Multiscale Methods for Shape Constraints in Deconvolution: Confidence Statements for Qualitative Features

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

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. 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.

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

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

*Department of Mathematics, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081 HV Amsterdam, Netherlands.

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

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

[§]Institut für mathematische Stochastik und Versicherungslehre, Universität Bern, Alpeneggstrasse 22, CH-3012 Bern, Switzerland.

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 (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 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 $f' - af \lesseqgtr 0$.

Hypothesis testing for deconvolution and related inverse problems is a relatively new area. Current methods cover testing of parametric assumptions (cf. [4, 33, 6]) and, more recently, testing for certain smoothness classes such as Sobolev balls in a Gaussian sequence model (Laurent *et al.* [32, 33] and Ingster *et al.* [27]). All these papers focused on regression deconvolution models. Exceptions for density deconvolution are Holzmann *et al.* [24], Balabdaoui *et al.* [3], and Meister [36] who developed tests for various global hypotheses, such as global monotonicity based on classical Fourier inversion (see e.g. Carroll and Hall [7]). The latter test has been derived for one fixed interval and allows to check whether a density is monotone on that interval at a preassigned level of significance.

Throughout this work let $\mathcal{F}(f) = \int_{\mathbb{R}} \exp(-ix\cdot) f(x) dx$ denote the Fourier transform of $f \in L^1(\mathbb{R})$ or $f \in L^2(\mathbb{R})$ (depending on the context). As shape constraints, we consider a general class of differential operators $\operatorname{op}(p)$ with symbol p, which can be written for nice f as

$$(\operatorname{op}(p)f)(x) = \frac{1}{2\pi} \int e^{ix\xi} p(x,\xi) \mathcal{F}(f)(\xi) d\xi. \tag{1.2}$$

This class will be an enlargement of (elliptic) pseudo-differential operators by fractional differentiation. Given data from model (1.1) the goal is to identify intervals at a controlled error level on which $\operatorname{Re}(\operatorname{op}(p)f) \not\leq 0$ or $\operatorname{Re}(\operatorname{op}(p)f) \not\geq 0$, where Re denotes the projection on the real part. If applied to $\operatorname{op}(p) = D$ or D^2 (i.e. $p(x,\xi) = i\xi$ and $p(x,\xi) = -\xi^2$, respectively) with the differentiation operator $Df := f'$, our method yields bounds for the number and confidence regions for the location of modes and inflection points of f. Moreover, we discuss an example related to Wiksell's problem with shape constraint described by fractional differentiation. Our work can be viewed as an extension of Chaudhuri and Marron [8] as well as Dümbgen and Walther [13] who treated the case $\operatorname{op}(p) = D^m$ (with $m = 1$ in [13]) in the direct case, i.e. when $\epsilon = 0$. However, the approach in [8] does not allow for sequences of bandwidths tending to zero and yields limit distributions depending on the unknown quantities again. The methods in [13] require a deterministic coupling result. This allows to consider the multiscale approximation for $f = \mathbb{I}_{[0,1]}$ only but cannot be transfered to deconvolution. Thus, a new theoretical framework as well as completely different proving strategies have to be developed.

The statistic introduced in this paper investigates shape constraints of the unknown density f on all scales simultaneously. Although qualitative hypotheses such as local monotonicity seem, at a first glance, not to be expressible in terms of the Fourier transform, we can make use of the following trick: Define $S_{t,h}(\cdot) = (\cdot - t)/h$ and for a sufficiently smooth, positive kernel ϕ supported on $[0,1]$, consider the test statistic $T_{t,h} := n^{-1/2} \sum_{k=1}^n \operatorname{Re}\ v_{t,h}(Y_k)$ with

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

and $\operatorname{op}(p)^\star$ is the adjoint of $\operatorname{op}(p)$ (in a certain space) with respect to the L^2-inner product $\langle h_1, h_2 \rangle := \int_{\mathbb{R}} h_1(x)\overline{h_2}(x)\, dx$. Then, in expectation, using Parseval's identity,

$$\begin{aligned}\mathbb{E}\, T_{t,h} &= \frac{\sqrt{n}}{2\pi}\ \operatorname{Re} \int \mathcal{F}\big(\operatorname{op}(p)^\star(\phi \circ S_{t,h})\big)(s)\overline{\mathcal{F}(f)(s)} ds \\ &= \sqrt{n}\ \operatorname{Re} \int \big(\operatorname{op}(p)^\star(\phi \circ S_{t,h})\big)(x) f(x) dx = \sqrt{n}\ \langle \phi \circ S_{t,h}, \operatorname{Re} \operatorname{op}(p) f \rangle \qquad (1.3)\end{aligned}$$

for sufficiently regular functions f and ϕ. As an example, consider $\operatorname{op}(p) = D$. Then, the functions $\phi \circ S_{t,h}$ can serve as localized test functions for local monotonicity in the following sense: Whenever we know that $\langle \phi \circ S_{t,h}, f' \rangle > 0$, we may conclude that $f(s_1) < f(s_2)$ for some points $s_1 < s_2$ in $[t, t+h]$. This gives rise to a multiscale statistic

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

where w_h and $\widetilde{w}_h$ are chosen in order to calibrate the different scales with equal weight, while $\widehat{\operatorname{Std}(T_{t,h})}$ 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 that allows us to compute critical values. As mentioned before, our multiscale calibration requires new techniques in order to determine the speed of convergence between T_n and its approximation. The main tool will be a strong approximation based on Hungarian construction. This allows us on the one hand to extend the approach of [13], 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 superpositions of confidence bands. These will be denoted as confidence rectangles and allow us to identify regions where the shape constraints expressed in terms of differential inequalities, as mentioned at the beginning of this section, hold with prescribed probability. The strength of this approach lies in the fact that in contrast to sup norm bands all scales can be used simultaneously and the control of the bias becomes dispensable. For a more precise statement see Section 3.

It is a well-known fact (cf. Delaigle and Gijbels [11]) 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. Furthermore, our multiscale statistic allows to construct estimators for the number of modes and inflection points which have a number of nice properties: On one hand, modes and inflection points are detected with the minimax rate of convergence (up to a log-factor). On the other hand, the probability that the true number is overestimated is very low, and completely controlled by the quantile of the multiscale statistic. To state it differently, it is highly unlikely that artefacts will be included in the reconstruction, which is a desirable property in many applications. It is worth to note that neither assumptions are made on the number of modes nor additional model selection penalties are necessary.

This paper deals with the moderately ill-posed case, meaning that the Fourier transform of the blurring distribution decays at polynomial rate. In fact, we work under the well-known assumption of Fan [16] (cf. Assumption 2), which essentially assures that the inversion operator, mapping $g \mapsto f$, is pseudo-differential. This nicely combines 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 $\epsilon = 0$ (i.e. no deconvolution or direct problem) can be treated as well, of course.

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 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 [10], 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.* [17], Kacperski *et al.* [28]).

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 transfered 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. In particular, for a number of cases, we are even able to identify the asymptotically optimal kernel function ϕ as a beta kernel, where the degree increases with the ill-posedness of the problem. Proofs and further technicalities are shifted to the appendix and a supplementary part, which contains additionally various lemmas, enumerated by $B.1, B.2, \ldots, C.1, C.2, \ldots$

Notation: We write $\mathcal{T}$ for the set $[0,1]\times(0,1]$. $\lesssim$ and $\gtrsim$ means larger (smaller) or equal up to a constant and $\lfloor x \rfloor$ is the largest integer which is not larger than x. $\operatorname{supp}\phi$ denotes the support of ϕ. In the following, $\mathbb{N}$ is the set of non-negative integers. $\langle\cdot,\cdot\rangle$ denotes the L^2-inner product and $\|\cdot\|_p$, the L^p norm on $\mathbb{R}$. Furthermore, set $\mathrm{TV}(\cdot)$ for the total variation of functions on $\mathbb{R}$. As custom in the theory of Sobolev spaces, we define $\langle s\rangle := (1+|s|^2)^{1/2}$. If it is clear from the context, we write $x^k\phi$ to denote the function $x\mapsto x^k\phi(x)$ and similar $\langle x\rangle^k\phi$ for the function $x\mapsto\langle x\rangle^k\phi(x)$. The Sobolev space H^r is defined as the class of functions with norm

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

For any q and $\ell\in\mathbb{N}$, define H^q_ℓ as the following Sobolev type space

$$H^q_\ell := \big\{\psi \mid x^k\psi\in H^q, \text{ for } k=0,1,\ldots,\ell\ \big\}.$$

The norm on H^q_ℓ is given by $\|\psi\|_{H^q_\ell} := \sum_{k=0}^{\ell}\|x^k\psi\|_{H^q}$ for $\psi\in H^q_\ell$.

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 (1.1). 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 in the class

$$\begin{aligned}\mathcal{G} := \mathcal{G}_{c,C,q} := \big\{G \;\big|\; &G \text{ is a distribution function with density } g,\\ &c\le g\big|_{[0,1]},\ \|g\|_\infty\le c^{-1},\ \text{and } g\in\mathcal{J}(C,q)\ \big\}\end{aligned} \tag{2.1}$$

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

$$\mathcal{J} := \mathcal{J}(C,q) := \big\{h \mid |\sqrt{h(x)} - \sqrt{h(y)}| \leq C(1+|x|+|y|)^q|x-y|, \text{ for all } x,y \in \mathbb{R}\big\}.$$

For a set of real-valued functions $(\psi_{t,h})_{t,h}$ define the test statistic (empirical process) $T_{t,h} = n^{-1/2}\sum_{k=1}^n \psi_{t,h}(Y_k)$. Note that $\operatorname{Std}(T_{t,h}) = (\int \psi_{t,h}^2(s)g(s)ds)^{1/2} \approx \|\psi_{t,h}\|_2\sqrt{g(t)}$ if $\psi_{t,h}$ is localized around 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

$$\sup_{G\in\mathcal{G}} \|\widehat{g}_n - g\|_\infty = O_P(1/\log n). \tag{2.2}$$

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

$$B_n \subset \big\{(t,h)\big|\ t \in [0,1],\ h \in [l_n, u_n]\big\} \tag{2.3}$$

and consider for $\nu > e$ and

$$w_h = \frac{\sqrt{\frac{1}{2}\log\frac{\nu}{h}}}{\log\log\frac{\nu}{h}}, \tag{2.4}$$

distribution-free approximations of the multiscale statistic

$$T_n := \sup_{(t,h)\in B_n} w_h\left(\frac{\big|T_{t,h} - \mathbb{E}[T_{t,h}]\big|}{V_{t,h}\ \sqrt{\widehat{g}_n(t)}} - \sqrt{2\log\frac{\nu}{h}}\right). \tag{2.5}$$

Assumption 1 (Assumption on test functions)**.** *Given functions $(\psi_{t,h})_{(t,h)\in\mathcal{T}}$, numbers $(V_{t,h})_{(t,h)\in\mathcal{T}}$, and a set B_n of the form* (2.3)*, 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

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

(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* (2.5)*. Work in model* (1.1) *under Assumption 1 and suppose that on $\mathcal{T}$ the process $(t,h)\mapsto\sqrt{h}V_{t,h}^{-1}\int\psi_{t,h}(s)dW_s$ has continuous sample paths. Assume that $l_n n\log^{-3}n\to\infty$ and $u_n=o(1)$. Then, there exists a (two-sided) standard Brownian motion W, such that for $\nu>e$,*

$$\sup_{G\in\mathcal{G}_{c,C,q}}\Big|T_n-\sup_{(t,h)\in B_n}w_h\left(\frac{\left|\int\psi_{t,h}(s)dW_s\right|}{V_{t,h}}-\sqrt{2\log\frac{\nu}{h}}\right)\Big|=O_P(r_n), \tag{2.6}$$

with

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

Moreover,

$$\sup_{(t,h)\in\mathcal{T}}w_h\left(\frac{\left|\int\psi_{t,h}(s)dW_s\right|}{V_{t,h}}-\sqrt{2\log\frac{\nu}{h}}\right)<\infty,\quad a.s. \tag{2.7}$$

Hence, the approximating statistic in (2.6) *is almost surely bounded from above by* (2.7)*.*

The proof of the coupling in this theorem (cf. Appendix A) is based on generalizing techniques developed by Giné *et al.* [18], 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. [19, 18, 5, 20]).

To give an example, let us assume that $\psi_{t,h}=\psi(\frac{\cdot-t}{h})$ is a kernel function. By Lemmas C.2 and C.5, 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], p. 144).

Remark 1. *As a side remark let us mention that it is also possible to choose B_n in order to construct (level-dependent) values for simultaneous wavelet thresholding. To this end observe that $\widehat{d}_{j,k}=T_{k2^{-j},2^{-j}}$ and $d_{j,k}=\mathbb{E}T_{k2^{-j},2^{-j}}=\int\psi_{k2^{-j},2^{-j}}(s)g(s)ds=\int\psi(2^js-k)g(s)ds$ are the (estimated) wavelet coefficients and if j_{0n} and j_{1n} are integers satisfying $2^{-j_{1n}}n\log^{-3}n\to\infty$ and $j_{0n}\to\infty$, then, for $\alpha\in(0,1)$, and*

$$B_n=\left\{(k2^{-j},2^{-j})\big|\ k=0,1,\ldots,2^j-1,\ j_{0n}\leq j\leq j_{1n},\ j\in\mathbb{N}\ \right\},$$

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

$$\lim_{n\to\infty} \mathbb{P}\Big(\big|\widehat{d}_{j,k} - d_{j,k}\big| \le q_{j,k}(\alpha), \text{ for all } j,k, \text{ with } (k2^{-j}, 2^{-j}) \in B_n\Big) = 1-\alpha.$$

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. Note that we have the trivial lower bound

$$T_n \ge -\inf_{(t,h)\in B_n} \frac{\log \frac{\nu}{h}}{\log\log \frac{\nu}{h}},$$

which converges to $-\infty$ in general and 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, 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

$$B^{\circ}_{K_n} := \Big\{\big(\tfrac{i}{K_n}, \tfrac{1}{K_n}\big) \;\Big|\; i = 0, \dots, K_n - 1\Big\} \subset B_n. \tag{2.8}$$

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. [18, 5]) 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 (1.2). However, before we make 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 space of functions $a : \mathbb{R}\times\mathbb{R} \to \mathbb{C}$ such that for all $\alpha, \beta \in \mathbb{N}$,

$$|\partial_x^{\beta}\partial_{\xi}^{\alpha} a(x,\xi)| \le C_{\alpha,\beta}(1+|\xi|)^{m-\alpha} \quad \text{for all } x,\xi \in \mathbb{R}. \tag{3.1}$$

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

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

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 (1.2). Throughout the paper, we write $\iota_s^\alpha = \exp(\alpha\pi i \operatorname{sign}(s)/2)$ 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 and for $\alpha > 0$, by

$$\big(I_+^\alpha h\big)(x) := \frac{1}{\Gamma(\alpha)} \int_{-\infty}^x \frac{h(t)}{(x-t)^{1-\alpha}} dt \quad \text{and} \quad \big(I_-^\alpha h\big)(x) := \frac{1}{\Gamma(\alpha)} \int_x^\infty \frac{h(t)}{(t-x)^{1-\alpha}} dt. \tag{3.2}$$

For $\beta \geq 0$, 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. [29], p.90),

$$\mathcal{F}\big(D_\pm^\beta h\big)(\xi) = (\pm i\xi)^\beta \mathcal{F}\big(h\big)(\xi) = \iota_\xi^{\pm\beta} |\xi|^\beta \mathcal{F}\big(h\big)(\xi). \tag{3.3}$$

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

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

Let us mention that we cannot allow for all $\gamma \geq 0$ 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, i.e. $\sum_{j=1}^J p_j$ 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 and axes-parallel 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 $[a_1, a_2] \times [b_1, b_2]$.

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 pre-specified 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 (2.3), 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 non-empty 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 $\left[\min(t,\widetilde{t}), \max(t+h, \widetilde{t}+\widetilde{h})\right]$. The maximal number of disjoint intervals on which we find zeros is then an estimator for the number of roots.

Example 1. *Suppose* $\mathrm{op}(p) = D$*. 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.*
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* $[t,t+h] \times [ah^{-1}, bh^{-1}]$ *in at least one point.*

Example 2. *Suppose that we want to analyze the convexity/concavity properties of* $U = q(X)$ *(for instance* $U = e^X$*), where* q *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*

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

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,*

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

implies

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

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

Example 3 (Noisy Wiksell problem)**.** *In the classical Wiksell problem, cross-sections of planes 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 [22]). Statistically speaking, we have observations* $X_1, \ldots, X_n$ *with density* f *satisfying the following relationship (cf. Golubev and Levit [21])*

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

where $\propto$ *means up to a positive constant and* $I_-^{1/2}$ *as in* (3.2)*. Suppose now, that we are interested in monotonicity properties of the density* $h = H'$ *on* $[0,1]$*. For* $x > 0$, $-h' \leqslant 0$ *iff the fractional derivative of order* $3/2$ *satisfies* $(D_-^{3/2} f)(x) = D^2(I_-^{1/2} f)(x) \leqslant 0$*. It is reasonable to assume in applications that the observations are corrupted by measurement errors, which means we only observe* $Y_i = X_i + \epsilon_i$, *as in model* (1.1)*. This means we are in the framework described above and the shape constraint is given by* $\operatorname{op}(p)f \leqslant 0$ *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 $\operatorname{Op}(a)$ with $a \in S^m$ and S^m as in (3.1), is elliptic, if there exists ξ_0, such that $|a(x,\xi)| > K|\xi|^m$ for a positive constant K and all ξ satisfying $|\xi| > |\xi_0|$. For instance in the framework of Example 2, ellipticity holds if $\|q'\|_\infty < \infty$. Furthermore, for an arbitrary symbol $p \in S^{\overline{m}}$ let us denote by $\operatorname{Op}(p^\star)$ the adjoint of $\operatorname{Op}(p)$ 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 [25]). 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 $u, v \in H^m$,

$$\begin{aligned}\langle \operatorname{op}(a|\xi|^\gamma \iota_\xi^\mu) u, v \rangle &= \langle \operatorname{Op}(a) \operatorname{op}(|\xi|^\gamma \iota_\xi^\mu) u, v \rangle = \langle \operatorname{op}(|\xi|^\gamma \iota_\xi^\mu) u, \operatorname{Op}(a^\star) v \rangle \\ &= \langle u, \operatorname{op}(|\xi|^\gamma \iota_\xi^{-\mu}) \operatorname{Op}(a^\star) v \rangle \qquad (3.4)\end{aligned}$$

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* $0 < C_l \leq C_u < \infty$ *such that the characteristic function of* ϵ *is bounded from below and above by*

$$C_l \langle s \rangle^{-r} \leq |\mathbb{E}\, e^{-is\epsilon}| = |\mathcal{F}(f_\epsilon)(s)| \leq C_u \langle s \rangle^{-r} \quad \textit{for all } s \in \mathbb{R}.$$

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

$$\langle s \rangle |D\mathcal{F}(f_\epsilon)(s)| + \langle s \rangle^2 |D^2 \mathcal{F}(f_\epsilon)(s)| \leq C_u \langle s \rangle^{-r} \quad \textit{for all } s \in \mathbb{R}.$$

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 [16]). Heuristically, we can think of $\mathcal{F}(f_\epsilon)$ as an elliptic symbol in S^{-r}.

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

$$T_{t,h} := \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathrm{Re}\ \ v_{t,h}(Y_k) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \mathrm{Re}\ \ v_{t,h}(G^{-1}(U_k)) \tag{3.5}$$

with

$$v_{t,h} = \mathcal{F}^{-1}\big(\lambda_\gamma^\mu(\cdot)\mathcal{F}\big(\,\mathrm{Op}(a^\star)(\phi \circ S_{t,h})\big)\big) \tag{3.6}$$

and

$$\lambda(s) = \lambda_\gamma^\mu(s) = \frac{|s|^\gamma \iota_s^{-\mu}}{\mathcal{F}(f_\epsilon)(-s)}. \tag{3.7}$$

From (1.3) and (3.4), we find that for $f \in H^m$,

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

Proceeding as in Section 2 we consider the multiscale statistic

$$T_n = \sup_{(t,h) \in B_n} w_h \left(\frac{\big|T_{t,h} - \mathbb{E}[T_{t,h}]\big|}{\sqrt{\widehat{g}_n(t)}\ \|v_{t,h}\|_2} - \sqrt{2 \log \tfrac{\nu}{h}} \right), \tag{3.8}$$

i.e. with the notation of (2.5), we set $\psi_{t,h} := \mathrm{Re}\, v_{t,h}$ and $V_{t,h} := \|v_{t,h}\|_2$. Define further

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

Theorem 2. *Given an operator* $\mathrm{op}(p)$ *with symbol* $p \in \underline{S}^m$ *and let* T_n *be as in* (3.8). *Work in model* (1.1) *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* $\nu > e$,

$$\sup_{G \in \mathcal{G}_{c,C,q}} \Big| T_n - T_n^{\infty}(W) \Big| = o_P(r_n), \tag{3.9}$$

with

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

Moreover,

$$\sup_{(t,h) \in \mathcal{T}} w_h \left(\frac{\left| \int \operatorname{Re} v_{t,h}(s) dW_s \right|}{\|v_{t,h}\|_2} - \sqrt{2 \log \tfrac{\nu}{h}} \right) < \infty, \quad a.s. \tag{3.10}$$

Hence, the approximating statistic $T_n^{\infty}(W)$ *is almost surely bounded from above by* (3.10).

One can easily show using Lemma C.1, that if B_n contains (2.8) and the symbol p does not depend on t, then, the approximating statistic is also bounded from below. Furthermore, the case $\epsilon = 0$ can be treated as well (we can define $\mathcal{F}(f_\epsilon) = 1$ in this case). In particular, our framework allows for the important case $\epsilon = 0$ 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 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 $\langle s \rangle^k |D^k \mathcal{F}(f_\epsilon)(s)| \leq C_k \langle s \rangle^{-r}$ for all $s \in \mathbb{R}$ and $k = 3, 4, \ldots$ Then $(x, \xi) \mapsto \mathcal{F}(f_\epsilon)(-\xi)$ defines a symbol in S^{-r}. Because of the lower bound in Assumption 2, $C_l \langle \xi \rangle^{-r} \leq |\mathcal{F}(f_\epsilon)(-\xi)|$, the corresponding pseudo-differential operator is elliptic and $(x, \xi) \mapsto 1/\mathcal{F}(f_\epsilon)(-\xi)$ is the symbol of a parametrix and consequently an element in S^r (cf. Hörmander [25], Theorem 18.1.9). If $\phi \in H^{r+m}$ and $p \in \underline{S}^m \cap S^m$, then

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

Pseudo-differential operators are closed under composition. More precisely, $p_j \in S^{m_j}$, $j = 1, 2$ 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 why (as it will turn out later) the ill-posedness of the detection problem $\mathrm{Re}\,\mathrm{op}(p)f \lesseqgtr 0$ in model (1.1) is determined by the sum $m+r$, i.e.

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

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

$$\sum_{k=1}^{m} a_k(x) D^k \tag{3.11}$$

with smooth functions a_k $k = 1, \ldots, m$ and a_m bounded uniformly away from zero. If $1/\mathcal{F}(f_\epsilon)(-\cdot)$ is a polynomial of degree r (which is true for instance if ϵ is Exponential, Laplace or Gamma distributed) then $\mathrm{Op}(\widetilde{p})$ is again of the form (3.11) 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 [15] and Bissantz et *al.* [5]. Recall that for any $\alpha, a \in \mathbb{R},\ s \neq 0,\ D\iota_s^\alpha |s|^a = D(is)^{a_1}(-is)^{a_2} = a i \iota_s^{\alpha-1}|s|^{a-1}$ with $a_1 = (a+\alpha)/2$ and $a_2 = (a-\alpha)/2$.

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

$$\left|A\iota_s^\rho |s|^r \mathcal{F}(f_\epsilon)(s) - 1\right| + \left|A r^{-1} i \iota_s^{\rho+1}|s|^{r+1} D\mathcal{F}(f_\epsilon)(s) - 1\right| \le C_\epsilon \langle s \rangle^{-\beta_0}, \quad \textit{for all } s \in \mathbb{R}.$$

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*

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

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 non-negative. 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 the following, we investigate the approximation of the multiscale test statistic

$$T_n^P := \sup_{(t,h)\in B_n} w_h \left(\frac{h^{r+m-1/2} \left|T_{t,h} - \mathbb{E}[T_{t,h}]\right|}{\sqrt{\widehat{g}_n(t)}\ |Aa_P(t)|\ \|D_+^{r+m}\phi\|_2} - \sqrt{2\log\tfrac{\nu}{h}} \right), \tag{3.12}$$

by

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

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) $\phi \in H_3^{\lfloor r+m+5/2 \rfloor}$, $\operatorname{supp}\phi \subset [0,1]$, *and* $\operatorname{TV}(D^{\lfloor r+m+5/2\rfloor}\phi) < \infty$,

(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 $\nu > e$,

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

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

$$\sup_{(t,h)\in\mathcal{T}} w_h \left(\frac{\left|\int D_+^\sigma D_-^\tau \phi\left(\frac{s-t}{h}\right) dW_s\right|}{\|D_+^{r+m}\phi\left(\frac{\cdot - t}{h}\right)\|_2} - \sqrt{2\log\tfrac{\nu}{h}} \right) < \infty, \quad a.s. \tag{3.13}$$

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 (3.9), 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 positivity, i.e. for the remaining part of this work, 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 $(t,h) \in B_n$,

$$\langle \phi_{t,h}, \operatorname{op}(p)f \rangle \in \Big[\frac{T_{t,h} - d_{t,h}}{\sqrt{n}}, \frac{T_{t,h} + d_{t,h}}{\sqrt{n}}\Big], \tag{4.1}$$

where

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

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 non-empty intersection with each of the rectangles

$$\big[t, t+h\big] \times \Big[\frac{T_{t,h} - d_{t,h}}{h\sqrt{n}}, \frac{T_{t,h} + d_{t,h}}{h\sqrt{n}}\Big]. \tag{4.2}$$

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

$$b_-(t,h,\alpha) := \frac{T_{t,h} - d_{t,h}}{h\sqrt{n}}, \quad b_+(t,h,\alpha) := \frac{T_{t,h} + d_{t,h}}{h\sqrt{n}}. \tag{4.3}$$

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 non-empty intersection with each of the rectangles

$$\big[t, t+h\big] \times \Big[\frac{T_{t,h} - d_{t,h}^P}{h\sqrt{n}}, \frac{T_{t,h} + d_{t,h}^P}{h\sqrt{n}}\Big] \tag{4.4}$$

with

$$d_{t,h}^P := \sqrt{\widehat{g}_n(t)} |Aa_P(t)| h^{1/2-m-r} \big\| D_+^{r+m} \phi \big\|_2 \sqrt{2 \log \tfrac{\nu}{h}} \Big(1 + q_\alpha(T_n^{P,\infty}(W)) \frac{\log \log \frac{\nu}{h}}{\log \frac{\nu}{h}}\Big) \tag{4.5}$$

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

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

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

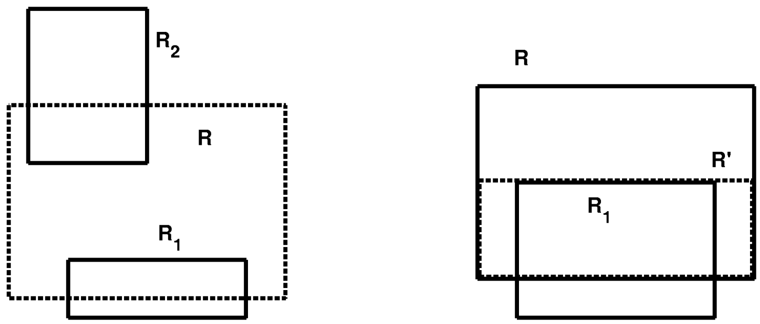

Figure 1: If the graph of $\mathrm{op}(p)f$ intersects R_1 and R_2, then also R (left). If $\mathrm{graph}(\mathrm{op}(p)f)$ intersects R and R_1, then also R' (right).

4.2 Structure on confidence rectangles

For any $(t, h) \in B_n$ the multiscale method returns a rectangle of the form (4.2) (or (4.4)). However, most of the rectangles are redundant since the fact that $\mathrm{graph}(\mathrm{op}(p)f)$ intersects these rectangles can be deduced already from the position of other rectangles (see for instance Figure 1) and the assumption that $\mathrm{op}(p)f$ is continuous. Naturally, we are interested in the set of 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 $\mathrm{op}(p)f$ is further assumed to be continuous, then by intermediate value theorem, $\mathrm{graph}(\mathrm{op}(p)f) \cap R_1 \neq \varnothing$ and $\mathrm{graph}(\mathrm{op}(p)f) \cap R_2 \neq \varnothing$ imply that $\mathrm{graph}(\mathrm{op}(p)f) \cap R \neq \varnothing$. 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 $R' := [t, t + h] \times [b_-(t, h, \alpha), b_+(t_1, h_1, \alpha)]$. Then, R' is contained in R and $\mathrm{graph}(\mathrm{op}(p)f) \cap R' \neq \varnothing$. 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 $R' := [t, t+h] \times [b_-(t_1, h_1, \alpha), b_+(t, h, \alpha)]$ we obtain $\mathrm{graph}(\mathrm{op}(p)f) \cap R' \neq \varnothing$. 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.

4.3 Comparison with confidence bands

Let us shortly comment on the relation between confidence rectangles and confidence bands. 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 (4.1), we obtain that

$$t \mapsto \Big[\frac{T_{t,h} - d_{t,h}}{h\sqrt{n}}, \frac{T_{t,h} + d_{t,h}}{h\sqrt{n}}\Big] \tag{4.6}$$

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 $|\,\mathrm{op}(p)f - \frac{1}{h}\langle \phi_{t,h}, \mathrm{op}(p)f\rangle|$ (for instance by slightly undersmoothing) we can, inflating (4.6) by a small amount, easily construct asymptotic confidence bands for $\mathrm{op}(p)f$ as well. Note that Theorem 2 does not require that $s^r \mathcal{F}(f_\epsilon)(s)$ converges to a constant and therefore we can construct confidence bands for situations which are not covered within the framework of [5]. For adaptive confidence bands in density deconvolution see the recent work by Lounici and Nickl [35]. 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. The figure on the right displays a band (4.6) computed for a large scale/bandwidth which obviously does not cover $\mathrm{op}(p)f$. Now, take a point, t_0 say, then (4.2) 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' 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.

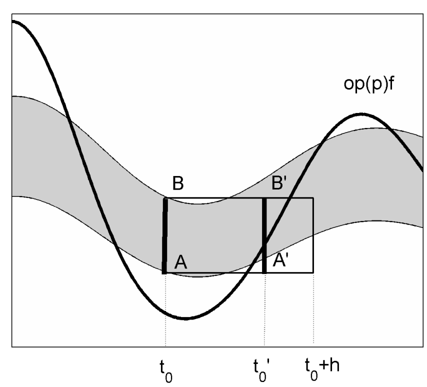

Figure: Obtaining confidence rectangles from bands.

5 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 by (1.3) the expectation of the statistic $T_{t,h}$ depends in general on 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.

5.1 Optimal choice of the kernel

In the following, we are going to study the problem of finding the optimal function ϕ. If $m+r \in \mathbb{N}$ and the confidence statements are formulated based on the conclusions of Theorem 3 this can be done explicitly.

Note that for given $(t,h) \in B_n$, the width of the rectangle (4.4) is given by $2d_{t,h}^P/(h\sqrt{n})$. Further, the choice of ϕ influences the value of $d_{t,h}^P$ in two ways, namely by the factor $\left\|D_+^{r+m}\phi\right\|_2 = \left\|D^{r+m}\phi\right\|_2$ as well as the quantile $q_\alpha(T_n^{P,\infty}(W))$ (cf. the definition of $d_{t,h}^P$ given in (4.5)). Since α is fixed, we have

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

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

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

This is in fact easy to solve if we additionally assume that $\phi \in H^q$ with $r+m \leq q < r+m+1/2$. 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 $k = 0, \ldots, r+m-1$, the solution ϕ_{m+r} has the form

$$\phi_{m+r}(x) = c_{m+r} x^{m+r}(1-x)^{m+r} \mathbb{I}_{(0,1)}(x). \tag{5.1}$$

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

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

In the particular case $r=0,\ m=1$ we obtain $\phi_1^{(1)}(x) \propto 1-2x$ and this is known from the work of Dümbgen and Walther [13], where the authors use 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 multiplicative inverse of the characteristic function of the noise density can be written as a polynomial, we were able to prove the theorems under weaker assumptions on ϕ including as a special case the optimal beta kernels.

5.2 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*). 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 $(t,h) \in B_n$,

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

The so constructed rectangles contain information on $\mathrm{op}(p)f$, where the amount of information is directly linked to the size of the rectangle. Therefore, it is natural to think of the area and the length of the diagonal as measures of localization quality. For the rectangle above, the area is given by

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

There is an interesting transition: Suppose that $m + r \leq 1$ (this includes for instance monotonicity in the direct case and exponential deconvolution). Then, $\operatorname{area}(t, h) \to 0$ uniformly in $h \in [l_n, u_n]$. In contrast, whenever $m + r > 1$,

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

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

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

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 $\operatorname{op}(p) f\big|_{[t,t+h]} < 0$.

Theorem 4. *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*

$$\operatorname{op}(p) f\big|_{[t,t+h]} > \frac{2 d_{t,h}}{h\sqrt{n}}.$$

Similar, define $M_n^+ := \{(t,h) \in B_n \mid \operatorname{op}(p) f|_{[t,t+h]} < -(2d_{t,h})/(h\sqrt{n})\}$. *Then, if* $b_+(t,h,\alpha)$ *and* $b_-(t,h,\alpha)$ *are given by (4.3), we obtain*

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

Proof. For all $(t,h) \in M_n^-$, conditionally on the event given by (4.1),

$$\operatorname{op}(p) f\big|_{[t,t+h]} > \frac{2d_{t,h}}{h\sqrt{n}} \ \Rightarrow\ \langle \phi_{t,h}, \operatorname{op}(p) f\rangle > \frac{2d_{t,h}}{\sqrt{n}} \ \Rightarrow\ T_{t,h} > d_{t,h} \ \Rightarrow\ b_-(t,h,\alpha) > 0.$$

Similar, one can argue for M_n^+. □

In order to formulate the next result, let us define

$$C_\alpha := \big(\sqrt{8\|f_\epsilon\|_\infty} h^{m+r-1/2} \|v_{t,h}\|_2 (1 + q_\alpha(T_n^\infty(W)))\big)^{2/(2m+2r+1)}. \tag{5.2}$$

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*

$$\mathrm{op}(p)f\big|_{[t,t+h]} > \left(\frac{\log n}{n}\right)^{\beta/(2\beta+2m+2r+1)} \tag{5.3}$$

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* (5.3) *replaced by* $\mathrm{op}(p)f|_{[t,t+h]} < -(\log n/n)^{\beta/(2\beta+2m+2r+1)}$. *Then, if* $b_-(t,h,\alpha)$ *and* $b_+(t,h,\alpha)$ *are given by* (4.3), *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

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

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

$$\frac{2d_{t,h}}{h\sqrt{n}} \leq \sqrt{8\left\|f_\epsilon\right\|_\infty}\left\|v_{t,h}\right\|_2\big(1+q_\alpha(T_n^\infty(W))\big)h^{-1/2}n^{-1/2}\sqrt{\log n} < \mathrm{op}(p)f\big|_{[t,t+h]}.$$

Similar for M_n^+. Now, the result follows by applying Theorem 4. □

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

$$\mathrm{op}(p)f\big|_{[t,t+h]} \gtrsim h^\beta. \tag{5.4}$$

We have three distinct regimes

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

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

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

under L^∞-risk assuming that $\mathrm{op}(p)f$ 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.

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 artefacts. It is worth mentioning that for this proper qualitative feature selection no additional penalization is required. 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 (for instance the distance between them shrinks not too fast), 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 $(x_{0,j})_{j=1,\ldots,K}$ of the zeros of $\mathrm{op}(p)f$ are fixed (but unknown) and $x_{0,j} \in (0,1)$ for $j = 1,\ldots,K$. For example, these roots can be extreme/saddle points if $\mathrm{op}(p) = D$ or points of inflection if $\mathrm{op}(p) = D^2$.

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

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

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

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

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,

$$\rho_n \leq \frac{k_{j,n}^-}{N_n} - x_{0,j} \leq 2\rho_n, \quad -2\rho_n \leq \frac{k_{j,n}^+}{N_n} - x_{0,j} \leq -\rho_n, \quad \text{and} \quad C_\alpha \gamma^{1/\beta} \rho_n \leq \frac{l_n}{N_n} \leq 2C_\alpha \gamma^{1/\beta} \rho_n.$$

Some calculations show that $(k_{j,n}^-/N_n, l_n/N_n) \in M_n^-$ and $((k_{j,n}^+ - l_n)/N_n, l_n/N_n) \in M_n^+$ for $j = 1,\ldots,K$. We can conclude from Corollary 1 and the construction, that for $j = 1,\ldots,K$, 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

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

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 [23]). 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.* [9].

For the special case of mode estimation in density deconvolution let us shortly comment on related work by Rachdi and Sabre [37] and Wieczorek [38]. In [38] 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], p.139, we can view the calibration of the multiscale statistics (2.5), (3.8), and (3.12) 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 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, i.e. $\mathrm{op}(p)f\big|_I \geq C > 0$, provided the length of the interval I is larger than $\mathrm{const.}\times(\log n/n)^{1/(2m+2r+1)}$. This can be also seen by numerical simulations, as outlined in the next section.

6 Numerical simulations

We will illustrate our method by investigating monotonicity of f ($\mathrm{op}(p) = D$, cf. Example 1) under Laplace-deconvolution, i.e. $f_\epsilon(x) = \theta^{-1} e^{-|x|/\theta}$ with $\theta = 0.075$. In this case, we

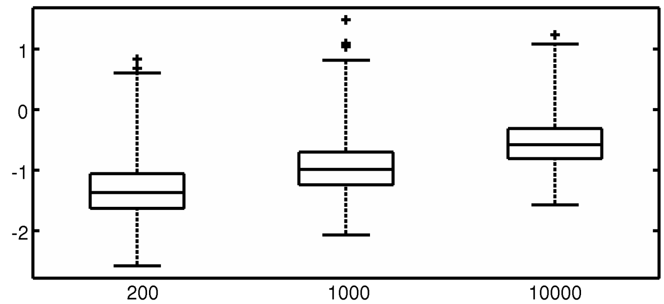

Figure 2: Boxplots for three different values ($n = 200$, $n = 1000$, $n = 10.000$) of the approximating statistic (6.1).

find

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

and the statistic (3.5) takes the explicit form

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

As kernel ϕ, we select the density of a Beta$(4,4)$ random variable (cf. Section 5). Moreover, we choose $u_n = 1/\log\log n$ for the multiscale statistic and define

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

Note that Assumptions 3 and 4 hold for $(A, \rho, r, \beta_0) = (\theta^2, 0, 2, 2)$ and $(\mu, m) = (1, 1)$, respectively. Then, the multiscale statistics

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

and

$$T_n^{P,\infty}(W) = \sup_{(t,h) \in B_n} w_h \left(\frac{\left| \int \phi^{(3)} \left(\frac{s-t}{h} \right) dW_s \right|}{\sqrt{h}\ \|\phi^{(3)}\|_2} - \sqrt{2 \log \left(\tfrac{\nu}{h} \right)} \right) \tag{6.1}$$

have a particular simple form.

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 2 for 10.000 simulations each. These plots show that the distribution is well-localized with only a few outliers. As proved, the approximating statistic is almost surely bounded as

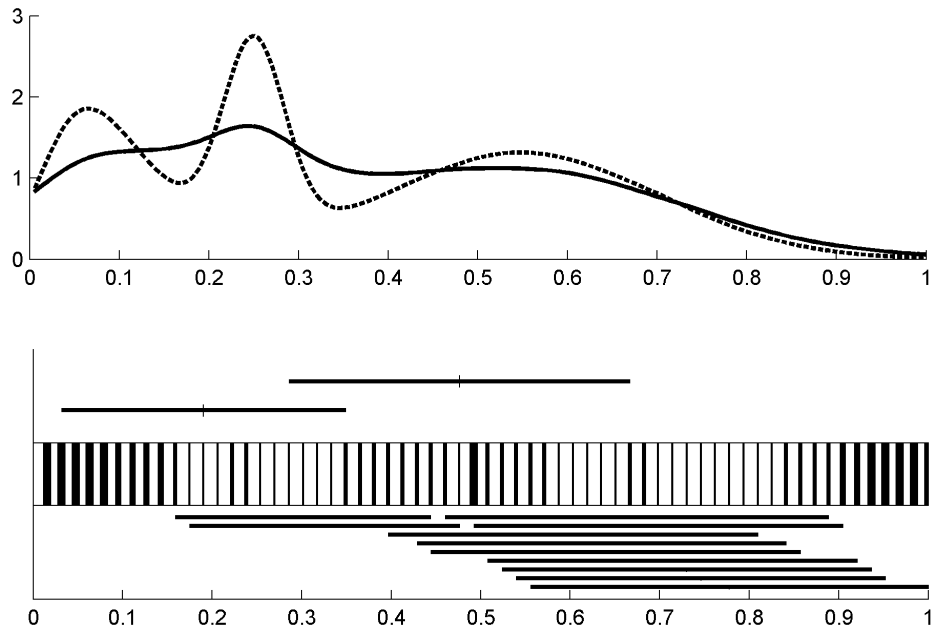

Figure 3: Simulation for sample size $n = 1000$ and 90%-quantile. *Upper display:* True density f (dashed) and convoluted density g (solid). *Lower display:* Line plot of the endpoints of intervals solving Problems (ii) and (ii') as well as minimal solutions to (ii) and (ii') (horizontal lines above/below)

$n \to \infty$. For increasing sample size, however, Figure 2 indicates, that the quantiles of the distributions $T_n^{P,\infty}(W)$ increase slowly.

In Figures 3 and 4, we give an example of a reconstruction based on a sample size of $n = 1000$ and confidence level equal to 90%. Based on 10.000 repetitions, the estimated quantile is $q_{0.1}(T_{1000}^{P,\infty}(W)) = -0.41$. For the simulation, we use $\nu = \exp(e^2)$ because then, $h \mapsto \sqrt{\log \nu/h}/(\log\log \nu/h)$ is monotone as long as $0 < h \leq 1$ (cf. Lemma C.3 (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, it becomes apparent how difficult it is to find the intervals on which f is monotone increasing/decreasing.

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

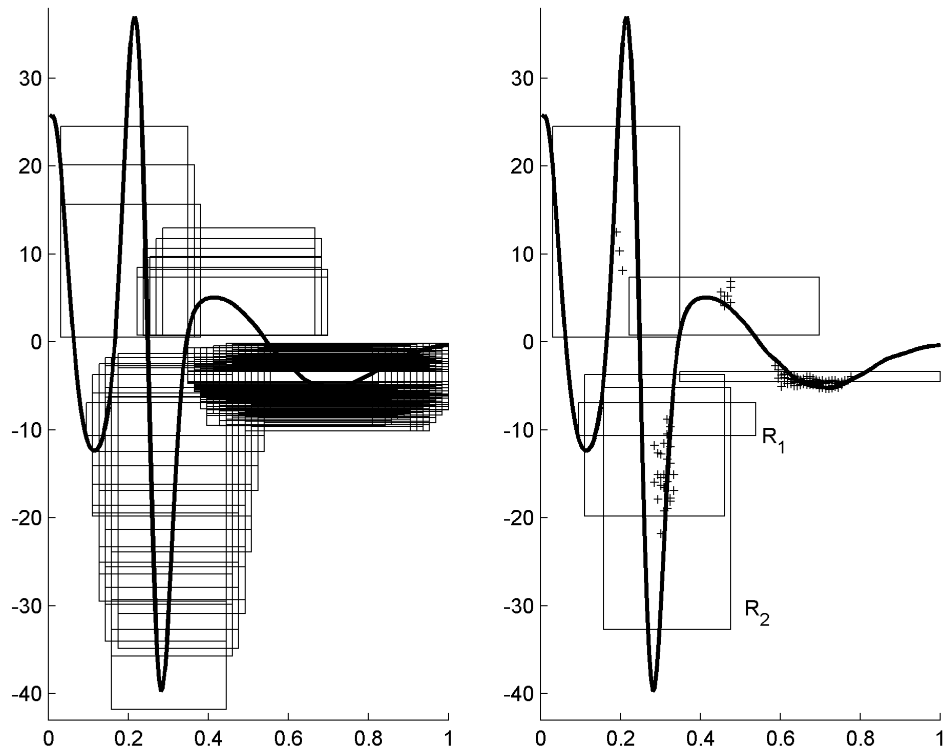

Figure 4: True (unobserved) derivative f' and minimal rectangles (left) as well as sparse minimal rectangles/ midpoints (right) for the same data set as in Figure 3.

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 [13].

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 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.

Figure 4 also shows nicely why a multiscale approach can provide additional insight compared to a one-scale method. Consider the rectangles R_1 and R_2 in the right display of Figure 4 and denote by (t_1, h_1) and (t_2, h_2) the corresponding indices in B_n (as in (4.4)). Note, R_1 and R_2 belong to more or less the same value in the time domain, i.e. $t_1 \approx t_2$ but different bandwidths h_1, h_2. Therefore, we may view R_1 and R_2 as a superposition of confidence statements on different scales. Since R_1 yields the better resolution in the t-coordinate and R_2 the better resolution in the y-coordinate, different qualitative statements can be inferred at the same time point. More practical, we would use R_2 in order to construct a confidence statement as in the lower display of Figure 3 and from R_1 we obtain the better bound for $\inf f'$. This would be impossible for any one-scale method.

7 Outlook and Discussion

We have investigated multiscale methods in order to analyze shape constraints expressed as pseudo-differential operator inequalities in deconvolution models. 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.

Our multiscale approach allows us to identify intervals such that for given significance level we know that $\mathrm{op}(p)f > 0$ 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 $\mathrm{op}(p)f$. Since we only required that $\mathrm{op}(p)f$ is continuous, $\mathrm{op}(p)f$ 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 $\mathrm{op}(p)f$ 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.

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 constraint which are associated with pseudo-differential operators appears to be a limitation of our method, since 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 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 operator. 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. This research was supported by the joint research grant FOR 916 of the German Science Foundation (DFG) and the Swiss National Science Foundation (SNF). The first author was partly funded by DFG postdoctoral fellowship SCHM 2807/1-1. The second author would like to acknowledge support by DFG grants CRC 755 and CRC 803. The authors are grateful for very helpful comments by two referees and an associate editor which led to a more general version of previous results.

Appendix A

Throughout the appendix, let

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

Furthermore, we often use the normalized differential $đ\xi := (2\pi)^{-1}d\xi$

Proof of Theorem 1. Let us study in a first step the statistic

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

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 (two-sided) Brownian motion W, such that with

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

we have

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

The main argument is based on the standard version of KMT (cf. [31]). 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

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

where G^{-1} denotes the quantile function of Y. For convenience, we restate the celebrated KMT inequality for the uniform empirical process.

Theorem 5 (KMT on $[0,1]$, cf. [31]). *There exist versions of $\mathbb{U}_n$ and a Brownian bridge B such that for all x*

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

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 [30], (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 6. *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 $C', K', \lambda' > 0$.

In Theorem 6, $\mathbb{P}^\star$ refers to the outer measure, however, for the function class considered in this paper, we have measurability of the corresponding event and hence may replace $\mathbb{P}^\star$ by $\mathbb{P}$. It is well-known (cf. Giné *et al.* [18], p. 172) that

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

Now, assume that $\rho : \mathbb{R} \to \mathbb{R}$ is such that $\mathrm{TV}(\rho) + 3|\rho(1)| \le 1$. Define $\widetilde{\rho} = (\rho - \rho(1)\mathbb{I}_{[0,1]})/(1 - |\rho(1)|)$ and observe that $\mathrm{TV}(\widetilde{\rho}) \le 1$ and $\widetilde{\rho}(1) = 0$. By (A.2) 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 $\sum |\lambda_i| \le 1$. Therefore, $\rho = (1 - |\rho(1)|)\widetilde{\rho} + \rho(1)\mathbb{I}_{[0,1]}$ can be written as linear combination of indicator functions, such that the sum of the absolute values of weights is bounded by 1. This shows

$$\big\{\rho \;\big|\; \rho : \mathbb{R} \to \mathbb{R},\ \operatorname{supp}\rho \subset [0,1],\ \mathrm{TV}(\rho) + 3|\rho(1)| \le 1\big\} \subset \overline{\mathrm{sc}}(\mathcal{F}_0).$$

Since $\mathrm{TV}(\psi_{t,h}\circ G^{-1}) \le \mathrm{TV}(\psi_{t,h})$ it follows by Assumption 1 (ii) that the function class

$$\mathcal{F}_n := \Big\{ C_\star V_{t,h}^{-1}\sqrt{h}\ \psi_{t,h}\circ G^{-1} : (t,h)\in B_n,\ G\in\mathcal{G}_{c,C,q}\Big\}$$

is a subset of $\overline{\mathrm{sc}}(\mathcal{F}_0)$ for sufficiently small constant $C_\star$. Combining Theorems 5 and 6 shows for $\widetilde{\delta}_n = n^{-1/2}$ that there are constants C', K', λ' and a Brownian bridge $(B(f))_{f\in\overline{\mathrm{sc}}(\mathcal{F}_0)}$ such that for $x>0$,

$$\mathbb{P}\Big(\sup_{(t,h)\in B_n,\ G\in\mathcal{G}} C_\star \frac{\sqrt{h}\big|\mathbb{U}_n\big(\psi_{t,h}\circ G^{-1}\big) - B\big(\psi_{t,h}\circ G^{-1}\big)\big|}{V_{t,h}} \ge n^{-1/2}(x + C'\log n)\Big) \le K' e^{-\lambda' x}.$$

Due to Lemma C.3 (i) and $l_n \ge \nu/n$ for sufficiently large n, we have that $w_{l_n} \le w_{\nu/n}$. This readily implies with $x=\log n$,

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

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 $B(f) = W(f) - (\int f)W(1)$. We have

$$\begin{aligned}
\sup_{(t,h)\in B_n,\ G\in\mathcal{G}} w_h \frac{\big|\int \psi_{t,h}(u)\ dG(u)\big|}{V_{t,h}\ \sqrt{g(t)}} &\le c^{-1} \sup_{(t,h)\in B_n,\ G\in\mathcal{G}} w_h \frac{\|\psi_{t,h}\|_1}{V_{t,h}\ \sqrt{g(t)}} \\
&\lesssim \sup_{h\in[l_n,u_n]} w_h h^{1/2} \le w_{u_n} u_n^{1/2},
\end{aligned}$$

where the second inequality follows from Assumption 1 (ii) and the last inequality from Lemma C.3 (ii). This implies further

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

and therefore

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

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

$$(W_t)_{t\in\mathbb{R}} = \Big(\int_0^t \frac{\mathbb{I}_{\{g>0\}}(s)}{\sqrt{g(s)}} dW_{G(s)}^{(1)}\Big)_{t\in\mathbb{R}}$$

are (two-sided) standard Brownian motions, proving $W(\psi_{t,h} \circ G^{-1}) = \int \psi_{t,h}(s)\sqrt{g(s)}dW_s$ and hence (A.1). Further note that Assumption 1 (iii) together with Lemma B.6 shows that

$$\sup_{G\in\mathcal{G}} \left| T_n^{(2)}(W) - \sup_{(t,h)\in B_n} w_h \frac{|\int \psi_{t,h}(s)dW_s|}{V_{t,h}} - \widetilde{w}_h \right| = O_P(\kappa_n).$$

In a final step let us show that (2.7) is almost surely bounded. In order to establish the result, we use Theorem 6.1 and Remark 1 of Dümbgen and Spokoiny [12]. We set $\rho\big((t,h),(t',h')\big) = (|t-t'| + |h-h'|)^{1/2}$. Further, let $X(t,h) = \sqrt{h}V_{t,h}^{-1}\int \psi_{t,h}(s)dW_s$ and $\sigma(t,h) = h^{1/2}$.

By assumption, X has continuous sample paths on $\mathcal{T}$ and obviously, for all $(t,h),(t',h') \in \mathcal{T}$,

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

Let $Z \sim \mathcal{N}(0,1)$. Since $X(t,h)$ is a Gaussian process and $V_{t,h} \ge \|\phi_{t,h}\|_2$, $\mathbb{P}(X(t,h) > \sigma(t,h)\eta) \le \mathbb{P}(Z > \eta) \le \exp(-\eta^2/2)$, for any $\eta > 0$. Further, denote by

$$A_{t,t',h,h'} := \left\| \frac{\psi_{t,h}\sqrt{h}}{V_{t,h}} - \frac{\psi_{t',h'}\sqrt{h}'}{V_{t',h'}} \right\|_2. \tag{A.3}$$

Because of $\mathbb{P}(|X(t,h) - X(t',h')| \ge A_{t,t',h,h'}\eta\Big) \le 2\exp\big(-\eta^2/2\big)$ we have by Lemma B.5 for a universal constant $K > 0$,

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

Finally, we can bound the entropy $\mathcal{N}((\delta u)^{1/2}, \{(t,h) \in \mathcal{T} : h \le \delta\})$ similarly as in [12], p. 145. Therefore, application of Remark 1 in [12] shows that

$$S := \sup_{(t,h)\in\mathcal{T}} \frac{\sqrt{\frac{1}{2}\log\frac{e}{h}}\ \big|\int \psi_{t,h}(s)dW_s\big|}{\log\big(e\log\frac{e}{h}\big)\ V_{t,h}} - \frac{\sqrt{\log(\frac{1}{h})\log(\frac{e}{h})}}{\log\big(e\log\frac{e}{h}\big)}$$

is almost surely bounded from above. Define

$$S' := \sup_{(t,h)\in\mathcal{T}} \frac{\sqrt{\frac{1}{2}\log\frac{\nu}{h}}\ \big|\int \psi_{t,h}(s)dW_s\big|}{\log\log\frac{\nu}{h}\ V_{t,h}} - \frac{\sqrt{\log(\frac{1}{h})\log(\frac{\nu}{h})}}{\log\log\frac{\nu}{h}}.$$

If $e < \nu \le e^e$, then

$$\log\log\tfrac{\nu}{h} = \log\left(\tfrac{\log\nu}{e}\log\tfrac{e^e}{h^{e/\log\nu}}\right) \ge \log\log\nu - 1 + \log\big(e\log\tfrac{e}{h}\big)$$

implies

$$\frac{\log\left(e\log\frac{e}{h}\right)}{\log\log\frac{\nu}{h}} \leq \frac{1}{\log\log\nu} + 1.$$

Furthermore, $\log\nu/h \leq (\log\nu)(\log e/h)$. Suppose now that $S' > 0$ (otherwise S' is bounded from below 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, (2.7) holds, i.e.

$$\sup_{(t,h)\in\mathcal{T}} w_h \frac{\left|\int \psi_{t,h}(s)dW_s\right|}{V_{t,h}} - \widetilde{w}_h$$

is almost surely bounded.

In the last step, let us prove that $\sup_{G\in\mathcal{G}_{c,C,q}} |T_n - T_n^{(1)}| = O_P(\sup_{G\in\mathcal{G}} \|\widehat{g}_n - g\|_\infty \log n/\log\log n)$. For sufficiently large n and because $G \in \mathcal{G}$, $\widehat{g}_n \geq c/2$ for all $t \in [0,1]$. Therefore using Lemma C.3 (i),

$$\begin{aligned}
\sup_{G\in\mathcal{G}} |T_n - T_n^{(1)}| &\leq \sup_{(t,h)\in B_n,\ G\in\mathcal{G}} w_h \frac{\left|T_{t,h} - \mathbb{E}[T_{t,h}]\right|}{V_{t,h}\ \sqrt{g(t)}} \frac{\sup_{G\in\mathcal{G}} \left\|\widehat{g}_n - g\right\|_\infty}{\widehat{g}_n(t)} \\
&\leq \frac{2\sup_{G\in\mathcal{G}} \left\|\widehat{g}_n - g\right\|_\infty}{c} \sup_{(t,h)\in B_n,\ G\in\mathcal{G}} w_h \frac{\left|T_{t,h} - \mathbb{E}[T_{t,h}]\right|}{V_{t,h}\ \sqrt{g(t)}} \\
&\leq \frac{2\sup_{G\in\mathcal{G}} \left\|\widehat{g}_n - g\right\|_\infty}{c} (T_n^{(1)} + \sup_{h\in[l_n,u_n]} \widetilde{w}_h) \\
&\leq \frac{2\sup_{G\in\mathcal{G}} \left\|\widehat{g}_n - g\right\|_\infty}{c} \big(T_n^{(1)} + O(\frac{\log n}{\log\log n})\big).
\end{aligned} \tag{A.4}$$

Since $T_n^{(1)}$ is a.s. bounded by Theorem 1, the result follows. □

Remark 2. *Next, we give a proof of Theorem 2. In fact we proof a slightly stronger version, which does not necessarily require the symbol a to be elliptic and $V_{t,h} = \|v_{t,h}\|_2$. It is only assumed that*

(i) $V_{t,h} \geq \|v_{t,h}\|_2$,

(ii) there exists constants c_V, C_V with $0 < c_V \leq h^{m+r-1/2}V_{t,h} \leq C_V < \infty$

(iii) for all $(t,h), (t',h') \in \mathcal{T}$ and whenever $h \leq h'$ it holds that $h^{m+r}|V_{t,h} - V_{t',h'}| \leq C_V(|t-t'| + |h-h'|)^{1/2}$.

Note, that as a special case these conditions are satisfied for $V_{t,h} = \|v_{t,h}\|_2$ *if* $\operatorname{op}(a)$ *is elliptic. This follows directly from Lemmas B.2 and B.4.*

Proof of Theorem 2. In order to prove the statements it is sufficient to check the conditions of Theorem 1. For $h > 0$, define the symbol

$$a^{\star}_{t,h}(x,\xi) := h^{\overline{m}} a^{\star}(xh + t, h^{-1}\xi). \tag{A.5}$$

Under the imposed conditions and by Remark B.1 we may apply Lemma B.3 for $\mathfrak{a}^{(t,h)} = a^{\star}_{t,h}$ and therefore, uniformly over $(t,h) \in \mathcal{T}$ and $u, u' \in \mathbb{R}$,

(I) $|v_{t,h}(u)| \lesssim h^{-m-r} \min\big(1, \frac{h^2}{(u-t)^2}\big)$.

(II) $|v_{t,h}(u) - v_{t,h}(u')| \lesssim h^{-m-r-1}|u - u'|$ and if $u, u' \neq t$,

$$|v_{t,h}(u) - v_{t,h}(u')| \lesssim h^{1-m-r}\frac{|u-u'|}{|u'-t|\;|u-t|} = h^{1-m-r}\Big|\int_{u'}^{u} \frac{1}{(x-t)^2}dx\Big|.$$

Using (I), we obtain $\|v_{t,h}\|_\infty \lesssim h^{-m-r}$ and $\|v_{t,h}\|_1 \lesssim h^{1-m-r}$. In order to show that the total variation is of the right order, let us decompose $v_{t,h}$ further into $v^{(1)}_{t,h} = v_{t,h}\mathbb{I}_{[t-h,t+h]}$ and $v^{(2)}_{t,h} = v_{t,h} - v^{(1)}_{t,h}$. By (II), $\operatorname{TV}(v^{(1)}_{t,h}) \lesssim h^{-m-r}$ and

$$\operatorname{TV}(v^{(2)}_{t,h}) \lesssim h^{-m-r} + h^{1-m-r}\int_{t+h}^{\infty} \frac{1}{(x-t)^2}dx \lesssim h^{-m-r}.$$

Since $\operatorname{TV}(v_{t,h}) \leq \operatorname{TV}(v^{(1)}_{t,h}) + \operatorname{TV}(v^{(2)}_{t,h}) \lesssim h^{-m-r}$, this shows together with Remark 2 that part (ii) of Assumption 1 is satisfied.

In the next step we verify Assumption 1, (iii) with $\kappa_n = \sup_{(t,h)\in B_n} w_h h^{1/2}\log(1/h) \lesssim u_n^{1/2}\log^{3/2} n$ (cf. Lemma C.3, (ii)), i.e. we show

$$\sup_{(t,h)\in B_n,\; G\in\mathcal{G}} w_h \frac{\operatorname{TV}\big(v_{t,h}(\cdot)[\sqrt{g(\cdot)} - \sqrt{g(t)}]\langle\cdot\rangle^{\alpha}\big)}{V_{t,h}} \lesssim u_n^{1/2}\log^{3/2} n.$$

By Lemma C.2, we see that this holds for $v_{t,h}$ replaced by $v^{(1)}_{t,h}$. Therefore, it remains to prove the statement for $v^{(2)}_{t,h}$. Let us decompose $v^{(2)}_{t,h}$ further into $v^{(2,1)}_{t,h} = v_{t,h}\mathbb{I}_{[t-1,t+1]\cap[t-h,t+h]^c}$ and $v^{(2,2)}_{t,h} = v^{(2)}_{t,h} - v^{(2,1)}_{t,h} = v_{t,h}\mathbb{I}_{[t-1,t+1]^c}$. For the remaining part, let u, u', such that $|u-t| \geq |u'-t| \geq h$. We have

$$\begin{aligned}\operatorname{TV}\big(v^{(2,1)}_{t,h}(\cdot)\big[\sqrt{g(\cdot)} - \sqrt{g(t)}\big]\langle\cdot\rangle^{\alpha}\big) &\lesssim \big\|v^{(2,1)}_{t,h}(\cdot)\big[\sqrt{g(\cdot)} - \sqrt{g(t)}\big]\big\|_\infty \\ &\quad + \operatorname{TV}\big(v^{(2,1)}_{t,h}(\cdot)\big[\sqrt{g(\cdot)} - \sqrt{g(t)}\big]\big).\end{aligned} \tag{A.6}$$

Using (I) and (II) together with the properties of the class $\mathcal{G}$ we can bound the variation $\big|v_{t,h}^{(2,1)}(u)\big[\sqrt{g(u)}-\sqrt{g(t)}\big]-v_{t,h}^{(2,1)}(u')\big[\sqrt{g(u')}-\sqrt{g(t)}\big]\big|$ by

$$\begin{aligned}&\big|v_{t,h}^{(2,1)}(u)-v_{t,h}^{(2,1)}(u')\big|\cdot\big|\sqrt{g(u')}-\sqrt{g(t)}\big|+\big|v_{t,h}^{(2,1)}(u)\big|\cdot\big|\sqrt{g(u)}-\sqrt{g(u')}\big|\\&\lesssim h^{1-m-r}\tfrac{|u-u'|}{|u-t|}+h^{2-m-r}\tfrac{|u-u'|}{|u-t|^2}\lesssim h^{1-m-r}\tfrac{|u-u'|}{|u-t|}\le h^{1-m-r}\Big|\int_{u'}^{u}\tfrac{1}{|x-t|}dx\Big|.\end{aligned}$$

This yields due to $h\ge l_n\gtrsim 1/n$,

$$\begin{aligned}\mathrm{TV}\left(v_{t,h}^{(2,1)}(\cdot)[\sqrt{g(\cdot)}-\sqrt{g(t)}]\right)&\lesssim h^{1-m-r}+h^{1-m-r}\int_{t+h}^{t+1}\frac{du}{|u-t|}\\&\lesssim h^{1-m-r}\log\tfrac{1}{h}\lesssim h^{1-m-r}\log n\end{aligned}$$

and with (A.6) also

$$\mathrm{TV}\left(v_{t,h}^{(2,1)}(\cdot)\big[\sqrt{g(\cdot)}-\sqrt{g(t)}\big]\langle\cdot\rangle^{\alpha}\right)\lesssim h^{1-m-r}\log n. \tag{A.7}$$

Finally, let us address the total variation term involving $v_{t,h}^{(2,2)}$. Given $\mathcal{G}_{c,C,q}$ we can choose α such that $\alpha>1/2$ and $\alpha+q<1$ (recall that $0\le q<1/2$). By Lemma B.7, we find that

$$\big|v_{t,h}^{(2,2)}(u)\langle u\rangle^{\alpha}-v_{t,h}^{(2,2)}(u')\langle u'\rangle^{\alpha}\big|\lesssim h^{1-m-r}\Big|\int_{u'}^{u}\frac{1}{(x-t)^{2-\alpha}}+\frac{1}{(x-t)^2}dx\Big|.$$

Moreover

$$\langle u\rangle^{\alpha}(1+|u'|+|u|)^q\le(1+|u'|+|u|)^{q+\alpha}\le(3+2|u-t|)^{q+\alpha}\le 3+2|u-t|^{q+\alpha}.$$

and thus

$$\big|v_{t,h}^{(2,2)}(u)\langle u\rangle^{\alpha}\big|\ \big|\sqrt{g(u)}-\sqrt{g(u')}\big|\lesssim h^{2-m-r}\frac{|u-t|^{q+\alpha}+1}{|u-t|^2}|u-u'|.$$

This allows us to bound the variation by

$$\begin{aligned}&\big|v_{t,h}^{(2,2)}(u)\big[\sqrt{g(u)}-\sqrt{g(t)}\big]\langle u\rangle^{\alpha}-v_{t,h}^{(2,2)}(u')\big[\sqrt{g(u')}-\sqrt{g(t)}\big]\langle u'\rangle^{\alpha}\big|\\&\le\big|v_{t,h}^{(2,2)}(u)\langle u\rangle^{\alpha}\big|\ \big|\sqrt{g(u)}-\sqrt{g(u')}\big|+\frac{2}{\sqrt{c}}\big|v_{t,h}^{(2,2)}(u)\langle u\rangle^{\alpha}-v_{t,h}^{(2,2)}(u')\langle u'\rangle^{\alpha}\big|\\&\lesssim h^{1-m-r}\Big|\int_{u'}^{u}\frac{1}{(x-t)^{2-q-\alpha}}+\frac{1}{(x-t)^{2-\alpha}}+\frac{1}{(x-t)^2}\ dx\ \Big|\end{aligned}$$

and therefore we conclude that

$$\begin{aligned}&\mathrm{TV}\left(v_{t,h}^{(2,2)}(\cdot)\big[\sqrt{g(\cdot)}-\sqrt{g(t)}\big]\langle\cdot\rangle^{\alpha}\right)\\&\lesssim h^{1-m-r}+h^{1-m-r}\int_{t+1}^{\infty}\frac{1}{(x-t)^{2-q-\alpha}}+\frac{1}{(x-t)^{2-\alpha}}+\frac{1}{(x-t)^2}dx\le h^{1-m-r}.\end{aligned}$$

Together with the bound for $v^{(1)}_{t,h}$ and (A.7) this shows that Assumption 1, (iii) holds.

Finally, Assumption 1 (iv) follows from Lemma B.4 and Remark 2 due to $\phi \in H^{\lceil r+m\rceil} \cap H^{r+m+1/2}$, $\operatorname{supp}\phi \subset [0,1]$ and $\phi \in \mathrm{TV}(D^{\lceil r+m\rceil}\phi) < \infty$. This shows that Assumption 1 holds for $(v_{t,h}, V_{t,h})$.

In the next step, we verify that $(t,h) \mapsto X(t,h) = \sqrt{h}V_{t,h}^{-1}\int v_{t,h}(s)dW_s$ has continuous sample paths. Note that in view of Lemma B.6, it is sufficient to show that there is an α with $1/2 < \alpha < 1$, such that

$$\mathrm{TV}\left(\left(\sqrt{h}V_{t,h}^{-1}v_{t,h} - \sqrt{h'}V_{t',h'}^{-1}v_{t',h'}\right)\langle\cdot\rangle^{\alpha}\right) \to 0,$$

whenever $(t',h') \to (t,h)$ on the space $\mathcal{T}$. Since Assumption 1 (iv) holds, we have

$$\left|\sqrt{h}V_{t,h}^{-1} - \sqrt{h'}V_{t',h'}^{-1}\right| \leq \frac{\sqrt{|h-h'|}}{V_{t,h}} + V_{t,h}^{-1}\frac{\sqrt{h'}|V_{t',h'} - V_{t,h}|}{V_{t',h'}} \to 0, \quad \text{for } (t',h') \to (t,h).$$

By Lemma B.7, $\mathrm{TV}(v_{t,h}(\cdot)\langle\cdot\rangle^{\alpha}) < \infty$. Therefore, it is sufficient to show that

$$\mathrm{TV}\left((v_{t,h} - v_{t',h'})\langle\cdot\rangle^{\alpha}\right) \to 0, \quad \text{whenever } (t',h') \to (t,h). \tag{A.8}$$

Using (B.3), we obtain

$$\begin{aligned}(K_{t,h}^{\gamma,\overline{m}}a_{t,h}^{\star})(u) &= v_{t,h} - v_{t',h'} \\ &= h^{-\overline{m}}\int \lambda_{\gamma}^{\mu}\left(\tfrac{s}{h}\right)\mathcal{F}\left(\mathrm{Op}(a_{t,h}^{\star})(\phi - \phi\circ S_{t',h'}\circ S_{t,h}^{-1})\right)(s)e^{is(u-t)/h}\,đs.\end{aligned}$$

Using Remark B.1, we can apply again Lemma B.3 (here ϕ should be replaced by $\phi - \phi \circ S_{t',h'} \circ S_{t,h}^{-1}$). In order to verify (A.8), we observe that by Lemma B.7 it is enough to show $\|\phi - \phi\circ S_{t',h'}\circ S_{t,h}^{-1}\|_{H_4^{\overline{q}}} \to 0$ for some $\overline{q} > r+m+3/2$ whenever $(t',h') \to (t,h)$ in $\mathcal{T}$. Note that

$$\begin{aligned}\left\|\phi - \phi\circ S_{t',h'}\circ S_{t,h}^{-1}\right\|_{H_4^{\overline{q}}}^2 &= \tfrac{1}{h}\sum_{j=0}^{4}\int \langle s\rangle^{2\overline{q}}\left|\mathcal{F}\left((x^j\phi)\circ S_{t,h}\right)(s) - \mathcal{F}\left((S_{t,h}(\cdot))^j(\phi\circ S_{t',h'})\right)(s)\right|^2 ds \\ &\leq \frac{2}{h}\sum_{j=0}^{4}\left\|(x^j\phi)\circ S_{t,h} - (x^j\phi)\circ S_{t',h'}\right\|_{H^{\overline{q}}}^2 \\ &\quad + \int \langle s\rangle^{2\overline{q}}\left|\mathcal{F}\left(\left[(S_{t',h'}(\cdot))^j - (S_{t,h}(\cdot))^j\right](\phi\circ S_{t',h'})\right)(s)\right|^2 ds\end{aligned} \tag{A.9}$$

with $(S_{t,h}(\cdot))^j := \left(\frac{\cdot - t}{h}\right)^j$. Note that for real numbers a, b we have the identity $a^j - b^j = \sum_{\ell=1}^{k}\binom{k}{\ell}b^{k-\ell}(a-b)^{\ell}$. Moreover, we can apply Lemma B.4 for $\overline{q}$ with $m + r + 3/2 < \overline{q} < \lfloor r+m+5/2\rfloor$ (and such a $\overline{q}$ clearly exists). Thus, with $a = S_{t,h}(\cdot)$, $b = S_{t',h'}(\cdot)$ and $S_{t,h} - S_{t',h'} = (h/h'-1)S_{t',h'} - (t'-t)/h$ the r.h.s. of (A.9) converges to zero if $(t',h') \to (t,h)$.

□

Proof of Theorem 3. By assumption, we can write $p_R(x,\xi) = a_R(x,\xi)|\xi|^{\gamma_1}\iota_\xi^{\mu_1}$ with $a_R \in S^{\overline{m}_1}$ and $\overline{m}_1+\gamma_1 = m'$. Recall that $p_P(x,\xi) = a_P(x)|\xi|^m\iota_\xi^\mu$. Since a_P is real-valued, $\operatorname{Op}(a_P)$ is self-adjoint. Taking the adjoint is a linear operator and therefore arguing as in (3.4) yields

$$\mathcal{F}\big(\operatorname{op}(p)^\star(\phi\circ S_{t,h})\big)(s) = |s|^m\iota_s^{-\mu}\mathcal{F}\big(a_P(\phi\circ S_{t,h})\big)(s) + |s|^{\gamma_1}\iota_s^{-\mu_1}\mathcal{F}\big(\operatorname{Op}(a_R^\star)(\phi\circ S_{t,h})\big)(s).$$

Decompose $v_{t,h} = v_{t,h}^{(1)} + v_{t,h}^{(2)}$ with

$$\begin{aligned}
v_{t,h}^{(1)}(u) &:= \int \lambda_m^\mu(s)\mathcal{F}\big(a_P(\phi\circ S_{t,h})\big)(s)e^{isu}đs \\
&= \int \lambda_m^\mu\big(\tfrac{s}{h}\big)\mathcal{F}\big(a_P(\cdot h+t)\phi\big)(s)e^{is(u-t)/h}đs \\
v_{t,h}^{(2)}(u) &:= \int \lambda_{\gamma_1}^{\mu_1}(s)\mathcal{F}\big(\operatorname{Op}(a_R^\star)(\phi\circ S_{t,h})\big)(s)e^{isu}đs \\
&= h^{-\overline{m}_1}\int \lambda_{\gamma_1}^{\mu_1}\big(\tfrac{s}{h}\big)\mathcal{F}\big(\operatorname{Op}(a_{t,h}^{(1)})\phi\big)(s)e^{is(u-t)/h}đs
\end{aligned}$$

using similar arguments as in (B.3) and $a_{t,h}^{(1)}(x,\xi) := h^{\overline{m}_1}a_R^\star(xh+t,h^{-1}\xi)$. For $j=1,2$, we denote by $T_{t,h}^{(j)}$ and $T_n^{P,(j)}$ the statistics $T_{t,h}$ and T_n^P with $v_{t,h}$ replaced by $v_{t,h}^{(j)}$, $j=1,2$, respectively. Recall the definitions of σ and τ and set

$$\begin{aligned}
v_{t,h}^P(u) &:= Aa_P(t)\int |s|^{r+m}\iota_s^{-\rho-\mu}\mathcal{F}(\phi\circ S_{t,h})(s)e^{isu}đs \\
&= Ah^{-r-m}a_P(t)\int |s|^{r+m}\iota_s^{-\rho-\mu}\mathcal{F}(\phi)(s)e^{is(u-t)/h}đs \\
&= Aa_P(t)D_+^\sigma D_-^\tau\phi\big(\tfrac{u-t}{h}\big). \qquad (A.10)
\end{aligned}$$

Further let $V_{t,h}^P := \|v_{t,h}^P\|_2 = |Aa_P(t)|\big\|D_+^{r+m}\phi((\cdot-t)/h)\big\|_2 = h^{1/2-r-m}|Aa_P(t)|\big\|D_+^{r+m}\phi\big\|_2$, and

$$T_n^{P,(1),\infty}(W) := \sup_{(t,h)\in B_n} w_h\left(\frac{\big|\int \operatorname{Re} v_{t,h}^{(1)}(s)dW_s\big|}{V_{t,h}^P} - \sqrt{2\log\tfrac{\nu}{h}}\right).$$

Note that for the approximation of T_n^P, we can write

$$T_n^{P,\infty}(W) = \sup_{(t,h)\in B_n} w_h\left(\frac{\big|\int \operatorname{Re} v_{t,h}^P(s)dW_s\big|}{V_{t,h}^P} - \sqrt{2\log\tfrac{\nu}{h}}\right).$$

Since $|T_n^P - T_n^{P,\infty}(W)| \le |T_n^P - T_n^{P,(1)}| + |T_n^{P,(1)} - T_n^{P,(1),\infty}(W)| + |T_n^{P,(1),\infty}(W) - T_n^{P,\infty}(W)|$ it is sufficient to show that there exists a Brownian motion W such that the terms on the right hand side converge to zero in probability. This will be done separately, and proofs for the single terms are denoted by $(I), (II)$ and (III). From (II) and (III), we will be able to conclude the boundedness of the approximating statistic.

(I): It is easy to see that for a constant K, $\|v_{t,h}^{(2)}\|_2 \leq K h^{1/2-m'-r} =: V_{t,h}^R$. By Remark 2 and

$$\begin{aligned}
&\big|T_n^P - T_n^{P,(1)}\big| \\
&\leq \sup_{h\in[l_n,u_n]} \frac{V_{t,h}^R}{V_{t,h}^P} \left(\sup_{(t,h)\in B_n} w_h \Big(\frac{|T_{t,h}^{(2)} - \mathbb{E}T_{t,h}^{(2)}|}{\sqrt{\widehat{g}_n(t)}\ V_{t,h}^R} - \sqrt{2\log\left(\tfrac{\nu}{h}\right)}\Big) + \sup_{h\in[l_n,u_n]} w_h \sqrt{2\log\left(\tfrac{\nu}{h}\right)} \right),
\end{aligned}$$

we can apply Theorem 2 (where m should be replaced by m', of course). Because of $u_n^{m-m'} \log n \to 0$, (I) is proved.

(II): We show that there is a Brownian motion W such that $|T_n^{P,(1)} - T_n^{P,(1),\infty}(W)| \leq |T_n^{P,(1)} - \widetilde{T}_n^{(1)}| + |\widetilde{T}_n^{(1)} - \widetilde{T}_n^{(1),\infty}(W)| + |\widetilde{T}_n^{(1),\infty}(W) - T_n^{P,(1),\infty}(W)| = o_P(1)$ with

$$\widetilde{T}_n^{(1)} := \sup_{(t,h)\in B_n} w_h \Big(\frac{\big|T_{t,h}^{(1)} - \mathbb{E}T_{t,h}^{(1)}\big|}{\sqrt{\widehat{g}_n(t)}\ \|v_{t,h}^{(1)}\|_2} - \sqrt{2\log\left(\tfrac{\nu}{h}\right)}\Big)$$

and

$$\widetilde{T}_n^{(1),\infty}(W) := \sup_{(t,h)\in B_n} w_h \Big(\frac{\big|\int \operatorname{Re} v_{t,h}^{(1)}(s) dW_s\big|}{\sqrt{\widehat{g}_n(t)}\ \|v_{t,h}^{(1)}\|_2} - \sqrt{2\log\left(\tfrac{\nu}{h}\right)}\Big).$$

Since by Assumption 4, $a_p \in S^0$ is elliptic and $p_P \in \underline{S}^m$, we find that $|\widetilde{T}_n^{(1)} - \widetilde{T}_n^{(1),\infty}(W)| = o_P(1)$ and

$$\widetilde{T}_n^{(1),\infty}(W) \leq \sup_{(t,h)\in \mathcal{T}} w_h \Big(\frac{\big|\int \operatorname{Re} v_{t,h}^{(1)}(s) dW_s\big|}{\sqrt{\widehat{g}_n(t)}\ \|v_{t,h}^{(1)}\|_2} - \sqrt{2\log\left(\tfrac{\nu}{h}\right)}\Big) < \infty \quad \text{a.s.} \tag{A.11}$$

by applying Theorem 2. Moreover, similar as in (A.4) and using $w_h \sqrt{2\log\left(\frac{\nu}{h}\right)} \geq 1$,

$$\sup_{G\in\mathcal{G}} \big|T_n^{P,(1)} - \widetilde{T}_n^{(1)}\big| \leq \sup_{(t,h)\in B_n} w_h \sqrt{2\log\left(\tfrac{\nu}{h}\right)} \frac{\big|V_{t,h}^P - \|v_{t,h}^{(1)}\|_2\big|}{V_{t,h}^P} \Big(1 + \sup_{G\in\mathcal{G}} \widetilde{T}_n^{(1)}\Big)$$

and

$$\big|\widetilde{T}_n^{(1),\infty}(W) - T_n^{P,(1),\infty}(W)\big| \leq \sup_{(t,h)\in B_n} w_h \sqrt{2\log\left(\tfrac{\nu}{h}\right)} \frac{\big|V_{t,h}^P - \|v_{t,h}^{(1)}\|_2\big|}{V_{t,h}^P} \Big(1 + \widetilde{T}_n^{(1),\infty}(W)\Big)$$

To finish the proof for (II), it remains to verify

$$\sup_{(t,h)\in B_n} w_h \sqrt{2\log\left(\tfrac{\nu}{h}\right)} \frac{\|v_{t,h}^P - v_{t,h}^{(1)}\|_2}{V_{t,h}^P} = o(1), \tag{A.12}$$

which will be done below.

(III): By Lemma B.6, we obtain $|T_n^{P,(1),\infty} - T_n^{P,\infty}| = o_P(1)$ if for some $\alpha > 1/2$,

$$\sup_{(t,h)\in B_n} w_h \frac{\mathrm{TV}\left((v_{t,h}^P - v_{t,h}^{(1)})\langle\cdot\rangle^\alpha\right)}{V_{t,h}^P} = o(1). \tag{A.13}$$

Let χ be a cut function, i.e. $\chi \in \mathcal{S}$ (the Schwartz space), $\chi(x) = 1$ for $x \in [-1, 1]$ and $\chi(x) = 0$ for $x \in (-\infty, -2] \cup [2, \infty)$ and define $p_{t,h}^{(1)}(x, \xi) = h^{-1}\chi(x)(a_P(xh+t) - a_P(t))$ and $p_{t,h}^{(2)}(x, \xi) = (xh)^{-1}(1 - \chi(x))(a_P(xh + t) - a_P(t))$. Then, $p_{t,h}^{(1)}, p_{t,h}^{(2)} \in S^0$ and $\big(a_P(\cdot h + t) - a_P(t)\big)\phi = h\,\mathrm{Op}(p_{t,h}^{(1)})\phi + h\,\mathrm{Op}(p_{t,h}^{(2)})(x\phi)$. Define the function

$$d_{t,h}(u) := \int e^{is(u-t)/h}\Big(\frac{1}{\mathcal{F}(f_\epsilon)(-\frac{s}{h})} - A\iota_s^{-\rho}\big|\tfrac{s}{h}\big|^r\Big)\iota_s^{-\mu}|s|^m\mathcal{F}(\phi)(s)đs \tag{A.14}$$

and note that

$$\|d_{t,h}\|_2^2 \lesssim h^{1+2m}\int \big\langle\tfrac{s}{h}\big\rangle^{2r+2m-2\beta_0}\big|\mathcal{F}(\phi)(s)\big|^2 ds \lesssim h^{1+2\beta_0^\star-2r}\|\phi\|_{H^{r+m}}^2$$

with $\beta_0^\star := \beta_0 \wedge (m + r)$. Using (B.2), we have now the decomposition

$$v_{t,h}^{(1)} - v_{t,h}^P = hK_{t,h}^{m,0}p_{t,h}^{(1)} + hK_{t,h}^{m,0}p_{t,h}^{(2)} + a_P(t)h^{-m}d_{t,h}, \tag{A.15}$$

where we have to replace ϕ by $x\phi$ in the second term of the right hand side. By assumption there exists $q > m + r + 3/2$ such that $\phi \in H_5^q$. Since the assumptions on $p_{t,h}^{(1)}$ and $p_{t,h}^{(2)}$ of Lemma B.3 can be easily verified, we may apply Lemma B.3 to the first two terms on the right hand side of (A.15). This yields together with Lemmas B.7, B.8, and B.9, uniformly over $(t, h) \in \mathcal{T}$,

$$\begin{aligned}\mathrm{TV}\left((v_{t,h}^P - v_{t,h}^{(1)})\langle\cdot\rangle^\alpha\right) \leq\ & \mathrm{TV}\left(\big(hK_{t,h}^{m,0}p_{t,h}^{(1)} + hK_{t,h}^{m,0}p_{t,h}^{(2)} + a_P(t)h^{-m}d_{t,h}\big)\langle\cdot\rangle^\alpha\mathbb{I}_{[t-1,t+1]}\right)\\ &+ \mathrm{TV}\left(v_{t,h}^P\langle\cdot\rangle^\alpha\mathbb{I}_{\mathbb{R}\setminus[t-1,t+1]}\right) + \mathrm{TV}\left(v_{t,h}^{(1)}\langle\cdot\rangle^\alpha\mathbb{I}_{\mathbb{R}\setminus[t-1,t+1]}\right)\\ \lesssim\ & h^{1-m-r} + h^{\beta_0^\star-m-r} + h^{1-r-m}.\end{aligned}$$

Since $m + r > 1/2$, this shows (A.13). From the decomposition (A.15) we obtain further $\|v_{t,h}^P - v_{t,h}^{(1)}\|_2 \lesssim h^{3/2-m-r} + h^{1/2+\beta_0^\star-m-r}$ and this shows (A.12). Thus the first part of the theorem is proved.

Finally with Lemma B.6 it is easy to check that (A.11) implies that (3.13) is bounded since (A.12) and (A.13) also hold with B_n and $o(1)$ replaced by $\mathcal{T}$ and $O(1)$, respectively. □

References

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

[2] Alinhac, S., and Gérard, P. *Pseudo-differential Operators and the Nash-Moser Theorem.* American Mathematical Society, 2007.

[3] Balabdaoui, F., Bissantz, K., Bissantz, N., and Holzmann, H. 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* (2010), 136–146.

[4] 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.

[5] 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.

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

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

[8] Chaudhuri, P., and Marron, J. Scale space view of curve estimation. *Ann. Statist. 28* (2000), 408–428.

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

[10] 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.

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

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

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

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

[15] Fan, J. Asymptotic normality for deconvolution kernel density estimators. *Sankhya 53* (1991), 97–110.

[16] Fan, J. On the optimal rates of convergence for nonparametric deconvolution problem. *Ann. Statist. 19* (1991), 1257–1272.

[17] 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.

[18] 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.

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

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

[21] Golubev, G., and Levit, B. Asymptotically efficient estimation in the Wicksell problem. *Ann. Statist. 26* (1998), 2407–2419.

[22] Groeneboom, P., and Jongbloed, G. Isotonic estimation and rates of convergence in Wiksell's problem. *Ann. Statist. 23* (1995), 1518–1542.

[23] 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.

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

[25] Hörmander, L. *The Analysis of Linear Partial Differential Operators III: Pseudo-Differential Operators.* Springer, Berlin, 2007.

[26] Hwang, I. L. The L^2-boundedness of pseudodifferential operators. *Transactionas of the American Mathematical Society 302* (1987), 55–76.

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

[28] 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.

[29] Kilbas, A., Srivastava, H., and Trujillo, J. *Theory and Applications of Fractional Differential Equations.* Elsevier, Amsterdam, 2006.

[30] 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.

[31] 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.

[32] 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.

[33] Laurent, B., Loubes, J., and Marteau, C. Testing inverse problems: a direct or indirect problem? *J. Statist. Plann. Inference 141* (2011), 1849–1861.

[34] Leadbetter, M., Lindgren, G., and Rootzen, H. *Extremes and related properties of random sequences and processes.* Springer, Berlin, 1983.

[35] Lounici, K., and Nickl, R. Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist. 39* (2011), 201–231.

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

[37] Rachdi, M., and Sabre, R. Consistent estimates of the mode of the probability density function in nonparametric deconvolution problems. *Statist. Probab. Lett. 47* (2000), 297–307.

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

[39] Wloka, J., Rowley, B., and Lawruk, B. *Boundary Value Problems for Elliptic Systems.* Cambridge University Press, 1995.

Appendix B Lemmas for the proof of the main theorems

We have the following uniform and continuous embedding of Sobolev spaces.

Lemma B.1. *Let $(p_i)_{i\in I} \subset S^m$ be a symbol class of pseudo-differential operators. Suppose further that for $\alpha \in \{0,1\}$, $k \in \mathbb{N}$, and finite constants C_k, only depending on k,*

$$\sup_{i\in I} |\partial_x^k \partial_\xi^\alpha p_i(x,\xi)| \le C_k(1+|\xi|)^m, \quad \forall x,\xi \in \mathbb{R}.$$

Then, for any $s \in \mathbb{R}$, there exists a finite constant $C = C(s, m, \max_{k\le 1+2|s|+2|m|} C_k)$, such that for all $\phi \in H^s$,

$$\|\operatorname{Op}(p_i)\phi\|_{H^{s-m}} \le C\|\phi\|_{H^s}.$$

Proof. This proof requires some subtle technicalities, appearing in the theory of pseudo-differential operators. First note that for any symbol $a \in S^0$ there exists a universal constant C_1 (which is in particular independent of a), such that

$$\|\operatorname{Op}(a)u\|_2 \le C_1 \max_{\alpha,\beta\in\{0,1\}} \left\|\partial_x^\beta \partial_\xi^\alpha a(x,\xi)\right\|_{L^\infty(\mathbb{R}^2)} \|u\|_2, \quad \text{for all } u \in L^2 \tag{B.1}$$

(cf. Theorem 2 in Hwang [26]). For $r \in \mathbb{R}$ denote by $\operatorname{Op}(\langle\xi\rangle^r)$ the pseudo-differential operator with symbol $(x,\xi) \mapsto \langle\xi\rangle^r$. It is well-known that this operator is indeed in S^r. Throughout the remaining proof let $C = C(s, m, \max_{k\le 1+2|s|+2|m|} C_k)$, denote a finite but unspecified constant, which may even change from line to line. Note that it is sufficient to show that uniformly in $\psi \in L^2$,

$$\left\|\operatorname{Op}(\langle\xi\rangle^{s-m}) \circ \operatorname{Op}(p_i) \circ \operatorname{Op}(\langle\xi\rangle^{-s})\psi\right\|_2 \le C\|\psi\|_2$$

(set $\phi = \langle D\rangle^{-s}\psi$). The composition of two operators with symbols in S^{m_1} and S^{m_2}, respectively is again a pseudo-differential operator and its symbol is in $S^{m_1+m_2}$. Therefore, $\operatorname{Op}(\langle\xi\rangle^{s-m}) \circ \operatorname{Op}(p_i) \circ \operatorname{Op}(\langle\xi\rangle^{-s}) \in S^0$. Set $p_{0,i}$ for its symbol. With (B.1) the lemma is proved, once we have established that

$$\sup_{i\in I} \max_{\alpha,\beta\in\{0,1\}} \left\|\partial_x^\beta \partial_\xi^\alpha p_{0,i}(x,\xi)\right\|_{L^\infty(\mathbb{R}^2)} \le C < \infty.$$

It is not difficult to see that $\operatorname{Op}(p_i) \circ \operatorname{Op}(\langle\xi\rangle^{-s}) = \operatorname{Op}(p_i\langle\xi\rangle^{-s})$. By Theorem 4.1 in [2], $p_{0,i} = \langle\xi\rangle^{s-m}\#(p_i\langle\xi\rangle^{-s})$, where $\#$ denotes the Leibniz product, i.e. for $p^{(1)} \in S^{m_1}$ and $p^{(2)} \in S^{m_2}$, $p^{(1)}\#p^{(2)}$ can be written as an oscillatory integral (cf. [2, 39])

$$\begin{aligned}
\left(p^{(1)}\#p^{(2)}\right)(x,\xi) &:= \operatorname{Os}-\int\int e^{-iy\eta} p^{(1)}(x,\xi+\eta)p^{(2)}(x+y,\xi)\,dy\,đ\eta \\
&:= \lim_{\epsilon\to 0}\int\int \chi(\epsilon y,\epsilon\eta)e^{-iy\eta}p^{(1)}(x,\xi+\eta)p^{(2)}(x+y,\xi)\,dy\,đ\eta,
\end{aligned}$$

for any χ in the Schwartz space of rapidly decreasing functions on $\mathbb{R}^2$ with $\chi(0,0) = 1$. Further for $a \in S^m$ and arbitrary $l \in \mathbb{N}$, $2l > 1 + m$,

$$\mathrm{Os} - \int\int e^{-iy\eta} a(y,\eta) dy đ\eta = \int\int e^{-iy\eta} \langle y\rangle^{-2} (1-\partial_\eta^2)\big[\langle \eta\rangle^{-2l}(1-\partial_y^2)^l a(y,\eta)\big] dy đ\eta$$

and the integrand on the r.h.s. is in L^1 (cf. [39], p.235). This can be also used to show that differentiation and integration commute for oscillatory integrals,

$$\partial_x^\alpha \partial_\xi^\beta \,\mathrm{Os} - \int\int e^{-iy\eta} a(x,y,\xi,\eta) dy đ\eta = \mathrm{Os} - \int\int e^{-iy\eta} \partial_x^\alpha \partial_\xi^\beta a(x,y,\xi,\eta) dy đ\eta.$$

Using Peetre's inequality, i.e. $\langle \xi+\eta\rangle^s \le 2^{|s|}\langle \xi\rangle^{|s|}\langle \eta\rangle^s$, we see that for $\alpha, \beta \in \{0,1\}$ and (x,ξ) fixed, the function $(y,\eta) \mapsto \partial_x^\beta \partial_\xi^\alpha \langle \xi+\eta\rangle^{s-m} p_i(x+y,\xi)\langle \xi\rangle^{-s}$ defines a symbol in S^{s-m}. Hence, for $\ell \in \mathbb{N}$, $1 < 2\ell - |s-m| \le 2$, $\alpha, \beta \in \{0,1\}$,

$$\begin{aligned} &\partial_x^\beta \partial_\xi^\alpha p_{0,i}(x,\xi) \\ &= \int\int e^{-iy\eta} \langle y\rangle^{-2}(1-\partial_\eta^2)\big[\langle \eta\rangle^{-2\ell}(1-\partial_y^2)^\ell \partial_x^\beta \partial_\xi^\alpha \langle \xi+\eta\rangle^{s-m} p_i(x+y,\xi)\langle \xi\rangle^{-s}\big] dy đ\eta. \end{aligned}$$

Using the imposed uniform bound on $\partial_x^k \partial_\xi^\alpha p(x,\xi)$, we obtain by treating the cases $\alpha = 0$ and $\alpha = 1$ separately,

$$\begin{aligned} &\sup_i \big|\partial_x^\beta \partial_\xi^\alpha p_{0,i}(x,\xi)\big| \\ &\le C\langle \xi\rangle^{m-s}\Big[\int \big|(1-\partial_\eta^2)\langle \eta\rangle^{-2\ell}\langle \xi+\eta\rangle^{s-m}\big| d\eta + \int \big|(1-\partial_\eta^2)\langle \eta\rangle^{-2\ell}\partial_\xi \langle \xi+\eta\rangle^{s-m}\big|\eta\Big] \\ &\le C + C\langle \xi\rangle^{m-s}\Big[\int \big|\partial_\eta^2\langle \eta\rangle^{-2\ell}\langle \xi+\eta\rangle^{s-m}\big| d\eta + \int \big|\partial_\eta^2\langle \eta\rangle^{-2\ell}\partial_\xi \langle \xi+\eta\rangle^{s-m}\big|\eta\Big] \end{aligned}$$

using Peetre's inequality again and $2\ell > 1 + |s-m|$ for the second estimate. Since for $q \in \mathbb{R}$, $\langle \xi\rangle^q \in S^q$, it follows that $|\partial_\xi^\alpha \langle \xi\rangle^q| \lesssim \langle \xi\rangle^{q-\alpha}$ and since $\langle . \rangle \ge 1$,

$$\partial_\eta^2 \langle \eta\rangle^{-2\ell}\langle \xi+\eta\rangle^{s-m} \lesssim \sum_{k=0}^{2} \langle \eta\rangle^{-2\ell-k}\langle \xi+\eta\rangle^{s-m-2+k} \lesssim \langle \eta\rangle^{-2\ell}\langle \xi+\eta\rangle^{s-m}.$$

Similar for the second term. Application of Peetre's inequality as above completes the proof. □

Lemma B.2. *Work under the assumptions of Theorem 2. If $v_{t,h}$ is given as in (3.6), then,*

$$\|v_{t,h}\|_2 \gtrsim h^{1/2-m-r}.$$

Proof. We only discuss the case $\gamma > 0$. If $\gamma = 0$ the proof can be done similarly. It follows from the definition that

$$\|v_{t,h}\|_2^2 = \int \frac{1+|s|^{2\gamma}}{|\mathcal{F}(f_\epsilon)(-s)|^2}\big|\mathcal{F}\big(\mathrm{Op}(a^\star)(\phi\circ S_{t,h})\big)(s)\big|^2 ds - \Big\|\frac{\mathcal{F}\big(\mathrm{Op}(a^\star)(\phi\circ S_{t,h})\big)}{\mathcal{F}(f_\epsilon)(-\cdot)}\Big\|_2^2.$$

Since the adjoint is given by $a^\star(x,\xi) = e^{\partial_x \partial_\xi}\overline{a}(x,\xi)$ in the sense of asymptotic summation, it follows immediately that $a^\star(x,\xi) = \overline{a}(x,\xi) + r(x,\xi)$ with $r \in S^{\overline{m}-1}$. From this we conclude that $\operatorname{Op}(a^\star)$ is an elliptic pseudo-differential operator. Because of $a^\star \in S^{\overline{m}}$ and ellipticity there exists a so called left parametrix $(a^\star)^{-1} \in S^{-\overline{m}}$ such that $\operatorname{Op}((a^\star)^{-1})\operatorname{Op}(a^\star) = 1 + \operatorname{Op}(a')$ and $a' \in S^{-\infty}$, where $S^{-\infty} = \bigcap_m S^m$ (cf. Theorem 18.1.9 in Hörmander [25]). In particular, $a' \in S^{-1}$. Moreover, $\operatorname{Op}((a^\star)^{-1}) : H^{r+\gamma} \to H^{r+m}$ is a continuous and linear and therefore bounded operator (cf. Lemma B.1). Furthermore, by convexity, $1 + |s|^{2\gamma} \geq 2\langle s\rangle^{2\gamma}$ and there exists a finite constant $c > 0$ such that

$$\begin{aligned}
&\int \frac{1+|s|^{2\gamma}}{|\mathcal{F}(f_\epsilon)(-s)|^2}\big|\mathcal{F}\big(\operatorname{Op}(a^\star)(\phi\circ S_{t,h})\big)(s)\big|^2 ds \\
&\quad \geq 2C_l^2 \|\operatorname{Op}(a^\star)(\phi\circ S_{t,h})\|_{H^{r+\gamma}}^2 \gtrsim \|\operatorname{Op}((a^\star)^{-1})\operatorname{Op}(a^\star)(\phi\circ S_{t,h})\|_{H^{r+m}}^2 \\
&\quad = \|(1+\operatorname{Op}(a'))(\phi\circ S_{t,h})\|_{H^{r+m}}^2 \geq \big(\|\phi\circ S_{t,h}\|_{H^{r+m}} - \|\operatorname{Op}(a')(\phi\circ S_{t,h})\|_{H^{r+m}}\big)^2 \\
&\quad \geq \big(\|\phi\circ S_{t,h}\|_{H^{r+m}} - c\|\phi\circ S_{t,h}\|_{H^{r+m-1}}\big)^2 \\
&\quad \geq h\int \big(1+\big|\tfrac{s}{h}\big|^2\big)^{m+r}\big|\mathcal{F}(\phi)(s)\big|^2 ds + O(h^{2(1-r-m)}) \\
&\quad \geq h^{1-2(r+m)}\int |s|^{2m+2r}\big|\mathcal{F}(\phi)(s)\big|^2 ds + O(h^{2(1-r-m)}).
\end{aligned}$$

On the other hand, we see immediately that

$$\left\|\frac{\mathcal{F}\big(\operatorname{Op}(a^\star)(\phi\circ S_{t,h})\big)}{\mathcal{F}(f_\epsilon)(-\cdot)}\right\|_2^2 \lesssim \|\operatorname{Op}(a^\star)(\phi\circ S_{t,h})\|_{H^r}^2 \lesssim \|\phi\circ S_{t,h}\|_{H^{r+\overline{m}}}^2 \lesssim h^{1-2(r+\overline{m})}.$$

Since $\phi \in L^2$ and h tends to zero the claim follows. □

Note that for bounded intervals $[a,b]$, partial integration holds $\int_a^b f'g = fg|_a^b - \int_a^b fg'$ whenever f and g are absolute continuous on $[a,b]$. As a direct consequence, we have $\int_{\mathbb{R}} f'g = -\int_{\mathbb{R}} fg'$ if f' and g' exist and $fg, f'g, fg' \in L^1$.

In order to formulate the key estimate for proving Theorems 2 and 3, let us introduce for fixed ϕ, a generic symbol $\mathfrak{a}^{(t,h)} \in S^{\overline{m}}$, and $\lambda = \lambda_\gamma^\mu$ as in (3.7)

$$(K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)})(u) = h^{-\overline{m}}\int \lambda\big(\tfrac{s}{h}\big)\mathcal{F}\big(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)(s)e^{is(u-t)/h}\,\text{đ}s. \tag{B.2}$$

From the context it will be always clear to which ϕ the operator $K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)}$ refers to. To simplify the expressions we do not indicate the dependence on ϕ and f_ϵ explicitly.

Remark B.1. *Recall* (A.5) *and note that if $a \in S^{\overline{m}}$ then also $a_{t,h}^\star \in S^{\overline{m}}$. Due to*

$$\big(\operatorname{Op}(a_{t,h}^\star)\phi\big)\circ S_{t,h} = h^{-\overline{m}}\operatorname{Op}(a^\star)(\phi\circ S_{t,h})$$

we obtain for $v_{t,h}$ in (3.6) *the representation,*

$$v_{t,h}(u) = h^{-\overline{m}} \int \lambda_\gamma^\mu\big(\tfrac{s}{h}\big)\mathcal{F}\big(\operatorname{Op}(a_{t,h}^\star)\phi\big)(s)e^{is(u-t)/h}đs = \big(K_{t,h}^{\gamma,\overline{m}} a_{t,h}^\star\big)(u). \tag{B.3}$$

Lemma B.3. *For $\mathfrak{a}^{(t,h)} \in S^{\overline{m}}$ and $\gamma + \overline{m} = m$ let $K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)}$ be as defined in* (B.2)*. Work under Assumption 2 and suppose that*

(i) $\phi \in H_4^q$ *with* $q > m + r + 3/2$,

(ii) $\gamma \in \{0\} \cup [1, \infty)$, *and*

(iii) for $k \in \mathbb{N}$, $\alpha \in \{0, 1, \ldots, 5\}$, there exist finite constants C_k, such that

$$\sup_{(t,h)\in\mathcal{T}} \big|\partial_x^k \partial_\xi^\alpha \mathfrak{a}^{(t,h)}(x, \xi)\big| \le C_k(1 + |\xi|)^{\overline{m}}, \quad \textit{for all } x, \xi \in \mathbb{R}.$$

Then, there exists a constant $C = C(q, r, \gamma, \overline{m}, C_l, C_u, \max_{k\le 4q} C_k)$ (C_l and C_u as in Assumption 2) such that for $(t, h) \in \mathcal{T}$,

(i) $|(K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)})(u)| \le C\|\phi\|_{H_4^q} h^{-m-r} \min\big(1, \frac{h^2}{(u-t)^2}\big)$,

(ii) $|(K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)})(u) - (K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)})(u')| \le C\|\phi\|_{H_4^q} h^{-m-r-1}|u - u'|$ *and for* $u, u' \ne t$,

$$\begin{aligned}|(K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)})(u) - (K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)})(u')| &\le C\|\phi\|_{H_4^q} \frac{h^{1-m-r}|u - u'|}{|u' - t|\ |u - t|} \\ &= C\|\phi\|_{H_4^q} h^{1-m-r}\Big|\int_{u'}^{u} \frac{1}{(x - t)^2}dx\Big|.\end{aligned}$$

Proof. During this proof, $C = C(q, r, \gamma, \overline{m}, C_l, C_u, \max_{k\le 4q} C_k)$ denotes an unspecified constant which may change in every line. The proof relies essentially on the well-known commutator relation for pseudo-differential operators $[x, \operatorname{Op}(p)] = i \operatorname{Op}(\partial_\xi p)$, with $\partial_\xi p : (x, \xi) \mapsto \partial_\xi p(x, \xi)$ (cf. Theorem 18.1.6 in [25]). By induction for $k \in \mathbb{N}$,

$$x^k \operatorname{Op}(\mathfrak{a}^{(t,h)}) = \sum_{r=0}^{k} \binom{k}{r} i^r \operatorname{Op}\big(\partial_\xi^r \mathfrak{a}^{(t,h)}\big)x^{k-r}. \tag{B.4}$$

As a preliminary result, let us show that for $k = 0, 1, 2$ the L^1-norms of

$$\langle s\rangle\ D_s^k\ \lambda\big(\tfrac{s}{h}\big)\mathcal{F}(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi)(s), \tag{B.5}$$

are bounded by $C\|\phi\|_{H_2^q} h^{-r-\gamma}$. Using Assumption 2 and Lemma B.1 this follows immediately for $k = 0$ and $q > r + m + 3/2$ by

$$\begin{aligned}\int \Big|\langle s\rangle\ \lambda\big(\tfrac{s}{h}\big)\mathcal{F}(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi)(s)\Big|ds &\le C_l^{-1}h^{-r-\gamma}\big\|\langle\cdot\rangle^{1+r+\gamma}\ \mathcal{F}(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi)\big\|_1 \\ &\le Ch^{-r-\gamma}\big\|\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big\|_{H^{q-\overline{m}}} \le Ch^{-r-\gamma}\big\|\phi\big\|_{H^q}.\end{aligned} \tag{B.6}$$

Now, $\mathfrak{a}^{(t,h)} \in S^{\overline{m}}$ implies that for $k \in \mathbb{N}$, $\partial_\xi^k \mathfrak{a}^{(t,h)} \in S^{\overline{m}-k} \subset S^{\overline{m}}$. Since by (B.4), Assumptions (i) and (iii), and Lemma B.1,

$$\|\langle x\rangle^2 \operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\|_1 \lesssim \|(1+x^4)\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\|_2 \le C\|\phi\|_{H_4^{\overline{m}}} < \infty, \tag{B.7}$$

we obtain for $j \in \{1,2\}$,

$$D_s^j \mathcal{F}(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi) = (-i)^j \mathcal{F}(x^j \operatorname{Op}(\mathfrak{a}^{(t,h)})\phi)(s)$$

by interchanging differentiation and integration. Explicit calculations thus show

$$D_s\lambda\big(\tfrac{s}{h}\big)\mathcal{F}\big(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)(s) = \big(D_s\lambda\big(\tfrac{s}{h}\big)\big)\mathcal{F}\big(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)(s) - i\lambda\big(\tfrac{s}{h}\big)\mathcal{F}\big(x\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)(s)$$

and

$$\begin{aligned} D_s^2\lambda\big(\tfrac{s}{h}\big)\mathcal{F}\big(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)(s) = \big(D_s^2\lambda\big(\tfrac{s}{h}\big)\big)\mathcal{F}\big(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)(s) - 2i\big(D_s\lambda\big(\tfrac{s}{h}\big)\big)\mathcal{F}\big(x\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)(s) \\ - \lambda\big(\tfrac{s}{h}\big)\mathcal{F}\big(x^2\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)(s). \end{aligned} \tag{B.8}$$

To finish the proof of (B.5) let us distinguish two cases, namely (I) $\gamma \in \{0\} \cup [2,\infty)$ and (II) $\gamma \in (1,2)$.

(I): For $k = 0,1,2$ and $s \neq 0$, we see by elementary calculations, $\big|\langle s\rangle D_s^k \lambda\big(\tfrac{s}{h}\big)\big| \le Ch^{-r-\gamma}\langle s\rangle^{r+\gamma+1}$. Using (B.4) and arguing similar as for (B.6) we obtain (replacing ϕ by $x\phi$ or $x^2\phi$ if necessary) bounds of the L^1-norms which are of the correct order $\|\phi\|_{H_4^q} h^{-r-\gamma}$.

(II): In principal we use the same arguments as in (I) but a singularity appears by expanding the first term on the r.h.s. of (B.8). In fact, it is sufficient to show that

$$\begin{aligned} \int_{-1}^{1} \Big|\frac{D_s^2\big|\frac{s}{h}\big|^\gamma \iota_s^{-\mu}}{\mathcal{F}(f_\epsilon)\big(-\frac{s}{h}\big)}\mathcal{F}\big(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)(s)\Big| ds \le C_l h^{-r-\gamma}\big\|\mathcal{F}\big(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)\big\|_\infty \int_1^1 |s|^{\gamma-2} ds \\ \lesssim C_l h^{-r-\gamma}\big\|\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big\|_1 \le Ch^{-r-\gamma}\|\phi\|_{H_4^{\overline{m}}}, \end{aligned}$$

where the last inequality follows from (B.7). Since this has the right order $h^{-r-\gamma}\|\phi\|_{H_4^q}$, (B.5) follows for $\gamma \in (1,2)$.

Together (I) and (II) prove (B.5). Hence we can apply partial integration twice and obtain for $t \neq u$,

$$(K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)})(u) = -\frac{h^{2-\overline{m}}}{(u-t)^2}\int e^{is(u-t)/h}\, D_s^2\lambda\big(\tfrac{s}{h}\big)\mathcal{F}\big(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)(s)\bar{d}s \tag{B.9}$$

and similarly, first interchanging integration and differentiation,

$$\begin{aligned} D_u(K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)})(u) &= ih^{-\overline{m}-1}\int e^{is(u-t)/h} s\lambda\big(\tfrac{s}{h}\big)\mathcal{F}\big(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)(s)\bar{d}s \\ &= -\frac{ih^{1-\overline{m}}}{(u-t)^2}\int e^{is(u-t)/h} D_s^2 s\lambda\big(\tfrac{s}{h}\big)\mathcal{F}\big(\operatorname{Op}(\mathfrak{a}^{(t,h)})\phi\big)(s)\bar{d}s \end{aligned} \tag{B.10}$$

(i): The estimates $|(K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)})(u)| \le C\|\phi\|_{H_4^q} h^{-m-r}$ and $|(K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)})(u)| \le C\|\phi\|_{H_4^q} h^{2-m-r}/(u-t)^2$ follow directly from (B.6) as well as (B.9) together with the L^1 bound of (B.5) for $k=2$.

(ii): To prove $|(K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)})(u) - (K_{t,h}^{\gamma,\overline{m}}\mathfrak{a}^{(t,h)})(u')| \le C\|\phi\|_{H_4^q} h^{-m-r-1}|u-u'|$ it is enough to note that $|e^{ix}-e^{iy}| \le |x-y|$. The result then follows from (B.6) again. For the second bound, see (B.10). The estimate for the L^1-norm of (B.5) with $k=2$ completes the proof. □

Let $\lceil x \rceil$ be the smallest integer which is not smaller than x.

Lemma B.4. *Let $0 \le \ell \le 1/2$ and $q \ge 0$. Assume that $\phi \in H^{\lceil q \rceil} \cap H^{q+\ell}$, $\operatorname{supp}\phi \subset [0,1]$ and $\mathrm{TV}(D^{\lceil q \rceil}\phi) < \infty$. Then, for $h \le h'$,*

$$\|\phi \circ S_{t,h} - \phi \circ S_{t',h'}\|_{H^q} \lesssim h^{-q}\sqrt{|t-t'|^{2\ell} + |h'-h|}.$$

In particular, for $\phi \in H^{\lceil r+m \rceil} \cap H^{r+m+1/2}$, $\operatorname{supp}\phi \subset [0,1]$ and $\mathrm{TV}(D^{\lceil r+m \rceil}\phi) < \infty$, $h \le h'$,

$$\|v_{t,h} - v_{t',h'}\|_2 \lesssim h^{-r-m}\sqrt{|t-t'| + |h'-h|}.$$

Proof. Since

$$\big\|\phi \circ S_{t,h} - \phi \circ S_{t',h'}\big\|_{H^q}^2 \lesssim \int \langle s\rangle^{2q}\big|1-e^{is(t-t')}\big|^2\big|\mathcal{F}\big(\phi\big(\tfrac{\cdot}{h}\big)\big)(s)\big|^2 ds + \big\|\phi\big(\tfrac{\cdot}{h}\big) - \phi\big(\tfrac{\cdot}{h'}\big)\big\|_{H^q}^2$$

and $|1-e^{is(t-t')}| \le 2\min(|s||t-t'|, 1) \le 2\min(|s|^\ell|t-t'|^\ell, 1) \le 2|s|^\ell|t-t'|^\ell$, we obtain (note that $\phi \in H^{q+\ell}$)

$$\big\|\phi \circ S_{t,h} - \phi \circ S_{t',h'}\big\|_{H^q}^2 \lesssim |t-t'|^{2\ell}h^{1-2q-2\ell} + \big\|\phi\big(\tfrac{\cdot}{h}\big) - \phi\big(\tfrac{\cdot}{h'}\big)\big\|_{H^q}^2.$$

Set $k = \lceil q \rceil$. Then,

$$\begin{aligned}\big\|\phi\big(\tfrac{\cdot}{h}\big) - \phi\big(\tfrac{\cdot}{h'}\big)\big\|_{H^{r+m}}^2 &\lesssim h^{1-2q}\big\|\phi - \phi\big(\tfrac{h}{h'}\,\cdot\big)\big\|_{H^q}^2 \\ &\lesssim h^{1-2q}\big\|\phi - \phi\big(\tfrac{h}{h'}\,\cdot\big)\big\|_2^2 + h^{1-2q}\big\|D^k\big(\phi - \phi\big(\tfrac{h}{h'}\,\cdot\big)\big)\big\|_2^2.\end{aligned}$$

For $j \in \{0, k\}$,

$$\begin{aligned}\big\|D^j\big(\phi - \phi\big(\tfrac{h}{h'}\,\cdot\big)\big)\big\|_2^2 &\le 2\big\|\phi^{(j)} - \phi^{(j)}\big(\tfrac{h}{h'}\,\cdot\big)\big\|_2^2 + 2\big(1-\big(\tfrac{h}{h'}\big)^j\big)^2\big\|\phi^{(j)}\big(\tfrac{h}{h'}\,\cdot\big)\big\|_2^2 \\ &\lesssim h^{-1}\big\|\phi^{(j)}\big(\tfrac{\cdot}{h}\big) - \phi^{(j)}\big(\tfrac{\cdot}{h'}\big)\big\|_2^2 + |h'-h|\ h^{-1}\|\phi^{(j)}\|_2^2.\end{aligned}$$

Now, application of Lemma C.5 completes the proof for the first part. The second claim follows from

$$\begin{aligned}\|v_{t,h} - v_{t',h'}\|_2^2 &= \int |\lambda(s)|^2\big|\mathcal{F}\big(\operatorname{Op}(a^\star)(\phi \circ S_{t,h} - \phi \circ S_{t',h'})\big)(s)\big|^2 ds \\ &\lesssim \big\|\phi \circ S_{t,h} - \phi \circ S_{t',h'}\big\|_{H^{r+m}}^2.\end{aligned}$$

□

Lemma B.5. *Let $A_{t,t',h,h'}$ be defined as in* (A.3) *and work under Assumption 1. Then, for a global constant* $K > 0$,

$$A_{t,t',h,h'} \leq K\sqrt{|t-t'|+|h-h'|}.$$

Proof. Without loss of generality, assume that for fixed (t,h), $V_{t,h} \geq V_{t',h'}$. We can write

$$\begin{aligned} A_{t,t',h,h'} &\leq \frac{\|\psi_{t,h}\sqrt{h}-\psi_{t',h'}\sqrt{h'}\|_2}{V_{t,h}} + \sqrt{h'}\|\psi_{t',h'}\|_2\Big|\frac{1}{V_{t,h}} - \frac{1}{V_{t',h'}}\Big| \\ &\leq \frac{\|\psi_{t,h}\sqrt{h}-\psi_{t',h'}\sqrt{h'}\|_2}{V_{t,h}} + \sqrt{h'}\frac{|V_{t,h}-V_{t',h'}|}{V_{t,h}}. \end{aligned}$$

By triangle inequality, $\|\psi_{t,h}\sqrt{h}-\psi_{t',h'}\sqrt{h'}\|_2 \leq \sqrt{h'}\|\psi_{t,h}-\psi_{t',h'}\|_2 + |\sqrt{h}-\sqrt{h'}|\ \|\psi_{t,h}\|_2$. Thus,

$$A_{t,t',h,h'} \leq \frac{\sqrt{h'}}{V_{t,h}}\Big(\|\psi_{t,h}-\psi_{t',h'}\|_2 + |V_{t,h}-V_{t',h'}|\Big) + \sqrt{|h-h'|}.$$

If $h' \leq h$, then the result follows by Assumption 1 (iv) and some elementary computations. Otherwise we can estimate $\sqrt{h'} \leq \sqrt{|h-h'|}+\sqrt{h}$ and so

$$A_{t,t',h,h'} \leq \frac{\sqrt{h}}{V_{t,h}}\Big(\|\psi_{t,h}-\psi_{t',h'}\|_2 + |V_{t,h}-V_{t',h'}|\Big) + 5\sqrt{|h-h'|}.$$

□

The next lemma extends a well-known bound for functions with compact support to general càdlàg functions. We found this result useful for estimating the supremum over a Gaussian process if entropy bounds are difficult.

Lemma B.6. *Let $(W_t)_{t\in\mathbb{R}}$ denote a two-sided Brownian motion. For $\alpha > 1/2$, a family of real-valued càdlàg functions $\{f_i|\ i\in I\}$, and a constant C_α, we have*

$$\sup_{i\in I}\Big|\int f_i(s)dW_s\Big| \leq C_\alpha \sup_{s\in[0,1]}|\overline{W}_s| \sup_{i\in I}\mathrm{TV}(\widetilde{f}_i),$$

where $\overline{W}$ is a standard Brownian motion on the same probability space and $\widetilde{f}_i(s) = f_i(s)\langle s\rangle^\alpha$.

Proof. The proof consists of two steps. First suppose that $\bigcup_{i\in I}\operatorname{supp} f_i \subset [0,1]$ and assume that the f_i are of bounded variation. Then, for all $i\in I$, there exists a function q_i with $\|q_i\|_\infty \leq \mathrm{TV}(f_i)$ 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 $u \in \mathbb{R}$, because f_i is càdlàg and thus $f_i(1) = 0$. With probability one,

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

Now let us consider the general case. If $C_\alpha := \|\langle\cdot\rangle^{-\alpha}\|_2$ then $h(s) = C_\alpha^{-2}\langle s\rangle^{-2\alpha}$ is a density of a random variable. Let H be the corresponding distribution function. Note that

$$(\overline{W}_t)_{t\in[0,1]} = \Big(\int_0^t \sqrt{h(H^{-1}(s))}dW_{H^{-1}(s)}\Big)_{t\in[0,1]}$$

is a standard Brownian motion satisfying $d\overline{W}_{H(s)} = \sqrt{h(s)}dW_s$ and thus

$$\sup_{i\in I}\Big|\int f_i(s)dW_s\Big| = C_\alpha \sup_{i\in I}\Big|\int \widetilde{f}_i(s)d\overline{W}_{H(s)}\Big| = C_\alpha \sup_{i\in I}\Big|\int_0^1 \widetilde{f}_i(H^{-1}(s))d\overline{W}_s\Big|.$$

Since $\mathrm{TV}(\widetilde{f}_i \circ H^{-1}) = \mathrm{TV}(\widetilde{f}_i)$ the result follows from the first part. □

Remark 3. *For the proofs of the subsequent lemmas, we make often use of elementary facts related to the function* $\langle\cdot\rangle^\alpha \in S^\alpha$ *with* $0 < \alpha < 1$*. Note that for* $t \in [0,1]$*,* $D_u\langle u\rangle^\alpha \leq \alpha\langle u\rangle^{\alpha-1} \in S^{\alpha-1}$*,* $D_u\langle u\rangle^\alpha \leq \alpha$*,*

$$\langle u\rangle^\alpha \leq \frac{1}{2}(1+|u|^\alpha) \leq 1 + |u-t|^\alpha, \quad \textit{and} \quad \langle u\rangle^{\alpha-1} \leq 2|u-t|^{\alpha-1}, \tag{B.11}$$

where the last inequality follows from $|u-t|^{1-\alpha}\langle u\rangle^{\alpha-1} \leq |u|^{1-\alpha}\langle u\rangle^{\alpha-1} + 1 \leq 2$.

Lemma B.7. *For* $(t,h) \in \mathcal{T}$ *let* $r_{t,h}$ *be a function satisfying the conclusions of Lemma B.3 for* r, m *and* ϕ*. Assume* $1/2 < \alpha < 1$*. Then, there exists a constant* K *independent of* $(t,h) \in \mathcal{T}$ *and* ϕ *such that*

$$\big|r_{t,h}(u)\langle u\rangle^\alpha - r_{t,h}(u')\langle u'\rangle^\alpha\big| \leq K\|\phi\|_{H_4^q}h^{1-m-r}\Big|\int_{u'}^{u} \frac{1}{(x-t)^{2-\alpha}} + \frac{1}{(x-t)^2}dx\Big|,$$

for all $u, u' \neq t$ *and*

$$\mathrm{TV}\left(r_{t,h}\langle\cdot\rangle^\alpha \mathbb{I}_{[t-1,t+1]}\right) \leq K\|\phi\|_{H_4^q}h^{-m-r},$$
$$\mathrm{TV}\left(r_{t,h}\langle\cdot\rangle^\alpha \mathbb{I}_{\mathbb{R}\setminus[t-1,t+1]}\right) \leq K\|\phi\|_{H_4^q}h^{1-m-r}.$$

Proof. Let C be as in Lemma B.3. In this proof $K = K(\alpha, C)$ denotes a generic constant which may change from line to line. Without loss of generality, we may assume that $|u-t| \geq |u'-t|$. Furthermore, the bound is trivial if $u' \leq t \leq u$ or $u \leq t \leq u'$. Therefore, let us assume further that $u \geq u' > t$ (the case $u \leq u' < t$ can be treated similarly). Together with the conclusions from Lemma B.3 and Remark 3 this shows that

$$\begin{aligned}\big|r_{t,h}(u)\langle u\rangle^\alpha - r_{t,h}(u')\langle u'\rangle^\alpha\big| &\leq \big|r_{t,h}(u)\big|\ \big|\langle u\rangle^\alpha - \langle u'\rangle^\alpha\big| + \langle u'\rangle^\alpha\big|r_{t,h}(u) - r_{t,h}(u')\big| \\ &\leq K\|\phi\|_{H_4^q}\Big[h^{2-m-r}\frac{1}{(u-t)^2} + h^{1-m-r}\frac{|u'-t|^\alpha+1}{|u'-t|\ |u-t|}\Big]|u-u'|.\end{aligned}$$

Clearly, the second term in the bracket dominates uniformly over $h \in (0,1]$. By Taylor expansion

$$\begin{aligned}\frac{|u-u'|}{|u'-t|^{1-\alpha}\ |u-t|} &= \frac{u-u'}{(u-t)^\alpha(u'-t)^{1-\alpha}(u-t)^{1-\alpha}} \\ &\leq \frac{(u-t)^{1-\alpha}-(u'-t)^{1-\alpha}}{(1-\alpha)(u'-t)^{1-\alpha}(u-t)^{1-\alpha}} = \int_{u'}^{u}\frac{1}{(x-t)^{2-\alpha}}dx.\end{aligned}$$

Hence

$$\frac{1}{|u'-t|\ |u-t|}|u-u'| = \Big|\int_{u'}^{u}\frac{1}{(x-t)^2}dx\Big|$$

completes the proof for the first part. For the second part decompose $r_{t,h}\mathbb{I}_{[t-1,t+1]}$ in $r_{t,h}^{(1)} = r_{t,h}\mathbb{I}_{[t-h,t+h]}$ and $r_{t,h}^{(2)} = r_{t,h}\mathbb{I}_{[t-1,t+1]} - r_{t,h}^{(1)}$. Observe that the conclusions of Lemma B.3 imply

$$\mathrm{TV}\left(r_{t,h}^{(1)}\langle\cdot\rangle^\alpha\right) \leq \|\langle\cdot\rangle^\alpha\mathbb{I}_{[t-h,t+h]}\|_\infty\,\mathrm{TV}(r_{t,h}^{(1)}) + \mathrm{TV}\left(\langle\cdot\rangle^\alpha\mathbb{I}_{[t-h,t+h]}\right)\|r_{t,h}^{(1)}\|_\infty \leq K\|\phi\|_{H_4^q}h^{-m-r}.$$

By using the first part of the lemma, we conclude that uniformly in $(t,h) \in \mathcal{T}$,

$$\mathrm{TV}\left(r_{t,h}\langle\cdot\rangle^\alpha\mathbb{I}_{[t-1,t+1]}\right) \leq \mathrm{TV}\left(r_{t,h}^{(1)}\langle\cdot\rangle^\alpha\right) + \mathrm{TV}\left(r_{t,h}^{(2)}\langle\cdot\rangle^\alpha\right) \lesssim K\|\phi\|_{H_4^q}(h^{-m-r} + h^{-m-r})$$

and also $\mathrm{TV}\left(r_{t,h}\langle\cdot\rangle^\alpha\mathbb{I}_{\mathbb{R}\setminus[t-1,t+1]}\right) \leq K\|\phi\|_{H_4^q}h^{1-m-r}$. □

Lemma B.8. *Work under Assumptions 2 and 3 and suppose that $m+r > 1/2$, $\langle x\rangle\phi \in L^1$, and $\phi \in H_1^{m+r+1}$. Let $d_{t,h}$ be as defined in* (A.14)*. Then, there exists a constant K independent of $(t,h) \in \mathcal{T}$, such that for $1/2 < \alpha < 1$,*

$$\mathrm{TV}(d_{t,h}\langle\cdot\rangle^\alpha\mathbb{I}_{[t-1,t+1]}) \leq Kh^{\beta_0\wedge(m+r)-r}\log\left(\tfrac{1}{h}\right).$$

Proof. For convenience let $\beta_0^\star := \beta_0 \wedge (m+r)$ and substitute $s \mapsto -s$ in (A.14), i.e.

$$d_{t,h}(u) := \int e^{-is(u-t)/h}\Big(\frac{1}{\mathcal{F}(f_\epsilon)(\frac{s}{h})} - A\iota_s^\rho\big|\tfrac{s}{h}\big|^r\Big)\iota_s^\mu|s|^m\mathcal{F}(\phi)(-s)\,d\!\!\!{}^-s.$$

Define

$$F_h(s) := \frac{1}{\mathcal{F}(f_\epsilon)(\frac{s}{h})} - A\iota_s^\rho\big|\tfrac{s}{h}\big|^r.$$

By Assumptions 2 and 3, we can bound the L^1-norm of

$$s \mapsto \langle s \rangle F_h(s) \iota_s^\mu |s|^m \mathcal{F}(\phi)(-s) \tag{B.12}$$

uniformly in (t,h) by $\int \langle s \rangle \langle \frac{s}{h} \rangle^{r-\beta_0} |s|^m \big|\mathcal{F}(\phi)(-s)\big| ds$. Bounding $\langle \frac{s}{h} \rangle^{r-\beta_0}$ by $\langle \frac{s}{h} \rangle^{r-\beta_0^\star}$ and considering the cases $r \le \beta_0^\star$ and $r > \beta_0^\star$ separately, we find $h^{\beta_0^\star - r} \int \langle s \rangle^{1+r+m-\beta_0^\star} |\mathcal{F}(\phi)(-s)| ds \lesssim h^{\beta_0^\star - r} \|\phi\|_{H^{r+m+1}}$ as an upper bound for (B.12), uniformly in $(t,h) \in \mathcal{T}$. Furthermore,

$$D_s F_h(s) = -\frac{D_s \mathcal{F}(f_\epsilon)(\frac{s}{h})}{\big(\mathcal{F}(f_\epsilon)(\frac{s}{h})\big)^2} - Ari\iota_s^{\rho-1} h^{-1} \big|\tfrac{s}{h}\big|^{r-1}$$

and by Assumptions 2 and 3,

$$\begin{aligned} \Big|sD_sF_h(s)\Big| &\le \big|sD_s\mathcal{F}(f_\epsilon)(\tfrac{s}{h})\big| \Big| A^2 \iota_s^{2\rho} \big|\tfrac{s}{h}\big|^{2r} - \frac{1}{\big(\mathcal{F}(f_\epsilon)(\frac{s}{h})\big)^2}\Big| \\ &\quad + |A| r \big|\tfrac{s}{h}\big|^r \Big| - A(ri)^{-1} \iota_s^{\rho+1} h \big|\tfrac{s}{h}\big|^{r+1} D_s \mathcal{F}(f_\epsilon)\big(\tfrac{s}{h}\big) - 1\Big| \\ &\lesssim \Big(\big|\tfrac{s}{h}\big| \big\langle \tfrac{s}{h} \big\rangle^{r-1} + \big|\tfrac{s}{h}\big|^r\Big) \big\langle \tfrac{s}{h} \big\rangle^{-\beta_0} \le 2 \big\langle \tfrac{s}{h} \big\rangle^{r-\beta_0^\star}. \end{aligned}$$

Similar as above, we can conclude that the L^1-norm of

$$s \mapsto D_s s F_h(s) \iota_s^\mu |s|^m \mathcal{F}(\phi)(-s)$$

is bounded by const.$\times h^{\beta_0^\star - r} \|\phi\|_{H_1^{r+m+1}}$, uniformly over all $(t,h) \in \mathcal{T}$. Therefore, we have by interchanging differentiation and integration first and partial integration,

$$\begin{aligned} D_u d_{t,h}(u) &= \frac{-i}{h} \int s e^{-is(u-t)/h} F_h(s) \iota_s^\mu |s|^m \mathcal{F}(\phi)(-s) đs \\ &= \frac{-1}{u-t} \int e^{-is(u-t)/h} D_s s F_h(s) \iota_s^\mu |s|^m \mathcal{F}(\phi)(-s) đs \end{aligned}$$

and the second equality holds for $u \ne t$. Together with (B.12) this shows that $|d_{t,h}(u)| \lesssim h^{\beta_0^\star - r}$ and $|D_u d_{t,h}(u)| \lesssim h^{\beta_0^\star - r - 1} \min(1, h/|u-t|)$. Using Remark 3 we find for the sets $A_{t,h}^{(1)} := [t-h, t+h]$ and $A_{t,h}^{(2)} := [t-1, t+1] \setminus A_{t,h}^{(1)}$,

$$\mathrm{TV}(d_{t,h} \mathbb{I}_{[t-1,t+1]}) \le 2\|d_{t,h}\|_\infty + \int_{A_{t,h}^{(1)}} |D_u d_{t,h}(u)| du + \int_{A_{t,h}^{(2)}} |D_u d_{t,h}(u)| du \lesssim h^{\beta_0^\star - r} \log\big(\tfrac{1}{h}\big).$$

Thus, $\mathrm{TV}(d_{t,h} \langle \cdot \rangle^\alpha \mathbb{I}_{[t-1,t+1]}) \lesssim \|d_{t,h}\|_\infty + \mathrm{TV}(d_{t,h} \mathbb{I}_{[t-1,t+1]}) \lesssim h^{\beta_0^\star - r} \log\big(\tfrac{1}{h}\big)$. □

Lemma B.9. *Work under the assumptions of Theorem 3 and let* $v_{t,h}^P$ *be defined as in* (A.10)*. Then, for* $1/2 < \alpha < 1$,

$$\mathrm{TV}(v_{t,h}^P \langle \cdot \rangle^\alpha \mathbb{I}_{\mathbb{R} \setminus [t-1,t+1]}) \le K h^{1-r-m},$$

where the constant K *does not depend on* (t,h).

Proof. The proof uses essentially the same arguments as the proof of Lemma B.3. Let $q := \lfloor r + m + 5/2 \rfloor$ and recall that by assumption $\langle x\rangle^2\phi \in L^1$. Decomposing the L^1 norm on $\mathbb{R}$ into $L^1([-1,1])$ and $L^1(\mathbb{R}\setminus[-1,1])$ and using Cauchy-Schwarz inequality and $\|\mathcal{F}(\phi)\|_\infty \leq \|\phi\|_1$, we see that for $j \in \{0,1\}$, the L^1-norm of $s \mapsto D_s^j|s|^{r+m}\iota_s^{-\rho-\mu}\mathcal{F}(\phi)(s)$ is bounded by const.$\times(\|\phi\|_{H_1^q} + \|\phi\|_1)$. Similar, for $k \in \{0,1,2\}$, the L^1-norms of $s \mapsto D_s^k|s|^{r+m+1}\iota_s^{-\rho-\mu+1}\mathcal{F}(\phi)(s)$ are bounded by a multiple of $\|\phi\|_{H_2^q} + \|\phi\|_1$. Hence, we have

$$v_{t,h}^P(u) = \frac{Ah^{1-r-m}ia_P(t)}{u-t}\int e^{is(u-t)/h}D_s|s|^{r+m}\iota_s^{-\rho-\mu}\mathcal{F}(\phi)(s)\bar{d}s$$

and

$$D_u v_{t,h}^P(u) = \frac{-Ah^{1-r-m}a_P(t)}{(u-t)^2}\int e^{is(u-t)/h}D_s^2|s|^{r+m+1}\iota_s^{-\rho-\mu+1}\mathcal{F}(\phi)(s)\bar{d}s.$$

Together with Remark 3 this shows that

$$\begin{aligned}\mathrm{TV}\left(v_{t,h}^P\langle\cdot\rangle^\alpha \mathbb{I}_{[t+1,\infty)}\right) &\leq \|v_{t,h}^P\langle\cdot\rangle^\alpha\mathbb{I}_{[t+1,\infty)}\|_\infty + \int_{t+1}^\infty |D_u v_{t,h}^P(u)\langle\cdot\rangle^\alpha|du \\ &\lesssim h^{1-r-m} + \int_{t+1}^\infty \frac{h^{1-r-m}}{|u-t|^{2-\alpha}} + \frac{h^{1-r-m}}{|u-t|^2}du \lesssim h^{1-r-m}.\end{aligned}$$

Similar, we can bound the total variation on $(-\infty, t-1]$. □

Appendix C Further technicalities

Lemma C.1. *Assume that $K_n \to \infty$, $\psi_{t,h} = \psi\big(\frac{\cdot-t}{h}\big)$ and $V_{t,h} = \|\psi_{t,h}\|_2 = \sqrt{h}\|\psi\|_2$. Suppose that $\lim_{j\to\infty}\log(j)|\int\psi(s-j)\psi(s)ds| \to 0$. Then, with w_h and $B_{K_n}^\circ$ as defined in (2.4) and (2.8), respectively,*

$$\sup_{(t,h)\in B_{K_n}^\circ} w_h\left(\frac{|\int\psi_{t,h}(s)dW_s|}{\|\psi_{t,h}\|_2} - \sqrt{2\log\tfrac{\nu}{h}}\right) \to -\frac{1}{4}, \quad \textit{in probability.}$$

Proof. Write $K := K_n$ and let $\xi_j := \|\psi_{t,h}\|_2^{-1}\int\psi_{j/K,1/K}(s)dW_s$ for $j = 0,\ldots,K-1$. Now, $(\xi_j)_j$ is a stationary sequence of centered and standardized normal random variables. In particular the distribution of $(\xi_j)_j$ does not depend on K and the covariance decays by assumption at a faster rate than logarithmically. By Theorem 4.3.3 (ii) in [34] the maximum behaves as the maximum of K independent standard normal r.v., i.e.

$$\mathbb{P}\big(\max(\xi_1,\ldots,\xi_K) \leq a_K + b_K t\big) \to \exp\big(-e^{-t}\big), \quad \text{for } t\in\mathbb{R} \text{ and } K\to\infty,$$

where

$$b_K := \frac{1}{\sqrt{2\log K}}, \quad \text{and} \quad a_K = \sqrt{2\log K} - \frac{\log\log K + \log(4\pi)}{\sqrt{8\log K}}.$$

Using the tail-equivalence criterion (cf. [14], Proposition 3.3.28), we obtain further

$$\lim_{K\to\infty} \mathbb{P}\big(\max(|\xi_1|,\ldots,|\xi_K|) \le a_K + b_K(t+\log 2)\big) = \exp\big(-e^{-t}\big), \quad \text{for } t\in\mathbb{R}.$$

Note that $T_n^\circ := \sup_{(t,h)\in B_n^\circ} w_h(\|\psi_{t,h}\|_2^{-1}|\int \psi_{t,h}(s)dW_s| - \sqrt{2\log(\nu/h)})$ has the same distribution as $w_{K^{-1}}\max(|\xi_1|,\ldots,|\xi_K|) - w_{K^{-1}}\sqrt{2\log(\nu K)}$. It is easy to show that

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

and

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

Assume that $\eta_n \to 0$ and $\eta_n \log\log K \to \infty$. Then for sufficiently large n,

$$\begin{aligned}
\mathbb{P}\big(T_n^\circ > -\tfrac{1}{4} + \eta_n\big) &= \mathbb{P}\Big(\max(|\xi_1|,\ldots,|\xi_K|) > \big(-\tfrac{1}{4}+\eta_n\big)/w_{K^{-1}} + \sqrt{2\log \nu K}\Big) \\
&= \mathbb{P}\Big(\max(|\xi_1|,\ldots,|\xi_K|) > \\
&\qquad \big(-1+4\eta_n\big)\frac{\log\log K}{\sqrt{8\log K}} + \sqrt{2\log K} + \frac{\log \nu}{\sqrt{2\log K}} + O\big(\frac{\log\log K}{\log^{3/2} K}\big)\Big) \\
&\le \mathbb{P}\Big(\max(|\xi_1|,\ldots,|\xi_K|) > a_K + b_K 2\eta_n \log\log K\Big) \to 0.
\end{aligned}$$

Similarly,

$$\mathbb{P}\big(T_n^\circ \le -\tfrac{1}{4} - \eta_n\big) \le \mathbb{P}\Big(\max(|\xi_1|,\ldots,|\xi_K|) \le a_K - b_K\eta_n \log\log K\Big) \to 0.$$

□

Lemma C.2. *Condition (iii) in Assumption 1 is fulfilled with* $\kappa_n = w_{u_n} u_n^{1/2}$, *whenever Condition (ii) of Assumption 1 holds and for all* $(t,h)\in B_n$, $\operatorname{supp}\psi_{t,h} \subset [t-h, t+h]$.

Proof. Let $1/2 < \alpha < 1$. Then $\langle\cdot\rangle^\alpha : \mathbb{R}\to\mathbb{R}$ is Lipschitz. Recall that $\mathrm{TV}(fg) \le \|f\|_\infty \mathrm{TV}(g) + \|g\|_\infty \mathrm{TV}(f)$. Since $\bigcup_{(t,h)\in B_n} \operatorname{supp}\psi_{t,h} \subset [-1,2]$ is bounded and contains the support of all functions $s\mapsto \psi_{t,h}(s)\big[\sqrt{g(s)} - \sqrt{g(t)}\big]\langle s\rangle^\alpha$ (indexed in $(t,h)\in B_n$), we obtain uniformly over $(t,h)\in B_n$ and $G\in\mathcal{G}$,

$$\begin{aligned}
&\mathrm{TV}\left(\psi_{t,h}(\cdot)\big[\sqrt{g(\cdot)} - \sqrt{g(t)}\big]\langle\cdot\rangle^\alpha\right) \\
&\quad \lesssim \big\|\psi_{t,h}(\cdot)\big[\sqrt{g(\cdot)} - \sqrt{g(t)}\big]\big\|_\infty + \mathrm{TV}\left(\psi_{t,h}(\cdot)\big[\sqrt{g(\cdot)} - \sqrt{g(t)}\big]\right)
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\mathrm{TV}\left(\psi_{t,h}(\cdot)\big[\sqrt{g(\cdot)}-\sqrt{g(t)}\big]\right)\\
&\le \|\psi_{t,h}\|_\infty \,\mathrm{TV}\left(\big[\sqrt{g(\cdot)}-\sqrt{g(t)}\big]\mathbb{I}_{[t-h,t+h]}(\cdot)\right)+\mathrm{TV}\left(\psi_{t,h}\right)\big\|\big[\sqrt{g(\cdot)}-\sqrt{g(t)}\big]\mathbb{I}_{[t-h,t+h]}(\cdot)\big\|_\infty\\
&\lesssim V_{t,h}h^{1/2},
\end{aligned}$$

where the last inequality follows from Assumption 1 (ii) as well as the properties of $\mathcal{G}$. With Lemma C.3 (ii) the result follows. □

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

Lemma C.3. *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 $x(h) \ge 2$, i.e. in $h \le \nu\exp(e^{-2})$. 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 C.4. *Suppose that* $\operatorname{supp} f \subset [0,\infty)$ *and let* $0 \le a \le 1$. *Then,*

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

and

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

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 increasing 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

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

In order to derive the second inequality, note that

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

□

Lemma C.5. *Suppose that* $\operatorname{supp}\psi \subset [0,1]$ *and* $\mathrm{TV}(\psi) < \infty$*. Let* $(t,h) \in \mathcal{T}$*. Then, there exists a constant* K *only depending on* ψ*, such that*

$$\left\|\psi\big(\tfrac{\cdot - t}{h}\big) - \psi\big(\tfrac{\cdot - t'}{h'}\big)\right\|_2 \leq K\sqrt{|h-h'| + |t-t'|}.$$

Proof. Note that

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

Without loss of generality assume $h' \leq h$. Using Lemma C.4 yields

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

Similarly, assuming $t \leq t'$,

$$\int_0^2 \left|\psi\big(\tfrac{s-t}{h'}\big) - \psi\big(\tfrac{s-t'}{h'}\big)\right| ds = h' \int_0^{(t'-t)/h'+1} \big|\psi(u) - \psi\big(u - \tfrac{t'-t}{h'}\big)\big| du \leq |t'-t|\,\mathrm{TV}\,(\psi).$$

□