J. Appl. Prob. Special Vol. **41A,** 81–98 (2004)

STOCHASTIC METHODS AND THEIR APPLICATIONS

Papers in honour of CHRIS HEYDE
Edited by J. GANI and E. SENETA

Part 2. Estimation methods

BOUNDARY COIFLETS FOR WAVELET SHRINKAGE IN FUNCTION ESTIMATION

IAIN M. JOHNSTONE, *Stanford University*

Department of Statistics, Sequoia Hall, Stanford University, Stanford, CA 94305, USA.
Email address: imj@stat.stanford.edu

BERNARD W. SILVERMAN, *Oxford University*

St Peter's College, Oxford University, Oxford OX1 2DL, UK.
Email address: bernard.silverman@spc.ox.ac.uk

APPLIED PROBABILITY TRUST
APRIL 2004

BOUNDARY COIFLETS FOR WAVELET SHRINKAGE IN FUNCTION ESTIMATION

IAIN M. JOHNSTONE
BERNARD W. SILVERMAN

Abstract

There are standard modifications of certain compactly supported wavelets that yield orthonormal bases on a bounded interval. We extend one such construction to those wavelets, such as 'coiflets', that may have fewer vanishing moments than had to be assumed previously. Our motivation lies in function estimation in statistics. We use these boundary-modified coiflets to show that the discrete wavelet transform of finite data from sampled regression models asymptotically provides a close approximation to the wavelet transform of the continuous Gaussian white noise model. In particular, estimation errors in the discrete setting of computational practice need not be essentially larger than those expected in the continuous setting of statistical theory.

Keywords: Coiflets; Gaussian white noise model; minimax risk; multiresolution analysis; nonparametric regression; preconditioning; quadrature error; vanishing moments; wavelet shrinkage

2000 Mathematics Subject Classification: Primary 62G08; 42C40
Secondary 62G20; 65T60

1. Introduction

The simplest setting for much of the theory of nonparametric function estimation is the Gaussian white noise model

$$Y(t) = \int_0^t f(s)\,\mathrm{d}s + \varepsilon W(t), \qquad 0 \le t \le 1, \tag{1}$$

in which $f \in L_2[0, 1]$ is unknown and W is standard Brownian motion. If ψ_λ is an orthonormal basis for $L_2[0, 1]$, then the model takes the sequence form

$$y_\lambda = \theta_\lambda + \varepsilon z_\lambda \tag{2}$$

by setting $y_\lambda = \int \psi_\lambda\,\mathrm{d}Y$, $\theta_\lambda = \int \psi_\lambda f$ and $z_\lambda = \int \psi_\lambda\,\mathrm{d}W$. Here $\{z_\lambda\}$ are independent and identically distributed $N(0, 1)$ and the Parseval relation

$$\int (\hat{f} - f)^2 = \sum_\lambda (\hat{\theta}_\lambda - \theta_\lambda)^2$$

expresses the integrated squared error of function estimation in terms of the ℓ_2 error of estimated coefficients.

If $\{\psi_\lambda\}$ is a *wavelet* basis for $L_2[0, 1]$, then quite simple estimators can be built from coordinatewise shrinkage or thresholding of the wavelet coefficients y_λ, and these estimators have strong mean square optimality properties. See, for example, [4], [5], [7], [8], [9].

In practice, however, instrumentally acquired data that is to be digitally processed is typically discrete. A common simplification is the equally spaced regression model

$$Y_i = f(t_i) + \varepsilon_i, \qquad i = 1, \ldots, N, \tag{3}$$

with $f \in L_2[0, 1]$ unknown, $t_i = i/N$ and ε_i independent and $N(0, \sigma^2)$ distributed. After taking a discrete wavelet transform, we might arrive at empirical wavelet coefficients $\{\tilde{y}_\lambda\}$ satisfying

$$\tilde{y}_\lambda = \tilde{\theta}_\lambda + \varepsilon_N \tilde{z}_\lambda, \qquad \lambda \in \Lambda_N, \tag{4}$$

with $|\Lambda_N| \doteq N$ and $\varepsilon_N = \sigma/\sqrt{N}$. Here the $\{\tilde{\theta}_\lambda\}$ are the corresponding discrete wavelet transform of the *sampled* function values $\{f(t_i)\}$.

It is common practice to take estimators $\hat{\theta}(y)$ that are motivated, derived and analysed in the Gaussian white noise model of (1) and (2) and apply them in computer code to discrete wavelet transform data (Y_i) and $(\tilde{y}_\lambda)$, whether or not assumptions (3) and (4) apply exactly.

What can be said about the mean squared error of the resulting estimator $\hat{\theta}(\tilde{y})$? For example, is $\mathrm{E}\,\|\hat{\theta}(\tilde{y}) - \theta\|_2^2$ approximately as good as $\mathrm{E}\,\|\hat{\theta}(y) - \theta\|_2^2$? The question was studied in [6], where Deslauriers–Dubuc interpolation was used to pass from discrete to continuous models. In this paper, we adopt a different approach. Motivated in part by a study of certain empirical Bayes wavelet shrinkage methods [9], we make the connection by restricting to certain orthonormal wavelet bases. One consequence is a certain simplification (see Sections 4, 5) of the proofs in [6].

Let us first concentrate on points t in the interior of [0, 1], temporarily ignoring boundary issues. Let $N = 2^J$ and $\phi_{J,l}(t) = 2^{J/2}\phi(2^J t - l)$ be the corresponding (interior) scaling functions at level J. The continuous wavelet coefficients θ_{jk} at coarser scales $j < J$ are related to the finest scaling coefficients $\langle f, \phi_{J,l}\rangle$ in exactly the same way that the discrete wavelet transform coefficients $\tilde{\theta}_{jk}$ are related to $f(t_l)$.

If the scaling function ϕ has $\int \phi = 1$ and vanishing moments $\int t^r \phi(t)\,\mathrm{d}t = 0$ for $1 \le r \le R - 1$, then by Taylor expansion

$$\langle f, \phi_{J,l}\rangle - 2^{-J/2} f(l 2^{-J}) = O(\|f^{(R)}\|_\infty 2^{-(R+1/2)J}). \tag{5}$$

Thus, if the *scaling* function ϕ possesses some vanishing moments, we can expect the continuous wavelet transform to be well approximated by the discrete wavelet transform.

A family of such scaling functions having vanishing moments and *compact support* was constructed by Daubechies [3], who christened them ‘coiflets’.

Now to the boundary issues. The statistical models (1) and (3) are intimately tied to a compact interval, conventionally [0, 1], and so the question of behaviour near 0 and 1 arises. Families of compactly supported orthonormal wavelet bases adapted to [0, 1] have been constructed by Cohen, Daubechies, Jawerth and Vial in [1] and [2] and others. A key consequence of the compact support is that only a constant number of the 2^j wavelets at resolution level j ‘feel’ the boundaries, providing good control of boundary effects.

However, the scaling functions used in these constructions do not have sufficient vanishing moments to rely on (5), and so in this paper we set out an extension of the argument of [2] to establish the existence of orthobases for $L_2[0, 1]$ based on, for example, coiflets.

Section 2 sets out the existence result for wavelet bases on [0, 1]. Section 3 derives the analogue of the quadrature error bound (5) that is valid also near the boundaries (after a preconditioning step). Section 4 uses both preceding sections to derive bounds on the differences $\tilde{\theta}_\lambda - \theta_\lambda$ between discrete and continuous wavelet coefficients. Finally Section 5 draws some conclusions for the estimation error in discrete data settings that address the issues raised earlier.

2. Adapting orthobases to [0, 1]

Assume that (ϕ, ψ) are orthonormal scaling and wavelet functions supported in $[-S+1, S]$, and that ψ has vanishing moments of degree up to $R-1$, where $R \geq 1$. (The proof of Theorem 1 below contains precise assumptions.) We describe a construction of an orthonormal basis of wavelets and scaling functions for $L_2[0, 1]$. Motivated by the example of coiflets, for which the support length is $3R - 1$, we are specifically concerned with the situation $R < S$, and so modify the construction of [1] given for $R = S$. However, we do not yet make any vanishing moments assumptions on ϕ.

The idea is to modify the wavelets and scaling functions developed for $L_2(\mathbb{R})$ near the boundaries 0 and 1. At each level j, the modified construction retains the $2^j - 2S + 2$ 'interior' scaling functions, which have support entirely within [0, 1]. In addition, there are R boundary scaling functions at each end. There are $S - 1$ boundary wavelets at each end and $2^j - 2S + 2$ interior wavelets. The orthonormality properties of the wavelet array are maintained, as are the vanishing moment properties of the wavelets themselves, but the filters used in the corresponding multiresolution analysis are modified at each end.

To be explicit, the construction is based on boundary scaling functions ϕ_k^{B} for $k = -R, -R+1, \ldots, R-2, R-1$, and boundary wavelets ψ_k^{B} for $k = -S+1, -S+2, \ldots, S-1, S-2$. The support of these functions is contained in $[0, 2S-2]$ for $k \geq 0$ and in $[-(2S-2), 0]$ for $k < 0$. We fix a coarse resolution level L such that $6S - 6 \leq 2^L$. At every level $j \geq L$, the scaling functions are then defined by

$$\phi_{jk}(x) = \begin{cases} 2^{j/2}\phi_k^{\mathrm{B}}(2^j x), & 0 \leq k \leq R-1, \\ 2^{j/2}\phi(2^j x - k), & S-1 \leq k \leq 2^j - S, \\ 2^{j/2}\phi_{k-2^j}^{\mathrm{B}}(2^j(x-1)), & 2^j - R \leq k \leq 2^j - 1. \end{cases} \tag{6}$$

The support properties of the boundary scaling functions, and of the original scaling function ϕ, are such that all these functions are supported on [0, 1]. Define V_j as the span of all these ϕ_{jk} as k runs over the three sets of indices in (6). A key point is that V_j contains polynomials of degree up to $R-1$; this will allow the Besov function space characterizations we need, so long as R is chosen appropriately. Notice that there are no functions defined when $R \leq k < S-1$ or when $2^j - S < k < 2^j - R$. There are no such gaps in the definition of the wavelets; we have

$$\psi_{jk}(x) = \begin{cases} 2^{j/2}\psi_k^{\mathrm{B}}(2^j x), & 0 \leq k \leq S-2, \\ 2^{j/2}\psi(2^j x - k), & S-1 \leq k \leq 2^j - S, \\ 2^{j/2}\psi_{k-2^j}^{\mathrm{B}}(2^j(x-1)), & 2^j - S + 1 \leq k \leq 2^j - 1. \end{cases} \tag{7}$$

The result is that we have a construction with 2^j wavelets at each level. Let W_j be their linear span. The $S-1$ wavelets at each end are boundary wavelets, which have the same smoothness, on [0, 1], and the same vanishing moments as the original wavelets but are otherwise modified. The $2^j - 2S + 2$ interior wavelets are not affected by the boundary construction, and depend only on the $2^J - 2S$ interior scaling functions at the finest scale. There will be $2^j - 2(S - R - 1)$ scaling functions and scaling coefficients at every level, so in particular at the coarsest level L there will be $2^L - 2(S - R - 1)$ scaling coefficients. It is convenient at every level $j \geq L$ to define $\mathcal{K}_j^{\mathrm{B}}$ to be the set of k for which ψ_{jk} is a scaled version of a boundary wavelet, and $\mathcal{K}_j^{\mathrm{I}}$ to be the set of k for which ψ_{jk} is a scaled version of ψ itself. Let $\mathcal{K}_{L-1}$ be the set of indices for which the scaling function ϕ_{Lk} is defined.

Theorem 1. *Assume that* $2^L \geq 6S-6$. *The construction can be carried out so that* (i) *for each* $j \geq L$, *the scaling functions* $\{\phi_{jk}\}$ *defined in (6) form an orthonormal basis for* V_j, *and* (ii) *the collection*

$$\{\phi_{Lk},\ k \in \mathcal{K}_L\} \cup \{\psi_{jk},\ 0 \leq k \leq 2^j - 1,\ j \geq L\} \tag{8}$$

defined in (6) and (7) forms an orthonormal basis for $L_2[0, 1]$

Remark 1. Let J be fixed. As usual, the discrete wavelet transform $\mathcal{W}$ accomplishes the change of basis in V_J from the scaling functions $\{\phi_{Jl}\}$ to the wavelet basis given by

$$\{\phi_{Lk}\} \cup \{\psi_{jk},\ L \leq j < J,\ 0 \leq k \leq 2^j - 1\}. \tag{9}$$

Thus, the entries of the orthogonal matrix $\mathcal{W}$ are given by

$$\mathcal{W}_{jk,l} = \langle \psi_{jk}, \phi_{Jl} \rangle$$

if we abuse notation and write $\psi_{L-1,k}$ for ϕ_{Lk}.

The proof of Theorem 1 adapts methods and notation from [1] and [2] to the case $R < S$, which leads to two changes. The first is described in Remark 2 below. The second is that we do not give an explicit construction of the $S-1$ boundary wavelets at each end, Proposition 2 merely establishes their existence.

For the proof of this theorem, it is convenient to change the convention concerning the support of ϕ and ψ. The assumptions we need on ϕ and ψ are set out using this temporary convention:

(i) We have $\operatorname{supp}\phi = \operatorname{supp}\psi = [0, 2S-1]$. The scaling function ϕ satisfies, for some set of filter coefficients h_n, the two-scale relation

$$\phi(x) = 2\sum_{n=0}^{2S-1} h_n \phi(2x-n), \tag{10}$$

and similarly, with $g_n = (-1)^n h_{2S-1-n}$, the wavelet ψ satisfies

$$\psi(x) = 2\sum_{n=0}^{2S-1} g_n \phi(2x-n). \tag{11}$$

(ii) Set as usual $\phi_{jk}(x) = 2^{j/2}\phi(2^j x - k)$, and similarly for $\psi_{jk}(x)$. For each $J \in \mathbb{Z}$ the collections $\{\phi_{Jk},\ k \in \mathbb{Z}\} \cup \{\psi_{jk},\ j \geq J,\ k \in \mathbb{Z}\}$ each form an orthonormal basis for $L_2(\mathbb{R})$.

(iii) The wavelet ψ has R vanishing moments, i.e.

$$\int t^k \psi(t)\,\mathrm{d}t = 0, \qquad k = 0, 1, \ldots, R-1.$$

The vanishing moment condition implies (see e.g. [10, Theorem 7.4]) that there exist polynomials $P_i(x)$ of exact degree i such that

$$\sum_k k^i \phi(x-k) = P_i(x), \qquad 0 \leq i \leq R-1.$$

Following [2], we use these relationships to define functions on $[0, \infty)$:

$$\phi^{\mathrm{l},i}(x) = \left\{ P_i(x) - \sum_{k \geq 0} k^i \phi(x-k) \right\} \mathbf{1}_{[0,\infty)} = \sum_{k<0} k^i \phi(x-k)\, \mathbf{1}_{[0,\infty)} \tag{12}$$

for $0 \leq i \leq R-1$. Clearly $\operatorname{supp} \phi^{\mathrm{l},i} \subset [0, 2S-2]$.

Remark 2. The two variations upon [1] and [2], the possibility of which is already mentioned there, are: (i) we again construct as many functions as there are vanishing moments, but now $R < S$, and (ii) the function $\phi(x-k)$ for $k = 0$ is not included in the right-hand sum defining $\phi^{\mathrm{l},i}$; when $R < S$, there is no need to 'sacrifice' the interior scaling functions whose supports touch the ends of the interval.

Similarly, we define R functions on $x \leq 0$ via

$$\phi^{\mathrm{r},i}(x) = \left\{ P_i(x) - \sum_{k<2-2S} k^i \phi(x-k) \right\} \mathbf{1}_{(-\infty,0]} = \sum_{k \geq 2-2S} k^i \phi(x-k)\, \mathbf{1}_{(-\infty,0]}, \tag{13}$$

and $\operatorname{supp} \phi^{\mathrm{r},i} \subset [2-2S, 0]$.

Note that the scaling functions that touch each end of the interval [0, 1] do not overlap: i.e. $2^{-L}(2S-1) \leq \frac{1}{2}$, as implied by the condition $2^L \geq 6S-6$.

The subspaces $V_j,\ j \geq L$, are defined as the spaces spanned by the union of the three sets of generators:

$$\begin{aligned} G_j^{\mathrm{l}} &= \{\phi^{\mathrm{l},i}(2^j x), \qquad 0 \leq i \leq R-1\}, \\ G_j^{\mathrm{int}} &= \{\phi_{jk}(x), \qquad 0 \leq k \leq 2^j - 2S + 1\}, \\ G_j^{\mathrm{r}} &= \{\phi^{\mathrm{r},i}(2^j(x-1)),\ 0 \leq i \leq R-1\}. \end{aligned}$$

Define V_j^{l}, V_j^{int} and V_j^{r} as the linear spans of G_j^{l}, G_j^{int} and G_j^{r} respectively. Generators in G_j^{l} are supported in

$$[0, 2^{-j}(2S-2)] \subset [0, \tfrac{1}{2}),$$

and those in G_j^{r} are supported in

$$[1 - 2^{-j}(2S-2), 1] \subset (\tfrac{1}{2}, 1].$$

With these definitions, we can verify from (12) and (13) that

$$P_i(2^j x) = \phi^{\mathrm{l},i}(2^j x) + 2^{-j/2} \sum_{k=0}^{2^j-2S+1} k^i \phi_{jk}(x) + \sum_{l=0}^{i} \binom{i}{l} 2^{j(i-l)} \phi^{\mathrm{r},l}(2^j(x-1)),$$

so that V_j contains all polynomials of degree at most $k-1$ restricted to [0, 1].

Proposition 1. *For $j \geq L$, $V_j \subset V_{j+1}$ and the generators $G_j = G_j^{\mathrm{l}} \cup G_j^{\mathrm{int}} \cup G_j^{\mathrm{r}}$ form a basis for V_j.*

Proof. Step 1. First we show that $G_j \subset \operatorname{span}(G_{j+1})$. The two-scale relation (10) implies that

$$\phi_{jk} = \sqrt{2} \sum_{n=0}^{2S-1} h_n \phi_{j+1,2k+n}.$$

The conditions $0 \leq k \leq 2^j - 2S + 1$ and $0 \leq n \leq 2S-1$ entail that $0 \leq 2k+n \leq 2^{j+1} - 2S + 1$, so that $\phi_{jk} \in \operatorname{span}(G_{j+1}^{\mathrm{int}})$.

Step 2. To see that $\phi^{l,i}(2^j x) \in V_{j+1}$, recall that the set $\{P_j(x),\ 0 \le j \le R-1\}$ spans polynomials of degree $R-1$ and that the monomials satisfy trivial scaling relations $x^j = 2^{-j}(2x)^j$. We therefore have

$$\begin{aligned}
\phi^{l,i}(x) &\in \overline{\text{span}}\{P_j(x), \quad \phi(x-k)\} \\
&= \overline{\text{span}}\{x^j, \quad \phi(x-k)\} \\
&\subset \overline{\text{span}}\{(2x)^j, \quad \phi(2x-k)\} \\
&= \overline{\text{span}}\{P_j(2x), \quad \phi(2x-k)\} \\
&= \overline{\text{span}}\{\phi^{l,j}(2x), \phi(2x-k)\},
\end{aligned}$$

where in each case the indices satisfy $0 \le j \le R-1$ and $k \ge 0$. Hence we have an expansion

$$\phi^{l,i}(x) = \sum_{j=0}^{R-1} a_j \phi^{l,j}(2x) + \sum_{k=0}^{\infty} b_k \phi(2x-k). \tag{14}$$

Now $\text{supp}\,\phi(2x-k) = 2^{-1}[k, k+2S-1] \subset [2S-2, \infty)$ whenever $k \ge 4S-4$. But all functions $\phi^{l,i}(x)$ and $\phi^{l,j}(2x)$ vanish on $\text{supp}\,\phi(2x-k)$ for such k, and so taking inner products with $\phi(2x-k)$ in (14) shows that $b_k = 0$ for $k \ge 4S-4$. Since $4S-4 < 2^L < 2^{j+1}-2S+1$ for $j \ge L$, we conclude that $\phi^{l,i}(2^j x) \in V_{j+1}$. A similar argument shows that $\phi^{r,i}(2^j x) \in V_{j+1}$. Hence $G_j \subset$ span (G_{j+1}), so $V_j \subset V_{j+1}$.

Step 3. The generator sets G_j^l, G_j^{int} and G_j^r are orthogonal, owing to the range of indices k used in the definitions (12) and (13) and because the support of generators $\phi_{jk} \in G_j^{int}$ lies within $[0, 1]$. Hence it remains to show that the $\phi^{l,i}$ are linearly independent. This is done as in [2] by remarking that $\{\phi^{l,i},\ 0 \le i \le R-1\}$ are obtained by applying a rank R transformation to the functions $\Phi = \{\phi(x-k),\ 1-2S \le k \le -1\}$, restricted to $[0, \infty)$. (The functions in Φ are linearly independent because of their 'embedded supports': we suppose that $\sum_{k=-1}^{1-2S} a_k \phi(x-k) = 0$ on $[0, \infty)$. Start with $k = -1$: on $[2S-3, 2S-2]$, all $\phi(x-k)$ in the sum vanish except for the $k = -1$ term, which itself does not vanish since $\text{supp}\,\phi(\cdot + 1) = [-1, 2S-2]$. Consequently $a_{-1} = 0$. Apply the same argument to $k = -2$ on the interval $[2S-4, 2S-3]$ and so forth, to show that each $a_k = 0$ in turn.)

From Proposition 1, it follows that V_j has dimension $2^j - 2D$ for $D = S-1-R$. Similarly, V_{j+1} has dimension $2^{j+1} - 2D$. We define W_j to be the orthogonal complement of V_j in V_{j+1}, and clearly it has dimension 2^j.

Proposition 2. *Assume that $2^L \ge 6S-6$. There exist $S-1$ linearly independent functions $\psi^{l,i}$, $0 \le i \le S-2$, on $[0, \infty)$ with support in $[0, 2S-2]$, and $S-1$ linearly independent functions $\psi^{r,i}$, $0 \le i \le S-2$, on $(-\infty, 0]$ with support in $[2-2S, 0]$ such that, for each $j \ge L$, W_j is spanned by the three sets of generators*

$$\begin{aligned}
H_j^l &= \{\psi^{l,i}(2^j x), \qquad 0 \le i \le S-2\}, \\
H_j^{int} &= \{\psi_{jk}(x), \qquad 0 \le k \le 2^j - 2S + 1\}, \\
H_j^r &= \{\psi^{r,i}(2^j(x-1)), 0 \le i \le S-2\}.
\end{aligned}$$

Proof. Step 1. The two-scale relation (11) implies that, if $0 \le k \le 2^j - 2S + 1$, then

$$\psi_{jk}(x) = \sqrt{2} \sum_{n=0}^{2S-1} g_n \phi_{j+1,2k+n}(x),$$

and so $\psi_{jk} \in V_{j+1}$. On the other hand, ψ_{jk} is orthogonal to each of the generator sets of V_j: for example, for G_j^{l} this follows because $\phi^{\mathrm{l},i}(2^j x)$ involves only ϕ_{jk} for $k < 0$ and $\operatorname{supp}\psi_{jk} \subset [0, 1]$. Consequently, $\psi_{jk} \perp V_j$, and so ψ_{jk} belongs to W_j.

Step 2. We now demonstrate the existence of the sets $\psi^{\mathrm{l},i}$ and $\psi^{\mathrm{r},i}$. Initially, we work with the case $j = L$: results for $j > L$ will then follow by scaling. Let $W_L^{\mathrm{int}} = \operatorname{span}\{\psi_{Lk},\ 0 \le k \le 2^L - 2S + 1\}$. Let $W_L^{\perp} = W_L \cap (W_L^{\mathrm{int}})^{\perp}$ denote the orthocomplement of W_L^{int} in W_L, where it is evident that $\dim W_L^{\perp} = 2S - 2$.

Our goal is to decompose $W_L^{\perp}$ into $W_L^{\mathrm{l}} \oplus W_L^{\mathrm{r}}$, with each space having dimension $S - 1$ and consisting of certain functions supported in $[0, 2^{-L}(2S-2)]$ and $[1 - 2^{-L}(2S-2), 1]$ respectively.

To understand the support of functions in $W_L^{\perp}$, we need to look more closely at the support of $\phi_{L+1,k} \in V_{L+1}^{\mathrm{int}}$. Thus, let $I_{L+1} = \{0, 1, \ldots, 2^{L+1} - 2S + 1\}$ denote the set of indices k for which $\phi_{L+1,k} \in G_{L+1}^{\mathrm{int}}$. Then decompose I_{L+1} as the disjoint union $I^{\mathrm{l}} \cup I^{\mathrm{o}} \cup I^{\mathrm{r}}$, where

$$\begin{aligned} I^{\mathrm{l}} &= \{k \in I_{L+1} : \sup\operatorname{supp}\phi_{L+1,k} \le \tfrac{1}{2}\} = \{0, \ldots, 2^L - 2S + 1\},\\ I^{\mathrm{o}} &= \{k \in I_{L+1} : \tfrac{1}{2} \in \operatorname{int}\operatorname{supp}\phi_{L+1,k}\} = \{2^L - 2S + 2, \ldots, 2^L - 1\},\\ I^{\mathrm{r}} &= \{k \in I_{L+1} : \inf\operatorname{supp}\phi_{L+1,k} \ge \tfrac{1}{2}\} = \{2^L, \ldots, 2^{L+1} - 2S + 1\}. \end{aligned}$$

Define orthogonal projections $f^{\mathrm{l}} = P^{\mathrm{l}} f$, $f^{\mathrm{m}} = P^{\mathrm{m}} f$ and $f^{\mathrm{r}} = P^{\mathrm{r}} f$ operating on $f \in V_{L+1}$ respectively by projection onto the spans of the three generator sets

$$\begin{aligned} &\{\phi_{L+1,k},\ k \in I^{\mathrm{l}}\} \cup \{\phi^{\mathrm{l},i}(2^{L+1}x)\},\\ &\{\phi_{L+1,k},\ k \in I^{\mathrm{o}}\},\\ &\{\phi_{L+1,k},\ k \in I^{\mathrm{r}}\} \cup \{\phi^{\mathrm{r},i}(2^{L+1}(x-1))\}, \end{aligned}$$

with $0 \le i \le R - 1$ as usual. Clearly $f = f^{\mathrm{l}} + f^{\mathrm{m}} + f^{\mathrm{r}}$, and P^{l}, P^{m} and P^{r} are mutually orthogonal. Also, by the definition of the index sets I^{l} and I^{r}, we have $\operatorname{supp} f^{\mathrm{l}} \subset (-\infty, \frac{1}{2}]$ and $\operatorname{supp} f^{\mathrm{r}} \subset [\frac{1}{2}, \infty)$.

Step 3. We show that $f^{\mathrm{m}} = 0$ whenever $f \perp V_L^{\mathrm{int}} \oplus W_L^{\mathrm{int}}$. Let A^{o} denote the interior of set A. First note that, if $2^L \ge 6S - 6$, then

$$(\operatorname{supp}\phi_{L+1,k})^{\mathrm{o}} \cap (\operatorname{supp}\phi_{Ll})^{\mathrm{o}} = \varnothing$$

for $k \in I^{\mathrm{o}}$ and $l \notin \{0, 1, \ldots, 2^L - 2S + 1\}$. Indeed, since $\operatorname{supp}\phi_{jk} = 2^{-j}[k, k + 2S - 1]$,

$$\bigcup_{k \in I^{\mathrm{o}}} \operatorname{supp}\phi_{L+1,k} = 2^{-L-1}[2^L - 2S + 2, 2^L + 2S - 2],$$

while

$$\bigcup_{l<0} \operatorname{supp}\phi_{Ll} = (-\infty, 2^{-L}(2S-2)]$$

and

$$\bigcup_{l \ge 2^L - 2S + 2} \operatorname{supp}\phi_{Ll} = [2^{-L}(2^L - 2S + 2), \infty).$$

For the interiors of these last two intervals to be disjoint from the first, it is necessary and sufficient that $2^L \ge 6S - 6$.

Since $\phi_{L+1,k} \in V_L(\mathbb{R}) \oplus W_L(\mathbb{R})$, we have $\phi_{L+1,k} = \sum_l c_l\phi_{Ll} + d_l\psi_{Ll}$ for some c_l, d_l, but the previous remark on disjoint supports shows that, for $k \in I^{\mathrm{o}}$, $\phi_{L+1,k}$ is orthogonal to those ϕ_{Ll} not in V_L^{int}. Since $\operatorname{supp}\psi_{Ll} = \operatorname{supp}\phi_{Ll}$, the same is true for the wavelets, and so $\phi_{L+1,k} \in V_L^{\mathrm{int}} \oplus W_L^{\mathrm{int}}$. It is now clear that $f \perp V_L^{\mathrm{int}} \oplus W_L^{\mathrm{int}}$ implies that $f^{\mathrm{m}} = 0$.

Step 4. We will now argue that P^{l} and P^{r} map $W_L^{\perp}$ into itself. First, f^{l} and f^{r} both lie in V_{L+1}, and

$$V_{L+1} = V_L^{\mathrm{l}} \oplus V_L^{\mathrm{int}} \oplus V_L^{\mathrm{r}} \oplus W_L^{\mathrm{int}} \oplus W_L^{\perp}.$$

So our task is to show that f^{l} and f^{r} are orthogonal to each of V_j^{int}, W_j^{int}, V_j^{l} and V_j^{r}.

We begin by showing orthogonality to V_j^{int} and W_j^{int}. We need yet more projections. Set

$$P_L^{\mathrm{l}} f = \sum_{k<0} \langle f, \phi_{Lk}\rangle\phi_{Lk} + \langle f, \psi_{Lk}\rangle\psi_{Lk}.$$

Note that $\operatorname{supp} P_L^{\mathrm{l}} f \subset (-\infty, 2^{-L}(2S-2)] \subset (-\infty, \frac{1}{2})$. Correspondingly, define P_L^{r} by taking the sum over $k \geq 2^L - 2S + 2$, and note similarly that $\operatorname{supp} P_L^{\mathrm{r}} f \subset [2^{-L}(2^L - 2S + 2), \infty) \subset (\frac{1}{2}, \infty)$.

Clearly, if $f \perp V_j^{\mathrm{int}} \oplus W_j^{\mathrm{int}}$, then $f = P_L^{\mathrm{l}} f + P_L^{\mathrm{r}} f$. Thus, for such f we have two equations,

$$f = f^{\mathrm{l}} + f^{\mathrm{r}} \quad \text{and} \quad f = P_L^{\mathrm{l}} f + P_L^{\mathrm{r}} f,$$

and hence

$$f^{\mathrm{l}} - P_L^{\mathrm{l}} f = P_L^{\mathrm{r}} f - f^{\mathrm{r}}.$$

The support of the left-hand side lies in $(-\infty, \frac{1}{2})$ while that of the right-hand side is contained in $(\frac{1}{2}, \infty)$. Hence, $f^{\mathrm{l}} = P_L^{\mathrm{l}} f \perp V_L^{\mathrm{int}} \oplus W_L^{\mathrm{int}}$. Similarly for f^{r}.

It remains, then, to show that f^{l} and f^{r} are orthogonal to both V_L^{l} and V_L^{r}. Suppose that $g \in V_L^{\mathrm{l}}$. Since $f \in W_L^{\perp}$, we have both $f = f^{\mathrm{l}} + f^{\mathrm{r}}$ and $\langle f, g\rangle = 0$. Since $\operatorname{supp} g \subset (-\infty, \frac{1}{2}]$ and $\operatorname{supp} f^{\mathrm{r}} \subset [\frac{1}{2}, \infty)$ have disjoint interiors, we also have $\langle f^{\mathrm{r}}, g\rangle = 0$ from which it follows that $\langle f^{\mathrm{l}}, g\rangle = 0$. Hence $f^{\mathrm{l}} \perp V_L^{\mathrm{l}}$. An analogous argument yields $f^{\mathrm{r}} \perp V_L^{\mathrm{r}}$.

Step 5. We may now define $W_L^{\mathrm{l}} = P^{\mathrm{l}} W_L^{\perp}$ and $W_L^{\mathrm{r}} = P^{\mathrm{r}} W_L^{\perp}$, with the assurance that

$$W_L^{\perp} = W_L^{\mathrm{l}} \oplus W_L^{\mathrm{r}}.$$

It remains to show that $\dim W_L^{\mathrm{l}}(\phi) = \dim W_L^{\mathrm{r}}(\phi) = S - 1$, where we now show the scaling function explicitly in the notation. We will do this by exhibiting an isormorphism of $W_L^{\mathrm{l}}(\phi)$ on $W_L^{\mathrm{r}}(\tilde{\phi})$, where $\tilde{\phi}(x) = \phi(2S - 1 - x)$ is the reflection of ϕ having the same support. Assuming this to be done, we note that $\dim W_L^{\mathrm{l}}(\phi) = \dim W_L^{\mathrm{l}}(\tilde{\phi})$ and so

$$2 \dim W_L^{\mathrm{r}}(\phi) = \dim W_L^{\mathrm{r}}(\tilde{\phi}) + \dim W_L^{\mathrm{r}}(\phi) = \dim W_L^{\mathrm{l}}(\phi) + \dim W_L^{\mathrm{r}}(\phi) = 2S - 2.$$

We construct the required isomorphism as follows. We first remark that, if $m_0(\xi) = \sum_{n=0}^{2S-1} h_n \mathrm{e}^{-\mathrm{i}n\xi}$, then the reflected coefficients $\tilde{h}_n = h_{2S-1-n}$ lead to $\tilde{m}_0(\xi) = \mathrm{e}^{\mathrm{i}(2S-1)\xi}\overline{m_0(\xi)}$. Consequently, the resulting scaling function is $\tilde{\phi}(x) = \phi(2S - 1 - x)$, and similarly for $\tilde{\psi}(x)$. The pair $(\tilde{\phi}, \tilde{\psi})$ has all the properties of (ϕ, ψ); in particular $\tilde{\psi}$ has R vanishing moments if and only if ψ does. Applying the definitions (13) and (12) to $\tilde{\phi}$ and ϕ, we find that

$$\begin{aligned}\tilde{\phi}^{\mathrm{r},i}(x) &= \sum_{k \geq 2-2S} k^i \phi(2S-1+k-x)\,\mathbf{1}_{[-\infty,0)}(x)\\ &= \sum_{k'<0} (1-2S-k')^i \phi(-x-k')\,\mathbf{1}_{(-\infty,0]}(x)\\ &= \sum_{0\leq l\leq i} T_{il}\phi^{\mathrm{l},i}(-x).\end{aligned}$$

The triangular matrix $T = (T_{il})$ is invertible, so this gives the required isomorphism.

Step 6. We may thus choose a set of functions $\{2^{L/2}\psi^{\mathrm{l},i}(2^L x),\ i = 0, \dots, S-2\}$ to be an orthonormal basis for W_L^{l} and similarly $2^{L/2}\psi^{\mathrm{r},i}(2^L(x-1))$ to be some orthobasis for W_L^{r}. We saw earlier for $f \in W_L^{\perp}$ that $P_L^{\mathrm{l}} f = P^{\mathrm{l}} f$ was supported in $(-\infty, 2^{-L}(2S-2)]$, and so this applies *a fortiori* to $\psi^{\mathrm{l},i}(2^L x)$. Similarly, $\operatorname{supp}\psi^{\mathrm{r},i}(2^L x) \subset [2^{-L}(2-2S), \infty)$.

For general $j > L$, we set $W_j^{\mathrm{l}} = \operatorname{span}\{\psi^{\mathrm{l},i}(2^j x)\}$ and $W_j^{\mathrm{r}} = \operatorname{span}\{\psi^{\mathrm{r},i}(2^j(x-1))\}$, and the decomposition

$$V_{j+1} = V_j \oplus W_j^{\mathrm{l}} \oplus W_j^{\mathrm{int}} \oplus W_j^{\mathrm{r}}$$

follows from our previous results by simple scaling. The proof of Proposition 2 is complete.

Proof of Theorem 1. It remains to combine Propositions 1 and 2 and to translate their conclusions to the notation used for (6) and (7).

Thus, the boundary scaling functions $\{\phi_k^{\mathrm{B}}\}$ may be taken as an orthonormalized version of the left group $\{\phi^{\mathrm{l},k}\}$, each set being indexed by $k = 0, \dots, R-1$. Similarly, $\{\phi_k^{\mathrm{B}}, k = -R, \dots, -1\}$ may be taken as an orthonormalized version of $\{\phi^{\mathrm{r},k},\ k = 0, \dots, R-1\}$.

The boundary wavelets ψ_k^{B} for $k = 0, \dots, S-2$ can be identified with $\psi^{\mathrm{l},k}$, while, for $k = -S+1, \dots, -1$, we may take $\psi^{\mathrm{r},S-1+k}$.

Finally, since $V_j(\mathbb{R}) \to L_2(\mathbb{R})$ as $j \nearrow \infty$, it is easy to show that V_j^{int}, and *a fortiori* $V_j \to L_2[0,1]$. Hence (8) is an orthonormal basis for $L_2[0,1]$.

3. Preconditioning and quadrature errors

Suppose now that we are given a vector $\boldsymbol{f} = (f(l/N))$ of values of a function. Assume that the scaling function ϕ has integral 1 and $R-1$ vanishing moments for $R \geq 1$. If f is a polynomial p of degree at most $R-1$, then for interior scaling functions ϕ_{Jl}, the property (5) states that the point values $p(l/N)$ *equal* the scaling coefficients $\langle p, \phi_{Jl}\rangle$. This property fails in general for the boundary scaling functions in (6): to achieve something similar, we need to describe (again extending [1]) preconditioning operations at the left and right edges. After doing so, we can establish a version of (5) that applies for smooth f to *all* scaling coefficients.

Consider the left boundary first. Define the $R \times R$ matrix T by

$$T_{kl} = \int_0^{\infty} x^l \phi_k^{\mathrm{B}}(x)\,\mathrm{d}x, \qquad k, l = 0, 1, \dots, R-1,$$

and the $(S-1) \times R$ matrix U by

$$U_{jl} = j^l, \qquad 0 \leq j < S-1,\ 0 \leq l < R.$$

Then the matrix T is of rank R as a consequence of the boundary wavelet construction. To see this, assume that $Ta = 0$ for some $a = (a_l) \neq 0$. Then the polynomial $p(x) = \sum a_l x^l$ has degree at most $R - 1$ and is orthogonal on $[0, \infty)$ to $\phi^{1,i}$ for $i = 0, \ldots, R - 1$. But from the construction (12),

$$p(x) = \sum_i b_i \phi^{1,i}(x) + \sum_k c_k \phi(x - k).$$

If the second sum is truncated to the range $0 \le k < 2S$, then equality still holds for $x \in [0, 2S]$. But the coefficients b_i vanish on our linear independence assumption. Hence $p(x) = \sum_{k=0}^{2S-1} c_k \phi(x - k)$ on $[0, 2S]$. In particular, owing to the embedded supports of $\phi(\cdot - k)$, on $[0, 1]$ we must have $p(x) = c_0 \phi(x)$. However, there does not exist a compactly supported polynomial spline (with knots at the integers) whose integer translates form an orthonormal sequence [12]. Hence $c_0 = 0$ and so $p \equiv 0$ and $a = 0$.

We now define the left preconditioning transform A^{L} to be any $R \times (S - 1)$ matrix such that $A^{\mathrm{L}} U = T$, i.e.

$$\sum_i A^{\mathrm{L}}_{ki} i^l = T_{kl} = \int_0^\infty x^l \phi^{\mathrm{B}}_k(x)\, \mathrm{d}x. \tag{15}$$

Each subset of R rows of U is linearly independent. Because T is of rank R, so will A^{L} be. Similarly, the matrix A^{R} is constructed to satisfy $A^{\mathrm{R}} \bar{U} = \bar{T}$, where

$$\begin{aligned} \bar{T}_{kl} &= \int_{-\infty}^0 x^l \phi^{\mathrm{B}}_{-k}(x)\, \mathrm{d}x, \qquad && k = 1, 2, \ldots, R,\ l = 0, 1, \ldots, R - 1, \\ \bar{U}_{jl} &= (-1)^l j^l, && j = 1, 2, \ldots, S,\ l = 0, 1, \ldots, R - 1. \end{aligned}$$

We explain the utility of the preconditioning transformations as follows. First, recall that the reason for using coiflets is the vanishing moment property: if P is a polynomial of degree at most $R - 1$, then for the interior scaling functions $\phi(x - k)$,

$$P(k) = \langle P, \phi_{0,k} \rangle. \tag{16}$$

This property is lost for the boundary scaling functions such as $\phi^{\mathrm{B}}_{0,k}$. However, the preconditioning yields analogues: for $k = 0, \ldots, R - 1$, $\sum_i A^{\mathrm{L}}_{ki} P(i) = \langle P, \phi^{\mathrm{B}}_k \rangle$ and similarly at the right edge.

3.1. Preconditioning function values

We now define evaluation operators S_{Jk} corresponding to the application of the preconditioning transform to a discrete sequence of values of a function. Given any function g on $[0, 1]$, define

$$S_{Jk} g = \begin{cases} \displaystyle\sum_{i=0}^{S-2} A^{\mathrm{L}}_{ki} g(i 2^{-J}), & 0 \le k \le R - 1, \\ g(k 2^{-J}), & S - 1 \le k \le 2^J - S, \\ \displaystyle\sum_{i=1}^{S-1} A^{\mathrm{R}}_{2^J - k, i} g(1 - i 2^{-J}), & 2^J - R \le k \le 2^J - 1. \end{cases}$$

Write $S_J g$ for the vector of values $S_{Jk} g$ for fixed J.

Given a function g, under suitable conditions, $2^{-J/2} S_J g$ gives a good approximation to the vector of scaling coefficients of g at level J. We state and prove a proposition that bounds the error of this approximation. Define $\Delta_J g = (\Delta_{Jk}(g))$ to be the vector of coefficients

$$\Delta_{Jk}(g) = 2^{-J/2} S_{Jk} g - \langle g, \phi_{Jk} \rangle,$$

where, for any two functions f and g, we write $\langle f, g\rangle = \int_0^1 f(t)g(t)\,\mathrm{d}t$. For any interval A, define $\|g\|_{\infty,A} = \sup\{|g(x)|,\ x \in A\}$.

Proposition 3. *Suppose that the function g has $R \geq 1$ continuous derivatives $g^{(1)}, \ldots, g^{(R)}$ on $[0, 1]$. Define $S(\phi_{Jk})$ to be the support of ϕ_{Jk} if $k \in \mathcal{K}_J^{\mathrm{I}}$, the interval $[0, (2S-2)2^{-J}]$ if ϕ_{Jk} is a boundary scaling function at the left of the interval $[0, 1]$ and $[1-(2S-2)2^{-J}, 1]$ if ϕ_{Jk} is a boundary scaling function at the right of the interval.*

With the above definitions of boundary scaling functions and preconditioning operators,

$$|\Delta_{Jk}(g)| \leq c2^{-J(R+1/2)}\|g^{(R)}\|_{\infty,S(\phi_{Jk})}.$$

Proof. We approximate g by a Taylor polynomial p of degree $R-1$. The Taylor expansion is carried out around the point x, where $x = k2^{-J}$ if ϕ_{Jk} is an interior scaling function, $x = 0$ if ϕ_{Jk} is a boundary scaling function on the left of the interval, and $x = 1$ if ϕ_{Jk} is a boundary scaling function on the right of the interval.

We first show that $\Delta_{Jk}(p) = 0$. If ϕ_{Jk} is an interior scaling function, then by the vanishing moment properties of ϕ, by expressing $p(t)$ as a polynomial in $t - x$ we have $2^{J/2}\langle p, \phi_{Jk}\rangle = p(x) = S_{Jk}p$. If ϕ_{jk} is a left boundary scaling function, write

$$p(t) = \sum_{l=0}^{R-1} p_l t^l.$$

Then

$$S_{Jk}p = \sum_{l=0}^{R-1} p_l 2^{-lJ} \sum_{i=0}^{S-2} A^{\mathrm{L}}_{ki} i^l.$$

We can make use of (15) and the identity $2^{-lJ}W_{kl} = 2^{J/2}\int y^l\phi_{Jk}(y)\,\mathrm{d}y$ to again conclude that $S_{Jk}p = 2^{J/2}\langle p, \phi_{Jk}\rangle$. Via a similar argument, if ϕ_{jk} is a right boundary scaling function, write $p(t) = \sum_{l=0}^{R-1} p_l(t-1)^l$ and repeat analogous steps using $A^{\mathrm{R}}\bar{U} = \bar{W}$ to arrive again at $\Delta_{Jk}(p) = 0$.

Write $g = p + r$. For each k, the interval $S(\phi_{Jk})$ is of length less than $2S2^{-J}$ and contains both the support of ϕ_{Jk} and the range of function values used by the evaluation operator S_{Jk}. Write $g = p + r$ and use the property $\Delta_{Jk}(p) = 0$ to obtain that

$$\begin{aligned}
|\Delta_{Jk}(g)| &= |\Delta_{Jk}(r)| \\
&\leq |2^{-J/2}S_{Jk}r| + |\langle r, \phi_{Jk}\rangle| \\
&\leq c2^{-J/2}\|r\|_{\infty,S(\phi_{Jk})} \\
&\leq c2^{-J(R+1/2)}\|g^{(R)}\|_{\infty,S(\phi_{Jk})},
\end{aligned}$$

applying Taylor's theorem on the interval $S(\phi_{Jk})$ to bound r in terms of $\|g^{(R)}\|_{\infty,S(\phi_{Jk})}$. This completes the proof.

Here is a variant of the preceding result for functions having bounded total variation, where, as usual

$$\|g\|_{\mathrm{TV}} = \sup \sum_{i=1}^{m-1} |g(x_{i+1}) - g(x_i)|, \tag{17}$$

with supremum taken over all finite sequences $x_1 < x_2 < \cdots < x_m$ in $[0, 1]$.

Proposition 4. *Assume that ϕ has $\int \phi = 1$ and $R - 1$ vanishing moments, $R \geq 1$. With the above definitions of boundary scaling functions and preconditioning operators,*

$$\|\Delta_J g\|_1 \leq c 2^{-J/2} \|g\|_{\mathrm{TV}}.$$

Proof. Write $\mathcal{I}$ for the interior range $S - 1, \ldots, 2^J - S$. For $k \in \mathcal{I}$, let $t_k = k2^{-J}$ and observe that, since $\int \phi = 1$,

$$-\Delta_{Jk}(g) = 2^{-J/2} \int [g(t_k + u2^{-J}) - g(t_k)]\phi(u)\,\mathrm{d}u.$$

From the definition (17),

$$\sum_{k \in \mathcal{I}} |\Delta_{Jk}(g)| \leq 2^{-J/2} \cdot 2S \cdot \|g\|_{\mathrm{TV}} \|\phi\|_1 \tag{18}$$

since the increments in the previous integral can be divided into at most $2S$ nonoverlapping collections. It remains to consider the boundary contributions. Define $\breve{g}(x) = g(x) - g(0)$. For $k \in \{0, \ldots, R - 1\}$,

$$\Delta_{Jk}(g) = 2^{-J/2} \left[S_{Jk} g - \int_0^\infty g(2^{-J}u)\phi_k^{\mathrm{B}}(u)\,\mathrm{d}u \right] = \Delta_{Jk}(\breve{g}).$$

Since (15) entails that $S_{Jk}(1) = \sum_i A^{\mathrm{L}}_{ki} = \int_0^\infty \phi_k^{\mathrm{B}}$ and $\|\breve{g}\|_\infty \leq \|g\|_{\mathrm{TV}}$, we conclude that

$$\sum_{k=0}^{R-1} |\Delta_{Jk}(g)| \leq c 2^{-J/2} \|g\|_{\mathrm{TV}}. \tag{19}$$

A similar bound works for $k \in \{2^J - R, \ldots, 2^J - 1\}$, so in combination with (18) and (19), we arrive at the result.

3.2. Preconditioning data

Given a sequence $Y_0, Y_1, \ldots, Y_{N-1}$ with $N = 2^J$, we define the preconditioned sequence $P_J Y$ by

$$(P_J Y)_k = \begin{cases} \sum_{i=0}^{S-2} A^{\mathrm{L}}_{ki} Y_i, & 0 \leq k \leq R - 1, \\ Y_k, & S - 1 \leq k \leq N - S, \\ \sum_{i=1}^{S-1} A^{\mathrm{R}}_{N-k,i} Y_{N-i}, & N - R \leq k \leq N - 1. \end{cases}$$

If the original sequence is uncorrelated with variance 1, then the variance matrix of the first part of the preconditioned sequence is $A^{\mathrm{L}}(A^{\mathrm{L}})^\top$ while that of the last part of the sequence with indices taken in reverse order is $A^{\mathrm{R}}(A^{\mathrm{R}})^\top$. Because A^{L} and A^{R} are both of full rank R, these variance matrices are both strictly positive definite.

Given the choice of A^{L} and A^{R}, let c_A be the maximum of the eigenvalues of $A^{\mathrm{L}}(A^{\mathrm{L}})^\top$ and $A^{\mathrm{R}}(A^{\mathrm{R}})^\top$. Suppose that the Y_i were independent normal random variables with unit variance, as occurs for example in the regression model (3).

Let $\tilde{y}$ be the boundary-corrected discrete wavelet transform of the sequence $P_J Y$. Thus $\tilde{y} = \mathcal{W} P_J Y$ where $\mathcal{W}$ is the discrete wavelet transform matrix (see Remark 1). Clearly $\mathrm{E}\,\tilde{y} = \mathcal{W} P_J \boldsymbol{f} = \mathcal{W} S_J f$, where $\boldsymbol{f} = (f(i/N))$ is the vector of sampled values of f.

Since $\mathcal{W}$ is orthogonal, both $P_J Y$ and $\tilde{y}$ have multivariate normal distributions whose variance matrices have eigenvalues bounded by c_A. In particular, the variance of all elements of $\tilde{y}$ is bounded by c_A.

However 'most' coefficients have unit variance. Let I denote the collection of indices λ corresponding to interior wavelets: $I = \{(jk) : L \le j < J,\ S-1 \le k \le 2^j - S\}$. Then the array $\tilde{Y}^I = \{\tilde{y}_\lambda : \lambda \in I\}$ of interior coefficients will depend only on Y_i for $S-1 \le i \le N-S$, in other words, those Y_i left unchanged by the preconditioning. To see this, let ψ_λ be an interior wavelet. We have

$$\tilde{y}_\lambda = \sum_l \langle \psi_\lambda, \phi_{Jl} \rangle (P_J Y)_l.$$

It follows by iterating the two-scale relation (11) that ψ_λ is a linear combination of those scaling functions ϕ_{Jl} whose support is contained within supp ψ_λ. In particular, $\langle \psi_\lambda, \phi_{Jl} \rangle = 0$ for the boundary indices l with $0 \le l \le R-1$ and $N-R \le l \le N-1$.

As a result, the interior coefficients $\tilde{Y}^I$ are independent Gaussian with variance 1.

4. Discrete versus continuous wavelet coefficients

The previous section studied the quality of approximation of the sampled function $N^{-1/2} S_J f$ by the fine-scale scaling coefficients $\langle f, \phi_{Jk} \rangle$. Now define $\tilde{\theta}$ to be the boundary-corrected discrete wavelet transform of $N^{-1/2} S_J f$. Let $\theta = (\theta_\lambda = \langle f, \psi_\lambda \rangle)$ be the coefficients of f in the wavelet basis $\mathcal{B} = \{\psi_\lambda\}$.

We can now turn to one of our key questions, namely bounding the difference between these two wavelet arrays. Some regularity of f (i.e. θ) is needed: we make assumptions of two types. First, we consider norm balls in Besov spaces, denoted $\Theta(C) = \Theta^\alpha_{p,\infty}(C)$, and defined as the set of functions $f \in L_2[0,1]$ whose coefficients in basis $\mathcal{B}$ satisfy

$$\|\theta_j\|_p \le C 2^{-aj} \quad \text{for all } j \ge L-1. \tag{20}$$

Here $a = \alpha + \frac{1}{2} - 1/p$ and θ_j denotes the vector $(\theta_{jk},\ 0 \le k < 2^j)$. We assume here that $\alpha > 1/p$, which ensures that the point evaluation functionals $f \to f(t_0)$ are continuous, so that the sampling model (3) makes sense. As usual, $\|\theta_j\|_p = (\sum_k |\theta_{jk}|^p)^{1/p}$.

The second smoothness model will be to assume that $f \in \mathrm{TV}(C)$, the set of functions f having total variation norm (17) bounded by C. It is well known that

$$\Theta^1_{1,1} \subset \mathrm{TV} \subset \Theta^1_{1,\infty},$$

in the sense of embedding of linear spaces, so that our smoothness assumption corresponds to $\alpha = 1 = p$. (In this case, we can make sense of point evaluation by agreeing to use, say, the left-continuous versions of $f \in \mathrm{TV}$.)

Proposition 5. *Assume that the scaling function ϕ and the mother wavelet ψ have R continuous derivatives and support $[-S+1, S]$ for some integer S, and that $\int x^m \phi(x)\,dx = \int x^m \psi(x)\,dx = 0$ for $m = 1, 2, \ldots, R-1$. Assume that the wavelets and scaling functions are modified by the boundary construction described above. Assume that either* (i) $\theta \in \Theta^\alpha_{p,\infty}(C)$ *with $1/p < \alpha < R$ or* (ii) $\theta \in \mathrm{TV}(C)$*, in which case set $\alpha = p = 1$. Then, for each j with $L-1 \le j < J$,*

$$2^{aj} \|\theta_j - \tilde{\theta}_j\|_p \le c C 2^{-\bar{\alpha}(J-j)},$$

where $\bar{\alpha} = \alpha - (1/p - 1)_+ > 0$ and $c = c(\alpha, p, \phi, \psi)$.

Proof. Step 0. We first recall a lemma for matrix norms (see e.g. [11, Theorem 4.1.2] for a more general statement). Let $T : \mathbb{R}^L \to \mathbb{R}^K$ be a linear transformation satisfying

$$\sum_k |T_{kl}| \le M_1 \quad \text{and} \quad \sum_l |T_{kl}| \le M_2$$

for each k, l. Then, for $p \ge 1$,

$$\|T\theta\|_p \le M_1^{1/p} M_2^{1-1/p} \|\theta\|_p$$

for all $\theta \in \mathbb{R}^L$. Of course, if $M_1 = M_2$, then $\|T\|_p \le M_1$. For $0 < p \le 1$, if instead $\sum_k |T_{kl}|^p \le M_p^p$ for each l, then $\|T\theta\|_p \le M_p \|\theta\|_p$. For convenience, we refer to these bounds as 'Young's inequality'.

Step 1. Now let W be that part of the matrix $\mathcal{W}$ that maps the sequence at level J to the jth level of its discrete wavelet transform. Thus $W_{kl} = \langle \psi_{jk}, \phi_{Jl} \rangle$ if $j \ge L$. For $j = L - 1$, we have $W_{kl} = \langle \phi_{Lk}, \phi_{Jl} \rangle$. We may write

$$\tilde{\theta}_j - \theta_j = W \Delta_J f.$$

Assume that $N = 2^J$. We first bound $\|\Delta_J f\|$ by decomposing $f = \sum_{j \ge L-1} e_j$, where $e_j = \sum_l \theta_{jl} \psi_{jl}$ for $j \ge L$ and $e_{L-1} = \sum_l \theta_{L-1,l} \phi_{Ll}$. We have

$$\Delta_{Jk} e_j = \sum_l T_{kl} \theta_{jl}, \tag{21}$$

with

$$T_{kl} = \Delta_{Jk} \psi_{jl} = 2^{-J/2} S_{Jk} \psi_{jl} - \langle \phi_{Jk}, \psi_{jl} \rangle.$$

Step 2. First consider $j \ge J$. The orthogonality properties imply that the inner product $\langle \phi_{Jk}, \psi_{jl} \rangle$ always vanishes, so in this case $T_{kl} = 2^{-J/2} S_{Jk}(\psi_{jl})$. For this range of j, the key property is the compact support of ψ_{jl}: even taking account of the possibility that ψ_{jl} may be a boundary wavelet,

$$\psi_{jl}(k2^{-J}) = 0 \quad \text{if } |l - 2^{(j-J)}k| \ge 2S.$$

This means that 'most' coefficients T_{kl} vanish. More precisely, suppose that k is a boundary coefficient at level J. Define

$$\mathcal{L}(j) = \{l : |l - 2^{(j-J)}k'| < 2S \text{ for some } k' \text{ in } [0, S-2] \text{ or } [2^J - S + 1, 2^J - 1].\}$$

Then $S_{Jk}\psi_{jl}$ is zero if $l \notin \mathcal{L}(j)$.

In addition, from the definition of ψ_{jl} we have

$$|S_{Jk}\psi_{jl}| \le c2^{j/2} \quad \text{for all } j \text{ and } l.$$

We conclude that $|T_{kl}| \le c2^{(j-J)/2} I_{kl}$, where

$$I_{kl} = \begin{cases} 1 & \text{if } k \in \mathcal{K}_J^{\mathrm{I}},\ |l - 2^{j-J}k| < 2S, \\ 1 & \text{if } k \in \mathcal{K}_J^{\mathrm{B}},\ l \in \mathcal{L}(j), \\ 0 & \text{otherwise.} \end{cases}$$

Thus, both $\sum_k I_{kl}$ and $\sum_l I_{kl}$ are bounded by $8S^2$ and so with $c = c(p, S)$ we may apply Young's inequality with $M_p = c(p, S)2^{(j-J)/2}$ to conclude that

$$\|\Delta_J e_j\|_p \leq c2^{(j-J)/2}\|\theta_j\|_p, \qquad j \geq J.$$

Step 3. For $j \in [L, J-1]$, using the smoothness of the wavelets ψ and the vanishing moments of ϕ, Proposition 3 yields that

$$|T_{kl}| \leq c2^{-(R+1/2)J} \sup |\psi_{jk}^{(R)}| \leq c2^{-(R+1/2)(J-j)}.$$

When $|l - 2^{-(J-j)}k| \geq 4S$, the bounded support properties of the wavelet and scaling function will ensure that the inner product $\langle \phi_{Jk}, \psi_{jl}\rangle$ and the evaluation $S_{Jk}\psi_{jl}$ are both zero, and hence $\Delta_{Jk}\psi_{jl} = 0$. The number of indices k for which $|l - 2^{-(J-j)}k| < 4S$ is bounded by $8S \cdot 2^{J-j}$. It therefore follows that

$$\sum_k |T_{kl}|^p \leq c2^{(J-j)}2^{-(R+1/2)(J-j)p},$$

while $\sum_l |T_{kl}|$ is no larger than this (with $p = 1$). Consequently, Young's inequality yields

$$\|\Delta_J e_j\|_p \leq c2^{-(R+1/2-1/p)(J-j)}\|\theta_j\|_p, \qquad L \leq j < J.$$

Step 4. Under the Besov smoothness assumptions, since $\|\theta_j\|_p \leq C2^{-aj}$ for all j, we may summarize the first two steps with the bounds

$$\|\Delta_J e_j\|_p \leq \begin{cases} cC2^{-aJ}2^{-(a-1/2)(j-J)}, & j \geq J, \\ cC2^{-aJ}2^{-(R-\alpha)(J-j)}, & L-1 \leq j < J, \end{cases}$$

which decay geometrically in j if $\alpha < 1/p < R$, and so

$$\|\Delta_J f\|_p \leq c_p C2^{-aJ}. \tag{22}$$

Under the total variation assumption, $a = \frac{1}{2}$, and (22) is precisely the conclusion of Proposition 4.

Step 5. We turn now to bounding the operator p-norm of the matrix W. First note that, uniformly in k and l,

$$|W_{kl}| = |\langle \phi_{Jk}, \psi_{jl}\rangle| \leq \|\psi_{jk}\|_\infty \|\phi_{Jl}\|_1 \leq c2^{-(J-j)/2}.$$

For each l, the number of values k for which W_{kl} is nonzero is uniformly bounded, so

$$\sum_k |W_{kl}|^p \leq c \max_k |W_{kl}|^p \leq c2^{-(J-j)p/2}.$$

In particular, when $0 < p \leq 1$, Young's inequality yields $\|W\|_p \leq c2^{-(J-j)/2}$.

As seen earlier, for each k, the number of values l for which W_{kl} is nonzero is bounded by $c2^{(J-j)}$, so

$$\sum_l |W_{kl}| \leq c2^{(J-j)} \max_l |W_{kl}| \leq c2^{(J-j)/2}.$$

Applying Young's inequality with $M_1 = c2^{(j-J)/2}$ and $M_2 = c2^{(J-j)/2}$, we obtain, for $p \geq 1$,

$$\|W\|_p \leq c2^{(1/2-1/p)(J-j)}.$$

Everything in the above argument works if we set $j = L - 1$ and replace ψ_{jk} by ϕ_{Lk}, so the preceding bound also holds for $j = L - 1$.

The vector $\tilde{\theta}_j - \theta_j$ is given by $W \Delta_J f$, so we can now complete the proof of Proposition 5 by combining the bounds on W and Δ_J:

$$2^{aj} \|\tilde{\theta}_j - \theta_j\|_p \le \|W\|_p \cdot 2^{aj} \|\Delta_J f\|_p.$$

From (22), we have $2^{aj} \|\Delta_J f\|_p \le cC2^{-a(J-j)}$ while $\|W\|_p \le c2^{-b(J-j)}$ with $b = \min(\frac{1}{2}, 1/p - \frac{1}{2})$. Since $a + b = \alpha - (1/p - 1)_+$, the proof is complete.

5. Estimation error

Finally, we return to the question of comparing estimation error from the continuous wavelet model of (1) and (2) with that from the sampled discrete data model of (3) and (4). In this section we make a comparison based on minimax risks. For a comparison using a specific estimator, see for example [9].

In the continuous data model (2), the minimax risk over parameter space Θ is

$$R_{\mathcal{E}}(\Theta, \varepsilon) = \inf_{\hat{\theta} \in \mathcal{E}} \sup_{\theta \in \Theta} \mathrm{E}_\theta \|\hat{\theta}(y) - \theta\|_2^2.$$

Here $\mathcal{E}$ stands for a class of estimators: if $\mathcal{E} = N$, then all estimators are allowed, while $\mathcal{E} = M$ restricts to coordinatewise (or 'marginal') estimators: the λth coordinate depends only on y_λ: namely $\hat{\theta}_\lambda(y) = \delta_\lambda(y_\lambda)$. When $\Theta = \Theta^\alpha_{p,\infty}(C)$, we write $M(C, \varepsilon)$ to abbreviate $R_M(\Theta^\alpha_{p,\infty}(C), \varepsilon)$.

In the discrete data model of (3) and (4), with $\varepsilon_N = \sigma/\sqrt{N}$,

$$\tilde{R}_{\mathcal{E}}(\Theta, \varepsilon_N) = \inf_{\hat{\theta} \in \mathcal{E}} \sup_{\theta \in \Theta} \mathrm{E}_\theta \|\hat{\theta}(\tilde{y}) - \theta\|_2^2.$$

Note that we still intend to estimate f and therefore θ, even though $\mathrm{E}\,\tilde{y}_\lambda = \tilde{\theta}_\lambda$, according to (4).

Suppose that $\mathcal{B} = \{\psi_\lambda\}$ is a boundary modified coiflet orthonormal basis for $L_2[0, 1]$ as described in Section 2.

Theorem 2. *Under the assumptions of Proposition 5, as $N \to \infty$,*

$$\tilde{R}_N(\Theta(C), \varepsilon_N) \le R_N(\Theta(C), \varepsilon_N)(1 + o(1)).$$

The result says that, over a wide range of smoothness classes, indexed by smoothness α, homogeneity p and radius C, the sampled data problem is not essentially harder than the continuous data problem. (The converse result, stating that the sampled data problem is not asymptotically easier, is demonstrated in [6].)

Our strategy will be to take a near-optimal (coordinatewise) estimator $\hat{\theta}(y)$ from the continuous model and apply it to suitable coordinates $\tilde{y}_\lambda$ in the sampled data model. In fact, we restrict attention to levels $j \le J_0$ where $J_0 = J_0(\alpha, p) < J$ will be specified below. We will need notation for the restriction of norms to these levels: $\|\theta\|^2_{2,J} = \sum_{j \le J} \|\theta_j\|_2^2$, while $\|\theta\|^2_{2,J^\perp}$ denotes the corresponding sum over $j > J$.

Since $\mathrm{E}\,\tilde{y}_\lambda = \tilde{\theta}_\lambda$, it is natural to decompose the error of estimation of θ in terms of $\tilde{\theta}$, noting again that $\hat{\theta}_\lambda = 0$ for $j > J_0$:

$$\|\hat{\theta}(\tilde{y}) - \theta\|_2 \le \|\hat{\theta}(\tilde{y}) - \tilde{\theta}\|_{2,J_0} + \|\tilde{\theta} - \theta\|_{2,J_0} + \|\theta\|_{2,J_0^\perp}. \tag{23}$$

We concentrate on the first term on the right-hand side: the other two will later be shown to be negligible.

For interior coordinates, $\lambda \in \Lambda^{\mathrm{I}}_{J_0}$, we use a coordinatewise estimator to be described below. For boundary coordinates $\lambda \in \Lambda^{\mathrm{B}}_{J_0}$ say, simply use the unbiased estimator: $\hat{\theta}_\lambda(\tilde{y}_\lambda) = \tilde{y}_\lambda$. Then

$$\mathrm{E}\,\|\hat{\theta}(\tilde{y}) - \tilde{\theta}\|^2_{2,J_0} = \sum_{\lambda \in \Lambda^{\mathrm{I}}_{J_0}} \mathrm{E}[\delta_\lambda(\tilde{y}_\lambda) - \tilde{\theta}_\lambda]^2 + \sum_{\lambda \in \Lambda^{\mathrm{B}}_{J_0}} \mathrm{E}[\tilde{y}_\lambda - \tilde{\theta}_\lambda]^2. \tag{24}$$

From the discussion of Section 3.2, for $\lambda \in \Lambda^{\mathrm{B}}_{J_0}$, we have $\mathrm{var}(\tilde{y}_\lambda) \le c_A \varepsilon_N^2$ and so the second term on the right-hand side of (24) is bounded by $|\Lambda^{\mathrm{B}}_{J_0}| c_A \varepsilon_N^2$. Since $|\Lambda^{\mathrm{B}}_{J_0}| \le 4S \log \varepsilon_N^{-1}$, we conclude that it is negligible, being $O((\log \varepsilon_N^{-1})\varepsilon_N^2)$.

Suppose that $J_0 = (1-\gamma)J$ for some $\gamma > 0$ to be specified later. The key role of Proposition 5 is to show that, for levels below J_0, the Besov norm of $\tilde{\theta}$ is not greatly inflated relative to θ. Thus, for $j \le J_0$ and $p \ge 1$,

$$2^{aj}\|\tilde{\theta}_j\|_p \le 2^{aj}\|\theta_j\|_p + 2^{aj}\|\tilde{\theta}_j - \theta_j\|_p \le C[1 + c2^{-\alpha\gamma J}] = C[1+\delta_N] = C_N,$$

say. Thus, if $\theta \in \Theta(C)$, then $\Pi_{J_0}\tilde{\theta} \in \Theta(C_N)$, where Π_J denotes projection onto coordinates with $j \le J_0$.

Suppose that

$$\hat{\theta}(y) = (\delta_\lambda(y_\lambda))$$

is a coordinatewise estimator attaining the minimax risk $M(C_N, \varepsilon_N)$ in the continuous data problem.

For the *interior* coordinates $\lambda \in \Lambda^{\mathrm{I}}_{J_0}$, the discrete data wavelet coefficients $\tilde{y}_\lambda$ have distribution $N(\tilde{\theta}_\lambda, \varepsilon_N^2)$ and can be regarded as a submodel of the sequence model (2), but with $C_N = C[1+\delta_N]$. Hence, since $\hat{\theta}$ operates coordinatewise,

$$\sup_{\theta \in \Theta(C)} \sum_{\lambda \in \Lambda^{\mathrm{I}}_{J_0}} \mathrm{E}[\delta_\lambda(\tilde{y}_\lambda) - \tilde{\theta}_\lambda]^2 \le \sup_{\theta \in \Theta(C')} \sum_{\lambda} \mathrm{E}[\delta_\lambda(y_\lambda) - \theta_\lambda]^2 = M(C_N, \varepsilon_N).$$

According to [6, Lemma 2.3], the coordinatewise minimax risk satisfies the scaling relation

$$M(C_N, \varepsilon_N) \le (1+\delta_N)^2 M(C, \varepsilon_N),$$

and according to [5, Theorems 3 and 5], for Besov spaces it is asymptotically equivalent to the unrestricted minimax risk, so that there exist $\eta_N \to 0$ such that

$$M(C, \varepsilon_N) \le (1+\eta_N) R_N(\Theta(C), \varepsilon_N).$$

Combining (24) and its successors, we find that

$$\mathrm{E}\,\|\hat{\theta}(\tilde{y}) - \tilde{\theta}\|^2_{2,J_0} \le (1+\eta_N)(1+\delta_N)^2 R_N(\Theta(C), \varepsilon_N) + c(\log \varepsilon_N^{-1})\varepsilon_N^2. \tag{25}$$

We turn to the remaining terms in (23). Recall first that

$$R_N = R_N(\Theta(C), \varepsilon_N) \asymp C^{2/(2\alpha+1)} \varepsilon_N^{4\alpha/(2\alpha+1)}. \tag{26}$$

We have $J_0 = (1-\gamma)J = c(1-\gamma)\log \varepsilon_N^{-1}$. Hence, applying again Proposition 5 but now with $p = 2$,

$$\|\tilde{\theta} - \theta\|^2_{2,J_0} = \sum_{j \le J_0} \|\tilde{\theta}_j - \theta_j\|_2^2 \le c^2 C^2 J_0 2^{-2\alpha J} \asymp C^2 (\log \varepsilon_N^{-1}) \varepsilon_N^{4\alpha} = o(R_N). \tag{27}$$

When $\theta \in \Theta(C)$,

$$\begin{aligned}\|\theta\|^2_{2,J_0^\perp} &= \sum_{j>J_0} \|\theta_j\|_2^2 \\ &\le \sum_{j>J_0} 2^{2j(1/2-1/p)_+} \|\theta_j\|_p^2 \\ &\le C^2 \sum_{j>J_0} 2^{-2j[a-(1/2-1/p)_+]} \\ &\le cC^2 2^{-2\alpha' J_0} \\ &\asymp C^2 \varepsilon_N^{4\alpha'(1-\gamma)}, \end{aligned} \tag{28}$$

where $\alpha' = \alpha$ for $p \ge 2$ and $\alpha' = \alpha - 1/p + 1/2$ for $p < 2$. Comparing (28) with (26), clearly $\alpha' > \alpha/(2\alpha + 1)$ either when $p \ge 2$, or when $p < 2$ and $\alpha \ge 1/p$. Under these conditions, there exists a $\gamma > 0$ sufficiently small that $\alpha'(1-\gamma) > \alpha/(2\alpha+1)$, and so $\|\theta\|^2_{2,J_0^\perp} = o(R_N)$ uniformly over $\Theta(C)$ as $N \to \infty$.

Since (27) and (28) are thus of smaller order than (25), Theorem 2 follows by applying the general bound $\mathrm{E}(\sum X_i)^2 \le (\sum \sqrt{\mathrm{E}\, X_i^2}\,)^2$ to (23).

Acknowledgements

The first author, IMJ, is grateful for this opportunity to thank Professor C. C. Heyde for his very generous guidance and mentorship at the start of IMJ's career, while an undergraduate and MSc student at the Australian National University and at CSIRO, for joint work leading to IMJ's first publication, and for encouragement and support ever since.

The paper was written while IMJ was at the University of California, Berkeley, and BWS at the University of Bristol. The research of IMJ was supported in part by NSF DMS 0072661, NIH R01 CA 72028 and EB001988-08.

References

[1] Cohen, A., Daubechies, I. and Vial, P. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Computational Harmonic Anal.* **1,** 54–81.

[2] Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993). Multiresolution analysis, wavelets and fast algorithms on an interval. *C. R. Acad. Sci. Paris* **316,** 417–421.

[3] Daubechies, I. (1993). Orthonormal bases of compactly supported wavelets. II. Variations on a theme. *SIAM J. Math. Anal.* **24,** 499–519.

[4] Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81,** 425–455.

[5] Donoho, D. L. and Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.* **26,** 879–921.

[6] Donoho, D. L. and Johnstone, I. M. (1999). Asymptotic minimaxity of wavelet estimators with sampled data. *Statistica Sinica* **9,** 1–32.

[7] Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: asymptopia? *J. R. Statist. Soc. B* **57,** 301–369.

[8] Johnstone, I. M. (2003). *Function Estimation in Gaussian Noise: Sequence Models*. Unpublished manuscript. Available at http://www-stat.stanford.edu/∼imj/.

[9] Johnstone, I. M. and Silverman, B. W. (2003). Empirical Bayes selection of wavelet thresholds. Submitted. Available at http://www.stats.ox.ac.uk/∼silverma/.

[10] Mallat, S. (1999). *A Wavelet Tour of Signal Processing*, 2nd edn. Academic Press, San Diego, CA.

[11] Okikiolu, G. O. (1971). *Aspects of the Theory of Bounded Integral Operators in L^p-Spaces*. Academic Press, London.

[12] Wojtaszczyk, P. (1997). *A Mathematical Introduction to Wavelets*. Cambridge University Press.