Multiscale Methods for Shape Constraints in Deconvolution

L. Dümbgen[*], A. Munk[†] and J. Schmidt-Hieber[‡§]

Abstract

We derive multiscale statistics for deconvolution in order to detect qualitative features of the unknown density. An important example covered within this framework is to test for local monotonicity on all scales simultaneously. The errors in the deconvolution model are restricted to a certain class of distributions that include Laplace, Gamma and Exponential random variables. Our approach relies on inversion formulas for deconvolution operators. For multiscale testing, we consider a calibration, motivated by the modulus of continuity of Brownian motion. We investigate the performance of our results from both the theoretical and simulation based point of view. A major consequence of our work is that the detection of qualitative features of a density in a deconvolution problem is a doable task although the minimax rates for pointwise estimation are very slow.

AMS 2010 Subject Classification: Primary 62G10; secondary 62G15, 62G20.

Keywords: Brownian motion; convexity; differential inequalities; ill-posed problems; mode detection; monotonicity; multiscale statistics; shape constraints.

[*]Institut für mathematische Stochastik und Versicherungslehre, Universität Bern, Sidlerstrasse 5, 3012 Bern, Switzerland.

[†]Institut für Mathematische Stochastik, Universität Göttingen, Goldschmidtstr. 7, 37077 Göttingen and Max-Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany.

[‡]Institut für Mathematische Stochastik, Universität Göttingen, Goldschmidtstr. 7, 37077 Göttingen, Germany and Department of Mathematics, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, Netherlands.

[§]For correspondence `schmidth@math.uni-goettingen.de`

1 Introduction and Notation

Assume that we observe $Y = (Y_1, \dots, Y_n)$ according to the deconvolution model

$$Y_i = X_i + \epsilon_i, \quad i = 1, \dots, n, \tag{1.1}$$

where $X_i, \epsilon_i,\ i = 1, \dots, n$ are assumed to be real valued and independent, $X_i \overset{i.i.d.}{\sim} X, \epsilon_i \overset{i.i.d.}{\sim} \epsilon$ and Y_1, X, ϵ have densities g, f and f_ϵ, respectively. Our goal is to develop multiscale test statistics for certain structural assumptions on f, where the density f_ϵ of the blurring distribution is assumed to be known.

Structural assumptions or shape constraints are conveniently expressed in this paper as linear differential inequalities of the density f in the *time domain,* assuming for the moment that f is sufficiently smooth. Important examples are $f' \gtrless 0$ to check local monotonicity properties as well as $f'' \gtrless 0$ for local convexity or concavity. To give another example, suppose that we are interested in local monotonicity properties of the density $\tilde{f}$ of $\exp(aX)$ for a given $a > 0$. Since $\tilde{f}(s) = (as)^{-1} f(a^{-1} \log(s))$, one can easily verify that local monotonicity properties of $\tilde{f}$ may be expressed in terms of the inequalities $f' - af \lesseqgtr 0$.

In general, we consider a differential operator A and want to identify intervals on which $Af \not\leq 0$ or $Af \not\geq 0$. If applied to $A = D$ or D^2 with the derivative operator $Df := f'$, our method yields bounds for the number and confidence regions for the location of modes and inflection points of f. Indeed, our work is an extension of Dümbgen and Walther [11] who treated the case $A = D$ in the direct case, i.e. when $\epsilon_i = 0$. It is not easy, however, to transfer the methods of [11] to the deconvolution setting. To this end it would much more convenient to express hypotheses on the local shape in the frequency domain.

Hypothesis testing for deconvolution and related inverse problems is a relatively new area. Current methods cover testing of parametric assumptions (cf. [3, 26, 5]) and, more recently, testing adaptively certain smoothness classes such as Sobolev balls in a Gaussian sequence model (Laurent *et al.* [25, 26] and Ingster *et al.* [20]). All these papers focused on regression deconvolution models. Exceptions for density deconvolution are Holzmann *et al.* [19] and Meister [27] who developed a test of (global) monotonicity based on classical Fourier inversion (see e.g. Carroll and Hall [6]). This test has been derived for one fixed interval, which allows to check whether a density is monotone on that interval at a preassigned level of significance.

In this paper we introduce a statistic for investigating shape constraints of the unknown density f on all scales simultaneously. As mentioned above, at a first glance, this appears to be a quite difficult task because qualitative hypotheses such as local monotonicity cannot be immediately expressed in terms of the Fourier coefficients. Let us illustrate the basic idea for the case f', i.e. $A = D$ and D is the differentiation operator. Define $\phi_{t,h}(\cdot) = \phi((\cdot - t)/h)$ with a sufficiently smooth and positive kernel ϕ supported on $[0, 1]$. The functions $\phi_{t,h}$ serve as local test functions for local monotonicity in the following sense: Whenever we know that $\langle \phi_{t,h}, f' \rangle > 0$, we may conclude that $f(s_1) < f(s_2)$ for some points $s_1 < s_2$ in $[t, t + h]$. Here and throughout the sequel we write $\langle h_1, h_2 \rangle := \int_{\mathbb{R}} h_1(x) h_2(x)\, dx$.

For simplicity we only consider the case where the so called inversion operators, i.e. the multiplicative inverse of the Laplace transform are polynomials, which leads us to the following assumption on the noise that also appears in [4]. Throughout this work let $\mathcal{F}(f) = \int_{\mathbb{R}} \exp(-ix\cdot) f(x) dx$ denote the Fourier transform of $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$.

Assumption 1. *We assume that the characteristic function of ϵ has the representation*

$$\psi_\epsilon(t) := \left(\mathbb{E}\, e^{it\epsilon}\right)^{-1} = \left(\mathcal{F}(f_\epsilon)(-t)\right)^{-1} = \sum_{j=0}^{r} q_j\, (it)^j$$

for some non-negative integer r and real coefficients $q_0 = 1, q_1, \ldots, q_r$.

Then, we obtain by partial integration and Plancherel's identity

$$\begin{aligned} -\langle \phi_{t,h}, f' \rangle = \langle D\phi_{t,h}, f \rangle &= \frac{1}{2\pi} \big\langle \mathcal{F}(D\phi_{t,h}), \mathcal{F}(f) \big\rangle = \frac{1}{2\pi} \big\langle \mathcal{F}(D\phi_{t,h}), \psi_\epsilon\, \mathcal{F}(g) \big\rangle \\ &= \frac{1}{2\pi} \big\langle \mathcal{F}(\psi_\epsilon(D) D\phi_{t,h}), \mathcal{F}(g) \big\rangle = \big\langle \psi_\epsilon(D) D\phi_{t,h}, g \big\rangle. \end{aligned} \tag{1.2}$$

Thus, the l.h.s. can be estimated directly via $Y_1, \ldots, Y_n$ and we find that it is possible to infer local properties of f' similar as in the case without convolution by

$$T_{t,h} = -\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \psi_\epsilon(D) D\phi_{t,h}(Y_k).$$

Since $\mathbb{E}\, T_{t,h} = \sqrt{n} \langle \phi_{t,h}, f' \rangle$, this gives rise to a multiscale statistic

$$T_n = \sup_{(t,h)} w_h \left(\frac{h^{r+1/2} \big| T_{t,h} - \mathbb{E}\, T_{t,h} \big|}{\sqrt{\widehat{g}_n(t)}} - \widetilde{w}_h \right),$$

where w_h and $\widetilde{w}_h$ are chosen in order to calibrate the different scales with equal weight, while $\widehat{g}_n$ is an estimator of g. Note that the additional factor $h^{r+1/2}$ is due to the ill-posedness of the problem (cf. Assumption 1) as well as to differentiation of f.

In this paper we will derive the limit distribution of T_n in order to determine the critical values, which turns out to be distribution free. Our multiscale calibration requires new techniques for proving convergence to a limit distribution. Furthermore, we will show how to extend methods introduced by Giné *et al.* [15, 16] for construction of confidence bands in order to prove multiscale results. This allows us on the one hand to extend the approach of [11], resulting for example in simultaneous confidence statements for the existence and location of regions of increase and decrease. On the other hand, our approach is statistically more informative than pure testing. In fact, for given shape constraint, we construct objects, which appear to be similar to confidence bands. For a more precise statement see Section 3.

Is is a well-known fact (cf. Delaigle and Gijbels [9]) that selection of an appropriate bandwidth is a delicate issue in deconvolution models. One of the main advantages of multiscale methods is that essentially no smoothing parameter is required. The main choice will be the quantile of the multiscale statistic, which has a clear probabilistic interpretation.

As illustrated above, our approach is based on inversion of differential operators to resolve the discrepancy between hypotheses, formulated in the time domain, and testing methods in the spectral domain, i.e. it nicely combines shape constraints given by differential inequalities and deconvolution. To give another example, consider the case where ϵ is exponential with density e^{-x} for $x \geq 0$. In this case we may recover f by $f = g + g'$ (cf. Jongbloed [21], and for more examples van Es and Kok [29]). The key advantages of this inversion method is the following locality property: $f(x)$ can be expressed as a linear combination of derivatives of g at x. Therefore, testing for a shape constraint on the interval I only requires observations falling into I.

Let us finally address Assumption 1. If X is gamma distributed, let us call $-X$ negative gamma distributed. The class of distributions satisfying Assumption 1 can be shown to be the class of finite mixtures of gamma and negative gamma distributed random variables with shape parameters ≥ 1. In particular, exponential, Laplace and gamma distributed r.v.s belong to this class. Moreover, the density f_ϵ is necessarily bounded if $r \geq 1$. The special case $\epsilon = 0$ (i.e. no deconvolution or direct problem) can be treated as well, of course.

For practical applications, we may use these models whenever the error variable ϵ is an independent waiting time. For example let X_i be the (unknown) time of infection of the i-th patient, ϵ_i the corresponding incubation time, and Y_i is the time when diagnosis is made. Then, it is convenient to assume $\epsilon \sim \Gamma(r, \theta)$ (see for instance [8], Section 3.5). By the techniques developed in this paper one will be able to identify for example time intervals where the number of infections increased and decreased for a specified confidence level.

Another application is single photon emission computed tomography (SPECT), where the detected scattered photons are blurred by Laplace distributed random variables (cf. Floyd *et al.* [14], Kacperski *et al.* [22]).

The paper is organized as follows. The general multiscale statistic and the main theorem are established in Section 2. This part applies to density estimation in general. Section 3 is devoted to examples as well as the construction and discussion of the multiscale method for shape constraints, using the theoretical results obtained in Section 2. Theoretical questions related to the performance of the multiscale method and numerical aspects are discussed in Sections 4 and 5. In particular, we are able to identify the asymptotically optimal kernel function as a Legendre polynomial. Proofs and further technicalities are shifted to the appendix.

2 A general multiscale test statistic

In this section, we shall give a fairly general convergence result. The presented result does not use the deconvolution structure of model (1.1). It only requires that we have observations $Y_i = G^{-1}(U_i),\ i = 1, \ldots, n$ with U_i i.i.d. uniform on $[0,1]$ and G an unknown distribution function with Lebesgue density g.

Let us summarize some notation, used throughout this work. $\langle \cdot, \cdot \rangle$ denotes the L^2-inner product. By slight abuse of notation, we write $\|\cdot\|_2,\ \|\cdot\|_\infty$, for the norms on $L^2([0,1])$ and $L^\infty([0,1])$, respectively. Suppose that $\mathrm{TV}(\cdot)$ denotes the total variation of functions on $\mathbb{R}$. Let us introduce the function classes

$$\begin{aligned}\mathcal{TV}^{(m)} := \big\{\phi \ \big| \ \operatorname{supp}\phi \subset [0,1],\ \phi^{(l)} \text{ is continuous for } 0 \le l < m, \\ \phi^{(m)} \text{ is càdlàg with } \ \mathrm{TV}(\phi^{(m)}) < \infty\big\},\end{aligned}$$

and for fixed $c, C \geq 0$,

$$\begin{aligned}\mathcal{G} := \mathcal{G}_{c,C} := \big\{G \ \big| \ G \text{ is a distribution function with density } g, \\ g\big|_{[0,1]} \geq c, \ \text{and } |g(x) - g(y)| \leq C|x-y|, \text{ for all } x, y \in [0,1] \ \big\}. \qquad (2.1)\end{aligned}$$

Concerning the definition of $\mathcal{TV}^{(m)}$, in case of $m = 0$, we simply assume that $\phi^{(0)} := \phi$ is càdlàg with finite total variation. In case of $m > 0$, $\phi^{(l)}$ with $1 \le l < m$ stands for the usual derivative, while $\phi^{(m-1)}$ is assumed to be absolutely continuous with L^1-derivative $\phi^{(m)}$.

Suppose that for $m \geq 0$,

$$L\phi(x) = \langle \phi, \alpha_{-1} \rangle + \sum_{l=0}^{m} \alpha_l(x) \phi^{(l)}(x) \qquad (2.2)$$

is a differential operator on $\mathcal{TV}^{(m)}$. Throughout this section, let m denote the order of L. We fix a function $\phi \in \mathcal{TV}^{(m)}$ and write

$$\phi_{t,h}(\cdot) = \phi\left(\frac{\cdot - t}{h}\right).$$

Now, consider the test statistic

$$T_{t,h} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} L\phi_{t,h}(Y_k) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} L\phi_{t,h}(G^{-1}(U_k)) \tag{2.3}$$

and note that

$$\mathbb{E}\, T_{t,h} = \sqrt{n} \int \big(L\phi_{t,h}\big)(s) g(s) ds = \sqrt{n} \int \phi_{t,h}(s) \big(L^\star g\big)(s) ds,$$

where $L^\star$ denotes the adjoint operator. One may think of $t \mapsto T_{t,h}$ as a kernel estimator of $L^\star g$ with bandwidth h. We combine the single test statistics for an arbitrary subset

$$B_n \subset \big\{(t,h) \big|\ t \in [0,1],\ h \in [l_n, u_n],\ t+h \leq 1\big\} \tag{2.4}$$

and consider for $\nu > e$ and

$$w_h = \frac{\sqrt{\frac{1}{2}\log\frac{\nu}{h}}}{\log\log\frac{\nu}{h}}, \tag{2.5}$$

the multiscale statistic

$$T_n := \sup_{(t,h)\in B_n} w_h \left(\frac{h^{m-1/2}\big|T_{t,h} - \mathbb{E}[T_{t,h}]\big|}{\sqrt{\widehat{g}_n(t)}\ \alpha_m(t)} - \sqrt{2\log\frac{\nu}{h}} \right), \tag{2.6}$$

where $\widehat{g}_n$ is an estimator of g, satisfying

$$\sup_{G\in\mathcal{G}} \|\widehat{g}_n - g\|_\infty = O_P(1/\log n). \tag{2.7}$$

Theorem 1. *Given a differential operator L of the form (2.2). Work under Assumption 1, where m is the order of L. Assume that $\phi \in \mathcal{TV}^{(m)}$ is normalized, such that $\|\phi^{(m)}\|_{L^2} = 1$. Further suppose that $\mathrm{TV}(\alpha_q) + \|\alpha_q\|_\infty < \infty$ for $q = 0, \ldots, m,$ α_m is Lipschitz and bounded away from zero, $l_n n \log^{-3} n \to \infty$ and $u_n = o(1)$. Then, there exists a standard Brownian motion W, such that for $\nu > e$,*

$$\sup_{G\in\mathcal{G}_{c,C}} \left| T_n - \sup_{(t,h)\in B_n} w_h \left(\frac{\big|\int \phi^{(m)}(\frac{s-t}{h}) dW_s\big|}{\sqrt{h}} - \sqrt{2\log\frac{\nu}{h}} \right) \right| = O_P(r_n),$$

with

$$r_n = \sup_{G\in\mathcal{G}} \big\|\widehat{g}_n - g\big\|_\infty \frac{\log n}{\log\log n} + l_n^{-1/2} n^{-1/2} \frac{\log^{3/2} n}{\log\log n} + \frac{\sqrt{u_n \log(1/u_n)}}{\log\log(1/u_n)}.$$

Moreover,

$$\sup_{t\in[0,1],\ h>0,\ t+h\leq 1} w_h \left(\frac{\big|\int \phi^{(m)}(\frac{s-t}{h}) dW_s\big|}{\sqrt{h}} - \sqrt{2\log\frac{\nu}{h}} \right) < \infty, \quad a.s. \tag{2.8}$$

Hence, the limit statistic is almost surely bounded from above.

The proof of the coupling in this theorem (cf. Appendix A) is based on generalizing techniques developed by Giné *et al.* [15], while finiteness of the limiting test statistic utilizes results of Dümbgen and Spokoiny [10]. Note that Theorem 1 can be understood as a multiscale analog of the L^∞-loss convergence for kernel estimators (cf. [16, 15, 4, 17]).

Let us give some interesting examples for the choice of B_n illuminating the wide range of applications of Theorem 1.

Example 1 (Confidence bands for $L^\star g$ with fixed bandwidth)**.** *Let $h = h_n$ be a sequence converging to zero and assume for simplicity that $h_n \lesssim n^{-\kappa},\ \kappa > 0$ and $h_n n \log^{-3} n \to \infty$. Consider $B_n := [0, 1-h_n] \times \{h_n\}$. Then, we obtain*

$$\sup_{t\in[0,1-h]} \frac{h^{m-1/2}\big|T_{t,h} - \mathbb{E}\, T_{t,h}\big|}{\sqrt{\widehat{g}_n(t)}\ \alpha_m(t)} = \sup_{t\in[0,1-h]} \frac{\Big|\int \phi^{(m)}\big(\frac{s-t}{h}\big) dW_s\Big|}{\sqrt{h}} \\ + O_p\big(\|\widehat{g}_n - g\|_\infty \sqrt{\log n} + (nh)^{-1/2}\log n + h^{1/2}\big).$$

Using Theorem A1 of [2], we recover essentially Corollary 2 in [4].

Example 2 (Wavelet thresholding)**.** *Suppose that ϕ is a wavelet with compact support on $[0,1]$, for instance, the Haar wavelet, i.e. $\phi(\cdot) = \mathbb{I}_{[0,1/2)}(\cdot) - \mathbb{I}_{[1/2,1)}(\cdot) \in \mathcal{TV}^{(0)}$. Then, the wavelet coefficients are given by*

$$d_{j,k} := \int \phi\big(2^j s - k\big)(L^\star g)(s) ds = \int \phi_{2^{-j}k, 2^{-j}}(s)(L^\star g)(s) ds.$$

Suppose that j_{0n} and j_{1n} are integers satisfying $2^{-j_{1n}} n \log^{-3} n \to \infty$ and $j_{0n} \to \infty$. Set

$$B_n = \big\{(t,h) = (2^{-j}k, 2^{-j})\big|\ k = 0,1,\ldots,2^j - 1,\ j_{0n} \le j \le j_{1n},\ j \ \textit{integer}\ \big\}.$$

Then for $\alpha \in (0,1)$, Theorem 1 yields in a natural way level-dependent thresholds $q_{j,k}(\alpha)$, such that asymptotically

$$\lim_{n\to\infty} \mathbb{P}\Big(\big|\widehat{d}_{j,k} - d_{j,k}\big| \le q_{j,k}(\alpha),\ \textit{for all } j,k,\ \textit{with } (2^{-j}k, 2^{-j}) \in B_n\Big) = 1 - \alpha.$$

Let us close this section with a number of remarks.

Theorem 1 shows that the limit statistic is almost surely bounded from above. Note that we have the trivial lower bound

$$T_n \ge -\frac{\log \frac{\nu}{h}}{\log\log \frac{\nu}{h}},$$

which describes the behavior of T_n, provided the cardinality of B_n is small (for instance if B_n contains only one element). However, if B_n is sufficiently rich, the limit is also bounded from below. Let us make this more precise. Assume, that for every n there exists a $K_n,\ K_n \to \infty$ such that

$$B^\circ_{K_n} := \left\{ \left(\tfrac{i}{K_n}, \tfrac{1}{K_n}\right) \mid i = 0, \ldots, K_n - 1 \right\} \subset B_n.$$

Then, the limit statistic is asymptotically bounded from below by $-1/4$. This follows from

Lemma 1. *Assume that $K_n \to \infty$ and $\|\phi^{(m)}\|_{L^2} = 1$. Then,*

$$\sup_{(t,h)\in B^\circ_{K_n}} w_h \left(\frac{\left|\int \phi^{(m)}(\frac{s-t}{h}) dW_s\right|}{\sqrt{h}} - \sqrt{2\log\tfrac{\nu}{h}} \right) \to -\frac{1}{4}, \quad \textit{in probability.}$$

It is a challenging problem to calculate the distribution for general index set B_n explicitly. Although the tail behavior has been studied for the one-scale case (cf. [15, 4]) this has not been addressed so far for the limit statistic in Theorem 1. For implementation, later on, our method relies therefore on Monte Carlo simulations.

3 Testing for shape constraints

We start by explaining the main idea of the test. Let D denote the differentiation operator. Suppose that for $m \geq 0$, we have a linear differential operator of the form

$$f \mapsto Af(x) := \alpha_{-1}(x) + \sum_{l=0}^{m} \alpha_l(x) D^l f(x). \tag{3.1}$$

Throughout the remaining part of the paper, we will always assume that Af is continuous. A rectangle in $\mathbb{R}^2$ with vertices $(a_1, b_1), (a_1, b_2), (a_2, b_1), (a_2, b_2),\ a_1 < a_2,\ b_1 < b_2$ will be denoted by $\mathrm{Rect}(a_1, a_2, b_1, b_2)$.

The main objective of this paper is to obtain uniform confidence statement of the following kinds:

(i) The number and location of the roots and maxima of Af.

(ii) Simultaneous identification of intervals of the form $[t_i, t_i + h_i],\ t_i \in [0,1], h_i > 0,\ i$ in some index set I, with the following property: For a pre-specified confidence level we can conclude that for all $i \in I$ the functions $(Af)|_{[t_i, t_i+h_i]}$ attain, at least on a subset of $[t_i, t_i + h_i]$, positive values.

(ii') Same as (ii), but we want to conclude that $(Af)|_{[t_i, t_i+h_i]}$ has to attain negative values.

(iii) For any pair $(t,h) \in B_n$ with B_n as in (2.4), we want to find $b_-(t,h,\alpha)$ and $b_+(t,h,\alpha)$, such that we can conclude that with overall confidence $1-\alpha$, the graph of Af (denoted as graph(Af) in the sequel) has a non-empty intersection with every rectangle $\mathrm{Rect}(t, t+h, b_-(t,h,\alpha), b_+(t,h,\alpha))$.

In the following we will refer to these goals as Problems (i), (ii), (ii') and (iii), respectively. Note that the answer to Problem (ii) follows from Problem (iii) by taking all intervals $[t,t+h]$ with $b_-(t,h,\alpha) > 0$. Analogously, $[t,t+h]$ satisfies (ii') whenever $b_+(t,h,\alpha) < 0$. The geometrical ordering of the intervals obtained by (ii) and (ii') yields in a straightforward way a lower bound for the number of roots of Af, solving Problem (i) (cf. also Dümbgen and Walther [11]). A confidence interval for the location of a root can be constructed as follows: If there exists $[t,t+h]$ such that $b_-(t,h,\alpha) > 0$ and $[\widetilde{t}, \widetilde{t}+\widetilde{h}]$ with $b_+(\widetilde{t},\widetilde{h},\alpha) < 0$, then, with confidence $1-\alpha$, Af has a zero in the interval $\left[\min(t,\widetilde{t}), \max(t+h, \widetilde{t}+\widetilde{h})\right]$.

Example 3. *Suppose $A = D$. In this case we want to find a collection of intervals $[t,t+h]$ such that with overall probability $1-\alpha$ for each such interval there exists a nondegenerate subinterval on which f is strictly monotonically increasing.*
To state it differently, suppose that f' is continuous and $\phi \geq 0$ is a kernel with support on $[0,1]$, i.e. $\phi \geq 0$ with $\int_0^1 \phi(x)dx = 1$. If $\int \phi_{t,h}(x) f'(x)dx > 0$, then there is a nondegenerate subinterval of $[t,t+h]$ on which $f' > 0$. In particular, we can reject the null hypothesis that $f' \leq 0$ on $[t,t+h]$ at level $1-\alpha$. More generally, $\int \phi_{t,h}(x) f'(x)dx \in [a,b]$ implies by the intermediate value theorem that the graph of f' intersects the rectangle $\mathrm{Rect}(t, t+h, ah^{-1}, bh^{-1})$ *in at least one point.*

Example 4. *Suppose that we want to analyze the convexity/concavity properties of $U = p(X)$ (for instance $U = e^X$), where p is a function, which is strictly monotone increasing on the support of the distribution of X. Let f_U denote the density of U. Then, by change of variables*

$$f_U(y) = \frac{1}{p'\big(p^{-1}(y)\big)} f\big(p^{-1}(y)\big),$$

and there is an A of the form (3.1), such that $f_U''(y) = (Af)(p^{-1}(y))$. Therefore, $\mathrm{graph}(Af) \cap \mathrm{Rect}(t, t+h, b_-(t,h,\alpha), b_+(t,h,\alpha)) \neq \varnothing$ *implies*

$$\mathrm{graph}(f_U'') \cap \mathrm{Rect}(p(t), p(t+h), b_-(t,h,\alpha), b_+(t,h,\alpha)) \neq \varnothing.$$

In particular, if $b_-(t,h,\alpha) > 0$ then, with confidence $1-\alpha$, we may conclude that f_U is strictly convex on a nondegenerate subinterval of $[p(t), p(t+h)]$.

Since in our deconvolution setting f is not directly accessible, we show that we can write (under the imposed boundary conditions) $\int \phi_{t,h}(x) Af(x)dx$ as a scalar product of some differential operator L and the density g (cf. Lemma 2). Then, we may apply the results from Section 2.

This gives rise to the following definitions. The so called formal adjoint operator of A is given by

$$\phi \mapsto A^\star \phi(x) := \langle \alpha_{-1}, \phi \rangle + \sum_{l=0}^{m} (-1)^l D^l (\alpha_l \phi)(x).$$

If it exists, we denote by ψ_ϵ the multiplicative inverse of the moment generating function of ϵ. Note that under Assumption 1, $\psi_\epsilon(t) = \sum_{j=0}^{r} q_j t^j$. For $\epsilon = 0$, we set $\psi_\epsilon(t) = 1$.

The following Lemma is the key result for our test.

Lemma 2. *Suppose that Assumption 1 holds and let $\phi \in \mathcal{TV}^{(m+r)}$. Then,*

$$\langle \phi_{t,h}, Af \rangle = \langle A^\star \phi_{t,h}, f \rangle = \langle \psi_\epsilon(D) A^\star \phi_{t,h}, g \rangle.$$

Proof. Note that $\phi \in \mathcal{TV}^{(m)}$ implies that $\phi^{(l)}(0) = \phi^{(l)}(1) = 0$ for $l < m$. The first equality follows by iterated partial integration and for the second see (1.2). □

In analogy to (2.3) let us define

$$T_{t,h} = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} q_r^{-1} \psi_\epsilon(D) A^\star \phi_{t,h}(Y_k). \tag{3.2}$$

By Lemma 2, $\mathbb{E}\, T_{t,h} = \sqrt{n} q_r^{-1} \langle \phi_{t,h}, Af \rangle$. Recall that $(-1)^m q_r^{-1} \psi_\epsilon(D) A^\star$ is a linear differential operator of the form (2.2). Following (2.6), we define

$$T_n := \sup_{(t,h) \in B_n} w_h \left(\frac{h^{m+r-1/2} \left| T_{t,h} - \sqrt{n} q_r^{-1} \langle \phi_{t,h}, Af \rangle \right|}{\sqrt{\widehat{g}_n(t)}\ \alpha_m(t)\ \|\phi^{(m+r)}\|_{L^2}} - \sqrt{2 \log \tfrac{\nu}{h}} \right), \tag{3.3}$$

where B_n, w_n and $\widehat{g}_n$ are as in (2.4), (2.5) and (2.7), respectively. Note that the order of $(-1)^m q_r^{-1} \psi_\epsilon(D) A^\star$ is $m + r$. We have convergence of T_n to

$$\widetilde{T}_n(W) := \sup_{(t,h) \in B_n} w_h \left(\frac{\left| \int \phi^{(m+r)}(\frac{s-t}{h}) dW_s \right|}{\sqrt{h}\ \|\phi^{(m+r)}\|_{L^2}} - \sqrt{2 \log \tfrac{\nu}{h}} \right), \tag{3.4}$$

as a direct consequence of Theorem 1. Recall the definition of $\mathcal{G}_{c,C}$ in (2.1). In order to formulate the next theorem, define $\mathcal{F} = \mathcal{F}_{\epsilon,c,C}$ as the space of densities f such that the corresponding distribution function of Y is in $\mathcal{G}_{c,C}$.

Theorem 2. *Work under Assumptions 1. Assume further that* $\phi \in \mathcal{TV}^{m+r}$, $\mathrm{TV}(\alpha_q^{(q+r)}) + \|\alpha_q^{(q+r)}\|_\infty < \infty$ *for* $q = 0, \ldots, m$, *and that* α_m *is Lipschitz and bounded away from zero. Suppose that* $l_n n \log^{-3} n \to \infty$ *and* $u_n = o(1)$. *Then, there exists a standard Brownian motion* W, *such that for* $\nu > e$,

$$\sup_{f \in \mathcal{F}} \left| T_n - \widetilde{T}_n(W) \right| = o_P(1). \tag{3.5}$$

Moreover, uniformly over B_n, $\widetilde{T}_n(W)$ *is almost surely bounded from above.*

Proof. Apply Theorem 1 to $\phi / \|\phi^{(m+r)}\|_{L^2}$ and $L = q_r^{-1} \psi_\epsilon(D) A^\star$. □

Clearly, the distribution of $\widetilde{T}_n(W)$ depends only on known quantities. By ignoring the $o_P(1)$ term on the right hand side of (3.5), we can therefore simulate the distribution of T_n. To formulate it differently, the $(1-\alpha)$-quantile of the statistic T_n is asymptotically given by the $(1-\alpha)$-quantile of $\widetilde{T}_n(W)$ (denoted by $q_\alpha(\widetilde{T}_n(W))$ in the sequel).

In order to obtain a confidence band one has to control the bias which requires a Hölder condition on Af. However, since we are more interested in a qualitative analysis, it suffices to assume that Af is continuous. Moreover, instead of a moment condition on the kernel ϕ, we require positivity, i.e. for the remaining part of this section, let us assume that $\phi \geq 0$ and $\int \phi(u) du = 1$. Therefore, we can conclude that asymptotically with probability $1 - \alpha$, for all $(t, h) \in B_n$,

$$\langle \phi_{t,h}, Af \rangle \in \Big[q_r \frac{T_{t,h} - a_{t,h}}{\sqrt{n}}, q_r \frac{T_{t,h} + a_{t,h}}{\sqrt{n}} \Big], \tag{3.6}$$

where

$$a_{t,h} := h^{1/2-m-r} \sqrt{\widehat{g}_n(t)} \alpha_m(t) \big\| \phi^{(m+r)} \big\|_{L^2} \sqrt{2 \log \tfrac{\nu}{h}} \Big(1 + q_\alpha(\widetilde{T}_n(W)) \frac{\log \log \frac{\nu}{h}}{\log \frac{\nu}{h}} \Big). \tag{3.7}$$

Using the continuity of Af, it follows that asymptotically with confidence $1 - \alpha$, for all $(t, h) \in B_n$, the graph of $x \mapsto Af(x)$ has a non-empty intersection with each of the rectangles

$$\mathrm{Rect} \Big(t, t+h, q_r \frac{T_{t,h} - a_{t,h}}{h\sqrt{n}}, q_r \frac{T_{t,h} + a_{t,h}}{h\sqrt{n}} \Big). \tag{3.8}$$

This means we find a solution of (iii) by setting

$$b_-(t, h, \alpha) := q_r \frac{T_{t,h} - a_{t,h}}{h\sqrt{n}}, \quad b_+(t, h, \alpha) := q_r \frac{T_{t,h} + a_{t,h}}{h\sqrt{n}}. \tag{3.9}$$

4 Choice of kernel and performance of the multiscale statistic

In this section, we investigate the size/area of the rectangles constructed in the previous paragraphs. Recall that the expectation of the statistic $T_{t,h}$ depends in general on all derivatives up to $\phi^{(m+r)}$ (cf. Lemma 2). In contrast, the variance of $T_{t,h}$ depends asymptotically only on the highest derivative $\phi^{(m+r)}$. Therefore, $\phi^{(m+r)}$ appears in the limit statistic $\widetilde{T}(W)$, but no other derivative does. In fact, we shall see in this section that our result can be compared to estimation of the $(m+r)$-th derivative of a density.

Optimal choice of the kernel: In the following, we are going to study the problem of finding the optimal function ϕ. It turns out that this can be done explicitly.

Note that for given $(t,h) \in B_n$, the width of the rectangle (3.8)) is given by $2q_r a_{t,h}/(h\sqrt{n})$. Further, the choice of ϕ influences the value of $a_{t,h}$ in two ways, namely by the factor $\|\phi^{(m+r)}\|_{L^2}$ as well as the quantile $q_\alpha(\widetilde{T}_n(W))$ (cf. the definition of $a_{t,h}$ given in (3.7)). Since α is fixed, we have for $n \to \infty$,

$$q_\alpha(\widetilde{T}_n(W)) \frac{\log\log \frac{\nu}{h}}{\log \frac{\nu}{h}} = o(1).$$

Therefore, $a_{t,h}$ depends in first order on $\|\phi^{(m+r)}\|_{L^2}$ and our optimization problem boils down to

$$\text{minimize } \|\phi^{(m+r)}\|_{L^2}, \quad \text{subject to } \int \phi(u)du = 1, \ \phi \in \mathcal{TV}^{(m+r)}.$$

This is in fact easy to solve. By Lagrange calculus, we find that on $(0,1)$, ϕ has to be a polynomial of order $2m+2r$. Under the boundary conditions, the solution ϕ_{m+r} has the form

$$\phi_{m+r}(x) = c_{m+r} x^{m+r}(1-x)^{m+r}\mathbb{I}_{(0,1)}(x). \tag{4.1}$$

Due to the normalization constraint $\int \phi_{m+r}(u)du = 1$, it follows that ϕ_{m+r} is the density of a Beta distributed random variable with parameters $\alpha = m+r+1$ and $\beta = m+r+1$, implying, $c_{m+r} = (2m+2r+1)!/((m+r)!)^2$. It is worth mentioning that $\phi_{m+r}^{(m+r)}$, restricted to the domain $[-1,1)$, is (up to translation/scaling) the $(m+r)$-th Legendre polynomial L_{m+r}, i.e.

$$\phi_{m+r}^{(m+r)} = (-1)^{m+r}\frac{(2m+2r+1)!}{(m+r)!} L_{m+r}(2\cdot -1)$$

(this is essentially Rodrigues' representation, cf. Abramowitz and Stegun [1], p. 785). For that reason, we even can compute

$$\left\|\phi_{m+r}^{(m+r)}\right\|_{L^2} = \frac{(2m+2r)!}{(m+r)!}\sqrt{2m+2r+1}.$$

In the particular case $r=0,\ m=1$ this is known from the work of Dümbgen and Walther [11], where the authors use locally most powerful tests to derive ϕ_1.

4.1 Performance of the method

In this part, we give some theoretical insights. We start by investigating Problem (*iii*) (cf. Section 3). After that, we will address issues related to (*ii*) and (*i*).

Problem (iii): Recall that with confidence $1-\alpha$, for all $(t,h)\in B_n$,

$$\operatorname{graph}(Af)\cap \operatorname{Rect}\Big(t, t+h, q_r\frac{T_{t,h}-a_{t,h}}{h\sqrt{n}}, q_r\frac{T_{t,h}+a_{t,h}}{h\sqrt{n}}\Big) \neq \varnothing.$$

The so constructed rectangles contain information on Af, where the amount of information is directly linked to the size of the rectangle. Therefore, it is natural to think of the area and the length of the diagonal as measures of localization quality. For the rectangle above, the area is given by

$$\operatorname{area}(t,h) := 2q_r a_{t,h} n^{-1/2} \sim h^{1/2-m-r} n^{-1/2}\sqrt{\log\frac{1}{h}}.$$

There is an interesting transition: Suppose that $m=r=0$ (density estimation). Then, $\operatorname{area}(t,h)\to 0$ for every h and $n\to\infty$. In contrast, whenever $m+r>0$,

$$\begin{aligned}
h \gg (\log n/n)^{1/(2m+2r-1)} \quad &\Rightarrow \quad \operatorname{area}(t,h)\to 0,\\
h \sim (\log n/n)^{1/(2m+2r-1)} \quad &\Rightarrow \quad \operatorname{area}(t,h) = O(1),\\
h \ll (\log n/n)^{1/(2m+2r-1)} \quad &\Rightarrow \quad \operatorname{area}(t,h)\to \infty.
\end{aligned}$$

On the other hand, the length of the diagonal behaves like $h \vee h^{-m-r-1/2} n^{-1/2}\sqrt{\log 1/h}$. If the rectangle is a square, then, $h\sim(\log n/n)^{1/(3+2m+2r)}$.

Problem (ii), (ii′): The following lemma gives a necessary condition in order to solve (*ii*). Loosely speaking, it states that whenever

$$Af\big|_{[t,t+h]} \gtrsim n^{-1/2} h^{-m-r-1/2}\sqrt{\log 1/h},$$

the multiscale test returns a rectangle $\operatorname{Rect}(t,t+h,b_-(t,h,\alpha),b_+(t,h,\alpha))$ which is in the upper half-plane with high-probability. Or, to state it differently, we can reject that $Af\big|_{[t,t+h]} < 0$.

Theorem 3. *Work under the assumptions of Theorem 2. Suppose that* $\phi \geq 0$*. Let* M_n^- *denote the set of tupels* $(t,h) \in B_n$ *for which*

$$A f\big|_{[t,t+h]} > \frac{2 q_r a_{t,h}}{h\sqrt{n}}.$$

Similar, define $M_n^+ := \{(t,h) \in B_n \mid Af|_{[t,t+h]} < -(2q_r a_{t,h})/(h\sqrt{n})\}$*. Then, if* $b_+(t,h,\alpha)$ *and* $b_-(t,h,\alpha)$ *are given by* (3.9)*, we obtain*

$$\lim_{n\to\infty} \mathbb{P}\Big((-1)^{\mp} b_{\pm}(t,h,\alpha) > 0, \text{ for all } (t,h) \in M_n^{\pm}\Big) \geq 1-\alpha.$$

Proof. For all $(t,h) \in M_n^-$, conditionally on the event given by (3.6),

$$A f\big|_{[t,t+h]} > \frac{2 q_r a_{t,h}}{h\sqrt{n}} \;\Rightarrow\; \langle \phi_{t,h}, Af\rangle > \frac{2 q_r a_{t,h}}{\sqrt{n}} \;\Rightarrow\; T_{t,h} > a_{t,h} \;\Rightarrow\; b_-(t,h,\alpha) > 0.$$

Similar, one can argue for M_n^+. □

In order to formulate the next result, let us define

$$C_\alpha := (8\|f_\epsilon\|_\infty q_r^2 \|\alpha_m\|_\infty^2 \|\phi^{(m+r)}\|_{L^2}^2 (1 + q_\alpha(\widetilde{T}_n(W)))^2)^{1/(2m+2r+1)}. \tag{4.2}$$

Corollary 1. *Work under the assumptions of Theorem 2. Suppose that* $\phi \geq 0$ *and* $\beta \in \mathbb{R}$*. Let* M_n^- *denote the set of tupels* $(t,h) \in B_n$ *satisfying*

$$A f\big|_{[t,t+h]} > \left(\frac{\log n}{n}\right)^{\beta/(2\beta+2m+2r+1)} \tag{4.3}$$

and

$$h \geq C_\alpha \left(\frac{\log n}{n}\right)^{1/(2\beta+2m+2r+1)}.$$

Let M_n^+ *be as* M_n^-*, with* (4.3) *replaced by* $Af|_{[t,t+h]} < -(\log n/n)^{\beta/(2\beta+2m+2r+1)}$*. Then, if* $b_-(t,h,\alpha)$ *and* $b_+(t,h,\alpha)$ *are given by* (3.9)*, we obtain*

$$\lim_{n\to\infty} \mathbb{P}\Big((-1)^{\mp} b_{\pm}(t,h,\alpha) > 0, \text{ for all } (t,h) \in M_n^{\pm}\Big) \geq 1-\alpha.$$

Proof. It holds that

$$a_{t,h} \leq h^{1/2-m-r}\, \|f_\epsilon\|_\infty^{1/2}\, \|\alpha_m\|_\infty \big\|\phi^{(m+r)}\big\|_{L^2} \sqrt{2\log \nu/h}\big(1 + q_\alpha(\widetilde{T}_n(W))\big)$$

For sufficiently large n, $h \geq l_n \geq \nu/n$. Therefore, we have for every $(t,h) \in M_n^-$,

$$\frac{2 q_r a_{t,h}}{h\sqrt{n}} \leq \sqrt{8\,\|f_\epsilon\|_\infty} q_r \|\alpha_m\|_\infty \big\|\phi^{(m+r)}\big\|_{L^2}\big(1 + q_\alpha(\widetilde{T}_n(W))\big) h^{-m-r-1/2} n^{-1/2}\sqrt{\log n} < A f\big|_{[t,t+h]}.$$

Similar for M_n^+. Now, the result follows by applying Theorem 3. □

The last result shows essentially that if $Af\big|_{[t,t+h]}$ is positive and $Af\big|_{[t,t+h]} \sim (\log n/n)^{\beta/(2\beta+2m+2r+1)}$ and $h \sim (\log n/n)^{1/(2\beta+2m+2r+1)}$ then with probability $1-\alpha$, our method returns a rectangle in the upper half-plane. Another way to state this is by imposing the condition

$$Af\big|_{[t,t+h]} \gtrsim h^{\beta}. \tag{4.4}$$

We have three distinct regimes

$$\begin{array}{lll} \beta > 0: & Af\big|_{[t,t+h]} \to 0 & h \to 0, \\ \beta = 0: & Af\big|_{[t,t+h]} = O(1) & h \sim (\log n/n)^{1/(2m+2r+1)} \to 0, \\ -m-r-1/2 < \beta < 0: & Af\big|_{[t,t+h]} \to \infty & h \to 0. \end{array}$$

It is of importance to compare the previous result to derivative estimation of a density. As it is well known, we could estimate Af with rate of convergence

$$\Big(\frac{\log n}{n}\Big)^{\beta/(2\beta+2m+2r+1)}$$

with respect to L^∞-norm assuming that Af is Hölder continuous with index $\beta > 0$ and that $h \sim (\log n/n)^{1/(2\beta+2m+2r+1)}$. This directly relates to the first case considered above.

Assuming that Af is smooth. If we want to use Theorem 2 for construction of confidence bands, we have to restrict us to scales $h \sim (\log n/n)^{1/(2\beta+2m+2r+1)}$, $\beta < \beta_0$, where β_0 denotes the Hölder index of Af.

Problem (i): Recall the construction of confidence bands given in Section 3. We will give a bound for the length of such a confidence interval, provided that Af has exactly one root. For example, this can be an extreme/saddle point if $A = D$ or a point of inflection if $A = D^2$.

In order to formulate the result, we need that B_n is sufficiently rich. Therefore, we assume that for all n, there exists an N_n, $N_n \gtrsim n^{1/(2m+2r+1)} \log^4 n$, such that

$$\Big\{\Big(\frac{k}{N_n}, \frac{l}{N_n}\Big) \mid k = 0, 1, \dots, \ l = 1, 2, \dots, \ k + l \le 1\Big\} \subset B_n.$$

Assume further that in a local neighborhood of the root x_0, Af behaves like

$$Af(x) = \gamma \operatorname{sign}(x - x_0)|x - x_0|^\beta + o(|x - x_0|^\beta),$$

for some positive β. Let $\rho_n = (\log n/n)^{1/(2\beta+2m+2r+1)} 2/\gamma^{1/\beta}$ and $C_\alpha, M^\pm$ as defined in Corollary 1. There exist integer sequences $(k_n^-)_n$, $(k_n^+)_n$, $(l_n)_n$ such that for all sufficiently large n,

$$\rho_n \le \frac{k_n^-}{N_n} - x_0 \le 2\rho_n, \quad -2\rho_n \le \frac{k_n^+}{N_n} - x_0 \le -\rho_n, \quad \text{and} \quad C_\alpha \gamma^{1/\beta} \rho_n \le \frac{l_n}{N_n} \le 2C_\alpha \gamma^{1/\beta} \rho_n.$$

Some calculations show that $(k_n^-/N_n, l_n/N_n) \in M_n^-$ and $((k_n^+ - l_n)/N_n, l_n/N_n) \in M_n^+$. We can conclude from Corollary 1 and the construction, that the confidence interval has to be a subinterval of

$$\left[\frac{k_n^+ - l_n}{N_n}, \frac{k_n^- + l_n}{N_n}\right].$$

Hence, the length of the confidence interval is bounded from above by

$$4(C_\alpha \gamma^{1/\beta} + 1)\rho_n \sim \left(\frac{\log n}{n}\right)^{1/(2\beta+2m+2r+1)}.$$

Observe that for localization of modes in density estimation $(m, r, \beta) = (1, 0, 1)$ the rate $(\log n/n)^{1/5}$ is optimal up to the log-factor (cf. Hasminskii [18]). The rate $(\log n/n)^{1/7}$ for localization of inflection points in density estimation $(m, r, \beta) = (2, 0, 1)$ coincides with the one found in Davis *et al.* [7].

4.2 On calibration of multiscale statistics

Let us shortly comment on the type of multiscale statistic, derived in Theorem 1. Following [10], p.139, we can view the calibration of the multiscale statistics (2.6) and (3.3) as a generalization of Lévy's modulus of continuity. In fact, the supremum is attained uniformly over different scales, making this calibration in particular attractive for construction of adaptive methods.

One of the restrictions of our method, compared to other works on multiscale statistics, is that we exclude the coarsest scales, i.e. $h > u_n = o(1)$ (cf. Theorem 2). Otherwise the limit statistic would not be distribution-free. However, excluding the coarsest scales is a very weak restriction since the important features of Af can be already detected at scales tending to zero with a certain rate. For instance in view of Corollary 1, the multiscale method detects a deviation from zero, i.e. $Af\big|_I \geq C > 0$, provided the length of the interval I is larger than const.$\times(\log n/n)^{1/(2m+2r+1)}$. This can be also seen by numerical simulations, as outlined in the next section.

5 Numerical simulations

For any $(t, h) \in B_n$ the multiscale method returns a rectangle of the form (3.8). However, most of the rectangles are redundant since the fact that graph(Af) intersects these rectangles can be deduced already from the position of other rectangles (see for instance Figure 1) and the assumption that Af is continuous. Naturally, we are interested in the set of

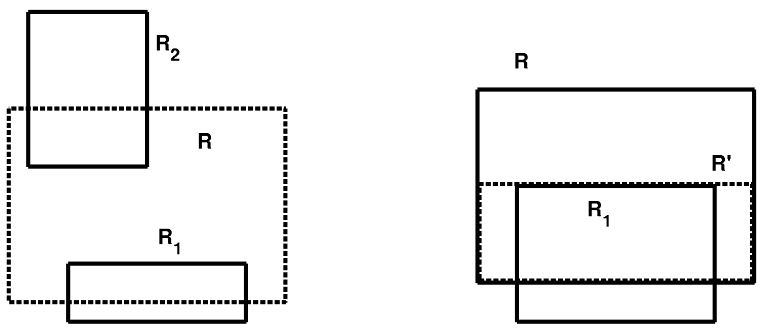

Figure 1: If the graph of Af intersects R_1 and R_2, then also R (left). If graph(Af) intersects R and R_1, then also R' (right).

rectangles, which are informative in the sense that they contain information on the signal, which cannot be deduced from other rectangles. Let us describe in three steps (A), (B), (B'), how to discard redundant rectangles.

(A) Fix $(t,h) \in B_n$. Suppose there exists $(t_1,h_1),(t_2,h_2) \in B_n$ ((t_1,h_1) and (t_2,h_2) not necessarily different) such that $[t_1,t_1+h_1],[t_2,t_2+h_2] \subset [t,t+h]$, $b_+(t_1,h_1,\alpha) \leq b_+(t,h,\alpha)$ and $b_-(t_2,h_2,\alpha) \geq b_-(t,h,\alpha)$. Denote by R, R_1, R_2 the rectangle obtained from $(t,h),(t_1,h_1)$ and (t_2,h_2), respectively (for an illustration see Figure 1). Since Af is further assumed to be continuous, then by intermediate value theorem, $\mathrm{graph}(Af) \cap R_1 \neq \varnothing$ and $\mathrm{graph}(Af) \cap R_2 \neq \varnothing$ imply that $\mathrm{graph}(Af) \cap R \neq \varnothing$. Hence, in this case, R is non-informative and will be discarded.

(B) Fix $(t,h) \in B_n$ and denote the induced rectangle by R. Suppose there exists $(t_1,h_1) \in B_n$, such that $[t_1,t_1+h_1] \subset [t,t+h]$ and $b_-(t_1,h_1,\alpha) \leq b_-(t,h,\alpha) \leq b_+(t_1,h_1,\alpha) < b_+(t,h,\alpha)$ (see Figure 1). Define $R' := \mathrm{Rect}(t,t+h,b_-(t,h,\alpha),b_+(t_1,h_1,\alpha))$. Then, R' is contained in R and $\mathrm{graph}(Af) \cap R' \neq \varnothing$. Therefore, we replace R by R'.

(B'): Same as (B), but consider the case $b_-(t,h,\alpha) < b_-(t_1,h_1,\alpha) \leq b_+(t,h,\alpha) \leq b_+(t_1,h_1,\alpha)$. With $R' := \mathrm{Rect}(t,t+h,b_-(t_1,h_1,\alpha),b_+(t,h,\alpha))$ we obtain $\mathrm{graph}(Af) \cap R' \neq \varnothing$. Therefore, we replace R by R'.

Throughout the following, let us refer to the remaining rectangles after application of (A), (B) and (B') as (set of) minimal rectangles.

We will illustrate our method by investigating monotonicity of f ($A = D$, cf. Example 3)

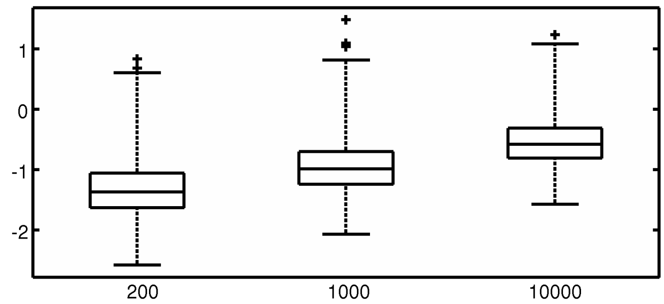

Figure 2: Boxplots for three different values ($n = 200,\ n = 1000,\ n = 10.000$) of the limit statistic (3.4).

under Laplace-deconvolution, i.e. $f_\epsilon(x) = e^{-|x|/\theta}/\theta$ with $\theta = 0.075$. In this case, we find

$$\psi_\epsilon(t) = 1 - \theta^2 t^2 \quad \text{and} \quad A^\star f = -f'$$

and the statistic (3.2) takes the explicit form

$$T_{t,h} = \frac{1}{h\sqrt{n}\,\theta^2} \sum_{k=1}^{n} \left(\frac{\theta^2}{h^2} \phi^{(3)} \left(\frac{Y_k - t}{h} \right) - \phi' \left(\frac{Y_k - t}{h} \right) \right).$$

As kernel ϕ, we select the density of a Beta$(4,4)$ random variable (cf. Section 4). Moreover, we choose $u_n = 1/\log\log n$ for the multiscale statistic and define

$$B_n = \Big\{ \Big(\frac{k}{N_n}, \frac{l}{N_n} \Big) \mid k = 0, 1, \ldots,\ l = 1, 2, \ldots, [N_n u_n],\ k + l \leq 1 \Big\}, \quad \text{for} \quad N_n = \big[n^{0.6}\big].$$

Boxplots for the corresponding limit distribution are displayed in Figure 2 for different values of n and 10.000 simulations each. These plots show that the distribution is well-localized with only a few outliers. As proved, the limit statistic is almost surely bounded for $n \to \infty$. For finite but increasing sample size, however, Figure 2 indicates, that the quantiles of the limit distribution grow slightly.

In Figures 3 and 4, we give an example of a reconstruction based on a sample size of $n = 1000$ and confidence level equals 90%. Based on 10.000 repetitions, the estimated quantile is $q_{0.1}(\widetilde{T}_{1000}(W)) = -0.41$. For the simulation, we use $\nu = \exp(e^2)$. Then, $h \mapsto \sqrt{\log \nu/h}/(\log\log \nu/h)$ is monotone as long as $0 < h \leq 1$ (cf. Lemma 4 (i)).

The upper display of Figure 3 shows the true density of f as well as the convoluted density g. Note that g is very smooth and as the other densities non-observable (we only have observations, which are distributed with density g). In fact, by visual inspection of g, we are not able to find the intervals on which f is monotone increasing/decreasing.

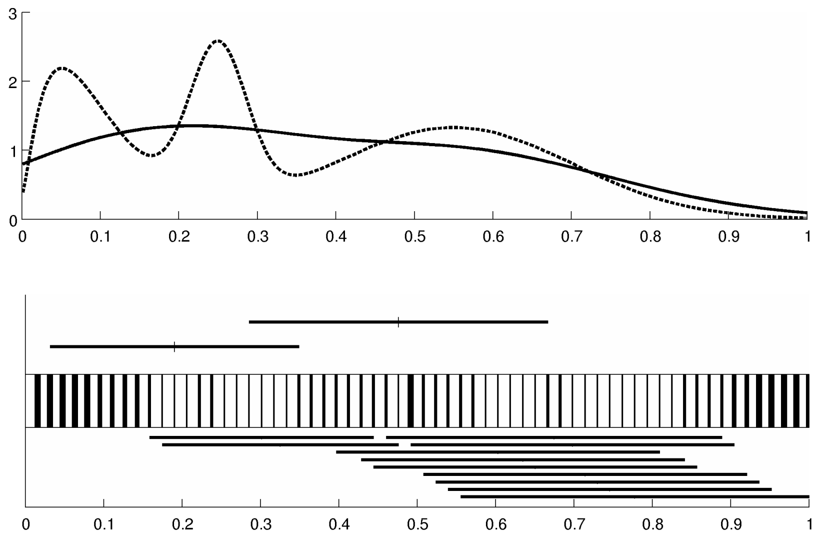

Figure 3: Simulation for sample size $n = 1000$ and 90%-quantile. *Upper display:* True density f (dashed) and convoluted density g (solid). *Lower display:* Line plot of the endpoints of intervals solving Problems (ii) and (ii') as well as minimal solutions to (ii) and (ii') (horizontal lines above/below)

The lower plot of Figure 3, displays minimal intervals which are solutions to Problems (ii) and (ii') (horizontal lines above and below the line plot, respectively). Here, minimal intervals for (ii) and (ii') denote the intervals for which no proper subinterval exists with the same property. The line plot itself depicts the endpoints of all intervals belonging to (ii) and (ii'). Note that the possible values for the endpoints are given by $k/N_n,\ k = 0, 1, \ldots, N_n$. If for given k there is more than one interval solving (ii) or (ii') with endpoint k/N_n the line width is increased accordingly. For more on this type of plotting, see Dümbgen and Walther [11].

The density f has been designed in order to investigate Corollary 1 numerically. Indeed, on $[0, 0.35]$, the signal (in this case $|f'|$) is in average large but the intervals on which f increases/decreases are comparably small. In contrast, on $[0.35, 1]$, $|f'|$ is small and there is only one increase/decrease.

The test is able to find two regions of increase and two regions, where the density decreases. The increase and decrease on the leftmost position are not detected by our test. Repetition of the simulation shows that the decrease on the intervals $[0.25, 0.35]$ and $[0.55, 1]$ is most of the time found while the increases (on $[0.17, 0.25]$ and $[0.35, 0.55]$) are less often detected. Furthermore, compared to the true function f, it can be seen that the difficulty lies in

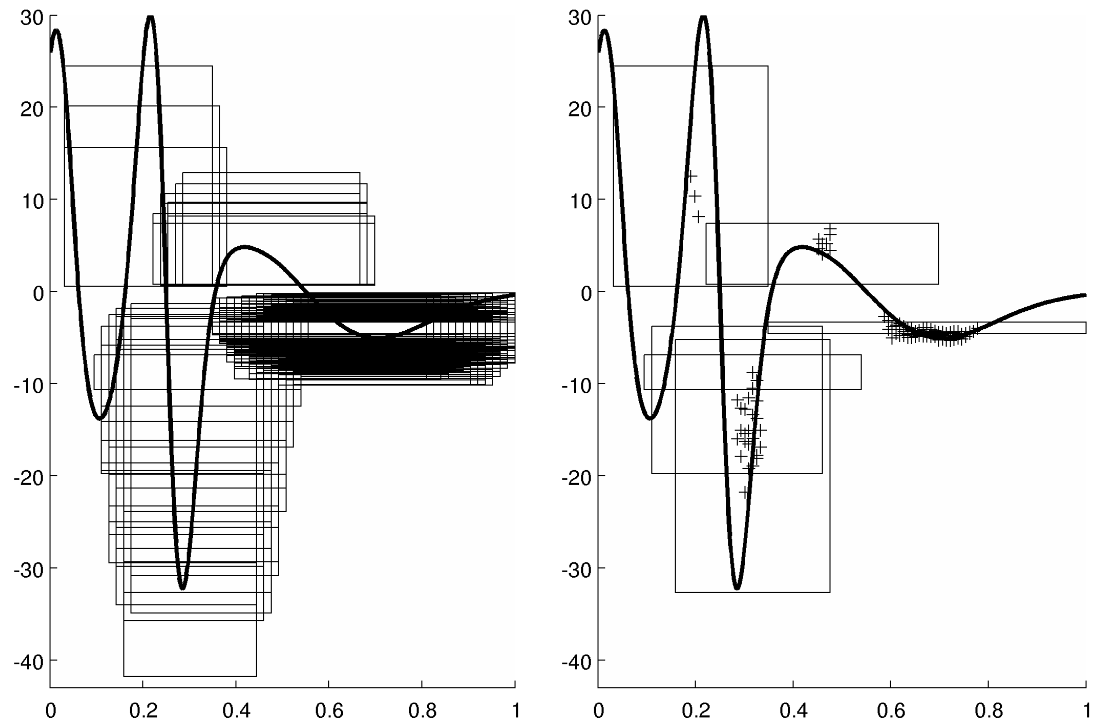

Figure 4: True (unobserved) derivative f' and minimal rectangles (left) as well as sparse minimal rectangles/ midpoints (right) for the same data set as in Figure 3.

precise localization of the regions of increase/decrease.

In Figure 4, the derivative of f as well as the minimal rectangles, additionally satisfying either $b_-(t,h,\alpha) > 0$ or $b_+(t,h,\alpha) < 0$, are displayed. For better visualization, we have depicted the midpoints of these rectangles and a sparse subset (right display in Figure 4) using the following reduction step:

(C): Let R be the rectangle with the smallest area and denote by S the set of rectangles having non-empty intersection with R. Find the rectangle in S minimizing the area of intersection with R. Display R and R' and discard R and all the rectangles in S. If there are rectangles left, start from the beginning.

By construction, we find as before two regions of increase and decrease. Compared to the multiscale solutions of Problems (ii) and (ii') (cf. Figure 3), we also obtain surprisingly precise information on the derivative of f. Observe that the graph of f' tends to cut the rectangles through the middle. Therefore, the midpoints of the rectangles (depicted as crosses in Figure 4) can be used for instance for estimation of maxima.

6 Outlook

We have investigated multiscale methods in order to analyze differential operators in deconvolution models. A more refined multiscale calibration has been considered using an idea of proof originally developed for construction of confidence bands. We believe that the same strategy can be applied to a variety of problems and more dimensional settings. In particular, similar results will hold for regression and spectral density estimation.

Our multiscale approach allows us to identify intervals such that for given significance level we know that $Af > 0$ at least on a subinterval. As outlined in Section 4, these results are sufficient for qualitative inference as for example construction of confidence bands for the roots of Af. Since we only required that Af is continuous, Af can be highly oscillating. In this framework, it is therefore impossible to obtain strong confidence statements in the sense that we find intervals on which Af is always positive. By adding bias controlling smoothness assumptions such as for instance Hölder conditions stronger results can be obtained resulting for instance in uniform confidence bands.

At the moment the proposed method is restricted to the class of blurring distributions introduced in Assumption 1 and extension to $r = \infty$ is not straightforward. Of particular interest is the case of Gaussian deconvolution. In this case the inversion formula is well known. It is basically the inverse Weierstrass transform (cf. Eddington [12], Pollard, [28], Widder [30]). Van Es and Kok [29] derive some heuristic arguments indicating that the inversion formula of a Gaussian can be approximated by the inversion formula of scaled sums of Laplacian distributed random variables satisfying Assumption 1.

Restricting to linear differential equations is a further drawback of our method, since very important shape constraints as for instance curvature cannot be handled within this framework and we may only work with linearizations (which is quite common in physics and engineering). Allowing for non-linearity however seems to be almost intractable.

We are aware of the fact that many other important qualitative features are also related to integral transforms (that are in general not of convolution type) and they even do not have a representation as differential inequality. For instance complete monotonicity and positive definiteness are by Bernstein's and Bochner's Theorem connected to the Laplace transform and Fourier transform, respectively. They cannot be handled with the methods proposed here and are subject to further investigations.

Acknowledgments. This research was supported by the German Science Foundation (DFG) and the Swiss National Science Foundation (SNF), grant FOR 916. The third author was partly funded by DFG postdoctoral fellowship SCHM 2807/1-1.

Appendix A

Throughout the appendix, let

$$w_h = \frac{\sqrt{\frac{1}{2}\log\frac{\nu}{h}}}{\log\log\frac{\nu}{h}}, \quad \widetilde{w}_h = \frac{\log\frac{\nu}{h}}{\log\log\frac{\nu}{h}}.$$

Proof of Theorem 1. Since the statistic only depends on $T_{t,h} - \mathbb{E}T_{t,h}$, we may assume that $\alpha_{-1} = 0$. Let us study in a first step the statistic

$$T_n^{(1)} = \sup_{(t,h)\in B_n} w_h \frac{h^{m-1/2}\left|T_{t,h} - \mathbb{E}T_{t,h}\right|}{\sqrt{g(t)}\alpha_m(t)} - \widetilde{w}_h.$$

Note that $T_n^{(1)}$ is the same as T_n, but $\widehat{g}_n$ is replaced by g. We will show that there exists a Brownian motion W, such that with

$$T_n^{(2)}(W) := \sup_{(t,h)\in B_n} w_h \frac{\left|\int (L\phi_{t,h})(s)\sqrt{g(s)}dW_s\right|}{\sqrt{g(t)}\alpha_m(t)} - \widetilde{w}_h,$$

we have

$$\sup_{G\in\mathcal{G}_{c,C}} \left|T_n^{(1)} - T_n^{(2)}(W)\right| = o_P(r_n). \tag{A.1}$$

The main argument is based on the standard version of KMT (cf. [24]). In order to state the result, let us define a Brownian bridge on the index set $[0,1]$ as a centered Gaussian process $(B(f))_{\{f\in\mathcal{F}\}}$, $\mathcal{F}\subset L^2([0,1])$ with covariance structure

$$\operatorname{Cov}\big(B(f), B(g)\big) = \langle f, g\rangle - \langle f, 1\rangle\langle g, 1\rangle.$$

Let $\mathcal{F}_0 := \{x \mapsto \mathbb{I}_{[0,s]}(x) : s\in[0,1]\}$. Note that $(B(f))_{\{f\in\mathcal{F}_0\}}$ coincides with the classical definition of a Brownian bridge. For $U_i \sim \mathcal{U}[0,1]$, i.i.d., the uniform empirical process on the function class $\mathcal{F}$ is defined as

$$\mathbb{U}_n(f) = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n f(U_i) - \int f(x)dx\right), \quad f\in\mathcal{F}.$$

In particular note that

$$T_{t,h} - \mathbb{E}\,T_{t,h} = \mathbb{U}_n\big((L\phi_{t,h})(G^{-1}(.))\big),$$

where G^{-1} denotes the quantile function of Y.

The key results is given by the following theorem.

Theorem 4 (KMT on $[0,1]$, cf. [24]). *There exist versions of $\mathbb{U}_n$ and a Brownian bridge B such that for all x*

$$\mathbb{P}\Big(\sup_{f\in\mathcal{F}_0}\big|\mathbb{U}_n(f)-B(f)\big|>n^{-1/2}(C\log n+x)\Big)<Ke^{-\lambda x},$$

where $C,K,\lambda>0$ are universal constants.

However, we need a functional version of KMT. We shall prove this by using the theorem above in combination with a result due to Koltchinskii [23], (Theorem 11.4, p. 112) stating that the supremum over a function class $\mathcal{F}$ behaves as the supremum over the symmetric convex hull $\overline{\mathrm{sc}}(\mathcal{F})$, defined by

$$\overline{\mathrm{sc}}(\mathcal{F}):=\Big\{\sum_{i=1}^{\infty}\lambda_i f_i : f_i\in\mathcal{F},\lambda_i\in[-1,1],\sum_{i=1}^{\infty}|\lambda_i|\le 1\Big\}.$$

Theorem 5. *Assume there exists a version B of a Brownian bridge, such that for a sequence $(\widetilde{\delta}_n)_n$ tending to 0,*

$$\mathbb{P}^*\Big(\sup_{f\in\mathcal{F}}|\mathbb{U}_n(f)-B(f)|\ge\widetilde{\delta}_n(x+C\log n)\Big)\le Ke^{-\lambda x},$$

where $C,K,\lambda>0$ are constants depending only on $\mathcal{F}$. Then, there exists a version $\widetilde{B}$ of a Brownian bridge, such that

$$\mathbb{P}^*\Big(\sup_{f\in\overline{\mathrm{sc}}(\mathcal{F})}|\mathbb{U}_n(f)-\widetilde{B}(f)|\ge\widetilde{\delta}_n(x+C'\log n)\Big)\le K'e^{-\lambda' x}$$

for constants $C',K',\lambda'>0$.

It is well-known (cf. Giné *et al.* [15], p. 172) that

$$\big\{\rho\;\big|\;\rho:[0,1]\to\mathbb{R},\rho(1)=0,\ \mathrm{TV}(\rho)\le 1\big\}\subset\overline{\mathrm{sc}}(\mathcal{F}_0). \tag{A.2}$$

This inclusion, will be the main ingredient in order to show

Lemma 3. *Under the assumptions of Theorem 1, there exists a positive constant $C_\star$, such that for every $C\ge 0$ we have the inclusion of function classes*

$$\Big\{C_\star h^m(L\phi_{t,h})(G^{-1}(.)):t\in[0,1],h\in(0,1],G\in\mathcal{G}_{0,C}\Big\}\subset\overline{\mathrm{sc}}(\mathcal{F}_0).$$

Let us denote

$$\mathcal{F}_n:=\Big\{C_\star h^m(L\phi_{t,h})(G^{-1}(.)):t\in[0,1],\ h\in[l_n,u_n],\ G\in\mathcal{G}_{c,C}\Big\}.$$

Combining Theorems 4 and 5 shows that there are constants C', K', λ' and a Brownian bridge $(B(f))_{f\in\overline{\mathrm{sc}}(\mathcal{F}_0)}$ such that for $x > 0$,

$$\mathbb{P}\Big(\sup_{t\in[0,1],\ h\in(0,1],\ G\in\mathcal{G}} C_\star h^m\big|(T_{t,h} - \mathbb{E}\,T_{t,h}) - B\big((L\phi_{t,h})(G^{-1}(.))\big)\big| \geq n^{-1/2}(x + C\log n)\Big) \\ \leq K'e^{-\lambda' x}.$$

Due to Lemma 4 (i) and $l_n \geq \nu/n$ for sufficiently large n, we have that $w_{l_n} \leq w_{\nu/n}$. This readily implies

$$\sup_{G\in\mathcal{G}}\sup_{(t,h)\in B_n} w_h \frac{h^{m-1/2}\Big|\big|T_{t,h} - \mathbb{E}\,T_{t,h}\big| - \big|B\big((L\phi_{t,h})(G^{-1}(.))\big)\big|\Big|}{\sqrt{g(t)}\ \alpha_m(t)} = O_P\Big(l_n^{-1/2}n^{-1/2}w_{\nu/n}\log n\Big).$$

Now, let us introduce the (general) Brownian motion $W(f)$ as a centered Gaussian process with covariance $\mathbb{E}[W(f)W(g)] = \langle f, g\rangle$. In particular, $W(f) = B(f) + (\int f)\xi$, $\xi \sim \mathcal{N}(0,1)$ and independent of B, defines a Brownian motion and hence there exists a version of $(W(f))_{f\in\overline{\mathrm{sc}}(\mathcal{F}_0)}$ such that $B(f) = W(f) - (\int f)W(1)$. By some calculations,

$$\sup_{G\in\mathcal{G}}\sup_{(t,h)\in B_n} w_h \frac{h^{m-1/2}\Big|\int L\phi_{t,h}(u)dG(u)\Big|}{\sqrt{g(t)}\ \alpha_m(t)} \lesssim \sup_{h\in[l_n,u_n]} w_h h^{1/2} \leq w_{u_n}u_n^{1/2},$$

where the last inequality follows from Lemma 4 (ii). This implies further

$$\mathbb{E}\Big[\Big\|w_h\frac{h^{m-1/2}}{\sqrt{g(t)}\ \alpha_m(t)}\Big[\big|B\big((L\phi_{t,h})(G^{-1}(.))\big)\big| - \big|W\big((L\phi_{t,h})(G^{-1}(.))\big)\big|\Big]\Big\|_{\mathcal{F}_n}\Big] = O(w_{u_n}u_n^{1/2}),$$

therefore

$$\sup_{G\in\mathcal{G}}\Big|T_n^{(1)} - \sup_{(t,h)\in B_n}\frac{w_h h^{m-1/2}\big|W\big((L\phi_{t,h})(G^{-1}(.))\big)\big|}{\sqrt{g(t)}\ \alpha_m(t)} - \widetilde{w}_h\Big| \\ = O_P(l_n^{-1/2}n^{-1/2}w_{1/n}\log n + w_{u_n}u_n^{1/2}),$$

and

$$\sup_{G\in\mathcal{G}}\Big|T_n^{(1)} - T_n^{(2)}(W)\Big| = O_P(l_n^{-1/2}n^{-1/2}w_{1/n}\log n + w_{u_n}u_n^{1/2}).$$

In the last equality we have used that $(W_t^{(1)})_{t\in[0,1]} = (W(\mathbb{I}_{[0,t]}(\cdot)))_{t\in[0,1]}$ and $(W_t)_{t\geq 0} = (\int_0^t 1/\sqrt{g(s)}dW_{G(s)}^{(1)})_{t\geq 0}$ are standard Brownian motions, proving

$$W((L\phi_{t,h})(G^{-1}(\cdot))) = \int (L\phi_{t,h})(s)\sqrt{g(s)}dW_s$$

and hence (A.1).

In the next step, we shall prove that

$$\sup_{G\in\mathcal{G}}\Big|T_n^{(2)}(W)-\sup_{(t,h)\in B_n} w_h\frac{\big|\int\phi^{(m)}(\frac{s-t}{h})dW_s\big|}{\sqrt{h}}-\widetilde{w}_h\Big|=O_P\big(w_{u_n}u_n^{1/2}\big). \tag{A.3}$$

Suppose that for a family of functions $\{f_i|\ i\in I\}$ with support in $[0,1]$ we want to bound $\sup_i|\int f_i(s)dW_s|$. Assume further that the f_i are of bounded variation. Then, for all $i\in I$, there exists a function q_i with $\|q_i\|_\infty\le \mathrm{TV}(f_i)$ and a probability measure P_i with $P_i[0,1[=1$ such that $f_i(u)=\int_{[0,u]}q_i(u)P_i(du)$ for all $u\in\mathbb{R}$, provided f_i is cadlag. With probability one,

$$\sup_{i\in I}\Big|\int f_i(s)dW_s\Big|=\sup_{i\in I}\Big|\int W_s q_i(s)P_i(ds)\Big|\le\sup_{s\in[0,1]}|W_s|\ \sup_{i\in I}\mathrm{TV}(f_i).$$

Let us define $\mathcal{F}_n^{(2)}$ as the class of functions

$$w_h h^{-1/2}\Big[h^m\sqrt{g(\cdot)}L\phi_{t,h}(\cdot)-\sqrt{g(t)}\alpha_m(t)\phi^{(m)}\big(\tfrac{\cdot-t}{h}\big)\Big]$$

with $t\in[0,1],\ h\in[l_n,u_n],\ t+h\le 1,\ G\in\mathcal{G}$. By the remark above, (A.3) is proved once we have established

$$\sup_{f\in\mathcal{F}_n^{(2)}}\mathrm{TV}(f)=O_P\big(w_{u_n}u_n^{1/2}\big).$$

In order to verify this, recall that $\mathrm{TV}(fg)\le\|f\|_\infty\,\mathrm{TV}(g)+\|g\|_\infty\,\mathrm{TV}(f)$ and therefore

$$\sup_{t\in[0,1],\ h\in[l_n,u_n],\ G\in\mathcal{G}}\mathrm{TV}\left(h^{m-1}\sqrt{g(\cdot)}\sum_{l=0}^{m-1}\alpha_l(\cdot)\frac{1}{h^l}\phi^{(l)}\big(\tfrac{\cdot-t}{h}\big)\right)<\infty.$$

Since by assumption g and α_m are Lipschitz,

$$\begin{aligned}&\mathrm{TV}\left(h^{-1}\big[\sqrt{g(\cdot)}\alpha_m(\cdot)-\sqrt{g(t)}\alpha_m(t)\big]\phi^{(m)}\big(\tfrac{\cdot-t}{h}\big)\right)\\&\quad\le h^{-1}\big\|\big[\sqrt{g(\cdot)}\alpha_m(\cdot)-\sqrt{g(t)}\alpha_m(t)\big]\mathbb{I}_{[t,t+h]}(\cdot)\big\|_\infty\,\mathrm{TV}\left(\phi^{(m)}\right)\\&\qquad+h^{-1}\big\|\phi^{(m)}\big\|_\infty\,\mathrm{TV}\left(\big[\sqrt{g(\cdot)}\alpha_m(\cdot)-\sqrt{g(t)}\alpha_m(t)\big]\mathbb{I}_{[t,t+h]}(\cdot)\right)\end{aligned}$$

is finite, uniformly in $t\in[0,1],\ h\in[l_n,u_n]$. Now, with Lemma 4 (ii), (A.3) follows.

In a final step let us show that (2.8) is almost surely bounded. In order to establish the result, we use Theorem 6.1 and Remark 1 of Dümbgen and Spokoiny [10]. Moreover, the proof is similar to the one for Theorem 2.1 in [10]. We set $\mathcal{T}=\{(t,h)\in[0,1]\times(0,1]\mid t+h\le 1\}$ and $\rho((t,h),(t',h'))=(|t-t'|+|h-h'|)^{1/2}$. Further, let $X(t,h)=\int\phi^{(m)}\big(\frac{s-t}{h}\big)dW_s$ and $\sigma(t,h)=h^{1/2}$.

Since $\phi^{(m)}$ is of bounded variation and càdlàg, there exist a function q with $\|q\|_\infty \leq \mathrm{TV}(\phi^{(m)}) < \infty$ and a probability measure P with $P[0,1] = 1$ such that $\phi^{(m)}(u) = \int_{[0,u]} q(x)P(dx)$ for all $u \geq 0$. By partial integration and due to $\phi^{(m)}(1) = 0$, for all $(t,h) \in \mathcal{T}$,

$$X(t,h) = -\int_{[0,1]} W_{uh+t} q(u)\, P(du).$$

Hence, by dominated convergence, $X(t,h)$ has continuous sample paths. Obviously, for all $(t,h), (t',h') \in \mathcal{T}$,

$$\sigma^2(t,h) \leq \sigma^2(t',h') + \rho^2((t,h),(t',h')).$$

Moreover, $\mathbb{P}(X(t,h) > h\eta) \leq \exp(-\eta^2/2)$, for any $\eta > 0$. Using Lemma 6, we obtain for a universal constant $K > 0$,

$$\mathbb{P}\Big(\big|X(t,h) - X(t',h')\big| \geq \rho((t,h),(t',h'))\eta\Big) \leq 2\exp\big(-\eta^2/(2K^2)\big).$$

Finally, we can bound the entropy $\mathcal{N}((\delta u)^{1/2}, \{(t,h) \in \mathcal{T} : h \leq \delta\})$ similarly as in [10], p. 145. Therefore, application of Remark 1 in [10] shows that

$$S := \sup_{t\in[0,1],\ h>0,\ t+h\leq 1} \frac{\sqrt{\frac{1}{2}\log\frac{e}{h}}\ \big|\int \phi^{(m)}(\frac{s-t}{h})dW_s\big|}{\sqrt{h}\log\big(e\log\frac{e}{h}\big)} - \frac{\sqrt{\log(\frac{1}{h})\log(\frac{e}{h})}}{\log\big(e\log\frac{e}{h}\big)}.$$

is almost surely bounded from above. Define

$$S' := \sup_{t\in[0,1],\ h>0,\ t+h\leq 1} \frac{\sqrt{\frac{1}{2}\log\frac{\nu}{h}}\ \big|\int \phi^{(m)}(\frac{s-t}{h})dW_s\big|}{\sqrt{h}\ \log\log\frac{\nu}{h}} - \frac{\sqrt{\log(\frac{1}{h})\log(\frac{\nu}{h})}}{\log\log\frac{\nu}{h}}.$$

Note that $\log \nu/h \leq (\log\nu)(\log e/h)$. Moreover, if $e < \nu \leq e^e$,

$$\log\log\tfrac{\nu}{h} = \log\Big(\tfrac{\log\nu}{e}\log\tfrac{e^e}{h^{e/\log\nu}}\Big) \geq \log\log\nu - 1 + \log\big(e\log\tfrac{e}{h}\big).$$

This implies

$$\frac{\log\big(e\log\frac{e}{h}\big)}{\log\log\frac{\nu}{h}} \leq \frac{2}{\log\log\nu} + 2.$$

Suppose that $S' > 0$ (otherwise S' is bounded by 0). Then, $S' \lesssim S$ and hence S' is almost surely bounded. Finally,

$$\sqrt{\log\tfrac{\nu}{h}}\Big|\sqrt{\log\tfrac{1}{h}} - \sqrt{\log\tfrac{\nu}{h}}\Big| \leq \log\nu.$$

Therefore, (2.8) hold, i.e.

$$\sup_{t\in[0,1],\ h>0,\ t+h\leq 1} w_h \frac{\big|\int \phi^{(m)}(\frac{s-t}{h})dW_s\big|}{\sqrt{h}} - \widetilde{w}_h$$

is almost surely bounded.

In the last step, let us prove that $\sup_{G\in\mathcal{G}_{c,C}} |T_n - T_n^{(1)}| = O_P(\sup_{G\in\mathcal{G}} \|\widehat{g}_n - g\|_\infty \log n / \log\log n)$. For sufficiently large n, we have by assumption $\sup_{G\in\mathcal{G}_{c,C}} \widehat{g}_n \geq c/2$. Therefore using Lemma 4 (i),

$$\begin{aligned}
\sup_{G\in\mathcal{G}} \left|T_n' - T_n\right| &\leq \sup_{(t,h)\in B_n,\ G\in\mathcal{G}} \frac{\sqrt{\frac{1}{2}\log\frac{\nu}{h}}\ h^{m-1/2}\left|T_{t,h} - \mathbb{E}[T_{t,h}]\right|}{\log\log\frac{\nu}{h}\ \sqrt{g(t)}} \frac{\sup_{G\in\mathcal{G}}\left\|\widehat{g}_n - g\right\|_\infty}{\sqrt{g(t)\widehat{g}(t)}} \\
&\leq \frac{\sqrt{2}\sup_{G\in\mathcal{G}}\left\|\widehat{g}_n - g\right\|_\infty}{c} \sup_{(t,h)\in B_n,\ G\in\mathcal{G}} \frac{\sqrt{\frac{1}{2}\log\frac{\nu}{h}}\ h^{m-1/2}\left|T_{t,h} - \mathbb{E}[T_{t,h}]\right|}{\log\log\frac{\nu}{h}\ \sqrt{g(t)}} \\
&\leq \frac{\sqrt{2}\sup_{G\in\mathcal{G}}\left\|\widehat{g}_n - g\right\|_\infty}{c} \Big(T_n + \frac{\log\frac{\nu}{l_n}}{\log\log\frac{\nu}{l_n}}\Big) \\
&\leq \frac{\sqrt{2}\sup_{G\in\mathcal{G}}\left\|\widehat{g}_n - g\right\|_\infty}{c} \Big(T_n + \frac{\log n}{\log\log n}\Big).
\end{aligned}$$

Since T_n is a.s. bounded by Theorem 1, the result follows. □

Appendix B Technical results

Proof of Lemma 3. Assume that $\rho : [0,1] \to \mathbb{R}$, $|\rho(1)| < 1$ and define $\widetilde{\rho} = (\rho - \rho(1))/(1 - |\rho(1)|)$. If $\mathrm{TV}(\widetilde{\rho}) \leq 1$, then there exists $\lambda_1, \lambda_2, \ldots \in \mathbb{R}$ and $t_1, t_2, \ldots \in [0,1]$ such that $\widetilde{\rho} = \sum \lambda_i \mathbb{I}_{[0,t_i]}$ and $\sum |\lambda_i| \leq 1$. Therefore, $\rho = (1 - |\rho(1)|)\widetilde{\rho} + \rho(1)$ can be written as linear combination of indicator functions, such that the sum of the absolute values of weights is bounded by 1. Since $\mathrm{TV}(\widetilde{\rho}) \leq 1 \Leftrightarrow \mathrm{TV}(\rho) + |\rho(1)| \leq 1$, we obtain using (A.2),

$$\left\{\rho \;\middle|\; \rho : [0,1] \to \mathbb{R}, \mathrm{TV}(\rho) + |\rho(1)| \leq 1\right\} \subset \overline{\mathrm{sc}}(\mathcal{F}_0).$$

Now, we interpret

$$\left(L\phi_{t,h}\right)\left(G^{-1}(\cdot)\right) = \langle \phi_{t,h}, \alpha_{-1}\rangle + \sum_{l=0}^{m} \alpha_l(G^{-1}(\cdot))h^{-l}\phi^{(l)}\left(\tfrac{G^{-1}(\cdot)-t}{h}\right)$$

as a function on $[0,1]$. For $\gamma : [0,1] \to \mathbb{R}$ define $\widetilde{\mathrm{TV}}(\gamma) := \mathrm{TV}(\gamma) + |\gamma(1)|$. By assumption and since G^{-1} is monotone increasing, $\widetilde{\mathrm{TV}}(\alpha_l(G^{-1}(.))) \leq \mathrm{TV}(\alpha_l) + \|\alpha_l\|_\infty$. Moreover,

$$\widetilde{\mathrm{TV}}\Big(\phi^{(l)}\big(\tfrac{G^{-1}(y)-t}{h}\big)\Big) \leq \mathrm{TV}\left(\phi^{(l)}\big(\tfrac{G^{-1}(y)-t}{h}\big)\right) + \|\phi^{(l)}\|_\infty \leq \mathrm{TV}(\phi^{(l)}) + \|\phi^{(l)}\|_\infty,$$

for all $t \in [0,1]$, $h > 0$. Note that $\widetilde{\mathrm{TV}}(f+g) \leq \widetilde{\mathrm{TV}}(f) + \widetilde{\mathrm{TV}}(g)$ and $\mathrm{TV}(fg) \leq \|f\|_\infty \mathrm{TV}(g) + \|g\|_\infty \mathrm{TV}(f)$. Hence, by the estimates above and for $0 < h \leq 1$, $\widetilde{\mathrm{TV}}(h^m(L\phi_{t,h})(G^{-1}(.)))$ is bounded by a constant C, which is independent of t, h and G. □

In the next lemma, we collect two facts about w_h.

Lemma 4. *For $h \in (0,1]$ and $\nu > e$ let $w_h := \sqrt{2^{-1}\log(\nu/h)}/\log\log(\nu/h)$. Then*

(i) $h \mapsto w_h$ is strictly decreasing on $\big(0, \nu\exp(e^{-2})\big]$, and

(ii) $h \mapsto w_h h^{1/2}$ is strictly increasing on $(0,1]$.

Proof. With $x = x(h) := \log\log(\nu/h) > 0$, we have $\log w_h = -\log(2)/2 + x/2 - \log x$. Since the derivative of this w.r.t. x equals $1/2 - 1/x$ and is strictly positive for $x > 2$, we conclude that $\log w_h$ is strictly increasing in $x(h) \geq 2$, i.e. in $h \leq \nu\exp(e^{-2})$. Moreover, $\log(w_h h^{1/2}) = \log(\nu/2)/2 + x/2 - \log x - e^x/2$, and the derivative of this w.r.t. $x > 0$ equals $1/2 - 1/x - e^x/2 < 0$. Thus $w_h h^{1/2}$ is strictly increasing in $h \in (0,1]$. □

Lemma 5. *Suppose that* $\operatorname{supp} f \subset [0,\infty)$ *and let $0 \leq a \leq 1$. Then,*

$$\int_0^{1+a} |f(x) - f(x-a)| dx \leq a\,\mathrm{TV}(f)$$

and

$$\int_0^1 |f(ax) - f(x)| dx \leq (1-a)\,\mathrm{TV}(f)$$

Proof. Without loss of generality, we can assume that f is of bounded variation, i.e.$\mathrm{TV}(f) < \infty$. Hence, there exist two positive and monotone functions f_1, f_2, such that $f = f_1 - f_2$, $f_1(u) = f_2(u) = 0$ for $u < 0$, and $f_1(\infty) + f_2(\infty) = \mathrm{TV}(f)$. Set $g = f_1 + f_2$. Then g is positive and monotone as well, and

$$\int_0^{1+a} |f(x) - f(x-a)| dx \leq \int_0^{1+a} \big(g(x+a) - g(x)\big) dx \leq \int_1^{1+a} g(x) dx \leq a\,\mathrm{TV}(f).$$

In order to derive the second inequality, note that

$$\begin{aligned}\int_0^1 |f(ax) - f(x)| dx &\leq \int_0^1 \big(g(x) - g(ax)\big) dx = \int_a^1 g(x) dx + (1 - 1/a)\int_0^a g(x) dx \\ &\leq \int_a^1 g(x) dx \leq (1-a)\,\mathrm{TV}(f).\end{aligned}$$

□

Lemma 6. *Suppose that* $\operatorname{supp}\psi \subset [0,1]$ *and* $\mathrm{TV}(\psi) < \infty$. *Let $t \in [0,1], h \in (0,1], t+h \leq 1$. Then, there exists a constant K only depending on ψ, such that*

$$\Big\|\psi\big(\tfrac{\cdot - t}{h}\big) - \psi\big(\tfrac{\cdot - t'}{h'}\big)\Big\|_{L^2} \leq K\sqrt{|h-h'| + |t-t'|}.$$

Proof. Note that

$$
\begin{aligned}
&\Big\|\psi\big(\tfrac{\cdot-t}{h}\big)-\psi\big(\tfrac{\cdot-t'}{h'}\big)\Big\|_{L^2}^2\\
&\leq 2\|\psi\|_\infty\int_0^1\Big|\psi\big(\tfrac{s-t}{h}\big)-\psi\big(\tfrac{s-t'}{h'}\big)\Big|ds\\
&\leq 2\|\psi\|_\infty\int_0^1\Big|\psi\big(\tfrac{s-t}{h}\big)-\psi\big(\tfrac{s-t}{h'}\big)\Big|ds+2\|\psi\|_\infty\int_0^1\Big|\psi\big(\tfrac{s-t}{h'}\big)-\psi\big(\tfrac{s-t'}{h'}\big)\Big|ds.
\end{aligned}
$$

Without loss of generality assume $h'\leq h$. Using Lemma 5 yields

$$
\begin{aligned}
&\int_t^{t+h}\Big|\psi\big(\tfrac{s-t}{h}\big)-\psi\big(\tfrac{s-t}{h'}\big)\Big|ds\leq\big\|\psi\big\|_\infty(h-h')+\int_t^{t+h'}\Big|\psi\big(\tfrac{s-t}{h}\big)-\psi\big(\tfrac{s-t}{h'}\big)\Big|ds\\
&=\big\|\psi\big\|_\infty(h-h')+h'\int_0^1\big|\psi\big(\tfrac{h'}{h}u\big)-\psi(u)\big|du\\
&\leq\big\|\psi\big\|_\infty(h-h')+h'\big(1-\tfrac{h'}{h}\big)\operatorname{TV}(\psi)\leq\Big[\big\|\psi\big\|_\infty+\operatorname{TV}(\psi)\Big]|h-h'|.
\end{aligned}
$$

Similarly, assuming $t\leq t'$,

$$
\int_0^1\Big|\psi\big(\tfrac{s-t}{h'}\big)-\psi\big(\tfrac{s-t'}{h'}\big)\Big|ds=h'\int_0^{(t'-t)/h'+1}\big|\psi(u)-\psi\big(u-\tfrac{t'-t}{h'}\big)\big|du\leq|t'-t|\operatorname{TV}(\psi).
$$

□

Proof of Lemma 1. The proof is based on the asymptotic behavior of the maximum $\max(\xi_1,\ldots,\xi_n)$ of i.i.d. standard Gaussian random variables $\xi_1,\xi_2,\ldots$, given by

$$
\mathbb{P}\big(\max(\xi_1,\ldots,\xi_n)\leq a_n+b_nt\big)\to\exp\big(-e^{-t}\big),\quad\text{for } t\in\mathbb{R}\text{ and } n\to\infty,
$$

where

$$
b_n:=\frac{1}{\sqrt{2\log n}},\quad\text{and}\quad a_n=\sqrt{2\log n}-\frac{\log\log n+\log(4\pi)}{\sqrt{8\log n}}.
$$

Using the tail-equivalence criterion (cf. [13], Proposition 3.3.28), we obtain further

$$
\lim_{n\to\infty}\mathbb{P}\big(\max(|\xi_1|,\ldots,|\xi_n|)\leq a_n+b_n(t+\log 2)\big)=\exp\big(-e^{-t}\big),\quad\text{for } t\in\mathbb{R}.
$$

Let

$$
T_n^\circ:=\sup_{(t,h)\in B_n^\circ}w_h\frac{\big|\int\phi^{(m)}\big(\frac{s-t}{h}\big)dW_s\big|}{\sqrt{h}}-\widetilde{w}_h.
$$

Note that T_n° has the same distribution as $w_{1/K_n}\max(|\xi_1|,\ldots,|\xi_{K_n}|)-\widetilde{w}_{1/K_n}$. It is easy to show that

$$
\Big|\frac{1}{w_{1/K_n}}-\frac{\log\log K_n}{\sqrt{\frac{1}{2}\log K_n}}\Big|=O\left(\frac{\log\log K_n}{\log^{3/2}K_n}\right)
$$

and

$$\sqrt{\log \nu K_n} = \sqrt{\log K_n} + \frac{\log \nu}{2\sqrt{\log K_n}} + O\Big(\frac{1}{\log^{3/2} K_n}\Big).$$

Assume that $\eta_n \to 0$ and $\eta_n \log\log K_n \to \infty$. For sufficiently large n,

$$\begin{aligned}
&\mathbb{P}\big(T_n^\circ > -\tfrac{1}{4} + \eta_n\Big) = \mathbb{P}\Big(\max(|\xi_1|, \ldots, |\xi_{K_n}|) > \big(-\tfrac{1}{4} + \eta_n\big)/w_{1/K_n} + \sqrt{2\log \nu K_n}\Big) \\
&= \mathbb{P}\Big(\max(|\xi_1|, \ldots, |\xi_{K_n}|) > \big(-1 + 4\eta_n\big)\frac{\log\log K_n}{\sqrt{8\log K_n}} + \sqrt{2\log K_n} + \frac{\log \nu}{\sqrt{2\log K_n}} + O\big(\frac{\log\log K_n}{\log^{3/2} K_n}\big)\Big) \\
&\le \mathbb{P}\Big(\max(|\xi_1|, \ldots, |\xi_{K_n}|) > a_n + b_n 2\eta_n \log\log K_n\Big) \to 0, \quad n \to \infty.
\end{aligned}$$

Similarly, for $n \to \infty$,

$$\mathbb{P}\big(T_n^\circ \le -\tfrac{1}{4} - \eta_n\Big) \le \mathbb{P}\Big(\max(|\xi_1|, \ldots, |\xi_{K_n}|) \le a_n - b_n \eta_n \log\log K_n\Big) \to 0.$$

□

References

[1] Abramowitz, M., and Stegun, I. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Courier Dover Publications, 1964.

[2] Bickel, P., and Rosenblatt, M. On some global measures of the deviations of density function estimation. *Ann. Statist. 1* (1973), 1071–1095.

[3] Bissantz, N., Claeskens, G., Holzmann, H., and Munk, A. Testing for lack of fit in inverse regression – with applications to biophotonic imaging. *J. Royal Statist. Society Ser. B 71* (2009), 25–48.

[4] Bissantz, N., Dümbgen, L., Holzmann, H., and Munk, A. Nonparametric confidence bands in deconvolution density estimation. *J. Royal Statist. Society Ser. B 69* (2007), 483–506.

[5] Bissantz, N., and Holzmann, H. Statistical inference for inverse problems. *Inverse Problems 24* (2008), 034009.

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

[7] Davis, L., Kovac, A., and Meise, M. Nonparametric regression, confidence regions and regularization. *Ann. Statist. 37* (2009), 2597–2625.

[8] De Angelis, D., Gilks, W. R., and Day, N. E. Bayesian projection of the acquired immune deficiency syndrome epidemic. *J. Royal Statist. Society, Ser. C 47* (1998), 449–498.

[9] Delaigle, A., and Gijbels, I. Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Stat. Data Anal. 45* (2004), 249–267.

[10] Dümbgen, L., and Spokoiny, V. G. Multiscale testing of qualitative hypothesis. *Ann. Statist. 29* (2001), 124–152.

[11] Dümbgen, L., and Walther, G. Multiscale inference about a density. *Ann. Statist. 26* (2008), 1758–1785.

[12] Eddington, A. S. On a formula for correcting statistics for the effect of a known probable error of observation. *Mon. Not. R. Astrom. Soc. 73* (1914), 359–360.

[13] Embrechts, P., Klüppelberg, C., and Mikosch, T. *Modelling Extremal Events.* Springer, Berlin, 1997.

[14] Floyd, C. E., Jaszczak, R. J., Greer, K. L., and E., C. R. Deconvolution of compton scatter in SPECT. *J. Nucl. Med. 26* (1985), 403–408.

[15] Giné, E., Koltchinskii, V., and Sakhanenko, L. Kernel density estimators: convergence in distribution for weighted sup-norms. *Probab. Theory Related Fields 130* (2004), 167–198.

[16] Giné, E., Koltchinskii, V., and Zinn, J. Weighted uniform consistency of kernel density estimators. *Ann. Probab. 32* (2004), 2570–2605.

[17] Giné, E., and Nickl, R. Confidence bands in density estimation. *Ann. Statist. 38* (2010), 1122–1170.

[18] Hasminskii, R. Z. Lower bounds for the risk of nonparametric estimates of the mode. In *Contributions to Statistics, Jaroslav Hajek Memorial Volume*, J. Jurechkova, Ed. Academia Prague, 2004, 91–97.

[19] Holzmann, H., Bissantz, N., and Munk, A. Density testing in a contaminated sample. *J. Multivariate Analysis 98* (2007), 57–75.

[20] Ingster, Y., Sapatinas, T., and Suslina, I. Minimax image detection from noisy tomographic data. arXiv:1103.3442v1, Math arXiv Preprint.

[21] Jongbloed, G. Exponential deconvolution: two asymptotically equivalent estimators. *Statist. Neerlandica 52* (1998), 6–17.

[22] KACPERSKI, K., ERLANDSSON, K., BEN-HAIM, S., AND HUTTON, B. Iterative deconvolution of simultaneous ^{99m} Tc and ^{201} Tl projection data measured on a CdZnTe-based cardiac SPECT scanner. *Phys. Med. Biol. 56* (2011), 1397–1414.

[23] KOLTCHINSKII, V. Komlos-Major-Tusnady approximation for the general empirical process and Haar expansions of classes of functions. *J. Theoret. Probab. 7* (1994), 73–118.

[24] KOMLÓS, J., MAJOR, P., AND TUSNÁDY, G. An approximation of partial sums of independent rv's and the sample df. *I. Z. Wahrsch. verw. Gebiete 32* (1975), 111–131.

[25] LAURENT, B., LOUBES, J., AND MARTEAU, C. Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. arXiv:0912.2423v2, Math arXiv Preprint.

[26] LAURENT, B., LOUBES, J., AND MARTEAU, C. Testing inverse problems: a direct or indirect problem? *J. Statist. Plann. Inference 141* (2011), 1849–1861.

[27] MEISTER, A. On testing for local monotinicity in deconvolution problems. *Statist. Probab. Lett. 79* (2009), 312–319.

[28] POLLARD, H. Distribution functions containing a Gaussian factor. *Proc. Amer. Math. Soc. 4* (1951), 578–582.

[29] VAN ES, A. J., AND KOK, A. R. Simple kernel estimators for certain nonparametric deconvolution problems. *Statist. Probab. Lett. 39* (1998), 151–160.

[30] WIDDER, D. V. Necessary and sufficient conditions for the representaion of a function by a Weierstrass transform. *Trans. Amer. Math. Soc. 71* (1951), 430–439.