Universität
Stuttgart

Fachbereich
Mathematik

Optimal Iterative Density Deconvolution: Upper and Lower Bounds

Christian H. Hesse, Alexander Meister

Stuttgarter Mathematische Berichte

Preprint 2004/002

Universität Stuttgart
Fachbereich Mathematik

Optimal Iterative Density Deconvolution: Upper and Lower Bounds

Christian H. Hesse, Alexander Meister

Stuttgarter Mathematische Berichte

Preprint 2004/002

Fachbereich Mathematik
Universität Stuttgart
Pfaffenwaldring 57
D-70 569 Stuttgart

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

ISSN **1613-8309**

Abstract

Assume that n independent copies of $Y = X + \epsilon$ are observed where ϵ is an unobservable measurement error with a known distribution. We consider the problem of estimating the unknown density of X when this density is known to lie in a given smoothness class. An iterative procedure for estimating the unknown density is introduced. Rates of convergence for mean integrated squared error are studied for smoothness classes arising from Fourier conditions. Minimax rates are derived for these classes. The sequence of estimators resulting from the iterative procedure is shown to attain the optimal rates both for smooth and for supersmooth error densities. The iterative scheme allows one to perform density estimation from contaminated observations by simple additive corrections to an appropriate ordinary kernel density estimator. In this way, the effect of the perturbation due to contamination by ϵ may be quantified. In addition, we demonstrate that the sequence of estimators converges exponentially fast to a specific estimator within the class of deconvoluting kernel density estimators. We also address the subtle theoretical issues that arise when the error density is not in $L_2(\mathbb{R})$ leading to a modification of the iterative procedure.

Keywords: density estimation, contaminated observations, estimator sequence, minimax rates.

AMS 1991 Mathematics Subject Classification: Primary 62G07; Secondary 62G20.

1 Introduction

Deconvolution has been a topic of intensive study during the last decade. Some recent additions to the literature are [24], [25], [20], [18], [32], [13], [23], [7], [8]. Early interest goes back to [6] and [14], when estimation of a cumulative distribution function is considered when data are contaminated by measurement error. More recently, the emphasis in the deconvolution context has been mostly on estimating density functions. [28], [2], [21], [30], [29], [9], [10], [11], [12], [22], [16], [17], [18] constitute examples. Related to our work are also [31], [4], [15], and [19] where adaptive density estimators for certain deconvolution problems are considered. Our results and the methods to prove them are different. Finally, [3] and [33] give book length introductions to deconvolution and density estimation.
The present paper aims to contribute to this line of work.

Assume that $(X_j, \epsilon_j), j = 1, 2, \ldots, n$ are iid. bivariate random vectors where X_j has an unknown density f, ϵ_j has a known density g and is independent of X_j. The aim is to estimate the density f based on the observations

$$Y_j = X_j + \epsilon_j \qquad\qquad j = 1, 2, \ldots, n \tag{1.1}$$

where it is known a priori that f belongs to some class of functions $\mathcal{F}$. Both the X_j and the ϵ_j are unobservable.

One might interpret the observations Y_j as measurements on the X_j that are corrupted by measurement error. But the measurement error context is not the only area where mixture models of type (1.1) arise. These models and the resulting density estimation problems appear naturally in many branches of statistics. Here, we only mention the empirical Bayes approach to compound decision problems as introduced in [27]. Using the notation employed earlier, let $\{X_1, X_2, \ldots, X_n\}$ be a set of parameters and let $f(x)$ be an unknown prior density. Conditionally on $X_j = x$ the observed Y_j are realisations of independent random variables with known parametric density $\tilde{g}(y/x)$. Then, unconditionally, the Y_j are realizations from a distribution with density $\int \tilde{g}(y/x)f(x)dx$. The empirical Bayes approach now consists of estimating $f(x)$ using $Y_1, Y_2, \ldots, Y_{n-1}$ and then utilizes this estimate to obtain the posterior distribution of X_n given Y_n. Thus, for the location model where $\tilde{g}(y/x) = g(y - x)$, we encounter the density estimation problem stated earlier.

When estimating the density of a mixture component, the issue of identifiability arises. We shall call a density f identifiable in a convolution with a density g, if the following implication holds:

$$\int g(y-x)f(x)dx = \int g(y-x)\tilde{f}(x)dx \Longrightarrow f(x) = \tilde{f}(x)\, a.e.$$

Making use of the characteristic functions $\Psi_f, \Psi_{\tilde{f}}, \Psi_g$, identifiability of f follows from

$$\Psi_f(t)\cdot\Psi_g(t) = \Psi_{\tilde{f}}(t)\cdot\Psi_g(t) \Longrightarrow \Psi_f(t) = \Psi_{\tilde{f}}(t)\ \forall t \in \mathbb{R}$$

This implication holds, for example, if $\Psi_g(t) \neq 0$ for all $t \in \mathbb{R}$. Hence, a non-vanishing characteristic function of error guarantees identifiability. This is not the weakest such assumption, but it is simple and we will use it throughout. In this paper we extend existing results in several directions. As in [12], we study minimax rates of convergence for the density estimation problem over certain smoothness classes. But while [12] focusses on Holder classes and L_p-risk on a compact interval, we deal with smoothness classes arising from Fourier conditions and with L_2-risk on the entire real line. [23] is also concerned with L_2-risk on the real line and with a similar function class but only considers smooth error densities. In contrast, we also study the problem for supersmooth error densities. In this context, deconvolution is more difficult. Some subtle theoretical issues arise depending on whether the density of error is square integrable or not. The latter necessitates some modifications in the iterative procedure. We study both cases in detail putting them on firm functional analytic ground. A further innovation is the introduction of an iterative procedure for density estimation. This procedure starts from an ordinary density estimator that ignores the effect of contamination and successively performs additive corrections. In this way, the effect of contamination by the error random variable ϵ may be quantified. In addition, we show that the sequence of estimators resulting from the iterative procedure converges exponentially fast to a specific member in the class of deconvoluting kernel density estimators.

2 Iterative deconvolution for square integrable error densities

In addition to the notation already introduced for the model (1.1), we here define the function class

$$\mathcal{F}^{\beta,a} := \{\text{densities } f | f \in L_2(\mathbb{R}) \text{ and } \int_{\omega}^{\infty} |\Psi_f(t)|^2 dt \leq a\,\omega^{1-2\beta}, \forall \omega > \omega_0\} \tag{2.1}$$

for fixed contants $a > 0, \omega_0 > 0$ and $\beta > \frac{1}{2}$. This is the smoothness class for which we aim to establish minimax results for mean integrated squared error (MISE) and develop an iterative procedure that attains the optimal rate. The iterative procedure turns out to depend on whether the density of error is square integrable or not. Also, the properties of the resulting estimator sequence depend on the rate of decrease of the characteristic function of error. In view of this, we distinguish several cases and first study the problem for densities of error belonging to the class

$$\mathcal{F}_{\epsilon,2,\eta} = \{\text{densities } g | g \in L_2(\mathbb{R}), |\Psi_g(t)| \neq 0, \forall t \in \mathbb{R}, \text{and} |\Psi_g(t)| \geq d|t|^{-\eta}, \quad \forall |t| > T\} \tag{2.2}$$

for some positive constants d, T and $\eta > \frac{1}{2}$.

In accordance with common practice, we refer to the densities in $\mathcal{F}_{\epsilon,2,\eta}$ as smooth densities. An example of such a density is the double-exponential for which $g(x) = \frac{1}{2}\exp(-|x|)$ with $\Psi_g(t) = (1+t^2)^{-1}$.

To introduce the iterative procedure conveniently, we first define several operators. Towards this end we write

$$L_2^{\omega}(\mathbb{R}) = \{f \in L_2(\mathbb{R}) | \Psi_f(t) = 0 \quad \lambda - \text{a.e.} \ \ \forall t \text{ with } |t| > \omega\} \tag{2.3}$$

where λ is Lebesgue measure. $L_2^\omega(\mathbb{R})$ is a closed linear subspace of the Hilbert space $L_2(\mathbb{R})$ and therefore is a Hilbert space itself. Linearity of $L_2^\omega(\mathbb{R})$ is elementary and closedness can be seen as follows: If $(f_n)_{n\in\mathbb{N}}$ is a sequence in $L_2^\omega(\mathbb{R})$ which converges to some $f \in L_2(\mathbb{R})$ in the $L_2(\mathbb{R})-$ Norm $\|\cdot\|_{L_2(\mathbb{R})}$, then $\Psi_f(t) = \Psi_f(t) - \Psi_{f_n}(t) = \Psi_{f-f_n}(t),\ \lambda-$a.e. $\forall t$ with $|t| > \omega$. Here, Ψ_f denotes the Fourier-transform of the function f. Hence,

$$\begin{aligned}
\int_{-\infty}^{-\omega} |\Psi_f(t)|^2 dt + \int_{\omega}^{\infty} |\Psi_f(t)|^2 dt &= \int_{-\infty}^{-\omega} |\Psi_{f-f_n}(t)|^2 |dt + \int_{\omega}^{\infty} |\Psi_{f-f_n}(t)|^2 dt \\
&\leq \int_{\mathbb{R}} |\Psi_{f-f_n}(t)|^2 dt = \|\Psi_{f-f_n}\|^2_{L_2(\mathbb{R})} \\
&= 2\pi \|f - f_n\|^2_{L_2(\mathbb{R})} \stackrel{n\to\infty}{\longrightarrow} 0 \qquad (2.4)
\end{aligned}$$

by Parseval's identity. Therefore, since the left side of (2.4) does not depend on n,
$\int_{-\infty}^{-\omega} |\Psi_f(t)|^2 dt + \int_{\omega}^{\infty} |\Psi_f(t)|^2 dt = 0$
from which it follows that $\Psi_f(t) = 0$ λ-a.e. for t with $|t| > \omega$ and, hence, that $f \in L_2^\omega(\mathbb{R})$.

For every closed linear subspace of a Hilbert space there is a unique operator that orthogonally projects on this subspace. In the case of $L_2^\omega(\mathbb{R})$, we denote this operator by P_ω. Its induced operator norm is $\|P_\omega\|_{L_2(\mathbb{R}),\, ind.} = 1$. Also, $I - P_\omega$, with I being the identity operator, is the unique orthogonal projection operator on the orthogonal complement

$$\left(L_2^\omega(\mathbb{R})\right)^\perp = \{f \in L_2(\mathbb{R}) | \Psi_f(t) = 0\ \lambda\text{ -a.e. for } t \in [-\omega, +\omega]\}$$

and $\|I - P_\omega\|_{L_2(\mathbb{R}),\, ind.} = 1$.

For the Fourier-transform of a projected function f, we have

$$\Psi_{P_\omega f}(t) = \Psi_f(t) \cdot 1_{[-\omega,+\omega]}(t) \qquad (2.5)$$

where $1_A(t)$ is the indicator function of the set A. For large ω and $f \in \mathcal{F}^{\beta,a}$ the distortion caused by an application of P_ω is small. A bound is given by

$$\|P_\omega f - f\|^2_{L_2(\mathbb{R})} \leq \frac{a}{\pi} \omega^{1-2\beta} \qquad (2.6)$$

as can be shown by Parseval's identity, (2.5), and the definition of $\mathcal{F}^{\beta,a}$.

We also introduce convolution operators. Let γ be any $L_2(\mathbb{R})$-density. Then, for any $L_2(\mathbb{R})$-density f

$$C_\gamma f := \gamma * f$$

where $*$ is convolution. C_γ is a continuous linear operator with $\|C_\gamma\|_{L_2(\mathbb{R}),ind.} \leq \|\gamma\|_{L_1(\mathbb{R})} = 1$. Also, since $\gamma \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ and $f \in L_2(\mathbb{R})$, the convolution identity

$$\Psi_{C_\gamma f} = \Psi_\gamma \cdot \Psi_f \qquad (2.7)$$

holds. This identity, however, is no longer valid if we merely assume that $\gamma \in L_1(\mathbb{R})$. This case, i.e. when the error densities are no longer in L_2, will be studied subsequently.

We collect some further properties of these operators, for later use. Clearly, for densities $\gamma, \tilde{\gamma} \in L_2(\mathbb{R})$, the corresponding operators C_γ and $C_{\tilde{\gamma}}$ commute and so do C_γ and P_ω. Furthermore, we have

Proposition 1 *Let $\gamma \in L_2(\mathbb{R})$ be a density and $g \in \mathcal{F}_{\epsilon,2,\eta}$ be the density of error.*

a. *The Hilbert-adjoint operator C_γ^t of C_γ is $C_\gamma^t = C_{\gamma^-}$ where $\gamma^-(x) = \gamma(-x), \forall x$.*

b. *If γ is an even function, then C_γ is Hermitian.*

c. *C_g and C_{g^-} are injective.*

d. *$C_g C_g^t = C_g^t C_g$ is Hermitian and positive-definite.*

Proof. a. Let $f, \tilde{f}$ be in $L_2(\mathbb{R})$. Denote by $\langle \cdot, \cdot \rangle$ the $L_2(\mathbb{R})$ inner product. Then, since for all $t \in \mathbb{R}$ $\Psi_{\gamma^-}(t) = \overline{\Psi_\gamma(t)}$, the complex-conjugate of $\Psi_\gamma(t)$, we have

$$\begin{aligned}
\langle C_\gamma f, \tilde{f} \rangle &= \frac{1}{2\pi} < \Psi_\gamma \Psi_f, \Psi_{\tilde{f}} \rangle = \frac{1}{2\pi} \int_{\mathbb{R}} \Psi_\gamma(t) \Psi_f(t) \overline{\Psi_{\tilde{f}}(t)} dt \\
&= \frac{1}{2\pi} \int\limits_{\mathbb{R}} \Psi_f(t) \overline{\overline{\Psi_\gamma(t)} \Psi_{\tilde{f}}(t)} dt = \frac{1}{2\pi} \langle \Psi_f, \overline{\Psi_\gamma} \Psi_{\tilde{f}} \rangle \\
&= \frac{1}{2\pi} \langle \Psi_f, \Psi_{\gamma -} \Psi_{\tilde{f}} \rangle = \frac{1}{2\pi} \langle \Psi_f, \Psi_{C_{\gamma^-} \tilde{f}} \rangle \\
&= \langle f, C_{\gamma^-} \tilde{f} \rangle
\end{aligned}$$

using Plancherel's identity. b. This follows from a. since for an even function $\gamma^- = \gamma$ and, hence, $C_\gamma^+ = C_{\gamma^-} = C_\gamma$. For c. it suffices to show that the null spaces of C_g, C_{g^-} consist of the zero element only. This holds since $|\Psi_g(t)| = |\Psi_{g^-}(t)| > 0$ for all $t \in \mathbb{R}$ by assumption. d. The stated identity is verified by

$$C_g^t C_g = C_{g^-} C_g = C_{g^- * g} = C_{g * g^-} = C_g C_g^t.$$

Positive-definiteness follows from

$$\langle f, C_g^t C_g f \rangle \; = \; \langle C_g f, C_g f \rangle \; = \; \|C_g f\|^2_{L_2(\mathbb{R})} \geq 0 \text{ for all } f \in L_2(\mathbb{R}) \tag{2.8}$$

with equality in (2.8) if and only if $C_g f = 0$ which by c. is equivalent to $f = 0$ in the $L_2(\mathbb{R})$-sense. Finally,

$$\langle f, C_g^t C_g \tilde{f} \rangle \; = \; \langle C_g f, C_g \tilde{f} \rangle \; = \; \langle C_g^t C_g f, f \rangle$$

so that $(C_g^t C_g)^t = C_g^t C_g$ and $C_g^t C_g$ is Hermitian.

■

The model (1.1) implies the relationship $h = f * g$ between the densities h, f and g of Y_j, X_j and ϵ_j. A possible strategy to obtain an estimator $\hat{f}$ of f is therefore to first estimate h by $\hat{h}$, say, based on the direct observations $Y_1, Y_2, \ldots, Y_n$ and then to solve the functional equation

$$C_g \hat{f} = \hat{h} \tag{2.9}$$

for $\hat{f}$. It follows from Proposition 1c. that (2.9) has at most one solution. Now, C_g is an $L_2(\mathbb{R}) \to L_2(\mathbb{R})$ operator. However, for a given $\hat{h} \in L_2(\mathbb{R})$ and $g \in \mathcal{F}_{\epsilon,2,\eta}$ a solution of (2.9) does not always exist. Setting

$$\hat{h} \text{ with } \psi_{\hat{h}}(t) = \begin{cases} 1 & \text{if } |t| \leq 1 \\ |t|^{-3/4} & \text{if } |t| > 1 \end{cases}$$

$$g(x) = \frac{1}{2}\, exp(-|x|),\; x \in \mathbb{R}$$

provides an example for this claim. A resolution of this difficulty is to restrict attention to certain subsets of $L_2(\mathbb{R})$ in the choice of the estimator $\hat{h}$. For our smoothness class $\mathcal{F}^{\beta,a}$, in view of the results on asymptotic optimality due to [5], it makes sense to restrict attention to estimators $\hat{h} \in L_2^\omega(\mathbb{R})$ for some appropriate sample size dependent $\omega = \omega(n)$. Now, (2.9) is equivalent to $C_g^t C_g \hat{f} = C_g^t \hat{h}$ and since both C_g and C_g^t commute with P_ω we arrive at

$$C_g^t C_g (P_\omega \hat{f}) = C_g^t P_\omega \hat{h} \tag{2.10}$$

to be solved. The operator $C_g^t C_g|_{L_2^\omega(\mathbb{R})}$ (i.e. $C_g^t C_g$ restricted to $L_2^\omega(\mathbb{R})$) is an $L_2^\omega(\mathbb{R}) \to L_2^\omega(\mathbb{R})$ operator and as such it is continuous. Its induced norm is not larger than the norm of $C_g^t C_g$ viewed as an $L_2(\mathbb{R}) \to L_2(\mathbb{R})$ operator. For any $\hat{h} \in L_2(\mathbb{R})$, the functional equation (2.10) possesses exactly one solution $P_\omega \hat{f}$ if ω is sufficiently large. This follows from

Proposition 2 *For $g \in \mathcal{F}_{\epsilon,2,\eta}$ the operator $C_g^t C_g|_{L_2^\omega(\mathbb{R})}$ is invertible and for ω sufficiently large it is*

$$\|I - C_g^t C_g\|_{L_2^\omega(\mathbb{R}),ind.} \leq 1 - d^2\omega^{-2\eta} < 1 \tag{2.11}$$

where

$$\|I - C_g^t C_g\|_{L_2^\omega(\mathbb{R}),ind.} = \sup_{f \in L_2^\omega(\mathbb{R})\setminus\{0\}} \frac{\|(I - C_g^t C_g)f\|_{L_2^\omega(\mathbb{R})}}{\|f\|_{L_2^\omega}(\mathbb{R})}$$

is the induced $L_2^\omega(\mathbb{R})$-norm. The Neumann series $\sum\limits_{j=0}^{\infty}(I - C_g^t C_g)^j$ converges in the induced $L_2^\omega(\mathbb{R})$-norm to $(C_g^t C_g|_{L_2^\omega(\mathbb{R})})^{-1}$.

Proof. It suffices to show (2.11). Then, both invertibility of the operator and convergence of the Neumann series follow from known results about Neumann series in Hilbert spaces (e.g. [26]). It is

$$\begin{aligned}
\|I - C_g^t C_g\|_{L_2^\omega,ind.} &= \sup_{f \in L_2^\omega(\mathbb{R})\setminus\{0\}} \frac{\|\Psi_f - \Psi_{g^-}\Psi_g\Psi_f\|_{L_2(\mathbb{R})}}{\|\Psi_f\|_{L_2^\omega(\mathbb{R})}} \\
&= \sup_{f \in L_2^\omega(\mathbb{R})\setminus\{0\}} \frac{(\int\limits_{-\omega}^{+\omega} |\Psi_f(t)|^2 |1 - |\Psi_g(t)|^2|^2 dt)^{\frac{1}{2}}}{(\int_{-\omega}^{+\omega} |\Psi_f(t)|^2 dt)^{\frac{1}{2}}} \\
&\leq \sup_{t \in [-\omega,+\omega]} \left\{|1 - |\Psi_g(t)|^2|\right\} \\
&= 1 - \min\{\inf_{t \in [-\omega,+\omega]\setminus[-T,+T]} \{|\Psi_g(t)|^2\}, \inf_{t \in [-T,+T]} \{|\Psi_g(t)|^2\}\} \ \text{ for } \omega > T. \\
&\leq 1 - d^2\omega^{-2\eta} \ \text{ for } \omega > \max\left\{\left(\frac{d}{|\Psi_g(t_{\min})|}\right)^{\frac{1}{\eta}}, T\right\}
\end{aligned}$$

where t_{min} is such that $\inf\limits_{t \in [-T,+T]} \{|\Psi_g(t)|^2\} = |\Psi_g(t_{min})|^2$. ■

Formally, the unique solution $P_\omega \hat{f}$ of (2.10) is

$$P_\omega \hat{f} = (C_g^t C_g)^{-1} C_g^t P_\omega \hat{h} \tag{2.12}$$

for ω large enough. The inverse operator $(C_g^t C_g)^{-1}$ is in general not easy to compute. But Proposition 2 allows us to represent the solution (2.12) by means of a Neumann series as

$$P_\omega \hat{f} = \sum_{j=0}^{\infty} (I - C_g^t C_g)^j C_g^t P_\omega \hat{h}$$

Truncation of this series then leads to the estimator sequence

$$\hat{f}^{(k)} = \sum_{j=0}^{k} (I - C_g^t C_g)^j C_g^t P_\omega \hat{h} \quad , \quad k \in \mathbb{N}_0 \tag{2.13}$$

or in recursive notation

$$\begin{aligned}
\hat{f}^{(0)} &= C_g^t P_\omega \hat{h} \\
\hat{f}^{(1)} &= 2\hat{f}^{(0)} - C_g^t C_g \hat{f}^{(0)} \\
\hat{f}^{(k+2)} &= 2\hat{f}^{(k+1)} - \hat{f}^{(k)} - C_g^t C_g(\hat{f}^{(k+1)} - \hat{f}^{(k)}) \quad , \quad k \in \mathbb{N}_0
\end{aligned} \tag{2.14}$$

The sequence $(\hat{f}^{(k)})_{k\in\mathbb{N}_0}$ converges to $P_\omega \hat{f}$ exponentially fast . Specifically, we have

Proposition 3 *For $(\hat{f}^{(k)})_{k\in\mathbb{N}_0}$ as defined by (2.14) and $P_\omega \hat{f}$ as given in (2.12)*

$$\|\hat{f}^{(k)} - P_\omega \hat{f}\|_{L_2(\mathbb{R})} \leq (1 - d^2\omega^{-2\eta})^{k+1} \|P_\omega \hat{f}\|_{L_2(\mathbb{R})}$$

for all k larger than some $k_0 \in \mathbb{N}_0$ if ω is sufficiently large.

Proof. Clearly,

$$\| \sum_{j=k+1}^{N} (I - C_g^t C_g)^j C_g^t P_\omega \hat{h} \|_{L_2(\mathbb{R})} \quad \stackrel{N\to\infty}{\longrightarrow} \quad \|\hat{f}^{(k)} - P_\omega \hat{f}\|_{L_2(\mathbb{R})}$$

as well as

$$\| \sum_{j=k+1}^{N} (I - C_g^t C_g)^j C_g^t P_\omega \hat{h} \|_{L_2(\mathbb{R})} \leq \|I - C_g^t C_g\|^{k+1}_{L_2^\omega(\mathbb{R}),ind.} \| \sum_{j=0}^{N-k-1} (I - C_g^t C_g)^j C_g^t P_\omega \hat{h} \|_{L_2(\mathbb{R})} \tag{2.15}$$

so that for $N \to \infty$ the right side of (2.15) converges to $\|I - C_g^t C_g\|^{k+1}_{L_2^\omega(\mathbb{R}),ind.} \|P_\omega \hat{f}\|$ and, therefore,

$$\begin{aligned}
\|\hat{f}^{(k)} - P_\omega \hat{f}\|_{L_2(\mathbb{R})} &\leq \|I - C_g^t C_g\|^{k+1}_{L_2^\omega(\mathbb{R}),ind.} \|P_\omega \hat{f}\|_{L_2(\mathbb{R})} \\
&\leq (1 - d^2\omega^{-2\eta})^{k+1} \|P_\omega \hat{f}\|_{L_2(\mathbb{R})}
\end{aligned}$$

for ω large enough by Proposition 2. ■

The iterative scheme (2.14) is somewhat awkward since a given sequence element depends on the two preceding elements. We, therefore, look for an alternative. Towards this end, consider that $C_g \hat{f} = \hat{h}$ implies that

$$P_\omega \hat{f} = P_\omega \hat{f} + U(P_\omega \hat{h} - C_g P_\omega \hat{f}) \tag{2.16}$$

for any linear continuous operator $U : L_2^\omega \to L_2^\omega$. If, in addition, the operator U is taken to be injective, then the solutions of the equations $C_g P_\omega \hat{f} - P_\omega \hat{h} = 0$ and $U(C_g P_\omega \hat{f} - P_\omega \hat{h}) = 0$ coincide. From (2.16) we deduce a fixed-point identity for $P_\omega \hat{f}$, namely

$$P_\omega \hat{f} = \Gamma(P_\omega \hat{f}) \tag{2.17}$$

with Γ defined by $\Gamma(\phi) = UP_\omega \hat{h} + (I - UC_g)\phi$ for $\phi \in L_2^\omega(\mathbb{R})$. Γ is an $L_2^\omega(\mathbb{R}) \to L_2^\omega(\mathbb{R})$ operator and it is a contraction if $\|I - UC_g\|_{L_2^\omega(\mathbb{R}),ind.} < 1$. In this case, by an application of the fixed-point theorem, (2.17) possesses a unique solution, namely (2.12), and the sequence of iterations

$$\tilde{f}^{(k+1)} := \Gamma(\tilde{f}^{(k)}) \tag{2.18}$$

converges to this solution for any initialization $\tilde{f}^{(0)} \in L_2^\omega(\mathbb{R})$.
The advantage of this alternative approach is that we may fine-tune our iterative scheme to the actual error density g and optimize with regard to speed of convergence by chosing an appropriate U. For example, the choice $U = C_g^t$, which is a linear continuous injective operator by Proposition

1 with $\|I - UC_g\|_{L_2^\omega(\mathbb{R}),ind.} < 1$ by Proposition 2 is seen to lead to the iteration (2.14) for the initialization $\hat{f}^{(0)} = C_g^t P_\omega \hat{h}$.
We now motivate a different choice for U resulting in enhanced speed of convergence of the iteration scheme. According to Proposition 1d, $C_g^t C_g$ is positive-definite and Hermitian. Consequently, there exists a unique Hermitian, positive-definite operator C, the root of $C_g^t C_g$, such that $C^2 = C_g^t C_g$.

Lemma 1 *a. The root C of $C_g^t C_g$ has the representation*

$$C = \Psi^{-1} M_{|\Psi_g|} \Psi$$

with

$$\begin{aligned} \Psi : \quad L_2(\mathbb{R}) &\longrightarrow L_2(\mathbb{R}) \\ f &\longmapsto \Psi_f \end{aligned}$$

and M_ξ is the multiplication operator, i.e. $M_\xi f = \xi \cdot f$.
b. For the root C of $C_g^t C_g$ we have

$$\Psi_{Cf} = |\Psi_g| \cdot \Psi_f \quad , \quad \forall f \in L_2(\mathbb{R})$$

Proof. Observe that

$$\Psi C^2 \Psi^{-1} = \Psi C_g^t C_g \Psi^{-1} = M_{|\Psi_g|^2} \Psi \Psi^{-1} = M_{|\psi_g|^2} \tag{2.19}$$

where $\Psi_{C_g^t C_g f} = |\Psi_g|^2 \Psi_f$, $\forall f \in L_2(\mathbb{R})$ was used.
Quite generally, multiplication operators are well-defined on $L_2(\mathbb{R})$. They are linear and continuous, if ξ is bounded. They are Hermitian and positive-definite, if ξ is $\mathbb{R}_+$-valued. Hence, in particular, since $0 < |\Psi_g(t)| \leq 1$, $\forall t \in \mathbb{R}$, $M_{|\Psi_g|}$ and $M_{|\Psi_g|^2}$ are linear, continuous, Hermitian, positive-definite $L_2(\mathbb{R}) \to L_2(\mathbb{R})$ operators. Furthermore,

$$M_{|\Psi_g|^2} = M_{|\Psi_g|}^2. \tag{2.20}$$

The operator $\Psi C \Psi^{-1}$ also is positive-definite and Hermitian. This is so since for any $f, f^* \in L_2(\mathbb{R})$

$$\begin{aligned} \langle \Psi C \Psi^{-1} f, f^* \rangle &= 2\pi \langle C \Psi^{-1} f, \Psi^{-1} f^* \rangle \quad = 2\pi \langle \Psi^{-1} f, C \Psi^{-1} f^* \rangle \\ &= \langle f, \Psi C \Psi^{-1} f^* \rangle. \end{aligned}$$

Setting $f^* = f$,

$$\langle \Psi C \Psi^{-1} f, f \rangle = 2\pi \langle C \Psi^{-1} f, \Psi^{-1} f \rangle \geq 0$$

due to positive-definiteness of C. Here, equality holds if $\Psi^{-1} f = 0$, which is equivalent to $f = 0$ (in the $L_2(\mathbb{R})$-sense) due to Ψ being an isomorphism. Furthermore,

$$(\Psi C \Psi^{-1})^2 = \Psi C^2 \Psi^{-1} = M_{|\Psi_g|^2} \tag{2.21}$$

by (2.19). As evidenced by (2.20) and (2.21) we thus have identified two positive-definite, Hermitian operators whose square is $M_{|\Psi_g|^2}$, namely $\Psi C \Psi^{-1}$ and $M_{|\Psi_g|}$. But since $M_{|\Psi_g|^2}$ is itself positive-definite and Hermitian, it possesses a unique positive-definite, Hermitian root. Hence,

$$M_{|\Psi_g|} = \Psi C \Psi^{-1}$$

from which we get

$$M_{|\Psi_g|} \Psi = \Psi C \tag{2.22}$$

and $\Psi^{-1}M_{|\Psi_g|}\Psi = C$. This proves a. Applying (2.22) to some $f \in L_2(\mathbb{R})$ results in $|\Psi_g|\Psi_f = \Psi_{Cf}$ which proves b. ■

The operator $C\,|_{L_2^\omega(\mathbb{R})}$ is invertible; its inverse may be approximated by a Neumann series. In fact, for ω large enough

$$\|I - C\|_{L_2^\omega(\mathbb{R}),ind.} \leq 1 - d\omega^{-\eta} \tag{2.23}$$

which may be proved similarly to Proposition 2. Now, we define $U : L_2^\omega(\mathbb{R}) \to L_2^\omega(\mathbb{R})$ as

$$U := C^{-1}C_g^t \tag{2.24}$$

Then, U is an isometry, e.g.

$$\begin{aligned} U^tU &= (C^{-1}C_g^t)^tC^{-1}C_g^t \\ &= C_g((C_gC_g^t)^{-1})C_g^t \\ &= C_gC_g^t(C_gC_g^t)^{-1} \\ &= I \end{aligned}$$

since C_g and C_g^t commute and so do $(C_gC_g^t)^{-1}$ and C_g^t. Being an isometry, U is also an injective, linear operator. Using Lemma 1a. and
$M_\xi M_{\xi^{-1}} = I$, U may be represented as

$$U = \Psi^{-1}M_{\overline{\Psi_g}/|\Psi_g|}\Psi$$

With this choice of U, the iterative scheme now reads

$$\overline{f}^{(k+1)} = C^{-1}C_g^tP_\omega\hat{h} + (I - C)P_\omega\overline{f}^{(k)} \tag{2.25}$$

$(\overline{f}^{(k)})_{k\in\mathbb{N}}$ is an estimator sequence in $L_2^\omega(\mathbb{R})$ since C, C^{-1}, $I - C$ are all $L_2^\omega(\mathbb{R}) \to L_2^\omega(\mathbb{R})$-operators. Compared to the previous choice $U = C_g^t$, speed of convergence has been increased, specifically we have

Proposition 4 *For the sequence of estimators $(\overline{f}^{(k)})_{k\in\mathbb{N}}$ defined by (2.25) with some $\overline{f}^{(0)} \in L_2(\mathbb{R})$ it is*

$$\|\overline{f}^{(k+1)} - P_\omega\hat{f}\|_{L_2(\mathbb{R})} \;\leq\; c(1 - d\omega^{-\eta})^{k+1} \;<\; 1$$

if ω is large enough. Here and throughout c denotes a generic positive constant that may change from one occurrence to another.

Proof. It is

$$\begin{aligned} \|\overline{f}^{(k+1)} - P_\omega\hat{f}\|_{L_2(\mathbb{R})} &= \|C^{-1}C_g^tP_\omega\hat{h} - (I - C)P_\omega\overline{f}^{(k)} - P_\omega\hat{f}\|_{L_2(\mathbb{R})} \\ &= \|(I - C)(P_\omega\overline{f}^{(k)} - P_\omega\hat{f})\|_{L_2^(\mathbb{R})} \quad \text{in view of (2.18)} \\ &\leq \|I - C\|_{L_2^\omega(\mathbb{R}),ind.}\|\overline{f}^{(k)} - P_\omega\hat{f}\|_{L_2(\mathbb{R})} \\ &\leq (1 - d\omega^{-\eta})^{k+1}\|\overline{f}^{(0)} - P_\omega\hat{f}\|_{L_2(\mathbb{R})} \end{aligned}$$

for ω large enough by (2.22). ■

In general, the operator C can be difficult to obtain. We consider two special cases where this is possible, where, in particular, C can be viewed as a convolution operator C_γ for some $L_2(\mathbb{R})$-density γ. This requires

$$C_\gamma^2 = C_{g^-*g} \;\Leftrightarrow\; \Psi_\gamma^2 = |\Psi_g|^2 \;\Leftrightarrow\; \Psi_\gamma = \pm|\Psi_g(t)|$$

which in turn is equivalent to

$$\Psi_\gamma = |\Psi_g(t)| \tag{2.26}$$

due to continuity of Ψ_γ, positivity of $|\Psi_g(t)|$ and the fact that $\Psi_\gamma(0) = 1$. Hence Ψ_γ is a real-valued positive function. If the error density g is even, then $C_g = C_g^t$ and C_g is both positive-definite and Hermitian. Since $C_g^2 = C_g C_g^t = C^2$ it follows by uniqueness of the positive-definite root that

$$C = C_g = C_g^t.$$

Hence, $\gamma = g$ can be chosen in (2.26) resulting in $U = C^{-1}C_g^t = I$. The iterative scheme (2.25) becomes

$$\overline{f}^{(k+1)} = P_\omega \hat{h} + (I - C_g) P_\omega \overline{f}^{(k)}.$$

Error densities g for which there exists an $L_2(\mathbb{R})$-density ζ such that $\zeta * \zeta = g$ constitute the second special case we here consider. Generalizing the concept of infinite divisibility, we call these densities divisible. Now, if g is divisible then $g * g^- = \zeta * \zeta * \zeta^- * \zeta^- = (\zeta * \zeta^-) * (\zeta * \zeta^-)$ leading to $\gamma = \zeta * \zeta^- \in L_2(\mathbb{R})$. In this case $C_\gamma = C = C_\zeta C_\zeta^t$. Without additional assumptions C^{-1} is difficult to compute explicitely. It may, however, be approximated by its Neumann series. Writing $S := I - C$ one obtains by induction

$$\overline{f}^{(k)} = S^k \overline{f}^{(0)} + \sum_{j=0}^{k-1} S^j C^{-1} C_g^t P_\omega \hat{h} \tag{2.27}$$

and replacing C^{-1} by $\sum_{l=0}^{m} S^l$ leads to

$$\overline{f}_m^{(k)} = S^k \overline{f}^{(0)} + \sum_{j=0}^{k-1}\sum_{l=0}^{m} S^{j+l} C_g^t P_\omega \hat{h}.$$

Concerning the rate of convergence, consider that

$$\begin{aligned} \|P_\omega \hat{f} - \overline{f}_m^{(k)}\|_{L_2(\mathbb{R})} &\le \|P_\omega \hat{f} - \hat{f}_m^{(k)}\|_{L_2(\mathbb{R})} + \|\hat{f}^{(k)} - \overline{f}_m^{(k)}\|_{L_2(\mathbb{R})} \\ &\le c(1 - d\omega^{-\eta})^k + \|\sum_{l=m+1}^{\infty} S^l \sum_{j=0}^{k-1} S^j C_g^t P_\omega \hat{h}\|_{L_2(\mathbb{R})} \end{aligned}$$

and

$$\begin{aligned} \|\sum_{l=m+1}^{\infty} S^l \sum_{j=0}^{k-1} S^j C_g^t P_\omega \hat{h}\|_{L_2(\mathbb{R})} &\le c \sum_{l=m+1}^{\infty} (1 - d\omega^{-\eta})^l \|C^{-1} C_g^t P_\omega \hat{h}\|_{L_2(\mathbb{R})} \\ &\le c \left(\frac{1}{1-(1-d\omega^{-\eta})} - \frac{1-(1-d\omega^{-\eta})^{m+1}}{1-(1-d\omega^{-\eta})} \right) \end{aligned}$$

so that

$$\|P_\omega \hat{f} - \overline{f}_m^{(k)}\|_{L_2(\mathbb{R})} \le c(1 - d\omega^{-\eta})^k + c(1 - d\omega^{-\eta})^{m+1}$$

With $m = k - 1$, the sequence $(\overline{f}_m^{(k)})_{k\in\mathbb{N}}$ has the same rate of convergence as $(\overline{f}^{(k)})_{k\in\mathbb{N}}$ in (2.27). Then, writing $\overline{f}_*^{(k)}$ for $\overline{f}_{k-1}^{(k)}$,

$$\begin{aligned} \overline{f}_*^{(k)} &= S^k \hat{f}^{(0)} + \sum_{j,l=0}^{k-1} S^{j+l} C_g^t P_\omega \hat{h} \\ &= S^k \hat{f}^{(0)} + \sum_{j=0}^{2k-2} [k - |j - k + 1|] S^j C_g^t P_\omega \hat{h}. \end{aligned}$$

This scheme works whenever the operator C can be computed.

3 Minimax result

In this section, we establish a minimax rate for the mean integrated squared error over the function class $\mathcal{F}^{\beta,a}$ and we prove that the estimator

$$P_\omega \hat{f} = (C_g^t C_g)^{-1} C_g^t P_\omega \hat{h} \tag{3.1}$$

for an appropriate ordinary density estimator $\hat{h}$ achieves the optimal rate. In particular, we take

$$\hat{h}_\lambda(x) = \frac{1}{n}\sum_{i=1}^{n} K_\lambda(x - Y_i) \tag{3.2}$$

where $K_\lambda(x) = \lambda K(\lambda x)$ is a kernel and $\lambda = \lambda(n)$ is a scaling parameter. [34] shows that the minimum MISE J_n^* within the class of kernel estimators is

$$J_n^* = \frac{1}{2\pi}\int \frac{|\Psi_h(t)|^2(1-|\Psi_h(t)|^2)}{1+(n-1)|\Psi_h(t)|^2} dt$$

The optimal kernel depends on the unknown density h. [5] studies MISE of the kernel density estimator based on the sync kernel $\tilde{K}(x) = (\pi x)^{-1}\sin x$ whose characteristic function is $\Psi_{\tilde{K}}(t) = 1$ for $|t| \leq 1$ and zero otherwise. For densities satisfying certain Fourier conditions similar to the ones considered here, the rate of decrease of its MISE is of the same order as J_n^*. Hence $\hat{h}_\lambda$ in (3.2) with the sync kernel $\tilde{K}_\lambda$ and $\lambda = \omega$ is a plausible choice for an ordinary density estimator for the class of functions here considered. We will show that with this choice the MISE-minimax rate for deconvolution is in fact attained: Using $\hat{h}_\lambda$ in (3.2) with $K_\lambda = \tilde{K}_\omega$, the sync kernel, we obtain

$$\hat{f}_\omega := P_\omega \hat{f} = (C_g^t C_g)^{-1} C_g^t \hat{h}_\omega \tag{3.3}$$

since $P_\omega \hat{h}_\omega = \hat{h}_\omega$. We first establish an upper bound for MISE uniform over the function class $\mathcal{F}^{\beta,a}$.

Theorem 1 *For error densities g from $\mathcal{F}_{\epsilon,2,\eta}$ and with $\omega = \omega_n = cn^{\frac{1}{2(\beta+\eta)}}$ it is*

$$\sup_{f\in\mathcal{F}^{\beta,a}} E_f \|\hat{f}_{\omega_n} - f\|^2_{L_2(\mathbb{R})} \leq c\, n^{-\frac{2\beta-1}{2(\beta+\eta)}}$$

with $\hat{f}_{\omega_n}$ as in (3.3).

Proof. First,

$$\sup_{f\in\mathcal{F}^\beta} E_f\|\hat{f}_{\omega_n} - f\|^2_{L_2(\mathbb{R})} \leq 2\sup_{f\in\mathcal{F}^\beta} E_f\|\hat{f}_{\omega_n} - P_{\omega_n} f\|^2_{L_2(\mathbb{R})} + 2\sup_{f\in\mathcal{F}^\beta}\|P_{\omega_n} f - f\|^2_{L_2(\mathbb{R})}$$

and the second summand on the right hand side can be handled by (2.6). To bound the remaining summand, consider that with $h = f * g$ it is $P_{\omega_n} h = P_{\omega_n} C_g f = C_g P_{\omega_n} f$ from which we get $C_g^t P_{\omega_n} h = C_g^t C_g P_{\omega_n} f$ and, hence,

$$P_{\omega_n} f = (C_g^t C_g)^{-1} C_g^t P_{\omega_n} h.$$

This gives

$$\sup_{f\in\mathcal{F}^{\beta,a}} E_f\|\hat{f}_{\omega_n} - P_{\omega_n} f\|^2_{L_2(\mathbb{R})} = \sup_{f\in\mathcal{F}^\beta} E_f\|(C_g^t C_g)^{-1} C_g^t P_{\omega_n}(\hat{h}_{\omega_n} - h)\|^2_{L_2(\mathbb{R})}$$

where $P_{\omega_n}(\hat{h}_{\omega_n} - h)$ is in $L_2^{\omega_n}(\mathbb{R})$. For any $\xi \in L_2^\omega(\mathbb{R})$, it is $(C_g^t C_g)^{-1} C_g^t \xi = \eta$ for some $\eta \in L_2^\omega(\mathbb{R})$. Since $\Psi_{C_g^t C_g \eta} = \Psi_{C_g^t \xi}$, this implies $\overline{\Psi_g}\Psi_\xi = |\Psi_g|^2\Psi_\eta$ and hence $\Psi_\eta = \Psi_\xi/\Psi_g$. As a result, we obtain

$$\begin{aligned}
&\sup_{f\in\mathcal{F}^{\beta,a}} E_f\|\hat{f}_{\omega_n}-P_{\omega_n}f\|_{L_2(\mathbb{R})}\\
&\le \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}^{\beta,a}}\int_{-\omega_n}^{+\omega_n}|\Psi_g(t)|^{-2}E_f|\Psi_{\hat{h}_{\omega_n}}(t)-\Psi_h(t)|^2dt\\
&\le \tfrac{1}{2\pi}\sup_{f\in\mathcal{F}^{\beta,a}}\int_{-\omega_n}^{+\omega_n}\left(\tfrac{1}{n}1_{[-\omega_n,+\omega_n]}(t)|\Psi_g(t)|^{-2}+\left|\tfrac{\Psi_h(t)}{\Psi_g(t)}\right|^2(1-1_{[-\omega_n,+\omega_n]})\right)dt
\end{aligned} \tag{3.4}$$

since

$$\begin{aligned}
&E_f|\Psi_{\hat{h}_{\omega_n}}(t)-\Psi_h(t)|^2 \;=\; E_f\left|\tfrac{1}{n}\sum_{j=1}^{n}e^{itY_j}(1_{[-\omega_n,+\omega_n]}(t)-\Psi_h(t))\right|^2\\
&=E_f\left|\tfrac{1}{n}\sum_{j=1}^{n}e^{itY_j}1_{[-\omega_n,+\omega_n]}(t)-\Psi_h(t)1_{[-\omega_n,+\omega_n]}(t)\right|^2\;+\;|\Psi_h(t)(1_{[-\omega_n,+\omega_n]}(t)-1)|^2\\
&\le \tfrac{1}{n}E_f|e^{itY_j}|^2\cdot 1_{[-\omega_n,+\omega_n]}(t)\;+\;|\Psi_h(t)(1_{[-\omega_n,+\omega_n]}(t)-1)|^2\\
&=\tfrac{1}{n}1_{[-\omega_n,+\omega_n]}(t)\;+\;|\Psi_h(t)|^2(1-1_{[-\omega_n,+\omega_n]}(t))
\end{aligned}$$

With $\Psi_h=\Psi_f\Psi_g$, we finally arrive at

$$\begin{aligned}
\sup_{f\in\mathcal{F}^{\beta,a}} E_f\|\hat{f}_{\omega_n}-f\|_{L_2(\mathbb{R})} \;&\le c\,\omega_n^{1-2\beta}\;+\;\tfrac{1}{\pi n}\int_{-\omega_n}^{+\omega_n}|\Psi_g(t)|^{-2}dt\\
&+\tfrac{1}{\pi}\sup_{f\in\mathcal{F}^{\beta,a}}\int_{-\omega_n}^{+\omega_n}|\Psi_f(t)|^2(1-1_{[-\omega_n,+\omega_n]}(t))dt\\
&=c\,\omega_n^{1-2\beta}\;+\;\tfrac{2}{\pi n}\int_0^T|\Psi_g(t)|^{-2}dt\;+\;\tfrac{2}{\pi n}\int_T^{\omega_n}|\Psi_g(t)|^{-2}dt\\
&\le c\,\omega_n^{1-2\beta}\;+\;c\,n^{-1}\;+\;c\,n^{-1}\omega_n^{2\eta+1}
\end{aligned}$$

which is optimized for $\omega_n=n^{\frac{1}{2(\beta+\eta)}}$ and then gives the bound stated in the theorem. ■

We can also obtain a corresponding lower bound of the same order if g is in the subset $\mathcal{F}'_{\epsilon,2,\eta}:=\{g\in\mathcal{F}_{\epsilon,2}\mid|\Psi_g^{(l)}(t)|\le\tilde{d}t^{-(\eta+l)}\,,\,l=0,1,\ldots\}$

Theorem 2 *For error densities g from $\mathcal{F}'_{\epsilon,2,\eta}$ we have*

$$\inf_{\hat{T}\in\mathcal{T}_n}\sup_{f\in\mathcal{F}^{\beta,a}}E_f\|\hat{T}-f\|^2_{L_2(\mathbb{R})}\;\ge\;c\,n^{-\frac{2\beta-1}{2(\beta+\eta)}}$$

where $\mathcal{T}_n$ is the set of all estimators based on n iid observations.

Proof.

Our strategy of proof leans on [12] in places. Then, in turn, the method of [1] is utilized, also known as Assouad's lemma. However, for our function class $\mathcal{F}^{\beta,a}$ the method requires significant modifications.

First, we introduce the Sobolev class of densities

$$\mathcal{H}_{s,a}:=\{L_2(\mathbb{R})\text{-densities }f\mid\||t|^s\Psi_f(t)\|^2_{L_2(\mathbb{R})}\le a\}$$

Since $f\in H_{s,a}$ implies

$$a\ge\int_\omega^\infty|t|^{2s}|\Psi_f(t)|^2dt\ge\omega^{2s}\int_\omega^\infty|\Psi_f(t)|^2dt$$

and, hence, $\int_{\omega}^{\infty} |\Psi_f(t)|^2 dt \leq a\omega^{-2s}$ for all $\omega \geq \omega_0 > 0$, we have the inclusion

$$\mathcal{H}_{\beta-\frac{1}{2},a} \subseteq \mathcal{F}^{\beta,a}.$$

Now, let $H \in \mathcal{H}_{\beta-\frac{1}{2},a/4q}$ (for some constant q still to be determined) be some bounded function with compact support integrating to zero. Also, define

$$\Phi(x) := \begin{cases} -\frac{1}{b-a} & \text{for } a-b < x < 0 \\ \frac{1}{b-a} & \text{for } 0 < x < b-a \\ 0 & \text{otherwise} \end{cases}$$

and

$$f_0(t) = (1+t^2)^{-r}$$

For appropriate r, $f_0 \in \mathcal{H}_{\beta-\frac{1}{2},a/4}$ can be assured. With these definitions in place, let

$$f_\Theta(t) := f_0(t) + a_n^{-1} \sum_{j=1}^{m_n} \Theta_j (\Phi * H)(m_n(t - x_{n,j}))$$

with $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_{m_n})$, $a_n = m_n^{\beta-\frac{1}{2}}$, $x_{n,j} = a + \frac{j}{n}(b-a)$, $j = 1, \ldots, m_n$ and set

$$\mathcal{F}_n := \{f_\Theta \mid \Theta \in \{0,1\}^{m_n}\}.$$

The elements of $\mathcal{F}_n$ are $L_2(\mathbb{R})$-functions. For later use we note that

$$\int_0^{b-a} |(\Phi * H)(x)|^2 dx > 0 \tag{3.5}$$

Next we will show that at least a sufficiently large subset of $\mathcal{F}_n$ lies in $\mathcal{H}_{\beta-\frac{1}{2},a}$ and, thus, in $\mathcal{F}^{\beta,a}$. Towards this end, write

$$b_{n,\Theta}(t) := \sum_{j=1}^{m_n} \Theta_j e^{i(b-a)tj} \Psi_\Phi(t) \tag{3.6}$$

and for $f \in \mathcal{F}_n$ consider that

$$\int |t|^{2\beta-1} |\Psi_f(t)|^2 dt \leq \frac{a}{2} + 2 \int |t|^{2\beta-1} m_n^{-1} |b_{n,\Theta}(t)|^2 |\Psi_H(t)|^2 dt \tag{3.7}$$

To evaluate this bound, let $\hat{\Theta}$ be an m_n-dimension random vector whose components $\hat{\Theta}_j$ are iid. $B(1, \frac{1}{2})$ random variables and show that

$$E(|b_{n,\hat{\Theta}}(t)|^2) \leq c\, m_n \quad \forall t \in \mathbb{R}, \forall n \in \mathbb{N} \tag{3.8}$$

This is true, since

$$\begin{aligned} E(|b_{n,\hat{\Theta}}(t)|^2) &= var(b_{n,\hat{\Theta}}(t)) + |E(b_{n,\hat{\Theta}}(t))|^2 \\ &\leq var\left(\sum_{j=1}^{m_n} \hat{\Theta}_j |e^{i(b-a)tj} \Psi_\Phi(t)| \right) + \frac{1}{4} \left| \sum_{j=1}^{m_n} e^{i(b-a)tj} \Psi_\Phi(t) \right|^2 \\ &\leq m_n + 4 \left| \frac{\cos((b-a)t) - 1}{(1 - e^{i(b-a)t})(b-a)t} \right|^2 \end{aligned}$$

and the second summand is a continuous function of t and uniformly bounded. Note that for a complex-valued random variable Z the variance is defined as $var(Z) := E(|Z - E(Z)|^2)$. Then

the identity $var\left(\sum_{j=1}^{n} Z_j\right) = \sum_{j=1}^{n} var(Z_j)$ for iid. Z_j continues to hold.
Now, using (3.8), we conclude with Markov's inequality that

$$\begin{aligned} & P\left(\int |t|^{2\beta-1} m_n^{-1} |b_{n,\hat{\Theta}}(t)|^2 |\Psi_H(t)|^2 dt \le a/4\right) \\ & \ge 1 - \tfrac{4}{a} E\left(\int |t|^{2\beta-1} m_n^{-1} |b_{n,\hat{\Theta}}(t)|^2 |\Psi_H(t)|^2 dt\right) \\ & \ge 1 - \tfrac{4c}{a} \int |t|^{2\beta-1} |\Psi_H(t)|^2 dt \\ & \ge 1 - \tfrac{c}{q} =: p > \tfrac{3}{4} \end{aligned} \tag{3.9}$$

by appropriate choice of the constant q.
Hence we have shown that there exists a set $\Lambda_n \subseteq \{0,1\}^{m_n}$ with $2^{-m_n} card(\Lambda_n) > \frac{3}{4}$ such that

$$\mathcal{F}'_n := \{f_\Theta \mid \Theta \in \Lambda_n\} \subseteq \mathcal{H}_{\beta-\frac{1}{2},a}$$

We will now establish that for this subset $\mathcal{F}'_n$ and with $-\infty < a < b < +\infty$

$$\sup_{f\in\mathcal{F}'_n} E_f \int_a^b |\hat{f}(x) - f(x)|^2 dx \ \ge\ c\, n^{-\frac{2\beta-1}{2(\beta+\eta)}}$$

where $\hat{f}$ is any estimator of f based on $Y_1, Y_2, \ldots, Y_n$. Then the statement of the theorem follows. First, observe that

$$\begin{aligned} & \sup_{f\in\mathcal{F}'_n} E_f \int_a^b |\hat{f}(x) - f(x)|^2 dx \\ & = \sup_{\Theta\in\Lambda_n} E_{f_\Theta} \int_a^b |\hat{f}(x) - f_\Theta(x)|^2 dx \\ & \ge P(\hat{\Theta} \in \Lambda_n) \cdot \int_a^b E_{\hat{\Theta}} \left(E_{f_{\hat{\Theta}}} |\hat{f}(x) - f_{\hat{\Theta}}(x)|^2 \mid \hat{\Theta} \in \Lambda_n\right) dx \end{aligned} \tag{3.10}$$

Next, define the sets

$$\begin{aligned} \Lambda^{(j,i)} &:= \{\Theta \in \{0,1\}^{m_n} \mid \Theta_j = i\} \quad ; i = 0,1 \\ \Lambda_n^{(j,i)} &:= \Lambda_n \cap \Lambda^{(j,i)} \end{aligned}$$

By simple computations, $P(\hat{\Theta} \in \Lambda_n^{(j,0)}) = P(\hat{\Theta} \in \Lambda_n^{(j,1)}) \ge p - \frac{1}{2}$ for the p used in (3.9). Furthermore, set

$$\begin{aligned} \overline{\Lambda}_n^{(j,1)} &:= \{\Theta_{j,1} \mid \Theta \in \Lambda_n^{(j,0)}\} \cap \Lambda_n^{(j,1)} \\ \overline{\Lambda}_n^{(j,0)} &:= \{\Theta_{j,0} \mid \Theta \in \overline{\Lambda}_n^{(j,1)}\} \end{aligned}$$

where for $\Theta = (\Theta_1, \Theta_2, \ldots, \Theta_{m_n})$ the corresponding

$$\begin{aligned} \Theta_{j,1} &:= (\Theta_1, \ldots, \Theta_{j-1}, 1, \Theta_{j+1}, \ldots, \Theta_{m_n}) \\ \Theta_{j,0} &:= (\Theta_1, \ldots, \Theta_{j-1}, 0, \Theta_{j+1}, \ldots, \Theta_{m_n}) \end{aligned}$$

are the vectors where the j-th components have been set to 1 or 0 , respectively. Similarly, the random vectors $\hat{\Theta}_{j,1}$ and $\hat{\Theta}_{j,0}$ are defined.
Clearly, $\overline{\Lambda}_n^{(j,1)} \subseteq \Lambda_n^{(j,1)}$ and $\overline{\Lambda}_n^{(j,0)} \subseteq \Lambda_n^{(j,0)}$ and

$$\begin{aligned} P(\hat{\Theta} \in \overline{\Lambda}_n^{(j,1)}) &= P(\hat{\Theta} \in \{\Theta_{j,1} \mid \Theta \in \Lambda_n^{(j,0)}\}) + P(\hat{\Theta} \in \Lambda_n^{(j,1)}) \\ &\quad - P\left(\hat{\Theta} \in \{\Theta_{j,1} \mid \Theta \in \Lambda_n^{(j,0)}\} \cup \Lambda_n^{(j,1)}\right) \\ &\ge P(\hat{\Theta} \in \Lambda_n^{(j,0)}) + P(\hat{\Theta} \in \Lambda_n^{(j,1)}) - \tfrac{1}{2} \\ &\ge 2p - \tfrac{3}{2}. \end{aligned}$$

Similarly, $P(\hat\Theta \in \overline{\Lambda}_n^{(j,0)}) \geq 2p - \frac{3}{2}$ is obtained.
The integrand on the right side of (3.10) may be lowerbounded by

$$(p-\tfrac{1}{2})\Big\{E_{\hat\Theta}\left(E_{f_{\hat\Theta}}|\hat f(x) - f_{\hat\Theta}(x)|^2 \mid \hat\Theta \in \Lambda_n^{(j,0)}\right) \\ + E_{\hat\Theta}\left(E_{f_{\hat\Theta}}|\hat f(x) - f_{\hat\Theta}(x)|^2 \mid \hat\Theta \in \Lambda_n^{(j,1)}\right)\Big\} \tag{3.11}$$

Furthermore, for $i = 0, 1$

$$E_{\hat\Theta}\left(E_{f_{\hat\Theta}}|\hat f(x) - f_{\hat\Theta}(x)|^2 \mid \hat\Theta \in \Lambda_n^{(j,i)}\right) \\ \geq \left(P(\hat\Theta \in \overline{\Lambda}_n^{(j,i)})/P(\hat\Theta \in \Lambda_n^{(j,i)})\right) E_{\hat\Theta}\left(E_{f_{\hat\Theta}}|\hat f(x) - f_{\hat\Theta}(x)|^2 \mid \hat\Theta \in \overline{\Lambda}_n^{(j,i)}\right)$$

Since $P(\hat\Theta \in \overline{\Lambda}_n^{(j,i)})/P(\hat\Theta \in \Lambda_n^{(j,i)}) \geq 2p - \frac{3}{2}$, (3.11) is not smaller than

$$(p-\tfrac{1}{2})(2p-\tfrac{3}{2})\{E_{\hat\Theta}\left(E_{f_{\hat\Theta}}|\hat f(x) - f_{\hat\Theta}(x)|^2 \mid \hat\Theta \in \overline{\Lambda}_n^{(j,1)}\right) \\ + E_{\hat\Theta}\left(E_{f_{\hat\Theta}}|\hat f(x) - f_{\hat\Theta}(x)|^2 \mid \hat\Theta \in \overline{\Lambda}_n^{(j,0)}\right)\} \\ = c\Big\{E_{\hat\Theta}\left(E_{f_{\hat\Theta}}|\hat f(x) - f_{\hat\Theta}(x)|^2 \mid \hat\Theta \in \overline{\Lambda}_n^{(j,1)}\right) \\ + E_{\hat\Theta}\left(E_{f_{\hat\Theta}}|\hat f(x) - f_{\hat\Theta}(x)|^2 \mid \hat\Theta \in \overline{\Lambda}_n^{(j,0)}\right)\Big\} \tag{3.12}$$

and in the second summand $\overline{\Lambda}_n^{(j,0)}$ may be replaced by $\overline{\Lambda}_n^{(j,1)}$ without changing the right hand side of (3.12). Hence (3.12) equals

$$c \cdot E_{\hat\Theta}\left(E_{f_{\hat\Theta_{j,1}}}|\hat f(x) - f_{\hat\Theta_{j,1}}(x)|^2 + E_{f_{\hat\Theta_{j,0}}}|\hat f(x) - f_{\hat\Theta_{j,0}}(x)|^2 \mid \hat\Theta \in \overline{\Lambda}_n^{(j,1)}\right) \\ \geq c \cdot E_{\hat\Theta}(a_{nj}^2(x) P_{\hat\Theta_{j,1}}(|\hat f(x) - f_{\hat\Theta_{j,1}}(x)| \geq a_{nj}(x))) \\ + a_{nj}^2(x) P_{\hat\Theta_{j,0}}(|\hat f(x) - f_{\hat\Theta_{j,0}}(x)| \geq a_{nj}(x))) \tag{3.13}$$

with $a_{nj}(x) = \frac{1}{2}|f_{\hat\Theta_{j,0}}(x) - f_{\hat\Theta_{j,1}}(x)| = \frac{1}{2a_n}|(\Phi * H)(m_n(t - x_{n,j}))|$.
Writing

$$R_0 := \{|\hat f(x) - f_{\hat\Theta_{j,0}}(x)| \geq a_{nj}(x)\}$$
$$R_1 := \{|\hat f(x) - f_{\hat\Theta_{j,1}}(x)| \geq a_{nj}(x)\}$$

we see that $R_0 = R_1^c =: R$ and the right hand side of (3.13) upperbounds

$$c a_{nj}^2(x)\left[1 - E_{\hat\Theta}(P_{\hat\Theta_{j,1}}(R) - P_{\hat\Theta_{j,0}}(R))\right] \tag{3.14}$$

Now, if the chi-square distance $\chi^2(h_{\hat\Theta_{j,0}}, h_{\hat\Theta_{j,1}}) \leq c/n$ where $h_{\hat\Theta} = g * f_{\hat\Theta}$, then for the expectation in (3.14)

$$|E_{\hat\Theta}(P_{\hat\Theta_{j,1}}(R) - P_{\hat\Theta_{j,0}}(R))| \\ \leq E\left(\int_R \left|\prod_{k=1}^n h_{\hat\Theta_{j,0}}(y_k) - \prod_{k=1}^n h_{\hat\Theta_{j,1}}(y_k)\right| dy_1 \cdots dy_n\right) \\ \leq 1 - e^{-c}$$

by the Bretagnolle-Huber inequality. For n sufficiently large

$$f_{\Theta_{j,0}}(x) \geq \frac{1}{2} f_0(x) \quad \forall x \in \mathbb{R}$$

and

$$f_0(x) \geq c \max_{x_{n,j} \in [a,b]} f_0(x - x_{n,j}) \quad \forall x \in \mathbb{R}$$

so that the argument of [12] may be applied to show that

$$\chi^2(f_{\hat{\Theta}_{j,0}} * g, f_{\hat{\Theta}_{j,1}} * g) \leq \frac{2}{c} m_n^{1-2\beta} \int \frac{[(\Phi * H)(m_n x) * g(x)]^2}{(f_0 * g)} dx$$

and this is smaller than or equal to c/n for $m_n = c \cdot n^{\frac{1}{2(\beta+\eta)+1}}$ by the local result of [9]. Hence (3.14) is not smaller than $c \cdot a_{nj}^2(x)$ and since j was arbitrary, we arrive at

$$E_{\hat{\Theta}} \left(E_{f_{\hat{\Theta}}} |\hat{f}(x) - f_{\hat{\Theta}}(x)|^2 \mid \hat{\Theta} \in \Lambda_n \right) \geq c \max_{1 \leq j \leq m_n} a_{nj}^2(x)$$

so that, finally,

$$\begin{aligned}
&\sup_{f \in \mathcal{F}'_n} E_f \int_a^b |\hat{f}(x) - f(x)|^2 dx \geq c \int_a^b \max_{1 \leq j \leq m_n} a_{nj}^2(x) dx \\
&\geq c \sum_{j=0}^{m_n - 1} \int_{x_{n,j}}^{x_{n,j+1}} |a_n^{-1} (\Phi * H)(m_n(x - x_{n,j}))|^2 dx \\
&= c \int_0^{b-a} |a_n^{-1} (\Phi * H)(y)|^2 dy \\
&= c\, a_n^{-2} \quad \text{by (3.5)} \\
&= c\, m_n^{1-2\beta} \quad \text{since } a_n = m_n^{\beta - 1/2} \\
&= c\, n^{-\frac{2\beta-1}{2(\beta+\eta)}}.
\end{aligned}$$

■

4 Error densities not in $L_2(\mathbb{R})$

In this section, we study the deconvolution problem for error densities taken from

$$\mathcal{F}_{\epsilon,1,\eta} = \{\text{densities } g \mid |\Psi_g(t)| \geq d|t|^{-\eta}, \forall t \text{ with } |t| \geq T \text{ and } |\Psi_g(t)| \neq 0, \forall t\}$$

for some positive constants d, T and $\eta \leq \frac{1}{2}$. For this case compared to the analysis in Sections 2 and 3 some subtle theoretical issues arise which we now address. First of all, to be precise, the Fourier transform Ψ_f of some function f is to be understood as the Fourier-Plancherel transform if f is in $L_2(\mathbb{R})$, and as the characteristic function if f is in $L_1(\mathbb{R})$. For $f \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ both interpretations coincide. Secondly, the convolution identity $\Psi_{g*f} = \Psi_g \Psi_f$ which was crucial in Sections 2 and 3 is valid only if both f and g are in $L_1(\mathbb{R})$, if $f \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ and $g \in L_2(\mathbb{R})$, and if $g \in L_1(\mathbb{R}) \cap L_2(\mathbb{R})$ and $f \in L_2(\mathbb{R})$. It is, however , not valid for $g \in L_1(\mathbb{R})$ and $f \in L_2(\mathbb{R})$. In the latter case, convolution of the two functions is not even well-defined. This has several implications. One of them is that the operator C_g can no longer be viewed as an $L_2(\mathbb{R}) \to L_2(\mathbb{R})$-operator. Instead, we consider it as an operator with domain $\mathcal{L} := L_1(\mathbb{R}) \cap L_2(\mathbb{R})$. But $\mathcal{L}$ equipped with the $L_2(\mathbb{R})$-norm though a normed linear space is not complete. A further problem arises from the fact that the operator P_ω for which $\Psi_{P_\omega f} = \Psi_f 1_{[-\omega,+\omega]}$ is not an $\mathcal{L} \to \mathcal{L}$-operator:
For $f \in \mathcal{L}$, $P_\omega f$ is in $L_2(\mathbb{R})$ but not necessarily also in $L_1(\mathbb{R})$. This is so since for $f \in L_1(\mathbb{R})$ the Fourier transform Ψ_f is continuous or at least is equal to a continuous function outside of a (Lebesgue-)null set. Now, if for the continuous modification $\tilde{\Psi}_f$ of Ψ_f , $\tilde{\Psi}_f(\omega) \neq 0$ or $\tilde{\Psi}_f(-\omega) \neq$

0 then $\Psi_{P_\omega f}$ is discontinuous at ω or at $-\omega$ and this discontinuity cannot be eliminated by modification of the function on a null set. In conclusion, application of the operator P_ω may lead to functions not in $\mathcal{L}$ and in the sequel the convolution formula may no longer be applied. Therefore, an alternative operator needs to be introduced with the purpose to restrict frequencies to a compact set similar to what is accomplished by P_ω. Towards this end, set

$$g_\omega(x) = \frac{2}{\omega\pi x^2}\left(\cos(\frac{\omega}{2}x) - \cos(\omega x)\right) \tag{4.1}$$

which is in $\mathcal{L}$ and whose characteristic function is given by

$$\Psi_{g_\omega}(t) = \frac{2(t+\omega)}{\omega}1_{(-\omega,-\omega/2)}(t) + 1_{[-\omega/2,+\omega/2]}(t) + \frac{2(\omega-t)}{\omega}1_{(\omega/2,\omega)}(t)$$

We utilize g_ω to replace the operator P_ω by the convolution operator P_ω^* defined by

$$P_\omega^* f \;=\; g_\omega * f \quad , \quad f \in \mathcal{L}.$$

Convolution of two $L_1(\mathbb{R})$-functions produces an $L_1(\mathbb{R})$-function and convolution of two $L_2(\mathbb{R})$-functions leads to an $L_2(\mathbb{R})$-function. Hence, P_ω^* is an $\mathcal{L} \to \mathcal{L}$-operator and for $f \in \mathcal{L}$

$$\Psi_{P_\omega^* f} = \Psi_{g_\omega * f} = \Psi_{g_\omega}\Psi_f$$

since both f and g_ω are in $\mathcal{L}$. We also have an analogy to (2.6), namely

$$\|P_\omega^* f - f\|^2_{L_2(\mathbb{R})} \;\leq\; \frac{a}{\pi 2^{1-2\beta}}\omega^{1-2\beta} \text{ for } f \in \mathcal{F}^{\beta,a} \subset \mathcal{L}.$$

Now, let us consider the convolution operator C_γ for densities γ not in $L_2(\mathbb{R})$. C_γ is an $\mathcal{L} \to \mathcal{L}$-operator: for $f \in \mathcal{L}$, $C_\gamma f \in L_1(\mathbb{R})$ is clear and $C_\gamma f \in L_2(\mathbb{R})$ follows from

$$\begin{aligned} |(C_\gamma f)(x)|^2 &= \left|\int_{\mathbb{R}} \gamma(y) f(x-y)dy\right|^2 \\ &= |E[f(x-Y)]|^2 && \text{, if the random variable } Y \text{ has density } \gamma. \\ &\leq E[|f(x-Y)|^2] \\ &= \int_{\mathbb{R}} |f(x-y)|^2\gamma(y)dy \end{aligned}$$

which is integrable. In fact,

$$\|f\|^2_{L_2(\mathbb{R})} \;=\; \int_{\mathbb{R}}\int_{\mathbb{R}} |f(x)|^2\gamma(y)dxdy \;=\; \int_{\mathbb{R}}\left[\int_{\mathbb{R}} |f(x-y)|^2\gamma(y)dy\right]dx.$$

Unlike before, in the present context, the convolution operator C_{g^-} cannot be considered the Hilbert-adjoint operator of C_g, since $\mathcal{L}$ equipped with the inner product norm of $L_2(\mathbb{R})$ fails to be complete and, hence, is not a Hilbert space.
In analogy to (2.13), we set

$$\hat{f}_*^{(k)} \;=\; \sum_{j=0}^{k}(I - C_{g^-}C_g)^j C_{g^-}P_\omega^*\hat{h} \quad , \quad k \in \mathbb{N}_0 \tag{4.2}$$

for some ordinary density estimator $\hat{h} \in \mathcal{L}$ of the density h based on direct observations. In analogy to Proposition 2, for $f \in \mathcal{L} \cap L_2^\omega(\mathbb{R})$ we have

$$\|(I - C_{g^-}C_g)f\|_{L_2(\mathbb{R})} \;\leq\; (1 - d^2\omega^{-2\eta})\|f\|_{L_2(\mathbb{R})} \tag{4.3}$$

from which we conclude that

$$\|(I - C_{g^-} C_g)^j C_{g^-} P_\omega^* \hat{h}\|_{L_2(\mathbb{R})} \leq (1 - d^2 \omega^{-2\eta})^j \|C_{g^-} P_\omega^* \hat{h}\|_{L_2(\mathbb{R})}$$

and so the sequence $(\hat{f}_*^{(k)})_{k \in \mathbb{N}_0}$ converges in $L_2(\mathbb{R})$-norm. Since $L_2^\omega(\mathbb{R})$ equipped with this norm is a Hilbert space, the $L_2(\mathbb{R})$-limit of the sequence in (4.2) is in $L_2^\omega(\mathbb{R})$. We denote this limit by $\hat{f}_\omega^*$. It can be represented as

$$\hat{f}_\omega^* := \sum_{j=0}^{\infty} (I - C_{g^-} C_g)^j C_{g^-} P_\omega^* \hat{h} \tag{4.4}$$

We mention that the sequence in (4.2) may not converge in $L_1(\mathbb{R})$-norm and $\hat{f}_\omega^*$ in (4.4) may not be an $L_1(\mathbb{R})$-function.
It remains to specify an appropriate estimator $\hat{h} \in \mathcal{L}$ for the sequence in (4.2). In the context considered in this Section, the sync-kernel estimator is not available as previously, since it fails to lie in $\mathcal{L}$. Instead, we take

$$\hat{h}_\omega^*(y) := \frac{1}{n} \sum_{j=1}^{n} g_\omega(y - Y_j) \tag{4.5}$$

with g_ω from (4.1). Then we can state

Theorem 3 *For error densities g from $\mathcal{F}_{\epsilon,1,\eta}$ and with $\omega = \omega_n = c\, n^{\frac{1}{2(\beta+\eta)}}$ we have*

$$\sup_{f \in \mathcal{F}^{\beta,a}} E_f \, \|\hat{f}_{\omega_n}^* - f\|_{L_2(\mathbb{R})}^2 \leq c\, n^{-\frac{2\beta-1}{2(\beta+\eta)}} \tag{4.6}$$

with $\hat{f}_{\omega_n}^*$ as in (4.4) using $\hat{h} = \hat{h}_{\omega_n}^*$ from (4.5).

The proof of Theorem 3 follows the line of argument of the proof of Theorem 1. But one needs to be mindful of the technical aspects outlined earlier in this Section and so it turns out to be more technically intricate. Also, the constant on the right hand side of (4.6) is different from the corresponding constant in Theorem 1.

5 Supersmooth error densities

In this Section we allow for error densities from the smoothness class

$$\begin{aligned} \mathcal{G}_{\epsilon,\eta,\xi} := \{\text{densities } g \mid g \in L_2(\mathbb{R}), |\Psi_g(t)| \geq b|t|^\eta \exp(-\delta|t|^\xi) \forall t \text{ with} \\ |t| \geq T \text{ for some } T > 0 \text{ and } |\Psi_g(t)| \neq 0\, \forall t\} \end{aligned}$$

Densities with an exponentially decreasing characteristic function are commonly referred to as supersmooth densities. In addition to smooth densities the class $\mathcal{G}_{\epsilon,\eta,\xi}$ does contain supersmooth densities. In analogy to the results in Section 3:

Proposition 5 *For the root C of the operator $C_g^t C_g$ with $g \in \mathcal{G}_{\epsilon,\eta,\xi}$ it is*

$$\|I - C\|_{L_2^\omega(\mathbb{R}),ind.} \leq 1 - b\omega^\eta \exp(-\delta\omega^\xi)$$

Proposition 6 *For $\overline{f}^{(k)}$ as defined in (2.25) with $g \in \mathcal{G}_{\epsilon,\eta,\xi}$ and $P_\omega \hat{f}$ as in (2.12) we have*

$$\|\overline{f}^{(k)} - P_\omega \hat{f}\|_{L_2(\mathbb{R})} \leq (1 - b\omega^\eta \exp(-\delta\omega^\xi))^{k+1} \|P_\omega \hat{f}\|_{L_2(\mathbb{R})}$$

Theorem 4 *For error densities g from $\mathcal{G}_{\epsilon,\eta,\xi}$ and with $\omega = \omega_n = \left[\left(\frac{1}{4\delta}\right)\ln n\right]^{1/\xi}$ we have*

$$\sup_{f\in\mathcal{F}^\beta} E_f\|\hat{f}_{\omega_n} - f\|^2_{L_2(\mathbb{R})} \le c(\ln n)^{\frac{1-2\beta}{\xi}}$$

with $\hat{f}_{\omega_n}$ as in (3.3).

The proofs are similar in each case to those of the corresponding previous results. Note that the optimal sequence ω_n in Theorem 4 can be determined exactly, not merely up to a constant. Similarly to Theorem 2 we can also obtain a lower bound result for error densities from

$$\mathcal{G}'_{\epsilon,\eta,\tilde{\eta},\xi} := \{g \in \mathcal{G}_{\epsilon,\eta,\xi} \mid |\Psi_g(t)| \le \tilde{b}|t|^{\tilde{\eta}}\exp(-\delta|t|^\xi)$$
$$\text{and } P(|\epsilon - u| \le |u|^{\alpha_0} = O(|u|^{-(\alpha-\alpha_0)}) \text{ as } |u| \to \infty\}.$$

Theorem 5 *For error densities g from $\mathcal{G}'_{\epsilon,\eta,\tilde{\eta},\xi}$ we have*

$$\inf_{\hat{T}\in\mathcal{T}_n}\sup_{f\in\mathcal{F}^\beta} E_f\|\hat{T} - f\|^2_{L_2(\mathbb{R})} \ge c(\ln n)^{\frac{1-2\beta}{\xi}}$$

For the definition of $\mathcal{T}_n$, see Theorem 2. Again, the proof follows the line of argument of the corresponding previous result.

6 Simulations

Now, we illustrate the iterative deconvolution estimator with some numerical experiments. We take the standard normal density ($N(0,1)$) and the standard Cauchy density as examples of the density f which is being estimated. For the error density g we also take two paradigmatic cases, one smooth and one supersmooth density: the double exponential density and $N(0,1)$. Hence, the summary of our simulations consists of four simulated cases:

Figure	f	g
1	$N(0,1)$	double exponential
2	Cauchy	double exponential
3	$N(0,1)$	$N(0,1)$
4	Cauchy	$N(0,1)$

Selecting the scaling parameter $\lambda_n = n^{1/8}$ in figure (1) and (2) according to Theorem 1 and $\lambda_n = \sqrt{\ln(n)/2}$ in figure (3) and (4) according to Theorem 4, we compute the iterative estimators based on $n = 1000$ independent contaminated observations in each case. This is a relatively small sample size for deconvolution tasks, especially if the error density is supersmooth. The estimators are calculated by the iterative scheme (2.4). We have plotted f (the dashed line) as well as the iterations $\hat{f}^{(0)}$, $\hat{f}^{(1)}$, $\hat{f}^{(2)}$, $\hat{f}^{(5)}$, $\hat{f}^{(10)}$, constituting successively improving (in each of the four figures) approximations to f.

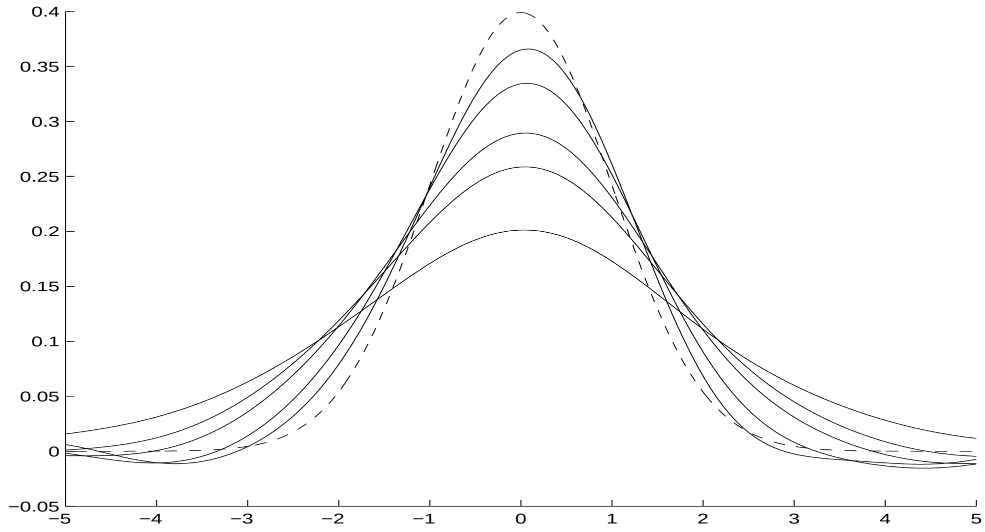

Figure 1: $X \sim N(0,1)$, $\epsilon \sim$ double exponential

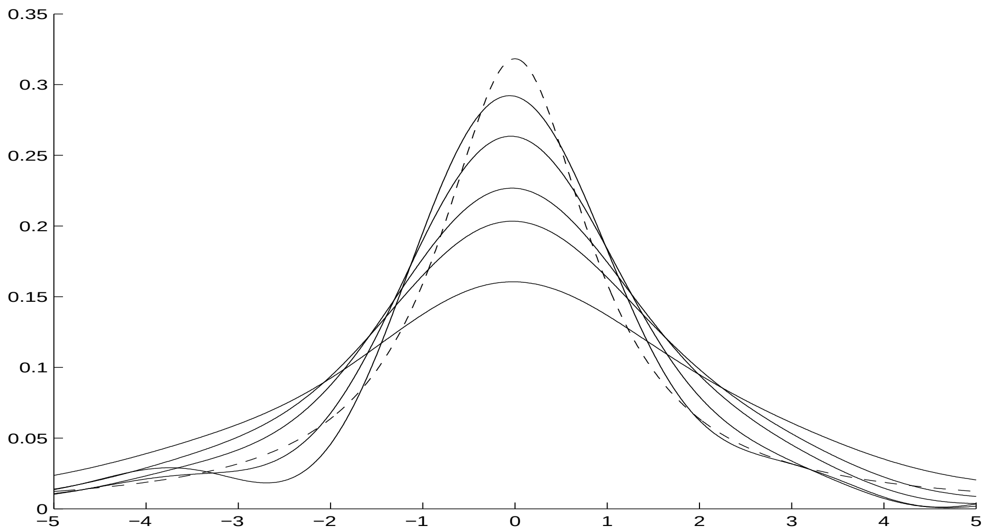

Figure 2: $X \sim$ Cauchy, $\epsilon \sim$ double exponential

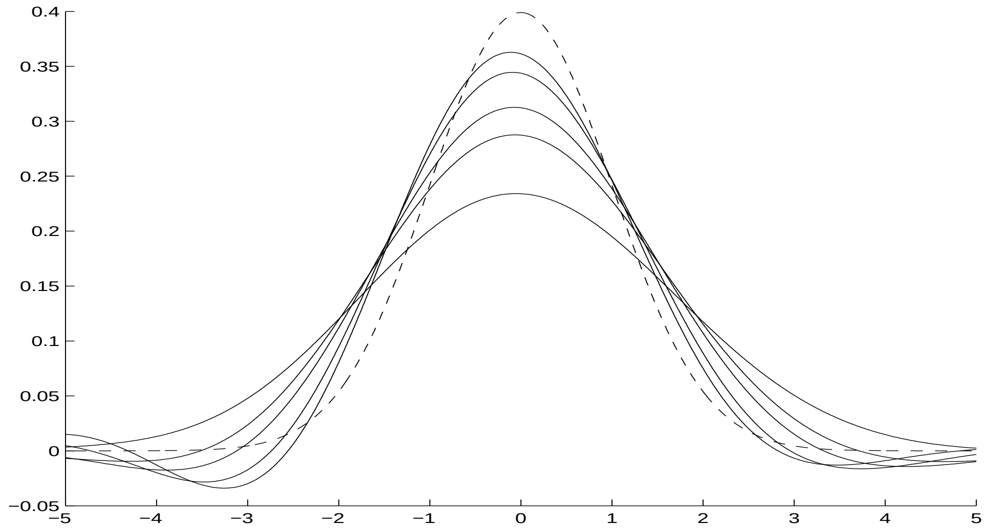

Figure 3: $X \sim N(0,1)$, $\epsilon \sim N(0,1)$

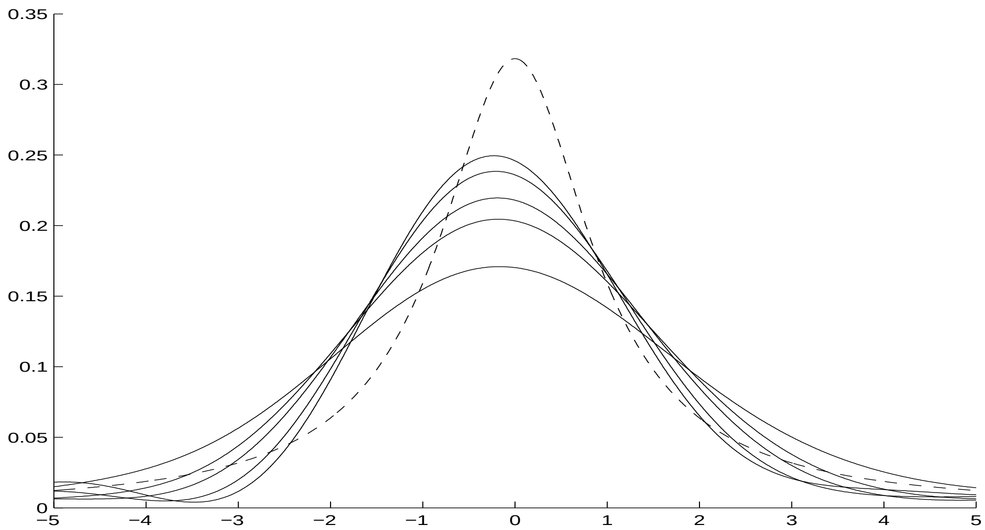

Figure 4: $X \sim$ Cauchy, $\epsilon \sim N(0,1)$

References

[1] Bretagnolle, J. and Huber, C., *Estimation des densités, risque minimax*, Z. Wahrsch. Verw. Gebiete 47, (1979), 119-137

[2] Carroll,R.J. and Hall,P., *Optimal rates of convergence for deconvolving a density*, J. Amer. Statist. Assoc. 83, (1988), 1184-1186.

[3] Carroll, R.J., Ruppert, D. and Stefanski, L.A., *Measurement Error in Nonlinear Models*, Chapman and Hall, London, (1995).

[4] Cavalier, L. and Tsybakov, A.B., *Sharp adaption for inverse problems with random noise*, Probab. Theory Relat. Fields 123, (2002), 323-354.

[5] DAVIS,K.B., *Mean integrated square error properties of density estimates*, Ann. Statist., 5 , (1977), 530-535.

[6] EDDINGTON, A.S., *On a formula for correcting statistics for the effects of a known probable error of observation*, Mon. Not. R. Astron. Soc., 73, (1913), 359-360.

[7] EFROMOVICH,S., *Robust and efficient recovery of a signal passed through a filter and then contaminated by non-Gaussian noise*, IEEE Trans. Inform. Theory, 43, (1997a), 1184-1191.

[8] EFROMOVICH,S., *Density estimation for the case of supersmooth measurement error*, J. Amer. Statist. Assoc., 92, (1997b), 526-535.

[9] FAN,J., *On the optimal rates of convergence for non-parametric deconvolution problems*, Ann. Statist., 19, (1991a), 1257-1272.

[10] FAN, J., *Global behaviour of deconvolution kernel estimates*, Statist. Sinica, 1, (1991b), 541-551.

[11] FAN,J., *Deconvolution with supersmooth distributions*, Canad. J. Statist., 20, (1992), 155-169.

[12] FAN,J., *Adaptively local one-dimensional subproblems with application to a deconvolution problem*, Ann. Statist., 21, (1993), 600-610.

[13] FOTOPOULOS,S.B., *Invariance principles for deconvoluting kernel density estimation under stationary sequences of random variables*, Journal of Statistical Planning and Inference, 86, (2000), 31-50.

[14] GAFFY,W.R., *A consistent estimator of a component of a convolution*, Ann. Math. Statist., 30, (1959), 198-205.

[15] GOLDENSHLUGER, A., *On pointwise adaptive nonparametric deconvolution*, Bernoulli, 5, (1998), 907-926.

[16] HESSE,C.H., *Deconvolving a density from partially contaminated observations*, J. Multivariate Anal., 55, (1995a), 246-260.

[17] HESSE,C.H., *How many good oberservations do you need for fast density deconvolution from supersmooth errors*, Sankhya A, 58, (1995b), 491-506.

[18] HESSE,C.H., *Data-driven deconvolution*, Nonparametric Statistics, 10, (1999), 343-373.

[19] JOHNSTONE, I.M., *Wavelet shrinkage for correlated data and inverse problems: adaptive results*, Statistica Sinica, 9, (1999) 51-83.

[20] KOO,J.-K., *Longspline deconvolution in Besov space*, Scand. J. Statist. 26, (1999), 73-86.

[21] LIU,M.C. AND TAYLOR,R.C., *A consistent nonparametric density estimator for the deconvolution problem*, Canad. J. Statist., 17, (1989), 427-438.

[22] MASRY,E. AND RICE,J.A., *Gaussian deconvolution via differentiation*, Canad. J. Statist., 20, (1992), 9-21.

[23] NEUMANN,M.H., *On the effect of estimating the error density in nonparametric deconvolution*, Nonparametric Statistics, 7, (1997), 307-330.

[24] PENSKY,M. AND VIDAKOVIC,B., *Adaptive wavelet estimator for nonparametric density deconvolution*, Ann. Statist., 27, (1999), 2033-2053.

[25] Rachdi,M. and Sabre,R., *Consistent estimates of the mode of the probability density function in nonparametric deconvolution problems*, Statist. and Probab. Lett., 47, (2000), 105-114.

[26] Reed,M. and Simon,B., *Methods of Modern Mathematical Physics, I Functional Analysis*, Academic Press, (1980)

[27] Robbins,H., *The empirical Bayes approach to statistical decision problems*, Ann. Math. Statist., 35, (1964), 1-20

[28] Snyder,D.L., Miller,M.I. and Schultz,T.J., *Constrained probability density estimation from noisy data*, Proceedings of the 1988 Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore, Maryland, (1988), 170-172.

[29] Stefanski, L.A., *Rates of convergence of some estimators in a class of deconvolution problems*, Statist. Prob. Lett., 9, (1990), 229-235

[30] Stefanski,L.A. and Carroll,R.J., *Deconvolving kernel density estimators*, Statistics, 21, (1990), 169-184.

[31] Tsybakov, A.B., *Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes*, Ann. Statist. 26, (1998), 2420-2469.

[32] Walter,G.G. and Shen,X., *Deconvolution using Meyer wavelets*, J. Integral Eq. Appl., 11, (1999), 515-534.

[33] Wand, M.P. and Jones, M.C., *Kernel Smoothing*, Chapman and Hall, London, (1995).

[34] Watson,G.S. and Leadbetter,M.R., *On the estimation of the probability density*, I. Ann. Math. Statist., 34, (1963), 480-491.

Christian H. Hesse
Pfaffenwaldring 57
70569 Stuttgart
Germany
E-Mail: `hesse@mathematik.uni-stuttgart.de`

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

Erschienene Preprints ab Nummer 2004/001

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

2004/001 *Walk, H.*: Strong Laws of Large Numbers by Elementary Tauberian Arguments.

2004/002 *Hesse, C.H., Meister, A.*: Optimal Iterative Density Deconvolution: Upper and Lower Bounds.

2004/003 *Meister, A.*: On the effect of misspecifying the error density in a deconvolution problem.

2004/004 *Meister, A.*: Deconvolution Density Estimation with a Testing Procedure for the Error Distribution.

2004/005 *Efendiev, M.A., Wendland, W.L.*: On the degree of quasiruled Fredholm maps and nonlinear Riemann-Hilbert problems.

2004/006 *Dippon, J., Walk, H.*: An elementary analytical proof of Blackwell's renewal theorem.