INVERTING NOISY INTEGRAL EQUATIONS USING WAVELET EXPANSIONS: A CLASS OF IRREGULAR CONVOLUTIONS

PETER HALL, FRITS RUYMGAART[1], ONNO VAN GAANS, AND ARNOUD VAN ROOIJ

Australian National University, Texas Tech. University, and Katholieke Universiteit Nijmegen

Suppose a random sample is observed from a density which is a known transformation of an unknown underlying density to be recovered. Expansion of this unknown density in a wavelet basis yields Fourier coefficients that can be reexpressed in terms of the sampled density and an extension of the adjoint of the inverse of the operator involved. This seems to yield a new approach to inverse estimation. Focusing on deconvolution optimal error rates are obtained in the case of certain irregular kernels like the boxcar that cannot easily be dealt with by classical techniques or by Donoho's (1995) wavelet-vaguelette method.

AMS subject classifications: 42C15, 45E10, 46N30, 62G07.
Keywords and phrases: inverse estimation, wavelet expansion, deconvolution, irregular kernel.

1 Introduction

When a smooth input signal is to be recovered from indirect, noisy measurements, Hilbert space methods based on a regularized inverse of the integral operator involved usually yield optimal rates of convergence of the mean integrated squared error (MISE). Statistical theory can be conveniently developed exploiting Halmos' (1963) version of the spectral theorem (van Rooij & Ruymgaart (1996), van Rooij, Ruymgaart & van Zwet (1998)). In practice, however, the input signal often is not regular like, for instance, in dynamical systems where it might be a pulse function. In such cases the traditional recovery technique may fail to capture the local irregularities of the input. Difficulties also arise in instances where the kernel of the integral operator itself displays a certain lack of smoothness. Whenever one has to deal with irregularities, wavelet methods seem pertinent. In classical, direct estimation of discontinuous densities Hall & Patil (1995) successfully apply a wavelet expansion. For certain inverse estimation models Donoho (1995), in a seminal paper, proposes a wavelet-vaguelette decomposition for optimal recovery of spatially inhomogeneous inputs. In both papers nonlinear

[1]Supported by NSF grant DMS 95-94485 and NSA grant MDA 904-99-1-0029. .

techniques are used. Donoho (1995) points out that his method remains essentially restricted to so-called renormalizable problems (Donoho & Low (1992)). Convolution with the "boxcar" (the indicator of the unit interval) is an example of an operator that is excluded.

In this paper we propose an alternative approach to statistical inverse problems with a special view towards irregularities, by using a wavelet expansion where an extension of the inverse operator appears in the Fourier coefficients. More precisely, let $K : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ be a bounded, injective integral operator and consider the equation

$$g = Kf, \quad f \in L^2(\mathbb{R}). \tag{1.1}$$

Suppose we can find an operator $\mathcal{D}$ with dense domain so that its adjoint $\mathcal{D}^*$ is well-defined. Let $\mathcal{F} \subset L^2(\mathbb{R})$ and suppose that the domain of $\mathcal{D}$ contains $K\mathcal{F}$ and that $\mathcal{D}$ satisfies $\mathcal{D}Kf = f$, $f \in \mathcal{F}$; in other words, let $\mathcal{D}$ be an extension of the restriction of K^{-1} to $K\mathcal{F}$. Given any orthonormal basis ψ_λ, $\lambda \in \Lambda$, we have for $f \in \mathcal{F}$ the expansion

$$f = \sum_{\lambda\in\Lambda} \langle f, \psi_\lambda\rangle \psi_\lambda = \sum_{\lambda\in\Lambda} \langle \mathcal{D}Kf, \psi_\lambda\rangle \psi_\lambda = \sum_{\lambda\in\Lambda} \langle g, \mathcal{D}^*\psi_\lambda\rangle \psi_\lambda, \tag{1.2}$$

provided that the ψ_λ are in the domain of $\mathcal{D}^*$. Since in practice g is imperfectly known, it may be much better to deal with $\mathcal{D}^*\psi_\lambda$ than with $\mathcal{D}g$. Also, the $\mathcal{D}^*\psi_\lambda$ are independent of the specific function f and hence can be used for the entire class $\mathcal{F}$. The use of generalized Fourier coefficients to obtain convergence rates as such is, of course, not new and appears for instance in Wahba (1977) and Wahba & Wang (1990).

In simpler situations it will be possible to choose ψ_λ in the domain of K^{-1}. Calculation of the Fourier coefficients will then be usually performed in the spectral domain by application of Halmos' (1963) spectral theorem, mentioned before, coupled with the polar decomposition (Riesz & Nagy (1990)).

It is far too ambitious to deal with (1.1) for arbitrary K and we will focus on examples of operators whose inverses are suitably related to certain differential operators. In Section 2 we will see that the boxcar convolution and the Abel type integral operator in Wicksell's problem are in this class, and that recovery of the forcing term in certain dynamical systems is a prototype. We will, moreover, restrict ourselves to indirect density estimation. Hence we will assume $\mathcal{F}$ to be a class of square integrable densities, and $g = Kf$ is also supposed to be a density. The data consist of an i.i.d. sample $X_1, ..., X_n$ from g, with generic sample element X. An estimator of f is obtained by replacing the Fourier coefficients in (1.2) with their estimators and by truncation and data-driven thresholding like in the direct case (Donoho, Johnstone, Kerkyacharian & Picard (1996), Hall & Patil (1995)).

The boxcar convolution $g = \mathbf{1}_{[0,1]} * f$, mentioned before, provides a good example of the difficulties, both statistical and analytical, that one may encounter when irregularities are involved. In Section 3 it will be shown that for smooth input functions the spectral cut-off type regularized inverse estimator does not yield optimal convergence rates of the MISE over most of the smoothness range. Furthermore, by directly expressing the convolution in terms of an indefinite integral of f, it can be easily seen that inversion boils down to a sum of shifted derivatives of the image g, provided that f has finite support. The generic way of solving a convolution, however, is by transformation to the frequency domain via the Fourier transform, where in this particular case the inverse reduces to division by the characteristic function of $\mathbf{1}_{[0,1]}$, i.e. to division by $e^{\frac{1}{2}it}\text{sinc}\frac{1}{2}t$, where

$$\text{sinc}\, x := \frac{\sin x}{x},\ x \neq 0,\ \ \text{sinc}\, 0 := 1. \tag{1.3}$$

Hence we divide by a function that has zeros at $2k\pi$, $k \in \mathbb{Z}$, and zeros at $\pm\infty$. It is a fair conjecture, corroborated by the recovery via the direct method in the time domain, that the latter zeros represent differentiation in the actual inversion procedure, but the interpretation of the other zeros is not at all immediate. As a referee pointed out, however, a much more sophisticated kind of harmonic analysis has been developed by S. Mallat and his students to deal with such transfer functions. This analysis involves wavelet bases with dyadic decomposition tailored to the problem at hand. Here we propose a different approach. Convolutions are a very important class of operators and it would be interesting to classify their inverses exploiting the properties in the frequency domain. This is still too ambitious but in Section 4 inverses are obtained for a subclass that contains the boxcar.

Finally, in Section 5 we compute the MISE for indirect density estimators constructed by means of the wavelet method (1.2). As has already been observed above, we will restrict ourselves to operators K for which K^{-1} is a kind of differential operator. Calculations for the MISE for both smooth and discontinuous input functions can be patterned on those in Hall & Patil (1995). Due to space limitations we will restrict ourselves to smooth input functions, and how to obtain optimal rates in the boxcar example.

2 Examples

It will often be convenient to precondition and replace the original equation with an equivalent equation

$$p := Tg = TKf =: Rf, \quad f \in L^2(\mathbb{R}), \tag{2.1}$$

where $T : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is a bounded injective operator that we can choose at our convenience. Setting $T = K^*$ would yield a strictly positive Hermitian operator R. In Example 2.2 preconditioning with K itself will be convenient. Preconditioning will not in general introduce extra, undue ill-posedness.

Example 2.1 Let us first consider the boxcar convolution and assume that f has support in a compact interval $[A, B]$. If F is an indefinite integral of f, it is immediate that $g(x) = (\mathbf{1}_{[0,1]} * f)(x) = F(x) - F(x-1)$, which entails $g'(x) = f(x) - f(x-1)$. Choosing an integer I with $I > B - A - 1$, it follows that

$$f(x) = \sum_{i=0}^{I} g'(x - i), \quad x \in [A, B]. \tag{2.2}$$

The condition on f may be weakened at the cost of more involved technicalities.

Example 2.2 Next let us consider Wicksell's unfolding problem where the density of the radii of not directly observable spherical particles is to be recovered from a sample of planar cuts. The relation between the density f of the squares of the radii of the spheres and the density g of the squares of the observed radii of the cuts is given by

$$g(x) = \mu \int_x^1 \frac{f(y)}{\sqrt{y - x}} dy =: \mu(Kf)(x), \quad 0 \le x \le 1, \quad f \in L^2([0, 1]), \tag{2.3}$$

assuming that the radii of the spheres are all smaller than 1, and where μ is a constant that for simplicity we will assume to be known. This model plays a role in stereology and medicine. For some recent results we refer to Nychka, Wahba, Goldfarb & Pugh (1984), Silverman, Jones, Wilson & Nychka (1990), and Groeneboom & Jongbloed (1995). We have already mentioned that inversion of (2.3) is included in Donoho (1995) as a special case. Yet we want to include it here to show that it also fits in our framework. This is due to the circumstance that the operator K represents a fractional integration, meaning that preconditioning with $T = K$ yields the equivalent equation $p := Kg = \mu K^2 f =: \mu Rf$, where

$$(Rf)(x) = \pi \int_0^1 \mathbf{1}_{[x,1]}(y) f(y) dy = \pi\{F(1) - F(x)\}, \ 0 \le x \le 1, \tag{2.4}$$

where F is an indefinite integral of f. This means that

$$f(x) = -\frac{1}{\mu\pi} p'(x), \quad 0 \le x \le 1. \tag{2.5}$$

Example 2.3 Let D^j be the j-th derivative and consider a dynamical system driven by the differential equation $\sum_{j=0}^{J} c_j (D^j g)(x) = bf(x)$, $x \geq 0$, under the usual initial conditions and conditions on the given numbers $b, c_0, ..., c_J$. The forcing term f is unknown and to be recovered from g. Again f is supposed to have support in a compact interval $[A, B] \subset [0, \infty)$. Although in this case no real inversion is involved, the relationship

$$f(x) = \frac{1}{b} \sum_{j=0}^{J} c_j (D^j g)(x) \mathbf{1}_{[A,B]}(x), \quad x \geq 0, \tag{2.6}$$

between f and g still involves an unbounded operator. In practice the noisy data on g usually lead to a regression model.

We intend to show that in all three examples the exact relation between f and g is of the form $f = \mathcal{D}g$ where $\mathcal{D}$ is a differential operator of the following type:

$$(\mathcal{D}g)(x) = \sum_{j=0}^{J} \sum_{i=0}^{I} c_{ji}(x)(D^j g)(x - a_{ji}), \quad x \in \mathbb{R}. \tag{2.7}$$

Here I and J are nonnegative integers, each a_{ji} is a real number, each c_{ji} is a real-valued compactly supported function with a continuous j-th derivative. (I, J, a_{ji} and c_{ji} may depend on the interval $[A, B]$ but not on the specific f.)

Example 2.1 is easy. Letting I be larger than $(B + 1) - (A - 1) - 1$ we have $f(x) = \sum_{i=0}^{I} g'(x-i)$ for all $x \in [A - 1, B + 1]$. Hence, if we choose any continuously differentiable function $c : \mathbb{R} \to \mathbb{R}$ with $c(x) = 1$ for $x \in [A, B]$ and $c(x) = 0$ for $x \notin [A - 1, B + 1]$, we find

$$f(x) = \sum_{i=0}^{I} c(x) g'(x-i), \quad x \in \mathbb{R}. \tag{2.8}$$

Example 2.2 requires some poetic license. Firstly, we want to recover f from p, not from g (see (2.5)). Secondly, we need to interpret f as a function defined not only on $[0, 1]$ but on all of $\mathbb{R}$ (and vanishing outside $[0, 1]$). Extending p to a continuous function on $\mathbb{R}$ that is constant on $(-\infty, 0]$ and on $[1, \infty)$ we have $f = -(\mu\pi)^{-1} p'$ (almost everywhere) on $\mathbb{R}$. Then, choosing any compactly supported, continuously differentiable function $c : \mathbb{R} \to \mathbb{R}$ with $c(x) = 1$ for all $x \in [0, 1]$, we have

$$f(x) = -(\mu\pi)^{-1} c(x) p'(x), \quad x \in \mathbb{R}. \tag{2.9}$$

Example 2.3 is quite similar. Again we view f as a function defined on $\mathbb{R}$. We extend g to a $J - 1$ times differentiable function on $\mathbb{R}$, satisfying $\sum_{j=0}^{J} c_j g^{(j)}(x) = 0$, $x < 0$. (This is always possible.) Then $f =$

$b^{-1}\sum_{j=0}^{J} c_j g^{(j)}$ a.e. on $\mathbb{R}$. If c is any compactly supported, J times continuously differentiable function with $c(x) = 1$ for $x \in [A, B]$, then

$$f(x) = \frac{1}{b}\sum_{j=0}^{J} c_j c(x) g^{(j)}(x), \quad x \in \mathbb{R}. \tag{2.10}$$

3 A traditional approach to the boxcar convolution problem

In this section we will elaborate on the statistical shortcomings of a traditional recovery of the input when convoluted with the boxcar. If K is convolution with $\mathbf{1}_{[0,1]}$ the adjoint K^* is convolution with $\mathbf{1}_{[-1,0]}$. Preconditioning with K^* yields the equivalent equation

$$p = \mathbf{1}_{[-1,0]} * g = \Delta * f, \quad \Delta := \mathbf{1}_{[-1,0]} * \mathbf{1}_{[0,1]}, \tag{3.1}$$

where f is an unknown density in $L^2(\mathbb{R})$, and where the sample is from g. Application of the Fourier-Plancherel transform $F : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ yields (see also (1.3))

$$Rf := F^{-1}(\tilde{\Delta} \cdot Ff), \quad f \in L^2(\mathbb{R}), \quad \tilde{\Delta}(t) = \operatorname{sinc}^2 \frac{1}{2}t, \quad t \in \mathbb{R}. \tag{3.2}$$

($\sim$ indicates the Fourier Transformation.)

An unbiased and $\sqrt{n}$-consistent estimator of p is given by

$$\hat{p}(x) := n^{-1}\sum_{k=1}^{n} \mathbf{1}_{[-1,0]}(x - X_k), \qquad x \in \mathbb{R},$$

with Fourier transform $(F\hat{p})(t) = \widetilde{\mathbf{1}_{[-1,0]}}(t)\hat{\chi}(t)$, $t \in \mathbb{R}$, where

$$\hat{\chi}(t) := \frac{1}{n}\sum_{k=1}^{n} e^{\mathrm{i}tX_k}, \quad t \in \mathbb{R}, \tag{3.3}$$

is the empirical characteristic function. Since $\hat{p}$ is only an approximation of p, the unbounded inverse of R requires regularization. Spectral cut-off regularization yields an estimator of type

$$\hat{f}_\alpha := F^{-1}\left(\frac{1}{\tilde{\Delta}}\mathbf{1}_{\{\tilde{\Delta} \geq \alpha\}} F\hat{p}\right) = \frac{1}{\sqrt{2\pi}} F^{-1}\left(\frac{\hat{\chi}}{\widetilde{\mathbf{1}_{[0,1]}}}\mathbf{1}_{\{\tilde{\Delta} \geq \alpha\}}\right), \quad \alpha > 0, \tag{3.4}$$

for the input f. The formula shows that preconditioning has not introduced undue ill-posedness, indeed, since eventually we divide only by $\widetilde{\mathbf{1}_{[0,1]}}$, which is the characteristic function of the original convolution kernel.

To compute the MISE we will assume that the density f is in the class $\mathcal{F}_\nu$ of all probability densities with Fourier transform bounded by $C(1+|t|)^{-\nu}$, $t \in \mathbb{R}$, for some $\nu > \frac{1}{2}$ and $0 < C < \infty$. We have

$$
\begin{aligned}
\sup_{f\in\mathcal{F}_\nu} \mathbf{E}||\hat{f}_\alpha - f||^2 &= \sup_{f\in\mathcal{F}_\nu} \{\mathbf{E}||\hat{f}_\alpha - \mathbf{E}\hat{f}_\alpha||^2 + ||\mathbf{E}\hat{f}_\alpha - f||^2\} \approx \\
&\approx \frac{1}{n} \int_{\{\tilde{\Delta}\geq\alpha\}} \frac{(1-|\tilde{g}(t))|^2}{\alpha} dt + \int_{\{\tilde{\Delta}\leq\alpha\}} |t|^{-2\nu} dt \approx \\
&\approx \frac{1}{n\alpha\sqrt{\alpha}} + \sum_{k=1}^{[1/\sqrt{\alpha}]} \int_{k\pi(1-\sqrt{\alpha})}^{k\pi(1+\sqrt{\alpha})} t^{-2\nu} dt + \int_{[1/\sqrt{\alpha}]}^{\infty} t^{-2\nu} dt \approx \\
&\approx \frac{1}{n\alpha\sqrt{\alpha}} + \sqrt{\alpha} \int_{1}^{[1/\sqrt{\alpha}]} t^{-2\nu} dt + \int_{[1/\sqrt{\alpha}]}^{\infty} t^{-2\nu} dt,
\end{aligned} \tag{3.5}
$$

where "$\approx$" indicates that the l.h.s. and r.h.s are of the same order as $\alpha \downarrow 0$. The following result is now immediate.

Theorem 3.1 *For suitable choice of $\alpha = \alpha_{\nu,n} \downarrow 0$, as $n \to \infty$, we have*

$$
\sup_{f\in\mathcal{F}_\nu} \mathbf{E}||\hat{f}_\alpha - f||^2 \approx \begin{cases} n^{-(2\nu-1)/(2\nu+2)}, & 1/2 < \nu < 1 \\ n^{-\frac{1}{4}}, & \nu \geq 1 \end{cases} . \tag{3.6}
$$

Next let us adopt the lower bound to the minimax MISE in van Rooij & Ruymgaart (1998), and let $\mathcal{T}$ denote the class of all $L^2(\mathbb{R})$-valued estimators with finite expected squared norm. The symbol C will be used as a generic constant. In the present case we then arrive at

$$
\begin{aligned}
\inf_{T\in\mathcal{T}} \sup_{f\in\mathcal{F}_\nu} \mathbf{E}||T-f||^2 &\geq C\int_{-\infty}^{\infty} \frac{(1+|t|)^{-2\nu}}{1+n(\text{sinc } \frac{1}{2}t)^2(1+|t|)^{-2\nu}} dt \geq \\
&\geq Cn^{-(2\nu-1)/(2\nu+2)}, \quad \nu \geq \tfrac{1}{2}.
\end{aligned} \tag{3.7}
$$

These results imply that for $\nu > 1$, i.e. for most of the smoothness range, the convergence rate of the MISE for the spectral cut-off type estimators is suboptimal. It should be noted that the smoothness class $\mathcal{F}_\nu$ is somewhat different from the smoothness classes in terms of derivatives that one usually finds in the literature. See also Section 5.

For regular kernels spectral cut-off estimators in general obtain the optimal rate (van Rooij & Ruymgaart (1996)). A regular kernel has a Fourier transform that decays monotonically to zero in the tails. In such a regular case the summation in the third line of (3.5), which is due to the oscillations of $\tilde{\Delta}$ between its zeros, would not have been present and the optimal rate would indeed have emerged.

4 Exact inverses of certain irregular convolution operators

In this section we present some convolution operators K that have inverses of the type described in (2.7). Our function f has its support contained in a compact interval $[A, B]$.

Example 4.1 First, let K be convolution with a kernel of the type

$$k(x) = \sum_{n=1}^{N} c_n \mathbf{1}_{[s_{n-1}, s_n]}(x), \quad x \in \mathbb{R}, \tag{4.1}$$

where $c_1, \ldots, c_N$ and $s_0, s_1, \ldots, s_N$ are given numbers and

$$c_1 = 1, \quad 0 = s_0 < s_1 < \ldots < s_N. \tag{4.2}$$

Set $\gamma_n = c_n - c_{n+1}$ for $n = 1, \ldots, N-1$ and $\gamma_N = c_N$. If F is an indefinite integral of f, then for $x \in \mathbb{R}$ we have
(4.3)

$$\begin{aligned} g(x) \;&:=\; (k * f)(x) = \sum_{n=1}^{N} c_n \int_{s_{n-1}}^{s_n} f(x-t)dt = \\ &= \sum_{n=1}^{N} c_n \left(F(x-s_{n-1}) - F(x-s_n)\right) = F(x) - \sum_{n=1}^{N} \gamma_n F(x-s_n). \end{aligned}$$

Thus,

$$g' = f - \sum_{n=1}^{N} \gamma_n f(\bullet - s_n) = f - \mu * f \tag{4.4}$$

if we define μ to be the real-valued measure concentrated on the finite set $\{s_1, \ldots, s_N\}$ with $\mu(\{s_n\}) = \gamma_n$ for each n.

Let $\mu^{*1}, \mu^{*2}, \ldots$ be the convolution powers of μ, i.e., $\mu^{*1} := \mu$, $\mu^{*(m+1)} := \mu * \mu^{*m}$ for $m \in \mathbb{N}$. It follows from (4.4) that

$$f = g' + \sum_{m=1}^{M-1} \mu^{*m} * g' + \mu^{*M} * f \tag{4.5}$$

for all M. Each μ^{*m} is concentrated on a finite subset of $[ms_1, \infty)$, and, of course, $ms_1 \to \infty$ as $m \to \infty$. It follows that it makes sense to speak of the infinite sum $\sum_{m=1}^{\infty} \mu^{*m}$ and that there exist numbers $\lambda_1, \lambda_2, \ldots$ and $0 < t_1 < t_2 < \ldots$ with

$$\sum_{m=1}^{\infty} \mu^{*m}(S) = \sum_{i=1}^{\infty} \lambda_i \mathbf{1}_S(t_i), \quad S \subset \mathbb{R} \text{ bounded.} \tag{4.6}$$

As f has compact support, it also follows that $\mu^{*m} * f \to 0$ a.e., as $m \to \infty$, so that, by (4.5)

$$\begin{aligned} f &= g' + \sum_{m=1}^{\infty} \mu^{*m} * g' = \\ &= g' + \sum_{i=1}^{\infty} \lambda_i g'(\bullet - t_i) = \sum_{i=0}^{\infty} \lambda_i g'(\bullet - t_i), \end{aligned} \tag{4.7}$$

where we introduce $\lambda_0 := 1$, $t_0 := 0$.

Actually, f is supported by $[A, B]$, and g, which is $k * f$, by $[A, B{+}s_N]$. Therefore, if I is a nonnegative integer with $t_I \geq B - A + 1$, then $g'(x - t_i) = 0$ if $i > I$ and $x \in [A{-}1, B{+}1]$, so that $f = \sum_{i=0}^{I} \lambda_i g'(\bullet - t_i)$ on $[A{-}1, B{+}1]$. Choosing an infinitely differentiable $c : \mathbb{R} \to \mathbb{R}$ with $c(x) = 1$ for $x \in [A, B]$ and $c(x) = 0$ for $x \notin [A{-}1, B{+}1]$, we see that

$$f(x) = \sum_{i=0}^{I} \lambda_i c(x) g'(x - t_i), \quad x \in \mathbb{R}, \tag{4.8}$$

a formula of the type described in (2.7).

Example 4.2 It is not difficult to generalize the above. Clearly, the condition $c_1 = 1$ may be replaced by $c_1 \neq 0$ without noticeable harm. The condition $s_0 = 0$ is not serious, either. Indeed, let k be as in (4.1) (without having $s_0 = 0$). Let K_0 be convolution with $k_0 := k(\bullet + s_0)$. Then k_0 is of the type considered above, and there exist $I \in \mathbb{N}$, numbers $\lambda_0, \ldots, \lambda_I$ and $t_0 < \ldots < t_I$, and a compactly supported, infinitely differentiable function c for which

$$f(x) = \sum_{i=0}^{I} \lambda_i c(x) (K_0 f)'(x - t_i) = \sum_{i=0}^{I} \lambda_i c(x) g'(x - t_i + s_0), \quad x \in \mathbb{R}, \tag{4.9}$$

which brings us back to (2.7).

Example 4.3 The preceding generalizes easily to a suitable class of spline functions. Suppose $s_0 < s_1 < \ldots < s_N$, $J \in \mathbb{N}$, and let $k : \mathbb{R} \to \mathbb{R}$ be such that k vanishes identically outside the interval $[s_0, s_N]$ and has $J - 1$ continuous derivatives, whereas for each $n \in \{1, \ldots, N\}$ the restriction of k to $[s_{n-1}, s_n]$ is a polynomial of degree at most J. Assume $k \neq 0$. The function k has a J-th derivative at all points except possibly $s_0, s_1, \ldots, s_N$, and $k^{(J)}$ is constant on each (s_{n-1}, s_n). For each $j \in \{1, \ldots, J\}$, $k^{(j-1)}$ is an indefinite integral of $k^{(j)}$. Hence, $k^{(J)} \neq 0$. By the previous examples, there exist $I \in \mathbb{N}$, numbers $\lambda_0, \ldots, \lambda_I$ and $t_0 < \ldots < t_I$, and a compactly supported, infinitely differentiable function c such that

$$f(x) = \sum_{i=0}^{I} \lambda_i c(x) (k^{(J)} * f)'(x - t_i) = \sum_{i=0}^{I} \lambda_i c(x) g^{(J+1)}(x - t_i), \quad x \in \mathbb{R} \tag{4.10}$$

and again we have an instance of (2.7).

5 Estimation and MISE when the wavelet expansion is used

Let $\mathcal{W}_J$ denote the class of all compactly supported functions on $\mathbb{R}$ with J continuous derivatives. Given any $\chi \in \mathcal{W}_J$ we write

$$\chi_{m,k}(x) := 2^{m/2}\chi(2^m x - k), \quad x \in \mathbb{R},\ m \in \mathbb{Z}, k \in \mathbb{Z}. \tag{5.1}$$

Let $\varphi \in \mathcal{W}_J$ be a scaling function and ψ the corresponding wavelet. For our purposes the wavelet $\psi \in \mathcal{W}_J$ must satisfy the additional property $\int_{-\infty}^{\infty} x^j \psi(x)dx = 0,\ j = 0, ..., r-1$ for some $r \in \mathbb{N}$ (see below). The resulting orthonormal wavelet basis is $\{\psi_{m,k}, (m,k) \in \mathbb{Z} \times \mathbb{Z}\}$. The existence of a wavelet with all these properties is shown in Daubechies (1992). At a given resolution level $M \in \mathbb{Z}$ the low frequency elements can be combined in the usual way to yield the orthonormal system $\{\varphi_{M,k}, k \in \mathbb{Z}\}$, that can be complemented to an orthonormal basis of $L^2(\mathbb{R})$ by adding the system of high frequency wavelets $\{\psi_{m,k}, m > M, k \in \mathbb{Z}\}$. Restricting ourselves in this section to operators K with $K^{-1} = \mathcal{D}$ as in (2.7), we see from (1.2) that we will need the adjoint, $\mathcal{D}^*$. The domain of $\mathcal{D}^*$ contains $\mathcal{W}_J$, and for $\chi \in \mathcal{W}_J$ we have

$$\begin{aligned} \mathcal{D}^*\chi &= \sum_{j=0}^{J}\sum_{i=0}^{I}(-1)^j (c_{ji}\,\chi)^{(j)}(\bullet + a_{ji}) = \\ &= \sum_{j=0}^{J}\sum_{i=0}^{I} d_{ji}\cdot \chi^{(j)}(\bullet + a_{ji}), \end{aligned} \tag{5.2}$$

for certain, easily obtained continuous functions d_{ji} that have compact supports.

Let us write, for brevity, $f_{M,k} := \langle f, \varphi_{M,k}\rangle$ and $f_{m,k} := \langle f, \psi_{m,k}\rangle$, $m > M$, so that we have the expansion

$$f = \sum_{k=-\infty}^{\infty} f_{M,k}\varphi_{M,k} + \sum_{m=M+1}^{\infty}\sum_{k=-\infty}^{\infty} f_{m,k}\psi_{m,k}.$$

Since X has density $g = Kf$ we have $\mathbf{E}(\mathcal{D}^*\chi)(X) = \langle g, \mathcal{D}^*\chi\rangle = \langle \mathcal{D}g, \chi\rangle = \langle f, \chi\rangle$, $\chi \in \mathcal{W}_J$, so that

$$\hat{f}_{M,k} := \frac{1}{n}\sum_{i=1}^{n}(\mathcal{D}^*\varphi_{M,k})(X_i), \quad \hat{f}_{m,k} := \frac{1}{n}\sum_{i=1}^{n}(\mathcal{D}^*\psi_{m,k})(X_i), \tag{5.3}$$

are unbiased estimators of $f_{M,k}$ and $f_{m,k}$. We are now in a position to present the general form of the wavelet-type inverse estimator

$$\hat{f}_{M,\nu,\delta} := \sum_{k=-\infty}^{\infty} \hat{f}_{M,k}\varphi_{M,k} + \sum_{m=M+1}^{M+\nu}\sum_{k=-\infty}^{\infty} 1_{\{|\hat{f}_{m,k}|>\delta\}}\hat{f}_{m,k}\psi_{m,k}, \tag{5.4}$$

for $M \in \mathbb{Z}$, $\nu \in \mathbb{N}$, and threshold $\delta > 0$. This kind of estimator has been introduced by Donoho (1995) in an inverse model and by Donoho, Johnstone, Kerkyacharian & Picard (1996) and Hall & Patil (1995) in a direct model. The difference is that here the estimated Fourier coefficients contain the exact inverse of the operator. In asymptotics the parameters M, ν, and δ will depend on n. Let us write

$$f_{M,\nu,\delta} := \mathbf{E}\hat{f}_{M,\nu,\delta}. \tag{5.5}$$

We will consider the details for the asymptotics of the MISE under the assumption that f is in the class $\mathcal{F}_r'$ of all functions on $\mathbb{R}$ that have $r \in \mathbb{N}$ square integrable derivatives. Results for nonsmooth f could be likewise patterned on Hall & Patil (1995) but require more technicalities that cannot be presented here due to space limitations. For the present choice of wavelet we have

$$\sum_{m=M+1}^{\infty} \sum_{k=-\infty}^{\infty} f_{m,k}^2 = O(2^{-2rM}), \text{ as } M \to \infty, \quad f \in \mathcal{F}_r'. \tag{5.6}$$

For such smooth functions there is no need to include the high frequency terms in the estimator which then reduces to

$$\hat{f}_M := \sum_{k=-\infty}^{\infty} \hat{f}_{M,k}\varphi_{M,k}, \text{ with } f_M := \mathbf{E}\hat{f}_M. \tag{5.7}$$

By (5.6) and because $\hat{f}_{M,k}$ is an unbiased estimator of $f_{M,k}$ it follows that the MISE equals

$$\begin{aligned} \mathbf{E}||\hat{f}_M - f||^2 &= \mathbf{E}||\hat{f}_M - f_M||^2 + ||f_M - f||^2 \\ &= \sum_{k=-\infty}^{\infty} \mathbf{Var}\hat{f}_{M,k} + O(2^{-2rM}). \end{aligned} \tag{5.8}$$

It will be convenient to set $\Phi_M(x) := \sum_{j=0}^{J}\sum_{i=0}^{I} \delta_{ji}|\varphi^{(j)}(x+2^M a_{ji})|$, $x \in \mathbb{R}$, $M \in \mathbb{Z}$, where for each j and i, δ_{ji} is the maximal absolute value of the function d_{ji}, introduced in (5.2). As φ is continuous and compactly supported, there is a constant C with $\sum_{k=-\infty}^{\infty} \Phi_M^2(x-k) \le C$, $x \in \mathbb{R}$, $M \in \mathbb{Z}$. Using the generic sample element X, from (5.3) we obtain:

$$\begin{aligned} &\sum_{k=-\infty}^{\infty} \mathbf{Var}\hat{f}_{M,k} \le \tfrac{1}{n} \sum_{k=-\infty}^{\infty} \mathbf{E}(\mathcal{D}^*\varphi_{M,k})^2(X) = \\ &= \tfrac{1}{n} \sum_{k=-\infty}^{\infty} \mathbf{E}\left\{ \sum_{j=0}^{J}\sum_{i=0}^{I} d_{ji}(X)\varphi_{M,k}^{(j)}(X + a_{ji}) \right\}^2 = \\ &= \tfrac{1}{n} \sum_{k=-\infty}^{\infty} \mathbf{E}\left\{ \sum_{j=0}^{J}\sum_{i=0}^{I} d_{ji}(X)2^{M/2}2^{Mj}\varphi^{(j)}(2^M(X + a_{ji}) - k) \right\}^2 \le \\ &\le \tfrac{1}{n}2^{M(1+2J)} \sum_{k=-\infty}^{\infty} \mathbf{E}\Phi_M^2(2^M x - k) \le \tfrac{1}{n}2^{M(1+2J)}C. \end{aligned} \tag{5.9}$$

In order to balance the variance and the bias part we should choose $M \sim {}^2\log n^{1/(2J+2r+1)} =: M(n)$, which yields the following result.

Theorem 5.1 *For $f \in \mathcal{F}_r'$ we have*

$$\mathbf{E}||\hat{f}_M - f||^2 = \mathrm{O}(n^{-r/(J+r+\frac{1}{2})}), \quad as\ n \to \infty, \tag{5.10}$$

provided that we choose $M \sim M(n)$ as defined above.

Example 5.1 Let us return once more to the boxcar convolution. If we do not want to make the assumption that f (and hence g) have bounded support we obtain an infinite sum on the right in (2.2). Because φ has compact support, however, $\mathcal{D}^*\varphi_{m,k}$ will involve only a fixed finite number of terms for all $k, m \geq M$, and any given number M. Apparently $J = 1$ and for $f \in \mathcal{F}_r'$ the asymptotic order of the MISE equals $O(\mathrm{n}^{-2r/(2r+3)})$. In order to compare with (3.7) for $f \in \mathcal{F}_\nu$, we should take $\nu = r + \frac{1}{2}$ and it follows that the wavelet-type estimators obtain the optimal rate for any number of derivatives and hence are superior to the regularized-inverse type estimators in (3.6).

Remark 5.1 A more general situation arises when preconditioning is applied. In Example 2.2, for instance, it is only after preconditioning that we arrive at an operator R with inverse of type (2.7). Expansion (1.2) generalizes to $f = \sum_{\lambda\in\Lambda} \langle f, \psi_\lambda\rangle \psi_\lambda = \sum_{\lambda\in\Lambda} \langle R^{-1}p, \psi_\lambda\rangle \psi_\lambda = \sum_{\lambda\in\Lambda} \langle \mathcal{D}Tg, \psi_\lambda\rangle \psi_\lambda = \sum_{\lambda\in\Lambda} \langle g, T^*\mathcal{D}^*\psi_\lambda\rangle \psi_\lambda$. For $f \in \mathcal{F}_r$ the estimator will again be given by

$$\hat{f}_{M,k} := \frac{1}{n}\sum_{i=1}^{n}(T^*\mathcal{D}^*\varphi_{M,k})(X_i). \tag{5.11}$$

The extra ill-posedness contained in $\mathcal{D}^*$ due to the preconditioning would be compensated by T^* which is a smoothing operator. In the calculation of the MISE this should be reflected in the order of the variance of $\hat{f}_{M,k}$. We will not consider this point here.

Remark 5.2 The main difficulty with the boxcar convolution are the zeros of the characteristic function of its kernel that prevent us from conveniently dealing with the deconvolution in the frequency domain. Many kernels have characteristic functions that don't have any zeros and that decay monotonically in the tails. The Fourier coefficients in expansion (1.2) can be computed in the frequency domain. In fact we have

$$f = \sum_{\lambda\in\Lambda} \langle Fg, F(K^{-1})^*\psi_\lambda\rangle \psi_\lambda = \sum_{\lambda\in\Lambda} \left\langle Fg, (1/\,\bar{\tilde{k}})F\psi_\lambda \right\rangle \psi_\lambda. \tag{5.12}$$

For suitable wavelets the Fourier coefficients can be unbiasedly estimated by substituting the empirical characteristic function, multiplied by $\frac{1}{2\pi}$, for Fg. All that matters for calculation of the MISE along these lines is that the order of the variance of these Fourier coefficients can be obtained with sufficient accuracy.

Acknowledgements. The authors are indebted to Roger Barnard for some very helpful discussions and to the referees for useful comments. The second author gratefully acknowledges the hospitality of the Australian National University.

REFERENCES

Brocket, R., & Mesarovic, M. (1965). The reproducibility of multivariate systems. *J. Math. Anal. Appl.* **11**, 548-563.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* SIAM, Philadelphia.

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

Donoho. D.L., Johnstone, I.M., Kerkyacharian, G. & Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist.* **24**, 508-539.

Donoho, D.L. & Low, M.G. (1992). Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.* **20**, 944-970.

Groeneboom, P. & Jongbloed, G. (1995). Isotonic estimation and rates of convergence in Wicksell's problem. *Ann. Statist.* **23**, 1518-1542.

Hall, P. & Patil, P. (1995). Formulae for mean integrated squared error of nonlinear wavelet-based density estimators. *Ann. Statist.* **23**, 905-928.

Halmos, P.R. (1963). What does the spectral theorem say? *Amer. Math. Monthly* **70**, 241-247.

Nychka, D., Wahba, G., Goldfarb, S. & Pugh, T. (1984). Cross-validated spline methods for the estimation of three-dimensional tumor size distributions from observations on two-dimensional cross sections. *J. Amer. Statist. Assoc.* **79**, 832-846.

Riesz, F. & Sz.-Nagy, B. (1990). *Functional Analysis.* Dover, New York.

van Rooij, A.C.M. & Ruymgaart, F.H. (1996). Asymptotic minimax rates for abstract linear estimators. *J. Statist. Pl. Inf.* **53**, 389-402.

van Rooij, A.C.M. & Ruymgaart, F.H. (1998). On inverse estimation. To appear in *Asymptotics, Nonparametrics, and Time Series* (S. Ghosh, Ed.), Dekker, New York.

van Rooij, A.C.M., Ruymgaart, F.H. & van Zwet, W.R. (1999). Asymptotic efficiency of inverse estimators. *Th. Probability Appl.*, to appear.

Silverman, B.W., Jones, M.C., Wilson, J.D. & Nychka, D.W. (1990). A smoothed EM approach to indirect estimation problems, with particular reference to stereology and emission tomography (with discussion). *J.R. Statist. Soc. B.* **52**, 271-324.

Wahba, G. (1977). Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.* **14**, 651-667.

Wahba, G. & Wang, J. (1990). When is the optimal regularization parameter insensitive to the choice of the loss function? *Comm. Statist. - Th. Meth.* **19**, 1685-1700.

DEPARTMENT OF MATHEMATICS
TEXAS TECH UNIVERSITY
LUBBOCK TX 79409
USA
ruymg@math.ttu.edu