J. R. Statist. Soc. B (2004)
66, *Part* 3, *pp.* 547–573

Wavelet deconvolution in a periodic setting

Iain M. Johnstone,

Stanford University, USA

Gérard Kerkyacharian,

Centre National de la Recherche Scientifique and Université de Paris X, France

Dominique Picard

Centre National de la Recherche Scientifique and Université de Paris VII, France

and Marc Raimondo

University of Sydney, Australia

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on 'Statistical approaches to inverse problems' on Wednesday, December 10th, 2003*, Professor J. T. Kent *in the Chair*]

Summary. Deconvolution problems are naturally represented in the Fourier domain, whereas thresholding in wavelet bases is known to have broad adaptivity properties. We study a method which combines both fast Fourier and fast wavelet transforms and can recover a blurred function observed in white noise with $O\{n \log(n)^2\}$ steps. In the periodic setting, the method applies to most deconvolution problems, including certain 'boxcar' kernels, which are important as a model of motion blur, but having poor Fourier characteristics. Asymptotic theory informs the choice of tuning parameters and yields adaptivity properties for the method over a wide class of measures of error and classes of function. The method is tested on simulated light detection and ranging data suggested by underwater remote sensing. Both visual and numerical results show an improvement over competing approaches. Finally, the theory behind our estimation paradigm gives a complete characterization of the 'maxiset' of the method: the set of functions where the method attains a near optimal rate of convergence for a variety of L^p loss functions.

Keywords: Adaptive estimation; Deconvolution; Meyer wavelet; Nonparametric regression

1. Deconvolution in white noise

Suppose that we observe the random process $Y_n(\cdot)$,

$$Y_n(\mathrm{d}t) = f * g(t)\,\mathrm{d}t + \sigma n^{-1/2}\,W(\mathrm{d}t), \qquad t \in T = [0,1], \tag{1}$$

where σ is a positive constant, $W(\cdot)$ is Gaussian white noise and

$$f * g(t) = \int_T f(t-u)\,g(u)\,\mathrm{d}u. \tag{2}$$

Our goal is to recover the unknown function f from the noisy blurred observations (1). The blurring function g in convolution (2) is assumed to be known. Further, we assume that the

Address for correspondence: Iain M. Johnstone, Department of Statistics, Stanford University, Stanford, CA 94305-4065, USA.
E-mail: imj@stat.stanford.edu

 1369–7412/04/66547

function f is periodic on the unit interval T and that g has a certain degree of smoothness. There is an extensive statistical literature on deconvolution problems; in particular wavelet methods have received considerable attention over the last decade. References that are particularly relevant to the present work include Donoho (1995), Abramovich and Silverman (1998), Pensky and Vidakovic (1999), Fan and Koo (2002), Kalifa and Mallat (2003) and Neelamani *et al.* (2003): these works in turn contain further references to previous literature.

An important application setting that is modelled by expression (1) is that of motion blur in signals or images; see for example Bertero and Boccacci (1998). Here g is taken as a 'boxcar' $g(x) = (2a)^{-1}\, \mathbb{I}_{[-a,a]}(x)$, of half-width a. Owing to oscillations in the Fourier coefficients of g, this situation escapes the assumptions of much wavelet literature, but recent work of Neelamani *et al.* (2004) studied it explicitly with their ForWaRD algorithm.

Our aim in this paper is to study a wavelet deconvolution algorithm which can be applied to many deconvolution problems including certain cases of boxcar blur. We are particularly interested in obtaining adaptivity properties relative to a variety of error measures and function classes. Our theoretical investigation is conducted by using model (1), but examples and software are provided for data sampled at a discrete set of n equally spaced points.

For *ordinary smooth convolution* where the Fourier coefficients of g decay in a polynomial fashion, $|g_l| \sim C|l|^{-\nu}$, our proposal can recover the unknown function f with an accuracy of order

$$\left\{\frac{\log(n)}{n}\right\}^{\alpha}, \qquad \alpha = \frac{sp}{2(s+\nu)+1}, \tag{3}$$

performance being measured in an integrated L^p-metric, for any $p > 1$. Here n denotes the usual sample size and s plays the role of a smoothness index for our target function f (taken in a large class which includes spatially inhomogeneous functions). For boxcar blur, we show that rate (3) holds with $\nu = \frac{3}{2}$, provided that the boxcar width is 'badly approximated' by rational numbers. This notion is recalled in Section 2.2; it includes boxcars of width that is given by quadratic irrational numbers such as $\sqrt{5}$; see also remark 8 following proposition 2.

Our rate result (3) is established for a class of convolution operators satisfying a decay condition when averaged over dyadic Fourier blocks. Thus, if (g_l) denote the Fourier coefficients of g, we suppose that for some fixed $r > 0$

$$\left(2^{-j} \sum_{l:|l|=2^j}^{2^{j+r}} |g_l|^{-2}\right)^{1/2} \asymp 2^{j\nu}$$

(condition C_ν). (The notation $a_j \asymp b_j$ means that there are constants c_0 and c_1 such that, for all j, $c_0 \leqslant a_j/b_j \leqslant c_1$.) Condition C_ν typically holds for ordinary 'smooth' convolution and also covers certain oscillatory cases, such as arise with irrational boxcar blur; see for example Fig. 2 in Section 2.1 and proposition 2. It does not include 'supersmooth' kernels, such as the Gaussian kernel, with exponential or faster Fourier decay.

For both theoretical and practical convenience, we use band-limited wavelet basis functions, and in particular the (periodized) Meyer wavelet basis for which fast algorithms are available; Kolaczyk (1994) and Donoho and Raimondo (2004). Our method can thus perform deconvolution in $O\{n\log(n)^2\}$ steps. The WaveD software package that was used to prepare most figures and tables in this paper is available at `http://www.usyd.edu.au:8000/u/marcr/`. It is intended for use with `Wavelab`; see Buckheit *et al.* (1995).

We begin in Section 2 by describing an application of statistical deconvolution to remote sensing. There follow short reviews of continued fractions, periodized Meyer wavelets, Besov spaces and wavelet shrinkage. Section 3 describes our method specifically, and its relationship

with the wavelet–vaguelette approach of Donoho (1995) and Abramovich and Silverman (1998). Section 4 is concerned with numerical performance and competing approaches. In the implementation of our method the choice of tuning parameters is informed by asymptotic minimax theory: this is discussed in Section 5. Proofs are summarized in Appendices A and B.

2. Motivations and preliminaries

2.1. Illustration from remote sensing

Deconvolution is a common problem in many areas of signal and image processing; see for example Jain (1989). Here we shall focus on light detection and ranging (LIDAR) remote sensing as in Je Park *et al*. (1997) and Harsdorf and Reuter (2000). LIDAR uses a laser device which emits pulses, reflections of which are gathered by a telescope that is aligned with the laser. The return signal is used to determine the distance and position of the reflecting material. Accordingly, the distance resolution is limited by the time resolution of the LIDAR instrument. If the system response function of the LIDAR detector is longer than the time resolution interval the measured LIDAR signal is blurred and the effective accuracy of the LIDAR decreases. This loss of precision can be corrected by deconvolution. In practice, measured LIDAR signals are corrupted by additional noise which renders direct deconvolution impossible. Borrowed from Harsdorf and Reuter (2000), we have depicted an ideal LIDAR signal in Fig. 1; this will be our target function f for numerical illustrations throughout this paper. The system response function of the LIDAR detector (denoted $g(t)$ in expression (1)) is calibrated *a posteriori* once the LIDAR instrument has been built. We follow Harsdorf and Reuter (2000) and use system response functions that have a strong low pass characteristic. In the WaveD software that was

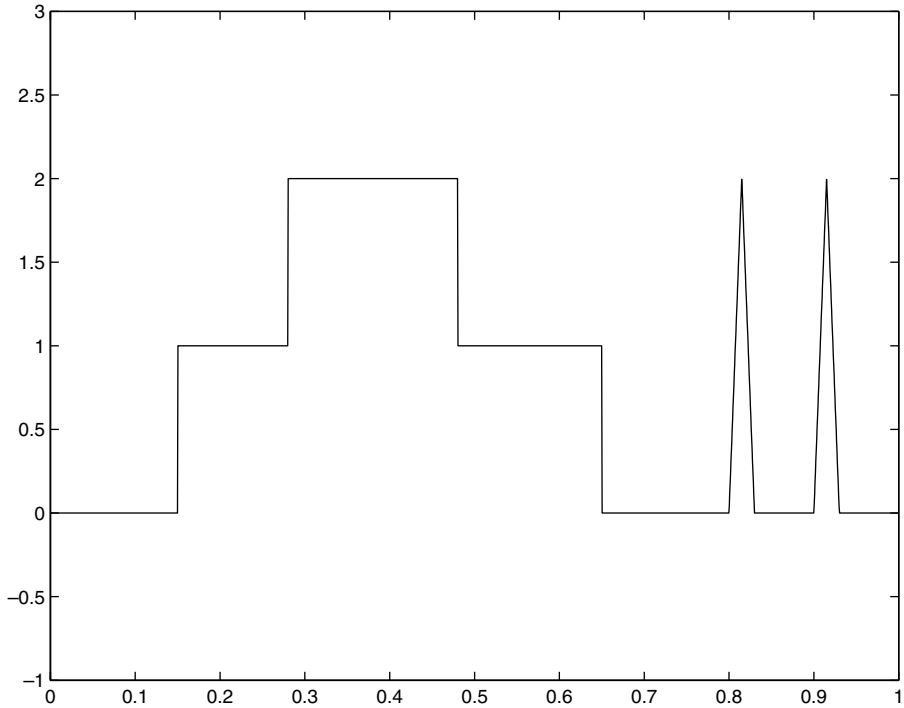

Fig. 1. Ideal LIDAR signal as in Harsdorf and Reuter (2000), corresponding to data for underwater LIDAR

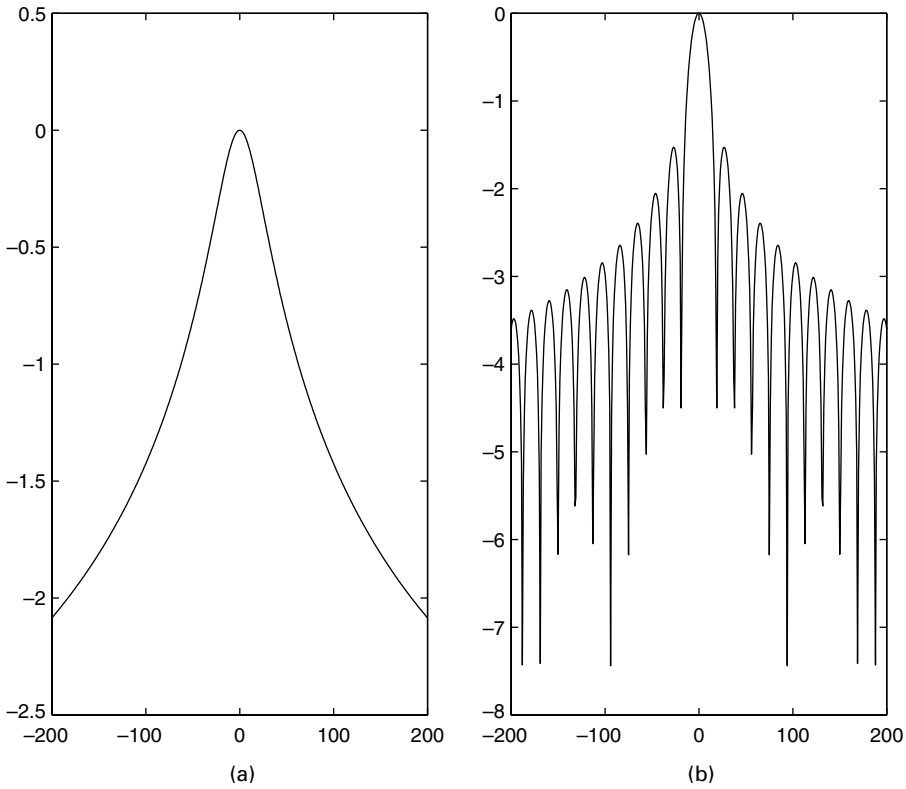

Fig. 2. (a) Log-spectrum of a $\Gamma(1, 0.0065)$ probability distribution function (smooth blur) and (b) log-spectrum of a boxcar function $g(x) = (1/2a)\,\mathbb{I}_{[-a,a]}(x)$ with $a = 1/\sqrt{353}$ (boxcar blur)

used to plot most figures in this paper, the system response function parameters can be changed by the user to accommodate different calibration settings. In Fig. 2, we illustrate a smooth blur and a boxcar blur scenario in the Fourier domain. These two examples of system response function shapes illustrate the possibilities that are offered by our assumption C_ν, under which near optimal rates are achievable. Finally, Fig. 3 shows artificial LIDAR data for a combination of different noise levels and system response functions.

2.2. Boxcar blur and the continued fractions algorithm

A boxcar function is the indicator of an interval $g(x) = (1/2a)\,\mathbb{I}_{[-a,a]}(x)$ where the parameter a indicates a preferred spatial scale. The Fourier coefficients of such a boxcar are given by

$$g_l = \frac{\sin(\pi l a)}{\pi l a}, \qquad l \in \mathbb{Z}. \tag{4}$$

The convolution problem (1) that is associated with the boxcar, later referred to as boxcar blur, has the problem that, for *rational* $a = p/q$, the coefficients g_k vanish for any integer k multiple of q. Hence, even without noise some frequencies are lost and f cannot be fully recovered. The problem is less severe for irrational numbers, and particularly for those which are 'badly approximable' (BA) by rational numbers. We briefly review the key tool in constructing such numbers.

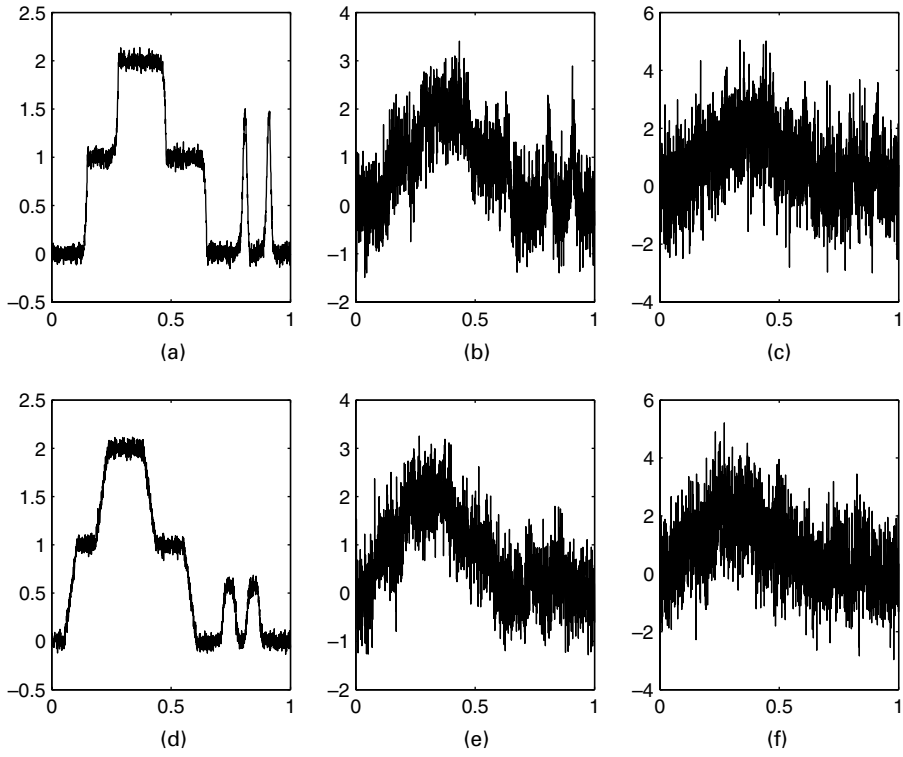

Fig. 3. Simulated LIDAR signals (1) with $t_i = i/n$, $n = 2048$, corresponding to the system response functions of Fig. 2: (a) smooth blur with low (standard deviation sd = 0.05) noise level; (b) smooth blur with medium (sd = 0.5) noise level; (c) smooth blur with high (sd = 1) noise level; (d) boxcar blur with low (sd = 0.05) noise level; (e) boxcar blur with medium (sd = 0.5) noise level; (f) boxcar blur with high (sd = 1) noise level

2.2.1. The continued fractions algorithm

Let a_0 be an integer and $a_1, a_2, \ldots$ be strictly positive integers. Define sequences (p_k) and (q_k) recursively by $p_0/q_0 = a_0$, $p_1/q_1 = a_0 + 1/a_1$ and $p_2/q_2 = a_0 + 1/(a_1 + 1/a_2)$, and for $n \geqslant 2$ let

$$\begin{aligned} q_n &= a_n q_{n-1} + q_{n-2}, \\ p_n &= a_n p_{n-1} + p_{n-2}. \end{aligned} \tag{5}$$

The sequence of rational numbers (p_k/q_k) that is constructed in this way has very special properties, the first of which being that

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{\ddots + \frac{1}{a_k}}} = [a_0; a_1, a_2, \ldots, a_k] = \frac{p_k}{q_k}. \tag{6}$$

In fact, any real number a that is not an integer may be uniquely determined by its continued fraction expansion. To every real number corresponds a unique sequence (a_k) and vice versa:

$$a = [a_0; a_1, a_2, \ldots] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \frac{1}{a_3 + \ldots}}}. \tag{7}$$

For rational numbers the expansion stops as in equation (6) and $a_k = 0$ for later k, whereas for

irrational number a the sequence of $(a_k), a_k > 0,$ is infinite. The rational numbers $(p_k(a)/q_k(a))$ defined at expression (5) are called the convergents of a. For any irrational number a, the convergents have the property of *best approximation*: for $n \geqslant 1$,

$$\inf_{1 \leqslant k \leqslant q_n} \|ka\| = |q_n a - p_n| = \|q_n a\|, \tag{8}$$

where $\|x\|$ denotes the distance from $x \in \mathbb{R}$ to the nearest integer. The study of such Diophantine approximations plays a central role in our analysis of the boxcar blur, since from equation (4)

$$\frac{2}{\pi} \frac{\|la\|}{la} \leqslant g_l \leqslant \frac{\|la\|}{la}. \tag{9}$$

We recall some basic properties, referring to Lang (1966) and Khinchin (1992) for further details. The quality of best approximation satisfies

$$\frac{1}{2q_{n+1}} < \|q_n a\| < \frac{1}{q_{n+1}}. \tag{10}$$

(a) The denominators q_n grow at least geometrically:

$$\begin{array}{ll} q_{n+i} \geqslant 2^{(i-1)/2} q_n, & i = 2k+1, \\ q_{n+i} \geqslant 2^{i/2} q_n, & i = 2k, \quad k > 0. \end{array} \tag{11}$$

(b) For all $n \geqslant 0$,

$$a_n < q_n / q_{n-1} \leqslant a_n + 1.$$

Hence, the size of the *elements* in the continued fractions algorithm (5) determines the quality of best rational approximation to a. It is customary to define families of irrational numbers a according to the size of their elements as follows.

Definition 1. An irrational number a is called BA if

$$\sup_n \{a_n(a)\} < \infty.$$

Definition 2. A rational number a is called BA of order n if a is the convergent of order k of a BA number ($a = p_k/q_k$) and if $q_{k-1} \leqslant n < q_k$.

The set of all BAs contains quadratic irrational numbers (e.g. $\sqrt{5}$). For the boxcar blur, we prove (proposition 2 in Section 4) that condition C_ν holds with $\nu = \frac{3}{2}$ for any scale a chosen in the set of BAs. In the finite sample implementation (of size n) of model (1), our method will remain numerically stable for any scale a that is chosen in the set of BA rational numbers of order at least n (see remark 8 below proposition 2) and satisfying a uniform bound (in a) on $\sup_n \{a_n(a)\}$. We refer to Johnstone and Raimondo (2002) for a discussion of cases outside the BA numbers.

2.3. Periodized Meyer wavelet transforms

Let (ϕ, ψ) denote the Meyer scaling and wavelet function; see Meyer (1992) or Mallat (1999). As usual,

$$\psi_\kappa(x) = \psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k), \qquad j, k \in \mathbb{Z}, \tag{12}$$

is the dilated and translated wavelet at resolution level j and time position $k/2^j$; here and below κ denotes the bivariate index (j, k). The functions ϕ_κ are defined similarly. Such wavelet functions

define a multiresolution analysis of $L^2(\mathbb{R})$; for any $f \in L^2(\mathbb{R})$ the following expansion holds:

$$f = \sum_k \alpha_{j_0,k}\phi_{j_0,k} + \sum_{j \geqslant j_0} \sum_k \beta_{j,k}\psi_{j,k} \tag{13}$$

where

$$\begin{aligned} \alpha_\kappa &= \int f\phi_\kappa, \\ \beta_\kappa &= \int f\psi_\kappa \end{aligned} \tag{14}$$

are the wavelet coefficients of f. Quite naturally, we can define a similar multiresolution analysis for periodic functions in $L^2(T)$, $T = [0, 1]$. This is done by periodizing the basis functions

$$\begin{aligned} \Phi_\kappa(x) &= \sum_{l \in \mathbb{Z}} \phi_\kappa(x + l), \\ \Psi_\kappa(x) &= \sum_{l \in \mathbb{Z}} \psi_\kappa(x + l). \end{aligned} \tag{15}$$

Here and in the rest of the paper (Φ, Ψ) will denote the periodized Meyer scaling and wavelet functions (Fig. 4). Thus, for any periodic function f an expansion that is similar to equation (13) holds with periodized basis functions $(\Phi_\kappa, \Psi_\kappa)$ and bivariate index κ restricted to the set $I = \{(j, k) : j \geqslant 0 \text{ and } k = 0, 1, \ldots, 2^j - 1\}$. We use this basis for the following reasons.

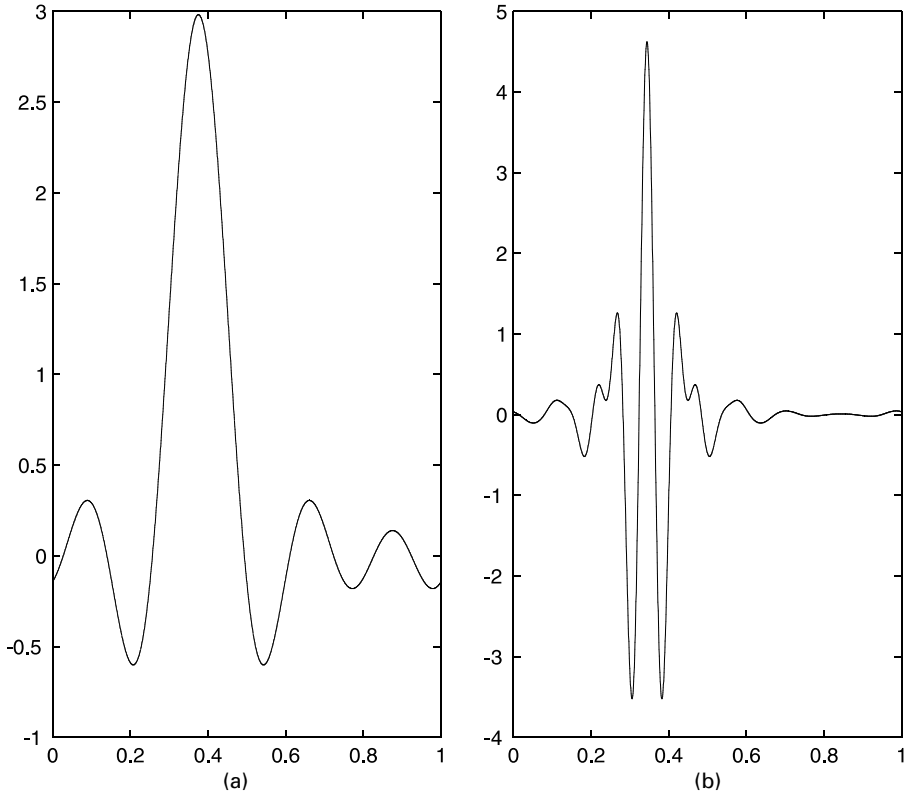

Fig. 4. Periodized Meyer scaling and wavelet function: (a) $\Phi_{3,4}$; (b) $\Psi_{4,5}$

(a) The Meyer wavelet is band limited. In particular, we have $\mathrm{Supp}(\hat{\psi}) = \{w : |w| \in [2\pi/3, 8\pi/3]\}$, where $\hat{\psi}(\xi) = \int \psi(x)\exp(-\mathrm{i}\xi x)\,\mathrm{d}x$ denotes the Fourier transform of ψ.
(b) An efficient algorithm, due to Kolaczyk (1994), is available to compute the periodized Meyer wavelet transforms. It requires only $O\{n\log(n)^2\}$ steps to derive an empirical version of the coefficients (14) from a sample of size n of f.

Band-limited wavelets have been used in the deconvolution setting by Walter and Shen (1999), Shen and Walter (2002), Pensky and Vidakovic (1999) and Fan and Koo (2002). General information on band-limited wavelet bases may be found, for example, in Mallat (1999), Hernàndez and Weiss (1996) and Walter (1994).

2.4. A wide class of target functions

Let us first introduce the standard Besov spaces of periodic functions $B^s_{\pi,r}(T), s > 0, \pi \geqslant 1$ and $r \geqslant 1$. For this, define for every measurable function f

$$\Delta_\varepsilon f(x) = f(x+\varepsilon) - f(x);$$

then, recursively, $\Delta^2_\varepsilon f(x) = \Delta_\varepsilon(\Delta_\varepsilon f)(x)$ and similarly $\Delta^N_\varepsilon f(x)$ for positive integer N. Let

$$\rho^N(t, f, \pi) = \sup_{|\varepsilon| \leqslant t} \left\{ \int_0^1 |\Delta^N_\varepsilon f(u)|^\pi \,\mathrm{d}u \right\}^{1/\pi}.$$

Then, for $N > s$, we define

$$B^s_{\pi,r}(T) = \left\{ f \text{ periodic: } \left[\int_0^1 \left\{ \frac{\rho^N(t, f, \pi)}{t^s} \right\}^r \frac{\mathrm{d}t}{t} \right]^{1/r} < \infty \right\}$$

(with the usual modifications for $r = \infty$ or $\pi = \infty$).

In what follows, we shall also write $\Psi_{-1} = \Phi$. In this setting, recall that the Besov spaces are characterized by the behaviour of the wavelet coefficients (as soon as the wavelet is periodic and has enough smoothness and vanishing moments). In particular, for $f \in L^\pi(T)$,

$$\begin{aligned} f &= \sum_{j,k} \beta_{j,k} \Psi_{j,k} \in B^s_{\pi,r}(T) \\ &\Leftrightarrow \sum_{j \geqslant 0} 2^{j(s+1/2-1/\pi)r} \Big(\sum_{0 \leqslant k \leqslant 2^j} |\beta_{j,k}|^\pi \Big)^{r/\pi} < \infty. \end{aligned} \tag{16}$$

The Besov spaces have proved to be an interesting scale for studying the properties of statistical procedures. The index s indicates the degree of smoothness of the function. Owing to the differential averaging effects of the integration parameters π and r, the Besov spaces capture a variety of smoothness features in a function including spatially inhomogeneous behaviour; see Donoho *et al*. (1995).

2.5. Wavelet shrinkage

Wavelet shrinkage is now a well-established statistical procedure for nonparametric estimation. A wavelet estimator of an unknown function $f \in L^2(T)$, based on hard thresholding, is given by

$$\hat{f} = \sum_{\kappa \in I_0} \hat{\alpha}_\kappa \, \mathbb{I}_{\{|\hat{\alpha}_\kappa| \geqslant \lambda_{j_0}\}} \Phi_\kappa + \sum_{\kappa \in I_1} \hat{\beta}_\kappa \, \mathbb{I}_{\{|\hat{\beta}_\kappa| \geqslant \lambda_j\}} \Psi_\kappa \tag{17}$$

where $\hat{\alpha}_\kappa$ and $\hat{\beta}_\kappa$ are estimated wavelet coefficients and I_0 and I_1 are sets of indices. $I_0 = \{(j_0, k) : k = 0, 1, \ldots, 2^{j_0} - 1\}$ corresponds to a coarse resolution level j_0 and $I_1 = \{(j, k) : k =$

$0, 1, \ldots, 2^j - 1, j_0 \leqslant j \leqslant j_1\}$ indexes details up to a fine resolution level j_1. The procedure (17) is non-linear since only large coefficients $|\hat{\beta}_\kappa| \geqslant \lambda_j$ are kept; here λ_j is a threshold parameter. The choices of parameters j_0, j_1 and λ_j as well as estimators $\hat{\alpha}_\kappa$ and $\hat{\beta}_\kappa$ depend on the problem at hand. For deconvolution problem (1) this will be discussed in the next section.

3. Wavelet deconvolution in the Fourier domain

3.1. Inverse estimation paradigm

Since the Fourier transform interchanges convolution and multiplication, it is natural to employ Fourier representations for the deconvoluation problem. Let $e_l(t) = \exp(2\pi i l t)$, $l \in \mathbb{Z}$, and write $f_l = \langle f, e_l\rangle$ and $g_l = \langle g, e_l\rangle$ for the Fourier coefficients of f and g respectively where $\langle f, g\rangle = \int_T f\bar{g}$. Letting $h = f * g$ we have

$$h_l = \langle f * g, e_l\rangle = f_l \times g_l. \tag{18}$$

For the (real-valued) random processes Y_n and W we write, with a slight abuse of notation, $y_l = \langle e_l, Y_n\rangle = \int e_l \, \mathrm{d}Y_n$ and $z_l = \langle e_l, W\rangle = \int e_l \, \mathrm{d}W$. Calculating Fourier coefficients in expression (1):

$$y_l = h_l + \sigma n^{-1/2} z_l \tag{19}$$

where z_l are zero-mean Gaussian random variables. We denote by Ψ_l^κ the Fourier coefficients of Ψ_κ, i.e. $\Psi_l^\kappa = \langle \Psi_\kappa, e_l\rangle$. Combining equation (18) with Plancherel's identity we obtain

$$\int_T f\bar{\Psi}_\kappa = \sum_l f_l \bar{\Psi}_l^\kappa = \sum_l \frac{h_l}{g_l}\bar{\Psi}_l^\kappa. \tag{20}$$

Noting that $\bar{\Psi}_\kappa = \Psi_\kappa$, we can recover wavelet coefficients

$$\beta_\kappa = \int_T f\Psi_\kappa = \sum_l \frac{h_l}{g_l}\bar{\Psi}_l^\kappa. \tag{21}$$

Here g_l and Ψ_l^κ are known Fourier coefficients but the h_ls are not directly observable; in equation (19) we take y_l as an (unbiased) estimator of h_l and let

$$\hat{\beta}_\kappa = \sum_l \frac{y_l}{g_l}\bar{\Psi}_l^\kappa \tag{22}$$

be our estimator of β_κ which can be computed from the observations (1). Of course, an estimator $\hat{\alpha}_\kappa$ of α_κ is defined in a similar fashion with Φ in place of Ψ.

Let $C_j = \{l : \Psi_l^\kappa \neq 0\}$—it is easily seen that this set does not depend on k. Indeed, from the compact support of the Meyer wavelet, we have

$$C_j \subset (2\pi/3)[-2^{j+2}, -2^j] \bigcup [2^j, 2^{j+2}].$$

3.2. The wavelet deconvolution method

For deconvolution problems (1) we shall use wavelet-based estimator (17) with coefficients (22). Estimator (17) requires three input parameters: j_0, j_1 and λ. The coarse scale has the default value $j_0 = 3$ in software and is not important in the asymptotic theory. To specify the more critical thresholds λ_j and finest scale j_1, set

$$\tau_j = \left(|C_j|^{-1} \sum_{l \in C_j} |g_l|^{-2}\right)^{1/2}, \tag{23}$$

where $|C_j|$ denotes the cardinality of C_j. Then, for the thresholds,

$$\lambda_j^{\mathrm{WaveD}} := \lambda_{n,j} = \eta\sigma\tau_j\sqrt{\{\log(n)/n\}}, \tag{24}$$

where the default value of η is $\sqrt{2}$ in software—this is discussed further below. Finally, using the decay parameter ν from assumption C_ν, the finest scale j_1 is determined from

$$2^{j_1} = \{n/\log(n)\}^{1/(1+2\nu)}. \tag{25}$$

If it is necessary to compute an estimate of the noise standard deviation σ, we adapt the method of Donoho *et al*. (1995) that was developed for direct data. If $y_{J,k} = \langle Y_n, \Psi_{J,k}\rangle$ denote the finest scale wavelet coefficients of the observed data, then $\hat{\sigma} = \mathrm{mad}(y_{J,k})/0.6745$, where mad is the median absolute deviation.

We summarize the main steps of our wavelet deconvolution method, and we illustrate it in Fig. 5. Here and in the rest of the paper we refer to this as the *WaveD* method.

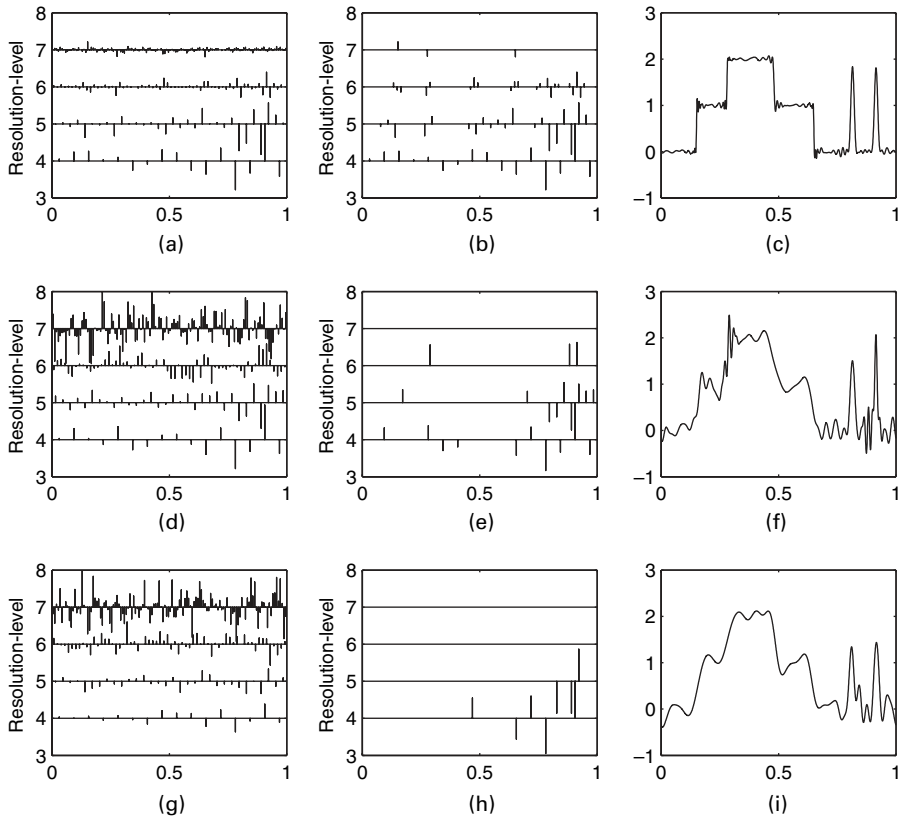

Fig. 5. WaveD method applied to the LIDAR signals of Fig. 3 (smooth blur) (the threshold values are summarized in Table 1; levels 4–7 are shown as 100% thresholding occurs at level 8 or below): (a) low noise, estimated wavelet coefficients (22); (b) low noise, estimated wavelet coefficients (22) after shrinkage (24) with $\eta_S = \sqrt{2}$; (c) low noise, estimated LIDAR signal; (d) medium noise, estimated wavelet coefficients (22); (e) medium noise, estimated wavelet coefficients (22) after shrinkage (24) with $\eta_S = \sqrt{2}$; (f) medium noise, estimated LIDAR signal; (g) high noise, estimated wavelet coefficients (22); (h) high noise, estimated wavelet coefficients (22) after shrinkage (24) with $\eta_S = \sqrt{2}$; (i) high noise, estimated LIDAR signal

Table 1. Level-by-level thresholds (smooth blur): $\lambda_j^{\mathrm{WaveD}}$ defined by expression (24)†

Noise level	η	*Thresholds for the following levels of j:*			
		$j=5$	$j=6$	$j=7$	$j=8$
Low ($\hat{\sigma}=0.047$)	$\sqrt{2}$	0.0036 (31.35%)	0.0066 (62.5%)	0.0128 (95.31%)	0.0242 (100%)
Medium ($\hat{\sigma}=0.491$)	$\sqrt{2}$	0.0375 (68.75%)	0.0688 (98.44%)	0.1328 (100%)	Maximum level: $j_1=6$
High ($\hat{\sigma}=0.955$)	$\sqrt{2}$	0.0729 (90.63%)	0.1338 (100%)	Maximum level: $j_1=5$	
Low ($\hat{\sigma}=0.047$)	$\sqrt{6}$	0.0063 (31.25%)	0.0115 (81.25%)	0.0222 (99.22%)	0.0418 (100%)
Medium ($\hat{\sigma}=0.491$)	$\sqrt{6}$	0.0649 (87.5%)	0.1192 (100%)	Maximum level: $j_1=5$	
High ($\hat{\sigma}=0.955$)	$\sqrt{6}$	0.1263 (96.88%)	0.2318 (100%)	Maximum level: $j_1=5$	

†Indicated in parentheses are the corresponding fractions of shrunken coefficients. The first three rows correspond to the smaller choice $\eta_S=\sqrt{2}$. The last three rows correspond to the conservative choice $\eta_L=\sqrt{\{2(2\nu+1)\}}$ with $\nu=1$. (In the WaveD software the (default) maximum resolution level (25) is set to be the level preceeding $j(100\%)$ where $j(100\%)$ is the smallest level where 100% of thresholding occurs.)

(a) Compute Fourier coefficients y_l and g_l and recover wavelet coefficients (22) by using Kolaczyck's algorithm (which requires only $O\{n\log(n)^2\}$ operations); see Figs 5(a), 5(d) and 5(g).
(b) Compute, if needed, an estimate of the noise standard deviation $\hat{\sigma}$ as described above. Find thresholds $\lambda_j := \lambda_{n,j}$ as shown in equation (24) and illustrated in Table 1 for the boxcar.
(c) Apply hard thresholding $\hat{\beta}_\kappa\, \mathbb{I}(|\hat{\beta}_\kappa| > \lambda_j)$; see Figs 5(b), 5(e) and 5(h). Finally, invert the wavelet transform to obtain an estimate of f; see Figs 5(c), 5(f) and 5(i). (In the WaveD software the (default) maximum resolution level j_1 is determined from the data as follows: j_1 is set to be the level preceeding $j(100\%)$ where $j(100\%)$ is the smallest level where 100% of thresholding occurs; see Table 1.)
(d) Cycle-spin the WaveD estimator in the fashion of Coifman and Donoho (1995) (optional). This improves visual and numerical performance and was used in Fig. 7 and Table 2 later. We refer to Donoho and Raimondo (2004) for an efficient algorithm which cycle-spins the WaveD estimator over all circulant shifts.

3.3. Connection with the wavelet–vaguelette decomposition

Donoho (1995) gave the first discussion of wavelet thresholding in linear inverse problems and introduced the wavelet–vaguelette decomposition (WVD). Specialized to convolution operators on the circle, the WVD structure in part postulates the existence of biorthogonal systems $(\mathcal{U}_\kappa)$ and $(\mathcal{V}_\kappa)$ and pseudosingular values κ_j (not depending on the spatial index k) such that

$$g * \Psi_\kappa = \kappa_j \mathcal{V}_\kappa,$$
$$\check{g} * \mathcal{U}_\kappa = \kappa_j \Psi_\kappa,$$

where $\check{g}(t) = g(-t)$. In terms of Fourier coefficients, and setting $s_{jl} = g_l/\kappa_j$,

$$\begin{aligned} \mathcal{V}_l^\kappa &= s_{jl}\Psi_l^\kappa, \\ \mathcal{U}_l^\kappa &= \Psi_l^\kappa / \bar{s}_{jl}. \end{aligned} \tag{26}$$

The WVD class of estimators takes a co-ordinatewise thresholding rule $\delta(x_\kappa, t_j)$ and level-dependent thresholds (t_j) and sets $\hat{f} = \Sigma_\kappa\, \delta(\langle \mathcal{U}_\kappa, Y_n\rangle/\kappa_j, t_j)\Psi_\kappa$. If we observe that

$$\langle \mathcal{U}_\kappa, Y_n\rangle/\kappa_j = \sum_l y_l \bar{\mathcal{U}}_l^\kappa/\kappa_j = \sum_l (y_l/g_l)\bar{\Psi}_l^\kappa = \hat{\beta}_\kappa$$

then it is clear that our estimator (17) *formally* can be viewed as being consistent with the WVD recipe. However, the *implementation* of the estimator differs here—in the WaveD scenario (22) the functions $\mathcal{U}_\kappa$ and $\mathcal{V}_\kappa$ are not constructed explicitly, and the coefficients $\hat{\beta}_\kappa = \langle \mathcal{U}_\kappa, Y_n\rangle/\kappa_j$ are instead evaluated in the Fourier domain using the original wavelets Ψ_κ and filter g.

The key additional property that is needed to establish a WVD is that the systems $(\mathcal{U}_\kappa)$ and $(\mathcal{V}_\kappa)$ each form Riesz bases—this property allows lower bounds and hence minimax rates of convergence to be established over the Besov classes that are considered in this paper. The lower bound arguments are given in detail in Donoho (1995) for L_2-loss and, as noted there, the methods can be extended to more general loss measures.

For the dilation homogeneous operators on $\mathbb{R}$ that were principally studied in Donoho (1995) and Abramovich and Silverman (1998), the vaguelette systems are multiples of translates and dilates of a single mother vaguelette $\mathcal{U}_{0,0}$ or $\mathcal{V}_{0,0}$, and the Riesz basis property can be established as in Donoho (1995). See also Lee and Lucier (2001).

This dilation structure is no longer available for the candidate vaguelettes (26) corresponding to convolution on the circle (e.g. Fig. 6). Nevertheless, we show in Appendix B that $(\mathcal{U}_\kappa)$ and $(\mathcal{V}_\kappa)$ are Riesz bases if $\kappa_j = \tau_j^{-1}$ and condition C_ν holds and if in addition we have, for constants C_0 and C that are independent of j,

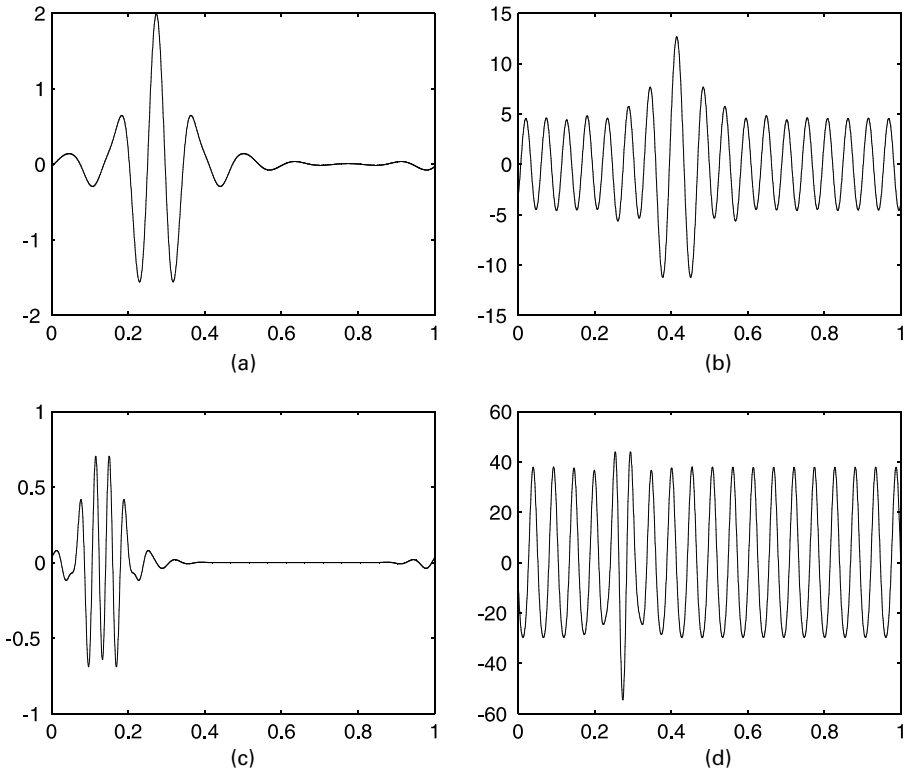

Fig. 6. Candidate vaguelettes corresponding to the boxcar convolution of Fig. 2: (a) $\mathcal{V}_{4,5}$; (b) $\mathcal{U}_{4,5}$; (c) $\mathcal{V}_{5,6}$; (d) $\mathcal{U}_{5,6}$

$$\left.\begin{aligned} &C_0 \leqslant s_{jl} \leqslant C, \\ &\Delta s_{jl} \leqslant 2^{-j} C, \\ &\Delta^2 s_{jl} \leqslant 2^{-2j} C, \end{aligned}\right\} \tag{27}$$

where the difference operator $\Delta s_{jl} = s_{j,l+1} - s_{jl}$ and $\Delta^2 s_{jl} = \Delta(\Delta s_{jl})$.

These conditions hold if $g_l \equiv C|l|^{-\nu}$ or more generally if $c_0|l|^{-\nu} \leqslant |g_l| \leqslant c_1|l|^{-\nu}$ and $\Delta^r g_l \leqslant |l|^{-\nu-r}$ for $r = 1$ and $r = 2$. For such convolution operators, we obtain lower bounds to the rates of convergence over Besov spaces via the methods of Donoho (1995).

However, the sufficient conditions (27) are not satisfied by the boxcar kernel, owing to the oscillations in g_l which inflate Δg_l. We do not yet know whether the systems $(\mathcal{U}_\kappa)$ and $(\mathcal{V}_\kappa)$ form Riesz bases in this case, and so yield vaguelettes; see Fig. 6. Thus the only lower bounds that are currently available for this kernel are those which were established for Fourier hyperrectangles and ellipsoids in Johnstone and Raimondo (2004).

3.4. Remarks on tuning

Asymptotic minimax theory gives insight into the choice of tuning parameters. For example, in direct estimation problems Donoho *et al.* (1995) have shown near optimality of

$$\lambda_j^{\mathrm{UNI}} := \lambda_n = \hat{\sigma}\sqrt{\{2\log(n)/n\}} \tag{28}$$

where $\hat{\sigma}$ is an estimated scale from the data and

$$2^{j_1} = O\{n/\log(n)\}. \tag{29}$$

For deconvolution problems (1), our main result (proposition 1 in Section 4.1) states that, for any constant $\eta \geqslant 2\sqrt{\{8\pi(p \vee 2)\}}$ that is sufficiently large, the choices (24) and (25) are near optimal for a wide variety of target function (Appendix B.1) and L^p loss functions.

It may be seen that the finest scale j_1 that is suggested by equation (25) is considerably smaller than that given by equation (29) in the direct case. The size of the thresholds in equation (24) may in principle be larger than in equation (28), but in practice smaller thresholds than suggested by the proof may be desirable.

Again, it is interesting to compare our results with the WVD approach. For dilation homogeneous operators with index ν, Abramovich and Silverman (1998) showed that, if (soft) thresholding is to be used at *all* levels $j < \log_2(n)$, then near optimality of expected mean-squared estimation error rates required the use of higher thresholds, which in our notation are given by

$$\lambda_j^{\mathrm{WVD}} := \lambda_{n,j} = \hat{\sigma}\tau_j\sqrt{\{2(2\nu+1)\log(n)/n\}}. \tag{30}$$

We have tested our method with two choices of the parameter η: $\eta_{\mathrm{L}} = \sqrt{\{2(2\nu+1)\}}$ as in Abramovich and Silverman (1998) and a smaller value $\eta_{\mathrm{S}} = \sqrt{2}$ that is similar to that of direct estimation (28). The values of corresponding thresholds (for smooth blur) are given in Table 1. In our simulation study (Section 4), we used the WaveD estimator with η_{S} as it led to slightly better results than the conservative choice η_{L}.

4. Numerical performances of the wavelet deconvolution method

We compare several approaches to deconvolution which differ in the degree to which Fourier and wavelet filtering are balanced. On the one hand, we have Wiener-filter-like methods which have no wavelet component but use only Fourier inversion together with a regularization param-

eter (below referred to as the FoRD method). On the other hand we have wavelet decomposition approaches (like the WaveD method) where we perform Fourier inversion with no regularization but use wavelet smoothing to remove noise. Between those two approaches lies the ForWaRD method which combines Fourier regularization with wavelet smoothing. We give only a brief description of the ForWaRD and FoRD methods, referring to Neelamani *et al*. (2004) and references therein for further details.

4.1. Fourier regularized deconvolution

The FoRD estimator of f is defined in the Fourier domain:

$$\tilde{f}_{\alpha l} := g_{\alpha l} y_l, \tag{31}$$

where

$$g_{\alpha l} := \frac{1}{g_l} \frac{|g_l|^2}{|g_l|^2 + \alpha\sigma^2}. \tag{32}$$

Then we take $\hat{f}_\alpha$ as an estimator of f by using the Fourier series with coefficients $\tilde{f}_{\alpha l}$. Here α is a regularization parameter which balances the suppression of noise with signal distortion. Small values of α give an unbiased but noisy estimate whereas large values of α suppress the noise but also distort the signal. We use the terminology FoRD of Neelamani *et al*. (2004) since our comparisons below use the regularization parameter choice in code that was graciously provided by R. Neelamani.

4.2. Wavelet-regularized deconvolution

First apply filtering (31), deriving an estimator $\hat{f}_\alpha$ of f. In the second step, we further smooth $\hat{f}_\alpha$ by using data-driven level-dependent wavelet thresholding (17). Here also α plays the role of a regularization parameter which balances the level of noise and signal distortion. For $\alpha = 0$ the ForWaRD estimator is similar to a compactly supported WVD estimator of f whereas for $\alpha > 0$ the ForWaRD estimator is a hybrid of the FoRD and WVD methods. Although it is difficult to derive an optimal choice of the regularization parameter α (see Neelamani *et al*. (2004)), a data-driven algorithm to compute the ForWaRD estimator is available at `http://www.dsp.rice.edu/software/ward.shtml`.

We compared the WaveD method with the FoRD and ForWaRD methods in a simulation study using the LIDAR target that is depicted in Fig. 1. Performance was tested with different blurring types and noise levels as illustrated in Fig. 3. For each combination of noise level and blurring type we computed the Monte Carlo approximation to the MISE $:= \mathbb{E}\|\hat{f} - f\|_2^2$. Our results are illustrated in Fig. 7 for smooth blur and summarized in Table 2.

4.3. Analysis of the results

Regularized Fourier filtering tends to distort the original signal and the superiority of the WaveD method becomes more apparent as the noise level increases. For smooth blur WaveD outperformed both ForWaRD and FoRD in all cases with larger margins in high noise scenarios. For all the methods we observe smaller margins and poorer performances in the boxcar blur scenario (which confirms a larger DIP). For boxcar blur WaveD outperformed ForWaRD for larger noise levels whereas ForWaRD outperformed WaveD for smaller noise levels. Both WaveD and ForWaRD outperformed FoRD, whose performance is limited because of its linear nature.

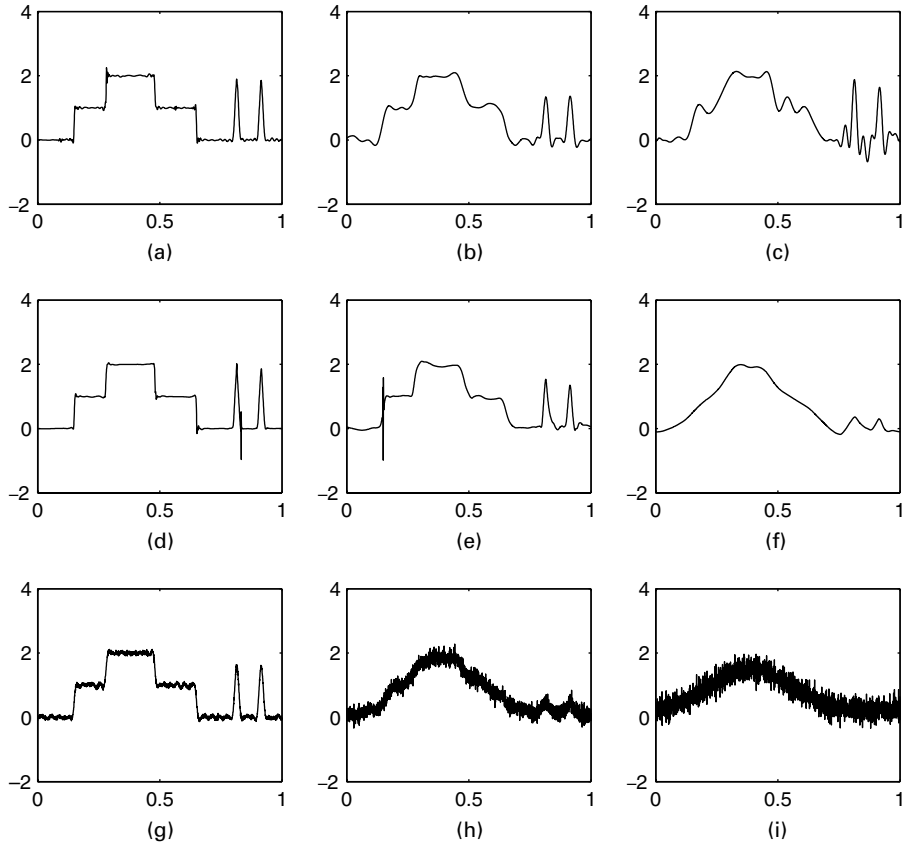

Fig. 7. (a) WaveD (low noise scenario); (b) WaveD (medium noise scenario); (c) WaveD (high noise scenario); (d) ForWaRD (low noise scenario); (e) ForWaRD (medium noise scenario); (f) ForWaRD (high noise scenario); (g) FoRD (low noise scenario); (h) FoRD (medium noise scenario); (i) FoRD (high noise scenario)

Remark 1. Cycle spinning the WaveD estimator mitigates its lack of translation invariance and leads to a better visual appearance and smaller MSE. We refer to Donoho and Raimondo (2004) for an efficient algorithm which cycle-spins the WaveD estimator over all circulant shifts (used in the computations for Table 2 and Fig. 7).

Remark 2. Boundary corrections to deal with non-periodic signals and an extension to the deconvolution of two-dimensional data are currently under investigation by the authors.

5. Asymptotic theory

5.1. Near optimality for a wide range of smoothness classes

Proposition 1. Suppose that we observe the random process (1) with $\sigma = 1$, under assumption C_ν. Let $p > 1$ be an arbitrary number. If f belongs to $B^s_{\pi,r}(T)$ with $\pi \geqslant 1$, $s \geqslant 1/\pi$ and

$$0 < r \leqslant \min\left\{ \frac{p(2\nu+1)}{2(\nu+s)+1}, \frac{(2\nu+1)p-2}{2(s+\nu)-2/\pi+1} \right\}$$

then, for $\eta \geqslant 2\sqrt{\{8\pi(p \vee 2)\}}$ the wavelet-based estimator (17) with threshold (24) with $\hat{\sigma} = 1$ and

Table 2. Monte Carlo approximations to MISE = $\mathbb{E}\|\hat{f} - f\|_2^2$†

Method	*Blur*	*Means for the following levels of noise:*			
		$\sigma_{\text{low}} = 0.05$	$\sigma_{\text{med}} = 0.5$	$\sigma_{\text{high}} = 1$	$\sigma_{\text{lim}} = 1.25$
WaveD	Smooth	**0.0024**	**0.0180**	**0.0388**	**0.0519**
ForWaRD	Smooth	0.0027	0.0208	0.0642	0.0950
FoRD	Smooth	0.0084	0.0906	0.3201	0.3352
WaveD	Boxcar	0.0223	0.0753	**0.0831**	**0.0900**
ForWaRD‡	Boxcar	**0.0110**	**0.0573**	0.0906	0.1030
FoRD	Boxcar	0.0237	0.0950	0.3470	0.3610

†The results are means of 1000 independent simulations of model (1) with $n = 2048$ as in Fig. 3. For each scenario, numbers in bold indicate the method which has the smallest MSE.

maximum level (25) is such that

$$\mathbb{E}\|\hat{f} - f\|_p^p \leqslant C\{n^{-1}\log(n)\}^{\alpha}, \tag{33}$$

$$\alpha = \frac{sp}{2(s+\nu)+1}, \qquad \text{if } s \geqslant (2\nu+1)\left(\frac{p}{2\pi} - \frac{1}{2}\right) \tag{34}$$

and

$$\alpha = \frac{(s - 1/\pi + 1/p)p}{1 + 2(s+\nu-1/\pi)}, \qquad \text{if } \frac{1}{\pi} - \frac{1}{2} - \nu \leqslant s < (2\nu+1)\left(\frac{p}{2\pi} - \frac{1}{2}\right). \tag{35}$$

Remark 3. There is an 'elbow effect' or 'phase transition' in the rates of convergence, switching from condition (34) to condition (35) as the assumed smoothness decreases. The existence of this effect is familiar from the direct observation case (e.g. Donoho *et al.* (1995) and references cited therein, where conditions (34) and (35) are respectively referred to as the 'dense' and 'sparse' cases). The additional presence of the parameter ν makes the sparse case (35) relevant even for quadratic loss, $p = 2$.

Remark 4. For $p = 2$ and smooth convolutions, the rate α and dense case condition (34) are consistent with results of Donoho (1995), Abramovich and Silverman (1998) and Fan and Koo (2002). Pensky and Vidakovic (1999) obtained similar rates (in the density model) with the additional restriction $\pi = 2$ (Hilbert–Sobolev function classes) so that constraint (34) does not appear.

Remark 5. For $p = 2$ Kalifa and Mallat (2003) proposed a related procedure which can be applied for *hyperbolic* convolution where the convolution kernel depends on the sample size. Such convolutions do not satisfy condition C_ν; hence their results are not directly comparable with ours.

Remark 6. For $p = 2$ and severely ill-posed convolutions such as boxcar blur, our results agree with the degree of ill-posedness $\nu = \frac{3}{2}$ that was derived in Johnstone and Raimondo (2002).

Remark 7. For $p \neq 2$ our result seems to be new in the deconvolution context.

The following proposition gives some examples of blurring type where near optimal results of proposition 1 are achievable by our estimator.

Proposition 2.

(a) For ordinary smooth convolution, $|g_l| \sim C|l|^{-\nu}$, where g_l denotes the Fourier coefficients of g and $\nu > 0$, assumption C_ν is satisfied.
(b) For boxcar blur, $g(x) = (1/2a)\mathbb{I}_{[-a,a]}(x)$, where a is a BA number, assumption C_ν is satisfied with $\nu = \frac{3}{2}$.

Remark 8. Combining the results of propositions 1 and 2 for boxcar blur, we see that rates (34) and (35) hold with $\nu = \frac{3}{2}$ provided that the boxcar width is a BA irrational number. In the finite sample implementation of model (1) on a computer, Fourier coefficients g_l are computed up to $l = n$ or $l = -n$, or more precisely for blocks C_j that are wholly contained in $[-n, n]$. Hence condition C_ν needs only to be satisfied for any $j > 0$ where $2^{j+r} \leqslant n$. An examination of the proof of proposition 2, part (b), shows that the latter condition holds for those BA rational numbers that are of order greater than n discussed in Section 2.2.

Remark 9. For almost all irrational numbers (i.e. except for a set of Lebesgue measure 0), the boxcar blur is also known to have a degree of ill-posedness of $\frac{3}{2}$, ignoring logarithmic terms (Johnstone and Raimondo, 2002). Whether the WaveD estimator can be tuned to achieve rates that are similar to expression (3) for boxcar blur in the almost all case remains open.

Remark 10. In the direct estimation setting, alternatives to co-ordinatewise non-linear thresholding also have broad adaptivity properties (e.g. Efromovich (1999)); whether such results extend to the deconvolution is an issue for further work.

5.2. The maxiset approach

Near optimal properties of our proposal are direct applications of the following theorem which has been borrowed from Kerkyacharian and Picard (2000). This theorem gives the 'maxiset' (condition (40)) for a general wavelet estimator of the form (39). It will be applied directly to our procedure as outlined in Appendix A. We refer to Appendix B for condition (63) (known as the Temlyakov property). First, we introduce some notation: μ will denote the measure such that, for $j \in \mathbb{N}$, $k \in \mathbb{N}$,

$$\mu\{(j,k)\} = \|\sigma_j \Psi_{j,k}\|_p^p = 2^{j(p/2-1)} \sigma_j^p \|\Psi\|_p^p,$$

$$l_{q,\infty}(\mu) = \left\{ f, \sup_{\lambda > 0} [\lambda^q \, \mu\{(j,k)/|\beta_{j,k}| > \sigma_j \lambda\}] < \infty \right\}.$$

Theorem 1. Let $p > 1$, $0 < q < p$, $\{\psi_{j,k}, j \geqslant -1, k = 0, 1, \ldots, 2^j\}$ be a periodized wavelet basis of $L^2(T)$ and σ_j be a positive sequence such that the heteroscedastic basis $\sigma_j \psi_{j,k}$ satisfies property (63) in Appendix B. Suppose that Λ_n is a set of pairs (j,k) and c_n is a deterministic sequence tending to 0 with

$$\sup_n [\mu\{\Lambda_n\} c_n^p] < \infty. \tag{36}$$

If, for any pair $\kappa = (j,k) \in \Lambda_n$, we have

$$\mathbb{E}|\hat{\beta}_\kappa - \beta_\kappa|^{2p} \leqslant C(\sigma_j c_n)^{2p}, \tag{37}$$

$$P(|\hat{\beta}_\kappa - \beta_\kappa| \geqslant \eta \sigma_j c_n / 2) \leqslant C(c_n^{2p} \wedge c_n^4) \tag{38}$$

for some positive constants η and C, then the wavelet-based estimator

$$\hat{f}_n = \sum_{\kappa \in \Lambda_n} \hat{\beta}_\kappa \psi_\kappa \, \mathbb{I}(|\hat{\beta}_\kappa| \geqslant \eta \sigma_j c_n) \tag{39}$$

is such that, for all positive integers n,

$$\mathbb{E}\|\hat{f}_n - f\|_p^p \leqslant C c_n^{p-q},$$

if and only if

$$f \in l_{q,\infty}(\mu) \quad \text{and} \quad \sup_n \left(c_n^{q-p} \| f - \sum_{\kappa \in \Lambda_n} \beta_\kappa \psi_\kappa \|_p^p \right) < \infty. \tag{40}$$

Remark 11. Through condition (40) and in the light of Appendix B.1 the theorem gives the maxiset of the method, i.e. the set of functions where the method attains a given rate of convergence. This way of measuring the performances of statistical procedures has been particularly successful in the nonparametric framework. It has often the advantage of giving less arbitrary and pessimistic comparisons of procedures than the minimax approach.

Remark 12. We shall prove (Appendix B.1) that the Besov spaces $B^s_{\pi,r}(T)$ are included in the maxiset defined in expression (40) for q chosen such that c_n^{p-q} provides the rate that is given in inequality (33). In particular, for appropriate choices of j_1 and Λ_n (see expression (42)), we have that

$$\| f - \sum_{\kappa \in \Lambda_n} \beta_\kappa \psi_\kappa \|_p \asymp \| f - P_{V_{j_1}} f \|_p \tag{41}$$

where $P_{V_{j_1}}$ denotes the projection on the space V_{j_1} of the multiresolution analysis that is associated with the wavelet basis. In this case it appears more clearly that the second part of condition (40) is directly linked to standard conditions for membership in Besov spaces. This part is responsible for the condition $s \geqslant 1/\pi$ in the assumptions of proposition 1.

Acknowledgements

We are grateful to all the referees, who provided us with helpful suggestions that have improved the original version significantly. This project began while Gérard Kerkyacharian and Dominique Picard visited the University of Sydney, partly funded by the University of Sydney. IMJ was supported in part by National Science Foundation grant DMS 00-72661 and National Institutes of Health grant ROI EB001988-08.

Appendix A: Proofs

A.1. Outline of the proof of proposition 1

We shall prove that proposition 1 follows from theorem 1. For this, we shall consider the wavelet-based estimator (17) with threshold (24) and maximum level (25). In the light of theorem 1, that is to say that $\sigma_j = \tau_j$ as in equation (23) and

$$\left.\begin{aligned} \Lambda_n &= \{(j,k), -1 \leqslant j \leqslant j_1, 0 \leqslant k \leqslant 2^j\}, \\ c_n &= \{\log(n)/n\}^{1/2}, \\ 2^{j_1} &= O\{n/\log(n)\}^{1/(1+2\nu)}. \end{aligned}\right\} \tag{42}$$

In this setting, and under assumption C_ν, we shall prove the following claims.

(a) Inequalities (37) and (38) hold with $\eta \geqslant 2\sqrt{\{8\pi(p \vee 2)\}}$ (claim 1).
(b) The basis $(\sigma_j \Psi_{jk})$ satisfies condition (63) (see Appendix B.2) as soon as there is a constant C such that, for any finite subset Λ of $\mathbb{N}$,

$$\sum_{j \in \Lambda} 2^j \sigma_j^2 \leqslant C \sup_{j \in \Lambda}(2^j \sigma_j^2), \qquad \text{if } 2 < p < \infty, \tag{43}$$

$$\sum_{j\in\Lambda} 2^{jp/2}\sigma_j^p \leqslant C\ \sup_{j\in\Lambda}(2^{jp/2}\sigma_j^p), \qquad \text{if } 1<p<2 \tag{44}$$

(claim 2). Note that for $p=2$ condition (63) holds without any condition on σ_j.

(c) Conditions (36), (43) and (44) are satisfied (claim 3).

Hence, under the assumptions of proposition 1, theorem 1 applies to the wavelet-based estimator (17) which combined with remarks following theorem 1 gives the rate (33). To complete the proof we shall now prove the claims.

A.1.1. Proof of claim 1

First, we derive the bias and variance of $\hat{\beta}_\kappa$. Taking the expectation in equation (19),

$$\mathbb{E}(\hat{\beta}_\kappa) = \sum_l \frac{h_l}{g_l}\bar{\Psi}_l^\kappa = \beta_\kappa, \tag{45}$$

under the assumptions of proposition 1 we have $\sigma=1$. It follows that

$$B_\kappa = \hat{\beta}_\kappa - \beta_\kappa = \sum_l \frac{y_l - h_l}{g_l}\bar{\Psi}_l^\kappa = n^{-1/2}\sum_l \frac{z_l}{g_l}\bar{\Psi}_l^\kappa = n^{-1/2}\sum_l \frac{\bar{\Psi}_l^\kappa}{g_l} z_l,$$

as the z_ls are independent and identically distributed standard Gaussian random variables:

$$\operatorname{var}(\hat{\beta}_\kappa) = \mathbb{E}(B_\kappa \bar{B}_\kappa) = n^{-1}\sum_l \left|\frac{\bar{\Psi}_l^\kappa}{g_l}\right|^2 \mathbb{E}(z_l)^2 = n^{-1}\sum_l \left|\frac{\Psi_l^\kappa}{g_l}\right|^2. \tag{46}$$

Note that $C_j = \{l : \Psi_l^\kappa \neq 0\}$ and that

$$\Psi_l^\kappa = 2^{-j/2}\theta_{jkl}\hat{\psi}(2^{-j}\cdot 2\pi l), \qquad \theta_{jkl} = \exp(-2\pi \mathrm{i} l k\cdot 2^{-j}), \tag{47}$$

so $|\Psi_l^\kappa| = 2^{-j/2}|\hat{\psi}|(2^{-j}\cdot 2\pi l) \leqslant 2^{-j/2}\|\hat{\psi}\|_\infty = 2^{-j/2}$ since, for the Meyer wavelet, $\|\hat{\psi}\|_\infty = 1$. Using definition (23) and recalling that for the Meyer wavelet $|C_j| = 4\pi\cdot 2^j$:

$$\operatorname{var}(\hat{\beta}_\kappa) \leqslant 2^{-j} n^{-1} \sum_{l\in C_j} |g_l|^{-2} = 4\pi n^{-1}\tau_j^2, \tag{48}$$

after recalling the notation of equation (23). As the $\hat{\beta}_\kappa$s are Gaussian,

$$\mathbb{E}|\hat{\beta}_\kappa - \beta_\kappa|^{2p} \leqslant C_{2p}\operatorname{var}(\hat{\beta}_\kappa)^p \tag{49}$$

which combined with expression (48) yields

$$\mathbb{E}|\hat{\beta}_\kappa - \beta_\kappa|^{2p} \leqslant C_{2p}\left(\frac{\tau_j^2}{n}\right)^p, \tag{50}$$

hence proving inequality (37). Let $Z\sim\mathcal{N}(0,1)$; by using expression (48) we have that

$$P\left(|\hat{\beta}_\kappa - \beta_\kappa| \geqslant \frac{\eta\tau_j c_n}{2}\right) \leqslant 2\ P\left\{Z > \frac{\eta\sqrt{\log(n)}}{2\sqrt{(4\pi)}}\right\} \leqslant \frac{4\sqrt{2}}{\eta\sqrt{\log(n)}} n^{-\eta^2/32\pi}.$$

Hence, for $\eta \geqslant 2\sqrt{\{8\pi(p\vee 2)\}}$ we have proved that

$$P(|\hat{\beta}_\kappa - \beta_\kappa| \geqslant \eta\tau_j c_n/2) = P(|\hat{\beta}_\kappa - \beta_\kappa| \geqslant \lambda_j^{\mathrm{WaveD}}/2) \leqslant c_2 n^{-(p\vee 2)} \tag{51}$$

which proves inequality (38) for the WaveD threshold (24) with $\sigma=1$.

A.1.2. Proof of claim 2

The proof of claim 2 is a direct application of theorem 2 (see Appendix B).

A.1.3. Proof of claim 3

Clearly conditions (43) and (44) will be true for any σ_j of the form $2^{j\nu}$, which follows from assumption C_ν if $\sigma_j = \tau_j$. Next we prove inequality (36); under assumption C_ν, $\sigma_j \asymp 2^{j\nu} C$, $\nu > 0$, we have

$$2^{j_1} \sum_{j=0}^{j_1} 2^{j\nu p} \cdot 2^{j(p/2-1)} \asymp 2^{j_1(\nu p + p/2)}.$$

For $p > 1$, $p\nu + p/2 > 1$ is equivalent to $\nu > 1/p - \frac{1}{2}$. Now by equation (25)

$$2^{j_1(p/2)(2\nu+1)} \asymp c_n^{-p} \asymp \{n/\log(n)\}^{p/2},$$

which proves inequality (36).

A.1.4. Proof of proposition 2

Recall definition (23):

$$\tau_j^2 = |C_j|^{-1} \sum_{C_j} |g_l|^{-2}.$$

Here Ψ is band limited: hence $C_j = \{l : 2^j \leqslant |l| \leqslant 2^{j+r}\}$, for some fixed $r > 0$. To simplify the exposition we shall further assume that $C_j = \{l : 2^j \leqslant l \leqslant 2^{j+r}\}$, noting that, by symmetry, bounds below hold for negative values of l also. Under assumption (*a*): $|g_l| \sim C|l|^\nu$,

$$\tau_j^2 \asymp |C_j|^{-1} \sum_{l=2^j}^{2^{j+r}} l^{2\nu} \asymp 2^{-j} \cdot 2^{j(2\nu+1)} \asymp 2^{2j\nu}, \tag{52}$$

which proves proposition 2, part (a). Under assumption (b), we shall prove that

$$\tau_j^2 = |C_j|^{-1} \sum_{l=2^j}^{2^{j+r}} g_l^{-2} \asymp 2^{3j}, \tag{53}$$

which by identification, $2^{2j\nu} = 2^{3j}$, shows that condition C_ν holds with $\nu = \frac{3}{2}$. Result (53) follows from condition (9) and the following lemma (see Johnstone and Raimondo (2002)). We refer to Section 2.2 for the notion of a BA number.

Lemma 1. Let p/q and p'/q' be successive principal convergents in the continued fraction expansion of a real number a. Let $q \geqslant 4$ and N be a non-negative integer with $N + q < q'$. Then, for BA number a,

$$c_0 q^2 \leqslant \sum_{l=N+1}^{N+q} \|la\|^{-2} \leqslant c_1(a) q^2. \tag{54}$$

Starting at equation (23) and using condition (9), we see that

$$\tau_j^2 \asymp 2^{-j} \sum_{l \in C_j} \left(\frac{l}{\|la\|} \right)^2 \asymp 2^{-j} \cdot 2^{2j} \sum_{l \in C_j} \|la\|^{-2} \asymp 2^j \sum_{l \in C_j} \|la\|^{-2}. \tag{55}$$

Our task, then, is to show that $\Sigma_{l \in C_j} \|la\|^{-2} \asymp 2^{2j}$.

We consider first the upper bound. Let m be the smallest index such that $q_m \geqslant 2^j$. Recall that $C_j = \{l : 2^j \leqslant l \leqslant 2^{j+2}\}$. The geometric growth of the denominators q_n (compare expression (11)) implies that $q_{m+2r} \geqslant 2^r q_m > 2^{j+r}$, so

$$C_j \subset \mathbb{N} \cap [1, q_{m+4}).$$

Introduce intervals $D_0 = \mathbb{N} \cap [1, q_m)$ and $D_\tau = \mathbb{N} \cap [q_{m+\tau-1}, q_{m+\tau})$ for $\tau = 1, \ldots, 4$ which together cover C_j. Since a is BA, there is an integer $K = K(a)$ such that $q_{n+1} \leqslant K q_n$ for all n. Hence there are at most K disjoint blocks of length $q_{m+\tau-1}$ that cover D_τ. Apply lemma 1 to each of these blocks:

$$\sum_{l \in D_\tau} \|la\|^{-2} \leqslant c_1 K q_{m+\tau-1}^2, \qquad 1 \leqslant \tau \leqslant 4,$$

whereas $\Sigma_{l\in D_0}\|la\|^{-2}\leqslant c_1(a)q_m^2$. Since $q_{m+\tau-1}\leqslant K^\tau q_{m-1}$, we combine over τ to obtain

$$\sum_{l\in C_j}\|la\|^{-2}\leqslant\sum_{\tau=0}^{4}\sum_{l\in D_\tau}\|la\|^{-2}\leqslant C\left(K^2+\sum_{\tau=1}^{4}K^{2\tau+1}\right)q_{m-1}^2.$$

Noting that $q_{m-1}\leqslant 2^j$, we recover the upper bound.

For the lower bound, a little care is needed to construct intervals $[N+1,N+q]\subset C_j$ on which to apply the lower bound of condition (54) in lemma 1. Define q_m as before. Set $N=2^j$ and consider the following three cases.

(a) For $q_m>2^{j+1}$, set $q=q_{m-1}$. Since $q_{m-1}\leqslant 2^j$, we have $N+q=2^j+q_{m-1}\leqslant 2^{j+1}<q_m=q'$ and so $[N+1,N+q]\subset C_j$. In addition, $q^2=q_{m-1}^2\geqslant K^{-2}q_m^2$.

(b) For $q_m\leqslant 2^{j+1}$ and $q_{m+1}\geqslant 2q_m+q_{m-1}$ (where the second condition corresponds to $a_{m+1}\geqslant 2$ in expression (5)), set $q=q_m$. We have $N+q=2^j+q_m\leqslant 3\times 2^j$ so $[N+1,N+q]\subset C_j$ and $N+q<2q_m<q_{m+1}=q'$.

(c) Finally, suppose that $q_m\leqslant 2^{j+1}$ and $q_{m+1}=q_m+q_{m-1}$. Now set $q=q_{m+1}$, so that $N+q=2^j+q_{m+1}\leqslant 2^j+q_m+q_{m-1}\leqslant 2^j+2^{j+1}+2^j\leqslant 4\times 2^j$. In addition $N+q<q_m+q_{m+1}\leqslant q_{m+2}=q'$, and $q^2=q_{m+1}^2\geqslant q_m^2$.

In each of cases (a)–(c), we have

$$\sum_{l\in C_j}\|la\|^{-2}\geqslant\sum_{N+1}^{N+q}\|la\|^{-2}\geqslant cq_m^2\geqslant 2^{2j}c.$$

Appendix B

B.1. Embedding of Besov spaces

Our aim here is to investigate which particular periodic Besov space may be embedded in the spaces $l_{q,\infty}(\mu)$ as well as imply the condition

$$\sup_n\left(c_n^{q-p}\left\|f-\sum_{(j,k)\in\Lambda_n}\beta_{j,k}\Psi_{j,k}\right\|_p^p\right)<\infty. \tag{56}$$

Let us recall that we shall concentrate on the case where

$$\mu(j,k)=2^{j(p/2-1)}\tau_j^p,\ \tau_j=2^{j\nu},\ 2^{j_1}=\{n/\log(n)\}^{1/(2\nu+1)},\ c_n=\{\log(n)/n\}^{1/2}\ \text{and}\ p-q=2\alpha.$$

First, we observe that condition (56) will be satisfied when f belongs to $B_{p,\infty}^{(\nu+1/2)(1-q/p)}(T)$. Hence, we only need to prove that $B_{\pi,r}^s(T)$ is included in $B_{p,\infty}^{(\nu+1/2)(1-q/p)}(T)$. For this we shall use two types of Besov embeddings, setting appropriate conditions on s, π, r and q.

(a) In the periodic setting, we have

$$B_{\pi,r}^s(T)\subset B_{\rho,r}^{s''}(T),\qquad\text{if } 0<\rho\leqslant\pi,\ s\geqslant s''. \tag{57}$$

(b) In the general case, we have the standard 'Sobolev embeddings'

$$B_{\pi,r}^s(T)\subset B_{\rho,r}^{\sigma'}(T),\qquad\text{if } \rho>\pi,\ \text{and } s-1/\pi=\sigma'-1/\rho. \tag{58}$$

To prove condition (56), we are interested in taking $\rho=p$. For the case $p\leqslant\pi$, $s>0$ implies that only the dense case (34) can occur; hence we need to prove that $s\geqslant(\nu+\frac{1}{2})(1-q/p)=(\nu+\frac{1}{2})2s/(1+2\nu+2s)$. This is always true for $s>0$ since $1-q/p=2s/(1+2\nu+2s)$.

For the case $p>\pi$, we must prove that, in the dense case (34), $\sigma'\geqslant(\nu+\frac{1}{2})2s/(1+2\nu+2s)$. This is equivalent to $2s\sigma'+(1+2\nu)(1/p-1/\pi)\geqslant 0$. But in this case the left-hand side is greater than $(1+2\nu)(p/\pi-1)(s-1/\pi)\geqslant 0$. In the sparse case (35), we must check that $\sigma'\geqslant(2\nu+1)s'/(1+2\nu+2s')$, but this is equivalent to $(2\nu+1)p/\{(2\nu+1)p-2+2p\sigma'\}\leqslant 1$ or $s\geqslant 1/\pi$.

Now let us turn to the problem of embedding a particular space $B_{\pi,r}^s(T)$ into $l_{q,\infty}(\mu)$. First let us mention that we shall simplify the problem by considering the embedding into

$$l_q(\mu)=\left\{f:f=\sum_{j,k}\left(\frac{|\beta_{j,k}|}{\tau_j}\right)^q\mu(j,k)<\infty\right\}.$$

Using Markov inequality, $l_q(\mu) \subset l_{q,\infty}(\mu)$. We observe that in the dense case where

$$s = (2\nu+1)\left(\frac{p}{2q} - \frac{1}{2}\right) \tag{59}$$

we have

$$l_q(\mu) = B^s_{q,q}(T),$$

hence deriving the advertised rate of convergence since here $p - q = 2sp/\{1 + 2(\nu + s)\}$.

It remains to study the more intricate cases where we do not have $\pi = r = q$.

Proposition 3.

(a) Let q be defined by the relation (59); if $0 < r \leqslant q$ and

$$s \geqslant (2\nu+1)\left(\frac{p}{2\pi} - \frac{1}{2}\right) \tag{60}$$

then

$$B^s_{\pi,r}(T) \hookrightarrow l_q(\mu) = B^s_{q,q}(T).$$

(b) Let q be defined by

$$p - q = \frac{2s'p}{1 + 2(\nu + s')}, \qquad s' = \frac{s - 1/\pi + 1/p}{1 - 2/\{(2\nu+1)p\}}; \tag{61}$$

if $0 < r \leqslant q$ and

$$\frac{1}{\pi} - \frac{1}{2} - \nu < s < (2\nu+1)\left(\frac{p}{2\pi} - \frac{1}{2}\right) \tag{62}$$

then

$$B^s_{\pi,r}(T) \hookrightarrow l_q(\mu) = B^{s'}_{q,q}(T).$$

Remark 13. Case (62) implies that

$$p > \frac{2}{1+2\nu}.$$

For $\nu \geqslant \frac{1}{2}$ this not a restriction, since we are considering $1 < p < \infty$. Moreover, in this case the first member of inequality (62) is always true if we deal with $1 \leqslant \pi$, as $1/\pi - \frac{1}{2} - \nu \leqslant 0$.

Proof. We have

$$\sum_{j,k}\left(\frac{|\beta_{j,k}|}{\tau_j}\right)^q \mu(j,k) = \sum_{j,k}\left(\frac{|\beta_{j,k}|}{\tau_j}\right)^q \tau_j^p \cdot 2^{j(p/2-1)} = \sum_{j,k}|\beta_{j,k}|^q \tau_j^{p-q} \cdot 2^{j(p/2-1)}$$
$$= \sum_{j,k}|\beta_{j,k}|^q \cdot 2^{j\{(\nu+1/2)p - \nu q - 1\}};$$

recalling that q has been chosen in such a way that

$$l_q(\mu) = B^s_{q,q}(T),$$

and using expression (16), we obtain the following characterization of $B^s_{q,q}(T)$:

$$\sum_{j,k}|\beta_{j,k}|^q \cdot 2^{j(s+1/2)q-1} < \infty.$$

We shall now use embeddings (57) and (58), taking $\rho = q$.

(a) If $s \geqslant (2\nu+1)(p/2\pi - \frac{1}{2})$ and $r \leqslant q$ we have $q \leqslant \pi$; hence, using condition (57),

$$B^s_{\pi,r}(T) \hookrightarrow l_q(\mu) = B^s_{q,q}(T).$$

(Let us observe in addition that $p > q \Leftrightarrow s > 0$.)

(b) If $s < (2\nu+1)(p/2\pi - \frac{1}{2})$ and $r \leqslant q$ we shall use condition (58) to find an embedding with a different order of smoothness. This explains our definition of q. Solving

$$s - \frac{1}{\pi} = s' - \frac{1}{q}, \qquad s' = (2\nu+1)\left(\frac{p}{2q} - \frac{1}{2}\right), \quad \pi < q,$$

and using condition (58) gives

$$B^s_{\pi,r}(T) \hookrightarrow B^{s'}_{q,q}(T) = l_q(\mu).$$

We now must check that $p > q$, but this is equivalent to $1/\pi - \frac{1}{2} - \nu < s$.

B.2. Temlyakov inequalities

Let us recall the Temlyakov property for a basis $e_n(x)$ in L^p: there are absolute constants c and C such that, for all $\Lambda \subset \mathbb{N}$,

$$c \sum_{n\in\Lambda} \int |e_n(x)|^p \mathrm{d}x \leqslant \int \Big\{ \sum_{n\in\Lambda} |e_n(x)|^2 \Big\}^{p/2} \mathrm{d}x \leqslant C \sum_{n\in\Lambda} \int |e_n(x)|^p \, \mathrm{d}x$$

or, equivalently,

$$c' \Big\| \Big\{ \sum_{n\in\Lambda} |e_n(x)|^p \Big\}^{1/p} \Big\|_p \leqslant \Big\| \Big\{ \sum_{n\in\Lambda} |e_n(x)|^2 \Big\}^{1/2} \Big\|_p \leqslant C' \Big\| \Big\{ \sum_{n\in\Lambda} |e_n(x)|^p \Big\}^{1/p} \Big\|_p. \tag{63}$$

Obviously the left-hand side is always true for $p \geqslant 2$ with $c = 1$, whereas the right-hand side is always true for $p \leqslant 2$ with $C = 1$. In this section, we shall prove the following result.

Theorem 2. Let ϕ be a scaling function of a multiresolution analysis and ψ the associated wavelet. Let us assume that

$$|\phi(x)| + |\psi(x)| \leqslant \frac{C}{1+|x|}.$$

If there is a constant $C < \infty$ such that for all $A \subset \mathbb{N}$

$$\Big\{ \sum_{j\in A} (2^{j/2}\sigma_j)^{p\wedge 2} \Big\}^{1/(p\wedge 2)} \leqslant C' \Big\{ \sum_{j\in A} (2^{j/2}\sigma_j)^{p\vee 2} \Big\}^{1/(p\vee 2)} \tag{64}$$

then the weighted wavelet basis $\{2^{j/2}\sigma_j \, \psi(2^j x - k),\ j \in \mathbb{N},\ k \in \mathbb{Z}\} \cup \{\sigma_0 \, \phi(x-k), k \in \mathbb{Z}\}$ satisfies the Temlyakov property.

Proof. We start by proving the theorem for the Haar basis. We introduce the weighted Haar basis $(2^{j/2}\sigma_j h_\kappa)$ where as usual $h_\kappa(x) = h_{j,k}(x) = h(2^j x - k)$ and $h(x) = \mathbf{1}_{[0,1]}(2x) - \mathbf{1}_{[0,1]}(2x-1)$.

Let us suppose first that $p \geqslant 2$ and there exists $C < \infty$ such that, for all $A \subset \mathbb{N}$,

$$\Big\{ \sum_{j\in A} (2^{j/2}\sigma_j)^2 \Big\}^{1/2} \leqslant C \Big\{ \sum_{j\in A} (2^{j/2}\sigma_j)^p \Big\}^{1/p}. \tag{65}$$

Typically this is true when $\sigma_j = 2^{j\nu}$. If inequality (65) is true, we have for all $\Lambda \subset \mathbb{N} \times \mathbb{Z}$, pointwise,

$$\Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j \, h_\kappa(x)|^2 \Big\}^{1/2} \leqslant C \Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j \, h_\kappa(x)|^p \Big\}^{1/p}$$

so in this case

$$\Big\| \Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j \, h_\kappa(x)|^2 \Big\}^{1/2} \Big\|_p \leqslant C \Big\| \Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j \, h_\kappa(x)|^p \Big\}^{1/p} \Big\|_p.$$

Using inequality (65) for $p \geqslant 2$,

$$\Big\| \Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j\, h_\kappa(x)|^p \Big\}^{1/p} \Big\|_p \leqslant \Big\| \Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j\, h_\kappa(x)|^2 \Big\}^{1/2} \Big\|_p \leqslant C \Big\| \Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j\, h_\kappa(x)|^p \Big\}^{1/p} \Big\|_p .$$

Now we suppose that $p \leqslant 2$ and that there exists $C' < \infty$ such that, for all $A \subset \mathbb{N}$,

$$\Big\{ \sum_{j\in A} (2^{j/2}\sigma_j)^p \Big\}^{1/p} \leqslant C' \Big\{ \sum_{j\in A} (2^{jp/2}\sigma_j)^2 \Big\}^{1/2}. \tag{66}$$

Then again we have pointwise, for all $\Lambda \subset \mathbb{N} \times \mathbb{Z}$,

$$\Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j\, h_\kappa(x)|^p \Big\}^{1/p} \leqslant C' \Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j\, h_\kappa(x)|^2 \Big\}^{1/2}$$

so in this case

$$\Big\| \Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j\, h_\kappa(x)|^p \Big\}^{1/p} \Big\|_p \leqslant C \Big\| \Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j\, h_\kappa(x)|^2 \Big\}^{1/2} \Big\|_p ;$$

using inequality (66) for $p \leqslant 2$,

$$\frac{1}{C'} \Big\| \Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j\, h_\kappa(x)|^p \Big\}^{1/p} \Big\|_p \leqslant \Big\| \Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j\, h_\kappa(x)|^2 \Big\}^{1/2} \Big\|_p \leqslant \Big\| \Big\{ \sum_{\kappa\in\Lambda} |2^{j/2}\sigma_j\, h_\kappa(x)|^p \Big\}^{1/p} \Big\|_p .$$

Now we shall extend this result to a general wavelet by using the *transfer* lemma (below). For any locally measurable function let us recall the definition of the Hardy–Littlewood maximal function. Let I denote an interval of $\mathbb{R}$ and $|I|$ its Lebesgue measure: for all $x \in \mathbb{R}$,

$$f^*(x) = \sup_{I, x\in I} \left\{ \frac{1}{|I|} \int_I |f(y)|\, \mathrm{d}y \right\}.$$

Lemma 2 (transfer). Let us consider two sequences of functions $(f_n(x))_{n\in\mathbb{N}}$ and $(e_n(x))_{n\in\mathbb{N}}$. Suppose that the sequence $(f_n(x))_{n\in\mathbb{N}}$ satisfies the Temlyakov property and that there exists $A < \infty$ such that for all $n \in \mathbb{N}$

$$|f_n(x)| \leqslant A\, e_n^*(x) \qquad \text{almost everywhere,}$$
$$|e_n(x)| \leqslant A\, f_n^*(x) \qquad \text{almost everywhere.}$$

Then the sequence $(e_n(x))_{n\in\mathbb{N}}$ satisfies the Temlyakov property also.

Theorem 2 follows from lemma 2 since, for $f = 1_{[0,1]}$, $f^*(x) \asymp c(1 \wedge 1/|x|)$, and obviously, for all $x \in \mathbb{R}$, $|f(x)| \leqslant f^*(x)$. Combining this with the assumption of theorem 2 we have that $|\psi(x)| \leqslant C\, h^*(x)$ and $|h(x)| \leqslant C\, \psi^*(x)$, which is obvious.

To complete the proof we derive the lemma.

Proof. The key tool for deriving the transfer lemma is the Fefferman–Stein inequality (Fefferman and Stein, 1971): for all $p, q, 1 < p < \infty, 1 < q \leqslant \infty$, there is a positive constant $C_{p,q} < \infty$ such that

$$\Big\| \Big\{ \sum_n |f_n(x)|^q \Big\}^{1/q} \Big\|_p \leqslant \Big\| \Big\{ \sum_n (f_n^*)^q(x) \Big\}^{1/q} \Big\|_p \leqslant C_{p,q} \Big\| \Big\{ \sum_n |f_n(x)|^q \Big\}^{1/q} \Big\|_p .$$

Using our assumption and the previous inequality we have, for all $q,\ 1<q\leqslant\infty$, and all $\Lambda,\ \Lambda\subset\mathbb{N}$,

$$\begin{aligned}\left\|\left\{\sum_{n\in\Lambda}|f_n(x)|^q\right\}^{1/q}\right\|_p &\leqslant A\left\|\left\{\sum_{n\in\Lambda}|e_n^*(x)|^q\right\}^{1/q}\right\|_p\\ &\leqslant AC_{p,q}\left\|\left\{\sum_{n\in\Lambda}|e_n(x)|^q\right\}^{1/q}\right\|_p\\ &\leqslant A^2C_{p,q}\left\|\left\{\sum_{n\in\Lambda}|f_n^*(x)|^q\right\}^{1/q}\right\|_p\\ &\leqslant A^2C_{p,q}^2\left\|\left\{\sum_{n\in\Lambda}|f_n(x)|^q\right\}^{1/q}\right\|_p.\end{aligned}$$

So for all $q, 1<q\leqslant\infty$,

$$\left\|\left\{\sum_{n\in\Lambda}|f_n(x)|^q\right\}^{1/q}\right\|_p\asymp\left\|\left\{\sum_{n\in\Lambda}|e_n(x)|^q\right\}^{1/q}\right\|_p.$$

Using the previous computation for $q=2$ and $q=p$, we have

$$\left\|\left\{\sum_{n\in\Lambda}|f_n(x)|^p\right\}^{1/p}\right\|_p\asymp\left\|\left\{\sum_{n\in\Lambda}|f_n(x)|^2\right\}^{1/2}\right\|_p$$

and so

$$\left\|\left\{\sum_{n\in\Lambda}|e_n(x)|^p\right\}^{1/p}\right\|_p\asymp\left\|\left\{\sum_{n\in\Lambda}|e_n(x)|^2\right\}^{1/2}\right\|_p.$$

B.3. Vaguelette properties

We use $(u_\kappa,\kappa\in I)$ to denote a generic system of candidate vaguelettes on the circle T. With $u_\kappa(t)=\Sigma u_l^\kappa e_l(t)$, then u_l^κ stands for $s_{jl}\Psi_l^\kappa$ in the case of the $(\mathcal{V}_\kappa)$ system and Ψ_l^κ/s_{jl} in the case of the $(\mathcal{U}_\kappa)$ system.

Adapting the definition of Meyer and Coifman (1997), chapter 8, page 56, we say that $\{u_\kappa,\kappa\in I\}$ is a system of periodic vaguelettes on T if there are exponents $0<\beta<\alpha$ and a constant C such that

(a) $|u_\kappa(t)|\leqslant 2^{j/2}C(1+|2^jt-k|)^{-1-\alpha},\ t\in T$,
(b) $\int_T u_\kappa(t)\,\mathrm{d}t=0$ and
(c) $|u_\kappa(s)-u_\kappa(t)|\leqslant 2^{j(1/2+\beta)}C|s-t|^\beta, s,t\in T$.

(In what follows, $\alpha=1$ and $0<\beta<1$.)

The proof of Meyer and Coifman (1997), theorem 2, page 56, goes through essentially unchanged for $L_2(T)$ under conditions (a)–(c) and so, for every sequence (α_κ),

$$\|\sum\alpha_\kappa\,u_\kappa(t)\|_2\leqslant C'\,\|\alpha_\kappa\|_{l_2}.$$

From the remarks around Donoho (1995), theorem 2, this is sufficient for the Riesz basis property.

It remains, then, to verify conditions (a)–(c). Condition (b) is immediate, since for the Meyer wavelet $\Psi_{0,0}^\kappa=\hat{\psi}_\kappa(0)=0$. For the Hölder condition (c),

$$\Delta=\frac{|u_\kappa(s)-u_\kappa(t)|}{|s-t|^\beta}\leqslant\sum_l|u_l^\kappa|\frac{|e_l(s)-e_l(t)|}{|s-t|^\beta}\leqslant 2\pi\sum_l|l|^\beta|u_l^\kappa|, \tag{67}$$

as follows by considering $|s-t|^{-1}\leqslant|l|$ and $|s-t|^{-1}>|l|$ separately. Writing $u_l^\kappa=\gamma_l\Psi_l^\kappa$, we have

$$|u_l^\kappa|=|\gamma_l|\cdot 2^{-j/2}|\hat{\psi}|(2^{-j}\cdot 2\pi l)\leqslant\begin{cases}C|\gamma_l|\cdot 2^{-j/2} & l\in C_j,\\ 0 & \text{otherwise.}\end{cases}$$

If $|g_l|\sim C|l|^{-\nu}$, then set $\kappa_j=2^{-j\nu}C$ and observe that, for $l\in C_j$, we have $s_{jl}\asymp 1$ and so for either $(\mathcal{U}_\kappa)$ or $(\mathcal{V}_\kappa)$ systems $|\gamma_l|\asymp 1$ for $l\in C_j$. Combining with the two previous displays,

$$\Delta\leqslant C\cdot 2^{-j/2}\sum_{l\in C_j}|l|^\beta\leqslant C\cdot 2^{-j/2}\cdot 2^j\cdot 2^{j\beta}=C\cdot 2^{j(1/2+\beta)}. \tag{68}$$

For condition (a), we first observe that $u_{jk}(t) = u_{j,0}(t - 2^{-j}k)$, and so it suffices to show, for $\kappa = (j, 0)$, that

$$(2^j t \wedge 1)^2 |u_\kappa(t)| \leqslant 2^{j/2} C, \qquad |t| \leqslant \tfrac{1}{2}. \tag{69}$$

For $|t| \leqslant \frac{1}{2}$, we have $t^2 \sim |1 - \exp(-2\pi i t)|^2$ and, setting $\Delta f_l = f_{l+1} - f_l$ and $\Delta^2 f_l = \Delta(\Delta f_l)$,

$$\{1 - \exp(-2\pi i t)\}^2 \sum_l u_l^\kappa \, e_l(t) = \sum_l (\Delta^2 u_l^\kappa) \, e_l(t).$$

Set $r_{jl} = \hat{\psi}(2^{-j} \cdot 2\pi l)$; from formula (47) for Ψ_l^κ, we obtain

$$u_l^\kappa = \begin{cases} 2^{-j/2} w_l & l \in C_j, \\ 0 & \text{otherwise}, \end{cases}$$

with

$$w_l = \begin{cases} r_{jl} s_{jl} & \text{for } \mathcal{V}_\kappa, \\ r_{jl} / s_{jl} & \text{for } \mathcal{U}_\kappa. \end{cases}$$

If we suppose that

$$|\Delta^2 w_l| \leqslant C \cdot 2^{-2j}, \qquad l \in C_j, \tag{70}$$

then from our previous remarks

$$t^2 |u_\kappa(t)| \leqslant C \sum_{l \in C_j} 2^{-j/2} |\Delta^2 w_l| \leqslant C \cdot 2^{-j/2} \cdot 2^{-2j} \cdot |C_j| \leqslant C \cdot 2^{-3j/2}.$$

If $|t| \leqslant 2^{-j}$, we simply bound $|u_\kappa(t)| \leqslant \Sigma |u_l^\kappa|$ and retracing the argument from condition (67) to condition (68) with $\beta = 0$ we obtain condition (69).

To establish condition (70), observe first that, since $r_l = \hat{\psi}(2^{-j} \cdot 2\pi l)$, we have $\Delta^r r_l \leqslant (2\pi \cdot 2^{-j})^r \|\hat{\psi}^{(r)}\|_\infty \leqslant 2^{-rj} C$ for $r = 0, 1, 2$. Some calculation shows that

$$\Delta^2(r_l s_l) = \Delta^2 r_l s_{l+2} + 2 \, \Delta r_l \, \Delta s_{l+1} + r_l \, \Delta^2 s_l.$$

$$\Delta^2\left(\frac{r_l}{s_l}\right) = \frac{\Delta^2 r_l}{s_{l+2}} - 2\frac{\Delta r_l}{s_{l+1}} \frac{\Delta s_{l+1}}{s_{l+2}} + 2\frac{r_l}{s_l} \frac{\Delta s_l}{s_{l+1}} \frac{\Delta s_{l+1}}{s_{l+2}} - \frac{r_l}{s_l} \frac{\Delta^2 s_l}{s_{l+2}},$$

and now condition (70) may be seen to follow from condition (27).

References

Abramovich, F. and Silverman, B. (1998) Wavelet decomposition approaches to statistical inverse problems. *Biometrika*, **85**, 115–129.

Bertero, M. and Boccacci, P. (1998) *Introduction to Inverse Problems in Imaging*. Philadelphia: Institute of Physics.

Buckheit, J., Chen, S., Donoho, D. and Johnstone, I. (1995) Wavelab reference manual. Stanford University, Stanford. (Available from `http://www-stat.stanford.edu/wavelab/`.)

Coifman, R. and Donoho, D. (1995) Translation-invariant de-noising. *Lect. Notes Statist.*, **103**, 125–150.

Donoho, D. (1995) Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Appl. Comput. Harm. Anal.*, **2**, 101–126.

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

Donoho, D. L. and Raimondo, M. (2004) Translation invariant deconvolution in a periodic setting. *Int. J. Wavlts Multiresoln Inform. Process.*, to be published.

Efromovich, S. (1999) Quasi-linear wavelet estimation. *J. Am. Statist. Ass.*, **94**, 189–204.

Fan, J. and Koo, J. (2002) Wavelet deconvolution. *IEEE Trans. Inform. Theory*, **48**, 734–747.

Fefferman, C. and Stein, E. (1971) Some maximal inequalities. *Am. J. Math.*, **93**, 107–115.

Harsdorf, S. and Reuter, R. (2000) Stable deconvolution of noisy lidar signals. Oldenburg University, Oldenburg. (Available from `http://las.physik.uni-oldenburg.de/projekte/earsel/4th_workshop_paper/harsdorf.pdf`.)

Hernàndez, E. and Weiss, G. (1996) *A First Course on Wavelets*. Boca Raton: CRC Press.

Jain, A. (1989) *Fundamentals of Digital Image Processing*. Englewood Cliffs: Prentice Hall.

Je Park, Y., Whoe Dho, S. and Jin Kong, H. (1997) Deconvolution of long-pulse lidar signals with matrix formulation. *Appl. Opt.*, **36**, 5158–5161.
Johnstone, I. M. and Raimondo, M. (2004) Periodic boxcar deconvolution and diophantine approximation. *Ann. Statist*., **32**, no. 5, in the press.
Kalifa, J. and Mallat, S. (2003) Thresholding estimators for linear inverse problems and deconvolutions. *Ann. Statist.*, **31**, 58–109.
Kerkyacharian, G. and Picard, D. (2000) Thresholding algorithms and well-concentrated bases. *Test*, **9**, 283–344.
Khinchin, A. Y. (1997) *Continued Fractions*. New York: Dover Publications.
Kolaczyk, E. (1994) Wavelet methods for the inversion of certain homogeneous linear operators in the presence of noisy data. *PhD Dissertation*. Department of Statistics, Stanford University, Stanford.
Lang, S. (1966) *Introduction to Diophantine Approximations*. New York: Springer.
Lee, N.-Y. and Lucier, B. J. (2001) Wavelet methods for inverting the radon transform with noisy data. *IEEE Trans. Image Process.*, **10**, 79–94.
Mallat, S. (1999) *A Wavelet Tour of Signal Processing*, 2nd, expanded, edn. New York: Academic Press.
Meyer, Y. (1992) *Wavelets and Operators*, vol. 1. Cambridge: Cambridge University Press.
Meyer, Y. and Coifman, R. (1997) *Wavelets: Calderón-Zygmund and Multilinear Operators*. Cambridge: Cambridge University Press.
Neelamani, R., Choi, H. and Baraniuk, R. (2004) Forward: Fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Trans. Signal Process.*, **52**, 418–433.
Pensky, M. and Vidakovic, B. (1999) Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, **27**, 2033–2053.
Shen, X. A. and Walter, G. G. (2002) Meyer wavelet regularization. *Numer. Funct. Anal. Optimzn*, **23**, 195–215.
Walter, G. G. (1994) *Wavelet and Other Orthogonal Systems with Applications*. Boca Raton: CRC Press.
Walter, G. and Shen, X. (1999) Deconvolution using the Meyer wavelet. *J. Integr. Eqns Appl.*, **11**, 515–534.