The Annals of Statistics
2013, Vol. 41, No. 3, 1299–1328
DOI: 10.1214/13-AOS1089

MULTISCALE METHODS FOR SHAPE CONSTRAINTS IN DECONVOLUTION: CONFIDENCE STATEMENTS FOR QUALITATIVE FEATURES[1]

BY JOHANNES SCHMIDT-HIEBER,[2] AXEL MUNK[3] AND LUTZ DÜMBGEN

Vrije Universiteit Amsterdam, Universität Göttingen and Universität Bern

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. We investigate the moderately ill-posed setting, where the Fourier transform of the error density in the deconvolution model is of polynomial decay. 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.

1. Introduction. We observe $Y = (Y_1, \ldots, Y_n)$ according to the deconvolution model

$$Y_i = X_i + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

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

Although estimation in deconvolution models has attracted a lot of attention during the last decades (cf. Fan [15], Diggle and Hall [11], Pensky and Vidakovic [35], Johnstone et al. [26], Butucea and Tsybakov [6] as well as Meister [32] for some selective references), inference about f and its qualitative features is rather less well studied. In fact, adaptive confidence bands would be desirable, but turn out to be very ambitious. First, they suffer from the bad convergence rates induced by the ill-posedness of the problem (cf. Bissantz et al. [4]), making confidence bands

Received March 2012; revised November 2012.

[1]Supported by the joint research Grant FOR 916 of the German Science Foundation (DFG) and the Swiss National Science Foundation (SNF).

[2]Supported in part by DFG postdoctoral fellowship SCHM 2807/1-1.

[3]Supported by DFG Grants CRC 755 and CRC 803.

MSC2010 subject classifications. Primary 62G10; secondary 62G15, 62G20.

Key words and phrases. Brownian motion, convexity, pseudo-differential operators, ill-posed problems, mode detection, monotonicity, multiscale statistics, shape constraints.

less attractive for applications. Second, one would need to circumvent the classical problems of honest adaptation over Hölder scales. To overcome these difficulties the aim of the paper is to derive simultaneous confidence statements for qualitative features of f.

Structural properties or shape constraints will be conveniently expressed as (pseudo)-differential inequalities of the density f, 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 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 $f' - af \lessgtr 0$.

This paper deals with the moderately ill-posed case, meaning that the Fourier transform of the blurring density f_ε decays at polynomial rate. In fact, we work under the well-known assumption of Fan [15] (cf. Assumption 2), which essentially assures that the inversion operator, mapping $g \mapsto f$, is pseudo-differential. This combines nicely with the assumption on the class of shape constraints. Our framework includes many important error distributions such as exponential, χ^2, Laplace and gamma distributed random variables. The special case $\varepsilon = 0$ (i.e., no deconvolution or direct problem) can be treated as well, of course.

1.1. *Example*: *Detecting trends in deconvolution*. To illustrate the key ideas, suppose that we are interested in detection of regions of increase and decrease of the true density in Laplace deconvolution; that is, the error density is given by $f_\varepsilon = (2\theta)^{-1}\exp(-|\cdot|/\theta)$. Let ϕ be a sufficiently smooth, nonnegative kernel function [i.e., $\int \phi(u)\,du = 1$], supported on $[0,1]$. Then, since $f = g - \theta^2 g''$ in this case, it follows by partial integration that

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

has expectation $\mathbb{E}T_{t,h} = \sqrt{n}\int_t^{t+h} \phi(\frac{s-t}{h}) f'(s)\,ds$. The construction of the multiscale test relies on the following analytic observation. Suppose that for a given pair (t,h) there is a number $d_{t,h}$ such that

$$|T_{t,h} - \mathbb{E}T_{t,h}| \le d_{t,h}. \tag{3}$$

If in addition $T_{t,h} > d_{t,h}$, then necessarily

$$\mathbb{E}T_{t,h} = \sqrt{n}\int_t^{t+h} \phi\left(\frac{s-t}{h}\right) f'(s)\,ds > 0 \tag{4}$$

and by the nonnegativity of ϕ, $f(s_1) < f(s_2)$ for some points $s_1 < s_2$ in $[t, t+h]$. On the contrary, $T_{t,h} < -d_{t,h}$ implies that there is a decrease on $[t, t+h]$.

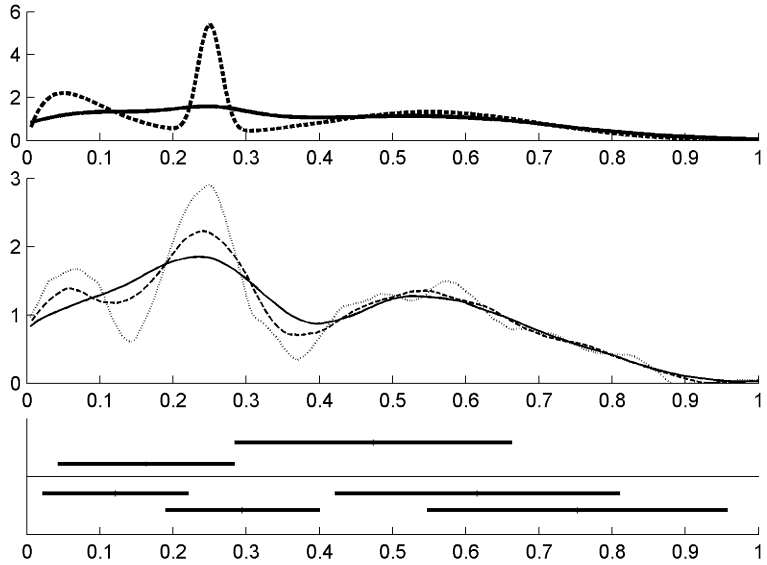

FIG. 1. *Simulation for sample size* $n = 2000$ *and* 90%*-quantile. Upper display*: *True density* f (*dashed*) *and convoluted density* g (*solid*). *Middle display*: *Kernel density estimates for* f *based on the bandwidths* $h = 0.22$ (“····”), $h = 0.31$ (“– –”) *and* $h = 0.40$ (“——”). *Lower display*: *Confidence statements. Thick horizontal lines are intervals with monotone increase/decrease* (*above/below the thin line*).

For a sequence $N_n = o(n/\log^3 n)$ tending to infinity faster than $\log^3 n$ and $u_n = 1/\log\log n$, define

$$B_n := \left\{ \left(\frac{k}{N_n}, \frac{l}{N_n} \right) \middle| k = 0, 1, \ldots, l = 1, 2, \ldots, [N_n u_n], k + l \le N_n \right\}.$$

Given $\alpha \in (0, 1)$, we will be able to compute bounds $d_{t,h}$ such that for all $(t, h) \in B_n$, inequality (3) holds simultaneously with asymptotic probability $1 - \alpha$. Taking into account that (3) implies (4), this allows us to identify regions of increase and decrease for prescribed probability.

Figure 1 shows a simulation result for $n = 2000$, $N_n = \lfloor n^{3/5} \rfloor$, $\theta = 0.075$ and confidence level 90%. The upper panel of Figure 1 displays the true density of f as well as the convoluted density g. Notice that we only have observations with density g. In fact, by visual inspection of g it becomes apparent how difficult it is to find segments on which f is monotone increasing/decreasing.

The lower panel of Figure 1 displays intervals for which we can conclude that there is a monotone increase/decrease. Let us give precise instructions on how to read this plot: Pick any of the thick horizontal lines. Then, with overall probability 90%, somewhere in this interval there is a monotone increase or decrease of f, depending on whether it is drawn above or below the thin line, respectively. In particular, the fact that intervals with monotone increase and decrease overlap does

not yield a contradiction, since the statement is that the monotonicity holds only on a nonempty subset of the corresponding interval. (The way the intervals are piled up in the plot, besides the fact that they are above or below the thin line, is arbitrary and does not contain information.) Recall that we have uniformity in the sense that with confidence 90% all these statements are true simultaneously; cf. also Dümbgen and Walther [13].

To illustrate our approach consider the middle panel in Figure 1. Here, we have displayed three reconstructions using $t \mapsto T_{t,h}/(h\sqrt{n})$ as kernel density estimator with the same unimodal kernel as for the test statistic and three different bandwidths $h \in \{0.22, 0.31, 0.40\}$. Not surprisingly (cf. Delaigle and Gijbels [10]), the reconstructions yield very different answers for what the shape of f could be. For instance, focus on the left-hand side of the graph. For $h = 0.22$ and $h = 0.31$, the density estimators have a mode at around 0.06, which is completely smoothed out under the larger bandwidth $h = 0.4$. As a practitioner, not knowing the truth, we might want to screen for modes by browsing through the plots for varying bandwidths and ask ourselves whether there is another mode or not. With the confidence statement in the lower display, we see that the true density f has to have a monotone decrease on $[0.02, 0.22]$ with confidence 90% (this is exactly the meaning of the leftmost horizontal line). This rules out the reconstruction without a mode at 0.06, since it is monotone increasing on the whole interval $[0, 0.25]$ and thus does not reflect the right shape behavior. The kernel density estimator corresponding to the smallest bandwidth $h = 0.22$ (although it is the best estimator in a pointwise sense) suggests that there could be another mode at around 0.58. However, since the confidence intervals do not support such a hypothesis, this could be merely an artifact. Combining the confidence statements in Figure 1, we conclude that with 90% confidence the true density has a local minimum and a local maximum on $[0, 1]$. Repetition of the simulation shows that often two, three or four segments of increases and decreases are detected, and at most one mode on $[0, 1]$ is found (in 69% of the cases). Therefore, sample size $n = 2000$ is not large enough to detect systematically the correct number of minima and maxima (2 and 3). Numerical simulations for larger sample size and more details are given in Section 6.

The derived confidence statements should be viewed as an additional tool for analyzing data, in particular for substantiating vague conclusions or visual impressions from point estimators.

1.2. *Pseudo-differential operators and multiscale analysis.* As mentioned at the beginning of the Introduction, we interpret shape constraints as pseudo-differential inequalities. Let $\mathcal{F}(f) = \int_{\mathbb{R}} \exp(-ix\cdot) f(x)\,dx$ always denote the Fourier transform of $f \in L^1(\mathbb{R})$ or $f \in L^2(\mathbb{R})$ (depending on the context). Consider a general class of differential operators $\mathrm{op}(p)$ with symbol p which can be written for nice f as

$$\big(\mathrm{op}(p) f\big)(x) = \frac{1}{2\pi} \int e^{ix\xi}\, p(x,\xi) \mathcal{F}(f)(\xi)\,d\xi. \tag{5}$$

This class will be an enlargement of (elliptic) pseudo-differential operators by fractional differentiation. Given data from model (1) the goal is then to identify intervals at a controlled error level on which $\operatorname{Re}(\operatorname{op}(p)f) \not\leq 0$ or $\operatorname{Re}(\operatorname{op}(p)f) \not\geq 0$. Here Re denotes the projection on the real part. In Section 1.1 we studied implicitly already the case of $\operatorname{op}(p)$ being the differentiation operator $Df = f'$ (monotonicity). If applied to $\operatorname{op}(p) = D^2$ [i.e., $p(x,\xi) = -\xi^2$], our method yields bounds for the number and confidence regions for the location of inflection points of f. We also discuss an example related to Wicksell's problem with shape constraint described by fractional differentiation.

The statistic introduced in this paper investigates shape constraints of the unknown density f on all scales simultaneously. Generalizing (4), we need to derive simultaneous confidence intervals for $\langle \phi \circ S_{t,h}, \operatorname{Re}(\operatorname{op}(p)f)\rangle$ with the scale-and-location shift $S_{t,h} = (\cdot - t)/h$ and the inner product $\langle h_1, h_2\rangle := \int_{\mathbb{R}} h_1(x)\overline{h}_2(x)\,dx$ in L^2. If $\operatorname{op}(p)^\star$ is the adjoint of $\operatorname{op}(p)$ (in a certain space) with respect to $\langle\cdot,\cdot\rangle$, then

$$\begin{aligned}
&\sqrt{n}\langle \phi \circ S_{t,h}, \operatorname{Re}\operatorname{op}(p) f\rangle \\
&\quad = \sqrt{n}\operatorname{Re}\int \big(\operatorname{op}(p)^\star(\phi \circ S_{t,h})\big)(x) f(x)\,dx \\
&\quad = \frac{\sqrt{n}}{2\pi}\operatorname{Re}\int \mathcal{F}\big(\operatorname{op}(p)^\star(\phi \circ S_{t,h})\big)(s)\overline{\mathcal{F}(f)(s)}\,ds,
\end{aligned} \tag{6}$$

and the RHS can be estimated unbiasedly by the test statistic $T_{t,h} := n^{-1/2} \times \sum_{k=1}^n \operatorname{Re} v_{t,h}(Y_k)$ with

$$v_{t,h}(u) := \frac{1}{2\pi}\int \mathcal{F}\big(\operatorname{op}(p)^\star(\phi \circ S_{t,h})\big)(s)\frac{e^{isu}}{\mathcal{F}(f_\varepsilon)(-s)}\,ds.$$

This gives rise to a multiscale statistic

$$T_n = \sup_{(t,h)} w_h\bigg(\frac{|T_{t,h} - \mathbb{E}T_{t,h}|}{\widehat{\operatorname{Std}(T_{t,h})}} - \widetilde{w}_h\bigg),$$

where w_h and $\widetilde{w}_h$ are chosen in order to calibrate the different scales with equal weight, while $\widehat{\operatorname{Std}(T_{t,h})}$ is an estimator of the standard deviation of $T_{t,h}$.

The key result in this paper is the approximation of T_n by a distribution-free statistic from which critical values can be inferred. Given the critical values, we can in a second step compute bounds $d_{t,h}$ such that a statement of type (3) holds. Following the same ideas as in Section 1.1, this is enough to identify intervals on which $\operatorname{Re}(\operatorname{op}(p)f) \not\leq 0$ or $\operatorname{Re}(\operatorname{op}(p)f) \not\geq 0$. In fact the multiscale method implies confidence statements which are stronger than the ones described up to now. These objects can be related to superpositions of confidence bands. For more precise statements, see Section 4.

1.3. *Comparison with related work and applications.* Hypothesis testing for deconvolution and related inverse problems is a relatively new area. Current methods cover testing of parametric assumptions (cf. [3, 5, 29]) and, more recently, testing for certain smoothness classes such as Sobolev balls in a Gaussian sequence model (Laurent et al. [29, 30] and Ingster et al. [25]). All of these papers focus on regression deconvolution models. Exceptions for density deconvolution are Holzmann et al. [23], Balabdaoui et al. [2] and Meister [33] who developed tests for various global hypotheses, such as global monotonicity. The latter test has been derived for one fixed interval and allows one to check whether a density is monotone on that interval at a preassigned level of significance.

Our work can also be viewed as an extension of Chaudhuri and Marron [7] as well as Dümbgen and Walther [13] who treated the case $\mathrm{op}(p) = D^m$ (with $m = 1$ in [13]) in the direct case, that is, when $\varepsilon = 0$. However, the approach in [7] does not allow for sequences of bandwidths tending to zero and yields limit distributions depending on unknown quantities again. The methods in [13] require a deterministic coupling result. The latter allows one to consider the multiscale approximation for $f = \mathbb{I}_{[0,1]}$ only, but it cannot be transferred to the deconvolution setting.

One of the main advantages of multiscale methods, making it attractive for applications, 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. Furthermore, our multiscale statistic allows us to construct estimators for the number of modes and inflection points which have a number of nice properties: First, modes and inflection points are detected with the minimax rate of convergence (up to a log-factor). Second, the probability that the true number is overestimated can be made small, since it is completely controlled by the quantile of the multiscale statistic. To state it differently, it is highly unlikely that artifacts are detected, which is a desirable property in many applications. It is worth noting that neither assumptions are made on the number of modes nor additional model selection penalties are necessary.

For practical applications, we may use these models if, for instance, the error variable ε is an independent waiting time. For example let X_i be the (unknown) time of infection of the ith patient, ε_i the corresponding incubation time, and Y_i is the time when diagnosis is made. Then, it is convenient to assume $\varepsilon \sim \Gamma(r, \theta)$; see, for instance, [9], 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. [16], Kacperski et al. [27].

The paper is organized as follows. In Section 2 we show how distribution-free approximations of multiscale statistics can be derived for general empirical processes under relatively weak conditions. For the precise statement, see Theorem 1.

These results are transferred to shape constraints and deconvolution models in Section 3. In Section 4 we discuss the statistical consequences and show how confidence statements can be derived. Theoretical questions related to the performance of the multiscale method and numerical aspects are discussed in Sections 5 and 6. Proofs and further technicalities are shifted to the Appendix and a supplementary part [37].

Notation: We write $\mathcal{T}$ for the set $[0,1] \times (0,1]$. The expression $\lfloor x \rfloor$ means the largest integer not exceeding x. The support of a function ϕ is $\operatorname{supp}\phi$, $\|\cdot\|_p$ denotes the norm in $L^p := L^p(\mathbb{R})$ and $\mathrm{TV}(\cdot)$ stands for the total variation of functions on $\mathbb{R}$. As customary in the theory of Sobolev spaces, put $\langle s\rangle := (1+|s|^2)^{1/2}$. One should not confuse this with $\langle\cdot,\cdot\rangle$, the L^2-inner product. If it is clear from the context, we write $x^k\phi$ and $\langle x\rangle^k\phi$ for the functions $x \mapsto x^k\phi(x)$ and $x \mapsto \langle x\rangle^k\phi(x)$, respectively. The ($L^2$-)Sobolev space H^r is defined as the class of functions with norm

$$\|\phi\|_{H^r} := \left(\int \langle s\rangle^{2r} \left|\mathcal{F}(\phi)(s)\right|^2 ds\right)^{1/2} < \infty.$$

For any q and $\ell \in \mathbb{N}$ ($\mathbb{N}$ is always the set of nonnegative integers) define H^q_ℓ as the Sobolev type space

$$H^q_\ell := \{\psi | x^k\psi \in H^q, \text{ for } k = 0,1,\dots,\ell\}$$

with norm $\|\psi\|_{H^q_\ell} := \sum_{k=0}^{\ell} \|x^k\psi\|_{H^q}$.

2. A general multiscale test statistic. In this section, we shall give a fairly general convergence result which is of interest on its own. The presented result does not use the deconvolution structure of model (1). It only requires that we have observations $Y_i = G^{-1}(U_i)$, $i = 1,\dots,n$ with U_i i.i.d. uniform on $[0,1]$ and G an unknown distribution function with Lebesgue density g in the class

$$\begin{aligned}\mathcal{G} := \mathcal{G}_{c,C,q} := \{&G | G \text{ is a distribution function with density} g,\\ &c \le g|_{[0,1]},\ \|g\|_\infty \le c^{-1}, \text{and} g \in \mathcal{J}(C,q)\}\end{aligned} \tag{7}$$

for fixed $c, C \ge 0$, $0 \le q < 1/2$ and the Lipschitz type constraint

$$\begin{aligned}\mathcal{J} &:= \mathcal{J}(C,q)\\ &:= \{h | |\sqrt{h(x)} - \sqrt{h(y)}| \le C(1+|x|+|y|)^q |x-y|, \text{ for all } x, y \in \mathbb{R}\}.\end{aligned}$$

For a set of real-valued functions $(\psi_{t,h})_{t,h}$ define the test statistic (empirical process) $T_{t,h} = n^{-1/2}\sum_{k=1}^n \psi_{t,h}(Y_k)$. If h is small and $\psi_{t,h}$ localized around t, then $\mathrm{Std}(T_{t,h}) \approx (\int \psi^2_{t,h}(s)g(s)\,ds)^{1/2} \approx \|\psi_{t,h}\|_2\sqrt{g(t)}$. It will turn out later on that one should allow for a slightly regularized standardization, and therefore we consider

$$\frac{|T_{t,h} - \mathbb{E}[T_{t,h}]|}{V_{t,h}\sqrt{\widehat{g}_n(t)}}$$

with $V_{t,h} \geq \|\psi_{t,h}\|_2$ and $\widehat{g}_n$ an estimator of g, satisfying

$$\sup_{G \in \mathcal{G}} \|\widehat{g}_n - g\|_\infty = O_P(1/\log n). \tag{8}$$

Unless stated otherwise, asymptotic statements refer to $n \to \infty$. We combine the single test statistics for an arbitrary subset

$$B_n \subset \{(t,h) | t \in [0,1], h \in [l_n, u_n]\} \tag{9}$$

and consider for $\nu > e$ and

$$w_h = \frac{\sqrt{1/2 \log \nu/h}}{\log\log \nu/h}, \tag{10}$$

distribution-free approximations of the multiscale statistic

$$T_n := \sup_{(t,h) \in B_n} w_h \Bigg(\frac{|T_{t,h} - \mathbb{E}[T_{t,h}]|}{V_{t,h}\sqrt{\widehat{g}_n(t)}} - \sqrt{2 \log \frac{\nu}{h}} \Bigg). \tag{11}$$

ASSUMPTION 1 (Assumption on test functions). Given a set B_n of the form (9), functions $(\psi_{t,h})_{(t,h)\in\mathcal{T}}$, and numbers $(V_{t,h})_{(t,h)\in\mathcal{T}}$, suppose that the following assumptions hold:

(i) For all $(t,h) \in \mathcal{T}$, $\|\psi_{t,h}\|_2 \leq V_{t,h}$.

(ii) We have uniform bounds on the norms

$$\sup_{(t,h)\in\mathcal{T}} \frac{\sqrt{h}\,\mathrm{TV}(\psi_{t,h}) + \sqrt{h}\|\psi_{t,h}\|_\infty + h^{-1/2}\|\psi_{t,h}\|_1}{V_{t,h}} \lesssim 1.$$

(iii) There exists $\alpha > 1/2$ such that

$$\kappa_n := \sup_{(t,h)\in B_n, G\in\mathcal{G}} w_h \frac{\mathrm{TV}(\psi_{t,h}(\cdot)[\sqrt{g(\cdot)} - \sqrt{g(t)}]\langle\cdot\rangle^\alpha)}{V_{t,h}} \to 0.$$

(iv) There exists a constant K such that for all $(t,h), (t',h') \in \mathcal{T}$,

$$\frac{\sqrt{h} \wedge \sqrt{h'}}{V_{t,h} \vee V_{t',h'}} \big[\|\psi_{t,h} - \psi_{t',h'}\|_2 + |V_{t,h} - V_{t',h'}|\big] \leq K\sqrt{|t-t'| + |h-h'|}.$$

THEOREM 1. *Given a multiscale statistic of the form* (11), *work in model* (1) *under Assumption* 1, *and suppose that* $l_n n \log^{-3} n \to \infty$ *and* $u_n = o(1)$. *If the process* $(t,h) \mapsto \sqrt{h} V_{t,h}^{-1} \int \psi_{t,h}(s)\, dW_s$ *has continuous sample paths on* $\mathcal{T}$, *then there exists a* (*two-sided*) *standard Brownian motion* W, *such that for* $\nu > e$,

$$\sup_{G \in \mathcal{G}_{c,C,q}} \Bigg| T_n - \sup_{(t,h)\in B_n} w_h \Bigg(\frac{|\int \psi_{t,h}(s)\, dW_s|}{V_{t,h}} - \sqrt{2 \log \frac{\nu}{h}} \Bigg) \Bigg| = O_P(r_n), \tag{12}$$

with

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

Moreover,

$$\sup_{(t,h)\in\mathcal{T}} w_h\Bigg(\frac{|\int \psi_{t,h}(s)\,dW_s|}{V_{t,h}} - \sqrt{2\log\frac{v}{h}}\Bigg) < \infty \qquad a.s. \tag{13}$$

Hence, the approximating statistic in (12) *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. [17], while finiteness of the approximating test statistic utilizes results of Dümbgen and Spokoiny [12]. Note that Theorem 1 can be understood as a multiscale analog of the L^∞-loss convergence for kernel estimators; cf. [4, 17–19].

To give an example, let us assume that $\psi_{t,h} = \psi(\frac{\cdot - t}{h})$ is a kernel function. By Lemmas B.12 and B.4, Assumption 1 holds for $V_{t,h} = \|\psi_{t,h}\|_2 = \sqrt{h}\|\psi\|_2$ whenever $\psi \neq 0$ on a Lebesgue measurable set, $\mathrm{TV}(\psi) < \infty$ and $\operatorname{supp}\psi \subset [0,1]$. Furthermore, by partial integration, we can easily verify that the process $(t,h) \mapsto \|\psi\|_2^{-1}\int \psi_{t,h}(s)\,dW_s$ has continuous sample paths; cf. [12], page 144.

For an application of Theorem 1 to wavelet thresholding, cf. Example C.1 in the supplementary material [37]. Let us close this section with a result on the lower bound of the approximating statistic.

Theorem 1 shows that the approximating statistic is almost surely bounded from above. On the contrary, we have the trivial lower bound

$$T_n \geq -\inf_{(t,h)\in B_n} \frac{\log v/h}{\log\log v/h},$$

which converges to $-\infty$ in general and describes the behavior of T_n, provided the cardinality of B_n is small (e.g., if B_n contains only one element). However, if B_n is sufficiently rich, T_n can be shown to be bounded from below, uniformly in n. Let us make this more precise. Assume, that for every n there exists a K_n such that $K_n \to \infty$ and

$$B_{K_n}^\circ := \Big\{\Big(\frac{i}{K_n}, \frac{1}{K_n}\Big)\Big| i = 0, \ldots, K_n - 1\Big\} \subset B_n. \tag{14}$$

Then the approximating statistic is asymptotically bounded from below by $-1/4$. This follows from Lemma C.1 in the Appendix. 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. [4, 17]) this has not been addressed so far for the approximating statistic in Theorem 1. For implementation, later on, our method relies therefore on Monte Carlo simulations.

3. Testing for shape constraints in deconvolution. We start by defining the class of differential operators in (5). However, before making this precise, let us

define pseudo-differential operators in dimension one as well as fractional integration and differentiation. Given a real m, consider S^m the class of functions $a : \mathbb{R} \times \mathbb{R} \to \mathbb{C}$ such that for all $\alpha, \beta \in \mathbb{N}$,

$$\left|\partial_x^\beta \partial_\xi^\alpha a(x,\xi)\right| \le C_{\alpha,\beta}\left(1+|\xi|\right)^{m-\alpha} \qquad \text{for all } x, \xi \in \mathbb{R}. \tag{15}$$

Then the pseudo-differential operator $\mathrm{Op}(a)$ corresponding to the symbol a can be defined on the Schwartz space of rapidly decreasing functions $\mathcal{S}$ by

$$\mathrm{Op}(a) : \mathcal{S} \to \mathcal{S},$$
$$\mathrm{Op}(a)\phi(x) := \frac{1}{2\pi} \int e^{ix\xi} a(x,\xi) \mathcal{F}(\phi)(\xi)\, d\xi.$$

It is well known that for any $s \in \mathbb{R}$, $\mathrm{Op}(a)$ can be extended to a continuous operator $\mathrm{Op}(a) : H^{m+s} \to H^s$. In order to simplify the readability, we only write Op for pseudo-differential operators and op in general for operators of the form (5). Throughout the paper, we write $\iota_s^\alpha = \exp(\alpha \pi i \,\mathrm{sign}(s)/2)$ and understand as usual $(\pm is)^\alpha = |s|^\alpha \iota_s^{\pm\alpha}$. The Gamma function evaluated at α will be denoted by $\Gamma(\alpha)$. Let us further introduce the Riemann–Liouville fractional integration operators on the real axis for $\alpha > 0$ by

$$\begin{aligned} \left(I_+^\alpha h\right)(x) &:= \frac{1}{\Gamma(\alpha)} \int_{-\infty}^x \frac{h(t)}{(x-t)^{1-\alpha}}\, dt \quad \text{and} \\ \left(I_-^\alpha h\right)(x) &:= \frac{1}{\Gamma(\alpha)} \int_x^\infty \frac{h(t)}{(t-x)^{1-\alpha}}\, dt. \end{aligned} \tag{16}$$

For $\beta \ge 0$, we define the corresponding fractional differentiation operators $(D_+^\beta h)(x) := D^n (I_+^{n-\beta} h)(x)$ and $(D_-^\beta h)(x) = (-D)^n (I_-^{n-\beta} f)(x)$, where $n = \lfloor \beta \rfloor + 1$. For any $s \in \mathbb{R}$, we can extend D_+^β and D_-^β to continuous operators from $H^{\beta+s} \to H^s$ using the identity (cf. [28], page 90)

$$\mathcal{F}\left(D_\pm^\beta h\right)(\xi) = (\pm i\xi)^\beta \mathcal{F}(h)(\xi) = \iota_\xi^{\pm\beta} |\xi|^\beta \mathcal{F}(h)(\xi). \tag{17}$$

In this paper, we consider operators $\mathrm{op}(p)$ which "factorize" into a pseudo-differential operator and a fractional differentiation in the Riemann–Liouville sense. More precisely, the symbol p is in the class

$$\underline{S}^m := \big\{(x,\xi) \mapsto p(x,\xi) = a(x,\xi)|\xi|^\gamma \iota_\xi^\mu \,|\, a \in S^{\overline{m}}, m = \overline{m} + \gamma, \\ \gamma \in \{0\} \cup [1,\infty), \mu \in \mathbb{R}\big\}.$$

Let us mention that we cannot allow for all $\gamma \ge 0$ since in our proofs it is essential that $\partial_\xi^2 p(x,\xi)$ is integrable. The results can also be formulated for finite sums of symbols, that is, $\sum_{j=1}^J p_j$ and $p_j \in \underline{S}^m$. However, for simplicity we restrict us to $J = 1$.

Throughout the remaining part of the paper, we will always assume that $\mathrm{op}(p) f$ is continuous. A closed rectangle in $\mathbb{R}^2$ parallel to the coordinate axes 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 $[a_1, a_2] \times [b_1, b_2]$.

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 $\mathrm{op}(p) f$.

(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 prespecified confidence level we can conclude that for all $i \in I$ the functions $(\mathrm{op}(p) f)|_{[t_i, t_i + h_i]}$ 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 $(\mathrm{op}(p) f)|_{[t_i, t_i + h_i]}$ has to attain negative values.

(iii) For any pair $(t, h) \in B_n$ with B_n as in (9), 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 $\mathrm{op}(p) f$ [denoted as $\mathrm{graph}(\mathrm{op}(p) f)$ in the sequel] has a nonempty intersection with every rectangle $[t, t + h] \times [b_-(t, h, \alpha), b_+(t, h, \alpha)]$.

In the following we will refer to these goals as problems (i), (ii), (ii′) and (iii), respectively. Note that (ii) follows from (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 $\mathrm{op}(p) f$, solving problem (i); cf. also Dümbgen and Walther [13]. 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 $1 - \alpha$, $\mathrm{op}(p) f$ has a zero in the interval $[\min(t, \widetilde{t}), \max(t + h, \widetilde{t} + \widetilde{h})]$. The maximal number of disjoint intervals on which we find zeros is then an estimator for the number of roots.

EXAMPLE 1. In the example in Section 1.1 we had $\mathrm{op}(p) = D$. 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/decreasing.

Instead of studying qualitative features of X directly, we might as well be interested in properties of the density of a transformed random variable $q(X)$. If X is nonnegative and $a > 0$, q could be, for instance, a (slightly regularized) log-transform $q = \log(\cdot + a)$.

EXAMPLE 2. Suppose that we want to analyze the convexity/concavity properties of $U = q(X)$, where q is a smooth 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

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

and there is a pseudo-differential operator $\mathrm{Op}(p)$ with symbol

$$p(x,\xi) = -\frac{1}{(q'(x))^2}\xi^2 - \frac{q''(x)q'(x)+2q''(x)}{(q'(x))^4}i\xi + \frac{3(q''(x))^2 - q'''(x)q'(x)}{(q'(x))^5},$$

such that $f_U''(y) = (\mathrm{op}(p)f)(q^{-1}(y))$. Therefore,

$$\mathrm{graph}\big(\mathrm{op}(p)f\big) \cap [t, t+h] \times \big[b_-(t,h,\alpha), b_+(t,h,\alpha)\big] \neq \varnothing$$

implies

$$\mathrm{graph}\big(f_U''\big) \cap \big[q(t), q(t+h)\big] \times \big[b_-(t,h,\alpha), b_+(t,h,\alpha)\big] \neq \varnothing.$$

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 $[q(t), q(t+h)]$.

EXAMPLE 3 (Noisy Wicksell problem). In the classical Wicksell problem, cross-sections of a plane with randomly distributed balls in three-dimensional space are observed. From these observations the distribution H or density $h = H'$ of the squared radii of the balls has to be estimated; cf. Groeneboom and Jongbloed [21]. Statistically speaking, we have observations $X_1, \ldots, X_n$ with density f satisfying the following relationship (cf. Golubev and Levit [20]):

$$1 - H(x) \propto \int_x^\infty \frac{f(t)}{(t-x)^{1/2}}\,dt = \Gamma\left(\frac{1}{2}\right)\big(I_-^{1/2} f\big)(x) \qquad \text{for all } x \in [0,\infty),$$

where $\propto$ means up to a positive constant and $I_-^{1/2}$ as in (16). Suppose now that we are interested in monotonicity properties of the density $h = H'$ on $[0,1]$. For $x > 0$, $-h' \lesseqgtr 0$ if and only if the fractional derivative of order $3/2$ satisfies $(D_-^{3/2} f)(x) = D^2(I_-^{1/2} f)(x) \lesseqgtr 0$. It is reasonable to assume in applications that the observations are corrupted by measurement errors, which means we only observe $Y_i = X_i + \varepsilon_i$ as in model (1). Hence we are in the framework described above and the shape constraint is given by $\mathrm{op}(p)f \lesseqgtr 0$ for $p(x,\xi) = \iota_\xi^{-3/2}|\xi|^{3/2}$.

In order to formulate our results in a proper way, let us introduce the following definitions. We say that a pseudo-differential operator $\mathrm{Op}(a)$ with $a \in S^m$ and S^m as in (15) is elliptic if there exists ξ_0 such that $|a(x,\xi)| > K|\xi|^m$ for a positive constant K and all ξ satisfying $|\xi| > |\xi_0|$. In the framework of Example 2, for instance, ellipticity holds if $\|q'\|_\infty < \infty$. It is well known that ellipticity is equivalent to a generalized invertibility of the operator. Furthermore, for an arbitrary symbol $p \in S^{\overline{m}}$, let us denote by $\mathrm{Op}(p^\star)$ the adjoint of $\mathrm{Op}(p)$ with respect to the inner product $\langle\cdot,\cdot\rangle$. This is again a pseudo-differential operator and $p^\star \in S^{\overline{m}}$. Formally, we can compute $p^\star$ by $p^\star(x,\xi) = e^{\partial_x \partial_\xi}\overline{p}(x,\xi)$, where $\overline{p}$ denotes the complex conjugate of p. Here the equality holds in the sense of asymptotic summation; for a precise statement see Theorem 18.1.7 in Hörmander [24]. Now, suppose that

we have a symbol in $\underline{S}^m$ of the form $a|\xi|^\gamma \iota_\xi^\mu = a(x,\xi)|\xi|^\gamma \iota_\xi^\mu$ with $a \in S^{\overline{m}}$ and $\overline{m} + \gamma = m$. Since for any $u, v \in H^m$,

$$\begin{aligned}\langle \operatorname{op}(a|\xi|^\gamma \iota_\xi^\mu)u, v\rangle &= \langle \operatorname{Op}(a)\operatorname{op}(|\xi|^\gamma \iota_\xi^\mu)u, v\rangle = \langle \operatorname{op}(|\xi|^\gamma \iota_\xi^\mu)u, \operatorname{Op}(a^\star)v\rangle \\ &= \langle u, \operatorname{op}(|\xi|^\gamma \iota_\xi^{-\mu})\operatorname{Op}(a^\star)v\rangle,\end{aligned} \tag{18}$$

we conclude that $\mathcal{F}(\operatorname{op}(a|\xi|^\gamma \iota_\xi^\mu)^\star \phi) = |\xi|^\gamma \iota_\xi^{-\mu} \mathcal{F}(\operatorname{Op}(a^\star)\phi)$ for all $\phi \in H^m$.

In order to formulate the assumptions and the main result, let us fix one symbol $p \in \underline{S}^m$ and one factorization $p(x,\xi) = a(x,\xi)|\xi|^\gamma \iota_\xi^\mu$ with a, γ, μ as in the definition of $\underline{S}^m$.

ASSUMPTION 2. We assume that there is a positive real number $r > 0$ and constants $0 < C_l \le C_u < \infty$ such that the characteristic function of ε is bounded from below and above by

$$C_l \langle s\rangle^{-r} \le \bigl|\mathbb{E}e^{-is\varepsilon}\bigr| = \bigl|\mathcal{F}(f_\varepsilon)(s)\bigr| \le C_u \langle s\rangle^{-r} \qquad \text{for all } s \in \mathbb{R}.$$

Moreover, suppose that the second derivative of $\mathcal{F}(f_\varepsilon)$ exists and

$$\langle s\rangle \bigl|D\mathcal{F}(f_\varepsilon)(s)\bigr| + \langle s\rangle^2 \bigl|D^2\mathcal{F}(f_\varepsilon)(s)\bigr| \le C_u \langle s\rangle^{-r} \qquad \text{for all } s \in \mathbb{R}.$$

These are the classical assumptions on the decay of the Fourier transform of the error density in the moderately ill-posed case; cf. Assumptions (G1) and (G3) in Fan [15]. Heuristically, we can think of $\mathcal{F}(f_\varepsilon)$ as an elliptic symbol in S^{-r}.

Let Re denote the projection on the real part. For sufficiently smooth ϕ consider the test statistic

$$T_{t,h} := \frac{1}{\sqrt{n}} \sum_{k=1}^n \operatorname{Re} v_{t,h}(Y_k) = \frac{1}{\sqrt{n}} \sum_{k=1}^n \operatorname{Re} v_{t,h}\bigl(G^{-1}(U_k)\bigr) \tag{19}$$

with

$$v_{t,h} = \mathcal{F}^{-1}\bigl(\lambda_\gamma^\mu(\cdot)\mathcal{F}\bigl(\operatorname{Op}(a^\star)(\phi \circ S_{t,h})\bigr)\bigr) \tag{20}$$

and

$$\lambda(s) = \lambda_\gamma^\mu(s) = \frac{|s|^\gamma \iota_s^{-\mu}}{\mathcal{F}(f_\varepsilon)(-s)}. \tag{21}$$

From (6) and (18), we find that for $f \in H^m$,

$$\mathbb{E}T_{t,h} = \sqrt{n} \int (\phi \circ S_{t,h})(x) \operatorname{Re}\bigl(\operatorname{op}(p) f\bigr)(x)\, dx.$$

Proceeding as in Section 2 we consider the multiscale statistic

$$T_n = \sup_{(t,h)\in B_n} w_h\biggl(\frac{|T_{t,h} - \mathbb{E}[T_{t,h}]|}{\sqrt{\widehat{g}_n(t)}\|v_{t,h}\|_2} - \sqrt{2\log\frac{\nu}{h}}\biggr), \tag{22}$$

that is, with the notation of (11), we set $\psi_{t,h} := \operatorname{Re} v_{t,h}$ and $V_{t,h} := \|v_{t,h}\|_2$. Define further

$$T_n^{\infty}(W) := \sup_{(t,h)\in B_n} w_h\Bigg(\frac{|\int \operatorname{Re} v_{t,h}(s)\,dW_s|}{\|v_{t,h}\|_2} - \sqrt{2\log\frac{\nu}{h}}\Bigg).$$

THEOREM 2. *Given an operator* $\operatorname{op}(p)$ *with symbol* $p \in \underline{S}^m$, *let* T_n *be as in* (22). *Work in model* (1) *under Assumption* 2. *Suppose that*:

(i) $l_n n\log^{-3} n \to \infty$ *and* $u_n = o(\log^{-3} n)$;
(ii) $\phi \in H_4^{\lfloor r+m+5/2\rfloor}$, $\operatorname{supp}\phi \subset [0,1]$, *and* $\operatorname{TV}(D^{\lfloor r+m+5/2\rfloor}\phi) < \infty$;
(iii) $\operatorname{Op}(a)$ *is elliptic*.

Then there exists a (*two-sided*) *standard Brownian motion* W, *such that for* $\nu > e$,

$$\sup_{G\in\mathcal{G}_{c,C,q}} |T_n - T_n^{\infty}(W)| = o_P(r_n), \tag{23}$$

with

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

Moreover,

$$\sup_{(t,h)\in\mathcal{T}} w_h\Bigg(\frac{|\int \operatorname{Re} v_{t,h}(s)\,dW_s|}{\|v_{t,h}\|_2} - \sqrt{2\log\frac{\nu}{h}}\Bigg) < \infty \qquad a.s. \tag{24}$$

Hence the approximating statistic $T_n^{\infty}(W)$ *is almost surely bounded from above by* (24).

One can easily show using Lemma C.1, that if B_n contains (14) and the symbol p does not depend on t, then the approximating statistic is also bounded from below. Furthermore, the case $\varepsilon = 0$ can be treated as well [we can define $\mathcal{F}(f_\varepsilon) = 1$ in this case]. In particular, our framework allows for the important case $\varepsilon = 0$ and $\operatorname{op}(p)$ the identity operator, which cannot be treated with the results from [13].

For special choices of p and f_ε the functions $(v_{t,h})_{t,h}$ have a much simpler form, which allows us to read off the ill-posedness of the problem from the index of the pseudo-differential operator associated with $v_{t,h}$. Let us shortly discuss this. Suppose Assumption 2 holds and additionally $\langle s\rangle^k |D^k\mathcal{F}(f_\varepsilon)(s)| \le C_k\langle s\rangle^{-r}$ for all $s\in\mathbb{R}$ and $k = 3, 4, \ldots$. Then $(x,\xi)\mapsto \mathcal{F}(f_\varepsilon)(-\xi)$ defines a symbol in S^{-r}. Because of the lower bound in Assumption 2, $C_l\langle\xi\rangle^{-r} \le |\mathcal{F}(f_\varepsilon)(-\xi)|$, the corresponding pseudo-differential operator is elliptic, and $(x,\xi)\mapsto 1/\mathcal{F}(f_\varepsilon)(-\xi)$ is the symbol of a parametrix and consequently an element in S^r; cf. Hörmander [24], Theorem 18.1.9. If $\phi \in H^{r+m}$ and $p \in \underline{S}^m \cap S^m$, then

$$\begin{aligned} v_{t,h}(u) &= \frac{1}{2\pi}\int \mathcal{F}\Bigg(\operatorname{Op}\Bigg(\frac{1}{\mathcal{F}(f_\varepsilon)(-\cdot)}\Bigg)\circ\operatorname{Op}(p^\star)(\phi\circ S_{t,h})\Bigg)(s)e^{isu}\,ds \\ &= \operatorname{Op}\Bigg(\frac{1}{\mathcal{F}(f_\varepsilon)(-\cdot)}\Bigg)\circ\operatorname{Op}(p^\star)(\phi\circ S_{t,h})(u). \end{aligned}$$

Pseudo-differential operators are closed under composition. More precisely, $p_j \in S^{m_j}$ for $j = 1, 2$ implies that the symbol of the composed operator is in $S^{m_1+m_2}$. Therefore, there is a symbol $\widetilde{p} \in S^{m+r}$ such that $v_{t,h} = \mathrm{Op}(\widetilde{p})(\phi \circ S_{t,h})$. Hence, for fixed h, the function $t \mapsto v_{t,h}$ can be viewed as a kernel estimator with bandwidth h. Furthermore, the problem is completely determined by the composition $\mathrm{Op}(\widetilde{p})$, and this yields a heuristic argument as to why (as it will turn out later) the ill-posedness of the detection problem $\mathrm{Re}\,\mathrm{op}(p) f \lesseqgtr 0$ in model (1) is determined by the sum $m + r$, that is,

ill-posedness of shape constraint + ill-posedness of deconvolution problem.

Suppose further that r and m are integers, and $\mathrm{Op}(p)$ is a differential operator of the form

$$\sum_{k=1}^{m} a_k(x) D^k \tag{25}$$

with smooth functions a_k $k = 1, \ldots, m$ and a_m bounded uniformly away from zero. If $1/\mathcal{F}(f_\varepsilon)(-\cdot)$ is a polynomial of degree r (which is true, e.g., if ε is Exponential, Laplace or Gamma distributed), then $\mathrm{Op}(\widetilde{p})$ is again of the form (25) but with degree $m + r$, and hence $v_{t,h}(u)$ is essentially a linear combination of derivatives of ϕ evaluated at $(u - t)/h$. However, these assumptions on the error density are far to restrictive. In the following paragraph we will show that even under more general conditions the approximating statistic has a very simple form.

Principal symbol. In order to perform our test, it is necessary to compute quantiles of the approximating statistic in Theorem 2. Since the approximating statistic has a relatively complex structure let us give conditions under which it can be simplified considerably. First, we impose a condition on the asymptotic behavior of the Fourier transform of the errors. Similar conditions have been studied by Fan [14] and Bissantz et al. [4]. Recall that for any $\alpha, a \in \mathbb{R}, s \neq 0$, $D\iota_s^\alpha |s|^a = D(is)^{a_1}(-is)^{a_2} = ai\iota_s^{\alpha-1}|s|^{a-1}$ with $a_1 = (a + \alpha)/2$ and $a_2 = (a - \alpha)/2$.

ASSUMPTION 3. Suppose that there exist $\beta_0 > 1/2$, $\rho \in [0, 4)$ and positive numbers A, C_ε such that

$$\left|A\iota_s^\rho |s|^r \mathcal{F}(f_\varepsilon)(s) - 1\right| + \left|Ar^{-1} i\iota_s^{\rho+1}|s|^{r+1} D\mathcal{F}(f_\varepsilon)(s) - 1\right| \leq C_\varepsilon \langle s\rangle^{-\beta_0} \qquad \forall s \in \mathbb{R}.$$

For instance the previous assumption holds with $A = \theta^r$ and $\rho \equiv r \bmod 4$ if f_ε is the density of a $\Gamma(r, \theta)$ distributed random variable. In this case $\mathcal{F}(f_\varepsilon)(s) = (1 + i\theta s)^{-r}$.

ASSUMPTION 4. Given $m = \{0\} \cup [1, \infty)$, suppose there exists a decomposition $p = p_P + p_R$ such that $p_R \in \underline{S}^{m'}$ for some $m' < m$, and

$$p_P(x, \xi) = a_P(x)|\xi|^m \iota_\xi^\mu \qquad \text{for all } x, \xi \in \mathbb{R},$$

with $(x, \xi) \mapsto a_P(x) \in S^0$, a_P real-valued and $|a_P(\cdot)| > 0$.

For $s \neq 0$, $\iota_s^2 = -1$. Assume that in the special case $m = 0$ we have $|\rho + \mu| \leq r$. Then, we can (and will) always choose ρ and μ in Assumptions 3 and 4 such that $\sigma = (r + m + \rho + \mu)/2$ and $\tau = (r + m - \rho - \mu)/2$ are nonnegative. The symbol p_P is called principal symbol. We will see that, together with the characteristics from the error density, it completely determines the asymptotics. The condition basically means that there is a smooth function b such that the highest order of the pseudo-differential operator coincides with $a_P(x)D^m$. Note that principal symbols are usually defined in a slightly more general sense; however Assumption 4 turns out to be appropriate for our purposes. In particular, the last assumption is verified for Examples 1–3.

In the following, we investigate the approximation of the multiscale test statistic

$$T_n^P := \sup_{(t,h)\in B_n} w_h\Bigg(\frac{h^{r+m-1/2}|T_{t,h} - \mathbb{E}[T_{t,h}]|}{\sqrt{\widehat{g}_n(t)}|Aa_P(t)|\|D_+^{r+m}\phi\|_2} - \sqrt{2\log\frac{v}{h}}\Bigg), \tag{26}$$

by

$$T_n^{P,\infty}(W) := \sup_{(t,h)\in B_n} w_h\Bigg(\frac{|\int D_+^\sigma D_-^\tau \phi((s-t)/h)\,dW_s|}{\|D_+^{r+m}\phi((\cdot - t)/h)\|_2} - \sqrt{2\log\frac{v}{h}}\Bigg).$$

THEOREM 3. *Work under Assumptions 2, 3 and 4. Suppose further, that*

(i) $l_n n \log^{-3} n \to \infty$ *and* $u_n = o(\log^{-(3\vee(m-m')^{-1})} n)$;
(ii) $\phi \in H_3^{\lfloor r+m+5/2\rfloor}$, $\operatorname{supp}\phi \subset [0, 1]$ $\mathrm{TV}(D^{\lfloor r+m+5/2\rfloor}\phi) < \infty$;
(iii) *if* $m = 0$, *assume that* $r > 1/2$ *and* $|\mu + \rho| \leq r$.

Then there exists a (two-sided) standard Brownian motion W, *such that for* $v > e$,

$$\sup_{G\in\mathcal{G}_{c,C,q}} |T_n^P - T_n^{P,\infty}(W)| = o_P(1),$$

and the approximating statistic $T_n^{P,\infty}(W)$ *is almost surely bounded from above by*

$$\sup_{(t,h)\in\mathcal{T}} w_h\Bigg(\frac{|\int D_+^\sigma D_-^\tau \phi((s-t)/h)\,dW_s|}{\|D_+^{r+m}\phi((\cdot - t)/h)\|_2} - \sqrt{2\log\frac{v}{h}}\Bigg) < \infty \qquad a.s. \tag{27}$$

4. Confidence statements.

4.1. *Confidence rectangles.* Suppose that Theorem 2 holds. The distribution of $T_n^\infty(W)$ depends only on known quantities. By ignoring the $o_P(1)$ term on the right-hand side of (23), we can therefore simulate the distribution of T_n. To formulate it differently, the distance between the $(1-\alpha)$-quantiles of T_n and $T_n^\infty(W)$ tends asymptotically to zero, although $T_n^\infty(W)$ does not need to have a weak limit. The $(1-\alpha)$-quantile of $T_n^\infty(W)$ will be denoted by $q_\alpha(T_n^\infty(W))$ in the sequel.

In order to obtain a confidence band one has to control the bias which requires a Hölder condition on $\operatorname{op}(p)f$. However, since we are more interested in a qualitative analysis, it suffices to assume that $\operatorname{op}(p)f$ is continuous [and $f \in H^m$ in order to define the scalar product of $\operatorname{op}(p)f$ properly]. Moreover, instead of a moment condition on the kernel ϕ, we require nonnegativity, that is, for the remaining part of this work, assume that $\phi \geq 0$ and $\int \phi(u)\,du = 1$. Theorem 2 implies that asymptotically with probability $1-\alpha$, for all $(t,h) \in B_n$,

$$\langle \phi_{t,h}, \operatorname{op}(p) f \rangle \in \left[\frac{T_{t,h} - d_{t,h}}{\sqrt{n}}, \frac{T_{t,h} + d_{t,h}}{\sqrt{n}} \right], \tag{28}$$

where

$$d_{t,h} := \sqrt{\widehat{g}_n(t)} \| v_{t,h} \|_2 \sqrt{2 \log \frac{\nu}{h}} \left(1 + q_\alpha\big(T_n^\infty(W)\big) \frac{\log\log \nu/h}{\log \nu/h} \right).$$

Using the continuity of $\operatorname{op}(p)f$, it follows that asymptotically with confidence $1-\alpha$, for all $(t,h) \in B_n$, the graph of $x \mapsto \operatorname{op}(p)f(x)$ has a nonempty intersection with each of the rectangles

$$[t, t+h] \times \left[\frac{T_{t,h} - d_{t,h}}{h\sqrt{n}}, \frac{T_{t,h} + d_{t,h}}{h\sqrt{n}} \right]. \tag{29}$$

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

$$b_-(t,h,\alpha) := \frac{T_{t,h} - d_{t,h}}{h\sqrt{n}}, \qquad b_+(t,h,\alpha) := \frac{T_{t,h} + d_{t,h}}{h\sqrt{n}}. \tag{30}$$

If instead Theorem 3 holds, we obtain by similar arguments that asymptotically with confidence $1-\alpha$, for all $(t,h) \in B_n$, the graph of $x \mapsto \operatorname{op}(p)f(x)$ has a nonempty intersection with each of the rectangles

$$[t, t+h] \times \left[\frac{T_{t,h} - d_{t,h}^P}{h\sqrt{n}}, \frac{T_{t,h} + d_{t,h}^P}{h\sqrt{n}} \right] \tag{31}$$

with

$$\begin{aligned} d_{t,h}^P := {} & \sqrt{\widehat{g}_n(t)} |A a_P(t)| h^{1/2-m-r} \| D_+^{r+m} \phi \|_2 \sqrt{2 \log \frac{\nu}{h}} \\ & \times \left(1 + q_\alpha\big(T_n^{P,\infty}(W)\big) \frac{\log\log \nu/h}{\log \nu/h} \right) \end{aligned} \tag{32}$$

and $q_\alpha(T_n^{P,\infty}(W))$ the $1-\alpha$-quantile of $T_n^{P,\infty}(W)$. Therefore we find a solution with

$$b_-(t,h,\alpha) := \frac{T_{t,h} - d_{t,h}^P}{h\sqrt{n}}, \qquad b_+(t,h,\alpha) := \frac{T_{t,h} + d_{t,h}^P}{h\sqrt{n}}.$$

Finally let us mention that instead of rectangles we can also cover $\mathrm{op}(p)f$ by ellipses. Note that in particular a rectangle is an ellipse with respect to the $\|\cdot\|_\infty$ vector norm on $\mathbb{R}^2$, that is, (up to translation) a set of the form $\{(x_1,x_2) : \max(a|x_1|, b|x_2|) = 1\}$ for positive a, b.

4.2. *Comparison with confidence bands.* Let us shortly comment on the relation between confidence rectangles and confidence bands, which for density deconvolution were studied by Bissantz et al. [4] and Lounici and Nickl [31]. Fix one scale $h = h_n$ and consider $B_n = [0,1] \times \{h\}$. For simplicity let us further restrict to the framework of Theorem 2. From (28), we obtain that

$$t \mapsto \left[\frac{T_{t,h} - d_{t,h}}{h\sqrt{n}}, \frac{T_{t,h} + d_{t,h}}{h\sqrt{n}}\right] \tag{33}$$

is a uniform $(1-\alpha)$-confidence band for the locally averaged function $t \mapsto \frac{1}{h}\langle \phi_{t,h}, \mathrm{op}(p)f\rangle$. Restricting to scales on which the stochastic error dominates the bias $|\mathrm{op}(p)f - \frac{1}{h}\langle \phi_{t,h}, \mathrm{op}(p)f\rangle|$ (e.g., by slightly undersmoothing) we can, inflating (33) by a small amount, easily construct asymptotic confidence bands for $\mathrm{op}(p)f$ as well. Note that Theorem 2 does not require that $s^r\mathcal{F}(f_\varepsilon)(s)$ converges to a constant, and therefore we can construct confidence bands for situations which are not covered within the framework of [4]. However, the construction of confidence bands described above will not work on scales where we oversmooth or if bias and stochastic error are of the same order. The strength of the multiscale approach lies in the fact that for confidence rectangles, all scales can be used simultaneously. This allows for another view on confidence rectangles. Figure 2 displays a band (33) computed for a large scale/bandwidth which obviously does not cover $\mathrm{op}(p)f$. Now, take a point, t_0 say, then (29) is equivalent to the existence of a point $t_0' \in [t_0, t_0+h]$ such that the confidence interval $[A, B]$ at t_0 shifted to t_0' (and denoted by $[A', B']$ in Figure 2) contains $\mathrm{op}(p)f(t_0')$. Thus, confidence rectangles also account for the uncertainty of $t \mapsto \mathrm{op}(p)f(t)$ along the t-axis.

5. Choice of kernel and theoretical properties of the multiscale statistic. In this section, we investigate the size/area of the rectangles constructed in the previous paragraphs. Recall that by (6) the expectation of the statistic $T_{t,h}$ depends in general on $\mathrm{op}(p)$. In contrast, Theorem 3 shows that the variance of $T_{t,h}$ depends asymptotically only on the principal symbol, which acts on ϕ as a differentiation operator of order $m+r$. Therefore, the $(m+r)$th derivative of ϕ appears in the approximating statistic $T_n^{P,\infty}(W)$, but no other derivative does. In fact, we shall see in this section that the scaling property of the confidence rectangles can be compared to the convergence rates appearing in estimation of the $(m+r)$th derivative of a density.

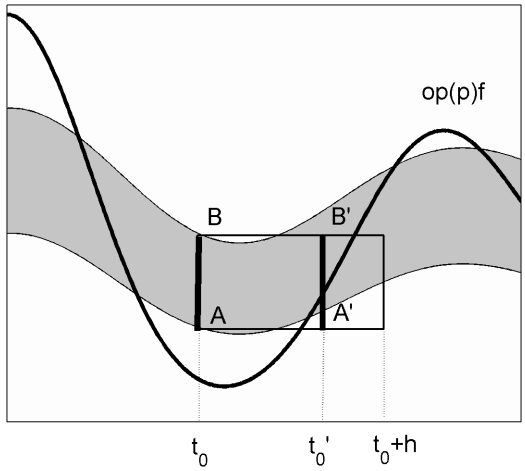

FIG. 2. *Obtaining confidence rectangles from bands.*

5.1. *Optimal choice of the kernel.* In what follows we are going to study the problem of finding an optimal function ϕ. If $m + r \in \mathbb{N}$ and the confidence statements are formulated via the conclusions of Theorem 3, this can be done explicitly.

Note that for given $(t, h) \in B_n$, the width of rectangle (31) is given by $2d^P_{t,h}/(h\sqrt{n})$. Further, the choice of ϕ influences the value of $d^P_{t,h}$ in two ways, namely by the factor $\|D^{r+m}_+\phi\|_2 = \|D^{r+m}\phi\|_2$ as well as the quantile $q_\alpha(T_n^{P,\infty}(W))$; cf. the definition of $d^P_{t,h}$ given in (32). Since α is fixed, we have

$$q_\alpha\big(T_n^{P,\infty}(W)\big)\frac{\log\log \nu/h}{\log \nu/h} = o(1).$$

Therefore, $d^P_{t,h}$ depends in first order on $\|D^{r+m}\phi\|_2$ and our optimization problem can be reformulated as

$$\text{minimize } \big\|D^{r+m}\phi\big\|_2 \qquad \text{subject to } \int \phi(u)\,du = 1.$$

This is in fact easy to solve if we additionally assume that $\phi \in H^q$ with $r + m \le q < r + m + 1/2$. By Lagrange calculus, we find that on $(0, 1)$, ϕ has to be a polynomial of order $2m + 2r$. Under the induced boundary conditions $\phi^{(k)}(0) = \phi^{(k)}(1) = 0$ for $k = 0, \dots, r + m - 1$, the solution ϕ_{m+r} is of the form

$$\phi_{m+r}(x) = c_{m+r}x^{m+r}(1 - x)^{m+r}\mathbb{I}_{(0,1)}(x). \tag{34}$$

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}, that is,

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

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

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

In the particular case of $r=0, m=1$ we obtain $\phi_1^{(1)}(x) \propto 1-2x$. This is known from the work of Dümbgen and Walther [13] who considered locally most powerful tests to derive $\phi_1^{(1)}$.

To summarize, we can find the "optimal" kernel, but it turns out that it has less smoothness than it is required by the conditions for Theorem 3 due to its behavior on the boundaries $\{0,1\}$. However, if the operator defining the shape constraint and the inversion operator $g \mapsto f$ are both differential operators (for an example see Section 1.1), then the theorems can be proved under weaker assumptions on ϕ including as a special case the optimal beta kernels.

5.2. *Theoretical properties 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). It is easy to see that $\|v_{t,h}\|_2 \lesssim h^{1/2-m-r}$, and thus $d_{t,h}$ and $d_{t,h}^P$ are of the same order. We can therefore restrict ourselves in the following to the situation, where the confidence statements are constructed based on the approximation in Theorem 2. In the other case, similar results can be derived.

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

$$\operatorname{graph}\big(\operatorname{op}(p) f\big) \cap [t, t+h] \times \left[\frac{T_{t,h} - d_{t,h}}{h\sqrt{n}}, \frac{T_{t,h} + d_{t,h}}{h\sqrt{n}}\right] \neq \varnothing.$$

The so constructed rectangles localize $\operatorname{op}(p) f$, where the amount of information is directly linked to the size of the rectangle. Therefore, it is natural to think of the length of the diagonal as a measure of localization quality. This length behaves like $h \vee h^{-m-r-1/2} n^{-1/2} \sqrt{\log 1/h}$. In particular, if the rectangle is a square, then $h \sim (\log n/n)^{1/(3+2m+2r)}$, and this coincides with the optimal bandwidth for a kernel density estimator under a Lipschitz assumption on f. This is no surprise, of course, since Lipschitz continuity allows a function to oscillate over an interval I by an amount that is proportional to the length $|I|$.

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

$$\operatorname{op}(p) f|_{[t,t+h]} \gtrsim n^{-1/2} h^{-m-r-1/2} \sqrt{\log 1/h},$$

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

In order to formulate the next theorem, recall the definition of $b_\pm(t,h,\alpha)$ given in (30). Further, set $r_{t,h,n} := 2d_{t,h}/(h\sqrt{n})$ and denote by M_n^- the set of tupels $(t,h) \in B_n$ for which $\operatorname{op}(p) f|_{[t,t+h]} > r_{t,h,n}$. Similarly define $M_n^+ := \{(t,h) \in B_n | \operatorname{op}(p) f|_{[t,t+h]} < -r_{t,h,n}\}$.

THEOREM 4. *Work under the assumptions of Theorem 2. If $\phi \geq 0$, then*

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

PROOF. For all $(t,h) \in M_n^-$, conditionally on the event given by (28),

$$\begin{aligned} \operatorname{op}(p) f|_{[t,t+h]} > r_{t,h,n} &\Rightarrow \langle \phi_{t,h}, \operatorname{op}(p) f \rangle > h r_{t,h,n}, \\ &\Rightarrow T_{t,h} > d_{t,h} \Rightarrow b_-(t,h,\alpha) > 0. \end{aligned}$$

One can argue similarly for M_n^+. □

Define

$$C_\alpha := \big(\sqrt{8\|f_\varepsilon\|_\infty} h^{m+r-1/2} \|v_{t,h}\|_2 \big(1 + q_\alpha\big(T_n^\infty(W)\big)\big)\big)^{2/(2m+2r+1)}, \tag{35}$$

and let $\widetilde{M}^\pm$ be the set of tupels $(t,h) \in B_n$ satisfying the pair of constraints

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

and

$$\operatorname{op}(p) f|_{[t,t+h]} \lessgtr \left(\frac{\log n}{n}\right)^{\beta/(2\beta+2m+2r+1)} \tag{36}$$

(with $>$ in the last equality corresponding to $\widetilde{M}_n^-$ and $<$ to $\widetilde{M}_n^+$).

COROLLARY 1. *Work under the assumptions of Theorem 2. If $\phi \geq 0$ and $\beta \in \mathbb{R}$, then*

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

PROOF. It holds that

$$d_{t,h} \leq \|f_\varepsilon\|_\infty^{1/2} \|v_{t,h}\|_2 \sqrt{2 \log \nu/h} \big(1 + q_\alpha\big(T_n^\infty(W)\big)\big).$$

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

$$r_{t,h,n} \leq \sqrt{8\|f_\varepsilon\|_\infty} \|v_{t,h}\|_2 \big(1 + q_\alpha\big(T_n^\infty(W)\big)\big) h^{-1/2} n^{-1/2} \sqrt{\log n} < \operatorname{op}(p) f|_{[t,t+h]}$$

and similarly for $\widetilde{M}_n^+$. Since $\widetilde{M}_n^\pm \subset M_n^\pm$, the result follows directly from Theorem 4. $\square$

Roughly speaking, the last result shows that if $h \sim (\log n/n)^{1/(2\beta+2m+2r+1)}$ and $\operatorname{op}(p) f|_{[t,t+h]} \sim (\log n/n)^{\beta/(2\beta+2m+2r+1)} = h^\beta$, then with probability $1-\alpha$, our method returns a rectangle in the upper half-plane. We have three distinct regimes,

$$\begin{aligned}
&\beta > 0: \quad \operatorname{op}(p) f|_{[t,t+h]} \to 0,\\
&\beta = 0: \quad \operatorname{op}(p) f|_{[t,t+h]} = O(1),\\
&-m-r-\tfrac{1}{2} < \beta < 0: \quad \operatorname{op}(p) f|_{[t,t+h]} \to \infty.
\end{aligned}$$

It is insightful to compare the previous result to derivative estimation of a density if $m+r$ is a positive integer. As it is well known, $D^{m+r} f$ can be estimated with rate of convergence

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

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

Problem (i): At the beginning of Section 3 we shortly addressed construction of confidence statements for the number of roots and their location. Note that estimators derived in this way have many interesting features. On the one hand, we know that with probability $1-\alpha$ the estimated number of roots is a lower bound for the true number of roots. Therefore, these estimates do not come from a trade-off between bias and variance, but they allow for a clear control on the probability to observe artifacts. In order to show that the lower bound for the number of roots is not trivial, we need to prove that whenever two roots are well separated (e.g., the distance between them does not shrink too quickly), they will be detected eventually by our test. This property follows if we can show that the simultaneous confidence intervals for a fixed number of roots, say, shrink to zero.

Therefore, assume for simplicity that the number K and the locations $(x_{0,j})_{j=1,\dots,K}$ of the zeros of $\operatorname{op}(p) f$ are fixed (but unknown) and $x_{0,j} \in (0,1)$ for $j = 1, \dots, K$. For example, these roots can be extreme/saddle points if $\operatorname{op}(p) = D$ or points of inflection if $\operatorname{op}(p) = D^2$.

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

$$\left\{\left(\frac{k}{N_n}, \frac{l}{N_n}\right) \middle| k = 0, 1, \dots, l = 1, 2, \dots, k+l \le N_n\right\} \subset B_n.$$

Assume further that in a neighborhood of the roots $x_{0,j}$, $\operatorname{op}(p) f$ behaves like

$$\operatorname{op}(p) f(x) = \gamma \operatorname{sign}(x - x_{0,j}) |x - x_{0,j}|^\beta + o\big(|x - x_{0,j}|^\beta\big)$$

for some positive $\beta \in (0,1]$. Let $\rho_n = (\log n/n)^{1/(2\beta+2m+2r+1)} 2/\gamma^{1/\beta}$ and $C_\alpha, M_n^\pm$ as defined in Corollary 1. There exist integer sequences $(k_{j,n}^-)_{j,n}$, $(k_{j,n}^+)_{j,n}$, $(l_n)_n$ such that for all sufficiently large n,

$$\rho_n \le \frac{k_{j,n}^-}{N_n} - x_{0,j} \le 2\rho_n, \qquad -2\rho_n \le \frac{k_{j,n}^+}{N_n} - x_{0,j} \le -\rho_n$$

and

$$C_\alpha \gamma^{1/\beta} \rho_n \le \frac{l_n}{N_n} \le 2C_\alpha \gamma^{1/\beta} \rho_n.$$

Direct calculations show $(k_{j,n}^-/N_n, l_n/N_n) \in M_n^-$ and $((k_{j,n}^+ - l_n)/N_n, l_n/N_n) \in M_n^+$ for $j = 1, \dots, K$. We can conclude from Corollary 1 and the construction that for $j = 1, \dots, K$, the confidence intervals have to be a subinterval of

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

Hence, the length for each confidence interval is bounded from above by

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

As $n \to \infty$ the confidence intervals shrink to zero and will therefore become disjoint eventually. This shows that our estimator for the number of roots picks asymptotically the correct number with high probability. Observe, that for localization of modes in density estimation $(m, r, \beta) = (1, 0, 1)$ the rate $(\log n/n)^{1/5}$ is indeed optimal up to the log-factor; cf. Hasminskii [22]. 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. [8].

For the special case of mode estimation in density deconvolution [here $(m, r, \beta) = (1, r, 1)$], let us shortly comment on related work by Rachdi and Sabre [36] and Wieczorek [39]. In [39] optimal estimation of the mode under relatively restrictive conditions on the smoothness of f is considered. In contrast, Rachdi and Sabre find the same rates of convergence $n^{-1/(2r+5)}$ (but with respect to the mean-square error). Under the stronger assumption that $D^3 f$ exists they also provide confidence bands which converge at a different rate, of course.

5.3. *On calibration of multiscale statistics.* Let us shortly comment on the type of multiscale statistic, derived in Theorems 1–3. Following [12], page 139, we can view the calibration of the multiscale statistics (11), (22) and (26) 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, that is, $h > u_n = o(1)$; cf. Theorem 1. Otherwise the approximating statistic would not be distribution-free. However, excluding the coarsest scales is a very weak restriction since the important features of $\mathrm{op}(p)f$ 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, that is, $\mathrm{op}(p)f|_I \geq C > 0$, 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.

6. Numerical simulations. In this section we provide further simulation results and discussion to the example from Section 1.1 (cf. also Example 1, Section 3), that is, studying monotonicity of the density f under Laplace-deconvolution. More precisely, the error density is $f_\varepsilon(x) = \theta^{-1}e^{-|x|/\theta}$ with $\theta = 0.075$. In this case,

$$\mathcal{F}(f_\varepsilon)(t) = \langle \theta t\rangle^{-2} \quad \text{and} \quad \mathrm{op}(p)^\star f = -Df.$$

One should notice that for Laplace deconvolution the inversion operator, mapping g to f, is given by $1 - \theta^2 D^2$ and therefore statistic (19) takes the simple form (2); cf. also the discussion following Theorem 2. The ill-posedness of the shape constraint and the deconvolution problem give $m = 1$, $r = 2$. Together with (34) it is therefore natural to choose ϕ as the density of a Beta(4, 4) random variable. Further, recall that $u_n = 1/\log\log n$, $N_n = [n^{3/5}]$ and

$$B_n = \left\{ \left(\frac{k}{N_n}, \frac{l}{N_n} \right) \middle| k = 0, 1, \ldots, l = 1, 2, \ldots, [N_n u_n], k + l \leq N_n \right\}.$$

Note that Assumptions 3 and 4 hold for $(A, \rho, r, \beta_0) = (\theta^2, 0, 2, 2)$ and $(\mu, m) = (1, 1)$, respectively. Thus, we might work in the framework of Theorem 3. The multiscale statistics

$$T_n^P = \sup_{(t,h)\in B_n} w_h\left(\frac{|T_{t,h} - \mathbb{E}T_{t,h}|}{\sqrt{\widehat{g}_n(t)}\theta^2\|\phi^{(3)}\|_2} - \sqrt{2\log\left(\frac{\nu}{h}\right)} \right)$$

and

$$T_n^{P,\infty}(W) = \sup_{(t,h)\in B_n} w_h\left(\frac{|\int \phi^{(3)}((s-t)/h)\,dW_s|}{\sqrt{h}\|\phi^{(3)}\|_2} - \sqrt{2\log\left(\frac{\nu}{h}\right)} \right) \tag{37}$$

have a particular simple form as well, and the rectangles in (31) can be computed via

$$d_{t,h}^P = h^{-5/2}\sqrt{\widehat{g}_n(t)}\theta^2\|\phi^{(3)}\|_2\sqrt{2\log\frac{\nu}{h}}\left(1 + q_{\alpha,n}\frac{\log\log \nu/h}{\log \nu/h}\right). \tag{38}$$

Boxplots for the distributions $T_{200}^{P,\infty}(W)$, $T_{1000}^{P,\infty}(W)$ and $T_{10,000}^{P,\infty}(W)$ are displayed in Figure 3 based on 10,000 repetitions each. The plot shows that the distribution is

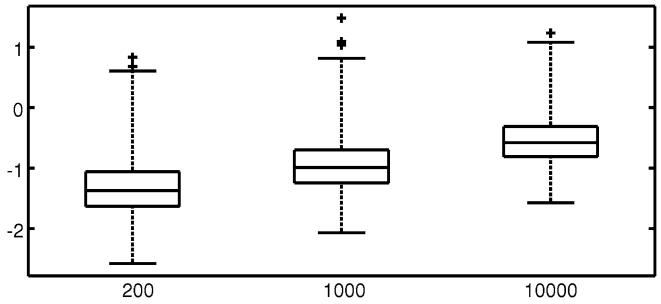

FIG. 3. *Boxplots for three different values* ($n = 200, n = 1000, n = 10{,}000$) *of the approximating statistic* (37).

well-concentrated with a few outliers only. Although our theoretical results imply boundedness of the multiscale statistic as $n \to \infty$, Figure 3 indicates that if n is in the range of a few thousand, $T_n^{P,\infty}(W)$ increases slowly.

In Section 1.1 we showed confidence statements for a simulated sample of size $n = 2000$. To complement our study, let us now investigate the case of large n, that is, $n = 10{,}000$. Again we choose the confidence level equal to 90%. The estimated quantile is $q_{0.1}(T_{10,000}^{P,\infty}(W)) = -0.04$. For all simulations, we use $\nu = \exp(e^2)$ because then $h \mapsto \sqrt{\log \nu / h}/(\log\log \nu / h)$ is monotone as long as $0 < h \le 1$; cf. Lemma B.11(i). The density f has been designed in order to investigate Corollary 1 numerically. Indeed, on $[0, 0.35]$ the signal $|f'|$ is large on average, but the intervals on which f increases/decreases are comparably small. By way of contrast, on $[0.35, 1]$ the signal $|f'|$ is small and there is only one increase/decrease.

The test is able to find all increases and decreases of f besides the increase on $[0, 0.04]$, which is not detected; cf. Figure 4. In contrast to the simulation in

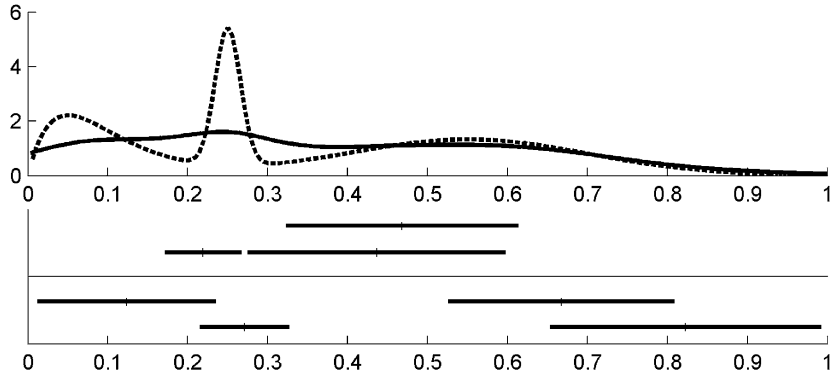

FIG. 4. *Simulation for sample size* $n = 10{,}000$ *and* 90%*-quantile. Upper display*: *True density* f (*dashed*) *and convoluted density* g (*solid*). *Lower display*: *Subset of minimal solutions to* (ii) *and* (ii$'$) (*horizontal lines above/below the thin line*).

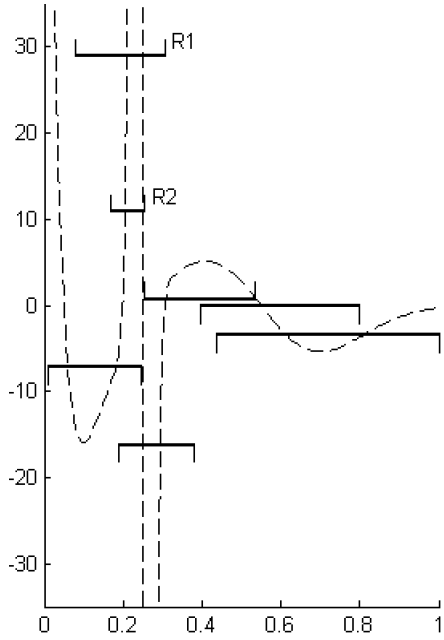

FIG. 5. *True* (*unobserved*) *derivative* f' (*dashed*) *and confidence statements for the level of* f'. *Computed for the same data set as in Figure* 4.

Figure 1, we see now a much better localization of the sharp increase/decrease on $[0.2, 0.25]$ and $[0.25, 0.3]$.

With the confidence rectangles at hand, we are able to say more about f than localizing regions of increase/decrease only. In fact, we also can provide some confidence statements about the value of f' close to a given point. Instead of plotting all confidence statements, we have displayed in Figure 5 the most prominent ones, allowing for a good characterization of the derivative f' and telling us something about the strength of the increases/decreases of f.

A bracket of type "$\sqcup$" means that f' has to be above the horizontal line, somewhere. To give an example, from the bracket $R1$ we can conclude that at least on a subset of $[0.07, 0.3]$, the derivative f' exceeds 29. Similarly, "$\sqcap$" means that somewhere f' has to be below the corresponding horizontal line. As always, these statements hold simultaneously with confidence 90%.

What we find is that in regions where the derivative does not oscillate much, we can achieve rather precise confidence statements about the value of f'. For example, from the rightmost bracket we can infer that with 90% confidence, the minimum of f' on $[0.45, 1]$ has to be below -4, coming close to the true minimum, which is approximately -6.

Figure 5 also shows nicely why a multiscale approach can provide additional insight compared to a one-scale method. Consider $R1$ and $R2$ in Figure 5 and denote by (t_1, h_1) and (t_2, h_2) the corresponding indices in B_n [as in (31)]. Note that

$R1$ and $R2$ belong to similar time points in the sense that $R2 \subset R1$ but different bandwidths h_1, h_2. Therefore we may view $R1$ and $R2$ as a superposition of confidence statements on different scales. This allows us to infer different qualitative and quantitative statements close to the same time point. We would use $R2$ in order to detect and localize an increase (as in Figure 4) or to construct a confidence band for a mode, whereas from $R1$ we obtain a better lower bound for $\sup f'$. Thus, for a qualitative analysis there is a real gain by taking into account all scales simultaneously.

7. Outlook and discussion. Given a density deconvolution model, we have investigated multiscale methods in order to analyze qualitative features of the unknown density which can be expressed as pseudo-differential operator inequalities. Compared to previous work, a more refined multiscale calibration has been considered using an idea of proof based on KMT results together with tools from the theory of pseudo-differential operators. We believe that the same strategy can be applied to a variety of other problems. In particular, it is to be expected that similar results will hold for regression and spectral density estimation.

In the formulation of the problem, but also in the proofs, it becomes apparent that modern tools from functional and harmonic analysis such as pseudo-differential operators are very helpful and to a certain extent unavoidable. In the same spirit, very recently, Nickl and Reiß [34] as well as Söhl and Trabs [38] used singular integral theory in order to prove Donsker theorems in deconvolution-type models. It is expected that reconsidering deconvolution theory from the viewpoint of harmonic analysis will lead to an improved understanding of the field.

Our multiscale approach allows us to identify intervals such that for given significance levels we know that $\operatorname{op}(p) f > 0$, at least on a subinterval. As outlined in Section 5, these results allow for qualitative inference as, for example, construction of confidence bands for the roots of $\operatorname{op}(p) f$. Since we only required that $\operatorname{op}(p) f$ is continuous, $\operatorname{op}(p) f$ 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 $\operatorname{op}(p) f$ is always positive. By adding bias controlling smoothness assumptions such as Hölder conditions, stronger results can be obtained resulting in, for instance, uniform confidence bands.

Obtaining multiscale results for error distributions as in Assumption 2 is already a very difficult topic on its own, and extension to the severely ill-posed case, including Gaussian deconvolution, becomes technically challenging since the theory of pseudo-differential operators has to the best of our knowledge not been formulated on the induced function spaces so far. Therefore we intend to treat this in a subsequent paper.

Restricting to shape constraints, which are associated with pseudo-differential operators, appears to be a limitation of our method, since important shape constraints such 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 nonlinearity is a very challenging task for further investigations. We are further aware of the fact that many other important qualitative features are related to integral transforms (that are, in general, not of convolution type), and they do not have a representation as pseudo-differential operators. 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 research.

Acknowledgments. The authors are grateful for very helpful comments by Steve Marron, Markus Reiß, Jakob Söhl, Mathias Trabs and Günther Walther as well as two referees and an Associate Editor which led to a more general version of previous results.

SUPPLEMENTARY MATERIAL

Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features (DOI: 10.1214/13-AOS1089SUPP; .pdf). All proofs can be found in the supplementary part, which contains additionally various lemmas, enumerated by $B.1, B.2, \ldots, C.1, C.2, \ldots$.

REFERENCES

[1] ABRAMOWITZ, M. and STEGUN, I. A., eds. (1992). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover Publications Inc., New York. MR1225604

[2] BALABDAOUI, F., BISSANTZ, K., BISSANTZ, N. and HOLZMANN, H. (2010). Demonstrating single and multiple currents through the *E. coli*-SecYEG-pore: Testing for the number of modes of noisy observations. *J. Amer. Statist. Assoc.* **105** 136–146. MR2757197

[3] BISSANTZ, N., CLAESKENS, G., HOLZMANN, H. and MUNK, A. (2009). Testing for lack of fit in inverse regression—with applications to biophotonic imaging. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 25–48. MR2655522

[4] BISSANTZ, N., DÜMBGEN, L., HOLZMANN, H. and MUNK, A. (2007). Non-parametric confidence bands in deconvolution density estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 483–506. MR2323764

[5] BISSANTZ, N. and HOLZMANN, H. (2008). Statistical inference for inverse problems. *Inverse Problems* **24** 034009, 17. MR2421946

[6] BUTUCEA, C. and TSYBAKOV, A. B. (2007). Sharp optimality in density deconvolution with dominating bias. I. *Theory Probab. Appl* **52** 111–128.

[7] CHAUDHURI, P. and MARRON, J. S. (2000). Scale space view of curve estimation. *Ann. Statist.* **28** 408–428. MR1790003

[8] DAVIES, P. L., KOVAC, A. and MEISE, M. (2009). Nonparametric regression, confidence regions and regularization. *Ann. Statist.* **37** 2597–2625. MR2541440

[9] DE ANGELIS, D., GILKS, W. R. and DAY, N. E. (1998). Bayesian projection of the acquired immune deficiency syndrome epidemic. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **47** 449–498.

[10] DELAIGLE, A. and GIJBELS, I. (2004). Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Statist. Data Anal.* **45** 249–267. MR2045631

[11] DIGGLE, P. J. and HALL, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **55** 523–531. MR1224414

[12] DÜMBGEN, L. and SPOKOINY, V. G. (2001). Multiscale testing of qualitative hypotheses. *Ann. Statist.* **29** 124–152. MR1833961

[13] DÜMBGEN, L. and WALTHER, G. (2008). Multiscale inference about a density. *Ann. Statist.* **36** 1758–1785. MR2435455

[14] FAN, J. (1991). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A* **53** 97–110. MR1177770

[15] FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272. MR1126324

[16] FLOYD, C. E., JASZCZAK, R. J., GREER, K. L. and COLEMAN, R. E. (1985). Deconvolution of compton scatter in SPECT. *J. Nucl. Med.* **26** 403–408.

[17] GINÉ, E., KOLTCHINSKII, V. and SAKHANENKO, L. (2004). Kernel density estimators: Convergence in distribution for weighted sup-norms. *Probab. Theory Related Fields* **130** 167–198. MR2093761

[18] GINÉ, E., KOLTCHINSKII, V. and ZINN, J. (2004). Weighted uniform consistency of kernel density estimators. *Ann. Probab.* **32** 2570–2605. MR2078551

[19] GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38** 1122–1170. MR2604707

[20] GOLUBEV, G. K. and LEVIT, B. Y. (1998). Asymptotically efficient estimation in the Wicksell problem. *Ann. Statist.* **26** 2407–2419. MR1700238

[21] GROENEBOOM, P. and JONGBLOED, G. (1995). Isotonic estimation and rates of convergence in Wicksell's problem. *Ann. Statist.* **23** 1518–1542. MR1370294

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

[23] HOLZMANN, H., BISSANTZ, N. and MUNK, A. (2007). Density testing in a contaminated sample. *J. Multivariate Anal.* **98** 57–75. MR2292917

[24] HÖRMANDER, L. (2007). *The Analysis of Linear Partial Differential Operators. III: Pseudo-Differential Operators*. Springer, Berlin. MR2304165

[25] INGSTER, Y. I., SAPATINAS, T. and SUSLINA, I. A. (2011). Minimax nonparametric testing in a problem related to the Radon transform. *Math. Methods Statist.* **20** 347–364. MR2886641

[26] JOHNSTONE, I. M., KERKYACHARIAN, G., PICARD, D. and RAIMONDO, M. (2004). Wavelet deconvolution in a periodic setting. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** 547–573. MR2088290

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

[28] KILBAS, A. A., SRIVASTAVA, H. M. and TRUJILLO, J. J. (2006). *Theory and Applications of Fractional Differential Equations*. *North-Holland Mathematics Studies* **204**. Elsevier Science B.V., Amsterdam. MR2218073

[29] LAURENT, B., LOUBES, J. M. and MARTEAU, C. (2011). Testing inverse problems: A direct or an indirect problem? *J. Statist. Plann. Inference* **141** 1849–1861. MR2763215

[30] LAURENT, B., LOUBES, J.-M. and MARTEAU, C. (2012). Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electron. J. Stat.* **6** 91–122. MR2879673

[31] LOUNICI, K. and NICKL, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist.* **39** 201–231. MR2797844

[32] MEISTER, A. (2009). *Deconvolution Problems in Nonparametric Statistics*. *Lecture Notes in Statistics* **193**. Springer, Berlin. MR2768576

[33] MEISTER, A. (2009). On testing for local monotonicity in deconvolution problems. *Statist. Probab. Lett.* **79** 312–319. MR2493014

[34] NICKL, R. and REISS, M. (2012). A Donsker theorem for Lévy measures. *J. Funct. Anal.* **263** 3306–3332. MR2973342

[35] PENSKY, M. and VIDAKOVIC, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27** 2033–2053. MR1765627

[36] RACHDI, M. and SABRE, R. (2000). Consistent estimates of the mode of the probability density function in nonparametric deconvolution problems. *Statist. Probab. Lett.* **47** 105–114. MR1747097

[37] SCHMIDT-HIEBER, J., MUNK, A. and DÜMBGEN, L. (2013). Supplement to "Multiscale methods for shape constraints in deconvolution: Confidence statements for qualitative features." DOI:10.1214/13-AOS1089SUPP.

[38] SÖHL, J. and TRABS, M. (2012). A uniform central limit theorem and efficiency for deconvolution estimators. Preprint. Available at arXiv:1208.0687v1.

[39] WIECZOREK, B. (2010). On optimal estimation of the mode in nonparametric deconvolution problems. *J. Nonparametr. Stat.* **22** 65–80. MR2598954

J. SCHMIDT-HIEBER
DEPARTMENT OF MATHEMATICS
VRIJE UNIVERSITEIT AMSTERDAM
DE BOELELAAN 1081A
1081 HV AMSTERDAM
THE NETHERLANDS
E-MAIL: j.schmidt-hieber@vu.nl

A. MUNK
INSTITUT FÜR MATHEMATISCHE STOCHASTIK
UNIVERSITÄT GÖTTINGEN
GOLDSCHMIDTSTR. 7
D-37077 GÖTTINGEN
AND
MAX-PLANCK INSTITUTE
FOR BIOPHYSICAL CHEMISTRY
AM FASSBERG 11
D-37077 GÖTTINGEN
GERMANY
E-MAIL: munk@math.uni-goettingen.de

L. DÜMBGEN
INSTITUT FÜR MATHEMATISCHE STOCHASTIK
UND VERSICHERUNGSLEHRE
UNIVERSITÄT BERN
ALPENEGGSTRASSE 22
CH-3012 BERN
SWITZERLAND
E-MAIL: lutz.duembgen@stat.unibe.ch