A New Method of Solving Noisy Abel-Type Equations

J. Garza

Department of Mathematics and Statistics, Texas Tech University,
Lubbock, Texas 79409

P. Hall

Centre for Mathematics and Its Applications, Australian National University,
Canberra, ACT 0200 Australia

and

F. H. Ruymgaart

Department of Mathematics and Statistics, Texas Tech University,
Lubbock, Texas 79409

Submitted by Ulrich Stadtmueller

Received January 4, 2000

A new approach to solving noisy integral equations of the first kind is applied to
the family of Abel equations. Such equations play a role in stereology (Wicksell’s
unfolding problem), medicine, engineering, and astronomy. The method is based
on an expansion in an arbitrary orthonormal basis, coupled with exact inversion of
the integral operator. The inverse appears in the Fourier coefficients of the
expansion, where it can be carried over to the usually well-behaved basis elements
in the form of the adjoint. This method is an alternative to Tikhonov regulariza-
tion, regularization of the inverse of the operator itself, or a wavelet-vaguelette/
singular-value decomposition. The method is particularly interesting in irregularity
of the kernel, the input, or both. Because knowledge of the spectral properties of
the operator is not required, the method is also of interest in regular cases where
these spectral properties are not sufficiently known or are hard to deal with. For
smooth input functions, the simple basis of trigonometric functions yields input
estimators whose mean integrated squared error converges at the optimal rate for
the entire family of Abel operators. This can be shown when smooth wavelets are
used for Abel operators with index smaller than 1/2, and when the Haar wavelet is
used for operators with index larger than 1/2.

1 The research was partially supported by National Security Agency grant MDA 904-99-1-0029.

0022-247X/01 $35.00
Copyright © 2001 by Academic Press
All rights of reproduction in any form reserved.
1. INTRODUCTION

Several methods have been proposed in the literature to solve noisy integral equations of the first kind. In such an integral equation, typically the inverse of the integral operator involved is unbounded, which is the source of ill-posedness. Because in practice information about the output function (i.e., the image of the unknown input function under the integral operator) is incomplete and corrupted by random noise, this ill-posedness is a serious problem. Therefore, in any procedure to recover the input from an imperfect output, some kind of regularization will be needed.

One such method is of the penalized least squares type, based on Tikhonov regularization and used by, for instance, Wahba [19] and Nychka and Cox [15]. Another recovery procedure exploits regularization of the inverse operator using Halmos’s [10] version of the spectral theorem. An overview of this approach was given by van Rooij and Ruymgaart [18]. Under standard regularity conditions these two methods will in general yield input estimators whose mean integrated square error (MISE) converges to 0 at the optimal rate.

Recently, Donoho [3] introduced a wavelet-vaguelette decomposition for optimal recovery of inhomogeneous input functions. This approach is reminiscent of the singular value decomposition for compact operators used by Johnstone and Silverman [11, 12], although the method applies to noncompact operators as well. Expansion in a suitable wavelet basis leads at once to “almost diagonalizing” the operator and to a convenient representation of prior knowledge regarding the input functions. However, this method is in essence adapted to scale-invariant operators and does not claim to discuss, for instance, convolution with kernels having a preferred spatial scale like the boxcar; that is the indicator of the interval $[-1, 1]$.

As an alternative, we may propose an expansion in an arbitrary orthonormal basis coupled with exact inversion of the integral operator. We show in Section 2 that the inverse appears in the Fourier coefficients of the expansion, where it can be carried over to the usually well-behaved basis elements in the form of the adjoint. This method was used by Hall et al. [9] for the aforementioned boxcar deconvolution problem, where it yielded optimal MISE rates that could not be obtained by spectral cutoff regularization of the inverse operator. Although only smooth input functions were considered in that paper, a wavelet basis was used. It is a fair conjecture that inputs with discontinuities of the first kind can be optimally recovered if high-resolution wavelets are included in the expansion with data-driven thresholding. Such thresholding was proposed by Donoho [3] in an inverse model and by Donoho et al. [3a] and Hall and Patil [8] in a direct model.
It seems, therefore, that this alternative might be quite successful if either the kernel or the input is irregular or if both are irregular. Because the method is entirely independent of the spectral properties of the operator, it will also be quite useful in regular cases, where these spectral properties are unknown or hard to deal with. In such regular cases, one might prefer an orthonormal system that suitably represents prior smoothness of the input. For dealing with irregular inputs, one may have to use a wavelet basis.

In this paper we want to illustrate the usefulness of this method by considering the class of generalized Abel equations \((5, 7)\) with index \(\alpha \), \(0 < \alpha < 1\) (see also \([13]\)). The noisy Abel equation with index \(\alpha = 1/2\) is related to Wicksell’s unfolding problem with applications in stereology, medicine, biology, and engineering, and has been extensively studied in the recent statistical literature \((6, 12, 14, 15, 16)\). An example from astronomy regarding binary orbits where the Abel \((1/2)\) equation occurs was reported by Feller \([4, p. 33]\).

The orthonormal basis chosen for the expansion appears to make a difference. As it turns out, the simple basis of trigonometric functions is not only convenient for specifying the smoothness of the input, but also yields estimators with MISE converging at the optimal rate (Sec. 3). Surprisingly, there are strong indications that for \(1/2 < \alpha < 1\), the optimal rate is not attained when the estimators are derived from a wavelet basis (Sec. 4). All this relates to the growth rate of the Fourier coefficients, which contain the adjoint inverse operator applied to the basis elements. Because of the unboundedness of the kernel, the localization property of the wavelets may have an adverse effect in the aforementioned range of the parameter. In Section 2 the model is introduced, and the inversion procedure with some of the ill-posedness issues is discussed.

2. PRELIMINARIES

In this paper we restrict ourselves to smooth input functions and deal mainly with trigonometric orthonormal systems. To better focus on the main, analytical aspects of the paper, we make certain unnecessarily restrictive assumptions regarding the statistical model. One of these is the assumption of a random design; another, that the input functions are symmetric about \(1/2\) with value 0 at 0 (and hence at 1). In Remarks 3.3 and 3.4 we briefly comment on how these restrictions might be alleviated at the cost of extra technicalities. As we explain in Remark 4.1, the foregoing restriction on the input function is not needed if smooth wavelets are used.
We introduce the set of functions
\[ \mathcal{L}_0 := \{ f \in C^1([0,1]) : f(0) = 0 \}. \] (2.1)

We are concerned with a noisy version of the integral equation
\[ g(x) = \frac{1}{\Gamma(\alpha)} \int_0^x f(y) \frac{dy}{(x-y)^\alpha}, 0 \leq x \leq 1, f \in \mathcal{L}_0, \] (2.2)
where \( 0 < \alpha < 1 \). Adopting a random design, let \( X \) denote a uniform \((0,1)\) design variable, and let \( \varepsilon \) denote a random error variable that has mean 0 and variance \( 0 < \sigma^2 < \infty \) and that is stochastically independent of \( X \). We observe a random sample \((X_1, Y_1), \ldots, (X_n, Y_n)\) consisting of independent copies of \((X, Y)\), where
\[ Y = (K_\alpha f)(X) + \varepsilon. \] (2.3)

The problem is to estimate \( f \) from the data.

Because we assess the quality of the estimator through the \( L^2([0,1]) \) norm, a Hilbert space perspective is pertinent. It is well known that for \( 0 < \alpha \leq 1/2 \), the linear transformation is bounded as an operator of \( L^2([0,1]) \) into itself. For each \( 0 < \alpha < 1 \), however, the effect of \( K_\alpha \) is some smoothing. Let us define, for \( \varphi \in C^1([0,1]) \),
\[ (K_\alpha^* \varphi)(x) := \frac{1}{\Gamma(\alpha)} \int_x^1 \frac{\varphi(y)}{(y-x)^\alpha} dy, 0 \leq x \leq 1, \] (2.4)
which will act as the adjoint of \( K_\alpha \). Let us also define
\[ \mathcal{L}_1 := \{ f \in C^1([0,1]) : f(1) = 0 \}. \] (2.5)

Combining Hackbusch (1989, Lemma 6.4.2, Theorem 6.2.1, and Theorem 6.4.4) the following result is immediate.

**Theorem 2.1.** For \( f \in \mathcal{L}_0 \) and any \( 0 < \alpha < 1 \), we have \( g := K_\alpha f \in \mathcal{L}_0 \) and Eq. (2.2) has solution
\[ f = DK_{1-\alpha} g = K_{1-\alpha} Dg. \] (2.6)

Similarly, if \( f \in \mathcal{L}_1 \), then we have \( K_\alpha^* f \in \mathcal{L}_1 \) and
\[ DK_{1-\alpha}^* f = K_{1-\alpha}^* Df. \] (2.7)

Because in practice \( g \) is only approximately known and differentiation is an unstable process, solving the noisy equation (2.3) requires regularization to cope with the ill-posedness. An estimated solution is obtained from
an exact orthonormal expansion for the input function, \( f \), by estimating the Fourier coefficients and suitable tapering. Let \( e_1, e_2, \ldots \) be an orthonormal system in \( L^2([0,1]) \), not necessarily a basis, such that [cf. (2.5)]

\[
e_k \in \mathcal{L}_1, \quad k \in \mathbb{N}. \tag{2.8}
\]

**Theorem 2.2.** Suppose that \( f \in \mathcal{L}_0 \) has the \( L^2 \) expansion \( f = \sum_{k=1}^{\infty} \langle f, e_k \rangle e_k \) for the \( e_k \) satisfying (2.8) and let \( g = K_1 f \). Then we have

\[
\langle f, e_k \rangle = -\langle g, K_1 e_k \rangle. \tag{2.9}
\]

**Proof.** We have the expansion \( f = \sum_{k=1}^{\infty} \langle f, e_k \rangle e_k = \sum_{k=1}^{\infty} \langle K_1^{-1} g, e_k \rangle e_k \) according to Theorem 2.1. We know that \( g \in \mathcal{L}_0 \) and hence \( K_1^{-1} g \in \mathcal{L}_0 \), so that in particular \( (K_1^{-1} g)(0) = 0 \). Because also \( e_k(1) = 0 \), it follows from integration by parts that \( \langle DK_1^{-1} g, e_k \rangle = -\langle K_1^{-1} g, De_k \rangle \). Application of Fubini’s theorem yields \(-\langle K_1^{-1} g, De_k \rangle = -\langle g, K_1^{-1} e_k \rangle \).

For the most part, we use the orthonormal basis \( 1_{[0,1]} e_1, \tilde{e}_1, e_2, \tilde{e}_2, \ldots \) of \( L^2([0,1]) \), where

\[
e_k(x) := \sqrt{2} \sin k(-\pi + 2\pi x), \quad 0 \leq x \leq 1 \tag{2.10}
\]

and

\[
\tilde{e}_k(x) := \sqrt{2} \cos k(-\pi + 2\pi x), \quad 0 \leq x \leq 1, \tag{2.11}
\]

\( k \in \mathbb{N} \). For convenience, we restrict \( f \) to a part of the linear span of \( e_1, e_2, \ldots \), thus restricting ourselves to functions that are symmetric about 1/2. Such \( f \) satisfy (2.1), and the \( e_k \) in (2.10) satisfy (2.8).

One way to understand the degree of ill-posedness of the recovery method based on the expansion in Theorem 2.2 is to consider the behavior of the Fourier coefficients \( k \mapsto \langle g, K_1^{\alpha} De_k \rangle \leq \|g\| \|K_1^{\alpha} De_k\| \). Let us first calculate the order of magnitude of \( \|K_\beta e_k\|, 0 < \beta < 1 \). The 0’s of the function \( e_k \) are at the points

\[
z_m := \frac{m}{2k}, \quad m = 0, 1, \ldots, 2k. \tag{2.12}
\]

As usual, we use the sign changes of \( e_k \) and the monotonicity of the kernel. First, we take \( k \) even, meaning that \( e_k \) will be positive between \( z_0 \)
and \( z_1 \), and assume that \( x \in [z_{2(m_x-1)}, z_{2m_x}] \) for some \( 1 \leq m_x \leq k \). It follows that

\[
\int_0^x \frac{e_k(y)}{(x-y)^\beta} \, dy = \sum_{l=1}^{m_x-1} \int_{z_{2l-1}}^{z_{2l}} \frac{e_k(y)}{(x-y)^\beta} \, dy + \int_{z_{2(m_x-1)}}^{x} \frac{e_k(y)}{(x-y)^\beta} \, dy. \tag{2.13}
\]

Because of the mean value theorem and the sign changes of \( e_k(y) \) we can observe that the first term on the right in (2.13) is equal to 0, whereas the second term is bounded by

\[
\frac{\sqrt{2}}{1 - \beta} \left\| (x-y)^{1-\beta} \right\|_{z_{2m_x-1}}^x \leq c_1 k^{\beta-1},
\]

where \( c_1 = \sqrt{2} / (1 - \beta) \). A similar result holds true for \( k \) odd, and a completely analogous argument shows that \( (K_\beta e_k)(x) \geq -c_2 k^{\beta-1} \) for some \( 0 < c_2 < \infty \).

Summarizing, we have shown that

\[
\left\| (K_\beta e_k)(x) \right\| \leq Ck^{\beta-1}, \quad 0 \leq x \leq 1, \tag{2.14}
\]

for some \( 0 < C < \infty \). The same orders would be obtained for the operator \( K_\beta^* \), and when the functions \( \hat{e}_k \) in (2.11) were considered. Because \( (De_k)(x) = 2\pi k\hat{e}_k(x), \) \( 0 \leq x \leq 1 \), the effect on the order of the norm by first differentiating the \( e_k \) is just an extra factor \( 2\pi k \). Combining all of this yields the following result.

**Theorem 2.3.** For the \( e_k \) given in (2.10) and \( 0 < \beta < 1 \), we have

\[
\| K_\beta e_k \|, \| K_\beta^* e_k \| = O(k^{\beta-1}) \quad \text{as} \quad k \to \infty \tag{2.15}
\]

and

\[
\| K_\beta De_k \|, \| K_\beta^* De_k \| = O(k^\beta) \quad \text{as} \quad k \to \infty. \tag{2.16}
\]

**Remark 2.1.** Although in particular (2.16) is very useful in Section 3 when we compute the rate of the MISE of the estimators and the rate of a lower bound to the risk, our original purpose was to gain insight into the ill-posedness from such an expression. Now the order obtained in (2.15) presents too optimistic a picture. It is well known that \( K_\beta \) is an unbounded operator in \( L^2([0,1]) \) for \( 1/2 < \beta < 1 \), but we still have \( \| K_\beta e_k \| \to 0 \), as \( k \to \infty \), for each \( 0 < \beta < 1 \). For the inversion of the noisy equation,
however, this conservative behavior will be favorable. In a sense, one might wish to find an orthonormal system where these rates or those in (2.16) are as small as possible. From the statistical results in Section 3, we may infer that the system (2.10) satisfies this property; see Remark 3.1. The situation will be different for $1/2 < \beta < 1$ if we use scaling functions as a suitable system for expanding smooth inputs.

3. OPTIMAL ERROR RATE FOR RECOVERY OF SMOOTH INPUT FUNCTIONS

Henceforth we restrict the input functions to a smoothness class of type

$$\mathcal{F}_\lambda := \left\{ f = \sum_{k=1}^{\infty} f_k e_k : |f_k| \leq \lambda_k \right\},$$

(3.1)

where the $e_k$ are given in (2.10) and

$$\lambda_k \geq 0, \sum_{k=1}^{\infty} \lambda_k < \infty.$$  (3.2)

Note that assumption (3.2) entails uniform convergence of the Fourier expansion for $f$, so that in particular,

$$\mathcal{F}_\lambda \subset \mathcal{L}_0,$$  (3.3)

because $e_k \in \mathcal{L}_0 \cap \mathcal{L}_1$. The summability condition in (3.2) is slightly more restrictive than is usually encountered in curve estimation, where summability of the squares suffices. But this restriction guarantees the pointwise convergence of the Fourier series in (3.1), which entails $f \in \mathcal{L}_0 \cap \mathcal{L}_1$. This restriction does not really play a role when we assume that $f$ has at least one square-integrable derivative, because this essentially implies (3.2).

As an estimator of $f$, we now propose

$$\hat{f}(x) := \sum_{k=1}^{N} \hat{f}_k e_k(x), \quad 0 \leq x \leq 1,$$  (3.4)

for suitable $N \in \mathbb{N}$, where the empirical Fourier coefficients are given by [cf. (2.9)]

$$\hat{f}_k := \frac{-1}{n} \sum_{i=1}^{n} Y_i \left( K_{1-a} D e_k \right)(X_i).$$  (3.5)

In several examples (e.g., Wicksell's problem), the parameter $\alpha$ is known. Whether knowledge of $\alpha$ is necessary and whether an estimator
adapted to unknown \( \alpha \) could be considered are interesting questions. But because apart from a constant, \( K_\alpha K_\beta \) equals \( K_{\alpha + \beta -} \), unknown \( \alpha \) leads to unidentifiability. We do not pursue here the question of how the input should be further restricted to restore identifiability.

The terms on the right side of (3.5) require (for large \( k \)) integration of highly oscillating functions. (For suitable numerical integration of such functions see, e.g., [1].) The estimators of the Fourier coefficients of \( f \) in (3.5) are unbiased because, according to (2.10),

\[
\mathbb{E} \hat{f}_k = - \mathbb{E} Y(K_{1-a}^- D e_k)(X) = - \mathbb{E}(K_\alpha f)(X) \cdot (K_{1-a}^- D e_k)(X) = - \langle g, K_{1-a}^- D e_k \rangle = \langle f, e_k \rangle =: f_k. \tag{3.6}
\]

Of course, this entails

\[
\mathbb{E} \hat{f}(x) = \sum_{k=1}^N f_k e_k(x). \tag{3.7}
\]

For the MISE, we now find

\[
\mathbb{E} \| \hat{f} - f \|^2 = \mathbb{E} \| \hat{f} - \mathbb{E} \hat{f} \|^2 + \| \mathbb{E} \hat{f} - f \|^2 \\
= \mathbb{E} \sum_{k=1}^N (\hat{f}_k - f_k)^2 + \sum_{k=N+1}^\infty f_k^2 \\
= \frac{1}{n} \sum_{k=1}^N \text{var} \{ Y(K_{1-a}^- D e_k)(X) \} + \sum_{k=N+1}^\infty f_k^2 \\
\leq \frac{1}{n} \sum_{k=1}^N \mathbb{E} \{ Y(K_{1-a}^- D e_k)(X) \}^2 + \sum_{k=N+1}^\infty \lambda_k^2. \tag{3.8}
\]

For positive numbers \( a_k > 0, b_k > 0 \), let us write

\[
a_k \asymp b_k \quad \text{as} \quad k \to \infty, \tag{3.9}
\]

if \( a_k = O(b_k) \) and \( b_k = O(a_k) \), as \( k \to \infty \). Let us now more specifically assume that the input function satisfies

\[
f \in \mathcal{F}_\lambda \text{ with } \lambda_k \asymp k^{-\nu} \text{ for some } \nu > 1. \tag{3.10}
\]
Theorem 3.1. For \( \lambda \) as in (3.10), the MISE satisfies
\[
\sup_{f \in \mathcal{F}} E\|\hat{f} - f\|^2 = O\left(n^{-2(1+\alpha)}\right) \quad \text{as } n \to \infty, \tag{3.11}
\]
provided that we choose \( N \asymp n^{1/2(1+\alpha)} \).

Proof. To exploit (3.8), we first observe that the functions in \( \mathcal{F} \) are uniformly bounded by \( \sqrt{2} \sum_{k=1}^{\infty} \lambda_k \), so that \( |g(x)| = |(K_0 f)(x)| \leq \sqrt{2} \sum_{k=1}^{\infty} \lambda_k / \int_0^1 (x-y)^{\alpha-1} dy \leq (\sqrt{2} \sum_{k=1}^{\infty} \lambda_k) / (1 - \alpha) =: C, \ 0 \leq x \leq 1 \). Hence the corresponding functions \( g \) are also uniformly bounded, and
\[
E[Y(K_{1-a}^* De_k)(X)]^2 \leq (C^2 + \sigma^2) E(K_{1-a}^* De_k)^2(X) \leq Ck^{2(1-a)} \quad \text{as } k \to \infty, \tag{3.12}
\]
using \( C \in (0, \infty) \) as a generic constant throughout the remainder of this proof. This follows from (2.16), and because \( \epsilon \) has mean 0, finite variance \( \sigma^2 \), and is stochastically independent of \( X \).

Because the upper bound in (3.12) and the \( \lambda_k \) are independent of \( f \in \mathcal{F} \), it follows that
\[
\sup_{f \in \mathcal{F}} E\|\hat{f} - f\|^2 \leq C \left\{ \frac{1}{n} \sum_{k=1}^{N} k^{2(1-a)} + \sum_{k=N+1}^{\infty} k^{-2\nu} \right\} \leq C \left\{ \frac{1}{n} N^{3-2a} + N^{1-2\nu} \right\}. \tag{3.13}
\]

The variance and bias contributions are balanced if we take \( N \asymp n^{1/2(1+\alpha)} \), which yields the overall rate as claimed in (3.11).

To obtain a lower bound to the risk, let \( \mathcal{F} \) denote the class of all estimators \( T \) of \( f \) with \( E\|T\|^2 < \infty \). Assume that the error variable \( \epsilon \) has a continuously differentiable density \( \psi \) with respect to Lebesgue measure with finite Fisher information
\[
I_\psi := \int_{-\infty}^{\infty} \frac{\psi'(x)^2}{\psi(x)} \, dx < \infty. \tag{3.14}
\]

Theorem 3.2. Under assumption (3.14), the MISE has a lower bound
\[
\inf_{T \in \mathcal{F}} \sup_{f \in \mathcal{F}} E\|T - f\|^2 \geq C \sum_{k=1}^{\infty} \frac{k^{-2\nu}}{nk^{-2(1+\alpha)} + 1}. \tag{3.15}
\]
The estimators \( \hat{f} \) in (3.4) with \( N \approx n^{1/2(1+\nu-\alpha)} \) are asymptotically optimal in the sense that their risk is of the same order as the lower bound in (3.15), as \( n \to \infty \).

**Proof.** To apply Theorem 3.1 of van Rooij and Ruymgaart (1996) along the lines of, for instance, their Example 4.2, we first note that under the present assumptions \( \sum_{k=1}^{\infty} f_k(K_a e_k)(x) \) is uniformly convergent for \( 0 \leq x \leq 1 \), as follows from (3.10) and calculations like (2.13). Because \( X \) is uniform on \([0,1]\), the joint density of \( X \) and \( Y \) equals

\[
p(x, y) := \psi \left( y - \sum_{k=1}^{\infty} f_k(K_a e_k)(x) \right), \quad (x, y) \in [0,1] \times \mathbb{R}, \tag{3.16}
\]

and similar to the example referred to in the beginning of this proof, we have

\[
\left\| \frac{\partial}{\partial f_k} \sqrt{P} \right\|^2 = \frac{1}{4} \|K_a e_k\|^2 \cdot I_0 \leq C k^{2(a-1)}, \tag{3.17}
\]

using (2.15). Application of the aforementioned theorem yields at once the lower bound in (3.15).

Asymptotically, for \( n \to \infty \), we have

\[
\sum_{k=1}^{\infty} \frac{k^{-2\nu}}{nk^{-2(1+\nu-\alpha)} + 1} \asymp \int_0^{\infty} \frac{x^{-2\nu}}{nx^{-2(1+\nu-\alpha)} + 1} \, dx \approx n^{-\frac{\nu+1}{2\nu+1}} \int_0^{\infty} \frac{y^{2(1-\alpha)}}{1 + y^{2(\nu+1-\alpha)}} \, dy, \tag{3.18}
\]

where the last integral is finite because \( \nu > 1 \).

**Remark 3.1.** We see from (3.12) that the rate of \( \|K_{1-\alpha}^* D e_k\| \), obtained in (2.16), is crucial for obtaining the optimal convergence rate of the MISE of the proposed estimators. For instance, if we would have found that \( \|K_{1-\alpha}^* D e_k\| \approx k^\gamma \), for some \( \gamma > 1 - \alpha \), then this rate would have been \( \asymp n^{-(3\nu-1)/2(\nu+\gamma)} \) and no longer optimal. In this sense the orthonormal system \( \{e_k\} \) is optimal.

**Remark 3.2.** In practice, it is of great importance to know how the regularization parameter \( N \) should be chosen for given, finite sample sizes. For some general results on data-driven selection of regularization parameters in statistical inverse problems see [2].

**Remark 3.3.** Instead of the random design, a deterministic design \( x_{a,i} \in [0,1], \ i = 1, \ldots, n \), might be used. We focus on a regular grid, as is
usually used in image analysis. In this case the estimators of the Fourier coefficients are as in (3.5), but with $X_i$ replaced by $x_{n,i} = i/n$. Now the estimators are no longer unbiased, because the integrals representing the inner products are replaced by Riemann sums. Because both the basis elements $e_i$ and the input function $f$ are smooth [it suffices to take $\nu > 2$ in (3.10)], these Riemann sum approximations are sufficiently accurate to ensure the same convergence rates as for the random design. The technical details are straightforward and are omitted.

REMARK 3.4. It is clear from the proof of Theorem 2.2 that the assumption $f \in \mathcal{L}_0, e_k \in \mathcal{L}_1$ simplifies the integration by parts. By adding the function $1_{[0,1]}$ to the orthonormal system $\{e_k\}$, we can deal with input functions $f$ with $f(0) \neq 0$. To recover such $f$, we have to estimate $\langle f, 1_{[0,1]} \rangle$ which boils down to estimating $(K_{1-a}g)(1)$. To remove symmetry involves further technicalities. We do not pursue this kind of generalization here.

4. SOME PROBLEMS WHEN WAVELETS ARE USED

Let $\varphi$ be a scaling function with compact support in $[0,1]$, wavelet $\psi$ and dilation-translation families $\varphi_{m,k}(x) := 2^m/\varphi(2^mx - k)$, $\psi_{m,k}(x) := 2^m/\psi(2^mx - k), 0 \leq x \leq 1$, where we tacitly restrict the integer indices $m$ and $k$ to yield different functions with support in $[0,1]$. At this moment, we do not specify any further properties of the wavelets, but will require that

$$f \in \mathcal{F}^r, \quad r \in \mathbb{N},$$

(4.1)

where $\mathcal{F}^r$ is the class of all functions in $L^2([0,1])$ that have a continuous $r$th derivative. A sufficiently smooth wavelet can and is chosen in such a way that

$$\sum_{m \geq M} \sum_k \langle f, \psi_{m,k} \rangle^2 = O(2^{-2rM}) \quad \text{as} \quad M \to \infty.$$

(4.2)

Throughout, $0 < C < \infty$ denotes a generic constant.

REMARK 4.1. Because a smooth scaling function $\varphi$ with support in the unit interval yields a dilation-translation family of functions that are in $\mathcal{L}_0 \cap \mathcal{L}_1$, relation (2.9) now holds true for $f \in \mathcal{F}^r$ without further restriction.
Because of the smoothness of \( f \), its estimator based on the wavelet expansion can be restricted to the low-frequency terms

\[
\hat{f}_M := \sum_k \hat{f}_{M,k} \varphi_{M,k}, \quad \hat{f}_{M,k} := -\frac{1}{n} \sum_{i=1}^n Y_i(K^*_1 a D \varphi_{M,k})(X_i) \tag{4.3}
\]

for suitable \( M \) [cf. (3.4) and (3.5)]. We write

\[
f_M := E \hat{f}_M = \sum_k (E \hat{f}_{M,k}) \varphi_{M,k} = \sum \langle f, \varphi_{M,k} \rangle \varphi_{M,k}. \tag{4.4}
\]

It follows easily as in the proof of Theorem 3.1 that

\[
E \|f_M - f\|^2 = E \|\hat{f}_M - f_M\|^2 + \|f_M - f\|^2
\]

\[
= \frac{1}{n} \sum_k \text{var}(Y(K^*_1 a D \varphi_{M,k})(X)) + \|f_M - f\|^2
\]

\[
\leq \frac{1}{n} \sum_k E(Y(K^*_1 a D \varphi_{M,k})(X))^2 + \|f_M - f\|^2
\]

\[
\leq \frac{C}{n} \sum_k \|K^*_1 a D \varphi_{M,k}\|^2 + O(2^{-2rM}). \tag{4.5}
\]

To further specify this upper bound, note that the function \( D \varphi_{M,k} \) has the same fixed number of sign changes for each \( M \) and \( k \). Unlike the situation for the sinus basis considered in Theorem 2.3, here we cannot expect cancellation to play a role in the determination of the order of the norms in the last line of (4.5). Because

\[
(D \varphi_{M,k})(x) = 2^{3M/2} \varphi'(2^M x - k), \quad (k - 1)2^{-M} \leq x \leq k2^{-M},
\]

and \( \varphi' \) is bounded, it seems not unreasonable to try the upper bound,

\[
\|K^*_1 a D \varphi_{M,k}\|^2 \leq C 2^{3M} \|K^*_1 a 1_{[(k-1)2^{-M}, k2^{-M})}\|^2. \tag{4.7}
\]

Throughout the remainder of this section, we use the following lemma to compute norms as in (4.7).

**Lemma 4.1.** Let \( 0 < \Delta < 1/2, a = k \Delta, b = (k + 1) \Delta, \) for \( k \in \mathbb{N} \) such that \( a, b \in [0, 1] \), \( \beta < 1 \) such that \( |\beta| \neq 1/2, \beta \neq 0 \), and

\[
h_{\Delta, \beta}(x) := \left( (b - x)^\beta - (a - x)^\beta \right) 1_{[0, a]}(x) + (b - a)^\beta 1_{[a, b]}(x),
\]

\[
x \in [0, 1]. \tag{4.8}
\]
Then we have (as $\Delta \downarrow 0$)

$$
\int_0^1 h_{\Delta, \beta}^2(x) \, dx = \begin{cases}
O(\Delta^{2\beta+1}), & \beta < \frac{1}{2} \\
O(k^{2-\beta} \Delta^{2\beta+1}), & \frac{1}{2} < \beta < 1.
\end{cases}
\tag{4.9}
$$

**Proof.** Let $0 < C < \infty$ denote a generic constant and set $p(x) := (b - x)^\beta$, so that $p(x + \Delta) = (a - x)^\beta$ and $p'(x) = -\beta(b - x)^{\beta-1}$. It follows from the mean value theorem that for $x < \xi(x) < x + \Delta$,

$$
\begin{align*}
|(b - x)^\beta - (a - x)^\beta| &= \Delta |\beta p'(\xi(x))| \\
&= C\Delta(b - \xi(x))^{\beta-1} \leq C\Delta(a - x)^{\beta-1},
\end{align*}
$$

where we restrict $x$ to the interval $[0, a - \Delta]$. Furthermore, we have

$$
(b - x)^\beta \begin{cases}
\leq (a - x)^\beta, & \beta < 0 \\
\geq (a - x)^\beta, & \beta > 0.
\end{cases}
\tag{4.11}
$$

It follows that for $\beta < 0$, $\beta \neq -1/2$,

$$
\begin{align*}
\int_0^1 h_{\Delta, \beta}^2(x) \, dx &\leq C \left( \Delta^2 \int_0^{a - \Delta} (a - x)^{2\beta - 2} \, dx \\
&\quad + \int_{a - \Delta}^a (a - x)^{2\beta} \, dx + \int_a^b (b - x)^{2\beta} \, dx \right) \\
&\leq C\Delta^2(a^{2\beta-1} + \Delta^{2\beta-1}) + C\Delta^{2\beta+1} \leq C\Delta^{2\beta+1}.
\end{align*}
$$

\tag{4.12}

For $0 < \beta < 1/2$, we need only replace $\int_{a - \Delta}^a (a - x)^{2\beta} \, dx$ with $\int_{a - \Delta}^a (b - x)^{2\beta} \, dx$, which yields the same bound, because still $2\beta - 1 < 0$. For $1/2 < \beta < 1$, just note that $a^{2\beta-1} = (k\Delta)^2\Delta^{2\beta-1} \geq \Delta^{2\beta-1}$, because $2\beta - 1 > 0$. Hence in this case, the overall order is given by $\Delta^2(k\Delta)^{2\beta-1} = k^{2\beta-1}\Delta^{2\beta+1}$.

**Remark 4.2.** It is tempting to use the elementary inequality [7, Lemma 6.2.2]

$$
|x^\beta - y^\beta| \leq |x - y|^\beta, \quad x \geq 0, \quad y \geq 0, \quad 0 \leq \beta \leq 1,
$$

\tag{4.13}

in (4.12). This would yield (for such $\beta$)

$$
\int_0^1 h_{\Delta, \beta}^2(x) \, dx \leq C(k\Delta^{2\beta+1} + \Delta^{2\beta+1}) = O(k\Delta^{2\beta+1}).
\tag{4.14}
$$
Because the upper bound is needed for large \( k \) (i.e., for \( k > \frac{1}{2} 2^M \)) this upper bound is essentially larger than the one in (4.9) for \( \beta < 1/2 \) or \( 1/2 < \beta < 1 \). Thus it will lead to a slower rate for the MISE.

Next, we observe that

\[
(K_{1-a}^r 1_{[a,b]})(x) = Ch_{\Delta,a}(x). \tag{4.15}
\]

Application of Lemma 4.1 with \( \Delta = 2^{-M} \) and \( 0 < \alpha = \beta < 1/2 \) yields

\[
\|K_{1-a}^r \varphi_{M,k}\|^2 = O(2^{-M(2\alpha+1)}),
\]

\[
\|K_{1-a}^r D\varphi_{M,k}\|^2 = O(2^{2M(1-a)}), 0 < \alpha < \frac{1}{2}, \tag{4.16}
\]

provided that \( k \geq 2 \). Combining this with (4.5), the following result is now immediate.

**THEOREM 4.1.** Choosing \( 2^M \approx n^{1/((3-2a+2r))} \), we find

\[
E\|\hat{f}_M - f\|^2 = O(n^{-2r/(3-2a+2r)}) \quad \text{as } n \to \infty, \tag{4.17}
\]

for \( f \in \mathcal{F}^r \), provided that \( 0 < \alpha < 1/2 \).

For \( 1/2 < \alpha = \beta < 1 \), the result in (4.9) leads to

\[
E\|\hat{f}_M - f\|^2 \leq \frac{C}{n} 2^M \sum_{k=1}^{2^M} k^{2a-1} 2^{2M(1-a)} + O(2^{-2rM})
\]

\[
= O\left( \frac{1}{n} 2^M \right) + O(2^{-2rM}). \tag{4.18}
\]

In this case, balancing the two terms yields

\[
E\|\hat{f}_M - f\|^2 = O(n^{-r/(1+r)}) \quad \text{as } n \to \infty. \tag{4.19}
\]

**REMARK 4.3.** Because for \( f \in \mathcal{F}^r \), the lower bound of the MISE reduces to the rate in (3.11) with \( \nu = r + 1/2 \), we see from (4.17) that for \( 0 < \alpha < 1/2 \), the wavelet estimators attain the optimal rate. For \( 1/2 < \alpha < 1 \), however, the present method does not yield the optimal rate. Therefore, we consider a specific simple wavelet that allows for a different approach.

The wavelet that we consider is the Haar wavelet with scaling function \( \varphi := 1_{[0,1]} \). The problem with this wavelet is that it is not smooth, although (4.2) remains true for \( r = 1 \). Indeed, the estimator \( \hat{f}_{M,k} \) of \( \langle f, \varphi_{M,k} \rangle \) as defined in (4.3) does not make sense, because \( \varphi_{M,k} \) is not differentiable.
For \( f \in T_1^a \cap \mathcal{L}_0 \), however, we see that
\[
\langle f, \varphi_{M,k} \rangle = \langle DK_{1-a}g, \varphi_{M,k} \rangle
= \langle K_{1-a}Dg, \varphi_{M,k} \rangle = \langle Dg, K_{1-a}^* \varphi_{M,k} \rangle,
\]
(4.20)
by applying Theorem 2.1. Note that for \( 1/2 < \alpha < 1 \), the operator \( K_{1-a}^* \) is defined on \( L^2([0,1]) \) with \( K_{1-a}^* \varphi_{M,k} = C2^{M/2}\tilde{h}_{\Delta,a} \in L^2([0,1]) \), where \( \tilde{h}_{\Delta,a} \) is defined in (4.8) (\( a = (k-1)2^{-M} \), \( b = k2^{-M}, \Delta = 2^{-M} \)). This function is sufficiently well behaved for the integration by parts formula to remain valid, so that we have
\[
\langle f, \varphi_{M,k} \rangle = -\langle g, DK_{1-a}^* \varphi_{M,k} \rangle
= -C\Delta^{-1/2}\langle g, h_{\Delta,a} \rangle,
\]
(4.21)
where \( h_{\Delta,a} \in L^2([0,1]) \). Hence in some “weak” sense \( D \) and \( K_{1-a}^* \) can be interchanged even when the Haar wavelet is involved.

It follows from these considerations that
\[
\tilde{f}_{M,k}^a = \frac{-1}{n} \sum_{i=1}^{n} Y_i(DK_{1-a}^* \varphi_{M,k})(X_i)
\]
(4.22)
is also an unbiased estimator of \( \langle f, \varphi_{M,k} \rangle \) so that
\[
\hat{f}_{M,k}^a = \sum_k \tilde{f}_{M,k}^a \varphi_{M,k}
\]
has expectation \( E\hat{f}_{M,k}^a = f_M \) as in (4.4).

Just as in (4.5), we need to compute \( \|DK_{1-a}^* \varphi_{M,k}\|^2 = C\Delta^{-1}\|h_{\Delta,a}^\prime\|^2 = C\Delta^{-1}\|h_{\Delta,a-1}\|^2 \). Applying (4.9) with \( \beta = \alpha - 1 < 0 \) yields
\[
\|DK_{1-a}^* \varphi_{M,k}\|^2 = O(2^{2M(1-a)}), \quad \frac{1}{2} < \alpha < 1.
\]
(4.24)
This yields at once the following.

**Theorem 4.2.** The estimators in (4.22), based on the Haar wavelet, satisfy
\[
E\|\hat{f}_{M,k}^a - f\|^2 = O(n^{-2r/(3-2\alpha+2r)}) \quad \text{as } n \to \infty
\]
(4.25)
for \( f \in T_1^a \cap \mathcal{L}_0 \) and \( 1/2 < \alpha < 1 \).

**Remark 4.2.** For both the trigonometric and smooth wavelet bases, it turned out to be convenient to apply \( (K_{a}^{-1})^* \) by first applying \( D^* \) and then \( K_{1-a}^* \); see (2.16) and (4.7). Because both bases have a frequency that increases with the index, it is easy to see that differentiation, when applied
first, produces this frequency as a multiplicative factor and hence determines the ill-posedness in an obvious quantitative manner. Because of the very simple explicit form of the Haar wavelet, it was possible to carry out the calculations with the operators in reverse order. One might consider interchanging these operators in the case of smooth wavelets in an attempt to arrive at a better upper bound for the MISE. But this approach does not seem to lead to the desired result either.

REMARK 4.5. When the input signal is known to be smooth, there is no reason to use a wavelet basis. We have seen that, for instance, a trigonometric system performs optimally over the entire range \(0 < \alpha < 1\). However, when the input signal may have irregularities like jump discontinuities, in direct curve estimation wavelet expansions, including high-resolution terms with thresholding, are known to be superior [8]. As the discussion in this section shows, however, one should be careful when using a wavelet expansion to recover irregular inputs in a noisy Abel equation with parameter \(1/2 < \alpha < 1\).

ACKNOWLEDGMENTS

The authors thank the referee for some useful remarks.

REFERENCES

A NEW METHOD OF SOLVING NOISY ABEL-TYPE EQUATIONS