Universität Stuttgart

Fachbereich Mathematik

Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions

Alexander Meister

Preprint 2007/001

Universität Stuttgart
Fachbereich Mathematik

Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions

Alexander Meister

Preprint 2007/001

Fachbereich Mathematik
Fakultät Mathematik und Physik
Universität Stuttgart
Pfaffenwaldring 57
D-70 569 Stuttgart

E-Mail: preprints@mathematik.uni-stuttgart.de
WWW: http://www.mathematik/uni-stuttgart.de/preprints

ISSN **xxxx-xxxx**

Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions

Alexander Meister

Institut für Stochastik und Anwendungen,
Universität Stuttgart
D-70569 Stuttgart, Germany

Abstract: This paper is concerned with deconvolution from error or blurring densities whose Fourier transforms have isolated zeros and show oscillatory behaviour; unlike conventional approaches where the Fourier transform decays about monotonously. We introduce specific estimation procedures based on local polynomial approximation in the Fourier domain. Under combined moment and smoothness conditions, we are able to improve the convergence rates compared to existing methods in density deconvolution. The corresponding minimax theory is derived. In compactly supported models as in signal deblurring and Berkson regression, nearly optimal rates are achieved under conditions which are significantly weaker than those assumed in earlier papers.

Keywords: Berkson regression; density estimation; image deblurring; nonparametric statistics; optimal convergence rates; statistical inverse problems.

MSC: 62G07; 62G08; 62H35.

1. Introduction

Many problems in nonparametric statistics require application of deconvolution procedures, e.g. density estimation based on contaminated data, errors-in-variables problems in nonparametric regression, image deblurring.

Fourier techniques are the dominating method in deconvolution problems as, in the Fourier domain, convolution with a known error density g changes into simple multiplication by the Fourier transform of g, which is denoted by g^{ft}. Therefore, most deconvolution procedures are based on dividing an empirical quantity by g^{ft}. More concretely, our goal is estimating a function $f \in L_2(\mathbb{R})$ where we have a data-based version $\hat{\psi}_h(t)$ of $h^{ft}(t) = f^{ft}(t)g^{ft}(t)$ with $h = f * g$ ($*$ denotes convolution).

The probably most obvious strategy is employing

$$\tilde{\psi}_f(t) \;=\; \hat{\psi}_h(t)/g^{ft}(t) \tag{1}$$

as the estimator of $f^{ft}(t)$. However, we are getting into trouble if g^{ft} has some zeros. Therefore, it has become common in deconvolution problems to assume that g^{ft} vanishes nowhere; although there are important densities which do not satisfy that condition, for example, uniform densities and convolution of uniform densities with any other density. So there is a comprehensive nonparametric class of error densities where classical deconvolution estimators are not applicable.

We focus on error densities g whose Fourier transforms have periodic zeros and show oscillatory bahaviour; unlike error densities considered in traditional approaches where g^{ft} decays about monotonously, i.e. the upper and the lower bound of g^{ft} coincide with each other up to different positive constants. We call an error density g Fourier-oscillating (FO) if it satisfies

$$C_2\big|\sin(\pi t/\lambda)\big|^{\mu}\,|t|^{-\nu} \;\geq\; |g^{ft}(t)| \;\geq\; C_1\big|\sin(\pi t/\lambda)\big|^{\mu}\,|t|^{-\nu}\,, \qquad \forall |t| \geq T\,, \tag{2}$$

with some constants $\lambda, C_2, C_1, T > 0$; while $|g^{ft}(t)| > 0$ is stipulated for $|t| \leq T$. Parameter $\mu \geq 1$ represents the order of the isolated zeros; while $\nu \geq \mu$ describes the tail behaviour of $|g^{ft}|$. Hence, all μ-fold self-convolved uniform densities are included into the framework of (2) for $\mu = \nu$; as well as convolutions of those densities with ordinary smooth densities, i.e. their Fourier transforms decay as a polynomial.

In terms of density deconvolution, Hall & Meister (2007) study the optimal convergence rates for FO error densities; they show that the rates can be kept from non-FO ordinary smooth error densities only in rare cases; e.g. parameter μ is sufficiently small with respect to ν and to the smoothness degree of f; also, the definition of smoothness classes for f is modified, compared to standard Sobolev conditions. In the related problem of image recovery, Johnstone et al. (2004) and Kerkyacharian et al. (2007) consider deconvolution from uniform blurring densities g when there is a certain relation between the variance of g and the support of f. In special deconvolution models, e.g. circular deconvolution (see Goldenshluger (2002)) or compactly supported f (see Meister (2005)), it is possible to recover f even if the set of all zeros of g^{ft} is open and non-void; however, the rates are slow then (logarithmic or at least sub-algebraic).

In the current paper, we consider deconvolution problems from FO error densities g under restrictions, which are satisfied by assumptions on the decay of f, in addition to usual smoothness assumptions on f. The methods for reconstructing $f^{ft}(t)$ when $g^{ft}(t)$ is equal or close to zero are based on a local polynomial approach in the Fourier domain; it is described in Section 2. Then, by using those specific estimation procedures, we are able to improve the speed of convergence compared to the rates derived in Hall & Meister (2007) in the field of density deconvolution (Section 3); data-driven bandwidth selection is discussed. In compactly supported models under discrete transform models, the rates are even very close to those derived for non-FO error densities (Section 4). We derive nearly optimal rates in signal deblurring under significantly weaker conditions on g, compared to Johnstone et al. (2004) and Kerkyacharian et al. (2007); more concretely, under uniform blurring densities g, we allow any positive scaling parameter of g; not only irrational parameters (Subsection 4.1). Also, our results have applications in Berkson regression problems (Subsection 4.2). The proofs are deferred to Section 5.

2. Methodology

In the sequel, the isolated zeros of g^{ft} are denoted by $t_j = j\lambda$ for $|j| \geq T/\lambda$, according to (2). We consider how to estimate $f^{ft}(t)$ when t is near or even equal t_j. Apparently, the function $\tilde{\psi}_f$ as in (1) has a pole at t_j. Therefore, reconstruction of f^{ft} in the neighbourhoods $A_{j,n} = [t_j - a_{j,n}, t_j + a_{j,n}]$, $a_{j,n} > 0$ becomes particularly challenging.

However, if f^{ft} is locally approximable by a polynomial around some neighbourhoods of t_j – say $T_j = (t_j - \tau, t_j + \tau)$ with some fixed $\tau \in (0, \lambda/2)$ – then $f^{ft}(t)$ shall be empirically accessible for $t \in A_{j,n}$ from $\tilde{\psi}_f$ on a domain outside $A_{j,n}$; hence we define the interval $B_{j,n} = [t_j - a_{j,n} - b_{j,n}, t_j - a_{j,n}]$, for $j > 0$, and $B_{j,n} = [t_j + a_{j,n}, t_j + a_{j,n} + b_{j,n}]$, otherwise; with $a_{j,n}, b_{j,n} > 0$ and $a_{j,n} + b_{j,n} \leq \tau$ so that $\tilde{\psi}_f$ does not have any poles on $B_{j,n}$.

We assume local smoothness conditions on f^{ft} by upper bounds on its derivatives,

$$\max_{k\in\{0,\ldots,m+1\}} \sup_{|j|\geq T/\lambda} \sup_{t\in T_j} \big|\big[f^{ft}\big]^{(k)}(t)\big| \leq C_3 , \tag{3}$$

where $C_3 > 0$, $m \geq 0$. Then we may apply the Taylor expansion around t_j, giving us

$$f^{ft}(t) = \sum_{k=0}^{m} \frac{1}{k!} \big[f^{ft}\big]^{(k)}(t_j) \cdot (t - t_j)^k + R_{t_j,m}(t) , \tag{4}$$

for all $t \in T_j$ where the residual term $R_{t_j,m}(t)$ satisfies $\big|R_{t_j,m}(t)\big| \leq C_3 \big[(m+1)!\big]^{-1} \cdot |t - t_j|^{m+1}$, due to Lagrange's representation. We notice that condition (3) holds if the first to the $(m+1)$th moment of f are bounded above,

$$\int |x|^k |f(x)| dx \leq C_4 , \qquad \forall k \in \{0, \ldots, m+1\} , \tag{5}$$

for some appropriate constant C_4. In problems where f is a density it suffices to assume the above inequality for $k = m + 1$ only. On the other hand, the Cauchy density $f_0(x) = 1/\big[\pi(1 + x^2)\big]$ is also included into the framework of (3) although it does not satisfy (5). Hence, condition (3) seems rather mild.

To construct an estimator of $f^{ft}(t), t \in A_{j,n}$, we apply a local polynomial approach in the Fourier domain. As a tool for projecting $\tilde{\psi}_f$ onto the space of all polynomials defined on $B_{j,n}$ with the degree $\leq m_n$, we introduce the polynomials $P_{k,j,n}$ by

$$P_{k,j,n}(t) = \begin{cases} \Big(\frac{4k+2}{b_{j,n}}\Big)^{1/2} L_k\big[2b_{j,n}^{-1}\big(t - t_j + a_{j,n} + b_{j,n}/2\big)\big] , & \text{for } j > 0 , \\ \Big(\frac{4k+2}{b_{j,n}}\Big)^{1/2} L_k\big[2b_{j,n}^{-1}\big(t - t_j - a_{j,n} - b_{j,n}/2\big)\big] , & \text{otherwise} , \end{cases} \tag{6}$$

on $t \in B_{j,n}$, writing χ_A for the indicator function of a set A and denoting the kth Legendre polynomial on $[-1, 1]$ by L_k. One can show that $\big\{P_{k,j,n}\big\}_{\text{integer } k\geq 0}$ are an orthonormal base of the Hilbert space of all squared-integrable functions defined on $B_{j,n}$, called $L_2(B_{j,n})$. We introduce

$$\tilde{\psi}^{\text{app.}}_{f;j,n}(t) = \sum_{k=0}^{m_n} P_{k,j,n}(t) \cdot \int_{B_{j,n}} P_{k,j,n}(s) \hat{\psi}_h(s)/g^{ft}(s)\, ds , \tag{7}$$

on $t \in B_{j,n}$. The sequence $(m_n)_n$ is still to be determined. Due to the constraint (3), we have to respect $m_n \leq m$. Note that $\tilde{\psi}^{\text{app.}}_{f;j,n}$ may uniquely be continued on $t \in \mathbb{R}\backslash B_{j,n}$ so that we have a polynomial on the whole real line. Hence, we employ that continuation, denoted by $\tilde{\psi}^{\text{con.}}_{f;j,n}(t)$, as the estimator of $f^{ft}(t)$ for $t \in A_{j,n}$.

The accuracy of estimator $\tilde{\psi}^{\text{con.}}_{f;j,n}(t)$ is studied in the following lemma. We write const. for a generic positive constant.

Lemma 1 *Assume that g is bounded and satisfies (2); choose $a_{j,n}, b_{j,n} > 0$ so that $a_{j,n}+b_{j,n} \leq \tau$, $\forall |j| > T/\lambda$, integer n. We assume $m_n^2 \leq const. \cdot b_{j,n}/a_{j,n}$ for all integers n and $|j| > T/\lambda$. If f satifies (3) and $m_n \leq m$, then we have, for any $|j| \geq T/\lambda$,*

$$\sup_{t\in A_{j,n}} E\big|\tilde{\psi}_{f;j,n}^{con.}(t) - f^{ft}(t)\big|^2 \leq O\big(m_n^2 j^{2\nu} a_{j,n}^{-2\mu}\big) \sup_{s\in B_{j,n}} E\big|\hat{\psi}_h(s) - h^{ft}(s)\big|^2 + O\big(C_3^2 a_{j,n}^{2m_n+2}/[(m_n+1)!]^2\big) + O\big(C_3^2 m_n^3 (b_{j,n}+a_{j,n})^{2m_n+2}/[(m_n+1)!]^2\big)\,,$$

where $O(\cdots)$ does not depend on C_3, m and f.

3. Density deconvolution

In this section, we apply the results derived in the previous section to the problem of density estimation under additive measurement error. In this model, we observe the data

$$Y_j \;=\; X_j + \varepsilon_j\,, \qquad\quad j \in \{1,\ldots,n\}$$

where all $X_1, \varepsilon_1, \ldots, X_n, \varepsilon_n$ are independent. The ε_j, which represent the contamination of the data, have the known density g; while our goal is estimating the density of the X_j, denoted by f.

That widely-studied problem – also known as density deconvolution – has received considerable attention during the last decades. Deconvolution kernel estimators were introduced by Stefanski & Carroll (1990) and Carroll & Hall (1988); the underlying minimax theory was developed by Fan (1991, 1993). For recent contributions to that topic see Delaigle & Gijbels (2002, 2004a,b), Carroll & Hall (2004), Butucea & Tsybakov (2007a,b). As further methods, we mention discrete transform approaches (Hall & Qiu (2005)) or wavelet-based methods (Pensky & Vidakovic (1999), Compte et al. (2006)).

Therefore, the problem as well as the notations f, g, $h = f * g$ are embedded into the general framework introduced in Section 1 and 2; where h is interpreted as the density of the observations Y_j. Hence, for $\hat{\psi}_h$ as used in Section 1 and 2, we choose the empirical characteristic function

$$\hat{\psi}_h(t) \;=\; \frac{1}{n}\sum_{j=1}^{n} \exp(itY_j)\,.$$

Inside the neighbourhoods $t \in A_{j,n}$ of the isolated zeros of g^{ft}, we employ the procedure derived in Section 2 to estimate $f^{ft}(t)$; while outside $A_{j,n}$, we apply the simple version according to (1). The specific parameters $a_{j,n}$, $b_{j,n}$ are still to be chosen suitably. More precisely, for integer $|j| \geq T/\lambda$, we define

$$\hat{\psi}_f(t) \;=\; \begin{cases} \tilde{\psi}_{f;j,n}^{\text{con.}}(t)\,, & \text{if } t \in A_{j,n} \text{ for some } |j| \geq T/\lambda\,, \\ \hat{\psi}_h(t)/g^{ft}(t)\,, & \text{otherwise,} \end{cases} \tag{8}$$

leading to the final density estimator

$$\hat{f}(x) \;=\; \frac{1}{2\pi}\int \exp(-itx) K^{ft}(th)\hat{\psi}_f(t)dt\,, \tag{9}$$

with a kernel function K and the bandwidth $h = h_n$. We stipulate that K^{ft} is compactly supported so that estimator (5) is well-defined.

With respect to the target density f, we assume (3) with fixed constants C_3, m, τ and usual Sobolev conditions, given by

$$\int |f^{ft}(t)|^2\big(1+|t|^{2\beta}\big)dt \;\leq\; C_5\,. \tag{10}$$

All those densities are collected into the class $\mathcal{F}_{\beta,C_5;C_3,m,\tau}$. Hence, our framework for f is a combination of smoothness (with degree β) and decay restrictions (with degree m). We feel that these assumptions are realistic for most densities of practical interest.

We will show that our estimator (9) achieves optimal rates of convergence in the considered problem. We define the mean integrated squared error (MISE) of an estimator $\hat{f}$ for f by

$$\mathrm{MISE}_n(\hat{f}, f) \;=\; E\|\hat{f} - f\|^2\,,$$

where $\|\cdot\|$ denotes the $L_2(\mathbb{R})$-norm. We give the following theorem.

Theorem 1 *Take estimator $\hat{f}$ as in (9). Assume that the error density g is bounded and satisfies (2). Select K so that K^{ft} is supported on $[-1,1]$; $|K^{ft}(t)| \leq 1$; and $|K^{ft}(t) - 1| = o(|t|^\beta)$. We choose $m_n = m$; $a_{j,n} = const.\cdot j^\kappa n^{-\eta}$ with $\kappa = \nu/(\mu+m+1)$; $b_{j,n} = 3a_{j,n}$; and $\eta = 1/(2\mu+2m+2)$; and $h = const.\cdot n^{-(2m+3)/\left[2m(2\beta+2\nu+1)+4\beta\mu+2\mu+6\nu+4\beta+2\right]}$. Then, we have*

$$\sup_{f\in\mathcal{F}_{\beta,C_5;C_3,m,\tau}} MISE_n(\hat{f}, f) \;=\; O\big(n^{-(2m+3)\beta/\left[m(2\beta+2\nu+1)+2\beta\mu+\mu+3\nu+2\beta+1\right]}\big)\,.$$

In our setting, the ridge-parameter estimator of Hall & Meister (2007) is unable to take advantage of (3) and loses its optimality. For instance, taking the Sobolev(β) class of densities with bounded first moment ($m = 0$) with respect to f and the uniform density on $[-1,1]$ as g, the supremum of the MISE of estimator (9) converges at the rate $n^{-3\beta/(4\beta+5)}$ according to Theorem 1; while the ridge-parameter method does not achieve rates faster than $n^{-1/2}$ (see Theorem 4.2 in Hall & Meister (2007)), which is significantly slower if $\beta > 5/2$. In the countercase $\beta \leq 5/2$, the more restrictive smoothness conditions of Hall & Meister (2007) become efficient.

On the other hand, we realize that the rates in Theorem 1 are slower than those derived in Fan (1991, 1993) for non-FO error densities whose Fourier transforms decay as $|t|^{-\nu}$; that rate is

$$n^{-2\beta/(2\beta+2\nu+1)}\,. \tag{11}$$

In the sequel, we will refer to those rates as classical deconvolution rates as they are also applicable in a couple of different deconvolution problems. Of course μ does not occur in (11) because there are no isolated zeros; but the tail behaviour of $g^{ft}(t)$ described by ν influences the speed of convergence. Obviously, we have to pay for the periodic zeros of g^{ft}. Hence, there is a necessity to prove optimality for the rate in Theorem 1. The following theorem shows that our estimator cannot be improved with respect to the convergence rates under mild additional assumptions on the derivative of g, which are indeed satisfied for μ-fold self-convolved uniform densities as well as their convolutions with appropriate ordinary smooth densities, for example.

Theorem 2 *Assume that g is bounded and satisfies (2) and, in addition,*

$$|g^{ft\prime}(t)| \;\leq\; D_2'|\sin(\pi t/\lambda)|^{\mu-1}|t|^{-\nu}$$

for some constant $D_2' > 0$; let $\beta > 1/(m+1)$; consider C_3, C_5, τ as sufficiently large. Then, for any estimator $\hat{f}$ based on the data $Y_1, \ldots, Y_n$, we have

$$\sup_{f\in\mathcal{F}_{\beta,C_5;C_3,m,\tau}} MISE_n(\hat{f}, f) \;\geq\; const.\cdot n^{-(2m+3)\beta/\left[m(2\beta+2\nu+1)+2\beta\mu+\mu+3\nu+2\beta+1\right]}\,.$$

As an unattractive feature, estimator (9) uses parameter m in its construction; and the optimal selection of h requires knowledge of the smoothness degree β, in addition. The previous aspect is justified if some a-priori information on the decay of f as in (5) is given. Further, under certain conditions, one may empirically test the assumption $\int |x|^{m+1} f(x) \leq C_k$ for any fixed m and f; for estimators of the $(m+1)$th moment of X_1 based on the contaminated data $Y_1, \ldots, Y_n$, see Meister (2006). With respect to the bandwidth choice, we propose a cross-validation (CV) procedure;

those methods are common in deconvolution problems (e.g. Hesse (1999), Hall & Meister (2007)). We define the quantity $\hat{\Xi}(h;t)$ by

$$\hat{\Xi}(h;t) \;=\; \frac{1}{\pi n(n-1)} \mathrm{Re} \sum_{j\neq k} K^{ft}(th)\,|g^{ft}(t)|^{-2} \exp\big(it(Y_j - Y_k)\big)\,, \qquad \forall |t| \notin \bigcup_p A_{p,n}$$

and

$$\hat{\Xi}(h;t) \;=\; \frac{1}{\pi n(n-1)} \mathrm{Re} \sum_{j\neq k} \sum_{l,l'}^{m_n} K^{ft}(th) P_{l,p,n}(t) P_{l',p,n}(t) \cdot \int_{B_{p,n}} P_{l,p,n}(s) \exp(-isY_j)/g^{ft}(-s)\, ds \cdot \int_{B_{p,n}} P_{l',p,n}(s) \exp(isY_k)/g^{ft}(s)\, ds\,, \tag{12}$$

for $t \in A_{p,n}$ and some integer p; while K^{ft} is supported on $[-1,1]$. Then we put our CV-function equal to

$$\mathrm{CV}(h) \;=\; \int \big|\hat{f}(h;t)\big|^2 dt \;-\; \int \hat{\Xi}(h;t) dt\,,$$

where $\hat{f}(h;\cdot)$ denotes estimator (9) with focus on its bandwidth h. A completely data-driven choice of the bandwidth $\hat{h}$ can be derived by taking the minimum of $\mathrm{CV}(\cdot)$ over an appropriately chosen subset of $[0,\infty)$.

4. Compactly supported functions f

Heuristically, we obtain the classical convergence rates for nonparametric deconvolution (11) when putting $m = \infty$ in Theorem 1. According to (5), that situation is identifiable with problems where all moments of f exist; that is satisfied by compactly supported f. Considering those f is realistic in other deconvolution problems than density estimation from corrupted data, e.g. signal recovery or Berkson regression.

In models with bounded support, discrete transform approaches are favorable, e.g. Hall & Qiu (2005), as square-integrable f can be represented by its Fourier series so that estimating f reduces to making the Fourier coefficients of f empirically accessible. We assume that f is supported on $[-\pi,\pi]$; simple rescaling techniques make our method applicable for more general support intervals. Given $f \in L_2([-\pi,\pi])$, we have

$$f(x) \;=\; \frac{1}{2\pi} \sum_{k\in\mathbb{Z}} f_k \exp(-ixk) \cdot \chi_{[-\pi,\pi]}(x)\,,$$

with the coefficients $f_k = \int \exp(itk) f(t) dt \;=\; f^{ft}(k)$. We notice that deconvolution is extremely difficult if e.g. $\lambda = 1$, hence $t_k = k$, in (2); then only a finite number of Fourier coefficients $g_k = g^{ft}(k)$ is not equal to zero. Therefore, standard methods as utilizing (1) for $t = k$ are not applicable. At that point of view, no consistent estimator of f seems to exist; however, we will show that this impression is wrong.

One could try to reelevate the troubles with $\lambda = 1$ by changing the support boundaries of f. Nevertheless, we should consider that the support domain cannot be chosen arbitrarily large as the endpoints of f are usually exempted from smoothness constraints on f, i.e. jump discontinuities at $\{\pi,\pi\}$ are allowed while f shall be smooth on (π,π). Therefore, the support is usually determined by the specific experiment. Those assumptions seem reasonable, for example considering a picture taken on a bounded domain. That reflects in the following definition of the smoothness assumptions, given by classifying the decay of the Fourier coefficients,

$$\sum_k \big(1 + |k|^{2\beta}\big)|f_k|^2 \;\le\; C_6\,, \tag{13}$$

for $\beta \geq 1/2$; the sum is to be taken over all integers k. In discrete transform models, (13) replaces the Sobolev condition (10). Similar conditions are also used in Korostelev & Tsybakov (1993), Hall & Qiu (2005), Delaigle et al. (2006), for instance.

Now, we are able to specify the function class $\mathcal{F}'_{C_6,\beta}$ by the totality of all functions f supported on $[-\pi,\pi]$; satisfying (13) and $\|f\|_{L_1(\mathbb{R})} \leq 1$. The latter condition can be relaxed by assuming an arbitrary but fixed upper bound instead of 1. In problems where f is a density, it need not be assumed explicitly, of course.

Let us study how condition (3) changes for $f \in \mathcal{F}'_{C_6,\beta}$. We have

$$\big|[f^{ft}]^{(m)}(t)\big| \;\leq\; \int_{[-\pi,\pi]} |x|^m |f(x)| dx \;\leq\; \pi^m\,, \qquad \forall\, t \in \mathbb{R}, \text{ integer } m > 0 \tag{14}$$

Therefore, C_3 in (3) may no longer be seen as a constant but as an exponentially increasing sequence; that becomes significant when applying Lemma 1. Now we are allowed to select $(m_n)_n$ as a sequence tending to infinity in the construction of estimator $\tilde{\psi}^{\text{con.}}_{f;j,n}$. Lemma 1, when setting $m = m_n$, provides, for $t \in A_{j,n}$,

$$\begin{aligned} E\big|\tilde{\psi}^{\text{con.}}_{f;j,n}(t) - f^{ft}(t)\big|^2 \leq O\big(m_n^2 j^{2\nu} a_{j,n}^{-2\mu}\big) \sup_{s\in B_{j,n}} E\big|\hat{\psi}_h(s) - h^{ft}(s)\big|^2 + O\big((\pi a_{j,n})^{2m_n+2}/[(m_n+1)!]^2\big) \\ + \; O\big(m_n^3[\pi(b_{j,n}+a_{j,n})]^{2m_n+2}/[(m_n+1)!]^2\big)\,, \end{aligned} \tag{15}$$

where constants do not depend on f; and $m_n^2 \leq \text{const.} \cdot b_{j,n}/a_{j,n}$ still has to be respected. Then we derive the estimator $\hat{f}_k$ of f_k by

$$\hat{f}_k \;=\; \begin{cases} \tilde{\psi}^{\text{con.}}_{f;j,n}(k)\,, & \text{if } k \in A_{j,n} \text{ for some } |j| > T/\lambda\,, \\ \hat{\psi}_h(k)/g^{ft}(k)\,, & \text{otherwise}\,, \end{cases}$$

and the function estimator

$$\hat{f}(x) \;=\; \frac{1}{2\pi} \sum_{|k|\leq J_n} \hat{f}_k \exp(-ixk) \cdot \chi_{[-\pi,\pi]}(x)\,, \tag{16}$$

where the integer sequence $(j_n)_n$ is the analogue of the reciprocal bandwidth in kernel estimation.

Yet, we leave $\hat{\psi}_h(t)$ undefined. Its specific selection depends on the statistical experiment. At this point, we only assume that

$$\sup_{t\in\mathbb{R}} E\big|\hat{\psi}_h(t) - h^{ft}(t)\big|^2 \;\leq\; \text{const.}\, n^{-1}\,, \tag{17}$$

where const. does not depend on f. Then, we give the following theorem.

Theorem 3 *Take estimator $\hat{f}$ as in (16). Assume that g is bounded and satisfies (2); and that (17) holds. Select $b_{j,n} = b \in (0,\lambda/4)$, $a_{j,n} = a_n = \text{const.}\,(\ln n\,/\ln\ln n)^{-2}$, $m_n = C_m \ln n\,/\ln\ln n$ with a constant $C_m > 1/2$, $J_n = \text{const.} \cdot n^{1/(2\beta+2\nu+1)} (\ln n/\ln\ln n)^{(-2-4\mu)/(2\nu+2\beta+1)}$. Then, we have*

$$\sup_{f\in\mathcal{F}'_{C_6,\beta}} MISE_n(\hat{f},f) \;=\; O\Big(n^{-2\beta/(2\beta+2\nu+1)}(\ln n/\ln\ln n)^{2\beta(2+4\mu)/(2\beta+2\nu+1)}\Big)\,.$$

Hence, estimator (16) achieves the classical rates (11) up to a logarithmic loss. We mention that in the papers of Johnstone et al. (2004) and Kerkyacharian et al. (2007), the consideration is also restricted to nearly optimal rates. Nevertheless, the question whether the logarithmic factor can be removed in the underlying model remains open.

As a data-driven procedure for selecting parameter J_n, we define the cross-validation function

$$\text{CV}(J) \;=\; \sum_{|k|\leq J} |\hat{f}_k|^2 \;-\; \sum_{|k|\leq J} \hat{\Xi}(1/J;k)\,,$$

where we take $\Xi(h;\cdot)$ from (12) with the sinc kernel as K. Minimizing $\mathrm{CV}(J)$ over an appropriate set of integers gives us an empirically accessible choice of J in the discrete-transform setting.

Further, in density deconvolution with compactly supported and continuous f ($\beta > 1/2$) where g is a uniform density $g_a = (2a)^{-1}\chi_{[-a,a]}$, we mention that, for $f(\pi) \geq$ const., the scaling parameter a is consistently estimable by $\hat{a} = \max\{Y_1, \ldots, Y_n\} - \pi$. Therefore, under those conditions, a need not be known. Nevertheless, our focus is directed to other problems in the compact support setting.

In the following subsections, we will apply the general result of Theorem 3 to two important statistical problems. All that remains to be verified is the existence of an estimator $\hat{\psi}_h$ satisfying (17).

4.1. Signal deblurring

We consider the famous model of image or signal reconstruction where the function $Y(x), x \in I$, driven by the stochastic differential equation,

$$dY(x) \,=\, (f * g)dx \,+\, n^{-1/2}dW(x)\,, \qquad \forall x \in I\,,$$

is observed on some compact interval I; while $W(x)$ denotes the standard Wiener process. For the minimax theory for non-FO blurring densities g (also known as pointspread function), see e.g. Korostelev & Tsybakov (1993). From there, we learn that the rates (11) are also obtainable and optimal in this setting.

Specific problems with uniform blurring densities g have become known as boxcar deconvolution, see Johnstone et al. (2004) and Kerkyacharian (2007) et al.. In those papers, deconvolution from uniform densities g is considered if – transferred to our notation – λ as in (2) is irrational. That means that none of the Fourier coefficients $g^{ft}(k)$ is equal to zero but some subsequences may accumulate to the zeros of g^{ft}. Another approach which might address the FO-case in signal deblurring is the ForWaRD algorithm introduced in Neelamani et al. (2004); however, the investigation of convergence rates is restricted to standard cases in that paper.

Theorem 3 provides a more general treatment of this problem for any $\lambda > 0$. Consider the situation where g is the μ-fold self-convolved uniform density supported on $[-\mu\pi/\lambda, \mu\pi/\lambda]$. We recall that $h = f * g$, by definition. As the estimator $\hat{\psi}_h(t)$, we propose

$$\hat{\psi}_h(t) \,=\, \int_{[-\pi-\mu\pi/\lambda,\pi+\mu\pi/\lambda]} \exp(itx)\, dY(x)\,, \tag{18}$$

assuming $[-\pi - \mu\pi/\lambda, \pi + \mu\pi/\lambda] \subseteq I$. Then, by the following lemma

Lemma 2

$$E\big|\hat{\psi}_h(t) - h^{ft}(t)\big|^2 \,\leq\, const. \cdot 1/n\,,$$

with constants independent of t and f ,

condition (17) is satisfied so that for $f \in \mathcal{F}'_{C_6,\beta}$ – a realistic condition in image deblurring – Theorem 3 is applicable. Therefore, estimator (16) establishes nearly optimal rates

4.2. Berkson regression

Since Berkson (1950), errors-in-variables regression problems have been studied where the design variables are contaminated after the corresponding data have been measured. In the mathematical model, one observes the i.i.d. data $(X_1, Y_1), \ldots, (X_n, Y_n)$ with

$$Y_j \,=\, f(X_j + \delta_j) + \eta_j\,,$$

where the X_j, δ_j, $j = 1, \ldots, n$ are independent; $E(\eta_j \mid X_j) = 0$ and $E(\eta_j^2 \mid X_j) = \sigma^2 \in (0, \infty)$, uniformly in f and n. Here, g must be interpreted as the density of the $-\delta_j$. The density of the X_j is called the design density f_X. Our goal is estimating the regression function f.

Note that this problem is different from the deconvolution regression model considered in Fan & Truong (1993) where the design points are observed with noise δ_j while $Y_j = f(X_j) + \eta_j$. A nonparametric treatment of the Berkson regression problem is given in Delaigle et al. (2006). In that paper, the authors mention that, in practice, both f and g are likely to be compactly supported. They propose an approach based on orthogonal series and derive optimal rates of convergence, which correspond to (11), indeed. Although the authors consider FO error densities their setting must be viewed at as a classical deconvolution problem as the Fourier coefficients $g^{ft}(j)$ are chosen so that $|g^{ft}(j)| \geq \text{const.}|j|^{-2\nu}$. That seems rather restrictive with respect to parameter λ. Cases where some of the $g^{ft}(k)$ are zero are not included; even accumulation of the $g^{ft}(k)$ at some zeros of g^{ft} is forbidden.

Again, we are able to consider the problem for arbitrary $\lambda > 0$, so we include even the extreme case where almost all $g^{ft}(k)$ are equal to zero. The conditions from Delaigle et al. (2006) may be adopted except the positive lower bound on the $g^{ft}(k)$. In detail, we assume that f is supported on $[-\pi, \pi]$; that f_X is bounded away from zero on $[-\pi - c, \pi + c]$ for some $c > 0$; that the η_j have mean zero and uniformly bounded variance; and that g is bounded, supported on $[-c, c]$ and satisfies (2). Hence, the contaminated regression function $h(x) := E(Y_1 \mid X_1 = x) = \big(f * g\big)(x)$ has its support on $[-\pi - c, \pi + c]$. Therefore, the uniform densities on $[-\pi/\lambda, \pi/\lambda]$ are included into our framework for any $\lambda \geq \pi/c$. Under the appropriate smoothness assumptions on f, we may verify $f \in \mathcal{F}'_{C_6,\beta}$. Estimating function $h(x)$ is a direct nonparametric regression problem, i.e. no deconvolution step is required. In rare cases, where the design density f_X is known and supported on $[-\pi - c, \pi + c]$, we may employ

$$\hat{\psi}_h(t) \; = \; \frac{1}{n} \sum_{j=1}^{n} Y_j \exp(itX_j) \, / \, f_X(X_j) \, ,$$

as an unbiased estimator for $h^{ft}(t)$, which satisfies (17). In a more realistic setting, we may define $\hat{\psi}_h$ as the Fourier transform of a truncated local linear smoother so that validity of (17) can also be ensured, see Delaigle et al. (2006) for the underlying theory.

Then, for fixed f_X, we are able to employ Theorem 3. Our procedure achieves nearly optimal convergence rates in the Berkson regression problem, too.

5. Proofs

Proof of Lemma 1: First we consider

$$\begin{aligned}
& E\Big|\tilde{\psi}^{\text{con.}}_{f;j,n}(t) \; - \; \sum_{k=0}^{m_n} P_{k,j,n}(t) \cdot \int_{B_{j,n}} P_{k,j,n}(s) f^{ft}(s) ds\Big|^2 \\
& \quad = \; E\Big| \int_{B_{j,n}} \big[\hat{\psi}_h(s) - h^{ft}(s)\big] \big[g^{ft}(s)\big]^{-1} \sum_{k=0}^{m_n} P_{k,j,n}(t) P_{k,j,n}(s) ds\Big|^2 \\
& \quad \leq \; b_{j,n} \cdot \int_{B_{j,n}} \big|g^{ft}(s)\big|^{-2} E\big|\hat{\psi}_h(s) - h^{ft}(s)\big|^2 \, \Big|\sum_{k=0}^{m_n} P_{k,j,n}(t) P_{k,j,n}(s)\Big|^2 \, ds \\
& \quad \leq \; O\big(b_{j,n} j^{2\nu} a_{j,n}^{-2\mu}\big) \sup_{s \in B_{j,n}} E\big|\hat{\psi}_h(s) - h^{ft}(s)\big|^2 \cdot \sum_{k=0}^{m_n} \big|P_{k,j,n}(t)\big|^2 \, , \qquad (19)
\end{aligned}$$

where we have used $f^{ft}(s) = h^{ft}(s)/g^{ft}(s)$ for $s \in B_{j,n}$, the Cauchy-Schwarz inequality, Fubini's theorem and the orthonormality of the $P_{k,j,n}$. To continue that chain of inequalities, we derive the following inequality for the Legendre polynomials L_k.

$$|L_k(t)| \;=\; \Big|\sum_{l=0}^{k} 2^{-k}\binom{k}{l}^2 (t-1)^l(t+1)^{k-l}\Big| \;\le\; \big|1+(|t|-1)/2\big|^k \cdot \Big|\sum_{l=0}^{k}\frac{k^{2l}}{(l!)^2}\Big(\frac{|t|-1}{|t|+1}\Big)^l\Big|$$
$$\le\; \big|1+(|t|-1)/2\big|^k \cdot \exp\big[k^2(|t|-1)\big]\,,$$

for all $|t| \ge 1$ (see e.g. Koepf (1998)). Hence, as $t \in A_{j,n}$,

$$\big|P_{k,j,n}(t)\big| \;\le\; \big[(4k+2)/b_{j,n}\big]^{1/2}\cdot(1+2a_{j,n}/b_{j,n})^k \cdot \exp(16k^2 a_{j,n}/b_{j,n})\,.$$

From there, we derive that (19) is bounded above by

$$O\big(m_n^2 j^{2\nu} a_{j,n}^{-2\mu}\big) \sup_{s\in B_{j,n}} E\big|\hat\psi_h(s)-h^{ft}(s)\big|^2\,, \tag{20}$$

as $m_n^2 \;\le\;$ const. $\cdot\, b_{j,n}/a_{j,n}$.

Further, we study under the condition $m_n \le m$,

$$\Big|f^{ft}(t) - \sum_{k=0}^{m_n} P_{k,j,n}(t)\cdot\int_{B_{j,n}} P_{k,j,n}(s) f^{ft}(s)ds\Big|^2$$
$$= \Big|R_{t_j,m_n}(t) \;-\; \sum_{k=0}^{m_n} P_{k,j,n}(t)\cdot\int_{B_{j,n}} P_{k,j,n}(s) R_{t_j,m_n}(s)ds\Big|^2$$
$$\le\; O\big(C_3^2|t-t_j|^{2m_n+2}/[(m_n+1)!]^2\big) \;+\; 2\sum_{k,k'=0}^{m_n}\big|P_{k,j,n}(t)P_{k',j,n}(t)\big|\int_{B_{j,n}} R^2_{t_j,m_n}(s)\,ds$$
$$\le\; O\big(C_3^2 a_{j,n}^{2m_n+2}/[(m_n+1)!]^2\big) \;+\; O\big(C_3^2 b_{j,n}[b_{j,n}+a_{j,n}]^{2m_n+2}/[(m_n+1)!]^2\big)\cdot\Big(\sum_{k=0}^{m_n}|P_{k,j,n}(t)|\Big)^2$$
$$\le\; O\big(C_3^2 a_{j,n}^{2m_n+2}/[(m_n+1)!]^2\big) \;+\; O\big(C_3^2 m_n^3[b_{j,n}+a_{j,n}]^{2m_n+2}/[(m_n+1)!]^2\big)\,, \tag{21}$$

where $f^{ft}(t)$ as well as $f^{ft}(s)$ inside the integral have been approximated by their m_nth Taylor polynomial (also see (4)) around t_j so that the polynomial part disappears; also, the bound for the Legendre polynomials has been used as above. Combining (20) and (21) gives us the result stated in the lemma. ∎

Proof of Theorem 1: By Parseval's identity and Fubini's theorem, we obtain

$$\mathrm{MISE}_n(\hat f, f) \;=\; \frac{1}{2\pi}\int E\big|\hat\psi_f(t)-f^{ft}(t)\big|^2\,dt \;\le\; \frac{1}{2\pi}\sum_{1/(\lambda h)\ge|j|\ge T/\lambda}\int_{A_{j,n}} E\big|\tilde\psi^{\text{con.}}_{f;j,n}(t)-f^{ft}(t)\big|^2 dt$$
$$+\;\frac{1}{2\pi}\sum_{1/(\lambda h)\ge|j|\ge T/\lambda}\int_{C_{j,n}} \mathrm{var}\,\big\{\hat\psi_h(t)/g^{ft}(t)\big\}\,dt \;+\; O(h^{2\beta}) \;+\; O(n^{-1})\,, \tag{22}$$

where $C_{j,n} = [t_j-\lambda/2, t_j+\lambda/2]\backslash A_{j,n}$. Further, we have used that $\hat\psi_f(t)$ is an unbiased estimator of $f^{ft}(t)$ for $t \in [-1/h, 1/h]\backslash\bigcup_j A_{j,n}$; K^{ft} is supported on $[-1,1]$; the integral restricted to $[t_{-1}+\lambda/2, t_1-\lambda/2]$ decays at the rate $O(n^{-1})$ and, hence, may be neglected. Condition $|K^{ft}(t)-1| = o(|t|^\beta)$ gives us the bound $O(h^{2\beta})$ for the bias term.

With respect to the second term in (22), we derive

$$\sum_{1/(\lambda h)\ge|j|\ge T/\lambda}\int_{C_{j,n}} \mathrm{var}\,\big\{\hat\psi_h(t)/g^{ft}(t)\big\}\,dt \;\le\; 4n^{-1}\sum_{1/(\lambda h)\ge j\ge T/\lambda}\int_{[t_j+a_{j,n},t_j+\lambda/2]}\big|g^{ft}(t)\big|^{-2}\,dt$$
$$\le\; O(n^{-1})\sum_{1/(\lambda h)\ge j\ge T/\lambda} j^{2\nu}\int_{[\pi a_{j,n}/\lambda,\pi/2]}\big|\sin t\big|^{-2\mu}dt \;\le\; O(n^{-1})\sum_{1/(\lambda h)\ge j\ge T/\lambda} j^{2\nu} a_{j,n}^{1-2\mu}\,. \tag{23}$$

Lemma 1 may be applied to the first term in (22), leading to

$$\sum_{1/(\lambda h)\ge|j|\ge T/\lambda} \int_{A_{j,n}} E\big|\tilde{\psi}^{\text{con.}}_{f;j,n}(t) - f^{ft}(t)\big|^2 dt$$
$$\le O(n^{-1}) \sum_{1/(\lambda h)\ge|j|\ge T/\lambda} |j|^{2\nu} a_{j,n}^{1-2\mu} + \sum_{1/(\lambda h)\ge|j|\ge T/\lambda} a_{j,n}^{2m+3}, \quad (24)$$

where $\sup_s E|\hat{\psi}_h(s) - h^{ft}(s)|^2 \le 1/n$ and $b_{j,n} = 3a_{j,n}$ have been considered. Inserting the specific choice of $a_{j,n}$, we obtain as an upper bound for (22),

$$O\Big(n^{-1+(2\mu-1)\eta} \sum_{1/(\lambda h)\ge j\ge T/\lambda} j^{2\nu+\kappa(1-2\mu)}, \, n^{-\eta(2m+3)} \sum_{1/(\lambda h)\ge j\ge T/\lambda} j^{(2m+3)\kappa}, \, h^{2\beta}\Big)$$
$$= O\Big(n^{-1+(2\mu-1)\eta} \max\{1, h^{-2\nu+(2\mu-1)\kappa-1}\}, \, n^{-\eta(2m+3)} \max\{1, h^{-\kappa(2m+3)-1}\}, \, h^{2\beta}\Big)$$

Selecting κ, η, h as stated in the theorem gives us the desired rate. ■

Proof of Theorem 2: As important tools, we introduce the supersmooth Cauchy density $f_0(x) = 1/\big[\pi(1+x^2)\big]$ and the density $f_1(x) = C_m\big[\big(1-\cos x\big)/(\pi x^2)\big]^{m+2}$ with the appropriate constant $C_m > 0$. We see that f_1 satisfies (5); hence, the Fourier transform f_1^{ft} is $(m+1)$-fold continuously differentiable and supported on $[-m-2, m+2]$. We consider the densities

$$f_\theta(x) = \frac{1}{2}\delta_n f_0(\delta_n x) + \frac{1}{2}\delta_n f_1(\delta_n x)\cdot\Big(1 + \text{const.}\cdot\delta_n^{m+1}\sum_{j=K_n}^{2K_n}\theta_j\cos(j\lambda x)\Big),$$

where $\delta_n \downarrow 0$, $\theta = (\theta_{K_n}, \ldots, \theta_{2K_n})$, $\theta_j \in \{0,1\}$ and K_n denotes an integer tending to infinity. As the corresponding Fourier transforms, we have

$$f_\theta^{ft}(t) = \frac{1}{2}f_0^{ft}(t/\delta_n) + \frac{1}{2}f_1^{ft}(t/\delta_n) + \text{const.}\,\delta_n^{m+1}\sum_{K_n\le|j|\le 2K_n}\theta_{|j|}f_1^{ft}\big((t-t_j)/\delta_n\big).$$

Under the selection

$$\delta_n = \text{const.}\cdot K_n^{(-2\beta-1)/(2m+3)} \quad (25)$$

with a suitable constant. we can verify the membership of f_θ in $\mathcal{F}_{\beta,C_5;C_3,m,\tau}$.

Considering the θ_j as independent random variables with $P(\theta_j = 0) = 1/2$, we obtain for an arbitrary estimator $\hat{f}$, which is based on $Y_1, \ldots, Y_n$, by Parseval's identity,

$$\sup_{f\in\mathcal{F}_{\beta,C_5;C_3,m,\tau}} E\|\hat{f} - f\|^2 \ge E_\theta E_{f_\theta}\|\hat{f} - f_\theta\|^2$$
$$= \frac{1}{2\pi}\sum_{K_n\le|j|\le 2K_n} E_\theta E_{f_\theta}\int_{(j-1/2)/\lambda}^{(j+1/2)/\lambda}\big|\hat{f}^{ft} - f_\theta^{ft}(t)\big|^2 dt$$
$$\ge \text{const.}\cdot\delta_n^{2m+2}\sum_{K_n\le|j|\le 2K_n}\int_{(j-1/2)/\lambda}^{(j+1/2)/\lambda}\big|f_1^{ft}\big((t-t_j)/\delta_n\big)\big|^2 dt$$
$$\ge \text{const.}\cdot K_n\delta_n^{2m+3}, \quad (26)$$

if the χ^2-distance between the observed densities $f_{\theta_{j,0}} * g$ and $f_{\theta_{j,1}} * g$ satisfies

$$\chi^2(f_{\theta_{j,0}} * g, f_{\theta_{j,1}} * g) = O(1/n), \quad (27)$$

for any $j \in \{K_n, \ldots, 2K_n\}$, where we write $\chi^2(f,g) = \int (f-g)^2/f\,dx$ and $\theta_{j,b} = (\theta_{K_n}, \ldots, \theta_{j-1}, b, \theta_{j+1}, \ldots, \theta_{2K_n})$ for $b \in \{0,1\}$. That result follows as in Fan (1991, 1993). So we have

$$\chi^2(f_{\theta_{j,0}} * g, f_{\theta_{j,1}} * g) \;\le\; 2\delta_n^{-1} \int \big|\big[\big(f_{\theta_{j,0}} - f_{\theta_{j,1}}\big) * g\big](x)\big|^2 \big[\big(f_0(\delta_n\cdot) * g\big)(x)\big]^{-1} dx\,. \tag{28}$$

With respect to the denominator in (28), we obtain

$$\big[f_0(\delta_n\cdot) * g\big](x) \;\ge\; \frac{1}{\pi}\int_{|y|\le q} g(y)\big[1 + 2\delta_n^2(x^2+y^2)\big]^{-1} dy \;\ge\; \text{const.}\cdot\big(1+\delta_n^2 x^2\big)^{-1}\,,$$

when choosing q sufficiently large so that $\int_{|y|\le q} g(y)dy > 0$. Therefore,

$$\chi^2(f_{\theta_{j,0}} * g, f_{\theta_{j,1}} * g) \;\le\; \text{const.}\cdot \int \big|\big[\big(f_{\theta_{j,0}} - f_{\theta_{j,1}}\big) * g\big](x)\big|^2 \big(\delta_n^{-1} + \delta_n x^2\big) dx\,.$$

In the following we employ the equality $\big(y^{ft}\big)' = i\big\{\cdot\, y(\cdot)\big\}^{ft}$ and Parseval's identity, leading to

$$\begin{aligned}\chi^2(f_{\theta_{j,0}} * g, f_{\theta_{j,1}} * g) \;\le\; \text{const.}\cdot \Big[&\delta_n^{2m+1}\int \big|f_1^{ft}\big((t\pm t_j)/\delta_n\big)\big|^2 \big|g^{ft}(t)\big|^2 dt \\ &+\; \delta_n^{2m+1}\int \big|[f_1^{ft}]'\big((t\pm t_j)/\delta_n\big)\big|^2 \big|g^{ft}(t)\big|^2 dt \\ &+\; \delta_n^{2m+3}\int \big|f_1^{ft}\big((t\pm t_j)/\delta_n\big)\big|^2 \big|[g^{ft}]'(t)\big|^2 dt\Big]\,.\end{aligned}$$

As both $(f_{\theta_{j,0}} - f_{\theta_{j,1}})^{ft}$ and its derivative are supported on $[\pm t_j - \delta_n, \pm t_j + \delta_n]$ we derive that $\chi^2(f_{\theta_{j,0}} * g, f_{\theta_{j,1}} * g)$ is bounded above by

$$O\bigg\{\delta_n^{2m+1}\int_{t_j-\delta_n}^{t_j+\delta_n} \big|g^{ft}(t)\big|^2 dt \;+\; \delta_n^{2m+3}\int_{t_j-\delta_n}^{t_j+\delta_n} \big|[g^{ft}]'(t)\big|^2 dt\bigg\}\,.$$

Using (10) and the additional conditions in Theorem 2, we obtain $O\big(\delta_n^{2+2m+2\mu} K_n^{-2\nu}\big)$ as an upper bound for $\chi^2(f_{\theta_{j,0}} * g, f_{\theta_{j,1}} * g)$. Combining that result with (25) we see that (27) is satisfied by the selection

$$K_n \;=\; n^{(2m+3)/[m(4\beta+2+4\nu)+2+4\beta\mu+2\mu+6\nu+4\beta]}\,,$$

and, by (26), we receive the lower bound on the MISE as stated. ■

Proof of Theorem 3: First note that the condition $m_n^2 \le \text{const.}\cdot b_{j,n}/a_{j,n}$ is satisfied by the specific parameter selection in the theorem. By Parseval's identity for Fourier series and (15), we derive

$$\begin{aligned} E\|\hat f - f\|_{L_2(\mathbb{R})}^2 \;\le\; &\frac{1}{2\pi}\sum_{k\in[-J_n,J_n]\cap\bigcup_j A_{j,n}} E\big|\tilde\psi^{\text{con.}}_{f;j,n}(k) - f_k\big|^2 \\ &+ \sum_{k\in[-J_n,J_n]\setminus\bigcup_j A_{j,n}} E\big|\hat\psi_h(k)/g^{ft}(k) - f_j\big|^2 \;+\; O(J_n^{-2\beta}) \\ \le\; &O\big(n^{-1}m_n^2 a_n^{-2\mu}\big)\sum_{|k|\le J_n}|k|^{2\nu} \;+\; O\big(J_n[\pi a_n]^{2m_n+2}/[(m_n+1)!]^2\big) \\ &+\; O\big(J_n m_n^3[\pi(b+a_n)]^{2m_n+2}/[(m_n+1)!]^2\big) \;+\; O(J_n^{-2\beta}) \\ \le\; &O\big(n^{-1}m_n^2 a_n^{-2\mu} J_n^{2\nu+1}\big) \;+\; O\big(J_n[\pi a_n]^{2m_n+2}/[(m_n+1)!]^2\big) \\ &+\; O\big(J_n m_n^3[\pi(b+a_n)]^{2m_n+2}/[(m_n+1)!]^2\big) \;+\; O(J_n^{-2\beta})\,. \end{aligned}$$

We have used that $k \in A_{j,n}$ implies that the ratio j/k is bounded above and below by some positive constants. Due to $C_m > 1/2$ and $b < \lambda/4$, the second and the third term are negligible in the equation above. Inserting the selection rules for b, a_n, m_n, J_n gives the convergence rates as stated. ■

Proof of Lemma 2: We utilize Ito's formula. Concerning the expectation, we obtain

$$E\hat{\psi}_h(t) = \int_{[-\pi-\mu\pi/\lambda,\pi+\mu\pi/\lambda]} \exp(itx)h(x)dx + n^{-1/2} E \int_{[-\pi-\mu\pi/\lambda,\pi+\mu\pi/\lambda]} \exp(itx)\, dW(x)$$
$$= h^{ft}(t)\,,$$

while considering that h is supported on $[-\pi - \mu\pi/\lambda, \pi + \mu\pi/\lambda]$. Therefore, we have unbiasedness of our estimator. Now we study the variance

$$\operatorname{var}\hat{\psi}_h(t) \le n^{-1} E\Big| \int_{[-\pi-\mu\pi/\lambda,\pi+\mu\pi/\lambda]} \exp(itx)\, dW(x)\Big|^2 = n^{-1} \int_{[-\pi-\mu\pi/\lambda,\pi+\mu\pi/\lambda]} |\exp(itx)|^2 dx$$
$$\le O(1/n)\,.$$

■

References

1. Berkson, J. (1950), Are there two regression problems?, *J. Amer. Statist. Assoc.* **45**, 164–180.
2. Butucea, C. & Tsybakov, A.B. (2007a), Sharp optimality for density deconvolution with dominating bias. I, *Theo. Probab. Appl.* **52**, to appear.
3. Butucea, C. & Tsybakov, A.B. (2007b), Sharp optimality for density deconvolution with dominating bias. II, *Theo. Probab. Appl.* **52**, to appear.
4. Carroll, R.J. & Hall, P. (1988), Optimal rates of convergence for deconvolving a density, *J. Amer. Statist. Assoc.* **83**, 1184–1186.
5. Carroll, R.J. & Hall, P. (2004), Low-order approximations in deconvolution and regression with errors in variables, *J. Roy. Statist. Soc., Ser. B* **66**, 31–46.
6. Comte, F., Rozenholc, Y. & Taupin, M.-L. (2006), Penalized contrast estimator for adaptive density deconvolution, *Canad. J. Statist.* **34**, 431–452.
7. Delaigle, A. & Gijbels, I. (2002), Estimation of integrated squared density derivatives from a contaminated sample, *J. Roy. Statist. Soc.* Ser. B **64**, 869–86.
8. Delaigle, A. & Gijbels, I. (2004a), Practical bandwidth selection in deconvolution kernel density estimation, *Comput. Statist. Data Anal.* **45**, 249–267.
9. Delaigle, A. & Gijbels, I. (2004b), Bootstrap bandwidth selection in kernel density estimation from a contaminated sample, *Ann. Inst. Statist. Math.* **56**, 19–47.
10. Delaigle, A., Hall, P. & Qiu, P. (2006), Nonparametric methods for solving the Berkson errors-in-variables problem, *J. Roy. Statist. Soc., Ser. B* **68**, 201–220.
11. Fan, J. (1991), On the optimal rates of convergence for non-parametric deconvolution problems, *Ann. Statist.* **19**, 1257–1272.
12. Fan, J. (1993), Adaptively local one-dimensional subproblems with application to a deconvolution problem, *Ann. Statist.* **21**, 600–610.

13. Fan, J. & Truong, Y. (1993), Nonparametric regression with errors in variables, *Ann. Statist.* **21**, 1900–1925.

14. Goldenshluger, A. (2002), Density deconvolution in the circular structural model, *J. Multivar. Anal.* **84**, 350–375.

15. Hall, P. & Meister, A. (2007), A ridge-parameter approach to deconvolution, *Ann. Statist.*, to appear.

16. Hesse, C.H. (1999), Data-driven deconvolution, *J. Nonparam. Stat.* **10**, 343–373.

17. Johnstone, I., Kerkyacharian, G., Picard, D. & Raimondo, M. (2004), Wavelet deconvolution in a periodic setting, *J. Roy. Statist. Soc., Ser. B* **66**, 547–573. with discussion pp. 627–652.

18. Hall, P. & Qiu, P. (2005), Discrete-transform approach to deconvolution problems, *Biometrika*, **92**, 135–148.

19. Kerkyacharian, G., Picard, D. & Raimondo, M. (2007), Adaptive boxcar deconvolution on full Lebesgue measure sets, *Statist. Sinica*, to appear.

20. Koepf, W. (1998), *Hypergeometric Summation: An Algorithmic Approach to Summation and Special Function Identities*, Vieweg, Braunschweig, Germany.

21. Korostelev, A.P. & Tsybakov, A.B. (1993), *Minimax Theory of Image Reconstruction*, Lecture Notes in Statistics **82**, Springer, New York.

22. Meister, A. (2005), Deconvolving compactly supported densities, preprint, available online: *http://www.mathematik.uni-stuttgart.de/preprints/downloads/2005/2005-007.pdf.*

23. Meister, A. (2006), Support estimation via moment estimation in presence of noise, *Statistics* **40**, 259–275.

24. Neelamani, R., Choi, H., Baraniuk, R. (2004), ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems, *IEEE Trans. Signal Process.* **52**, 418–433.

25. Pensky, M. & Vidakovic, B. (1999), Adaptive wavelet estimator for nonparametric density deconvolution, *Ann. Statist.* **27**, 2033–2053.

26. Stefanski, L.A. & Carroll, R.J. (1990), Deconvoluting kernel density estimators, *Statistics* **21**, 169–184.

Alexander Meister
Pfaffenwaldring 57
70569 Stuttgart
Germany
E-Mail: `meistear@mathematik.uni-stuttgart.de`

Erschienene Preprints ab Nummer 2005/001

Komplette Liste: `http://www.mathematik.uni-stuttgart.de/preprints`

2005/001 *Mielke, A.; Schmid, F.:* Vortex pinning in super-conductivity as a rate-independent model

2005/002 *Kimmerle, W.; Luca, F., Raggi-Cárdenas, A.G.:* Irreducible Components of the Burnside Ring

2005/003 *Höfert, C.; Kimmerle, W.:* On Torsion Units of Integral Group Rings of Groups of Small Order

2005/004 *Röhrl, N.:* A Least Squares Functional for Solving Inverse Sturm-Liouville Problems

2005/005 *Borisov, D.; Ekholm, T; Kovarik, H.:* Spectrum of the Magnetic Schrödinger Operator in a Waveguide with Combined Boundary Conditions

2005/006 *Zelik, S.:* Spatially nondecaying solutions of 2D Navier-Stokes equation in a strip

2005/007 *Meister, A.:* Deconvolving compactly supported densities

2005/008 *Förster, C., Weidl, T.:* Trapped modes for an elastic strip with perturbation of the material properties

2006/001 *Dippon, J., Schiemert, D.:* Stochastic differential equations driven by Gaussian processes with dependent increments

2006/002 *Lesky, P.A.:* Orthogonale Polynomlösungen von Differenzengleichungen vierter Ordnung

2006/003 *Dippon, J., Schiemert, D.:* Option Pricing in a Black-Scholes Market with Memory

2006/004 *Banchoff, T., Kühnel, W.:* Tight polyhedral models of isoparametric families, and PL-taut submanifolds

2006/005 *Walk, H.:* A universal strong law of large numbers for conditional expectations via nearest neighbors

2006/006 *Dippon, J., Winter, S.:* Smoothing spline regression estimates for randomly right censored data

2006/007 *Walk, H.:* Almost sure Cesàro and Euler summability of sequences of dependent random variables

2006/008 *Meister, A.:* Optimal convergence rates for density estimation from grouped data

2006/009 *Förster, C.:* Trapped modes for the elastic plate with a perturbation of Young's modulus

2006/010 *Teufel, E.:* A contribution to geometric inequalities in Euclidean space forms

2006/011 *Kovarik, H.; Vugalter, S.; Weidl, T.:* Spectral estimates for two-dimensional Schrödinger operators with applications to quantum layers

2006/012 *Kovarik, H.; Sacchetti, A.:* Resonances in twisted waveguides

2006/013 *Brehm, U; Kühnel, W.:* Equivelar maps on the torus

2006/014 *Birman, M.Sh.; Suslina, T.A.:* Homogenization with corrector for periodic differential operators. Approximation of solutions in the Sobolev class $H^1(\mathbb{R}^d)$

2007/001 *Meister, A.:* Deconvolution from Fourier-oscillating error densities under decay and smoothness restrictions