Adaptive procedure for Fourier estimators: application to deconvolution and decompounding

Céline Duval

MAP5 UMR CNRS 8145
e-mail: celine.duval@parisdescartes.fr

and

Johanna Kappus

Institut für Mathematik, Universität Rostock
e-mail: johanna.kappus@uni-rostock.de

Abstract: The purpose of this paper is twofold. First, introduce a new adaptive procedure to select the optimal – up to a logarithmic factor – cutoff parameter for Fourier density estimators. Two inverse problems are considered: deconvolution and decompounding. Deconvolution is a typical inverse problem, for which our procedure is numerically simple and stable, a comparison is performed with penalized techniques. Moreover, the procedure and the proof of oracle bounds do not rely on any knowledge on the noise term. Second, for decompounding, *i.e.* non-parametric estimation of the jump density of a compound Poisson process from the observation of n increments at timestep Δ, build an unified adaptive estimator which is optimal – up to a logarithmic factor – regardless the behavior of Δ.

MSC 2010 subject classifications: Primary 62C12, 62C20; secondary 62G07.
Keywords and phrases: Adaptive density estimation, deconvolution, decompounding, model selection.

1. Introduction and motivations

1.1. Adaptive procedure

In the literature on non-parametric statistics a lot of space is dedicated to adaptive procedures. Adaptivity may be understood as minimax-adaptivity, *i.e.* optimal rates of convergence are attained simultaneously over a collection of class of densities, such as Sobolev-balls. Adaptivity may also refer to proving non-asymptotic oracle bounds, *i.e.* having a procedure that mimics, up to a constant, the estimator that minimizes a given loss function. It is this last notion of adaptivity we adopt here.

Hereafter, we propose an approach that is relevant for inverse problems when the estimator relies on Fourier techniques. This method is inspired from the one introduced in Duval and Kappus [20] and is generalized to deconvolution and decompounding inverse problems. We present this procedure below in a general context, even though in this article we study oracle bounds for two specific inverse problems; deconvolution and decompounding.

Heuristic of the adaptive procedure

Notations. We first introduce some notations which are used throughout the rest of the text. Given a random variable Z, $\varphi_Z(u) = \mathbb{E}[e^{iuZ}]$ denotes the characteristic function of Z. For $f \in \mathrm{L}^1(\mathbb{R})$, $\mathcal{F}f(u) = \int e^{iux} f(x) \mathrm{d}x$ is understood to be the Fourier transform of f. Moreover, we denote by $\|\cdot\|$ the L^2-norm of functions, $\|f\|^2 := \int |f(x)|^2 \mathrm{d}x$. Given some function $f \in \mathrm{L}^1(\mathbb{R}) \cap \mathrm{L}^2(\mathbb{R})$, we denote by f_m the uniquely defined function with Fourier transform $\mathcal{F}f_m = (\mathcal{F}f)\mathbf{1}_{[-m,m]}$.

General statistical setting. Consider n i.i.d. realizations Y_j, $1 \le j \le n$, of a random variable Y with Lebesgue-density f_Y. Suppose Y is related to a variable X, with Lebesgue-density f through a known transformation $\mathbf{T}$ relating their characteristic functions: $\varphi_Y = \mathbf{T}(\varphi_X)$, where $\mathbf{T}$ admits a continuous inverse. To estimate the density f of X from the (Y_j), we build an estimator $\widehat{\varphi}_{X,n}$ of φ_X as follows

$$\widehat{\varphi}_{X,n}(u) = \mathbf{T}^{-1}\big(\widehat{\varphi}_{Y,n}\big)(u) \quad \text{where } \widehat{\varphi}_{Y,n}(u) := \frac{1}{n}\sum_{j=1}^{n} e^{iuY_j}, \ u \in \mathbb{R}.$$

Cutting off in the spectral domain and applying Fourier inversion gives an estimator of f

$$\widehat{f}_m(x) = \frac{1}{2\pi} \int_{-m}^{m} e^{-iux} \widehat{\varphi}_{X,n}(u) \mathrm{d}u, \quad \forall m > 0, \ x \in \mathbb{R}.$$

Its performance is measured with a L^2-loss, the choice of the cutoff parameter m is crucial. The optimal cutoff $m^\star$ which minimizes the L^2-risk is such that

$$\mathbb{E}\big[\|\widehat{f}_{m^\star} - f\|^2\big] = \inf_{m \ge 0}\Big\{\frac{1}{2\pi} \int_{[-m,m]^c} |\varphi_X(u)|^2 \mathrm{d}u + \frac{1}{2\pi} \int_{-m}^{m} \mathbb{E}\big[|\widehat{\varphi}_{X,n}(u) - \varphi_X(u)|^2\big] \mathrm{d}u\Big\}.$$

(1.1)

This optimal value m^* usually depends on the unknown regularity of f and is hence not feasible. An adaptive optimal procedure consistst in selecting a random cutoff $\widehat{m}_n$, calculated from the observations, for which the L^2-risk is close to the one of $\widehat{f}_{m^*}$, meaning that one can establish an oracle bound

$$\mathbb{E}\big[\|\widehat{f}_{\widehat{m}_n} - f\|^2\big] \le C \inf_{m \ge 0}\Big\{\frac{1}{2\pi} \int_{[-m,m]^c} |\varphi_X(u)|^2 \mathrm{d}u + \frac{1}{2\pi} \int_{-m}^{m} \mathbb{E}\big[|\widehat{\varphi}_{X,n}(u) - \varphi_X(u)|^2\big] \mathrm{d}u\Big\} + r_n,$$

for a positive constant C and r_n a negligible remainder. Then, $\widehat{f}_{\widehat{m}_n}$ is called adaptive rate optimal estimator of f.

Heuristic of the adaptive procedure. Suppose there exists a function F_{φ_Y}, possibly depending on $\mathbf{T}$ and φ_Y, such that for some positive constant C, it holds

$$\mathbb{E}\big[|\widehat{\varphi}_{X,n}(u) - \varphi_X(u)|^2\big] \leq C|F_{\varphi_Y}(u)|^2\mathbb{E}\big[|\widehat{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2\big],$$

which is the case *e.g.* if $\mathbf{T}^{-1}$ is Fréchet differentiable: the quantity $F_{\varphi_Y} = (\mathbf{T}^{-1})'(\varphi_Y)$ is explicit in the deconvolution case. Then, it follows

$$\mathbb{E}\big[\|\widehat{f}_m - f\|^2\big] \leq \frac{1}{2\pi}\int\limits_{[-m,m]^c} |\varphi_X(u)|^2\mathrm{d}u + \frac{C}{n}\int\limits_{-m}^{m} \big|F_{\varphi_Y}(u)\big|^2\mathrm{d}u, \quad m \geq 0. \quad (1.2)$$

The second term in (1.2) is a majorant of the integrated variance of the estimator. If the upper bound (1.2) is optimal, meaning that it has the same order as (1.1), asymptotically we get $m^\star \asymp \overline{m}_n$, where $\overline{m}_n$ is such that the bias-variance compromise in the right hand side of (1.2) is realized. To compute $\overline{m}_n$, we differentiate in m the right hand side of (1.2) giving that $\overline{m}_n$ satisfies

$$|\varphi_X(\overline{m}_n)|^2 = \frac{C}{n}\big|F_{\varphi_Y}(\overline{m}_n)\big|^2 \iff |\mathbf{T}^{-1}(\varphi_Y)(\overline{m}_n)|^2 = \frac{C}{n}\big|F_{\varphi_Y}(\overline{m}_n)\big|^2. \quad (1.3)$$

Clearly, (1.3) has an empirical version, we select $\widehat{m}_n$ accordingly, let

$$\widehat{m}_n \in \left\{m \in [0,n], \left|\frac{\mathbf{T}^{-1}(\widehat{\varphi}_{Y,n})(m)}{F_{\widehat{\varphi}_{Y,n}}(m)}\right| = \frac{\kappa}{\sqrt{n}}\right\}, \quad (1.4)$$

for some $\kappa > 0$, possibly depending on n. As, the solution of (1.4) may not be unique we consider one element in this ensemble, such as its maximum.

Relation to other works   There exist numerous techniques for adaptivity, we mention some of them together with a non exhaustive list of references. Loosely speaking there exist three main approaches; thresholding techniques for wavelet density estimators (see *e.g.* [16, 17, 36]), penalized estimators (see *e.g.* [3, 1, 31, 29]) and pair wise comparison of estimators such as the Goldenshluger and Lepskii's procedure (see *e.g.* [23, 24, 27]). These techniques have been developed for different inverse problems and in anisotropic multidimensional settings.

All the afore mentioned methods rely on the choice of a parameter to be calibrated by the practitioner, such as κ in (1.4). Numerical performances of the selected estimator are sensitive to this choice and many studies have been devoted to the calibration of this parameter (see *e.g.* Baudry *et al.* [2] and Lacour *et al.* [27]). An advantage of the procedure presented here is that, in the cases considered, for all the values of κ such that the oracle bound bound is valid, the corresponding estimator is reasonable.

Many adaptive procedures such as penalization methods minimizes an empirical version of the upper bound (1.2), while the spirit of (1.4) consists in finding the zeroes of an empirical version of the derivative in m of the upper bound (1.2). Roughly speaking, the difference between our procedure and a penalization procedure is the same as the difference between Z-estimators and M-estimators.

Adaptation to the deconvolution problem We consider the deconvolution problem as it is an archetype inverse problem that has been extensively studied in the literature (see the references in Section 2). Moreover, it is a building brick of many inverse problems. One observes n i.i.d. realizations of $Y_i = X_i + \varepsilon_i$, where X_i and ε_i are independent, the density of ε_1 is known and the density f of X_1 is the quantity of interest. Optimal rates of convergence depend on the asymptotic decay of the characteristic function φ_ε of the noise ε, usually *ordinary smooth cases* – when φ_ε decays polynomially to 0 – and *super smooth cases* – when φ_ε decays exponentially to 0.

In that first case our procedure presents many advantages. On a theoretical point of view our procedure and the proof to establish oracle inequalities are the same in both ordinary smooth and super smooth cases, whereas usual adaptive procedures study these two cases separately. Moreover, the proof to get the oracle bound is rather elegant and relies on a fine cutting of the quadratic risk: the most powerful result involved is an Hoeffding concentration inequality. Usually, tools used to establish oracle inequalities rely on more demanding concentration results. On a numerical point of view, our procedure is simple and for all the possible choices of the hyper parameter κ in (1.4) predicted by the theory, the associated estimators are relevant. We conduct an extensive simulation study which illustrates the stability of the procedure. We compare our results with a penalization procedure described in Comte and Lacour [12], which are known to be rapid and efficient in deconvolution contexts.

1.2. A unified estimator for decompounding

Consider a compound Poisson process Z,

$$Z_t = \sum_{j=1}^{N_t} X_j, \quad t \geq 0,$$

where N is a Poisson process with intensity λ independent of the i.i.d. random variables $(X_j)_{j \in \mathbb{N}}$ with common density f. In the decompounding problem, one discretely observes one trajectory of a Z at sampling rate $\Delta > 0$ over $[0, T]$: $(Z_{i\Delta}, \ i = 1, \ldots, n)$, where $n = \lfloor T/\Delta \rfloor$. The aim is to estimate f from these observations. This model is central in many applied fields *e.g.* statistical physics, biology, financial series or mathematical insurance as it is well adapted to study phenomena where random independent events occur at random times. For instance, in insurance failure theory these events can model the claims that insurance companies have to pay to the subscribers, it is the Cramér-Lundberg model (see Embrechts *et al.* [21]).

In the literature, cases $\Delta \to 0$ (high frequency observations) or Δ fixed (low frequency observations, often $\Delta = 1$) have received a lot of attention (see the references given in Section 3) and are usually considered separately. Here we propose a unified strategy which is valid regardless the behavior of $\Delta := \Delta_n \to \Delta_0 \in [0, \infty)$. The dependency in Δ_0 of the upper bound is made

explicit and shows a deterioration as Δ_0 increases. Those results complement the knowledge on decompounding. Moreover, the estimator remains consistant in cases where Δ grows to infinity slowly. This latter result is not straight-forward, if the jump density is centered and has unit variance, it holds that $Z_\Delta = \sqrt{\lambda\Delta}\zeta_\Delta$, where $\zeta_\Delta \to \mathcal{N}(0,1)$ as $\Delta \to \infty$. Therefore, one would expect that in these regimes non-parametric estimation of f is impossible as each in-crement is close, in law, to a parametric Gaussian variable. When Δ goes too rapidly to infinity, namely as a power of $n\Delta_n$, Duval [19] shows that consistent non-parametric estimation of f is impossible, regardless of the choice of the loss function. Having a consistant estimator for f when Δ gets large is inter-esting, usually a Gaussian approximation is made to simplify computations, at the expense of loosing the specificities of the jump density (see e.g. Cont and de Larrard [15]).

Finally, we show that our adaptive procedure can be extended to this case and leads to an adaptive and rate optimal estimator of the jump density f, up to a logarithmic loss, for all sampling rates such that $\Delta_n < 1/4\log(n\Delta_n)$ as $n \to \infty$, this condition is fulfilled for fixed or vanishing Δ.

Organisation of the article In Section 2 we establish and prouve an oracle inequality based on our procedure for the deconvolution problem. We illustrate numerically its performances and compare it with a penalized adaptive optimal estimator given in [12]. Section 3 is dedicated to the decompounding problem. Finally, Section 4 gathers the proofs of the results of Section 3.

2. Deconvolution

2.1. Statistical setting

Suppose that $X_1, \ldots, X_n$ are i.i.d. with density f and are accessible through the noisy observations
$$Y_j = X_j + \varepsilon_j, \ j = 1, \ldots, n.$$
Assume that the (ε_j) are i.i.d., independent of the (X_j) and such that $\forall u \in \mathbb{R}$, $\varphi_\varepsilon(u) \neq 0$. Suppose that the distribution of ε_1 is known. This last assumption can be softened, the procedure allows a straightforward generalization to the case where the distribution of ε_1 can be estimated from an additional sample, see Neumann [32]. Then, the mapping $\mathbf{T}$ defined in Section 1.1 is given by $\mathbf{T} : \varphi \mapsto \varphi\varphi_\varepsilon$, which is a continuous mapping of inverse $\mathbf{T}^{-1} : \varphi \mapsto \varphi/\varphi_\varepsilon$ and $(\mathbf{T}^{-1})'$ is equal to $1/\varphi_\varepsilon$.

A deconvolution estimator of the characteristic function φ_X of X is given by

$$\frac{\widehat{\varphi}_{Y,n}(u)}{\varphi_\varepsilon(u)}, \quad \text{with} \ \ \widehat{\varphi}_{Y,n}(u) := \frac{1}{n}\sum_{j=1}^n e^{iuY_j}, \quad u \in \mathbb{R},$$

denoting the empirical characteristic function. Since φ_X is a characteristic func-tion, its absolute value is bounded by 1 and the estimator can hence be improved

by using the definition

$$\widehat{\varphi}_{X,n}(u) := \frac{\widehat{\varphi}_{Y,n}(u)}{\varphi_\varepsilon(u)} \frac{1}{\max\{1, \left|\frac{\widehat{\varphi}_{Y,n}(u)}{\varphi_\varepsilon(u)}\right|\}}, \quad u \in \mathbb{R}. \tag{2.1}$$

Cutting off in the spectral domain and applying Fourier inversion gives the estimator

$$\widehat{f}_m(x) = \frac{1}{2\pi} \int_{-m}^{m} e^{-iux} \widehat{\varphi}_{X,n}(u) \mathrm{d}u, \quad x \in \mathbb{R}. \tag{2.2}$$

This estimator and adaptation techniques have been extensively studied in the literature, including in more general settings than above. Optimal rates of convergence and adaptive procedures are well known in dimension $d = 1$ (see *e.g.* [8, 37, 22, 5, 6, 7, 34, 14] for L^2-loss functions or [30] for the L^∞-loss). Results have also been established for multivariate anisotropic densities (see *e.g.* [13] for L^2-loss functions or [35, 28] for L^p-loss functions, $p \in [1, \infty]$).

2.2. Risk bounds and adaptive bandwidth selection

The following risk bound is well known in the literature on deconvolution.

$$\mathbb{E}\big[\|\widehat{f}_m - f\|^2\big] \le \frac{1}{2\pi} \Big(\int_{[-m,m]^c} |\varphi_X(u)|^2 \mathrm{d}u + \frac{1}{n} \int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2} \Big). \tag{2.3}$$

This upper bound is the sum of a bias term that decreases with m and a variance term, increasing with m. We select $\overline{m}_n$, the optimal cutoff parameter, such that the upper bound (2.3) is minimal

$$\overline{m}_n \in \underset{m \ge 0}{\operatorname{arginf}}\Big\{ \int_{[-m,m]^c} |\varphi_X(u)|^2 \mathrm{d}u + \frac{1}{n} \int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2} \Big\}.$$

Differentiating the right hand side with respect to m, we find that the following holds for the optimal cutoff parameter:

$$|\varphi_X(\overline{m}_n)|^2 = \frac{1}{n|\varphi_\varepsilon(\overline{m}_n)|^2} \iff |\varphi_X(\overline{m}_n)\varphi_\varepsilon(\overline{m}_n)| = |\varphi_Y(\overline{m}_n)| = \frac{1}{\sqrt{n}}. \tag{2.4}$$

This equality has an empirical version and we select $\widehat{m}_n$ accordingly. In order to ensure adaptivity the following heuristic consideration is helpful. When the characteristic function is replaced by its empirical version, the standard deviation is of the order $n^{-1/2}$. Consequently, estimating φ_Y by $\widehat{\varphi}_{Y,n}$ makes sense for $|\varphi_Y| \ge n^{-1/2}$. If $|\varphi_Y| < n^{-1/2}$, the noise is dominant so the estimator can be set to zero. This inspires to re-define the estimator of φ_Y as follows:

$$\widetilde{\varphi}_{Y,n}(u) = \widehat{\varphi}_{Y,n}(u)\mathbf{1}_{\{|\widehat{\varphi}_{Y,n}(u)| \ge \kappa_n n^{-1/2}\}}, \quad u \in \mathbb{R}, \tag{2.5}$$

with the threshold value $\kappa_n := (1 + \kappa\sqrt{\log n})$. The constant $\kappa > 0$ is specified below and the additional log term is added to ensure good concentration properties on the event considered. Then, the estimator of f given in (2.2) is modified using (2.5) instead of (2.1). We obtain

$$\widetilde{\varphi}_{X,n}(u) := \frac{\widetilde{\varphi}_{Y,n}(u)}{\varphi_\varepsilon(u)} \frac{1}{\max\{1, |\frac{\widetilde{\varphi}_{Y,n}(u)}{\varphi_\varepsilon(u)}|\}}, \quad u \in \mathbb{R}$$

and

$$\widetilde{f}_m(x) = \frac{1}{2\pi} \int_{-m}^{m} e^{-iux} \widetilde{\varphi}_{X,n}(u) \mathrm{d}u, \quad x \in \mathbb{R}.$$

We define the empirical cutoff parameter $\widehat{m}_n$ as follows. Since $\widehat{\varphi}_{Y,n}$ may show an oscillatory behavior and the solution of (2.4) may not be unique, we consider

$$\widehat{m}_n = \max\left\{m > 0 : |\widehat{\varphi}_{Y,n}(m)| = \kappa_n n^{-1/2}\right\} \wedge n^\alpha, \tag{2.6}$$

for some $\alpha \in (0, 1]$. It is worth emphasizing that the calculation of $\widehat{m}_n$ does only rely on the empirical characteristic function $\widehat{\varphi}_{Y,n}$ which can be estimated from the direct observations, and does not require the evaluation of penalty terms depending on the (perhaps unknown) φ_ε. Moreover, this procedure is the same in the super smooth case as well as in the ordinary smooth case. In (2.6) if we set $\alpha = 1$, $\widehat{m}_n$ is thresholded to the value n, which is natural as if $m \geq n$, the variance term in (2.3) no longer vanishes. It is possible to set $0 < \alpha < 1$, if one has additional knowledge on the regularity of f (see the discussion below).

Theorem 2.1. *Let $\widehat{m}_n$ defined as in (2.6), with $\kappa > \sqrt{2}$ and $\alpha \in (0, 1]$. Then, there exist a positive constant C_1 depending only on the choice of κ and a universal positive constants C_2 such that*

$$\mathbb{E}[\|f - \widetilde{f}_{\widehat{m}_n}\|^2] \leq C_1 \inf_{m \in [0, n^\alpha]} \left\{ \int_{[-m,m]^c} |\varphi_X(u)|^2 \mathrm{d}u + \frac{\log n}{n} \int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2} \right\} + C_2 n^{\alpha - \frac{\kappa^2}{2}}.$$

Proof of Theorem 2.1. **Step 1: An upper bound for $\widetilde{f}_m$.** Let $m > 0$, we first establish an upper bound for the estimator $\widetilde{\widetilde{f}}_m$ of f defined as $\widetilde{f}_m$ but whose characteristic function is given by (2.5). Parseval equality and the definition (2.5) of $\widetilde{\varphi}_{Y,n}$ give

$$\mathbb{E}[\|\widetilde{\widetilde{f}}_m - f\|^2] \leq \|f_m - f\|^2 + \frac{1}{2\pi} \int_{-m}^{m} \frac{\mathbb{E}[|\widetilde{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2]}{|\varphi_\varepsilon(u)|^2} \mathrm{d}u,$$

where $\mathbb{E}[|\widetilde{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2]$ equals

$$\mathbb{E}\left[|\widehat{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2 \mathbf{1}_{|\widehat{\varphi}_{Y,n}(u)| \geq \frac{\kappa_n}{\sqrt{n}}}\right] + |\varphi_Y(u)|^2 \mathbb{P}\left(|\widehat{\varphi}_{Y,n}(u)| < \frac{\kappa_n}{\sqrt{n}}\right).$$

The first term in the right hand side is bounded by $\frac{1}{n}$. Recall that $\kappa_n = 1 + \kappa\sqrt{\log n}$, we decompose the second term on the set $\mathcal{A} = \{u, |\varphi_Y(u)| < \frac{1+2\kappa\sqrt{\log n}}{\sqrt{n}}\}$ this leads to

$$|\varphi_Y(u)|^2 \mathbb{P}\left(|\widehat{\varphi}_{Y,n}(u)| < \frac{\kappa_n}{\sqrt{n}}\right) \leq \frac{(1+2\kappa\sqrt{\log n})^2}{n} + \mathbb{P}\left(|\widehat{\varphi}_{Y,n}(u) - \varphi_Y(u)| \geq \frac{\kappa\sqrt{\log n}}{\sqrt{n}}\right)$$

$$\leq \frac{(1+2\kappa\sqrt{\log n})^2}{n} + n^{-\frac{\kappa^2}{2}} \leq \frac{1+(1+2\kappa\sqrt{\log n})^2}{n},$$

where we used the Hoeffding inequality and $\kappa > \sqrt{2}$. Finally, gathering all the above inequalities, together with the fact that $\mathbb{E}[\|\widetilde{f}_m - f\|^2] \leq \mathbb{E}[\|\widetilde{\widetilde{f}}_m - f\|^2]$, we get the following upper bound for $\widetilde{f}_m$, $0 < m \leq n$,

$$\mathbb{E}[\|\widetilde{f}_m - f\|^2] \leq \|f_m - f\|^2 + \frac{2 + (1+2\kappa\sqrt{\log n})^2}{2\pi n} \int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2}. \qquad (2.7)$$

Step 2: Adaptation. It holds using the Parseval equality, that

$$\mathbb{E}[\|\widetilde{f}_{\widehat{m}_n} - f\|^2] = \mathbb{E}\left[\int_{u \in [-\widehat{m}_n, \widehat{m}_n]^c} |\varphi_X(u)|^2 \mathrm{d}u\right] + \mathbb{E}\left[\int_{u \in [-\widehat{m}_n, \widehat{m}_n]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \mathrm{d}u\right]$$

$$:= T_1 + T_2.$$

Let $0 < m \leq n$ be fixed. First, consider the event $\mathcal{E} = \{\widehat{m}_n < m\}$, on this event we have the straightforward upper bound $T_2 \leq \int_{[-m,m]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \mathrm{d}u$, we recover a usual variance term. Now we control the surplus in the bias of the estimator $\widetilde{f}_{\widehat{m}_n}$ decomposing T_1 as

$$T_1 = \int_{[-m,m]^c} |\varphi_X(u)|^2 \mathrm{d}u + \mathbb{E}\left[\int_{|u| \in [\widehat{m}_n, m]} |\varphi_X(u)|^2 \mathrm{d}u\right].$$

It is the sum of a usual bias term and an additional term controled using the inequality

$$|\varphi_X|^2 \leq 2\frac{|\widehat{\varphi}_{Y,n}|^2}{|\varphi_\varepsilon|^2} + 2\frac{|\varphi_Y - \widehat{\varphi}_{Y,n}|^2}{|\varphi_\varepsilon|^2}.$$

Along with the definition of $\widehat{m}_n$, it gives

$$\mathbb{E}\left[\mathbf{1}_\mathcal{E} \int_{|u| \in [\widehat{m}_n, m]} |\varphi_X(u)|^2 \mathrm{d}u\right] \leq 2\mathbb{E}\left[\mathbf{1}_\mathcal{E} \int_{|u| \in [\widehat{m}_n, m]} \frac{\kappa_n^2 n^{-1}}{|\varphi_\varepsilon(u)|^2} \mathrm{d}u\right] + \int_{-m}^{m} \frac{\mathbb{E}[|\widehat{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2]}{|\varphi_\varepsilon(u)|^2} \mathrm{d}u$$

$$\leq \int_{-m}^{m} \frac{2(\kappa_n^2 + 1)n^{-1}}{|\varphi_\varepsilon(u)|^2} \mathrm{d}u.$$

Then, the definition of κ_n and (2.7) imply that, on the event $\mathcal{E}$, for a positive constant C depending only on the choice of κ,

$$\mathbb{E}[\|\widetilde{f}_{\widehat{m}_n} - f\|^2 \mathbf{1}_\mathcal{E}] \leq \|f_m - f\|^2 + C\frac{\log n}{n} \int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2}.$$

Second, on the complement set $\mathcal{E}^c$, we immediately have $T_1 \leq \int_{[-m,m]^c} |\varphi_X(u)|^2 \mathrm{d}u$. It remains to control the surplus in the variance of $\widetilde{f}_{\widehat{m}_n}$ using the decomposition

$$T_2 = \int_{u \in [-m,m]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \mathrm{d}u + \mathbb{E}\left[\int_{|u| \in [m,\widehat{m}_n]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \mathrm{d}u \right].$$

By the definition of $\widehat{m}_n$, it holds

$$\mathbb{E}\left[\int_{|u| \in [m,\widehat{m}_n]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \mathbf{1}_{\{|\varphi_Y(u)|>n^{-1/2}\}} \mathrm{d}u \mathbf{1}_{\mathcal{E}^c} \right]$$

$$\leq \int_{|u| \in [m,n^\alpha]} \frac{\mathbb{E}[|\widetilde{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2]}{|\varphi_\varepsilon(u)|^2} \mathbf{1}_{\{|\varphi_Y(u)|>n^{-1/2}\}} \mathrm{d}u.$$

On the event $\{|\varphi_Y(u)| > n^{-1/2}\}$, we derive that

$$\mathbb{E}[|\widetilde{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2] \leq |\varphi_Y(u)|^2 + \mathbb{E}[|\widehat{\varphi}_{Y,n}(u) - \varphi_Y(u)|^2]$$

$$\leq |\varphi_Y(u)|^2 + \frac{1}{n} \leq 2|\varphi_Y(u)|^2.$$

Consequently, we get

$$\mathbb{E}\left[\int_{|u| \in [m,\widehat{m}_n]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \, \mathbf{1}_{\{|\varphi_Y(u)|>n^{-1/2}\}} \mathrm{d}u \, \mathbf{1}_{\mathcal{E}^c} \right] \leq 2 \int_{[-m,m]^c} |\varphi_X(u)|^2 \mathrm{d}u.$$

Next, using that $|\widetilde{\varphi}_{X,n}(u)| \leq 1$, we derive that

$$\mathbb{E}\left[\int_{|u| \in [m,\widehat{m}_n]} |\widetilde{\varphi}_{X,n}(u) - \varphi_X(u)|^2 \mathbf{1}_{\{|\varphi_Y(u)|\leq n^{-1/2}\}} \mathrm{d}u \, \mathbf{1}_{\mathcal{E}^c} \right]$$

$$\leq \int_{|u| \in [m,n^\alpha]} |\varphi_X(u)|^2 \mathrm{d}u + 4 \int_{|u| \in [m,n^\alpha]} \mathbb{P}(|\widehat{\varphi}_{Y,n}(u)| \geq \kappa_n n^{-1/2}) \mathbf{1}_{\{|\varphi_Y(u)|\leq n^{-1/2}\}} \mathrm{d}u$$

$$\leq \int_{|u| \in [m,n^\alpha]} |\varphi_X(u)|^2 \mathrm{d}u + 4 \int_{|u| \in [m,n^\alpha]} \mathbb{P}(|\widehat{\varphi}_{Y,n}(u) - \varphi_Y(u)| > \kappa(\log n/n)^{1/2}) \mathrm{d}u$$

$$\leq \int_{u \in [m,m]^c} |\varphi_X(u)|^2 \mathrm{d}u + 8n^{\alpha - \kappa^2/2}.$$

The last inequality is a direct consequence of the Hoeffding inequality. Putting the above together, we have shown that for universal positive constants C_1 and C_3 and a constant C_2 depending only on κ, for all $m \geq 0$,

$$\mathbb{E}[\|\widetilde{f}_{\widehat{m}_n} - f\|^2] \leq C_1 \|f - f_m\|^2 + C_2 \frac{\log n}{n} \int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\varepsilon(u)|^2} + C_3 n^{\alpha - \kappa^2/2}.$$

Taking the infimum over m completes the proof. $\qquad\qquad\qquad\qquad\qquad\square$

Discussion Theorem 2.1 is non asymptotic and ensures that the estimator $\widetilde{f}_{\widehat{m}_n}$ automatically reaches the bias-variance compromise, up to a logarithmic factor and the multiplicative constant C_1.

Comments on the adaptive procedure. Similarly to the grouped data setting (see [20]), the computation of the adaptive cutoff (2.6) involves only the set $\left\{|\widehat{\varphi}_{Y,n}(u)| = (1 + \kappa\sqrt{\log n})/\sqrt{n}\right\}$. Therefore, $\widehat{m}_n$ depends only on the empirical characteristic function of the direct observations Y, and not the one of the errors φ_ε nor φ_X and the adaptive estimator and the proof of the oracle bound are the same in the super smooth case and the ordinary smooth case. Generalizing the result to the case where the distribution of the error is unknown but where we have *e.g.* independent i.i.d. relalizations $(\varepsilon_1, \ldots, \varepsilon_N)$ of ε should therefore be straightforward.

Comments on the proof. Proof of Theorem 2.1 is self contained and relies on fine cuttings of the quadratic risk. The more involved tool used is an Hoeffding inequality, whereas usual techniques involve stronger results such as Thalagrand inequalities. The interest is that it should be robust to small changes of in the modelisation.

Note that compared to [20], the proof relies on more direct arguments. Moreover, it permits to derive a stronger result, namely an oracle type inequality, whereas in [20] we ensured that *given some regularity class* the optimal rate is achieved on this class.

Choice of the hyper parameters in (2.6). Regarding the choice of α and κ in (2.6), it is always possible to take $\alpha = 1$. Note that the case $\alpha > 1$ is not interesting as, even in the direct problem $\varepsilon = 0$ a.s., if $m > n$ the variance term in (2.3) no longer tends to 0. Taking $\alpha < 1$ is possible only if one has additional information on the target density f. For instance, if one knowns that f is in a Sobolev class of regularity β, for some $\beta \geq \beta_0 > 0$,

$$f \in \mathscr{S}(\beta, L) := \left\{ f \in \mathbf{F}, \ \int_{\mathbb{R}} (1 + |u|)^{2\beta} |\mathcal{F}f(u)|^2 \mathrm{d}u \leq L \right\} \tag{2.8}$$

where $\mathbf{F}$ is the set of densities with respect to the Lebesgue measure. Then, it holds that $\|f - f_m\|^2 \asymp m^{-2\beta}$ and straightforward computations lead to $m^\star \lesssim (n/\log n)^{\frac{1}{2\beta+1}}$ (regardless the the asymptotic decay of φ_ε). Then, one may restrict the interval for $\widehat{m}_n$ to $[0, n^\alpha]$ where $1 > \alpha > \frac{1}{2\beta_0+1}$. Second, the choice of κ must be such that $n^{\alpha - \kappa^2/2}$ is negligible, the choice $\kappa > 2$ always works. The following numerical study illustrates that the procedure is stable in the choice of κ.

2.3. Numerical results

Stability of the procedure To illustrate the performances of the method and the influence of the parameter κ we proceed as follows. Fix $\alpha = 1$, therefore
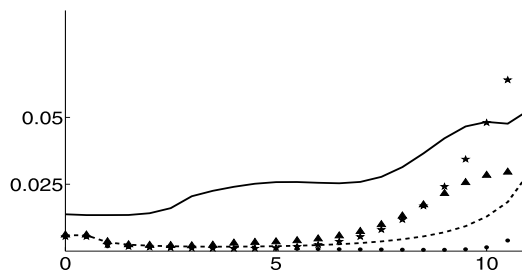
FIG 1. **Direct problem:** *Computations by $M = 1000$ Monte Carlo iterations of the L^2-risks (y axis) for different values of $\kappa \leq (\sqrt{n} - 1)\sqrt{\log(n)}^{-1} = 11.6$ (x axis). Estimation of f from $n = 1000$ i.i.d. direct realizations for different distributions: Uniform $\mathcal{U}[1,3]$ (plain line), Gaussian $\mathcal{N}(2,1)$ (dots), Cauchy (stars), Gamma $\Gamma(2,1)$ (dotted line) and the mixture $0.7\mathcal{N}(4,1) + 0.3\Gamma(2, \frac{1}{2})$ (triangles).*

we do not assume any regularity on the considered examples. For different densities f, namely, Uniform $\mathcal{U}[1,3]$, Gaussian $\mathcal{N}(2,1)$, Cauchy, Gamma $\Gamma(2,1)$ and the mixture $0.7\mathcal{N}(4,1) + 0.3\Gamma(2, \frac{1}{2})$, and for different values of κ we compute the adaptive L^2 risks from $M = 1000$ Monte Carlo iterations. The results are displayed on Figures 1, 2 and 3. We consider three different settings:

- The direct density estimation problem (Figure 1): we observe i.i.d. realizations of f. It is a particular deconvolution problem where $\varepsilon = 0$ a.s.
- Deconvolution problem with ordinary smooth noise (Figure 2): the error ε is Gamma $\Gamma(2,1)$ *i.e.* $|\varphi_\varepsilon|$ decays as $|u|^{-2}$ asymptotically.
- Deconvolution problem with super smooth noise (Figure 3): the error ε is Cauchy *i.e.* $|\varphi_\varepsilon|$ decays as $e^{-|u|}$ asymptotically.

On Figures 1, 2 and 3 we observe that the adaptive rates are small and that the procedure is stable in the choice of kappa. We observe, on these three cases, that the value of κ should not be chosen too large but that for a wide range of values the performances are similar. In practice, the value of n is fixed and there is a natural boundary for κ, indeed observe that it is useless to increase κ if $(1 + \kappa \log n)n^{-1/2} \geq 1$ as the selection rule (2.6) will be constant equal to n^α. Moreover, we expect that if $(1 + \kappa \log n)n^{-1/2}$ gets too large, *e.g.* larger than $1/2$ the performances of the adaptive estimator should deteriorate. This practical consideration encourages to choose κ smaller than $(\sqrt{n} - 1)\sqrt{\log(n)}^{-1}$. In Figures 1, 2 and 3 it appears that for all the meaningful values of κ, *e.g.* smaller than $\frac{1}{2}(\sqrt{n} - 1)\sqrt{\log(n)}^{-1}$ for instance, the performances of the adaptive estimator are similar.

Comparison with a penalization procedure   We compare the performances of our procedure for $\kappa = 8$, with a penalization procedure and with an oracle. For the penalization procedure, we follow Comte and Lacour [12] and consider
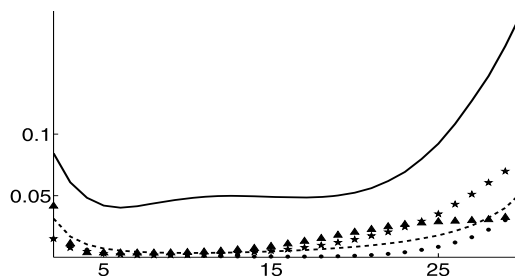
FIG 2. **Deconvolution problem (ordinary smooth case):** *Computations by $M = 1000$ Monte Carlo iterations of the* L^2*-risks (y axis) for different values of* $\kappa \leq (\sqrt{n} - 1)\sqrt{\log(n)}^{-1} = 32.6$ *(x axis). Estimation of f from $n = 10000$ i.i.d. direct of $X + \varepsilon$ where ε has distribution $\Gamma(2,1)$ (i.e. $\varphi_\varepsilon(u) = (1 - iu)^{-2}$) and for different distributions for X: Uniform $\mathcal{U}[1,3]$ (plain line), Gaussian $\mathcal{N}(2,1)$ (dots), Cauchy (stars), Gamma $\Gamma(2,1)$ (dotted line) and the mixture $0.7\mathcal{N}(4,1) + 0.3\Gamma(2, \frac{1}{2})$ (triangles).*

the adaptive estimator $\widehat{f}_{\widetilde{m}_n}$ which is the estimator defined in (2.2) where

$$\widetilde{m}_n = \operatorname*{argmin}_{m \in [0, M_n]} \{-\|\widehat{f}_m\|^2 + \mathrm{pen}(m)\}, \ \ \mathrm{pen}(m) = K\Big(\frac{\Delta(m)}{\log(m+1)}\Big)^2 \frac{\Delta(m)}{n},$$

where $M_n > 0$, $K > 0$ and $\Delta(m) = \frac{1}{2\pi} \int_{[-m,m]} |\varphi_\varepsilon(u)|^{-2} \mathrm{d}u$, which is known in our setting. The parameter M_n is chosen as the maximal integer such that $1 \leq \frac{\Delta(m)}{n} \leq 2$. For the parameter K it is calibrated by preliminary simulation experiments. For calibration strategies (dimension jump and slope heuristics), the reader is referred to Baudry *et al.* [2]. Here, we test a grid of values of the K's from the empirical error point of view, to make a relevant choice; the tests are conducted on a set of densities which are different from the one considered hereafter, to avoid overfitting. After these preliminary experiments, K is chosen equal to 2 which is the same value as the one considered in Comte and Lacour [12]. The standard errors are given in parenthesis. The running times for each risks of the penalization procedure and our procedure are similar. However, one should take into account that a preliminary calibration step seems obsolete in our case. In deconvolution problems, the theoretical optimal K can be in some cases far away from the practically optimal K and may vary with the sample size explaining the nessecity of this calibration step (see *e.g.* Kappus and Mabon [26] where the practical optimal value of K was much smaller than the value predicted by the theory).

Second, an oracle "estimator" is computed $\widehat{f}_{m^\star}$, which is the estimator defined in (2.2) where $m^\star$ corresponds to the following oracle bandwidth

$$m^\star = \operatorname*{argmin}_{m > 0} \mathbb{E}[\|f - \widehat{f}_m\|^2].$$

This oracle can be explicitly evaluated when f is known. We denote these different risks by R, for the risk of our procedure, R_{pen} for the penalized estimator
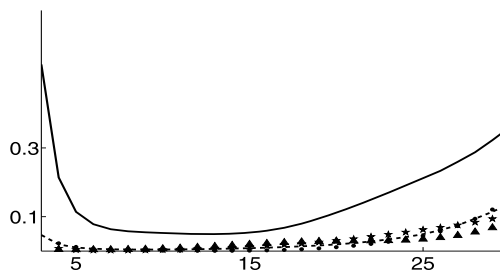
FIG 3. **Deconvolution problem (super smooth case):** *Computations by $M = 1000$ Monte Carlo iterations of the* L^2*-risks (y axis) for different values of* $\kappa \leq (\sqrt{n} - 1)\sqrt{\log(n)}^{-1} = 32.6$ *(x axis). Estimation of f from $n = 10000$ i.i.d. realizations of $X + \varepsilon$ where ε has Cauchy distribution (i.e. $\varphi_\varepsilon(u) = e^{-|u|}$) and for different distributions for X: Uniform $\mathcal{U}[1,3]$ (plain line), Gaussian $\mathcal{N}(2,1)$ (dots), Cauchy (stars), Gamma $\Gamma(2,1)$ (dotted line) and the mixture $0.7\mathcal{N}(4,1) + 0.3\Gamma(2, \frac{1}{2})$ (triangles). For the uniform distribution, the rates where stable around the value 0.5, they do not appear on the Figure not to spoil the readability of the other curves.*

TABLE 1

Comparaison of the different adaptive estimators for the Gamma distribution.

f_ε \ f	n	R	$\widehat{m}$	$\Gamma(2,1)$ R_{pen}	$\widetilde{m}$	R_{or}	$m^\star$
$\Gamma(2,1)$	500	4.31×10^{-2} (0.20)	1.05 (0.07)	1.97×10^{-2} (0.01)	0.80 (0.05)	0.74×10^{-2} (0.36×10^{-2})	0.66 (0.14)
	1000	1.74×10^{-2} (0.13)	0.98 (0.04)	1.70×10^{-2} (0.03)	0.94 (0.03)	0.59×10^{-2} (0.28×10^{-2})	0.72 (0.14)
	5000	0.40×10^{-2} (0.06)	0.85 (0.01)	1.30×10^{-2} (0.01)	1.32 (0.05)	0.31×10^{-2} (0.13×10^{-2})	0.91 (0.15)
$\mathcal{C}$	500	5.27×10^{-2} (0.23)	0.90 (0.07)	1.21×10^{-2} (0.69×10^{-2})	0.56 (0.04)	0.92×10^{-2} (0.48×10^{-2})	0.61 (0.12)
	1000	1.84×10^{-2} (0.13)	0.84 (0.04)	0.98×10^{-2} (0.61×10^{-2})	0.70 (0.03)	0.70×10^{-2} (0.34×10^{-2})	0.67 (0.13)
	5000	0.51×10^{-2} (0.07)	0.70 (0.01)	0.71×10^{-2} (0.01×10^{-2})	1.10 (0.02)	0.39×10^{-2} (0.17×10^{-2})	0.82 (0.13)

and R_{or} for the oracle procedure. All these risks are computed on 1000 Monte Carlo iterations. The results are gathered in Tables 1 for the Gamma density, 2 for the mixture and 3 for the Cauchy density where $\mathcal{C}$ stands for the Cauchy distribution. In each case both an ordinary smooth and a super smooth errors are considered.

Comparison of the different methods. Tables 1, 2 and 3 show that all the procedures behave as expected; the L^2-risks decreases with n and are smaller in the case of an ordinary smooth deconvolution problem than in the case of a super smooth deconvolution problem. The estimator with the smallest risk is the oracle, and the penalized risks are most of the time smaller than our procedure which is consistent with the fact that our procedure has a logarithmic loss and is asymptotic. More precisely for small values of n our procedure does not perform as well as the penalized method. But for larger values of n it is competitive. We can exhibit particular cases where our procedure is more stable in the choice of the hyper parameter than the penalized procedure, even on large sample sizes

TABLE 2
Comparaison of the different adaptive estimators on a mixture.

f f_ε	n	$0.7\mathcal{N}(4,1) + 0.3\Gamma(4,\frac{1}{2})$					
		R	$\widehat{m}$	R_{pen}	$\widetilde{m}$	R_{or}	$m^\star$
$\Gamma(2,1)$	500	1.77×10^{-2}	0.92	0.78×10^{-2}	0.78	0.33×10^{-2}	0.59
		(0.13)	(0.05)	(0.65×10^{-2})	(0.03)	(0.17×10^{-2})	(0.16)
	1000	0.74×10^{-2}	0.89	0.73×10^{-2}	0.87	0.26×10^{-2}	0.67
		(0.08)	(0.03)	(0.64×10^{-2})	(0.03)	(0.15×10^{-2})	(0.14)
	5000	0.13×10^{-2}	0.79	0.38×10^{-2}	1.10	0.01×10^{-2}	0.82
		(0.03)	(0.01)	(0.30×10^{-2})	(0.01)	(0.06×10^{-3})	(0.11)
$\mathcal{C}$	500	2.78×10^{-2}	0.82	0.73×10^{-2}	0.55	0.43×10^{-2}	0.50
		(0.15)	(0.05)	(0.50×10^{-2})	(0.05)	(0.20×10^{-2})	(0.14)
	1000	1.02×10^{-2}	0.77	0.65×10^{-2}	0.67	0.34×10^{-2}	0.58
		(0.10)	(0.03)	(0.52×10^{-2})	(0.03)	(0.16×10^{-2})	(0.15)
	5000	0.21×10^{-2}	0.66	0.59×10^{-2}	0.94	0.16×10^{-2}	0.74
		(0.04)	(0.01)	(0.48×10^{-2})	(0.02)	(0.10×10^{-2})	(0.11)

TABLE 3
Comparaison of the different adaptive estimators for the Cauchy distribution.

f f_ε	n	$\mathcal{C}$					
		R	$\widehat{m}$	R_{pen}	$\widetilde{m}$	R_{or}	$m^\star$
$\Gamma(2,1)$	500	2.69×10^{-2}	0.59	1.00×10^{-2}	0.68	0.67×10^{-2}	0.62
		(0.16)	(0.07)	(0.67×10^{-2})	(0.03)	(0.29×10^{-2})	(0.10)
	1000	1.00×10^{-2}	0.84	0.97×10^{-2}	0.82	0.49×10^{-2}	0.67
		(0.09)	(0.04)	(0.70×10^{-2})	(0.05)	(0.21×10^{-2})	(0.10)
	5000	0.27×10^{-2}	0.69	0.81×10^{-2}	1.18	0.21×10^{-2}	0.82
		(0.05)	(0.01)	(0.57×10^{-2})	(0.04)	(0.10×10^{-2})	(0.10)
$\mathcal{C}$	500	2.50×10^{-2}	0.74	1.12×10^{-2}	0.45	0.86×10^{-2}	0.56
		(0.16)	(0.07)	(0.27×10^{-2})	(0.01)	(0.34×10^{-2})	(0.09)
	1000	1.00×10^{-2}	0.68	0.74×10^{-2}	0.59	0.62×10^{-2}	0.62
		(0.10)	(0.04)	(0.32×10^{-2})	(0.03)	(0.24×10^{-2})	(0.09)
	5000	0.52×10^{-2}	0.54	0.73×10^{-2}	0.93	0.29×10^{-2}	0.74
		(0.07)	(0.01)	(0.52×10^{-2})	(0.03)	(0.11×10^{-2})	(0.09)

(see Figure 4 for example). This is due to the fact that the penalized constant K that is suitable for small values of n is different than for larger values of n. In practice a logarithmic term in n is added in the penalty term, that is theoretically unnecessary and entails a logarithmic loss but improves the numerical results. If we add this logarithmic term (we replace $K = 2$ with $\widetilde{K} \log(n)^{2.5}$ with $\widetilde{K} = 0.3$ and the multiplying $\log(n)^{2.5}$ factor as suggested in Comte *et al.* [14]). This second penalty procedure performs well for all values of n and when n gets large it has similar performances as our procedure (see Table 4). For our procedure, changing κ for smaller values of n does not improve the results.

3. Decompounding

3.1. *Statistical setting*

Let Z be a compound Poisson process with intensity $\lambda > 0$ and jump density f, *i.e.*
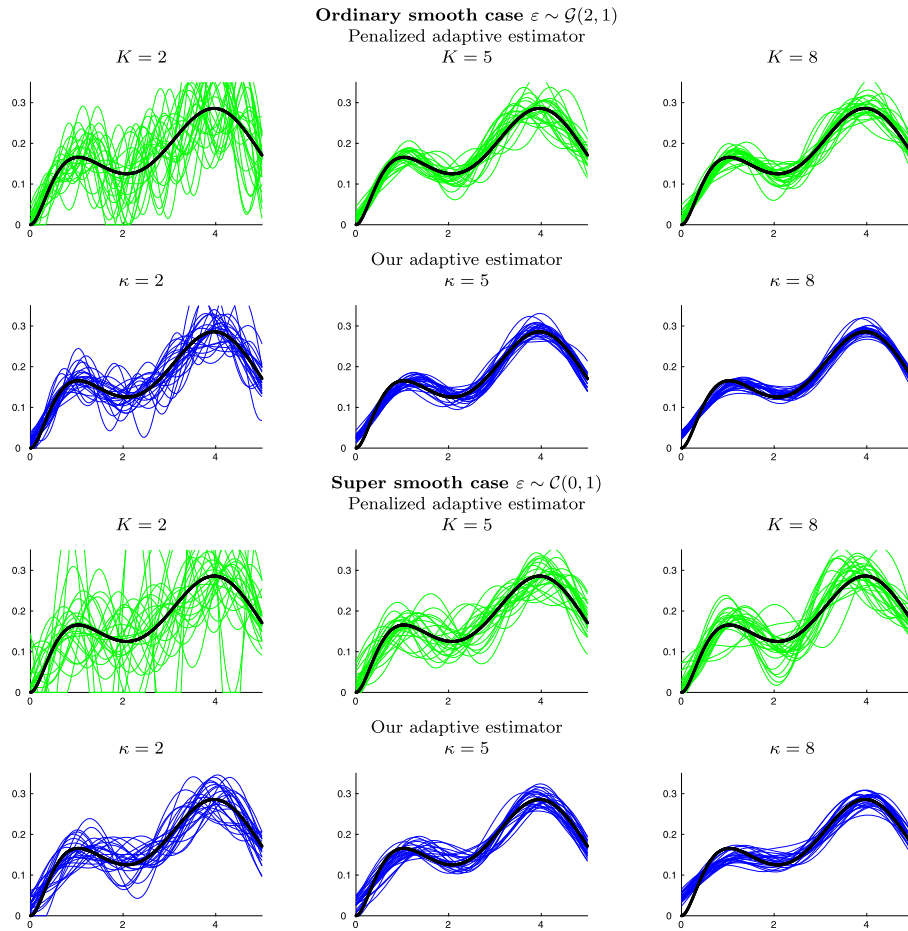
$$Z_t := \sum_{j=1}^{N_t} X_j, \quad t \geq 0$$

Ordinary smooth case $\varepsilon \sim \mathcal{G}(2,1)$
Penalized adaptive estimator

Our adaptive estimator

Super smooth case $\varepsilon \sim \mathcal{C}(0,1)$
Penalized adaptive estimator

Our adaptive estimator

FIG 4. *Comparison for different values of K and κ the penalized estimator (green) and our adaptive estimator (blue). Estimation of $f \sim \left(0.3\mathcal{G}(3,\frac{1}{2}) + 0.7\mathcal{G}(4,1)\right)$ (bold black) from $n = 10000$ observations.*

where N is an homogeneous Poisson process with intensity λ and independent of the i.i.d. variables (X_j) with common density f. One trajectory of Z is observed at sampling rate Δ over $[0, T]$, $T = n\Delta$, $n \in \mathbb{N}$. Non-parametric estimation of f, or its Lévy density λf has been the subject of many papers, among others, [4, 9, 11, 18, 38] and [25] for the multidimensional setting.

We observe Z at the time points $j\Delta$, $j = 1, \dots, n$, for $\Delta > 0$, denote the j-th increment by $Y_{j\Delta} = Z_{j\Delta} - Z_{(j-1)\Delta}$. We aim at estimating f from the increments $(Y_{j\Delta}, j = 1, \dots, n)$. Consider φ the characteristic function of X_1 and φ_Δ the characteristic function of $Z_\Delta = Y_\Delta$. The Lévy-Kintchine formula relates them as follows

$$\varphi_\Delta(u) = \exp\left(\Delta\lambda(\varphi(u) - 1)\right), \quad u \in \mathbb{R}.$$

TABLE 4
Risks and selected cutoff of the penalized procedure with an additional logarithmic term in the penalty. Notation $\mathcal{M}$ stands for the mixture $0.7\mathcal{N}(4,1) + 0.3\Gamma(4, \frac{1}{2})$.

f		$\mathcal{G}$		$\mathcal{M}$		$\mathcal{C}$	
f_ε	n	$R_{\widetilde{pen}}$	$\widetilde{m}_{\widetilde{pen}}$	$R_{\widetilde{pen}}$	$\widetilde{m}_{\widetilde{pen}}$	$R_{\widetilde{pen}}$	$\widetilde{m}_{\widetilde{pen}}$
$\Gamma(2,1)$	500	1.17×10^{-2} (0.81×10^{-2})	0.72 (0.03)	0.63×10^{-2} (0.51×10^{-2})	0.69 (0.02)	0.83×10^{-2} (0.46×10^{-2})	0.62 (0.005)
	1000	0.94×10^{-2} (0.63×10^{-2})	0.82 (0.04)	0.40×10^{-2} (0.30×10^{-2})	0.75 (0.02)	0.59×10^{-2} (0.32×10^{-2})	0.69 (0.03)
	5000	0.62×10^{-2} (0.62×10^{-2})	1.08 (0.01)	0.20×10^{-2} (0.16×10^{-2})	0.94 (0.02)	0.31×10^{-2} (0.19×10^{-2})	0.91 (0.02)
$\mathcal{C}$	500	1.34×10^{-2} (0.43×10^{-2})	0.45 (0.01)	0.60×10^{-2} (0.25×10^{-2})	0.45 (0.01)	1.31×10^{-2} (0.24×10^{-2})	0.41 (0.02)
	1000	0.91×10^{-2} (0.43×10^{-2})	0.59 (0.02)	0.74×10^{-2} (0.51×10^{-2})	0.51 (0.003)	0.94×10^{-2} (0.14×10^{-2})	0.45 (0.002)
	5000	0.56×10^{-2} (0.33×10^{-2})	0.81 (0.05)	0.21×10^{-2} (0.15×10^{-2})	0.75 (0.001)	0.35×10^{-2} (0.11×10^{-2})	0.65 (0.02)

Then, the mapping $\mathbf{T}$ defined in Section 1.1 is given by $\mathbf{T} : \varphi \mapsto \varphi_\Delta$. As Z is a compound Poisson process, $|\varphi_\Delta|$ is bounded from below by $e^{-2\lambda\Delta}$, which remains bounded away from 0 as long as $\Delta < \infty$. Moreover, if $\mathbb{E}[|X_1|] < \infty$ it holds that φ is differentiable and we can then define the distinguished logarithm of φ_Δ (see Lemma 1 in [20])

$$\varphi(u) = 1 + \frac{\text{Log}\left(\varphi_\Delta(u)\right)}{\lambda\Delta}, \quad \text{where } \text{Log}(\varphi_\Delta(u)) = \int_0^u \frac{\varphi'_\Delta(z)}{\varphi_\Delta(z)}\text{d}z,\, u \in \mathbb{R}. \quad (3.1)$$

For simplicity, we assume that the intensity λ is known: $\lambda = 1$. Following (3.1), an estimator of φ is hence given by

$$\widehat{\varphi}_n(u) = 1 + \frac{1}{\Delta}\text{Log}(\widehat{\varphi}_{\Delta,n}(u)), \quad u \in \mathbb{R} \quad (3.2)$$

with

$$\text{Log}(\widehat{\varphi}_{\Delta,n}(u)) := \int_0^u \frac{\widehat{\varphi}'_{\Delta,n}(z)}{\widehat{\varphi}_{\Delta,n}(z)}\text{d}z, \quad \text{where } \widehat{\varphi}^{(k)}_{\Delta,n}(z) = \frac{1}{n}\sum_{j=1}^n (iY_{j\Delta})^k e^{izY_{j\Delta}}, \quad (3.3)$$

$k \in \{0, 1\}$. The quantity $\text{Log}(\widehat{\varphi}_{\Delta,n})$: if φ_Δ never cancels, it may not be the case of its estimator $\widehat{\varphi}_{\Delta,n}$. Usually, to prevent this issue a local threshold is used and $(\widehat{\varphi}_{\Delta,n}(z))^{-1}$ is replaced with $(\widehat{\varphi}_{\Delta,n}(z))^{-1}\mathbf{1}_{|\widehat{\varphi}_{\Delta,n}(z)|>r_n}$, for some vanishing sequence r_n (see *e.g.* Neumann and Reiß [33]). Here we do not use a local threshold inside the integral in (3.3), replacing $(\widehat{\varphi}_{\Delta,n}(z))^{-1}$ by $(\widehat{\varphi}_{\Delta,n}(z))^{-1}\mathbf{1}_{|\widehat{\varphi}_{\Delta,n}(z)|>r_n}$, instead, we will threshold the integral so that our estimator $\widetilde{\text{Log}(\varphi_{\Delta,n})}$ of $\text{Log}(\varphi_{\Delta,n})$ satisfies $\widetilde{\text{Log}(\varphi_{\Delta,n})} = \text{Log}(\widehat{\varphi}_{\Delta,n})$. For that, consider

$$\widetilde{\varphi}_n(u) := \widehat{\varphi}_n(u)\mathbf{1}_{|\widehat{\varphi}_n(u)|\leq 4}, \quad u \in \mathbb{R}, \quad (3.4)$$

where $\widehat{\varphi}_n$ is given by (3.2). The choice of a threshold equal to 4 is technical (see the proof of Theorem 3.1). Cutting off in the spectral domain an applying a

Fourier inversion provides the estimator if f

$$\widehat{f}_{m,\Delta}(x) = \frac{1}{2\pi} \int_{-m}^{m} e^{-iux} \widetilde{\varphi}_n(u)\mathrm{d}u, \quad x \in \mathbb{R}. \qquad (3.5)$$

3.2. Adaptive upper bound

3.2.1. Upper bound and discussion on the rate

Theorem 3.1. *Assume that* $\mathbb{E}[X_1^2] < \infty$, $\Delta < \frac{1}{4}\log(n\Delta)$ *and* $n\Delta \to \infty$ *as* $n \to \infty$. *Then, for any* $m \geq 0$ *it holds*

$$\mathbb{E}\big[\|\widehat{f}_{m,\Delta} - f\|^2\big] \leq \|f_m - f\|^2 + \frac{2}{n\Delta} \int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2} + 2 \; 5^2 \mathbb{E}[X_1^2]\frac{m}{n\Delta} + 2^3 5^2 \frac{m^2}{(n\Delta)^2}.$$

The contraint $\Delta < \frac{1}{4}\log(n\Delta)$ is fulfilled for any bounded Δ as $n\Delta \to \infty$. Moreover it allows Δ to be such that $\Delta := \Delta_n \to 0$ and $\Delta_n \to \infty$, not too fast. This last point is interesting. An estimator that is optimal simultaneously when Δ is fixed or vanishing and consistant, optimal up to a logarithmic loss, when Δ tends to infinity, has scarcely been investigated, nor the estimation problem when the sampling rate goes to infinity. To the knowledge of the authors, the only similar result was released shortly after our result in Coca [10]. In [10], a $L_p, p \geq 1$ adaptive optimal nonparametric estimation of the Lévy density (which is related to the jump density) is studied. Both results are complementary, our estimator is adapted to L_2 and has the advantage that its definition (both adaptive and non adaptive) is simpler, leading to more succinct proofs and we provide a numerical study. In the remaining of this paragraph, we discuss the different rates of convergence implied by Theorem 3.1 according to the behavior of Δ.

Discussion on the rates The upper bound derived in Theorem 3.1 is the sum of four terms: a bias, two variance terms $V \asymp \frac{e^{4\Delta}m}{n\Delta}$ (using that $|\varphi_\Delta(u)| \geq e^{-2\Delta}$) and $V' \asymp \frac{m}{n\Delta}$, which is always smaller or of the same order as V, and a remainder. Assume that f lies is the Sobolev ball $\mathscr{S}(\beta, L)$ (see (2.8)). Then, the bias $\|f - f_m\|^2$ has asymptotic order $m^{-2\beta}$ and we may derive the following rates of convergence.

- **Microscopic and mesoscopic regimes.** Let $\Delta = \Delta_n$ be such that $\Delta_n \to \Delta_0 \in [0, \infty)$ such that $n\Delta_n \to \infty$. Then, the bias variance compromise leads to the choice $m^\star = (e^{-4\Delta_0}n\Delta_0)^{\frac{1}{2\beta+1}}$ and to the rate of convergence $\big(e^{-4\Delta_0}n\Delta_0\big)^{-\frac{2\beta}{2\beta+1}}$ that matches the optimal rates of convergence as Δ_0 is fixed or tending to 0. Indeed, the rate is in $T^{-\frac{2\beta}{2\beta+1}}$, with $T = n\Delta_n$ denoting the time horizon, it is clearly rate optimal as it corresponds to the optimal rate of convergence to estimate the jump density of a compound Poisson process from continuous observations ($\Delta = 0$). The

constant $e^{-4\Delta_0}$ appearing in the rate depends exponentially on Δ_0, which asymptotically as little effect but in practice deteriorates the numerical performances.

- **Macroscopic regime.** Let $\Delta = \Delta_n \to \infty$ such that $\Delta_n < \frac{1}{4}\log(n\Delta_n)$ The variance term V tends to 0, so that the estimator is consistent. Heuristically, if Δ goes to infinity the central limit theorem states that Y_Δ is close in law to a parametric Gaussian variable, *e.g.* if f is centered and with unit variance it holds that: $\sqrt{\Delta}^{-1}Y_\Delta \xrightarrow[\Delta\to\infty]{d} \mathcal{N}(0,1)$. Consequently, the fact that f can be constantly estimated is non trivial. Duval [19] establishes that if $\Delta = O((n\Delta)^\delta)$, for some $\delta \in (0,1)$, *i.e.* when Δ_n goes rapidly to infinity, there exists no consistent non-parametric estimator of f. The fact that estimation is impossible when Δ goes too rapidly to infinity was established through an asymptotic equivalence result. In this case it is always possible to build two different compound Poisson processes for which the statistical experiments generated by their increments are asymptotically equivalent. Therefore, the result of Theorem 3.1 is new in that context. We may distinguish two additional regimes:

  1. Slow macroscopic regime. If $\Delta_n = o\big(\log(n\Delta_n)\big)$, the choice $m^\star = \big(e^{-4\Delta_n}n\Delta_n\big)^{\frac{1}{2\beta+1}}$ leads to the rate of convergence $\big(e^{-4\Delta_n}n\Delta_n\big)^{-\frac{2\beta}{2\beta+1}}$. There is no lower bound in the literature to ensure if this rate is optimal. However if Δ goes slowly to infinity, for example if $\Delta_n = \log(\log(n\Delta_n))$, then the rate is $\big((\log(n\Delta_n))^{-4}n\Delta_n\big)^{-\frac{2\beta}{2\beta+1}}$, which is rate optimal, up to the logarithmic loss that may not be optimal.

  2. Intermediate macroscopic regime. Let $\Delta_n = \delta\log(n\Delta_n), 0 < \delta < 1/4$, then $m^\star = (n\Delta_n)^{\frac{1-4\delta}{2\beta+1}}$, leading to the rate $(n\Delta_n)^{-\frac{2\beta(1-4\delta)}{2\beta+1}}$. This rate deteriorates as δ increases. The limit $\delta = 1/4$ imposed by Theorem 3.1 may not be optimal, no lower bound adapted to this case exists in the literature.

The interest of the macroscopic regime is mainly theoretical as in practice if Δ is a large constant to get $e^{-4\Delta}n\Delta$ large one should consider a huge amount n of observations. However, this regime enlightens the role of the sampling rate Δ in the non-parametric estimation of the jump density. Using [19], consistent non-parametric estimation of the jump density is impossible if $\exists \delta > 0, \ \Delta_n = O(n\Delta_n)^\delta$, the remaining questions are what happens in between and if the log loss in the upper bound that appears when $\Delta_n \to \infty$ is avoidable or not. The constant $1/4$ in the bound $\Delta_n < 1/4\log(n\Delta_n)$ of Theorem 3.1 can probably be improved.

3.2.2. *Adaptive choice of the cutoff parameter*

We consider the optimal cutoff $\overline{m}_n$ given by

$$\overline{m}_n \in \operatorname*{arginf}_{m\geq 0}\Big\{\|f_m - f\|^2 + \frac{2}{n\Delta}\int_{-m}^{m}\frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2} + 2\ 5^2\mathbb{E}[X_1^2]\frac{m}{n\Delta} + 2^3 5^2\frac{m^2}{(n\Delta)^2}\Big\}.$$

Following the previous strategy, the upper bound given by Theorem 3.1 is optimal, at least for $\Delta \to [0, \infty)$. The leading variance terms is in $\frac{m e^{4\Delta}}{n\Delta}$, we differentiate in m the upper bound to find that the optimal cutoff $m^\star \asymp \overline{m}_n$ is such that:

$$|\varphi(\overline{m}_n)|^2 = \frac{e^{4\Delta}}{n\Delta},$$

which has an empirical version, we select $\widehat{m}_n$ accordingly. As in the deconvolution setting, we modify the estimator $\widetilde{\varphi}_n$ in (3.4) which is set to 0 when the estimator of $|\varphi|$ is smaller that $1/\sqrt{n\Delta}$, meaning that the noise is dominant. Define

$$\overline{\varphi}_n(u) := \widetilde{\varphi}_n(u) \mathbf{1}_{|\widetilde{\varphi}_n(u)| \geq \kappa_{n,\Delta}/\sqrt{n\Delta}}, \quad u \in \mathbb{R} \tag{3.6}$$

where $\kappa_{n,\Delta} := (e^{2\Delta} + \kappa\sqrt{\log(n\Delta)})$, $\kappa > 0$, and the new the estimator of f

$$\overline{f}_{m,\Delta}(x) = \frac{1}{2\pi} \int_{-m}^{m} e^{-iux} \overline{\varphi}_n(u) \mathrm{d}u, \quad x \in \mathbb{R}.$$

Finally, we introduce the empirical threshold, for some $\alpha \in (0, 1]$ and $\kappa > 0$

$$\widehat{m}_n = \max\left\{ m \geq 0 : |\overline{\varphi}_n(m)| = \frac{\kappa_{n,\Delta}}{\sqrt{n\Delta}} \right\} \wedge (n\Delta)^\alpha.$$

Theorem 3.2. *Assume that* $\mathbb{E}[X_1^4] < \infty$, $\kappa > \frac{e^{2\Delta}}{\Delta}$, $\Delta < \frac{1}{4}\log(n\Delta)$ *and* $n\Delta \to \infty$ *as* $n \to \infty$. *Then, for a positive constant* C_1, *depending on* κ, $\mathbb{E}[X_1^2]$ *and* $\mathbb{E}[X_1^4]$, *and* C_2 *a constant depending on* $\mathbb{E}[X_1^2]$ *and* $\mathbb{E}[X_1^4]$, *it holds*

$$\mathbb{E}[\|\overline{f}_{\widehat{m}_n,\Delta} - f\|^2]$$

$$\leq C_1 \inf_{m \in [0,(n\Delta)^\alpha]} \left\{ \|f_m - f\|^2 + \frac{\log(n\Delta)m}{n\Delta} + \frac{1}{n\Delta} \int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2} + \frac{m^2}{(n\Delta)^2} \right\}$$

$$+ C_2\left(\frac{1}{n\Delta} + (n\Delta)^{\alpha - \kappa^2 \Delta^2 e^{-4\Delta}} \right).$$

If $\kappa > \frac{\sqrt{2}e^{2\Delta}}{\Delta}$ the last additional term is negligible, regardless the value $\alpha \leq 1$ and Theorem 3.2 ensures that the adaptive estimator $\overline{f}_{\widehat{m}_n,\Delta}$ satisfies the same upper bound as in Theorem 3.1. Therefore, it is adaptive and rate optimal, up to a logarithmic term and the multiplicative constant C_1, in the microscopic and mesoscopic regimes defined above. In the macroscopic regimes such that $\Delta := \Delta_n \to \infty$ such that $\Delta_n < \frac{1}{4}\log(n\Delta_n)$ as $n \to \infty$ the estimator is consistent. Note that to establish the adaptive upper bound we imposed a stronger assumption that $\mathbb{E}[X_1^4] < \infty$. In the following numerical study, we recover that the procedure is stable in the choice of κ.

3.3. Numerical results

As for the deconvolution problem, we illustrate the performance of this adaptive estimator for different densities f. We consider the same densities as for the
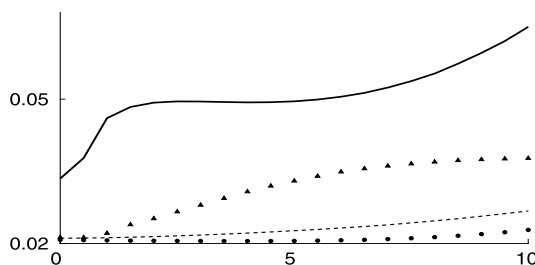
FIG 5. **Decompounding:** *Computations by $M = 1000$ Monte Carlo iterations of the L^2-risks (y axis) for different values of $\kappa \leq \frac{1}{2}(\sqrt{n} - 1)\sqrt{\log(n)}^{-1} = 11.9$ (x axis). Estimation of f from $n = 5000$ ($T = 5000$ and $\Delta = 1$) increments of a compound Poisson process with intensity $\lambda = 1$ and jump density f: Uniform $\mathcal{U}[1,3]$ (plain line), Gaussian $\mathcal{N}(2,1)$ (dots), Gamma $\Gamma(2,1)$ (dotted line) and the mixture $0.7\mathcal{N}(4,1) + 0.3\Gamma(2, \frac{1}{2})$ (triangles).*

deconvolution problem, the Cauchy density excepted as it is not covered by our procedure: it has infinite moments. We compute the adaptive L^2-risks of our procedure over 1000 Monte Carlo iterations for various values of κ. We consider $n = 5000$ and the sampling interval $\Delta = 1$. The results are represented on Figure 5, we observe that the rates are small and stable regardless the value of κ and the density considered.

4. Proofs of Section 3

4.1. Proof of Theorem 3.1

4.1.1. Preliminaries

We establish two technical Lemmas used in the proof of Theorem 3.1.

Lemma 4.1. *Let $m > 0$ and $\zeta > 0$ and define the event*

$$\Omega_{\zeta,\Delta}(m) := \left\{ \forall u \in [-m,m], \ \big|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)\big| \leq \zeta \sqrt{\frac{\log(n\Delta)}{n\Delta}} \right\}.$$

1. *If $\mathbb{E}[X_1^2]$ is finite, then, the following holds for $\eta > 0$ and any $\zeta > \sqrt{\Delta(1 + 2\eta)}$,*

$$\mathbb{P}\big(\Omega_{\zeta,\Delta}(m)^c\big) \leq \frac{\mathbb{E}[X_1^2]}{n\Delta} + 4\frac{m}{(n\Delta)^\eta}.$$

2. *If $\mathbb{E}[X_1^4]$ is finite, then, the following holds for $\eta > 0$ and any $\zeta > \sqrt{\Delta(1 + 2\eta)}$,*

$$\mathbb{P}\big(\Omega_{\zeta,\Delta}(m)^c\big) \leq \frac{C}{(n\Delta)^2} + 4\frac{m}{(n\Delta)^\eta},$$

where C depends on $\mathbb{E}[X_1^4]$ and $\mathbb{E}[X_1^2]$.

Proof of Lemma 4.1. Consider the events

$$A(c) := \left\{ \left| \frac{1}{n} \sum_{j=1}^{n} |Y_{j\Delta}| - \mathbb{E}[|Y_\Delta|] \right| \le c \right\}$$

$$B_{h,\tau}(m) := \left\{ \forall |k| \le \left\lceil \frac{m}{h} \right\rceil, \ |\widehat{\varphi}_{\Delta,n}(kh) - \varphi_\Delta(kh)| \le \tau \sqrt{\frac{\log(n\Delta)}{n\Delta}} \right\}$$

for some positive constants c, h and τ to be determined. First, using that $x \to e^{iux}$ is 1-Lipschitz and that $\mathbb{E}[|Y_\Delta|] \le \Delta\mathbb{E}[|X_1|]$ we get on the event $A(c)$

$$\left| \widehat{\varphi}_{\Delta,n}(u) - \widehat{\varphi}_{\Delta,n}(u+h) \right| \mathbf{1}_{A(c)} \le h\big(\Delta\mathbb{E}[|X_1|] + c \big), \quad \forall u \in \mathbb{R}, \ h > 0. \tag{4.1}$$

If $\mathbb{E}[X_1^2]$ is finite the Markov inequality and the bound $\mathbb{V}[|Y_\Delta|] \le \mathbb{V}[Y_\Delta] = \Delta\mathbb{E}[X_1^2]$ lead to

$$\mathbb{P}\big(A(c)^c \big) \le \frac{\Delta\mathbb{E}[X_1^2]}{c^2 n}. \tag{4.2}$$

If $\mathbb{E}[X_1^4]$ is finite (4.2) can be improved using that

$$\mathbb{E}\left[\left(\sum_{j=1}^{n} (|Y_{j\Delta}| - \mathbb{E}[|Y_\Delta|]) \right)^4 \right] \le n\Delta^2\mathbb{E}[X_1^4] + 3n(n-1)\Delta^2\mathbb{E}[X_1^2],$$

leading to

$$\mathbb{P}\big(A(c)^c \big) \le C \frac{\Delta^2}{c^4 n^2}, \tag{4.3}$$

where C is a constant depending on $\mathbb{E}[X_1^4]$ and $\mathbb{E}[X_1^2]$. Second, we have that

$$\mathbb{P}\big(B_{h,\tau}(m)^c \big) \le \mathbb{P}\left(\exists \ |k| \le \left\lceil \frac{m}{h} \right\rceil, \ |\widehat{\varphi}_{\Delta,n}(kh) - \varphi_\Delta(kh)| > \tau\sqrt{\frac{\log(n\Delta)}{n\Delta}} \right)$$

$$\le \sum_{k=-\lceil m/h \rceil}^{\lceil m/h \rceil} \mathbb{P}\left(|\widehat{\varphi}_{\Delta,n}(kh) - \varphi_\Delta(kh)| > \tau\sqrt{\frac{\log(n\Delta)}{n\Delta}} \right)$$

$$\le \sum_{k=-\lceil m/h \rceil}^{\lceil m/h \rceil} 2\exp\left(-\frac{\tau^2 \log(n\Delta)}{2\Delta} \right) = 4\left\lceil \frac{m}{h} \right\rceil (n\Delta)^{-\tau^2/(2\Delta)}$$

where the last inequality is obtained applying the Hoeffding inequality. Let $|u| \le m$, there exists k such that $u \in [kh - \frac{h}{2}, kh + \frac{h}{2}]$ and we can write that

$$\mathbf{1}_{A(c) \cap B_{h,\tau}(m)} |\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)|$$

$$\le \mathbf{1}_{A(c) \cap B_{h,\tau}(m)} \left(|\widehat{\varphi}_{\Delta,n}(u) - \widehat{\varphi}_{\Delta,n}(kh)| + |\widehat{\varphi}_{\Delta,n}(kh) - \varphi_\Delta(kh)| + |\varphi_\Delta(kh) - \varphi_\Delta(u)| \right).$$

Using (4.1), the definition of $B_{h,\tau}(m)$ and that $x \to e^{iux}$ is 1-Lipschitz, lead to

$$\mathbf{1}_{A(c) \cap B_{h,\tau}(m)} \sup_{u \in [-m,m]} |\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)| \le 2h\Delta\mathbb{E}[|X_1|] + hc + \tau\sqrt{\frac{\log(n\Delta)}{n\Delta}}. \tag{4.4}$$

Taking $c = \Delta$, $h = o\left(\sqrt{\frac{\log(n\Delta)}{n\Delta}}\right)$ such that $h > 1/\sqrt{n\Delta}$ and $\zeta > \tau$, (4.4) shows that $A(c) \cap B_{h,\tau}(m) \subset \Omega_{\zeta,\Delta}(m)$. Moreover, it follows from $h > 1/\sqrt{n\Delta}$, (4.1) and (4.2) that, for all $\eta > 0$

$$\mathbb{P}\left(\Omega^c_{\zeta,\Delta}(m)\right) \leq \mathbb{P}\left(A^c(\Delta)\right) + \mathbb{P}\left(B^c_{h,\tau}(m)\right) \leq \frac{\mathbb{E}[X_1^2]}{n\Delta} + 4\left\lceil \frac{m}{h} \right\rceil (n\Delta)^{-\frac{\tau^2}{2\Delta}}$$
$$\leq \frac{\mathbb{E}[X_1^2]}{n\Delta} + 4m(n\Delta)^{\frac{\Delta-\tau^2}{2\Delta}}.$$

Finally, choosing $\tau^2 = \Delta(1 + 2\eta)$ leads to the result. The second inequality is obtained follows from similar arguments using (4.2) instead of (4.3). □

Lemma 4.2. *Let $\gamma > 0$, define*

$$M^{(\gamma)}_{n,\Delta} := \min\left\{m \geq 0 : |\varphi_\Delta(m)| = \gamma\sqrt{\log(n\Delta)/(n\Delta)}\right\},$$

with the convention $\inf\{\emptyset\} = +\infty$. *Take $\gamma > \zeta > 0$, then, we have*

$$\mathbf{1}_{|u| \leq M^\gamma_{n,\Delta} \wedge m, \Omega_{\zeta,\Delta}(m)}\left|\mathrm{Log}(\widehat{\varphi}_{\Delta,n}(u)) - \mathrm{Log}(\varphi_\Delta(u))\right| \leq \frac{\gamma}{\zeta} \log\left(\frac{\gamma}{\gamma - \zeta}\right) \frac{|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)|}{|\varphi_\Delta(u)|}.$$

Proof of Lemma 4.2. First note that for $|u| \leq M^\gamma_{n,\Delta}$, the ratio $\frac{\varphi'_\Delta}{\varphi_\Delta}$ is well defined. Moreover, on the event $\Omega_{\zeta,\Delta}(m)$ then we have that

$$|\widehat{\varphi}_{\Delta,n}(u)| \geq |\varphi_\Delta(u)| - |\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)| \geq (\gamma - \zeta)\sqrt{\frac{\log(n\Delta)}{n\Delta}} > 0, \quad \forall |u| \leq m.$$

Then, the quantity $\frac{\widehat{\varphi}'_{\Delta,n}}{\widehat{\varphi}_{\Delta,n}}$ is also well defined if $\gamma > \zeta$. For $v \in \mathbb{R}$, notice that

$$\frac{\widehat{\varphi}'_{\Delta,n}(v)}{\widehat{\varphi}_{\Delta,n}(v)} - \frac{\varphi'_\Delta(v)}{\varphi_\Delta(v)} = \frac{\left(-\frac{(\widehat{\varphi}_{\Delta,n}(v) - \varphi_\Delta(v))}{\varphi_\Delta(v)}\right)'}{\left(1 - \frac{(\widehat{\varphi}_{\Delta,n}(v) - \varphi_\Delta(v))}{\varphi_\Delta(v)}\right)}. \tag{4.5}$$

On the event $\Omega_{\zeta,\Delta}(m)$, it holds $\forall u \in [-m \wedge M^\gamma_{n,\Delta}, m \wedge M^\gamma_{n,\Delta}]$

$$|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)| \leq \zeta\sqrt{\log(n\Delta)}(n\Delta)^{-\frac{1}{2}} \leq \frac{\zeta}{\gamma}|\varphi_\Delta(u)|, \tag{4.6}$$

where $\gamma > \zeta$. Then, a Neumann series expansion, with (4.5) and (4.6) gives for $|v| \leq m \wedge M^\gamma_{n,\Delta}$,

$$\frac{\widehat{\varphi}'_{\Delta,n}(v)}{\widehat{\varphi}_{\Delta,n}(v)} - \frac{\varphi'_\Delta(v)}{\varphi_\Delta(v)} = -\sum_{\ell=0}^\infty \left(\frac{\widehat{\varphi}_{\Delta,n}(v) - \varphi_\Delta(v)}{\varphi_\Delta(v)}\right)'\left(\frac{\widehat{\varphi}_{\Delta,n}(v) - \varphi_\Delta(v)}{\varphi_\Delta(v)}\right)^\ell,$$

where

$$\left(\frac{\widehat{\varphi}(v) - \varphi_\Delta(v)}{\varphi_\Delta(v)}\right)'\left(\frac{\widehat{\varphi}_{\Delta,n}(v) - \varphi_\Delta(v)}{\varphi_\Delta(v)}\right)^\ell = \frac{1}{\ell+1}\left[\left(\frac{\widehat{\varphi}_{\Delta,n}(v) - \varphi_\Delta(v)}{\varphi_\Delta(v)}\right)^{\ell+1}\right]'.$$

Using $\widehat{\varphi}_\Delta(0) - \varphi_\Delta(0) = 0$ and (4.6), we get

$$\mathbf{1}_{|u| \le m \wedge M_{n,\Delta}^\gamma, \Omega_{\zeta,\Delta}(m)} \Big| \int_0^u \Big(\frac{\widehat{\varphi}'_{\Delta,n}(v)}{\widehat{\varphi}_{\Delta,n}(v)} - \frac{\varphi'_\Delta(v)}{\varphi_\Delta(v)} \Big) \mathrm{d}v \Big|$$

$$\le \sum_{\ell=0}^\infty \frac{1}{\ell+1} \frac{|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)|^{\ell+1}}{|\varphi_\Delta(u)|^{\ell+1}} \le \frac{|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)|}{|\varphi_\Delta(u)|} \sum_{\ell=0}^\infty \frac{(\zeta/\gamma)^\ell}{\ell+1}$$

$$= \frac{\gamma}{\zeta} \log\Big(\frac{\gamma}{\gamma - \zeta} \Big) \frac{|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)|}{|\varphi_\Delta(u)|}, \tag{4.7}$$

which completes the proof. □

4.1.2. Proof of Theorem 3.1

We have the decomposition

$$\|\widehat{f}_{m,\Delta} - f\|^2 = \|f_m - f\|^2 + \|\widehat{f}_{m,\Delta} - f_m\|^2 = \|f_m - f\|^2 + \frac{1}{2\pi} \int_{-m}^m |\widetilde{\varphi}_n(u) - \varphi(u)|^2 \mathrm{d}u.$$

Let $\gamma > \zeta$, we decompose the second term on the events $\{m \le M_{n,\Delta}^\gamma\}$ and $\Omega_{\zeta,\Delta}(m)$ of Lemma 4.2,

$$\int_{-m}^m |\widetilde{\varphi}_n(u) - \varphi(u)|^2 \mathrm{d}u = \int_{-m \wedge M_{n,\Delta}^\gamma}^{m \wedge M_{n,\Delta}^\gamma} \mathbf{1}_{\Omega_{\zeta,\Delta}(m)} |\widetilde{\varphi}_n(u) - \varphi(u)|^2 \mathrm{d}u$$

$$+ \mathbf{1}_{m > M_{n,\Delta}^\gamma, \Omega_{\zeta,\Delta}(m)} \int_{|u| \in [M_{n,\Delta}^\gamma, m]} |\widetilde{\varphi}_n(u) - \varphi(u)|^2 \mathrm{d}u$$

$$+ \mathbf{1}_{\Omega_{\zeta,\Delta}(m)^c} \int_{-m}^m |\widetilde{\varphi}_n(u) - \varphi(u)|^2 \mathrm{d}u$$

$$:= T_{1,n} + T_{2,n} + T_{3,n}.$$

Fix $\gamma_\Delta = \frac{2\zeta}{1 \wedge \Delta} > \zeta$. On the event $\{|u| \le m \wedge M_{n,\Delta}^{\gamma_\Delta}, \Omega_{\zeta,\Delta}(m)\}$, Lemma 4.2 and equations (3.1), (3.2) and (4.7), along with (4.6), imply

$$|\widetilde{\varphi}_n(u)| \le 1 + \frac{|\mathrm{Log}(\widehat{\varphi}_{\Delta,n}(u)) - \mathrm{Log}(\varphi_\Delta(u))| + |\mathrm{Log}(\varphi_\Delta(u))|}{\Delta} \le 3 + \frac{1}{\Delta} \log\Big(\frac{\gamma}{\gamma - \zeta} \Big) \le 4,$$

consequently $\widetilde{\varphi}_n(u) = \widehat{\varphi}_n(u)$. Then, we get from Lemma 4.2 and the definition of γ_Δ, that

$$\mathbb{E}[T_{1,n}] = \frac{1}{\Delta^2} \int_{-m \wedge M_{n,\Delta}^{\gamma_\Delta}}^{m \wedge M_{n,\Delta}^{\gamma_\Delta}} \mathbb{E}\Big[\mathbf{1}_{\Omega_{\zeta,\Delta}(m)} \big| \mathrm{Log}(\widehat{\varphi}_{\Delta,n}(u)) - \mathrm{Log}(\varphi_\Delta(u)) \big|^2 \Big] \mathrm{d}u$$

$$\le \frac{1}{\Delta^2} \int_{-m}^m \frac{\mathbb{E}\big[|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)|^2 \big]}{|\varphi_\Delta(u)|^2} \mathrm{d}u.$$

Direct computations together with the Lévy-Kintchine formula lead to

$$\mathbb{E}\big[|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)|^2\big] = \frac{1 - |\varphi_\Delta(u)|^2}{n} \leq \frac{\big(2\Delta|\mathrm{Re}(\varphi(u)) - 1|\big) \wedge 1}{n} \leq \frac{2\Delta \wedge 1}{n}.$$

We derive that

$$\mathbb{E}[T_{1,n}] \leq \frac{2\Delta \wedge 1}{n\Delta^2} \int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2} \leq \frac{2}{n\Delta} \int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2}.$$

Next, fix $\zeta > \sqrt{5\Delta}$, Lemma 4.1 with $\eta = 2$ gives

$$\mathbb{E}[T_{3,n}] \leq 2 \; 5^2 m \Big(\frac{\mathbb{E}[X_1^2]}{n\Delta} + 4\frac{m}{(n\Delta)^2} \Big).$$

Moreover, using $|\varphi_\Delta(u)| \geq e^{-2\Delta}$, $\forall u \in \mathbb{R}$ together with the constraint $\Delta \leq \delta \log(n\Delta)$, $\delta < \frac{1}{4}$, we get

$$|\varphi_\Delta(u)| \geq (n\Delta)^{-2\delta} > \gamma_\Delta \sqrt{\log(n\Delta)/(n\Delta)}, \quad \forall u \in \mathbb{R}.$$

Finally, $M_{n,\Delta}^{\gamma_\Delta} = +\infty$, $\forall \zeta > 0$ and $T_{2,n} = 0$ almost surely. Gathering all terms completes the proof. □

4.2. Proof of Theorem 3.2

Preliminary

Lemma 4.3. *Assume that $\mathbb{E}[X_1^4] < \infty$. Let $\eta > 0$, $\alpha \in (0,1]$ and $c(\Delta) = \kappa\Delta e^{-2\Delta}$. Then, for some positive constant C depending only on $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$, it holds for $\widehat{\varphi}_n$ defined in (3.2) that, for all $u \in [-(n\Delta)^\alpha, (n\Delta)^\alpha]$,*

$$\mathbb{P}\Big(|\widehat{\varphi}_n(u) - \varphi(u)| \geq \frac{\kappa\sqrt{\log n\Delta}}{\sqrt{n\Delta}}\Big) \leq 2(n\Delta)^{-c(\Delta)^2} + \frac{C}{(n\Delta)^2} + 4(n\Delta)^{\alpha-\eta}.$$

Proof. We use Lemmas 4.1 and 4.2 with $\gamma_\Delta = \frac{2\zeta}{1\vee\Delta}$ and $\zeta > \sqrt{\Delta(1+2\eta)}$, $\eta > 0$. First it holds that

$$\mathbb{P}\Big(|\widehat{\varphi}_n(u) - \varphi(u)| \geq \kappa\sqrt{\frac{\log(n\Delta)}{n\Delta}}\Big)$$

$$= \mathbb{P}\Big(|\mathrm{Log}(\widehat{\varphi}_{\Delta,n}(u)) - \mathrm{Log}(\varphi_\Delta(u))| \geq \kappa\Delta\sqrt{\frac{\log(n\Delta)}{n\Delta}}\Big)$$

$$\leq \mathbb{P}\Big(|\widehat{\varphi}_{\Delta,n}(u) - \varphi_\Delta(u)| \geq |\varphi_\Delta(u)|\kappa\Delta\sqrt{\frac{\log(n\Delta)}{n\Delta}}\Big) + \mathbb{P}(\Omega_{\zeta,\Delta}^c((n\Delta)^\alpha)).$$

Then, we derive from the Hoeffding inequality and Lemma 4.1 that

$$\mathbb{P}\Big(|\widehat{\varphi}_n(u) - \varphi(u)| \geq \kappa\sqrt{\frac{\log(n\Delta)}{n\Delta}}\Big) \leq 2(n\Delta)^{-c(\Delta)^2} + \frac{C}{(n\Delta)^2} + 4(n\Delta)^{\alpha-\eta},$$

where C depends on $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$. □

Proof of Theorem 3.2

   Step 1: An upper bound for $\overline{f}_{m,\Delta}$. Let $0 < m < (n\Delta)^\alpha$, Parseval equality and (3.6) leads to

$$\mathbb{E}[\|\overline{f}_{m,\Delta} - f\|^2] \leq \|f_m - f\|^2 + \frac{1}{2\pi} \int_{-m}^{m} \mathbb{E}\big[|\overline{\varphi}_n(u) - \varphi(u)|^2\big] \mathrm{d}u,$$

where

$$\mathbb{E}\big[|\overline{\varphi}_n(u) - \varphi(u)|^2\big]$$
$$= \mathbb{E}\big[|\widetilde{\varphi}_n(u) - \varphi(u)|^2 \mathbf{1}_{|\widetilde{\varphi}_n(u)| \geq \frac{\kappa_{n,\Delta}}{\sqrt{n\Delta}}}\big] + |\varphi(u)|^2 \mathbb{P}\Big(|\widetilde{\varphi}_n(u)| < \frac{\kappa_{n,\Delta}}{\sqrt{n\Delta}}\Big).$$

The first term in the right hand side is bounded using Theorem 3.1. For the second term, recall that $\kappa_{n,\Delta} = e^{2\Delta} + \kappa\sqrt{\log(n\Delta)}$ and decompose it on the set $\mathcal{A} = \{u, |\varphi(u)| < \frac{e^{2\Delta} + 2\kappa\sqrt{\log(n\Delta)}}{\sqrt{n\Delta}}\}$ this leads to

$$|\varphi(u)|^2 \mathbb{P}\Big(|\widetilde{\varphi}_n(u)| < \frac{\kappa_{n,\Delta}}{\sqrt{n\Delta}}\Big)$$
$$\leq \frac{(e^{2\Delta} + 2\kappa\sqrt{\log(n\Delta)})^2}{n\Delta} + \mathbb{P}\Big(|\widetilde{\varphi}_n(u) - \varphi(u)| \geq \frac{\kappa\sqrt{\log n\Delta}}{\sqrt{n\Delta}}\Big)$$
$$\leq \frac{1 + \big(1 + 2\kappa\sqrt{\log(n\Delta)}\big)^2}{n\Delta} + \mathbb{P}\Big(|\widehat{\varphi}_n(u) - \varphi(u)| \geq \frac{\kappa\sqrt{\log n\Delta}}{\sqrt{n\Delta}}\Big)$$
$$\leq Ce^{4\Delta} \frac{\log(n\Delta)}{n\Delta}$$

from Lemma 4.3 with $\eta > 2$, $\kappa > e^{2\Delta}/\Delta$ and where C depends on κ, $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$. Finally, gathering all the above inequalities, we get the following upper bound for $\overline{f}_{m,\Delta}$, $0 < m \leq (n\Delta)^\alpha$,

$$\mathbb{E}[\|\overline{f}_{m,\Delta} - f\|^2] \leq \|f_m - f\|^2 + Ce^{4\Delta} \frac{\log(n\Delta)}{n\Delta} m + \frac{2}{n\Delta} \int_{-m}^{m} \frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2} + \frac{2^3 5^2 m^2}{(n\Delta)^2},$$

(4.8)

where C depends on κ, $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$.

   In the sequel C denotes a constant depending on κ, $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$ whose value may change from line to line.

   Step 2: Adaptation. Let $0 < m \leq (n\Delta)^\alpha$ be fixed. Consider the event $\mathcal{E} = \{\widehat{m}_n < m\}$, on this event we control the surplus in the bias of the estimator $\widetilde{f}_{\widehat{m}_n}$. Using the inequality $|\varphi|^2 \leq 2|\widetilde{\varphi}_n|^2 + 2|\varphi - \widetilde{\varphi}_n|^2$, along with the definition of $\widehat{m}_n$ and (4.8), give

$$\mathbb{E}\Big[\mathbf{1}_{\mathcal{E}} \int_{|u| \in [\widehat{m}_n, m]} |\varphi(u)|^2 \mathrm{d}u\Big] \leq 2\mathbb{E}\Big[\mathbf{1}_{\mathcal{E}} \int_{|u| \in [\widehat{m}_n, m]} \frac{\kappa_{n,\Delta}^2}{n\Delta} \mathrm{d}u\Big] + 2\int_{-m}^{m} \mathbb{E}[|\widetilde{\varphi}_n(u) - \varphi(u)|^2] \mathrm{d}u$$

$$\leq 4\frac{\kappa_{n,\Delta}^2 m}{n\Delta} + \frac{4}{n\Delta}\int_{-m}^{m}\frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2}$$
$$+ 2^2 5^2 \mathbb{E}[X_1^2]\frac{m}{n\Delta} + 2^4 5^2\frac{m^2}{(n\Delta)^2} + Ce^{4\Delta}\frac{\log(n\Delta)}{n\Delta}m.$$

Recall that $\kappa_{n,\Delta} = e^{2\Delta} + \kappa\sqrt{\log(n\Delta)}$ together with (4.8), this implies immediately that, on the event $\mathcal{E}$,

$$\mathbb{E}[\|\overline{f}_{\widehat{m}_n,\Delta} - f\|^2 \mathbf{1}_{\mathcal{E}}] \leq \|f_m - f\|^2 + Ce^{4\Delta}\frac{\log(n\Delta)m}{n\Delta} + \frac{5}{n\Delta}\int_{-m}^{m}\frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2} + \frac{2^5 5^2 m^2}{(n\Delta)^2}.$$

Second, consider the complement set $\mathcal{E}^c$, where we control the surplus in the variance of $\widetilde{f}_{\widehat{m}_n}$. By the definition of $\widehat{m}_n$, it holds

$$\mathbb{E}\Big[\int_{|u|\in[m,\widehat{m}_n]} |\overline{\varphi}_n(u) - \varphi(u)|^2\mathrm{d}u\mathbf{1}_{\mathcal{E}^c}\Big] \leq \int_{|u|\in[m,(n\Delta)^\alpha]} \mathbb{E}[|\overline{\varphi}_n(u) - \varphi(u)|^2]\mathrm{d}u.$$

Let $\eta > 2$, such that $\alpha - \eta < -1$, and $\zeta > \sqrt{\Delta(1+2\eta)}$ and γ_Δ as in the proof of Theorem 3.1 (leading to $M_{n,\Delta}^{\gamma_\Delta} = +\infty$). Then, Lemmas 4.1 (decomposing on $\Omega_{\zeta,\Delta}((n\Delta)^\alpha)$) and 4.2 lead to

$$\mathbb{E}[|\overline{\varphi}_n(u) - \varphi(u)|^2] \leq |\varphi(u)|^2 + \mathbb{E}[|\widehat{\varphi}_n(u) - \varphi(u)|^2]$$
$$\leq |\varphi(u)|^2 + \frac{2}{n\Delta|\varphi_\Delta(u)|^2} + \frac{\mathbb{E}[X_1^2]}{n\Delta} + \frac{4}{n\Delta}.$$

First, on the event $\{|\varphi(u)| > e^{2\Delta}/\sqrt{n\Delta}\}$, we obtain

$$\mathbb{E}[|\overline{\varphi}_n(u) - \varphi(u)|^2] \leq |\varphi(u)|^2(6 + \mathbb{E}[X_1^2]).$$

Consequently, define $C_0 := 6 + \mathbb{E}[X_1^2]$, then,

$$\mathbb{E}\Big[\int_{|u|\in[m,\widehat{m}_n]} |\overline{\varphi}_n(u) - \varphi(u)|^2\,\mathbf{1}_{\{|\varphi(u)|>e^{2\Delta}/\sqrt{n\Delta}\}}\mathrm{d}u\,\mathbf{1}_{\mathcal{E}^c}\Big] \leq C_0\int_{[-m,m]^c} |\varphi(u)|^2\mathrm{d}u.$$

Next, using that $|\overline{\varphi}_n(u)| \leq 4$ and the definition of $\widehat{m}_n$, we derive that

$$\mathbb{E}\Big[\int_{|u|\in[m,\widehat{m}_n]} |\overline{\varphi}_n(u) - \varphi(u)|^2\,\mathbf{1}_{\{|\varphi(u)|\leq e^{2\Delta}/\sqrt{n\Delta}\}}\mathrm{d}u\,\mathbf{1}_{\mathcal{E}^c}\Big]$$
$$\leq \int_{|u|\in[m,(n\Delta)^\alpha]} |\varphi(u)|^2\mathrm{d}u + 5^2\int_{|u|\in[m,(n\Delta)^\alpha]} \mathbb{P}\big(|\widehat{\varphi}_n(u)| \geq \kappa_{n,\Delta}/\sqrt{n\Delta}\big)\mathbf{1}_{\{|\varphi(u)|\leq e^{2\Delta}/\sqrt{n\Delta}\}}\mathrm{d}u$$
$$\leq \int_{|u|\in[m,(n\Delta)^\alpha]} |\varphi(u)|^2\mathrm{d}u + 5^2\int_{|u|\in[m,(n\Delta)^\alpha]} \mathbb{P}\big(|\widehat{\varphi}_n(u) - \varphi(u)| > \kappa\sqrt{\log(n\Delta)/(n\Delta)}\big)\mathrm{d}u.$$

Finally, we bound the last term using Lemma 4.3 with $\eta > 3$, such that $2\alpha - \eta < -1$ and $\zeta > \sqrt{\Delta(1 + 2\eta)}$, it follows that

$$\int_{|u| \in [m, (n\Delta)^\alpha]} \mathbb{P}\Big(|\widehat{\varphi}_n(u) - \varphi(u)| > \kappa\sqrt{\frac{\log(n\Delta)}{n\Delta}}\Big) \mathrm{d}u \leq 4(n\Delta)^{\alpha - c(\Delta)^2} + \frac{C'}{n\Delta},$$

where C' depends on $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$. Putting the above together, we have shown that for a positive constant C_1, depending on κ, $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$, and C_2 a constant depending on $\mathbb{E}[X_1^2]$ and $\mathbb{E}[X_1^4]$

$$\mathbb{E}\big[\|\overline{f}_{\widehat{m}_n, \Delta} - f\|^2\big] \leq C_1\Big(\|f_m - f\|^2 + \frac{\log(n\Delta)m}{n\Delta} + \frac{1}{n\Delta}\int_{-m}^m \frac{\mathrm{d}u}{|\varphi_\Delta(u)|^2} + \frac{m^2}{(n\Delta)^2}\Big)$$
$$+ C_2\Big(\frac{1}{n\Delta} + (n\Delta)^{\alpha - c(\Delta)^2}\Big).$$

Taking the infimum in m completes the proof. $\qquad\square$

References

[1] A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probability theory and related fields*, 113(3):301–413, 1999. MR1679028

[2] J.-P. Baudry, C. Maugis, and B. Michel. Slope heuristics: overview and implementation. *Statistics and Computing*, 22(2):455–470, 2012. MR2865029

[3] L. Birgé, P. Massart, et al. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998. MR1653272

[4] B. Buchmann and R. Grübel. Decompounding: an estimation problem for Poisson random sums. *Ann. Statist.*, 31(4):1054–1074, 2003. MR2001642

[5] C. Butucea. Deconvolution of supersmooth densities with smooth noise. *Canadian Journal of Statistics*, 32(2):181–192, 2004. MR2064400

[6] C. Butucea and A. B. Tsybakov. Sharp optimality in density deconvolution with dominating bias. I. *Teor. Veroyatn. Primen.*, 52(1):111–128, 2007. MR2354572

[7] C. Butucea and A. B. Tsybakov. Sharp optimality in density deconvolution with dominating bias. II. *Teor. Veroyatn. Primen.*, 52(2):336–349, 2007. MR2742504

[8] R. J. Carroll and P. Hall. Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186, 1988. MR0997599

[9] A. J. Coca. Efficient nonparametric inference for discretely observed compound poisson processes. *Probability Theory and Related Fields*, 170(1-2):475–523, 2018. MR3748329

[10] A. J. Coca. Adaptive nonparametric estimation for compound poisson processes robust to the discrete-observation scheme. 1803.09849, [math.ST].

[11] F. Comte, C. Duval, and V. Genon-Catalot. Nonparametric density estimation in compound poisson processes using convolution power estimators. *Metrika*, 77(1):163–183, 2014. MR3152023

[12] F. Comte and C. Lacour. Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):601–627, 2011. MR2853732

[13] F. Comte and C. Lacour. Anisotropic adaptive kernel deconvolution. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 49, pages 569–609. Institut Henri Poincaré, 2013. MR3088382

[14] F. Comte, Y. Rozenholc, and M.-L. Taupin. Finite sample penalization in adaptive density deconvolution. *J. Stat. Comput. Simul.*, 77(11-12):977–1000, 2007. MR2416478

[15] R. Cont and A. De Larrard. Price dynamics in a markovian limit order market. *SIAM Journal on Financial Mathematics*, 4(1):1–25, 2013. MR3032934

[16] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, pages 425–455, 1994. MR1311089

[17] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: asymptopia? *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 301–369, 1995. MR1323344

[18] C. Duval. Density estimation for compound Poisson processes from discrete data. *Stochastic Process. Appl.*, 123(11):3963–3986, 2013. MR3091096

[19] C. Duval. When is it no longer possible to estimate a compound poisson process? *Electronic Journal of Statistics*, 8(1):274–301, 2014. MR3189556

[20] C. Duval and J. Kappus. Nonparametric adaptive estimation for grouped data. *Journal of Statistical Planning and Inference*, 182:12–28, 2017. MR3574505

[21] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media, 2013. MR1458613

[22] J. Fan. On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, pages 1257–1272, 1991. MR1126324

[23] A. Goldenshluger and O. Lepski. Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics*, pages 1608–1632, 2011. MR2850214

[24] A. Goldenshluger and O. Lepski. On adaptive minimax density estimation on $\mathbb{R}^d$. *Probability Theory and Related Fields*, 159(3-4):479–543, 2014. MR3230001

[25] S. Gugushvili, F. Van der Meulen, and P. Spreij. Nonparametric bayesian inference for multidimensional compound Poisson processes. *Modern Stochastics: Theory and Applications*, 2(1):1–15, Mar 2015. MR3356921

[26] J. Kappus and G. Mabon. Adaptive density estimation in deconvolution problems with unknown error distribution. *Electronic Journal of Statistics*, 8(2):2879–2904, 2014. MR3299125

[27] C. Lacour, P. Massart, and V. Rivoirard. Estimator selection: a new method with applications to kernel density estimation. *Sankhya A*, 79(2):298–335,

2017. MR3707423

[28] O. V. Lepski and T. Willer. Oracle inequalities and adaptive estimation in the convolution structure density model. *Ann. Statist.*, 47(1):233–287, 02 2019. MR3909933

[29] M. Lerasle. Optimal model selection in density estimation. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 48, pages 884–908. Institut Henri Poincaré, 2012. MR2976568

[30] K. Lounici and R. Nickl. Uniform risk bounds and confidence bands in wavelet deconvolution. *Annals of Statistics*, 39:201–231, 2011. MR2797844

[31] P. Massart. *Concentration inequalities and model selection*, volume 6. Springer, 2007. MR2319879

[32] M. H. Neumann. On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist.*, 7(4):307–330, 1997. MR1460203

[33] M. H. Neumann and M. Reiß. Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15(1):223–248, 2009. MR2546805

[34] M. Pensky, B. Vidakovic, et al. Adaptive wavelet estimator for nonparametric density deconvolution. *The Annals of Statistics*, 27(6):2033–2053, 1999. MR1765627

[35] G. Rebelles. Structural adaptive deconvolution under l_p-losses. *Mathematical Methods of Statistics*, 25(1):26–53, 2016. MR3480609

[36] P. Reynaud-Bouret, V. Rivoirard, and C. Tuleau-Malot. Adaptive density estimation: a curse of support? *Journal of Statistical Planning and Inference*, 141(1):115–139, 2011. MR2719482

[37] L. A. Stefanski. Rates of convergence of some estimators in a class of deconvolution problems. *Statistics & Probability Letters*, 9(3):229–235, 1990. MR1045189

[38] B. van Es, S. Gugushvili, and P. Spreij. A kernel type nonparametric density estimator for decompounding. *Bernoulli*, 13(3):672–694, 2007. MR2348746