Adaptive wavelet decompositions of stationary time series [*†‡]

Gustavo Didier and Vladas Pipiras
University of North Carolina at Chapel Hill

June 19, 2006

Abstract

We introduce and examine particular wavelet-based decompositions of stationary time series in discrete time. The decompositions are essentially characterized by uncorrelated detail coefficients, possibly correlated approximation coefficients and a practically small length of associated low and high pass filters. The covariance structure of a time series, assumed to be known, enters into associated wavelet bases. Applications to simulation and maximum likelihood estimation are also presented. The focus is somewhat on long memory time series.

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, Johnstone and others (Donoho and Johnstone (1994, 1995)) 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, Flandrin, Taqqu and Veitch (2003), Pipiras (2005), Moulines, Roueff and Taqqu (2006)). Other applications but in continuous time, concern locally stationary time series (Mallat, Papanicolaou and Zhang (1998), Nason, von Sachs and Kroisandt (2000)), multifractal processes (Ossiander and Waymire (2000), Resnick, Samorodnitsky, Gilbert and Willinger (2003), Jaffard, Lashermes and Abry (2005)).

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

[*]The second author was supported in part by the NSF grant DMS-0505628.

[†]*AMS Subject classification.* Primary 60G10, 42C40; secondary 62M10, 62F10.

[‡]*Keywords and phrases*: stationary time series, circular time series, wavelets, conjugate mirror filters, zero moments, adaptive wavelet decompositions, long memory, simulation, maximum likelihood estimation.

exploring it in depth. The most studied is probably the case of long memory time series. See, for example, Dijkerman and Mazumdar (1994). But even this case, as seen from the reference, 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)).

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. 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.

MLE, in particular, has greatly motivated this work. 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, Vaidyanthan 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, Angulo and Anh (2003), Meyer, Sellan and Taqqu (1999).

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.

Correlated approximation coefficients allow, in particular, to have associated low and high pass filters of practically small length. The number of zero moments of the underlying 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 as the above situations have gained increased attention fairly recently). But 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 Didier and Pipiras (2006). Despite some similarities, however, the focus and contents of this work are very different from those in Didier and Pipiras (2006).

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 paper is organized as follows. In Section 2, we gather some basic notions and facts on time series and wavelets that will be used throughout the paper. In Section 3, we introduce and examine Adaptive Wavelet Decompositions (AWD) of stationary time series. Examples are considered in Section 4. Applications of AWD can be found in Section 5.

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 $h \in \mathbb{Z}$,

$$EX_{k+h}X_h = EX_kX_0 =: r(k), \quad k \in \mathbb{Z}, \tag{2.1}$$

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

$$X_n = \sum_{k=-\infty}^{\infty} a_k \epsilon_{n-k} = (a * \epsilon)_n, \quad n \in \mathbb{Z}, \tag{2.2}$$

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 (2.2) (see e.g. Brockwell and Davis (1991)):

$$X_n = \int_0^{2\pi} e^{inw} dW(w) = \int_0^{2\pi} e^{inw} \widehat{a}(w) dZ(w), \quad n \in \mathbb{Z}, \tag{2.3}$$

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 $EdZ(w)\overline{dZ(w')} = dw 1_{\{w=w'\}}/2\pi$, and

$$\widehat{a}(w) = \sum_{k=-\infty}^{\infty} a_k e^{-ikw}, \quad w \in (0, 2\pi), \tag{2.4}$$

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 $w \in \mathbb{R}$,

$$|\widehat{u}(w)|^2 + |\widehat{u}(w+\pi)|^2 = 2, \tag{2.5}$$

$$\widehat{v}(w) = e^{-iw}\overline{\widehat{u}(w+\pi)} \tag{2.6}$$

and hence

$$|\widehat{v}(w)|^2 + |\widehat{v}(w+\pi)|^2 = 2, \tag{2.7}$$

$$\widehat{u}(w)\overline{\widehat{v}(w)} + \widehat{u}(w+\pi)\overline{\widehat{v}(w+\pi)} = 0. \tag{2.8}$$

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,

$$\widehat{u}(w) = (1 + e^{-iw})^N \widehat{u}_{0,N}(w), \quad \widehat{v}(w) = (1 - e^{-iw})^N \widehat{v}_{0,N}(w), \tag{2.9}$$

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

$$a_j = \downarrow_2 (\overline{u} * a_{j-1}), \quad d_j = \downarrow_2 (\overline{v} * a_{j-1}), \quad j = 1, 2, \ldots, \tag{2.10}$$

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

$$a_j = u * \uparrow_2 a_{j+1} + v * \uparrow_2 d_{j+1}, \quad j = 0, 1, \ldots, \tag{2.11}$$

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

$$\widehat{(\downarrow_2 x)}(w) = \frac{1}{2}\Big(\widehat{x}\Big(\frac{w}{2}\Big) + \widehat{x}\Big(\frac{w}{2} + \pi\Big)\Big), \quad \widehat{(\uparrow_2 x)}(w) = \widehat{x}(2w). \tag{2.12}$$

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 $0, 1, \ldots, T-1$, with $T = 2^J$. In this case, the convolution $*$ above is replaced by the circular convolution $\circledast$, and the discrete Fourier transform of $x = \{x_0, x_1, \ldots, x_{T-1}\}$ becomes

$$\widehat{x}(w) = \sum_{k=0}^{T-1} x_k e^{-ikw}, \quad \text{at } w = \frac{2\pi j}{T},\ j = 0, \ldots, T-1. \tag{2.13}$$

In particular, with these modifications, (2.10) is considered for $j = 1, \ldots, J$, and (2.11) continues to hold for $j = 0, 1, \ldots, J-1$. 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 $Y = \{Y_0 \ldots, Y_{T-1}\}$ that are natural in the context of circular convolutions, are

$$Y_n = (b \circledast \epsilon)_n, \quad n = 0, \ldots, T-1, \tag{2.14}$$

where $\epsilon = \{\epsilon_0, \ldots, \epsilon_{T-1}\}$ are independent, $\mathcal{N}(0,1)$ random variables and $b = \{b_0, \ldots, b_{T-1}\}$ is a vector. These time series vectors are also stationary but not every stationary vector can be written this way. The covariance matrix $(E(Y_i Y_j),\ i, j = 0, \ldots, T-1)$ is, in fact, circular. Conversely, under mild assumptions, a Gaussian vector Y with a circular covariance matrix can be written as (2.14). If r_Y is the autocovariance function of Y, observe also that

$$r_Y = b \circledast \overline{b}, \quad \widehat{r}_Y(w) = |\widehat{b}(w)|^2. \tag{2.15}$$

3 Definition and basic properties of AWD

We shall use below the following general result of its own interest.

Proposition 3.1 *Let $a, b \in l^2(\mathbb{Z})$ be arbitrary filters and $u, v \in l^2(\mathbb{Z})$ be CMF. Define*

$$\widehat{U}_d(\omega) = \overline{\left(\frac{\widehat{b}(2\omega)}{\widehat{a}(\omega)}\right)}\widehat{u}(\omega), \quad \widehat{V}_d(\omega) = \overline{\left(\frac{1}{\widehat{a}(\omega)}\right)}\widehat{v}(\omega), \tag{3.1}$$

and

$$\widehat{U}_r(\omega) = \frac{\widehat{a}(\omega)}{\widehat{b}(2\omega)}\widehat{u}(\omega), \quad \widehat{V}_r(\omega) = \widehat{a}(\omega)\widehat{v}(\omega). \tag{3.2}$$

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

$$U_d, V_d, U_r, V_r \in l^1(\mathbb{Z}). \tag{3.3}$$

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

$$Y = \downarrow_2 (\overline{U}_d * X), \quad \eta = \downarrow_2 (\overline{V}_d * X), \tag{3.4}$$

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

$$X = U_r * \uparrow_2 Y + V_r * \uparrow_2 \eta. \tag{3.5}$$

Proof: The condition (3.3) ensures that the time series in (3.4) and (3.5) 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 (2.3) of the time series X. By Theorem 4.10.1 in Brockwell and Davis (1991), we obtain that

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

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

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

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 (2.5), (2.7) and (2.8) 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 (3.5) 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

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

Remark 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, the Fourier transforms of the R.H.S. of (3.4) are

$$\downarrow_2 \widehat{(\overline{U}_d * X)}(w) = \widehat{b}(w)\downarrow_2 \widehat{(u*\epsilon)}(w), \quad \downarrow_2 \widehat{(\overline{V}_d * X)}(w) = \downarrow_2 \widehat{(v*\epsilon)}(w). \tag{3.6}$$

Similarly, the Fourier transform of the R.H.S. of (3.5) 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 3.2 Another interpretation of Proposition 3.1 is to say that (U_d, V_d, U_r, V_r) forms a perfect reconstruction filter bank (see, for example, Mallat (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 3.1 is not a consequence of perfect reconstruction because filtering involves (random) time series.

The following result is a simple consequence of Proposition 3.1.

Corollary 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*

$$a^j \in l^2(\mathbb{Z}) \tag{3.7}$$

and

$$\widehat{U}_d^j(\omega) = \overline{\left(\frac{\widehat{a}^j(2\omega)}{\widehat{a}^{j-1}(\omega)}\right)}\widehat{u}(\omega), \quad \widehat{V}_d^j(\omega) = \overline{\left(\frac{1}{\widehat{a}^{j-1}(\omega)}\right)}\widehat{v}(\omega), \tag{3.8}$$

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

$$U_d^j, V_d^j \in l^1(\mathbb{Z}), \quad j \geq 1. \tag{3.9}$$

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

$$X^j = \downarrow_2 (\overline{U}_d^j * X^{j-1}), \quad \xi^j = \downarrow_2 (\overline{V}_d^j * X^{j-1}). \tag{3.10}$$

Then, for $j \geq 1$,

$$X^j = a^j * \epsilon^j \tag{3.11}$$

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,

$$\widehat{U}_r^j(\omega) = \frac{\widehat{a}^j(\omega)}{\widehat{a}^{j+1}(2\omega)}\widehat{u}(\omega), \quad \widehat{V}_r^j(\omega) = \widehat{a}^j(\omega)\widehat{v}(\omega) \tag{3.12}$$

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

$$U_r^j, V_r^j \in l^1(\mathbb{Z}), \tag{3.13}$$

then

$$X^j = U_d^j * \uparrow_2 X^{j+1} + V_r^j * \uparrow_2 \xi^{j+1}, \quad j \geq 0. \tag{3.14}$$

Definition 3.1 The decomposition of a stationary time series $X = X^0$ into the series X^j, ξ^j, $j \geq 1$, in Corollary 3.1 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 Observe that CMF u and v play reversible roles in Proposition 3.1 and Corollary 3.1. Possibly interchanging u and v, we can consider the decomposition

$$X^j = \downarrow_2 (\overline{\mathcal{U}}_d^j * X^{j-1}), \quad \xi^j = \downarrow_2 (\overline{\mathcal{V}}_d^j * X^{j-1}), \tag{3.15}$$

where

$$\widehat{\mathcal{U}}_d^j(\omega) = \overline{\left(\frac{\widehat{a}^j(2\omega)}{\widehat{a}^{j-1}(\omega)}\right)}\widehat{r}_j(\omega), \quad \widehat{\mathcal{V}}_d^j(\omega) = \overline{\left(\frac{1}{\widehat{a}^{j-1}(\omega)}\right)}\widehat{q}_j(\omega), \tag{3.16}$$

with r_j being either u or v, and q_j being the other. In this case, ξ^j are still independent, Gaussian white noise sequences, and $X^j = a^j * \epsilon^j$ with a Gaussian white noise ϵ^j.

The decomposition (3.15)–(3.16) partitions the frequency range $(0, 2\pi)$ into a different collection of subbands. In the orthogonal context, such decompositions are generally known as wavelet packets. Contrary to AWD, however, "details" ξ^j in these decompositions do not necessarily annihilate polynomial trends.

Remark 3.4 AWD can be easily extended to cyclic time series $Y = Y^0 = a^0 \circledast \epsilon^0$ given by (2.14) of length 2^J. Consider

$$Y^j = \downarrow_2 (\overline{U}_d^j \circledast Y^{j-1}), \quad \xi^j = \downarrow_2 (\overline{V}_d^j \circledast Y^{j-1}), \quad j = 1, \dots, J, \tag{3.17}$$

at decomposition, and

$$Y^j = U_r^j \circledast \uparrow_2 Y^{j+1} + V_r^j \circledast \uparrow_2 \xi^{j+1}, \quad j = 0, \dots, J-1, \tag{3.18}$$

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:

$$\widetilde{X}^j = \downarrow_2 (\overline{U}_d^j \circledast \widetilde{X}^{j-1}), \quad \widetilde{\xi}^j = \downarrow_2 (\overline{V}_d^j \circledast \widetilde{X}^{j-1}), \quad j = 1, \dots, J. \tag{3.19}$$

These relations differ from those in (3.10) 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 (3.9). Moreover, it can be verified that

$$\widetilde{X}^j = U_r^j \circledast \uparrow_2 \widetilde{X}^{j+1} + V_r^j \circledast \uparrow_2 \widetilde{\xi}^{j+1}, \quad j = 0, \dots, J-1. \tag{3.20}$$

The idea behind (3.19) 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 alike 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 3.2 The decomposition of a stationary vector $\widetilde{X}^0 = (X_0, X_1, \dots, X_{T-1})$ with $T = 2^J$ into the vectors $\widetilde{X}^j, \widetilde{\xi}^j$, $j = 1, \dots, J$, in (3.19) will be called *approximate AWD*.

Remark 3.5 Using circular convolutions in (3.19) at decomposition can be viewed as one way of dealing with the boundary when having finite data. More precisely, approximate AWD of $X = \widetilde{X}^0 = (X_0, X_1, \dots, X_{T-1})$ 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 $(X_0, X_1, \dots, X_{T-2}, X_{T-1}, X_{T-2}, \dots, X_1)$. Using circular convolutions is convenient analytically.

Remark 3.6 Another perspective on approximate AWD concerns covariance factorization. If $\widetilde{X}^0 = (X_0, X_1, \dots, X_{T-1})$ with $T = 2^J$ is a Gaussian stationary sequence, let

$$\widetilde{Y} = (\widetilde{\xi}^1, \widetilde{\xi}^2, \dots, \widetilde{\xi}^J, \widetilde{Y}^J) \tag{3.21}$$

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

$$\widetilde{Y} = \widetilde{X}^0 M \tag{3.22}$$

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

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

where Id is the identity matrix, and the variance of $\widetilde{Y}^J$ is ignored for simplicity. By using (3.22),

$$E\widetilde{X}^{0\prime}\widetilde{X}^0 = (M^{-1})'\Big(E\widetilde{Y}'\widetilde{Y}\Big)M^{-1} \approx (M^{-1})'M^{-1}, \tag{3.23}$$

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 3.1 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 1). In the applications considered here, we were motivated by (b), in view of approximate AWD (see the discussion preceding Definition 3.2). In regard to (c), one natural approximation of a series X^0 at scale 2^j is

$$X^j = \{X^0_{2^jk}\}_{k\in\mathbb{Z}}. \tag{3.24}$$

In particular, if $\widehat{a}^0(w) = \widehat{a}(w)$ enters the spectral representation (2.3) of X^0, then

$$\widehat{a}^j(w) = \frac{1}{2}\Big(\widehat{a}^{j-1}\Big(\frac{w}{2}\Big) + \widehat{a}^{j-1}\Big(\frac{w}{2}+\pi\Big)\Big) \tag{3.25}$$

is associated with the spectral representation of X^j. See Example 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.

Proposition 3.2 *Suppose that the underlying orthogonal MRA has N zero moments with factorization (2.9). 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 $|U_{d,n}^j|, |V_{d,n}^j| \leq C_j|n|^{-D-2}$, where C_j is a constant. Then, for any $j \geq 1$,*

$$\xi^j(p) = 0, \tag{3.26}$$

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

Proof: 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

$$|X^j_n| \le C(1+|n|)^D, \tag{3.27}$$

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 (3.27) holds for $j-1$ and consider it with j. Then,

$$|X^j_n| \le \sum_k |U^j_{d,k} X^{j-1}_{n-k}| \le C_1 \sum_k (1+|k|)^{-D-2}(1+|n-k|)^D$$

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

where constants C_i may depend on j. Using (3.27) and the assumed bound for $V^j_{d,n}$, the argument above also shows that ξ^j is well-defined.

To prove (3.26), we will first establish the formula

$$\widehat{X}^j(\omega) = \frac{1}{2^j}\sum_{n=0}^{2^j-1}\left\{\overline{\prod_{k=1}^{j}\widehat{U}^k_d\Big(\frac{\omega}{2^{j+1-k}}+b_{n,k}\Big)}\right\}\widehat{p}\Big(\frac{\omega}{2^j}+\frac{n\pi}{2^{j-1}}\Big), \tag{3.28}$$

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 (3.28) is trivial for $j=1$. Assume it holds for $j-1$ and consider it for j. Then,

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

$$= \frac{1}{2^j}\sum_{n=0}^{2^{j-1}-1}\prod_{k=1}^{j}\left(\overline{\widehat{U}^k_d\Big(\frac{\omega}{2^{j+1-k}}+b_{n,k}\Big)}\widehat{p}\Big(\frac{\omega}{2^j}+\frac{n\pi}{2^{j-2}}\Big) + \overline{\widehat{U}^k_d\Big(\frac{\omega}{2^{j+1-k}}+b'_{n,k}\Big)}\widehat{p}\Big(\frac{\omega}{2^j}+\frac{\pi}{2^{j-1}}+\frac{n\pi}{2^{j-2}}\Big)\right)$$

$$= \frac{1}{2^j}\sum_{n=0}^{2^{j-1}-1}\prod_{k=1}^{j}\left(\overline{\widehat{U}^k_d\Big(\frac{\omega}{2^{j+1-k}}+b_{n,k}\Big)}\widehat{p}\Big(\frac{\omega}{2^j}+\frac{2n\pi}{2^{j-1}}\Big) + \overline{\widehat{U}^k_d\Big(\frac{\omega}{2^{j+1-k}}+b'_{n,k}\Big)}\widehat{p}\Big(\frac{\omega}{2^j}+\frac{(2n+1)\pi}{2^{j-1}}\Big)\right)$$

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

Since

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

and

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

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

$$\overline{\widehat{v}\Big(\frac{\omega}{2}\Big)}\widehat{p}\Big(\frac{\omega}{2^j}+\frac{2n\pi}{2^{j-1}}\Big) = 0 \tag{3.29}$$

and

$$\overline{\widehat{v}\Big(\frac{\omega}{2}+\pi\Big)}\widehat{p}\Big(\frac{\omega}{2^j}+\frac{(2n+1)\pi}{2^{j-1}}\Big)=0. \tag{3.30}$$

Observe that, by using (3.28), the relation (3.29) 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 (3.30). □

Remark 3.7 AWD introduced here are related to those in continuous time, considered by Didier and Pipiras (2006). Suppose that a stationary, Gaussian time series X^0 is thought as

$$X_n^0\approx 2^{-j_0/2}X\Big(\frac{n}{2^{j_0}}\Big),\quad n\in\mathbb{Z}, \tag{3.31}$$

where $X=\{X(t)\}_{t\in\mathbb{R}}$ is a stationary Gaussian process in continuous time and $j_0\in\mathbb{Z}$ is fixed. Suppose that X is given by it spectral representation

$$X(t)=\int_{\mathbb{R}}e^{itx}\widehat{g}(x)d\widehat{B}(x),\quad t\in\mathbb{R}, \tag{3.32}$$

where $\widehat{g}(x)=\int_{\mathbb{R}}e^{-ixu}g(u)du$ is the Fourier transform of g, 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 $\{B_1(x)\}_{x\geq0}$ and $\{B_2(x)\}_{x\geq0}$ such that $EB_1(1)^2=EB_2(1)^2=(4\pi)^{-1}$.

In this case, by the results of Didier and Pipiras (2006), and under suitable conditions, we have

$$X_n^j = \int_{\mathbb{R}}X(t)\Phi_{j_0-j}(t-2^{-(j_0-j)}n)dt, \tag{3.33}$$

$$\xi_n^j = \int_{\mathbb{R}}X(t)\Psi_{j_0-j}(t-2^{-(j_0-j)}n)dt, \tag{3.34}$$

where

$$\widehat{\Phi}_{j_0-j}(x) = \overline{\Big(\frac{\widehat{a}_j(2^{-(j_0-j)}x)}{\widehat{g}(x)}\Big)}2^{-(j_0-j)/2}\widehat{\phi}(2^{-(j_0-j)}x), \tag{3.35}$$

$$\widehat{\Psi}_{j_0-j}(x) = \overline{\widehat{g}(x)^{-1}}2^{-(j_0-j)/2}\widehat{\psi}(2^{-(j_0-j)}x) \tag{3.36}$$

with a wavelet ψ and a scaling function ϕ of an orthogonal MRA. In other words, X^j and ξ^j can be viewed as approximation and detail terms in the AWD of the process X considered by Didier and Pipiras (2006). The continuous time perspective can be useful, for example, in showing easier that the details ξ^j eliminate polynomials of degree D as long as wavelet ψ has $N>D$ zero moments.

4 Examples of AWD

We provide here several examples of AWD. We would like the associated filters $U_d^j, V_d^j, U_r^j, V_r^j$ to decay to zero fast (see (b) following Remark 3.6). 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 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

$$\widehat{a}(w) = (1 - e^{-iw})^{-s} \tag{4.1}$$

(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

$$\widehat{a^j}(w) = \widehat{a}(w), \tag{4.2}$$

for any $j \geq 1$, and focus on the definition (3.12) of U_r^j, V_r^j. Note that

$$\frac{\widehat{a}(w)}{\widehat{a}(2w)} = \frac{(1-e^{-iw})^{-s}}{(1-e^{-i2w})^{-s}} = (1+e^{-iw})^{s} = \sum_{k=0}^{\infty} f_k^{(s)} e^{-iwk}, \tag{4.3}$$

$$\widehat{a}(w) = (1-e^{-iw})^{-s} = \sum_{k=0}^{\infty} g_k^{(-s)} e^{-iwk} \tag{4.4}$$

are the two filters entering (3.12). These filters, in fact, decay extremely slowly: one can show that, as $k \to \infty$,

$$f_k^{(s)} \sim (-1)^k \frac{k^{-s-1}}{\Gamma(-s)}, \quad g_k^{(-s)} \sim \frac{k^{s-1}}{\Gamma(s)}. \tag{4.5}$$

(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 (2.9), observe that

$$\widehat{U}_r(w) \equiv \widehat{U}_r^j(w) = (1+e^{-iw})^{s+N}\widehat{u}_{0,N}(w), \quad \widehat{V}_r(w) \equiv \widehat{V}_r^j(w) = (1-e^{-iw})^{-s+N}\widehat{v}_{0,N}(w). \tag{4.6}$$

By (4.3)–(4.5), we now have

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

$$(1-e^{-iw})^{-s+N} = \sum_{k=0}^{\infty} g_k^{(-s+N)} e^{-iwk} \quad \text{with} \quad g_k^{(-s+N)} \sim \frac{k^{s-N-1}}{\Gamma(s-N)}, \tag{4.7}$$

as $k \to \infty$. Comparing (4.7) with (4.5), 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 (4.6) 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 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 $s = 0.25$

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 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.

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 1 are not sensitive to the value of s. In particular, the change in the results is small as s approaches $1/2$.

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 (3.8) 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 (4.1) appears to be a suitable choice for FARIMA$(0,s,0)$ time series.

Remark 4.1 The faster decay in (4.7) has also the following simple explanation that is useful more generally. According to (4.3)–(4.5), 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 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

$$\widehat{a}(w)=(1-a_1e^{-iw})^{-1}, \tag{4.8}$$

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

$$a^j(w)\equiv 1,\quad j\geq 1. \tag{4.9}$$

Then,

$$\widehat{U}_d^1(w)=(1-a_1e^{iw})\widehat{u}(w),\quad \widehat{V}_d^1(w)=(1-a_1e^{iw})\widehat{v}(w) \tag{4.10}$$

and

$$\widehat{U}_d^j(w)=\widehat{u}(w),\quad \widehat{V}_d^j(w)=\widehat{v}(w),\quad j\geq 2. \tag{4.11}$$

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 (4.9),

$$\widehat{U}_r^0(w) = \frac{\widehat{u}(w)}{1 - a_1 e^{-iw}}, \quad \widehat{V}_r^0(w) = \frac{\widehat{v}(w)}{1 - a_1 e^{-iw}} \tag{4.12}$$

and

$$\widehat{U}_r^j(w) = \widehat{u}(w), \quad \widehat{V}_r^j(w) = \widehat{v}(w), \quad j \geq 1. \tag{4.13}$$

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 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

$$\widehat{a}^j(w) = (1 - a_1^{2^j} e^{-iw})^{-1} \tag{4.14}$$

so that

$$\widehat{U}_r^j(w) = (1 + a_1^{2^j} e^{-iw})\widehat{u}(w), \quad \widehat{V}_r^j(w) = \frac{\widehat{v}(w)}{1 - a_1^{2^j} e^{-iw}}, \quad j \geq 0. \tag{4.15}$$

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 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 $(1, a_1, a_1^2, \ldots)$ with the filter v. Note that the $(2N + j)$th nonzero element of the convolution is

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

and decays as a geometric sequence. In Table 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 2), this effect is no longer present.

Note also that, with the choice (4.14) for AWD, the approximations X^j become AR(1) time series with the parameters $a_1^{2^j}$. The decomposition filters associated with (4.14) are

$$\widehat{U}_d^j(w) = \frac{\widehat{u}(w)}{1 + a_1^{2^j} e^{iw}}, \quad \widehat{V}_d^j(w) = (1 - a_1^{2^j} e^{iw})\widehat{v}(w). \tag{4.16}$$

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 (4.14) 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 4.1). In this case, the AWD with (4.9) is probably as best as one can do (though see also Remark 4.3). Note that, with (4.9), increasing the number of zero moments make the filters U_r^0

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 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.

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

$$\widehat{a}(w) = 1 + b_1 e^{-iw}, \tag{4.17}$$

where $-1 < b_1 < 1$ ($b_1 \neq 0$). Since $\widehat{a}(w)$ in (4.17) is reciprocal to that in (4.8), 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 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 (3.24) are all (up to a constant) Gaussian white noise sequences or

$$\widehat{a^j}(w) \equiv 1. \tag{4.18}$$

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 (3.24) are associated with

$$\widehat{a^j}(w) = (1 - a_1^{2^j} e^{-iw})^{-1}. \tag{4.19}$$

Observe that (4.18) and (4.19) are exactly what was proposed for AWD at reconstruction for MA(1) and AR(1) time series in Example 4.2 above.

Remark 4.3 Let X^0 be an MA(1) time series with $\widehat{a}(w) = 1 + b_1 e^{-iw}$, $-1 < b_1 < 1$ ($b_1 \neq 0$). Recall from Example 4.2 that, when $0 < b_1 < 1$ and especially when b_1 is close to 1, the number of zero moments of the underlying MRA does not help for decay of associated filters in any AWD. The extension considered in Remark 3.3 seems to overcome this, though at the expense of annihilation of polynomial trends for "detail" sequences. More precisely, for the MA(1) time series taken above, consider

$$\widehat{a}^j(w) = 1 - b_1^{2^j} e^{-iw} \tag{4.20}$$

and

$$r_1 = v,\ q_1 = u, \quad r_j = u,\ q_j = v,\ j \geq 2. \tag{4.21}$$

in (3.15)–(3.16). Then,

$$\widehat{\mathcal{U}}_d^1(\omega) = (1 - b_1 e^{iw})\widehat{v}(w), \quad \widehat{\mathcal{V}}_d^1(\omega) = \frac{\widehat{u}(w)}{1 + b_1 e^{iw}} \tag{4.22}$$

and

$$\widehat{\mathcal{U}}_d^j(\omega) = (1 + b_1^{2^{j-1}} e^{iw})\widehat{u}(w), \quad \widehat{\mathcal{V}}_d^j(\omega) = \frac{\widehat{v}(w)}{1 - b_1^{2^{j-1}} e^{iw}}, \quad j \geq 2. \tag{4.23}$$

When $0 < b_1 < 1$ and especially when b_1 is close to 1, the zero moments make the decay of the filters $\mathcal{V}_d^j$ faster. (The filters $\mathcal{U}_d^1$ are already finite and short if CMF u and v are such.)

Remark 4.4 Observe from the examples above that AWD are based on easy manipulations with $\widehat{a}(w)$. This may not be the case for many other stationary time series. An example is fractional Gaussian noise (fGn) sequence X_n defined as the increments of fractional Brownian motion $\{B_H(t)\}_{t\in\mathbb{R}}$, $H \in (0,1)$, that is, $X_n = B_H(n) - B_H(n-1)$, $n \in \mathbb{Z}$. One can show (Proposition 2.1 in Beran (1994)) that

$$|\widehat{a}(w)|^2 = C_H(1 - \cos w) \sum_{n=\infty}^{\infty} |2\pi n + w|^{-2H-1}, \quad w \in (0, 2\pi), \tag{4.24}$$

where $C_H = EBHH^2(1)\sin(\pi H)\Gamma(2H+1)$. But the difficulty in dealing with (4.24) is not intrinsic to AWD – one would have as much trouble in dealing with (4.24) with conventional Fourier methods.

5 Applications of AWD

We consider here applications of AWD to simulation (Section 5.1) and MLE (Section 5.2). Simulation uses AWD at reconstruction and MLE uses AWD at decomposition.

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 (3.14) 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 depend on the time series to simulate. For example, in the case of (4.6), the reconstruction filters are the same for all j. The second step refers to the fact that application of the reconstruction scheme (3.14) 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 (3.14). 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 (3.14) with $\widetilde{U}_r^{J-1}, \widetilde{V}_r^{J-1}$ to X_J of length L_J+1, we obtain $2(L_J+1)-1-L_J = L_J+2$ number of 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 $O(2^J \log 2^J)$.

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 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 (3.18), that is,

$$\widetilde{X}^k = U_r^k \circledast \uparrow_2 \widetilde{X}^{k+1} + V_r^k \circledast \uparrow_2 \widetilde{\xi}^{k+1}, \quad k = 0, \dots, K-1, \tag{5.1}$$

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 (5.1) 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 3.4 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 (2.15) 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

$$\widetilde{r}^0(n) = \frac{1}{2^K} \sum_{m=0}^{2^K-1} e^{i\frac{2\pi mn}{2^K}} \Big|\widehat{a}^0\Big(\frac{2\pi m}{2^K}\Big)\Big|^2, \tag{5.2}$$

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

It may appear from (5.2) that, as $K \to \infty$,

$$\widetilde{r}^0(n) \approx \frac{1}{2\pi} \int_0^{2\pi} e^{inw} |\widehat{a}^0(w)|^2 dw = r^0(n), \tag{5.3}$$

where r^0 is the autocovariance function of X^0. The approximation (5.3) 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

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

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

$$\leq \sup_{w\in(0,2\pi)}\Big|\frac{\partial|\widehat{a}^0(w)|^2}{\partial w}\Big|\frac{2\pi}{2^K}+\sup_{w\in(0,2\pi)}|\widehat{a}^0(w)|^2\frac{2\pi T}{2^K}, \tag{5.4}$$

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 (5.4). Making use of the first generated values in the context of orthogonal wavelet decompositions can also be found in Parcival and Walden (2000), Section 9.2, but without the explicit connection to circular time series and the resulting error (5.4) 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 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 (5.1) 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 (5.4) 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 $T/2^J$.

5.2 Maximum likelihood estimation

The approximate covariance factorization (3.23) discussed in Remark 3.6 naturally leads to the following Gaussian MLE based on AWD. Given the observations $\widetilde{X}^0=(X_0,X_1,\ldots,X_{T-1})$, the negative log-likelihood is (up to additive and multiplicative constants)

$$\log|\widetilde{\Sigma}_\theta|+\widetilde{X}^0\widetilde{\Sigma}_\theta^{-1}\widetilde{X}^{0\prime}, \tag{5.5}$$

where $\widetilde{\Sigma}_\theta$ is the covariance matrix of the model with unknown parameters θ, and $|\cdot|$ denotes the determinant. As in Remark 3.6, a vector $\widetilde{Y}_\theta$ of detail coefficients in approximate AWD can be written as

$$\widetilde{Y}_\theta=\widetilde{X}^0 M_\theta \tag{5.6}$$

for a matrix M_θ which depends on the model parameters θ. By (3.23), $\widetilde{\Sigma}_\theta^{-1}\approx M_\theta M_\theta'$ and hence the expression (5.5) is approximately equal to

$$\log|\widetilde{\Sigma}_\theta|+\widetilde{Y}_\theta\widetilde{Y}_\theta'. \tag{5.7}$$

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 Szegö (1958), and consider

$$\frac{T}{\pi}\int_0^{2\pi}\log|\widehat{a}_\theta(w)|dw+\widetilde{Y}_\theta\widetilde{Y}_\theta'. \tag{5.8}$$

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

In Tables 3–4 at the end of the paper, 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). We also consider a non-Gaussian, exponential distribution for the generated error terms in the MA(1) case (Table 4) and, for FARIMA$(0, s, 0)$ model, we report results with a superimposed linear trend $-1+0.5t$ (Table 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, $j = 1, \ldots, J$, is approximated by

$$\frac{2^{j+1}}{2\pi} \int_{2\pi/2^j}^{2\pi/2^{j+1}} |\widehat{a}(w)|^2 dw. \tag{5.9}$$

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 3–4 suggest that MLE based on AWD works quite well. It is generally comparable to Whittle MLE and is superior over it in in the AR(1) case with $a_1 = \pm 0.9$. It is generally superior over MLE based on OWD which is likely the result of the approximation (5.9). 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 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 3.5) 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.

References

Abry, P., Flandrin, P., Taqqu, M. S. & Veitch, D. (2003), Self-similarity and long-range dependence through the wavelet lens, *in* 'Theory and applications of long-range dependence', Birkhäuser Boston, Boston, MA, pp. 527–556.

Akansu, A. N. & Haddad, R. A. (2001), *Multiresolution Signal Decomposition*, second edn, Academic Press Inc., San Diego, CA.

Benassi, A. & Jaffard, S. (1994), Wavelet decomposition of one- and several-dimensional Gaussian processes, *in* 'Recent advances in wavelet analysis', Vol. 3 of *Wavelet Anal. Appl.*, Academic Press, Boston, MA, pp. 119–154.

Beran, J. (1994), *Statistics for Long-Memory Processes*, Chapman & Hall, New York.

Brockwell, P. J. & Davis, R. A. (1991), *Time Series: Theory and Methods*, 2nd edn, Springer-Verlag, New York.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, SIAM Philadelphia. CBMS-NSF series, Volume 61.

Didier, G. & Pipiras, V. (2006), Gaussian stationary processes: adaptive wavelet decompositions, discrete approximations and their convergence, Preprint. Available at http://www.stat.unc.edu/faculty/pipiras.

Dietrich, C. R. & 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**(4), 1088–1107.

Dijkerman, R. W. & Mazumdar, R. R. (1994), 'On the correlation structure of the wavelet coefficients of fractional Brownian motion', *IEEE Transactions on Information Theory* **40**(5), 1609–1612.

Donoho, D. L. (1995), 'Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition', *Applied and Computational Harmonic Analysis* **2**(2), 101–126.

Donoho, D. L. & Johnstone, I. M. (1994), 'Ideal spatial adaptation by wavelet shrinkage', *Biometrika* **81**(3), 425–455.

Donoho, D. L. & Johnstone, I. M. (1995), 'Adapting to unknown smoothness via wavelet shrinkage', *Journal of the American Statistical Association* **90**(432), 1200–1224.

Grenander, U. & Szego, G. (1958), *Toeplitz Forms and their Applications*, Chelsea, New York.

Jaffard, S., Lashermes, B. & 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.

Mallat, S. (1998), *A Wavelet Tour of Signal Processing*, Academic Press Inc., San Diego, CA.

Mallat, S., Papanicolaou, G. & Zhang, Z. (1998), 'Adaptive covariance estimation of locally stationary processes', *The Annals of Statistics* **26**(1), 1–47.

Meyer, Y., Sellan, F. & 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**(5), 465–494.

Moulines, E., Roueff, F. & 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. & Kroisandt, G. (2000), 'Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum', *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **62**(2), 271–292.

Ossiander, M. & Waymire, E. C. (2000), 'Statistical estimation for multiplicative cascades', *The Annals of Statistics* **28**(6), 1533–1560.

Percival, D. B. & Walden, A. T. (2000), *Wavelet Methods for Time Series Analysis*, Vol. 4 of *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, Cambridge.

Pipiras, V. (2005), 'Wavelet-based simulation of fractional Brownian motion revisited', *Applied and Computational Harmonic Analysis* **19**(1), 49–60.

Resnick, S., Samorodnitsky, G., Gilbert, A. & Willinger, W. (2003), 'Wavelet analysis of conservative cascades', *Bernoulli* **9**(1), 97–135.

Ruiz-Medina, M. D., Angulo, J. M. & Anh, V. V. (2003), 'Fractional-order regularization and wavelet approximation to the inverse estimation problem for random fields', *Journal of Multivariate Analysis* **85**(1), 192–216.

Vaidyanathan, P. P. & Akkarakaran, S. (2001), 'A review of the theory and applications of optimal subband and transform coders', *Applied and Computational Harmonic Analysis* **10**(3), 254–289.

Veitch, D. & Abry, P. (1999), 'A wavelet-based joint estimator of the parameters of long-range dependence', *IEEE Transactions on Information Theory* **45**(3), 878–897.

Zhang, J. & Walter, G. (1994), 'A wavelet-based KL-like expansion for wide-sense stationary random processes', *IEEE Transactions on Signal Processing* **42**(7), 1737–1745.

Gustavo Didier and Vladas Pipiras
Dept. of Statistics and Operations Research
UNC at Chapel Hill
CB#3260, Smith Bldg.
Chapel Hill, NC 27599, USA
didier@unc.edu, pipiras@email.unc.edu

Model	θ_0	T	Whittle		N	AWD		OWD	
			bias	rMSE		bias	rMSE	bias	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
FARIMA$(1, s, 0)$[§]	$a_1 = 0.5$	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$(1, s, 0)$[†]	$a_1 = 0.5$	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$(1, s, 0)$[†]	$a_1 = 0.5$	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: 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).

Model	θ_0	T	Whittle bias	Whittle rMSE	N	AWD bias	AWD rMSE	OWD bias	OWD rMSE
MA(1)* $\widehat{a}^j(w) \equiv 1$	$b_1 = 0.5$	2^{10}	-0.0002	0.0283	2	-0.0006	0.0277	-0.0814	0.0949
					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)* $\widehat{a}^j(w) = 1 - (-b_1)^{2^j} e^{-iw}$	$b_1 = 0.5$	2^{10}			2	-0.0014	0.0277		
					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		
MA(1)* $\widehat{a}^j(w) \equiv 1$	$b_1 = 0.5$ (exp)	2^{10}	0.0006	0.0276	2	-0.0003	0.0264	-0.0808	0.0960
					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

Table 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).