Electronic Journal of Statistics
ISSN: 1935-7524

Deconvolution of repeated measurements corrupted by unknown noise

Jérémie Capitao-Miniconi[1], Elisabeth Gassiat[1] and Luc Lehéricy[2]

[1] *Université Paris-Saclay,*
CNRS, Laboratoire de mathématiques d'Orsay,
91405, Orsay, France, e-mail: jeremie.capitao-miniconi@universite-paris-saclay.fr*;*
elisabeth.gassiat@universite-paris-saclay.fr

[2] *Université Côte d'Azur,*
CNRS, LJAD, France, e-mail: luc.lehericy@univ-cotedazur.fr

Abstract: Recent advances have demonstrated the possibility of solving the deconvolution problem without prior knowledge of the noise distribution. In this paper, we study the repeated measurements model, where information is derived from two measurements of X perturbed independently by additive errors. Our contributions include establishing identifiability without any assumption on the noise except for coordinate independence. We propose an estimator of the density of the signal for which we provide rates of convergence, and prove that it reaches the minimax rate in the case where the support of the signal is compact. Additionally, we propose a model selection procedure for adaptive estimation. Numerical simulations demonstrate the effectiveness of our approach even with limited sample sizes.

Keywords and phrases: deconvolution, repeated measurements, unknwon noise.

Contents

1. Introduction

Density deconvolution is one of the classical topics in nonparametric statistics and has garnered significant attention over the past decades. The aim is to identify the density of some random variable X, the signal, which cannot be observed directly, but is contaminated by some additive error ε, the noise, independent of X. Most of the literature on the deconvolution problem considers the situation where the distribution of the noise is perfectly known, see [2], [7], [23], [24], for early references, see also the book [20]. However, knowledge of the noise distribution is unrealistic in practice. In some situations, it is possible to get a pure noise sample so that the noise distribution may be estimated separately and plugged into methods that assume the noise distribution known, see [12], [21]. It has been proved recently [8] that, under very mild assumptions, it is possible to solve the deconvolution problem without knowing the noise distribution and with no sample of the noise, for multivariate signals. This paper is a continuation of [8] in a specific structured model.

In the present work, we are interested in the case where information can be drawn from repeated measurements of X, so that the multivariate signal is the repetition of X. This framework is known as model of repeated measurements. The observations are

$$Y_j^{(i)} = X_j + \varepsilon_j^{(i)} \ ; \ i = 1, \dots, p \ ; \ j = 1, \dots, n$$

where the random variables X_j and $\varepsilon_j^{(i)}$, $j = 1, \dots, p$, $i = 1, \dots, n$ are independently distributed. We consider in this work the case $p = 2$, though our results can be extended to any p, see also the discussion in Section 7.

By making substractions of the coordinates of the observations, it is possible to estimate the noise distribution consistently when it is assumed to be symmetric, and consequently estimate the signal density, see [5], [6] and [13]. The symmetry assumption can be dropped using Kotlarski's identity [14]. Further works use this identity to propose estimation strategies not relying on the symmetry assumption, see [17], [3] and [15]. However, all these works require that the characteristic function of the signal and the characteristic function of the noise vanish nowhere. The identifiability result of [22] relaxes this non-vanishing assumption, although it still prevents the characteristic function from vanishing on open non empty sets.

The aim of our work is to prove that it is not needed to assume the characteristic functions do not vanish to be able to estimate the signal and the noise distributions. Our main contributions are as follows.

- We first prove an identifiability result in general settings, see Theorem 2.1. Our only assumption on the noise is that its coordinates are independently distributed. We do not assume anything on its characteristic function, which may vanish anywhere. Our assumption on the signal is that its Laplace transform is finite everywhere and grows at most as the exponential of a power function, whatever the power ρ is. This is equivalent to an assumption on the tails of the distribution of the signal. The repeated observations setting allows to improve on the assumptions in [8].
- We further develop for the repeated observations model the estimation methodology introduced in [8]. We propose an estimation procedure for the density of the signal X in $\mathbb{R}^d$ and we get upper bounds on the integrated quadratic risk of our estimator in Theorem 4.2. We get an upper bound decreasing as $(\log\log(n)/\log(n))^{2\beta/\rho}$ for ordinary smooth signals and as $\exp\left\{-c(\gamma)\left(\log(n)/\log\log(n)\right)^{\frac{\alpha}{\rho}\wedge 1}\right\}$ for supersmooth signals. Ordinary smooth distributions are the ones for which the characteristic function decays polynomially at infinity, and supersmooth distributions are the ones for which the characteristic function decays as the exponential of a power, as introduced in early literature on deconvolution, see [7]. Here $\rho \geq 1$ is a parameter that depends on the tail of the distribution ($\rho = 1$ corresponds to a compactly supported distribution, and $\rho = 2$ to a sub-Gaussian distribution), β is the power in the polynomial decrease of the ordinary smooth distribution, α is the power and $-\gamma$ the constant in the exponential decay of the supersmooth distribution, and $c(\gamma)$ is a constant depending on γ. In the compact case ($\rho = 1$), we show in Theorem 4.4 that the rate for ordinary smooth signals is minimax.
- We propose a model selection procedure to obtain an adaptive estimator with the same rate of convergence as the estimator with a known tail parameter ρ, see Theorem 4.3. To cover situations in which the estimation rate may be improved when the characteristic function of the noise does not vanish and has specific decay, we construct a data-driven combination of our estimator $\widehat{f}$ and of the estimator $\widehat{f}^{CK}$ built in [3]. We prove in Theorem 5.1 that this combination achieves the best rate of convergence among the two, in all situations studied in [3].
- Finally, we present numerical simulations in Section 6. In various settings for the signal distribution and for the noise distribution, we find that our estimator has surprisingly good behaviour even with a small sample size ($n = 100$). We compare our results with the experiments of [3] and find that our estimator outperforms the one from [3] for Gaussian signals, which are supersmooth and with light tails. Moreover, our simulations indicate

that our procedure performs well even when the tails of the distribution of the signal are too heavy for our theoretical results to apply, as its results are comparable to the ones of [3] for Gamma and Bigamma signals. Finally, we propose a data-driven method to select the hyper parameters involved in our procedure.

Possible further works are discussed in Section 7. Detailed proofs can be found in Section 8. In particular, part of the proof of Proposition 4.1 corrects an error in the original part of the proof of the analogous proposition in [8].

2. Setting and identifiability Theorem

Consider the repeated measurements model with 2 repetitions:

$$Y = \begin{pmatrix} Y^{(1)} \\ Y^{(2)} \end{pmatrix} = \begin{pmatrix} X \\ X \end{pmatrix} + \begin{pmatrix} \varepsilon^{(1)} \\ \varepsilon^{(2)} \end{pmatrix} = r(X) + \varepsilon, \tag{2.1}$$

in which, for $i \in \{1, 2\}$, $Y^{(i)}, X, \varepsilon^{(i)} \in \mathbb{R}^d$, and $r : x \in \mathbb{R}^d \mapsto (x, x) \in \mathbb{R}^d \times \mathbb{R}^d$. We assume that the random variables X and ε are independent, and we consider independent and identically distributed observations Y_j, $j = 1, \dots, n$, following model (2.1).

We shall not assume that the distribution of the noise ε is known, instead the only assumption we will make on the noise is the following.

(H1) $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are independent random variables.

Let us now introduce our assumption on the Laplace transform of X.

(H2) There exists $\rho > 0$, $a > 0$ and $b > 0$ such that for all $\lambda \in \mathbb{R}^d$, $\mathbb{E}\left[\exp\left(\lambda^\top X\right)\right] \leq a \exp\left(b\|\lambda\|^\rho\right)$, where $\|\cdot\|$ denotes the Euclidean norm.

Note that by Chernoff's bound, this is equivalent to assuming that the tails of the distribution of X satisfy $\mathbb{P}(\|X\| \geq t) = O(\exp(ct^{1+1/(\rho-1)}))$ when $\rho > 1$, X a.s. bounded when $\rho = 1$, and $X = 0$ when $\rho < 1$.

Under **(H2)**, the characteristic function of the signal X can be extended into a multivariate analytic function denoted by

$$\begin{aligned} \Phi_X : \mathbb{C}^d &\longrightarrow \mathbb{C} \\ z &\longmapsto \mathbb{E}\left[\exp\left(iz^\top X\right)\right] \end{aligned}$$

Obviously, if no centering constraint is put on the signal or on the noise, it is possible to translate the signal by a fixed vector $m \in \mathbb{R}^d$ and the noise by $-m$ without changing the observation. The model can thus be identifiable only up to translation. We prove the following identifiability theorem.

Theorem 2.1. *Assume that X and X' satisfy **(H2)** and $\varepsilon \sim \mathbb{Q}$, $\tilde{\varepsilon} \sim \tilde{\mathbb{Q}}$ satisfy **(H1)**. Then $\mathbb{P}_{r(X)} * \mathbb{Q} = \mathbb{P}_{r(X')} * \tilde{\mathbb{Q}}$ implies $\mathbb{P}_X = \mathbb{P}_{X'}$ and $\mathbb{Q} = \tilde{\mathbb{Q}}$ up to translation.*

The proof of Theorem 2.1 is detailed in Section 8.1. It may be seen as starting similarly as the proof of Theorem 2.1 in [8] and then taking into account the particular form of the characteristic function of the observations in model (2.1).

Our result improves on earlier results concerning the assumption on the noise distribution. Indeed, the identifiability result of [14] (further used in [17], [3] and [15]) requires that the characteristic functions of both the noise and the signal vanish nowhere. Our approach differs from the identifiability Lemma 2.1 of [22] in the sense that they do not make any assumption on the tails of the distribution of the signal but use some intricate assumption on the non zero sets of the characteristic functions of the noise and the signal, which excludes noise characteristic functions vanishing on non empty open subsets of $\mathbb{R}$.

3. Estimation procedure

From now on, we assume that the distribution $\mathbb{P}_X$ of the signal has a density f with respect to the Lebesgue measure on $\mathbb{R}^d$. We shall assume that X satisfies **(H2)**, and we also assume that an upper bound ρ_0 is known on ρ, where ρ is given by **(H2)**. Moreover, because the distribution of X may be identified only up to translation, we assume that X is centered with no loss of generality.

The first step in the estimation procedure is the estimation of the characteristic function of the signal by a method inspired by the proof of the identifiability theorem. A key step in the proof of Theorem 2.1 is the fact that, if a multivariate analytic function ϕ has growth as in **(H2)**, is such that $\phi(0) = 1$, and for all $t \in \mathbb{R}^d$, $\overline{\phi(t)} = \phi(-t)$, and satisfies, for all t_1, t_2 in a neighborhood of 0 in $\mathbb{R}^d$,

$$\phi(t_1 + t_2)\Phi_X(t_1)\Phi_X(t_2) = \Phi_X(t_1 + t_2)\phi(t_1)\phi(t_2),$$

then $\phi = \Phi_X$ up to translation, that is up to multiplication by a factor $\exp(ic^\top t)$ for some constant vector $c \in \mathbb{R}^d$. We constrain the gradient function of ϕ to be null at 0 to impose $c = 0$.

In other words, if we define, for $\nu > 0$,

$$M(\phi;\nu|\Phi_X) = \int_{[-\nu,\nu]^d\times[-\nu,\nu]^d} |\phi(t_1+t_2)\Phi_X(t_1)\Phi_X(t_2) - \Phi_X(t_1+t_2)\phi(t_1)\phi(t_2)|^2 |\Phi_{\varepsilon^{(1)}}(t_1)\Phi_{\varepsilon^{(2)}}(t_2)|^2 dt_1 dt_2,$$

then Φ_X is the only minimizer of $M(\cdot;\nu|\Phi_X)$ over a well chosen set of multivariate analytic functions. We will construct an estimator of the criterion $M(\cdot;\nu|\Phi_X)$ based on the observations and minimize it to get an estimator of the characteristic function of the signal. Let us now describe the details of this procedure.

For any $S > 0$, let $\Upsilon_{\rho,S}$ be the subset of multivariate analytic functions from

$\mathbb{C}^d$ to $\mathbb{C}$ defined as follows:

$$\Upsilon_{\rho,S} = \Bigg\{ \phi \text{ analytic s.t. } \forall z \in \mathbb{R}^d, \overline{\phi(z)} = \phi(-z), \phi(0) = 1, \forall a, \partial_a^1 \phi(0) = 0 \\ \text{and } \forall j \in \mathbb{N}^d \setminus \{0\}, \left| \frac{\partial^j \phi(0)}{j!} \right| \leq \frac{S^{\|j\|_1}}{(\|j\|_1)^{\|j\|_1/\rho}} \Bigg\},$$

where $j! = \prod_{a=1}^d j_a!$ and $\partial^j = \partial_1^{j_1} \ldots \partial_d^{j_d}$. By Lemma 3.1 in [8], if Φ_X satisfies Assumption **(H2)**, then there exists S (depending only on a, b and ρ) such that $\Phi_X \in \Upsilon_{\rho,S}$, and conversely all $\phi \in \Upsilon_{\rho,S}$ satisfy Assumption **(H2)**for some a and b depending only on ρ and S. Thus for large enough S, Φ_X is the only minimizer of $M(\cdot;\nu|\Phi_X)$ over $\Upsilon_{\rho,S}$.

Fix some constant $\nu_{\text{est}} > 0$ and define M_n as, for all $\phi \in \Upsilon_{\rho,S}$,

$$M_n(\phi) = \int_{[-\nu_{\text{est}},\nu_{\text{est}}]^d \times [-\nu_{\text{est}},\nu_{\text{est}}]^d} |\phi(t_1 + t_2)\tilde{\phi}_n(t_1, 0)\tilde{\phi}_n(0, t_2) \\ - \tilde{\phi}_n(t_1, t_2)\phi(t_1)\phi(t_2)|^2 dt_1 dt_2,$$

where for all $(t_1, t_2) \in \mathbb{C}^d \times \mathbb{C}^d$,

$$\tilde{\phi}_n(t_1, t_2) = \frac{1}{n} \sum_{\ell=1}^n \exp\left\{ i t_1^\top Y_\ell^{(1)} + i t_2^\top Y_\ell^{(2)} \right\}.$$

As n tends to infinity, $M_n(\phi)$ converges a.s. to $M(\phi; \nu_{\text{est}}|\Phi_X)$. We shall minimize M_n over multivariate polynomials. We thus introduce, for all $m \in \mathbb{N}$, the set $\mathbb{C}_m[X_1, \ldots, X_d]$ of multivariate polynomials in d variables with total degree at most m and coefficients in $\mathbb{C}$. For all n, let us choose m_n, and define $\widehat{\Phi}_{n,\rho}$ as a (up to $1/n$) measurable minimizer of the functional $\phi \mapsto M_n(\phi)$ over $\mathbb{C}_{m_n}[X_1, \ldots, X_d] \cap \Upsilon_{\rho,S}$, that is,

$$\widehat{\Phi}_{n,\rho} \in \mathbb{C}_{m_n}[X_1, \ldots, X_d] \cap \Upsilon_{\rho,S}$$

and

$$\forall \phi \in \mathbb{C}_{m_n}[X_1, \ldots, X_d] \cap \Upsilon_{\rho,S},\ M_n(\widehat{\Phi}_{n,\rho}) \leq M_n(\phi) + \frac{1}{n}.$$

The following proposition states some very general consistency property.

Proposition 3.1. *If*

$$\lim_{n\to+\infty} m_n = +\infty,$$

then

$$\int_{[-\nu_{est},\nu_{est}]^d} |\widehat{\Phi}_{n,\rho}(t) - \Phi_X(t)|^2 dt = o_{\mathbb{P}}(1)$$

for any $\rho > 0$, $S > 0$, $\Phi_X \in \Upsilon_{\rho,S}$, and any distribution of ε provided $\varepsilon^{(1)}$ and $\varepsilon^{(2)}$ are independent random variables having finite first moment.

The proof of Proposition 3.1 can be found in Section 8.2. We shall prove in Proposition 4.1 that $\widehat{\Phi}_{n,\rho}$ converges to Φ_X in $L^2([-\nu_{\text{est}}, \nu_{\text{est}}]^d)$ at almost parametric rate, uniformly over ρ for ρ in the compact set $[1, \rho_0]$ if we set $m_n = \left\lceil \rho_0 \frac{\log(n)}{\log\log(n)} \right\rceil$, where ρ_0 is the *a priori* upper bound on ρ. From now on m_n is set to this value.

To define our estimator of the density of X, we fix a specific polynomial expansion. Let us introduce the truncation operator T_m as follows. If ϕ is a multivariate analytic function defined in a neighborhood of 0 in $\mathbb{C}^D$ written as $\phi : x \mapsto \sum_{(i_1,\ldots,i_d)\in\mathbb{N}^d} c_i \prod_{a=1}^d x_a^{i_a}$, define

$$T_m\phi : x \mapsto \sum_{(i_1,\ldots,i_d)\in\mathbb{N}^d : i_1+\ldots+i_d\le m} c_i \prod_{a=1}^{d} x_a^{i_a}.$$

We finally define the estimator of the density of the signal as follows. Fix some integer $m_{n,\rho}$ and positive real number $h_{n,\rho}$. Then for all $t \in \mathbb{R}^d$,

$$\widehat{f}_{n,\rho}(t) = \frac{1}{(2\pi)^d} \int_{[-h_{n,\rho}, h_{n,\rho}]^d} \exp(-it^\top u)\; T_{m_{n,\rho}} \widehat{\Phi}_{n,\rho}(u) du. \tag{3.1}$$

Note that the function $\widehat{f}_{n,\rho}$ is in general not a probability density function. In particular it is not positive everywhere. It is possible to further project that function onto the convex subset of probability densities, and all theoretical guarantees stay true since projection decreases distance. Note also that the construction of the estimator of the density uses first a polynomial expansion of large enough degree m_n to build an estimator of the characteristic function with small enough risk on $[-\nu, \nu]^d$, then a further truncation to a smaller degree $m_{n,\rho}$ to be able to control the Fourier inversions on a growing set $[-h_{n,\rho}, h_{n,\rho}]^d$.

We prove in Theorem 4.2 that a good choice of $m_{n,\rho}$ and $h_{n,\rho}$ allows to control the integrated quadratic risk over regularity classes of densities, and construct an estimator that is adaptive in ρ in Theorem 4.3. The rates are shown to be minimax optimal for compactly supported ordinary smooth signals in Theorem 4.4.

In the theroretical results that follow, ν_{est} is a fixed non random constant. We propose in Section 6.4 a method to choose all parameters including ν_{est} in a data driven way.

4. Rates of convergence

The first step to control the quadratic risk of the estimated density is to control the quadratic risk of the estimator of the characteristic function over some small set in $\mathbb{R}^d$. The constants in the proposition below depend on the signal through ρ and S, and on the noise through its second moment and the quantity

$$c_\nu = \inf\{|\Phi_{\varepsilon^{(i)}}(t)|,\ t \in [-\nu, \nu]^d,\ i = 1, 2\}, \tag{4.1}$$

provided it is positive, which holds for any noise distribution for small enough ν by continuity of the characteristic function. For any $\nu > 0$, $c(\nu) > 0$, $E > 0$, define $\mathcal{Q}^{(2d)}(\nu, c(\nu), E)$ the set of distributions $\mathbb{Q} = \otimes_{j=1}^{2} \mathbb{Q}_j$ on $\mathbb{R}^{2d}$ such that $c_\nu \geq c(\nu)$ and $\int_{\mathbb{R}^{2d}} \|x\|^2 d\mathbb{Q}(x) \leq E$.

Proposition 4.1. *Assume there exists $a > 0$ such that for all integer n,*

$$m_n \geq a \frac{\log(n)}{\log\log(n)}.$$

Then for all $\nu \in (0, \nu_{est} \wedge 1/(2Sd^2)]$, $\rho_0 \geq 1$, $S, c(\nu), E > 0$ and $\delta \in (0,1)$, there exist positive constants c and n_0 such that the following holds. For any $\rho \in [1, \rho_0]$, for all $\Phi_X \in \Upsilon_{\rho,S}$ and $\mathbb{Q} \in \mathcal{Q}^{(2d)}(\nu, c(\nu), E)$, for all $n \geq n_0$ and $x \geq 1$, with probability at least $1 - e^{-x}$,

$$\forall \rho' \in [\rho, \rho_0], \quad \int_{[-\nu,\nu]^d} |\widehat{\Phi}_{n,\rho'}(t) - \Phi_X(t)|^2 dt \leq c \frac{x}{n^{(1 \wedge \frac{a}{\rho}) - \delta}}.$$

The proof of Proposition 4.1 builds upon intermediate results in [8] and [9] and is detailed in Section 8.2.

4.1. Upper bound

The aim of this section is to give an upper bound of the maximum $L_2(\mathbb{R}^d)$-risk for the estimation of f. We shall denote $\|\cdot\|_2$ the norm in $L_2(\mathbb{R}^d)$. For all $\rho \geq 1$, $\beta > 0$, $S > 0$, $c_\beta > 0$, we denote $\Psi(\rho, S, \beta, c_\beta)$ the set of distributions $\mathbb{P}_X$ with a density f on $\mathbb{R}^d$ such that the characteristic function Φ_X is in $\Upsilon_{S,\rho}$ and satisfies

$$\int_{\mathbb{R}^d} |\Phi_X(u)|^2 (1 + \|u\|^2)^\beta du \leq c_\beta.$$

Also, for all $\rho \geq 1$, $\alpha > 0$, $\gamma > 0$, $S > 0$, $c_{\alpha,\gamma} > 0$, we denote $\Gamma(\rho, S, \alpha, \gamma, c_{\alpha,\gamma})$ the set of distributions $\mathbb{P}_X$ with a density f on $\mathbb{R}^d$ such that the characteristic function Φ_X is in $\Upsilon_{S,\rho}$ and satisfies

$$\int_{\mathbb{R}^d} |\Phi_X(u)|^2 \exp(2\gamma \|u\|^\alpha) du \leq c_\gamma.$$

Theorem 4.2. *For $c_h < (3\sqrt{e})^{-1}$, define for any $\rho \geq 1$*

$$m_{n,\rho} = \left\lfloor \frac{\rho}{4} \frac{\log(n)}{\log\log(n)} \right\rfloor, \quad h_{n,\rho} = c_h \frac{m_{n,\rho}^{1/\rho}}{S}.$$

Then for any $\rho_0 \geq 1$, $\nu \in (0, \nu_{est}]$, $c(\nu) > 0$, $E > 0$, $S > 0$, $\beta > 0$ and $c_\beta > 0$, there exists n_0 and $C > 0$ such that for all $n \geq n_0$,

$$\sup_{\rho \in [1,\rho_0]} \sup_{\substack{\mathbb{P}_X \in \Psi(\rho,S,\beta,c_\beta) \\ \mathbb{Q} \in \mathcal{Q}^{(2d)}(\nu,c(\nu),E)}} \left(\frac{\log(n)}{\log\log(n)} \right)^{\frac{2\beta}{\rho}} \mathbb{E}_{(\mathbb{P}_{r(X)} * \mathbb{Q})^{\otimes n}} [\|\widehat{f}_{n,\rho} - f\|_2^2] \leq C,$$

and for any $\rho_0 \geq 1$*,* $\nu \in (0, \nu_{est}]$*,* $c(\nu) > 0$*,* $E > 0$*,* $S > 0$*,* $\alpha > 0$*,* $\gamma > 0$ *and* $c_{\alpha,\gamma} > 0$*, letting*

$$c_{\exp} = \begin{cases} 2\gamma \left(\frac{c_h}{S}\right)^{\alpha} & \text{if } \alpha < \rho, \\ \left(2\gamma \left(\frac{c_h}{S}\right)^{\alpha}\right) \wedge 1 & \text{if } \alpha = \rho, \\ 1 & \text{if } \alpha > \rho, \end{cases} \tag{4.2}$$

there exists n_0 *and* $C > 0$ *such that for all* $n \geq n_0$*,*

$$\sup_{\rho \in [1,\rho_0]} \sup_{\substack{\mathbb{P}_X \in \Gamma(\rho,S,\alpha,\gamma,c_{\alpha,\gamma}) \\ \mathbb{Q} \in \mathcal{Q}^{(2d)}(\nu,c(\nu),E)}} \mathbb{E}_{(\mathbb{P}_{r(X)} * \mathbb{Q})^{\otimes n}}[\|\widehat{f}_{n,\rho} - f\|_2^2] \times \exp\left\{ c_{\exp} \left(\frac{\log(n)}{\log\log(n)}\right)^{\frac{\alpha}{\rho} \wedge 1} \right\} \leq C.$$

The proof of Theorem 4.2 is detailed in Section 8.6. It is interesting to note that the estimator is adaptive to the class of regularity of the signal.

Let us compare this upper bound with the previous results in the literature. The best rates of convergence without the symmetry assumption on the noise are obtained in [3], who improve upon the results in [17] regarding both the assumptions on the noise and signal distributions and the rates of convergence. Though the earlier work [22] has weaker assumptions than [3], the author only proves consistency of his estimator but does not provide rates of convergence.

The rates in [3] depend on the tail of the characteristic function of the signal and on the tail of the characteristic function of the noise, as is usual in the deconvolution literature. The authors prove polynomial rates of convergence when the noise is ordinary smooth (that is with a characteristic function decreasing as a power function), up to a power of $\log(n)$ factor when the signal is not ordinary smooth but supersmooth. For supersmooth errors the rate is a power of $\log(n)$. Our assumptions are not an exact generalization of the ones in [3], as the authors of [3] do not need assumptions on the tails of the distribution of the signal, and instead assume that both the characteristic function of the signal and the characteristic function of the noise do not vanish anywhere and have some controlled behaviour near infinity. In contrast, the class upon which our upper bound applies includes both ordinary smooth and supersmooth noise distributions, and allows the characteristic functions to vanish, even on open sets. We propose in Section 5 a method to choose between our estimator and the one proposed in [3] to get the best of both worlds.

Although we were not able to prove it, we believe it should be possible to adapt our estimation procedure to the tails of the characteristic functions by selecting the parameter ν_{est} based on the observations to obtain improved rates on the classes where the characteristic functions do not vanish and are ordinary smooth. Our simulations indicate that this idea seems relevant, see Section 6. We discuss this point in Section 7.

4.2. Adaptivity in ρ

The construction of the estimator above assumes the tail parameter ρ known. Unfortunately, this tail parameter is typically unknown in practice. We now propose a data-driven model selection procedure to choose ρ and we prove that the resulting estimator has a rate of convergence corresponding to the smallest ρ such that $\Phi_X \in \Upsilon_{\rho,S}$ for some $S > 0$. Our strategy is based on Goldenshluger and Lepski's methodology [11]. As usual, the idea is to perform a bias-variance trade off. Although we have an upper bound for the variance term, the bias is not easily accessible. The variance bound in the ordinary smooth case is

$$\sigma_n^{OS}(\rho,\beta) = c_\sigma \left(\frac{\log\log(n)}{\log(n)} \right)^{\frac{\beta}{\rho}}$$

and in the supersmooth case

$$\sigma_n^{SS}(\rho,\gamma,\alpha) = c_\sigma \exp\left\{ -\frac{c_{\exp}}{2} \left(\frac{\log(n)}{\log\log(n)} \right)^{\frac{\alpha}{\rho}\wedge 1} \right\}$$

for all $\rho \in [1,\rho_0]$ and for some constant $c_\sigma > 0$, where $c_{\exp}$ is defined in (4.2).

In the ordinary smooth case, for each $\beta > 0$ and $\rho \in [1,\rho_0]$, the proxy of the bias is defined as

$$A_n(\rho,\beta) = 0 \vee \sup_{\rho' \in [\rho,\rho_0]} \left\{ \|\widehat{f}_{n,\rho'} - \widehat{f}_{n,\rho}\|_2 - \sigma_n^{OS}(\rho',\beta) \right\}.$$

Define the estimator of ρ for $\beta > 0$ by

$$\widehat{\rho}_\beta \in \arg\min\{A_n(\rho,\beta) + \sigma_n^{OS}(\rho,\beta),\ \rho \in [1,\rho_0]\}$$

and let $\widetilde{f}_\beta = \widehat{f}_{n,\widehat{\rho}_\beta}$. Then, for each $\beta > 0$, let

$$B_n(\beta) = 0 \vee \sup_{\beta' \in (0,\beta]} \left\{ \|\widetilde{f}_{\beta'} - \widetilde{f}_\beta\|_2 - 4\sigma_n^{OS}(\widehat{\rho}_{\beta'},\beta') \right\}$$

and select

$$\widehat{\beta} \in \arg\min\{B_n(\beta) + 4\sigma_n^{OS}(\widehat{\rho}_\beta,\beta),\ \beta > 0\}.$$

Finally, let $\widehat{f}_n^{OS} = \widetilde{f}_{\widehat{\beta}} = \widehat{f}_{n,\widehat{\rho}_{\widehat{\beta}}}$.

In the supersmooth case, for each $\gamma,\alpha > 0$ and $\rho \in [1,\rho_0]$, the proxy of the bias is defined as

$$A_n(\rho,\gamma,\alpha) = 0 \vee \sup_{\rho' \in [\rho,\rho_0]} \left\{ \|\widehat{f}_{n,\rho'} - \widehat{f}_{n,\rho}\|_2 - \sigma_n^{SS}(\rho',\gamma,\alpha) \right\}.$$

Define the estimator of ρ for $\gamma,\alpha > 0$ by

$$\widehat{\rho}_{\gamma,\alpha} \in \arg\min\{A_n(\rho,\gamma,\alpha) + \sigma_n^{SS}(\rho,\gamma,\alpha),\ \rho \in [1,\rho_0]\}$$

and let $\widetilde{f}_{\gamma,\alpha} = \widehat{f}_{n,\widehat{\rho}_{\gamma,\alpha}}$. Then, for each $\gamma, \alpha > 0$, let

$$B_n(\gamma,\alpha) = 0 \vee \sup_{\gamma' \in (0,\gamma]} \left\{ \|\widetilde{f}_{\gamma',\alpha} - \widetilde{f}_{\gamma,\alpha}\|_2 - 4\sigma_n^{SS}(\widehat{\rho}_{\gamma',\alpha}, \gamma', \alpha) \right\},$$

select

$$\widehat{\gamma}_\alpha \in \arg\min\{B_n(\gamma,\alpha) + 4\sigma_n^{SS}(\widehat{\rho}_{\gamma,\alpha}, \gamma, \alpha),\ \gamma > 0\}.$$

and let $\overline{f}_\alpha = \widetilde{f}_{\widehat{\gamma}_\alpha,\alpha}$. Then, for each $\alpha > 0$, let

$$C_n(\alpha) = 0 \vee \sup_{\alpha' \in (0,\alpha]} \{ \|\overline{f}_{\alpha'} - \overline{f}_\alpha\|_2 - 24\sigma_n^{SS}(\widehat{\rho}_{\widehat{\gamma}_{\alpha'},\alpha'}, \widehat{\gamma}_{\alpha'}, \alpha') \}$$

and select

$$\widehat{\alpha} \in \arg\min\{C_n(\alpha) + 24\sigma_n^{SS}(\widehat{\rho}_{\widehat{\gamma}_\alpha,\alpha}, \widehat{\gamma}_\alpha, \alpha),\ \alpha > 0\}.$$

Finally, let $\widehat{f}_n^{SS} = \overline{f}_{\widehat{\alpha}} = \widehat{f}_{n,\widehat{\rho}_{\widehat{\gamma}_{\widehat{\alpha}},\widehat{\alpha}}}$.

The following theorem states that these estimators are rate adaptive.

Theorem 4.3. *For any $\rho_0 \geq 1$, $\nu \in (0, \nu_{est}]$, $c(\nu) > 0$, $E > 0$, $S > 0$, $\beta > 0$ and $c_\beta > 0$, there exists $c_\sigma > 0$ and $C > 0$ such that*

$$\limsup_{n\to+\infty} \sup_{\rho \in [1,\rho_0]} \sup_{\substack{\mathbb{P}_X \in \Psi(\rho,S,\beta,c_\beta) \\ \mathbb{Q} \in \mathcal{Q}^{(2d)}(\nu,c(\nu),E)}} \left(\frac{\log(n)}{\log\log(n)} \right)^{\frac{2\beta}{\rho}} \mathbb{E}_{(\mathbb{P}_{r(X)} * \mathbb{Q})^{\otimes n}} [\|\widehat{f}_n^{OS} - f\|_2^2] \leq C.$$

For any $\rho_0 \geq 1$, $\nu \in (0, \nu_{est}]$, $c(\nu) > 0$, $E > 0$, $S > 0$, $\alpha > 0$, $\gamma > 0$ and $c_{\alpha,\gamma} > 0$, taking $c_{\exp}$ as in (4.2), there exists $c_\sigma > 0$ and $C > 0$ such that

$$\limsup_{n\to+\infty} \sup_{\rho \in [1,\rho_0]} \sup_{\substack{\mathbb{P}_X \in \Gamma(\rho,S,\alpha,\gamma,c_{\alpha,\gamma}) \\ \mathbb{Q} \in \mathcal{Q}^{(2d)}(\nu,c(\nu),E)}} \exp\left(c_{\exp} \left(\frac{\log(n)}{\log\log(n)} \right)^{\frac{\alpha}{\rho} \wedge 1} \right) \times \mathbb{E}_{(\mathbb{P}_{r(X)} * \mathbb{Q})^{\otimes n}} [\|\widehat{f}_n^{SS} - f\|_2^2] \leq C.$$

The proof of Theorem 4.3 is detailed in Section 8.7.

4.3. Lower bound

In this section, we provide a lower bound of the minimax risk in the case $\rho = 1$, that is for compactly supported signals. The lower bound in Theorem 4.4 matches the rate given in Theorem 4.2 for our estimator, proving that our estimator, together with its adaptive version, achieves the minimax adaptive rate in this case.

Theorem 4.4. *For all $S > 0$, $\beta \geq 1/2$, $c_\beta > 0$, and $\nu > 0$, there exists a constant $c > 0$ such that*

$$\inf_{\widehat{f}} \sup_{\substack{\mathbb{P}_X \in \Psi(1,S,\beta,c_\beta) \\ \mathbb{Q} \in \mathcal{Q}^{(2d)}(\nu,c(\nu),E)}} \mathbb{E}_{(\mathbb{P}_{r(X)} * \mathbb{Q})^{\otimes n}} [\|\widehat{f} - f\|_2^2] \geq c \left(\frac{\log \log(n)}{\log(n)} \right)^{2\beta},$$

where the infimum is taken over all the estimators $\widehat{f}$, that is all measurable functions of $Y_1, \ldots, Y_n$.

The proof of Theorem 4.4 is detailed in Section 8.9. It is based on Le Cam's method, also known as the two-points method (see for instance [16]), one of the most widespread techniques to derive lower bounds, and adapts ideas from [19] and [8] to the repeated measurements setting.

5. Adaptativity to unknown noise regularity

In this section, we propose a method to choose between our estimator and the ones developed in [3]. We thus consider the same setting as [3] and assume that $d = 1$, that is, observations are in one dimension. We also assume that both coordinates of the noise have the same distribution, with density f_ε.

The method used is a special case of the following, more general methodology. Given two estimators $\hat{f}_1$ and $\hat{f}_2$ of a parameter f and two bounded metric sets $\mathcal{F}_1 \subset \mathcal{F}_2$ endowed with the distance d such that the estimators satisfy

$$\begin{cases} \sup_{f \in \mathcal{F}_1} \mathbb{E}_f[d(\hat{f}_1, f)^2] \leq R_1, \\ \sup_{f \in \mathcal{F}_2} \mathbb{P}_f(d(\hat{f}_2, f)^2 \geq R_2) \leq c_2 R_1 \end{cases} \tag{5.1}$$

for some $c_2 > 0$ and with $R_1 \leq R_2$, it is possible to construct an estimator $\hat{f}$ that combines the best of both estimators with no *a priori* knowledge on f, that is, for some constant $c > 0$,

$$\begin{cases} \sup_{f \in \mathcal{F}_1} \mathbb{E}_f[d(\hat{f}, f)^2] \leq cR_1, \\ \sup_{f \in \mathcal{F}_2} \mathbb{E}_f[d(\hat{f}, f)^2] \leq cR_2. \end{cases} \tag{5.2}$$

The idea is to use $\hat{f}_1$ to get improved convergence rates when in $\mathcal{F}_1$, while keeping $\hat{f}_2$ as a safeguard in the general case where $\hat{f}_1$ may produce aberrant results. Its construction is as follows: if both estimators are close (distance of the order of the largest of the error upper bounds), take the estimator with the smallest domain of validity, otherwise, take the more general one:

$$\hat{f} = \begin{cases} \hat{f}_1 & \text{if } d(\hat{f}_1, \hat{f}_2)^2 \leq 4R_2, \\ \hat{f}_2 & \text{otherwise.} \end{cases} \tag{5.3}$$

Let us prove that this estimator satisfies (5.2).

General f. Assume $f \in \mathcal{F}_2$. Let $C_{\mathcal{F}} > 0$ be the diameter of $\mathcal{F}_2$. The error of $\hat{f}$ is

$$\begin{aligned} d(\hat{f}, f)^2 &= d(\hat{f}_1, f)^2 \mathbf{1}_{\hat{f}=\hat{f}_1} + d(\hat{f}_2, f)^2 \mathbf{1}_{\hat{f}=\hat{f}_2} \\ &\leq 2(d(\hat{f}_1, \hat{f}_2)^2 + d(\hat{f}_2, f)^2)\mathbf{1}_{\hat{f}=\hat{f}_1} + d(\hat{f}_2, f)^2 \mathbf{1}_{\hat{f}=\hat{f}_2} \\ &\leq 8R_2 + 2d(\hat{f}_2, f)^2 \end{aligned}$$

and by (5.1), $\mathbb{E}[d(\hat{f}_2, f)^2] \leq c_2 C_{\mathcal{F}} R_1 + R_2 \leq (1 + c_2 C_{\mathcal{F}}) R_2$, hence the second line of (5.2).

Specific f. Assume $f \in \mathcal{F}_1$, then likewise

$$\begin{aligned} d(\hat{f}, f)^2 &= d(\hat{f}_1, f)^2 \mathbf{1}_{\hat{f}=\hat{f}_1} + d(\hat{f}_2, f)^2 \mathbf{1}_{\hat{f}=\hat{f}_2} \\ &\leq d(\hat{f}_1, f)^2 + d(\hat{f}_2, f)^2 \mathbf{1}_{d(\hat{f}_2, \hat{f}_1)^2 \geq 4R_2}, \end{aligned}$$

and note that

$$\mathbf{1}_{d(\hat{f}_2, \hat{f}_1)^2 \geq 4R_2} \leq \mathbf{1}_{d(\hat{f}_2, f) \geq R_2} + \mathbf{1}_{d(\hat{f}_1, f) \geq d(\hat{f}_2, f)},$$

so that

$$\begin{aligned} d(\hat{f}, f)^2 &\leq 2d(\hat{f}_1, f)^2 + d(\hat{f}_2, f)^2 \mathbf{1}_{d(\hat{f}_2, f)^2 \geq R_2} \\ &\leq 2d(\hat{f}_1, f)^2 + C_{\mathcal{F}} \mathbf{1}_{d(\hat{f}_2, f)^2 \geq R_2} \end{aligned}$$

and therefore by (5.1), $\mathbb{E}[d(\hat{f}, f)^2] \leq 2R_1 + C_{\mathcal{F}} c_2 R_1 \leq (2 + c_2 C_{\mathcal{F}}) R_1$, hence the first line of (5.2).

For positive constants C_1, C_2, c, β and η, denote by $\mathcal{F}^u(C_1, C_2, c, \beta, \eta)$ the class of square integrable probability densities f whose characteristic function Φ satisfies

$$\forall u, v \in \mathbb{R}^+, \quad u \geq v \Rightarrow |\Phi(u)| \leq C_2 |\Phi(v)| \tag{5.4}$$

and

$$\forall u \in \mathbb{R}, \ |\Phi(u)| \leq \left(1 + C_1 |u|^2\right)^{-\beta} e^{-c|u|^\eta}.$$

Write $\mathcal{F}^l(C_1, C_2, c, \beta, \eta)$ the class of square integrable probability densities for which the condition (5.4) holds and

$$\forall u \in \mathbb{R}, \ |\Phi(u)| \geq \left(1 + C_1 |u|^2\right)^{-\beta} e^{-c|u|^\eta}.$$

In the following, we call a distribution–or a random variable–upper ordinary smooth if its density belongs to $\mathcal{F}^u(C_1, C_2, 0, \beta, \eta)$ for $\beta > 1/2$, upper supersmooth if it belongs to $\mathcal{F}^u(C_1, C_2, c, \beta, \eta)$ for $c > 0$, and likewise for lower ordinary / supersmooth with $\mathcal{F}^l$.

Note that these definitions of upper smoothness are closely related to the ones used in Section 4.1. In particular, an upper ordinary smooth distribution in $\Upsilon_{\rho,S}$ will belong to a $\Psi(\rho, S, \beta', c_\beta)$ for any $\beta' < \beta - 1/2$ (since $d = 1$) and some c_β

that depend on the smoothness parameters, and likewise an upper supersmooth distribution in $\Upsilon_{\rho,S}$ will belong to a $\Gamma(\rho, S, \alpha, \gamma, c_{\alpha,\gamma})$.

For $C_3 > 0$, denote by $\mathcal{F}(C_3, p)$ the class of pairs (f_X, f_ε) of square integrable probability densities for which the following conditions are met:

$$\left(\|\Phi_X''\Phi_\varepsilon + \mathbb{E}[\varepsilon^2]\Phi_X\Phi_\varepsilon\|_1 + \|\Phi_X'\Phi_\varepsilon\|_2^2\right)^p + \mathbb{E}[|X+\varepsilon|^{2p}] \leq C_3,$$

and $(\log(\Phi_X\Phi_\varepsilon)')$ is square integrable with

$$\|\log(\Phi_X\Phi_\varepsilon)'\|_2^{2p} \leq C_3.$$

Let us summarize the results of [3] concerning the estimation of the signal density with lightened formulas: they construct an estimator $\hat{f}^{CK}$ of the density of the signal such that, assuming the parameters are in $\mathcal{F}(C_3, p)$, for the oracle choice of their bandwidth parameter,

- for upper ordinary smooth signals and lower ordinary smooth errors (**Case I** in their article), $\mathbb{E}[\|\hat{f}^{CK} - f\|_2^2] = O(n^{-\zeta})$ for some $\zeta > 0$,
- for upper ordinary smooth signals and lower supersmooth errors (**Case II**), $\mathbb{E}[\|\hat{f}^{CK} - f\|_2^2] = O(\log(n)^{-\zeta})$ for some $\zeta > 0$,
- for upper supersmooth signals and lower ordinary smooth errors (**Case III**), $\mathbb{E}[\|\hat{f}^{CK} - f\|_2^2] = O(n^{-p/(p+1)}\log(n)^{\zeta})$ for some ζ,

all of this uniformly over classes of (f, f_ε) and with an explicit formula for ζ that depends on the regularity of the noise in addition to the regularity of the signal. We will use this as the first line of (5.1).

On the other hand, by Proposition 8.4, the estimator $\widehat{f}_{n,\rho}$ defined in Section 3 satisfies the second line of (5.1) with the risk

$$R_2 = \begin{cases} c_\sigma \left(\frac{\log(n)}{\log\log(n)}\right)^{\frac{-2\beta}{\rho}} & \text{in the ordinary smooth case,} \\ c_\sigma \exp\left(-c_{\exp}\left(\frac{\log(n)}{\log\log(n)}\right)^{\frac{\alpha}{\rho}\wedge 1}\right) & \text{in the supersmooth case.} \end{cases} \quad (5.5)$$

Let $\widehat{f}^{A,\mathrm{I}}$ be the estimator defined as in (5.3) from $\hat{f}_1 = \hat{f}^{\mathrm{CK}}$ (the estimator from **Case I**) and $\hat{f}_2 = \widehat{f}_{n,\rho}$, with R_2 as above. Likewise, define $\widehat{f}^{A,\mathrm{II}}$ and $\widehat{f}^{A,\mathrm{III}}$ for **Case II** and **Case III** respectively. The resulting estimator has improved rates of convergence compared to ours when the densities do belong to the presumed **Case**, while still having a controlled error even if they do not.

Theorem 5.1. *Assume that $(f, f_\varepsilon) \in \mathcal{F}(C_3, p)$ for some $C_3, p > 0$. If the signal is upper ordinary smooth and the noise is lower ordinary smooth (**Case I**), there exists $\zeta > 0$ such that*

$$\mathbb{E}[\|\widehat{f}^{A,I} - f\|_2^2] = O(n^{-\zeta}),$$

and in all cases, with the notations of Section 4.1,

$$\mathbb{E}[\|\widehat{f}^{A,I} - f\|_2^2] = \begin{cases} O\left(\left(\frac{\log(n)}{\log\log(n)}\right)^{\frac{-2\beta}{\rho}}\right) & \textit{if } f \textit{ ordinary smooth,} \\ O\left(\exp\left(-c_{\exp}\left(\frac{\log(n)}{\log\log(n)}\right)^{\frac{\alpha}{\rho}\wedge 1}\right)\right) & \textit{if } f \textit{ supersmooth.} \end{cases}$$

The same general case holds for $\widehat{f}^{A,II}$ and $\widehat{f}^{A,III}$, and for their particular cases:

- *for upper ordinary smooth signals and lower supersmooth errors (**Case II**),* $\mathbb{E}[\|\widehat{f}^{A,II} - f\|_2^2] = O(\log(n)^{-\zeta})$ *for some* $\zeta > 0$*,*
- *for upper supersmooth signals and lower ordinary smooth errors (**Case III**),* $\mathbb{E}[\|\widehat{f}^{A,III} - f\|_2^2] = O(n^{-p/(p+1)}\log(n)^{\zeta})$ *for some* ζ*.*

Note that we do not have an explicit formula for the constant c_σ in the definition of R_2 (5.5). To use the methodology in practice would require a method to choose c_σ based on the observations, or alternatively one based on the hold-out approach discussed in Section 6.4.

6. Simulations

The aim of this section is to assess the performance of our method on synthetic datasets.

Although theoretical values for the parameters m and h are given in Theorem 4.2, in practice, other values may produce much better results. As such, a practical implementation of our method consists of two steps: constructing an estimator for several possible parameters (m, ν_{est}, h), and then selecting the "best" parameters $(\hat{m}, \hat{\nu}_{\text{est}}, \hat{h})$, in the hope that the resulting estimator performs comparably to the best parameter. Data-driven selection of the parameters (m, ν_{est}, h) is discussed in Section 6.4. For the rest of this section, we select the parameters (m, ν_{est}, h) ourselves, to get an idea of how the method performs when the best, or at least good, parameters are selected.

In Section 6.1, we explain how the estimator of the characteristic function and the target density of the signal are computed. We then introduce in Section 6.2 the four synthetic datasets covering several smoothness scenarios on which our method is assessed and show the graphs of the resulting estimators. In Section 6.3, we compare the empirical risk of our estimator with the empirical risk of the estimator defined in [3]. Finally, we discuss how to select the parameters (m, ν_{est}, h) in Section 6.4.

All simulation codes are available in Python at `https://www.normalesup.org/~llehericy/repeated_simfiles`.

6.1. Procedure

We consider real-valued signals, that is $d = 1$. For each dataset, we may consider several set of parameters (m, ν_{est}, h). For each of them, the estimators are computed as follows.

For any integer $m > 0$, $\phi \in \Upsilon_{\rho,S}$ and $t \in \mathbb{C}$, $T_m\phi(t) = 1 + \sum_{k=2}^m \phi_k (it)^k$, where the coefficients satisfy $\phi_k \in \mathbb{R}$, due to the fact that for all $\phi \in \Upsilon_{\rho,S}$ and $t \in \mathbb{R}$, $\overline{\phi(t)} = \phi(-t)$. Recall that the absence of degree one term is due to the assumption that the signal is centered.

Given grids of values of m and ν_{est}, we compute in increasing order for m and ν_{est} an estimator $\widehat{\Phi}$ for each (m, ν_{est}). Then, for each (m, ν_{est}) used and each h in a grid of values, we compute the corresponding estimated density $\widehat{f}$ of X. In our simulations, $m \in \{4, 8, 12, 16, 20, 25, 30\}$, $\nu_{\text{est}} \in \{1.5, 2.5, 4, 5.5, 7\}$ and $h \in \{1.5, 2.5, 4, 5.5, 7, 8\}$.

The computation of $\widehat{\Phi}$ is done as follows. For any polynomial $T_m\phi$ with coefficients as above, the integral $M_n(T_m\phi)$ is approximated by a Riemann sum over a regular grid with 101×101 points covering $[-\nu_{\text{est}}, \nu_{\text{est}}]^2$. We minimize $M_n(T_m\phi)$ as a function of the coefficients $(\phi_k)_{2\leq k\leq m}$ with the function *optimize.minimize* from the package *scipy* (version 1.13.1) of Python (version 3.12.7), using the Newton-CG method. We provide it with numerical estimates of the gradient and hessian of M_n and otherwise default parameters, other than the initial point which we detail later. The estimates of the derivatives of M_n are obtained through approximating the following integrals by Riemann sums over the same grid as M_n: when writing $P_k(t) = (it)^k$, $\phi^{(1)}(t_1) = \phi(t_1, 0)$, $\phi^{(2)}(t_2) = \phi(0, t_2)$ and likewise for $\widetilde{\phi}^{(1)}$ and $\widetilde{\phi}^{(2)}$, it holds

$$\frac{\partial M_n(\phi)}{\partial \phi_k} = 2\int_{[-\nu,\nu]^2} \text{Re}\Bigg(\overline{\left(\widetilde{\phi}\phi^{(1)}\phi^{(2)} - \phi(t_1+t_2)\widetilde{\phi}^{(1)}\widetilde{\phi}^{(2)}\right)} \times \left(\widetilde{\phi}P_k(t_1)\phi^{(2)} + \widetilde{\phi}\phi^{(1)}P_k(t_2) - P_k(t_1+t_2)\widetilde{\phi}^{(1)}\widetilde{\phi}^{(2)}\right)\Bigg) dt_1 dt_2$$

and

$$\begin{aligned}\frac{\partial^2 M_n(\phi)}{\partial \phi_k \partial \phi_\ell} = 2\int_{[-\nu,\nu]^2} \text{Re}\Bigg(&\overline{\left(\widetilde{\phi}P_k(t_1)\phi^{(2)} + \widetilde{\phi}\phi^{(1)}P_k(t_2) - P_k(t_1+t_2)\widetilde{\phi}^{(1)}\widetilde{\phi}^{(2)}\right)} \\ &\times \left(\widetilde{\phi}P_\ell(t_1)\phi^{(2)} + \widetilde{\phi}\phi^{(1)}P_\ell(t_2) - P_\ell(t_1+t_2)\widetilde{\phi}^{(1)}\widetilde{\phi}^{(2)}\right) \\ &+ \overline{\left(\widetilde{\phi}\phi^{(1)}\phi^{(2)} - \phi(t_1+t_2)\widetilde{\phi}^{(1)}\widetilde{\phi}^{(2)}\right)} \\ &\times \widetilde{\phi}\Big(P_k(t_1)P_\ell(t_2) + P_\ell(t_1)P_k(t_2)\Big)\Bigg) dt_1 dt_2.\end{aligned}$$

Given the well known issue of potentially finding a non-global minimum of M_n, we run the approximate minimization algorithm with several initial points ϕ:

- $\phi_k = (T_m\Phi_X)_k(1 + 0.05U_k)$ with i.i.d. $U_k \sim \mathcal{U}([-1, 1])$, a perturbation of the truncation of the Taylor expansion of the true signal distribution. Note that this Taylor expansion does not converge when $X + 2 \sim \Gamma(4, 2)$, in which case this initialization is ignored,
- $\phi_k = (P_{m,\nu_{\text{est}}}\Phi_X)_k(1 + 0.05U_k)$ with i.i.d. $U_k \sim \mathcal{U}([-1, 1])$, a perturbation of the orthogonal projection of Φ_X in $L^2([-\nu_{\text{est}}, \nu_{\text{est}}])$ on the set of ϕ considered, i.e. of polynomials of degree at most m with degree zero coefficient equal to one and no degree one coefficients such that $\overline{\phi(t)} = \phi(-t)$,

- using the previous value m^{prev} of m with corresponding estimator $\widehat{\phi}^{\text{prev } m}$ (or $m^{\text{prev}} = 0$ if m is the smallest one in the grid of possible m), take $\phi_k = \widehat{\phi}_k^{\text{prev } m}$ for $k \leq m^{\text{prev}}$ and $\phi_k = (P_{m,\nu_{\text{est}}}\Phi_X)_k(1 + 0.05U_k)$ a perturbation of the coefficients of the projection of Φ_X otherwise,
- using the previous value m^{prev} of m with corresponding estimator $\widehat{\phi}^{\text{prev } m}$, take $\phi_k = \widehat{\phi}_k^{\text{prev } m}$ for $k \leq m^{\text{prev}}$ and $\phi_k = 0$ otherwise,
- using the previous value $\nu_{\text{est}}^{\text{prev}}$ of ν_{est} with corresponding estimator $\widehat{\phi}^{\text{prev } \nu}$, take $\phi = \widehat{\phi}^{\text{prev } \nu}$.

After minimizing M_n for each initial point, we keep the one with the lowest M_n. Choosing initializations that rely on $T_m\Phi_X$ or $P_{m,\nu_{\text{est}}}\Phi_X$ is not doable in practical scenarios (where only the last two initializations are usable), but for a proof of concept of our method, it should reduce the likelihood of getting stuck in suboptimal minima and thus help properly assess the best-case performance of our method.

Finally, the estimator $\widehat{f}_{(m,\nu_{\text{est}},h)} = \max(0, \text{Re}(\widehat{f}_n))$ is computed, where $\widehat{f}_n$ is computed as in (3.1) (where the integral is approximated by a Riemann sum over a regular grid with 1000 points) over a regular grid of 10000 points covering the interval $[-15, 15]$. This estimate will be used to compute the L_2 loss $\|\widehat{f}_{(m,\nu_{\text{est}},h)} - f\|_2^2$ in Section 6.3 through a Riemann sum on the aforementioned regular grid on $[-15, 15]$.

Note that taking $\max(0, \text{Re}(\widehat{f}_n))$ as our estimator instead of $\widehat{f}_n$ has negligible computational cost and can only improve the quadratic loss. Indeed, the loss $\|\max(0, \text{Re}(\widehat{f}_n)) - f\|_2^2$ is upper bounded by the loss $\|\widehat{f}_n - f\|_2^2$ since $\max(0, \text{Re}(\widehat{f}_n))$ is the projection of $\widehat{f}_n$ on the convex set of real, nonnegative functions. The projection of $\widehat{f}_n$ on the set of probability densities, which is also convex, would reduce the quadratic loss even further, but computing it is more involved and the result depends on the tails of $\widehat{f}_n$.

6.2. Synthetic datasets

Write $\Gamma(\alpha, \beta)$ the gamma distribution with shape α and rate β (such that its density is $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbf{1}_{(0,\infty)}(x)$). We also denote $b\Gamma(\alpha, \beta, \gamma, \delta)$ the bilateral Gamma distribution with parameters $\alpha, \beta, \gamma, \delta$ which is the distribution of $U - V$ where U and V are independent random variables such that U has distribution $\Gamma(\alpha, \beta)$ and V has distribution $\Gamma(\gamma, \delta)$. We use the same target densities and errors than [3], up to centering, that is:

`GammaBiGamma`: $X + 2 \sim \Gamma(4, 2)$ and $\varepsilon^{(i)} \sim b\Gamma(2, 2, 3, 3)$ for $i = 1, 2$ (the location shift ensures that $\mathbb{E}[X] = 0$),

`BiGammaGamma`: $X \sim b\Gamma(1, 1, 2, 2)$ and $\varepsilon^{(i)} + 2 \sim \Gamma(4, 2)$ for $i = 1, 2$ (the location shift ensures that $\mathbb{E}[\varepsilon] = 0$),

`GaussianBiGamma`: $X \sim \mathcal{N}(0, 1)$ and $\varepsilon^{(i)} \sim b\Gamma(2, 2, 3, 3)$ for $i = 1, 2$,

`GaussianMixtureGaussian`: $X \sim \mathcal{N}(0, 1)$ and $\varepsilon^{(i)}$ is the mixture of two normal distributions with parameters $(-2, 1)$ and $(2, 2)$ and equal weights $1/2$ for $i = 1, 2$. We use the notation $\varepsilon^{(i)} \sim m\mathcal{N}(-2, 1, 2, 2)$.

These four scenarios cover three of the possible cases for the smoothness of the signal and noise distributions: the first and second ones have ordinary smooth signal and noise (**Case I**), the third one supersmooth signal and ordinary smooth noise (**Case III**), and the last one supersmooth signal and noise.

For each scenario, we generate three sets of 100 i.i.d. samples $(Y_t)_{1\leq t\leq n}$: 100 of size $n = 100$, 100 of size of $n = 1000$ and 100 of size $n = 10000$. For each n, each sample of size n (indexed by $1 \leq i \leq 100$) and each value of m, ν_{est} and h in the grids introduced in Section 6.1, we compute an estimator $\widehat{f}^{i,n}_{(m,\nu_{\text{est}},h)}$ and compute its quadratic loss $E_n(i, m, \nu_{\text{est}}, h) = \|\widehat{f}^{i,n}_{(m,\nu_{\text{est}},h)} - f\|_2^2$ as described in Section 6.1.

Figures 1, 2, 3 and 4 show the estimators in each scenario for the first sample of size $n = 100$ ($i = 1$) and for (m, ν_{est}, h) chosen manually as an approximate minimizer of the median and/or mean value of $\{E_n(i, m, \nu_{\text{est}}, h), 1 \leq i \leq 100\}$, as well as the corresponding estimator of Φ_X. Even with a small amount of data ($n = 100$), our estimator manages to recover the density of the signal accurately, provided that m, ν_{est} and h are chosen properly.

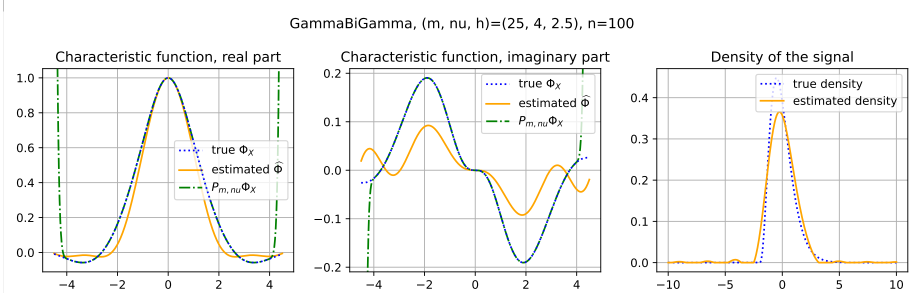

Fig 1: $X + 2 \sim \Gamma(4, 2)$, for $i \in \{1, 2\}$, $\varepsilon^{(i)} \sim b\Gamma(2, 2, 3, 3)$.

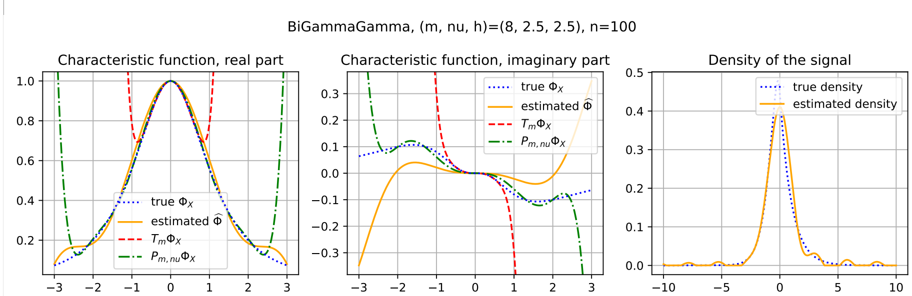

Fig 2: $X \sim b\Gamma(1, 1, 2, 2)$, for $i \in \{1, 2\}$, $\varepsilon^{(i)} + 2 \sim \Gamma(4, 2)$.

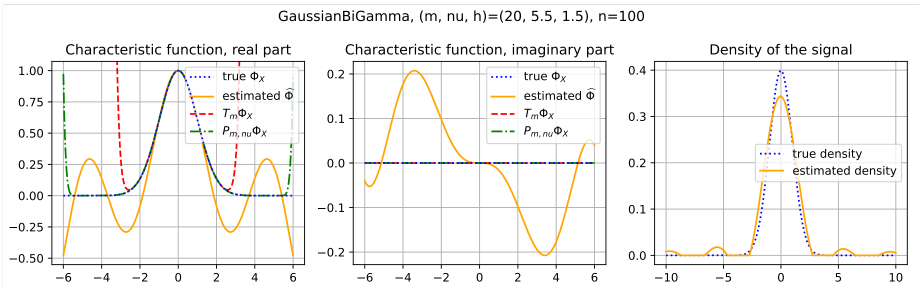

Fig 3: $X \sim \mathcal{N}(0,1)$, for $i \in \{1,2\}$, $\varepsilon^{(i)} \sim b\Gamma(2,2,3,3)$.

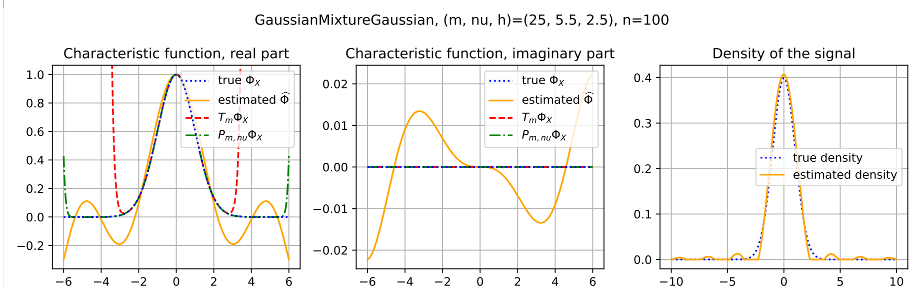

Fig 4: $X \sim \mathcal{N}(0,1)$, for $i \in \{1,2\}$, $\varepsilon^{(i)} \sim m\mathcal{N}(-2,1,2,2)$.

6.3. Comparison with the estimator of [3]

In the following, we compare our estimator with the density estimator presented in [3]. To do so, we compute the empirical risk

$$\widehat{r}_n(m,\nu_{\text{est}},h) = \frac{1}{100}\sum_{i=1}^{100} \|\widehat{f}^{i,n}_{(m,\nu_{\text{est}},h)} - f\|_2^2 = \frac{1}{100}\sum_{i=1}^{100} E_n(i,m,\nu_{\text{est}},h). \tag{6.1}$$

Table 1 shows the comparison to the empirical risk of the oracle estimator of [3], denoted $\widehat{r}_{CK}$ in what follows, reported in their Table 4 (column $\widehat{r}^{\text{or}}$).

Computing our estimator requires choosing the parameters (m,ν_{est},h). This is done by picking a minimizer of $\widehat{r}_n(m,\nu_{\text{est}},h)$. We call this value the oracle value of (m,ν_{est},h). A data-driven selection of these parameters is presented in the next section.

Our procedure outperforms the estimator of [3] for the Gaussian signals. For the $\Gamma(4,2)$ or $b\Gamma(1,1,2,2)$ signals, even though our theory does not apply because these distributions have only sub-exponential tails, it should be noted that our procedure still gives sensible results and performs comparably to the estimator of [3].

n	$100\widehat{r}$	$100\widehat{r}_{CK}^{\mathrm{or}}$
$X \sim \Gamma(4,2)-2,\ \varepsilon \sim b\Gamma(2,2,3,3)$		
100	**1.32**	1.51
1000	0.43	**0.34**
10,000	0.18	**0.15**
$X \sim b\Gamma(1,1,2,2),\ \varepsilon \sim \Gamma(4,2)-2$		
100	1.45	**1.35**
1000	0.61	**0.27**
10,000	0.44	**0.13**
$X \sim \mathcal{N}(0,1),\ \varepsilon \sim b\Gamma(2,2,3,3)$		
100	**0.66**	1.04
1000	**0.12**	0.19
10,000	**0.02**	0.07
$X \sim \mathcal{N}(0,1),\ \varepsilon \sim m\mathcal{N}(-2,1,2,2)$		
100	**0.50**	3.10
1000	**0.54**	1.18
10,000	**0.09**	0.40

TABLE 1

Comparison of the empirical risks of our oracle estimator $\widehat{r}$, defined as the minimum of the empirical risk $\widehat{r}_n$ defined in (6.1), and the risk $\widehat{r}_{CK}^{or}$ of the oracle estimator of [3].

6.4. Data-driven selection of the parameters (m, ν_{est}, h)

Theorem 4.2 states that m has to be chosen of the order of $\log(n)/\log\log(n)$ when n is large enough, with a constant depending on ρ. In practice, and especially for small values of n, such a choice might not produce the best results. For instance, for $n = 1000$, Theorem 4.2 gives $m \simeq \rho$, but the algorithm works better when we significantly increase the value of m. Likewise, while the theoretical result suggests using $h_{n,\rho} \gg \nu_{\mathrm{est}}$, in practice, that is not always the best choice.

Thus, there is a need for a data driven choice of $(m, \nu_{\mathrm{est}}, h)$. Here is a proposition for a hold-out estimator. First, partition the set of observations in three blocks with indices respectively in E_1, E_2 and T. For each set of parameters (m, ν_{est}), compute an estimator $\widehat{\Phi}^X_{(m,\nu_{\mathrm{est}})}$ of the Fourier transform of the distribution of X based on the observations with indices in E_1, obtained by minimizing $M_n(\phi)$ with ϕ polynomial of degree at most m as in Section 3.

Then, considering now the noise as the unknown signal, use the methodology proposed in [4] to estimate the distribution of the noise: given

$$\tilde{\phi}_n^E(t_1, t_2) = \frac{1}{|E_2|} \sum_{i \in E_2} \exp\left\{ i t_1^\top Y_i^{(1)} + i t_2^\top Y_i^{(2)} \right\},$$

let

$$\widehat{\Phi}^{\varepsilon}_{(m,\nu_{\text{est}}),1}(t_1) = \frac{\tilde{\phi}^E_n(t_1,0)}{\widehat{\Phi}^X_{(m,\nu_{\text{est}})}(t_1)} \mathbf{1}_{|\widehat{\Phi}^X_{(m,\nu_{\text{est}})}(t_1)| \geq |E_2|^{-1/2}}$$

$$\text{and} \quad \widehat{\Phi}^{\varepsilon}_{(m,\nu_{\text{est}}),2}(t_2) = \frac{\tilde{\phi}^E_n(0,t_2)}{\widehat{\Phi}^X_{(m,\nu_{\text{est}})}(t_2)} \mathbf{1}_{|\widehat{\Phi}^X_{(m,\nu_{\text{est}})}(t_2)| \geq |E_2|^{-1/2}}.$$

For each $h > 0$, compute the estimator of the density of $(Y^{(1)}, Y^{(2)})$

$$\widehat{p}_{(m,\nu_{\text{est}},h)}(y_1,y_2) = 0 \vee \frac{1}{(2\pi)^{2d}} \int_{[-h,h]^d \times [-h,h]^d} \exp(-it_1^\top y_1 - it_2^\top y_2) \widehat{\Phi}^X_{(m,\nu_{\text{est}})}(t_1+t_2) \widehat{\Phi}^{\varepsilon}_{(m,\nu_{\text{est}}),1}(t_1) \widehat{\Phi}^{\varepsilon}_{(m,\nu_{\text{est}}),2}(t_2) dt_1 dt_2 \quad (6.2)$$

and then compute the hold-out criterion

$$HO(m,\nu_{\text{est}},h) = \|\widehat{p}_{(m,\nu_{\text{est}},h)}\|_2^2 - \frac{2}{|T|} \sum_{i \in T} \widehat{p}_{(m,\nu_{\text{est}},h)}(Y_i^{(1)}, Y_i^{(2)}).$$

This is the empirical version of the L^2 distance $\|p - \widehat{p}\|_2^2 = \|p\|_2^2 + \|\widehat{p}\|_2^2 - 2\mathbb{E}_p[\widehat{p}]$, up to the unknown but constant term $\|p\|^2$. Finally, choose the parameters minimizing $(m,\nu_{\text{est}},h) \mapsto HO(m,\nu_{\text{est}},h)$.

Note that if not for taking $\widehat{p}_{(m,\nu_{\text{est}},h)}$ nonnegative, we could use that by Parseval's identity,

$$\|\widehat{p}_{(m,\nu_{\text{est}},h)}\|_2^2 = \|(t_1,t_2) \mapsto \widehat{\Phi}^X_{(m,\nu_{\text{est}})}(t_1+t_2) \widehat{\Phi}^{\varepsilon}_{(m,\nu_{\text{est}}),1}(t_1) \widehat{\Phi}^{\varepsilon}_{(m,\nu_{\text{est}}),2}(t_2)\|_{2,h}^2 \quad (6.3)$$

where $\|\cdot\|_{2,\nu}$ denotes the norm in $L^2([-\nu,\nu]^d)$. Computing the right hand term is less computationally demanding as it avoids the Fourier inversion.

Let us mention that oracle inequalities can be obtained for hold-out procedures in general settings, see for instance Chapter 8 of [18]. Although this goes beyond the scope of this article, it suggests that such a hold-out estimator could satisfy a result analog to Theorem 4.3 and be natively adaptive in the signal tail and regularity.

As an illustration, estimates of the characteristic function and density of the noise and the signal in one case are shown in Figure 5, with $|E_1| = |E_2| = 1000$ and $|T| = 333$. Following (6.3) and to perform faster simulations, we used the L^2 norm of the characteristic function of the observations instead of the L^2 norm of the positive density. When applying this method as is–without fine-tuning the threshold in the estimation of Φ^ε, taking other values of $|T|$ or computing the true L^2 norm of the estimated density of Y instead of using the approximation of (6.3)–, it tends to always select the more conservative parameters $(m,\nu_{\text{est}},h) = (4,7,1.5)$. This means the risk of the data-driven estimator is between 7.10^{-3} and 4.10^{-2} in all cases of Table 1: while it avoids degenerate choices of parameters, it does not yet perform comparably to the oracle.

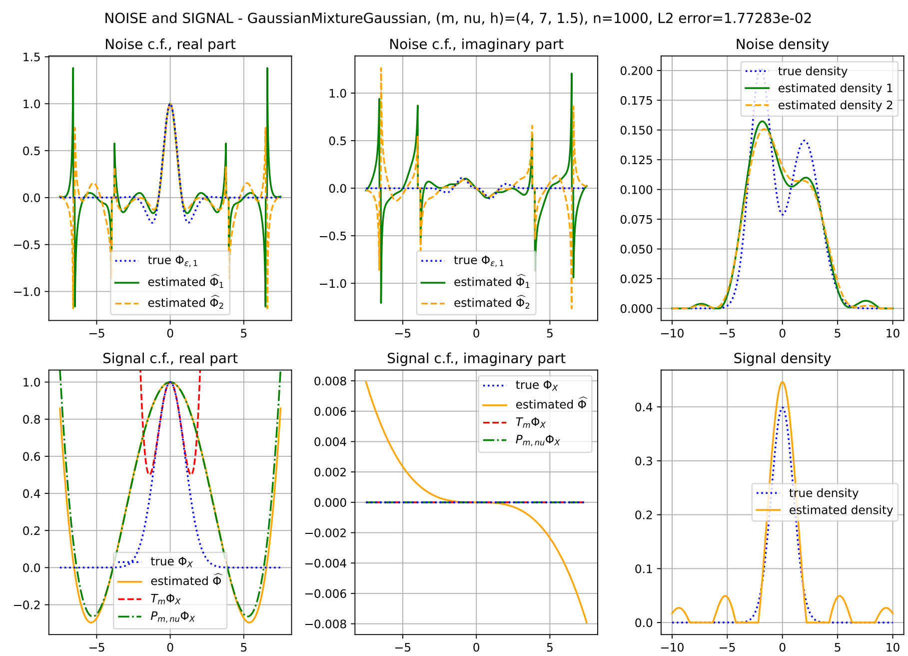

Fig 5: Top row: noises' estimated characteristic functions and densities obtained in the data-driven procedure. Bottom row: signal's estimated characteristic function and density, similar to Figures 1-4. Both rows use the parameters minimizing the hold-out criterion.

7. Discussion and perspectives

In this paper, we considered the repeated measurements model with two repetitions of the unknown signal, and we proposed a new estimation procedure for the distribution of the signal. Our estimator of the characteristic function of the signal is consistent in a neighborhood of 0 without any assumption on the noise distribution provided the noise has independent components and finite first moment, and for signals having a Laplace transform at λ with growth at most $\exp(b\|\lambda\|^\rho)$ for some $\rho > 0$. We provided theoretical results about minimax rates and adaptation procedures and presented simulation experiments.

In the case where more repetitions are available, it is possible to extend our procedure to take the additional information into account. We have investigated the possibility of using the sum of elementary criteria built using only groups of two repetitions as criterion to estimate the characteristic function of the signal. With this approach, taking $p > 2$ repetitions does not improve the rates of convergence of the estimators, only the constants. What could change the rates would be to take $p_n \to +\infty$ together with n; depending on the relationship between n and p_n, we expect a transition to occur between our regime (which has logarithmic rates in n) and the limit where p_n is considerably larger than n ($p_n \gg n^\alpha$ for some large enough $\alpha > 0$), where the signal can be recovered by simple average and the problem devolves into standard density estimation (which has polynomial rates in n).

One important result of our findings is the surprisingly good behaviour of our estimator in simulations. These findings open up new avenues of research, based on the following questions:

- Is it possible to extend the identifiability result to nonparametric classes of distributions having heavier tails?
- Is it possible to adapt the choice of (m, ν_{est}, h) on classes of distributions with known decay of the characteristic function for the noise and / or the signal? (For instance, taking lower ordinary or supersmooth signals instead of only upper smooth as in Section 4.1.) In particular, slower decay should allow one to take larger values of ν_{est} and h. The question would be to prove better convergence rates on these smaller classes of distributions when using oracle parameters.
- As explained in Section 6.4, when an oracle inequality holds, the data-driven hold-out procedure to select (m, ν_{est}, h) allows to mimic the oracle parameters in terms of risk. It would be interesting to prove an oracle inequality for our data-driven selection method, and investigate how to calibrate it in practice to recover similar performances to the oracle estimators. Combined with the previous point, this could produce estimators that are adaptive on the smaller classes of regularity.

8. Proofs

8.1. Proof of Theorem 2.1

In the following, we will write ϕ_i (resp. $\tilde{\phi}_i$) the characteristic function of $\varepsilon^(i)$ (resp $\tilde{\varepsilon}(i)$) for $i \in \{1, 2\}$. Since the distribution of Y is the same under $\mathbb{P}_{r(X)} * \mathbb{Q}$ and $\mathbb{P}_{r(X')} * \tilde{\mathbb{Q}}$, for any $(t_1, t_2) \in \mathbb{R}^d \times \mathbb{R}^d$,

$$\Phi_X(t_1 + t_2)\phi_1(t_1)\phi_2(t_2) = \Phi_{X'}(t_1 + t_2)\tilde{\phi}_1(t_1)\tilde{\phi}_2(t_2), \tag{8.1}$$

and by taking $t_2 = 0$, and $t_1 = 0$ in (8.1), also

$$\Phi_X(t_1)\phi_1(t_1) = \Phi_{X'}(t_1)\tilde{\phi}_1(t_1), \tag{8.2}$$

and

$$\Phi_X(t_2)\phi_2(t_2) = \Phi_{X'}(t_2)\tilde{\phi}_2(t_2). \tag{8.3}$$

There exists a neighborhood V of 0 in $\mathbb{R}^d \times \mathbb{R}^d$ such that for all $(t_1, t_2) \in \mathbb{R}^d \times \mathbb{R}^d$, $\phi_1(t_1) \neq 0$, $\tilde{\phi}_1(t_1) \neq 0$, $\phi_2(t_2) \neq 0$ and $\tilde{\phi}_2(t_2) \neq 0$, so that (8.1), (8.2) and (8.3) imply that for any $(t_1, t_2) \in V$,

$$\Phi_X(t_1 + t_2)\Phi_{X'}(t_1)\Phi_{X'}(t_2) = \Phi_{X'}(t_1 + t_2)\Phi_X(t_1)\Phi_X(t_2).$$

Since the function $(z_1, z_2) \in \mathbb{C}^d \times \mathbb{C}^d \mapsto \Phi_X(z_1+z_2)\Phi_{X'}(z_1)\Phi_{X'}(z_2) - \Phi_{X'}(z_1+z_2)\Phi_X(z_1)\Phi_X(z_2)$ is a multivariate analytic function which is zero on a purely real neighborhood of 0, then it is the null function on the whole multivariate complex space, see Lemma 25 in [8], so that, for any z_1, z_2 in $\mathbb{C}^d$,

$$\Phi_X(z_1+z_2)\Phi_{X'}(z_1)\Phi_{X'}(z_2) = \Phi_{X'}(z_1+z_2)\Phi_X(z_1)\Phi_X(z_2). \tag{8.4}$$

Taking $z_2 = -z_1$, since $\Phi_X(0) = 1$ and $\Phi_X(-z_1) = \overline{\Phi_X(z_1)}$, we get that, for all $z \in \mathbb{C}^d$,

$$|\Phi_{X'}(z)| = |\Phi_X(z)|. \tag{8.5}$$

We set $R(z) = |\Phi_{X'}(z)| = |\Phi_X(z)|$ and define $\Theta(z) \in (-\pi, \pi]$ and $\tilde{\Theta}(z) \in (-\pi, \pi]$ such that $\Phi_X(z) = R(z)\exp(i\Theta(z))$ and $\Phi_{X'}(z) = R(z)\exp(i\tilde{\Theta}(z))$. The functions R, θ and $\tilde{\Theta}$ are continuous on the open set where $R(z) \neq 0$, which includes a neighborhood of 0 since $R(0) = 1$. Note that $\Theta(0) = 0$ and $\tilde{\Theta}(0) = 0$. Thus there exists δ such that $0 < \delta < \pi/6$, such that there there exist a neighborhood A_δ of 0 in $\mathbb{C}^d$, such that for all $(z_1, z_2) \in A_\delta^2$ and all $z \in A_\delta$, $R(z \neq 0$, $R(z_1+z_2) \neq 0$, and also $\Theta(z) \in (-\delta, \delta)$, $\tilde{\Theta}(z) \in (-\delta, \delta)$, $\Theta(z_1+z_2) \in (-\delta, \delta)$, and $\tilde{\Theta}(z_1+z_2) \in (-\delta, \delta)$. Using equations (8.4) and (8.5), we get that for all $z_1 \in A_\delta$ and $z_2 \in A_\delta$,

$$\exp\{i(\Theta(z_1+z_2) + \tilde{\Theta}(z_1) + \tilde{\Theta}(z_2))\} = \exp\{i(\tilde{\Theta}(z_1+z_2) + \Theta(z_1) + \Theta(z_2))\},$$

which gives

$$\exp\{i(\Theta(z_1+z_2) + \tilde{\Theta}(z_1) + \tilde{\Theta}(z_2)) - i(\tilde{\Theta}(z_1+z_2) + \Theta(z_1) + \Theta(z_2))\} = 1. \tag{8.6}$$

But since for all $z_1 \in A_\delta$ and $z_2 \in A_\delta$,

$$-6\varepsilon < \Theta(z_1+z_2) - \tilde{\Theta}(z_1+z_2) + \tilde{\Theta}(z_1) - \Theta(z_1) + \tilde{\Theta}(z_2) - \Theta(z_2) < 6\varepsilon,$$

(8.6) implies that for all $z_1 \in A_\delta$ and $z_2 \in A_\delta$,

$$\Theta(z_1+z_2) - \tilde{\Theta}(z_1+z_2) = \Theta(z_1) - \tilde{\Theta}(z_1) + \Theta(z_2) - \tilde{\Theta}(z_2).$$

Now using Theorem 2 in [1] we get that there exists $\alpha \in \mathbb{R}^d$ such that for all $z \in A_\delta \cap \mathbb{R}^d$, $\Theta(z) = \tilde{\Theta}(z) + \alpha^\top z$. Thus, for all $z \in A_\delta \cap \mathbb{R}^d$, $\Phi_X(z) - \Phi_{X'}(z)\exp(i\alpha^\top z) = 0$. Since the function $z \in \mathbb{C}^d \mapsto \Phi_X(z) - \Phi_{X'}(z)\exp(i\alpha^\top z)$ is a multivariate analytic function of d variables which is zero on a purely real neighborhood of 0, then it is the null function on the whole multivariate complex space, that is,

$$\forall z \in \mathbb{C}^d,\ \Phi_X(z) = \Phi_{X'}(z)\exp(i\alpha^\top z) = \Phi_{X'+\alpha}(z)$$

which ends the proof.

8.2. Proof of Proposition 3.1 and Proposition 4.1

For any $\nu > 0$, denote $\|\cdot\|_{2,\nu}$ the norm in $L^2([-\nu,\nu]^d)$ and $\|\cdot\|_{\infty,\nu}$ the norm in $L^\infty([-\nu,\nu]^d)$.

We first prove Proposition 3.1. For any ϕ and Φ_X in $\Upsilon_{\rho,S}$,

$$|M_n(\phi) - M(\phi;\nu_{\text{est}}|\Phi_X)| \leq C\frac{\|Z_n\|_{\infty,\nu_{\text{est}}}}{\sqrt{n}}$$

for a constant C depending only on ν_{est}, ρ and S, and where $Z_n(t) = \sqrt{n}(\tilde{\phi}_n(t) - \Phi_X(t)\Phi_{\varepsilon^{(1)}}(t_1)\Phi_{\varepsilon^{(2)}}(t_2))$. This leads as usual to

$$\begin{aligned} M(\widehat{\Phi}_{n,\rho};\nu_{\text{est}}|\Phi_X) &\leq M_n(\widehat{\Phi}_{n,\rho}) + C\frac{\|Z_n\|_{\infty,\nu_{\text{est}}}}{\sqrt{n}} \\ &\leq M_n(T_{m_n}\Phi_X) + C\frac{\|Z_n\|_{\infty,\nu_{\text{est}}}}{\sqrt{n}} \\ &\leq M(T_{m_n}\Phi_X;\nu_{\text{est}}|\Phi_X) + 2C\frac{\|Z_n\|_{\infty,\nu_{\text{est}}}}{\sqrt{n}} \\ &\leq 2C\frac{\|Z_n\|_{\infty,\nu_{\text{est}}}}{\sqrt{n}} + 6\|T_{m_n}\Phi_X - \Phi_X\|_{\infty,\nu_{\text{est}}} + \frac{1}{n}, \end{aligned} \tag{8.7}$$

since $M(\Phi_X;\nu_{\text{est}}|\Phi_X) = 0$ and using Lipschitz properties of M.

Note that $(\exp\{it^T\cdot\})_{t\in[-\nu_{\text{est}},\nu_{\text{est}}]^d}$ is a Glivenko-Cantelli class as soon as ε has a finite first moment (recall that all moments of the signal are finite thanks to **(H1)**), so that $\frac{\|Z_n\|_{\infty,\nu_{\text{est}}}}{\sqrt{n}} = o_{\mathbb{P}}(1)$. Proposition 3.1 follows then from (8.13).

Let us now prove Proposition 4.1 To get convergence rates in $L^2([-\nu,\nu]^d)$ for estimators minimizing an empirical criterion M_n converging to a criterion M, the usual way is first to prove consistency, in order to focus on a local analysis of the criterion M. This is the aim of Proposition 3.1. The key point is then to relate M to the error in $L^2([-\nu,\nu]^d)$ of the estimator of Φ_X. To obtain the key local relationship between $M(\phi;\nu|\Phi_X)$ and $\|\phi-\Phi_X\|_{2,\nu}$, we follow the same path as in [8].

1. Define M^{lin} the part of M which is a quadratic function as a function of $\phi - \Phi_X$. When restricted to functions in $\Upsilon_{\rho,S} \cap \mathbb{C}_m[X_1,\ldots,X_d]$, prove a lower bound relating $M^{\text{lin}}(T_m\phi;T_m\Phi_X)$ to $\|T_m(\phi-\Phi_X)\|_{2,\nu}$ for any integer m.
2. Using the fact that any ϕ in $\Upsilon_{\rho,S}$ may be approximated using $T_m\phi$ with an error decreasing very fast in m, choose m so that the lower bound becomes true for all ϕ in $\Upsilon_{\rho,S}$ as soon as $\|\phi-\Phi_X\|_{2,\nu}$ is small enough.

In [8], there is a problem in the proof of 1., see [10]. However, the main stream of the proof holds true, so that we detail now the proof of Proposition 4.1, pointing where we follow [8] and where the differences are. To ease the comparison with [8] we introduce similar notations. For any ϕ and Φ_X in $\Upsilon_{\rho,S}$, define $\Delta\phi = \phi - \Phi_X$, and $N : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{C}$ such that for all $(t_1, t_2) \in \mathbb{R}^d \times \mathbb{R}^d$,

$$N(t_1, t_2; \Delta\phi, \Phi_X) = \Delta\phi(t_1 + t_2)\Phi_X(t_1)\Phi_X(t_2) - \Phi_X(t_1 + t_2)\Delta\phi(t_1)\Phi_X(t_2) \\ - \Phi_X(t_1 + t_2)\Phi_X(t_1)\Delta\phi(t_2),$$

and $M^{\text{lin}}(\Delta\phi, \Phi_X; \nu) = \int_{[-\nu,\nu]^{2d}} |N(t_1, t_2; \Delta\phi, \Phi_X)|^2 \, dt_1 dt_2$.

Fix any $\rho_0 \geq 1$, any $\nu \in (0, \nu_{\text{est}} \wedge 1/2Sd^2)$, any $S, c(\nu), E > 0$. The following lemmas are key steps for the proof of Proposition 4.1.

Lemma 8.1. *There exists a constant c_1 depending only on ρ_0, ν, and S such that for all $\rho \leq \rho_0$, for all ϕ and Φ_X in $\Upsilon_{\rho,S}$, for all integer m,*

$$M^{lin}(T_m\Delta\phi, T_m\Phi_X; \nu) \geq c_1^m \int_{[-\nu,\nu]^d} |N(t, t; T_m\Delta\phi, T_m\Phi_X)|^2 \, dt.$$

The proof of Lemma 8.1 is given in Section 8.4. The idea is that because of the repeated observations setting, useful information in the N function can be found on the diagonal. Lemma 8.1 allows, in the context of repeated observations, to apply the minoration of the quadratic form of point 1. with the strategy of [8]. This is done in the following lemma.

Lemma 8.2. *There exists a constant c_2 depending only on ρ_0, ν, and S such that for all $\rho \leq \rho_0$, for all ϕ and Φ_X in $\Upsilon_{\rho,S}$, for all integer m,*

$$\int_{[-\nu,\nu]^d} |N(t, t; T_m\Delta\phi, T_m\Phi_X)|^2 \, dt \geq c_2^m \|T_m\Delta\phi\|_{2,\nu}^2.$$

The proof of Lemma 8.2 is given in Section 8.5.

Using Lemma 8.1 and Lemma 8.2 we get that for all $\rho \leq \rho_0$, for all ϕ and Φ_X in $\Upsilon_{\rho,S}$, for all integer m,

$$M^{\text{lin}}(T_m\Delta\phi, T_m\Phi_X; \nu) \geq (c_1c_2)^m \|T_m\Delta\phi\|_{2,\nu}^2.$$

Now, using Lemma H.3 in [9] and easy computations to upper bound the difference $|M^{\text{lin}}(\Delta\phi, \Phi_X; \nu) - M^{\text{lin}}(T_m\Delta\phi, T_m\Phi_X; \nu)|$ we get that there exists constants c_3 and c_4 depending only on ρ_0, ν, and S such that for all $\rho \leq \rho_0$, for all ϕ and Φ_X in $\Upsilon_{\rho,S}$, for all integer m,

$$M^{\text{lin}}(\Delta\phi, \Phi_X; \nu) \geq c_3^m \|\Delta\phi\|_{2,\nu}^2 - c_4^m m^{-2m/\rho}. \tag{8.8}$$

Now, using (8.8), (8.12) in Section 8.3 and following the proof of Proposition A.2 in [8] we get the following lemma.

Lemma 8.3. *For any $\delta > 0$, there exists $\eta > 0$ and a constant c_5 depending only on δ, ρ_0, ν, and S such that for all $\rho \leq \rho_0$, for all ϕ and Φ_X in $\Upsilon_{\rho,S}$, as soon as $\|\phi - \Phi_X\|_{2,\nu} \leq \eta$,*

$$M(\phi; \nu|\Phi_X) \geq c_5\|\phi - \Phi_X\|_{2,\nu}^{2+2\delta}.$$

The end of the proof of Proposition 4.1 follows now, let us explain how. Recall that for all S, all $\rho \leq \rho_0$, all $\Phi_X \in \Upsilon_{\rho,S}$ and $\rho' \in [\rho, \rho_0]$,

$$M(\widehat{\Phi}_{n,\rho'}; \nu|\Phi_X) \leq 2C\frac{\|Z_n\|_{\infty,\nu_{\text{est}}}}{\sqrt{n}} + \frac{C_{bias}}{n^{\frac{a}{\rho}}} + \frac{1}{n},$$

for a constant C_{bias} depending only on d, S, ν_{est} and ρ_0, using (8.7), (8.13) and $m_n \geq a\frac{\log(n)}{\log\log(n)}$. Concentration properties of the process Z_n are proved in Appendix G of [9]: there exists a numerical constant B depending only on d, ρ, S, ν, E, such that for all $n \geq 1$, all $x > 0$, all Φ_X in $\Upsilon_{\rho,S}$, all $\mathbb{Q}^{(2d)} \in \mathcal{Q}(\nu, c(\nu), E)$, with probability at least $1 - 4e^{-x}$,

$$\frac{\|Z_n\|_{\infty,\nu_{\text{est}}}}{\sqrt{n}} \leq B\left(\sqrt{\frac{x}{n}} \wedge \frac{x}{n}\right). \tag{8.9}$$

As a consequence, following Section A.1 in [8] we get that for all $\eta > 0$, there exists n_0 such that for all $\rho \in [1, \rho_0]$, for all $\Phi_X \in \Upsilon_{\rho,S}$, $\mathbb{Q} \in \mathcal{Q}^{(2d)}(\nu, c(\nu), E)$, for all $n \geq n_0$, with probability $1 - e^{-c_\eta n}$,

$$\forall \rho' \in [\rho, \rho_0],\ \|\widehat{\Phi}_{n,\rho'} - \Phi_X\|_{2,\nu} \leq \eta.$$

Now, using direct computations, Lipschitz properties of M_n, and norm comparisons in [9] Section J, similar to (8.12)), we get that there exists a universal constant c_M, and positive b and η_1 such that, if we write $\epsilon(u) = b/\log\log(1/u)$ for all $u > 0$, then for any $\rho' \in [1, \rho_0]$ and any $\phi \in \Upsilon_{\rho',S}$ such that $\|\phi - \Phi_X\|_{2,\nu} \leq \eta_1$,

$$\begin{aligned}(M_n(T_{m_n}\Phi_X) - M(T_{m_n}\Phi_X); \nu|\Phi_X)) - (M_n(\phi) - M(\phi; \nu|\Phi_X))\\ \leq c_M\frac{\|Z_n\|_{\infty,\nu_{\text{est}}}}{\sqrt{n}}\|\phi - \Phi_X\|_{2,\nu_{\text{est}}}^{1-\epsilon(\|\phi-\Phi_X\|_{2,\nu_{\text{est}}})},\end{aligned}$$

leading to: for all $\rho' \in [\rho, \rho_0]$,

$$M(\widehat{\Phi}_{n,\rho'}; \nu|\Phi_X) \leq c_M\frac{\|Z_n\|_{\infty,\nu_{\text{est}}}}{\sqrt{n}}\|\widehat{\Phi}_{n,\rho'} - \Phi_X\|_{2,\nu_{\text{est}}}^{1-\epsilon(\|\widehat{\Phi}_{n,\rho'}-\Phi_X\|_{2,\nu_{\text{est}}})} + \frac{C_{bias}}{n^{\frac{a}{\rho}}} + \frac{2}{n}.$$

Using (25) in Section A.3 of [8], (8.13) and Lemma 8.3, then using Proposition 3.1 and (8.9), Proposition 4.1 follows.

8.3. Useful inequalities

We collect in this section results borrowed from [9]. For this purpose, for any $\rho \geq 1$ and $S > 0$, we introduce $\Upsilon_{\rho,S,2}$ the subset of multivariate analytic functions from $\mathbb{C}^{2d}$ to $\mathbb{C}$ defined as follows:

$$\Upsilon_{\rho,S,2} = \Bigg\{\phi \text{ analytic s.t. } \forall z \in \mathbb{R}^{2d}, \overline{\phi(z)} = \phi(-z), \phi(0) = 1 \\ \text{and } \forall j \in \mathbb{N}^{2d} \setminus \{0\}, \left|\frac{\partial^j \phi(0)}{j!}\right| \leq \frac{S^{\|j\|_1}}{(\|j\|_1)^{\|j\|_1/\rho}}\Bigg\}.$$

We first state norm comparisons that can be deduced from [9]. Following Section J.3, page 19 line 4, we get that there exists a constant c_6 depending only on d, ρ, S, ν, such that for all integer m and all $R \in \Upsilon_{\rho,S,2} \cap \mathbb{C}_m[X_1, \ldots, X_d, X_{d+1}, \ldots, X_{2d}]$,

$$\int_{[-\nu,\nu]^d} |R(u,0)|^2 \, dt \leq c_6 m^d \int_{[-\nu,\nu]^{2d}} |R(u,v)|^2 \, dudv. \tag{8.10}$$

Define, for any $R \in \mathbb{C}_m[X_1, \ldots, X_d, X_{d+1}, \ldots, X_{2d}]$, $\|R\|_{\mathfrak{M}}$ as the euclidian norm of the polynomial coefficients of R. Then Lemmas I.3 and I.4 in [9] state that there exist c_7 and c_8 depending only on d, ρ, S, ν, such that for all $R \in \Upsilon_{\rho,S,2} \cap \mathbb{C}_m[X_1, \ldots, X_d, X_{d+1}, \ldots, X_{2d}]$,

$$c_7^m \|R\|_{L^2([-\nu,\nu]^{2d})} \leq \|R\|_{\mathfrak{M}} \leq c_8^m \|R\|_{L^2([-\nu,\nu]^{2d})}. \tag{8.11}$$

Also, following the proof of Proposition B.5 in [8], we get that there exists constants b, c_9 and η_2 depending only on d, ρ, S, ν, such that for all $R \in \Upsilon_{\rho,S,2}$, as soon as $\|R\|_{L^2([-\nu,\nu]^{2d})} \leq \eta_2$,

$$\int_{[-\nu,\nu]^d} |R(u,0)|^2 \, du \vee \int_{[-\nu,\nu]^d} |R(0,v)|^2 \, dv \leq c_9 \|R\|_{L^2([-\nu,\nu]^{2d})}^{2-\epsilon_b(\|R\|_{L^2([-\nu,\nu]^{2d})})}, \tag{8.12}$$

where for all $u > 0$, $\epsilon_b(u) = b/\log\log(1/u)$.

We finally recall that thanks to Lemma H.3 in [9] there exists a constant C depending only on d such that for any S, ρ, any $\alpha > 0$, any integer m, any $\phi \in \Upsilon_{\rho,S}$,

$$\|\phi - T_m\phi\|_{\infty,\alpha} \leq C(S\alpha)^{m+\rho} m^{d-m/\rho} \exp\left\{\frac{(S\alpha)^\rho}{\rho}\right\}. \tag{8.13}$$

8.4. Proof of Lemma 8.1

Let ϕ and Φ_X in $\Upsilon_{\rho,S}$, and let m be an integer. Define the function R, for all $(u,v) \in \mathbb{R}^{2d}$ by $R(u,v) = N(u+v, u-v; T_m\Delta\phi, T_m\Phi_X)$, so that

$$\int_{[-\nu,\nu]^d} |R(u,0)|^2 \, du = \int_{[-\nu,\nu]^d} |N(t,t;T_m\Delta\phi, T_m\Phi_X)|^2 \, dt,$$

and

$$\begin{aligned}\int_{[-\nu,\nu]^{2d}} |R(u,v)|^2 \, dudv &= \int_{[-\nu,\nu]^d} |N(u+v, u-v; T_m\Delta\phi, T_m\Phi_X)|^2 \, dudv \\ &\leq \int_{[-2\nu,2\nu]^d} |N(t_1,t_2;T_m\Delta\phi, T_m\Phi_X)|^2 \, dt_1 dt_2.\end{aligned}$$

Now, using the fact that for any integers $i \leq j$, $\binom{j}{i} \leq 2^j$, we get that $R \in \Upsilon_{\rho,4S,2} \cap \mathbb{C}_{3m}[X_1, \ldots, X_d, X_{d+1}, \ldots, X_{2d}]$ and $N(\cdot,\cdot;T_m\Delta\phi, T_m\Phi_X) \in \Upsilon_{\rho,4S,2} \cap$

$\mathbb{C}_{3m}[X_1, \dots, X_d, X_{d+1}, \dots, X_{2d}]$. We may thus use (8.10) to get that for some constant $c > 0$ depending only on d, ρ, S, ν,

$$\int_{[-\nu,\nu]^d} |R(u,0)|^2 \, du \le c(3m)^d \int_{[-\nu,\nu]^{2d}} |R(u,v)|^2 \, dudv,$$

then by the second inequality in (8.11), for some constant $C > 0$ depending only on d, ρ, S, ν,

$$\int_{[-2\nu,2\nu]^d} |N(t_1,t_2;T_m\Delta\phi, T_m\Phi_X)|^2 \, dt_1 dt_2 \le C^{3m} \|N(\cdot,\cdot;T_m\Delta\phi, T_m\Phi_X)\|_{\mathfrak{M}},$$

and by the first inequality in (8.11), for some constant $\tilde{C} > 0$ depending only on d, ρ, S, ν,

$$\|N(\cdot,\cdot;T_m\Delta\phi, T_m\Phi_X)\|_{\mathfrak{M}} \le \tilde{C}^{3m} \int_{[-\nu,\nu]^d} |N(t_1,t_2;T_m\Delta\phi, T_m\Phi_X)|^2 \, dt_1 dt_2,$$

and Lemma 8.1 follows.

8.5. Proof of Lemma 8.2

Let A and D be such that $\Phi_X(t) = \sum_i A_i t^i$ and $\Delta\phi(t) = \sum_i D_i t^i$, so that $T_m\Phi_X(t) = \sum_{i:\|i\|_1 \le m} A_i t^i$ and $T_m\Delta\phi(t) = \sum_{i:\|i\|_1 \le m} D_i t^i$

$$\begin{aligned} N(t,t;T_m\Delta\phi, T_m\Phi_X) &= T_m\Delta\phi(2t) T_m\Phi_X(t)^2 - 2T_m\Delta\phi(t) T_m\Phi_X(2t) T_m\Phi_X(t) \\ &= T_m\Phi_X(t) \left[\sum_{i:\|i\|_1 \le m} D_i 2^{\|i\|_1} t^i \sum_{j:\|j\|_1 \le m} A_j t^j - 2 \sum_{i:\|i\|_1 \le m} D_i t^i \sum_{j:\|j\|_1 \le m} A_j 2^{\|j\|_1} t^j \right] \\ &= T_m\Phi_X(t) \left[\sum_{k:\|k\|_1 \le 2m} \left(\sum_{\substack{0 \preccurlyeq u \preccurlyeq k, \\ \|u\|_1 \le m, \|k-u\|_1 \le m}} D_u \underbrace{(2^{\|u\|_1} - 2^{\|k-u\|_1 + 1}) A_{k-u}}_{B_{k,u}} \right) t^k \right] \end{aligned}$$

(where $u \preccurlyeq v$ is component-wise inequality, that is $u_a \le v_a$ for all a.) When $\nu \le 1/2Sd^2$, for all $t \in [-\nu,\nu]^d$, $|T_m\Phi_X(t)] \ge 1/2$, so that

$$\int_{[-\nu,\nu]^d} |N(t,t;T_m\Delta\phi, T_m\Phi_X)|^2 \, dt \ge \frac{1}{2} \int_{[-\nu,\nu]^d} \left| \frac{N(t,t;T_m\Delta\phi, T_m\Phi_X)}{T_m\Phi_X(t)} \right|^2 dt.$$

Now, the matrix B is triangular inferior with diagonal coefficients $B_{k,k} = (2^{\|k\|_1} - 2)$ (since $A_0 = 1$ because Φ_R is the characteristic function of a probability measure), which are all non-zero except for all $B_{k,k}$ when $\|k\|_1 = 1$. However, $\nabla(\Delta\phi)(0) = 0$, which reduces the set of possible $\Delta\phi$ to a complement of the kernel of B. Using exactly the same reasoning as in Lemma I.5 of [9], we get that there exists a constant $c_B > 0$ such that

$$\inf_{D \text{ of degree } m \text{ without degree 1 term}} \frac{\|BD\|_2}{\|D\|} \ge c_B^{2m},$$

that is

$$\frac{\|t \mapsto (N/\Phi_R)(t,t)\|_{\mathfrak{M}}}{\|\Delta\phi\|_{\mathfrak{M}}} \geq c_B^{2m}.$$

Combining with (8.11) we get Lemma 8.2.

8.6. *Proof of Theorem 4.2*

The whole proof follows the proof of Theorem 3.2 of [8], after noticing that all arguments go through with unbounded ρ_0. We recall the main arguments. For all $\Phi_X \in \Psi(\rho, S, \beta, c_\beta)$,

$$\|\widehat{f}_{n,\rho} - f\|_2^2 = \frac{1}{(4\pi^2)^d}\|T_{m_{n,\rho}}\widehat{\Phi}_{n,\rho} - \Phi_X\|_{2,h_{n,\rho}}^2 + \frac{1}{(4\pi^2)^d}\int_{\mathbb{R}^d\setminus[-h_{n,\rho},h_{n,\rho}]^d} |\Phi_X(u)|^2 du \tag{8.14}$$

Now,

$$\|T_{m_{n,\rho}}\widehat{\Phi}_{n,\rho} - \Phi_X\|_{2,h_{n,\rho}}^2 \leq 2\|T_{m_{n,\rho}}\widehat{\Phi}_{n,\rho} - T_{m_{n,\rho}}\Phi_X\|_{2,h_{n,\rho}}^2 + 2\|T_{m_{n,\rho}}\Phi_X - \Phi_X\|_{2,h_{n,\rho}}^2.$$

By (25) in Section A.3 of [8],

$$\|T_{m_{n,\rho}}\widehat{\Phi}_{n,\rho} - T_{m_{n,\rho}}\Phi_X\|_{2,h_{n,\rho}}^2 \leq m_{n,\rho}^d \left(2 + 2\frac{h_{n,\rho}}{\nu}\right)^{2m_{n,\rho}+d} \|T_{m_{n,\rho}}\widehat{\Phi}_{n,\rho} - T_{m_{n,\rho}}\Phi_X\|_{2,\nu}^2,$$

and

$$\begin{aligned}\|T_{m_{n,\rho}}\widehat{\Phi}_{n,\rho} - T_{m_{n,\rho}}\Phi_X\|_{2,\nu}^2 &\leq 2\|\widehat{\Phi}_{n,\rho} - \Phi_X\|_{2,\nu}^2 + 4\|T_{m_{n,\rho}}\widehat{\Phi}_{n,\rho} - \widehat{\Phi}_{n,\rho}\|_{2,\nu}^2 \\ &\quad + 4\|T_{m_{n,\rho}}\Phi_X - \Phi_X\|_{2,\nu}^2.\end{aligned}$$

Using (8.13) we arrive at

$$\begin{aligned}&\|T_{m_{n,\rho}}\widehat{\Phi}_{n,\rho} - \Phi_X\|_{2,h_{n,\rho}}^2 \\ &\leq 2C^2 h_{n,\rho}^d (Sh_{n,\rho})^{2m_{n,\rho}+2\rho} m_{n,\rho}^{2(d-m_{n,\rho}/\rho)} \exp\left(2\frac{(Sh_{n,\rho})^\rho}{\rho}\right) \\ &\quad + 2m_{n,\rho}^d \left(\frac{3h_{n,\rho}}{\nu}\right)^{2m_{n,\rho}+d} \Big[2\|\widehat{\Phi}_{n,\rho} - \Phi_X\|_{2,\nu}^2 \\ &\qquad + 8C^2\nu^d (S\nu)^{2m_{n,\rho}+2\rho} m_{n,\rho}^{2(d-m_{n,\rho}/\rho)} \exp\left(2\frac{(S\nu)^\rho}{\rho}\right)\Big],\end{aligned}$$

which altogether and with the definition of $h_{n,\rho}$ and $m_{n,\rho}$ shows that for some constant C, if $c_h < 1/3\sqrt{e}$,

$$\begin{aligned}\|\widehat{f}_{n,\rho} - f\|_2^2 \leq C\max\Bigg\{&\exp(-m_{n,\rho})\left(1 + \sqrt{n}\|\widehat{\Phi}_{n,\rho} - \Phi_X\|_{2,\nu}^2\right), \\ &\int_{\mathbb{R}^d\setminus[-h_{n,\rho},h_{n,\rho}]^d} |\Phi_X(u)|^2 du\Bigg\}.\end{aligned} \tag{8.15}$$

Now, for any $\Phi_X \in \Psi(\rho, S, \beta, c_\beta)$,

$$\int_{\mathbb{R}^d \setminus [-h_{n,\rho}, h_{n,\rho}]^d} |\Phi_X(u)|^2 du \leq \frac{c_\beta}{(1 + h_{n,\rho}^2)^\beta},$$

so that using Proposition 4.1, there exists $C' > 0$ such that with large probability,

$$\|\widehat{f}_{n,\rho} - f\|_2^2 \leq C' m_{n,\rho}^{-2\beta/\rho},$$

and for any $\Phi_X \in \Gamma(\rho, S, \alpha, \gamma, c_{\alpha,\gamma})$,

$$\int_{\mathbb{R}^d \setminus [-h_{n,\rho}, h_{n,\rho}]^d} |\Phi_X(u)|^2 du \leq c_\gamma \exp(-2\gamma h_{n,\rho}^\alpha),$$

so that using Proposition 4.1 with large probability

$$\|\widehat{f}_{n,\rho} - f\|_2^2 \leq C \exp\left(-\left[m_{n,\rho} \wedge \left(2\gamma \left(\frac{c_h}{S}\right)^\alpha m_{n,\rho}^{\alpha/\rho}\right)\right]\right).$$

Then using Proposition 4.1, Theorem 4.2 follows using the values for $m_{n,\rho}$ and $h_{n,\rho}$ in the theorem.

8.7. *Proof of Theorem 4.3*

This theorem relies on the following uniform control of the (non-adaptive) estimators.

Proposition 8.4. *For all $\nu \in (0, \nu_{est}]$, $c(\nu) > 0$, $E > 0$, $S > 0$, $\beta > 0$ and $c_\beta > 0$, there exists n_0 and $c_\sigma > 0$ such that for all $n \geq n_0$,*

$$\sup_{\rho \in [1,\rho_0]} \inf_{\substack{\mathbb{P}_X \in \Psi(\rho,S,\beta,c_\beta) \\ \mathbb{Q} \in \mathcal{Q}^{(2d)}(\nu,c(\nu),E)}} \mathbb{P}_{(\mathbb{P}_{r(X)} * \mathbb{Q})^{\otimes n}} \left(\sup_{\rho' \in [\rho,\rho_0]} \left\{ \left(\frac{\log(n)}{\log\log(n)} \right)^{\frac{2\beta}{\rho'}} \|\widehat{f}_{n,\rho'} - f\|_2^2 \right\} \leq c_\sigma \right) \geq 1 - \frac{1}{n}.$$

For all $\nu \in (0, \nu_{est}]$, $c(\nu) > 0$, $E > 0$, $S > 0$, $\alpha > 0$, $\gamma > 0$ and $c_{\alpha,\gamma} > 0$, there exists $c_\sigma > 0$ such that

$$\sup_{\rho \in [1,\rho_0]} \sup_{\substack{\mathbb{P}_X \in \Gamma(\rho,S,\alpha,\gamma,c_{\alpha,\gamma}) \\ \mathbb{Q} \in \mathcal{Q}^{(2d)}(\nu,c(\nu),E)}} \mathbb{P}_{(\mathbb{P}_{r(X)} * \mathbb{Q})^{\otimes n}} \left(\sup_{\rho' \in [\rho,\rho_0]} \left\{ \exp\left(c_{\exp} \left(\frac{\log(n)}{\log\log(n)} \right)^{1 \wedge \frac{\alpha}{\rho'}} \right) \times \|\widehat{f}_{n,\rho'} - f\|_2^2 \right\} \leq c_\sigma \right) \geq 1 - \frac{1}{n}.$$

The proof of Proposition 8.4 mirrors the proof of Theorem 4.2 with a uniform control of the errors, and is detailed in Section 8.8.

Fix $\nu \in (0, \nu_{\text{est}}]$, $c(\nu) > 0$, $E > 0$, $S > 0$, $\beta > 0$, $c_\beta > 0$, $\rho \in [1, \rho_0]$, $\mathbb{P}_X \in \Psi(\rho, S, \beta, c_\beta)$ and $\mathbb{Q} \in \mathcal{Q}^{(2d)}(\nu, c(\nu), E)$. Let c_σ be as in Proposition 8.4, let $\sigma_n^{OS}(\rho', \beta') = c_\sigma \left(\frac{\log\log(n)}{\log(n)}\right)^{\beta'/\rho'}$, and assume we are in the event of probability $1 - 1/n$ where $\|\widehat{f}_{n,\rho'} - f\|_2 \le \sigma_n^{OS}(\rho', \beta')$ for all $\rho' \in [\rho, \rho_0]$ and $\beta' \in (0, \beta]$.

First, let $\beta' \in (0, \beta]$, and let us control the error $\|\widetilde{f}_{n,\beta'} - f\|_2$. Note that

$$A_n(\rho', \beta') \le 0 \vee \sup_{\rho'' \in [\rho', \rho_0]} \{\|\widehat{f}_{n,\rho'} - f\|_2 + \|\widehat{f}_{n,\rho''} - f\|_2 - \sigma_n^{OS}(\rho'', \beta')\} \le \sigma_n^{OS}(\rho', \beta).$$

To control the error of $\widehat{f}_{n,\widehat{\rho}_{\beta'}}$, first use the triangle inequality, then the fact that by definition of A_n, for all $\rho', \rho'' \in [1, \rho_0]$ and $\beta' > 0$,

$$\|\widehat{f}_{n,\rho'} - \widehat{f}_{n,\rho''}\|_2 \le A_n(\rho' \wedge \rho'', \beta') + \sigma_n^{\text{OS}}(\rho' \vee \rho'', \beta'),$$

and lastly that by definition of $\widehat{\rho}_{\beta'}$,

$$A_n(\widehat{\rho}_{\beta'}, \beta') \vee \sigma_n^{\text{OS}}(\widehat{\rho}_{\beta'}, \beta') \le A_n(\widehat{\rho}_{\beta'}, \beta') + \sigma_n^{\text{OS}}(\widehat{\rho}_{\beta'}, \beta') \le A_n(\rho, \beta') + \sigma_n^{\text{OS}}(\rho, \beta'),$$

so that

$$\begin{aligned}
\|\widehat{f}_{n,\widehat{\rho}_{\beta'}} - f\|_2 &\le \|\widehat{f}_{n,\rho} - f\|_2 + \|\widehat{f}_{n,\widehat{\rho}_{\beta'}} - \widehat{f}_{n,\rho}\|_2 \\
&\le \sigma_n^{OS}(\rho, \beta) + \begin{cases} A_n(\rho, \beta') + \sigma_n^{OS}(\widehat{\rho}_{\beta'}, \beta') & \text{if } \widehat{\rho}_{\beta'} \ge \rho, \\ A_n(\widehat{\rho}_{\beta'}, \beta') + \sigma_n^{OS}(\rho, \beta') & \text{otherwise,} \end{cases} \\
&\le \sigma_n^{OS}(\rho, \beta) + \begin{cases} 2A_n(\rho, \beta') + \sigma_n^{OS}(\rho, \beta') & \text{if } \widehat{\rho}_{\beta'} \ge \rho, \\ A_n(\rho, \beta') + 2\sigma_n^{OS}(\rho, \beta') & \text{otherwise,} \end{cases} \\
&\le 4\sigma_n^{OS}(\rho, \beta') \quad \text{since } \beta' \mapsto \sigma_n^{OS}(\rho, \beta') \text{ is nonincreasing.}
\end{aligned} \tag{8.16}$$

Now, let us control the error on $\widehat{f}_n^{OS}$: on the same event, with the same reasoning,

$$\begin{aligned}
\|\widehat{f}_n^{OS} - f\|_2 &\le \|\widetilde{f}_\beta - f\|_2 + \|\widetilde{f}_{\widehat{\beta}} - \widetilde{f}_\beta\|_2 \\
&\le 4\sigma_n^{OS}(\rho, \beta) + \begin{cases} B_n(\beta) + 4\sigma_n^{OS}(\widehat{\rho}_{\widehat{\beta}}, \widehat{\beta}) & \text{if } \widehat{\beta} \le \beta, \\ B_n(\widehat{\beta}) + 4\sigma_n^{OS}(\widehat{\rho}_\beta, \beta) & \text{otherwise,} \end{cases} \\
&\le 4\sigma_n^{OS}(\rho, \beta) + \begin{cases} 2B_n(\beta) + 4\sigma_n^{OS}(\widehat{\rho}_\beta, \beta) & \text{if } \widehat{\beta} \le \beta, \\ B_n(\beta) + 8\sigma_n^{OS}(\widehat{\rho}_\beta, \beta) & \text{otherwise.} \end{cases}
\end{aligned}$$

Note that

$$\begin{aligned}
B_n(\beta) &\le 0 \vee \sup_{\beta' \in (0, \beta]} \{\|\widehat{f}_{n,\widehat{\rho}_\beta} - f\|_2 + \|\widehat{f}_{n,\widehat{\rho}_{\beta'}} - f\|_2 - 4\sigma_n^{OS}(\widehat{\rho}_{\beta'}, \beta')\} \\
&\le \|\widehat{f}_{n,\widehat{\rho}_\beta} - f\|_2 \le 4\sigma_n^{OS}(\rho, \beta) \quad \text{by (8.16)}
\end{aligned}$$

and that by definition of $\widehat{\rho}_\beta$,

$$\sigma_n^{OS}(\widehat{\rho}_\beta,\beta) \le A_n(\widehat{\rho}_\beta,\beta) + \sigma_n^{OS}(\widehat{\rho}_\beta,\beta) \le A_n(\rho,\beta) + \sigma_n^{OS}(\rho,\beta) \le 2\sigma_n^{OS}(\rho,\beta),$$

so that

$$\|\widehat{f}_n^{OS} - f\|_2 \le 24\,\sigma_n^{OS}(\rho,\beta).$$

In the supersmooth case with parameters ρ,γ,α, the proof above works as is when replacing $\sigma_n^{OS}(\rho',\beta')$ by $\sigma^{SS}(\rho',\beta',\alpha')$ for any $\rho' \in [1,\rho_0]$ and $\alpha',\beta' > 0$, showing that on the same event, for all $\alpha' \in (0,\alpha]$,

$$\|\overline{f}_{\alpha'} - f\|_2 \le 24\,\sigma_n^{SS}(\rho,\gamma,\alpha'),$$

and the same proof as the control of $\widehat{f}_n^{OS}$ leads to

$$\|\widehat{f}_n^{SS} - f\|_2 \le C\,\sigma_n^{SS}(\rho,\gamma,\alpha)$$

for some numerical constant C.

Finally, on the event of probability at most $1/n$ where at least one $\|\widehat{f}_{n,\rho'} - f\|_2$ is not upper bounded by $\sigma_n^{OS}(\rho',\beta)$ in the ordinary smooth case or $\sigma_n^{SS}(\rho',\alpha,\gamma)$ in the supersmooth case, we use that by construction, for all m, $T_m\widehat{\Phi}_{n,\rho'} \in \Upsilon_{\rho_0,S}$, which is bounded in $L^2([-h_{n,1},h_{n,1}]^d)$, and therefore all $\widehat{f}_{n,\rho'}$ as well as f are uniformly bounded in $L^2(\mathbb{R}^d)$. The error introduced by this event of probability at most $1/n$ is therefore $O(1/n)$, which is dominated by $\sigma_n^{OS}(\rho,\beta)$ and $\sigma_n^{SS}(\rho,\alpha,\gamma)$.

8.8. *Proof of Proposition 8.4*

With the notations of the Proposition, by (8.15), there exists constants $C > 0$ and n_0 such that for all $n \ge n_0$ and $\rho' \ge 1$,

$$\|\widehat{f}_{n,\rho'} - f\|_2^2 \le C\max\left\{\exp(-m_{n,\rho'})\left(1+\sqrt{n}\|\widehat{\Phi}_{n,\rho'} - \Phi_X\|_{2,\nu}^2\right),\right.$$
$$\left.\int_{\mathbb{R}^d\setminus[-h_{n,\rho'},h_{n,\rho'}]^d} |\Phi_X(u)|^2du\right\}.$$

By Proposition 4.1, taking $x = \log(n)$, we obtain that up to changing the constant C, there exists n_0 such that for $n \ge n_0$, with probability at least $1-n^{-1}$, for all $\rho' \in [\rho,\rho_0]$,

$$\|\widehat{f}_{n,\rho'} - f\|_2^2 \le C\max\left\{\exp(-m_{n,\rho'}), \int_{\mathbb{R}^d\setminus[-h_{n,\rho'},h_{n,\rho'}]^d} |\Phi_X(u)|^2du\right\}$$
$$\le C\begin{cases}\max\left(\exp(-m_{n,\rho'}), m_{n,\rho'}^{-2\beta/\rho'}\right)\\ \max\left(\exp(-m_{n,\rho'}), \exp(-2\gamma(c_h/S)^\alpha m_{n,\rho'}^{\alpha/\rho'})\right)\end{cases}$$

up to changing C, depending on the set of f considered. Therefore, given the choice of $m_{n,\rho'}$, up to changing C,

$$\sup_{\substack{\mathbb{P}_X \in \Psi(\rho,S,\beta,c_\beta)\\ \mathbb{Q}\in\mathcal{Q}^{(2d)}(\nu,c(\nu),E)}} \mathbb{P}_{(\mathbb{P}_{r(X)}*\mathbb{Q})^{\otimes n}}\left[\sup_{\rho'\in[\rho,\rho_0]}\left(\frac{\log(n)}{\log\log(n)}\right)^{2\beta/\rho'} \|\widehat{f}_{n,\rho'}-f\|_2^2 > C\right] \leq \frac{1}{n}$$

and

$$\sup_{\substack{\mathbb{P}_X \in \Gamma(\rho,S,\alpha,\gamma,c_{\alpha,\gamma})\\ \mathbb{Q}\in\mathcal{Q}^{(2d)}(\nu,c(\nu),E)}} \mathbb{P}_{(\mathbb{P}_{r(X)}*\mathbb{Q})^{\otimes n}}\left[\sup_{\rho'\in[\rho,\rho_0]}\exp\left(c_{\exp}\left(\frac{\log(n)}{\log\log(n)}\right)^{1\wedge\frac{\alpha}{\rho'}}\right)\right.$$
$$\left.\times\|\widehat{f}_{n,\rho'}-f\|_2^2 > C\right] \leq \frac{1}{n},$$

and the Proposition follows.

8.9. Proof of Theorem 4.4

The proof uses Le Cam's two-point method. For both points, we assume that the coordinates ε_j, $j = 1,\dots,2d$, are independent identically distributed with density

$$g : x \mapsto c_g \frac{1+\cos(cx)}{(\pi^2-(cx)^2)^2}$$

for some $c > 0$, where c_g is such that g is a probability density with characteristic function

$$\mathcal{F}[g] : t \mapsto \left[\left(1-\left|\frac{t}{c}\right|\right)\cos\left(\pi\frac{t}{c}\right)+\frac{1}{\pi}\sin\left(\pi\left|\frac{t}{c}\right|\right)\right]1_{[-c,c]}(t).$$

With an adequate choice of c, $Q \in \mathcal{Q}^{(2d)}(\nu, c(\nu), E)$. Let f_0 and f_n be the two different probability densities of X and $\mathbb{Q}$ the distribution of the noise ε. We write $\mathbb{P}_0$ and $\mathbb{P}_n$ the distribution of $r(X)$ for X having probability density f_0 and f_n. Then, the minimax risk is lower bounded by

$$\frac{1}{4}\|f_0-f_n\|^2_{L_2(\mathbb{R}^d)}\left[1-\frac{1}{2}\|(\mathbb{P}_0*\mathbb{Q})^{\otimes n}-(\mathbb{P}_n*\mathbb{Q})^{\otimes n}\|_{TV}\right], \tag{8.17}$$

with $\|\cdot\|_{TV}$ the total variation norm. The remaining of the proof follows the ideas in [19], proof of Theorem 2. As in [19], we introduce $u : x \in \mathbb{R} \mapsto c_u \exp(-1/(1-x^2))1_{(-1,1)}(x)$ with c_u such that u is a probability density, and for all integer $k \geq 0$, $R_k : x \mapsto \mathcal{L}_k(x)1_{(-1/2,1/2)}(x)$ where $\mathcal{L}_k$ is the Legendre polynomial of order k in $(-1,1)$. We now define the densities f_0 and f_n. For all $b > 0$ and $x \in \mathbb{R}$, define $u_b(x) = bu(bx)$. Fix some $b > 0$. For $x \in \mathbb{R}$ and sequences $(\alpha_n)_n$, $(b_n)_n$ and $(K_n)_n$, define $h_n(x) = u_b(x) + \alpha_n(R_{K_n} * u_{b_n})(x)$. Finally, for all $x \in \mathbb{R}^d$,

$$f_0(x) = \prod_{j=1}^{d} u_b(x_j) \text{ and } f_n(x) = h_n(x_1) \prod_{j=2}^{d} u_b(x_j).$$

As soon as $\alpha_n \to 0$ and $b_n \to +\infty$ when $n \to +\infty$, for large enough n, h_n is a positive probability density with compact support. Moreover, using equations (25) and (26) in [19] and noticing that the characteristic function of $\mathbb{P}_0$ (resp. $\mathbb{P}_n$) at point $t = (t_1, t_2) \in \mathbb{R}^d \times \mathbb{R}^d$ is the characteristic function of f_0 (resp. f_n) at point $t_1 + t_2$, we know that we can chose $b > 0$ such that $\mathbb{P}_1 \in \Psi(1, S, \beta, c_\beta)$ and $\mathbb{P}_n \in \Psi(1, S, \beta, c_\beta)$ as soon as

$$\alpha_n^2 b_n^{2\beta} \leq C K_n \tag{8.18}$$

for some $C > 0$. As it is used in [8], we have

$$1 - \frac{1}{2}\|(\mathbb{P}_0 * \mathbb{Q})^{\otimes n} - (\mathbb{P}_n * \mathbb{Q})^{\otimes n}\|_{TV} \geq \left(1 - \frac{1}{2}\|(\mathbb{P}_0 * \mathbb{Q}) - (\mathbb{P}_n * \mathbb{Q})\|_{TV}\right)^n.$$

Thus if $(\alpha_n)_n$, $(b_n)_n$ and $(K_n)_n$ are chosen such that

$$\|(\mathbb{P}_0 * \mathbb{Q}) - (\mathbb{P}_n * \mathbb{Q})\|_{TV} = O\left(\frac{1}{n}\right), \tag{8.19}$$

we get that the minimax risk is lower bounded by $c\|f_0 - f_n\|_{L_2(\mathbb{R}^d)}$ for some constant $c > 0$. Now,

$$\|f_0 - f_n\|_2^2 = \alpha_n^2 \|u_b\|_2^{2(d-1)} \|R_{K_n} * u_{b_n}\|_2^2.$$

It is proved in [19] that if for well chosen $c_1 > 0$,

$$b_n = c_1 K_n, \tag{8.20}$$

then

$$\|R_{K_n} * u_{b_n}\|_2^2 \geq \frac{c_3}{K_n}$$

for some $c_3 > 0$, so that

$$\|f_0 - f_n\|_2^2 \geq c_4 \frac{\alpha_n^2}{K_n} \tag{8.21}$$

for some $c_4 > 0$. Now, assume that (8.20) holds, $\alpha_n^2 = C K_n b_n^{-2\beta}$, and

$$K_n = c_2 \frac{\log(n)}{\log\log(n)} \tag{8.22}$$

for some $c_2 > 0$. Following closely the proof in [19] equations (29), (30), (31) and its upper bounding, we get that (8.19) holds. Theorem 4.4 follows then from (8.21).

Acknowledgments

The authors wants to thank the Associate Editor and a referee for their insightful comments and questions that helped improve the manuscript. Jeremie Capitao-Miniconi would like to acknowledge support from the UDOPIA-ANR-20-THIA-0013. Élisabeth Gassiat would like to acknowledge the Institut Universitaire de France and the ANR ASCAI: ANR-21-CE23-0035-02.

References

[1] Aczél, J. (1966). *Lectures on functional equations and their applications. Mathematics in Science and Engineering* **Vol. 19**. Academic Press, New York-London.

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

[3] Comte, F. and Kappus, J. (2015). Density deconvolution from repeated measurements without symmetry assumption on the errors. *J. Multivariate Anal.* **140** 31–46.

[4] Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73** 601–627.

[5] Comte, F., Samson, A. and Stirnemann, J. J. (2014). Deconvolution estimation of onset of pregnancy with replicate observations. *Scand. J. Stat.* **41** 325–345.

[6] Delaigle, A., Hall, P. and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.* **36** 665–685.

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

[8] Gassiat, E., Le Corff, S. and Lehéricy, L. (2022). Deconvolution with unknown noise distribution is possible for multivariate signals. *Ann. Statist.* **50** 303–323.

[9] Gassiat, E., Le Corff, S. and Lehéricy, L. (2022). Supplementary to "Deconvolution with unknown noise distribution is possible for multivariate signals". *Ann. Statist.*

[10] Gassiat, E., Lehéricy, L. and Le Corff, S. (2025). Erratum: Deconvolution with unknown noise distribution is possible for multivariate signals. preprint.

[11] Goldenshluger, A. and Lepski, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** 1150–1190.

[12] Johannes, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist.* **37** 2301–2323.

[13] Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution. *Electron. J. Stat.* **8** 2879–2904.

[14] Kotlarski, I. (1967). On characterizing the gamma and the normal distribution. *Pacific J. Math.* **20** 69–76.

[15] Kurisu, D. and Otsu, T. (2022). On linearization of nonparametric deconvolution estimators for repeated measurements model. *J. Multivariate Anal.* **189** Paper No. 104921, 19.

[16] Le Cam, L. (1986). *Asymptotic methods in statistical decision theory. Springer Series in Statistics*. Springer-Verlag, New York.

[17] Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivariate Anal.* **65** 139–165.

[18] Massart, P. (2007). *Concentration inequalities and model selection: Ecole d'Eté de Probabilités de Saint-Flour XXXIII-2003*. Springer.

[19] Meister, A. (2007). Deconvolving compactly supported densities. *Math. Methods Statist.* **16** 63–76.

[20] Meister, A. (2009). *Deconvolution problems in nonparametric statistics. Lecture Notes in Statistics* **193**. Springer-Verlag, Berlin.

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

[22] Neumann, M. H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.* **98** 1955–1968.

[23] Stefanski, L. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21** 169–184.

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