also, studying the problem in the original continuous domain would be theoretically very valuable.

3.A   Proof of Theorem 1

Proof of Part 1 (Equation (3.40))

According to (3.24), we have that

$$
\begin{aligned}
\mathbf{R}(\mathbf{H}) &= \frac{1}{N} \sum_{n=1}^{N} \log \bar{h}_n = \frac{1}{N} \sum_{n=1}^{N} \log \left(\frac{1}{\bar{h}_n^{-1}} \right) \\
&= \int_{\mathbb{R}} \log \left(\frac{1}{\gamma} \right) p\left(\gamma \right) \mathrm{d}\gamma
\end{aligned}
\tag{3.42}
$$

in which $p(\cdot)$ is the empirical distribution of $\bar{h}_n^{-1}$.

According to SVD, we can write $\mathbf{L}^{-1} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^\top$ where $\mathbf{\Lambda}$ is a diagonal matrix with λ_i as diagonal entries. Taking $\mathbf{s}$ in the KLT domain is equivalent to multiplying it by $\mathbf{U}^\top$. The eigenvalues of the covariance of AR(1) matrices are known in closed form and are given

by [52] and [53], for $\kappa \geq 0$, as

$$|\lambda_i|^{-1} = \sqrt{\left(1 - e^{-\kappa T}\right)^2 + 4e^{-\kappa T} \sin^2\left(\frac{\omega_i}{2}\right)} \qquad (3.43)$$

and

$$v_{ij} = \frac{\sqrt{2}}{\sqrt{N + \left(1 - e^{-2\kappa T}\right)\lambda_i^2}} \, \sin\left(\omega_i\left(j - \frac{N+1}{2}\right) + i\frac{\pi}{2}\right) \qquad (3.44)$$

in which ω_i, $i = 1, \ldots, N$, is the ith positive root of

$$\tan\left(N\omega\right) = -\frac{\left(1 - e^{-2\kappa T}\right)\sin\omega}{\cos\omega - 2e^{-\kappa T} + e^{-2\kappa T}\cos\omega}. \qquad (3.45)$$

Since $\tan\left(N\omega\right)$ is an injective function that sweeps the whole domain of the real numbers while $\omega \in \left[\frac{i-1}{N}\pi, \frac{i}{N}\pi\right]$, for $i = 1, \ldots, N$, (3.45) has a single root in each of such intervals. Thus, as N tends to infinity, the empirical distribution of the ω_i tends to the uniform distribution on $[0, \pi]$. Then, starting from (3.43), one can obtain the limit empirical distribution of $|\lambda_i|$ as

$$p_\lambda(\lambda) = \frac{2}{\pi} \frac{\lambda}{\sqrt{\lambda^2 - \left(1 - e^{-\kappa T}\right)^2}\sqrt{\left(1 + e^{-\kappa T}\right)^2 - \lambda^2}}. \qquad (3.46)$$

Now, $\sum_{j=1}^{N} v_{ij}^2 = 1$ means that

$$\sum_{j=1}^{N} \left|\sin\left(\omega_i\left(j - \frac{N+1}{2}\right) + i\frac{\pi}{2}\right)\right|^2 \sim \mathcal{O}\left(N\right) \qquad (3.47)$$

as N tends to infinity. But, for $\alpha < 2$, we have that

$$\left(\sum_{j=1}^{N} \left|\sin\left(\omega_i\left(j - \frac{N+1}{2}\right) + i\frac{\pi}{2}\right)\right|^\alpha\right)^{\frac{1}{\alpha}} \geq \qquad (3.48)$$

$$\left(\sum_{j=1}^{N} \left|\sin\left(\omega_i\left(j - \frac{N+1}{2}\right) + i\frac{\pi}{2}\right)\right|^2\right)^{\frac{1}{\alpha}} \sim \mathcal{O}(N^{\frac{1}{\alpha}}).$$

Thus, for $\alpha < 2$, $\left(\sum_{j=1}^{N} |v_{ij}|^{\alpha} \right)^{\frac{1}{\alpha}}$ grows faster than $\mathcal{O}(N^{\frac{1}{\alpha} - \frac{1}{2}})$ and thus tends to infinity as N tends to infinity. Consequently, the limit empirical distribution of $\bar{h}_i^{-1}$ can be represented as

$$p(\gamma) = \begin{cases} \frac{2}{\pi} \dfrac{\gamma}{\sqrt{\gamma^2 - (1 - e^{-\kappa T})^2} \sqrt{(1 + e^{-\kappa T})^2 - \gamma^2}} & \alpha = 2 \\ \delta(\gamma) & \alpha \neq 2. \end{cases} \tag{3.49}$$

By plugging this result into (3.42), we conclude that, for $\alpha < 2$, $\lim_{N \to \infty} R(\text{KLT}) = \infty$. This completes the proof of the right-hand side.

Now, for the proof of the left-hand side, we need to specify the matrix $\mathbf{H}$ for the operator-like wavelet transform. This matrix is given by the recursive construction

$$\mathbf{H}_k = \begin{bmatrix} \dfrac{\sqrt{1 - e^{-2\kappa T}}}{\sqrt{1 - e^{-2^{k+1}\kappa T}}} & 0 & \vec{0}_{k-1} \\ 0 & \dfrac{\sqrt{1 - e^{-2\kappa T}}}{\sqrt{1 - e^{-2^{k+1}\kappa T}}} & \vec{0}_{k-1} \\ \vec{0}_{k-1}^{\top} & \vec{0}_{k-1}^{\top} & \mathbf{I}_{k-1} \end{bmatrix} \times \begin{bmatrix} \ell_{k-1} & e^{-2^{k-1}\kappa T} \ell_{k-1} \\ -e^{-2^{k-1}\kappa T} \ell_{k-1} & \ell_{k-1} \\ \mathbf{H}'_{k-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{H}'_{k-1} \end{bmatrix} \tag{3.50}$$

in which $\vec{0}_{k-1}$ is the 2^{k-1}-dimension zero row vector, $\mathbf{I}_{k-1}$ is the $2^{k-1} \times 2^{k-1}$ identity matrix, $\mathbf{H}'_{k-1}$ is the matrix $\mathbf{H}_{k-1}$ omitting the first row and

$$\ell_{k-1} = \begin{bmatrix} 1 & e^{-\kappa T} & \dots & e^{-(2^{k-1}-1)\kappa T} \end{bmatrix}. \tag{3.51}$$

Also, $\mathbf{H}_0 = [1]$. Let us denote the empirical distribution of $\bar{h}_i^{-1}$ (the reciprocal of the α-(pseudo) norm of the rows of $\mathbf{H}_k \mathbf{L}_{2^k}$) by $p_k(\gamma) = \sum_{i=1}^{k} p_i \delta(\gamma - \gamma_i)$. Now, for the sequence of p_i and γ_i, with respect to k, we have the following recursive relation:

- Replace p_{k-1} by $\left(\frac{p_{k-1}}{2}, \frac{p_{k-1}}{2} \right)$

- Remove γ_{k-1}. Then, if $\kappa > 0$, set

$$\gamma_{k-1} = \frac{\sqrt{1 - e^{-2^{k+1}\kappa T}}}{\sqrt{1 - e^{-2\kappa T}}} \left(\sum_{i=-2^{k-1}+1}^{2^{k-1}} \left(\frac{e^{-|i|\kappa T} - e^{-(2^k - |i|)\kappa T}}{1 - e^{-2\kappa T}} \right)^{\alpha} \right)^{-\frac{1}{\alpha}} \tag{3.52}$$

and

$$\gamma_k = \frac{\sqrt{1 - e^{-2^{k+1}\kappa T}}}{\sqrt{1 - e^{-2\kappa T}}} \left(\sum_{i=1}^{2^k} \left(\frac{1 - e^{-2i\kappa T}}{1 - e^{-2\kappa T}} \right)^\alpha \right)^{-\frac{1}{\alpha}} \tag{3.53}$$

else, if $\kappa = 0$, set

$$\gamma_{k-1} = 2^{\frac{k}{2}} \left(\sum_{i=-2^{k-1}+1}^{2^{k-1}} \left(2^{k-1} - |i| \right)^\alpha \right)^{-\frac{1}{\alpha}} \tag{3.54}$$

and

$$\gamma_k = 2^{\frac{k}{2}} \left(\sum_{i=1}^{2^k} i^\alpha \right)^{-\frac{1}{\alpha}}. \tag{3.55}$$

Consequently, according to (3.42), we have that

$$\lim_{n \to \infty} \text{R(HWT)} = \sum_{k=1}^{\infty} 2^{-k} \log \gamma_k^{-1}. \tag{3.56}$$

However, for the case $\kappa > 0$ and $k < N$,

$$\begin{aligned}
\gamma_k^{-1} &\leq \frac{2 \left(\left(2^k - 1 \right) \left(1 - e^{-2^k \kappa T} \right)^\alpha \right)^{\frac{1}{\alpha}}}{\sqrt{\left(1 - e^{-2\kappa T} \right) \left(1 - e^{-2^{k+1}\kappa T} \right)}} \\
&\leq \frac{2}{\sqrt{1 - e^{-2\kappa T}}} \frac{\sqrt{1 - e^{-2^k \kappa T}}}{\sqrt{1 + e^{-2^k \kappa T}}} \left(2^k - 1 \right)^{\frac{1}{\alpha}} \\
&\leq \frac{2^{1+\frac{k}{\alpha}}}{\sqrt{1 - e^{-2\kappa T}}}. \tag{3.57}
\end{aligned}$$

Thus,

$$\begin{aligned}
\lim_{n \to \infty} \text{R(HWT)} &\leq \sum_{k=1}^{\infty} 2^{-k} \log \frac{2^{1+\frac{k}{\alpha}}}{\sqrt{1 - e^{-2\kappa T}}} \\
&= \left(\frac{2}{\alpha} + \frac{1}{2} \log \frac{1}{1 - e^{-2\kappa T}} \right) \log 2. \tag{3.58}
\end{aligned}$$

For the case $\kappa = 0$ and $k < N$,

$$\gamma_k^{-1} \leq 2^{-\frac{k}{2}} \left(\left(2^k - 1\right) \left(2^{k-1}\right)^\alpha \right)^{\frac{1}{\alpha}} \leq 2^{\frac{k}{2} + \frac{k}{\alpha} - 1}. \tag{3.59}$$

Thus,

$$\lim_{n \to \infty} \mathrm{R(HWT)} \leq \sum_{k=1}^{\infty} 2^{-k} \log 2^{\frac{k}{2} + \frac{k}{\alpha} - 1} = \frac{2}{\alpha} \log 2. \tag{3.60}$$

Therefore, the proof is complete.

Proof of Part 2 (Equation (3.41))

Proof: We have that

$$\mathrm{MSE}(\mathbf{H}) = \frac{1}{N} \sum_{n=1}^{N} \nu(\bar{h}_n^{-1}) = \int_{\mathbb{R}} \nu(\gamma^{-1}) p(\gamma) \, \mathrm{d}\gamma \tag{3.61}$$

in which $\nu(\gamma^{-1})$ is the MMSE of the estimating w from s in the scalar problem

$$s = \gamma^{-1} w + z, \tag{3.62}$$

where w is a stable random variable with characteristic function $\hat{p}_w(\omega) = \exp\left(-|\omega|^\alpha\right)$ and z is a Gaussian random variable with variance σ^2. We know that $\nu(\cdot)$ is a monotone continuous function that vanishes at zero and tends to σ^2 asymptotically. Also, $p(\cdot)$ is the empirical distribution of the reciprocals of $\bar{h}_i$ in (3.23). The proof is then essentially the same as the one of Theorem 1 but simpler since the function $\nu(\cdot)$ is bounded.

For $\mathbf{H}$ equal to Fourier transform, the limiting $p(\gamma)$ was given in (3.49). Thus, for $\alpha < 2$, as n tends to infinity, $\mathrm{MSE}(\mathbf{H})$ tends to σ^2. This completes the proof of the right-hand side.

For the case that $\mathbf{H}$ is the operator-like wavelet transform, the limmit is $p(\gamma) = \sum_{k=1}^{\infty} p_k \delta \left(\gamma - \gamma_k\right)$ where $p_k = 2^{-k}$ and γ_k were given in (3.52) – (3.55). Thus, we have that

$$\mathrm{MSE(OpWT)} = \sum_{k=1}^{\infty} 2^{-k} \nu(\gamma_k^{-1}) \leq \frac{1}{2} \nu(\gamma_1^{-1}) + \frac{\sigma^2}{2}. \tag{3.63}$$

But, obviously, $\gamma_1^{-1} < \infty$; hence, $\nu(\gamma_1^{-1}) < \sigma^2$, which completes the proof.

Chapter 4

Denoising Performance of Wavelets for Self-Similar SαS Processes

In this chapter, we investigate the performance of wavelet shrinkage methods for the denoising of 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 to 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 $\alpha = 2$.

4.1   Introduction

In the previous chapter, we showed that operator-like wavelets are optimal for representation of SαS AR(1) processes for the compression and denoising applications. Also, we observed that conventional wavelets, specifically Haar wavelets exhibit a near optimal performance. In this chapter, we are going to study the transform-domain performance of denoising of another important classes of stochastic processes i.e. self-similar processes. Inspired by the result of the previous chapter, we focus on the wavelet domain denoising performance. Also, it makes the mathematical calculations doable. In contrast with the previous section, we study these processes in the continuous domain rather than discretizing them.

A classical example of a self-similar process is the fractional Brownian motion (fBm) [54]. It can be interpreted as the fractional integral of a continuous-domain white Gaussian noise (a.k.a. innovation) [16, Section 7.5.2]. The order of integration γ provides a direct control of the degree of fractality, which explains why the fBm is a popular model of real-world signals such as images [55, 56], traffic in communication networks [57], and financial processes [58]. Higher-order generalizations of an fBm are characterized in [59].

The non-Gaussian counterpart of an fBm is the fractional stable motion [60, 61], 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 the properties of long-range dependency and self-similarity and enjoys a wide range of applications [62, 23].

As discussed in Section 2.2.2, 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. Again, this is very relevant to modern signal processing and to the development of algorithms for the recovery of sparse signals, including compressed sensing [63, 64].

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 [65, 66, 67, 68]. Regarding the α-stable processes, there is some prior work on the determination of the statistical distribution of wavelet coefficients [69], 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 $\alpha < 1$ which is presented in the previous

chapter. 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 [7, 70], signal reconstruction [71, 72], and denoising [73, 74, 75]. This is reinforced by fundamental results from approximation theory on the optimality of wavelets for the N-term approximation of functions in Besov spaces [13].

Our objective in this chapter 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 [76, 77, 78, 79]. 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 γth 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 [29].

Similar to the discussions in Section 3.2.2, 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 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 [80]. Here, the fact that the variance of the signal is unbounded requires a more technical treatment.

The chapter is organized as follows: In Section 4.2, we review the properties of self-similar SαS processes. Our signal-estimation problem is then formulated in Section 4.3. In Section 4.4, we address the issue of the calculation of the average energy of a signal from its wavelet coefficients. In Section 4.5, 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 chapter is presented in Section 4.6. In Section 4.7, 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 4.8.

4.2   Fractional SαS Lévy Processes

We now start with some preliminaries on self-similar SαS processes that are necessary for understanding the chapter. We again use the framework introduced in [17, 18, 16] 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 [60].

We recall the notion of SαS white noise which was described in Section 2.2.2. Suppose that w is an SαS white noise with the Lévy exponent $-|a\omega|^\alpha$ for a given $a > 0$. A process s is a self-similar SαS process of order $\gamma \geq 0$ if

$$\mathrm{D}^\gamma s = w, \tag{4.1}$$

where D^γ is the γth order derivative operator that is defined as [16, Chapter 7.5]

$$\mathrm{D}^\gamma\{f\}(x) = \frac{1}{2\pi}\int_\mathbb{R}(\mathrm{j}\omega)^\gamma \hat{f}(\omega)\mathrm{e}^{\mathrm{j}\omega x}\mathrm{d}\omega. \tag{4.2}$$

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

$$\langle s, \theta \rangle = \langle \mathrm{D}^{-\gamma}w, \theta \rangle = \langle w, \mathrm{D}^{-\gamma*}\theta \rangle \tag{4.3}$$

in which $\mathrm{D}^{-\gamma*}$ is the L_α-stable adjoint of the inverse operator of D^γ that is defined by (see [16, Chapter 7.5])

$$\mathrm{D}^{-\gamma*}\{\theta\}(x) = \frac{1}{2\pi}\int_\mathbb{R}\frac{\hat{\theta}(\omega) - \sum_{k=0}^{\lfloor\gamma+\frac{1}{\alpha}\rfloor-1}\frac{\hat{\theta}^{(k)}(0)\omega^k}{k!}}{(\mathrm{j}\omega)^\gamma}\mathrm{e}^{\mathrm{j}\omega x}\mathrm{d}\omega \tag{4.4}$$

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

$$\mathrm{D}^{-\gamma*}\{\theta\}(x) = \int_\mathbb{R} K(x, t)\theta(t)\mathrm{d}t, \tag{4.5}$$

where the kernel of the inverse operator is

$$K(x,t) = \frac{1}{\Gamma(\gamma)}(t-x)_+^{\gamma-1} - \sum_{k=0}^{\lfloor \gamma + \frac{1}{\alpha} \rfloor - 1} \frac{t^k}{k!} \frac{(-x)_+^{\gamma-1-k}}{\Gamma(\gamma-k)} \tag{4.6}$$

in which $\Gamma(\cdot)$ is Euler's Gamma function where $(x)_+ = \max\{0,x\}$. Consequently, according to (2.13) and (4.3), $\langle s, \theta \rangle$ is an SαS random variable with the characteristic function

$$\hat{p}_{\langle s,\theta \rangle}(\omega) = \exp\left(-|a|\|D^{-\gamma*}\theta\|_\alpha^\alpha |\omega|^\alpha\right). \tag{4.7}$$

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 ($\alpha < 2$). 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 [60, Chapter 10]. For the case $\gamma = 1$, which corresponds to Lévy processes, the sample paths are right-semicontinuous with left limit [81, Chapter 2]. On the asymptotic behavior of the sample paths, according to [81, Chapter 9] and [82, Theorem 1.3], we know that, if $\gamma \geq 1$, then there exists a constant $C_\eta \in \mathbb{R}$ for which

$$|s(x)| \leq C_\eta \left(1 + |x|\right)^\eta \tag{4.8}$$

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

We mainly need two properties to describe the effect of the operator $D^{-\gamma*}$. The first one is that the restriction of $D^{-\gamma*}$ 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 (4.4)). More precisely, when θ is in this subspace, we have that

$$D^{-\gamma*}\{\theta(c \cdot -b)\}(x) = c^{-\gamma}D^{-\gamma*}\{\theta(\cdot)\}(cx - b) \tag{4.9}$$

for any $b, c \in \mathbb{R}$. If the test functions do not have enough vanishing moments, then $D^{-\gamma*}$ 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*

$$\lim_{j \to \pm\infty} \|D^{-\gamma*}\theta(\cdot - j)\|_\alpha = \infty. \tag{4.10}$$

The proof is given in Appendix 4.A.

Lemma 1 means that the dispersion of $\langle s, \theta(\cdot - j)\rangle$ tends to infinity by letting j tend to infinity (see (4.7)). This fact results in significant simplifications in the derivation of our main result in Section 4.6.

4.3   Performance Measure of Denoising of Continuous Processes

Assume now that s is the self-similar SαS process defined by (4.1). Also, assume that z is an AWGN independent from s with variance σ^2. Then, the continuous-time stochastic process

$$\tilde{s} = s + z \tag{4.11}$$

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

$$\text{MAEE}(s, \hat{s}) = \lim_{T \to \infty} \mathbb{E}_{s,\hat{s}} \left[\frac{1}{2T} \int_{-T}^{T} \left(s(x) - \hat{s}(x) \right)^2 \mathrm{d}x \right]. \tag{4.12}$$

Now, recalling the discussions in Section 2.3.1, assume that φ and ψ are the father and mother wavelets of an orthonormal wavelet family, respectively. Then, for any $i_0 \in \mathbb{Z}$,

$$\{\varphi_{i_0 j}\}_{j \in \mathbb{Z}} \cup \left\{ \{\psi_{ij}\}_{j \in \mathbb{Z}} \right\}_{i=i_0}^{\infty} \tag{4.13}$$

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

$$\varphi_{i_0 j}(x) = 2^{\frac{i_0}{2}} \varphi(2^{i_0} x - j) \tag{4.14}$$

and

$$\psi_{ij}(x) = 2^{\frac{i}{2}} \psi(2^i x - j). \tag{4.15}$$

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 4.4, the equality

$$s(x) = \sum_{j \in \mathbb{Z}} \Pi_{i_0 j} \varphi_{i_0 j}(x) + \sum_{i \geq i_0} \sum_{j \in \mathbb{Z}} \Xi_{ij} \psi_{ij}(x) \tag{4.16}$$

holds almost everywhere for almost every realization of s when

$$\Pi_{i_0 j} = \langle s, \varphi_{i_0 j} \rangle, \tag{4.17}$$

$$\Xi_{ij} = \langle s, \psi_{ij} \rangle. \tag{4.18}$$

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

$$\widetilde{\Pi}_{i_0 j} = \langle \tilde{s}, \varphi_{i_0 j} \rangle, \tag{4.19}$$

$$\widetilde{\Xi}_{ij} = \langle \tilde{s}, \psi_{ij} \rangle, \tag{4.20}$$

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

$$\hat{s}(x) = \sum_{j \in \mathbb{Z}} \widehat{\Pi}_{i_0 j} \varphi_{i_0 j}(x) + \sum_{i \geq i_0} \sum_{j \in \mathbb{Z}} \widehat{\Xi}_{ij} \psi_{ij}(x). \tag{4.21}$$

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 chapter is to calculate $\text{MAEE}(s, \hat{s})$ 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 $\text{MAEE}_{i_0}(\varphi, \psi)$.

First, Theorem 2 tells us that the optimal coefficient-wise denoiser $\big($the denoiser that minimizes $\text{MAEE}_{i_0}(\varphi, \psi)\big)$ is the coefficient-wise MMSE estimator. Then, we establish in Theorem 3 some properties for the MMSE function of denoising an SαS random variable.

Based on these two theorems, we transform (4.12) 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 4.6 that for this denoiser the series on the right-hand side of (4.21) 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 (4.12). Therefore, the derived formula can also measure the ability of a given wavelet basis to decouple self-similar SαS stochastic processes.

4.4 Calculating Average Energy of a Signal 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 (4.16).

Consider the wavelet family defined in (4.13)-(4.14). For a function $f \in L_2(\mathbb{R})$, we have that

$$f(x) = \sum_{j \in \mathbb{Z}} \langle f, \varphi_{i_0 j} \rangle \varphi_{i_0 j}(x) + \sum_{i \geq i_0} \sum_{j \in \mathbb{Z}} \langle f, \psi_{ij} \rangle \psi_{ij}(x) \qquad (4.22)$$

and

$$\|f\|_2^2 = \sum_{j \in \mathbb{Z}} \langle f, \varphi_{i_0 j} \rangle^2 + \sum_{i \geq i_0} \sum_{j \in \mathbb{Z}} \langle f, \psi_{ij} \rangle^2. \qquad (4.23)$$

This means that we can calculate the energy of a square-integrable function from its wavelet coefficients. In Theorem 2, 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 2. *Suppose that the wavelets are such that*

$$|\varphi(x)|, |\psi(x)| \leq \eta(x) = \frac{A}{(1 + |x|)^{n+1+\epsilon}} \qquad (4.24)$$

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

$$|f(x)| \leq g(x) = B(1 + |x|)^n, \tag{4.25}$$

we have that

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(x)^2 \mathrm{d}x = \tag{4.26}$$

$$\lim_{N \to \infty} \frac{1}{2N} \left(\sum_{|j| \leq 2^{i_0} N} \langle f, \varphi_{i_0 j} \rangle^2 + \sum_{i \geq i_0} \sum_{|j| \leq 2^i N} \langle f, \psi_{ij} \rangle^2 \right).$$

The proof is given in Appendix 4.B.

This theorem helps us evaluate (4.12) for a wavelet-domain denoising method. We can also establish the complementary convergence result whose proof is given in Appendix 4.C.

Proposition 1. *Under the assumptions of Theorem 2,*

$$f(x) = \sum_{j \in \mathbb{Z}} \langle f, \varphi_{i_0 j} \rangle \varphi_{i_0 j}(x) + \sum_{i \geq i_0} \sum_{j \in \mathbb{Z}} \langle f, \psi_{ij} \rangle \psi_{ij}(x) \tag{4.27}$$

holds almost everywhere.

4.5 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

$$Y = X + Z, \tag{4.28}$$

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 [48]

$$\widehat{X} = \mathbb{E}[X|Y] = Y + \sigma^2 \frac{p'_Y(Y)}{p_Y(Y)} \tag{4.29}$$

in which

$$p_Y = p_X * p_Z. \tag{4.30}$$

There, 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 this optimal denoiser as (see (3.7)) [83, 84]

$$\mathbb{E}\big[(X - \widehat{X})^2\big] = \sigma^2 - \sigma^4 \int \frac{\big(p_Y'(y)\big)^2}{p_Y(y)} \mathrm{d}y. \tag{4.31}$$

We call this function MMSE(a, σ), where a and σ are the dispersion of the signal and the standard deviation of the noise, respectively.

In Theorem 3, we prove some properties of MMSE(a, σ) that we are going to use in the derivation of our main result. A similar function is studied in [80, 85], mostly in the scenario where X is a finite-variance random variable. The difficulty here is that SαS random variables for $\alpha < 2$ have an infinite variance.

Theorem 3. *For any fixed $\sigma \geq 0$, the following properties hold for MMSE(a, σ) as a function of the dispersion a:*

1. *Increasing over $[0, +\infty)$.*

2. *Towards infinity, we have that*

$$\lim_{a \to \infty} MMSE(a, \sigma) = \sigma^2. \tag{4.32}$$

3. *Around zero, we have that*

$$\lim_{a \to 0} \frac{MMSE(a, \sigma)}{a^{\alpha - \epsilon}} = 0 \tag{4.33}$$

    for any $\epsilon > 0$.

For the proof, refer to Appendix 4.D.
Also, notice that we trivially have that

$$\mathrm{MMSE}(a, \sigma) \leq \sigma^2 \tag{4.34}$$

for any a and σ since the identity estimator, which is $\widehat{X} = Y$, has the MSE of σ^2. This can also be deduced by using (4.31).

Additionally, we propose Conjecture 1. This conjecture is not required by our analysis but leads to a nice intuitive interpretation.

Conjecture 1. *Around zero, we have that*

$$\lim_{a \to 0} \frac{MMSE(a, \sigma)}{a^{\alpha + \epsilon}} = \infty \tag{4.35}$$

for any $\epsilon > 0$.

4.6   MAEE of Component-Wise Wavelet-Based Denoising

We now present our main result on the solution of the problem formulated in Section 4.3. 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

$$|\varphi(x)|, |\psi(x)| \leq \frac{A}{(1 + |x|)^\eta} \tag{4.36}$$

for an $\eta > \gamma + \frac{1}{\alpha}$ and $A \in \mathbb{R}$. These assumptions, along with (4.8) and Proposition 1, yield

$$s(x) = \sum_{j \in \mathbb{Z}} \langle s, \varphi_{i_0 j} \rangle \varphi_{i_0 j}(x) + \sum_{i \geq i_0} \sum_{j \in \mathbb{Z}} \langle s, \psi_{ij} \rangle \psi_{ij}(x) \tag{4.37}$$

almost surely.

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

$$\phi = \mathrm{D}^{-\gamma *} \psi. \tag{4.38}$$

As discussed in Section 4.2 and as a result of the vanishing moments of ψ, we can write

$$\phi_{ij}(x) = \mathrm{D}^{-\gamma *}\{\psi_{ij}\}(x) = 2^{i(\frac{1}{2} - \gamma)} \phi(2^i x - j). \tag{4.39}$$

Therefore, according to (4.3) and (4.7), the characteristic function of Ξ_{ij} (see also (4.18)) is

$$\hat{p}_{\Xi_{ij}}(\omega) = \exp\left(-\left|a\|\phi_{ij}\|_\alpha \omega\right|^\alpha\right). \tag{4.40}$$

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

$$\|\phi_{ij}\|_\alpha = 2^{i(\frac{1}{2} - \frac{1}{\alpha} - \gamma)} \|\phi\|_\alpha. \tag{4.41}$$

Similarly, we find that

$$\hat{p}_{\Pi_{i_0 j}}(\omega) = \exp\left(-\left|a\|\mathrm{D}^{-\gamma*}\varphi_{i_0 j}\|_\alpha \omega\right|^\alpha\right). \tag{4.42}$$

But, since φ is orthogonal to ψ, it does not have any vanishing moments. Thus, according to Section 4.2, $\mathrm{D}^{-\gamma*}\varphi_{i_0 j}$ with j varying are not shifted versions of each other. However, Lemma 1 is all what we need about them.

The combination of (4.11) with (4.19)-(4.18) implies that

$$\widetilde{\Pi}_{i_0 j} = \Pi_{i_0 j} + Z'_{i_0 j}, \tag{4.43}$$

$$\widetilde{\Xi}_{ij} = \Xi_{ij} + Z_{ij}, \tag{4.44}$$

where

$$Z'_{i_0 j} = \langle z, \varphi_{i_0 j}\rangle, \tag{4.45}$$

$$Z_{ij} = \langle z, \psi_{ij}\rangle. \tag{4.46}$$

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 4.5.

In Section 4.3, $\mathrm{MAEE}_{i_0}(\varphi, \psi)$ is the mean average error energy of the optimal coefficient-wise denoising. Thanks to Theorem 2, it is reformulated in the wavelet domain as

$$\mathrm{MAEE}_{i_0}(\varphi, \psi) =$$

$$\lim_{N\to\infty} \frac{1}{2N}\mathbb{E}\Bigg[\sum_{|j|\leq 2^{i_0}N}\left(\Pi_{i_0 j} - \widehat{\Pi}_{i_0 j}\right)^2 + \sum_{i\geq i_0}\sum_{|j|\leq 2^i N}\left(\Xi_{ij} - \widehat{\Xi}_{ij}\right)^2\Bigg]$$

$$= \lim_{N\to\infty} \frac{1}{2N}\Bigg(\sum_{|j|\leq 2^{i_0}N}\mathbb{E}\left[\left(\Pi_{i_0 j} - \widehat{\Pi}_{i_0 j}\right)^2\right] + \sum_{i\geq i_0}\sum_{|j|\leq 2^i N}\mathbb{E}\left[\left(\Xi_{ij} - \widehat{\Xi}_{ij}\right)^2\right]\Bigg). \tag{4.47}$$

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 4.5 and (4.40)-(4.42), we have that

$$\mathbb{E}\left[(\Pi_{i_0 j} - \widehat{\Pi}_{i_0 j})^2\right] = \mathrm{MMSE}\left(a\|\mathrm{D}^{-\gamma*}\varphi_{i_0 j}\}\|_\alpha, \sigma\right), \tag{4.48}$$

$$\mathbb{E}\left[(\Xi_{ij} - \widehat{\Xi}_{ij})^2\right] = \mathrm{MMSE}\left(2^{i(\frac{1}{2}-\frac{1}{\alpha}-\gamma)}a\|\phi\|_\alpha, \sigma\right). \tag{4.49}$$

Replacing (4.48) and (4.49) in (4.47), we get

$$\text{MAEE}_{i_0}(\varphi, \psi) = \tag{4.50}$$

$$\lim_{N \to \infty} \frac{1}{2N} \left(\sum_{|j| \leq 2^{i_0} N} \text{MMSE} \left(a \| \text{D}^{-\gamma*} \varphi_{i_0 j} \} \|_\alpha, \sigma \right) \right.$$

$$\left. + \sum_{i \geq i_0} (2^{i+1} N + 1) \text{MMSE} \left(2^{i(\frac{1}{2} - \frac{1}{\alpha} - \gamma)} a \| \phi \|_\alpha, \sigma \right) \right).$$

Now, (4.41) gives us

$$\text{MAEE}_{i_0}(\varphi, \psi) = \lim_{N \to \infty} \frac{1}{2N} \sum_{|j| \leq 2^{i_0} N} \text{MMSE} \left(a \| \text{D}^{-\gamma*} \varphi_{i_0 j} \} \|_\alpha, \sigma \right)$$

$$+ \sum_{i \geq i_0} 2^i \text{MMSE} \left(2^{i(\frac{1}{2} - \frac{1}{\alpha} - \gamma)} a \| \phi \|_\alpha, \sigma \right). \tag{4.51}$$

Also, the existing limit in the right-hand side of (4.51) can be calculated by combining Lemma 1 and Part 1 of Theorem 3. Since $\varphi_{i_0 j}$ has no vanishing moments, $\| \text{D}^{-\gamma*} \varphi_{i_0 j} \|_\alpha$ tends to infinity as j goes to infinity. Thus, we conclude that

$$\lim_{j \to \infty} \text{MMSE} \left(a \| \text{D}^{-\gamma*} \varphi_{i_0 j} \|_\alpha, \sigma \right) = \sigma^2. \tag{4.52}$$

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 (4.52) in (4.51), we finally obtain

$$\text{MAEE}_{i_0}(\varphi, \psi) = 2^{i_0} \sigma^2 + \sum_{i \geq i_0} 2^i \text{MMSE} \left(2^{i(\frac{1}{2} - \frac{1}{\alpha} - \gamma)} a \| \phi \|_\alpha, \sigma \right). \tag{4.53}$$

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

$$\left(\frac{1}{2} - \frac{1}{\alpha} - \gamma \right) (\alpha - \epsilon) + 1 < 0. \tag{4.54}$$

Thus, according to Part 3 of Theorem 3, the summation in (4.53) 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 $\text{MAEE}_{i_0}(\varphi, \psi)$ implies that the series in (4.21) 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 (4.47) as our original definition of performance measure as motivated by Theorem 2, 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 $\text{MMSE}(a, \sigma) \leq \sigma^2$ for any a, a decrease in i_0 results in a decrease of (4.53). 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

$$\text{MAEE}(\psi) = \sum_{i \in \mathbb{Z}} 2^i \text{MMSE}\left(2^{i(\frac{1}{2} - \frac{1}{\alpha} - \gamma)} a \|\phi\|_\alpha, \sigma\right) \tag{4.55}$$

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 (4.53) and (4.55) is that, although the function $\text{MMSE}(\cdot, \cdot)$ is not known analytically, a smaller $\|\phi\|_\alpha$ results in a smaller $\text{MAEE}(\psi)$. This is because we know that $\text{MMSE}(\cdot, \cdot)$ is an increasing function of its first argument (Part 2 of Theorem 3). 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., $\|D^{-\gamma *}\psi\|_\alpha$. 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.

4.7 Optimal Meyer Wavelets for Denoising Self-Similar SαS Processes

Based on the result of Section 4.6 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 [29].

According to [29, 33], for a given function $v : [0, 1] \to \mathbb{R}$ such that

$$v(\rho) + v(1 - \rho) = 1, \tag{4.56}$$

the profile

$$\mathcal{W}\{v\}(\omega) = \begin{cases} \sin\left(\frac{\pi}{2}v(\frac{3}{2\pi}|\omega| - 1)\right), & \frac{2\pi}{3} < |\omega| \le \frac{4\pi}{3} \\ \cos\left(\frac{\pi}{2}v(\frac{3}{4\pi}|\omega| - 1)\right), & \frac{4\pi}{3} < |\omega| \le \frac{8\pi}{3} \\ 0, & \text{otherwise} \end{cases} \tag{4.57}$$

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

$$Q_\alpha^\gamma(v) = \int_\mathbb{R} \left| \frac{1}{2\pi} \int_\mathbb{R} \frac{\mathcal{W}\{v\}(\omega)}{(\mathrm{j}\omega)^\gamma} e^{\mathrm{j}\omega x} \mathrm{d}\omega \right|^\alpha \mathrm{d}x \tag{4.58}$$

$$= \frac{1}{\pi^\alpha} \int_\mathbb{R} \left| \int_{\frac{2\pi}{3}}^{\frac{8\pi}{3}} \frac{\mathcal{W}\{v\}(\omega)}{\omega^\gamma} \cos\left(\omega x - \frac{\pi}{2}\gamma\right) \mathrm{d}\omega \right|^\alpha \mathrm{d}x$$

is a predictor of its denoising performance. Next, we apply a projected-gradient-descent algorithm with adaptive step size to find the function v that minimizes $Q_\alpha^\gamma(v)$. An adaptive step size is specially important for $\alpha \le 1$ for which the functional does not have a Lipschitz gradient. The pseudo-code of our optimization method is given in Algorithm 3. 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, $\mathcal{P}\{\tilde{v}\}$ is the projector that maps $\tilde{v}$ to the nearest function that satisfies (4.56).

Algorithm 3: Optimal Meyer Wavelet

1: **input:** α, γ
2: **initialize:** $v \in L_2([0, 1])$
3: **initialize:** $\eta > 0$, $\kappa^+ \geq 1$ and $\kappa^- \leq 1$
4: $Q \leftarrow Q_\alpha^\gamma(v)$
5: **repeat**
6:     $v_{old} \leftarrow v$ and $Q_{old} \leftarrow Q$
7:     $\tilde{v} \leftarrow v - \eta \nabla Q_\alpha^\gamma\{v\}$
8:     $v \leftarrow \mathcal{P}\{\tilde{v}\}$
9:     $Q \leftarrow Q_\alpha^\gamma(v)$
10:    **if** $Q \leq Q_{old}$ **then**
11:        $\eta \leftarrow \kappa^+ \cdot \eta$
12:    **else**
13:        $v \leftarrow v_{old}$ and $Q \leftarrow Q_{old}$
14:        $\eta \leftarrow \kappa^- \cdot \eta$
15:    **end if**
16: **until** v converges
17: **return** v

According to Appendix 4.E, ∇Q_α^γ is calculated as

$$\nabla Q_\alpha^\gamma\{v\}(\rho) = \tag{4.59}$$

$$\frac{1}{\pi^\alpha} \frac{\pi^2 \alpha}{3} \left(\frac{3}{2\pi}\right)^\gamma \frac{\cos\left(\frac{\pi}{2} v(\rho)\right)}{(\rho + 1)^\gamma}$$

$$\times \int_\mathbb{R} \lambda \left(\int_{\frac{2\pi}{3}}^{\frac{8\pi}{3}} \frac{\mathcal{W}\{v\}(\omega)}{\omega^\gamma} \cos\left(\omega x - \frac{\pi}{2}\gamma\right) \mathrm{d}\omega \right) \cos\left(\frac{2\pi}{3}(\rho + 1)x - \frac{\pi}{2}\gamma\right) \mathrm{d}x$$

$$- \frac{2}{\pi^\alpha} \frac{\pi^2 \alpha}{3} \left(\frac{3}{4\pi}\right)^\gamma \frac{\sin\left(\frac{\pi}{2} v(\rho)\right)}{(\rho + 1)^\gamma}$$

$$\times \int_\mathbb{R} \lambda \left(\int_{\frac{2\pi}{3}}^{\frac{8\pi}{3}} \frac{\mathcal{W}\{v\}(\omega)}{\omega^\gamma} \cos\left(\omega x - \frac{\pi}{2}\gamma\right) \mathrm{d}\omega \right) \cos\left(\frac{4\pi}{3}(\rho + 1)x - \frac{\pi}{2}\gamma\right) \mathrm{d}x,$$

where $\lambda(x) = \text{sgn}(x)|x|^{\alpha-1}$. Also, in Appendix 4.E, we prove that

$$\mathcal{P}\{\tilde{v}\}(\rho) = \frac{\tilde{v}(\rho) - \tilde{v}(1-\rho) + 1}{2}. \tag{4.60}$$

Thus, we have all the ingredients to implement Algorithm 3.

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 [86]. The formal statement of this result is given in Proposition 2.

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

$$v(\rho) = \mathbf{1}_{\left[\frac{1}{2},1\right]}(\rho), \tag{4.61}$$

which corresponds to the wavelet with the Fourier profile

$$\mathcal{W}\{v\}(\omega) = \mathbf{1}_{[-2\pi,-\pi]}(\omega) + \mathbf{1}_{[\pi,2\pi]}(\omega), \tag{4.62}$$

minimizes $Q_2^\gamma(v)$.

The proof is provided in Appendix 4.F.

4.8 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 4.7. Additionally, we compare the performance of Meyer wavelets which are compactly supported in the Fourier domain with the Daubechies wavelets [28] that are compactly supported in the time domain.

To implement Algorithm 3, we take the samples of v and $\mathcal{W}\{v\}$ 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 4.1 the outcome of the algorithm for $\gamma = 1$ and $\alpha = 1.2, 1.8$, and 2. We show in Figure 4.2 the plots for the optimal wavelet when $\alpha = 1.2$ and $\gamma = 1, 2, 4$.

Figure 4.1: Optimal $v(\rho)$ and corresponding wavelet profile $\mathcal{W}\{v\}(\omega)$ for $\gamma = 1$ and $\alpha = 1.2, 1.8$, and 2.

An interesting phenomenon that is observed in these plots is that, by letting either α or γ increase, the wavelet approaches the Shannon wavelet ((4.61) and (4.62)). 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 $Q_\alpha^\gamma(v)$, the frequency content tends toward higher frequencies and v will have less weight on $[0, \frac{1}{2}]$. Therefore, the optimal wavelet approaches the Shannon wavelet which is vanishing on this interval (see (4.61)).

Conversely, the optimal wavelet tends to the Held wavelet ($v(\rho) = \rho$) when α decreases [36]. Thus, we can roughly say that, for very sparse cases (small α), it is better to use the
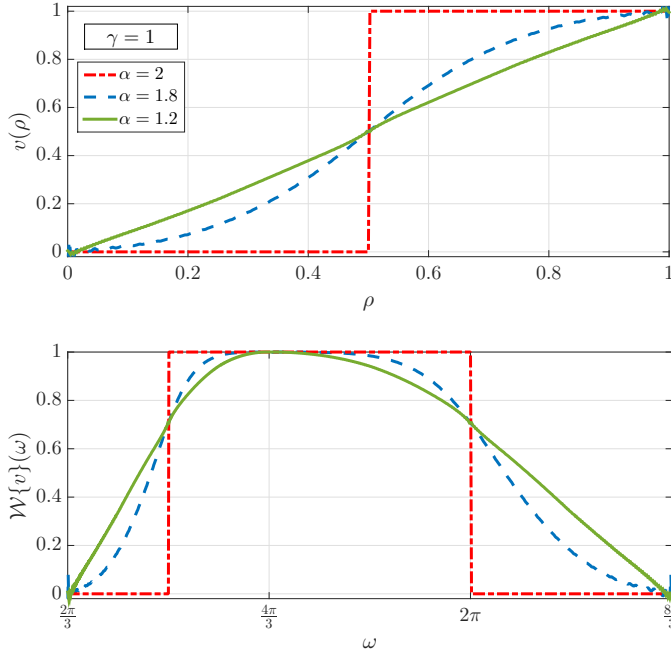
Figure 4.2: Optimal $v(\rho)$ and corresponding wavelet profile $\mathcal{W}\{v\}(\omega)$ for $\alpha = 1.2$ and $\gamma = 1, 2$, and 4.

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 notorious for their minimal support in the time domain, which is valued in many applications.

According to the discussions in Section 4.6, comparing the MAEE obtained by different wavelets for denoising γth-order SαS self-similar processes is equivalent to comparing the α-(pseudo)norm of the γth-order integration of their mother wavelet. We plotted this
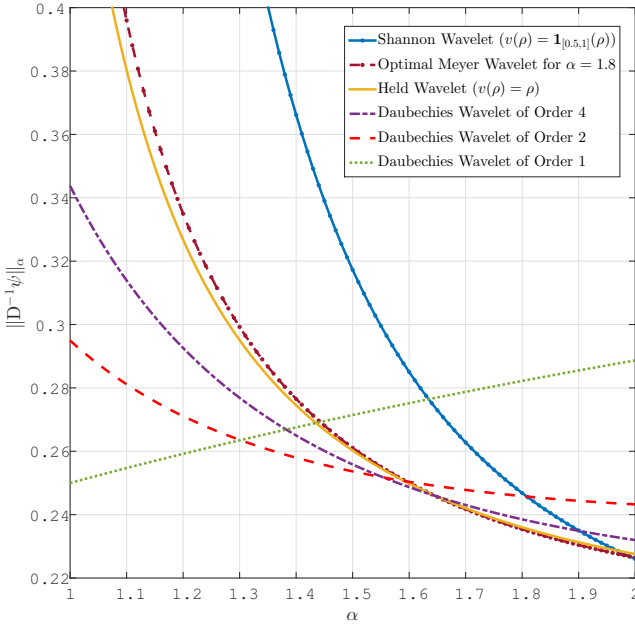
Figure 4.3: The α-norm of the first-order integral of different mother wavelets versus α.

quantity versus α in Figure 4.3 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 [32]) 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.

4.9   Summary

In this chapter, 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 γth-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 2). 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 3).

4.A   Proof of Lemma 1

First, notice that

$$\mathrm{D}^{-\gamma*}\{\theta(-\cdot)\}(x) = \mathrm{D}^{-\gamma*}\{\theta(\cdot)\}(-x). \tag{4.63}$$

Thus, it is enough to prove (4.10) when letting j tend to $+\infty$. According to (4.5) and (4.6), for $x > 0$, we have that

$$\mathrm{D}^{-\gamma*}\{\theta(\cdot - j)\}(x) = \frac{1}{\Gamma(\gamma)}\left((-\cdot)_+^{\gamma-1} * \theta(\cdot)\right)(x - j) \tag{4.64}$$

which implies that

$$\|D^{-\gamma*}\{\theta(\cdot - j)\}\|_\alpha \geq \frac{1}{\Gamma(\gamma)} \left\| \left((-\cdot)_+^{\gamma-1} * \theta(\cdot) \right) \cdot \mathbf{1}_{[-j,0]} \right\|_\alpha, \tag{4.65}$$

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

$$\lim_{x \to -\infty} \frac{\left((-\cdot)_+^{\gamma-1} * \theta(\cdot) \right)(x)}{(-x)_+^{\gamma-1}} = C, \tag{4.66}$$

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

$$\frac{\left((-\cdot)_+^{\gamma-1} * \theta(\cdot) \right)(x)}{(-x)_+^{\gamma-1}} = \frac{\int_\mathbb{R} (t-x)_+^{\gamma-1} \theta(t) dt}{(-x)^{\gamma-1}} \tag{4.67}$$

$$= \int_\mathbb{R} \left(1 - \frac{t}{x} \right)_+^{\gamma-1} \theta(t) dt$$

$$= \int_{|t| \leq |x|^r} \left(1 - \frac{t}{x} \right)_+^{\gamma-1} \theta(t) dt + \int_{|t| > |x|^r} \left(1 - \frac{t}{x} \right)_+^{\gamma-1} \theta(t) dt.$$

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

$$\left(1 - \frac{1}{|x|^r} \right)^{\gamma-1} \int_{|t| \leq |x|^r} \theta(t) dt$$

$$\leq \int_{|t| \leq |x|^r} \left(1 - \frac{t}{x} \right)_+^{\gamma-1} \theta(t) dt \leq \tag{4.68}$$

$$\left(1 + \frac{1}{|x|^r} \right)^{\gamma-1} \int_{|t| \leq |x|^r} \theta(t) dt.$$

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

$$\lim_{x \to -\infty} \int_{|t| \leq |x|^r} \left(1 - \frac{t}{x} \right)_+^{\gamma-1} \theta(t) dt = \int_\mathbb{R} \theta(t) dt. \tag{4.69}$$

As for the second term in the right-hand side of (4.67), we write

$$\int_{-\infty}^{-|x|^r} \left(1 - \frac{t}{x}\right)_+^{\gamma-1} \theta(t) \mathrm{d}t \leq \left(1 - \frac{1}{|x|^r}\right)^{\gamma-1} \int_{-\infty}^{-|x|^r} \theta(t) \mathrm{d}t \tag{4.70}$$

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

$$\int_{|x|^r}^{\infty} \left(1 - \frac{t}{x}\right)_+^{\gamma-1} \theta(t) \mathrm{d}t \leq \int_{|x|^r}^{\infty} \left(1 + t\right)_+^{\gamma-1} \theta(t) \mathrm{d}t. \tag{4.71}$$

Since both integrals on the right-hand side of (4.70) and (4.71) are finite by assumption, by letting x tend to $-\infty$ we obtain that

$$\lim_{x \to -\infty} \int_{|t| > |x|^r} \left(1 - \frac{t}{x}\right)_+^{\gamma-1} \theta(t) \mathrm{d}t = 0. \tag{4.72}$$

Up to now, we have proved that,

$$\lim_{x \to -\infty} \frac{\left((-\cdot)_+^{\gamma-1} * \theta(\cdot)\right)(x)}{(-x)_+^{\gamma-1}} = \int_{\mathbb{R}} \theta(t) \mathrm{d}t, \tag{4.73}$$

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 (4.73) and using the bilinearity of the convolution operator, we can generalize (4.73) to any $\theta \in L_1(\mathbb{R})$. Therefore, since $\int_{\mathbb{R}} \theta(x) \mathrm{d}x \neq 0$, the argument (4.66) 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 (4.68). Thus, (4.69) still holds. In (4.70), the inequality converts to

$$\int_{-\infty}^{-|x|^r} \left(1 - \frac{t}{x}\right)_+^{\gamma-1} \theta(t) \mathrm{d}t$$

$$\leq \sup_{[x, -|x|^r]} |\theta(t)| \times \int_x^{-|x|^r} \left(1 - \frac{t}{x}\right)_+^{\gamma-1} \mathrm{d}t$$

$$\leq \frac{A}{(1 + |x|^r)^n} \times \frac{-x}{\gamma} \left(1 - \frac{1}{|x|^r}\right)^{\gamma}. \tag{4.74}$$

Since $nr > 1$, the bound tends to zero as x tends to $-\infty$. The corresponding inequality for (4.71) with $\gamma < 1$ is

$$\int_{|x|^r}^{\infty} \left(1 - \frac{t}{x}\right)_+^{\gamma-1} \theta(t)\mathrm{d}t \leq \left(1 + \frac{1}{|x|^r}\right)^{\gamma-1} \int_{|x|^r}^{\infty} \theta(t)\mathrm{d}t, \tag{4.75}$$

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

4.B   Proof of Theorem 2

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

$$\mathrm{P}_T\{f\}(t) = f(t) \cdot \mathbf{1}_{[-T,T]} \tag{4.76}$$

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

$$\mathrm{Q}_N\{f\}(t) = \sum_{|j|\leq N} \langle f, \varphi_{0j}\rangle \varphi_{0j}(t) + \sum_{i\geq 0} \sum_{|j|\leq 2^i N} \langle f, \psi_{ij}\rangle \psi_{ij}(t) \tag{4.77}$$

for $f : \mathbb{R} \to \mathbb{R}$ satisfying the conditions of the theorem. Later in the proof, we show that, for a fixed N, (4.77) 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 $\{\varphi_{0j}\}_{|j|\leq N} \cup \{\{\varphi_{ij}\}_{|j|\leq 2^i N}\}_{i=0}^{\infty}$. Hence, we want to prove that

$$\lim_{T\to\infty} \frac{1}{2T} \|\mathrm{P}_T f\|_2^2 = \lim_{N\to\infty} \frac{1}{2N} \|\mathrm{Q}_N f\|_2^2. \tag{4.78}$$

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

$$\begin{aligned}
\|\mathrm{P}_T f\|_2 &= \|\mathrm{P}_T \mathrm{Q}_N f + \mathrm{P}_T(1 - \mathrm{Q}_N)f\|_2 \\
&\leq \|\mathrm{P}_T \mathrm{Q}_N f\|_2 + \|\mathrm{P}_T(1 - \mathrm{Q}_N)f\|_2 \\
&\leq \|\mathrm{Q}_N f\|_2 + \|\mathrm{P}_T(1 - \mathrm{Q}_N)f\|_2.
\end{aligned} \tag{4.79}$$

Similarly, we have that

$$\|\mathrm{Q}_N f\|_2 \leq \|\mathrm{P}_T f\|_2 + \|\mathrm{Q}_N(1 - \mathrm{P}_T)f\|_2. \tag{4.80}$$

Therefore, we can write

$$\sqrt{\frac{T}{N}} \Big(\frac{\|P_T f\|_2}{\sqrt{T}} - \frac{\|P_T(1 - Q_N)f\|_2}{\sqrt{T}} \Big)$$
$$\leq \frac{\|Q_N f\|_2}{\sqrt{N}} \leq \quad\quad (4.81)$$
$$\sqrt{\frac{T}{N}} \Big(\frac{\|P_T f\|_2}{\sqrt{T}} + \frac{\|Q_N(1 - P_T)f\|_2}{\sqrt{T}} \Big).$$

Now, assume that $\delta > 0$ and let $N^+ = (1 + \delta)^2 T$ and $N^- = (1 - \delta)^2 T$. Using the inequalities in (4.81), by letting T tend to infinity we obtain that

$$(1 - \delta) \limsup_{N \to \infty} \frac{\|Q_N f\|_2}{\sqrt{N}} \leq$$
$$\lim_{T \to \infty} \frac{\|P_T f\|_2}{\sqrt{T}} + \limsup_{T \to \infty} \frac{\|Q_{N^-}(1 - P_T)f\|_2}{\sqrt{T}} \quad\quad (4.82)$$

and

$$(1 + \delta) \limsup_{N \to \infty} \frac{\|Q_N f\|_2}{\sqrt{N}} \geq$$
$$\lim_{T \to \infty} \frac{\|P_T f\|_2}{\sqrt{T}} - \limsup_{T \to \infty} \frac{\|P_T(1 - Q_{N^+})f\|_2}{\sqrt{T}}. \quad\quad (4.83)$$

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

$$\limsup_{T \to \infty} \frac{\|Q_N(1 - P_T)f\|_2^2}{T} = 0 \text{ if } N \leq (1 - \delta)T \quad\quad (4.84)$$

and

$$\limsup_{T \to \infty} \frac{\|P_T(1 - Q_N)f\|_2^2}{T} = 0 \text{ if } N \geq (1 + \delta)T. \quad\quad (4.85)$$

Notice that (4.84), at first hand, yields that (4.77) is a converging series in $L_2(\mathbb{R})$ and

$Q_N f$ is well-defiend. The reason is that

$$\sum_{|j|\leq N} \langle f, \varphi_{0j}\rangle^2 + \sum_{i\geq 0} \sum_{|j|\leq 2^i N} \langle f, \psi_{ij}\rangle^2 =$$

$$\sum_{|j|\leq N} (\langle P_T f, \varphi_{0j}\rangle + \langle (1-P_T)f, \varphi_{0j}\rangle)^2$$

$$+ \sum_{i\geq 0} \sum_{|j|\leq 2^i N} (\langle P_T f, \psi_{ij}\rangle + \langle (1-P_T)f, \psi_{ij}\rangle)^2. \tag{4.86}$$

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

$$\sum_{|j|\leq N} \langle f, \varphi_{0j}\rangle^2 + \sum_{i\geq 0} \sum_{|j|\leq 2^i N} \langle f, \psi_{ij}\rangle^2 \leq$$

$$2\Big(\sum_{|j|\leq N} \langle P_T f, \varphi_{0j}\rangle^2 + \sum_{i\geq 0} \sum_{|j|\leq 2^i N} \langle P_T f, \psi_{ij}\rangle^2 \Big)$$

$$+ 2\Big(\sum_{|j|\leq N} \langle (1-P_T)f, \varphi_{0j}\rangle^2 + \sum_{i\geq 0} \sum_{|j|\leq 2^i N} \langle (1-P_T)f, \psi_{ij}\rangle^2 \Big)$$

$$= 2\|Q_N P_T f\|_2^2 + 2\|Q_N (1-P_T)f\|_2^2. \tag{4.87}$$

Since $|P_T f(x)| \leq P_T g(x)$ 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 (4.86) is finite. Consequently, the left-hand side of (4.86) is finite, too, and thus $Q_N f$ belongs to $L_2(\mathbb{R})$.

To prove (4.84), we write

$$\|Q_N(1-P_T)f\|_2^2 = \sum_{|j|\leq N} \Big(\int_{|x|>T} f(x)\varphi(x-j)\mathrm{d}x \Big)^2$$

$$+ \sum_{i\geq 0} \sum_{|j|\leq 2^i N} \Big(\int_{|x|>T} f(x)2^{\frac{i}{2}}\psi(2^i x - j)\mathrm{d}x \Big)^2$$

$$= \sum_{|j|\leq N} \Big(\int_{|x+j|>T} f(x+j)\varphi(x)\mathrm{d}x \Big)^2$$

$$+ \sum_{i\geq 0} \sum_{|j|\leq 2^i N} 2^{-i} \Big(\int_{|\frac{x+j}{2^i}|>T} f\Big(\frac{x+j}{2^i}\Big)\psi(x)\mathrm{d}x \Big)^2. \tag{4.88}$$

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 (4.88), we deduce that

$$
\begin{aligned}
\|Q_N(1 - P_T)f\|_2^2 &\le 4T \Big(\int_{x > \delta T} g(x + T)\eta(x)\mathrm{d}x \Big)^2 \\
&\quad + \sum_{i \ge 0} 2^{-i} \cdot 4 \cdot 2^i T \Big(\int_{x > 2^i \delta T} g(x + T)\eta(x)\mathrm{d}x \Big)^2 \\
&\le 4T \Big[\Big(\int_{x > \delta T} g\big((1 + \tfrac{1}{\delta})x\big)\eta(x)\mathrm{d}x \Big)^2 \\
&\quad + \sum_{i \ge 0} \Big(\int_{x > 2^i \delta T} g\big((1 + \tfrac{1}{\delta})x\big)\eta(x)\mathrm{d}x \Big)^2 \Big].
\end{aligned}
\tag{4.89}
$$

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

$$
\begin{aligned}
\|Q_N(1 - P_T)f\|_2^2 &\le 4T \Big[\Big(\int_{\delta T}^{\infty} \frac{c}{x^{1+\epsilon}}\mathrm{d}x \Big)^2 + \sum_{i \ge 0} \Big(\int_{2^i \delta T}^{\infty} \frac{c}{x^{1+\epsilon}}\mathrm{d}x \Big)^2 \Big] \\
&= 4T \Big(\frac{c^2}{\delta^2 T^{2\epsilon}} + \sum_{i \ge 0} \frac{c^2}{2^{2i}\delta^2 T^{2\epsilon}} \Big) = c'T^{1-2\epsilon}
\end{aligned}
\tag{4.90}
$$

with c' independent of T, which completes the proof of (4.84).

To prove (4.85), we write

$$
\begin{aligned}
\|P_T(1 - Q_N)f\|_2^2 &= \int_{-T}^{T} |(1 - Q_N)\{f\}(x)|^2 \mathrm{d}x \\
&\le 2T \Big[\sup_{|x| \le T} |(1 - Q_N)\{f\}(x)| \Big]^2.
\end{aligned}
\tag{4.91}
$$

We have that

$$
\begin{aligned}
(1 - Q_N)\{f\}(x) &= \sum_{|j| > N} \Big(\int_{\mathbb{R}} f(u)\varphi(u - j)\mathrm{d}u \Big)\varphi(x - j) \tag{4.92} \\
&\quad + \sum_{i \ge 0} \sum_{|j| > 2^i N} \Big(\int_{\mathbb{R}} f(u)2^{\frac{i}{2}}\psi(2^i u - j)\mathrm{d}u \Big)2^{\frac{i}{2}}\psi(2^i x - j).
\end{aligned}
$$

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

$$|(1 - Q_N)\{f\}(x)| \leq \sum_{|j|>N} \left(\int_{\mathbb{R}} g(u + j)\eta(u)\mathrm{d}u \right)\eta(x - j) \tag{4.93}$$

$$+ \sum_{i \geq 0} \sum_{|j|>2^i N} \left(\int_{\mathbb{R}} g(\frac{u + j}{2^i})\eta(u)\mathrm{d}u \right)\eta(2^i x - j).$$

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

$$|(1 - Q_N)\{f\}(x)| \leq \sum_{|j|>N} \left(\int_{\mathbb{R}} g(u + j)\eta(u)\mathrm{d}u \right)\eta(\delta' j) \tag{4.94}$$

$$+ \sum_{i \geq 0} \sum_{|j|>2^i N} \left(\int_{\mathbb{R}} g(\frac{u + j}{2^i})\eta(u)\mathrm{d}u \right)\eta(2^i \delta' j),$$

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

$$(x + y)^n \leq C(|x|^n + |y|^n). \tag{4.95}$$

The same C satisfies

$$g(x + y) \leq C(g(x) + g(y)). \tag{4.96}$$

Thus, we write

$$|(1 - Q_N)\{f\}(x)| \leq C \sum_{|j|>N} \left(\int_{\mathbb{R}} (g(u) + g(j))\eta(u)\mathrm{d}u \right)\eta(\delta' j) \tag{4.97}$$

$$+ C \sum_{i \geq 0} \sum_{|j|>2^i N} \left(\int_{\mathbb{R}} (g(\frac{u}{2^i}) + g(\frac{j}{2^i}))\eta(u)\mathrm{d}u \right)\eta(2^i \delta' j).$$

By expanding the summations, we obtain

$$
\begin{aligned}
|(1 - Q_N)\{f\}(x)| \;\leq\; & C\Big(\int_{\mathbb{R}} g(u)\eta(u)\mathrm{d}u\Big) \sum_{|j|>N} \eta(\delta'j) \\
& + C\Big(\int_{\mathbb{R}} \eta(u)\mathrm{d}u\Big) \sum_{|j|>N} g(j)\eta(\delta'j) \\
& + C\sum_{i\geq 0}\Big[\Big(\int_{\mathbb{R}} g(\tfrac{u}{2^i})\eta(u)\mathrm{d}u\Big) \sum_{|j|>N} \eta(2^i\delta'j) \\
& \qquad\qquad + \Big(\int_{\mathbb{R}} \eta(u)\mathrm{d}u\Big) \sum_{|j|>N} g(\tfrac{j}{2^i})\eta(2^i\delta'j)\Big]. \quad (4.98)
\end{aligned}
$$

However, we have that

$$
\int_{\mathbb{R}} g(\tfrac{u}{2^i})\eta(u)\mathrm{d}u \leq BC\Big(\int_{\mathbb{R}} \eta(u)\mathrm{d}u + \frac{1}{2^{in}}\int_{\mathbb{R}} u^n\eta(u)\mathrm{d}u\Big), \quad (4.99)
$$

where B and C are as in (4.25) and (4.95), respectively. Likewise, we have that

$$
\begin{aligned}
\int_{|u|>N} g(\tfrac{u}{2^i})\eta(2^i\delta'u)\mathrm{d}u \;\leq\; & BC\Big(\int_{|u|>N} \eta(2^i\delta'u)\mathrm{d}u + \frac{1}{2^{in}}\int_{|u|>N} u^n\eta(2^i\delta'u)\mathrm{d}u\Big) \\
\leq\; & \frac{BC}{2^i\delta'}\Big(\int_{|u|>2^i\delta'N} \eta(u)\mathrm{d}u + \frac{1}{(2^{2i}\delta')^n}\int_{|u|>2^i\delta'N} u^n\eta(u)\mathrm{d}u\Big) \\
\leq\; & \frac{BC}{2^i\delta'}\Big(\frac{\dot{c}}{(2^i\delta'N)^{n+\epsilon}} + \frac{1}{(2^{2i}\delta')^n}\frac{\ddot{c}}{2^i\delta'N^\epsilon}\Big). \quad (4.100)
\end{aligned}
$$

Therefore, using the inequalities in (4.100) and bounding the summations in (4.98) by integrals, for a large enough N, we get

$$
\begin{aligned}
|(1 - Q_N)\{f\}(x)| \;\leq\; & \frac{c_1}{N^{n+\epsilon}} + \frac{c_2}{N^\epsilon} + \sum_{i=0}^{\infty}\Big(\frac{c_3}{2^{(n+2)i}}\cdot\frac{1}{N^{n+\epsilon}} + \frac{c_4}{2^{(2n+1)i}N^\epsilon}\Big) \\
\leq\; & \frac{c''}{T^\epsilon}, \quad (4.101)
\end{aligned}
$$

in which c_1 to c_4 and c'' are constants independent of N and T. This completes the proof of (4.85) and hence the proof of the theorem.

4.C   Proof of Proposition 1

Repeating the proof of (4.85), we show that

$$\limsup_{N\to\infty} \|\mathrm{P}_T(1-\mathrm{Q}_N)f\|_2^2 = 0 \tag{4.102}$$

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.

4.D   Proof of Theorem 3

1) Assume that $a_1 \geq a_2 \geq 0$. We are going to show that $M(a_1, b) \geq M(a_2, b)$. 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

$$\begin{aligned} p_{X_1}(x) &= \int p_{X_2}(x - t) p_{X_3}(t) \mathrm{d}t \\ &= \mathbb{E}_{X_3}[p_{X_2}(x - X_3)]. \end{aligned} \tag{4.103}$$

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 [85].

2) This is a direct implication of [80, Theorem 11] since the distribution of an SαS random variable is absolutely continuous.

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

$$M(a, \sigma) = \frac{2a^2\sigma^2}{2a^2 + \sigma^2}, \tag{4.104}$$

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

$$T(y) = \begin{cases} 0, & |y| < \Delta \\ y, & |y| \geq \Delta \end{cases} \tag{4.105}$$

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

$$f_a(x) = \frac{1}{a} f\left(\frac{x}{a}\right), \tag{4.106}$$

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

$$F_a(s, t) = \int_s^t f_a(x)\mathrm{d}x. \tag{4.107}$$

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

$$G(s, t) = \int_s^t g(x)\mathrm{d}x. \tag{4.108}$$

Then, we have that

$$\begin{aligned}
\mathrm{MSE}(T) &= \mathbb{E}_{X,Z}\left[(X - T(X + Z))^2\right] \\
&= \mathbb{P}\left(|X + Z| < \Delta\right) \mathbb{E}_{X,Z}\left[X^2 \mid |X + Z| < \Delta\right] \\
&\quad + \mathbb{P}\left(|X + Z| \geq \Delta\right) \mathbb{E}_{X,Z}\left[Z^2 \mid |X + Z| \geq \Delta\right].
\end{aligned} \tag{4.109}$$

According to Bayes' rule, we can write

$$\begin{aligned}
f_a\left(x \big| |X + Z| < \Delta\right) &= \frac{\mathbb{P}\left(|x + Z| < \Delta\right) f_a(x)}{\mathbb{P}\left(|X + Z| < \Delta\right)} \\
&= \frac{G\left(-\Delta - x, \Delta - x\right) f_a(x)}{\mathbb{P}\left(|X + Z| < \Delta\right)}
\end{aligned} \tag{4.110}$$

and, similarly,

$$g\left(z \big| |X + Z| \geq \Delta\right) = \frac{\left(1 - F_a\left(-\Delta - z, \Delta - z\right)\right) g(z)}{\mathbb{P}\left(|X + Z| \geq \Delta\right)}. \tag{4.111}$$

Incorporating (4.110) and (4.111) in (4.109), we obtain

$$\begin{aligned}
\mathrm{MSE}(T) &= \int_{\mathbb{R}} x^2 f_a(x) G(-\Delta - x, \Delta - x)\mathrm{d}x \\
&\quad + \int_{\mathbb{R}} z^2 g(z) \left(1 - F_a\left(-\Delta - z, \Delta - z\right)\right)\mathrm{d}z.
\end{aligned} \tag{4.112}$$

We investigate the behavior of the two terms on the right-hand side of (4.112) separately. For the first term, we know that, for $\alpha < 2$, we have that

$$\lim_{x \to \infty} \frac{f(x)}{x^{-(1+\alpha)}} = C, \tag{4.113}$$

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

$$x^{\frac{1}{2}+\alpha-\delta} f_a(x) \in L_2(\mathbb{R}) \tag{4.114}$$

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

$$\int x^2 f_a(x) G(-\Delta - x, \Delta - x) \mathrm{d}x \leq$$
$$\left\| x^{\frac{1}{2}+\alpha-\delta} f_a(x) \right\|_2 \cdot \left\| x^{\frac{3}{2}-\alpha+\delta} G(-\Delta - x, \Delta - x) \right\|_2. \tag{4.115}$$

Notice that

$$G(-\Delta - x, \Delta - x) \leq \begin{cases} 1, & |x| \leq \Delta \\ \frac{\sigma^2}{\Delta} g(x), & |x| > \Delta \end{cases}. \tag{4.116}$$

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

$$\left\| x^{\frac{3}{2}-\alpha+\delta} G(-\Delta - x, \Delta - x) \right\|_2^2 \leq \frac{\Delta^{4-2\alpha+2\delta}}{2-\alpha+\delta} + \frac{\sigma}{2\Delta\sqrt{\pi}}. \tag{4.117}$$

Additionally, we can write

$$\left\| x^{\frac{1}{2}+\alpha-\delta} f_a(x) \right\|_2 = \left(\int x^{1+2\alpha-2\delta} f_a^2(x) \mathrm{d}x \right)^{\frac{1}{2}}$$
$$= \left(\int x^{1+2\alpha-2\delta} \frac{1}{a^2} f^2 \left(\frac{x}{a} \right) \mathrm{d}x \right)^{\frac{1}{2}}$$
$$= a^{\alpha-\delta} \left(\int x^{1+2\alpha-2\delta} f^2 (x) \, \mathrm{d}x \right)^{\frac{1}{2}}$$
$$= a^{\alpha-\delta} \left\| x^{\frac{1}{2}+\alpha-\delta} f(x) \right\|_2. \tag{4.118}$$

Now, we investigate the second term of the right-hand side of (4.112). We specifically write

$$\int_{\mathbb{R}} z^2 g(z) \left(1 - F_a \left(-\Delta - z, \Delta - z\right)\right) \mathrm{d}z =$$

$$\int_{\mathbb{R}} z^2 g(z) \int_{\mathbb{R}} \left(1 - \mathbf{1}_{[-\Delta, \Delta]}(x + z)\right) f_a(x) \mathrm{d}x \mathrm{d}z$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} \left(1 - \mathbf{1}_{[-\Delta, \Delta]}(x + z)\right) z^2 g(z) f_a(x) \mathrm{d}z \mathrm{d}x$$

$$= 2 \int_{\Delta}^{\infty} \int_{\mathbb{R}} z^2 g(z) f_a(t - z) \mathrm{d}z \mathrm{d}t, \tag{4.119}$$

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

$$\int_{\mathbb{R}} z^2 g(z) f_a(t - z) \mathrm{d}z =$$

$$\int_{|z| < \frac{t}{2}} z^2 g(z) f_a(t - z) \mathrm{d}z + \int_{|z| \geq \frac{t}{2}} z^2 g(z) f_a(t - z) \mathrm{d}z$$

$$\leq f_a(\frac{t}{2}) \int_{\mathbb{R}} z^2 g(z) \mathrm{d}z + \frac{t^2}{4} g(\frac{t}{2}) \int_{\mathbb{R}} f_a(z) \mathrm{d}z$$

$$= \sigma^2 f_a(\frac{t}{2}) + \frac{t^2}{4} g(\frac{t}{2}) \tag{4.120}$$

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

$$2 \int_{\Delta}^{\infty} \int_{\mathbb{R}} z^2 g(z) f_a(t - z) \mathrm{d}z \mathrm{d}t \leq 2\sigma^2 \int_{\Delta}^{\infty} f_a(\frac{t}{2}) \mathrm{d}t + 2 \int_{\Delta}^{\infty} \frac{t^2}{4} g(\frac{t}{2}) \mathrm{d}t \tag{4.121}$$

for $\Delta \geq 2\sigma\sqrt{2}$. According to (4.113), there exists $C' \in \mathbb{R}$ for which

$$f(t) \leq \frac{C'}{t^{1+\alpha}}, \tag{4.122}$$

and thus

$$f_a(t) \leq a^{\alpha} \frac{C'}{t^{1+\alpha}}. \tag{4.123}$$

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

$$t^2 \mathrm{e}^{-\frac{t^2}{2}} \leq A \mathrm{e}^{-t}. \tag{4.124}$$

Hence, relying on (4.121), we get

$$2 \int_\Delta^\infty \int_\mathbb{R} z^2 g(z) f_a(t-z) \mathrm{d}z \mathrm{d}t \leq a^\alpha 2^{2+\alpha} C' \sigma^2 \frac{1+\alpha}{\Delta^\alpha} + 2A\sigma^2 \sqrt{\frac{2}{\pi}} \mathrm{e}^{-\frac{\Delta}{2\sigma}}. \tag{4.125}$$

Now, incorporating (4.115) and (4.125) in (4.112), we obtain

$$\begin{aligned}
\mathrm{MSE}(T) \leq {} & a^{\alpha-\delta} \left\| x^{\frac{1}{2}+\alpha-\delta} f(x) \right\|_2 \left(\frac{\Delta^{4-2\alpha+2\delta}}{4-2\alpha+2\delta} + 2\sigma\sqrt{\pi} \right) \\
& + a^\alpha 2^{2+\alpha} C' \sigma^2 \frac{1+\alpha}{\Delta^\alpha} + 2A\sigma^2 \sqrt{\frac{2}{\pi}} \mathrm{e}^{-\frac{\Delta}{2\sigma}}.
\end{aligned} \tag{4.126}$$

Finally, setting

$$\Delta = 2\alpha\sigma \left| \log a \right| \tag{4.127}$$

completes the proof.

4.E   Calculation of the Gradient of Q_α^γ

According to the definition of the gradient, we have that

$$\begin{aligned}
\mathrm{D}_u Q_\alpha^\gamma(v) & = \frac{\partial}{\partial\epsilon} Q_\alpha^\gamma(v+\epsilon u) \Big|_{\epsilon=0} \\
& = \int_0^\infty u(\rho) \, \nabla Q_\alpha^\gamma\{v\}(\rho) \mathrm{d}\rho,
\end{aligned} \tag{4.128}$$

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

$$\mathrm{D}_u Q_\alpha^\gamma(v) =$$

$$\frac{1}{(2\pi)^\alpha} \int_\mathbb{R} \frac{\partial}{\partial\epsilon}\bigg|\bigg|\int_\mathbb{R} \frac{\mathcal{W}\{v+\epsilon u\}(\omega)}{(\mathrm{j}\omega)^\gamma}\mathrm{e}^{\mathrm{j}\omega x}\mathrm{d}\omega\bigg|^\alpha\bigg|_{\epsilon=0}\mathrm{d}x$$

$$= \frac{\alpha}{(2\pi)^\alpha}\int_\mathbb{R}\lambda\left(\int_\mathbb{R}\frac{\mathcal{W}\{v\}(\omega)}{(\mathrm{j}\omega)^\gamma}\mathrm{e}^{\mathrm{j}\omega x}\,\mathrm{d}\omega\right)\left(\int_\mathbb{R}\frac{\mathcal{W}'\{v\}(r)\bar{u}(\rho)}{(\mathrm{j}\rho)^\gamma}\mathrm{e}^{\mathrm{j}\rho x}\,\mathrm{d}\rho\right)\mathrm{d}x$$

$$= \frac{\alpha}{(2\pi)^\alpha}\int_\mathbb{R}\bar{u}(\rho)\frac{\mathcal{W}'\{v\}(\rho)}{(\mathrm{j}\rho)^\gamma}\left(\int_\mathbb{R}\lambda\left(\int_\mathbb{R}\frac{\mathcal{W}\{v\}(\omega)}{(\mathrm{j}\omega)^\gamma}\mathrm{e}^{\mathrm{j}\omega x}\,\mathrm{d}\omega\right)\mathrm{e}^{\mathrm{j}\rho x}\,\mathrm{d}x\right)\mathrm{d}\rho \quad (4.129)$$

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

$$\mathcal{W}'\{v\}(\rho) = \frac{\pi}{2}\times\begin{cases}\cos(\frac{\pi}{2}v(\frac{3}{2\pi}|\rho|-1)), & \frac{2\pi}{3}<|\rho|\leq\frac{4\pi}{3}\\ -\sin(\frac{\pi}{2}v(\frac{3}{4\pi}|\rho|-1)), & \frac{4\pi}{3}<|\rho|\leq\frac{8\pi}{3}\\ 0, & \text{otherwise}\end{cases} \quad (4.130)$$

and

$$\bar{u}(\rho) = \begin{cases}u(\frac{3}{2\pi}\rho-1), & \frac{2\pi}{3}<|\rho|\leq\frac{4\pi}{3}\\ u(\frac{3}{4\pi}\rho-1), & \frac{4\pi}{3}<|\rho|\leq\frac{8\pi}{3}\end{cases}. \quad (4.131)$$

By breaking the outer integral of (4.129) 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

$$\mathrm{D}_u Q_\alpha^\gamma(v) = \frac{(2\pi)^{1-\alpha}\alpha}{3}\int_0^1 u(\rho)\frac{\mathcal{W}'\{v\}(\frac{2\pi}{3}(\rho+1))}{(\mathrm{j}\frac{2\pi}{3}(\rho+1))^\gamma}$$

$$\times\left(\int_\mathbb{R}\lambda\left(\int_\mathbb{R}\frac{\mathcal{W}\{v\}(\omega)}{(\mathrm{j}\omega)^\gamma}\mathrm{e}^{\mathrm{j}\omega x}\,\mathrm{d}\omega\right)\mathrm{e}^{\mathrm{j}\frac{2\pi}{3}(\rho+1)x}\,\mathrm{d}x\right)\mathrm{d}\rho$$

$$+ \frac{2(2\pi)^{1-\alpha}\alpha}{3}\int_0^1 u(\rho)\frac{\mathcal{W}'\{v\}(\frac{4\pi}{3}(\rho+1))}{(\mathrm{j}\frac{4\pi}{3}(\rho+1))^\gamma}$$

$$\times\left(\int_\mathbb{R}\lambda\left(\int_\mathbb{R}\frac{\mathcal{W}\{v\}(\omega)}{(\mathrm{j}\omega)^\gamma}\mathrm{e}^{\mathrm{j}\omega x}\,\mathrm{d}\omega\right)\mathrm{e}^{\mathrm{j}\frac{4\pi}{3}(\rho+1)x}\,\mathrm{d}x\right)\mathrm{d}\rho.$$

$$(4.132)$$

Now, incorporating (4.57), (4.130) and some algebra, according to (4.128), we obtain (4.59).

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

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

$$J(v) = \|v(\rho) - \tilde{v}(\rho)\|_2^2 \tag{4.133}$$

and

$$\bar{J}(v) = \frac{1}{2}\left(\|v(\rho) - \tilde{v}(\rho)\|_2^2 + \|(1 - v(1 - \rho)) - \tilde{v}(\rho)\|_2^2\right) \tag{4.134}$$

We are interested in

$$\mathcal{P}\{\tilde{v}\} = \operatorname{argmin} J(v) \tag{4.135}$$

subject to the constraint

$$v(x) + v(1 - x) = 1. \tag{4.136}$$

Notice that $J(v) = \bar{J}(v)$ for any function v that satisfies (4.136). 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 (4.136). 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

$$\nabla \bar{J}\{v\}(\rho) = (v(\rho) - \tilde{v}(\rho)) - (1 - v(\rho) - \tilde{v}(1 - \rho)). \tag{4.137}$$

By solving $\nabla \bar{J}\{v\} = 0$, we obtain (4.60).

4.F   Proof of Proposition 2

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

$$Q_2^\gamma(v) = \int_{\mathbb{R}} \left|\frac{\mathcal{W}\{v\}(\omega)}{(\mathrm{j}\omega)^\gamma}\right|^2 \mathrm{d}\omega. \tag{4.138}$$

Based on (4.57), we get

$$Q_2^\gamma(v) = 2\int_{\frac{2\pi}{3}}^{\frac{4\pi}{3}} \frac{\sin^2\left(\frac{\pi}{2}v(\frac{3}{2\pi}\omega - 1)\right)}{\omega^{2\gamma}} + 2\int_{\frac{4\pi}{3}}^{\frac{8\pi}{3}} \frac{\cos^2\left(\frac{\pi}{2}v(\frac{3}{4\pi}\omega - 1)\right)}{\omega^{2\gamma}}\mathrm{d}\omega. \tag{4.139}$$

Performing ordinary algebraic manipulations, (4.139) leads to

$$Q_2^\gamma(v) = 2\Big(\big(\frac{2\pi}{3}\big)^{1-2\gamma} - \big(\frac{4\pi}{3}\big)^{1-2\gamma}\Big) \int_0^1 \frac{\sin^2\big(\frac{\pi}{2}v(\rho)\big)}{(\rho+1)^{2\gamma}} \mathrm{d}\rho$$
$$+ \frac{2}{2\gamma-1}\Big(\big(\frac{4\pi}{3}\big)^{1-2\gamma} - \big(\frac{8\pi}{3}\big)^{1-2\gamma}\Big). \tag{4.140}$$

Now, using (4.56), we get

$$Q_2^\gamma(v) = 2\Big(\big(\frac{2\pi}{3}\big)^{1-2\gamma} - \big(\frac{4\pi}{3}\big)^{1-2\gamma}\Big)$$
$$\times \int_0^{\frac{1}{2}} \sin^2\big(\frac{\pi}{2}v(\rho)\big)\Big(\frac{1}{(\rho+1)^{2\gamma}} - \frac{1}{(2-\rho)^{2\gamma}}\Big)\mathrm{d}\rho$$
$$+ \frac{2}{2\gamma-1}\Big(\big(\frac{\pi}{2}\big)^{1-2\gamma} - \pi^{1-2\gamma}\Big). \tag{4.141}$$

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

Chapter 5

Localized Isotropic Wavelets for Image Analysis and Reconstruction

In this chapter, we use the previous theoretical results to gain a practical advantage. We propose two classes of optimality criteria based on which we design new isotropic wavelets for image processing applications. The first class specifies the spatial localization of the wavelet profile, and the second that of the resulting wavelet coefficients. Also, we present an infinite-dimensional optimization scheme that helps us find the optimal profile for a given criterion over the space of tight frames. From these metrics and the proposed algorithm, we construct tight wavelet frames that are optimally localized and provide their analytical expression. In particular, one of the considered criterion helps us finding back the popular Simoncelli wavelet profile. Finally, the investigation of local orientation estimation, image reconstruction from detected contours in the wavelet domain, and denoising, indicate that optimizing wavelet localization improves the performance of steerable wavelets, since our new wavelets outperform the traditional ones.

5.1   Introduction

Isotropic wavelets described in Section 2.3.4 are purely radial functions that allow for an orientation-free decomposition of images, while retaining all other popular features of classical wavelet bases such as multiresolution analysis. The steerable pyramid [87, 88, 89, 90, 91] is a well-known construction that relies on such wavelets. In this setting, a purely angular element is included in order to rotate derivatives of the wavelets and impose a preferred directionality. Some well-known algorithms for denoising ([92], including the widely-used Bayesian least-squares Gaussian-scale-mixture (BLS-GSM) algorithm [93]), texture analysis (or synthesis) [94, 95], and regularization with sparsity constraints for inverse problems [96, 97] rely on the steerable pyramid, although methods that do not exploit steerability are also available for these tasks. Steerability is a crucial aspect in many other image-processing applications such as finding the dominant orientation at each image location, detecting contours [98], or identifying features in a rotation-invariant fashion [99]. More recently, algorithms for image reconstruction from the small subset of wavelet coefficients called the "primal sketch" have been proposed relying on the steerable pyramid [100, 101]. In this work, we study the design of wavelet profiles for use in applications relying on steerable tight frames. The specification of steerable wavelet frames includes two components: a radial profile and a directional components can be optimized separately. The angular component is represented using circular harmonics [90]. Here, we concentrate on the radial profile, which determines the localization.

In order to generate an isotropic wavelet transform, the underlying basis functions must satisfy several properties, the main ones being isotropy and perfect reconstruction of the image. Another desirable feature is that the basis functions form a tight frame. In this way, the wavelet transform is self-reversible, enabling simpler and faster algorithms. As discussed in Section 2.3.4, the isotropy and perfect-reconstruction conditions are ensured by choosing a radially bandlimited mother wavelet that satisfies some partition of unity in the frequency domain [35, 36]. Many such bases have been proposed, which include the Meyer [33], Papadakis [102], and Simoncelli [103] wavelets. Inspired by the biological visual system, the Simoncelli wavelet is the one implemented in the original version of the steerable pyramid and the BLS-GSM denoising algorithm. Due to its good performance in a wide range of practical applications, it remains a commonly used profile.

It is always interesting to have a simple measure which quantifies the performance of a wavelet in practical applications. The studies in previous chapters give us a clue about what the measure of performance should be. In Chapters 3 and 4, we considered two different classes of stochastic processes and in both cases we deduced that the best wavelets

are the ones that are most localized ones after applying the inverse of the whitening operators. However, the meaning of localization were different for different processes. As we saw, the localization meant having shortest support for AR(1) processes, and it meant having smallest L_α-norm for self-similar processes (notice that finite L_α-norm of a function controls its decay rate). Thus, we can conclude that, in general, localization is the best candidate for measuring the performance of a wavelet. However, we have to pick a general meaning for localization for a general purpose wavelet, i.e. when it is not design for a specific stochastic process. On the other hand, there are also intuitive justifications about appropriateness of localization measure. On the practical side, steerable wavelets are bandlimited with infinitely many vanishing moments, which tends to induce oscillations that can be visually displeasing. It is observed that more-localized wavelets result in fewer oscillations and are less subject to truncation artifacts.

In this chapter, we introduce a method to design radial profiles for steerable tight frames. Since the frequency response of steerable wavelets is polar-separable, we can concentrate on the task of optimizing the radial frequency profile. We focus on moment-based measures of localization and propose two different classes of criteria depending on whether we consider the localization in the spatial domain or in the wavelet domain. Two criteria can be derived within each class, depending on whether one wants to consider localization over the whole space or in each radial direction. Ultimately, the choice of a particular criterion among these is guided by the application. Similar to Section 4.7, we describe an algorithm using the calculus of variations to optimize the wavelet corresponding to each measure through gradient descent. We then obtain analytical expressions for the optimally localized profiles. We then show the benefit of our optimized design in three practical applications, namely, local orientation estimation, image reconstruction from edges, and denoising. These experiments highlight different use-cases in which each of the proposed type of localization (spatial versus wavelet domain) is desirable. In particular, we provide additional results on the image-reconstruction problem compared to [104], as well as further study of the performance of our wavelets for local orientation estimation and for the BLS-GSM denoising algorithm.

The organization of the chapter is as follows: In the next section, we specify our measures of localization and propose a step-by-step algorithm to design optimally localized profiles corresponding to each measure. We then provide the closed-form expression of the resulting optimal wavelets. Finally, we focus on three practical applications in Section 5.3, namely, local orientation estimation in filamentous structures, image reconstruction from a primal sketch and image denoising using BLS-GSM. We use our novel optimally localized wavelet profiles and compare them against well-known wavelet profiles such as the

Papadakis, Meyer, and Simoncelli wavelets.

5.2 Localized Isotropic Wavelets

In this section, we present a general framework that relies on the calculus of variations to find the optimal wavelet with respect to a given localization measure. We restrict ourselves to wavelets specified in Section 2.3.4 and focus on two natural classes of moment-based measures. Due to lack of our knowledge about the underlying process in practice, the choice of these natural measures of localization is the most justifiable. Additionally, the theoretical justifications that come in the sequel along with their nice properties from optimization view point are the other reasons for proposing these criteria.

First, regarding Section 2.3.4, if $\psi(r)$ is the radial profile of the mother wavelet of an isotropic tight wavelet frame with a fast filterbank implementation, then we have

$$\psi(r) = \mathcal{H}\{h\}(r) = \int_0^\infty h(t) \, J_0(rt) \, t \, \mathrm{d}t. \tag{5.1}$$

in which $h(\rho)$ is the radial profile of the Fourier transform of ψ that is supported on $[\frac{\pi}{4}, \pi]$ and satisfies

$$\sum_{i \in \mathbb{Z}} |h(2^i \rho)|^2 = 1 \quad \text{for} \quad \forall \rho \in \mathbb{R}^+ \backslash \{0\}. \tag{5.2}$$

Now, assume that the functional V is a given measure of localization. We shall consider that this measure operates in the Fourier domain. When V is Gâteaux differentiable, the natural method of minimizing it would be to use a variation of the steepest-descent algorithm [105]. To do this, we need the gradient of V. As V is a functional on an infinite-dimensional space, we have to rely on the calculus of variations to obtain its gradient. The second issue which should be taken care of is that, during the optimization steps, we have to be careful not to leave the set of tight frames. Thus, we need to characterize the projector onto the space of tight wavelet frames. Having these two major components, the outline of the optimization algorithm is given in Algorithm 4, in which $\nabla V(h)(\rho)$ denotes the gradient of V at ρ, and $\mathcal{P}$ denotes the orthogonal projector onto the space of tight wavelet frames. This corresponds to the standard projected gradient descent algorithm in an infinite dimensional space [106].

Algorithm 4: Most Localized Wavelet

1: **initialize:** $h \in L_2([0, \infty])$
2: **initialize:** $\eta > 0$
3: **repeat**
4:    $\tilde{h} \leftarrow h - \eta \, \nabla V\{h\}$
5:    $h \leftarrow \mathcal{P}\{\tilde{h}\}$
6: **until** h converges
7: **return** h

5.2.1 Measures of Localization

We now propose four measures of localization split between two natural classes. The first class consists of measures of the variance. Variance is the most well-known measure of localization as less variance implies more concentration around the center. In addition, we know from the uncertainty principle that the best achievable localization of a function is inversely proportional to the localization of its Fourier transform, and vice versa. More precisely, for a function $\psi : \mathbb{R}^2 \to \mathbb{R}$, we have [107]

$$\frac{\int_{\mathbb{R}^2} \|\mathbf{x}\|^2 \psi(\mathbf{x})^2 \mathrm{d}\mathbf{x}}{\int_{\mathbb{R}^2} \psi(\mathbf{x})^2 \mathrm{d}\mathbf{x}} \cdot \frac{\int_{\mathbb{R}^2} \|\boldsymbol{\omega}\|^2 |\mathcal{F}\{\psi\}(\boldsymbol{\omega})|^2 \mathrm{d}\boldsymbol{\omega}}{\int_{\mathbb{R}^2} |\mathcal{F}\{\psi\}(\boldsymbol{\omega})|^2 \mathrm{d}\boldsymbol{\omega}} \geq \frac{1}{16\pi^4}. \tag{5.3}$$

The first term of the left-hand side is the variance of the wavelet itself, and the second term is the variance of its Fourier transform. Thus, setting an upper bound on the variance in one domain imposes a lower bound on the variance in the other domain. In practice, we are interested in bandlimited wavelets, which implies that the variance of the wavelet in the Fourier domain is bounded from above. Thus, the variance of the wavelet in the space domain is bounded from below. Since we are interested in finding the wavelet profile that attains the minimum value for the variance, our first measure of localization is given by

$$V_{2\mathrm{D}}(\psi) = \frac{\int_{\mathbb{R}^2} \|\mathbf{x}\|^2 \psi(\mathbf{x})^2 \mathrm{d}\mathbf{x}}{\int_{\mathbb{R}^2} \psi(\mathbf{x})^2 \mathrm{d}\mathbf{x}} = \frac{\int_0^\infty r^2 \psi(r)^2 \, r \mathrm{d}r}{\int_0^\infty \psi(r)^2 \, r \mathrm{d}r}. \tag{5.4}$$

Isotropic wavelets are often used in a directional framework, for instance by applying the Riesz transform or by applying an angular mask. This suggests that the variance of the one-dimensional radial profile of the isotropic wavelet can also be a good candidate

for the measure of localization. Hence, we propose the second variance-based measure of localization

$$V_{1D}(\psi) = \frac{\int_0^\infty r^2 \psi(r)^2 \, dr}{\int_0^\infty \psi(r)^2 \, dr}. \tag{5.5}$$

It measures the spread of the wavelet along each radial line.

The second class of measures focuses on the localization of the wavelet coefficients rather than that of the wavelet profile. More precisely, the energy of a function computed over some spatial neighbourhood should be well represented by the wavelet coefficients associated to that neighbourhood and its vicinity. According to [108], if $f = \sum_{m \in \mathbb{Z}^2} f_m$ is an L_2-function from $\mathbb{R}^2$ to $\mathbb{R}$ and f_m is the restriction of f to the unit square centered at m, then

$$|\langle f_m, \psi_{i,\mathbf{k}} \rangle| \leq C_i \left(\int_{\mathbb{R}^2} \|\mathbf{x}\|^2 \psi(\mathbf{x})^2 d\mathbf{x} \right)^{1/2} \|f_m\|_2 \, |\mathbf{k} - m|^{-1}, \tag{5.6}$$

where C_i is a constant that only depends on the scale i and is independent of ψ and f. The ℓ_1-norm of a vector is denoted as $|\cdot|$ and the standard L_2-norm of a function as $\|\cdot\|_2$. Thus, as the wavelet $\psi_{i,\mathbf{k}}$ gets further from position m, the contribution of f_m in the corresponding wavelet coefficient decays. Moreover, the rate of decay is controlled by a constant that corresponds to the unnormalized variance of the wavelet profile ψ. We hence propose this value as a third measure of localization, this time for the wavelet coefficients. We define

$$U_{2D}(\psi) = \int_{\mathbb{R}^2} \|\mathbf{x}\|^2 \psi(\mathbf{x})^2 d\mathbf{x} = \int_0^\infty r^2 \psi(r)^2 \, r dr. \tag{5.7}$$

Accounting for the fact that isotropic wavelets are often used in a directional setting, we define the unnormalized variance of the one-dimensional radial profile of the wavelet in analogy to (5.5) as our last measure of localization

$$U_{1D}(\psi) = \int_0^\infty r^2 \psi(r)^2 \, dr. \tag{5.8}$$

To summarize, the first class of measures focuses on the localization of the shape of the wavelet in the space domain while the second class of measures describes the localization of the wavelet coefficients. We thus expect the first class to match applications that benefit from a local wavelet analysis. The second class should, on the contrary, be more appropriate in the context of applications that involve some form of wavelet-domain N-term

approximation. For both classes of measures, the index 2D indicates that the spread of the wavelet is a measure over the whole space. The index 1D, conversely, measures the spread in each radial direction. In applications where a directionality component is imposed on top of the isotropic profile (*e.g.*, in detection tasks using steerable filters), we predict that profiles with the best radial localization should exhibit the best performances.

5.2.2 Gradient of the Functionals and Projector onto the Space of Tight Frames

In this subsection, we give the explicit expressions of the two major components that are required for implementing Algorithm 4: the gradient of the localization criterion and the operator $\mathcal{P}$ that projects a solution onto the space of tight wavelet frames.

For formalization purpose, we consider h as a function in the weighted L_2-space of $[0, \infty]$, $\mathcal{L}_w$, whose inner product is defined as

$$\langle f, g \rangle_w = \int_0^\infty f(\rho)g(\rho)w(\rho)\mathrm{d}\rho, \tag{5.9}$$

where w is a strictly positive weighting function.

We shall see that the inclusion of such a weight will provides us with some degrees of freedom to design the projection operator, which can then be used advantageously to simplify the implementation. Since we are operating in a Hilbert space, the choice of w specifies the metric and hence the form of the orthogonal projection operator $\mathcal{P}_w : \mathcal{L}_w \to \mathcal{T}$, where $\mathcal{T}$ is the set of functions satisfying the tight frame property (5.2). Now, if we switch to another weighting function v, we can define another "orthogonal" projection operator $\mathcal{P}_v : \mathcal{L}_v \to \mathcal{T}$ which is such that $\mathcal{P}_v h = h$ for all $h \in \mathcal{T}$ as well as $\mathcal{P}_v \mathcal{P}_v f = \mathcal{P}_v f$ for all $f \in \mathcal{L}_w$. In other words, $\mathcal{P}_v$ also constitutes a valid projector for the space $\mathcal{L}_w$ (including $L_2([0, \infty])$ with $w = 1$), albeit not necessarily the one that minimizes the corresponding approximation error. The important point here is that using $\mathcal{P}_v$ rather than $\mathcal{P}_w$ will not modify the outcome of the optimization process. Another way to put it is that the underlying Karush-Kuhn-Tucker conditions of optimality in the Lagrange multiplier method of optimization (see [109]) are independent of the actual choice of the Hilbert space $\mathcal{L}_w$.

Now, defining

$$A_m(h) = \int_0^\infty r^m \mathcal{H}\{h\}(r)^2 \mathrm{d}r, \tag{5.10}$$

we see that $V_{2\mathrm{D}}$, $V_{1\mathrm{D}}$, $U_{2\mathrm{D}}$, and $U_{1\mathrm{D}}$ can be easily written in terms of A_m for different m. Thus, we only need to compute the gradient of A_m. Using the basic rules of differentiation, we have

$$\nabla V_{2\mathrm{D}}\{h\} = \frac{1}{A_1(h)}\nabla A_3\{h\} - \frac{A_3(h)}{A_1(h)^2}\nabla A_1\{h\}, \tag{5.11}$$

$$\nabla V_{1\mathrm{D}}\{h\} = \frac{1}{A_0(h)}\nabla A_2\{h\} - \frac{A_2(h)}{A_0(h)^2}\nabla A_0\{h\}, \tag{5.12}$$

$$\nabla U_{2\mathrm{D}}\{h\} = \nabla A_3\{h\} \tag{5.13}$$

and

$$\nabla U_{1\mathrm{D}}\{h\} = \nabla A_2\{h\}. \tag{5.14}$$

The main point for our purpose is that the functional $A_m : \mathcal{L}_w \to \mathbb{R}$ is Gâteaux differentiable and that its infinite-dimensional gradient can be obtained explicitly as shown in Appendix 5.A. The ultimate outcome is

$$\nabla A_m\{h\}(\rho) = \frac{2\rho}{w(\rho)}\int_0^\infty r^m \mathcal{H}\{h\}(r) J_0(\rho r)\mathrm{d}r, \tag{5.15}$$

where $\mathcal{H}\{h\}$. Note that this functional gradient depends on the weighting function of the space, w.

    The final ingredient for our algorithm is the operator that projects a function onto the set of tight-frames. Here, unlike in the case of the gradient, the computational complexity of the orthogonal projector is strongly dependent on the choice of w. In fact, in the following theorem, we will see that there is a very specific weighting function $w = v$ for which we can have a closed-form formula for the required projector.

Theorem 4. *Let $\mathcal{L}_w$ be the Hilbert space whose inner product is specified by (5.9) and let $\mathcal{T}$ be the set of functions in $\mathcal{L}_w$ satisfiying the tight frame property (2.49). Then, the operator*

$$\mathcal{P}\{\tilde{h}\}(\rho) = \frac{\tilde{h}(\rho)}{\sqrt{\sum_{i\in\mathbb{Z}}\tilde{h}^2(2^i\rho)}} \tag{5.16}$$

is a projector from $\mathcal{L}_w \to \mathcal{T}$. In particular, it is the orthogonal projector $\mathcal{L}_v \to \mathcal{T}$ for the weighting function

$$v(t) = 2^i \quad for \quad \frac{\pi}{2^{i+1}} \le t \le \frac{\pi}{2^i}. \tag{5.17}$$

The proof is given in Appendix 5.B.

The form (5.16) of the projector is intuitively very reasonable. The simplification results from the choice of the appropriate metric in the proof of the theorem. The result is non-obvious a priori because this is the only instance of w for which we are able to carry out the computation to the end. Constraint (5.2) is the equation of the unit infinite-dimensional sphere for each value of ρ. The projector (5.16) is therefore projecting $\tilde{h}$ on the unit sphere for each value of ρ. The theorem ensures that we have an equivalence between the ℓ_2-norm projection in the space of sequences corresponding to each value of ρ and a projection in the weighted L_2-space of functions on $\mathbb{R}$ for the weighting function v specified by (5.17).

5.2.3 Numerical Optimization

We apply our numerical optimization algorithm for each of the proposed measures of localization. First of all, it has been shown analytically [108] that the Simoncelli wavelet minimizes the criterion $U_{2\mathrm{D}}$. Thus, we already know the optimal profile with respect to the measure $U_{2\mathrm{D}}$. Regarding $V_{2\mathrm{D}}$, $V_{1\mathrm{D}}$, and $U_{1\mathrm{D}}$, we run Algorithm 4 by uniformly taking 512 samples of $h(\rho)$ for ρ in $\pi/4$ to π on a logarithmic scale, hence simplifying the computation of the projection map (5.16). To calculate $\mathcal{H}\{h\}(r)$, we compute the integral of (2.48) from 0 to 300 relying on the trapezoid method with 3000 intervals. The algorithm is left to run until absolute variations of $V(h)$ fall under 10^{-3}.

In this optimization settings, we obtain the minimum values 1.73, 0.39, and 1.64 for $V_{2\mathrm{D}}(h)$, $V_{1\mathrm{D}}(h)$, and $U_{1\mathrm{D}}(h)$, respectively. However, due to the Gibbs phenomenon that results from the truncation of the Fourier transform, the resulting wavelets exhibit ringing artifacts. In order to remove these effects and obtain a smooth profile for practical applications, we fitted a closed-form formula to the numerically obtained wavelets. We thus propose four wavelets named $h_{V_{2\mathrm{D}}}$, $h_{U_{2\mathrm{D}}}$, $h_{V_{1\mathrm{D}}}$, and $h_{U_{1\mathrm{D}}}$, which correspond to each of the considered measures of localization. From (2.51), (2.52), and (2.54), it is sufficient to specify the wavelet profile either on $\left[\frac{\pi}{4}, \frac{\pi}{2}\right]$ or on $\left[\frac{\pi}{2}, \pi\right]$ to describe it entirely. The expressions of $h_{V_{2\mathrm{D}}}$ and $h_{U_{2\mathrm{D}}}$ are more easily given on the interval $\left[\frac{\pi}{4}, \frac{\pi}{2}\right]$ as

$$h_{V_{2\mathrm{D}}}(\rho)\big|_{\rho\in[\frac{\pi}{4},\frac{\pi}{2}]} = \frac{\sqrt{6} - \sqrt{1 + 20(\frac{2\rho}{\pi} - 1)^2}}{\sqrt{6} - 1}, \tag{5.18}$$

$$h_{U_{2\mathrm{D}}}(\rho)\big|_{\rho\in[\frac{\pi}{4},\frac{\pi}{2}]} = \cos\left(\frac{\pi}{2}\log_2\frac{2\rho}{\pi}\right). \tag{5.19}$$

We recall that $h_{U_{2D}}$ corresponds to the Simoncelli wavelet. The profiles $h_{V_{1D}}$ and $h_{U_{1D}}$ are better expressed on the interval $[\frac{\pi}{2}, \pi]$ as

$$h_{V_{1D}}(\rho)\big|_{\rho \in [\frac{\pi}{2}, \pi]} = \left(\log_2 \frac{\pi}{\rho} - 0.005 \sqrt{\frac{\pi}{\rho}} \sin \left(\pi \log_2 \frac{\pi}{\rho} \right) \right)^{\frac{2}{5}}, \qquad (5.20)$$

$$h_{U_{1D}}(\rho)\big|_{\rho \in [\frac{\pi}{2}, \pi]} = \sqrt{\frac{(\log_2 \frac{\pi}{\rho} + 0.6)^4 - 0.6^4}{1.6^4 - 0.6^4}}. \qquad (5.21)$$

The radial profiles of these wavelets are shown in Figures 5.1 and 5.2 in Fourier and space domains, respectively. We have that $V_{2D}(h_{V_{2D}}) = 1.74$, $V_{1D}(h_{V_{1D}}) = 0.40$, and $U_{1D}(h_{U_{1D}}) = 1.65$. These values are only marginally suboptimal. The values of the different measures of localization for each of these wavelets as well as for more traditional ones are given in Table 5.1. Moreover, we note that Figure 5.2 confirms our expectation that $h_{V_{1D}}$ has the most localized profile shape.

Table 5.1: Localization of Different Wavelets Measured by V_{2D}, V_{1D}, U_{2D}, and U_{1D} (Equations (5.4)-(5.5) and (5.7)-(5.8))

Wavelet type	Localization			
	V_{2D}	V_{1D}	U_{2D}	U_{1D}
$h_{V_{2D}}$ (5.18)	**1.74**	0.44	3.88	2.19
$h_{V_{1D}}$ (5.20)	$+\infty$	**0.40**	$+\infty$	2.03
$h_{U_{2D}}$ (Simonceli) (5.19)[94]	1.84	0.46	**3.55**	1.93
$h_{U_{1D}}$ (5.21)	$+\infty$	0.52	$+\infty$	**1.65**
Papadakis [102]	2.06	0.49	4.93	2.52
Meyer [33]	2.88	0.66	6.04	2.61
Shannon [89]	$+\infty$	$+\infty$	$+\infty$	$+\infty$

The measures V_{2D} and U_{2D} can also be interpreted as the normalized and unnormalized third-order moment of the radial profile of the wavelet, respectively, while V_{1D} and U_{1D} correspond to its normalized and unnormalized second-order moment. Furthermore, having finite values for higher-order moments in the space domain implies being smoother in the Fourier domain. Thus, the minimiziers of V_{2D} and U_{2D} necessarily have finite V_{1D} and U_{1D} values. However, the converse is not always true.

As we see in Table 5.1, the minimizers of V_{1D} and U_{1D} have infinite values for V_{2D} and U_{2D}. This is in accordance with the roughness of the profiles at points $\frac{\pi}{4}$, $\frac{\pi}{2}$, and π
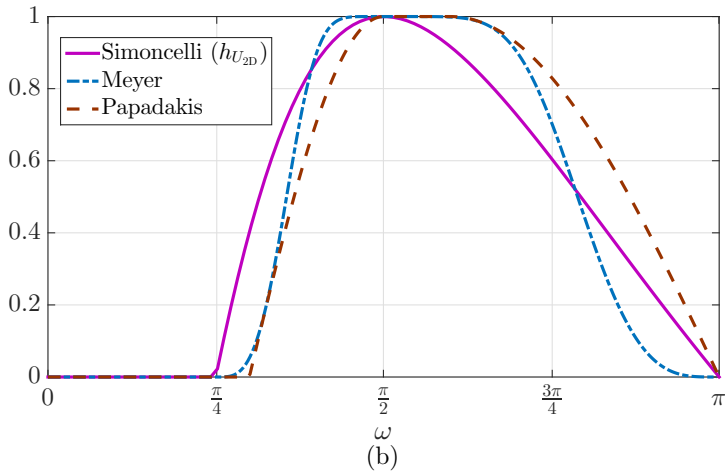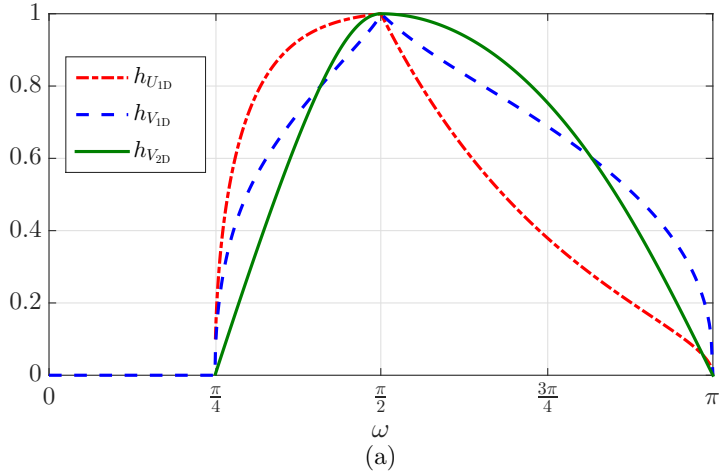
Figure 5.1: Wavelet profiles in Fourier domain for (a) the proposed optimal profiles, and (b) existing ones (Simoncelli, Meyer, and Papadakis).
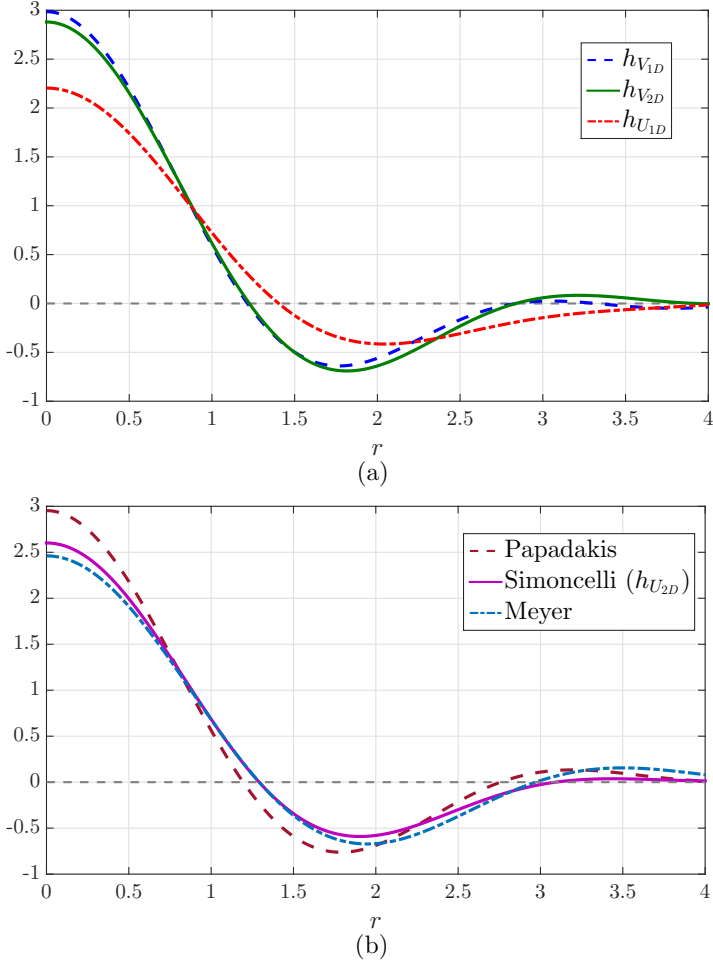
Figure 5.2: Wavelet profiles in space domain for (a) the proposed optimal profiles, and (b) existing ones (Simoncelli, Meyer, and Papadakis).

(see Figure 5.1). In the case of the Shannon wavelet, the discontinuities of the frequency-domain profile results in a slow decay in the spatial domain that brings V_{1D} and U_{1D} to infinity. From Table 5.1, we observe that the Simoncelli wavelet is not only optimal for the criterion U_{2D}, but also exhibits reasonable values for other measures of localization.

5.3   Applications

We now demonstrate the benefit of well-localized wavelet profiles for practical applications. In particular, we study the performance of our wavelets and compare them against other existing popular profiles for the problems of local orientation estimation, image reconstruction from edges and denoising.

5.3.1   Estimation of Local Orientation

In this section, we focus on the task of estimating the local orientation of ridge-like objects (*e.g.*, filaments) using a steerable ridge detector. We construct Hessian-like wavelets, the design of which is made easy by selecting an isotropic kernel and applying the generalized Riesz-wavelet transform using the appropriate shaping matrix, as described in [90, Section 5.1.3]. We refer the reader to [110] for a detailed description on how to steer Hessian filters to retrieve the orientation corresponding to the best response of the ridge detector at every point of the image. To perform a multiscale ridge detection at every location using the Hessian filter, we go through every scale of the wavelet pyramid and select the one where the strongest filter response can be found. The final output of our experiment is therefore an angle map with the same dimensions as the input image, and which contains at each pixel the estimated local orientation yielding the best ridge filter response. The Riesz-wavelet transform [90] and the extraction of local orientation estimation for each point of the input image have been implemented as a Java-based plug-in for the open-source image-analysis software ImageJ [111].

In our experiment, we rely on a 512×512 pixels 8-bits image (Figure 5.3a) in which several regions of interest (ROIs) made of short line segments have been manually selected by an expert, and where local orientation should be estimated. The angle that each of the manually placed ROI form with the horizontal direction is considered as ground truth and corresponds to the orientation that shall automatically be retrieved. We estimate the local orientation of each ROI with the trivial isotropic profile (Shannon [89]), several popular isotropic wavelet profiles (Simoncelli [94], which corresponds to $h_{U_{2D}}$ (5.19), Pa-

padakis [102], and Meyer [33]), as well as with the wavelets we propose ($h_{U_{1D}}$ (5.21), $h_{V_{2D}}$ (5.18) and $h_{V_{1D}}$ (5.20)). To obtain an estimation of the local orientation of each ROI, we average the orientation estimates provided by the steerable filter under the ROI (*i.e.*, we average the values of the pixels that belong to the line segment composing the ROI). We investigate the quality of each of the local orientation estimate by comparing the ground truth orientation with the automatically retrieved one (Table 5.2). We also report the absolute error between the ground truth and each of the estimates in Table 5.3. The experiment is conducted using 4 scales of wavelet decomposition.

(a)   (b)
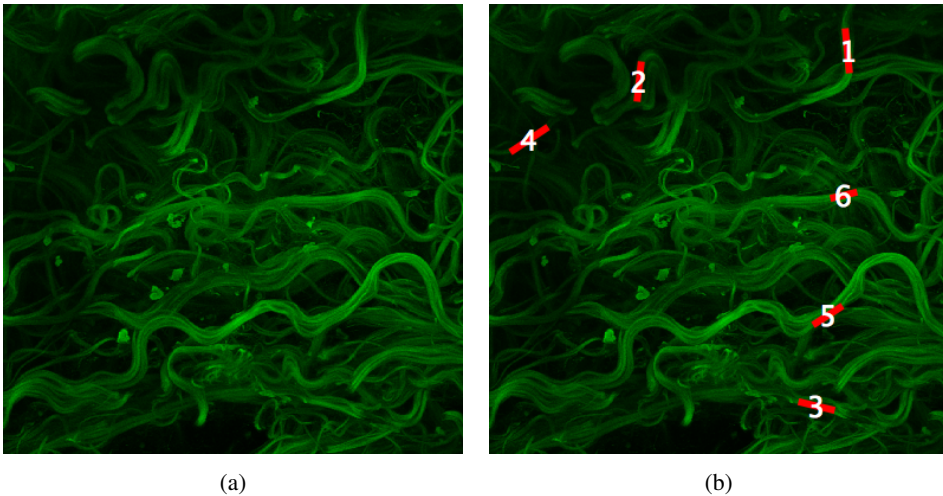
Figure 5.3: Estimation of local orientation. (a) Original filaments image from [112], and (b) overlaid local regions of interest with their label.

In this experiment, we observe that best results are obtained with $h_{U_{1D}}$. This can be explained by the fact that the two classes of wavelets h_U and h_V are optimized for different applications. The profiles of $h_{V_{2D}}$ and $h_{V_{1D}}$ are most localized in the spatial domain as they optimize V_{2D} (5.4) and V_{1D} (5.5), while $h_{U_{2D}}$ and $h_{U_{1D}}$ optimally localizes wavelet coefficients following U_{2D} (5.7) and U_{1D} (5.8). The estimation of local orientation is better when the wavelet response is strong and well localized, as ridges (here, the filaments) are then more accurately detected. In the present experiment, a profile maximizing wavelet

coefficients localization, and hence a criterion of the class U, is therefore desirable. This is confirmed by the results and the good performance obtained by $h_{U_{1D}}$ and $h_{U_{2D}}$. Also, among the class of h_U wavelets, one observes that $h_{U_{1D}}$ outperforms $h_{U_{2D}}$. An argument for this is that a 1-dimensional design is more suitable for steerable wavelets since they are inherently directional. In the current application, the wavelets align themselves with ridges so that the U_{1D} criterion, which measures the spread in each radial direction, is the most appropriate one.

5.3.2 Image Reconstruction from Edges

The experiment we study here is image approximation from a reduced set of wavelet-based edges. First, a multiscale *primal sketch* [100], or *edge map* [101], is extracted from the set of wavelet coefficients of the image. An approximation of the original image is then recovered from this small subset of coefficients relying on constrained optimization.

To extract a multiscale edge map from the input image, we rely on a gradient-like wavelet framework. It is implemented with the help of the generalized Riesz-wavelet transform [90], and of an appropriate shaping matrix [90, Section 5.1.1] that yields a pair of x- and y-derivative wavelets. Edges in the multiscale gradient signal are then detected based on a wavelet-domain version of the Canny edge detector, which includes non-maximum suppression and hysteresis thesholding [100]. Note that the Canny edge detector requires an estimation of the strength and orientation of the gradient for each point of the image, which is obtained by steering the gradient-like wavelets at every scale and location. The final edge map is composed of the wavelet coefficients retained by the multiscale edge detector. To preserve the graylevel information of the image, all coefficients of the lowpass residual of the wavelet decomposition are saved.

Reconstruction is then formulated as the constrained optimization problem

$$\text{minimize } \|\boldsymbol{z}\|_1 \tag{5.22}$$

$$\text{subject to } \boldsymbol{z} = \mathbf{W}^H \boldsymbol{f} \tag{5.23}$$

$$z_{\boldsymbol{i}} = q_{\boldsymbol{i}}, \, \forall \, \boldsymbol{i} \in S, \tag{5.24}$$

with variables $\boldsymbol{z}$ and $\boldsymbol{f}$, where S is the set of indices for the wavelet coefficients that are part of the edge map, $\mathbf{W}^H$ the wavelet-analysis operator, and $\boldsymbol{f}$ an image. Finally, $q_{\boldsymbol{i}}$ denotes the wavelet coefficient of the original image at location $\boldsymbol{i}$, where $\boldsymbol{i}$ is an index of 2-D position and scale. This formulation is motivated by two main principles. First, we aim at conserving the elements of the edge-map in order to reconstruct the image. We refer

Table 5.2: Estimation of Local Orientations for the Regions of Interest depicted in Figure 5.3b

Wavelet type	Angle [°]					
	1	2	3	4	5	6
Ground Truth	*95.36*	*129.09*	*83.88*	*166.50*	*33.02*	*13.24*
$h_{V_{2D}}$ (5.18)	95.88	133.19	80.48	164.66	36.74	15.95
$h_{V_{1D}}$ (5.20)	96.08	133.59	79.14	163.13	37.44	15.90
$h_{U_{2D}}$ (Simonceli) (5.19)[94]	96.42	133.75	78.37	164.55	37.72	16.67
$h_{U_{1D}}$ (5.21)	**95.57**	**132.88**	**83.19**	**165.07**	**36.12**	**15.53**
Papadakis	97.31	134.91	79.46	163.39	41.65	18.04
Meyer	97.15	134.67	79.65	164.25	41.61	18.75
Shannon	96.78	136.61	76.97	136.10	70.73	22.16

Table 5.3: Error in Local Orientation Estimates for the Regions of Interest depicted in Figure 5.3b

Wavelet type	Absolute error [°]					
	1	2	3	4	5	6
$h_{V_{2D}}$ (5.18)	0.52	4.10	3.41	1.85	3.72	2.71
$h_{V_{1D}}$ (5.20)	0.73	4.49	4.75	3.38	4.41	2.66
$h_{U_{2D}}$ (Simonceli) (5.19)[94]	1.07	4.66	5.52	1.95	4.69	3.43
$h_{U_{1D}}$ (5.21)	**0.22**	**3.79**	**0.69**	**1.43**	**3.09**	**2.29**
Papadakis	1.95	5.82	4.42	3.11	8.62	4.80
Meyer	1.79	5.57	4.24	2.26	8.59	5.51
Shannon	1.43	7.51	6.91	30.41	37.71	8.92

to them as wavelet-based edges, as they are the output of a Canny edge detector applied on the wavelet transform of the image. This gives us constraint (5.24), which imposes the elements q_i, $i \in S$ to be fixed during the optimization process. Second, we want the estimated missing wavelet coefficients to project back onto an image. Knowing that images are sparse in the wavelet domain, we impose sparsity by minimizing the ℓ_1-norm of the wavelet transform z of the image, yielding (5.22). Our problem thus amounts to minimizing a convex functional under a finite set of linear constraints, which guarantees the existence of a feasible minimum that can be reached using appropriate optimization algorithms.

Here, we propose an improvement of the reconstruction algorithm in [101] that relies on a gradient descent of the augmented Lagrangian. Our new algorithm is based on the alternating-direction method of multipliers (ADMM), which motivates the introduction of the auxiliary variable z. ADMM is a method known to converge very fast to an acceptable solution and that guarantees the residual to be brought to zero. In practice, it is observed that the fast and moderately good estimate provided after 30 iterations of the algorithm is already visually satisfactory. We refer the reader to [113] for a complete description of the method.

To reconstruct the image with ADMM, we first form the augmented Lagrangian

$$\mathcal{L}(z, f, \lambda) = \|z\|_1 + \lambda^T \left(z - \mathbf{W}^H f\right) + \frac{\mu}{2} \left\|z - \mathbf{W}^H f\right\|_2^2, \tag{5.25}$$

where μ is a step size that can be adapted to influence the speed of convergence. We rewrite (5.25) in terms of the scaled dual variable $u = \frac{\lambda}{\mu}$ in order to obtain simpler mathematical expressions, yielding

$$\mathcal{L}(z, f, u) = \|z\|_1 + \frac{\mu}{2} \left\|z - \mathbf{W}^H f + u\right\|_2^2 - \frac{\mu}{2} \|u\|_2^2. \tag{5.26}$$

The ADMM algorithm for our problem thus consists of the three successive iterations

$$z^{(k+1)} = \arg\min_{z} \mathcal{L}(z, f^{(k)}, u^{(k)}), \tag{5.27}$$

$$f^{(k+1)} = \arg\min_{f} \mathcal{L}(z^{(k+1)}, f, u^{(k)}), \tag{5.28}$$

$$u^{(k+1)} = u^{(k)} + \left(z^{(k+1)} - \mathbf{W}^H f^{(k+1)}\right). \tag{5.29}$$

The update for $\boldsymbol{z}$ can be rewritten in an element-wise manner as

$$\boldsymbol{z}^{(k+1)} = \arg\min_{\boldsymbol{z}} \left(\|\boldsymbol{z}\|_1 + \frac{\mu}{2} \left\| \boldsymbol{z} - \mathbf{W}^H \boldsymbol{f}^{(k)} + \boldsymbol{u}^{(k)} \right\|_2^2 \right) \tag{5.30}$$

$$= \arg\min_{\boldsymbol{z}} \left(\sum_{\boldsymbol{i} \in I} |z_{\boldsymbol{i}}| + \frac{\mu}{2} \sum_{\boldsymbol{i} \in I} \left| z_{\boldsymbol{i}} - v_{\boldsymbol{i}}^{(k)} \right|^2 \right), \tag{5.31}$$

where we denote by I the set of indices for all wavelet coefficients of the image. We introduced $v_i^{(k)} = [\mathbf{W}^H \boldsymbol{f}^{(k)}]_i - u_i^{(k)}$ for convenience. For all elements $\boldsymbol{i} \in S$, (5.24) imposes that $z_{\boldsymbol{i}} = q_{\boldsymbol{i}}$, and no further computations are required. For $\boldsymbol{i} \notin S$, $z_{\boldsymbol{i}}$ should be colinear with $v_{\boldsymbol{i}}$ in order to annihilate the second term in (5.31). Plugging $z_{\boldsymbol{i}} = Cv_{\boldsymbol{i}}$ into (5.31) and solving for the optimal constant C brings us to the component-wise expression of the $\boldsymbol{z}$ update

$$z_{\boldsymbol{i}}^{(k+1)} = \begin{cases} q_{\boldsymbol{i}}, & \text{if } \boldsymbol{i} \in S \\ \left(1 - \frac{1}{\mu \left| v_{\boldsymbol{i}}^{(k)} \right|} \right)_+ v_{\boldsymbol{i}}^{(k)}, & \text{if } \boldsymbol{i} \notin S, \end{cases} \tag{5.32}$$

where $(\cdot)_+ = \max(0, \cdot)$ corresponds to the shrinkage of $v_{\boldsymbol{i}}^{(k)}$. Then, updating $\boldsymbol{f}$ boils down to an unconstrained quadratic optimization problem. It can hence be performed by taking the partial derivative of the augmented Lagrangian and solving for zero. This yields

$$\boldsymbol{f}^{(k+1)} = \arg\min_{\boldsymbol{f}} \left\| \boldsymbol{z} - \mathbf{W}^H \boldsymbol{f} + \boldsymbol{u} \right\|_2^2 \tag{5.33}$$

$$= \left(\mathbf{W}\mathbf{W}^H \right)^{-1} \mathbf{W} \left(\boldsymbol{z}^{(k+1)} + \boldsymbol{u}^{(k)} \right). \tag{5.34}$$

As $\mathbf{W}$ forms a tight frame, (5.34) can be further simplified by observing that $\mathbf{W}\mathbf{W}^H = \mathbf{I}$. Finally, $\boldsymbol{u}$ is simply modified through a linear update.

In practice, the algorithm is initialized with a $\boldsymbol{z}^{(0)}$ composed of all the wavelet coefficients $q_{\boldsymbol{i}}$, $\boldsymbol{i} \in S$ retained in the edge map and the complete lowpass residual of the image to reconstruct. Then, $\boldsymbol{f}^{(0)}$ is initialized as an image entirely composed of pixels with value zero and $\mathbf{W}^H \boldsymbol{f}^{(0)}$ is obtained by taking its wavelet transform. Finally, $\boldsymbol{u}^{(0)}$ is set as a pyramid of images composed only of zeros and having the same number of scales and dimensions as $\mathbf{W}^H \boldsymbol{f}^{(0)}$. The parameter μ is empirically set to 10^6. With these settings, between thirty and fifty ADMM iterations were observed to be sufficient to reach "visual

convergence", which corresponds to a situation where additional iterations bring unnoticeable visual improvements. Note that, through the whole optimization procedure, the left-multiplication by $\mathbf{W}^H$ or $\mathbf{W}$ corresponds to performing a wavelet analysis or synthesis, respectively. The algorithm can thus be executed in reasonable time as these two operations can be performed efficiently with the help of a filterbank implementation. To illustrate this experiment, we give in Figure 5.4 the original Cameraman image, its reconstruction from 7% of the wavelet coefficients, as well as the mask containing the coefficients retained by our multiscale wavelet-based edge detector.

We implemented the Riesz-wavelet transform [90] as well as the edge-map extraction and subsequent image reconstruction using ADMM as a Java-based plug-in for the open-source image-analysis software ImageJ [111].

We gather results on a set of 5 standard test images, namely, House, Pirate, Peppers, Lena, Bridge, Cameraman, and Einstein, all being 512×512 pixels grayscale images. We run the same experiments with the trivial isotropic profile (Shannon [89]), several popular isotropic wavelet profiles (Simoncelli [94], which happens to be $h_{U_{2D}}$ (5.19), Papadakis [102], and Meyer [33]), and finally with our proposed wavelets ($h_{U_{1D}}$ (5.21), $h_{V_{2D}}$ (5.18) and $h_{V_{1D}}$ (5.20)). We investigate the reconstruction performance of the different wavelets in terms of the PSNR of the reconstructed image. We start by retaining 7% of the total number of wavelet coefficients in the image. These 7% are chosen among the set of wavelet-based edges retained by our multi-scale Canny edge detector[1]. Note that, as our test images all have the same size, this percentage corresponds to the same absolute number of coefficients in each case. All experiments are conducted using 4 scales of decomposition. Reconstruction results obtained after 30 iterations of the ADMM algorithm are shown in Table 5.4. In order to allow for a visual comparison of the performance, we also show in Figure 5.5 close-ups of the Lena image reconstructed using the different wavelet profiles. We here observe that $h_{U_{1D}}$ outperforms the other wavelets. Further experiments of reconstruction using a set of edge coefficients corresponding to 1 to 7% of the total number of wavelet coefficients in the image allows us to reach similar conclusions, as seen in Figure 5.6. Here, only results on Lena and Cameraman are displayed, as they are representative of the results observed in the remaining test images. This confirms that the proposed $h_{U_{1D}}$ profile is better for reconstruction than the other wavelets considered in this experiment.

Notice that, in this application, $h_{U_{1D}}$ followed by $h_{U_{2D}}$ outperforms in particular $h_{V_{2D}}$

---

[1]In practice, we adapt the percentage of coefficients retained by the multiscale Canny edge detector by changing the hysteresis thresholding parameters.
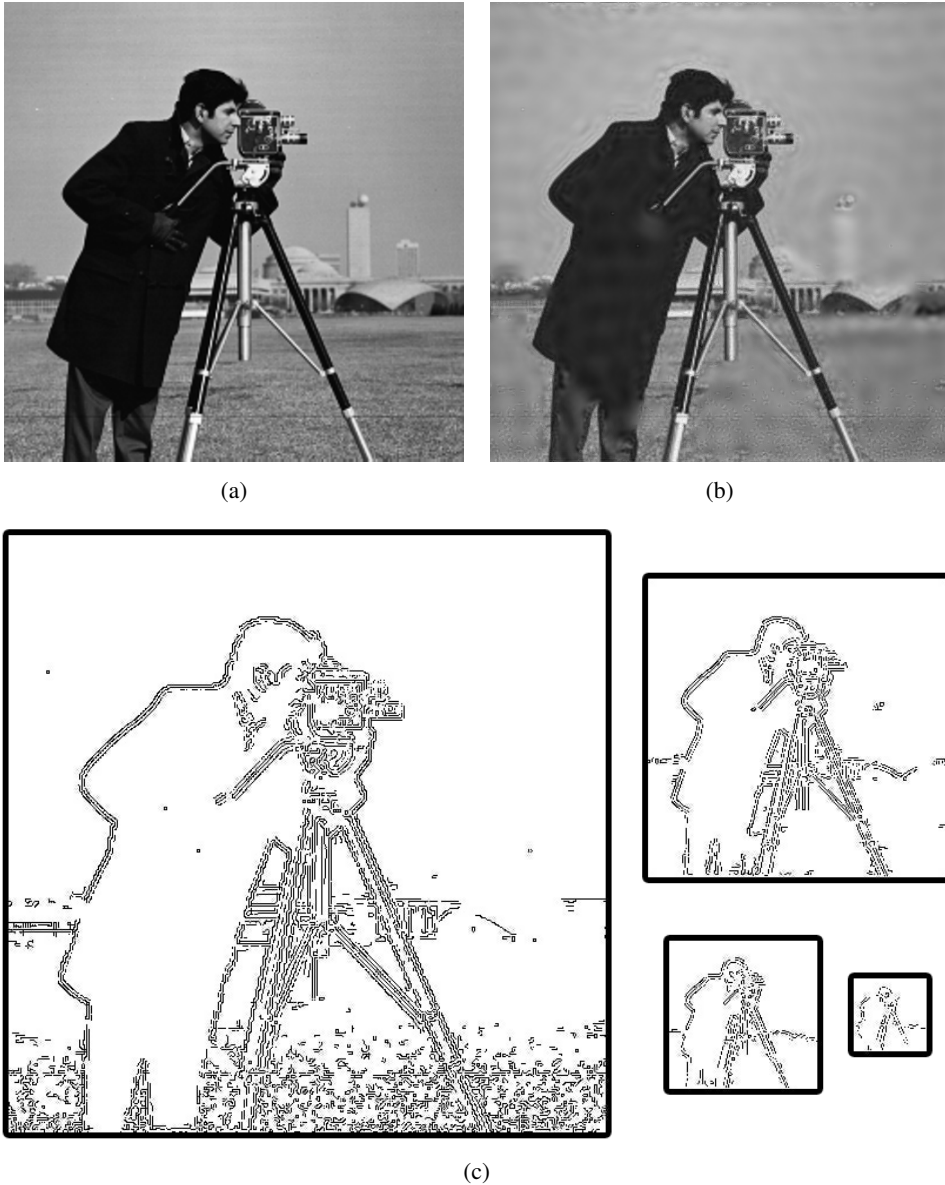
(a)   (b)

(c)

Figure 5.4: Wavelet-based edge reconstruction. (a) Original Cameraman image, (b) final result after reconstruction using $h_{U_{1\mathrm{D}}}$, and (c) binary masks featuring the wavelet coefficients saved for reconstruction at different scales. Here, 4 scales were used and 7% of the total number of coefficients were retained.

Table 5.4: Reconstruction from Wavelet-Based Edges (7% of Coefficients)

Wavelet type	House	Pirate	Peppers	PSNR [dB] Lena	Bridge	Cameraman	Einstein
$h_{V_{2D}}$ (5.18)	28.90	27.33	27.61	29.14	23.42	29.01	28.97
$h_{V_{1D}}$ (5.20)	29.72	27.32	27.96	30.04	23.99	29.70	29.28
$h_{U_{2D}}$ (5.19) (Simoncelli [94])	29.15	27.37	27.96	29.95	23.91	30.14	29.22
$h_{U_{1D}}$ (5.21)	**30.01**	**28.24**	**28.42**	**30.37**	**24.07**	**30.67**	**29.40**
Papadakis [102]	27.98	26.55	27.34	28.99	23.67	28.52	28.89
Meyer [33]	27.72	26.29	26.70	28.88	23.50	28.97	28.65
Shannon [89]	26.80	24.91	25.53	27.20	23.03	25.60	27.61

(a)   (b)   (c)
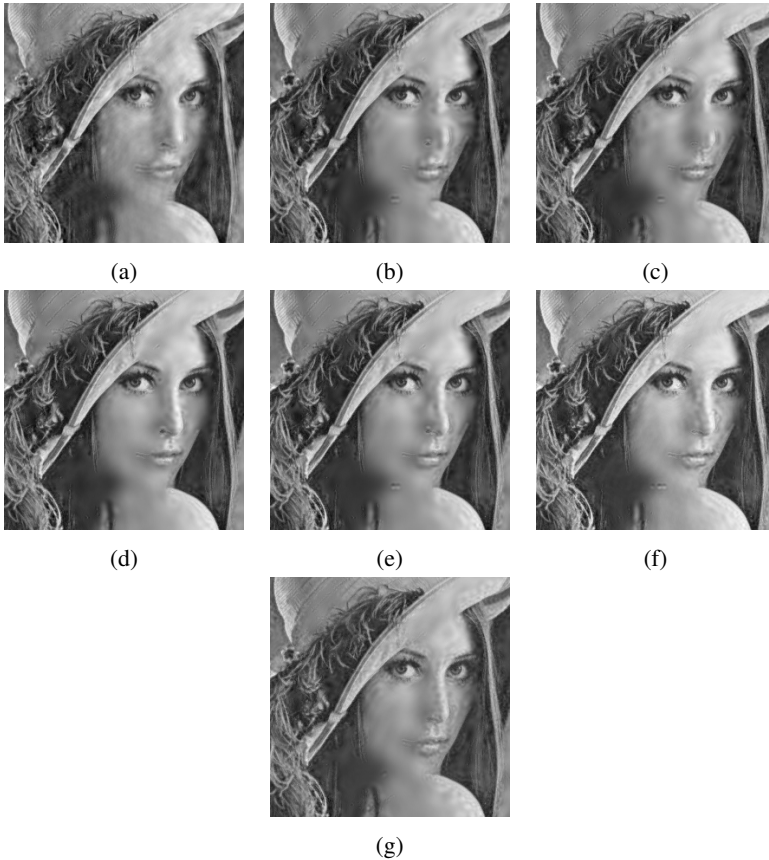
(d)   (e)   (f)

(g)

Figure 5.5: Wavelet-based edge reconstruction. Close-up of reconstruction of Lena relying on (a) Shannon, (b) Meyer, (c) Papadakis, (d) $h_{U_{1D}}$, (e) $h_{U_{2D}}$ (Simoncelli), (f) $h_{V_{1D}}$, and (g) $h_{V_{2D}}$. The best PSNR is achieved by $h_{U_{1D}}$ shown in (e). Here, 4 scales were used and 7% of the total number of coefficients were retained.

and $h_{V_{1D}}$. As the construction of the edge map in the current experiment relies on the same kind of framework as the filaments detection for local orientation estimation in Section 5.3.1 (namely, multiscale steerable filters), the same arguments hold for explaining these results. The reconstruction task obtains better results when the edge map contains sharper elements, which corresponds to better sets of edges. What matters most in this experiment is therefore again the optimal localization of the wavelet coefficients. This explains why the best performance is obtained with the profiles optimizing criterion of the class U ($h_{U_{1D}}$ and $h_{U_{2D}}$). Results can actually directly be interpreted from the values of U_{1D} provided in Table 5.1. Starting from the optimal $h_{U_{1D}}$, the next most localized profiles sorted by distance to the optimum are Simoncelli ($h_{U_{2D}}$), $h_{V_{1D}}$, $h_{V_{2D}}$, Papadakis and Meyer. The quality of reconstruction obtained by the different profiles and shown in Figure 5.6 follow the same pattern, with Papadakis and Meyer being the worst and followed by $h_{V_{2D}}$ and $h_{V_{1D}}$, $h_{U_{2D}}$ (Simoncelli), and finally $h_{U_{1D}}$, which achieves the best results.

5.3.3 BLS-GSM Denoising

The BLS-GSM algorithm [93] is a famous and very efficient approach for recovering noise-corrupted images. The motivation behind this method is the observed strong correlation between wavelet coefficients located at similar positions at various orientations and scales. More specifically, the properties of the image (*i.e.*, the neighborhood of each wavelet coefficient) are modeled by a Gaussian scale mixture model (GSM) in the multiresolution wavelet transform space. The original, noise-free value of each coefficient is then estimated using Bayesian least squares (BLS) under the Gaussian scale mixture model, and making use of the correlation between coefficients in the pyramid. The algorithm therefore improves the denoising by taking advantage of local similarities. In practice, BLS-GSM is performed on subbands of an oriented multiresolution transformation of the noise-corrupted image, which corresponds in the original implementation to the steerable pyramid with Simoncelli's wavelet profile. We orient readers interested in a more detailed description of BLS-GSM to the very comprehensive paper of Rajaei [114].

A Matlab implementation of BLS-GSM has been released by the authors of the initial paper [93]. In order to perform the following experiments, we modify this original implementation[2] by replacing the Simoncelli wavelet by other wavelet profiles.

---

[2]BLS-GSM Image Denoising Matlab Toolbox 1.0.3 (latest revision: February 23, 2005), available from `http://www.io.csic.es/PagsPers/JPortilla/software/section/3-bayesian-denoising-in-the-wavelet-domain9/`
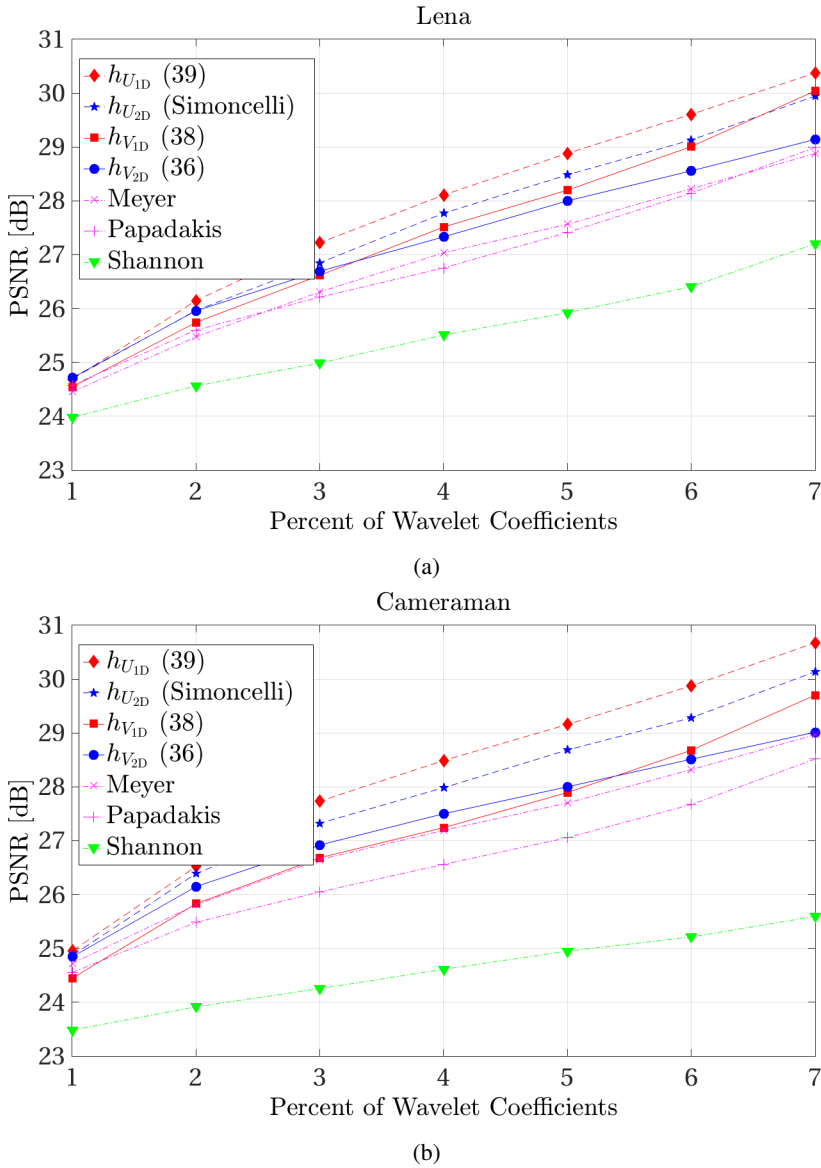
Figure 5.6: Wavelet-based edge reconstruction. Evolution of the PSNR as a function of the percentage of retained wavelet coefficients using different wavelet profiles on the (a) Lena and (b) Cameraman images. The legend is sorted by decreasing order of performance.

Table 5.5: BLS-GSM Denoising at $\sigma = 40$ with 4 Orientations

Wavelet type	PSNR [dB]						
	House	Pirate	Peppers	Lena	Bridge	Cameraman	Einstein
$h_{V_{2D}}$ (5.18)	28.80	26.06	26.60	29.23	23.97	29.66	28.77
$h_{V_{1D}}$ (5.20)	**28.85**	**26.09**	**26.62**	**29.28**	**24.01**	**29.71**	**28.80**
$h_{U_{2D}}$ (5.19) (Simoncelli [94])	28.80	26.03	26.55	29.22	23.96	29.64	28.75
$h_{U_{1D}}$ (5.21)	28.70	25.98	26.38	29.13	23.94	29.60	28.69
Papadakis [102]	28.67	26.01	26.51	29.11	23.94	29.54	28.71
Meyer [33]	28.57	25.93	26.38	29.03	23.89	29.41	28.65
Shannon [89]	27.89	25.71	25.84	28.63	23.78	28.87	28.36

We perform several experiments in order to compare performance of different isotropic wavelet profiles when denoising with the BLS-GSM algorithm. We use the same image set as previously in 5.3.2. We also compare results between the same collection of wavelet profiles, namely Shannon, Meyer, Papadakis, Simoncelli ($h_{U_{2D}}$), $h_{U_{1D}}$, $h_{V_{2D}}$ and $h_{V_{1D}}$. In our first experiment, we corrupt the images with additive Gaussian noise of standard deviation $\sigma = 40$ and use the default number of orientations for the construction of the steerable pyramid, namely 4. We evaluate results in terms of the PSNR of the denoised image, as presented in Table 5.5. From this, one observes that $h_{V_{1D}}$ (5.20) outperforms all other profiles. In order to further investigate this, we perform two additional experiments by varying the parameters of BLS-GSM. We first run the algorithm for each wavelet profile on images corrupted by noise levels with $\sigma \in [20, 70]$. Results are displayed in Figure 5.7. As expected, higher noise levels yield lower PSNR as the image becomes harder to retrieve. We also study the influence on the number of orientations chosen to built the steerable pyramid[3]. Increasing the number of orientations yield better results, but also significantly increases computation time. We show results for each of the studied wavelet profile in Figure 5.8. From these two experiments, we observe that $h_{V_{1D}}$, followed by $h_{V_{2D}}$, yields consistently better results than all the other popular profiles we tested, outperforming state-of-the-art results using the steerable pyramid. Although we only show here results on Lena and Cameraman due to space constraints, the same observation can be made using House, Pirate, Peppers, Bridge and Einstein.

Unlike $h_{V_{2D}}$ and $h_{V_{1D}}$, we observe that $h_{U_{1D}}$ yield less impressive results, and in particular does not compete with the Simoncelli profile ($h_{U_{2D}}$) initially used in BLS-GSM, which gives similar results than $h_{V_{2D}}$. This observation is consistent with the way each of the wavelets are constructed. As explained in Section 5.3.1, the h_U maximally localizes the wavelet coefficients, while the h_V have a profile which is optimally localized in spatial domain. The most desirable feature here, in comparison with local orientation estimation and image reconstruction from edges, is a spatially localized profile for the steerable pyramid. A transformation yielding very localized wavelet coefficients is actually even counter-productive as it concentrates the neighborhood of each coefficient, and thus reduces the amount of information that can be exploited by BLS-GSM. The profiles of choice for this application are hence of the class h_V. The values of V_{1D} that can be found in Figure 5.1 for the different profiles are consistent with these observations: $h_{V_{1D}}$ is optimal, then comes $h_{V_{2D}}$ and $h_{U_{2D}}$ (Simoncelli), which is the third closest to the optimal value. The Papadakis and $h_{U_{1D}}$ follow with V_{1D} values that are about equivalently far from

---

[3]The maximum number of orientations allowed by the Matlab implementation is 16.

the optimum, and the Meyer wavelet finally gets the worse V_{1D} localization value. This trend is conserved in our results, as seen in Figures 5.7 and 5.8. Sorting the profiles by result quality (from best to worse) indeed yields $h_{V_{1D}}$, $h_{V_{2D}}$, $h_{U_{2D}}$ (Simoncelli), Papadakis, $h_{U_{1D}}$ and Meyer.

Note that although the difference in PSNR are marginal, the improvement comes at no cost as the only modification to the algorithm is a change of the radial wavelet profile. We also emphasize that these results do not imply that the proposed design should outperform denoising results based on other non-steerable wavelet frames. Rather, they indicate that the signal-domain localization of the wavelets is beneficial to the BLS-GSM algorithm.

5.4 Summary

In this chapter, we have introduced a method for designing maximally localized isotropic tight-frame wavelets. The key ingredient is a measure of localization that can be optimized in order to derive the corresponding profile. We provide two classes of criterion for measuring localization either in the spatial or in the wavelet domain and express the resulting profiles optimizing each criterion either over the whole space or along each radial direction. We then consider three experimental settings in which we compare results obtained with our wavelets against state-of-the-art. First, we focus on the problem of estimating local orientation of filamentous structures, and then on the task of reconstructing an image from a small subset of edges in the wavelet domain. Both of these experiments rely on steerable filters, either Hessian- or gradient-based. In this setting, the wavelets obtained by optimizing the localization of wavelet coefficients outperforms existing isotropic wavelet profiles. Then, we demonstrate the efficiency of the most localized profiles in spatial domain in a denoising experiment using the popular BLS-GSM algorithm. These different use-cases show that both of our classes of localization criterion are relevant depending on the kind of application being considered, and that the proposed wavelet profiles are interesting candidates for image-processing tasks involving isotropic wavelets. Moreover, our results experimentally confirm the validity of the proposed localization criterion. The good or bad performance of the studied wavelet profiles can indeed be interpreted in the light of these two metrics. This further hints at the fact that our localization criterion are useful in order to study the localization of a given wavelet either in terms of its spatial profile, or of the coefficients it generates. In fact, the two proposed localization criteria can serve as quick estimates to assess the relative performance of any given isotropic tight-frame wavelet profile based on a simple calculation.
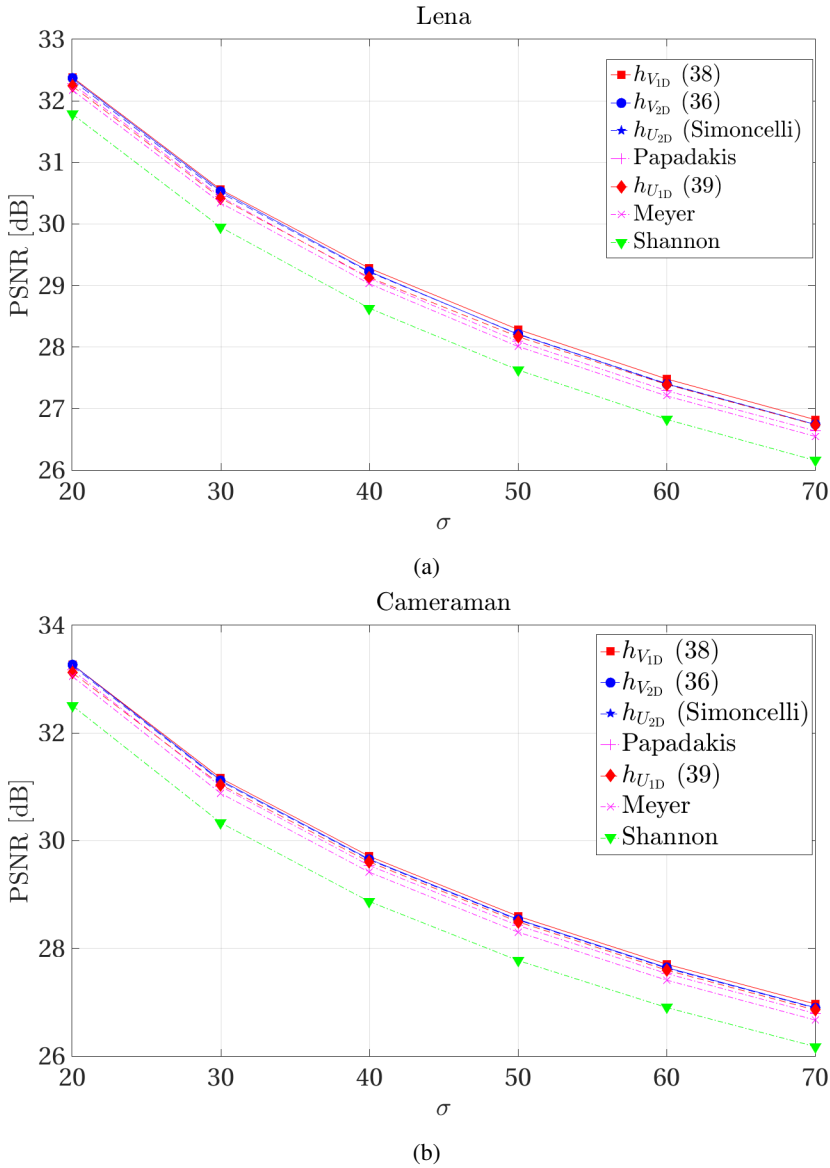
Figure 5.7: BLS-GSM denoising. Evolution of the PSNR as a function of noise using different wavelet profiles on the (a) Lena and (b) Cameraman images. The legend is sorted by decreasing order of performance.
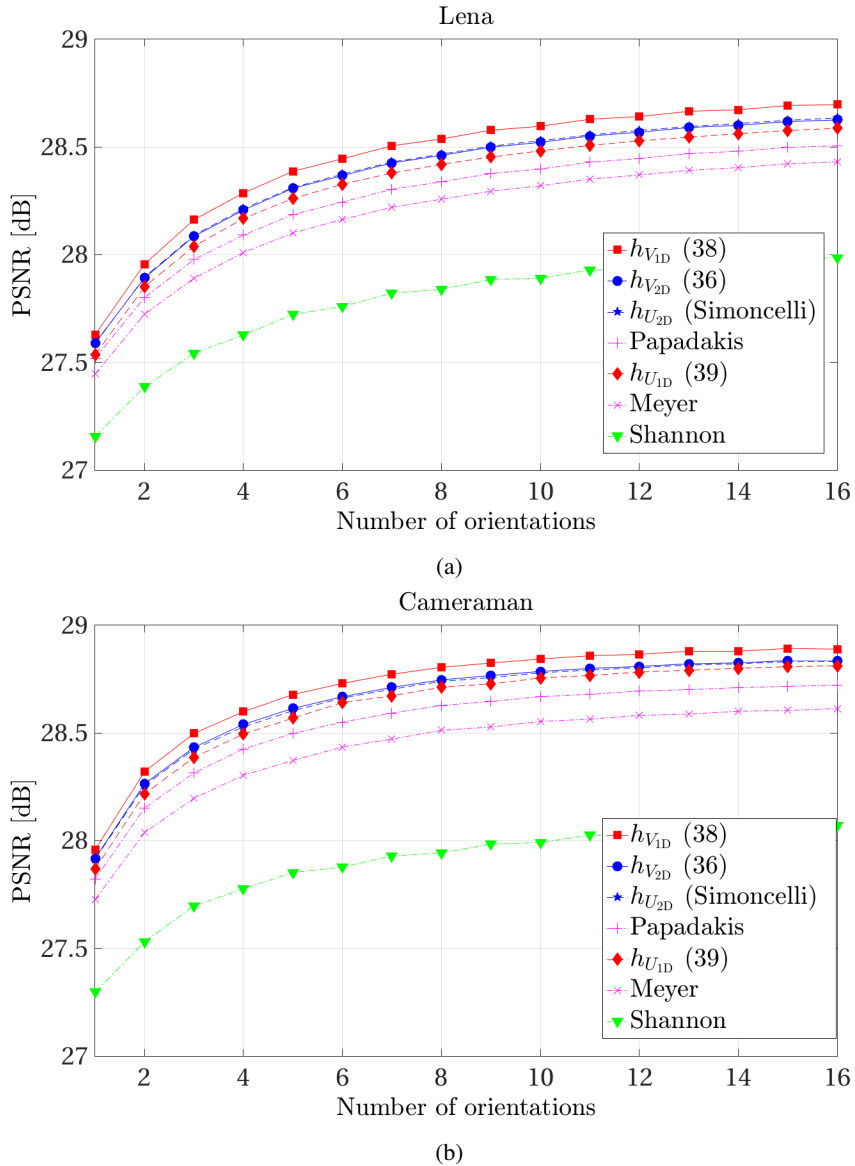
(a)

(b)

Figure 5.8: BLS-GSM denoising. Evolution of the PSNR as a function of the number of angles using different wavelet profiles on the (a) Lena and (b) Cameraman images. The legend is sorted by decreasing order of performance.

5.A Computation of the Gradient

In order to obtain the gradient of the functional $A_m : \mathcal{L}_w \to \mathbb{R}$, we first calculate the Gâteaux derivative of A_m in the direction of a given function $g \in \mathcal{L}_w$ as

$$D_g A_m(h) = \frac{\partial}{\partial \epsilon} A_m(h + \epsilon g)\Big|_{\epsilon=0} \tag{5.35}$$

$$= \int_0^\infty r^m \frac{\partial}{\partial \epsilon} \mathcal{H}\{h + \epsilon g\}(r)^2\Big|_{\epsilon=0} \mathrm{d}r.$$

Due to the linearity of $\mathcal{H}\{\cdot\}$, we can write

$$D_g A_m(h) = \int_0^\infty r^m \frac{\partial}{\partial \epsilon}\left(\mathcal{H}\{h\}(r) + \epsilon \mathcal{H}\{g\}(r)\right)^2\Big|_{\epsilon=0} \mathrm{d}r$$

$$= 2 \int_0^\infty r^m\, \mathcal{H}\{h\}(r)\mathcal{H}\{g\}(r)\mathrm{d}r$$

$$= 2 \int_0^\infty r^m \mathcal{H}\{h\}(r) \left(\int_0^\infty g(t) J_0(tr) t \mathrm{d}t\right) \mathrm{d}r$$

$$= 2 \int_0^\infty g(t) t \left(\int_0^\infty r^m \mathcal{H}\{h\}(r) J_0(tr) \mathrm{d}r\right) \mathrm{d}t. \tag{5.36}$$

According to the definition of the gradient, we have that

$$D_g A_m(h) = \int_0^\infty g(t)\, \nabla A_m\{h\}(t)\, w(t)\mathrm{d}t. \tag{5.37}$$

We therefore obtain

$$\nabla A_m\{h\}(\rho) = \frac{2\rho}{w(\rho)} \int_0^\infty r^m \mathcal{H}\{h\}(r) J_0(\rho r)\mathrm{d}r. \tag{5.38}$$

5.B Characterization of the Projector onto the Space of Tight Frames

In order to characterize the orthogonal projector $\mathcal{P} : \mathcal{L}_w \to \mathcal{T}$, we have to solve a minimization problem. For a given $\tilde{h} \in \mathcal{L}_w$, we are looking for a function $h \in \mathcal{T}$ that satisfies

$$\sum_{i \in \mathbb{Z}} |h(2^i \rho)|^2 = 1 \quad \text{for} \quad \rho \in [\frac{\pi}{2}, \pi] \tag{5.39}$$

and that minimizes the functional

$$\|\tilde{h} - h\|_2^2 = \int_0^\infty (\tilde{h}(\rho) - h(\rho))^2 w(\rho) \mathrm{d}\rho. \tag{5.40}$$

We solve this problem using an infinite-dimensional Lagrange-multiplier method [115]. Similar to (5.35)-(5.38), we calculate the gradient of constraint (5.39) for each value of ρ as

$$\nabla \left\{ \sum_{i \in \mathbb{Z}} |h(2^i \rho)|^2 \right\} (t) = 2 \frac{h(t)}{w(t)} \sum_{i \in \mathbb{Z}} \delta(t - 2^i \rho), \tag{5.41}$$

where δ is the Dirac delta distribution. The gradient of (5.40) is given by

$$\nabla \left\{ \|\tilde{h} - h\|_2^2 \right\} (t) = 2(h(t) - \tilde{h}(t)). \tag{5.42}$$

According to the Lagrange-multiplier method, there exists a function k supported on $\left[\frac{\pi}{2}, \pi\right]$ at the minimizer of (5.40) for which we have that [115]

$$\begin{aligned}
h(t) - \tilde{h}(t) &= \int_0^\pi k(\rho) \frac{h(t)}{w(t)} \sum_{i \in \mathbb{Z}} \delta(t - 2^i \rho) \mathrm{d}\rho \\
&= \frac{h(t)}{w(t)} \sum_{i \in \mathbb{Z}} 2^i k(2^i t).
\end{aligned} \tag{5.43}$$

Therefore,

$$h(t) = \frac{\tilde{h}(t)}{1 - \frac{1}{w(t)} \sum_{i \in \mathbb{Z}} 2^i k(2^i t)}. \tag{5.44}$$

Now, we fix the weight function w as

$$w(t) = 2^i \quad \text{for} \quad \frac{\pi}{2^{i+1}} \le t \le \frac{\pi}{2^i}. \tag{5.45}$$

Equation (5.44) can now be simplified to

$$h(t) = \frac{\tilde{h}(t)}{1 - \sum_{i \in \mathbb{Z}} k(2^i t)}. \tag{5.46}$$

To determine k, we substitute (5.46) in (5.39) and obtain

$$\sum_{j\in\mathbb{Z}} \frac{\tilde{h}^2(2^j t)}{(1 - \sum_{i\in\mathbb{Z}} k(2^{i+j}t))^2} = 1. \qquad (5.47)$$

Since the denominator of (5.46) is invariant to dilations by powers of 2, the denominator of (5.47) does not depend of j. It means that

$$1 - \sum_{i\in\mathbb{Z}} k(2^i t) = \left(\sum_{j\in\mathbb{Z}} \tilde{h}^2(2^j t)\right)^{\frac{1}{2}}. \qquad (5.48)$$

By substituting (5.48) in (5.46), we get that

$$h(t) = \frac{\tilde{h}(t)}{\sqrt{\sum_{i\in\mathbb{Z}} \tilde{h}^2(2^i t)}}. \qquad (5.49)$$

Chapter 6

Summary and Future Studies

In this thesis, we theoretically substantiated the optimality of wavelets for representing sparse signals in a stochastic framework. Also, we obtained a criteria to distinguish between different wavelet families. At the end, we employed our theoretical results to gain practical advantages. In the sequel, we give the contributions in more detail.

6.1   Summary of Results

Optimality of operator-like wavelets for decoupling AR(1) SαS processes: In the first chapter of the main body of the thesis, we studied AR(1) processes that are driven by SαS noises. First, we showed that by sampling a continuous-domain AR(1) process, we obtain a discrete-domain AR(1) process. Then, we focused on discrete-domain SαS AR(1) processes and derived an exact formulation for the mutual information of their coefficient in a transform domain. This formula enabled us to run an optimization algorithm to find the orthonormal basis which maximally decouples such processes. Ultimately, we saw that for small α, the optimal basis is exactly the operator-like wavelets matched to the process under consideration. Additionally, we proposed a criterion to measure the coefficient-wise denoising in a transform domain based on the Stein's formula. Using that, we also saw that the optimal basis for small enough α is the matched operator-like wavelet. Moreover, we proved that for any α less than 2, operator-like wavelets outperform DCT-like bases for both compression and denoising applications. In fact, we showed that for sparse signals,

119

the representation in DCT-like bases have no advantage over the original time domain.

Characterizing the performance of a wavelet basis for denoising self-similar SαS processes: In the second part, we studied another important family of sparse stochastic processes, namely self-similar processes. In contrast with the previous chapter, we studied their wavelet decomposition in the continuous domain rather than sampling them. Our goal was to determine the performance of an arbitrary wavelet basis for the application of denoising. To achieve this goal, we first proved a theorem about wavelet decomposition of locally-L_2 signals (rather than L_2 signals). In the second step, we extended some existing results for the MMSE function of finite variance signals to SαS signals. Combining these two theorems, we could exactly determine the denoising performance of any given orthonormal wavelet basis. Also, we could simplify and obtain a very concise criteria to compare different wavelets. At the end, using calculus of variations, we ran an infinite-dimensional optimization algorithm to find the optimal Meyer wavelet for denoising self-similar SαS processes.

Designing steerable wavelets that outperform the existing ones in image processing applications: In the theoretical chapters while studying the wavelet decomposition of AR(1) and self-similar processes, we observed in both cases that localization of the mother wavelet plays the main role in the performance of the basis. In the last chapter, we exploited this observation to design wavelets for practical image processing applications. First, we used our intuition about images to propose appropriate criteria for different image processing applications. Then, similar to the previous chapter, we proposed an infinite-dimensional optimization algorithm for designing maximally localized steerable tight wavelet frames. These wavelets have a wide usage in different image processing applications. Finally, we could demonstrate the benefits of an optimal wavelet design in several applications.

6.2   Future Studies

Finally, we list some extensions of the work presented in this thesis. First, we may extend the framework of Chapter 3 from AR(1) processes to general discrete sparse stationary processes. Also, we can impose the multi-resolution constraint on the matrix under optimization to directly derive a wavelet-like basis. Another direction altogether would be abandon finding the transformation that maximally decouples the signal. While this decoupling allows us to apply a coefficient-wise manipulation to get a near-optimal coding or denoising performance, there are also other patterns of dependency for which we can

perform the optimal or near-optimal manipulations. For example, for signals with Markov dependency, we can implement the exact MMSE estimator by using graphical models and belief propagation [53, 116]. Based on this fact, we can try to find a representation domain in which the coefficients have maximum Markov dependency, instead of trying to find a representation with the maximally independent coefficients.

We could also investigate the coding of continuous-domain stochastic processes. In that case, we actually need to perform a rate-distortion analysis, because in the continuous domain, a wavelet can have small average mutual-information but at the same time a high approximation error. Therefore, the study should take into account both statistical dependency of the coefficients and approximation power of the wavelets. Deriving criteria for stochastic processes more general than self-similar ones and considering the situations in which partial dependency between the coefficients are taken into account are two other interesting subjects to work on.

Another open topic is to find the optimal frames rather than bases. This problem can be seen as a generalized dictionary learning where we try to design overcomplete dictionaries to represent finite dimensional signals. Finally, a valuable theoretical study is interpreting the already existing results in SCA as the ICA for heavy-tailed, or more specifically, α-stable random variables. This way, we can obtain a statistical interpretation for the existing deterministic methods and then we can use statistical tools to improve them.

Bibliography

[1] R. M. Gray and L. D. Davisson, *An Introduction to Statistical Signal Processing*. Cambridge University Press, 2004, cambridge Books Online. [Online]. Available: http://dx.doi.org/10.1017/CBO9780511801372

[2] J. Pearl, "On coding and filtering stationary signals by discrete Fourier transforms," *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 229–232, March 1973.

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

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

[5] I. Daubechies, "The wavelet transform, time-frequency localization and signal analysis," *IEEE Transactions on Information Theory*, vol. 36, no. 5, pp. 961–1005, 1990.

[6] D. L. Donoho, "Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data," in *Proceedings of symposia in Applied Mathematics*, vol. 47. American Mathematical Society, 1993, pp. 173–205.

[7] M. Antonini, M. Barlaud, P. Mathieu, and I. Daubechies, "Image coding using wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 205–220, April 1992.

[8] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic press, 2008.

[9] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, pp. 607–609, 13 June 1996.

[10] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.

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

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

[13] R. A. Devore, "Nonlinear approximation," *Acta Numerica*, vol. 7, pp. 51–150, January 1998.

[14] J. F. Cardoso and D. L. Donoho, "Some experiments on independent component analysis of non-Gaussian processes," in *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics*, Caesarea, 14-16 June 1999, pp. 74 –77.

[15] P. Abry, P. Gonçalvés, and P. Flandrin, "Wavelets, spectrum analysis and $1/f$ processes," *Wavelets and statistics, Springer*, vol. 15-29, 1995.

[16] M. Unser and P. D. Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge, UK: Cambridge University Press, 2014.

[17] M. Unser, P. D. Tafti, and Q. Sun, "A unified formulation of Gaussian vs. sparse stochastic processes—Part I: Continuous-domain theory," *IEEE Transactions on Information Theory*, vol. 60, no. 3, pp. 1945–1962, March 2014.

[18] M. Unser, P. D. Tafti, A. Amini, and H. Kirshner, "A unified formulation of Gaussian vs. sparse stochastic processes—Part II: Discrete-domain theory," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 3036–3051, May 2014.

[19] I. Gelfand and N. Y. Vilenkin, *Generalized Functions*. New York, USA: Academic Press, 1964, vol. 4.

[20] D. Applebaum, *Lévy processes and stochastic calculus*. Cambridge university press, 2009.

[21] Y. Meyer and D. H. Salinger, *Wavelets and operators*. Cambridge university press, 1995, vol. 1.

[22] M. Unser and P. D. Tafti, *An Introduction to Sparse Stochastic Processes*. Cambridge University Press, 2014, cambridge Books Online. [Online]. Available: http://dx.doi.org/10.1017/CBO9781107415805

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

[24] E. E. Kuruoglu, W. J. Fitzgerald, and P. J. Rayner, "Near optimal detection of signals in impulsive noise modeled with a symmetric/spl alpha/-stable distribution," *IEEE Communications Letters*, vol. 2, no. 10, pp. 282–284, 1998.

[25] D. Middleton, "Non-Gaussian noise models in signal processing for telecommunications: new methods an results for class A and class B noise models," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1129–1149, 1999.

[26] A. Achim and E. E. Kuruoglu, "Image denoising using bivariate α-stable distributions in the complex wavelet domain," *IEEE Signal Processing Letters*, vol. 12, no. 1, pp. 17–20, 2005.

[27] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, July 1989.

[28] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Communications on Pure and Applied Mathematics*, vol. 41, no. 7, pp. 909–996, October 1988.

[29] Y. Meyer, "Principe d'incertitude, bases hilbertiennes et algèbres d'opérateurs," *Séminaire Bourbaki*, vol. 38, no. 662, pp. 209–223, 1985-86.

[30] M. Unser and T. Blu, "Wavelet theory demystified," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 470–483, Feb 2003.

[31] I. Khalidov and M. Unser, "From differential equations to the construction of new wavelet-like bases," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1256–1267, April 2006.

[32] A. Haar, "Zur Theorie der orthogonalen Funktionensysteme," *Mathematische Annalen*, vol. 69, no. 3, pp. 331–371, September 1910.

[33] I. Daubechies, *Ten Lectures on Wavelets*.   SIAM, 1992.

[34] M. D. Buhmann, *Radial Basis Functions*.   Cambridge, United Kingdom: Cambridge University Press, 2003.

[35] M. Papadakis, G. Gogoshin, I. A. Kakadiaris, D. J. Kouri, and D. K. Hoffman, "Nonseparable radial frame multiresolution analysis in multidimensions," *Numerical Functional Analysis and Optimization*, vol. 24, no. 7-8, pp. 907–928, January 2003.

[36] S. Held, M. Storath, P. Massopust, and B. Forster, "Steerable wavelet frames based on the Riesz transform," *IEEE Transactions on Image Processing*, vol. 19, no. 3, pp. 653–667, March 2010.

[37] N. Ahmed, "Discrete cosine transform," *IEEE Transactions on Communications*, vol. 23, no. 1, pp. 90–93, September 1974.

[38] M. Hamidi and J. Pearl, "Comparison of the cosine and Fourier transforms of Markov-1 signals," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 24, no. 5, pp. 428–429, 1976.

[39] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Standards and Practice*.   Kluwer Academic Publishers, 2001.

[40] C. Taswell, "The what, how, and why of wavelet shrinkage denoising," *Computing in Science Engineering*, vol. 2, no. 3, pp. 12 –19, May/June 2000.

[41] Y. Meyer, *Wavelets and Applications*, Masson, France, 1992.

[42] P. Flandrin, "On the spectrum of fractional Brownian motions," *IEEE Transactions on Information Theory*, vol. 35, no. 1, pp. 197–199, January 1989.

[43] S. I. Resnick, "Heavy tail modeling and teletraffic data: special invited paper," *The Annals of Statistics*, vol. 25, no. 5, pp. 1805–1869, 1997.

[44] C. M. Gallagher, "A method for fitting stable autoregressive models using the autocovariation function," *Statistics & probability letters*, vol. 53, no. 4, pp. 381–390, 2001.

[45] S. Ling, "Self-weighted least absolute deviation estimation for infinite variance autoregressive models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 3, pp. 381–393, 2005.

[46] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, November 2012, vol. 2.

[47] J. V. Stone, *Independent Component Analysis*. The MIT Press, September 2004.

[48] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, November 1981.

[49] M. Sahmoudi, K. Abed-Meraim, and M. Benidir, "Blind separation of impulsive alpha-stable sources using minimum dispersion criterion," *IEEE Signal Processing Letters*, vol. 12, no. 4, pp. 281–284, April 2005.

[50] A. R. Soltani and R. Moeanaddin, "On dispersion of stable random vectors and its application in the prediction of multivariate stable processes," *Journal of Applied Probability*, vol. 31, no. 3, pp. 691–699, September 1994.

[51] H. Stark, J. W. Woods, and H. Stark, *Probability and random processes with applications to signal processing*. Upper Saddle River, N.J: Prentice Hall, 2002.

[52] W. D. Ray and R. M. Driver, "Further decomposition of the Karhunen-Loève series representation of stationary random process," *IEEE Transactions on Information Theory*, vol. IT-16, no. 6, pp. 663–668, November 1970.

[53] U. S. Kamilov, P. Pad, A. Amini, and M. Unser, "MMSE estimation of sparse Lévy processes," *IEEE Transactions on Signal Processing*, vol. 61, no. 1, pp. 137–147, January 2013.

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

[55] 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 Transactions on Medical Imaging*, vol. 5, no. 3, pp. 152–161, September 1986.

[56] B. Pesquet-Popescu and J. L. Véhel, "Stochastic fractal models for image processing," *IEEE Signal Processing Magazine*, vol. 19, no. 5, pp. 48–62, September 2002.

[57] K. Park and W. Willinger, *Self-Similar Network Traffic and Performance Evaluation*. Wiley Online Library, 2000.

[58] R. Elliott and J. Van der Hoek, "Fractional Brownian motion and financial modelling," in *Mathematical Finance*. Springer, 2001, pp. 140–151.

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

[60] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes*. Chapman and Hall, 1994.

[61] P. Embrechts and M. Maejima, "An introduction to the theory of self-similar stochastic processes," *International Journal of Modern Physics B*, vol. 14, no. 12n13, pp. 1399–1420, May 2000.

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

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

[64] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[65] P. Flandrin, "Wavelet analysis and synthesis of fractional Brownian motion," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 910–917, March 1992.

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

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

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

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

[70] A. S. Lewis and G. Knowles, "Image compression using the 2-D wavelet transform," *IEEE Transactions on Image Processing*, vol. 1, no. 2, pp. 244–250, April 1992.

[71] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 959–971, October 1996.

[72] R. H. Chan, T. F. Chan, L. Shen, and Z. Shen, "Wavelet algorithms for high-resolution image reconstruction," *SIAM Journal on Scientific Computing*, vol. 24, no. 4, pp. 1408–1432, February 2003.

[73] 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, July 1997.

[74] M. K. Mihcak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 300–303, December 1999.

[75] A. Achim, P. Tsakalides, and A. Bezerianos, "SAR image denoising via Bayesian wavelet shrinkage based on heavy-tailed modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 8, pp. 1773–1784, August 2003.

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

[77] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[78] A. Achim, A. Bezerianos, and P. Tsakalides, "Novel Bayesian multiscale method for speckle removal in medical ultrasound images," *IEEE Transactions on Medical Imaging*, vol. 20, no. 8, pp. 772–783, August 2001.

[79] 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 Recognition Letters*, vol. 27, no. 12, pp. 1370–1382, September 2006.

[80] Y. Wu and S. Verdu, "MMSE dimension," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4857–4879, August 2011.

[81] K. Sato, *Lévy Processes and Infinite Divisibility*.   Cambridge University Press Cambridge, 1999.

[82] A. Ayache, F. Roueff, and Y. Xiao, "Local and asymptotic properties of linear fractional stable sheets," *Comptes Rendus Mathématique*, vol. 344, no. 6, pp. 389–394, March 2007.

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

[84] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261 –1282, April 2005.

[85] Y. Wu and S. Verdu, "Functional properties of minimum mean-square error and mutual information," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1289–1301, March 2012.

[86] M. Unser, "On the optimality of ideal filters for pyramid and wavelet signal approximation," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3591–3596, December 1993.

[87] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, September 1991.

[88] E. P. Simoncelli and W. T. Freeman, "The steerable pyramid: A flexible architecture for multi-scale derivative computation," in *Proceedings of the 1995 IEEE International Conference on Image Processing (ICIP'95)*, vol. 3, Washington D.C., USA, October 23-26, 1995, pp. 3444–3444.

[89] M. Unser, N. Chenouard, and D. Van De Ville, "Steerable pyramids and tight wavelet frames in $L_2(\mathbb{R}^d)$," *IEEE Transactions on Image Processing*, vol. 20, no. 10, pp. 2705–2721, October 2011.

[90] M. Unser and N. Chenouard, "A unifying parametric framework for 2D steerable wavelet transforms," *SIAM Journal on Imaging Sciences*, vol. 6, no. 1, pp. 102–135, January 2013.

[91] S. Fischer, F. Šroubek, L. Perrinet, R. Redondo, and G. Cristóbal, "Self-invertible 2d log-gabor wavelets," *International Journal of Computer Vision*, vol. 75, no. 2, pp. 231–246, November 2007.

[92] A. A. Bharath and J. Ng, "A steerable complex wavelet construction and its application to image denoising," *IEEE Transactions on Image Processing*, vol. 14, no. 7, pp. 948–959, July 2005.

[93] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, November 2003.

[94] J. Portilla and E. P. Simoncelli, "A parametric texture model based on joint statistics of complex wavelet coefficients," *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–70, October 2000.

[95] G. Tzagkarakis, B. Beferull-Lozano, and P. Tsakalides, "Rotation-invariant texture retrieval with Gaussianized steerable pyramids," *IEEE Transactions on Image Processing*, vol. 15, no. 9, pp. 2702–2718, September 2006.

[96] I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, pp. 1413–1457, November 2004.

[97] H. Rabbani, "Image denoising in steerable pyramid domain based on a local laplace prior," *Pattern Recognition*, vol. 42, no. 9, pp. 2181–2193, September 2009.

[98] F. Denis and A. Baskurt, "Multidirectional curvilinear structures detection using steerable pyramid," *Journal of Electronic Imaging*, vol. 13, no. 4, pp. 756–765, October 2004.

[99] Z. Puspoki, V. Uhlmann, C. Vonesch, and M. Unser, "Design of steerable wavelets to detect multifold junctions," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 643–657, Feb 2016.

[100] D. Van De Ville and M. Unser, "Complex wavelet bases, steerability, and the Marr-like pyramid," *IEEE Transactions on Image Processing*, vol. 17, no. 11, pp. 2063–2080, November 2008.

[101] N. Chenouard and M. Unser, "3D steerable wavelets in practice," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4522–4533, November 2012.

[102] J. R. Romero, S. K. Alexander, S. Baid, S. Jain, and M. Papadakis, "The geometry and the analytic properties of isotropic multiresolution analysis," *Advances in Computational Mathematics*, vol. 31, no. 1-3, pp. 283–328, October 2009.

[103] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable multiscale transforms," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, March 1992.

[104] P. Pad, V. Uhlmann, and M. Unser, "VOW: Variance-optimal wavelets for the steerable pyramid," in *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP'14)*, Paris, France, October 27-30, 2014, pp. 2973–2977.

[105] H. Sagan, *Introduction to the Calculus of Variations*.   New York, NY, USA: Dover Publications, 2012.

[106] Y. I. Alber, A. N. Iusem, and M. V. Solodov, "On the projected subgradient method for nonsmooth convex optimization in a hilbert space," *Mathematical Programming*, vol. 81, no. 1, pp. 23–35, March 1998.

[107] G. B. Folland and A. Sitaram, "The uncertainty principle: A mathematical survey," *Journal of Fourier Analysis and Applications*, no. 3, pp. 207–238, May 1997.

[108] J. P. Ward, P. Pad, and M. Unser, "Optimal isotropic wavelets for localized tight frame representations," *IEEE Signal Processing Letters*, vol. 22, no. 11, pp. 1918–1921, November 2015.

[109] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, United Kingdom: Cambridge University Press, 2004.

[110] M. Jacob and M. Unser, "Design of steerable filters for feature detection using Canny-like criteria," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1007–1019, August 2004.

[111] M. D. Abràmoff, P. J. Magalhães, and S. J. Ram, "Image processing with ImageJ," *Biophotonics International*, vol. 11, no. 7, pp. 36–42, July 2004.

[112] R. Rezakhaniha, A. Agianniotis, J. Schrauwen, A. Griffa, D. Sage, C. Bouten, F. van de Vosse, M. Unser, and N. Stergiopulos, "Experimental investigation of collagen waviness and orientation in the arterial adventitia using confocal laser scanning microscopy," *Biomechanics and Modeling in Mechanobiology*, vol. 11, no. 3-4, pp. 461–473, March 2012.

[113] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, January 2011.

[114] B. Rajaei, "An analysis and improvement of the bls-gsm denoising method," *Image Processing on Line*, vol. 4, pp. 44–70, April 2014.

[115] D. G. Luenberger, *Optimization by Vector Space Methods*. New York, NY, USA: John Wiley & Sons, 1969.

[116] P. Tohidi, E. Bostan, P. Pad, and M. Unser, "MMSE denoising of sparse and non-gaussian AR(1) processes," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 4333–4337.

Curriculum Vitæ

PEDRAM PAD

EPFL STI LIB, Station 17 ⋄ 1015 Lausanne, Switzerland

(+41) · 78 · 8085979 ⋄ pedram.pad@epfl.ch

EDUCATION

- **École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland**   *2011 – Present*
  PhD student at Biomedical Imaging Group
  Supervisor: Michael Unser

- **Sharif University of Technology, Tehran, Iran**   *2009 – 2011*
  MSc in Communications Systems at Multimedia Laboratory
  Supervisor: Farokh Marvasti

- **Sharif University of Technology, Tehran, Iran**   *2004 – 2009*
  Two Degrees: BSc in Electrical Engineering
  BSc in Mathematical Sciences

HONORS AND AWARDS

- **Awarded $50k Research Grant for Research in Machine Learning**   *2016*
  Hasler Foundation, Switzerland

- **Best Student Paper Award**   *2014*
  IEEE International Conference on Image Processing (ICIP)

- **Awarded Fellowship for Graduate Studies**   *2011*
  Alcatel-Lucent, Bell-Labs Mobile Network Center, Paris, France

- **Best Bachelor Thesis Award**   *2009*
  IEEE Iran Section

- **Exceptional Talents Admission for BSc and MSc Program**   *2004 & 2009*
  Offered Direct Entrance for Sharif Univ. of Tech. (Bypassing the Entrance Exam)

- **Member of Iran National Elite Foundation**   *2005*

- **Gold Medalist of National Math Olympiad**   *2003*

- **Exceptional Talents Admission for High School**   *2000*
  National Organization of Development of Exceptional Talents, Esfahan, Iran

PUBLICATIONS

Journal Papers

▶ **P. Pad**, K. Alishahi, and M. Unser, "Optimized Wavelet Denoising for Self-Similar α-Stable Processes," accepted in IEEE Transactions on Information Theory, 2017.

▶ **P. Pad**, V. Uhlmann, and M. Unser, "Maximally Localized Radial Profiles for Tight Steerable Wavelet Frames," IEEE Transactions on Image Processing, vol. 25, no. 5, pp. 2275–2287, 2016.

▶ J. P. Ward, **P. Pad**, and M. Unser, "Optimal isotropic wavelets for localized tight frame representations," IEEE Signal Processing Letters, vol. 22, no. 11, pp. 1918-1921, 2015.

▶ **P. Pad** and M. Unser, "Optimality of operator-like wavelets for representing sparse AR(1) processes," IEEE Transactions on Signal Processing, vol. 63, no. 18, pp. 4827–4837, 2015.

▶ U.S. Kamilov, **P. Pad**, A. Amini, and M. Unser, "MMSE estimation of sparse Lévy processes," IEEE Transactions on Signal Processing, vol. 61, no. 1, pp. 137–147, 2013.

▶ K. Alishahi, S. Dashmiz, **P. Pad** and F. Marvasti, "Design of signature sequences for overloaded CDMA and bounds on the sum capacity with arbitrary symbol alphabets," IEEE Transactions on Information Theory, vol. 58, no. 3, pp. 1441–1469, 2012.

- ► A. H. Hosseini, O. Javidbakht, **P. Pad**, and F. Marvasti, "A review on synchronous CDMA systems: optimum overloaded codes, channel capacity, and power control," EURASIP Journal on Wireless Communications and Networking, no. 1, pp. 1–22, 2011.
- ► **P. Pad**, A. Mousavi, A. Goli, and F. Marvasti, "Simplified MAP-MUD for Active User CDMA," IEEE Communications Letters, vol. 15, no. 6, pp. 599–601, 2011.
- ► **P. Pad**, M. J. Faraji, and F. Marvasti, "Constructing and Decoding GWBE Codes Using Kronecker Products," IEEE Communications Letters, vol. 14, no. 1, pp. 1–3, Jan. 2010.
- ► K. Alishahi, F. Marvasti, V. Aref, and **P. Pad**, "Bounds on the Sum Capacity of Synchronous Binary CDMA Channels," IEEE Transactions on Information Theory, vol. 55, no. 8, pp. 3577–3593, 2009.
- ► **P. Pad**, F. Marvasti, K. Alishahi, and S. Akbari, "A class of errorless codes for over-loaded synchronous wireless and optical CDMA systems," IEEE Transactions on Information Theory, vol. 55, no. 6, pp. 2705–2715, 2009.

Conference Papers

- ► P. Tohidi, E. Bostan, **P. Pad**, and M. Unser, "MMSE Denoising of sparse and non-Gaussian AR(1) Processes," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2016, Shanghai, China.
- ► A. Depeursinge, **P. Pad**, A. S. Chin, A. N. Leung, D. L. Rubin, H. Muller, and M. Unser, "Optimized steerable wavelets for texture analysis of lung tissue in 3-D CT: Classification of usual interstitial pneumonia," IEEE 12th International Symposium on Biomedical Imaging (ISBI) 2015, Brooklyn, NY, USA.
- ► **P. Pad**, V. Uhlmann, and M. Unser, "VOW: Variance-optimal wavelets for the steerable pyramid," IEEE International Conference on Image Processing (ICIP) 2014, Paris, France.
- ► **P. Pad** and M. Unser, "On the optimality of operator-like wavelets for sparse AR(1) processes," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2013, Vancouver, BC, Canada.
- ► A. Makhdoumi-Kakhaki, H. Karkeh-Abadi, **P. Pad**, H. Saeedi, F. Marvasti, and K. Alishahi, "Capacity achieving linear codes with random binary sparse generating matrices over the Binary Symmetric Channel," IEEE International Symposium on Information Theory (ISIT) 2012, Cambridge, MA, USA.
- ► M. Heidari-Khoozani, A. Rashidinejad, M. H. Lotfi-Froushani, **P. Pad**, and F. Marvasti, "Almost-Optimum Signature Matrices in Binary-Input Synchronous Overloaded CDMA," IEEE International Conference on Telecommunications (ICT) 2011, Ayia Napa, Cyprus.
- ► A. Mousavi, **P. Pad**, P. Delgosha, and F. Marvasti, "A New Decoding Scheme for Errorless Codes for Overloaded CDMA with Active User Detection," IEEE International Conference on Telecommunications (ICT) 2011, Ayia Napa, Cyprus.
- ► M. Kafashan, H. Hosseini, S. Beygiharchegani, **P. Pad** and F. Marvasti, "New rectangular partitioning methods for lossless binary image compression," IEEE International Conference on Signal Processing (ICSP) 2010, Beijing, China.
- ► S. Dashmiz, M. R. Takapoui, **P. Pad**, and F. Marvasti, "New Bounds for the Sum Capacity of Binary and Nonbinary Synchronous CDMA Systems," IEEE International Symposium on Information Theory (ISIT) 2010, Austin, TX, USA.
- ► M. H. Shafinia, P. Kabir, **P. Pad**, S. M. Mansouri, and F. Marvasti, "Errorless Codes For CDMA Systems with Near-Far Effect," IEEE International Conference on Communications (ICC) 2010, Cape Town, South Africa.
- ► **P. Pad**, M. Soltanolkotabi, S. Hadikhanlou, A. Enayati, and F. Marvasti, "Errorless Codes for Over-loaded CDMA with Active User Detection," IEEE International Conference on Communications (ICC) 2009, Dresden, Germany.
- ► **P. Pad**, F. Marvasti, K. Alishahi, and S. Akbari, "Errorless Codes for Over-loaded CDMA Systems and Evaluation of Channel Capacity Bounds," International Symposium on Information Theory (ISIT) 2008, Toronto, Canada.
- ► A. Javanmard, **P. Pad**, M. Babaie-Zadeh, and C. Jutten, "Estimating the Mixing Matrix in Underdetermined Sparse Component Analysis (SCA) Using Consecutive independent Component Analysis," European Signal Processing Conference (EUSIPCO) 2008, Lausanne, Switzerland.
- ► S. M. R. Farahi, F. Khademolhosseini, and **P. Pad**, "A System Dynamical Approach to Circumstance of Infectious Diseases Outbreak in Society," The 26th International Conference of the System Dynamics Society, Greece, 2008.

WORK AND TEACHING EXPERIENCES

Teacher Assistant of Prof. Unser in Image Processing Course *2013 – 2015*
Image Processing Engineer at Dideh-Pardaz-Saba Company *2011*
Developed a real-time automatic surface inspection system for steel sheets by combining several stages of image processing and machine learning for feature extraction and classification
Mathematics Tutor of Iran Mathematics Olympiad medalists *2004 – 2011*

RESEARCH INTERESTS

Machine Learning, Image Processing, Statistical Signal Processing, Information Theory

COMPUTER AND LANGUAGE SKILLS

Expert	C++, Java, Matlab	English	*Fluent*
Familiar	Phython	French	*Basic*
		Persian	*Mother Tongue*

EXTRACURRICULAR

Playing Frisbee
Member of Organiziation of Iranian Students (Organizing Social Events for 400+ People)