Statistica Sinica **9**(1999), 1-32

ASYMPTOTIC MINIMAXITY OF WAVELET ESTIMATORS WITH SAMPLED DATA

David L. Donoho and Iain M. Johnstone

Stanford University

Abstract: Donoho and Johnstone (1998) studied a setting where data were obtained in the continuum white noise model and showed that scalar nonlinearities applied to wavelet coefficients gave estimators which were asymptotically minimax over Besov balls. They claimed that this implied similar asymptotic minimaxity results in the sampled-data model. In this paper we carefully develop and fully prove this implication.

Our results are based on a careful definition of an empirical wavelet transform and precise bounds on the discrepancy between empirical wavelet coefficients and the theoretical wavelet coefficients.

Key words and phrases: Besov spaces, bounded operators between Besov spaces, Minimax estimation, thresholding, wavelet transforms of sampled data, wavelets, white noise equivalence.

1. Introduction

Suppose we have noisy sampled data

$$\tilde{y}_i = f(t_i) + \sigma \tilde{z}_i \qquad i = 0, \ldots, n, \tag{1.1}$$

where: $t_i = i/n$ are equispaced sampling points; $f(t)$, $t \in [0,1]$, is a smooth function we would like to recover; and $\tilde{z}_i \overset{i.i.d.}{\sim} N(0,1)$ is a Gaussian noise term. The unknown f belongs to a known class $\mathcal{F}$ of smooth functions, and we wish to construct an estimator achieving the minimax mean-squared-error

$$\tilde{M}(n, \mathcal{F}) = \inf_{\hat{f}} \sup_{f \in \mathcal{F}} E\|\hat{f} - f\|^2_{L^2[0,1]}. \tag{1.2}$$

In earlier work Donoho and Johnstone ((1990) and (1998)) studied this problem in the case where $\mathcal{F}$ is a ball in a Besov or Triebel space (this assumption includes the more familiar Sobolev and Hölder classes) and concluded that certain wavelet-domain estimators, based on applying scalar nonlinearities to the empirical wavelet coefficients of the data $(y_i)_{i=0}^n$, would be asymptotically minimax for certain $\mathcal{F}$. In that work, use was made of a result we intend to prove fully in this article.

In fact Donoho and Johnstone ((1990) and (1998)) did not study the problem (1.1)-(1.2) directly . Their attention was directed to observations from the so-called "white-noise model"

$$Y_\epsilon(dt) = f(t)dt + \epsilon W(dt) \qquad t \in [0,1], \tag{1.3}$$

where now $\epsilon > 0$ is a small parameter, f is an unknown smooth function which we would like to recover, and $W(t)$ is a standard Wiener process, so that $W(dt)$ is white noise. Donoho and Johnstone ((1990) and (1998)) studied the minimax mean-squared-error in the white noise model,

$$M(\epsilon, \mathcal{F}) = \inf_{\hat{f}} \sup_{f\in\mathcal{F}} E\|\hat{f} - f\|^2_{L^2[0,1]}, \tag{1.4}$$

where the infimum is taken over measurable procedures $\hat{F}$ yielding estimates $\hat{f} = \hat{F}(Y_\epsilon)$. They studied the asymptotic behavior of the minimax risk as $\epsilon \to 0$, where again $\mathcal{F}$ is a ball of certain Triebel or Besov spaces, and found that, for a significant range of such spaces, appropriate estimators built from nonlinear functions of the wavelet coefficients were asymptotically minimax over $\mathcal{F}$. Such estimators take the form

$$\hat{f} = \sum_I \eta_I(\langle \psi_I, Y_\epsilon \rangle)\psi_I, \tag{1.5}$$

where (ψ_I) is a nice wavelet orthonormal basis of $[0,1]$ and (η_I) is a collection of scalar nonlinearities, depending on ϵ and $\mathcal{F}$. Hence, in the white noise model, scalar nonlinearities applied to wavelet coefficients yield asymptotically minimax estimates over certain Besov and Triebel balls. Donoho and Johnstone ((1990) and (1998)) then argued that the results in the white noise model (1.3)-(1.4) implied similar phenomena for the sampled data model (1.1)-(1.2).

It is well established that one may prove results in the sampled-data model (1.1)-(1.2) by first proving them in the white noise model (1.3)-(1.4) and then arguing that this implies parallel results for sampled data. Examples where this has been done in detail include Donoho and Nussbaum (1990) and Donoho (1994). The general equivalence result of Brown and Low (1996) shows that for bounded loss function $\ell(.)$ and for collections $\mathcal{F}$ which are bounded subsets of Hölder classes $C^{1/2+\delta}, \delta > 0$, we have, under the calibration $\epsilon = \sigma/\sqrt{n}$,

$$\inf_{\hat{f}} \sup_{f\in\mathcal{F}} E_{Y_\epsilon}\ell\Big(\|\hat{f} - f\|^2_{L^2[0,1]}\Big) \asymp \inf_{\hat{f}} \sup_{f\in\mathcal{F}} E_{\mathbf{y}_n}\ell\Big(\|\hat{f} - f\|^2_{L^2[0,1]}\Big), \tag{1.6}$$

the expectation on the left-hand side being with respect to white noise observations Y_ϵ and on the right hand-side being with respect to $\mathbf{y}_n = (\tilde{y}_0, \ldots, \tilde{y}_n)$. Hence there is considerable tradition to support the approach of Donoho and

Johnstone ((1990) and (1998)). Our goal in this paper is to establish, as explicitly as possible, for as wide a scale of $\mathcal{F}$ as possible, for the unbounded loss function $\|\hat{f} - f\|^2_{L^2[0,1]}$, that results of Donoho and Johnstone (1998) for the white noise model imply corresponding results for the sampled-data model.

In the article to follow we establish two main types of results. First, we establish lower bounds showing that the sampled-data problem is not easier than the white noise problem. Second, we establish upper bounds showing that the sampled-data problem is not harder than the white-noise problem.

1.1. Lower bounds

Theorem 1.1. *Let $\alpha > 1/p$ and $1 \le p, q \le \infty$, or $\alpha = p = q = 1$. Let F be either a Triebel space $F^{\alpha}_{p,q}[0,1]$ or a Besov space $B^{\alpha}_{p,q}[0,1]$, normed in the way we describe in Section 2 below. For $C > 0$, define the ball*

$$\mathcal{F} = \{f : \|f\|_F \le C\}. \tag{1.7}$$

Then, with $\epsilon_n = \sigma/\sqrt{n}$ we have

$$\tilde{M}(n, \mathcal{F}) \ge M(\epsilon_n, \mathcal{F})(1 + o(1)), \quad n \to \infty. \tag{1.8}$$

In words, there is no measurable estimator giving a worst-case performance in the sampled-data-problem (1.2) which is substantially better than what we can get for the worst-case performance of measurable procedures in the white-noise-problem (1.4).

Remark. Although Donoho and Johnstone (1998) gives more detailed information, we recall here that

$$M(\epsilon_n, \mathcal{F}) \asymp C^{2(1-r)}\sigma^{2r}n^{-r}, \qquad n \to \infty, \qquad r = 2\alpha/(2\alpha+1). \tag{1.9}$$

1.2. Upper bounds

We will specialize to estimators derived by applying certain coordinatewise mappings to the noisy wavelet coefficients.

For the white-noise model, this means the estimate is of the form (1.5) where each function $\eta_I(y)$ either belongs to one of three specific families – *Linear, Soft Thresholding, or Hard Thresholding* – or else is a general scalar function of a scalar argument. The families are described in the following table:

Family	Description	Form of $\eta_I(y)$
$\mathcal{E}_L$	Diagonal linear procedures in the wavelet domain	$\eta_I^L(y) = c_I \cdot y$
$\mathcal{E}_S$	Soft thresholding of wavelet coefficients	$\eta_I^S(y) = (\lvert y\rvert - \lambda_I)_+ \mathrm{sgn}(y)$
$\mathcal{E}_H$	Hard thresholding of wavelet coefficients	$\eta_I^H(y) = y 1_{\{\lvert y\rvert \ge \lambda_I\}}$
$\mathcal{E}_N$	Scalar nonlinearities of wavelet coefficients	Arbitrary $\eta_I^N(y)$

For the sampled-data problem, this means that the estimate is of the form

$$\hat{f} = \sum_I \eta_I(y_I^{(n)})\psi_I, \tag{1.10}$$

where $y_I^{(n)}$ is an empirical wavelet coefficient based on the sampled data $(\tilde{y}_i)$ defined in Section 4.2 below, and the η_I belong to one of the families $\mathcal{E}$. Then define the $\mathcal{E}$-minimax risks in the two problems:

$$M_{\mathcal{E}}(\epsilon, \mathcal{F}) = \inf_{\hat{f}\in\mathcal{E}} \sup_{f\in\mathcal{F}} E_{Y_\epsilon}\|\hat{f} - f\|^2_{L^2[0,1]} \tag{1.11}$$

and

$$\tilde{M}_{\mathcal{E}}(n, \mathcal{F}) = \inf_{\hat{f}\in\mathcal{E}} \sup_{f\in\mathcal{F}} E_{\mathbf{y}_n}\|\hat{f} - f\|^2_{L^2[0,1]}. \tag{1.12}$$

We write M_L for the linear minimax risk $M_{\mathcal{E}_L}$ etc.

This definition is really four definitions – one for each choice of $\mathcal{E}$. To avoid confusion, we spell this out in a particular case. Let $\eta^S_\lambda(y) = (|y| - \lambda)_+\mathrm{sgn}(y)$ denote the soft thresholding nonlinearity. Let

$$M_S(\epsilon, \mathcal{F}) = \inf_{(\lambda_I)} \sup_{\mathcal{F}} E\|\hat{f} - f\|^2_{L^2[0,1]}, \tag{1.13}$$

where the infimum is over choice of thresholds (λ_I), and the estimator $\hat{f}$ takes the form

$$\hat{f} = \sum_I \eta^S_{\lambda_I}(\langle\psi_I, Y_\epsilon\rangle)\psi_I. \tag{1.14}$$

With this notation established, we have

Theorem 1.2. *Let $\alpha > 1/p$ and $1 \le p, q \le \infty$ or $\alpha = p = q = 1$. For each of the four classes $\mathcal{E}$ of coordinatewise estimators,*

$$\tilde{M}_{\mathcal{E}}(n, \mathcal{F}) \le M_{\mathcal{E}}(\epsilon_n, \mathcal{F})(1 + o(1)), \quad n \to \infty. \tag{1.15}$$

Our approach is to make an explicit construction transforming a sampled-data problem into a quasi-white-noise problem in which estimates from the white-noise model can be employed. We then show that these estimates on the quasi-white-noise-model data behave nearly as well as on the truly-white-noise-model data. The construction is described in detail in Section 3 and its properties are established in Section 4.

1.3. Implications

Several conclusions follow immediately from these bounds.

First, *asymptotic minimaxity of scalar nonlinearities.* Donoho and Johnstone (1998) show that for Besov balls $\mathcal{F}$ with $p \le q$ and normed as described in Section 2 below,

$$M_N(\epsilon, \mathcal{F}) \sim M(\epsilon, \mathcal{F}) \quad \text{as } \epsilon \to 0 \tag{1.16}$$

This is, appropriate scalar nonlinearities of the wavelet coefficients are asymptotically minimax among all measurable procedures.

Corollary 1.3. *Let $\alpha > 1/p$ and $1 \le p \le q \le \infty$. For $\mathcal{F}$ a ball (1.7) in the Besov scale,*

$$\tilde{M}_N(n, \mathcal{F}) \sim \tilde{M}(n, \mathcal{F}), \quad n \to \infty. \tag{1.17}$$

Second, *near-asymptotic minimaxity of hard/soft thresholding.*

Donoho and Johnstone (1998) show that for Besov and Triebel balls,

$$M_S(\epsilon, \mathcal{F}) \le \Lambda^S(p, q) M(\epsilon, \mathcal{F})(1 + o(1)) \quad \epsilon \to 0, \tag{1.18}$$

$$M_H(\epsilon \mathcal{F}) \le \Lambda^H(p, q) M(\epsilon, \mathcal{F})(1 + o(1)) \quad \epsilon \to 0, \tag{1.19}$$

where Λ^S and Λ^H are constants depending on p and q only, and can be bounded explicitly; e.g., $\Lambda^S \le 2.2$, $\forall\, 1 \le p, q \le \infty$.

Corollary 1.4. *Let $\alpha > 1/p$ and $1 \le p, q \le \infty$ or let $\alpha = p = q = 1$. Then*

$$\tilde{M}_S(n, \mathcal{F}) \le \Lambda^S(p, q) \tilde{M}(n, \mathcal{F})(1 + o(1)) \quad n \to \infty, \tag{1.20}$$

$$\tilde{M}_H(n, \mathcal{F}) \le \Lambda^H(p, q) \tilde{M}(n, \mathcal{F})(1 + o(1)) \quad n \to \infty. \tag{1.21}$$

Third, *near-asymptotic minimaxity of linear estimates.* Donoho, Liu and MacGibbon (1990) and Donoho and Johnstone (1998) show that for Besov bodies and $p, q \ge 2$,

$$M_L(\epsilon, \mathcal{F}) \le 1.25 \cdot M(\epsilon, \mathcal{F})(1 + o(1)), \quad \epsilon \to 0. \tag{1.22}$$

Corollary 1.5. *Let $\alpha > 1/p$ and $2 \le p, q \le \infty$. Then*

$$\tilde{M}_L(n, \mathcal{F}) \le 1.25 \cdot \tilde{M}(n, \mathcal{F})(1 + o(1)) \quad n \to \infty.$$

1.4. Precision of the empirical wavelet transform

Our approach to the upper bounds is based on defining and studying a certain empirical wavelet transform, a scheme for transforming finite sequences into approximate wavelet coefficients

$$(f(t_i))_{i=0}^n \stackrel{\mathcal{W}_n}{\to} (\tilde{\theta}_I)_I.$$

This transform is defined using Deslauriers-Dubuc Interpolation, which starts from $n+1$ data $f(t_i)$, and produces a smooth interpolating function $\tilde{P}_n f(t)$. The coefficients are obtained by setting $\tilde{\theta}_I = \langle \tilde{P}_n f, \psi_I \rangle$.

This transform is "accurate" in the sense that for each fixed I, the empirical coefficient $\tilde{\theta}_I = \tilde{\theta}_I^{(n)}$ converges to the theoretical counterpart θ_I as $n \to \infty$. More significantly, we can establish approximation in norm for the vector of coefficients out to a certain cutoff scale. Suppose that (α, p, q) are fixed, and we set

$$\gamma = \begin{cases} \frac{1}{2\alpha+1} & p \geq 2 \\ \frac{1}{2\alpha+1} \frac{\alpha}{\alpha+1/2-1/p} & p < 2 \end{cases}. \tag{1.23}$$

Fix a small positive number η so that $\lambda = \gamma + \eta < 1$, and then define the *cutoff scale*

$$j_0 = j_0(n; \alpha, p, q, \eta) = \lceil \lambda \log_2(n) \rceil. \tag{1.24}$$

We will be interested only in the degree of approximation achieved by the empirical wavelet coefficients at scales $j \leq j_0$. Our results can be stated in terms of the partial reconstruction

$$\tilde{f} = \sum_{|I| \geq 2^{-j_0}} \tilde{\theta}_I \psi_I; \tag{1.25}$$

we will establish, in Section 5 below, the following result.

Theorem 1.6. *Let $\alpha > 1/p$ and $1 \leq p, q \leq \infty$ or $\alpha = p = q = 1$. Then for F a Triebel or Besov space with the prescribed (α, p, q) and $\mathcal{F}$ the corresponding ball* (1.7)*,*

$$\sup_{f \in \mathcal{F}} \|\tilde{f} - f\|_F \to 0. \tag{1.26}$$

In short, the empirical wavelet coefficients are uniformly accurate over Besov or Triebel balls in a certain range of values (α, p, q), in a strong norm. *It is the range of validity of this result that governs the range of validity of (α, p, q) in our earlier theorems.*

Of course, the issue of the accuracy of empirical wavelet coefficients is not new. Strang has called the lack of attention to this issue one of the great "wavelet crimes".

Previous studies of sampling include Donoho (1992b), Donoho (1993) and Delyon and Juditsky (1996). Those articles addressed a different program than our current one. The results of Donoho (1992b) show that for the empirical transform coefficients $(\bar{\theta}_I^{(n)})$ defined in that work, the object

$$\bar{f}_n = \sum_{j \leq \log_2(n)} \bar{\theta}_I^{(n)} \psi_I$$

satisfies bounds of the form

$$\|\bar{f}_n\|_{\mathcal{F}} \leq A \cdot \|f\|_{\mathcal{F}}, \tag{1.27}$$

for an unknown constant A not necessarily close to 1. Such inequalities are useful for establishing that the most natural definition of empirical wavelet transform for sampled data and the most natural implementation of wavelet thresholding achieve nearly- (but not precisely-) minimax behavior (compare Donoho, Johnstone, Kerkyacharian and Picard (1995), Donoho (1992a)). In the context of this paper, (1.27) would lead easily to the conclusion

$$\tilde{M}(n, \mathcal{F}) \leq A \cdot M(\epsilon_n, \mathcal{F}) \tag{1.28}$$

for a constant A which is unknown and not expected to be close to 1 – but not to anything sharper. However, this conclusion is considerably weaker than what we prove here, which might be viewed as of the same form as (1.28), only with $A = A_n \to 1$ as $n \to \infty$. Inequalities like (1.28) with unknown $A > 1$ would not yield the corollaries given Section 1.3.

Earlier work does not seem to us to be oriented towards developing uniform bounds (1.26). In consequence, we do not see that that work gives information at the level of precise constants – our goal here.

1.5. Contents

The article begins in Section 2 with some background on the wavelet bases we are using and the Besov/Triebel scales they induce. Lower bounds on the difficulty of the sampled-data problem are given in Section 3. Upper bounds on the difficulty are given in Section 4, where a specific construction is provided. Technical lemmas are proved in Section 5, and a key technical result implying Theorem 1.6 is proven in Section 6.

2. Sequence Space

In the background, we are always assuming that we have fixed the choice (α, p, q) and the choice of the Triebel or Besov scale, so that we will be interested in a fixed Triebel space $F^{\alpha}_{p,q}[0, 1]$ or a Besov space $B^{\alpha}_{p,q}[0, 1]$. Once these choices have been made, we choose a wavelet basis $(\psi_I)_I$ compatible with this space. In this paper, this will always be a nice orthonormal wavelet basis, consisting of wavelets of compact support, with elements having R continuous derivatives ($\psi_I \in C^R$) and $D + 1$ vanishing moments ($\int t^{\ell} \psi_I(t) dt = 0$, $\ell = 0, \dots, D$). We impose the vanishing moment condition only at sufficiently fine scales; see below. The basis is chosen so that $\min(R, D) \geq \alpha$. Then the wavelet basis will be an unconditional basis of the corresponding space of interest, which is the essential point for us. Compare Frazier, Jawerth and Weiss (1991), Meyer (1990).

A word about our indexing scheme. Wavelets are indexed by dyadic intervals $I = I_{j,k} = [k/2^j, (k + 1)/2^j)$, $j \geq 0$ and $0 \leq k < 2^j$. The indexing scheme reflects the fact that wavelet ψ_I is localized so that for an appropriate $c > 0$, $supp(\psi_I) \subset cI$ for every I, where cI denotes the interval with the same center as I dilated by a factor c.

We include in our index set an extra interval $I_{-1,0} = [0, 2)$ (which doesn't fit in $[0, 1)$). Hence our indexing set $\mathcal{I} = \{I_{-1,0}, I_{0,0}, I_{1,0}, I_{1,1}, I_{2,0}, \ldots\}$. It will be convenient at times to break this down into 'resolution levels' $\mathcal{I}_j = \{I_{j,k}, 0 \leq k < 2^j\}$ based on the length 2^{-j} of the intervals. We occasionally write $\mathcal{I}_j{}'$ for $\cup_{j' \leq j} \mathcal{I}_{j'}$.

For the standard Haar basis, the exceptional interval $I_{-1,0}$ indexes the 'father' wavelet $\psi_{-1,0}(t) = 1_{[0,1]}$, which has integral 1, while the other intervals index wavelets $\psi_I(t) = |I|^{-1/2} h(2^j t - k)$, where $h(t) = 1_{[1/2,1)}(t) - 1_{[0,1/2)}(t)$ is the Haar function with integral 0. For modern wavelet bases, the single exceptional interval becomes an exceptional layer: for a certain counting number $m \geq 0$, we gather together $I_{-1,0}$, $I_{0,0}$, and the levels $\mathcal{I}_j$ for $j < m$ into a set $\mathcal{I}'$ of cardinality 2^m. These 'first' 2^m wavelets ψ_I, $I \in \mathcal{I}'$ do not necessarily have vanishing moments. At all 'finer' scales $|I| \leq 2^{-m}$ the wavelets do have vanishing moments.

Since the wavelet basis is required to live on the interval $[0, 1]$, we assume it has the following structure (for details, see Cohen, Daubechies, Jawerth and Vial (1993)). Suppose that the wavelet ψ has support $[-k_0, k_0 + 1]$. For $k_0 \leq k < 2^j - k_0, \psi_I(t) = \psi^0_{jk}(t) = 2^{j/2}\psi(2^j t - k)$. In the boundary cases, ψ_{jk} are obtained from translates of ψ^0_{jk} by orthonormalization on $[0, 1]$. In particular, there exist $k_0 \times k_0$ matrices $E^\sharp = (e^\sharp_{kk'})$, $E^\flat = (e^\flat_{kk'})$, *not depending on* j, such that for $0 \leq k < k_0$:

$$\psi_{jk}(t) = \sum_{k'=0}^{k_0-1} e^\sharp_{kk'} \psi^0_{jk'}(t), \qquad \psi_{j,2^j-k}(t) = \sum_{k'=0}^{k_0-1} e^\flat_{kk'} \psi^0_{j,2^j-k'}(t). \tag{2.1}$$

In particular, the 'boundary' wavelets have the same smoothness properties as the interior wavelets.

2.1. Balls in sequence space

Let f be a square-integrable function on $[0, 1]$. Then f has wavelet coefficients $\theta_I = \langle f, \psi_I \rangle$ and these coefficients obey a Parseval relation $\|\theta\|_{\ell^2} = \|f\|_{L^2}$.

We have a similar relation for our Besov/Triebel norm. By our assumption that the wavelet basis has been constructed to be an unconditional basis for $B^\alpha_{p,q}$ or $F^\alpha_{p,q}$ of interest, we can use it to define the corresponding Besov/Triebel norm. Define

$$\|\theta\|_{\mathbf{b}^\alpha_{p,q}} = \Big(\sum_j 2^{jaq} \Big(\sum_{I \in \mathcal{I}_j} |\theta_I|^p \Big)^{q/p} dt \Big)^{1/q}, \tag{2.2}$$

with $a = \alpha + 1/2 - 1/p$ and, for $\chi_I(t) = 1_I(t)$,

$$\|\theta\|_{\mathbf{f}^\alpha_{p,q}} = \Big(\int_0^1 (\sum_I |\theta_I|^q 2^{jaq} \chi_I(t))^{p/q} dt\Big)^{1/p},$$

where now $a = \alpha + 1/2$. Then we may use the correspondence $f \leftrightarrow \theta$ to define norms for Besov and Triebel classes;

$$\|f\|_{B^\alpha_{p,q}} = \|\theta\|_{\mathbf{b}^\alpha_{p,q}} \tag{2.3}$$

$$\|f\|_{F^\alpha_{p,q}} = \|\theta\|_{\mathbf{f}^\alpha_{p,q}}, \tag{2.4}$$

norms for Besov/Triebel space equivalent to the usual ones. For example the Hölder space $\Lambda^\alpha = B^\alpha_{\infty,\infty}[0,1]$, $0 < \alpha < 1$, has traditionally used the norm

$$\|f\|_{\Lambda^\alpha} = \|f\|_{L^\infty} + \sup_{t,s\in[0,1]} \frac{|f(s) - f(t)|}{|t-s|^\alpha},$$

but we use instead the equivalent norm

$$\|f\|_{B^\alpha_{\infty,\infty}} = \sup_I |I|^{\alpha+1/2} |\theta_I|.$$

These Besov and Triebel norms agree when $p = q$: $\|\theta\|_{\mathbf{f}^\alpha_{p,p}} = \|\theta\|_{\mathbf{b}^\alpha_{p,p}}$. Otherwise, norms from the Triebel scale are bracketed by norms from the Besov scale with the same σ and p, but different q:

$$a_0 \|\theta\|_{\mathbf{b}^\alpha_{p,p\vee q}} \le \|\theta\|_{\mathbf{f}^\alpha_{p,q}} \le a_1 \|\theta\|_{\mathbf{b}^\alpha_{p,p\wedge q}}, \tag{2.5}$$

with $a_i = a_i(\alpha, p, q)$ (see, e.g., Peetre (1975), p. 261 or Triebel (1992), p. 96).

Below, by a Besov or Triebel Ball $\mathcal{F}$ with parameters (α, p, q) we mean that $\mathcal{F}$ is of the form

$$\mathcal{F} = \{f : \|f\|_F \le C\},$$

where $F = F^\alpha_{p,q}$ or $F = B^\alpha_{p,q}$. Ultimately, this amounts to saying that $\mathcal{F} = \{f : \|\theta\|_{\mathbf{f}} \le C\}$ where $\mathbf{f} = \mathbf{f}^\alpha_{p,q}$ or $\mathbf{b}^\alpha_{p,q}$. In that sense $\mathcal{F}$ is just the image under an (ℓ^2, L^2) isometry of the ball of coefficients $\Theta(C) = \{\theta : |\theta\|_{\mathbf{f}} \le C\}$.

We will frequently need two facts. The first is implicitly used throughout. It is well known and we give it without proof.

Lemma 2.1. *Point evaluation $f \mapsto f(t_0)$ is a continuous linear functional on spaces $F^\alpha_{p,q}$ or $B^\alpha_{p,q}$ where $\alpha > 1/p$, $1 \le p, q \le \infty$, or $\alpha = p = q = 1$ (and not for $\alpha < 1/p$, or $\alpha = 1/p$, $q > 1$).*

This may provide the reader with an early understanding of the reason that the condition "$\alpha > 1/p$, $1 \le p, q \le \infty$ or $\alpha = p = q = 1$" occurs throughout: the sampling model (1.1) makes sense over this scale and no larger one.

A second point is that in the range of (α, p, q) we are considering, the wavelet coefficients decay at increasingly fine scales.

Lemma 2.2. *Let* $P_j f = \sum_{\mathcal{I}_j{}'} \theta_I \psi_I$ *where* $\mathcal{I}_j{}'$ *is the collection of all* I *with scale parameter* j *or less,*

$$\sup_{f \in \mathcal{F}} \|f - P_j f\|_{L^2} = O(2^{-j\alpha'}),$$

where

$$\alpha' = \begin{cases} \alpha & p \geq 2 \\ \alpha + 1/2 - 1/p & p < 2 \end{cases}. \tag{2.6}$$

Beginning with this lemma, we collect the proofs in Section 5 below.

2.2. Estimation in sequence space

Donoho and Johnstone (1998) studied a sequence space estimation problem of the following form: we suppose observations

$$\bar{y}_I = \theta_I + \epsilon \bar{z}_I, \quad I \in \mathcal{I}, \tag{2.7}$$

where $\bar{z}_I \overset{i.i.d}{\sim} N(0,1)$ and θ is an unknown vector which we would like to recover with small squared-error loss. They suppose $\theta \in \Theta$, a fixed set, and consider the minimax risk

$$\overline{M}(\epsilon, \Theta) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} E\|\hat{\theta} - \theta\|_{\ell^2}^2. \tag{2.8}$$

This sequence space problem is closely connected with the white-noise problem (1.4). Let $Y_\epsilon(t)$ be the white-noise model data of (1.3); the expression $\bar{y}_I = \langle Y_\epsilon, \psi_I \rangle$ makes sense. We can use either stochastic integration theory, or, more prosaically, integration by parts to sensibly interpret it:

$$\int \psi_I(t) Y_\epsilon(dt) = \int \psi_I(t) f(t) dt + \epsilon \int \psi_I(t) W(dt) = \langle \psi_I, f \rangle - \epsilon \cdot \langle (\psi_I)', W \rangle.$$

The resulting data obey

$$\bar{y}_I = \theta_I + \epsilon \bar{z}_I, \quad \forall I,$$

where the θ_I are precisely the wavelet coefficients of f and $\bar{z}_I \overset{i.i.d.}{\sim} N(0,1)$.

Under the correspondence $f = \sum \theta_I \psi_I$, with $\mathcal{F} = \{f : f = \sum \theta_I \psi_I, \theta \in \Theta\}$, the sequence problem is in isometric correspondence with the original white noise problem, and the minimax risk (1.4) from observations (1.3) obeys

$$\overline{M}(\epsilon, \Theta) = M(\epsilon, \mathcal{F}), \qquad \forall \epsilon > 0. \tag{2.9}$$

Similarly, if attention is restricted to co-ordinatewise estimators of class $\mathcal{E}$, then $\overline{M}_{\mathcal{E}}(\epsilon, \Theta) = M_{\mathcal{E}}(\epsilon, \mathcal{F})$. Given an estimator class $\mathcal{E}$ and ball $\Theta = \Theta(\alpha, p, q, C)$, we denote by

$$\hat{\theta}_I = \delta_I(\bar{y}_I; \mathcal{E}, \Theta, \epsilon), \quad I \in \mathcal{I},$$

the minimax co-ordinatewise estimator corresponding to noise level ϵ:

$$\sup_{\theta\in\Theta} E||\hat{\theta}-\theta||_{\ell^2}^2 = \overline{M}_{\mathcal{E}}(\epsilon,\Theta).$$

We will need to study what happens in cases where Θ and ϵ are inflated slightly: for this purpose we use

Lemma 2.3. *If $\epsilon_1 \geq \epsilon_0$ and $C_1 \geq C_0$, then*

$$\overline{M}_{\mathcal{E}}(\epsilon_1, C_1) \leq (\epsilon_1/\epsilon_0)^2 (C_1/C_0)^2 \overline{M}_{\mathcal{E}}(\epsilon_0, C_0).$$

3. Sampling Does not Make Things Easier

Donoho and Johnstone (1998) derived information about the structure of least-favorable priors for the sequence-space problem (2.8) when Θ is a Besov ball or a Triebel ball. Let μ be a prior distribution on Θ, and let

$$\overline{B}(\mu,\epsilon) = E\|E\{\theta|y\}-\theta\|_{\ell^2}^2 \tag{3.1}$$

denote the Bayes Risk, where the expectations are taken with respect to the joint distribution $\theta\sim\mu, \quad \bar{y}=\theta+\epsilon\bar{z}, \quad \bar{z}\perp\theta$. They studied an asymptotically least-favorable prior $\mu^{(\epsilon)}$, obeying

$$\overline{B}(\mu^{(\epsilon)},\epsilon) = \sup\{\overline{B}(\mu,\epsilon)\ : supp(\mu)\subset\Theta\}(1+o(1)) \quad \epsilon\to 0; \tag{3.2}$$

for this prior, the minimax theorem says that

$$\overline{B}(\mu^{(\epsilon)},\epsilon) = \overline{M}(\epsilon,\Theta)(1+o(1)), \quad \epsilon\to 0. \tag{3.3}$$

Our proof of Theorem 1 begins with the assertion it is not necessary to consider coefficients at scales finer than our cutoff scale $j_0(n)$ defined by (1.24) in order to achieve an asymptotically least favorable prior.

Lemma 3.1. *Let $1\leq p,q\leq\infty$ and $\alpha\geq 1/p$ or $\alpha=p=q=1$. Let $\theta^{(\epsilon)}$ be distributed according to the (asymptotically) least favorable prior $\mu^{(\epsilon)}$. Let $\theta^{[n]}$ be a random sequence with*

$$\theta_I^{[n]} = \begin{cases} \theta_I^{(\epsilon_n)} & 0\leq j\leq j_0(n),\ k=0,\dots,2^j-1, \\ 0 & \textit{otherwise.} \end{cases}$$

Let $\mu^{[n]}$ be the distribution of $\theta^{[n]}$. Then
(a) $supp(\mu^{[n]})\subset\Theta$,
(b) $\overline{B}(\mu^{[n]},\epsilon_n)\sim\overline{B}(\mu^{(\epsilon_n)},\epsilon_n)$, $n\to\infty$.

The implication is that there is a prior distribution $\mu^{[n]}$ using a limited range of scales and yet achieving

$$\overline{B}(\mu^{[n]}, \epsilon_n) \sim \overline{M}(\epsilon_n, \Theta), \quad n \to \infty. \tag{3.4}$$

The proof is in the appendix: here we explain only why $j_0(n)$ has the form (1.23) - (1.24). Since $\theta^{(\epsilon_n)} \in \Theta$ always and $|\theta_I^{[n]}| \leq |\theta_I^{(\epsilon_n)}|$ for all I, orthosymmetry of Θ implies that $\theta^{[n]} \in \Theta$ always. Hence, by Lemma 2.2,

$$\begin{aligned} \|\theta^{[n]} - \theta^{(\epsilon_n)}\|^2 &\leq \sup_{\Theta} \|P_{j_0} f - f\|^2 \\ &= O(2^{-2j_0\alpha'}) = O(n^{-r'}), \end{aligned} \tag{3.5}$$

where (compare (1.23) and (2.6)): $r' = 2\lambda\alpha' > 2\gamma\alpha' = 2\alpha/(2\alpha+1) = r$, so that the error in ignoring terms beyond j_0 is of smaller order than the minimax risk (compare (1.9)).

We will focus attention on the prior $\mu^{[n]}$ and show that its Bayes risk for estimation from sampled data is asymptotically at least as bad as the Bayes risk for estimation from sequence data.

The idea is as follows. Let $\theta^{[n]}$ denote the random variable constructed in Lemma 3.1. Consider the sampling operator

$$T(\theta^{[n]}) = (f^{[n]}(t_0), \ldots, f^{[n]}(t_n))', \tag{3.6}$$

where $f^{[n]}(t) = \sum_{|I| \geq 2^{-j_0}} \theta_I^{[n]} \psi_I(t)$. Thus T is a linear mapping from $\mathcal{R}^m$ into $\mathcal{R}^{n'}$, where $m = 2^{j_0+1}$ and $n' = n+1$. Think of $\mathcal{R}^m$ as (an initial segment of) sequence space with norm $\|\theta\|^2_{\ell^2_m} = \sum_1^m \theta_i^2$, whereas $\mathcal{R}^{n'}$ corresponds to a discretization of $[0,1]$, and so is naturally normed by $\|\xi\|_n^2 = (1/n)\sum_0^n \xi_i^2$. If we have sampled data (1.1) with f the random object $f^{[n]}$, we are equivalently observing

$$\tilde{y} = T(\theta^{[n]}) + \sigma\tilde{z}, \tag{3.7}$$

where $\tilde{y}, \tilde{z} \in \mathcal{R}^{n'}$, $\tilde{z}_i \overset{i.i.d}{\sim} N(0,1)$. Now in comparison, the sequence data

$$\overline{y}_I = \theta_I^{[n]} + \epsilon_n \bar{z}_I \quad , \quad |I| \geq 2^{-j_0}. \tag{3.8}$$

Note the difference between the two: in one case, one observes $\theta^{[n]}$ with noise; in the other case, one observes a linear transform of $\theta^{[n]}$ with noise. Now *if* T were a partial isometry, mapping m-vectors θ into n'-vectors of equal norm, then the Bayes risk from observations (3.7),

$$\tilde{B}(\tilde{\mu}^{[n]}, \sigma) = E\|E\{T(\theta^{[n]})|\tilde{y}\} - T(\theta^{[n]})\|_n^2,$$

(where $\tilde{\mu}^{[n]}$ stands for the distribution of $T(\theta^{[n]})$ when $\theta^{[n]}$ is distributed $\mu^{[n]}$) would obey

$$\tilde{B}(\tilde{\mu}^{[n]}, \sigma) = \overline{B}(\mu^{[n]}, \epsilon_n). \tag{3.9}$$

Indeed, if T were a partial isometry, then $T^*\tilde{y} =_D \overline{y}$ and $T\overline{y} + \sigma(I - TT^*)z =_D \tilde{y}$ show that there are randomized mappings turning observations according to one model into observations according to the other model. Consequently $(\overline{y}, \Theta)$, $(\tilde{y}, \Theta)$ would be equivalent experiments in Le Cam's sense, and (3.9) would follow.

Now T is **not** in general a partial isometry, but it is close.

Lemma 3.2. *Let $\alpha > 1/p$ and $1 \le p, q \le \infty$, or $\alpha = p = q = 1$. Define T as above.*

(a) *There exists a partial isometry $I^{[n]}$ from $(\mathcal{R}^m, \|\cdot\|_{\ell^2_m})$ to $(\mathcal{R}^{n'}, \|\cdot\|_n)$ so that*

$$\delta_n^2 := \sup_{\theta \in \Theta} \quad \|(T - I^{[n]})\theta\|_n^2 = o(n^{-r}), \tag{3.10}$$

where $r = 2\alpha/(2\alpha + 1)$.

(b) *The largest singular value of $T^{(n)} = 1 + o(1)$; i.e.*

$$\lambda_{\max}(T^*T) = \|T^*T\| = 1 + o(1). \tag{3.11}$$

With this degree of approximation of T by a partial isometry, everything goes asymptotically as if T were itself a partial isometry.

Lemma 3.3. *Let $\overline{y}$ and $\tilde{y}$ be as in* (3.7) *and* (3.8) *above. Let $\theta^{[n]}$ be as in Lemma* 3.1*, so that $\theta^{[n]} \in \Theta$ w.p.1. Suppose*

$$\sup_{\theta \in \Theta} \quad \|(T - I^{[n]})\theta\|_n^2 \le \delta^2. \tag{3.12}$$

Then

$$\sqrt{\tilde{B}}(\tilde{\mu}, \sigma) \ge [\sqrt{\overline{B}}(\mu, \epsilon_n) - \delta] / \max(1, \lambda_{\max}(T^*T)).$$

It follows from (3.10) and (3.11) and Lemma 3.3 that

$$\tilde{B}(\tilde{\mu}^{[n]}, \sigma) \ge \overline{B}(\mu^{[n]}, \epsilon_n)(1 + o(1)). \tag{3.13}$$

The final step is to relate the frequentist minimax MSE $\tilde{M}(n, \mathcal{F})$ which uses the $L^2[0,1]$ norm to the Bayes risk $\tilde{B}(\tilde{\mu}^{[n]}, \epsilon_n)$ which is based on the discretized norm $\|\cdot\|_n$ on $\mathcal{R}^{n'}$.

First we write $\tilde{M}(n, \mathcal{F})$ in sequence space terms using Parseval's relation. In what follows, $\tilde{y}$ denotes the vector of observations from model (1.1):

$$\tilde{M}(n, \mathcal{F}) = \inf_{\hat{f}(\tilde{y})} \sup_{f \in \mathcal{F}} E\|\hat{f}(\tilde{y}) - f\|^2_{L^2[0,1]}$$

$$
\begin{aligned}
&= \inf_{\hat\theta(\tilde y)} \sup_{\theta\in\Theta} E\|\hat\theta(\tilde y) - \theta\|_{\ell^2}^2 \\
&\geq \inf_{\hat\theta(\tilde y)} \sup_{\theta\in\Theta} E\|\hat\theta(\tilde y) - \theta\|_{\ell^2_m}^2 \\
&\geq E_{\mu^{[n]}}\|E(\theta^{[n]}|\tilde y) - \theta^{[n]}\|_{\ell^2_m}^2, \qquad (3.14)
\end{aligned}
$$

where $E_{\mu^{[n]}}$ denotes expectation with respect to the joint distribution of $(\theta^{[n]}, \tilde y)$, in which $\theta^{[n]} \sim \mu^{[n]}$ and $\tilde y$ follows model (3.7).

Second, we use the definition of Bayes risk to write

$$
\begin{aligned}
\tilde B(\tilde\mu^{[n]}, \sigma) &= \inf_{\hat\xi(\tilde y)} E_{\mu^{[n]}}\|\hat\xi(\tilde y) - T\theta^{[n]}\|_n^2 \qquad (3.15)\\
&\leq \inf_{\hat\theta(\tilde y)} E_{\mu^{[n]}}\|I^{[n]}\hat\theta(\tilde y) - T\theta^{[n]}\|_n^2, \qquad (3.16)
\end{aligned}
$$

where $I^{[n]}$ is the partial isometry of Lemma 3.2.

To link these two quantities, we write

$$
\begin{aligned}
\|I^{[n]}\hat\theta - T\theta\|_n &= \|I^{[n]}(\hat\theta - \theta) + (I^{[n]} - T)(\theta)\|_n \\
&\leq \|\hat\theta - \theta\|_{\ell^2_m} + \delta_n,
\end{aligned}
$$

where δ_n^2 is defined as in (3.12). Consequently

$$
(E_{\mu^{[n]}}\|I^{[n]}\hat\theta - T\theta^{[n]}\|_n^2)^{1/2} \leq (E_{\mu^{[n]}}\|\hat\theta - \theta^{[n]}\|_{\ell^2_m}^2)^{1/2} + \delta_n. \qquad (3.17)
$$

Picking $\hat\theta = E(\theta^{[n]}|\tilde y)$ in (3.17) and exploiting (3.15) and (3.14) yields

$$
\sqrt{\tilde B}(\tilde\mu^{[n]}, \sigma) \leq \sqrt{\tilde M}(n, \mathcal{F}) + \delta_n.
$$

In combination with (3.13) and Lemma 3.1,

$$
\begin{aligned}
(\sqrt{\tilde M}(n, \mathcal{F}) + \delta_n)^2 &\geq \overline{B}(\mu^{[n]}, \epsilon_n)(1 + o(1)) \\
&\sim \overline{B}(\mu^{(\epsilon_n)}, \epsilon_n) \\
&\sim \overline{M}(\epsilon_n, \Theta) \\
&= M(\epsilon_n, \mathcal{F}),
\end{aligned}
$$

from which Theorem 1.1 finally follows.

4. Sampling Does not Make Things Harder

We now show how to use estimators in the white-noise model to construct estimators with nearly equal worst-case performance in the sampled-data model.

4.1. Deslauriers-Dubuc interpolation

Dubuc (1986) and Deslauriers and Dubuc ((1987) and (1989)) have proposed a method of interpolating sampled data $f(i/n)$, $i = 0, \ldots, n$ to produce a smooth function $\tilde{P}_n f(t)$, $t \in [0,1]$. The method is based on the use of local polynomial interpolation applied in a recursive multiscale fashion. We will avoid discussion of the details here, save only to say that the method defines a fundamental function $\tilde{\varphi}$ satisfying the interpolation conditions

$$\tilde{\varphi}(i) = \delta_{i0},$$

and such that

$$\int \tilde{\varphi}(t)dt = 1, \qquad \int t\tilde{\varphi}(t)dt = 0,$$

and $\tilde{\varphi}$ has $\tilde{R}$ continuous derivatives. Actually, the Deslauriers-Dubuc algorithm is a family of algorithms, indexed by a parameter $\tilde{D}$, the degree of local polynomial interpolation; by choice of $\tilde{D}$ sufficiently large we may arrange to make $\tilde{R} > \alpha$. We assume that $\tilde{D}$ has been chosen in this fashion.

The smooth interpolant $\tilde{P}_n f(t)$ is constructed as follows. The scaled fundamental functions $\tilde{\varphi}_i = \tilde{\varphi}(nt - i), i = 0, \ldots, n$ satisfy

$$\tilde{\varphi}_i(j/n) = 1_{\{i=j\}}, \quad 0 \le i, j \le n.$$

Let then

$$\tilde{P}_n f(t) = \sum_{i=0}^{n} f(i/n)\tilde{\varphi}_i(t)$$

be the resulting Deslauriers-Dubuc interpolant and define its wavelet coefficients

$$\tilde{\theta}_I = \langle \psi_I, \tilde{P}_n f \rangle \quad I \in \mathcal{I}.$$

How do these coefficients compare to the true wavelet coefficients of f? We have

Lemma 4.1. *Let $\alpha > 1/p$ and $1 \le p, q \le \infty$. Then*

$$\sup_{f \in \mathcal{F}} ||f - \tilde{P}_n f||_{L^2}^2 = O(n^{-r'}),$$

where

$$r' = \begin{cases} 2\alpha & p \ge 2 \\ 2(\alpha + 1/2 - 1/p) & p < 2 \end{cases},$$

and so $r' > r = \frac{2\alpha}{2\alpha+1}$.

It follows that the wavelet coefficients of $\tilde{P}_n f$, taken as a whole, will give an approximation to those of f in which the error is negligible compared to the order $O(n^{-r})$ of the minimax risk.

This remains true if we cut off scales beyond $j_0(n)$. Define the wavelet coefficients $\tilde\theta^{(n)}$ by

$$\tilde\theta_I^{(n)} = \begin{cases} \langle \psi_I, \tilde P_n f\rangle & j \le j_0(n;\alpha,p,q,\eta) \\ 0 & \text{else} \end{cases},$$

then

$$\sup_{\theta\in\Theta} ||\tilde\theta^{(n)} - \theta||_{l^2}^2 = O(n^{-r'}). \tag{4.1}$$

Important for us will also be the fact that Deslauriers-Dubuc interpolation is in some sense non-expansive for the (α,p,q) norm. To make this precise we need to discuss the wavelet coefficients with a scale cut-off.

Lemma 4.2. *Let $\alpha > 1/p$, $1 \le p,q \le \infty$, or $\alpha = p = q = 1$. Then*

$$||\tilde\theta^{(n)}||_{\mathbf{b}^\alpha_{p,q}} \le ||\theta||_{\mathbf{b}^\alpha_{p,q}}(1+\Delta_n(\alpha,p,q,\eta))$$
$$||\tilde\theta^{(n)}||_{\mathbf{f}^\alpha_{p,q}} \le ||\theta||_{\mathbf{f}^\alpha_{p,q}}(1+\Delta_n(\alpha,p,q,\eta)),$$

where

$$\Delta_n \to 0 \text{ as } n \to \infty.$$

In particular if we set $\Theta = \Theta(\alpha,p,q,C)$ and

$$C^{(n)} = \sup_{\theta\in\Theta} ||\tilde\theta^{(n)}||_{\mathbf{f}},$$

where $\mathbf{f} = \mathbf{b}^\alpha_{p,q}$ or $\mathbf{f}^\alpha_{p,q}$, then $C^{(n)} \to C$ as $n\to\infty$.

Our approach will use the observed noisy data $\tilde y_i$ from model (1.1) to create the noisy interpolant

$$\tilde y^{(n)}(t) = \sum_{i=0}^{n} \tilde y_i \tilde\varphi_i(t).$$

We shall need to relate the variance of the wavelet coefficients of $\tilde y^{(n)}$ to the benchmark $\epsilon_n^2 = \sigma^2 n^{-1}$, so define positive scalars λ_{In} by the relation

$$\mathrm{Var}\,\langle \tilde y^{(n)}, \psi_I\rangle = \lambda_{In}^2 \cdot \epsilon_n^2.$$

Lemma 4.3. *We have*

$$\lambda^{(n)} = \sup_{j\le j_0} \lambda_{In} \le 1 + o(1),$$

as $n\to\infty$.

4.2. The construction

We are now in a position to obtain an estimator from sampled data. We assume n, (α,p,q,C), γ, η, j_0 are specified, and the existence of constants $\Delta^{(n)}$, $\lambda^{(n)}$, and $(\lambda_{In})_I$ associated with the interpolation scheme are all available. We have the following steps.

1. *Preparation.* Obtain $\epsilon_I = \lambda_{In}\epsilon_n$, $\epsilon^{(n)} = \lambda^{(n)}\epsilon_n$ and $C^{(n)} = C(1+\Delta_n)$.
2. *Interpolation.* Using the observed noisy data $\tilde{y}_i$ from model (1.1), $i = 0, \ldots, n$, interpolate:
$$\tilde{y}^{(n)}(t) = \sum_{i=0}^{n} \tilde{y}_i \tilde{\varphi}_i(t)$$
where the $\tilde{\varphi}_i(t)$ are the Deslauriers-Dubuc fundamental functions.
3. *Inflation.* For all $I_{j,k}$ satisfying $j \le j_0(n)$, set
$$y_I^{(n)} = \langle \psi_I, \tilde{y}^{(n)} \rangle + \sqrt{(\epsilon^{(n)})^2 - \epsilon_I^2} \cdot \tilde{z}_I,$$
where the $\tilde{z}_I$ are i.i.d. $N(0,1)$ independent of the $\{y_i\}$. This process of adding noise 'inflates' the covariance matrix of the noise terms so that all diagonal entries are equal to $(\epsilon^{(n)})^2$, and we may write
$$y_I^{(n)} = \tilde{\theta}_I + \epsilon^{(n)} z_I^{(n)} \quad j \le j_0,$$
where $z_I^{(n)}$ are zero mean, unit variance, and jointly normally distributed, but may possibly be correlated.
4. *Estimation.* We apply, to all I with $j \le j_0$, the minimax-$\mathcal{E}$ family $(\delta_I(\cdot; \mathcal{E}, \Theta_n, \epsilon^{(n)})_I)$ where $\Theta_n = \Theta(\alpha, p, q, C^{(n)})$, getting
$$\hat{\theta}_I^{[n]} = \begin{cases} \delta_I(y_I^{(n)}) & j \le j_0 \\ 0 & \text{else} \end{cases}.$$
5. *Reconstruction.* We return to the original domain, getting
$$\hat{f}(t) = \sum_{j \le j_0} \hat{\theta}_I^{[n]} \psi_I.$$

4.3. Risk properties of the estimator

Given the substantial effort we have made in setting things up, the proof of Theorem 1.2 is easy. Let $\tilde{f}$ be the partial reconstruction defined in (1.25). We begin by writing $\hat{f} - f = (\hat{f} - \tilde{f}) + (\tilde{f} - f)$ and then decomposing

$$(\sup_{f \in \mathcal{F}} E||\hat{f} - f||^2_{L^2[0,1]})^{1/2} \le (\sup_{f \in \mathcal{F}} E||\hat{f} - \tilde{f}||^2)^{1/2} + (\sup_{f \in \mathcal{F}} ||\tilde{f} - f||^2)^{1/2}. \tag{4.2}$$

The first term on the RHS of (4.2) obeys

$$\begin{aligned} \sup_{f \in \mathcal{F}} E||\hat{f} - \tilde{f}||^2_{L^2[0,1]} &= \sup_{\theta \in \Theta} E||\hat{\theta}^{[n]} - \tilde{\theta}^{(n)}||^2_{\ell^2} && (4.3) \\ &\le \overline{M}_{\mathcal{E}}(\epsilon^{(n)}, C^{(n)}) && (4.4) \\ &\le (\epsilon^{(n)}/\epsilon_n)^2 (C^{(n)}/C)^2 \overline{M}_{\mathcal{E}}(\epsilon_n, C) && (4.5) \\ &\sim \overline{M}_{\mathcal{E}}(\epsilon_n, C) && (4.6) \\ &= M_{\mathcal{E}}(\epsilon_n, \mathcal{F}(C)). && (4.7) \end{aligned}$$

We explain the steps in more detail. Equality (4.3) is just the isometry property of the wavelet transform. (4.3) $\Rightarrow$ (4.4) makes use of the fact that $\{\tilde{\theta}^{(n)} : \theta \in \Theta(\alpha, p, q, ; C)\} \subset \Theta(\alpha, p, q, ; C^{(n)})$ and that all the estimators in question are constructed coordinatewise. (4.4) $\Rightarrow$ (4.5) uses Lemma 2.3. (4.5) $\Rightarrow$ (4.6) uses the crucial results of Lemmas 4.2 and Lemma 4.3. Finally (4.6) $\Rightarrow$ (4.7) is just the risk equivalence of white noise with sequence space.

The second term on the RHS of (4.2) can be handled by (4.1):

$$\sup_{f \in \mathcal{F}} ||\tilde{f} - f||^2_{L^2[0,1]} = o(n^{-r}),$$

where $r = 2\alpha/(2\alpha + 1)$.

Comparing the sizes of the two terms on the RHS of (4.2), and noting that $M_{\mathcal{E}}(\epsilon^{(n)}, C) \geq M(\epsilon^{(n)}, C) \geq c \cdot n^{-r}$ we conclude that the second is asymptotically negligible, and so

$$\begin{aligned} \tilde{M}_{\mathcal{E}}(n, \mathcal{F}) &\leq \sup_{f \in \mathcal{F}} E||\hat{f} - f||^2_{L^2[0,1]} \\ &\leq (\sup_{f \in \mathcal{F}} E||\hat{f} - \tilde{f}||^2)(1 + o(1)) \leq M_{\mathcal{E}}(\epsilon_n, \mathcal{F}(C))(1 + o(1)). \end{aligned}$$

This completes the proof of Theorem 1.2.

5. Proofs of Lemmas

We have completed the proof of all the announced results, modulo the lemmas they depend on. We now turn to those lemmas. It will turn out that all but one of the lemmas involve routine estimates. Lemma 4.2 requires a further development in a section of its own, and in some sense embodies the intellectual issue at the heart of the paper.

A word on notation: constants depending only on α, p, q and the choice of wavelet will be denoted by c, not necessarily the same at each occurrence. Uppercase C is reserved for the size of the norm ball $\Theta(C)$.

5.1. Proof of Lemma 2.2

First note that whether $\mathcal{F}$ corresponds to a Besov or Triebel sequence ball $\Theta(C)$,

$$\|\theta_{j\cdot}\|_p := (\sum_{I \in \mathcal{I}_j} |\theta_I|^p)^{1/p} \leq cC2^{-ja}, \qquad a = \alpha + 1/2 - 1/p. \tag{5.1}$$

(In the Besov case, this follows from definition (2.2) and the fact that ℓ_q norms decrease as q increases. The Triebel case follows from the Besov using (2.5).) Secondly, for $\mathbf{v} \in \mathcal{R}^m$,

$$||\mathbf{v}||_{l^2} \leq m^{(1/2-1/p)_+} ||\mathbf{v}||_{l^p}. \tag{5.2}$$

We conclude that, for $\alpha' = a - (1/2 - 1/p)_+$ and $K = (1 - 2^{-2\alpha'})^{-1}$,

$$\sum_{j'>j} \sum_k |\theta_{j',k}|^2 \le C^2 \sum_{j'>j} 2^{-2j'\alpha'} = KC^2 2^{-2j\alpha'}.$$

5.2. Proof of Lemma 2.3

The invariance relation

$$\overline{M}_{\mathcal{E}}(\epsilon, C) = r^2 \overline{M}_{\mathcal{E}}(\epsilon/r, C/r)$$

holds for all $\epsilon, C, r > 0$, as is easily seen by defining a new problem from (2.7) via $y' = y/r, \theta' = \theta/r$, etc., and noting that $\hat{\theta}(y) \in \mathcal{E}$ iff $r^{-1}\hat{\theta}(ry) \in \mathcal{E}$. Taking $\epsilon = \epsilon_1, C = C_1$ and $r = (\epsilon_1/\epsilon_0)(C_1/C_0)$, one sees that it suffices to show that $\overline{M}_{\mathcal{E}}(\epsilon, C)$ is increasing in C (trivial) and ϵ.

Given a scalar estimator $\eta(y) = \eta_I(y) \in \mathcal{E}$, we define

$$r(\mu, \eta, \epsilon) = E[\eta(\mu + \epsilon z) - \mu]^2, \qquad z \sim N(0,1).$$

Suppose that $\epsilon' < \epsilon$: to show that $\overline{M}_{\mathcal{E}}(\epsilon', C) \le \overline{M}_{\mathcal{E}}(\epsilon, C)$, it will suffice to exhibit, for each $\eta \in \mathcal{E}$, an estimator $\eta' \in \mathcal{E}$ such that

$$r(\mu, \eta', \epsilon') \le r(\mu, \eta, \epsilon) \qquad \text{for all } \mu. \tag{5.3}$$

For $\mathcal{E}_N$, set $y' = \mu + \epsilon' z$ and let $w \sim N(0, \epsilon^2 - (\epsilon')^2)$ be independent of z. Define $\eta'(y') = E[\eta(y' + w)|y']$: now (5.3) follows from Jensen's inequality.

The estimators in the classes $\mathcal{E}_L, \mathcal{E}_S, \mathcal{E}_H$ are all indexed by parameters λ_I, so we write $r(\mu, \lambda, \epsilon)$ for $E[\eta(\mu + \epsilon z, \lambda) - \mu]^2$. When $\epsilon = 1$, we write simply $r(\mu, \lambda)$. Now (5.3) will be satisfied if we exhibit a threshold modification $\lambda' = g(\rho)\lambda$ such that

$$\rho \mapsto r(\mu, g(\rho)\lambda, \rho) \qquad \text{is increasing for all } \lambda > 0, \mu \in \mathcal{R}. \tag{5.4}$$

Indeed, simply use $\eta'(y', \lambda')$ with $\lambda' = (g(\epsilon')/g(\epsilon))\lambda$.

For $\mathcal{E}_L$, $r(\mu, \lambda, \rho) = \lambda^2\rho^2 + (1-\lambda)^2\theta^2$, so the trivial choice $g(\rho) \equiv 1$ suffices for (5.4).

For $\mathcal{E}_S$, we take $g(\rho) = \rho$. From the invariance relation

$$r(\mu, \lambda, \epsilon) = \epsilon^2 r(\mu/\epsilon, \lambda/\epsilon), \tag{5.5}$$

we obtain

$$(\partial/\partial\rho)\ r(\mu, \rho\lambda, \rho) = (2\rho r - \mu r_\mu)(\mu/\rho, \lambda),$$

where, using the risk formulas for soft thresholding in Donoho and Johnstone ((1994), eq. A2.8), $r_\mu(\mu, \lambda) = 2\mu P_\mu\{|y| < \lambda\}$. Since $\rho > 1$, and

$$(2r - \mu r_\mu)(\mu, \lambda) = 2E[\eta(y, \lambda) - \mu]^2 I\{|y| > \lambda\} > 0,$$

we obtain (5.4).

For $\mathcal{E}_H$, it seems least inconvenient to take $g(\rho) = (\rho+1)/2$. Again using (5.5),

$$(\partial/\partial\rho)\ r(\mu,(1+\rho)\lambda/2,\rho) = (2\rho r - \mu r_\mu - (\lambda/2)r_\lambda)(\bar{\mu},\bar{\lambda}),$$

where r_μ, r_λ denote the corresponding partial derivatives of $r(\mu,\lambda)$ and $\bar{\mu} = \mu/\rho, \bar{\lambda} = (1+\rho)\lambda/2\rho$. Using the expression for $r(\mu,\lambda)$ in Donoho and Johnstone ((1994), eq. A2.2), one obtains

$$\begin{aligned} r_\lambda(\lambda,\mu) &= \phi(\lambda-\mu)\{\mu^2-(\lambda-\mu)^2\} + \phi(\lambda+\mu)\{\mu^2-(\lambda+\mu)^2\}, \\ r_\mu(\lambda,\mu) &= \phi(\lambda-\mu)\{(\lambda-\mu)^2-\mu^2\} + \phi(\lambda+\mu)\{\mu^2-(\lambda+\mu)^2\} \\ &\quad +2\mu\{\Phi(\lambda-\mu)-\Phi(-\lambda-\mu)\}. \end{aligned}$$

After some algebra, and defining $H(y) = \tilde{\Phi}(y) + y\phi(y) = \int_{-\infty}^{-y} z^2\phi(z)dz > 0$, we obtain for $\rho > 1$ and all (λ,μ):

$$\begin{aligned} &(2\rho r - \mu r_\mu - (\lambda/2)r_\lambda)(\mu,\lambda) \\ &= 2(\rho-1)\mu^2\{\Phi(\lambda-\mu)-\Phi(-\lambda-\mu)\} + 2\rho\{H(\lambda-\mu)+H(\lambda+\mu)\} \\ &\quad +\phi(\lambda-\mu)(\lambda-2\mu)^2\lambda/2 + \phi(\lambda+\mu)(\lambda+2\mu)^2\lambda/2 > 0. \end{aligned}$$

5.3. Completion of Proof of Lemma 3.1

We saw that $||\theta^{[n]} - \theta^{(\epsilon_n)}||_{l^2}^2 \le cC^2 \cdot n^{-r'}$ where $r' > r$. Jensen's inequality gives

$$||E\{\theta^{[n]}|y\} - E\{\theta^{(\epsilon_n)}|y\}||^2 \le E\{||\theta^{[n]} - \theta^{(\epsilon_n)}||_{l^2}^2|y\} \le cC^2n^{-r'},$$

and so

$$E||E\{\theta^{[n]}|y\} - \theta^{[n]}||^2 \le ((E||E\{\theta^{(\epsilon_n)}|y\} - \theta^{(\epsilon_n)}||_{l^2}^2)^{1/2} + cCn^{-r'/2})^2$$

Symmetrically,

$$E||E\{\theta^{(\epsilon_n)}|y\} - \theta^{(\epsilon_n)}||^2 \le ((E||E\{\theta^{[n]}|y\} - \theta^{[n]}||_{l^2}^2)^{1/2} + cCn^{-r'/2})^2,$$

as $\overline{B}(\mu^{(\epsilon_n)},\epsilon_n) \le C^{2(1-r)}n^{-r}$, where $r < r'$. The lemma follows.

5.4. Proof of Lemma 3.2

Part a. As at (3.7), for $\theta \in \mathcal{R}^m$, let $f = \sum_{|I|\ge 2^{-j_0}} \theta_I\psi_I$ and $T\theta = (f(t_k))_0^n$. The partial isometry $I^{(n)}$ is given by projection onto an orthonormal basis $\{\phi_{nk}\}_{k=0}^n$ for $L^2[0,1]$: $I^{(n)}\theta = (\sqrt{n}\langle f,\phi_{nk}\rangle)_0^n$. To construct this basis, let $\phi(t)$ be a $C^{\bar{\alpha}}$ scaling function supported on $[0,k_0]$ such that $\{\phi(t-k), k\in\mathcal{Z}\}$ is an orthonormal set in $L^2(\mathcal{R})$ and $\int t^k\phi(t)dt = \delta_{k0}$ for $0 \le k \le \lfloor\bar{\alpha}\rfloor$. Then set $\phi_{nk}(t) = \sqrt{n}\phi(nt -$

k). For $0 \le k < k_0$ and $n - k_0 < k \le n$, the ϕ_{nk} are not orthonormal in $L^2[0,1]$, but can be made so by Gram-Schmidt orthogonalization while retaining support in $[0,1]$ in the intervals $[0, c \cdot k_0/n]$ and $[1 - c \cdot k_0/n, 1]$ respectively, for a constant c not depending on n.

For those ϕ_{nk} supported entirely in $(0,1)$, a standard argument using $\bar{\alpha}-$Hölder smoothness of f and $\lfloor \bar{\alpha} \rfloor$ vanishing moments of ϕ gives

$$|f(t_k) - \sqrt{n}\langle f, \phi_{nk}\rangle| \le c\|f\|_{\dot{C}^{\bar{\alpha}}} n^{-\bar{\alpha}} \tag{5.6}$$

where c depends on ϕ only. Using supremum norm bounds for the boundary ϕ_{jk},

$$\begin{aligned} \|(T - I^{(n)})\theta\|_n^2 &= n^{-1} \sum_0^n [f(t_k) - \sqrt{n}\langle f, \phi_{nk}\rangle]^2 \\ &\le 2c^2 \|f\|_\infty^2 k_0 n^{-1} + c^2 \|f\|_{\dot{C}^{\bar{\alpha}}}^2 n^{-2\bar{\alpha}}. \end{aligned} \tag{5.7}$$

Now (5.1) shows that $\theta \in \Theta$ implies $|\theta_{jk}| \le \|\theta_{j\cdot}\|_p \le cC 2^{-aj}$, and hence

$$\begin{aligned} \|f\|_{C^{\bar{\alpha}}} &\le c \sup_{j \le j_0, k} 2^{(\bar{\alpha}+1/2)} |\theta_{jk}| \\ &\le cC \sup_{j \le j_0} 2^{(\bar{\alpha}+1/2-a)j} \le cCn^{[\bar{\alpha}-(\alpha-1/p)]\lambda}, \end{aligned}$$

since $a - 1/2 = \alpha - 1/p$, so long as $\bar{\alpha} > (\alpha - 1/p)$. Also

$$|f(t)| \le \sum |\theta_I||\psi_I(t)| \le cC \sum_{j \le j_0} 2^{-aj} 2^{j/2},$$

and hence

$$\|f\|_\infty \le \begin{cases} cC & \text{if } \alpha > 1/p \\ cC \log n & \text{if } \alpha = 1/p. \end{cases}$$

Substituting these embedding bounds into (5.7) yields

$$\delta_n^2 \le cC^2 n^{2[\bar{\alpha}-(\alpha-1/p)]\lambda - 2\bar{\alpha}} + O(n^{-1} \log^2 n).$$

It is now easy to verify that $\delta_n^2 = o(n^{-2\alpha/(2\alpha+1)})$ so long as $\bar{\alpha} > 1/2$ (if $p \ge 2$) and if $\bar{\alpha} > \alpha$ (if $p < 2$).

Note: Were we concerned only with the case $\alpha > 1/p$, a much shorter proof could be based on building $I^{(n)}$ from the singular value decomposition of T and the inequalities

$$\delta_n^2 \le c\|T - I^{(n)}\|_2^2 \le c\|T^*T - I\|_2^2 \le c(2^{j_0}/n)^2,$$

where the last bound is proved below. However, when $\alpha = p = 1$, negligibility of this bound relative to n^{-r} is incompatible with the corresponding negligibility in (3.5).

Part b. We show that

$$\lambda_{max}(T^*T) = \|T^*T\|_2 \le 1 + O(2^{j_0}/n). \tag{5.8}$$

The matrix $(t_{II'} : I, I' \in \mathcal{I}_{\le j_0})$ representation of T^*T in the basis $(\psi_I, I \in \mathcal{I}_{\le j_0})$ is given by

$$t_{II'} = \langle \psi_I, \psi_{I'} \rangle_n = (1/n) \sum_i \psi_I(t_i)\psi_{I'}(t_i).$$

The weighted form of Schur's lemma (e.g. Meyer (1990), Vol. 2, Sec 8.4) states, in the case of a symmetric matrix $M = (m_{II'})$, that if there exist positive weights w_I such that for all indices I

$$\sum_{I'} |m_{II'}| w_{I'} \le A w_I$$

then $\|M\|_2 \le A$. Thus (5.8) will follow if we show that

$$\sum_{I'} |t_{II'}| 2^{-j'/2} \le 2^{-j/2}(1 + c2^{j_0}/n) \qquad \forall I \in \mathcal{I}^{j_0}. \tag{5.9}$$

The next lemma gives elementwise bounds on the distance of T^*T from the identity matrix.

Lemma 5.1. *Let* $\chi(I, I') = 1$ *if* supp $\psi_I \cap$ supp $\psi_{I'} \ne \emptyset$ *and* 0 *otherwise. If* ψ *is Lipschitz and has compact support, then*

$$|\langle \psi_I, \psi_{I'} \rangle_n - \delta_{II'}| \le c 2^{(j+j')/2}/n \cdot \chi(I, I'). \tag{5.10}$$

Assuming the truth of the lemma for a moment, we have from (5.10),

$$\sum_{I'} |t_{II'}| 2^{-j'/2} \le 2^{-j/2}(1 + c2^j/n) + (c/n) S_I,$$

where $S_I = \sum_{I' \ne I} 2^{j/2} \chi(I, I')$. Since $\sum_{I' \in \mathcal{I}_{j'}} \chi(I, I') \le c_2 2^{(j'-j)_+}$,

$$S_I \le 2^{j/2} \sum_{j' \le j_0} c_2 2^{(j'-j)_+} \le c_3 2^{j_0} 2^{-j/2},$$

which establishes (5.9).

Finally we pass to the proof of the Lemma. Since ψ is Lipschitz,

$$|\psi_I(s) - \psi_I(t)| \le c \cdot 2^{j/2} 2^j |s - t|, \tag{5.11}$$

(whether ψ_I is a boundary or an interior wavelet) and so

$$\begin{aligned} |\psi_I \psi_{I'}(s) - \psi_I \psi_{I'}(t)| &\le |\psi_I(s) - \psi_I(t)| \|\psi_{I'}\|_\infty + \|\psi_I\|_\infty |\psi_{I'}(s) - \psi_{I'}(t)| \\ &\le c 2^{(j+j')/2} 2^{j \vee j'} |s - t|. \end{aligned}$$

Now if $I_i = [t_i, t_{i+1})$, and $\|f\|_L$ denotes the Lipschitz constant of f,then

$$|n^{-1}f(t_i) - \int_{I_i} f| \le \int_{I_i} \|f\|_L(t-t_i)dt = (1/2)n^{-2}\|f\|_L,$$

which implies that

$$|n^{-1}\sum_i \psi_I\psi_{I'}(t_i) - \int \psi_I\psi_{I'}| \le (1/2)cn^{-2}\cdot 2^{(j+j')/2}\cdot 2^{j\vee j'}\cdot \#\{i\in \; supp\; \psi_I\cap\psi_{I'}\}$$
$$= cn^{-1}2^{(j+j')/2}. \qquad (5.12)$$

5.5. Proof of Lemma 3.3

We first show how to pass from model (3.8), using randomization, to a version of model (3.7) with a slightly modified variance. Apply the sampling operator to (3.8) to obtain $T\bar{y} = T\theta + \epsilon_n T\bar{z}$. The variance of each component $(\epsilon_n T\bar{z})_i$ is bounded by

$$\begin{aligned}\sup_{\|\xi\|_{\ell^2_n}=1} \mathrm{Var}\,(\epsilon_n \xi^t T\bar{z}) &= \sigma^2 \sup_\xi \mathrm{Var}\,\langle \xi, T\bar{z}\rangle_n / \langle \xi,\xi\rangle_n \\ &= \sigma^2 \sup_\xi \; \|T^*\xi\|^2_{\ell^2_m}/\|\xi\|^2_n = \sigma\|T^*\|^2 \\ &= \sigma^2\lambda_{max}(TT^*) = \sigma^2\lambda_{max}(T^*T) = \sigma^2\lambda^2, \quad \text{say,}\end{aligned}$$

where $\langle\cdot,\cdot\rangle_n$ is the inner product corresponding to $\|\cdot\|_n$ and, again, T is regarded as an operator from $(\mathcal{R}^m, \|\cdot\|_{\ell^2_m})$ to $(\mathcal{R}^{n'}, \|\cdot\|_n)$. Hence, we may choose a Gaussian vector $\tilde{w}$ in $\mathcal{R}^{n'}$, independent of $\bar{z}$, so that

$$\epsilon_n T\bar{z} + \tilde{w} \sim N(0, \sigma^2\lambda^2 I_{n'}).$$

Thus $T\bar{y} + \tilde{w} = T\theta + \epsilon_n T\bar{z} + \tilde{w} \overset{\mathcal{D}}{=} T\theta + \lambda\sigma\tilde{z}$.

Consider now observations $\tilde{y} = \xi + \lambda\sigma\tilde{z}$, and for a prior distribution $\tilde{\mu}(d\xi)$ on ξ, let $\hat{\xi}_{\tilde{\mu},\lambda\sigma}$ denote the corresponding Bayes estimator. In particular, let $\tilde{\mu}$ correspond to $T\theta^{[n]}$, where $\theta^{[n]}$ is the random variable constructed in Lemma 3.1. Define a randomized estimator for θ in model (3.8) at noise level ϵ_n by

$$\hat{\theta}(\bar{y}, \tilde{w}) = I^{[n]*}\hat{\xi}_{\tilde{\mu},\lambda\sigma}(T\bar{y} + \tilde{w}).$$

Then using the partial isometry property of $I^{[n]}$,

$$\begin{aligned}\overline{B}(\mu^{[n]}, \epsilon_n) &\le E_{\mu^{[n]}}\|\hat{\theta}(\bar{y},\tilde{w}) - \theta^{[n]}\|^2_{\ell^2_m} \\ &\le E_{\mu^{[n]}}\|\hat{\xi}_{\tilde{\mu},\lambda\sigma}(T\theta + \lambda\sigma\tilde{z}) - T\theta^{[n]} + (T - I^{[n]})\theta^{[n]}\|^2_n \\ &\le (\sqrt{\tilde{B}}(\tilde{\mu}^{[n]}, \lambda\sigma) + \delta)^2.\end{aligned}$$

We now use some standard properties of Bayes rules: set $\nu = \tilde{\mu}^{[n]}$. Then

$$\tilde{B}(\nu, \lambda\sigma) \le \tilde{B}(\nu, \sigma) \qquad \text{if } \lambda \le 1, \tag{5.13}$$

$$\tilde{B}(\nu, \lambda\sigma) = \lambda^2 \tilde{B}(\nu_{\lambda^{-1}}, \sigma) \le \lambda^2 \tilde{B}(\nu, \sigma), \qquad \text{if } \lambda \ge 1, \tag{5.14}$$

where $\nu_a(d\xi)$ denotes the scaled prior $\nu(d\xi/a)$ and the final inequality follows from properties of Fisher information. Combining (5.13) and (5.14) yields

$$\tilde{B}(\tilde{\mu}^{[n]}, \lambda\sigma) \le (\lambda^2 \vee 1)\tilde{B}(\tilde{\mu}^{[n]}, \sigma),$$

and hence the conclusion of Lemma 3.3.

5.6. Proof of Lemma 4.1

It will be convenient to work in $L^2(\mathcal{R})$ so as to use results on interpolating wavelet transforms on $\mathcal{R}$ from Donoho (1992b). We extend $f = \sum \theta_I \psi_I \in \mathcal{F}$ (defined on $[0,1]$) to $\mathcal{R}$ using (2.1): $f = \sum \theta'_I \psi^0_I$ where $\theta'_I = \theta_I$ unless $\text{supp}\psi_I \cup [0,1]^c$ is non-empty. Hence $f \in B^{\alpha}_{p,q}[0,1]$ implies $f \in B^{\alpha}_{p,q}(\mathcal{R})$ (and the embedding has bounded norm). Let $e_n(\mathcal{F}) = \sup_{\theta\in\Theta} \|f - \tilde{P}_n f\|_{L^2(\mathcal{R})}$: simple scaling arguments show that if $2^J \le n < 2^{J+1}$, then $e_n(\mathcal{F}) \le c e_{2^J}(\mathcal{F})$, so that it suffices to take $n = 2^J$. Donoho (1992b) constructs an interpolating multiresolution with detail spaces W_j, wavelets $\tilde{\psi}_{jk}$ and associated (non-orthogonal) projections $\tilde{Q}_j$. Thus $f - \tilde{P}_n f = \sum_{j\ge J} \tilde{Q}_j f$, with $\tilde{Q}_j f = \sum_k \tilde{\theta}_{jk}\tilde{\psi}_{jk}$. Donoho ((1995), Theorem 2.7) shows that $f \in B^{\alpha}_{p,q}$ (resp. $F^{\alpha}_{p,q}$) iff $\tilde{\theta} \in \mathbf{b}^{\alpha}_{p,q}$ (resp $\mathbf{f}^{\alpha}_{p,q}$)and his Lemma 7.3 shows that $\|\tilde{Q}_j f\|_{L^2} \le c\|\tilde{\theta}_{j\cdot}\|_{\ell^2}$. Putting all this together with (5.2) and (5.1),

$$\begin{aligned} e_n(\mathcal{F}) &\le \sup_{\mathcal{F}} \sum_{j\ge J} \|\tilde{Q}_j f\|_{L^2} \le c \sup_{\Theta} \sum_{j\ge J} \|\tilde{\theta}_{j\cdot}\|_2 \\ &\le cC \sum_{j\ge J} 2^{j(1/2-1/p)_+} 2^{-ja} \le cC 2^{-J\alpha'} = cCn^{-\alpha'}, \end{aligned}$$

where $\alpha' = a - (1/2 - 1/p)_+ = \alpha$ if $p \ge 2$ and $\alpha' = \alpha + 1/2 - 1/p$ if $p \le 2$. In either case $\alpha' > \alpha/(2\alpha+1)$ so that on setting $r' = 2\alpha'$, we have $e^2_n(\mathcal{F}) = O(n^{-r'}) = o(n^{-r})$.

We note that the cited norm equivalence result Theorem 2.7 is proved only for $\alpha > 1/p$, but remarks on the critical case $\alpha = 1/p$ in Section 6 in that paper show that the argument above in fact extends – with additional bookkeeping – to cover the Bump Algebra $\mathbf{b}^1_{1,1}$.

5.7. Proof of Lemma 4.3

We have, on setting $\xi_i = \sqrt{n}\tilde{\varphi}_i$,

$$\text{Var}\,\langle \tilde{y}^{(n)}, \psi_I\rangle = \sigma^2 \sum_i \langle \tilde{\varphi}_i, \psi_I\rangle^2 = \epsilon^2_n \sum_i \langle \xi_i, \psi_I\rangle^2. \tag{5.15}$$

Since $\int \xi_i(t)dt = 1/\sqrt{n}$, and $\|\psi_I\|_L \leq c2^{3j/2}$, we have from (5.6), since $j_1 = \log_2(n)$,

$$|\langle\psi_I, \xi_i\rangle - \psi_I(i/n)/\sqrt{n}| \leq c\, 2^{3/2(j-j_1)}.$$

As a result,

$$\lambda_{In} = \Big(\sum_{i=0}^{n}\langle\psi_I, \xi_i\rangle^2\Big)^{1/2} \leq \Big(n^{-1}\sum_{i=0}^{n}\psi_I^2(i/n)\Big)^{1/2} + \Big(c^2\sum_{i=0}^{n}2^{3(j-j_1)}1_{S_I}(i)\Big)^{1/2}, \tag{5.16}$$

where $S_I = \{i : \mathrm{supp}\psi_I \cup \mathrm{supp}\xi_i \neq \emptyset\}$. Evidently, for $0 \leq j \leq j_0$,

$$\sum_{i=0}^{n}2^{3(j-j_1)}1_{S_I}(i) \leq c2^{2(j-j_1)} \leq c2^{2(j_0-j_1)}.$$

From (5.12), $n^{-1}\sum\psi_I^2(i/n) \leq 1 + cn^{-1}2^j \leq 1 + c2^{j_0-j_1}$. Hence, since $j_0 = \lambda j_1$ and $2^{j_1} = n$,

$$\sup_{j\leq j_0}\lambda_{In} \leq 1 + c2^{j_0-j_1} = 1 + o(1).$$

6. Precision of Empirical Wavelet Coefficients

Let again $\tilde\theta^{(n)}$ denote the sequence of empirical wavelet coefficients $\tilde\theta_I^{(n)} = \langle\psi_I, \tilde P_n f\rangle 1_{\{|I|\geq 2^{-j_0}\}}$. As $f = \sum\theta_J\psi_J$, we may write

$$\tilde\theta_I^{(n)} = \sum_J \theta_J\langle\psi_I, \tilde P_n\psi_J\rangle := \sum_J T_{I,J}^{(n)}\theta_J,$$

here $T_{I,J}^{(n)} = \langle\psi_I, \tilde P_n\psi_J\rangle\; 1_{\{|I|\geq 2^{-j_0}\}}$. Thinking of $T^{(n)} = (T_{I,J}^{(n)} \;:\; I, J \in \mathcal{I})$ as a linear operator on sequences $(\theta_J \;: J \in \mathcal{I})$ we have

$$\tilde\theta^{(n)} = T^{(n)}\theta.$$

Define the partial identity operator

$$I_{I,J}^{(n)} = \begin{cases} 1_{\{I=J\}} & |I| \geq 2^{-j_0} \\ 0 & |I| < 2^{-j_0} \end{cases}.$$

Our principal result on the accuracy of $\tilde\theta^{(n)}$ can be stated as follows:

Theorem 6.1. *Let $\alpha > 1/p$ and $1 \leq p, q \leq \infty$, or let $\alpha = p = q = 1$. Let $\tilde D \geq \tilde R > R > \alpha$. Then $T^{(n)}$ is a bounded operator on $\mathbf{f} = \mathbf{b}_{p,q}^{\alpha}$ or $\mathbf{f}_{p,q}^{\alpha}$ and*

$$\Delta_n = \Delta_n(\alpha, p, q) = \|T^{(n)} - I^{(n)}\|_{(\mathbf{f},\mathbf{f})} \to 0,$$

as $n \to \infty$.

Corollary 6.2. *Let $\alpha > 1/p$ and $1 \le p, q \le \infty$, or let $\alpha = p = q = 1$. Then*

$$\|\tilde{\theta}^{(n)}\|_{\mathbf{f}} \le \|\theta\|_{\mathbf{f}} \cdot (1 + \Delta_n),$$

where $\Delta_n \to 0$ as $n \to \infty$.

Lemma 4.2 follows immediately.

Our method of proof for Theorem 6.1 is based on interpolation from various bounding cases. Details for the outline below follow in succeeding subsections.

1°. Let $U^{(n)} = T^{(n)} - I^{(n)}$: crucially, $U^{(n)}$ is non-trivial on at most $\log_2 n$ resolution levels:

$$(U^{(n)}\theta)_I = 0 \qquad \text{for } |I| < 2^{-j_0(\alpha,p,q,n)}. \tag{6.17}$$

In the case $\alpha > 1/p$, a crude bound then suffices to reduce to the case $q = \infty$:

$$\max(\|U^{(n)}\|_{\mathbf{b}^\alpha_{p,q}}, \|U^{(n)}\|_{\mathbf{f}^\alpha_{p,q}}) \le c(\log n)^{1/q} \|U\|_{\mathbf{b}^\alpha_{p,\infty}}. \tag{6.18}$$

When $\alpha = p = q = 1$, we give a separate bound for $\|U\|_{\mathbf{b}^1_{1,1}} (= \|U\|_{\mathbf{f}^1_{1,1}})$.

2°. We recall for the reader's convenience that

$$\begin{aligned} \|U^{(n)}\|_{\mathbf{b}^\alpha_{p,q}} &= \sup \|U^{(n)}\theta\|_{\mathbf{b}^\alpha_{p,q}} \,/\, \|\theta\|_{\mathbf{b}^\alpha_{p,q}} \\ &= \sup \|\tilde{U}\xi\|_{\ell^q(\ell^p)} / \|\xi\|_{\ell^q(\ell^p)} := N_{pq}(\tilde{U}). \end{aligned}$$

Here $\tilde{U}_{IJ} = 2^{a(i-j)} U^{(n)}_{IJ}$ and $\ell^q(\ell^p)$ denotes the besov space $b^0_{p,q}$; equivalently

$$\|\theta\|_{\ell^q(\ell^p)} = \Big(\sum_j \Big(\sum_{k=0}^{2^j-1} |\theta_{j,k}|^p\Big)^{q/p}\Big)^{1/q}$$

(with the usual modification if $p = \infty$ or $q = \infty$).

We emphasize that $\tilde{U} = \tilde{U}^{(n)}(a, \lambda)$ depends on (α, p) through $a = \alpha + 1/2 - 1/p$ and $\lambda = \lambda(\alpha, p)$ defined at (1.23).

3°. We show that $\tilde{U}$ is bounded on $\ell^\infty(\ell^p)$ by interpolation from the endpoints $p = 1$ and ∞:

$$N_{p\infty}(\tilde{U}) \le \max\{N_{1\infty}(\tilde{U}), N_{\infty\infty}(\tilde{U})\}. \tag{6.19}$$

For the extreme cases $(\bar{p}, \bar{q}) = (1,1), (1,\infty)$ and (∞,∞), we establish bounds

$$N_{\bar{p}\bar{q}}(\tilde{U}^{(n)}(a,\lambda)) \le c n^{-\Delta_{\bar{p}\bar{q}}(a,\lambda)}, \tag{6.20}$$

where the values of $\Delta_{\bar{p}\bar{q}}$ are summarized in the table below. Using $a = \alpha + 1/2 - 1/p$, we have also shown in the table sufficient conditions for positivity of $\Delta_{\bar{p}\bar{q}}$. Note that the values of $a = a(\alpha, p)$ and $\lambda = \lambda(\alpha, p)$ remain fixed in computation of $\Delta_{\bar{p}\bar{q}}(a, \lambda)$ over the boundary values of $(\bar{p}, \bar{q})$.

Quantity	Lower Bound	Validity
$\Delta^{\alpha}_{1,1}$	$(1-\lambda)a$	$\alpha \geq \frac{1}{p}$
$\Delta^{\alpha}_{1,\infty}$	$(1-\lambda)a$	$\alpha > \frac{1}{p}$
$\Delta^{\alpha}_{\infty,\infty}$	$(1-\lambda)(a-\frac{1}{2})$	$\alpha > \frac{1}{p}$

4°. Assembling the results of steps 1° through 3°, we obtain, if $\alpha > 1/p$,

$$\|U^{(n)}\|_{\mathbf{f}} \leq c(\log n)^{1/q} n^{-(1-\lambda)(\alpha-1/p)} \to 0, \quad \text{as } n \to \infty,$$

while if $\alpha = p = q = 1$,

$$\|U^{(n)}\|_{\mathbf{b}} \leq cn^{-(1-\lambda)a} \to 0.$$

This completes the outline for the proof of Theorem 6.1. Note, however, that nothing is said about $\alpha = p = 1, q > 1$. On other grounds we know that $N^1_{1,q}$ does not converge to 0 if $q > 1$. In this sense our results are complete and maximal.

6.1. Norm Bounds

Proof of (6.18). Indeed, if $\theta_I = 0$ for $|I| < 2^{-j_0}$, then

$$\begin{aligned}\|\theta\|^q_{\mathbf{b}^{\alpha}_{p,q}} &\leq \sum_{j \leq j_0(n)} (2^{aj}\|\theta_{j\cdot}\|_p)^q \qquad\qquad (a = \alpha + 1/2 - 1/p)\\ &\leq j_0(n)[\max_j 2^{aj}\|\theta_{j\cdot}\|_p]^q \leq (\log n)\|\theta\|_{\mathbf{b}^{\alpha}_{p,\infty}}.\end{aligned}$$

Since ℓ^q norms decrease as q increases, (2.5) implies that

$$\begin{aligned}\|U^{(n)}\theta\|_{\mathbf{f}^{\alpha}_{p,q}}/\|\theta\|_{\mathbf{f}^{\alpha}_{p,q}} &\leq (a_1/a_0)\|U^{(n)}\theta\|_{\mathbf{b}^{\alpha}_{p,p\wedge q}}/\|\theta\|_{\mathbf{b}^{\alpha}_{p,p\vee q}}\\ &\leq (a_1/a_0)(\log n)^{1/q}\|U^{(n)}\theta\|_{\mathbf{b}^{\alpha}_{p,\infty}}/\|\theta\|_{\mathbf{b}^{\alpha}_{p,\infty}}.\end{aligned}$$

The argument for $\mathbf{f} = \mathbf{b}^{\alpha}_{p,q}$ is similar, but simpler.

Proof of (6.19). This result can be obtained by mechanical application of interpolation theory - see, e.g., DeVore and Lorentz ((1993), Chapter 7) (DL below) for details and definitions. The tool is a version of the Riesz-Thorin theorem (DL Theorem 7.1): If U is a bounded operator on Banach spaces $X_i, i = 0, 1$, with norm M_i, then U is also bounded on the interpolation space $X_{\theta,q} = (X_0, X_1)_{\theta,q}$ with norm $\leq M_0^{1-\theta} M_1^{\theta} \leq \max(M_0, M_1)$. In our case, since $\ell^p = (\ell^1, \ell^{\infty})_{1-1/p,p}$, $\ell^{\infty}(\ell^p)$ is also an interpolation space and DL's equation(7.21) gives

$$\ell^{\infty}(\ell^p) = (\ell^{\infty}(\ell^1), \ell^{\infty}(\ell^{\infty}))_{1-1/p,p},$$

and hence $N_{p\infty} \leq \max\{N_{1\infty}, N_{\infty\infty}\}$.

6.2. Proof of (6.20)

Our development of these bounds will use various notational conventions. Both I and J will denote dyadic intervals; we use i as a scale parameter for interval I in the same way we used j before, ℓ as a location parameter in the same way we used k before; we use j and k as scale and location parameters associated with intervals labelled J. Sums over I will now typically run over scales $i \leq j_0$ only; while sums over J will typically run over all scales j. Let $j_1 = \log_2(n)$; j_1 is not necessarily integral; we let $\tilde{j}_1 = \lfloor j_1 \rfloor$, which is integral.

We note also the basic facts

$$N_{11} = \|\tilde{U}\|_{\ell^1(\ell^1)} = \sup_J \sum_I |\tilde{U}_{IJ}| \tag{6.21}$$

$$N_{\infty\infty} = \|\tilde{U}\|_{\ell^\infty(\ell^\infty)} = \sup_I \sum_J |\tilde{U}_{IJ}| \tag{6.22}$$

$$N_{1\infty} = \|\tilde{U}\|_{\ell^\infty(\ell^1)} \leq \sup_i \sum_j \sup_k \sum_\ell |\tilde{U}_{I,J}| \tag{6.23}$$

In establishing our bounds, we use four key inequalities, proved in this section.

Lemma 6.3. *For $j \leq j_1$,*

$$|\langle \psi_I, \psi_J - \tilde{P}_n \psi_J \rangle| \leq c \cdot 2^{(j-j_1)R}. \tag{6.24}$$

For $j > j_1$,

$$|\langle \psi_I, \tilde{P}_n \psi_J \rangle| \leq c \cdot 2^{-j_1} \cdot 2^{(i+j)/2} \quad 1_{\{cJ \cap \mathbf{Z}/2^{j_1} \neq \emptyset\}} \cdot 1_{\{cJ \cap cI \neq \emptyset\}}. \tag{6.25}$$

Lemma 6.4. *Let $w_J = \psi_J - \tilde{P}_n \psi_J$. Then if $j \leq j_1$,*

$$\sum_{I \in \mathcal{I}_i} |\langle \psi_I, w_J \rangle| \leq c 2^{(j-j_1)R} \cdot 2^{(i-j)_+/2}. \tag{6.26}$$

$$\sum_{J \in \mathcal{I}_j} |\langle \psi_I, w_J \rangle| \leq c 2^{(j-j_1)R} \cdot 2^{(j-i)_+/2}. \tag{6.27}$$

Proof of (6.24). By the Cauchy-Schwartz inequality, it suffices to bound $\|(I - \tilde{P}_n)\psi_J\|_2$. Although $\|\cdot\|_2$ here refers to the norm on $L^2([0,1])$, it is of course bounded by the norm of $L^2(\mathcal{R})$. Donoho ((1992b), Lemma 7.2) studied the norms of $I - \tilde{P}_n$ on $L^2(\mathcal{R})$ under restriction to $\bar{V}_j = \text{span}\{\psi^0_{jk}, k \in \mathcal{Z}\}$, obtaining

$$\|I - \tilde{P}_n\|_{\bar{V}_j} \leq c 2^{(j-j_1)\min(\tilde{D},R)}.$$

In particular, since $\tilde{D} > R$, and using relations (2.1) for the boundary wavelets, we obtain

$$\|(I - \tilde{P}_n)\psi_J\|_2 \leq c 2^{(j-j_1)R}. \tag{6.28}$$

Proof of (6.25). Now

$$\langle \psi_I, \tilde{P}_n \psi_J \rangle = \sum_u \psi_J(u/n) \langle \psi_I, \tilde{\phi}_u \rangle,$$

and for such u that $u/n \in cI$,

$$\langle \psi_I, \tilde{\phi}_u \rangle \leq \|\psi_I\|_\infty \|\tilde{\phi}_u\|_1 \leq c 2^{i/2 - j_1}.$$

Since $\psi_J(u/n) \leq c2^{j/2}$ if $u/n \in cJ$, we obtain

$$|\langle \psi_I, \tilde{P}_n \psi_J \rangle| \leq c 2^{(i+j)/2} 2^{-j_1} \sum_{u=0}^{n} 1_{\{u/n \in cJ \cap cI\}}.$$

Proof of (6.26). Now

$$|\langle \psi_I, w_J \rangle| \leq \|\psi_I\|_{L^2(cI)} \|w_J\|_{L^2(cI)}$$

and $\|\psi_I\|_{L^2(cI)} = 1$, and since $\#\{\ell \;:\; cI \cap \operatorname{supp} w_J \neq \emptyset\} = O(2^{(i-j)_+})$, we have

$$\begin{aligned}
\sum_{I \in \mathcal{I}_i} |\langle \psi_I, w_J \rangle| &\leq \sum_{I \in \mathcal{I}_i} \|w_J\|_{L^2(cI)} \\
&\leq c 2^{(i-j)/2} \quad \big(\sum \|w_J\|^2_{L^2(cI)}\big)^{1/2} \\
&= c \|w_J\|_{L^2(\mathcal{R})} \; 2^{(i-j)/2}.
\end{aligned}$$

From (6.28) we get $\|w_J\|_{L^2} \leq c 2^{(j-j_1)R}$.

Proof of (6.27). The argument is essentially the same as for (6.26):

$$\begin{aligned}
\sum_J |\langle \psi_I, w_J \rangle| &\leq \sum_J \|\psi_I\|_{L^2(cJ)} \cdot c 2^{(j-j_1)R} \\
&\leq c 2^{(j-j_1)R} \; [\; 2^{(j-i)_+} \sum_J \|\psi_I\|^2_{L^2(cJ)} \;]^{1/2} \\
&= c 2^{(j-j_1)R} \; 2^{(j-i)_+/2}.
\end{aligned}$$

We may summarize the consequences of Lemmas 6.3 and 6.4 as follows.

(i) Uniformly in $k = 1, \ldots, 2^j$,

$$\sum_l |\langle \psi_I, w_J \rangle| \leq \begin{cases} c 2^{(j-j_1)R} 2^{(i-j)_+/2} & j \leq j_1 \\ c 2^{-j_1} 2^{(i+j)/2} & j > j_1 \end{cases}. \tag{6.29}$$

(ii) Uniformly in $l = 1, \ldots, 2^i$,

$$\sum_k |\langle \psi_I, w_J \rangle| \leq c 2^{(j_1 - j)_+ R} 2^{(j-i)_+/2}. \tag{6.30}$$

6.2.1. Estimates for (1, 1)

We are assuming that $\tilde{R} > R > 1 + \alpha$. Now

$$N_{11} \le \sup_j S_j, \qquad S_j = \sup_{k \in \mathcal{I}_j} \sum_{i \le j_0} 2^{a(i-j)} \sum_l |\langle \psi_I, w_J \rangle|.$$

If $j \le j_1$, from (6.29),

$$\begin{aligned} S_j &\le c \sum_{i \le j_0} 2^{a(i-j)} 2^{(j-j_1)R} 2^{(i-j)_+/2} \\ &\le c 2^{a(j-j_0)+(j-j_1)R+(j_0-j)_+/2} = c2^{\ell(j)}, \quad \text{say.} \end{aligned}$$

If $j > j_1$, then from (6.29) and noting that $a \ge 1/2$,

$$\begin{aligned} S_j &\le c2^{-j_1 - j(a-1/2)} \sum_{i \le j_0} 2^{(a+1/2)i} \\ &\le c2^{-j_1 - j_1(a-1/2) + j_0(a+1/2)} = c2^{-j_1(1-\lambda)(a+1/2)} \end{aligned}$$

after noting that $j_0 = \lambda j_1$. Finally, since ℓ is a piecewise linear function of j,

$$\begin{aligned} \max_{0 \le j \le j_1} \ell(j) &= \max\{\ell(0), \ell(j_0), \ell(j_1)\} \\ &= \max\{(a+1/2)j_0 - j_1 R, (j_0 - j_1)R, a(j_0 - j_1)\}. \end{aligned}$$

Thus $N_{11} \le cn^{-\Delta_{11}}$, where since $R > 1 + \alpha$,

$$\Delta_{11} = \min\{R - \lambda(a+1/2), (1-\lambda)R, (1-\lambda)a\} = (1-\lambda)a.$$

6.2.2. Estimates for $(1, \infty)$

We have

$$\begin{aligned} N_{1\infty} &\le \sup_{i \le j_0} \sum_j 2^{a(i-j)} \tilde{\eta}(i,j), \\ \tilde{\eta}(i,j) &= \sup_k \sum_l |\langle \psi_I, w_J \rangle|, \end{aligned}$$

and we again use (6.29) to bound $\tilde{\eta}(i,j)$. Hence, on noting that the coefficients of i in the exponents are positive and that $R > a + 1/2$,

$$\begin{aligned} \sup_{i \le j_0} \sum_{j \le j_1} 2^{a(i-j)} \tilde{\eta}(i,j) &\le c \sum_{j \le j_1} 2^{a(j_0-j)+(j-j_1)R+(j_0-j)_+/2} \\ &\le c2^{a(j_0-j_1)} = c2^{-j_1(1-\lambda)a}. \end{aligned}$$

Similarly, now noting that $a > 1/2$,

$$\begin{aligned}\sup_{i\le j_0}\sum_{j>j_1} 2^{a(i-j)}\tilde{\eta}(i,j) &\le c2^{-j_1}\sum_{j>j_1} 2^{a(j_0-j)+(j_0+j)/2}\\ &\le c2^{-j_1+a(j_0-j_1)+(j_0+j_1)/2}\\ &= c2^{-j_1(1-\lambda)(a+1/2)}.\end{aligned}$$

Thus $N_{1\infty} \le cn^{-\Delta_{1\infty}}$ where $\Delta_{1\infty} = (1-\lambda)a$.

6.2.3. Estimates for (∞,∞).

Immediately applying (6.30), we have

$$\begin{aligned}N_{\infty\infty} &\le \sup_I \sum_j 2^{a(i-j)}\sum_k |\langle\psi_I, w_J\rangle|\\ &\le c\sup_{i\le j_0}\sum_j 2^{a(i-j)-(j_1-j)_+R+(j-i)_+/2}\\ &\le c\sum_j 2^{a(j_0-j)-(j_1-j)_+R+(j-j_0)_+/2}\\ &\le c\sum_j 2^{m(j)} \le c2^{\sup_j m(j)},\end{aligned}$$

since the slope of $m(j)$ is never zero. Since $a > 1/2$,

$$\begin{aligned}\sup_j m(j) &= \max\{m(0), m(j_0), m(j_1)\}\\ &= \max\{aj_0 - j_1R, -(j_1-j_0)R, -(j_1-j_0)(a-1/2)\},\end{aligned}$$

so that $N_{\infty\infty} \le cn^{-\Delta_{\infty\infty}}$ with (since $R > a+1/2$)

$$\Delta_{\infty\infty} = \min\{R-\lambda a, (1-\lambda)R, (1-\lambda)(a-1/2)\} = (1-\lambda)(a-1/2).$$

Acknowledgements

The authors would like to thank Michael Frazier for discussions on boundedness of operators and sampling.

This research was partially supported by NSF DMS-92-09130, NSF DMS-95-05151 and the Guggenheim Foundation (IMJ).

References

Brown, L. D. and Low, M. G. (1996). Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.* **3**, 2384-2398.

Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993). Multiresolution analysis, wavelets, and fast algorithms on an interval. *Comptes Rendus Acad. Sci. Paris* (*A*) **316**, 417-421.

Delyon, B. and Juditsky, A. (1996). On minimax wavelet estimators. *Applied Computational and Harmonic Analysis* **3**, 215-228.

Deslauriers, G. and Dubuc, S. (1987). Interpolation dyadique. In *Fractals, Dimensions non-entières et applications*. Masson, Paris.

Deslauriers, G. and Dubuc, S. (1989). Symmetric iterative interpolation processes. *Constr. Approx.* **5**, 49-68.

DeVore, R. and Lorentz, G. (1993). *Constructive Approximation*. Springer Verlag, New York.

Donoho, D. (1992a). De-noising via soft-thresholding. *IEEE Trans. Inform. Theory* **41**, 613-627.

Donoho, D. (1992b). Interpolating wavelet transforms. Technical Report 408, Department of Statistics, Stanford University.

Donoho, D. (1993). Smooth wavelet decompositions with blocky coefficient kernels. In *Recent Advances in Wavelet Analysis* (Edited by L. Schumaker). Academic Press, Boston.

Donoho, D. (1994). Asymptotic minimax risk for sup-norm loss; solution via optimal recovery. *Probab. Theory Related Fields* **99**, 145-170.

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

Donoho, D. L. and Johnstone, I. M. (1990). Wavelets and optimal nonparametric function estimation. Technical Report, Department of Statistics, University of California Berkeley.

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

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

Donoho, D. L., Johnstone, I. M., Kerkyacharian, G. and Picard, D. (1995). Wavelet shrinkage: Asymptopia? (With Discussion). *J. Roy. Statist. Soc. Ser. B* **57**, 301-369.

Donoho, D. L., Liu, R. C. and MacGibbon, K. B. (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.* **18**, 1416-1437.

Donoho, D. L. and Nussbaum, M. (1990). Minimax quadratic estimation of a quadratic functional. *J. Complexity* **6**, 290-323.

Dubuc, S. (1986). Interpolation through an iterative scheme. *J. Math. Anal. Appl.* **114**, 185-204.

Frazier, M., Jawerth, B. and Weiss, G. (1991). *Littlewood-Paley Theory and the study of function spaces*. NSF-CBMS Regional Conf. Ser in Mathematics, **79**, American Mathematical Society, Providence, RI.

Meyer, Y. (1990). *Ondelettes et Opérateurs, I: Ondelettes, II: Opérateurs de Calderón-Zygmund, III:* (with R. Coifman)*, Opérateurs multilinéaires*. Hermann, Paris. English translation of first volume is published by Cambridge University Press.

Peetre, J. (1975). *New Thoughts on Besov Spaces*. I. Duke University Mathematics Series, Raleigh, Durham.

Triebel, H. (1992). *Theory of Function Spaces* II. Birkhäuser Verlag, Basel.

Statistics Department, Stanford University, Stanford CA 94305-4065, U.S.A.

E-mail: imj@stat.stanford.edu

E-mail: donoho@stat.stanford.edu

(Received January 1998; accepted September 1998)