Studies in Stochastic Processes: Adaptive Wavelet Decompositions and Operator Fractional Brownian Motions

Gustavo de Vasconcellos Didier

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research (Statistics).

Chapel Hill
2007

Approved by

Advisor: Vladas Pipiras

Reader: M. Ross Leadbetter

Reader: Chuanshu Ji

Reader: Eric Renault

Reader: Murad S. Taqqu

ABSTRACT

GUSTAVO DE VASCONCELLOS DIDIER: Studies in Stochastic Processes: Adaptive Wavelet Decompositions and Operator Fractional Brownian Motions
(Under the direction of Vladas Pipiras)

The thesis is centered around the themes of wavelet methods for stochastic processes, and of operator self-similarity. It comprises three parts. The first two parts concern particular wavelet-based decompositions of stationary processes, in either continuous or discrete time. The decompositions are essentially characterized by uncorrelated detail coefficients and possibly correlated approximation coefficients. This is of interest, for example, in simulation and maximum likelihood estimation. In discrete time, the focus is somewhat on long memory time series. The last part of the thesis concerns operator fractional Brownian motions. These are Gaussian operator self-similar processes with stationary increments, and are multivariate analogues of the one-dimensional fractional Brownian motion. We establish integral representations of operator fractional Brownian motions, study their basic properties and examine questions of uniqueness.

ACKNOWLEDGEMENTS

My command of words falls short of properly expressing my gratitude to my advisor, Professor Vladas Pipiras, for directing my dissertation work and being my mentor in the Ph.D. program. It has certainly been a privilege to closely work with him, and benefit from his incessant inspiration, intellectual depth, personal enthusiasm, kindness and dedication. I can only be deeply grateful for this invaluable experience that will surely exert a lifelong influence on my career.

I am sincerely thankful to the other members of my committee, Professors M. Ross Leadbetter, Chuanshu Ji, Eric Renault and Murad Taqqu for patiently reading this dissertation and providing new insights into the material herein. I am also thankful to all the faculty and staff for a great experience in the department. Special thanks go to Professors Budhiraja and Carlstein for their guidance and help during the Ph.D. program and with the job market.

I am grateful to Dr. Robert Rodriguez for the opportunity of being a research assistant at SAS.

I am grateful to Professors Aloisio Araujo, Maria E. Vares, Jorge Chami and Getulio B. da Silveira for their help and encouragement with the idea of obtaining the Ph.D. degree abroad.

It would be impossible to list here the names of all the relatives and friends in the US and Brazil whose support have enabled me to complete my Ph.D. studies. I am fortunate to have a loving family and unique friends. Of course, I am especially indebted to my parents, Penha and Aluisio, for their self-sacrifice, love and enthusiasm for my project to pursue my doctoral studies abroad. I am also deeply indebted to my dear siblings Suzana, Carlos and Aluisio, and also to my wonderful aunts, uncles, in-laws, nieces and nephew, cousins, and to Rosa, all of them for their affection and unconditional support in the good and bad moments. Even though physically distant, they have always been present during the journey. In Chapel Hill, I am very grateful to my dear friends Brian, Chamu, John, Fernando, Myung Hee, Ping, Xuxin, Lingsong, Levi, Melissa and so many others for making my personal and academic experience in North Carolina unique. In Brazil, I am sincerely grateful to José, Marcos, João Pedro, Henrique, João Pedro Azevedo, Mariana, Juan Pablo, Ivan, Pablo, Leo, Pedro, Helio, Eduardo, Alex, Cacá (to name a few!) for their friendship and encouragement. Special thanks go to Alexandre and Flavio also for their friendship and their ever-inspiring presence in the last five years.

The devil is in the details.

Proverb

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

The thesis is centered around the themes of wavelet methods for stochastic processes (Chapters 2 and 3), and of operator self-similarity (Chapter 4). The wavelet analysis of a random process involves expressing it in terms of a wavelet basis. Orthogonal wavelet bases usually provide expansions with "almost" uncorrelated coefficients. Several other non-orthogonal wavelet bases were constructed leading to exactly uncorrelated coefficients, for example, for fractional Brownian motion. Contributing to this body of work, we introduce here novel wavelet-based decompositions for stationary processes, in either continuous or discrete time. Called Adaptive Wavelet Decompositions (AWD), their detail coefficients are also uncorrelated but approximation coefficients are possibly correlated. Chapter 2 concerns AWD in continuous time. Approximation coefficients in these AWD have to be taken correlated as only such will approximate a stationary process at hand, an important property known as the "wavelet crime". In Chapter 3, we extend AWD to discrete time processes. Correlated approximation coefficients in these decompositions allow to have shorter filters in the associated Fast Wavelet Transform-like algorithm. This is particularly relevant when dealing with long memory and near unit root time series. In either continuous or discrete time, because of uncorrelated detail coefficients, AWD can be used in simulation of Gaussian stationary processes, and in maximum likelihood estimation.

In discrete time, AWD are especially suitable when dealing with long range dependent time series. When exploring multivariate analogues of AWD, we found that, surprisingly, multivariate long range dependence and related multivariate fractional Brownian motions have been little explored. Fractional Brownian motion (FBM) is a generalization of Brownian motion to the case where the increments are correlated. It is closely related to long range dependence, since its increments are often used as discrete-time models for long range dependent data. FBM is characterized by three properties: it is a Gaussian process; its increments are stationary; and its distribution scales across time according to a (fractional) parameter, a property called self-similarity. The appropriate multivariate version of FBM is the so-called operator fractional Brownian motion (OFBM). This process is also Gaussian, has stationary increments, and its distribution scales across time according to a matrix, a property appropriately called operator self-similarity. Chapter 4 is a more systematic study of OFBMs. We establish spectral and time-domain representations of OFBMs, and look into the relation between the (operator) self-similarity parameter and the characterization of the law of the process. With a view toward the analysis of multivariate long range dependent time series, we also study the cross spectrum of OFBMs. Finally, we analyze questions of uniqueness of the representation of OFBMs, and explore the symmetry structure of bivariate operator self-similar Gaussian processes.

CHAPTER 2

Gaussian stationary processes: adaptive wavelet decompositions, discrete approximations and their convergence

2.1 Introduction

Consider a real-valued Gaussian stationary process $X = \{X(t)\}_{t\in\mathbb{R}}$ having the integral representation

$$X(t) = \int_{\mathbb{R}} g(t-u)dB(u) = \int_{\mathbb{R}} e^{itx}\widehat{g}(x)d\widehat{B}(x), \tag{2.1}$$

where $g \in L^2(\mathbb{R})$ is a real-valued function, called a *kernel function*, $\widehat{g} \in L^2(\mathbb{R})$ is its Fourier transform defined by convention as

$$\widehat{g}(x) = \int_{\mathbb{R}} e^{-ixu}g(u)du,$$

$\{B(u)\}_{u\in\mathbb{R}}$ is a standard Brownian motion and $\{\widehat{B}(x)\}_{x\in\mathbb{R}} = \{B_1(x) + iB_2(x)\}_{x\in\mathbb{R}}$ is a complex-valued Brownian motion satisfying $B_1(x) = B_1(-x)$, $B_2(x) = -B_2(-x)$, $x \geq 0$, with two independent Brownian motions $\{B_1(x)\}_{x\geq 0}$ and $\{B_2(x)\}_{x\geq 0}$ such that $EB_1(1)^2 = EB_2(1)^2 = (4\pi)^{-1}$. (The latter conditions on $\widehat{B}(x)$ ensure that the second integral in (2.1) is real-valued and has the same covariance structure as the first integral in (2.1).) Many Gaussian stationary processes, especially those of practical interest, can be represented by (2.1). See, for example, Rozanov (1967) and others. The covariance function $R(t) = EX(t)X(0)$ of X and its Fourier transform are given by

$$R(t) = (g * g^{\vee})(t) = \frac{1}{2\pi}\widehat{|\widehat{g}|^2}(-t), \quad \widehat{R}(x) = |\widehat{g}(x)|^2, \tag{2.2}$$

where $g^{\vee}(u) = g(-u)$ is the time reversion operation and $*$ stands for convolution. The Fourier transform $\widehat{R}(x)$ is also known as the spectral density of X. Note, however, that the two rightmost expressions in (2.2) are not meaningful for general $g \in L^2(\mathbb{R})$ because the function R may be neither in $L^2(\mathbb{R})$ nor $L^1(\mathbb{R})$.

Under mild assumptions on g and in a special Gaussian case, Theorem 1 of Zhang and Walter (1994) states that a Gaussian process X in (2.1) has a wavelet-based expansion

$$X(t) = \sum_{n=-\infty}^{\infty} a_{J,n}\theta^J(t - 2^{-J}n) + \sum_{j=J}^{\infty}\sum_{n=-\infty}^{\infty} d_{j,n}\Psi^j(t - 2^{-j}n), \tag{2.3}$$

for any $J \in \mathbb{Z}$, with convergence in the $L^2(\Omega)$-sense for each t. Here, $a_J = \{a_{J,n}\}_{n\in\mathbb{Z}}$, $d_j = \{d_{j,n}\}_{j\geq J, n\in\mathbb{Z}}$ are independent $\mathcal{N}(0,1)$ random variables. The functions θ^j and Ψ^j are defined through their Fourier transforms as

$$\widehat{\theta^j}(x) = \widehat{g}(x)\, 2^{-j/2}\, \widehat{\phi}(2^{-j}x), \quad \widehat{\Psi^j}(x) = \widehat{g}(x)\, 2^{-j/2}\, \widehat{\psi}(2^{-j}x), \tag{2.4}$$

where ϕ and ψ are scaling and wavelet functions, respectively, associated with a suitable orthogonal Multiresolution Analysis (MRA, in short). For more information on scaling function, wavelet and MRA, see for example Mallat (1998), Daubechies (1992), or many others. Moreover, the coefficients $a_{j,n}$ and $d_{j,n}$ in (2.3) can be expressed as

$$a_{j,n} = \int_{\mathbb{R}} X(t)\theta_j(t - 2^{-j}n)dt, \quad d_{j,n} = \int_{\mathbb{R}} X(t)\Psi_j(t - 2^{-j}n)dt, \tag{2.5}$$

with the functions θ_j and Ψ_j, "dual" to θ^j and Ψ^j, defined through

$$\widehat{\theta}_j(x) = \overline{\widehat{g}(x)^{-1}}\, 2^{-j/2}\, \widehat{\phi}(2^{-j}x), \quad \widehat{\Psi}_j(x) = \overline{\widehat{g}(x)^{-1}}\, 2^{-j/2}\, \widehat{\psi}(2^{-j}x). \tag{2.6}$$

Zhang and Walter (1994) call (2.3) a Karhunen-Loève-like (KL-like) wavelet-based expansion. It is discussed in several textbooks, for example, Walter and Shen (2001), and Vidakovic (1999). The sum $\sum_n a_{J,n}\theta^J(t - 2^{-J}n)$ in (2.3) is interpreted as an approximation term at scale 2^{-J}, and the sums $\sum_n d_{j,n}\Psi^j(t - 2^{-j}n)$, $j \geq J$, are interpreted as detail terms at finer scales 2^{-j}, $j \geq J$. The KL-like expansion is related to the wavelet-vaguelette expansions of Donoho (1995), the expansions of Benassi and Jaffard (1994), and others, where $J = -\infty$ in (2.3) and hence the first approximation term in (2.3) is absent.

Though the approximation term $\sum_n a_{J,n}\theta^J(t - 2^{-J}n)$ in (2.3) involves independent $\mathcal{N}(0,1)$ random variables $a_{J,n}$ which are convenient to deal with in theory, the term is also unnatural in one important respect. It is customary with wavelet bases that not only an approximation term but also the respective approximation coefficients, the sequence $a_{J,n}$ in this case, approximate the signal at hand. The sequence $a_{J,n}$ does not have this property because it consists of independent random variables and hence cannot approximate a typically dependent stationary process $X(t)$. In this work, we modify the approximation terms as

$$\sum_{n=-\infty}^{\infty} a_{J,n}\theta^J(t - 2^{-J}n) = \sum_{n=-\infty}^{\infty} X_{J,n}\Phi^J(t - 2^{-J}n) \tag{2.7}$$

so that the new approximation coefficients $X_J = \{X_{J,n}\}_{n\in\mathbb{Z}}$ now have this property, namely,

$$2^{J/2}X_{J,[2^J t]} \approx X(t) \tag{2.8}$$

in a suitable sense, as $J \to \infty$, where $[x]$ denotes the integer part of $x \in \mathbb{R}$.

In the relation (2.7) above,

$$\widehat{\Phi}^J(x) = \frac{\widehat{g}(x)}{\widehat{g}_J(2^{-J}x)}2^{-J/2}\widehat{\phi}(2^{-J}x) = \frac{\widehat{\theta}^J(x)}{\widehat{g}_J(2^{-J}x)} \tag{2.9}$$

with the discrete Fourier transform $\widehat{g}_J(y)$ of a sequence $g_J = \{g_{J,n}\}$. (A discrete Fourier transform of $g = \{g_n\}$ is defined by

$$\widehat{g}(x) = \sum_{n=-\infty}^{\infty} g_n e^{-inx}, \quad x \in \mathbb{R},$$

and is periodic with the period 2π.) The random sequence $X_J = \{X_{J,n}\}$ in (2.7) is defined as

$$\widehat{X}_J(x) = \widehat{g}_J(x)\widehat{a}_J(x) \tag{2.10}$$

in the frequency domain. Moreover, we expect that

$$X_{J,n} = \int_{\mathbb{R}} X(t)\Phi_J(t - 2^{-J}n)dt, \qquad (2.11)$$

where

$$\widehat{\Phi}_J(x) = \overline{\left(\frac{\widehat{g}_J(2^{-J}x)}{\widehat{g}(x)}\right)} 2^{-J/2}\widehat{\phi}(2^{-J}x). \qquad (2.12)$$

The relation (2.7) can be informally and easily verified by taking Fourier transforms on both sides of the expression.

It is well-known (e.g. Daubechies (1992) in the deterministic context) that (2.8) is a property of the corresponding wavelet basis functions. When

$$G_J(2^{-J}x) := \frac{\widehat{g}_J(2^{-J}x)}{\widehat{g}(x)} \approx 1, \qquad (2.13)$$

we have $\widehat{\Phi}_J(x) \approx 2^{-J/2}\widehat{\phi}(2^{-J}x) \approx 2^{-J/2}$ for large J (typically, $\widehat{\phi}(0) = \int_{\mathbb{R}} \phi(t)dt = 1$) and hence, by (2.11), we expect that

$$\begin{aligned} 2^{J/2}X_{J,n} &= \frac{2^{J/2}}{2\pi}\int_{\mathbb{R}} \widehat{X}(x)e^{-ix2^{-J}n}\overline{\widehat{\Phi}_J(x)}dx \\ &\approx \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{X}(x)e^{-ix2^{-J}n}dx = X(2^{-J}n). \end{aligned}$$

The conditions for (2.7) and (2.8) will thus involve the function G_J given in (2.13).

Though the modification (2.7) appears small, it is fundamental and important in several ways, and surprisingly leads to many research questions. First, the convergence allows for several applications, for example, simulation and maximum likelihood estimation, at the reconstruction and decomposition use of (2.3) with (2.7), respectively. In this chapter, we study only the issue of simulation. We show in Section 2.6 that there is a nonstandard Fast Wavelet Transform algorithm relating the sequences $X_j = \{X_{j,n}\}$ across scales. It is nonstandard in the sense that the low- and high-pass filters entering the algorithm depend on the scale parameter j. The algorithm is convenient in simulation since independent, $\mathcal{N}(0,1)$ random variables $d_{j,n}$ (the detail coefficients in (2.3)) need to be generated to produce an approximation at finer scale. Convergence of X_j to the process X is ensured by the property (2.8). In fact, as shown in Section 2.7, this convergence is exponentially fast in j and almost sure uniformly on compact intervals. Dependence of the convergence speed and the type of approximation involving X_j on the smoothness of X is also studied in Section 2.7, and turns out to be quite complex.

Maximum likelihood estimation not considered here, refers to the following. The property (2.8) is known in the wavelet literature as "wavelet crime". It is used, in practice, to replace the approximation coefficients $X_{j,n}$ at finest scale by the normalized observations $2^{-j/2}X(2^{-j}n)$. Assuming a model for X and hence for X_j, as in maximum likelihood estimation, the approximation sequence X_j can be transformed (by the corresponding wavelet transformation) into independent, $\mathcal{N}(0,1)$ detail coefficients at coarser scales and approximation coefficients at coarsest scale. This can be viewed as a factorization of the covariance matrix of X_j and could be used in maximum likelihood estimation. For more details, see Chapter 3, where we study analogous wavelet decompositions in discrete time. Let us also note that none of these applications are possible having the decomposition (2.3) alone.

Second, the wavelet-based decomposition (2.3) with (2.7) can be viewed as a generalization to stationary Gaussian processes of a particular wavelet decomposition of fractional Brownian motion established in Sellan (1995), Meyer *et al.* (1999). This extension is significant for several reasons. Self-similarity (of fractional Brownian motion, for instance) and wavelets have long been considered closely related, with the articles above being one example. Our extension shows that self-similarity (though an important special case) is not necessary to make some of these connections. Also, we work in the general framework of Gaussian stationary processes. We formulate conditions on sequences X_j to have (2.7) and (2.8) in general. This is quite nontrivial by itself. In particular, we want our conditions to include some natural discrete time approximations X_j to continuous time processes X such as AR(1) time series approximations X_j to the Ornstein-Uhlenbeck process X. Most of the conditions used in this chapter are stated in Section 2.3 and several examples are considered in Section 2.4.

Third, more generally, the decompositions (2.3) with (2.7) are examples of decompositions of stationary Gaussian processes with independent coefficients such as the usual Fourier representation or the Karhunen-Loève expansion. These decompositions have a convenient multiresolution structure where a process is viewed as an approximation term superimposed by finer and finer details, and are characterized by other nice properties such as (2.8). Such wavelet decompositions (apart from (2.3)) have largely been missing in the literature at a fundamental level. We hope that our work will help filling in the current gap. Comparing (2.3) to (2.3) with (2.7), we have already noted that none of the above applications are possible having (2.3) alone. In defense of (2.3), these decompositions are in the spirit of decompositions of discrete time signals used in Signal Processing, where signals are decomposed into subbands with uncorrelated coefficients. For example, this is a necessary condition to achieve a suitable optimality in coding. But because the decompositions (2.3) lack the "wavelet crime" property (2.8), they are not that useful in practice.

Fourth, we study whether the wavelet bases in (2.3) and (2.3) with (2.7) are Riesz, which are the bases preferred in the nonorthogonal context. We show in Section 2.9 that both bases, in fact, are Riesz under additional assumptions. This provides a partial answer to the above question which was asked but kept open since Zhang and Walter (1994). Though the results on Riesz bases may appear to bear little relation to Probability, we see them as key if one has to manipulate with the decompositions (2.3) and (2.3) with (2.7).

Fifth and last, this work raises many more questions. As mentioned above, in Chapter 3 we study analogous wavelet decompositions in discrete time. Pipiras (2004) explored a similar decomposition for a non-Gaussian self-similar process called the Rosenblatt process. We also plan to consider multidimensional $X(t)$, with either $t \in \mathbb{R}^m$ (or $\mathbb{Z}^m$) or $X(t) \in \mathbb{R}^n$.

The decomposition (2.3) with (2.7) can be viewed as being more general than (2.3) – becoming (2.3) when $X_j = a_j$ are independent, $\mathcal{N}(0,1)$ random variables. For this reason, both decompositions should be viewed under one framework. This is the view taken in the following definition and in Chapter 3.

Definition 2.1.1. Decompositions (2.3) and (2.3) with (2.7) will be called *Adaptive Wavelet Decompositions.*

Adaptiveness refers to the fact that the basis functions are chosen based on the dependence structure of the underlying stationary Gaussian process.

The rest of the chapter is organized as follows. In Section 2.2, we briefly introduce a wavelet basis to be used in wavelet-based decompositions. In Section 2.3, we state the assumptions on the discrete deterministic approximations g_J and the functions g. In Section 2.4, we consider several examples of Gaussian stationary processes and their discrete approximations. The KL-like wavelet decomposition (2.3) and its modification (2.7) are proved in Section 2.5. In particular, we reprove the decomposition (2.3) because inaccurate assumptions were used in Zhang and Walter (1994). We show that there is a FWT-like algorithm relating $\{X_{j,n}\}$ across different scales in Section 2.6. In Sections 2.7 and 2.8, we examine convergence of discrete random approximations X_J and illustrate simulation in practice. Section 2.9 concerns questions on Riesz bases. Finally, in Appendix A, we consider integration of stationary Gaussian processes.

2.2 Wavelet bases of $L^2(\mathbb{R})$

We specify here a scaling function ϕ and a wavelet ψ which will be used below. There are many choices for these functions. We shall work with particular Meyer wavelets (Meyer (1992), Mallat (1998)) because of their nice theoretical properties. The results of this chapter and their proofs rely on specific nice properties of the selected Meyer wavelets. Other wavelet bases could be taken, e.g., the celebrated Daubechies wavelets, and are being currently investigated. Meyer wavelets are also used in Zhang and Walter (1994), Meyer *et al.* (1999) and others.

Let $\mathcal{S}(\mathbb{R})$ be the Schwartz class of $C^\infty(\mathbb{R})$ functions f that decay faster than any polynomial at infinity and so do their derivatives, that is,

$$\lim_{|t|\to\infty} t^m \frac{d^n f(t)}{dt^n} = 0,$$

for any $m, n \geq 1$. We can choose a scaling function $\phi \in \mathcal{S}(\mathbb{R})$ satisfying

$$\widehat{\phi}(x) \in [0,1], \quad \widehat{\phi}(x) = \widehat{\phi}(-x),$$

$$\widehat{\phi}(x) = \begin{cases} 1, & |x| \leq 2\pi/3, \\ 0, & |x| > 4\pi/3, \end{cases} \qquad \widehat{\phi}(x) \text{ decreases on } [0,\infty).$$

The corresponding CMF u has the discrete Fourier transform

$$\widehat{u}(x) = \begin{cases} \sqrt{2}\,\widehat{\phi}(2x), & |x| \le 2\pi/3, \\ 0, & |x| > 2\pi/3. \end{cases}$$

The wavelet function ψ associated with ϕ is such that $\psi \in \mathcal{S}(\mathbb{R})$ and

$$\widehat{\psi}(x) = \frac{1}{\sqrt{2}}\,\widehat{v}\left(\frac{x}{2}\right)\widehat{\phi}\left(\frac{x}{2}\right) \quad \text{with} \quad \widehat{v}(x) = e^{-ix}\overline{\widehat{u}(x+\pi)}, \tag{2.14}$$

where v is the other CMF. One can verify that, for the Meyer wavelets,

$$\widehat{\psi}(x) = e^{-\frac{ix}{2}}\left(\widehat{\phi}\left(\frac{x}{2}\right)^2 - \widehat{\phi}(x)^2\right)^{1/2}. \tag{2.15}$$

In particular, $\widehat{\psi}(x) = 0$ for $|x| \le 2\pi/3$ and $|x| \ge 8\pi/3$. The collection of functions $\phi(t-k), 2^{j/2}\psi(2^j t - k)$, $k \in \mathbb{Z}, j \ge 0$, makes an orthonormal basis of $L^2(\mathbb{R})$.

2.3 Basis functions and discrete approximations

Let $g \in L^2(\mathbb{R})$ be a kernel function appearing in (2.1), and $g_J = \{g_{J,n}\}_{n\in\mathbb{Z}}$, $J \in \mathbb{Z}$, be sequences of real numbers such that $g_J \in l^2(\mathbb{Z})$. Following Section 2.1 (see, in particular, (2.13)), we shall think of g_J as a *discrete (deterministic) approximation* of g at scale 2^{-J}.

A discrete approximation $g_J \in l^2(\mathbb{Z})$ induces a *discrete (random) approximation* $X_J = \{X_{J,n}\}$ defined by (2.10), that is,

$$X_{J,n} = \sum_{k=-\infty}^{\infty} g_{J,k} a_{J,n-k} \tag{2.16}$$

in the time domain, or symbolically

$$\widehat{X}_J(x) = \widehat{g}_J(x)\widehat{a}_J(x) \tag{2.17}$$

in the frequency domain, where $a_J = \{a_{J,n}\}$ are independent $\mathcal{N}(0,1)$ random variables (Gaussian white noise). As $J \to \infty$, we expect that $2^{J/2}X_{J,[2^J t]}$ approximates $X(t)$ defined by (2.1). Conversely, we may think that a random discrete approximation X_J of X

given by (2.1) can be represented by (2.16) with a sequence g_J. Hence, X_J also induces a deterministic discrete approximation g_J of g.

We will make some of the following assumptions on g and g_J. Let $L^p_{loc}(\mathbb{R})$ consist of functions which are in L^p on any compact interval of $\mathbb{R}$. Set also

$$G_J(x) = \frac{\widehat{g}_J(x)}{\widehat{g}(2^J x)}, \quad x \in \mathbb{R}. \tag{2.18}$$

Note that, with the notation (2.18), expressions (2.9) and (2.12) become

$$\widehat{\Phi}^J(x) = (G_J(2^{-J}x))^{-1}\, 2^{-J/2}\widehat{\phi}(2^{-J}x), \quad \widehat{\Phi}_J(x) = \overline{G_J(2^{-J}x)} 2^{-J/2}\widehat{\phi}(2^{-J}x) \tag{2.19}$$

Assumption 1: Suppose that

$$\widehat{g}^{-1} \in L^2_{loc}(\mathbb{R}). \tag{2.20}$$

Assumption 2: Suppose that, for any $J \in \mathbb{Z}$,

$$G_J,\ G_J^{-1} \in L^2_{loc}(\mathbb{R}). \tag{2.21}$$

Assumption 3: Suppose that, for any $J_0 \in \mathbb{Z}$,

$$\max_{p=-1,1} \max_{k=0,1,2} \sup_{J \geq J_0} \sup_{|x| \leq 4\pi/3} \left| \frac{\partial^k (G_J(x))^p}{\partial x^k} \right| < \infty. \tag{2.22}$$

Assumption 4: Suppose that, for large $|x|$,

$$\left| \frac{\partial^k \widehat{g}(x)}{\partial x^k} \right| \leq \frac{\text{const}}{|x|^{k+1}}, \quad k = 0, 1, 2. \tag{2.23}$$

Assumption 5: Assume that, for large J,

$$|G_J(0) - 1| \leq \text{const}\ 2^{-J}. \tag{2.24}$$

As explained below, Assumptions 1 and 2 ensure that the basis functions used in decompositions are well-defined. Assumptions 3, 4 and 5 will be used to establish the modification (2.7) and to show that X_J is an approximation sequence for X in the sense of (2.8).

Observe that the functions θ^j and Ψ^j in (2.4) are well-defined pointwise through the inverse Fourier transform since $\widehat{\theta}^j, \widehat{\Psi}^j \in L^1(\mathbb{R})$ for $\widehat{g} \in L^2(\mathbb{R})$. By using Assumptions 1 and 2, the functions θ_j and Ψ_j in (2.6), Φ^j in (2.9) and Φ_j in (2.12) (see also (2.19)) are well-defined pointwise through the inverse Fourier transform as well. Moreover, $\theta_j, \Psi_j, \Phi^j, \Phi_j$ are in $L^2(\mathbb{R})$ because their Fourier transforms are in $L^2(\mathbb{R})$.

Appendix A contains some results on defining integrals $\int X(t)f(t)dt$. See, in particular, the definition of a related function space $L^2_g(\mathbb{R})$ in (A.7) of integrands $f(t)$. Since $\theta_j, \Psi_j \in L^2_g(\mathbb{R})$, the coefficients $a_{j,n}$ and $d_{j,n}$ in (2.5) are well-defined. Using properties of integrals developed in Appendix A, it is easy to see that $a_{j,n}$ and $d_{j,n}$ are independent $\mathcal{N}(0,1)$ random variables. Since $\Phi_j \in L^2_g(\mathbb{R})$, the integral in (2.11) is well-defined as well.

Another consequence of the above assumptions are useful bounds on the functions Φ^j, Ψ^j. We will use these bounds several times below.

Lemma 2.3.1. *Under Assumptions 3 and 4 above, we have*

$$|2^{-j/2}\Phi_j(2^{-j}u)|, |2^{-j/2}\Phi^j(2^{-j}u)| \leq \frac{C}{1+|u|^2}, \quad u \in \mathbb{R}, \tag{2.25}$$

$$|\Psi^j(2^{-j}u)| \leq \frac{C2^{-j/2}}{1+|u|^2}, \quad u \in \mathbb{R}, \tag{2.26}$$

where a constant C does not depend on $j \geq j_0$, for fixed j_0.

Proof. By definition of Φ^j in (2.9) (see also (2.19)) and after a change of variables, observe that

$$2^{-j/2}\Phi^j(2^{-j}u) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{iux}(G_j(x))^{-1}\widehat{\phi}(x)dx, \quad u \in \mathbb{R}. \tag{2.27}$$

Since $\text{supp}\{\widehat{\phi}\} \subset \{|x| \leq 4\pi/3\}$, we obtain by Assumption 3 that

$$|2^{-j/2}\Phi^j(2^{-j}u)| \leq C, \quad u \in \mathbb{R}, \tag{2.28}$$

for a constant C which does not depend on $j \geq j_0$, for fixed j_0. Using integration by parts in (2.27) twice and Assumption 3, we have

$$2^{-j/2}\Phi^j(2^{-j}u) = -\frac{1}{2\pi u^2}\int_{\mathbb{R}} e^{iux}\frac{\partial^2}{\partial x^2}\Big((G_j(x))^{-1}\widehat{\phi}(x)\Big)dx, \quad u \in \mathbb{R}.$$

By Assumption 3 and properties of $\widehat{\phi}$, for any $j \geq j_0$,

$$|2^{-j/2}\Phi^j(2^{-j}u)| \leq \frac{C}{|u|^2}\int_{|x|\leq 4\pi/3}\Bigg(\left|\frac{\partial^2}{\partial x^2}(G_j(x))^{-1}\right| + \left|\frac{\partial}{\partial x}(G_j(x))^{-1}\right| + \left|(G_j(x))^{-1}\right|\Bigg)dx \leq \frac{C}{|u|^2}, \quad u \in \mathbb{R}.$$

The bound (2.25) for Φ^j follows from (2.28) and (2.29). The case of Φ_j is proved similarly.

To show the bound (2.26), observe from (2.4) that

$$\Psi^j(2^{-j}u) = \frac{2^{j/2}}{2\pi}\int_{\mathbb{R}} e^{iux}\widehat{g}(2^jx)\widehat{\psi}(x)dx, \quad u \in \mathbb{R}. \tag{2.29}$$

Since $\text{supp}\{\widehat{\psi}\} \subset \{2\pi/3 \leq |x| \leq 8\pi/3\}$, we obtain by Assumption 4 that

$$|\Psi^j(2^{-j}u)| \leq C2^{j/2}\int_{2\pi/3}^{8\pi/3}\frac{dx}{1+2^jx} \leq C'2^{-j/2}, \quad u \in \mathbb{R}, \tag{2.30}$$

for constants C, C' which do not depend on $j \geq j_0$, for fixed j_0. Using integration by parts in (2.29) and Assumption 4, we have

$$\Psi^j(2^{-j}u) = -\frac{2^{j/2}}{2\pi u^2}\int_{\mathbb{R}} e^{iux}\frac{\partial^2}{\partial x^2}\Big(\widehat{g}(2^jx)\widehat{\psi}(x)\Big)dx, \quad u \in \mathbb{R}. \tag{2.31}$$

Hence, by using properties of $\widehat{\psi}$ and Assumption 4, for $j \geq j_0$,

$$\begin{aligned}|\Psi^j(2^{-j}u)| &\leq \frac{C2^{j/2}}{|u|^2}\int_{2\pi/3\leq|x|\leq 8\pi/3}\left(2^{2j}\left|\frac{\partial^2\widehat{g}}{\partial x^2}(2^jx)\right| + 2^j\left|\frac{\partial\widehat{g}}{\partial x}(2^jx)\right| + |\widehat{g}(2^jx)|\right)dx \\ &\leq \frac{C'2^{j/2}}{|u|^2}\int_{2\pi/3}^{8\pi/3}\left(\frac{2^{2j}}{1+2^{3j}x^3} + \frac{2^j}{1+2^{2j}x^2} + \frac{1}{1+2^jx}\right)dx \\ &\leq C''\frac{2^{-j/2}}{|u|^2}, \quad u \in \mathbb{R}. \end{aligned} \tag{2.32}$$

The bound (2.26) follows from (2.30) and (2.32). □

2.4 Examples

We consider here several examples of Gaussian stationary processes together with their possible discrete approximations.

Example 2.4.1. The Ornstein-Uhlenbeck (OU) process X is perhaps the best-known Gaussian stationary process. It is the only Gaussian stationary process which is Markov. The OU process can be represented by (2.1) with

$$g(t) = \sigma e^{-\lambda t} 1_{\{t \geq 0\}}, \quad \widehat{g}(x) = \frac{\sigma}{\lambda + ix}, \tag{2.33}$$

for some $\lambda > 0$ and $\sigma > 0$.

At this point, one can approximate either g or X. We do so for the process X because it has a well-known discrete approximation. Observe from (2.1) and (2.33) that, for $J, n \in \mathbb{Z}$,

$$X(2^{-J}(n+1)) = e^{-\lambda 2^{-J}} X(2^{-J}n) + \sigma \sqrt{\frac{1 - e^{-2\lambda 2^{-J}}}{2\lambda}}\, a_{J,n+1},$$

where $\{a_{J,n}\}_{n \in \mathbb{Z}}$ is a Gaussian white noise. Therefore, since we expect $2^{J/2} X_{J,[2^J t]} \approx X(t)$, it appears natural to consider the discrete approximation

$$X_{J,n} = 2^{-J/2} \sigma \sqrt{\frac{1 - e^{-2\lambda 2^{-J}}}{2\lambda}}\, (I - e^{-\lambda 2^{-J}} B)^{-1}\, a_{J,n}, \tag{2.34}$$

where B denotes the backshift operator (not to be confused with Bm) and $I = B^0$. In other words, X_J is an AR(1) time series (see Brockwell and Davis (1991)).

In view of (2.33) and (2.34), the deterministic discrete approximations g_J have the discrete Fourier transforms

$$\widehat{g}_J(x) = 2^{-J/2} \sigma \sqrt{\frac{1 - e^{-2\lambda 2^{-J}}}{2\lambda}}\, (1 - e^{-\lambda 2^{-J}} e^{-ix})^{-1}. \tag{2.35}$$

Furthermore, $\widehat{g}$ and $\widehat{g}_J$ satisfy Assumptions 1 to 5. Indeed, Assumptions 1 and 2 hold because, for every $J \in \mathbb{Z}$,

$$\widehat{g}^{-1}(x) = \frac{\lambda + ix}{\sigma}, \quad G_J(x) = 2^{J/2}\sqrt{\frac{1 - e^{-2\lambda 2^{-J}}}{2\lambda}}\,\frac{2^{-J}\lambda + ix}{1 - e^{-\lambda 2^{-J}}e^{-ix}} \tag{2.36}$$

and G_J^{-1} are continuous functions on $\mathbb{R}$, and thus square-integrable on compact sets.

To show Assumption 3, consider the domain $D^{J_0} = \{z \in \mathbb{C} :\ 0 \leq \mathrm{Re}(z) \leq 2^{-J_0}\lambda, |\mathrm{Im}(z)| \leq 4\pi/3\}$. The functions

$$F(z) = \begin{cases} \frac{z}{1-e^{-z}}, & z \in \mathbb{C}\backslash\{i2k\pi,\ k \in \mathbb{Z}\}, \\ 1, & z = 0, \end{cases}$$

and $F(z)^{-1}$ are holomorphic and different from zero on the open set $D_\epsilon^{J_0} = \{w \in \mathbb{C} : \inf_{z \in D^{J_0}} |z - w| < \epsilon\} \supset D^{J_0}$. By setting $z = 2^{-J}\lambda + ix \in D^{J_0}$, we have $G_J(x) = C_J F(z)$ for all $J \geq J_0$ and $|x| \leq 4\pi/3$, where $0 < c_1 \leq C_J \leq c_2 < +\infty$ for some c_1, c_2. Hence, Assumption 3 must hold.

Assumption 4 follows from the relation

$$\frac{\partial^k \widehat{g}(x)}{\partial x^k} = \frac{\sigma(-i)^k k!}{(\lambda + ix)^{k+1}}, \quad k = 0, 1, 2, ...$$

Finally, Assumption 5 is also satisfied because

$$|G_J(0) - 1| = \left|2^{J/2}\sqrt{\frac{1 - e^{-2\lambda 2^{-J}}}{2\lambda}}\left(\frac{2^{-J}\lambda}{1 - e^{-\lambda 2^{-J}}}\right) - 1\right| = \sqrt{\frac{\lambda 2^{-J}}{1 - e^{-\lambda 2^{-J}}}}\ \left|\sqrt{\frac{1 + e^{-\lambda 2^{-J}}}{2}}\right.$$

$$\left. - \sqrt{\frac{1 - e^{-\lambda 2^{-J}}}{\lambda 2^{-J}}}\right| \leq C_1 \left(\left|\sqrt{\frac{1 + e^{-\lambda 2^{-J}}}{2}} - 1\right| + \left|\sqrt{\frac{1 - e^{-\lambda 2^{-J}}}{\lambda 2^{-J}}} - 1\right|\right) \leq C_2\, 2^{-J}$$

for constants $C_1, C_2 > 0$.

Example 2.4.2. Consider a Gaussian stationary process (2.1) with a kernel function g having the Fourier transform

$$\widehat{g}(x) = \frac{f(x)}{h(x)}. \tag{2.37}$$

Here,

$$f(x) = \prod_{k\in\mathcal{P}_1} p(a_k, b_k; x)\, p(-a_k, b_k; x) \prod_{m\in\mathcal{P}_2} p(0, c_m; x), \tag{2.38}$$

$$h(x) = \prod_{k\in\mathcal{Q}_1} p(d_k, e_k; x)\, p(-d_k, e_k; x) \prod_{m\in\mathcal{Q}_2} p(0, f_m; x) \tag{2.39}$$

with

$$p(a, b; x) = ix + ia + b, \tag{2.40}$$

where $\mathcal{P}_1, \mathcal{P}_2, \mathcal{Q}_1$ and $\mathcal{Q}_2$ are finite sets of indices. It is assumed that polynomials $f(x)$ and $h(x)$ have no common roots, and also that $\forall k \in \mathcal{Q}_1$, $e_k \neq 0$, and $\forall m \in \mathcal{Q}_2$, $f_m \neq 0$. Note that the polynomials f and h are Hermitian symmetric. Hence, $\widehat{g}$ is also Hermitian symmetric and thus g is real-valued. Kernel functions $\widehat{g}$ as in (2.37) correspond to rational spectral densities (Rozanov (1967)).

To define a discrete approximation $\widehat{g}_J$ of $\widehat{g}$, consider first $p(a, b; x)$, which is a "building block" of f in (2.38) and h in (2.39). Define a discrete approximation of $p(a, b; x)$ as

$$p_J(a, b; x) = 2^J \left(1 - e^{-2^{-J}b - 2^{-J}ia - ix}\right) \tag{2.41}$$

and also, in analogy to (2.18), set

$$P_J(x) = \frac{p_J(a, b; x)}{p(a, b; 2^J x)} = \frac{1 - e^{-2^{-J}b - 2^{-J}ia - ix}}{ix + 2^{-J}ia + 2^{-J}b}. \tag{2.42}$$

The form (2.41) ensures that $p_J(a, b; x)p_J(-a, b; x)$ and $p_J(0, b; x)$ are Hermitian symmetric functions. Define now a discrete approximation $\widehat{g}_J$ of $\widehat{g}$ by (2.37), where p's in (2.38) and (2.39) are replaced by p_J's. The function G_J is then given by (2.37), where p's in (2.38) and (2.39) are replaced by P_J's.

We shall now verify that $\widehat{g}$ and $\widehat{G}_J$ satisfy Assumptions 1-5. Assumptions 1 and 2 are satisfied because the "building blocks" p^{-1}, P_J and P_J^{-1} for $\widehat{g}^{-1}$, G_J and G_J^{-1} are continuous functions on the real line. To show Assumption 3, it is enough to prove (2.22) for the function P_J. Similarly to the case of the Ornstein-Uhlenbeck process, we are interested in the behavior of F and F^{-1} for $z = i(x + 2^{-J}a) + 2^{-J}b$, where $|x| \leq 4\pi/3$ and $J \geq J_0$. So, define the set

$$D^{J_0} = \Big\{ z \in \mathbb{C} : \ 0 \leq \Re(z) \leq 2^{-J_0}b, |\Im(z)| \leq \frac{4\pi}{3} + \Re(z)\Big|\frac{a}{b}\Big| \Big\},$$

and note that $z = i(x + 2^{-J}a) + 2^{-J}b \in D^{J_0}$ when $|x| \leq 4\pi/3$ and $J \geq J_0$. Also, consider the set $D^{J_0}_{\epsilon} = \{w \in \mathbb{C} : \ \inf_{z \in D^{J_0}} |z - w| < \epsilon\}$. The functions F and F^{-1} are holomorphic on $D^{J_0}_{\epsilon} \supset D^{J_0}$ for small enough ϵ, and thus Assumption 3 holds.

Consider now Assumption 4. The condition (2.23) is satisfied for $k = 0$ by the definition of $\widehat{g}$ and the implicit assumption $\widehat{g} \in L^2(\mathbb{R})$ (that is, the polynomial h has a higher degree than the polynomial f). When $k = 1$, note that

$$\frac{\partial \widehat{g}(x)}{\partial x} = \frac{f'(x)}{h(x)} - \frac{f(x)h'(x)}{(h(x))^2}$$

and the condition (2.23) follows since the difference between the degrees of $f'(x)$ and $h(x)$, and those of $f(x)h'(x)$ and $(h(x))^2$ increased by 1. The case $k = 2$ can be argued in a similar way.

To show (2.24) in Assumption 5, it is enough to prove it for

$$P_J(0) = \frac{1 - e^{-2^{-J}b - 2^{-J}ia}}{2^{-J}ia + 2^{-J}b}.$$

This can be done by using standard properties of exponentials and using their Taylor expansions.

Finally, let us note that the discrete approximations g_J based on (2.41) correspond to ARMA time series X_J (Brockwell and Davis (1991)).

(Non)Example 2.4.1. Let $B_H(t)$, $H \in (0,1)$, be fractional Brownian motion (fBm, in short), that is, a Gaussian H-self-similar process with stationary increments (see, for example, Embrechts and Maejima (2002), Samorodnitsky and Taqqu (1994)). Consider a stationary Gaussian process $\{X(t)\}_{t \in \mathbb{R}}$ defined by $X(t) = B_H(t) - B_H(t-1)$ and known as fractional Gaussian noise (fGn). FGn has the representation (2.1) with

$$g(t) = \frac{\sigma}{C(H)}\left(t_+^{H-\frac{1}{2}} - (t-1)_+^{H-\frac{1}{2}}\right), \quad \widehat{g}(x) = \frac{\sigma\Gamma(H+\frac{1}{2})}{C(H)}\left(\frac{e^{-ix}-1}{ix}\right)(ix)^{\frac{1}{2}-H}, \quad (2.43)$$

where $\sigma > 0$, $C(H)^2 = \int_0^\infty ((1+t)^{H-1/2} - t^{H-1/2})^2 dt + (2H)^{-1}$ and $C(1/2) = 1$. With this choice of $C(H)$, $EX(1)^2 = \sigma^2$.

Since $\widehat{g}(x)^{-1} = \mathrm{const}(ix/e^{-ix} - 1)(ix)^{H-1/2}$ is not in L^2_{loc} (nor in L^1_{loc}) around the points $\{2k\pi, k \in \mathbb{Z}\backslash\{0\}\}$, the function $\widehat{g}$ in (2.43) does not satisfy Assumption 1. Hence, the functions θ_j and Ψ_j in (2.6) cannot be computed through their Fourier transforms. This is somewhat surprising because the wavelet-based representation analogous to (2.3) with (2.7) has been established for fBm by Meyer *et al.* (1999). However, it seems that one cannot do much about this. Assumption 1 already appears to be weak.

2.5 Adaptive wavelet decompositions

We first reestablish the decomposition (2.3) of Zhang and Walter (1994) by providing a more rigorous proof.

Theorem 2.5.1. *(Zhang and Walter (1994)) Let X be a Gaussian stationary process given by (2.1). Suppose that Assumptions 1 and 2 of Section 2.3 hold. Then, with the notation of Section 2.1, the process X admits the following wavelet-based decomposition: for any $J \in \mathbb{Z}$,*

$$\begin{aligned} X(t) &= \sum_{n=-\infty}^{\infty} a_{J,n}\theta^J(t - 2^{-J}n) + \sum_{j=J}^{\infty}\sum_{n=-\infty}^{\infty} d_{j,n}\Psi^j(t - 2^{-j}n) \quad (2.44)\\ &= \sum_{j=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} d_{j,n}\Psi^j(t - 2^{-j}n), \quad (2.45)\end{aligned}$$

with the convergence in the $L^2(\Omega)$-sense for each t, and independent $\mathcal{N}(0,1)$ random variables $a_{J,n}, d_{j,n}$ that are expressed through (2.5).

Proof. (Zhang and Walter (1994)) Under Assumptions 1 and 2, the basis functions θ^J and Ψ^j in (2.44) and (2.45) are well-defined pointwise (Section 2.3). The coefficients $a_{j,n}$, $d_{j,n}$ are well-defined, independent $\mathcal{N}(0,1)$ random variables (Section 2.3). Except for more rigor, the rest of the proof follows that of Zhang and Walter (1994). Since the proof is short, we provide it for the reader's convenience.

Observe that

$$E\left(X(t)-\sum_{n=-N_1}^{N_2}a_{J,n}\theta^J(t-2^{-J}n)-\sum_{j=J}^{K}\sum_{n=-M_1}^{M_2}d_{j,n}\Psi^j(t-2^{-j}n)\right)^2$$

$$=E\Bigg(X(t)^2-2\sum_{n=-N_1}^{N_2}X(t)a_{J,n}\theta^J(t-2^{-J}n)-2\sum_{j=J}^{K}\sum_{n=-M_1}^{M_2}X(t)d_{j,n}\Psi^j(t-2^{-j}n)$$

$$+\left(\sum_{n=-N_1}^{N_2}a_{J,n}\theta^J(t-2^{-J}n)+\sum_{j=J}^{K}\sum_{n=-M_1}^{M_2}d_{J,n}\Psi^j(t-2^{-j}n)\right)^2\Bigg). \tag{2.46}$$

By using Appendix A and the definition of function θ^j (Sections 2.1 and 2.3), we have

$$EX(t)a_{J,n}=EX(t)\int_{\mathbb{R}}X(s)\theta_J(s-2^{-J}n)ds$$

$$=\frac{1}{2\pi}\int_{\mathbb{R}}e^{itx}|\widehat{g}(x)|^2\theta_J(\widehat{\cdot-2^{-J}}n)(x)dx=\frac{1}{2\pi}\int_{\mathbb{R}}e^{i(t-2^{-J}n)x}|\widehat{g}(x)|^2\widehat{\theta}_J(x)dx$$

$$=\frac{1}{2\pi}\int_{\mathbb{R}}e^{i(t-2^{-J}n)x}2^{-J/2}\widehat{g}(x)\widehat{\phi}_J(2^{-J}x)dx=\theta^J(t-2^{-J}n). \tag{2.47}$$

Similarly, we have

$$EX(t)d_{j,n}=\Psi^j(t-2^{-j}n). \tag{2.48}$$

Using (2.47), (2.48) and independence of $a_{J,n}$, $d_{j,n}$, relation (2.46) becomes

$$R(0)-\sum_{n=-N_1}^{N_2}\theta^J(t-2^{-J}n)^2-\sum_{j=J}^{K}\sum_{n=-M_1}^{M_2}(\Psi^j(t-2^{-j}n))^2. \tag{2.49}$$

Observe from the definition of θ^J that

$$\theta^J(t-2^{-J}n)=\frac{1}{2\pi}\int_{\mathbb{R}}e^{i(t-2^{-J}n)x}2^{-J/2}\widehat{g}(x)\widehat{\phi}_J(2^{-J}x)dx$$

$$=\frac{1}{2\pi}\int_{\mathbb{R}}\widehat{g}(x)2^{J/2}(\phi(2^J\widehat{(\cdot+t)}-n))(x)dx=\frac{1}{2\pi}\int_{\mathbb{R}}g(u)2^{J/2}\phi(n-2^J(u+t))du, \tag{2.50}$$

and similarly

$$\Psi^j(t-2^{-j}n)=\frac{1}{2\pi}\int_{\mathbb{R}}g(u)2^{j/2}\psi(n-2^j(u+t))du. \tag{2.51}$$

Since the collection of functions $2^{J/2}\phi(n-2^J(u+t))$, $2^{j/2}\phi(n-2^j(u+t))$, $j \geq J$, $n \in \mathbb{Z}$, makes an orthonormal basis of $L^2(\mathbb{R})$ for any $t \in \mathbb{R}$, and since $R(0) = \int_{\mathbb{R}} |g(t)|^2 dt$, we obtain from (2.50) and (2.51) that relation (2.49) converges to 0 as N_i, M_i ($i = 1, 2$) and K approach infinity. □

In the next result, we modify the approximation term in the decomposition (2.44) according to (2.7).

Theorem 2.5.2. *Let X be a Gaussian stationary process given by (2.1). Suppose that Assumptions 1 and 2 of Section 2.3 hold. Then, with the notation of Section 2.1, the process X admits the following wavelet-based decomposition: for any $J \in \mathbb{Z}$,*

$$X(t) = \sum_{n=-\infty}^{\infty} X_{J,n}\Phi^J(t - 2^{-J}n) + \sum_{j=J}^{\infty}\sum_{n=-\infty}^{\infty} d_{j,n}\Psi^j(t-2^{-j}n). \tag{2.52}$$

The convergence in (2.52) is in the $L^2(\Omega)$-sense for each t under Assumption 3, and it is almost sure, uniform over compact intervals of t under Assumptions 3 and 4. The sequence $X_J = \{X_{J,n}\}_{n\in\mathbb{Z}}$ is defined by either (2.11) or (2.16).

Proof. We first argue that the definitions (2.11) and (2.16) of X_j are equivalent. By using Appendix A, observe that, for $X_{J,n}$ defined by (2.11) and $a_{J,n}$ defined by (2.5),

$$E\Big(X_{J,n} - \sum_{k=-N_1}^{N_2} g_{J,k}a_{J,n-k}\Big)^2$$

$$= E\Bigg(\int_{\mathbb{R}} X(t)\Bigg(\Phi_J(t-2^{-J}n) - \sum_{k=-N_1}^{N_2} g_{J,k}\theta_J(t-2^{-J}(n-k))\Bigg)dt\Bigg)^2$$

$$= \frac{1}{2\pi}\int_{\mathbb{R}} \Big|\widehat{g}_J(2^{-J}x) - \sum_{k=-N_1}^{N_2} g_{J,k}e^{ix2^{-J}k}\Big|^2 2^{-J}\left|\widehat{\phi}(2^{-J}x)\right|^2 dx \longrightarrow 0,$$

as $N_i \to \infty$ ($i = 1, 2$), since $\sum_k g_{J,k}e^{ixk}$ converges to $\widehat{g}_J(x)$ in $L^2(-\pi,\pi)$ and $\widehat{\phi}$ has a compact support.

To show (2.52), we start with (2.44) and modify its first sum as (2.7), that is,

$$\sum_{n=-\infty}^{\infty} a_{J,n}\theta^J(t-2^{-J}n) = \sum_{n=-\infty}^{\infty} X_{J,n}\Phi^J(t-2^{-J}n). \tag{2.53}$$

We first show that, under Assumption 3, the R.H.S. converges in the $L^2(\Omega)$-sense for fixed t and, under Assumptions 3 and 4, almost surely, uniformly over compacts of t. Observe that, by Lemma 2.3.1, $|\Phi^J(t-2^{-J}n)| \le C/(1+|t-2^{-J}n|^2)$ and, by Lemma 3 in Meyer *et al.* (1999), $|X_{J,n}| \le A\sqrt{\log(2+|n|)}$ a.s., where a random variable A does not depend on n. The almost sure convergence uniformly on compacts $t \in K$ follows since

$$\sup_{t\in K}\sum_{n=-\infty}^{\infty}|X_{J,n}||\Phi^J(t-2^{-J}n)| \le A\sup_{t\in K}\sum_{n=-\infty}^{\infty}\frac{\sqrt{\log(2+|n|)}}{1+|t-2^{-J}n|^2} < \infty \quad \text{a.s.}$$

For the convergence in $L^2(\Omega)$, observe that, for fixed t,

$$E\left(\sum_{n=-\infty}^{\infty}|X_{J,n}||\Phi^J(t-2^{-J}n)|\right)^2 \le CE\sum_{n=-\infty}^{\infty}\frac{|X_{J,n}|^2}{1+|n|^2}\sum_{n=-\infty}^{\infty}\frac{1}{1+|n|^2} < \infty.$$

We shall now prove the equality in (2.53). Observe that, for each u,

$$\theta^J(u) = \sum_{k=-\infty}^{\infty} g_{J,k}\Phi^J(u-2^{-J}k). \tag{2.54}$$

Indeed, arguing as above,

$$F_m(u) = \sum_{k=-m}^{m} g_{J,k}\Phi^J(u-2^{-J}k) \longrightarrow F(u) = \sum_{k=-\infty}^{\infty} g_{J,k}\Phi^J(u-2^{-J}k) \tag{2.55}$$

pointwise, and

$$\widehat{F}_m(x) = \left(\sum_{k=-m}^{m} g_{J,k}e^{-i2^{-J}kx}\right)\frac{\widehat{g}(x)}{\widehat{g}_J(2^{-J}x)}2^{-J/2}\widehat{\phi}(2^{-J}x) \longrightarrow \widehat{\theta}^J(x) \tag{2.56}$$

in $L^2(\mathbb{R})$, since $\sum_{k=-m}^{m} g_{J,k}e^{-ikx}$ converges to $\widehat{g}_J(x)$ in $L^2(-\pi,\pi)$, and $\widehat{g}(x)/\widehat{g}_J(2^{-J}x)$ is bounded by Assumption 3 on the compact support of $\widehat{\phi}(2^{-J}x)$. Hence, $F_m \to \theta^J$ in $L^2(\mathbb{R})$ and $\theta^J = F$ a.e. Since both F and θ^J are continuous, we obtain (2.54).

Set now, for $m \geq 1$,

$$a_{J,n}^{(m)} = \begin{cases} a_{J,n}, & |n| \leq m, \\ 0, & |n| > m, \end{cases} \qquad X_{J,n}^{(m)} = \sum_{k=-\infty}^{\infty} g_{J,k} a_{J,n-k}^{(m)}.$$

By using (2.54), we obtain that

$$\sum_{n=-\infty}^{\infty} a_{J,n}^{(m)} \theta^J(t - 2^{-J}n) = \sum_{n=-\infty}^{\infty} X_{J,n}^{(m)} \Phi^J(t - 2^{-J}n). \tag{2.57}$$

The L.H.S. of (2.57) converges in $L^2(\Omega)$ to the L.H.S. of (2.53) (and, in fact, also almost surely by the Three Series Theorem). Let us show that the R.H.S. of (2.57) converges to the R.H.S. of (2.53). We want to argue next that

$$\sup_{m \geq 1} |X_{J,n}^{(m)}| \leq A\sqrt{\log(2 + |n|)} \quad \text{a.s.} \tag{2.58}$$

for a random variable A which only depends on J. By using the Lévy-Octaviani inequality (e.g. Proposition 1.1.1 in Kwapień and Woyczyński (1992)), we have

$$P\left(\sup_{m=1,\dots,M} |X_{J,n}^{(m)}| > a\right) \leq 2P\Big(|X_{J,n}^{(M)}| > a\Big), \tag{2.59}$$

for any $a > 0$ and $M \geq 1$. By the Three Series Theorem, $X_{J,n}^{(M)} \to X_{J,n}$ almost surely, as $M \to \infty$. Hence, passing to the limit with M in (2.59), we have

$$P\left(\sup_{m \geq 1} |X_{J,n}^{(m)}| > a\right) \leq 2P(|X_{J,n}| > a).$$

The bound (2.58) now follows as in the proof of Lemma 3 in Meyer *et al.* (1999). By using Lemma 2.3.1 and the bound (2.58), the R.H.S. of (2.57) converges a.s. to the R.H.S. of (2.53).

It is left to show that the second term in (2.52) converges almost surely and uniformly on compacts. By Lemma 3 in Meyer *et al.* (1999),

$$|d_{j,n}| \leq A\sqrt{\log(2 + |j|)}\sqrt{\log(2 + |n|)} \quad \text{a.s.},$$

where a random variable A does not depend on j, n. By Lemma 2.3.1, we have

$$|\Psi^j(t - 2^{-j}n)| = |\Psi^j(2^{-j}(2^j t - n))| \leq \frac{C 2^{-j/2}}{1 + |2^j t - n|^2},$$

for $j \geq J$. Then, as in the proof of Theorem 2 in Meyer *et al.* (1999),

$$\begin{aligned}\sum_{j=J}^{\infty} \sum_{n=-\infty}^{\infty} |d_{j,n}||\Psi^j(t - 2^{-j}n)| &\leq A' \sum_{j=J}^{\infty} 2^{-j/2} \sqrt{\log(2 + |j|)} \sum_{n=-\infty}^{\infty} \frac{\sqrt{\log(2 + |n|)}}{1 + |2^j t - n|^2} \\ &\leq A'' \sum_{j=J}^{\infty} 2^{-j/2} \sqrt{\log(2 + |j|)} \sqrt{\log(2 + |2^j t|)} < \infty\end{aligned}$$

a.s. uniformly over compact intervals of t. □

2.6 FWT-like algorithm

We show here that discrete approximation sequences X_j are related across different scales by a FWT-like algorithm.

Proposition 2.6.1. *Let X_j and d_j be the sequences appearing in (2.52), and let u and v denote the CMFs associated with the orthogonal Meyer MRA. Then, under Assumptions 1–4 of Section 2.3:*

(i) (Reconstruction step)

$$X_{j+1} = u_j * \uparrow_2 X_j + v_j * \uparrow_2 d_j, \tag{2.60}$$

where the filters u_j and v_j are defined through their discrete Fourier transforms

$$\widehat{u}_j(x) = \frac{\widehat{g}_{j+1}(x)}{\widehat{g}_j(2x)} \widehat{u}(x), \quad \widehat{v}_j(x) = \widehat{g}_{j+1}(x) \widehat{v}(x); \tag{2.61}$$

(ii) (Decomposition step)

$$X_j = \downarrow_2 (\overline{u}_j^d * X_{j+1}), \quad d_j = \downarrow_2 (\overline{v}_j^d * X_{j+1}), \tag{2.62}$$

where $\overline{x}$ stand for the time reversal of a sequence x, and the filters u_j^d and v_j^d are defined *through their discrete Fourier transforms by*

$$\widehat{u}_j^d(x) = \overline{\left(\frac{\widehat{g}_j(2x)}{\widehat{g}_{j+1}(x)}\right)}\widehat{u}(x), \quad \widehat{v}_j^d(x) = \overline{\left(\frac{1}{\widehat{g}_{j+1}(x)}\right)}\widehat{v}(x). \tag{2.63}$$

The convergence in (2.60) and (2.62) is in the $L^2(\Omega)$-sense, and also absolute almost surely.

Proof. Observe first that the filters u_j, v_j, u_j^d, v_j^d are well-defined since $\widehat{u}_j, \widehat{v}_j, \widehat{u}_j^d, \widehat{v}_j^d \in L^2(-\pi,\pi)$. The latter follows by writing

$$\widehat{u}_j(x) = G_{j+1}(x)\Big(G_j(2x)\Big)^{-1}\widehat{u}(x), \quad \widehat{v}_j(x) = G_{j+1}(x)\widehat{g}(2^{j+1}x)\widehat{v}(x),$$

$$\widehat{u}_j^d(x) = \overline{G_{j+1}(x)^{-1}}\;\overline{G_j(2x)}\widehat{u}(x), \quad \widehat{v}_j^d(x) = \overline{G_{j+1}(x)^{-1}}\;\overline{\widehat{g}(2^{j+1}x)^{-1}}\widehat{v}(x) \tag{2.64}$$

(see (2.18)), and using Assumptions 1 and 3.

(i) To show (2.60), we need to prove

$$X_{j+1,n} = \sum_{k=-\infty}^{\infty} X_{j,k}u_{j,n-2k} + \sum_{k=-\infty}^{\infty} d_{j,k}v_{j,n-2k}. \tag{2.65}$$

We first prove the convergence in (2.65) in the $L^2(\Omega)$-sense. Observe that, by using (2.11), (2.5) and Appendix A,

$$E\left(X_{j+1,n} - \left(\sum_{k=-K}^{K}\left(X_{j,k}u_{j,n-2k} + d_{j,k}v_{j,n-2k}\right)\right)\right)^2$$

$$= E\left(\int_{\mathbb{R}} X(t)\left(\Phi_{j+1}(t-2^{-j-1}n) - \sum_{k=-K}^{K}\Phi_j(t-2^{-j}k)u_{j,n-2k}\right.\right.$$

$$\left.\left. - \sum_{k=-K}^{K}\Psi_j(t-2^{-j}k)v_{j,n-2k}\right)dt\right)^2$$

$$= \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{g}(x)|^2 \Big| e^{-i2^{-j-1}nx}\widehat{\Phi}_{j+1}(x) - \widehat{\Phi}_j(x)\sum_{k=-K}^{K} e^{-i2^{-j}kx}u_{j,n-2k}$$

$$-\widehat{\Psi}_j(x) \sum_{k=-K}^{K} e^{-i2^{-j}kx} v_{j,n-2k}\Big|^2 dx$$

$$= \frac{1}{2\pi} \int_{\mathbb{R}} \Big| e^{-i2^{-j-1}nx} \overline{\widehat{g}_{j+1}(2^{-j-1}x)} 2^{-(j+1)/2} \widehat{\phi}(2^{-j-1}x) - \overline{\widehat{g}_j(2^{-j}x)} 2^{-j/2} \widehat{\phi}(2^{-j}x) \cdot$$

$$\cdot \sum_{k=-K}^{K} e^{-i2^{-j}kx} u_{j,n-2k} - 2^{-j/2} \widehat{\psi}(2^{-j}x) \sum_{k=-K}^{K} e^{-i2^{-j}kx} v_{j,n-2k}\Big|^2 dx.$$

Hence, it is sufficient to prove that

$$\overline{\widehat{g}_j(2^{-j}x)} 2^{-j/2} \widehat{\phi}(2^{-j}x) \sum_{k=-\infty}^{\infty} e^{-i2^{-j}kx} u_{j,n-2k} + 2^{-j/2} \widehat{\psi}(2^{-j}x) \sum_{k=-\infty}^{\infty} e^{-i2^{-j}kx} v_{j,n-2k}$$

$$= e^{-i2^{-j-1}nx} \overline{\widehat{g}_{j+1}(2^{-j-1}x)} 2^{-(j+1)/2} \widehat{\phi}(2^{-j-1}x) \tag{2.66}$$

with the convergence in $L^2(\mathbb{R})$. We only consider the case $n = 2p$ (the case $n = 2p+1$ may be treated in an analogous fashion). Then, relation (2.66) becomes

$$\overline{\widehat{g}_j(2^{-j}x)}\; 2^{-j/2}\; \widehat{\phi}(2^{-j}x) \sum_{m=-\infty}^{\infty} e^{i2^{-j}mx} u_{j,2m} + 2^{-j/2} \widehat{\psi}(2^{-j}x) \sum_{m=-\infty}^{\infty} e^{i2^{-j}mx} v_{j,2m}$$

$$= \overline{\widehat{g}_{j+1}(2^{-j-1}x)}\; 2^{-(j+1)/2}\; \widehat{\phi}(2^{-j-1}x). \tag{2.67}$$

The L.H.S. of (2.67) is

$$\overline{\widehat{g}_j(2^{-j}x)}\; 2^{-j/2}\; \widehat{\phi}(2^{-j}x) \frac{\overline{\widehat{u}_j(2^{-j-1}x)} + \overline{\widehat{u}_j(2^{-j-1}x+\pi)}}{2}$$

$$+2^{-j/2} \widehat{\psi}(2^{-j}x) \frac{\overline{\widehat{v}_j(2^{-j-1}x)} + \overline{\widehat{v}_j(2^{-j-1}x+\pi)}}{2}$$

$$= 2^{-1} 2^{-j/2}\;\; \widehat{\phi}(2^{-j}x) \left(\overline{\widehat{g}_{j+1}(2^{-j-1}x)}\; \overline{\widehat{u}(2^{-j-1}x)} + \overline{\widehat{g}_{j+1}(2^{-j-1}x+\pi)}\; \overline{\widehat{u}(2^{-j-1}x+\pi)} \right)$$

$$+2^{-1} 2^{-j/2}\; \widehat{\psi}(2^{-j}x) \left(\overline{\widehat{g}_{j+1}(2^{-j-1}x)}\; \overline{\widehat{v}(2^{-j-1}x)} + \overline{\widehat{g}_{j+1}(2^{-j-1}x+\pi)}\; \overline{\widehat{v}(2^{-j-1}x+\pi)} \right)$$

$$= \overline{\widehat{g}_{j+1}(2^{-j-1}x)} \Big(2^{-j/2}\; \widehat{\phi}(2^{-j}x) \frac{\overline{\widehat{u}(2^{-j-1}x)} + \overline{\widehat{u}(2^{-j-1}x+\pi)}}{2}$$

$$+2^{-j/2} \widehat{\psi}(2^{-j}x) \frac{\overline{\widehat{v}(2^{-j-1}x)} + \overline{\widehat{v}(2^{-j-1}x+\pi)}}{2} \Big)$$

$$+2^{-1}2^{-j/2}\Big(\widehat{\phi}(2^{-j}x)\overline{\widehat{u}(2^{-j-1}x+\pi)}+\widehat{\psi}(2^{-j}x)\overline{\widehat{v}(2^{-j-1}x+\pi)}\Big)\cdot$$

$$\cdot\Big(\widehat{g}_{j+1}(2^{-j-1}x+\pi)-\widehat{g}_{j+1}(2^{-j-1}x)\Big).$$

This is also R.H.S. of (2.67) since

$$2^{-j/2}\,\widehat{\phi}(2^{-j}x)\frac{\overline{\widehat{u}(2^{-j-1}x)}+\overline{\widehat{u}(2^{-j-1}x+\pi)}}{2}+2^{-j/2}\widehat{\psi}(2^{-j}x)\frac{\overline{\widehat{v}(2^{-j-1}x)}+\overline{\widehat{v}(2^{-j-1}x+\pi)}}{2}$$

$$=2^{-(j+1)/2}\widehat{\phi}(2^{-j-1}x)$$

(this is the Fourier transform of the last relation in the proof of Theorem 7.7 in Mallat (1998)) and, with $y=2^{-j-1}x$,

$$\widehat{\phi}(2^{-j}x)\overline{\widehat{u}(2^{-j-1}x+\pi)}+\widehat{\psi}(2^{-j}x)\overline{\widehat{v}(2^{-j-1}x+\pi)}=\widehat{\phi}(2y)\overline{\widehat{u}(y+\pi)}+\widehat{\psi}(2y)\overline{\widehat{v}(y+\pi)}$$

$$=2^{-1/2}\widehat{\phi}(y)\Big(\widehat{u}(y)\overline{\widehat{u}(y+\pi)}+\widehat{v}(y)\overline{\widehat{v}(y+\pi)}\Big)=0,$$

where we used the relations $\widehat{\phi}(2y)=2^{-1/2}\widehat{\phi}(y)\widehat{u}(y)$ ((7.30) in Mallat (1998)), $\widehat{\psi}(2y)=2^{-1/2}\widehat{\phi}(y)\widehat{v}(y)$ ((7.57) in Mallat (1998)) and $\widehat{u}(y)\overline{\widehat{u}(y+\pi)}+\widehat{v}(y)\overline{\widehat{v}(y+\pi)}=0$ (Theorem 7.8 in Mallat (1998)).

We now show that the convergence in (2.65) is also absolute almost surely. By using Assumptions 3, 4, and integration by parts twice, we may conclude that

$$|u_{j,k}|,|v_{j,k}|\le C(1+|k|^2)^{-1},\quad k\in\mathbb{Z}.$$

By Lemma 3 in Meyer *et al.* (1999), $|X_{j,k}|,|d_{j,k}|\le A\sqrt{\log(2+|k|)}$ a.s., where a random variable A does not depend on k. The absolute convergence a.s. now follows.

(ii) The proof of (2.62) follows by similar arguments. We need to prove that

$$X_{j,n}=\sum_{k=-\infty}^{\infty}X_{j+1,k}u^d_{j,k-2n}\tag{2.68}$$

and

$$d_{j,n}=\sum_{k=-\infty}^{\infty}X_{j+1,k}v^d_{j,k-2n}.\tag{2.69}$$

To show (2.68) with convergence in the $L^2(\Omega)$-sense, it suffices to prove that

$$e^{-i2^{-j}nx}\overline{\widehat{g}_j(2^{-j}x)}2^{-j/2}\widehat{\phi}(2^{-j}x)$$

$$= \overline{\widehat{g}_{j+1}(2^{-(j+1)}x)}2^{-(j+1)/2}\widehat{\phi}(2^{-(j+1)}x) \sum_{k=-\infty}^{\infty} e^{-ik2^{-(j+1)}x}u^d_{j,k-2n}. \tag{2.70}$$

But the R.H.S. of (2.70) is

$$\overline{\widehat{g}_{j+1}(2^{-(j+1)}x)}2^{-(j+1)/2}\widehat{\phi}(2^{-(j+1)}x) \sum_{m=-\infty}^{\infty} e^{-i(2n+m)2^{-(j+1)}x}u^d_{j,m}$$

$$= \overline{\widehat{g}_{j+1}(2^{-(j+1)}x)}2^{-(j+1)/2}\widehat{\phi}(2^{-(j+1)}x)e^{-in2^jx}\widehat{u}^d_j(2^{-(j+1)}x)$$

$$= 2^{-(j+1)/2}\widehat{\phi}(2^{-(j+1)}x)e^{-in2^jx}\overline{\widehat{g}_j(2^{-j}x)}\widehat{u}(2^{-(j+1)}x), \tag{2.71}$$

which is also the L.H.S. of (2.70) by using $\widehat{\phi}(2y) = 2^{-1/2}\widehat{\phi}(y)\widehat{u}(y)$. The proof of the equality (2.69) in the $L^2(\Omega)$-sense is similar. The absolute almost surely convergence of (2.68) and (2.69) may be deduced by arguments analogous to those for the absolute almost surely convergence of (2.65). □

2.7 Convergence of random discrete approximations

We will also assume the following:

Assumption 6: Suppose that there are $\beta \in \mathbb{N} \cup \{0\}$ and $\alpha \in (0, 1]$ such that, for any compact K,

$$\left| X(t) - X(s) - X^{(1)}(s)(t-s) - \ldots - X^{(\beta)}(s)\frac{(t-s)^\beta}{\beta!} \right| \leq A|t-s|^{\beta+\alpha}, \tag{2.72}$$

for all $t \in \mathbb{R}$, $s \in K$, a.s., where a random variable A depends only on K. (As usual, $f^{(k)}$ denotes the kth derivative of f.)

Note that (2.72) implies, for some random variable B,

$$|X(t) - X(s)| \leq B|t-s|^\gamma \quad \text{for all } t \in \mathbb{R},\, s \in K, \quad \text{with } \gamma = 1 \wedge (\beta + \alpha). \tag{2.73}$$

Condition (2.72) in Assumption 6 is satisfied by many Gaussian stationary processes. It follows, in particular, from the two conditions:

$$X^{(\beta)} \quad \text{is } \alpha\text{-Hölder} \quad \text{a.s.} \tag{2.74}$$

and

$$|X(t)| \le C(1 + |t|)^{\beta+\alpha} \quad \text{a.s.} \tag{2.75}$$

By Theorem and a discussion on pp. 181-182 in Cramér and Leadbetter (1967), (2.74) follows from

$$\int_0^\infty x^{2\beta+2\alpha} \log(1 + x)|\widehat{g}(x)|^2 dx < \infty. \tag{2.76}$$

There is also an equivalent condition in terms of the autocovariance function of a stationary Gaussian process.

Condition (2.75) is always satisfied for stationary Gaussian processes that are bounded on compact intervals, such as for those satisfying (2.74). In fact, a stronger condition holds:

$$|X(t)| \le C\sqrt{\log(2 + |t|)} \quad \text{a.s.}, \tag{2.77}$$

where C is a random variable. To see this, note that the discrete-time sequence $X_k = \sup_{t\in[k,k+1)} |X(t)|$ is stationary. Moreover, by Theorem 2 in Lifshits (1995), p. 142, for some $m \in \mathbb{R}$ and $\sigma > 0$,

$$P(X_0 \ge m + \tau) \le 2(1 - \Phi(\tau/\sigma)), \quad \tau > 0,$$

where Φ is the distribution function of standard normal law. In other words, the right tail of the distribution function of X_n decays at least as fast as that of the distribution function of standard normal law. The bound (2.77) can then be obtained as (3.15) in Lemma 3 of Meyer *et al.* (1999).

We shall need some assumptions stronger than parts of Assumptions 3 and 5.

Assumption 3*: Suppose that, for any $J_0 \in \mathbb{Z}$,

$$\max_{k=0,1,\ldots,\beta+[\alpha]+2} \sup_{J \geq J_0} \sup_{|x| \leq 4\pi/3} \left| \frac{\partial^k G_J(x)}{\partial x^k} \right| < \infty. \tag{2.78}$$

Assumption 5*: Assume that, for large J,

$$|G_J(0) - 1| \leq \text{const } 2^{-(\beta+1)J}. \tag{2.79}$$

As in Lemma 2.3.1, under Assumption 3*, we have

$$|2^{-j/2}\Phi_j(2^{-j}u)| \leq \frac{C}{1 + |u|^{\beta+[\alpha]+2}}, \quad u \in \mathbb{R}, \tag{2.80}$$

where a constant C does not depend on $j \geq j_0$, for fixed j_0.

In addition, we will suppose the following:

Assumption 7: If $\beta \geq 1$ in Assumption 6, suppose that

$$G_J^{(n)}(0) = \frac{\partial^n G_J}{\partial x^n}(0) = 0, \quad n = 1, \ldots, \beta. \tag{2.81}$$

The next result establishes convergence of random discrete approximations.

Proposition 2.7.1. *Under Assumptions 2,3,5 of Section 2.3 and Assumption 6 above, we have*

$$\sup_{t \in K} |2^{J/2} X_{J,[2^J t]} - X(t)| \leq A_1 2^{-J\gamma} \quad \text{a.s.}, \tag{2.82}$$

where K is a compact interval and A_1 is a random variable that does not depend on J. If, in addition, Assumptions 3,5* and 7 above hold, then*

$$\sup_{t \in K} |2^{J/2} X_{J,[2^J t]} - X([2^J t]2^{-J})| \leq A_2 2^{-J(\beta+\alpha)} \quad \text{a.s.}, \tag{2.83}$$

where a random variable A_2 does not depend on J.

Proof. Suppose without loss of generality that $K = [0, 1]$. In view of Assumption 6, it is enough to show (2.83) or that

$$\sup_{k=0,\ldots,2^J} \left| 2^{J/2} X_{J,k} - X(k2^{-J}) \right| \leq A\, 2^{-J(\beta+\alpha)} \quad \text{a.s.}$$

Note by Assumption 7 and the properties of the scaling function ϕ that

$$\int_{\mathbb{R}} u^n \Phi_J(u) du = (-i)^{-n} \widehat{\Phi}_J^{(n)}(0) = (-i)^{-n} \frac{\partial^n}{\partial x^n} \left(\overline{G_J(x)} 2^{-J/2} \widehat{\phi}(2^{-J}x) \right) \Big|_{x=0} = 0,$$

for $n = 1, \ldots, \beta$. By using Appendix A, Assumptions 5* and 6 (with (2.80)), we have

$$|2^{J/2} X_{J,k} - X(k2^{-J})|$$

$$\leq 2^{J/2} \int_{\mathbb{R}} \left| X(t) - X(k2^{-J}) - \ldots - X^{(\beta)}(k2^{-J}) \frac{(t-k2^{-J})^\beta}{\beta!} \right| |\Phi_J(t-k2^{-J})| dt$$

$$+ X(k2^{-J}) |G_J(0) - 1| \leq A 2^{J/2} \int_{\mathbb{R}} |t - k2^{-J}|^{\beta+\alpha} |\Phi_J(t-k2^{-J})| dt + B 2^{-(\beta+1)J}$$

$$= A 2^{J/2} \int_{\mathbb{R}} |u|^{\beta+\alpha} |\Phi_J(u)| du + B 2^{-(\beta+1)J} \quad (\text{setting } u = 2^{-J}v)$$

$$= A' 2^{-J(\beta+\alpha)} \int_{\mathbb{R}} |v|^{\beta+\alpha} |2^{-J/2} \Phi_J(2^{-J}v)| dv$$

$$\leq A'' 2^{-J(\beta+\alpha)} \int_{\mathbb{R}} \frac{|v|^{\beta+\alpha}}{1 + |v|^{\beta+[\alpha]+2}} dv = A''' 2^{-J(\beta+\alpha)}.$$

□

According to Proposition 2.7.1, the discrete approximations $2^{J/2} X_{J,[2^J t]}$ converge to the process $X(t)$. Note also that, when $\beta \geq 1$, the convergence is faster on the dyadics than on the whole interval. An interesting question is whether the faster convergence rate $\beta + \alpha$ can be obtained on an interval for some other approximation based on $X_{J,[2^J t]}$.

For a function f, defined on either $\mathbb{R}$ or $\mathbb{Z}$, consider the operator

$$(\Delta_h^p f)(a) = \sum_{k=0}^{p} \binom{p}{k} (-1)^{p-k} f(a + kh), \quad p \in \mathbb{N},$$

where a, h are in either $\mathbb{R}$ or $\mathbb{Z}$, respectively. When $f = f_k$ is a function on $\mathbb{Z}$, we write

$$\Delta^p f_k = (\Delta_1^p f)(k) \quad \text{and} \quad \Delta^p f = (\Delta_1^p f)(0).$$

In view of the condition (2.72), to obtain the faster rate $\beta + \alpha$ on a whole interval, it is natural to try an approximation which includes the terms mimicking the β derivatives in (2.72). Thus, for $\beta \geq 1$, consider the approximations

$$\widehat{X}_{\beta,J}(t) = 2^{J/2} X_{J,[2^J t]} + 2^{J/2} \sum_{p=1}^{\beta} \frac{\Delta^p X_{J,[2^J t]}}{2^{-Jp}} \frac{(t - [2^J t]2^{-J})^p}{p!}, \tag{2.84}$$

with the idea that $2^{J/2} \Delta^p X_{J,[2^J t]} \approx X^{(p)}(t) 2^{-Jp}$ for large J. For example, when $\beta = 1$,

$$\widehat{X}_{1,J}(t) = 2^{J/2} X_{J,[2^J t]} + 2^{J/2} \frac{X_{J,[2^J t]+1} - X_{J,[2^J t]}}{2^{-J}} (t - [2^J t]2^{-J}).$$

When $\beta = 0$, we get $\widehat{X}_{0,J}(t) = 2^{J/2} X_{J,[2^J t]}$.

Although intuitive, the approximation $\widehat{X}_{\beta,J}$ in (2.84) may not converge to $X(t)$ at the faster rate $\beta + \alpha$ on compact intervals (see Remark 2.7.1 below). It turns out, though, that a modification of (2.84) does attain that rate. In order to build such approximation, we will make use of two auxiliary results below. For any $x \in \mathbb{R}$, define the function s_x on $\mathbb{Z}$ by

$$s_x(k) = x + k, \quad k \in \mathbb{Z}.$$

Note that $\Delta^p s_0^j = \sum_{k=0}^{p} \binom{p}{k} (-1)^{p-k} k^j$, $j \in \mathbb{N}$ (recall from above that $\Delta^p s_0^j = (\Delta_1^p s_0^j)(0)$).

Lemma 2.7.1. *For any $x \in \mathbb{R}$, $n \in \mathbb{N}$, $j = 0, 1, ..., n$, we have*

$$\Delta^n s_x^j = \begin{cases} 0 & , \text{ if } j < n, \\ n! & , \text{ if } j = n. \end{cases}$$

Proof. The relation (2.7.1) is trivial for $j = 0$ by basic combinatorics. Suppose by induction that it holds for $j-1<n$ and consider the case of $j<n$. Then, with $x=0$,

$$\Delta^n s_0^j = \sum_{k=1}^{n-1} \binom{n}{k} (-1)^{n-k} k^j + \binom{n}{n} (-1)^0 n^j. \tag{2.85}$$

The right-hand side of (2.85) is equal to

$$n \sum_{k=1}^{n-1} \binom{n-1}{k-1} (-1)^{(n-1)-(k-1)} k^{j-1} + n^j = n \sum_{k=0}^{n-2} \binom{n-1}{k} (-1)^{(n-1)-k} (k+1)^{j-1} + n^j$$

$$= n \sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^{(n-1)-k} (k+1)^{j-1} = n \sum_{i=0}^{j-1} \binom{n-1}{i} \left[\sum_{k=0}^{n-1} \binom{n-1}{k} (-1)^{(n-1)-k} k^i \right].$$

By the induction hypothesis, the terms in the brackets above equal 0 since $i \le j-1 < n$. Similarly, one can use induction and the result for $j<n$ to show that $\Delta^n s_0^n = n!$.

For all $x \in \mathbb{R}$, and $j \le n \in \mathbb{N}$, we have

$$\Delta^n s_x^j = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} (x+k)^j = \sum_{i=0}^{j} \binom{j}{i} k^i x^{j-i} \left[\sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} k^i \right].$$

If $j<n$, the term in the brackets above equals 0 when $i \le j$. If $j=n$, the bracketed term equals $n!$ for $i=n$, which concludes the proof. □

Observe from (2.83) that replacing $2^{J/2}\Delta^p X_{J,[2^J t]}$ by $(\Delta^p_{2^{-J}} X)([2^J t]/2^J)$ in the approximation (2.84) makes an error of the desired faster rate $\alpha+\beta$. The next lemma shows that, after suitable correction, $(\Delta^p_{2^{-J}} X)([2^J t]/2^J)$ approximates $X^{(p)}([2^J t]/2^J)$ (and then $X^{(p)}(t)$) at the desired rate $\alpha+\beta$. This correction needs to be taken into account when considering a modification to $\widehat{X}_{\beta,J}$. The modification is considered in the proposition below.

Lemma 2.7.2. *Let $\beta \in \mathbb{N}$, $\alpha \in (0,1)$ and $G \subseteq \mathbb{R}$ be an open interval. If $f: G \to \mathbb{R}$ is a Lipschitz function of order $\beta+\alpha$ in the sense of (2.72), then, for $\mathbb{N} \ni p \le \beta$, $a \in G$, we have*

$$\frac{\Delta_h^p f(a)}{h^p} - f^{(p)}(a) = \sum_{j=1}^{\beta-p} \frac{f^{(p+j)}(a)}{(p+j)!} \, h^j \, \Delta^p s_0^{p+j} + O(h^{\beta+\alpha-p}), \tag{2.86}$$

as $h \to 0$.

Proof. By using Lemma 2.7.1, we can write

$$\Delta_h^p f(a) = \sum_{k=0}^{p} \binom{p}{k} (-1)^{p-k} \left[f(a+kh) - f(a) - \sum_{i=1}^{p} \frac{f^{(i)}(a)}{i!} (kh)^i \right] + f^{(p)}(a) h^p.$$

Thus, by (2.72),

$$\Delta_h^p f(a) - f^{(p)}(a) h^p = \sum_{k=0}^{p} \binom{p}{k} (-1)^{p-k} \left[f(a+kh) - f(a) - \sum_{i=1}^{\beta} \frac{f^{(i)}(a)}{i!} (kh)^i \right]$$

$$+ \sum_{k=0}^{p} \binom{p}{k} (-1)^{p-k} \sum_{j=1}^{\beta-p} \frac{f^{(p+j)}(a)}{(p+j)!} (kh)^{p+j}$$

$$= O(h^{\beta+\alpha}) + \sum_{k=0}^{p} \binom{p}{k} (-1)^{p-k} \sum_{j=1}^{\beta-p} \frac{f^{(p+j)}(a)}{(p+j)!} h^{p+j} \Delta^p s_0^{p+j}.$$

□

Define the approximation function $\widetilde{X}_{\beta,J}(t)$ by

$$\widetilde{X}_{\beta,J}(t) = \widetilde{X}_{(0),J} + \sum_{p=1}^{\beta} \frac{\widetilde{X}_{(p),J}(t)}{p!} \, (t - [2^J t] 2^{-J})^p, \tag{2.87}$$

where

$$\widetilde{X}_{(0),J} := 2^{J/2} X_{J,[2^J t]}, \quad \widetilde{X}_{(\beta),J} := \frac{2^{J/2} \Delta^\beta X_{J,[2^J t]}}{2^{-J\beta}}$$

and

$$\widetilde{X}_{(p),J} := \frac{2^{J/2} \Delta^p X_{J,[2^J t]}}{2^{-J}} + \sum_{j=1}^{\beta-p} \frac{\widetilde{X}_{(p+j),J}}{(p+j)!} \, 2^{-Jj} \, \Delta^p s_0^{p+j}, \quad p = 1, 2, ..., \beta - 1.$$

Proposition 2.7.2. *Under stronger assumptions of Proposition 2.7.1, we have*

$$\sup_{t \in K} |\widetilde{X}_{\beta,J}(t) - X(t)| \le A 2^{-J(\beta+\alpha)} \quad \text{a.s.}, \tag{2.88}$$

where K is a compact interval and A is random variable that does not depend on J.

Proof. If the relation

$$\widetilde{X}_{(p),J} - X^{(p)}([2^J t] 2^{-J}) = O(2^{-J(\beta+\alpha-p)}) \tag{2.89}$$

holds for $p = 0, 1, 2, ..., \beta$, then, by Assumption 6,

$$|\widetilde{X}_{\beta,J}(t) - X(t)| \leq \left|\widetilde{X}_{\beta,J}(t) - X([2^J t]2^{-J}) - \sum_{p=1}^{\beta} \frac{X^{(p)}([2^J t]2^{-J})}{p!} (t - [2^J t]2^{-J})^p\right|$$

$$+\left|X(t) - X([2^J t]2^{-J}) - \sum_{p=1}^{\beta} \frac{X^{(p)}([2^J t]2^{-J})}{p!} (t - [2^J t]2^{-J})^p\right| = O(2^{-J(\beta+\alpha)}),$$

which proves (2.88).

Relation (2.89) holds for $p = 0$ by Proposition 2.7.1. To show (2.89) for $\beta \geq 1$, we argue by backward induction. For $p = \beta$, by Proposition 2.7.1 and Lemma 2.7.2, we have

$$|\widetilde{X}_{(\beta),J} - X^{(\beta)}([2^J t]2^{-J})| \leq \left|\frac{2^{J/2}\Delta^\beta X_{J,[2^J t]}}{2^{-J\beta}} - \frac{\Delta^\beta X([2^J t]2^{-J})}{2^{-J\beta}}\right|$$

$$+\left|\frac{\Delta^\beta X([2^J t]2^{-J})}{2^{-J\beta}} - X^{(\beta)}([2^J t]2^{-J})\right| = O(2^{-J(\beta+\alpha-\beta)}).$$

Assume by induction that (2.89) holds for $p + 1, ..., \beta - 1, \beta$ (with $p \geq 1$). Then, by Proposition 2.7.1 and Lemma 2.7.2, we obtain that

$$|\widetilde{X}_{(p),J} - X^{(p)}([2^J t]2^{-J})| \leq \left|\frac{2^{J/2}\Delta^p X_{J,[2^J t]}}{2^{-Jp}} - \frac{\Delta^p X([2^J t]2^{-J})}{2^{-Jp}}\right|$$

$$+\left|\frac{\Delta^p X([2^J t]2^{-J})}{2^{-Jp}} - X^{(p)}([2^J t]2^{-J}) - \sum_{j=1}^{\beta-p} \frac{X^{(p+j)}([2^J t]2^{-J})}{(p+j)!} 2^{-Jj}\Delta^p s_0^{p+j}\right|$$

$$+\left|\sum_{j=1}^{\beta-p} \frac{X^{(p+j)}([2^J t]2^{-J})}{(p+j)!} 2^{-Jj}\Delta^p s_0^{p+j} - \sum_{j=1}^{\beta-p} \frac{\widetilde{X}_{(p+j),J}}{(p+j)!} 2^{-Jj}\Delta^p s_0^{p+j}\right| = O(2^{-J(\beta+\alpha-p)}).$$

□

Remark 2.7.1. When $\beta = 2$, the approximation $\widetilde{X}_{\beta,J}$ becomes

$$\widetilde{X}_{2,J} = 2^{J/2} X_{J,[2^J t]}$$

$$+2^{J/2}\left(\frac{X_{J,[2^J t]+1} - X_{J,[2^J t]}}{2^{-J}} + \frac{X_{J,[2^J t]+2} - 2X_{J,[2^J t]+1} + X_{J,[2^J t]}}{2\, 2^{-J}}\right)(t - [2^J t]2^{-J})$$

$$+2^{J/2}\frac{X_{J,[2^Jt]+2}-2X_{J,[2^Jt]+1}+X_{J,[2^Jt]}}{(2^{-J})^2}\,(t-[2^Jt]2^{-J})^2. \qquad (2.90)$$

Compare (2.90) with the approximations $\widehat{X}_{2,J}$ given in (2.84). Observe that, if $\widehat{X}_{2,J}$ also converges to X at the rate $2+\alpha$, then

$$O(2^{-J(2+\alpha)}) = \widetilde{X}_{2,J} - \widehat{X}_{2,J}$$

$$= 2^{J/2}\frac{X_{J,[2^Jt]+2}-2X_{J,[2^Jt]+1}+X_{J,[2^Jt]}}{2\,2^{-J}}\,(t-[2^Jt]2^{-J})$$

or, by using (2.83) in Proposition 2.7.1,

$$O(2^{-J(2+\alpha)}) = \frac{X(([2^Jt]+2)2^{-J})-2X(([2^Jt]+1)2^{-J})+X([2^Jt]2^{-J})}{2^{-J}}\,(t-[2^Jt]2^{-J})$$

or, by using Taylor expansions,

$$O(2^{-J(2+\alpha)}) = (2X''(t_1)-X''(t_2))2^{-J}\,(t-[2^Jt]2^{-J}),$$

with $t_1 = t_1(J)$ and $t_2 = t_2(J)$ that are close to t. The last relation may not be satisfied under our assumptions, showing that one cannot expect $\widehat{X}_{2,J}$ to converge to X at the rate $2+\alpha$.

Although the approximations $\widetilde{X}_{\beta,J}$ converge to X at the faster rate $\beta+\alpha$, these approximations do not necessarily have continuous paths. Indeed, it can be easily verified that $\widetilde{X}_{\beta,J}$ is continuous when $\beta = 1,2$ but not so when $\beta = 3$. For a fixed $\beta \geq 2$, it may be desirable to have not only a continuous but also a $C^{\beta-1}$ approximation $\overline{X}_{\beta,J}$. Moreover, in analogy to (2.87), in order to have the faster convergence, we would expect the p-th derivative of the approximation $\overline{X}_{\beta,J}$ at $[2^Jt]2^{-J}$ to approximate the p-th derivative of the process X at t.

We generally found such $C^{\beta-1}$ approximations difficult to construct. One difficulty is the following. As in (2.87), we may seek an approximation $\overline{X}_{\beta,J}$ which is a polynomial of order β on an interval $([2^Jt]2^{-J}, [2^Jt]2^{-J}+1)$. Since $\overline{X}_{\beta,J}$ is globally $C^{\beta-1}$, we would require its derivatives $\overline{X}^{p}_{\beta,J}$, $p = 0,1,...,\beta-1$, to be equal to prescribed values at the endpoints

$[2^J t]2^{-J}$ and $[2^J t]2^{-J}+1$. Requiring this yields 2β equations that a polynomial $\overline{X}_{\beta,J}$ must satisfy. Since a polynomial of order β has only $\beta+1$ coefficients, this is not possible in general. Despite this difficulty, we have found the following general scheme to yield $C^{\beta-1}$ approximations, at least for the first several values of $\beta \geq 2$.

To construct a C^1 approximation $\overline{X}_{2,J}$, we could require first that its derivative

$$2^{-J/2}\overline{X}^{(1)}_{2,J}(t) = 2^{-J/2}\widehat{X}_{1,J}(t) \text{ based on the sequence } \frac{\Delta X_{J,[2^J t]}}{2^{-J}}$$

$$= \frac{\Delta X_{J,[2^J t]}}{2^{-J}} + \frac{1}{2^{-J}}\left(\frac{\Delta X_{J,[2^J t]+1}}{2^{-J}} - \frac{\Delta X_{J,[2^J t]}}{2^{-J}}\right)(t-[2^J t]2^{-J})$$

$$= \frac{\Delta X_{J,[2^J t]}}{2^{-J}} + \frac{\Delta^2 X_{J,[2^J t]}}{(2^{-J})^2}(t-[2^J t]2^{-J}). \tag{2.91}$$

Observe that, by construction using continuous approximation $\widehat{X}_{1,J}$, $\overline{X}^{(1)}_{2,J}$ is continuous. Moreover, $\overline{X}^{(1)}_{2,J}$ approximates $X^{(1)}(t)$, and $\overline{X}^{(2)}_{2,J}$ on the interval $([2^J t]2^{-J}, [2^J t]2^{-J}+1)$ approximates $X^{(2)}(t)$. Integrating (2.91) and requiring it to be continuous yields the following approximation

$$2^{-J/2}\overline{X}_{2,J}(t) = \frac{X_{J,[2^J t]+1} + X_{J,[2^J t]}}{2} + \frac{\Delta X_{J,[2^J t]}}{2^{-J}}(t-[2^J t]2^{-J})$$

$$+\frac{\Delta^2 X_{J,[2^J t]}}{(2^{-J})^2}(t-[2^J t]2^{-J})^2. \tag{2.92}$$

Note that $\overline{X}_{2,J}$ differs from $\widehat{X}_{2,J}$ by the constant term.

Similarly, to construct a C^2 approximation $\overline{X}_{3,J}$, we could require that

$$\overline{X}^{(1)}_{3,J}(t) = \overline{X}_{2,J}(t) \text{ based on the sequence } \frac{\Delta X_{J,[2^J t]}}{2^{-J}}.$$

Integrating the resulting expression and requiring it to be continuous yields

$$\overline{X}_{3,J}(t) = \frac{1}{6}X_{J,[2^J t]+2} + \frac{4}{6}X_{J,[2^J t]+1} + \frac{1}{6}X_{J,[2^J t]} + \frac{1}{2}\frac{\Delta_2 X_{J,[2^J t]}}{2^{-J}}(t-[2^J t]2^{-J})$$

$$+\frac{1}{2}\frac{\Delta^2 X_{J,[2^J t]}}{(2^{-J})^2}(t-[2^J t]2^{-J})^2 + \frac{1}{6}\frac{\Delta^3 X_{J,[2^J t]}}{(2^{-J})^3}(t-[2^J t]2^{-J})^3 \tag{2.93}$$

(the subindex 2 in $\Delta_2 X_{J,[2^J t]}$ is not a typo). The approximation (2.93) is C^2 and its derivatives of orders $p = 0, 1, 2, 3$ approximate those of the process X.

We expect that the above scheme yields $C^{\beta-1}$ approximations $\overline{X}_{\beta,J}$ for any $\beta \geq 2$. However, as explained in Remark 2.7.2 below, we cannot expect these approximations to converge at the faster rate $\beta + \alpha$. This is perhaps not surprising, because the discontinuous approximations in (2.87) are already nontrivial.

Remark 2.7.2. One cannot expect the approximation $\overline{X}_{2,J}$ in (2.93) to converge to X at the faster rate $2 + \alpha$. Indeed, if this rate were achieved, we would have (see (2.90))

$$O(2^{-J(2+\alpha)}) = \widetilde{X}_{2,J}(t) - \overline{X}_{2,J}(t)$$

$$= X_{J,[2^J t]+1} - X_{J,[2^J t]} - \frac{X_{J,[2^J t]+2} - 2X_{J,[2^J t]+1} + X_{J,[2^J t]}}{2\, 2^{-J}} \,(t - [2^J t]2^{-J})$$

or, by using (2.82),

$$O(2^{-J(2+\alpha)}) = X(([2^J t] + 1)2^{-J}) - X([2^J t]2^{-J})$$

$$- \frac{X(([2^J t] + 2)2^{-J}) - 2X(([2^J t] + 1)2^{-J}) + X([2^J t]2^{-J})}{2^{-J}} \,(t - [2^J t]2^{-J})$$

or, by Taylor expansions,

$$O(2^{-J(2+\alpha)}) = X'([2^J t]2^{-J})2^{-J} + \frac{1}{2}X''(t_1)2^{-2J} - X''(t_2)2^{-J}\,(t - [2^J t]2^{-J}),$$

with $t_1 = t_1(J)$ and $t_2 = t_2(J)$ close to t, or, by expanding $X'([2^J t]2^{-J})$ further,

$$O(2^{-J(2+\alpha)}) = X'(t)2^{-J} + \frac{1}{2}X''(t_1)2^{-2J} - X''(t_2)2^{-J}\,(t - [2^J t]2^{-J}).$$

This relation may not be satisfied under our assumptions on X.

2.8 Simulation: the case of the OU process

We will illustrate here how the results of Sections 2.6 and 2.7 can be used to simulate a stationary process X. We consider only the case of the OU process in Example 2.4.1.

Recall from that example that the discrete approximations taken for the OU process are

$$\widehat{g}_J(x) = 2^{-J/2}\sigma\sqrt{\frac{1-e^{-2\lambda 2^{-J}}}{2\lambda}}\,(1-e^{-\lambda 2^{-J}}e^{-ix})^{-1} \tag{2.94}$$

and the corresponding discrete random approximations X_J are suitable AR(1) time series in (2.34). With the choice (2.94) of approximations, observe that the filters u_j and v_j used in reconstruction (2.60) become

$$\widehat{u}_j(x) = \frac{2^{-1/2}}{\sqrt{1+e^{-2\lambda 2^{-(j+1)}}}}(1+e^{-\lambda 2^{-(j+1)}}e^{-ix})\widehat{u}(x), \tag{2.95}$$

$$\widehat{v}_j(x) = 2^{-(j+1)/2}\sigma\sqrt{\frac{1-e^{-2\lambda 2^{-(j+1)}}}{2\lambda}}\,(1-e^{-\lambda 2^{-(j+1)}}e^{-ix})^{-1}\widehat{v}(x). \tag{2.96}$$

Suppose one wants to simulate the OU process on the interval $[0,1]$. The idea is to begin by generating a discrete approximation X_0 at scale 2^0. This step is easy as X_0 is an AR(1) time series. Then, substituting X_0 into (2.60), one may get the approximation X_1, and continuing recursively from X_1 now, the approximation X_J for arbitrary fixed $J \geq 1$. Note that applying (2.60) recursively each time essentially involves just simulating independent $\mathcal{N}(0,1)$ random variables and computing filters u_j and v_j. Proposition 2.7.1 ensures that the properly normalized X_J approximate the OU process uniformly over $[0,1]$ and exponentially fast in J.

We illustrate this in Figure 2.1 for the OU process with $\lambda = 1,\ \sigma = 1$. The plot on the left depicts the consecutive approximations X_j from X_0 at scale 2^0 to X_J at the finest scale 2^{-J} with $J = 11$. In the right plot, we present the sup-differences between consecutive approximations X_{j-1} and X_j, $j = 2,\ldots,11$, on the log scale. The decay in that plot confirms that normalized approximations X_J converge to the OU process exponentially fast in J.

Several comments should be made on how approximations X_j are obtained in Figure 1. Though theoretically unjustified, we use not Meyer but the celebrated Daubechies CMFs with $N = 8$ zero moments. The advantage of these CMFs is that they have finite length (equal to $2N$). In particular, the filters u_j in (2.96) are then also finite (of length $2N+2$)

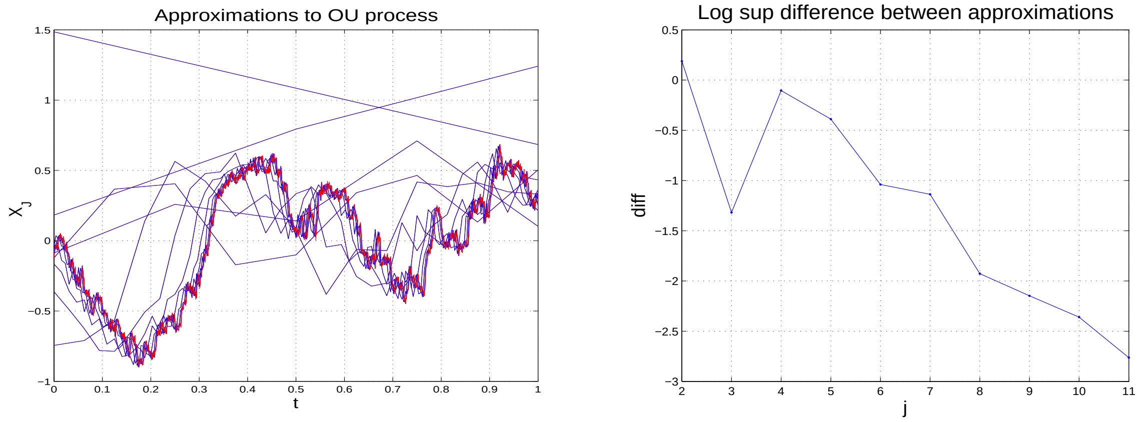

Figure 2.1: Approximations X_J and the logarithms of their sup differences.

for any j. The filters v_j, however, are not finite and are truncated in practice, disregarding those elements that are smaller than a prescribed level $\delta = 10^{-10}$. Let us also note that applications of (2.60) involve more elements of X_j than those plotted in Figure 1. This is achieved by taking the initial approximation X_0 of suitable length. Some indication on how this is done can be seen from the analogous simulation of fractional Brownian motion in Pipiras (2005).

Finally, let us indicate another interesting feature of the above simulation. Focus on the filters v_j defined by (2.96). They have infinite length and are truncated in practice. It may seem from the definition (2.96) that v_j have to be taken of very long length as j increases because the elements of the filter

$$(1 - e^{-\lambda 2^{-(j+1)}} e^{-ix})^{-1} = \sum_{k=0}^{\infty} e^{-\lambda 2^{-(j+1)} k} e^{-ixk}$$

decay extremely slowly for larger j. In fact, the opposite turns out to be true. As j increases, the filters v_j can essentially be taken of finite length $2N - 2$, and things get even better for larger j in a way!

To explain why this happens, recall (e.g. Mallat (1998), Theorem 7.4) that N zero moments translates into the factorization

$$\widehat{v}(x) = (1 - e^{-ix})^N \widehat{v}_{0,N}(x), \tag{2.97}$$

where, in the case of Daubechies CMF v, the filters $v_{0,N}$ have also finite length. An explanation follows by observing that

$$\frac{1-e^{-ix}}{1-e^{-\lambda 2^{-(j+1)}}e^{-ix}} = \sum_{k=0}^{\infty} a_k^{(j)} e^{-ixk} \to 1,$$

or $a_0^{(j)} \to 1$, $a_0^{(k)} \to 0$, $k \geq 1$, as $j \to \infty$. More precisely,

$$\frac{1-e^{-ix}}{1-e^{-\lambda 2^{-(j+1)}}e^{-ix}} - 1 = \frac{-e^{-ix}(1-e^{-\lambda 2^{-(j+1)}})}{1-e^{-\lambda 2^{-(j+1)}}e^{-ix}} = -(1-e^{-\lambda 2^{-(j+1)}}) \sum_{k=1}^{\infty} e^{-\lambda 2^{-(j+1)}(k-1)} e^{-ixk},$$

so that the elements $a_k^{(j)}$, $k \geq 1$, are bounded by $1-e^{-\lambda 2^{-(j+1)}} \leq \lambda 2^{-(j+1)} \to 0$, as $j \to \infty$.

2.9 Riesz bases

Both the decomposition (2.44) of Zhang and Walter and its modification (2.52) appear to be ordinary decompositions of signals X into corresponding "bases". These "bases" are not orthogonal. We will show, however, that under quite general assumptions *both* of them are Riesz bases of $L^2(\mathbb{R})$. Several remarks are in order at this point.

Remarks

1. Riesz bases (or frames, more generally) are often desirable because of numerical stabilities associated with them (Daubechies (1992)).

2. One may ask why the space $L^2(\mathbb{R})$ is taken here whereas stationary processes X do not have their sample paths in $L^2(\mathbb{R})$. One reason is that a basis is often Riesz not only for $L^2(\mathbb{R})$ but also for other spaces. Hence, proving it for $L^2(\mathbb{R})$ is a good indication of a Riesz basis in other spaces. (We avoided proving that they are Riesz bases in suitable function spaces associated with X and opted for a direct proof for simplicity.)

3. Our result extends that of Meyer *et al.* (1999), who showed that the particular wavelet bases used for fractional Brownian motion, analogous to (2.52), are Riesz. It also seems that Zhang and Walter have already asked whether their bases in (2.44) are Riesz but were not able to provide an affirmative answer (Zhang and Walter (1994), Walter and Shen (2001)). Our result thus provides a partial answer to their open question.

4. As indicated above, bases in both (2.44) and (2.52) turn out to be Riesz. This is perhaps not surprising as wavelet functions are the same in both (2.44) and (2.52). From the perspective of Riesz bases, the modification (2.52) therefore is not different from (2.44).

We will focus on the bases associated with (2.52), and then discuss those associated with (2.44). Recall also that a set $\{e_l\}_{l\in\mathbb{Z}}$ is a Riesz basis of a Hilbert space $\mathcal{H}$ if

(i) span$\{e_l\}_{l\in\mathbb{Z}}$ is dense in $\mathcal{H}$;

(ii) there are constants $C_2 \geq C_1 > 0$ such that

$$C_1\Big(\sum_{l\in\mathbb{Z}}|a_l|^2\Big)^{1/2} \leq \Big|\Big|\sum_{l\in\mathbb{Z}} a_l e_l\Big|\Big|_{\mathcal{H}} \leq C_2\Big(\sum_{l\in\mathbb{Z}}|a_l|^2\Big)^{1/2}, \tag{2.98}$$

for all sequences $\{a_l\}_{l\in\mathbb{Z}} \in l^2(\mathbb{Z})$.

We will assume some additional regularity conditions on $\widehat{g}(x)$, namely,

Assumption 8: Suppose that

(a) $|\widehat{g}(x)| > 0$ for all $x \in \mathbb{R}$;

(b) $\widehat{g} \in C(\mathbb{R})$;

(c) for any $\epsilon > 0$, there are $d \in \mathbb{R}$ and constants $C_4 \geq C_3 > 0$ such that

$$C_3|x|^{-d} \leq |\widehat{g}(x)| \leq C_4|x|^{-d}, \quad \text{for } |x| > \epsilon. \tag{2.99}$$

For instance, in the case of the Ornstein-Uhlenbeck process, $|\widehat{g}(x)|^2 = (1+x^2)^{-1}$ satisfies (2.99) with $d = 1$. But (2.99) also does not cover a seemingly simple case where $\widehat{g}(x) = e^{-x^2}$.

We consider properly normalized functions of (2.52), namely, the family of functions

$$\{\Phi^0(t-k), \eta^j(t-2^{-j}k) : k \in \mathbb{Z}, j \geq 0\} \tag{2.100}$$

as well as their biorthogonal counterparts

$$\{\Phi_0(t-k), \eta_j(t-2^{-j}k) : k \in \mathbb{Z}, j \geq 0\}, \tag{2.101}$$

where

$$\eta^j(t-2^{-j}k) = 2^{jd}\Psi^j(t-2^{-j}k), \quad \eta_j(t-2^{-j}k) = 2^{-jd}\Psi_j(t-2^{-j}k) \tag{2.102}$$

(the exponent d in the normalization (2.102) is the same as in the condition (2.99)). To simplify the exposition, we will focus on (2.100), but all the upcoming arguments can be adapted for (2.101).

The proof that (2.100) is a Riesz basis of $L^2(\mathbb{R})$ uses some of the arguments in Meyer *et al.* (1999), and whenever convenient we will refer the reader to their original paper. We will need the following two lemmas.

Lemma 2.9.1. *Under Assumption 3, the family* $\{\Phi^0(t-k)\}_{k\in\mathbb{Z}}$ *is a Riesz basis of its closed linear span* V^0 *in* $L^2(\mathbb{R})$.

Proof. Condition (i) of the definition of a Riesz basis is immediately satisfied. As for (ii), we have that, for any $\{a_k\}_{k\in\mathbb{Z}} \in l^2(\mathbb{Z})$,

$$\Big\| \sum_k a_k \widehat{\Phi}^0(t-k)\Big\|^2_{L^2(\mathbb{R})} = \frac{1}{2\pi}\int_{\mathbb{R}} \Big| \sum_k a_k e^{-ikx} G_0(x)\widehat{\phi}(x)\Big|^2 dx,$$

and thus

$$\begin{aligned} 0 < C \inf_{x\in[-\pi,\pi]} |G_0(x)|^2 \ \sum_k |a_k|^2 &\leq \Big\| \sum_k a_k \widehat{\Phi}^0(t-k)\Big\|^2_{L^2(\mathbb{R})} \\ &\leq 2 \sup_{x\in[-4\pi/3,4\pi/3]} |G_0(x)|^2 \ \sum_k |a_k|^2, \end{aligned}$$

for some constant C, where the infimum and the supremum above are finite by Assumption 3 on G_0. □

Lemma 2.9.2. *Under Assumptions 3 and 8, for* $j \in \mathbb{Z}$ *and* $\{a_k\}_{k\in\mathbb{Z}}$, $\{b_k\}_{k\in\mathbb{Z}} \in l^2(\mathbb{Z})$, *there exists a unique sequence* $\{c_k\}_{k\in\mathbb{Z}} \in l^2(\mathbb{Z})$ *such that*

$$\sum_k a_k \Phi^j(t-2^{-j}k) + \sum_k b_k \eta^j(t-2^{-j}k) = \sum_k c_k \Phi^{j+1}(t-2^{-(j+1)}k). \tag{2.103}$$

Moreover, the induced map from $l^2(\mathbb{Z}) \times l^2(\mathbb{Z})$ *to* $l^2(\mathbb{Z})$ *is an isomorphism.*

Proof. We adapt here the proof of Lemma 5.3 in Meyer *et al.* (1999). In terms of Fourier transforms, (2.103) may be expressed as

$$\widehat{a}\Big(\frac{x}{2^j}\Big)\widehat{\Phi}^j(x) + \widehat{b}\Big(\frac{x}{2^j}\Big)\widehat{\eta}^j(x) = \widehat{c}\Big(\frac{x}{2^{j+1}}\Big)\widehat{\Phi}^{j+1}(x), \tag{2.104}$$

where $\widehat{a}$, $\widehat{b}$ and $\widehat{c}$ are, respectively, the 2π-periodic extensions of the discrete Fourier transforms of $\{a_k\}$, $\{b_k\}$ and $\{c_k\}$. Set

$$\widehat{\Phi}^j(x) = U^j(x)\widehat{\Phi}^{j+1}(x), \quad \widehat{\eta}^j(x) = V^j(x)\widehat{\Phi}^{j+1}(x),$$

where, for $x \in (-\pi, \pi)$,

$$U^j(x) = \frac{\widehat{g}_{j+1}(2^{-(j+1)}x)}{\widehat{g}_j(2^{-j}x)}\widehat{u}(2^{-(j+1)}x) \quad \text{and} \quad V^j(x) = 2^{jd}\widehat{g}_{j+1}(2^{-(j+1)}x)\widehat{v}(2^{-(j+1)}x)$$

with the Meyer CMFs u and v. Then, the relation (2.104) may be rewritten as

$$\widehat{a}(2x)U^j(2^{j+1}x) + \widehat{b}(2x)V^j(2^{j+1}x) = \widehat{c}(x),$$

which implies that $\widehat{c}$ can be obtained from $\widehat{a}$ and $\widehat{b}$. Moreover, by Assumption 3, $U^j(2^{j+1}x) = G_{j+1}(x)G_j(2x)^{-1}\widehat{u}(x)$ and $V^j(2^{j+1}x) = 2^{jd}G_{j+1}(x)\widehat{g}(2^{j+1}x)\widehat{v}(x)$ are $L^2(-\pi,\pi)$ functions, and thus so is $\widehat{c}$.

Conversely, consider the family of matrices $\{M(x)\}_{x\in(-\pi,\pi)}$, where

$$M(x) = \begin{pmatrix} U^j(2^{j+1}x) & V^j(2^{j+1}x) \\ \\ U^j(2^{j+1}(x+\pi)) & V^j(2^{j+1}(x+\pi)) \end{pmatrix}.$$

These matrices are invertible, since

$$\begin{aligned} \det[M(x)] &= 2^{jd}\,\frac{\widehat{g}_{j+1}(x)\widehat{g}_{j+1}(x+\pi)}{\widehat{g}_j(2x)}\big(\widehat{u}(x)\widehat{v}(x+\pi) - \widehat{u}(x+\pi)\widehat{v}(x)\big) && (2.105)\\ &= 2^{jd}\,\frac{G_{j+1}(x)G_{j+1}(x+\pi)}{G_j(2x)}\widehat{g}(2^{j+1}(x+\pi))(-2e^{-ix}) && (2.106) \end{aligned}$$

is bounded away from zero by Assumption 3 on G_j and Assumption 8 on $\widehat{g}$. Thus, $\widehat{a}$ and $\widehat{b}$ can also be recovered from $\widehat{c}$ by using

$$\begin{pmatrix} \widehat{a}(2x) \\ \widehat{b}(2x) \end{pmatrix} = \begin{pmatrix} \widehat{a}(2(x+\pi)) \\ \widehat{b}(2(x+\pi)) \end{pmatrix} = M(x)^{-1} \begin{pmatrix} \widehat{c}(x) \\ \widehat{c}(x+\pi) \end{pmatrix}.$$

Equivalently, for example,

$$\widehat{a}(2x) = G_j(2x)(-2e^{-ix})\Big(G_{j+1}(x)^{-1}\widehat{v}(x+\pi)\widehat{c}(x) + G_{j+1}(x+\pi)^{-1}\widehat{v}(x)\widehat{c}(x+\pi)\Big)$$

and thus by Assumptions 3 and 8, $\widehat{c} \in L^2(-\pi,\pi)$, we get that $\widehat{a} \in L^2(-\pi,\pi)$. Similarly, $\widehat{b} \in L^2(-\pi,\pi)$. $\square$

The following proposition is the main result of this section.

Proposition 2.9.1. *Under Assumptions 3 and 8, the family (2.100) is a Riesz basis of* $L^2(\mathbb{R})$.

Proof. For $j \in \mathbb{Z}$, denote by V^j the closure of $\operatorname{span}\{\Phi^j(t-2^{-j}k), k \in \mathbb{Z}\}$ and by W^j the closure of $\operatorname{span}\{\eta^j(t-2^{-j}k), k \in \mathbb{Z}\}$. By Lemma 2.9.2,

$$V^j \oplus W^j = V^{j+1}, \tag{2.107}$$

which is a direct but not orthogonal sum. By using the fact that $\widehat{\Phi}^j(x) \neq 0$ for $|x| \leq \frac{2\pi}{3}$ and $\widehat{\Phi}^j(x) = 0$ for $|x| \geq \frac{4\pi}{3}$, and by proceeding exactly as in Meyer *et al.* (1999), Lemma 5.3, we have that

$$V^j \subseteq V^{j+1}, \quad \bigcap_{j=0}^{\infty} V^j = V^0, \quad \text{and} \quad \bigcup_{j=0}^{\infty} V^j \text{ is dense in } L^2(\mathbb{R}). \tag{2.108}$$

Therefore, from (2.107) and (2.108), the space $V^0 \bigoplus_{j\geq 0} W^j$ is dense in $L^2(\mathbb{R})$, which gives us part (i) of the definition of a Riesz basis.

Suppose at the moment that there exist constants $C_2 \geq C_1 > 0$ such that

$$C_1\Big(\sum_k \sum_{j\geq 0} b_{j,k}^2\Big)^{1/2} \leq \Big|\Big| \sum_k \sum_{j\geq 0} b_{j,k}\eta^j(t-2^{-j}k)\Big|\Big| \leq C_2\Big(\sum_k \sum_{j\geq 0} b_{j,k}^2\Big)^{1/2} \qquad (2.109)$$

for any sequence $\{b_{j,k}\} \in l^2(\mathbb{Z})$, where for simplicity we write $\|\cdot\|$ instead of $\|\cdot\|_{L^2(\mathbb{R})}$. Then, since the family $\{\Phi^0(t-k), k \in \mathbb{Z}\}$ is a Riesz basis of V^0 by Lemma 2.9.1, we have

$$\Big|\Big| \sum_k a_k \Phi^0(t-k) + \sum_k \sum_{j\geq 0} b_{j,k}\eta^j(t-2^{-j}k)\Big|\Big|$$

$$\leq \sqrt{2}\Big(\Big|\Big| \sum_k a_k \Phi^0(t-k)\Big|\Big|^2 + \Big|\Big|\sum_k \sum_{j\geq 0} b_{j,k}\eta^j(t-2^{-j}k)\Big|\Big|^2\Big)^{1/2} \leq C\,\Big(\sum_k a_k^2 + \sum_k \sum_{j\geq 0} b_{j,k}^2\Big)^{1/2},$$

for some constant C, which establishes the R.H.S. inequality of (2.98). The L.H.S. inequality of (2.98) may be shown in the following way. As proved in Zhang and Walter (1994), Lemma 1, $\{\theta^0(t-k), k \in \mathbb{Z}\}$ with θ^0 in (2.4) is a Riesz basis of the space U^0 it generates. Moreover, the functions $\{\theta^0(t-k), \theta_0(t-k), \eta^j(t-2^{-j}k), \eta_j(t-2^{-j}k), k \in \mathbb{Z}, j \geq 0\}$ satisfy the relations

$$\int_{\mathbb{R}} \theta^0(t-k)\theta_0(t-k')dt = \delta_{\{k=k'\}}, \quad \int_{\mathbb{R}} \eta^j(t-2^{-j}k)\eta_{j'}(t-2^{-j'}k')dt = \delta_{\{j=j'\}}\delta_{\{k=k'\}},$$

$$\int_{\mathbb{R}} \theta^0(t-k)\eta_j(t-2^{-j}k')dt = 0 \quad \text{and} \quad \int_{\mathbb{R}} \eta^j(t-2^{-j}k)\theta_0(t-k')dt = 0, \quad j \geq 0.$$

Then, for any sequences $\{a_k\}, \{b_{j,k}\} \in l^2(\mathbb{Z})$, we can write

$$\sum_k a_k^2 + \sum_k \sum_{j\geq 0} b_{j,k}^2 = \int_{\mathbb{R}} \Bigg(\sum_k a_k\theta^0(t-k) + \sum_k \sum_{j\geq 0} b_{j,k}\eta^j(t-2^{-j}k)\Bigg)\cdot$$

$$\cdot\Bigg(\sum_k a_k\theta_0(t-k) + \sum_k \sum_{j\geq 0} b_{j,k}\eta_j(t-2^{-j}k)\Bigg)dt$$

$$\leq \Big|\Big|\sum_k a_k\theta^0(t-k) + \sum_k \sum_{j\geq 0} b_{j,k}\eta^j(t-2^{-j}k)\Big|\Big| \cdot \Big|\Big|\sum_k a_k\theta_0(t-k) + \sum_k \sum_{j\geq 0} b_{j,k}\eta_j(t-2^{-j}k)\Big|\Big|$$

$$\leq C\Big|\Big|\sum_k a_k\theta^0(t-k) + \sum_k \sum_{j\geq 0} b_{j,k}\eta^j(t-2^{-j}k)\Big|\Big|\Big(\sum_k a_k^2 + \sum_k \sum_{j\geq 0} b_{j,k}^2\Big)^{1/2}$$

for some constant $C > 0$, and hence

$$\frac{1}{C}\Big(\sum_k a_k^2 + \sum_k \sum_{j\geq 0} b_{j,k}^2\Big)^{1/2} \leq \Big|\Big|\sum_k a_k \theta^0(t-k) + \sum_k \sum_{j\geq 0} b_{j,k}\eta^j(t-2^{-j}k)\Big|\Big|. \qquad (2.110)$$

Consider now a sequence $\{r_k\} \in l^2(\mathbb{Z})$ and define

$$\widehat{a}(x) = \widehat{r}(x)\widehat{g_0}^{-1}(x), \quad x \in (-\pi, \pi).$$

Since

$$||\widehat{a}||_{L^2(-\pi,\pi)} = ||\widehat{r}\,\widehat{g_0}^{-1}||_{L^2(-\pi,\pi)} = ||\widehat{r}\,G_0^{-1}\,\widehat{g}^{-1}||_{L^2(-\pi,\pi)},$$

Assumptions 3 and 8 imply that there exist constants C and C' such that

$$C||\widehat{r}||_{L^2(-\pi,\pi)} \leq ||\widehat{a}||_{L^2(-\pi,\pi)} \leq C'||\widehat{r}||_{L^2(-\pi,\pi)} \qquad (2.111)$$

(in particular, the R.H.S. inequality shows that the corresponding sequence $\{a_k\}$ is in $l^2(\mathbb{Z})$). Consider now the extensions of $\widehat{a}$, $\widehat{r}$ and $\widehat{g}_0$ to $\mathbb{R}$ by 2π-periodicity. From the equality $\widehat{a} = \widehat{r}\,\widehat{g_0}^{-1}$ and Assumption 8, we get $\widehat{r}\,\widehat{\Phi}^0 = \widehat{a}\,\widehat{\theta}^0$, and thus

$$\sum_k r_k \Phi^0(t-k) = \sum_k a_k \theta^0(t-k), \qquad (2.112)$$

where the above equality is in the $L^2(\mathbb{R})$ sense. So, from (2.110), (2.111) and (2.112), we have

$$\Big|\Big|\sum_k r_k \Phi^0(t-k) + \sum_k \sum_{j\geq 0} b_{j,k}\eta^j(t-2^{-j}k)\Big|\Big| \geq C\Big(\sum_k r_k^2 + \sum_k \sum_{j\geq 0} b_{j,k}^2\Big)$$

for some constant $C > 0$, and thus we have established (2.98).

It remains to prove (2.109). Observe that

$$\Big|\Big|\sum_k \sum_{j\geq 0} a_{j,k}\eta^j(t-2^{-j}k)\Big|\Big|^2 = \frac{1}{2\pi}\int_{\mathbb{R}}\Big|\sum_k \sum_{j\geq 0} a_{j,k}2^{-jd}e^{-ik2^{-j}x}2^{-j/2}\widehat{\psi}(2^{-j}x)\Big|^2|\widehat{g}(x)|^2dx,$$

which, by using (2.99), can be bounded from above and below (up to a constant) by

$$\frac{1}{2\pi}\int_{\mathbb{R}}\Big|\sum_k\sum_{j\geq 0}a_{j,k}2^{-jd}e^{-ik2^{-j}x}2^{-j/2}\widehat{\psi}(2^{-j}x)\Big|^2|(ix)^{-d}|^2dx$$

$$=\frac{1}{2\pi}\int_{\mathbb{R}}\Big|\sum_k\sum_{j\geq 0}a_{j,k}e^{-ik2^{-j}x}2^{-j/2}\widehat{\psi}_{-d}(2^{-j}x)\Big|^2dx=\Big|\Big|\sum_k\sum_{j\geq 0}a_{j,k}2^{j/2}\psi_{-d}(2^jt-k)\Big|\Big|^2,$$

where $\widehat{\psi}_{-d}(x)=(ix)^{-d}\widehat{\psi}(x)$. Then, (2.109) follows from the relation (5.9) in Meyer *et al.* (1999). □

Proposition 2.9.2. *Under Assumptions 3 and 8, the family*

$$\{\theta^0(t-k),\eta^j(t-2^{-j}k),k\in\mathbb{Z},j\geq 0\}$$

is a Riesz basis of $L^2(\mathbb{R})$.

Proof. In the proof of Proposition (2.9.1), we have already established part *(ii)* of the definition of Riesz basis. Part *(i)* is given in Zhang and Walter (1994), Lemmas 1-3. □

CHAPTER 3

Adaptive wavelet decompositions of stationary time series

3.1 Introduction

Wavelet methods generally refer to an array of concepts, ideas and techniques that are used in Signal Processing, Pure Mathematics, Theoretical Physics, and many other areas. Initially developed under various names and by different research communities, these methods started to converge in the 1980's producing a genuine revolution in their understanding, use and applications (Daubechies (1992), Mallat (1998), Akansu and Haddad (2001)). These developments were also greatly intertwined with those in Statistics where wavelet shrinkage of Donoho and Johnstone (1994, 1995) and others has become commonplace in problems of denoising.

Time Series Analysis, viewed rather as a subdiscipline of Statistics than a part of Signal Processing, has benefitted from wavelet methods as well (see a nice monograph on the subject by Percival and Walden (2000)). Despite a lengthy wavelet theory of treating time series as general signals, truly successful applications of wavelets oriented to Time Series Analysis are not many. Several studies examine the wavelet variance of stationary or stationary increments time series (Section 8 in Percival and Walden (2000)). Wavelets also proved useful to analyze and synthesize long memory time series (Section 9 in Percival and Walden (2000), as well as Abry *et al.* (2003), Pipiras (2005), Moulines *et al.* (2006)) and in connection to unit roots (Fan and Gençay (2006)). Other applications but in continuous time, concern locally stationary time series (Mallat *et al.* (1998), Nason *et al.* (2000)), multifractal processes (Ossiander and Waymire (2000), Resnick *et al.* (2003), Jaffard *et al.*

(2005)). Wavelet analysis of quite general stationary and nonstationary random processes can be found in Cambanis and Masry (1994), Cambanis and Houdré (1995), Krim and Pesquet (1995), Averkamp and Houdré (1998).

An appealing property when using wavelets in Time Series Analysis is the decorrelation property of detail (wavelet) coefficients. Though this fact has by now become an integral part of the "folklore" (and can be formalized to some degree), there are not too many statistical studies exploring it in depth. The most studied is probably the case of long memory time series. See, for example, Dijkerman and Mazumdar (1994), Craigmile and Percival (2005). But even this case, as seen from these references, is not quite simple. A related difficulty with decorrelation is that dependence, though weak(er), is still present and needs to be taken into account in rigorous studies. For example, for a continuous-time stationary process $\{X(t)\}_{t\in\mathbb{R}}$ and orthogonal wavelets $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$, the correlation structure of detail coefficients $d_{j,k} = \int_{\mathbb{R}} X(t)\psi_{j,k}(t)dt$ can be expressed (under mild assumptions) as

$$Ed_{j,k}d_{j',k'} = \int_{\mathbb{R}}\int_{\mathbb{R}} R(t-s)\psi_{j,k}(t)\psi_{j',k'}(t)dtds = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{R}(x)\widehat{\psi}_{j,k}(x)\overline{\widehat{\psi}_{j',k'}(x)}dx,$$

where $R(u) = EX(u)X(0)$ is the autocovariance function and $\widehat{f}(x) = \int_{\mathbb{R}} e^{-iux}f(u)du$ is the Fourier transform of f. Dealing with such covariance structures exactly is generally quite difficult and hence one often opts for assuming complete decorrelation (see, for example, Veitch and Abry (1999)).

In this chapter, we introduce and examine here particular wavelet-based decompositions of time series where detail coefficients are uncorrelated. We focus on stationary times series in discrete time. As in Chapter 2, the resulting decompositions will generally be called *Adaptive Wavelet Decompositions* (*AWD*, in short). The adaptiveness refers to the fact that the wavelet basis (or associated filters) is chosen based on the correlation structure of a time series. In particular, we suppose in this work that the correlation structure of a time series is *known*. This is also reflected in our applications, namely, Maximum Likelihood Estimation (MLE) and Simulation based on AWD. In MLE, the known correlation structure is that of a fitted time series model. Knowing the correlation structure, however, may be too restrictive for other applications.

MLE, in particular, has greatly motivated this chapter. Several authors have previously considered wavelet-based MLE for stationary or stationary increment time series (Section 9 in Percival and Walden (2000), Jensen (1999), Moulines *et al.* (2006)). These MLE (except Moulines *et al.* (2006)) use orthogonal wavelet decompositions, and are approximate in the sense that a complete decorrelation of detail coefficients is assumed, and the variance of detail coefficients at a scale (octave) is taken approximate. We sought to provide a wavelet-based MLE which removes these assumptions or such that

- detail coefficients are decorrelated,
- their variance is taken exact

and also, as in the previous cases, MLE such that it is

- practical to implement,
- computationally efficient,
- not affected by polynomial trends.

MLE based on AWD is a *step toward* obtaining such MLE. It is not totally satisfactory yet because dealing with polynomial trends and some types of stationary time series presents difficulties. (Difficulties with polynomial trends result from the boundary effect when applying AWD to finite data.)

The idea of seeking particular wavelet or other bases with uncorrelated coefficients is obviously not new. The classical, non-wavelet example is that of the Karhunen-Loève (KL) bases, possessing other optimal properties as well. But except special cases, the KL bases are not found explicitly and they do not annihilate polynomial trends. The Signal Processing literature offers a number of alternative decompositions in both wavelet (subband) and other contexts. It is typically assumed that all coefficients in these decompositions are uncorrelated because this is generally considered a necessary condition for coding optimality. (With uncorrelated coefficients, coding gain is no longer possible.) See, for example, Vaidyanathan and Akkarakaran (2001) and the references therein. Similar decompositions oriented to Statistics and Probability, and in continuous time can also be found in Zhang and Walter (1994), Benassi and Jaffard (1994), Donoho (1995), Ruiz-Medina *et al.* (2003), Meyer *et al.* (1999).

As in Chapter 2, AWD considered here have uncorrelated detail coefficients but also allow approximation coefficients to be correlated. This extension appears to be particularly relevant in at least two situations of interest, namely,

1. long memory,
2. near unit roots.

It is quite intriguing that these are exactly the two situations where orthogonal wavelet decompositions were found particularly useful (see the discussion with the references above). Correlated approximation coefficients allow, in particular, to have associated low and high pass filters (which can be thought of as AWD basis in discrete time) of practically small length. The number of zero moments of the underlying orthogonal wavelet basis plays here a fundamental role. Having small filter length is important at the boundary (border) when dealing with finite data. The gain in length is minimal, if any, in other situations that we know of (explaining perhaps why AWD were not considered earlier, since the above situations have gained increased attention fairly recently). The extension provided by AWD is also interesting for several other reasons discussed below. In its approach, this study is also closest to our parallel work on AWD in continuous time in Chapter 2. Despite some similarities, however, the focus and contents of this work are very different from those in Chapter 2.

Another conspicuous example of representations with uncorrelated coefficients are spectral (Fourier) representations. We were also motivated by the question of what their appropriate counterparts in the "wavelet domain" are. AWD introduced here offer one such possibility.

The rest of the chapter is organized as follows. In Section 3.2, we gather some basic notions and facts on time series and wavelets that will be used throughout the chapter. In Section 3.3, we introduce and examine Adaptive Wavelet Decompositions (AWD) of stationary time series. Examples are considered in Section 3.4. Applications of AWD and proofs can be found in Sections 3.5 and 3.6, respectively.

3.2 Preliminaries on time series and wavelets

We focus throughout on stationary time series $X = \{X_n\}_{n\in\mathbb{Z}}$ in discrete time. Stationarity refers to the 2nd order (wide-sense) stationarity, that is, the case when, for any $h \in \mathbb{Z}$,

$$EX_{k+h}X_h = EX_kX_0 =: r(k), \quad k \in \mathbb{Z}, \tag{3.1}$$

where r is the autocovariance function. We suppose, in addition, that a time series X is *Gaussian*. (In this case, decorrelation is equivalent to independence.) This assumption is not restrictive. Since the law of a Gaussian time series is determined by second moments, our arguments can be based only on the second moment considerations. After removing Gaussianity, the same arguments then apply to 2nd order stationary time series. Most of our applications, however, assume Gaussianity.

We will also work only with linear time series

$$X_n = \sum_{k=-\infty}^{\infty} a_k\epsilon_{n-k} = (a * \epsilon)_n, \quad n \in \mathbb{Z}, \tag{3.2}$$

where $a = \{a_k\} \in l^2(\mathbb{Z})$ and $*$ denotes the usual convolution. In the Gaussian case, $\epsilon = \{\epsilon_n\}$ are independent, $\mathcal{N}(0,1)$ random variables. We will refer to such ϵ as a *Gaussian white noise* (sequence). One of the main tools we will use is the spectral representation of X in (3.2) (see e.g. Brockwell and Davis (1991)):

$$X_n = \int_0^{2\pi} e^{inw}dW(w) = \int_0^{2\pi} e^{inw}\widehat{a}(w)dZ(w), \quad n \in \mathbb{Z}, \tag{3.3}$$

where $W(w)$, $w \in (0, 2\pi)$, is a Gaussian, orthogonal (independent) increment, complex-valued process such that $EdW(w)\overline{dW(w')} = |\widehat{a}(w)|^2 dw 1_{\{w=w'\}}/2\pi$, $Z(w)$, $w \in (0, 2\pi)$, is a Gaussian, orthogonal (independent) increment process such that $EdZ(w)\overline{dZ(w')} = dw1_{\{w=w'\}}/2\pi$, and

$$\widehat{a}(w) = \sum_{k=-\infty}^{\infty} a_k e^{-ikw}, \quad w \in (0, 2\pi), \tag{3.4}$$

is the discrete Fourier transform of a sequence $a \in l^2(\mathbb{Z})$. The quantity $|\widehat{a}(w)|^2/2\pi$ is known as a spectral density of X. Observe also that

$$r = a * \overline{a}, \quad \widehat{r}(w) = |\widehat{a}(w)|^2,$$

where $\{\overline{x}_k\} = \{x_{-k}\}$ stands for reversal in time of a sequence $\{x_k\}$.

In regard to wavelets, since we work in discrete time, we will use the so-called Conjugate Mirror Filters (CMF) associated with an orthogonal Multiresolution Analysis (MRA). See, for example, Mallat (1998). These are a low pass filter $u = \{u_n\}$ and a high pass filter $v = \{v_n\}$ satisfying a number of properties. In particular, for any $w \in \mathbb{R}$,

$$|\widehat{u}(w)|^2 + |\widehat{u}(w+\pi)|^2 = 2, \tag{3.5}$$

$$\widehat{v}(w) = e^{-iw}\overline{\widehat{u}(w+\pi)} \tag{3.6}$$

and hence

$$|\widehat{v}(w)|^2 + |\widehat{v}(w+\pi)|^2 = 2, \tag{3.7}$$

$$\widehat{u}(w)\overline{\widehat{v}(w)} + \widehat{u}(w+\pi)\overline{\widehat{v}(w+\pi)} = 0. \tag{3.8}$$

Popular CMF are those of Daubechies with N zero moments, $N \geq 1$. For fixed N, these filters are of finite length $2N$. It is also known (e.g., Mallat (1998), p. 241) that, with N zero moments and finite length CMF,

$$\widehat{u}(w) = (1 + e^{-iw})^N \widehat{u}_{0,N}(w), \quad \widehat{v}(w) = (1 - e^{-iw})^N \widehat{v}_{0,N}(w), \tag{3.9}$$

with $u_{0,N}, v_{0,N}$ of finite length as well.

CMF u and v appear in the (orthogonal) Fast Wavelet Transform (FWT) of a deterministic sequence $x = \{x_n\}$. Setting $a_0 = x$, at the decomposition step, one defines the approximation and detail coefficients as

$$a_j = \downarrow_2 (\overline{u} * a_{j-1}), \quad d_j = \downarrow_2 (\overline{v} * a_{j-1}), \quad j = 1, 2, \ldots, \tag{3.10}$$

where $(\downarrow_2 x)_k = x_{2k}$ is the downsampling (decimation) by factor 2 operation. At the reconstruction step, one has

$$a_j = u * \uparrow_2 a_{j+1} + v * \uparrow_2 d_{j+1}, \quad j = 0, 1, \dots, \tag{3.11}$$

where $(\uparrow_2 x)_k = x_{k/2} 1_{\{\text{even } k\}} + 0 1_{\{\text{odd } k\}}$ is the upsampling by factor 2 operation. One can easily verify that

$$\widehat{(\downarrow_2 x)}(w) = \frac{1}{2}\Big(\widehat{x}\Big(\frac{w}{2}\Big) + \widehat{x}\Big(\frac{w}{2} + \pi\Big)\Big), \quad \widehat{(\uparrow_2 x)}(w) = \widehat{x}(2w). \tag{3.12}$$

The time series and wavelet decompositions are considered above on the index set $\mathbb{Z}$. We shall also consider below the case of a finite index set $0, 1, \dots, T-1$, with $T = 2^J$. In this case, the convolution $*$ above is often replaced by the circular convolution $\circledast$, and the discrete Fourier transform of $x = \{x_0, x_1, \dots, x_{T-1}\}$ becomes

$$\widehat{x}(w) = \sum_{k=0}^{T-1} x_k e^{-ikw}, \quad \text{at } w = \frac{2\pi j}{T}, \ j = 0, \dots, T-1. \tag{3.13}$$

In particular, with these modifications, (3.10) is considered for $j = 1, \dots, J$, and (3.11) continues to hold for $j = 0, 1, \dots, J-1$. When x and y are of arbitrary (possibly infinite) length, the circular convolution is defined as

$$x \circledast y = x^{per} \circledast y^{per} \quad \text{with, e.g.,} \quad x_k^{per} = \sum_n x_{k+nT}.$$

One has $\widehat{(x \circledast y)}(w) = \widehat{x}(w)\widehat{y}(w)$, where x and y can be of arbitrary length.

The time series vectors $Y = \{Y_0 \dots, Y_{T-1}\}$ that are natural in the context of circular convolutions, are

$$Y_n = (b \circledast \epsilon)_n, \quad n = 0, \dots, T-1, \tag{3.14}$$

where $\epsilon = \{\epsilon_0, \dots, \epsilon_{T-1}\}$ are independent, $\mathcal{N}(0,1)$ random variables and $b = \{b_0, \dots, b_{T-1}\}$ is a vector. These time series vectors are also stationary (in the sense that $EY_iY_j = EY_0Y_{j-i}$ with $0 \le i \le j \le T-1$) but not every stationary vector can be written this way. The covariance matrix $(E(Y_iY_j),\ i,j = 0,\dots,T-1)$ is, in fact, circular. Conversely, under mild assumptions, a Gaussian vector Y with a circular covariance matrix can be written as (3.14). If r_Y is the autocovariance function of Y, observe also that

$$r_Y = b \circledast \overline{b}, \quad \widehat{r}_Y(w) = |\widehat{b}(w)|^2. \tag{3.15}$$

3.3 Definition and basic properties of AWD

We shall use below the following general result of its own interest. See Section 3.6 for a proof.

Proposition 3.3.1. *Let $a, b \in l^2(\mathbb{Z})$ be arbitrary filters and $u, v \in l^2(\mathbb{Z})$ be CMF. Define*

$$\widehat{U}_d(\omega) = \overline{\left(\frac{\widehat{b}(2\omega)}{\widehat{a}(\omega)}\right)}\widehat{u}(\omega), \quad \widehat{V}_d(\omega) = \overline{\left(\frac{1}{\widehat{a}(\omega)}\right)}\widehat{v}(\omega), \tag{3.16}$$

and

$$\widehat{U}_r(\omega) = \frac{\widehat{a}(\omega)}{\widehat{b}(2\omega)}\widehat{u}(\omega), \quad \widehat{V}_r(\omega) = \widehat{a}(\omega)\widehat{v}(\omega). \tag{3.17}$$

Suppose that $\widehat{U}_d, \widehat{V}_d, \widehat{U}_r, \widehat{V}_r \in L^2(0, 2\pi)$ and the corresponding filters

$$U_d, V_d, U_r, V_r \in l^1(\mathbb{Z}). \tag{3.18}$$

(i) (Decomposition step) If $X = a * \epsilon$ is a stationary time series with a Gaussian white noise ϵ, then

$$Y = \downarrow_2 (\overline{U}_d * X), \quad \eta = \downarrow_2 (\overline{V}_d * X), \tag{3.19}$$

are such that $Y = b * \xi$ is a stationary Gaussian time series with a Gaussian white noise ξ, and η is a Gaussian white noise, independent of ξ and hence of Y.

(ii) (Reconstruction step) If Y and η are the independent time series obtained in (i) above, then

$$X = U_r * \uparrow_2 Y + V_r * \uparrow_2 \eta. \tag{3.20}$$

Remark 3.3.1. The results (i) and (ii) can be informally explained as follows. Writing

$\widehat{X}(w) = \widehat{a}(w)\widehat{\epsilon}(w)$, observe that

$$\overline{\widehat{U_d}(w)}\widehat{X}(w) = \widehat{b}(2w)\widehat{u}(w)\widehat{\epsilon}(w), \quad \overline{\widehat{V_d}(w)}\widehat{X}(w) = \widehat{v}(w)\widehat{\epsilon}(w).$$

Hence, by using (3.12), the Fourier transforms of the R.H.S. of (3.19) are

$$\downarrow_2 \widehat{(\overline{U}_d * X)}(w) = \widehat{b}(w)\downarrow_2 \widehat{(u * \epsilon)}(w), \quad \downarrow_2 \widehat{(\overline{V}_d * X)}(w) = \downarrow_2 \widehat{(v * \epsilon)}(w). \tag{3.21}$$

Similarly, the Fourier transform of the R.H.S. of (3.20) is

$$\widehat{U}_r(w)\widehat{Y}(2w) + \widehat{V}_r(w)\widehat{\eta}(2w) = \widehat{a}(w)\Big(\widehat{u}(w)\widehat{\xi}(2w) + \widehat{v}(w)\widehat{\eta}(2w)\Big)$$

$$= \widehat{a}(w)\Big(\widehat{u* \uparrow_2 \xi} + \widehat{v* \uparrow_2 \eta}\Big)(w).$$

If ϵ is a Gaussian white noise, it is easy to verify that its discrete (orthogonal) wavelet transform leads to approximation coefficients $\xi = \downarrow_2 (u*\epsilon)$ and detail coefficients $\eta = \downarrow_2 (v*\epsilon)$ which are two independent Gaussian white noise sequences. The equation $u* \uparrow_2 \xi + v* \uparrow_2 \eta$ is just the usual reconstruction of ϵ.

Remark 3.3.2. Another interpretation of Proposition 3.3.1 is to say that the set of filters (U_d, V_d, U_r, V_r) form a perfect reconstruction filter bank (see, for example, Brockwell and Davis (1991), p. 259). Indeed, by Theorem 7.8 in Mallat (1998), this is so if and only if

$$\overline{\widehat{U_d}(w)}\widehat{U}_r(w+\pi) + \overline{\widehat{V_d}(w)}\widehat{V}_r(w+\pi) = 0,$$

$$\overline{\widehat{U_d}(w)}\widehat{U}_r(w) + \overline{\widehat{V_d}(w)}\widehat{V}_r(w) = 0.$$

The L.H.S. of the first relation is

$$\frac{\widehat{a}(w+\pi)}{\widehat{a}(w)}\left(\overline{\widehat{u}(w)}\widehat{u}(w+\pi) + \overline{\widehat{v}(w)}\widehat{v}(w+\pi)\right),$$

which is 0, since the term in the parentheses is 0. The second relation can be proved similarly. Note also that Proposition 3.3.1 is not a consequence of perfect reconstruction because filtering involves (random) time series.

The following result is a simple consequence of Proposition 3.3.1.

Corollary 3.3.1. *Let $X^0 = a^0 * \epsilon^0$ be a Gaussian, stationary time series with $a^0 \in l^2(\mathbb{Z})$ and a Gaussian white noise ϵ^0. For $j \geq 1$, let also*

$$a^j \in l^2(\mathbb{Z}) \tag{3.22}$$

and

$$\widehat{U}_d^j(\omega) = \overline{\left(\frac{\widehat{a}^j(2\omega)}{\widehat{a}^{j-1}(\omega)}\right)}\widehat{u}(\omega), \quad \widehat{V}_d^j(\omega) = \overline{\left(\frac{1}{\widehat{a}^{j-1}(\omega)}\right)}\widehat{v}(\omega), \tag{3.23}$$

where u, v are CMF. Suppose that $\widehat{U}_d^j, \widehat{V}_d^j \in L^2(0, 2\pi)$ and the corresponding filters

$$U_d^j, V_d^j \in l^1(\mathbb{Z}), \quad j \geq 1. \tag{3.24}$$

(i) (Decomposition step) For $j \geq 1$, let

$$X^j = \downarrow_2 (\overline{U}_d^j * X^{j-1}), \quad \xi^j = \downarrow_2 (\overline{V}_d^j * X^{j-1}). \tag{3.25}$$

Then, for $j \geq 1$,

$$X^j = a^j * \epsilon^j \tag{3.26}$$

with a Gaussian white noise ϵ^j, and ξ^j, $j \geq 1$, are independent, Gaussian white noise sequences, and ϵ^J (hence X^J) and ξ^j, $j \leq J$, are independent.

(ii) (Reconstruction step) If, in addition,

$$\widehat{U}_r^j(\omega) = \frac{\widehat{a}^j(\omega)}{\widehat{a}^{j+1}(2\omega)}\widehat{u}(\omega), \quad \widehat{V}_r^j(\omega) = \widehat{a}^j(\omega)\widehat{v}(\omega) \tag{3.27}$$

are such that $\widehat{U}_r^j, \widehat{V}_r^j \in L^2(0, 2\pi)$ and the corresponding filters

$$U_r^j, V_r^j \in l^1(\mathbb{Z}), \tag{3.28}$$

then

$$X^j = U_d^j * \uparrow_2 X^{j+1} + V_r^j * \uparrow_2 \xi^{j+1}, \quad j \geq 0. \tag{3.29}$$

Definition 3.3.1. The decomposition of a stationary time series $X = X^0$ into the series X^j, ξ^j, $j \geq 1$, in Corollary 3.3.1 will be called *Adaptive Wavelet Decomposition* (AWD, in short) of a stationary time series X. We will refer to X^j as *approximations* and to ξ^j as *details.*

Remark 3.3.3. AWD can be easily extended to cyclic time series $Y = Y^0 = a^0 \circledast \epsilon^0$ given by (3.14) of length 2^J. Consider

$$Y^j = \downarrow_2 (\overline{U}_d^j \circledast Y^{j-1}), \quad \xi^j = \downarrow_2 (\overline{V}_d^j \circledast Y^{j-1}), \quad j = 1, \dots, J, \tag{3.30}$$

at decomposition, and

$$Y^j = U_r^j \circledast \uparrow_2 Y^{j+1} + V_r^j \circledast \uparrow_2 \xi^{j+1}, \quad j = 0, \dots, J-1, \tag{3.31}$$

at reconstruction. Then, ξ^j are independent, Gaussian white noise sequences of length 2^{J-j}, and $Y^j = a^j \circledast \epsilon^j$ are circular time series with Gaussian white noise sequences ϵ^j of length 2^{J-j}.

In practice, only finite data $X_0, X_1, \dots, X_{T-1}$ are available and hence AWD cannot be applied (supposing also that a is known). For finite data $\widetilde{X}^0 = (X_0, X_1, \dots, X_{T-1})$ with $T = 2^J$, consider the following time series vectors:

$$\widetilde{X}^j = \downarrow_2 (\overline{U}_d^j \circledast \widetilde{X}^{j-1}), \quad \widetilde{\xi}^j = \downarrow_2 (\overline{V}_d^j \circledast \widetilde{X}^{j-1}), \quad j = 1, \dots, J. \tag{3.32}$$

These relations differ from those in (3.25) by the presence of circular convolution $\circledast$. In particular, observe that the series $\widetilde{X}^j, \widetilde{\xi}^j$ have now length 2^{J-j}. Observe also that $\widetilde{X}^j, \widetilde{\xi}^j$ are well-defined as long as $U_d^j, V_d^j \in l^1(\mathbb{Z})$ which is the assumption (3.24). Moreover, it can be verified that

$$\widetilde{X}^j = U_r^j \circledast \uparrow_2 \widetilde{X}^{j+1} + V_r^j \circledast \uparrow_2 \widetilde{\xi}^{j+1}, \quad j = 0, \dots, J-1. \tag{3.33}$$

The idea behind (3.32) is the following. If U_d^j, V_d^j have short length (or decay fast to 0), and the length T is large, then most elements of $\widetilde{X}^1, \widetilde{\xi}^1$ are computed as in AWD (and hence those of $\widetilde{X}^1$ are akin to $a^1 * \epsilon^1$, and those in ξ^1 are independent). Only those few coefficients that are at the end of the time series vector $\widetilde{X}^0$ (affected by the border, or under the border effect) are different from those in AWD. More generally, the elements of $\widetilde{X}^j, \widetilde{\xi}^j$ unaffected by the border are computed as in AWD.

Definition 3.3.2. The decomposition of a stationary vector $\widetilde{X}^0 = (X_0, X_1, \ldots, X_{T-1})$ with $T = 2^J$ into the vectors $\widetilde{X}^j, \widetilde{\xi}^j$, $j = 1, \ldots, J$, in (3.32) will be called *approximate AWD.*

Remark 3.3.4. Using circular convolutions in (3.32) at decomposition can be viewed as one way of dealing with the boundary when having finite data. More precisely, approximate AWD of $X = \widetilde{X}^0 = (X_0, X_1, \ldots, X_{T-1})$ is the usual AWD applied to the infinite time series obtained by extending observations periodically outside the boundary. Other ways are, for example, to consider observations outside the boundary as zero, or to extend periodically the vector $(X_0, X_1, \ldots, X_{T-2}, X_{T-1}, X_{T-2}, \ldots, X_1)$. Using circular convolutions is convenient analytically.

Remark 3.3.5. Another perspective on approximate AWD concerns covariance factorization. If $\widetilde{X}^0 = (X_0, X_1, \ldots, X_{T-1})$ with $T = 2^J$ is a Gaussian stationary sequence, let

$$\widetilde{Y} = (\widetilde{\xi}^1, \widetilde{\xi}^2, \ldots, \widetilde{\xi}^J, \widetilde{Y}^J) \tag{3.34}$$

be a $1 \times T$ vector consisting of details $\widetilde{\xi}^j$ and last approximation $\widetilde{Y}^J$ in approximate AWD. Write

$$\widetilde{Y} = \widetilde{X}^0 M \tag{3.35}$$

for an invertible matrix M. Most of the details $\widetilde{\xi}^j$ are approximately independent, $\mathcal{N}(0, 1)$ random variables. Hence,

$$E\widetilde{Y}'\widetilde{Y} \approx \mathrm{Id},$$

where Id is the identity matrix, and the variance of $\widetilde{Y}^J$ is ignored for simplicity. By using

(3.35),

$$E\widetilde{X}^{0\prime}\widetilde{X}^0 = (M^{-1})'\Big(E\widetilde{Y}'\widetilde{Y}\Big)M^{-1} \approx (M^{-1})'M^{-1}, \tag{3.36}$$

which is an approximate factorization of the covariance matrix of $\widetilde{X}^0$. Observe also that the matrix M is not orthogonal.

Note from Definition 3.3.1 that AWD are quite general in the choice of moving average filters a^j, and hence the corresponding time series X^j. In fact, AWD can be defined for many different choices of a^j's but only some of them will have desired properties. These properties can be suggested by an application at hand or other considerations, for example,

(*a*) X^J and ξ^j, $j \geq J \geq 1$, consisting of uncorrelated (independent) variables,

(*b*) $U_d^j, V_d^j, U_r^j, V_r^j$ decaying to zero fast, or

(*c*) X^j being natural approximations to X^0 at scale 2^j.

The property (*a*) is important in Signal Processing as it is typically associated with optimality in coding (see Section 3.1). In the applications considered here, we were motivated by (*b*), in view of approximate AWD (see the discussion preceding Definition 3.3.2). In regard to (*c*), one natural approximation of a series X^0 at scale 2^j is

$$X^j = \{X^0_{2^j k}\}_{k\in\mathbb{Z}}. \tag{3.37}$$

In particular, if $\widehat{a}^0(w) = \widehat{a}(w)$ enters the spectral representation (3.3) of X^0, then

$$\widehat{a}^j(w) = \frac{1}{2}\Big(\widehat{a}^{j-1}\Big(\frac{w}{2}\Big) + \widehat{a}^{j-1}\Big(\frac{w}{2}+\pi\Big)\Big) \tag{3.38}$$

is associated with the spectral representation of X^j. See Example 3.4.2 below for further discussion on (*c*).

An important property of any AWD is that details ξ^j ignore polynomial trends up to the order of the number of zero moments. Analogous fact is well-known for orthogonal wavelet decompositions. (In discrete time, this follows immediately from Theorem 7.4, (*iv*), in Mallat (1998).) We show that it continues to hold here as well (see Section 3.6 for a proof).

Proposition 3.3.2. *Suppose that the underlying orthogonal MRA has N zero moments with factorization (3.9). Let $p_n = p(n)$ where a polynomial p is of degree $D < N$. Consider AWD with decomposition filters U_d^j, V_d^j such that $|U_{d,n}^j|, |V_{d,n}^j| \le C_j |n|^{-D-2}$, where C_j is a constant. Then, for any $j \ge 1$,*

$$\xi^j(p) = 0, \tag{3.39}$$

where $\xi^j(p)$ are details in AWD when applied to the polynomial p.

3.4 Examples of AWD

We provide here several examples of AWD. We would like the associated set of filters $U_d^j, V_d^j, U_r^j, V_r^j$ to decay to zero fast (see (b) following Remark 3.3.5). For some time series, this turns out to be possible when the number of zero moments of the underlying MRA increases. Note from the examples below that we use the term "decay" in a rather loose sense.

Example 3.4.1. (FARIMA(0,s,0)) Let X be a Gaussian FARIMA(0,s,0) time series with $s \in (-1/2, 1/2)$ ($s \neq 0$), that is, $X = a * \epsilon$ with a Gaussian white noise ϵ and

$$\widehat{a}(w) = (1 - e^{-iw})^{-s} \tag{3.40}$$

(see, for example, Brockwell and Davis (1991), p. 520, or Beran (1994)). The case $s \in (0, 1/2)$ corresponds to the so-called long memory, generally considered more difficult to deal with.

Consider AWD with

$$\widehat{a^j}(w) = \widehat{a}(w), \tag{3.41}$$

for any $j \ge 1$, and focus on the definition (3.27) of U_r^j, V_r^j. Note that

$$\frac{\widehat{a}(w)}{\widehat{a}(2w)} = \frac{(1 - e^{-iw})^{-s}}{(1 - e^{-i2w})^{-s}} = (1 + e^{-iw})^{s} = \sum_{k=0}^{\infty} f_k^{(s)} e^{-iwk}, \tag{3.42}$$

$$\widehat{a}(w) = (1 - e^{-iw})^{-s} = \sum_{k=0}^{\infty} g_k^{(-s)} e^{-iwk} \tag{3.43}$$

are the two filters entering (3.27). These filters, in fact, decay extremely slowly: one can show by using the Stirling's formula that, as $k \to \infty$,

$$f_k^{(s)} \sim (-1)^k \frac{k^{-s-1}}{\Gamma(-s)}, \quad g_k^{(-s)} \sim \frac{k^{s-1}}{\Gamma(s)}. \tag{3.44}$$

(For example, when $s \in (0, 1/2)$, the second filter is not even summable.)

It is therefore quite surprising that, in fact, the resulting filters U_r^j, V_r^j may decay to 0 very rapidly. As mentioned above, this results from the number of zero moments of the underlying orthogonal MRA. Letting N denote the number of zero moments and using (3.9), observe that

$$\widehat{U}_r(w) \equiv \widehat{U}_r^j(w) = (1+e^{-iw})^{s+N}\widehat{u}_{0,N}(w), \quad \widehat{V}_r(w) \equiv \widehat{V}_r^j(w) = (1-e^{-iw})^{-s+N}\widehat{v}_{0,N}(w). \tag{3.45}$$

By (3.42)–(3.44), we now have

$$(1+e^{-iw})^{s+N} = \sum_{k=0}^{\infty} f_k^{(s+N)} e^{-iwk} \quad \text{with} \quad f_k^{(s+N)} \sim (-1)^k \frac{k^{-s-N-1}}{\Gamma(-s-N)},$$

$$(1-e^{-iw})^{-s+N} = \sum_{k=0}^{\infty} g_k^{(-s+N)} e^{-iwk} \quad \text{with} \quad g_k^{(-s+N)} \sim \frac{k^{s-N-1}}{\Gamma(s-N)}, \tag{3.46}$$

as $k \to \infty$. Comparing (3.46) with (3.44), we see that these filters now decay rapidly when N is large.

The latter observation by itself does not show that the resulting filters U_r, V_r in (3.45) decay faster as N increases because $u_{0,N}$ and $v_{0,N}$ also grow in size (not length). To see that U_r, V_r indeed decrease faster with N, consider Table 3.1. In this table, we provide lengths of U_r, V_r truncated at a priori specified cutoff levels δ for various choices of N and Daubechies CMF. The value $s = 0.25$ is considered. The filters $u_{0,N}$ can be found in Table 6.2 of Daubechies (1992), p. 196. Observe from Table 1 that the effect of increasing N is really substantial. For example, when $\delta = 10^{-7}$, the length of truncated V_r goes from 4066 with $N = 1$ to 40 when $N = 10$. It should also be noted that the results of Table 3.1 are not sensitive to the value of s. In particular, the change in the results is small as s approaches 1/2.

Length of truncated filters					
Filters	Cutoff ϵ	$N=1$	$N=3$	$N=6$	$N=10$
U_r	10^{-7}	706	77	38	35
	10^{-10}	$\approx 1.5\times 10^4$	375	89	56
	10^{-15}	$\approx 2.5\times 10^6$	5577	440	140
V_r	10^{-7}	4066	114	44	40
	10^{-10}	$\approx 2\times 10^5$	696	108	63
	10^{-15}	$\approx 1.5\times 10^8$	$\approx 1.4\times 10^4$	557	160

Table 3.1: Lengths of truncated filters U_r and V_r at cutoff δ with $s = 0.25$ and the Daubechies MRA with N zero moments.

We discussed above the decay of reconstruction filters U_r^j, V_r^j. Similar conclusions can be reached for decomposition filters U_d^j, V_d^j in (3.23) by writing, for example, in the case of U_d^j,

$$\overline{\left(\frac{\widehat{a}(2w)}{\widehat{a}(w)}\right)}(1+e^{-iw})^N = (1+e^{iw})^s(1+e^{-iw})^N = (1+e^{iw})^{s+N}e^{-iwN}.$$

In conclusion, if fast decaying filters $U_d^j, V_d^j, U_r^j, V_r^j$ are needed, the AWD with (3.40) appears to be a suitable choice for FARIMA$(0,s,0)$ time series.

Remark 3.4.1. The faster decay in (3.46) has also the following simple explanation that is useful more generally. According to (3.42)–(3.44), the elements $f_k^{(s)}$ of $(1+e^{-iw})^s = \widehat{a}(w)/\widehat{a}(2w)$ decay as

$$f_k^{(s)} \sim (-1)^k \frac{k^{-s-1}}{\Gamma(-s)}.$$

Application of the filter $(1+e^{-iw})^N$ to $(1+e^{-iw})^s$ corresponds to taking sums in blocks of size N. Since $f_k^{(s)}$ oscillates and decays, the sums will become smaller. A similar explanation with difference instead of sums applies to the elements $g_k^{(-s)}$ of $(1-e^{-iw})^{-s}$.

Example 3.4.2. (AR(1),MA(1)) Let X be a Gaussian AR(1) time series, that is, $X = a * \epsilon$ with a Gaussian white noise ϵ and

$$\widehat{a}(w) = (1 - a_1 e^{-iw})^{-1}, \tag{3.47}$$

where $-1 < a_1 < 1$ ($a_1 \neq 0$). The case of $a_1 = \pm 1$, not considered here, corresponds to unit roots, and the case of a_1 close to ± 1 ($-1 < a_1 < 1$) is referred to as near unit roots.

If only the decomposition of X is of interest (as, for example, in maximum likelihood estimation), consider AWD with

$$a^j(w) \equiv 1, \quad j \geq 1. \tag{3.48}$$

Then,

$$\widehat{U}_d^1(w) = (1 - a_1 e^{iw})\widehat{u}(w), \quad \widehat{V}_d^1(w) = (1 - a_1 e^{iw})\widehat{v}(w) \tag{3.49}$$

and

$$\widehat{U}_d^j(w) = \widehat{u}(w), \quad \widehat{V}_d^j(w) = \widehat{v}(w), \quad j \geq 2. \tag{3.50}$$

Hence, the corresponding filters U_d^j, V_d^j are of short and finite length (supposing that u and v are such). Note also that, in this case, all approximations X^j and details ξ^j are Gaussian white noise sequences.

Suppose now that the reconstruction of X is also of interest. With the choice (3.48),

$$\widehat{U}_r^0(w) = \frac{\widehat{u}(w)}{1 - a_1 e^{-iw}}, \quad \widehat{V}_r^0(w) = \frac{\widehat{v}(w)}{1 - a_1 e^{-iw}} \tag{3.51}$$

and

$$\widehat{U}_r^j(w) = \widehat{u}(w), \quad \widehat{V}_r^j(w) = \widehat{v}(w), \quad j \geq 1. \tag{3.52}$$

When a_1 is close to 0, the elements of $(1 - a_1 e^{-iw})^{-1} = \sum_{k=0}^{\infty} a_1^k e^{-iwk}$ decay to zero rapidly and hence the filters U_r^0, V_r^0 can be taken of short length in practice. When a_1 is close to ± 1, however, the decay of a_1^k is much slower, resulting in longer filters U_r^0, V_r^0. Zero moments are not helpful for U_r^0 when $0 < a_1 < 1$, and for V_r^0 when $-1 < a_1 < 0$ (see Remark 3.4.1 above).

When $0 < a_1 < 1$, the decay of U_r^0 can be improved by considering a different AWD. Take AWD with

$$\widehat{a}^j(w) = (1 - a_1^{2^j} e^{-iw})^{-1} \tag{3.53}$$

so that

$$\widehat{U}_r^j(w) = (1 + a_1^{2^j} e^{-iw})\widehat{u}(w), \quad \widehat{V}_r^j(w) = \frac{\widehat{v}(w)}{1 - a_1^{2^j} e^{-iw}}, \quad j \geq 0. \tag{3.54}$$

In this case, U_r^j are also of finite and short length. The larger number of zero moments make the filter V_r^j decay faster, especially when a_1 is close to 1. We illustrate this in Table 3.2 in the following way. Let v be the Daubechies CMF with N zero moments so that its length is $2N$. The filter V_r^0 is obtained by convolving the sequence $(1, a_1, a_1^2, \dots)$ with the filter v. Note that the $(2N+j)$th nonzero element of the convolution is

$$a_1^j c := a_1^j (1, a_1, \dots, a_1^{2N-1}) v', \quad j \geq 0,$$

and decays as a geometric sequence. In Table 3.2, we provide the absolute values of the $(2N)$th nonzero element of the filter V_r^0 for various choices of the parameter a_1 and the number of zero moments N. In parentheses, we provide the value of a_1^{2N} for comparison. Note that, when a_1 is closer to 1, the filter V_r^0 indeed decays much faster (in the sense of being closer to 0 overall) with the increasing number of zero moments. For smaller a_1 ($a_1 = 0.5$ in Table 3.2), this effect is no longer present.

Note also that, with the choice (3.53) for AWD, the approximations X^j become AR(1) time series with the parameters $a_1^{2^j}$. The decomposition filters associated with (3.53) are

$$\widehat{U}_d^j(w) = \frac{\widehat{u}(w)}{1 + a_1^{2^j} e^{iw}}, \quad \widehat{V}_d^j(w) = (1 - a_1^{2^j} e^{iw}) \widehat{v}(w). \tag{3.55}$$

When a_1 is close to 1, the filters U_d^j can also be seen to decay faster with the increasing number of zero moments.

When $-1 < a_1 < 0$ and especially when a_1 is close to -1, the AWD with (3.53) is not helpful because the decay of V_r^0 (V_r^j with $j = 0$) is not affected by the increasing number of zero moments. This occurs because, in simple terms, the elements of $(1 - a_1 e^{-iw})^{-1} = \sum_{k=1}^{\infty} (-1)^k |a_1|^k e^{-iwk}$ oscillate and the difference operator $(1 - e^{-iw})^N$ does not make them decrease to 0 faster (see Remark 3.4.1). In this case, the AWD with (3.48) is probably the best one can do. Note that, with (3.48), increasing the number of zero moments make the filters U_r^0 decay faster. This does not affect V_r^0 and, the closer a_1 is to -1, the longer V_r^0 should be taken in practice.

We discussed above the case of AR(1) time series. Suppose now that X is an MA(1)

Size of the $(2N)$th nonzero element				
a_1	$N=1$	$N=3$	$N=6$	$N=10$
0.5	0.3535 (0.25)	0.0267 (0.0156)	0.0007 (0.0002)	7.9×10^{-6} (9.5×10^{-7})
0.7	0.2121 (0.49)	0.0089 (0.1176)	0.0001 (0.0138)	2.8×10^{-7} (0.0007)
0.9	0.0707 (0.81)	0.0004 (0.5314)	3.2×10^{-7} (0.2824)	8.5×10^{-11} (0.1215)
0.999	0.0007 (0.998)	5.5×10^{-10} (0.994)	3.9×10^{-15} (0.988)	3.8×10^{-10} (0.9801)

Table 3.2: The $(2N)$th nonzero element of the filter V_r^0 for various choices of a_1 and the Daubechies MRA with N zero moments.

time series, that is, $X = a * \epsilon$ with

$$\widehat{a}(w) = 1 + b_1 e^{-iw}, \tag{3.56}$$

where $-1 < b_1 < 1$ ($b_1 \neq 0$). Since $\widehat{a}(w)$ in (3.56) is reciprocal to that in (3.47), our discussion above also covers the case of MA(1) time series. For example, reconstruction filters for AR(1) time series now become decomposition filters for MA(1) time series. Equivalently, AWD for MA(1) time series is applied at decomposition with either

$$\widehat{a^j}(w) \equiv 1, \quad j \geq 1,$$

or

$$\widehat{a^j}(w) = 1 - (-b_1)^{2^j} e^{-iw}, \quad j \geq 1.$$

It is also clear that our discussion can be extended to more general ARMA(p, q) time series.

Remark 3.4.2. If X^0 is an MA(1) time series with $\widehat{a}(w) = 1 + b_1 e^{-iw}$, $-1 < b_1 < 1$ ($b_1 \neq 0$), or $X_n^0 = \epsilon_n + b_1 \epsilon_{n-1}$ with a Gaussian white noise $\{\epsilon_n\}$, then X^j in (3.37) are all (up to a constant) Gaussian white noise sequences or

$$\widehat{a^j}(w) \equiv 1. \tag{3.57}$$

If X^0 is an AR(1) time series with $\widehat{a}(w) = (1 - a_1 e^{-iw})^{-1}$, $-1 < a_1 < 1$ ($a_1 \neq 0$), or

$X_n^0 = \epsilon_n + a_1\epsilon_{n-1} + a_1^2\epsilon_{n-2} + \ldots$, then X^j in (3.37) are associated with

$$\widehat{a}^j(w) = (1 - a_1^{2^j} e^{-iw})^{-1}. \tag{3.58}$$

Observe that (3.57) and (3.58) are exactly what was proposed for AWD at reconstruction for MA(1) and AR(1) time series in Example 3.4.2 above.

3.5 Applications of AWD

We consider here applications of AWD to simulation (Section 3.5.1) and MLE (Section 3.5.2). Simulation uses AWD at reconstruction and MLE uses AWD at decomposition.

3.5.1 Simulation

Suppose that the time series X of length 2^J is desired. It can be simulated using AWD through the following steps:

1. For $j = 0, 1, \ldots, J-1$, determine the largest length L_J of the reconstruction filters U_r^j, V_r^j truncated at a chosen cutoff level $\delta > 0$. Let $\widetilde{U}_r^j, \widetilde{V}_r^j$, $j = 0, 1, \ldots, J-1$, be the reconstruction filters U_r^j, V_r^j truncated to have length L_J each.

2. Use some simulation method to generate the time series vector X^J of length $L_J + 1$.

3. Apply the reconstruction scheme (3.29) recursively J times with the truncated reconstruction filters $\widetilde{U}_r^j, \widetilde{V}_r^j$ and taking into account the border effect to obtain the time series X^0 of length 2^J.

Several observations regarding these steps are in order. Implementation of the first step depends on the time series to simulate. For example, in the case of (3.45), the reconstruction filters are the same for all j. The second step refers to the fact that the application of the reconstruction scheme (3.29) requires some initial approximation X^j. We take $j = J$ because X^J can be taken of the smallest possible length $L_J + 1$ in order to apply the simulation scheme (3.29). The time series X^J can be simulated by a popular Circular Matrix Embedding (CME) method (Dietrich and Newsam (1997)) or, since L_J is often small, by the Durbin-Levinson algorithm (Brockwell and Davis (1991)). For the third step, observe that applying the scheme (3.29) with $\widetilde{U}_r^{J-1}, \widetilde{V}_r^{J-1}$ to X_J of length L_J+1, we obtain $2(L_J+1)-1-L_J = L_J+2$ observations of the time series X^{J-1} which are unaffected by the border. Here, $2(L_J+1)-1$ is the number of observations after the operation $\uparrow_2$ and $(-L_J)$ takes into account the border effect. By repeating this argument, the number of observations of the resulting time series X^0 which are unaffected by the border, is $L_J + 2^J > 2^J$.

Simulation based on AWD is of interest because it is very fast. Modulo computation of the truncated reconstruction filters $\widetilde{U}_r^j, \widetilde{V}_r^j$ and simulation of the initial time series X^J, the simulation algorithm based on AWD is of the computational order $O(2^J)$. The CME method based on FFT is of the slower order $O(2^J \log 2^J)$. This obviously is relevant only for simulation of really *long* time series.

In simulation above, however, it is necessary to generate a time series at initial coarsest scale (by some other method) and to deal with boundary in a quite nontrivial way. This could be avoided at the expense of making an approximation if convolutions in AWD are replaced by circular convolutions. In other words, consider a time series $\widetilde{X}^0$ of length 2^K defined recursively by (3.31), that is,

$$\widetilde{X}^k = U_r^k \circledast \uparrow_2 \widetilde{X}^{k+1} + V_r^k \circledast \uparrow_2 \widetilde{\xi}^{k+1}, \quad k = 0, \dots, K-1, \tag{3.59}$$

where $\widetilde{\xi}^k$ are independent, Gaussian white noise sequences of length 2^{K-k}, and $\widetilde{X}^K = a^K \circledast \widetilde{\epsilon}^K = (\sum_n a_n^K)\widetilde{\epsilon}_0^K$ is of length 1. The scheme (3.59) is easy to implement. But is $\widetilde{X}^0$ close to the desired time series $X = X^0$ in any way?

To answer this question, note by Remark 3.3.3 that $\widetilde{X}^0$ can, in fact, be represented as $\widetilde{X}^0 = a^0 \circledast \widetilde{\epsilon}^0$ with a Gaussian, white noise sequence $\widetilde{\epsilon}^0$ of length 2^K. As $\widetilde{X}^0$ is a cyclic time series, it does not approximate a stationary time series X^0. Observe also by (3.15) that

$$\widehat{\widetilde{r}^0}(w) = |\widehat{a}^0(w)|^2, \quad w = \frac{2\pi m}{2^K},\ m = 0, \dots, 2^K - 1,$$

and

$$\widetilde{r}^0(n) = \frac{1}{2^K} \sum_{m=0}^{2^K-1} e^{i\frac{2\pi mn}{2^K}} \Big|\widehat{a}^0\Big(\frac{2\pi m}{2^K}\Big)\Big|^2, \tag{3.60}$$

where $\widetilde{r}^0$ is the autocovariance function of $\widetilde{X}^0$.

It may appear from (3.60) that, as $K \to \infty$,

$$\widetilde{r}^0(n) \approx \frac{1}{2\pi} \int_0^{2\pi} e^{inw} |\widehat{a}^0(w)|^2 dw = r^0(n), \tag{3.61}$$

where r^0 is the autocovariance function of X^0. The approximation (3.61) indeed occurs but only at n sufficiently smaller than 2^K. For example, if $n < T$ and $|\widehat{a}^0(w)|^2$ is smooth in w, then

$$|r^0(n) - \widetilde{r}^0(n)| \le \frac{1}{2\pi} \int_0^{2\pi} \sum_{m=0}^{2^K-1} \Big| |\widehat{a}^0(w)|^2 - \Big|\widehat{a}^0\Big(\frac{2\pi m}{2^K}\Big)\Big|^2 \Big| 1_{[\frac{2\pi m}{2^K}, \frac{2\pi(m+1)}{2^K})}(w) dw$$

$$+ \frac{1}{2\pi} \int_0^{2\pi} \sum_{m=0}^{2^K-1} \Big| e^{iwn} - e^{i\frac{2\pi mn}{2^K}} \Big| \Big|\widehat{a}^0\Big(\frac{2\pi m}{2^K}\Big)\Big|^2 1_{[\frac{2\pi m}{2^K}, \frac{2\pi(m+1)}{2^K})}(w) dw$$

$$\le \sup_{w \in (0,2\pi)} \Big| \frac{\partial |\widehat{a}^0(w)|^2}{\partial w} \Big| \frac{2\pi}{2^K} + \sup_{w \in (0,2\pi)} |\widehat{a}^0(w)|^2 \frac{2\pi T}{2^K}, \tag{3.62}$$

which is small when $T/2^K$ is small. This suggests that the first T values of $\widetilde{X}^0$ can be used to approximate X^0, with the resulting error in autocovariance being of the order $T/2^K$ by (3.62). The use of the first generated values in the context of orthogonal wavelet decompositions can also be found in Percival and Walden (2000), Section 9.2, but without the explicit connection to circular time series and the resulting error (3.62) above.

As we expect $\widehat{a}^j(0) = \sum_n a_n^j = \infty$ for long memory time series (this is the case, for example, for FARIMA$(0, s, 0)$ time series in Example 3.4.1), the discussion and arguments above need to be modified. One way to do this is to set $\widehat{a}^j(0) = 1$. Application of (3.59) then yields $\widetilde{X}^0$ with

$$\widehat{a}^0(w) = \begin{cases} 1, & w = 0, \\ \widehat{a}^0(w), & \text{otherwise.} \end{cases}$$

The error (3.62) could be studied in a similar way though, because $|\widehat{a}^0(w)|^2$ is no longer smooth at $w = 0$, its decay would be slower than $T/2^J$.

3.5.2 Maximum likelihood estimation

The approximate covariance factorization (3.36) discussed in Remark 3.3.5 naturally leads to the following Gaussian MLE based on AWD. Given the vector of observations

$\widetilde{X}^0 = (X_0, X_1, \ldots, X_{T-1})$, the negative log-likelihood is (up to additive and multiplicative constants)

$$\log|\widetilde{\Sigma}_\theta| + \widetilde{X}^0 \widetilde{\Sigma}_\theta^{-1} \widetilde{X}^{0\prime}, \tag{3.63}$$

where $\widetilde{\Sigma}_\theta$ is the covariance matrix of the model with unknown parameters θ, and $|\cdot|$ denotes the determinant. As in Remark 3.3.5, a vector $\widetilde{Y}_\theta$ of detail coefficients in approximate AWD can be written as

$$\widetilde{Y}_\theta = \widetilde{X}^0 M_\theta \tag{3.64}$$

for a matrix M_θ which depends on the model parameters θ. By (3.36), $\widetilde{\Sigma}_\theta^{-1} \approx M_\theta M_\theta'$ and hence the expression (3.63) is approximately equal to

$$\log|\widetilde{\Sigma}_\theta| + \widetilde{Y}_\theta \widetilde{Y}_\theta'. \tag{3.65}$$

Observe that $|\widetilde{\Sigma}_\theta|$ cannot be immediately simplified because the matrices M_θ are not orthogonal. To simplify this determinant, one can make a classical approximation of Grenander and Szego (1958), and consider

$$\frac{T}{\pi} \int_0^{2\pi} \log|\widehat{a}_\theta(w)| dw + \widetilde{Y}_\theta \widetilde{Y}_\theta'. \tag{3.66}$$

MLE based on AWD is achieved by minimizing this expression with respect to unknown parameters θ.

In Tables 3.3–3.4, we present MLE results based on AWD in several time series models, namely, AR(1), FARIMA$(0, s, 0)$, FARIMA$(1, s, 0)$ and MA(1). In the Model column, we indicate the AWD used for MLE through $\widehat{a}^j(w)$, and the type of optimization method used (grid search or the `Matlab` functions `fminsearch`, `fminbnd`). We tried different optimization methods because some results were sensitive to their choice, in particular, for FARIMA$(1, s, 0)$ models when using AWD (Table 3.3). We also consider a non-Gaussian, exponential distribution for the generated error terms in the MA(1) case (Table 3.4) and, for FARIMA$(0, s, 0)$ model, we report results with a superimposed linear trend $-1 + 0.5t$ (Table 3.3). The results are reported throughout in terms of the bias and the square root of the mean squared error of the estimators. (These are computed based on 1000 Monte Carlo replications.) For comparison, we also present MLE results based on standard Whittle approximations (Chapter 6 in Beran (1994)) and orthogornal wavelet decompositions (Percival and Walden (2000), Jensen (1999)). In the latter case, in particular, the variance of the detail terms at scale 2^j, $j = 1, \dots, J$, is approximated by

$$\frac{2^{j+1}}{2\pi} \int_{2\pi/2^j}^{2\pi/2^{j+1}} |\widehat{a}(w)|^2 dw. \tag{3.67}$$

The sample size T is the length of the considered time series, and N denotes the number of zero moments of the underlying Daubechies MRA.

The results of Tables 3.3–3.4 suggest that MLE based on AWD works quite well. It is generally comparable to Whittle MLE and is superior to it in the AR(1) case with $a_1 = \pm 0.9$. It is generally superior to MLE based on OWD which is likely to be the result of the approximation (3.67). Note also that increasing the number of zero moments (from 2 to 6) have generally made little difference in the results for AWD. Observe from Table 3.4 that trend is not ignored by MLE based on AWD. This occurs because of the boundary effect. We have tried several other ways of dealing with the boundary (mentioned in Remark 3.3.4) but the results did not lead to improvement. We are presently exploring finer MLE based on AWD where only coefficients unaffected by the boundary are considered, or where proper adjustments to the coefficients at the boundary are made.

3.6 Proofs of the main results

Proof of Proposition 3.3.1: The condition (3.18) ensures that the time series in (3.19) and (3.20) are well-defined (Theorem 4.10.1 and Remark 1 in Brockwell and Davis (1991), p. 154-155).

(i) We shall use the spectral representation (3.3) of the time series X. By Theorem 4.10.1 in Brockwell and Davis (1991), we obtain that

$$\Big(\downarrow_2 (\overline{U}_d * X) \Big)_n = \int_0^{2\pi} e^{i2nw} \widehat{b}(2w) \overline{\widehat{u}(w)} dZ(w) = \left(\int_0^{\pi} + \int_{\pi}^{2\pi} \right) e^{i2nw} \widehat{b}(2w) \overline{\widehat{u}(w)} dZ(w)$$

$$= \int_0^{2\pi} e^{inw}\widehat{b}(w)\overline{\widehat{u}\Big(\frac{w}{2}\Big)}dZ\Big(\frac{w}{2}\Big) + \int_0^{2\pi} e^{inw}\widehat{b}(w+2\pi)\overline{\widehat{u}\Big(\frac{w}{2}+\pi\Big)}dZ\Big(\frac{w}{2}+\pi\Big)$$

$$= \int_0^{2\pi} e^{i2nw}\widehat{b}(w)dZ_1(w)$$

with

$$dZ_1(w) = \overline{\widehat{u}\Big(\frac{w}{2}\Big)}dZ\Big(\frac{w}{2}\Big) + \overline{\widehat{u}\Big(\frac{w}{2}+\pi\Big)}dZ\Big(\frac{w}{2}+\pi\Big), \quad w \in (0, 2\pi).$$

Similarly,

$$\Big(\downarrow_2 (\overline{V}_d * X)\Big)_n = \int_0^{2\pi} e^{i2nw}dZ_2(w)$$

with

$$dZ_2(w) = \overline{\widehat{v}\Big(\frac{w}{2}\Big)}dZ\Big(\frac{w}{2}\Big) + \overline{\widehat{v}\Big(\frac{w}{2}+\pi\Big)}dZ\Big(\frac{w}{2}+\pi\Big), \quad w \in (0, 2\pi).$$

To prove (i), it is enough to show that Z_1 and Z_2 are orthogonal increment processes with $E|dZ_1(w)|^2 = E|dZ_2(w)|^2 = dw/2\pi$ and satisfying $EdZ_1(w)\overline{dZ_2(w')} = 0$. This follows by using the properties (3.5), (3.7) and (3.8) and orthogonal increments of Z as

$$EdZ_1(w)\overline{dZ_1(w')}$$

$$= \overline{\widehat{u}\Big(\frac{w}{2}\Big)}\widehat{u}\Big(\frac{w'}{2}\Big)EdZ\Big(\frac{w}{2}\Big)\overline{dZ\Big(\frac{w'}{2}\Big)} + \overline{\widehat{u}\Big(\frac{w}{2}+\pi\Big)}\widehat{u}\Big(\frac{w'}{2}+\pi\Big)EdZ\Big(\frac{w}{2}+\pi\Big)\overline{dZ\Big(\frac{w'}{2}+\pi\Big)}$$

$$= \left(\Big|\widehat{u}\Big(\frac{w}{2}\Big)\Big|^2 + \Big|\widehat{u}\Big(\frac{w}{2}+\pi\Big)\Big|^2\right)\frac{dw}{4\pi}1_{\{w=w'\}} = \frac{dw}{2\pi}1_{\{w=w'\}},$$

$$EdZ_2(w)\overline{dZ_2(w')} = \frac{dw}{2\pi}1_{\{w=w'\}},$$

by similar arguments, and

$$EdZ_1(w)\overline{dZ_2(w')} = \left(\overline{\widehat{u}\Big(\frac{w}{2}\Big)}\widehat{v}\Big(\frac{w'}{2}\Big) + \overline{\widehat{u}\Big(\frac{w}{2}+\pi\Big)}\widehat{v}\Big(\frac{w'}{2}+\pi\Big)\right)\frac{dw}{4\pi}1_{\{w=w'\}} = 0.$$

(ii) We establish (3.20) only at even times $n = 2s$. (The case $n = 2s+1$ can be proved in a similar way.) Using the spectral representation of Y above, we obtain that

$$(U_r * \uparrow_2 Y)_n = (\downarrow_2 U_r * Y)_s = \frac{1}{2}\int_0^{2\pi} e^{isw}\Big(\widehat{U}_r\Big(\frac{w}{2}\Big) + \widehat{U}_r\Big(\frac{w}{2}+\pi\Big)\Big)\widehat{b}(w)dZ_1(w)$$

$$= \frac{1}{2}\int_0^{2\pi} e^{isw}\Big(\widehat{a}\Big(\frac{w}{2}\Big)\widehat{u}\Big(\frac{w}{2}\Big) + \widehat{a}\Big(\frac{w}{2}+\pi\Big)\widehat{u}\Big(\frac{w}{2}+\pi\Big)\Big)dZ_1(w)$$

$$= \frac{1}{2}\int_0^{\pi} e^{i2sw}\widehat{a}(w)\widehat{u}(w)dZ_1(2w) + \frac{1}{2}\int_0^{\pi} e^{i2sw}\widehat{a}(w+\pi)\widehat{u}(w+\pi)dZ_1(2w)$$

$$= \frac{1}{2}\int_0^{2\pi} e^{inw}\widehat{a}(w)\widehat{u}(w)dZ_1(2w).$$

Similarly,

$$(V_r * \uparrow_2 Y)_n = \frac{1}{2}\int_0^{2\pi} e^{inw}\widehat{a}(w)\widehat{v}(w)dZ_2(2w).$$

Hence,

$$(U_r * \uparrow_2 Y)_n + (V_r * \uparrow_2 Y)_n =$$

$$\int_0^{2\pi} e^{inw}\widehat{a}(w)\left(\frac{1}{2}\widehat{u}(w)dZ_1(2w) + \frac{1}{2}\widehat{v}(w)dZ_2(2w)\right) = \int_0^{2\pi} e^{inw}\widehat{a}(w)\widehat{u}(w)dZ(w) = X_n,$$

since

$$\widehat{u}(w)dZ_1(2w) + \widehat{v}(w)dZ_2(2w) = |\widehat{u}(w)|^2 dZ(w) + \widehat{u}(w)\overline{\widehat{u}(w+\pi)}dZ(w+\pi)$$

$$+|\widehat{v}(w)|^2 dZ(w) + \widehat{v}(w)\overline{\widehat{v}(w+\pi)}dZ(w+\pi) = 2dZ(w).$$

□

PROOF OF PROPOSITION 3.3.2: We will establish first that approximations $X^j = X^j(p)$ and details $\xi^j = \xi^j(p)$ are well-defined. In fact, we will show that

$$|X_n^j| \le C(1+|n|)^D, \tag{3.68}$$

where a constant C may depend on j. This bound is trivial for $j=0$ since $X^0 = p$ is a polynomial of degree D. Suppose that (3.68) holds for $j-1$ and consider it with j. Then,

$$|X_n^j| \le \sum_k |U_{d,k}^j X_{n-k}^{j-1}| \le C_1 \sum_k (1+|k|)^{-D-2}(1+|n-k|)^D$$

$$\le C_2 \sum_k (1+|k|)^{-D-2}(1+|n|^D + |k|^D) \le C_3(1+|n|)^D,$$

where constants C_i may depend on j. Using (3.68) and the assumed bound for $V_{d,n}^j$, the argument above also shows that ξ^j is well-defined.

To prove (3.39), we will first establish the formula

$$\widehat{X}^j(\omega) = \frac{1}{2^j} \sum_{n=0}^{2^j-1} \left\{ \prod_{k=1}^{j} \overline{\widehat{U}_d^k\Big(\frac{\omega}{2^{j+1-k}} + b_{n,k}\Big)} \right\} \widehat{p}\Big(\frac{\omega}{2^j} + \frac{n\pi}{2^{j-1}}\Big), \tag{3.69}$$

where $b_{n,k} \in [0, 2\pi)$. Since p is not in $l^2(\mathbb{R})$, the use of $\widehat{p}$ has to be clarified. Here and below, equations in the "spectral domain" should be interpreted through the "time domain" where, in particular, all products of Fourier transforms should be regarded as convolutions. The relation (3.69) is trivial for $j = 1$. Assume it holds for $j-1$ and consider it for j. Then,

$$\downarrow_2 \widehat{(\overline{U_d^j} * X^{j-1})}(w) = \frac{1}{2}\left(\overline{\widehat{U}_d^j\Big(\frac{\omega}{2}\Big)} \widehat{X}^{j-1}\Big(\frac{\omega}{2}\Big) + \overline{\widehat{U}_d^j\Big(\frac{\omega}{2}+\pi\Big)} \widehat{X}^{j-1}\Big(\frac{\omega}{2}+\pi\Big) \right)$$

$$= \frac{1}{2^j} \sum_{n=0}^{2^{j-1}-1} \prod_{k=1}^{j} \Bigg(\overline{\widehat{U}_d^k\Big(\frac{\omega}{2^{j+1-k}} + b_{n,k}\Big)} \widehat{p}\Big(\frac{\omega}{2^j} + \frac{n\pi}{2^{j-2}}\Big) +$$

$$\overline{\widehat{U}_d^k\Big(\frac{\omega}{2^{j+1-k}} + b'_{n,k}\Big)} \widehat{p}\Big(\frac{\omega}{2^j} + \frac{\pi}{2^{j-1}} + \frac{n\pi}{2^{j-2}}\Big) \Bigg)$$

$$= \frac{1}{2^j} \sum_{n=0}^{2^{j-1}-1} \prod_{k=1}^{j} \Bigg(\overline{\widehat{U}_d^k\Big(\frac{\omega}{2^{j+1-k}} + b_{n,k}\Big)} \widehat{p}\Big(\frac{\omega}{2^j} + \frac{2n\pi}{2^{j-1}}\Big)$$

$$+ \overline{\widehat{U}_d^k\Big(\frac{\omega}{2^{j+1-k}} + b'_{n,k}\Big)} \widehat{p}\Big(\frac{\omega}{2^j} + \frac{(2n+1)\pi}{2^{j-1}}\Big) \Bigg)$$

$$= \frac{1}{2^j} \sum_{n=0}^{2^j-1} \prod_{k=1}^{j} \overline{\widehat{U}_d^k\Big(\frac{\omega}{2^{j+1-k}} + c_{n,k}\Big)} \widehat{p}\Big(\frac{\omega}{2^j} + \frac{n\pi}{2^{j-1}}\Big).$$

Since

$$\widehat{\xi}^j(\omega) = \frac{1}{2}\left(\overline{\widehat{V}_d^j\Big(\frac{\omega}{2}\Big)} \widehat{X}^{j-1}\Big(\frac{\omega}{2}\Big) + \overline{\widehat{V}_d^j\Big(\frac{\omega}{2}+\pi\Big)} \widehat{X}^{j-1}\Big(\frac{\omega}{2}+\pi\Big) \right)$$

and

$$\widehat{V}_d^j(\omega) = \overline{\Big(\frac{1}{\widehat{a}^{j-1}(\omega)}\Big)} \widehat{v}(\omega),$$

it suffices to prove that, for $n = 0, 1, ..., 2^{j-1}-1$,

$$\overline{\widehat{v}\Big(\frac{\omega}{2}\Big)} \widehat{p}\Big(\frac{\omega}{2^j} + \frac{2n\pi}{2^{j-1}}\Big) = 0 \tag{3.70}$$

and

$$\overline{\widehat{v}\Big(\frac{\omega}{2}+\pi\Big)}\widehat{p}\Big(\frac{\omega}{2^j}+\frac{(2n+1)\pi}{2^{j-1}}\Big)=0. \tag{3.71}$$

Observe that, by using (3.69), the relation (3.70) follows from

$$\overline{\widehat{v}\Big(2^{j-1}\Big(\frac{\omega}{2^j}+\frac{2n\pi}{2^{j-1}}\Big)\Big)}\widehat{p}\Big(\frac{\omega}{2^j}+\frac{2n\pi}{2^{j-1}}\Big)=\overline{\widehat{v}(2^{j-1}\omega')}\widehat{p}(\omega')=\overline{\widehat{v}(2^{j-1}\omega')}\widehat{p}(\omega')$$

$$=\overline{\widehat{v}_{0,N}(2^{j-1}\omega')}\prod_{k=2}^{j}(1+e^{i2^{j-k}w'})^N(1-e^{i\omega'})^N\widehat{p}(\omega')=0,$$

since $(1-e^{-i\omega'})^N\widehat{p}(\omega')=0$. A similar argument applies to (3.71). □

Model	θ_0	T	Whittle bias	Whittle rMSE	N	AWD bias	AWD rMSE	OWD bias	OWD rMSE
AR(1)§	$a_1 = 0.5$	2^{10}	-0.0010	0.0279	2	-0.0027	0.0277	-0.0347	0.0453
$\widehat{a}^j(w) \equiv 1$					6	-0.0036	0.0262	-0.0117	0.0299
		2^{14}	0.0002	0.0067	2	-0.0001	0.0070	-0.0344	0.0351
					6	-0.0005	0.0069	-0.0105	0.0126
	$a_1 = -0.9$	2^{10}	-0.0192	0.0711	2	0.0034	0.0140	-0.1620	0.2002
					6	0.0025	0.0140	-0.1374	0.1763
		2^{14}	0.0001	0.0033	2	0.0001	0.0035	-0.2314	0.2315
					6	0.0001	0.0034	-0.1886	0.2023
	$a_1 = 0.9$	2^{10}	0.0699	0.1238	2	-0.0044	0.0150	-0.0078	0.0490
					6	-0.0041	0.0146	0.0237	0.0811
		2^{14}	0.0387	0.0906	2	-0.0002	0.0034	-0.0152	0.0157
					6	-0.0004	0.0035	-0.0006	0.0265
FARIMA$(0, s, 0)$§	$s = 0.3$	2^{10}	0.0003	0.0246	2	-0.0069	0.0270	-0.0079	0.0256
$\widehat{a}^j(w) = (1 - e^{-iw})^{-s}$					6	-0.0055	0.0258	-0.0082	0.0253
		2^{14}	0.0002	0.0062	2	-0.0028	0.0068	-0.0026	0.0065
					6	-0.0027	0.0067	-0.0028	0.0068
FARIMA$(0, s, 0)$§	$s = 0.3$	2^{10}	0.6985	0.6999	6	0.6914	0.6970	0.6890	0.7062
$\widehat{a}^j(w) = (1 - e^{-iw})^{-s}$	$+(-1 + 0.5t)$	2^{14}	0.6985	0.6999	6	0.6914	0.6970	0.6890	0.7062
FARIMA$(1, s, 0)$§	$a_1 = 0.5$	2^{10}	-0.0215	0.1087	2	-0.3700	0.3799	0.0105	0.0772
$\widehat{a}^j(w) = (1 - e^{-iw})^{-s}$	$s = 0.3$				6	-0.2585	0.2662	0.0000	0.0836
$\widehat{s}$ results		2^{14}	0.0004	0.0173	2	-0.3470	0.3475	0.0243	0.0325
					6	-0.2310	0.2316	0.0074	0.0198
	$a_1 = -0.5$	2^{10}	-0.0010	0.0305	2	-0.2387	0.2452	-0.0238	0.0466
	$s = 0.3$				6	0.0083	0.0392	-0.0108	0.0403
		2^{14}	0.0003	0.0073	2	-0.2317	0.2321	-0.0210	0.0227
					6	0.0204	0.0229	-0.0079	0.0116

Model	θ_0	T	Whittle bias	Whittle rMSE	N	AWD bias	AWD rMSE	OWD bias	OWD rMSE
FARIMA$(1,s,0)^{\S}$	$a_1 = 0.5$	2^{10}	0.0174	0.1082	2	0.2622	0.2691	-0.0772	0.1163
$\widehat{a}^j(w) = (1-e^{-iw})^{-s}$	$s = 0.3$				6	0.2968	0.3015	-0.0282	0.0942
$\widehat{a}_1$ results		2^{14}	-0.0003	0.0194	2	0.2554	0.2559	-0.0763	0.0802
					6	0.2964	0.2968	-0.0227	0.0306
	$a_1 = -0.5$	2^{10}	0.0018	0.0324	2	0.2504	0.2587	0.0479	0.0727
	$s = 0.3$				6	0.0609	0.0726	0.0236	0.0660
		2^{14}	0.0001	0.0080	2	0.2475	0.2480	0.0417	0.0430
					6	0.0561	0.0569	0.0141	0.0174
FARIMA$(1,s,0)^{\dagger}$	$a_1 = 0.5$	2^{10}			2	-0.0815	0.1326		
$\widehat{a}^j(w) = (1-e^{-iw})^{-s}$	$s = 0.3$				6	-0.0882	0.1389		
$\widehat{s}$ results	$a_1 = -0.5$	2^{10}			2	-0.0092	0.0321		
	$s = 0.3$				6	-0.0084	0.0319		
FARIMA$(1,s,0)^{\dagger}$	$a_1 = 0.5$	2^{10}			2	0.0737	0.1296		
$\widehat{a}^j(w) = (1-e^{-iw})^{-s}$	$s = 0.3$				6	0.0809	0.1345		
$\widehat{a}_1$ results	$a_1 = -0.5$	2^{10}			2	0.0084	0.0355		
	$s = 0.3$				6	0.0063	0.0348		

Table 3.3: MLE results (§: **fminsearch** optimization; †: grid search; θ_0: true parameters; T: sample size with $2^{10} = 1024$ and $2^{14} = 16,384$; N: the number of zero moments in Daubechies MRA).

			Whittle			AWD		OWD	
Model	θ_0	T	bias	rMSE	N	bias	rMSE	bias	rMSE
MA(1)*	$b_1 = 0.5$	2^{10}	-0.0002	0.0283	2	-0.0006	0.0277	-0.0814	0.0949
$\widehat{a}^j(w) \equiv 1$					6	-0.0003	0.0267	-0.0294	0.0610
		2^{14}	0.0001	0.0065	2	0.0002	0.0066	-0.0853	0.0861
					6	0.0001	0.0067	-0.0323	0.0351
	$b_1 = 0.9$	2^{10}	-0.0062	0.0184	2	-0.0074	0.0191	-0.2932	0.3058
					6	-0.0069	0.0189	-0.1245	0.1908
		2^{14}	-0.0003	0.0035	2	-0.0005	0.0035	-0.3051	0.3057
					6	-0.0004	0.0034	-0.1645	0.1678
	$b_1 = -0.5$	2^{10}	0.0021	0.0273	2	-0.0014	0.0268	0.0389	0.0494
					6	-0.0016	0.0274	0.0083	0.0303
		2^{14}	0.0000	0.0070	2	0.0001	0.0067	0.0383	0.0390
					6	0.0001	0.0069	0.0068	0.0099
	$b_1 = -0.9$	2^{10}	0.0074	0.0194	2	0.0040	0.0185	0.0677	0.0749
					6	0.0053	0.0190	0.0110	0.0213
		2^{14}	0.0004	0.0036	2	0.0004	0.0036	0.0536	0.0541
					6	0.0004	0.0035	0.0027	0.0048
MA(1)*	$b_1 = 0.5$	2^{10}	0.0006	0.0276	2	-0.0003	0.0264	-0.0808	0.0960
$\widehat{a}^j(w) \equiv 1$	(exp)				6	-0.0013	0.0271	-0.0284	0.0683
		2^{14}	-0.0000	0.0068	2	0.0001	0.0066	-0.0854	0.0863
					6	0.0002	0.0067	-0.0308	0.0340
	$b_1 = -0.5$	2^{10}	-0.0000	0.0286	2	-0.0007	0.0274	0.0384	0.0529
					6	-0.0019	0.0271	0.0086	0.0296
		2^{14}	0.0002	0.0068	2	-0.0002	0.0068	0.0385	0.0396
					6	-0.0003	0.0066	0.0068	0.0098

Model	θ_0	T	Whittle		N	AWD		OWD	
			bias	rMSE		bias	rMSE	bias	rMSE
MA(1)*	$b_1 = 0.5$	2^{10}			2	-0.0014	0.0277		
$\widehat{a}^j(w) = 1 - (-b_1)^{2^j} e^{-iw}$					6	0.0007	0.0279		
		2^{14}			2	0.0002	0.0070		
					6	0.0001	0.0069		
	$b_1 = 0.9$	2^{10}			2	-0.0056	0.0173		
					6	-0.0076	0.0193		
		2^{14}			2	-0.0004	0.0035		
					6	-0.0005	0.0034		
	$b_1 = -0.5$	2^{10}			2	-0.0003	0.0280		
					6	0.0009	0.0272		
		2^{14}			2	-0.0004	0.0069		
					6	0.0003	0.0067		
	$b_1 = -0.9$	2^{10}			2	0.0046	0.0181		
					6	0.0049	0.0182		
		2^{14}			2	0.0003	0.0035		
					6	0.0004	0.0035		

Table 3.4: MLE results (*: **fminbnd** optimization; θ_0: true parameters; T: sample size with $2^{10} = 1024$ and $2^{14} = 16,384$; N: the number of zero moments in Daubechies MRA).

CHAPTER 4

On operator fractional Brownian motions

4.1 Introduction

Fractional Brownian motion (FBM), denoted $B_H = \{B_H(t)\}_{t\in\mathbb{R}}$ with $H \in (0,1)$, is a stochastic process characterized by the following three properties:

(i) Gaussianity;

(ii) self-similarity with parameter H;

(iii) stationarity of the increments.

By self-similarity, it is meant that the law of B_H scales as

$$\{B_H(ct)\}_{t\in\mathbb{R}} \stackrel{d}{=} \{c^H B_H(t)\}_{t\in\mathbb{R}}, \tag{4.1}$$

where $c > 0$. By stationary increments, it is meant that the process

$$\{B_H(t+h) - B_H(h)\}_{t\in\mathbb{R}}$$

has the same distribution for any increment size $h \in \mathbb{R}$. It may be shown that these three properties actually characterize FBM in the sense that it is the *unique* (up to a constant) such process for a given $H \in (0,1)$. FBM plays an important role in both theory and applications, especially in connection to long range dependence (Embrechts and Maejima (2002), Doukhan *et al.* (2003)).

We are interested here in multivariate counterparts of FBM, called operator fractional Brownian motions (OFBMs). In the multivariate context, OFBM $B_H = (B_{1,H}, \ldots, B_{n,H})^*$

$= \{(B_{1,H}(t), ..., B_{n,H}(t))^* \in \mathbb{R}^n, t \in \mathbb{R}\}$ is a collection of random vectors. It is also Gaussian and has stationary increments. But self-similarity is now replaced by

(ii') operator self-similarity.

A multivariate process B_H is called operator self-similar (o.s.s.) if (4.1) holds, where H is an invertible operator (for a general discussion, see Subsection 4.2.3). Operator self-similarity extends the usual self-similarity and was first studied thoroughly in Laha and Rohatgi (1982) and Hudson and Mason (1982). The theory of operator self-similarity bears a resemblance to that of operator stable measures (see Jurek and Mason (1993) and Meerschaert and Scheffler (2001)).

Examples of OFBMs have been studied in the past. They arise and are used in the context of multivariate time series and long range dependence (see, for example, Chung (2002), Marinucci and Robinson (1999, 2000)). Another context is that of queueing systems, where reflected OFBMs model the size of multiple queues in particular classes of queueing models, and are studied in problems related to, for example, large deviations (see Konstantopoulos and Lin (1996), Delgado (2007), Majewski (2003, 2005)). Other related papers study particular classes of OFBMs from a theoretical perspective (see, for example, Mason and Yimin (2002), Maejima (1994)).

Despite a growing interest in OFBMs, there is not a work that examines the general class of OFBMs, and a number of questions for OFBMs remain open. We address some of these questions here. More specifically, we establish integral representations of OFBMs (Section 4.3) and study their basic properties. In the multivariate case, the three properties (i), (ii') and (iii) do *not* characterize the distribution of OFBM. The derivation of integral representations of OFBMs is therefore quite different from the univariate case. We prove that OFBMs have a rigid dependence structure among components which we call Dichotomy Principle (Section 4.4). Finally, we also study questions of uniqueness for OFBMs (Section 4.5). It is known since the fundamental work of Hudson and Mason (1982) that the exponent H for the same o.s.s. process is typically not unique. We will examine here the results of Hudson and Mason (1982) for particular classes of OFBMs. Appendix A contains some known results on commutativity of operators, and Appendix B concerns the exponential of a matrix in Jordan normal form.

4.2 Preliminaries

We begin by introducing some notation and by considering some preliminaries on the exponential map and operator self-similarity that are used throughout the paper.

4.2.1 Some notation

In this paper, the notation and terminology for finite-dimensional operator theory will be prevalent over their matrix analogues. However, whenever convenient the latter will be used.

All with respect to the field $\mathbb{R}$, $M(n)$ or $M(n,\mathbb{R})$ is the vector space of all $n\times n$ operators (endomorphisms), $GL(n)$ or $GL(n,\mathbb{R})$ is the general linear group (invertible operators, or automorphisms), $O(n)$ is the orthogonal group of operators O such that $OO^* = I = O^*O$ (i.e., the adjoint operator is the inverse), $SO(n) \subseteq O(n)$ is the special orthogonal group of operators with determinant equal to 1, and $so(n)$ is the vector space of skew-symmetric operators (i.e., $A^* = -A$). The sign * always indicates the adjoint operator, regardless of whether the underlying field is $\mathbb{R}$ or $\mathbb{C}$. Matrix-wise, it should be interpreted as transposition or Hermitian transposition, accordingly. Otherwise, the notation will indicate the change to the field $\mathbb{C}$. For instance, $M(n,\mathbb{C})$ is the vector space of complex endomorphisms. Whenever it is said that $A \in M(n)$ has a complex eigenvalue or eigenspace, one is considering the operator embedding $M(n) \hookrightarrow M(n,\mathbb{C})$. We will say that two endomorphisms $A, B \in M(n)$ are *conjugate* (or similar) when there exists $P \in GL(n)$ such that $A = PBP^{-1}$. In this case, P is called a *conjugacy*. The expression $\text{diag}(\lambda_1, ..., \lambda_n)$ denotes the operator whose matrix expression has the values $\lambda_1, ..., \lambda_n$ on the diagonal and zeros elsewhere. An operator $U \in M(n,\mathbb{C})$ is said to be unitary when $UU^* = U^*U = I$. An operator $A \in M(n,\mathbb{C})$ is said to be normal if it commutes with its adjoint, that is, $AA^* = A^*A$. By the Spectral Theorem, an operator $A \in M(n,\mathbb{C})$ is normal if and only if there exists an orthonormal basis of eigenvectors of A for the underlying vector space. If the normal operator $A \in M(n)$ is self-adjoint, then such basis can be written with purely real coordinates.

4.2.2 The exponential map

The meaning of the expression c^H in (4.1) with $c > 0$ and $H \in M(n)$ is given through the notion of exponential map by setting $c^H := \exp(\log(c)H)$, where

$$\exp(A) = \sum_{k=0}^{\infty} \frac{A^k}{k!},$$

and this infinite series converges for all $A \in M(n)$ (see also Hausner and Schwartz (1968), pp. 59-60). A few remarks about the exponential map are of importance here.

(R1) Loosely speaking, an exponential map

$$\exp : g \to G$$

takes a vector space of operators $g \subseteq M(n)$ into a closed subgroup $G \subseteq GL(n)$ of operators. In this sense, it de-linearizes the vector space. For example,

$$\exp(M(n)) \subseteq GL(n), \quad \exp(so(n)) = SO(n). \tag{4.2}$$

In other words, the exponential of any operator is invertible, and the exponential of any skew-symmetric operator is an orthogonal operator with det = 1 (and vice-versa). Whenever well-defined (as in (4.2)), the inverse of the exponential map, appropriately called log map, may be considered.

(R2) More precisely, let G be a closed (sub)group of operators. Denote by $g = T(G)$ the tangent space of G, i.e., the set of $A \in M(n)$ such that

$$A = \lim_{n\to\infty} \frac{G_n - I}{d_n}, \quad \text{for some } \{G_n\} \subseteq G \text{ and some } 0 < d_n \to 0.$$

In this sense, g is, in fact, a linearization of G in a vicinity of I.

It can be shown (Jurek and Mason (1993), pp. 15-16) that the exp map takes g into G. The relation between G and g may be pictured as a hyperplane (the latter) touching a manifold (the former) at I. The group operations on $GL(n)$ are infinitely differentiable, so $GL(n)$ is a Lie group. The tangent space $M(n)$ endowed with the Lie Bracket $[A, B] = AB - BA$ is a Lie Algebra.

(R3) It is *not* true in general that $\exp(A + B) = \exp(A)\exp(B)$. This relation holds if A and B commute; however, commutativity is not a necessary condition (Horn and Johnson (1991), p. 435).

(R4) It is easily seen that, for invertible P, $e^{PAP^{-1}} = Pe^{A}P^{-1}$.

4.2.3 Operator self-similar processes

The definition of operator self-similarity is as follows.

Definition 4.2.1. A stochastic process $\{X(t)\}_{t\in\mathbb{R}}$ on a finite-dimensional vector space V (typically, $\mathbb{R}^n$) is said to be operator self-similar (o.s.s.) if it is continuous in law at each $t \neq 0$ and if for every $c > 0$ there exists a linear operator $A(c)$ on V and a vector $a(c)$ in V such that

$$\{X(ct)\}_{t\in\mathbb{R}} \stackrel{d}{=} \{A(c)X(t) + a(c)\}_{t\in\mathbb{R}}. \tag{4.3}$$

Throughout the paper, we will assume all processes to be proper, i.e., for each t the distribution is not contained in a proper subspace of V. Furthermore, we will only consider what is called strictly o.s.s. processes, in the sense that $a(c) \equiv 0$ (see Corollary 3, Hudson and Mason (1982)).

Theorems 1, 2 and 3 in Hudson and Mason (1982) give the general relation between $A(c)$ in (4.3) and an (operator) exponent H for the o.s.s. process X. They provide the conditions for the existence, the non-uniqueness and the restrictions on such operator H. For the reader's convenience, we will state and briefly relate them here.

The first theorem says that, just like in the univariate case, $A(c)$ in (4.3) can be interpreted in terms of a scaling law.

Theorem 4.2.1. *(Hudson and Mason (1982):* Existence of H*) Let $\{X(t)\}_{t\in\mathbb{R}}$ be a proper o.s.s. process. Then, there exists an operator H such that, for each $c > 0$, (4.1) holds.*

An operator H that satisfies (4.1) is called an *exponent* of the process X, and the set of all such H is denoted by $\mathcal{E}(X)$.

The non-uniqueness of H satisfying (4.1) depends on the symmetry group G_1 of X, which is defined as follows.

Definition 4.2.2. The symmetry group of an o.s.s. process X is the set G_1 of operators $A \in GL(n)$ such that

$$\{X(t)\}_{t\in\mathbb{R}} \stackrel{d}{=} \{AX(t)\}_{t\in\mathbb{R}} \tag{4.4}$$

Theorem 4.2.2. *(Hudson and Mason (1982):* Non-uniqueness of H*) Let* $\{X(t)\}_{t\in\mathbb{R}}$ *be a proper o.s.s. process. Then, for any* $H \in \mathcal{E}(X)$,

$$\mathcal{E}(X) = H + T(G_1), \tag{4.5}$$

where $T(G_1) = W\mathcal{L}_0W^{-1}$ *for some positive-definite operator* W *and some subspace* $\mathcal{L}_0$ *of* $so(n)$*. Consequently,* X *has a unique exponent if and only if* G_1 *is finite.*

It turns out that the symmetry group G_1 is always compact, which implies that there exists a positive definite self-adjoint operator W and a closed subgroup $\mathcal{O}_0$ of $O(n)$ such that $G_1 = W\mathcal{O}_0W^{-1}$ (see, for instance, Hudson and Mason (1982) pp. 285, 289). A process X that has maximal symmetry, i.e., such that $G_1 = WO(n)W^{-1}$, is called *elliptically symmetric.*

Theorem 4.2.3. *(Hudson and Mason (1982):* Admissibility of H*)* $H \in M(n)$ *is an exponent for some o.s.s. process* X *if and only if*

(i) *every eigenvalue of* H *has non-negative real part;*

(ii) *every eigenvalue of* H *having null real part is a simple root of the minimal polynomial of* H*.*

If $H \in M(n)$ satisfies the conditions in Theorem 4.2.3, it is called *admissible.*

4.3 Integral representations of OFBMs

In the univariate case, for fixed $H \in (0,1)$, the law of FBM is unique up to a constant. This follows in a standard way by using H-self-similarity and stationarity of the increments as

$$EX(s)X(t) = \frac{1}{2}(EX(t)^2 + EX(s)^2 - E(X(t) - X(s))^2)$$

$$= \frac{EX(1)^2}{2}\{|t|^{2H} + |s|^{2H} - |t-s|^{2H}\}.$$

The same arguments cannot be applied in the case of OFBM. In fact, for OFBM,

$$EX(t)X(s)^* + EX(s)X(t)^*$$

$$= EX(t)X(t)^* + EX(s)X(s)^* - E(X(t) - X(s))(X(t) - X(s))^*$$

$$= |t|^H\Gamma(1,1)|t|^{H^*} + |s|^H\Gamma(1,1)|s|^{H^*} - |t-s|^H\Gamma(1,1)|t-s|^{H^*},$$

and it is not true in general that $EX(t)X(s)^* = EX(s)X(t)^*$. In this sense, a given operator H does not characterize the law of OFBM. This also does not exclude the case where two different Hs lead to the same OFBM, and we will see in Section 4.5 below that this may happen.

Even though a fixed H does not determine the law of OFBM, an alternative characterization can be sought through integral representations of OFBMs. In the univariate case, it is well-known that FBM has the spectral representation

$$B_H(t) = \frac{1}{C_2(H)}\int_{\mathbb{R}} \frac{e^{ixt} - 1}{ix}|x|^{-(H-1/2)}\widetilde{B}(dx), \tag{4.6}$$

where $\widetilde{B}(x) = \widetilde{B}_1(x) + i\widetilde{B}_2(x)$ is a complex-valued Brownian motion such that $\widetilde{B}_1(-x) = \widetilde{B}_1(x)$ and $\widetilde{B}_2(-x) = -\widetilde{B}_2(x)$, and $C_2(H)$ is a normalizing constant (see, for instance, Samorodnitsky and Taqqu (1994), p. 328). The representation (4.6) also yields the law of FBM, and sheds light on its structure (that is, it says how it can be built from the usual BM). It is therefore natural to try to obtain integral representations for OFBMs. This is done through a number of results given next.

Definition 4.3.1. We will say a function $f : \mathbb{R} \to \mathbb{C}^n$ is *operator-homogeneous of degree* $K \in M(n)$ if, for $c > 0$,

$$f(cx) = c^K f(x), \quad x \in \mathbb{R}. \tag{4.7}$$

As with ordinary homogeneity, all operator-homogeneous functions of the same degree differ only by an operator constant, which is their value on the sphere. In fact, from (4.7),

$$f(c) = c^K f(1) =: c^K A, \quad c > 0, \quad A \in M(n, \mathbb{C}),$$

$$f(c) = (-c)^K f(-1) =: (-c)^K B, \quad c < 0, \quad B \in M(n, \mathbb{C}).$$

For our purposes, the value of f at zero is defined arbitrarily. Consequently, any operator-homogeneous function of degree K can be written in the form

$$f(x) = x_+^K A + x_-^K B. \tag{4.8}$$

In Theorem 4.3.1 below, we establish integral representations of OFBMs in the spectral domain. Before that, we state a technical lemma.

Lemma 4.3.1. *Let $\{\widetilde{Y}(x)\}_{x\in\mathbb{R}} \in \mathbb{C}^n$ be an orthogonal-increment process, and set $F_{ij}(dx) = E\widetilde{Y}_i(dx)\overline{\widetilde{Y}_j(dx)}$, where $i, j = 1, ..., n$ and $\widetilde{Y}_i$ is a component of $\widetilde{Y}$. If $F_{ii}(dx)$ and $F_{jj}(dx)$ are absolutely continuous with respect to the Lebesgue measure over a given interval, then so is $F_{ij}(dx)$.*

Proof. A consequence of the Cauchy-Schwartz Inequality. □

Theorem 4.3.1. *Let $\{B_H(t)\}_{t\in\mathbb{R}}$ be OFBM with o.s.s. exponent H, where the real parts of the characteristic roots of H are in the interval $(0, 1)$. Then,*

$$\{B_H(t)\}_{t\in\mathbb{R}} \stackrel{d}{=} \Big\{ \int_{\mathbb{R}} \frac{e^{itx} - 1}{ix} (x_+^{-D} A + x_-^{-D} \overline{A}) d\widetilde{B}(x) \Big\}_{t\in\mathbb{R}}, \tag{4.9}$$

where $D = H - I(1/2)$, $A \in GL(n, \mathbb{C})$, $\overline{A}$ is the matrix whose entries are the complex conjugates of the entries of A, and $\widetilde{B}(x) := \widetilde{B}_1(x) + i\widetilde{B}_2(x)$ is a complex-valued multivariate Brownian motion satisfying $\widetilde{B}_1(-x) = \widetilde{B}_1(x)$, $\widetilde{B}_2(-x) = -\widetilde{B}_2(x)$ and $Ed\widetilde{B}(x)d\widetilde{B}(x)^* = dx$.

Proof. For notational simplicity, set $X = B_H$. Since X has stationary increments, we have

$$X(t) - X(s) = \int_{\mathbb{R}} \frac{e^{itx} - e^{isx}}{ix} \widetilde{Y}(dx), \tag{4.10}$$

where $\widetilde{Y}(dx)$ is an orthogonal-increment random measure in $\mathbb{C}^n$ (see Doob (1990)). Since $\mathrm{Re}(h_k) > 0$ for all $k = 1, ..., n$, then $X(0) = 0$ a.s. (see Maejima and Mason (1994)). Therefore, X can be represented as

$$X(t) = \int_{\mathbb{R}} \frac{e^{itx} - 1}{ix} \widetilde{Y}(dx). \tag{4.11}$$

Moreover, since X is Gaussian, $\widetilde{Y}(dx)$ is a Gaussian random measure. Let

$$F_X(dx) = E\widetilde{Y}(dx)\widetilde{Y}(dx)^*$$

be the multivariate spectral distribution of $\widetilde{Y}(dx)$. The rest of the proof goes in three steps:

(i) showing the existence of a spectral density function,

(ii) decorrelating the measure $\widetilde{Y}(dx)$ by finding a filter based upon the spectral density function,

(iii) showing that the filter is an operator-homogeneous function.

Step (i): Since X is o.s.s. with exponent H,

$$X(ct) \overset{d}{=} c^H \int_{\mathbb{R}} \frac{e^{itx} - 1}{ix} \widetilde{Y}(dx). \tag{4.12}$$

On the other hand, through a change of variables $v = cx$,

$$X(ct) \overset{d}{=} \int_{\mathbb{R}} \frac{e^{itv} - 1}{iv} c\widetilde{Y}(c^{-1}dv). \tag{4.13}$$

In differential form, this means that

$$c^H \widetilde{Y}(dx) \overset{d}{=} c\widetilde{Y}(c^{-1}dx) \tag{4.14}$$

or, equivalently,

$$\widetilde{Y}(cdx) \overset{d}{=} c^{I-H}\widetilde{Y}(dx). \tag{4.15}$$

Thus, $F_X([0,c])$ can be written as

$$E\widetilde{Y}([0,c])\widetilde{Y}([0,c])^* = c^{I-H}F_X([0,1])(c^{I-H})^*, \tag{4.16}$$

for $c > 0$ without loss of generality. By Lemma 4.3.1, it suffices to prove that the individual F_{ii} are absolutely continuous. By the explicit formula for c^{I-H}, the individual entries $F_X([0,c])_{ij}$ in the expression on the right-hand side of (4.16) are linear combinations (with complex weights) of terms of the form

$$\frac{(\log(c))^l}{l!}c^{1-h_k}, \quad k = 1, ..., n, \quad l \in \mathbb{N} \tag{4.17}$$

(or their respective conjugate), or identically zero for $c > 0$. Thus, $F_X(c)$ is differentiable in c over $(0,\infty)$ since $F_X([0,c])_{ij} = F_X(c)_{ij} - F_X(0)_{ij}$.

We want to prove that $\widetilde{Y}(0) = 0$ a.s. We now proceed as in Maejima and Mason (1994). Since the real part of the eigenvalues of $I - H$ are strictly greater than zero, then by Proposition 2.1.(ii) in Maejima and Mason (1994) we have $\| t^{I-H} \| \to 0$ as $t \to 0$, where $\| . \|$ is the (complex) operator norm. Thus, by equation (4.15),

$$\| \widetilde{Y}(0) \| \stackrel{d}{=} \| c^{I-H}\widetilde{Y}(0) \| \leq \| c^{I-H} \| \ \| \widetilde{Y}(0) \| \ \to 0 \quad \text{as } t \to 0.$$

So, $\widetilde{Y}(0) = 0$ a.s., as claimed.

Now note that

$$\widetilde{Y}(c) \stackrel{d}{=} c^{I-H}Y(1), \tag{4.18}$$

and since

$$\| c^{I-H}\widetilde{Y}(1) \| \leq \| c^{I-H} \| \ \| \widetilde{Y}(1) \| \ \to 0 \quad \text{as } c \to 0,$$

we also have

$$c^{I-H}\widetilde{Y}(1) \to 0 \qquad \text{as } c \to 0 \tag{4.19}$$

(the same argument holds for $\widetilde{Y}(-c)$). Thus, (4.18), (4.19) and the fact that $\widetilde{Y}$ is Gaussian imply that

$$\widetilde{Y}(c) \xrightarrow{L^2} 0 = \widetilde{Y}(0) \quad \text{as } c \to 0$$

(i.e., $\widetilde{Y}$ is L^2-stochastically continuous at zero). Therefore,

$$F_X([-c, c]) \to 0 \quad \text{as } c \to 0,$$

because

$$F_X([-c, c]) = E\Big(\int_{-c}^{c} d\widetilde{Y}(x)\Big)\Big(\int_{-c}^{c} d\widetilde{Y}(x)\Big)^*$$

$$= E\Big(\int_{-c}^{0} d\widetilde{Y}(x) + \int_{0}^{c} d\widetilde{Y}(x)\Big)\Big(\int_{-c}^{0} d\widetilde{Y}(x) + \int_{0}^{c} d\widetilde{Y}(x)\Big)^*$$

$$= E\Big(\int_{-c}^{0} d\widetilde{Y}(x)\Big)\Big(\int_{-c}^{0} d\widetilde{Y}(x)\Big)^* + E\Big(\int_{0}^{c} d\widetilde{Y}(x)\Big)\Big(\int_{0}^{c} d\widetilde{Y}(x)\Big)^* \to 0 \quad \text{as } c \to 0,$$

where the third equality follows by the orthogonal increments of $\widetilde{Y}$. This implies that

$$F_X(0) - F_X(0^-) = \lim_{c \to 0} F_X(c) - F_X(-c) = \lim_{c \to 0} F_X([-c, c]) = 0.$$

As a consequence, for all $i = 1, ..., n$ we have that $F_{ii}(c)$ is differentiable for $c \neq 0$ and continuous at zero. Thus, a multivariate spectral density function $f_X(x)$ exists.

Step (ii): Since $f_X(x)$ is positive definite Hermitian-symmetric for every x, the Spectral Theorem yields a square root $\widehat{a}(x)$ of $f_X(x)$. Let $d\widetilde{B}(x)$ be a complex-valued multivariate Brownian motion as in the statement of the theorem. The random measure $\widehat{a}(x)d\widetilde{B}(x)$ is equal (in distribution) to $\widetilde{Y}(dx)$, since

$$E(\widehat{a}(x)d\widetilde{B}(x)d\widetilde{B}(x)^*\widehat{a}(x)^*) = \widehat{a}(x)\widehat{a}(x)^*dx = f_X(x)dx = F_X(dx).$$

This implies that X can also be represented as

$$X(t) \stackrel{d}{=} \int_{\mathbb{R}} \frac{e^{itx} - 1}{ix} \widehat{a}(x)d\widetilde{B}(x). \tag{4.20}$$

Step (iii): By rewriting (4.14) with $\widehat{a}(x)d\widetilde{B}(x)$, we obtain that

$$c^H\widehat{a}(x)d\widetilde{B}(x) \stackrel{d}{=} \widehat{a}\Big(\frac{x}{c}\Big)c^{I(1/2)}d\widetilde{B}(x),$$

whence

$$\widehat{a}(cx) = c^{-D}\widehat{a}(x). \tag{4.21}$$

This means that $\widehat{a}$ is operator-homogeneous of degree $K = -D$, which implies it has the form (4.8) and representation (4.9) holds. □

We next obtain integral representations of OFBMs in the time domain. We will use the following elementary result. We write $f \in L^2(\mathbb{R}, \mathbb{R}^{n^2})$ for a matrix-valued function f when $\int_{\mathbb{R}}[f(u) \circ f(u)]du < +\infty$, where $A \circ B := \text{trace}(A^*B)$. The Fourier transform of $f \in L^2(\mathbb{R}, \mathbb{R}^{n^2})$ is defined as $\widehat{f}(x) = \int_{\mathbb{R}} e^{-ixu}f(u)du$.

Lemma 4.3.2. *Let $f, g \in L^2(\mathbb{R}, \mathbb{R}^{n^2})$. Then, the Plancherel identity holds, i.e.,*

$$\int_{\mathbb{R}} f(u)g(u)^*du = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}(x)\widehat{g}(x)^*dx, \tag{4.22}$$

where $\widehat{f}$ and $\widehat{g}$ are the component-wise Fourier transforms of f and g.

Theorem 4.3.2. *Let $\{B_H(t)\}_{t\in\mathbb{R}}$ be OFBM with o.s.s. exponent H. Then,*

$$\{B_H(t)\}_{t\in\mathbb{R}} \stackrel{d}{=} \Big\{\int_{\mathbb{R}} \Big(((t-u)_+^D - (-u)_+^D)M + ((t-u)_-^D - (-u)_-^D)N\Big)\, dB(u)\Big\}_{t\in\mathbb{R}}, \tag{4.23}$$

where $D = H - I(1/2)$, $(M, N) \in (GL(n) \cup \{0\}) \times (GL(n) \cup \{0\})\backslash\{(0, 0)\}$ and $B(u)$ is a real-valued, multivariate Brownian motion.

Proof. Let X and $\widetilde{X}$ denote the processes on the right-hand side of (4.9) and (4.23), respectively. It suffices to show that the covariance structures of X and $\widetilde{X}$ are the same. For simplicity, we only consider $\widetilde{X}$ in (4.23) with $N = 0$ and show that it has the representation (4.9).

As in the univariate case, one can show that

$$\int_{\mathbb{R}} \Big((t-u)_+^D - (-u)_+^D\Big) e^{-iux} du = (e^{-itx}-1)|x|^{-(D+I)}\Gamma(D+I)e^{i\mathrm{sign}(x)\pi(D+I)/2},$$

where

$$\Gamma(K) = \int_0^\infty e^{-x}x^{K-I}dx$$

converges absolutely if the characteristic roots of the operator K are greater than zero. Then, by Lemma 4.3.2,

$$E\widetilde{X}(s)\widetilde{X}(t)^* = \int_{\mathbb{R}} ((s-u)_+^D - (-u)_+^D)MM^*((t-u)_+^{D^*} - (-u)_+^{D^*})du$$

$$= \frac{1}{2\pi}\int_{\mathbb{R}} \frac{(e^{-isx}-1)(e^{itx}-1)}{|x|^2}(|x|^{-D}\Gamma(D+I)e^{i\mathrm{sign}(x)\pi(D+I)/2})MM^*\cdot$$

$$\cdot(e^{-i\mathrm{sign}(x)\pi(D^*+I)/2}\Gamma(D+I)^*|x|^{-D^*})dx.$$

This is also the covariance structure for X in (4.9) with $A := \Gamma(D+I)e^{i\pi(D+I)/2}M$.

Note also that for $\widetilde{X}$ to take values in $\mathbb{R}^n$, it is necessary that $(M,N) \in (GL(n) \cup \{0\}) \times (GL(n)\cup\{0\})\backslash\{(0,0)\}$. In fact, any operators $\widetilde{M}, \widetilde{N} \in M(n,\mathbb{C})$ have the form $\widetilde{M} = M_1 + iM_2$, $\widetilde{N} = N_1 + iN_2$, where $M_1, M_2, N_1, N_2 \in M(n)$ (actually, we must have M_1 or $N_1 \in GL(n)$, otherwise $\widetilde{X}$ cannot be a proper process in $\mathbb{R}^n$). By considering the expression (4.23) with $M := \widetilde{M}$, $N := \widetilde{N}$, it follows that $\widetilde{X}(t) \in \mathbb{R}^n$ for a given t if and only if $\int_{\mathbb{R}} \Big(((t-u)_+^D - (-u)_+^D)M_2 + ((t-u)_-^D - (-u)_-^D)N_2\Big)\, dB(u) = 0$, which does not hold a.s. unless $M_2 = N_2 = 0$. □

Remark 4.3.1. For what operators H is the time domain representation (4.23) of OFBM well-defined? Let $D = H - (1/2)I$. The integral (4.23) is well-defined as long as the integrand is in $L^2(\mathbb{R})$. Using the Jordan form of $D = PJP^{-1}$, where $P \in GL(n,\mathbb{C})$ and J is in Jordan normal form with the eigenvalues d_l, $l = 1,...,n$ of D, the square-integrability follows if $|t-u|^J - |-u|^J$ is in $L^2(\mathbb{R})$. By Appendix B.2, it is enough to have the functions

$$(\log|t-u|)^m|t-u|^{d_l} - (\log|-u|)^m|-u|^{d_l} \tag{4.24}$$

in $L^2(\mathbb{R})$, where $m = 1, ..., n_{J_{d_l}}$ and $n_{J_{d_l}}$ is the size of the Jordan block J_{d_l} of D. The functions (4.24) are in $L^2(\mathbb{R})$ when $d_l \in (-1/2, 1/2)$.

4.4 Dichotomy principle

As in the univariate case, increments of OFBM are stationary and have a special name.

Definition 4.4.1. Let $\{B_H(t)\}_{t\in\mathbb{R}}$ be an OFBM. The increment process

$$\{Y_H(t)\}_{t\in T} \stackrel{d}{=} \{B_H(t+1) - B_H(t)\}_{t\in T}, \quad \text{where } T = \mathbb{Z} \text{ or } \mathbb{R},$$

is called Operator Fractional Gaussian Noise (OFGN).

From Theorem 4.3.1, the spectral representation of OFGN in continuous time is

$$\{Y_H(t)\}_{t\in\mathbb{R}} \stackrel{d}{=} \Big\{ \int_{\mathbb{R}} e^{itx} \frac{e^{ix}-1}{ix} (x_+^{-D} A + x_-^{-D} \overline{A}) d\widetilde{B}(x) \Big\}_{t\in\mathbb{R}}, \tag{4.25}$$

where $D = H - (1/2)I$. Then, the spectral density of $\{Y_H(t)\}_{t\in\mathbb{R}}$ is

$$f_H(x) = \frac{|e^{ix}-1|^2}{|x|^2} (x_+^{-D} AA^* x_+^{-D^*} + x_-^{-D} \overline{AA^*} x_-^{-D^*}), \quad x \in \mathbb{R}, \tag{4.26}$$

since the cross terms are zero.

In discrete time, observe that

$$EY_H(0)Y_H(n) = \int_0^{2\pi} e^{inx} \sum_{k=-\infty}^{\infty} f_{Y_H}(x + 2\pi k) dx, \quad n \in \mathbb{Z}. \tag{4.27}$$

Then, the spectral density $\{Y_H(n)\}_{n\in\mathbb{R}}$ is

$$g_{Y_H}(x) = 2(1-\cos(x)) \sum_{k=-\infty}^{\infty} \frac{1}{|x+2\pi k|^2} \Big((x+2\pi k)_+^{-D} AA^* (x+2\pi k)_+^{-D^*}$$

$$+ (x+2\pi k)_-^{-D} \overline{AA^*} (x+2\pi k)_-^{-D^*} \Big), \quad x \in (0, 2\pi). \tag{4.28}$$

The form (4.28) of the spectral density leads to the following result.

Theorem 4.4.1. *Let H be a normal operator with eigenvalues h_l, $l = 1, ..., n$, such that*

$$1/2 < \mathrm{Re}(h_l) < 1, \quad l = 1, ..., n. \tag{4.29}$$

Let $g_{Y_H}(x) = \{g_{Y_H}(x)_{ij}\}$ be the spectral density (4.28) of OFGN in discrete time. Then, either

(i) $g_{Y_H}(x)_{ij}$ *diverges as* $x \to 0$, *or*

(ii) $g_{Y_H}(x)_{ij} \equiv 0$, $x \in (0, 2\pi)$.

Proof. Let $D = H - (1/2)I$ and denote the eigenvalues of D by $d_1, ..., d_n \in \mathbb{C}$. By the assumption, $0 < \mathrm{Re}(d_l) < 1/2$. Since D is normal,

$$x^{-D} = P \,\mathrm{diag}(x^{-d_1}, ..., x^{-d_n})\, P^*,$$

where $P \in U(n)$. Therefore, each term of the summation in (4.28) involves the matrix expression

$$P \,\mathrm{diag}((x + 2\pi k)^{-d_1}, ..., (x + 2\pi k)^{-d_n})\, P^* A A^* P \,\mathrm{diag}(\overline{(x + 2\pi k)^{-d_1}}, ..., \overline{(x + 2\pi k)^{-d_n}})\, P^*,$$

whose entries are linear combinations of products of the complex power functions $(x + 2\pi k)^{-d_1}, ..., (x + 2\pi k)^{-d_n}$ and their complex conjugates. The behavior of $g_{Y_H}(x)$ as $x \to 0^+$ is governed by the term

$$\frac{2(1 - \cos(x))}{x^2} P \,\mathrm{diag}(x^{-d_1}, ..., x^{-d_n})\, P^* A A^* P \,\mathrm{diag}(\overline{x^{-d_1}}, ..., \overline{x^{-d_n}})\, P^*.$$

As $x \to 0^+$, $\frac{2(1-\cos(x))}{x^2} \to 1$. Therefore, since $\mathrm{Re}(d_l) > 0$ for $l = 1, ..., n$, $g_{Y_H}(x)_{ij}$ diverges as a power function as $x \to 0^+$ unless it is identically zero over the entire spectral domain. □

Remark 4.4.1. We expect Theorem 4.4.1 to hold in the general case where the characteristic roots $h_1, ..., h_n$ of H all have real parts between 1/2 and 1. Indeed, using the explicit form for x^J, where J is a Jordan block (see (B.9)), each term in the summation (4.28) is a linear combination of functions of the form $(\log |x|)^m |x|^{-d_l}$, where $d_l = h_l - 1/2$.

The range $1/2 < h_l < 1$ is known as that of long range dependence. Theorem 4.4.1 thus states that, if OFGN is long range dependent in the sense of (4.29), then cross correlation between any two components is characterized by the following dichotomy: it is either long range dependent (with diverging cross spectra at zero) or identically equal to zero. From a practical perspective, this means that the class of OFGN may not be flexible enough to capture multivariate long range dependence structures.

4.5 On the non-uniqueness of exponents

Theorem 4.2.2 states that the class $\mathcal{E}(X)$ of exponents of an o.s.s. process X may contain more than one operator, and that this depends on the symmetry group G_1 of X through its tangent space $T(G_1)$. We examine here G_1 and related questions of (non-)uniqueness for particular classes of OFBMs (Section 4.5.2). We start with some preliminary remarks.

4.5.1 Preliminary remarks

The idea that operator exponents are not unique can be understood from at least two inter-related perspectives: properties of operator (matrix) exponents and distributional properties of o.s.s. processes. From the first perspective, consider for example matrices of the form

$$\begin{pmatrix} 0 & s \\ -s & 0 \end{pmatrix} \in so(2),$$

where $s \in \mathbb{R}$. Being normal, these operators can be diagonalized as $L_s = P\Lambda_s P^*$, where $P \in O(2)$ and $\Lambda_s = \text{diag}(is, -is)$. In particular, $\exp\{L_{2\pi k}\} = I$, since $\exp\{i2\pi k\} = 1$. Since L_s and $L_{s'}$ commute for any $s, s' \in \mathbb{R}$, this yields

$$\exp(L_s) = \exp(L_{2\pi k})\exp(L_s) = \exp(L_{2\pi k} + L_s), \tag{4.30}$$

and shows the potential non-uniqueness of operator exponents from purely operator (matrix) properties. Note also that the situation here is quite different from the 1-dimensional case: in one dimension, the same is possible but only with complex exponents, whereas here the operators $L_{2\pi k}$ have all real entries.

From the perspective of distributional properties, we illustrate several ideas through the following simple example.

Example 4.5.1. (Single parameter OFBM) Consider OFBM B_H with exponent $H = \mathrm{diag}(h, ..., h)$, $h \in (0,1)$, and $M = I$, $N = 0$ in the representation (4.23). It will be called a single parameter OFBM. Note that, in this case, $EB_H(t)B_H(s) =: \Gamma(t,s) = \Gamma_h(t,s)I$, where $\Gamma_h(t,s)$ is the covariance structure of a univariate FBM with parameter h. Since B_H is Gaussian, $O \in G_1$ if and only if $O\Gamma(t,s)O^* = \Gamma(t,s)$. In the case of single parameter OFBM, this is equivalent to $OO^* = I$ or, since O has an inverse (B_H is assumed proper), $OO^* = O^*O = I$. In other words, $G_1 = O(n)$, that is, single parameter OFBM is elliptically symmetric, and

$$\mathcal{E}(B_H) = H + so(n).$$

Thus, the exponents for a single parameter OFBM are not unique. From another angle, for a given $c > 0$ and $L \in so(n)$, we have $L\log(c) \in so(n)$ and hence $\exp\{L\log(c)\} = c^L \in O(n) = G_1$. Then,

$$\{B_H(ct)\}_{t\in\mathbb{R}} \stackrel{d}{=} \{c^H B_H(t)\}_{t\in\mathbb{R}} \stackrel{d}{=} \{c^H c^L B_H(t)\}_{t\in\mathbb{R}} \stackrel{d}{=} \{c^{H+L} B_H(t)\}_{t\in\mathbb{R}},$$

which also shows that the exponents are not unique.

4.5.2 Symmetry group and non-uniqueness of exponents in the case $n = 2$

We study here questions of non-uniqueness in the case $n = 2$. This case is natural to consider first because $G_1 \subseteq WO(n)W^{-1}$ (see Section 4.2.3) and orthogonal operators in $O(n)$ can be quasi-diagonalized in terms of 1- and 2-dimensional orthogonal operators. The case $n = 2$ has also been studied separately in a related work on operator stable measures (Hudson and Mason (1981)).

We have already remarked that the symmetry group G_1 of o.s.s. processes is contained in a set $WO(n)W^{-1}$ for some positive definite self-adjoint operator W. This implies that the symmetry group of the o.s.s. process

$$\widetilde{X}(t) := W^{-1}X(t) \tag{4.31}$$

is contained in $O(n)$. Theorems 4.5.1, 4.5.2 and Lemmas 4.5.1, 4.5.2 below shed light on the structure of Gaussian o.s.s. processes of the form $\widetilde{X}$ or, equivalently, for the cases where $W = I$. In particular, these results also apply to OFBMs, which are Gaussian. The proofs of the results below often use Appendix B.1 on commutativity of operators. Note also that, for a Gaussian process $\widetilde{X}$ with $G_1 \subseteq O(n)$, we have $O \in G_1$ if and only if $O\Gamma(t,s) = \Gamma(t,s)O$ for $s, t \in \mathbb{R}$, where

$$\widetilde{\Gamma}(t,s) = E\widetilde{X}(t)\widetilde{X}(s)$$

is the covariance structure of $\widetilde{X}$.

Theorem 4.5.1. *For a 2-dimensional, Gaussian, o.s.s. process $\widetilde{X}$ as in (4.31), $SO(2) \cap G_1$ is:*

(i) $\{I, -I\}$, *or*

(ii) $SO(2)$.

Proof. Note that the eigenvectors of *any* rotation $SO(2)\backslash\{I, -I\}$ must be of the form

$$u = \frac{\sqrt{2}}{2}\; e^{i\tau}\begin{pmatrix} 1 \\ -i \end{pmatrix}, \quad v = \frac{\sqrt{2}}{2}\; e^{i\beta}\begin{pmatrix} 1 \\ i \end{pmatrix}, \tag{4.32}$$

where τ and β are arbitrary angles in $[0, 2\pi)$. Assume there is a rotation $O \in SO(2)\backslash\{I, -I\}$ such that $O \in G_1$. This O must commute with $\widetilde{\Gamma}(t,s)$ for every $s, t \in \mathbb{R}$. Since the eigenvalues of O are different, then by Corollary B.1.2 in Appendix B.1.2, $\widetilde{\Gamma}(t,s)$ must have the same Jordan canonical form structure as O. Therefore, $\widetilde{\Gamma}(t,s) = U\mathrm{diag}(\widetilde{\Gamma}_1(t,s), \widetilde{\Gamma}_2(t,s))U^*$, for $U := (u, v)$ and two univariate functions $\widetilde{\Gamma}_1(t,s)$ and $\widetilde{\Gamma}_2(t,s)$. This shows that $\widetilde{\Gamma}(t,s)$ commutes with any other rotation in $SO(2)$. □

Theorem 4.5.2. *For a 2-dimensional, Gaussian, o.s.s. process $\widetilde{X}$ as defined in (4.31), $(O(2)\backslash SO(2)) \cap G_1$ is:*

(i) $\emptyset$, *or*

(ii) $\{R_1, R_2\}$, *where R_1, R_2 are the reflections around two given orthogonal axes, or*

(iii) $O(2)\backslash SO(2)$.

Proof. Assume $\widetilde{\Gamma}(t,s)$ commutes with a reflection R_1. Then, it also commutes with the corresponding reflection R_2. If $\widetilde{\Gamma}(t,s)$ commutes with a third reflection R_3, it must commute with all $O(2)\backslash SO(2)$, since R_3 must have different (real, orthonormal) eigenspaces. □

Lemma 4.5.1. *There is no 2-dimensional, Gaussian, o.s.s. process $\widetilde{X}$ as in (4.31) such that $G_1 = SO(2) \cup \{R_1, R_2\}$, where $R_1, R_2 \in O(2)\backslash SO(2)$.*

Proof. If $G_1 = SO(2) \cup \{R_1, R_2\}$, then $\widetilde{\Gamma}(t,s)$ must have the same eigenspaces as $SO(2)$. If $\widetilde{\Gamma}(t,s)$ also commutes with R_1, then it must also commute with all $O(2)\backslash SO(2)$, since in this case it must be diagonalizable with two equal real eigenvalues. □

Lemma 4.5.2. *There is no 2-dimensional, Gaussian, o.s.s $\widetilde{X}$ as in (4.31) such that $G_1 = \{I, -I\} \cup O(2)\backslash SO(2)$.*

Proof. If $G_1 = \{I, -I\} \cup (O(2)\backslash SO(2))$, then $\Gamma(t,s)$ cannot have more than two reflections R_1, R_2 without being diagonalizable with two equal eigenvalues. This implies $G_1 = O(2)$. □

Theorems 4.5.1, 4.5.2 and Lemmas 4.5.1, 4.5.2 combined give the following theorem on the classification of 2-dimensional, Gaussian, o.s.s. processes.

Theorem 4.5.3. *2-dimensional, Gaussian, o.s.s. processes can be classified according to the symmetry group G_1 under four types, namely, the ones whose G_1 is conjugate by a positive definite operator W to*

(I.a) $\{I, -I\}$;

(I.b) $SO(2)$;

(II.a) $\{I, -I, R_1, R_2\}$, *where R_1 and R_2 are the two reflection operators associated with a pair of orthogonal eigenspaces;*

(II.b) $O(2)$.

Only processes of types (I.b) *and* (II.b) *have non-unique exponents.*

Definition 4.5.1. When a symmetry group G_1 is of the type (I.a), (I.b), (II.a) or (II.b), we will say that it (or the corresponding o.s.s. process) is minimal, rotational, trivial or maximal, respectively.

The next results will provide additional insight into the structure of exponents of Gaussian, o.s.s. process X. We shall use the following theorem due to Maejima (1998).

Theorem 4.5.4. *(Maejima (1998)) There exists $H_0 \in \mathcal{E}(X)$ such that*

$$H_0 A = A H_0$$

for all $A \in G_1$.

The following simple result will also be useful.

Lemma 4.5.3. *If $\mathcal{E}(X)$ is not unique, then $T(G_1) = W so(2) W^{-1}$ for some positive definite operator W.*

Proof. If $\mathcal{E}(X)$ is not unique, Theorem 4.5.3 implies that $WSO(2)W^{-1} \subseteq G_1$ for some positive definite W. Therefore, $T(G_1) = W\mathcal{L}W^{-1}$ is a non-trivial subspace of $W so(2) W^{-1}$. The only subspaces of $W so(2) W^{-1}$ are $\{0\}$ and $W so(2) W^{-1}$ itself, which implies the result. □

The next result clarifies the structure of exponents when $\mathcal{E}(X)$ is not unique.

Theorem 4.5.5. *Let H_0 be the commuting operator in Theorem 4.5.4, and let W be a positive definite operator G_1 such that $G_1 = W\mathcal{O}W^{-1}$ for some $\mathcal{O} \subseteq O(2)$. If $\mathcal{E}(X)$ is not unique, then*

$$H_0 = WU\mathrm{diag}(h, \overline{h})U^* W^{-1}, \tag{4.33}$$

where the columns of $U \in U(2)$ are eigenvectors of $SO(2)$. In particular,

$$\mathcal{E}(X) = W(U\mathrm{diag}(h, \overline{h})U^* + so(2))W^{-1}. \tag{4.34}$$

Moreover, for any $H \in \mathcal{E}(X)$, $W^{-1}HW$ is normal, and $H = \mathrm{Re}(h)I \in \mathcal{E}(X)$.

Proof. If $\mathcal{E}(X)$ is not unique, then by Theorem 4.5.3, H_0 commutes with $WSO(2)W^{-1}$. In particular, H_0 commutes with WOW^{-1} for $O \in SO(2)\backslash\{I, -I\}$. Such O is diagonalizable with two complex conjugate eigenvalues, which implies that the eigenspaces of WOW^{-1} have dimension one. By Corollary B.1.2 in Appendix B.1.2, the eigenspaces of WOW^{-1} are also eigenspaces of the operator H_0. Thus, H_0 can be written as $WU\mathrm{diag}(h_1, h_2)U^*W^{-1}$. Note that $W^{-1}H_0W \in GL(2, \mathbb{R})$. Therefore, since h_1, h_2 are also the characteristic roots of the operator $U\mathrm{diag}(h_1, h_2)U^*$, we have $h_1 = h_2$, and thus (4.33) holds. Since $\mathcal{E}(X)$ is not unique, Lemma 4.5.3 yields $T(G_1) = Wso(2)W^{-1}$, which gives (4.34).

For $H \in \mathcal{E}(X)$, $W^{-1}HW$ is normal by using (4.33). In particular, we may choose the operator exponent $H := H_0 + WL_{-\mathrm{Im}(h)}W^{-1} = \mathrm{Re}(h)I$, where

$$L_{-\mathrm{Im}(h)} = \begin{pmatrix} 0 & -\mathrm{Im}(h) \\ \mathrm{Im}(h) & 0 \end{pmatrix}.$$

□

The unique exponent of the trivial case is described next.

Theorem 4.5.6. *Let H_0 be the commuting operator in Theorem 4.5.4, and let W be a positive definite operator G_1 such that $G_1 = W\mathcal{O}_0W^{-1}$ for some $\mathcal{O} \subseteq O(2)$. If G_1 is trivial, then*

$$H_0 = WO\mathrm{diag}(h_1, h_2)O^*W^{-1}.$$

where $O \in SO(2)$, and h_1, h_2 are the two eigenvalues of H_0. In particular, $W^{-1}H_0W$ is normal.

Proof. As in the proof of Theorem 4.5.5, H_0 must commute with WR_1W^{-1} and WR_2W^{-1}, where R_1 and R_2 are two reflections as in Theorem 4.5.3. Finally, note that R_1 and R_2 can both be diagonalized with the same real orthonormal eigenvectors, and eigenvalues 1 and -1. □

The following proposition and several examples specifically concern OFBM.

Proposition 4.5.1. *Up to a positive definite operator, every maximal symmetry 2-dimensional OFBM is a single-parameter OFBM.*

Proof. Let X be a maximal OFBM and $\Gamma(t,s) = EX(t)X(s)^*$. We may suppose without loss of generality that $G_1 = O(2)$. Since $\Gamma(t,s)$ commutes with all $O(2)$, there is $\gamma(t,s) \in \mathbb{R}$ such that

$$\Gamma(t,s) = \gamma(t,s)I. \tag{4.35}$$

Observe next that

$$\Gamma(ct,cs) = c^{H_0}\Gamma(t,s)c^{H_0^*},$$

where H_0 is the operator given by Theorem 4.5.4. Since H_0 must commute with all $O(2)$, it must be of the form $H_0 = hI$ for some $h \in \mathbb{R}$. As a consequence, $c^{H_0}c^{H_0^*} = c^{2H_0}$, $\gamma(ct,cs) = c^{2h}\gamma(t,s)$ and $\gamma(t,t) = t^{2h}\gamma(1,1) =: t^{2h}\sigma^2$. By using the symmetry of $\Gamma(t,s)$ and the stationarity of the increments, when (without loss of generality) $t > s > 0$,

$$(t-s)^{2h}\sigma^2 I = E[X(t-s)X(t-s)^*] = E[(X(t)-X(s))(X(t)-X(s))^*]$$

$$= EX(t)X(t)^* - EX(s)X(t)^* - EX(t)X(s)^* + EX(s)X(s)^*$$

$$= \Gamma(t,t) - \Gamma(s,t) - \Gamma(t,s) + \Gamma(s,s) = t^{2h}\sigma^2 I - 2\Gamma(s,t) + s^{2h}\sigma^2 I.$$

This yields

$$\gamma(t,s) = \frac{\sigma^2}{2}(|t|^{2h} + |s|^{2h} - |t-s|^{2h}), \quad s,t \in \mathbb{R},$$

which proves the result in view of (4.35). □

Example 4.5.2. If the covariance structure $\Gamma(t,s)$ of OFBM can be diagonalized as

$$\mathrm{diag}(\gamma_1(t,s), \gamma_2(t,s)),$$

where $\gamma_1(t,s) \neq \gamma_2(t,s)$ for some $s,t \in \mathbb{R}$, then G_1 is of trivial type. Indeed, for $A \in GL(2)$, the equation $A\mathrm{diag}(\gamma_1(t,s),\gamma_2(t,s))A^* = \mathrm{diag}(\gamma_1(t,s),\gamma_2(t,s))$ gives the solutions $G_1 = \{I, -I, \mathrm{diag}(1,-1), \mathrm{diag}(-1,1)\}$. We obtain this kind of $\Gamma(t,s)$, for example, by taking the time-domain representation of OFBM with $M = I$, $N = 0$, and $H = \mathrm{diag}(h_1, h_2) \in GL(2)$, $h_1 \neq h_2$, and $0 < h_l < 1$, $l = 1, 2$.

Theorems 4.5.5 and 4.5.6 show that if G_1 is trivial, rotational or maximal, then there exists positive definite W such that $W^{-1}HW$ is normal, where H is any operator in $\mathcal{E}(X)$. This can be used in the construction of a simple example of OFBM with minimal symmetry group.

Example 4.5.3. If

$$H = \begin{pmatrix} h & 0 \\ 1 & h \end{pmatrix},$$

then OFBM has the minimal symmetry group. Indeed, if there exists positive definite W such that $WHW^{-1} = A$, where A is normal, then $H = W^{-1}AW$ is diagonalizable over $\mathbb{C}$ (contradiction).

Observe that OFBMs in Examples 4.5.3, 4.5.2 and 4.5.1 are of minimal, trivial and maximal types, respectively. The next example provides OFBM of rotational type. Thus, classes of all four types of o.s.s. processes in Theorem 4.5.3 are non-empty.

Example 4.5.4. Consider OFBM given by the integral representation (4.23) with $H = \mathrm{diag}(h, h)$, $M \in SO(2)$ and $N = I$. Let $f_{h,+}(t, u) = (t-u)_+^{h-1/2} - (-u)_+^{h-1/2}$, $f_{h,-}(t, u) = (t-u)_-^{h-1/2} - (-u)_-^{h-1/2}$ and

$$g_1(t, s) = \int_{\mathbb{R}} f_{h,+}(t, u) f_{h,+}(s, u) du = \int_{\mathbb{R}} f_{h,-}(t, u) f_{h,-}(s, u) du,$$

$$g_2(t, s) = \int_{\mathbb{R}} f_{h,+}(t, u) f_{h,-}(s, u) du.$$

Note that, for suitable constants C_h, $\widetilde{C}_h$, and $s, t > 0$,

$$g_1(t, s) = C_h(t^{2h} + s^{2h} - |t-s|^{2h}), \quad g_2(t, s) = \widetilde{C}_h(-t^{2h} + |t-s|^{2h} 1_{\{t>s\}}).$$

The covariance structure of such OFBM can be expressed as

$$\Gamma(t, s) = 2g_1(t, s)I + g_2(t, s)M + g_2(s, t)M^*.$$

Then, $\widetilde{O} \in G_1$ if and only if $\widetilde{O}\Gamma(t,s)\widetilde{O}^* = \Gamma(t,s)$, or

$$2g_1(t,s)(\widetilde{O}\widetilde{O}^* - I) + g_2(t,s)(\widetilde{O}M\widetilde{O}^* - M) + g_2(s,t)(\widetilde{O}M^*\widetilde{O}^* - M^*)^* = 0. \tag{4.36}$$

For $t > s$, $\frac{\partial}{\partial t} g_2(s,t) = 0$. Then,

$$2\frac{\partial}{\partial t} g_1(t,s)(\widetilde{O}\widetilde{O}^* - I) + \frac{\partial}{\partial t} g_2(t,s)(\widetilde{O}M\widetilde{O}^* - M) = 0.$$

By integrating this back from 0 to t, we obtain that

$$2g_1(t,s)(\widetilde{O}\widetilde{O}^* - I) + g_2(t,s)(\widetilde{O}M\widetilde{O}^* - M) = 0. \tag{4.37}$$

In particular, comparing (4.36) and (4.37), $M^* = \widetilde{O}M^*\widetilde{O}^*$. This yields $M = \widetilde{O}M\widetilde{O}^*$, and hence, in view of (4.36), $\widetilde{O}\widetilde{O}^* = I$. The last relation implies that $\widetilde{O} \in O(2)$. Hence, $M\widetilde{O} = \widetilde{O}M$ and, since $M \in SO(2)$, this happens only with $\widetilde{O} \in SO(2)$. Thus, $G_1 = SO(2)$.

APPENDIX A

On the integration of continuous-time stationary Gaussian processes

Let $\{X(t)\}_{t\in\mathbb{R}}$ be a Gaussian stationary process given by (2.1). We define here the integral

$$\int_{\mathbb{R}} X(t)f(t)dt, \tag{A.1}$$

for suitable functions f and state its properties as used throughout the paper. Our strategy will be to define (A.1) both pathwise and as an $L^2(\Omega)$ limit and to show that the two definitions coincide in relevant cases. In the pathwise case, the integral (A.1) will be denoted by $\mathcal{I}_\omega(f)$ (i.e. defined ω-wise), and, in the $L^2(\Omega)$ case, it will be denoted by $\mathcal{I}_2(f)$.

For simplicity, we assume that the sample paths of X are continuous. Path continuity is not a stringent assumption since, by Belayev's alternative (Belayev (1960)), either the sample paths of a Gaussian stationary process are continuous or very badly-behaved in the sense of possessing discontinuities of the second type.

Assume first that $f(t) = \sum_{i=1}^n f_i 1_{[a_i,b_i)}(t)$ is a step function. For such function, the stochastic integral (A.1) may be defined pathwise as the ordinary Riemann integral

$$\mathcal{I}_\omega\Big(\sum_{i=1}^n f_i 1_{[a_i,b_i]}\Big) = \sum_{i=1}^n \int_{a_i}^{b_i} X(t)dt. \tag{A.2}$$

Lemma A.0.4. *The integral (A.2) has the following properties: for step functions f, f_1 and f_2, and with the notation $\mathcal{I}(f) = \mathcal{I}_\omega(f)$:*

(P1) $\mathcal{I}(f)$ is a Gaussian random variable with mean zero.

(P2) The following moment formulae hold:

$$E\mathcal{I}(f)^2 = \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{g}(x)|^2 |\widehat{f}(x)|^2 dx; \tag{A.3}$$

$$E\Big[\mathcal{I}(f_1)\mathcal{I}(f_2)\Big] = \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{g}(x)|^2 \widehat{f}_1(x)\overline{\widehat{f}_2(x)} dx; \tag{A.4}$$

$$E\Big[\mathcal{I}(f)X(t)\Big] \quad = \quad \frac{1}{2\pi}\int_{\mathbb{R}} e^{itx}|\widehat{g}(x)|^2\widehat{f}(x)dx. \tag{A.5}$$

(P3) For real c and d, $\mathcal{I}(cf_1 + df_2) = c\mathcal{I}(f_1) + d\mathcal{I}(f_2)$.

Proof. Property (P3) is elementary. It is enough to prove properties (P1) and (P2) in the case of indicator functions $f = 1_{[a,b)}$, $f_1 = 1_{[a_1,b_1)}$ and $f_2 = 1_{[a_2,b_2)}$. By using Lemma A.0.6 below, we have

$$\mathcal{I}_\omega(1_{[a,b]}) = \int_{\mathbb{R}}\left[\int_a^b g(t-u)dt\right]dB(u) \quad \text{a.s.} \tag{A.6}$$

Property (P1) is immediate since $\mathcal{I}_\omega(1_{[a,b)})$ is an integral with respect to Brownian motion.

We now turn to property (P2) and show first (A.4), of which (A.3) is a special case. By using (A.6) and the notation $f_1 = 1_{[a_1,b_1)}$, $f_2 = 1_{[a_2,b_2)}$, (A.4) follows from

$$E\mathcal{I}_\omega(1_{[a_1,b_1)})\mathcal{I}_\omega(1_{[a_2,b_2)}) = \int_{\mathbb{R}}\Bigg(\int_{a_1}^{b_1} g(t-u)dt\Bigg)\Bigg(\int_{a_2}^{b_2} g(t-u)dt\Bigg)du$$

$$= \int_{\mathbb{R}}(f_1 * g^{\vee})(u)(f_2 * g^{\vee})(u)du = \frac{1}{2\pi}\int_{\mathbb{R}}\widehat{f_1}(x)\overline{\widehat{f_2}(x)}|\widehat{g}(x)|^2dx.$$

To show (A.5), note that, by using (2.1), (A.6) and the notation $f = 1_{[a,b)}$,

$$EX(t)\mathcal{I}_\omega(1_{[a,b)}) = \int_{\mathbb{R}} g(t-u)\Bigg(\int_a^b g(s-u)ds\Bigg)du$$

$$= \int_{\mathbb{R}} g(t-u)(f * g^{\vee})(u)du = \frac{1}{2\pi}\int_{\mathbb{R}} e^{itx}\widehat{f}(x)|\widehat{g}(x)|^2dx.$$

□

An extension of the integral (A.1) to more general functions f can be achieved by an argument of approximation in $L^2(\Omega)$. Consider the space of deterministic functions

$$L^2_g := \Big\{f \in L^2(\mathbb{R}) : \int_{\mathbb{R}}|\widehat{f}(x)|^2|\widehat{g}(x)|^2dx < \infty\Big\} \tag{A.7}$$

with the inner product

$$\langle f_1, f_2\rangle_{L^2_g} := \frac{1}{2\pi}\int_{\mathbb{R}}\widehat{f_1}(x)\overline{\widehat{f_2}(x)}|\widehat{g}(x)|^2dx. \tag{A.8}$$

Denote also

$$\mathcal{I}_X^s = \Big\{ \mathcal{I}_\omega(f) : \ f \text{ is a step function} \Big\}, \tag{A.9}$$

equipped with the ordinary $L^2(\Omega)$ inner product

$$E\mathcal{I}_\omega(f_1)\mathcal{I}_\omega(f_2). \tag{A.10}$$

The space $\mathcal{I}_X^s$ and the restriction of L_g^2 to step functions are isometric since, for elementary functions f_1 and f_2,

$$E\mathcal{I}_\omega(f_1)\mathcal{I}_\omega(f_2) = \frac{1}{2\pi}\int_{\mathbb{R}} \widehat{f}_1(x)\overline{\widehat{f}_2(x)}|\widehat{g}(x)|^2 dx = \langle f_1, f_2 \rangle_{L_g^2}. \tag{A.11}$$

Thus, a natural way to define the integral $\mathcal{I}_2$ for a given $f \in L_g^2$ is to take a sequence of step functions l_n that approximate f in the L_g^2 norm, and set $\mathcal{I}_2(f)$ as the corresponding $L^2(\Omega)$ limit of $\mathcal{I}_\omega(l_n)$. We address the question of the existence of such a sequence of step functions in the following lemma.

Lemma A.0.5. *For every function $f \in L_g^2(\mathbb{R})$, there is a sequence $\{l_n\}$ of step functions such that $\|f - l_n\|_{L_g^2} \longrightarrow 0$.*

Proof. This result can be proved as Lemma 5.1 in Pipiras and Taqqu (2000). For the reader's convenience, we indicate here the main steps of the proof. Moreover, the proof of Lemma 5.1 in Pipiras and Taqqu (2000) contains a small error (see the argument before Case 2 on p. 274 in that paper) and needs to be modified slightly.

As in Pipiras and Taqqu (2000), it is enough to show the result in Case 1: f is an even function and, more specifically, such that $\widehat{f}(x) = 1_{[-1,1]}(x)$, and Case 2: f is an odd function and, more specifically, such that $\widehat{f}(x) = i(1_{[0,1]}(x) - 1_{[-1,0]}(x))$. We briefly consider Case 1 only.

In Case 1, write first

$$2\pi\|f - l_n\|_{L_g^2}^2 = \int_{\mathbb{R}} |x1_{[-1,1]}(x) - x\widehat{l}_n(x)|^2 \frac{|\widehat{g}(x)|^2}{x^2} dx.$$

Let $U(x)$ be the function on $x \in \mathbb{R}$ such that $U(x) = x1_{[-1,1]}(x)$ for $x \in [-k, k]$ and $U(x)$ is periodic with period $2k$, where $k \geq 2$. Suppose $\epsilon > 0$ is arbitrarilly small. Since $\widehat{g} \in L^2$ and $|U(x)| \leq 1$, we can fix k such that

$$\int_{|x|>k} |U(x)|^2 \frac{|\widehat{g}(x)|^2}{x^2} dx < \epsilon.$$

Then,

$$2\pi \|f - l_n\|_{L_g^2}^2 \leq \int_{\mathbb{R}} |U(x) - x\widehat{l_n}(x)|^2 \frac{|\widehat{g}(x)|^2}{x^2} dx + \epsilon. \tag{A.12}$$

The functions l_n are now constructed as follows. As shown in Pipiras and Taqqu (2000), there is a sequence of trigonometric functions $U_n(x) = \sum_{j=-n}^{n} u_j e^{i\pi jx/k}$ such that

(*i*) $\sup_{n,x} |U_n(x)| \leq \text{const}$,

(*ii*) $\sup_n |U_n(x)| \leq \text{const } |x|$, for small $|x|$,

(*iii*) $U_n(x) \to U(x)$ except at discontinuity points of $U(x)$.

Then, by the Dominated Convergence Theorem, we obtain

$$\int_{\mathbb{R}} |U(x) - U_n(x)|^2 \frac{|\widehat{g}(x)|^2}{x^2} dx \to 0. \tag{A.13}$$

In view of (A.12) and (A.13), it is enough to observe that $U_n(x) = x\widehat{l_n}(x)$ for some step functions l_n. □

Given $f \in L_g^2$, we may use Lemma A.0.5 to define (A.1) as

$$\mathcal{I}_2(f) = \lim(L^2(\Omega))\mathcal{I}_\omega(l_n), \tag{A.14}$$

where $\{l_n\}$ is a sequence of step functions such that $\|f - l_n\|_{L_g^2} \to 0$. This definition does not depend on the approximating sequence of f. The integral $\mathcal{I}_2(f)$ has the following properties.

Theorem A.0.7. *The map $\mathcal{I}_2 : f \longrightarrow \mathcal{I}_2(f)$ defined by (A.14) is an isometry between the spaces L_g^2 and $\mathcal{I}_X = \{\mathcal{I}_2(f) : f \in L_g^2\}$. Moreover, $\mathcal{I}_2(f) = \mathcal{I}_\omega(f)$ a.s. for step functions f, and the integral $\mathcal{I}_2(f)$ satisfies the properties (P1), (P2) and (P3) of Lemma A.0.7 with $\mathcal{I}(f) = \mathcal{I}_2(f)$ and $f, f_1, f_2 \in L_g^2$.*

Proof. The proof is omitted as being standard once we have Lemma A.0.5. □

Remark. Relation (A.5) in property (P2) can be seen as a particular case of (A.4) with $f_2(u) := \delta_t(u)$, where the latter stands for the Dirac delta at $u = t$. For such f_2, $\widehat{f_2}(x) = \widehat{\delta_t}(x) = e^{-itx}$ and note that $\int_{\mathbb{R}} |\widehat{f_2}(x)|^2 |\widehat{g}(x)|^2 dx = \int_{\mathbb{R}} |\widehat{g}(x)|^2 dx < \infty$.

It is possible to define (A.1) also pathwise for more general integrand functions. As discussed in Section 2.7, for a Gaussian stationary process $\{X(t)\}_{t\in\mathbb{R}}$, we have, almost surely,

$$|X(t)| \leq C\sqrt{\log(2 + |t|)}, \quad t \in \mathbb{R}, \tag{A.15}$$

where C is a random variable. Consider the space

$$\mathcal{L} := \{f \in L^2(\mathbb{R}) : \int_{\mathbb{R}} \sqrt{\log(2 + |t|)}|f(t)|dt < \infty\}. \tag{A.16}$$

For $f \in \mathcal{L}$, in view of (A.15) we may define

$$\mathcal{I}_\omega(f) = \int_{\mathbb{R}} X(t)f(t)dt$$

pathwise as an improper Riemann integral. It is reasonable to expect the integrals $\mathcal{I}_\omega(f)$ and $\mathcal{I}_2(f)$ to coincide a.s. at least for suitable integrands f.

Proposition A.0.2. *For $f \in \mathcal{L} \cap L^2_g$, $\mathcal{I}_2(f) = \mathcal{I}_\omega(f)$ a.s.*

Proof. Note that $\mathcal{I}_2(f) = \mathcal{I}_\omega(f)$ for a step function f. Take a sequence of step functions $\{l_n\}$ such that $\|f - l_n\|_{L^2_g} \to 0$. We know that $\|\mathcal{I}_2(f) - \mathcal{I}_2(l_n)\|_{L^2(\Omega)} \to 0$. It is therefore enough to show that $\|\mathcal{I}_\omega(f) - \mathcal{I}_\omega(l_n)\|_{L^2(\Omega)} \to 0$ as well. This follows by using Lemma A.0.6 below since

$$\|\mathcal{I}_\omega(f) - \mathcal{I}_\omega(l_n)\|_{L^2(\Omega)} = E\left[\int_{\mathbb{R}} X(t)(l_n(t) - f(t))dt\right]^2$$

$$= E\Big(\int_{\mathbb{R}}\Big(\int_{\mathbb{R}} g(t-u)dB(u)\Big)(l_n(t) - f(t))dt\Big)^2$$

$$= E\Big(\int_{\mathbb{R}}\Big(\int_{\mathbb{R}} g(t-u)(l_n(t) - f(t))dt\Big)dB(u)\Big)^2$$

$$= \int_{\mathbb{R}}\Big(\int_{\mathbb{R}} g(t-u)(l_n(t) - f(t))dt\Big)^2 du = \int_{\mathbb{R}}\Big(g^\vee * (l_n - f)(u)\Big)^2 du$$

$$= \frac{1}{2\pi}\int_{\mathbb{R}} |\widehat{g}(x)|^2 |(\widehat{l_n - f})(x)|^2 dx = \|l_n - f\|_{L_g^2} \to 0.$$

□

The next lemma was used several times in the appendix above.

Lemma A.0.6. *Let $\{X(t)\}_{t\in\mathbb{R}}$ be as in (2.1) with continuous sample paths and $f \in \mathcal{L}$ be an a.e. continuous bounded function, where $\mathcal{L}$ is defined in (A.16). Then,*

$$\mathcal{I}_\omega(f) = \int_{\mathbb{R}} X(t)f(t)dt = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} g(t-u)dB(u)\Big)f(t)dt$$

$$= \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} g(t-u)f(t)dt\Big)dB(u) \quad \text{a.s.} \tag{A.17}$$

Proof. Suppose first that f is bounded. From the definition of improper integral, $\mathcal{I}_\omega(f) = \lim_{a\to+\infty}\mathcal{I}_\omega(f1_{[-a,a]})$ a.s. We will first show that (A.17) holds for $f1_{[-a,a]}$, where $a > 0$ is fixed. Let $\Pi = \{-a = t_0 \le t_1 \le ... \le t_n = a\}$ denote a partition of the interval $[-a, a]$. By the a.e. sample path continuity of $X(t)f(t)$, the discretization $\sum_{t_k\in\Pi} X(t_k)f(t_k)(t_{k+1}-t_k)$ converges to $\mathcal{I}_\omega(f1_{[-a,a]})$ a.s. as $\|\Pi\| \to 0$. The discretization is an $L^2(\Omega)$ random variable, and it suffices to prove that it also converges in $L^2(\Omega)$ to the integral on the R.H.S. of (A.17). Write

$$\sum_{t_k\in\Pi} X(t_k)f(t_k)(t_{k+1}-t_k) = \int_{\mathbb{R}} G_\Pi(u)dB(u),$$

where $G_\Pi(u) = \sum_{t_k\in\Pi} g(t_k-u)f(t_k)(t_{k+1}-t_k)$. Observe that

$$\widehat{G_\Pi}(x) = \sum_{t_k\in\Pi} e^{-it_kx}\widehat{g}(-x)f(t_k)(t_{k+1}-t_k). \tag{A.18}$$

As $\|\Pi\| \to 0$, $\widehat{G_\Pi}(x)$ converges pointwise to $\widehat{g}(-x)(\widehat{f1_{[-a,a]}})(x)$, which is the Fourier transform of $(g^\vee * (f1_{[-a,a]}))(u) = \int_{\mathbb{R}} g(t-u)f(t)1_{[-a,a]}dt =: G(u)$. Furthermore,

$$\left|\widehat{G_\Pi}(x)\right| \le \left|\widehat{g}(-x)\sum_{t_k\in\Pi} e^{-it_kx}f(t_k)(t_{k+1}-t_k)\right| \le C\,|\widehat{g}(x)|\,, \tag{A.19}$$

since f is bounded. Since $\widehat{g} \in L^2(\mathbb{R})$ by assumption, the Dominated Convergence Theorem implies that $\widehat{G}_\Pi(x)$ converges to $\widehat{G}(x)$ in $L^2(\mathbb{R})$. This yields that $\int_{\mathbb{R}} G_\Pi(u)dB(u)$ converges to $\int_{\mathbb{R}} G(u)dB(u)$ in $L^2(\Omega)$, and proves (A.17) for $f1_{[-a,a]}$. To show (A.17) in general, it is enough to prove that

$$\int_{\mathbb{R}} \Big(\int_{-a}^{a} g(t-u)f(t)dt \Big) dB(u) \overset{L^2(\Omega)}{\longrightarrow} \int_{\mathbb{R}} \Big(\int_{\mathbb{R}} g(t-u)f(t)dt \Big) dB(u). \tag{A.20}$$

Taking Fourier transforms, this is equivalent to

$$\int_{\mathbb{R}} |\widehat{g}(x)|^2 |(\widehat{f1}_{[-a,a]} - \widehat{f})(x)|^2 dx \to 0.$$

The convergence follows from the Dominated Convergence Theorem since $\widehat{f1_{[-a,a]}}(x) \to \widehat{f}(x)$ for all $x \in \mathbb{R}$ (use $\mathcal{L} \subseteq L^1(\mathbb{R})$) and $|\widehat{f1}_{[-a,a]}(x)| \leq \|f\|_{L^1}$.

For the case of unbounded f, just consider a sequence of truncated integrands $f_n := f1_{\{|f|\leq n\}}$ and apply again the Dominated Convergence Theorem.

$\square$

APPENDIX B

Supplementary material on linear operators

B.1 On the commutativity of operators

The characterization of the commutativity of operators is a well-known problem in Linear Algebra (see, for instance, MacDuffee (1946), p. 89, or Taussky (1953)). More precisely, given an operator A, the problem is to find the set $\mathcal{C}(A)$ of operators that commute with A. $\mathcal{C}(A)$ is called the *centralizer of A*.

From now on, E represents an n-dimensional vector space with field $\mathbb{F}$, and $\mathcal{L}(E, \mathbb{F})$ is the space of endomorphisms on E ($\mathbb{F}$ is included in the notation to stress what particular field is taken). In particular, $\mathcal{L}(E, \mathbb{R})$ is isomorphic to $M(n)$. A vector $v \in E\backslash\{0\}$ is said to be an eigenvector for $A \in \mathcal{L}(E, \mathbb{F})$ if there exists $\lambda \in \mathbb{F}$ such that $Av = \lambda v$. $\lambda \in \mathbb{F}$ is said to be an eigenvalue when there exists a vector $v \in E\backslash\{0\}$ such that $Av = \lambda v$. For a given eigenvalue $\lambda \in \mathbb{F}$, the subspace $E_\lambda := \{v \in E; Av = \lambda v\}$ is said to be the eigenspace of A corresponding to λ.

Note that sufficient conditions for commutativity are usually easy to obtain.

Example B.1.1. Assume $A, X \in M(n)$ are two diagonalizable operators with (individually) distinct eigenvalues. A and X commute if they have the same eigenspaces, since, in this case,

$$A = PD_AP^{-1}, \quad X = PD_XP^{-1} \tag{B.1}$$

for diagonal D_A and D_X, and diagonal matrices commute. Note that P in (B.1) is not unique.

Eigenvalues with multiplicity greater than 1 introduce the multi-dimensionality of eigenspaces. For instance, the Identity commutes with every (e.g., diagonalizable) operator A because it can be diagonalized through any basis of $\mathbb{R}^n$, and in particular, the eigenvector basis of A.

Still in the context of diagonalizable operators, the sharing of eigenvectors - equivalently, of 1-dimensional invariant subspaces - is also a necessary condition, as Theorem B.1.3 below shows. The intuition behind it is clear: the order of the application of operators does not matter if and only if they act as scalars - which are algebraic entities that commute - upon the same 1-dimensional (invariant) subspaces of $\mathbb{R}^n$.

In the general case of any two operators $A, X \in M(n)$, the complexity of the matter increases, because the dimensions of the eigenspaces may not add up to n (see Subsection B.1.2). The next proposition give a general necessary condition for commutativity.

Proposition B.1.1. *Let $A, X \in \mathcal{L}(E, \mathbb{F})$. If X commutes with A, then each eigenspace of A is invariant by X.*

Proof. Let E_λ be an eigenspace of A associated with the eigenvalue $\lambda \in \mathbb{F}$. Then, $Av = \lambda v$ implies that

$$A(Xv) = X(Av) = X(\lambda v) = \lambda(Xv),$$

i.e., $Xv \in E_\lambda$.

□

A case of particular interest is when the eigenspace of A is 1-dimensional. Then, one can immediately obtain an eigenvector for X, which, depending on the context, may be used in the construction of an eigenvector basis of X (as in Section 4.5).

Corollary B.1.1. *Under the assumptions of Proposition B.1.1, if the eigenspace E_λ of A is 1-dimensional, then there exists η such that $Xw = \eta w$ for all $w \in E_\lambda$.*

Proof. Since E_λ is unidimensional, then $X(E_\lambda)$ has dimension either zero or one. In the former case, E_λ is an eigenspace of X with eigenvalue $\eta = 0$. In the latter case, choose some arbitrary v in E_λ. The vector Xv can be written as ηv for some $\eta \neq 0$. Likewise, any other vector $w \in E_\lambda$ can be written as $\alpha(w)v$ for some $\alpha(w) \in \mathbb{F}$. Therefore,

$$Xw = X\alpha(w)v = \alpha(w)Xv = \alpha(w)\eta v = \eta\alpha(w)v = \eta w.$$

□

Note, however, that the subspace E_λ in Corollary B.1.1 may not be the eigenspace of the operator X associated with the eigenvalue λ. For instance, if $X =$ Identity, the entire E is the eigenspace associated with the eigenvalue 1.

In Subsection B.1.1, necessary and sufficient conditions for commutativity are obtained in the classical setting of self-adjoint operators. This case is of interest not only because it is familiar to most readers but also because the discussion of commutativity can be carried out directly in terms of eigenvectors and eigenspaces, a fact related to the Spectral Theorem. In Subsection B.1.2, necessary and sufficient conditions for the commutativity of any two operators are obtained. The matrix perspective is predominant because it facilitates understanding of the issues involved. Subsection B.1.1 is based on Lima (1996), chapters 12 and 13, whereas Subsection B.1.2 is based on Gantmacher (1959), chapter 8, and Lima (1996), appendix.

B.1.1 The case of self-adjoint operators

Since our general discussion of commutativity involves the use of invariant subspaces, we opted for not directly using the Spectral Theorem in this subsection. This will make more clear at what point self-adjointness is indeed necessary (see also Remark B.1.2).

We begin by showing (Proposition B.1.2) that *every* $A \in \mathcal{L}(E, \mathbb{R})$ has an invariant subspace of dimension 1 or 2. Equivalently, either there exists a non-null vector $u \in E$ such that $Au = \lambda u$ or there exist linearly independent $u, v \in E$ such that Au and Av are both linear combinations of u and v, i.e., $Au = \alpha u + \beta v$, $Av = \gamma u + \delta v$.

To show this, we first prove Lemma B.1.1, which states there exists an irreducible monic polynomial p of degree 1 or 2 such that the $\mathrm{Ker}(p(A))$ is non-empty (a monic polynomial is a polynomial whose coefficient of the highest order term is 1). The proof of Lemma B.1.1 makes use of the Fundamental Theorem of Algebra, which implies that every monic real polynomial is decomposable as the product of irreducible monic polynomials of the first and second degrees. Here, one should remember that an irreducible second degree polynomial does not have real roots.

Denote $p(x) = a_0 + a_1 x + ... + a_n x^n$, and $p(A) = a_0 I + a_1 A + ... + a_n A^n$.

Lemma B.1.1. *Let $A \in \mathcal{L}(E, \mathbb{R})$. There exists an irreducible monic polynomial p of degree*

1 or 2 and a non-null vector v such that $p(A)v = 0$.

Proof. The space $\mathcal{L}(E, \mathbb{R})$ has dimension n^2, and therefore the operators $I, A, ..., A^{n^2}$ are linearly dependent. This means there exist $\alpha_0, \alpha_1, ..., \alpha_{n^2}$, of which at least one is not zero, such that

$$\alpha_0 I + \alpha_1 A + ... + \alpha_{n^2} A^{n^2} = 0.$$

Let α_m be the highest-indexed non-zero coefficient. If we set $\beta_i = \alpha_i / \alpha_m$, we obtain a monic polynomial

$$q(x) := \beta_0 + \beta_1 x + ... + \beta_{m-1} x^{m-1} + x^m$$

such that $q(A) = 0$. By the Fundamental Theorem of Algebra, we can factor $q(x) = q_1(x)...q_k(x)$, where each $q_i(x)$ is a monic irreducible polynomial of degree 1 or 2. Therefore,

$$q(A) = q_1(A)...\, q_k(A) = 0,$$

which implies there exists $i \in \{1, ..., k\}$ such that $q_i(A)$ is not invertible. Therefore, there exists a non-null v such that $q_i(A)v = 0$. To finish the proof, just set $p = q_i$. □

Proposition B.1.2. *Any $A \in \mathcal{L}(E, \mathbb{R})$ has an invariant subspace of dimension 1 or 2.*

Proof. Let p be the polynomial given by Lemma B.1.1. If $p(x) = x - \lambda$, then $p(A)v = (A - I\lambda)v = 0$, and thus we obtain a 1-dimensional invariant subspace.

Alternatively, if p is of degree 2, then we can write it as $p(x) = x^2 + ax + b$, $a, b \in \mathbb{R}$. This means that $p(A)v = A^2 v + aAv + bv = 0$, and thus, $A(Av) = -a(Av) - bv$. Thus, the subspace generated by v and Av is invariant by A. Furthermore, this subspace must be 2-dimensional. In fact, assume by contradiction that v and Av are linearly dependent. Then, there exists $\lambda \in \mathbb{R}$ such that $Av = \lambda v$, and thus

$$0 = A^2 v + aAv + bv = \lambda^2 v + a\lambda v + bv = (\lambda^2 + a\lambda + b)v,$$

which implies $\lambda^2 + a\lambda + b = 0$. This is impossible, since the irreducible second-degree polynomial p has no real root.

□

Although Proposition B.1.2 proves the existence of a 1- or 2-dimensional invariant subspace for an operator A, it is not clear whether A has an eigenvector *basis*. This is where self-adjointness comes into play. We now prove a simple fact about self-adjoint operators.

Lemma B.1.2. *Let E be a vector space with inner product, and let $A \in \mathcal{L}(E, \mathbb{R})$ be self-adjoint. If λ and λ' are two distinct eigenvalues of A, their respective eigenvectors v and v' are orthogonal.*

Proof. This follows by self-adjointness and the fact that $\lambda - \lambda' \neq 0$, since

$$(\lambda - \lambda')\langle v, v'\rangle = \langle \lambda v, v'\rangle - \langle v, \lambda' v'\rangle = \langle Av, v'\rangle - \langle Av, v'\rangle = 0.$$

□

The next proposition shows the existence of an orthonormal basis of eigenvectors for a self-adjoint A in the case where E is 2-dimensional. Note that the existence of an invariant subspace as stated in Proposition B.1.2 is, in fact, necessary for the argument to work.

Proposition B.1.3. *Let E be a 2-dimensional vector space with inner product, and let $A \in \mathcal{L}(E, \mathbb{R})$ be a self-adjoint operator. There exists an orthonormal basis $\{u_1, u_2\} \subseteq E$ of eigenvectors of A.*

Proof. Let $\{v, w\}$ be an arbitrary orthonormal basis of E. Due to the symmetry of the matrix representation of A, we have

$$Av = av + bw \quad \text{and} \quad Aw = bv + cw.$$

Thus, the eigenvalues of A are the roots of the polynomial $p(\lambda) = \lambda^2 - (a+c)\lambda + (ac - b^2)$. If the discriminant is zero, then $b = 0$, $a = c$ and thus $A = aI$, which implies that every non-null vector in E is an eigenvector of A. If the discriminant is greater than zero, then λ_1 and λ_2 are real and distinct roots. Thus, $A - \lambda_1 I$ and $A - \lambda_2 I$ are both non-invertible. Therefore, there exist eigenvectors u_1, u_2 of A, i.e., $Au_1 = \lambda_1 u_1$ and $Au_2 = \lambda_2 u_2$ (without loss of generality, we can assume u_1 and u_2 have norm 1). Since the eigenvectors corresponding to distinct eigenvalues of a self-adjoint operator are orthogonal (Lemma B.1.2), $\{u_1, u_2\} \subseteq E$

is an orthonormal basis of eigenvectors of A.

□

Proposition B.1.4. *Let E be a vector space with inner product. Every self-adjoint operator $A \in \mathcal{L}(E, \mathbb{R})$ has an eigenvector.*

Proof. By Proposition B.1.2, there exists a 1- or 2-dimensional subspace $V \subseteq E$ which is invariant by A. If $\dim(V) = 1$, then every non-null vector $v \in V$ is an eigenvector of A. If $\dim(V) = 2$, then by applying Proposition B.1.3 to the restriction $A : V \to V$ of A to the invariant subspace V, we obtain an eigenvector of A. □

Remark B.1.1. What Proposition B.1.4 ensures is the existence of an eigenvector when we are restricted to the field $\mathbb{R}$ (for instance, in this context a rotation in $SO(2)\backslash\{I, -I\}$ does not have an eigenvector, although Proposition B.1.2 still holds). Over the field $\mathbb{C}$, the existence of an eigenvector is an immediate consequence of applying the Fundamental Theorem of Algebra to the polynomial $\det(A - \lambda I)$, and does not depend on specific assumptions on A such as self-adjointness.

Proposition B.1.5. *Let E be a vector space with inner product. If the subspace $V \subseteq E$ is invariant by the linear operator $A \in \mathcal{L}(E, \mathbb{F})$, then $V^{\perp}$ is invariant by the adjoint A^*.*

Proof. Let $u \in V$, $v \in V^{\perp}$. Note that $\langle A^* v, u\rangle = \langle v, Au\rangle = 0$, since V is invariant by A. Thus, $A^* v \in V^{\perp}$. □

Proposition B.1.5 yields the following result.

Proposition B.1.6. *Let E be a vector space with inner product, and let $A \in \mathcal{L}(E, \mathbb{R})$ be a self-adjoint operator. If the subspace V is invariant by A, then so is $V^{\perp}$.*

We can now prove the main result of this subsection.

Theorem B.1.1. *Let E be a vector space with inner product, and let $A, X \in \mathcal{L}(E, \mathbb{R})$ be self-adjoint, linear operators. A and X commute if and only if there exists a basis of common eigenvectors.*

Proof. Let $u_1, ..., u_n$ be a basis of common eigenvectors of A and X. Let $\lambda_1^A, ..., \lambda_n^A$ be their respective (possibly repeated) A eigenvalues, and let $\lambda_1^X, ..., \lambda_n^X$ be their respective (possibly repeated) X eigenvalues. Take a vector $v = \sum_{i=1}^n \alpha_i u_i \in E$, $\alpha_i \in \mathbb{R}$, and write

$$XAv = XA\Big(\sum_{i=1}^n \alpha_i u_i\Big) = \Big(\sum_{i=1}^n \alpha_i \lambda_i^X \lambda_i^A u_i\Big) = \Big(\sum_{i=1}^n \alpha_i \lambda_i^A \lambda_i^X u_i\Big)$$

$$= AX\Big(\sum_{i=1}^n \alpha_i u_i\Big) = AXv.$$

For the converse, as a consequence of Proposition B.1.4, there exists an eigenspace E_{λ_1} of A with associated eigenvalue λ_1. Now assume A and X commute. By Proposition B.1.1, E_{λ_1} is invariant by X. By Proposition B.1.4, X has an eigenvector $w \in E_{\lambda_1}$, which must also be an eigenvector of A. Thus, w is a common eigenvector of A and X. By Proposition B.1.6, the subspace $\text{span}(w)^\perp \subseteq E$ is invariant by both A and X, so the argument can repeated to obtain a new common eigenvector in this subspace. So, by repeatedly applying Proposition B.1.6, we obtain a basis of common eigenvectors.

□

From the matrix perspective, Theorem B.1.1 states that two self-adjoint linear operators A, X commute if and only if there is a basis from which we can construct a matrix $O \in O(n)$ that simultaneously diagonalizes A and X, i.e.,

$$A = OD_AO^* \quad \text{and} \quad X = OD_XO^*.$$

As in the more general case of diagonalizable operators, the commutativity of A and X is related to the fact that diagonal matrices commute.

Remark B.1.2. The Spectral Theorem for E with inner-product and field $\mathbb{R}$ states that $A \in \mathcal{L}(E, \mathbb{R})$ is self-adjoint if and only if there exists an orthonormal basis of eigenvectors of A. So, one could have proved Theorem B.1.1 by directly employing the Spectral Theorem in place of Proposition B.1.4.

Remark B.1.3. All the discussion in this subsection may be easily extended to the case of normal operators. Of course, this involves dealing with complex vector spaces. See, for instance, Gantmacher (1959), chapter 9.

B.1.2 The general case

Over complex vector spaces, eigenvalues and eigenvectors always exist (see Remark B.1.1; in particular, 1-dimensional invariant subspaces always exist, as an extension of Proposition B.1.2). However, the dimensions of the eigenspaces of $A \in M(n, \mathbb{C})$ do not generally add up to n since the geometric dimension of a characteristic root (i.e., the dimension of the associated eigenspace) may be less than its algebraic dimension (i.e., its multiplicity). This implies that in general operators are not diagonalizable and the Spectral Theorem (even for normal operators) does not hold. The closest one can get to diagonalization is the so-called Jordan canonical form (also known as Jordan normal form; see, for instance, Lima (1996), p. 340, or Lang (1987), p. 262). Every matrix $A \in M(n, \mathbb{C})$ is conjugate to a matrix J whose diagonal is made up of so-called Jordan blocks. Each Jordan block J_{λ_i} has the form

$$J_{\lambda_i} = \begin{pmatrix} \lambda_i & 0 & 0 & \dots & 0 \\ 1 & \lambda_i & 0 & \dots & 0 \\ 0 & 1 & \lambda_i & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & \lambda_i \end{pmatrix}, \tag{B.2}$$

where λ_i is a root of the characteristic polynomial of A, and there can be more than one block with the same value λ_i on the diagonal. Jordan blocks commute, since they are lower triangular Toeplitz operators. This already points to the general form of $\mathcal{C}(A)$, in the sense that this set must encompass more matrices than only those that can be reduced to Jordan canonical form through the same conjugacy $P \in GL(n, \mathbb{C})$ as A. For instance, for

$$A := \begin{pmatrix} \lambda & 0 & 0 \\ 1 & \lambda & 0 \\ 0 & 1 & \lambda \end{pmatrix},$$

$\mathcal{C}(A)$ must include all 3×3 lower-triangular Toeplitz matrices

$$\begin{pmatrix} a & 0 & 0 \\ b & a & 0 \\ c & b & a \end{pmatrix}, \quad a, b, c \in \mathbb{C}.$$

The problem of finding commuting matrices is a particular case of that of finding the non-trivial solutions $X \in M(m, n, \mathbb{C})$ to the equation

$$AX = XB, \quad A \in M(m, \mathbb{C}), \quad B \in M(n, \mathbb{C}). \tag{B.3}$$

We can write the elementary divisors of A and B as

$$(\lambda - \lambda_1)^{p_1}, (\lambda - \lambda_2)^{p_2}, ..., (\lambda - \lambda_u)^{p_u}, \quad p_1 + p_2 + ... + p_u = m,$$

$$(\lambda - \mu_1)^{q_1}, (\lambda - \mu_2)^{q_2}, ..., (\lambda - \mu_v)^{q_v}, \quad q_1 + q_2 + ... + q_v = n$$

(for the definition of elementary divisors, see Gantmacher (1959), p. 193). Let $I^{(k)}$ denote the k-dimensional Identity, and $H^{(k)}$ denote the (nilpotent) matrix with ones on the first subdiagonal and zeros elsewhere. The reduction to Jordan canonical form yields

$$A = U\widetilde{A}U^{-1}, \quad B = V\widetilde{B}V^{-1} \tag{B.4}$$

for conjugacies U,V, where

$$\widetilde{A} = \mathrm{diag}(\lambda_1 I^{(p_1)} + H^{(p_1)}, ..., \lambda_u I^{(p_u)} + H^{(p_u)}),$$

$$\widetilde{B} = \mathrm{diag}(\mu_1 I^{(q_1)} + H^{(q_1)}, ..., \mu_v I^{(q_v)} + H^{(q_v)}).$$

If we set

$$\widetilde{X} = U^{-1}XV,$$

then (B.3) can be written as

$$\widetilde{A}\widetilde{X} = \widetilde{X}\widetilde{B}, \tag{B.5}$$

which is simpler to deal with, since $\widetilde{A}$ and $\widetilde{B}$ are in Jordan canonical form. Now, $\widetilde{X}$ can be partitioned into blocks $X_{\alpha\beta}$ (without the "$\sim$" for notational simplicity), $\alpha = 1, ..., u, \beta = 1, ..., v$, corresponding to the quasi-diagonal form of $\widetilde{A}$ and $\widetilde{B}$. Accordingly, $X_{\alpha\beta}$ is of dimension $p_\alpha \times q_\beta$, since it right multiplies a Jordan block of dimension $p_\alpha \times p_\alpha$ on the left-hand side of (B.5), and left multiplies a Jordan block of dimension $q_\beta \times q_\beta$ on the right-hand side of (B.5).

By block multiplication, we obtain

$$(\lambda_\alpha I^{(p_\alpha)} + H^{(p_\alpha)})X_{\alpha\beta} = X_{\alpha\beta}(\mu_\beta I^{(q_\beta)} + H^{(q_\beta)}), \quad \alpha = 1, ..., u,\ \beta = 1, ..., v.$$

Equivalently,

$$(\mu_\beta - \lambda_\alpha)X_{\alpha\beta} = H_\alpha X_{\alpha\beta} - X_{\alpha\beta}G_\beta, \tag{B.6}$$

where $H_\alpha := H^{(p_\alpha)}$, $G_\beta := H^{(q_\beta)}$.

Thus, for given α, β, there are two cases to consider.

(i) $\lambda_\alpha \neq \mu_\beta$: By iterating equation (B.6) $r - 1$ times, we get

$$(\mu_\beta - \lambda_\alpha)^r X_{\alpha\beta} = \sum_{\sigma+\tau=r} (-1)^\tau \binom{r}{\tau} H_\alpha^\sigma X_{\alpha\beta} G_\beta^\tau.$$

By the nilpotence of H_α and G_β, if we take $r \geq p_\alpha + q_\beta - 1$, then each term of the summation has either $H_\alpha^\sigma = 0$ or $G_\beta^\tau = 0$. Since $\lambda_\alpha \neq \mu_\beta$, then $X_{\alpha\beta} = 0$.

(ii) $\lambda_\alpha = \mu_\beta$: In this case, we can rewrite (B.6) as

$$H_\alpha X_{\alpha\beta} = X_{\alpha\beta}G_\beta. \tag{B.7}$$

Set $X_{\alpha\beta} = [x_{ik}]$, $i = 1, ..., p_\alpha$, $k = 1, ..., q_\beta$. For the sake of illustration, assume without loss of generality that $p_\alpha > q_\beta$. Since

$$H_\alpha = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}_{p_\alpha \times p_\alpha}, \quad G_\beta = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}_{q_\beta \times q_\beta},$$

we have

$$H_\alpha X_{\alpha\beta} = \begin{pmatrix} 0 & 0 & \dots & 0 \\ x_{11} & x_{12} & \dots & x_{1,q_\beta} \\ x_{21} & x_{22} & \dots & x_{2,q_\beta} \\ \vdots & \vdots & \ddots & \vdots \\ x_{p_\alpha-1,1} & x_{p_\alpha-1,2} & \dots & x_{p_\alpha-1,q_\beta} \end{pmatrix}_{p_\alpha \times q_\beta},$$

$$X_{\alpha\beta} G_\beta = \begin{pmatrix} x_{12} & x_{13} & \dots & x_{1,q_\beta} & 0 \\ x_{22} & x_{23} & \dots & x_{2,q_\beta} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{p_\alpha,2} & x_{p_\alpha,3} & \dots & x_{p_\alpha,q_\beta} & 0 \end{pmatrix}_{p_\alpha \times q_\beta}$$

and hence

$$X_{\alpha\beta} = \begin{pmatrix} x_{11} & 0 & 0 & \dots & 0 \\ x_{21} & x_{11} & 0 & \dots & 0 \\ x_{31} & x_{21} & x_{11} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{p_\alpha,1} & x_{p_\alpha-1,1} & \dots & x_{21} & x_{11} \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}_{p_\alpha \times q_\beta}.$$

In particular, when $p_\alpha = q_\beta$,

$$T_{p_\alpha} := X_{\alpha\beta} = \begin{pmatrix} c_{\alpha\beta} & 0 & \dots & 0 \\ c'_{\alpha\beta} & c_{\alpha\beta} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ c_{\alpha\beta}^{(p_\alpha-1)} & \dots & c'_{\alpha\beta} & c_{\alpha\beta} \end{pmatrix}.$$

Therefore, for $p_\alpha \neq q_\beta$ we can write

$$X_{\alpha\beta} := \begin{pmatrix} T_{q_\beta} \\ 0 \end{pmatrix} \text{ when } p_\alpha > q_\beta, \text{ and } X_{\alpha\beta} := (T_{p_\alpha}, 0) \text{ when } p_\alpha < q_\beta.$$

We will say that in any of these cases $X_{\alpha\beta}$ is in *regular triangular form.*

As for the count of number of arbitrary parameters, let $d_{\alpha\beta}(\lambda)$ be the greatest common divisor of the elementary divisors $(\lambda - \lambda_\alpha)^{p_\alpha}$, $(\lambda - \mu_\beta)^{q_\beta}$. Also, let $\delta_{\alpha\beta}$ be the degree of $d_{\alpha\beta}(\lambda)$. In the case $\lambda_\alpha \neq \mu_\beta$, $\delta_{\alpha\beta} = 0$, and in the case $\lambda_\alpha = \mu_\beta$, $\delta_{\alpha\beta} = \min(p_\alpha, q_\beta)$. Therefore, $\delta_{\alpha\beta}$ gives the number of arbitrary parameters in $X_{\alpha\beta}$. Thus, the number N of arbitrary parameters in X is $\sum_{\alpha=1}^{u} \sum_{\beta=1}^{v} \delta_{\alpha\beta}$. We proved the following theorem.

Theorem B.1.2. *Let*

$$A := U\widetilde{A}U^{-1} = U\mathrm{diag}(\lambda_1 I^{(p_1)} + H^{(p_1)}, ..., \lambda_u I^{(p_u)} + H^{(p_u)})U^{-1},$$

$$B := V\widetilde{B}V^{-1} = V\mathrm{diag}(\mu_1 I^{(q_1)} + H^{(q_1)}, ..., \mu_v I^{(q_v)} + H^{(q_v)})V^{-1}.$$

The general solution of $AX = XB$ is given by

$$X = UX_{\widetilde{A}\widetilde{B}}V^{-1},$$

where $X_{\widetilde{A}\widetilde{B}}$ is the general solution to the equation

$$\widetilde{A}\widetilde{X} = \widetilde{X}\widetilde{B}.$$

$X_{\widetilde{A}\widetilde{B}}$ is decomposed into blocks $X_{\alpha\beta}$ of size $p_\alpha \times q_\beta$, where $\alpha = 1, ..., u$, $\beta = 1, ..., v$.

If $\lambda_\alpha \neq \mu_\beta$, then $X_{\alpha\beta} = 0$. If $\lambda_\alpha = \mu_\beta$, then $X_{\alpha\beta}$ is a lower triangular matrix.

$X_{\widetilde{A}\widetilde{B}}$, and therefore also X, depends linearly on $N = \sum_{\alpha=1}^{u} \sum_{\beta=1}^{v} \delta_{\alpha\beta}$ arbitrary parameters $c_1, ..., c_N$, where $\delta_{\alpha\beta}$ is the degree of the greatest common divisor of $(\lambda - \lambda_\alpha)^{p_\alpha}$ and $(\lambda - \mu_\beta)^{q_\beta}$. In particular,

$$X = \sum_{j=1}^{N} c_j X_j.$$

Each matrix X_j is a solution to $AX = XB$ by setting c_j to 1 and the remaining terms c to 0.

Note that if A and B do not have common characteristic roots, i.e., if the polynomials $\det(\lambda I - A)$ and $\det(\lambda I - B)$ are co-prime, then $N = 0$, and thus the only solution is $X = 0$.

We now apply the theorem to an example in the case $A = B$.

Example B.1.2. Assume A has the elementary divisors

$$(\lambda - \lambda_1)^4, (\lambda - \lambda_1)^3, (\lambda - \lambda_2)^2, (\lambda - \lambda_2), \quad \lambda_1 \neq \lambda_2. \tag{B.8}$$

Then, $\mathcal{C}(A)$ is made up of operators conjugate to

$$\begin{pmatrix}
a & 0 & 0 & 0 & | & 0 & 0 & 0 & | & 0 & 0 & | & 0 \\
b & a & 0 & 0 & | & h & 0 & 0 & | & 0 & 0 & | & 0 \\
c & b & a & 0 & | & k & h & 0 & | & 0 & 0 & | & 0 \\
d & c & b & a & | & l & k & h & | & 0 & 0 & | & 0 \\
- & - & - & - & | & - & - & - & | & - & - & | & - \\
e & 0 & 0 & 0 & | & m & 0 & 0 & | & 0 & 0 & | & 0 \\
f & e & 0 & 0 & | & p & m & 0 & | & 0 & 0 & | & 0 \\
g & f & e & 0 & | & q & p & m & | & 0 & 0 & | & 0 \\
- & - & - & - & | & - & - & - & | & - & - & | & - \\
0 & 0 & 0 & 0 & | & 0 & 0 & 0 & | & r & 0 & | & 0 \\
0 & 0 & 0 & 0 & | & 0 & 0 & 0 & | & s & r & | & w \\
- & - & - & - & | & - & - & - & | & - & - & | & - \\
0 & 0 & 0 & 0 & | & 0 & 0 & 0 & | & t & 0 & | & z
\end{pmatrix}$$

by some conjugacy $P \in GL(n, \mathbb{C})$, where the blocks on the diagonal above correspond to the Jordan blocks of $A = PJP^{-1}$ in the block diagonal matrix J.

The general form of $\mathcal{C}(A)$ is given in the theorem below.

Theorem B.1.3. *Let A be an operator in $M(n, \mathbb{C})$ whose representation in Jordan form* *contains m Jordan blocks, with conjugacy $P \in GL(n, \mathbb{C})$. Then, $\mathcal{C}(A)$ is made up of operators P-conjugate to*

$$\widetilde{A} = [\widetilde{A}_{i,j}]_{i,j=1,\ldots,m},$$

where $\widetilde{A}_{i,j}$ is either the null matrix or an arbitrary regular lower triangular matrix depending on whether $\lambda_i \neq \lambda_j$ or $\lambda_i = \lambda_j$.

Remark B.1.4. The cases described in Example B.1.1 and Theorem B.1.1 follow from Theorem B.1.3 by imposing the pertinent restrictions both on A and on the set of solutions X.

The following result - an immediate consequence of Theorem B.1.3 - complements Example B.1.1 and Corollary B.1.1.

Corollary B.1.2. *Assume $A \in M(n, \mathbb{C})$ has pairwise different characteristic roots. Then, if we denote by $P \in GL(n, \mathbb{C})$ the matrix whose columns are non-null eigenvectors $p_1, ..., p_n$ of A, we have*

$$\mathcal{C}(A) = \{X \in M(n, \mathbb{C});\ X = P\mathrm{diag}(\lambda_1, ..., \lambda_n)P^{-1},\ \lambda_i \in \mathbb{C}\}.$$

B.2 The closed form of the exponential of a matrix in Jordan canonical form

We develop here the expression for

$$z^J = \exp(J \log z) = \sum_{k=0}^{\infty} \frac{J^k (\log z)^k}{k!},$$

where J is a matrix in Jordan canonical form. Let J_λ be a Jordan block of size n_λ, whose expression is given in (B.2). It can be shown that

$$J_\lambda^k = \begin{pmatrix} \lambda^k & 0 & 0 & 0 & \dots & 0 \\ \binom{k}{1}\lambda^{k-1} & \lambda^k & 0 & 0 & \dots & 0 \\ \binom{k}{2}\lambda^{k-2} & \binom{k}{1}\lambda^{k-1} & \lambda^k & 0 & \dots & 0 \\ \binom{k}{3}\lambda^{k-3} & \binom{k}{2}\lambda^{k-2} & \binom{k}{1}\lambda^{k-1} & \lambda^k & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ \binom{k}{n_\lambda-1}\lambda^{k-n_\lambda+1} & \binom{k}{n_\lambda-2}\lambda^{k-n_\lambda+2} & \dots & \dots & \binom{k}{1}\lambda^{k-1} & \lambda^k \end{pmatrix},$$

where, by convention, $\binom{k}{j} = 0$ when $k < j$ (see, for instance, Lütkepohl (1993), p. 460).

Now, note that

$$\sum_{k=0}^{\infty} \binom{k}{j} \frac{\lambda^{k-j}(\log z)^k}{k!} = \sum_{k=j}^{\infty} \binom{k}{j} \frac{\lambda^{k-j}(\log z)^k}{k!} = \frac{(\log z)^j}{j!} \sum_{k=j}^{\infty} \frac{\lambda^{k-j}(\log z)^{k-j}}{(k-j)!}$$

$$= \frac{(\log z)^j}{j!} z^\lambda.$$

Therefore,

$$z^{J_\lambda} = \begin{pmatrix} z^\lambda & 0 & 0 & 0 & \dots & 0 \\ (\log z)z^\lambda & z^\lambda & 0 & 0 & \dots & 0 \\ \frac{(\log z)^2}{2!}z^\lambda & (\log z)z^\lambda & z^\lambda & 0 & \dots & 0 \\ \frac{(\log z)^3}{3!}z^\lambda & \frac{(\log z)^2}{2!}z^\lambda & (\log z)z^\lambda & z^\lambda & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ \frac{(\log z)^{n_\lambda-1}}{(n_\lambda-1)!}z^\lambda & \frac{(\log z)^{n_\lambda-2}}{(n_\lambda-2)!}z^\lambda & \dots & \dots & (\log z)z^\lambda & z^\lambda \end{pmatrix}. \qquad \text{(B.9)}$$

.

BIBLIOGRAPHY

Abry P., Flandrin P., Taqqu M.S. and Veitch D. (2003). Self-similarity and long-range dependence through the wavelet lens. In: *Theory and applications of long-range dependence*, pp. 527–556. Birkhäuser Boston, Boston, MA.

Akansu A.N. and Haddad R.A. (2001). *Multiresolution Signal Decomposition*. Academic Press Inc., San Diego, CA, second edition.

Averkamp R. and Houdré C. (1998). Some distributional properties of the continuous wavelet transform. *IEEE Transactions on Information Theory* **44(3)**, 1111–1124.

Belayev Y.K. (1960). Continuity and hölder's conditions for sample functions of stationary gaussian processes. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. Vol. II: Contributions to probability theory*. University of California Press.

Benassi A. and Jaffard S. (1994). Wavelet decomposition of one- and several-dimensional Gaussian processes. In: *Recent advances in wavelet analysis*, volume 3 of *Wavelet Anal. Appl.*, pp. 119–154. Academic Press, Boston, MA.

Beran J. (1994). *Statistics for Long-Memory Processes*. Chapman & Hall, New York.

Brockwell P.J. and Davis R.A. (1991). *Time Series: Theory and Methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition.

Cambanis S. and Houdré C. (1995). On the continuous wavelet transform of second-order random processes. *IEEE Transactions on Information Theory* **41(3)**, 628–642.

Cambanis S. and Masry E. (1994). Wavelet approximation of deterministic and random signals: Convergence properties and rates. *IEEE Transactions on Information Theory* **40(4)**, 1013–1029.

Chung C.F. (2002). Sample means, sample autocovariances, and linear regression of stationary multivariate long memory processes. *Econometric Theory* **18**, 51–78.

Craigmile P.F. and Percival D.B. (2005). Asymptotic decorrelation of between-scale wavelet coefficients. *IEEE Transactions on Information Theory* **51**, 1039–1048.

Cramér H. and Leadbetter M.R. (1967). *Stationary and Related Stochastic Processes*. Dover Publications Inc., Mineola, NY.

Daubechies I. (1992). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Delgado R. (2007). A reflected fbm limit for fluid models with on/off sources under heavy traffic. *Stochastic Processes and Their Applications* **117**, 188–201.

Dietrich C.R. and Newsam G.N. (1997). Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix. *SIAM Journal on Scientific Computing* **18**, 1088–1107.

Dijkerman R.W. and Mazumdar R.R. (1994). On the correlation structure of the wavelet coefficients of fractional Brownian motion. *IEEE Transactions on Information Theory* **40**, 1609–1612.

Donoho D.L. (1995). Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis. Time-Frequency and Time-Scale Analysis, Wavelets, Numerical Algorithms, and Applications* **2**, 101–126.

Donoho D.L. and Johnstone I.M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455.

Donoho D.L. and Johnstone I.M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association* **90**, 1200–1224.

Doob J.L. (1990). *Stochastic processes*. Wiley Classics Library. John Wiley & Sons Inc., New York. Reprint of the 1953 original, A Wiley-Interscience Publication.

Doukhan P., Oppenheim G. and Taqqu M. (2003). *Theory and Applications of Long-Range Dependence*. Birkhäuser, Boston, MA.

Embrechts P. and Maejima M. (2002). *Selfsimilar Processes*. Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ.

Fan Y. and Gençay R. (2006). Unit root and cointegration tests with wavelets. Preprint.

Gantmacher F.R. (1959). *The Theory of Matrices. Vol. 1*. AMS Chelsea Publishing, Providence, RI. Translated from the Russian by K. A. Hirsch, Reprint of the 1959 translation.

Grenander U. and Szego G. (1958). *Toeplitz Forms and their Applications*. Chelsea, New York.

Hausner M. and Schwartz J. (1968). *Lie Groups; Lie Algebras*. Notes on mathematics and its applications. Gordon and Breach Science Publishers, New York.

Horn R. and Johnson C. (1991). *Topics in Matrix Analysis*. Cambridge University Press, New York, NY.

Hudson W. and Mason J. (1981). Operator-stable distribution on $\mathbb{R}^2$ with multiple exponents. *The Annals of Probability* **9**, 482–489.

Hudson W. and Mason J. (1982). Operator-self-similar processes in a finite-dimensional space. *Transactions of the American Mathematical Society* **273**, 281–297.

Jaffard S., Lashermes B. and Abry P. (2005). Wavelet leaders in multifractal analysis. Preprint.

Jensen M.J. (1999). An approximate wavelet mle of short and long memory parameters. *Studies in Nonlinear Dynamics and Econometrics* **3**, 239–253.

Jurek Z. and Mason J. (1993). *Operator-Limit Distributions in Probability Theory*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, NY.

Konstantopoulos T. and Lin S.J. (1996). Fractional Brownian approximations of queueing networks. In: *Stochastic Processes Networks. Lecture Notes in Statistics*, 117, pp. 257–273. Springer, New York.

Krim H. and Pesquet J.C. (1995). Multiresolution analysis of a class of nonstationary processes. *IEEE Transactions on Information Theory* **41(4)**, 1010–1020.

Kwapień S. and Woyczyński W. (1992). *Random Series and Stochastic Integrals: Simple and Multiple*. Birkhäuser, Boston, MA.

Laha R.G. and Rohatgi V.K. (1982). Operator self-similar stochastic processes in $\mathbf{R}^d$. *Stochastic Processes and Their Applications* **12**, 73–84.

Lang S. (1987). *Linear Algebra*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, NY.

Lifshits M. (1995). *Gaussian Random Functions*. Kluwer Academic Publishers, Boston, MA.

Lima E.L. (1996). *Álgebra Linear*. Coleção Matemática Universitária. Instituto de Matemática Pura e Aplicada, Rio de Janeiro, Brazil.

Lütkepohl H. (1993). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin; New York.

MacDuffee C.C. (1946). *The Theory of Matrices*. Chelsea Publishing Company, New York.

Maejima M. (1994). Operator self-similar stable processes. *Stochastic Processes and Their Applications* **54**, 139–163.

Maejima M. (1998). Norming operators for operator self-similar processes. In: *Stochastic Processes and Related Topics*, Trends Math., pp. 287–295. Birkhäuser, Boston, MA.

Maejima M. and Mason J. (1994). Operator-self-similar stable processes. *Stochastic Processes and their Applications* **54**, 139–163.

Majewski K. (2003). Large deviations for multi-dimensional reflected fractional Brownian motion. *Stochastics* **75**, 233–257.

Majewski K. (2005). Fractional Brownian heavy traffic approximations of multiclass feed-forward queueing networks. *Queueing Systems* **50**, 199–230.

Mallat S. (1998). *A Wavelet Tour of Signal Processing*. Academic Press Inc., San Diego, CA.

Mallat S., Papanicolaou G. and Zhang Z. (1998). Adaptive covariance estimation of locally stationary processes. *The Annals of Statistics* **26**, 1–47.

Marinucci D. and Robinson P. (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference* **80**, 111–122.

Marinucci D. and Robinson P. (2000). Weak convergence of multivariate fractional processes. *Stochastic Processes and Their Applications* **86**, 103–120.

Mason J. and Yimin X. (2002). Sample path properties of operator-self-similiar Gaussian random fields. *Theory of Probability and Its Applications* **46**, 58–78.

Meerschaert M.M. and Scheffler H.P. (2001). *Limit Distributions for Sums of Independent Random Vectors: Heavy Tails in Theory and Practice*. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., New York.

Meyer Y. (1992). *Wavelets and Operators*, volume 37 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Meyer Y., Sellan F. and Taqqu M.S. (1999). Wavelets, generalized white noise and fractional integration: the synthesis of fractional Brownian motion. *The Journal of Fourier Analysis and Applications* **5**, 465–494.

Moulines E., Roueff F. and Taqqu M.S. (2006). A wavelet Whittle estimator of the memory parameter of a nonstationary Gaussian time series. Preprint.

Nason G.P., von Sachs R. and Kroisandt G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **62**, 271–292.

Ossiander M. and Waymire E.C. (2000). Statistical estimation for multiplicative cascades. *The Annals of Statistics* **28**, 1533–1560.

Percival D.B. and Walden A.T. (2000). *Wavelet Methods for Time Series Analysis*, volume 4 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge.

Pipiras V. (2004). Wavelet-type expansion of the Rosenblatt process. *The Journal of Fourier Analysis and Applications* **10**, 599–634.

Pipiras V. (2005). Wavelet-based simulation of fractional Brownian motion revisited. *Applied and Computational Harmonic Analysis* **19**, 49–60.

Pipiras V. and Taqqu M.S. (2000). Integration questions related to fractional Brownian motion. *Probability Theory and Related Fields* **118**, 251–291.

Resnick S., Samorodnitsky G., Gilbert A. and Willinger W. (2003). Wavelet analysis of conservative cascades. *Bernoulli* **9**, 97–135.

Rozanov Y.A. (1967). *Stationary Random Processes*. Holden-Day Inc., San Francisco, Calif.

Ruiz-Medina M.D., Angulo J.M. and Anh V.V. (2003). Fractional-order regularization and wavelet approximation to the inverse estimation problem for random fields. *Journal of Multivariate Analysis* **85**, 192–216.

Samorodnitsky G. and Taqqu M.S. (1994). *Stable Non-Gaussian Random Processes*. Stochastic Modeling. Chapman & Hall, New York.

Sellan F. (1995). Synthèse de mouvements browniens fractionnaires à l'aide de la transformation par ondelettes. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique* **321**, 351–358.

Taussky O. (1953). Commutativity in finite matrices. *The American Mathematical Monthly* **64**, 229–235.

Vaidyanathan P.P. and Akkarakaran S. (2001). A review of the theory and applications of optimal subband and transform coders. *Applied and Computational Harmonic Analysis* **10**, 254–289.

Veitch D. and Abry P. (1999). A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Transactions on Information Theory* **45**, 878–897.

Vidakovic B. (1999). *Statistical Modeling by Wavelets*. John Wiley & Sons Inc., New York.

Walter G.G. and Shen X. (2001). *Wavelets and Other Orthogonal Systems*. Studies in Advanced Mathematics. Chapman & Hall/CRC, Boca Raton, FL, second edition.

Zhang J. and Walter G. (1994). A wavelet-based KL-like expansion for wide-sense stationary random processes. *IEEE Transactions on Signal Processing* **42**, 1737–1745.